SigmaXL® Version 10.0 Workbook

Contact Information:

Technical Support: 1-866-475-2124 (Toll Free in North America) or 1-519-579-5877

Sales: 1-888-SigmaXL (888-744-6295)

E-mail: support@SigmaXL.com Web: www.SigmaXL.com

Published April, 2024

Table of Contents

SigmaXL® What's New in Version 10.0,
System Requirements and Getting Help

What's New in Version 10.0

See **SigmaXL Version 10.0 Feature List Summary** for a complete feature list summary. New features in SigmaXL Version 10.0 include:

- **Graphical Tool Selection Guide**

- **Hypothesis Test Selection Guide**

- **Revised Control Chart Selection Guide**

- **New Data Manipulation Tools**
  - Convert Raw Data to Frequency (Tally)
  - Convert Frequency to Raw Data

- **New Graphical Tools**
  - Heatmap
  - Interval Plots and Multiple X Interval Plots
  - Empirical/Normal CDF Plots
  - XYZ Contour/Surface Plot
    - Automatic Interpolation Method Selection and XY Standardization using Cross-Validation
    - Inverse Distance, Akima's Polynomial and Biharmonic Spline Interpolation

- **New MSA Templates**
  - GLM GageRR (Crossed) Metrics with Interaction
  - GLM GageRR (Crossed) Metrics without Interaction
  - GLM GageRR (Nested) Metrics
  - GLM GageRR (Expanded) Metrics
  - Orthogonal (Deming) Regression

- **New Statistical Tools**
  - Advanced Multiple Regression: Box-Tidwell Test and Power Transformation Recommendation for Continuous Predictors (New in Version 10.02).
  - Nonparametric Friedman and Friedman's Exact
  - Nonparametric Two Sample Kolmogorov-Smirnov (KS) Test and KS Exact (option in Mann-Whitney)
  - General Linear Model. Extends Advanced Multiple Regression to include:
    - Fixed and Random Factors
    - Nested Factors
    - Covariates (can be Nested)
    - For Random or Mixed Random/Fixed Factors with a balanced design, the ANOVA and Variance Components (VC) report is given based on

Expected Mean Squares. VC confidence intervals using Restricted Maximum Likelihood (REML) are included.

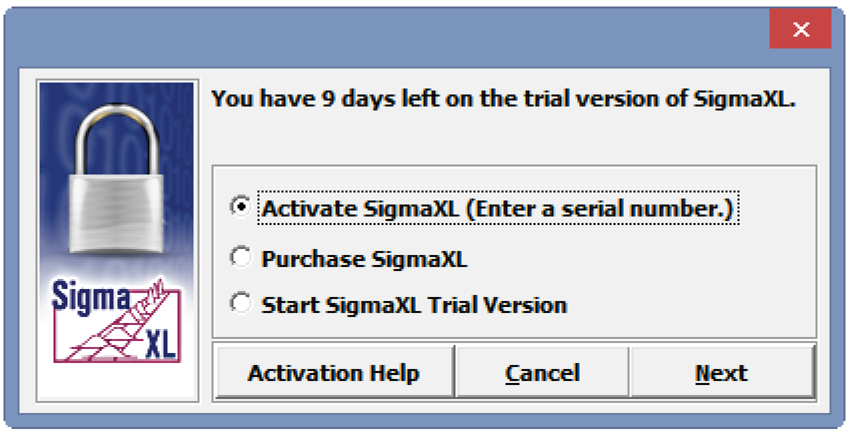
- If the design is unbalanced or model is non-hierarchical, REML is used to compute the VC values and confidence intervals. Fixed Effects Tests are based on Satterthwaite approximation degrees of freedom.
- Main Effects with Confidence Intervals and Interaction Plots of Fitted Means for Non-Nested Fixed Factors
- Tukey and Fisher Pairwise Comparison of Means for Non-Nested Fixed Factors
- Predicted Response Calculator
- Multiple Response Optimization for Nested or Non-Nested Fixed Factors

Activating SigmaXL

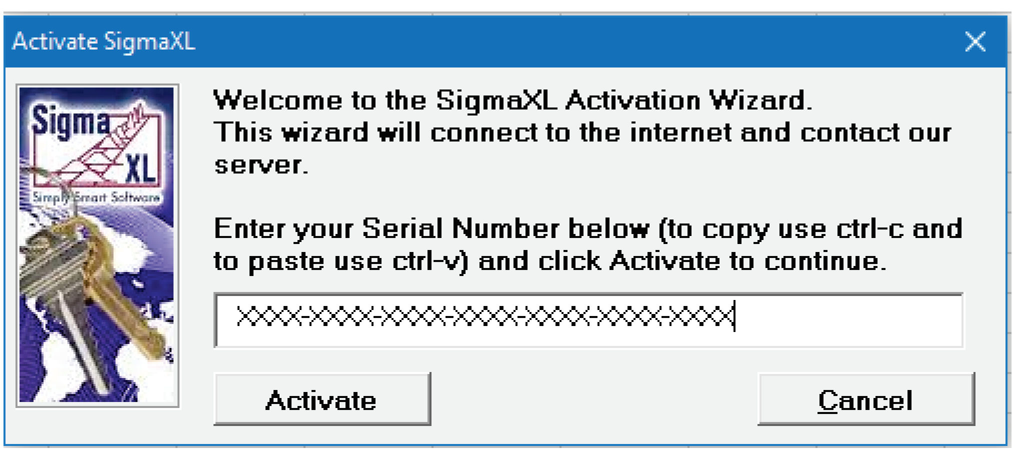
Please proceed with the following steps if you have a valid serial number and your computer is connected to the Internet. If you do not have an Internet connection, please email Support@SigmaXL.com from a different system or call 1-888-SigmaXL (1-888-744-6295) or 1-519-579-5877.

If your trial has timed out and you do not have a serial number but wish to purchase a SigmaXL license, please click **Purchase SigmaXL** in the **Activation Wizard Box** and this will take you to SigmaXL's order page http://www.sigmaxl.com/Order%20SigmaXL.shtml. You can also call 1-888-SigmaXL (1-888-744-6295) or 1-519-579-5877 to place an order.

1. In the **Activation Wizard** box select **Activate SigmaXL (Enter a serial number.)**

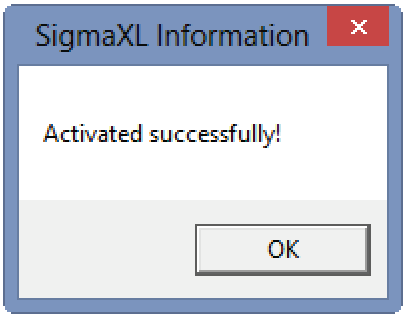


2. Click **Next**. Enter your serial number. We recommend that you copy and paste the serial number into the serial number field to avoid entry errors (note, Mac requires manual entry - copy/paste does not work):

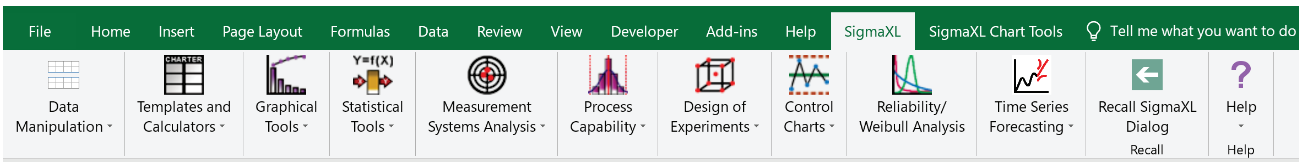


Tip: The number "1" is not used in a serial number.

3. Click **Next**. The activation process continues and is confirmed as shown:

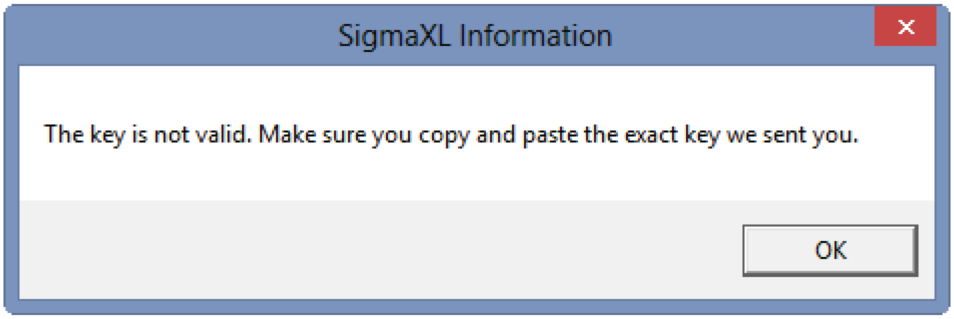


4. Click **OK**. SigmaXL appears as a new Ribbon in Excel:



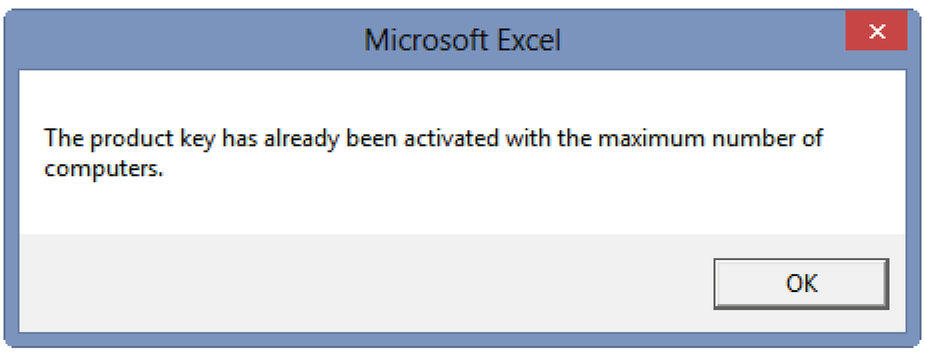
Error Messages

1. Incorrect Serial Number:



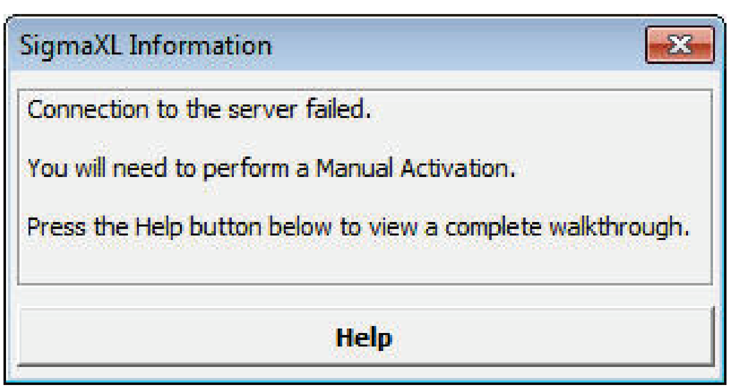
This is due to an incorrect serial number entry in the previous registration window – click **OK** and please reenter your serial number.

2. Serial Number Used in Previous Activation:



The license is currently activated on the maximum allowable number of computers. You will need to deactivate SigmaXL from your old computer while connected to the Internet. To deactivate a SigmaXL license go to **SigmaXL** > **Help** > **Deactivate SigmaXL**.

3. Internet Connection Problem:



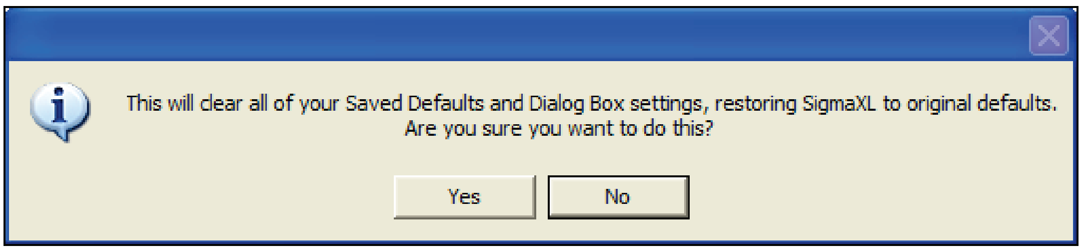
If you do not have an Internet connection, please email Support@SigmaXL.com from a different system or call 1-888-SigmaXL (1-888-744-6295) or 1-519-579-5877.

SigmaXL® Defaults and Menu Options

Clear Saved Defaults

Clear Saved Defaults will reset all saved defaults such as Pareto and Multi-Vari Chart settings, saved control limits, and dialog box settings. All settings are restored to the original installation defaults.

Click **SigmaXL > Help > SigmaXL Defaults > Clear Saved Defaults**. A warning message is given prior to clearing saved defaults.

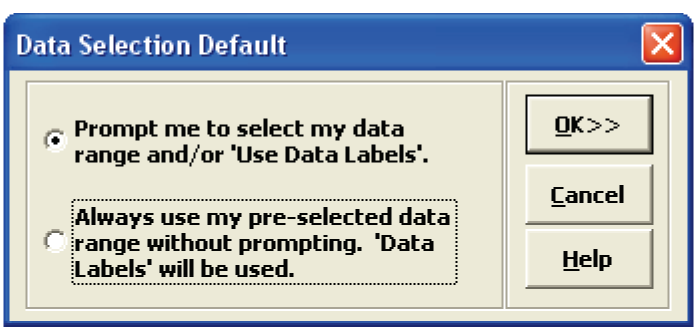


Data Selection Default

The **Data Selection Default** setting is: **Prompt me to select my data range and/or 'Use Data Labels'**.

This can be changed to: **Always use my pre-selected data range without prompting. 'Data Labels' will be used**. This setting saves you from having to click **Next** at the start of every function, but the user is responsible to ensure that the proper data selection is made prior to starting any menu item.

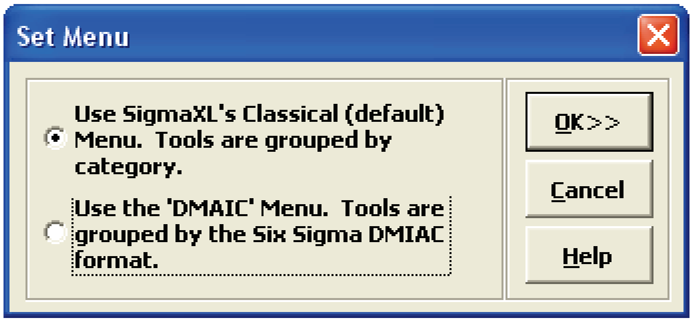
Click **SigmaXL > Help > SigmaXL Defaults > Data Selection Default** to make this change. This will apply permanently unless you revert back to the **Prompt me** setting or click **Clear Saved Defaults** shown above.



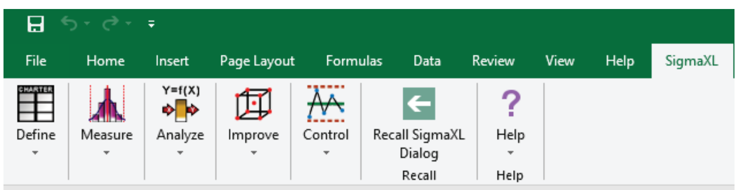
Menu Options (Classical or DMAIC)

The default SigmaXL menu system groups tools by category, but this can be changed to the Six Sigma DMAIC format.

Click **SigmaXL > Help > SigmaXL Defaults > Menu Options – Set SigmaXL's Menu to Classical** or **DMAIC**. The Set Menu dialog allows you to choose between **Classical** (default) and **DMAIC**:



If you select the **DMAIC** Menu, the **DMAIC** Menu Ribbon appears as shown:



All SigmaXL tools are available with this menu format, but they are categorized using the Six Sigma DMAIC phase format. Note that some tools will appear in more than one phase.

This workbook uses the classical (default) menu format, but the chapters are organized as Measure, Analyze, Improve and Control.

SigmaXL® System Requirements

Minimum System Requirements:

Computer and processor: 1.6 gigahertz (GHz) or faster, 2-core.

Memory: 4 GB of RAM or greater.

Hard disk: 1 GB of available hard-disk space.

Display: 1280x768 or higher resolution monitor.

Windows Operating System: Microsoft Windows 10 with current Service Packs, or later operating system.

Microsoft Windows Excel version: Excel 2016+, Office 365 with latest service packs installed.

Mac Operating System: Three most recent versions of macOS. Some systems may require the use of Parallels software.

Mac Excel Version: Excel 2019+, Office 365 with latest service packs installed.

Administrative rights to install software.

Getting Help and Product Registration

To access the help system, please click **SigmaXL > Help > Help**.

Technical support is available by phone at 1-866-475-2124 (toll-free in North America) or 1-519-579-5877 or by e-mail support@sigmaxl.com.

Please note that users obtain free technical support and upgrades for one year from date of purchase. Optional maintenance is available for purchase prior to the anniversary date.

To register by web, simply click **SigmaXL > Help > Register SigmaXL**.

Introduction to SigmaXL®
Data Format and Tools Summary

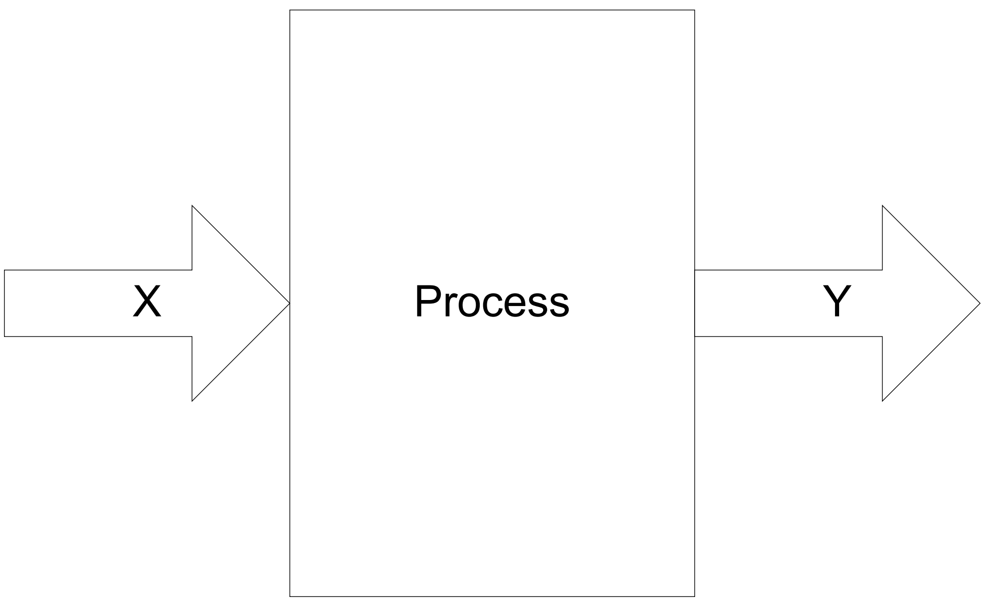
Introduction

SigmaXL is a powerful but easy to use Excel Add-In that will enable you to Measure, Analyze, Improve and Control your service, transactional, and manufacturing processes. This is the ideal cost-effective tool for Six Sigma Green Belts and Black Belts, Quality and Business Professionals, Engineers, and Managers.

SigmaXL will help you in your problem solving and process improvement efforts by enabling you to easily slice and dice your data, quickly separating the "vital few" factors from the "trivial many". This tool will also help you to identify and validate root causes and sources of variation, which then helps to ensure that you develop permanent corrective actions and/or improvements.

The Y=f(X) Model

SigmaXL utilizes the "Y=f(X)" model in its dialog boxes. Y denotes a key process output metric; X denotes a key process input metric. This process is shown pictorially as:



The mathematical expression Y = f(X) denotes that the variable Y is a function of X. Y can also be viewed as the effect of interest and X is the cause. For example, Y could be customer satisfaction as measured on a survey and X could be location or responsiveness to calls (also measured on a survey). The goal is to figure out which X's from among many possible are the key X's and to what extent do they impact the Y's of interest. Solutions and improvements then focus on those key X's.

Data Types: Continuous Versus Discrete

X and Y metrics can each be continuous or discrete. A continuous measure will have readings on a continuous scale where a mid-point has meaning. For example, in a customer satisfaction survey using a 1 to 5 score, the value 3.5 has meaning. Survey results may be averaged to obtain non-integer results. Other examples of continuous measures include cycle time, thickness, and weight. A discrete measure is categorical in nature. If we have Customer Types 1, 2, and 3, customer type 1.5 has no meaning. Other examples of discrete measures include defect counts and number of customer complaints.

It is possible to have various combinations of discrete/continuous X's and discrete/continuous Y's. Some examples are given below:

Examples of Discrete (Category) X and Discrete Y

- X = Customer Type, Y = Number of Complaints
- X = Product Type, Y = Number of Defects
- X = Day Shift vs. Night Shift, Y = Proportion of Defective Units

Examples of Discrete (Category) X and Continuous Y

- X = Customer Type, Y = Customer Satisfaction (1-5)
- X = Before Improvement vs. After Improvement, Y = Customer Satisfaction (1-5)
- X = Location, Y = Order to Delivery Time

Examples of Continuous X and Discrete Y

- X = Responsiveness to Calls (1-5), Y = Number of Complaints
- X = Process Temperature, Y = Number of Defects

Examples of Continuous X and Continuous Y

- X = Responsiveness to Calls (1-5), Y = Customer Satisfaction (1-5)
- X = Amount of Loan ($), Y = Cycle Time (Loan Application to Approval)

Note that in SigmaXL, a discrete X can be text or numeric, but a continuous X must be numeric. Y's must be numeric. If Y is discrete, count data will be required. If the data of interest is discrete text, it should be referenced as X1 and SigmaXL will automatically search through the text data to obtain a count (applicable for Pareto, Chi-Square and EZ-Pivot tools).

Stacked Data Column Format versus Unstacked Multiple Column Format

SigmaXL can accommodate two data formats: stacked column and unstacked multiple column. The stacked column format has an X column also referred to as the "Group Category" column and a Y column that contains the data of interest. The following is an example of data in stacked column format, with three unique groups of Customer Type:

Customer Type (X)	Overall Satisfaction (Y)
Type 2	3.5
Type 3	3.2
Type 3	3.3
Type 2	4.1
Type 1	3.2
Type 1	2.9
Type 1	1.9
Type 2	3.7
Type 3	4.0
Type 1	2.0
Type 3	2.6
Type 1	3.0
Type 2	4.1
Type 3	3.5
Type 2	5.0
Type 2	4.0
Type 3	4.4
Type 2	4.6
Type 1	2.5

If the data is in unstacked multiple column format, each unique group of X corresponds to a different column. The above data is now shown in unstacked format with customer satisfaction scores for each customer type in separate columns:

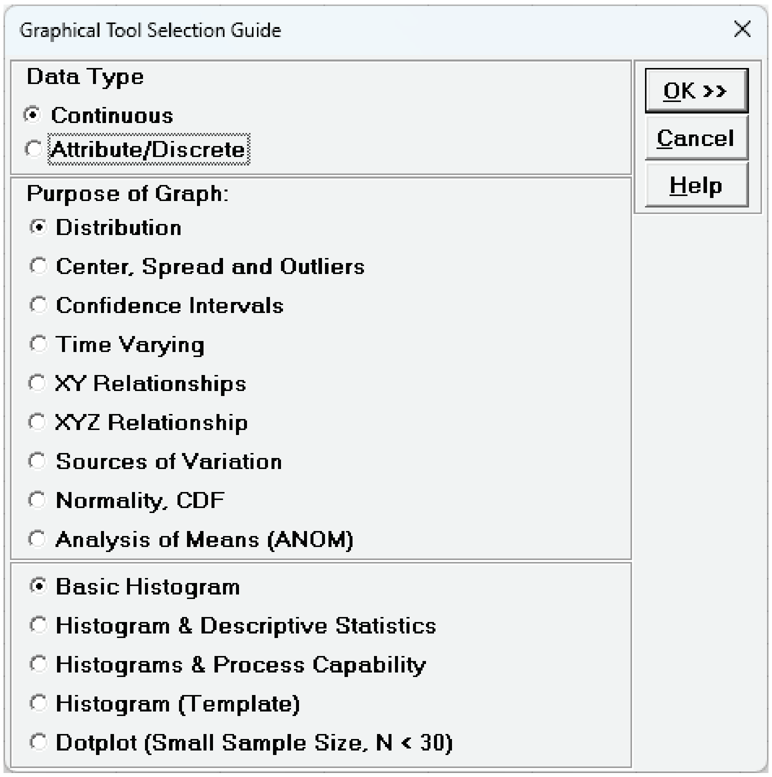
Sat_Type 1	Sat_Type 2	Sat_Type 3
3.2	3.5	3.2
2.9	4.1	3.3
1.9	3.7	4
2	4.1	2.6
3	5	3.5
2.5	4	4.4
	4.6	4.2

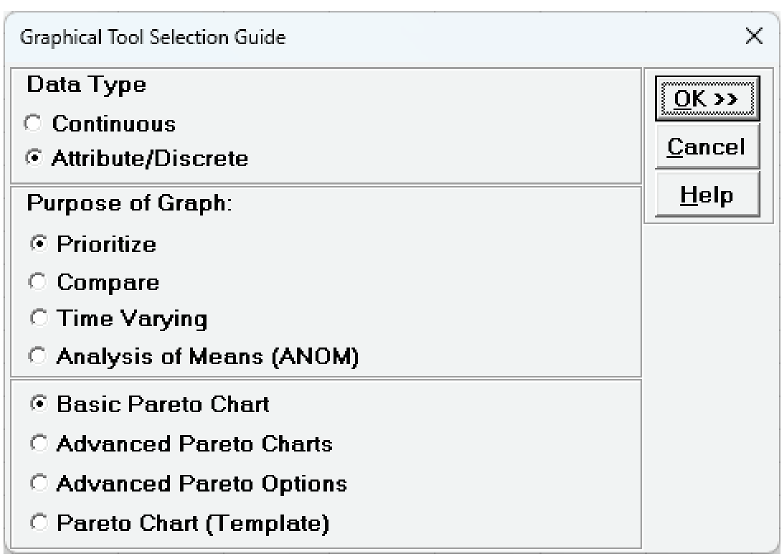
Summary of Graphical Tools

Tool	What	Type of Data	When to Use	Location in SigmaXL
Pareto Chart	Plots a bar chart of the response in descending order with a cumulative sum line.	Y=Discrete (e.g., Defect Count) or Continuous (e.g., Cost; must be additive) X=Discrete (Category)	To separate the vital few from the trivial many, help specify a problem statement, and prioritize potential root causes. This chart is based on the Pareto principle, which states that typically 80% of the defects in a process or product are caused by only 20% of the possible causes.	➢ SigmaXL > Templates & Calculators > Basic Graphical Templates > Pareto Chart ➢ SigmaXL > Graphical Tools > Basic Pareto Chart (Single) ➢ SigmaXL > Graphical Tools > Advanced Pareto Charts (Multiple)
Pivot Chart	Plots a stacked (or clustered) bar chart from an Excel Pivot Table.	Y=Discrete or Continuous X=Discrete (Category)	To easily 'slice and dice' your data, quickly look at different X factors and their contribution to the total. It is similar to the Pareto Chart, but without the descending bar order.	➢ SigmaXL > Graphical Tools > EZ-Pivot/Pivot Charts
Heatmap	Display counts or summary statistics in tabular format with results gradient color coded: minimum is dark blue to maximum dark red.	Y=Discrete or Continuous X=Discrete (Category)	To easily 'slice and dice' your data, quickly look at different X factors and their contribution to the total or summary statistics (typically the mean), aided by the color coding.	➢ SigmaXL > Graphical Tools > Heatmap
Histogram	Visual display of one variable showing data center, spread, shape and outliers.	Y=Continuous X=Discrete (Category)	1. Summarize large amounts of data 2. To get a 'feel for the data' 3. To compare actual description to customer specifications	➢ SigmaXL > Templates & Calculators > Basic Graphical Templates > Histogram ➢ SigmaXL > Graphical Tools > Basic Histogram (Single) ➢ SigmaXL > Graphical Tools > Histograms & Descriptive Statistics ➢ SigmaXL > Graphical Tools > Histograms & Process Capability
Dotplots	Visual display of one variable showing data center, spread, shape and outliers.	Y=Continuous X=Discrete (Category)	1. Small sample size (n < 30) 2. To get a 'feel for the data'	➢ SigmaXL > Graphical Tools > Dotplots
Boxplots	Visual display of the summary of Y data grouped by category of X.	Y=Continuous X=Discrete (Category)	Summary display to visualize differences in data center, spread and outliers across categories.	➢ SigmaXL > Graphical Tools > Boxplots ➢ SigmaXL > Graphical Tools > Multiple X Boxplots
Interval Plots	Plots the data mean and confidence intervals for each category group.	Y=Continuous X=Discrete (Category)	Quickly compare group means and their confidence intervals.	➢ SigmaXL > Graphical Tools > Interval Plots ➢ SigmaXL > Graphical Tools > Multiple X Interval Plots
Normal Probability Plot	Plots data in a straight line if the data is normally distributed.	Y=Continuous X=Discrete (Category)	To check for Normality and Outliers.	➢ SigmaXL > Graphical Tools > Normal Probability Plots
Empirical/Normal CDF Plots	Plots data from lowest to highest against their percentiles; compare against the same for the fitted Normal Distribution.	Y=Continuous X=Discrete (Category)	To check for Normality using the Cumulative Distribution Function (CDF).	➢ SigmaXL > Graphical Tools > Empirical/Normal CDF Plots
Run Charts	Plots observations in time sequence	Y=Continuous or Discrete	To view process performance over time for trends, shifts or cycles. To test for Randomness using the Nonparametric Runs Test	➢ SigmaXL > Templates & Calculators > Basic Graphical Templates > Run Chart ➢ SigmaXL > Graphical Tools > Run Chart ➢ SigmaXL > Graphical Tools > Overlay Run Chart
Multi-Vari Charts	Plots vertical lines with dots to allow comparison of subgroups on one variable.	Y=Continuous X=Discrete (Category)	To visually compare subgroups by individual data points and the mean. To identify major sources of variation (e.g., within a subgroup, between subgroups, or over time).	➢ SigmaXL > Graphical Tools > Multi-Vari Charts
Scatter Plot (Diagram)	Plots a response Y versus a predictor X.	Y=Continuous X=Continuous	To understand the possible relationships between two variables. To identify possible root causes which are related to Y.	➢ SigmaXL > Graphical Tools > Scatter Plot ➢ SigmaXL > Graphical Tools > Scatter Plot Matrix
XYZ Contour/Surface Plots	Plots a response Z versus two predictors Y and X.	Z=Continuous Y=Continuous X=Continuous	To understand the possible relationship between a response variable and two predictor variables without a regression model.	➢ SigmaXL > Graphical Tools > XYZ Contour/Surface Plot
Analysis of Means (ANOM)	Plots response Y mean for each level of X category with decision limits.	Y=Continuous X=Discrete (Category)	ANOM is a complement to ANOVA showing which group means are significantly different than the grand mean.	➢ SigmaXL > Graphical Tools > Analysis of Means (ANOM)
Control Charts	Plots observations in time sequence against a mean and control limits.	Y=Continuous or Discrete	To monitor the process over time for trends, shifts or cycles in order to control and improve process performance. To identify special causes.	➢ SigmaXL > Control Charts > ➢ SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart ➢ SigmaXL > Time Series Forecasting > ARIMA Control Chart

Graphical Tool Selection Guide

Click **SigmaXL > Graphical Tools > Graphical Tool Selection Guide** to start the guide. Please have the data that you wish to analyze as the active sheet with data preselected or ready for selection. Select **Data Type** as **Continuous** or **Attribute/Discrete**. Select **Purpose of Graph** and then choose from the available graphical tools. Click **OK >>** to start the data selection and creation of the graph. Note that this is equivalent to selecting a graphical tool from the menu.



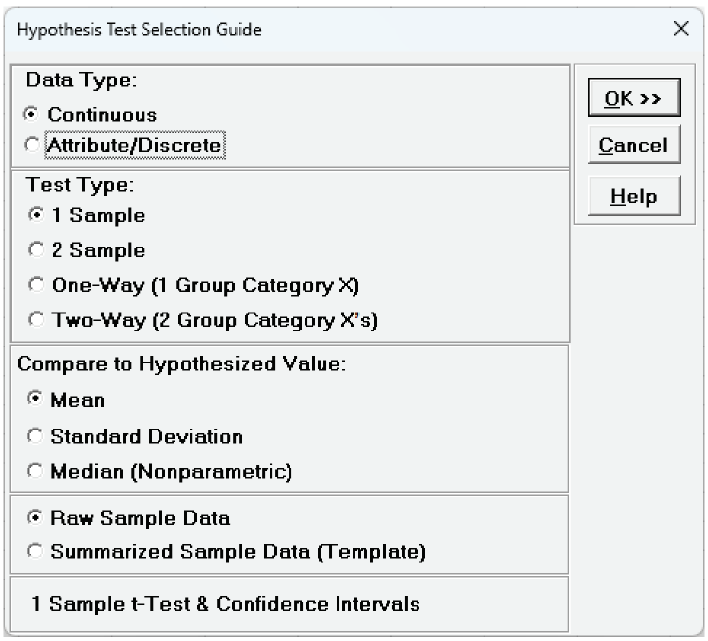


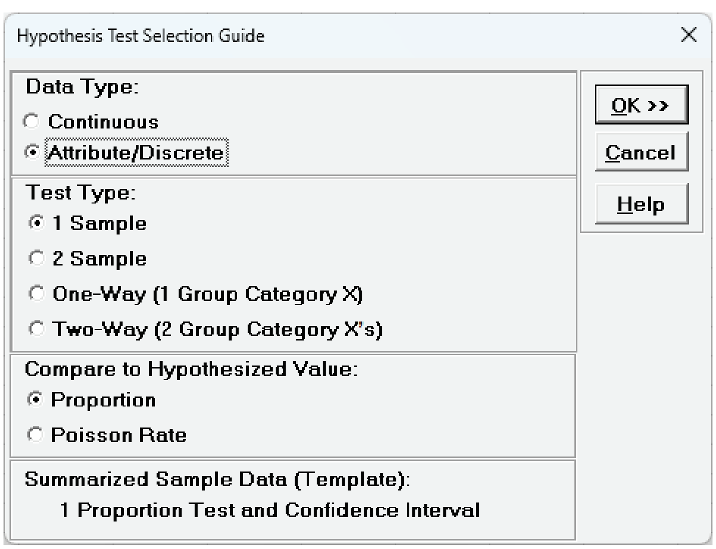
Summary of Statistical Tools

Tool	What	Type of Data	When to Use	Location in SigmaXL
t-Test	Determine if there is a significant difference between two group means or if the true mean of the data is equal to a standard value.	Y=Continuous X=Discrete (Category)	1. Test if mean = specified value 2. Test if 2 sample means are equal 3. Paired t: to reduce variation when comparing two sample means 4. Multiple pairwise comparisons	SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample t-Test and Confidence Interval for Mean 2 Sample t-Test and Confidence Interval (Compare 2 Means) SigmaXL > Statistical Tools > 1 Sample t-Test & Confidence Intervals / Paired t-Test 2 Sample t-Test / 2 Sample Comparison Tests One-Way ANOVA & Means Matrix
One-Way ANOVA (Analysis of Variance)	Determine if there is a difference in mean among many groups.	Y=Continuous X=Discrete (Category)	Determine if there is a statistically significant difference in means among the groups.	SigmaXL > Statistical Tools > One-Way ANOVA & Means Matrix (for equal variance) SigmaXL > Statistical Tools > Equal Variance Tests > Welch's ANOVA (for unequal variance)
Two-Way ANOVA (Analysis of Variance)	Determine if there is a difference in mean among many groups for two factors plus their interaction.	Y=Continuous X1=Discrete (Category) X2=Discrete (Category)	Determine if there is a statistically significant difference in means among the groups.	SigmaXL > Statistical Tools > Two-Way ANOVA
Nonparametric Tests	Determine if there is a difference between two or more group medians or if the median of the data is equal to a standard value.	Y=Continuous X=Discrete (Category)	1. Test if median = specified value: 1 Sample Sign Test or Wilcoxon 2. Test if 2 sample medians are equal: 2 Sample Mann-Whitney 3. Test if there is a difference in medians among the groups: Kruskal-Wallis or Mood's Median	SigmaXL > Statistical Tools > Nonparametric Tests SigmaXL > Statistical Tools > Nonparametric Tests – Exact (for small sample)
CI for Standard Deviation/ F-test / Bartlett's Test/ Levene's Test	Determine if there is a difference between two or more group variances or determine the confidence interval of a single standard deviation.	Y=Continuous X=Discrete (Category)	1. Determine the confidence interval for a single standard deviation. 2. Test if 2 sample variances (standard deviations) are equal. 3. Determine if there is a statistically significant difference for the variances among the groups. Use Bartlett's test for normal data. Use Levene's test for non-normal data.	SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample Chi-Square Test and CI for Standard Deviation 2 Sample F-Test and CI (Compare 2 Standard Deviations) SigmaXL > Statistical Tools > Two Sample Comparison Tests SigmaXL > Statistical Tools > Equal Variance Tests > Bartlett / Levene
Proportions Test	Determine if there is a difference between two proportions or determine the confidence interval of a single proportion.	Y=Discrete (Proportion) X=Discrete (Category)	1. Determine the confidence interval for a single proportion. 2. Determine if there is a statistically significant difference for two proportions.	SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Proportion Test & Confidence Interval 2 Proportions Test & Confidence Interval
Poisson Rate Test	Determine if there is a difference between two rates or determine the confidence interval of a single rate.	Y=Discrete (Count) X=Discrete (Category)	1. Determine the confidence interval for a single rate. 2. Determine if there is a statistically significant difference for two rates.	SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Poisson Rate Test and Confidence Interval 2 Poisson Rates Test and Confidence Interval
One-Way Chi-Square Goodness-of-Fit	Determine if the observed frequencies for one discrete variable are distributed equally	Y=Discrete (Count)	Test the distribution of observed frequency counts against expected (typically uniform discrete).	SigmaXL > Templates & Calculators > Basic Statistical Templates > One-Way Chi-Square Goodness-of-Fit Test One-Way Chi-Square Goodness-of-Fit Test Exact (for small sample)
Chi Square	Determine if there is a difference for observed frequencies of two discrete variables.	Y=Discrete (Count) X=Discrete (Category)	Determine if there is a relationship between two discrete variables.	SigmaXL > Statistical Tools > Chi-Square Tests > Chi-Square Test (for raw discrete data in stacked column format) Chi-Square Test – Two Way Table Data (for pivot or contingency table) SigmaXL > Statistical Tools > Chi-Square Tests – Exact (for small sample)
Anderson-Darling Normality Test	Determine if the data is normally distributed.	Y=Continuous	Test if the sample data is normally distributed.	SigmaXL > Graphical Tools > Histograms & Descriptive Statistics SigmaXL > Statistical Tools > Descriptive Statistics
Correlation	Quantify strength of relationships.	Y=Continuous X=Continuous	Determine if there is evidence of a relationship between Xs and Ys, quantify the relationship, identify root causes.	SigmaXL > Statistical Tools > Correlation Matrix
Regression (Simple Linear & Multiple)	Summarizes, describes, predicts and quantifies relationships.	Y=Continuous X=Continuous or Discrete (Category)	1. Determine if there is evidence of a relationship between Xs and Ys. 2. Model data to develop a mathematical equation to quantify the relationship. 3. Identify root causes. 4. Make predictions using the model.	SigmaXL > Graphical Tools > Scatter Plots (for simple linear regression) SigmaXL > Statistical Tools > Regression > Multiple Regression SigmaXL > Statistical Tools > Advanced Multiple Regression
Logistic Regression	Summarizes, describes, predicts and quantifies relationships.	Y=Discrete (Binary or Ordinal) X=Continuous or Discrete		SigmaXL > Statistical Tools > Regression > Binary Logistic Regression Ordinal Logistic Regression
General Linear Model	Summarizes, describes, predicts and quantifies relationships.	Y=Continuous X=Discrete (Category - Fixed or Random) or Continuous (Covariate)	1. Model nested data to develop a mathematical equation to quantify the relationship. 2. Determine Variance Components (Sources of Variation)	SigmaXL > Statistical Tools > General Linear Model
Design of Experiments (DOE)	Systematic and efficient proactive approach to testing relationships.	Y=Continuous or Discrete X=Continuous or Discrete	To establish cause and effect relationship between Ys and Xs. To identify 'vital few' Xs.	SigmaXL > Design of Experiments
Time Series Forecasting	Make predictions with time series data.	Y=Continuous	Analyze historical data in order to predict future data. Models include Exponential Soothing and ARIMA.	SigmaXL > Time Series Forecasting

Hypothesis Test Selection Guide

Click **SigmaXL > Statistical Tools > Hypothesis Test Selection Guide** to start the guide. Please have the data that you wish to analyze as the active sheet with data preselected or ready for selection. Select **Data Type** as **Continuous** or **Attribute/Discrete**. Select **Test Type** as **1 Sample**, **2 Sample**, **One-Way (1 Group Category X)**, or **Two-Way (2 Group Category X's)**. Select **Compare to Hypothesized Value** or **Compare** options. If available, select **Raw Sample Data** or **Summarized Sample Data (Template)**, then choose from the available hypothesis tests. Click **OK >>** to start the data selection and perform the hypothesis test. This is equivalent to selecting a hypothesis test from the menu. Note that Regression, Correlation and General Linear Model are not included in this guide.





SigmaXL: Measure Phase Tools

Part A - Basic Data Manipulation

Introduction to Basic Data Manipulation

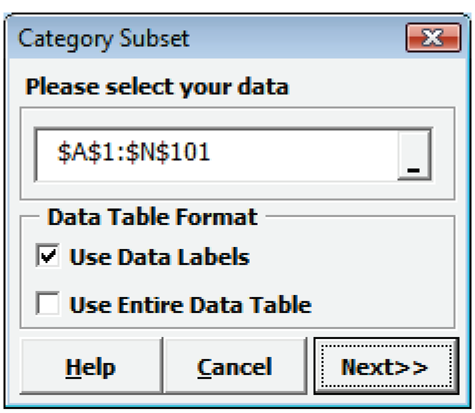
Open **Customer Data.xlsx** (to access, click **SigmaXL > Help > Sample Data** or **Start > Programs > SigmaXL > Sample Data**). This data is in **stacked** column format. This format is highly recommended for use with SigmaXL. Note that all pertinent information is provided in each record (row). Also note that only one row is used for column headings (labels) and there are no blank rows or columns. Each column contains a consistent format of either numeric, text, or date. This is also the data format used by other major statistical software packages.

Customer Record No	Order Date	Customer Type	Avg No. of orders per mo	Avg days Order to delivery time	Loyalty - Likely to Recommend	Overall Satisfaction	Responsive to Calls	Ease of Communications	Staff Knowledge	Size of Customer	Major-Complaint	Product Type
1	1/5/2001	2	24	38	4	3.54	3.02	4.07	1.65	Small	Return-calls	Consumer
2	1/5/2001	3	36.4	42	4	3.16	3.21	3.11	3.8	Large	Difficult-to-order	Consumer
3	1/5/2001	2	32.8	44	2	2.42	1.93	2.9	2.88	Small	Return-calls	Manufacturer
4	1/5/2001	2	47.6	48	3	2.7	1.88	2.52	4.08	Large	Difficult-to-order	Manufacturer
5	1/5/2001	3	30.6	51	3	3.31	3.75	2.86	3.88	Small	Not-available	Consumer
6	1/5/2001	2	52.2	55	4	4.12	4.31	3.93	1.12	Large	Return-calls	Consumer
7	1/5/2001	1	35.8	49	4	3.24	4.06	2.42	4.64	Large	Return-calls	Manufacturer
8	1/5/2001	2	36.5	39	4	4.47	4.75	4.2	4.98	Large	Return-calls	Manufacturer
9	1/5/2001	2	39.9	44	4	3.83	3.18	4.48	3.16	Large	Difficult-to-order	Consumer
10	1/5/2001	1	28	43	3	2.94	2.03	3.85	4.01	Small	Return-calls	Consumer
11	1/8/2001	2	25.9	44	3	3.24	3.05	4.43	4.72	Small	Return-calls	Manufacturer
12	1/8/2001	2	23.9	50	3	4.18	3.67	4.69	4.86	Small	Difficult-to-order	Manufacturer
13	1/8/2001	2	37.9	58	5	4.53	4.29	4.77	1.9	Large	Return-calls	Consumer

Note that Loyalty, Overall Satisfaction, Responsive to Calls, Ease of Communications, and Staff Knowledge were obtained from surveys. A Likert scale of 1 to 5 was used, with 1 being very dissatisfied, and 5 very satisfied. Survey results were averaged to obtain non-integer results.

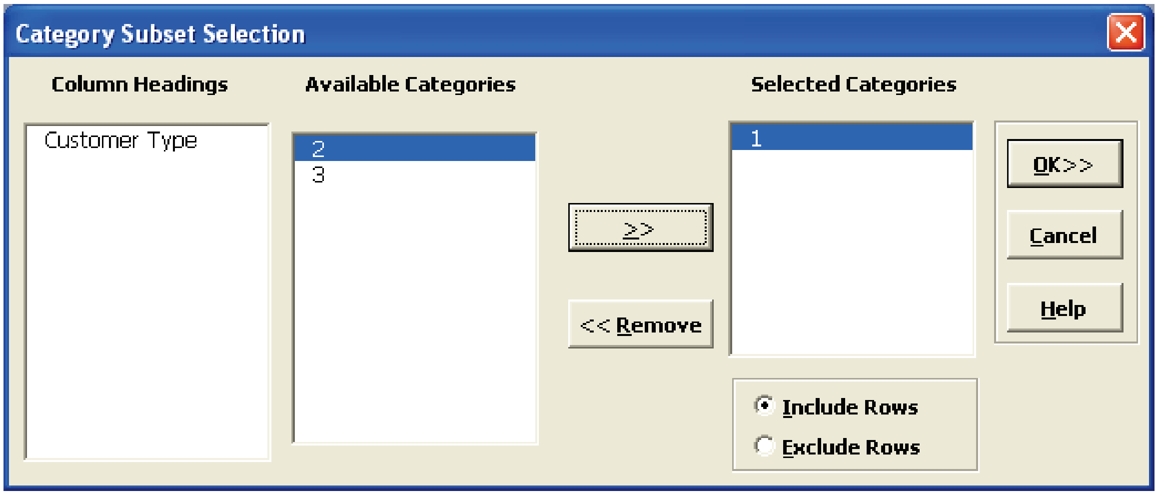
Category Subset

1. Click **SigmaXL > Data Manipulation > Category Subset**.
2. If you are working with a portion of a dataset, specify the appropriate range, otherwise check **Use Entire Data Table**.



3. Click **Next**.

4. Select *Customer Type*, *1*, **>>** as shown:

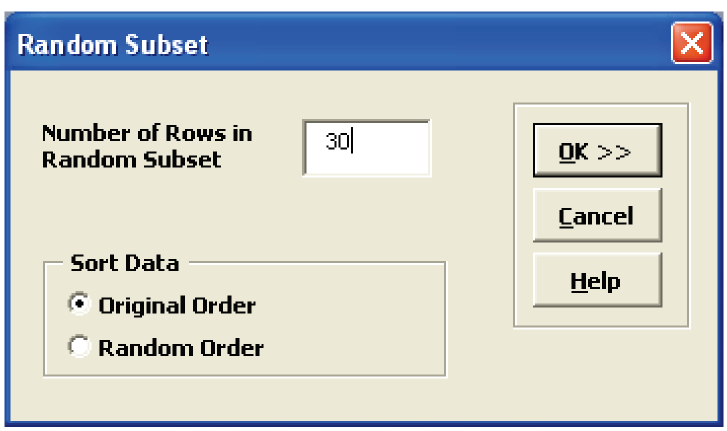


5. Click **OK**.

   A new subset worksheet is created containing only Customer Type 1.
   Note: We could have chosen more than one Customer Type and had the option to create a subset which included or excluded these Customer Types.

Random Subset

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.
2. Click **SigmaXL > Data Manipulation > Random Subset**.
3. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
4. Enter **Number of Rows in Random Subset** as 30. The default **Sort Data** selection is **Original Order**.



5. Click **OK**. A new worksheet is created that contains a random subset of 30 rows. This feature is useful for data collection to ensure a random sample, e.g., given a list of transaction numbers select a random sample of 30 transactions.

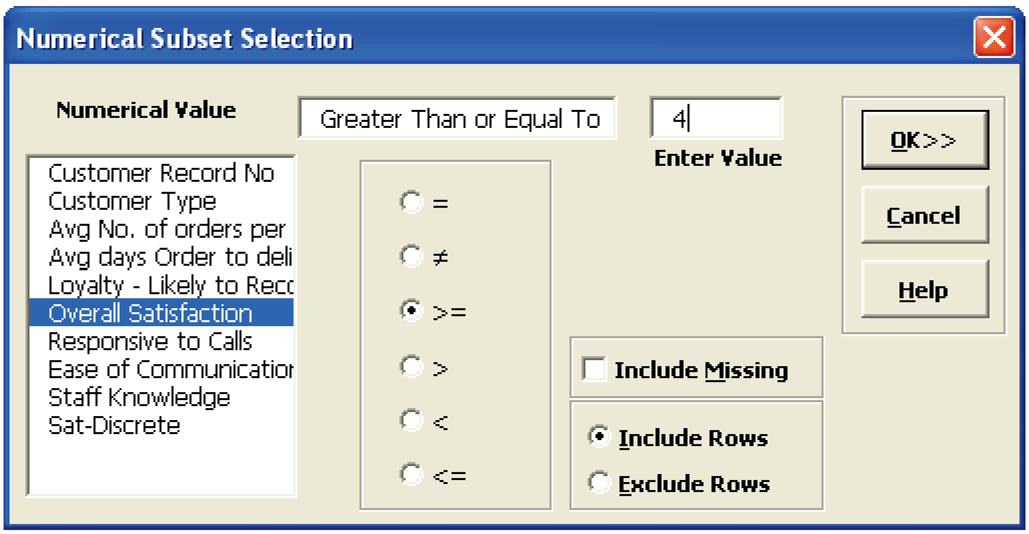
Numerical Subset

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.

2. Click **SigmaXL > Data Manipulation > Numerical Subset**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

4. Select *Overall Satisfaction*, **>=**, **Enter Value** as 4.



5. Click **OK**.

A new subset worksheet is created containing only those rows with Overall Satisfaction >= 4.

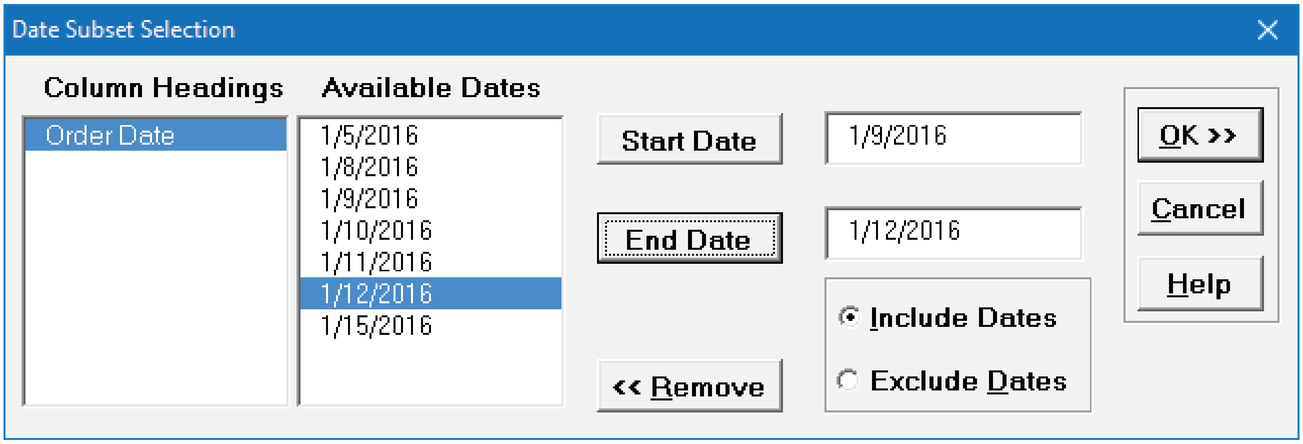
Date Subset

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.

2. Click **SigmaXL > Data Manipulation > Date Subset**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

4. Select *Order Date*, select *1/9/2016*, click **Start Date**, select *1/12/2016*, click **End Date**.



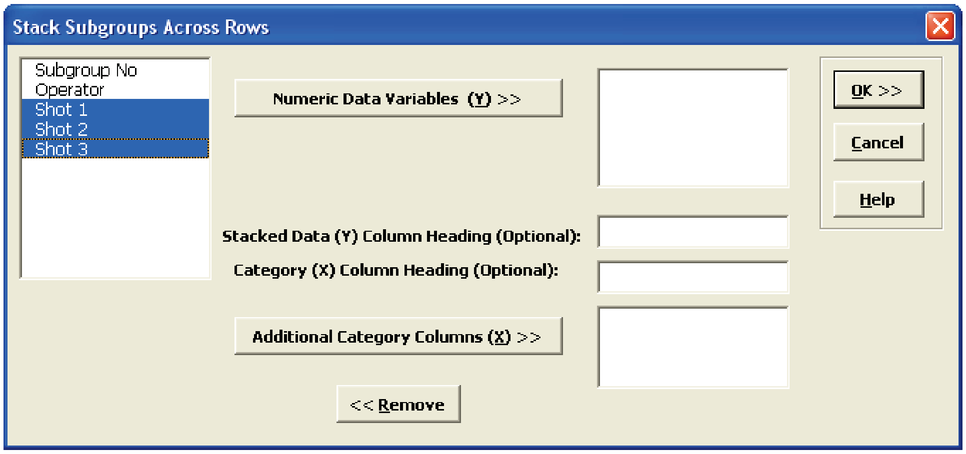
5. Click **OK**. A new subset worksheet is created containing only those rows with Order Date between 1/9/2016 to 1/12/2016.

Transpose Data

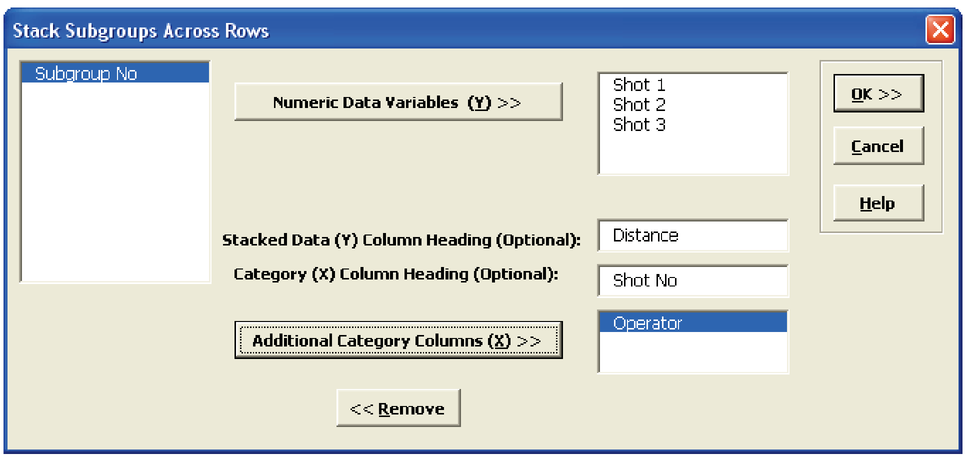
1. Open **Catapult Data Row Format.xlsx**.
2. Manually select the entire data table if the data is not already selected.
3. Click **SigmaXL > Data Manipulation > Transpose Data**.
4. This will transpose rows to columns or columns to rows. It is equivalent to **Copy**, **Paste Special**, **Transpose**.

Stack Subgroups Across Rows

1. Now we will stack the subgroups across rows for the transposed data. Ensure that the **Transposed Data Sheet** of **Catapult Data Row Format.xlsx** is active.
2. Click **SigmaXL > Data Manipulation > Stack Subgroups Across Rows**.
3. Check **Use Entire Data Table**, click **Next**.
4. Click on *Shot 1*. Shift Click on *Shot 3* to highlight the three columns of interest.



5. Click **Numeric Data Variables (Y) >>** to select *Shot 1 to Shot 3*. Enter *Distance* as the **Stacked Data (Y) Column Heading** and *Shot No* as the **Category (X) Column Heading**. Select *Operator,* click **Additional Category Columns (X) >>:**



Note that any selected column may be removed by highlighting and double-clicking or clicking the **Remove** button.

6. Click **OK**. The resulting stacked data is shown:

Operator	Shot No	Distance
John	Shot 1	105
John	Shot 2	104
John	Shot 3	99
Moe	Shot 1	105
Moe	Shot 2	99
Moe	Shot 3	91
Sally	Shot 1	96
Sally	Shot 2	96
Sally	Shot 3	99
Sue	Shot 1	102
Sue	Shot 2	105
Sue	Shot 3	103
David	Shot 1	100
David	Shot 2	96
David	Shot 3	100

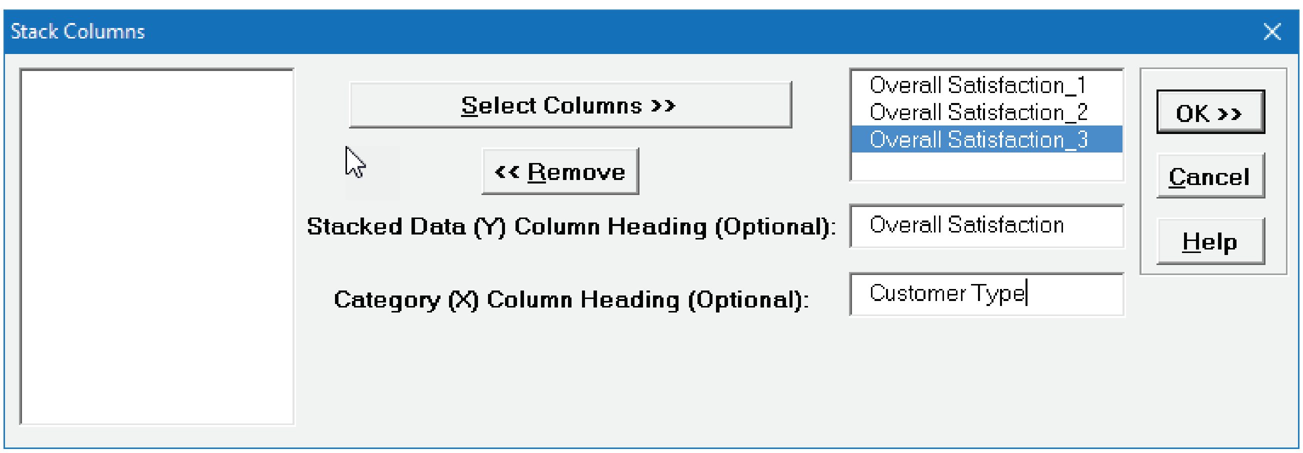
Stack Columns

1. Open **Customer Satisfaction Unstacked.xlsx**.

2. Click **SigmaXL > Data Manipulation > Stack Columns**.

3. Check **Use Entire Data Table**, click **Next**.

4. Shift Click on *Overall Satisfaction_3* to highlight all three column names. Click **Select Columns >>**. Enter the **Stacked Data (Y) Column Heading (Optional)** as *Overall Satisfaction*. Enter the **Category (X) Column Heading (Optional)** as *Customer Type*.



5. Click **OK**. Shown is the resulting stacked column format:

Customer Type	Overall Satisfaction
Overall Satisfaction_1	3.24
Overall Satisfaction_1	2.94
Overall Satisfaction_1	1.86
Overall Satisfaction_1	2.04
Overall Satisfaction_1	2.96
Overall Satisfaction_1	2.53
Overall Satisfaction_1	4.67
Overall Satisfaction_1	4.67
Overall Satisfaction_1	2.57
Overall Satisfaction_1	3.09
Overall Satisfaction_1	3.57
Overall Satisfaction_1	4.25
Overall Satisfaction_1	4.05
Overall Satisfaction_1	3.58
Overall Satisfaction_1	3.82
Overall Satisfaction_1	3.8
Overall Satisfaction_1	2.81
Overall Satisfaction_1	3.99
Overall Satisfaction_1	4.15
Overall Satisfaction_1	3.56
Overall Satisfaction_1	3.26
Overall Satisfaction_1	4.8
Overall Satisfaction_1	1.72
Overall Satisfaction_1	3.01
Overall Satisfaction_1	2.65
Overall Satisfaction_1	3.92
Overall Satisfaction_1	4.24
Overall Satisfaction_1	3.97
Overall Satisfaction_1	4.02
Overall Satisfaction_1	2.56
Overall Satisfaction_1	2.9
Overall Satisfaction_2	3.54
Overall Satisfaction_2	2.42
Overall Satisfaction_2	2.7

6. Data that is in stacked column format can be unstacked using **Data Manipulation > Unstack Columns**.

Random Data

The normal random data generator is used to produce normal random data. Column headings are automatically created with Mean and Standard Deviation values (e.g. 1: Mean = 0; StDev = 1). This utility works with Recall SigmaXL Dialog (**F3**) to append columns to the current Normal Random Data worksheet. An example is shown in **Measure Phase Tools, Part G – Normal Probability Plots**.

Additional random number generators include Uniform (Continuous & Integer), Lognormal, Exponential, Weibull and Triangular. The column headings show the specified parameter values.

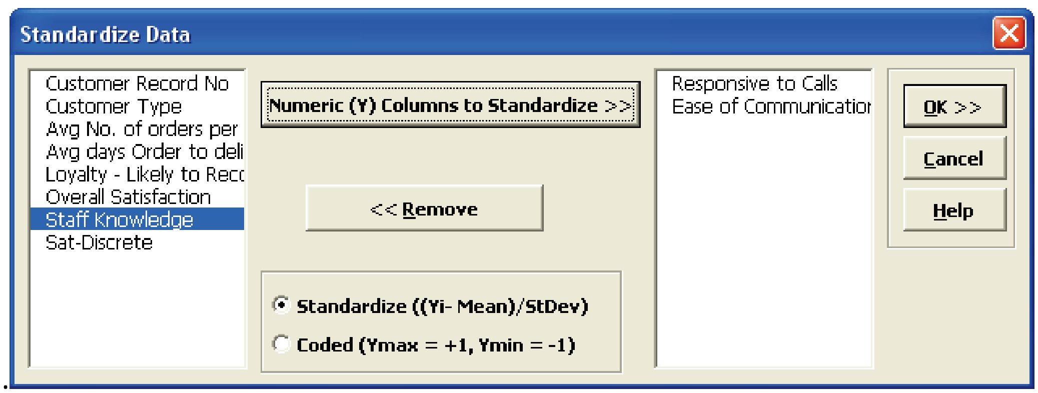
Box-Cox Transformation

This tool is used to convert nonnormal data to normal by applying a power transformation. Examples of use are given in **Measure Phase Tools, Part J – Process Capability for Nonnormal Data** and **Control Phase Tools, Part A – Individuals Charts for Nonnormal Data**.

Standardize Data

This tool is used to Standardize ((Yi – Mean)/StDev) or Code (Ymax = +1, Ymin = -1) your data. This is particularly useful when performing Multiple Regression. Standardized Predictors have better statistical properties. For example, the importance of model coefficients can be determined by the relative size because units are removed. Another statistical benefit is reduced multicollinearity when investigating two-factor interactions.

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.
2. Click **SigmaXL > Data Manipulation > Standardize Data**.
3. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
4. Select *Responsive to Calls* and *Ease of Communications* and click **Numeric (Y) Columns to Standardize >>**



5. Click **OK**. The results are given on the **Standardize** sheet:

Responsive to Calls	Responsive to Calls_Standardize	Ease of Communications	Ease of Communications_Standardize
3.02	-0.740511661	4.07	0.353207938
3.21	-0.573887768	3.11	-0.700161496
1.93	-1.696406629	2.9	-0.93058606
1.88	-1.740255022	2.52	-1.347544794
3.75	-0.100325123	2.86	-0.974476453
4.31	0.390776879	3.93	0.199591563

Convert to Discrete

This tool is used to convert continuous data to discrete (a.k.a. discretization) using percentile based equal frequency. It is useful for graphical and statistical tools that require group category variables.

The percentiles are selected to produce the specified number of discrete levels. For example, with the default number of discrete levels = 4, SigmaXL uses the 25^{th} percentile, 50^{th} percentile and 75^{th} percentile as bin intervals. The numerical values are converted as follows:

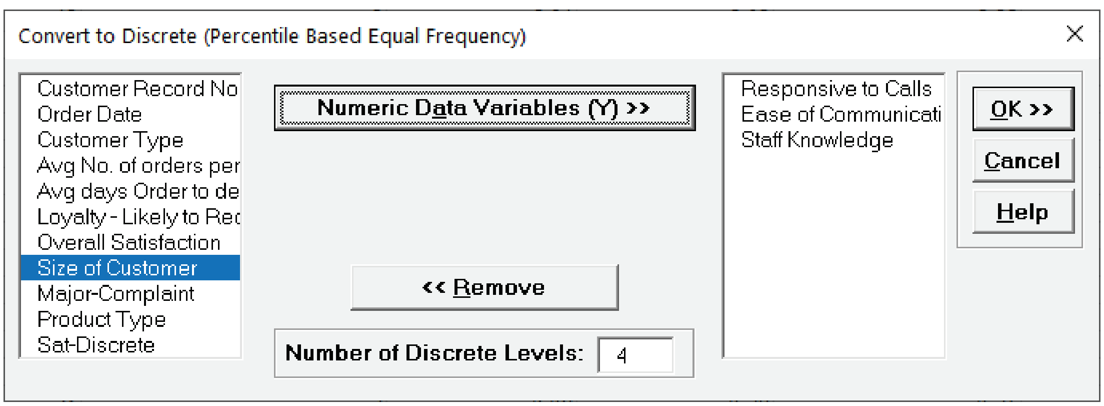
Minimum to < 25^{th} percentile -> 1
25^{th} percentile to < 50^{th} percentile -> 2
50^{th} percentile to < 75^{th} percentile -> 3
75^{th} percentile to Maximum -> 4

These percentiles will produce an approximate equal frequency if the data are continuous. Ties will result in some unequal frequency counts.

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.

2. Click **SigmaXL > Data Manipulation > Convert to Discrete**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Responsive to Calls*, *Ease of Communications* and *Staff Knowledge*. Click **Numeric Data Variables (Y) >>**



4. Click **OK**. The results are given on the **Convert to Discrete** sheet:

Responsive to Calls	Ease of Communications	Staff Knowledge	Responsive to Calls_discrete	Ease of Communications_discrete	Staff Knowledge_discrete
3.02	4.07	1.65	2	3	1
3.21	3.11	3.8	2	1	3
1.93	2.9	2.88	1	1	2
1.88	2.52	4.08	1	1	4
3.75	2.86	3.88	2	1	3
4.31	3.93	1.12	2	3	1
4.06	2.42	4.64	2	1	4
4.75	4.2	4.98	3	3	4
3.18	4.48	3.16	2	3	2
2.03	3.85	4.01	1	3	3
3.05	4.43	4.72	2	3	4
3.67	4.69	4.86	2	4	4
4.29	4.77	1.9	2	4	1
3.03	3.42	4.02	2	2	3
2.23	1.48	3.96	1	1	3
2.73	4.6	3.88	1	3	3
5	4.03	3.34	4	3	3
1.11	2.97	3.92	1	1	3
4.99	3.61	5	4	2	4

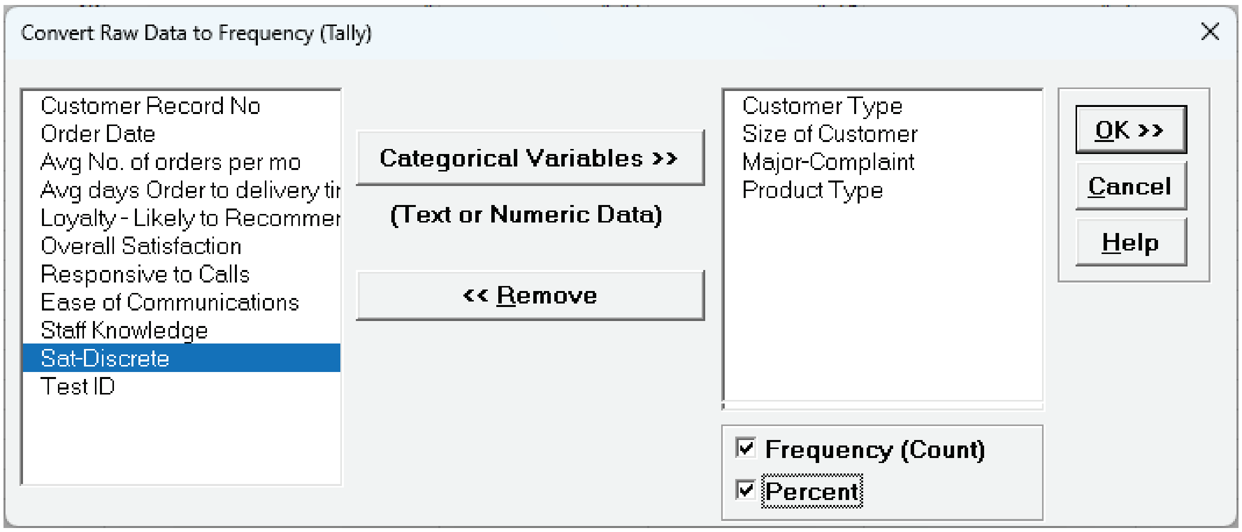
Frequency Conversion - Convert Raw Data to Frequency (Tally)

This tool is used to convert raw categorical data to frequency data, a.k.a. Tally. It performs the same function as Count in Excel's Pivot Table, but is easier to use.

1. Click **Sheet 1** Tab of **Customer Data.xlsx**.

2. Click **SigmaXL > Data Manipulation > Frequency Conversion > Convert Raw Data to Frequency (Tally)**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Customer Type*, *Size of Customer*, *Major-Complaint* and *Product Type*. Click **Categorical Variables >>.** Check both **Frequency (Count)** and **Percent.**



If a variable is selected with more than 50 unique levels, a warning message is given. Typically, this occurs when the user has incorrectly selected a continuous variable as categorical.

4. Click **OK**. The results are given on the **Raw Data to Frequency (Tally)** sheet:

Customer Type	Frequency (Count)	Percent		Size of Customer	Frequency (Count)	Percent		Major-Complaint	Frequency (Count)	Percent		Product Type	Frequency (Count)	Percent
1	31	31.00		Large	45	45.00		Difficult-to-order	19	19.00		Consumer	50	50.00
2	42	42.00		Small	55	55.00		Not-available	4	4.00		Manufacturer	50	50.00
3	27	27.00						Order-takes-too-long	10	10.00				
				Total=	100			Return-calls	60	60.00		Total=	100	
Total=	100							Wrong-color	7	7.00				
								Total=	100					

Frequency Conversion - Convert Frequency to Raw Data

This tool is used to convert frequency data to raw data.

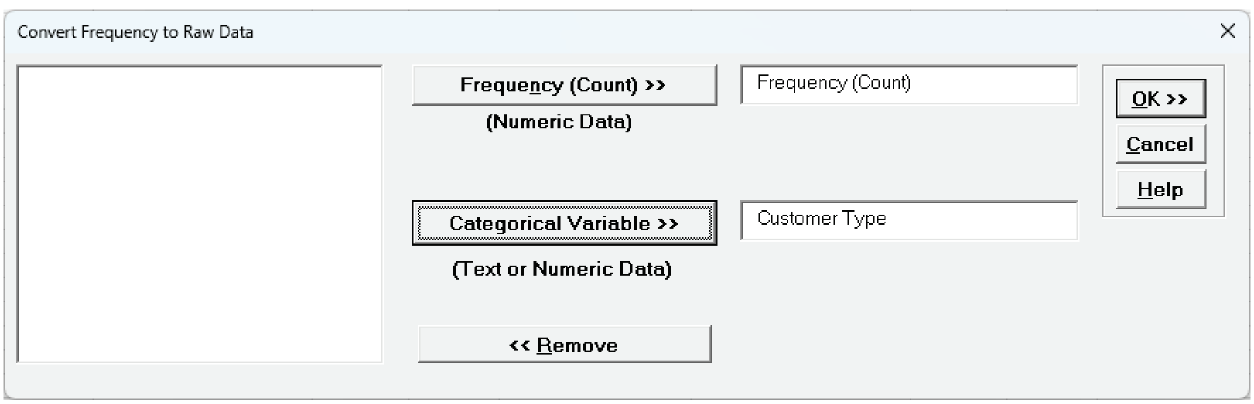
1. Follow the steps above to create a **Raw Data to Frequency (Tally)** sheet.

2. Select cells **A1:B4** as shown.

	A	B	C
1	Customer Type	Frequency (Count)	Percent
2	1	31	31.00
3	2	42	42.00
4	3	27	27.00
5			
6	Total=	100	
7			

3. Click **SigmaXL > Data Manipulation > Frequency Conversion > Convert Frequency to Raw Data**. Click **Next**.

4. Select *Frequency (Count)*. Click **Frequency (Count) >>**. Select *Customer Type*. Click **Categorical Variables >>**.



5. Click **OK**. The results are given on the **Convert Frequency to Raw Data** sheet:

Customer Type
1
1
1
1
1
1
1
1
1
1

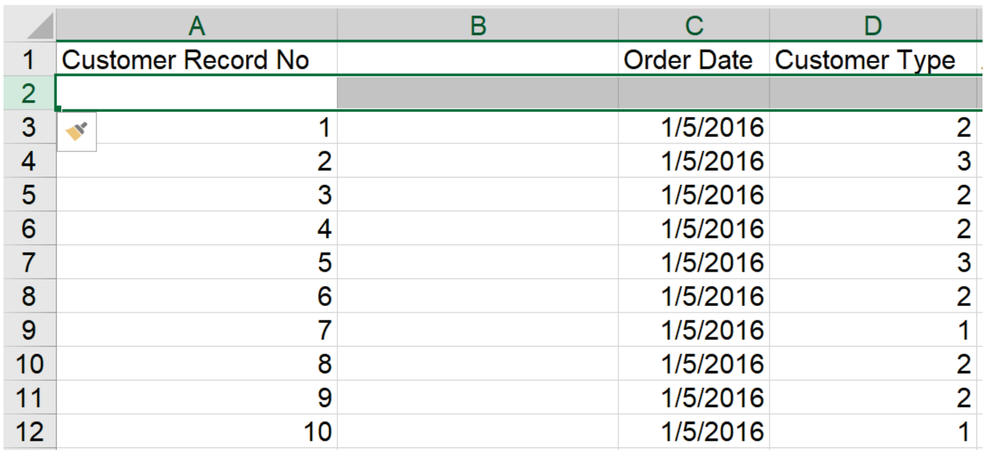
Data Preparation – Remove Blank Rows and Columns

This data preparation utility is provided as a convenient way to prepare data for analysis by deleting any empty rows and/or columns.

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab.

2. Insert a new column in B; Click **Column B heading**, **Right-Click > Insert > Columns**.

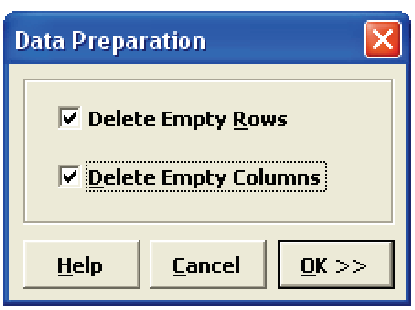
3. Insert a new row in row 2. Click **Row 2 label**, **Right-Click > Insert > Rows** as shown:



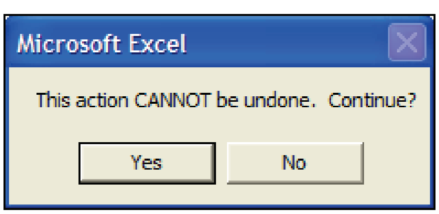
	A	B	C	D
1	Customer Record No		Order Date	Customer Type
2				
3	1		1/5/2016	2
4	2		1/5/2016	3
5	3		1/5/2016	2
6	4		1/5/2016	2
7	5		1/5/2016	3
8	6		1/5/2016	2
9	7		1/5/2016	1
10	8		1/5/2016	2
11	9		1/5/2016	2
12	10		1/5/2016	1

4. This is now an example of a data set that requires deletion of empty rows and columns. Click **SigmaXL >Data Manipulation >Data Preparation >Remove Blank Rows and Columns**.

5. Check **Delete Empty Rows** and **Delete Empty Columns**.



6. Click **OK**. A warning message is given prior to the deletion step.



7. Click **Yes**. The empty rows and columns are deleted automatically.

	A	B	C
1	Customer Record No	Order Date	Customer Type
2	1	1/5/2001	2
3	2	1/5/2001	3
4	3	1/5/2001	2
5	4	1/5/2001	2
6	5	1/5/2001	3
7	6	1/5/2001	2
8	7	1/5/2001	1
9	8	1/5/2001	2
10	9	1/5/2001	2
11	10	1/5/2001	1

Data Preparation – Change Text Data Format to Numeric

This **Data Preparation** utility will convert data that represents numeric values but are currently in text format. This sometimes occurs when importing data into Excel from another application or text file.

Recall SigmaXL Dialog

Recall SigmaXL Dialog is used to activate the last data worksheet and recall the last dialog, making it very easy to do repetitive analysis. To access, in Excel 2013 and newer, click the **Recall SigmaXL Dialog** menu button as shown below:

Alternatively, you can use the Hot Key **F3** (not available in Excel Mac). This feature can also be accessed by clicking **SigmaXL > Help > Hot Keys > Recall SigmaXL Dialog**.

Note that **Recall SigmaXL Dialog** may not be available for all functions.

Activate Last Worksheet

Activate Last Worksheet is used to activate the last data worksheet without recalling the dialog. To access, press hot key **F4** (not available in Excel Mac). This feature can also be accessed by clicking **SigmaXL > Help > Hot Keys > Activate Last Sheet**.

Part B – Templates & Calculators

Introduction to Templates & Calculators

To use SigmaXL templates, select the appropriate template, enter the inputs and the resulting outputs are produced immediately. If the template does not automatically perform the calculations, click Excel **Formulas > Calculation Options** and select **Automatic** or simply click **Formulas > Calculate Now**. Alternatively, you can click **File > Options > Formulas**, select **Workbook Calculation**, **Automatic**, and click **OK.**

Some Templates and Calculators are protected worksheets, but this may be modified by clicking **SigmaXL > Help > Unprotect Worksheet**. Alternatively, you can click Excel **File > Info > Unprotect** or **Home > Format > Unprotect Sheet**.

Click **SigmaXL > Templates & Calculators** to access the templates and calculators. Basic Graphical Templates can also be accessed by clicking **SigmaXL > Graphical Tools > Basic Graphical Templates**. Basic Statistical Templates can be accessed by clicking **SigmaXL > Statistical Tools > Basic Statistical Templates**. Basic MSA Templates can be accessed by clicking **SigmaXL > Measurement Systems Analysis > Basic MSA Templates**. Basic Process Capability Templates can be accessed by clicking **SigmaXL > Process Capability > Basic Process Capability Templates**. Basic DOE Templates can also be accessed by clicking **SigmaXL > Design of Experiments > Basic DOE Templates**. Basic Taguchi DOE Templates can also be accessed by clicking **SigmaXL > Design of Experiments > Basic Taguchi DOE Templates**. Basic Control Chart Templates can also be accessed by clicking **SigmaXL > Control Charts > Basic Control Chart Templates**.

SigmaXL Templates & Calculators

- DMAIC & DFSS Templates:
  - Team/Project Charter
  - SIPOC Diagram
  - Data Measurement Plan
  - Cause & Effect (Fishbone) Diagram and Quick Template
  - Cause & Effect (XY) Matrix
  - Failure Mode & Effects Analysis (FMEA)
  - Quality Function Deployment (QFD)
  - Pugh Concept Selection Matrix
  - Control Plan
- Lean Templates:
  - Takt Time Calculator
  - Value Analysis/Process Load Balance
  - Value Stream Mapping
- Basic Graphical Templates:
  - Pareto Chart, Histogram and Run Chart
- Basic Statistical Templates:
  - Sample Size – Discrete and Continuous
  - Minimum Sample Size for Robust t-Tests and ANOVA
  - 1 Sample Z-Test and Confidence Interval for Mean
  - 1 Sample t-Test and Confidence Interval for Mean
  - 2 Sample t-Test and Confidence Interval (Compare 2 Means)
  - 1 Sample Equivalence Test for Mean
  - 2 Sample Equivalence Test (Compare 2 Means)
  - 1 Sample Chi-Square Test and CI for Standard Deviation
  - 2 Sample F-Test and CI (Compare 2 Standard Deviations)
  - 1 Proportion Test and Confidence Interval

- 2 Proportions Test and Confidence Interval
- 2 Proportions Equivalence Test
- 1 Poisson Rate Test and Confidence Interval
- 2 Poisson Rates Test and Confidence Interval
- 2 Poisson Rates Equivalence Test
- One-Way Chi-Square Goodness-of-Fit Test
- One-Way Chi-Square Goodness-of-Fit Test Exact

- Probability Distribution Calculators:
  - Normal, Inverse Normal, Lognormal, Exponential, Weibull
  - Binomial, Poisson, Hypergeometric
- Basic MSA Templates:
  - Type 1 Gage Study
  - Gage Bias and Linearity Study
  - Gage R&R Study – with Multi-Vari Analysis
  - Attribute Gage R&R (Attribute Agreement Analysis)
- Basic Process Capability Templates:
  - Process Sigma Level – Discrete and Continuous
  - Process Capability & Confidence Intervals
  - Tolerance Interval Calculator (Normal Exact)
- Basic DOE Templates:
  - 2 to 5 Factors; Main Effects & Interaction Plots
- Basic Taguchi DOE Templates:
  - Taguchi L4, L8, L9, L12, L16, L18, L27
- Control Chart Templates:
  - Basic: Individuals, C (Count)
  - Rare Events: G, T, Prob G
  - Time Weighted: EWMA, CUSUM
  - Trend
  - Average Run Length (ARL) Calculators: Shewhart, Attribute P & C, EWMA, CUSUM

Team/Project Charter

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Team/Project Charter** to access the Team/Project Charter template.

TEAM/PROJECT CHARTER

Project Name:	
Date (Last Revision):	
Prepared By:	
Approved By:	

Business Case:	Opportunity Statement (High Level Problem Statement):
	Defect Definition:
Goal Statement:	**Project Scope:**
	Process Start Point:
	Process End Point:
Expected Savings/Benefits:	In Scope:
	Out of Scope:

Project Plan:

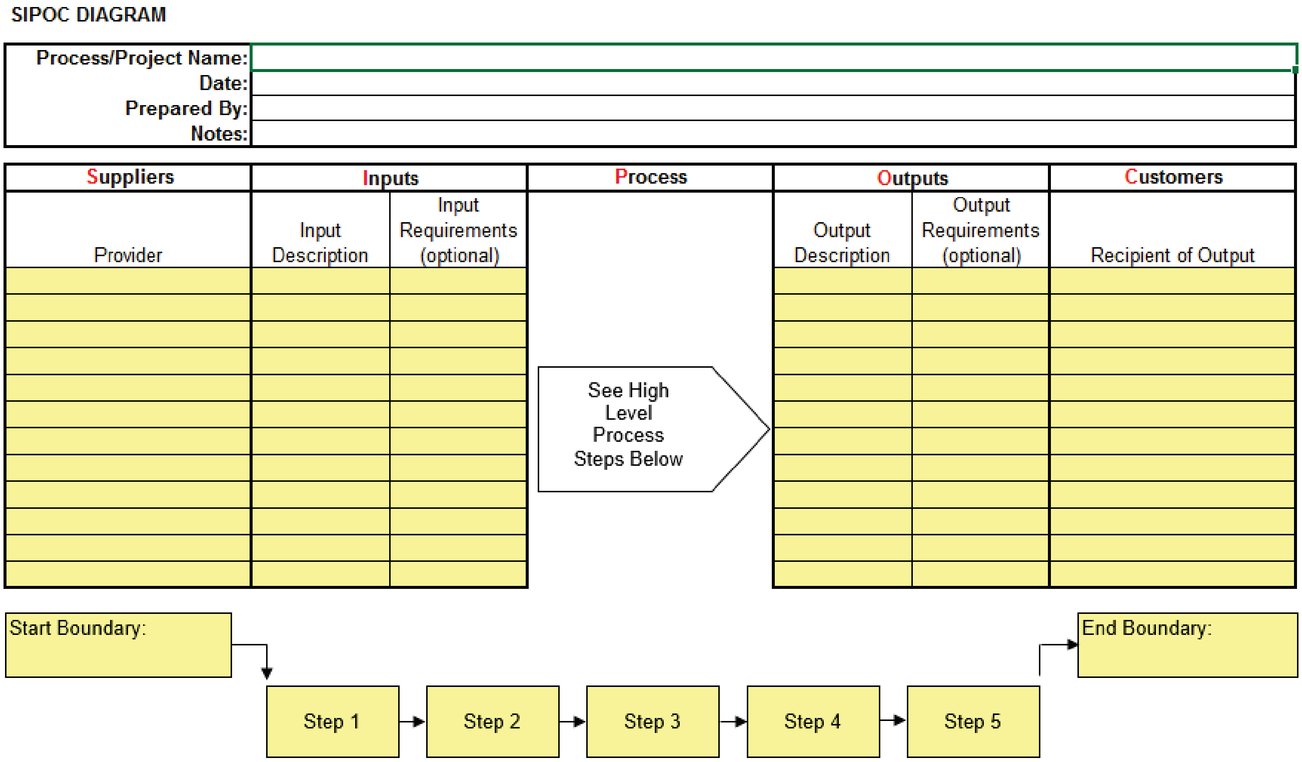
Task/Phase	Start Date	End Date	Actual End

Team:

Name:	Role:	Commitment (%):

SIPOC Diagram

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > SIPOC Diagram** to access the SIPOC Diagram template.



Data Measurement Plan

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Data Measurement Plan** to access the Data Measurement Plan template.

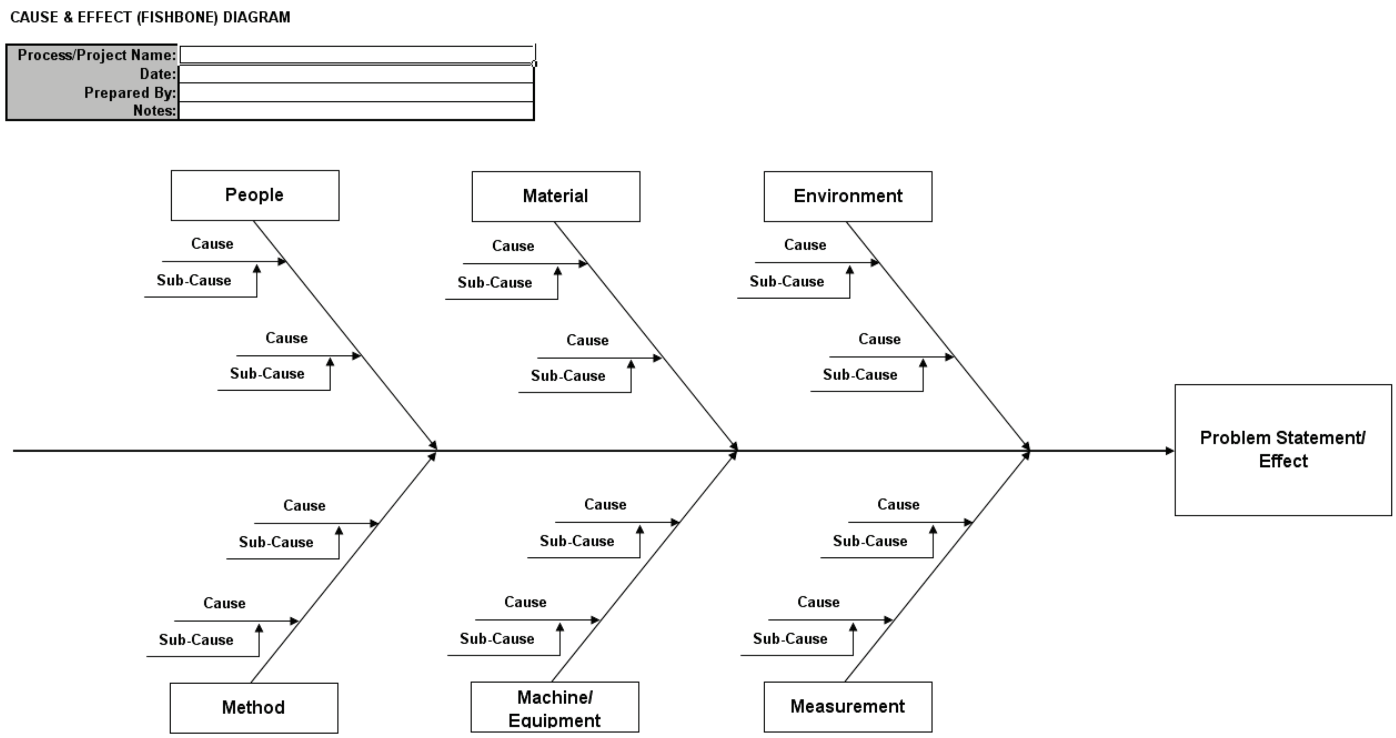
DATA MEASUREMENT PLAN

Project Name:	
Date:	
Prepared By:	
Notes:	

Measurement/Metric	X or Y	Operational Definition	Type of Data (Discrete/Continuous)	Data Source and Location	Sample Size	Who Will Collect the Data?	When Will Data be Collected?	How Will Data be Collected?	Is the Measurement System Capable?	Graphical and/or Statistical Tools to be Used

Cause & Effect (Fishbone) Diagram

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Cause & Effect (Fishbone) Diagram** to access the Cause & Effect Diagram form.



Notes:

1. Overwrite the text in the Cause, Sub-Cause and Problem Statement/Effect as appropriate.

2. Use copy and paste to create additional causes or sub-causes.

3. The arrows with text are grouped. To ungroup use Excel's Draw > Ungroup tool.

Cause & Effect (Fishbone) Template

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Cause & Effect (Fishbone) Template** to access the Cause & Effect Diagram template.

Cause & Effect (Fishbone) Template

Process/Project Name:	
Date:	
Prepared By:	
Notes:	

Fishbone Diagram

Problem Statement/Effect:	

Variable	People	Sub Cause 1	Sub Cause 2
X1			
X2			
X3			
X4			
X5			
X6			
X7			
X8			
X9			
X10			

Variable	Method	Sub Cause 1	Sub Cause 2
X1			
X2			
X3			
X4			
X5			
X6			
X7			
X8			
X9			
X10			

Variable	Material	Sub Cause 1	Sub Cause 2
X1			
X2			
X3			
X4			
X5			
X6			
X7			
X8			
X9			
X10			

Simply fill in the template table and click the Fishbone Diagram button to create the diagram.

Cause & Effect (XY) Matrix Example

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Cause & Effect (XY) Matrix** to access the Cause & Effect (XY) Matrix template.

The following example is given in **SigmaXL > Help > Template Examples > C&E Matrix**. This is a simple Cause and Effect Matrix example for a Call Center. If prompted, please ensure that macros are enabled in order to allow Pareto Charts to be created.

CAUSE & EFFECT (XY) MATRIX

Process/Project Name:	Call Center Example
Date:	
Performed By:	
Notes:	

Pareto Chart

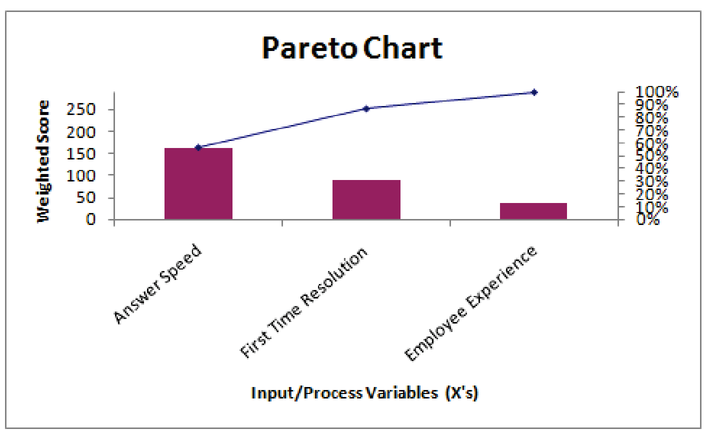
Output Variables (Y's):	Call Abandon Rate	Customer Satisfaction	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Importance Score (1-10):	8	10								

Input/Process Variables (X's)	Table of Association Scores (X's to Y's)										Weighted Score
Answer Speed	9	9									162
Employee Experience	1	3									38
First Time Resolution	0	9									90

Notes for use of the Cause and Effect Matrix template, also known as the XY Matrix:

1. Weight the Output Variables (Y's) on a scale of 1 to 10 with 10 indicating most important to the Customer.
2. For Root Cause Analysis, assign the association/effect multiplier score for each X to Y using a scale of 0, 1, 3, 9, where 0 = None, 1 = Weak, 3 = Moderate, and 9 = Strong. Initially this assignment will likely be a team subjective assessment. Data should be collected and the degree of association should be validated with Graphical and Statistical Tools.
3. For Project Selection or Solution Selection, assign the association multiplier score for each X to Y using a scale of 1 to 10, with 10 indicating strong association.

Click the Pareto Chart button to create a Pareto Chart of Weighted Scores by Input/Process Variables:



Failure Mode & Effects Analysis (FMEA) Example

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Failure Mode & Effects Analysis (FMEA)** template.

The following example is given in **SigmaXL > Help > Template Examples > FMEA**. This is a simple Failure Mode and Effects example for Stocking Inventory. If prompted, please ensure that macros are enabled in order to allow sort by **RPN**.

Potential Failure Mode & Effects Analysis

Process/Product:	Stock Inventory
FMEA Team:	
Responsibility:	
Prepared By:	

Row Number Sort	Process								
	Process Steps or Product Functions	Potential Failure Mode	Potential Effects of Failure	Severity (1-10)	Potential Cause(s) of Failure	Occurrence (1-10)	Current Controls	Detection (1-10)	Risk Priority Number (RPN) Sort
1	Stock Inventory	Stock in wrong location	Unable to locate Stock	5	Correct location is full	7	Stock checked twice a year	9	315
2	Stock Inventory	Damaged	Insufficient	7	Supplier Defect	3	Incoming	8	168
3	Stock Inventory	Damaged	Insufficient product	7	Handling Error	5	Standard Operating	9	315

If you hover the mouse cursor over the **Severity**, **Occurrence** or **Detection** heading the recommended scale will appear as a comment:

Score	Severity Guidelines		
	Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma	
10	**Failure to Meet Safety and/or Regulatory Requirements:** May endanger operator (machine or assembly) without warning.	Injure a customer or employee	Bad
9	**Failure to Meet Safety and/or Regulatory Requirements:** May endanger operator (machine or assembly) with warning.	Be illegal	
8	**Major Disruption:** 100% of product may have to be scrapped. Line shutdown or stop ship.	Render product or service unfit for use	
7	**Significant Disruption:** A portion of the production run may have to be scrapped. Deviation from primary process including decreased line speed or added manpower.	Cause extreme customer dissatisfaction	
6	**Moderate Disruption:** 100% of production run may have to be reworked off line and accepted.	Result in partial malfunction	
5	**Moderate Disruption:** A portion of the production run may have to be reworked off line and accepted.	Cause a loss of performance which is likely to result in a complaint	
4	**Moderate Disruption:** 100% of production run may have to be reworked in station before it is processed.	Cause minor performance loss	
3	**Moderate Disruption:** A portion of the production run may have to be reworked in station before it is processed.	Cause a minor nuisance but can be overcome with no performance loss	
2	**Minor Disruption:** Slight inconvenience to process, operation, or operator.	Be unnoticed and have only minor effect on performance	
1	**No Effect:** No discernible effect.	Be unnoticed and not affect the performance	Good

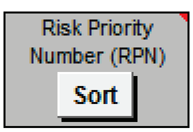
The recommended scales for Severity, Occurrence, and Detection are shown below:

Score	Severity Guidelines		
	Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma	
10	**Failure to Meet Safety and/or Regulatory Requirements:** May endanger operator (machine or assembly) without warning.	Injure a customer or employee	Bad
9	**Failure to Meet Safety and/or Regulatory Requirements:** May endanger operator (machine or assembly) with warning.	Be illegal	
8	**Major Disruption:** 100% of product may have to be scrapped. Line shutdown or stop ship.	Render product or service unfit for use	
7	**Significant Disruption:** A portion of the production run may have to be scrapped. Deviation from primary process including decreased line speed or added manpower.	Cause extreme customer dissatisfaction	
6	**Moderate Disruption:** 100% of production run may have to be reworked off line and accepted.	Result in partial malfunction	
5	**Moderate Disruption:** A portion of the production run may have to be reworked off line and accepted.	Cause a loss of performance which is likely to result in a complaint	
4	**Moderate Disruption:** 100% of production run may have to be reworked in station before it is processed.	Cause minor performance loss	
3	**Moderate Disruption:** A portion of the production run may have to be reworked in station before it is processed.	Cause a minor nuisance but can be overcome with no performance loss	
2	**Minor Disruption:** Slight inconvenience to process, operation, or operator.	Be unnoticed and have only minor effect on performance	
1	**No Effect:** No discernible effect.	Be unnoticed and not affect the performance	Good

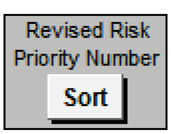
Score	Occurrence Guidelines			
	Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma		
10	**Very High:** ≥ 100 per thousand; ≥ 1 in 10	More than once per day	> 30%	Bad
9	**High:** 50 per thousand; 1 in 20	Once every 3-4 days	< 30%	
8	**High:** 20 per thousand; 1 in 50	Once every week	< 5%	
7	**High:** 10 per thousand; 1 in 100	Once per month	< 1%	
6	**Moderate:** 2 per thousand; 1 in 500	Once every 3 months	< 0.03%	
5	**Moderate:** 0.5 per thousand; 1 in 2,000	Once every 6 months	< 1 per 10,000	
4	**Moderate:** 0.1 per thousand; 1 in 10,000	Once per year	< 6 per 100,000	
3	**Low:** 0.01 per thousand; 1 in 100,000	Once every 1-3 years	< 6 per million	
2	**Low:** ≤ 0.001 per thousand; 1 in 1,000,000	Once every 3-6 years	< 3 per 10 million	
1	**Very Low:** Failure is eliminated through preventive control.	Once every 6-100 years	< 2 per billion	Good

Score	Detection Guidelines Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma	
10	**Almost Impossible**. No detection opportunity: No current process control; cannot detect or is not analyzed.	Defect caused by failure is not detectable	Bad
9	**Very Remote**. Not likely to detect at any stage: Failure Mode and/or Error (Cause) is not easily detected (e.g., random audits).	Occasional units are checked for defects	
8	**Remote**. Problem Detection Post Processing: Failure Mode detection post-processing by operator through visual/tactile/audible means.	Units are systematically sampled and inspected	
7	**Very Low**. Problem Detection at Source: Failure Mode detection in-station by operator through visual/tactile/audible means or post-processing through use of attribute gauging (go/no-go, manual torque check/clicker wrench, etc.).	All units are manually inspected	
6	**Low**. Problem Detection Post Processing: Failure Mode detection post-processing by operator through use of variable gauging or in-station by operator through use of attribute gauging (go/no-go, manual torque check/clicker wrench, etc.).	Manual inspection with mistake-proofing modifications	
5	**Moderate**. Problem Detection at Source: Failure Mode or Error (Cause) detection in-station by operator through use of variable gauging or by automated controls in-station that will detect discrepant part and notify operator (light, buzzer, etc.). Gauging performed on setup and first-piece check (for set-up causes only).	Process is monitored (SPC) and manually inspected	
4	**Moderately High**. Problem Detection Post Processing: Failure Mode detection post-processing by automated controls that will detect discrepant part and lock part to prevent further processing.	SPC is used with an immediate reaction to out of control conditions	
3	**High**. Problem Detection at Source: Failure Mode detection in-station by automated controls that will detect discrepant part and automatically lock part in station to prevent further processing.	SPC as above, 100% inspection surrounding out of control conditions	
2	**Very High**. Error Detection and/or Problem Prevention: Error (Cause) detection in-station by automated controls that will detect error and prevent discrepant part from being made.	All units are automatically inspected	
1	**Almost Certain**. Detection not applicable; Error Prevention: Error (Cause) prevention as a result of fixture design, machine design or part design. Discrepant parts cannot be made because item has been error-proofed by process/product design.	Defect is obvious and can be kept from affecting the customer	Good

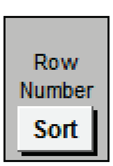
Click **Risk Priority Number (RPN) Sort** to sort the rows by RPN in descending order.



Click **Revised Risk Priority Number Sort** to sort the rows by Revised RPN in descending order.

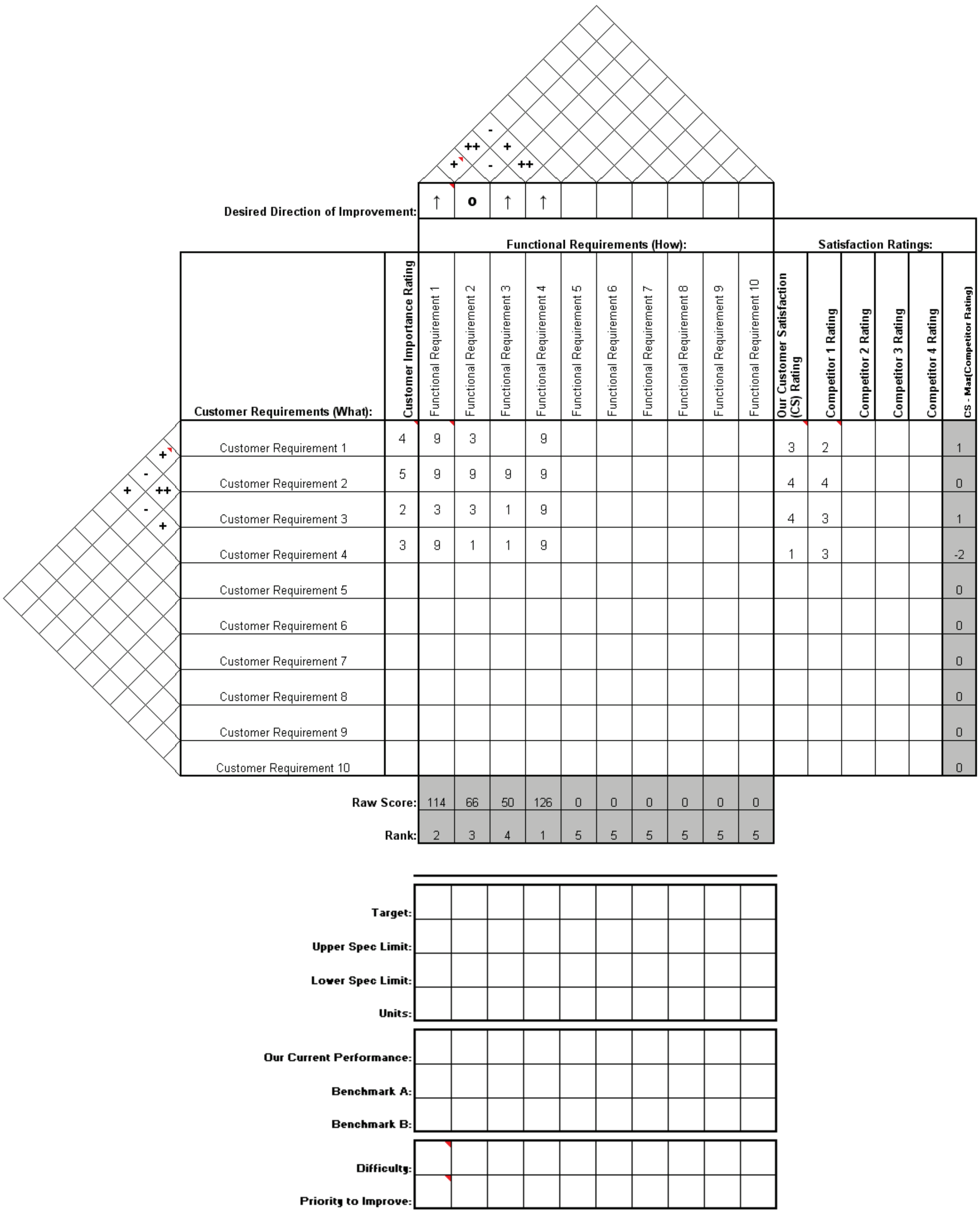


Click **Row Number Sort** to restore the FMEA worksheet to the original row order (ascending).



Quality Function Deployment (QFD) Example

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Quality Function Deployment (QFD) > 10 by 10 QFD** (or 20 by 20, 30 by 30) to access the Quality Function Deployment (QFD) template. Enter data as shown.



Desired Direction of Improvement:		↑	o	↑	↑												
		Functional Requirements (How):										**Satisfaction Ratings:**					
Customer Requirements (What):	**Customer Importance Rating**	Functional Requirement 1	Functional Requirement 2	Functional Requirement 3	Functional Requirement 4	Functional Requirement 5	Functional Requirement 6	Functional Requirement 7	Functional Requirement 8	Functional Requirement 9	Functional Requirement 10	**Our Customer Satisfaction (CS) Rating**	**Competitor 1 Rating**	**Competitor 2 Rating**	**Competitor 3 Rating**	**Competitor 4 Rating**	**CS - Max(Competitor Rating)**
Customer Requirement 1	4	9	3		9							3	2				1
Customer Requirement 2	5	9	9	9	9							4	4				0
Customer Requirement 3	2	3	3	1	9							4	3				1
Customer Requirement 4	3	9	1	1	9							1	3				-2
Customer Requirement 5																	0
Customer Requirement 6																	0
Customer Requirement 7																	0
Customer Requirement 8																	0
Customer Requirement 9																	0
Customer Requirement 10																	0
Raw Score:		114	66	50	126	0	0	0	0	0	0						
Rank:		2	3	4	1	5	5	5	5	5	5						

Target:										
Upper Spec Limit:										
Lower Spec Limit:										
Units:										
Our Current Performance:										
Benchmark A:										
Benchmark B:										
Difficulty:										
Priority to Improve:										

Scroll down to view the QFD drop-down lists. The values in yellow highlight may be modified, which changes the available options for the drop-down lists.

Correlations	
Strong Positive	++
Positive	+
None	
Negative	-
Strong Negative	--

Direction of Improvement	
Maximize	↑
Target	o
Minimize	↓

Relationships	
None	
Weak	1
Moderate	3
Strong	9

Customer Importance Rating	
Low	1
	2
Medium	3
	4
High	5

Satisfaction Ratings	
Low	1
	2
Medium	3
	4
High	5

Difficulty	
Very Easy	1
	2
	3
	4
Very Difficult	5

Priority To Improve	
Low Priority	1
	2
Medium Priority	3
	4
High Priority	5

Pugh Concept Selection Matrix Example

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Pugh Concept Selection Matrix** to access the Pugh Concept Selection Matrix template. Enter data as shown.

Key Criteria	Weight	Concept A	Concept B	Concept C	Concept D	Concept E	Concept F	Concept G	Concept H	Concept I	Concept J	Current Baseline Datum
Criterion 1	4	+	S	-								S
Criterion 2	5	+	S	S								S
Criterion 3	3	S	+	-								S
Criterion 4	2	+	S	S								S
Criterion 5	5	+	+	S								S
Criterion 6												S
Criterion 7												S
Criterion 8												S
Criterion 9												S
Criterion 10												S
Criterion 11												S
Criterion 12												S
Criterion 13												S
Criterion 14												S
Criterion 15												S
Criterion 16												S
Criterion 17												S
Criterion 18												S
Criterion 19												S
Criterion 20												S

Sum of Positives (+):	4	2	0	0	0	0	0	0	0	0
Sum of Negatives(-):	0	0	2	0	0	0	0	0	0	0
Sum of Sames (S):	1	3	3	0	0	0	0	0	0	0
Positives - Negatives:	4	2	-2	0	0	0	0	0	0	0

Weighted Sum of Positives (+):	16	8	0	0	0	0	0	0	0	0
Weighted Sum of Negatives (-):	0	0	7	0	0	0	0	0	0	0
Weighted Sum of Sames (S):	3	11	12	0	0	0	0	0	0	0
Weighted Positives - Weighted Negatives:	16	8	-7	0	0	0	0	0	0	0

The **Current Baseline Datum** is a reference column provided for convenience; it is not used in the calculations.

Scroll down to view the Weights and Concept Selection drop-down lists. The values in yellow highlight may be modified, which changes the available options for the drop-down lists.

Weights	
Low	1
	2
Medium	3
	4
High	5

Concept Selection	
Positive (1)	+
Same (0)	S
Negative (-1)	-

Control Plan

Click **SigmaXL > Templates & Calculators > DMAIC & DFSS Templates > Control Plan** to access the Control Plan template.

CONTROL PLAN

Project Name:	
Process Description:	
Date:	
Revision:	
Prepared By:	
Approved By:	

Process Step	Key Indicator	X (control) or Y (monitor)	Product/Process Specifications/Target	Evaluation/ Measurement Technique	%P/Total (R&R) %P/Tolerance	Sample Size	Sample Frequency	Responsibility	Control Method	Contingency Action Plan	Misc. Information

Lean – Takt Time Calculator Example

Click **SigmaXL > Templates & Calculators > Lean > Takt Time Calculator** to access the Takt Time calculator. The template gives the following default example:

SigmaXL Lean Templates: Takt Time Calculator

Daily Customer Demand:	units per day	22
Scheduled Work:	hours per shift	8
Shifts per Day:		2
Lunch:	minutes per shift	30
Breaks:	minutes per shift	30
Planned Downtime:	minutes per shift	30
Staff/Operator Cycle Time:	minutes per unit	226
Available Time:	minutes per day	780.0
Takt Time:	minutes per unit	35.5
Required Number of Staff/Operators:		6.4

Lean – Value Analysis/Process Load Balance

Click **SigmaXL > Templates & Calculators > Lean > Value Analysis/Process Load Balance** to access the Value Analysis/Process Load Balance template. The template gives the following default example:

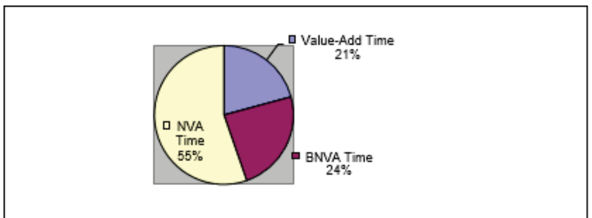
SigmaXL Lean Templates: Value Analysis/Process Load Balance

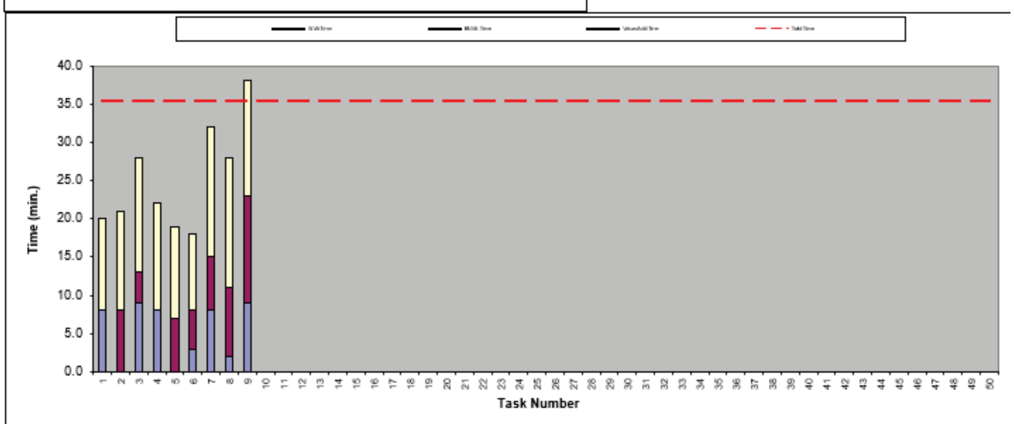
Process Name:	
Date:	
Prepared By:	
Notes:	

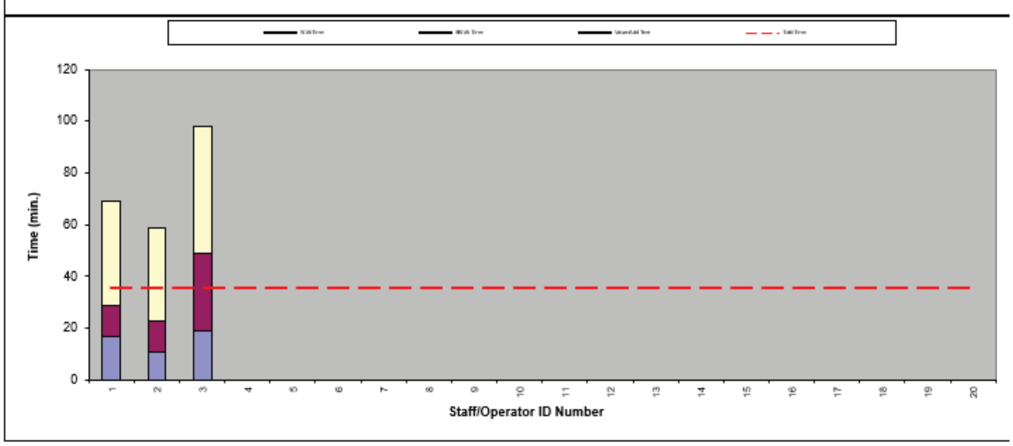
SigmaXL Copyright © 2008

Enter Takt Time:	35.5

Task Number	Staff/Operator ID	Task Description	Value-Add Time	BNVA Time	NVA Time	Total Time	Comments
1	1	Task 1	8.0	0.0	12.0	20.0	
2	1	Task 2	0.0	8.0	13.0	21.0	
3	1	Task 3	9.0	4.0	15.0	28.0	
4	2	Task 4	8.0	0.0	14.0	22.0	
5	2	Task 5	0.0	7.0	12.0	19.0	
6	2	Task 6	3.0	5.0	10.0	18.0	
7	3	Task 7	8.0	7.0	17.0	32.0	
8	3	Task 8	2.0	9.0	17.0	28.0	
9	3	Task 9	9.0	14.0	15.0	38.0	
10						0.0	

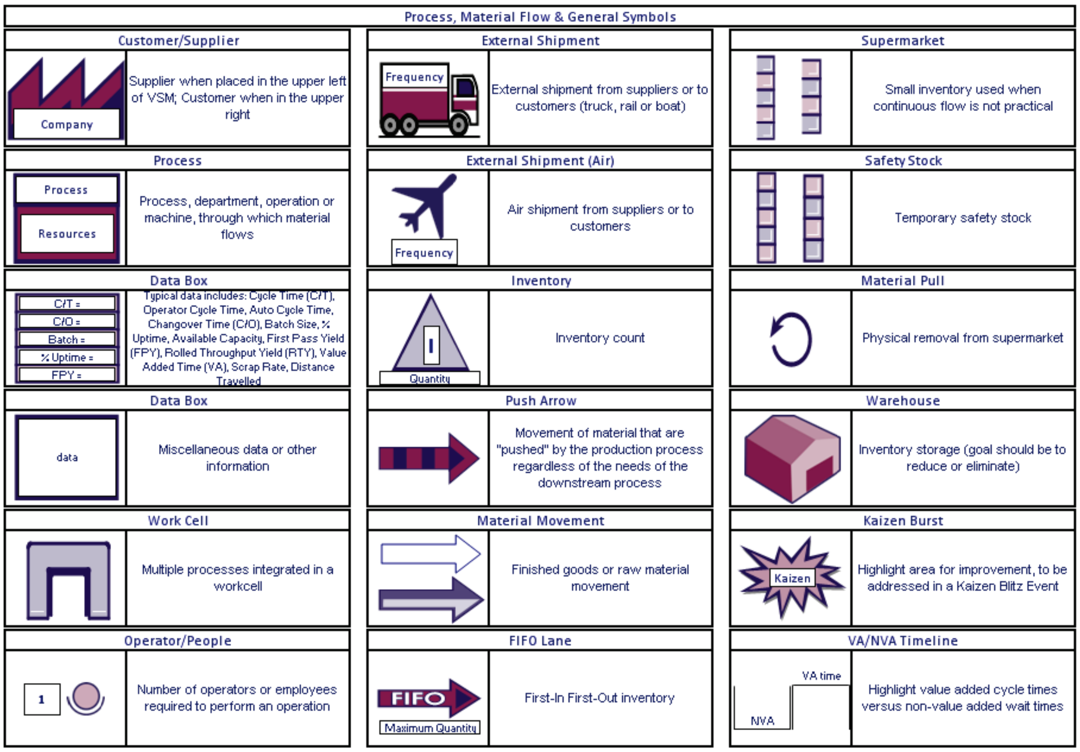






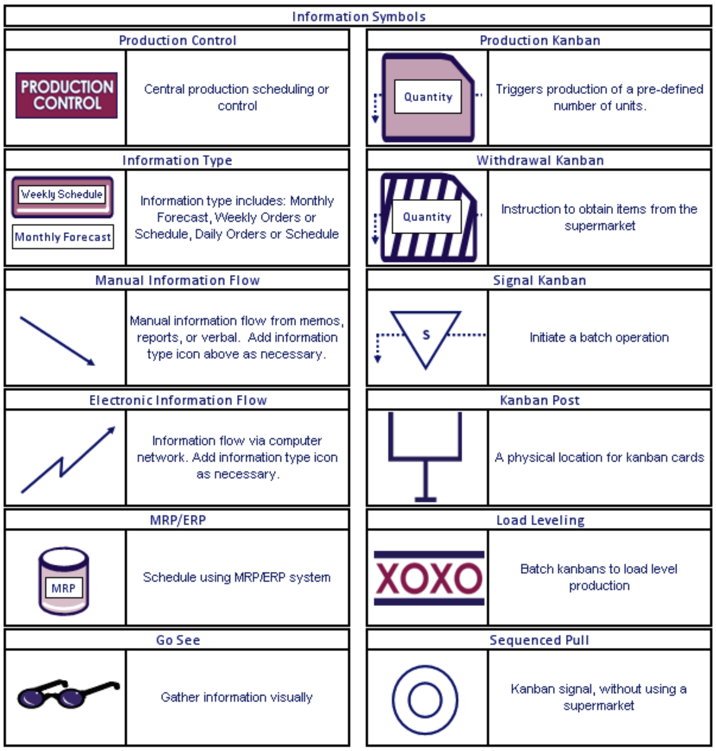
Lean – Value Stream Mapping

Click **SigmaXL > Templates & Calculators > Lean > Value Stream Mapping** to access the Value Stream Mapping template. The example on the following page is given in **SigmaXL > Help > Template Examples > Value Stream Mapping.**



Process, Material Flow & General Symbols

Customer/Supplier	Company	Supplier when placed in the upper left of VSM; Customer when in the upper right
Process	Process; Resources	Process, department, operation or machine, through which material flows
Data Box	C/T =; C/O =; Batch =; % Uptime =; FPY =	Typical data includes: Cycle Time (C/T), Operator Cycle Time, Auto Cycle Time, Changover Time (C/O), Batch Size, % Uptime, Available Capacity, First Pass Yield (FPY), Rolled Throughput Yield (RTY), Value Added Time (VA), Scrap Rate, Distance Travelled
Data Box	data	Miscellaneous data or other information
Work Cell		Multiple processes integrated in a workcell
Operator/People	1	Number of operators or employees required to perform an operation
External Shipment	Frequency	External shipment from suppliers or to customers (truck, rail or boat)
External Shipment (Air)	Frequency	Air shipment from suppliers or to customers
Inventory	I; Quantity	Inventory count
Push Arrow		Movement of material that are "pushed" by the production process regardless of the needs of the downstream process
Material Movement		Finished goods or raw material movement
FIFO Lane	FIFO; Maximum Quantity	First-In First-Out inventory
Supermarket		Small inventory used when continuous flow is not practical
Safety Stock		Temporary safety stock
Material Pull		Physical removal from supermarket
Warehouse		Inventory storage (goal should be to reduce or eliminate)
Kaizen Burst	Kaizen	Highlight area for improvement, to be addressed in a Kaizen Blitz Event
VA/NVA Timeline	VA time; NVA	Highlight value added cycle times versus non-value added wait times



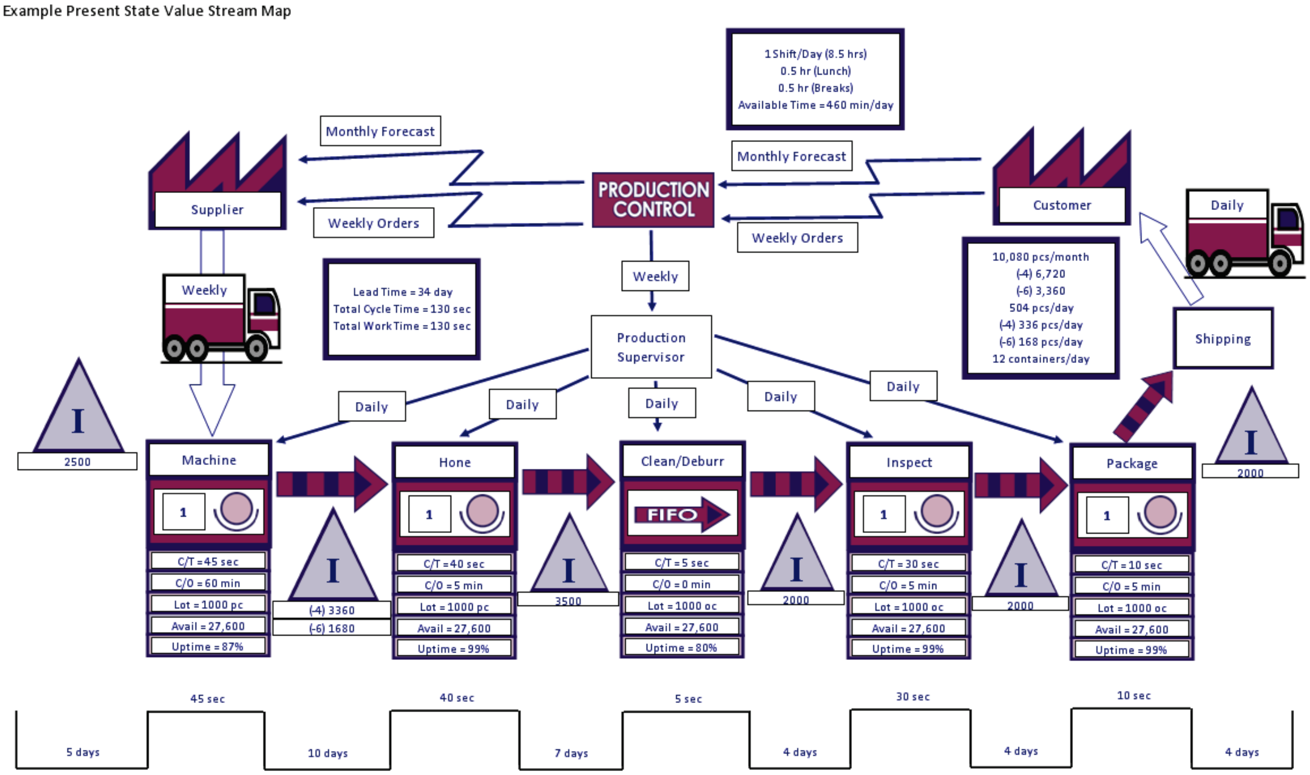
Notes:

1. Copy and Paste the above symbols into a blank worksheet as needed to create the value stream map. Resize if necessary.

2. Recommended resource for Value Stream Mapping: Quarterman Lee and Brad Snyder, "The Strategos Guide to Value Stream & Process Mapping", Enna Products Corp., 2006.

This is a present state value stream map of a manufacturing operation. This example is courtesy and copyright of Strategos, Inc. www.strategosinc.com.

Example Present State Value Stream Map



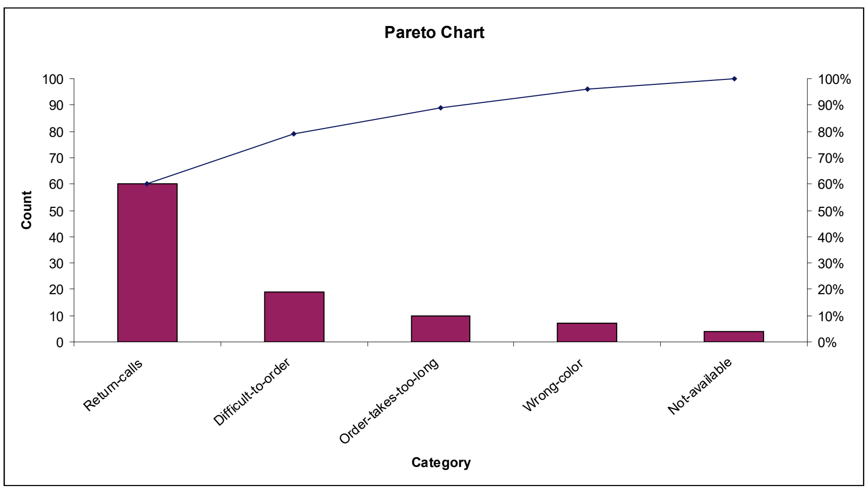
Basic Graphical Templates – Pareto Example

Click **SigmaXL > Templates & Calculators > Basic Graphical Templates > Pareto Chart**. Enter data as shown:

Category	Count
Difficult-to-order	19
Not-available	4
Order-takes-too-long	10
Return-calls	60
Wrong-color	7

Pareto Chart

Click the **Pareto Chart** button to produce the Pareto Chart:

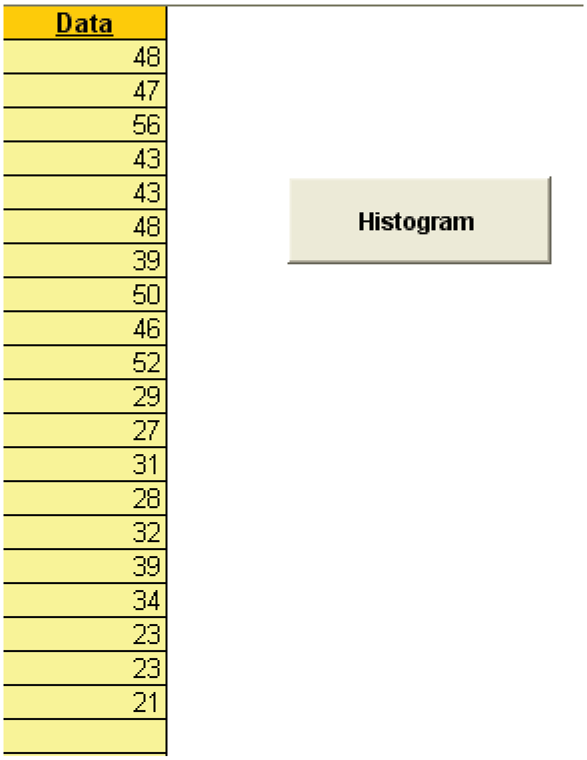


Notes:

1. This Pareto Chart template should be used with count data like number of errors. You can also create Pareto charts with cost data.
2. You can replace the **Category** and **Count** column headings with any headings that you wish.
3. Enter the Pareto categories in the **Category** column. These can be Name, Location, Error Type or other text information. Pareto categories are required and will appear on the horizontal X-Axis of the Pareto Chart.
4. Enter your count (or cost) data in the **Count** column.
5. Click the **Pareto Chart** button to create a Pareto Chart.

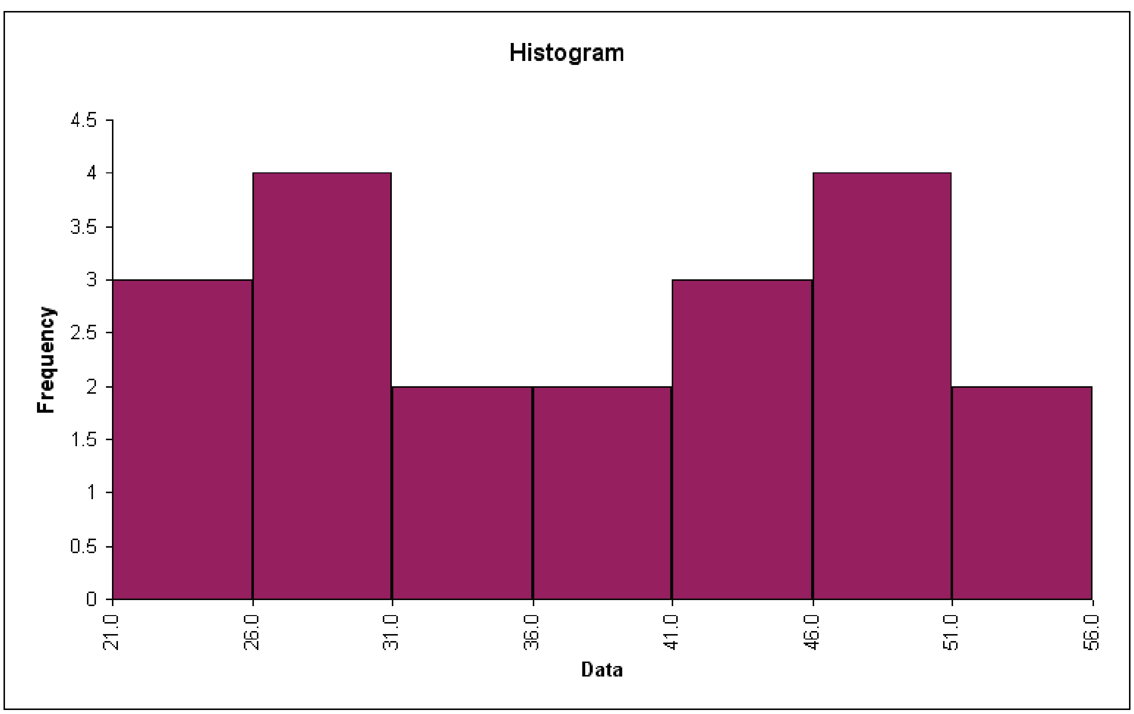
Basic Graphical Templates – Histogram Example

Click **SigmaXL > Templates & Calculators > Basic Graphical Templates > Histogram**. Enter data as shown:



Data
48
47
56
43
43
48
39
50
46
52
29
27
31
28
32
39
34
23
23
21

Click the **Histogram** button to produce the Histogram:



Notes:

1. This Histogram template should be used with continuous data like cycle time.
2. You can replace the **Data** column heading with any heading that you wish.
3. Enter your data in the **Data** column.
4. Click the **Histogram** button to create a Histogram.

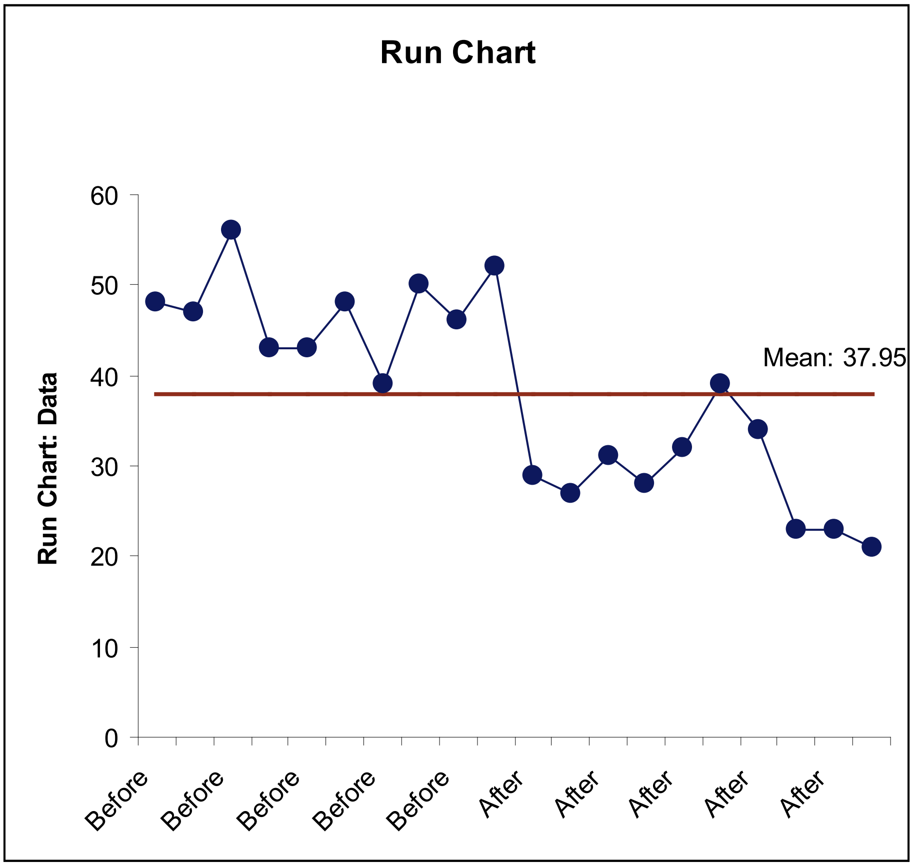
Basic Graphical Templates – Run Chart Example

Click **SigmaXL > Templates & Calculators > Basic Graphical Templates > Run Chart**. Enter the data as shown (copy and paste the histogram data):

Before/After	Data
Before	48
Before	47
Before	56
Before	43
Before	43
Before	48
Before	39
Before	50
Before	46
Before	52
After	29
After	27
After	31
After	28
After	32
After	39
After	34
After	23
After	23
After	21

Run Chart

Click the **Run Chart** button to produce the Run Chart:



Notes:

1. This Run Chart template can be used with a variety of data types: continuous data like cycle time, count data like number of errors, or cost data. The data must be in chronological time-sequence order.
2. You can replace the **X-Axis Label** and **Data** column headings with any headings that you wish.
3. Enter your data in the **Data** column.
4. Enter labels in **X-Axis Label** column. Labels can be Date, Time, Name, or other text information. These labels are optional and will appear on the horizontal X-Axis of the Run Chart.
5. Click the **Run Chart** button to create a Run Chart.

Basic Statistical Templates – Sample Size – Discrete

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > Sample Size – Discrete** to access the Sample Size – Discrete calculator. The template gives the following default example:

Sample Size Calculator - Discrete Data		
Sample Data (user inputs):		
Estimate of Proportion	P	0.5
Desired margin of error	delta / half-interval	0.03
Population Size (optional)	N	
Confidence level (enter .95 for 95%)	100*(1-α)%	95.0%
Results:		
Minimum Sample Size	n	1068
	n (adjusted for small N)	
	np check (should be ≥ 5)	534

Notes for use of the Sample Size – Discrete Calculator:

1. P is estimate of proportion for outcome of interest. Use P = 0.5 if unknown.
2. Delta is desired proportion margin of error. Enter as the half-width, i.e. if the desired margin of error is +/- 3%, enter 0.03.
3. Enter population size N to adjust for small populations (N < 10000).
4. np should be >= 5. If necessary, reduce delta to adjust.
5. Power (1 – Beta) is not considered in these calculations. Power and Sample Size may be calculated using **SigmaXL > Statistical Tools > Power & Sample Size Calculators**.

This sample size calculator is based on a confidence interval approach, where you enter the desired half-interval value (delta). Let's say that you desire the proportion margin of error to be +/- 3%. As noted above, you simply enter a delta (half-interval) value of .03.

The challenge with this sample size calculator is that we need an estimate of the population proportion. This is a bit of a chicken and egg situation – which comes first? I want to calculate an appropriate sample size to determine an outcome, but I am being asked to enter an estimate of the population proportion. First, keep in mind that this tool is a planning tool – the true confidence intervals will be determined after you collect your data. Estimating the population proportion can be done if you have good historical data to draw from, for example, in historical customer surveys the percentage of satisfied customers was 80%. In this case, you could use P=0.8. If you do not have *a priori* knowledge, then leave P=0.5 which gives the most conservative value (i.e., largest estimate of sample size). If you enter a value other than 0.5 it will result in a smaller sample size requirement.

Basic Statistical Templates – Sample Size – Continuous

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > Sample Size – Continuous** to access the Sample Size – Continuous calculator. The template gives the following default example:

Sample Size Calculator - Continuous Data		
Sample Data (user inputs):		
Estimate of Standard Deviation	S	1
Desired margin of error	delta / half-interval	0.25
Population Size (optional)	N	
Confidence level (enter .95 for 95%)	100*(1-α)%	95.0%
Results:		
Minimum Sample Size	n	62
	n (adjusted for small N)	

Notes for use of the Sample Size – Continuous Calculator:

1. Delta uses the same units as the standard deviation. Enter as the half-width, i.e., if the desired margin of error is +/- 0.25, enter 0.25.
2. Enter (optional) population size N to adjust for small populations (N < 1000).
3. Power (1 – Beta) is not considered in these calculations. Power and Sample Size may be calculated using **SigmaXL > Statistical Tools > Power & Sample Size Calculators**.

Similar to the Sample Size – Discrete calculator, this sample size calculator is based on a confidence interval approach, where you enter the desired half-interval value (delta). If a survey had responses that were on a continuous scale of 1 to 5, and you desired a margin of error on the mean to be +/- .25, then you would use .25 as the delta value. Note that in this continuous case the units are not percentages but level of satisfaction.

Here we need an *a priori* estimate of the population standard deviation. If you have no idea what the standard deviation will be, then you could take a small sample to get a rough estimate of the standard deviation.

Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA

It is well known that the central limit theorem enables the t-Test and ANOVA to be fairly robust to the assumption of normality. A question that invariably arises is, “How large does the sample size have to be?” A popular rule of thumb answer for the one sample t-Test is “n = 30.” While this rule of thumb often does work well, the sample size may be too large or too small depending on the degree of nonnormality as measured by the Skewness and Kurtosis. Furthermore, it is not applicable to a One Sided t-Test, 2 Sample t-Test or One-Way ANOVA.

To address this issue, we have developed a unique template that gives a minimum sample size needed for a hypothesis test to be robust.

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > Minimum Sample Size for Robust t-Tests and ANOVA** to access this template. It includes minimum sample size for robustness for the 1 Sample t-Test, 2 Sample t-Test and the One-Way ANOVA.

The user may specify the alternative hypothesis as “Less Than” (one sided), “Not Equal To” (two sided) or “Greater Than” (one sided). Confidence levels of 90% (α = 0.1), 95% (α = .05) or 99% (α = .01) may also be specified:

Minimum Sample Size for Robust Hypothesis Testing		
Sample Data (user inputs):		
Hypothesis Test:		1 Sample t-Test
Alternative Hypothesis :	Ha	Not Equal To
Confidence Level:	100*(1-α)%	95%
Skewness:	Skew	1
Kurtosis:	Kurt	-0.48
Results:		
Minimum sample size for each sample/group:	n	30

To use the template, simply select the appropriate **Hypothesis Test**, **Alternative Hypothesis** and **Confidence Level** using the drop-down selection. Enter **Skewness** and **Kurtosis** values as shown in the yellow highlighted cells.

Note that in the example shown, the rule of thumb for a 1 Sample t-Test “n = 30” is confirmed with a moderate skew value of 1.

Now change the **Alternative Hypothesis** to "Less Than" and **Confidence Level** to 99% as shown:

Minimum Sample Size for Robust Hypothesis Testing		
Sample Data (user inputs):		
Hypothesis Test:		1 Sample t-Test
Alternative Hypothesis :	Ha	Less Than
Confidence Level:	100*(1-α)%	99%
Skewness:	Skew	1
Kurtosis:	Kurt	-0.48
Results:		
Minimum sample size for each sample/group:	n	752

The minimum sample size required for robustness is now 752!

On the other hand, if you want to perform a standard One-Way ANOVA, enter the values as shown:

Minimum Sample Size for Robust Hypothesis Testing		
Sample Data (user inputs):		
Hypothesis Test:		3 Sample One Way ANOVA
Alternative Hypothesis :	Ha	Not Equal To
Confidence Level:	100*(1-α)%	95%
Skewness:	Skew	1
Kurtosis:	Kurt	-0.48
Results:		
Minimum sample size for each sample/group:	n	3

Now the minimum sample size requirement is only 3. This value applies to each sample or group, so for the 3 Sample ANOVA that would mean each sample has *n* = 3 for a total number of observations = 9.

Note that this calculator is strictly addressing the question of alpha robustness to nonnormality. Power is not considered here.

If the minimum sample size requirements cannot be met, you should use a nonparametric equivalent to the parametric hypothesis test (i.e. One Sample Sign or Wilcoxon, Two Sample Mann-Whitney, Kruskal-Wallis or Mood's Median: SigmaXL > Statistical Tools > Nonparametric Tests).

Skewness and Kurtosis

Sample Skewness and Kurtosis values can be obtained from SigmaXL's descriptive statistics: SigmaXL > Statistical Tools > Descriptive Statistics.

A slight Skew is +/- 0.5, moderate Skew is +/- 1, severe Skew is +/- 2 and extreme Skew is +/- 5. The Skewness range used should be -5 to +5. Values beyond this range are extrapolated so may be inaccurate.

Kurtosis for a normal distribution is 0. Kurtosis must be greater than or equal to:
(Skew ^ 2 - 1.48).

Kurtosis "delta" is Kurt - Skew^2 and is used in the regression equation. Kurtosis delta range should be -1.48 to +1.48. A kurtosis delta less than -1.48 denotes a bimodal distribution so this is a lower boundary. Values above +1.48 are extrapolated so may be inaccurate.

The calculator assumes that all samples have the same Skewness and Kurtosis.

Monte Carlo Simulation

The data for minimum sample size formulas are derived from extensive Monte Carlo simulations for *n* = 2 to 2000. Observed alpha values were determined empirically from the P-Values of 100,000 replicate hypothesis tests for each *n*. Nonnormal data with Skew = -5 to +5 and Kurt delta = -1.48 to +1.48 was generated using the Pearson Family function (see SigmaXL DiscoverSim Workbook: Appendix for details).

Minimum sample size for robustness occurs when the simulated observed alpha is within +/- 20% of the specified alpha (Bradley 1980 and Rhiel 1996).

Regression Models

A separate regression model was constructed for each hypothesis test, alternative hypothesis, and confidence level (total of 21 models), using coding as shown.

	Coded for Regression	Notes for Regression Coding
1 Sample t-Test	1	1=1 Sample t; 2=2 Sample t; 3=3 Sample ANOVA
Not Equal To	2	1=LT for +ve Skew or GT for -ve Skew; 2=NE; 3=GT for +ve Skew or LT for -ve Skew
95%	2	1=90% (α=0.1); 2=95% (α=.05); 3=99% (α=.01)
1	1	Skew^2
-0.48	-1.48	Kurt_delta = Kurt - Skew^2. Minimum = -1.48.

The model predictor terms are Skewness^2 and Kurtosis Delta. The response is *n* minimum. The term coefficient values are stored in the template as shown.

Hypothesis Test:	1	1	1	1	1
Alternative Hypothesis :	1	1	1	2	2
Confidence Level:	1	2	3	1	2
Description:	1 Sample t; Ha = LT +, GT -; α=0.1	1 Sample t; Ha = LT +, GT -; α=.05	1 Sample t; Ha = LT +, GT -; α=.01	1 Sample t; Ha = NE; α=0.1	1 Sample t; Ha = NE; α=.05

Predictor Term	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant:	2.682	-2.940	-257.06	-0.658037	-0.864317
Skew^2:	44.257	137.32	1008.2	9.440	22.848
Kurt Delta:	-2.431	-10.157	-0.564855	-1.356	-5.280

Regression model n:	51	149	752	11	30
Selected n:	0	0	0	0	30
Model R-Square:	99.48%	99.34%	99.93%	99.81%	99.85%

Model R-Square values are typically over 99%, with some exceptions (96%) due to small estimated sample sizes.

The model used is the one which matches the selected hypothesis test, alternative hypothesis, and confidence level. The selected *n* is highlighted in red and is also displayed in the template Results cell. If *n* > 2000, a text display "> 2000" is shown in the Results cell.

References

Boos, D.D. and Hughes-Oliver, J.M. (2000), "How Large Does *n* Have to be for Z and t Intervals?" *The American Statistician*, 54(2), 121-128.

Bradley, J. V. (1980), "Nonrobustness in Z, t, and F Tests at Large Sample Sizes," *Bulletin of the Psychonomics Society*, 16(5), 333-336.

Rhiel, G. S., and Chaffin, W. W. (1996), "An Investigation of the Large-Sample/Small-Sample Approach to the One-Sample Test for a Mean (Sigma Unknown)," *Journal of Statistics Education,* 4, No. 3 (www.amstat.org/publications/jse).

SigmaXL, Inc., DiscoverSim Version 1.1 Workbook.

Basic Statistical Templates – 1 Sample Z-Test and Confidence Interval for Mean

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample Z-Test and Confidence Interval for Mean** to access the 1 Sample Z-Test calculator. The template gives the following default example.

1 Sample Z - Test and Confidence Interval for Mean

Sample Data (user inputs):		
Sample Size	n	30
Sample Mean	x-bar	1
Known Population Standard Deviation	σ	1
Null Hypothesis (hypothesized mean)	H_0: Mean (μ) =	0.5
Alternative Hypothesis	H_a: Mean (μ)	Not Equal To
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%

Results:	
Standard Error Mean	0.1826
DF	29
Z-statistic	2.7386
alpha	0.0500
P-Value (2 sided)	0.0062
Upper Confidence Limit (2 Sided)	1.3578
Lower Confidence Limit (2 Sided)	0.6422

Test Information	
Null Hypothesis H_0: Mean (μ) = 0.5	Reject
Alternative Hypothesis H_a: Mean (μ) ≠ 0.5	Conclude true at 95.0% confidence level

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of obtaining a Z-statistic at least as extreme as 2.7386, given that the null hypothesis is true. Since the P-Value is less than alpha (0.05), we reject the null hypothesis: Mean (μ) = 0.5 and conclude that the alternative hypothesis Ha: Mean (μ) ≠ 0.5 is true at the confidence level of 95.0%.
We are 95% confident that the true population mean lies within the interval (0.6422 to 1.3578).

Notes:

1. Enter summarized Sample Data, Known Population Standard Deviation, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Alternative Hypothesis using drop-down.

3. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

Basic Statistical Templates – 1 Sample t-Test and Confidence Interval for Mean

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample t-Test and Confidence Interval for Mean** to access the 1 Sample t-Test calculator. The template gives the following default example.

1 Sample t - Test and Confidence Interval for Mean		
Sample Data (user inputs):		
Sample Size	n	30
Sample Mean	x-bar	1
Sample Standard Deviation	s	1
Null Hypothesis (hypothesized mean)	H_0: Mean (μ) =	0.5
Alternative Hypothesis	H_a: Mean (μ)	Not Equal To
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%
Results:		
Standard Error Mean		0.1826
DF		29
t-statistic		2.7386
alpha		0.0500
P-Value (2 sided)		0.0104
Upper Confidence Limit (2 Sided)		1.3734
Lower Confidence Limit (2 Sided)		0.6266

Test Information	
Null Hypothesis H_0: Mean (μ) = 0.5	Reject
Alternative Hypothesis H_a: Mean (μ) ≠ 0.5	Conclude true at 95.0% confidence level

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of obtaining a t-statistic at least as extreme as 2.7386, given that the null hypothesis is true. Since the P-Value is less than alpha (0.05), we reject the null hypothesis: Mean (μ) = 0.5 and conclude that the alternative hypothesis Ha: Mean (μ) ≠ 0.5 is true at the confidence level of 95.0%.
We are 95% confident that the true population mean lies within the interval (0.6266 to 1.3734).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Alternative Hypothesis using drop-down.

3. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

Basic Statistical Templates – 2 Sample t-Test and Confidence Interval (Compare 2 Means)

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Sample t-Test & Confidence Interval (Compare 2 Means)** to access the 2 Sample t-Test calculator. The template gives the following default example.

2 Sample t-Test and Confidence Interval

Sample Data (user inputs):		Sample 1	Sample 2
Sample Size	n	10	10
Sample Mean	x-bar	0.0000	1.0000
Sample Standard Deviation	s	1.0000	1.0000
Null Hypothesis (hypothesized mean difference)	H_0: $\mu_1 - \mu_2$ =	0.0	
Alternative Hypothesis	H_a: $\mu_1 - \mu_2$	Not Equal To	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Assume Equal Variances		Yes	

Results:	
Sample Mean Difference	-1.0000
Standard Error Difference (using pooled variance)	0.4472
df (equal variance)	18
t-statistic	-2.2361
alpha	0.0500
P-Value (2-sided)	0.0382
Upper Confidence Limit (2 Sided)	-0.0604
Lower Confidence Limit (2 Sided)	-1.9396

Test Information	
Null Hypothesis H_0: Mean Difference (μ1 - μ2) = 0	Reject
Alternative Hypothesis H_a: Mean Difference (μ1 - μ2) ≠ 0	Conclude true at 95.0% confidence level

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of obtaining a t-statistic at least as extreme as -2.2361, given that the null hypothesis is true. Since the P-Value is less than alpha (0.05), we reject the null hypothesis H0: Mean Difference (μ1 - μ2) = 0 and conclude that the alternative hypothesis Ha: Mean Difference (μ1 - μ2) ≠ 0 is true at the confidence level of 95.0%.
We are 95% confident that the true population mean difference lies within the interval (-1.9396 to -0.0604).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Alternative Hypothesis and "Assume Equal Variances" using drop-down.

3. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

Basic Statistical Templates – 1 Sample Equivalence Test for Mean

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample Equivalence Test for Mean** to access the 1 Sample Equivalence Test (Two One-Sided t-Tests TOST) calculator. The template gives the following default example.

1 Sample Equivalence Test for Mean (Two One-Sided t-Tests TOST)		
Sample Data (user inputs):		
Sample Size	n	30
Sample Mean	x-bar	0
Sample Standard Deviation	s	1
Target	T	0
Upper Equivalence Limit	UEL	0.5
Lower Equivalence Limit	LEL	-0.5
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%
Results:		
Standard Error Mean		0.1826
DF		29
Difference (Mean - Target)		0
alpha		0.0500
t1-statistic (upper)		-2.7386
t2-statistic (lower)		2.7386
P1-Value (upper)		0.0052
P2-Value (lower)		0.0052
Equivalence P-Value (maximum of P1 and P2)		0.0052
Upper Confidence Limit Difference		0.3102
Lower Confidence Limit Difference		-0.3102

Interpretation of Equivalence P-Value and Confidence Intervals
Since the Equivalence P-Value is less than alpha (0.05), conclude that equivalence is true.
Since the 95% confidence interval is within the equivalence interval, conclude that equivalence is true.

Notes:

1. Enter summarized Sample Data, Target, Upper & Lower Equivalence Limits and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Null hypothesis for P1: Difference >= UEL; Null hypothesis for P2: Difference <= LEL. Both null hypotheses must be rejected to conclude that equivalence is true.

3. UEL and LEL establish the zone or region of equivalence and are determined by what size difference is considered practically significant.

4. The Upper and Lower Confidence Limit Difference values use the 100*(1-α)% method by Hsu, which is generally more powerful than the classical 100*(1-2α)% method. Upper is the maximum of UC, (UEL+LEL)/2 and Lower is the minimum of LC, (UEL+LEL)/2. For details, see Hsu, J.C., Hwang, J.T.G., Liu, H. K., and Ruberg, S. J. (1994), "Confidence Intervals Associated with Tests for Bioequivalence," *Biometrika* 81, 103-114.

5. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

Basic Statistical Templates – 2 Sample Equivalence Test (Compare 2 Means)

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Sample Equivalence Test (Compare 2 Means)** to access the 2 Sample Equivalence Test (Two One-Sided t-Tests TOST) calculator. The template gives the following default example.

2 Sample Equivalence Test (Two One-Sided t-Tests TOST)

Sample Data (user inputs):		Test/Treatment Sample	Reference/Control Sample
Sample Size	n	30	30
Sample Mean	x-bar	0.0000	0.5000
Sample Standard Deviation	s	1.0000	1.0000
Upper Equivalence Limit	UEL	1.0	
Lower Equivalence Limit	LEL	-1.0	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Assume Equal Variances		Yes	

Results:	
Sample Mean Difference	-0.5000
Standard Error Difference (using pooled variance)	0.2582
df (equal variance)	58
alpha	0.0500
t1-statistic (upper)	-5.8095
t2-statistic (lower)	1.9365
P1-Value (upper)	0.0000
P2-Value (lower)	0.0288
Equivalence P-Value (maximum of P1 and P2)	0.0288
Upper Confidence Limit Mean Difference	0.0000
Lower Confidence Limit Mean Difference	-0.9316

Interpretation of Equivalence P-Value and Confidence Intervals
Since the Equivalence P-Value is less than alpha (0.05), conclude that equivalence is true.
Since the 95% confidence interval is within the equivalence interval, conclude that equivalence is true.

Notes:

1. Enter summarized Sample Data, Upper & Lower Equivalence Limits and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select "Assume Equal Variances" using drop-down.

3. Null hypothesis for P1: Mean Difference >= UEL; Null hypothesis for P2: Mean Difference <= LEL. Both null hypotheses must be rejected to conclude that equivalence is true.

4. UEL and LEL establish the zone or region of equivalence and are determined by what size mean difference is considered practically significant.

5. The Upper and Lower Confidence Limit Mean Difference values use the 100*(1-α)% method by Hsu, which is generally more powerful than the classical 100*(1-2α)% method. Upper is the maximum of UC, (UEL+LEL)/2 and Lower is the minimum of LC, (UEL+LEL)/2. For details, see Hsu, J.C., Hwang, J.T.G., Liu, H. K., and Ruberg, S. J. (1994), "*Confidence Intervals Associated with Tests for Bioequivalence,*" *Biometrika* 81, 103-114.

Basic Statistical Templates – 1 Sample Chi-Square Test and CI for Standard Deviation

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Sample Chi-Square Test and CI for Standard Deviation** to access the 1 Sample Chi-Square-Test for Standard Deviation calculator. The template gives the following default example.

1 Sample Standard Deviation Chi-Square Test and Confidence Interval

Sample Data (user inputs):		
Sample Size	n	10
Sample Standard Deviation	s	1
Null Hypothesis (hypothesized StDev)	H_0: Sigma (σ) =	0.5
Alternative Hypothesis	H_a: Sigma (σ)	Not Equal To
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%

Results:	
Sample Variance	1.0000
DF	9
Chi-Square test statistic	36.0000
alpha	0.0500
P-Value (2 sided)	0.0001
Upper Confidence Limit (2 Sided)	1.8256
Lower Confidence Limit (2 Sided)	0.6878

Test Information	
Null Hypothesis H_0: Mean Difference (μ1 - μ2) = 0	Reject
Alternative Hypothesis H_a: Mean Difference (μ1 - μ2) ≠ 0	Conclude true at 95.0% confidence level

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of obtaining a test statistic at least as extreme as 36.0000, given that the null hypothesis is true. Since the P-Value is less than alpha (0.05), we reject the null hypothesis: Sigma (σ) = 0.5 and conclude that the alternative hypothesis Ha: Sigma (σ) ≠ 0.5 is true at the confidence level of 95.0%.
We are 95% confident that the true population standard deviation lies within the interval (0.6878 to 1.8256).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.
2. Select Alternative Hypothesis using drop-down.
3. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

Basic Statistical Templates – 2 Sample F-Test and CI (Compare 2 Standard Deviations)

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Sample F-Test and CI (Compare 2 Standard Deviations)** to access the 2 Sample F-Test calculator. The template gives the following default example.

2 Sample Standard Deviations / Variances F-Test and Ratio CI

Sample Data (user inputs):		Sample 1	Sample 2
Sample Size	n	10	10
Sample Standard Deviation	s	2.0000	1.0000
Null Hypothesis (hypothesized ratio of StdDevs)	H_0: σ_1/σ_2 =	1.0	
Alternative Hypothesis	H_a: σ_1/σ_2	Not Equal To	
Confidence Level (enter .95 for 95%)		95.0%	

Results:		
Ratio of Standard Deviations	2.0000	
Sample Variance	4.0000	1.0000
Ratio of Variances	4.0000	
df	9	9
F-statistic	4.0000	
alpha	0.0500	
P-Value (2-sided)	0.0510	
Upper Confidence Limit (2 Sided)	4.0130	
Lower Confidence Limit (2 Sided)	0.9968	

Test Information	
Null Hypothesis H_0: $\sigma1/\sigma2 = 1$	Fail to Reject
Alternative Hypothesis H_a: $\sigma1/\sigma2 \neq 1$	

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of obtaining an F-statistic at least as extreme as 4.0000, given that the null hypothesis is true. Since the P-Value is greater than alpha (0.05), we fail to reject the null hypothesis H0: σ1/σ2 = 1.
We are 95% confident that the true population sigma ratio lies within the interval (0.9968 to 4.0130).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Alternative Hypothesis using drop-down.

3. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

4. The F-test for two sample standard deviations/variances assumes that samples are normally distributed.

Basic Statistical Templates – 1 Proportion Test and Confidence Interval

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Proportion Test and Confidence Interval** to access the 1 Proportion Test calculator. The template gives the following default example.

1 Proportion Test and Confidence Interval

Sample Data (user inputs):		
Number of Events	x	1
Sample Size	n	10
Null Hypothesis (hypothesized proportion)	H_0: Proportion =	0.5
Alternative Hypothesis	H_a: Proportion	Not Equal To
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%
Hypothesis Test Method		Binomial Exact
Confidence Interval Method		Exact (Clopper-Pearson Beta)

Results:	
Sample proportion (x/n)	0.1000
alpha	0.0500
npq (npq should be >= 5 for normal approximation; q = 1-p)	0.9000
Z-statistic (normal)	-2.5298
Binomial exact probability P-Value (2-sided)	0.0215
Upper Confidence Limit (2-sided)	0.4450
Lower Confidence Limit (2-sided)	0.0025

Test Information	
Null Hypothesis H_0: Proportion (p) = 0.5	Reject
Alternative Hypothesis H_a: Proportion (p) ≠ 0.5	Conclude true at 95.0% confidence level

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of observing a sample proportion at least as extreme as 0.1000, given that the null hypothesis is true. Since the P-Value is less than alpha (0.05), we reject the null hypothesis: Proportion (p) = 0.5 and conclude that the alternative hypothesis: Proportion (p) ≠ 0.5 is true at the confidence level of 95.0%.
We are 95% confident that the true population proportion lies within the interval (0.0025 to 0.4450).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.
2. Select Alternative Hypothesis and Methods using drop-down. If x = 0, it is recommended that the Alternative Hypothesis be set to one-sided "Less Than." If x = n, set the Alternative Hypothesis to one-sided "Greater Than."

3. Hypothesis Test Method:

   - Binomial Exact is recommended.

   - Use Normal Approximation only when n is large and proportion is close to 0.5 (check that np >= 5 and nq >= 5, or the more conservative npq, shown above, >= 5). See Rosner (2010).

4. Confidence Interval Method (recommendations based on literature review, see references below):

   - Confidence intervals for binomial proportions have an "oscillation" phenomenon where the coverage probability varies with n and p.

   - Exact (Clopper-Pearson) is strictly conservative and will guarantee the specified confidence level as a minimum coverage probability, but results in wide intervals. Recommended only for applications requiring strictly conservative intervals.

   - Wilson Score and Jeffreys Beta have mean coverage probability matching the specified confidence interval. We recommend Wilson for two-sided confidence intervals and Jeffreys for one-sided.

   - Normal Approximation (Wald) is not recommended, but included here to validate hand calculations. Use only when n is large and proportion is close to 0.5 (check that npq, shown above, >= 5).

5. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

REFERENCES

1. Agresti, A. and Coull, B.A. (1998), Approximate is Better than “Exact” for Interval Estimation of Binomial Proportions. *The American Statistician*, 52,119–126.

2. Brown, L. D., Cai, T. T. and DasGupta, A. (2001), Interval estimation for a binomial proportion. *Statistical Science*, 16, 101-133. With comments and a rejoinder by the authors.

3. Cai, T. T. (2005), One-sided confidence intervals in discrete distributions. *J. Statist. Plann. Inference*, 131, 63-88.

4. Clopper, C.J., and Pearson, E.S. (1934), The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26, 404–413.

5. Newcombe, R. (1998a), Two-sided confidence intervals for the single proportion: Comparison of seven methods, *Statistics in Medicine*, 17, 857-872.

6. Rosner, B. (2010), *Fundamentals of Biostatistics*. Seventh Edition. Equation 5.15, page 133. Duxbury Press.

7. Vollset, S.E. (1993). Confidence intervals for a binomial proportion, *Statistics in Medicine*, 12, 809–824.

Basic Statistical Templates – 2 Proportions Test and Confidence Interval

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Proportions Test and Confidence Interval** to access the 2 Proportions Test calculator. The template gives the following default example.

2 Proportions Test and Confidence Interval

Sample Data (user inputs):		Sample 1	Sample 2
Number of Events	x	1	2
Sample Size	n	10	10
Null Hypothesis (hypothesized difference)	H_0: P_1 - P_2 =	0	
Alternative Hypothesis	H_a: P_1 - P_2	Not Equal To	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Hypothesis Test Method		Fisher's Exact	
Confidence Interval Method		Newcombe-Wilson Score	
Results:			
Sample proportion (x/n)		0.1000	0.2000
Sample proportion difference		-0.1000	
alpha		0.0500	
Minimum expected value (should be >= 5 for normal approximation)		1.5000	
Fisher's Exact probability P-Value (2-sided)		1.0000	
Upper Confidence Limit (2-sided)		0.2362	
Lower Confidence Limit (2-sided)		-0.4205	

Test Information		
Null Hypothesis H_0:	$P_1 - P_2 = 0$	Fail to Reject
Alternative Hypothesis H_a:	$P_1 - P_2 \neq 0$	

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of observing a difference in proportions at least as extreme as -0.1000, given that the null hypothesis is true. Since the P-Value is greater than alpha (0.05), we fail to reject the null hypothesis H0: P1 - P2 = 0.
We are 95% confident that the true difference in population proportions lies within the interval (-0.4205 to 0.2362).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Alternative Hypothesis and Methods using drop-down.

3. Hypothesis Test Method:

   - Fisher's Exact is recommended, but is only valid for H0: P1 - P2 = 0.
   - Use Normal Approximation only when minimum expected cell value, shown above, is >= 5. Unpooled is used if H0: P1 - P2 <> 0. Use pooled if H0: P1 - P2 = 0. (See Peltier).

4. Confidence Interval Method:

   - Confidence intervals for difference in binomial proportions have an "oscillation" phenomenon where the coverage probability varies with n and p.
   - Newcombe-Wilson Score is recommended and has a mean coverage probability that is close to the specified confidence interval. (See Newcombe method 10).
   - Newcombe-Wilson Score (CC = Continuity Corrected) is conservative and will typically meet the specified confidence level as a minimum coverage probability, but results in wider intervals. (See Newcombe method 11).
   - Jeffreys-Perks is similar to Newcombe-Wilson, and included as an option because it is the preferred interval for some practitioners. (See Radhakrishna; Newcombe method 4).
   - Normal Approximation is not recommended, but included here to validate hand calculations. Use only when minimum expected cell value, shown above, is >= 5.
   - Exact methods that are strictly conservative (like Clopper-Pearson for the one proportion case) are not included in this template because they are computationally intensive and slow.

5. Due to the complexity of calculations, this template uses vba macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

REFERENCES

1. Beal, S. L. (1987), Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics*, 43, 941-950.

2. Newcombe, R. G. (1998b), Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine,* 17:873–890.

3. Peltier, C., Why Do We Pool for the Two-Proportion z-Test? http://apcentral.collegeboard.com/apc/members/courses/teachers_corner/49013.html.

4. Radhakrishna S., Murthy B.N., Nair N.G., Jayabal P., Jayasri R. (1992), Confidence intervals in medical research. *Indian J Med Res.*, Jun;96:199-205.

Basic Statistical Templates – 2 Proportions Equivalence Test

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Proportions Equivalence Test** to access the 2 Proportions Equivalence Test (Two One-Sided Tests TOST) calculator. The template gives the following default example.

2 Proportions Equivalence Test (Two One-Sided Tests TOST)			
Sample Data (user inputs):		Test/Treatment Sample	Reference/Control Sample
Number of Events	x	10	11
Sample Size	n	100	100
Upper Equivalence Limit	UEL	0.1	
Lower Equivalence Limit	LEL	-0.1	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Confidence Interval Method		Newcombe-Wilson Score	
Results:			
Sample proportion (x/n)		0.1000	0.1100
Sample proportion difference		-0.0100	
alpha		0.0500	
Minimum expected value (should be >= 5 for normal approximation)		10.5000	
Z1-statistic (upper, normal unpooled)		-2.5376	
Z2-statistic (lower, normal unpooled)		2.0762	
P1-Value (upper)		0.0056	
P2-Value (lower)		0.0189	
Equivalence P-Value (maximum of P1 and P2)		0.0189	
Upper Confidence Limit Proportion Difference		0.0633	
Lower Confidence Limit Proportion Difference		-0.0835	

Interpretation of Equivalence P-Value and Confidence Intervals
Since the Equivalence P-Value is less than alpha (0.05), conclude that equivalence is true.
Since the 95% confidence interval is within the equivalence interval, conclude that equivalence is true.

Notes:

1. Enter summarized Sample Data, Upper & Lower Equivalence Limits and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Select Confidence Interval Method using drop-down.

   - Confidence intervals for difference in binomial proportions have an "oscillation" phenomenon where the coverage probability varies with n and p.
   - Newcombe-Wilson Score is recommended and has a mean coverage probability that is close to the specified confidence interval. (See Newcombe method 10).
   - Newcombe-Wilson Score (CC = Continuity Corrected) is conservative and will typically meet the specified confidence level as a minimum coverage probability, but results in wider intervals. (See Newcombe method 11).
   - Jeffreys-Perks is similar to Newcombe-Wilson, and included as an option because it is the preferred interval for some practitioners. (See Radhakrishna; Newcombe method 4).
   - Normal Approximation will match the P-Value calculations, but should only be used when minimum expected cell value, shown above, is >= 5.

- Exact methods that are strictly conservative (like Clopper-Pearson for the one proportion case) are not included in this template because they are computationally intensive and slow.

3. Null hypothesis for P1: Proportion Difference >= UEL; Null hypothesis for P2: Proportion Difference <= LEL. Both null hypotheses must be rejected to conclude that equivalence is true. The P-Values are based on the normal approximation unpooled method.

4. UEL and LEL establish the zone or region of equivalence and are determined by what size proportion difference is considered practically significant.

5. Since the confidence interval options use different methods than the hypothesis test (except for Normal option), it is possible that the conclusion from the Equivalence P-Value will be different from that of the confidence interval. In this case, we recommend using just the confidence interval method.

6. The Upper and Lower Confidence Limit Proportion Difference values use the 100*(1-α)% method by Hsu, which is generally more powerful than the classical 100*(1-2α)% method. Upper is the maximum of UC, (UEL+LEL)/2 and Lower is the minimum of LC, (UEL+LEL)/2.

7. Due to the complexity of calculations, this template uses VBA macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

REFERENCES

1. Beal, S. L. (1987), Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics*, 43, 941-950.

2. Hsu, J.C., Hwang, J.T.G., Liu, H. K., and Ruberg, S. J. (1994), Confidence Intervals Associated with Tests for Bioequivalence, *Biometrika* 81, 103-114.

3. Newcombe, R. G. (1998b), Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine,* 17:873–890.

4. Radhakrishna S., Murthy B.N., Nair N.G., Jayabal P., Jayasri R. (1992), Confidence intervals in medical research. *Indian J Med Res.*, Jun;96:199-205.

Basic Statistical Templates – 1 Poisson Rate Test and Confidence Interval

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 1 Poisson Rate Test and Confidence Interval** to access the 1 Poisson Rate calculator. The template gives the following default example.

1 Sample Poisson Rate Test and Confidence Interval		
Sample Data (user inputs):		
Number of occurrences	x	1
Sample Size (area of opportunity)	nt	10.0
Null Hypothesis (hypothesized rate)	H_0: Rate (λ) =	0.5
Alternative Hypothesis	H_a: Rate (λ)	Not Equal To
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%
Hypothesis Test Method		Poisson Exact
Confidence Interval Method		Exact (Garwood Chi-Square)
Results:		
Sample mean rate (x/(nt))		0.1000
alpha		0.0500
Z-statistic (normal)		-1.7889
Poisson exact probability P-Value (2-sided)		0.0809
Upper Confidence Limit (2-sided)		0.5572
Lower Confidence Limit (2-sided)		0.0025

Test Information	
Null Hypothesis H_0: Rate (λ) = 0.5	Fail to Reject
Alternative Hypothesis H_a: Rate (λ) ≠ 0.5	

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of observing a sample rate at least as extreme as 0.1000, given that the null hypothesis is true. Since the P-Value is greater than alpha (0.05), we fail to reject the null hypothesis: Rate (λ) = 0.5.
We are 95% confident that the true population rate lies within the interval (0.0025 to 0.5572).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Sample Size (area of opportunity) can be a count, time, length or other unit that defines the area of opportunity for an occurrence.

3. Select Alternative Hypothesis and Methods using drop-down. If x = 0, it is recommended that the Alternative Hypothesis be set to one-sided "Less Than."

4. Hypothesis Test Method:

    - Poisson Exact is recommended.
    - Use Normal Approximation only when the number of occurrences, x > 10.

5. Confidence Interval Method (recommendations based on literature review, see references below):

    - Confidence intervals for Poisson rates have an "oscillation" phenomenon where the coverage probability varies with x and n.
    - Exact (Garwood Chi-Square) is strictly conservative and will guarantee the specified confidence level as a minimum coverage probability, but results in wide intervals. Recommended only for applications requiring strictly conservative intervals.
    - Jeffreys has mean coverage probability matching the specified confidence interval and is recommended.
    - Normal Approximation (Wald) is not recommended, but included here to validate hand calculations. Use only when the number of occurrences, x > 10.

6. This calculator will automatically detect the version of Excel and use the appropriate statistical functions. If Excel 2010 or higher, then the Excel 2010 statistical functions are used.

REFERENCES

1. Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika* 28:437–442.

2. Swift, M.B. (2009). Comparison of confidence intervals for a Poisson Mean-Further considerations, *Communication in Statistics — Theory and Methods*, 38, 748–759.

Basic Statistical Templates – 2 Poisson Rates Test and Confidence Interval

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Poisson Rates Test and Confidence Interval** to access the 2 Poisson Rates calculator. The template gives the following default example.

2 Poisson Rates Test and Confidence Interval

Sample Data (user inputs):		Sample 1	Sample 2
Number of occurrences	x	1	2
Sample Size (area of opportunity)	nt	10.0	10.0
Null Hypothesis (hypothesized difference)	H_0: $Rate_1$ - $Rate_2$ ($\lambda_1 - \lambda_2$) =	0	
Alternative Hypothesis	H_a: $Rate_1$ - $Rate_2$ ($\lambda_1 - \lambda_2$)	Not Equal To	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Hypothesis Test Method		Exact (Binomial Method)	
Confidence Interval Method		Jeffreys Hybrid	

Results:		
Sample mean rate (x/(nt))	0.1000	0.2000
Sample rate difference	-0.1000	
alpha	0.0500	
Exact probability P-Value (2-sided)	1.0000	
Upper Confidence Limit (2-sided)	0.3001	
Lower Confidence Limit (2-sided)	-0.5505	

Test Information		
Null Hypothesis H_0:	$\lambda_1 - \lambda_2 = 0$	Fail to Reject
Alternative Hypothesis H_a:	$\lambda_1 - \lambda_2 \neq 0$	

Interpretation of P-Value and Confidence Intervals
The P-Value is the two-sided (two-tail) probability of observing a difference in rates at least as extreme as -0.1000, given that the null hypothesis is true. Since the P-Value is greater than alpha (0.05), we fail to reject the null hypothesis H0: λ1 - λ2 = 0.
We are 95% confident that the true difference in population rates lies within the interval (-0.5505 to 0.3001).

Notes:

1. Enter summarized Sample Data, Null Hypothesis and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Sample Size (area of opportunity) can be a count, time, length or other unit that defines the area of opportunity for an occurrence.

3. Select Alternative Hypothesis and Methods using drop-down.

4. Hypothesis Test Method:

   - Exact (Binomial Method) is recommended, but is only valid for H0: $\lambda1 - \lambda2 = 0$.
   - Use Normal Approximation only when the number of occurrences for each sample is greater than 10. Unpooled is used if H0: $\lambda1 - \lambda2 <> 0$. Use pooled if H0: $\lambda1 - \lambda2 = 0$.

5. Confidence Interval Method:

   - Confidence intervals for difference in Poisson rates have an "oscillation" phenomenon where the coverage probability varies with x and n.
   - Jeffreys Hybrid is recommended and has a mean coverage probability that is close to the specified confidence interval. (See Li, 2011).
   - Normal Approximation is not recommended, but included here to validate hand calculations. Use only when the number of occurrences for each sample is greater than 10.

6. Due to the complexity of calculations, this template uses vba macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

REFERENCES

1. Li, H.Q., Tang, M.L., Poon, W.Y. and Tang, N.S. (2011), Confidence intervals for difference between two Poisson rates, *Comm. Stat. Simul. Comput.* 40, pp. 1478–1491.

2. Przyborowski, J., Wilenski, H., (1940), Homogeneity of results in testing samples from Poisson series. *Biometrika* 31, 313–323.

Basic Statistical Templates – 2 Poisson Rates Equivalence Test

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Poisson Rates Equivalence Test** to access the 2 Poisson Rates Equivalence Test (Two One-Sided Tests TOST) calculator. The template gives the following default example.

2 Poisson Rates Equivalence Test (Two One-Sided Tests TOST)

Sample Data (user inputs):		Test/Treatment Sample	Reference/Control Sample
Number of occurrences	x	1	2
Sample Size (area of opportunity)	nt	10.0	10.0
Upper Equivalence Limit	UEL	0.5	
Lower Equivalence Limit	LEL	-0.5	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Confidence Interval Method		Jeffreys Hybrid	
Results:			
Sample mean rate (x/(nt))		0.1000	0.2000
Sample rate difference		-0.1000	
alpha		0.0500	
Z1-statistic (upper, normal unpooled)		-3.4641	
Z2-statistic (lower, normal unpooled)		2.3094	
P1-Value (upper)		0.0003	
P2-Value (lower)		0.0105	
Equivalence P-Value (maximum of P1 and P2)		0.0105	
Upper Confidence Limit Rate Difference		0.2239	
Lower Confidence Limit Rate Difference		-0.4630	

Interpretation of Equivalence P-Value and Confidence Intervals
Since the Equivalence P-Value is less than alpha (0.05), conclude that equivalence is true.
Since the 95% confidence interval is within the equivalence interval, conclude that equivalence is true.

Notes:

1. Enter summarized Sample Data, Upper & Lower Equivalence Limits and Confidence Level in cells with yellow highlight. Do not modify any other part of this worksheet.

2. Sample Size (area of opportunity) can be a count, time, length or other unit that defines the area of opportunity for an occurrence.

3. Select Confidence Interval Method using drop-down.

   - Confidence intervals for difference in Poisson rates have an "oscillation" phenomenon where the coverage probability varies with x and n.
   - Jeffreys Hybrid is recommended and has a mean coverage probability that is close to the specified confidence interval. (See Li, 2011).
   - Normal Approximation will match the P-Value calculations, but should only be used when the number of occurrences for each sample is greater than 10.

4. Null hypothesis for P1: Rate Difference >= UEL; Null hypothesis for P2: Rate Difference <= LEL. Both null hypotheses must be rejected to conclude that equivalence is true. The P-Values are based on the normal approximation unpooled method.

5. LEL and UEL establish the zone or region of equivalence and are determined by what size rate difference is considered practically significant.

6. Since the Jeffreys Hybrid confidence interval option uses a different method than the normal based hypothesis test, it is possible that the conclusion from the Equivalence P-Value will be different from that of the confidence interval. In this case, we recommend using just the confidence interval method.

7. The Upper and Lower Confidence Limit Rate Difference values use the 100*(1-α)% method by Hsu, which is generally more powerful than the classical 100*(1-2α)% method. Upper is the maximum of UC, (UEL+LEL)/2 and Lower is the minimum of LC, (UEL+LEL)/2.

8. Due to the complexity of calculations, this template uses VBA macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

REFERENCES

1. Hsu, J.C., Hwang, J.T.G., Liu, H. K., and Ruberg, S. J. (1994), Confidence Intervals Associated with Tests for Bioequivalence, *Biometrika* 81, 103-114.

2. Li, H.Q., Tang, M.L., Poon, W.Y. and Tang, N.S. (2011), Confidence intervals for difference between two Poisson rates, *Comm. Stat. Simul. Comput.* 40, pp. 1478–1491.

3. Przyborowski, J., Wilenski, H., (1940), Homogeneity of results in testing samples from Poisson series. *Biometrika* 31, 313–323.

Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > One-Way Chi-Square Goodness-of-Fit Test** to access the One-Way Chi-Square Goodness-of-Fit Test calculator. The template gives the following default example.

One-Way Chi-Square Test

Sample Data (user inputs):

Category ID	Observed Count	Specify Historical Count (optional)
1	1	1
2	2	2
3	3	3
4	10	4
5	10	5

Results:

Percentage of cells with expected count < 5	40.0%
Sum of Observed Counts	26
Chi-Square	3.4231
df	4
p-value	0.4897

Show Calculations

Notes:

1. Enter Category ID, Observed Count values and (optional) Historical Counts. Do not modify any other part of this worksheet.
2. If optional Historical Counts are not specified, chi-square is calculated using equal expected proportions.
3. If optional Historical Counts are specified, a value must be entered for each observed count.
4. Chi-Square statistic requires that no more than 20% of cells have an expected count less than 5.

Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test - Exact

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > One-Way Chi-Square Goodness-of-Fit Test - Exact** to access the One-Way Chi-Square Goodness-of-Fit Test – Exact calculator. The template gives the following default example.

One-Way Chi-Square Test

Sample Data (user inputs):		
Category ID	Observed Count	Specify Historical Count (optional)
1	1	1
2	2	2
3	3	3
4	10	4
5	10	5
6		

Results:	
Percentage of cells with expected count < 5	40.0%
Sum of Observed Counts	26
Chi-Square	3.4231
DF	4
Chi-Square P-Value	0.4897
Exact P-Value	0.5013

Show/Hide Calculations

Exact P-Value

Monte Carlo P-Value

Notes:

1. Enter Category ID, Observed Count values and (optional) Historical Counts. Do not modify any other part of this worksheet.

2. If optional Historical Counts are not specified, chi-square is calculated using equal expected proportions.

3. If optional Historical Counts are specified, a value must be entered for each observed count.

4. The Chi-Square statistic requires that no more than 20% of cells have an expected count less than 5 (and none of the cells have an expected count less than 1). If this assumption is not satisfied, the Chi-Square approximation may be invalid and Exact or Monte Carlo P-Values should be used. In this example 40% of the cells have an expected count less than 5 so Exact should be used.

5. This example shows that the Exact P-Value is 0.5013 and the "large sample" Chi-Square P-Value is 0.4897.

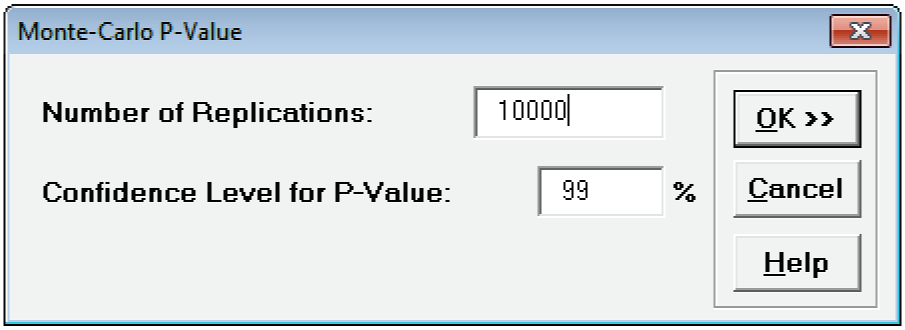
6. Chi-Square Exact solves the permutation problem using enhanced enumeration. For further details refer to the Appendix Exact and Monte Carlo P-Values for Nonparametric and Contingency Test.

7. It is important to note that while exact P-Values are "correct," they do not increase (or decrease) the power of a small sample test, so they are not a solution to the problem of failure to reject the null due to inadequate sample size.

8. For data that requires more computation time than specified, Monte Carlo P-Values provide an approximate (but unbiased) P-Value that typically matches exact to two decimal places using 10,000 replications. One million replications give a P-Value that is typically accurate to three decimal places. A confidence interval (99% default) is given for the Monte Carlo P-Values.

Monte Carlo Example:

1. The Exact P-Value for this example is solved very quickly, so Monte Carlo is not needed, but we will run it for continuity in the example. Click **Monte Carlo P-Value.** Select **Number of Replications** = 10000 and **Confidence Level for P-Value** = 99%.



Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

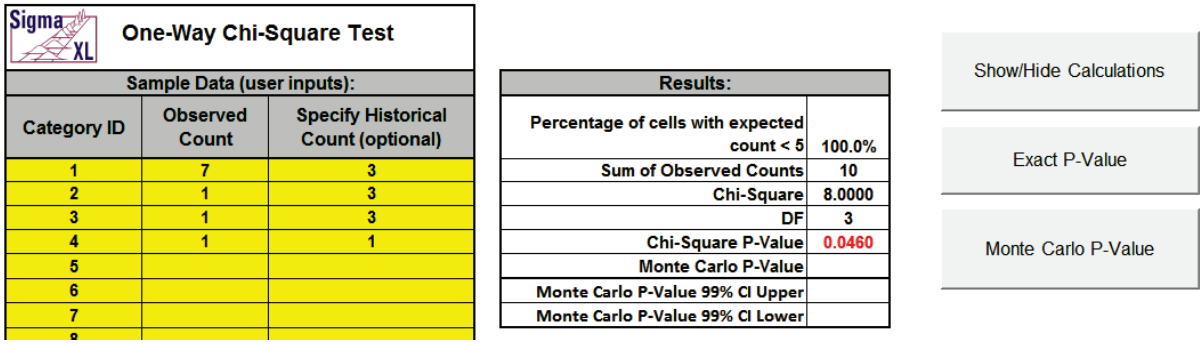
2. Click **OK**. Results:

Results:	
Percentage of cells with expected count < 5	40.0%
Sum of Observed Counts	26
Chi-Square	3.4231
DF	4
Chi-Square P-Value	0.4897
Monte Carlo P-Value	0.5012
Monte Carlo P-Value 99% CI Upper	0.5128
Monte Carlo P-Value 99% CI Lower	0.4896

The Monte Carlo P-Value here is 0.5012 with a 99% confidence interval of 0.4896 to 0.5128. This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). The true Exact P-Value = 0.5013 lies within this confidence interval.

Small Sample Exact Example:

3. Now we will consider a small sample problem. Enter the following values for sample data in the yellow highlight region. Note that the displayed Monte Carlo (or Exact) P-Values are cleared when new data is entered in the template:



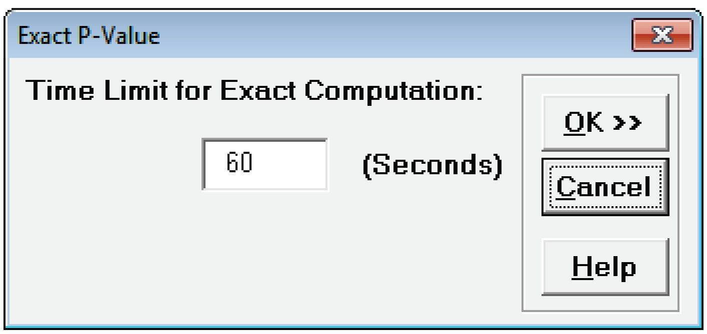
One-Way Chi-Square Test

Sample Data (user inputs):		
Category ID	Observed Count	Specify Historical Count (optional)
1	7	3
2	1	3
3	1	3
4	1	1
5		
6		
7		
8		

Results:	
Percentage of cells with expected count < 5	100.0%
Sum of Observed Counts	10
Chi-Square	8.0000
DF	3
Chi-Square P-Value	0.0460
Monte Carlo P-Value	
Monte Carlo P-Value 99% CI Upper	
Monte Carlo P-Value 99% CI Lower	

This example is adapted from Mehta, C.R. and Patel, N. R., *IBM SPSS Exact Tests*, IBM Corp., page 44.
ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/en/client/Manuals/IBM_SPSS_Exact_Tests.pdf

4. Click **Exact P-Value.** Select **Time Limit for Exact Computation** = 60 seconds**.**



5. Click **OK**. Results:

Results:	
Percentage of cells with expected count < 5	100.0%
Sum of Observed Counts	10
Chi-Square	8.0000
DF	3
Chi-Square P-Value	0.0460
Exact P-Value	0.0523

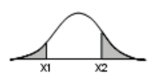
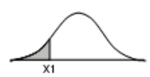
Note that the Exact P-Value is 0.0523 which is a “fail-to-reject” of the null hypothesis (H0), but the “large sample” or “asymptotic” Chi-Square P-Value incorrectly rejected H0 with a P-Value of 0.046. The exact P-Value matches that given in the reference. The error can also go the other way, where a large sample Chi-Square P-Value is a “fail-to-reject” of the null hypothesis and the Exact P-Value is a rejection of H0.

In conclusion, always use the Exact (or Monte Carlo) P-Value when the Chi-Square large sample assumptions are not met.

Probability Distribution Calculators – Normal

Click **SigmaXL > Templates & Calculators > Probability Distribution Calculators > Normal** to access the Normal Distribution Probability calculator. The template gives the following default example:

Normal Distribution Probability Calculator		
Input the following information:		
Mean	μ	**0**
Standard Deviation	σ	**1**
Lower Bound (or LSL)	**X1**	**1**
Upper Bound (or USL)	**X2**	**2**

Areas:	**Select the Probability of Interest from the following probabilities:**		
between X1 and X2		**Prob(1 ≤ X ≤ 2)**	**0.135905122**
outside X1 and X2		**Prob(X ≤ 1 AND X ≥ 2)**	**0.864094878**
below X1		**Prob(X ≤ 1)**	**0.841344746**
above X1		**Prob(X ≥ 1)**	**0.158655254**
below X2		**Prob(X ≤ 2)**	**0.977249868**
above X2		**Prob(X ≥ 2)**	**0.022750132**

Basic MSA Templates – Type 1 Gage Study

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Type 1 Gage Study** to access the Type 1 Gage Study template.

The following example is given in **SigmaXL > Help > Template Examples > MSA > Type 1 Gage Study**.

Type 1 Gage Study

Gage Name:	Example Type 1 Gage Study
Date of Study:	
Performed By:	
Notes:	

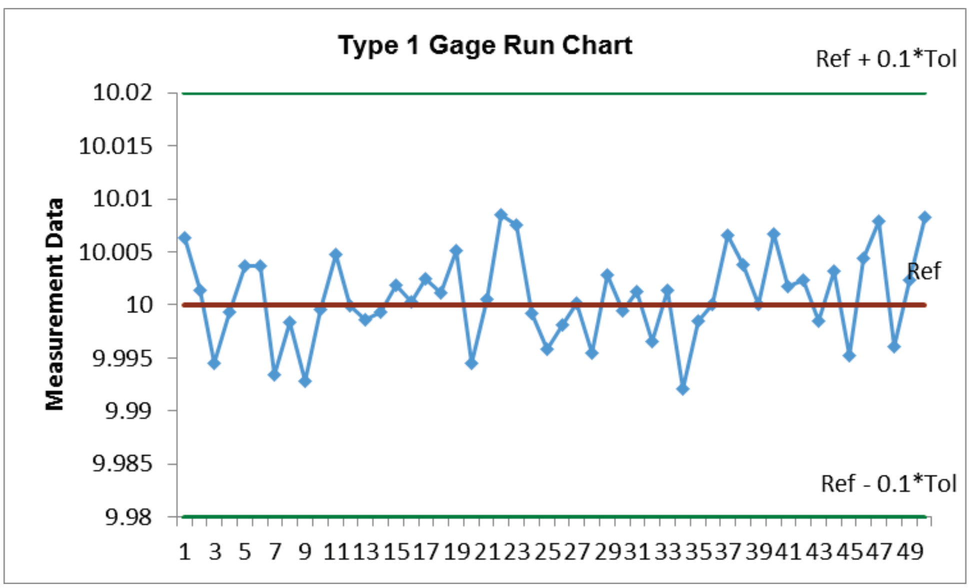
Tolerance (Tol, USL - LSL):	0.2
Part Reference Value (Ref):	10
Gage Resolution:	0.0001
StDev Multiplier:	6.00
Percent of Tolerance for Cg/Cgk:	20

Count	50
Mean	10.00071
StDev	0.004149
StDev*Multiplier (SV)	0.024893
Ref + 0.1*Tol	10.02
Ref - 0.1*Tol	9.98

Bias	0.00071
Bias t-Statistic	1.210106
Bias P-Value	0.2320
Gage Capability Cg	1.61
Gage Capability Cgk	1.55
Gage Repeatability % (100*SV/Tol)	12.45%
Gage Repeatability & Bias % (20/Cgk)	12.90%
Resolution % of Tolerance	0.05%

Trial:	Measurement Data
1	10.0063
2	10.0014
3	9.9945

Type 1 Gage Run Chart



Notes:

1. This calculator assumes that the data are normally distributed.
2. Enter your data in the **Measurement Data** column. Minimum recommended sample size is 25, but 50 is preferred. This template has a maximum sample size of 1000, suitable for automated test systems.
3. Enter **Tolerance** (Upper Specification Limit USL - Lower Specification Limit LSL).
4. Enter **Part Reference Value**. If missing, Gage Capability Cgk is not computed. Bias will assume a 0 value.
5. Enter **Gage Resolution**. If missing, Resolution % of Tolerance is not computed.
6. The default **StDev Multiplier** is 6 (99.73% coverage). Change this to 5.15 if AIAG convention is being used (99% coverage). Dietrich [1] recommends using 4 for approximate 95% coverage. We recommend 6 for typical and critical measurement systems, and 4 for non-critical.
7. The default **Percent of Tolerance** for Cg/Cgk is 20. Stacey [2] suggests using 10 for critical measurement systems.
8. If **Cg** and **Cgk** are >= 1.33, the measurement device is capable. If **Cg** or **Cgk** are < 1.33, the measurement device is not capable and needs to be improved.
9. If **Resolution % of Tolerance** (RE%) is <= 5%, the resolution is acceptable. If RE% is > 5% the resolution is not acceptable.
10. Click the **Type 1 Gage Run Chart** button to create a Run Chart with center line = Ref, upper limit = Ref + 0.1*Tol and lower limit = Ref - 0.1*Tol (assuming 20% of Tolerance).
11. This calculator requires Excel 2010 or higher.

REFERENCES

1. Dietrich, E. (2002), "Measurement System Capability" Reference Manual, http://www.q-das.com/uploads/tx_sbdownloader/Leitfaden_v21_me.pdf.
2. Stacey, J., Mathematics of Measurement Systems Analysis, Rolls-Royce plc., http://www.mei.org.uk/files/pdf/Mathematics_of_Measurement_Systems_Analysis.pdf

Basic MSA Templates – Gage Bias and Linearity Study

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Gage Bias and Linearity Study** to access the Gage Bias and Linearity Study template.

The following example is given in **SigmaXL > Help > Template Examples > MSA > Gage Bias and Linearity Study**.

Gage Bias and Linearity Study

Gage Name:	Example Gage Bias and Linearity Study - AIAG MSA Reference Manual, 4th Edition, Page 99
Date of Study:	
Performed By:	
Notes:	

Process Variation (SV from Gage R&R or 6 * Historical StDev):	10

Gage Linearity Report

Part:	1	2	3	4	5
Reference Value:	2	4	6	8	10
Trial					
1	2.7	5.1	5.8	7.6	9.1
2	2.5	3.9	5.7	7.7	9.3
3	2.4	4.2	5.9	7.8	9.5
4	2.5	5	5.9	7.7	9.3
5	2.7	3.8	6	7.8	9.4
6	2.3	3.9	6.1	7.8	9.5
7	2.5	3.9	6	7.8	9.5
8	2.5	3.9	6.1	7.7	9.5
9	2.4	3.9	6.4	7.8	9.6
10	2.4	4	6.3	7.5	9.2
11	2.6	4.1	6	7.6	9.3
12	2.4	3.8	6.1	7.7	9.4
13					
14					
15					
16					
17					
18					
19					
20					

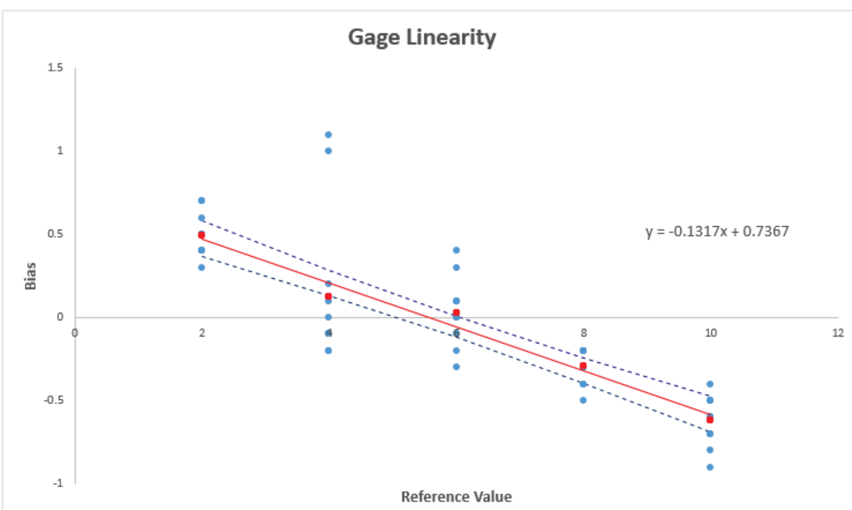
Part:	1	2	3	4	5
Reference Value:	2	4	6	8	10
Trial	Bias Calculations				
1	0.7	1.1	-0.2	-0.4	-0.9
2	0.5	-0.1	-0.3	-0.3	-0.7
3	0.4	0.2	-0.1	-0.2	-0.5
4	0.5	1	-0.1	-0.3	-0.7
5	0.7	-0.2	0	-0.2	-0.6
6	0.3	-0.1	0.1	-0.2	-0.5
7	0.5	-0.1	0	-0.2	-0.5
8	0.5	-0.1	0.1	-0.3	-0.5
9	0.4	-0.1	0.4	-0.2	-0.4
10	0.4	0	0.3	-0.5	-0.8
11	0.6	0.1	0	-0.4	-0.7
12	0.4	-0.2	0.1	-0.3	-0.6
13					
14					
15					
16					
17					
18					
19					
20					

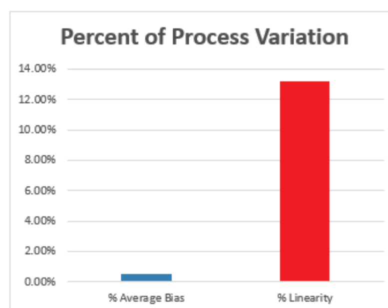
Bias	0.491666667	0.125	0.025	-0.291666667	-0.616666667
% Bias	4.92%	1.25%	0.25%	2.92%	6.17%
P-Value	0.0000	0.3540	0.6671	0.0000	0.0000

Average Bias	-0.053333333
% Average Bias	0.53%
P-Value	0.0894

	Coeff	SE Coeff	P-Value
Constant	0.736666667	0.072524273	0.0000
Slope	-0.131666667	0.010933445	0.0000

R-Square	71.43%
Adj, R-Square	70.94%
StDev	0.239539789
Linearity	1.316666667
% Linearity	13.17%





Notes:

1. Enter **Process Variation** (SV from Gage R&R or 6 * Historical StDev). If neither are available, use Process Tolerance. % Average Bias will be blank if not specified. Note that % Linearity does not require Process Variation.

2. Enter part **Reference Values**. These should be relative to a traceable standard, but if not available, see AIAG MSA Reference Manual [1], pp. 92-93 for suggestions on how to obtain reference values. Reference values should cover the operating range of the gage.

3. Enter part measurement data in the yellow highlight region below the respective part/reference value. AIAG recommends a minimum of 10 trials per part/reference. **It is important that the parts be selected at random for each trial in order to minimize appraiser "recall" bias.**

4. Click **Gage Linearity Report** button to generate the Linearity Report.

5. If a P-Value is < .05, it is highlighted in bold red to indicate significance. If the Average Bias P-Value or Linearity Slope P-Value are < .05, the respective bar in Percent of Process Variation will also be red to highlight significance.

6. Statistical significance is the best indicator of a problem with bias or linearity. Some practitioners also use a rule of thumb that a good measurement system will have % Average Bias < 5% and % Linearity < 5%.

7. The dashed lines on the Linearity scatter plot are the regression fit 95% confidence intervals. If the zero reference line is not within the intervals, the linearity slope is significant, indicating that the measurement system linearity is not acceptable.

8. This calculator requires Excel 2010 or higher. Linearity regression uses vba macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

9. Rows 34-39 and 42 are hidden to simplify the output report. These may be unhidden to view how the calculations are performed.

REFERENCE

1. Automotive Industry Action Group AIAG (2010). Measurement Systems Analysis MSA Reference Manual, 4th Edition

Basic MSA Templates – Gage R&R Study (MSA)

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Gage R&R Study (MSA)** to access the Gage R&R Study template.

The following example is given in **SigmaXL > Help > Template Examples > MSA > Gage R&R**. If prompted, please ensure that macros are enabled.

Gage Name:	Example Gage R&R
Date of Study:	
Performed By:	
Notes:	

Process Tolerance	
USL:	2
LSL:	1

StDev Multiplier:	6.00

Operator A	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8	Part 9	Part 10
Reading 1	1.34	1.31	1.52	1.44	1.44	1.29	1.52	1.49	1.44	1.52
Reading 2	1.31	1.36	1.52	1.42	1.49	1.23	1.52	1.49	1.47	1.52
Reading 3										

Operator B	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8	Part 9	Part 10
Reading 1	1.29	1.29	1.55	1.42	1.42	1.21	1.52	1.49	1.39	1.52
Reading 2	1.29	1.26	1.49	1.39	1.39	1.21	1.55	1.47	1.36	1.49
Reading 3										

Operator C	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8	Part 9	Part 10
Reading 1	1.26	1.44	1.55	1.42	1.42	1.23	1.52	1.49	1.42	1.55
Reading 2	1.29	1.42	1.52	1.42	1.42	1.26	1.55	1.49	1.42	1.55
Reading 3										

Gage R&R Metrics	Variance Component	% Contribution of Variance Component	StDev	StDev * Multiplier	% Total Variation (TV)	% Tolerance
Gage R&R:	0.001289167	11.44	0.035904967	0.215429803	33.83	21.54
Operator (AV Appraiser Variation):	0.000258426	2.29	0.016075631	0.096453789	15.15	9.65
Operator * Part (INT Interaction):	0.000627407	5.57	0.025048102	0.150288611	23.60	15.03
Reproducibility (SQRT(AV^2 + INT^2)):	0.000885833	7.86	0.029762952	0.178577714	28.04	17.86
Repeatability (EV Equipment Variation):	0.000403333	3.58	0.02008316	0.120498963	18.92	12.05
Part Variation (PV):	0.009977222	88.56	0.099886046	0.599316277	94.10	59.93
Total Variation (TV):	0.011266389	100.00	0.106143247	0.636859482	100.00	63.69

Create Stacked Column Format for "Analyze Gage R&R" >>

Click **Create Stacked Column Format for "Analyze Gage R&R" >>** if you wish to analyze the above data using SigmaXL's menu Gage R&R Analysis

Notes for use of the Gage R&R Template:

1. The Automotive Industry Action Group (AIAG) recommended study includes 10 Parts, 3 Operators and 3 Replicates. The template calculations will work with a minimum of 2 Operators, 2 Parts and 2 Replicates. The data should be balanced with each operator measuring the same number of parts and the same number of replicates. Use **SigmaXL > Measurement Systems Analysis** to specify up to 30 Parts, 10 Operators and 10 Replicates.

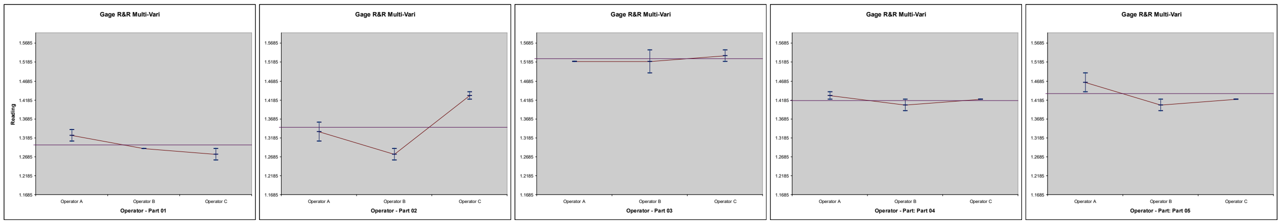
2. Enter process Upper Specification Limit (USL) and Lower Specification Limit (LSL) in the Process Tolerance window. This is used to determine the % Tolerance metrics. If the specification is single-sided, leave both entries blank.

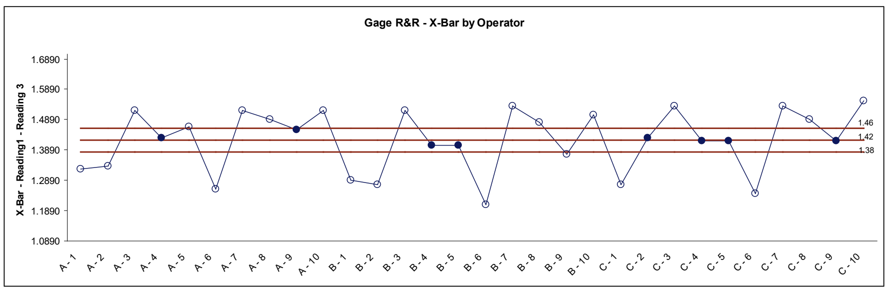
3. The default StDev multiplier is 6. Change this to 5.15 if AIAG convention is being used.

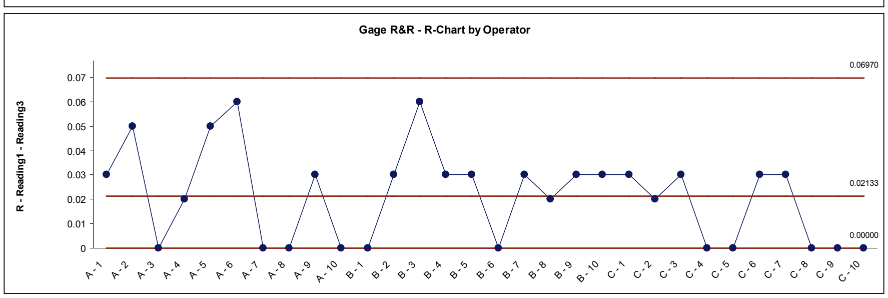
4. The cells shaded in light blue highlight the critical metrics Gage R&R % Total Variation (also known as %R&R) and %Tolerance: < 10% indicates a good measurement system; > 30% indicates an unacceptable measurement system.

Basic MSA Templates - Gage R&R: Multi-Vari & X-bar R Charts

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Gage R&R: Multi-Vari & X-bar R Charts** to access the Multi-Vari & X-bar R Charts template. An example is given in **SigmaXL > Help > Template Examples > MSA > Gage R&R - Multi-Vari** and **Gage R&R – X-Bar R:**







Notes for use of the Gage R&R: Multi-Vari & X-bar R Charts:

1. The Gage R&R Multi-Vari and X-bar & R charts can only be generated if a Gage R&R template has been completed and is the active worksheet. Select the Gage R&R worksheet and click **SigmaXL > Templates & Calculators > Gage R&R: Multi-Vari & X-Bar R Charts** to create the above charts.

2. The Multi-Vari chart shows each Part as a separate graph. Each Operator's response readings are denoted as a vertical line with the top tick corresponding to the Maximum value, bottom tick is the Minimum, and the middle tick is the Mean. The horizontal line across each graph is the overall average for each part.

3. When interpreting the X-bar and R chart for a Gage R&R study, it is desirable that the X-bar chart be out-of-control, and the Range chart be in-control. The control limits are derived from within Operator repeatability.

Basic MSA Templates – Attribute Gage R&R (MSA)

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Attribute MSA** to access the Attribute MSA Study template. An example is given in **SigmaXL > Help > Template Examples > Attribute MSA.**

MEASUREMENT SYSTEM STUDY FOR ATTRIBUTE DATA (Recommend 3 Appraisers, 2 Repeats, Minimum of 10 Good Parts and 10 Bad Parts)

Product/Unit Name:	Example Attribute MSA
Date of Study:	
Performed By:	
Notes:	

Good Part or Unit:	G
Bad Part or Unit:	NG

Part	True Standard	Appraiser A Trial # 1	Appraiser A Trial #2	Appraiser B Trial # 1	Appraiser B Trial #2	Appraiser C Trial # 1	Appraiser C Trial #2
1	G	G	G	NG	G	G	G
2	G	NG	G	G	G	G	G
3	NG	NG	NG	NG	NG	NG	NG
4	G	G	G	G	NG	G	G
5	NG	NG	NG	NG	NG	NG	NG
6	NG	NG	NG	NG	NG	NG	NG
7	G	G	G	NG	G	G	G
8	NG	NG	NG	NG	NG	NG	NG
9	G	G	G	G	G	G	G
10	NG	NG	NG	NG	NG	NG	NG
11	G	G	G	NG	G	G	G
12	NG	NG	NG	NG	NG	NG	NG
13	NG	NG	NG	NG	NG	NG	NG
14	G	NG	G	NG	G	G	G
15	NG	NG	NG	NG	G	NG	NG
16	G	G	NG	G	G	G	NG
17	G	G	G	G	G	G	G
18	NG	NG	NG	NG	G	NG	NG
19	NG	NG	NG	NG	NG	NG	NG
20	G	G	G	NG	G	G	G

Create Stacked Column format to Analyze with "Attribute MSA (Binary)" >>

Attribute MSA Analysis:

Within Appraiser Agreement:	# Inspected	# Matched	Percent	95% LC	95% UC	Cohen's Kappa
A	20	17	85.00%	62.11%	96.79%	0.6939
B	20	12	60.00%	36.05%	80.88%	0.2381
C	20	19	95.00%	75.13%	99.87%	0.9000

Type I = False Reject (Appraiser Rejected Good Part)
Type II = False Accept (Appraiser Accepted Bad Part)
Mixed = Assessments across trials are not identical

Each Appraiser vs Standard Agreement:	# Inspected	# Matched	Percent	95% LC	95% UC	Type I Errors	Type I Error %	Type II Errors	Type II Error %	Mixed Error	Mixed Error %	Cohen's Kappa
A	20	17	85.00%	62.11%	96.79%	0	0.00%	0	0.00%	3	15.00%	0.8500
B	20	12	60.00%	36.05%	80.88%	0	0.00%	0	0.00%	8	40.00%	0.6000
C	20	19	95.00%	75.13%	99.87%	0	0.00%	0	0.00%	1	5.00%	0.9500

Between Appraiser Agreement:	# Inspected	# Matched	Percent	95% LC	95% UC
	20	10	50.00%	27.20%	72.80%

All Appraisers vs Standard Agreement:	# Inspected	# Matched	Percent	95% LC	95% UC	Cohen's Kappa
	20	10	50.00%	27.20%	72.80%	0.8000

Click **Create Stacked Column Format to Analyze with "Attribute MSA (Binary)" >>** if you wish to analyze the above data using SigmaXL's menu Attribute MSA Analysis tool.

Notes for use of the Attribute Gage R&R (MSA) Template:

1. Attribute Gage R&R is also known as Attribute Agreement Analysis.

2. Recommend for study: 3 Appraisers, 2 to 3 Replicates, Minimum of 10 Good Parts and 10 Bad Parts. The data should be balanced with each appraiser evaluating the same number of parts and the same number of replicates.

3. Specify the Good Part or Unit as G or other appropriate text (P, Y, etc.). Specify the Bad Part or Unit as NG or other appropriate text (F, N, etc.). Be careful to avoid typing or spelling errors when entering the results. A space accidentally inserted after a character will be treated as a different value leading to incorrect results.

Basic MSA Templates – Orthogonal (Deming) Regression

Click **SigmaXL > Templates & Calculators > Basic MSA Templates > Orthogonal (Deming) Regression** to access the Orthogonal (Deming) Regression template.

The following example is given in **SigmaXL > Help > Template Examples > MSA > Orthogonal (Deming) Regression**.

X	Y
3.96	2.55
5.45	6.88
9.72	10.09
1.97	2.46
3.73	4.85
7.78	9.54
7.3	6.5
2.55	3.49
1.41	2.22
7.23	6.71
1.35	2.84
8.33	8.46
5.23	5.46
3.63	3.71
9	9.16
3.43	3.23
5.13	5.77
9.18	8.54
6.54	6.34
6.14	6.1
7.38	7.97
3.25	2.28
8.71	8.66
8.79	8.05
1.63	1.09
2.62	0.86
5.67	4.19
7.11	6.97
8.25	6.49
2.25	1.69

Measurement Variance Ratio (Y/X):	1
Confidence Level:	95
H0: Slope =	1

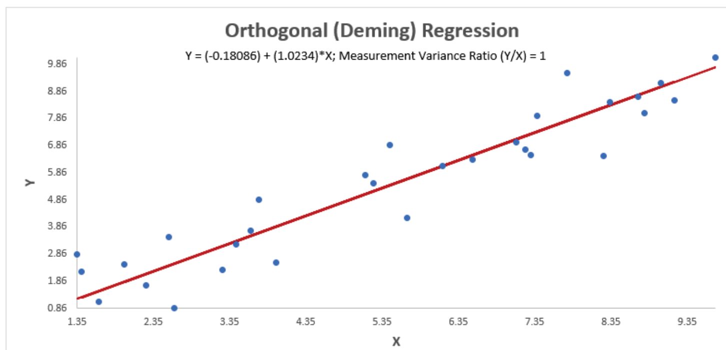
Orthogonal Regression Report

Orthogonal (Deming) Regression Model: Y = (-0.18086) + (1.0234)*X

Model Summary	
Pearson Correlation Squared	88.48%
Generalized R-Square	93.93%

Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	Z	P	Lower 95% CI	Upper 95% CI
Constant	-0.180857154	0.421385169	-0.4292	0.6678	-1.006756909	0.645042602
X	1.023407689	0.069824406	0.3352	0.7374	0.886554368	1.160261011

Error Variances	
Variable	Variance
Y	0.432683719
X	0.432683719



X	Y	Predicted Y	Residuals	Standardized Residuals	X Fit	X Residuals	Y Fit	Y Residuals	Optimized Residuals
3.96	2.55	3.871837297	-1.321837297	-1.404414381	3.299258228	0.660741772	3.195629087	-0.645629087	-0.923805502
5.45	6.88	5.396714754	1.483285246	1.575948217	6.191444143	-0.741444143	6.155514391	0.724485609	1.036638227
9.72	10.09	9.766665588	0.323334412	0.343533579	9.88162394	-0.16162394	9.932072771	0.157927229	0.225971918
1.97	2.46	1.835255995	0.624744005	0.663772666	2.282288405	-0.312288405	2.15485435	0.30514565	0.436621021

Notes:

1. Orthogonal (Deming) Regression differs from the simple linear regression in that it accounts for errors in observations on both the x- and the y- axis. Use it to compare the same parts on two measurement systems.
2. Enter your Measurement data in the **X** and **Y** columns. The **X** and **Y** column headings may be modified. This template has a maximum sample size of 1000. Rows with missing data will be deleted.
3. Enter **Measurement Variance Ratio (Y/X)**. A value of 1 assumes equal variance. This ratio is assumed to be constant.
4. The variances for this ratio may be estimated from separate Gage R&R studies or from data replicates. Note that you cannot simply use the variances from the data.
5. Enter **Confidence Level** and **H0: Slope =** . Typically H0:Slope = 1, so you are expecting a p-value > alpha, i.e., fail to reject H0.
6. Click the **Orthogonal Regression Report** button to produce the regression report and scatterplot with orthogonal best fit line.
7. Residuals are given in the report and include Predicted (Raw Y) Residuals, Standardized (Raw Y) Residuals, X Fit Residuals, Y Fit Residuals and Deming Optimized Residuals. See [2] for formulas.
8. Model Summary includes Pearson Correlation Squared and Generalized R-Square [1].
9. See SigmaXL Workbook Appendix **Orthogonal (Deming) Regression** for formula details.
10. SigmaXL must be initialized and appear on the menu in order for this template to function.

REFERENCES

[1] Bossé, M., Marland, E., Rhoads, G., Sanqui, J.A. & BeMent, Z. (2022), "Generalizing R^2 for deming regressions", *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2022.2059678

[2] NCSS Procedures, "Deming Regression", https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Deming_Regression.pdf. Note that the error variance ratio is defined as Variance(X)/Variance(Y).

[3] Wikipedia "Deming Regression", https://en.wikipedia.org/wiki/Deming_regression.

ACKNOWLEDGEMENT

Special thanks to the authors of "Generalizing R^2 for deming regressions", Michael Bossé, Eric Marland, Gregory Rhoads, Jose Almer Sanqui & Zack BeMent, for providing R code used to validate Generalized R-Square.

Basic Process Capability Templates – Process Sigma Level – Discrete Data

Click **SigmaXL > Templates & Calculators > Basic Process Capability Templates > Process Sigma Level – Discrete** to access the Process Sigma Level – Discrete calculator. The template gives the following default example.

Process Sigma Level Calculator - Discrete Data		
Sample Data (user inputs):		
Number of units	n	500
Total number of defects observed	d	5
Number of defect opportunities per unit	o	1
Sigma Shift (typically +1.5)		1.5
Results:		
Defects per Unit	dpu	0.01
Defects per Million Opportunities	dpmo	10,000.0
Defects per Opportunity	dpo%	1.00%
Yield	yield%	99.00%
Process Sigma Level	sigma	3.826

Notes for use of the Process Sigma Calculator for Discrete Data:

1. Total number of defects should include defects made and later fixed.
2. Sample size should be large enough to observe 5 defects.

Basic Process Capability Templates – Process Sigma Level – Continuous Data

Click **SigmaXL > Templates & Calculators > Basic Process Capability Templates > Process Sigma Level – Continuous** to access the Process Sigma Level – Continuous calculator. The template gives the following default example.

Process Sigma Level Calculator - Continuous Data (Assumes that data are normally distributed)		
Sample Data (user inputs):		
Mean	x-bar	0
Standard Deviation	s	1
Upper Specification Limit	USL	3
Lower Specification Limit	LSL	-3
Sigma Shift (typically +1.5)		1.5
Results:		
Expected ppm > USL		1349.9
Expected % > USL		0.13%
Expected ppm < LSL		1349.9
Expected % < LSL		0.13%
Expected ppm (overall)		2699.8
Expected yield (overall) %		99.73%
Process Sigma Level		4.282

Note: This calculator assumes that the Mean and Standard Deviation are computed from data that are normally distributed.

Basic Process Capability Templates – Process Capability Indices

Click **SigmaXL > Templates & Calculators > Basic Process Capability Templates > Process Capability** to access the Process Capability Indices calculator. The template gives the following default example.

Process Capability Indices - Continuous Data (Assumes that data are normally distributed)		
Sample Data (user inputs):		
Mean	x-bar	0
Standard Deviation	s	1
Upper specification limit	USL	3
Lower specification limit	LSL	-3
Results:		
	Cp, Pp	1.00
	Cpu, Ppu	1.00
	Cpl, Ppl	1.00
	Cpk, Ppk	1.00

Notes for use of the Process Capability Indices Calculator:

1. This calculator assumes that the Mean and Standard Deviation are computed from data that are normally distributed.
2. Reports Cp, Cpk if entered S is Within or Short Term (using a control chart).
3. Reports Pp, Ppk if entered S is Overall or Long Term.

Basic Process Capability Templates – Process Capability & Confidence Intervals

Click **SigmaXL > Templates & Calculators > Basic Process Capability Templates > Process Capability & Confidence Intervals** to access the Process Capability Indices & Confidence Intervals calculator. The template gives the following default example.

Process Capability Indices and Confidence Intervals - Continuous Data (Assumes that data are normally distributed)					
Sample Data (user inputs):					
Mean	x-bar	1	Upper Specification Limit	USL	3
Standard Deviation	s	1	Lower Specification Limit	LSL	-3
Sample Size	n	30	Confidence Level (enter .95 for 95%)		95.0%
Sigma Shift (typically +1.5)		1.5			
Process Sigma Level Results:			Process Capability Results:		
Expected ppm > USL		22,750.1	Cp, Pp		1.00
Expected % > USL		2.28%	Lower Limit Cp, Pp		0.74
Expected ppm < LSL		31.7	Upper Limit Cp, Pp		1.26
Expected % < LSL		0.00%	Cpu, Ppu		0.67
Expected ppm (overall)		22,781.8	Cpl, Ppl		1.33
Expected yield (overall)		97.72%	Cpk, Ppk		0.67
			Lower Limit Cpk, Ppk		0.46
Process Sigma Level		3.499	Upper Limit Cpk, Ppk		0.88

Note: Cp, Cpk if S is Within or Short Term; Pp, Ppk if S is Overall or Long Term

Notes for use of the Process Capability & Confidence Intervals Calculator:

1. This calculator assumes that the Mean and Standard Deviation are computed from data that are normally distributed.
2. Reports Cp, Cpk if entered S is Within or Short Term (using a control chart).
3. Reports Pp, Ppk if entered S is Overall or Long Term.

Basic Process Capability Templates – Tolerance Interval Calculator (Normal Exact)

Click **SigmaXL > Templates & Calculators > Basic Process Capability Templates > Tolerance Interval Calculator (Normal Exact)** to access the Tolerance Interval Calculator (Normal Exact). The template gives the following default example.

Tolerance Interval Calculator (Normal Exact)		
Sample Data (user inputs):		
Sample Size	n	30
Sample Mean	x-bar	1
Sample Standard Deviation	s	1
Population Coverage (enter .99 for 99%)	100*(p)%	99.0%
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%
Tolerance Interval Type		Two-Sided

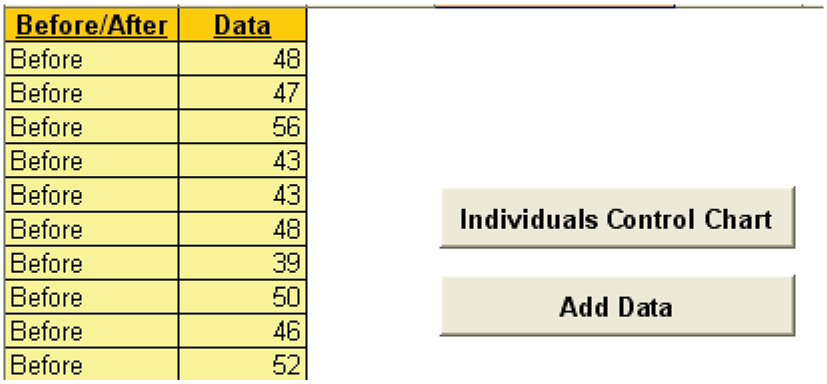
Results:	
Upper Tolerance Limit	4.3546
Lower Tolerance Limit	-2.3546

Notes:

1. Enter summarized Sample Data, Population Coverage and Confidence Level in cells with yellow highlight.
2. Select Tolerance Interval Type using drop-down.
3. This calculator assumes that the summary statistics come from data that is normally distributed. Tolerance intervals are exact, one-sided uses the noncentral t distribution, two-sided uses the noncentral chi-square distribution. See Appendix Tolerance Interval Calculator (Normal Exact) for statistical details.
4. Due to the complexity of calculations, this template uses vba macros rather than Excel formulas. SigmaXL must be initialized and appear on the menu in order for this template to function.

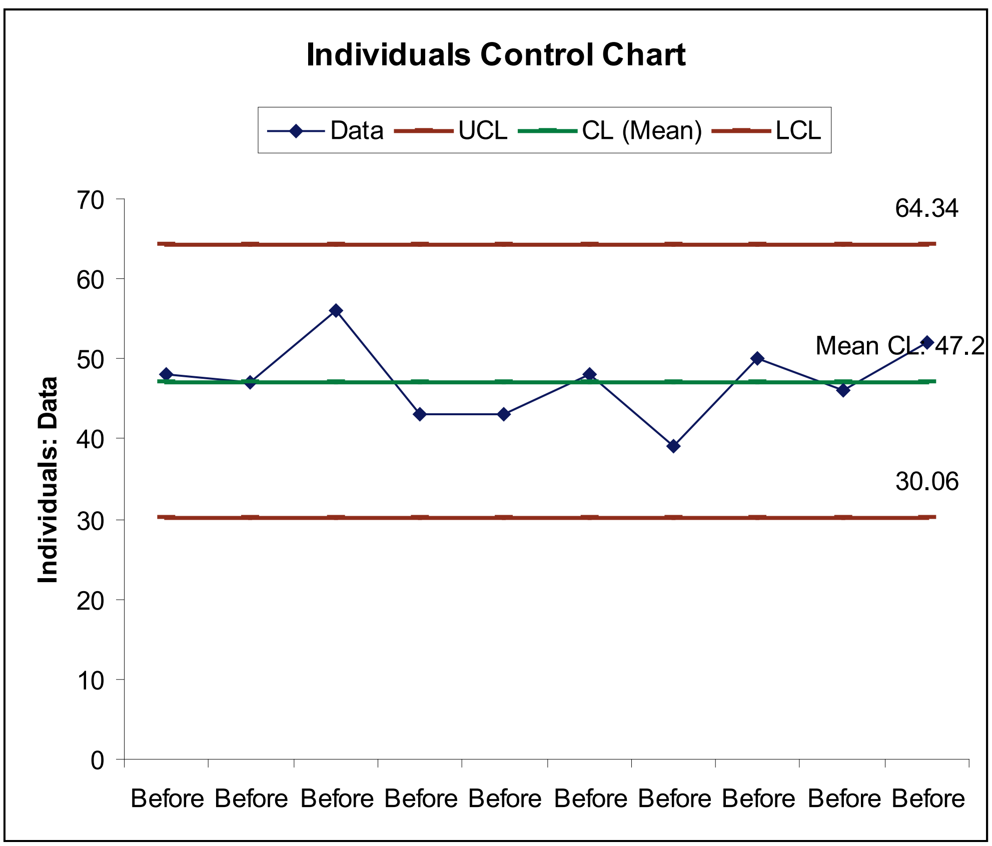
Basic Control Chart Templates – Individuals Chart

Click **SigmaXL > Templates & Calculators > Control Chart Templates > Basic > Individuals**. Enter the data as shown (or copy and paste the “Before” data from the histogram data above):



Before/After	Data
Before	48
Before	47
Before	56
Before	43
Before	43
Before	48
Before	39
Before	50
Before	46
Before	52

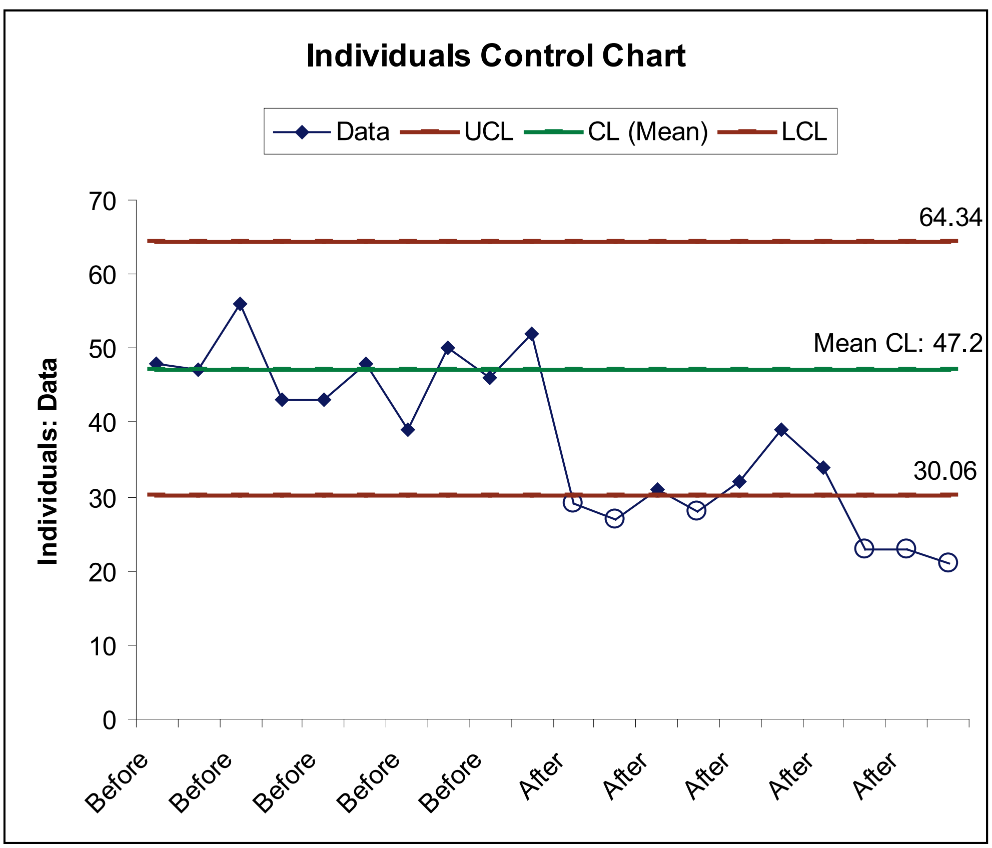
Click the Individuals Control Chart button to produce the Individuals Chart (note that in actual application, one should use a minimum of 20 to 30 data points to compute the initial control limits).



Now enter the data as shown (or copy and paste the "After" data from the histogram data used earlier):

Before/After	Data
Before	48
Before	47
Before	56
Before	43
Before	43
Before	48
Before	39
Before	50
Before	46
Before	52
After	29
After	27
After	31
After	28
After	32
After	39
After	34
After	23
After	23
After	21

Click the Add Data button to add the "After" data to the Individuals chart.



Notes:

1. This Individuals Control Chart template should be used with continuous data like cycle time. The data must be in chronological time-sequence order.
2. You can replace the **X-Axis Label** and **Data** column headings with any headings that you wish.
3. Enter your data in the **Data** column.
4. Enter labels in **X-Axis Label** column. Labels can be Date, Time, Name, or other text information. These labels are optional and will appear on the horizontal X-Axis of the Individuals Control Chart.
5. Click the **Individuals Control Chart** button to create an Individuals Control Chart.
6. After the control chart has been created and additional new data entered into the **Data** column, click the **Add Data** button to add the data to the existing chart.

Part C – Descriptive/Summary Statistics

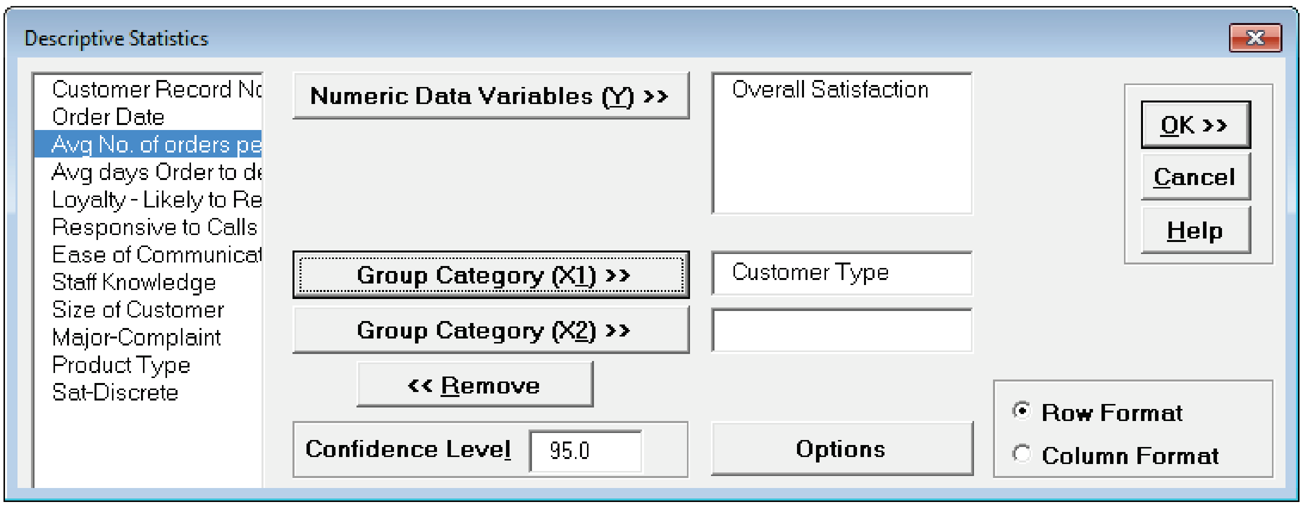
Descriptive Statistics

1. Open **Customer Data.xlsx**. (To access, click **SigmaXL** > **Help** > **Sample Data** or **Start > Programs > SigmaXL > Sample Data**). Click **Sheet 1** Tab.

2. Click **SigmaXL > Statistical Tools > Descriptive Statistics**.

3. Check **Use Entire Data Table**, click **Next**.

4. Select *Overall Satisfaction*, click **Numeric Data Variables (Y) >>**. Select *Customer Type*, click **Group Category (X1) >>** as shown:



Tip: Be careful to not select a continuous variable for **Group Category**, as each unique value will be considered as a category level.

5. Click **OK**.

6. Descriptive Statistics are given for Customer Satisfaction grouped by Customer Type:

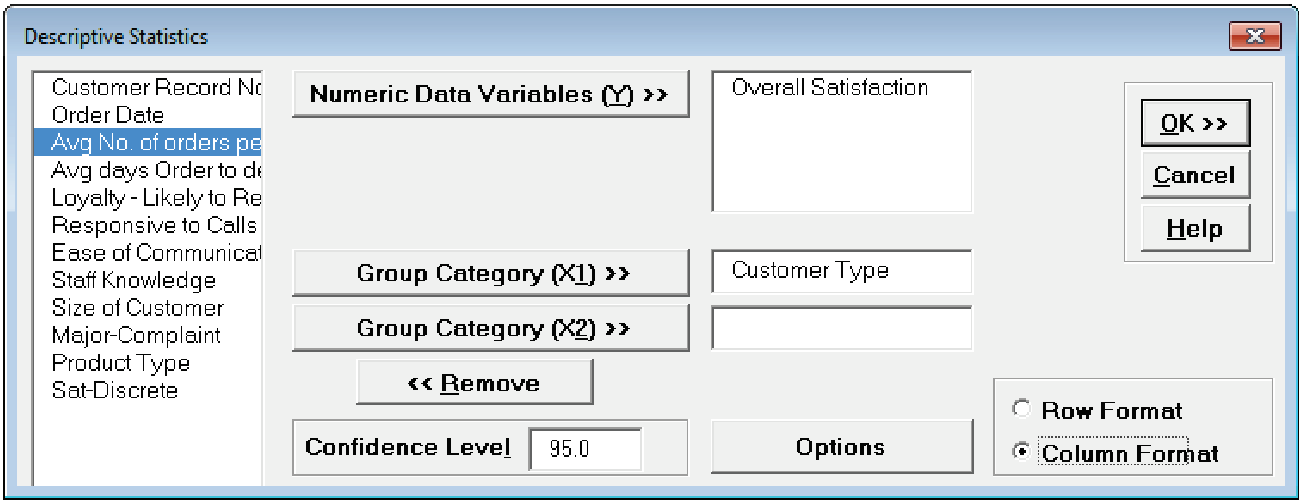
Overall Satisfaction by Customer Type			
Descriptive Statistics	**Customer Type = 1**	**Customer Type = 2**	**Customer Type = 3**
Count	31	42	27
Mean	3.394	4.205	3.641
Stdev	0.824680	0.621200	0.670478
Range	3.080	2.560	2.740
Minimum	1.720	2.420	2.190
25th Percentile (Q1)	2.810	3.828	3.240
50th Percentile (Median)	3.560	4.340	3.510
75th Percentile (Q3)	4.020	4.725	4.170
Maximum	4.800	4.980	4.930
95.0% CI Mean	3.091 to 3.696	4.012 to 4.399	3.376 to 3.906
95.0% CI Sigma	0.659012 to 1.102328	0.511126 to 0.792132	0.528013 to 0.918845
Anderson-Darling Normality Test	0.312776	0.826259	0.389291
p-value (A-D Test)	0.5306	0.0302	0.3600
Skewness	-0.235169	-0.967994	0.139571
p-value (Skewness)	0.5557	0.0121	0.7411
Kurtosis	-0.67169	0.679609	-0.313701
p-value (Kurtosis)	0.3705	0.2865	0.8435

Which Customer Type has the highest mean satisfaction score? Clearly Type 2. However, we have to be careful concluding that there is a significant difference in satisfaction solely by looking at the

Means. In the Analyze Phase, we will run tests of hypothesis to validate that Type 2 Customers are, in fact, significantly more satisfied.

Tip: Click on **Column B**, click **View | Window | Split, Freeze Panes**. This freezes **Column A** and allows you to scroll across the Descriptive Statistics for each level of the Group Category. This is particularly beneficial when there are a large number of columns.

7. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Change the format selected to **Column Format** as shown:



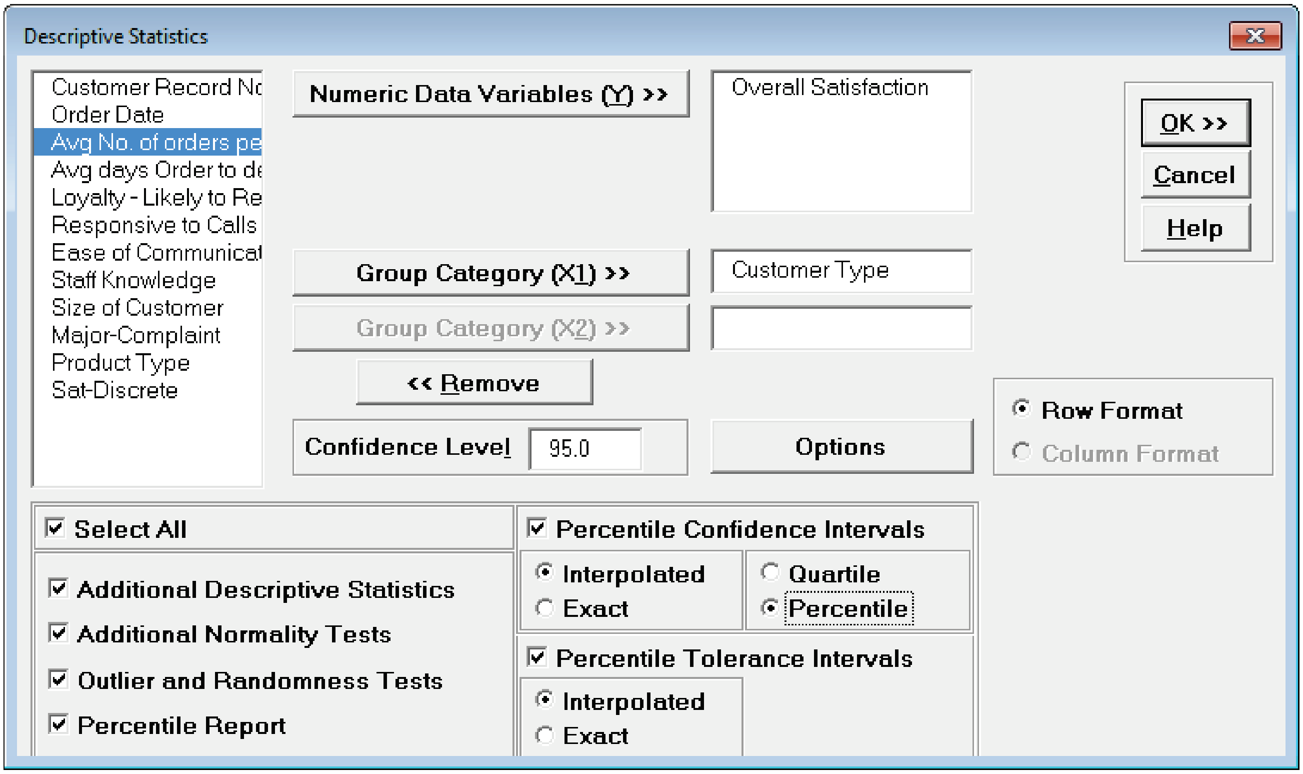
8. Click **OK**. Descriptive Statistics are given for Customer Satisfaction broken out by Customer Type in Column Format:

Overall Satisfaction by Customer Type																		
Descriptive Statistic	Count	Mean	Stdev	Range	Minimum	25th Percentile (Q1)	50th Percentile (Median)	75th Percentile (Q3)	Maximum	95.0% CI Mean	95.0% CI Sigma	Anderson-Darling Normality Test	p-value (A-D Test)	Skewness	p-value (Skewness)	Kurtosis	p-value (Kurtosis)
Customer Type = 1	31	3.394	0.824680	3.080	1.720	2.810	3.560	4.020	4.800	3.091 to 3.696	0.659012 to 1.102328	0.312776	0.5306	-0.235169	0.5557	-0.67169	0.3705
Customer Type = 2	42	4.205	0.621200	2.560	2.420	3.828	4.340	4.725	4.980	4.012 to 4.399	0.511126 to 0.792132	0.826259	0.0302	-0.967994	0.0121	0.679609	0.2865
Customer Type = 3	27	3.641	0.670478	2.740	2.190	3.240	3.510	4.170	4.930	3.376 to 3.906	0.528013 to 0.918845	0.389291	0.3600	0.139571	0.7411	-0.313701	0.8435

Tip: This column format is useful to create subsequent graphs on the summary statistics.

Descriptive Statistics - Options

1. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Click **Options.** Check **Select All** and change Percentile Confidence Intervals to **Percentile** to display all Percentile values in the report.



Tip: Select only those options that are of interest in order to minimize the size of the report. Here we are selecting all options for demonstration purposes. Note that when any option is checked, **Row Format** is automatically selected, **Column Format** and **Group Category (X2)** are greyed out. These display options are limited due to the amount of information displayed in the extended report.

2. Click **OK**. Extended Descriptive Statistics are given for Customer Satisfaction grouped by Customer Type:

Overall Satisfaction by Customer Type			
Descriptive Statistics	Customer Type = 1	Customer Type = 2	Customer Type = 3
Count	31	42	27
Mean	3.394	4.205	3.641
Stdev	0.824680	0.621200	0.670478
Range	3.080	2.560	2.740
Minimum	1.720	2.420	2.190
25th Percentile (Q1)	2.810	3.828	3.240
50th Percentile (Median)	3.560	4.340	3.510
75th Percentile (Q3)	4.020	4.725	4.170
Maximum	4.800	4.980	4.930
95.0% CI Mean	3.0911 to 3.696	4.0117 to 4.3988	3.3759 to 3.9063
95.0% CI Sigma	0.65901 to 1.1023	0.51113 to 0.79213	0.52801 to 0.91884
Anderson-Darling Normality Test	0.312776	0.826259	0.389291
P-Value (A-D Test)	0.5306	0.0302	0.3600
Skewness	-0.235169	-0.967994	0.139571
P-Value (Skewness)	0.5557	0.0121	0.7411
Kurtosis	-0.671690	0.679609	-0.313701
P-Value (Kurtosis)	0.3705	0.2865	0.8435
Additional Descriptive Statistics			
5% Trimmed Mean	3.413	4.251	3.648
Standard Error of Mean	0.148117	0.095853182	0.129034
Variance	0.680097	0.385889	0.449541
Coefficient of Variation	24.301	14.772	18.414
StDev (Within, Short Term)	0.678487	0.521969	0.683988
Additional Normality Tests			
Shapiro-Wilk/KSL	0.969383	0.923744	0.971664
P-Value (Shapiro-Wilk/KSL)	0.5022	0.0080	0.6463
Doornik-Hansen	0.813602	8.554	0.127399
P-Value (Doornik-Hansen)	0.6658	0.0139	0.9383

3. The **Additional Descriptive Statistics** are:
   - 5% Trimmed Mean. The highest 5% and lowest 5% are excluded and mean calculated with the rest of the data. This gives a robust alternative to the Median as a measure of central tendency in the presence of outliers.
   - Standard Error of Mean (StDev/$\sqrt{N}$)
   - Variance (StDev^2)
   - Coefficient of Variation (100 * StDev/Mean)
   - Short Term StDev (MR-bar/d2)

4. The **Additional Normality Tests** are:
   - Shapiro-Wilk (n <= 5000) and Kolmogorov-Smirnov-Lilliefors (KSL, n > 5000)
     - This is a popular alternative to Anderson Darling.
   - Doornik-Hansen (DH)
     - Univariate omnibus test based on Skewness and Kurtosis. (Note, the bivariate DH test is used in Correlation Matrix to test bivariate normality).
     - Best for data with ties, i.e. “chunky” data. Anderson-Darling, Shapiro-Wilk and KSL are severely affected by ties in the data and will trigger a low P-Value even if the data are normal.
     - See Appendix Doornik-Hansen (DH) Normality Test for further details and references.

Overall Satisfaction by Customer Type			
Descriptive Statistics	**Customer Type = 1**	**Customer Type = 2**	**Customer Type = 3**
Percentile Report			
0.135	1.72	2.42	2.19
0.5	1.72	2.42	2.19
1	1.72	2.42	2.19
2.5	1.72	2.441	2.19
5	1.804	2.778	2.37
10	2.138	3.285	2.832
15	2.554	3.518	3.056
20	2.602	3.65	3.144
25	2.81	3.8275	3.24
30	2.924	3.981	3.272
35	2.97	4.091	3.298
40	3.074	4.132	3.33
45	3.248	4.3035	3.422
50	3.56	4.34	3.51
55	3.576	4.4195	3.566
60	3.804	4.464	3.644
65	3.9	4.5275	3.988
70	3.978	4.558	4.074
75	4.02	4.725	4.17
80	4.11	4.786	4.294
85	4.242	4.881	4.428
90	4.586	4.91	4.688
95	4.722	4.9655	4.878
97.5	4.8	4.97925	4.93
99	4.8	4.98	4.93
99.5	4.8	4.98	4.93
99.865	4.8	4.98	4.93
75 - 25 (50%, Interquartile Range IQR)	1.21	0.8975	0.93
90 - 10 (80%, Interdecile Range IDR)	2.448	1.625	1.856
95 - 5 (90%, Span)	2.918	2.1875	2.508
97.5 - 2.5 (95%, +/- 1.96 Sigma Equivalent)	3.08	2.53825	2.74
99 - 1 (98%)	3.08	2.56	2.74
99.5 - 0.5 (99%)	3.08	2.56	2.74
99.865 - 0.135 (99.73%, +/- 3 Sigma Equivalent)	3.08	2.56	2.74

5. The **Percentile Report** gives 27 values from 0.135 to 99.865.

6. The **Percentile Ranges** are:
    - 75 - 25 (50%, Interquartile Range IQR)
    - 90 - 10 (80%, Interdecile Range IDR)
    - 95 - 5 (90%, Span)
    - 97.5 - 2.5 (95%, +/- 1.96 Sigma Equivalent)
    - 99 - 1 (98%)
    - 99.5 - 0.5 (99%)
    - 99.865 - 0.135 (99.73%, +/- 3 Sigma Equivalent)

Overall Satisfaction by Customer Type			
Descriptive Statistics	Customer Type = 1	Customer Type = 2	Customer Type = 3
Percentile Confidence Intervals (Interpolated 95.0%)			
0.135	Min. sample size = 2731	Min. sample size = 2731	Min. sample size = 2731
0.5	Min. sample size = 736	Min. sample size = 736	Min. sample size = 736
1	Min. sample size = 368	Min. sample size = 368	Min. sample size = 368
2.5	Min. sample size = 146	Min. sample size = 146	Min. sample size = 146
5	Min. sample size = 72	Min. sample size = 72	Min. sample size = 72
10	Min. sample size = 36	2.5826 to 3.6589	Min. sample size = 36
15	1.8393 to 2.9095	3.1206 to 3.8676	2.4389 to 3.2707
20	2.0087 to 2.9711	3.2747 to 4.0232	2.7108 to 3.3014
25	2.4809 to 3.1075	3.4847 to 4.1163	2.9143 to 3.3899
30	2.5606 to 3.2586	3.5891 to 4.3005	3.0606 to 3.4556
35	2.6047 to 3.5658	3.7308 to 4.3359	3.1354 to 3.5275
40	2.7749 to 3.6297	3.8487 to 4.4153	3.2052 to 3.6277
45	2.906 to 3.817	3.9942 to 4.4601	3.2554 to 3.862
50	2.9535 to 3.9362	4.0946 to 4.5184	3.2891 to 4.0227
55	3.0221 to 3.987	4.1397 to 4.5458	3.3036 to 4.1239
60	3.2061 to 4.0266	4.3049 to 4.6959	3.3845 to 4.1874
65	3.387 to 4.1066	4.3441 to 4.7459	3.4401 to 4.3391
70	3.5707 to 4.2345	4.4284 to 4.8516	3.5236 to 4.4274
75	3.7743 to 4.2921	4.4737 to 4.8928	3.626 to 4.6107
80	3.8978 to 4.67	4.53 to 4.91	3.9334 to 4.7587
85	3.9853 to 4.6892	4.6717 to 4.9457	4.0779 to 4.8581
90	Min. sample size = 36	4.77 to 4.9742	Min. sample size = 36
95	Min. sample size = 72	Min. sample size = 72	Min. sample size = 72
97.5	Min. sample size = 146	Min. sample size = 146	Min. sample size = 146
99	Min. sample size = 368	Min. sample size = 368	Min. sample size = 368
99.5	Min. sample size = 736	Min. sample size = 736	Min. sample size = 736
99.865	Min. sample size = 2731	Min. sample size = 2731	Min. sample size = 2731
Percentile Tolerance Intervals (Interpolated 95.0%)			
50%	2.56 to 4.1129	3.62 to 4.8262	3.04 to 4.2764
80%	1.7453 to 4.67	3.108 to 4.97	2.4552 to 4.93
90%	Min. sample size = 46	Min. sample size = 46	Min. sample size = 46
95%	Min. sample size = 93	Min. sample size = 93	Min. sample size = 93
98%	Min. sample size = 236	Min. sample size = 236	Min. sample size = 236
99%	Min. sample size = 473	Min. sample size = 473	Min. sample size = 473
99.73%	Min. sample size = 1756	Min. sample size = 1756	Min. sample size = 1756

7. The **Percentile Confidence Intervals** give 27 values from 0.135 to 99.865.

8. The **Quartile Confidence Intervals** give 3 values: 25, 50 and 75.

9. The **Percentile Tolerance Intervals** are 50%, 80%, 90%, 95%, 98%, 99%, and 99.73%.

10. Confidence Intervals and Tolerance Intervals can be exact or interpolated. If exact, the actual exact confidence level will be a value greater than or equal to specified, due to percentile values being discrete in nature. The actual exact level will also be reported in this case. If interpolated, the result will be an interpolated estimate of the specified confidence level (typically 95.0%) and is the recommended setting. See Appendix **Percentile (Nonparametric) Confidence and Tolerance Intervals** for further details.

    Tip: The Tolerance Intervals given here are nonparametric, so the data does not have to be normal. However, if you have normal data, you can use the Tolerance Interval Calculator for Normal data, which will allow smaller sample sizes to be used. See **Basic Process Capability Templates – Tolerance Interval Calculator (Normal Exact).**

11. If the Confidence Interval or Tolerance Interval cannot be computed due to inadequate sample size, a minimum sample size is reported.

Overall Satisfaction by Customer Type			
Descriptive Statistics	Customer Type = 1	Customer Type = 2	Customer Type = 3
Outlier and Randomness Tests			
Outliers (Boxplot Rules)	No outliers found.	Potential (1.5*IQR) outlier lower count = 1.	No outliers found.
Grubbs Outlier Test	Grubbs' Test P-Value = 1.000. Fail to reject null hypothesis: "There are no outliers in the data set." Note that Grubbs' Test assumes normality and tests only if the maximum or minimum is an outlier.	Grubbs' Test P-Value = 0.105. Fail to reject null hypothesis: "There are no outliers in the data set." Note that Grubbs' Test assumes normality and tests only if the maximum or minimum is an outlier.	Grubbs' Test P-Value = 0.654. Fail to reject null hypothesis: "There are no outliers in the data set." Note that Grubbs' Test assumes normality and tests only if the maximum or minimum is an outlier.
Randomness Runs Test	Nonparametric Runs Test (Exact) P-Value = 0.066. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.

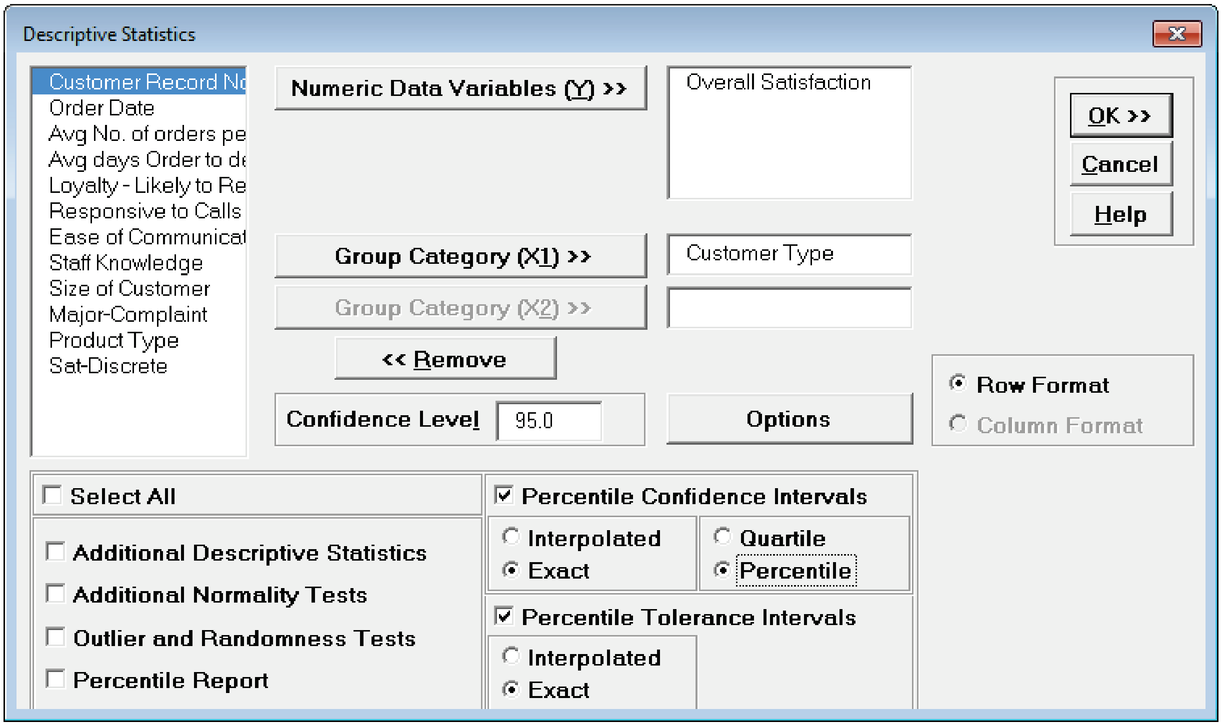
12. The **Outlier (Boxplot Rules)** Tests are: Potential 1.5(IQR), Likely 2.2(IQR), Extreme 3.0(IQR)

13. **Grubbs Outlier Test** is more powerful at detecting a single outlier as maximum or minimum but assumes that the remainder of the data are normally distributed.

14. The **Randomness Runs Test** is a nonparametric exact runs test. If there are values equal to the median, they are set to both "Counted as Below" and "Counted as Above", with the reported P-Value being the larger of the two. This results in a more conservative test to minimize false alarms.

15. The Outlier and Randomness Tests use the same Green, Yellow, Red highlight given in the automatic assumptions report that are included in t-tests and ANOVA.

16. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Click **Options**. Uncheck **Select All** to clear the selections and check **Percentile Confidence Intervals**, select **Exact** and **Percentile**. Check **Percentile Tolerance Intervals**, and select **Exact** as shown:



17. Click **OK**. The Percentile Confidence Intervals and Tolerance Intervals are displayed:

Overall Satisfaction by Customer Type			
Descriptive Statistics	Customer Type = 1	Customer Type = 2	Customer Type = 3
Percentile Confidence Intervals (Exact 95.0% Minimum)			
0.135	Min. sample size = 2731	Min. sample size = 2731	Min. sample size = 2731
0.5	Min. sample size = 736	Min. sample size = 736	Min. sample size = 736
1	Min. sample size = 368	Min. sample size = 368	Min. sample size = 368
2.5	Min. sample size = 146	Min. sample size = 146	Min. sample size = 146
5	Min. sample size = 72	Min. sample size = 72	Min. sample size = 72
10	Min. sample size = 36	2.42 to 3.67 (96.7%)	Min. sample size = 36
15	1.72 to 2.94 (98.1%)	2.7 to 3.9 (97.3%)	2.19 to 3.29 (97.4%)
20	1.86 to 3.01 (97.9%)	3.24 to 4.09 (96.6%)	2.64 to 3.31 (97%)
25	2.04 to 3.24 (98%)	3.39 to 4.12 (96.9%)	2.88 to 3.41 (95.8%)
30	2.56 to 3.26 (95.2%)	3.54 to 4.31 (97.2%)	3.04 to 3.51 (96.6%)
35	2.57 to 3.57 (96.4%)	3.67 to 4.35 (96.6%)	3.12 to 3.53 (95.9%)
40	2.65 to 3.8 (97.4%)	3.83 to 4.43 (96.1%)	3.16 to 3.65 (97.1%)
45	2.9 to 3.82 (95.5%)	3.99 to 4.47 (95.7%)	3.24 to 3.98 (96.8%)
50	2.94 to 3.97 (97.1%)	4.09 to 4.53 (95.6%)	3.26 to 4.11 (98.1%)
55	3.01 to 3.99 (95.5%)	4.12 to 4.55 (95.7%)	3.3 to 4.17 (96.8%)
60	3.09 to 4.05 (97.4%)	4.3 to 4.72 (96.1%)	3.31 to 4.21 (97.1%)
65	3.26 to 4.15 (96.4%)	4.33 to 4.75 (96.6%)	3.43 to 4.42 (95.9%)
70	3.57 to 4.24 (95.2%)	4.4 to 4.87 (97.2%)	3.51 to 4.43 (96.6%)
75	3.58 to 4.67 (98%)	4.47 to 4.91 (96.9%)	3.62 to 4.66 (95.8%)
80	3.82 to 4.67 (97.9%)	4.53 to 4.91 (96.6%)	3.65 to 4.8 (97%)
85	3.97 to 4.8 (98.1%)	4.63 to 4.97 (97.3%)	4.02 to 4.93 (97.4%)
90	Min. sample size = 36	4.75 to 4.98 (96.7%)	Min. sample size = 36
95	Min. sample size = 72	Min. sample size = 72	Min. sample size = 72
97.5	Min. sample size = 146	Min. sample size = 146	Min. sample size = 146
99	Min. sample size = 368	Min. sample size = 368	Min. sample size = 368
99.5	Min. sample size = 736	Min. sample size = 736	Min. sample size = 736
99.865	Min. sample size = 2731	Min. sample size = 2731	Min. sample size = 2731
Percentile Tolerance Intervals (Exact 95.0% Minimum)			
50%	2.56 to 4.15 (96.5%)	3.62 to 4.84 (95.6%)	3.04 to 4.42 (97.4%)
80%	1.72 to 4.67 (96.3%)	2.7 to 4.97 (97.9%)	2.19 to 4.93 (98.1%)
90%	Min. sample size = 46	Min. sample size = 46	Min. sample size = 46
95%	Min. sample size = 93	Min. sample size = 93	Min. sample size = 93
98%	Min. sample size = 236	Min. sample size = 236	Min. sample size = 236
99%	Min. sample size = 473	Min. sample size = 473	Min. sample size = 473
99.73%	Min. sample size = 1756	Min. sample size = 1756	Min. sample size = 1756

18. The specified 95% is a guaranteed minimum. The exact confidence level is given with each reported interval.

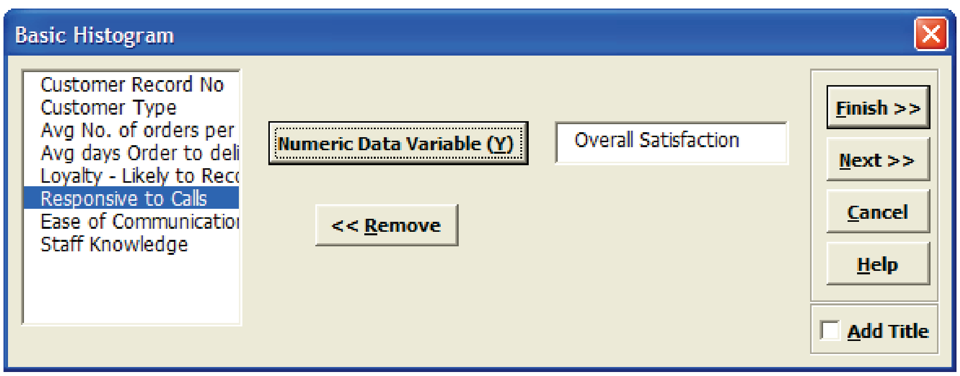
Part D – Histograms

Basic Histogram Template

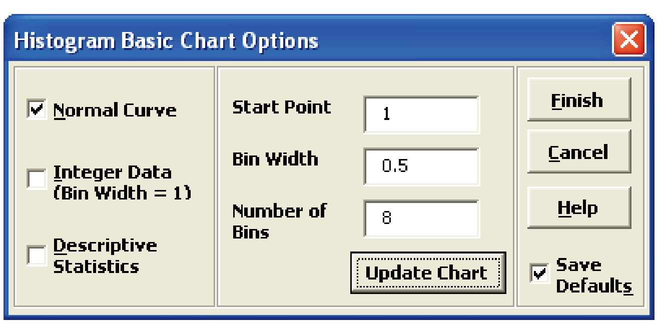
Click **SigmaXL > Templates and Calculators** > **Basic Graphical Templates > Histogram** or **SigmaXL > Graphical Tools > Basic Graphical Templates > Histogram**. See **Part B – Templates and Calculators** for a Histogram Template example.

Single (Basic) Histogram

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Graphical Tools > Basic Histogram**.
3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
4. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>** as shown:

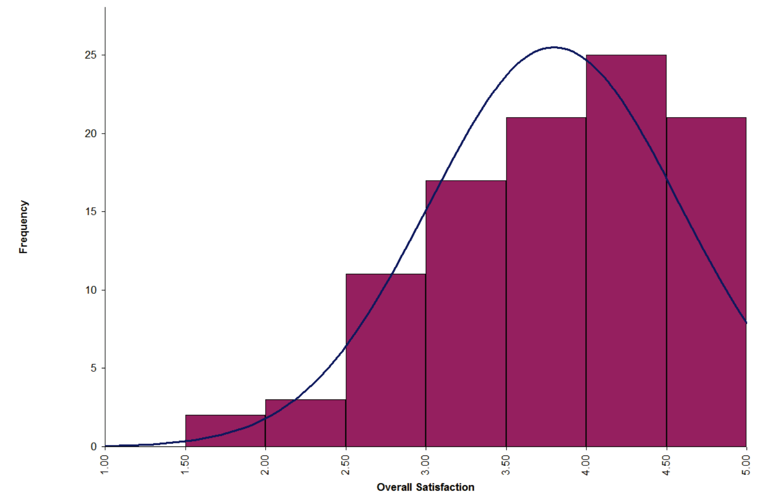


5. Click **Next**. Ensure that **Normal Curve** is checked. Set **Start Point** = 1. Change the **Bin Width** to 0.5, and the **Number of Bins** to 8. Click **Update Chart** to view the histogram. (If the survey satisfaction data was pure integer format, we would have checked the **Integer Data** option).



Note: For SigmaXL Mac version, click **SigmaXL > Graphical Tools > Basic Histogram Options** to set options. The Histogram Basic Chart is displayed after you click **Finish**.

6. Click **Finish**. A histogram of Overall Customer Satisfaction is produced.

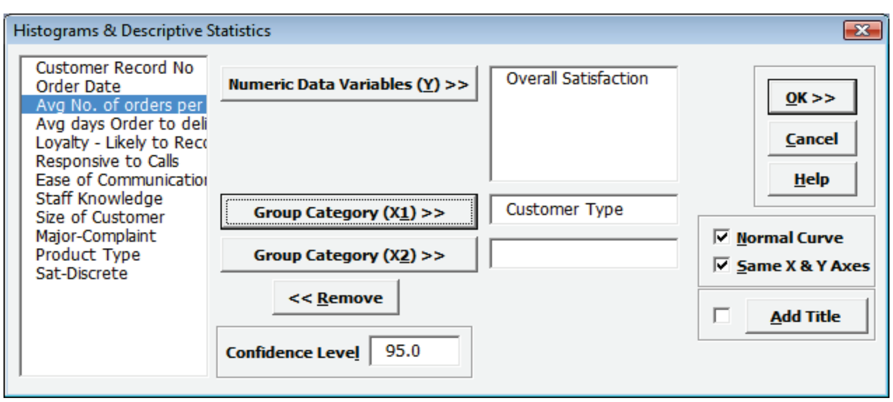


7. Note that bin one is 1 to < 1.5, bin 2 is 1.5 to < 2, etc.

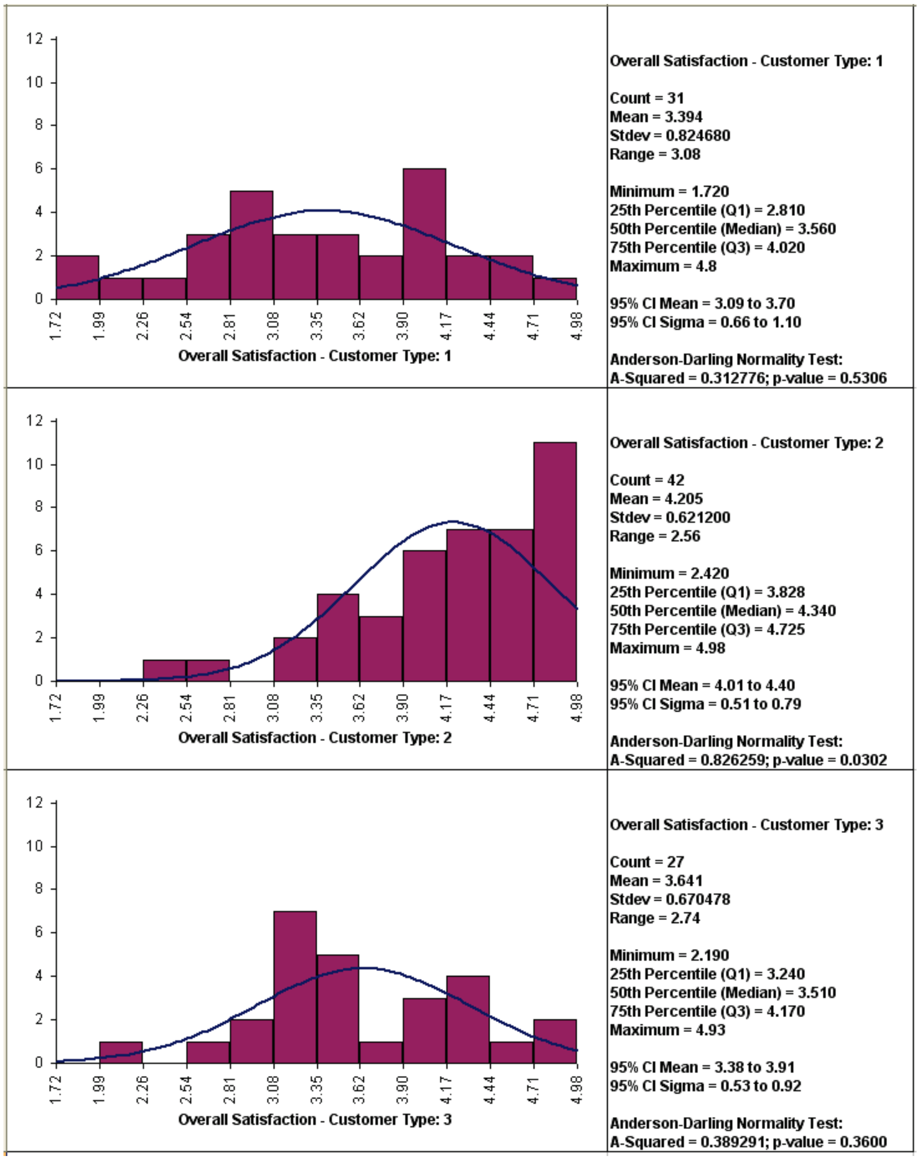
   Tip: Any graph produced by SigmaXL can be Copied/Pasted into Word. It can also be enlarged by clicking on the graph and dragging the corner. The number of decimal places displayed can be modified by clicking on the Axis Label and selecting the **Number** tab to adjust. The text label alignment can also be modified by: Select **Axis**, Right Click, **Format Axis.**

Multiple Histograms

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Graphical Tools > Histograms & Descriptive Statistics**.
3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
4. Select *Overall Satisfaction*, click **Numeric Data Variables (Y) >>**, select *Customer Type*, click **Group Category (X1) >>** as shown:



5. Click **OK**. Multiple Histograms and Descriptive Statistics of Customer Satisfaction By Customer Type are produced:



Clearly Customer Type 2 shows a higher level of overall satisfaction, with the data skewed left. Note that Customer Type 1 and 3 have data that is normally distributed, but this is not desirable when the response is a satisfaction score!

6. Note that bin one is 1.72 to < 1.99, bin 2 is 1.99 to < 2.26, etc. The number of decimals displayed can be changed by double-clicking on the X axis, click Number tab, and adjust decimal places.

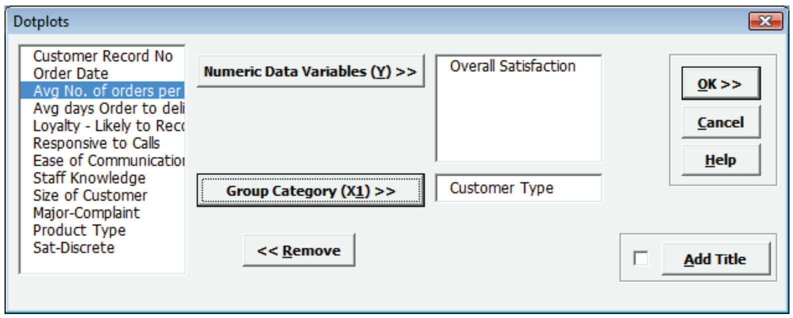
Part E – Dotplots

Dotplots

1. Click Sheet 1 Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet). Click **SigmaXL > Graphical Tools > Dotplots**.

2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**, select *Customer Type*, click **Group Category (X1) >>**:



4. Click **OK**. Multiple dotplots of Customer Satisfaction By Customer Type are produced:

5. Since all data points are shown, dotplots are a useful alternative to Histograms, particularly when the group sample sizes are small (n < 30).

Part F – Boxplots & Multiple X Boxplots

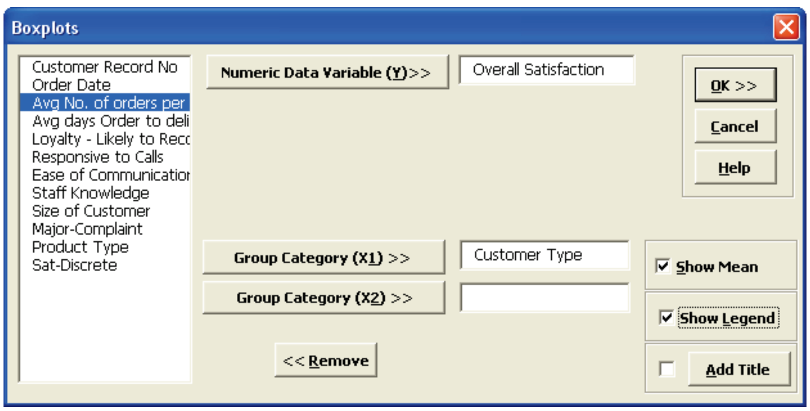
Boxplots

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).

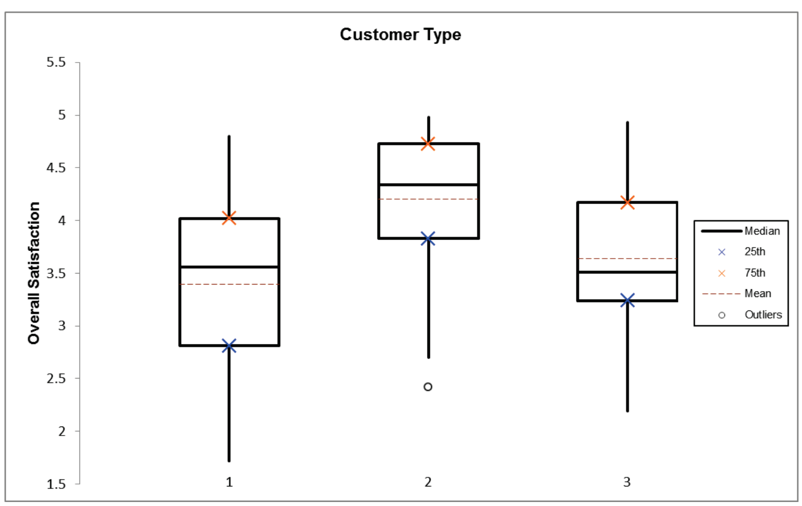
2. Click **SigmaXL > Graphical Tools > Boxplots**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

4. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**, select *Customer Type*, click **Group Category (X1) >>**, check **Show Mean**, check **Show Legend**:

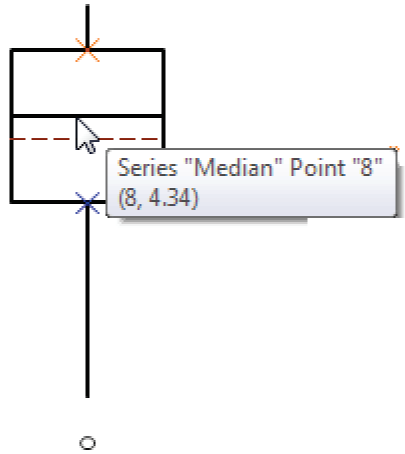


5. Click **OK**. A boxplot of Customer Satisfaction by Customer Type is produced:

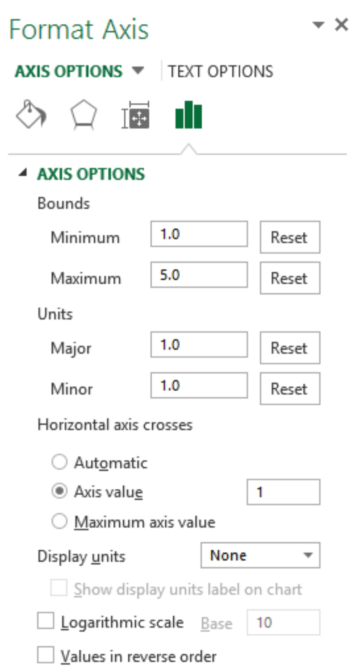


6. The legend indicates that the solid center line is the median. The dashed red line shows the sample mean. The top of the box is the 75th percentile (Q3). The bottom of the box is the 25th percentile (Q1). The height of the box is called the Inter-Quartile Range (IQR) and is a robust measure of spread or sample variability. The data point highlighted for Customer Type 2 is a potential outlier (< Q1 – 1.5 * IQR or > Q3 + 1.5 * IQR). Note that extreme outliers are highlighted with a solid dot (< Q1 – 3 * IQR or > Q3 + 3 * IQR).

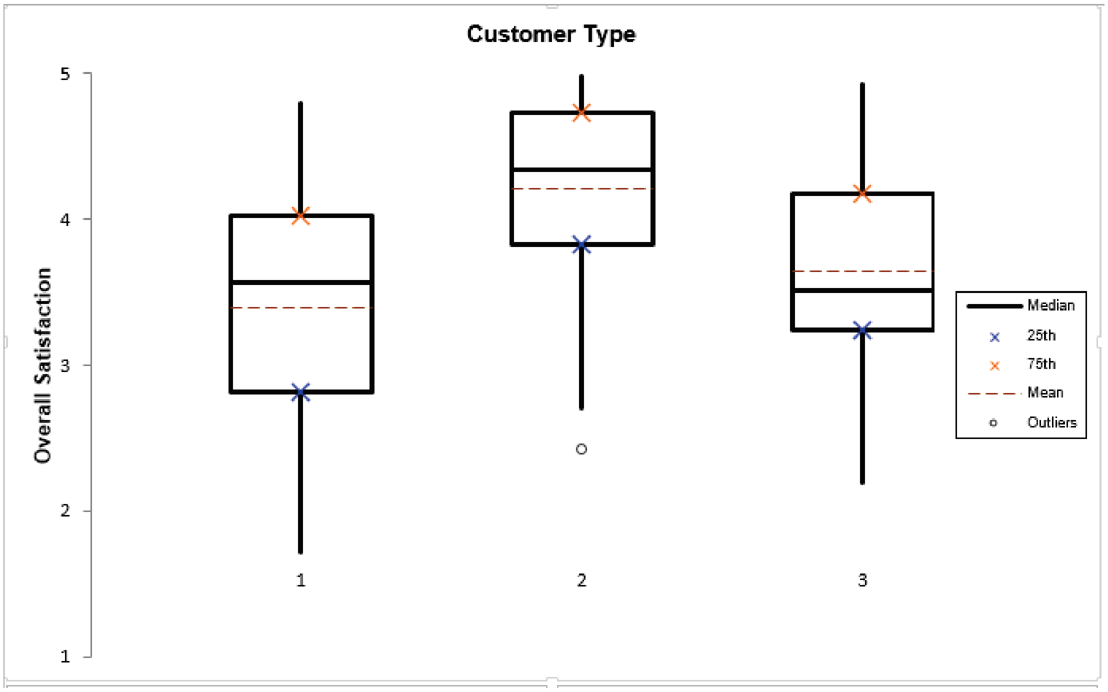
Tip: If you hover your mouse cursor in the middle of any of the Boxplot lines as shown, you will see the balloon help indicating what the line is and its numeric Y value.



7. Now we will modify the Y axis scale, showing 1 as minimum and 5 as maximum (given that the response data comes from a survey with 1-5 scale). To do this right click on the Y axis and select "Format Axis", modify the minimum value and maximum value. Change **Horizontal axis crosses > Axis value** to 1 as shown:

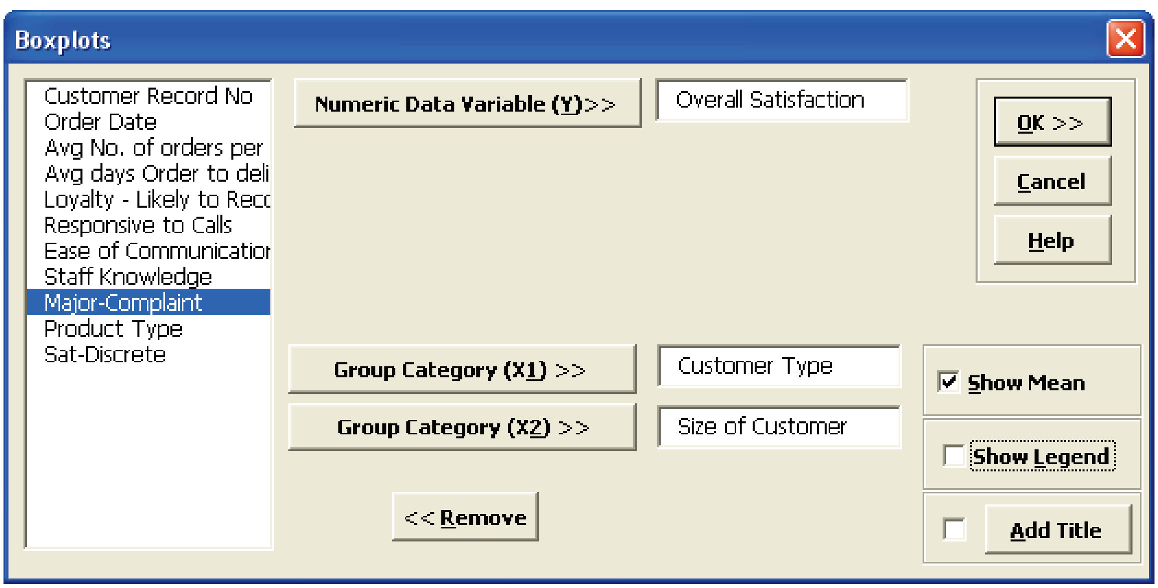


8. The Boxplot axis is modified as shown below:

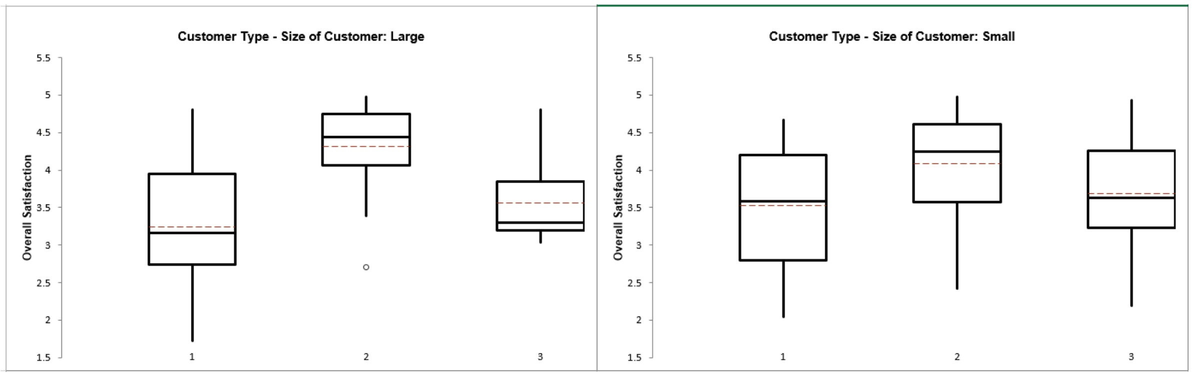


9. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.

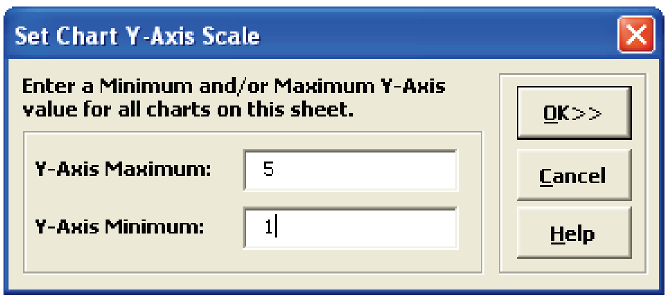
10. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X1) >>**; select *Size of Customer*, click **Group Category (X2) >>**; check **Show Mean;** uncheck **Show Legend**:



11. Click **OK**. Boxplots of Customer Satisfaction By Customer Type and Size are produced:



12. In order to adjust the Y-axis scale for both charts, click **SigmaXL Chart Tools > Set Chart Y-Axis Max/Min**.



13. Click **OK**. The Y-axis scale maximum and minimum are now modified for both charts.

Multiple X Boxplots

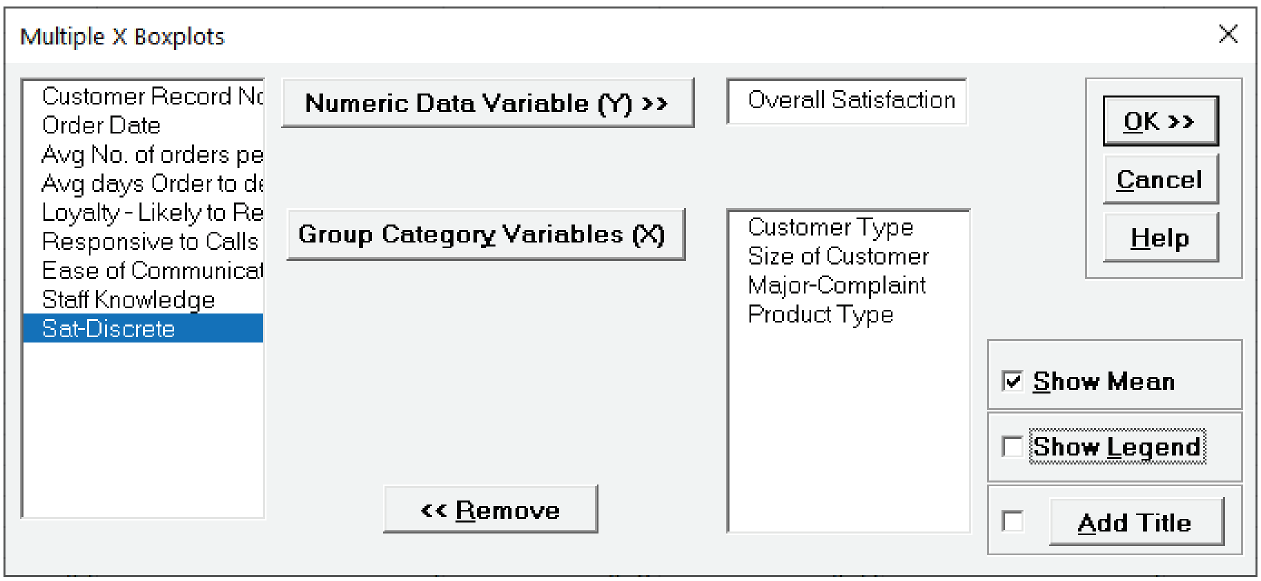
Multiple X Boxplots allow you to create boxplots with one Y variable and multiple group category X's. A row of boxplots will be created, one for each X variable. This is useful for easy comparison of the effect of each category X.

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).

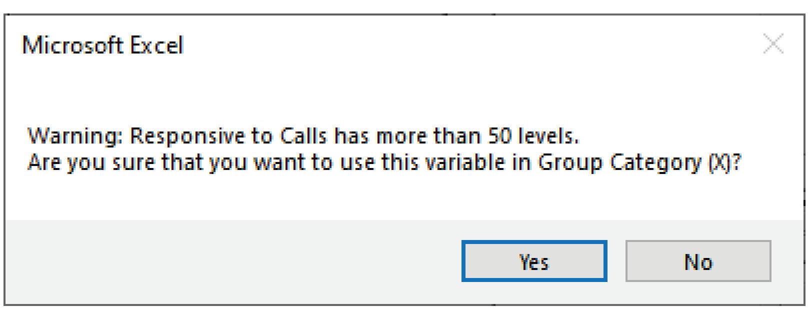
2. Click **SigmaXL > Graphical Tools > Multiple X Boxplots**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

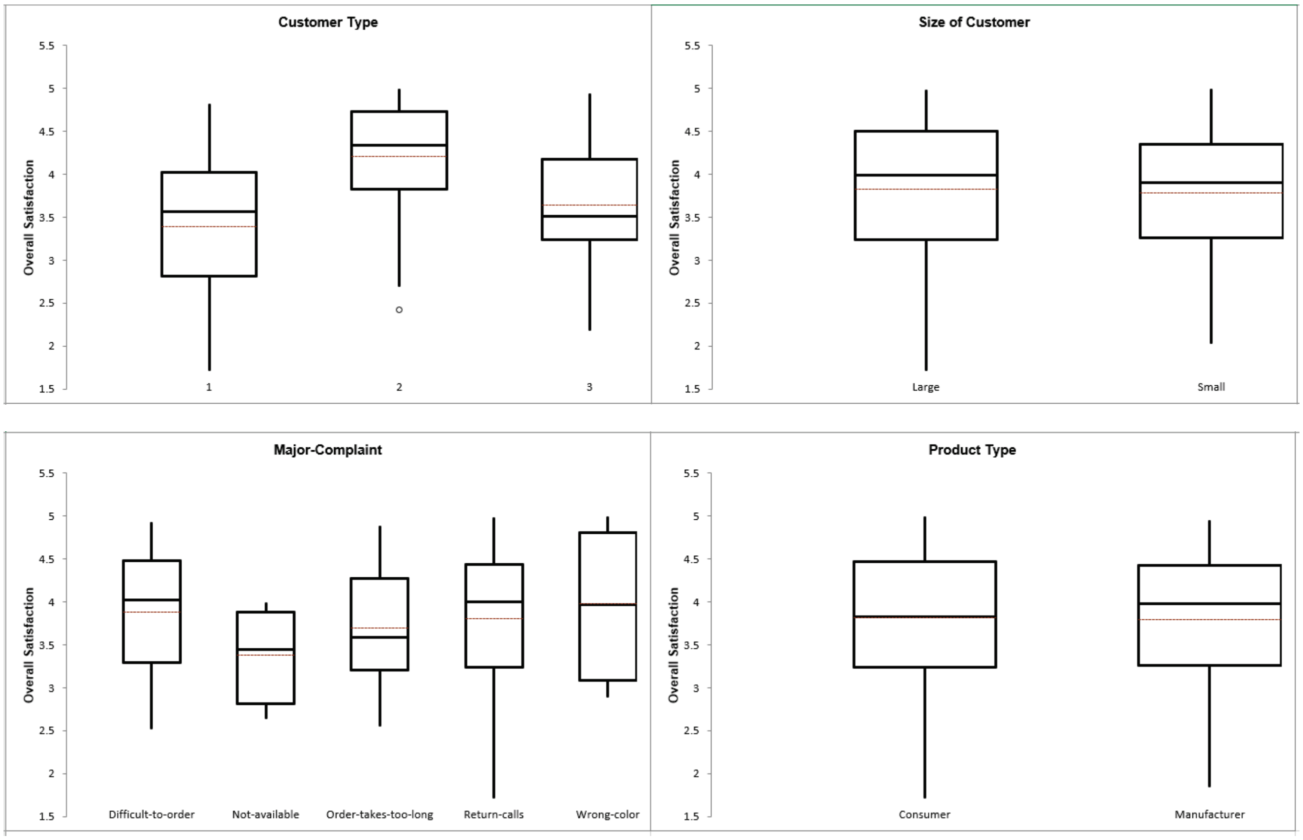
4. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**. Select *Customer Type*, *Size of Customer, Major-Complaint,* and *Product Type.* Click **Group Category Variables (X) >>**. Check **Show Mean**, uncheck **Show Legend**:



Tip: Be careful to not select a continuous variable for **Group Category Variables (X)**, as each unique value will be considered as a category level. If SigmaXL detects a variable with more than 50 levels a warning is given. For example, if *Responsive to Calls* was selected:



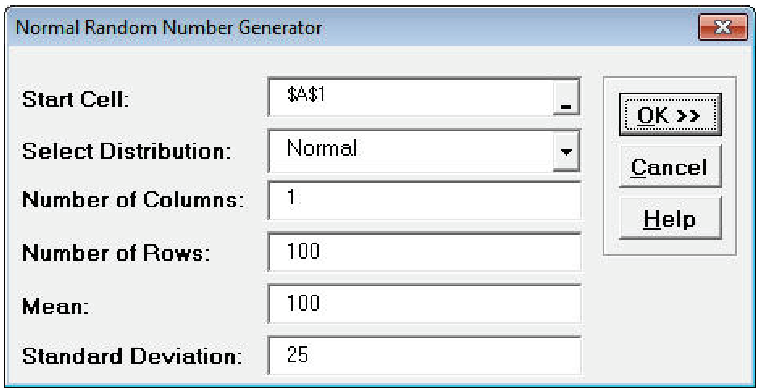
5. Click **OK**. A row of boxplots showing Overall Satisfaction by Customer Type, Size of Customer, Major-Complaint and Product Type are produced (note that this will take about one minute):



Part G – Normal Probability Plots and Empirical/Normal CDF Plots

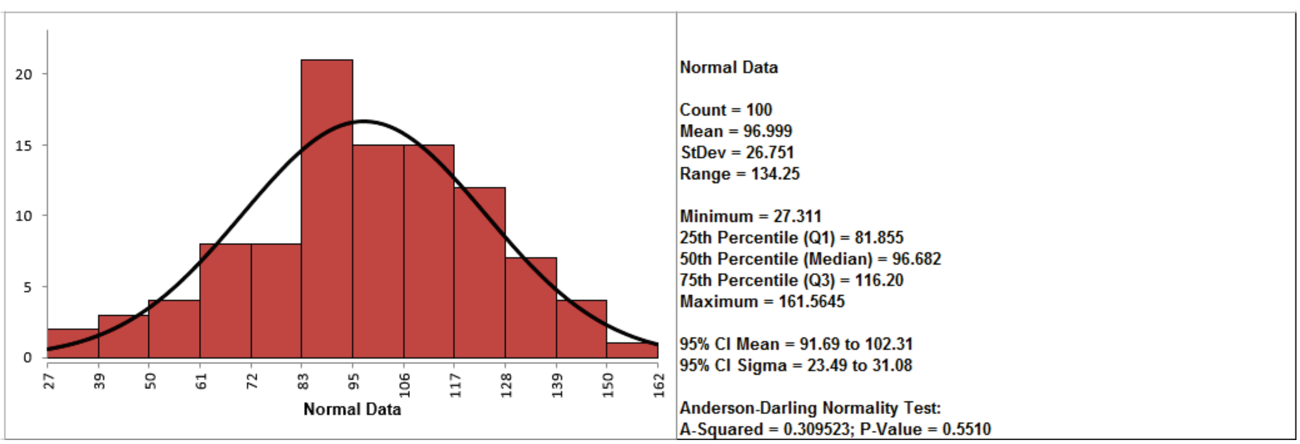
Normal Probability Plots

1. Create 100 random normal values as follows: **Click SigmaXL > Data Manipulation > Random Data > Normal**. Specify 1 Column, 100 Rows, Mean of 100 and Standard Deviation of 25 as shown below:



2. Click **OK**. Change Column heading to *Normal Data*.

3. Create a Histogram & Descriptive Statistics for this data. Your data will be slightly different due to the random number generation:



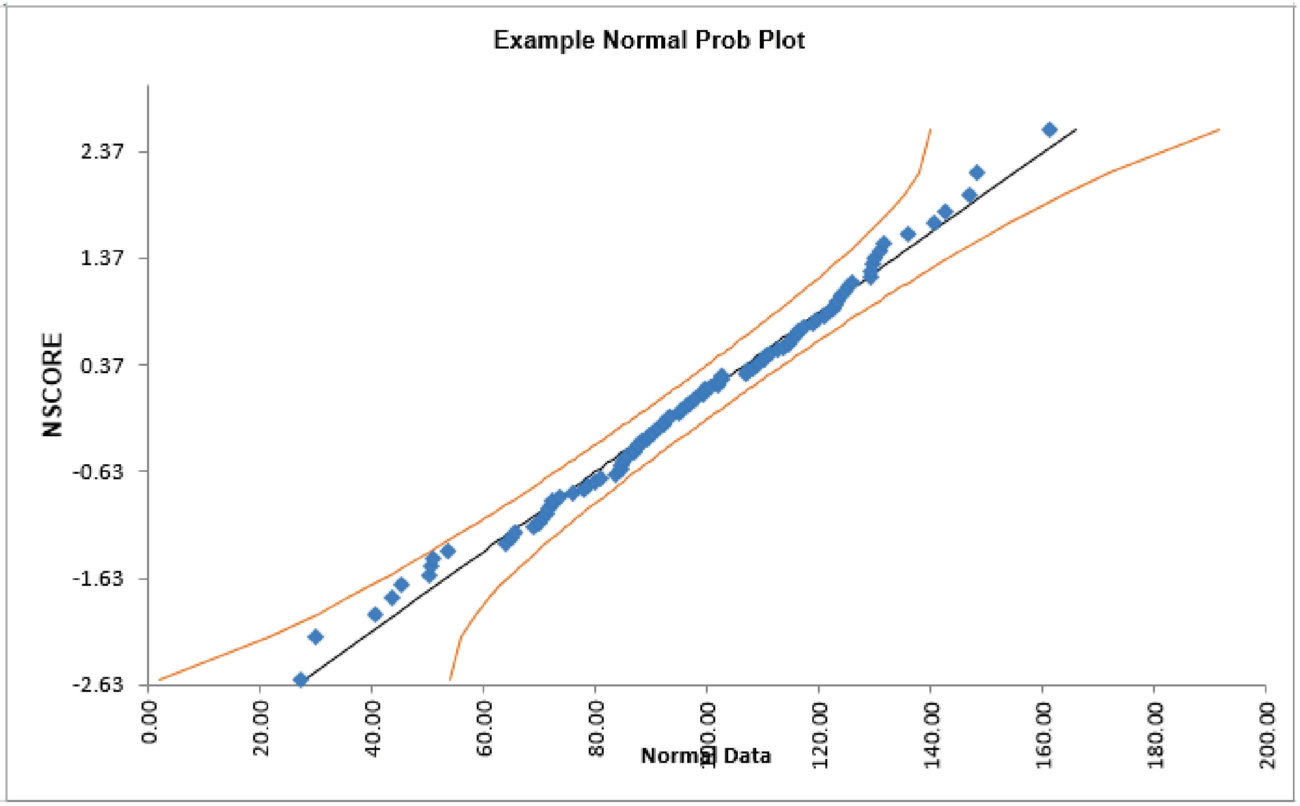
If the P-Value of the Anderson-Darling Normality test is greater than or equal to .05, the data is considered to be normal (interpretation of P-Values will be discussed further in Analyze).

4. Create a normal probability plot of this data: Click **SigmaXL Random Data (1)** Sheet, Click **SigmaXL > Graphical Tools > Normal Probability Plots**.

5. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

6. Select *Normal Data*, click **Numeric Data Variable (Y) >>**. Check **Add Title**. Enter *Example Normal Prob Plot*. Click **OK**.

7. Click **OK**. A Normal Probability Plot of simulated random data is produced (again, your plot will be slightly different due to the random number generation):



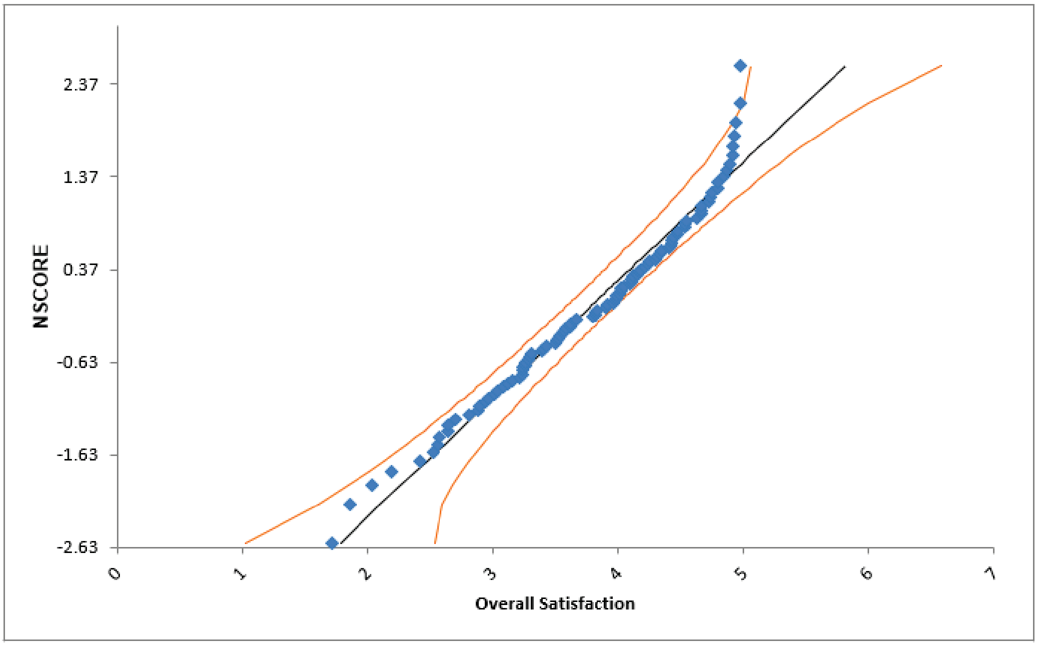
The data points follow the straight line fairly well, indicating that the data is normally distributed. Note that the data will not likely fall in a perfectly straight line. The eminent statistician George Box uses a "Fat Pencil" test where the data, if covered by a fat pencil, can be considered normal! We can also see that the data is normal since the points fall within the normal probability plot 95% confidence intervals (confidence intervals will be discussed further in Analyze).

8. Click **Sheet 1** Tab of **Customer Data.xlsx**.

9. Click **SigmaXL > Graphical Tools > Normal Probability Plots**.

10. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

11. Select *Overall Satisfaction*; click **Numeric Data Variable (Y) >>**. Click **OK**. A Normal Probability Plot of Customer Satisfaction data is produced:

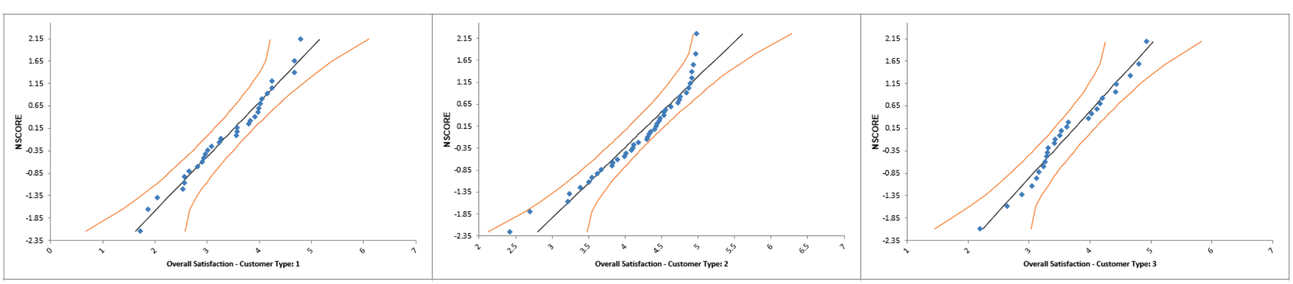


Is this data normally distributed? See earlier histogram and descriptive statistics of Customer Satisfaction data.

12. Now we would like to stratify the customer satisfaction score by customer type and look at the normal probability plots.

13. Click **Sheet 1** of **Customer Data.xlsx**. Click **SigmaXL > Graphical Tools > Normal Probability Plots**. Ensure that **Entire Table** is selected, click **Next**. (Alternatively, press **F3** or click **Recall SigmaXL Dialog** to recall last dialog).

14. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type* as **Group Category (X) >>**. Click **OK**. Normal Probability Plots of Overall Satisfaction by Customer Type are produced:



Reviewing these normal probability plots, along with the previously created histograms and descriptive statistics, we see that the satisfaction data for customer type 2 is not normal, and skewed left, which is desirable for satisfaction data! Note that although the customer type 2 data falls within the 95% confidence intervals, the Anderson Darling test from descriptive statistics shows $p < .05$ indicating nonnormal data. Smaller sample sizes tend to result in wider confidence intervals, but we still see that the curvature for customer type 2 is quite strong.

Tip: Use the Normal Probability Plot (NPP) to distinguish reasons for nonnormality. If the data fails the Anderson-Darling (AD) test (with $p < 0.05$) and forms a curve on the NPP, it is inherently nonnormal or skewed. Calculations such as Sigma Level, Pp, Cp, Ppk and Cpk assume normality and will therefore be affected. Consider transforming the data using LN(Y) or SQRT(Y) or using the Box-Cox Transformation tool (**SigmaXL > Data Manipulation > Box-Cox Transformation**) to make the data normal. Of course, whatever transformation you apply to your data, you must also apply to your specification limits. See also the Process Capability for nonnormal data tools.

If the data fails the AD normality test, but the bulk of the data forms a straight line and there are some outliers, the outliers are driving the nonnormality. Do not attempt to transform this data! Determine the root cause for the outliers and take corrective action on those root causes.

Empirical/Normal CDF Plots

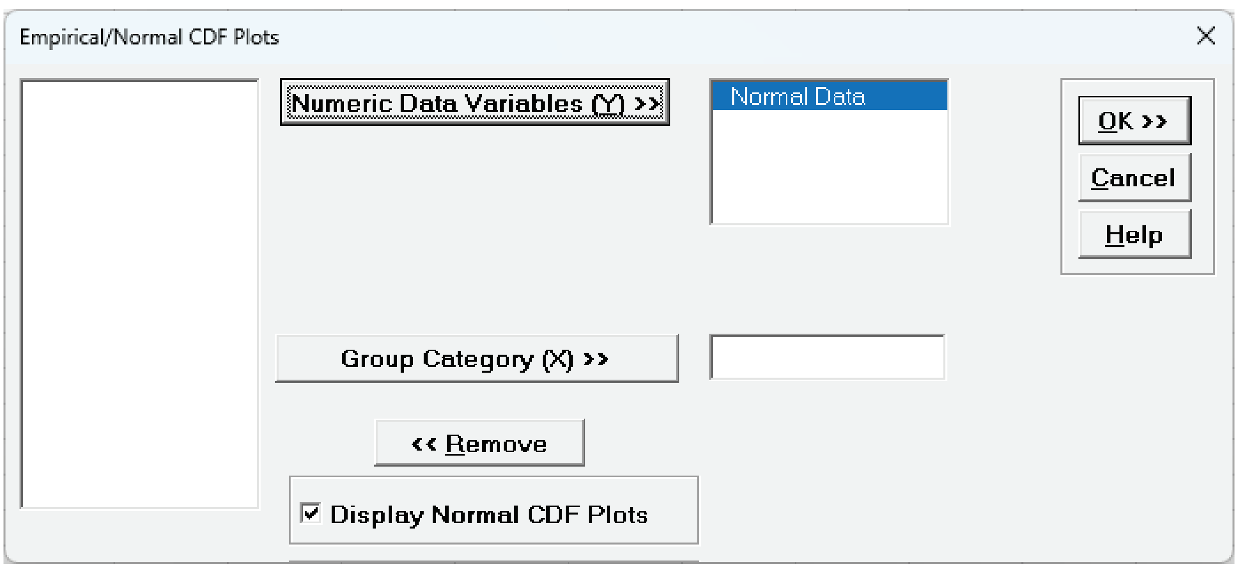
The Empirical/Normal CDF (Cumulative Distribution Function) Plot shows the data values sorted from lowest to highest on the X axis with the respective percentiles (percentages) on the Y axis and may be compared against the same for the fitted Normal Distribution. It is similar to the Normal Probability Plot but instead of a straight line it will form an "S-shaped" curve. The empirical data is plotted as a blue stepped line, whereas the fitted normal distribution is shown as a smooth red line. Large deviations between the two indicate that the data are not normally distributed.

Examples - Empirical/Normal CDF Plots

1. We will repeat the examples used above for Normal Probability Plots. Click **SigmaXL Random Data (1)** Sheet (if not available please do Steps 1 and 2 given above).

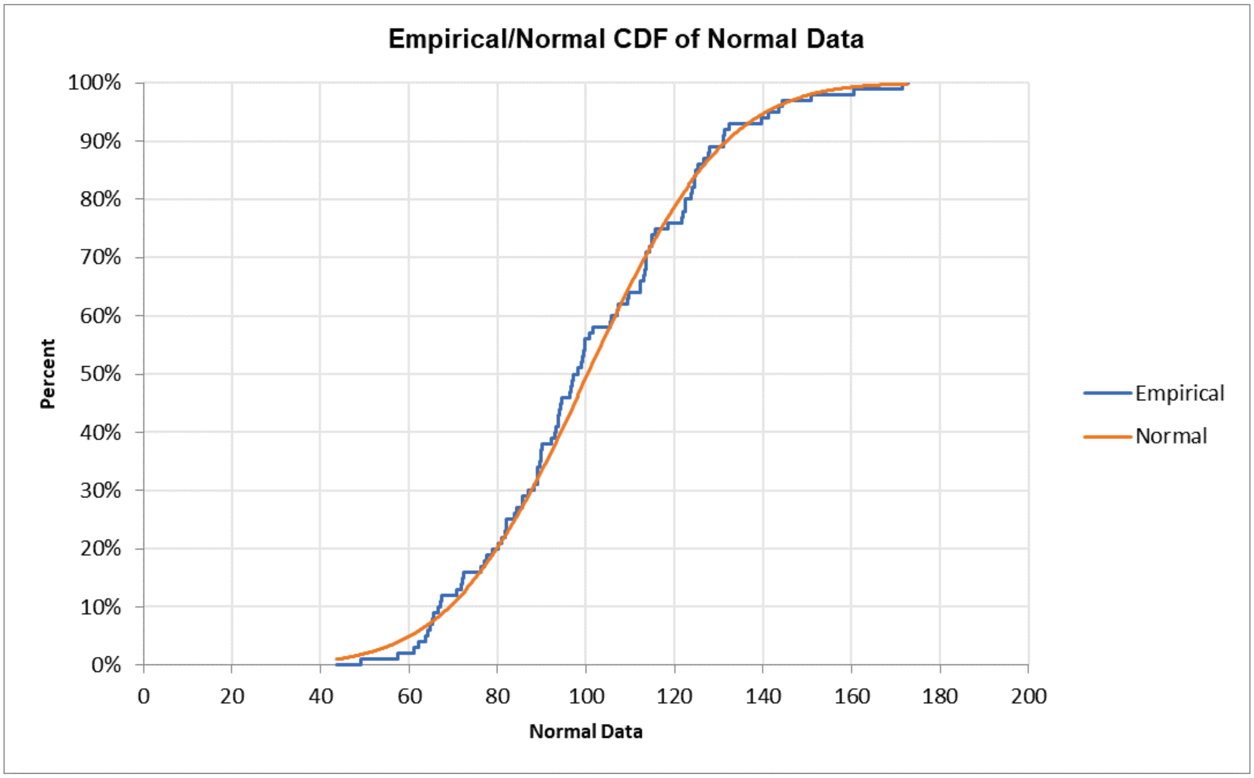
2. Click **SigmaXL > Graphical Tools > Empirical/Normal CDF Plots**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Normal Data*, click **Numeric Data Variables (Y) >>**. Check **Display Normal CDF Plots**.



Note that Empirical/Normal CDF Plots permit multiple Y variables. If more than one Y is selected **Group Category (X)** is greyed out.

4. Click **OK**. An Empirical/Normal CDF Plot of the simulated random data is produced (your plot will be slightly different due to the random number generation):

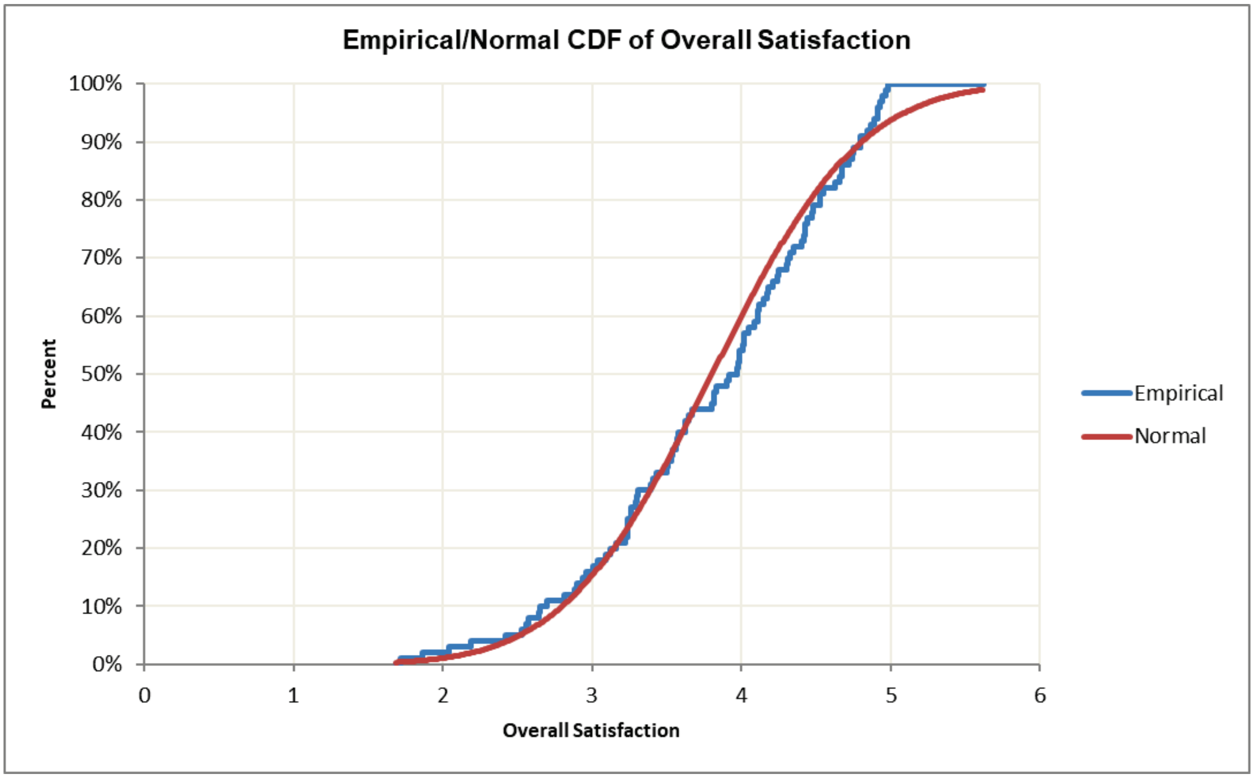


The Empirical CDF plot follows the Normal CDF fairly well, indicating that the data is normally distributed.

5. Click **Sheet 1** Tab of **Customer Data.xlsx**.

6. Click **SigmaXL > Graphical Tools > Empirical/Normal CDF Plots**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

7. Select *Overall Satisfaction*; click **Numeric Data Variables (Y) >>**. Click **OK**. An Empirical/Normal CDF Plot of the Overall Satisfaction data is produced:

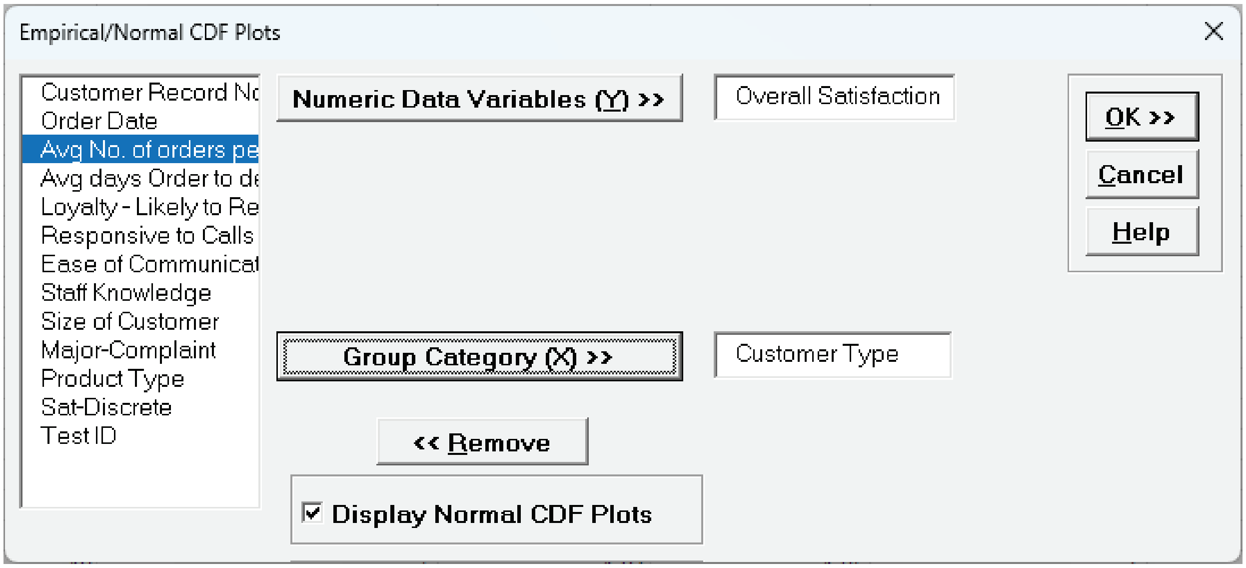


The Empirical CDF plot does not follow the Normal CDF well, indicating that the data is not normally distributed. The Empirical curve hits 100% at Overall Satisfaction = 5 since that was the maximum possible survey value, resulting in a skewed left distribution as noted in the histogram.

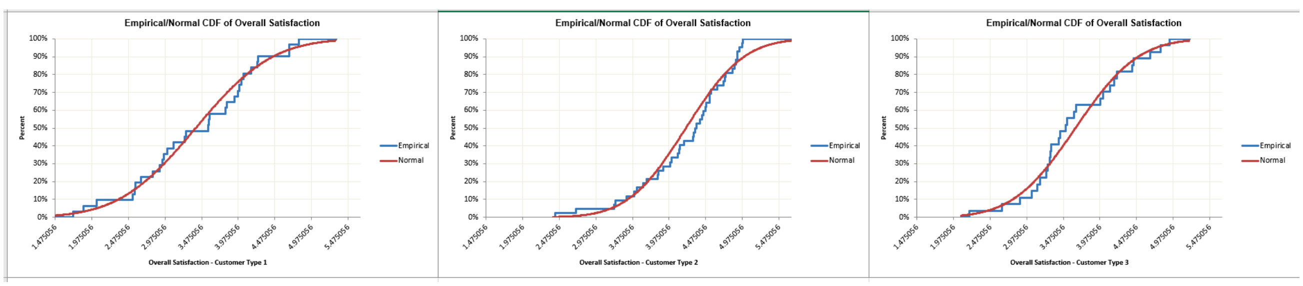
8. Now we would like to stratify the customer satisfaction score by customer type and look at the Empirical/Normal CDF plots.

9. Click **Sheet 1** of **Customer Data.xlsx**. Click **SigmaXL > Graphical Tools > Empirical/Normal CDF Plots**. Ensure that **Entire Table** is selected, click **Next**. (Alternatively, press **F3** or click **Recall SigmaXL Dialog** to recall last dialog).

10. Select *Overall Satisfaction*, click **Numeric Data Variables (Y) >>**; select *Customer Type* as **Group Category (X) >>**.



11. Click **OK**. Empirical/Normal CDF Plots of Overall Satisfaction by Customer Type are produced:



We can see that the Empirical CDF plot for Customer Type 2 does not follow the Normal CDF well, indicating that the data is not normally distributed. The Empirical curve hits 100% at Overall Satisfaction = 5 since that was the maximum possible survey value, resulting in a skewed left distribution as noted in the histogram.

Customer Types 1 and 3 are harder to interpret, so we would use normal probability plots and normality tests to complement these plots to assess normality.

Tip: Empirical CDF Plots may also be compared against each other by using Excel's copy/paste for the graphs. In this case, it is recommended to uncheck the **Display Normal CDF Plots** option when creating the plots. The Two Sample KS Test is a formal test used to compare two empirical CDFs. See **Two Sample Mann-Whitney Test (with 2 Sample KS Option)**.

Part H– Run Charts

Run charts, also known as trend charts and time series plots, add the dimension of time to the graphical tools. They allow us to see trends and outliers in the data. Run Charts are a precursor to control charts, which add calculated control limits. Note that Run Charts should be used only on unsorted data, in its original chronological sequence.

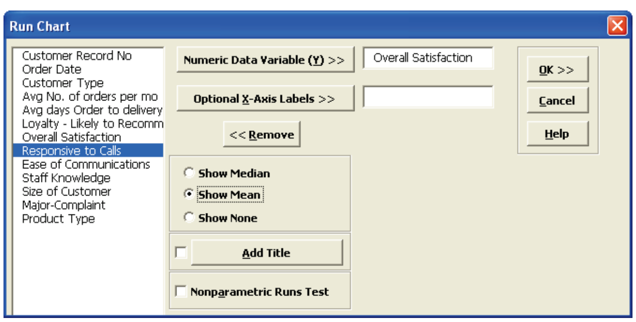
Basic Run Chart Template

Click **SigmaXL > Templates and Calculators** > **Basic Graphical Templates > Run Chart** or **SigmaXL > Graphical Tools > Basic Graphical Templates > Run Chart**. See **Part B – Templates and Calculators** for a Run Chart Template example.

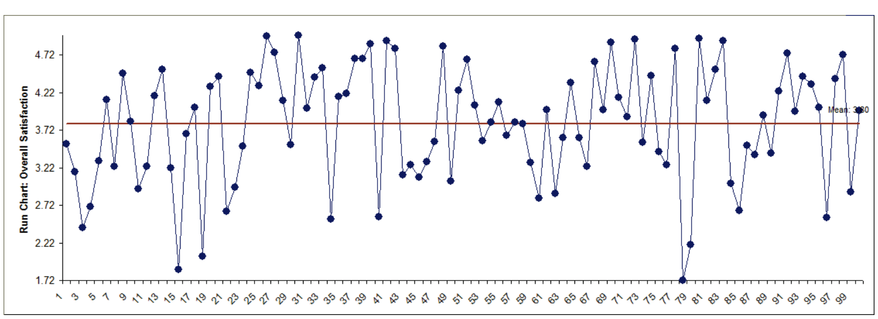
Run Charts

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet). Click **SigmaXL > Graphical Tools > Run Chart**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

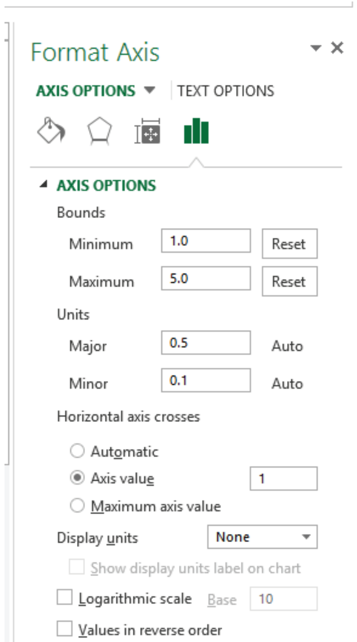
2. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test** (to be discussed later in Part N of Analyze Phase).



3. Click **OK**. A Run Chart of Overall Satisfaction with Mean center line is produced.

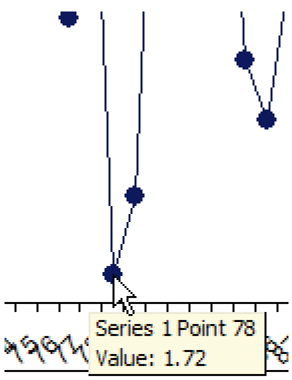


4. Select the Y axis, Right Click, **Format Axis**, to activate the **Format Axis** dialog. Change **Minimum** to 1, **Maximum** to 5, **Horizontal axis crosses > Axis value** to 1:

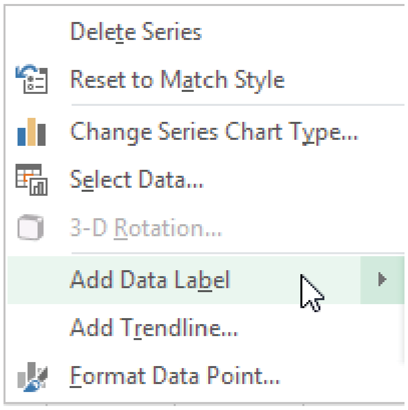


5. The chart is automatically updated.

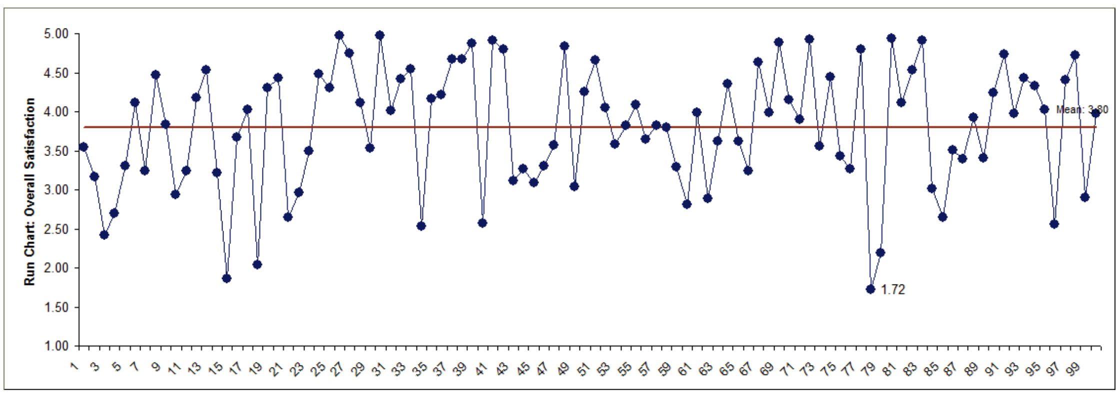
6. Are there any obvious trends? Some possible cycling, but nothing clearly stands out. It may be interesting to look more closely at a specific data point. Any data point value can be identified by simply moving the cursor over it:



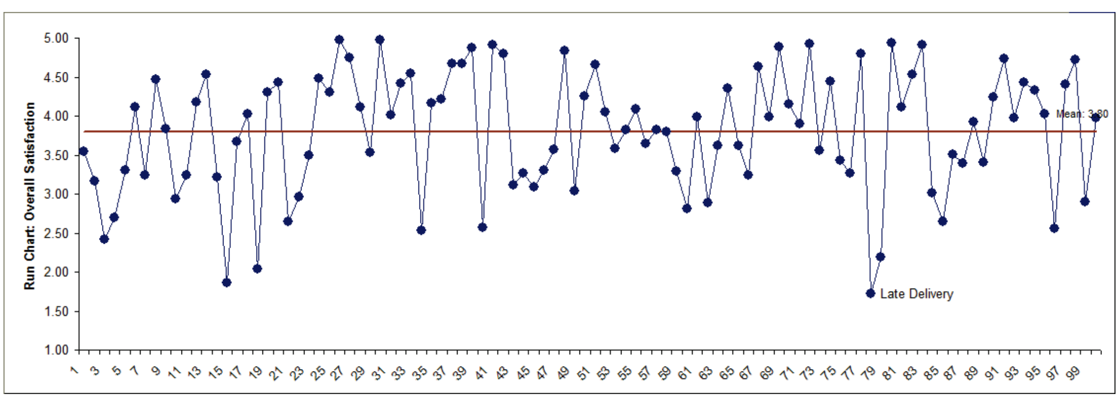
7. A label can be added to a data point by one single-click on the data point, followed by a Right mouse click, and select **Add Data Label**. See also **SigmaXL Chart Tools > Add Data Label** in Control Phase Tools, Part B - X-Bar & Range Charts.



8. Click **OK**. Resulting Run Chart with label attached to data point:



9. This label can be changed to a text comment. Single-click three times on the label and type in a comment as shown:

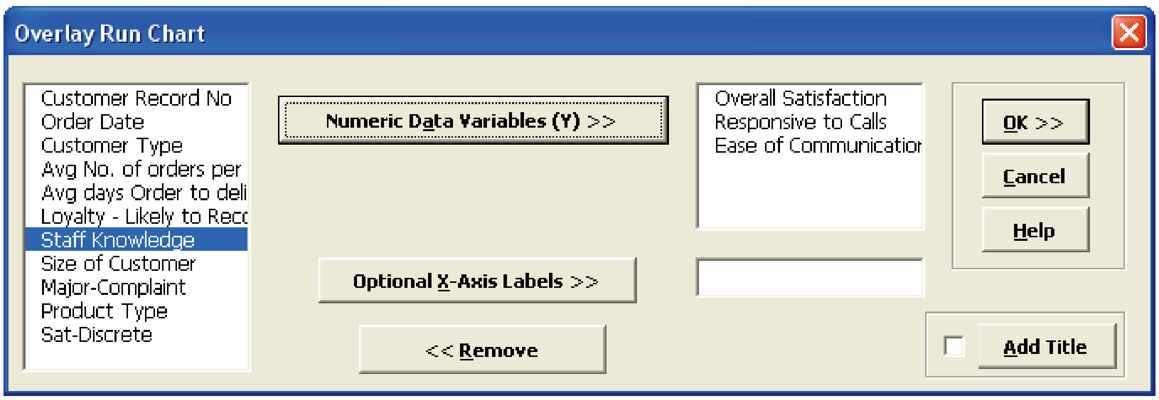


Overlay Run Charts

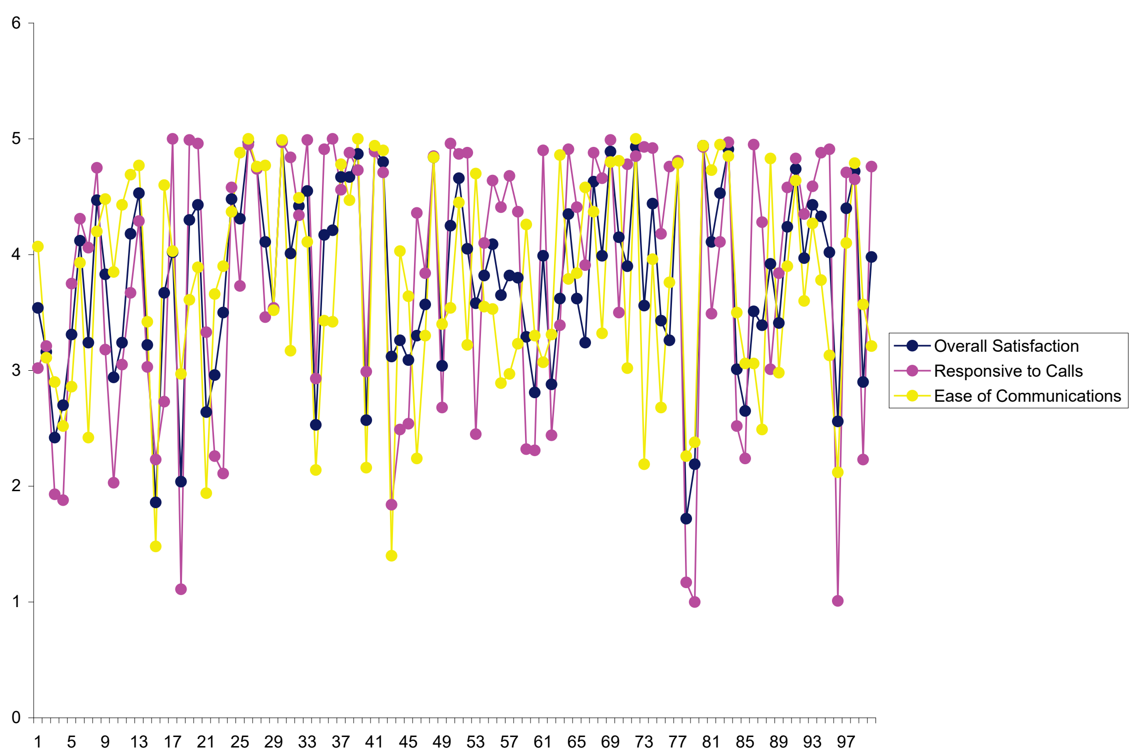
1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press F4 to activate last worksheet). Click **SigmaXL > Graphical Tools > Overlay Run Chart**.

2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Overall Satisfaction, Responsive to Calls and Ease of Communications*. Click **Numeric Data Variable (Y) >>**.



4. Click **OK**. An Overlay Run Chart of Overall Satisfaction, Responsive to Calls and Ease of Communications is produced.



Part I – Measurement Systems Analysis

Type 1 Gage Study Template

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Type 1 Gage Study** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Type 1 Gage Study**.

See **Basic MSA Templates – Type 1 Gage Study** for an example of the Type 1 Gage Study template.

Gage Bias and Linearity Study Template

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Gage Bias and Linearity Study** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Gage Bias and Linearity Study**.

See **Basic MSA Templates – Gage Bias and Linearity Study** for an example of the Gage Bias and Linearity Study template.

Gage R&R Study (MSA) Template

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Gage R&R Study (MSA)** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Gage R&R Study (MSA)**.

See **Basic MSA Templates – Gage R&R Study (MSA)** for an example of the Gage R&R Study (MSA) template. If you wish to analyze the template data with the Gage R&R Analysis Tool, click Create Stacked Column Format for “Analyze Gage R&R” >>.

After entering the data into the Gage R&R Study template, you can create Gage Multi-Vari & X-bar R charts. Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Gage R&R: Multi-Vari & X-bar R Charts** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Gage R&R: Multi-Vari & X-bar R Charts**. See **Basic MSA Templates - Gage R&R: Multi-Vari & X-bar R Charts** for an example.

Attribute Gage R&R (MSA) Template

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Attribute Gage R&R (MSA)** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Attribute Gage R&R (MSA)**.

See **Basic MSA Templates – Attribute Gage R&R (MSA)** for an example of the Attribute Gage R&R (MSA) template.

Orthogonal (Deming) Regression Template

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > Orthogonal (Deming) Regression** or **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > Orthogonal (Deming) Regression**.

See **Basic MSA Templates – Orthogonal (Deming) Regression** for an example of the Orthogonal (Deming) Regression template.

See the Appendix **Orthogonal (Deming) Regression** for formula details.

GLM GageRR (Crossed) Metrics with/without Interaction

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > GLM GageRR (Crossed) Metrics with Interaction** or **SigmaXL > Measurement Systems Analysis > GLM GageRR (Crossed) Metrics with Interaction**. Use this template to compute Gage R&R metrics from the Variance Components of General Linear Model (GLM) when the model includes a Part*Operator interaction term.

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > GLM GageRR (Crossed) Metrics without Interaction** or **SigmaXL > Measurement Systems Analysis > GLM GageRR (Crossed) Metrics without Interaction**. Use this template to compute Gage R&R metrics from the Variance Components of General Linear Model (GLM) when the model does not include a Part*Operator interaction term.

See **Example 3: Classical Gage R&R Study** for an example of the GLM GageRR (Crossed) Metrics without Interaction template used after a GLM analysis of a Gage R&R study. This is used to understand the relationship between a Classical Gage R&R study and the GLM analysis.

See **Example 4: Gage R&R Study with Operator as Fixed Factor** for an example of the GLM GageRR (Crossed) Metrics without Interaction template used after a GLM analysis of a Gage R&R study where Operator is treated as Fixed Factor not a Random Factor.

GLM GageRR (Nested) Metrics

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > GLM GageRR (Nested) Metrics** or **SigmaXL > Measurement Systems Analysis > GLM GageRR (Nested) Metrics**. Use this template to compute Gage R&R metrics from the Variance Components of General Linear Model (GLM) when the model is nested, for example, in a Destructive Gage R&R study.

See **Example 5: Destructive (Nested) Gage R&R** for an example of the GLM GageRR (Nested) Metrics template used after a GLM analysis of a Destructive Gage R&R study.

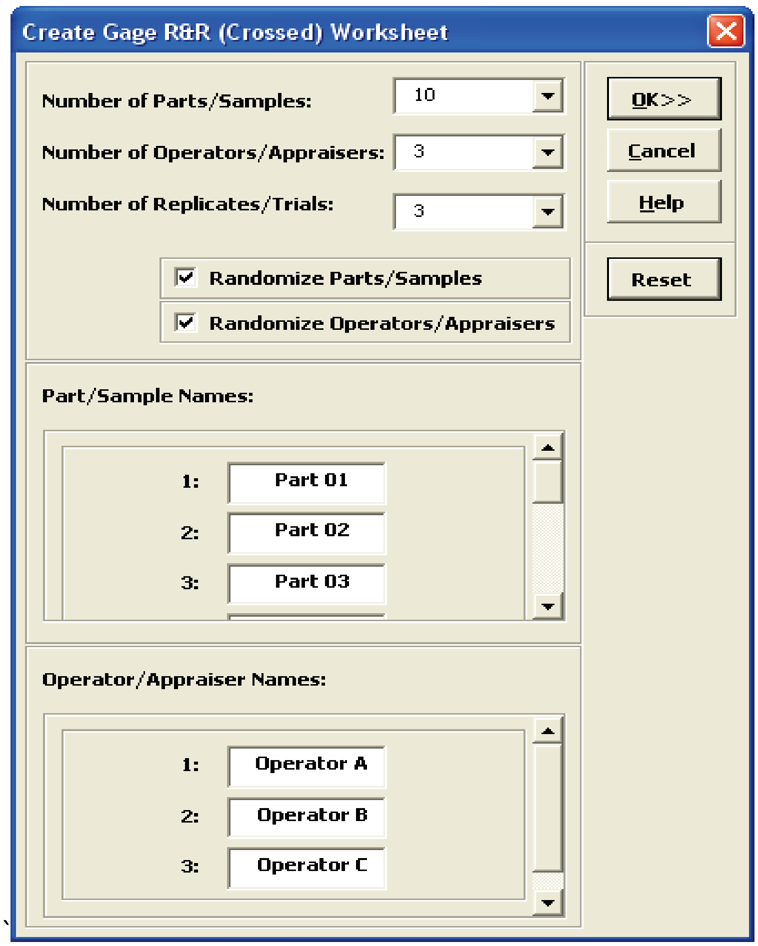
GLM GageRR (Expanded) Metrics

Click **SigmaXL > Templates and Calculators** > **Basic MSA Templates > GLM GageRR (Expanded) Metrics** or **SigmaXL > Measurement Systems Analysis > GLM GageRR (Expanded) Metrics**. Use this template to compute Gage R&R metrics from the Variance Components of General Linear Model (GLM) when the model is expanded and includes more than just Operator and Part variables.

See **Example 6: Expanded Gage R&R** for an example of the GLM GageRR (Expanded) Metrics template used after a GLM analysis of an Expanded Gage R&R study.

Create Gage R&R (Crossed) Worksheet

1. To create a blank Gage R&R worksheet, click **SigmaXL > Measurement Systems Analysis > Create Gage R&R (Crossed) Worksheet**. A crossed Gage R&R study means that every operator/appraiser will measure every part/sample.



2. The default settings are 10 Parts, 3 Operators, 3 Replicates with Parts and Operators randomized. Clicking the **Reset** button will restore the dialog to these default settings.

3. Part and Operator names can be edited.

4. Click **OK**. The following worksheet is produced:

Gage R&R Study (Crossed) Worksheet

Gage Name:	
Date of Study:	
Performed By:	
Notes:	

Run Order	Std. Order	Part	Operator	Measurement
1	86	Part 09	Operator C	
2	61	Part 01	Operator C	
3	65	Part 02	Operator C	
4	74	Part 05	Operator C	
5	64	Part 02	Operator C	
6	83	Part 08	Operator C	
7	71	Part 04	Operator C	
8	66	Part 02	Operator C	
9	62	Part 01	Operator C	
10	90	Part 10	Operator C	
11	76	Part 06	Operator C	
12	80	Part 07	Operator C	
13	73	Part 05	Operator C	
14	79	Part 07	Operator C	
15	63	Part 01	Operator C	

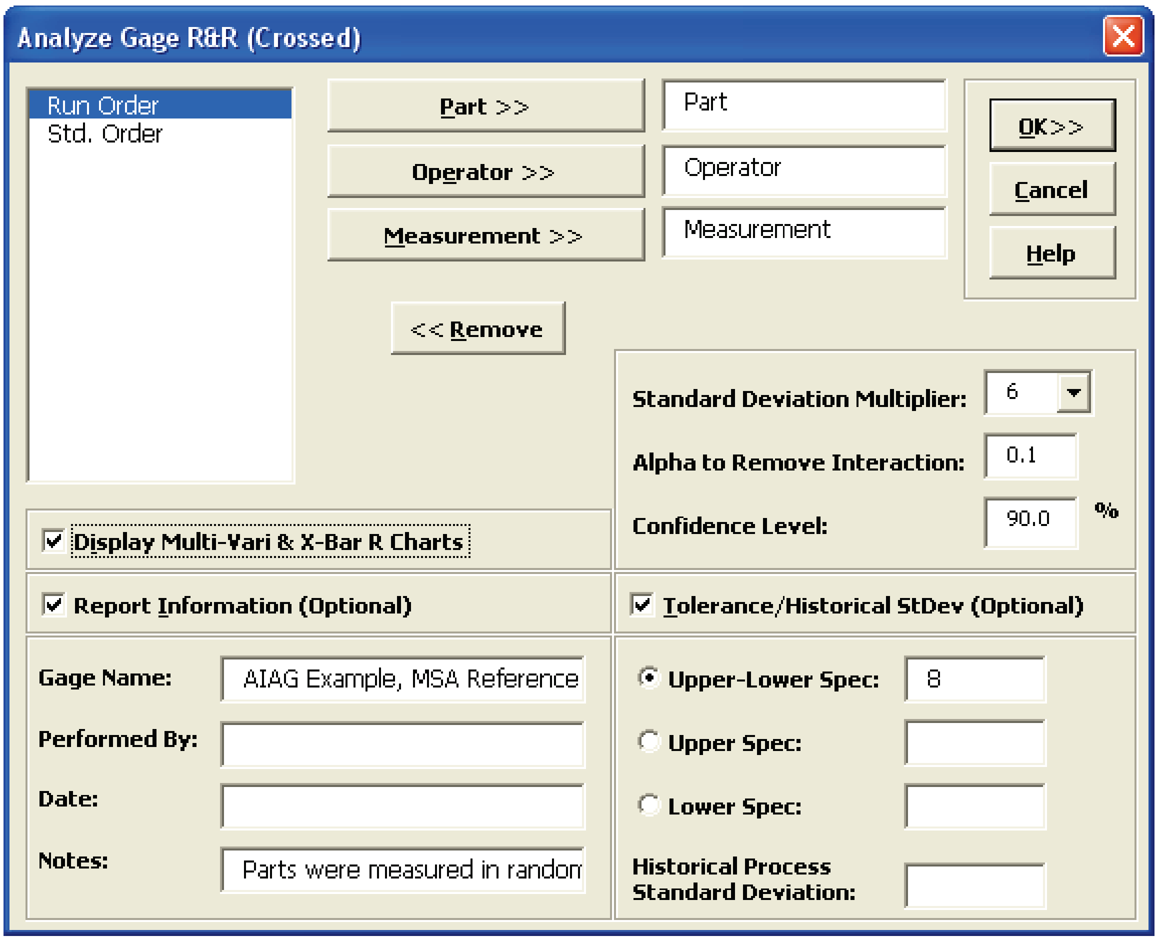
5. Your worksheet order will be different due to the randomization. The results of the Gage R&R study are entered in the yellow highlighted Measurement cells.

Analyze Gage R&R (Crossed)

1. Open the file **Gage RR – AIAG.xlsx**. This is an example from the Automotive Industry Action Group (AIAG) MSA Reference Manual, 3rd Edition, page 101. Note that parts were measured in random order, but the worksheet is given in standard order.

2. Click **SigmaXL > Measurement Systems Analysis > Analyze Gage R&R (Crossed)**. The data worksheet is recognized by SigmaXL and highlighted automatically. Click **Next**.

3. Select *Part*, *Operator* and *Measurement* as shown. Check **Display Multi-Vari & X-Bar R Charts**. Check **Tolerance/Historical StDev (Optional)** and enter 8 for the Process Tolerance **Upper-Lower Spec**.



4. Click **OK**. The following Gage R&R Study Report is produced:

Gage R&R Study (Crossed) Report

Gage Name:	AIAG Example, MSA Reference Manual, 3rd Edition, Page 101
Date of Study:	
Performed By:	
Notes:	Parts were measured in random order, but worksheet is given in standard order.

Process Tolerance (USL - LSL):	8
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6
Alpha to Remove Interaction:	0.1
Confidence Level:	90.0
Number of Parts:	10
Number of Operators:	3
Number of Replicates:	3
Design Type:	Balanced

Analysis of Variance with Part * Operator Interaction:

Source	DF	SS	MS	F	P
Part:	9	88.362	9.8180	492.29	0.0000
Operator:	2	3.1673	1.5836	79.406	0.0000
Part * Operator:	18	0.358982	0.019943457	0.433721	0.9741
Repeatability:	60	2.7589	0.045982222		
Total:	89	94.647	1.06		

Analysis of Variance without Part * Operator Interaction (P for Interaction >= 0.1):

Source	DF	SS	MS	F	P
Part:	9	88.362	9.8180	245.61	0.0000
Operator:	2	3.1673	1.5836	39.617	0.0000
Repeatability:	78	3.1179	0.039973276		
Total:	89	94.64711222	1.063450699		

Gage R&R Metrics	StDev	StDev Lower 90% CI	StDev Upper 90% CI	6 * StDev	% Total Variation (TV)	% TV Lower 90% CI	% TV Upper 90% CI
Gage R&R:	0.302372	0.235108	1.03	1.8142	27.86	15.46	70.93
Operator (AV Appraiser Variation):	0.226838	0.127545	1.01	1.3610	20.90		
Part * Operator (INT Interaction):	0	0	0	0	0.00		
Reproducibility (SQRT(AV^2 + INT^2)):	0.226838	0.127545	1.01	1.3610	20.90		
Repeatability (EV Equipment Variation):	0.199933	0.176915	0.230560	1.1996	18.42		
Part Variation (PV):	1.04	0.758821	1.7170	6.2540	96.04		
Total Variation (TV):	1.09	0.816099	1.8111	6.5118	100.00		

Gage R&R Metrics	% Tolerance	% Tolerance Lower 90% CI	% Tolerance Upper 90% CI
Gage R&R:	22.68	17.63	77.50
Operator (AV Appraiser Variation):	17.01	9.57	76.03
Part * Operator (INT Interaction):	0.00	0.00	0.00
Reproducibility (SQRT(AV^2 + INT^2)):	17.01	9.57	76.03
Repeatability (EV Equipment Variation):	14.99	13.27	17.29
Part Variation (PV):	78.17	56.91	128.78
Total Variation (TV):	81.40	61.21	135.83

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.091428538	7.76
Operator:	0.051455261	4.37
Part * Operator:	0	0.00
Reproducibility:	0.051455261	4.37
Repeatability:	0.039973276	3.39
Part Variation:	1.09	92.24
Total Variation:	1.1779	100.00

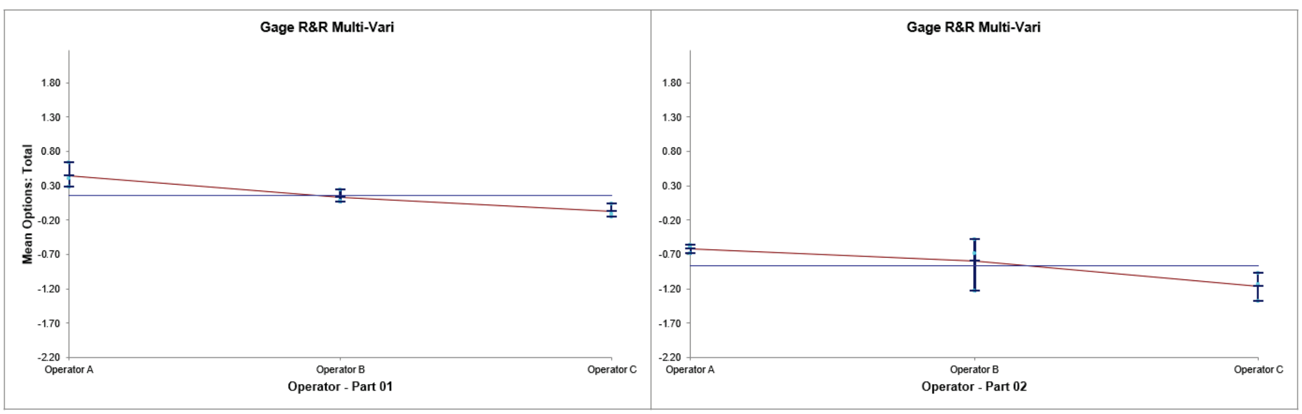
Gage R&R Metrics	NDC	NDC Lower 90% CI	NDC Upper 90% CI
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	4.9	1.4	9.0

5. The first ANOVA table shows that the Operator * Part Interaction is not significant with the P-Value = 0.9741 (greater than the alpha-to-remove value of 0.1). The second ANOVA table shows the results excluding the Part * Operator Interaction.

6. The cells shaded in light blue highlight the critical metrics: Gage R&R % Total Variation (also known as %R&R) and %Tolerance: < 10% indicates a good measurement system; > 30% indicates an unacceptable measurement system. The Number of Distinct Categories should be at least 5. Initial analysis shows that this is a marginal measurement system. (Traditionally NDC is truncated to an integer value, but SigmaXL reports a more informative one decimal place).

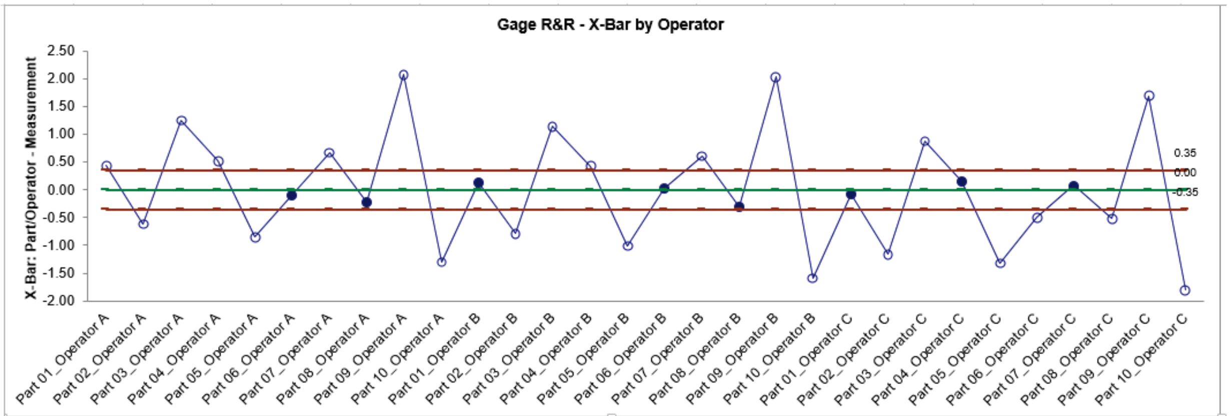
7. SigmaXL also reports the confidence intervals for the Gage R&R metrics. Note that when the confidence intervals are taken into account, it is possible that this measurement system is totally inadequate. This strongly suggests that the measurement system needs to be improved but it also points out the weakness of the traditional Gage R&R study with 10 Parts, 3 Operators, and 3 Replicates. These sample sizes will typically yield wide confidence intervals. For statistical details of the Gage R&R confidence intervals see Burdick, R. K., Borror, C. M., and Montgomery, D. C., “Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models”, ASA-SIAM Series on Statistics and Applied Probability, 2005.

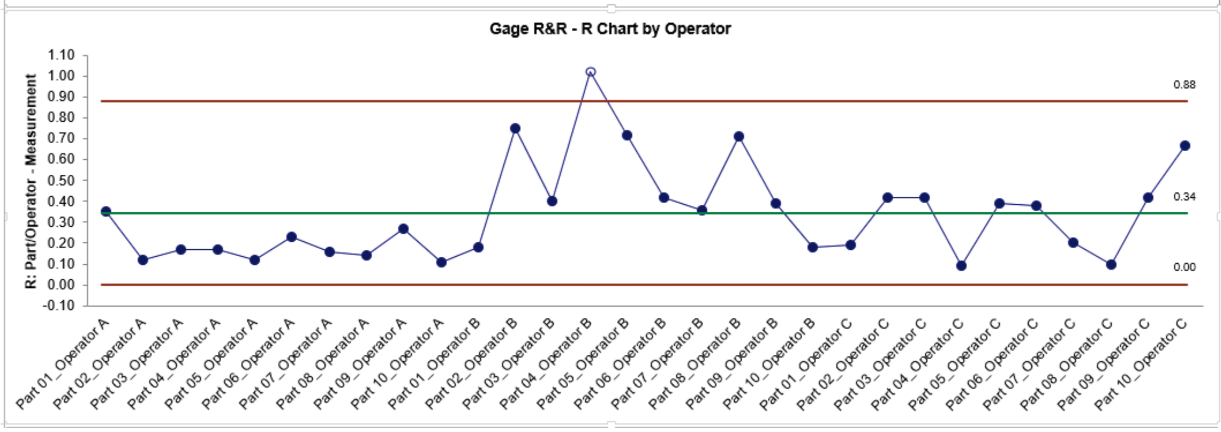
8. Click on the **Gage R&R – Multi-Vari Sheet** tab to view the Multi-Vari Chart:



The Multi-Vari chart shows each Part as a separate graph. Each Operator’s response readings are denoted as a vertical line with the top tick corresponding to the Maximum value, bottom tick is the Minimum, and the middle tick is the Mean. The horizontal line across each graph is the overall average for each part.

9. Click on the **Gage R&R – X-Bar R Sheet** tab to view the X-Bar & R Chart:





When interpreting the X-bar and R chart for a Gage R&R study, it is desirable that the X-bar chart be out-of-control, and the Range chart be in-control. The control limits are derived from within Operator repeatability. The Range chart indicates a problem with Operator B being inconsistent.

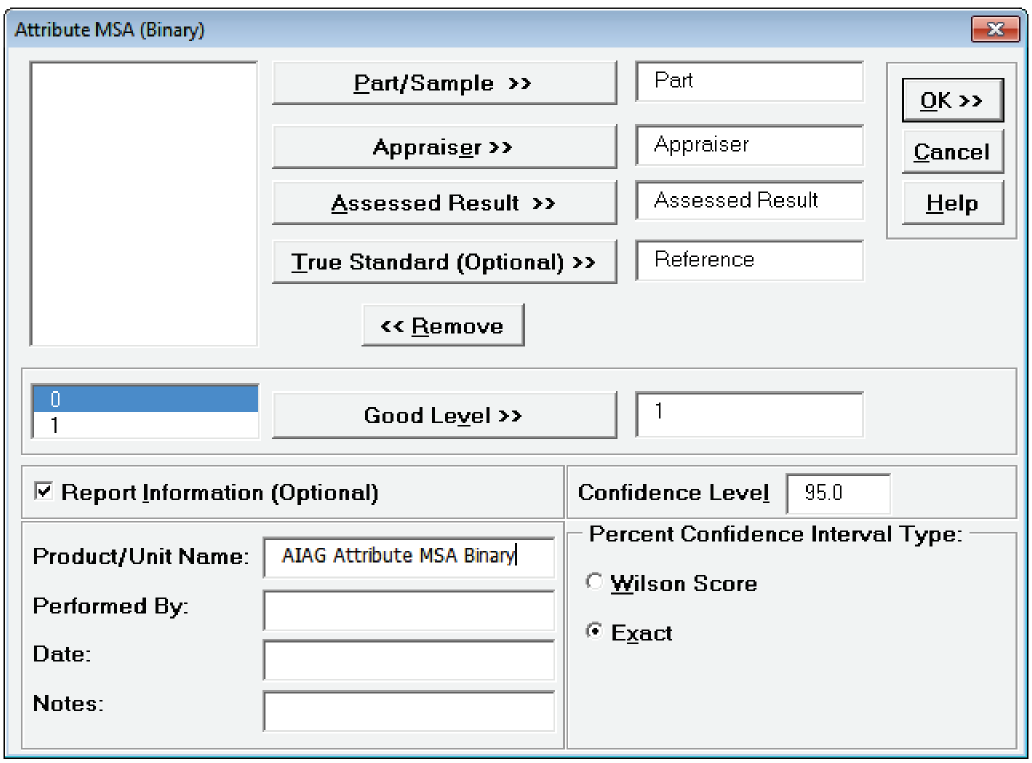
Attribute MSA (Binary)

Attribute MSA is also known as Attribute Agreement Analysis. Use the Binary option if the assessed result is text or numeric binary (e.g., 0/1, Pass/Fail, Good/Bad, G/NG, Yes/No).

1. Open the file **Attribute MSA – AIAG.xlsx**. This is an example from the Automotive Industry Action Group (AIAG) MSA Reference Manual, 3rd edition, page 127 (4th Edition, page 134). There are 50 samples, 3 appraisers and 3 trials with a 0/1 response. A "good" sample is denoted as a 1. A "bad" sample is denoted as a 0. Note that the worksheet data must be in stacked column format and the known reference values must be consistent for each sample.

2. Click **SigmaXL > Measurement Systems Analysis > Attribute MSA (Binary)**. Ensure that the entire data table is selected. Click **Next.**

3. Select *Part*, *Appraiser*, *Assessed Result* and *Reference* as shown. Check **Report Information** and enter *AIAG Attribute MSA Binary* for **Product/Unit Name**. Select **Percent Confidence Interval Type – Exact**. The default **Good Level** of "1" will be used as specified in the AIAG manual:



Tip: The **Good Level** definition is used to determine Type I and Type II error rates. It is applicable only when a **True Standard** is selected.

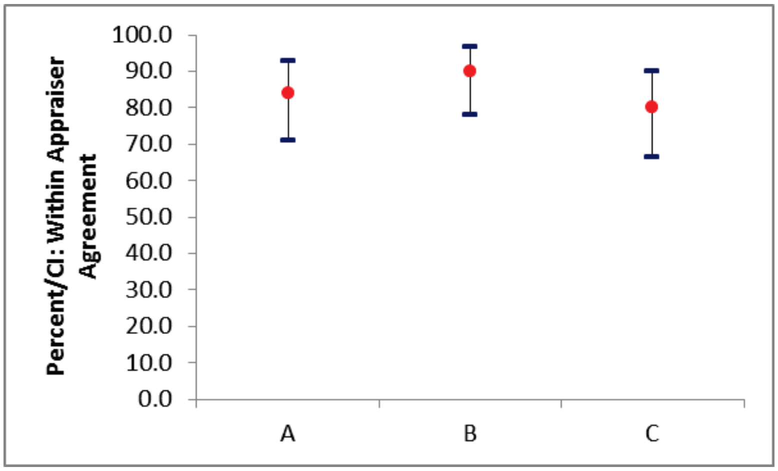
Tip: **Percent Confidence Interval Type** applies to the Percent Agreement and Percent Effectiveness Confidence Intervals. These are binomial proportions that have an "oscillation" phenomenon where the coverage probability varies with the sample size and proportion value. **Exact** is strictly conservative and will guarantee the specified confidence level as a minimum coverage probability, but results in wider intervals. **Wilson Score** has a mean coverage probability that matches the specified confidence interval. Since the intervals are narrower and thereby more powerful, Wilson Score is recommended for use in attribute MSA studies due to the small sample sizes typically used. **Exact** is selected in this example for continuity with the results from SigmaXL Version 6.

4. Click **OK**. The Attribute MSA Binary Analysis Report is produced. The tables and associated graphs are described separately by section for clarity.

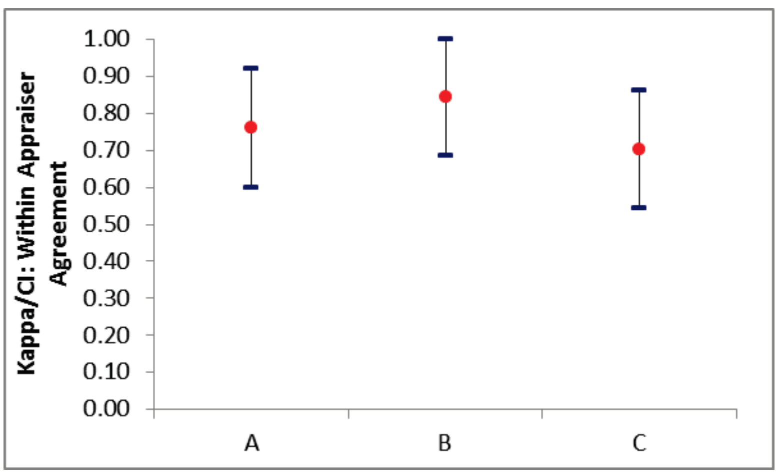
   Tip: While this report is quite extensive, a quick assessment of the attribute measurement system can be made by viewing the Kappa color highlights: **Green** - very good agreement (Kappa >= 0.9); **Yellow** - marginally acceptable, improvement should be considered (Kappa 0.7 to < 0.9); **Red** - unacceptable (Kappa < 0.7). Further details on Kappa are given below.

Within Appraiser Agreement is an assessment of each appraiser's consistency of ratings across trials and requires at least two trials. This is analogous to Gage R&R Repeatability. Note that the reference standard is not considered, so an appraiser may be perfectly consistent but consistently wrong!

Percent/CI: Within Appraiser Agreement Graph:



Kappa/CI: Within Appraiser Agreement Graph:



Within Appraiser Agreement Table:

Attribute Agreement Report:

Within Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
A	50	42	84.00	70.89	92.83	0.7600	0.0000	0.6000	0.9200
B	50	45	90.00	78.19	96.67	0.8451	0.0000	0.6850	1.0000
C	50	40	80.00	66.28	89.97	0.7029	0.0000	0.5429	0.8629

Within Appraiser Percent Agreement will decrease as the number of trials increase because a match occurs only if an appraiser is consistent across all trials. Use the **Kappa/CI: Within Appraiser Agreement Graph** to determine adequacy of the Within Appraiser agreement. See below for additional interpretation guidelines.

Tip: Hover the mouse pointer over the heading cells to view the following report comments.

 # Inspected: Number of parts or samples.

 # Matched: A match occurs only if an appraiser is consistent across all trials.

Percent Agreement = (**# Matched** / **# Inspected**) * 100

LC = **Percent** Lower Confidence Limit. **UC** = **Percent** Upper Confidence Limit. Confidence intervals (CI) for binomial proportions have an "oscillation" phenomenon where the coverage probability varies with n and p. **Exact** is strictly conservative and will guarantee the specified confidence level as a minimum coverage probability, but results in wide intervals. Wilson **Score** has a mean coverage probability that matches the specified confidence interval. Since the intervals are narrower and thereby more powerful, they are recommended for use in attribute MSA studies due to the small sample sizes typically used. See Appendix **Percent Confidence Intervals (Exact Versus Wilson Score)** for references.

Fleiss' Kappa statistic is a measure of agreement that is analogous to a "correlation coefficient" for discrete data. Kappa ranges from -1 to +1: A Kappa value of +1 indicates perfect agreement. If Kappa = 0, then agreement is the same as would be expected by chance. If Kappa = -1, then there is perfect disagreement**. "Rule-of-thumb" interpretation guidelines: >= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable, improvement should be considered (yellow); < 0.7 unacceptable (red)**. See Appendix **Kappa** for further details on the Kappa calculations and "rule-of-thumb" interpretation guidelines.

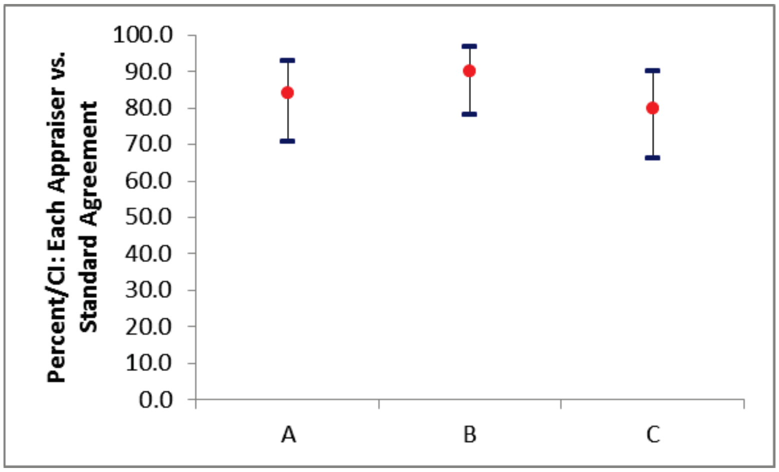
Fleiss' Kappa P-Value: H0: Kappa = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

Fleiss' Kappa LC (Lower Confidence) and **Fleiss' Kappa UC** (Upper Confidence) limits use a kappa normal approximation. Interpretation Guidelines: Kappa lower confidence limit >= 0.9: very good agreement. Kappa upper confidence limit < 0.7: the attribute agreement is unacceptable. Wide confidence intervals indicate that the sample size is inadequate.

In this example, we have marginal **Within Appraiser Agreement** for each of the appraisers.

Each Appraiser vs. Standard Agreement is an assessment of each appraiser's ratings across trials compared to a known reference standard. This is analogous to Gage R&R Accuracy.

Percent/CI: Each Appraiser vs. Standard Agreement Graph:



Each Appraiser vs. Standard Agreement Table:

Each Appraiser vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
A	50	42	84.00	70.89	92.83	0.8802	0.0000	0.7202	1.0000
B	50	45	90.00	78.19	96.67	0.9226	0.0000	0.7626	1.0000
C	50	40	80.00	66.28	89.97	0.7747	0.0000	0.6147	0.9347

Tip: The **Percent/CI Each Appraiser vs. Standard Agreement Graph** can be used to compare agreement to standard across the appraisers, but should not be used as an absolute measure of agreement. **Each Appraiser vs. Standard Agreement** will decrease as the number of trials increase because a match occurs only if an appraiser agrees with the standard consistently across all trials. Use **Fleiss' Kappa** in the **Each Appraiser vs. Standard Agreement Table** to determine the adequacy of Each Appraiser versus Standard agreement.

 # Inspected: Number of parts or samples.

 # Matched: A match occurs only if an appraiser agrees with the standard consistently across all trials.

Percent Agreement = (**# Matched** / **# Inspected**) * 100

Kappa is interpreted as above: **>= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable, improvement should be considered (yellow); < 0.7 unacceptable (red)**.

Appraisers A and C have marginal agreement versus the standard values. Appraiser B has very good agreement to the standard.

Each Appraiser vs. Standard Disagreement is a breakdown of each appraiser's rating misclassifications (compared to a known reference standard). This table is applicable only to binary two-level responses (e.g., 0/1, G/NG, Pass/Fail, True/False, Yes/No).

Each Appraiser vs. Standard Disagreement	Type I Error	Type I Error %	Type II Error	Type II Error %	Mixed Error	Mixed Error %
A	0	0.00	0	0.00	8	16.00
B	0	0.00	0	0.00	5	10.00
C	0	0.00	0	0.00	10	20.00

Disagreement Legend		
Error Type	Assessment	True Standard
Type I:	0	1
Type II:	1	0
Mixed:	Assessments across trials are not identical	

A **Type I Error** occurs when the appraiser consistently assesses a good part/sample as bad. "Good" is defined by the user in the Attribute MSA analysis dialog.

Type I Error % = (Type I Error / # Good Parts or Samples) * 100

A **Type II Error** occurs when a bad part/sample is consistently assessed as good.

Type II Error % = (Type II Error / # Bad Parts or Samples) * 100

A **Mixed Error** occurs when the assessments across trials are not identical.

Mixed Error % = (Mixed Error / # Parts or Samples) * 100

Between Appraiser Agreement is an assessment of the appraisers' consistency of ratings across trials and between each other. At least two appraisers are required. This is analogous to Gage R&R Reproducibility. Note that the reference standard is not considered, so the appraisers may be perfectly consistent, but consistently wrong!

Between Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
	50	39	78.00	64.04	88.47	0.7936	0.0000	0.7474	0.8398

All Appraisers vs. Standard Agreement is an assessment of all appraisers' ratings across trials compared to a known reference standard. This is analogous to Gage R&R Accuracy.

All Appraisers vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
	50	39	78.00	64.04	88.47	0.8592	0.0000	0.7668	0.9516

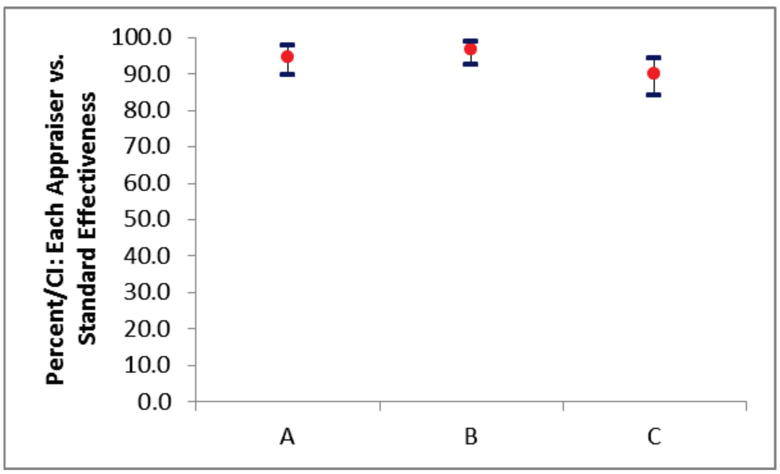
Kappa is interpreted as above: **>= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable (yellow); < 0.7 unacceptable (red)**.

Since the **Between Appraiser Agreement** and **All Appraisers vs. Standard Agreement** are marginally acceptable, improvements to the attribute measurement should be considered. Look for unclear or confusing operational definitions, inadequate training, operator distractions or poor lighting. Consider the use of pictures to clearly define a defect.

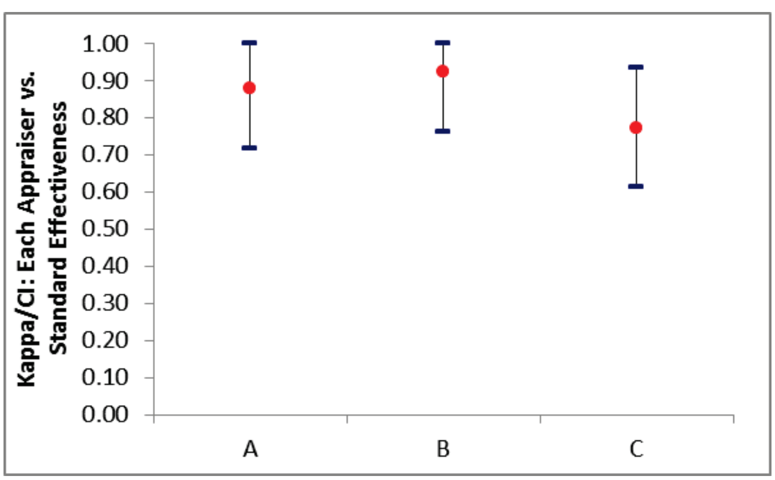
The **Attribute Effectiveness Report** is similar to the **Attribute Agreement Report**, but treats each trial as an opportunity. Consistency across trials or appraisers is not considered. This has the benefit of providing a Percent measure that is unaffected by the number of trials or appraisers. Also, the increased sample size for # Inspected results in a reduction of the width of the Percent confidence interval. The Misclassification report shows all errors classified as Type I or Type II. Mixed errors are not relevant here. This report requires a known reference standard and includes: **Each Appraiser vs. Standard Effectiveness**, **All Appraisers vs. Standard Effectiveness**, and **Effectiveness and Misclassification Summary**.

Each Appraiser vs. Standard Effectiveness is an assessment of each appraiser's ratings compared to a known reference standard. This is analogous to Gage R&R Accuracy. Unlike the **Each Appraiser vs. Standard Agreement** table above, consistency across trials is not considered here - each trial is considered as an opportunity. This has the benefit of providing a Percent measure that is unaffected by the number of trials. Also, the increased sample size for # Inspected results in a reduction of the width of the Percent confidence interval.

Percent/CI: Each Appraiser vs. Standard Effectiveness Graph:



Kappa/CI: Each Appraiser vs. Standard Effectiveness Graph:



Each Appraiser vs. Standard Effectiveness Table:

Attribute Effectiveness Report:

Each Appraiser vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
A	150	142	94.67	89.76	97.67	0.8788	0.0000	0.7187	1.0000
B	150	145	96.67	92.39	98.91	0.9230	0.0000	0.7629	1.0000
C	150	135	90.00	84.04	94.29	0.7739	0.0000	0.6138	0.9339

 # Inspected = # Parts or Samples * # Trials

 # Matched: A match occurs if an appraiser agrees with the standard (consistency across trials is not considered here).

Percent Effectiveness = (# Matched / # Inspected) * 100. Interpretation Guidelines for Percent Effectiveness: => 95% very good; 85% to <95% marginal, may be acceptable but improvement should be considered; < 85% unacceptable. These guidelines assume an equal number of known good and known bad parts/samples.

Kappa is interpreted as above: **>= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable (yellow); < 0.7 unacceptable (red)**.

Tip: The Kappa values in the Effectiveness tables are very similar to those in the Agreement tables (the slight difference is due to average Kappa for unstacked versus Kappa for stacked data). This is why the **Kappa/CI Each Appraiser vs. Standard Agreement** graph is not shown. It would essentially be a duplicate of the **Kappa/CI Each Appraiser vs. Standard Effectiveness** graph.

Appraisers A and C have marginal agreement versus the standard values, with less than 95% Effectiveness and Kappa < 0.9. Appraiser B has very good agreement to the standard.

All Appraisers vs. Standard Effectiveness is an assessment of all appraisers' ratings compared to a known reference standard. This is analogous to Gage R&R Accuracy. Unlike the **All Appraiser vs. Standard Agreement** table above, consistency across trials and appraisers is not considered here - each trial is considered as an opportunity. This has the benefit of providing a Percent measure that is unaffected by the number of trials or appraisers. Also, the increased sample size for # Inspected results in a reduction of the width of the Percent confidence interval.

All Appraisers vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Exact)	95.0% UC (Exact)	Fleiss' Kappa	Fleiss' Kappa P-Value	Fleiss' Kappa 95.0% LC	Fleiss' Kappa 95.0% UC
	450	422	93.78	91.13	95.83	0.8581	0.0000	0.7657	0.9505

 # Inspected = # Parts or Samples * # Trials * # Appraisers

 # Matched: A match occurs if an appraiser agrees with the standard (consistency across trials and appraisers is not considered here).

Percent Effectiveness = (**# Matched** / **# Inspected**) * 100.

The interpretation guidelines for Kappa and Percent Effectiveness are the same as noted above. This measurement system is marginally acceptable.

Each Appraiser vs. Standard Misclassification is a breakdown of each appraiser's rating misclassifications (compared to a known reference standard). This table is applicable only to binary two-level responses (e.g., 0/1, G/NG, Pass/Fail, True/False, Yes/No). Unlike the **Each Appraiser vs. Standard Disagreement** table above, consistency across trials is not considered here. All errors are classified as Type I or Type II. Mixed errors are not relevant.

Each Appraiser vs. Standard Misclassification	Type I Error	# Inspected (1)	Type I Error %	Type II Error	# Inspected (0)	Type II Error %
A	5	102	4.90	3	48	6.25
B	2	102	1.96	3	48	6.25
C	9	102	8.82	6	48	12.50

Misclassification Legend		
Error Type	Assessment	True Standard
Type I:	0	1
Type II:	1	0

A **Type I** error occurs when the appraiser assesses a good part/sample as bad (consistency across trials is not considered here). "Good" is defined by the user in the Attribute MSA analysis dialog. See **Misclassification Legend** for specific definition of **Type I** and **Type II** Errors.

 # Inspected = # Good Parts or Samples * # Trials

Type I Error % = (Type I Error / # Inspected Good) * 100

A **Type II** error occurs when a bad part/sample is assessed as good. See **Misclassification Legend** for specific definition of **Type I** and **Type II** Errors.

 # Inspected = # Bad Parts or Samples * # Trials

Type II Error % = (Type II Error / # Inspected Bad) * 100

All Appraisers vs. Standard Misclassification is a breakdown of all appraisers' rating misclassifications (compared to a known reference standard). This table is applicable only to binary two-level responses (e.g., 0/1, G/NG, Pass/Fail, True/False, Yes/No). Unlike the **All Appraisers vs. Standard Disagreement table** above, consistency across trials and appraisers is not considered here. All errors are classified as Type I or Type II. Mixed errors are not relevant.

All Appraisers vs. Standard Misclassification	Type I Error	# Inspected (1)	Type I Error %	Type II Error	# Inspected (0)	Type II Error %
	16	306	5.23	12	144	8.33

A **Type I** error occurs when the appraiser assesses a good part/sample as bad (consistency across trials is not considered here). "Good" is defined by the user in the Attribute MSA analysis dialog. See **Misclassification Legend** for specific definition of **Type I** and **Type II** Errors.

 # Inspected = # Good Parts or Samples * # Trials * # Appraisers

Type I Error % = (Type I Error / # Inspected Good) * 100

A **Type II** error occurs when a bad part/sample is assessed as good. See **Misclassification Legend** for specific definition of **Type I** and **Type II** Errors.

 # Inspected = # Bad Parts or Samples * # Trials * # Appraisers

Type II Error % = (Type II Error / # Inspected Bad) * 100

Effectiveness and Misclassification Summary is a summary table of all appraisers' correct rating counts and misclassification counts compared to the known reference standard values.

Effectiveness and Misclassification Summary	Standard (0)	Standard (1)
Appraiser (0)	132	16
Appraiser (1)	12	290

Attribute MSA Data is a summary showing the original data in unstacked format. This makes it easy to visually compare appraiser results by part. If a reference standard is provided, the cells are color highlighted as follows: agrees with reference (green); Type I error (yellow); Type II error (red):

Attribute MSA Data:

Part	Reference	A 1	A 2	A 3	B 1	B 2	B 3	C 1	C 2	C 3
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	1	1	1	0	1	1	0	1	0	0
7	1	1	1	1	1	1	1	1	0	1
8	1	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	0	0	0
10	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	1
13	1	1	1	1	1	1	1	1	1	1
14	1	1	1	0	1	1	1	1	0	0
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1
21	1	1	1	0	1	0	1	0	1	0
22	0	0	0	1	0	1	0	1	1	0
23	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	0	0	0	0	0	0	0	0	0	0
26	0	0	1	0	0	0	0	0	0	1
27	1	1	1	1	1	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1
29	1	1	1	1	1	1	1	1	1	1
30	0	0	0	0	0	0	1	0	0	0
31	1	1	1	1	1	1	1	1	1	1
32	1	1	1	1	1	1	1	1	1	1
33	1	1	1	1	1	1	1	1	1	1
34	0	0	0	1	0	0	1	0	1	1
35	1	1	1	1	1	1	1	1	1	1
36	1	1	1	0	1	1	1	1	0	1
37	0	0	0	0	0	0	0	0	0	0
38	1	1	1	1	1	1	1	1	1	1
39	0	0	0	0	0	0	0	0	0	0
40	1	1	1	1	1	1	1	1	1	1
41	1	1	1	1	1	1	1	1	1	1
42	0	0	0	0	0	0	0	0	0	0
43	1	1	0	1	1	1	1	1	1	0
44	1	1	1	1	1	1	1	1	1	1
45	0	0	0	0	0	0	0	0	0	0
46	1	1	1	1	1	1	1	1	1	1
47	1	1	1	1	1	1	1	1	1	1
48	0	0	0	0	0	0	0	0	0	0
49	1	1	1	1	1	1	1	1	1	1
50	0	0	0	0	0	0	0	0	0	0

In conclusion, with the Kappa scores in the "yellow zone" (< 0.9) and Percent Effectiveness less than 95% this measurement system is marginal and should be improved. Appraiser B is the exception and does well against the standard. Look for unclear or confusing operational definitions, inadequate training, operator distractions or poor lighting. Consider the use of pictures to clearly define a defect. Use Attribute MSA as a way to "put your stake in the ground" and track the effectiveness of improvements to the measurement system.

Attribute MSA (Ordinal)

Attribute MSA is also known as Attribute Agreement Analysis. Use the Ordinal option if the assessed result is numeric ordinal (e.g., 1, 2, 3, 4, 5). There must be at least 3 response levels in the assessed result, otherwise it is binary. Examples of ordinal responses used elsewhere in this workbook include:

- *Customer Loyalty –Likely to Recommend* score which contains ordinal integer values from 1 to 5, where a 1 indicates that the customer is very unlikely to recommend and a 5 indicates that the customer is very likely to recommend. This example is used in **Ordinal Logistic Regression**.

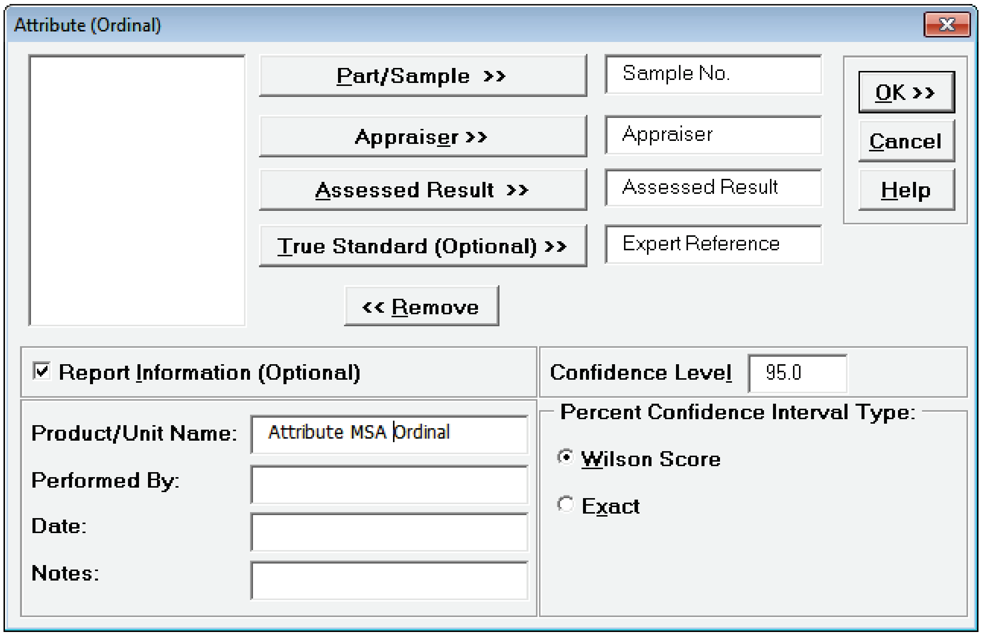
- *Taste Score* on a scale of 1-7 where 1 is "awful" and 7 is "delicious." This is used in the cake bake taste test Design of Experiments: **Part B – Three Factor Full Factorial Example Using DOE Template**.

An Ordinal Attribute MSA study should be done prior to formal ordinal data collection for use in hypothesis testing, regression or design of experiments.

1. Open the file **Attribute MSA – Ordinal.xlsx**. This is an Ordinal MSA example with 50 samples, 3 appraisers and 3 trials. The response is 1 to 5, grading product quality. One denotes “Very Poor Quality,” 2 is “Poor,” 3 is “Fair,” 4 is “Good” and a 5 is “Very Good Quality.” The *Expert Reference* column is the reference standard from an expert appraisal. Note that the worksheet data must be in stacked column format and the reference values must be consistent for each sample.

2. Click **SigmaXL > Measurement Systems Analysis > Attribute MSA (Ordinal)**. Ensure that the entire data table is selected. Click **Next.**

3. Select *Sample No.*, *Appraiser*, *Assessed Result* and *Expert Reference* as shown. Check **Report Information** and enter *Attribute MSA Ordinal* for **Product/Unit Name**. Select **Percent Confidence Interval Type – Wilson Score**:

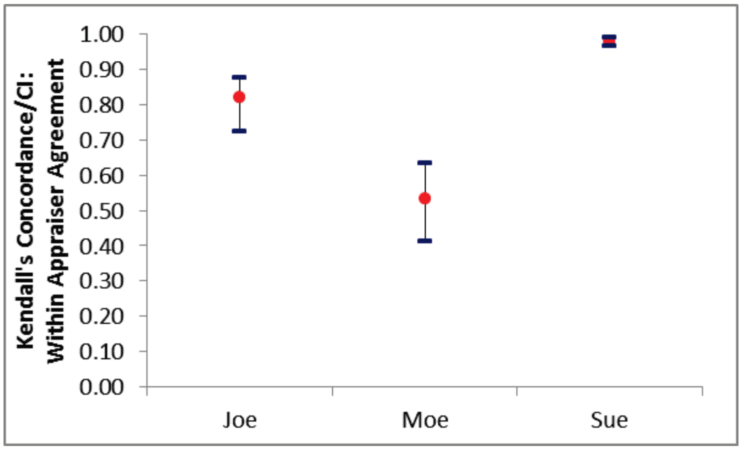


4. Click **OK**. The Attribute MSA Ordinal Analysis Report is produced.

   Tip: While this report is quite extensive, a quick assessment of the attribute measurement system can be made by viewing the **Kendall Concordance** and **Kendall Correlation** color highlights: **Green** - very good agreement; **Yellow** - marginally acceptable, improvement should be considered; **Red** - unacceptable. Further details on the Kendall Coefficients are given below.

   Tip: **Fleiss' Kappa** and **Percent Agreement** are included in the report for completeness but not recommended for use with Ordinal response data because they treat each response level as nominal. Kendall's Concordance and Correlation take the order of the data into account, so a deviation of 1 is not as bad as a deviation of 2 or more. See **Attribute MSA – Nominal** for a discussion of the **Fleiss' Kappa** report.

Kendall’s Concordance/CI Within Appraiser Agreement Graph:



Within Appraiser Agreement Table:

Attribute Agreement Report:

Within Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Coeff. Concordance	Kendall's Concordance P-Value	Kendall's Concordance 95.0% LC	Kendall's Concordance 95.0% UC
Joe	50	14	28.00	17.47	41.67	0.8204	0.0000	0.7234	0.8772
Moe	50	2	4.00	1.10	13.46	0.5332	0.0048	0.4112	0.6351
Sue	50	40	80.00	66.96	88.76	0.9784	0.0000	0.9670	0.9908

Between Appraiser Agreement Table:

Between Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Coeff. Concordance	Kendall's Concordance P-Value	Kendall's Concordance 95.0% LC	Kendall's Concordance 95.0% UC
	50	1	2.00	0.35	10.50	0.6495	0.0000	0.5649	0.7214

Kendall's Coefficient of Concordance (Kendall's W) is a measure of association for discrete ordinal data, used for assessments that do not include a known reference standard. Kendall’s coefficient of concordance ranges from 0 to 1: A coefficient value of 1 indicates perfect agreement. If the coefficient = 0, then the agreement is random, i.e., the same as would be expected by chance. **“Rule-of-thumb” interpretation guidelines: >= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable, improvement should be considered (yellow); < 0.7 unacceptable (red)**. See Appendix Kendall’s Coefficient of Concordance for further details on the Kendall Concordance calculations and “rule-of-thumb” interpretation guidelines.

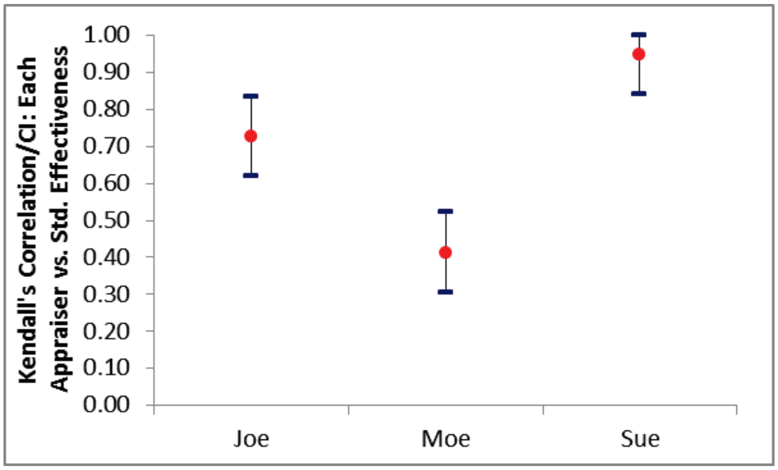
Kendall's Concordance P-Value: H0: Kendall's Coefficient of Concordance = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

Kendall's Concordance LC (Lower Confidence) limit and **Kendall's Concordance UC** (Upper Confidence) limit cannot be solved analytically, so are estimated using bootstrapping. See Appendix Kendall’s Coefficient of Concordance for further details on the bootstrap confidence intervals. Interpretation Guidelines: Concordance lower confidence limit >= 0.9: very good agreement. Concordance upper confidence limit < 0.7: the attribute agreement is unacceptable. Wide confidence intervals indicate that the sample size is inadequate.

The **Within Appraiser Agreement** for Joe is marginal, Moe is unacceptable and Sue is very good.

The **Between Appraiser Agreement** is unacceptable.

Kendall's Correlation/CI Each Appraiser vs. Standard Effectiveness Graph:



Each Appraiser vs. Standard Agreement Table:

Each Appraiser vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Correlation Coeff.	Kendall's Correlation P-Value	Kendall's Correlation 95.0% LC	Kendall's Correlation 95.0% UC
Joe	50	11	22.00	12.75	35.24	0.7323	0.0000	0.6218	0.8427
Moe	50	2	4.00	1.10	13.46	0.4161	0.0000	0.3057	0.5265
Sue	50	40	80.00	66.96	88.76	0.9495	0.0000	0.8391	1.0000

All Appraisers vs. Standard Agreement Table:

All Appraisers vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Correlation Coeff.	Kendall's Correlation P-Value	Kendall's Correlation 95.0% LC	Kendall's Correlation 95.0% UC
	50	1	2.00	0.35	10.50	0.6993	0.0000	0.6355	0.7631

Each Appraiser vs. Standard Effectiveness Table:

Each Appraiser vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Correlation Coeff.	Kendall's Correlation P-Value	Kendall's Correlation 95.0% LC	Kendall's Correlation 95.0% UC
Joe	150	81	54.00	46.02	61.78	0.7280	0.0000	0.6201	0.8359
Moe	150	47	31.33	24.45	39.14	0.4137	0.0000	0.3058	0.5217
Sue	150	134	89.33	83.38	93.33	0.9478	0.0000	0.8399	1.0000

All Appraisers vs. Standard Effectiveness Table:

All Appraisers vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Kendall's Correlation Coeff.	Kendall's Correlation P-Value	Kendall's Correlation 95.0% LC	Kendall's Correlation 95.0% UC
	450	262	58.22	53.61	62.69	0.6830	0.0000	0.6211	0.7448

Kendall's Correlation Coefficient (Kendall's tau-b) is a measure of association for discrete ordinal data, used for assessments that include a known reference standard. Kendall's correlation coefficient ranges from -1 to 1: A coefficient value of 1 indicates perfect agreement. If the coefficient = 0, then the agreement is random, i.e., the same as would be expected by chance. A coefficient value of -1 indicates perfect disagreement. "**Rule-of-thumb**" **interpretation guidelines: >= 0.8 very good agreement (green); 0.6 to < 0.8 marginally acceptable, improvement should be considered (yellow); < 0.6 unacceptable (red)**. See Appendix Kendall's Correlation Coefficient for further details on the Kendall Correlation calculations and "rule-of-thumb" interpretation guidelines.

Kendall's Correlation P-Value: H0: Kendall's Correlation Coefficient = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

Kendall's Correlation LC (Lower Confidence) and **Kendall's Correlation UC** (Upper Confidence) limit use a normal approximation. Interpretation Guidelines: Correlation lower confidence limit >= 0.8: very good agreement. Correlation upper confidence limit < 0.6: the attribute agreement is unacceptable. Wide confidence intervals indicate that the sample size is inadequate.

Tip: Kendall's Correlation values in the Effectiveness tables are very similar to those in the Agreement tables (the slight difference is due to average Kendall for unstacked versus Kendall for stacked data). This is why the **Kendall's Correlation/CI Each Appraiser vs. Standard Agreement** graph is not shown. It would essentially be a duplicate of the **Kendall's Correlation/CI Each Appraiser vs. Standard Effectiveness** graph.

Appraiser Joe has marginal agreement versus the standard values. Appraiser Moe has unacceptable agreement to the standard. Sue has very good agreement to the standard.

Overall, the appraisers have marginal agreement to the standard.

Note that the Percent Agreement results in **All Appraisers vs. Standard Agreement Table** show only 2% agreement! This is due to the requirement that all appraisers agree with the standard across all trials for a 5-level response, which is very unlikely to occur. This highlights the problem with using Percent Agreement in an Ordinal MSA. Kendall's coefficients are the key metric to assess an Ordinal MSA.

Effectiveness and Misclassification Summary is a summary table of all appraisers' correct rating counts and misclassification counts compared to the known reference standard values.

Effectiveness and Misclassification Summary	Standard (1)	Standard (2)	Standard (3)	Standard (4)	Standard (5)
Appraiser Response (1)	102	20	3	2	0
Appraiser Response (2)	38	37	3	6	5
Appraiser Response (3)	8	14	19	20	6
Appraiser Response (4)	12	7	6	26	19
Appraiser Response (5)	2	3	5	9	78

Attribute MSA Data is a summary showing the original data in unstacked format. This makes it easy to visually compare appraiser results by part. If a reference standard is provided, the cells are color highlighted as follows: absolute deviation = 0 (green); absolute deviation = 1 (yellow); absolute deviation >=2 (red):

Attribute MSA Data:

Sample No.	Expert Reference	Joe 1	Joe 2	Joe 3	Moe 1	Moe 2	Moe 3	Sue 1	Sue 2	Sue 3
1	4	5	4	5	4	4	3	3	4	4
2	2	3	3	2	3	3	2	2	2	2
3	5	5	5	5	5	4	4	5	5	5
4	4	3	3	3	3	3	4	3	3	4
5	3	4	3	3	2	1	5	3	3	3
6	1	2	1	2	2	4	2	1	1	1
7	1	1	1	1	2	1	1	1	1	1
8	1	1	2	1	2	1	1	1	1	1
9	5	4	5	4	5	5	5	5	5	5
10	1	1	2	3	1	4	1	1	1	1

In conclusion, this measurement system is marginal and should be improved. Appraiser Moe needs training and Appraiser Joe needs a refresher. Sue has very good agreement based on Kendall's Concordance and Correlation, but would have been considered marginal based on Kappa (< .9) and Percent Effectiveness (< 95%). As discussed above, Kappa and Percent Effectiveness do not take the order of the response data into account, so are not as useful as Kendall's coefficients in an Ordinal MSA study.

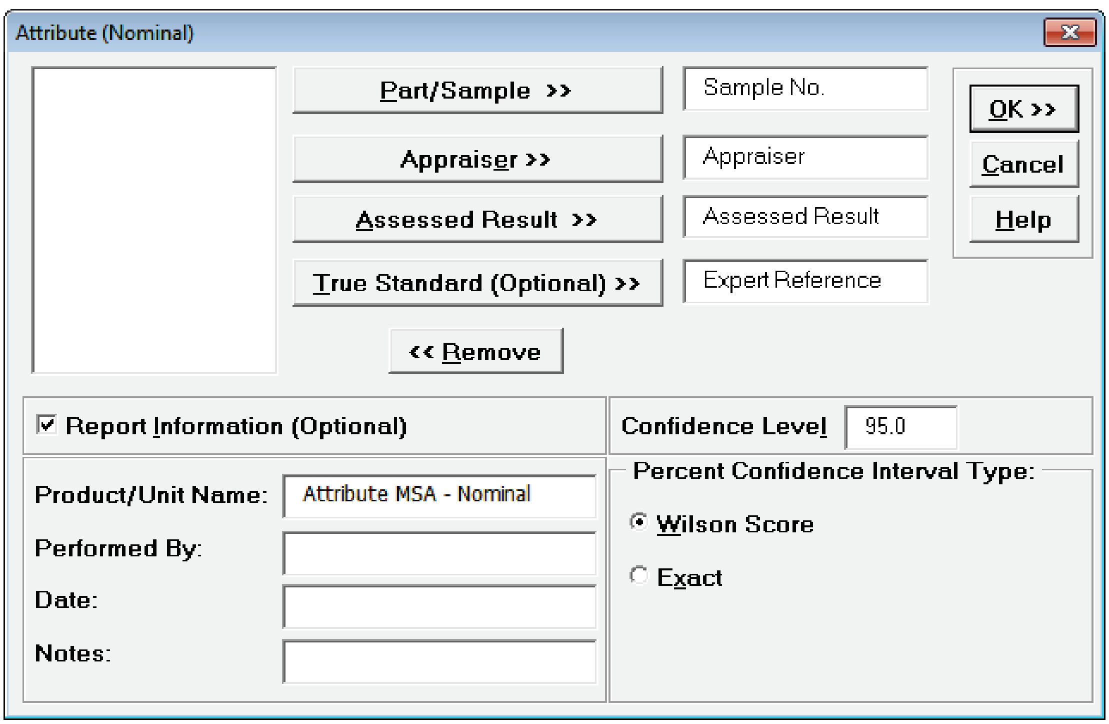
Attribute MSA (Nominal)

Attribute MSA is also known as Attribute Agreement Analysis. Use the Nominal option if the assessed result is numeric or text nominal (e.g., Defect Type 1, Defect Type 2, Defect Type 3). There must be at least 3 response levels in the assessed result, otherwise it is binary.

1. Open the file **Attribute MSA – Nominal.xlsx**. This is a Nominal MSA example with 30 samples, 3 appraisers and 2 trials. The response is text "Type_1", "Type_2" and "Type_3." The *Expert Reference* column is the reference standard from an expert appraisal. Note that the worksheet data must be in stacked column format and the reference values must be consistent for each sample.

2. Click **SigmaXL > Measurement Systems Analysis > Attribute MSA (Nominal)**. Ensure that the entire data table is selected. Click **Next.**

3. Select *Sample No.*, *Appraiser*, *Assessed Result* and *Expert Reference* as shown. Check **Report Information** and enter *Attribute MSA – Nominal* for **Product/Unit Name**. Select **Percent Confidence Interval Type – Wilson Score**:



4. Click **OK**. The Attribute MSA Nominal Analysis Report is produced.

Tip: While this report is quite extensive, a quick assessment of the attribute measurement system can be made by viewing the Kappa color highlights: **Green** - very good agreement (Kappa >= 0.9); **Yellow** - marginally acceptable, improvement should be considered (Kappa 0.7 to < 0.9); **Red** - unacceptable (Kappa < 0.7). See **Attribute MSA (Binary)** for a detailed discussion of the report graphs and tables. Here we will just look at Kappa.

Attribute Agreement Report:

Attribute Agreement Report:

Within Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
A	30	30	100.00	88.65	100.00	1.0000	0.0000	0.7333	1.0000	1.0000
B	30	29	96.67	83.33	99.41	0.9467	0.0000	0.6799	1.0000	1.0000
C	30	26	86.67	70.32	94.69	0.7849	0.0000	0.5184	1.0000	1.0000

Each Appraiser vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
A	30	29	96.67	83.33	99.41	0.9467	0.0000	0.7581	1.0000	1.0000
B	30	29	96.67	83.33	99.41	0.9733	0.0000	0.7847	1.0000	1.0000
C	30	25	83.33	66.44	92.66	0.8391	0.0000	0.6506	1.0000	1.0000

Between Appraiser Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
	30	25	83.33	66.44	92.66	0.9110	0.0000	0.8421	0.9798	1.0000

All Appraisers vs. Standard Agreement	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
	30	25	83.33	66.44	92.66	0.9197	0.0000	0.8108	1.0000	1.0000

Attribute Effectiveness Report:

Attribute Effectiveness Report:

Each Appraiser vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
A	60	58	96.67	88.64	99.08	0.9467	0.0000	0.7581	1.0000	1.0000
B	60	59	98.33	91.14	99.71	0.9733	0.0000	0.7847	1.0000	1.0000
C	60	54	90.00	79.85	95.34	0.8399	0.0000	0.6513	1.0000	1.0000

All Appraisers vs. Standard Effectiveness	# Inspected	# Matched	Percent	95.0% LC (Score)	95.0% UC (Score)	Fleiss' Kappa Overall Response	Fleiss' Kappa Overall Response P-Value	Fleiss' Kappa Overall Response 95.0% LC	Fleiss' Kappa Overall Response 95.0% UC	Fleiss' Kappa Response (Type_1)
	180	171	95.00	90.77	97.35	0.9200	0.0000	0.8111	1.0000	1.0000

Effectiveness and Misclassification Summary	Standard (Type_1)	Standard (Type_2)	Standard (Type_3)
Appraiser Response (Type_1)	30	0	0
Appraiser Response (Type_2)	0	70	1
Appraiser Response (Type_3)	0	8	71

Fleiss' Kappa statistic is a measure of agreement that is analogous to a "correlation coefficient" for discrete data. Kappa ranges from -1 to +1: A Kappa value of +1 indicates perfect agreement. If Kappa = 0, then agreement is the same as would be expected by chance. If Kappa = -1, then there is perfect disagreement**. "Rule-of-thumb" interpretation guidelines: >= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable, improvement should be considered (yellow); < 0.7 unacceptable (red)**. See Appendix **Kappa** for further details on the Kappa calculations and "rule-of-thumb" interpretation guidelines.

Fleiss' Kappa P-Value: H0: Kappa = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

Fleiss' Kappa LC (Lower Confidence) and **Fleiss' Kappa UC** (Upper Confidence) limits use a kappa normal approximation. Interpretation Guidelines: Kappa lower confidence limit >= 0.9: very good agreement. Kappa upper confidence limit < 0.7: the attribute agreement is unacceptable. Wide confidence intervals indicate that the sample size is inadequate.

Fleiss' Kappa Overall is an overall Kappa for all of the response levels.

Fleiss' Kappa Individual gives Kappa for each response level (*Type_1*, *Type_2* and *Type_3*). This is useful to identify if an appraiser has difficulty assessing a particular defect type.

This is a very good attribute measurement system, but Appraiser C is marginal so a refresher would be helpful.

Part J – Process Capability

Process Capability Templates

Click **SigmaXL > Templates and Calculators > Basic Process Capability Templates > Process Sigma Level – Discrete Data** or
Process Sigma Level – Continuous Data or
Process Capability Indices or
Process Capability & Confidence Intervals or
Tolerance Interval Calculator (Normal Exact).

These templates are also located at **SigmaXL > Process Capability > Basic Process Capability Templates**.

See **Measure Phase Part B – Templates and Calculators** for Process Capability Template examples:

Basic Process Capability Templates – Process Sigma Level – Discrete Data

Basic Process Capability Templates – Process Sigma Level – Continuous Data

Basic Process Capability Templates – Process Capability Indices

Basic Process Capability Templates – Process Capability & Confidence Intervals

Basic Process Capability Templates – Tolerance Interval Calculator (Normal Exact)

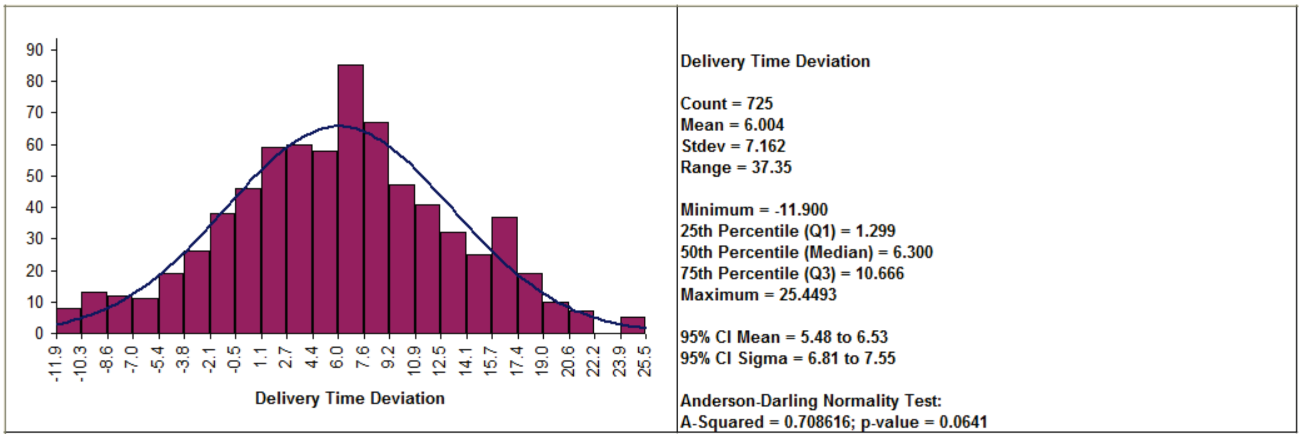
Histograms and Process Capability

1. Open the file **Delivery Times.xlsx**. This contains continuous data of hotel breakfast delivery times. Deviation Time is the deviation around targeted delivery time in minutes. The Critical Customer Requirements (CCR's) are as follows: USL = 10 minutes late, LSL = -10 minutes (early).

2. Let's begin with a view of the data using Histograms and Descriptive Statistics. Click **SigmaXL > Graphical Tools > Histograms & Descriptive Statistics**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

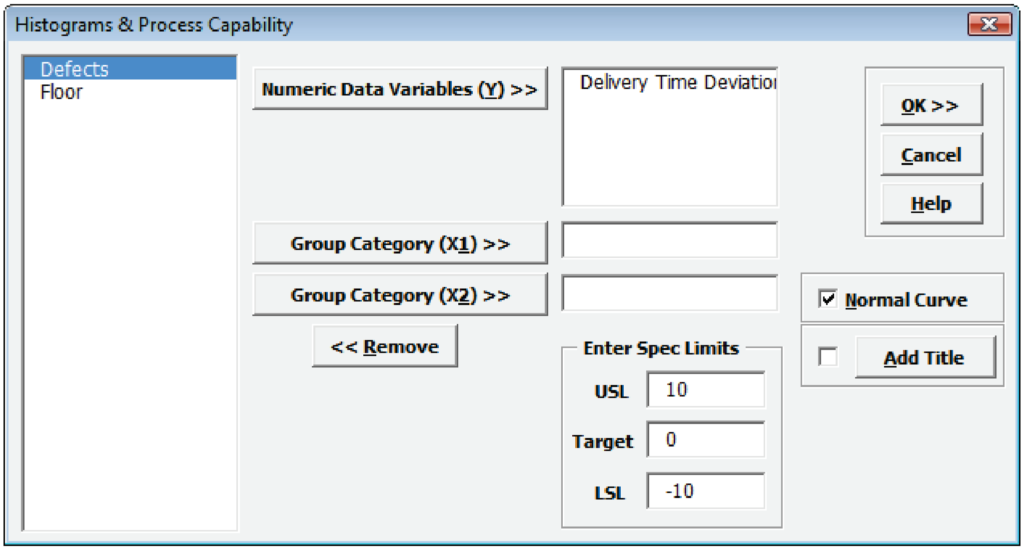
4. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**. Click **OK**.



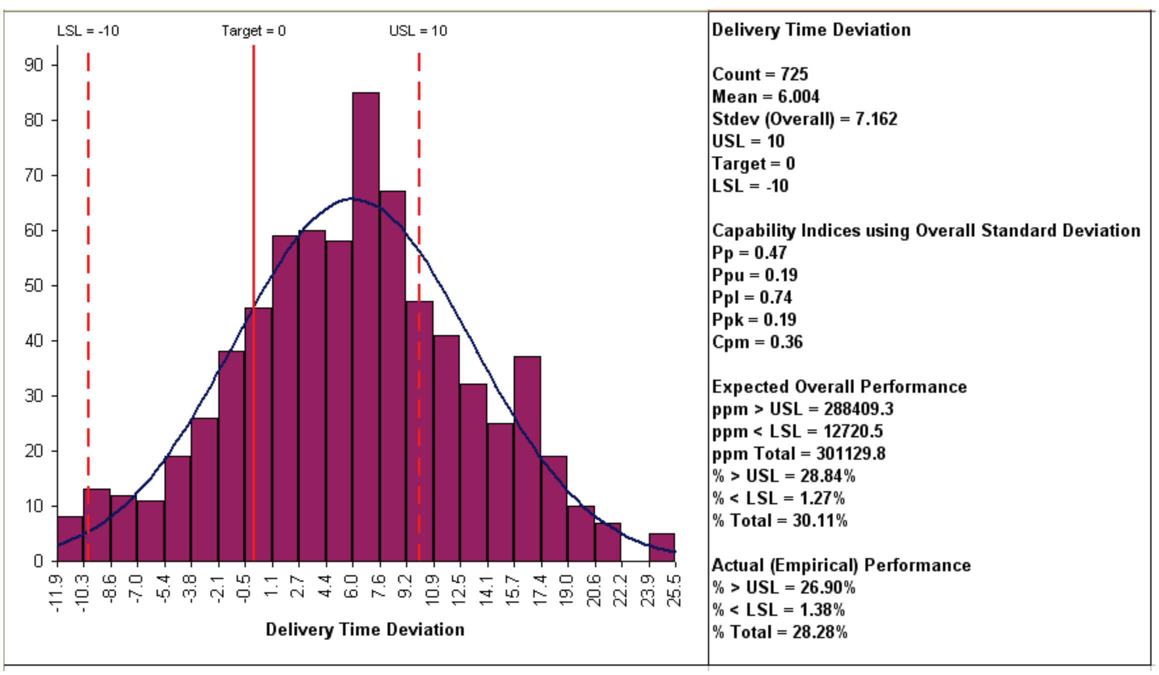
Note that the data has an average of 6, so on average the breakfast delivery is 6 minutes late. Also note the wide variation in delivery times (StDev = 7.2 minutes). With the AD P-Value > .05, this data can be assumed to have a normal distribution.

5. Click **Sheet 1** Tab. Click **SigmaXL > Process Capability > Histograms & Process Capability** (**or SigmaXL > Graphical Tools > Histograms & Process Capability**).

6. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**. Enter **USL** = 10, **Target** = 0, **LSL** = -10, check **Normal Curve** as shown below:



7. Click **OK**. Resulting Histogram & Process Capability report is shown below:



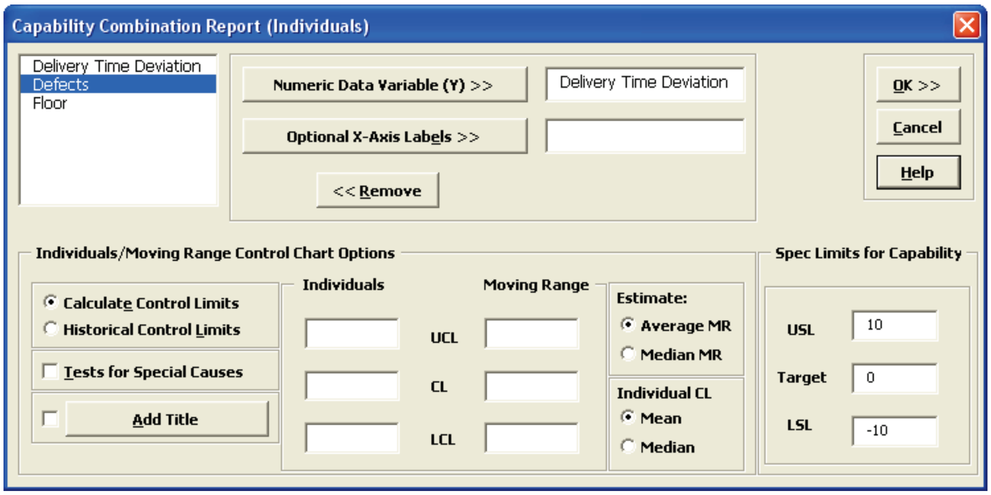
With the Process Performance indices Pp and Ppk < 1, this process is clearly in need of improvement. Note that the difference between Pp and Ppk is due to the off-center process mean. (Cp and Cpk indices are provided in the Capability Combination Report or optionally when creating Control Charts).

Capability Combination Report (Individuals)

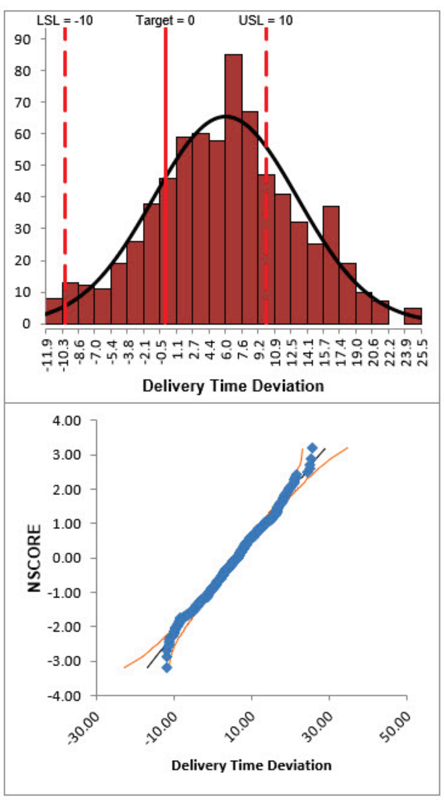
1. Open the file **Delivery Times.xlsx**.

2. Click **SigmaXL >Process Capability > Capability Combination Report (Individuals)**.

3. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**. Enter **USL** = 10, **Target** = 0, **LSL** = -10, as shown below:



4. Click **OK**. The resulting Process Capability Combination report is shown below:



Process Capability Report: Delivery Time Deviation	
Count	725
Mean	6.004
StDev (Overall, Long Term)	7.162
StDev (Within, Short Term)	7.202
USL	10
Target	0
LSL	-10

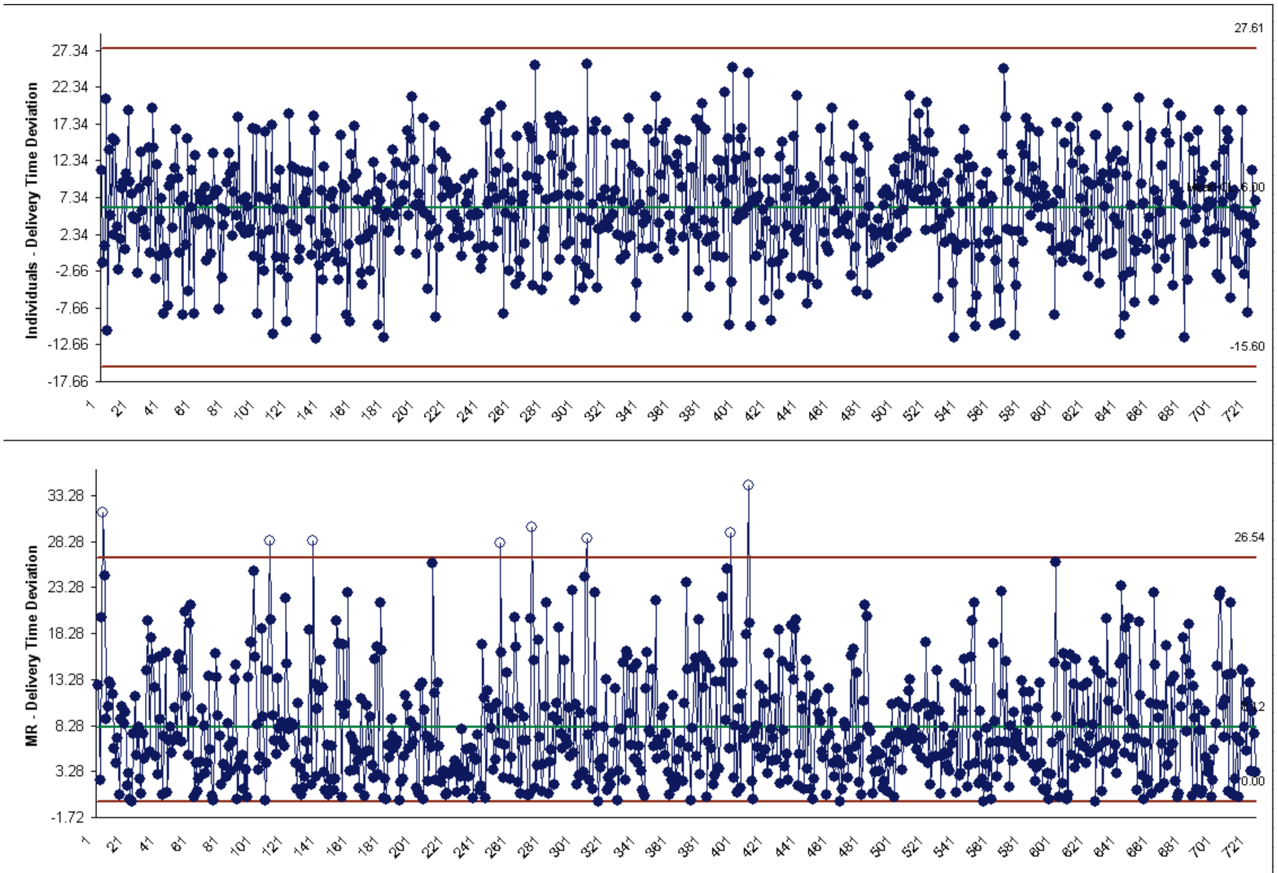
Capability Indices using Overall StDev	
Pp	0.47
Ppu	0.19
Ppl	0.74
Ppk	0.19
Cpm	0.36

Potential Capability Indices using Within StDev	
Cp	0.46
Cpu	0.18
Cpl	0.74
Cpk	0.18

Expected Overall Performance	
ppm > USL	288409.3
ppm < LSL	12720.5
ppm Total	301130
% > USL	28.84%
% < LSL	1.27%
% Total	30.11%

Actual (Empirical) Performance	
% > USL	26.90%
% < LSL	1.38%
% Total	28.28%

Anderson-Darling Normality Test	
A-Squared	0.708616
P-Value	0.0641



5. Cp is the "best case" potential capability index. Ppk is the "worst case" actual performance. If Cp is excellent (say Cp =2) and Ppk is poor (say Ppk < 1), this indicates that the process can be dramatically improved through centering and stabilizing. If Cp is poor, you will have to implement fundamental improvements using DMAIC to address common cause variation. If Ppk = Cp, this indicates that the process is centered and stable, and the short-term (within subgroup) variation is the same as the long-term (overall) variation.

6. Cpm is a similar statistic but it incorporates a penalty for deviation from target.

7. On a technical note regarding compatibility with other software tools, please be advised that SigmaXL does not use unbiasing constants when calculating Overall StDev, Pp, Ppk and Cpm. This is done in order to ensure that the calculated Overall StDev matches the results given in Descriptive Statistics. Note that Within StDev, Cp and Cpk are based on control chart techniques (MR-bar/d2 for Individuals Chart, R-bar/d2 for X-Bar & R, S-bar/c4 for X-Bar & S).

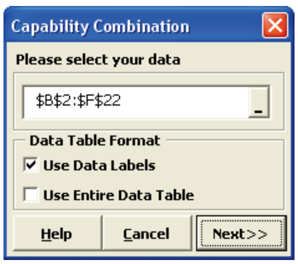
Capability Combination Report (Subgroups)

1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the catapult ball 3 times. The target distance is 100 inches. The Upper Specification Limit (USL) is 108 inches. The Lower Specification Limit (LSL) is 92 inches.

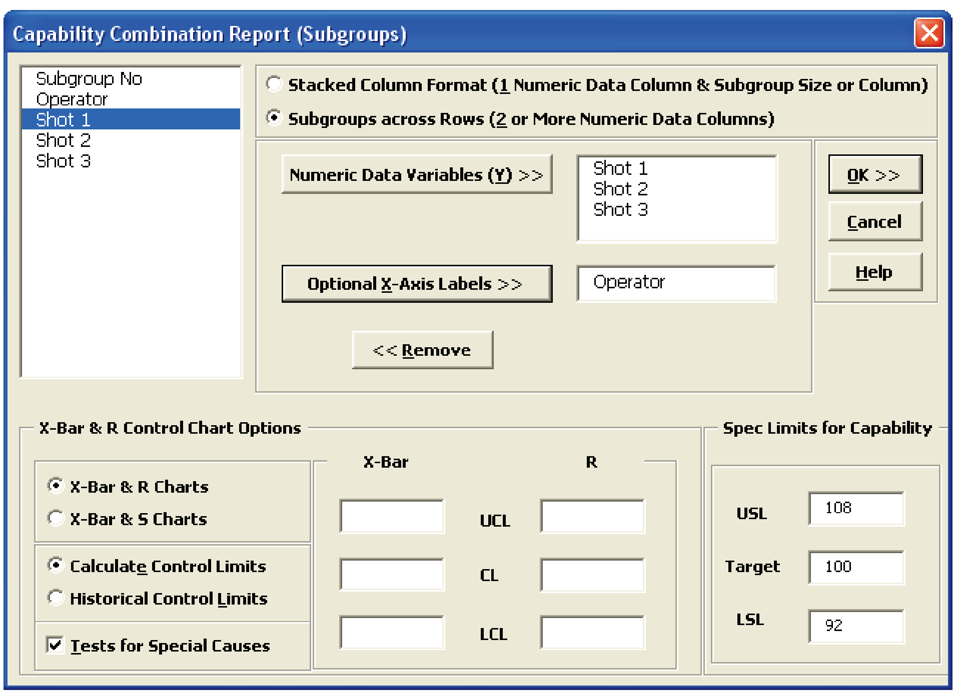
2. Select **B2:F22**; here, we will only use the first 20 subgroups to determine the process capability (subgroups 21 to 25 are studied later in **Control Phase Tools: Statistical Process Control (SPC) Charts: Part B - X-Bar & Range Charts**).

3. Select **SigmaXL > Process Capability > Capability Combination Report (Subgroups)**.

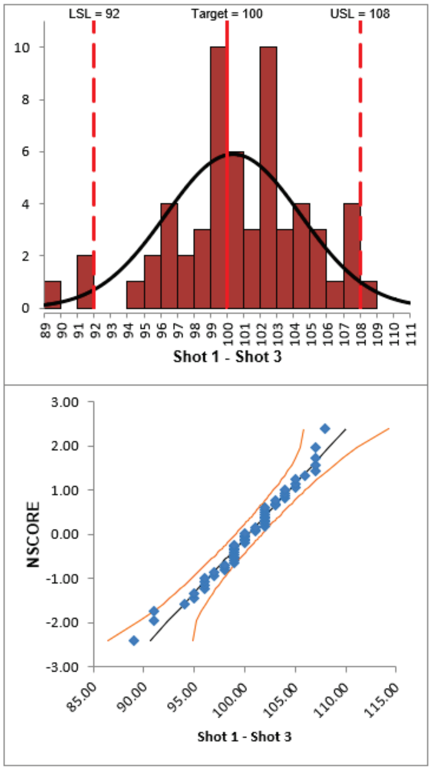
4. Do not check **Use Entire Data Table**!



5. Click **Next**. Select **Subgroups across Rows**, select *Shot 1*, *Shot 2*, *Shot 3*, click **Numeric Data Variables (Y)** >>. Select *Operator*, click **Optional X-Axis Labels** >>. Select **X-Bar & R Charts**. Check **Tests for Special Causes**. Enter **USL** = 108, **Target** = 100, **LSL** = 92, as shown:



6. Click **OK**. The resulting Process Capability Combination report is shown below:



Process Capability Report: Shot 1 - Shot 3	
Count	60
Mean	100.37
StDev (Overall, Long Term)	4.055
StDev (Within, Short Term)	3.721
USL	108
Target	100
LSL	92

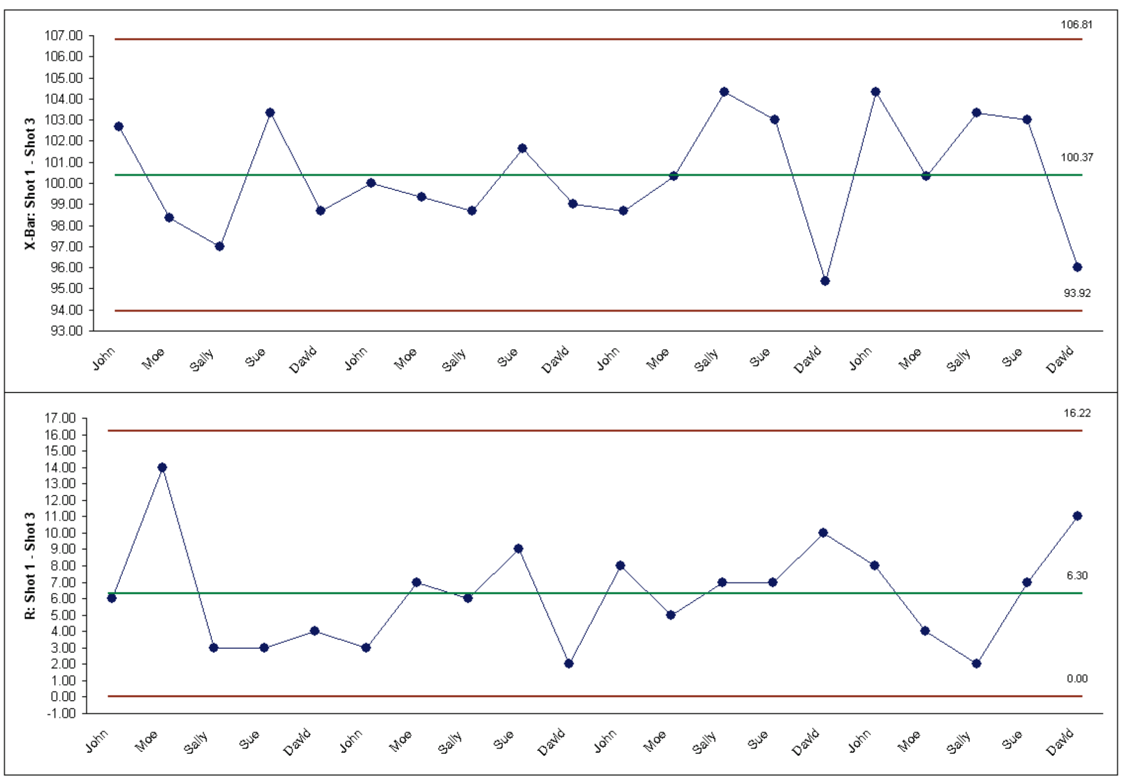
Capability Indices using Overall StDev	
Pp	0.66
Ppu	0.63
Ppl	0.69
Ppk	0.63
Cpm	0.66

Potential Capability Indices using Within StDev	
Cp	0.72
Cpu	0.68
Cpl	0.75
Cpk	0.68

Expected Overall Performance	
ppm > USL	29874.0
ppm < LSL	19531.9
ppm Total	49406
% > USL	2.99%
% < LSL	1.95%
% Total	4.94%

Actual (Empirical) Performance	
% > USL	
% < LSL	5.00%
% Total	5.00%

Anderson-Darling Normality Test	
A-Squared	0.530155
P-Value	0.1689



Tests for Special Causes - X-Bar: Shot 1 - Shot 3
Number of Data Points Failing Tests = 0

Capability Combination Report (Individuals Nonnormal)

An important assumption for process capability analysis is that the data be normally distributed. The Capability Combination Report (Individuals Nonnormal) allows you to transform the data to normality or utilize nonnormal distributions, including:

Box-Cox Transformation (includes an automatic threshold option so that data with negative values can be transformed)

Johnson Transformation

Distributions supported:

- Half-Normal
- Lognormal (2 & 3 parameter)
- Exponential (1 & 2 parameter)
- Weibull (2 & 3 parameter)
- Beta (2 & 4 parameter)
- Gamma (2 & 3 parameter)
- Logistic
- Loglogistic (2 & 3 parameter)
- Largest Extreme Value
- Smallest Extreme Value

Automatic Best Fit based on AD P-Value

For technical details, see Appendix: **Statistical Details for Nonnormal Distributions and Transformations**. See also Andrew Sleeper, *Six Sigma Distribution Modeling*, for further information on these methods.

Note that these transformations and distributions are particularly effective for inherently skewed data but should not be used with bimodal data or where the nonnormality is due to outliers (typically identified with a Normal Probability Plot). In these cases, you should identify the reason for the bimodality or outliers and take corrective action. Another common reason for nonnormal data is poor measurement discrimination leading to “chunky” data. In this case, attempts should be made to improve the measurement system.

Box-Cox Transformation

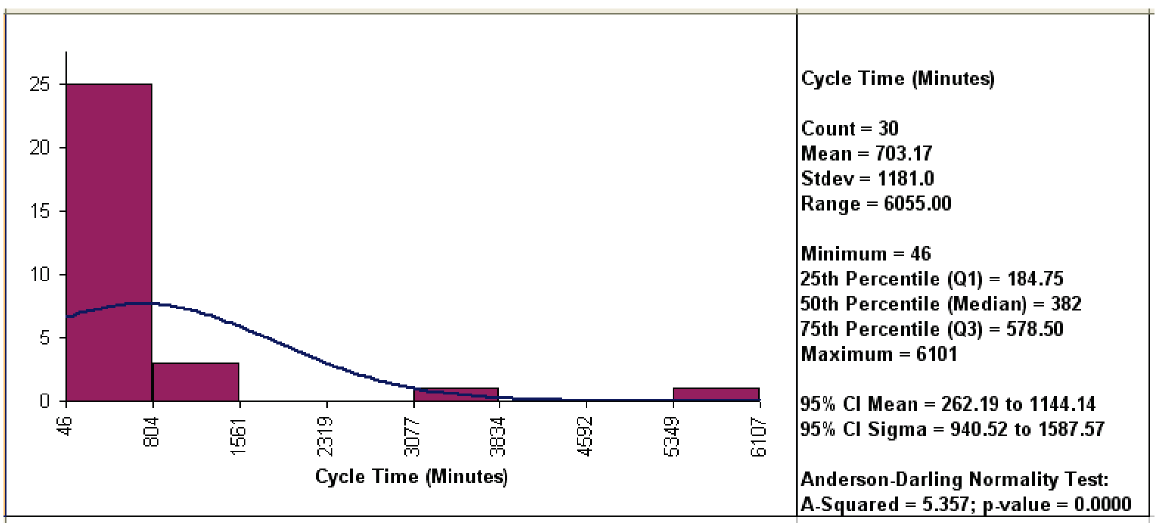
SigmaXL's default setting is to use the Box-Cox transformation which is the most common approach to dealing with nonnormal data. Box-Cox is used to convert nonnormal data to normal by applying a power transformation, Y^lambda, where lambda varies from -5 to +5. You may select rounded or optimal lambda. Rounded is typically preferred since it will result in a more "intuitive" transformation such as Ln(Y) (lambda=0) or SQRT(Y) (lambda=0.5). If the data includes zero or negative values, select Lambda & Threshold. SigmaXL will solve for an optimal threshold which is a shift factor on the data so that all of the values are positive.

1. Open the file **Nonnormal Cycle Time2.xlsx**. This contains continuous data of process cycle times. The Critical Customer Requirement is: USL = 1000 minutes.

2. Let's begin with a view of the data using Histograms and Descriptive Statistics. Click **SigmaXL > Graphical Tools > Histograms & Descriptive Statistics**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

4. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. Click **OK**.

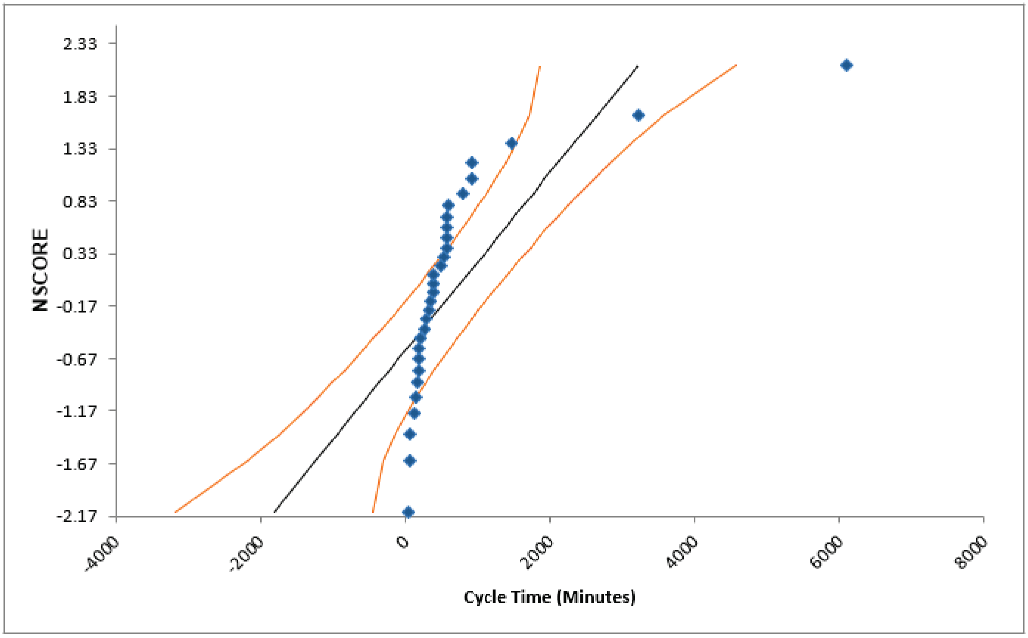


Clearly this is a process in need of improvement. To start, we would like to get a baseline process capability. The problem with using regular Capability analysis is that the results will be incorrect due to the nonnormality in the data. The Histogram and AD P-Value < .05 clearly show that this data is not normal.

5. We will confirm the nonnormality by using a Normal Probability Plot. Click **Sheet 1** Tab (or **F4**). Click **SigmaXL > Graphical Tools > Normal Probability Plots**.

6. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

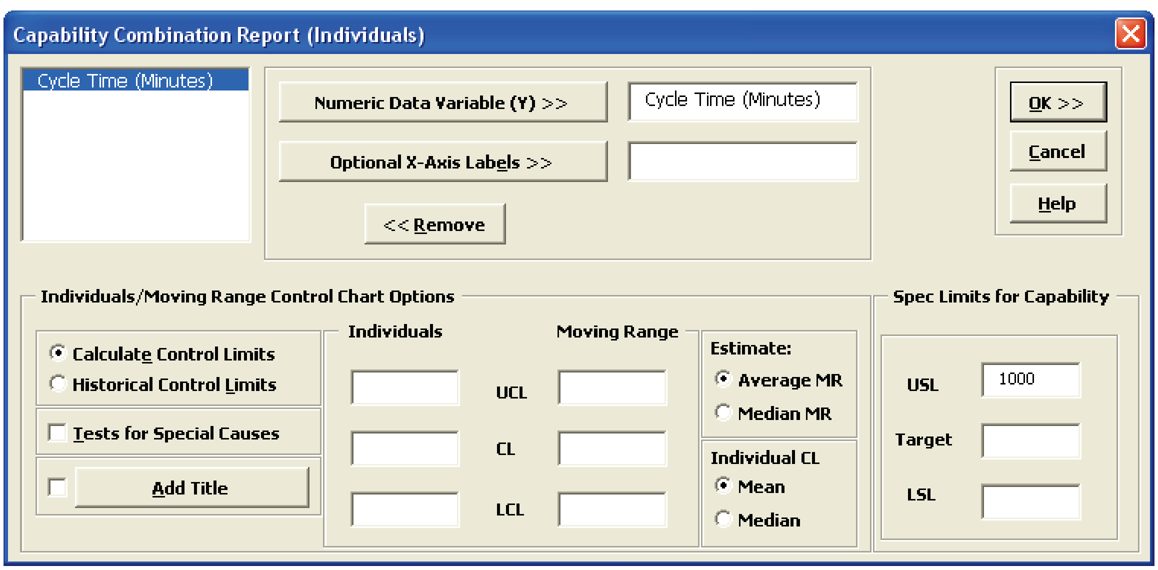
7. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. Click **OK**. A Normal Probability Plot of Cycle Time data is produced:



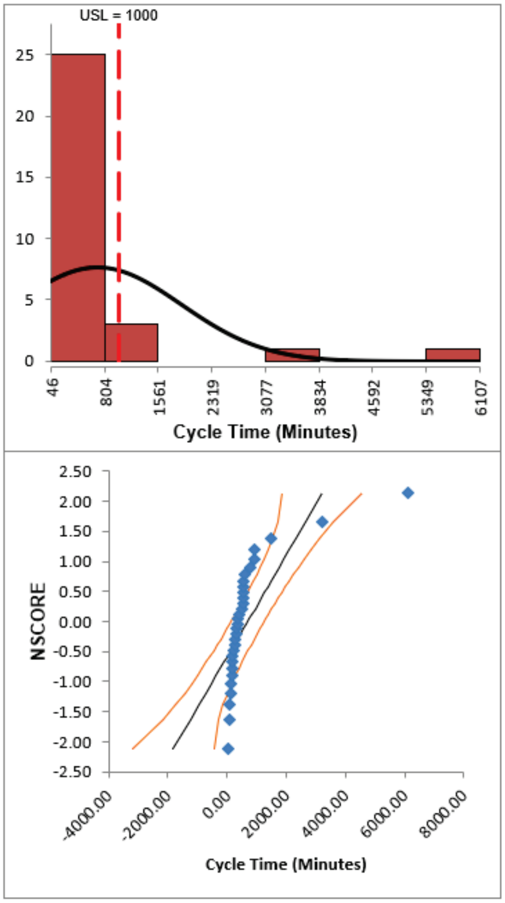
8. The curvature in this normal probability plot confirms that this data is not normal.

9. For now, let us ignore the nonnormal issue and perform a Process Capability study assuming a normal distribution. Click **Sheet 1** Tab. Click **SigmaXL > Process Capability > Capability Combination Report (Individuals)**.

10. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. Enter **USL** = 1000; delete previous **Target** and **LSL** settings.



11. Click **OK**. The resulting Process Capability Report is shown below:



Process Capability Report: Cycle Time (Minutes)	
Count	30
Mean	703.17
StDev (Overall, Long Term)	1181.0
StDev (Within, Short Term)	709.10
USL	1000
Target	
LSL	

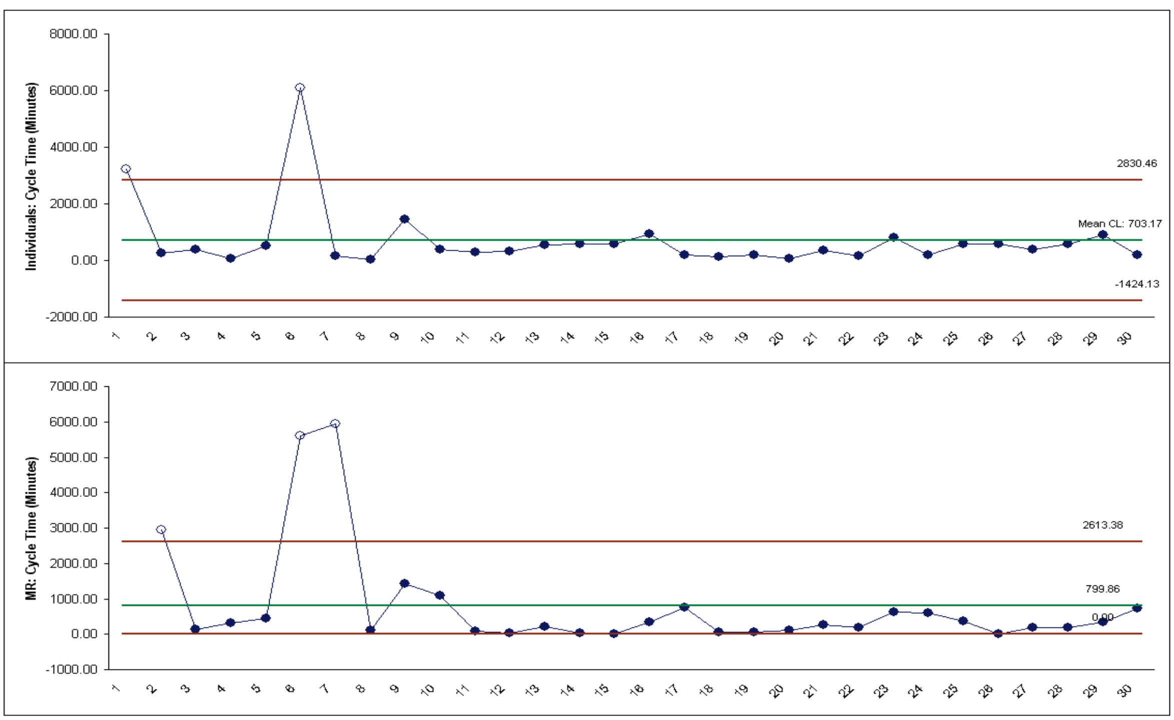
Capability Indices using Overall StDev	
Pp	
Ppu	0.08
Ppl	
Ppk	0.08
Cpm	

Potential Capability Indices using Within StDev	
Cp	
Cpu	0.14
Cpl	
Cpk	0.14

Expected Overall Performance	
ppm > USL	400771.3
ppm < LSL	
ppm Total	400771
% > USL	40.08%
% < LSL	
% Total	40.08%

Actual (Empirical) Performance	
% > USL	10.00%
% < LSL	
% Total	10.00%

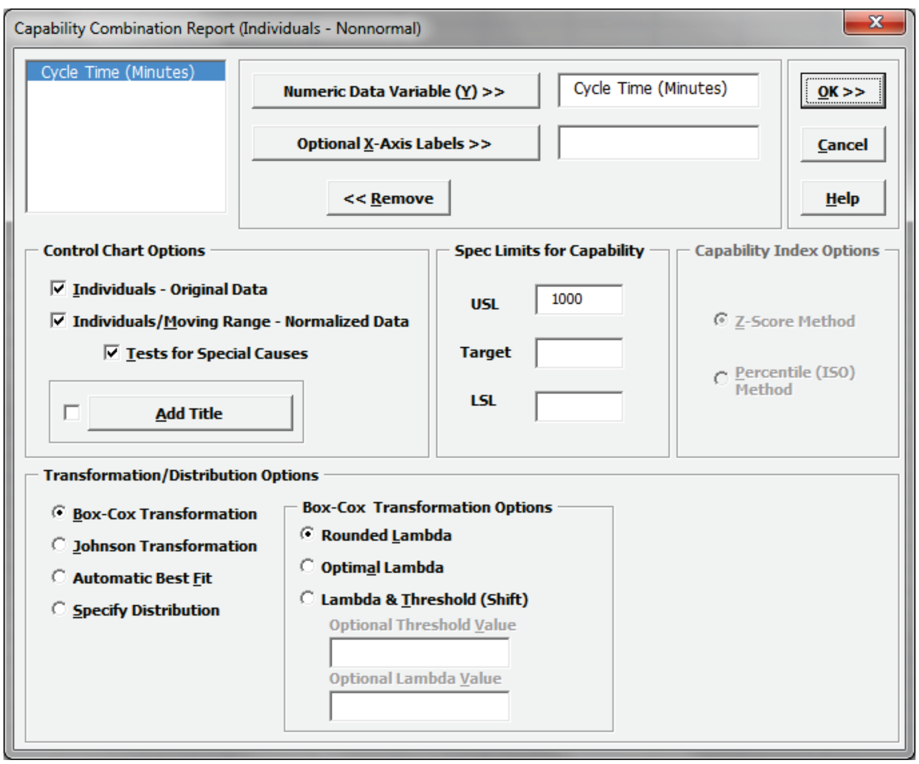
Anderson-Darling Normality Test	
A-Squared	5.357
P-Value	0.0000



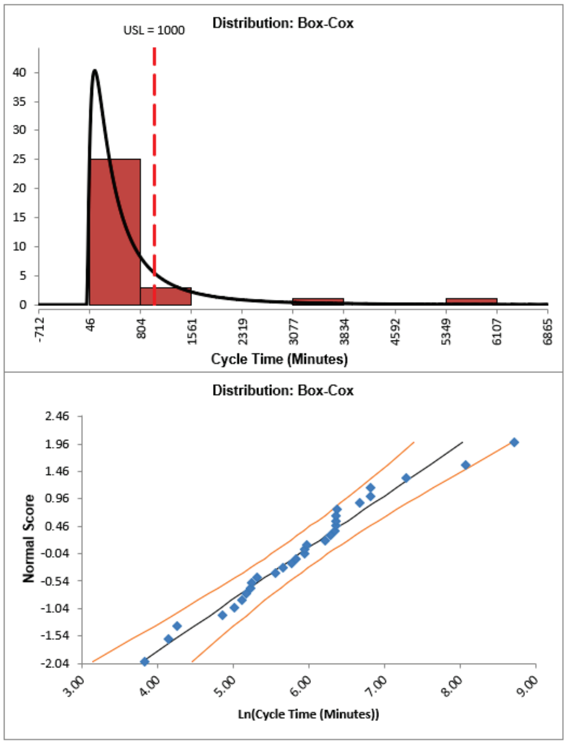
Notice the discrepancy between the Expected Overall (Theoretical) Performance and Actual (Empirical) Performance. This is largely due to the nonnormality in the data, since the expected performance assumes that the data is normal. So why not just use the actual performance and disregard the expected? This would not be reliable because the sample size, n = 30, is too small to estimate performance using pass/fail (discrete) criteria.

Also note that the process appears to be out-of-control on both the individuals and moving range charts.

12. We will now perform a process capability analysis using the Capability Combination Report for Nonnormal Individuals. Click **Sheet 1** Tab (or **F4**). Click **SigmaXL > Process Capability > Nonnormal > Capability Combination Report (Individuals Nonnormal).** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
13. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. Enter **USL** = 1000. We will use the default selection for **Transformation/Distribution Options: Box-Cox Transformation** with **Rounded Lambda**. Check **Tests for Special Causes** as shown:



14. Click **OK**. The resulting Process Capability Combination report is shown below:



Process Capability Report (Nonnormal): Cycle Time (Minutes)	
Box-Cox Transformation: Z = Ln(Y)	
Sample Count	30
Sample Mean	703.17
USL	1000
Target	
LSL	
Lambda (Rounded)	0
Mean (Transformed)	5.920
StDev (Transformed Overall, Long Term)	1.061030703
StDev (Transformed Within, Short Term)	1.004837519
USL (Transformed)	6.908
Target (Transformed)	
LSL (Transformed)	

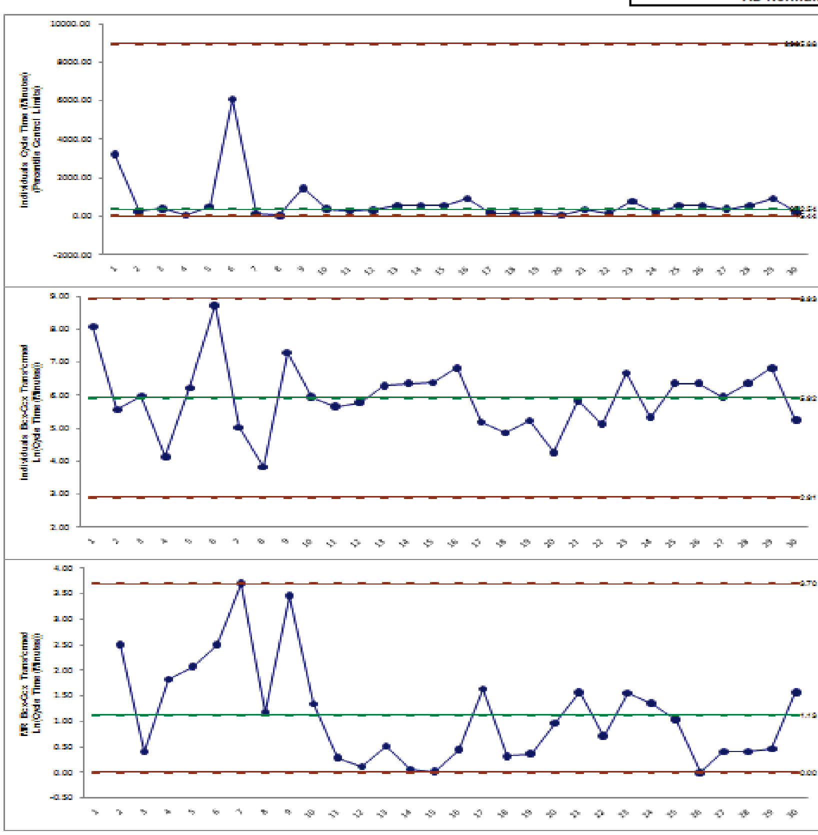
Capability Indices using Transformed Overall StDev	
Pp	
Ppu	0.31
Ppl	
Ppk	0.31

Potential Capability Indices using Transformed Within StDev	
Cp	
Cpu	0.33
Cpl	
Cpk	0.33

Expected Overall Performance	
ppm > USL	176026.1
ppm < LSL	
ppm Total	176026.1
% > USL	17.60%
% < LSL	
% Total	17.60%

Actual (Empirical) Performance	
% > USL	10.00%
% < LSL	
% Total	10.00%

Anderson-Darling Normality Tests	
AD Normality Transformed Data	0.369554
AD Normality P-Value Transformed Data	0.4041
AD Normality Original Data	5.357
AD Normality P-Value Original Data	0.0000



Tests for Special Causes - Indv: Cycle Time (Minutes)
Number of Data Points Failing Tests = 0

The **AD Normality P-Value Transformed Data** value of 0.404 confirms that the Box-Cox transformation to normality was successful. The process capability indices and expected performance can now be used to establish a baseline performance. Note that there are no out-of-control signals on the control charts, so the signals observed earlier when normality was assumed were false alarms.

The **Individuals – Original Data** chart displays the untransformed data with control limits calculated as:

UCL = 99.865 percentile

CL = 50th percentile

LCL = 0.135 percentile

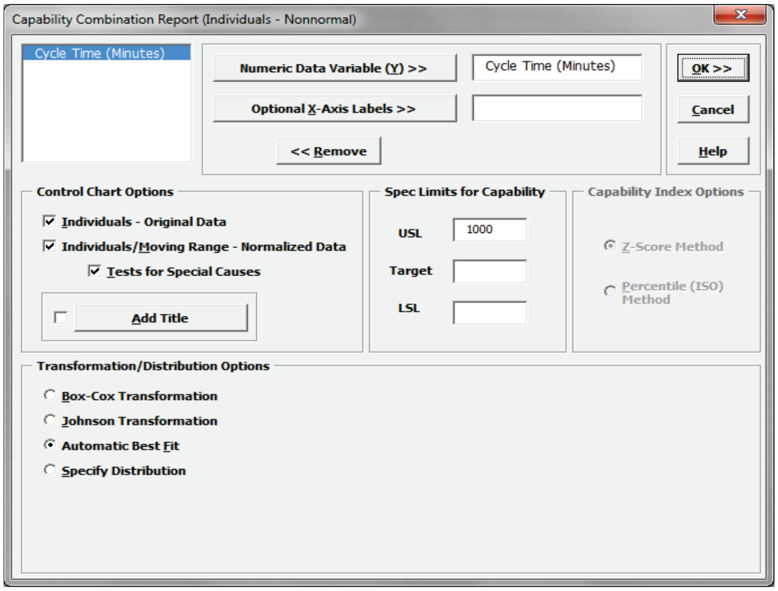
The benefit of displaying this chart is that one can observe the original untransformed data. Since the control limits are based on percentiles, this represents the overall, long term variation rather than the typical short-term variation. The limits will likely be nonsymmetrical.

The **Individuals/Moving Range – Normalized Data** chart displays the transformed z-values with control limits calculated using the standard Shewhart formulas for Individuals and Moving Range charts. The benefit of using this chart is that tests for special causes can be applied and the control limits are based on short term variation. The disadvantage is that one is observing transformed data on the chart rather than the original data.

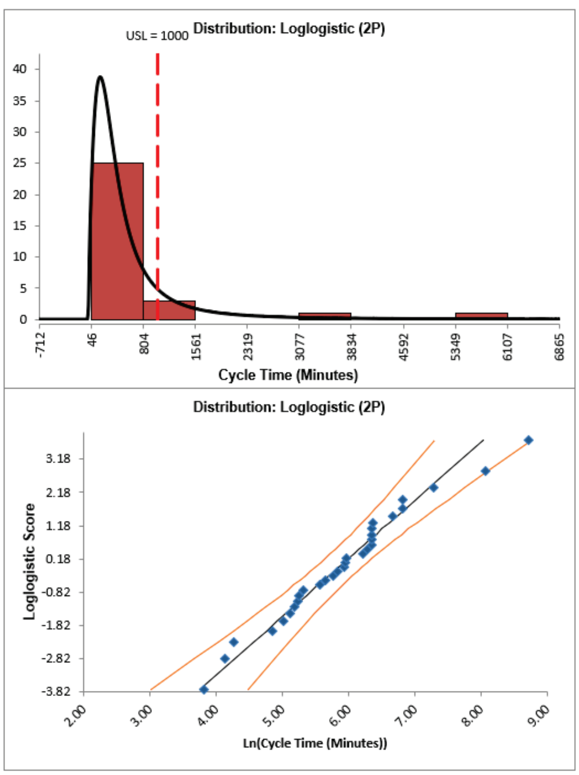
Automatic Best Fit

Now we will redo the capability analysis using the **Automatic Best Fit option**.

15. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select **Automatic Best Fit** as shown:



16. Click **OK**. The resulting Process Capability Combination report is shown below. Please note that due to the extensive computations required, this could take up to 1 minute (or longer for large datasets):



Process Capability Report (Nonnormal): Cycle Time (Minutes)	
Distribution: Loglogistic (2P)	
Sample Count	30
Sample Mean	703.17
Location	5.895
Scale	0.571471221
StDev (Transformed Overall, Long Term)	1
StDev (Normalized Within, Short Term)	0.964388
USL	1000
Target	
LSL	

Capability Indices (Z-Score Method)	
Pp	
Ppu	0.35
Ppl	
Ppk	0.35

Potential Capability Indices using Normalized Within StDev	
Cp	
Cpu	0.37
Cpl	
Cpk	0.37

Expected Overall Performance	
ppm > USL	145274.6
ppm < LSL	
ppm Total	145274.6
% > USL	14.53%
% < LSL	
% Total	14.53%

Actual (Empirical) Performance	
% > USL	10.00%
% < LSL	
% Total	10.00%

Anderson-Darling Goodness-of-Fit Tests	
AD Loglogistic (2P)	0.245264
AD Loglogistic (2P) P-Value	> 0.2500
AD Normality Original Data	5.357
AD Normality P-Value Original Data	0.0000

The 2 Parameter Loglogistic distribution was selected as the best fit distribution. For details on how this selection was made, see Appendix: **Statistical Details for Nonnormal Distributions and Transformations**.

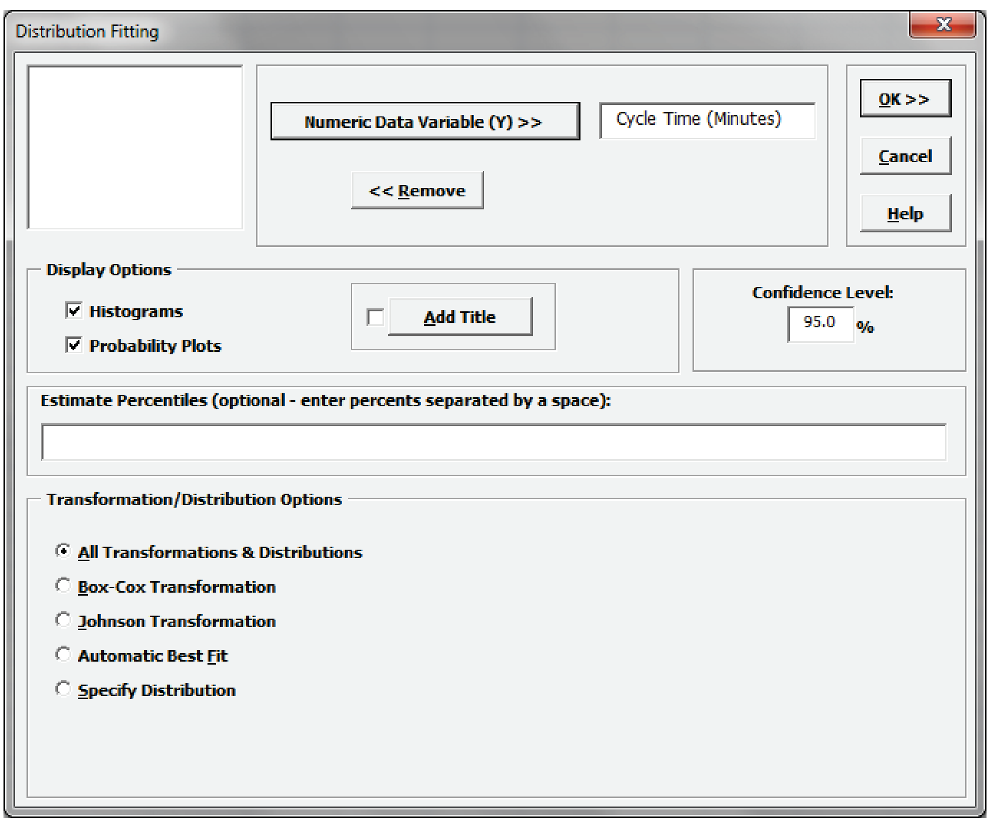
The Anderson Darling statistic for the Loglogistic distribution is 0.245 which is less than the 0.37 value for the AD Normality test of the Box-Cox transformation indicating a better fit. (Note that published AD P-Values for this distribution are limited to a maximum value of 0.25. The best fit selection uses a P-Value estimate that is obtained by transforming the data to normality and then using a modified Anderson Darling Normality test on the transformed data).

Distribution Fitting

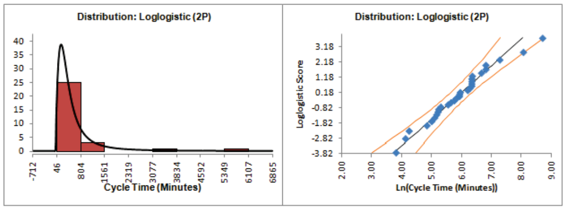
Another helpful tool to evaluate transformations and distributions is **Distribution Fitting.**

1. Click **Sheet 1** Tab (or **F4**). Click **SigmaXL > Process Capability > Nonnormal > Distribution Fitting**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

2. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. We will use the default selection for **Transformation/Distribution Options: All Transformations & Distributions** as shown:



3. Click **OK**. The resulting Distribution Fitting report is shown below. Please note that due to the extensive computations required, this could take up to 1 minute (or longer for large datasets):

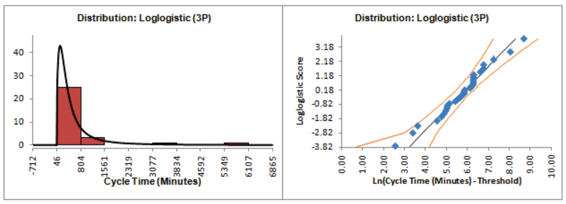


Distribution Fitting Report: Cycle Time (Minutes)	
Distribution: Loglogistic (2P)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	-221.011
AD Loglogistic (2P)	0.245264
AD Loglogistic (2P) p-value	> 0.2500

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.895	0.179714	5.543	6.247
Scale	0.571471	0.087949896	0.422662	0.772671

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	8.329	5.067	2.528	27.442
50	363.22	65.277	255.39	516.59
99.865	15841	9631.1	4811.0	52156

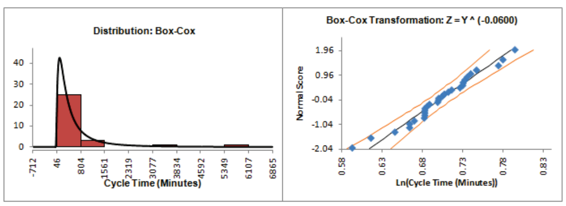


Distribution Fitting Report: Cycle Time (Minutes)	
Distribution: Loglogistic (3P)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	-220.631
AD Loglogistic (3P)	0.299483
AD Loglogistic (3P) p-value	0.562 (Z-Score Est.)

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.771	0.232556	5.315	6.227
Scale	0.671824	0.151455	0.431879	1.045078097
Threshold (ML Estimate)	33.025	23.549	-13.129	79.180

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	36.816	20.605	12.293	110.27
50	353.84	67.889	242.94	515.38
99.865	27183	25811	4227.5	174793

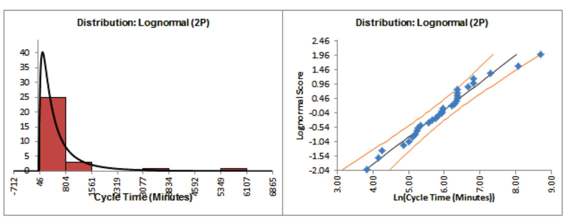


Distribution Fitting Report: Cycle Time (Minutes)	
Box-Cox Transformation: Z = Y ^ (-0.0600)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	51.498
AD Box-Cox	0.335933
AD Box-Cox p-value	0.485000

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Lambda (Optimal)	-0.06		-0.573963	0.423205
Mean (Transformed Data)	0.702382	0.007937578	0.686825	0.717939
StDev (Transformed Data)	0.044219135	0.00570856	0.034333861	0.056950541

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	11810		36158	4138.7
50	360.67		523.88	250.34
99.865	20.179		42.944	9.797

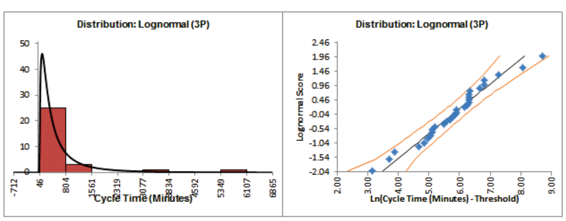


Distribution Fitting Report: Cycle Time (Minutes)	
Distribution: Lognormal (2P)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	-221.456
AD Lognormal (2P)	0.369554
AD Lognormal (2P) p-value	0.425000

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.920	0.193717	5.541	6.300
Scale	1.061030703	0.140537	0.818434	1.376

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	15.445	7.166	6.221	38.346
50	372.54	72.167	254.85	544.58
99.865	8985.9	4169.3	3619.2	22310

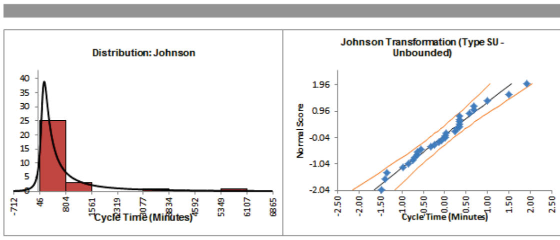


Distribution Fitting Report: Cycle Time (Minutes)	
Distribution: Lognormal (3P)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	-221.179
AD Lognormal (3P)	0.385496
AD Lognormal (3P) p-value	0.370 (Z-Score Est.)

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.805	0.254377	5.307	6.304
Scale	1.160	0.221685	0.797715	1.687
Threshold (ML Estimate)	22.909	23.784	-23.708	69.525

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	33.134	18.057	11.386	96.418
50	354.92	73.985	235.88	534.03
99.865	10804	6560.8	3285.8	35521



Distribution Fitting Report: Cycle Time (Minutes)	
Johnson Transformation (Type SU - Unbounded)	
Z = Shape1 + Shape2 * Asinh((Y - Location) / Scale)	
Sample Count	30
Sample Mean	703.17

Model Summary and Goodness-of-Fit	
Log-Likelihood	-35.536
AD Johnson	0.286526
AD Johnson p-value	0.599000

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Shape1	-1.0451			
Shape2	0.613400			
Location	119.33			
Scale	96.387			
Mean (Transformed Data)	-0.047151462	0.144433	-0.330234	0.235931
StDev (Transformed Data)	0.804576	0.103876532	0.624699	1.036246798

Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	-360.45		-1331.1	-28.013
50	355.08		258.88	502.38
99.865	12664		4305.3	37712

The distributions and transformations are sorted in descending order using the AD Normality P-Value on the transformed z-score values. Note that the first distribution shown may not be the selected "best fit", because the best fit procedure also looks for models that are close but with fewer parameters.

The reported AD P-Values are those derived from the particular distribution. The AD P-Value is not available for distributions with a threshold (except Weibull), so the AD Normality P-Value on the transformed z-score values is used (labeled as Z-Score Est.).

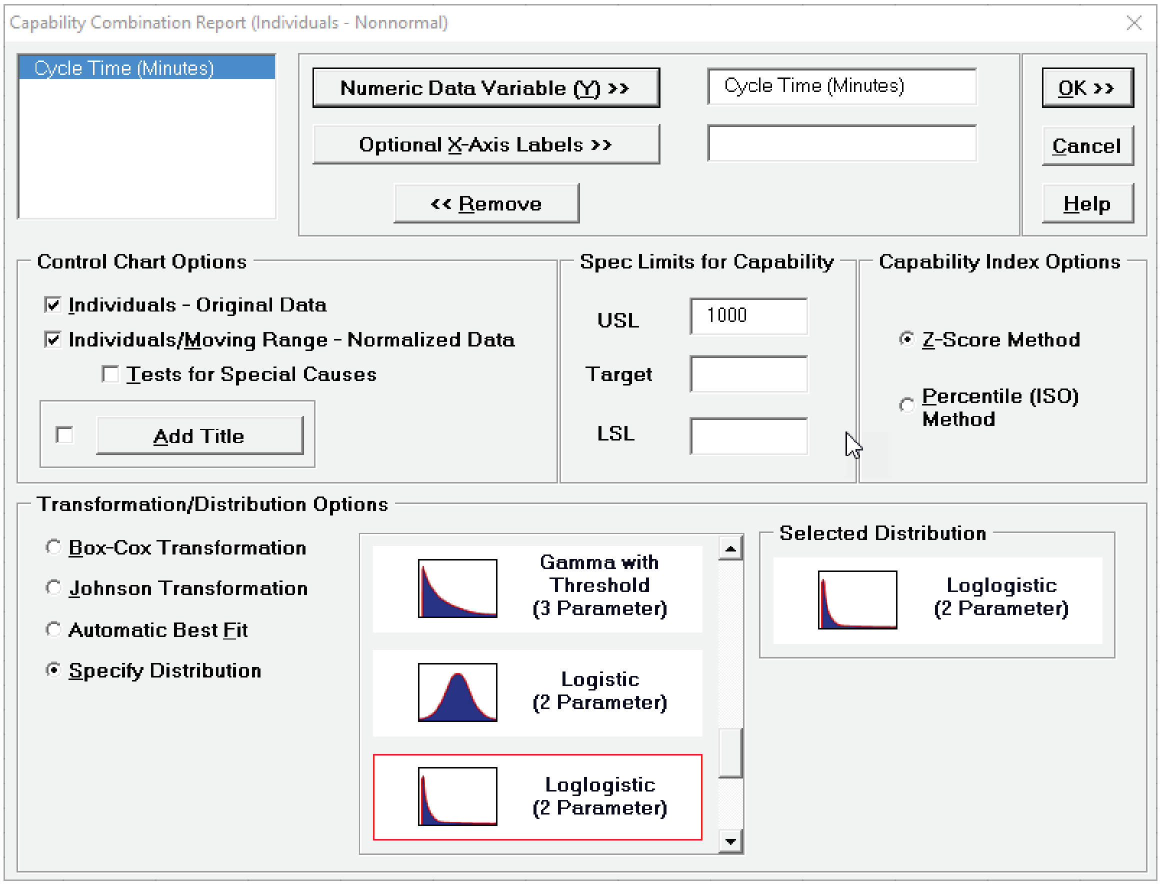
Since the sort order is based on the AD P-Values from Z-Score estimates, it is possible that the reported distribution-based AD P-Values may not be in perfect descending order. However, any discrepancies based on sort order will likely not be statistically or practically significant.

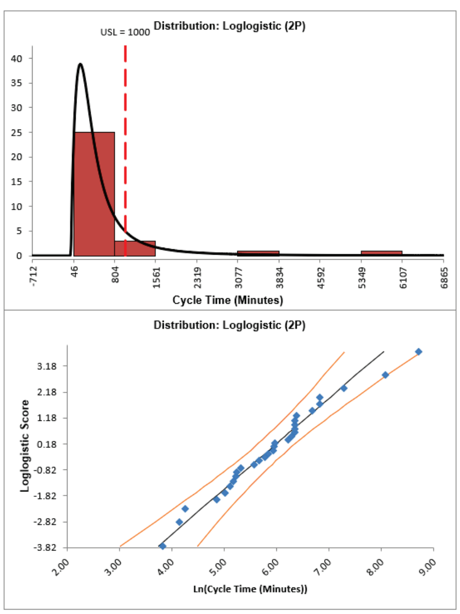
Some data will have distributions and transformations where the parameters cannot be solved (e.g., 2-parameter Weibull with negative values). These are excluded from the Distribution Fitting report.

The parameter estimates and percentile report includes a confidence interval as specified in the **Distribution Fitting** dialog, with 95% being the default. Note that the wide intervals here are due to the small sample size, n = 30.

The control limits for the percentile-based Individuals chart will be the 0.135% (lower control limit), 50% (center line, median) and 99.865% (upper control limit). Additional percentiles may be entered in the **Distribution Fitting** dialog.

After reviewing this report, if you wish to perform a process capability analysis with a particular transformation or distribution, simply select **Specify Distribution** from the **Transformation/Distribution Options** in the **Capability Combination Report (Individuals - Nonnormal)** dialog as shown below (using 2 Parameter Loglogistic):





Process Capability Report (Nonnormal): Cycle Time (Minutes)	
Distribution: Loglogistic (2P)	
Sample Count	30
Sample Mean	703.17
Location	5.895
Scale	0.571471221
StDev (Transformed Overall, Long Term)	1
StDev (Normalized Within, Short Term)	0.964388
USL	1000
Target	
LSL	

Capability Indices (Z-Score Method)	
Pp	
Ppu	0.35
Ppl	
Ppk	0.35

Potential Capability Indices using Normalized Within StDev	
Cp	
Cpu	0.37
Cpl	
Cpk	0.37

Expected Overall Performance	
ppm > USL	145274.6
ppm < LSL	
ppm Total	145274.6
% > USL	14.53%
% < LSL	
% Total	14.53%

Actual (Empirical) Performance	
% > USL	10.00%
% < LSL	
% Total	10.00%

Anderson-Darling Goodness-of-Fit Tests	
AD Loglogistic (2P)	0.245264
AD Loglogistic (2P) P-Value	> 0.2500
AD Normality Original Data	5.357
AD Normality P-Value Original Data	0.0000

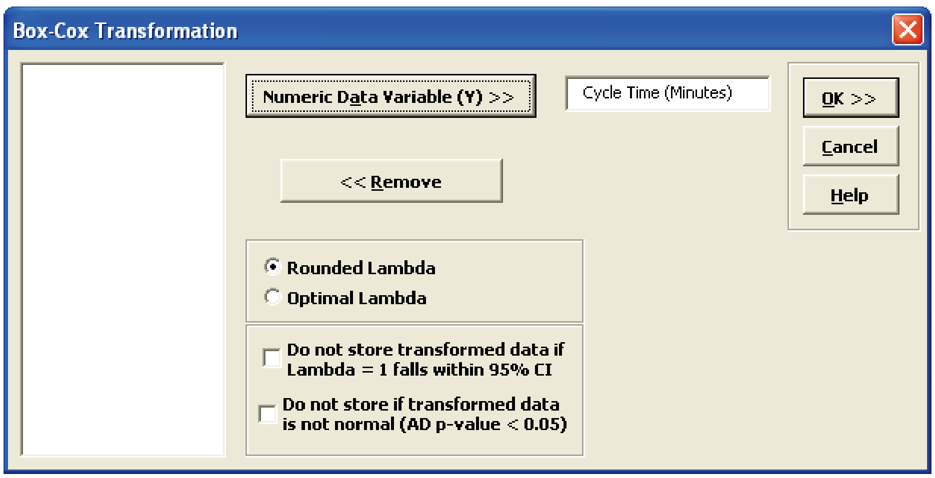
Box-Cox Transformation

This is a standalone tool that allows you to visually see how the Box-Cox transformation selects a rounded or optimal lambda value.

1. Open the file **Nonnormal Cycle Time2.xlsx**. Select **Sheet 1** Tab.

2. Click **SigmaXL > Process Capability > Nonnormal > Box-Cox Transformation (**or **SigmaXL > Data Manipulation > Box-Cox Transformation** or **SigmaXL > Control Charts > Nonnormal > Box-Cox Transformation**). Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

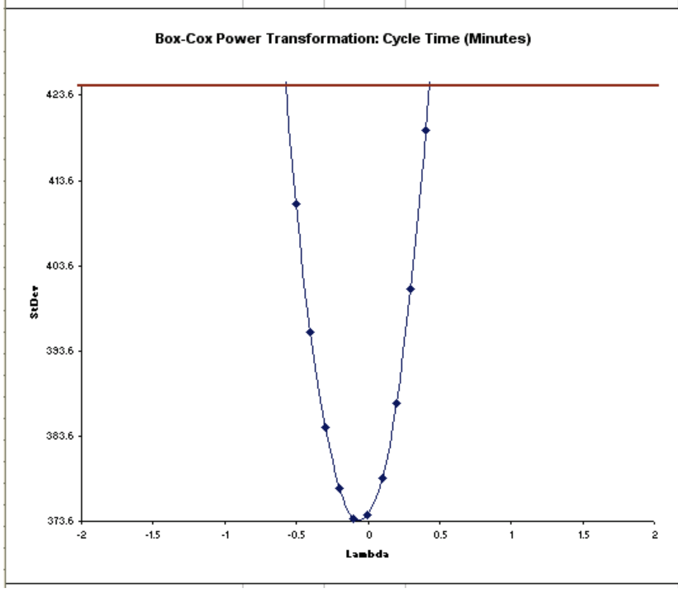
3. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. The selected variable must contain all positive values.



Tip: Note that while this tool is often successful to transform the data to normality, there may not be a suitable transformation to make the data normal. The output report indicates the Anderson-Darling P-Value for the transformed data. You may wish to check **Do not store if transformed data is not normal**. Another option is **Do not store transformed data if Lambda = 1 falls within 95%CI**. This latter option prevents you from using transformations that do not result in a statistically significant improvement towards normality.

4. Click **OK**. The resulting report is shown:

Box-Cox Power Transformation: Cycle Time (Minutes)	
Optimal Lambda	-0.06
Final (Rounded) Lambda	0
UC Lambda (95%)	0.423205
LC Lamda (95%)	-0.573963
Anderson Darling Normality Test for Transformed Data:	
A-Squared	0.369554
AD p-value	0.4041



Cycle Time (Minutes)	Transformed Data (Ln(Y))
3216	8.076
261	5.565
392	5.971
63	4.143
501	6.217
6101	8.716
151	5.017
46	3.829
1465	7.290
383	5.948
287	5.659
322	5.775
538	6.288
574	6.353
586	6.373
920	6.824
178	5.182
129	4.860
187	5.231
71	4.263
342	5.835
167	5.118
794	6.677
205	5.323
576	6.356
575	6.354
381	5.943
576	6.356
918	6.822
190	5.247

5. The fact that Lambda=1 falls outside of the 95% confidence interval tells us that the transformation is statistically significant. The Anderson-Darling P-Value of 0.4041 indicates that we cannot reject the null hypothesis that the transformed data is normal, so the Ln transformation has successfully converted the data to normality.

   SigmaXL uses the Box-Cox Confidence Interval formula recommended by Draper & Smith (1998), *Applied Regression Analysis*, 3rd ed., Wiley, pp. 282-283, Formula 13.2.10.

Part K – Reliability/Weibull Analysis

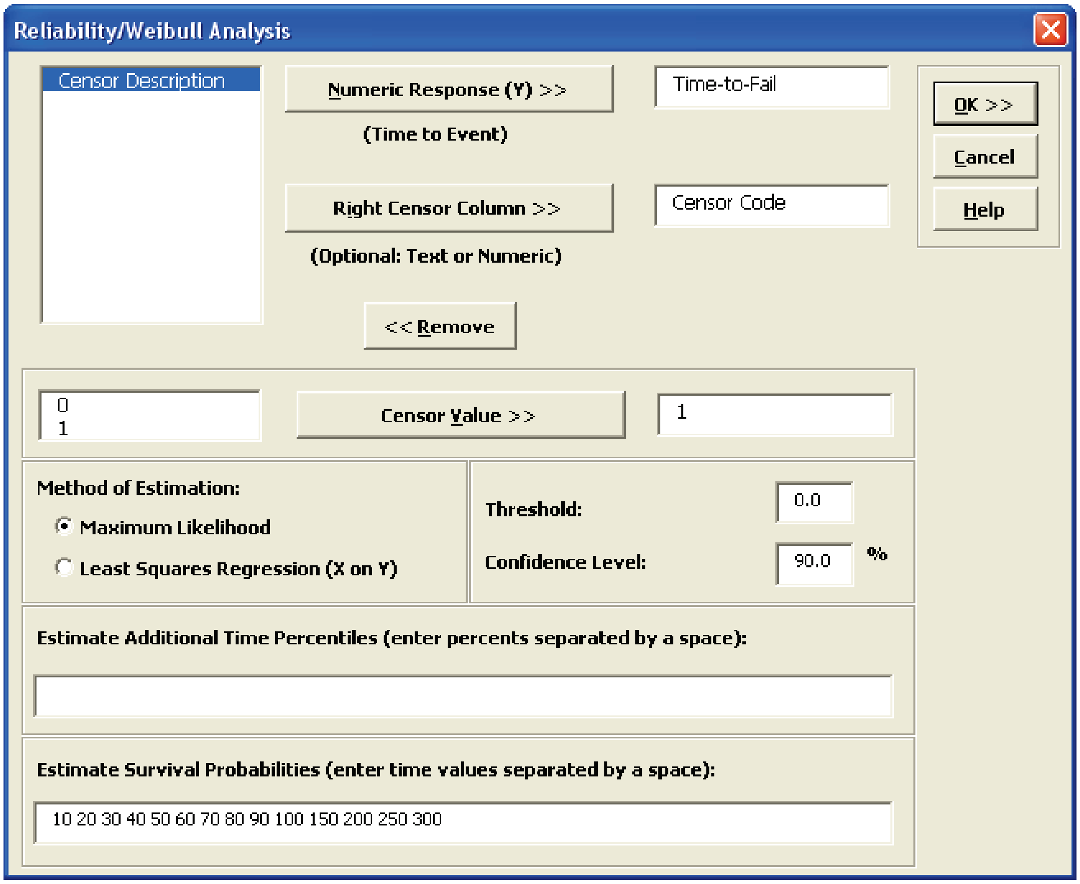
Reliability/Weibull Analysis

The Weibull distribution is a continuous distribution that was publicized by Waloddi Weibull in 1951. It has become widely used, especially in the reliability field. The Weibull distribution's popularity resulted from its ability to be used with small sample sizes and its flexibility. See Dodson, B. "The Weibull Analysis Handbook" Second Edition, ASQ, for further information. For SigmaXL technical details, see Appendix: **Statistical Details for Nonnormal Distributions and Transformations**.

The Weibull distribution can also be used as an alternative to the Box-Cox Transformation to determine Process Capability Indices or Control Limits for nonnormal data.

Weibull is particularly effective when the data are censored, with exact failure times unknown or suspension of test. Suspension of test is also referred to as right censored. SigmaXL can analyze complete or right censored data.

1. Open the file **Weibull Dodson.xlsx**. This contains right censored time-to-failure data from Dodson, page 28.
2. Click **SigmaXL > Reliability/Weibull Analysis**. Ensure that the entire data table is selected. Click **Next**.
3. Select *Time-to-Fail*, click **Numeric Response (Y) >>**. Note that the selected variable must contain all positive values. Select *Censor Code*, click **Right Censor Column >>**. The default censor value of 1 will be used. We will use the default method of estimation, **Maximum Likelihood** (some reliability practitioners prefer **Least Squares Regression (X on Y)** for small sample sizes). Other defaults include **Threshold** = 0, and **Confidence Level** = 90%. Enter time values to estimate survival probabilities as shown:



4. The resulting Weibull analysis report is shown:

Weibull Analysis: Time-to-Fail

User Settings:

Estimation Method	Maximum Likelihood
Confidence Level	90.0
Threshold	0

Censoring Information:

Number of Uncensored Observations	11
Number of Right Censored Observations	9
Total	20

Model Summary and Goodness-of-Fit:

Log-Likelihood	-66.176
Hollander-Proschan Test	0.029
Hollander-Proschan p-value	0.9771

Parameter Estimates:

Parameter	Estimate	SE Estimate	Lower 90% CI	Upper 90% CI
Shape	2.9744	0.716995	2.0008	4.4218
Scale	203.29	20.851	171.74	240.65

Distribution Characteristics:

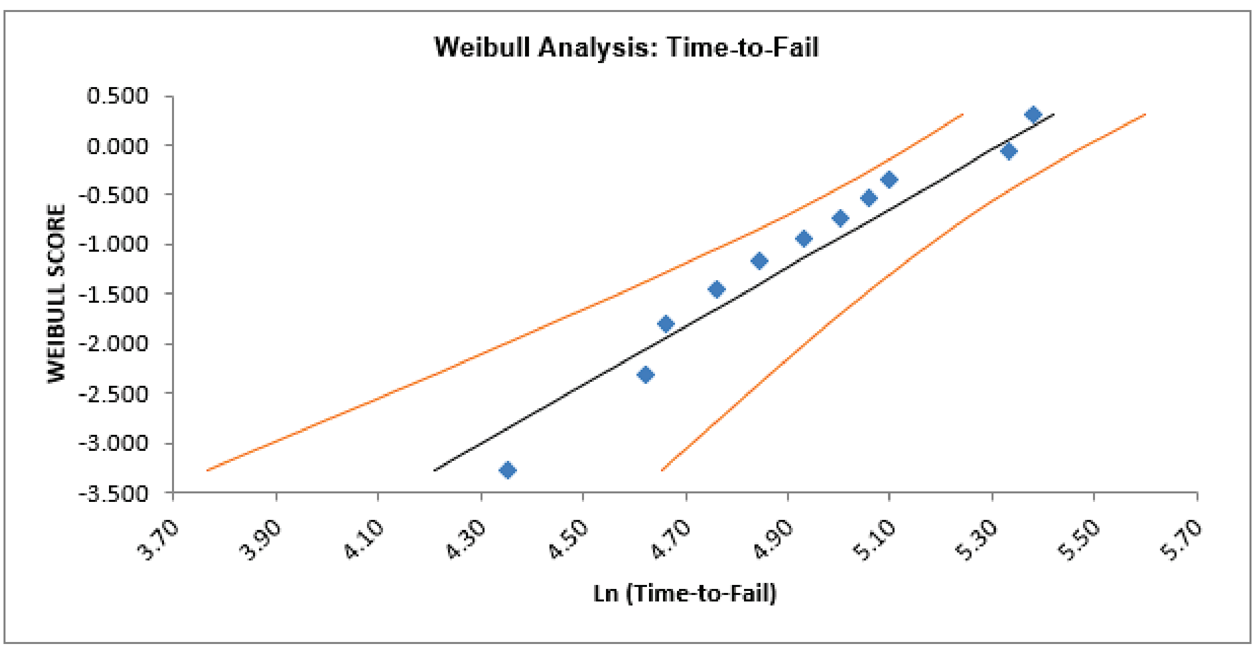
	Estimate	SE Estimate	Lower 90% CI	Upper 90% CI
Mean (MTTF)	181.47	18.420	153.57	214.44
Standard Deviation	66.463	16.172	44.541	99.174

Percentile Report:

Percentage	Percentile (Time)	SE Percentile	Lower 90% CI	Upper 90% CI
0.1	19.935	11.035	8.0201	49.549
0.135	22.052	11.680	9.2279	52.699
0.5	34.268	14.592	17.010	69.037
1	43.298	16.077	23.508	79.747
5	74.895	18.491	49.898	112.41
10	95.401	18.620	69.204	131.52
25	133.73	17.724	107.53	166.30
50	179.73	18.395	151.88	212.68
75	226.89	24.902	189.42	271.78
90	269.09	35.286	216.88	333.87
95	293.99	42.803	231.38	373.54
99	339.71	58.540	255.86	451.03
99.5	356.11	64.682	264.14	480.10
99.865	383.55	75.460	277.51	530.11
99.9	389.32	77.800	280.26	540.83

Survival Probability Report:

Time	Survival Probability	Lower 90.0% CI	Upper 90.0% CI
10	0.999871	0.995704	0.999996
20	0.998990	0.985014	0.999932
30	0.996631	0.968919	0.999639
40	0.992091	0.947956	0.998821
50	0.984698	0.922545	0.997055
60	0.973826	0.893053	0.993800
70	0.958916	0.859814	0.988415
80	0.939499	0.823131	0.980189
90	0.915219	0.783272	0.968381
100	0.885858	0.740474	0.952288
150	0.667092	0.488546	0.795498
200	0.385752	0.208191	0.560906
250	0.157255	0.035498215	0.358758
300	0.04151304	0.00136017	0.215707



5. The Hollander-Proschan test is used when there are right censored observations (see Dodson). If the P-Value is less than 0.05, this indicates significant lack of fit, so Weibull would not be an appropriate distribution. Given the P-Value of 0.9771, there is no lack-of-fit with this time-to-fail data. This is confirmed in the Weibull probability plot with the data roughly following a straight line.

6. If the data is uncensored then the Anderson-Darling unadjusted test is used.

7. The shape parameter, also referred to as Beta, is the slope of the line on the Weibull probability plot. The Weibull distribution can be used in a wide variety of situations and dependent on the value of Beta, is equal to or can approximate several other distributions. For example, if Beta = 1, the Weibull distribution is identical to the exponential distribution; if Beta = 2.5, the Weibull distribution approximates the lognormal distribution; and if Beta = 3.6, the Weibull distribution approximates the normal distribution. Because of this flexibility, there are few observed failure rates that cannot be accurately modeled by the Weibull distribution. Furthermore, if Beta is < 1, the system is suffering from infant mortality failures; if Beta is = 1, failures are random – they are occurring at a constant failure rate and are independent of time; if Beta is > 1, components are wearing out with age.

8. The scale parameter is the characteristic life. This is the point at which we could expect 63.2% of the population under study to have failed. The survival probability report shows that the survival probability at time 200 = 0.386. The cumulative failure probability at 200 is 1 – 0.386 = 0.614.

9. The percentile report tells us that we can expect a 50% failure rate (one-half of the population will fail) at time = 179.73, i.e. the Median = 179.73.

10. The percentile report can also be used to determine control limits for nonnormal data. The lower control limit is the 0.135 % value = 22.052; the center line is the 50% value = 179.73; the upper control limit is the 99.865 % value = 383.55.

SigmaXL: Analyze Phase Tools

Part A – Stratification with Pareto

SigmaXL's Pareto tool allows you to create Basic (Single) or Advanced (Multiple) Pareto Charts. Advanced Pareto charts are particularly useful in the Analyze Phase because of the ease with which you can slice and dice (or stratify) your data. Of course, Pareto charts are not limited to the Analyze Phase – they can also be used to aid project selection and to prioritize in the Measure Phase.

Consider the following guidelines to help ensure that your Pareto analysis is successful:

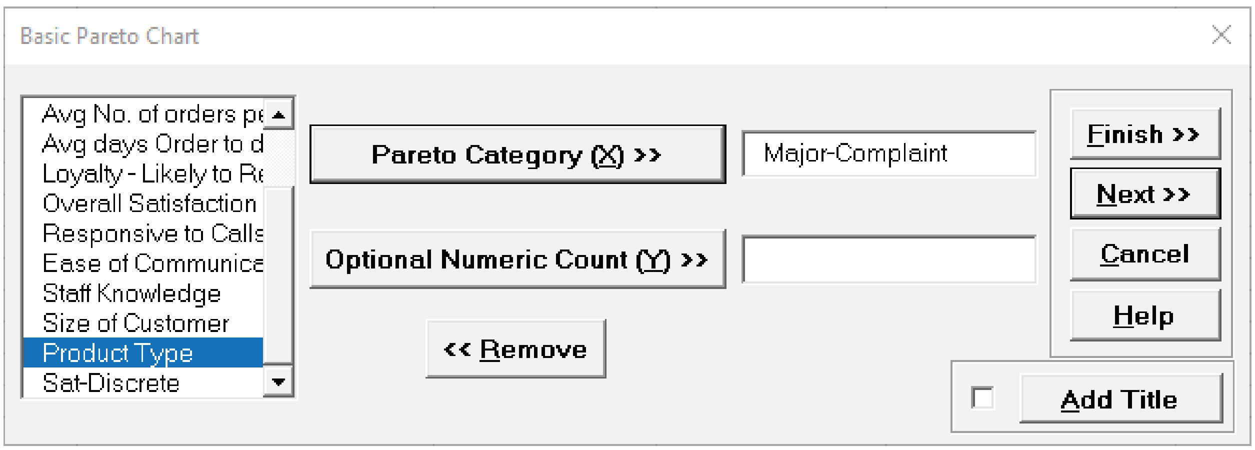
- Your Pareto analysis will only be as good as the quality of the data collected. Ensure that you have the right data and that the data is correct. Use other graphs such as run charts to apply a sanity check to your data.
- Check process stability using appropriate control charts. If the process is not in control, your prioritization of defects and root causes could be invalid.
- Avoid collecting data over too short a time period. Your data may not be representative of the process as a whole. Also keep in mind that since the data is discrete, a minimum sample size of 500 is recommended with 1000 preferred.
- Conversely, data gathered over too long a time period may include process changes that could lead to incorrect conclusions. SigmaXL provides a date subsetting feature that allows you to easily explore different time periods.
- If your initial Pareto analysis does not yield useful results, explore other categories that may be important. SigmaXL's Advanced Charts makes it easy for you to 'slice and dice' your data with different X categories.
- Consider Pareto charting measures such as cost and severity, in addition to defect counts. SigmaXL enables you to chart multiple Y responses.

Basic Pareto Chart Template

Click **SigmaXL > Templates and Calculators** > **Basic Graphical Templates > Pareto Chart** or **SigmaXL > Graphical Tools > Basic Graphical Templates > Pareto Chart**. See **Measure Phase Part B – Templates and Calculators** for a Pareto Template example.

Basic (Single) Pareto Charts

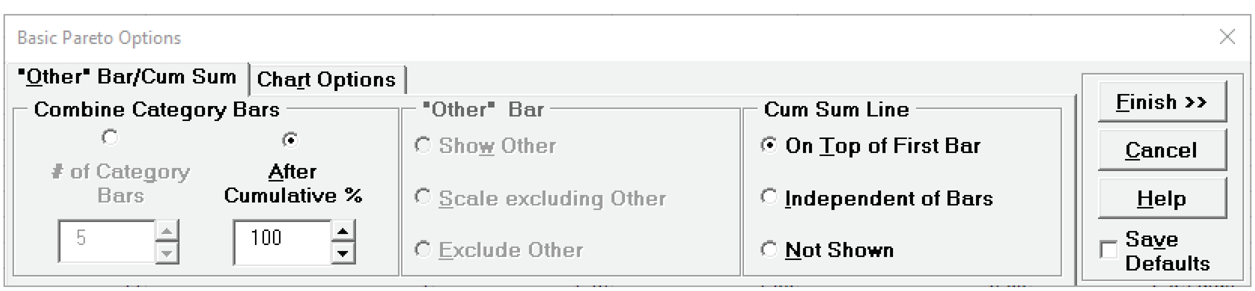
1. Open the file **Customer Data.xlsx**. Click **SigmaXL>Graphical Tools >Basic Pareto Chart**.
2. Ensure that entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.
3. Select *Major-Complaint*, click **Pareto Category (X) >>**.



Tip: SigmaXL will automatically count the number of unique items in the Pareto Category. If we had a separate column with a count (or cost), this count column would be selected as the Optional Numeric Count (Y).

4. Click **Next**. Set **Basic Chart Options** as follows:

   Tab **"Other" Bar/Cum Sum**:
   Cum Sum Line – On Top of First Bar

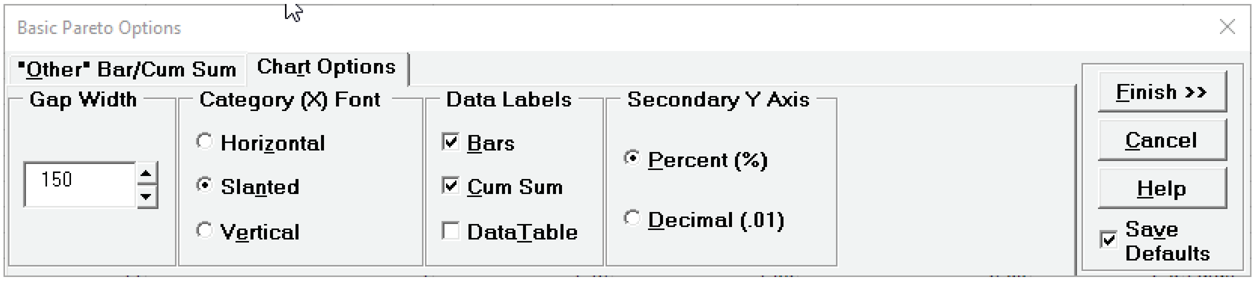


Tab **Chart Options**:

Category (X) Font – Slanted

Data Labels – Check Bars, Cum Sum

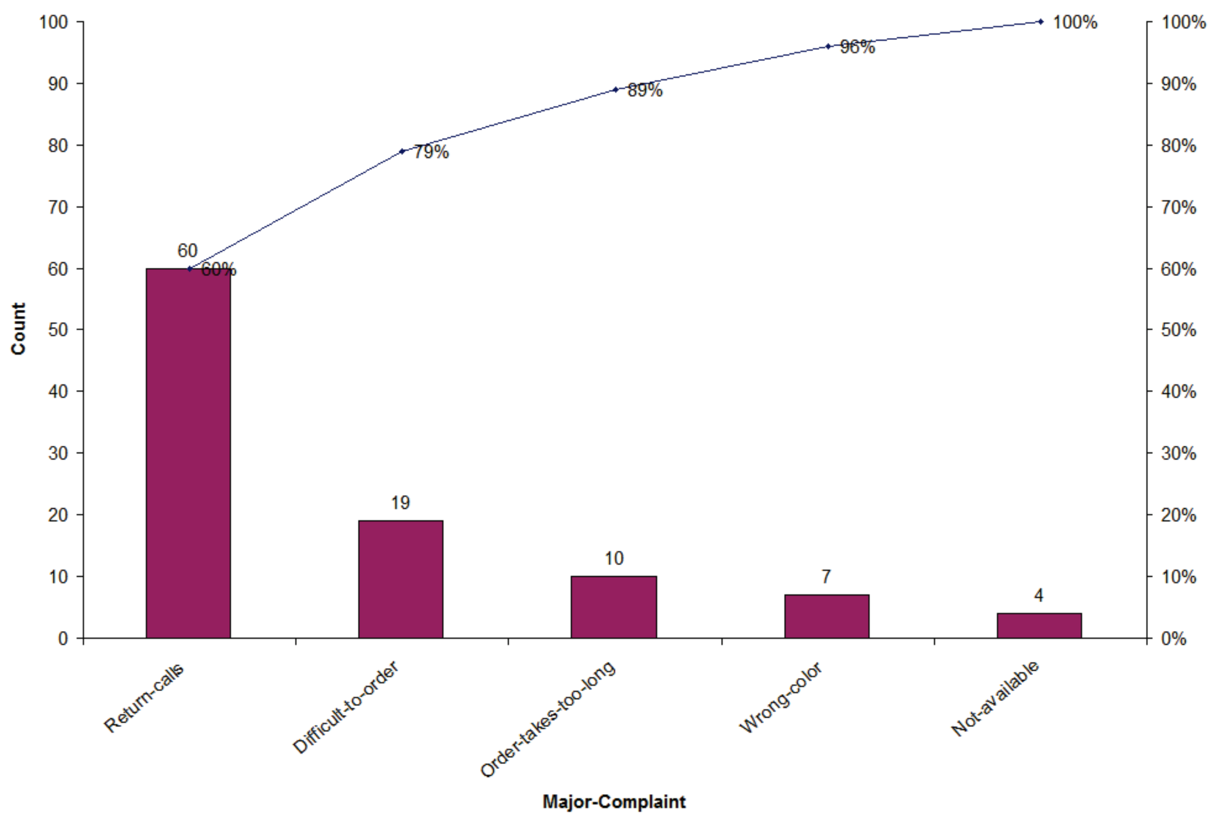
Check **Save Defaults**.



Tip: After you have saved your defaults, you can bypass the above options, by clicking **Finish** instead of **Next** at the original Basic Pareto Chart dialog box. The saved defaults will automatically be applied.

Note: For SigmaXL Mac version, click **SigmaXL > Graphical Tools > Basic Pareto Chart Options** to set options. The Basic Pareto Chart is displayed after you click **Finish**.

5. Click **Finish**. The Pareto Chart is produced:

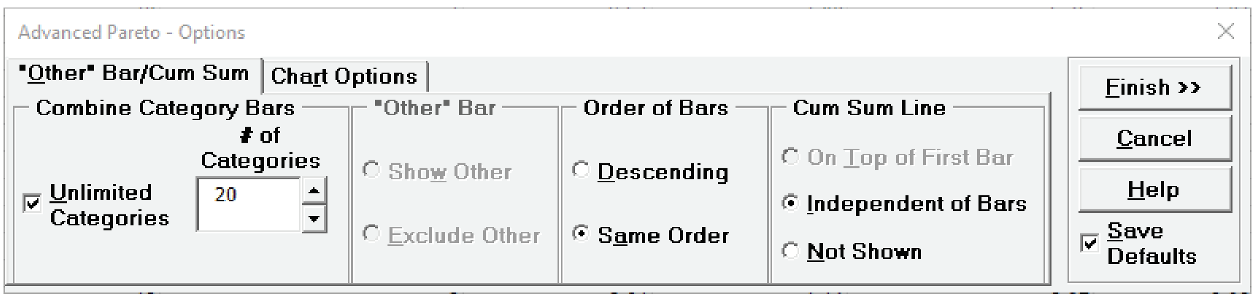


Advanced (Multiple) Pareto Charts

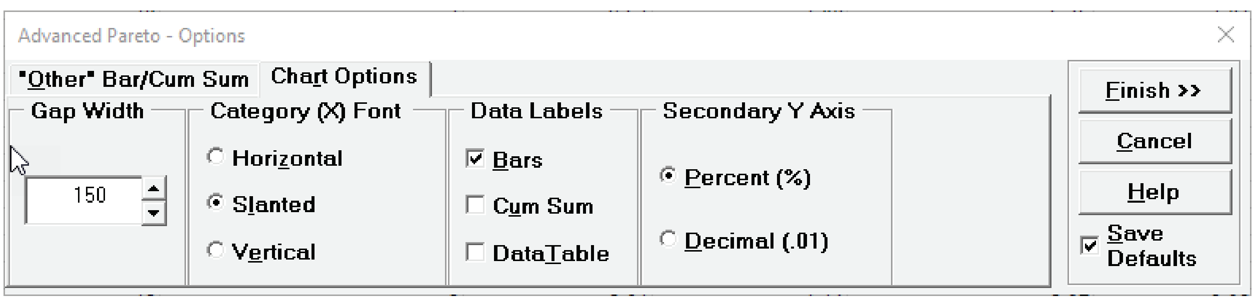
1. Click **Sheet1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Graphical Tools > Advanced Pareto Options**.
   Note that the Sample Charts have nothing to do with the data set being evaluated. They are used to dynamically illustrate how your options affect the charts to be produced. (The Sample Charts are not displayed in Excel for Mac).

3. Set **Order of Bars** to **Same Order** on the **"Other Bar"/Cum Sum** options tab.
   This is typically used for comparative purposes. The **Descending Order** option makes each Chart a true Pareto Chart, but is less useful for comparison.



4. Click **Chart Options** tab. Set according to choice – in this case we have selected **Data Labels** for the **Bars** but not for the **Cum Sum** line.

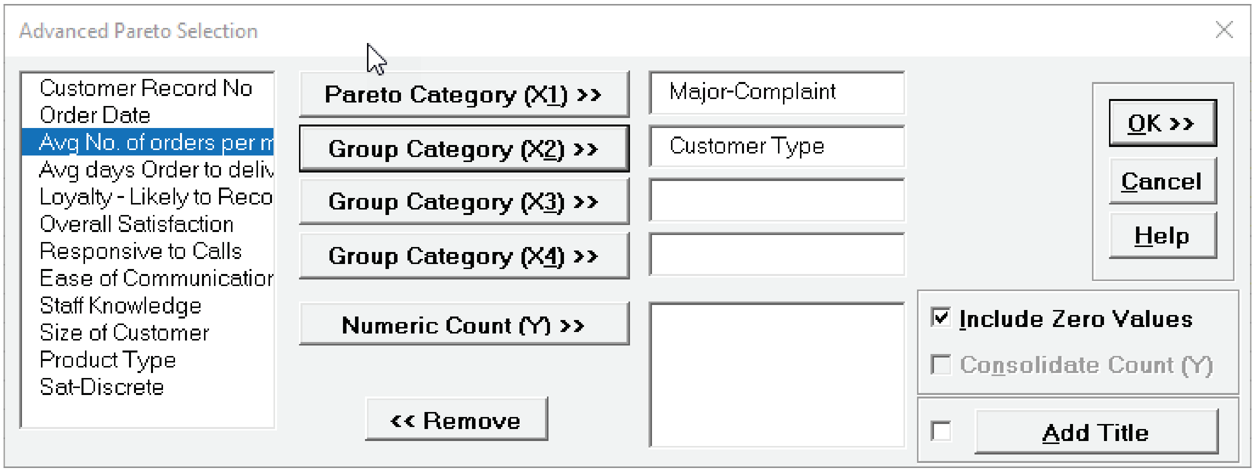


5. Ensure that **Save Defaults** is checked. Note that these options will be saved and applied to all Advanced Pareto Charts. Click **Finish**

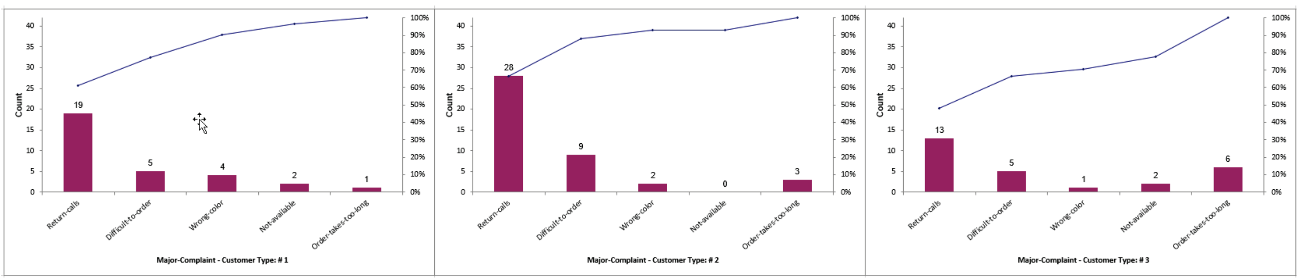
6. SigmaXL automatically takes you to the next step of Chart Generation (This is equivalent to clicking **SigmaXL > Graphical Tools > Advanced Pareto Charts**). If necessary, check **Use Entire Data Table**.

7. Click **Next**.

8. Select *Major Complaint*, click **Pareto Category (X1) >>**; select *Customer Type*, click **Group Category** (**X2) >>**.



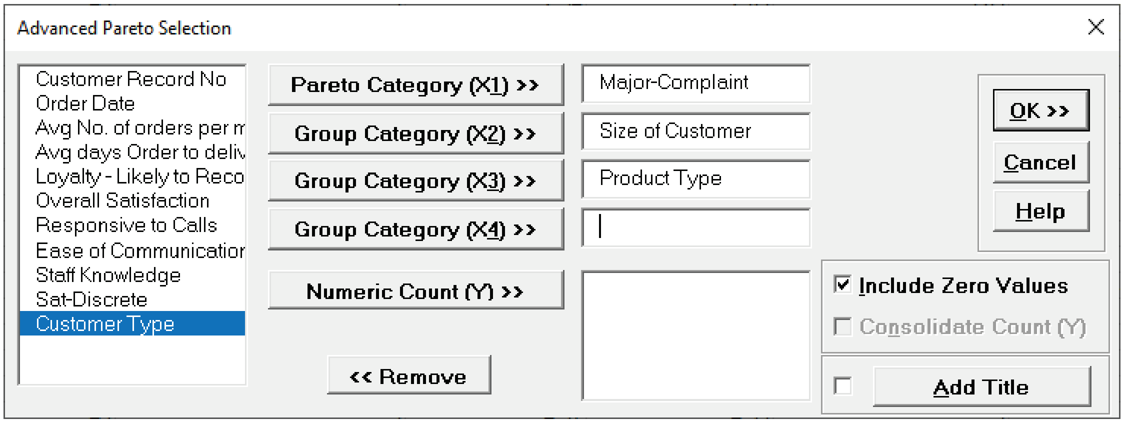
9. Click **OK**. A Pareto Chart of Major Customer Complaints is produced for each Customer Type.



10. Click **Sheet 1** Tab, Click **SigmaXL > Graphical Tools > Advanced Pareto Charts**.

11. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**. (Steps 10 and 11 can be bypassed with the **Recall SigmaXL Dialog** menu or by pressing **F3** to recall last dialog).

12. Select *Major Complaint*, click **Pareto Category (X1) >>**; select *Size of Customer*, click **Group Category (X2) >>**; select *Product Type*, click **Group Category (X3) >>**.

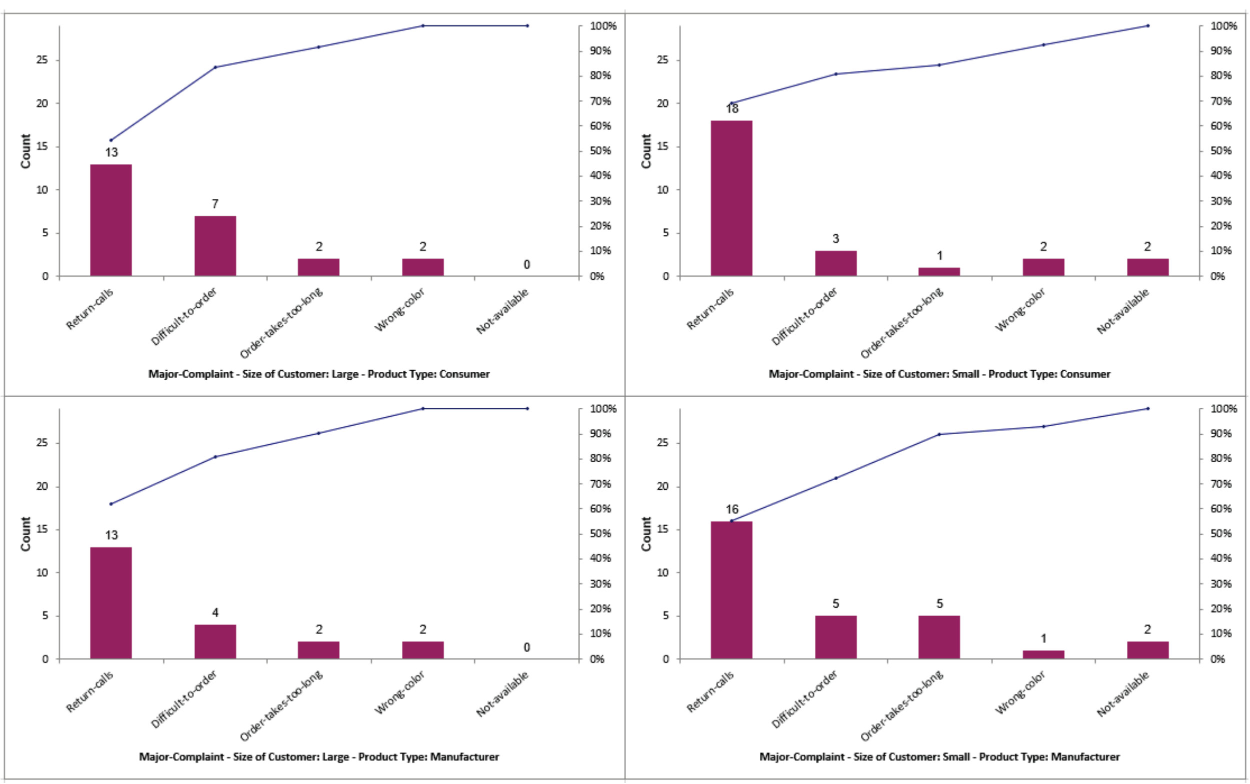


If a **Numeric Count (Y)** variable is not specified, SigmaXL automatically determines the counts from the **Pareto Category (X1)**.

Normally we would use a text column of discrete Xs, but be aware that numeric columns are also allowed. **Be careful here – this could easily generate a very large number of charts.**

The total number of charts generated = (# of levels in X2) * (# of levels in X3) * (# of levels in X4) * (# of Y variables). Please note, Group Category (X3) and Group Category (X4)

13. Click **OK**. Multiple Paretos are generated:



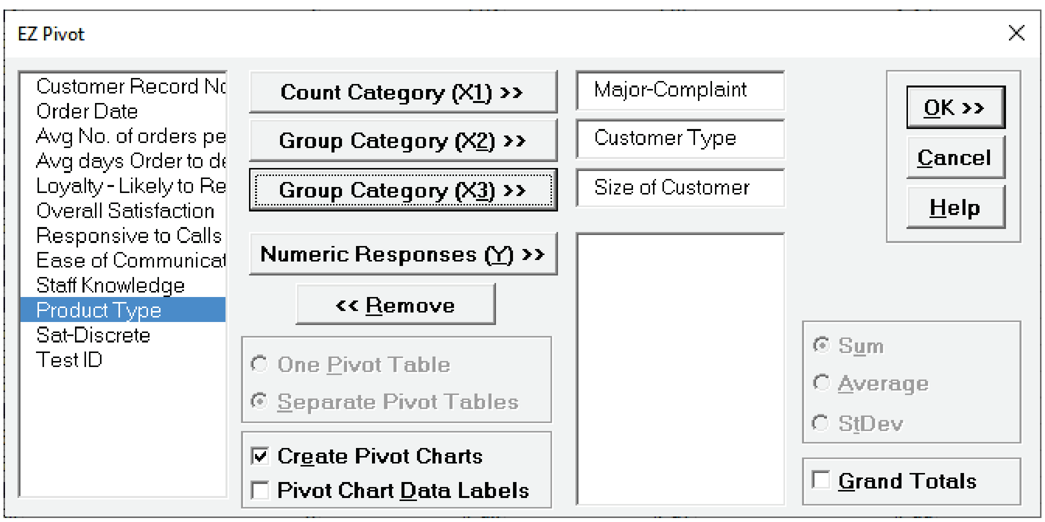
Part B - EZ-Pivot and Heatmap

EZ-Pivot/Pivot Charts

One of the most powerful features in Excel is the Pivot table. SigmaXL's EZ-Pivot tool simplifies the creation of Pivot tables using the familiar X and Y dialog box found in the previous Pareto tools.

Example of Three X's, No Response Y's

1. Open **Customer Data.xlsx**, click **Sheet 1** (or press **F4** to activate last worksheet). Select **SigmaXL > Graphical Tools > EZ-Pivot/Pivot Charts**.
2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
3. Select *Major Complaint*, click **Count Category (X1) >>**. Note that if Y is not specified, the Pivot Table Data is based on a count of X1, hence the name **Count Category**.
4. Select *Customer Type*, click **Group Category (X2) >>**; select *Size of Customer*, click **Group Category** **(X3) >>** as shown.

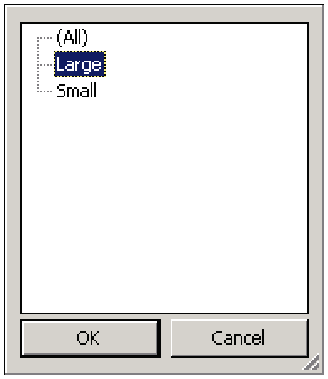


5. Click **OK**. Resulting Pivot Table of Major Complaint by Customer Type is shown:

Size of Customer	(All)		
Count of Major-Complaint	Customer Type		
Major-Complaint	1	2	3
Difficult-to-order	5	9	5
Not-available	2		2
Order-takes-too-long	1	3	6
Return-calls	19	28	13
Wrong-color	4	2	1

6. This Pivot table shows the counts for each Major Complaint (X1), broken out by Customer Type (X2), for all Sizes of Customers (X3). (Grand Totals can be added to the Pivot Table by using **Pivot Table Toolbar > Table Options**. Check **Grand Totals for Columns**, **Grand Totals for Rows**).

7. To display counts for a specific Customer Size, click the arrow adjacent to **Size of Customer (All)**. Select **Large**.

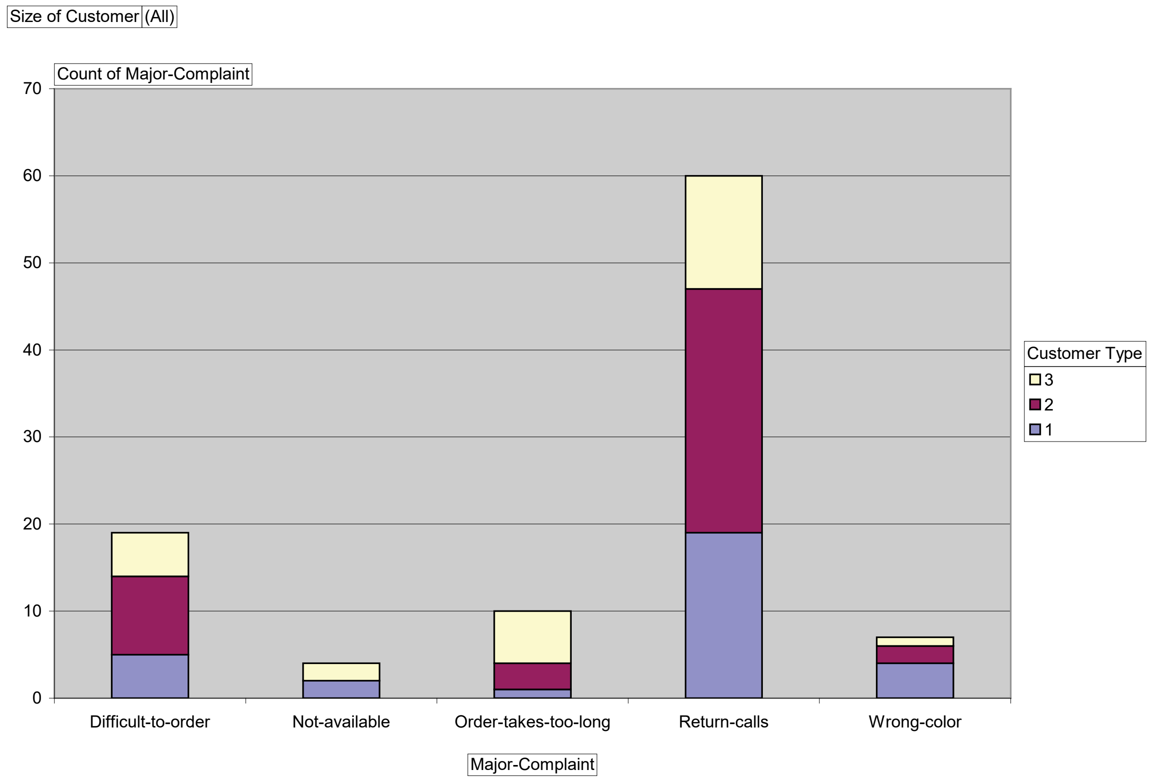


8. Click **OK**. Resulting Pivot Table is:

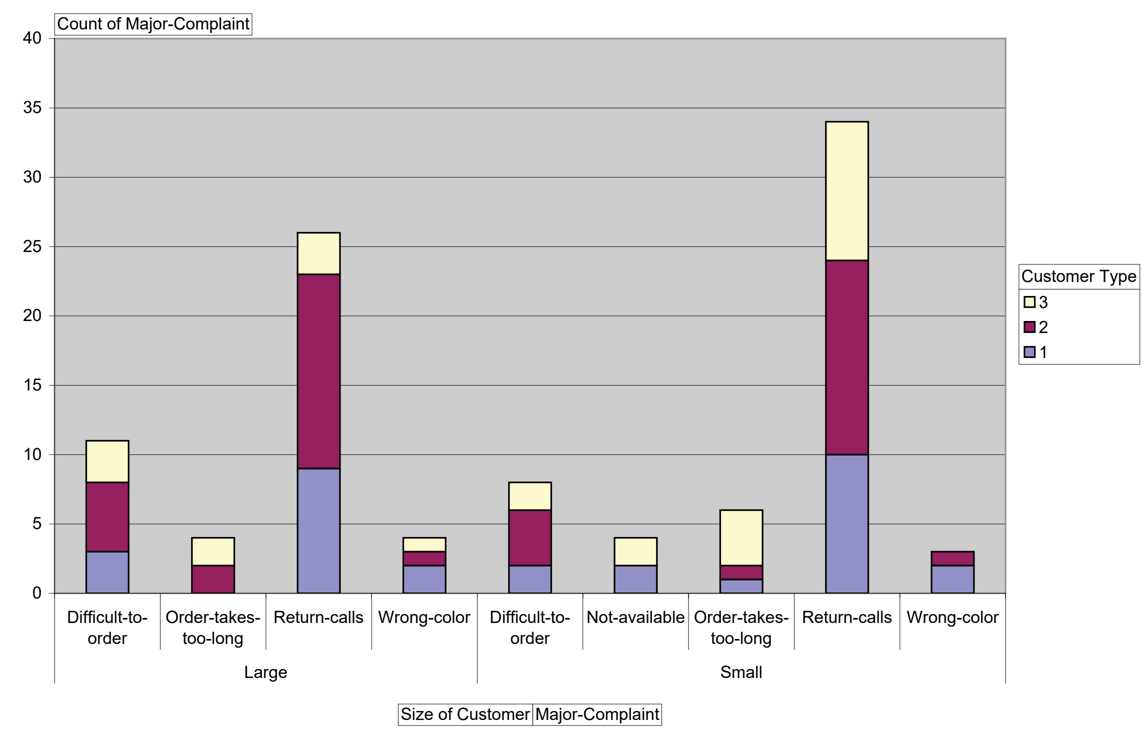
Size of Customer	Large		
Count of Major-Complaint	Customer Type		
Major-Complaint	1	2	3
Difficult-to-order	3	5	3
Order-takes-too-long		2	2
Return-calls	9	14	3
Wrong-color	2	1	1

Note that the Major Complaint "Not-Available" is not shown. Pivot tables only show rows where there is at least a count of one.

9. The Pivot Chart can be seen by clicking the **EZ Pivot Chart (1)** tab; reset **Size of Customer to All** as shown below:



10. Drag **Size of Customer** from **Report Filter** to **Axis Fields** (shown under **PivotChart Fields** to the right of the chart) and Excel will automatically split the Pivot Chart showing both Large and Small Customers.

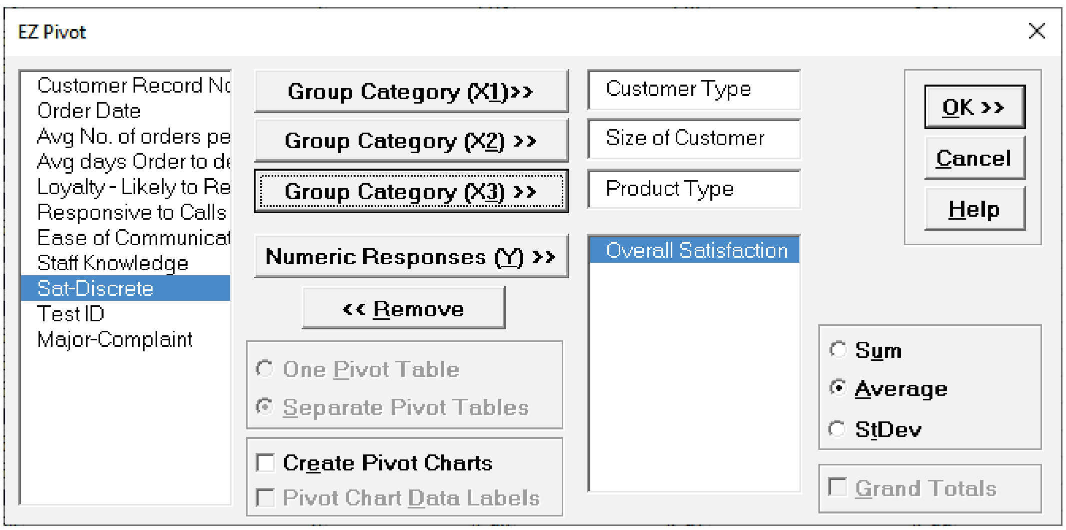


Example of Three X's and One Y

1. Select **Sheet 1** of **Customer Data.xlsx**; click **SigmaXL > Graphical Tools > EZ-Pivot/Pivot Charts**; click **Next** (alternatively, click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog).

2. Select *Customer Type*, click **Count Category (X1) >>**; select *Size of Customer*, click **Group Category (X2) >>**; select *Product Type*, click **Group Category (X3)**; select *Overall Satisfaction*, click **Numeric Responses (Y) >>**. Note that the Label for **X1** changed from **Count Category** to **Group Category**. The Pivot Table data will now be based on Y data.

3. The Response default uses a Sum of Y. This however can be changed to Average or Standard Deviation. Select **Average**. Uncheck **Create Pivot Charts** (Since we are looking at averages, the stacked bar Pivot Charts would not be very useful, unless they are changed to **clustered column** format using **Chart > Chart Type**).



4. Click **OK**. The resulting Pivot Table is:

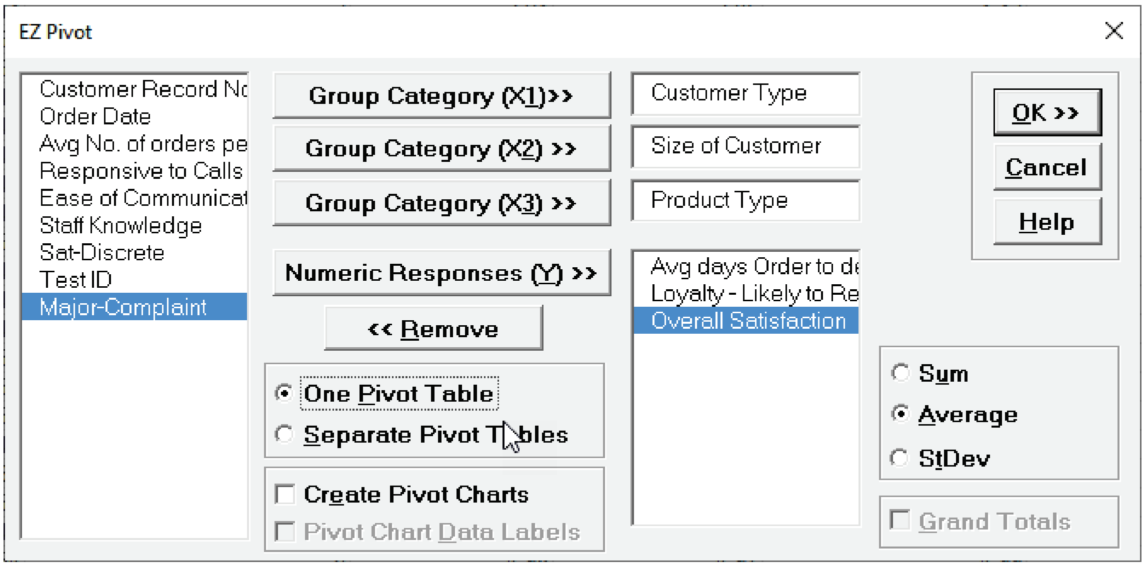
Product Type	(All)	
Average of Overall Satisfaction	Size of Customer	
Customer Type	Large	Small
1	3.237857143	3.521764706
2	4.309090909	4.091
3	3.56	3.681666667

5. Note that the table now contains Averages of the Customer Satisfaction scores (Y). Again, Product Type (X3) can be varied to show Consumer, Manufacturer, or All. Double clicking on **Average of Overall Satisfaction** allows you to switch to Standard Deviation (StDev).

Example of 3 X's and 3 Y's

1. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.

2. Select *Customer Type*, click **Group Category (X1) >>**; select *Size of Customer,* click **Group Category (X2) >>**; select *Product Type,* click **Group Category (X3) >>.** Select Avg *Days Order to Delivery*, *Loyalty – Likely to Respond*, *Overall Satisfaction,* click **Numeric Responses (Y) >>**. Select **Average** and **One Pivot Table** (default is separate Pivot Tables for each Y). Uncheck **Create Pivot Charts**.



3. Click **OK**. Resulting Pivot Table:

Product Type	(All)		
		Size of Customer	
Customer Type	Data	Large	Small
1	Average of Avg days Order to delivery time	50.85714286	49.76470588
	Average of Loyalty - Likely to Recommend	2.857142857	3.470588235
	Average of Overall Satisfaction	3.237857143	3.521764706
2	Average of Avg days Order to delivery time	48.63636364	49.65
	Average of Loyalty - Likely to Recommend	3.818181818	3.85
	Average of Overall Satisfaction	4.309090909	4.091
3	Average of Avg days Order to delivery time	47	47.66666667
	Average of Loyalty - Likely to Recommend	3.333333333	3.5
	Average of Overall Satisfaction	3.56	3.681666667

Again, Product Type (X3) can be varied.

Heatmap

Use the Heatmap graphical tool to display counts or summary statistics in Pivot table format with results gradient color coded: minimum is dark blue to maximum dark red. This allows you to easily 'slice and dice' your data, quickly look at different X factors and their contribution to the total or summary statistics (typically the mean), aided by the color coding. Up to Three Row Categories and Three Column Categories are permitted.

If an Optional Numeric Response is not specified, SigmaXL will use Counts as the statistic to display. If a Numeric Response is specified the following statistics are available:

- Mean
- Sum
- Median
- Percentile
- Standard Deviation
- Minimum
- Maximum
- Range
- Interquartile Range
- Percent >= value 1 and <= value 2
- Percent < value 1 or > value 2
- Percent = specified value
- Percent <> specified value
- Percent <= specified value
- Percent < specified value
- Percent >= specified value
- Percent > specified value
- Count

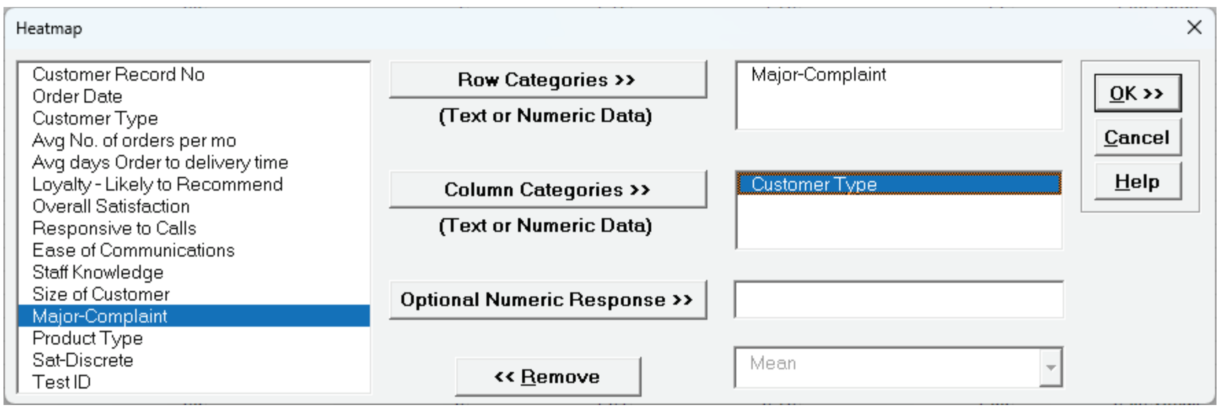
This provides a much more versatile set of descriptive statistics than are available in SigmaXL's EZ-Pivot or Excel's Pivot table.

Example: Customer Data – No Response

1. We will use the Heatmap tool to analyze the Customer Data as done with EZ-Pivot Example of Three X's, No Response Y's. Open **Customer Data.xlsx**, click **Sheet 1** (or press **F4** to activate last worksheet). Select **SigmaXL > Graphical Tools > Heatmap**.

2. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Major Complaint*, click **Row Categories >>**. Select *Customer Type*, click **Column Categories >>.** Note that if **Optional Numeric Response** is not specified, the Heatmap Table Data is based on counts.



4. Click **OK**. The resulting Heatmap of Count for Major Complaint by Customer Type is shown:

Heatmap: Count				
		Customer Type		
		1	2	3
Major-Complaint	Difficult-to-order	5	9	5
	Not-available	2	0	2
	Order-takes-too-long	1	3	6
	Return-calls	19	28	13
	Wrong-color	4	2	1

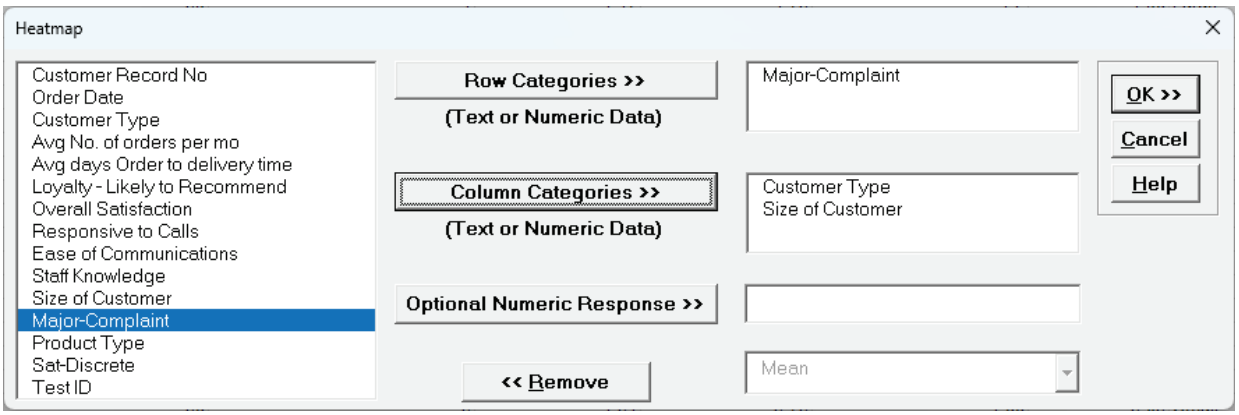
Dark red highlights the maximum count; dark blue highlights the minimum count.

Tip: Heatmap colors may be changed using Excel's Conditional Formatting: Select the data excluding Row and Column headers, click **Home > Conditional Formatting > Color Scales**.

Note that this matches the Pivot Table given in the example:

Size of Customer	(All)		
Count of Major-Complaint	Customer Type		
Major-Complaint	1	2	3
Difficult-to-order	5	9	5
Not-available	2		2
Order-takes-too-long	1	3	6
Return-calls	19	28	13
Wrong-color	4	2	1

5. Now we will add *Size of Customer* to the Column Categories. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select *Size of Customer*, click **Column Categories >>.**



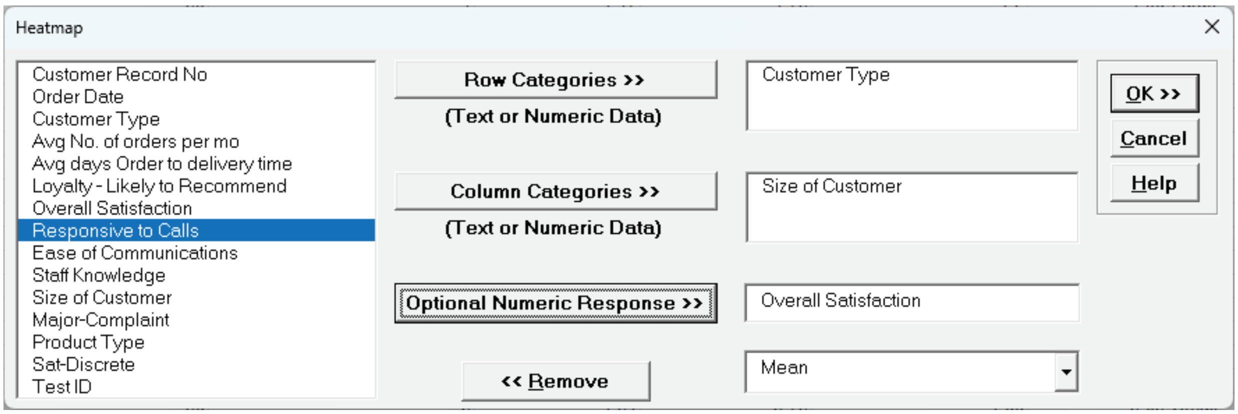
6. Click **OK**. The resulting Heatmap of Count for Major Complaint by Customer Type/Size of Customer is shown:

		Heatmap: Count					
		Customer Type/Size of Customer					
		1		2		3	
		Large	Small	Large	Small	Large	Small
Major-Complaint	Difficult-to-order	3	2	5	4	3	2
	Not-available	0	2	0	0	0	2
	Order-takes-too-long	0	1	2	1	2	4
	Return-calls	9	10	14	14	3	10
	Wrong-color	2	2	1	1	1	0

Example: Customer Data – Mean of Overall Satisfaction

1. We will now use the Heatmap tool to analyze the Customer Data as done with EZ-Pivot Example of Three X's and One Y. Select **Sheet 1** of **Customer Data.xlsx**; click **SigmaXL > Graphical Tools > Heatmap**; click **Next** (alternatively, click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog).

2. Select *Customer Type*, click **Row Categories >>**; select *Size of Customer*, click **Column Categories >>**; select *Overall Satisfaction*, click **Optional Numeric Response >>**. Use the default statistic *Mean* as shown:



3. Click **OK**. The resulting Heatmap of Mean of Overall Satisfaction by Customer Type and Size of Customer is shown:

Heatmap: Mean of Overall Satisfaction			
		Size of Customer	
		Large	Small
Customer Type	1	3.237857143	3.521764706
	2	4.309090909	4.091
	3	3.56	3.681666667

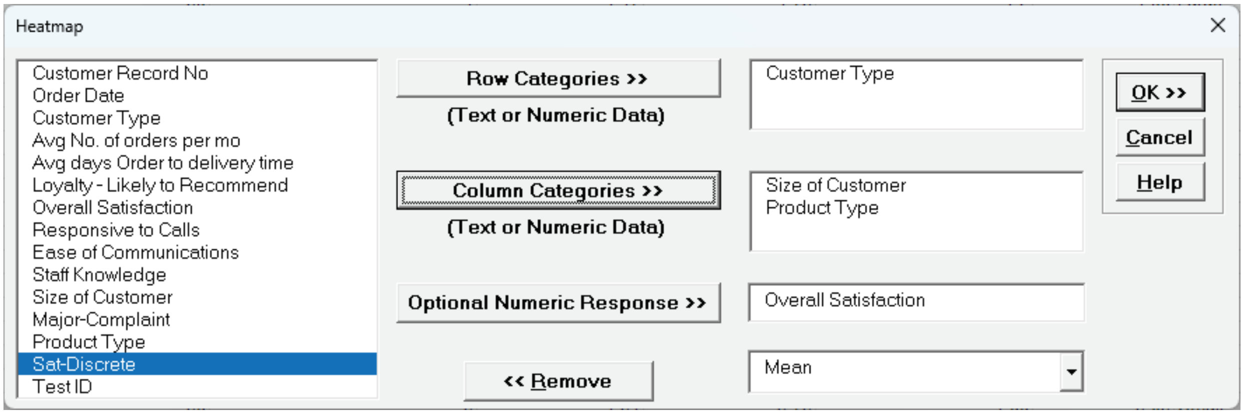
Dark red highlights the maximum mean; dark blue highlights the minimum mean.

Tip: Heatmap colors may be changed using Excel's Conditional Formatting: Select the data excluding Row and Column headers, click **Home > Conditional Formatting > Color Scales**.

Note that this matches the Pivot Table given in the example:

Product Type	(All)	
Average of Overall Satisfaction	Size of Customer	
Customer Type	Large	Small
1	3.237857143	3.521764706
2	4.309090909	4.091
3	3.56	3.681666667

4. Now we will add *Product Type* to the Column Categories. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select *Product Type*, click **Column Categories >>.**



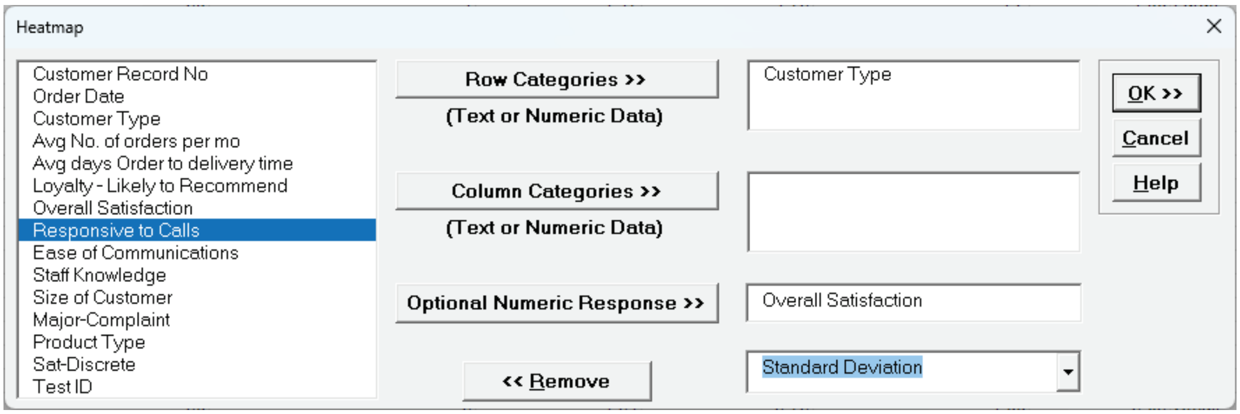
5. Click **OK**. The resulting Heatmap of Mean of Overall Satisfaction by Customer Type and Size of Customer/Product Type is shown:

Heatmap: Mean of Overall Satisfaction					
		Size of Customer/Product Type			
		Large		Small	
		Consumer	Manufacturer	Consumer	Manufacturer
Customer Type	1	3.253333333	3.22625	3.631818182	3.32
	2	4.484545455	4.133636364	4.107142857	4.082307692
	3	3.511428571	3.73	3.5525	3.785

Example: Customer Data – Other Statistics

1. We will now use the Heatmap tool to analyze other statistics for Overall Satisfaction by Customer Type. Select **Sheet 1** of **Customer Data.xlsx**; click **SigmaXL > Graphical Tools > Heatmap**; click **Next** (alternatively, click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog).

2. Select *Customer Type*, click **Row Categories >>**; select *Overall Satisfaction*, click **Optional Numeric Response >>**. Select the statistic *Standard Deviation* as shown:

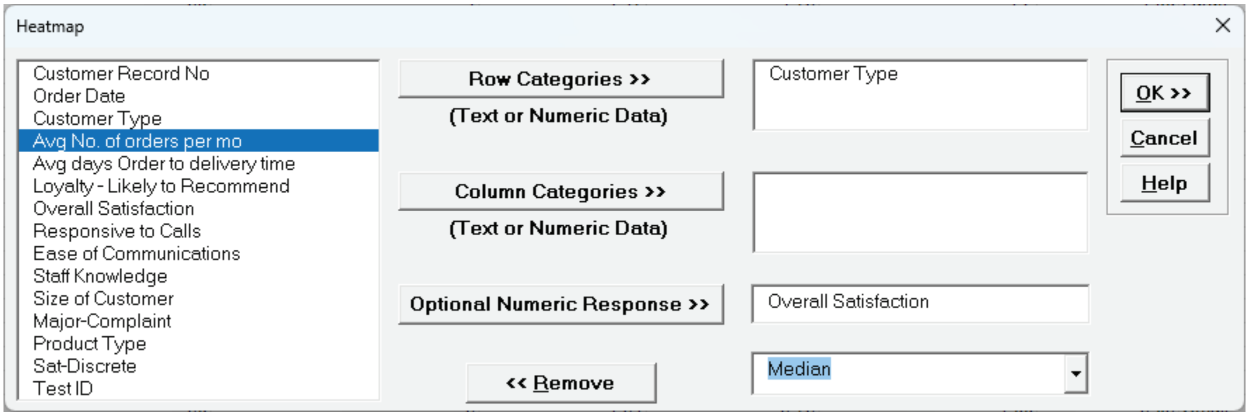


3. Click **OK**. The resulting Heatmap of Standard Deviation of Overall Satisfaction by Customer Type is shown:

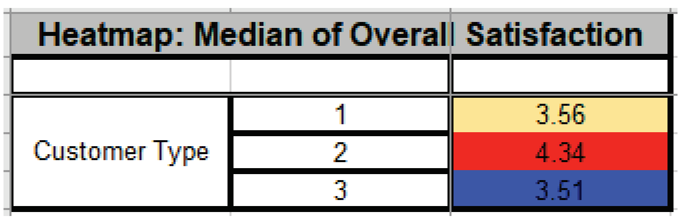
Heatmap: Standard Deviation of Overall Satisfaction		
Customer Type	1	0.824679931
	2	0.621199619
	3	0.670478207

Dark red highlights the maximum standard deviation; dark blue highlights the minimum standard deviation.

4. Now we will analyze the data using Medians. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select the statistic *Median* as shown:



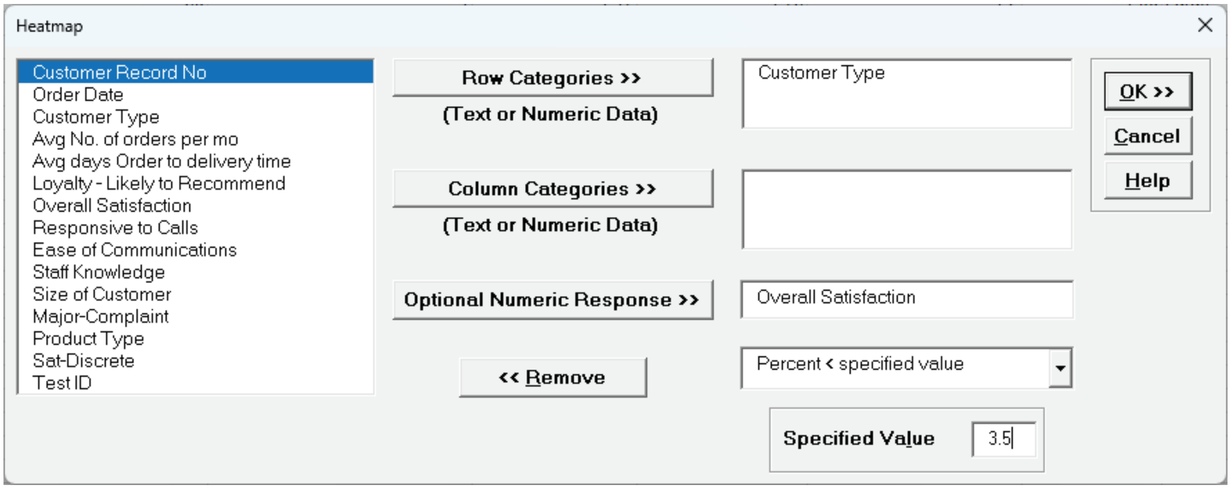
5. Click **OK**. The resulting Heatmap: Median of Overall Satisfaction by Customer Type is shown:



Heatmap: Median of Overall Satisfaction		
Customer Type	1	3.56
	2	4.34
	3	3.51

Dark red highlights the maximum median; dark blue highlights the minimum median.

6. Next we will analyze Percent with Overall Satisfaction < 3.5 (i.e., percent dissatisfied customers). Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select the statistic *Percent < specified value*; enter 3.5 for **Specified Value** as shown:



7. Click **OK**. The resulting Heatmap of Percent < 3.5 for Overall Satisfaction by Customer Type is shown:

Heatmap: Percent < 3.5 for Overall Satisfaction		
Customer Type	1	48.38709677
	2	11.9047619
	3	48.14814815

Dark red highlights the maximum percent; dark blue highlights the minimum percent. Since this is percent dissatisfied customers, lower is better.

Part C – Confidence Intervals

One Sample Confidence Interval Templates

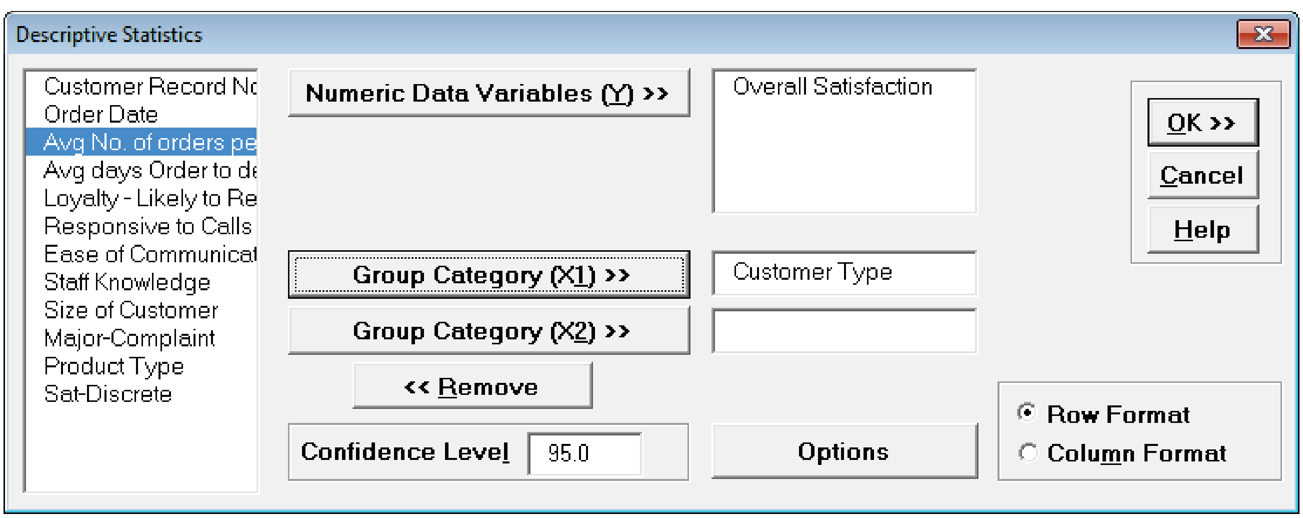
Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
1 Sample Z-Test and Confidence Interval for Mean or
1 Sample t-Test and Confidence Interval for Mean or
1 Sample Chi-Square Test and CI for Standard Deviation or
1 Proportion Test and Confidence Interval or
1 Poisson Rate Test and Confidence Interval.

These confidence interval templates are also located at **SigmaXL > Statistical Tools > Basic Statistical Templates**.

See **Measure Phase Part B – Templates and Calculators** for Confidence Interval template examples.

Confidence Intervals – Descriptive Statistics

1. Open **Customer Data.xlsx.** Click **Sheet 1** Tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Descriptive Statistics**.
3. Check **Use Entire Data Table**, click **Next**.
4. Select *Overall Satisfaction*, click **Numeric Data Variables (Y) >>**. Select *Customer Type*, click **Group Category (X1) >>**; **Confidence Level** default is 95%:



5. Click **OK**. Descriptive Statistics are given for Customer Satisfaction grouped by Customer Type:

Overall Satisfaction by Customer Type			
Descriptive Statistics	**Customer Type = 1**	**Customer Type = 2**	**Customer Type = 3**
Count	31	42	27
Mean	3.394	4.205	3.641
StDev	0.824680	0.621200	0.670478
Range	3.080	2.560	2.740
Minimum	1.720	2.420	2.190
25th Percentile (Q1)	2.810	3.828	3.240
50th Percentile (Median)	3.560	4.340	3.510
75th Percentile (Q3)	4.020	4.725	4.170
Maximum	4.800	4.980	4.930
95.0% CI Mean	3.0911 to 3.696	4.0117 to 4.3988	3.3759 to 3.9063
95.0% CI Sigma	0.65901 to 1.1023	0.51113 to 0.79213	0.52801 to 0.91884
Anderson-Darling Normality Test	0.312776	0.826259	0.389291
P-Value (A-D Test)	0.5306	0.0302	0.3600
Skewness	-0.235169	-0.967994	0.139571
P-Value (Skewness)	0.5557	0.0121	0.7411
Kurtosis	-0.671690	0.679609	-0.313701
P-Value (Kurtosis)	0.3705	0.2865	0.8435

We are given the 95% confidence interval for each sample Mean (95% CI Mean) as well as the 95% confidence interval for the Standard Deviation (95% CI Sigma – do not confuse this with Process Sigma Quality Level).

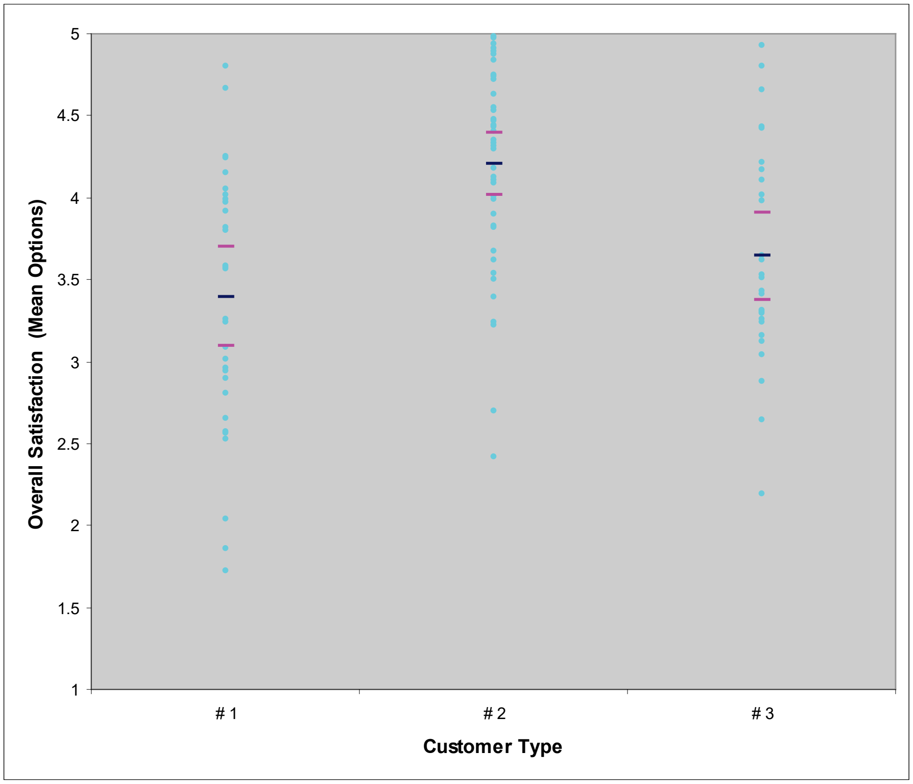
These confidence intervals are very important in understanding our data and making decisions from the data. How often are we driven by sample estimates only and fail to consider the confidence interval or margin of error? For example, newspapers will often fail to take into account the confidence interval when reporting opinion poll results. (To calculate confidence intervals for discrete proportion data, use **SigmaXL > Templates and Calculators** > **Basic Statistical Templates > 1 Proportion Test and Confidence Interval**).

Note that a confidence interval of 95% implies that, on average, the true population parameter (Mean, Median, Standard Deviation, or Proportion) will lie within the interval 19 times out of 20.

A confidence interval or margin of error does not take into account measurement error or survey bias, so the actual uncertainty may be greater than stated. This should be addressed with good data collection, reliable measurement systems, and good survey design.

Confidence Intervals for the Mean can be obtained in several ways with SigmaXL: Descriptive Statistics, Histograms & Descriptive Statistics, 1-Sample t-test & Confidence Intervals, One-Way ANOVA, Multi-Vari Charts and Interval Plots.

To illustrate confidence intervals for the mean of *Overall Satisfaction* graphically, we have generated a Multi-Vari Chart (with 95% CI Mean Options) using the **Customer Data.xlsx** data. This chart type will be covered later (Part Q).

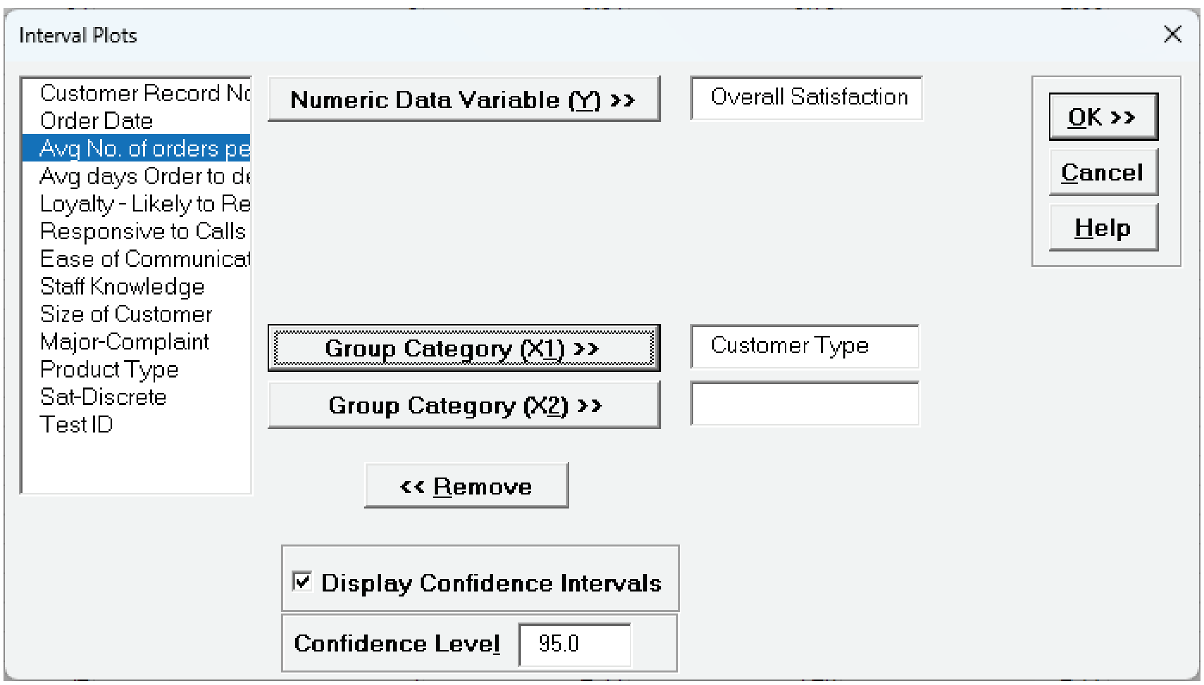


The dots correspond to individual data points. The tick marks show the 95% upper confidence limit, mean, and 95% lower confidence limit. Clearly we can see that Customer Type 2 has a significantly higher level of mean satisfaction; the lower limit does not overlap with the upper limit for Types 1 and 3. On the other hand, we see overlap of the CI's when comparing types 1 and 3. Hypothesis testing will be used to compare the mean satisfaction scores more precisely and determine statistical significance for the results.

Interval Plots

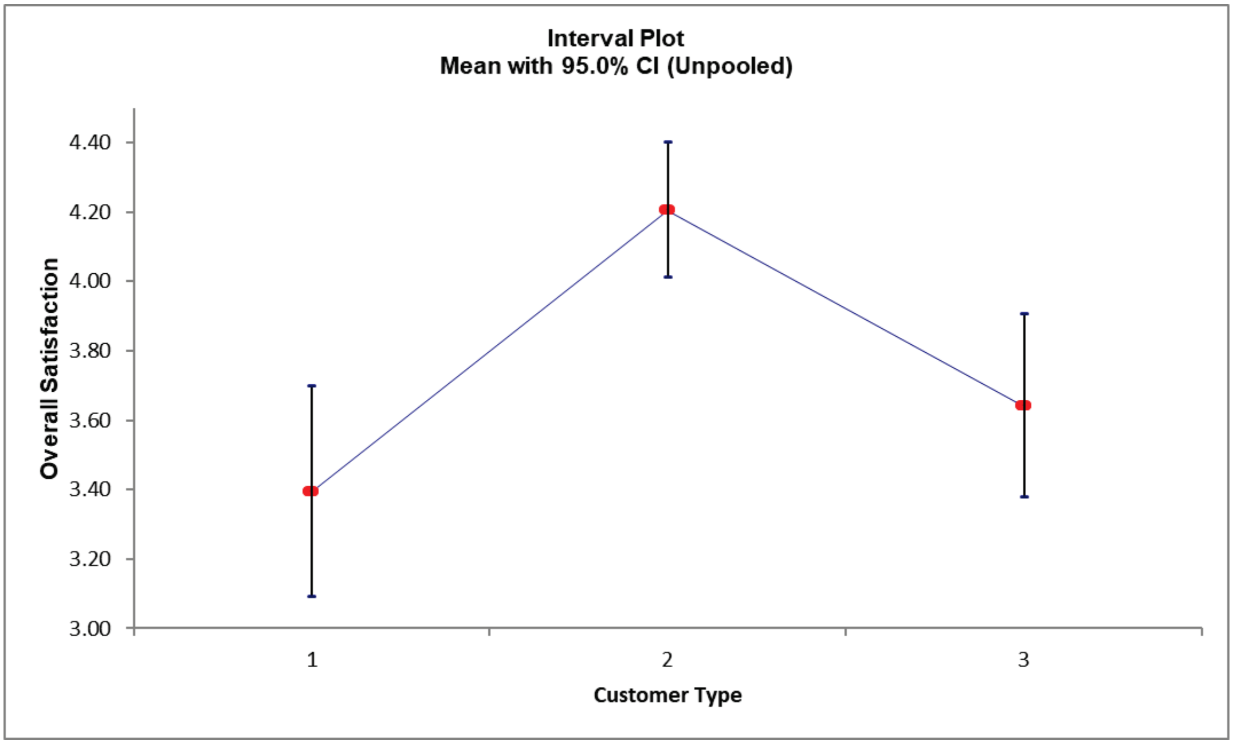
Interval Plots show the mean and confidence interval, similar to the Multi-Vari chart above, but do not display the individual data points. The confidence intervals are computed as done in Descriptive Statistics with the standard deviation calculated separately for each group, so they are labelled “Unpooled”. (Interval Plots with confidence intervals based on pooled standard deviation are given in One Way ANOVA).

1. Open **Customer Data.xlsx.** Click **Sheet 1** Tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Graphical Tools > Interval Plots**.
3. Check **Use Entire Data Table**, click **Next**.
4. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**. Select *Customer Type*, click **Group Category (X1) >>**; check **Display Confidence Intervals**; **Confidence Level** default is 95%:



Tip: Uncheck **Display Confidence Intervals** to produce a Main Effects Plot.

5. Click **OK**. The Interval Plot for Customer Satisfaction grouped by Customer Type is shown:



The tick marks show the 95% upper and lower confidence limits. The red dot shows the mean. As noted in the Multi-Vari chart, we can see that Customer Type 2 has a significantly higher level of mean satisfaction; the lower limit does not overlap with the upper limit for Types 1 and 3. On the other hand, we see overlap of the CI's when comparing types 1 and 3. Hypothesis testing will be used to compare the mean satisfaction scores more precisely and determine statistical significance for the results.

Multiple X Interval Plots

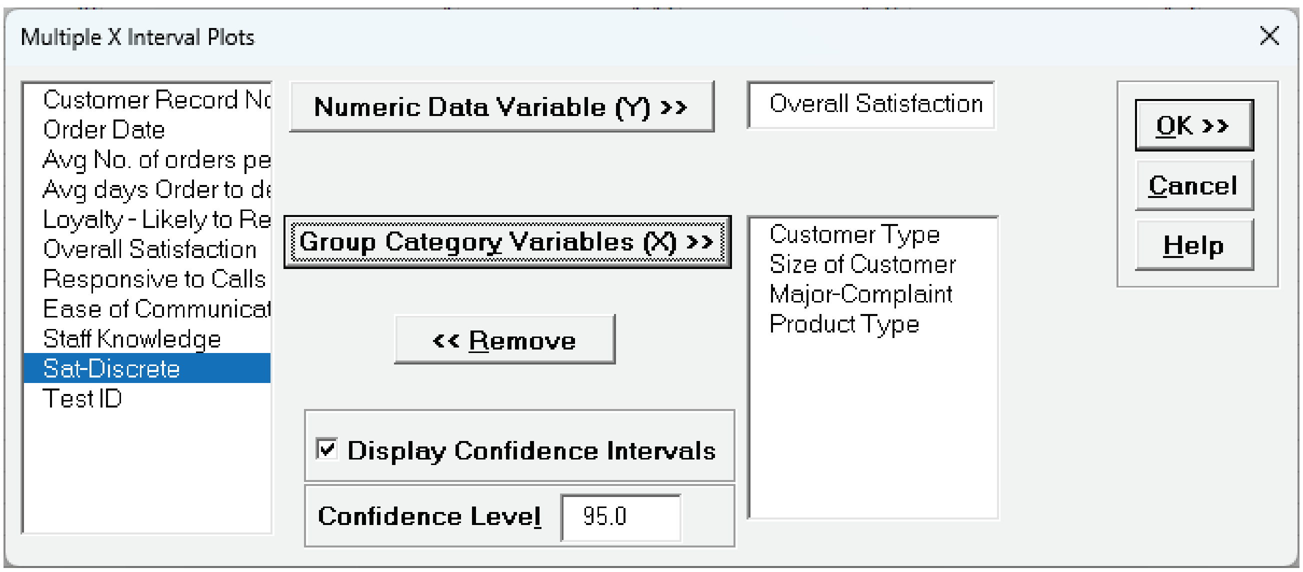
Multiple X Interval Plots are similar to Multiple X Boxplots and allow you to create interval plots with one Y variable and multiple group category X's. A row of interval plots will be created, one for each X variable. This is useful for easy comparison of the effect of each category X. Confidence Intervals may be unchecked, in which case the plots are equivalent to Main Effects plots for data means.

1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet).

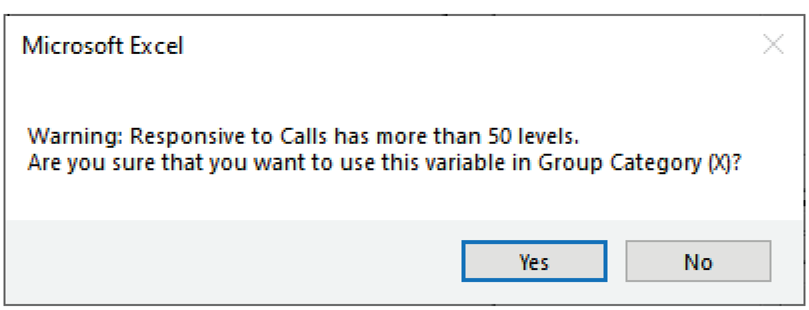
2. Click **SigmaXL > Graphical Tools > Multiple X Interval Plots**.

3. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

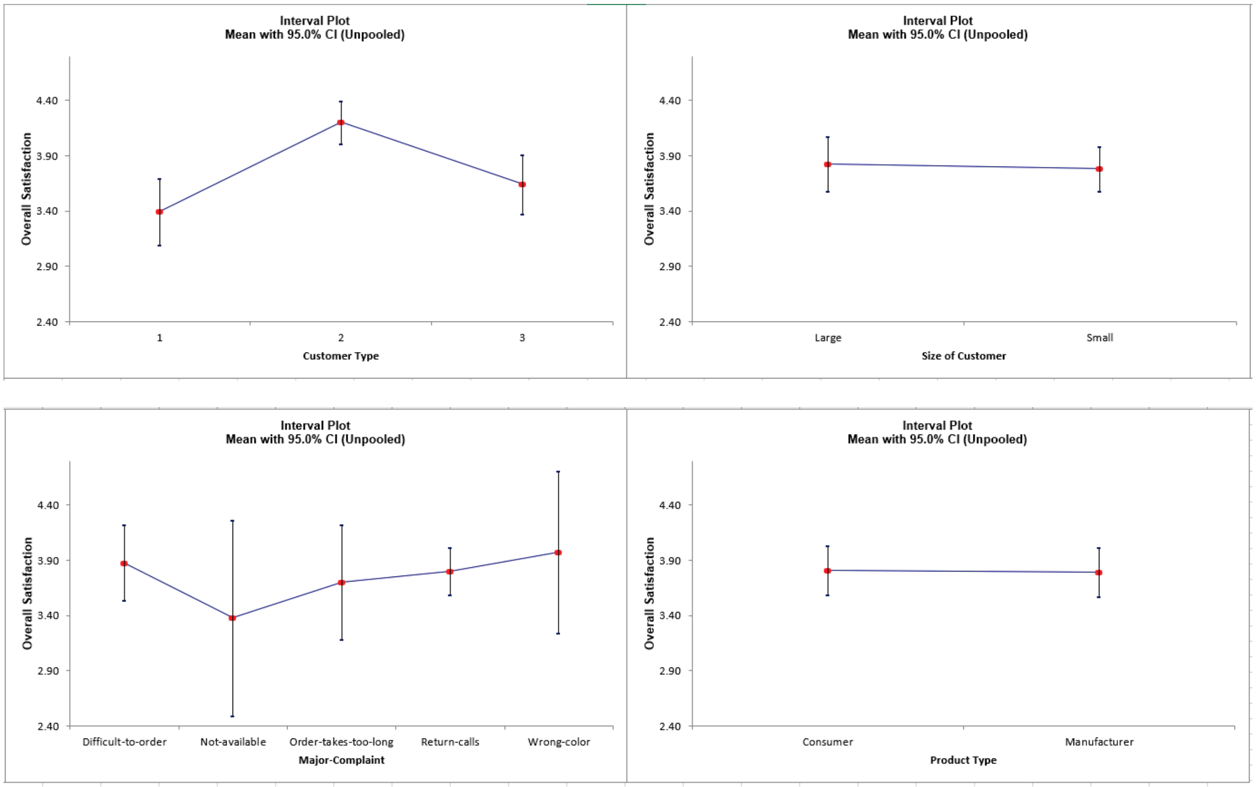
4. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**. Select *Customer Type*, *Size of Customer, Major-Complaint* and *Product Type.* Click **Group Category Variables (X) >>**. Check **Display Confidence Intervals** with **Confidence Level** 95%:



Tip: Be careful to not select a continuous variable for **Group Category Variables (X)**, as each unique value will be considered as a category level. If SigmaXL detects a variable with more than 50 levels a warning is given. For example, if *Responsive to Calls* was selected:

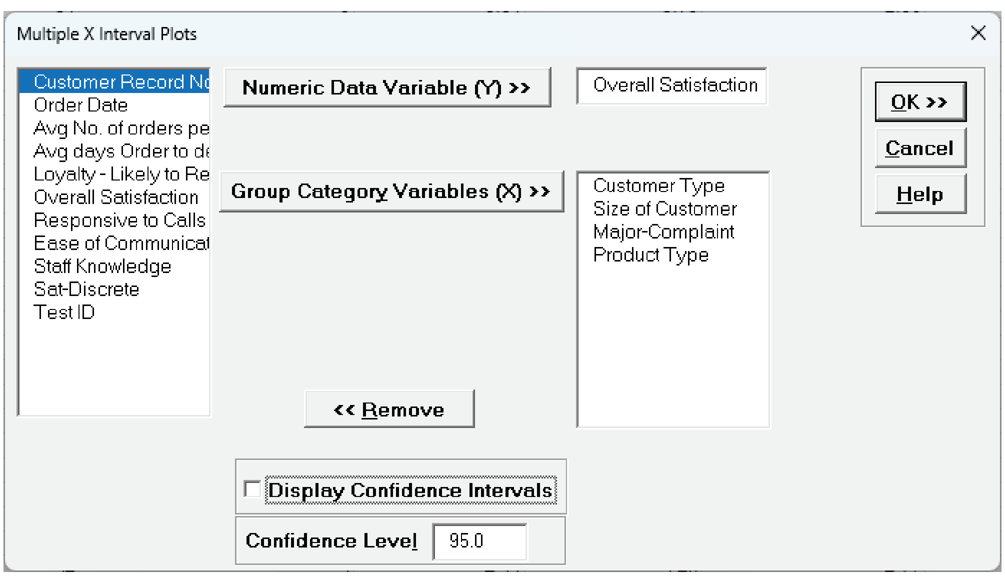


5. Click **OK**. A row of Interval Plots showing Overall Satisfaction by Customer Type, Size of Customer, Major-Complaint and Product Type are produced:

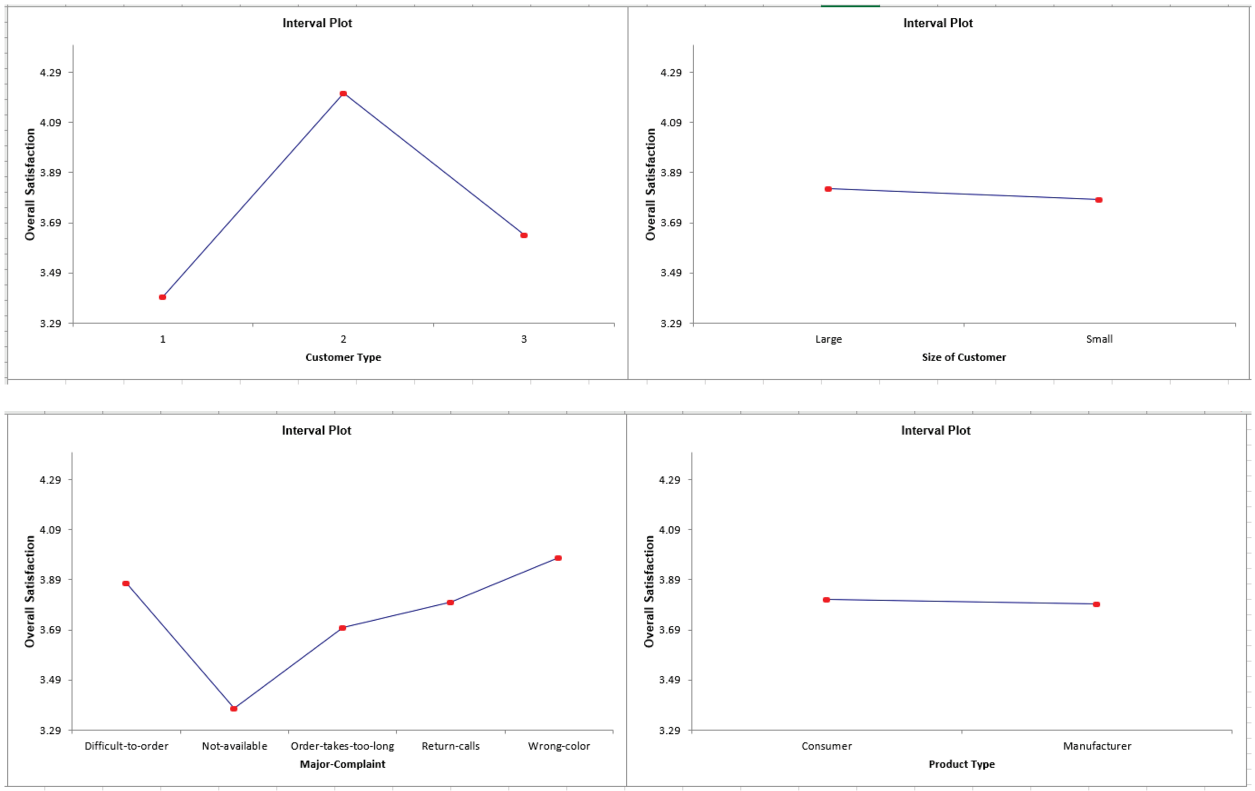


The only factor with non-overlapping confidence intervals is Customer Type. Statistical significance for these factors can be formally evaluated individually using One-Way ANOVA or as a group in General Linear Model (GLM).

6. Now we will create Interval Plots without confidence intervals. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Uncheck Display Confidence Intervals.



7. Click **OK**. A row of Interval Plots showing Overall Satisfaction by Customer Type, Size of Customer, Major-Complaint and Product Type are produced:



These are also known as Main Effects Plots for data means, with steeper slopes noted for Customer Type and Major Complaint.

Part D – Hypothesis Testing – One Sample Z and t-Test

One Sample Z, t and Equivalence Test Templates

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
1 Sample Z-Test and Confidence Interval for Mean or
1 Sample t-Test and Confidence Interval for Mean or
1 Sample Equivalence Test for Mean

These templates are also located at **SigmaXL > Statistical Tools > Basic Statistical Templates**.

See **Measure Phase Part B – Templates and Calculators** for One Sample Z, t and Equivalence Test template examples:

Basic Statistical Templates – 1 Sample Z-Test and Confidence Interval for Mean

Basic Statistical Templates – 1 Sample t-Test and Confidence Interval for Mean

Basic Statistical Templates – 1 Sample Equivalence Test for Mean

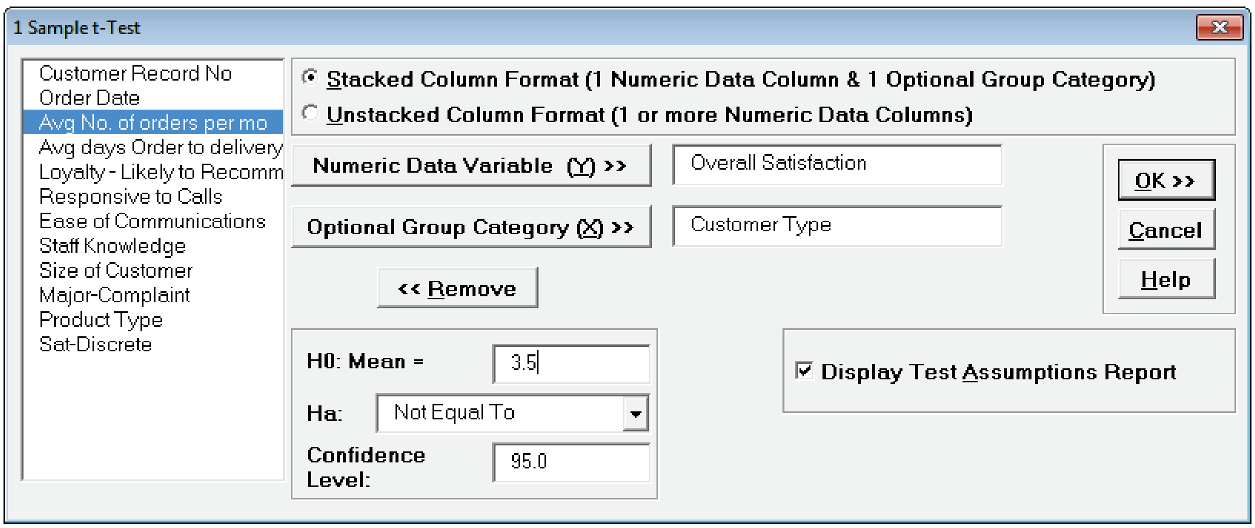
One Sample t-Test

1. Open **Customer Data.xlsx**, select **Sheet 1** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > 1 Sample t-Test & Confidence Intervals**. If necessary, check **Use Entire Data Table**, click **Next**.

2. Ensure that **Stacked Column Format** is selected. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**, select *Customer Type*, click **Optional Group Category (X) >>**.

3. Historically, our average customer satisfaction score has been 3.5. We would like to see if this has changed, with the results grouped by customer type. Null Hypothesis H0: μ=3.5; Alternative Hypothesis Ha: μ≠3.5

4. Enter 3.5 for the **Null Hypothesis H0: Mean** value. Set **Ha** as *Not Equal To*, **Confidence Level** = 95.0%, and check **Display Test Assumptions Report**.



5. Click **OK**. Results:

1 Sample t-Test - Overall Satisfaction			
Test Information			
H_0: Mean (Mu) = 3.5			
H_a: Mean (Mu) Not Equal To 3.5			
Customer Type	**1**	**2**	**3**
Count	31	42	27
Mean	3.394	4.205	3.641
StDev	0.824680	0.621200	0.670478
SE Mean	0.148117	0.095853182	0.129034
t	-0.718700	7.357	1.093599814
P-Value (2-sided)	0.4779	0.0000	0.2842
UC (2-sided, 95%)	3.696	4.399	3.906
LC (2-sided, 95%)	3.091	4.012	3.376

6. Note the P-Values. Customer Type 2 shows a significant change (increase) in Satisfaction Mean (P-Value < .05), whereas Customer Types 1 and 3 show no change (P-Value ≥ .05). Also note the confidence intervals around each mean match the results from Descriptive Statistics.

7. In the Measure Phase we determined that Overall Satisfaction for Customer Type 2 has nonnormal data but this does not imply that the P-Value for the 1 Sample t-test is wrong. The Central Limit Theorem applies here: the distribution of averages tends to be normal, even if the individual observations are not-normal. With a sample size of 42, the t-test is fairly robust against skewed data.

8. 1 Sample t-Test Assumptions Report:

1 Sample t-Test Assumptions Report			
Normality:	Anderson Darling P-Value = 0.531. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.	Anderson Darling P-Value = 0.030. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.9680 and Kurtosis value = 0.6796. See robustness and outliers.	Anderson Darling P-Value = 0.360. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.
Robustness:	Not applicable for normal data.	Minimum sample size for a robust t-test = 22. Since the sample size is greater than this, the t-test is robust to nonnormality.	Not applicable for normal data.
Outliers (Boxplot Rules):	No outliers found.	Potential (1.5*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 0.066. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.

This is a text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation).

Each sample is tested for Normality using the Anderson Darling test. If not normal, the minimum sample size for robustness of the 1 sample t-Test is determined utilizing Monte Carlo regression equations (see **Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA**). If the sample size is inadequate, a warning is given and a suitable Nonparametric Test is recommended (Wilcoxon if symmetric, Sign Test if not symmetric).

Each sample is tested for Outliers defined as: Potential: Tukey's Boxplot (> Q3 + 1.5*IQR or < Q1 – 1.5*IQR); Likely: Tukey's Boxplot 2.2*IQR; Extreme: Tukey's Boxplot 3*IQR. If outliers are present, a warning is given and recommendation to review the data with a Boxplot and Normal Probability Plot. Here we have a potential outlier for Customer Type 2.

Tip: If the removal of outlier(s) result in an Anderson Darling P-Value that is > 0.1, a notice is given that excluding the outlier(s), the sample data are inherently normal.

Each sample is also tested for Randomness using the Exact Nonparametric Runs Test. If the sample data is not random, a warning is given and recommendation to review the data with a Run Chart.

See Appendix **Hypothesis Test Assumptions Report** for further details.

Part E – Power and Sample Size

Power and Sample Size – One Sample t-Test – Customer Data

Using the One Sample t-Test (above, Part D), we determined that Customer Types 1 and 3 resulted in “Fail to reject H0: μ=3.5”. A failure to reject H0 does not mean that we have proven the null to be true. The question that we want to consider here is “What was the power of the test?” Restated, “What was the likelihood that given Ha: μ≠3.5 was true, we would have rejected H0 and accepted Ha?” To answer this, we will use the Power and Sample Size Calculator.

Tip: Typical sample size rules of thumb address confidence interval size and robustness to normality (e.g. n=30). Computing power is more difficult because it involves the magnitude of change in mean to be detected, so one needs to use the power and sample size calculator.

Please use the following guidelines when using the power and sample size calculator:

Power >= .99 (Beta Risk is <= .01) is considered Very High Power

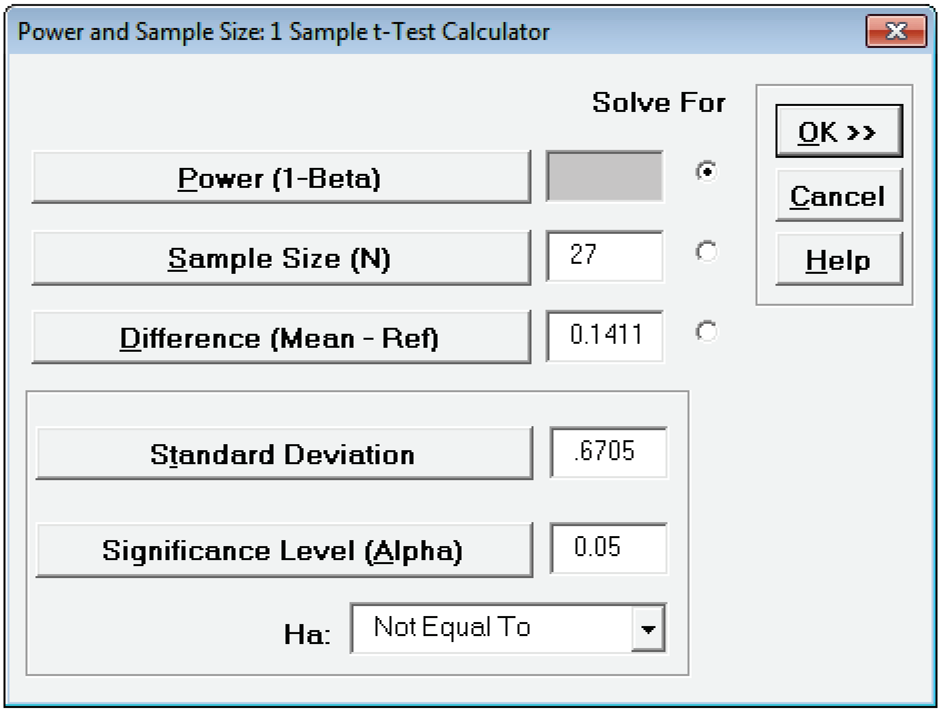
Power >= .95 and < .99 (Beta Risk is <= .05) is High Power

Power >= .8 and < .95 (Beta Risk is <= .2) is Medium Power. Typically, a power value of .9 to detect a difference of 1 standard deviation is considered adequate for most applications. If the data collection is difficult and/or expensive, then .8 might be used.

Power >= .5 and < .8 (Beta Risk is <= .5) is Low Power – not recommended.

Power < .5 (Beta Risk is > .5) is Very Low Power – do not use!

1. Click **SigmaXL > Statistical Tools > Power and Sample Size Calculators > 1 Sample t-Test Calculator**. We will only consider the statistics from Customer Type 3 here. We will treat the problem as a two-sided test with **Ha**: *Not Equal To* to be consistent with the original test.
2. Enter 27 in **Sample Size (N)**. The difference to be detected in this case would be the difference between the sample mean and the hypothesized value i.e. 3.6411 – 3.5 = 0.1411. Enter 0.1411 in **Difference**. Leave **Power** value blank, with **Solve For Power** selected (default). Given any two values of Power, Sample size, and Difference, SigmaXL will solve for the remaining selected third value. Enter the sample standard deviation value of 0.6705 in **Standard Deviation**. Keep **Alpha** and **Ha** at the default values as shown:



3. Click **OK**. The resulting report is shown:

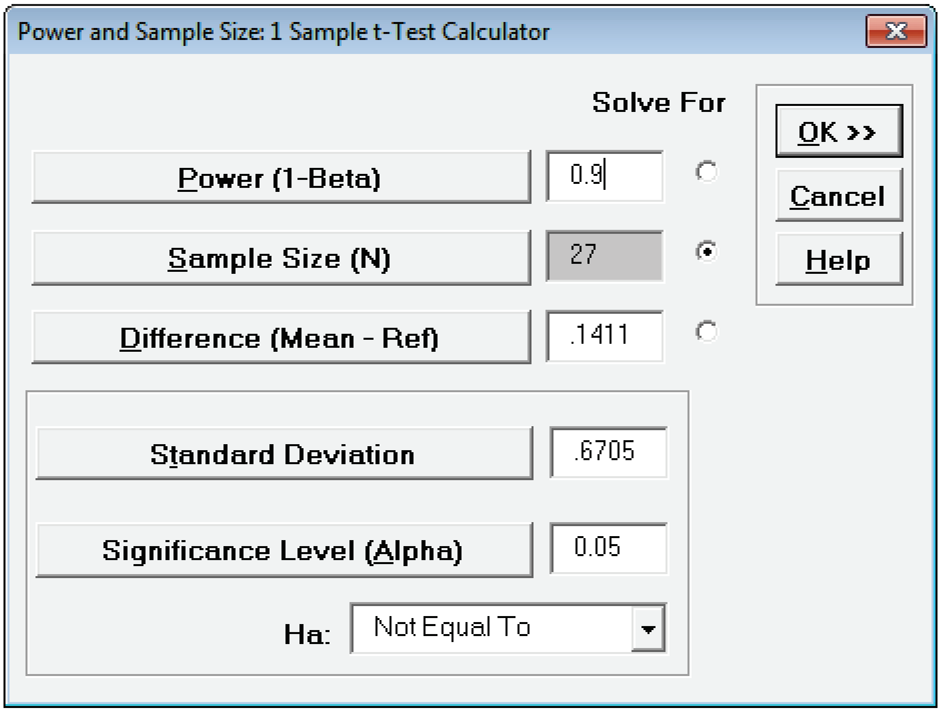
Power and Sample Size: 1 Sample t Test

H0: Mean (μ) = Reference
Ha: Mean (μ) ≠ Reference
Solve For: Power (1 - Beta)

Sample Size (N)	Difference	Standard Deviation	Significance Level (Alpha)	Power (1 - Beta)
27	0.1411	0.6705	0.05	0.183614

4. A power value of 0.1836 is very poor. It is the probability of detecting the specified difference. Alternatively, the associated Beta risk is 1-0.1836 = 0.8164 which is the probability of failure to detect such a difference. Typically, we would like to see Power > 0.9 or Beta < 0.1. In order to detect a difference this small we would need to increase the sample size. We could also set the difference to be detected as a larger value.

5. First, we will determine what sample size would be required in order to obtain a Power value of 0.9. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select the **Solve For Sample Size** button as shown. It is not necessary to delete the entered sample size of 27 – it will be ignored. Enter a **Power Value** of .9:



6. Click **OK**. The resulting report is shown:

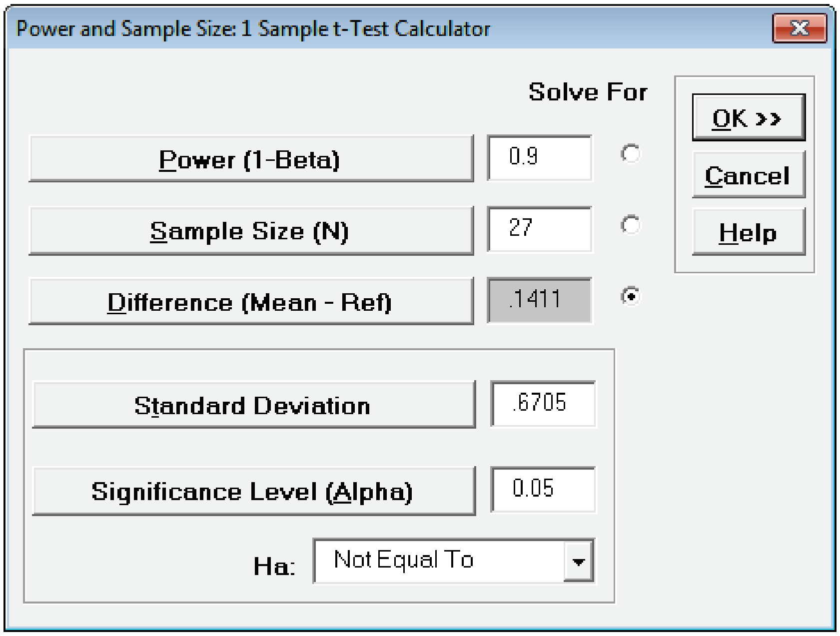
Power and Sample Size: 1 Sample t Test

H0: Mean (μ) = Reference
Ha: Mean (μ) ≠ Reference
Solve For: Sample Size (N)

Power (1 - Beta)	Difference	Standard Deviation	Significance Level (Alpha)	Sample Size (N)	Actual Power
0.9	0.1411	0.6705	0.05	240	0.900958197

7. A sample size of 240 would be required to obtain a power value of 0.9. The actual power is rarely the same as the desired power due to the restriction that the sample size must be an integer. The actual power will always be greater than or equal to the desired power.

8. Now we will determine what the difference would have to be to obtain a Power value of 0.9, given the original sample size of 27. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select the **Solve For Difference** button as shown:



9. Click **OK**. The resulting report is shown:

Power and Sample Size: 1 Sample t Test

H0: Mean (μ) = Reference
Ha: Mean (μ) ≠ Reference
Solve For: Difference (Mean - Reference)

Power (1 - Beta)	Sample Size (N)	Standard Deviation	Significance Level (Alpha)	Difference
0.9	27	0.6705	0.05	0.434637

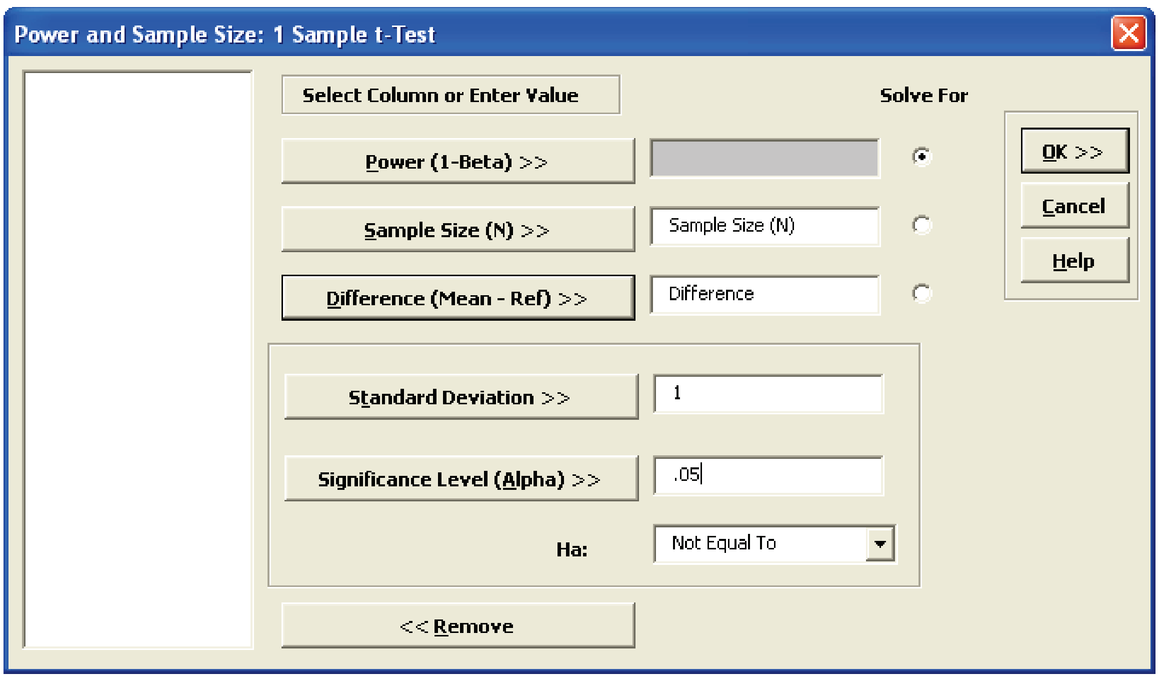
10. A difference of 0.435 would be required to obtain a Power value of 0.9, given the sample size of 27.

Power and Sample Size – One Sample t-Test – Graphing the Relationships between Power, Sample Size, and Difference

In order to provide a graphical view of the relationship between Power, Sample Size, and Difference, SigmaXL provides a tool called **Power and Sample Size with Worksheets**. Similar to the Calculators, **Power and Sample Size with Worksheets** allows you to solve for Power (1 – Beta), Sample Size, or Difference (specify two, solve for the third). You must have a worksheet with Power, Sample Size, or Difference values. Other inputs such as Standard Deviation and Alpha can be included in the worksheet or manually entered.

1. Open the file **Sample Size and Difference Worksheet.xlsx**, select the **Sample Size & Diff** sheet tab. Click **SigmaXL > Statistical Tools > Power & Sample Size with Worksheets** > **1 Sample t-Test**. If necessary, check **Use Entire Data Table**. Click **Next**.

2. Ensure that **Solve For Power (1 – Beta)** is selected. Select *Sample Size (N)* and *Difference* columns as shown. Enter the **Standard Deviation** value of 1. Enter .05 as the **Significance Level** value:



Note: By setting Standard Deviation to 1, the Difference values will be a multiple of Standard Deviation.

Click **OK**. The output report is shown below:

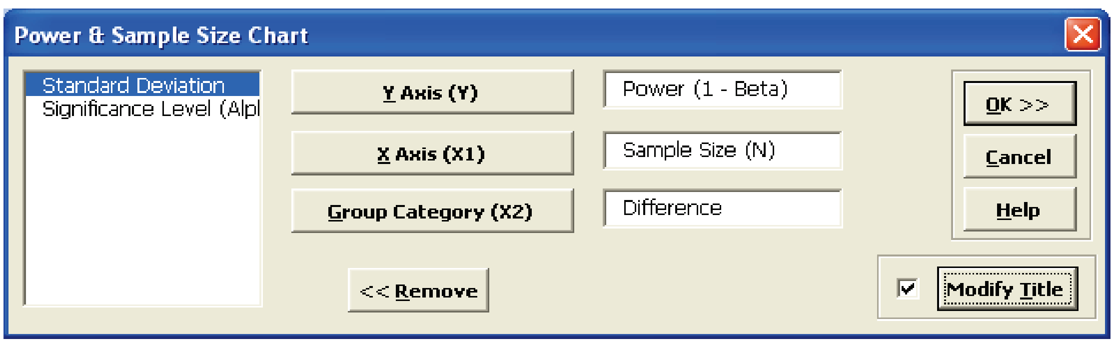
Power and Sample Size: 1 Sample t Test

H0: Mean (μ) = Reference
Ha: Mean (μ) ≠ Reference
Solve For: Power (1 - Beta)

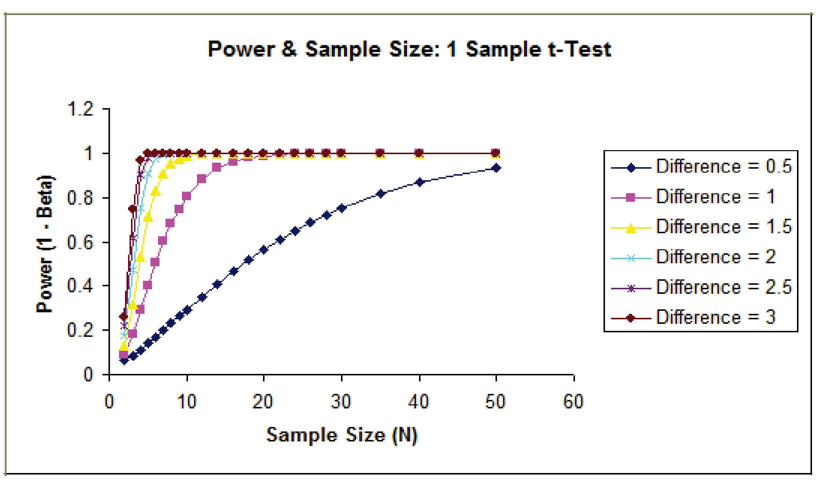
Sample Size (N)	Difference	Standard Deviation	Significance Level (Alpha)	Power (1 - Beta)
2	0.5	1	0.05	0.061948607
3	0.5	1	0.05	0.084107056
4	0.5	1	0.05	0.111274916
5	0.5	1	0.05	0.140516692
6	0.5	1	0.05	0.17070708
7	0.5	1	0.05	0.201327803
8	0.5	1	0.05	0.232077006
9	0.5	1	0.05	0.262746095
10	0.5	1	0.05	0.293175607

3. To create a graph showing the relationship between Power, Sample Size and Difference, click **SigmaXL > Statistical Tools > Power & Sample Size Chart**. Check **Use Entire Data Table**. Click **Next**.

4. Select *Power (1 – Beta)*, click **Y Axis (Y)**; select *Sample Size (N)*, click **X Axis (X1)**; select *Difference*, click **Group Category (X2)**. Click **Add Title**. Enter *Power & Sample Size: 1 Sample t-Test*:



5. Click **OK**. Click **OK**. The resulting Power & Sample Size Chart is displayed:



Since the Standard Deviation was specified as 1, the Difference values are multiples of the Standard Deviation.

Part F – One Sample Nonparametric Tests

Introduction to Nonparametric Tests

Nonparametric tests make fewer assumptions about the distribution of the data compared to parametric tests like the t-Test. Nonparametric tests do not rely on the estimation of parameters such as the mean or the standard deviation. They are sometimes called distribution-free tests.

Nonparametric tests use Medians and Ranks, thus they are robust to outliers in the data. If, however, the data are normal and free of outliers, nonparametric tests are less powerful than normal based tests to detect a real difference when one exists.

Nonparametric tests should be used when the data are nonnormal, the data cannot be readily transformed to normality, and the sample size is too small for robustness due to the Central Limit Theorem.

Exact Nonparametric Tests

Nonparametric tests do not assume that the sample data are normally distributed, but they do assume that the test statistic follows a Normal or Chi-Square distribution when computing the "large sample" or "asymptotic" P-Value. The One-Sample Sign Test, Wilcoxon Signed Rank, Two Sample Mann-Whitney and Runs Test assume a Normal approximation for the test statistic. Kruskal-Wallis and Mood's Median use Chi-Square to compute the P-Value. With very small sample sizes (see recommendations below), these approximations may be invalid, so exact methods should be used. SigmaXL computes the exact P-Values utilizing permutations and fast network algorithms.

It is important to note that while exact P-Values are "correct," they do not increase (or decrease) the power of a small sample test, so they are not a solution to the problem of failure to detect a change due to inadequate sample size.

For data that require more computation time than specified, Monte Carlo P-Values provide an approximate (but unbiased) P-Value that typically matches exact to two decimal places using 10,000 replications. One million replications give a P-Value that is typically accurate to three decimal places. A confidence interval (99% default) is given for the Monte Carlo P-Values. Note that the Monte Carlo confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

Recommended Sample Sizes for use of Exact Nonparametric Tests

Sign Test: N <= 50

Wilcoxon Signed Rank: N <= 15

Mann-Whitney: Each sample N <= 10

Kruskal-Wallis: Each sample N <= 5

Mood's Median: Each sample N <= 10

Runs Test (Above/Below) or Runs Test (Up/Down) Test: N <= 50

These are sample size guidelines for when exact nonparametric tests should be used rather than "large sample" asymptotic based on the Normal or Chi-Square approximation. It is always acceptable to use an exact test, but computation time can become an issue especially for tests with two or more samples. In those cases, one can always use a Monte Carlo P-Value with 99% confidence interval.

Validation of SigmaXL Exact P-Values

SigmaXL Exact P-Values are validated by comparison to textbook examples, published paper examples and other exact software such as StatXact, SPSS Exact, SAS, and Matlab.

Monte Carlo P-Values are validated using 1e6 replications and compared against exact. Repeated simulations are used to validate the confidence intervals.

For further details and references refer to the Appendix **Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests**.

One Sample Sign Test

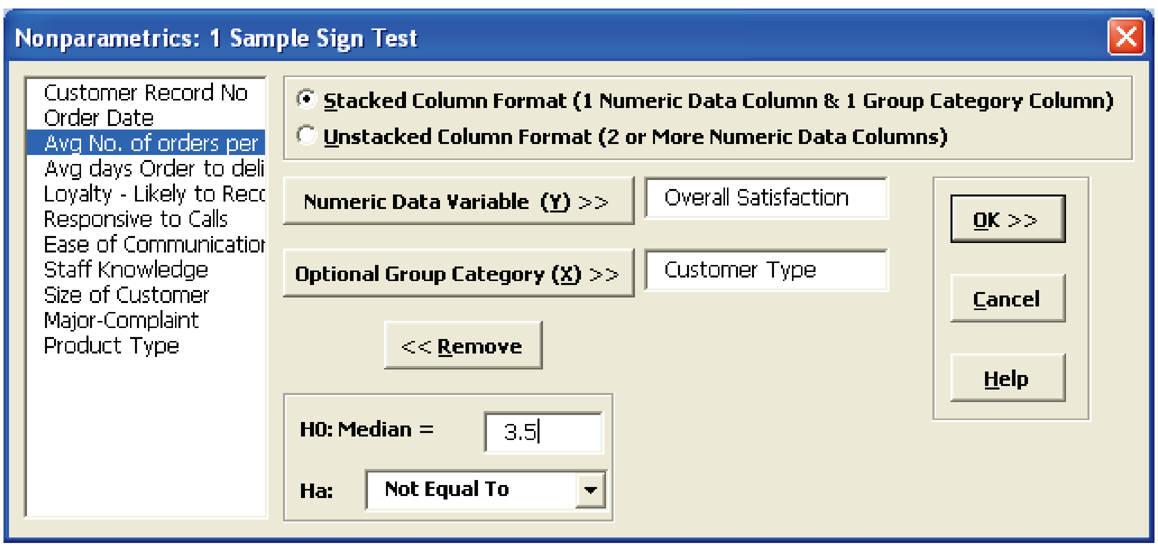
The sign test is the simplest of the nonparametric tests, and is similar to testing if a two-sided coin is fair. Count the number of positive values (larger than hypothesized median), the number of negative values (smaller than the hypothesized median), and test whether there are significantly more positives (or negatives) than expected. The One Sample Sign Test is a nonparametric equivalent to the parametric One Sample t-Test.

Historically, our Median customer satisfaction score has been 3.5. We would like to see if this has changed, with the results grouped by customer type (H0: Median=3.5, Ha: Median≠3.5, α = 0.05).

1. Open **Customer Data.xlsx**, select **Sheet 1** tab. Click **SigmaXL > Statistical Tools > Nonparametric Tests>1 Sample Sign**. If necessary, check **Use Entire Data Table**, click **Next**.

2. Ensure that **Stacked Column Format** is selected. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Optional Group Category (X) >>**.

3. Enter 3.5 for the **Null Hypothesis H0: Median** value. Set **Ha** as *Not Equal To*.



4. Click **OK**. Results:

1 Sample Sign Test for Medians: Overall Satisfaction			
Test Information			
H_0: Median = 3.5			
H_a: Median Not Equal To 3.5			
Customer Type	**1**	**2**	**3**
Count (N)	31	42	27
Median	3.560	4.340	3.510
Points Below 3.5	15	5	13
Points Equal To 3.5	0	1	0
Points Above 3.5	16	36	14
P-Value (2-sided)	1.0000	0.0000	1.0000

Note the P-Values. Customer Type 2 shows a significant change (increase) in Satisfaction Median (P-Value < .05), whereas Customer Types 1 and 3 show no change (P-Value ≥ .05). While the P-Values are not the same as those given by the 1 sample t-Test, the conclusions do match.

If Count (N) is less than or equal to 50, the Sign Test computes an exact P-Value using the binomial distribution. For N > 50, the P-Value is estimated using a normal approximation. Since this is always done automatically and is very fast, the Sign Test is not included in the separate Nonparametric Exact menu.

One Sample Wilcoxon Signed Rank Test

The Wilcoxon Signed Rank test is a more powerful nonparametric test than the Sign Test, but it adds an assumption that the distribution of values is symmetric around the median. An example of a symmetric distribution is the uniform distribution. Symmetry can be observed with a histogram, or by checking to see if the Skewness is large (> .5 or < - .5).

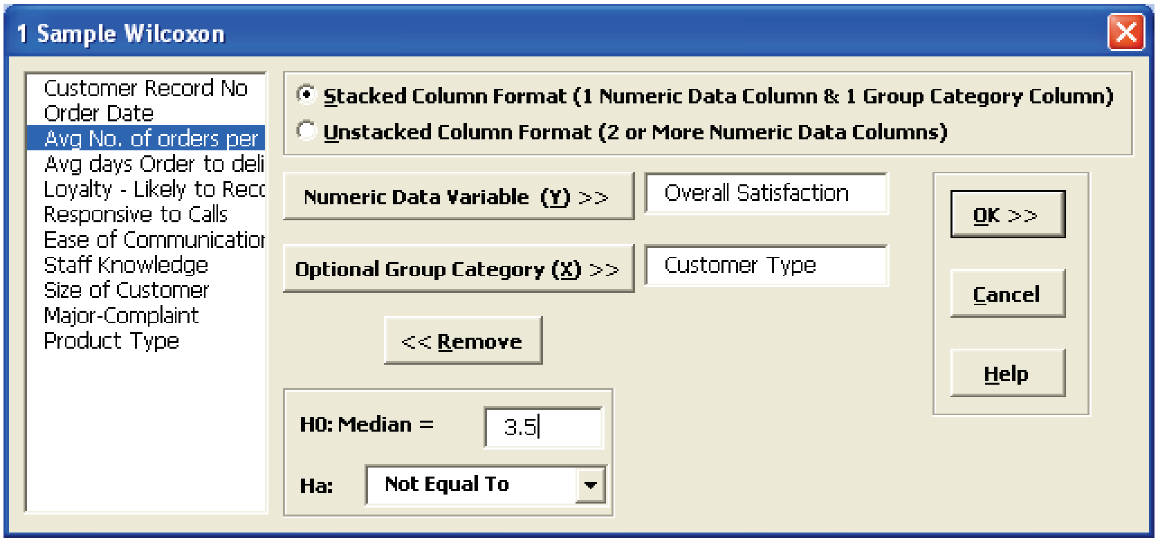
The One Sample Wilcoxon Test is a nonparametric equivalent to the parametric One Sample t-Test (i.e., One Sample t-Test on Ranks).

Historically, our Median customer satisfaction score has been 3.5. We would like to see if this has changed, with the results grouped by customer type (H0: Median=3.5, Ha: Median≠3.5, α = 0.05).

1. Open **Customer Data.xlsx**, select Sheet 1 tab (or press F4 to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests > 1 Sample Wilcoxon**. If necessary, check Use Entire Data Table, click Next.

2. Ensure that **Stacked Column Format** is selected. Select *Overall Satisfaction*, click **Numeric Data Variable (Y)** >>; select *Customer Type*, click **Optional Group Category (X) >>**.

3. Enter 3.5 for the **Null Hypothesis H0: Median** value. Keep **Ha** as *Not Equal To*.



4. Click **OK**. Results:

1 Sample Wilcoxon Test: Overall Satisfaction			
Test Information			
H_0: Median = 3.5			
H_a: Median Not Equal To 3.5			
Customer Type	**1**	**2**	**3**
Count (N)	31	42	27
Count for Test	31	41	27
Median	3.56	4.34	3.51
Wilcoxon Statistic	217.50	802.50	222.00
P-Value (2-sided)	0.5566	0.0000	0.4349

Note the P-Values. Customer Type 2 shows a significant change (increase) in Satisfaction Median (P-Value < .05), whereas Customer Types 1 and 3 show no change (P-Value ≥ .05). Although the P-Values are not identical to the sign test and t-Test, the conclusions match. (Note, in the case of Customer Type 2, the Sign Test is preferred since the data is not symmetrical but skewed).

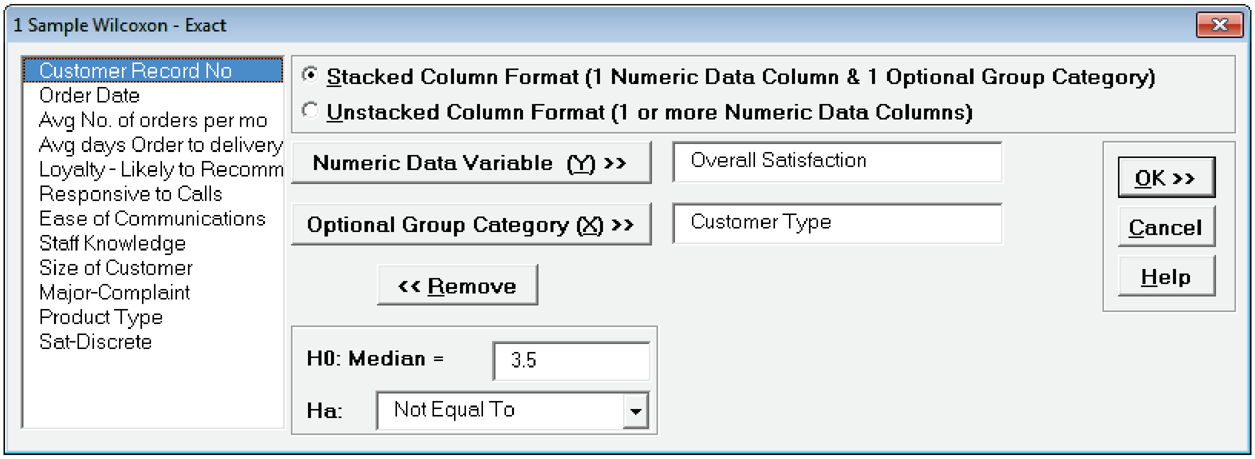
One Sample Wilcoxon Signed Rank Test - Exact

We will now redo the above example to compute exact P-Values. Typically, this would not be necessary unless the sample sizes were smaller (N <= 15 for Wilcoxon), but this gives us continuity on the example. We will consider a small sample problem later.

1. Open **Customer Data.xlsx**, select Sheet 1 tab (or press F4 to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > 1 Sample Wilcoxon - Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

2. Ensure that **Stacked Column Format** is selected. Select *Overall Satisfaction*, click **Numeric Data Variable (Y)** >>; select *Customer Type*, click **Optional Group Category (X) >>**.

3. Enter 3.5 for the **Null Hypothesis H0: Median** value. Keep **Ha** as *Not Equal To*.



4. Click **OK**. Results:

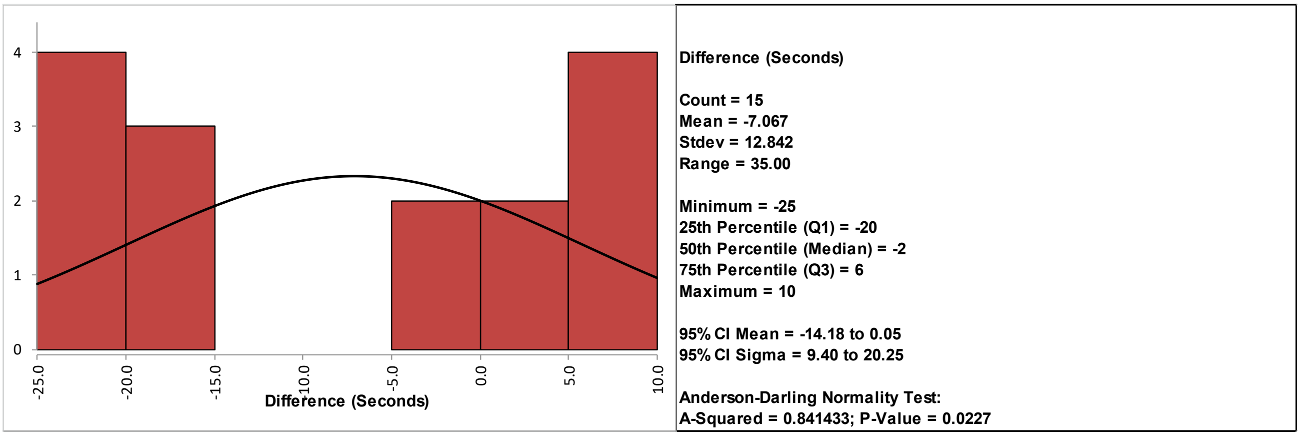
1 Sample Wilcoxon Test - Exact: Overall Satisfaction			
Test Information			
H_0: Median = 3.5			
H_a: Median Not Equal To 3.5			
Customer Type	**1**	**2**	**3**
Count (N)	31	42	27
Count for Test	31	41	27
Median	3.56	4.34	3.51
Wilcoxon Statistic	217.50	802.50	222.00
Exact P-Value (2-sided)	0.5584	0.0000	0.4410

The Wilcoxon Statistics are identical to the above "large sample" or "asymptotic" results. The Exact P-Values are close but slightly different. This was expected because the sample sizes are reasonable (N > 15), so the "large sample" P-Values are valid using a normal approximation for the Wilcoxon Statistic.

Note, if Count (N) is greater than 1000, the Exact P-Value is estimated using a continuity - corrected normal approximation. Since the Wilcoxon Exact P-Value is computed very quickly for sample sizes as large as 1000, Monte Carlo P-Values are not required.

5. Now we will consider a small sample problem. Open **Nonnormal Task Time Difference – Small Sample.xlsx**. A study was performed to determine the effectiveness of training to reduce the time required to complete a short but repetitive process task. Fifteen operators were randomly selected and the difference in task time was recorded in seconds (after training – before training). A negative value denotes that the operator completed the task in less time after training than before.

6. Click **SigmaXL > Graphical Tools > Histograms & Descriptive Statistics.** If necessary, check **Use Entire Data Table**. Click **Next**. Select *Difference (Seconds)*, click **Numeric Data Variables (Y) >>.** Click **OK**.

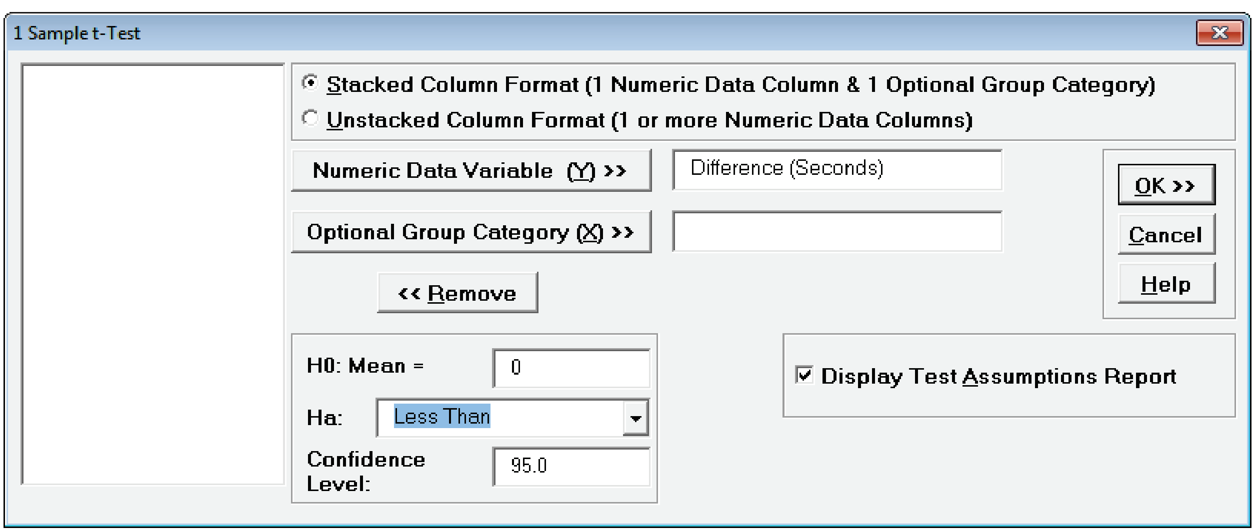


This small sample data fails the Anderson Darling Normality Test (P-Value = .023). Note that this is due to the data being uniform or possibly bimodal, not due to a skewed distribution. Now we will perform a 1 Sample t-Test and review the assumptions.

7. Select **Task Time Difference** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > 1 Sample t-Test & Confidence Intervals**. If necessary, check **Use Entire Data Table**, click **Next**.

8. Ensure that **Stacked Column Format** is selected. Select *Difference (Seconds)*, click **Numeric Data Variable (Y) >>**.

9. This is a one-sided test because we have no reason to expect that the task time will increase, so the Null Hypothesis H0: μ = 0; and the Alternative Hypothesis Ha: μ < 0. Enter 0 for the **H0: Mean** value. Select **Ha** as *Less Than*, **Confidence Level** = 95.0%, and check **Display Test Assumptions Report**.



10. Click **OK**.

1 Sample t-Test	
Test Information	
H_0: Mean (Mu) = 0	
H_a: Mean (Mu) Less Than 0	
Results:	**Difference (Seconds)**
Count	15
Mean	-7.067
StDev	12.842
SE Mean	3.316
t	-2.131
P-Value (1-sided)	0.0256
UC (1-sided, 95%)	-1.226

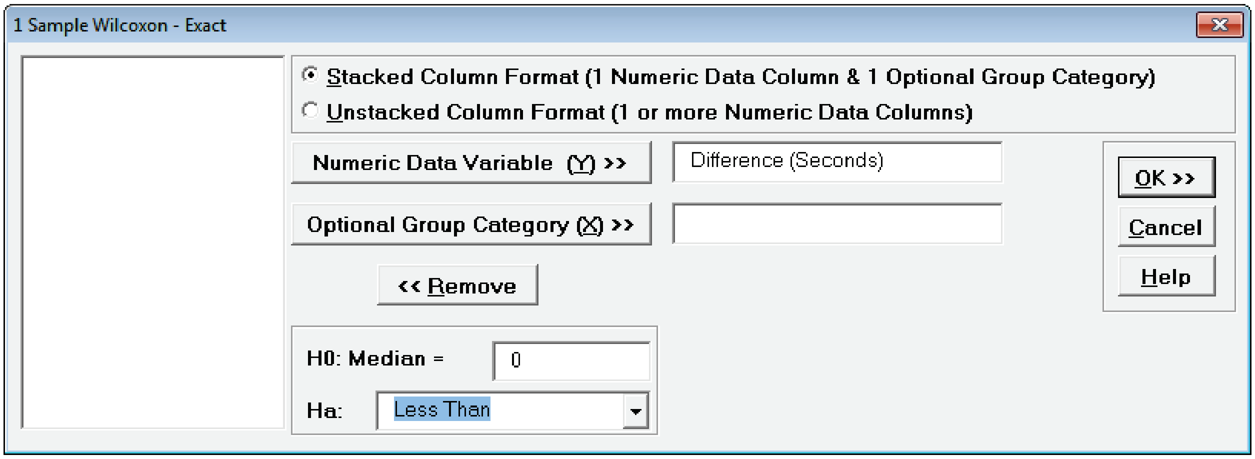
1 Sample t-Test Assumptions Report	
Normality:	Anderson Darling P-Value = 0.023. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.1369 and Kurtosis value = -1.8385. See robustness and outliers.
Robustness:	Minimum sample size for a robust t-test = 16. It is recommended that the Exact Nonparametric One Sample Wilcoxon Test be used (SigmaXL > Statistical Tools > Nonparametric Tests - Exact > 1 Sample Wilcoxon - Exact).
Outliers (Boxplot Rules):	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.

The **1 Sample t-Test Assumptions Report** highlights that the data are not normal, but note that Kurtosis equal to -1.84 is the issue here, not Skewness. This was observed in the Histogram above with the data being uniform or possibly bimodal.

The sample size is too small for a robust t-Test, so the Exact One Sample Wilcoxon Test is recommended. The Wilcoxon Test is recommended over the Sign Test because it is a more powerful test and meets the requirement that the data be symmetrical. The Exact test is recommended because the sample size is very small (N <= 15).

11. Select **Task Time Difference** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > 1 Sample Wilcoxon - Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

12. Ensure that **Stacked Column Format** is selected. Select *Difference (Seconds)*, click **Numeric Data Variable (Y)** >>. Enter 0 for the **H0: Median** value. Select **Ha** as *Less Than*.



13. Click **OK**. Results:

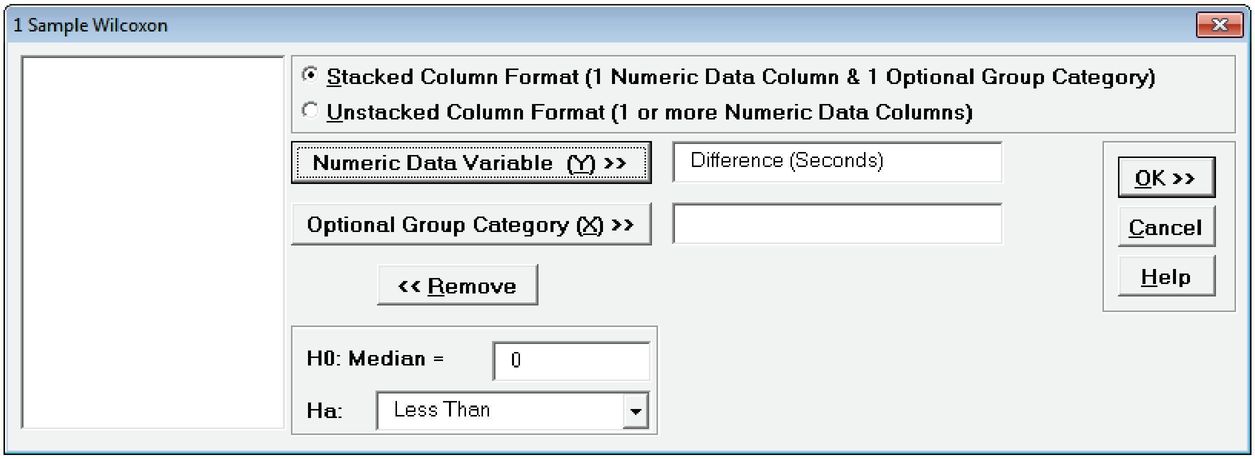
1 Sample Wilcoxon Test - Exact	
Test Information	
H_0: Median = 0	
H_a: Median Less Than 0	
Results:	**Difference (Seconds)**
Count (N)	15
Count for Test	14
Median	-2
Wilcoxon Statistic	26.00
Exact P-Value (1-sided)	0.0497

With the P-Value = .0497 we reject H0 and conclude that the Median Task Time Difference is significantly less than 0, so the training is effective.

By way of comparison we will now rerun the analysis using the large sample (asymptotic) Wilcoxon test.

14. Select **Task Time Difference** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests > 1 Sample Wilcoxon**. If necessary, check **Use Entire Data Table**, click **Next**.

15. Ensure that **Stacked Column Format** is selected. Select *Difference (Seconds)*, click **Numeric Data Variable (Y)** >>. Enter 0 for the **H0: Median** value. Select **Ha** as *Less Than*.



16. Click **OK**. Results:

1 Sample Wilcoxon Test	
Test Information	
H_0: Median = 0	
H_a: Median Less Than 0	
Results:	**Difference (Seconds)**
Count (N)	15
Count for Test	14
Median	-2
Wilcoxon Statistic	26.00
P-Value (1-sided)	0.0513

Now with the P-Value = .0513 we **incorrectly fail to reject** H0.

The difference between exact and large sample P-Value is small but it was enough to lead us to falsely conclude that the training is ineffective.

In conclusion, whenever you have a small sample size and are performing a nonparametric test, always use the Exact option.

Part G – Two Sample t-Test

Two Sample t and Equivalence Test Templates

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
2 Sample t-Test and Confidence Interval (Compare 2 Means) or
2 Sample Equivalence Test (Compare 2 Means).

These templates are also located at **SigmaXL > Statistical Tools > Basic Statistical Templates**.

See **Measure Phase Part B – Templates and Calculators** for Two Sample t-Test template examples:

Basic Statistical Templates – 2 Sample t-Test and Confidence Interval (Compare 2 Means)

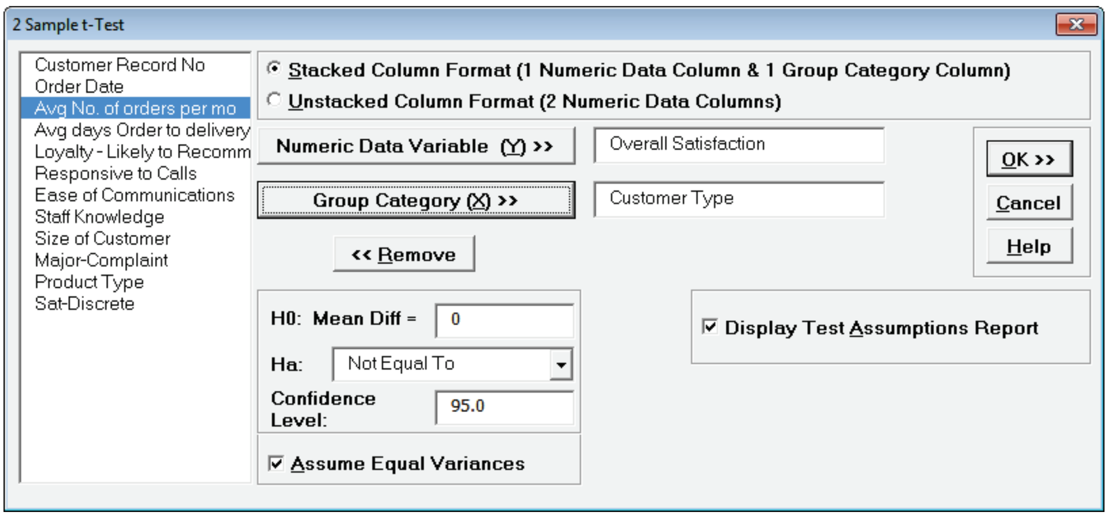
Basic Statistical Templates – 2 Sample Equivalence Test (Compare 2 Means)

Two Sample t-Test

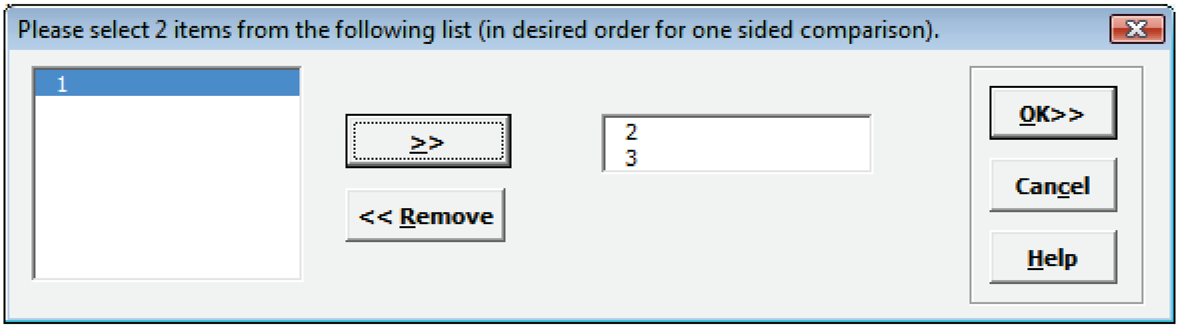
1. Open **Customer Data.xlsx**, click **Sheet 1** tab (or press **F4** to activate last worksheet). We will look at comparing means of Customer Satisfaction by Customer Type (2 vs. 3), using the Two Sample t-test. H0: μ2=μ3, Ha: μ2≠μ3

2. Click **SigmaXL > Statistical Tools > 2 Sample t-test**. If necessary, check **Use Entire Data Table**, click **Next**.

3. With stacked column format checked, select *Overall Satisfaction*, click **Numeric Data Variable Y >>**; select *Customer Type*, click **Group Category X >>**; **H0: Mean Diff =** 0; **Ha:** *Not Equal To*; **Confidence Level:** 95%; ensure that **Assume Equal Variances** and **Display Test Assumption Report** are checked:



4. Click **OK**. Select *Customer Type 2* and *3*.



5. Click **OK**. Resulting output:

2 Sample t-Test: Overall Satisfaction		
Test Information		
H_0: Mean Difference = 0		
H_a: Mean Difference Not Equal To 0		
Assume Equal Variance		
Customer Type	**2**	**3**
Count	42	27
Mean	4.205	3.641
Standard Deviation	0.621200	0.670478
Mean Difference	0.564127	
Std Error Difference	0.158060	
DF	67	
t	3.569	
P-Value (2-sided)	0.0007	
UC (2-sided, 95%)	0.879616	
LC (2-sided, 95%)	0.248638	

Given the P-Value of .0007 we reject H0 and conclude that Mean Customer Satisfaction is significantly different between Customer type 2 and 3. This confirms previous findings.

6. 2 Sample t-Test Assumptions Report:

2 Sample t-Test Assumptions Report		
Normality:	Anderson Darling P-Value = 0.030. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.9680 and Kurtosis value = 0.6796. See robustness and outliers.	Anderson Darling P-Value = 0.360. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.
Robustness:	Minimum sample size for a robust two sample t-test = 2. Since each sample size is greater than this, the two sample t-test is robust to nonnormality.	
Outliers (Boxplot Rules):	Potential (1.5*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.	(Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.
Equal Variance:	Levene's test for Equal Variances P-Value = 0.622. Fail to reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is not violated.	

This is a text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation).

Each sample is tested for Normality using the Anderson-Darling test. If not normal, the minimum sample size for robustness of the 2 sample t-Test is determined utilizing Monte Carlo regression equations (see Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA). If the sample size is inadequate, a warning is given and the Nonparametric Mann-Whitney Test is recommended.

Each sample is tested for Outliers defined as: Potential: Tukey's Boxplot (> Q3 + 1.5*IQR or < Q1 – 1.5*IQR); Likely: Tukey's Boxplot 2.2*IQR; Extreme: Tukey's Boxplot 3*IQR. If outliers are present, a warning is given and recommendation to review the data with a Boxplot and Normal Probability Plot. Here we have a potential outlier for Customer Type 2.

Tip: If the removal of outlier(s) result in an Anderson Darling P-Value that is > 0.1, a notice is given that excluding the outlier(s), the sample data are inherently normal.

Each sample is tested for Randomness using the Exact Nonparametric Runs Test. If the sample data is not random, a warning is given and recommendation to review the data with a Run Chart.

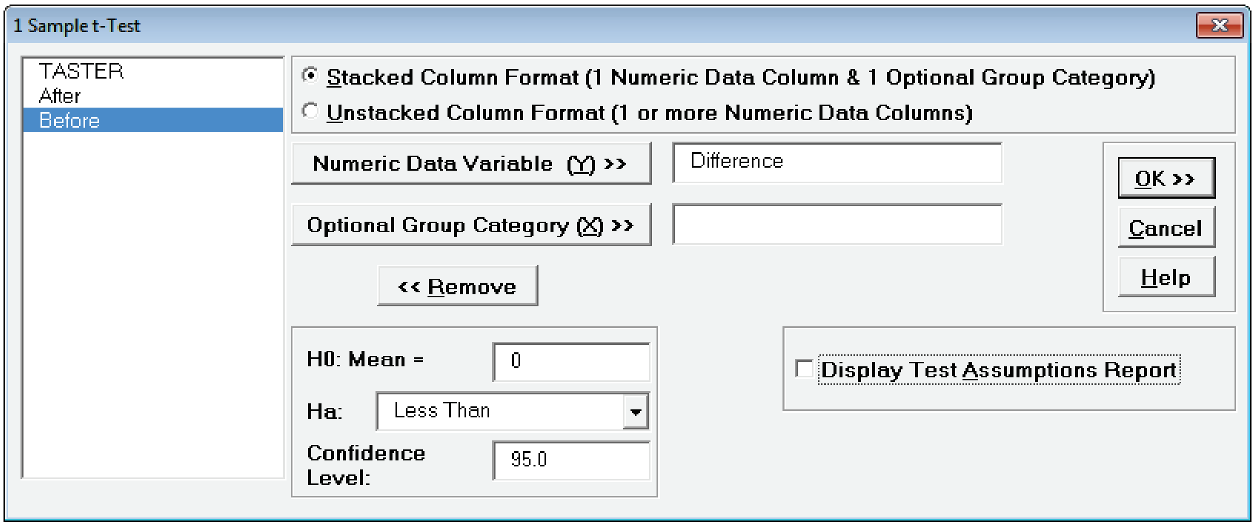
A test for Equal Variances is also applied. If all sample data are normal, the F-Test is utilized, otherwise Levene's Test is used. If the variances are unequal and the test being used is the equal variance option, then a warning is given and Welch's test is recommended.

See Appendix **Hypothesis Test Assumptions Report** for further details.

Paired t-Test

1. Open the file **Dietcola.xlsx**. These are the results of a Before and After taste test on sweetness for diet cola. Ten tasters were used and one month elapsed with the cola in warm storage between the before and after results. Do a one sample t-test on the column of differences.

2. Click **SigmaXL > Statistical Tools > 1-Sample t-test & Confidence Intervals**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**. Select *Difference*, click **Numeric Data Variable (Y) >>**, set **H0: Mean** = 0, **Ha:** *Less Than* (this is a one-sided or one-tail test – sweetness cannot increase). Uncheck **Display Test Assumptions Report** for now:



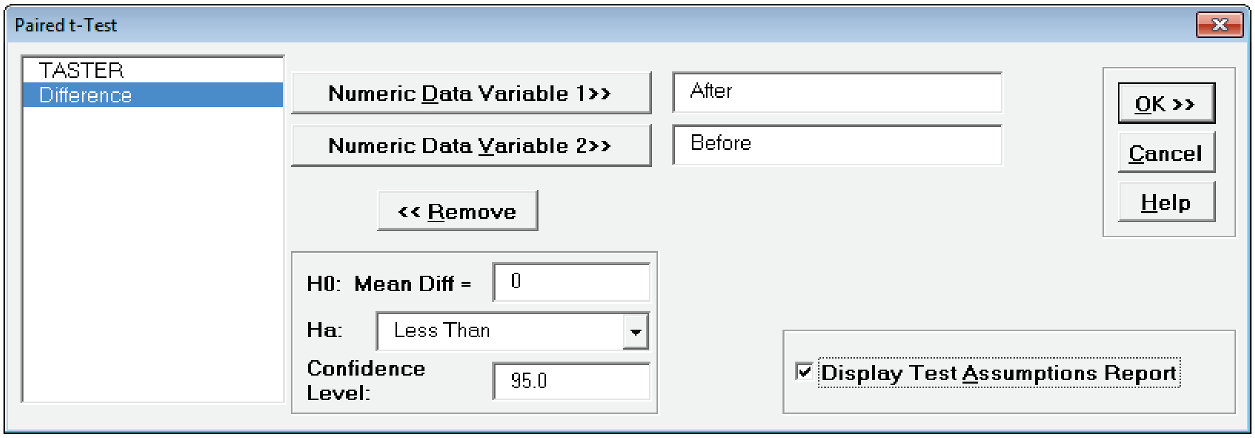
3. Click **OK**. Result:

1 Sample t-Test	
Test Information	
H_0: Mean (Mu) = 0	
H_a: Mean (Mu) Less Than 0	
Results:	**Difference**
Count	10
Mean	-1.02
StDev	1.196
SE Mean	0.378242
t	-2.697
P-Value (1-sided)	0.0123
UC (1-sided, 95%)	-0.326641

Given the P-Value of .012, we reject H0 and conclude that the sweetness has in fact decreased.

4. Now redo the analysis using the paired t-test: Click **Sheet 1** Tab; Click **SigmaXL > Statistical Tools > Paired t-Test**; Click **Next**; Select *After*, click **Numeric Data Variable 1 >>**, select *Before* as

Numeric Data Variable 2 >>, **H0: Mean Diff =** 0, **Ha:** *Less Than*. Check **Display Test Assumptions Report**.



5. Click **OK**. Results are identical to the One sample t-test analysis of difference column, with the added assumptions report showing that all assumptions are met:

Paired t-Test	
Test Information	
H_0: Mean Difference = 0	
H_a: Mean Difference Less Than 0	
Results:	**After - Before**
Count	10
Mean	-1.02
StDev	1.196
SE Mean	0.378242
t	-2.697
P-Value (1-sided)	0.0123
UC (1-sided, 95%)	-0.326641

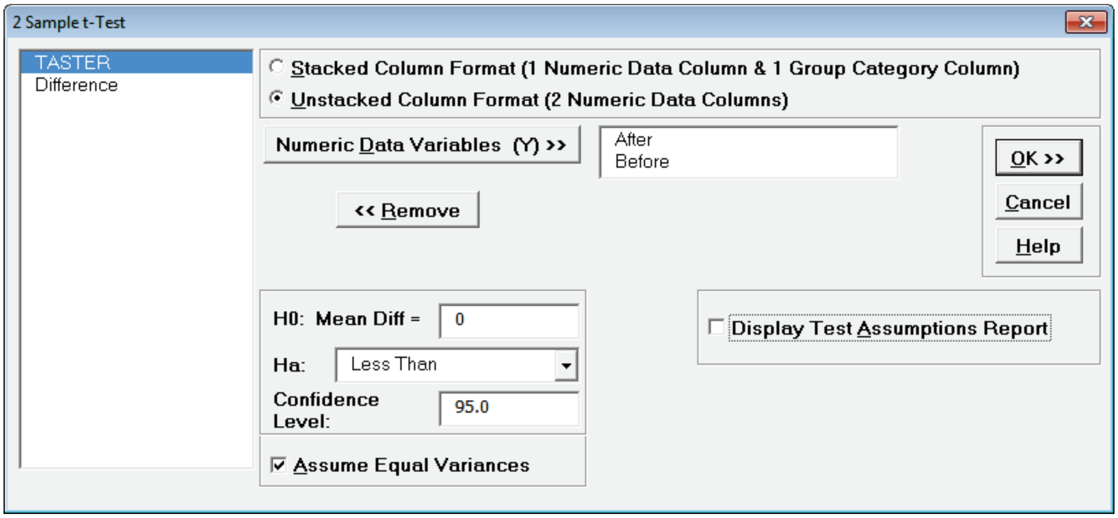
Paired t-Test Assumptions Report	
Normality:	Anderson Darling P-Value = 0.369. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.
Robustness:	Not applicable for normal data.
Outliers (Boxplot Rules):	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 0.079. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.

Unpaired 2 Sample t-Test vs. Paired t-Test

1. Open the **Dietcola.xlsx** file, click the **Sheet 1** tab (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Statistical Tools > 2 Sample t-Test**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Check **Unstacked Column Format**. Select *After*, *Before* and click **Numeric Data Variables (Y) >>**. **H0: Mean Diff** = 0, **Ha:** *Less Than*; check **Assume Equal Variances.** Uncheck **Display Test Assumptions Report**:



4. Click **OK**. Results:

2 Sample t-Test		
Test Information		
H_0: Mean Difference = 0		
H_a: Mean Difference Less Than 0		
Assume Equal Variance		
Results:	**After**	**Before**
Count	10	10
Mean	6.580	7.600
Standard Deviation	1.871	0.951023
Mean Difference	-1.02	
Std Error Difference	0.663626	
DF	18	
t	-1.537	
P-Value (1-sided)	0.0708	
UC (1-sided, 95%)	0.130770	

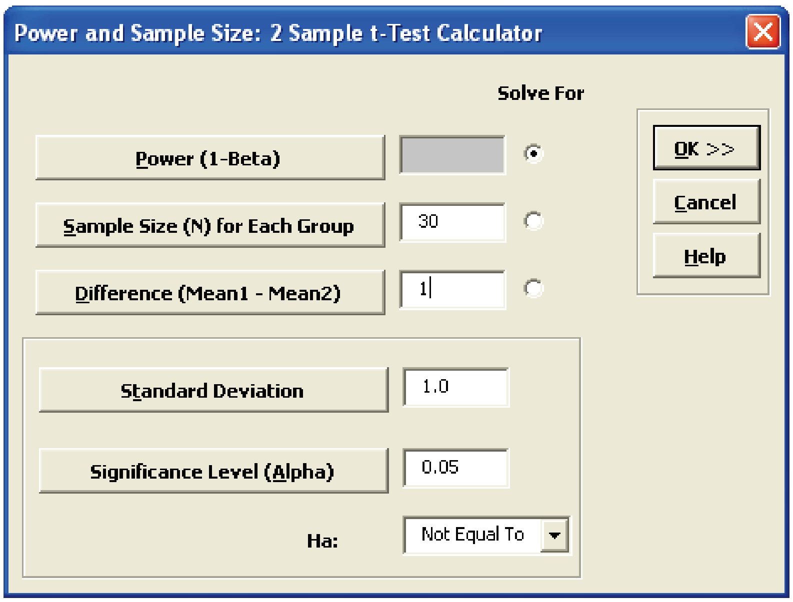
Now the P-Value is .07, indicating a fail to reject H0. What changed? Hint: Compare the SE Mean of the Paired t-test to the Std Error Difference of the unpaired two-sample t-test. Where does the additional variability come from in the two-sample t-test? The paired t-test is the appropriate test to use here.

Power & Sample Size for 2 Sample T-Test

To determine Power & Sample Size for a 2 Sample t-Test, you can use the Power & Sample Size Calculator or Power & Sample Size with Worksheet.

1. Click **SigmaXL > Statistical Tools > Power & Sample Size Calculators > 2 Sample t-Test Calculator**.

2. Select **Solve For Power (1 – Beta)**. Enter **Sample Size** and **Difference** as shown:



Note that we are calculating the power or likelihood of detection given that Mean1 – Mean2 = 1, with sample size for each group = 30, standard deviation = 1, significance level = .05, and Ha: Not Equal To (two-sided test).

3. Click **OK**. The resulting report is displayed:

Power and Sample Size: 2 Sample t Test

H0: Mean 1 = Mean 2
Ha: Mean 1 ≠ Mean 2
Solve For: Power (1 - Beta)

Sample Size (N)	Difference	Standard Deviation	Significance Level (Alpha)	Power (1 - Beta)
30	1	1	0.05	0.967708259

A power value of 0.97 is good, hence we have the basis for the “minimum sample size n=30” rule of thumb used for continuous data.

4. To determine Power & Sample Size using a Worksheet, click **SigmaXL > Statistical Tools > Power & Sample Size with Worksheet > 2 Sample t-Test**.

5. A graph showing the relationship between Power, Sample Size and Difference can then be created using **SigmaXL > Statistical Tools > Power & Sample Size Chart**. See Part E for an example using the 1 Sample t-Test.

Part H – Two Sample F and Comparison Test

Two Sample F-Test Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates > 2 Sample F-Test and CI (Compare 2 Standard Deviations)** or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Sample F-Test and CI (Compare 2 Standard Deviations)** to access the 2 Sample F-Test calculator. The template gives the following default example.

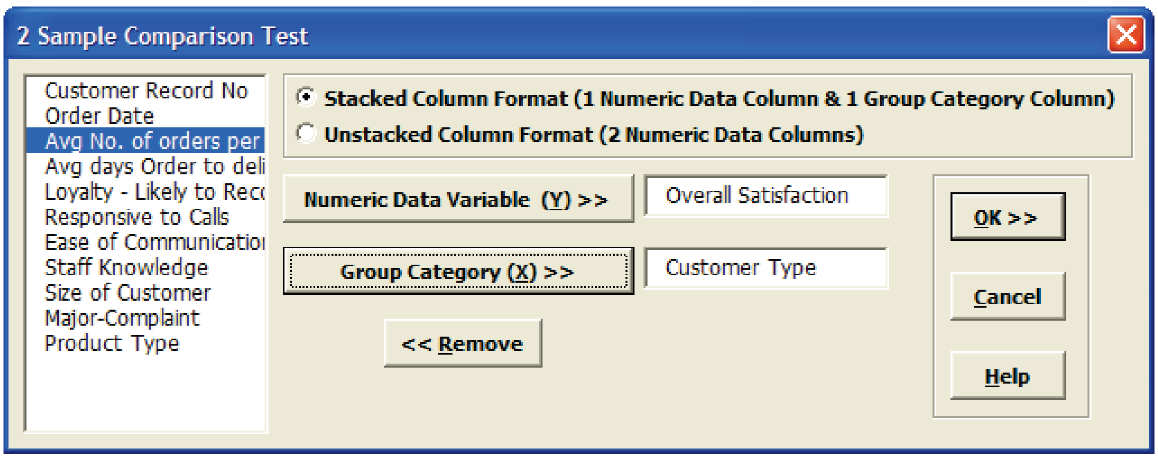
2 Sample F-Test and Ratio Confidence Interval (Compare 2 Standard Deviations)			
Sample Data (user inputs):		Sample 1	Sample 2
Sample Size	n	10	10
Sample Standard Deviation	s	2.0000	1.0000
Null Hypothesis (hypothesized ratio of StDevs)	H_0: σ_1/σ_2 =	1.0	
Alternative Hypothesis	H_a: σ_1/σ_2	Not Equal To	
Confidence Level (enter .95 for 95%)		95.0%	
Results:			
Ratio of Standard Deviations		2.0000	
Sample Variance		4.0000	1.0000
Ratio of Variances		4.0000	
df		9	9
F-statistic		4.0000	
alpha		0.0500	
P-Value (2-sided)		0.0510	
Upper Confidence Limit (2 Sided)		4.0130	
Lower Confidence Limit (2 Sided)		0.9968	

Two Sample Comparison Test

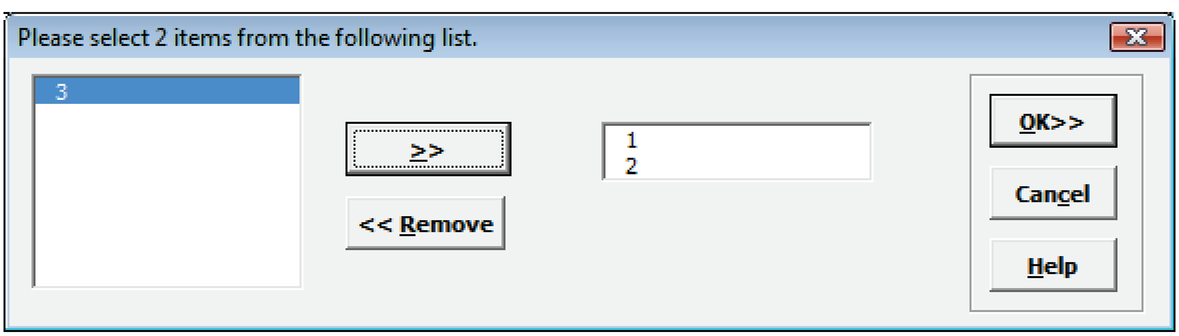
We will now do a full comparison test of Customer Satisfaction for Customer Type 1 and 2. This test checks each sample for normality, equal variance (F-test and Levene's), 2 sample t-test (assuming equal and unequal variance), and Mann Whitney test for equal Medians. Depending on the normality, variance, and sample size results, the appropriate P-Values are highlighted in yellow.

1. Open **Customer Data.xlsx**, click on **Sheet 1** Tab.

2. Click **SigmaXL > Statistical Tools > 2 Sample Comparison Tests**. Click **Next**. Check **Stacked Column Format**. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**, select *Customer Type*, click **Group Category (X) >>**.



3. Click **OK**. Select *Customer Type 1* and *2*.



4. Click **OK**.

2 Sample Comparison - Overall Satisfaction		
Customer Type	**1**	**2**
Count	31	42
Mean	3.394	4.205
Median	3.560	4.340
Standard Deviation	0.824680	0.621200
AD Normality Test p-value	0.5306	0.0302
Test for Equal Variances:		
F-test (use with normal data):		
F (test statistic)	1.762	
p-value (2-sided)	0.0916	
Levene's test (use with non-normal data):		
p-value (2-sided)	0.0443	
2 Sample t-Test for means:		
Assume Equal Variance:		
t (test statistic)	-4.799	
p-value (2-sided)	0.0000	
p-value (1-sided)	0.0000	
Assume Unequal Variance:		
t (test statistic)	-4.601	
p-value (2-sided)	0.0000	
p-value (1-sided)	0.0000	
2 Sample Mann-Whitney test for medians:		
p-value (2-sided)	0.0000	
p-value (1-sided)	0.0000	

Test 1: Anderson Darling
H0: Data are Normal
Ha: Data are Not Normal

Test 2: Equal Variances
H0: $\sigma_1^2 = \sigma_2^2$
Ha: $\sigma_1^2 \neq \sigma_2^2$

Test 3: Equal Means
H0: $\mu_1 = \mu_2$
Ha: $\mu_1 \neq \mu_2$

Test 4: Equal Medians
H0: $\text{Median}_1 = \text{Median}_2$
Ha: $\text{Median}_1 \neq \text{Median}_2$

Customer Type 2 has nonnormal data. This makes Levene's test the appropriate test for unequal variance. Levene's test indicates that Customer type 2 has a significantly lower variance, or standard deviation. The lower standard deviation translates to a **consistent** level of satisfaction.

Since Levene's test indicates unequal variance, the appropriate t-test assumes unequal variance. The t-test indicates that Customer Type 2 has a significantly higher mean satisfaction.

The Two Sample Mann-Whitney test also shows a significant difference in Medians. This test is highlighted when the data are not normal and the sample size is small (as determined by the formulas used in **Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA**).

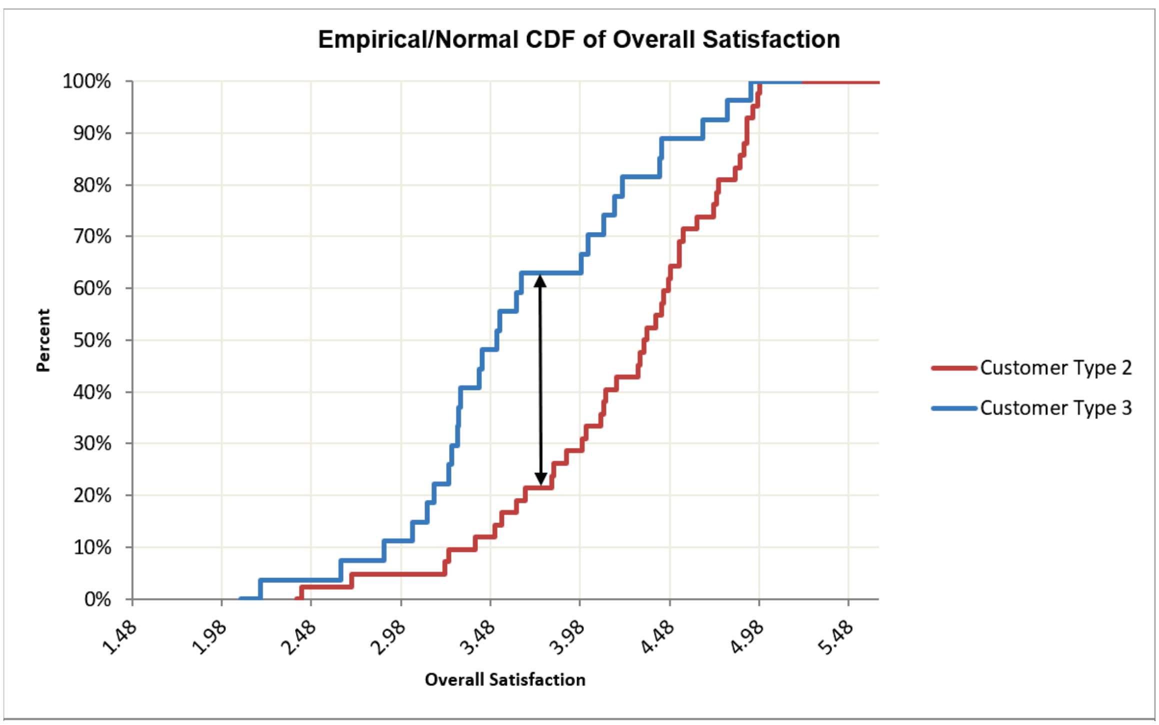
Clearly the next step would be to determine a root cause or best practices to reduce the variability in overall satisfaction and increase the mean for all customer types.

Part I – Two Sample Nonparametric Test: Mann-Whitney

Two Sample Mann-Whitney Test (with 2 Sample KS Option)

We will look at comparing medians of Customer Satisfaction by Customer Type, using the Two Sample Mann-Whitney test with H0: Median Difference = 0, Ha: Median Difference ≠ 0. The Two Sample Mann-Whitney Test is the nonparametric equivalent to the parametric Two Sample t-Test (i.e., Two Sample t-Test on Ranks). The test does not assume sample normality but does assume that the samples have equal shapes. If the shapes are different, the null hypothesis is that the distributions are the same.

The optional Two Sample KS (Kolmogorov-Smirnov) test is used to compare the distributions of two samples. The test is H0: Distribution (CDF) Difference = 0, Ha: Distribution (CDF) Difference ≠ 0. The CDF is the Cumulative Distribution Function. The two-sided test statistic is the maximum absolute difference between the CDF values as shown:

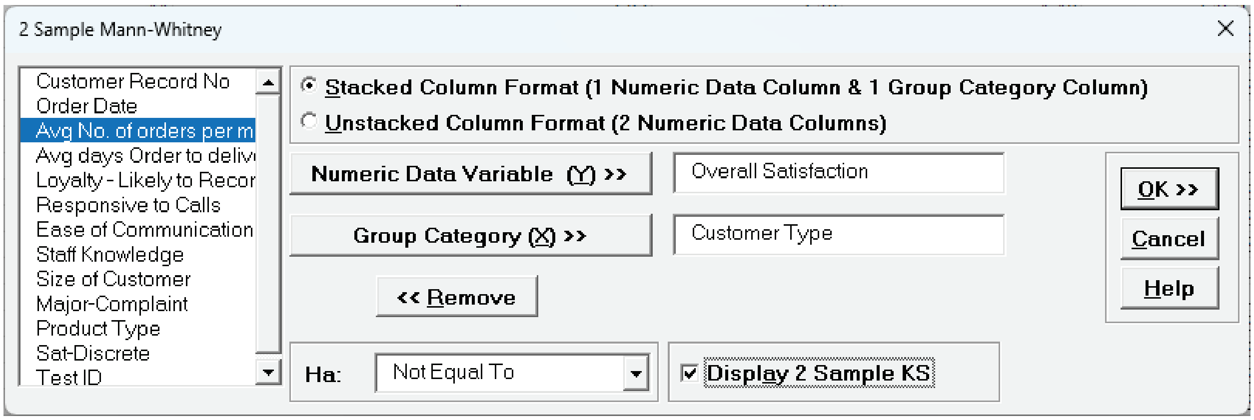


Note this graph may be recreated using **SigmaXL > Graphical Tools > Empirical/Normal CDF Plots**. **Display Normal CDF Plots** is unchecked. Copy and Paste Customer Type 3 plot into Customer Type 2. Adjust legend and axis labels.

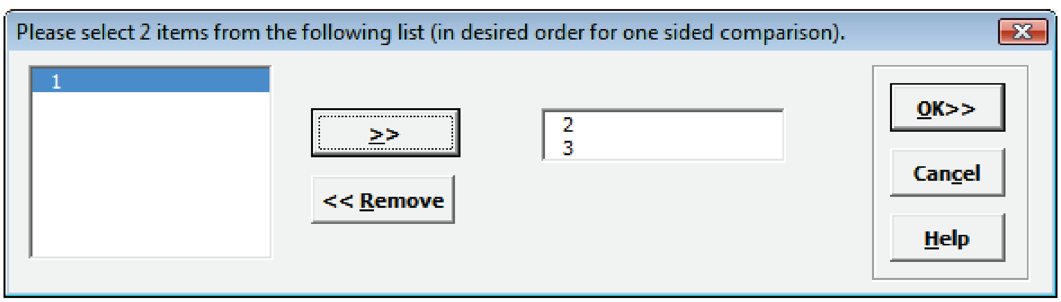
1. Open **Customer Data.xlsx**, click **Sheet 1** tab (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Statistical Tools > Nonparametric Tests > 2 Sample Mann-Whitney**. If necessary, check **Use Entire Data Table**, click **Next**.

3. With **Stacked Column Format** checked, select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**; and **Ha:** *Not Equal To*. Check **Display 2 Sample KS**.



4. Click **OK**. Select *Customer Type 2* and *3*.



5. Click **OK**. The resulting output for the 2 Sample Mann-Whitney test is:

2 Sample Mann-Whitney - Overall Satisfaction		
Test Information		
H_0: Median Difference = 0		
H_a: Median Difference ≠ 0		
Customer Type	**2**	**3**
Count	42	27
Median	4.340	3.510
Mann-Whitney Statistic	1744.00	
P-Value (2-sided, adjusted for ties)	0.0008	

Given the P-Value of .0008 we reject H0 and conclude that Median Customer Satisfaction is significantly different between Customer types 2 and 3. This confirms previous findings and matches the results of the 2 Sample t-Test.

6. The resulting output for 2 Sample KS test is:

2 Sample KS: Overall Satisfaction		
Test Information		
H_0: Distribution (CDF) Difference = 0		
H_a: Distribution (CDF) Difference ≠ 0		
Customer Type	**2**	**3**
Count	42	27
Median	4.340	3.510
KS Statistic	0.4392	
P-Value (2-sided)	0.0035	

Given the P-Value of .0035 we reject H0 and conclude that the Satisfaction distributions are significantly different between Customer types 2 and 3. This agrees with the results for the Mann-Whitney test.

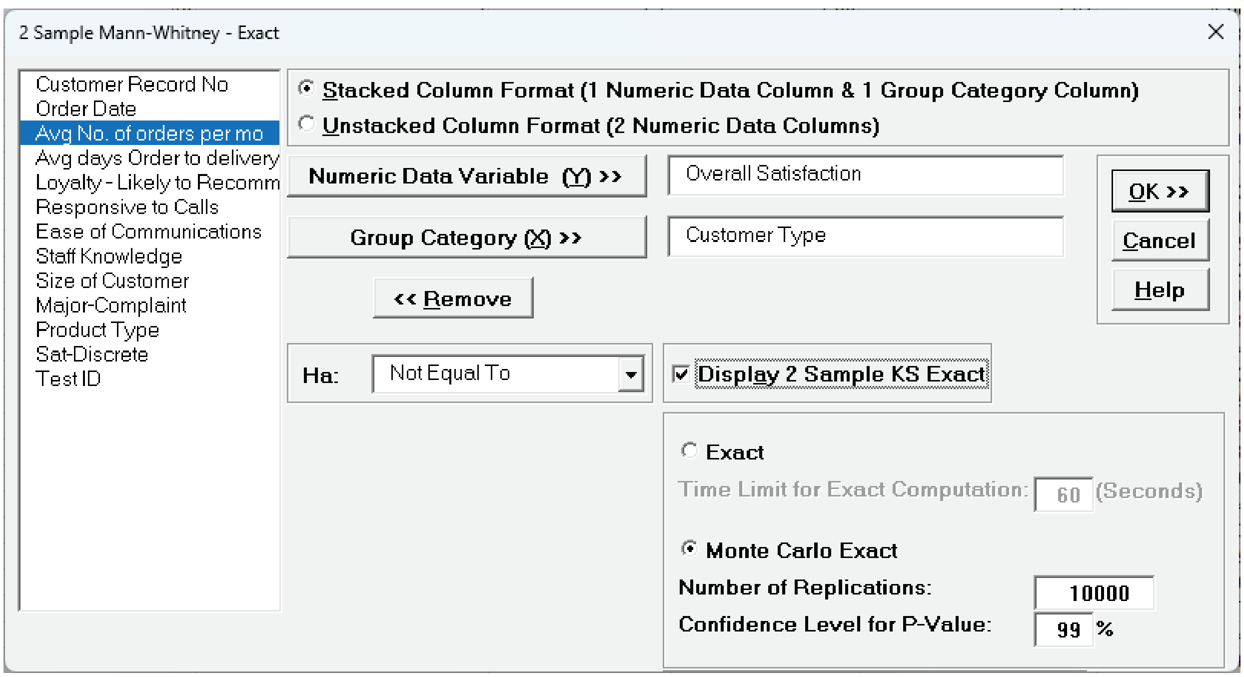
Two Sample Mann-Whitney Test – Exact (with 2 Sample KS Option)

We will now redo the above example to compute exact P-Values. Typically, this would not be necessary unless the sample sizes were smaller (each sample N <= 10 for Mann-Whitney), but this gives us continuity on the example. Due to the large number of permutations, we will use Monte Carlo Exact for this analysis and will consider a small sample problem later.

1. Open **Customer Data.xlsx**, click **Sheet 1** tab (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > 2 Sample Mann-Whitney – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

3. With **Stacked Column Format** checked, select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>** and **Ha:** *Not Equal To*. Select **Monte Carlo Exact** with the default **Number of Replications** = 10000 and **Confidence Level for P-Value** = 99%. Check **Display 2 Sample KS Exact**.



Tip 1: If **Exact** is selected and the computation time limit is exceeded, a dialog will prompt you to use Monte Carlo or to increase the computation time.

Tip 2: 10,000 replications will result in a Monte Carlo P-Value that is typically correct to two decimal places. One million (1e6) replications will result in three decimal places of accuracy and typically require less than 60 seconds to solve for any data set.

Tip 3: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

The KS (Kolmogorov Smirnov) Exact test is only available for Monte Carlo Exact.

4. Click **OK**. Select *Customer Type 2* and *3*. Click **OK**. The resulting output for the 2 Sample Mann-Whitney - Monte Carlo test is:

2 Sample Mann-Whitney - Monte Carlo: Overall Satisfaction		
Test Information		
H_0: Median Difference = 0		
H_a: Median Difference ≠ 0		
Customer Type	**2**	**3**
Count	42	27
Median	4.340	3.510
Mann-Whitney Statistic	1744.00	
Monte Carlo P-Value (2-sided)	0.0004	
Monte Carlo P-Value 99% CI Upper	0.0009	
Monte Carlo P-Value 99% CI Lower	0.0000	

Given the Monte Carlo P-Value of .0004 we reject H0 and conclude that Median Customer Satisfaction is significantly different between Customer types 2 and 3. The Monte Carlo P-Value is very close to the above "large sample" or "asymptotic" result. This was expected because the sample size is reasonable (each sample N > 10), so the "large sample" P-Values are valid using a normal approximation for the Mann-Whitney Statistic.

The Monte Carlo P-Value 99% confidence interval is 0.0000 to 0.0009. Note that the Monte Carlo P-Value will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). This was demonstrated using 10,000 replications, but with a P-Value this low, it is recommended that the number of replications be increased to 1e5 or 1e6 to get a better estimate.

5. The resulting output for the 2 Sample KS - Monte Carlo test is:

2 Sample KS - Monte Carlo: Overall Satisfaction		
Test Information		
H_0: Distribution (CDF) Difference = 0		
H_a: Distribution (CDF) Difference ≠ 0		
Customer Type	**2**	**3**
Count	42	27
Median	4.340	3.510
KS Statistic	0.439153439	
Monte Carlo P-Value (2-sided)	0.0040	
Monte Carlo P-Value 99% CI Upper	0.0091	
Monte Carlo P-Value 99% CI Lower	0.0000	

Given the Monte Carlo P-Value of .004 with 99% confidence interval 0.0000 to 0.0091, we reject H0 and conclude that Satisfaction distributions are significantly different between Customer types 2 and 3. As with the Mann-Whitney, the Monte Carlo P-Value is very close to the above “large sample” or “asymptotic” result.

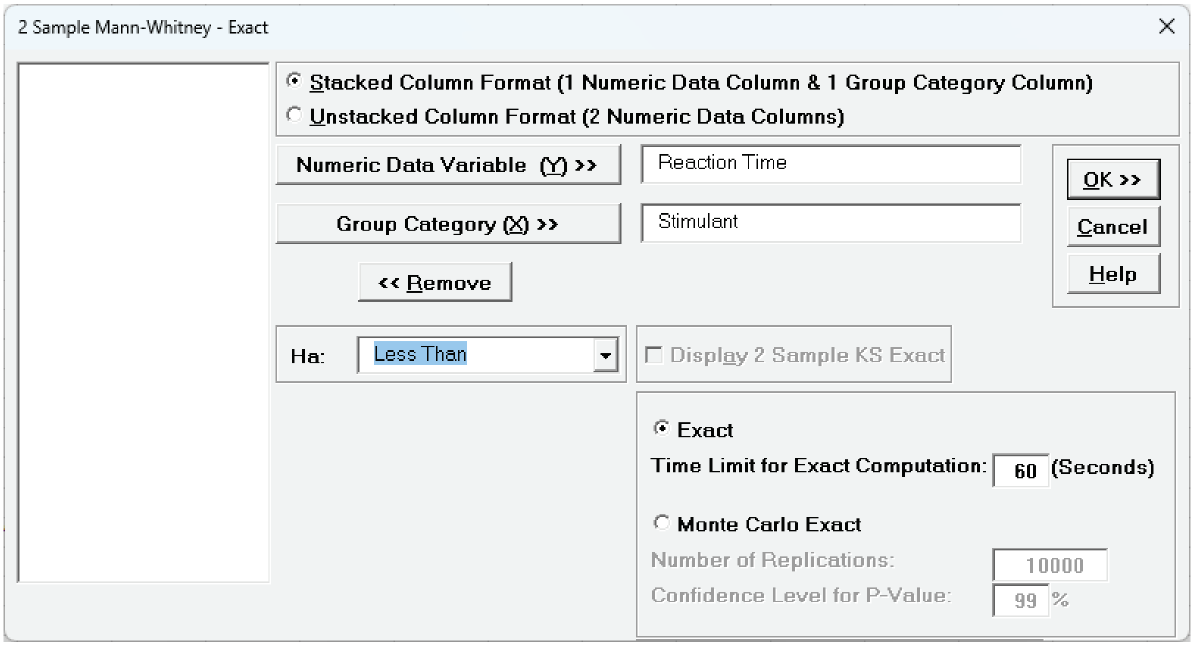
6. Now we will consider a small sample problem. Open **Stimulant Test.xlsx**. This data is from:

   Narayanan, A. and Watts, D. “Exact Methods in the NPAR1WAY Procedure,” SAS Institute Inc., Cary, NC. http://support.sas.com/rnd/app/stat/papers/exact.pdf

   Researchers conducted an experiment to compare the effects of two stimulants. Thirteen randomly selected subjects received the first stimulant, and six randomly selected subjects received the second stimulant. The reaction times are in minutes. We will test the null hypothesis of no difference between the medians of the two stimulants against the alternative that stimulant 1 has smaller median reaction time than stimulant 2.

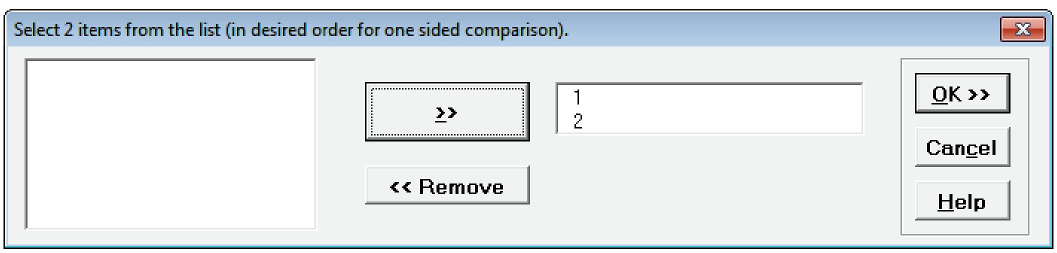
7. Select **Reaction Time** tab. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > 2 Sample Mann-Whitney - Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

8. With **Stacked Column Format** checked, select *Reaction Time*, click **Numeric Data Variable (Y) >>**; select *Stimulant*, click **Group Category (X) >>**; and **Ha:** *Less Than*. Select **Exact** with the default **Time Limit for Exact Computation** = 60 seconds.



The 2 Sample KS Exact option is only available for Monte Carlo, so is greyed out.

9. Click **OK**. Select *Stimulant 1* and *2*.



This sets the order for the one-sided test, so the alternative hypothesis Ha is Median 1 < Median 2.

10. Click **OK**. Results:

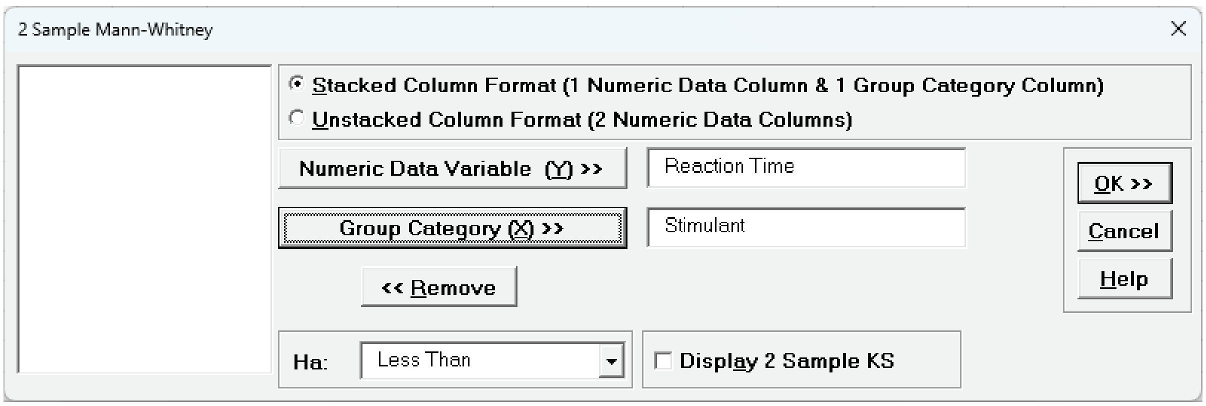
2 Sample Mann-Whitney - Exact - Reaction Time		
Test Information		
H_0: Median Difference = 0		
H_a: Median Difference Less Than 0		
Stimulant	**1**	**2**
Count	13	6
Median	3.270	3.485
Mann-Whitney Statistic	110.50	
Exact P-Value (1-sided)	0.0527	

With the P-Value = .0527 we fail to reject H0, so cannot conclude that there is a difference in median reaction times. This exact P-Value matches that given in the reference paper.

By way of comparison, we will now rerun the analysis using the "large sample" or "asymptotic" Mann-Whitney test.

11. Select **Reaction Time** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests > 2 Sample Mann-Whitney**. If necessary, check **Use Entire Data Table**, click **Next**.

12. With **Stacked Column Format** checked, select *Reaction Time*, click **Numeric Data Variable (Y) >>**; select *Stimulant*, click **Group Category (X) >>**; and **Ha:** *Less Than*. **Display 2 Sample KS** is left unchecked.



13. Click **OK**. Select *Stimulant 1* and *2*. Click **OK**. Results:

2 Sample Mann-Whitney - Reaction Time		
Test Information		
H_0: Median Difference = 0		
H_a: Median Difference Less Than 0		
Stimulant	**1**	**2**
Count	13	6
Median	3.270	3.485
Mann-Whitney Statistic	110.50	
P-Value (1-sided, adjusted for ties)	0.0421	

Now with the P-Value = .0421 we **incorrectly reject** H0.

The difference between exact and large sample P-Value is small but it was enough to lead us to falsely conclude that stimulant 1 resulted in a reduced median reaction time.

In conclusion, whenever you have a small sample size and are performing a nonparametric test, always use the Exact option.

Part J – One-Way ANOVA & Means Matrix

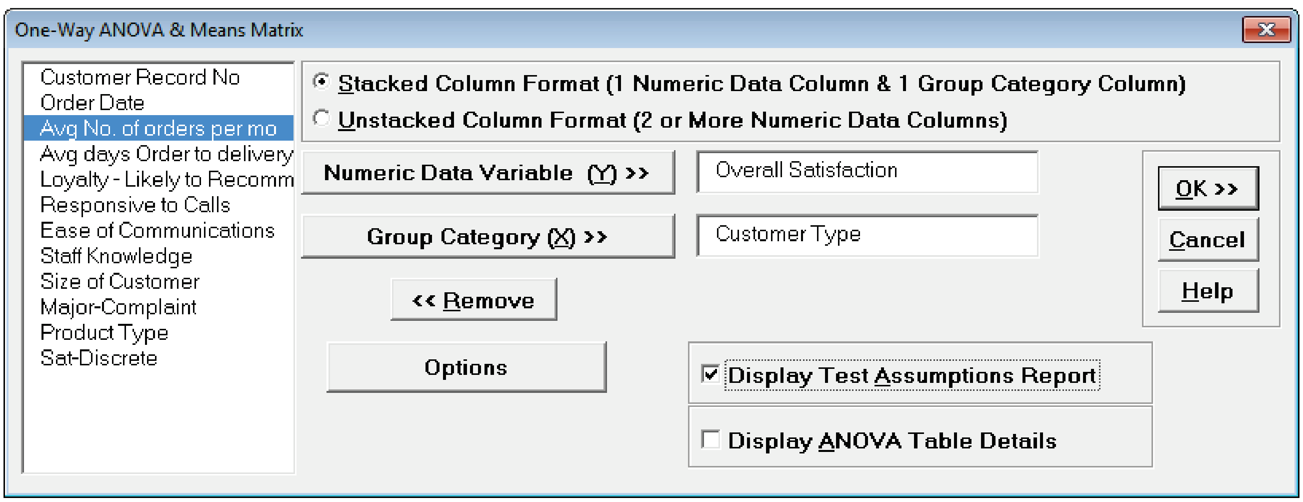
One-Way ANOVA & Means Matrix

1. One-Way ANOVA and Means Matrix allows you to quickly do multiple pairwise comparisons. The One-Way ANOVA tests H0: µ1 = µ2 = µ3; Ha: at least one pairwise set of means are not equal.

2. Open **Customer Data.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).

3. Click **SigmaXL > Statistical Tools > One-Way ANOVA & Means Matrix**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

4. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category** (**X) >>**. Uncheck **Display ANOVA Table Details**. Check **Display Test Assumptions Report**.



5. Click **OK**. The results are shown below:

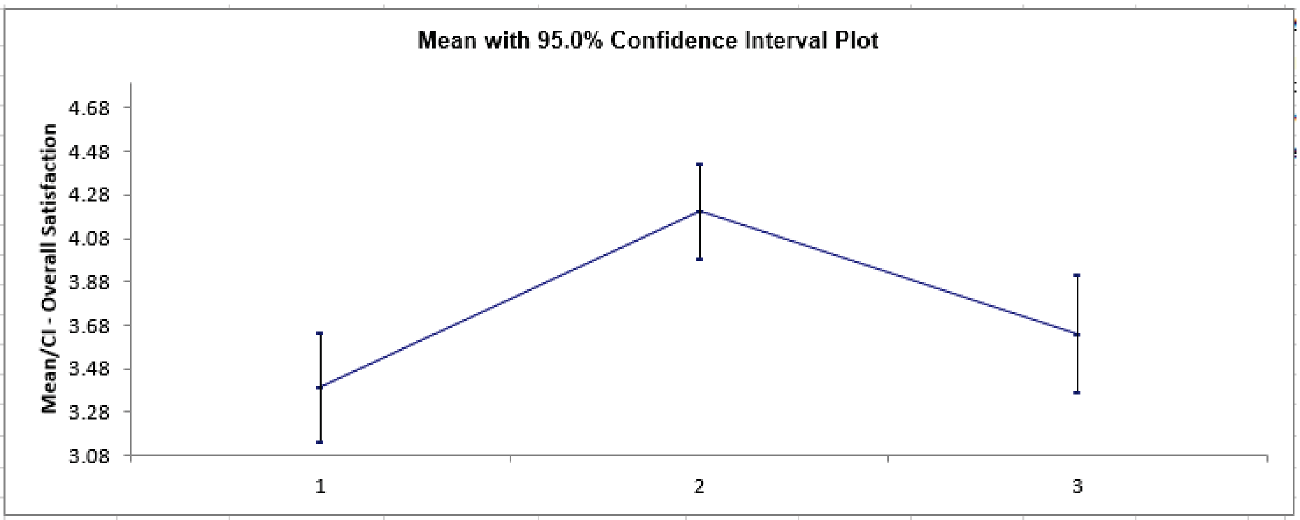
One-Way ANOVA & Means Matrix: Overall Satisfaction			
H_0: Mean 1 = Mean 2 = ... = Mean k			
H_a: At least one pair Mean i ≠ Mean j			

Customer Type	1	2	3
Count	31	42	27
Mean	3.394	4.205	3.641
Standard Deviation	0.824680	0.621200	0.670478
UC (2-sided, 95.0%, pooled)	3.644	4.420	3.910
LC (2-sided, 95.0%, pooled)	3.143	3.990	3.373

Pooled Standard Deviation =	0.702810		R-Sq =	20.95%
DF =	97		R-Sq adj. =	19.32%
F =	12.856			
P-Value =	0.0000			

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.811690	-0.247563
2		0	0.564127
3			0

Fisher Pairwise Probabilities	1	2	3
1		0.0000	0.1840
2			0.0016
3			



6. The ANOVA P-Value of 0.0000 tells us that at least one pairwise set of means are not equal. From the Pairwise Mean Difference (Means Matrix), we conclude that Mean Overall Satisfaction is significantly different between Customer Type 2 and 3, as well as 1 and 2. Note that the default probabilities are Fisher Pairwise. The P-Values will be slightly different than the previous 2 Sample t-Test results because the variances from all 3 customer types are “pooled” here. This also results in slightly different confidence intervals. See below for more details on the multiple comparison options.

7. A graphical view of the Overall Satisfaction Mean and 95% Confidence Intervals are given to complement the Means Matrix. The fact that the CI’s for Customer Type 2 do not overlap those of Type 1 or 3, clearly shows that Customer Type 2 has a significantly higher mean satisfaction score. The overlap of CI’s for Type 1 and 3 shows that the mean satisfaction scores for 1 and 3 are not significantly different.

8. Note that the Confidence Level shown in the graph can be modified by clicking the **Options** button and setting **Confidence Level**. This also changes the alpha value used to highlight the P-Values in red:

alpha = (100 – Confidence Level)/100

so the default Confidence Level = 95.0%, gives an alpha = 0.05.

9. The R-Square (R-Sq) value of 20.95% indicates that Customer Type "explains" approximately 21% of the variation in Overall Satisfaction. We need to "drill down" to understand the root causes and best practices associated with Customer Type 2.

10. One-Way ANOVA Assumptions Report:

One-Way ANOVA Assumptions Report			
Normality:	Anderson Darling P-Value = 0.531. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.	Anderson Darling P-Value = 0.030. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.9680 and Kurtosis value = 0.6796. See robustness and outliers.	Anderson Darling P-Value = 0.360. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.
Robustness:	Minimum sample size for a robust ANOVA test = 2. Since each sample size is greater than this, the ANOVA test is robust to nonnormality.		
Outliers (Boxplot Rules):	No outliers found.	Potential (1.5*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 0.066. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.	Nonparametric Runs Test (Exact) P-Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.	Nonparametric Runs Test (Exact) P Value = 1.000. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.
Equal Variance:	Levene's test for Equal Variances P-Value = 0.115. Fail to reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is not violated.		

This is a text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation).

Each sample is tested for Normality using the Anderson Darling test. If not normal, the minimum sample size for robustness of the ANOVA Test is determined utilizing Monte Carlo regression equations (see **Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA**). If the sample size is inadequate, a warning is given and the appropriate Nonparametric test is recommended (Kruskal-Wallis if there are no extreme outliers, Mood's Median if there are extreme outliers).

Each sample is tested for Outliers defined as: Potential: Tukey's Boxplot (> Q3 + 1.5*IQR or < Q1 – 1.5*IQR); Likely: Tukey's Boxplot 2.2*IQR; Extreme: Tukey's Boxplot 3*IQR. If outliers are present, a warning is given and recommendation to review the data with a Boxplot and Normal Probability Plot. Here we have a potential outlier for Customer Type 2.

Tip: If the removal of outlier(s) result in an Anderson Darling P-Value that is > 0.1, a notice is given that excluding the outlier(s), the sample data are inherently normal.

Each sample is tested for Randomness using the Exact Nonparametric Runs Test. If the sample data is not random, a warning is given and recommendation to review the data with a Run Chart.

A test for Equal Variances is also applied. If all sample data are normal, Bartlett's Test is utilized, otherwise Levene's Test is used. If the variances are unequal, then a warning is given and Welch's ANOVA is recommended.

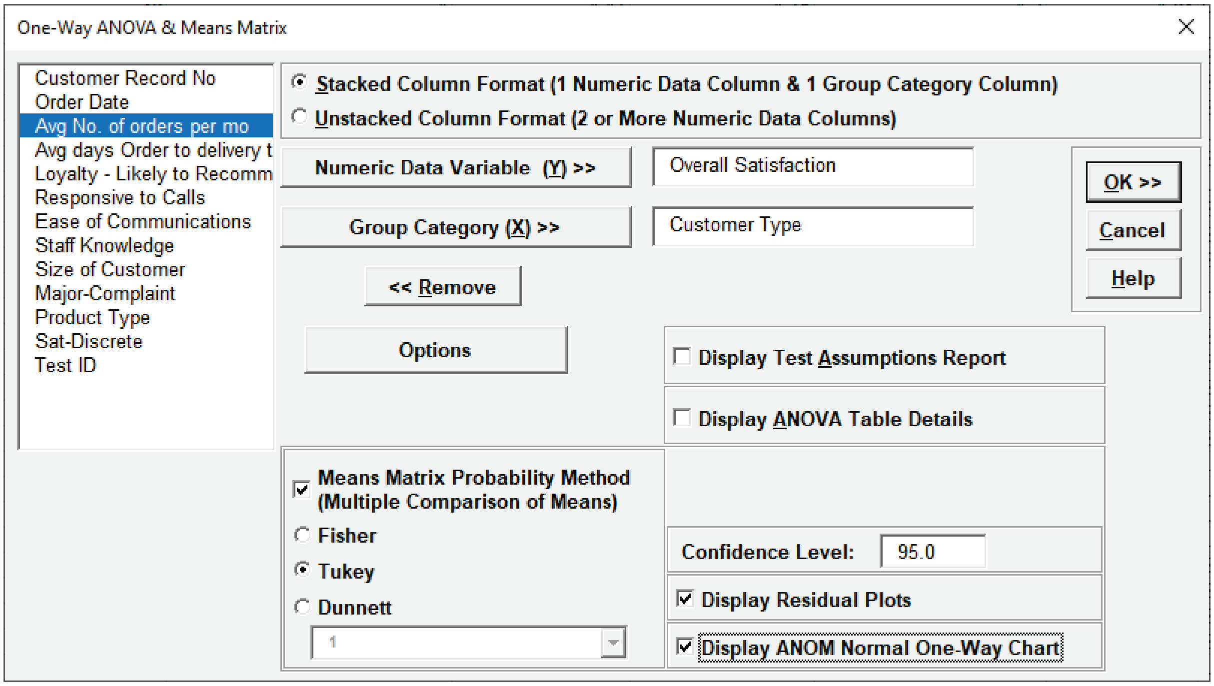
See Appendix **Hypothesis Test Assumptions Report** for further details.

One-Way ANOVA & Means Matrix - Options

Notes on Means Matrix Probability Method (Multiple Comparison of Means) and ANOM Chart:

- Fisher:
  - Also known as Fisher's Least Significant Difference (LSD)
  - Pairwise 2 sample t-tests with pooled standard deviation
  - Does not correct for family-wise error rate, so should only be used for k = 3 means and in the restricted case where the ANOVA P-Value is < alpha (this is also known as Protected Fisher LSD). For k = 3 means, Protected Fisher LSD is more powerful than Tukey.
- Tukey:
  - Similar to LSD, uses pairwise tests with pooled standard deviation, but is a studentized range statistic that corrects for family-wise error rate. Recommended for k > 3.
- Dunnett with Control:
  - If one of the groups are a control reference group, Dunnett with Control is more powerful than Tukey because it is doing fewer pairwise comparisons (only considers those pairwise against the control group).
  - Uses pooled standard deviation and a multivariate t distribution that corrects for family-wise error rate.
- See Appendix **Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests)** for further details and references.
- Display ANOM Normal One-Way Chart:
  - The ANOM alpha = (100 – Confidence Level)/100.
  - This chart is also available at **SigmaXL > Graphical Tools > Analysis of Means (ANOM)**. See **Part P – Analysis of Means (ANOM) Charts**.

1. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Uncheck **Display Test Assumptions Report**. Click the **Options** button. Select **Tukey**. Check **Display Residual Charts** and **Display ANOM Normal One-Way Chart**; **Confidence Level** = 95.0 as shown:



Note: The **Confidence Level** is used to set the level in the Mean/Confidence Interval Plot, the alpha level (alpha = (100 – CI)/100) used to highlight the P-Values and the alpha level for the ANOM chart. However, the confidence level used in the Residuals Normal Probability Plot is always 95%.

2. Click **OK**. The Pairwise Means Difference (Means Matrix) and Tukey Probability results are:

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.811690	-0.247563
2		0	0.564127
3			0

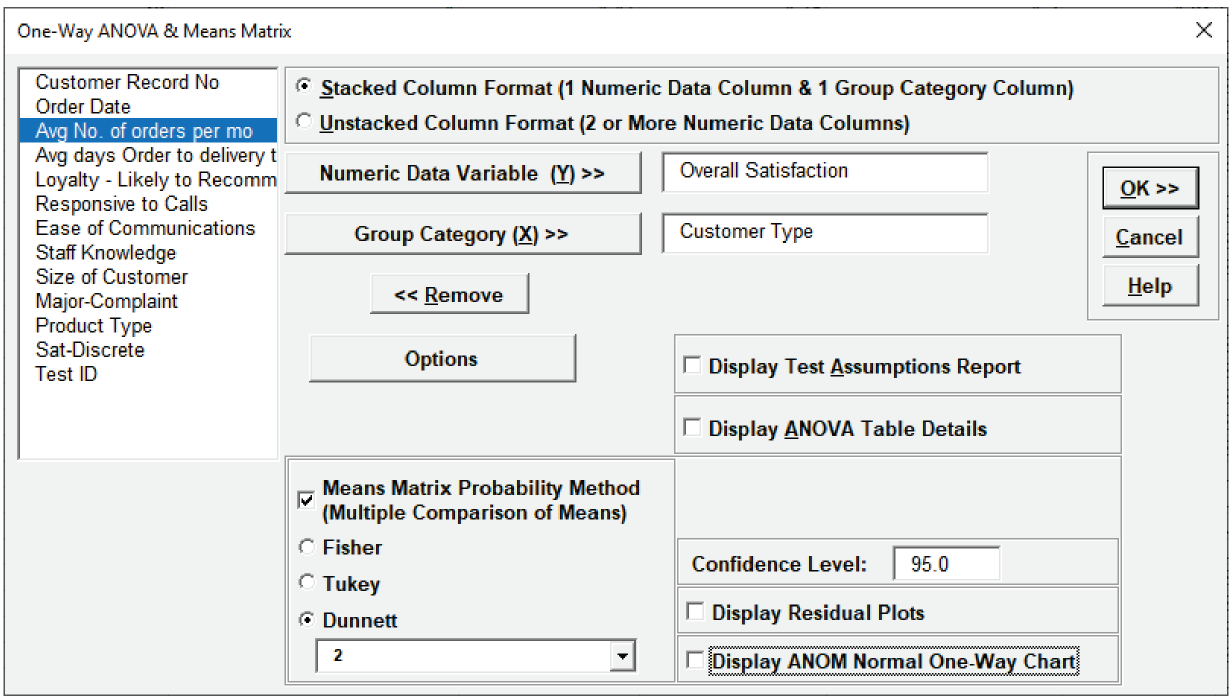
Tukey Probabilities	1	2	3
1		0.0000	0.3777
2			0.0044
3			

3. The significant Tukey Probability values above have not changed as compared to Fisher, but note that they are larger than Fisher to compensate for the family-wise error rate:

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.811690	-0.247563
2		0	0.564127
3			0

Fisher Pairwise Probabilities	1	2	3
1		0.0000	0.1840
2			0.0016
3			

4. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Click **Options**. Uncheck all **Display** options and select **Dunnett** with Control Level = 2 (i.e. we are treating Customer Type 2 as the Control Group):



5. Click **OK**. The Pairwise Means Difference (Means Matrix) and Dunnett Probability results are:

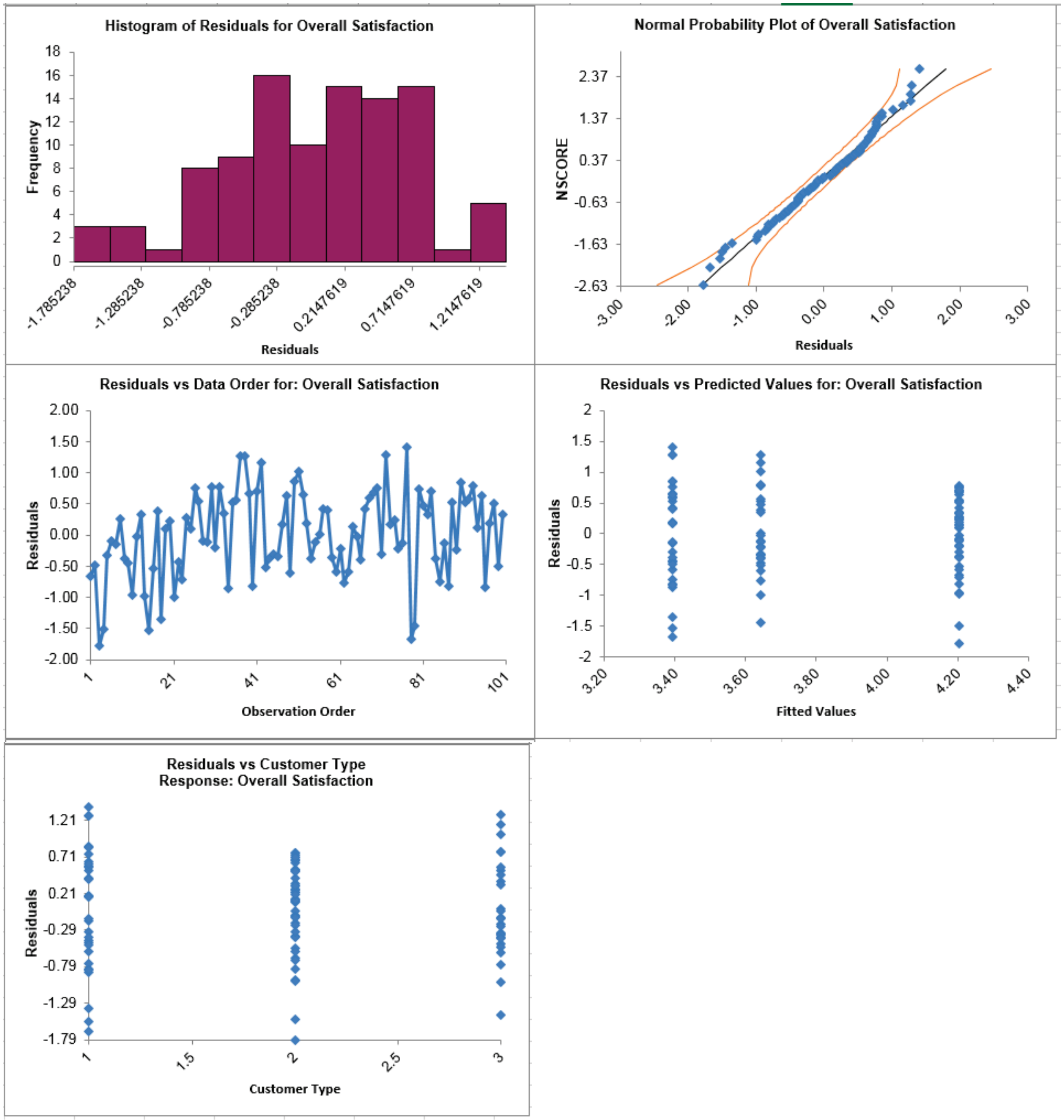
Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.811690	-0.247563
2		0	0.564127
3			0

Dunnett Probabilities (Control = 2)	1	2	3
1		0.0000	
2			0.0031
3			

All three mean differences are shown but only two probability values are calculated: Customer Type 1 – Control Type 2 and Control Type 2 – Type 3. (Typically, one would show 1 – 2 and 3 – 2 for Dunnett, but we are displaying the upper triangle for consistency with the other options).

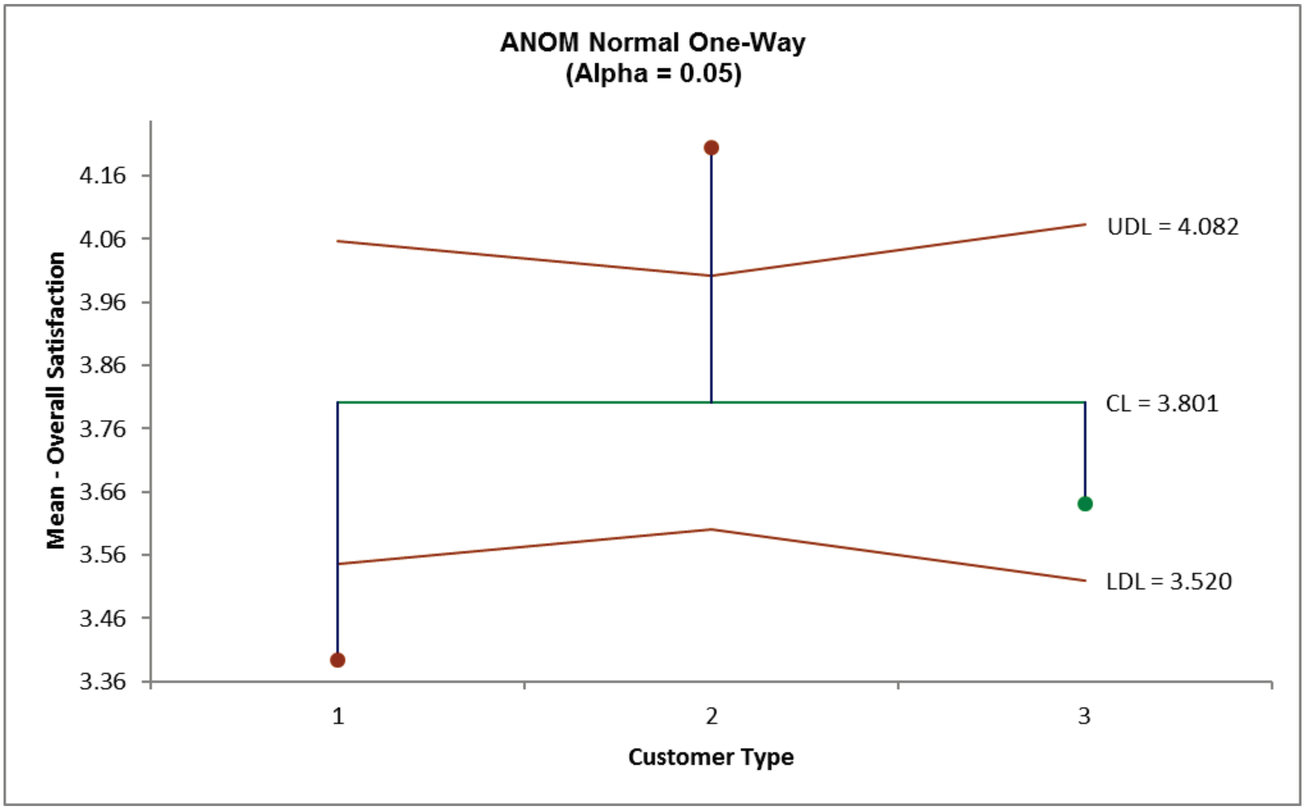
Note that the Dunnett 2 – 3 P-Value is .003 which is smaller than the Tukey .004 but larger than the Fisher .002. Dunnett is more powerful than Tukey (due to having fewer tests), but less powerful than Fisher. It does, however, have an advantage over Fisher because it protects the family-wise error rate for the comparisons being made.

6. Click the **ANOVA Residuals** sheet tab to display the Residual Plots (created at steps 1, 2):



Residuals are the unexplained variation from the ANOVA model (Actual – Predicted/Fitted values). We expect to see the residuals approximately normally distributed with no obvious patterns in the above graphs, which is the case here.

7. Click **ANOM_Normal_One_Way** sheet tab to display the ANOM Chart (created at steps 1, 2):



Here we see that Customer Type 1 mean satisfaction score is significantly below the overall mean and Customer Type 2 is significantly higher. This is consistent with the results that are observed in the Pairwise Means Difference (Means Matrix), but is easier to interpret.

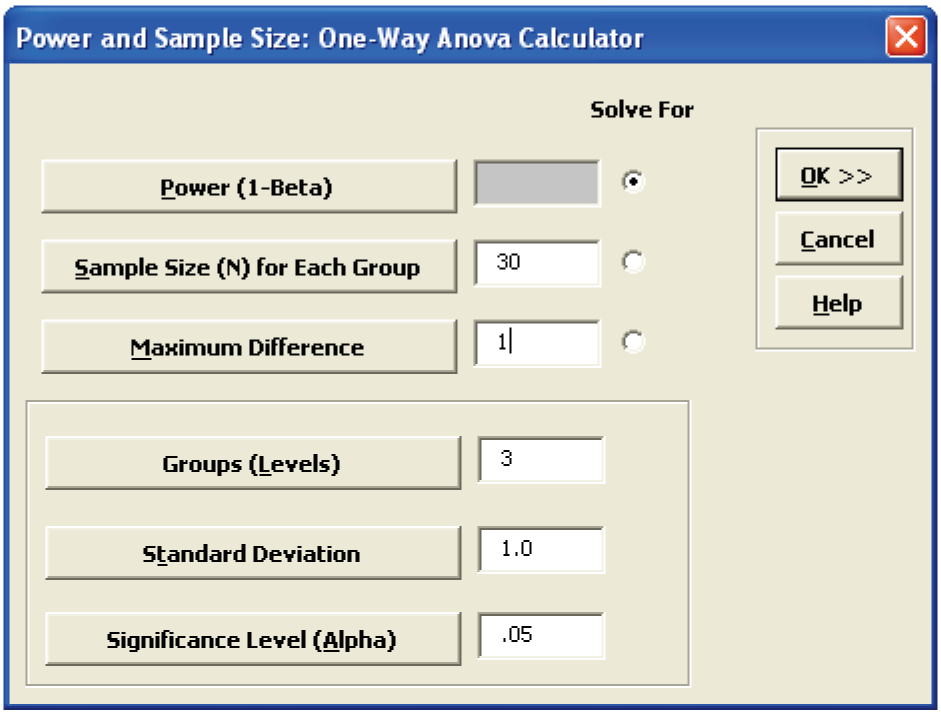
The varying decision limits are due to the varying sample sizes for each Customer Type, with smaller sample size giving wider limits in a manner similar to a control chart. If the data are balanced the decision limit lines will be constant.

Power & Sample Size for One-Way ANOVA

To determine Power & Sample Size for a One-Way ANOVA, you can use the Power & Sample Size Calculator or Power & Sample Size with Worksheet.

1. Click **SigmaXL > Statistical Tools > Power & Sample Size Calculators > One-Way ANOVA Calculator**.

2. Select **Solve For Power (1 – Beta)**. Enter **Sample Size** and **Maximum Difference** as shown:



Note that we are calculating the power or likelihood of detection given that the maximum difference between group means = 1, with sample size for each group = 30, 3 groups, standard deviation = 1, significance level = .05, and Ha: Not Equal To (two-sided test).

3. Click **OK**. The resulting report is displayed:

Power and Sample Size: One-Way ANOVA

H0: Mean 1 = Mean 2 = ... = Mean k
Ha: At least one pair Mean i ≠ Mean j
Solve For: Power (1 - Beta)

Sample Size (N)	Maximum Difference	Groups	Standard Deviation	Significance Level (Alpha)	Power (1 - Beta)
30	1	3	1	0.05	0.936276828

A power value of 0.94 is acceptable. Note that this value is less than the power value of 0.97 obtained with the two-sample t-Test.

4. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Change the number of groups to 4. Click **OK**. Note that the power value is 0.907. If the number of groups (levels) increases, you will have to increase the sample size in order to maintain statistical power.

5. To determine Power & Sample Size using a Worksheet, click **SigmaXL > Statistical Tools > Power & Sample Size with Worksheet > One-Way ANOVA**.

6. A graph showing the relationship between Power, Sample Size and Maximum Difference can then be created using **SigmaXL > Statistical Tools > Power & Sample Size Chart**. See Part E for an example using the 1 Sample t-Test.

Part K – Two-Way ANOVA

Two-Way ANOVA

Two-Way ANOVA tests the following:

H0 (Factor X1): $\mu_1 = \mu_2 = \ldots = \mu_k$
Ha (Factor X1): at least one pairwise set of means are not equal ($\mu_i \neq \mu_j$);

H0 (Factor X2): $\mu_1 = \mu_2 = \ldots = \mu_k$
Ha (Factor X2): at least one pairwise set of means are not equal ($\mu_i \neq \mu_j$);

H0 (Interaction): There is **no** interaction between factors X1 and X2
Ha (Interaction): There is an interaction between factors X1 and X2.

A Two-Way ANOVA analysis will typically have balanced data from a designed experiment, with an equal number of observations for each combination level of X1 and X2. SigmaXL will also accommodate unbalanced data. The minimum requirement is one observation per combination level of X1 and X2. An error message will be produced if this minimum requirement is not met.

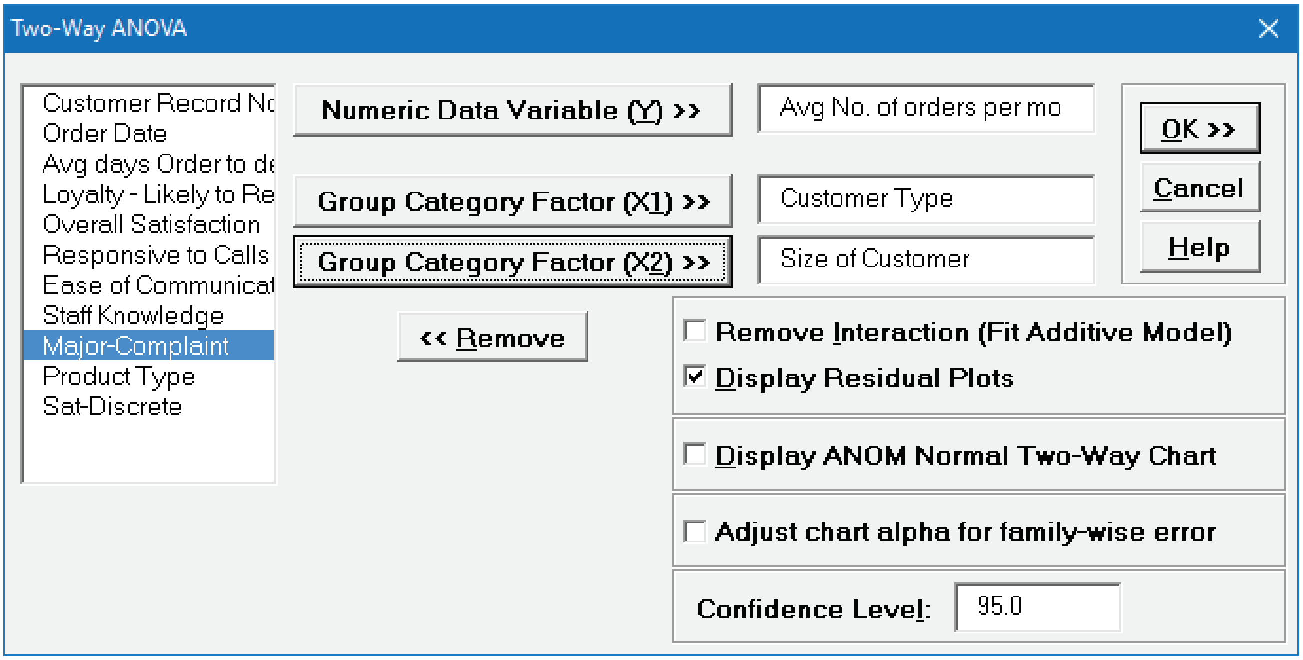
1. Open **Customer Data.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Statistical Tools > Two-Way ANOVA**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Select *Avg No of Orders per Mo,* click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category Factor (X1) >>**; select *Size of Customer*, click **Group Category Factor (X2) >>. Remove Interaction (Fit Additive Model)** should remain unchecked.

   Tip: If the Two-Way ANOVA report gives a P-Value for the interaction that is high (> 0.1), you should then press **F3** or click **Recall SigmaXL Dialog** to recall this dialog, check **Remove Interaction (Fit Additive Model)** and rerun the analysis.

   Display ANOM Normal Two-Way Chart should be unchecked. We will discuss this tool later in Part P – Analysis of Means (ANOM) Charts. Note that this option is only available when **Remove Interaction (Fit Additive Model)** is unchecked because ANOM Normal Two-Way always includes the interaction term in the model. The ANOM alpha is (100 – Confidence Level)/100).



4. Click **OK**. The results are shown below:

Two-Way ANOVA: Avg No. of orders per mo

H0 (Factor Customer Type): Mean 1 = Mean 2 = ... = Mean k
Ha (Factor Customer Type): At least one pair Mean i ≠ Mean j

H0 (Factor Size of Customer): Mean 1 = Mean 2 = ... = Mean k
Ha (Factor Size of Customer): At least one pair Mean i ≠ Mean j

H0 (Interaction): There is no interaction between factors X1 and X2
Ha (Interaction): There is an interaction between factors X1 and X2

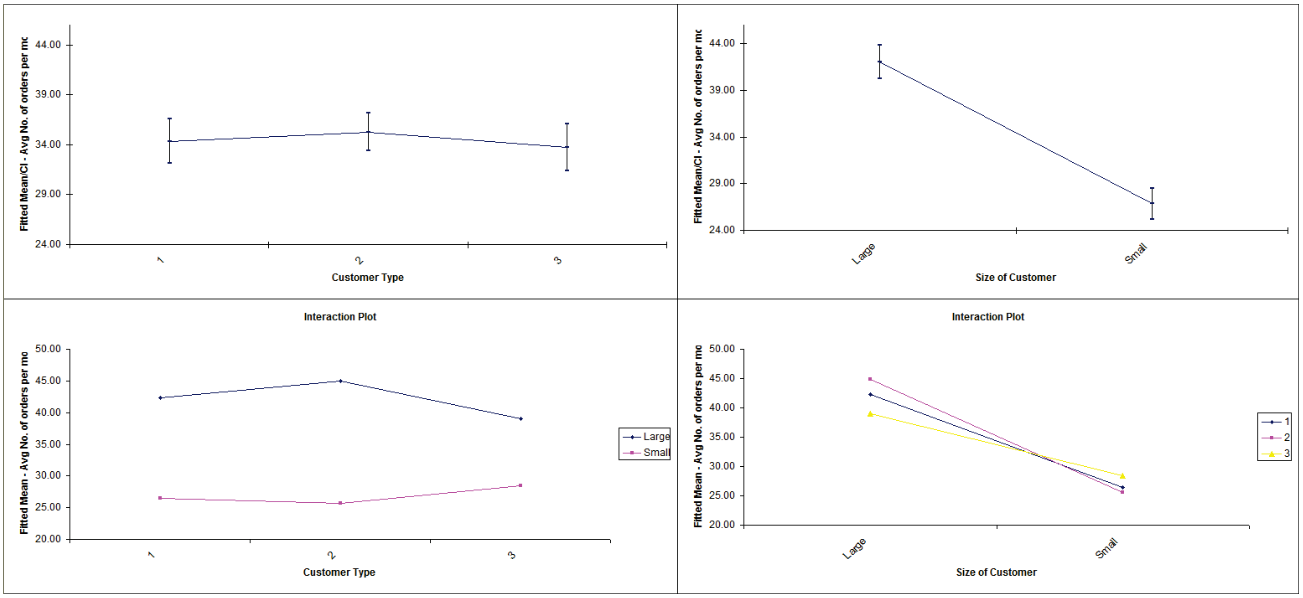
Factor and Model Summary:

Number of Levels - Customer Type	3
Number of Levels - Size of Customer	2
Number of Replicates	N/A
Design Type:	Unbalanced
Confidence Level	95.00%
R-Square	65.31%
R-Square Adjusted	63.46%
S (Pooled Standard Deviation)	6.162

Analysis of Variance:

Source	DF	SS	MS	F	P
Customer Type	2	39.443	19.721	0.519375	0.5966
Size of Customer	1	5298.9	5298.9	139.55	0.0000
Interaction	2	294.72	147.36	3.881	0.0240
Error	94	3569.3	37.971		
Total	99	10288	103.92		

5. Scroll down to view the Main Effects (with confidence intervals) and Interaction Plots:

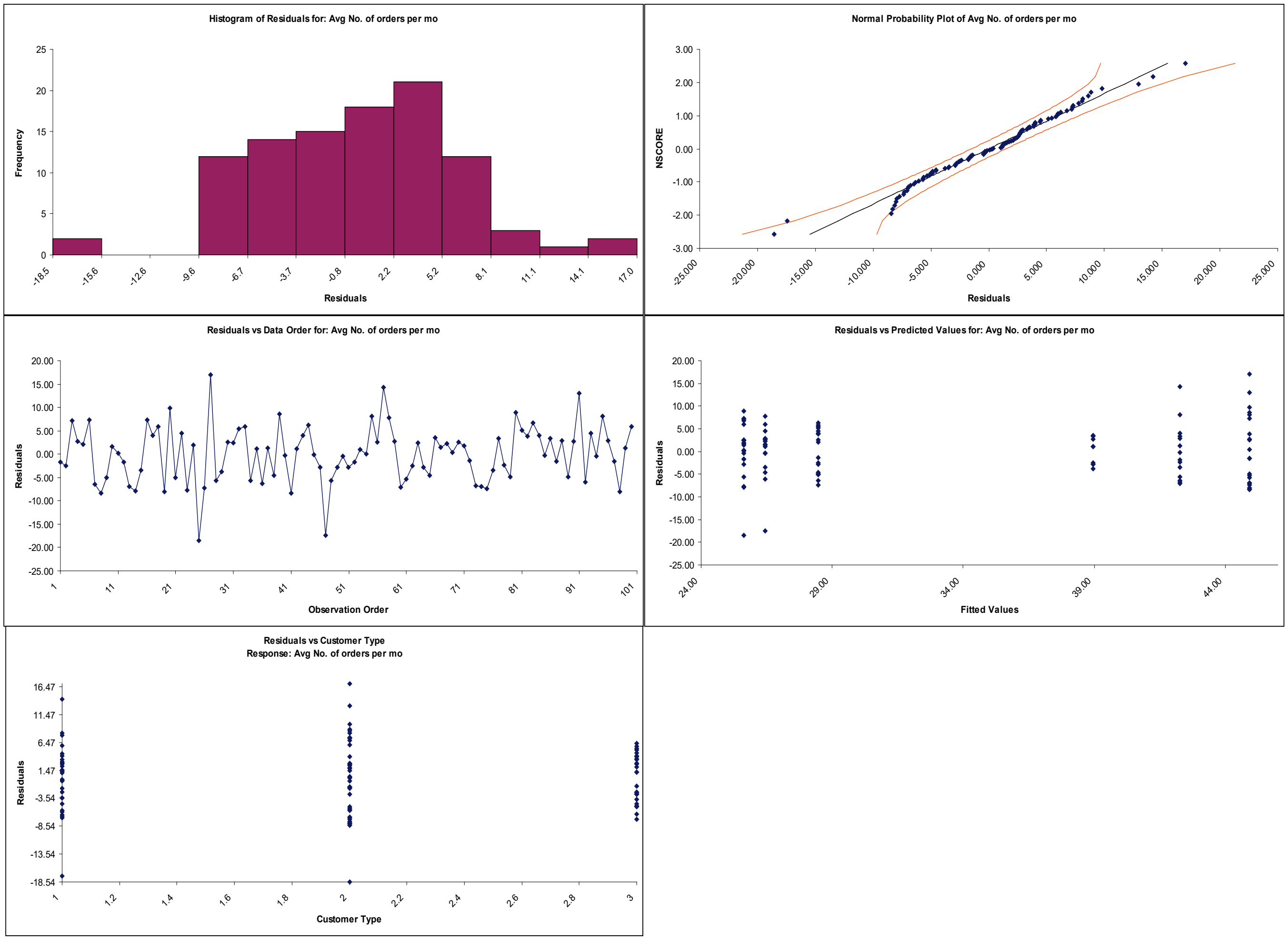


Note that the mean values shown are fitted (predicted least squares) means not data means. This compensates for unbalanced data.

6. From the ANOVA table, we can see that the Size of Customer term is significant with a P-Value less than .05. Customer Type is not significant by itself, but the Interaction term is significant. This indicates that the effect of Size of Customer on Average Number of Orders per Month depends on Customer Type (we could also equivalently say that the effect of Customer Type depends on Size of Customer).

7. This is also confirmed looking at the Main Effects and Interaction plots. Customer Type by itself is not significant. Size of Customer is obviously significant. Looking at the Interaction plot, the different slopes illustrate that the change in Average Number of Orders per Month across Customer Types depends on Customer Size, albeit this is a relatively small effect.

8. Click the **Two-Way Residuals** sheet to view the residual graphs:



Residuals are the unexplained variation from the ANOVA model (Actual – Predicted/Fitted values). We expect to see the residuals approximately normally distributed with no obvious patterns in the above graphs, which is the case here.

The Residuals versus Size of Customer graph is not shown because Size of Customer is text. In order to display this plot, Size could be coded numerically with 1 = Small and 2 = Large. Simply create a new column called Size-Coded and use the following Excel formula to create the coded values for the first record:

=IF(K2="Small",1,IF(K2="Large",2))

Copy and Paste this formula to obtain coded values for all 100 records. Rerun the Two-Way ANOVA analysis to create the residual graphs

Part L – Tests for Equal Variance & Welch's ANOVA

Bartlett's Test

Bartlett's Test is similar to the 2 sample F-Test (**SigmaXL > Statistical Tools > 2 Sample Comparison Test**) but allows for multiple group comparison of variances (or standard deviations). Like the F-Test, Bartlett's requires that the data from each group be normally distributed but is more powerful than Levene's Test.

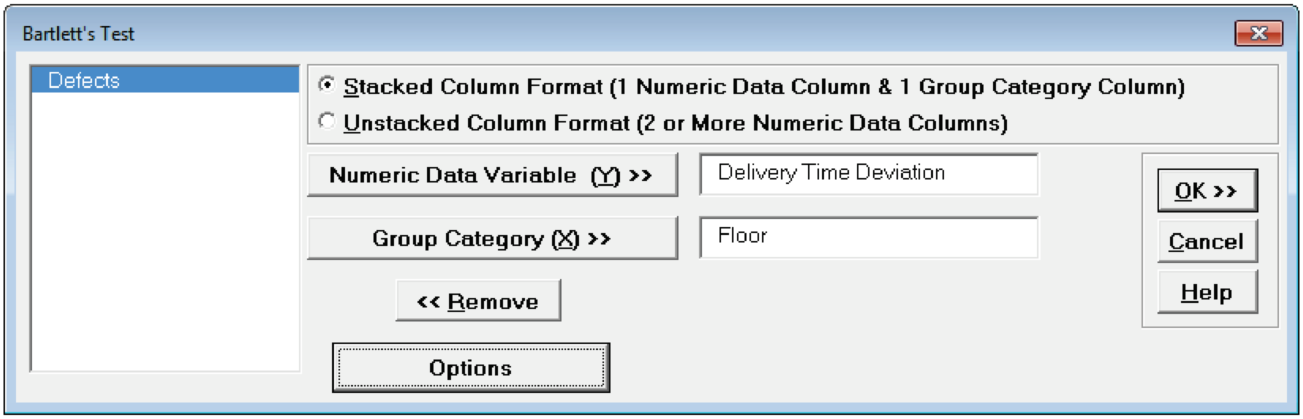
Notes on Multiple Comparison of Variances Probability Method and ANOM Chart:

- F-Test Pairwise
    - Pairwise 2 sample F-tests
    - Does not correct for family-wise error rate, so should only be used for k = 3 groups and in the restricted case where the Bartlett P-Value is < alpha.
- F-Test with Bonferroni Correction
    - Pairwise 2 sample F-tests with Bonferroni correction
    - Recommended for k > 3
    - Bonferroni P-Value' = P-Value * m
        - m = number of pairwise comparisons k(k-1)/2
- See Appendix Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests) for further details and references.
- Display ANOM Variances Chart:
    - The ANOM alpha = (100 – Confidence Level)/100.
    - This chart is also available at **SigmaXL > Graphical Tools > Analysis of Means (ANOM)**. See Part P – Analysis of Means (ANOM) Charts.

1. Open **Delivery Times.xlsx**, click on **Sheet 1** tab.

2. Click **SigmaXL > Statistical Tools > Equal Variance Tests > Bartlett**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**; select *Floor*, click **Group Category (X) >>**.



4. Click **OK**. The results are shown below:

Bartlett's Test For Equal Variance: Delivery Time Deviation										
(Use with normal data)										
Test Information										
H_0: Variance 1 = Variance 2 = ... = Variance k										
Ha: At least one pair Variance i ≠ Variance j										
Floor	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Count	73	73	73	73	73	72	72	72	72	72
Mean	4.989	6.064	6.242	5.581	6.250	8.004	6.647	3.649	6.539	6.084
Median	4.847	6.411	6.476	6.061	7.148	7.261	5.906	3.619	6.799	6.393
StDev	7.019	7.612	7.782	6.827	7.608	6.223	7.015	6.682	7.634	6.653
AD Normality Test P-Value	0.8084	0.9515	0.4024	0.7642	0.0693	0.4844	0.1543	0.4230	0.4014	0.1475
Bartlett's Test Statistic	7.073									
P-Value	0.6295									

5. All 10 Anderson-Darling Test P-Values are > .05 indicating that all group data are normal. Since the assumption of normality is met, Bartlett's is the appropriate test to use. If any one of the groups have a low P-Value for the Normality test, then Levene's test should be used.

6. With the P-Value = 0.63 we fail to reject H0; we do not have evidence to show that the group variances are unequal (practically speaking we will assume that the variances are equal).

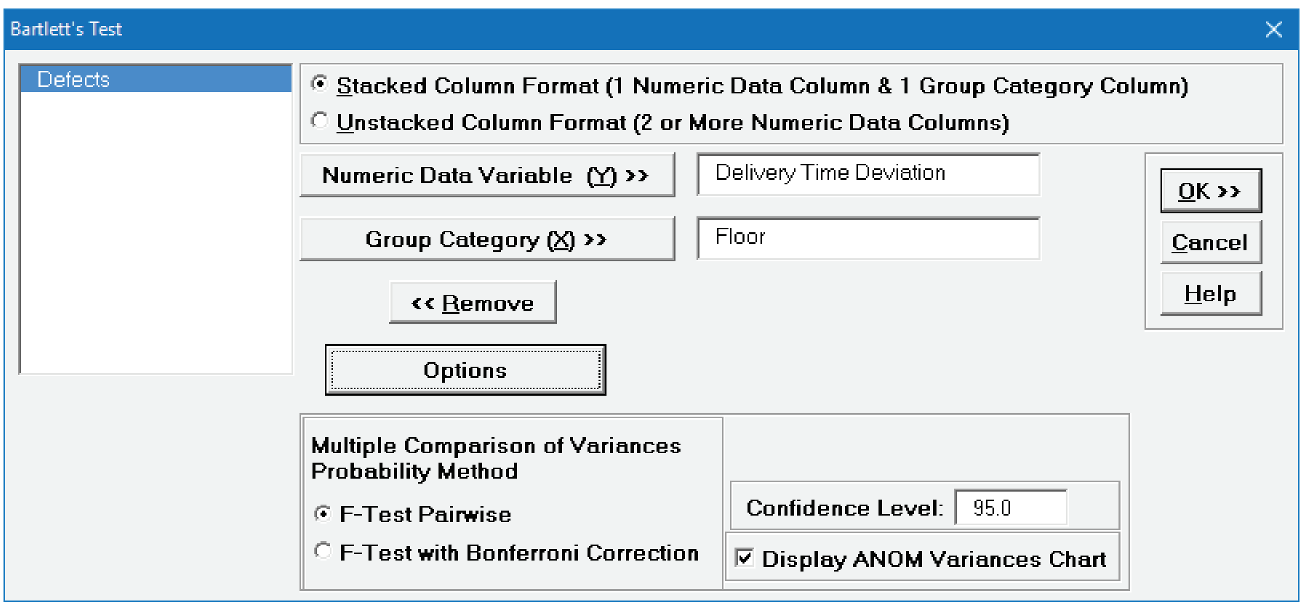
7. If the Equal Variance test is only being used to test the assumption for use in ANOVA then it is not necessary to examine the Multiple Comparison of Variances. However, in the context of a process improvement project, we often do want to know which groups are significantly different. This can give us important clues to identify opportunities for variance reduction.

8. With a "Fail-to-Reject H0" it is unnecessary to review the Multiple Comparison of Variances or the ANOM Variances Chart, but we will do so here for demonstration purposes.

9. The default Multiple Comparison of Variances is a matrix of F-Test Pairwise Probabilities:

Multiple Comparison of Variances										
F-Test Pairwise Probabilities	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1		0.4934	0.3832	0.8147	0.4958	0.3110	0.9967	0.6789	0.4790	0.6518
2			0.8513	0.3583	0.9970	0.0909	0.4925	0.2735	0.9798	0.2575
3				0.2689	0.8483	0.0607	0.3828	0.2003	0.8717	0.1875
4					0.3602	0.4352	0.8185	0.8568	0.3467	0.8275
5						0.0916	0.4948	0.2751	0.9768	0.2591
6							0.3146	0.5498	0.0871	0.5749
7								0.6830	0.4781	0.6558
8									0.2642	0.9703
9										0.2486
10										

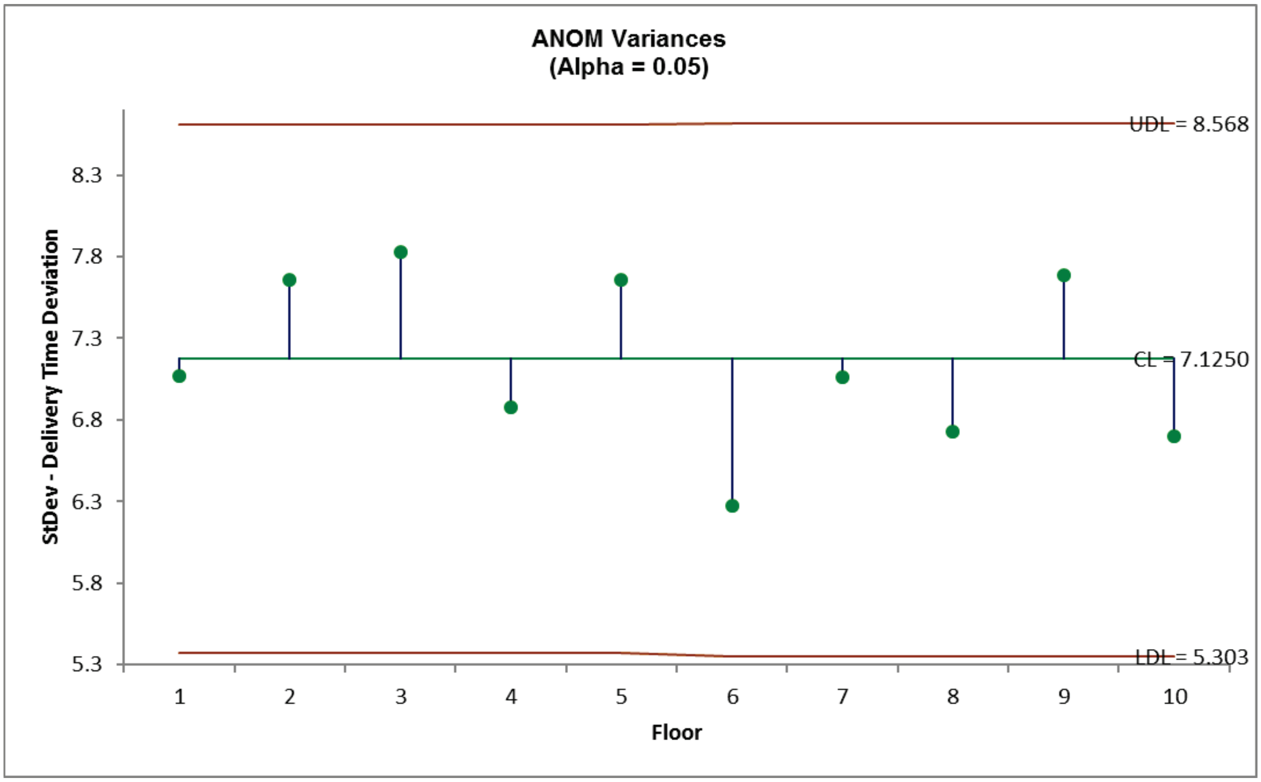
10. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Click **Options**. Check **Display ANOM Variances Chart**.



Note: The **Confidence Level** determines the alpha level (alpha = (100 – CI)/100) used to highlight the P-Values and the alpha level for the ANOM chart. However, the alpha level used to highlight P-Values in the Anderson-Darling Normality Test is always 0.05.

11. We will not run **F-Test with Bonferroni Correction** in this example, but typically that would be used when there are more than 3 groups.

12. Click **OK**. Click **ANOM_Variances** sheet tab to display the ANOM Chart:



The ANOM Variances chart visually shows that none of the group standard deviations are significantly different from the grand mean of all the standard deviations. It is called an ANOM Variances Chart but displays Standard Deviations for ease of interpretation (similar to a Standard Deviation S Control Chart). This does, however, result in non-symmetrical decision limits. The ANOM Variances chart in **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Variances** has an option to display Variances.

Levene's Test

Levene's Test for multiple group comparison of variances is less powerful that Bartlett's Test, but is robust to the assumption of normality. (This is a modification of the original Levene's Test, sometimes referred to as the Browne-Forsythe Test that uses Absolute Deviations from the Median or ADM).

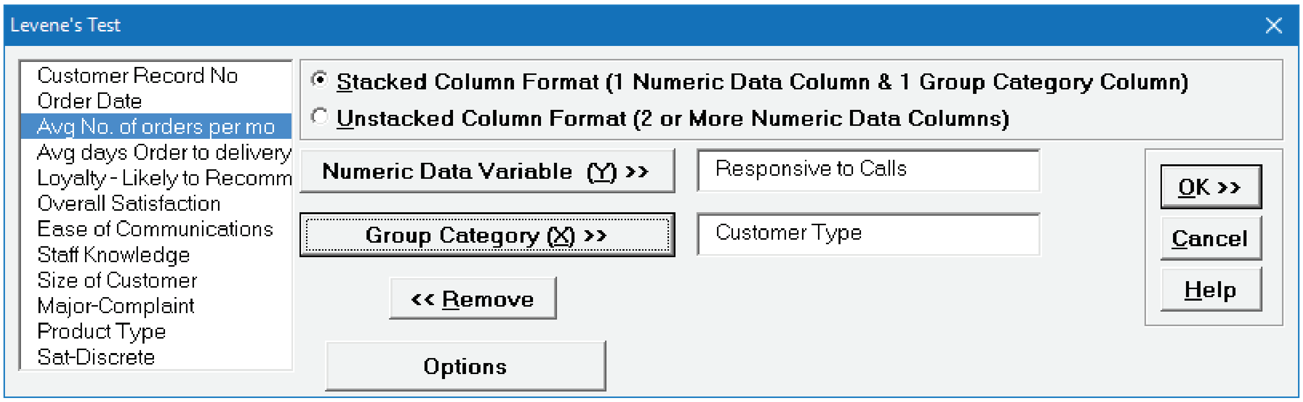
Notes on Multiple Comparison of Variances Probability Method and ANOM Chart:

- Levene Pairwise
    - Pairwise 2 sample Levene tests
    - Does not correct for family-wise error rate, so should only be used for k = 3 groups and in the restricted case where the Levene P-Value is < alpha.
- Tukey ADM (Absolute Deviations from the Median)
    - Application of Tukey on ADM
    - Recommended for k > 3
    - This post-hoc test is unique to SigmaXL, inspired by the method used in ANOM Levene Variances.
- See Appendix **Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests)** for further details and references.
- Display ANOM Levene Robust Variances Chart:
    - The ANOM alpha = (100 – Confidence Level)/100.
    - This chart is also available at **SigmaXL > Graphical Tools > Analysis of Means (ANOM)**. See **Part P – Analysis of Means (ANOM) Charts**.

1. Open **Customer Data.xlsx**, click on **Sheet 1** tab.

2. Click **SigmaXL > Statistical Tools > Equal Variance Tests > Levene**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Responsive to Calls*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**.



4. Click **OK**. The results are shown below:

Levene's Test For Equal Variance: Responsive to Calls			
(Use with non-normal data)			
Test Information			
H_0: Variance 1 = Variance 2 = ... = Variance k			
Ha: At least one pair Variance i ≠ Variance j			
Customer Type	**1**	**2**	**3**
Count	31	42	27
Mean	3.412	4.226	3.821
Median	3.500	4.720	4.180
StDev	1.304	0.921232	1.091543381
AD Normality Test P-Value	0.0021	0.0000	0.0190
Levene's Test Statistic	4.433		
DF Num	2		
DF Den	97		
P-Value	0.0144		

5. The Levene's Test P-Value of 0.0144 tells us that we reject H0. At least one pairwise set of variances are not equal. The normality test P-Values indicate that all 3 groups have nonnormal data (P-Values < .05). Since Levene's Test is robust to the assumption of normality, it is the correct test for equal variances (rather than Bartlett's Test).

6. If the Equal Variances test is only being used to test the assumption for use in ANOVA then it is not necessary to examine the Multiple Comparison of Variances. However, in the context of a

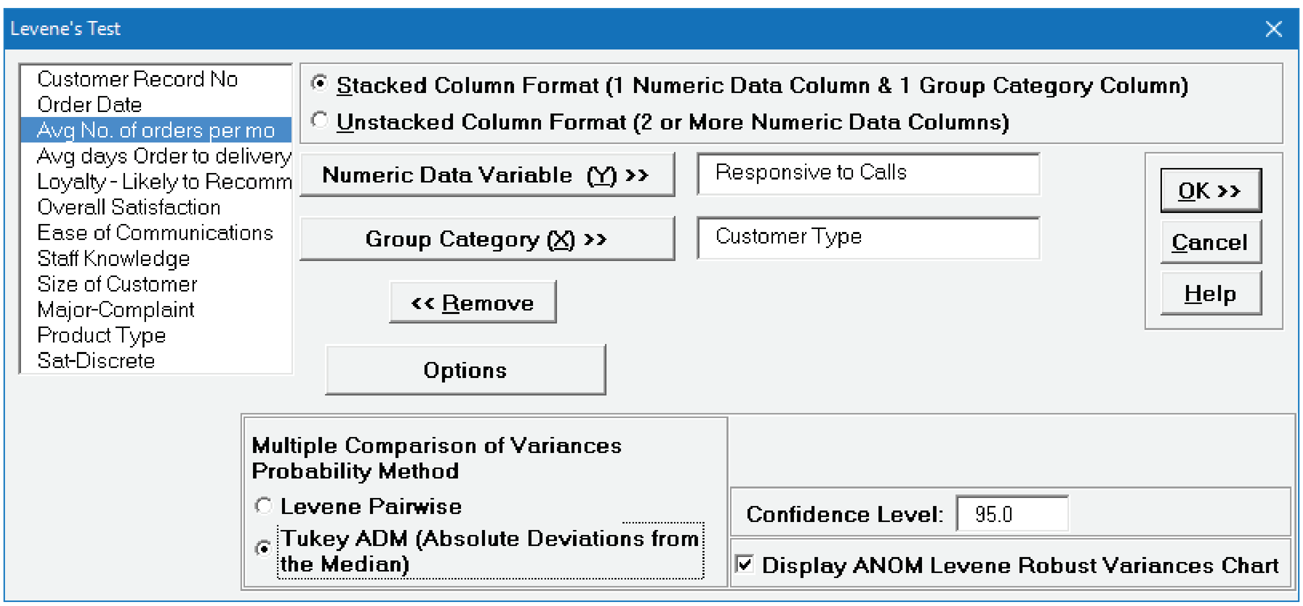
process improvement project, we often do want to know which groups are significantly different. This can give us important clues to identify opportunities for variance reduction.

7. The default Multiple Comparison of Variances is a matrix of Levene Pairwise Probabilities:

Multiple Comparison of Variances			
Levene Pairwise Probabilities	1	2	3
1		0.0036	0.0789
2			0.2991
3			

Customer Type 1 versus Customer Type 2 shows a significant difference in variance.

Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Click **Options**. Select **Tukey ADM (Absolute Deviations from Median)**. Check **Display ANOM Levene Robust Variances Chart**.



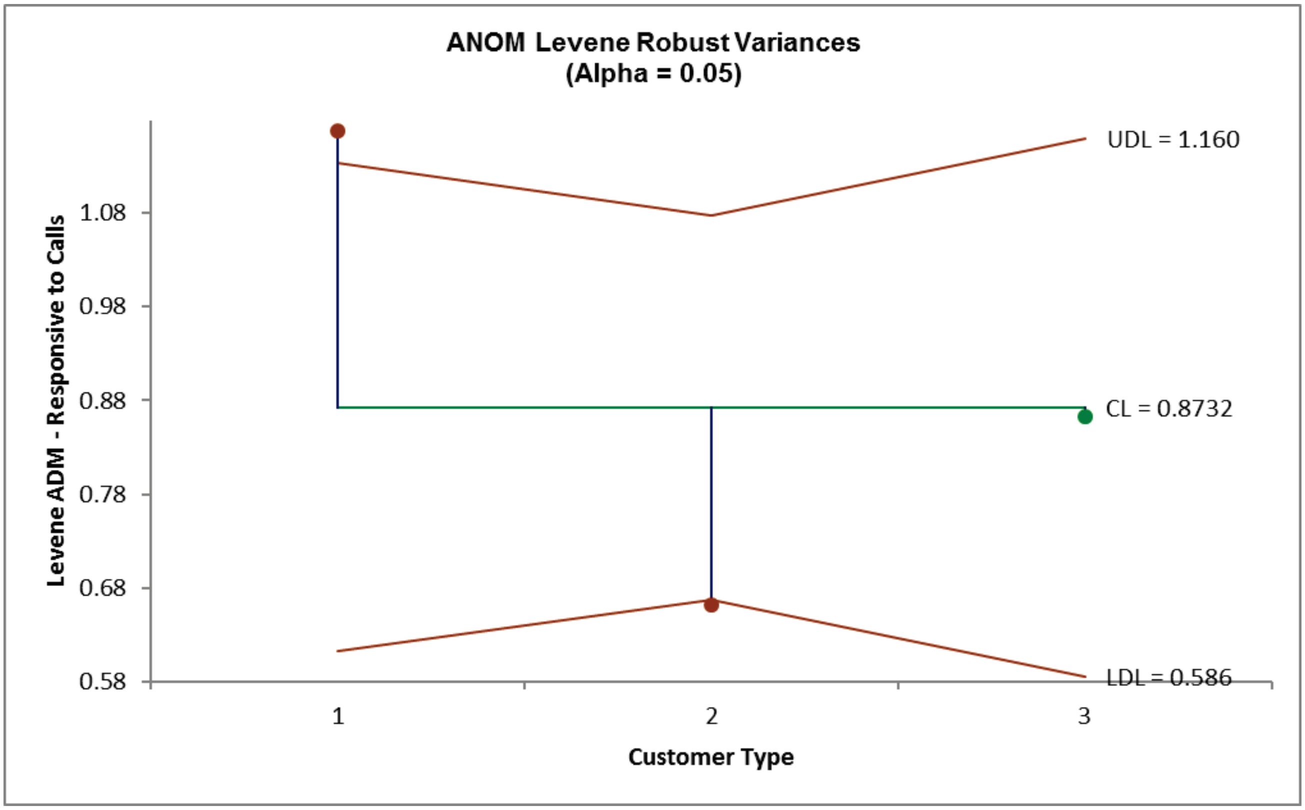
Note: The **Confidence Level** determines the alpha level (alpha = (100 – CI)/100) used to highlight the P-Values and the alpha level for the ANOM chart. However, the alpha level used to highlight P-Values in the Anderson-Darling Normality Test is always 0.05.

8. Click **OK**. The Multiple Comparison of Variances is a matrix of Tukey ADM Probabilities:

Multiple Comparison of Variances			
Tukey ADM Probabilities	1	2	3
1		0.0102	0.2466
2			0.4920
3			

The 1 – 2 P-Value is significant but larger than the Levene Pairwise because it adjusts for the family-wise error rate. Note that it is smaller than the Bonferroni corrected value = .0036 * 3 = .011, so more powerful than Bonferroni. The difference in power between Tukey and Bonferroni becomes more prominent with a larger number of groups, so Bonferroni is not included as an option.

9. Click the **ANOM_Levene** sheet tab to display the ANOM Chart:



10. The ANOM chart clearly shows Customer Type 1 has significantly higher variance (ADM) than overall and Customer Type 2 has significantly lower variance.

11. The varying decision limits are due to the varying sample sizes for each Customer Type, with smaller sample size giving wider limits in a manner similar to a control chart. If the data are balanced, the decision limit lines will be constant.

12. Now that we have determined that the variances are not equal, we are presented with a problem if we want to test for equal group means. Classical ANOVA assumes that the group variances are equal, so should not be used. A modified ANOVA called Welch’s ANOVA is robust to the assumption of equal variances and will be demonstrated next.

Welch's ANOVA Test (Assume Unequal Variance)

Welch's ANOVA is a test for multiple comparison of means. It is a modified One-Way ANOVA that is robust to the assumption of equal variances. Welch's ANOVA is an extension of the 2 sample t-test for means, assuming unequal variance. Nonparametric methods could also be used here but they are not as powerful as Welch's ANOVA.

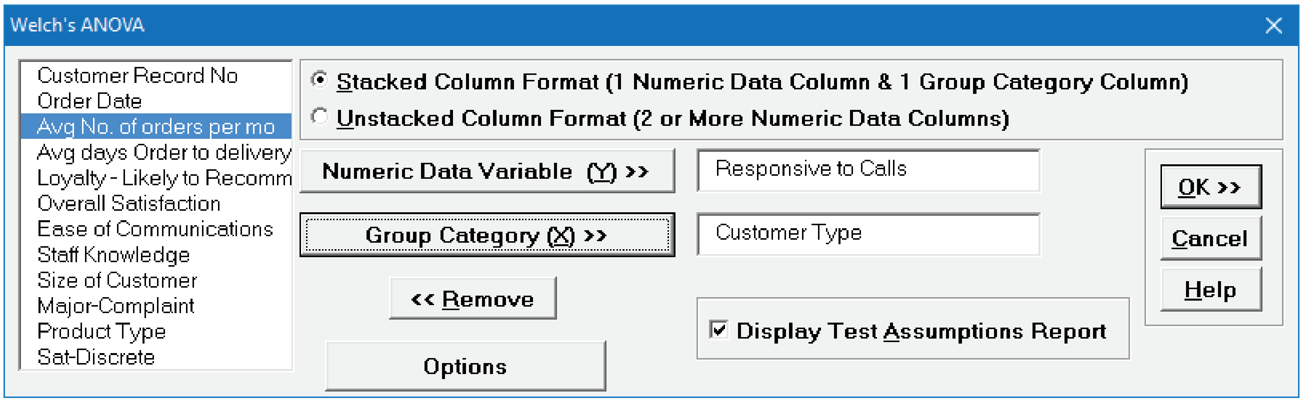
Notes on Means Matrix Probability Method (Multiple Comparison of Means):

- Welch Pairwise:
  - Pairwise 2 sample t-tests with unpooled standard deviation and weighted degrees of freedom (2 sample t-test for unequal variance)
  - Does not correct for family-wise error rate, so should only be used for k = 3 means and in the restricted case where the Welch ANOVA P-Value is < alpha.
- Games-Howell:
  - Similar to Welch Pairwise, uses unpooled standard deviation and weighted degrees of freedom, but is a studentized range statistic that corrects for family-wise error rate. Recommended for k > 3.
  - It is an extension of the Tukey test, but does not assume equal variance.
- See Appendix **Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests)** for further details and references.
- ANOM Chart for Welch ANOVA is not included in SigmaXL since it requires two-stage sampling (see Nelson, Wludyka, Copeland, 2005, Chapter 8, *Heteroscedastic Data*) **References for Analysis of Means (ANOM) Charts**.

1. Open **Customer Data.xlsx**, click on Sheet 1 tab (or press **F4** to activate last worksheet).

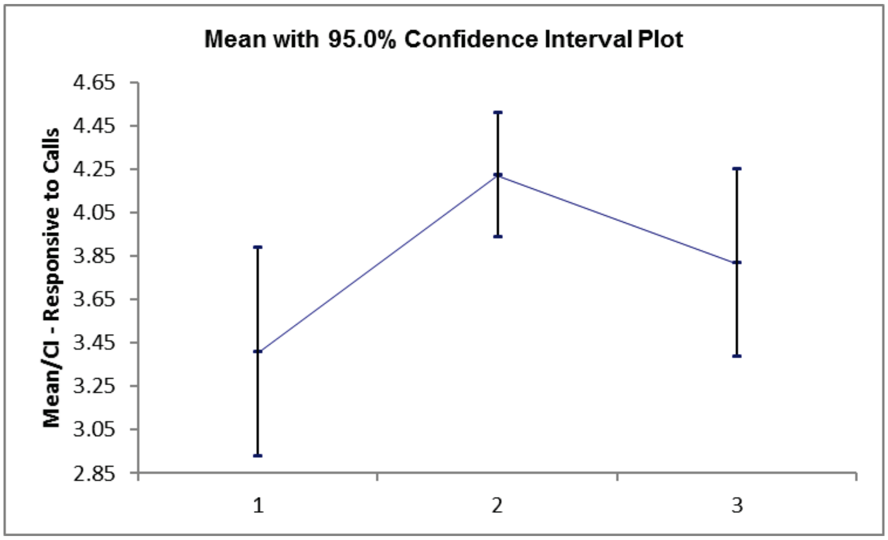
2. Click **SigmaXL > Statistical Tools > Equal Variance Tests > Welch's ANOVA (Assume Unequal Variance)** or **SigmaXL > Statistical Tools > Welch's ANOVA (Assume Unequal Variance)**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Responsive to Calls*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**. Check **Display Test Assumptions Report**.



4. Click **OK**. The results are shown below:

Welch's ANOVA: Responsive to Calls			
(Use if groups have unequal variance)			
Test Information			
H_0: Mean 1 = Mean 2 = ... = Mean k			
Ha: At least one pair Mean i ≠ Mean j			
Customer Type	**1**	**2**	**3**
Count	31	42	27
Mean	3.412	4.226	3.821
Median	3.500	4.720	4.180
StDev	1.304	0.921232	1.091543381
AD Normality Test P-Value	0.0021	0.0000	0.0190
Welch's F Ratio	4.654		
DF Num	2		
DF Den	55.753		
P-Value	0.0135		
Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.814578	-0.409869
2		0	0.404709
3			0
Welch Pairwise Probabilities	1	2	3
1		0.0045	0.1981
2			0.1170
3			



5. The P-Value for Welch's ANOVA is 0.0135, therefore we reject H0 and conclude that the group means for Responsive to Calls are not equal.

6. From the Pairwise Mean Difference (Means Matrix), we conclude that Mean Responsive to Calls is significantly different between Customer Type 1 and 2. Note that the default probabilities are Welch Pairwise. See below for more details on the multiple comparison options.

8. A graphical view of the Responsive to Calls Mean and 95% Confidence Intervals are given to complement the Means Matrix. Note that the standard deviations are unpooled, resulting in different CI widths for each group. The fact that the CI's for Customer Type 1 do not overlap those of Type 2, visually shows that there is a significant difference in mean Responsive to Calls. The overlap of CI's for Type 2 and 3 shows that the mean scores for 2 and 3 are not significantly different.

9. Later, we will explore the relationship between Responsive to Calls and Overall Satisfaction.

10. Welch's ANOVA Assumptions Report:

Welch's ANOVA Assumptions Report			
Normality:	Anderson Darling P-Value = 0.002. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.3047 and Kurtosis value = -1.2945. See robustness and outliers.	Anderson Darling P-Value = 0.000. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -1.2693 and Kurtosis value = 0.5291. See robustness and outliers.	Anderson Darling P-Value = 0.019. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.9436 and Kurtosis value = 0.1684. See robustness and outliers.
Robustness:	Minimum sample size for a robust ANOVA test = 4. Since each sample size is greater than this, the ANOVA test is robust to nonnormality.		
Outliers (Boxplot Rules):	No outliers found.	No outliers found.	No outliers found.
Randomness (Independence):	Nonparametric Runs Test (Exact) P-Value = 0.465. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.	Nonparametric Runs Test (Exact) P Value = 0.158. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.	Nonparametric Runs Test (Exact) P Value = 0.568. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (serial independence) is not violated.
Equal Variance:	Levene's Test for Equal Variances P-Value = 0.014. Reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is violated (at 95% confidence level). Welch's ANOVA is appropriate.		

This is a text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation).

Each sample is tested for Normality using the Anderson-Darling test. If not normal, the minimum sample size for robustness of the ANOVA Test is determined utilizing Monte Carlo regression equations (see **Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA**). If the sample size is inadequate, a warning is given and the appropriate Nonparametric test is recommended (Kruskal-Wallis if there are no extreme outliers, Mood's Median if there are extreme outliers).

Each sample is tested for Outliers defined as: Potential: Tukey's Boxplot (> Q3 + 1.5*IQR or < Q1 – 1.5*IQR); Likely: Tukey's Boxplot 2.2*IQR; Extreme: Tukey's Boxplot 3*IQR. If outliers are present, a warning is given and recommendation to review the data with a Boxplot and Normal Probability Plot

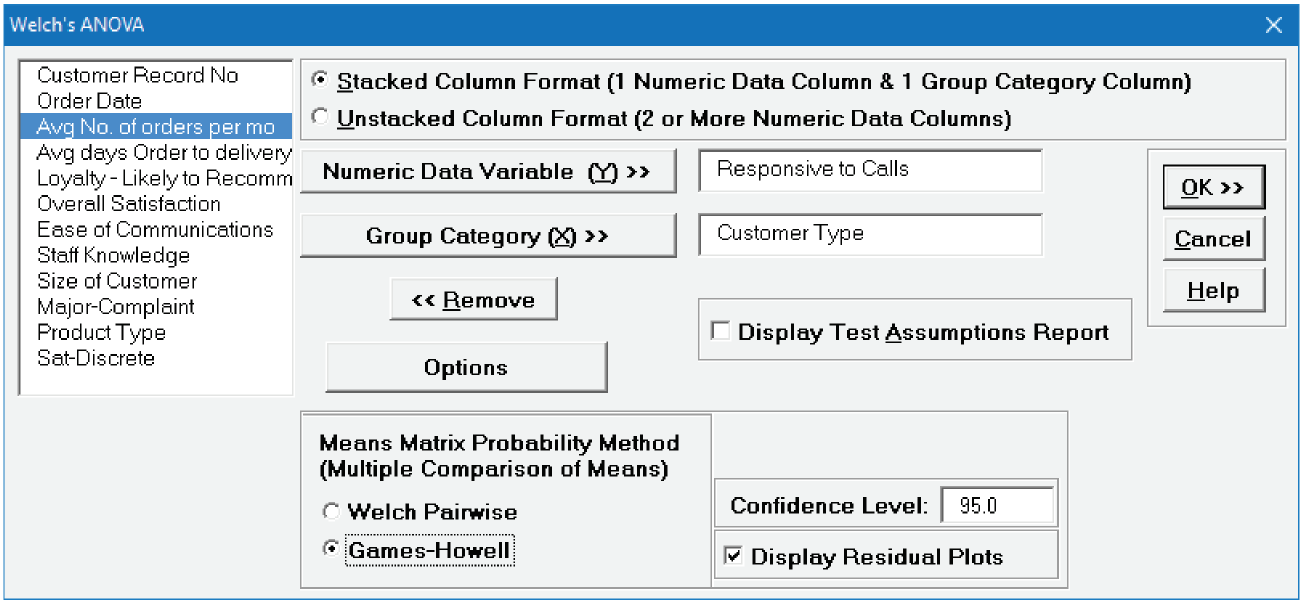
Tip: If the removal of outlier(s) result in an Anderson-Darling P-Value that is > 0.1, a notice is given that excluding the outlier(s), the sample data are inherently normal.

Each sample is tested for Randomness using the Exact Nonparametric Runs Test. If the sample data is not random, a warning is given and recommendation to review the data with a Run Chart.

A test for Equal Variances is also applied. If all sample data are normal, Bartlett's Test is utilized, otherwise Levene's Test is used. Since we are using Welch's ANOVA, it is confirmed as appropriate.

See Appendix **Hypothesis Test Assumptions Report** for further details.

11. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Uncheck **Display Test Assumptions Report**. Click the **Options** button. Select **Games-Howell**. Check **Display Residual Charts** as shown:



Note: The **Confidence Level** is used to set the level in the Mean/Confidence Interval Plot and the alpha level (alpha = (100 – CI)/100) used to highlight the P-Values. However, the confidence level used in the Residuals Normal Probability Plot is always 95%.

12. Click **OK**. The Pairwise Means Difference (Means Matrix) and Games-Howell Probability results are:

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.814578	-0.409869
2		0	0.404709
3			0

Games-Howell Probabilities	1	2	3
1		**0.0123**	0.3998
2			0.2572
3			

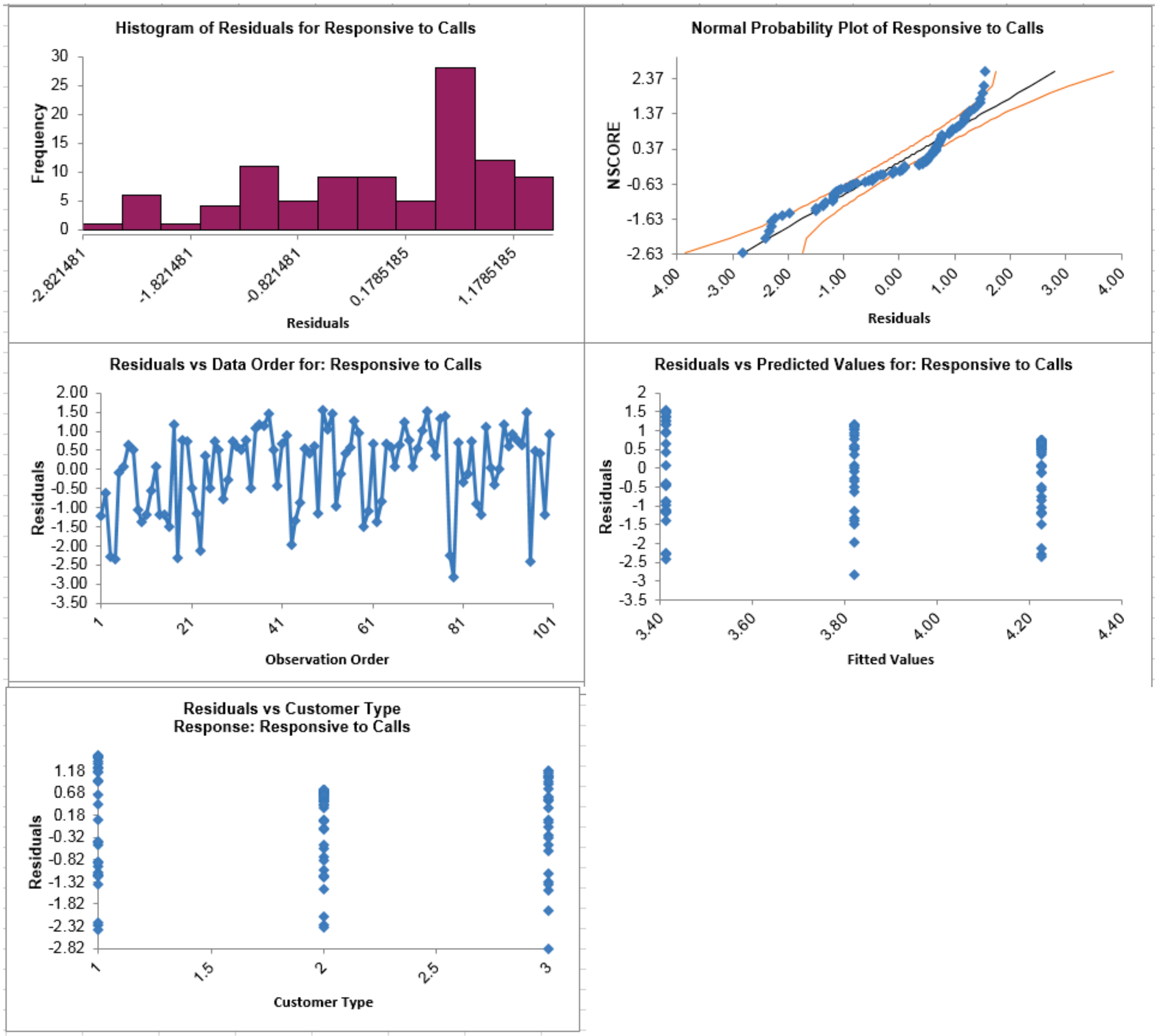
13. The significant Games-Howell Probability values above have not changed as compared to Welch Pairwise, but note that they are larger to compensate for the family-wise error rate:

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.814578	-0.409869
2		0	0.404709
3			0

Welch Pairwise Probabilities	1	2	3
1		**0.0045**	0.1981
2			0.1170
3			

Note also that the 1 – 2 Games-Howell probability is smaller than the Bonferroni corrected value = .0045 * 3 = .0135, so more powerful than Bonferroni. The difference in power between Games-Howell and Bonferroni becomes more prominent with a larger number of groups, so Bonferroni is not included as an option.

14. Click the **Welch Residuals** sheet tab to display the Residual Plots:



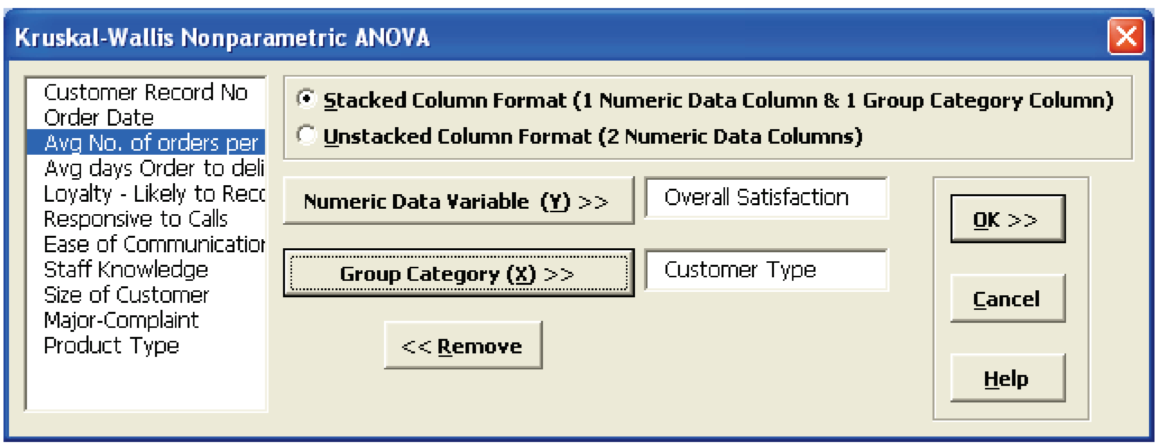
Residuals are the unexplained variation from the ANOVA model (Actual – Predicted or Fitted values). Note that the residuals are not normally distributed - as expected from the assumptions report - but like the regular ANOVA, Welch’s ANOVA is quite robust to the assumption of normality. Also, as expected from Levene’s test for equal variances, the variability for Type 2 is less than Type 1, but Welch’s ANOVA is robust to the assumption of equal variances, so we can trust that the P-Values are valid.

Part M – Nonparametric Multiple Comparison

Kruskal-Wallis

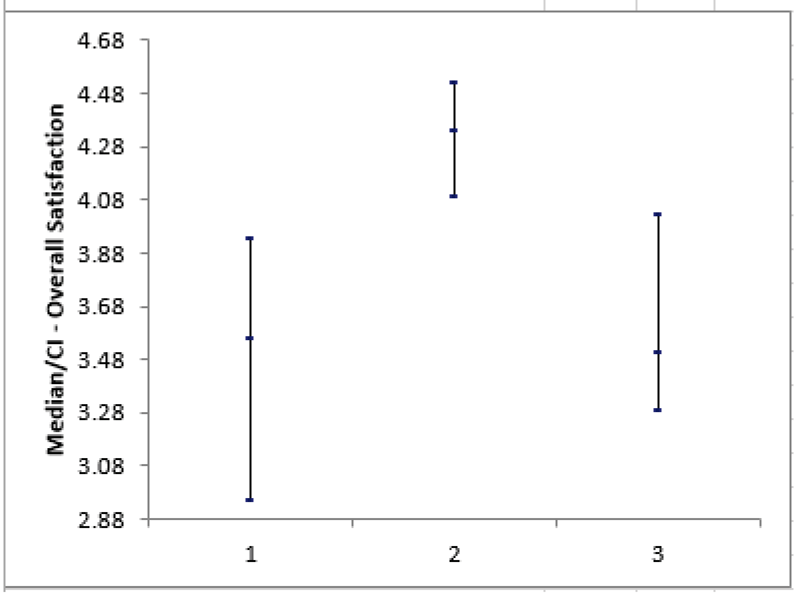
The Kruskal-Wallis test is an extension of the Mann-Whitney Rank test, allowing for more than 2 samples. It is a nonparametric equivalent to the parametric One-Way ANOVA (i.e., One-Way ANOVA on Ranks). The Null Hypothesis is: H0: Median1 = Median2 = … = MedianK. Ha: At least two Medians are different. The test does not assume sample normality but does assume that the samples have equal shapes. If the shapes are different, the null hypothesis is that the distributions are the same.

1. Open **Customer Data.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests > Kruskal-Wallis**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.
3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**.



4. Click **OK**. The results are shown below:

Kruskal-Wallis Nonparametric ANOVA: Overall Satisfacti			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Customer Type	**1**	**2**	**3**
Count (N)	31	42	27
Median	3.56	4.34	3.51
UC Median (2-sided, 95%)	3.936	4.518	4.023
LC Median (2-sided, 95%)	2.954	4.095	3.289
Z	-3.339	4.529	-1.557
Kruskal-Wallis Statistic (H)	21.360		
DF	2		
P-Value (2-sided, adjusted for ties)	0.0000		



Click on cell B16 to view the P-Value with more decimal place precision (or change the cell format to scientific notation). The P-Value of 0.000023 (2.3 e-5) tells us that we reject H0. At least one pairwise set of medians are not equal.

5. The Kruskal-Wallis Statistic is based on comparing mean ranks for each group versus the mean rank for all observations. The Z value for Customer Type 3 is -1.56, the smallest absolute Z-value. This size indicates that the mean rank for Type 3 differed least from the mean rank for all observations. The Z value for Customer Type 2 is 4.53, the largest absolute Z-value. This size indicates that the mean rank differed most from the mean rank for all observations.

6. A graphical view of the Overall Satisfaction Median and 95% Confidence Intervals are given to complement the Z scores. The fact that the CI's for Customer Type 2 do not overlap those of Type 1 or 3, clearly shows that Customer Type 2 has a significantly higher median satisfaction score. The overlap of CI's for Type 1 and 3 shows that the median satisfaction scores for 1 and 3 are not significantly different.

Kruskal-Wallis – Exact

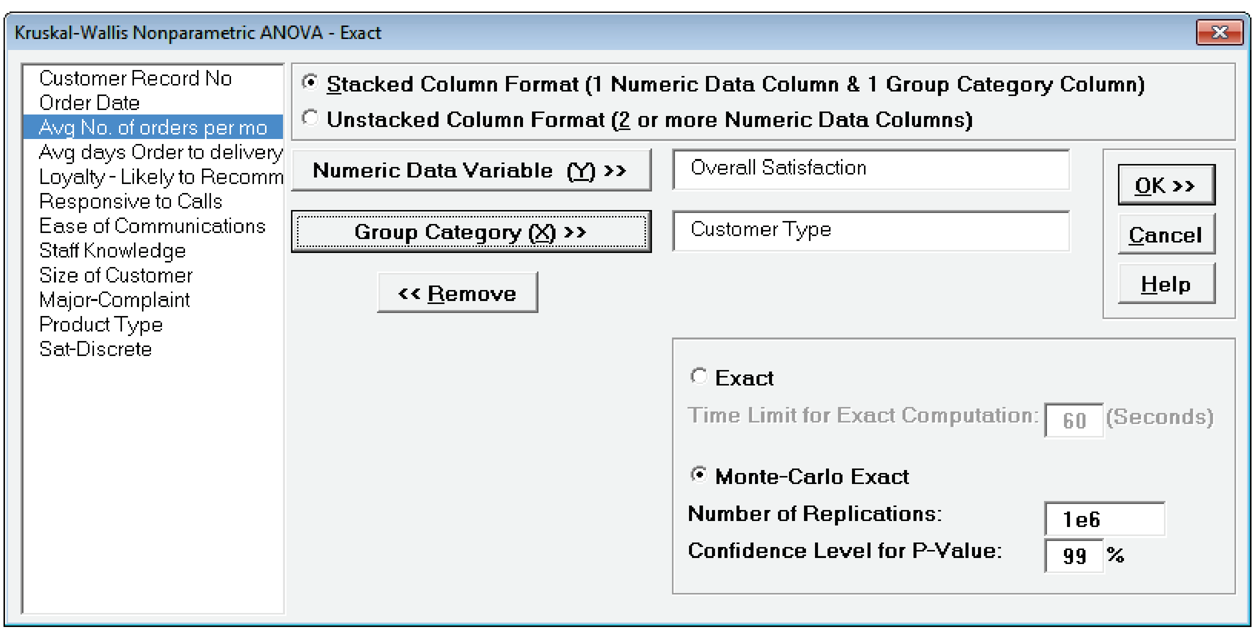
We will now redo the above example to estimate the exact P-Value using Monte Carlo. Typically, this would not be necessary unless the sample sizes were smaller (each sample N <= 5 for Kruskal-Wallis), but this gives us continuity on the example. We will consider a small sample problem later.

Computing an exact P-Value for Kruskal-Wallis is very computationally intensive. The Network Model by Mehta and Patel cannot be used for this test (see Appendix Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests). In this example, the total number of permutations are:

(31+42+27)! / (31! * 42! * 27!) = 7.42 E44

(i.e., more than the number of stars in the observable universe). So, we will not attempt to compute the exact, but rather use Monte Carlo!

1. Open **Customer Data.xlsx**, click **Sheet 1** tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Kuskal-Wallis – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.
3. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**. Select **Monte Carlo Exact** with the **Number of Replications** = 1e6 and **Confidence Level for P-Value** = 99%. One million replications are used because the expected P-Value is very small as estimated from the "large sample" Kruskal-Wallis above. This will take up to a minute to run, so if you have a slow computer, use 1e5 replications instead of 1e6.



Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

4. Click **OK**.

Kruskal-Wallis Nonparametric ANOVA - Monte Carlo: Overall Satisfaction			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Customer Type	**1**	**2**	**3**
Count (N)	31	42	27
Median	3.56	4.34	3.51
UC Median (2-sided, 95% approx.)	3.936	4.518	4.023
LC Median (2-sided, 95% approx.)	2.954	4.095	3.289
Z	-3.339	4.529	-1.557
Kruskal-Wallis Statistic (H)	21.360		
DF	2		
Monte Carlo P-Value (2-sided)	0.0000		
Monte Carlo P-Value 99% CI Upper	0.0000		
Monte Carlo P-Value 99% CI Lower	0.0000		

Click on cell B16 to view the P-Value with more decimal place precision (or change the cell format to scientific notation). The Monte Carlo P-Value here is 0.000009 (9 e-6) with a 99% confidence interval of .000002 (2 e-6) to 0.000016 (1.6 e-5). This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). So, we reject H0: at least one pairwise set of medians are not equal.

Note that the large sample (asymptotic) P-Value of 2.3 e-5 lies outside of the Monte Carlo exact confidence interval.

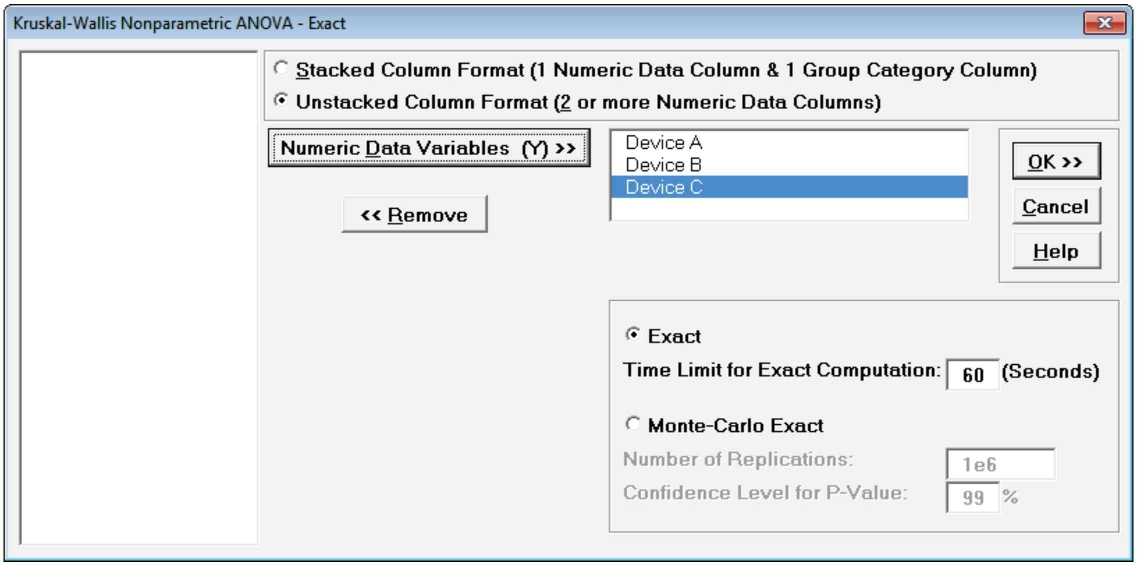
5. Now we will consider a small sample problem. Open **Snore Study.xlsx**. This data is from:

   Gibbons, J.D. and Chakraborti, S. (2010). *Nonparametric Statistical Inference* (5th Edition). New York: Chapman & Hall, (Example 10.2.1 data, page 347; Example 10.4.2 analysis, pp. 360 – 362).

   An experiment was conducted to determine which device is the most effective in stopping snoring or at least in reducing it. Fifteen men who are habitual snorers were divided randomly into three groups to test the devices. Each man's sleep was monitored for one night by a machine that measures the amount of snoring on a 100-point scale while using a device.

6. Select **Snore Study Data** tab. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Kruskal-Wallis - Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

7. With **Unstacked Column Format** checked, select *Device A, Device B and Device C*, click **Numeric Data Variables (Y) >>**. Select **Exact** with the default **Time Limit for Exact Computation** = 60 seconds.



8. Click **OK**. Results:

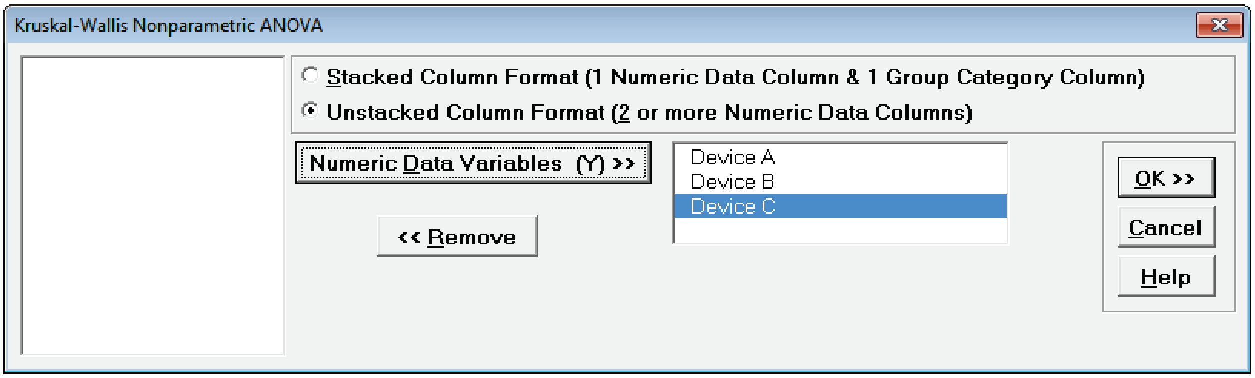
Kruskal-Wallis Nonparametric ANOVA - Exact			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Results:	**Device A**	**Device B**	**Device C**
Count (N)	5	5	5
Median	79	92	26
UC Median (2-sided, 95% approx.)	91	96	78
LC Median (2-sided, 95% approx.)	35	76	8
Z	0.244949	2.449	-2.694
Kruskal-Wallis Statistic (H)	8.880		
DF	2		
Exact P-Value (2-sided)	0.0042		

With the Exact P-Value = 0.0042 we reject H0, and conclude that there is a significant difference in median snore study scores. This exact P-Value matches that given in the reference textbook using SAS and StatXact.

By way of comparison, we will now rerun the analysis using the "large sample" or "asymptotic" Kruskal-Wallis test.

9. Select **Snore Study Data** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Nonparametric Tests > Kruskal-Wallis**. If necessary, check **Use Entire Data Table**, click **Next**.

10. With **Unstacked Column Format** checked, select *Device A, Device B and Device C*, click **Numeric Data Variables (Y) >>**.



11. Click **OK**. Results:

Kruskal-Wallis Nonparametric ANOVA			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Results:	**Device A**	**Device B**	**Device C**
Count (N)	5	5	5
Median	79	92	26
UC Median (2-sided, 95%)	91	96	78
LC Median (2-sided, 95%)	35	76	8
Z	0.244949	2.449	-2.694
Kruskal-Wallis Statistic (H)	8.880		
DF	2		
P-Value (2-sided, adjusted for ties)	0.0118		

With the P-Value = .0118 we reject H0 (using alpha = .05), but note that if we were using alpha = 0.01, we would have incorrectly failed to reject the null hypothesis. This "large sample" P-Value matches that given in the reference textbook using Minitab.

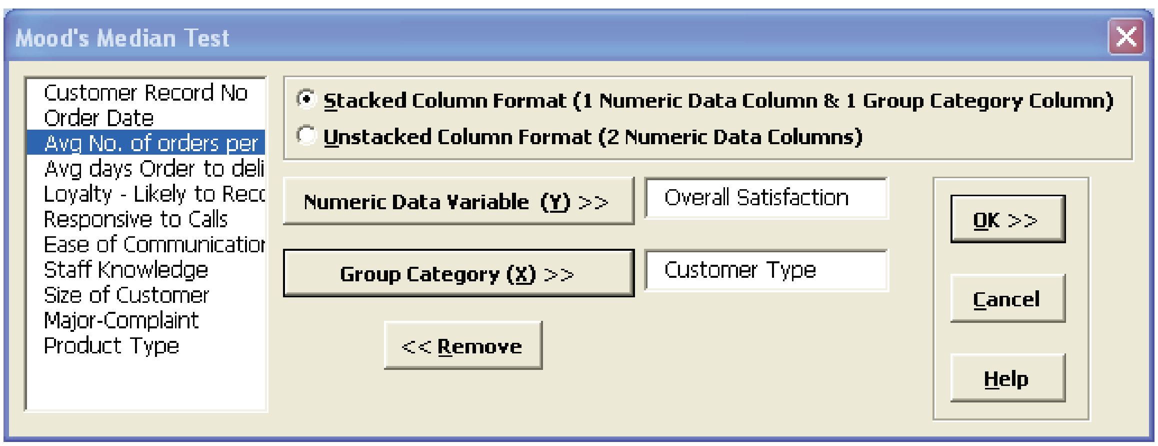
In conclusion, whenever you have a small sample size and are performing a nonparametric test, always use the Exact option.

Mood's Median Test

Mood's Median Test is an extension of the One Sample Sign Test, using Chi-Square as the test statistic. Like the Kruskal-Wallis test, Mood's median test can be used to test the equality of medians from multiple samples. It provides a nonparametric alternative to the one-way analysis of variance. The Null Hypothesis is: H0: Median1 = Median2 = … = MedianK. Ha: At least two Medians are different.

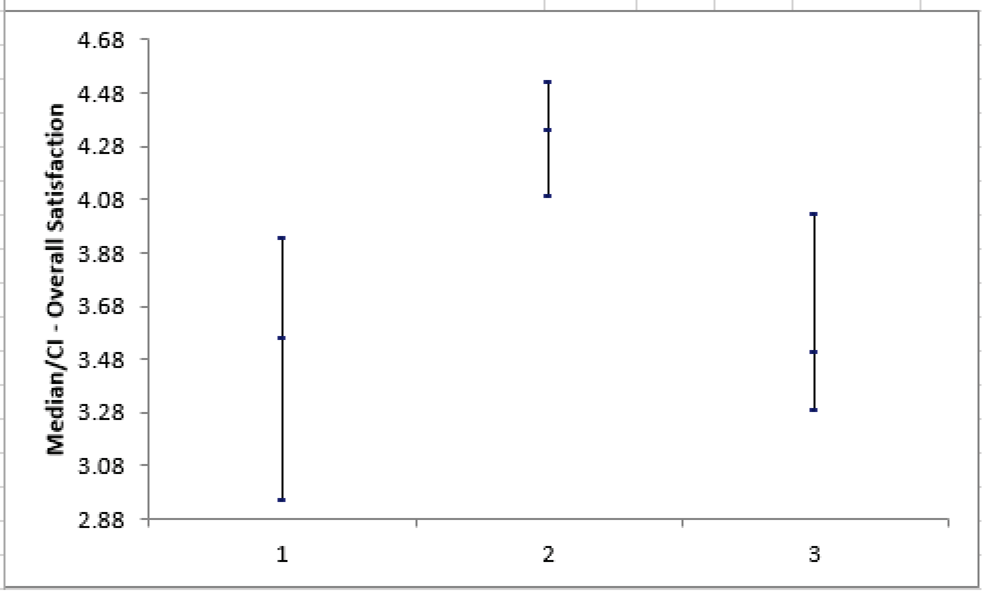
Mood's median test is more robust to outliers than the Kruskal-Wallis test, but is less powerful in the absence of outliers. You should first look at your data with Boxplots. If there are extreme outliers, then Mood's Median should be used rather than Kruskal-Wallis.

1. Open **Customer Data.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests > Mood's Median Test**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.
3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**.



4. Click **OK**. The results are shown below:

Mood's Median Test: Overall Satisfaction			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Customer Type	**1**	**2**	**3**
Count (N <= Overall Median)	21	12	17
Count (N > Overall Median)	10	30	10
Median	3.560	4.340	3.510
UC Median (2-sided, 95%)	3.936	4.518	4.023
LC Median (2-sided, 95%)	2.954	4.095	3.289
Overall Median	3.945		
Chi-Square	13.432		
DF	2		
P-Value (2-sided)	0.0012		



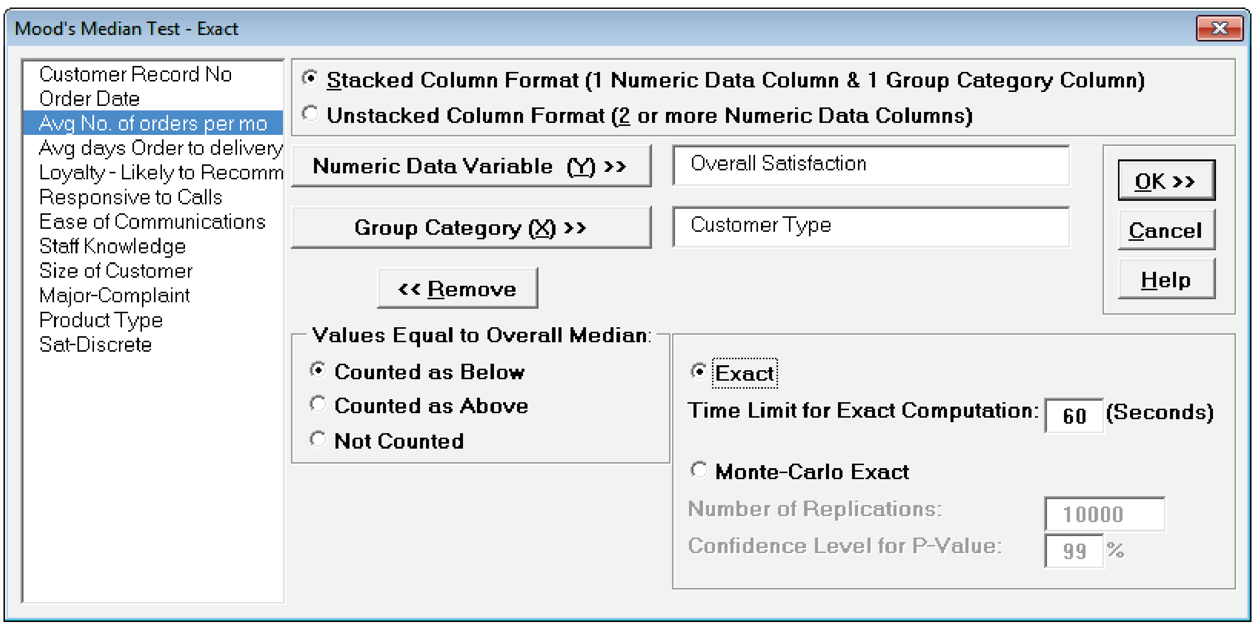
The P-Value of 0.0012 tells us that we reject H0. At least one pairwise set of medians are not equal.

5. A graphical view of the Overall Satisfaction Median and 95% Confidence Intervals are given. This is the same graph provided in the Kruskal-Wallis test report.

Mood's Median Test - Exact

We will now redo the above example to compute exact P-Values. Typically, this would not be necessary unless the sample sizes were smaller (each sample N <= 10 for Mood's Median), but this gives us continuity on the example. We will consider a small sample problem later. Unlike Kruskal-Wallis, computing an exact P-Value is efficient due to the reduction of the data to a Chi-Square table and the use of a Network Model (see Appendix Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests)

1. Open **Customer Data.xlsx**, click **Sheet 1** tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Mood's Median Test – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.
3. With **Stacked Column Format** checked, select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**. Select **Exact** with the default **Time Limit for Exact Computation** = 60 seconds. Set **Values Equal to Overall Median:** to **Counted as Below**.



Tip: If none of the observations are equal to the overall median, the options for "**Values Equal to Overall Median:**" do not affect the exact P-Value. That is the case with this Customer Data example. In cases where there are observations equal to the overall median, then the exact P-Value will be different for each option selected. For regular Mood's Median, SigmaXL uses **Counted as Below** (in agreement with Minitab) but will use **Counted as Above** for small sample cases if the average cell expected value is higher. StatXact uses **Counted as Below** for Mood's Median. Matlab uses **Not Counted**. Users should try **Counted as Below** and **Counted as Above** to ensure that the P-Values agree on reject or fail to reject H0. Use of this option will be demonstrated in the small sample example.

Tip: If the exact computation time limit is exceeded a dialog will prompt you to use Monte Carlo or to increase the computation time. When this occurs, Monte Carlo is recommended.

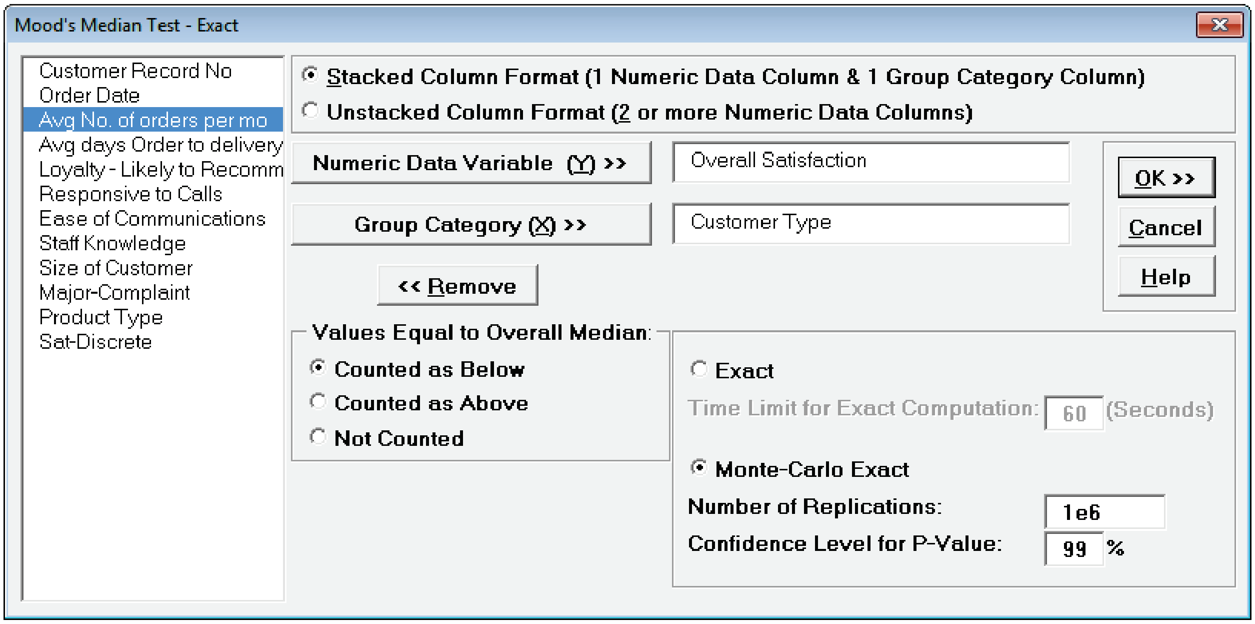
4. Click **OK**. Resulting output:

Mood's Median Test - Exact: Overall Satisfaction			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Customer Type	**1**	**2**	**3**
Count (N <= Overall Median)	21	12	17
Count (N > Overall Median)	10	30	10
Median	3.560	4.340	3.510
UC Median (2-sided, 95% approx.)	3.936	4.518	4.023
LC Median (2-sided, 95% approx.)	2.954	4.095	3.289
Overall Median	3.945		
Chi-Square	13.432		
DF	2		
Exact P-Value (2-sided)	0.0012		

Click on cell **B17** to view the Exact P-Value with more decimal place precision. Given the P-Value of .00124 we reject H0 and conclude that at least one pairwise set of medians are not equal. The Exact P-Value is very close to the above "large sample" or "asymptotic" result (.00121). This was expected because the sample size is reasonable (N > 10), so the "large sample" Mood's Median P-Values are valid using a chi-square approximation. Note also that the counts in **B8:D9** form a two-way contingency table and the expected counts are all greater than 5.

The Exact P-Value was computed in seconds, but if the data set was larger, the required computation time could become excessive, and Monte Carlo would be required.

5. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Select **Monte Carlo Exact** with the **Number of Replications** = 1e6 and **Confidence Level for P-Value** = 99%. One million replications are used because the expected P-Value is small as estimated from the "large sample" Mood's Median above. This will take up to a minute to run, so if you have a slow computer, use 1e5 replications instead of 1e6.



Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

6. Click **OK**.

Mood's Median - Monte Carlo: Overall Satisfaction			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Customer Type	**1**	**2**	**3**
Count (N <= Overall Median)	21	12	17
Count (N > Overall Median)	10	30	10
Median	3.560	4.340	3.510
UC Median (2-sided, 95% approx.)	3.936	4.518	4.023
LC Median (2-sided, 95% approx.)	2.954	4.095	3.289
Overall Median	3.945		
Chi-Square	13.432		
DF	2		
Monte Carlo P-Value (2-sided)	0.0012		
Monte Carlo P-Value 99% CI Upper	0.0013		
Monte Carlo P-Value 99% CI Lower	0.0012		

Click on cell **B17** to view the Monte Carlo P-Value with more decimal place precision. The Monte Carlo P-Value here is .001249 with a 99% confidence interval of .00117 (**B19**) to .00133 (**B18**). This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). The true Exact P-Value = .00124 lies within this confidence interval.

So, we reject H0: at least one pairwise set of medians are not equal.

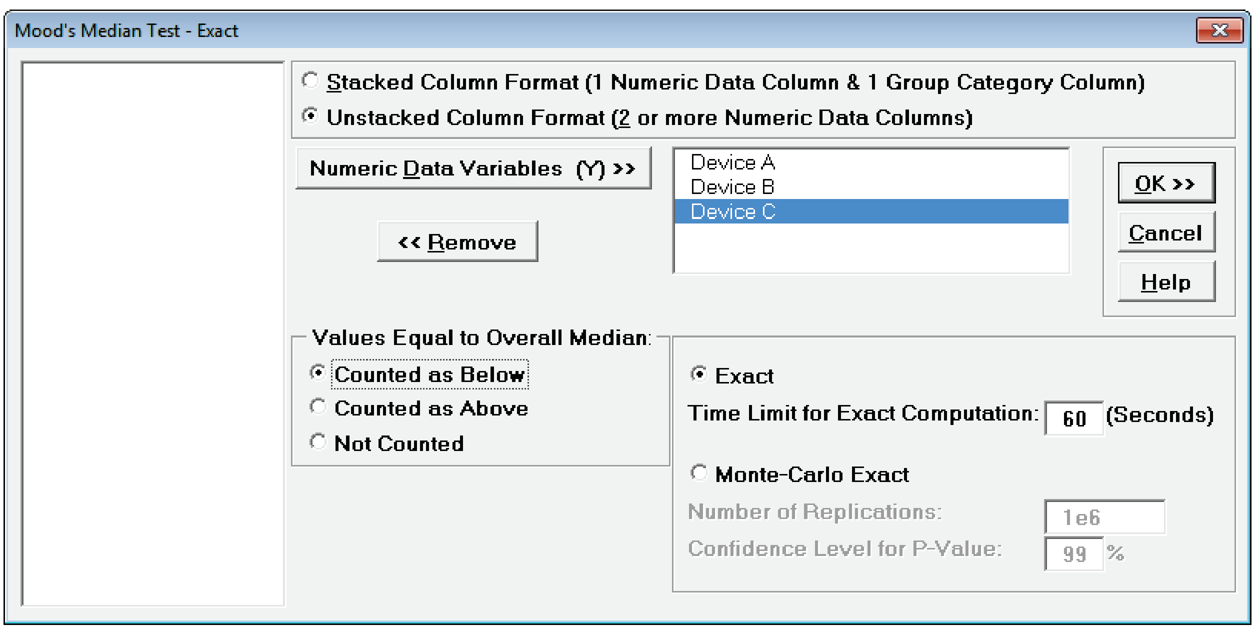
7. Now we will consider the small sample problem used in Kruskal-Wallis. Open **Snore Study.xlsx**. This data is from:

   Gibbons, J.D. and Chakraborti, S. (2010). *Nonparametric Statistical Inference* (5th Edition). New York: Chapman & Hall, (Example 10.2.1, pp. 347-348).

   An experiment was conducted to determine which device is the most effective in stopping snoring or at least in reducing it. Fifteen men who are habitual snorers were divided randomly into three groups to test the devices. Each man's sleep was monitored for one night by a machine that measures the amount of snoring on a 100-point scale while using a device.

8. Select **Snore Study Data** tab. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Mood's Median Test - Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

9. With **Unstacked Column Format** checked, select *Device A, Device B and Device C*, click **Numeric Data Variables (Y) >>**. Select **Exact** with the default **Time Limit for Exact Computation** = 60 seconds.



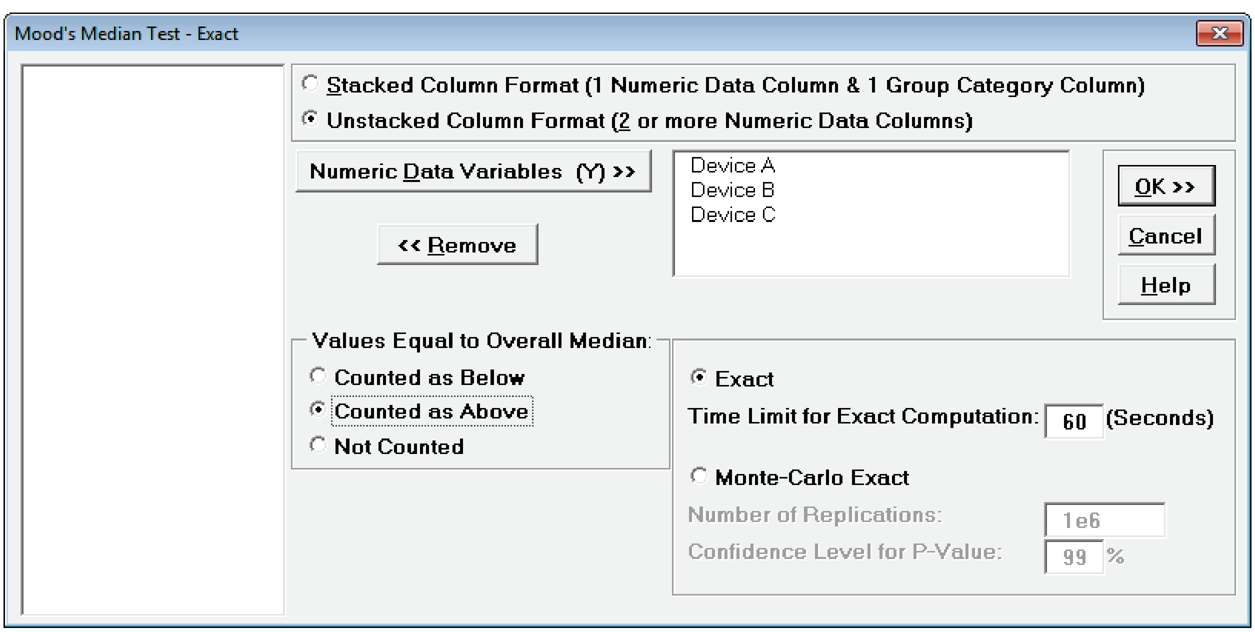
10. Click **OK**. Results:

Mood's Median Test - Exact			
Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Results:	**Device A**	**Device B**	**Device C**
Count (N <= Overall Median)	2	1	5
Count (N > Overall Median)	3	4	0
Median	79	92	26
UC Median (2-sided, 95% approx.)	91	96	78
LC Median (2-sided, 95% approx.)	35	76	8
Overall Median	78		
Chi-Square	6.964		
DF	2		
Exact P-Value (2-sided)	0.0676		

With the Exact P-Value = .0676 we fail to reject H0, and cannot conclude that there is a significant difference in median snore study scores. This exact P-Value matches that given in the reference textbook using StatXact. Comparing Mood's Median to Kruskal-Wallis we see that the Exact P-Value is higher because Mood's Median is not as powerful as Kruskal-Wallis.

We will now rerun the analysis using the **Counted as Above** option.

11. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Set **Values Equal to Overall Median:** to **Counted as Above**.



12. Click **OK**. Results:

Mood's Median Test - Exact

Test Information			
H_0: Median 1 = Median 2 = ... = Median k			
H_a: At least one pair Median i ≠ Median j			
Results:	**Device A**	**Device B**	**Device C**
Count (N < Overall Median)	2	1	4
Count (N >= Overall Median)	3	4	1
Median	79	92	26
UC Median (2-sided, 95% approx.)	91	96	78
LC Median (2-sided, 95% approx.)	35	76	8
Overall Median	78		
Chi-Square	3.750		
DF	2		
Exact P-Value (2-sided)	0.3007		

With the Exact P-Value = 0.3007 we fail to reject H0, and cannot conclude that there is a significant difference in median snore study scores. Note that setting **Values Equal to Overall Median** to **Counted as Above** resulted in a different count for Device C, which results in the dramatic difference in Exact P-Value.

In conclusion, when using Mood's Median Exact, always try **Counted as Below** and **Counted as Above** to ensure that the P-Values agree with each other.

Friedman Test

The Friedman test is an extension of the Sign test and the nonparametric equivalent of a One-Way Repeated-Measures ANOVA (i.e., One-Way ANOVA on Within Subject Ranks). The concept is the same as that of the Paired t-test, we are interested in the within subject variability, not the subject-to-subject variability.

It can also be considered as a nonparametric alternative to a DOE model with blocks or GLM with one fixed factor and one random factor. The Friedman test requires exactly 1 non-missing observation for each level of the treatment, within each block.

The null and alternative hypothesis are:

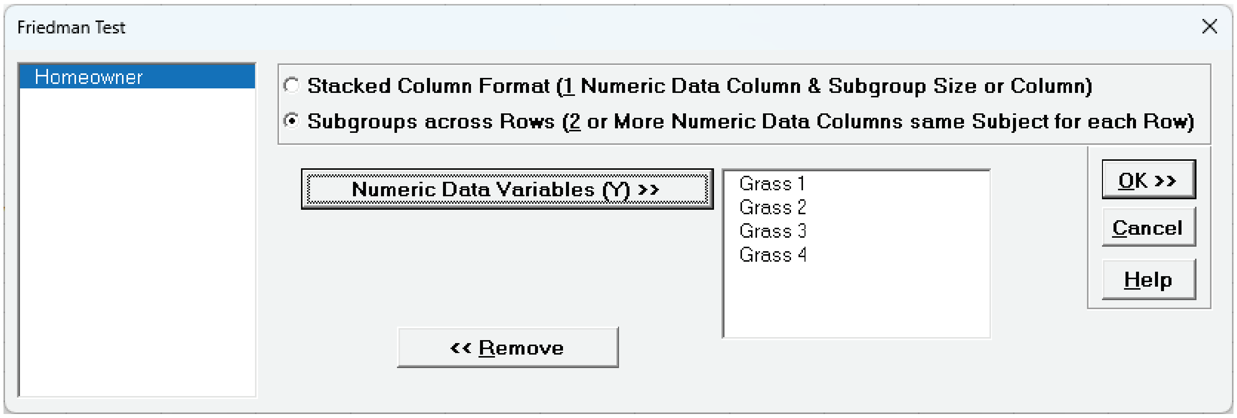
H_0: Median 1 = Median 2 = ... = Median k (apart from Subject/Block Effect)
H_a: At least one pair Median i ≠ Median j (apart from Subject/Block Effect)

1. Open **Conover Grass Type Experiment.xlsx**, click on **Sheet 1** tab. This data is from: Conover, W.J. (1999), *Practical Nonparametric Statistics* (3rd Edition). New York: John Wiley & Sons, pp. 371-373, Example 1. Twelve homeowners are randomly selected to participate in an experiment with a plant nursery. Each homeowner was asked to select four fairly identical areas in their yard and to plant four different types of grass, one in each area. At the end of a specified length of time each homeowner was asked to rank the grass types in order of preference. The data are given in tabular format (unstacked):

Homeowner	Grass 1	Grass 2	Grass 3	Grass 4
1	4	3	2	1
2	4	2	3	1
3	3	1.5	1.5	4
4	3	1	2	4
5	4	2	1	3
6	2	2	2	4
7	1	3	2	4
8	2	4	1	3
9	3.5	1	2	3.5
10	4	1	3	2
11	4	2	3	1
12	3.5	1	2	3.5

2. Click **SigmaXL > Statistical Tools > Nonparametric Tests > Friedman Test**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Subgroups across Rows (2 or More Numeric Data Columns same Subject for each Row)** is selected. Select *Grass 1* to *Grass 4,* click **Numeric Data Variables (Y) >>**.



4. Click **OK**. The results are shown below:

Friedman Test				
Test Information				
H_0: Median 1 = Median 2 = ... = Median k (apart from Subject/Block Effect)				
H_a: At least one pair Median i ≠ Median j (apart from Subject/Block Effect)				
Results:	**Grass 1**	**Grass 2**	**Grass 3**	**Grass 4**
Count	12	12	12	12
Median	3.5	2	2	3.25
Chi-Square	8.097			
DF	3			
P-Value (2-sided)	0.0440			

The P-Value of 0.044 tells us that we reject H0. At least one pairwise set of medians are not equal, so there is a difference in the preference of grass types (i.e., there is a difference in treatment effects).

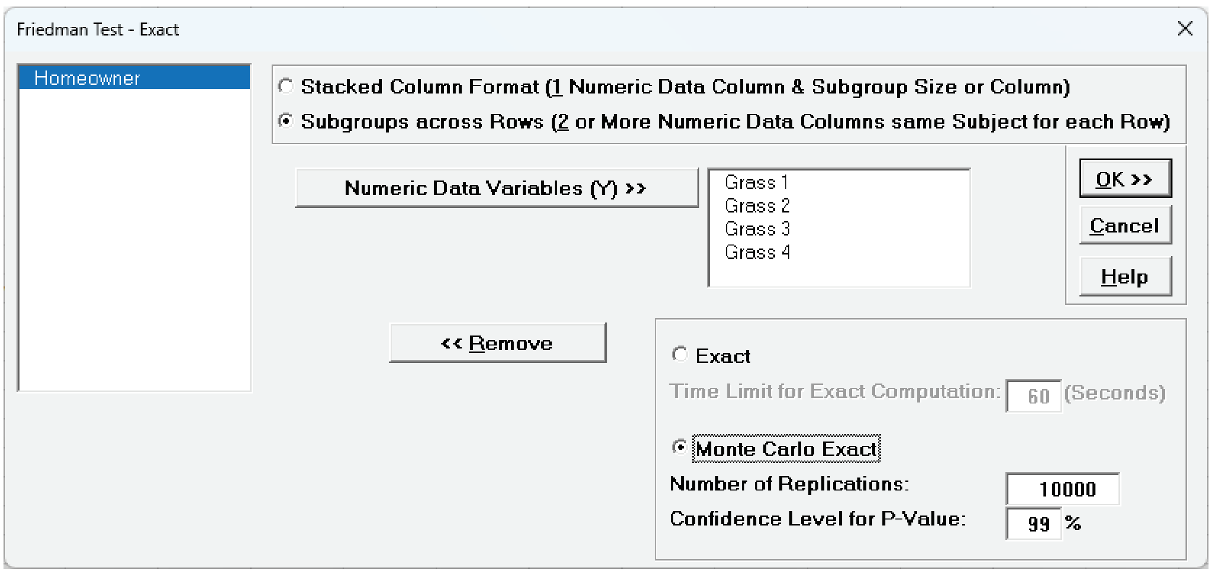
Friedman Test – Exact

We will now redo the above example to estimate the exact P-Value using Monte Carlo. Typically, this would not be necessary unless the sample sizes were smaller (N subjects <= 5), but this gives us continuity on the example.

1. Open **Conover Grass Type Experiment.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).

2. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Friedman Test – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

3. Ensure that **Subgroups across Rows (2 or More Numeric Data Columns same Subject for each Row)** is selected. Select *Grass 1* to *Grass 4,* click **Numeric Data Variables (Y) >>**. Select **Monte Carlo Exact** with the **Number of Replications** = 10000 and **Confidence Level for P-Value** = 99%.



Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

4. Click **OK**.

Friedman Test - Monte Carlo				
Test Information				
H_0: Median 1 = Median 2 = ... = Median k (apart from Subject/Block Effect)				
Ha: At least one pair Median i ≠ Median j (apart from Subject/Block Effect)				
Results:	Grass 1	Grass 2	Grass 3	Grass 4
Count	12	12	12	12
Median	3.5	2	2	3.25
Chi-Square	8.097			
DF	3			
Monte Carlo P-Value (2-sided)	0.0423			
Monte Carlo P-Value 99% CI Upper	0.0475			
Monte Carlo P-Value 99% CI Lower	0.0371			

The Monte Carlo P-Value is 0.0423 with a 99% confidence interval of 0.0371 to 0.0475. This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). So, we reject H0: at least one pairwise set of medians are not equal, i.e., there is a difference in the preference of grass types (or there is a difference in treatment effects).

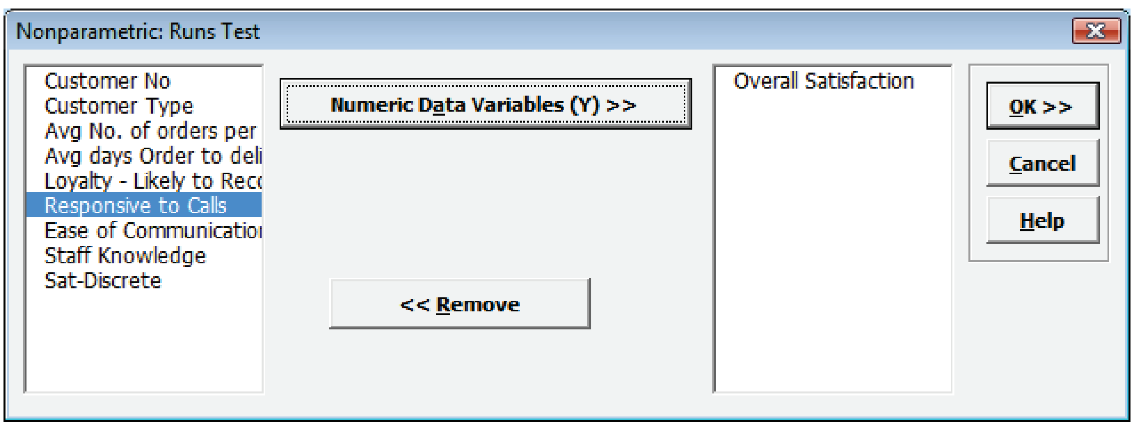
Note that the large sample (asymptotic) P-Value of 0.044 also lies within the Monte Carlo exact confidence interval.

Part N – Nonparametric Runs Test

Nonparametric Runs Test for Randomness

The nonparametric runs test provides a test for randomness or independence. The null hypothesis is H0: The data is random (or independent). The alternative hypothesis is Ha: The data is not random (or independent). Note that this test is also provided as an option in Run Charts (**SigmaXL > Graphical Tools > Run Chart**). In addition to providing an overall test for randomness, 4 tests are performed to detect Clustering, Mixtures, Trends, and Oscillations. If any of these patterns are significant (typically using α = 0.01), we would need to take corrective action before proceeding with further statistical analysis. (Note that SigmaXL will highlight any P-Values < .05 in red.)

1. Open **Customer Data.xlsx**, click on **Sheet 1** tab (or press **F4** to activate last worksheet).
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests > Runs Test**. Ensure that the entire data table is selected. If not, check Use **Entire Data Table**.
3. Click **Next**. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**:



4. Click **OK**. The resulting report is shown:

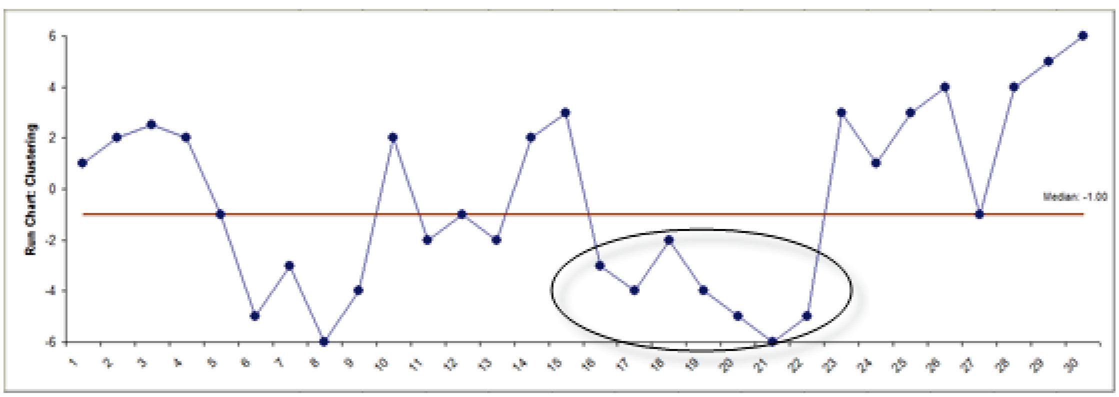
Nonparametric Runs Test	
Test Information	**Overall Satisfaction**
Number of Runs about Median:	44
Expected Number of Runs about Median:	51
Number of Points above Median:	50
Number of Points equal to or below Median:	50
P-Value for Clustering:	0.0797
P-Value for Mixtures:	0.9203
P-Value for Lack of Randomness (2-Sided):	0.1594
Number of Runs Up or Down:	64
Expected Number of Runs Up or Down:	66.333
P-Value for Trends:	0.2883
P-Value for Oscillation:	0.7117

5. With all of the P-Values being greater than 0.01, we fail to reject H0, and conclude that the data is random (or statistically independent). Recall from the run chart of this data that there were no obvious trends or patterns.

Nonparametric Runs Test for Randomness - Examples

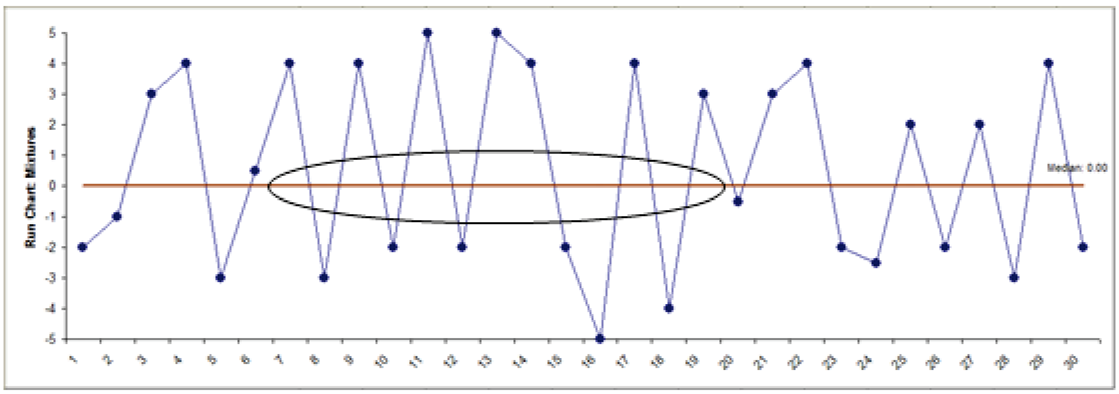
Examples of Clustering, Mixtures, Trends, and Oscillations are given below using the Run Chart to illustrate. The data for these examples are given in the file **Runs Test Example Data.xlsx**. (Use **SigmaXL > Graphical Tools > Run Charts** to create the run charts).

a. **Clustering** appears as a group of points in one area of the chart. It may indicate special cause variation such as sampling or measurement problems.



Nonparametric Runs Test: Clustering	
Number of Runs about Median:	9
Expected Number of Runs about Median:	15.933
Number of Points above Median:	14
Number of Points equal to or below Median:	16
P-Value for Clustering:	0.0048
P-Value for Mixtures:	0.9952
P-Value for Lack of Randomness (2-Sided):	0.0096
Number of Runs Up or Down:	17
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.1168
P-Value for Oscillation:	0.8832

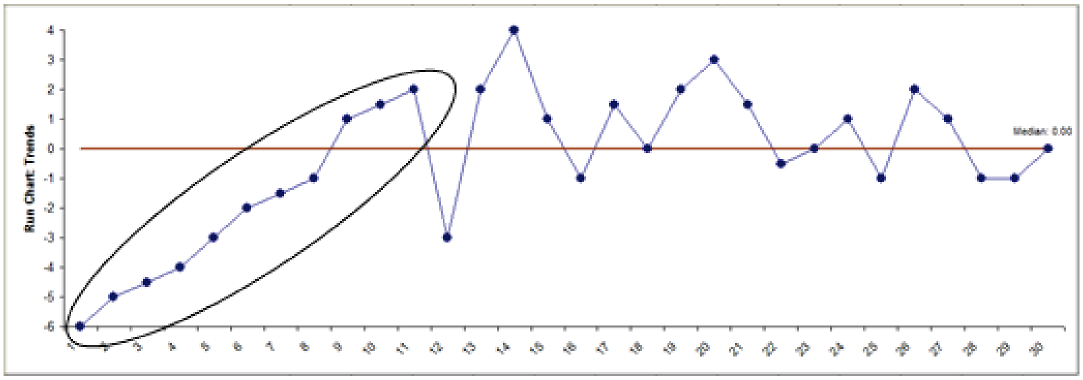
b. **Mixtures** appear as an absence of data points near the center line. A mixture may indicate a bimodal distribution due to a regular change of shift, machinery, or raw materials**.**



Nonparametric Runs Test: Mixtures	
Number of Runs about Median:	23
Expected Number of Runs about Median:	16
Number of Points above Median:	15
Number of Points equal to or below Median:	15
P-Value for Clustering:	0.9954
P-Value for Mixtures:	0.0046
P-Value for Lack of Randomness (2-Sided):	0.0093
Number of Runs Up or Down:	22
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.8514
P-Value for Oscillation:	0.1486

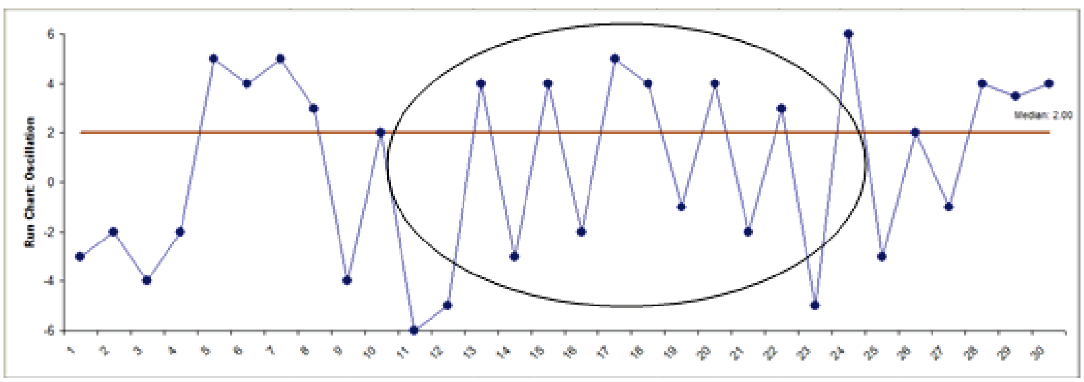
Note that the P-Value for Mixtures = 1 – P-Value for Clustering. They are mutually exclusive. The P-Value for Lack of Randomness = 2 * minimum of (P-Value Clustering, P-Value Mixtures).

c. **Trends** appear as an upward or downward drift in the data and may be due to special causes such as tool wear.



Nonparametric Runs Test: Trends	
Number of Runs about Median:	13
Expected Number of Runs about Median:	15.733
Number of Points above Median:	13
Number of Points equal to or below Median:	17
P-Value for Clustering:	0.1504
P-Value for Mixtures:	0.8496
P-Value for Lack of Randomness (2-Sided):	0.3008
Number of Runs Up or Down:	13
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.0015
P-Value for Oscillation:	0.9985

d. **Oscillations** appear as rapid up/down fluctuations indicating process instability.



Nonparametric Runs Test: Oscillation	
Number of Runs about Median:	16
Expected Number of Runs about Median:	15.933
Number of Points above Median:	14
Number of Points equal to or below Median:	16
P-Value for Clustering:	0.5099
P-Value for Mixtures:	0.4901
P-Value for Lack of Randomness (2-Sided):	0.9801
Number of Runs Up or Down:	25
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.9914
P-Value for Oscillation:	0.0086

Note that the P-Value for Trends = 1 – P-Value for Oscillation. They are mutually exclusive.

Nonparametric Runs Test for Randomness - Exact

The Nonparametric Runs test does not assume that the sample data are normally distributed, but it does assume that the test statistic follows a Normal distribution when computing the "large sample" or "asymptotic" P-Value. With a small sample size (N <= 50), this approximation may be invalid, so exact methods should be used. SigmaXL computes the exact P-Values utilizing permutations.

It is important to note that while exact P-Values are "correct," they do not increase (or decrease) the power of a small sample test, so they are not a solution to the problem of failure to detect a trend due to inadequate sample size.

Clustering, Mixtures and Lack of Randomness (Runs Above/Below)

If Count (N) is greater than 1000, the exact P-Value is estimated using a continuity-corrected normal approximation. Since the Runs Test Exact P-Value is computed very quickly for sample sizes as large as 1000, Monte Carlo P-Values are not required.

Trends and Oscillation (Runs Up/Down)

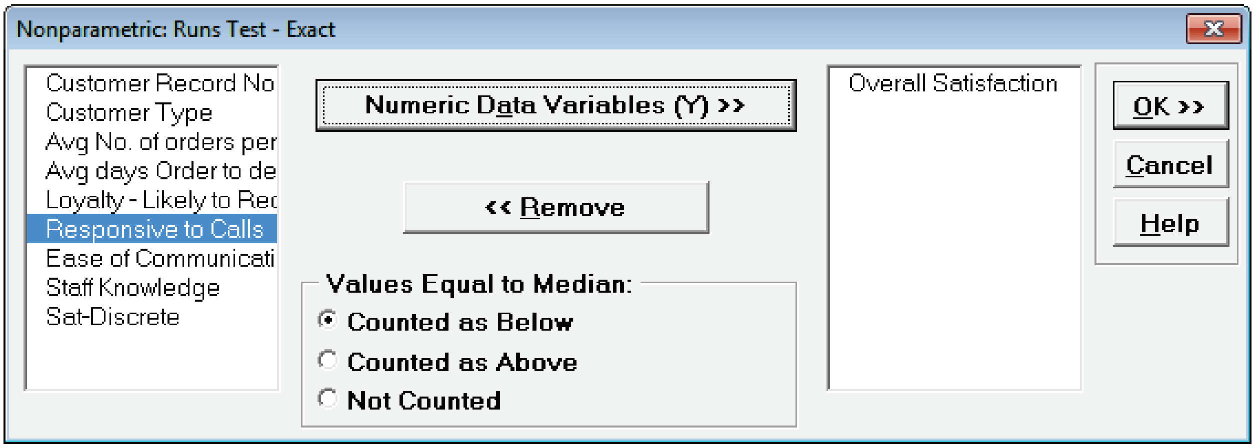
Exact P-Values are derived from published tables, given the sample size and the number of up/down runs. The exact tables apply to N <= 50. If N > 50, a continuity-corrected normal approximation is used.

Exact P-Values are not available in Run Charts.

For further details and references refer to the Appendix **Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests**.

We will now redo the Customer Data Overall Satisfaction example to compute exact P-Values. Typically, this would not be necessary unless the sample sizes were smaller (N <= 50), but this gives us continuity on the example. We will also consider the above small sample examples later.

1. Open **Customer Data.xlsx**, click **Sheet 1** tab.
2. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Runs Test – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.
3. Select *Overall Satisfaction*, click **Numeric Data Variables (Y) >>.** Set **Values Equal to Median:** to **Counted as Below**.



Note on Options for Clustering, Mixtures and Lack of Randomness (Runs Above/Below)

If none of the observations are equal to the sample median, the options for "**Values Equal to Median:**" do not affect the exact P-Value. That is the case with this Customer Data example. In cases where there are observations equal to the sample median, then the exact P-Value will be different for each option selected. For regular Runs Test, SigmaXL uses **Counted as Below** (in agreement with Minitab). StatXact uses **Counted as Above** for the Runs Test. Matlab uses **Not Counted.** Users should try **Counted as Below** and **Counted as Above** to ensure that the P-Values agree on reject or fail to reject H0.

Note on Options for Trends and Oscillation (Runs Up/Down)

The **Values Equal to Median** options also apply to the Runs Up/Down Test: zeros of the first order differences (i.e., two consecutive values that are the same). **Counted as Below** denotes "counted as negative." **Counted as Above** denotes "counted as positive." **Not Counted** denotes that zero first order differences are deleted. This consolidation of options was done to keep the dialog simple.

Missing values are ignored for all exact runs tests.

Use of these options will be demonstrated in the small sample example.

4. Click **OK**. Resulting output:

Nonparametric Runs Test - Exact	
Test Information	**Overall Satisfaction**
Number of Runs about Median:	44
Expected Number of Runs about Median:	51
Number of Points equal to or below Median:	50
Number of Points above Median:	50
Exact P-Value for Clustering:	0.0957
Exact P-Value for Mixtures:	0.9345
Exact P-Value for Lack of Randomness (2-Sided):	0.1914
Number of Runs Up or Down (zeros are negative):	64
Expected Number of Runs Up or Down:	66.333
P-Value for Trends:	0.3304
P-Value for Oscillation:	0.7512

With all of the P-Values being greater than 0.01 (alpha = .01 is preferred for the Runs Test to minimize false alarms), we fail to reject H0, and conclude that the data is random (or statistically independent).

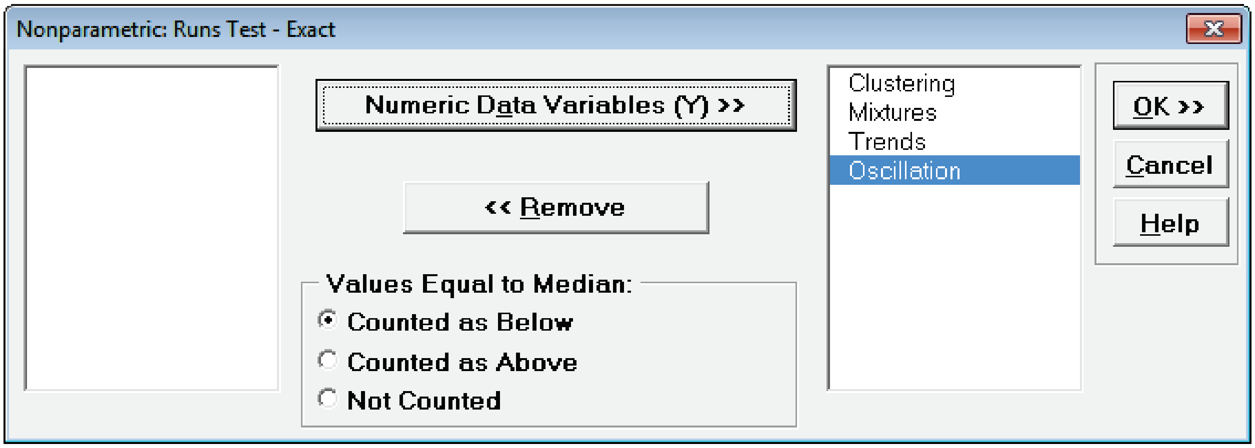
The Above/Below and Up/Down Runs counts are identical to the above “large sample” or “asymptotic” results. The Clustering/Mixtures/Randomness (Above/Below) Exact P-Values are close but slightly different. This was expected because the sample size is reasonable (N > 50), so the “large sample” P-Values are valid using a normal approximation.

The P-Values for Trends and Oscillation (Up/Down) use a normal approximation because the sample size is greater than 50. They are, however, slightly different than the “large sample” above because a continuity correction is now applied to the normal approximation.

5. Now we will consider the small sample examples. Open **Runs Test Example Data.xlsx**, click **Runs Test Example Data** tab.

6. Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Runs Test – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.

7. Select *Clustering* and shift click to *Oscillation*, click **Numeric Data Variables (Y) >>.** Set **Values Equal to Median:** to **Counted as Below**.



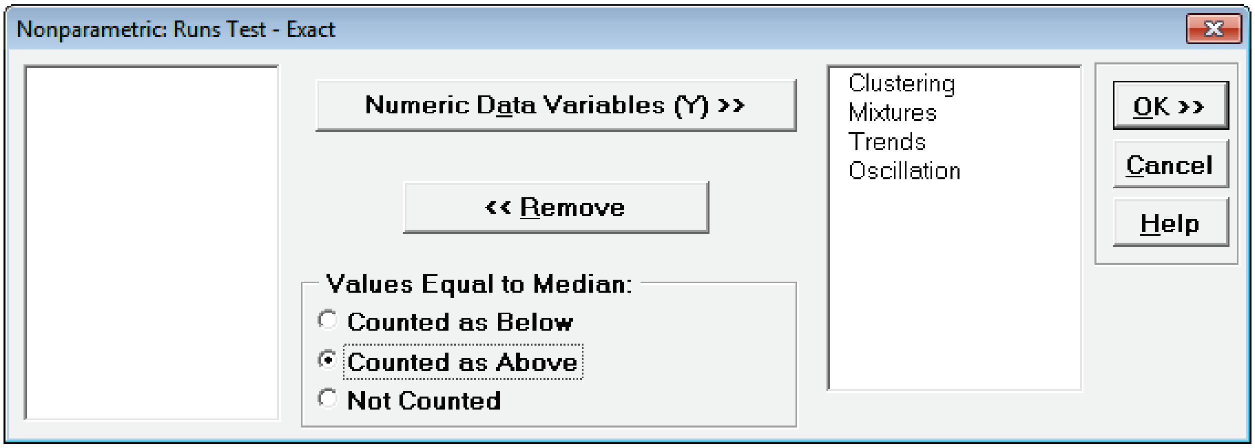
8. Click **OK**. Resulting output:

Nonparametric Runs Test - Exact				
Test Information	**Clustering**	**Mixtures**	**Trends**	**Oscillation**
Number of Runs about Median:	9	23	13	16
Expected Number of Runs about Median:	15.933	16	15.733	15.933
Number of Points above Median:	14	15	13	14
Number of Points equal to or below Median:	16	15	17	16
Exact P-Value for Clustering:	0.0073	0.9977	0.1980	0.5854
Exact P-Value for Mixtures:	0.9976	0.0070	0.8887	0.5664
Exact P-Value for Lack of Randomness (2-Sided):	0.0135	0.0139	0.3451	1.0000
Number of Runs Up or Down (zeros are negative):	17	22	13	25
Expected Number of Runs Up or Down:	19.667	19.667	19.667	19.667
Exact P-Value for Trends:	0.1662	0.8988	0.0028	0.9968
Exact P-Value for Oscillation:	0.9213	0.2072	0.9994	0.0130

The Exact P-Values are close to the above “large sample” or “asymptotic” results, but note that some of the values are now greater than .01 so they would fail-to reject H0. Note that this does not imply that the large sample runs test is more powerful, but rather we cannot conclude some of the previously identified patterns at the 99% confidence level.

We will now rerun the analysis using the **Counted as Above** option.

9. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Set **Values Equal to Median:** to **Counted as Above**.



10. Click **OK**. Results:

Nonparametric Runs Test - Exact				
Test Information	**Clustering**	**Mixtures**	**Trends**	**Oscillation**
Number of Runs about Median:	9	23	12	20
Expected Number of Runs about Median:	15.733	16	15.933	15.933
Number of Points equal to or above Median:	17	15	16	16
Number of Points below Median:	13	15	14	14
Exact P-Value for Clustering:	0.0084	0.9977	0.1007	0.9574
Exact P-Value for Mixtures:	0.9973	0.0070	0.9524	0.0919
Exact P-Value for Lack of Randomness (2-Sided):	0.0127	0.0139	0.1926	0.1394
Number of Runs Up or Down (zeros are positive):	17	22	13	25
Expected Number of Runs Up or Down:	19.667	19.667	19.667	19.667
Exact P-Value for Trends:	0.1662	0.8988	0.0028	0.9968
Exact P-Value for Oscillation:	0.9213	0.2072	0.9994	0.0130

Note the change in **Trends** and **Oscillation** Number of Runs and resulting dramatic change to the Exact P-Values.

In conclusion, when using Runs Test Exact, always try **Counted As Below** and **Counted as Above** to ensure that the P-Values agree with each other. Also, whenever you have a small sample size and are performing a nonparametric test, always use Exact.

Part O – Attribute/Discrete Data Tests

1 Proportion Test and Confidence Interval Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates > 1 Proportion Test and Confidence Interval** or **SigmaXL > Statistical Tools > Basic Statistical Templates > 1 Proportion Test and Confidence Interval**.

See **Basic Statistical Templates – 1 Proportion Test and Confidence Interval** for an example of the 1 Proportion Test and Confidence Interval template.

2 Proportions Test and Confidence Interval Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates > 2 Proportions Test and Confidence Interval** or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Proportions Test and Confidence Interval**.

See below for an example of the 2 Proportions Test and Confidence Interval template.

2 Proportions Equivalence Test Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates > 2 Proportions Equivalence Test** or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Proportions Equivalence Test**.

See **Basic Statistical Templates – 2 Proportions Equivalence Test** for an example of the 2 Proportions Equivalence Test template.

One-Way Chi-Square Goodness-of-Fit Template

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > One-Way Chi-Square Goodness-of-Fit Test** (or **SigmaXL > Statistical Tools > Chi-Square Tests > One Way Chi-Square Goodness-of-Fit Template**) to access the One-Way Chi-Square Goodness-of-Fit Test calculator.

See **Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test** for an example of the One-Way Chi-Square Goodness-of-Fit Test template.

One-Way Chi-Square Goodness-of-Fit Template - Exact

Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > One-Way Chi-Square Goodness-of-Fit Test Exact** (or **SigmaXL > Statistical Tools > Chi-Square Tests – Exact > One Way Chi-Square Goodness-of-Fit Exact Template**) to access the One-Way Chi-Square Goodness-of-Fit Test – Exact calculator.

See **Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test - Exact** for an example of the One-Way Chi-Square Goodness-of-Fit Test - Exact template.

1 Poisson Rate Test and Confidence Interval Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
1 Poisson Rate Test and Confidence Interval or **SigmaXL > Statistical Tools > Basic Statistical Templates > 1 Poisson Rate Test and Confidence Interval**.

See **Basic Statistical Templates – 1 Poisson Rate Test and Confidence Interval** for an example of the 1 Poisson Rate Test and Confidence Interval template.

2 Poisson Rates Test and Confidence Interval Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
2 Poisson Rates Test and Confidence Interval or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Poisson Rates Test and Confidence Interval**.

See **Basic Statistical Templates – 2 Poisson Rates Test and Confidence Interval** for an example of the 2 Poisson Rates Test and Confidence Interval template.

2 Poisson Rates Equivalence Test Template

Click **SigmaXL > Templates and Calculators** > **Basic Statistical Templates >**
2 Poisson Rates Equivalence Test or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Poisson Rates Equivalence Test**.

See **Basic Statistical Templates – 2 Poisson Rates Equivalence Test** for an example of the 2 Poisson Equivalence Test template.

2 Proportions Test and Confidence Interval Template Example

1. We begin with a scenario where Day Shift is running at 80% Yield and Night Shift has 70% Yield. This data is based on a random sample of 100 units for each Shift, each unit is either good or bad. Question: Is Day Shift running differently than Night Shift? Statistically, we call the Null Hypothesis, H0: Proportion P1 = Proportion P2; the alternative hypothesis, Ha is: P1≠P2. If the calculated P-Value < .05, then we reject the null hypothesis and conclude that Day Shift and Night Shift are different.

2. Click **SigmaXL > Templates & Calculators > Basic Statistical Templates > 2 Proportions Test and Confidence Interval** (or **SigmaXL > Statistical Tools > Basic Statistical Templates > 2 Proportions Test and Confidence Interval**).

3. Enter **x1** = 80, **n1** = 100, **x2** = 70, **n2** = 100 as shown:

2 Proportions Test and Confidence Interval			
Sample Data (user inputs):		**Sample 1**	**Sample 2**
Number of Events	x	80	70
Sample Size	n	100	100
Null Hypothesis (hypothesized difference)	H_0: P_1 - P_2 =	0	
Alternative Hypothesis	H_a: P_1 - P_2	Not Equal To	
Confidence Level (enter .95 for 95%)	100*(1-α)%	95.0%	
Hypothesis Test Method		Fisher's Exact	
Confidence Interval Method		Newcombe-Wilson Score	
Results:			
Sample proportion (x/n)		0.8000	0.7000
Sample proportion difference		0.1000	
alpha		0.0500	
Minimum expected value (should be >= 5 for normal approximation)		25.0000	
Fisher's Exact probability P-Value (2-sided)		0.1412	
Upper Confidence Limit (2-sided)		0.2167	
Lower Confidence Limit (2-sided)		-0.0202	

Since the Fisher's exact P-Value of 0.141 is greater than .05, we fail to reject H0. We do not have enough evidence to show that there is a significant difference between Day Shift and Night shift. This does not mean that we have proven that they are the same. In practice however, we either assume that they are the same or we collect more data.

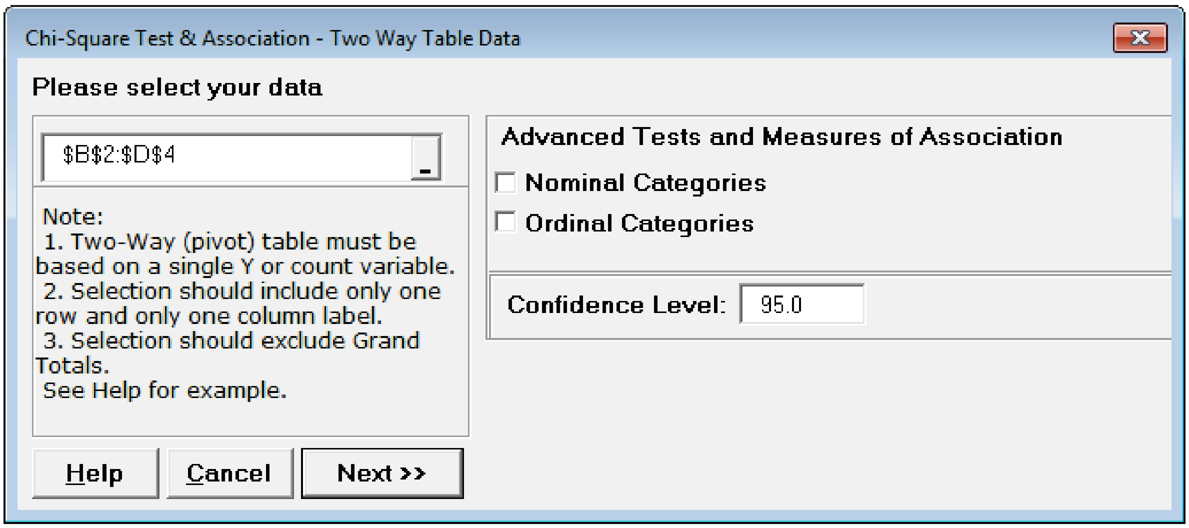
Note: Fisher's exact P-Values should be used for any real-world problem. The approximate P-Values based on the normal distribution are provided for instructional purposes, e.g., comparing to hand calculations.

4. Now enter **x1** = 160, **n1** = 200, **x2** = 140, **n2** = 200. Note that the Fisher's exact P-Value is now .028, so we reject H0.

Chi-Square Test – Two-Way Table Data

1. Open the file **Attribute Data.xlsx**, ensure that **Example 1** Sheet is active. This data is in Two – Way Table format, or pivot table format. Note that cells B2:D4 have been pre-selected.

2. Click **SigmaXL > Statistical Tools > Chi-Square Tests > Chi-Square Test & Association – Two-Way Table Data**. Note the selection of data includes the Row and Column labels (if we had Row and Column Totals these would NOT be selected). Do not check **Advanced Tests and Measures of Association**.



See Appendix [Chi-Square Tests and (Contingency) Table Associations](#) for the Chi-Square Table Statistics formula details.

3. Click **Next**. The resulting output is:

Chi-Square 2 Way Table Statistics		
Observed Counts	**Day Shift**	**Night Shift**
Pass	80	70
Fail	20	30
Expected Counts	**Day Shift**	**Night Shift**
Pass	75	75
Fail	25	25
Std. Residuals	**Day Shift**	**Night Shift**
Pass	0.577350	-0.577350
Fail	-1	1
Chi-Square	2.667	
DF	1	
P-Value	0.1025	

The P-Value matches that of the 2 Proportions test. Since the P-Value of 0.1 is greater than .05, we fail to reject H0.

4. Now click **Example 2** Sheet tab. The Yields have not changed but we have doubled the sample size. Repeat the above analysis. The resulting output is:

Chi-Square 2 Way Table Statistics		
Observed Counts	**Day Shift**	**Night Shift**
Pass	160	140
Fail	40	60
Expected Counts	**Day Shift**	**Night Shift**
Pass	150	150
Fail	50	50
Std. Residuals	**Day Shift**	**Night Shift**
Pass	0.816497	-0.816497
Fail	-1.414	1.414
Chi-Square	5.333	
DF	1	
P-Value	0.0209	

Since the P-Value is < .05, we now reject the Null Hypothesis, and conclude that Day Shift and Night Shift are significantly different. The Std. (Standardized) Residuals tell us that Day Shift failures are less than expected (assuming equal proportions), and Night Shift failures are more than expected.

Note, by doubling the sample size, we improved the power or sensitivity of the test.

5. Click the **Example 3** Sheet tab. In this scenario we have 3 suppliers and an additional marginal level. A random sample of 100 units per supplier is tested. The null hypothesis here is: No relationship between Suppliers and Pass/Fail/Marginal rates, but in this case, we can state it as No difference across suppliers. Redoing the above analysis (for selection B2:E5) yields the following:

Chi-Square 2 Way Table Statistics			
Observed Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	80	70	75
Fail	10	15	18
Marginal	10	15	7
Expected Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	75	75	75
Fail	14.333	14.333	14.333
Marginal	10.667	10.667	10.667
Std. Residuals	**Supplier A**	**Supplier B**	**Supplier C**
Pass	0.577350	-0.577350	0
Fail	-1.145	0.176090	0.968496
Marginal	-0.204124	1.327	-1.123
Chi-Square	6.008		
DF	4		
P-Value	0.1985		

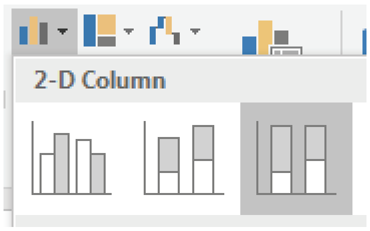
The P-Value tells us that we do not have enough evidence to show that there is a difference across the 3 suppliers.

6. Click the **Example 4** Sheet tab. Here we have doubled the sample size to 200 per supplier. Note that the percentages are identical to example 3. Redoing the above analysis yields the following:

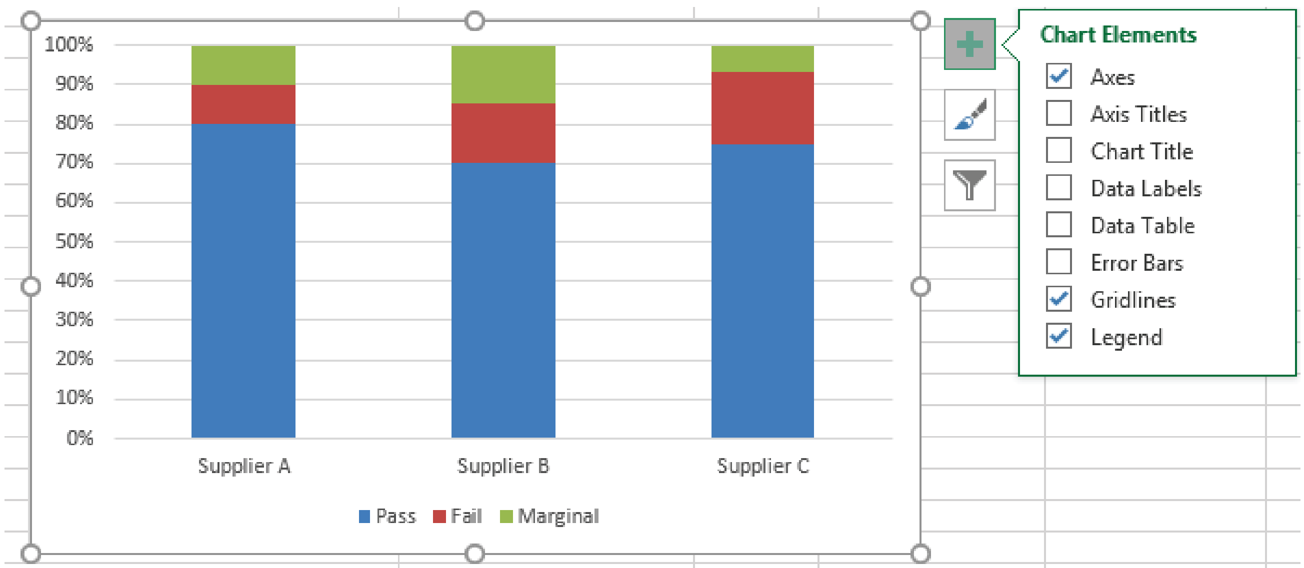
Chi-Square 2 Way Table Statistics			
Observed Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	160	140	150
Fail	20	30	36
Marginal	20	30	14
Expected Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	150	150	150
Fail	28.667	28.667	28.667
Marginal	21.333	21.333	21.333
Std. Residuals	**Supplier A**	**Supplier B**	**Supplier C**
Pass	0.816497	-0.816497	0
Fail	-1.619	0.249029	1.370
Marginal	-0.288675	1.876	-1.588
Chi-Square	12.016		
DF	4		
P-Value	0.0172		

With the P-Value < .05 we now conclude that there is a significant difference across suppliers. Examining the Std. (Standardized) Residuals tells us that Supplier A has fewer failures than expected (if there was no difference across suppliers), Supplier B has more marginal parts than expected and Supplier C has fewer marginal parts than expected.

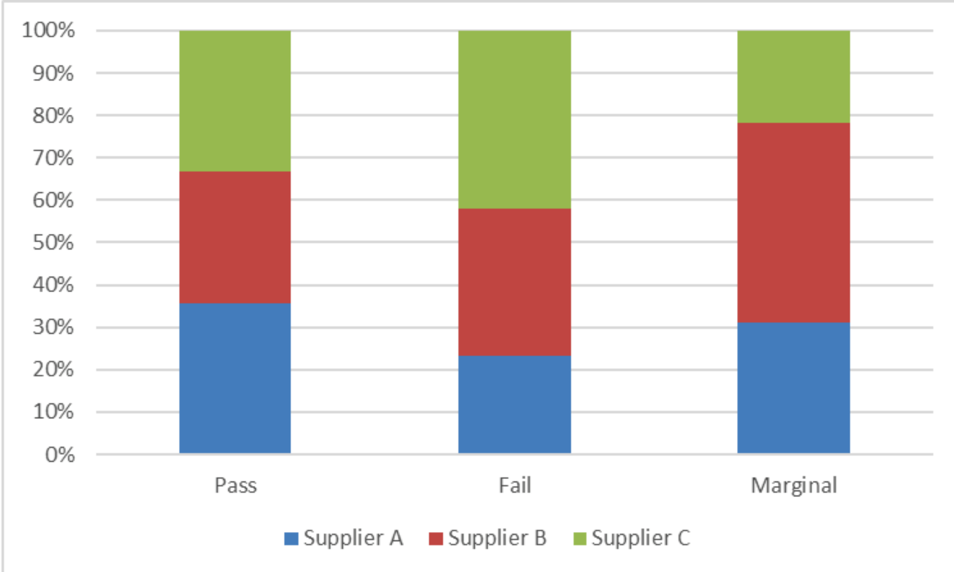
7. The table row and column cell percentages can be visualized using Excel's 100% Stacked Column Chart. Select cells **A3:D6** of the **Chi-Square** sheet. Click Excel's **Insert > Insert Column or Bar Chart** and select **100% Stacked Column** as shown.



8. Click to create the 100% stacked column chart. Uncheck the Chart Title as shown.



9. The rows and columns can easily be switched by clicking **Design > Switch Row/Column**



10. These charts make it easy to visualize the cell row and column percentages.

Chi-Square Test – Two-Way Table Data: Advanced Tests and Measures of Association – Nominal Categories

Checking the **Nominal Categories** option provides additional chi-square statistics and measures of association, including:

- Adjusted Residuals
    - Equivalent to normal z score
    - Red font highlight denotes significant cell residual value
    - Bold red highlight denotes significant cell residual value with Bonferroni adjustment
    - Note: red highlight is only active if Chi-Square P-Value is significant
- Cell's Contribution to Chi-Square
- Additional Chi-Square Tests
    - Likelihood Ratio
    - McNemar-Bowker Symmetry (Square Table)
        - For a 2x2 table, McNemar's test is equivalent to a paired two-proportions test, for example applicable to studying before versus after change in proportion on the same subject. The returned P-Value is exact, based on the binomial distribution.
        - Bowker extended McNemar's test for square tables larger than 2x2. The null hypothesis is that the table is symmetrical (i.e., symmetry of disagreement). This uses Chi-Square as the test statistic.
- Measures of Association for Nominal Categories
    - Pearson's Phi
        - Pearson's Phi is equivalent to Pearson's correlation coefficient for a 2x2 table. It is the most popular measure of association for 2x2 tables.
        - We recommend the following rules-of-thumb, adapted from Cohen (1988):
            - < 0.1 = Very Weak
            - 0.1 to < 0.3: Weak ("Small" Effect)
            - 0.3 to < 0.5: Moderate ("Medium" Effect)
            - > 0.5: Strong ("Large" Effect)
        - Although Phi is equivalent to Pearson's correlation for a 2x2 table, we recommend these rules-of-thumb for use in typical contingency tables, rather than those commonly used for correlation (i.e., > 0.9 = Strong).

- Cramer's V
  - Cramer's V is an extension of Phi for larger tables. It is the most popular measure of association for tables of any size.
  - It varies from 0 to 1, with 0 = no association and 1 = perfect association.
  - Use Cohen's rules-of-thumb given above for Phi.
- Contingency Coefficient
  - An alternative to Phi, varies from 0 to < 1.
  - Use Cohen's rules-of-thumb given above for Phi.
- Cohen's Kappa (Agreement - Square Table)
  - Kappa is used to measure agreement between two assessors evaluating the same parts or items.
  - For an extended Attribute Measurement Systems Analysis use **SigmaXL > Measurement Systems Analysis > Attribute MSA**.
    - For Attribute MSA used in Six Sigma quality, the recommendation is Kappa > 0.9 is strong agreement and < 0.7 is weak agreement, but for general use, the less stringent guidelines by Fleiss are recommended:

      Kappa: >= 0.75 or so signifies excellent agreement, for most purposes, and <= 0.40 or so signifies poor agreement.
    - See Appendix **Kappa** for further details.
- Goodman-Kruskal Lambda & Tau and Theil's Uncertainty
  - Measures of Proportional Reduction in Predictive Error. The basic concept is a measure that indicates how much knowing the value of the independent variable improves our ability to estimate the value of the dependent variable.
  - They are Directional Measures. If the Y dependent variable is in the **Rows Category**, then use the *Rows Dependent* measure. If the Y dependent variable is in the **Columns Category**, then use the *Cols Dependent* measure. If there is no clear X-Y dependent-independent relationship, then use the *Symmetric* measures (not available for Tau).
  - Use Cohen's rules-of-thumb for these measures.

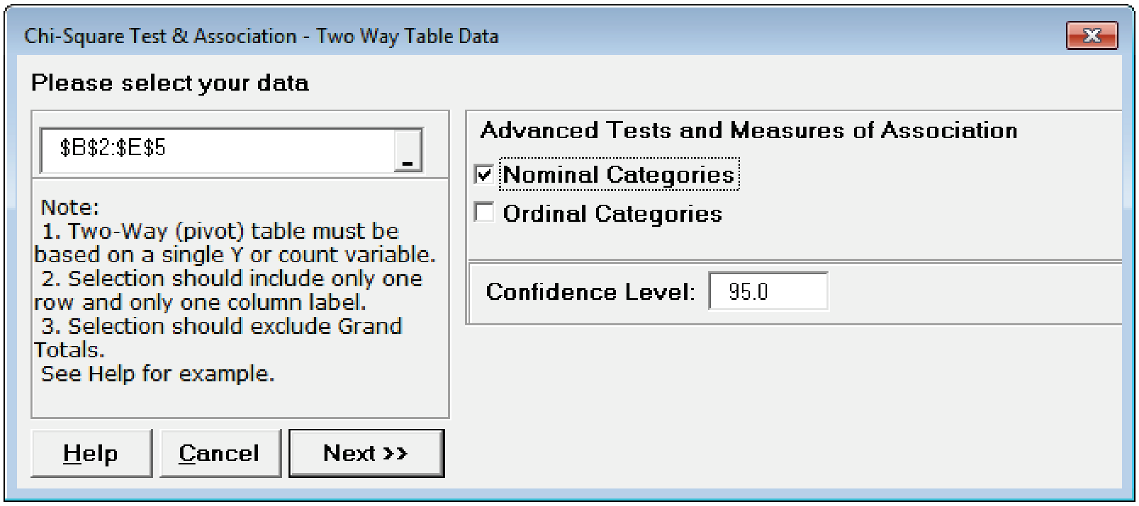
See Appendix **Chi-Square Tests and (Contingency) Table Associations** for further formula details and references. The following are external links with helpful presentations on measures of association for contingency tables:

- http://course1.winona.edu/bdeppa/STAT%20701%20Online/stat_701%20home.htm
  Click **4a) Measures of Association**. This is a narrated presentation, but a regular PowerPoint

is also available on the web site. It is part of a Biostatistics course by Dr. Brant Deppa, Department of Mathematics and Statistics, Winona State University.

- http://uregina.ca/~gingrich/ch11a.pdf by Dr. Paul Gingrich, University of Regina.

1. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Check **Nominal Categories** as shown:



2. Click **Next**. The resulting output is:

Chi-Square 2 Way Table Statistics			
Observed Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	160	140	150
Fail	20	30	36
Marginal	20	30	14
Expected Counts	**Supplier A**	**Supplier B**	**Supplier C**
Pass	150	150	150
Fail	28.667	28.667	28.667
Marginal	21.333	21.333	21.333
Std. Residuals	**Supplier A**	**Supplier B**	**Supplier C**
Pass	0.816497	-0.816497	0
Fail	-1.619	0.249029	1.370
Marginal	-0.288675	1.876	-1.588
Adjusted Residuals	**Supplier A**	**Supplier B**	**Supplier C**
Pass	2	-2	0
Fail	-2.14191972	0.329526111	1.812393613
Marginal	-0.374066	2.431429007	-2.05736301
Cell's Contribution to Chi-Square	**Supplier A**	**Supplier B**	**Supplier C**
Pass	0.66666667	0.666666667	0
Fail	2.62015504	0.062015504	1.875968992
Marginal	0.08333333	3.520833333	2.520833333

The adjusted residuals are equivalent to normal z values, so for a specified 95% confidence level, any value greater than 1.96 (or less than -1.96) is highlighted in red. This results in a slight difference in interpretation from that of the standardized residuals, but the 3 largest magnitude residuals are consistent.

Chi-Square	12.016			
DF	4			
P-Value	0.0172			

Additional Chi-Square Tests			
Test	**Chi-Square**	**DF**	**P-Value**
Likelihood Ratio	12.1425	4	0.0163
McNemar-Bowker Symmetry (Square Table)	189.9572	3	0.0000

Measures of Association for Nominal Categories				
Measure	**Value**	**Std. Error**	**95% Lower Bound**	**95% Upper Bound**
Pearson's Phi	0.1415			
Cramer's V	0.1001			
Contingency Coefficient	0.1401			
Cohen's Kappa (Agreement - Square Table)	0.0100	0.0211	-0.0314	0.0514
Goodman-Kruskal Lambda (Cols Dependent)	0.0650	0.0249	0.0162	0.1138
Goodman-Kruskal Lambda (Rows Dependent)	0.0000	0.0000	0.0000	0.0000
Goodman-Kruskal Lambda (Symmetric)	0.0473	0.0181	0.0118	0.0827
Goodman-Kruskal Tau (Cols Dependent)	0.0100	0.0057	0.0000	0.0211
Goodman-Kruskal Tau (Rows Dependent)	0.0095	0.0058	0.0000	0.0209
Theil's Uncertainty (Cols Dependent)	0.0092	0.0052	-0.0010	0.0194
Theil's Uncertainty (Rows Dependent)	0.0138	0.0078	-0.0015	0.0291
Theil's Uncertainty (Symmetric)	0.0110	0.0062	-0.0012	0.0233

As noted above, the Chi-Square P-Value tells us that there is a significant difference across suppliers, in other words, there is association between Supplier and Pass/Fail/Marginal, but it does not tell us the degree or strength of that association. Cramer's V is used for tables larger than 2x2 and from the rules-of-thumb, the 0.1 value is considered weak (or small effect).

Computing Odds Ratio and Confidence Interval for 2x2 Table

Odds Ratio and Confidence Intervals are not directly available for 2x2 Tables but can obtained using Logistic Regression. The table data must be rearranged to stacked column format and the response changed to “Event/Trial” format. **Example 2** of **Attribute Data.xlsx**:

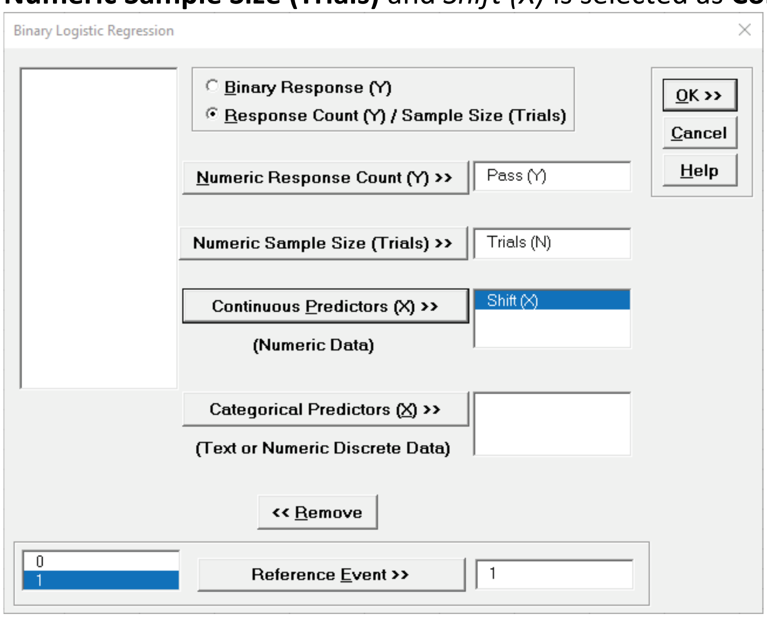
	Day Shift	Night Shift
Pass	160	140
Fail	40	60

would be changed to:

Shift (X)	Pass (Y)	Trials (N)
1	160	200
2	140	200

Note, coding Shift as continuous numeric is easier to interpret and is valid because it has a range of 1.

Analyze using **SigmaXL > Statistical Tools > Regression > Binary Logistic Regression**. Select **Response Count (Y) / Sample Size**. *Pass(Y)* is selected as **Numeric Response Count (Y),** *Trials (N)* is **Numeric Sample Size (Trials)** and *Shift (X)* is selected as **Continuous Predictors (X)**:



Click **OK**. The Odds Ratio and Confidence Limits are given as:

Odds Ratio	Lower 95% Odds Ratio	Upper 95% Odds Ratio
0.583333	0.368282	0.923960

The shift change 1 to 2 is 0.583 times as likely to produce passed (good) product with a 95% confidence interval of 0.368 to 0.924.

Chi-Square Test – Two-Way Table Data: Advanced Tests and Measures of Association – Ordinal Categories

Checking the **Ordinal Categories** option provides statistics and measures of association appropriate when both row and column category variables are ordinal:

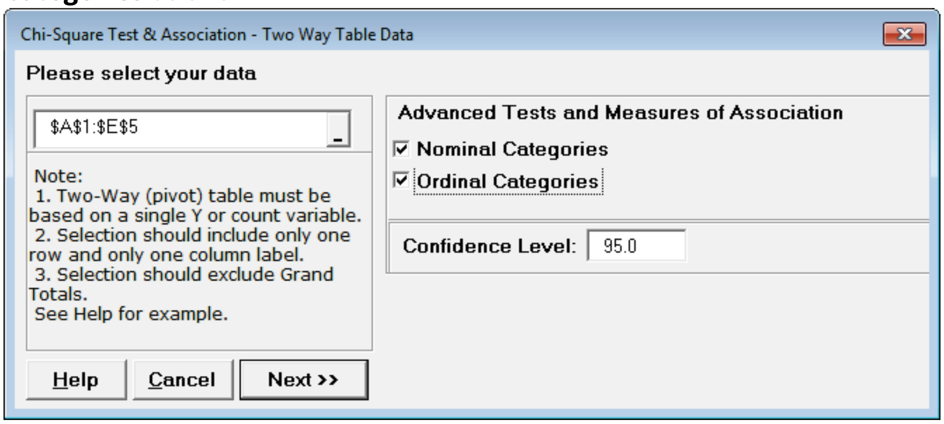
- Adjusted Residuals and Cell's Contribution to Chi-Square
- Tests of Association for Ordinal Categories
    - Concordant – Discordant
        - The P-Value for this hypothesis test is from Kendall's Tau-B, but is the same for all of the Concordant – Discordant ordinal measures: Tau-C, Gamma and Somers' D. See Agresti (2010). This may differ from other software using an approximation formula.
    - Spearman Rank Correlation
- Measures of Association for Ordinal Categories with Confidence Intervals
    - Spearman Rank Correlation
        - Equivalent to Pearson's correlation on ranks
    - Kendall's Tau-B (Square Table), Kendall-Stuart Tau-C (Rectangular Table), Goodman-Kruskal Gamma
        - Use Tau-B for square tables (no. rows = no. columns) and Tau-C for rectangular tables (no. rows <> no. columns).
        - Tau-B, Tau-C and Gamma are asymmetric measures, so will give the same result regardless of variable assignment to Rows and Columns.
        - These are all Concordant – Discordant measures.
    - Somers' D (Cols & Rows Dependent, Symmetric)
        - Also a Concordant – Discordant measure but directional. If the Y dependent variable is in the Rows Category, then use the Rows Dependent measure. If the Y dependent variable is in the Columns Category, then use the Cols Dependent measure. If there is no clear X-Y dependent-independent relationship, then use the Symmetric measure.
- SigmaXL provides rules-of-thumb for Kendall's Correlation in Ordinal Attribute MSA (strong association is > 0.8) and Pearson or Spearman Correlation (strong association is > 0.9), however these are in the context of measurement systems analysis, design of experiments or a controlled process study. For typical contingency table applications, we recommend the rules-of-thumb, adapted from Cohen (1988):

- 0.5+: Strong (Large Effect)
- 0.3 to < 0.5: Moderate (Medium Effect)
- 0.1 to < 0.3: Weak (Small Effect)
- < 0.1: Very Weak

See Appendix **Chi-Square Tests and (Contingency) Table Associations** for further formula details and references. See also Agresti (2010, Chapter 7 *Non-Model-Based Analysis of Ordinal Association*) and Deppa's presentation on Measures of Association at: http://course1.winona.edu/bdeppa/STAT%20701%20Online/stat_701%20home.htm.

1. Open the file **Attribute Data.xlsx**, click **Example 5 – Salary Sat** Sheet tab. This data is in two-way table format and has ordinal categories: *Salary* in the Rows and *Satisfaction Level* in the Columns. Note that cells A1:E5 have been pre-selected.

2. Click **SigmaXL > Statistical Tools > Chi-Square Tests > Chi-Square Test & Association – Two-Way Table Data**. Note the selection of data includes the Row and Column labels (if we had Row and Column Totals these would NOT be selected). Check **Nominal Categories** and **Ordinal Categories** as shown:



Tip: Even if the categories are ordinal, it is sometimes useful to select nominal categories as well for comparison purposes.

3. Click **Next**. The resulting output is:

Chi-Square 2 Way Table Statistics

Observed Counts	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
< 20K	5	6	19	12
20 to < 30K	5	7	20	14
30 to < 40K	3	12	28	24
>= 40K	1	2	18	26
Expected Counts	**Very Dissatisfied**	**Somewhat Dissatisfied**	**Somewhat Satisfied**	**Very Satisfied**
< 20K	2.911	5.614	17.673	15.802
20 to < 30K	3.188	6.149	19.356	17.307
30 to < 40K	4.644	8.955	28.193	25.208
>= 40K	3.257	6.282	19.777	17.683
Std. Residuals	**Very Dissatisfied**	**Somewhat Dissatisfied**	**Somewhat Satisfied**	**Very Satisfied**
< 20K	1.224	0.162972	0.315591	-0.956432
20 to < 30K	1.014757961	0.343393	0.146278	-0.794905
30 to < 40K	-0.762713	1.017372859	-0.036361522	-0.240586
>= 40K	-1.251	-1.708	-0.399632	1.978
Adjusted Residuals	**Very Dissatisfied**	**Somewhat Dissatisfied**	**Somewhat Satisfied**	**Very Satisfied**
< 20K	1.426134732	0.196736411	0.465932976	-1.360693153
20 to < 30K	1.196940639	0.419818923	0.218713418	-1.145298952
30 to < 40K	-0.967089216	1.33704372	-0.05844315	-0.37262257
>= 40K	-1.480071878	-2.095444235	-0.599449996	2.858762784
Cell's Contribution to Chi-Square	**Very Dissatisfied**	**Somewhat Dissatisfied**	**Somewhat Satisfied**	**Very Satisfied**
< 20K	1.499326463	0.026559799	0.099597859	0.914762153
20 to < 30K	1.029733719	0.117919038	0.02139728	0.631873485
30 to < 40K	0.581730668	1.035047535	0.00132216	0.057881515
>= 40K	1.564416624	2.918900046	0.15970582	3.911611766

Note: 4 out of 16 cells have expected counts less than 5.

Chi-Square	14.572
DF	9
P-Value	0.1034

Additional Chi-Square Tests			
Test	**Chi-Square**	**DF**	**P-Value**
Likelihood Ratio	15.4156	9	0.0801
McNemar-Bowker Symmetry (Square Table)	32.8921	6	0.0000

Measures of Association for Nominal Categories				
Measure	**Value**	**Std. Error**	**95% Lower Bound**	**95% Upper Bound**
Pearson's Phi	0.2686			
Cramer's V	0.1551			
Contingency Coefficient	0.2594			
Cohen's Kappa (Agreement - Square Table)	0.0752	0.0428	-0.0087	0.1592
Goodman-Kruskal Lambda (Cols Dependent)	0.0684	0.0547	0.0000	0.1756
Goodman-Kruskal Lambda (Rows Dependent)	0.0296	0.0556	0.0000	0.1385
Goodman-Kruskal Lambda (Symmetric)	0.0476	0.0483	0.0000	0.1424
Goodman-Kruskal Tau (Cols Dependent)	0.0230	0.0136	0.0000	0.0497
Goodman-Kruskal Tau (Rows Dependent)	0.0233	0.0115	0.0008	0.0458
Theil's Uncertainty (Cols Dependent)	0.0322	0.0153	0.0022	0.0621
Theil's Uncertainty (Rows Dependent)	0.0279	0.0134	0.0017	0.0541
Theil's Uncertainty (Symmetric)	0.0299	0.0143	0.0019	0.0578

Tests of Association for Ordinal Categories		
Test	**Value**	**P-Value**
Concordant - Discordant	2707	0.0009
Spearman Rank Correlation	0.2205	0.0016

Measures of Association for Ordinal Categories				
Measure	**Value**	**Std. Error**	**95% Lower Bound**	**95% Upper Bound**
Spearman Rank Correlation	0.2205	0.0669	0.0834	0.3495
Kendall's Tau-B (Square Table)	0.1899	0.0570	0.0782	0.3017
Kendall-Stuart Tau-C (Rectangular Table)	0.1769	0.0535	0.0721	0.2817
Goodman-Kruskal Gamma	0.2700	0.0795	0.1141	0.4258
Somers' D (Cols Dependent)	0.1791	0.0540	0.0732	0.2850
Somers' D (Rows Dependent)	0.2014	0.0603	0.0833	0.3196
Somers' D (Symmetric)	0.1896	0.0569	0.0781	0.3012

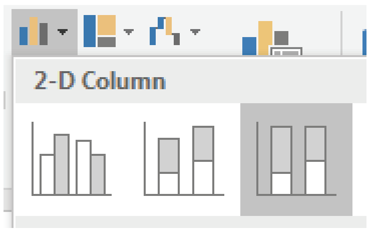
Note that the Chi-Square P-Value is 0.1, indicating that there is no significant association between Salary and Satisfaction when they are treated as nominal categories (although the significant result for McNemar-Bowker does show that there is lack of symmetry in the off diagonals).

Since the Chi-Square P-Value is not significant, the **Adjusted Residuals** are not highlighted, even though some values are greater than 1.96 (and less than -1.96). This follows the concept used in ANOVA called "Fisher Protected" where one considers the significance of post-hoc tests only when the overall test is significant.

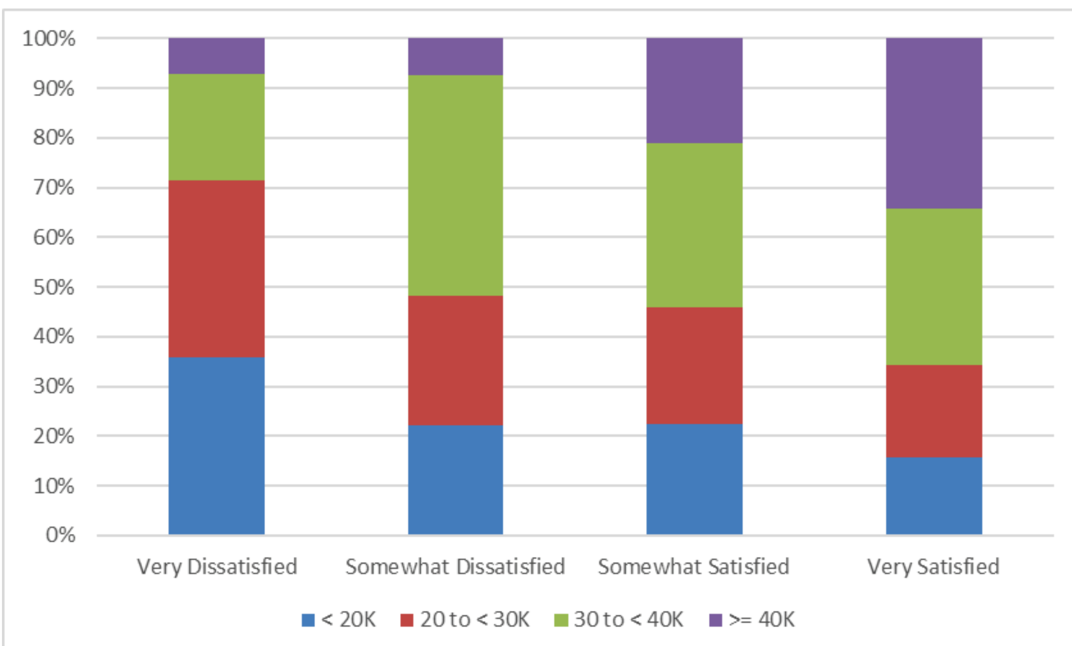
Note: 4 out of 16 cells have expected counts less than 5. If more than 20% of the cells have expected counts less than 5 (or if any of the cells have an expected count less than 1), the Chi-Square approximation may be invalid, and Fisher's Exact should be used. This will be discussed later, but for this example the Fisher's Monte-Carlo Exact P-Value = 0.095 so does not change the interpretation of the results for the above Chi-Square analysis.

When Salary and Satisfaction are treated as ordinal categories, the more powerful **Concordant – Discordant** and **Spearman Rank Correlation** P-Values clearly show that there is a significant association. The **Measures of Association for Ordinal Categories** show that this is positive, i.e., an increase in Salary is associated with an increase in Satisfaction. However, using the rules-of-thumb given above, we see that the association is weak.

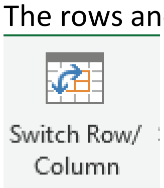
4. The table row and column cell percentages can be visualized using Excel's 100% Stacked Column Chart. Select cells **A3:E7** of the **Chi-Square** sheet. Click Excel's **Insert > Insert Column or Bar Chart** and select **100% Stacked Column** as shown.

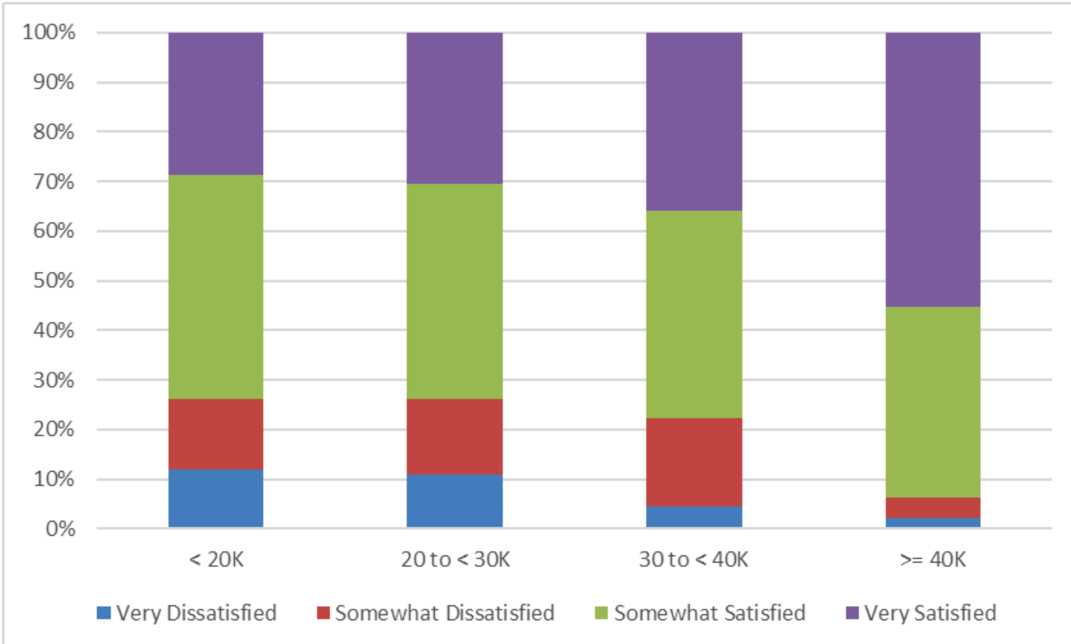


5. Click to create the 100% stacked column chart (uncheck the Chart Title):



6. The rows and columns can easily be switched by clicking **Design > Switch Row/Column**



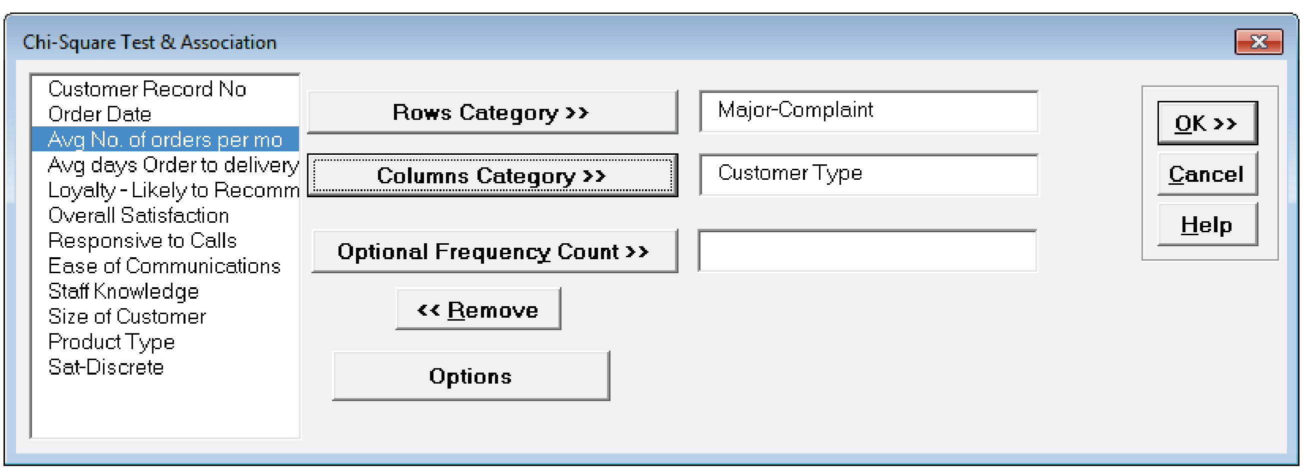


Chi-Square Test (Stacked Column Format Data)

1. Open **Customer Data.xlsx**. Click **Sheet 1** tab. The discrete data of interest is Complaints and Customer Type, i.e., does the type of complaint differ across customer type? Formally the Null Hypothesis is that there is no relationship (or independence) between Customer Type and Complaints.

2. Click **SigmaXL > Statistical Tools > Chi-Square Tests > Chi-Square Test & Association**. Ensure that the entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.

3. Select *Major-Complaint*, click **Rows Category >>**; select *Customer Type*, click **Columns Category >>**. We will not be using **Options** in this example.



4. Click **OK**. Results:

Chi-Square Test			
Major-Complaint - Customer Type			
Observed Counts	**1**	**2**	**3**
Difficult-to-order	5	9	5
Not-available	2	0	2
Order-takes-too-long	1	3	6
Return-calls	19	28	13
Wrong-color	4	2	1
Expected Counts	**1**	**2**	**3**
Difficult-to-order	5.890	7.980	5.130
Not-available	1.240	1.680	1.08
Order-takes-too-long	3.100	4.200	2.700
Return-calls	18.600	25.200	16.200
Wrong-color	2.170	2.940	1.890
Std. Residuals	**1**	**2**	**3**
Difficult-to-order	-0.366718	0.361076	-0.057396402
Not-available	0.682500	-1.296	0.885270
Order-takes-too-long	-1.193	-0.585540	2.008
Return-calls	0.092747779	0.557773	-0.795046
Wrong-color	1.242	-0.548219	-0.647380
Chi-Square	12.211		
DF	8		
p-value	0.1420		
Note: 9 out of 15 cells have expected counts less than 5.			

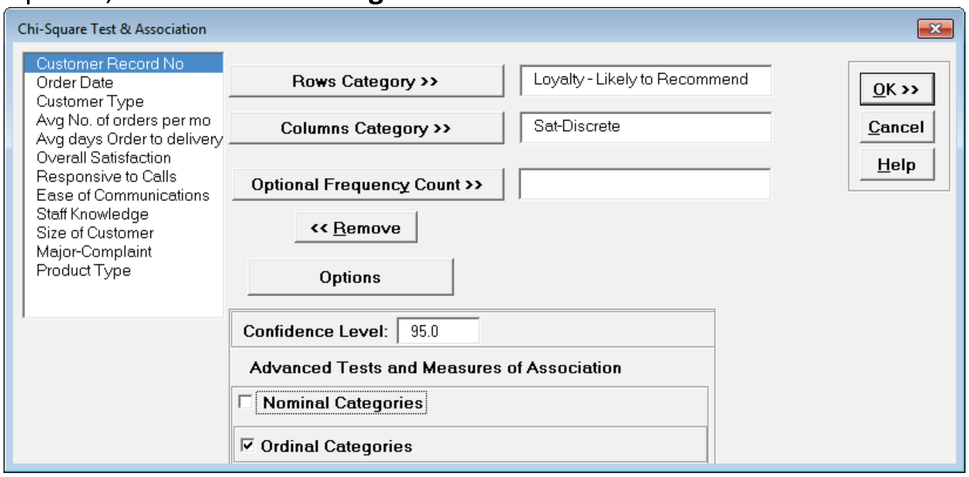
With the P-Value = 0.142 we fail to reject H0, so we do not have enough evidence to show a difference in customer complaints across customer types.

Note: 9 out of 15 cells have expected counts less than 5. If more than 20% of the cells have expected counts less than 5 (or if any of the cells have an expected count less than 1), the Chi-Square approximation may be invalid. Use Chi-Square Test – Fisher's Exact).

Tip: Use Advanced Pareto Analysis and Excel's 100% Stacked Column Chart to complement Chi-Square Analysis.

Caution: When using stacked column format data with Ordinal Category variables that are text, SigmaXL will sort alphanumerically which may not result in the correct ascending order for analysis. We recommend coding text Ordinal variables as numeric (e.g., 1,2,3) or modified text (e.g., Sat_0, Sat_1).

5. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Select *Loyalty – Likely to Recommend*, click **Rows Category >>**; select *Sat-Discrete*, click **Columns Category >>**. Click Options, check **Ordinal Categories**.



Sat-Discrete is derived from Overall Satisfaction where a score >= 3.5 is considered a 1, and scores < 3.5 are considered a 0.

6. Click **OK**. Results:

Chi-Square Test			
Loyalty - Likely to Recommend - Sat-Discrete			
Observed Counts		**0**	**1**
	1	2	0
	2	8	2
	3	17	16
	4	6	37
	5	0	12
Expected Counts		**0**	**1**
	1	0.660000	1.340
	2	3.300	6.700
	3	10.890	22.110
	4	14.190	28.810
	5	3.960	8.040
Std. Residuals		**0**	**1**
	1	1.649	-1.158
	2	2.587	-1.816
	3	1.852	-1.299
	4	-2.174	1.526
	5	-1.990	1.397
Adjusted Residuals		**0**	**1**
	1	2.03555	-2.036
	2	**3.33183**	**-3.332**
	3	2.76346	-2.763
	4	**-3.51818**	**3.518**
	5	-2.59161	2.592
Cell's Contribution to Chi-Square		**0**	**1**
	1	2.72061	1.34
	2	6.69394	3.297
	3	3.42811	1.688
	4	4.727	2.328
	5	3.96	1.95
Note: 4 out of 10 cells have expected counts less than 5.			

Chi-Square	32.134
DF	4
P-Value	0.0000

Tests of Association for Ordinal Categories		
Test	Value	P-Value
Concordant - Discordant	1417	0.0000
Spearman Rank Correlation	0.5558	0.0000

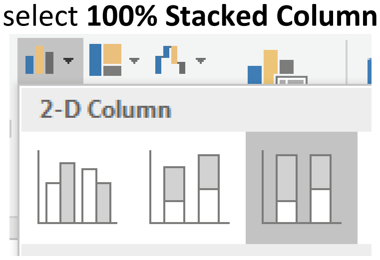
Measures of Association for Ordinal Categories				
Measure	Value	Std. Error	95% Lower Bound	95% Upper Bound
Spearman Rank Correlation	0.5558	0.0943	0.3910	0.6861
Kendall's Tau-B (Square Table)	0.5163	0.0637	0.3915	0.6410
Kendall-Stuart Tau-C (Rectangular Table)	0.5668	0.0814	0.4073	0.7263
Goodman-Kruskal Gamma	0.8330	0.0678	0.7001	0.9660
Somers' D (Cols Dependent)	0.4159	0.0553	0.3076	0.5243
Somers' D (Rows Dependent)	0.6409	0.0795	0.4850	0.7967
Somers' D (Symmetric)	0.5044	0.0622	0.3826	0.6263

As expected, there is a strong positive association between Loyalty and Discrete Satisfaction. Note: Kendall-Stuart Tau-C should be used here rather than Tau-B because it is a rectangular table. Since Satisfaction leads to Loyalty, Loyalty is the Dependent variable, so Rows Dependent Somers' D should be used rather than Cols Dependent or Symmetric.

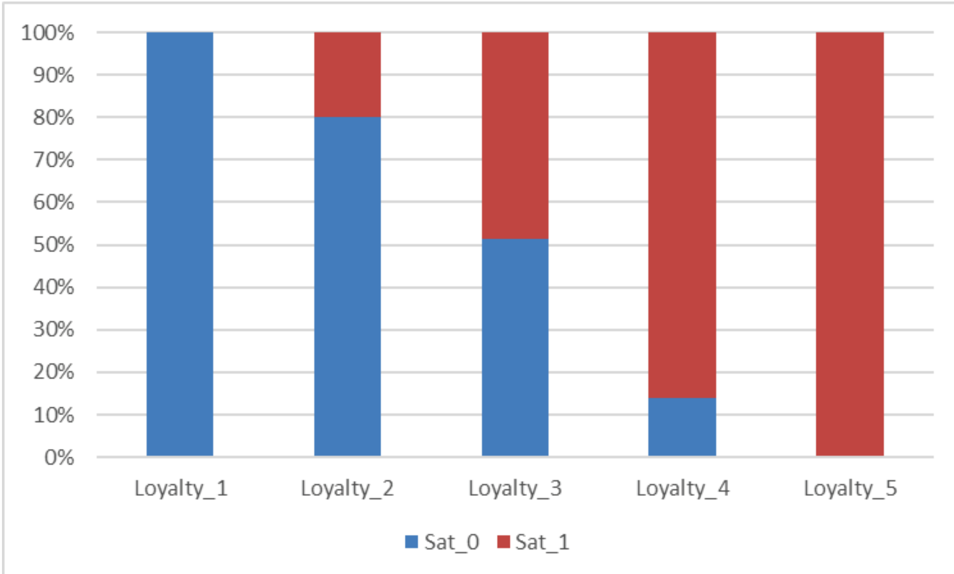
7. In order to visualize the row and column percentages with Excel's 100% Stacked Column Chart, we will need to modify the numeric row and column labels as shown, converting to text as shown:

Observed Counts	Sat_0	Sat_1
Loyalty_1	2	0
Loyalty_2	8	2
Loyalty_3	17	16
Loyalty_4	6	37
Loyalty_5	0	12

8. Select cells **A3:C8** of the **Chi-Square** sheet. Click Excel's **Insert > Insert Column or Bar Chart** and select **100% Stacked Column** as shown.



9. Click to create the 100% stacked column chart (uncheck the Chart Title):



This clearly shows the strong relationship between the Loyalty scores and Discrete Satisfaction.

Chi-Square Test – Fisher's Exact

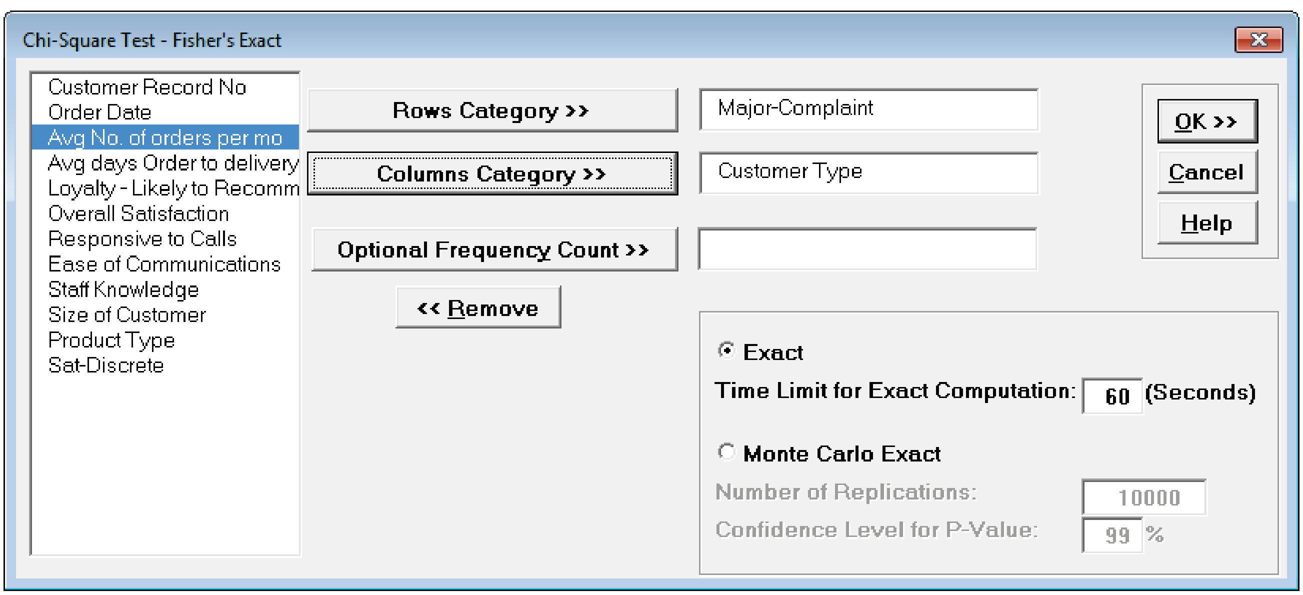
The **Chi-Square Test – Fisher's Exact** utilizes permutations and fast network algorithms to solve the Exact Fisher P-Value for contingency (two-way row*column) tables. This is an extension of the Fisher Exact option provided in the Two Proportion Test template. For data that requires more computation time than specified, Monte Carlo P-Values provide an approximate (but unbiased) P-Value that typically matches exact to two decimal places using 10,000 replications. One million replications give a P-Value that is typically accurate to three decimal places. A confidence interval (99% default) is given for the Monte Carlo P-Values. For further details refer to the Appendix **Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests**.

It is important to note that while exact P-Values are "correct," they do not increase (or decrease) the power of a small sample test, so they are not a solution to the problem of failure to detect a change due to inadequate sample size!

1. We will now re-analyze the above Major Complaint by Customer Type data. Open **Customer Data.xlsx**. Click **Sheet 1** tab. The discrete data of interest is Complaints and Customer Type, i.e., does the type of complaint differ across customer type? Formally the Null Hypothesis is that there is no relationship (or independence) between Customer Type and Complaints.

2. Click **SigmaXL > Statistical Tools > Chi-Square Tests – Exact > Chi-Square Test – Fisher's Exact**. Ensure that the entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.

3. Select *Major-Complaint*, click **Rows Category >>**; select *Customer Type*, click **Column Category >>**. Select **Exact** with the default **Time Limit for Exact Computation** = 60 seconds.



Tip: If the exact computation time limit is exceeded a dialog will prompt you to use Monte Carlo or to increase the computation time. When this occurs, Monte Carlo is recommended.

4. Click **OK**. Results:

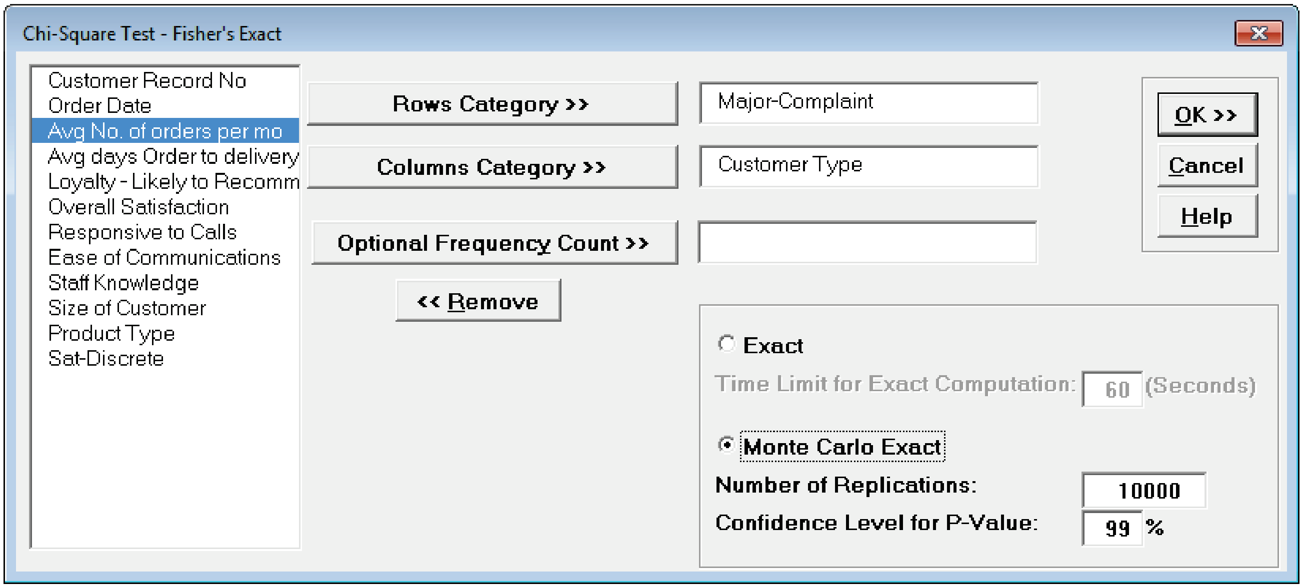
Chi-Square - Fisher's Exact			
Major-Complaint - Customer Type			
Observed Counts	**1**	**2**	**3**
Difficult-to-order	5	9	5
Not-available	2	0	2
Order-takes-too-long	1	3	6
Return-calls	19	28	13
Wrong-color	4	2	1
Expected Counts	**1**	**2**	**3**
Difficult-to-order	5.890	7.980	5.130
Not-available	1.240	1.680	1.08
Order-takes-too-long	3.100	4.200	2.700
Return-calls	18.600	25.200	16.200
Wrong-color	2.170	2.940	1.890
Std. Residuals	**1**	**2**	**3**
Difficult-to-order	-0.366718	0.361076	-0.057396402
Not-available	0.682500	-1.296	0.885270
Order-takes-too-long	-1.193	-0.585540	2.008
Return-calls	0.092747779	0.557773	-0.795046
Wrong-color	1.242	-0.548219	-0.647380
Chi-Square	12.211		
DF	8		
Chi-Square P-Value	0.1420		
Fisher's Exact P-Value	0.1469		
Note: 9 out of 15 cells have expected counts less than 5.			

With Fisher's Exact P-Value = 0.1469 we fail to reject H0, so we cannot conclude that there is a difference in customer complaints across customer types. This is close to the approximate Chi-Square P-Value of 0.142, so either P-Value results in the same conclusion.

However, there are cases where the "large sample" Chi-Square P-Value leads to one conclusion but Fisher's Exact P-Value gives another. This will be demonstrated later using **Two Way Table Data – Fisher's Exact**.

The Exact P-Value was computed very quickly, but if the data set was larger, the required computation time could become excessive, and Monte Carlo would be required. We will rerun this analysis with Monte Carlo and discuss the output report.

5. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Select **Monte Carlo Exact** with the default **Number of Replications** = 10000 and **Confidence Level for P-Value** = 99%.



Tip: As discussed above 10,000 replications will result in a Monte Carlo P-Value that is correct to two decimal places. One million (1e6) replications will result in three decimal places of accuracy and typically require less than 60 seconds to solve for any data set.

Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

6. Click **OK**. Results:

Chi-Square Test - Fisher's Monte-Carlo			
Major-Complaint - Customer Type			
Observed Counts	**1**	**2**	**3**
Difficult-to-order	5	9	5
Not-available	2	0	2
Order-takes-too-long	1	3	6
Return-calls	19	28	13
Wrong-color	4	2	1
Expected Counts	**1**	**2**	**3**
Difficult-to-order	5.890	7.980	5.130
Not-available	1.240	1.680	1.08
Order-takes-too-long	3.100	4.200	2.700
Return-calls	18.600	25.200	16.200
Wrong-color	2.170	2.940	1.890
Std. Residuals	**1**	**2**	**3**
Difficult-to-order	-0.366718	0.361076	-0.057396402
Not-available	0.682500	-1.296	0.885270
Order-takes-too-long	-1.193	-0.585540	2.008
Return-calls	0.092747779	0.557773	-0.795046
Wrong-color	1.242	-0.548219	-0.647380
Chi-Square	12.211		
DF	8		
Chi-Square P-Value	0.1420		
Fisher's Monte-Carlo P-Value	0.1486		
Fisher's Monte-Carlo P-Value 99% CI Upper	0.1569		
Fisher's Monte-Carlo P-Value 99% CI Lower	0.1403		
Note: 9 out of 15 cells have expected counts less than 5.			

Fisher's Monte Carlo P-Value here is 0.1486 with a 99% confidence interval of 0.1403 to 0.1569. This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). The true Exact P-Value = 0.1469 lies within this confidence interval. If the confidence interval is too wide (e.g., upper "fail-to-reject" H0 and lower "reject" H0), simply rerun the Monte Carlo option using a larger number of replications (use 1e5 or 1e6).

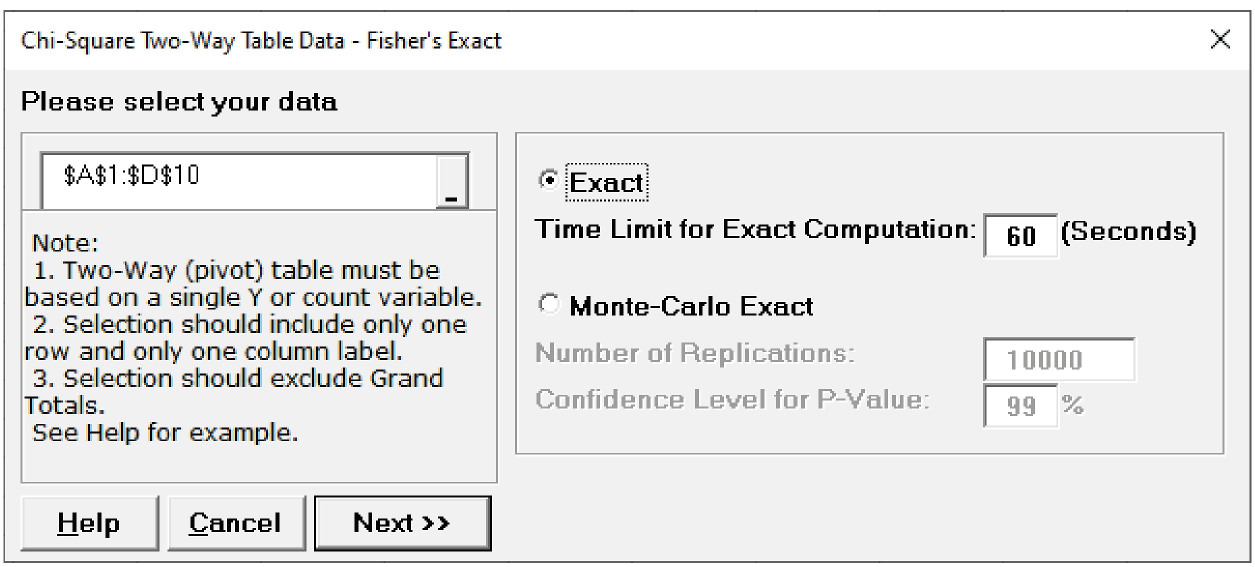
Chi-Square Test – Two Way Table Data – Fisher's Exact

Now we will consider a sparse data set where the Chi-Square approximation fails and Fisher's Exact is required to give a correct conclusion for the hypothesis test. This is adapted from a subset of dental health data (oral lesions) obtained from house to house surveys that were conducted in three geographic regions of rural India [1, 2]. The Fisher's Exact P-Value obtained with SigmaXL may be validated using these references. The data labels have been modified to a generic "A", "B", "C", etc. for the oral lesions location (rows) and "Region1", "Region2" and "Region3" for the geographic regions (columns).

1. Open the file **Oral_Lesions.xlsx.** This data is in Two –Way Contingency Table (or pivot table) format. Note that cells **A1:D10** have been pre-selected.

Site of Oral Lesion (Damage)	Region1	Region2	Region3
A	0	1	0
B	8	1	8
C	0	1	0
D	0	1	0
E	0	1	0
F	0	1	0
G	0	1	0
H	1	0	1
I	1	0	1

2. Click **SigmaXL > Statistical Tools > Chi-Square Tests – Exact > Chi-Square Test – Two-Way Table Data – Fisher's Exact**. Note the selection of data includes the Row and Column labels.



3. Click **Next**. Results:

Chi-Square 2 Way Table Statistics - Fisher's Exact

Observed Counts	Region1	Region2	Region3
A	0	1	0
B	8	1	8
C	0	1	0
D	0	1	0
E	0	1	0
F	0	1	0
G	0	1	0
H	1	0	1
I	1	0	1

Expected Counts	Region1	Region2	Region3
A	0.370370	0.259259	0.370370
B	6.296	4.407	6.296
C	0.370370	0.259259	0.370370
D	0.370370	0.259259	0.370370
E	0.370370	0.259259	0.370370
F	0.370370	0.259259	0.370370
G	0.370370	0.259259	0.370370
H	0.740741	0.518519	0.740741
I	0.740741	0.518519	0.740741

Std. Residuals	Region1	Region2	Region3
A	-0.608581	1.455	-0.608581
B	0.678971	-1.623	0.678971
C	-0.608581	1.455	-0.608581
D	-0.608581	1.455	-0.608581
E	-0.608581	1.455	-0.608581
F	-0.608581	1.455	-0.608581
G	-0.608581	1.455	-0.608581
H	0.301232	-0.720082	0.301232
I	0.301232	-0.720082	0.301232

Chi-Square	22.099
DF	16
Chi-Square P-Value	0.1400
Fisher's Exact P-Value	0.0101

Note: 25 out of 27 cells have expected counts less than 5.

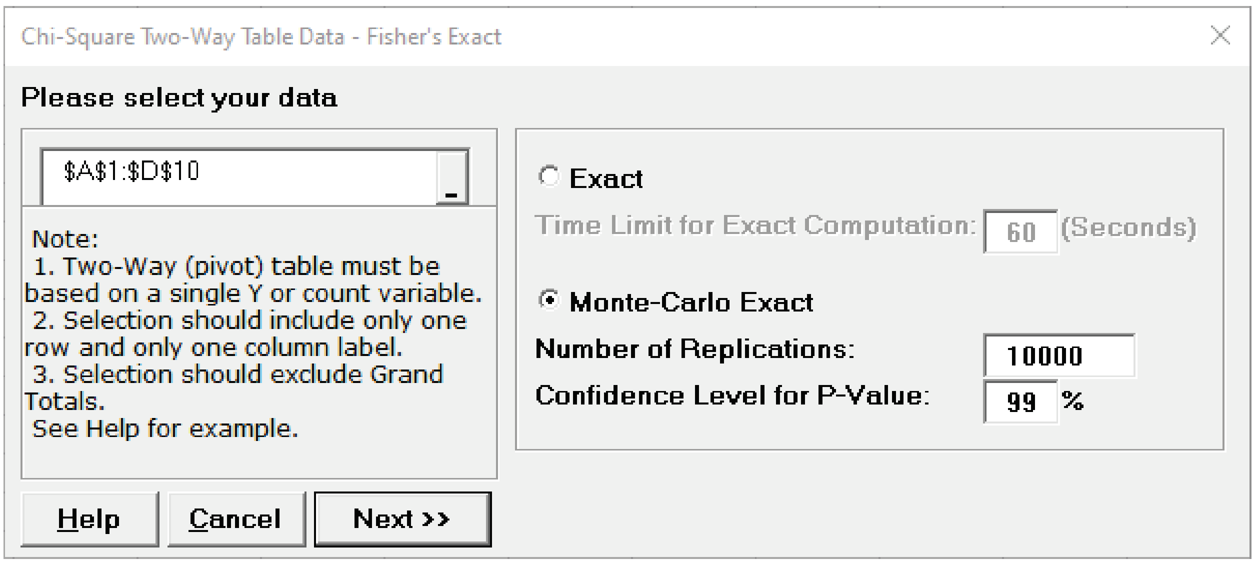
Note: 25 out of 27 cells have expected counts less than 5. Since more than 20% of the cells have expected counts less than 5 and several cells have an expected count less than 1, the Chi-Square P-Value = 0.14 is invalid and leads to an incorrect conclusion: Fail to Reject H0, Site of Oral Lesion and Geographic Region are independent. However, with Fisher's Exact P-Value =

0.0101, we have strong evidence to reject H0, and conclude that there is indeed a relationship between Site of Oral Lesion and Region.

While this example is originally from a dental health study, it could have been Defect Type versus Supplier and highlights that the use of Fisher's Exact for contingency tables with sparse data can make the difference between a good business decision and a bad business decision!

The Exact P-Value was computed very quickly, but if the data set was larger, the required computation time could become excessive, and Monte Carlo would be required. We will rerun this analysis with Monte Carlo and discuss the output report.

4. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Select **Monte Carlo Exact** with the default **Number of Replications** = 10000 and **Confidence Level for P-Value** = 99%.



Tip: As discussed above 10,000 replications will result in a Monte Carlo P-Value that is correct to two decimal places. One million (1e6) replications will result in three decimal places of accuracy and typically require less than 60 seconds to solve for any data set.

Tip: The Monte Carlo 99% confidence interval for P-Value is **not** the same as a confidence interval on the test statistic due to data sampling error. The confidence level for the hypothesis test statistic is still 95%, so **all reported P-Values less than .05 will be highlighted in red** to indicate significance. The 99% Monte Carlo P-Value confidence interval is due to the uncertainty in Monte Carlo sampling, and it becomes smaller as the number of replications increases (irrespective of the data sample size). The Exact P-Value will lie within the stated Monte Carlo confidence interval 99% of the time.

5. Click **Next**. Results:

Chi-Square 2 Way Table Statistics - Fisher's Monte Carlo

Observed Counts	Region1	Region2	Region3
A	0	1	0
B	8	1	8
C	0	1	0
D	0	1	0
E	0	1	0
F	0	1	0
G	0	1	0
H	1	0	1
I	1	0	1

Expected Counts	Region1	Region2	Region3
A	0.370370	0.259259	0.370370
B	6.296	4.407	6.296
C	0.370370	0.259259	0.370370
D	0.370370	0.259259	0.370370
E	0.370370	0.259259	0.370370
F	0.370370	0.259259	0.370370
G	0.370370	0.259259	0.370370
H	0.740741	0.518519	0.740741
I	0.740741	0.518519	0.740741

Std. Residuals	Region1	Region2	Region3
A	-0.608581	1.455	-0.608581
B	0.678971	-1.623	0.678971
C	-0.608581	1.455	-0.608581
D	-0.608581	1.455	-0.608581
E	-0.608581	1.455	-0.608581
F	-0.608581	1.455	-0.608581
G	-0.608581	1.455	-0.608581
H	0.301232	-0.720082	0.301232
I	0.301232	-0.720082	0.301232

Chi-Square	22.099
DF	16
Chi-Square P-Value	0.1400
Fisher's Monte Carlo P-Value	**0.0094**
Fisher's Monte Carlo P-Value 99% CI Upper	**0.0116**
Fisher's Monte Carlo P-Value 99% CI Lower	**0.0072**

Fisher’s Monte Carlo P-Value here is 0.0094 with a 99% confidence interval of 0.0072 to 0.0116. This will be slightly different every time it is run (the Monte Carlo seed value is derived from the system clock). The true Exact P-Value = 0.0101 lies within this confidence interval. If the confidence interval is too wide (e.g., upper is a "fail-to-reject" H0 and lower is a "reject" H0), simply rerun the Monte Carlo option using a larger number of replications (use 1e5 or 1e6).

References for Fisher's Exact

[1] Mehta, C. R.; Patel, N. R. (1997) "Exact inference in categorical data," unpublished preprint, https://www.cytel.com/hs-fs/hub/1670/file-2413766266-pdf/Pdf/Exact-Inference---MEHTA-PATEL-Exact-Inference-for-Categorical-Data-HARVARD-UNIVERSITY-AND-CYTEL-SOFTWARE-CORPORATION-1997.pdf. See Table 7 (p. 33) for validation of Fisher's Exact P-Value.

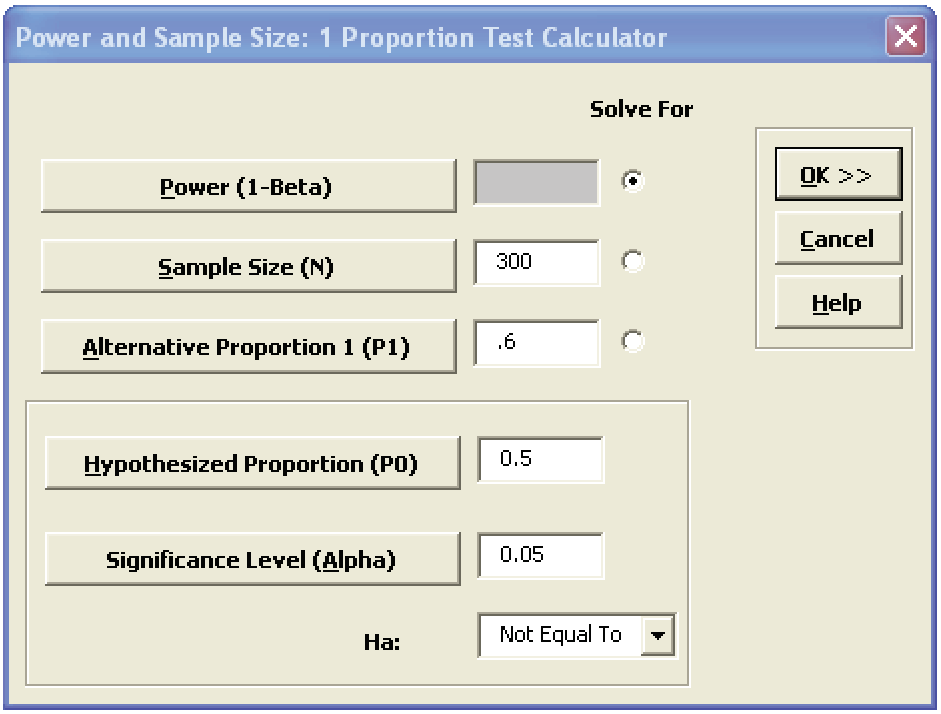
[2] Mehta, C.R. ; Patel, N.R. (1998). "Exact Inference for Categorical Data." In P. Armitage and T. Colton, eds., *Encyclopedia of Biostatistics*, Chichester: John Wiley, pp. 1411–1422.

Power & Sample Size for One Proportion Test

To determine Power & Sample Size for a 1 Proportion Test, you can use the Power & Sample Size Calculator or Power & Sample Size with Worksheet.

1. Click **SigmaXL > Statistical Tools > Power & Sample Size Calculators > 1 Proportion Test Calculator**.

2. Select **Solve For Power (1 – Beta)**. Enter **Sample Size** and **Alternative Proportion** as shown:



Note that we are calculating the power or likelihood of detection given that the hypothesized proportion is 0.5, but the alternative proportion is 0.6, sample size = 300, significance level = .05, and Ha: Not Equal To (two-sided test).

3. Click **OK**. The resulting report is displayed:

Power and Sample Size: 1 Proportion Test

H0: P0 = 0.5
Ha: P0 ≠ 0.5
Solve For: Power (1 - Beta)

Sample Size (N)	Alternative Proportion (P1)	Hypothesized Proportion (P0)	Significance Level (Alpha)	Power (1 - Beta)
300	0.6	0.5	0.05	0.937627018

A power value of 0.94 is acceptable, but note that the sample size n = 300, and the difference in proportion value is 0.1 or 10%! The sample size requirements for discrete data are much higher than those for continuous data.

4. To determine Power & Sample Size using a Worksheet, click **SigmaXL > Statistical Tools > Power & Sample Size with Worksheets > 1 Proportion Test**.

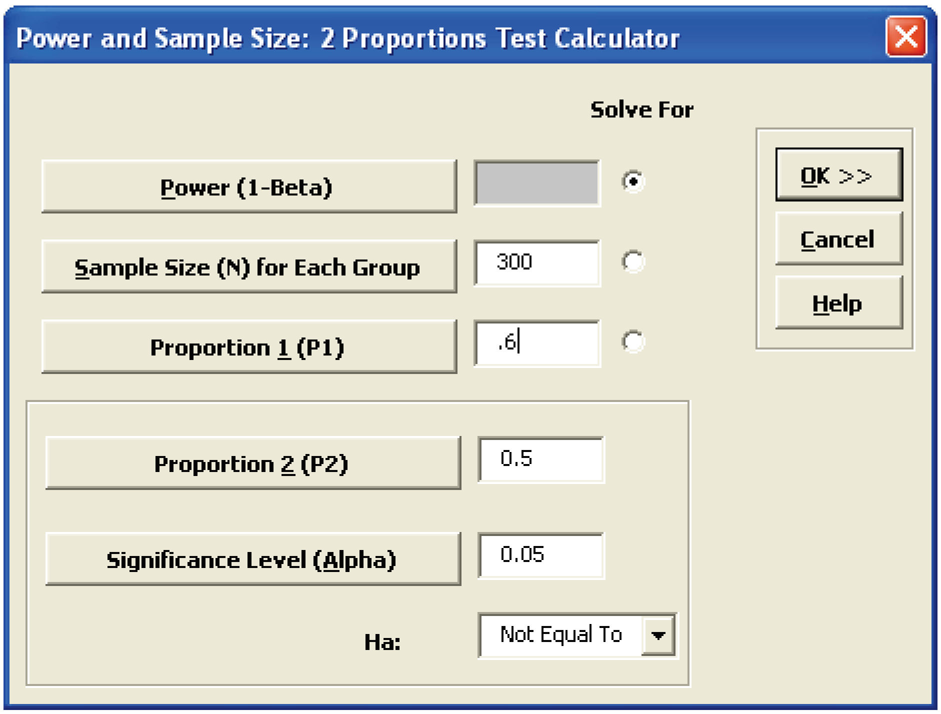
5. A graph showing the relationship between Power, Sample Size and Proportion Value can then be created using **SigmaXL > Statistical Tools > Power & Sample Size Chart**. See Part E for an example using the 1 Sample t-Test.

Power & Sample Size for Two Proportions Test

To determine Power & Sample Size for a 2 Proportions Test, you can use the Power & Sample Size Calculator or Power & Sample Size with Worksheet.

1. Click **SigmaXL > Statistical Tools > Power & Sample Size Calculators > 2 Proportions Test Calculator**.

2. Select **Solve For Power (1 – Beta)**. Enter **Sample Size** and **Proportion 1** values as shown:



Note that we are calculating the power or likelihood of detection given that P1 = 0.5 and P2 = 0.6, sample size for each group = 300, significance level = .05, and Ha: Not Equal To (two-sided test).

3. Click **OK**. The resulting report is displayed:

Power and Sample Size: 2 Proportions Test				
H0: P1 = P2				
Ha: P1 ≠ P2				
Solve For: Power (1 - Beta)				
Sample Size (N)	Proportion 1 (P1)	Proportion 2 (P2)	Significance Level (Alpha)	Power (1 - Beta)
300	0.6	0.5	0.05	0.693021232

A power value of 0.69 is unacceptable. Note that this value is much less than the power for the one proportion test (0.94).

4. To compensate, we will double the sample size per group. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Change the sample size per group from 300 to 600. Note that the power value is now 0.94.

5. To determine Power & Sample Size using a Worksheet, click **SigmaXL > Statistical Tools > Power & Sample Size with Worksheets > 2 Proportions Test**.

6. A graph showing the relationship between Power, Sample Size and Proportion Values can then be created using **SigmaXL > Statistical Tools > Power & Sample Size Chart**. See Part E for an example using the 1 Sample t-Test.

Part P – Analysis of Means (ANOM) Charts

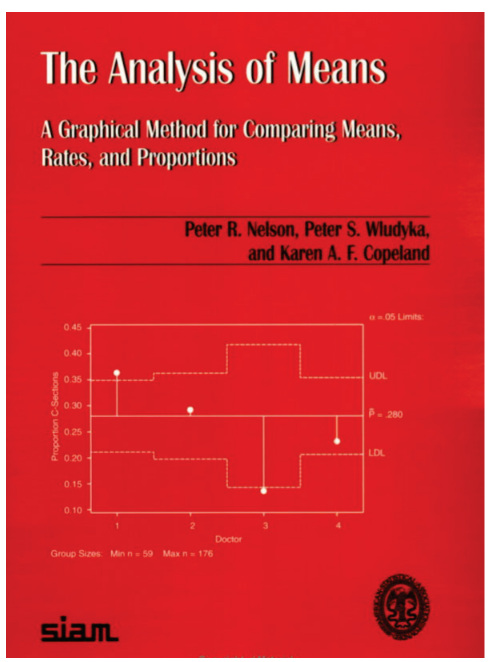
Introduction – What is ANOM

From http://asq.org/glossary/a.html:

> A statistical procedure for troubleshooting industrial processes and analyzing the results of experimental designs with factors at fixed levels. It provides a graphical display of data. Ellis R. Ott developed the procedure in 1967 because he observed that nonstatisticians had difficulty understanding analysis of variance. Analysis of means is easier for quality practitioners to use because it is (like) an extension of the control chart.

From the preface of the book, "The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions" by Peter R. Nelson, Peter S. Wludyka, and Karen A. F. Copeland:

> The goal of statistical data analysis is to use data to gain and communicate knowledge about processes and phenomena. Comparing means is often part of an analysis, for data arising in both experimental and observational studies.
>
> The analysis of means (ANOM) is an alternative procedure (to ANOVA) for comparing means.
>
> ANOM has the advantages of being much more intuitive and providing an easily understood graphical result, which clearly indicates any means that are different (from the overall mean) and allows for easy assessment of practical as well as statistical significance.



Formula for One-Way Balanced* Normal (from ANOM book 2.4)

$$\begin{aligned} \text{UDL} &= \overline{y}_{\bullet\bullet} + h(\alpha; I, N-I)\sqrt{\text{MS}_e}\sqrt{\tfrac{I-1}{N}} \\ \text{LDL} &= \overline{y}_{\bullet\bullet} - h(\alpha; I, N-I)\sqrt{\text{MS}_e}\sqrt{\tfrac{I-1}{N}} \end{aligned}$$

where:

$\bar{y}$.. = overall mean

h = critical value from multivariate *t* distribution – SigmaXL uses table exact critical values (Table B.1)

N = sample size

I = number of levels

$\sqrt{MSe}$ SQRT (Mean Square Error) = pooled standard deviation.

* Unbalanced uses critical values from studentized maximum modulus (SMM) distribution. SigmaXL uses table exact critical values (Table B.2). An adjustment is also made for varying sample size that results in varying decision limit values. See Appendix **Analysis of Means (ANOM) Charts**.

Formula for One-Way Balanced Binomial Proportions (from ANOM book 2.6)

$$\begin{aligned} \text{UDL} &= \overline{p} + h(\alpha; I, \infty)\sqrt{\overline{p}(1-\overline{p})}\sqrt{\tfrac{I-1}{N}} \\ \text{LDL} &= \overline{p} - h(\alpha; I, \infty)\sqrt{\overline{p}(1-\overline{p})}\sqrt{\tfrac{I-1}{N}} \end{aligned}$$

This is based on the normal approximation to the Binomial, so requires sample sizes large enough that np and n(p-1) > 5. SigmaXL automatically checks this and gives a warning message if the condition is not met.

Formula for One-Way Balanced Poisson Rates (from ANOM book 2.9)

$$\begin{aligned} \text{UDL} &= \bar{u} + h(\alpha; I, \infty)\sqrt{\bar{u}}\sqrt{\frac{I-1}{N}} \\ \text{LDL} &= \bar{u} - h(\alpha; I, \infty)\sqrt{\bar{u}}\sqrt{\frac{I-1}{N}} \end{aligned}$$

This is based on the normal approximation to the Poisson, so requires sample sizes large enough that nu > 5. SigmaXL automatically checks this and gives a warning message if the condition is not met.

Formula for Nonparametric Transformed Ranks (from ANOM book 9.3)

The values from the combined sample are ranked from smallest to largest (R_{ij}) and transformed to a z-score using the inverse normal:

$$E_{ij} = \Phi^{-1}[0.5 + R_{ij}/(2N + 1)]$$

and the above formula for balanced (or unbalanced) Normal is applied to compute decision limits.

Formula for Balanced Variances (from ANOM book 4.4)

$$\begin{aligned} \text{UDL} &= U(\alpha; I, n-1) I \text{MS}_e \\ \text{CL} &= \text{MS}_e \\ \text{LDL} &= L(\alpha; I, n-1) I \text{MS}_e \end{aligned}$$

where *U, L* are critical values from Table B.4 (*n* <= 35). For unbalanced and large sample size see Appendix **Analysis of Means (ANOM) Charts**.

Formula for Levene Robust Variances (from ANOM book 4.16)

The Levene Absolute Deviation from Median (ADM) is a simple transformation:

$$y_{ij} = \left| x_{ij} - \widetilde{x}_i \right|$$

and the above formula for balanced (or unbalanced) Normal is applied to compute decision limits.

Two-Way Normal

Main Effects for Two-Way ANOM are similar to One-Way but the mean square error (MSE) is derived from the ANOVA.

Slice Charts are a modified ANOM chart developed by Dr. Peter Wludyka that enables one to easily interpret the effects in the presence of an interaction (Wludyka 2013, 2015). The basic idea is to compare the levels of one factor for each level of the other factor with MSE derived from the Two-Way ANOVA. A yellow highlight automatically recommends Main Effects (if interaction is not significant, i.e., P-Value >= 0.1) or Slice Chart (if interaction is significant, P-Value < 0.1). The interaction P-Value is determined from ANOVA

There is an option to specify correction to alpha for multiple chart family-wise error rate:

Bonferroni alpha' = alpha/m; m = number of charts

Note that this is strictly to adjust for multiple charts. As seen in the formulas above, the individual charts also have correction built-in for the number of levels.

A Two-Way ANOM analysis will typically have balanced data from a designed experiment, with an equal number of observations for each combination level of X1 and X2. SigmaXL will allow unbalanced data, but we recommend that you use balanced data or slightly unbalanced data. ANOM charts use actual data means, not fitted (predicted least squares) means as used in Two-Way ANOVA Main Effects and Interaction Plots.

Two-Way Binomial Proportions and Poisson Rates

In collaboration, Peter Wludyka and John Noguera of SigmaXL extended the Slice Charts to Binomial and Poisson (Wludyka and Noguera 2016).

As with Normal, the basic idea is to compare the levels of one factor for each level of the other factor. MSE is derived from the whole model

A yellow highlight automatically recommends Main Effects (if interaction is not significant) or Slice Chart (if interaction is significant).

The interaction P-Value is automatically determined from Logistic regression for Binomial Proportions and Poisson regression for Poisson Rates.

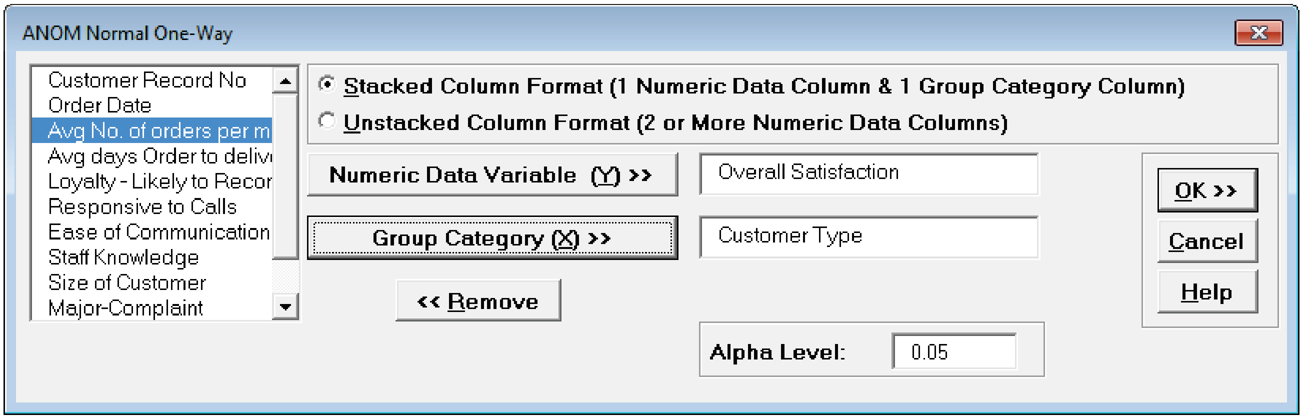
See Appendix **Analysis of Means (ANOM) Charts** for further formula details and references.

ANOM Normal One-Way

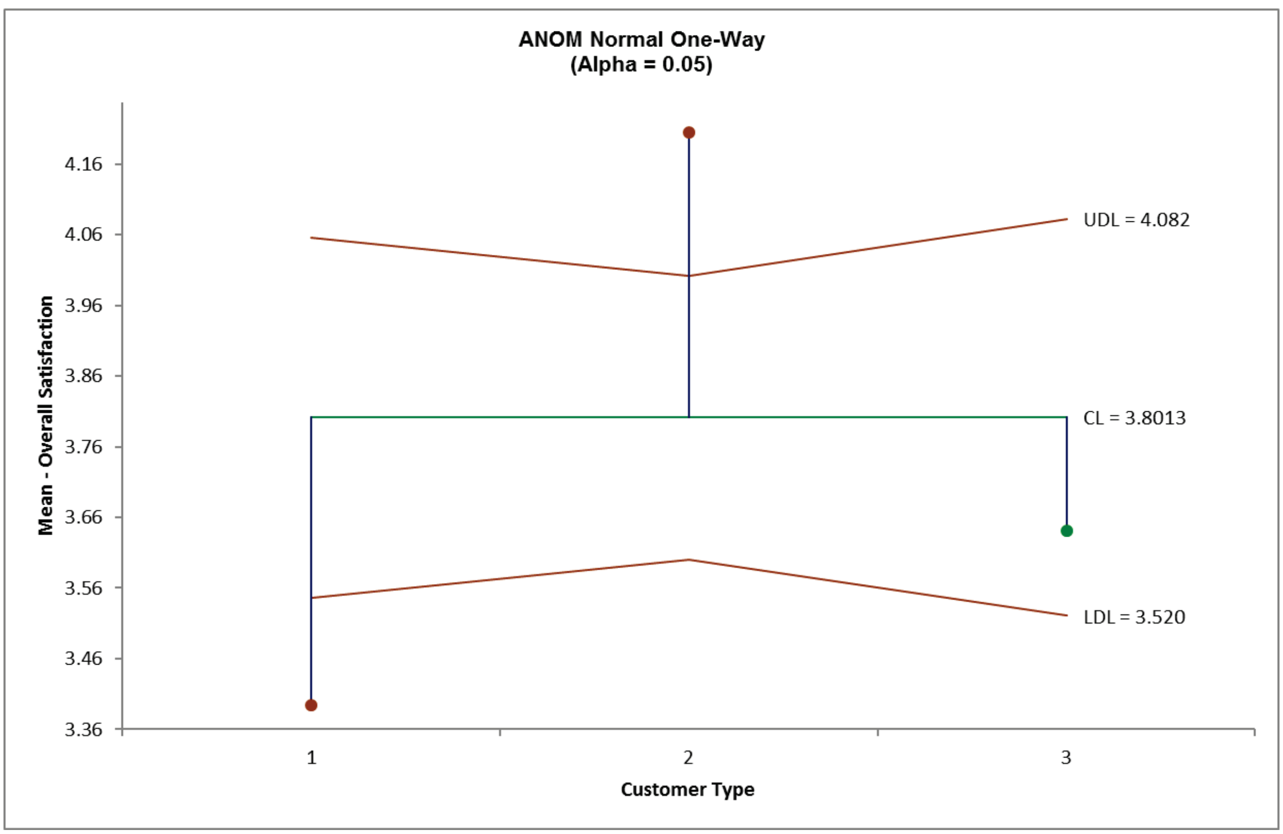
1. Open **Customer Data.xlsx**, click on **Sheet 1** tab.

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Normal One-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>**. Set **Alpha Level** = 0.05.



4. Click **OK**. The ANOM Normal One-Way chart is shown below:



5. Here we see that Customer Type 1 mean satisfaction score is significantly below the overall mean and Customer Type 2 is significantly higher. This is consistent with the results that were observed in the ANOVA analysis, but is easier to interpret.

6. The varying decision limits are due to the varying sample sizes for each Customer Type, with smaller sample size giving wider limits in a manner similar to a control chart. If the data are balanced the decision limit lines will be constant.

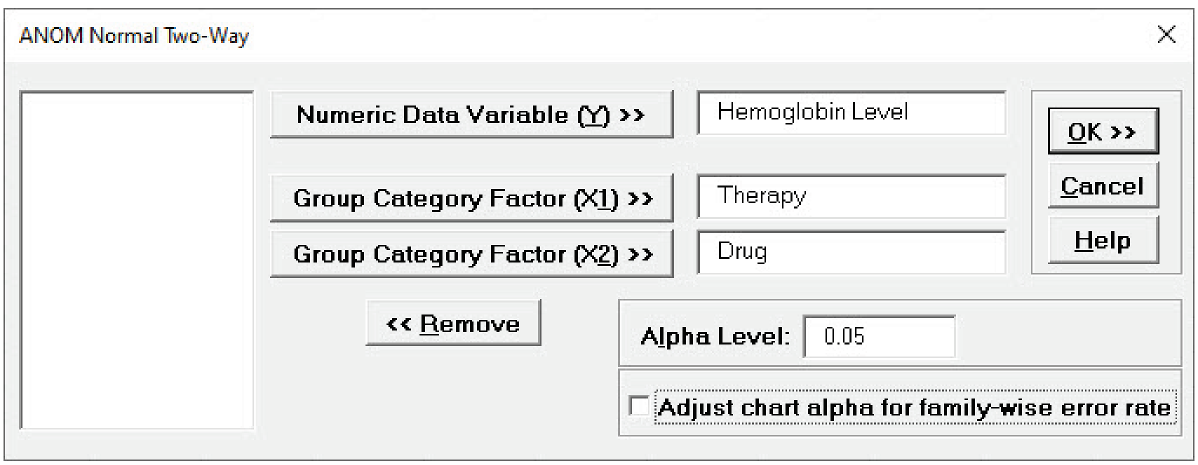
ANOM Normal Two-Way (with Main Effects and Slice Charts)

1. We will look at two balanced examples from the ANOM book (used with author permission), one with no interaction and another with a strong interaction, and then a slightly unbalanced version of the latter (Wludyka, 2013).

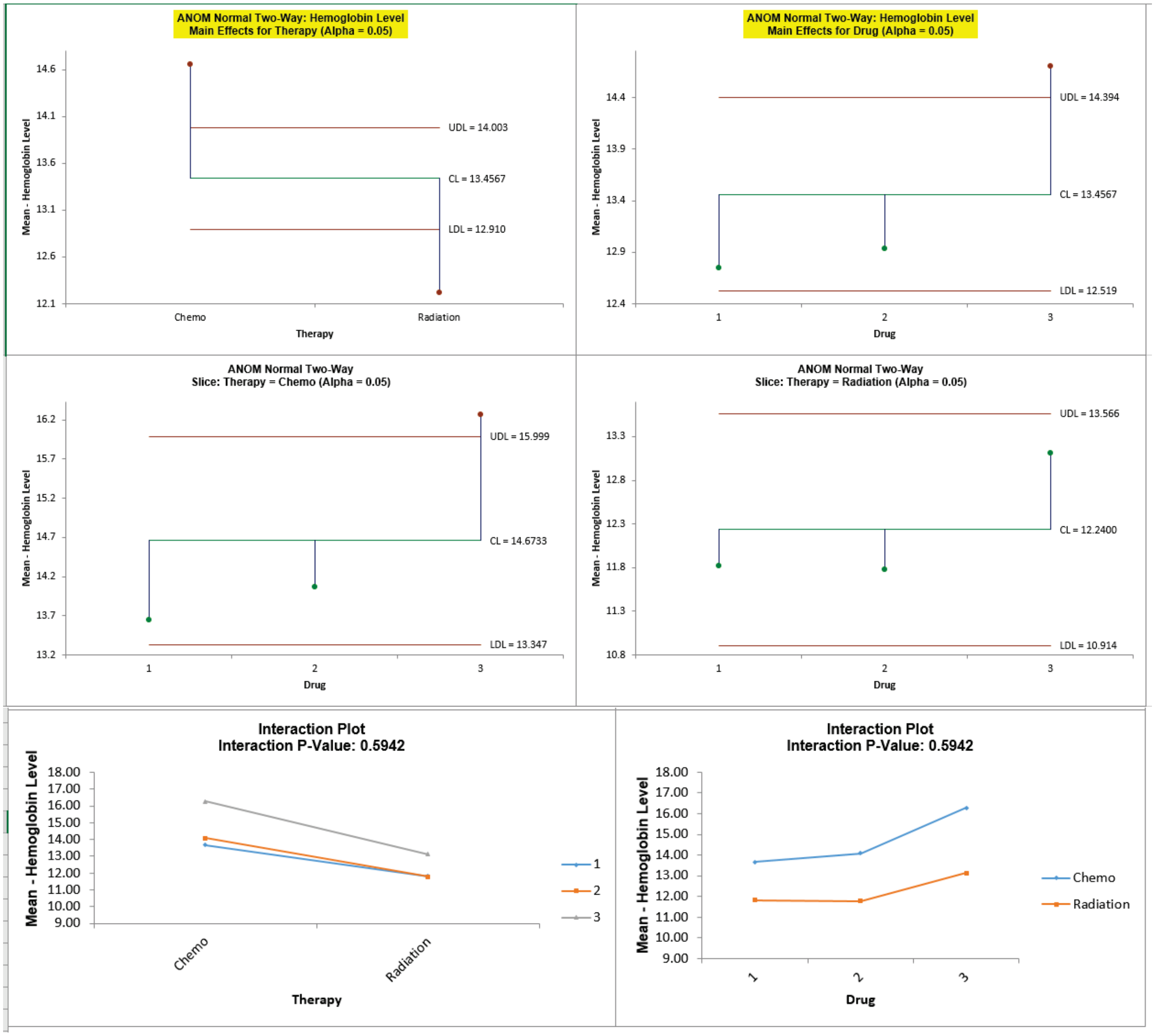
2. Open **ANOM Examples.xlsx**, click on **Hemoglobin – Normal** tab. This data is from a factorial design to study hemoglobin levels in males (Example 5.1). The factors are Therapy (Chemo or Radiation) and Drug (Type 1, 2, 3).

3. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Normal Two-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

4. Click **Next**. Select *Hemoglobin Level,* click **Numeric Data Variable (Y) >>**; select *Therapy*, click **Group Category Factor (X1) >>**; select *Drug*, click **Group Category Factor (X2) >>**. **Alpha Level** = 0.05, Adjust chart alpha for family-wise error rate is unchecked:



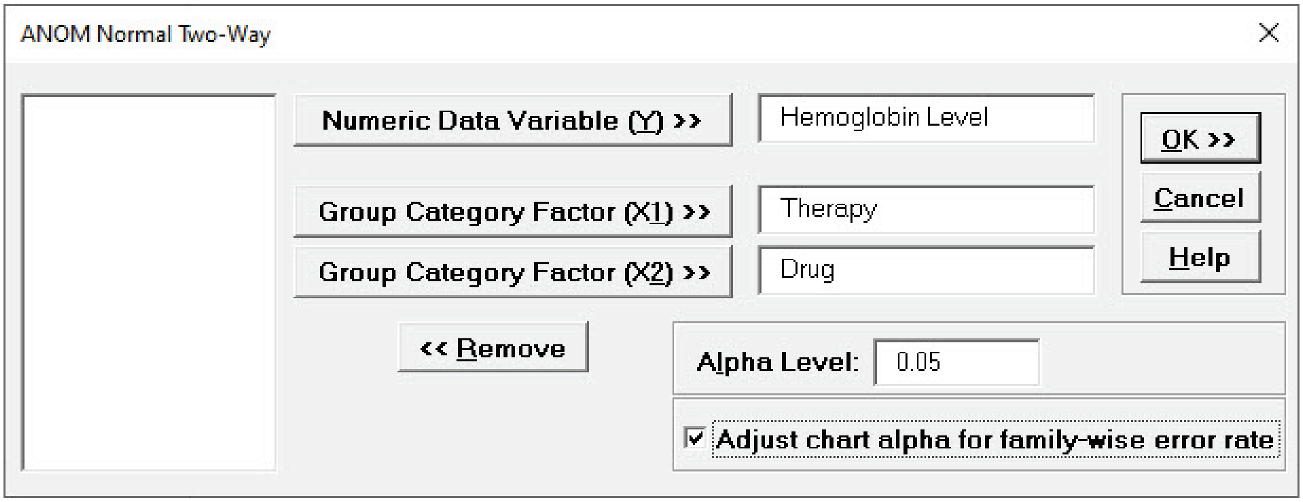
5. Click **OK**. The ANOM Normal Two-Way Main Effects, Slice charts and Interaction plots are shown below:



6. Since the Interaction is not significant (P-Value is >= 0.1), the Main Effects charts are highlighted. The decision limits match the manual calculations given in the ANOM book (pages 87, 88). Since the average hemoglobin level for radiation falls below the lower decision limit, and the average for chemotherapy falls above the upper decision limit, therapy has an (α = 0.05) effect. There is also an effect on hemoglobin level due to the drug since the average hemoglobin level for drug 3 is above the upper decision line.

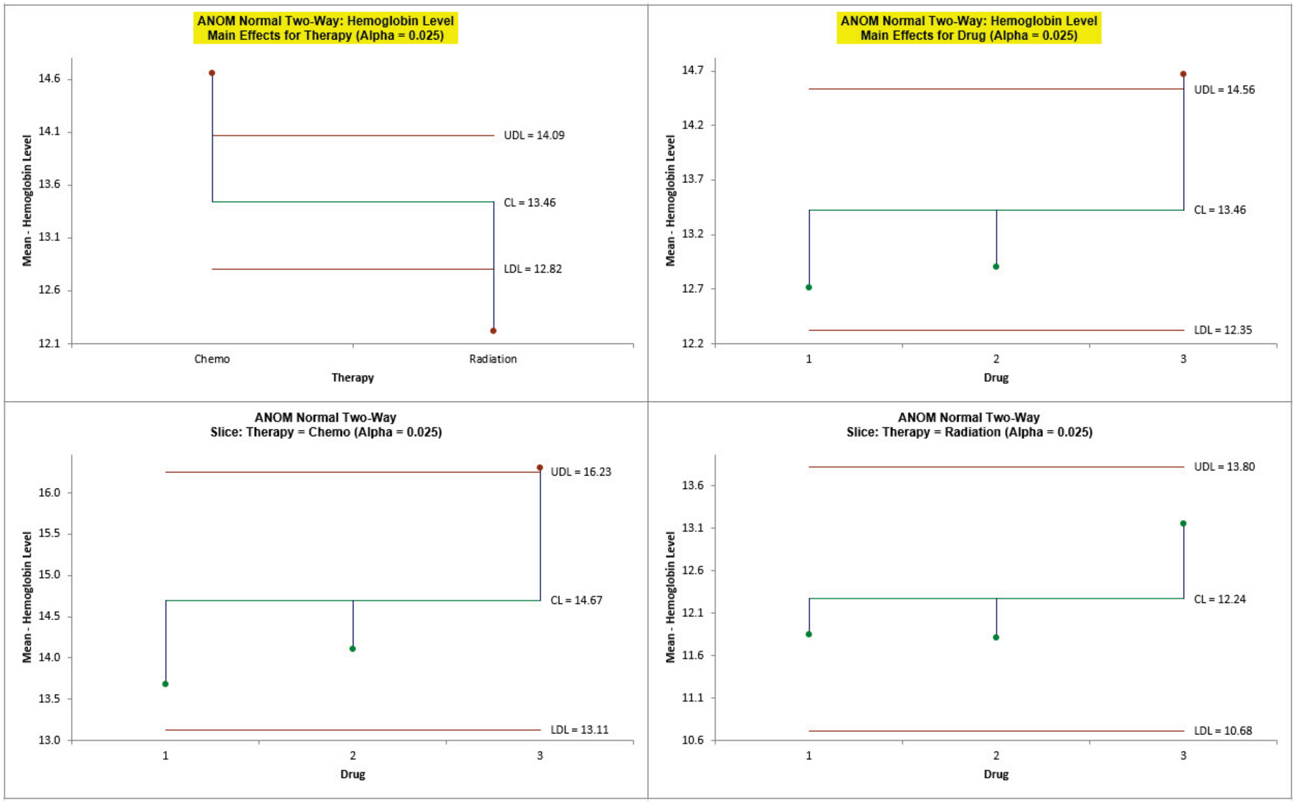
   Tip: Two-Way ANOM Normal is complimentary to Two-Way ANOVA. Analyzing the above data using Two-Way ANOVA shows that Therapy is significant with a P-Value = 0.0001, Drug is also significant with a P-Value = 0.0105 and the Interaction is not significant, with a P-Value = 0.5942.

7. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Check **Adjust chart alpha for family-wise error rate**.



This will apply a Bonferroni correction to the specified alpha values to control the overall family-wise error rate for the charts. So, the Main Effects alpha will be = 0.05/2 = 0.025 and the Slice Charts alpha will be = 0.05/2 = 0.025 as well.

8. Click **OK**. The ANOM Normal Two-Way Main Effects and Slice charts are shown below:

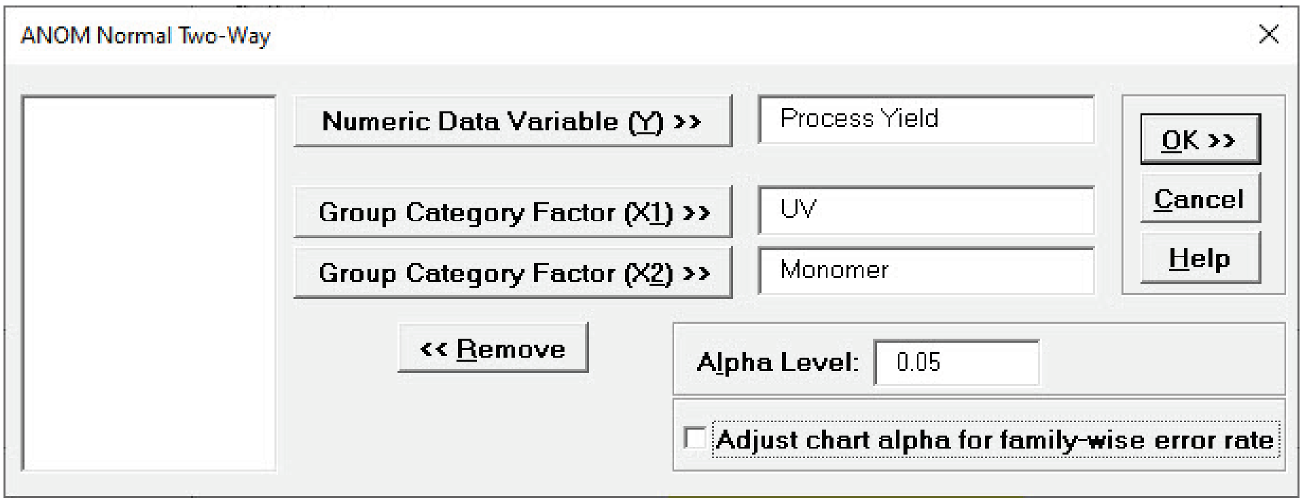


9. With the Bonferroni corrected alpha of 0.025, the significant results do not change, but they are closer to the decision lines. The trade-off for this chart family-wise adjustment is loss of power.

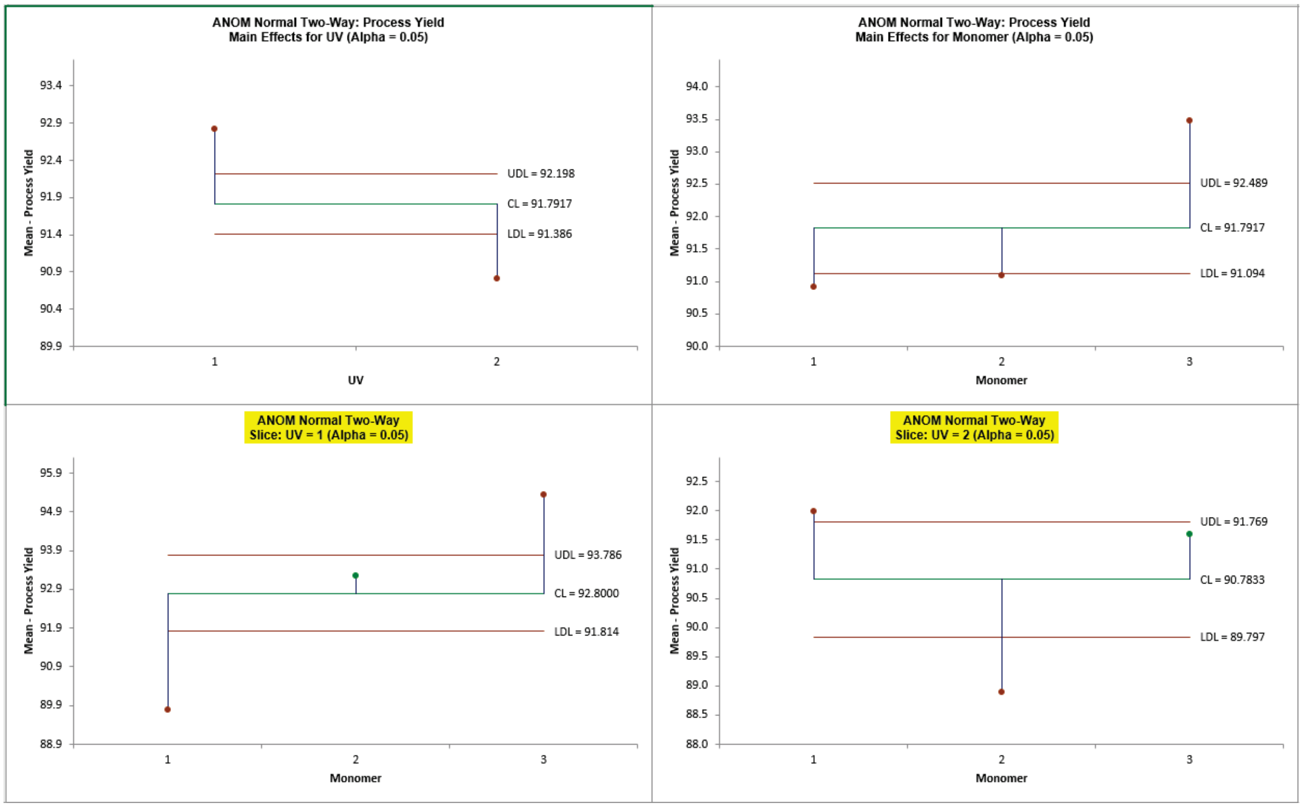
10. In **ANOM Examples.xlsx**, click on **Process Yield – Normal** tab. This data is from a factorial design to study the effect of three monomers and two levels of UV exposure (1 = UV; 2 = No UV) on the percent yield for contact lens production (Example 5.3).

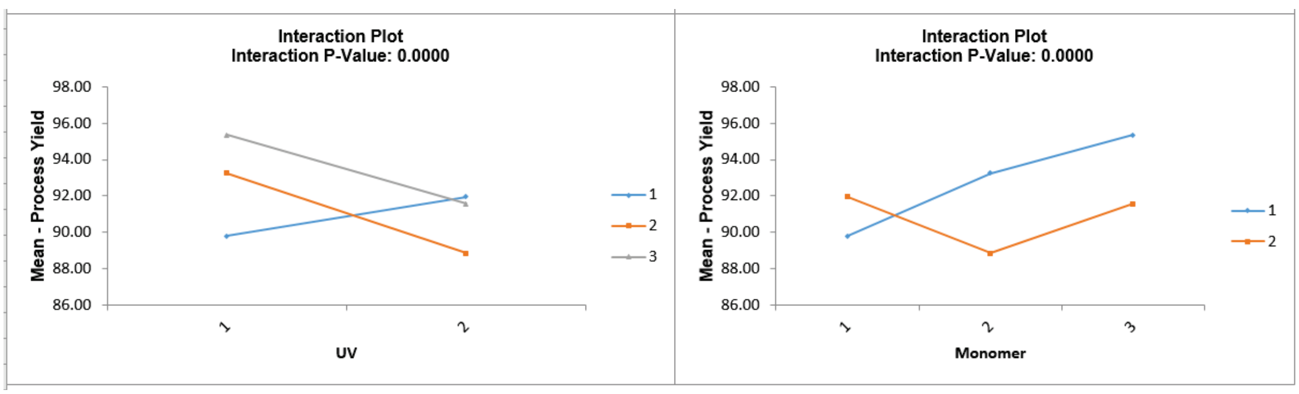
11. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Normal Two-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

12. Click **Next**. Select *Process Yield,* click **Numeric Data Variable (Y) >>**; select *UV*, click **Group Category Factor** (**X1) >>**; select *Monomer*, click **Group Category Factor** (**X2) >>. Alpha Level** = 0.05, **Adjust chart alpha for family-wise error rate** is unchecked:



13. Click **OK**. The ANOM Normal Two-Way Main Effects, Slice charts and Interaction plots are shown below:





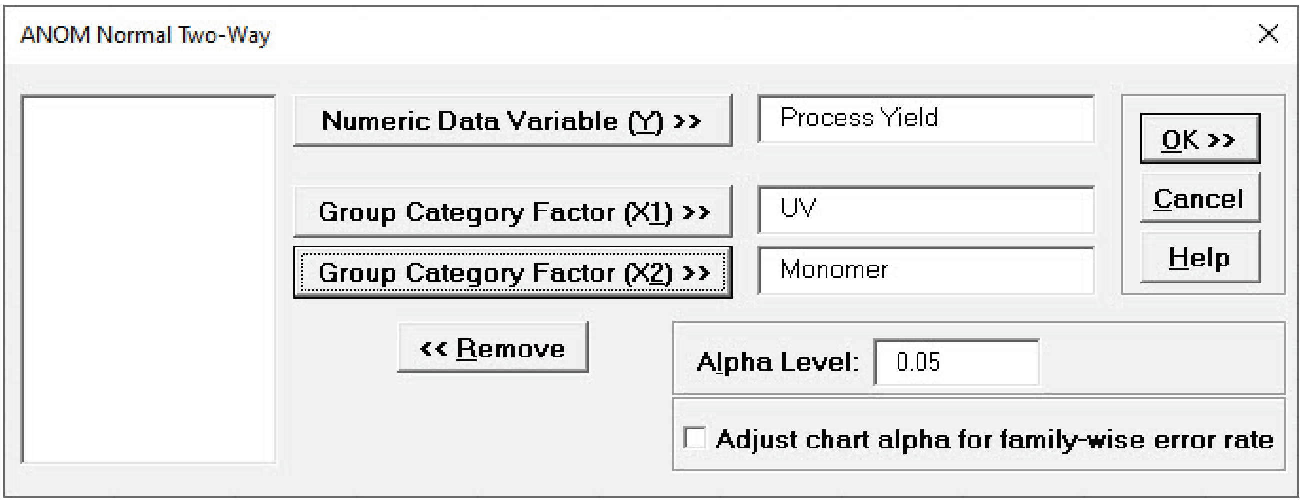
14. Since the Interaction P-Value is < 0.1, the Slice charts are highlighted. The decision limits for the Slice charts match the calculations given in Wludyka (2013). See Appendix **References for Analysis of Means (ANOM) Charts**. The Slice plot clearly shows the effect of the strong interaction. When UV = 1 (UV Exposure), the three monomers differ. In particular, the average for monomer 1 is below the overall average and the average for monomer 3 is above the overall average. However, for UV = 2 (no UV), the monomers also differ but in an entirely different manner than in the UV = 1 case, revealing the nature of the interaction.

    Tip: As previously noted, Two-Way ANOM Normal is complimentary to Two-Way ANOVA. Analyzing the above data using Two-Way ANOVA shows that UV is significant with a P-Value = 0.0001, Monomer is significant with a P-Value = 0.0000 (4.4 e-5) and the Interaction is also significant, with P-Value = 0.0000 (2.2e-6).

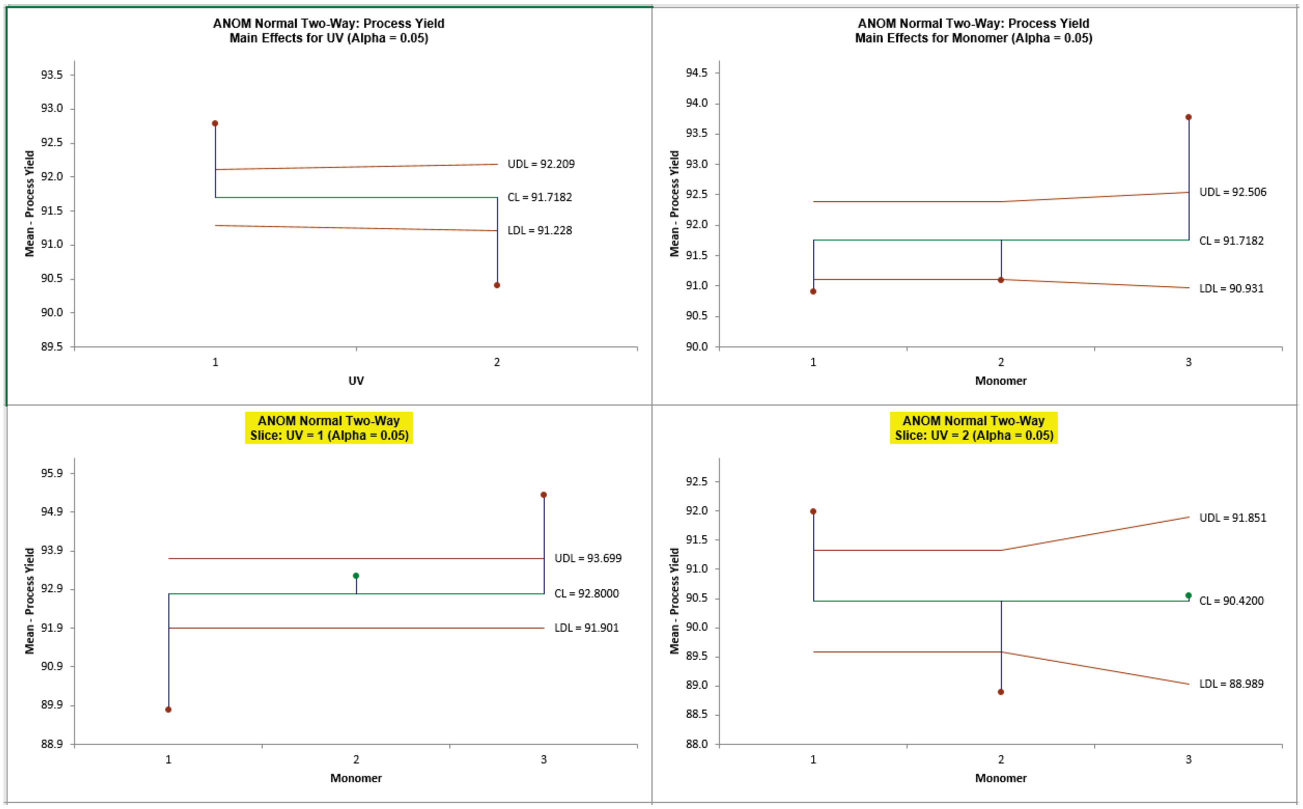
15. In **ANOM Examples.xlsx**, click on **Process Yield – Normal Unbal** tab. This data is from a factorial design to study the effect of three monomers and two levels of UV exposure (1 = UV; 2 = No UV) on the percent yield for contact lens production, but two of the observations are missing for A = 2 (no UV) and B = 3 (monomer 3).

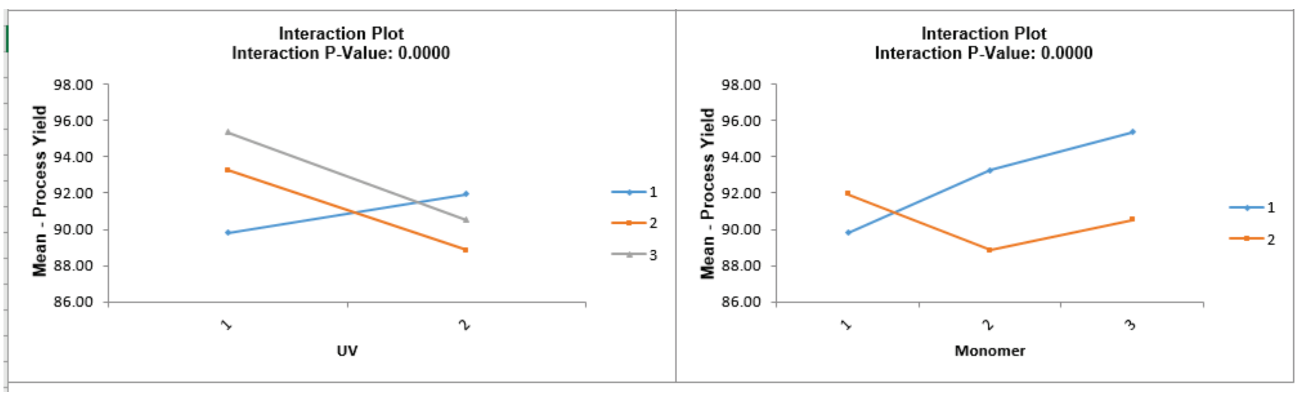
16. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Normal Two-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

17. Click **Next**. Select *Process Yield,* click **Numeric Data Variable (Y) >>**; select *UV*, click **Group Category Factor** (**X1) >>**; select *Monomer*, click **Group Category Factor** (**X2) >>. Alpha Level** = 0.05, **Adjust chart alpha for family-wise error rate** is unchecked:



18. Click **OK**. The ANOM Normal Two-Way Main Effects, Slice charts and Interaction plots are shown below:





19. The significant results of the experiment have not changed, but the decision limits are now varying due to the unbalanced data and use critical values for unbalanced data. The decision limits for the Slice charts match the calculations given in Wludyka (2013).

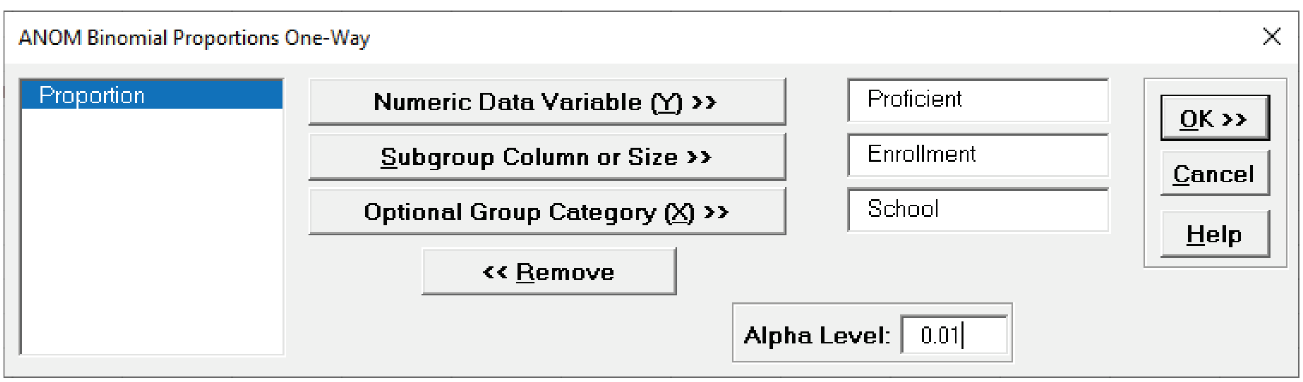
ANOM Binomial Proportions One-Way

1. Open **ANOM Examples.xlsx**, click on the **Test Scores – Binomial Prop** tab. This is standardized math test score data from 10 elementary schools (Example 3.4 from the ANOM book – used with author permission). We are testing to see if there is a difference between schools at an alpha = 0.01 level.

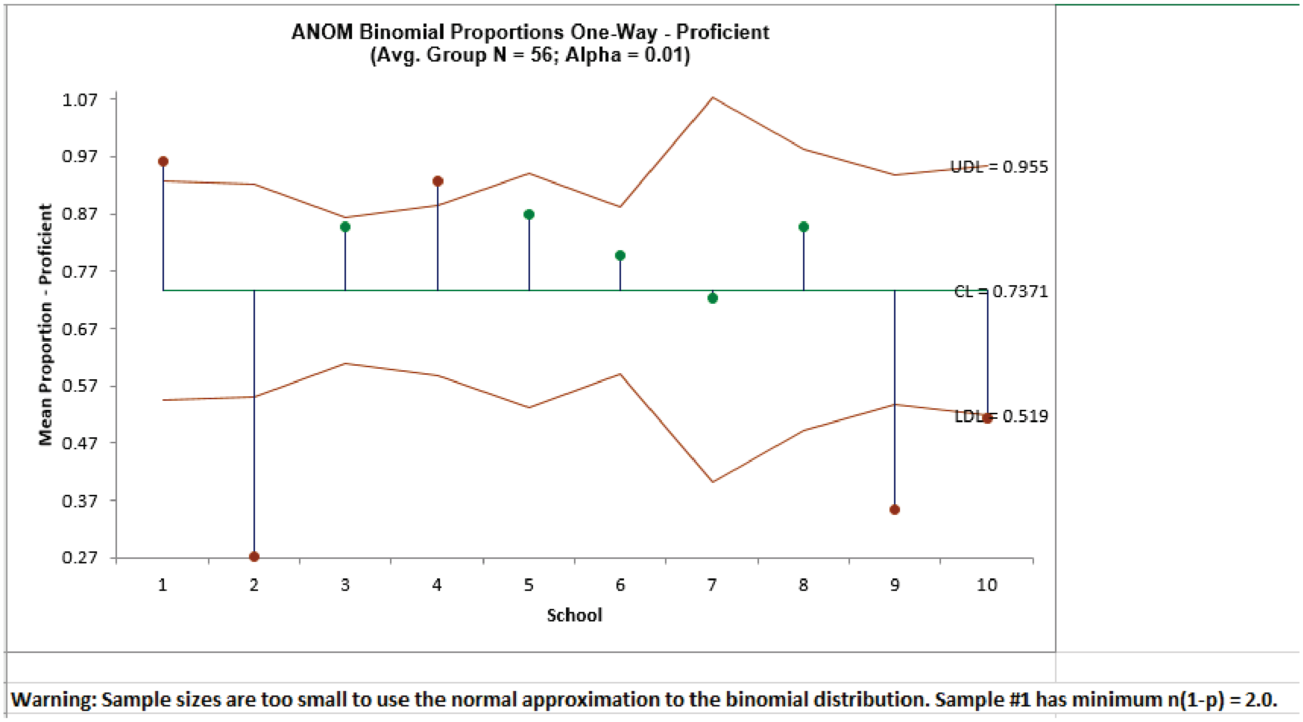
School	Proficient	Enrollment	Proportion
1	45	52	0.8654
2	15	55	0.2727
3	89	105	0.8476
4	75	81	0.9259
5	40	46	0.8696
6	67	84	0.7976
7	13	18	0.7222
8	28	33	0.8485
9	17	48	0.3542
10	21	41	0.5122

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Binomial Proportions One-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Select *Proficient,* click **Numeric Data Variable (Y) >>**; select *Enrollment*, click **Subgroup Column or Size >>**; select *School*, click **Optional Group Category (X) >>**. Set **Alpha Level** = 0.01:



4. Click **OK**. The ANOM Binomial Proportions One-Way chart is shown below:



5. The resulting ANOM decision chart shows that three schools are performing at significantly low levels and two schools are performing at significantly high levels.

6. SigmaXL automatically checks to see if the sample sizes are large enough for the normal approximation to the Binomial to be valid, i.e., np and n(p-1) are > 5. Here we see: **Warning: Sample sizes are too small to use the normal approximation to the binomial distribution. Sample #1 has minimum n(1-p) = 2.0.** Note that the warning does not show all occurrences, just the sample(s) with smallest failed np or n(1-p).

7. This does not mean that the chart results are invalid, an obvious "out" or "in" will not likely be affected, but the results should be used with caution, and if possible, more data collected.

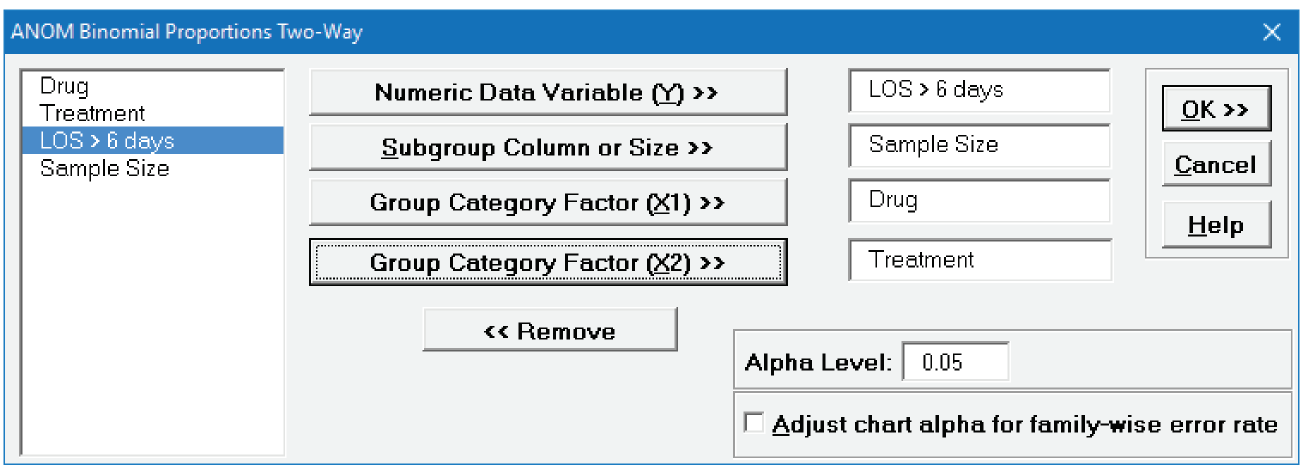
ANOM Binomial Proportions Two-Way

1. Open **ANOM Examples.xlsx**, click on the **Length of Stay - Binomial Prop** tab. This is hospital length of stay (LOS > 6 days) after bypass surgery. The data is from a factorial design with Factors: Drug (A, B) and Treatment (1, 2, 3). (Example 5.15 from the ANOM book – used with author permission).

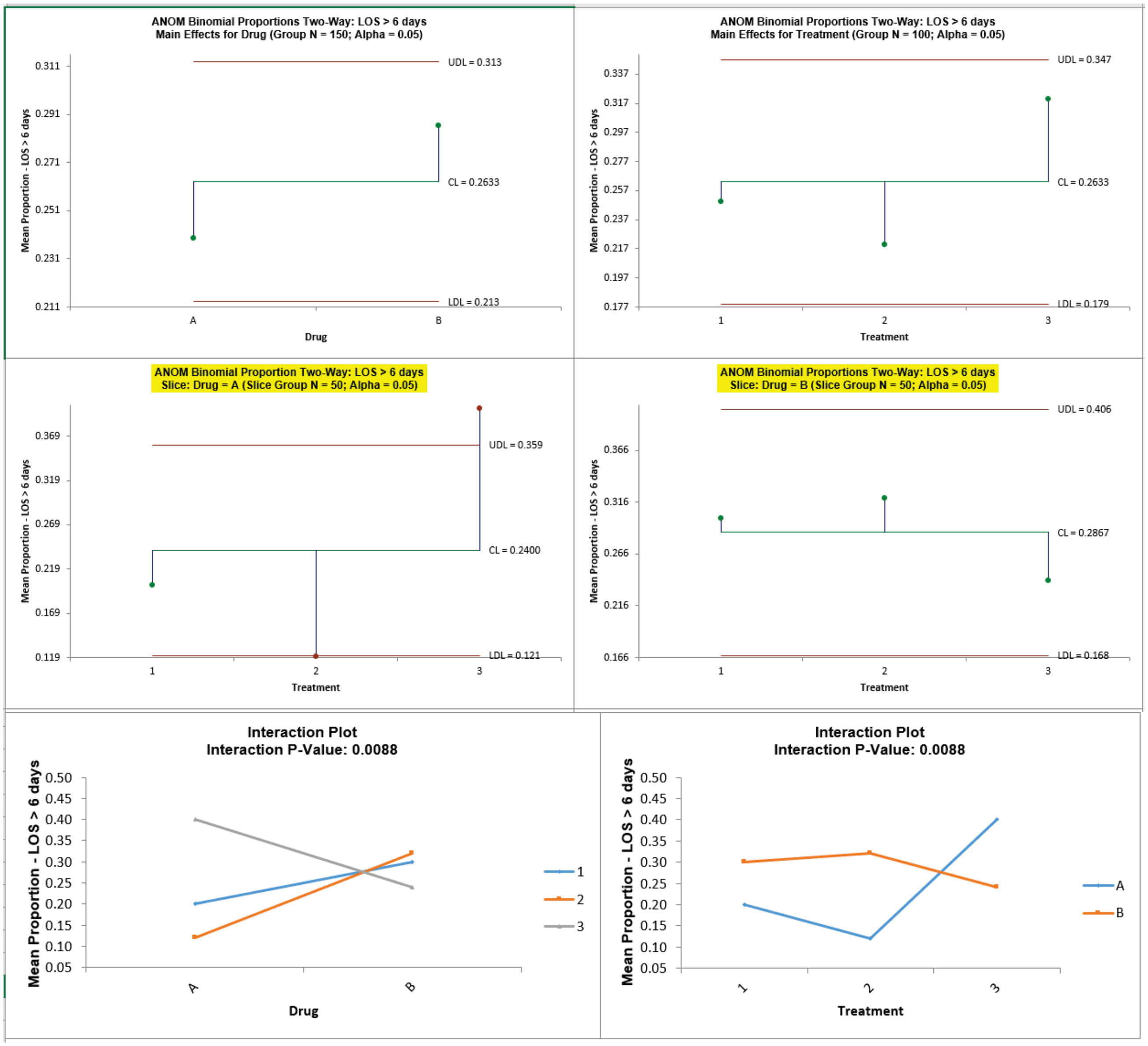
Drug	Treatment	LOS > 6 days	Sample Size
A	1	10	50
A	2	6	50
A	3	20	50
B	1	15	50
B	2	16	50
B	3	12	50

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Binomial Proportions Two-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Select *LOS > 6 days,* click **Numeric Data Variable (Y) >>**; select *Sample Size*, click **Subgroup Column or Size >>**; select *Drug*, click **Group Category Factor (X1) >>**; select *Treatment*, click **Group Category Factor (X2) >>**. **Alpha Level** = 0.05, **Adjust chart alpha for family-wise error rate** is unchecked:



4. Click **OK**. The ANOM Binomial Proportions Two-Way Main Effects, Slice charts and Interaction plots are shown below:



5. Since the Interaction P-Value is < 0.1 (automatically determined from Logistic regression), the Slice charts are highlighted. The Slice plot clearly shows the effect of the strong interaction. When Drug = A, Treatment 2 results in a significantly lower mean proportion LOS and Treatment 3 results in a significantly higher LOS. When Drug = B, the Treatment does not have a significant effect on LOS.

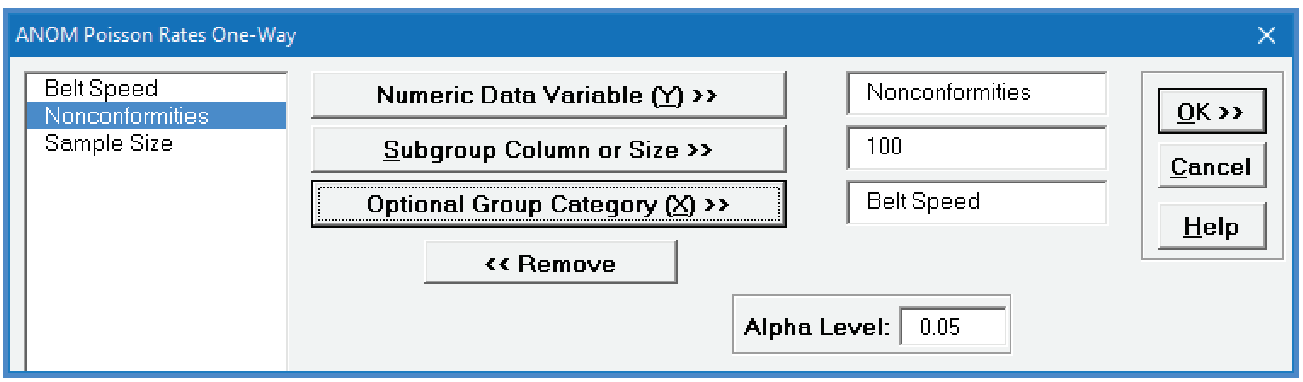
ANOM Poisson Rates One-Way

1. Open **ANOM Examples.xlsx**, click on the **Injection Molding-Poisson Rate** tab. This is injection molding data for contact lenses. It is from a study on the extent to which belt speed (cure time) influences the incidence of nonconformities (such as bubbles and tears) in the lenses. Since there can be more than one nonconformity per unit, Poisson ANOM is appropriate, similar to the U Control Chart (Example 2.16 from the ANOM book – used with author permission).

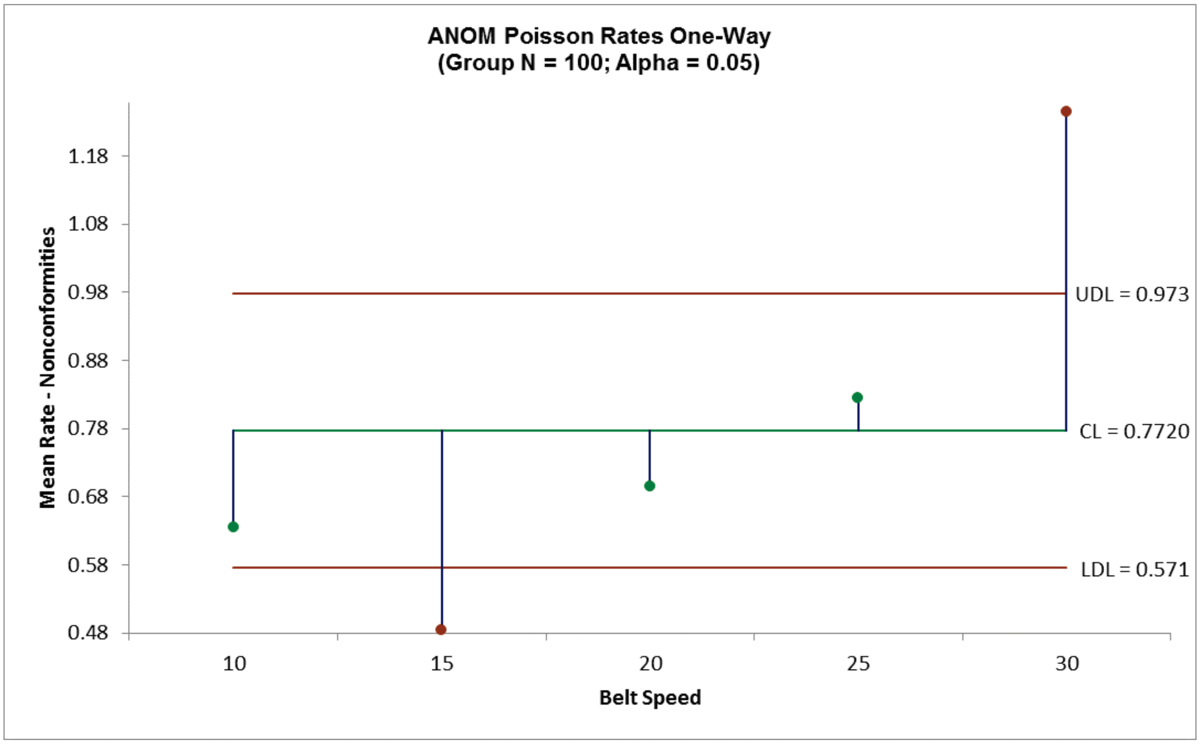
Belt Speed	Nonconformities	Sample Size
10	63	100
15	48	100
20	69	100
25	82	100
30	124	100

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Poisson Rates One-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Select *Nonconformities*, click **Numeric Data Variable (Y) >>**; enter 100 (or select *Sample Size*) for **Subgroup Column or Size >>**; select *Belt Speed*, click **Optional Group Category (X) >>**. Set **Alpha Level** = 0.05:



4. Click **OK**. The ANOM Poisson Rates One-Way chart is shown below:



5. From the ANOM chart, one sees that a belt speed of 15 results in a significantly low number of nonconformities and a belt speed of 30 results in a significantly high number of nonconformities.

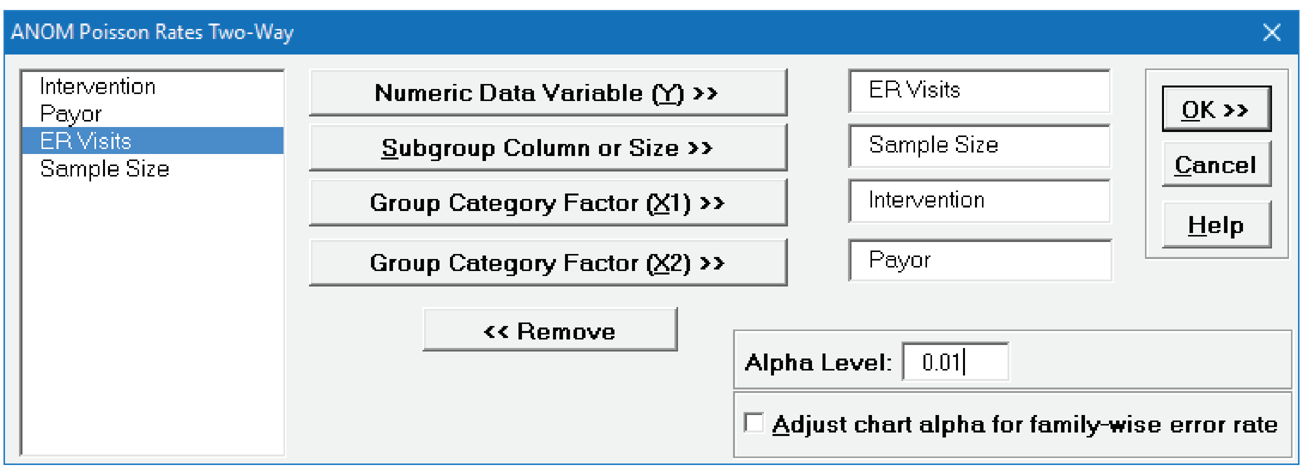
ANOM Poisson Rates Two-Way

1. Open **ANOM Examples.xlsx**, click on the **ER Visits – Poisson Rate** tab. This is Emergency Room Visits Data. It is a factorial design study to investigate the effect of nurse intervention (1 = home nurse assigned, 2 = home nurse not assigned) and payor groups (1, 2, 3) on the ER utilization rates for patients with COPD (a serious lung disease) with an alpha level = 0.01 (Example 5.16 from the ANOM book – used with author permission).

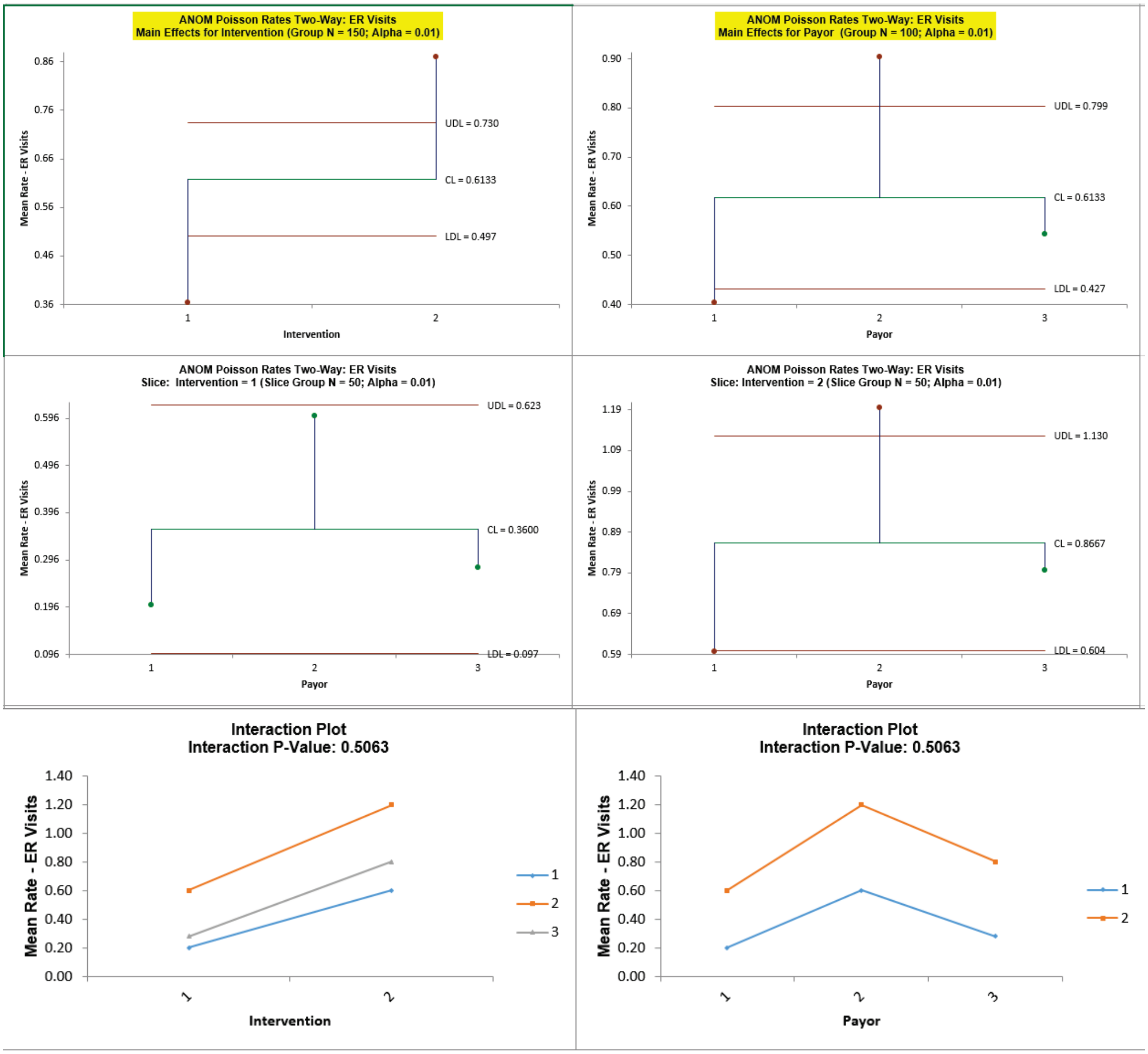
Intervention	Payor	ER Visits	Sample Size
1	1	10	50
1	2	30	50
1	3	14	50
2	1	30	50
2	2	60	50
2	3	40	50

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Poisson Rates Two-Way**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Select *ER Visits*, click **Numeric Data Variable (Y) >>**; select *Sample Size*, click **Subgroup Column or Size >>**; select *Intervention*, click **Group Category Factor (X1) >>**; select *Payor*, click **Group Category Factor (X2) >>**. **Alpha Level** = 0.01, **Adjust chart alpha for family-wise error rate** is unchecked:



4. Click **OK**. The ANOM Poisson Rates Two-Way Main Effects, Slice charts and Interaction plots are shown below:



5. Since the Interaction is not significant (P-Value >= 0.1, automatically determined from Poisson regression), the Main Effects charts are highlighted. Both factors Intervention and Payor are significant at alpha = 0.01. “Home nurse assigned” results in a significantly lower rate of ER visits versus “no home nurse assigned.” Payor Type 1 has a significantly lower rate of ER visits and Type 2 has a significantly higher rate.

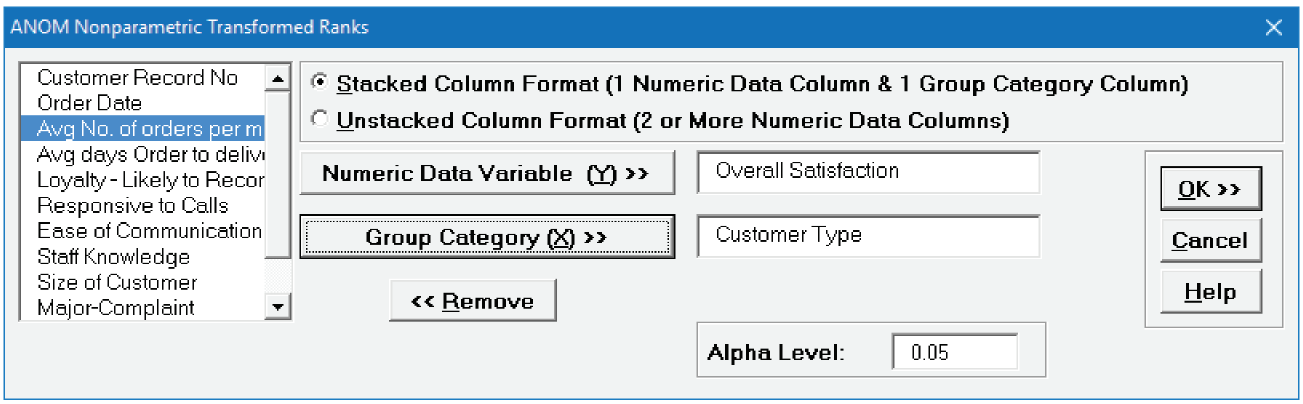
ANOM Nonparametric Transformed Ranks

ANOM Nonparametric Transformed Ranks is complementary to a Kruskal-Wallis test. Note that it is not included as an option in the Kruskal-Wallis dialog because there is a difference in the statistic used (Kruskal-Wallis uses ANOVA on ranks, whereas the ANOM utilizes a normal transformed rank).

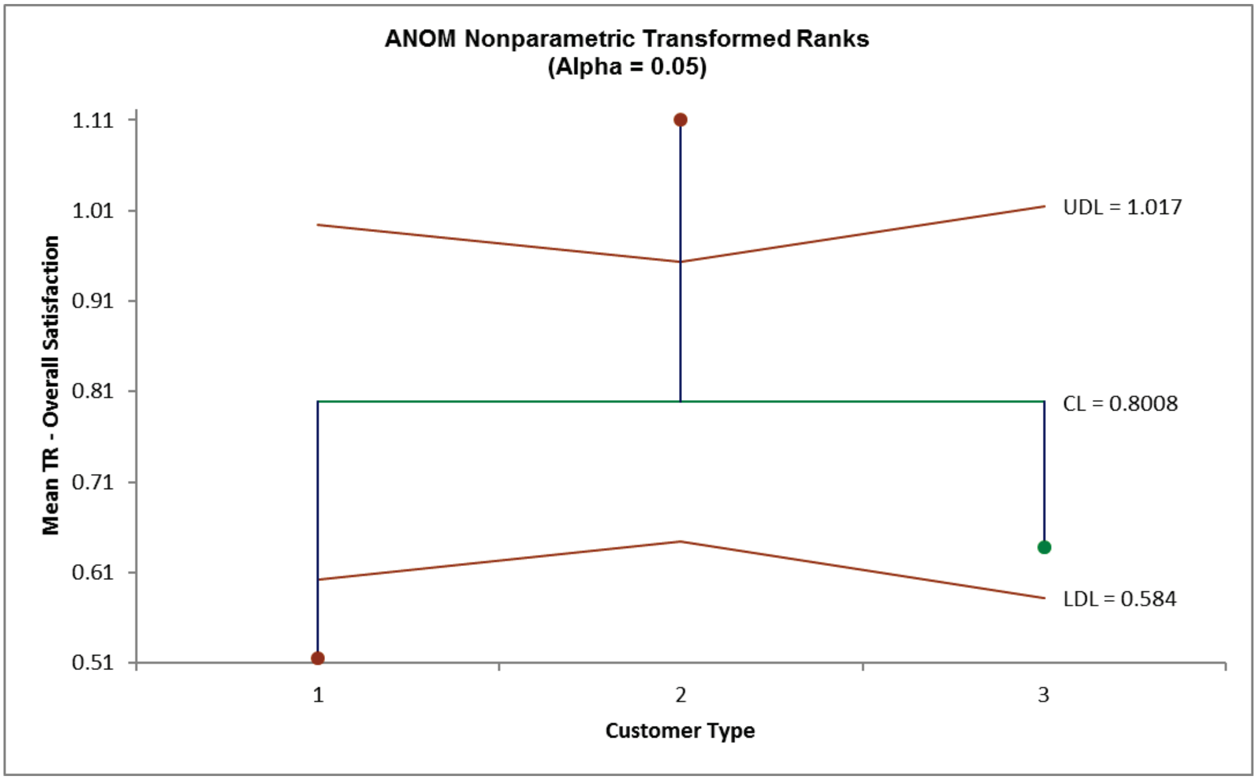
1. Open **Customer Data.xlsx**, click on **Sheet 1** tab.

2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Nonparametric Transformed Ranks**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category** **(X) >>**. Set **Alpha Level** = 0.05.



4. Click **OK**. The ANOM Nonparametric Transformed Ranks chart is shown below:



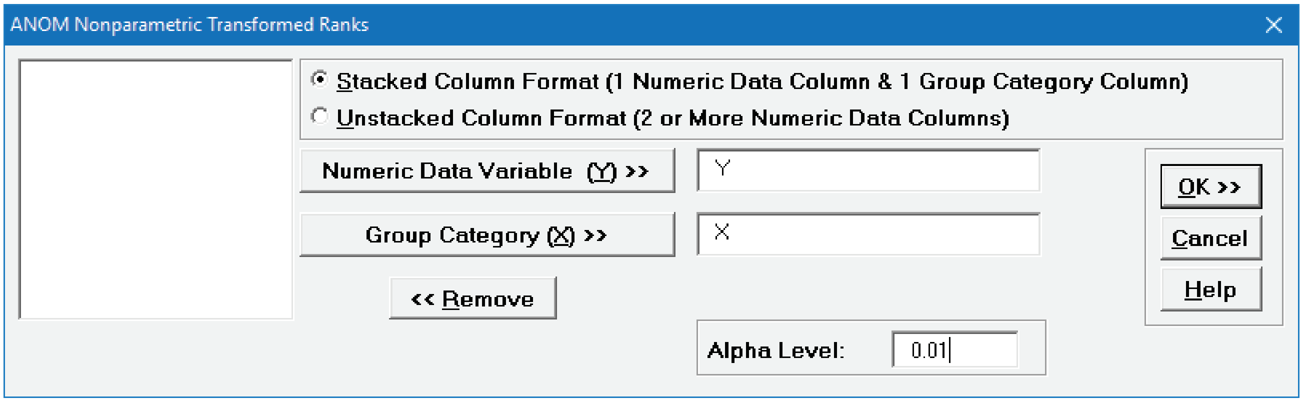
5. Here we see that Customer Type 1 mean transformed rank satisfaction score is significantly below the grand mean and Customer Type 2 is significantly higher. This is consistent with the results that were observed in the Kruskal-Wallis analysis (see Kruskal-Wallis).

6. The varying decision limits are due to the varying sample sizes for each Customer Type, with smaller sample size giving wider limits in a manner similar to a control chart. If the data are balanced, the decision limit lines will be constant.

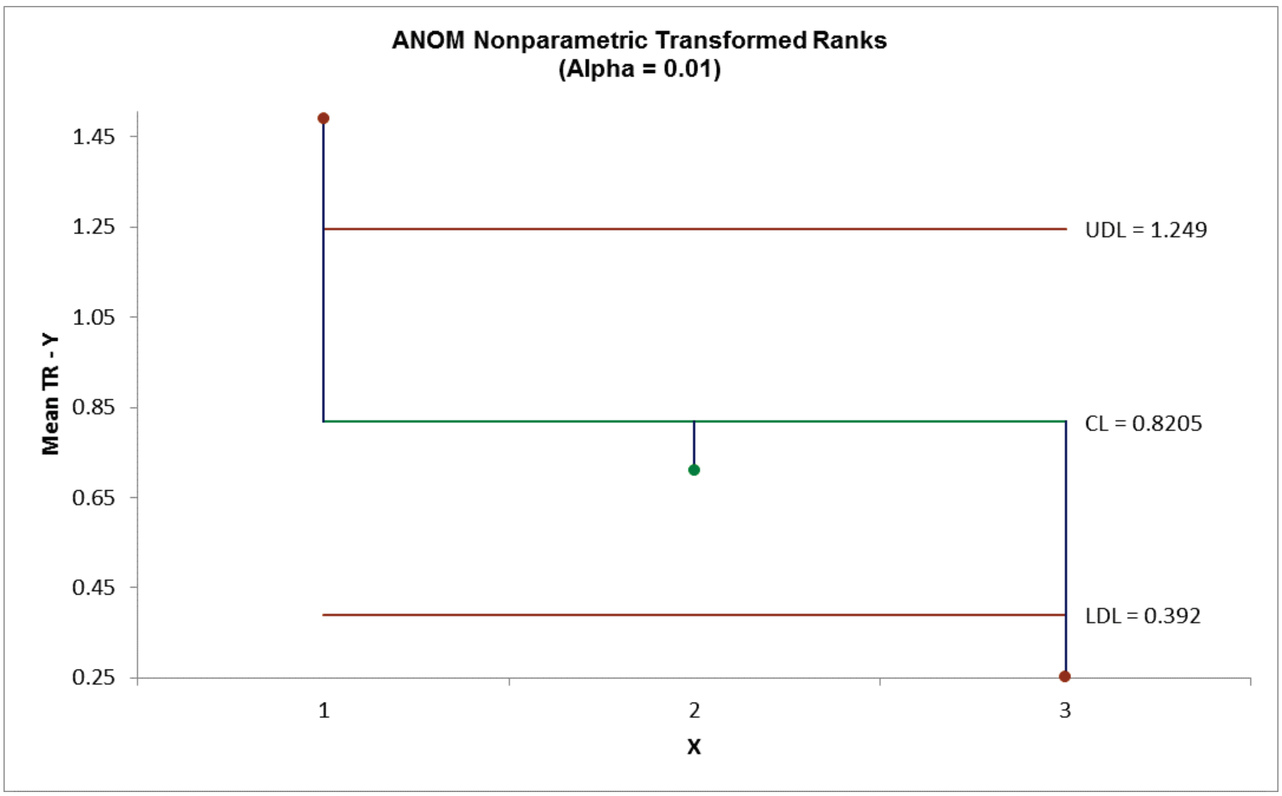
7. Open **ANOM Examples.xlsx**, click on the **Exponential Data – TR** tab. This is exponential data, Example 9.16 from the ANOM book with alpha level = 0.01 (used with author permission).

8. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Nonparametric Transformed Ranks**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

9. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Y*, click **Numeric Data Variable (Y) >>**; select *X*, click **Group Category (X) >>**. Set **Alpha Level** = 0.01.

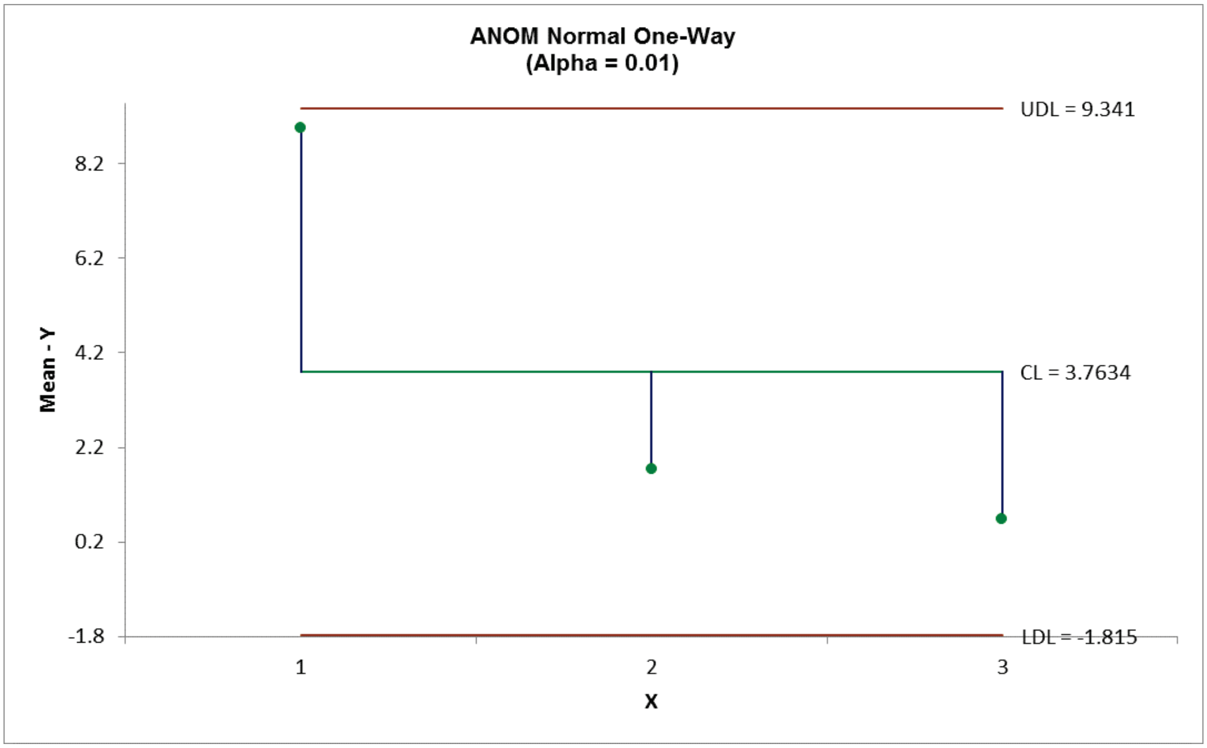


10. Click **OK**. The ANOM Nonparametric Transformed Ranks chart is shown below:



11. X = 1 shows as significantly higher than the grand mean transformed rank and X = 3 is significantly lower (at alpha = 0.01). These results match those given in the ANOM book.

12. By way of comparison, rerun the analysis using ANOM One-Way Normal (with alpha = 0.01). The results are:



13. ANOM Normal fails to detect the significant differences in mean noted above.

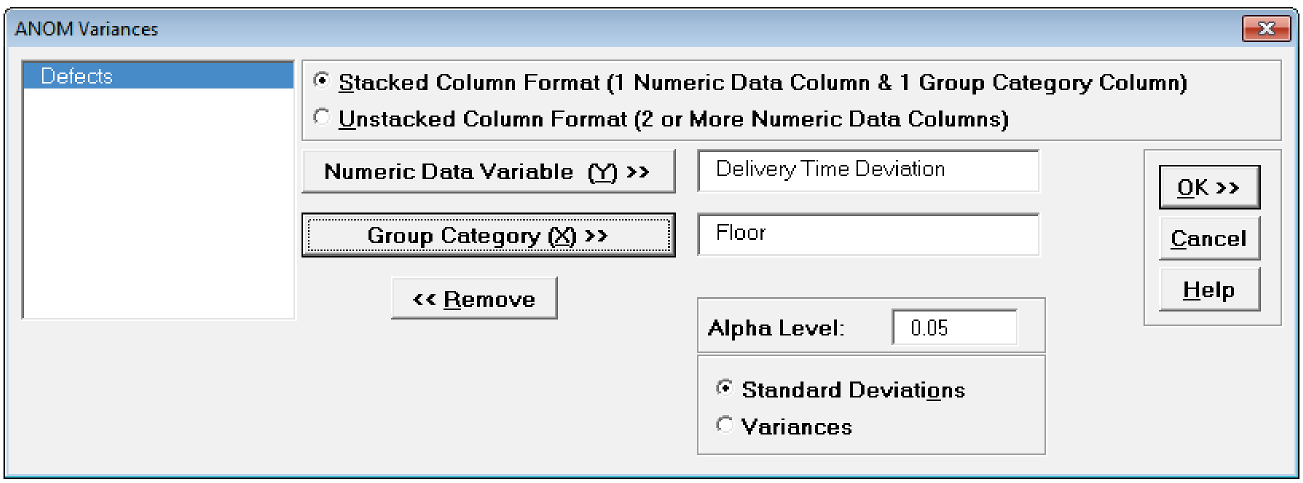
ANOM Variances

The ANOM Variances chart is complementary to Bartlett's Test for Equal Variance and is also available as an option in the Bartlett's Test dialog. It assumes that the data are normally distributed.

14. Open **Delivery Times.xlsx**, click on **Sheet 1** tab.

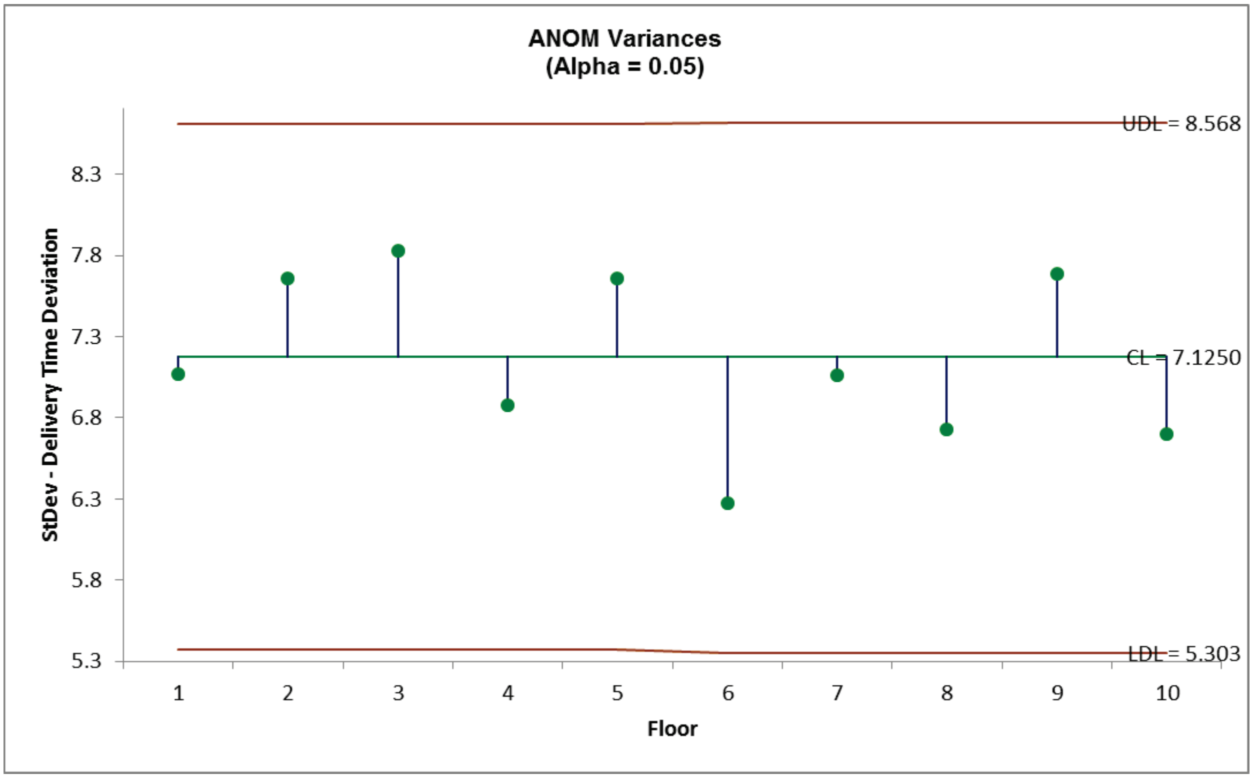
15. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Variances.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

16. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**; select *Floor*, click **Group Category (X) >>**; Select **Standard Deviations**; Set **Alpha Level** = 0.05.



We are analyzing the same normal data used in Bartlett's Test for Equal Variances. See [Bartlett's Test](#).

17. Click **OK**. The ANOM Variances chart is shown below:



18. The ANOM Variances chart visually shows that none of the group standard deviations are significantly different from the grand mean of all the standard deviations. It is called an ANOM Variances Chart but displays Standard Deviations for ease of interpretation (similar to a Standard Deviation S Control Chart). This does however result in non-symmetrical decision limits. To display Variances rerun the above analysis with option **Variances** selected.

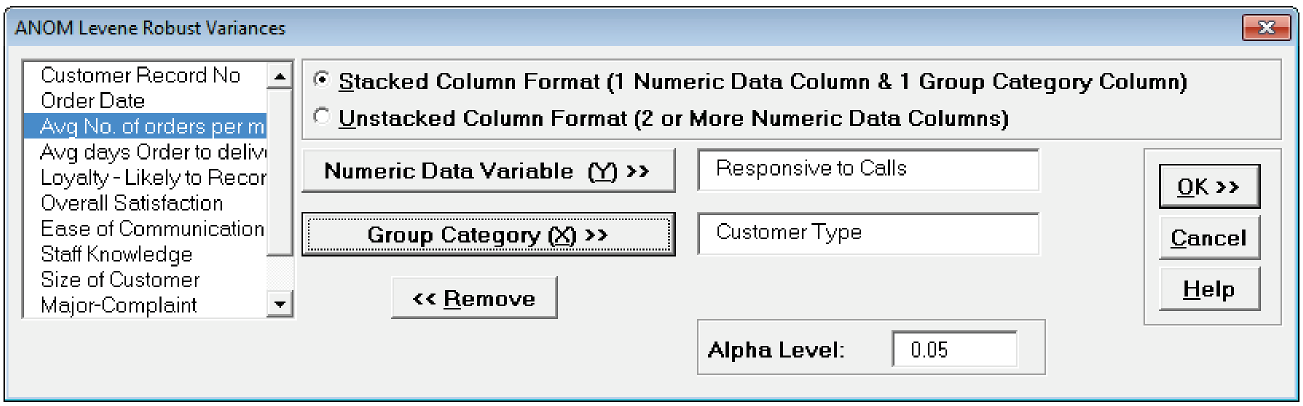
ANOM Levene Robust Variances

The ANOM Levene Robust Variances chart is complementary to Levene's Test for Equal Variance and is also available as an option in the Levene's Test dialog. Since it uses absolute deviations from the median, it is robust to the assumption of normality.

1. Open **Customer Data.xlsx**, click on **Sheet 1** tab.

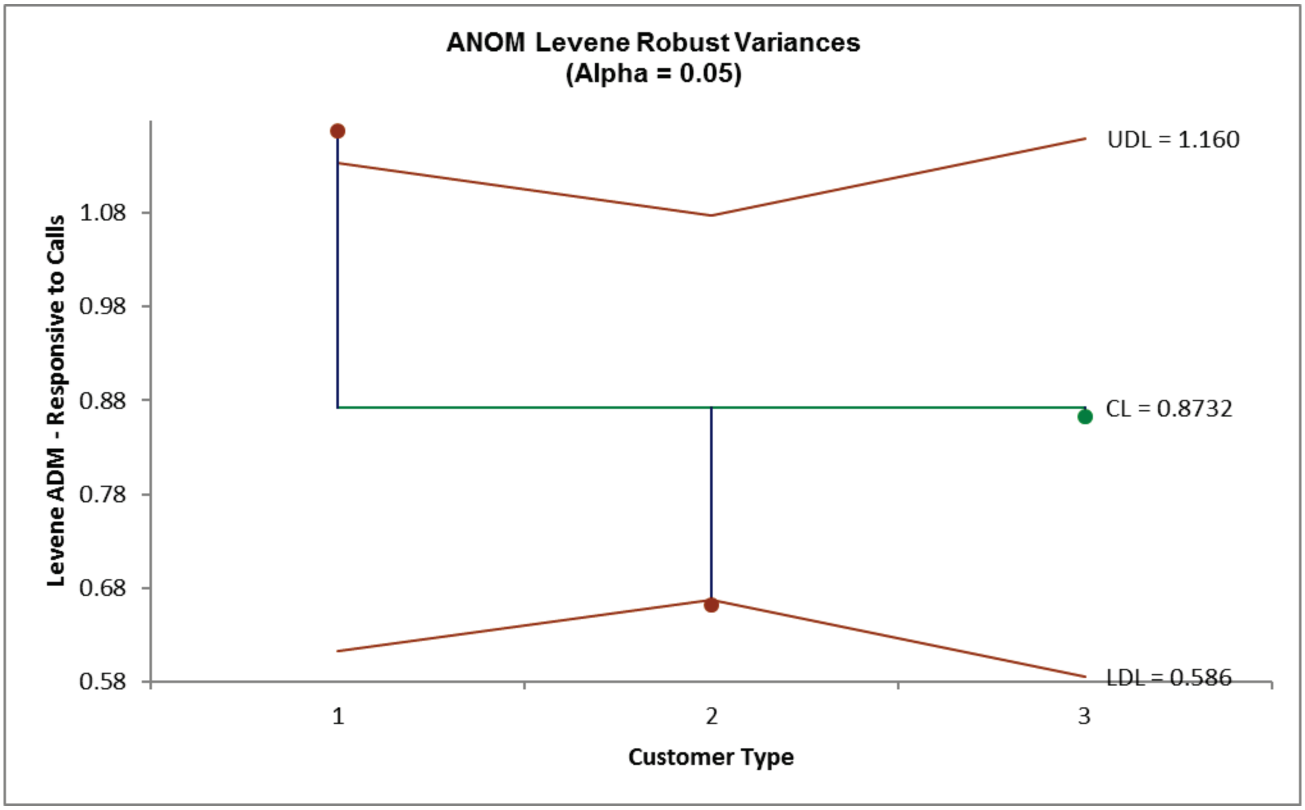
2. Click **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Levene Robust Variances.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**.

3. Click **Next**. Ensure that **Stacked Column Format** is checked. Select *Responsive to Calls*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X) >>;** Set **Alpha Level** = 0.05.



We are analyzing the same nonnormal data used in Levene's Test for Equal Variances. See **Levene's Test**.

4. Click OK. The ANOM Levene Robust Variances chart is displayed:



5. The ANOM chart clearly shows Customer Type 1 has significantly higher variance (ADM) than overall and Customer Type 2 has significantly lower variance.

6. The varying decision limits are due to the varying sample sizes for each Customer Type, with smaller sample size giving wider limits in a manner similar to a control chart. If the data are balanced the decision limit lines will be constant.

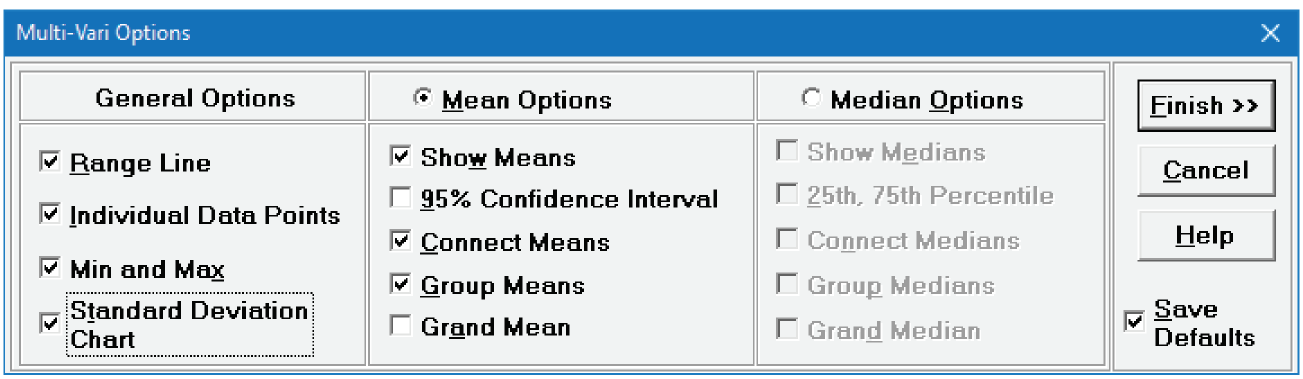
Part Q – Multi-Vari Charts

Multi-Vari Charts

The Multi-Vari chart is a powerful tool to identify dominant Sources of Variation (SOV). The three major “families” of variation are: Within Unit, Between Unit, and Temporal (Over Time). We will look at examples of each type of SOV and then use the Multi-Vari Chart to study Overall Satisfaction in the **Customer Data.xlsx** file.

1. Open **Multi-Vari Data.xlsx**, click Sheet **Within**. Select **SigmaXL > Graphical Tools > Multi-Vari Options**.

2. The charts shown will be updated as options are selected. Note that they are for demonstration purposes and are not based on the **Multi-Vari Data.xlsx** data (sample charts are not displayed in Excel for Mac). Ensure that all general **Options** are selected (**Range Line**, **Individual Data Points**, **Min and Max**, **Standard Deviation Chart**). Select **Mean Options**, ensure that **Show Means**, **Connect Means**, and **Group Means** are checked. Ensure that **Save Defaults** is checked. These settings would be typical for a Multi-Vari chart. (The Median options provide the ability to display percentiles as an alternative to the Means).



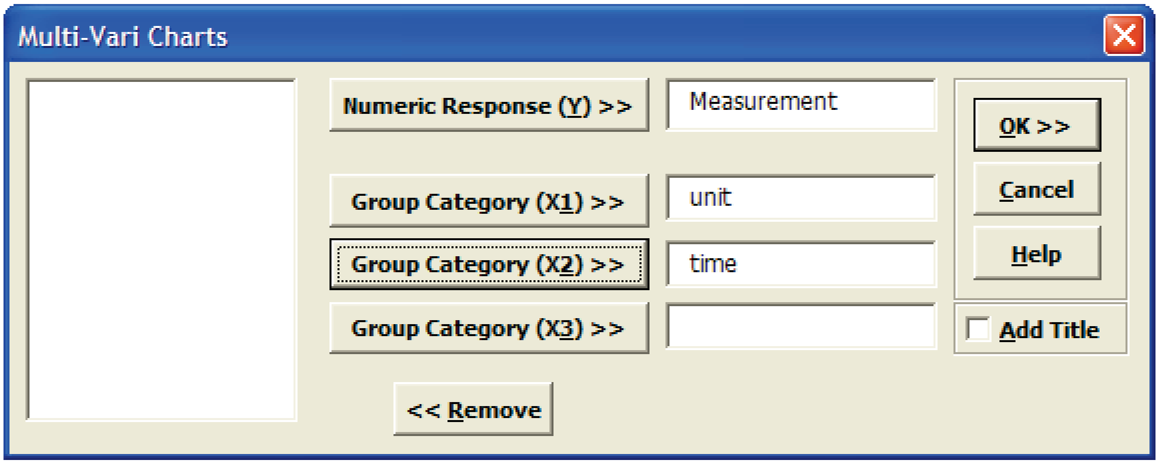
   Tip: Multi-Vari Charts can be used to display Confidence Intervals (as we did earlier in Part C). To do this, check the 95% Confidence Interval.

3. Click **Finish**. SigmaXL automatically starts the Multi-Vari Chart procedure (this is equivalent to clicking **SigmaXL > Graphical Tools > Multi-Vari Charts**).

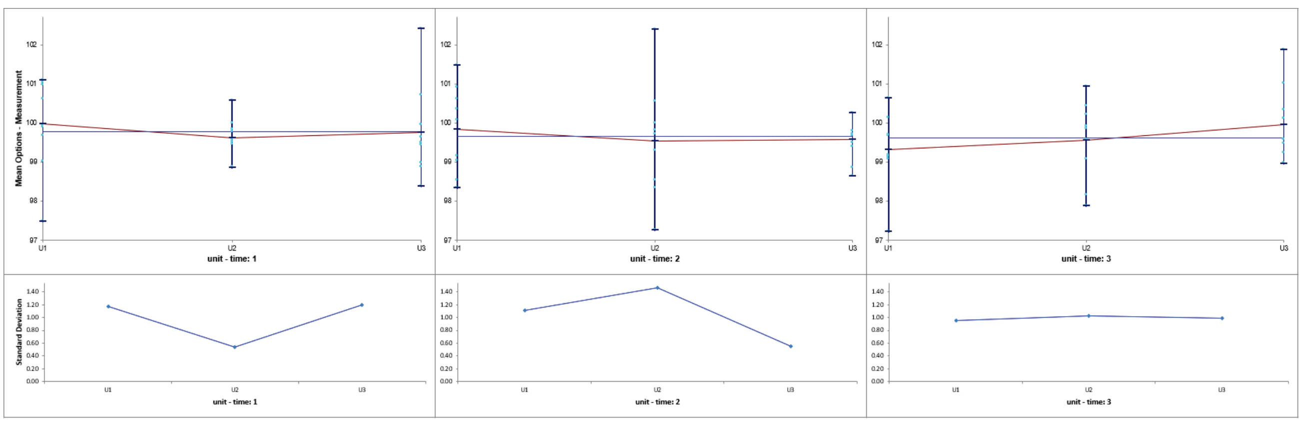
4. Check **Use Entire Data Table**. Click **Next**.

5. Note that the input X’s can be text or numeric but should be discrete. Y’s must be numeric - typically continuous, but can also be count or proportion data.

6. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1)** >>; select *time*, click **Group Category (X2) >>**.



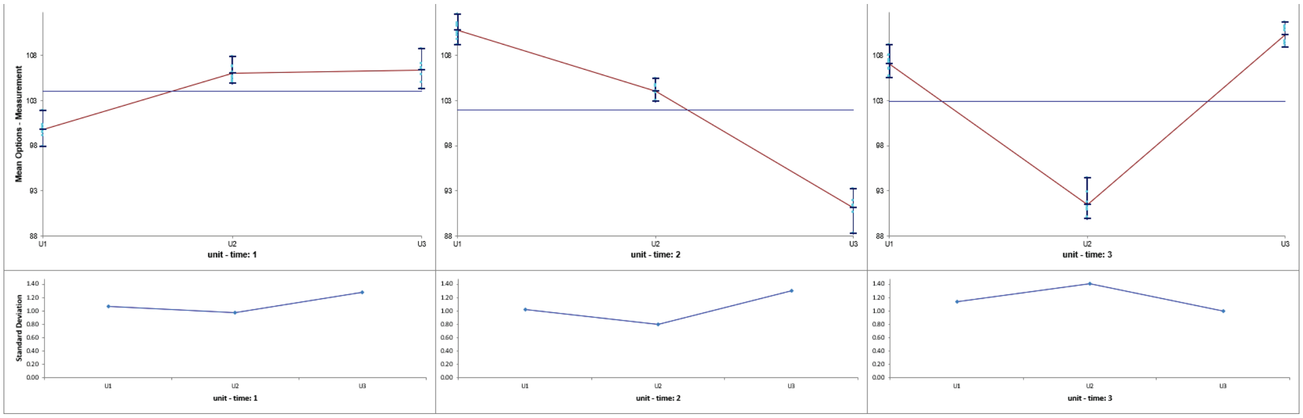
7. Click **OK**. Resulting Multi-Vari Chart illustrating dominant "Within Unit" Source of Variation:



8. Using **Multi-Vari Data.xlsx**, click Sheet **Between**. Select **SigmaXL > Graphical Tools > Multi-Vari Charts**. Check **Use Entire Data Table**. Click **Next**.

9. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1) >>**; select *time*, click **Group Category (X2) >>**.

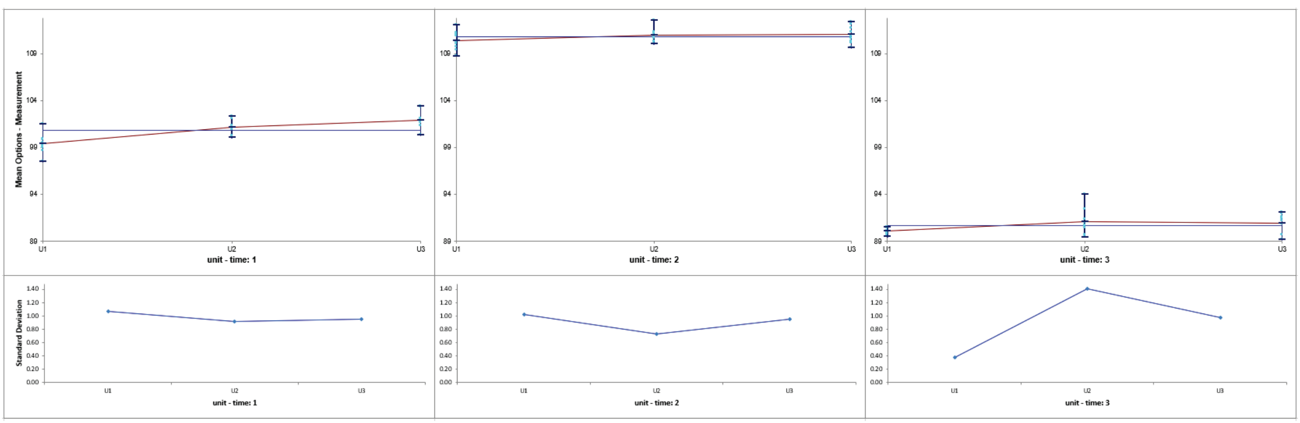
10. Click **OK**. Resulting Multi-Vari Chart illustrating dominant "Between Unit" Source of Variation:



11. Using **Multi-Vari Data.xlsx**, click Sheet **OverTime**. Select **SigmaXL > Graphical Tools > Multi-Vari Charts**. Check **Use Entire Data Table**. Click **Next**.

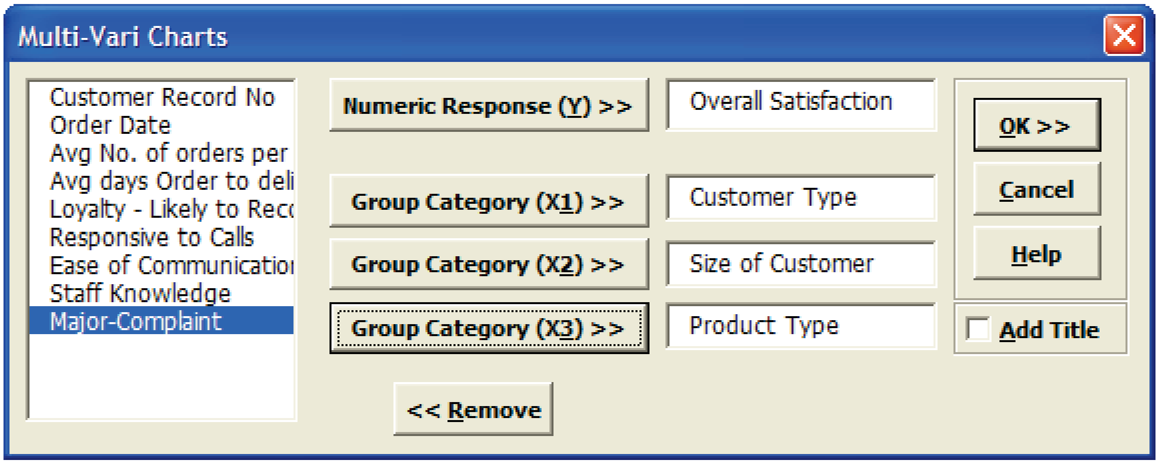
12. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1) >>**; select *time*, click **Group Category (X2) >>**.

13. Click **OK**. Resulting Multi-Vari Chart illustrating dominant "Over Time" Source of Variation:

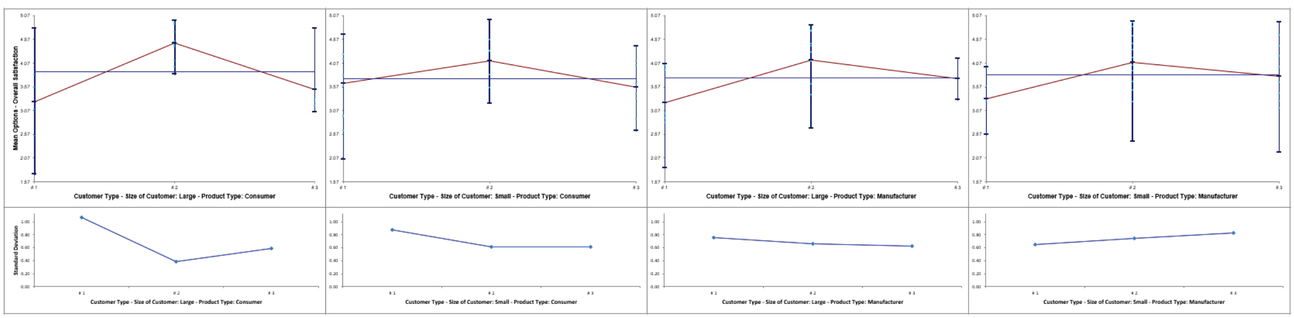


14. Open **Customer Data.xlsx**, click **Sheet 1**. Select **SigmaXL > Graphical Tools > Multi-Vari Charts**.

15. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**; select *Customer Type*, click **Group Category (X1) >>**; select *Size of Customer*, click **Group Category** (**X2) >>**; select *Product Type*, click **Group Category** (**X3) >>**.



16. Click **OK**. Resulting Multi-Vari chart:



Examining this Multi-Vari chart reveals that the dominant Source of Variation is "within" Customer Type, followed by "between" Customer Type. Furthermore, it would be worthwhile to examine the combination of Customer Type 2, Customer Size Large, and Product Type Consumer.

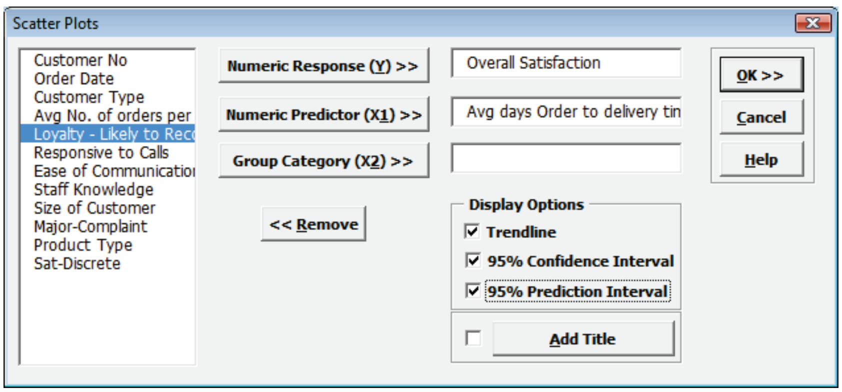
Other tools that can help us identify potential X factors that may explain some of the large "Within" variability are the Scatter Plot, Scatter Plot Matrix and Correlation Matrix.

Part R – Scatter Plots and XYZ Contour/Surface Plots

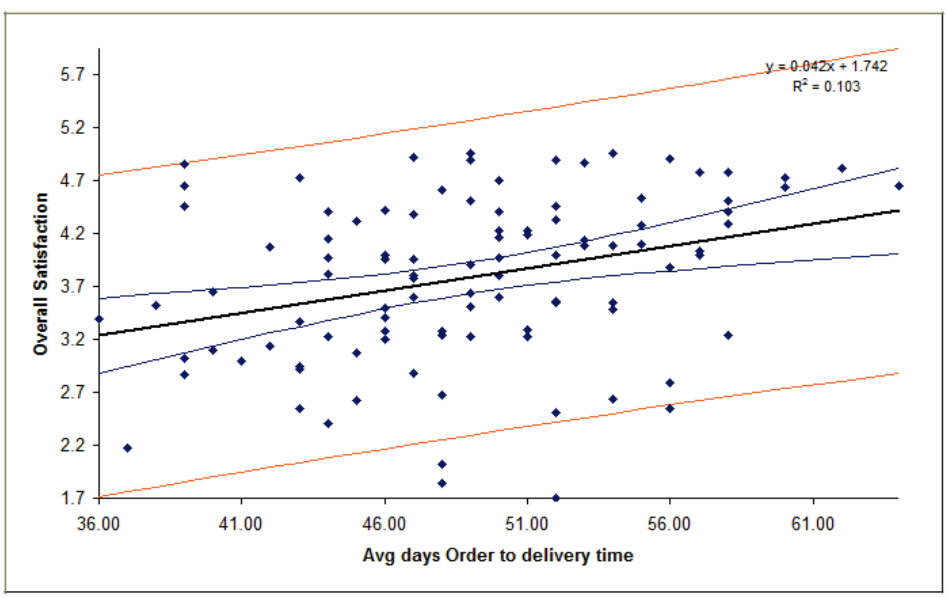
Scatter Plots with Trendline

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab. Click **SigmaXL > Graphical Tools > Scatter Plots**; if necessary, check **Use Entire Data Table**, click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**; select *Avg Days Order Time to Delivery*, click **Numeric Predictor (X1) >>**. Check **Trendline**, **95% Confidence Interval** and **95% Prediction Interval** as shown:



3. Click **OK**. The resulting Scatter Plot is shown with equation, trendline, 95% confidence interval (blue lines – for a given X value this is the 95% confidence interval for predicted mean Overall Satisfaction) and 95% prediction interval (red lines – for a given X value this is the 95% confidence interval for predicted individual values of Overall Satisfaction).



The equation is based on linear regression, using the method of least squares. R-squared * 100 is the percent variation of Y explained by X (here 10.3%).

4. Now we want to redo the Scatter Plot and stratify by Customer Type. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. (Or, Click **Sheet 1** Tab; Click **SigmaXL > Graphical Tools > Scatter Plots**; click **Next**.)

5. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**; select *Average Days Order Time to Delivery*, click **Numeric Predictor (X1) >>**; select *Customer Type*, click **Group Category (X2) >>**, Uncheck **95% Confidence Interval** and **95% Prediction Interval**. Click **OK**:

Clearly, according to the analysis, Customer Type 3 is happier if orders take longer! But, does this make sense? Of course not! Customer Sat scores should not increase with Order to Delivery time. What is happening here? This is a coincidental association. Something else is driving customer satisfaction. Later, we will look at the Scatter Plot Matrix to help us investigate other factors influencing Customer Satisfaction.

Tip: Be careful when interpreting scatter plots (and other statistical tools): Y versus X correlation or statistical significance does not always mean that we have a causal relationship. Umbrella sales are highly correlated to traffic accidents, but we cannot reduce the rate of traffic accidents by purchasing fewer umbrellas! The best way to validate a relationship is to perform a Design of Experiments (see Improve Phase).

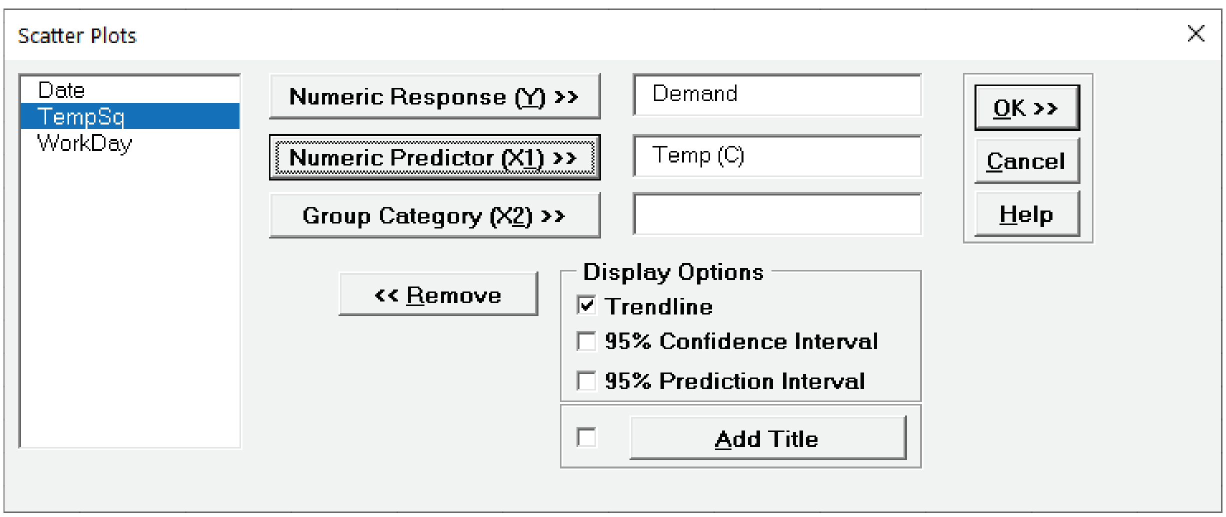
Scatter Plot with Quadratic Trendline

SigmaXL does not include nonlinear or polynomial curve fitting, so this example shows how a trendline in a **Scatter Plot with Trendline** may be modified in Excel to show a quadratic relationship.

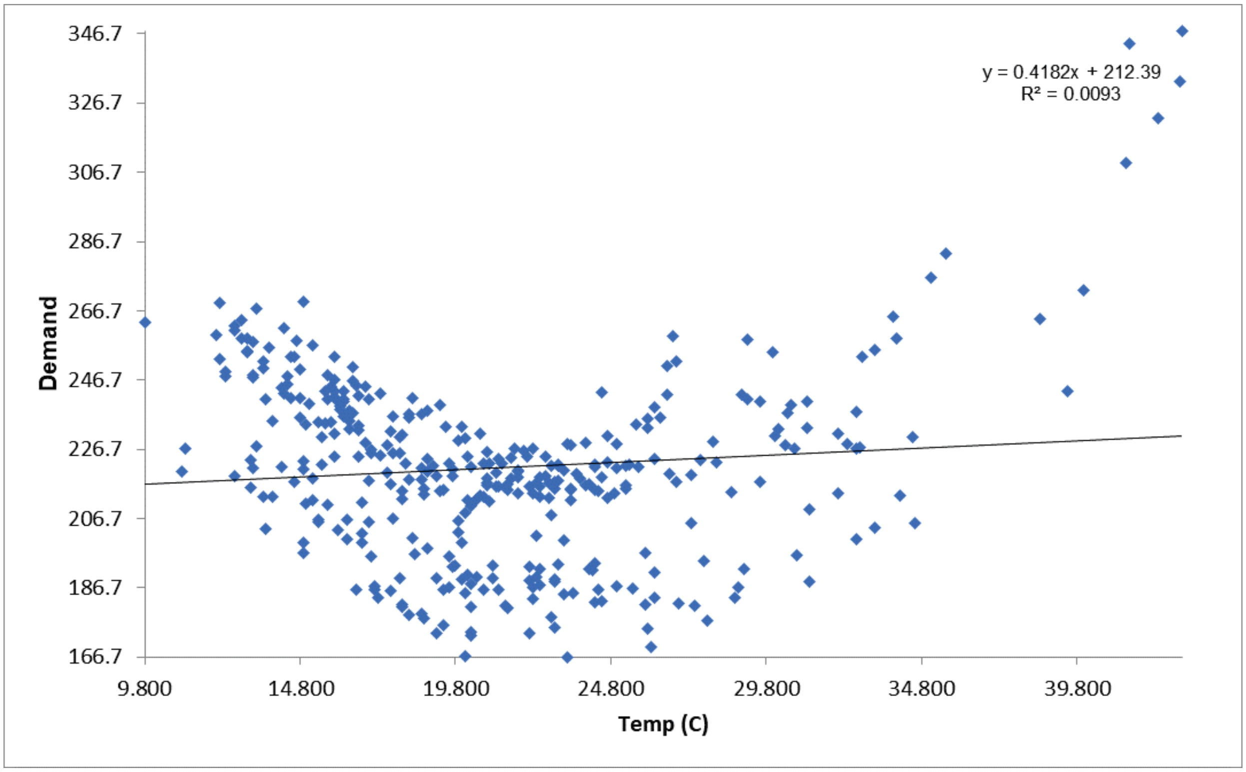
1. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014. Temp (C) is the maximum daily temperature in degrees Celsius for the city of Melbourne. (The additional columns are not used here: TempSq is Temperature squared, WorkDay takes on the value 1 on work days and 0 otherwise. This data is analyzed later using Time Series Forecasting, see ARIMA Forecast with Predictors).

2. Click **SigmaXL > Graphical Tools > Scatterplots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

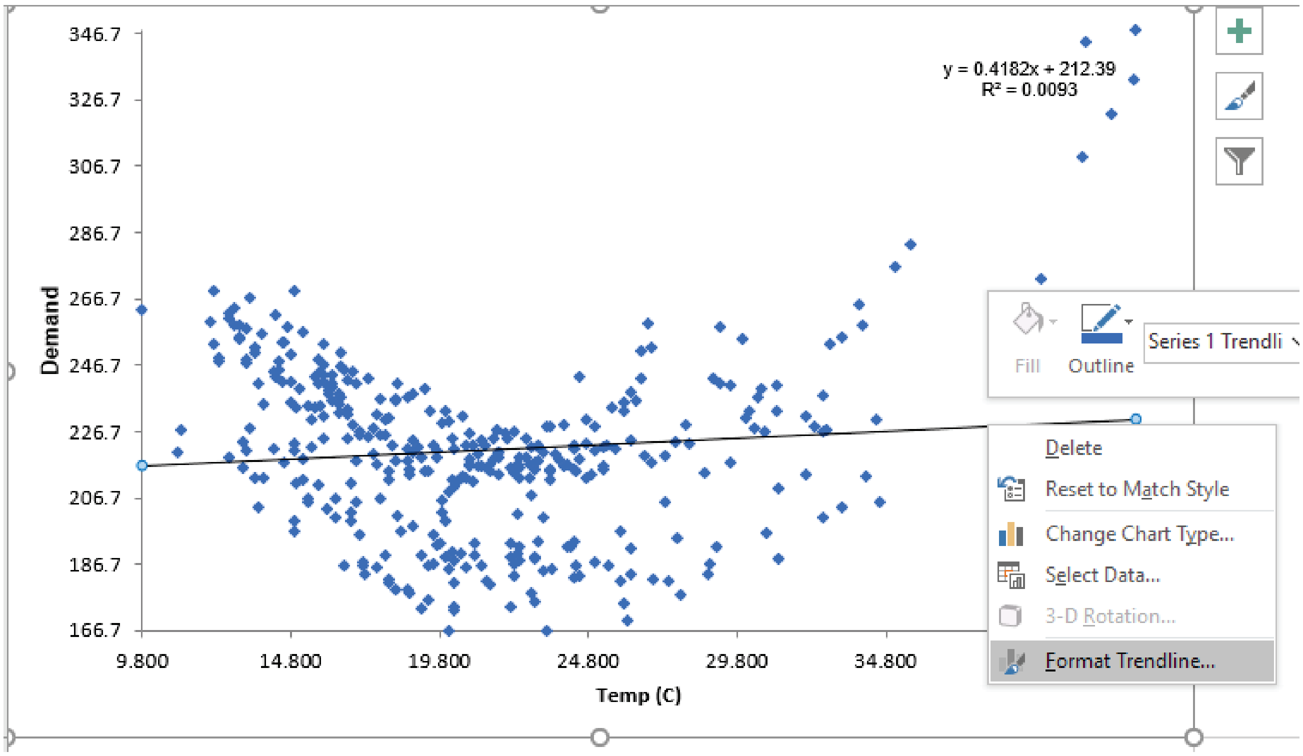
3. Select *Demand*, click **Numeric Response (Y) >>**; select *Temp (C)*, click **Numeric Predictor (X1) >>**. Check **Trendline**.



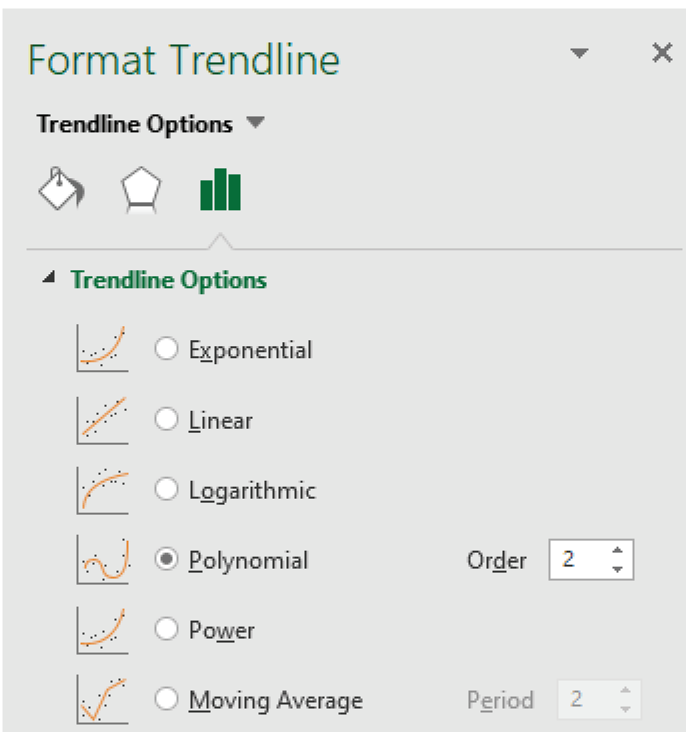
4. Click **OK**. A Scatter Plot of Electricity Demand versus Temperature is produced.



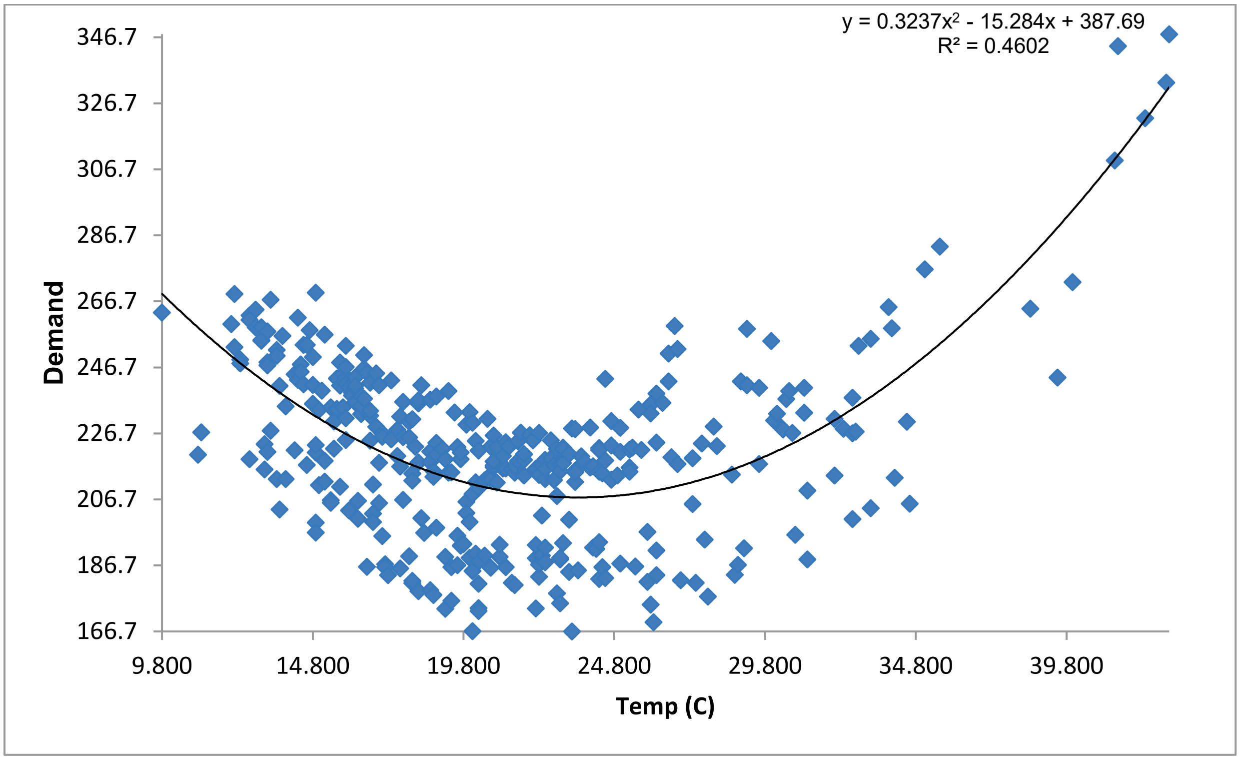
5. This shows a quadratic relationship: high temperatures in the summer cause electricity demand for air conditioning, low temperatures in the winter cause demand for heating. We will modify the trendline in Excel to a quadratic fit. Click on the Trendline, right click and select Format Trendline as shown:



The **Format Trendline** options are given. Select **Polynomial** with **Order** 2 as shown.



6. The Trendline is now a quadratic function as shown:



7. P-Values for coefficients and model residuals are not available. For a quadratic model like this, these may be obtained using Multiple Linear Regression with Temp and TempSq as model predictors.

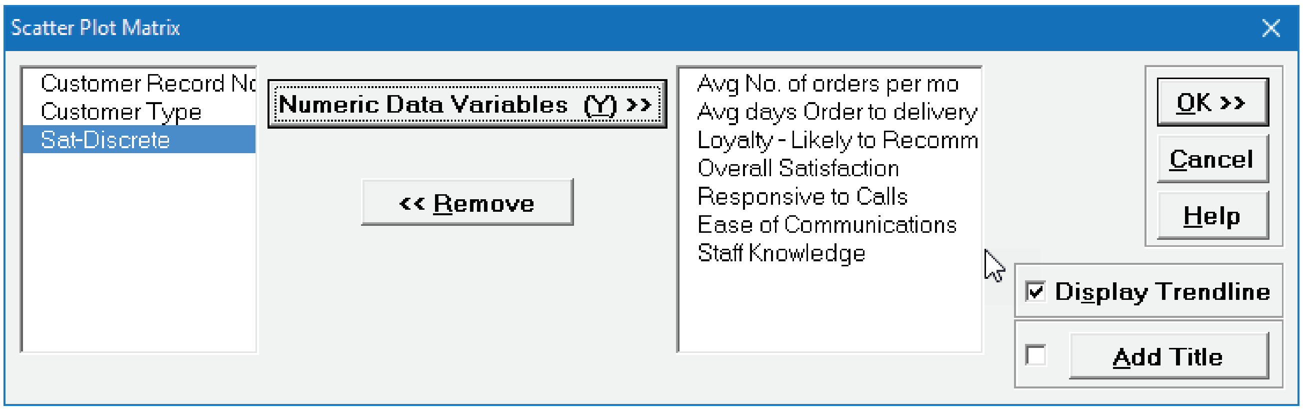
8. Nonlinear curve fitting may also be performed using Excel's Solver. See, for example, this video tutorial by Taylor Sparks: https://youtu.be/Ewp5CF5ba_w.

Scatter Plot Matrix

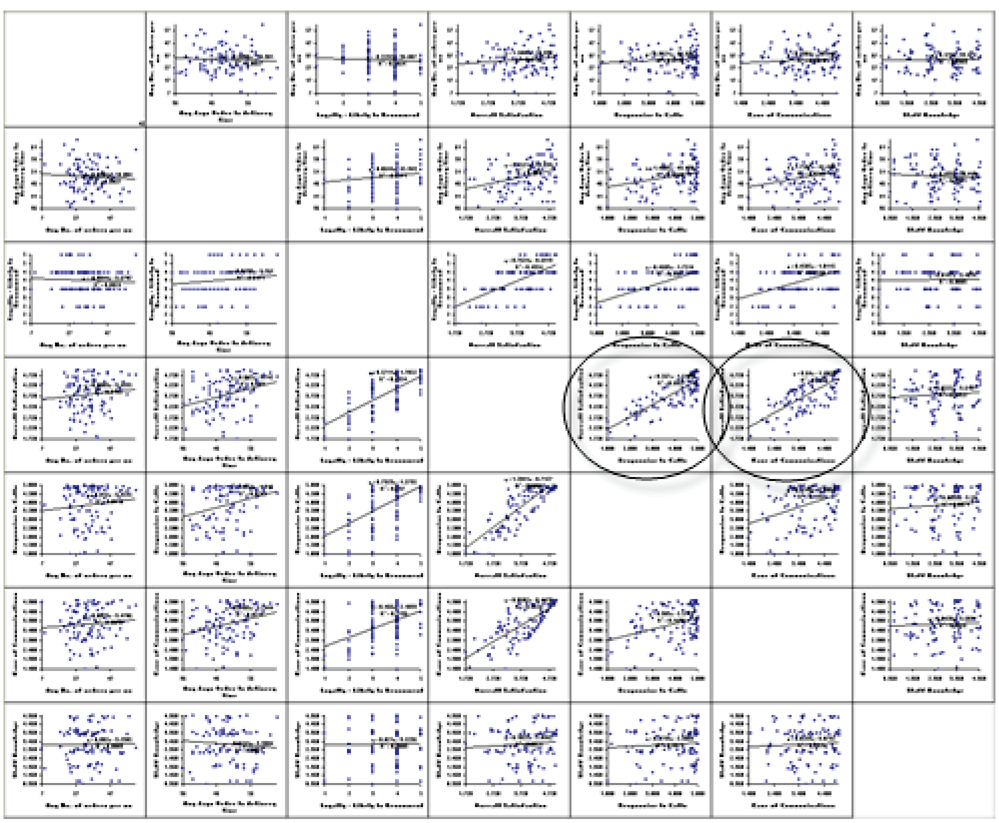
1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet). Click **SigmaXL > Graphical Tools > Scatter Plot Matrix**.

2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select the variable *Avg No. of orders per month*; shift-click on *Staff Knowledge* and click **Numeric Data Variables (Y) >>** as shown:



4. Click **OK**. Resulting Scatter Plot Matrix:



Of particular interest is Overall Satisfaction versus Responsiveness to Calls and Ease of Communications. These will be explored further with Multiple Linear Regression.

XYZ Contour/Surface Plot

The XYZ Contour/Surface Plot is used to plot a continuous response Z versus two continuous predictors Y and X in order to understand the possible relationship without a regression model.

Bivariate interpolation or extrapolation is used to create a mesh or grid of regularly spaced x- and y-values on which the contour or surface plot is based. The default mesh size is 20 x 20, but this may be modified.

Available interpolation/extrapolation methods include:

- Automatic with Cross Validation (default)
- Inverse Distance Weighting with Power
- Akima's Polynomial with Extrapolation
- Akima's Polynomial with Boundary at Min Z Value (no extrapolation)
- Biharmonic Spline.

The X and Y values may also be standardized which keeps both variables using the same units. An XY Scatterplot may be produced to view the coverage region (a.k.a. convex hull), beyond which extrapolation is required.

The choice of which interpolation/extrapolation method and settings are best to use is dependent on the data. Inverse Distance Weighting is more robust to outliers and sudden transitions than Akima or Biharmonic, but will not be as accurate for a smooth response.

Leave-One-Out Cross Validation may be used to assess the interpolation/extrapolation accuracy. Cross Validation R-Square and RMSE (root mean square error) are reported. If the maximum number of data points is exceeded (100 default), a random sample is used to keep the computation time reasonable.

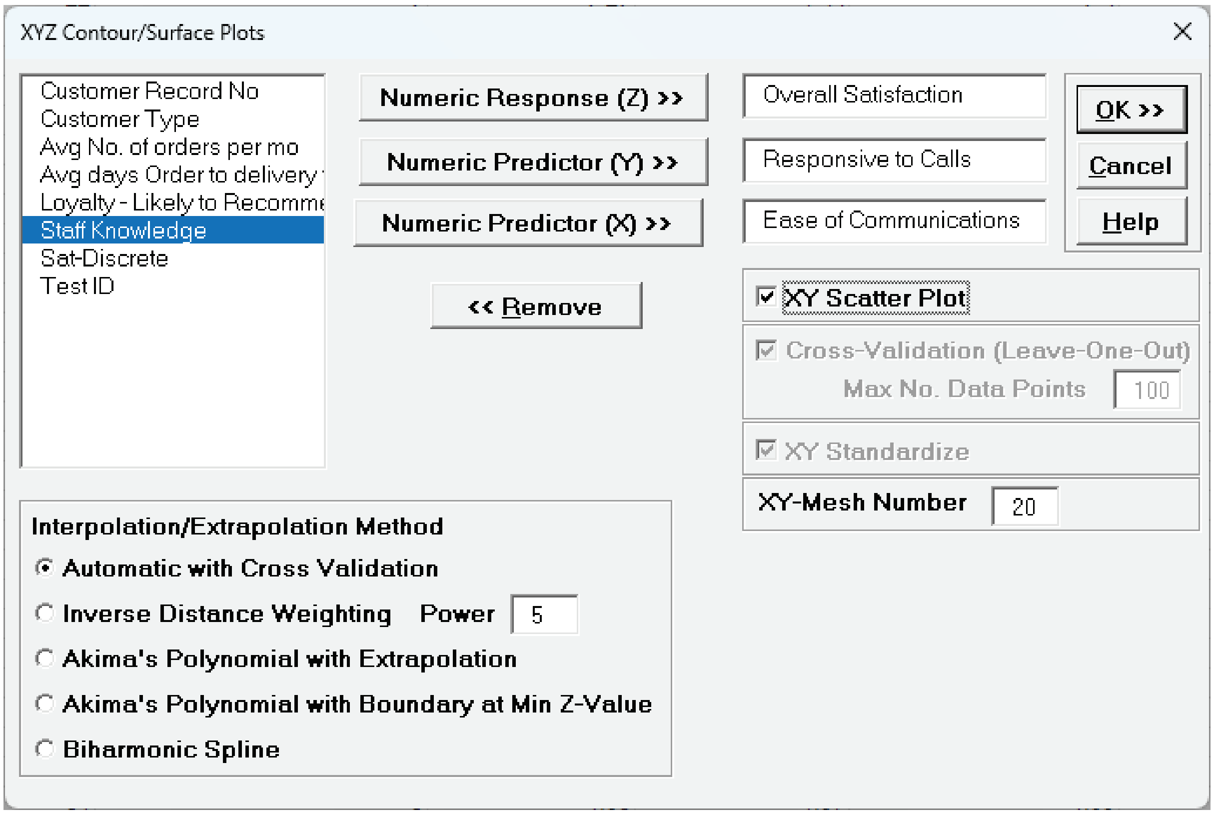
The default Automatic with Cross Validation option will attempt Inverse Distance Weighting with Power values from 1 to 12, Akima's Polynomial with Extrapolation and Biharmonic Spline, each with and without XY Standardization. Leave-One-Out Cross Validation is used to measure the prediction accuracy and the method with lowest RMSE is selected.

For formula details and references, see the Appendix: **XYZ Contour/Surface Plot**.

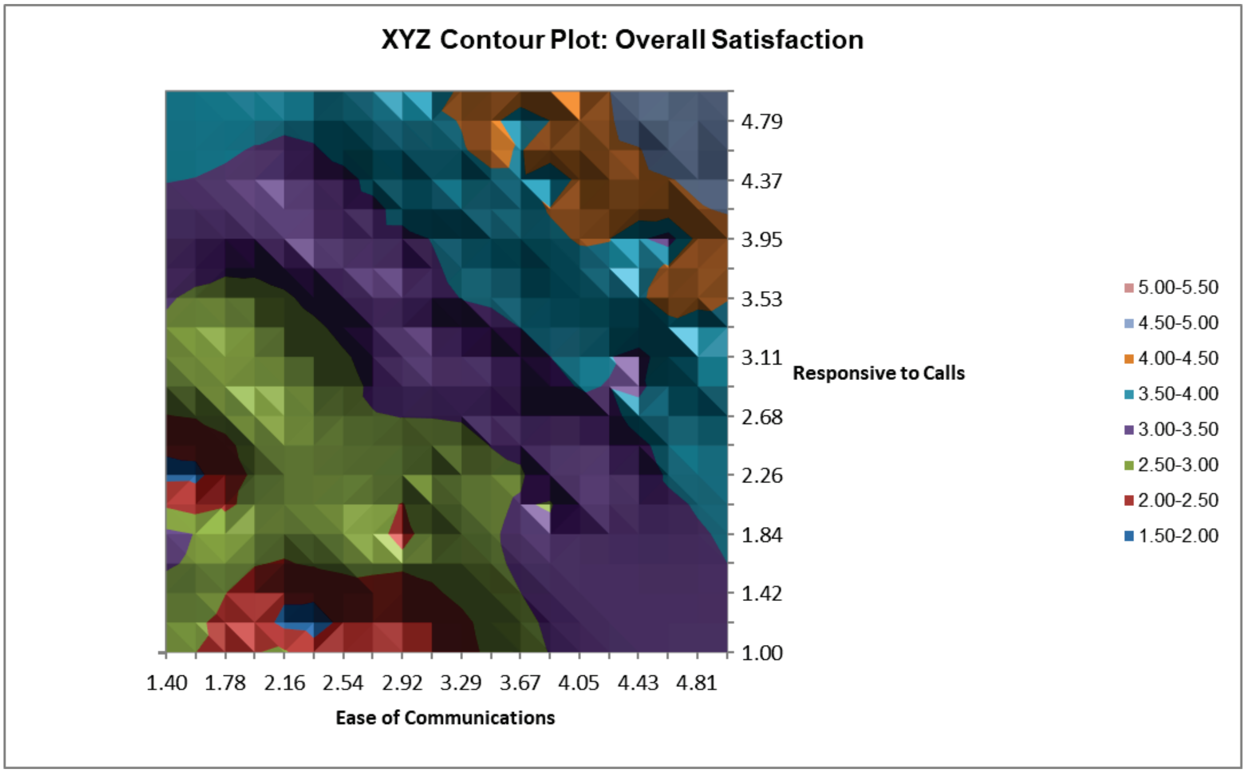
XYZ Contour/Surface Plot – Example

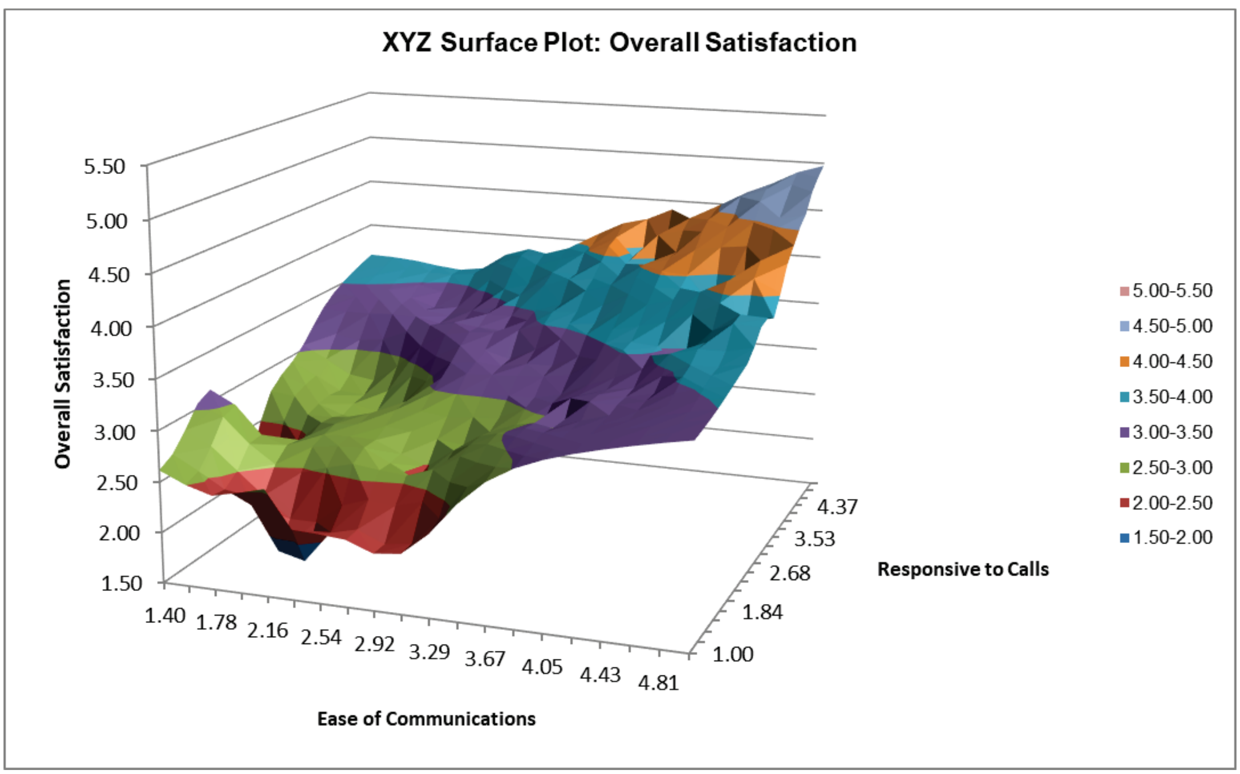
1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab. Click **SigmaXL > Graphical Tools > XYZ Contour/Surface Plot**; if necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Response (Z) >>**; select *Responsive to Calls*, click **Numeric Predictor (Y)**; select *Ease of Communications*, click **Numeric Predictor (X)**. Check **XY Scatter Plot**. Use the default **XY-Mesh Number** = 20 and **Interpolation/Extrapolation Method** as **Automatic with Cross Validation**.



3. Click **OK**. The XYZ Contour and Surface Plots are produced:





Clearly to maximize *Overall Satisfaction*, both *Responsive to Calls* and *Ease of Communications* must be maximized.

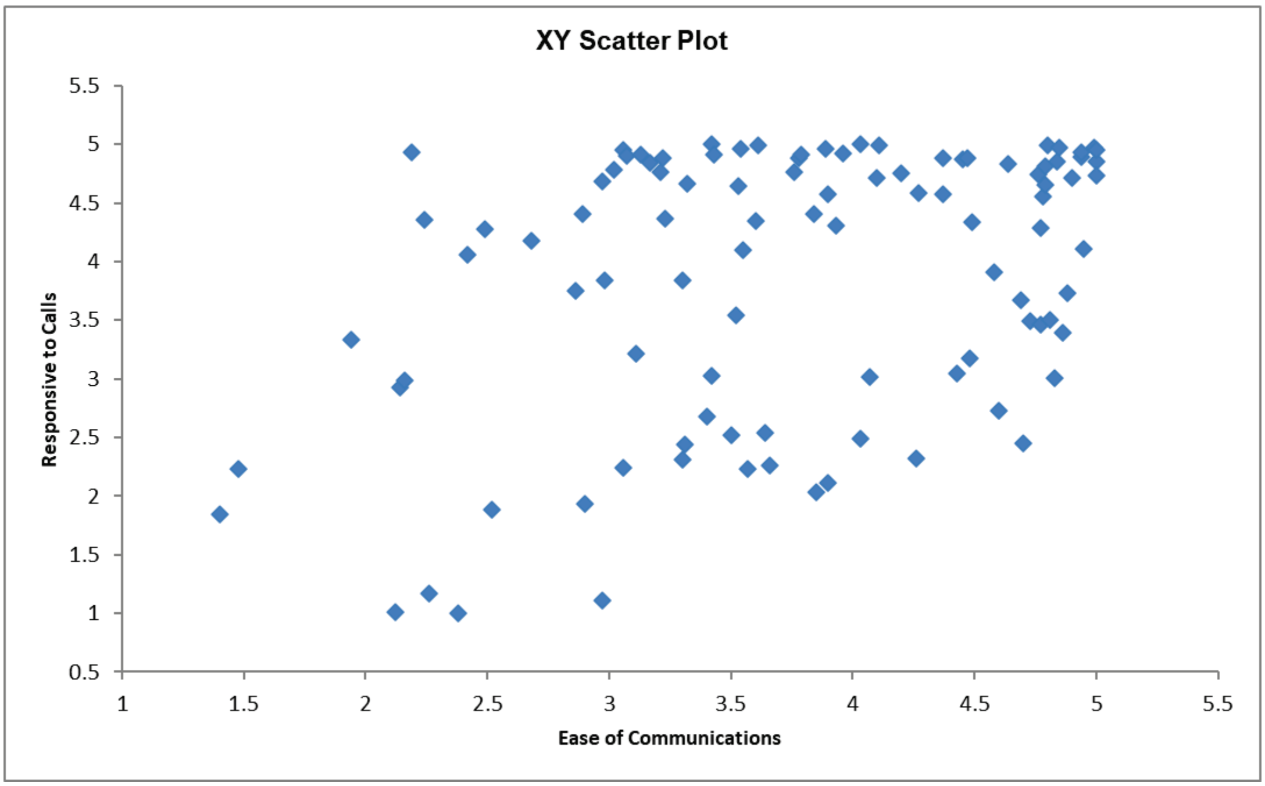
Tip: Excel's 3D Rotation Tool may be used by clicking on the chart, right-click and select 3D Rotation.

The Interpolation/Extrapolation and Cross-Validation report are given as:

Interpolation/Extrapolation Method (Automatic): Inverse Distance Weighting IDW (Power = 3)
XY is not Standardized; XY Mesh Number = 20 * 20
Cross-Validation (Leave-One-Out): R-Square = 83.91%; RMSE = 0.294403
Cross-Validation No. Data Points = 100

Inverse Distance Weighting with Power = 3 was selected by the Automatic Cross Validation method to produce the lowest RMSE. You can use **Recall Last Dialog** (or press **F3**) to experiment with other options.

The XY Scatter Plot is displayed as:



It is important to see where we do and do not have data coverage. High *Ease of Communications* with high *Responsive to Calls* are dense with data points so interpolation in this region will be more accurate. High *Ease of Communications* with low *Responsive to Calls* and High *Responsive to Calls* with low *Ease of Communications* have no data points so extrapolation is used and will be less accurate.

Part S – Correlation Matrix

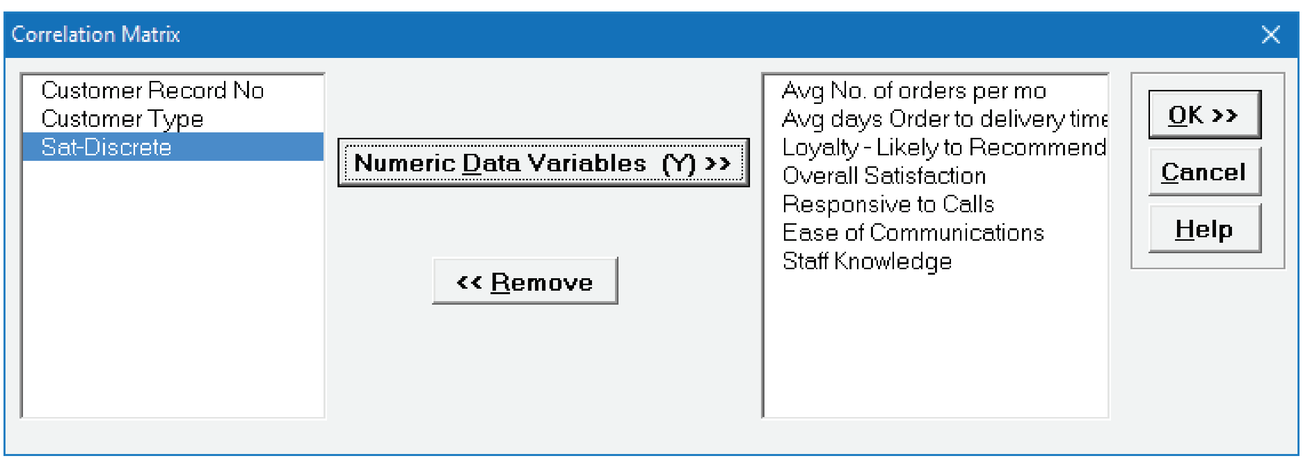
The correlation matrix complements the scatterplot matrix by quantifying the degree of association. The following table shows the approximate relationship between r, R-squared, and degree of association:

Pearson Correlation Coefficient (r)	R-Squared (%)	Degree of association
0.9 <= |r| <=1	> 80 %	Strong
0.7 <= |r| < 0.9	50 % to 80 %	Moderate
|r| < 0.7	< 50 %	Weak
Pearson Probability, p > 0.05		None

Correlation Matrix

1. Open **Customer Data.xlsx**. Click **Sheet 1** tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Correlation Matrix**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

2. Select the variable *Avg No. of orders per month*; shift-click on *Staff Knowledge* and click **Numeric Data Variables (Y) >>** as shown:



3. Click **OK**. Resulting Correlation Matrix is shown:

Pearson Correlations	Avg No. of orders per mo	Avg days Order to delivery time	Loyalty - Likely to Recommend	Overall Satisfaction	Responsive to Calls	Ease of Communications	Staff Knowledge
Avg No. of orders per mo	1	-0.0518	-0.0491	0.1155	0.1076	0.0885	0.0186
Avg days Order to delivery time		1	0.1307	0.3210	0.2725	0.2681	-0.0781
Loyalty - Likely to Recommend			1	0.6599	0.5805	0.4622	0.0176
Overall Satisfaction				1	0.8262	0.7454	0.0766
Responsive to Calls					1	0.3791	0.0845
Ease of Communications						1	0.0506
Staff Knowledge							1

Pearson Probabilities	Avg No. of orders per mo	Avg days Order to delivery time	Loyalty - Likely to Recommend	Overall Satisfaction	Responsive to Calls	Ease of Communications	Staff Knowledge
Avg No. of orders per mo		0.6090	0.6279	0.2527	0.2865	0.3812	0.8541
Avg days Order to delivery time			0.1949	0.0011	0.0061	0.0070	0.4398
Loyalty - Likely to Recommend				0.0000	0.0000	0.0000	0.8622
Overall Satisfaction					0.0000	0.0000	0.4490
Responsive to Calls						0.0001	0.4035
Ease of Communications							0.6171
Staff Knowledge							

Spearman Rank Correlations	Avg No. of orders per mo	Avg days Order to delivery time	Loyalty - Likely to Recommend	Overall Satisfaction	Responsive to Calls	Ease of Communications	Staff Knowledge
Avg No. of orders per mo	1	-0.0305	-0.0917	0.1006	0.0738	0.1000	0.0187
Avg days Order to delivery time		1	0.1097	0.3407	0.2489	0.2613	-0.0828
Loyalty - Likely to Recommend			1	0.6167	0.5507	0.4071	-0.0190
Overall Satisfaction				1	0.7782	0.7509	0.0890
Responsive to Calls					1	0.3204	0.0895
Ease of Communications						1	0.0716
Staff Knowledge							1

Spearman Rank Probabilities	Avg No. of orders per mo	Avg days Order to delivery time	Loyalty - Likely to Recommend	Overall Satisfaction	Responsive to Calls	Ease of Communications	Staff Knowledge
Avg No. of orders per mo		0.7629	0.3643	0.3192	0.4655	0.3222	0.8532
Avg days Order to delivery time			0.2774	0.0005	0.0125	0.0087	0.4127
Loyalty - Likely to Recommend				0.0000	0.0000	0.0000	0.8514
Overall Satisfaction					0.0000	0.0000	0.3786
Responsive to Calls						0.0012	0.3758
Ease of Communications							0.4792
Staff Knowledge							

Correlations highlighted in red are considered significant (P-Values < .05). The corresponding correlation coefficients above the P-Values are also highlighted in red. (Compare these to the Scatterplot Matrix.)

Note that Spearman's Rank Correlation complements Pearson's Correlation, in that it provides a robust measure of association. Spearman's rank is based on correlated ranks, which are not sensitive to outliers or departures from normality.

An automatic normality check is applied, utilizing the powerful Doornik-Hansen bivariate normality test. A yellow highlight recommends Pearson or Spearman correlations be used (but only if it is significant). Pearson is highlighted if the data are bivariate normal, otherwise Spearman is highlighted. Always review the data graphically with the scatterplot matrix as well.

Reference

Doornik, J.A. and Hansen, H. "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin Of Economics And Statistics*, 70, Supplement (2008).

Part T – Multiple Regression

Multiple Regression

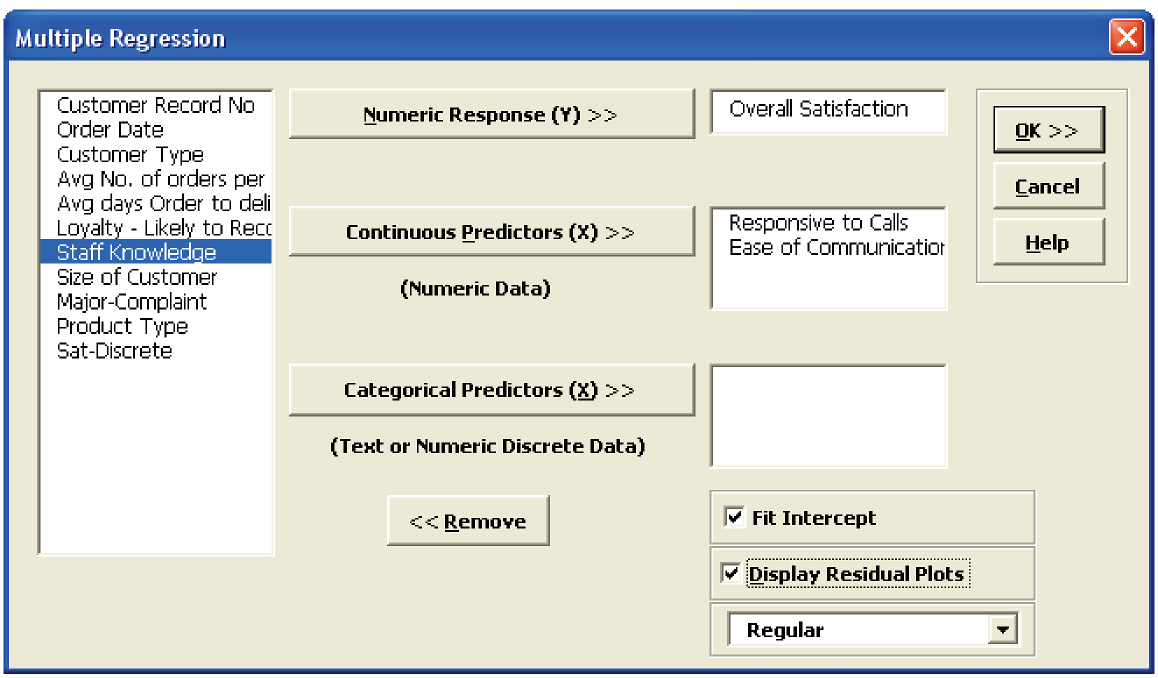
Multiple Regression analyzes the relationship between one dependent variable (Y) and multiple independent variables (X's). It is used to discover the relationship between the variables and create an empirical equation of the form:

Y = b0 + b1*X1 + b2*X2 + ... + bn*Xn

This equation can be used to predict a Y value for a given set of input X values. SigmaXL uses the method of least squares to solve for the model coefficients and constant term. Statistical tests of hypothesis are provided for the model coefficients.

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Regression > Multiple Regression**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**, select *Responsive to Calls* and *Ease of Communications*, click **Continuous Predictors (X) >>**. **Fit Intercept** and **Display Residual Plots** are checked by default.



Note: Fit Intercept is optional. If unchecked the model will not fit an intercept (constant) for the model. This may be useful when you have theoretical reasons to believe that Y = 0 when the X or X's are equal to 0 and the relationship is linear.

3. Click **OK**. The resulting Multiple Regression report is shown:

Multiple Regression Model: Overall Satisfaction = (0.493463) + (0.435673) * Responsive to Calls + (0.433346) * Ease of Communications

Model Summary:

R-Square	90.08%
R-Square Adjusted	89.88%
S (Root Mean Square Error)	0.248942

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	0.493463	0.116857	4.223	0.0001		
Responsive to Calls	0.435673	0.023710993	18.374	0.0000	1.168	0.856307
Ease of Communications	0.433346	0.029667131	14.607	0.0000	1.168	0.856307

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	2	54.601	27.301	440.53	0.0000
Error	97	6.011	0.061972154		
Total (Model + Error)	99	60.612	0.612246		

Durbin-Watson Test for Autocorrelation in Residuals:

DW Statistic	1.815
P-Value Positive Autocorrelation	0.1744
P-Value Negative Autocorrelation	0.8209

This model of Overall Satisfaction as a function of Responsiveness to Calls and Ease of Communications looks very good with an R-Square value of 90%. Both Predictors are shown to be significant with their respective P-Values < .05. Clearly, with the positive coefficient values, we need to focus on increasing these two X factors in order to improve customer satisfaction.

4. The variance inflation factor (VIF) and Tolerance scores are used to measure multicollinearity or correlation among predictors. VIF = 1 indicates no relation among predictors (which is highly desirable – you will see examples of this in Design of Experiments); VIF > 1 indicates that the predictors are correlated; VIF > 5 indicates that the predictors are strongly correlated and this will lead to poor estimates of the coefficients. Tolerance = 1/VIF.

5. The Durbin-Watson Test is used to determine if the residuals are Lag 1 autocorrelated. If either P-Value is < .05, then there is significant autocorrelation. This is likely due to an external bias factor affecting your process over time (e.g. "warm-up" effect, seasonality). This will also be evident in the plot of residuals versus observation order. Autocorrelation in the residuals is not a problem for this model.

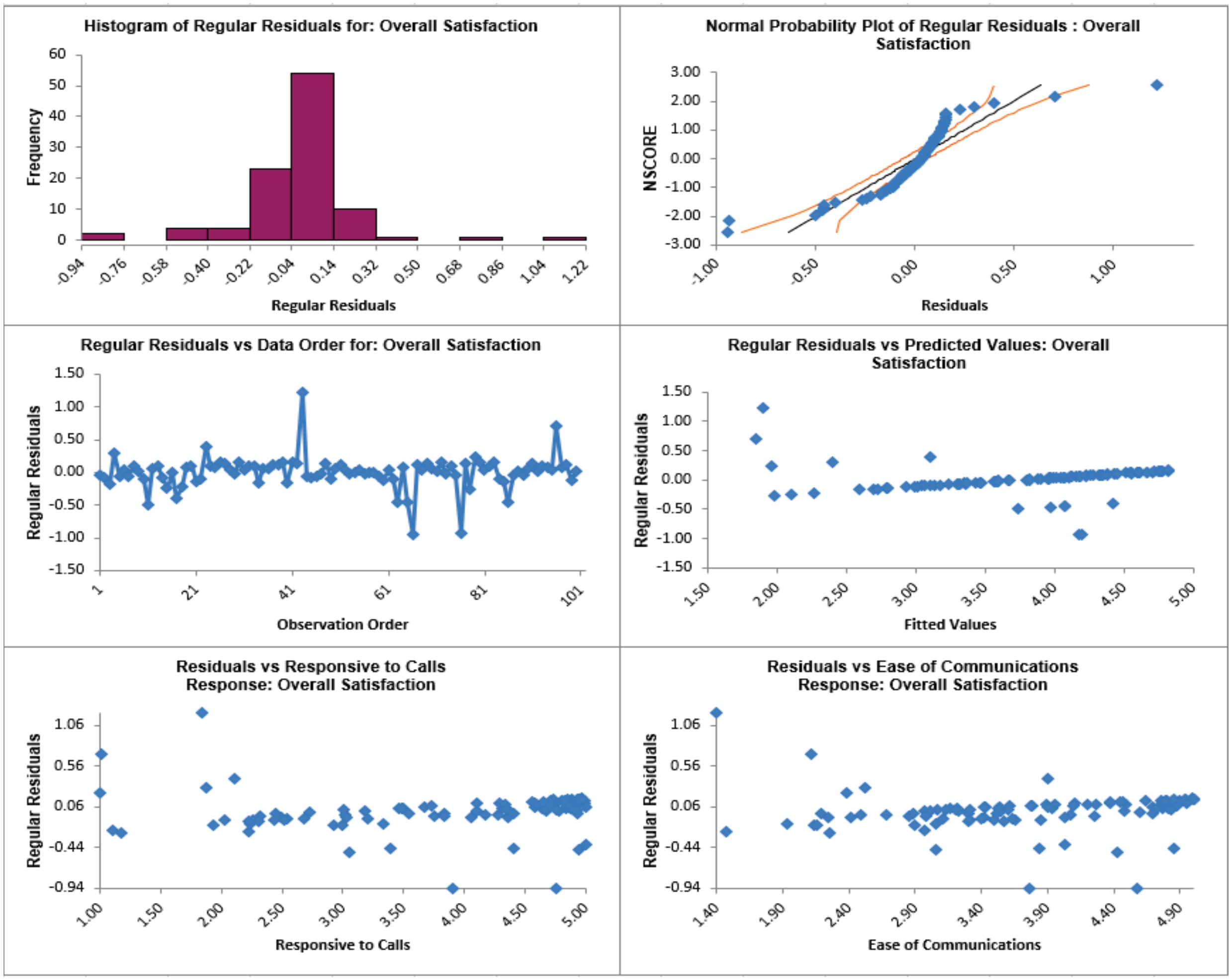
6. If the Fit Intercept option is unchecked, the following diagnostics are not reported due to statistical issues associated with the removal of the constant. These are: R-Square, R-Square Adjusted, VIF (Variance Inflation Factor) and Tolerance collinearity diagnostics, DW (Durbin-Watson) autocorrelation report, Anova report for Discrete factors (discrete factors will not be permitted when "Fit Intercept" is unchecked). All other tests including residual diagnostics are reported. Users can run the model with the "Fit Intercept" on to get the above statistics and then run with "Fit Intercept" off.

7. A predicted response calculator allows you to enter X values and obtain the predicted response value, including the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Responsive to Calls	5	4.838555471	4.750554344	4.926556597	4.336698591	5.34041235
Ease of Communications	5					

8. Click the **Mult Reg Residuals** Sheet and scroll right to view the Residual Plots shown below:

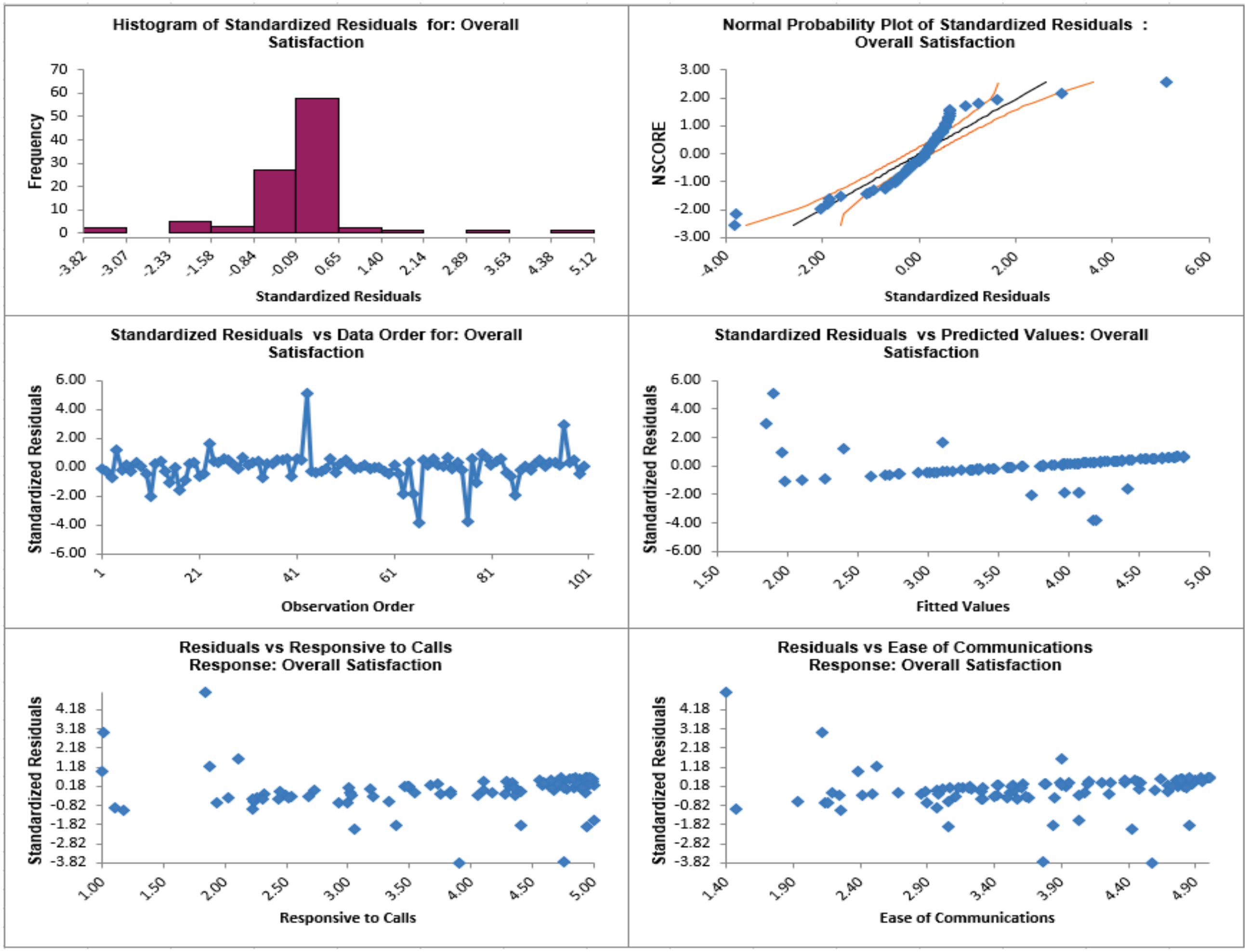


Residuals are the unexplained variation from the regression model (Y - Ŷ). We expect to see the residuals normally distributed with no obvious patterns in the above graphs. Clearly this is not the case here, with the Residuals versus Predicted Values indicating there is likely some other X factor influencing the Overall Satisfaction. It would be appropriate to consider other factors in the model.

9. SigmaXL also provides Standardized Residuals and Studentized (Deleted t) Residuals. The standardized residual is the residual, divided by an estimate of its standard deviation. This makes it easier to detect outliers. Standardized residuals greater than 3 and less than -3 are considered large (these outliers are highlighted in red). Studentized deleted residuals are

computed in the same way that standardized residuals are computed, except that the i^th^ observation is removed before performing the regression fit. This prevents the i^th^ observation from influencing the regression model, resulting in unusual observations being more likely to stand out.

10. Click **Recall Last Dialog** (or press **F3**). Change the **Residual** type to *Standardized*. Click **OK**. The resulting Standardized Residual plots are displayed:

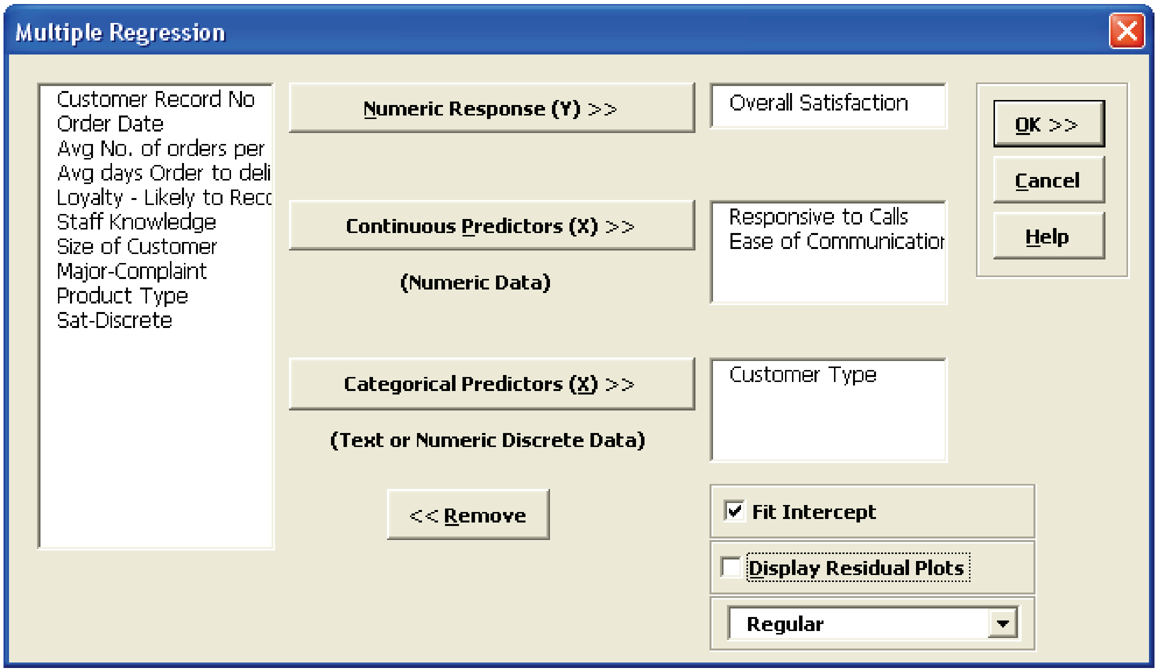


11. Other diagnostic measures reported but not plotted include Leverage, Cook's Distance (Influence), and DFITS. Leverage is a measure of how far an individual X predictor value deviates from its mean. High leverage points can **potentially** have a strong effect on the estimate of regression coefficients. Leverage values fall between 0 and 1. Cook's distance and DFITS are overall measures of influence. An observation is said to be influential if removing the observation substantially changes the estimate of coefficients. Cook's distance can be thought of as the product of leverage and the standardized residual squared; DFITS as the product of leverage and the studentized residual.

12. The easiest way to identify observations with high leverage and/or influence is to plot the measures on a run chart.

Multiple Regression with Continuous and Categorical Predictors

13. Click **Recall Last Dialog** (or press **F3**). Select *Customer Type*, click **Categorical Predictors (X) >>**. We will not discuss residuals further so you may wish to uncheck **Display Residual Plots**. Keep *Overall Satisfaction* as the **Numeric Response (Y)** and *Responsive to Calls* and *Ease of Communications* as the **Continuous Predictors (X)**:



14. Click **OK**. The resulting Multiple Regression report is shown:

Multiple Regression Model: Overall Satisfaction = (0.552345) + (0.427400) * Responsive to Calls + (0.409625) * Ease of Communications + (0.1

Model Summary:	
R-Square	90.58%
R-Square Adjusted	90.18%
S (Root Mean Square Error)	0.245199

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	0.552345	0.120148	4.597	0.0000		
Responsive to Calls	0.427400	0.023788018	17.967	0.0000	1.212	0.825379
Ease of Communications	0.409625	0.031120872	13.162	0.0000	1.325	0.754950
Customer Type_2	0.132728	0.063914154	2.077	0.0406	1.655	0.604180
Customer Type_3	0.023141785	0.065217411	0.354841	0.7235	1.394	0.717173

Analysis of Variance for Categorical (Discrete) Predictors:

Predictor Term	DF	SS	MS	F	P
Customer Type	2	0.299651	0.149825574	2.492000584	0.0881

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	4	54.901	13.725	228.29	0.0000
Error	95	5.712	0.060122608		
Total (Model + Error)	99	60.612	0.612246		

Durbin-Watson Test for Autocorrelation in Residuals:

DW Statistic	1.730
P-Value Positive Autocorrelation	0.0888
P-Value Negative Autocorrelation	0.9137

15. The parameter estimates table now includes Customer Type 2 and Customer Type 3, but where is Customer Type 1? Since Customer Type is a discrete predictor, SigmaXL applies "dummy coding" and the first alphanumerically sorted value, Customer Type 1, becomes the "hidden" or reference value. Another possible coding scheme is based on -1, 0, +1 instead of 0, 1, but the advantage of a 0, 1 coding scheme is the relative ease of interpretation when making predictions with the model.

16. In the **Predicted Response Calculator** enter the settings as shown:

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Responsive to Calls	5	4.870198931	4.778492731	4.961905131	4.374854056	5.365543806
Ease of Communications	5					
Customer Type_2	1					
Customer Type_3	0					

Note that to select Customer Type 1, you would enter **Customer Type 2** = 0 and **Customer Type 3** = 0; for Customer Type 2, as shown, you entered **Customer Type 2** = 1 and **Customer Type 3** = 0; for Customer Type 3 you would enter **Customer Type 2** = 0 and **Customer Type 3** = 1.

17. Note the addition of the Analysis of Variance for Categorical (Discrete) Predictors. The Customer Type P-Value is .09, so we do not have strong evidence to keep this term in the model. However, many practitioners will use an alpha value of 0.1 as a criterion for removal. You are probably wondering why the P-Value for Customer Type is not lower given the results we saw earlier using ANOVA. The change in P-Value is due to the inclusion of *Responsive to Calls* and *Ease of Communications* in the model. We have higher scores for *Ease of Communications* and *Responsive to Calls* with Customer Type 2. Statistically, Customer Type is somewhat correlated to *Responsive to Calls* and *Ease of Communications* (VIF for Customer Type = 1.66 and 1.39).

Part U – Advanced Multiple Regression

Summary of Features in Advanced Multiple Regression

- **Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**
    - Standardization and coding of continuous predictors
    - Option to display regression equation with unstandardized coefficients
    - (1, 0) or (-1,0,+1) coding of categorical predictors
    - Box-Cox Transformation
    - Specify confidence level
    - Residual Plots: Regular, Standardized and Studentized – Deleted t
    - Diagnostic measures: Cook's Distance (Influence), Leverage and DFITS
    - Storage of model design matrix
    - Main Effects and Interaction Plots (Fitted Means)
    - Contour and Surface Plots
    - Optimization with optional constraints including integer continuous
    - Automatic removal of extreme VIF or collinear terms (with alias and removal report)
    - Specify interactions, quadratic and higher orders (all interactions or up to 3-Way)
    - ANOVA Type I and/or Type III Sum-of-Squares with Pareto of Percent Contribution and Standardized Effects
    - Lenth Pseudo Standard Error for Saturated Models (Orthogonal or Non-Orthogonal) with Monte Carlo or Student T P-Values
    - Specify Test/Withhold Sample for R-square Test & StDev Test Validation
    - R-Square Predicted (Leave-One-Out Cross Validation)
    - R-Square K-Fold & StDev K-Fold (K-Fold Cross Validation)
    - Test for Constant Variance: Breusch-Pagan. Anderson-Darling Normality test is applied to residuals in order to automatically select Normal or Koenker (Robust) version. Report includes the Overall test and Individual predictors as well.
    - White robust standard errors for non-constant variance (Heteroskedasticity-Consistent)
    - Durbin-Watson test for autocorrelation in residuals with P-Values
    - Newey-West robust standard errors for non-constant variance with autocorrelation (Heteroskedasticity and Autocorrelation-Consistent)
    - White or Newey-West automatically selected based on Durbin-Watson P-Values
    - Box-Tidwell Test and Power Transformation Recommendation for Continuous Predictors
    - Stepwise/Best Subsets Regression:
        - Forward/Backward with alpha-to-enter, alpha-to-remove
        - Forward Selection with alpha-to-enter
        - Backward Elimination with alpha-to-remove
        - Forward, Backward Criterion: Minimize AICc, BIC; Maximize R-Square Adjusted, R-Square Predicted, R-Square K-Fold

        - Best Subsets utilizes the powerful MIDACO Solver (Mixed Integer Distributed Ant Colony Optimization) to solve best subsets with up to hundreds of continuous or categorical variables, including interactions and higher order terms. This feature gives SigmaXL a significant advantage over competitors with Best Subsets limited to 30 continuous variables.
        - Best Subsets Criterion: Minimize AICc, BIC; Maximize R-Square Adjusted
        - Hierarchical option
        - Detailed report with additional statistics such as Condition Number and Mallows' Cp.

- **Statistical Tools > Advanced Multiple Regression > Multiple Response Optimization**
    - Multiple Response Optimization with Desirability
        - Multistart Nelder-Mead Simplex
        - MIDACO

For details on the statistical methods, formulas and references, see the Appendix: **Advanced Multiple Regression**.

Multiple Response Optimization is introduced in Design of Experiments: **Part F – Multiple Response Optimization with Advanced Multiple Regression**

For further reading, the following books are recommended:

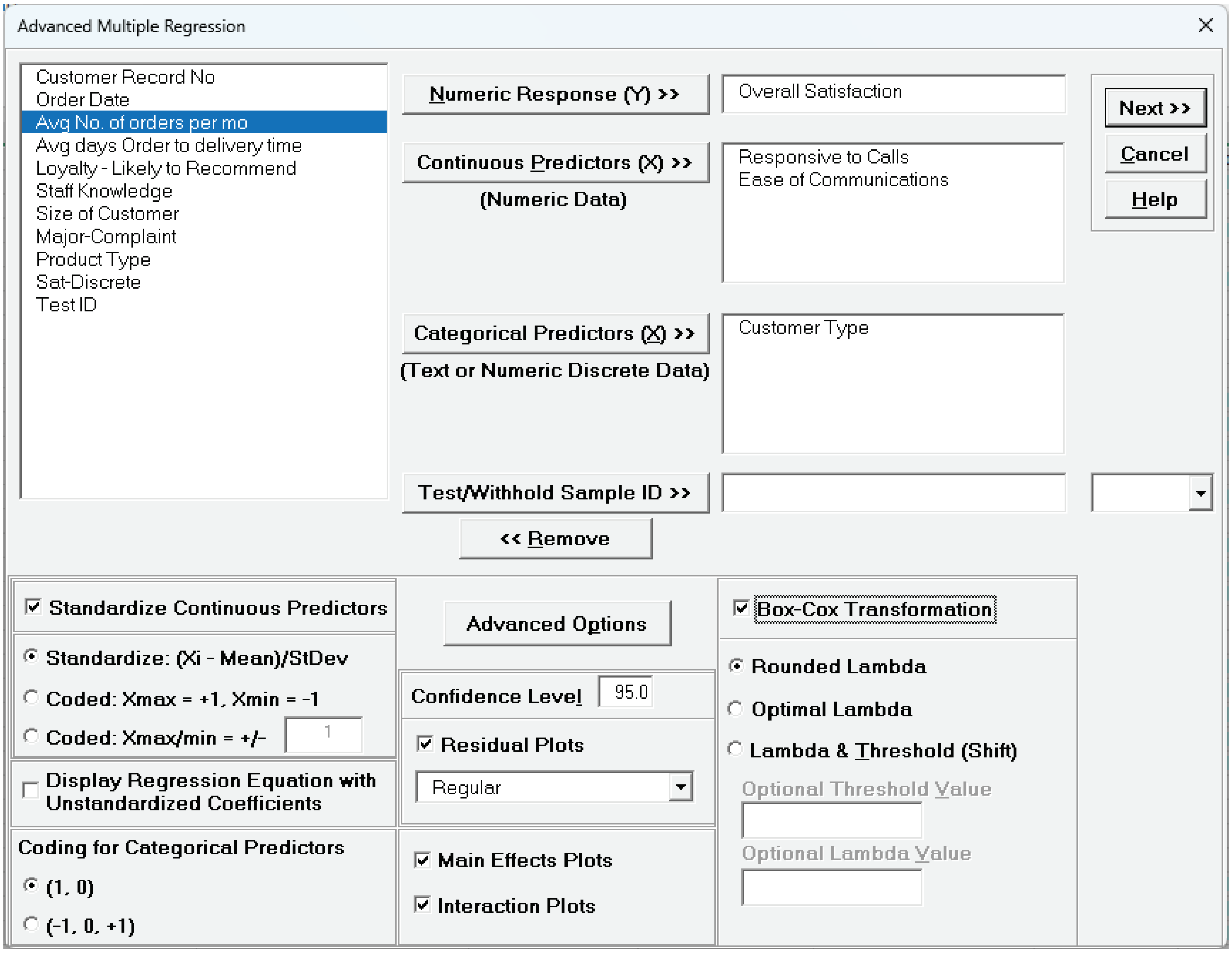
Frost, Jim (2019). *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*, Statistics By Jim Publishing. This is an excellent introduction to multiple regression that is concise, clear and inexpensive.

Montgomery, D.C. , E.A. Peck and G.G. Vining (2021). *Introduction to Linear Regression Analysis*, 6th Edition, John Wiley & Sons. This is an intermediate level book.

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*, Third Edition. Sage. This book includes advanced topics such as the White and Newey West robust standard errors and the Box-Tidwell test.

Advanced Multiple Regression Dialogs and Options

Fit Multiple Regression Model Dialog



- **Numeric Response** – select the response variable. Only one response may be selected at a time, but use **Recall SigmaXL Dialog** or Press **F3** to repeat an analysis with different options or to select a different response. The regression reports will be created on sheets MReg1 – Model *Y1name*, MReg2 - Model *Y2name*, etc., but truncated to fit the 31-character limit for Excel sheet names.

- **Continuous Predictors** – select continuous numeric predictors. Selections with data as text are error trapped. Note that the character "*" cannot be in the predictor name as this is used to denote a cross product term and will be error trapped.

- **Categorical Predictors** – select categorical predictors. Numeric predictors can be used but they will be converted to text and an underscore "_" will be appended to the number. If there are more than 50 unique levels, a warning message is given. Typically, this occurs when the user has incorrectly selected a continuous predictor as categorical.

- **Test/Withhold Sample ID** splits the data into a training and test/withhold sample for validation. To create a Test ID column with random 0/1 values, use the Excel function =IF(RAND()<=0.3,1,0), where 0.3 is the fraction desired for the test/withhold sample, 0 denotes training data, and 1 denotes test data. (Be sure to copy/paste values to freeze the random 0/1 numbers). The combo drop down is used to specify what rows are assigned to the test/withhold sample. Note, if a response value is missing from the test/withhold sample, a predicted response value will still be given in the Test report. Use of **Test/Withhold Sample ID** is recommended for large datasets (N >= 1000).

- **Standardize Continuous Predictors** with **Standardize: (Xi – Mean)/Stdev** will convert continuous predictors to Z-scores. This has two benefits: the predictors are scaled to the same units so coefficients can be meaningfully compared and it dramatically reduces the multicollinearity VIF scores when interactions and/or quadratic terms are specified.

- **Standardize Continuous Predictors** with **Coded: Xmax = +1, Xmin = -1** scales the continuous predictors so that Xmax is set to +1 and Xmin is set to -1. This is particularly useful for analyzing data from a full or fractional-factorial design of experiments.

- **Standardize Continuous Predictors** with **Coded: Xmax/Xmin = +/-** *value* scales the continuous predictors so that Xmax is set to +*value* and Xmin is set to -*value*. This is particularly useful if one is analyzing data from a response surface design of experiments, where *value* is set to the alpha axial value such as 1.414 for a two-factor rotatable design.

- **Display Regression Equation with Unstandardized Coefficients** displays the prediction equation with unstandardized/uncoded coefficients but the Parameter Estimates table will still show the standardized coefficients. This format is easier to interpret since there is only one coefficient value for each predictor.

- **Coding for Categorical Predictors (1, 0)** is the standard dummy coding used for categorical predictors, with the hidden reference value being the first alpha-numerically sorted level.

- **Coding for Categorical Predictors (-1, 0, +1)** is a coding scheme suitable for categorical predictors when the continuous predictors are **Coded: Ymax = +1, Ymin = -1**. For two levels, the coefficients are magnitude consistent with the continuous predictors. The hidden reference level is the last alpha-numerically sorted level.

- **Confidence Level** is used to determine what alpha value is used to highlight P-Values in red, the significance reference line in the Pareto Chart of Standardized Effects, and the percent confidence and prediction interval used in the Predicted Response Calculator.

- **Residual Plots** *Regular* display the raw residuals ($Y - \hat{Y}$) with a Histogram, Normal Probability Plot, Residuals vs Data Order, Residuals vs Predicted Values, Residuals vs Continuous Predictors and Residuals vs Categorical Predictors.

- **Residual Plots** *Standardized* display the residuals, divided by an estimate of its standard deviation. This makes it easier to detect outliers. Standardized residuals greater than 3 and less than -3 are considered large (these outliers are highlighted in red).

- **Residual Plots** *Studentized (Deleted t)* display studentized deleted residuals which are computed in the same way that standardized residuals are, except that the i^{th} observation is removed before performing the regression fit. This prevents the i^{th} observation from influencing the regression model, resulting in unusual observations being more likely to stand out.

- The Residuals report is provided on a separate sheet and includes a table with all residual types to the left of the plots. Other diagnostic measures included, but not plotted are: Cook's Distance (Influence), Leverage and DFITS. Leverage is a measure of how far an individual X predictor value deviates from its mean. High leverage points can *potentially* have a strong effect on the estimate of regression coefficients. Leverage values fall between 0 and 1. Cook's distance and DFITS are overall measures of influence. An observation is said to be influential if removing the observation substantially changes the estimate of model coefficients. Cook's distance can be thought of as the product of leverage and the standardized residual squared; DFITS as the product of leverage and the studentized residual. These diagnostic measures can be manually plotted using a Run Chart to identify unusually large values. Commonly used rough cutoff criterion for Cook's distance are: > 0.5, potentially influential and > 1, likely influential. A more accurate cutoff is the median of the F distribution: $> \mathrm{F}(0.5, p, n-p)$, where n is the sample size and p is the number of terms in the model design matrix, including the constant. A commonly used cutoff criterion for the absolute value of DFITS is: $> 2\sqrt{p/n}$. An observation that is an outlier and influential should be examined for measurement error or possible assignable cause. You could also try refitting the model excluding that observation to assess the influence.

- The Residuals report also includes a table to the right of the plots with the stored model design matrix and residuals. This can be used to manually create additional residual plots such as residuals versus interaction or quadratic terms.

- **Tip:** For large datsets (> 1K) you may want to uncheck the **Residual Plots** in order to speed up the analysis.

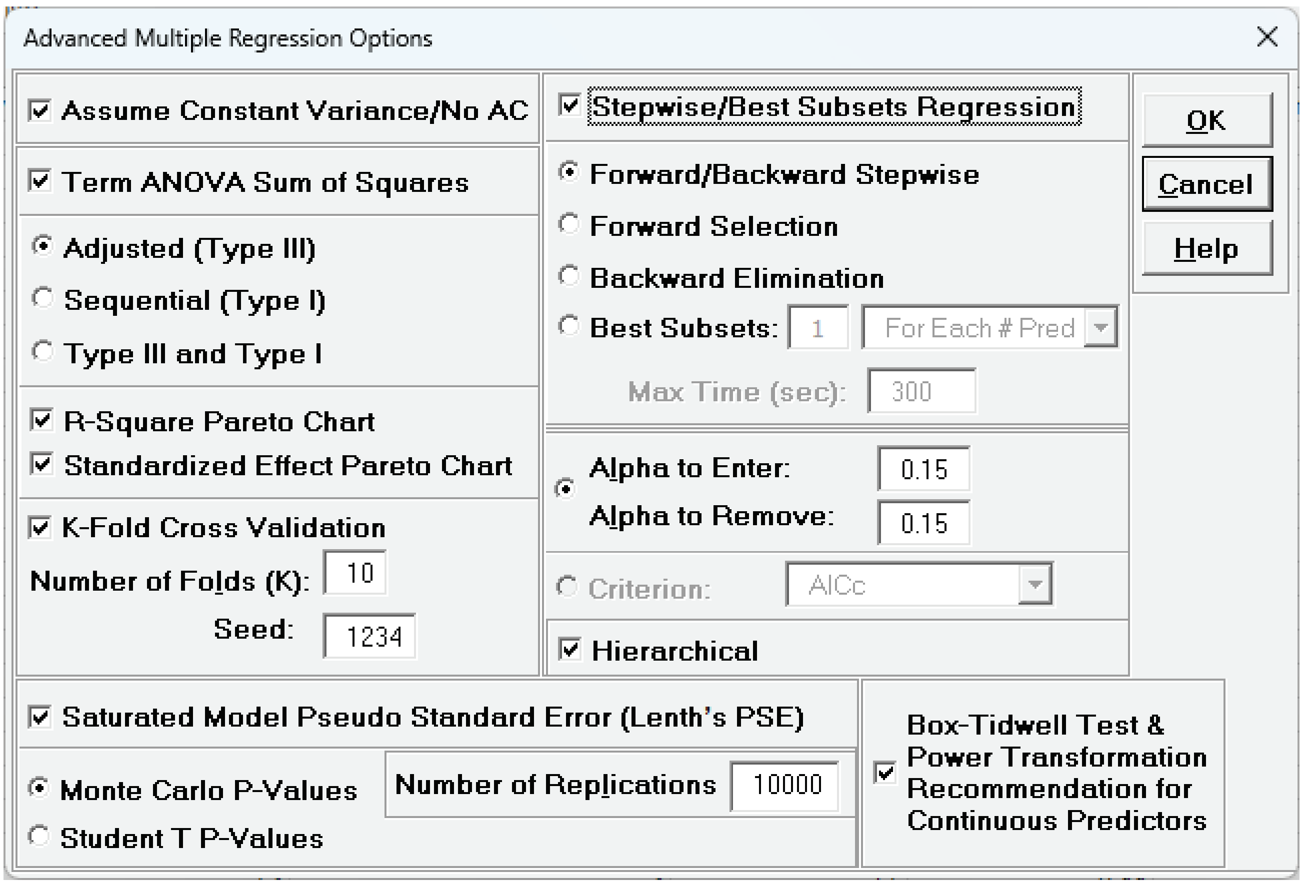
- **Main Effects Plots** and **Interaction Plots** use fitted means, not data means. If an interaction term is not in the model, the interaction plot is still displayed, but it is shaded grey.

- **Box-Cox Transformation** with **Rounded Lambda** will solve for an optimal lambda and is rounded to the nearest value of: -5, -4, -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 5. A 0 denotes a Ln(Y) transformation, 0.5 is the SQRT(Y), and 1 is untransformed. Threshold (Shift) is computed automatically if the response data includes 0 or negative values, otherwise it is 0. Note that the threshold is subtracted from the data so the value will be negative in order to provide positive response values. Solving lambda is also supported in Stepwise Regression. The reported Parameter Estimates, Model Summary, Information Criteria, Validation, Test Statistics and Residuals are for the Box-Cox transformed response. The Predicted Response Calculator automatically applies an inverse transformation so that the predicted response, confidence and prediction intervals are given in the original untransformed units. Note, Lambda is solved to normalize the regression residuals, not the raw data. It is solved using the classical Box-Cox formula but the actual transformation uses a simple power transformation.

- **Box-Cox Transformation** with **Optimal Lambda** uses the range of -5 to +5 for Lambda. Threshold is computed automatically if the response data includes 0 or negative values.

- **Box-Cox Transformation** with **Lambda & Threshold (Shift)** allows the user to specify Lambda and Threshold. Threshold is typically 0, but if the response data includes 0 or negative values, a negative threshold value should be entered, such that when subtracted from the data, results in positive response values.

Advanced Multiple Regression Options Dialog



- **Assume Constant Variance/No AC** (no autocorrelation in the residuals), if unchecked, SigmaXL will use either White robust standard errors for non-constant variance or Newey-West robust standard errors for non-constant variance with autocorrelation. If either of the Durbin-Watson P-Values are < .05 (i.e., significant positive or negative autocorrelation), Newey-West for Lag 1 is used, otherwise White HC3 is used. This will affect SE Coefficients, P-Values, ANOVA F and P-Values, and Prediction CI/PI. ANOVA F and P-Values are Wald estimates. ANOVA SS Type I Table and Pareto Charts are not available. Note: Stepwise P-Values are not adjusted.

- **Term ANOVA Sum of Squares** with **Adjusted (Type III)** provides a detailed ANOVA table for continuous and categorical predictors. Adjusted Type III is the reduction in the error sum of squares (SS) when the term is added to a model that contains all the remaining terms. Note, categorical terms are considered as a group, unlike the parameter estimates table which uses coding.

- **Term ANOVA Sum of Squares** with **Sequential (Type I)** provides a detailed ANOVA table for continuous and categorical predictors. Sequential Type I is the reduction in the error sum of squares (SS) when a term is added to a model that contains only the terms before it. This is affected by the order that they are entered in the model, so the user must be careful to specify model terms in the order of importance based on process knowledge. Note, if the terms are orthogonal then Type III and Type I SS will be the same.

- **R-Square Pareto Chart** displays a Pareto chart of term R-Square values ($100*SS_{term}/SS_{total}$). A separate Pareto Chart is produced for Type III and Type I SS. If there is only one predictor term, a Pareto Chart is not displayed.

- **Standardized Effect Pareto Chart** displays a Pareto chart of term T values (=T.INV(1-P/2,df_{error})). A separate Pareto Chart is produced for Type III and Type I SS. A significance reference line is include (=T.INV(1-alpha/2,df_{error})).

- **K-Fold Cross Validation:** In K-Fold cross-validation, the data is randomly partitioned into K (approximately equal) subsets. The model coefficients are estimated using K-1 partitions, i.e., (100*(K-1)/K)% of the data – the training set, and then statistical metrics are evaluated on the remaining data – the validation set. This is repeated for each of the K-Fold validation sets with R-Square K-Fold and S (Standard Deviation) K-Fold calculated as an average across the K samples, which results in a more accurate estimate of model prediction performance. The default K=10 is a popular choice, but some practitioners prefer K=5. Note that the final model parameter coefficients are based on all of the data, so K-Fold Cross Validation is used strictly to obtain R-Square K-Fold and S K-Fold. The fixed seed allows for replicable results, but the user may wish to re-run the analysis with a different seed a few times to see how much variation occurs in R-Square K-Fold and S K-Fold. If categorical predictors are used and the training sample does not include all of the levels, the K-Fold statistics cannot be computed.

- **Stepwise/Best Subsets Regression** with **Forward/Backward Stepwise:** Starting with an empty model, terms are added or removed from the model, one at a time, until all variables in the model have p-values that are less than (or equal to) the specified **Alpha-to-Remove** and all variables that are not in the model have p-values greater than the specified in **Alpha-to-Enter**. The stepwise process either adds the term which is most significant (largest F statistic, smallest p-value), or removes the term that is least significant (smallest F statistic, largest p-value). It does not consider all possible regression models. The independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Forward Selection:** Starting with an empty model, the most significant terms are added to the model, one at a time, until all variables that are not in the model have p-values greater than the specified in **Alpha-to-Enter**. Terms that are in the model are not removed regardless of p-value. Alternatively, criterion-based selection may be used. The most significant terms are added, one at a time, while at each stage the value of a measure, such as AICc or R-Square is monitored. If a minimum AICc is observed at step i, and this remains the minimum after 10 additional steps (or the model includes all terms), then the model at the minimum AICc is selected. If a maximum R-Square is observed at step i, and this remains the maximum after 10 additional steps, then the model with the maximum R-Square is selected. **Criterion** options are: *AICc*, *BIC*, *R-Square Adjusted*, *R-Square Predicted* and *R-Square K-Fold*. AICc is the Akaike Information Criterion corrected for small sample sizes, BIC is the Bayesian Information Criterion. For details on these metrics, see the Appendix: **Advanced Multiple Regression**. Note that for *R-Square K-Fold*, the F-statistic to decide which term to enter is based on all of the data. The K-Fold model is computed using the specified model, but a subset of the data is used as training data to estimate parameters and R-square is calculated using the out-of-sample validation data. As with forward/backward stepwise, the independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Backward Elimination:** Starting with all terms in the model, the least significant terms are removed from the model, one at a time, until all variables in the model have p-values that are less than (or equal to) the specified **Alpha-to-Remove**. Terms that are removed from the model are not entered again regardless of p-value. Alternatively, criterion-based selection may be used, as described above, but the least significant terms are removed, one at a time. It stops after 10 additional steps or if the model includes only one term. As with forward/backward stepwise, the independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Best Subsets:** With Best Subsets, for any given model with p terms, there are $2^p - 1$ possible combinations (non-hierarchical models). A criterion such as AICc is specified, and the model which results in the minimum AICc is selected. If $p \leq 15$, all possible combinations are explored - this is called exhaustive. Otherwise, the best model is derived using discrete optimization with the powerful MIDACO Solver (Mixed Integer Distributed Ant Colony Optimization). Start values are obtained using forward selection with the AICc criterion. MIDACO does not guarantee a best solution as we have in exhaustive, but will be close to best, even for hundreds of terms! Best Subsets **Criterion** options are: *AICc*, *BIC* and *R-Square Adjusted*. *R-Square Predicted* and *R-Square K-Fold* are not feasible as criterion here due to the computation times, but they are reported on the best selected models. Best Subsets report options are: Best *For Each # of Pred* (default) or Best *Overall*. Best For Each # of Predictors provides the most information, but takes longer to compute than Best Overall. The user may specify how many models to include (per # predictors or overall) in the report, with the default = 1. The default **Max Time (sec)** = 300 is the maximum total computation time allotted for either option. The independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** – **Hierarchical:** The Hierarchical option constrains the model so that all lower order terms that comprise the higher order terms are included in the model. This is checked by default. In Forward/Backward Stepwise and Forward Selection, a hierarchical model is required at each step, but extra terms can enter to maintain hierarchy.

For Backward Elimination and Best Subsets, extra terms are not permitted.

- **Saturated Model Pseudo Standard Error (Lenth's PSE):** For saturated models with $df_{error} = 0$, Lenth's method is used compute a pseudo standard error. For each term, a t ratio is computed by dividing the coefficient by the PSE. Since the distribution of the t ratio is not analytic, the probability is evaluated using Monte Carlo simulation. Student T P-Values are also available for comparison purposes. Lenth's PSE in the SigmaXL DOE Templates and DOE Analysis use Student T P-Values.

- **Box-Tidwell Test and Power Transformation Recommendation for Continuous Predictors:** Multiple linear regression assumes that relationships between the predictors and the response variable are linear. The Box-Tidwell procedure aims to find an optimal power transformation of the predictor variables to satisfy the linearity assumption. This transformation can be crucial for improving the model fit and prediction accuracy. For details, see the Appendix: **Advanced Multiple Regression**. Note, in SigmaXL:

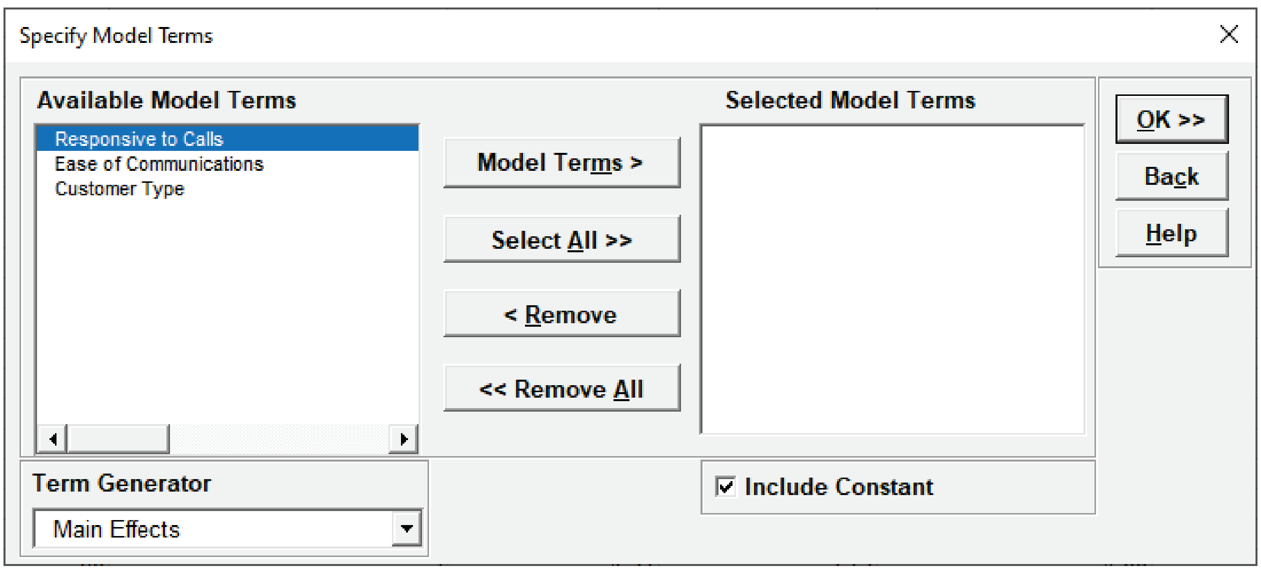
    - At least one continuous predictor with all positive data values must be included in the model.
    - Do not use standardization or coding as this will introduce 0 or negative values in the predictors.
    - Continuous predictors with values <= 0, categorical factors, interactions and higher order terms are included in the model but excluded from the Box-Tidwell (BT) test and transformation.
    - The constant must be included in the model.
    - Box-Tidwell cannot be performed with Error df = 0.
    - Box-Tidwell power transformations are calculated only for significant continuous predictors detected by the BT Test. This improves the overall robustness of the procedure.
    - If Box-Cox is used, Box-Tidwell power transformations are calculated using the Box-Cox transformed response. Box-Cox Lambda may not be optimal after refit.
    - Optimal and rounded power values are reported. Rounded is recommended for ease of interpretation.
    - Power values of -5 or +5 are limits and considered unstable, so rounded is set to 1.
    - Sheet BoxTidwell contains the original data with new columns for the transformed continuous predictors using rounded power. The model should be refit with these transformed predictors. If optimal power is desired, please use the Excel formula "=X^(Power)"; if Power = 0, use "=LN(X)".

Tip: There are a lot of options here, giving the user flexibility for model refinement, but this can also be overwhelming to someone starting out with these tools. We recommend using the following settings for **Stepwise/Best Subsets Regression**:

1. **Forward Selection**, **Criterion***: AICc, BIC or R-Square Predicted*, **Hierarchical** checked. This is fast and will build a model that minimizes AICc, BIC or maximizes R-Square Predicted. AICc or R-Square Predicted are recommended for the best model prediction accuracy, *BIC* is recommended for model parsimony. Note, however, this does not consider all possible models.

2. **Best Subsets**, *1 For Each # of Pred*, **Max Time (sec)** = 300, **Criterion:** *AICc* or *BIC*, **Hierarchical** checked. This can be slow, but gives a very useful report of the best model for each number of predictors in the model.

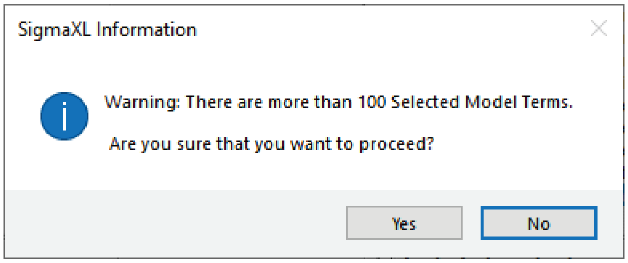
Specify Model Terms Dialog



- **Term Generator** – select any of the following to build a list of **Available Model Terms**:
  - *Main Effects* – default – no change to original specified terms.
  - *ME + 2-Way Interactions* – use this to include 2-way interactions in the model, for example, analyzing data from a Res IV or Res V fractional-factorial DOE. When specifying interactions or higher order terms, standardization of continuous predictors is highly recommended.
  - *ME + 2-Way Interactions + Quadratic* – use this to include 2-way interactions and quadratic terms in the model, for example, analyzing data from a response surface DOE. Categorical terms will not be squared.
  - *ME + All Interactions* – use this to include all possible interaction terms in the model, for example analyzing data from a full-factorial DOE.
  - *All up to 3-Way* – use this to include 2-way interactions, quadratic, 3-way interactions, quadratic*main effect and cubic terms in the model. Categorical terms will not be squared or cubed.

- **Model Terms:** Select from highlighted **Available Model Terms**.

- **Select All:** Select all **Available Model Terms.** Caution, the number of selected terms can become quite large, especially for the last two options in the **Term Generator**. If more than 100 terms are selected, a warning is given after clicking **OK**:



- **Include Constant:** If unchecked, the model will not fit a constant (intercept) for the model. This should only be used when you have strong *a priori* theoretical reasons to believe that Y = 0 when the X or X's are equal to 0 and the relationship is linear. Note, if this is not the case, R-Square values will be artificially inflated, so can be very misleading. If **Include Constant** is unchecked, the Breusch-Pagan Test for Constant Variance and Stepwise/Best Subsets Regression are not computed.

Tip: It is also important to ensure that the number of rows/observations are sufficient to estimate the number of selected model terms. A rule of thumb (excluding data from a designed experiment) is that for every term selected, there should be a minimum of 10 rows of data. This rule holds for potential terms used in stepwise and best subsets as well, otherwise one can easily produce a model that is highly significant but a meaningless model of noise. This is what Jim Frost calls "Data Dredging" in chapter 8 of his book *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*.

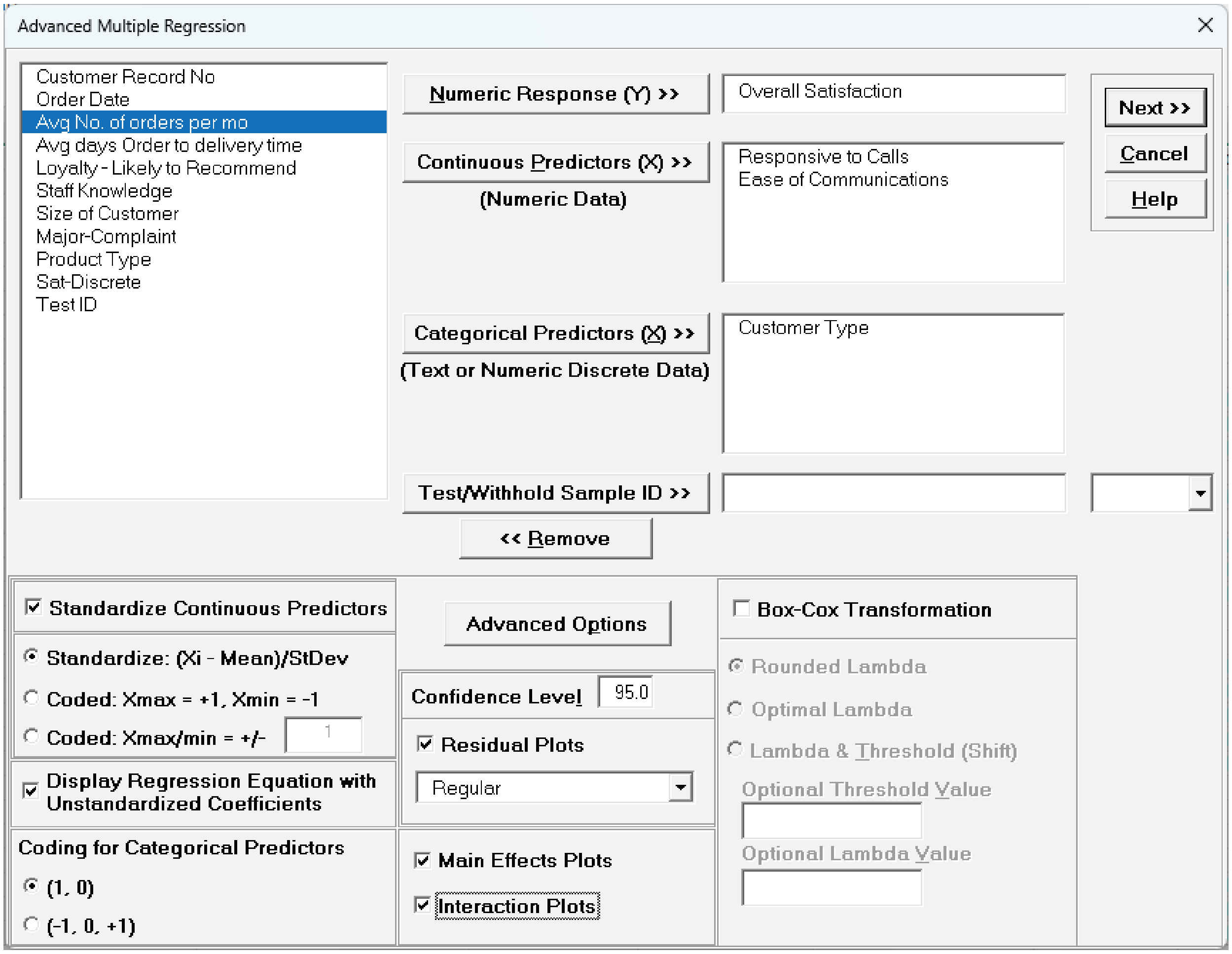
Example: Advanced Multiple Regression with Two-Way Interactions

We will revisit the previous regression analysis of *Overall Satisfaction* versus *Responsive to Calls*, *Ease of Communications* and *Customer Type* but now, using advanced multiple regression, we will:

- standardize the continuous predictors
- include two-way interactions
- discuss prediction equation with continuous and categorical predictors
- assess R-Square Predicted and R-Square K-fold for validation
- review the Parameter Estimates and ANOVA for Predictors report with Pareto of Percent Contribution and Pareto of Standardized Effects
- review the Durbin-Watson Test for Autocorrelation in Residuals and Breusch-Pagan Test for Constant Variance
- examine Residual Plots
- look at Main Effects and Interaction Plots with fitted means
- use Forward Selection with R-Square Predicted criterion to refine the model
- refit the model using Box-Cox Transformation
- find optimum values for predictors that maximize *Overall Satisfaction*
- create a Contour and Surface Plot to visualize the relationship of *Overall Satisfaction* versus *Responsive to Calls* and *Ease of Communications*
- demonstrate Best Subsets with AICc criterion
- demonstrate the use of Test/Withhold Sample ID

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, click **Use Entire Data Table**, click **Next**.

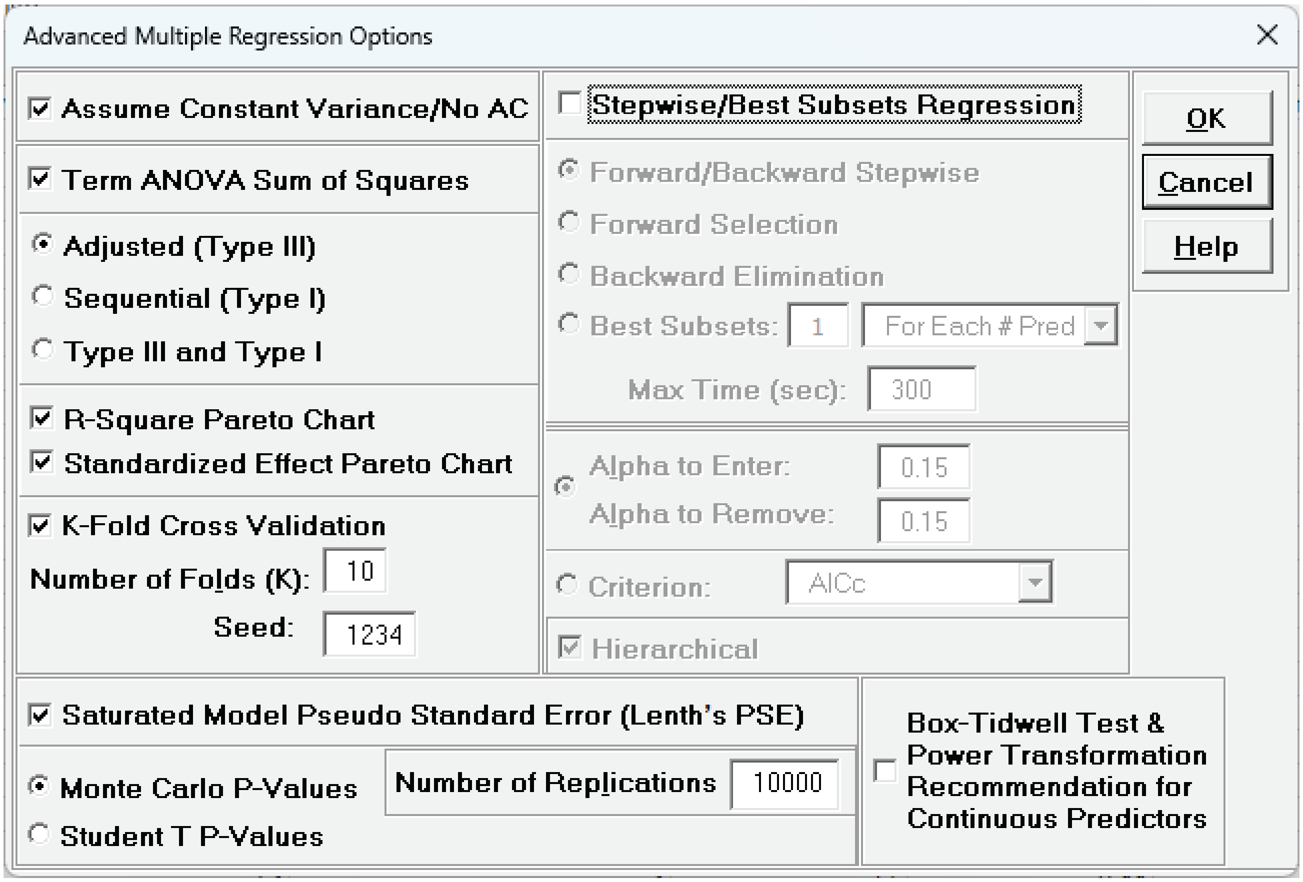
2. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**, select *Responsive to Calls* and *Ease of Communications*, click **Continuous Predictors (X) >>**, select *Customer Type*, click **Categorical Predictors (X) >>**. Check **Standardize Continuous Predictors** with default option **Standardize: (Xi-Mean)/StDev**. Check **Display Regression Equation with Unstandardized Coefficients**. We will use the default **Coding for Categorical Predictors (1,0).** Use the default **Confidence Level** = 95.0%. Regular **Residual Plots** are checked by default. Check **Main Effects Plots** and **Interaction Plots**. Leave **Box-Cox Transformation** unchecked for now - this will be used later. We will not use **Test/Withhold Sample ID** at this time, it will be demonstrated later.



Tip: If you are planning to include interaction terms in the model, always ensure that **Standardize Continuous Predictors** is checked. This has two benefits: the predictors are scaled to the same units so coefficients can be meaningfully compared and it dramatically reduces the multicollinearity VIF scores.

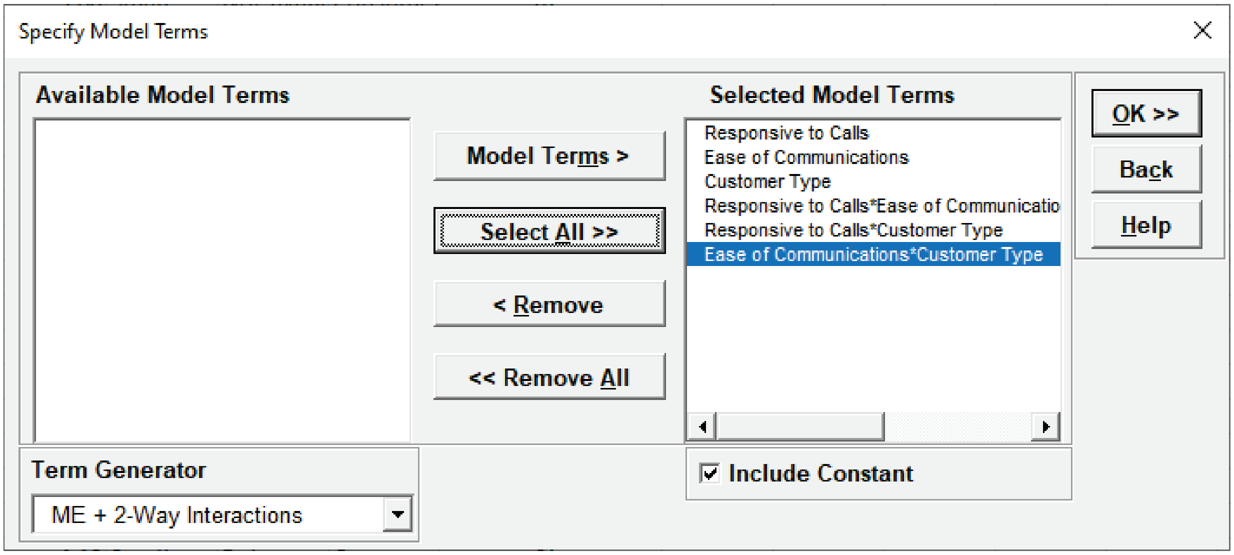
Tip: Check **Display Regression Equation with Unstandardized Coefficients** to display the prediction equation with unstandardized/uncoded coefficients. This format is easier to interpret with only one coefficient value for each predictor.

3. Click **Advanced Options.** Check **K-Fold Cross Validation** with default **Number of Folds (K)** = 10 and **Seed** = 1234. **Assume Constant Variance/No AC**, **Term ANOVA Sum of Squares** with **Adjusted (Type III)**, **R-Square Pareto Chart**, **Standardized Effect Pareto Chart** are checked by default. Keep **Stepwise/Best Subsets Regression** unchecked for now - this will be used later. **Saturated Model Pseudo Standard Error (Lenth's PSE)** is checked by default, but is not used here, as this is only applicable to saturated models with 0 error degrees of freedom.



4. Click **OK.** Click **Next >>.**

5. Using **Term Generator**, select *ME + 2-Way Interactions*. Click **Select All >>**. **Include Constant** is checked by default.



We are adding the three possible 2-way interactions. Note that the three interaction terms are: continuous*continuous, continuous*categorical and continuous*categorical.

6. Click **OK >>**. The Advanced Multiple Regression report is given.

7. The Regression Equation with unstandardized/uncoded coefficients is:

Multiple Regression Model (Uncoded): Overall Satisfaction = (1.38507)
+ (0.111481)*Responsive_to_Calls
+ (0.126645)*Ease_of_Communications
+ (0.505067)*(IF(Customer_Type="2_",1,0))
+ (0.821635)*(IF(Customer_Type="3_",1,0))
+ (0.101385)*Responsive_to_Calls*Ease_of_Communications
+ (-0.0184748)*Responsive_to_Calls*(IF(Customer_Type="2_",1,0))
+ (-0.0728492)*Responsive_to_Calls*(IF(Customer_Type="3_",1,0))
+ (-0.0997064)*Ease_of_Communications*(IF(Customer_Type="2_",1,0))
+ (-0.156844)*Ease_of_Communications*(IF(Customer_Type="3_",1,0))

Note blanks and special characters in the predictor names are converted to the underscore character "_". The numeric Customer Type 1, 2, 3 has also been converted to text so appears as "1_", "2_", "3_", where "1_" is the hidden reference level.

For categorical predictors, IF statements are used, but exclude the hidden reference level.

This is the display version of the prediction equation given at cell **L14** (which has more precision for the coefficients and predictor names are converted to legal Excel range names by padding with the underscore "_" character). If the equation exceeds 8000 characters (Excel's legal limit for a formula is 8192), then a truncated version is displayed and cell **L14** does not show the formula.

Note, the coefficients in the regression equation do not match those given in the Parameter Estimates table, since the table values are Standardized. If consistency is desired, one can always rerun the analysis with **Display Regression Equation with Unstandardized Coefficients** unchecked.

8. The Model Summary is:

Model Summary	
R-Square	92.88%
R-Square Adjusted	92.17%
R-Square Predicted	89.88%
S (Root Mean Square Error)	0.2190

R-Square = 92.88%, the percent variation explained by the model is quite good. R-Square Adjusted = 92.17%, which includes a penalty for the number of terms in the model design matrix, is also good, and S = 0.2190. The addition of the three interaction terms result in an improvement over the previous analysis with R-Square = 90.58%, R-Square Adjusted = 90.18% and S = 0.2452. R-Square Predicted = 89.88%, also known as Leave-One-Out Cross-Validation, indicates how well a regression model predicts responses for new observations and is typically less than R-Square Adjusted. This is also good.

9. The Model Information is given as:

Model Information	
Continuous Predictor Standardization/Coding	((Xi - Mean)/Stdev)
Categorical Predictor Coding	(1,0)
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

This summarizes the selected options for the model and, if applicable, will include Box-Cox Lambda and Threshold values.

10. The Information Criteria and Validation table is given as:

Information Criteria and Validation	
AICc	-5.4884
BIC	20.1685
R-Square 10-Fold	89.54%
S 10-Fold	0.2518

The information criteria AICc and BIC are reported here, but are used for model comparison. We will revisit these when we do model refinement with Forward Selection. R-Square 10-Fold = 89.54%, the R-Square from K-Fold Cross-Validation is good. It is approximately the same as the R-Square Predicted. S 10-Fold = 0.2518 while slightly higher than S as expected, is also good.

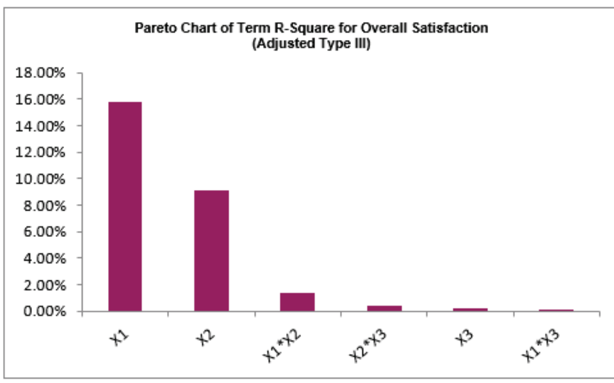
Repeating the analysis with a different K and/or seed will produce slightly different results for R-Square and S K-Fold.

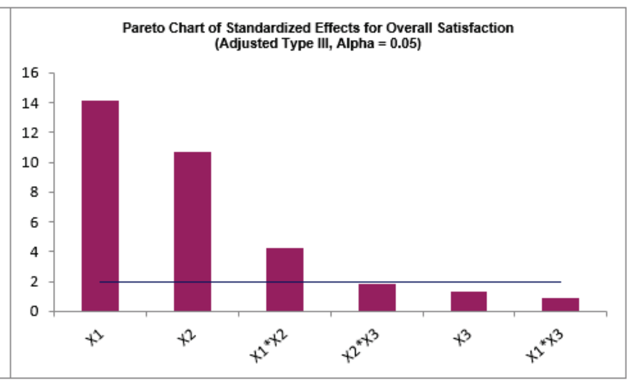
11. The Parameter Estimates – Standardized, Analysis of Variance for Model, Analysis of Variance for Predictors (Adjusted Type III) tables, Pareto of Term R-Square and Pareto of Standardized Effects are shown:

Parameter Estimates - Standardized						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	3.759039371	0.043723303	85.9734	0.0000		
Responsive to Calls	0.560435726	0.039622585	14.1444	0.0000	3.2408	0.3086
Ease of Communications	0.472484939	0.044071453	10.7209	0.0000	4.0094	0.2494
Customer Type_2_	0.05996351	0.059787508	1.0029	0.3186	1.8156	0.5508
Customer Type_3_	-0.04775003	0.06226883	-0.7668	0.4452	1.5935	0.6275
Responsive to Calls*Ease of Communications	0.105361601	0.024925911	4.2270	0.0001	1.5582	0.6418
Responsive to Calls*Customer Type_2_	-0.02106671	0.059332001	-0.3551	0.7234	2.1442	0.4664
Responsive to Calls*Customer Type_3_	-0.0830694	0.060658881	-1.3695	0.1743	1.8300	0.5465
Ease of Communications*Customer Type_2_	-0.09086854	0.063703126	-1.4264	0.1572	2.6544	0.3767
Ease of Communications*Customer Type_3_	-0.14294155	0.062520483	-2.2863	0.0246	2.3318	0.4288

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	9	56.29598423	6.255109359	130.4251	0.0000
Error	90	4.316346768	0.047959409		
Total (Model + Error)	99	60.612331	0.612245768		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Responsive to Calls	1	9.594886344	9.594886344	200.0626	0.0000	15.83%	14.1444
Ease of Communications	1	5.512327654	5.512327654	114.9374	0.0000	9.09%	10.7209
Customer Type	2	0.15808785	0.079043925	1.6481	0.1982	0.26%	1.2964
Responsive to Calls*Ease of Communications	1	0.856912444	0.856912444	17.8675	0.0001	1.41%	4.2270
Responsive to Calls*Customer Type	2	0.091194214	0.045597107	0.9507	0.3903	0.15%	0.8632
Ease of Communications*Customer Type	2	0.257370015	0.128685007	2.6832	0.0738	0.42%	1.8089





Pareto Legend
X1 = Responsive to Calls
X2 = Ease of Communications
X3 = Customer Type

The P-Values for continuous predictors are the same in the Parameter Estimates and ANOVA for Predictors tables. For categorical predictors and interactions involving a categorical predictor, the Parameter Estimates table gives a Coefficient and P-Value for each level, excluding the hidden reference level. The ANOVA for Predictors table gives an overall P-Value for each of the categorical terms.

Customer Type is insignificant as a main effect, but is potentially significant in the *Ease of Communications***Customer Type* interaction (Term P-Value = .0738).

The *Responsive to Calls*Customer Type* interaction is not significant (Term P-Value = 0.3903). However, in this example, we will not use P-Values to decide what terms are removed and what terms remain in the model, rather model refinement will be based on the criterion R-Square Predicted.

The Pareto Chart of Term R-Square and Pareto Chart of Standardized Effects graphically present the information given in the Predictor ANOVA table and show that *Responsive to Calls*, *Ease of Communications* and *Responsive to Calls*Ease of Communications* are the dominant contributors to the variability in Overall Satisfaction. Note that the Term R-Square values do not add up to R-Square = 92.88%, since these are Type III Adjusted Sum-of-Squares, but they are still useful to assess relative contribution. If we included Type I Sequential Sum-of-Squares, the Term R-Square values would add up to 92.88%, but have the disadvantage that the individual Term R-Square values depend on model order, whereas Type III do not. If the predictors are orthogonal, then Type III and Type I are the same.

The Variance Inflation Factor (VIF) scores are all < 5, so acceptable, but they are higher than those we saw in the previous analysis, so this is something that we will want to keep an eye on as we progress with model refinement.

12. The Durbin-Watson Test for Autocorrelation in Residuals table is:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	1.5800
P-Value Positive Autocorrelation	0.0157
P-Value Negative Autocorrelation	0.9827

The Durbin Watson (DW) test is used to detect the presence of positive or negative autocorrelation in the residuals at Lag 1. If either P-Value is < .05, then there is significant autocorrelation.

This may be evident in the plot of residuals versus observation order. We will look at the residuals plots shortly.

Note that DW does not test for higher order lags. If that is a concern, then use the Time Series Forecasting tools, particularly the Autocorrelation ACF/PACF Plots on the residuals and model using ARIMA with Predictors.

Here we do see significant positive autocorrelation, which is a problem because it violates the regression assumption of independence.

Tip: Note that this status may change with model refinement, as we will see this is the case in this example. If after model refinement, the Durbin Watson test still shows significant autocorrelation, then refit the model using **Recall Last Dialog**, click **Advanced Options** in the **Advanced Multiple Regression** dialog, and uncheck **Assume Constant Variance/No AC**. SigmaXL will apply the Newey-West (Lag 1) robust standard errors for non-constant variance with autocorrelation. For details, see the Appendix: **Advanced Multiple Regression**.

13. The Breusch-Pagan Test for Constant Variance is:

Breusch-Pagan Test for Constant Variance (Koenker Studentized - Robust)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	**Chi-Square**	**DF**	**P-Value**
All Terms	11.2222	9	0.2608
Responsive to Calls	1.97856	1	0.1595
Ease of Communications	1.22053	1	0.2693
Customer Type	3.55179	2	0.1693
Responsive to Calls*Ease of Communications	3.84677	1	0.0498
Responsive to Calls*Customer Type	4.2728	2	0.1181
Ease of Communications*Customer Type	0.700644	2	0.7045

There are two versions of the Breusch-Pagan (BP) test for Constant Variance: Normal and Koenker Studentized – Robust. SigmaXL applies an Anderson-Darling Normality test to the residuals in order to automatically select which version to use. If the AD P-Value < 0.05, Koenker Studentized – Robust is used.

The report includes the test for *All Terms* and for individual predictors. *All Terms* denotes that all terms are in the model. This should be used to decide whether or not to take corrective action. The individual predictor terms are evaluated one-at-a-time and provide supplementary information for diagnostic purposes. Note, this should always be used in conjunction with an examination of the residual plots.

Here we see that the *All Terms* test is not significant, but the *Responsive to Calls*Ease of Communications* interaction is significant.

Tip: As with the Durbin-Watson test, this status may change with model refinement. If the *All Terms* test is significant after model refinement, including a Box-Cox transformation, then refit the model using **Recall Last Dialog,** click **Advanced Options** in the **Advanced Multiple Regression** dialog, and uncheck **Assume Constant Variance/No AC**. SigmaXL will apply the White robust standard errors for non-constant variance. For details, see the Appendix: Advanced Multiple Regression.

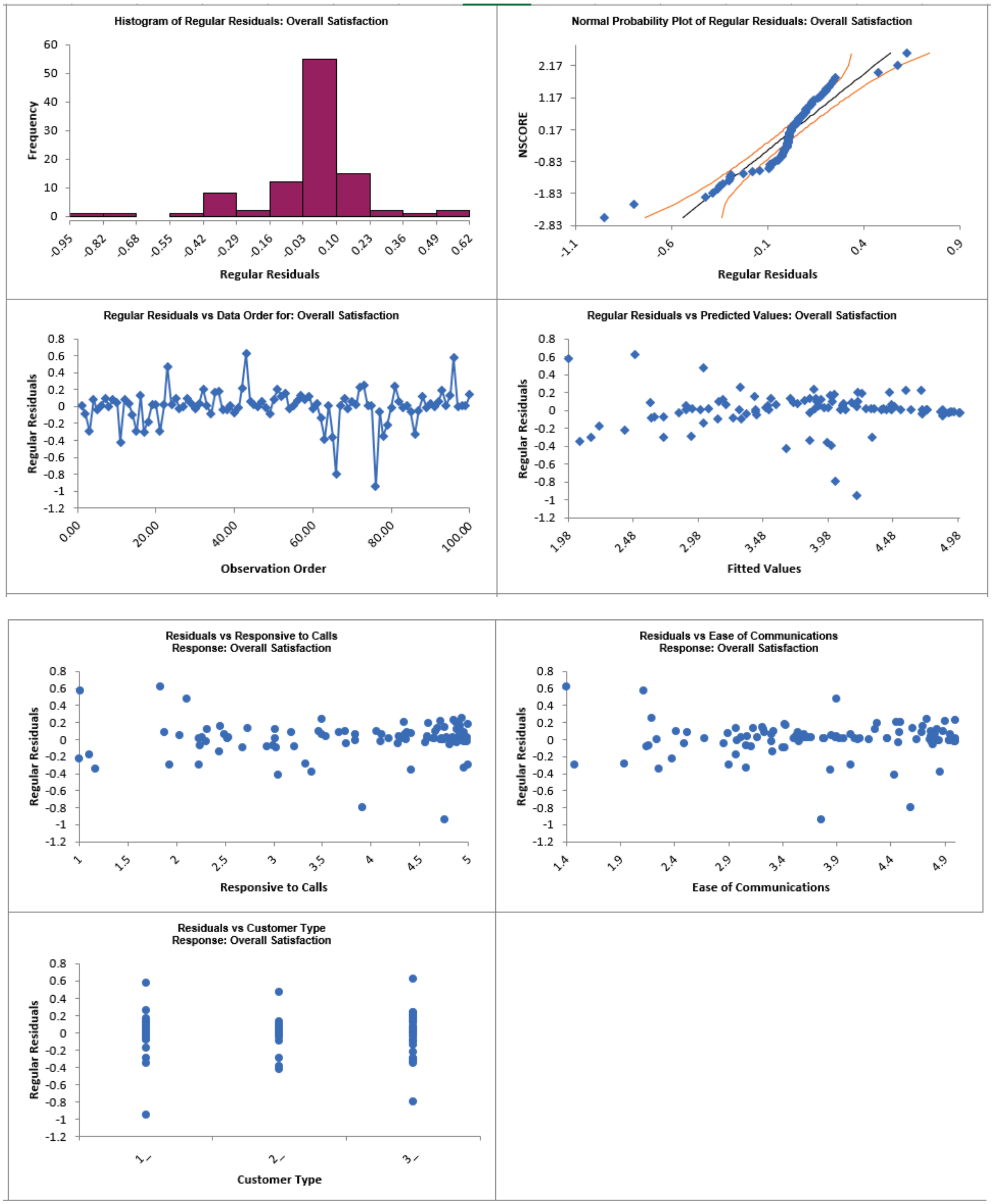
Tip: Lack of Constant Variance (a.k.a. Heteroskedasticity) is a nuisance for regression modelling but is also an opportunity. Examining the residual plots and BP individual predictors may yield process knowledge that identifies variance reduction opportunities.

14. Scroll to the Predicted Response Calculator. Enter *Responsive to Calls* and *Ease of Communication* values = 5 and select Customer Type = 2_ from the dropdown list to predict Overall Satisfaction including the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator							
Predictors	**Enter Settings:**	**Predicted Response**	**SE**	**Lower 95% CI**	**Upper 95% CI**	**Lower 95% PI**	**Upper 95% PI**
Responsive to Calls	5	5.024498619	0.059057617	4.907170354	5.141826883	4.573881545	5.475115693
Ease of Communications	5						
Customer Type	2_						

The use of a dropdown list for categorical predictors is easier than having to enter coded 0,1 values.

15. Click on Sheet **MReg1 – Residuals** to view the Residual Plots. Note, Sheet MReg# will increment every time a model is refitted.



Regular Residual Plots display the raw residuals (Y - Ŷ), the unexplained variation from the regression model, with a Histogram, Normal Probability Plot, Residuals vs Data Order, Residuals vs Predicted Values, Residuals vs Continuous Predictors and Residuals vs Categorical Predictors.

We expect to see the residuals normally distributed with no obvious patterns in the above graphs. This is not the case here, but they are better than the residual plots in the previous analysis, which showed a strong trend in the Residuals versus Predicted Values.

The table of Residuals, Standardized Residuals, Studentized (Deleted t) Residuals, Cook's Distance (Influence), Leverage and DFITS are given to the left of the Residual Plots:

Obs. No.	Responsive to Calls	Ease of Communications	Customer Type	Overall Satisfaction	Predicted (Fitted) Values	Residuals	Standardized Residuals	Studentized (Deleted t) Residuals	Cook's Distance (Influence)	Leverage	DFITS
1	3.02	4.07	2_	3.54	3.526825967	0.013174033	0.062366481	0.062020373	2.91048E-05	0.069618247	0.016965448
2	3.21	3.11	3_	3.16	3.248936833	-0.088936833	-0.417290881	-0.415368144	0.000971936	0.052865383	-0.098132536
3	1.93	2.9	2_	2.42	2.715216143	-0.295216143	-1.549801613	-1.562153591	0.077278101	0.243421198	-0.886085939
4	1.88	2.52	2_	2.7	2.613198544	0.086801456	0.484932998	0.482862646	0.01168433	0.331938061	0.34036412
5	3.75	2.86	3_	3.31	3.352565853	-0.042565853	-0.200400142	-0.199328176	0.000253146	0.05929637	-0.050044498
6	4.31	3.93	2_	4.12	4.114160994	0.005839006	0.027045305	0.026894743	2.11502E-06	0.02810287	0.00457333

Standardized Residuals are the residuals, divided by an estimate of its standard deviation. This makes it easier to detect outliers. Standardized residuals greater than 3 and less than -3 are considered outliers and highlighted in red. Standardized residuals for observation numbers 43, 66 and 76 are highlighted. Excel's filter by Font Color can be used to view just the outliers:

Obs. No.	Responsive to Ca	Ease of Communicatio	Customer Ty	Overall Satisfacti	Predicted (Fitted) Valu	Residual	Standardized Residua	Studentized (Deleted t) Residua	Cook's Distance (Influenc	Leverage	DFITS
43	1.84	1.4	3_	3.12	2.496679199	0.623320801	3.575137902	3.838198517	0.738448226	0.366183086	2.917391148
66	3.91	4.58	3_	3.24	4.035035759	-0.795035759	-3.816203238	-4.145104335	0.152920246	0.095025138	-1.343186952
76	4.76	3.76	1_	3.26	4.206462499	-0.946462499	-4.485433106	-5.061977077	0.155216836	0.071623294	-1.40599962

Note that this filter action will cause the Residual Plots to be hidden; clear the Filter to view the Residual Plots again.

Studentized (Deleted t) Residuals are computed in the same way that standardized residuals are, except that the i^{th} observation is removed before performing the regression fit. This prevents the i^{th} observation from influencing the regression model, resulting in unusual observations being more likely to stand out.

Leverage is a measure of how far an individual X predictor value deviates from its mean. High leverage points can *potentially* have a strong effect on the estimate of regression coefficients. Leverage values fall between 0 and 1.

Cook's distance and DFITS are overall measures of influence. An observation is said to be influential if removing the observation substantially changes the estimate of model coefficients. Cook's distance can be thought of as the product of leverage and the standardized residual squared; DFITS as the product of leverage and the studentized residual. These diagnostic measures can be manually plotted using a Run Chart to identify unusually large values.

Commonly used rough criterion/cutoff for Cook's distance are: > 0.5, potentially influential and > 1, likely influential. A more accurate cutoff is the median of the F distribution: $> \mathrm{F}(0.5, p, n - p)$, where n is the sample size and p is the number of terms in the model design matrix, including the constant (i.e., the number of rows in the Parameter Estimates table). So in this example using the Excel formula, the cutoff is: =F.DIST(0.5, 10,(100-10),TRUE) = 0.114. Excel's filter > 0.114 can be used to view influential observations based on Cook's Distance:

Obs. No.	Responsive to Ca	Ease of Communicatio	Customer Ty	Overall Satisfacti	Predicted (Fitted) Valu	Residual	Standardized Residua	Studentized (Deleted t) Residua	Cook's Distance (Influenc	Leverage	DFITS
43	1.84	1.4	3_	3.12	2.496679199	0.623320801	3.575137902	3.838198517	0.738448226	0.366183086	2.917391148
66	3.91	4.58	3_	3.24	4.035035759	-0.795035759	-3.816203238	-4.145104335	0.152920246	0.095025138	-1.343186952
76	4.76	3.76	1_	3.26	4.206462499	-0.946462499	-4.485433106	-5.061977077	0.155216836	0.071623294	-1.40599962
96	1.01	2.12	1_	2.56	1.983240249	0.576759751	2.989630137	3.132589657	0.257949383	0.223965289	1.682880561

Note that this filter action will cause the Residual Plots to be hidden; clear the Filter to view the Residual Plots again.

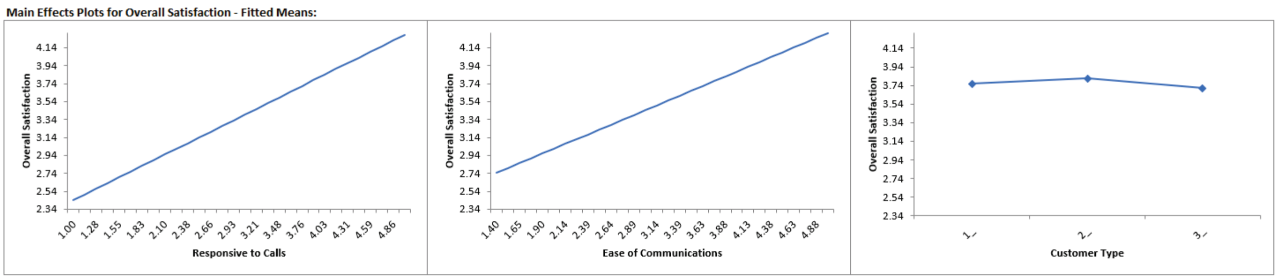
A commonly used criterion for absolute value of DFITS is $> 2\sqrt{p/n}$. So in this example, $|DFITS| > \ 2\sqrt{10/100} = 0.632$.

An observation that is an outlier and influential should be examined for measurement error or possible assignable cause. Observation 43 is clearly the most influential outlier, so you could try refitting the model excluding that observation to assess the influence. This would need to be done manually, but we will not do that in this demonstration.

The table to the right of the Residual Plots is the stored model design matrix with residuals. This can be used to manually create additional residual plots (use **SigmaXL > Graphical Tools > Scatter Plots**) for residuals versus interaction terms:

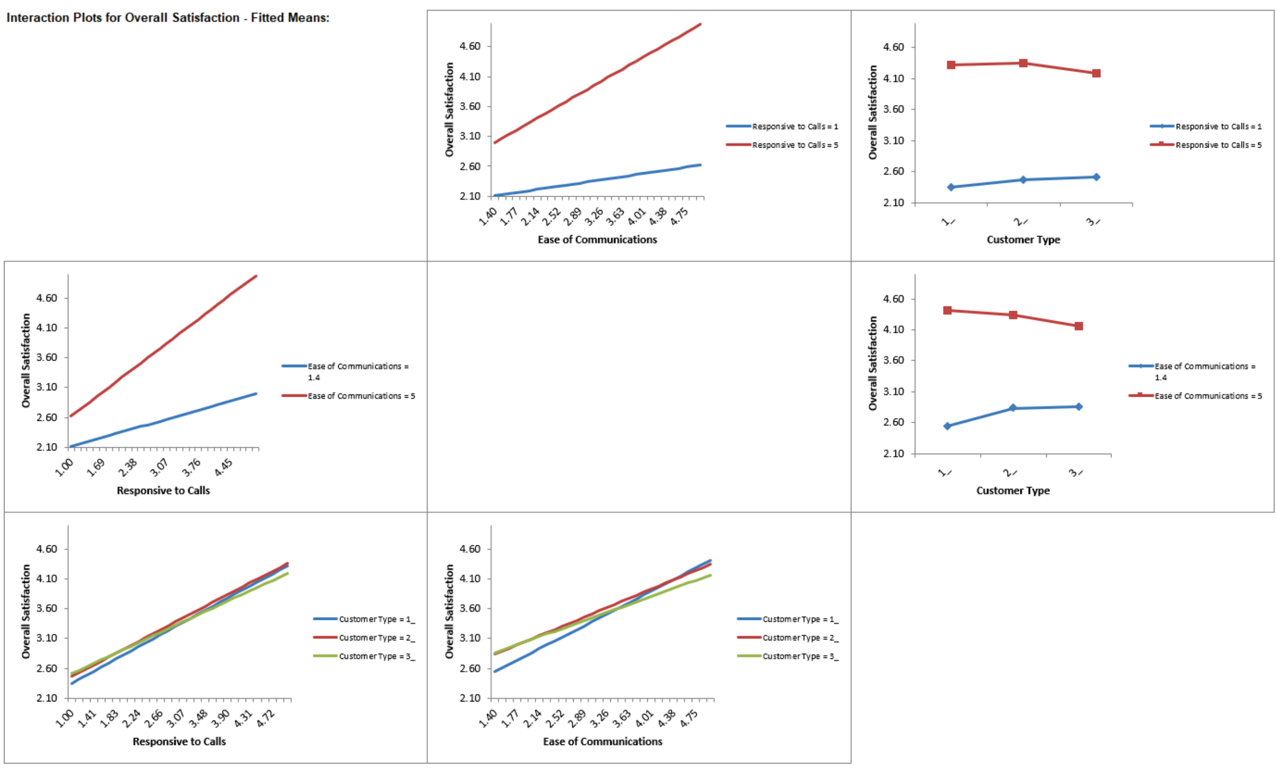
Model/Design Matrix for Overall Satisfaction:	Constant	Responsive to Calls	Ease of Communications	Customer Type_2_	Customer Type_3_	Responsive to Calls*Ease of Communications	Responsive to Calls*Customer Type_2_	Responsive to Calls*Customer Type_3_
Continuous Predictor Standardization/Coding: ((Yi - Mean)/Stdev)	1	-0.740511661	0.353207938	1	0	-0.261554597	-0.740511661	0
Categorical Predictor Coding: (1,0)	1	-0.573887768	-0.700161496	0	1	0.401814118	0	-0.573887768
	1	-1.696406629	-0.93058606	1	0	1.578652361	-1.696406629	0
	1	-1.740255022	-1.347544794	1	0	2.345071596	-1.740255022	0
	1	-0.100325123	-0.974476453	0	1	0.09776447	0	-0.100325123

16. Click on Sheet **MReg1 – Plots**. The Main Effects Plots and Interaction Plots for Overall Satisfaction are shown. These are based on Fitted Means as predicted by the model, not Data Means (as used in SigmaXL's Design of Experiments).



Main Effects Plots with Fitted Means use the predicted value for the response versus input predictor value, while holding all other variables constant. Continuous are held at their respective means and categorical are weighted equally.

Here we see that *Responsive to Calls* has the steepest slope followed by *Ease of Communications*. *Customer Type* does not appear to be an important factor, however, we also need to consider the interaction plots:



As with Main Effects Plots for Fitted Means, all continuous predictors not being plotted are held at their respective means and categorical are weighted equally.

Here we can clearly see a moderate interaction effect with the different slopes in *Responsive to Calls***Ease of Communications*, i.e., the effect that *Responsive to Calls* has on *Overall*

Satisfaction depends on the value of *Ease of Communications*. Similarly, the effect that *Ease of Communications* has on *Overall Satisfaction* depends on the value of *Responsive to Calls*.

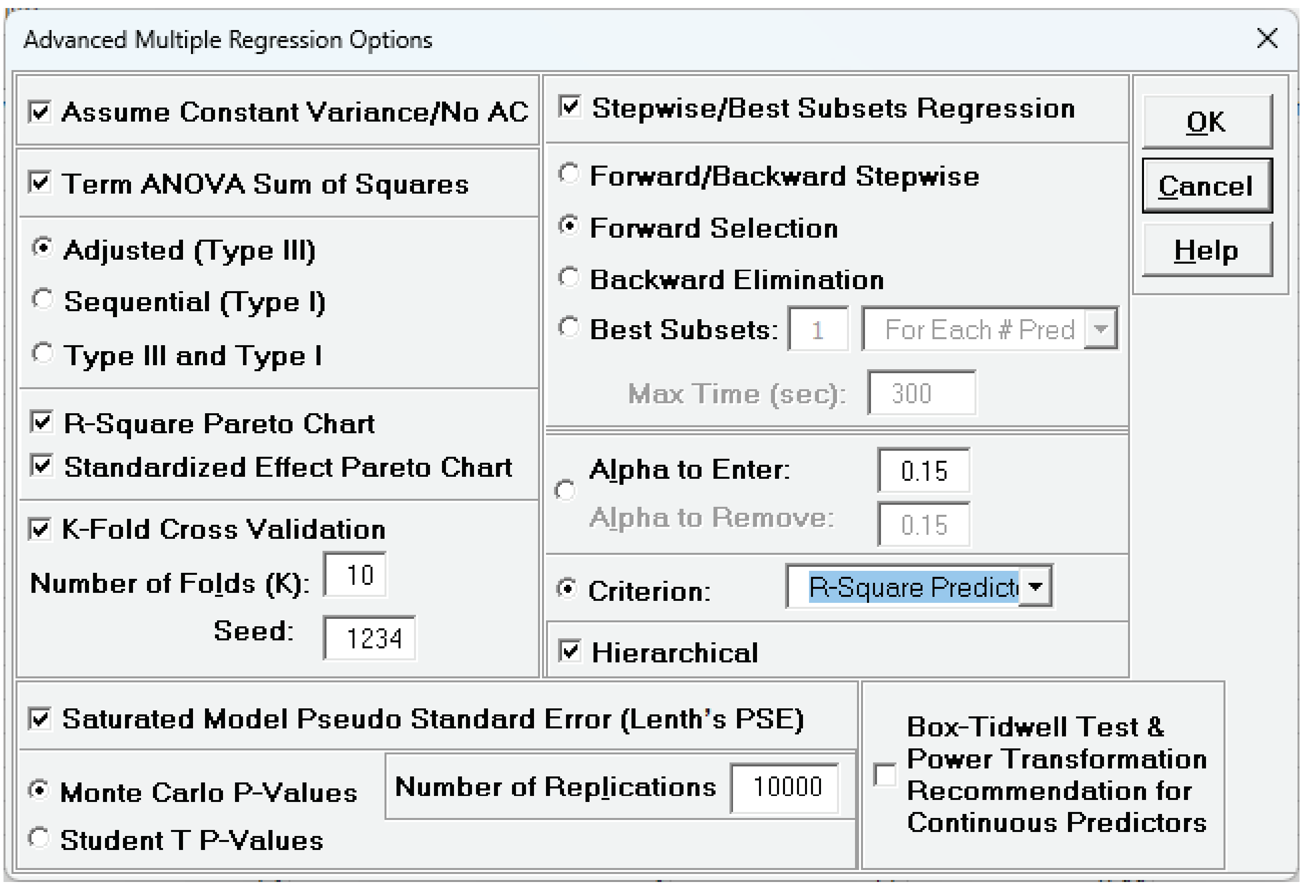
There also appears to be a slight interaction effect with the different slopes in *Customer Type* by *Ease of Communications*.

These plots should always be used in conjunction with the above Parameter Estimates table and Pareto Charts to determine significance.

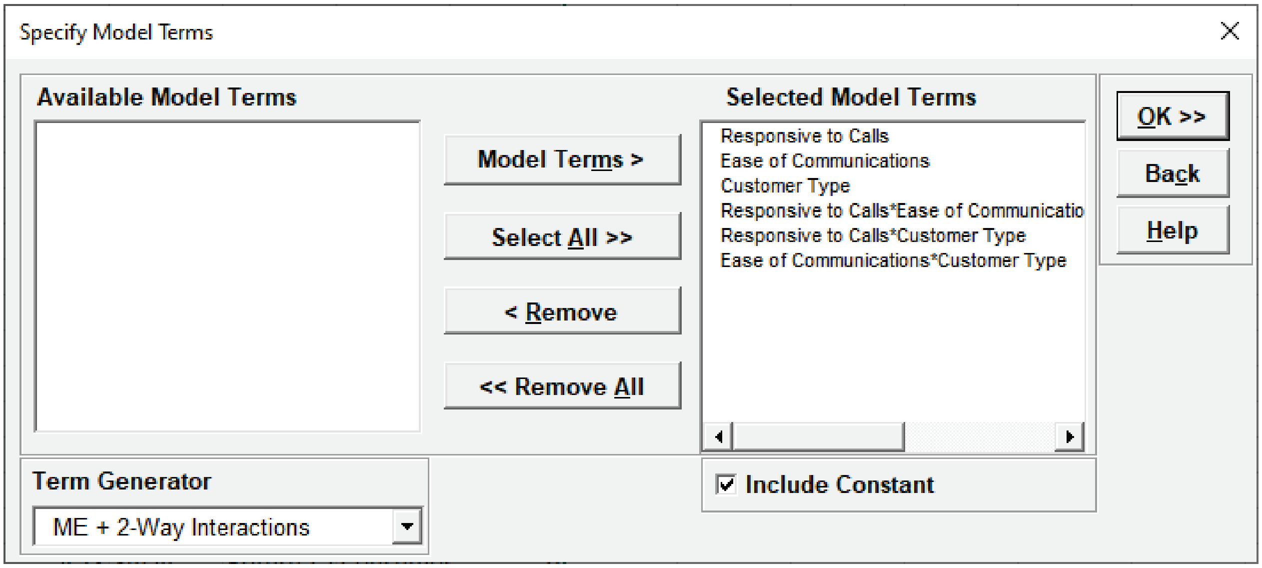
Note, if an interaction term is not in the model, the interaction plot is still displayed, but it is shaded grey.

17. Click **Recall Last Dialog** (or press **F3**). Click **Advanced Options** in the **Advanced Multiple Regression** dialog.

18. Check **Stepwise/Best Subsets Regression**. Select **Forward Selection** and **Criterion:** *R-Square Predicted*. **Hierarchical** is checked by default.



19. Click **OK**. Click **Next >>**.



20. Click **OK >>**.

21. Click on Sheet **MReg2 – Report**. Note, Sheet MReg# will increment every time a model is refitted. The Forward Selection report is given:

Step	Predictor Term	Mode	# Predictors	# Model Terms	P	S	R-Sq	R-Sq(Adj)	** R-Sq(Pred) **	PRESS	AICc	BIC	Mallows' Cp	Condition #
1	Responsive to Calls	Add	1	2	0.0000	0.4430	68.27%	67.94%	67.05%	19.9737	125.1878	132.7533	305.0440	1.0000
2	Ease of Communications	Add	2	3	0.0000	0.2489	90.08%	89.88%	89.08%	6.6194	11.0558	21.0555	31.3414	2.2210
** 3	Responsive to Calls*Ease of Communications	Add	3	4	0.0000	0.2281	91.76%	91.50%	90.35%	5.8495	-5.2260	7.1616	12.1727	3.1094
4	Ease of Communications*Customer Type	Add	4	6	0.0177	0.2209	92.43%	92.03%	90.42%	5.8070	-9.2250	7.7938	7.6093	13.9156
5	Customer Type	Add_H	5	8	0.1621	0.2189	92.73%	92.18%	90.25%	5.9080	-8.3976	13.0489	7.9015	18.6310
6	Responsive to Calls*Customer Type	Add	6	10	0.3903	0.2190	92.88%	92.17%	89.88%	6.1333	-5.4884	20.1685	10.0000	22.7277

Responsive to Calls	Ease of Communications	Customer Type	Responsive to Calls*Ease of Communications	Responsive to Calls*Customer Type	Ease of Communications*Customer Type
1	0	0	0	0	0
1	1	0	0	0	0
1	1	0	1	0	0
1	1	0	1	0	1
1	1	1	1	0	1
1	1	1	1	1	1

Model at Step 1	Model at Step 2	Model at Step 3	Model at Step 4	Model at Step 5	Model at Step 6
Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls
	Ease of Communications	Ease of Communications	Ease of Communications	Ease of Communications	Ease of Communications
		Responsive to Calls*Ease of Communications	Responsive to Calls*Ease of Communications	Customer Type	Customer Type
			Ease of Communications*Customer Type	Responsive to Calls*Ease of Communications	Responsive to Calls*Ease of Communications
				Ease of Communications*Customer Type	Responsive to Calls*Customer Type
					Ease of Communications*Customer Type

Scroll across to view the report tables. Forward Selection identifies the model at Step 3 (denoted with double asterisk ** and highlighted in yellow), as having the largest R-Square Predicted value = 90.35%. **** R-Sq(Pred) **** denotes that this is the criterion used to select the final model.

Note that while the model at Step 4 has a slightly higher R-Square Predicted, it is not selected because it does not include *Customer Type*, which is needed for the model to be Hierarchical. When *Customer Type* is added at Step 5, the R-Square Predicted value is less than that of Step 3.

Mode can be *Add*, *Add_H*, or *Remove*. *Add_H* denotes a predictor being added for Hierarchy. *Remove* is not applicable for Forward Selection.

 # Predictors is the number of continuous or categorical predictors in the model, excluding the constant.

 # Model Terms is the number of columns in the model design matrix, including the constant and coded columns for categorical predictors. This is the value used by all of the metrics.

P is the P-Value for the predictor. This is the P-Value for the term at that particular step and is subject to change when other terms are added. P-Values are used for model selection (i.e., stopping rule) only when Alpha-to-Enter has been specified. Since we selected Criterion, they are not used for model selection, but the decision of what term to consider at a particular step is still based on the P-Value.

S is the model standard deviation or root-mean-square error at that step.

R-Sq is the R-Square, i.e., the percent variation in response that is explained by the model at that step.

R-Sq(Adj) is the R-Square Adjusted for the model at that step. The model at Step 5 has the maximum R-Square Adjusted = 92.18%.

R-Sq(Pred) is the R-Square Predicted for the model at that step.

R-Sq(10-Fold) is the R-Square K-Fold and reported only if it is selected as the criterion.

PRESS is the prediction error sum of squares and is used to calculate R-Square Predicted.

AICc is the Akaike Information Criterion corrected for small sample sizes. If we had selected AICc as the criterion, then the model at Step 5 would have been selected with minimum AICc = -8.3976 (Step 4 is excluded due to the Hierarchical requirement).

BIC is the Bayesian Information Criterion and includes a stronger penalty for the number of terms in the model design matrix. The model at Step 3 has the minimum BIC = 7.1616.

Mallows' Cp is a measure that compares the full model to candidate models and is similar to the Akaike Information Criterion. An unbiased model has the Mallows' Cp close to the number of model terms. Mallows' Cp for the model with all terms will exactly match the number of model terms, which in this example is 10. The model with optimal Mallows' Cp value for this example is at Step 5, Mallows' Cp = 7.9015, which is close to the # Model Terms = 8.

Condition # measures whether a model is well conditioned. An ill conditioned model will have a large change in coefficient values for a small change in the input data. A rule of thumb is that $CN > 100$ indicates moderate multicollinearity, so all of the models considered here have an acceptable condition number.

For details on these metrics, see the Appendix: **Advanced Multiple Regression**.

In conclusion, we could use either Step 3 or Step 5 as our final model, but we will proceed with the simpler model at Step 3. It may be beneficial to refit using a different criterion for Forward such as AICc or BIC and/or a different method. Later, we will demonstrate the use of Best Subsets with the AICc criterion.

22. Click on Sheet **MReg2 – Model Overall**.

23. The Model Summary is:

Model Summary	
R-Square	91.76%
R-Square Adjusted	91.50%
R-Square Predicted	90.35%
S (Root Mean Square Error)	0.2281

R-Square = 91.76% and R-Square Adjusted = 91.5% have been reduced slightly (versus the full model R-Square = 92.88% and R-Square Adjusted = 92.17%). S = 0.2281 is slightly higher than the full model S = 0.219. However, R-Square Predicted = 90.35% is an improvement over the full model R-Square Predicted = 89.88% and we have simplified the model by removing all terms with *Customer Type*. This will make interpretation much easier.

24. The Model Information is given as:

Model Information	
Continuous Predictor Standardization/Coding	((Xi - Mean)/Stdev)
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method: Forward - Hierarchical	Max Rsq-Pred

This summarizes the selected options for the model showing that the Stepwise Method is Forward Selection, Hierarchical and the Criterion is *Max Rsq-Pred*, Maximize R-Square Predicted.

25. The Information Criteria and Validation table is given as:

Information Criteria and Validation	
AICc	-5.2260
BIC	7.1616
R-Square 10-Fold	90.07%
S 10-Fold	0.2453

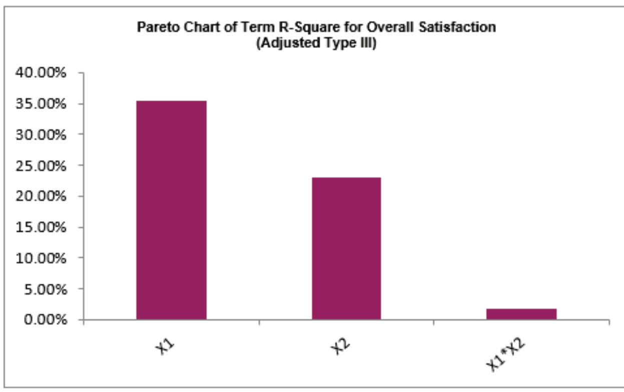
R-Square 10-Fold = 90.07% is an improvement over the full model R-Square 10-Fold = 89.54%. S 10-Fold = 0.2453 is slightly less than the full model S 10-Fold = 0.2518.

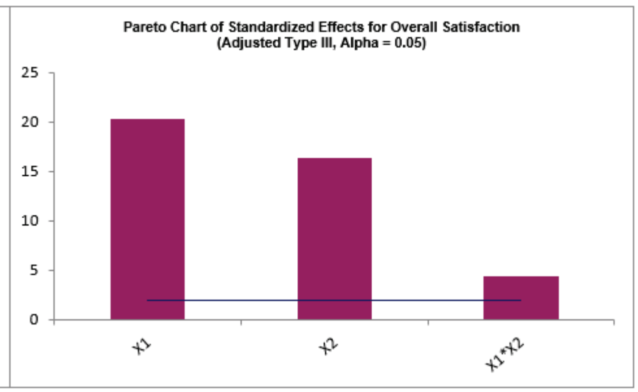
26. The Parameter Estimates – Standardized, Analysis of Variance for Model, Analysis of Variance for Predictors (Adjusted Type III) tables, Pareto of Term R-Square and Pareto of Standardized Effects are shown:

Parameter Estimates - Standardized						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	3.762957198	0.024408729	154.1644	0.0000		
Responsive to Calls	0.538373144	0.026504958	20.3122	0.0000	1.3364	0.7483
Ease of Communications	0.409954766	0.025009143	16.3922	0.0000	1.1898	0.8405
Responsive to Calls*Ease of Communications	0.102171588	0.023132627	4.4168	0.0000	1.2368	0.8086

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	3	55.61626813	18.53875604	356.2246	0.0000
Error	96	4.99606287	0.052042322		
Total (Model + Error)	99	60.612331	0.612245768		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Responsive to Calls	1	21.47183594	21.47183594	412.5841	0.0000	35.42%	20.3122
Ease of Communications	1	13.98398363	13.98398363	268.7041	0.0000	23.07%	16.3922
Responsive to Calls*Ease of Communications	1	1.015236039	1.015236039	19.5079	0.0000	1.67%	4.4168





Pareto Legend
X1 = Responsive to Calls
X2 = Ease of Communications

The continuous predictors *Responsive to Calls*, *Ease of Communications* and interaction *Responsive to Calls*Ease of Communications* all show as clearly significant.

The Pareto Chart of Term R-Square and Pareto Chart of Standardized Effects graphically present the information given in the Predictor ANOVA table and show that main effects *Responsive to Calls*, *Ease of Communications* are the dominant contributors to the variability in Overall Satisfaction. The interaction term *Responsive to Calls*Ease of Communications* is a smaller contributor, but still significant.

The Variance Inflation Factor (VIF) scores have been reduced from the full model, for example, *Ease of Communications* VIF = 1.19 versus the full model VIF = 4.01.

27. The Durbin-Watson Test for Autocorrelation in Residuals table is:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	1.8120
P-Value Positive Autocorrelation	0.1669
P-Value Negative Autocorrelation	0.8226

Here, there is no significant positive or negative autocorrelation.

Note that this is a change from the Durbin Watson Test for the full model which was:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	1.5800
P-Value Positive Autocorrelation	0.0157
P-Value Negative Autocorrelation	0.9827

This shows that the Residual diagnostics can change as the model changes with model refinement.

28. The Breusch-Pagan Test for Constant Variance is:

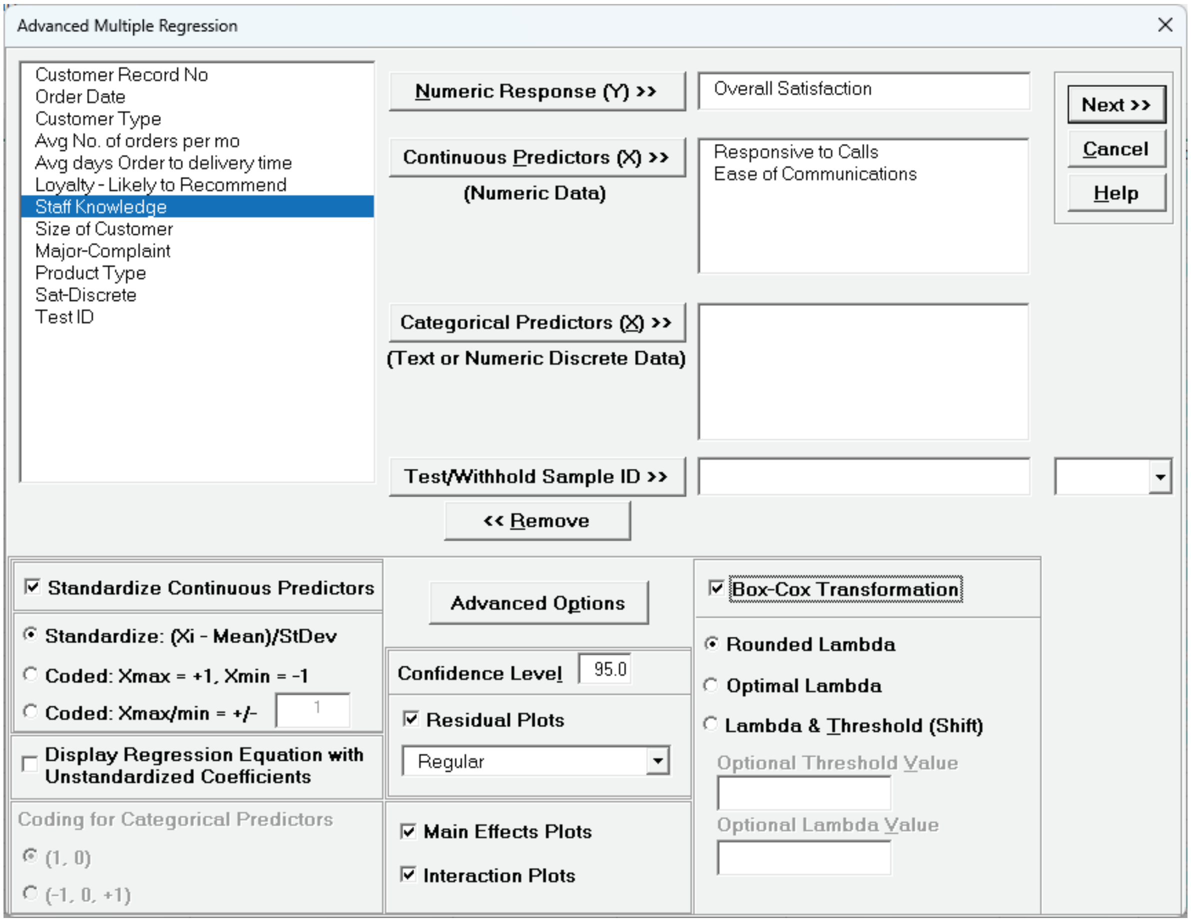
Breusch-Pagan Test for Constant Variance (Koenker Studentized - Robust)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	Chi-Square	DF	P-Value
All Terms	8.50973	3	0.0366
Responsive to Calls	3.16086	1	0.0754
Ease of Communications	2.93917	1	0.0865
Responsive to Calls*Ease of Communications	7.30293	1	0.0069

Here we see that the *All Terms* test is significant and that the *Responsive to Calls*Ease of Communications* interaction is also significant. This has changed from the full model which showed that *All Terms* was not significant:

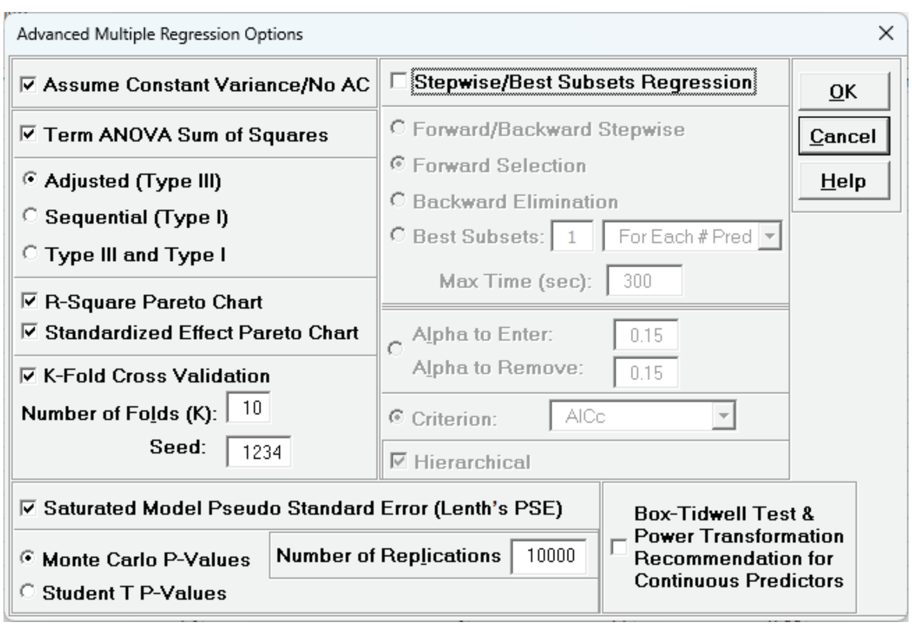
Breusch-Pagan Test for Constant Variance (Koenker Studentized - Robust)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	Chi-Square	DF	P-Value
All Terms	11.2222	9	0.2608
Responsive to Calls	1.97856	1	0.1595
Ease of Communications	1.22053	1	0.2693
Customer Type	3.55179	2	0.1693
Responsive to Calls*Ease of Communications	3.84677	1	0.0498
Responsive to Calls*Customer Type	4.2728	2	0.1181
Ease of Communications*Customer Type	0.700644	2	0.7045

Next, we will refit the model using a Box-Cox Transformation in order to deal with this violation of the regression assumption of constant variance.

29. Click **Recall Last Dialog** (or press **F3**). Select *Customer Type* and click **<< Remove**. Uncheck **Display Regression Equation with Unstandardized Coefficients**. Check **Box-Cox Transformation** with **Rounded Lambda** option.

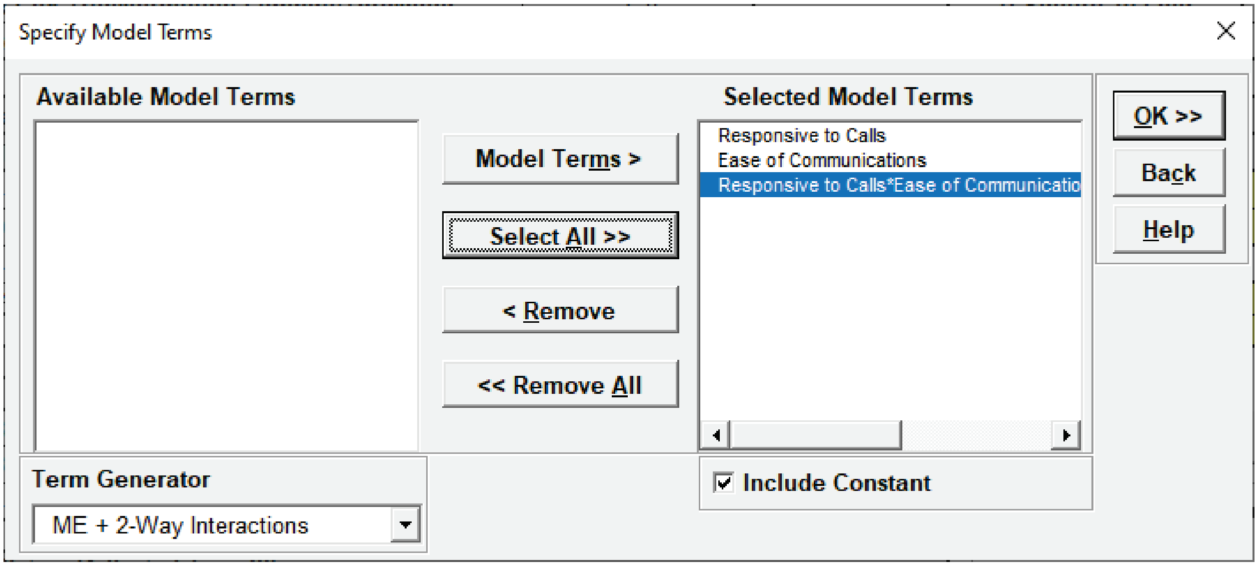


30. Click **Advanced Options**. Uncheck **Stepwise/Best Subsets Regression**.



Note, Box-Cox Transformation can be used with **Stepwise/Best Subsets Regression** but here we want to manually specify the model.

31. Click **OK**. Click **Next >>**. Select *ME + 2-Way Interactions*. Click **Select All >>**.



32. Click **OK >>**. The Advanced Multiple Regression report is given for the revised model. Note that the Model Summary, Information Criteria and Validation, Parameter Estimates, ANOVA, DW and BP Tests, and Residuals are for the Box-Cox transformed response. The Main Effects and Interaction Plots, Predicted Response Calculator, Optimize and Contour/Surface Plots all use an inverse transformation to return to the original units.

33. The Regression Equation with standardized coefficients is:

    Multiple Regression Model: Overall Satisfaction = ((14.5676)
    + (3.99525)*((Responsive_to_Calls-3.8644)/1.14029)
    + (3.14334)*((Ease_of_Communications-3.7481)/0.911361)
    + (1.30149)*((Responsive_to_Calls-3.8644)/1.14029)*((Ease_of_Communications-3.7481)/0.911361))^(1/2)

    This is the display version of the prediction equation given at cell **L14** (which has more precision for the coefficients).

    Note, these coefficients match those given in the Parameter Estimates table since they are standardized. If a simpler form of the prediction equation is desired, one can always rerun the analysis with **Display Regression Equation with Unstandardized Coefficients** checked.

34. The Model Summary is:

Model Summary	
R-Square	92.77%
R-Square Adjusted	92.54%
R-Square Predicted	91.95%
S (Root Mean Square Error)	1.5567

R-Square, R-Square Adjusted and R-Square Predicted have all increased over the previous untransformed model:

Model Summary	
R-Square	91.76%
R-Square Adjusted	91.50%
R-Square Predicted	90.35%
S (Root Mean Square Error)	0.2281

Note that the S = 1.5567 is reported for the Box-Cox Transformed response, so cannot be compared to the untransformed S = 0.2281, but the R-Square statistics can be compared.

35. The Model Information is given as:

Model Information	
Continuous Predictor Standardization/Coding	((Xi - Mean)/Stdev)
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	2; 0
Stepwise Method	N/A

This summarizes the selected options for the model showing the Box-Cox Lambda value = 2 (i.e., the response values are squared) and Stepwise is not used.

36. The Information Criteria and Validation table is given as:

Information Criteria and Validation	
AICc	378.8559
BIC	391.2435
R-Square 10-Fold	91.81%
S 10-Fold	1.6228

R-Square 10-Fold has increased over the previous untransformed model:

Information Criteria and Validation	
AICc	-5.2260
BIC	7.1616
R-Square 10-Fold	90.07%
S 10-Fold	0.2453

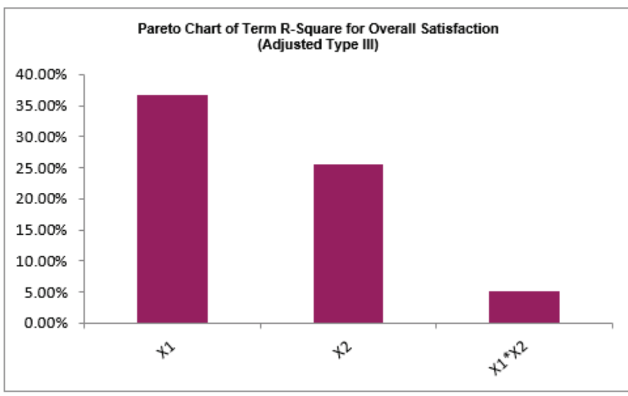
As noted above, the S 10-Fold = 1.6228 is reported for the Box-Cox Transformed response, so cannot be compared to the untransformed S 10-Fold = 0.2453, but the R-Square 10-Fold statistic can be compared.

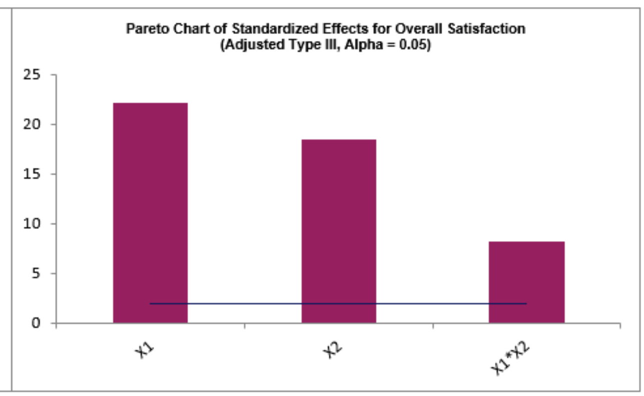
37. The Parameter Estimates – Standardized, Analysis of Variance for Model, Analysis of Variance for Predictors (Adjusted Type III) tables, Pareto of Term R-Square and Pareto of Standardized Effects are shown:

Parameter Estimates - Standardized						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	14.56758356	0.16655912	87.4619	0.0000		
Responsive to Calls	3.99525069	0.180863267	22.0899	0.0000	1.3364	0.7483
Ease of Communications	3.143344979	0.170656196	18.4192	0.0000	1.1898	0.8405
Responsive to Calls*Ease of Communications	1.301490517	0.157851315	8.2450	0.0000	1.2368	0.8086

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	3	2984.352023	994.7840075	410.5118	0.0000
Error	96	232.63464	2.4232775		
Total (Model + Error)	99	3216.986663	32.49481477		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Responsive to Calls	1	1182.470937	1182.470937	487.9635	0.0000	36.76%	22.0899
Ease of Communications	1	822.1349683	822.1349683	339.2657	0.0000	25.56%	18.4192
Responsive to Calls*Ease of Communications	1	164.7360922	164.7360922	67.9807	0.0000	5.12%	8.2450





Pareto Legend
X1 = Responsive to Calls
X2 = Ease of Communications

The interaction *Responsive to Calls*Ease of Communications* has increased in R-Square % and Standardized Effect versus the untransformed model, so is a stronger interaction effect.

38. Scroll down to view the Durbin-Watson Test for Autocorrelation in Residuals table:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	1.7952
P-Value Positive Autocorrelation	0.1466
P-Value Negative Autocorrelation	0.8438

The DW values have changed slightly from the untransformed model but still show no significant autocorrelation.

39. The Breusch-Pagan Test for Constant Variance is:

Breusch-Pagan Test for Constant Variance (Koenker Studentized - Robust)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	**Chi-Square**	**DF**	**P-Value**
All Terms	0.208641	3	0.9762
Responsive to Calls	0.104744	1	0.7462
Ease of Communications	0.0437594	1	0.8343
Responsive to Calls*Ease of Communications	0.180433	1	0.6710

This is a dramatic change for the BP Test for Constant Variance. The *All Terms* test now shows as insignificant, as well the *Responsive to Calls*Ease of Communications* interaction is insignificant. Recall that for the untransformed model, the BP Test was:

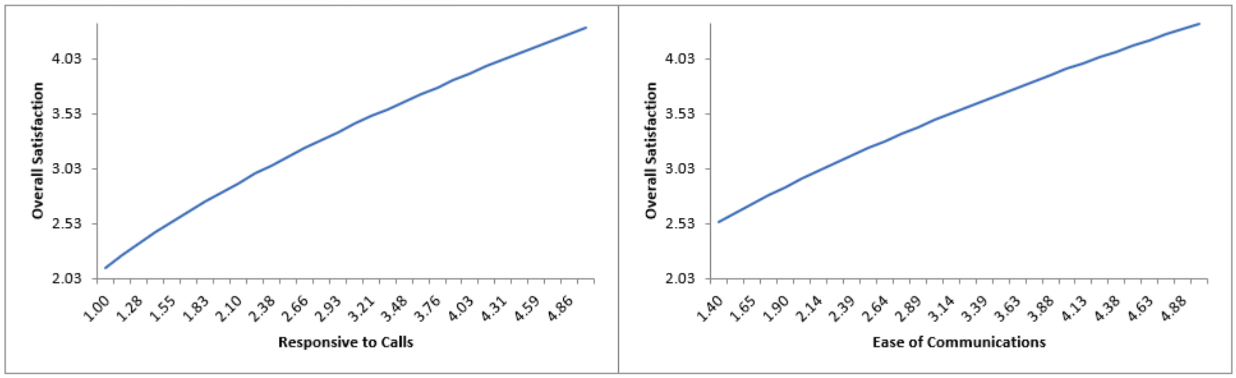
Breusch-Pagan Test for Constant Variance (Koenker Studentized - Robust)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	**Chi-Square**	**DF**	**P-Value**
All Terms	8.50973	3	**0.0366**
Responsive to Calls	3.16086	1	0.0754
Ease of Communications	2.93917	1	0.0865
Responsive to Calls*Ease of Communications	7.30293	1	**0.0069**

Tip: The use of the Box-Cox Transformation has eliminated the problem of non-constant variance in the model. Had that not been the case, we would have then proceeded to refit the model with **Assume Constant Variance/No AC** unchecked.

A review of the Residual Plots in Sheet **MReg3 – Residuals** shows that they still do not exhibit the properties of normally distributed, with no patterns and no outliers, so while not ideal, we will consider this as our final model for the purposes of this demonstration.

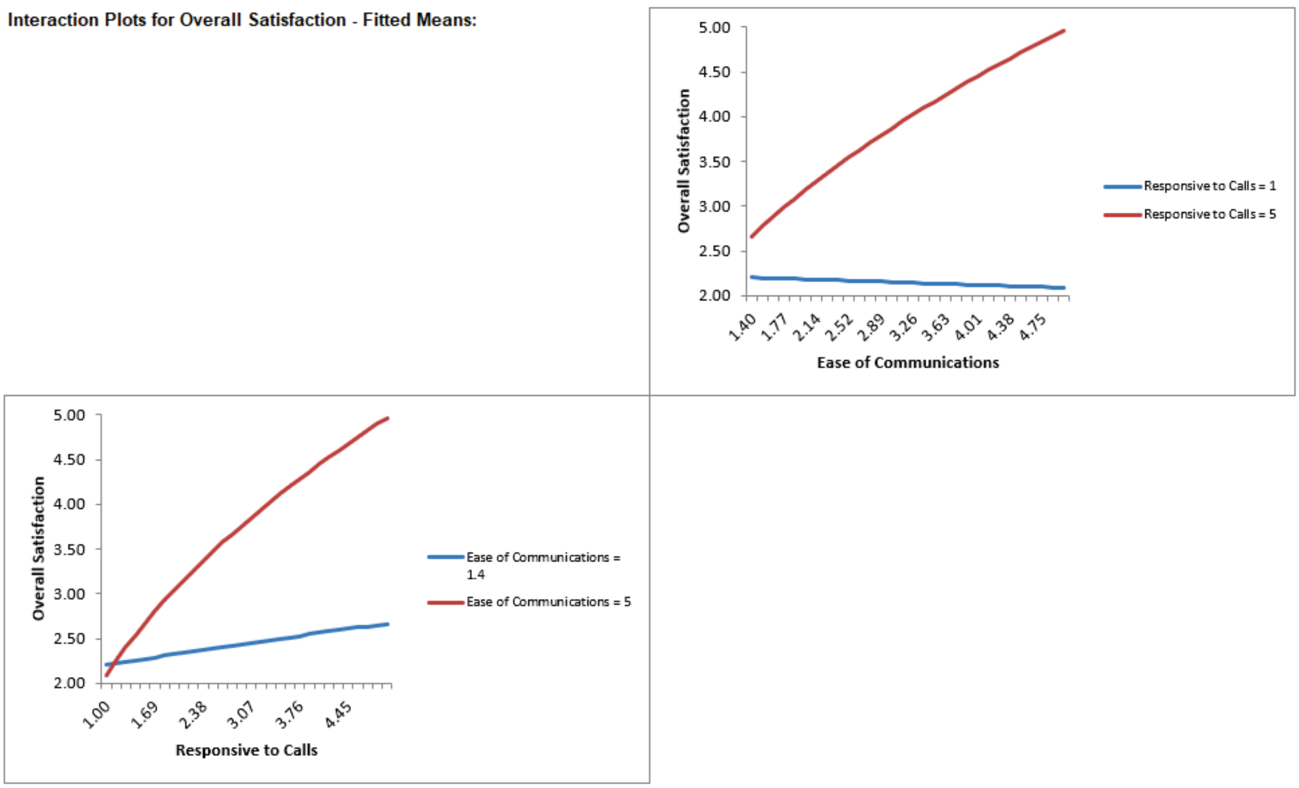
40. Click on Sheet **MReg3 – Plots**. Note, Sheet MReg# will increment every time a model is refitted. The Main Effects Plots and Interaction Plots for Overall Satisfaction are shown. These are based on Fitted Means as predicted by the model.

Main Effects Plots for Overall Satisfaction - Fitted Means:



Here we see that *Responsive to Calls* has the steepest slope but now there is some curvature due to the Box-Cox transformation.

Interaction Plots for Overall Satisfaction - Fitted Means:



We also see the curvature due to the Box-Cox Transformation in the Interaction Plots.

Here we can clearly see a strong interaction effect with the different slopes in *Responsive to Calls***Ease of Communications*, i.e., the effect that *Responsive to Calls* has on *Overall Satisfaction* depends on the value of *Ease of Communications*. Similarly, the effect that *Ease of Communications* has on *Overall Satisfaction* depends on the value of *Responsive to Calls*.

41. Click on Sheet **MReg3 – Model**. Scroll to the Predicted Response Calculator. Enter *Responsive to Calls* and *Ease of Communication* values = 5 to predict Overall Satisfaction with the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Responsive to Calls	5	4.964345487	0.612408292	4.888790221	5.038767944	4.633175042	5.274764568
Ease of Communications	5						

Note the formula at cell **L14** is an Excel formula (if the formula exceeds Excel's limit of 8192 characters, it is not given, but a predicted response value is still computed). Since we are applying an inverse transformation to the Box-Cox transformed prediction, the Confidence and Prediction Intervals are not symmetric.

42. Next, we will use SigmaXL's built in Optimizer. Scroll to view the Optimize Options:

Optimize Options			
Continuous Predictors	Lower Bound	Upper Bound	Integer
Responsive to Calls	1	5	0
Ease of Communications	1.4	5	0

Here we can constrain the continuous predictors, and if there was a categorical predictor, specify a level to use for optimization. If a continuous predictor is integer, change the **Integer** 0 to 1, and the Optimizer will return only integer values for that predictor.

We will leave the Optimize Option settings as is.

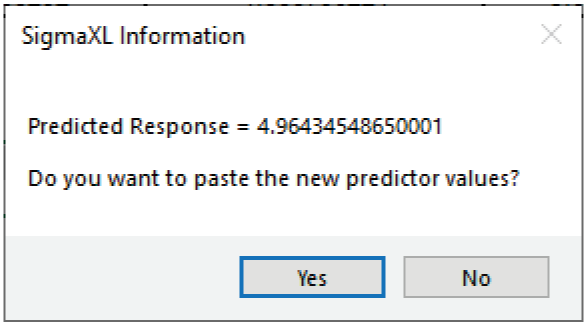
43. Scroll back to view the Goal setting and Optimize button. Select Goal = *Maximize*.

Goal:	Maximize
Target:	

Optimize | Contour/Surface Plots

The optimizer uses Multistart Nelder-Mead Simplex to solve for the desired response goal with given constraints. For more information see the Appendix: Single Response Optimization.

44. Click **Optimize**. The response solution and prompt to paste values into the Input Settings of the Predicted Response Calculator is given:



45. Click **Yes** to paste the values.

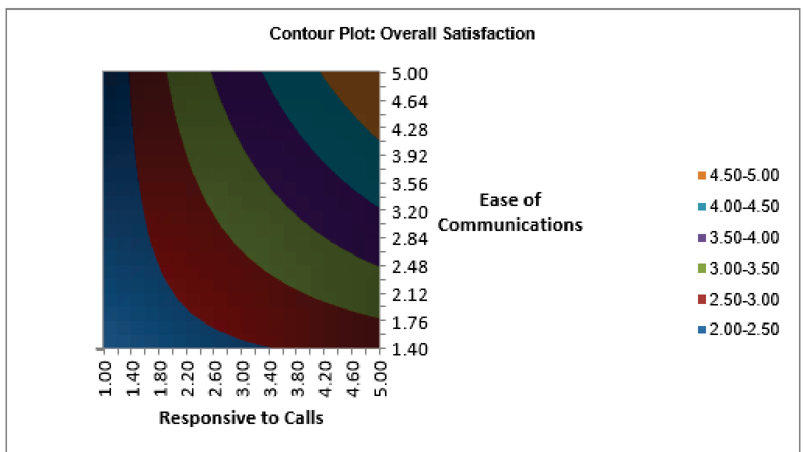
Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Responsive to Calls	5	4.964345487	0.612408292	4.888790221	5.038767944	4.633175042	5.274764568
Ease of Communications	5						

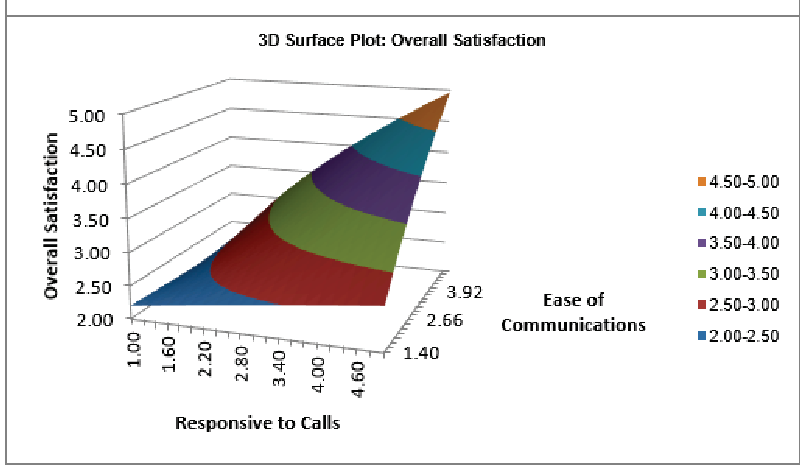
This confirms that our manual settings were correct to provide the maximum Overall Satisfaction.

46. Next, we will create a Contour/Surface Plot. Click the **Contour/Surface Plots** button. Note this button is not available if there are fewer than two continuous predictors.

47. A new sheet is created, **MReg3 – Contour** that displays the plots:

Predictors	Hold Values
Responsive to Calls	5
Ease of Communications	5



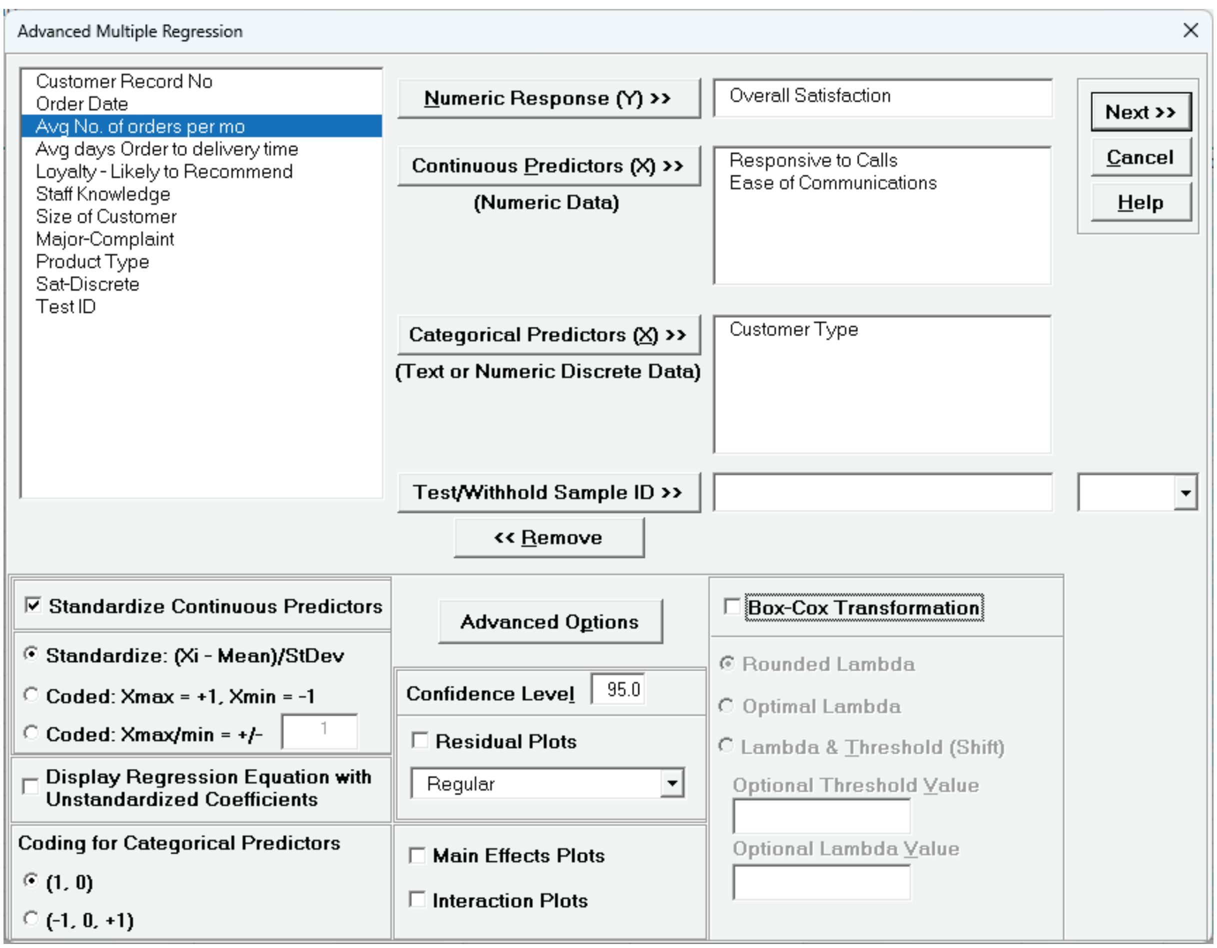


The curvature in the response is due to the two-way interaction (and the Box-Cox transformation). We clearly see that maximizing both Response to Calls and Ease of Communications is necessary to maximize Overall Satisfaction.

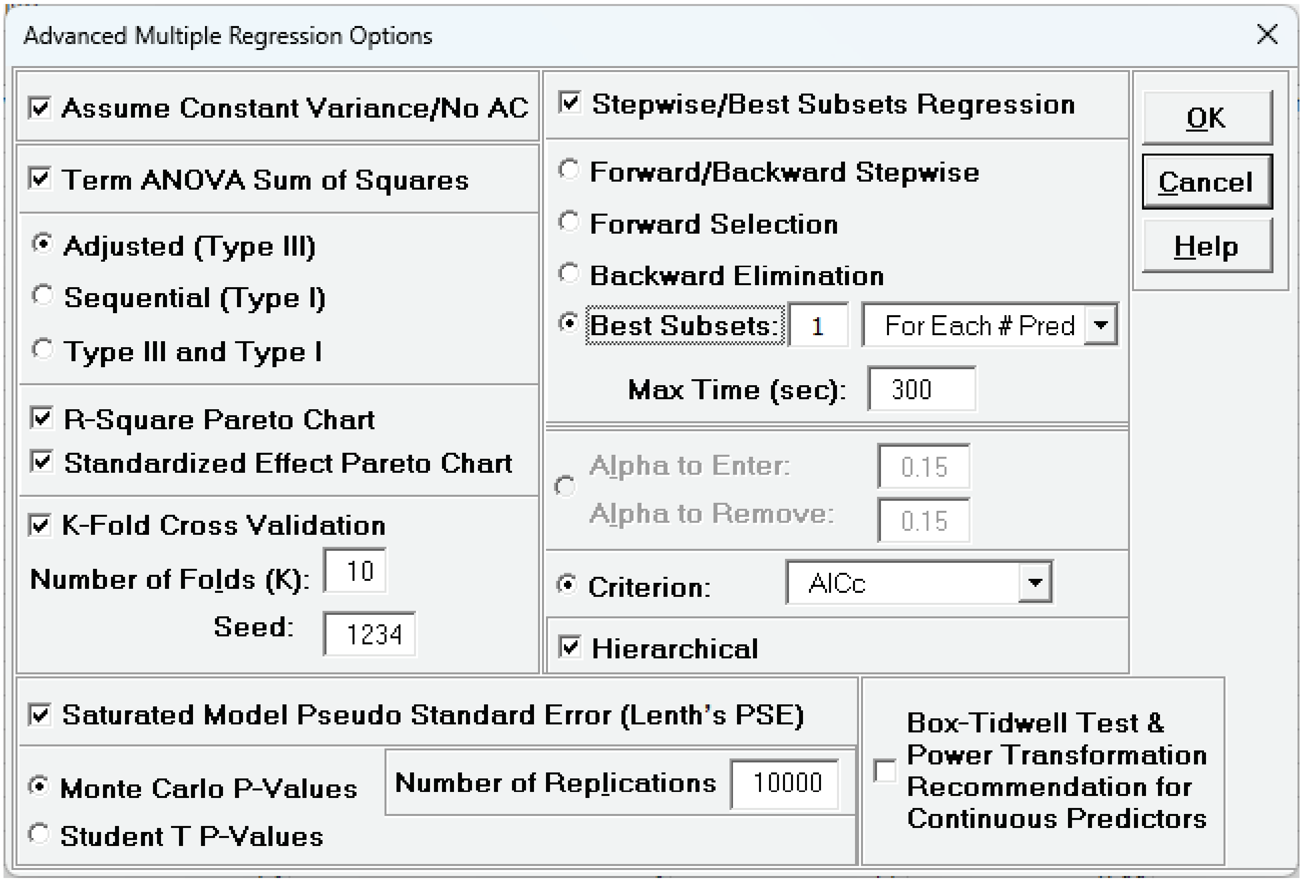
The table with the **Hold Values**, gives the values used to hold a predictor constant if it is not in the plot, so is not applicable here with only one plot based on the two continuous predictors.

Tip: These values are obtained from the Predicted Response Calculator settings, so if you wish to use different **Hold Values**, simply select the Model sheet, change the **Enter Settings** values and recreate the plots.

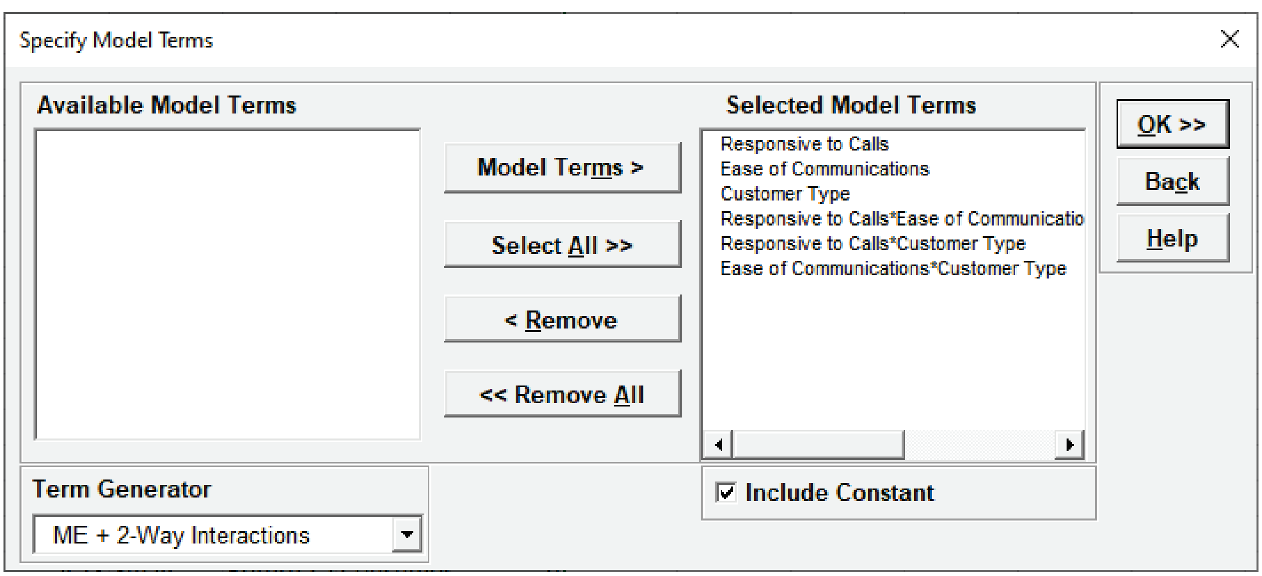
48. Next, we will demonstrate the use of Best Subsets Regression. Click **Recall Last Dialog** (or press **F3**). Select *Customer Type* and click **Categorical Predictors** (X) >>. Uncheck **Box-Cox Transformation** so that we can compare the Best Subsets report to the earlier Forward Selection report. Uncheck **Residual Plots**, **Main Effects** and **Interaction Plots** as we will not revisit these.



49. Click **Advanced Options**. Check **Stepwise/Best Subsets Regression**. Select **Best Subsets** with default *1 For Each # Pred*, **Max Time (sec)** = 300 and **Criterion** as *AICc*. **Hierarchical** is checked by default.



50. Click **OK**. Click **Next >>**. Select *ME + 2-Way Interactions,* click **Select All >>**.



51. Click **OK >>**.

52. Click on Sheet **MReg4 – Report**. Note, Sheet MReg# will increment every time a model is refitted. The Best Subsets report is given:

Model	# Predictors	# Model Terms	S	R-Sq	R-Sq(Adj)	R-Sq(Pred)	PRESS	R-Sq(10-Fold)	S(10-Fold)	** AICc **	BIC	Mallows Cp	Condition #
1	1	2	0.4430	68.27%	67.94%	67.05%	19.9737	66.36%	0.4561	125.1878	132.7533	305.0440	1.0000
2	2	3	0.2489	90.08%	89.88%	89.08%	6.6194	88.74%	0.2653	11.0558	21.0555	31.3414	2.2210
3	3	4	0.2281	91.76%	91.50%	90.35%	5.8495	90.07%	0.2504	-5.2260	7.1616	12.1727	3.1094
4	4	6	0.2264	92.05%	91.63%	90.14%	5.9770	89.54%	0.2597	-4.2400	12.7788	12.4962	5.3734
5	** 5	8	0.2189	92.73%	92.18%	90.25%	5.9080	89.57%	0.2621	-8.3976	13.0489	7.9015	18.6310
6	6	10	0.2190	92.88%	92.17%	89.88%	6.1333	89.54%	0.2655	-5.4884	20.1685	10.0000	22.7277

Responsive to Calls	Ease of Communications	Customer Type	Responsive to Calls*Ease of Communications	Responsive to Calls*Customer Type	Ease of Communications*Customer Type
1	0	0	0	0	0
1	1	0	0	0	0
1	1	0	1	0	0
1	1	1	1	0	0
1	1	1	1	0	1
1	1	1	1	1	1

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls	Responsive to Calls
	Ease of Communications	Ease of Communications	Ease of Communications	Ease of Communications	Ease of Communications
		Responsive to Calls*Ease of Communications	Customer Type	Customer Type	Customer Type
			Responsive to Calls*Ease of Communications	Responsive to Calls*Ease of Communications	Responsive to Calls*Ease of Communications
				Ease of Communications*Customer Type	Responsive to Calls*Customer Type
					Ease of Communications*Customer Type

This is similar to the Forward Selection report but instead of steps, shows the "best" model that minimizes the criterion AICc for each number of predictors. The double asterisk ** and yellow highlight show the selected overall best model. Model 5 has the minimum AICc.

Model 3 has the minimum BIC, but refitting the model with the BIC criterion may result in different models being selected for each # predictors, so it may be beneficial to run Best Subsets twice. We will not do that in this demonstration.

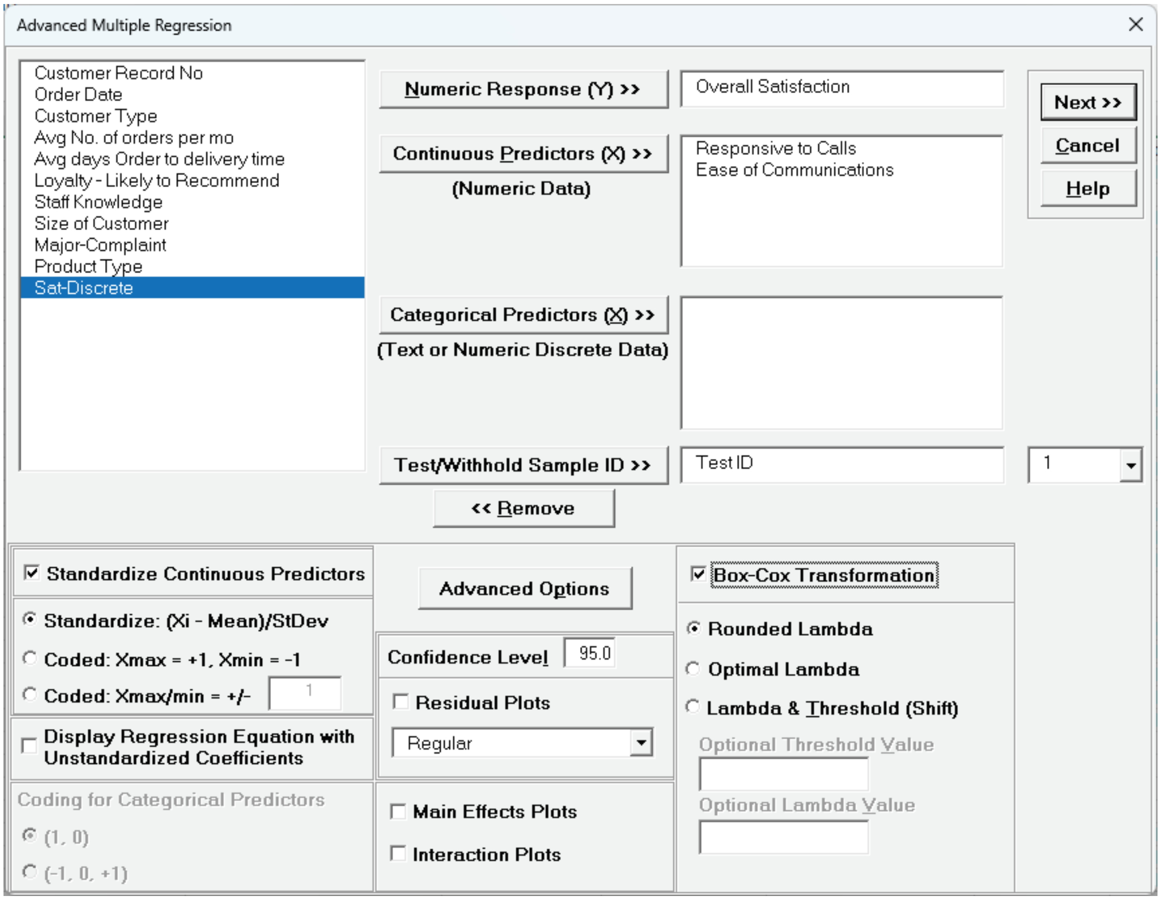
Note that Best Subsets evaluates only models that satisfy Hierarchy, those that do not are not considered. This simplifies the interpretation of the report.

R-Square Predicted and R-Square 10-Fold cannot be specified as criterion in Best Subsets due to their computation times, but they are reported here.

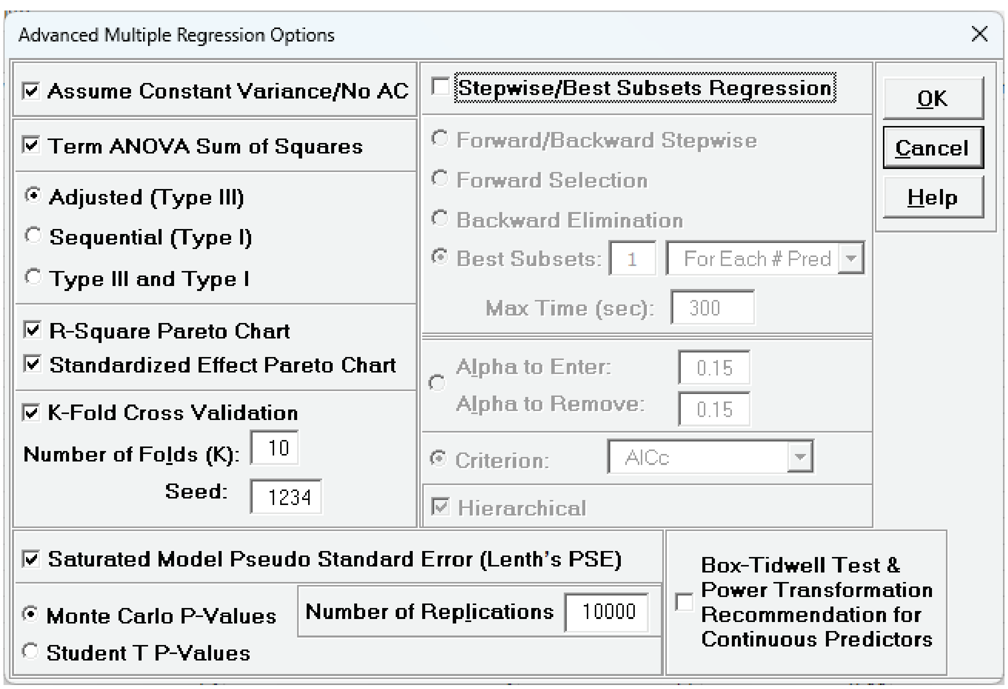
Mode and **P** given in the Forward Selection report are not applicable for Best Subsets.

53. Finally, we will demonstrate the use of **Test/Withhold Sample ID**. This splits the data into a training and test/withhold sample for validation. For this demonstration, we will use the *TestID* column given in **Customer Data.xlsx**, but a column of random 0/1 values can also be created manually using the Excel function =IF(RAND()<=0.3,1,0), where 0.3 is the fraction desired for the test/withhold sample, 0 denotes training data, and 1 denotes test data. If this is used, be sure to copy/paste values to freeze the random 0/1 numbers, since RAND is a volatile function which will recalculate every time Excel recalculates. The use of Test/Withhold Sample ID should be with large datasets, so while N=100 doesn't really qualify as large, we will use this for demonstration purposes and continuity in the example.

54. Click **Recall Last Dialog** (or press **F3**). Select *Customer Type* and click **<< Remove**. Select *Test ID* and click **Test/Withhold Sample ID >>**. We will use the default combo drop down selection = 1, which is used to specify what rows are assigned to the test/withhold sample. Check **Box-Cox Transformation** with **Rounded Lambda** option.

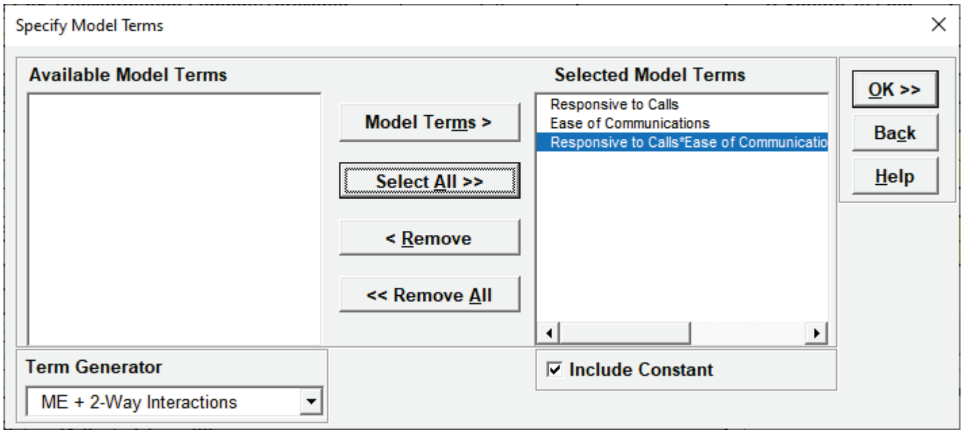


55. Click **Advanced Options**. Uncheck **Stepwise/Best Subsets Regression**.



Test/Withhold Sample ID and **Box-Cox Transformation** can be used with **Stepwise/Best Subsets Regression** but here we want to manually specify the model.

56. Click **OK**. Click **Next >>**. Select *ME + 2-Way Interactions*. Click **Select All >>**.



57. Click **OK >>**. The Test/Withhold Sample report is shown along with the Information Criteria and Validation.

Information Criteria and Validation		Test/Withhold Sample	
AICc	261.2901	ID/Column Level	Test ID = 1
BIC	271.2383	Test Sample %	34.00%
R-Square 10-Fold	91.02%	R-Square Test	92.65%
S 10-Fold	1.7714	S Test	1.3647

ID/Column Level indicates what text or numeric value in the Test ID column is used to specify the test sample. **Test Sample %** shows that the test sample comprises 34% of the data. **R-Square Test** = 92.65% and **S Test** = 1.365 are quite good, but note that they will vary if a different random Test ID is used.

58. Click on Sheet **MReg5 – Test**. Note, Sheet MReg# will increment every time a model is refitted. The detailed Test report is given:

Obs. No.	Responsive to Calls	Ease of Communications	Overall Satisfaction	Predicted Response	SE
4	1.88	2.52	2.7	2.484811867	0.68667515
6	4.31	3.93	4.12	4.109652881	0.477963686
13	4.29	4.77	4.53	4.483283119	0.567802001
17	5	4.03	4.02	4.474585718	0.558541005
18	1.11	2.97	2.04	2.13336235	0.736801836
25	3.73	4.88	4.31	4.2102128	0.606852881

Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI	Box-Cox Transformed Residuals
2.287297411	2.667742601	1.649085936	3.103078408	1.115709985
4.053712272	4.164842184	3.67889656	4.499357032	0.085153199
4.410823153	4.55459045	4.088298127	4.84618133	0.421072471
4.40435036	4.543735534	4.079237497	4.837732541	-3.861517344
1.861730207	2.374116777	1.027044765	2.836837832	-0.389634917
4.121859986	4.296749225	3.784389121	4.596757839	0.850208176

This gives the **Obs. No.** (Observation Number) for the **Test/Withhold Sample** data along with the predictor values, response values, **Predicted Response**, **SE**, **CI**, **PI** and **Box-Cox Transformed Residuals**. Residual plots are not provided but they can be created manually using SigmaXL's graphical tools.

Tip: If a row specified in the Test/Withhold Sample does not include a response value, it will still appear in this report with the **Predicted Response**, **SE**, **CI** and **PI** values given. It would be excluded from the **Test Sample %**, **R-Square Test** and **S Test** results.

Example: Advanced Multiple Regression with Box-Tidwell Test and Recommended Power Transformation – One X

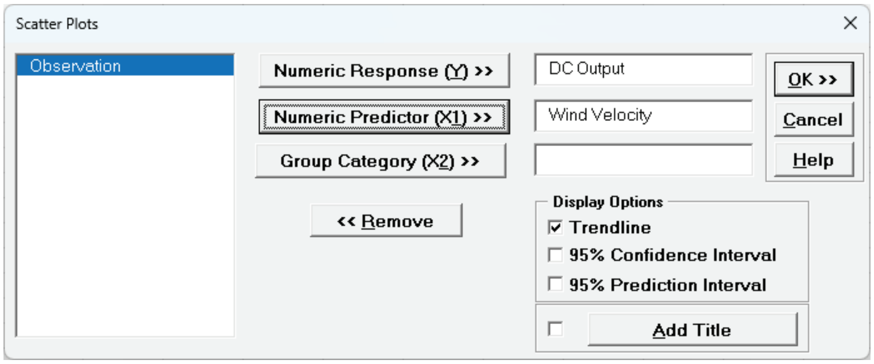
Multiple linear regression assumes that relationships between the predictors and the response variable are linear. The Box-Tidwell (BT) procedure aims to find an optimal power transformation of the predictor variables to satisfy the linearity assumption. This transformation can be crucial for improving the model fit and prediction accuracy. For details, see the Appendix: **Advanced Multiple Regression**. Note, in SigmaXL:

- At least one continuous predictor with all positive data values must be included in the model.
- Do not use standardization or coding as this will introduce 0 or negative values in the predictors.
- Continuous predictors with values <= 0, categorical factors, interactions and higher order terms are included in the model but excluded from the Box-Tidwell (BT) test and transformation.
- The constant must be included in the model.
- Box-Tidwell cannot be performed with Error df = 0.
- Box-Tidwell power transformations are calculated only for significant continuous predictors detected by the BT Test. This improves the overall robustness of the procedure.
- If Box-Cox is used, Box-Tidwell power transformations are calculated using the Box-Cox transformed response. Box-Cox Lambda may not be optimal after refit.
- Optimal and rounded power values are reported. Rounded is recommended for ease of interpretation.
- Power values of -5 or +5 are limits and considered unstable, so rounded is set to 1.
- Sheet BoxTidwell contains the original data with new columns for the transformed continuous predictors using rounded power. The model should be refit with these transformed predictors. If optimal power is desired, please use the Excel formula "=X^(Power)"; if Power = 0, use "=LN(X)".

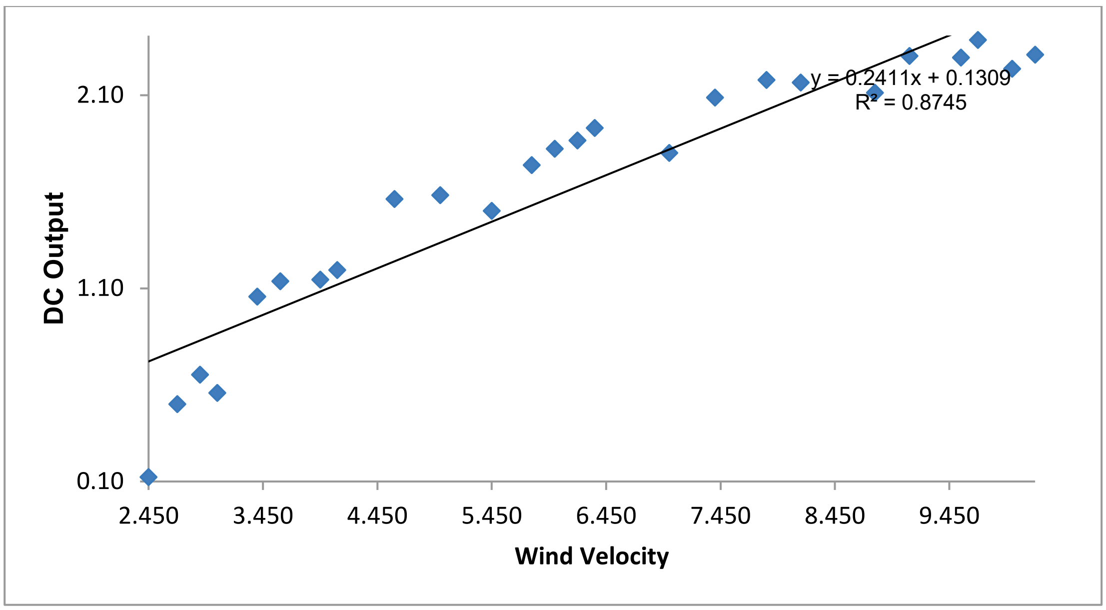
1. Open **Montgomery Table 5.5 Windmill Data.xlsx** (**Sheet 1** tab). This data is from Table 5.5, Montgomery, D.C., E.A. Peck and G.G. Vining (2021), *Introduction to Linear Regression Analysis*, 6th Edition, John Wiley & Sons. A research engineer is investigating the use of a windmill to generate electricity. He has collected data on the DC Output from his windmill and the corresponding wind velocity (mph).

2. First, we will create a Scatter Plot of the data with a Trendline. Click **SigmaXL > Graphical Tools > Scatter Plots**. If necessary, check **Use Entire Data Table**, click **Next.**

3. Select *DC Output*, click **Numeric Response (Y) >>**; select *Wind Velocity*, click **Numeric Predictor (X1) >>**. Check **Trendline** as shown:

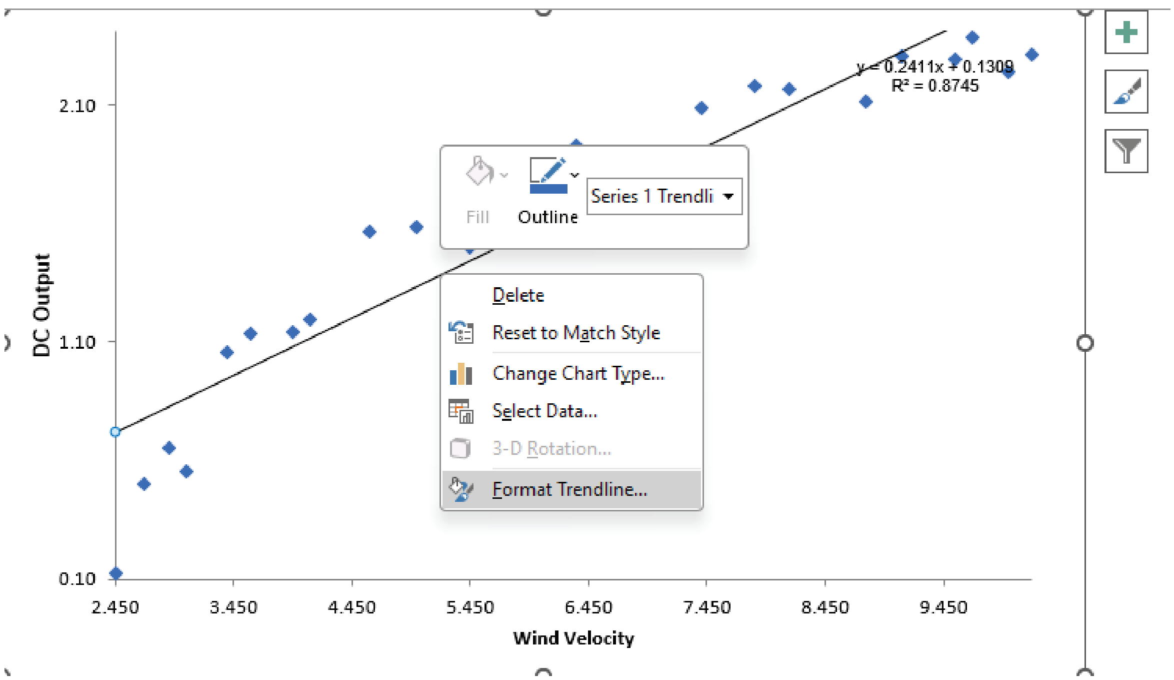


4. Click **OK**. The resulting Scatter Plot is shown with equation and trendline.

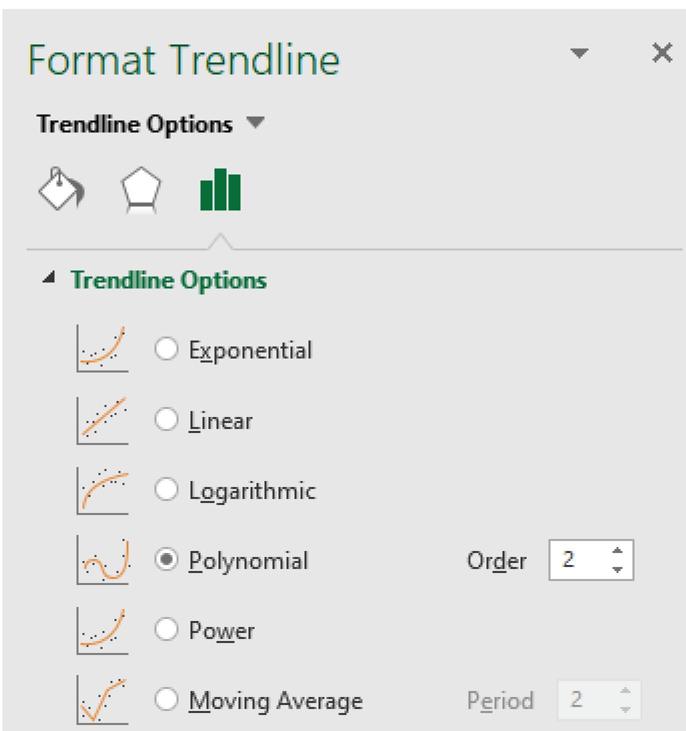


The equation is based on linear regression, using the method of least squares. R-squared * 100 is the percent variation of Y explained by X (here 87.45%). Curvature is apparent, so we will try to improve the fit by modifying the trendline in Excel to a quadratic fit.

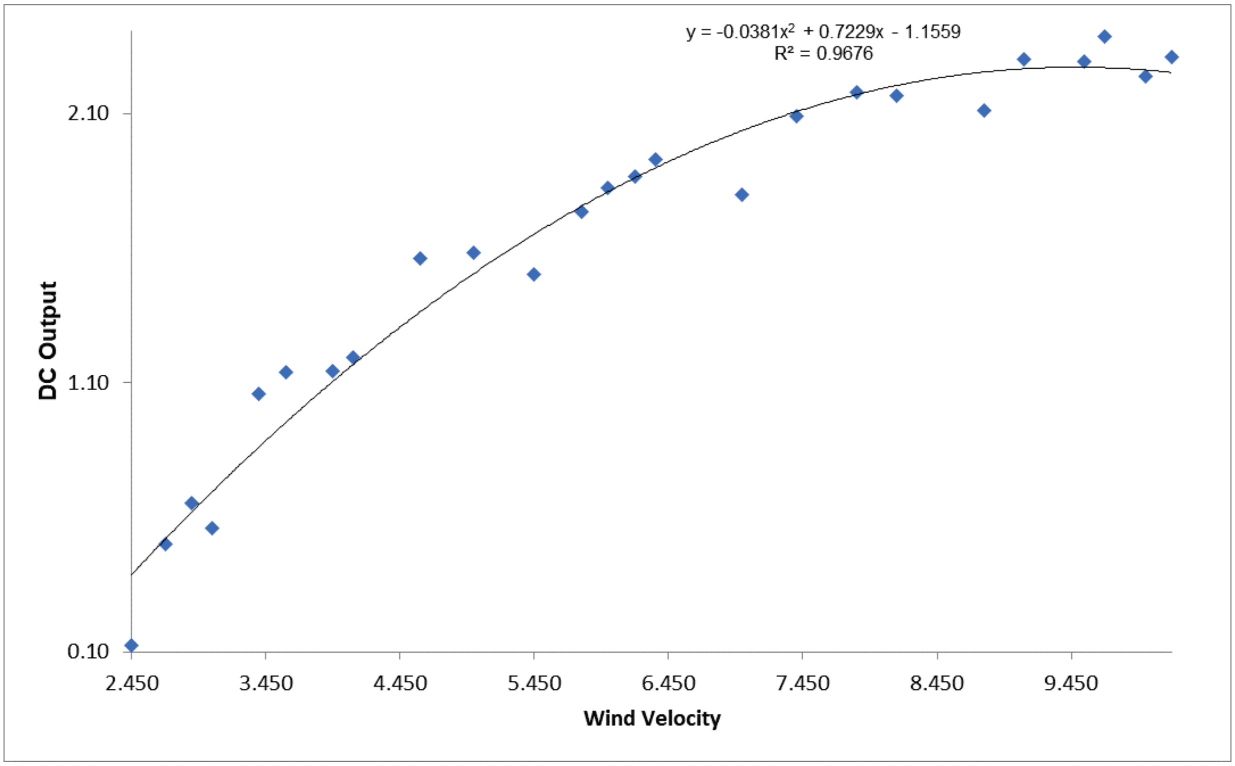
5. Click on the Trendline, right click and select Format Trendline as shown:



The **Format Trendline** options are given. Select **Polynomial** with **Order** 2 as shown.



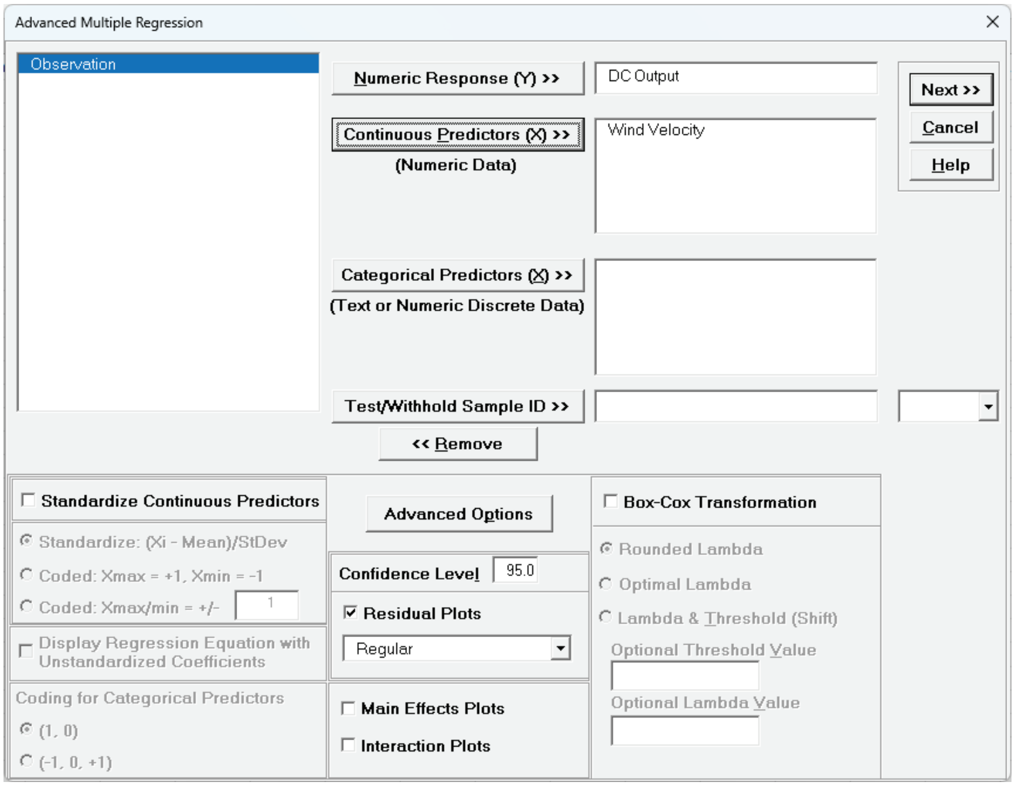
6. The Trendline is now a quadratic function as shown:



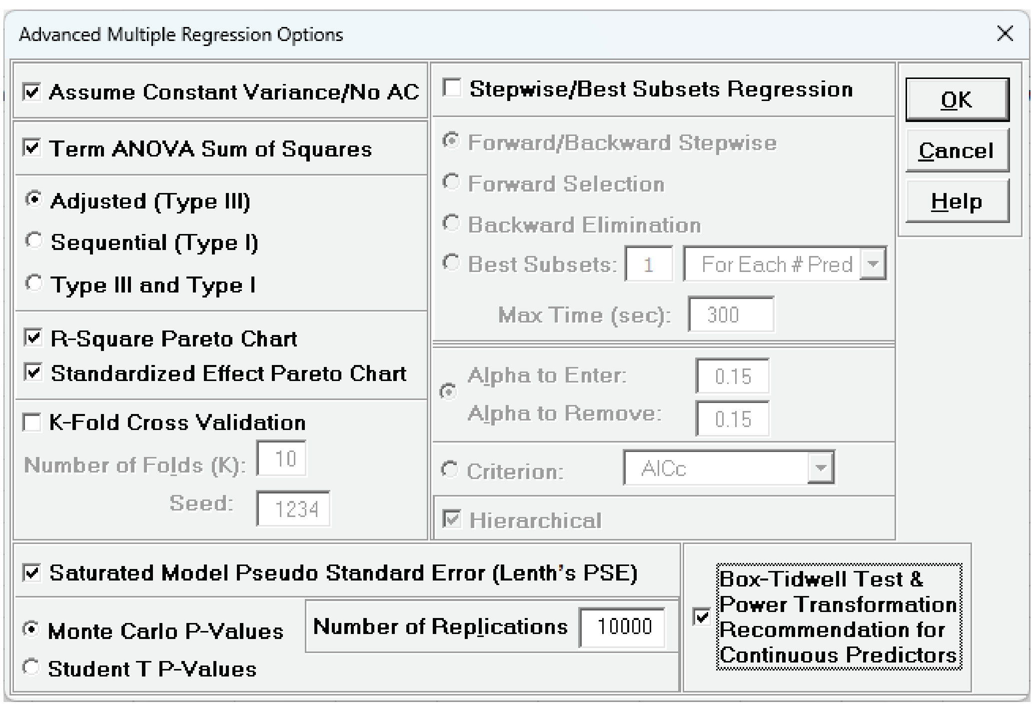
R-Square has improved from 87.45% to 96.76%. However, Montgomery et al. note that the quadratic model starts to bend down with higher wind velocity which is contrary to theory for a windmill operation. A cubic polynomial model could be used to improve on the quadratic (giving R-square = 97.69%), or nonlinear curve fitting using Excel's Solver, but we will now consider a model using the Box-Tidwell power transformation.

7. Click on the **Sheet 1** tab. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, check **Use Entire Data Table**, click **Next.**

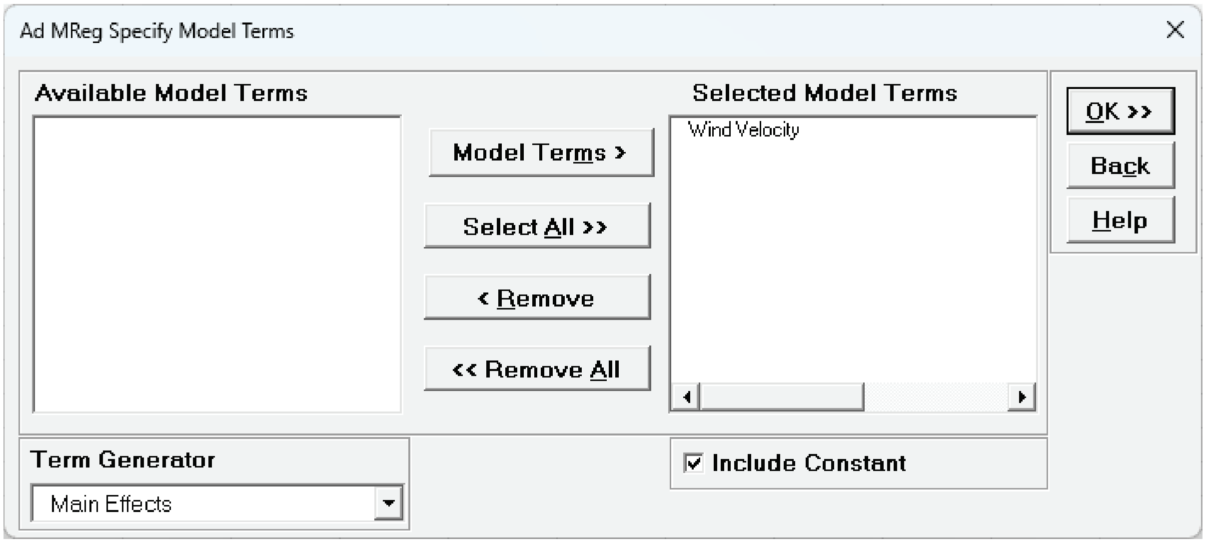
8. Select *DC Output*, click **Numeric Response (Y) >>**; select *Wind Velocity*, click **Continuous Predictors (X) >>** with **Residual Plots** checked as shown:



9. Click **Advanced Options**. Check **Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors**.



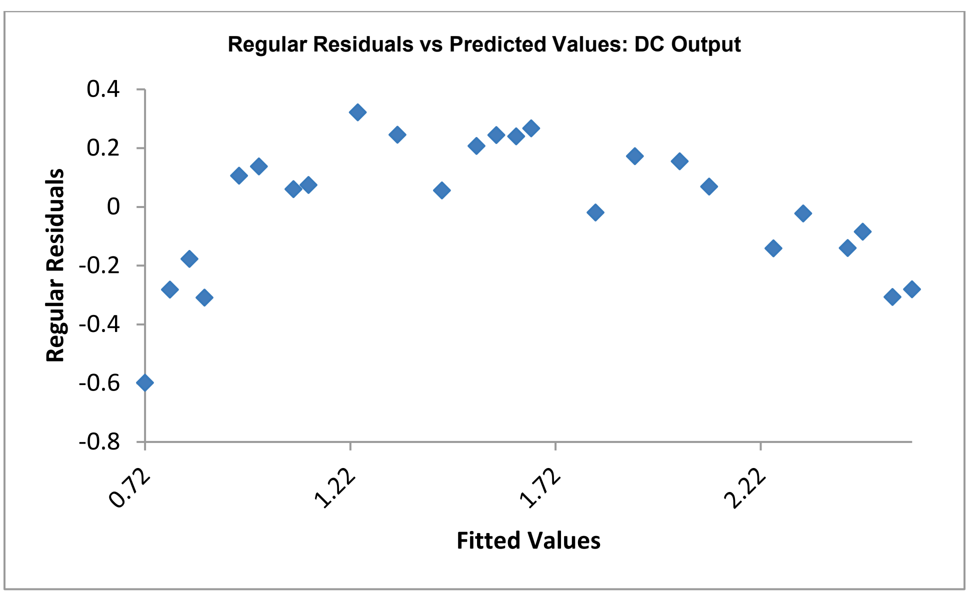
10. Click **OK**. Click **Next >>**. Select *Wind Velocity*. Click **Model Terms >>**.



11. Click **OK >>**. The Model Summary table shows the R-Square = 87.45% as we saw in the Scatter Plot Trendline equation. R-Square Predicted (Leave-One-Out Cross-Validation) is 84.58%.

Model Summary	
R-Square	87.45%
R-Square Adjusted	86.90%
R-Square Predicted	84.58%
S (Root Mean Square Error)	0.2361

12. Clicking on the **Residuals DC Output**, the Residuals vs Predicted Values plot shows strong curvature, indicating that the simple linear regression model is inadequate.



As noted above, we could refit this as a quadratic or cubic model (to do so, click **Recall SigmaXL Dialog**, **Next >>**, **Term Generator** *All up-to 3-Way*, Select **Model Terms**, click **OK>>**), but we will now examine the Box-Tidwell report.

13. Click the **MReg1 Model** Sheet tab and scroll down to view the **Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors** report:

Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors			
H0: Power is equal to 1; Ha: Power is not equal to 1.			
Predictor Term	P-Value	Optimal Power	Rounded Power
Wind Velocity	0.0000	-0.8333	-1

Given the low P-Value (6.13E-09), we reject the null hypothesis H0: Power = 1. The Box-Tidwell Optimal Power = -0.833 and Rounded Power = -1 (i.e., a reciprocal transformation of 1/X is recommended). This agrees with the Box-Tidwell power transformation given in Montgomery et al. We will discuss the calculation details later in this section.

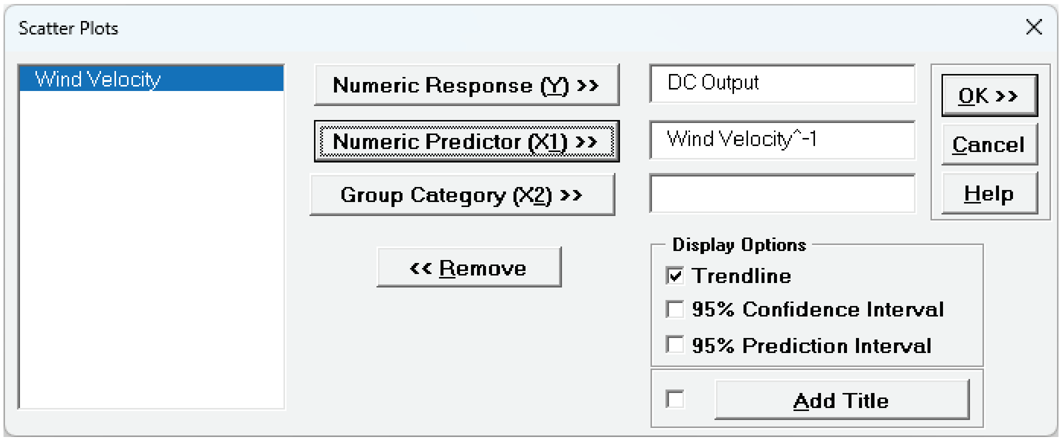
Report Notes:
Box-Tidwell power transformations are calculated for significant continuous predictors.
Sheet BoxTidwell contains the original data with new columns for the transformed continuous predictors using rounded power. The model should be refit with these transformed predictors.
If optimal power is desired, please use the Excel formula "=X^(Power)"; if Power = 0, use "=LN(X)".

14. Click on the Sheet **MReg1 – BoxTidwell DC Output**. The following table with BT Power transformation of Wind Velocity^-1 is given:

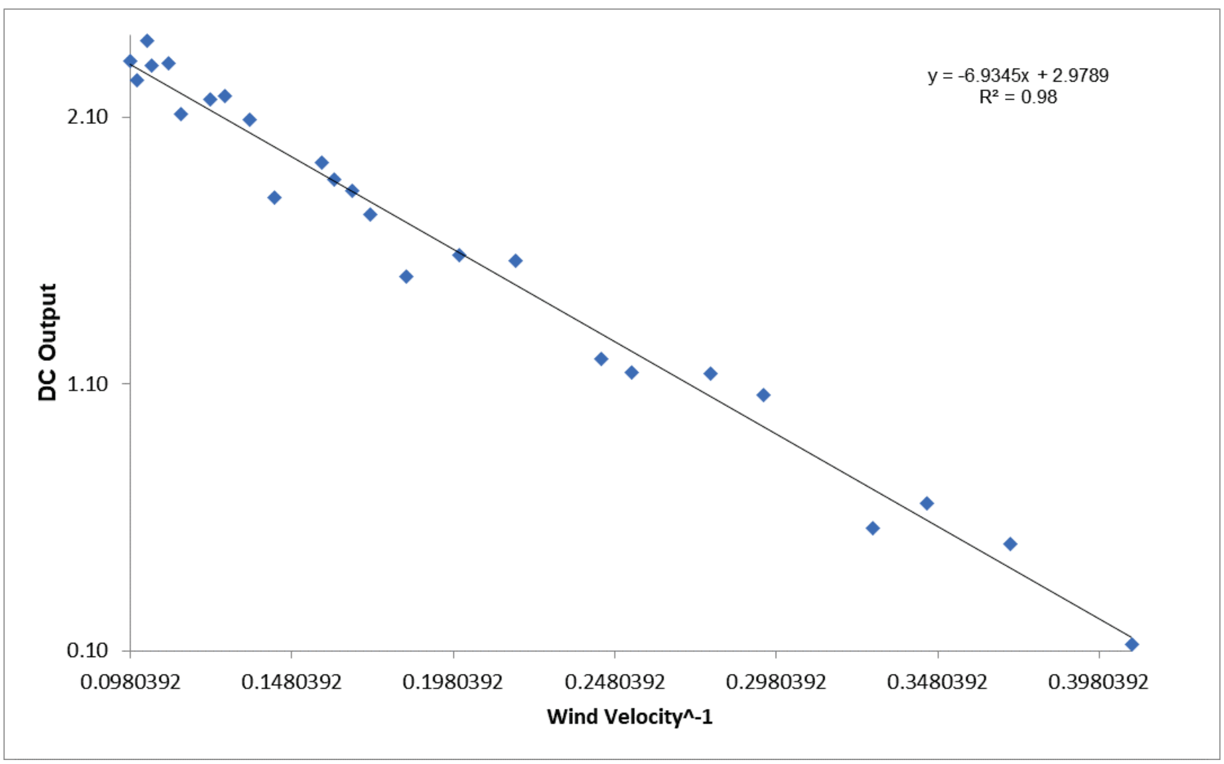
DC Output	Wind Velocity	Wind Velocity^-1
1.582	5	0.2
1.822	6	0.166666667
1.057	3.4	0.294117647
0.5	2.7	0.37037037
2.236	10	0.1
2.386	9.7	0.103092784
2.294	9.55	0.104712042
0.558	3.05	0.327868852
2.166	8.15	0.122699387
1.866	6.2	0.161290323
0.653	2.9	0.344827586
1.93	6.35	0.157480315
1.562	4.6	0.217391304
1.737	5.8	0.172413793
2.088	7.4	0.135135135
1.137	3.6	0.277777778
2.179	7.85	0.127388535
2.112	8.8	0.113636364
1.8	7	0.142857143
1.501	5.45	0.183486239
2.303	9.1	0.10989011
2.31	10.2	0.098039216
1.194	4.1	0.243902439
1.144	3.95	0.253164557
0.123	2.45	0.408163265

15. Now we will create a Scatter Plot with Trendline for the transformed Wind Velocity data. Click **SigmaXL > Graphical Tools > Scatter Plots**. If necessary, check **Use Entire Data Table**, click **Next.**

16. Select *DC Output*, click **Numeric Response (Y) >>**; select *Wind Velocity^-1*, click **Numeric Predictor (X1) >>**. Check **Trendline** as shown:



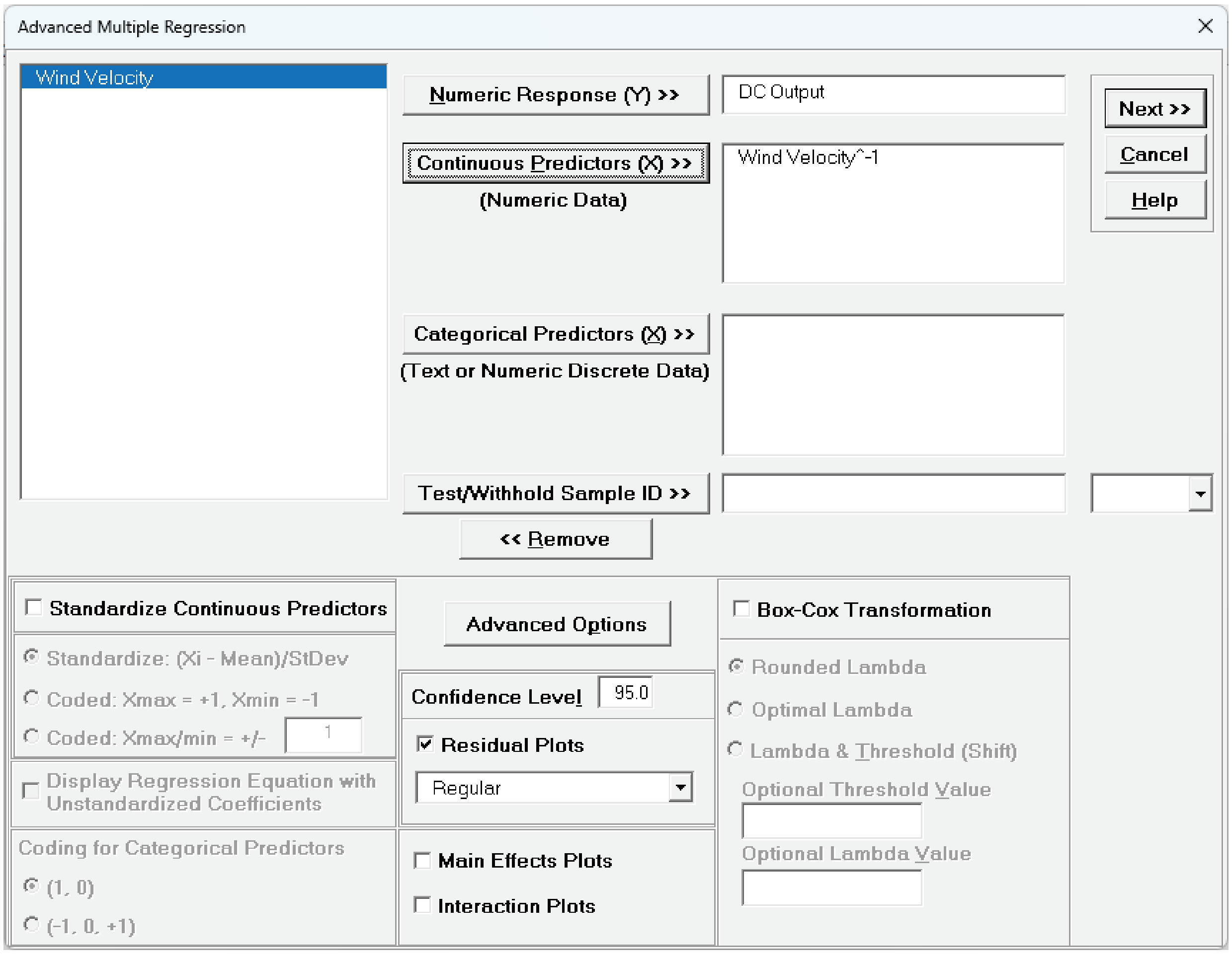
17. Click **OK**. The resulting Scatter Plot is shown with equation and trendline.



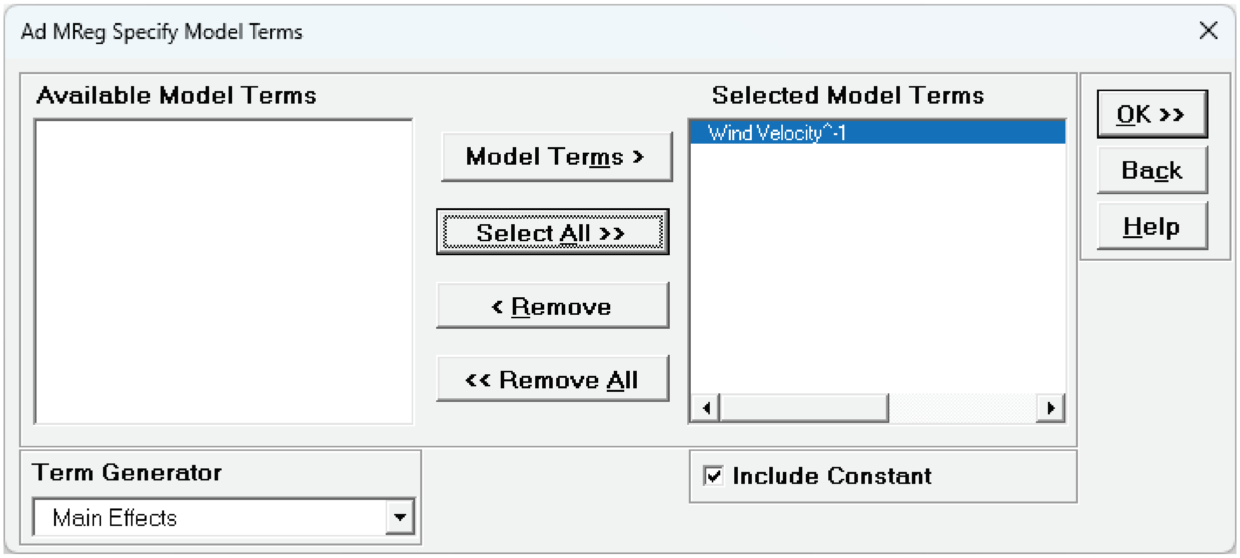
The Box-Tidwell Power Transformation has successfully linearized the relationship between DC Output and Wind Velocity. The straight-line fit looks good and R-Square * 100 is now 98% versus the original untransformed 87.5% (or the quadratic with 96.8%). While the model fit is dramatically improved, the slope is now negative due to the reciprocal relationship, making the interpretation less intuitive. An inverse transformation is required to revert to the original units.

18. Click on the Sheet **MReg1 – BoxTidwell DC Output**. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, check **Use Entire Data Table**, click **Next.**

19. Select *DC Output*, click **Numeric Response (Y) >>**; select *Wind Velocity*, click **Continuous Predictors (X) >>** with **Residual Plots** checked as shown:



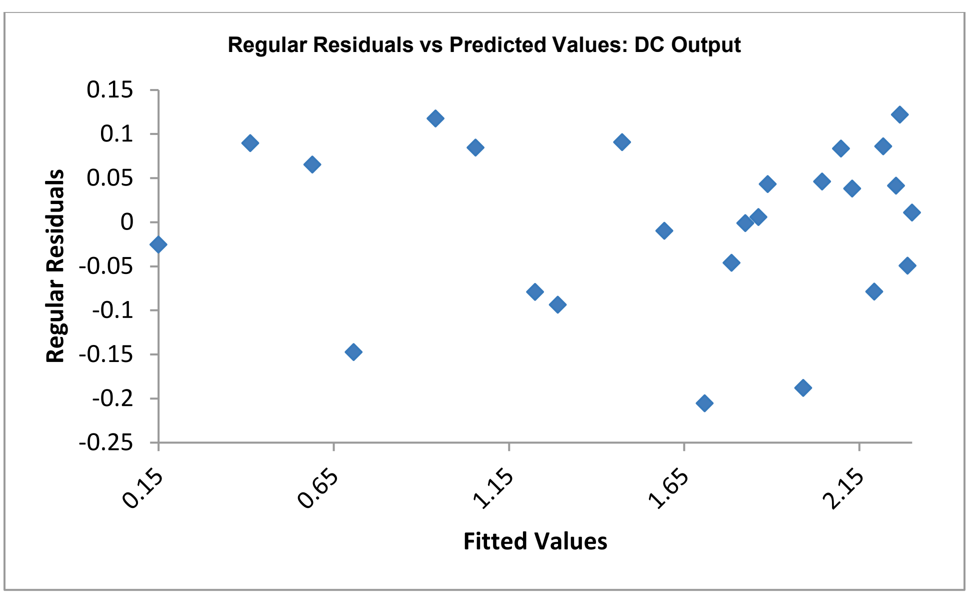
20. Click **Next >>**. Select *Wind Velocity^-1*. Click **Model Terms >>**.



21. Click **OK >>**. The Model Summary table shows the R-Square = 98.0% as we saw in the Scatter Plot Trendline equation. R-Square Predicted (Leave-One-Out Cross-Validation) is 97.67% (compared to the original untransformed 84.58%).

Model Summary	
R-Square	98.00%
R-Square Adjusted	97.92%
R-Square Predicted	97.67%
S (Root Mean Square Error)	0.0942

22. Clicking on the **Residuals DC Output**, the Residuals vs Predicted Values plot shows no obvious pattern, indicating that the BT transformed model has successfully removed the nonlinearity in the model.



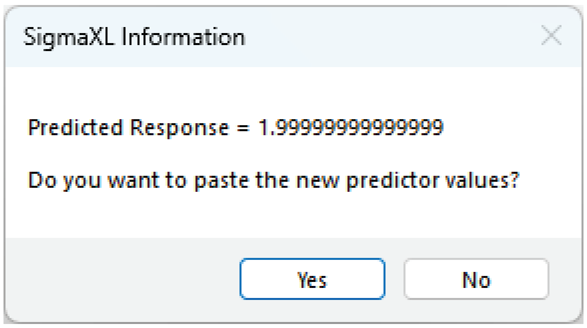
23. Now we will consider optimization with the transformed model. Click the **Model** Sheet tab. Scroll to the Predicted Response Calculator. Enter a Target = 2.

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Wind Velocity^-1	0.197454874	1.6096					

Goal:	Target
Target:	2

Optimize

24. Click **Optimize**.



Click **Yes**.

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Wind Velocity^-1	0.141157039	2	0.022131317	1.954217884	2.045782116	1.799884338	2.200115662

25. The Wind Velocity^-1 setting = 0.1412 gives a DC Output = 2.

26. In order to determine the original Wind Velocity, an inverse transform must be applied. To retrieve the original value of X from $X_{\text{transformed}}$, we use the following inverse transformation formula:

If $BT\ Power$ is not zero, the inverse formula is:

$$X = (X_{\text{transformed}})^{1/(\text{BT Power})}$$

If $BT\ Power$ is zero, implying a logarithmic transformation was applied ($\ln(X)$), then the inverse transformation is:

$$X = \exp(X_{\text{transformed}})$$

27. Click the mouse cursor in an empty cell and use the following formula (note that the X value at cell **K14** references the Excel Range Name _Wind_Velocity__1_ for use in the prediction equation at cell **L14**):

= K14^(1/-1)

=_Wind_Velocity__1_^(1/-1)

Gives the Wind Velocity = 7.084.

In this case, we could have also simply used the formula =1/K14 since it is a reciprocal transform.

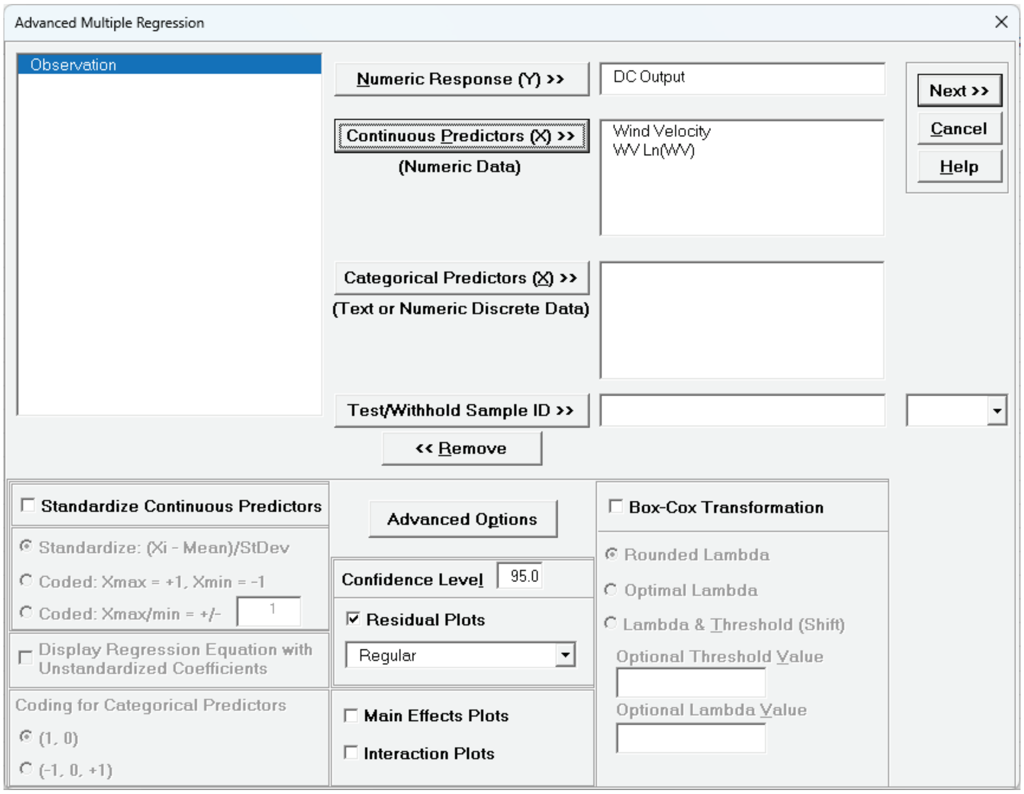
28. Finally, we will demonstrate the steps to compute the Box-Tidwell P-Value and initial estimate of the BT power transformation (see Appendix: Advanced Multiple Regression):

- **Original Regression Model**:
  - Fit a linear regression model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
  - $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated coefficients for the intercept and the slope of X, respectively.
- **Extended Regression Model with Interaction Term**:
  - Enhance the original model by adding an "interaction" term $X\ln(X)$: $\hat{Y} = \hat{\beta}_0^* + \hat{\beta}_1^* X + \hat{\beta}_2^*(X\ln(X))$
  - $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and $\hat{\beta}_2^*$ are the adjusted coefficients in the extended model, where $\hat{\beta}_2^*$ is specifically for the interaction term.
- **Statistical Testing**:
  - Perform a hypothesis test on $\hat{\beta}_2^*$.
  - The null hypothesis $H_0: \hat{\beta}_2^* = 0$ tests if the interaction term is necessary for improving model linearity.
- **Estimate** $BT\ Power$:
  - If the null hypothesis is rejected, suggesting a transformation is beneficial, compute $BT\ Power$ using: $BT\ Power_{\text{estimated}} = 1 + \frac{\hat{\beta}_2^*}{\hat{\beta}_1}$
- **Iterate**:
  - This transformation is applied to X and the above steps are iterated until $BT\ Power_{\text{estimated}}$ converges, but we will only consider the initial estimate here.

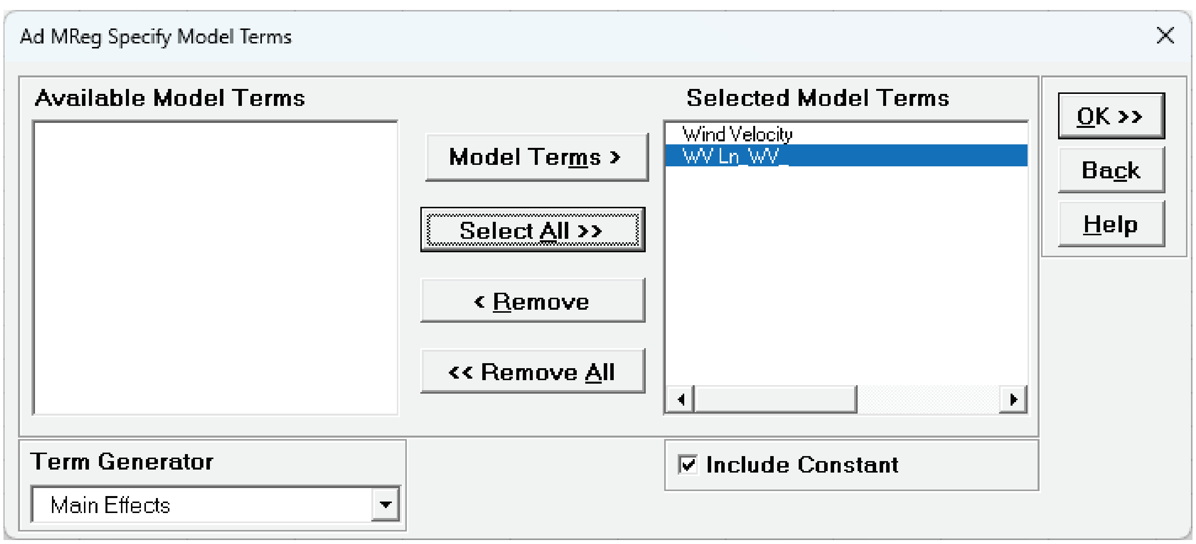
29. Since we already have the initial model, we will now construct the extended model. Click the **Sheet 2** tab which includes an extra column for Wind Velocity*Ln(Wind Velocity), denoted as WV Ln(WV).

30. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, check **Use Entire Data Table**, click **Next.**

31. Select *DC Output*, click **Numeric Response (Y) >>**; select *Wind Velocity* and WV Ln(WV) click **Continuous Predictors (X) >>** as shown:



32. Click **Next >>**. Click **Select All >>**.



Note the parenthesis characters are not permitted in the model so they are automatically changed to underscores. The parenthesis characters are reserved to denote nesting in a model.

33. Click **OK >>**.

Multiple Regression Model: DC Output = (-2.41684) + (1.53443)*Wind_Velocity + (-0.462596)*WV_Ln_WV_

Model Summary	
R-Square	97.38%
R-Square Adjusted	97.14%
R-Square Predicted	96.48%
S (Root Mean Square Error)	0.1103

Model Information	
Continuous Predictor Standardization/Coding	N/A
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	-2.416843762	0.285115225	-8.4767	0.0000		
Wind Velocity	1.534434869	0.141893885	10.8140	0.0000	254.2684	0.0039
WV Ln_WV_	-0.462596342	0.050654217	-9.1324	0.0000	254.2684	0.0039

The P-Value for the Interaction Predictor Term WV Ln_WV__ (Wind Velocity * Ln(Wind Velocity)) is the BT P-Value = 6.13E-09 (cell **F16**), which agrees with the BT P-Value reported earlier.

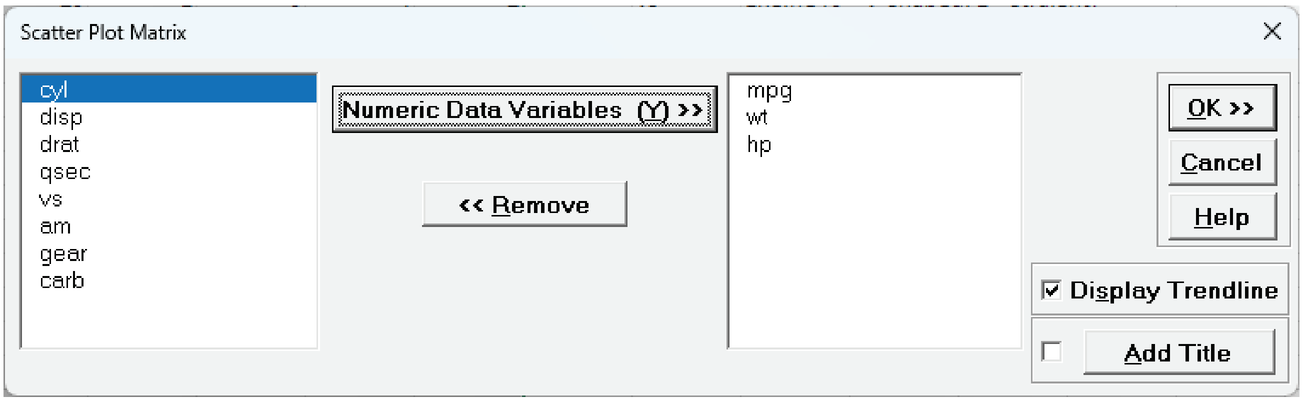
34. The original coefficient estimate for Wind Velocity was 0.2411. The initial estimate for BT Power is then 1 + (-.4626/0.2411) = -0.92, which is close to the final BT Power optimal value of -.8333.

Example: Advanced Multiple Regression with Box-Tidwell Test and Recommended Power Transformation – Two X's

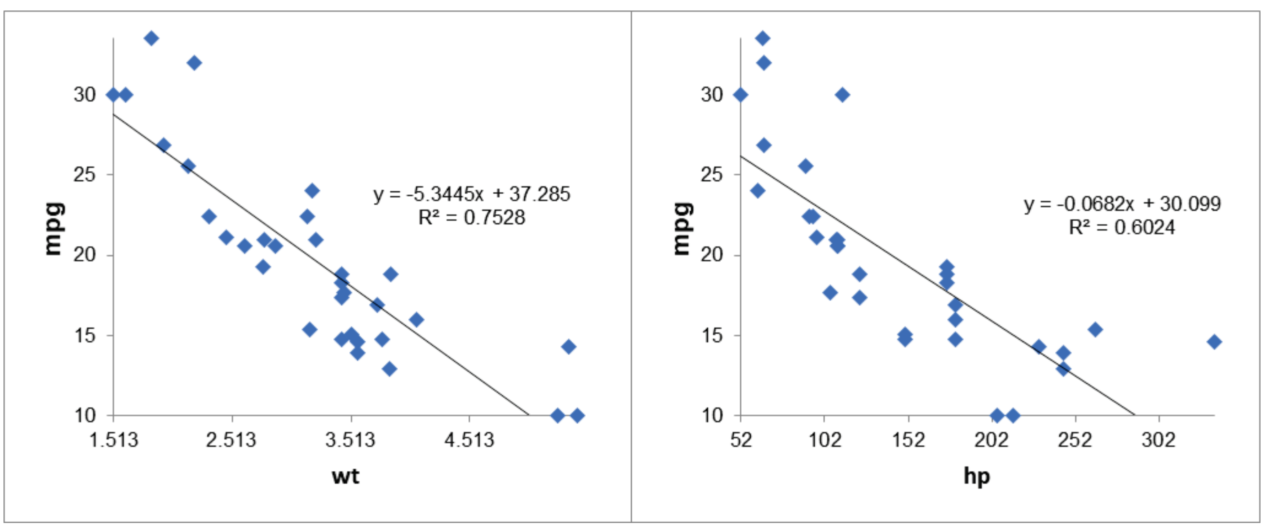
1. Open **MT Cars.xlsx** (**mtcars** tab). The data was extracted from the 1974 Motor Trend US magazine and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). The response of interest is mpg (Miles/US gallon). The continuous predictors to be considered here are wt (Weight (1000 lbs)) and hp (Gross horsepower).

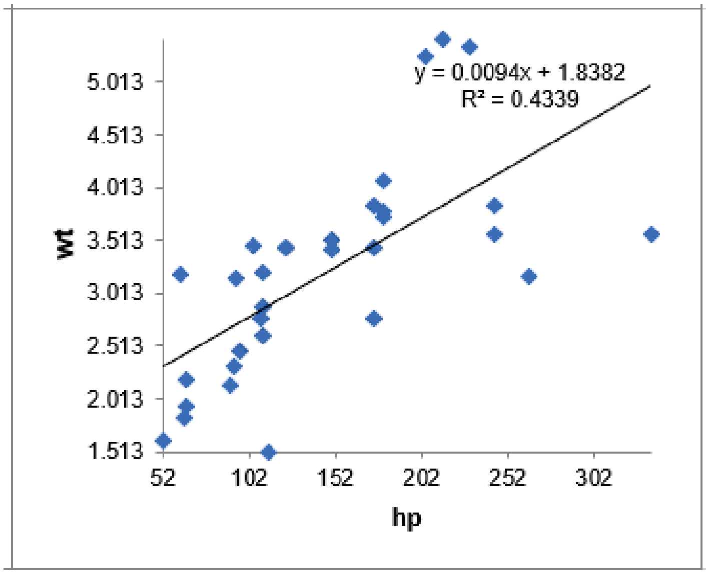
2. First, we will create a Scatter Plot Matrix of the data. Click **SigmaXL > Graphical Tools > Scatter Plot Matrix**. If necessary, check **Use Entire Data Table**, click **Next.**

3. Select *mpg, wt, hp*, click **Numeric Data Variables (Y) >>**. Check **Display Trendline** as shown:



4. Click **OK**. The mpg vs wt, mpg vs hp and wt vs hp are given as:

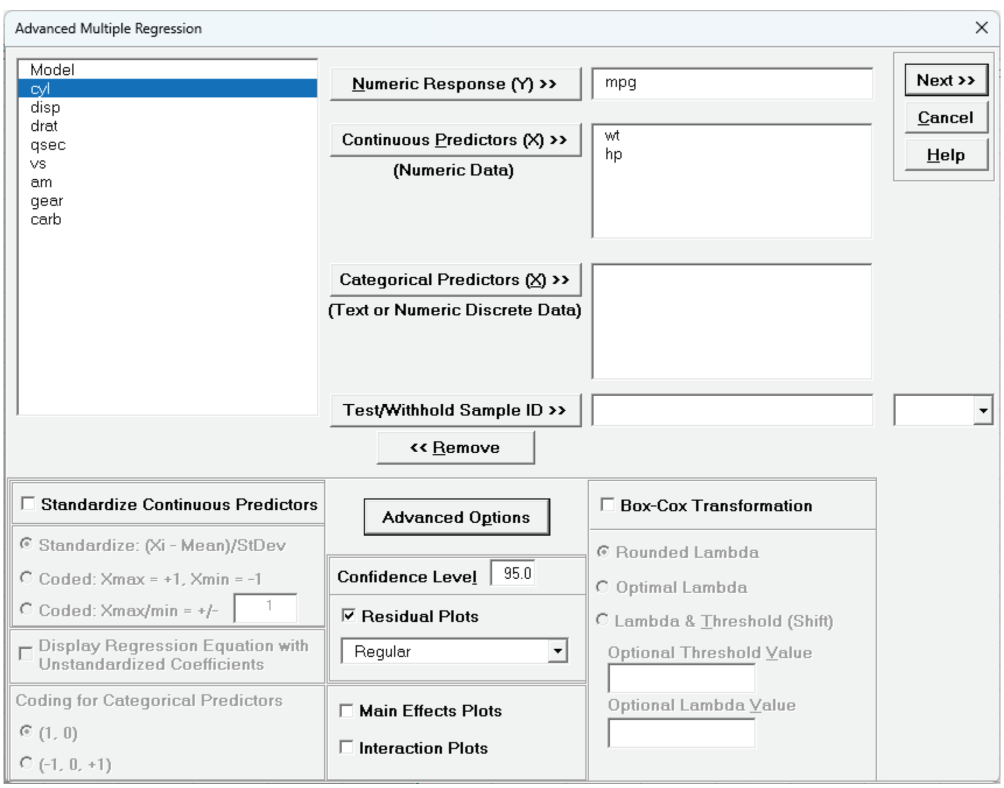




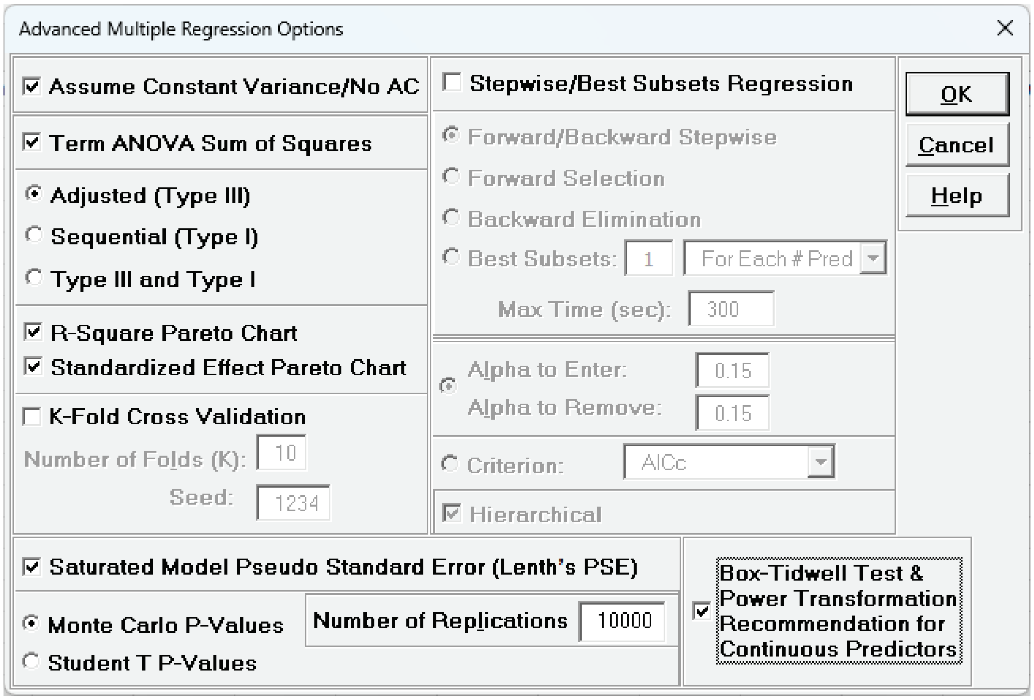
Some curvature is apparent in mpg vs wt and mpg vs hp. The wt vs hp plot also shows that the predictors are correlated, so we anticipate that VIF will be > 1.

5. Click on the **mtcars** tab. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, check **Use Entire Data Table**, click **Next.**

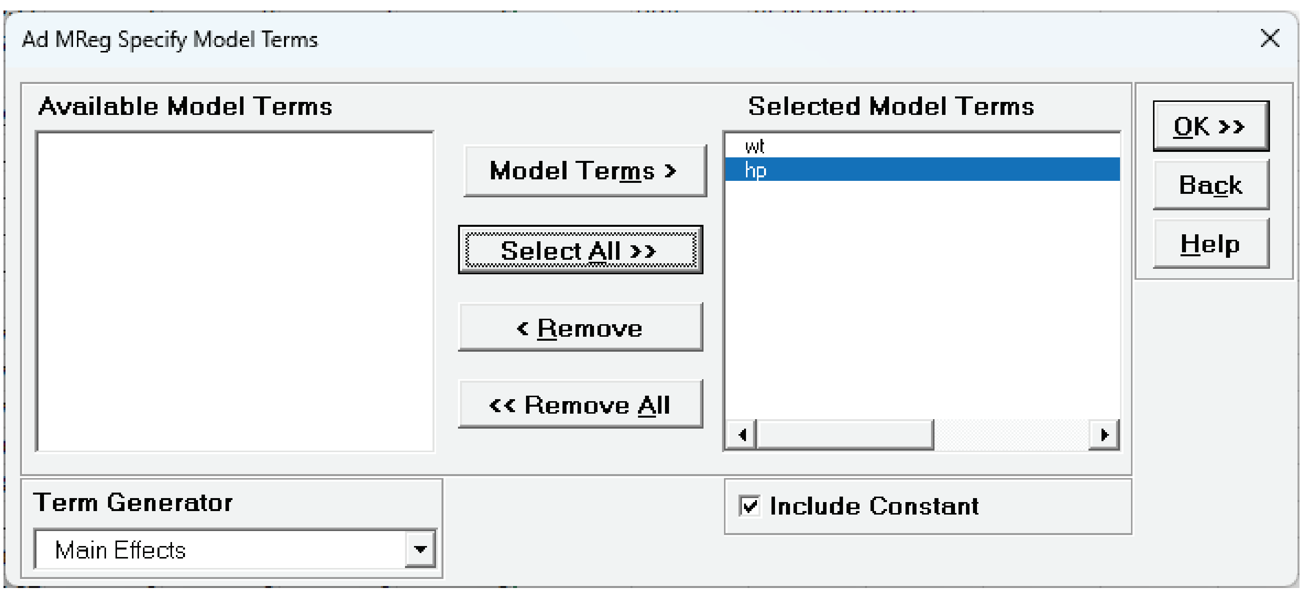
6. Select *mpg*, click **Numeric Response (Y) >>**; select *wt, hp*, click **Continuous Predictors (X) >>** with **Residual Plots** checked as shown:



7. Click **Advanced Options**. Check **Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors**.



8. Click **OK**. Click **Next >>**. Click **Select All >>**.



9. Click **OK >>**.

Multiple Regression Model: mpg = (37.2273) + (-3.87783)*wt + (-0.0317729)*hp

Model Summary	
R-Square	82.68%
R-Square Adjusted	81.48%
R-Square Predicted	78.11%
S (Root Mean Square Error)	2.5934

Model Information	
Continuous Predictor Standardization/Coding	N/A
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	37.22727012	1.598787538	23.2847	0.0000		
wt	-3.877830742	0.632733494	-6.1287	0.0000	1.7666	0.5661
hp	-0.031772947	0.00902971	-3.5187	0.0015	1.7666	0.5661

The Model Summary table shows the R-Square = 82.68% and R-Square Predicted (Leave-One-Out Cross-Validation) is 78.11%. The VIF scores are 1.77 so moderate but still acceptable at less than 5.

10. Scroll down to view the **Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors** report:

Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors			
H0: Power is equal to 1; Ha: Power is not equal to 1.			
Predictor Term	P-Value	Optimal Power	Rounded Power
wt	0.0112	-0.4174	-0.5
hp	0.0250	-0.5682	-0.5

Given the low P-Values for wt and hp, we reject the null hypothesis H0: Power = 1 for each predictor. The BT Rounded Power is -0.5 for wt and hp.

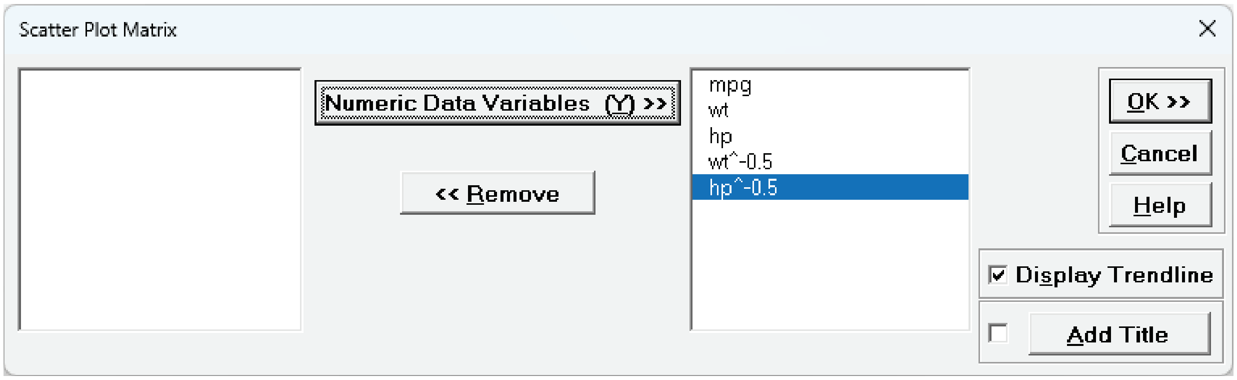
Report Notes:
Box-Tidwell power transformations are calculated for significant continuous predictors.
Sheet BoxTidwell contains the original data with new columns for the transformed continuous predictors using rounded power. The model should be refit with these transformed predictors.
If optimal power is desired, please use the Excel formula "=X^(Power)"; if Power = 0, use "=LN(X)".

11. Click on the Sheet **MReg1 – BoxTidwell mpg**. The following table with BT Power transformation of wt^-0.5 and hp^-0.5 is given:

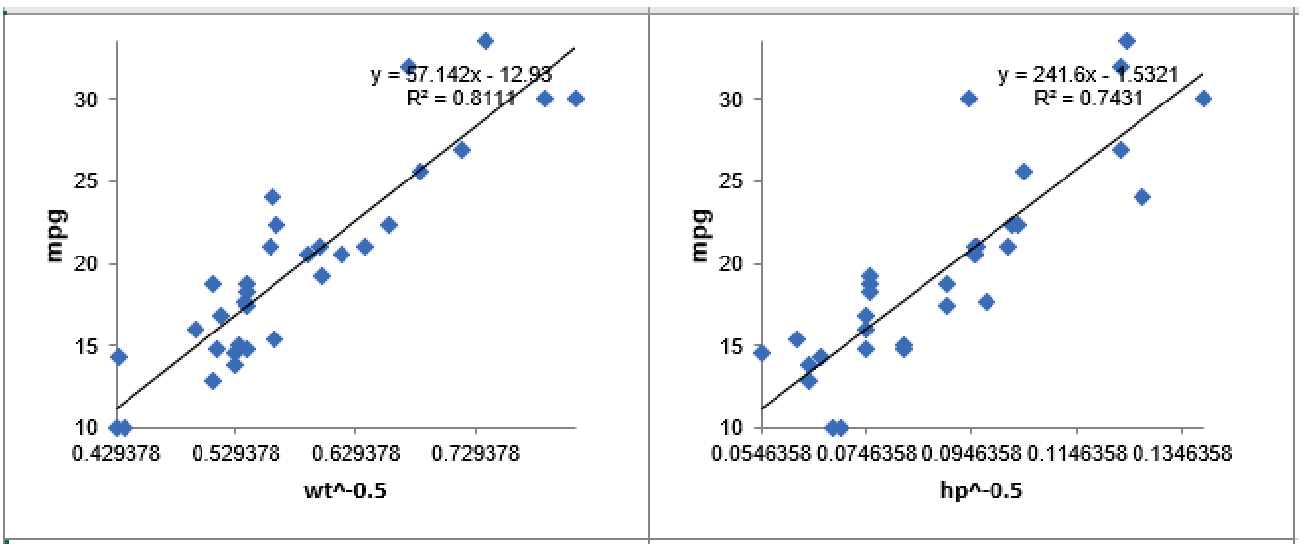
mpg	wt	hp	wt^-0.5	hp^-0.5
21	2.62	110	0.617802063	0.095346259
21	2.875	110	0.589767825	0.095346259
22.8	2.32	93	0.656532164	0.103695169
21.4	3.215	110	0.557711387	0.095346259
18.7	3.44	175	0.539163866	0.075592895
18.1	3.46	105	0.537603331	0.097590007
14.3	3.57	245	0.529256124	0.063887656
24.4	3.19	62	0.559892511	0.127000127
22.8	3.15	95	0.56343617	0.102597835
19.2	3.44	123	0.539163866	0.090166963
17.8	3.44	123	0.539163866	0.090166963
16.4	4.07	180	0.495681597	0.074535599
17.3	3.73	180	0.517780373	0.074535599
15.2	3.78	180	0.5143445	0.074535599
10.4	5.25	205	0.43643578	0.06984303
10.4	5.424	215	0.429378367	0.068199434
14.7	5.345	230	0.43253987	0.065938047
32.4	2.2	66	0.674199862	0.123091491
30.4	1.615	52	0.786889475	0.138675049
33.9	1.835	65	0.738213471	0.124034735
21.5	2.465	97	0.636929755	0.101534617
15.5	3.52	150	0.533001791	0.081649658
15.2	3.435	150	0.539556128	0.081649658
13.3	3.84	245	0.510310363	0.063887656
19.2	3.845	175	0.509978454	0.075592895
27.3	1.935	66	0.718885155	0.123091491
26	2.14	91	0.683585927	0.104828484
30.4	1.513	113	0.812981262	0.094072087
15.8	3.17	264	0.561655956	0.061545745
19.7	2.77	175	0.600841768	0.075592895
15	3.57	335	0.529256124	0.054635836
21.4	2.78	109	0.599760144	0.095782629

12. Now we will create a Scatter Plot Matrix of the transformed data. Click **SigmaXL > Graphical Tools > Scatter Plot Matrix**. If necessary, check **Use Entire Data Table**, click **Next.**

13. Select *mpg* to *hp^-0.5*, click **Numeric Data Variables (Y) >>.** Check **Display Trendline** as shown:



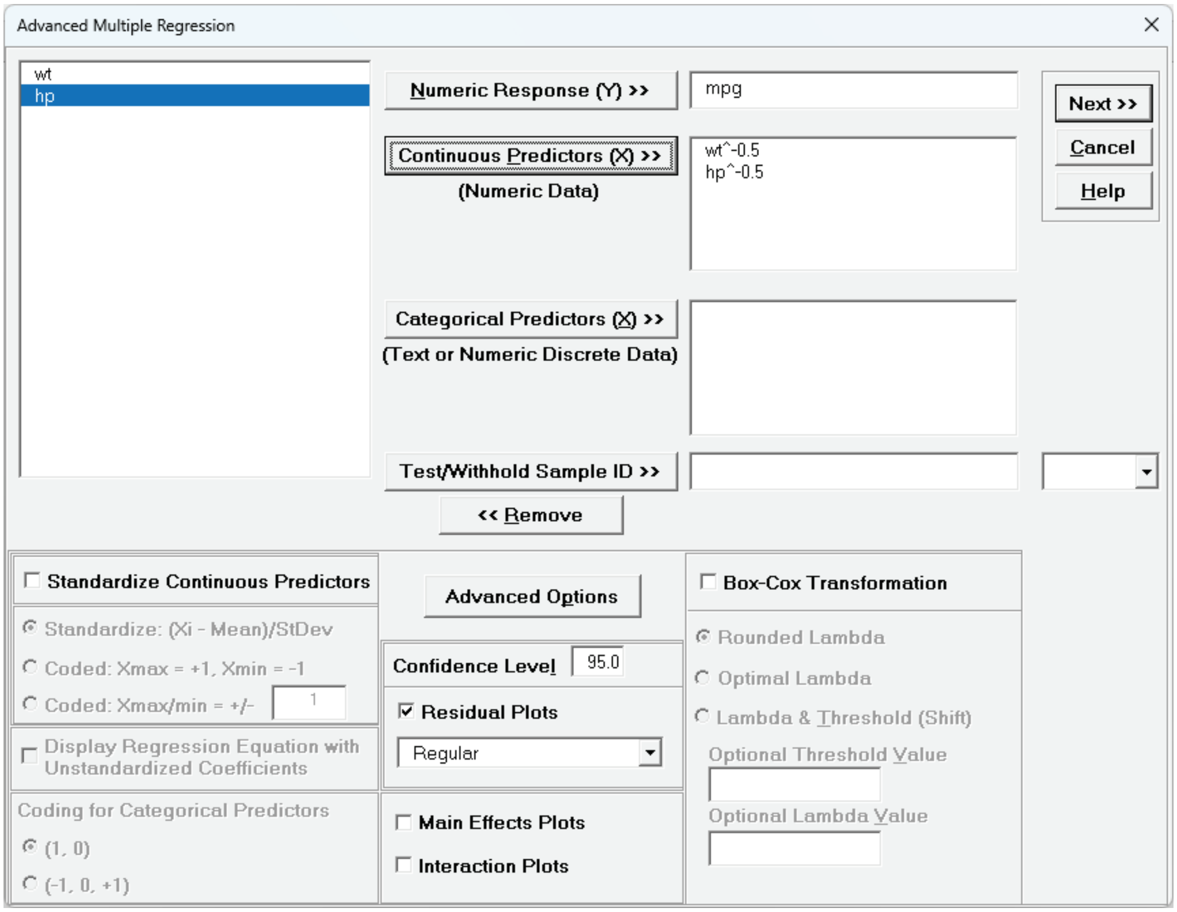
14. Click **OK**. The mpg vs wt^-0.5 and mpg vs hp^-0.5 are given as:



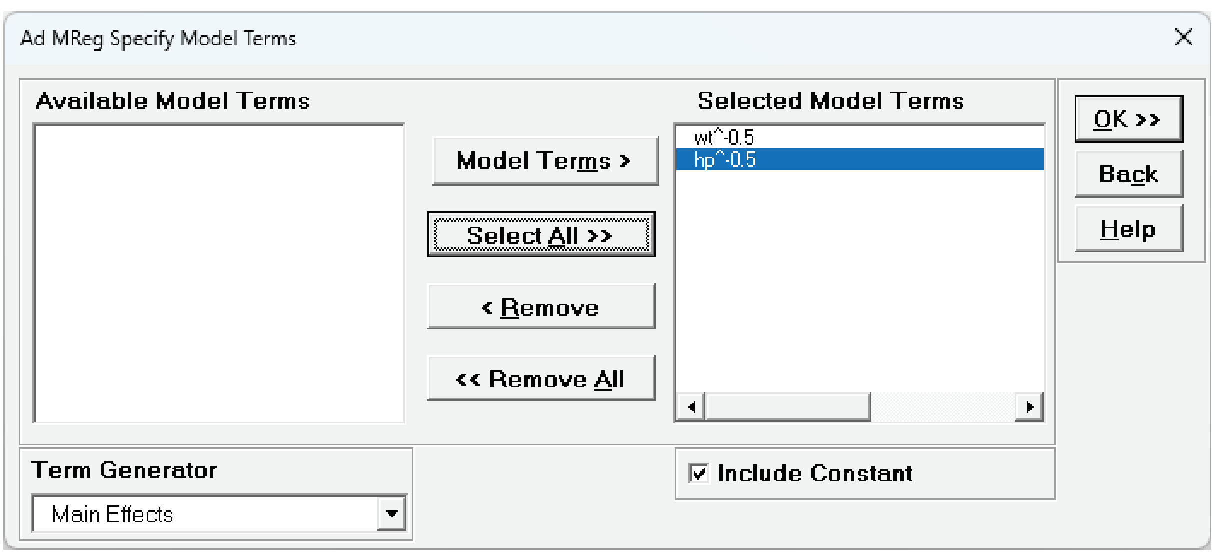
The Box-Tidwell Power Transformation has successfully linearized the relationship between mpg vs wt and mpg vs hp. R-Square * 100 is now 81.1% versus the original untransformed 75.3% for wt and 74.3% vs 60.2% for hp. While the model fits are dramatically improved, the slope is now positive due to the reciprocal SQRT relationship, making the interpretation less intuitive. An inverse transformation is required to revert to the original units.

15. Click on the Sheet **MReg1 – BoxTidwell mpg**. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, check Use Entire Data Table, click Next.

16. Select *mpg*, click **Numeric Response (Y) >>**; select *wt^-0.5* and *hp^-0.5*, click **Continuous Predictors (X) >>** with **Residual Plots** checked as shown:



17. Click **Next >>**. Click **Select All >>**.



18. Click **OK >>**.

Multiple Regression Model: mpg = (-12.0626) + (36.8451)*wt__0_5 + (121.36)*hp__0_5

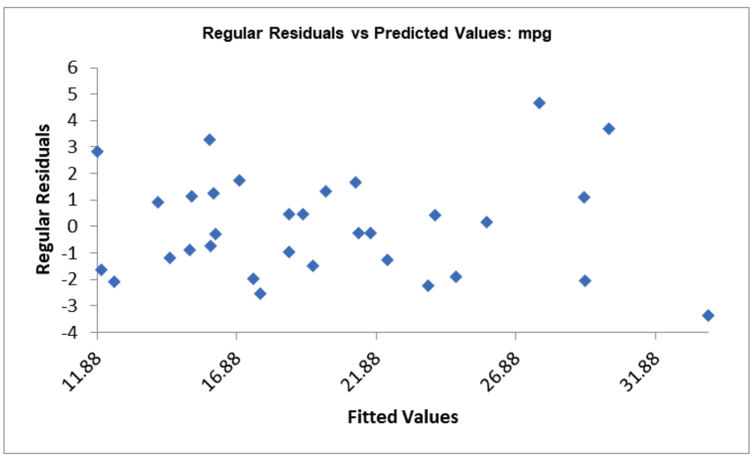
Model Summary	
R-Square	89.63%
R-Square Adjusted	88.91%
R-Square Predicted	86.91%
S (Root Mean Square Error)	2.0071

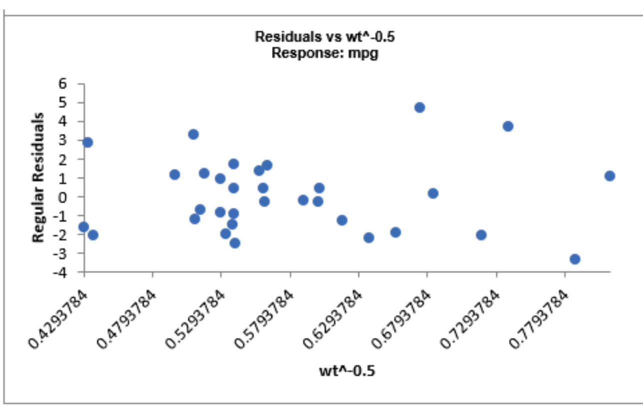
Model Information	
Continuous Predictor Standardization/Coding	N/A
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

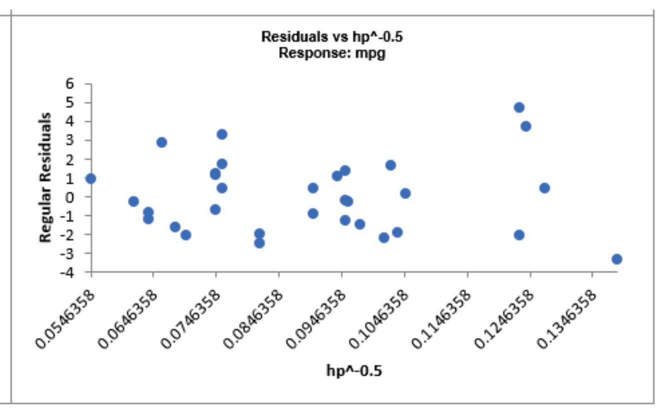
Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	-12.06261174	2.228651324	-5.4125	0.0000		
wt^-0.5	36.8451208	5.630928282	6.5433	0.0000	2.2015	0.4542
hp^-0.5	121.3603568	24.87296902	4.8792	0.0000	2.2015	0.4542

The Model Summary table shows the R-Square = 89.63% (versus the original untransformed 82.68%). R-Square Predicted is 86.91% (compared to the original untransformed 78.11%). The model fit has dramatically improved.

19. Clicking on the **Residuals mpg** sheet, the Residuals vs Predicted Values plot and Residuals vs wt^-0.5 and hp^-0.5 show no obvious patterns, indicating that the BT transformed model has successfully removed the nonlinearity in the model.







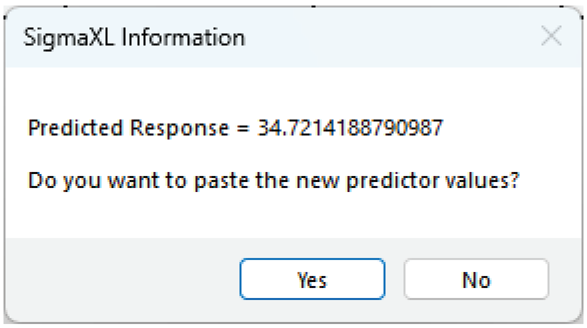
20. Now we will consider optimization with the transformed model. Click the **Model** Sheet tab. Scroll to the Predicted Response Calculator. Select Goal as **Maximize**.

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
wt^-0.5	0.577865603	20.090625					
hp^-0.5	0.089499645						

Goal:	Maximize
Target:	

Optimize   Contour/Surface Plots

21. Click **Optimize**.



Click **Yes**.

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
wt^-0.5	0.812981	34.72141888	0.991015788	32.69456401	36.74827374	30.14329218	39.29954558
hp^-0.5	0.138675						

22. In order to determine the original wt and hp, inverse transforms must be applied. To retrieve the original value of X from $X_{\text{transformed}}$, we use the following inverse transformation formula for each predictor:

If $BT\ Power$ is not zero, the inverse formula is:

$$X = (X_{\text{transformed}})^{1/(\text{BT Power})}$$

If $BT\ Power$ is zero, implying a logarithmic transformation was applied ($\ln(X)$), then the inverse transformation is:

$$X = \exp(X_{\text{transformed}})$$

23. Click the mouse cursor in an empty cell and use the following formulas (note that the X values at cells **K14** and **K15** reference Excel Range Names for use in the prediction equation at cell **L14**):

=_wt__0_5_^(1/-0.5) gives wt = 1.513

=_hp__0_5_^(1/-0.5) gives hp = 52.

Part V – General Linear Model

Summary of Features in General Linear Model

Statistical Tools > General Linear Model > Fit General Linear Model

Extends Advanced Multiple Regression to include:

- Fixed and Random Factors
- Nested Factors
- Covariates (can be Nested)
- For Random or Mixed Random/Fixed Factors with a balanced design, the ANOVA and Variance Components (VC) report is given based on Expected Mean Squares. VC confidence intervals using Restricted Maximum Likelihood (REML) are included.
- If the design is unbalanced or model is non-hierarchical, REML is used to compute the VC values and confidence intervals. Fixed Effects Tests are based on Satterthwaite approximation degrees of freedom.
- Main Effects with Confidence Intervals and Interaction Plots of Fitted Means for Non-Nested Fixed Factors
- Tukey and Fisher Pairwise Comparison of Means for Non-Nested Fixed Factors
- Predicted Response Calculator

Statistical Tools > General Linear Model > GLM Multiple Response Optimization

- Multiple Response Optimization for Nested or Non-Nested Fixed Factors

For details on the statistical methods, formulas and references, see the Appendix: **General Linear Model**. Multiple Response Optimization is introduced in Design of Experiments: **Part F – Multiple Response Optimization with Advanced Multiple Regression**. Note that if a factor is nested in one model and used in another, it must be nested the same way for all models.

Case study examples using GLM will be used to demonstrate the analysis of:

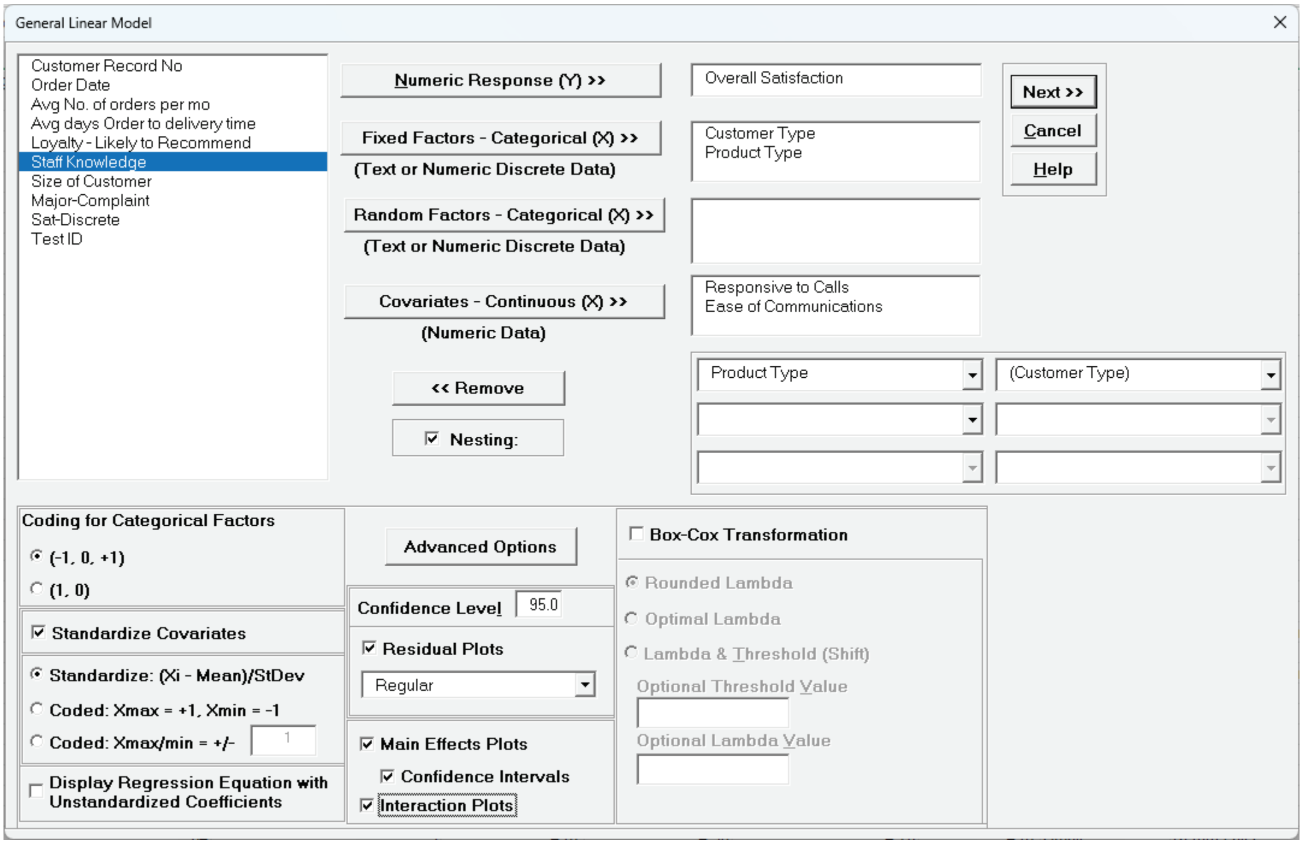
- Fixed Factors with Nested Variable
- Sources of Variation Study
- Classical Gage R&R
- Gage R&R Study with Operator as Fixed Factor
- Destructive (Nested) Gage R&R
- Expanded Gage R&R
- Unbalanced Nested Factorial Experiment with Fixed and Random Factors

For further reading, the following book is recommended:

Montgomery, D.C. (2020). *Design and Analysis of Experiments*, 10th Edition, John Wiley & Sons. See chapter 13 “Experiments with Random Factors” and chapter 14 “Nested and Split-Plot Designs”.

General Linear Model Dialogs and Options

Fit General Linear Model Dialog



- **Numeric Response (Y)** – select the response variable. Only one response may be selected at a time, but you can use **Recall SigmaXL Dialog** or Press **F3** to repeat an analysis with different options or to select a different response. The regression reports will be created on sheets GLM1 – Model *Y1name*, GLM2 - Model *Y2name*, etc., but truncated to fit the 31-character limit for Excel sheet names.

- **Fixed Factors - Categorical (X)** – select fixed categorical predictors. A fixed factor includes data for all levels of interest. Note, in previous analysis using categorical predictors for One & Two-Way ANOVA and Multiple Regression, the factors were assumed to be fixed.

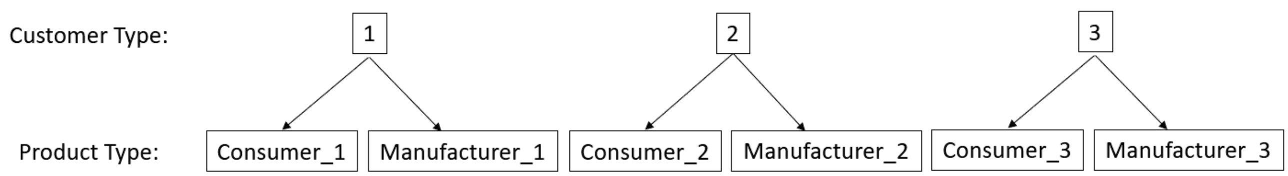
- **Random Factors - Categorical (X)** – select random categorical predictors. A random factor has levels that are randomly sampled from a larger number of possible levels but interest is in all possible levels. If a Random Factor is selected, **Advanced Options** and **Confidence Intervals** for **Main Effects Plots** are greyed out. A variance components report will be produced in sheet GLM# – VarComp Y*name*. Note that the GLM regression report treats random factors as fixed. ANOVA for Predictors, Pareto Charts and CI/PI for Predicted Response Calculator are not available in the regression report. See the variance components report for analysis of random or random/fixed factors. Confidence Intervals for Main Effects Plots, Stepwise/Best

Subsets Regression, K-Fold Cross Validation, and Pairwise Comparison of Means for Fixed Factors are not available.

- For Fixed and Random Factors, numeric variables can be used but they will be converted to text and an underscore "_" will be appended to the number. If there are more than 50 unique levels, a warning message is given. Typically, this occurs when the user has incorrectly selected a continuous variable as categorical. Note that the character "*" cannot be in the name as this is used to denote a cross product term and will be error trapped. Parenthesis characters "(" and ")" will be converted to underscore "_" since these are used to denote a nested term.

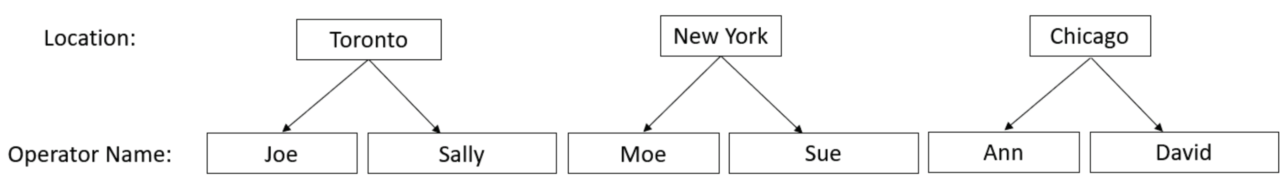
- **Covariates - Continuous (X)** – select continuous numeric variable of interest. Selections with data as text are error trapped. Note that the character "*" cannot be in the Covariate name as this is used to denote a cross product term and will be error trapped. Parenthesis characters "(" and ")" will be converted to underscore "_" since these are used to denote a nested term.

- **Nesting** – check to include nested terms in the model. The left-side combo dropdown selection can be a Factor or Covariate. The right-side dropdown selection is a Factor. This will create a term with the notation A(B) where A is nested in B. For the example above, that would produce Product Type(Customer Type). The graphical representation of this is:



The product type "Consumer_1" is unique to Customer Type 1, etc.

Another example of a nested term would be Operator(Location) where the operators in each plant location are different:



A maximum of three levels of nesting are permitted with up to four factors. In the model, this would appear as "D, C(D), B(C(D)), A(B(C(D)))" or "A, B(A), C(B(A)), D(C(B(A)))" depending on the nesting assignment. One should exercise caution when using three levels of nesting as the number of coefficients in the model is equal to the number of factor combinations = $levels_A * levels_B * levels_C * levels_D$.

A factor cannot be nested within itself, either directly or indirectly, so model "A(B), B(A)" or "A(B), B(C), C(A)" would be illegal.

Note, “B(A), C(A)” is legal, but “A(B), A(C)” is illegal. A workaround to this limitation would be to create a new factor “B_C” by concatenating the levels of B and C and using “A(B_C)”.

If B is nested in A, or A nested in B, i.e., B(A) or A(B), the interaction term A*B is not permitted. SigmaXL automatically excludes these interactions in the Specify Model Terms dialog. Interactions involving A or B with other factors are permitted.

Term hierarchy in a model is recommended, but not required.

For further information on nesting see the Appendix: **Nested Factors and Coding** and **Nested Covariates**.

- **Coding for Categorical Factors (-1, 0, +1)**, also known as effects coding, estimates the difference between each factor level mean and the overall mean. It results in lower multicollinearity VIF scores than **(1, 0)** coding. The reference level is the last alpha-numerically sorted level and is hidden in the Parameter Estimates table.

- **Coding for Categorical Factors (1, 0)**, also known as dummy coding, is the coding scheme typically used for categorical predictors in a regression analysis. The hidden reference value is the first alpha-numerically sorted level.

- **Standardize Continuous Predictors** with **Standardize: (Xi – Mean)/Stdev** will convert continuous predictors to Z-scores. This has two benefits: the predictors are scaled to the same units so coefficients can be meaningfully compared, and it dramatically reduces the multicollinearity VIF scores when interactions and/or quadratic terms are specified.

- **Standardize Continuous Predictors** with **Coded: Xmax = +1, Xmin = -1** scales the continuous predictors so that Xmax is set to +1 and Xmin is set to -1. This is particularly useful for analyzing data from a full or fractional-factorial design of experiments.

- **Standardize Continuous Predictors** with **Coded: Xmax/Xmin = +/-** *value* scales the continuous predictors so that Xmax is set to +*value* and Xmin is set to -*value*. This is particularly useful if one is analyzing data from a response surface design of experiments, where *value* is set to the alpha axial value such as 1.414 for a two-factor rotatable design.

- **Display Regression Equation with Unstandardized Coefficients** displays the prediction equation with unstandardized/uncoded coefficients but the Parameter Estimates table will still show the standardized coefficients. This format is easier to interpret since there is only one coefficient value for each predictor.

- **Confidence Level** is used to determine what alpha value is used to highlight P-Values in red, the significance reference line in the Pareto Chart of Standardized Effects, and the percent confidence and prediction interval used in the Predicted Response Calculator.

- **Residual Plots** *Regular* display the raw residuals (Y - Ŷ) with a Histogram, Normal Probability Plot, Residuals vs Data Order, Residuals vs Predicted Values, Residuals vs Continuous Predictors and Residuals vs Categorical Predictors.

- **Residual Plots** *Standardized* display the residuals, divided by an estimate of its standard deviation. This makes it easier to detect outliers. Standardized residuals greater than 3 and less than -3 are considered large (these outliers are highlighted in red).

- **Residual Plots** *Studentized (Deleted t)* display studentized deleted residuals which are computed in the same way that standardized residuals are, except that the i^{th} observation is removed before performing the regression fit. This prevents the i^{th} observation from influencing the regression model, resulting in unusual observations being more likely to stand out.

- The Residuals report is provided on a separate sheet and includes a table with all residual types to the left of the plots. Other diagnostic measures included, but not plotted are Cook's Distance (Influence), Leverage and DFITS. Leverage is a measure of how far an individual X predictor value deviates from its mean. High leverage points can *potentially* have a strong effect on the estimate of regression coefficients. Leverage values fall between 0 and 1. Cook's distance and DFITS are overall measures of influence. An observation is said to be influential if removing the observation substantially changes the estimate of model coefficients. Cook's distance can be thought of as the product of leverage and the standardized residual squared; DFITS as the product of leverage and the studentized residual. These diagnostic measures can be manually plotted using a Run Chart to identify unusually large values. Commonly used rough cutoff criterion for Cook's distance are: > 0.5, potentially influential and > 1, likely influential. A more accurate cutoff is the median of the F distribution: $> \mathrm{F}(0.5, p, n-p)$, where n is the sample size and p is the number of terms in the model design matrix, including the constant. A commonly used cutoff criterion for the absolute value of DFITS is: $> 2\sqrt{p/n}$. An observation that is an outlier and influential should be examined for measurement error or possible assignable cause. You could also try refitting the model excluding that observation to assess the influence.

- The Residuals report also includes a table to the right of the plots with the stored model design matrix and residuals. This can be used to manually create additional residual plots such as residuals versus interaction or quadratic terms.

- **Tip:** For large datsets (> 1K) you may want to uncheck the **Residual Plots** in order to speed up the analysis.

- **Main Effects Plots** with **Confidence Intervals** and **Interaction Plots** use fitted means, not data means. If an interaction term is not in the model, the interaction plot is still displayed, but it is shaded grey. If the model is not hierarchical, these plots are not displayed.

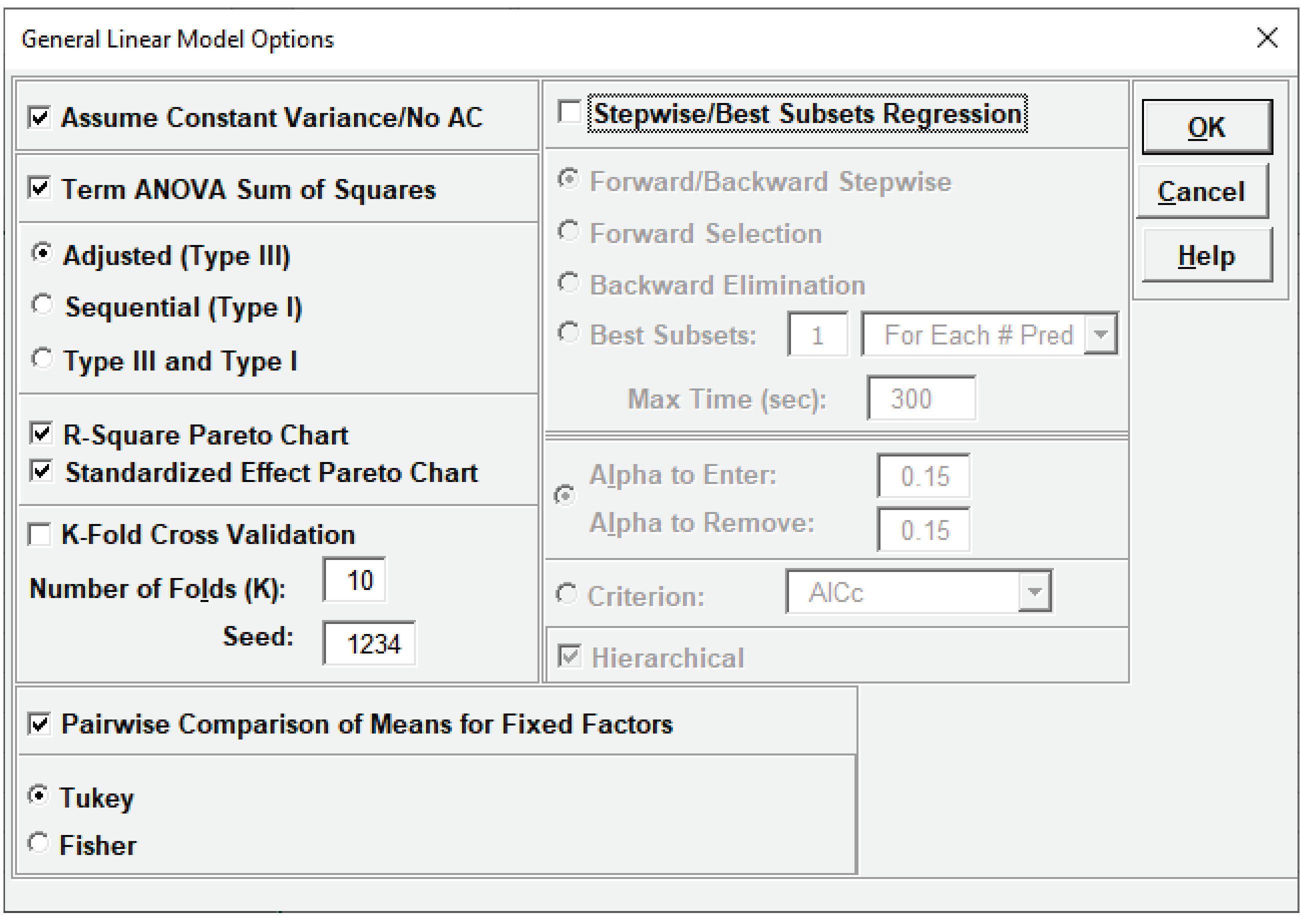
- **Box-Cox Transformation** with **Rounded Lambda** will solve for an optimal lambda and is rounded to the nearest value of: -5, -4, -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 5. A 0 denotes a Ln(Y) transformation, 0.5 is the SQRT(Y), and 1 is untransformed. Threshold (Shift) is computed automatically if the response data includes 0 or negative values, otherwise it is 0. Note that the threshold is subtracted from the data so the value will be negative in order to provide positive response values. Solving lambda is also supported in Stepwise Regression. The reported Parameter Estimates, Model Summary, Information Criteria, Validation, Test

Statistics and Residuals are for the Box-Cox transformed response. The Predicted Response Calculator automatically applies an inverse transformation so that the predicted response, confidence and prediction intervals are given in the original untransformed units. Note, Lambda is solved to normalize the regression residuals, not the raw data. It is solved using the classical Box-Cox formula but the actual transformation uses a simple power transformation.

- **Box-Cox Transformation** with **Optimal Lambda** uses the range of -5 to +5 for Lambda. Threshold is computed automatically if the response data includes 0 or negative values.

- **Box-Cox Transformation** with **Lambda & Threshold (Shift)** allows the user to specify Lambda and Threshold. Threshold is typically 0, but if the response data includes 0 or negative values, a negative threshold value should be entered, such that when subtracted from the data, results in positive response values.

General Linear Model Options Dialog (not available if Random Factor selected)



- **Assume Constant Variance/No AC** (no autocorrelation in the residuals), if unchecked, SigmaXL will use either White robust standard errors for non-constant variance or Newey-West robust standard errors for non-constant variance with autocorrelation. If either of the Durbin-Watson P-Values are < .05 (i.e., significant positive or negative autocorrelation), Newey-West for Lag 1 is used, otherwise White HC3 is used. This will affect SE Coefficients, P-

Values, ANOVA F and P-Values, and Prediction CI/PI. ANOVA F and P-Values are Wald estimates. ANOVA SS Type I Table and Pareto Charts are not available. Note: Stepwise P-Values are not adjusted.

- **Term ANOVA Sum of Squares** with **Adjusted (Type III)** provides a detailed ANOVA table for continuous and categorical predictors. Adjusted Type III is the reduction in the error sum of squares (SS) when the term is added to a model that contains all the remaining terms. Note, categorical terms are considered as a group, unlike the parameter estimates table which uses coding.

- **Term ANOVA Sum of Squares** with **Sequential (Type I)** provides a detailed ANOVA table for continuous and categorical predictors. Sequential Type I is the reduction in the error sum of squares (SS) when a term is added to a model that contains only the terms before it. This is affected by the order that they are entered in the model, so the user must be careful to specify model terms in the order of importance based on process knowledge. Note, if the terms are orthogonal then Type III and Type I SS will be the same.

- **R-Square Pareto Chart** displays a Pareto chart of term R-Square values ($100*SS_{term}/SS_{total}$). A separate Pareto Chart is produced for Type III and Type I SS. If there is only one predictor term, a Pareto Chart is not displayed.

- **Standardized Effect Pareto Chart** displays a Pareto chart of term T values (=T.INV(1-P/2,df_{error})). A separate Pareto Chart is produced for Type III and Type I SS. A significance reference line is included (=T.INV(1-alpha/2,df_{error})).

- **K-Fold Cross Validation:** In K-Fold cross-validation, the data is randomly partitioned into K (approximately equal) subsets. The model coefficients are estimated using K-1 partitions, i.e., (100*(K-1)/K)% of the data – the training set, and then statistical metrics are evaluated on the remaining data – the validation set. This is repeated for each of the K-Fold validation sets with R-Square K-Fold and S (Standard Deviation) K-Fold calculated as an average across the K samples, which results in a more accurate estimate of model prediction performance. The default K=10 is a popular choice, but some practitioners prefer K=5. Note that the final model parameter coefficients are based on all of the data, so K-Fold Cross Validation is used strictly to obtain R-Square K-Fold and S K-Fold. The fixed seed allows for replicable results, but the user may wish to re-run the analysis with a different seed a few times to see how much variation occurs in R-Square K-Fold and S K-Fold. If categorical predictors are used and the training sample does not include all of the levels, the K-Fold statistics cannot be computed.

- **Pairwise Comparison of Means for Fixed Factors: Tukey** or **Fisher.** Pairwise comparisons of means examines the difference between all combinations of the estimated means for each category of a factor, along with the standard error and confidence band for the difference. Tukey provides protection against false positives due to multiple comparisons so is the default selection. If the model is not hierarchical, the pairwise comparison report is not available. For further information see the Appendix: **Pairwise Comparison of Means for Fixed Factors**.

- **Stepwise/Best Subsets Regression** with **Forward/Backward Stepwise:** Starting with an empty model, terms are added or removed from the model, one at a time, until all variables in the model have p-values that are less than (or equal to) the specified **Alpha-to-Remove** and all variables that are not in the model have p-values greater than the specified in **Alpha-to-Enter**. The stepwise process either adds the term which is most significant (largest F statistic, smallest p-value), or removes the term that is least significant (smallest F statistic, largest p-value). It does not consider all possible regression models. The independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Forward Selection:** Starting with an empty model, the most significant terms are added to the model, one at a time, until all variables that are not in the model have p-values greater than the specified in **Alpha-to-Enter**. Terms that are in the model are not removed regardless of p-value. Alternatively, criterion-based selection may be used. The most significant terms are added, one at a time, while at each stage the value of a measure, such as AICc or R-Square is monitored. If a minimum AICc is observed at step i, and this remains the minimum after 10 additional steps (or the model includes all terms), then the model at the minimum AICc is selected. If a maximum R-Square is observed at step i, and this remains the maximum after 10 additional steps, then the model with the maximum R-Square is selected. **Criterion** options are: *AICc*, *BIC*, *R-Square Adjusted*, *R-Square Predicted* and *R-Square K-Fold*. AICc is the Akaike Information Criterion corrected for small sample sizes, BIC is the Bayesian Information Criterion. For details on these metrics, see the Appendix: **Advanced Multiple Regression**. Note that for *R-Square K-Fold*, the F-statistic to decide which term to enter is based on all of the data. The K-Fold model is computed using the specified model, but a subset of the data is used as training data to estimate parameters and R-square is calculated using the out-of-sample validation data. As with forward/backward stepwise, the independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Backward Elimination:** Starting with all terms in the model, the least significant terms are removed from the model, one at a time, until all variables in the model have p-values that are less than (or equal to) the specified **Alpha-to-Remove**. Terms that are removed from the model are not entered again regardless of p-value. Alternatively, criterion-based selection may be used, as described above, but the least significant terms are removed, one at a time. It stops after 10 additional steps or if the model includes only one term. As with forward/backward stepwise, the independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** with **Best Subsets:** With Best Subsets, for any given model with p terms, there are $2^p - 1$ possible combinations (non-hierarchical models). A criterion such as AICc is specified, and the model which results in the minimum AICc is selected. If $p \leq 15$, all possible combinations are explored - this is called exhaustive. Otherwise, the best model is derived using discrete optimization with the powerful MIDACO Solver (Mixed Integer Distributed Ant Colony Optimization). Start values are obtained using forward selection with the AICc criterion. MIDACO does not guarantee a best solution as we have in exhaustive, but will be close to best, even for hundreds of terms! Best Subsets **Criterion** options are: *AICc*, *BIC* and *R-Square Adjusted*. *R-Square Predicted* and *R-Square K-Fold* are not feasible as criterion here due to the computation times, but they are reported on the best selected models. Best Subsets report options are: Best *For Each # of Pred* (default) or Best *Overall*. Best For Each # of Predictors provides the most information but takes longer to compute than Best Overall. The user may specify how many models to include (per # predictors or overall) in the report, with the default = 1. The default **Max Time (sec)** = 300 is the maximum total computation time allotted for either option. The independent variables can be continuous and/or categorical. A categorical predictor is treated as a group, so is either all in or all out.

- **Stepwise/Best Subsets Regression** – **Hierarchical:** The Hierarchical option constrains the model so that all lower order terms that comprise the higher order terms are included in the model. This is checked by default. In Forward/Backward Stepwise and Forward Selection, a hierarchical model is required at each step, but extra terms can enter to maintain hierarchy. For Backward Elimination and Best Subsets, extra terms are not permitted.

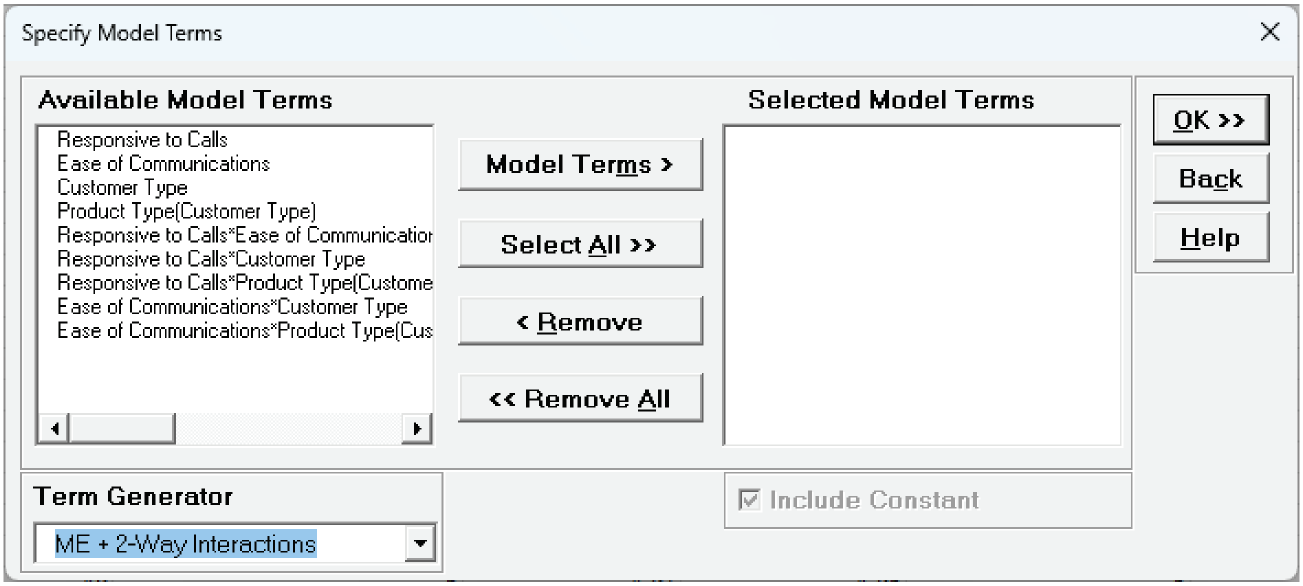
- **Saturated Model Pseudo Standard Error (Lenth's PSE):** For saturated models with $df_{error} = 0$, Lenth's method is used compute a pseudo standard error. For each term, a t ratio is computed by dividing the coefficient by the PSE. Since the distribution of the t ratio is not analytic, the probability is evaluated using Monte Carlo simulation. Student T P-Values are also available for comparison purposes. Lenth's PSE in the SigmaXL DOE Templates and DOE Analysis use Student T P-Values.

Tip: There are a lot of options here, giving the user flexibility for model refinement, but this can also be overwhelming to someone starting out with these tools. We recommend using the following settings for **Stepwise/Best Subsets Regression**:

1. **Forward Selection**, **Criterion***: AICc, BIC or R-Square Predicted*, **Hierarchical** checked. This is fast and will build a model that minimizes AICc, BIC or maximizes R-Square Predicted. AICc or R-Square Predicted are recommended for the best model prediction accuracy, *BIC* is recommended for model parsimony. Note, however, this does not consider all possible models.

2. **Best Subsets**, *1 For Each # of Pred*, **Max Time (sec)** = 300, **Criterion:** *AICc* or *BIC*, **Hierarchical** checked. This can be slow but gives a very useful report of the best model for each number of predictors in the model.

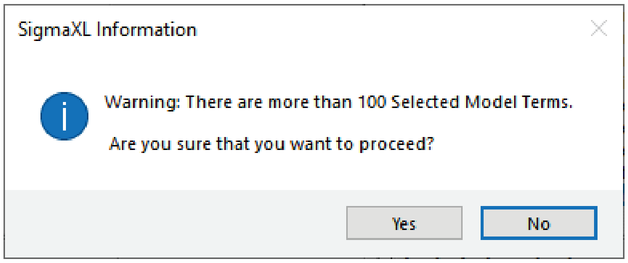
Specify Model Terms Dialog



- **Term Generator** – select any of the following to build a list of **Available Model Terms**:
  - *Main Effects* – default – no change to original specified terms.
  - *ME + 2-Way Interactions* – use this to include 2-way interactions in the model, for example, analyzing data from a Res IV or Res V fractional-factorial DOE. When specifying interactions or higher order terms, standardization of continuous predictors is highly recommended.
  - *ME + 2-Way Interactions + Quadratic* – use this to include 2-way interactions and quadratic terms in the model, for example, analyzing data from a response surface DOE. Categorical terms will not be squared.
  - *ME + All Interactions* – use this to include all possible interaction terms in the model, for example analyzing data from a full-factorial DOE.
  - *All up to 3-Way* – use this to include 2-way interactions, quadratic, 3-way interactions, quadratic*main effect and cubic terms in the model. Categorical terms will not be squared or cubed.

- **Model Terms:** Select from highlighted **Available Model Terms**.

- **Select All:** Select all **Available Model Terms.** Caution, the number of selected terms can become quite large, especially for the last two options in the **Term Generator**. If more than 100 terms are selected, a warning is given after clicking **OK**:



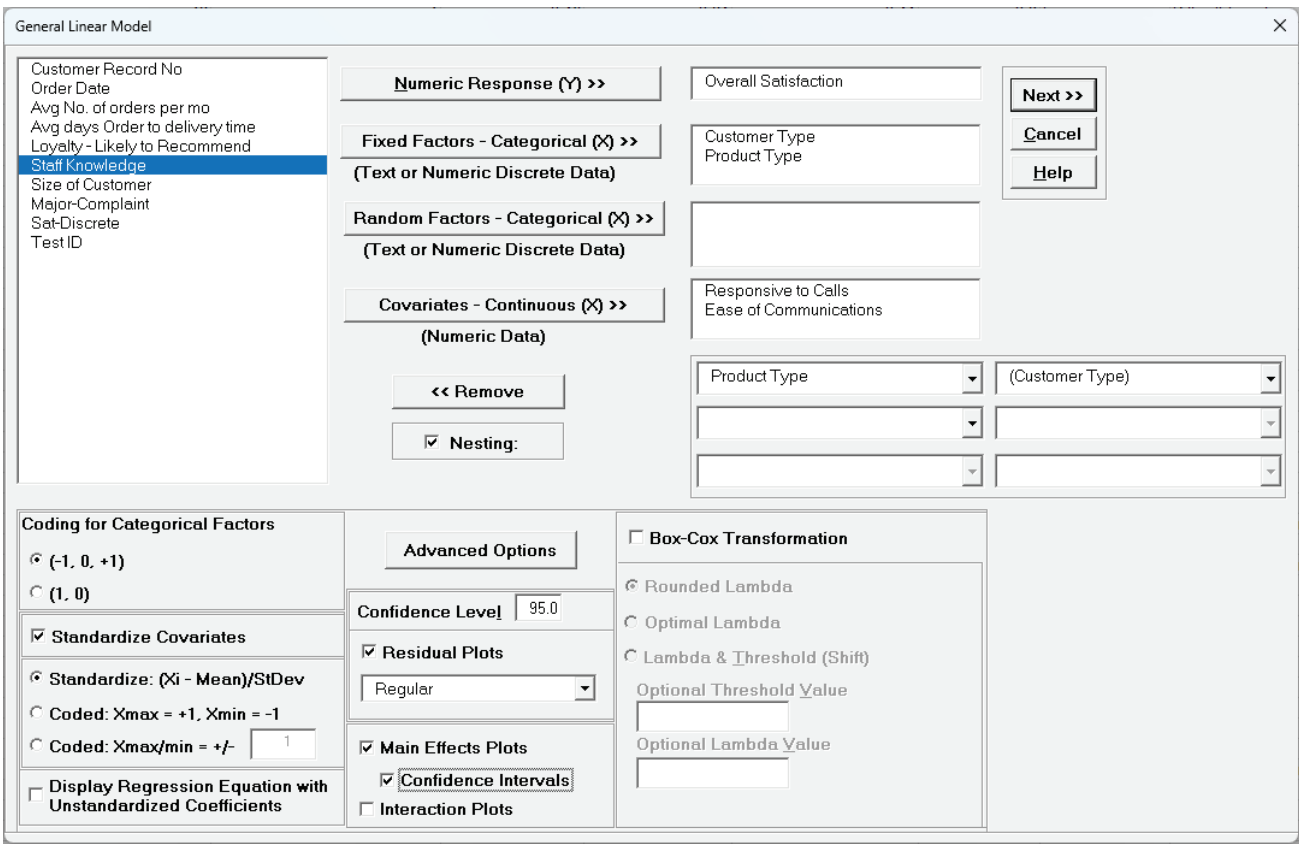
- **Include Constant:** Always checked in GLM**.**

Tip: It is also important to ensure that the number of rows/observations are sufficient to estimate the number of selected model terms. A rule of thumb (excluding data from a designed experiment) is that for every term selected, there should be a minimum of 10 rows of data. This rule holds for potential terms used in stepwise and best subsets as well, otherwise one can easily produce a model that is highly significant but a meaningless model of noise. This is what Jim Frost calls “Data Dredging” in chapter 8 of his book *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*.

Example 1: Fixed Factors with Nested Variable

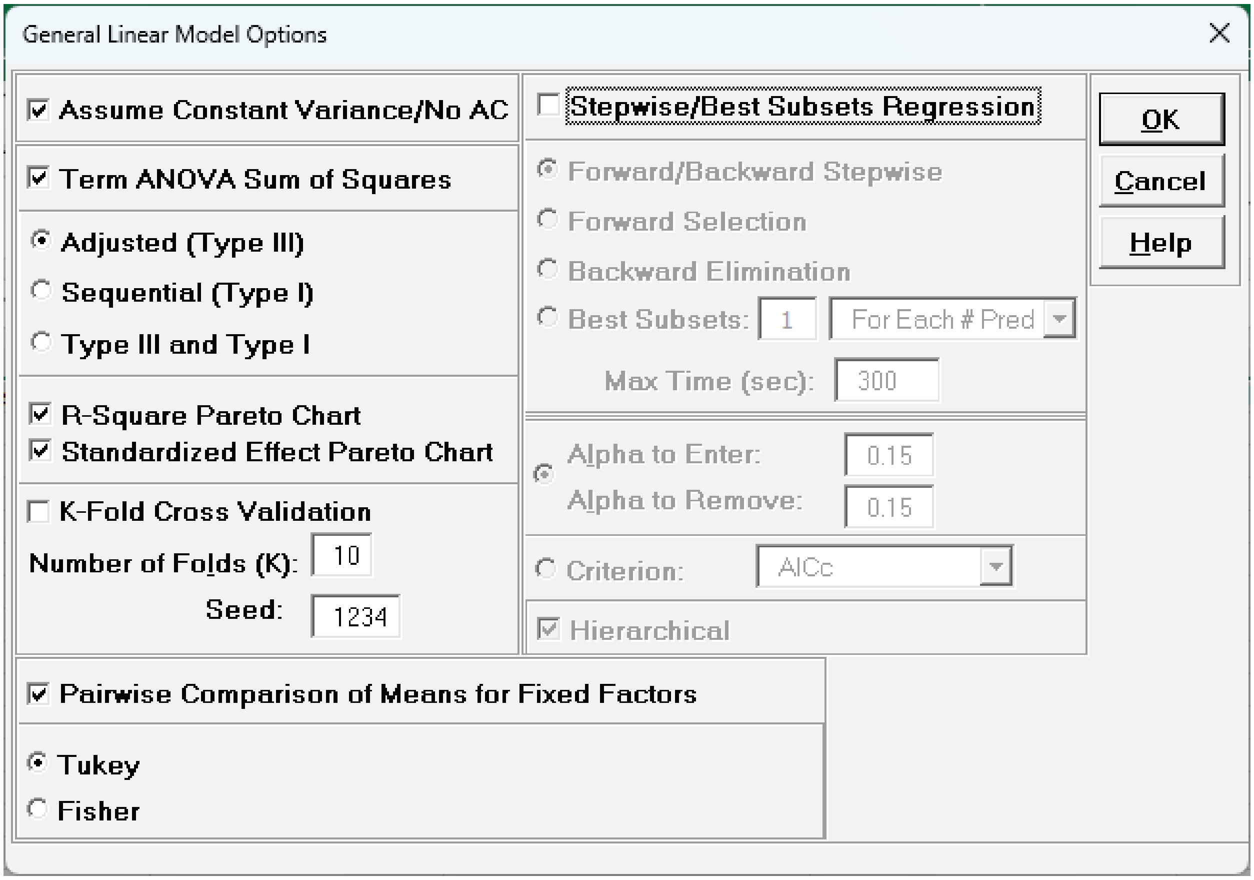
1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Response (Y) >>**, select *Customer Type* and *Product Type*, click **Fixed Factors - Categorical (X) >>**, select *Responsive to Calls* and *Ease of Communications*, click **Covariates - Continuous (X) >>**. Check **Nesting**. For the left-side drop-down "Select a Factor or Covariate", select *Product Type*. For the right-side drop-down "Select a Factor to Nest in:", select *(Customer Type)*. We will use the default **Coding for Categorical Predictors (-1, 0, +1)**. Check **Standardize Covariates** with default option **Standardize: (Xi-Mean)/StDev**. Use the default **Confidence Level** = 95.0%. Regular **Residual Plots** are checked by default. Check **Main Effects Plots** with **Confidence Intervals.** Leave **Interaction Plots** and **Box-Cox Transformation** unchecked.



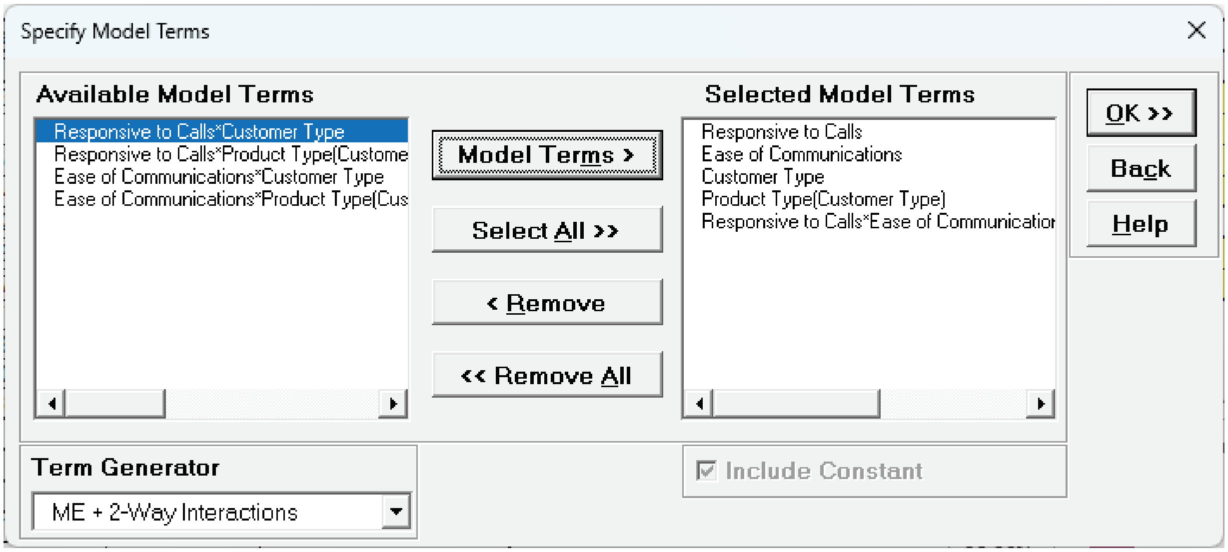
Tip: If you have covariates (continuous predictors) and are planning to include interaction terms in the model, always ensure that **Standardize Covariates** is checked. This has two benefits: the covariates are scaled to the same units so coefficients can be meaningfully compared, and it dramatically reduces the multicollinearity VIF scores.

3. Click **Advanced Options.** We will use the default options as shown with **K-Fold Cross Validation** and **Stepwise/Best Subsets Regression** unchecked. **Pairwise Comparison of Means for Fixed Factors** with **Tukey** option is checked.



4. Click **OK.** Click **Next >>.**

5. Using **Term Generator**, select *ME + 2-Way Interactions*. Select *Responsive to Calls* to *Responsive to Calls*Ease of Communication*. Click **Model Terms >**. **Include Constant** is always checked in General Linear Model.



Product Type(Customer Type) denotes that *Product Type* is nested within *Customer Type*. Note, this data is not actually nested, but we are using this as an example to demonstrate nesting with familiar data.

Term Generator, *ME + 2-Way Interactions* produces only legal two-way interactions, so *Product Type*Customer Type* is not available for selection. Based on previous regression analysis for *Overall Satisfaction,* we are only including the *Responsive to Calls*Ease of Communications* 2-way interaction.

6. Click **OK >>**. The General Linear Model report is given. The Parameter Estimates – Standardized and ANOVA tables are shown:

Parameter Estimates - Standardized						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	3.764860358	0.023698168	158.8671	0.0000		
Responsive to Calls	0.528193474	0.025885408	20.4051	0.0000	1.4144	0.7070
Ease of Communications	0.388197318	0.025827726	15.0303	0.0000	1.4081	0.7102
Customer Type_1_	-0.047490903	0.033504621	-1.4174	0.1598	1.3844	0.7223
Customer Type_2_	0.061041475	0.033164135	1.8406	0.0689	1.5653	0.6388
Product Type(Customer Type)_Consumer_1_	0.025052486	0.039622623	0.6323	0.5288	1.0347	0.9665
Product Type(Customer Type)_Consumer_2_	0.009162637	0.034001909	0.2695	0.7882	1.0264	0.9742
Product Type(Customer Type)_Consumer_3_	-0.141778394	0.04220058	-3.3596	0.0011	1.0218	0.9787
Responsive to Calls*Ease of Communications	0.088559804	0.022327097	3.9665	0.0001	1.2784	0.7822

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	8	56.34432657	7.043040822	150.1678	0.0000
Error	91	4.268004426	0.046901148		
Total (Model + Error)	99	60.612331	0.612245768		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Responsive to Calls	1	19.52807844	19.52807844	416.3668	0.0000	32.22%	20.4051
Ease of Communications	1	10.59537152	10.59537152	225.9086	0.0000	17.48%	15.0303
Customer Type	2	0.174136651	0.087068325	1.8564	0.1621	0.29%	1.4095
Product Type(Customer Type)	3	0.551731846	0.183910615	3.9212	0.0111	0.91%	2.5932
Responsive to Calls*Ease of Communications	1	0.737891042	0.737891042	15.7329	0.0001	1.22%	3.9665

With (-1, 0, 1) coding, the reference level is the last alpha-numerically sorted level. The hidden reference level for Customer Type is 3.

"Product Type(Customer Type)_Consumer_1_" denotes Product Type Consumer nested within Customer Type 1. The hidden reference level for Product Type is Manufacturer. For further information on nesting see the Appendix: **Nested Factors and Coding.**

The **Analysis of Variance for Predictors (Adjusted Type III)** table shows that Customer Type is not significant with P-Value = 0.162, but we will leave it in the model to maintain hierarchy since Product Type is nested within Customer Type. Product Type is significant with P-Value = 0.011.

7. The GLM Equation with standardized coefficients is:

   General Linear Model: Overall Satisfaction = (3.76486)
   + (0.528193)*((Responsive_to_Calls-3.8644)/1.14029)
   + (0.388197)*((Ease_of_Communications-3.7481)/0.911361)
   + (-0.0474909)*(IF(Customer_Type="1_",1,0)+IF(Customer_Type="3_",-1,0))
   + (0.0610415)*(IF(Customer_Type="2_",1,0)+IF(Customer_Type="3_",-1,0))
   + (0.0250525)*(IF(Product_Type="Consumer",1,0) + IF(Product_Type="",-1,0) + IF(Product_Type="Manufacturer",-1,0))*IF(Customer_Type="1_",1,0)
   + (0.00916264)*(IF(Product_Type="Consumer",1,0) + IF(Product_Type="",-1,0) + IF(Product_Type="Manufacturer",-1,0))*IF(Customer_Type="2_",1,0)
   + (-0.141778)*(IF(Product_Type="Consumer",1,0) + IF(Product_Type="",-1,0) + IF(Product_Type="Manufacturer",-1,0))*IF(Customer_Type="3_",1,0)
   + (0.0885598)*((Responsive_to_Calls-3.8644)/1.14029)*((Ease_of_Communications-3.7481)/0.911361)

   Note blanks and special characters in the predictor names are converted to the underscore character "_". The numeric Customer Type 1, 2, 3 has also been converted to text so appear as "1_", "2_", "3_".

   For categorical predictors, IF statements are used.

   This is the display version of the prediction equation given at cell **L14** (which has more precision for the coefficients and predictor names are converted to legal Excel range names by padding with the underscore "_" character). If the equation exceeds 8000 characters (Excel's legal limit for a formula is 8192), then a truncated version is displayed and cell **L14** does not show the formula.

8. Scroll to the Predicted Response Calculator. Enter *Responsive to Calls* and *Ease of Communication* values = 5 with Customer Type = 1_ and Product Type = Consumer from the dropdown lists to predict Overall Satisfaction including the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator							
Predictors	**Enter Settings:**	**Predicted Response**	**SE**	**Lower 95% CI**	**Upper 95% CI**	**Lower 95% PI**	**Upper 95% PI**
Responsive to Calls	5	4.922843097	0.080100122	4.763734045	5.081952149	4.464178431	5.381507763
Ease of Communications	5						
Customer Type	1_						
Product Type	Consumer						

9. Next, we will use SigmaXL's built in Optimizer. Scroll to view the Optimize Options:

Optimize Options			
Continuous Predictors	**Lower Bound**	**Upper Bound**	**Integer**
Responsive to Calls	1	5	0
Ease of Communications	1.4	5	0

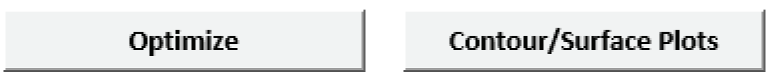
Categorical Predictors	All Levels/Hold Level
Customer Type	All Levels
Product Type	All Levels

Here we can constrain the continuous predictors and specify a level to use for optimization of the categorical predictors. If a continuous predictor is integer, change the **Integer** 0 to 1, and the Optimizer will return only integer values for that predictor.

We will leave the Optimize Option settings as is.

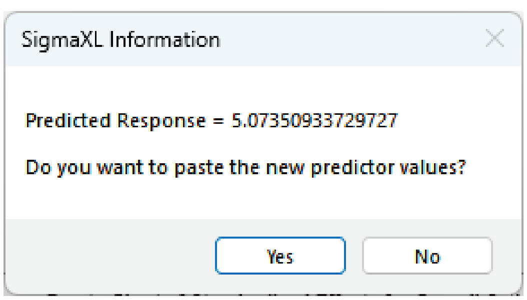
10. Scroll back to view the Goal setting and Optimize button. Select Goal = *Maximize*.

Goal:	Maximize
Target:	



The optimizer uses Multistart Nelder-Mead Simplex to solve for the desired response goal with given constraints. For more information see the Appendix: **Single Response Optimization**.

11. Click **Optimize**. The response solution and prompt to paste values into the Input Settings of the Predicted Response Calculator is given:



12. Click **Yes** to paste the values.

Predicted Response Calculator								
Predictors	**Enter Settings:**	**Predicted Response**	**SE**	**Lower 95% CI**	**Upper 95% CI**	**Lower 95% PI**	**Upper 95% PI**
Responsive to Calls	5	5.073509337	0.083358486	4.907927946	5.239090729	4.612559467	5.534459208
Ease of Communications	5						
Customer Type	3_						
Product Type	Manufacturer						

The optimizer has selected Responsive to Calls = 5, Ease of Communications = 5, Customer Type = 3_ and Product Type = Manufacturer to maximize the Overall Satisfaction predicted value.

Note that the optimizer does not test the validity of a nested combination, so if the goal is to optimize, it is best to use generic level names as done in this example. For Product Type, we use the generic levels "Consumer" and "Manufacturer", not "Consumer_Type1", "Consumer_Type2", "Consumer_Type3"; "Manufacturer_Type1", "Manufacturer_Type2", "Manufacturer_Type3".

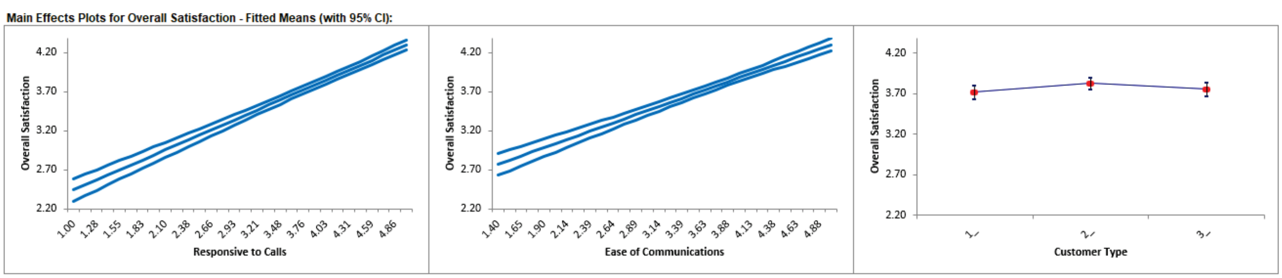
13. Click on Sheet **GLM1 – Pair Comp**. The Tukey Pairwise Comparison of Means for Customer Type is shown.

Tukey Pairwise Comparison of Means for Customer Type						
Diff of Levels	Diff of Means	SE	T	Adj P-Value	Lower 95% CI	Upper 95% CI
2_ - 1_	0.108532378	0.056187303	1.9316	0.1357	-0.025342993	0.242407749
3_ - 1_	0.033940332	0.059648688	0.5690	0.8369	-0.10818235	0.176063013
3_ - 2_	-0.074592046	0.057550679	-1.2961	0.4010	-0.211715881	0.062531789

Pairwise comparisons of means examines the difference between all combinations of the estimated means for each category of a factor, along with the standard error and confidence band for the difference. Tukey provides protection against false positives due to multiple comparisons. Nested factors are not shown in the pairwise comparison report, so Product Type is not included here. For further information see the Appendix: **Pairwise Comparison of Means for Fixed Factors**.

As expected from the ANOVA report, none of the pairwise comparisons are significant.

Click on Sheet **GLM1 – Plots**. The Main Effects Plots with 95% Confidence Intervals are shown.



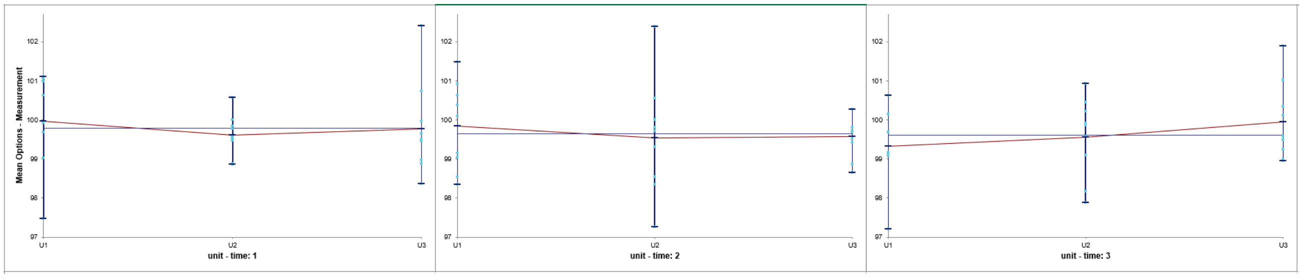
These are based on Fitted Means as predicted by the model, not Data Means. They use the predicted value for the response versus input predictor value, while holding all other variables constant. Continuous are held at their respective means and categorical are weighted equally.

Here we see that *Responsive to Calls* has the steepest slope followed by *Ease of Communications*. As expected, *Customer Type* does not appear to be an important factor.

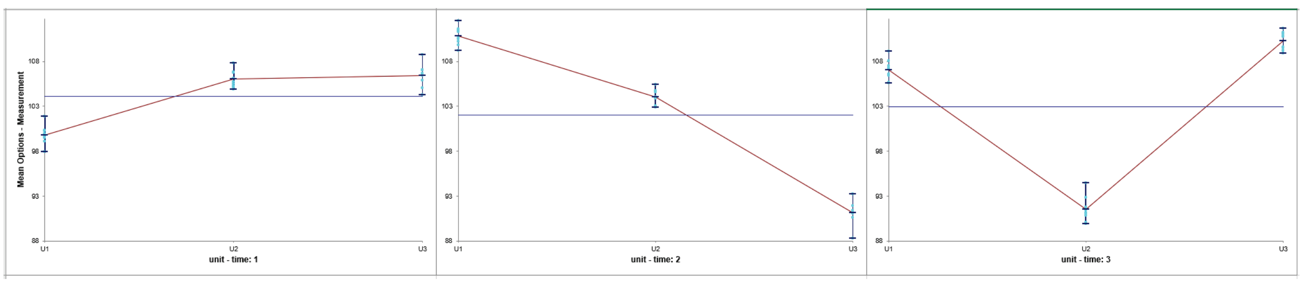
Example 2: Sources of Variation Study

We will now reanalyze the Multi-Vari data presented in **Multi-Vari Charts**. The Multi-Vari chart was used to identify dominant Sources of Variation (SOV), with the three major "families" of variation being Within Unit, Between Unit, and Temporal (Over Time).

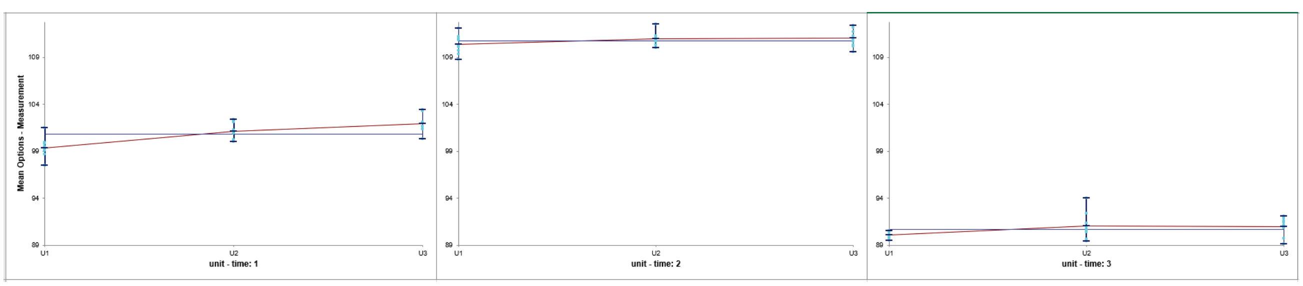
Dominant "Within Unit" Source of Variation:



Dominant "Between Unit" Source of Variation:



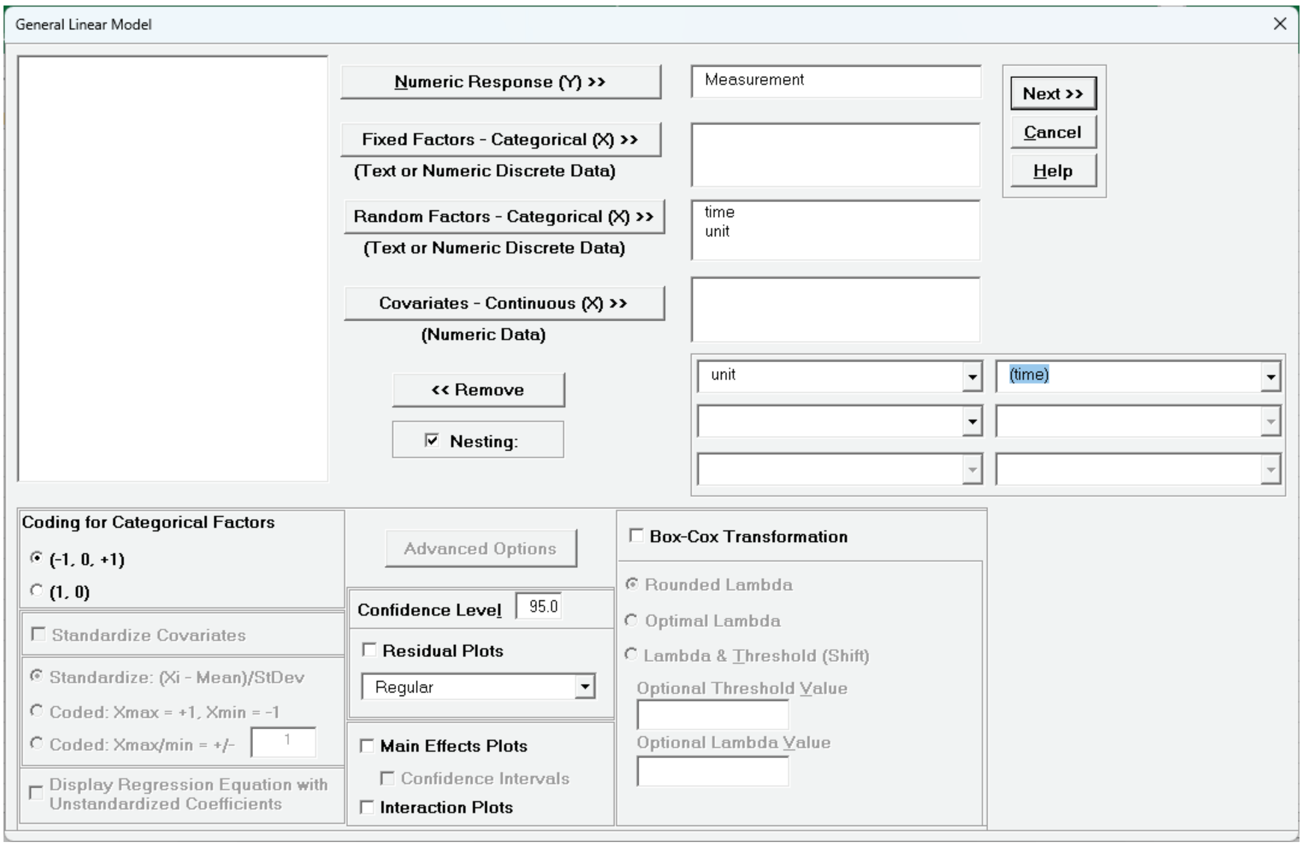
Dominant "Over Time" Source of Variation:



Unit and *Time* are Random Factors and *Unit* is nested within *Time*. The Variance Components will be calculated to quantify the percent contribution to total variation for each factor.

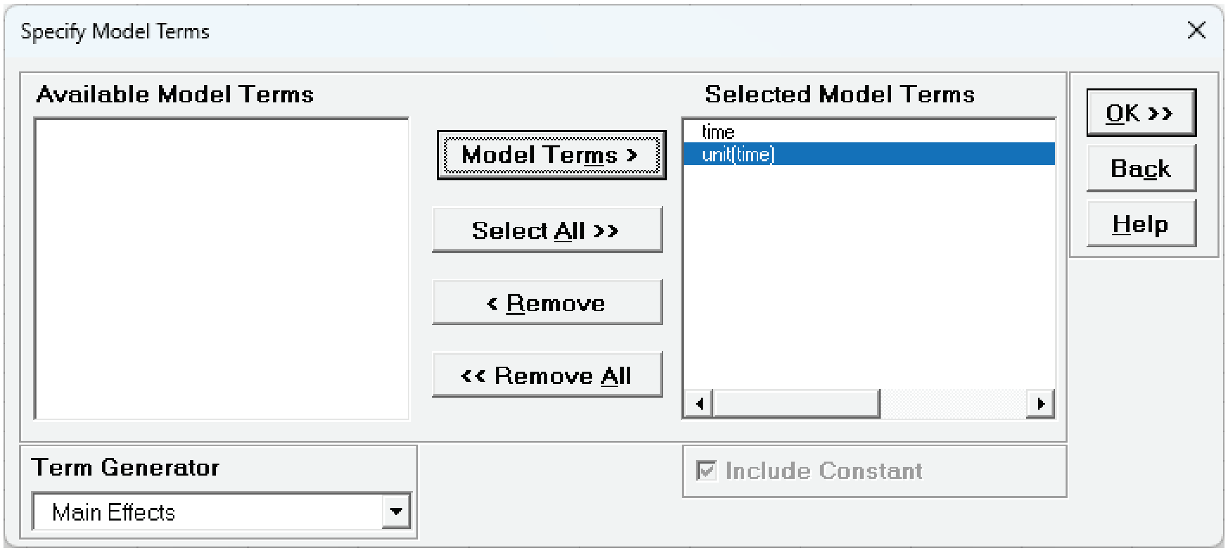
1. Open **Multi-Vari Data.xlsx**, click Sheet **Within**. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Measurement*, click **Numeric Response (Y) >>**, select *time*, click **Random Factors - Categorical (X) >>**, select *unit*, click **Random Factors - Categorical (X) >>.** Check **Nesting**. For the left-side drop-down "Select a Factor or Covariate", select *unit*. For the right-side drop-down "Select a Factor to Nest in:", select *(time)*. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Leave **Term Generator** as *Main Effects*. Click **Select All >>.** (If the order is reversed, select *time*, click **Model Terms >,** select *unit(time),* click **Model Terms >**.)



The term *unit(time)* denotes that *unit* is nested within *time*. Since *time* is the top level of the nesting, we place it in the model first.

5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
time	Random	3
unit(time)	Random	9

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
time	2	0.429364963	0.214682481	6.0000	0.486026173	0.4417	0.6623
unit(time)	6	2.916157037	0.486026173	72.0000	1.078104306	0.4508	0.8421
Error	72	77.62351	1.078104306				
Total	80	80.969032	1.0121129				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
time	-0.010049766	0.00%			
unit(time)	-0.065786459	0.00%			
Error	1.078104306	100.00%	0.158657639	0.744381519	1.376139036
Total	1.078104306				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
time	0	0.00%		
unit(time)	0	0.00%		
Error	1.038318018	100.00%	0.862775474	1.173089526
Total	1.038318018			

The *Error* term in the variance components table is the "Within Unit" variation. It is contributing to 100% of the total variation. The REML 95% confidence interval for this variance component is also given.

The terms *time* and *unit(time)* have negative Variance Component values, so are treated as 0. The ANOVA table shows that both terms are not significant random factors (alpha = 0.05).

Standard Deviation components are provided for convenience and useful in Measurement Systems Analysis, but they will not be discussed here.

The **GLM Model** sheet is the regression analysis but we will not review that here as our focus is on the variance components analysis. Note that the regression analysis treats random factors as fixed.

6. Now click sheet **Between**. Repeat steps 1 to 5 to produce the Variance Components report:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
time	Random	3
unit(time)	Random	9

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
time	2	58.53733519	29.26866759	6.0000	643.4351105	0.0455	0.9559
unit(time)	6	3860.610663	643.4351105	72.0000	1.260599358	510.4200	0.0000
Error	72	90.76315378	1.260599358				
Total	80	4009.911152	50.1238894				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
time	-22.74690529	0.00%			
unit(time)	71.35272347	98.26%	27.21631493	20.32550204	145.0237399
Error	1.260599358	1.74%	0.210099869	0.909306176	1.747607792
Total	72.61332282				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
time	0	0.00%		
unit(time)	8.447054129	99.13%	4.50838131	12.04258028
Error	1.12276416	13.18%	0.95357547	1.321971177
Total	8.52134513			

The *unit(time)* term in the variance components table is the “Between Unit” variation. It is contributing to 98.3% of the total variation. The “Within Unit” *Error* term is contributing to 1.7% of the total variation. The REML 95% confidence intervals for these variance components are also given.

The term *time* has a negative Variance Component value, so is treated as 0. The ANOVA table shows that *time* is not a significant random factor, but *unit(time)* is significant (alpha = 0.05).

7. Now click sheet **Over Time**. Repeat steps 1 to 5 to produce the Variance Components report:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
time	Random	3
unit(time)	Random	9

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
time	2	5436.654367	2718.327184	6.0000	6.359464037	427.4460	0.0000
unit(time)	6	38.15678422	6.359464037	72.0000	0.935748056	6.7961	0.0000
Error	72	67.37386	0.935748056				
Total	80	5542.185012	69.27731264				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
time	100.4432489	98.49%	100.6788273	14.08388582	716.3393283
unit(time)	0.602635109	0.59%	0.408327667	0.159700223	2.274067423
Error	0.935748056	0.92%	0.155957999	0.674981702	1.297256447
Total	101.981632				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
time	10.02213794	99.24%	3.75285036	26.76451622
unit(time)	0.776295761	7.69%	0.399625103	1.508001135
Error	0.967340713	9.58%	0.8215727	1.138971662
Total	10.09859555			

The *time* term in the variance components table is the "Over Time" variation. It is contributing to 98.5% of the total variation. The "Between Unit" *unit(time)* term is contributing to 0.6% of the total variation. The "Within Unit" *Error* term is contributing to 0.9% of the total variation. The REML 95% confidence intervals for these variance components are also given. Note that the confidence interval for *time* is very wide. This is due to only having 3 levels. If possible, additional time value data should be collected in order to reduce the uncertainty of the variance component value.

The ANOVA table shows that *time* and *unit(time)* are significant random factors (alpha = 0.05).

Example 3: Classical Gage R&R Study

We will now reanalyze the AIAG Gage R&R data presented in **Analyze Gage R&R (Crossed)**. The Gage R&R Study Report was given as:

Gage R&R Study (Crossed) Report

Gage Name:	AIAG Example, MSA Reference Manual, 3rd Edition, Page 101
Date of Study:	
Performed By:	
Notes:	Parts were measured in random order, but worksheet is given in standard order.

Process Tolerance (USL - LSL):	8
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6
Alpha to Remove Interaction:	0.1
Confidence Level:	90.0
Number of Parts:	10
Number of Operators:	3
Number of Replicates:	3
Design Type:	Balanced

Analysis of Variance with Part * Operator Interaction:

Source	DF	SS	MS	F	P
Part:	9	88.362	9.8180	492.29	0.0000
Operator:	2	3.1673	1.5836	79.406	0.0000
Part * Operator:	18	0.358982	0.019943457	0.433721	0.9741
Repeatability:	60	2.7589	0.045982222		
Total:	89	94.647	1.06		

Analysis of Variance without Part * Operator Interaction (P for Interaction >= 0.1):

Source	DF	SS	MS	F	P
Part:	9	88.362	9.8180	245.61	0.0000
Operator:	2	3.1673	1.5836	39.617	0.0000
Repeatability:	78	3.1179	0.039973276		
Total:	89	94.64711222	1.063450699		

Gage R&R Metrics	StDev	StDev Lower 90% CI	StDev Upper 90% CI	6 * StDev	% Total Variation (TV)	% TV Lower 90% CI	% TV Upper 90% CI
Gage R&R:	0.302372	0.235108	1.03	1.8142	27.86	15.46	70.93
Operator (AV Appraiser Variation):	0.226838	0.127545	1.01	1.3610	20.90		
Part * Operator (INT Interaction):	0	0	0	0	0.00		
Reproducibility (SQRT(AV^2 + INT^2)):	0.226838	0.127545	1.01	1.3610	20.90		
Repeatability (EV Equipment Variation):	0.199933	0.176915	0.230560	1.1996	18.42		
Part Variation (PV):	1.04	0.758821	1.7170	6.2540	96.04		
Total Variation (TV):	1.09	0.816099	1.8111	6.5118	100.00		

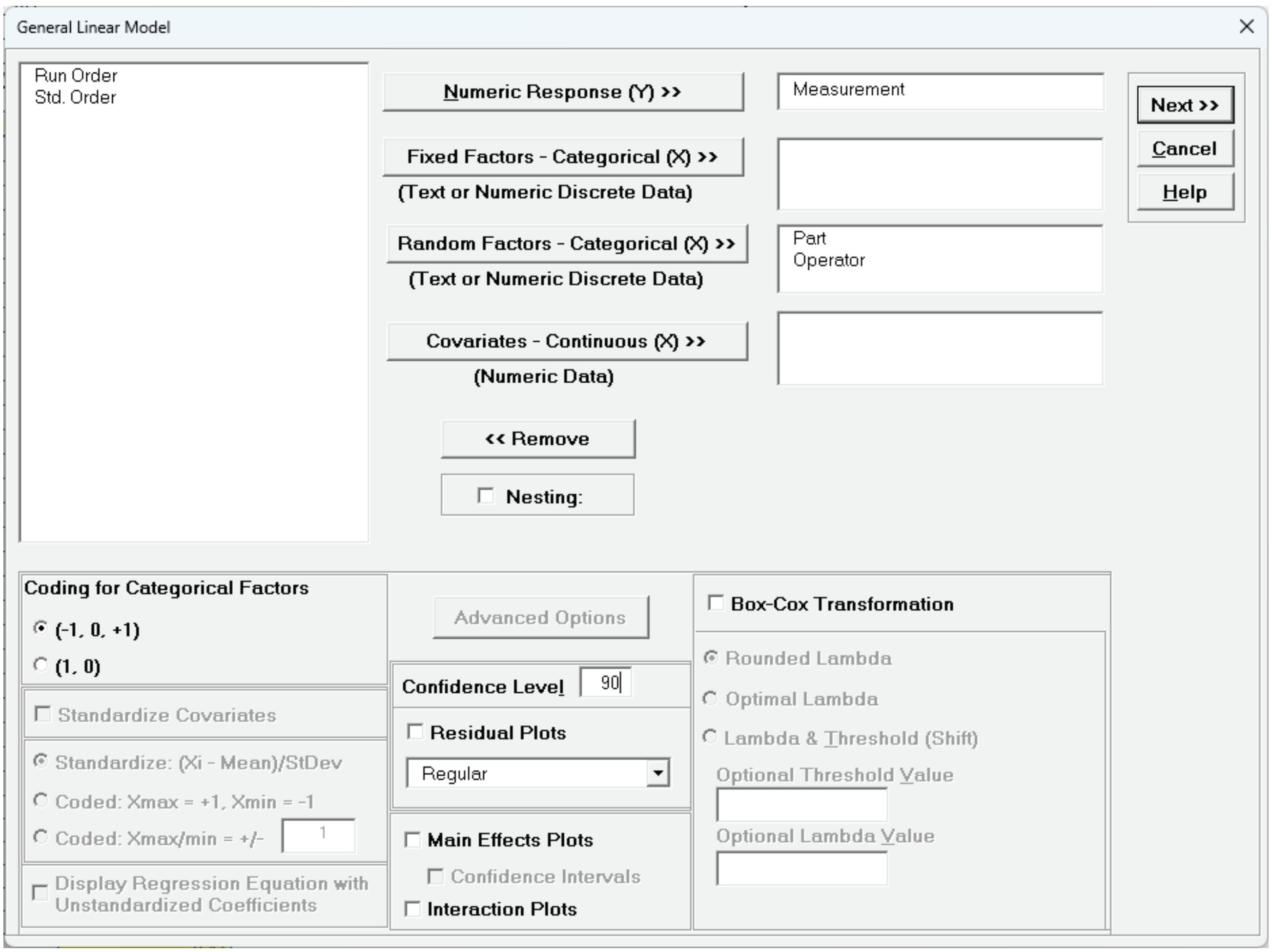
Gage R&R Metrics	% Tolerance	% Tolerance Lower 90% CI	% Tolerance Upper 90% CI
Gage R&R:	22.68	17.63	77.50
Operator (AV Appraiser Variation):	17.01	9.57	76.03
Part * Operator (INT Interaction):	0.00	0.00	0.00
Reproducibility (SQRT(AV^2 + INT^2)):	17.01	9.57	76.03
Repeatability (EV Equipment Variation):	14.99	13.27	17.29
Part Variation (PV):	78.17	56.91	128.78
Total Variation (TV):	81.40	61.21	135.83

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.091428538	7.76
Operator:	0.051455261	4.37
Part * Operator:	0	0.00
Reproducibility:	0.051455261	4.37
Repeatability:	0.039973276	3.39
Part Variation:	1.09	92.24
Total Variation:	1.1779	100.00

Gage R&R Metrics	NDC	NDC Lower 90% CI	NDC Upper 90% CI
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	4.9	1.4	9.0

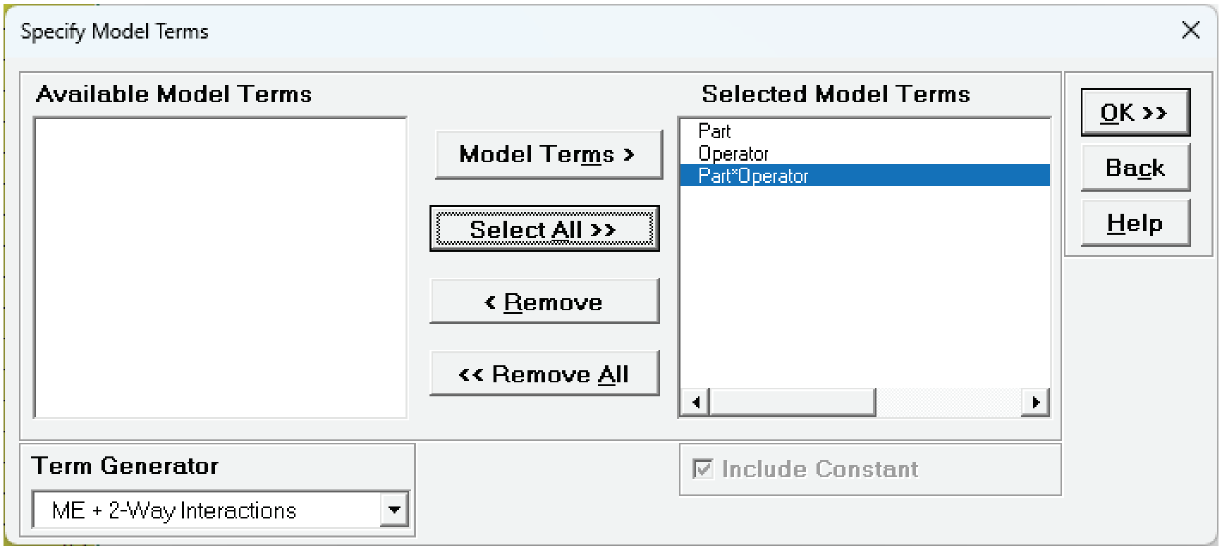
1. Open the file **Gage RR – AIAG.xlsx**. This is an example from the Automotive Industry Action Group (AIAG) MSA Reference Manual, 3rd Edition, page 101. Note that parts were measured in random order, but the worksheet is given in standard order. Preselect the worksheet data including column headings. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. Click **Next >>**.

2. Select *Measurement*, click **Numeric Response (Y) >>**, select *Part* and *Operator*, click **Random Factors - Categorical (X) >>**. Leave **Nesting** unchecked. We will use the default **Coding for Categorical Predictors (-1, 0, +1).** Set the **Confidence Level** = 90% to match the previous Gage R&R analysis. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Using **Term Generator**, select *ME + 2-Way Interactions*. Click **Select ALL >>**. **Include Constant** is always checked in General Linear Model.



5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Part	Random	10
Operator	Random	3

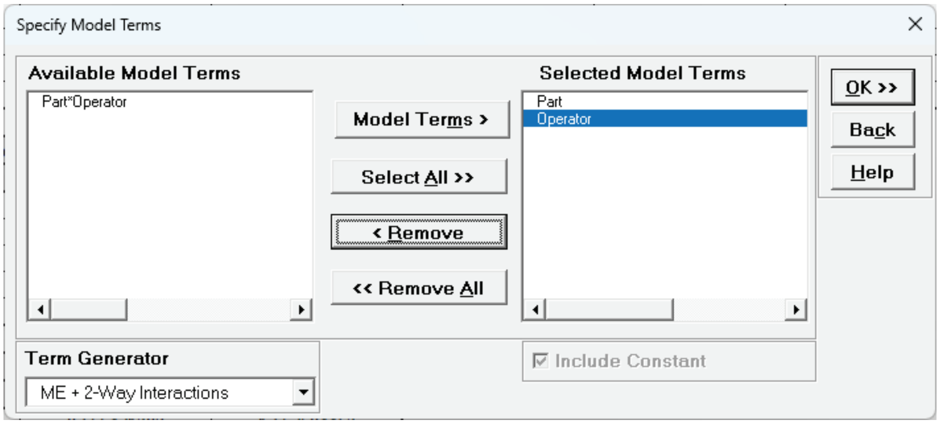
GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Part	9	88.36193444	9.817992716	18.0000	0.019943457	492.2914	0.0000
Operator	2	3.167262222	1.583631111	18.0000	0.019943457	79.4060	0.0000
Part*Operator	18	0.358982222	0.019943457	60.0000	0.045982222	0.4337	0.9741
Error	60	2.758933333	0.045982222				
Total	89	94.64711222	1.063450699				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 90% CI	Var Upper 90% CI
Part	1.08867214	91.73%	0.514250068	0.498750323	2.366647556
Operator	0.052122922	4.39%	0.052788135	0.009518636	0.278153715
Part*Operator	-0.008679588	0.00%			
Error	0.045982222	3.87%	0.006399252	0.030719246	0.052015037
Total	1.186777284				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 90% CI	StDev Upper 90% CI
Part	1.043394527	95.78%	0.706222573	1.538391223
Operator	0.22830445	20.96%	0.097563496	0.527402802
Part*Operator	0	0.00%		
Error	0.214434657	19.68%	0.175269069	0.228068053
Total	1.089393081			

6. Since the *Part*Operator* interaction term is not significant (alpha = 0.1), we will refit the model excluding this term. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog.

7. Click **Next >>**. Select *Part*Operator*, click **< Remove**.



8. Click **OK>>**. The revised Variance Components report is shown:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Part	Random	10
Operator	Random	3

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Part	9	88.36193444	9.817992716	78.0000	0.039973276	245.6139	0.0000
Operator	2	3.167262222	1.583631111	78.0000	0.039973276	39.6172	0.0000
Error	78	3.117915556	0.039973276				
Total	89	94.64711222	1.063450699				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 90% CI	Var Upper 90% CI
Part	1.086446604	92.24%	0.514250068	0.498750323	2.366647556
Operator	0.051455261	4.37%	0.052788135	0.009518636	0.278153715
Error	0.039973276	3.39%	0.006400847	0.030717231	0.052018452
Total	1.177875142				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 90% CI	StDev Upper 90% CI
Part	1.042327494	96.04%	0.706222573	1.538391223
Operator	0.226837521	20.90%	0.097563496	0.527402802
Error	0.19993318	18.42%	0.175263318	0.22807554
Total	1.085299563			

9. Now we will convert the GLM Variance Components to Gage R&R Metrics. Click on **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > GLM GageRR (Crossed) Metrics without Interaction** to open the conversion template.

10. Copy the Variance Component values in cells **C18:C20** (Sheet **GLM2 – VarComp**) and paste the values into the yellow highlighted cells **C15:C17** of the template. For Gage Name, enter the information as shown. Enter Process Tolerance =8, use Standard Deviation Multiplier = 6.

GLM Gage R&R Metrics without Interaction

Gage Name:	AIAG Example, Convert GLM to Gage R&R
Date of Study:	
Performed By:	
Notes:	

Process Tolerance (USL - LSL):	8
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6

GLM Variance Components	
Source	Variance
Part	1.086446604
Operator	0.051455261
Error	0.039973276

Source	Positive Variance
Part	1.086446604
Operator	0.051455261
Error	0.039973276
Total	1.177875142

Gage R&R Metrics	StDev	6 * StDev	% Total Variation (TV)	% Tolerance	% Process (Historical StDev)
Gage R&R:	0.302371522	1.814229134	27.86%	22.68%	
Reproducibility (AV Appraiser Variation):	0.226837521	1.361025129	20.90%	17.01%	
Repeatability (EV Equipment Variation):	0.19993318	1.199599078	18.42%	14.99%	
Part Variation (PV):	1.042327494	6.253964963	96.04%	78.17%	
Total Variation (TV):	1.085299563	6.511797379	100.00%	81.40%	

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.091428538	7.76%
Reproducibility:	0.051455261	4.37%
Repeatability:	0.039973276	3.39%
Part Variation:	1.086446604	92.24%
Total Variation:	1.177875142	100.00%

Gage R&R Metrics	NDC
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	4.9

Warning: It is crucial to ensure that the GLM Variance Components Source Names match those given in the template. If the GLM model order was different than that of the template, each entry would have to be copy/pasted individually.

11. This template converts the GLM Variance Components to Gage R&R metrics. The results match those given in the original analysis, excluding the confidence intervals and ANOVA table. With %Total Variation and %Tolerance less than 30% but greater than 10%, this is a marginal measurement system.

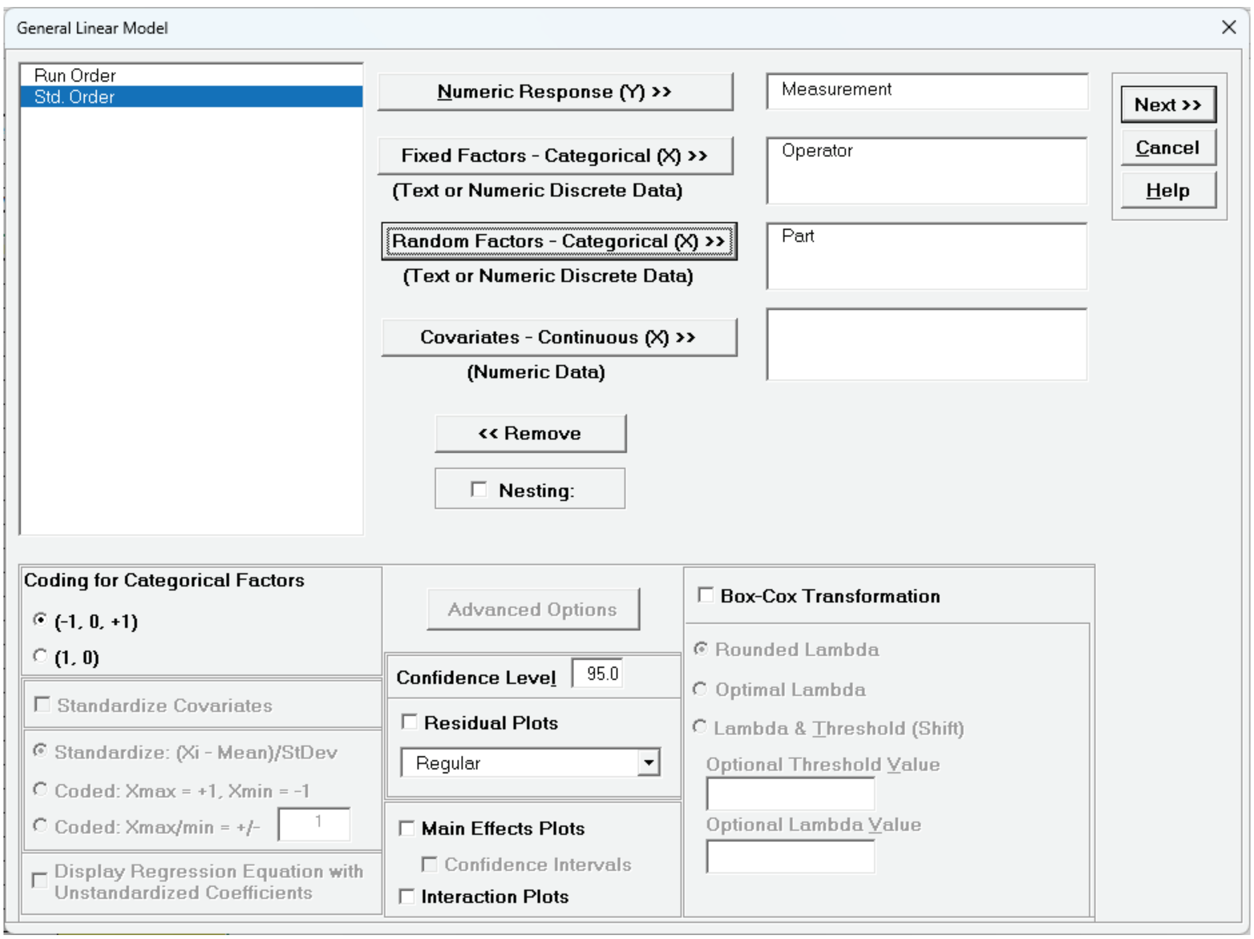
12. Next, we will show examples where the classical Gage R&R Analysis does not work and GLM is required to do the analysis.

Example 4: Gage R&R Study with Operator as Fixed Factor

We will reanalyze the AIAG Gage R&R data above, but now we will treat Operator as a Fixed Factor. In this case "Operator" could denote a test fixture and there are only three of them in the plant.

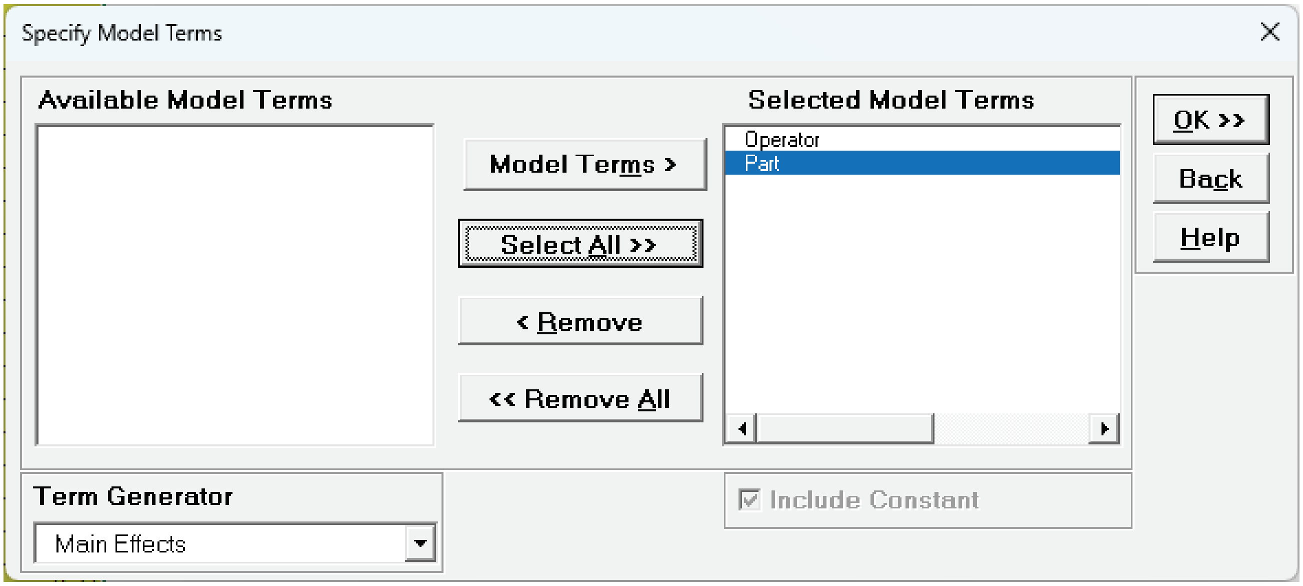
1. Open the file **Gage RR – AIAG.xlsx**. Preselect the worksheet data including column headings. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. Click **Next >>**.

2. Select *Measurement*, click **Numeric Response (Y) >>**, select *Operator*, click **Fixed Factors - Categorical (X) >>**, select *Part*, click **Random Factors - Categorical (X) >>**. Leave **Nesting** unchecked. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Leave **Term Generator** as *Main Effects*. Click **Select ALL >>**.



5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Operator	Fixed	3
Part	Random	10

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Operator	2	3.167262222	1.583631111	78.0000	0.039973276	39.6172	0.0000
Part	9	88.36193444	9.817992716	78.0000	0.039973276	245.6139	0.0000
Error	78	3.117915556	0.039973276				
Total	89	94.64711222	1.063450699				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
Part	1.086446604	96.45%	0.514250034	0.429642627	2.747320868
Error	0.039973276	3.55%	0.006400847	0.029205754	0.054710551
Total	1.126419881				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
Part	1.042327494	98.21%	0.655471301	1.65750441
Error	0.19993318	18.84%	0.17089691	0.233902868
Total	1.061329299			

Note that *Operator* is no longer in the Variance Components table. We will need to manually calculate the VC value to be entered into the GLM GageRR template.

6. Click on Sheet **GLM# Model** for the current model. The Parameter Estimates are given as:

Parameter Estimates						
Predictor Term	**Coefficient**	**SE Coefficient**	**T**	**P**	**VIF**	**Tolerance**
Constant	0.001444444	0.021074808	0.0685	0.9455		
Operator_Operator A	0.188888889	0.029804279	6.3376	0.0000	1.3333	0.7500
Operator_Operator B	0.066888889	0.029804279	2.2443	0.0276	1.3333	0.7500
Part_Part 01	0.167444444	0.063224423	2.6484	0.0098	1.8000	0.5556
Part_Part 02	-0.852555556	0.063224423	-13.4846	0.0000	1.8000	0.5556
Part_Part 03	1.097444444	0.063224423	17.3579	0.0000	1.8000	0.5556
Part_Part 04	0.365222222	0.063224423	5.7766	0.0000	1.8000	0.5556
Part_Part 05	-1.065888889	0.063224423	-16.8588	0.0000	1.8000	0.5556
Part_Part 06	-0.187	0.063224423	-2.9577	0.0041	1.8000	0.5556
Part_Part 07	0.453	0.063224423	7.1650	0.0000	1.8000	0.5556
Part_Part 08	-0.343666667	0.063224423	-5.4357	0.0000	1.8000	0.5556
Part_Part 09	1.938555556	0.063224423	30.6615	0.0000	1.8000	0.5556

7. The coded **(-1, 0, 1)** coefficients for Operator will be used to estimate the Fixed Factor Variance Component. It is the average of the coefficients squared, including the hidden reference value. (This is from Formula 6.2 in Burdick, R. K., Borror, C. M., and Montgomery, D. C., "Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models", ASA-SIAM Series on Statistics and Applied Probability, 2005.) Note, this calculation cannot be done with coded **(0, 1)** coefficients.

   Operator C coefficient (hidden reference) = -(0.18889 + 0.06689) = - 0.25578

   Fixed Operator VC = ((0.18889)^2 + (0.06689)^2 + (-0.25578)^2)/3 = 0.035192.

   Use Excel formulas to calculate this to full precision.

8. Now we will now convert the GLM and Manual Variance Components to Gage R&R Metrics. Click on **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > GLM GageRR (Crossed) Metrics without Interaction** to open the conversion template.

9. Enter the Variance Component values into the yellow highlighted cells **C15:C17** of the template as shown. For Gage Name, enter the information as shown. Enter Process Tolerance =8, use Standard Deviation Multiplier = 6.

GLM Gage R&R Metrics without Interaction

Gage Name:	AIAG Example with Fixed Operator, Convert GLM to AIAG R&R
Date of Study:	
Performed By:	
Notes:	

Process Tolerance (USL - LSL):	8
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6

GLM Variance Components	
Source	Variance
Part	1.086446604
Operator	0.035191802
Error	0.039973276

Source	Positive Variance
Part	1.086446604
Operator	0.035191802
Error	0.039973276
Total	1.161611683

Gage R&R Metrics	StDev	6 * StDev	% Total Variation (TV)	% Tolerance	% Process (Historical StDev)
Gage R&R:	0.274162504	1.644975026	25.44%	20.56%	
Reproducibility (AV Appraiser Variation):	0.187594783	1.125568696	17.41%	14.07%	
Repeatability (EV Equipment Variation):	0.19993318	1.199599078	18.55%	14.99%	
Part Variation (PV):	1.042327494	6.253964963	96.71%	78.17%	
Total Variation (TV):	1.077780907	6.466685441	100.00%	80.83%	

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.075165079	6.47%
Reproducibility:	0.035191802	3.03%
Repeatability:	0.039973276	3.44%
Part Variation:	1.086446604	93.53%
Total Variation:	1.161611683	100.00%

Gage R&R Metrics	NDC
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	5.4

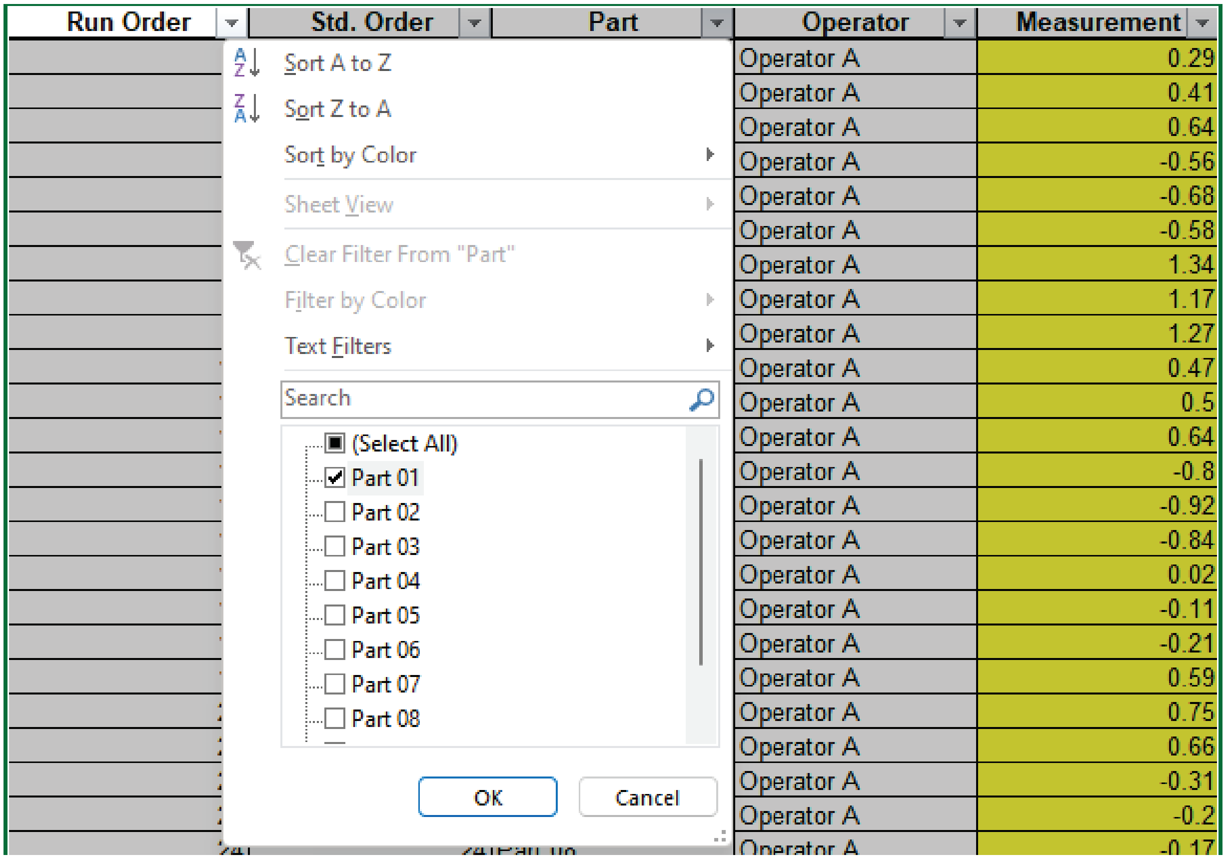
Warning: It is crucial to ensure that the GLM Variance Components Source Names match those given in the template.

10. This template converts the GLM Variance Components to Gage R&R metrics. Note that the %Total Variation has been reduced from the original 27.86% to 25.44%, %Tolerance from 22.68% to 20.56% and NDC increased from 4.9 to 5.4. However, it is still considered a marginal measurement system. Note that these steps should only be done when Operator is a fixed factor.

Example 5: Gage R&R Study with Operator and One Part

We will reanalyze the AIAG Gage R&R data, but now we will consider the case where there is only one part, i.e., like a Type 1 Gage R&R Study but with multiple operators.

1. Open the file **Gage RR – AIAG.xlsx**. Preselect the worksheet data including column headings. Click **Excel > Data > Filter**, select Part, uncheck Select All, check Part 01 as shown.



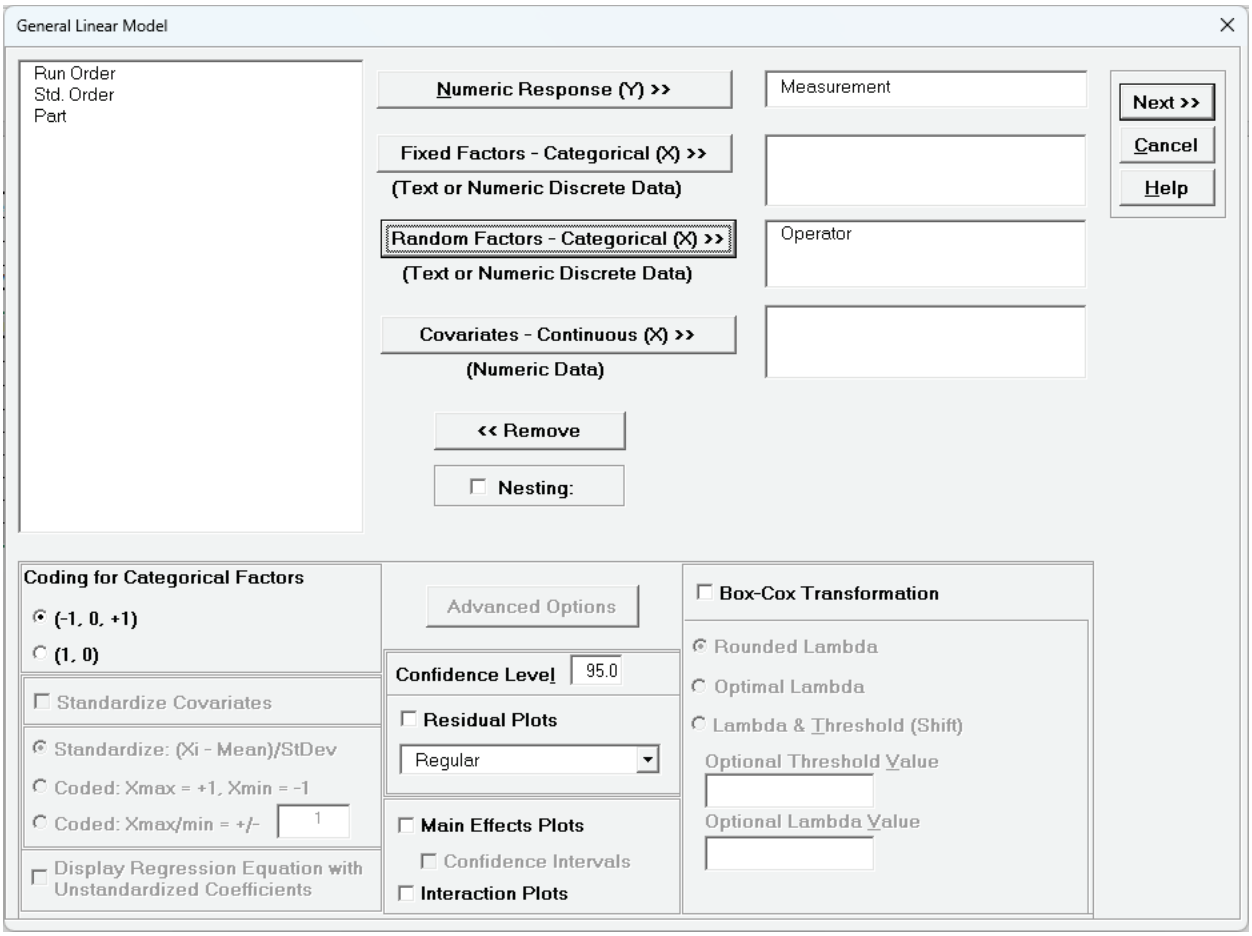
2. Click OK. This hides the rows for Parts 02 to 10.

Run Order	Std. Order	Part	Operator	Measurement
1	1	Part 01	Operator A	0.29
2	2	Part 01	Operator A	0.41
3	3	Part 01	Operator A	0.64
31	31	Part 01	Operator B	0.08
32	32	Part 01	Operator B	0.25
33	33	Part 01	Operator B	0.07
61	61	Part 01	Operator C	0.04
62	62	Part 01	Operator C	-0.11
63	63	Part 01	Operator C	-0.15

Note: Three measurement readings per operator for a study with one part is not recommended. We are using this example with the AIAG data for convenience.

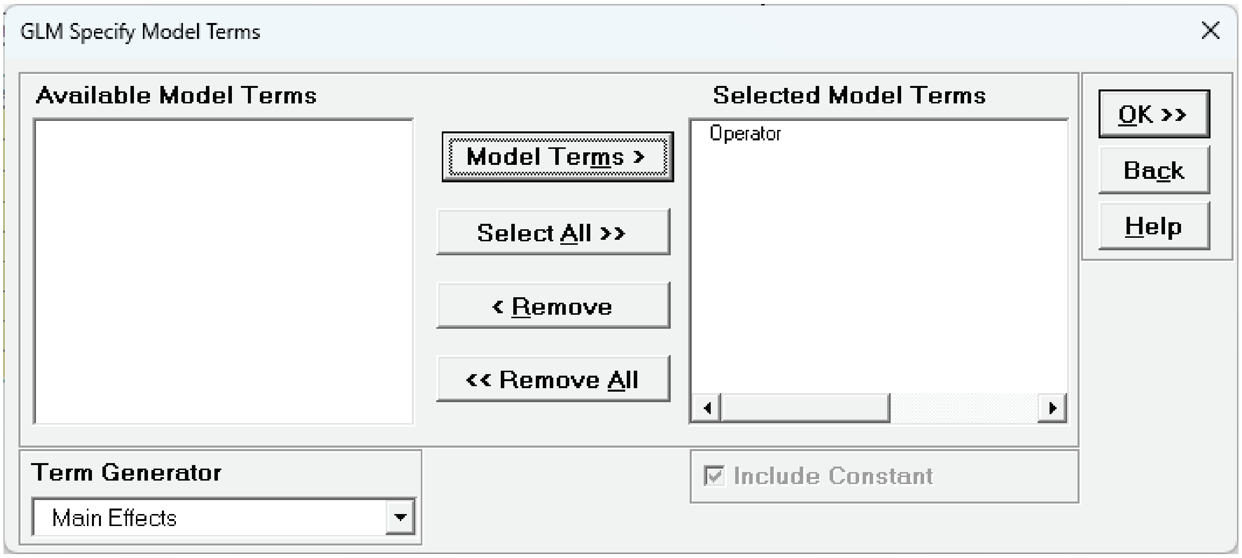
3. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. Click **Next >>**.

4. Select *Measurement*, click **Numeric Response (Y) >>**, select *Operator*, click **Random Factors - Categorical (X) >>**. Leave **Nesting** unchecked. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



5. Click **Next >>.**

6. Click **Model Terms >>**.



7. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Measurement
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Operator	Random	3

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Operator	2	0.411288889	0.205644444	6.0000	0.0173	11.8870	0.0082
Error	6	0.1038	0.0173				
Total	8	0.515088889	0.064386111				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
Operator	0.062781481	78.40%	0.068628955	0.00736799	0.534951115
Error	0.0173	21.60%	0.00998816	0.005579621	0.053639849
Total	0.080081481				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
Operator	0.250562331	88.54%	0.085836996	0.731403524
Error	0.131529464	46.48%	0.074696858	0.231602782
Total	0.282986716			

8. Now we will now convert the GLM and Manual Variance Components to Gage R&R Metrics. Click on **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > GLM GageRR (Crossed) Metrics without Interaction** to open the conversion template.

9. Copy the Variance Component values in cells **C16:C17** (Sheet **GLM# – VarComp** for the current model) and paste the values into the yellow highlighted cells **C16:C17** of the template. Enter 0 for Part Variance. For Gage Name, enter the information as shown. Enter Process Tolerance =8, use Standard Deviation Multiplier = 6.

GLM Gage R&R (Crossed) Metrics without Interaction

Gage Name:	AIAG Example, Operator with One Part
Date of Study:	
Performed By:	
Notes:	

Process Tolerance (USL - LSL):	8
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6

GLM Variance Components	
Source	**Variance**
Part	0
Operator	0.062781481
Error	0.0173

Source	Positive Variance
Part	0
Operator	0.062781481
Error	0.0173
Total	0.080081481

Gage R&R Metrics	StDev	6 * StDev	% Total Variation (TV)	% Tolerance	% Process (Historical StDev)
Gage R&R:	0.282986716	1.697920297	100.00%	21.22%	
Reproducibility (AV Appraiser Variation):	0.250562331	1.503373983	88.54%	18.79%	
Repeatability (EV Equipment Variation):	0.131529464	0.789176786	46.48%	9.86%	
Part Variation (PV):	0	0	0.00%	0.00%	
Total Variation (TV):	0.282986716	1.697920297	100.00%	21.22%	

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.080081481	100.00%
Reproducibility:	0.062781481	78.40%
Repeatability:	0.0173	21.60%
Part Variation:	0	0.00%
Total Variation:	0.080081481	100.00%

Gage R&R Metrics	NDC
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	0.0

Warning: It is crucial to ensure that the GLM Variance Components Source Names match those given in the template.

10. This template converts the GLM Variance Components to Gage R&R metrics. The %Total Variation is not applicable here because there is no part variation. With %Tolerance less than 30% but greater than 10%, this is a marginal measurement system.

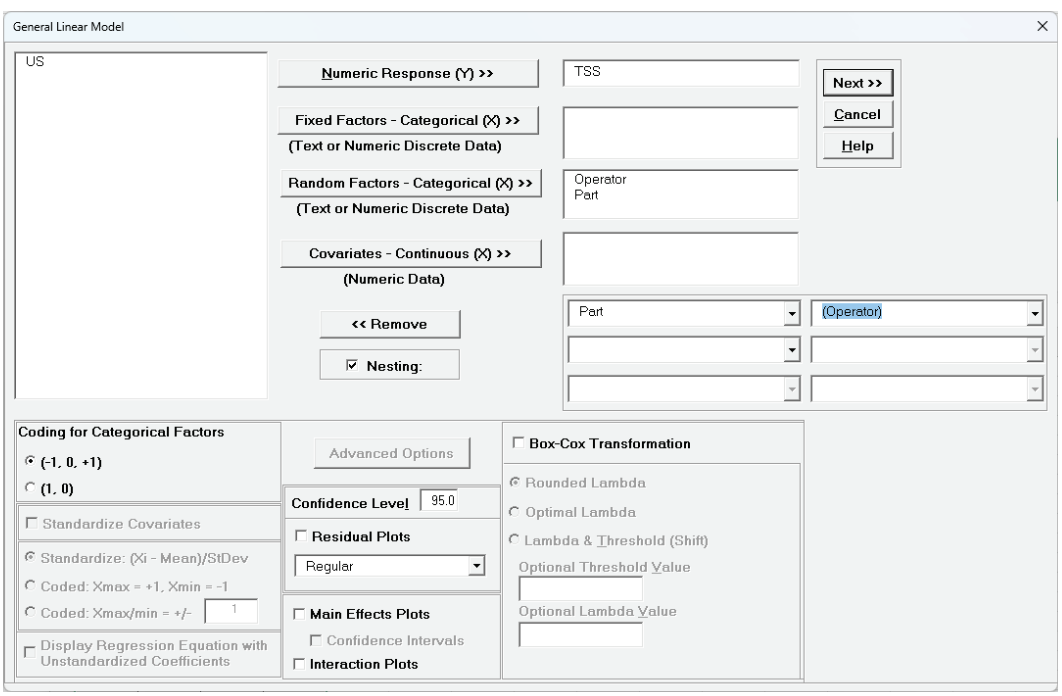
Example 6: Destructive (Nested) Gage R&R

We will now analyze data from a Destructive (Nested) Gage R&R found in the paper referenced below with link to the PDF. This is a shear test for a resistance spot welding process. Design of Experiments were used to produce the parts by varying preheat temperature, preheat current, welding temperature and welding current. Each run of a 4 Factor 8-Run Fractional Factorial DOE is a "Part" and this was replicated 6 times for a total of 48 welds. Two Shear Test machines were used so each machine tested 8 Parts with 3 Replicates. The shear test machines are the "Operator" and treated as a random factor in the study. Tensile Shear Strength (TSS Newtons) and Ultimate Strain (US mm) were measured, but we will only consider TSS here.

Reference: Almeida, F., Gomes, G., Sabioni, R., Gomes, J., de Paula, V., de Paiva, A., & da Costa, S. (2018). "A Gage Study Applied in Shear Test to Identify Variation Causes from a Resistance Spot Welding Measurement System", *Strojniški vestnik - Journal of Mechanical Engineering*, 64(10), 621-631. https://www.sv-jme.eu/?ns_articles_pdf=/ns_articles/files/ojs/5235/public/5235-30318-1-PB.pdf&id=6160

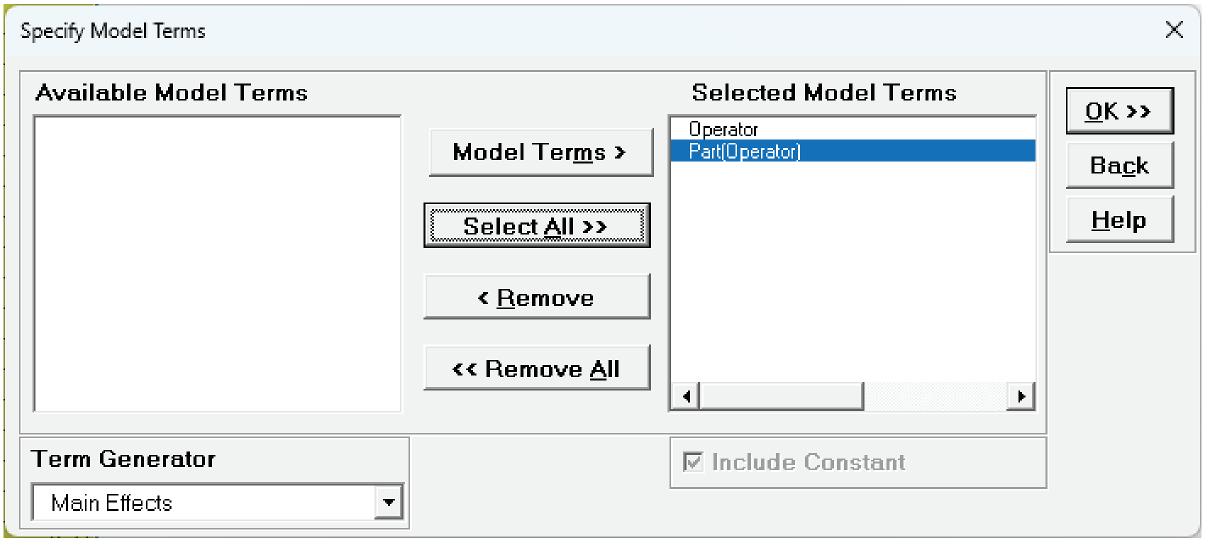
1. Open the file **Tensile Shear Strength.xlsx**. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *TSS*, click **Numeric Response (Y) >>**, select *Operator*, click **Random Factors - Categorical (X) >>**, select *Part*, click **Random Factors - Categorical (X) >>.** Check **Nesting**. For the left-side drop-down "Select a Factor or Covariate", select *Part*. For the right-side drop-down "Select a Factor to Nest in:", select *(Operator)*. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Leave **Term Generator** as *Main Effects*. Click **Select ALL >>**.



The term *Part(Operator)* denotes that *Part* is nested within *Operator*. Since *Operator* is the top level of the nesting, we place it in the model first.

5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	TSS
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Operator	Random	2
Part(Operator)	Random	16

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Operator	1	1952798.89	1952798.89	14.0000	5516671.154	0.3540	0.5614
Part(Operator)	14	77233396.16	5516671.154	32.0000	174940.3914	31.5346	0.0000
Error	32	5598092.524	174940.3914				
Total	47	84784287.57	1803921.012				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
Operator	-148494.6777	0.00%			
Part(Operator)	1780576.921	91.05%	642712.0241	811431.7198	3567389.766
Error	174940.3914	8.95%	43735.05349	107174.0297	285555.2436
Total	1955517.312				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
Operator	0	0.00%		
Part(Operator)	1334.382599	95.42%	900.7950487	1888.753495
Error	418.2587613	29.91%	327.3744488	534.3736928
Total	1398.398124			

Note that *Operator* is a negative variance component. It will be treated as zero.

6. Now we will now convert the GLM Variance Components to Gage R&R Metrics. Click on **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > GLM GageRR (Nested) Metrics** to open the conversion template.

7. Copy the Variance Component values in cells **C18:C20** (Sheet **GLM1 – VarComp**) and paste the values into the yellow highlighted cells **C15:C17** of the template. For Gage Name, enter the information as shown. Use Standard Deviation Multiplier = 6.

GLM Gage R&R (Nested) Metrics

Gage Name:	Tensile Shear Strength, Convert GLM to Gage R&R
Date of Study:	
Performed By:	
Notes:	

Process Tolerance (USL - LSL):	
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6

GLM Variance Components	
Source	**Variance**
Operator	-148494.6777
Part(Operator)	1780576.921
Error	174940.3914

Source	Positive Variance
Operator	0
Part(Operator)	1780576.921
Error	174940.3914
Total	1955517.312

Gage R&R Metrics	StDev	6 * StDev	% Total Variation (TV)	% Tolerance	% Process (Historical StDev)
Gage R&R:	418.2587613	2509.552568	29.91%		
Reproducibility (AV Appraiser Variation):	0	0	0.00%		
Repeatability (EV Equipment Variation):	418.2587613	2509.552568	29.91%		
Part Variation (PV):	1334.382599	8006.295595	95.42%		
Total Variation (TV):	1398.398124	8390.388742	100.00%		

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	174940.3914	8.95%
Reproducibility:	0	0.00%
Repeatability:	174940.3914	8.95%
Part Variation:	1780576.921	91.05%
Total Variation:	1955517.312	100.00%

Gage R&R Metrics	NDC
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	4.5

Warning: It is crucial to ensure that the GLM Variance Components Source Names match those given in the template. If the GLM model order was different than that of the template, each entry would have to be copy/pasted individually.

8. This template converts the GLM Variance Components to Gage R&R metrics. With %Total Variation less than 30% but greater than 10%, this is a marginal measurement system. The results match those given in the paper.

Example 7: Expanded Gage R&R

We will now analyze data from an Expanded Gage R&R found in the paper referenced below with link to the PDF. This is a wafer thickness measurement (Angstroms) for a semiconductor process. The design includes random factors: Batch, Wafer, Location and Operator. Wafer is nested within Batch and Location is nested within Wafer. There are 3 Batches, 3 Wafers, 4 Locations, 3 Operators and 2 Replicates for a total of 216 thickness measurements. The variance components for Part/Process include Batch, Wafer(Batch) and Location(Wafer(Batch)). The variance components for Reproducibility include Operator, Batch*Operator, Wafer(Batch)*Operator and Location(Wafer(Batch))*Operator. The variance component for Repeatability is the Error term from Replicates.

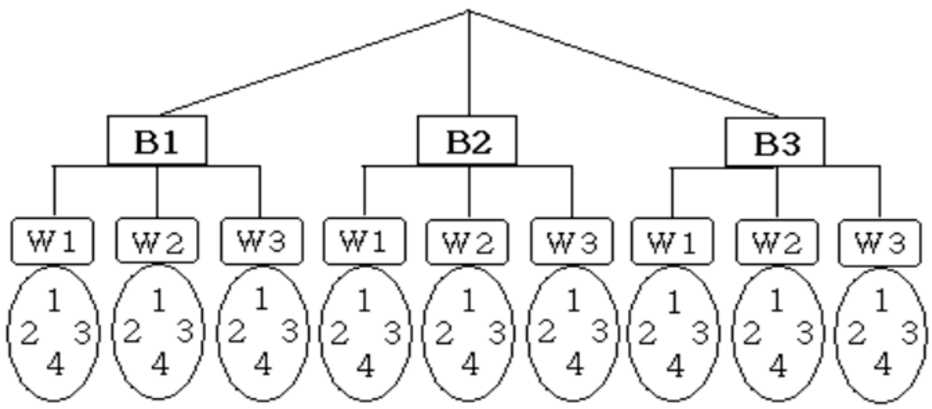
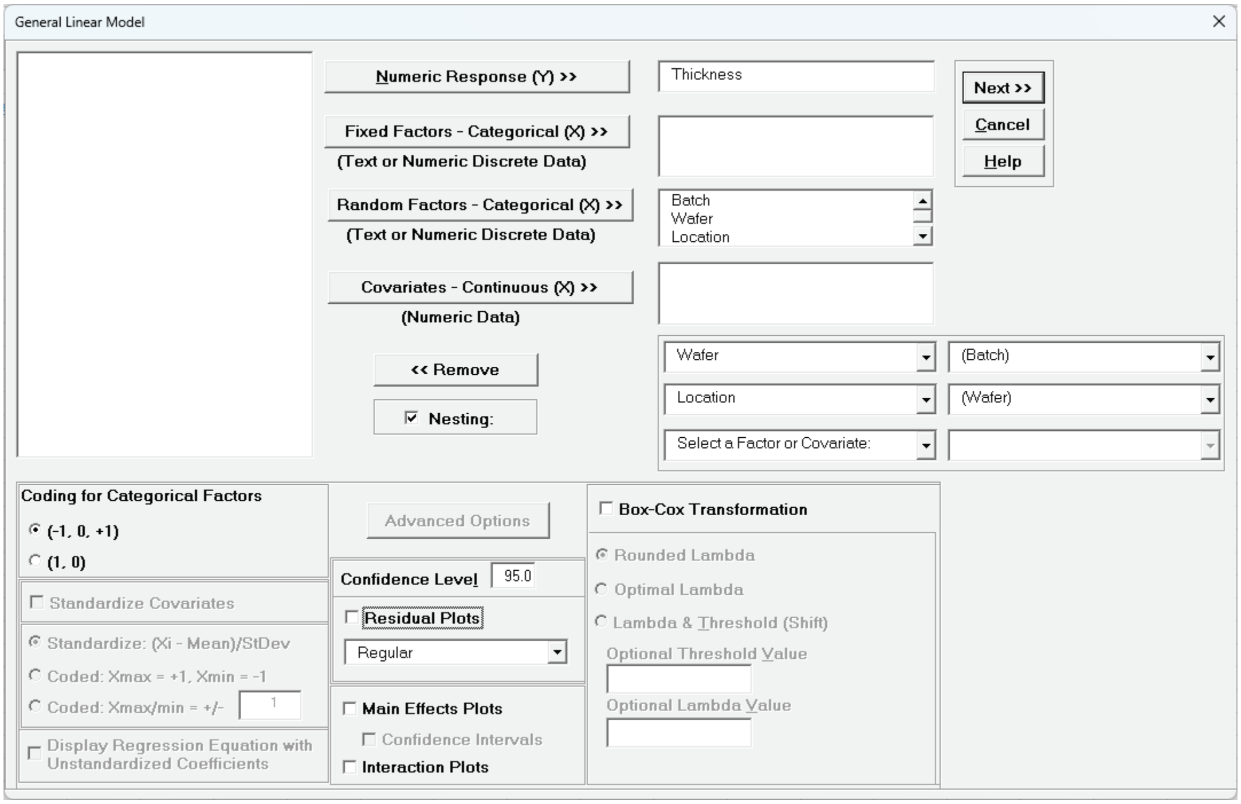


Figure 1. A measurement experiment.

Reference: Lee, S.H., Lee, C.W. (2005), "A Study of Gage R&R Analysis Considering the Variations of Between-Within Group and Within Part" *IE Interfaces* 18(4), 444-453. https://koreascience.kr/article/JAKO200529256810322.pdf. Written in Korean but abstract, formulas, figures and tables are in English.

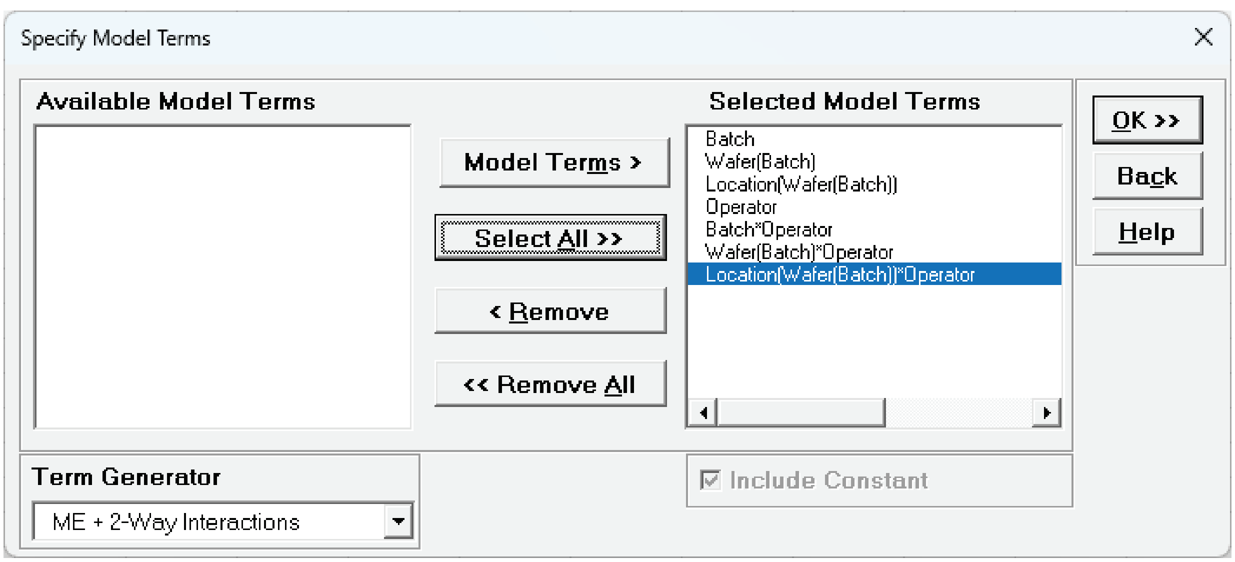
1. Open the file **Wafer Thickness.xlsx**. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Thickness*, click **Numeric Response (Y) >>**, select *Batch* to *Operator*, click **Random Factors - Categorical (X) >>.** Check **Nesting**. For the left-side drop-down "Select a Factor or Covariate", select *Wafer*. For the right-side drop-down "Select a Factor to Nest in:", select *(Batch)*. For the second left-side drop-down "Select a Factor or Covariate", select *Location*. For the right-side drop-down "Select a Factor to Nest in:", select *(Wafer)*. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Leave **Residual Plots**, **Main Effects Plots, Interaction Plots** and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Using **Term Generator**, select *ME + 2-Way Interactions*. Click **Select ALL >>**.



The term *Wafer(Batch)* denotes that *Wafer* is nested within *Batch*. The term *Location(Wafer(Batch))* denotes that *Location* is nested within *Wafer* which is nested within *Batch*. Only legal 2-Way Interactions are available, so *Wafer*Batch*, *Location*Wafer* or *Location*Batch* are not available.

5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Thickness
Factor Coding	(-1,0,+1)
Design/Model Type	Balanced; Hierarchical
Estimation Method	Expected Mean Squares with REML CI

Factor Information		
Factor	Type	Levels
Batch	Random	3
Wafer(Batch)	Random	9
Location(Wafer(Batch))	Random	36
Operator	Random	3

GLM Analysis of Variance							
Source	DF	SS	MS	Error DF	Error MS	F	P
Batch	2	8628.898148	4314.449074	6.0007	586.2546296	7.3593	0.0243
Wafer(Batch)	6	3517.305556	586.2175926	27.0499	61.17824074	9.5821	0.0000
Location(Wafer(Batch))	27	1649.833333	61.10493827	54.0000	0.804012346	76.0000	0.0000
Operator	2	6.731481481	3.365740741	4.0000	0.914351852	3.6810	0.1239
Batch*Operator	4	3.657407407	0.914351852	12.0000	0.877314815	1.0422	0.4256
Wafer(Batch)*Operator	12	10.52777778	0.877314815	54.0000	0.804012346	1.0912	0.3861
Location(Wafer(Batch))*Operator	54	43.41666667	0.804012346	108.0000	0.796296296	1.0097	0.4733
Error	108	86	0.796296296				
Total	215	13946.37037	64.86683893				

Variance Components					
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI
Batch	51.7804784	61.24%	60.10698551	5.322157438	503.7838921
Wafer(Batch)	21.87663966	25.87%	14.11922791	6.174665307	77.50823099
Location(Wafer(Batch))	10.05015432	11.89%	2.771896253	5.853369628	17.25597529
Operator	0.034047068	0.04%	0.047601087	0.002197951	0.52740171
Batch*Operator	0.00154321	0.00%	0.030796716	1.59E-20	1.4969E+14
Wafer(Batch)*Operator	0.009162809	0.01%	0.048769596	2.70E-07	310.8856182
Location(Wafer(Batch))*Operator	0.003858025	0.00%	0.09445163	5.59E-24	2.66E+18
Error	0.796296296	0.94%	0.108362204	0.609875016	1.039701242
Total	84.55217978				

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
Batch	7.195865368	78.26%	2.306980156	22.4451307
Wafer(Batch)	4.677247017	50.87%	2.484887383	8.803875907
Location(Wafer(Batch))	3.170197836	34.48%	2.419373809	4.154031209
Operator	0.184518476	2.01%	0.04688231	0.726224284
Batch*Operator	0.03928371	0.43%	1.26E-10	12234769.12
Wafer(Batch)*Operator	0.095722561	1.04%	0.000519671	17.63194879
Location(Wafer(Batch))*Operator	0.062112999	0.68%	2.36E-12	1631689106
Error	0.892354356	9.70%	0.780944951	1.019657414
Total	9.195225923			

Note that the variance component confidence intervals for the interaction terms with Operator are very wide and the P-Values are insignificant (alpha = 0.05), so the model could be refit excluding these terms but we will not do so here.

6. Now we will now convert the GLM Variance Components to Gage R&R Metrics. Click on **SigmaXL > Measurement Systems Analysis > Basic MSA Templates > GLM GageRR (Expanded) Metrics** to open the conversion template.

7. Copy the Variance Component values in Sheet **GLM1 – VarComp** and paste the values into the yellow highlighted cells of the template as shown. For Gage Name, enter the information as shown. Use Standard Deviation Multiplier = 6.

GLM Gage R&R (Expanded) Metrics

Gage Name:	Semiconductor Wafer Thickness, GLM to Gage R & R Metrics
Date of Study:	
Performed By:	
Notes:	

Process Tolerance (USL - LSL):	
Historical Process Standard Deviation:	
Standard Deviation Multiplier:	6

GLM Variance Components					
Error Source	Variance	Part/Process Source	Variance	Operator/Measurement Source	Variance
Error	0.796296296	Batch	51.7804784	Operator	0.034047068
		Wafer(Batch)	21.87663966	Batch*Operator	0.00154321
		Location(Wafer(Batch))	10.05015432	Wafer(Batch)*Operator	0.009162809
				Location(Wafer(Batch))*Operator	0.003858025

Gage R&R Metrics	StDev	6 * StDev	% Total Variation (TV)	% Tolerance	% Process (Historical StDev)
Gage R&R:	0.91918845	5.515130703	10.00%		
Reproducibility (AV Appraiser Variation):	0.220479276	1.322875656	2.40%		
Repeatability (EV Equipment Variation):	0.892354356	5.354126135	9.70%		
Part Variation (PV):	9.149167852	54.89500711	99.50%		
Total Variation (TV):	9.195225923	55.17135554	100.00%		

Gage R&R Metrics	Variance Component	% Contribution of Variance Component
Gage R&R:	0.844907407	1.00%
Reproducibility:	0.048611111	0.06%
Repeatability:	0.796296296	0.94%
Part Variation:	83.70727238	99.00%
Total Variation:	84.55217978	100.00%

Gage R&R Metrics	NDC
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	14.0

8. This template converts the GLM Variance Components to Gage R&R metrics. With %Total Variation = 10%, this is an acceptable measurement system. The results match those given in the paper.

Example 8: Unbalanced Nested Factorial Experiment with Fixed and Random Factors

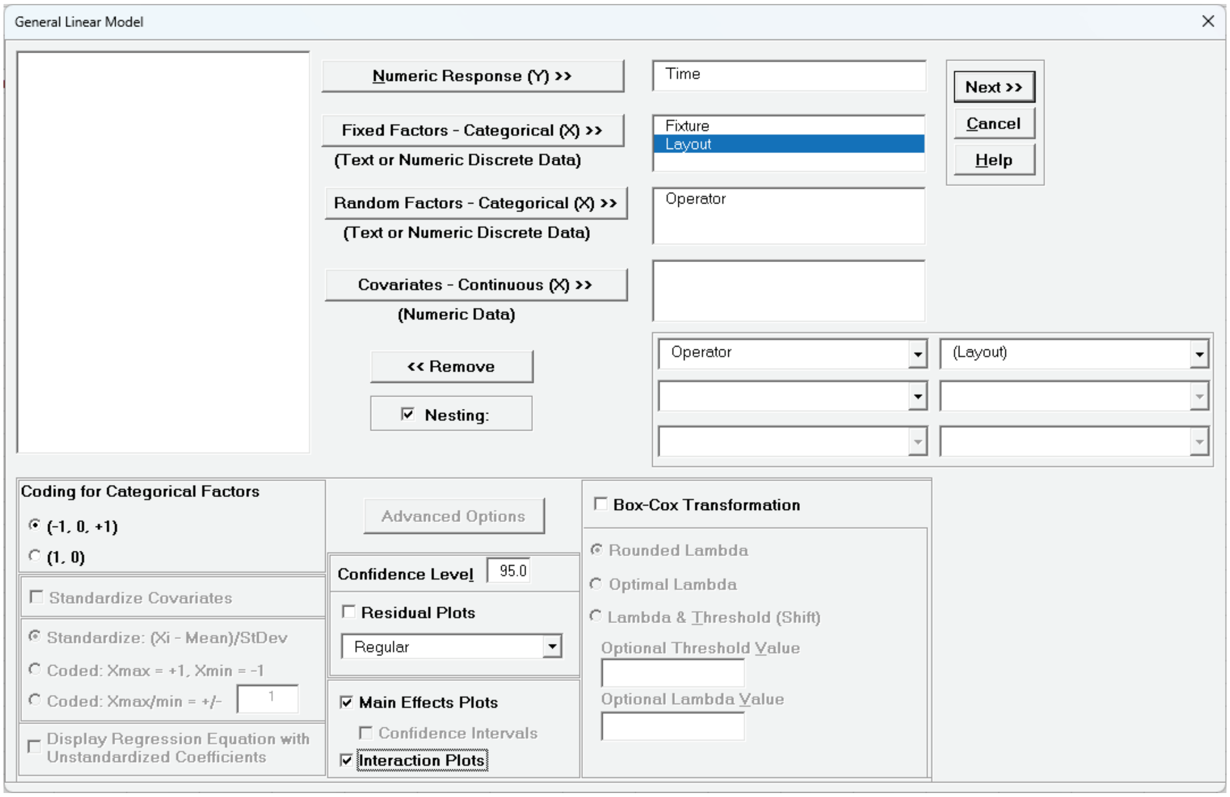
We will now analyze a nested factorial experiment adapted from Example 14.2 in Montgomery's Design and Analysis of Experiments book. The process is the hand insertion of electronic components on printed circuit boards and the goal of the study is to improve the speed of the assembly operation. The design includes three assembly fixtures and two workplace layouts. Operators are required to perform the assembly.

In the book, four operators are randomly selected for each fixture–layout combination, but here we will use 4 operators for layout 1, but three for layout 2 making it an unbalanced design. The operators chosen for layout 1 are different individuals from those chosen for layout 2, so operators are nested within layout. Because there are only three fixtures and two layouts, these are fixed factors but the operators are random factors, so this is a mixed model. The treatment combinations in this design are run in random order, but the worksheet is shown in standard order. Two replicates are obtained. The assembly times are measured in seconds.

Since the design is unbalanced, SigmaXL will automatically use Restricted Maximum Likelihood (REML) to calculate variance components rather than the traditional ANOVA Expected Means Squares (EMS) method.

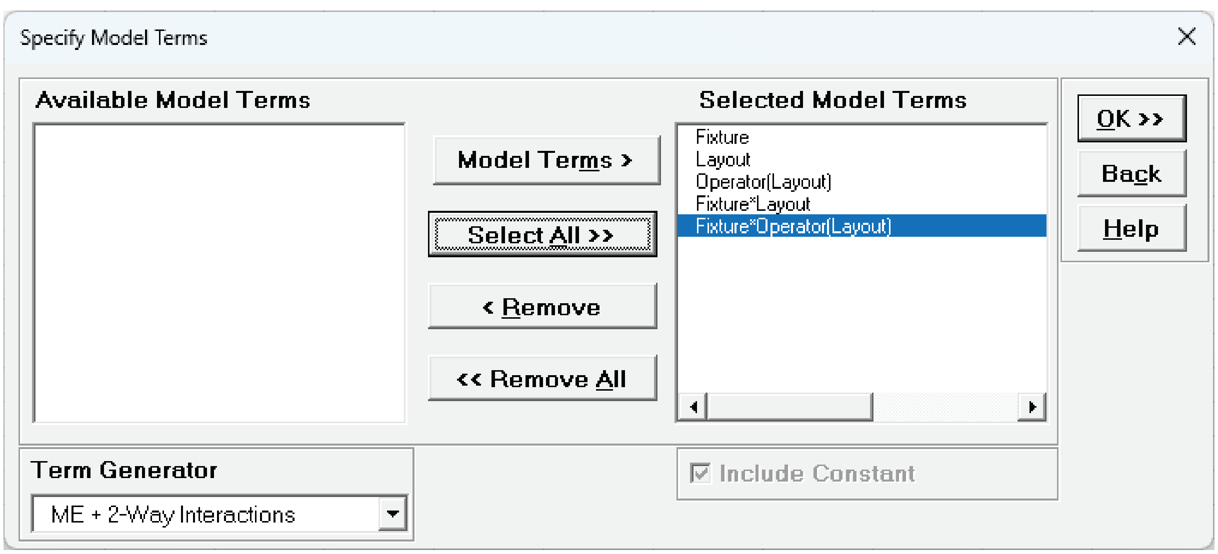
Reference: Montgomery, D.C. (2020). *Design and Analysis of Experiments*, 10th Edition, John Wiley & Sons.

1. Open the file **Assembly Time Nested Factorial.xlsx**. Click **SigmaXL > Statistical Tools > General Linear Model > Fit General Linear Model**. If necessary, click **Use Entire Data Table**, click **Next**.
2. Select *Time*, click **Numeric Response (Y) >>**, select *Fixture* and *Layout*, click **Fixed Factors - Categorical (X) >>**, select *Operator*, click **Random Factors - Categorical (X) >>**. Check **Nesting**. For the left-side drop-down "Select a Factor or Covariate", select *Operator*. For the right-side drop-down "Select a Factor to Nest in:", select *(Layout)*. We will use the default **Coding for Categorical Predictors (-1, 0, +1)** and default **Confidence Level** = 95.0%. Check **Main Effects Plots** and **Interaction Plots**. Leave **Residual Plots**, and **Box-Cox Transformation** unchecked. **Advanced Options** are not available with Random Factors.



3. Click **Next >>.**

4. Using **Term Generator**, select *ME + 2-Way Interactions*. Click **Select ALL >>**.



The term *Operator(Layout)* denotes that *Operator* is nested within *Layout.* Only legal 2-Way Interactions are available, so *Operator*Layout* is not available.

5. Click **OK >>**. The Variance Components report is given:

GLM Information	
Response	Time
Factor Coding	(-1,0,+1)
Design/Model Type	Unbalanced; Hierarchical
Estimation Method	Restricted Maximum Likelihood (REML)
Fixed Effects Test Method	Satterthwaite Approximation

Factor Information		
Factor	Type	Levels
Fixture	Fixed	3
Layout	Fixed	2
Operator(Layout)	Random	7

Tests of Fixed Effects				
Source	DF Num	DF Den	F	P
Fixture	2.0000	10.0000	7.4193	0.0106
Layout	1.0000	5.0000	0.1706	0.6967
Fixture*Layout	2.0000	10.0000	3.0115	0.0947

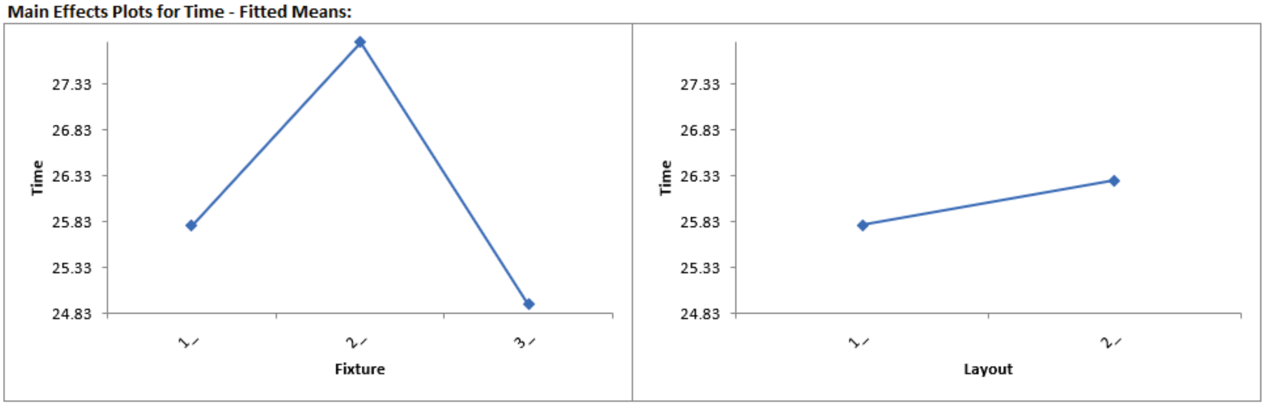
Variance Components							
Source	Variance	Var % of Total	SE Var for CI	Var Lower 95% CI	Var Upper 95% CI	Z	P
Operator(Layout)	1.708333204	34.38%	1.531057059	0.294918097	9.89563667	1.1158	0.1323
Fixture*Operator(Layout)	0.736706428	14.83%	0.974960065	0.055056821	9.857749794	0.7556	0.2249
Error	2.523809753	50.79%	0.77886467	1.378383581	4.621076279	3.2404	0.0006
Total	4.968849385						

StDev Components				
Source	StDev	StDev % of Total	StDev Lower 95% CI	StDev Upper 95% CI
Operator(Layout)	1.307032212	58.64%	0.543063621	3.14573309
Fixture*Operator(Layout)	0.858316042	38.51%	0.234641899	3.139705367
Error	1.588650293	71.27%	1.174045817	2.149668877
Total	2.229091605			

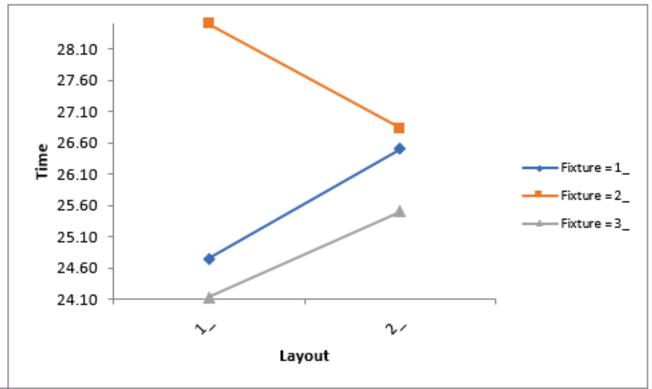
Since this is an unbalanced design, Restricted Maximum Likelihood (REML) is used to estimate the variance components. In **Tests of Fixed Effects**, we see that *Fixture* is significant but *Layout* and *Fixture**Layout* are not significant. In **Variance Components**, *Operator(Layout)* contributes to 34.4% of the total variance and the interaction *Fixture**Operator(Layout)* contributes to 14.8% of the total variance. Together they contribute to almost 50% of the total variance indicating that operator training upgrade may be necessary. The P-Values for these variance components are not significant but this test is known to be underpowered when there are only a few levels (see Appendix **Random Factors**).

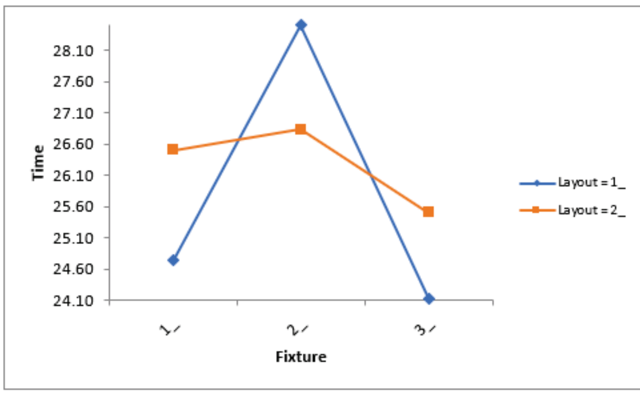
6. Click Sheet **GLM1 – Plots** to view the Main Effects and Interaction Plots.

Main Effects Plots for Time - Fitted Means:



Interaction Plots for Time - Fitted Means:





Clearly Fixture 2 results in an increase in assembly time. This may be an operator training issue but possibly the fixture can be modified to help the operators perform the assembly task more quickly.

Note: The Main Effects and Interaction Plots are computed from the predicted values of the regression model which assumes that all factors are fixed. Confidence Intervals for the Main Effects plots are not available when there is a Random Factor.

Part W – Logistic Regression

Binary Logistic Regression

Binary Logistic Regression is used to analyze the relationship between one binary dependent variable (Y) and multiple independent numeric and/or discrete variables (X's). It is used to discover the relationship between the variables and create an empirical equation of the form:

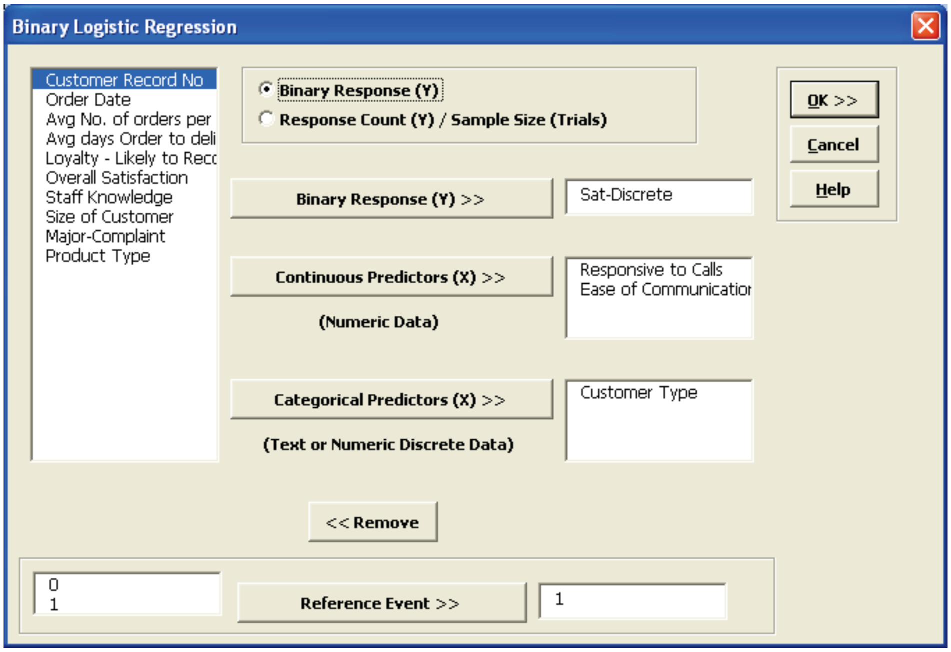
Ln(Py/(1-Py)) = b0 + b1*X1 + b2*X2 + ... + bn*Xn

This equation can be used to predict an event probability Y value for a given set of input X values. SigmaXL uses the method of maximum likelihood to solve for the model coefficients and constant term. Statistical tests of hypothesis and odds-ratios are provided for the model coefficients. The odds-ratios identify change in likelihood of the event for one-unit change in X.

An example application from medical research would be Y=Disease (Yes/No) and X's = Age, Smoker (Yes/No), Number Years of Smoking and Weight. The model coefficient P-Values would indicate which X's are significant and the odds-ratios would provide the relative change in risk for each unit change in X.

We will analyze the familiar Customer Satisfaction data using Y=Discrete Satisfaction where the values have been coded such that an Overall Satisfaction score >= 3.5 is considered a 1, and scores < 3.5 are considered a 0. Please note, we are not advising that continuous data be converted to discrete data in actual practice, but simply using the Discrete Satisfaction score for continuity with the previous analysis.

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Regression > Binary Logistic Regression**. If necessary, click **Use Entire Data Table**, click **Next**.
2. Select S*at-Discrete*, click **Binary Response (Y) >>**; select *Responsive to Calls* and *Ease of Communications*, click **Continuous Predictors (X) >>**; select *Customer Type*, click **Categorical Predictors (X) >>**.



Note that **Response Count (Y)/Sample Size (Trials)** should be used when each record contains both the number of occurrences along with associated sample size. This is common when tracking daily quality data or performing design of experiments where each run contains a response of the number of defects and sample size.

3. Click **OK**. The resulting Binary Logistic report is shown:

Binary Logistic Regression Model: ln(Py/(1-Py)) = (-16.647) + (2.454) * Responsive to Calls + (2.366) * Ease of Communications + (0.546089) *

Logit Link

Response Summary: Sat-Discrete

Value	Count	Proportion	Reference Event
0	33	0.33	
1	67	0.67	x
Total	100		

Parameter Estimates:

Term	Coefficient	SE Coefficient	Z	P	Odds Ratio	Lower 95% Odds Ratio	Upper 95% Odds Ratio
Constant	-16.647	3.837	-4.339	0.0000			
Responsive to Calls	2.454	0.558678	4.392	0.0000	11.630	3.891	34.765
Ease of Communications	2.366	0.673792	3.512	0.0004	10.660	2.846	39.929
Customer Type_2	0.546089	1.076518129	0.507273	0.6120	1.726	0.209321	14.240
Customer Type_3	-1.131	0.990522	-1.141	0.2537	0.322850	0.046328807	2.250

Wald Estimates for Categorical (Discrete) Predictors:

Term	Chi-Square	DF	P
Customer Type	2.630	2	0.2685

Model Summary and Goodness-of-Fit Statistics:

Log-Likelihood	-19.883
Test that all slope coefficients are equal to zero:	
Likelihood Ratio Chi-Square (G)	87.070
DF	4
P-Value	0.0000
McFadden's Pseudo R-Square	68.65%

Observed and Predicted Outcomes:

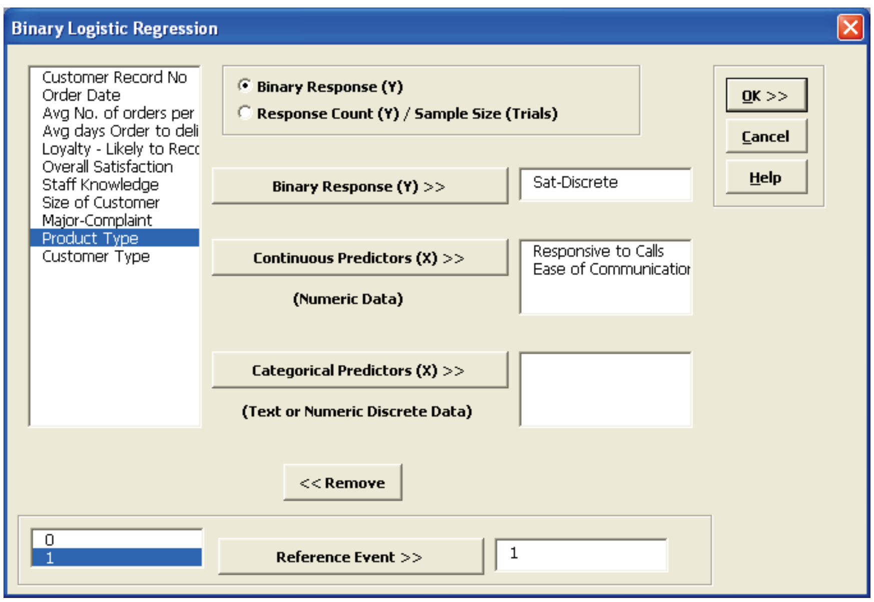
Observed Outcome	Predicted Outcome $\hat{Y}=0$	Predicted Outcome $\hat{Y}=1$	Row Total
Y = 0	29	4	33
Y = 1	3	64	67
Column Total	32	68	100
Percent Correctly Predicted:	93.00%		

4. The Likelihood Ratio P-Value < .05 tells us that the model is significant. The P-Values for the coefficients in the Parameter Estimates table confirm that Responsive to Calls and Ease of Communications are significant.

5. The P-Value in Wald Estimates for Categorical (Discrete) Predictors table tells us that Customer Type is not significant here.
   Tip: Significance for categorical predictors should be based on the Wald Estimates not the P-Values given in the Parameter Estimates table.

6. Note that Customer Type 1 is not displayed in the Parameter Estimates table. This is the "hidden" reference value for Customer Type. Categorical predictors must have one level selected as a reference value. SigmaXL sorts the levels alphanumerically and selects the first level as the reference value.

7. Now we will rerun the binary logistic regression but remove *Customer Type* as a predictor. Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Remove *Customer Type* by highlighting *Customer Type* and double-clicking (or press the **Remove** button).



8. Click **OK**. The resulting Binary Logistic report is shown:

Binary Logistic Regression Model: ln(Py/(1-Py)) = (-17.703) + (2.462) * Responsive to Calls + (2.589) * Ease of Communications
Logit Link

Response Summary: Sat-Discrete

Value	Count	Proportion	Reference Event
0	33	0.33	
1	67	0.67	x
Total	100		

Parameter Estimates:

Term	Coefficient	SE Coefficient	Z	P	Odds Ratio	Lower 95% Odds Ratio	Upper 95% Odds Ratio
Constant	-17.703	3.947	-4.486	0.0000			
Responsive to Calls	2.462	0.559880	4.397	0.0000	11.726	3.913	35.133
Ease of Communications	2.589	0.668700	3.871	0.0001	13.314	3.590	49.375

Model Summary and Goodness-of-Fit Statistics:

Log-Likelihood	-21.275
Test that all slope coefficients are equal to zero:	
Likelihood Ratio Chi-Square (G)	84.285
DF	2
P-Value	0.0000
McFadden's Pseudo R-Square	66.45%
Goodness-of-Fit Tests (P-Value < .05 indicates Lack-of-Fit):	
Pearson Residuals Chi-Square	112.63
DF	97
P-Value	0.1325
Deviance Residuals Chi-Square	42.551
DF	97
P-Value	1.0000
Hosmer-Lemeshow Chi-Square	20.187
DF	8
P-Value	0.0097
Measures of Association	
Concordant	2123
Discordant	86
Ties	2
Total	2211
Concordant Percent	96.02
Discordant Percent	3.89
Ties Percent	0.09
Goodman-Kruskal Gamma	0.922137
Somers' D	0.921303
Kendall's Tau-a	0.411515

Observed and Predicted Outcomes:

Observed Outcome	Predicted Outcome Ŷ = 0	Predicted Outcome Ŷ = 1	Row Total
Y = 0	30	3	33
Y = 1	1	66	67
Column Total	31	69	100
Percent Correctly Predicted:	96.00%		

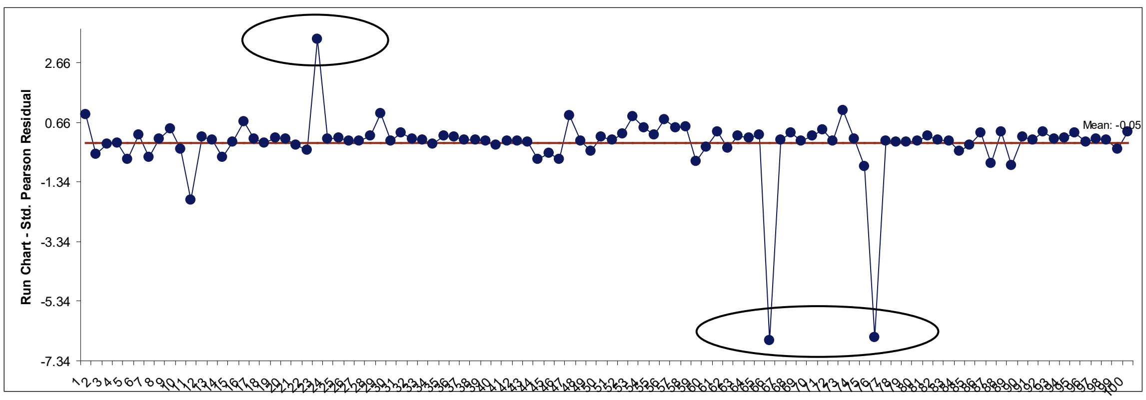
9. The Odds Ratios in the Parameter Estimates table tell us that for every unit increase in Responsive to Calls we are 11.7 times more likely to obtain a satisfied customer. For every unit increase in Ease of Communications we are 13.3 times more likely to obtain a satisfied customer.

10. McFadden's Pseudo R-Square mimics the R-square found in linear regression. This value varies between 0 and 1 but is typically much lower than the traditional R-squared value. A value less than 0.2 indicates a weak relationship; 0.2 to 0.4 indicates a moderate relationship; greater than 0.4 indicates a strong relationship. Here we have an R-square value of 0.66 indicating a strong relationship. This is also confirmed with the Percent Correctly Predicted value of 96%.

11. The Pearson, Deviance and Hosmer-Lemeshow Goodness of Fit tests are used to confirm if the binary logistic model fits the data well. P-Values < .05 for any of these tests indicate a significant lack of fit. Here the Hosmer-Lemeshow test is indicating lack of fit. Residuals analysis will help us to see where the model does not fit the data well.

12. The measures of association are used to indicate the relationship between the observed responses and the predicted probabilities. Larger values for Goodman-Kruskal Gamma, Somers' D and Kendall's Tau-a indicate that the model has better predictive ability.

13. The residuals report is given on the Sheet **Binary Logistic Residuals**. Three types of residuals are provided: Pearson, Standardized Pearson and Deviance. The Standardized Pearson Residual is most commonly used and is shown here plotted on a Run Chart. To create, click **SigmaXL > Graphical Tools > Run Chart**, check **Use Entire Data Table** to select the *Residuals* data, click **Next**, select *Std. Pearson Residual* as the **Numeric Data Variable (Y)**. Click **OK**.



Any Standardized Pearson Residual value that is less than -3 or greater than +3 is considered extreme and should be investigated. There are 3 such outliers here: rows 24, 67, and 77 in the residuals table. The +3.4 value indicates that the predicted event probability was low (.08) but the actual result was a 1. The -6.6 value indicates that the predicted event probability was high (.98) but the actual result was a 0. The large negative residuals have high Responsive to Calls and Ease of Communications but dissatisfied customers. The reasons for these discrepancies should be explored further but we will not do so here.

14. Reselect the **Binary Logistic** sheet. Scroll over to display the Event Probability calculator:

Response Event Probability:

Predictors	Enter Settings:	Predicted Event Probability
Responsive to Calls		
Ease of Communications		

This calculator provides a predicted Event Probability for given values of X (in this case the probability of a satisfied customer). Enter the values 3,3; 4,4; 5,5 as shown:

Response Event Probability:

Predictors	Enter Settings:	Predicted Event Probability
Responsive to Calls	3	0.072336996
Ease of Communications	3	

Response Event Probability:

Predictors	Enter Settings:	Predicted Event Probability
Responsive to Calls	4	0.924088278
Ease of Communications	4	

Response Event Probability:

Predictors	Enter Settings:	Predicted Event Probability
Responsive to Calls	5	0.999474064
Ease of Communications	5	

If Responsive to Calls and Ease of Communications are both equal to 3, the probability of a satisfied customer is only .07 (7%); if Responsive to Calls and Ease of Communications are both equal to 5, the probability of a satisfied customer is .9995 (99.95%)

15. Note that if the calculator includes predictors that are categorical (discrete), enter a 0 or 1 to denote the selected level as shown below (using the original analysis which included Customer Type):

Response Event Probability:

Predictors	Enter Settings:	Predicted Event Probability
Responsive to Calls	3.02	0.719225458
Ease of Communications	4.07	
Customer Type_2	1	
Customer Type_3	0	

If we wanted to select Customer Type 1, enter a 0 for both **Customer Types 2** and **3**. Customer Type 1 is the hidden reference value.

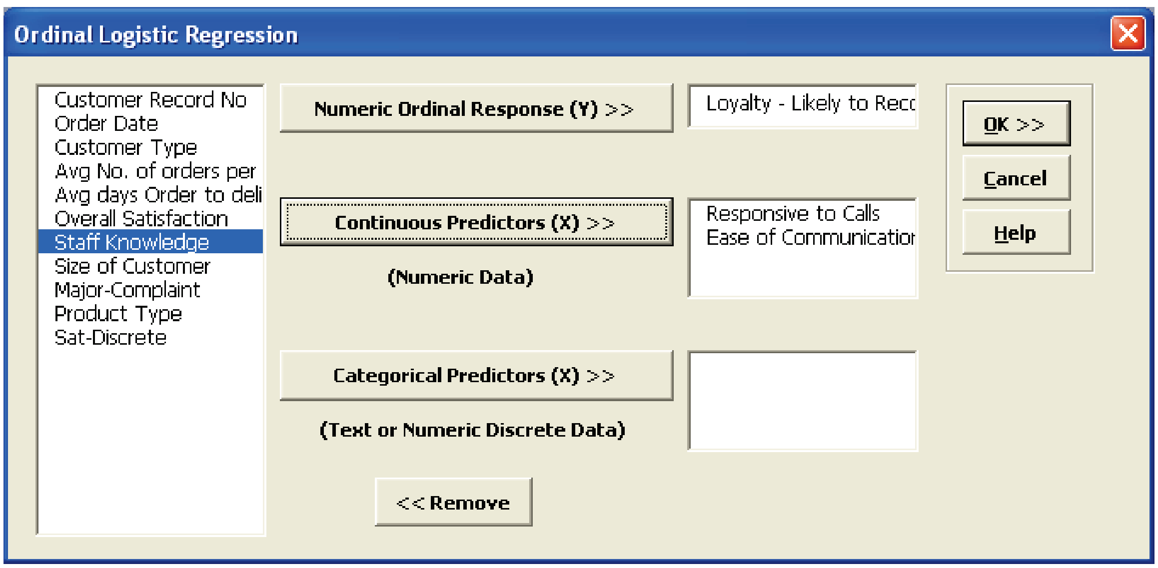
Ordinal Logistic Regression

Ordinal Logistic Regression is used to analyze the relationship between one ordinal dependent variable (Y) and multiple independent continuous and/or discrete variables (X's).

We will analyze the Customer Satisfaction data using Y=Loyalty –Likely to Recommend score which contains ordinal integer values from 1 to 5, where 5 indicates that the customer is very likely to recommend us and 1 indicates that they are very likely to not recommend us.

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab (or press **F4** to activate last worksheet). Click **SigmaXL > Statistical Tools > Regression > Ordinal Logistic Regression**. If necessary, click **Use Entire Data Table**, click **Next**.

2. Select *Loyalty – Likely to Recommend*, click **Numeric Ordinal Response (Y) >>**; select *Responsive to Calls* and *Ease of Communications*, click **Continuous Predictors (X) >>**.



3. Click **OK**. The resulting Ordinal Logistic report is shown:

Ordinal Logistic Regression Model: ln(Py/(1-Py)) = Constant + (-1.011305417) * Responsive to Calls + (-0.790025) * Ease of Communications
Logit Link

Response Summary: Loyalty - Likely to Recommend

Value	Count	Proportion
1	2	0.02
2	10	0.1
3	33	0.33
4	43	0.43
5	12	0.12
Total	100	

Parameter Estimates:

Term	Coefficient	SE Coefficient	Z	P	Odds Ratio	Lower 95% Odds Ratio	Upper 95% Odds Ratio
Constant 1	1.661	1.110374236	1.496	0.1348			
Constant 2	4.138	1.049364975	3.944	0.0001			
Constant 3	6.721	1.199	5.607	0.0000			
Constant 4	9.519	1.351	7.048	0.0000			
Responsive to Calls	-1.01130542	0.208727	-4.845	0.0000	0.363744	0.241615	0.547605
Ease of Communications	-0.790025	0.253997	-3.110	0.0019	0.453833	0.275861	0.746626

Model Summary and Goodness-of-Fit Statistics:

Log-Likelihood	-105.24
Test that all slope coefficients are equal to	
Likelihood Ratio Chi-Square (G)	47.850
DF	2
P-Value	0.0000
McFadden's Pseudo R-Square	0.185221
Goodness-of-Fit Tests	
Pearson Residuals Chi-Square	279.41
DF	394
P-Value	1.0000
Deviance Residuals Chi-Square	210.49
DF	394
P-Value	1.0000
Measures of Association	
Concordant	2702
Discordant	694
Ties	11
Total	3407
Concordant Percent	79.307
Discordant Percent	20.370
Ties Percent	0.322865
Goodman-Kruskal Gamma	0.591284
Somers' D	0.589375
Kendall's Tau-a	0.405657

Observed and Predicted Outcomes:

Observed Outcome	Predicted Outcomes					
	Y = 1	Y = 2	Y = 3	Y = 4	Y = 5	Row Total
Y = 1	0	2	0	0	0	2
Y = 2	0	2	6	2	0	10
Y = 3	0	2	14	17	0	33
Y = 4	0	1	10	32	0	43
Y = 5	0	0	0	12	0	12
Column Total	0	7	30	63	0	100
Percent Correctly Predicted:	48.00%					

4. The Likelihood Ratio P-Value < .05 tells us that the model is significant. The low P-Values for the coefficients confirm that Responsive to Calls and Ease of Communications are significant.

5. The Odds Ratios tell us that for every one-unit increase in Responsive to Calls the chance of a Loyalty score of 1 versus 2 (or 2 versus 3, etc.) is reduced by a multiple of 0.36. This is not very intuitive but will be easy to see when we use the Response Outcome Probability calculator.

6. McFadden's Pseudo R-Square value is 0.185 indicating that this is a weak (but close to moderate) degree of association. This is also confirmed with the Percent Correctly Predicted value of 48%.

7. The Pearson and Deviance Goodness of Fit (GOF) tests are used to confirm if the ordinal logistic model fits the data well. P-Values < .05 would indicate a significant lack of fit. Given that the GOF P-Values are greater than .05, we conclude that there is no significant lack of fit.

8. Scroll across to the Response Outcome Probability calculator. This calculator provides predicted outcome (event) probabilities for given values of X (in this case the probability of a satisfied customer). Enter the values 3,3; 4,4; 5,5 as shown:

Response Outcome Probability:

Predictors	Enter Settings:	Outcome	Predicted Cumulative Probability	Predicted Probability for each Level
Responsive to Calls	3	1	0.023126734	0.023126734
Ease of Communications	3	2	0.2199757	0.196848965
		3	0.788637058	0.568661358
		4	0.983934474	0.195297416
		5	1	0.016065526

Response Outcome Probability:

Predictors	Enter Settings:	Outcome	Predicted Cumulative Probability	Predicted Probability for each Level
Responsive to Calls	4	1	0.003892909	0.003892909
Ease of Communications	4	2	0.044483294	0.040590386
		3	0.381166233	0.336682938
		4	0.90999329	0.528827057
		5	1	0.09000671

Response Outcome Probability:

Predictors	Enter Settings:	Outcome	Predicted Cumulative Probability	Predicted Probability for each Level
Responsive to Calls	5	1	0.000644733	0.000644733
Ease of Communications	5	2	0.007626511	0.006981777
		3	0.092294817	0.084668306
		4	0.625327323	0.533032506
		5	1	0.374672677

9. Referring to the Predicted Probability for each Level, if **Responsive to Calls** and **Ease of Communications** are both equal to 3, we would expect to see typical loyalty scores of 3 (57%) with some at 2 (20%) and 4 (20%); if **Responsive to Calls** and **Ease of Communications** are both equal to 4, we would expect typical loyalty scores of 4 (53 %) with some at 3 (34%) and 5 (9%); if **Responsive to Calls** and **Ease of Communications** are both equal to 5, we would expect typical loyalty scores of 4 (53 %) with some at 3 (8.5%) and 5 (38%).

SigmaXL: Improve Phase Tools: Design of Experiments (DOE)

Part A – Overview of Basic Design of Experiments (DOE) Templates

The DOE templates are similar to the other SigmaXL templates: simply enter the inputs and resulting outputs are produced immediately. The DOE templates provide common 2-level designs for 2 to 5 factors. These basic templates are ideal for training, but use **SigmaXL > Design of Experiments > 2-Level Factorial/Screening Designs** to accommodate up to 19 factors with randomization, replication and blocking.

Click **SigmaXL > Design of Experiments > Basic DOE Templates** to access these templates:

- Two-Factor, 4-Run, Full-Factorial
- Three-Factor, 4-Run, Half-Fraction, Res III
- Three-Factor, 8-Run, Full-Factorial
- Four-Factor, 8-Run, Half-Fraction, Res IV
- Four-Factor, 16-Run, Full-Factorial
- Five-Factor, 8-Run, Quarter-Fraction, Res III
- Five-Factor, 16-Run, Half-Fraction, Res V

After entering the template data, main effects and interaction plots may be created by clicking **SigmaXL > Basic DOE Templates > Main Effects & Interaction Plots**. The DOE template must be the active worksheet.

DOE Templates are protected worksheets by default, but this may be modified by clicking **SigmaXL > Help > Unprotect Worksheet**.

Advanced analysis is available, but this requires that you unprotect the DOE worksheet. The following example shows how to use Excel's Equation Solver and SigmaXL's Multiple Regression in conjunction with a DOE template. **Caution:** If you unprotect the worksheet, do not change the worksheet title (e.g. **Three-Factor, Two-Level, 8-Run, Full-Factorial Design of Experiments).** This title is used by the Main Effects & Interaction Plots to determine appropriate analysis. Also, do not modify any cells with formulas.

Part B – Three Factor Full Factorial Example Using DOE Template

1. Open the file **DOE Example - Robust Cake.xlsx**. This is a Robust Cake Experiment adapted from the Video "Designing Industrial Experiments" by Box, Bisgaard and Fung.

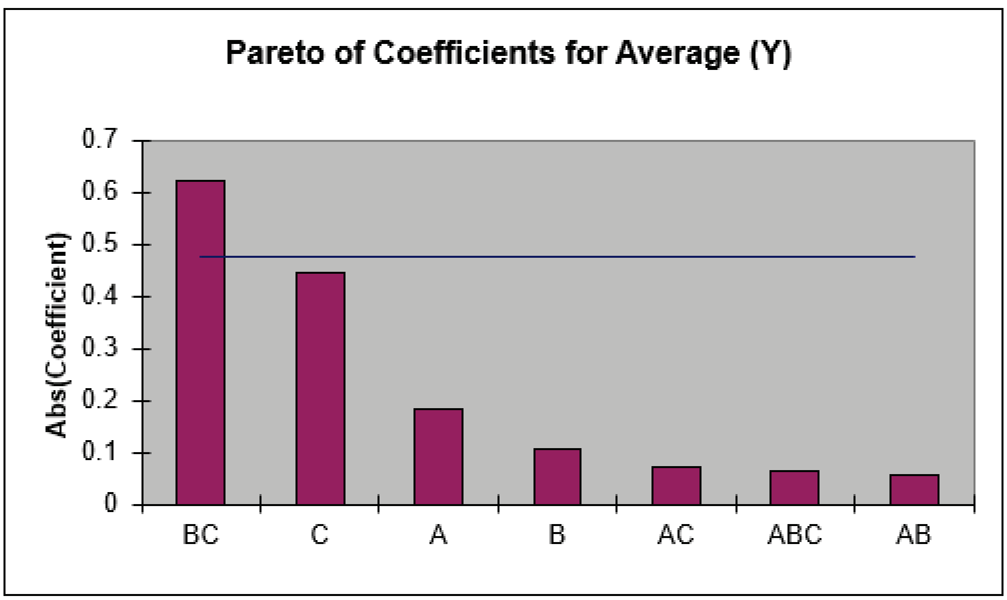
2. The response is Taste Score (on a scale of 1-7 where 1 is "awful" and 7 is "delicious").

3. The five Outer Array Reps have different Cooking Time and Temperature Conditions. The outer array was a two-factor, full-factorial plus center point, hence 5 replications.

4. The goal is to Maximize Mean and Minimize StDev of the Taste Score.

5. The X factors are Flour, Butter, and Egg. Actual low and high settings are not given in the video, so we will use coded -1 and +1 values. We are looking for a combination of Flour, Butter, and Egg that will not only taste good, but consistently taste good over a wide range of Cooking Time and Temperature conditions.

6. Scroll down to view the Pareto of Abs. Coefficients for Average (Y).



7. The BC (Butter * Egg) interaction is clearly the dominant factor. The bars above the 95% confidence blue line indicate the factors that are statistically significant; in this case only BC is significant. Keep in mind that this is an initial analysis. Later, we will show how to do a more powerful Multiple Regression analysis on this data. (Also, the Rule of Hierarchy states that if an interaction is significant, we must include the main effects in the model used.)

8. The significant BC interaction is also highlighted in red in the table of Effects and Coefficients:

Calculation of Effects and Coefficients for Average (Y):

	Constant	A	B	C	AB	AC	BC	ABC
Avg(Avg(Y)) @ +1:		4.865	4.79	5.125	4.74	4.755	4.06	4.62
Avg(Avg(Y)) @ -1:		4.5	4.575	4.24	4.625	4.61	5.305	4.745
Effect (Delta):		0.365	0.215	0.885	0.115	0.145	-1.245	-0.125
Coefficient (Delta/2):	4.6825	0.1825	0.1075	0.4425	0.0575	0.0725	-0.6225	-0.0625
SE Coefficient:	0.2325	0.2325	0.2325	0.2325	0.2325	0.2325	0.2325	0.2325
T-value	20.1397849	0.7849	0.4624	1.9032	0.24731	0.31183	-2.6774	-0.2688
P-value	9.6928E-20	0.4383	0.6469	0.066	0.80625	0.75719	0.01161	0.7898

9. The R-Square value is given as 27%. This is very poor for a Designed Experiment. Typically, we would like to see a minimum of 50%, with > 80% desirable.

R-Square:	27.03%
R-Square Adj.:	11.06%
S	1.4705

The reason for the poor R-square value is the wide range of values over the Cooking Temperature and Time conditions. In a robust experiment like this, it is more appropriate to analyze the mean response as an individual value rather than as five replicate values. The Standard Deviation as a separate response will also be of interest.

10. If the Responses are replicated, SigmaXL draws the blue line on the Pareto Chart using an estimate of experimental error from the replicates. If there are no replicates, an estimate called Lenth's Pseudo Standard Error is used (see Lenth, R.V. (1989). "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, Vol 31, pp. 469-473).

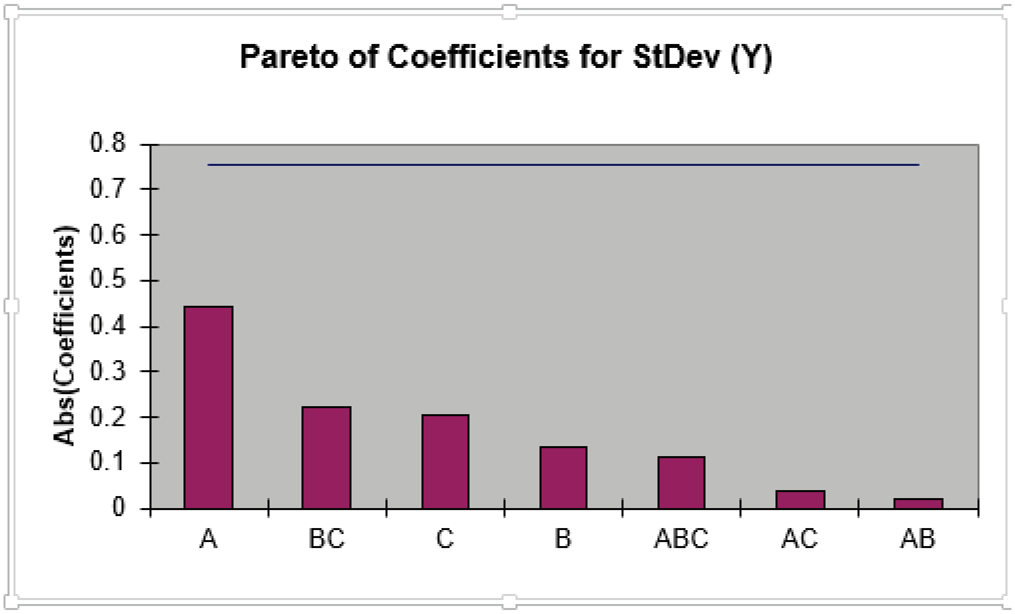
11. If the 95% Confidence line for coefficients were to be drawn using Lenth's method, the value would be 0.409 as given in the table:

Lenth's Pseudo Standard Error (PSE) Analysis for Unreplicated Data:

Lenth's PSE for Coefficients:	0.10875
Lenth's Margin of Error for Coefficients (95% Conf. Level):	0.40935
Lenth's Margin of Error for Effects (95% Conf. Level):	0.8187

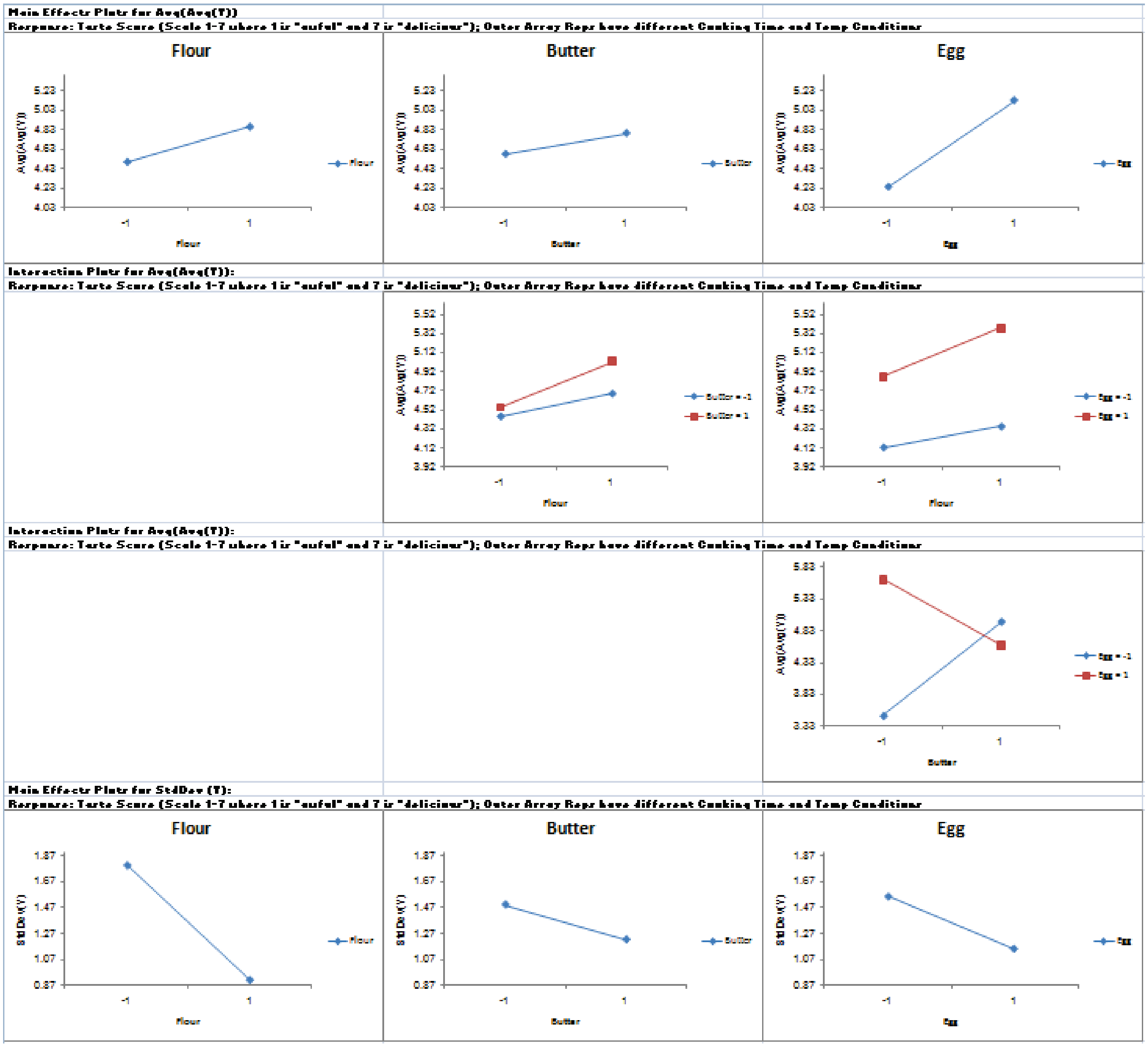
This would show factor C as significant.

12. Scroll down to view the Pareto of Coefficients for StDev(Y).



13. The A (Flour) main effect is clearly the dominant factor, but it does not initially appear to be statistically significant (based on Lenth's method). Later, we will show how to do a more powerful Regression analysis on this data.

14. The Pareto chart is a powerful tool to display the relative importance of the main effects and interactions, but it does not tell us about the direction of influence. To see this, we must look at the main effects and interaction plots. Click **SigmaXL > Design of Experiments > Basic DOE Templates > Main Effects & Interaction Plots**. The resulting plots are shown below:



15. The Butter*Egg two-factor interaction is very prominent here. Looking at only the Main Effects plots would lead us to conclude that the optimum settings to maximize the average taste score would be Butter = +1, and Egg = +1, but the interaction plot tells a very different story. The correct optimum settings to maximize the taste score is Butter = -1 and Egg = +1.

16. Since Flour was the most prominent factor in the Standard Deviation Pareto, looking at the Main Effects plots for StDev, we would set Flour = +1 to minimize the variability in taste scores.

17. Click on the Sheet **Three-Factor 8-Run DOE**. At the **Predicted Output for Y**, enter *Flour* = 1, *Butter* = -1, *Egg* = 1 as shown:

Factor	Factor Name	Low	High
A	Flour	-1	1
B	Butter	-1	1
C	Egg	-1	1

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - uncoded	Factor setting coded	Y-hat:	S-hat:
	1	1	5.9	0.68191
	-1	-1		
	1	1		

The predicted average (Y-hat) taste score is 5.9 with a predicted standard deviation (S-hat) of 0.68. Note that this prediction equation includes all main effects, two-way interaction, and the three-way interaction.

Multiple Regression and Excel Solver (Advanced Topics):

18. In order to run Multiple Regression analysis, we will need to unprotect the worksheet. Click **SigmaXL > Help > Unprotect Worksheet**.

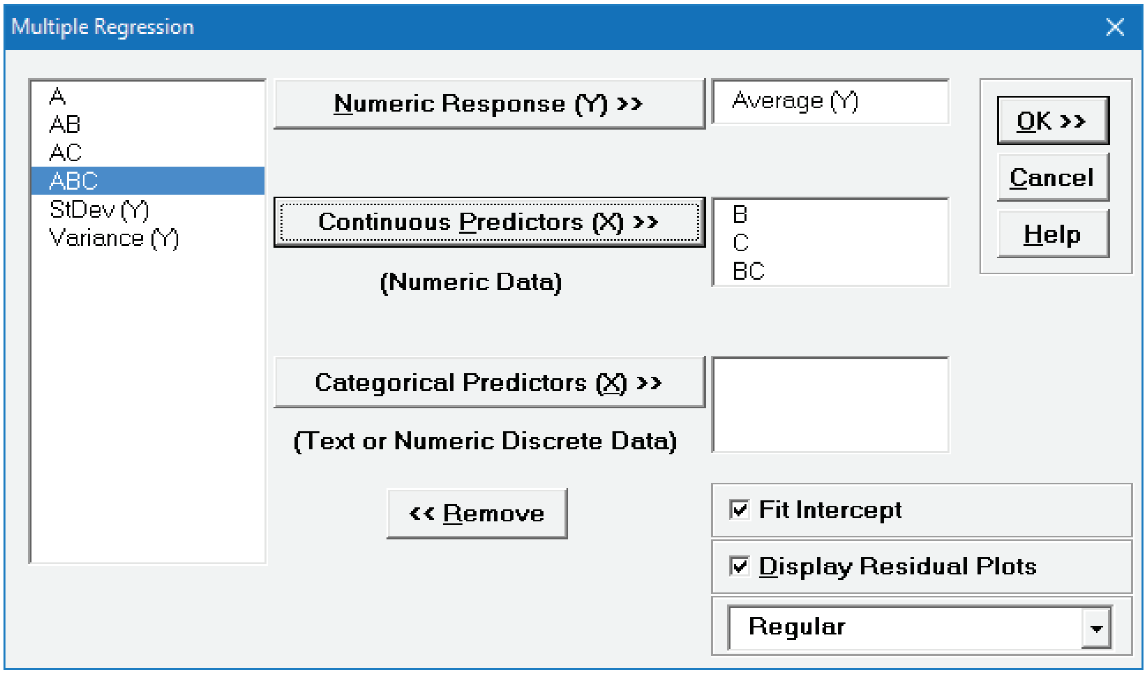
19. In the **Coded Design Matrix**, highlight columns A to ABC, and the calculated responses as shown:

Coded Design Matrix:

Standard Run Order	Actual Run Order	A	B	C	AB	AC	BC	ABC	Average (Y)	StDev (Y)	Variance (Y)
1	6	-1	-1	-1	1	1	1	-1	3.52	2.4499	6.002
2	2	1	-1	-1	-1	-1	1	1	3.5	1.38022	1.905
3	7	-1	1	-1	-1	1	-1	1	4.74	1.47919	2.188
4	1	1	1	-1	1	-1	-1	-1	5.2	0.93541	0.875
5	5	-1	-1	1	1	-1	-1	1	5.38	1.45499	2.117
6	8	1	-1	1	-1	1	-1	-1	5.9	0.68191	0.465
7	3	-1	1	1	-1	-1	1	-1	4.36	1.81742	3.303
8	4	1	1	1	1	1	1	1	4.86	0.66558	0.443

20. Click **SigmaXL > Statistical Tools > Regression > Multiple Regression**. Click **Next**.

21. Select *Average (Y)*, click **Numeric Response (Y) >>**; holding the CTRL key, select *B*, *C*, and *BC*; click **Continuous Predictors (X)** >> as shown:



22. Click **OK**. The resulting regression report is shown:

Multiple Regression Model: Average (Y) = (4.683) + (0.1075) * B + (0.442500) * C + (-0.622500) * BC

Model Summary:

R-Square	92.85%
R-Square Adjusted	87.50%
S (Root Mean Square Error)	0.302572

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	4.683	0.106975464	43.772	0.0000		
B	0.1075	0.106975464	1.005	0.3718	1	1.00000
C	0.442500	0.106975464	4.136	0.0144	1	1.00000
BC	-0.622500	0.106975464	-5.819	0.0043	1	1.00000

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	3	4.759	1.586	17.327	0.0093
Error	4	0.366200	0.09155		
Pure Error	4	0.366200	0.09155		
Total (Model + Error)	7	5.125	0.732164		

Durbin-Watson Test for Autocorrelation in Residuals:

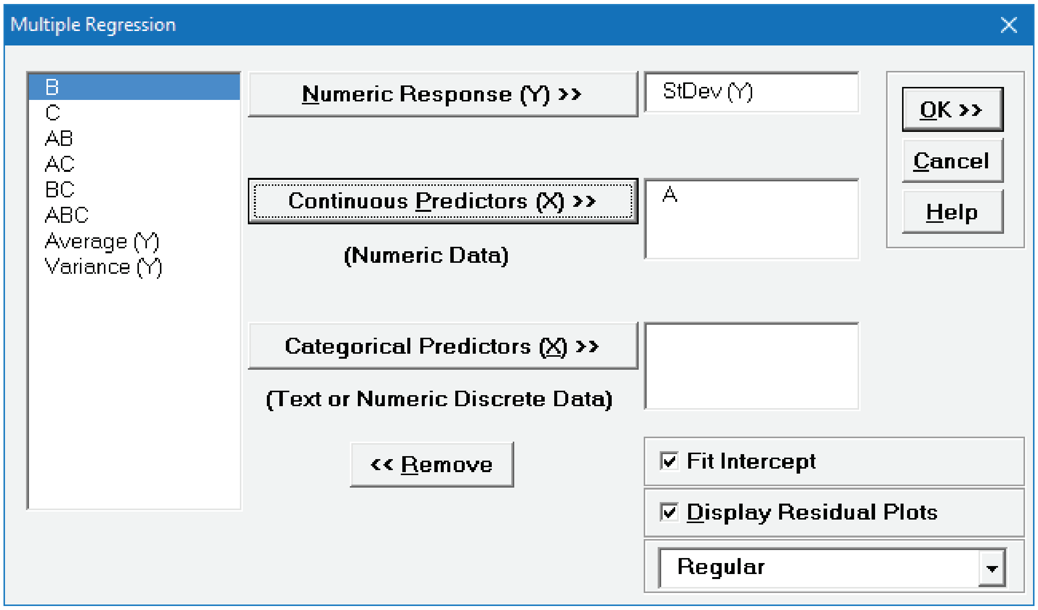
DW Statistic	3.498
P-Value Positive Autocorrelation	0.9670
P-Value Negative Autocorrelation	0.0000

Note that the R-square value of 92.85% is much higher than the earlier result of 27%. This is due to our modeling the mean response value rather than considering all data in the outer array. Note also that the C main effect now appears as significant.

23. Click on the Sheet **Three-Factor 8-Run DOE**.

24. With the **Coded Design Matrix** highlighted as before, click **SigmaXL > Statistical Tools > Regression > Multiple Regression**. Click **Next**.

25. Select *StDev (Y)*, click **Numeric Response (Y) >>**; select *A*, click **Continuous Predictors (X) >>** as shown:



26. Click **OK**. The resulting regression report is shown:

Multiple Regression Model: StDev (Y) = (1.358) + (-0.442296) * A

Model Summary:

R-Square	61.54%
R-Square Adjusted	55.13%
S (Root Mean Square Error)	0.403733

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	1.358	0.142741	9.514	0.0001		
A	-0.442296	0.142741	-3.099	0.0212	1	1.00000

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	1	1.565	1.565	9.601	0.0212
Error	6	0.978003	0.163000		
Pure Error	6	0.978003	0.163000		
Total (Model + Error)	7	2.543	0.363287		

Durbin-Watson Test for Autocorrelation in Residuals:

DW Statistic	1.071225365
P-Value Positive Autocorrelation	0.1278
P-Value Negative Autocorrelation	0.9821

Note that Factor A (Flour) now shows as a statistically significant factor affecting the Standard Deviation of Taste Score.

27. Now we will use Excel's Equation Solver to verify the optimum settings determined using the Main Effects and Interaction Plots.

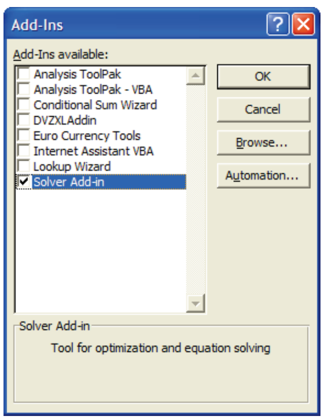
28. Click on the Sheet **Three-Factor 8-Run DOE**. At the **Predicted Output for Y**, enter 1 for *Flour*. We are setting this as a constraint, because Flour = +1 minimizes the Standard Deviation. Reset the *Butter* and *Egg* to 0 as shown:

Factor	Factor Name	Low	High
A	Flour	-1	1
B	Butter	-1	1
C	Egg	-1	1

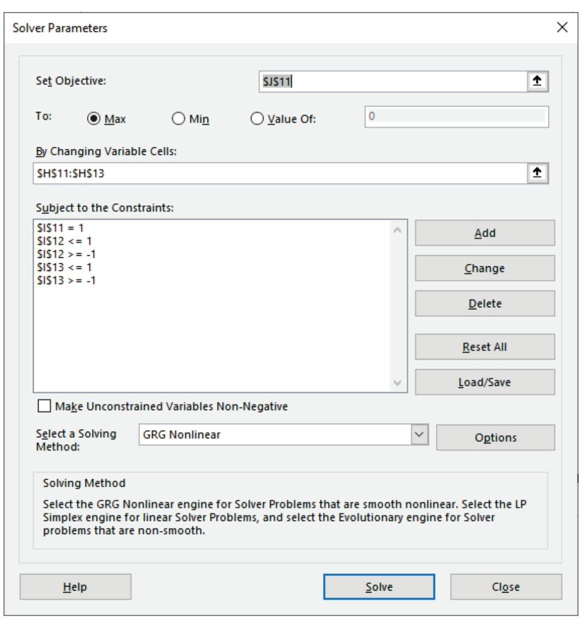
Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - uncoded	Factor setting coded	Y-hat:	S-hat:
	1	1	4.865	0.91578
	0	0		
	0	0		

29. In Excel, click **File > Options > Add-Ins > Manage Excel Add-Ins**, click **Go.** Ensure that the **Solver Add-in** is checked. If the Solver Add-in does not appear in the Add-ins available list, you will need to re-install Excel to include all add-ins.



30. Click **OK**. Click **Data > Solver**. Set the **Solver Parameters** as shown:



Cell J11 is the Y-hat, predicted average taste score. Solver will try to maximize this value. Cells H11 to H13 are the Actual Factor Settings to be changed. Cells I11 to I13 are the Coded Factor settings where the following constraints are given: I11=1; I12 >= -1; I12 <= 1; I13 >= -1; I13 <=1.

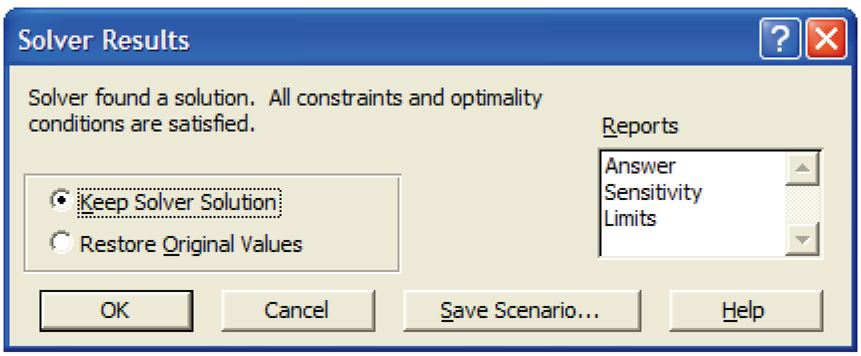
31. Click **Solve**. The solver results are given in the **Predicted Output for Y** as *Butter* = -1 and *Egg* = 1.

Factor	Factor Name	Low	High
A	Flour	-1	1
B	Butter	-1	1
C	Egg	-1	1

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - uncoded	Factor setting coded	Y-hat:	S-hat:
	1	1	5.9	0.68191
	-1	-1		
	1	1		

32. Solver indicates that a solution is found:



33. Click **OK** to keep the solution.

Tip: Advanced Multiple Regression with Multiple Response Optimization may also be used to analyze this data and replicate the above results. See **Multiple Response Optimization Example: Robust Cake**.

Part C – Design and Analysis of Catapult Full Factorial Experiment

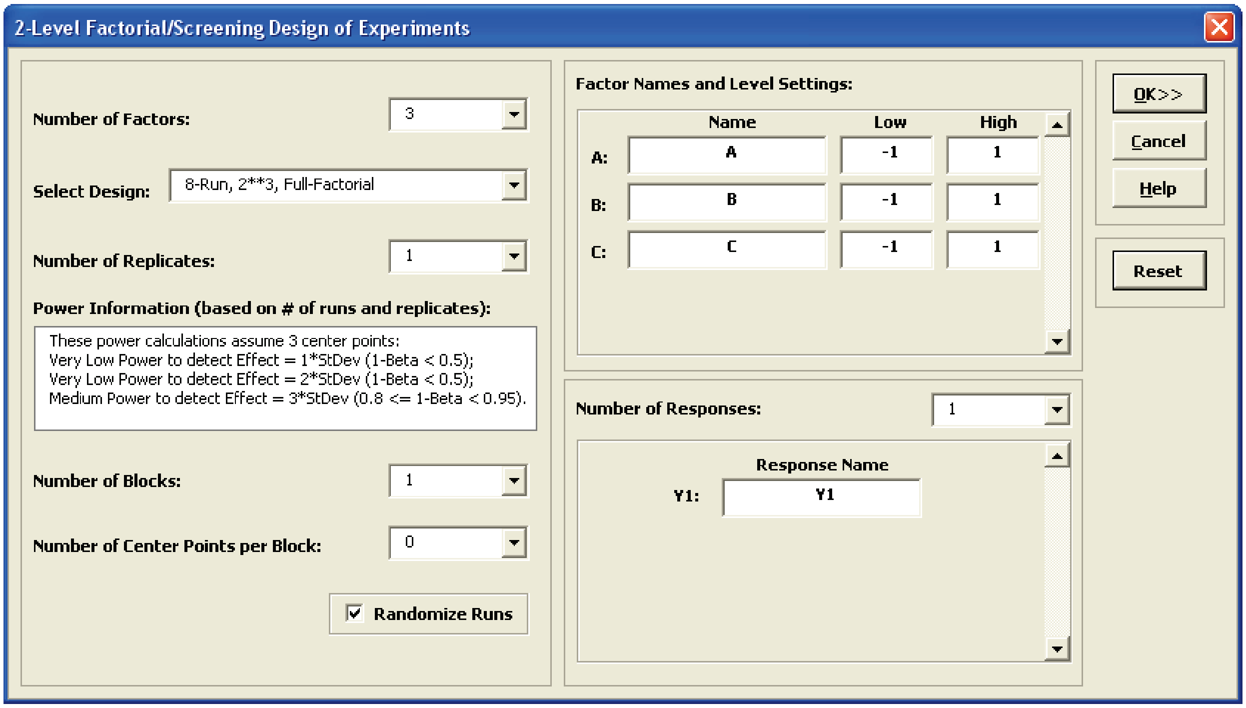
Catapults are frequently used in Six-Sigma or Design of Experiments training. They are a powerful teaching tool and make the learning fun. If you have access to a catapult, we recommend that you perform the actual experiment and use your own data. Of course, you can also follow along using the data provided. The response variable (Y) is distance, with the goal being to consistently hit a target of 100 inches.

1. Click **SigmaXL > Design of Experiments > 2-Level Factorial/Screening > 2-Level Factorial/Screening Designs**.

2. The Number of X Factors can be 2 to 19. Using process knowledge, we will limit ourselves to 3 factors: *Pull Back Angle*, *Stop Pin* and *Pin Height*. Pull Back will be varied from 160 to 180 degrees, Stop Pin will be positions 2 and 3 (count from the back), and Pin Height will be positions 2 and 3 (count from the bottom).

3. Select **Number of Factors** = 3.

4. The available designs are then given as: **4-Run, 2**(3-1), 1/2 Fraction, Res III** and **8-Run, 2**3, Full-Factorial**. If we had more than 5 factors, a Resolution III or Plackett-Burman Screening design would typically be used. Here we will choose the **8-Run, 2**3, Full-Factorial** design.



Notes: Design Generators and Aliasing of Effects will be reported for Fractional Factorial designs. When the Number of Factors is 9 or higher, Factor Name "I" is not used, in order to avoid confusion with the Fractional Factorial Defining Relation "I".

The Power Information is presented to assist the user with selection of number of runs and replicates for the design, so that one can see the trade-off between experimental cost and sensitivity to detect Effects of interest.

Power (1-Beta) < 0.5 is considered as *Very Low Power* to detect Effect.
Power (1-Beta) >= 0.5 and < 0.8 is considered as *Low Power* to detect Effect.
Power (1-Beta) >= 0.8 and < 0.95 is considered as *Medium Power* to detect Effect.
Power (1-Beta) >= 0.95 and < 0.99 is considered as *High Power* to detect Effect.
Power (1-Beta) >= 0.99 is considered as *Very High Power* to detect Effect.

The power calculations require an estimate of experimental error, so when Replicates = 1 an assumption of 3 center points is used in order to give an approximate reference level of power (regardless of the value for Number of Center Points per Block).

5. This design currently shows the following:

   These power calculations assume 3 center points:
   Very Low Power to detect Effect = 1*StDev (1-Beta < 0.5);
   Very Low Power to detect Effect = 2*StDev (1-Beta < 0.5);
   Medium Power to detect Effect = 3*StDev (0.8 <= 1-Beta < 0.95).

   We would like to have medium power to detect an Effect = 2*StDev. Change the **Number of Replicates** to 2. The Power Information is now:

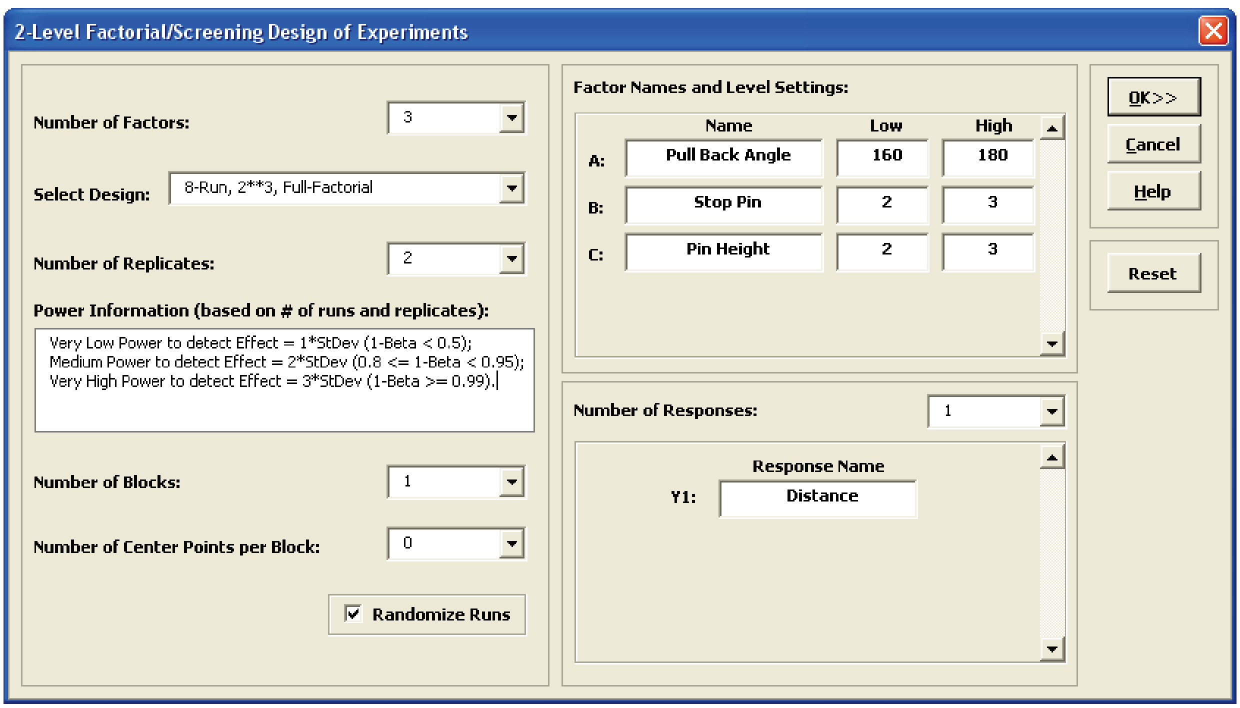
Very Low Power to detect Effect = 1*StDev (1-Beta < 0.5);
Medium Power to detect Effect = 2*StDev (0.8 <= 1-Beta < 0.95);
Very High Power to detect Effect = 3*StDev (1-Beta >= 0.99).

We will therefore choose two replicates. The number of replicates will always be a tradeoff between the desired power and the cost of the experimental runs.

Specify 2 or more blocks if there are constraints such as the number of runs per day or some other known external “nuisance” variable (like 2 different catapults or 2 operators). Here we will keep **Blocks** = 1 (i.e., no Blocking).

Center Points are useful to provide an estimate of experimental error with unreplicated designs, and allow detection of curvature. Typically, 3 to 5 center points are used. Here we will not use center points because we have replicated the design twice and do not expect significant curvature in the distance response. Furthermore, center points could not be set for Pin Height and Stop Pin (without drilling additional holes!).

6. Complete the **Factor Names**, **Level Settings** and **Response Name** as shown:



7. Click **OK**. The following worksheet is produced:

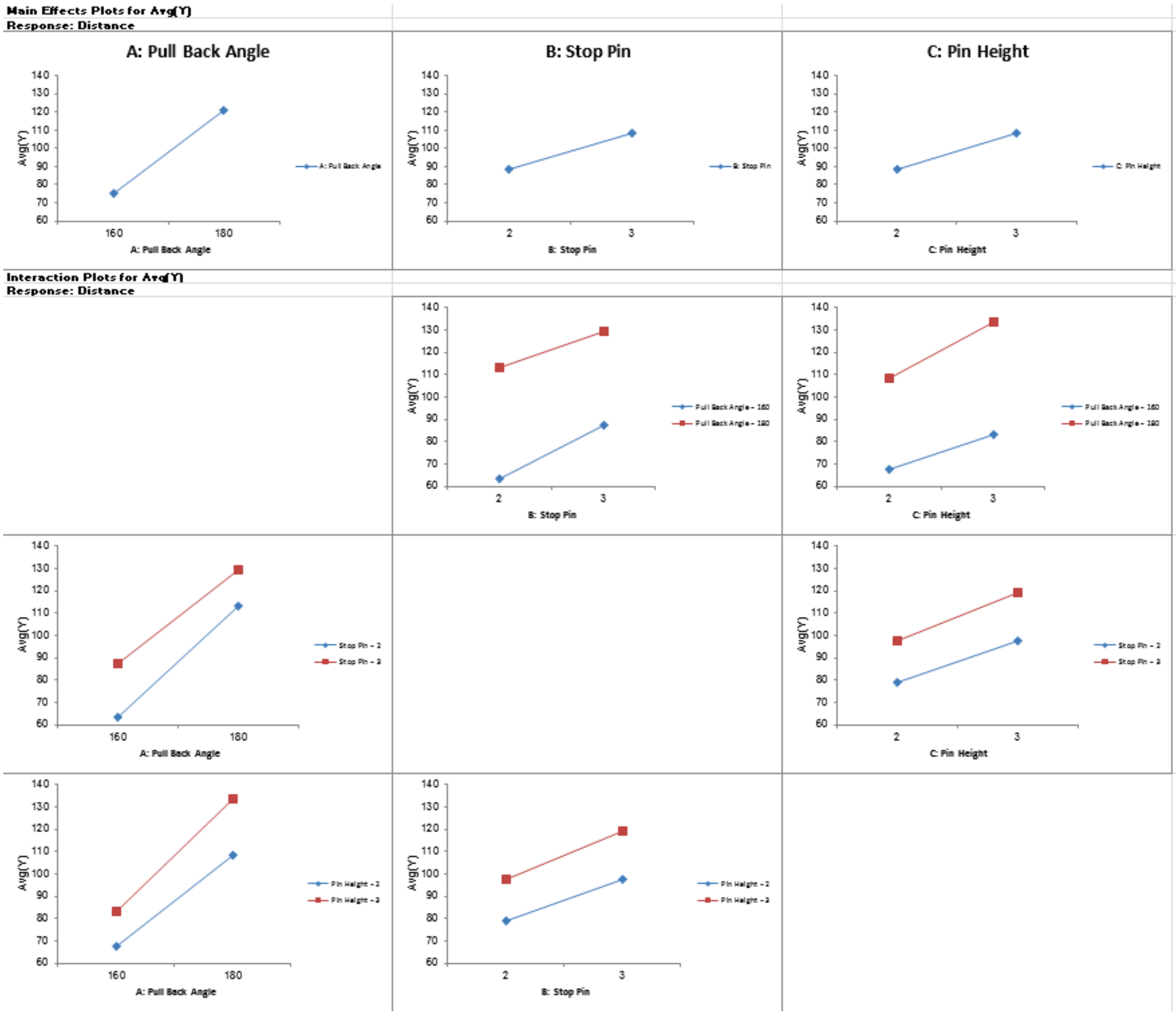
Design Type:	3 Factor, 8-Run, 2**3, Full-Factorial
Number of Replicates:	2
Number of Blocks:	1
Number of Center Points per Block:	0
Number of Responses:	1

Run Order	Std. Order	Center Points	Blocks	A: Pull Back Angle	B: Stop Pin	C: Pin Height	Distance
1	14	1	1	180	2	3	
2	4	1	1	180	3	2	
3	12	1	1	180	3	2	
4	9	1	1	160	2	2	
5	1	1	1	160	2	2	
6	5	1	1	160	2	3	
7	2	1	1	180	2	2	
8	15	1	1	160	3	3	
9	10	1	1	180	2	2	
10	7	1	1	160	3	3	
11	11	1	1	160	3	2	
12	13	1	1	160	2	3	
13	16	1	1	180	3	3	
14	6	1	1	180	2	3	
15	8	1	1	180	3	3	
16	3	1	1	160	3	2	

8. You can enter information about the experiment in the fields provided. If you have access to a catapult, perform the experimental runs in the given randomized sequence, and enter the distance values in the Distance column.

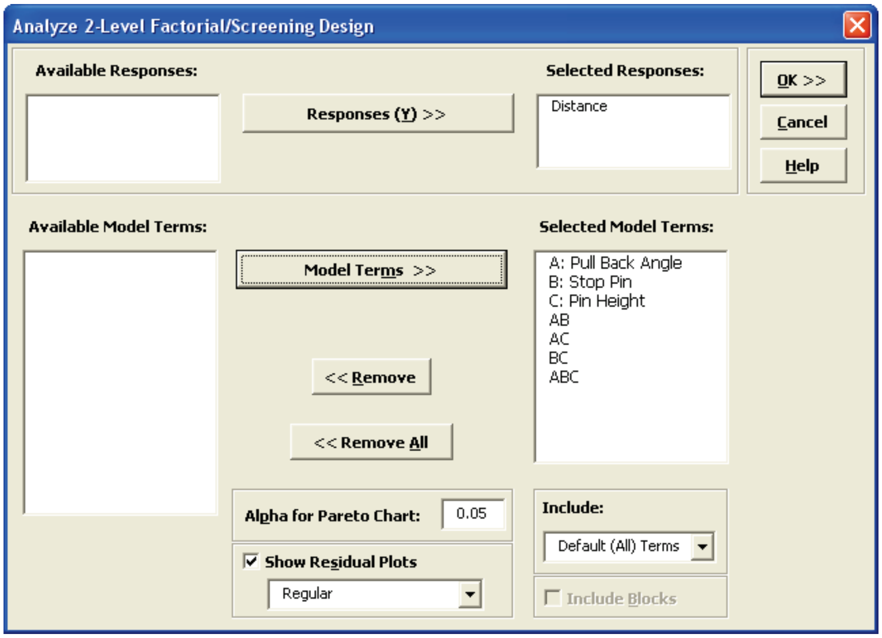
9. If you are not able to perform the catapult experiment, open the file **Catapult DOE V6.xlsx**.

10. Before we begin the regression analysis, we will have a quick look at the Main Effects and Interaction Plots. Click **SigmaXL > Design of Experiments > 2 Level Factorial/Screening > Main Effects & Interaction Plots**. The resulting plots are shown below:



Pull Back Angle is the dominant factor having the steepest slope. We can also see that the interaction terms are weak with the almost parallel lines.

11. Click **SigmaXL > Design of Experiments > 2-Level Factorial/Screening > Analyze 2-Level Factorial/Screening Design**.



12. We will use the default analyze settings (all terms in the model) to start. Click **OK**. The resulting Analysis report is shown:

Design of Experiments Analysis

DOE Multiple Regression Model: Distance = (98.3125) + (22.8125) * A: Pull Back Angle + (10.0625) * B: Stop Pin + (10.1875) * C: Pin

Title:	Catapult Example (SigmaXL Version 6)
Date:	
Name of Experimenter:	
Notes:	

Design Type:	3 Factor, 8-Run, 2**3, Full-Factorial
Number of Replicates:	2
Number of Blocks:	1
Number of Center Points per Block:	0
Response:	Distance

Model Summary:

R-Square	99.93%
R-Square Adjusted	99.86%
S (Root Mean Square Error)	1.030776406

Parameter Estimates (Coded Units):

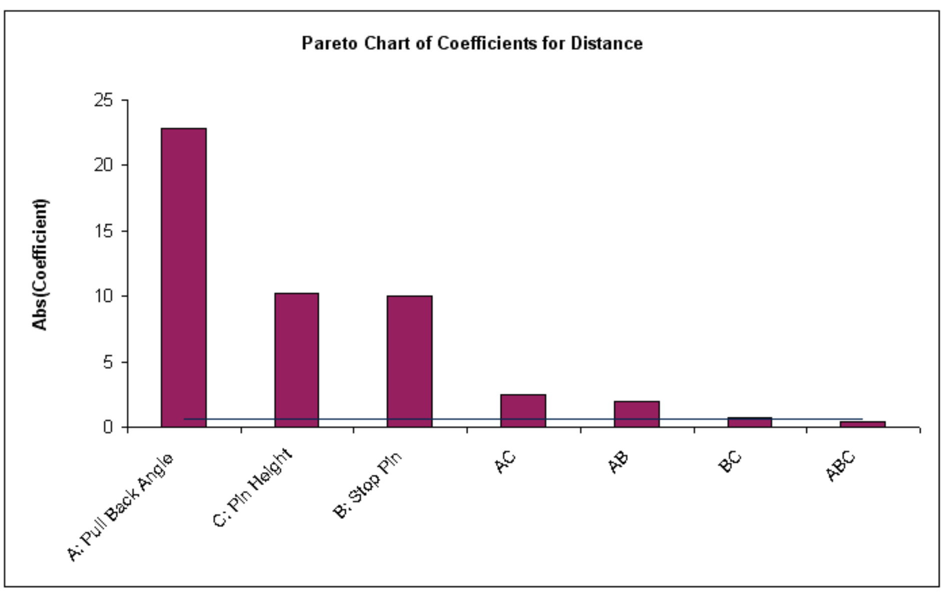
Term	**Coefficient**	**SE Coefficient**	**T**	**P**	**VIF**	**Tolerance**
Constant	98.3125	0.257694102	381.51	**0.0000**		
A: Pull Back Angle	22.8125	0.257694102	88.526	**0.0000**	1	1
B: Stop Pin	10.0625	0.257694102	39.048	**0.0000**	1	1
C: Pin Height	10.1875	0.257694102	39.533	**0.0000**	1	1
AB	-1.9375	0.257694102	-7.519	**0.0001**	1	1
AC	2.4375	0.257694102	9.459	**0.0000**	1	1
BC	0.6875	0.257694102	2.668	**0.0285**	1	1
ABC	0.4375	0.257694102	1.698	0.1280	1	1

Analysis of Variance for Model:

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	7	11773	1681.8	1582.9	**0.0000**
Error	8	8.500	1.0625		
Pure Error	8	8.500	1.0625		
Total (Model + Error)	15	11781	785.43		

Durbin-Watson Test for Autocorrelation in Residuals:

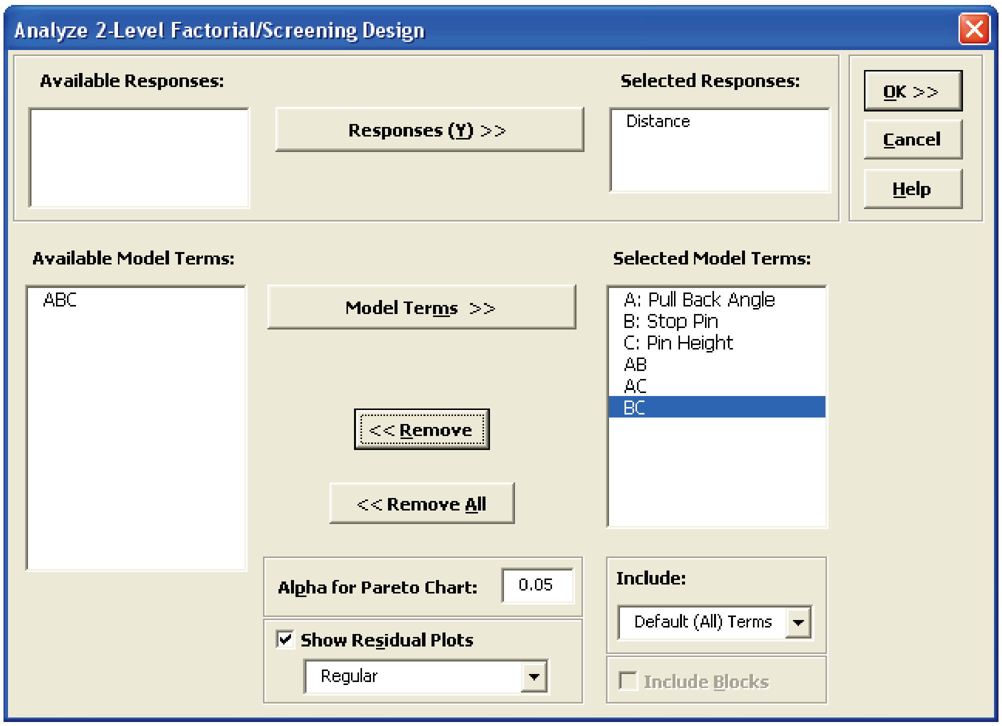
DW Statistic	2.176
P-Value Positive Autocorrelation	0.5491
P-Value Negative Autocorrelation	0.2336



13. The model looks very good with an R-Square value of 99.9%! The standard deviation (experimental error) is only 1.03 inches. Clearly *Pull Back Angle* is the most important predictor (X factor), but all the main effects and two-way interaction are significant. However, the three-way interaction is not significant, so it should be removed from the model.

14. Click **Recall Last Dialog** (or press **F3**).

15. Remove the *ABC* interaction term as shown:



16. Click **OK**. The revised report is shown below:

Design of Experiments Analysis

DOE Multiple Regression Model: Distance = (98.3125) + (22.8125) * A: Pull Back Angle + (10.0625) * B: Stop Pin + (10.1875) * C: Pin

Title:	Catapult Example (SigmaXL Version 6)
Date:	
Name of Experimenter:	
Notes:	

Design Type:	3 Factor, 8-Run, 2**3, Full-Factorial
Number of Replicates:	2
Number of Blocks:	1
Number of Center Points per Block:	0
Response:	Distance

Model Summary:

R-Square	99.90%
R-Square Adjusted	99.84%
S (Root Mean Square Error)	1.133

Parameter Estimates (Coded Units):

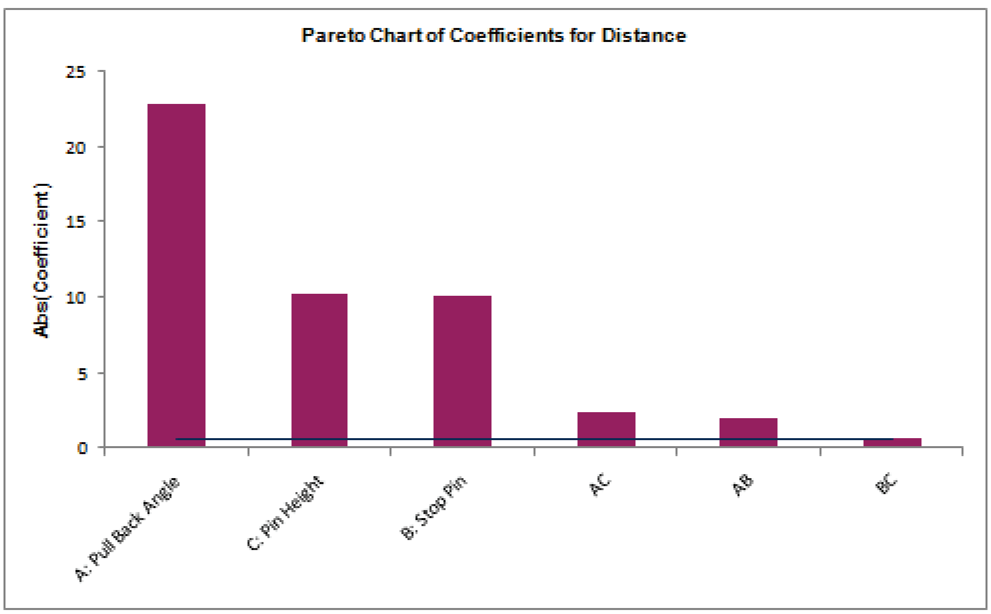
Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	98.3125	0.283363969	346.95	0.0000		
A: Pull Back Angle	22.8125	0.283363969	80.506	0.0000	1	1
B: Stop Pin	10.0625	0.283363969	35.511	0.0000	1	1
C: Pin Height	10.1875	0.283363969	35.952	0.0000	1	1
AB	-1.9375	0.283363969	-6.837	0.0001	1	1
AC	2.4375	0.283363969	8.602	0.0000	1	1
BC	0.6875	0.283363969	2.426	0.0382	1	1

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	6	11770	1961.6	1526.9	0.0000
Error	9	11.563	1.285		
Lack of Fit	1	3.063	3.063	2.882	0.1280
Pure Error	8	8.500	1.0625		
Total (Model + Error)	15	11781	785.43		

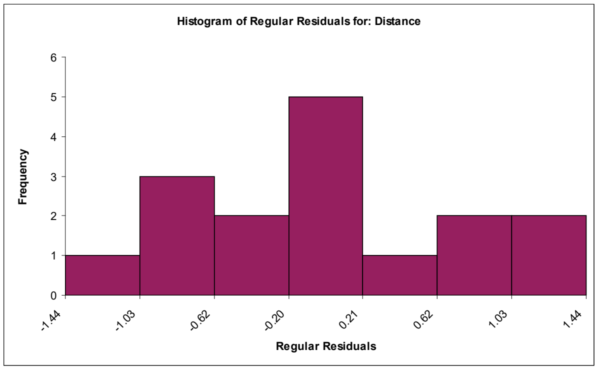
Durbin-Watson Test for Autocorrelation in Residuals:

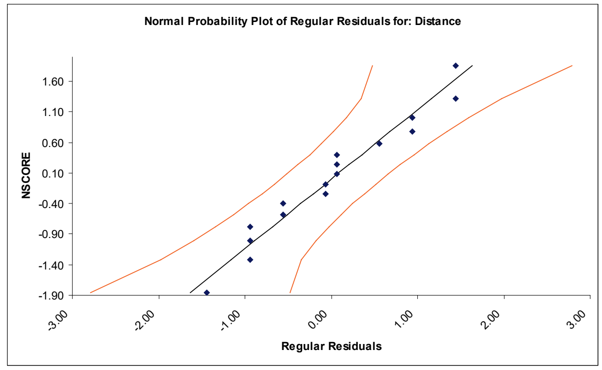
DW Statistic	1.609
P-Value Positive Autocorrelation	0.1339
P-Value Negative Autocorrelation	0.7654

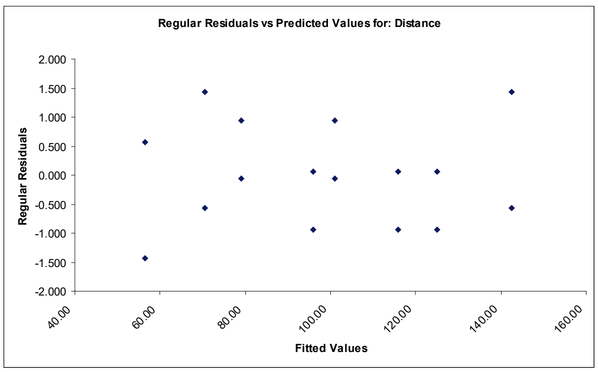


17. All the terms in the model are now significant, and there is no evidence of lack of fit (P-Value for lack-of-fit is 0.128 which is > .05).

18. Scroll down to view the Residual Plots. They also look very good, approximately normal, with no obvious patterns:







19. Scroll up to the **Predicted Response Calculator**. Enter the predicted values shown. These initial settings were determined by trial and error.

Predicted Response Calculator:

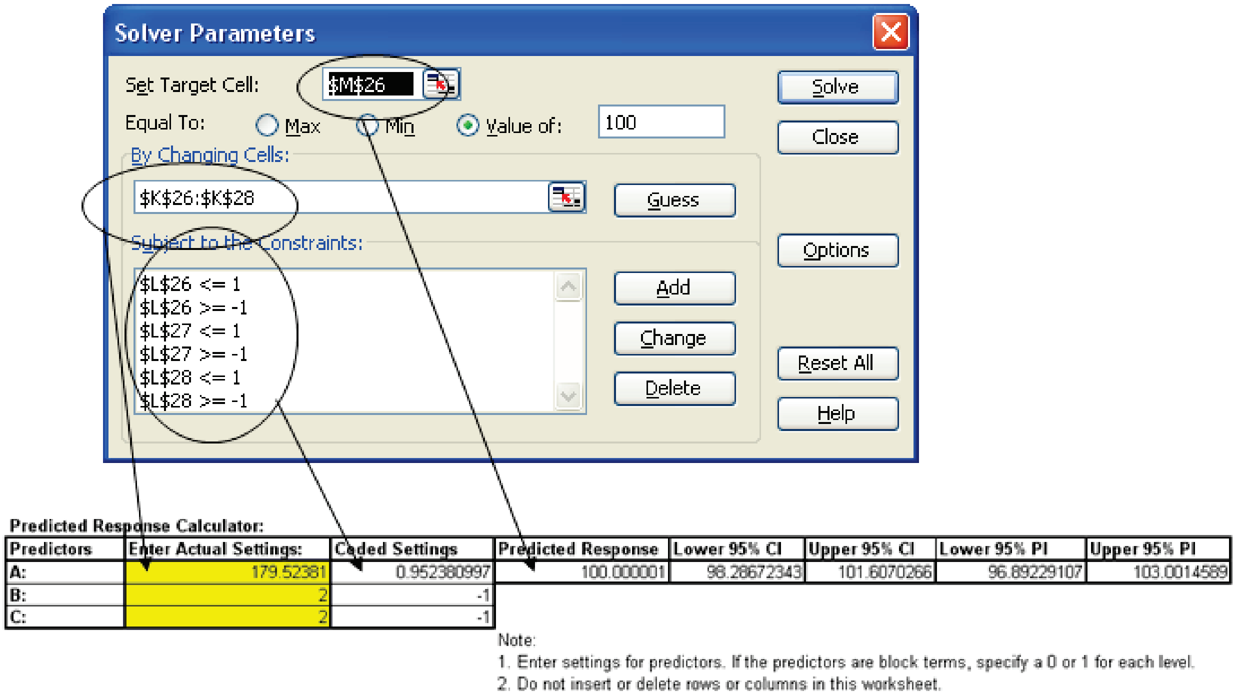
Predictors	Enter Actual Settings:	Coded Settings	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A:	179.5	0.95	99.946875	98.28672343	101.6070266	96.89229107	103.0014589
B:	2	-1					
C:	2	-1					

Note:
1. Enter settings for predictors. If the predictors are block terms, specify a 0 or 1 for each level.
2. Do not insert or delete rows or columns in this worksheet.

Note: CI is the 95% confidence interval for the long term mean and PI is the 95% prediction interval for individual values.

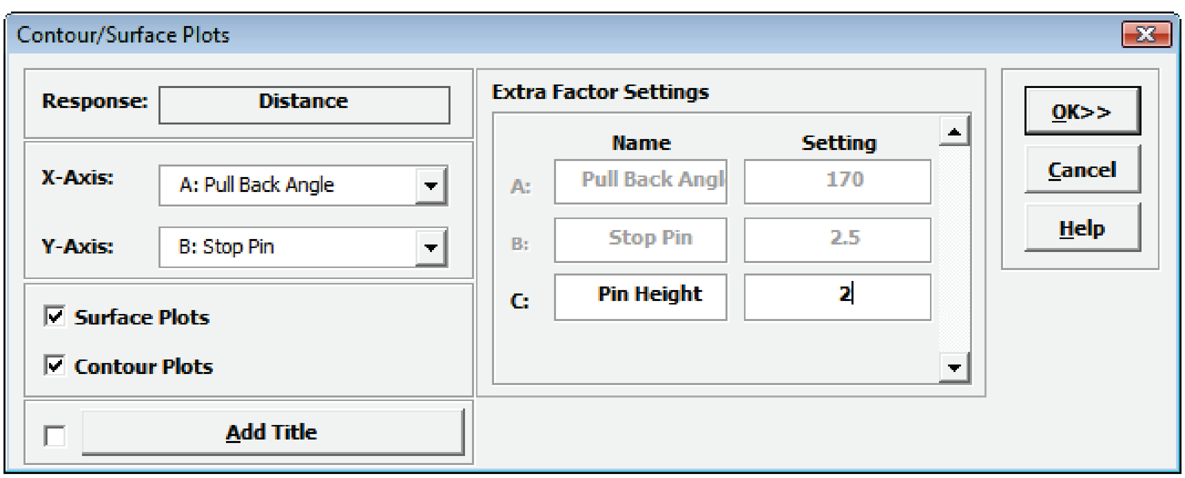
Note: The **DOE Multiple Regression Model: Distance** equation given above uses the coded coefficients, so if this equation is used to do predictions, the inputs must first be coded as done in the Predicted Response Calculator.

20. Excel's **Solver** may also be used to get a more exact solution:

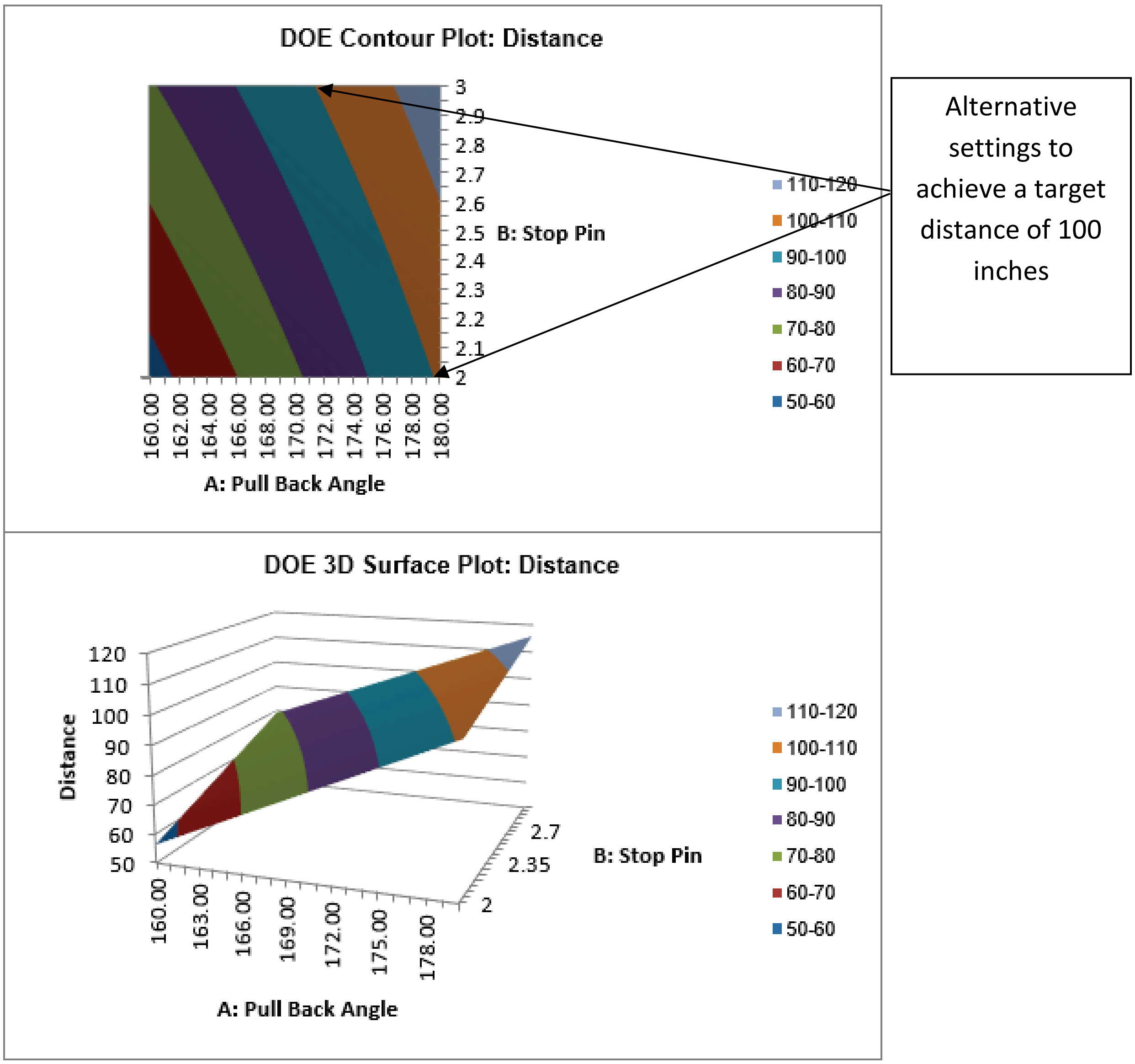


21. The model prediction must then be confirmed with actual experimental runs at the given settings of Pull Back Angle = 179.5, Stop Pin = 2, and Pin Height = 2.

22. Alternative settings to achieve the target distance may be obtained with Contour/Surface Plots. Click **SigmaXL > Design of Experiments > 2-Level Factorial/Screening > Contour/Surface Plots**. Set the **Pin Height** to 2 as shown (after clicking OK, you can use **Recall SigmaXL Dialog** to create another Contour/Surface plot with **Pin Height** set to 3):



23. Click **OK**. The following Contour and Surface Plots are displayed (with Pin Height = 2). Note the contour line with Catapult target distance = 100 inches. Although pin settings are discrete, they appear as continuous, so this will be a constraint in our selection of alternative settings. In addition to Pull Back Angle = 179.5, Stop Pin = 2, Pin Height = 2, we see that Pull Back Angle = approx. 171, Stop Pin = 3, Pin Height = 2 is also a valid setting. Alternative setting options are valuable in a designed experiment because they allow you to select lowest cost optimum settings, or settings that are easier to control.



Extra Factor Settings: Pin Height = 2

Tip: Use the contour/surface in conjunction with the predicted response calculator to determine optimal settings.

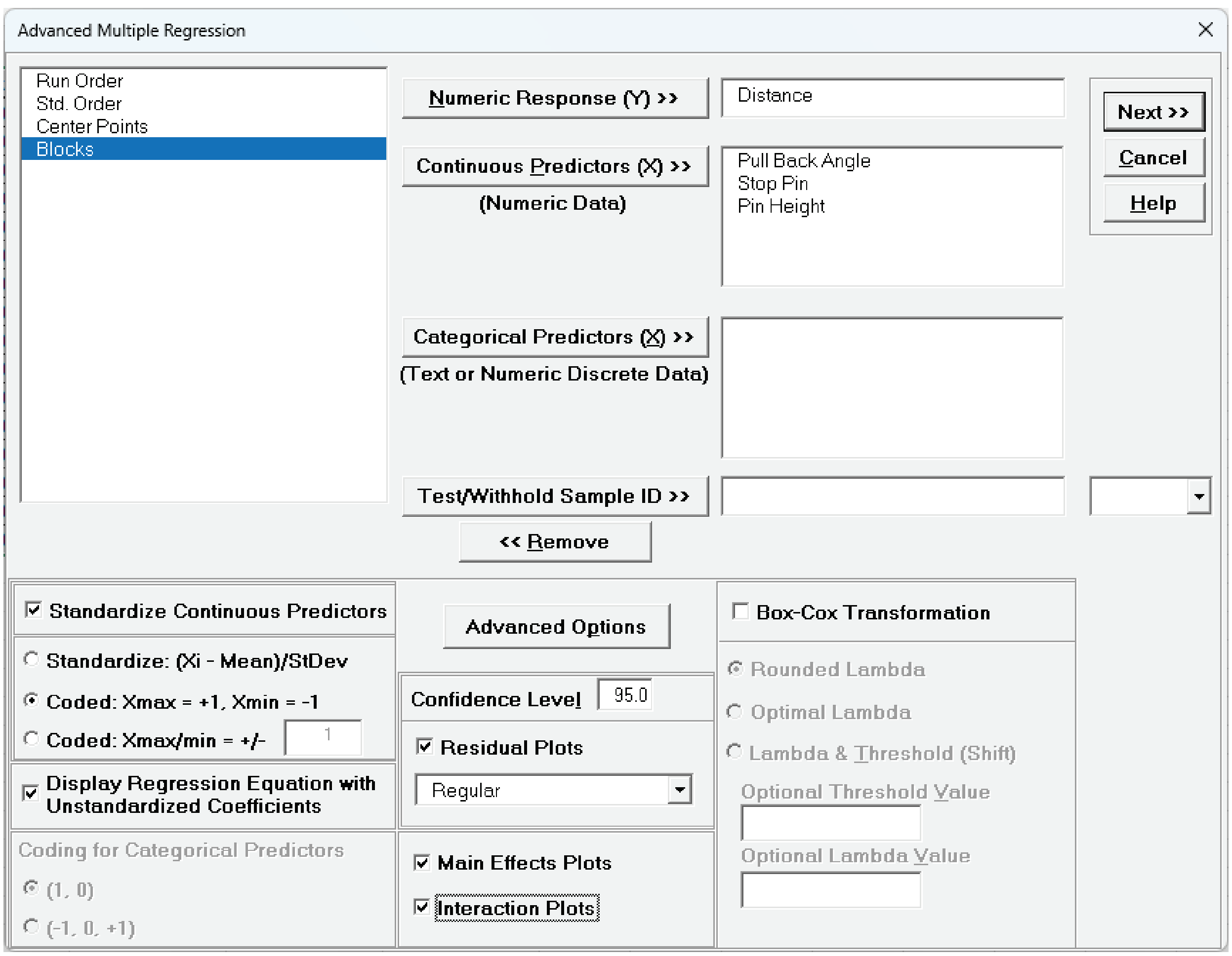
Analysis of Catapult Full Factorial Experiment with Advanced Multiple Regression

We will now redo the above analysis and optimization using Advanced Multiple Regression.

1. Open the file **Catapult DOE Data for Adv MReg.xlsx**. This is the Catapult DOE data copied into a workbook with *A:*, *B:* and *C:* removed from the Factor Names as they are not needed for Advanced Multiple Regression**.**

2. Click **Sheet 1** Tab. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, click **Use Entire Data Table**, click **Next**.

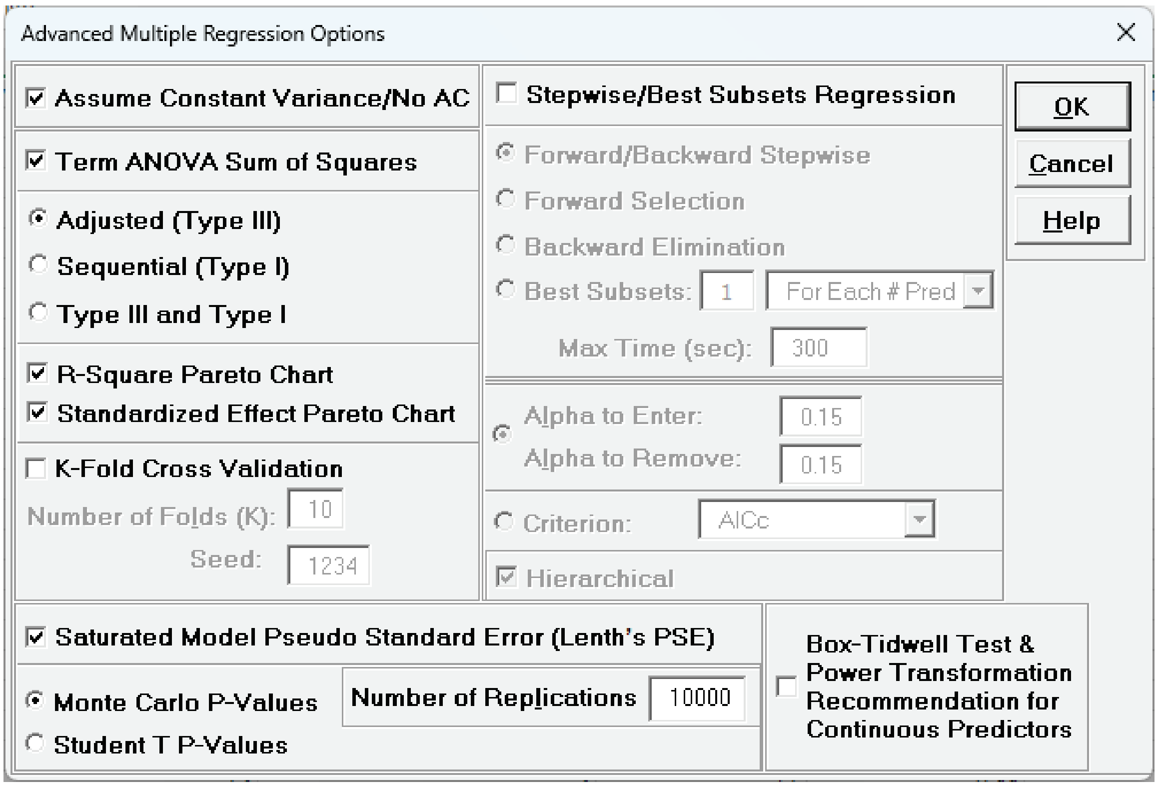
3. Select *Distance*, click **Numeric Response (Y) >>**; select *Pull Back Angle*, *Stop Pin*, and *Pin Height*; click **Continuous Predictors (X) >>**. Check **Standardize Continuous Predictors** with option **Coded: Xmax = +1, Xmin = -1**. Check **Display Regression Equation with Unstandardized Coefficients**. Use the default **Confidence Level** = 95.0%. Regular **Residual Plots** are checked by default. Check **Main Effects Plots** and **Interaction Plots**. Leave **Box-Cox Transformation** unchecked.



- **Standardize Continuous Predictors** with **Coded: Xmax = +1, Xmin = -1** scales the continuous predictors so that Xmax is set to +1 and Xmin is set to -1. This is particularly useful for analyzing data from a factorial design of experiments as we are doing here.

- **Display Regression Equation with Unstandardized Coefficients** displays the prediction equation with unstandardized/uncoded coefficients but the Parameter Estimates table will still show the standardized coefficients. This format is easier to interpret since there is only one coefficient value for each predictor.

4. Click **Advanced Options**. We will use the defaults as shown. Ensure that **Stepwise/Best Subsets Regression** is unchecked.



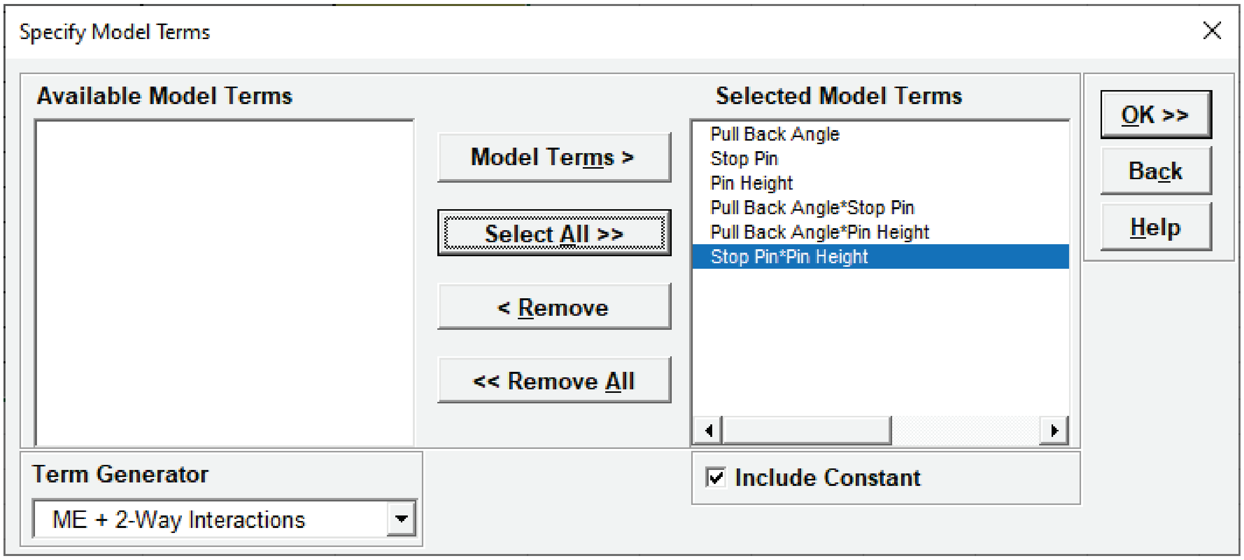
- **Term ANOVA Sum of Squares** with **Adjusted (Type III)** provides a detailed ANOVA table for continuous and categorical predictors. Adjusted Type III is the reduction in the error sum of squares (SS) when the term is added to a model that contains all the remaining terms.

- **R-Square Pareto Chart** displays a Pareto chart of term R-Square values ($100*SS_{term}/SS_{total}$). A separate Pareto Chart is produced for Type III and Type I SS. If there is only one predictor term, a Pareto Chart is not displayed.

- **Standardized Effect Pareto Chart** displays a Pareto chart of term T values (=T.INV(1-P/2,df_{error})). A separate Pareto Chart is produced for Type III and Type I SS. A significance reference line is include (=T.INV(1-alpha/2,df_{error})).

- **Saturated Model Pseudo Standard Error (Lenth's PSE)** is checked by default, but is not used here, as this is only applicable to saturated models with 0 error degrees of freedom.

5. Click **OK**. Using **Term Generator**, select *ME + 2-Way Interactions*. Click **Select All >>**. **Include Constant** is checked by default.



This matches the final model used in the original analysis for Distance. If we wanted to include the 3-Way Interaction, then *ME + All Interactions* would have been selected.

6. Click **OK**. The Advanced Multiple Regression report for Distance is given:

Multiple Regression Model (Uncoded): Distance = (-331.063) + (2.03125)*Pull_Back_Angle + (79.125)*Stop_Pin + (-69.375)*Pin_Height + (-0.3875)*Pull_Back_Angle*Stop_Pin + (0.4875)*Pull_Back_An

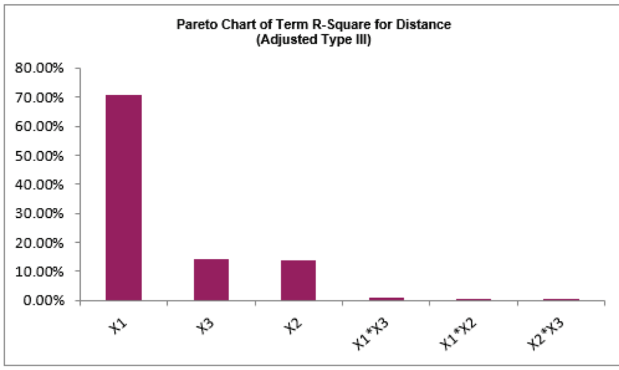
Model Summary	
R-Square	99.90%
R-Square Adjusted	99.84%
R-Square Predicted	99.69%
S (Root Mean Square Error)	1.1335

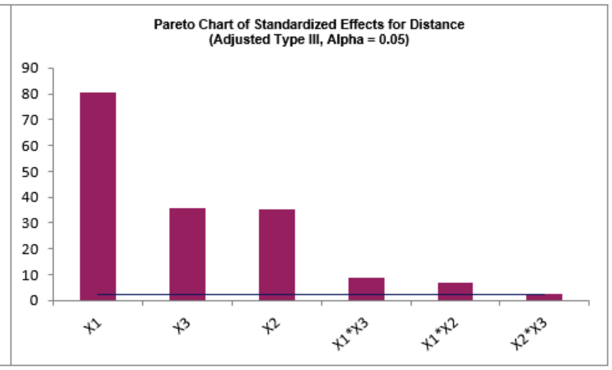
Model Information	
Continuous Predictor Standardization/Coding	Xmax = +1, Xmin = -1
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates - Coded						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	98.3125	0.283363969	346.9478	0.0000		
Pull Back Angle	22.8125	0.283363969	80.5060	0.0000	1.0000	1.0000
Stop Pin	10.0625	0.283363969	35.5109	0.0000	1.0000	1.0000
Pin Height	10.1875	0.283363969	35.9520	0.0000	1.0000	1.0000
Pull Back Angle*Stop Pin	-1.9375	0.283363969	-6.8375	0.0001	1.0000	1.0000
Pull Back Angle*Pin Height	2.4375	0.283363969	8.6020	0.0000	1.0000	1.0000
Stop Pin*Pin Height	0.6875	0.283363969	2.4262	0.0382	1.0000	1.0000

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	6	11769.875	1961.645833	1526.9027	0.0000
Error	9	11.5625	1.284722222		
Error: Lack-of-Fit	1	3.0625	3.0625	2.8824	0.1280
Error: Pure Error	8	8.5	1.0625		
Total (Model + Error)	15	11781.4375	785.4291667		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Pull Back Angle	1	8326.5625	8326.5625	6481.2162	0.0000	70.68%	80.5060
Stop Pin	1	1620.0625	1620.0625	1261.0216	0.0000	13.75%	35.5109
Pin Height	1	1660.5625	1660.5625	1292.5459	0.0000	14.09%	35.9520
Pull Back Angle*Stop Pin	1	60.0625	60.0625	46.7514	0.0001	0.51%	6.8375
Pull Back Angle*Pin Height	1	95.0625	95.0625	73.9946	0.0000	0.81%	8.6020
Stop Pin*Pin Height	1	7.5625	7.5625	5.8865	0.0382	0.06%	2.4262





Pareto Legend
X1 = Pull Back Angle
X2 = Stop Pin
X3 = Pin Height

Note, the prediction equation is uncoded so the coefficients do not match the coded coefficients given in the Parameter Estimates table. If consistency is desired, one can always rerun the analysis with **Display Regression Equation with Unstandardized Coefficients** unchecked. Blanks and special characters in the predictor names of the equation are converted to the underscore character "_".

The model summary statistics match the previous analysis. R-Square Predicted = 99.69%, also known as Leave-One-Out Cross-Validation, indicates how well a regression model predicts responses for new observations and is typically less than R-Square Adjusted. This is also very good.

The Parameter Estimates and ANOVA match the previous analysis. The Pareto Chart of Standardized Effects for Distance with significance line is similar to the Pareto Chart of Abs(Coefficient) but is based on the term T statistic.

Since this is an orthogonal design, Adjusted (Type III) Sum-of-Squares are the same as Sequential (Type I) Sum-of-Squares (not shown), so the Term R-Square Pareto shows the percent contribution to variabity in the Distance and sums to R-Square = 99.9%.

7. The Durbin-Watson Test for Autocorrelation in Residuals table is:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	1.6095
P-Value Positive Autocorrelation	0.1339
P-Value Negative Autocorrelation	0.7654

The Durbin Watson (DW) test is used to detect the presence of positive or negative autocorrelation in the residuals at Lag 1. If either P-Value is < .05, then there is significant autocorrelation. Here, there is no significant autocorrelation in the residuals, which is what we would expect in a randomized design of experiments.

8. The Breusch-Pagan Test for Constant Variance is:

Breusch-Pagan Test for Constant Variance (Normal)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	Chi-Square	DF	P-Value
All Terms	1.88494	6	0.9300
Pull Back Angle	0.134638	1	0.7137
Stop Pin	0.134638	1	0.7137
Pin Height	0.134638	1	0.7137
Pull Back Angle*Stop Pin	1.21175	1	0.2710
Pull Back Angle*Pin Height	0.134638	1	0.7137
Stop Pin*Pin Height	0.134638	1	0.7137

There are two versions of the Breusch-Pagan (BP) test for Constant Variance: Normal and Koenker Studentized – Robust. SigmaXL applies an Anderson-Darling Normality test to the residuals in order to automatically select which version to use. If the AD P-Value < 0.05, Koenker Studentized – Robust is used.

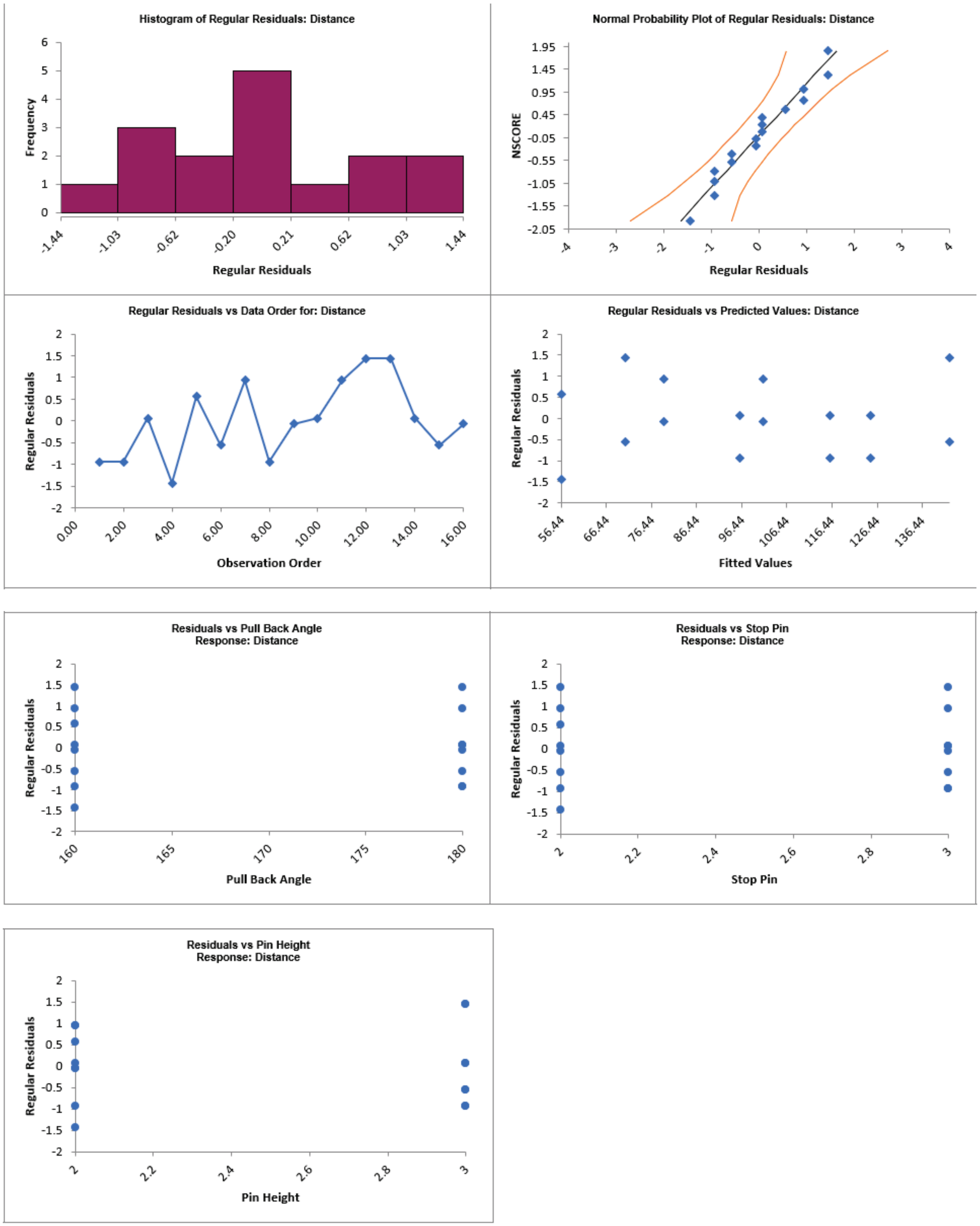
The report includes the test for *All Terms* and for individual predictors. *All Terms* denotes that all terms are in the model. This should be used to decide whether or not to take corrective action. The individual predictor terms are evaluated one-at-a-time and provide supplementary information for diagnostic purposes. Note, this should always be used in conjunction with an examination of the residual plots.

Here we see that the *All Terms* test is not significant, so we conclude that the variance is constant.

Tip: If the *All Terms* test is significant after model refinement, try a Box-Cox transformation. If that does not work, refit the model using **Recall Last Dialog,** click **Advanced Options** in the **Advanced Multiple Regression** dialog, and uncheck **Assume Constant Variance/No AC**. SigmaXL will apply the White robust standard errors for non-constant variance. For details, see the Appendix: **Advanced Multiple Regression**.

Tip: Lack of Constant Variance (a.k.a. Heteroskedasticity) is a nuisance for regression modelling but is also an opportunity. Examining the residual plots and individual predictors may yield process knowledge that identifies variance reduction opportunities.

9. Click on Sheet **MReg1 – Residuals** to view the Residual Plots. Note, Sheet MReg# will increment every time a model is refitted.

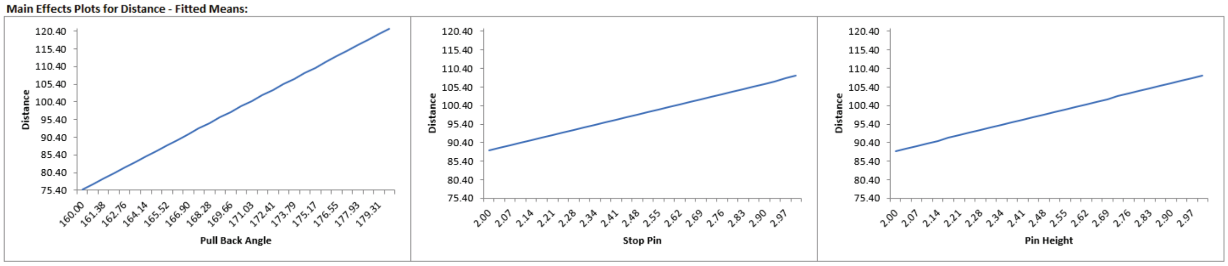


The Residual Plots are similar to those in the previous analysis and look very good, approximately normal, with no obvious patterns.

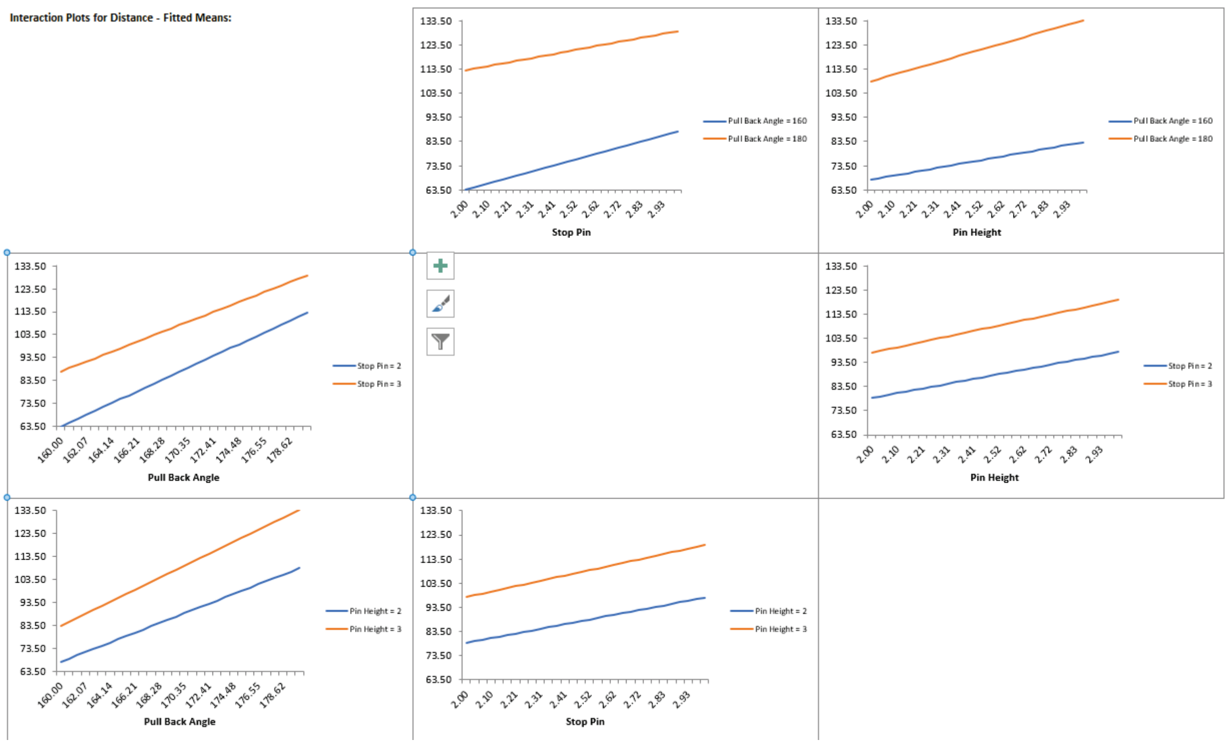
Note: Residuals versus interaction terms are not plotted, but they can be manually created using the model design matrix to the right of the Residual Plots (use **SigmaXL > Graphical Tools > Scatter Plots**).

10. Click on Sheet **MReg1 – Plots**. The Main Effects Plots and Interaction Plots for Overall Satisfaction are shown.

Main Effects Plots for Distance - Fitted Means:



Interaction Plots for Distance - Fitted Means:



These are based on Fitted Means as predicted by the model, not Data Means as used in the previous analysis. Main Effects Plots with Fitted Means use the predicted value for the response versus input predictor value, while holding all other variables at their respective means. Similarly for Interaction Plots, all predictors not being plotted are held at their respective means.

Pull Back Angle is the dominant factor having the steepest slope. We can also see that the interaction terms are weak with the almost parallel lines.

11. Click on Sheet **MReg1 – Model**. Scroll to the Predicted Response Calculator. Enter *Pull Back Angle* = 179.5, *Stop Pin* = 2, *Pin Height* = 2 to predict Overall Satisfaction with the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator								
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Pull Back Angle	179.5	99.946875	0.733879857	98.28672343	101.6070266	96.89229107	103.0014589
Stop Pin	2						
Pin Height	2						

Note the formula at cell **L14** is an Excel formula.

This matches the previous initial settings. Here the full predictor names are used making it easier to use and interpret. The Coded Settings are calculated as part of the Excel formula. Also, the prediction standard error **SE** is given.

12. Next, we will use SigmaXL's built in Optimizer. Scroll right to view the Optimize Options:

Optimize Options			
Continuous Predictors	Lower Bound	Upper Bound	Integer
Pull Back Angle	160	180	0
Stop Pin	2	3	0
Pin Height	2	3	0

Here we can constrain the lower and upper bounds of the continuous predictors. (If there was a categorical predictor, e.g., different ball type, you could also specify a ball type to use for optimization). Stop Pin and Pin Height are constrained to integers so these should be changed from 0 to 1 as shown.

Optimize Options			
Continuous Predictors	Lower Bound	Upper Bound	Integer
Pull Back Angle	160	180	0
Stop Pin	2	3	1
Pin Height	2	3	1

The Optimizer will return only integer values for Stop Pin and Pin Height.

13. Scroll back to view the Goal setting and Optimize button. Specify Target = 100 as shown.

Goal:	Target
Target:	100

Optimize   Contour/Surface Plots

The optimizer uses Multistart Nelder-Mead Simplex to solve for the desired response goal with given constraints. For more information see the Appendix: Single Response Optimization.

14. Click **Optimize**. The response solution and prompt to paste values into the Input Settings of the Predicted Response Calculator is given:

SigmaXL Information

Predicted Response = 99.9999999999986

Do you want to paste the new predictor values?

Yes No

15. Click **Yes** to paste the values.

Predicted Response Calculator								
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Pull Back Angle	170.8275862	100	0.568181641	98.71468383	101.2853162	97.13182682	102.8681732
Stop Pin	2						
Pin Height	3						

This agrees with the alternative solution that was obtained using the Contour Plot in the previous analysis. In order to match the previous Solver solution which used *Stop Pin* = 2 and *Pin Height* = 2, we can constrain the Optimizer to those pin settings.

16. Scroll to view the Optimize Options, and change the *Stop Pin* **Lower Bound** = 2, **Upper Bound** = 2; *Pin Height* **Lower Bound** = 2, **Upper Bound** = 2 as shown:

Optimize Options			
Continuous Predictors	Lower Bound	Upper Bound	Integer
Pull Back Angle	160	180	0
Stop Pin	2	2	1
Pin Height	2	2	1

17. Click **Optimize**. The response solution and prompt to paste values into the Input Settings of the Predicted Response Calculator is given:

SigmaXL Information

Predicted Response = 99.9999999999989

Do you want to paste the new predictor values?

Yes No

18. Click **Yes** to paste the values.

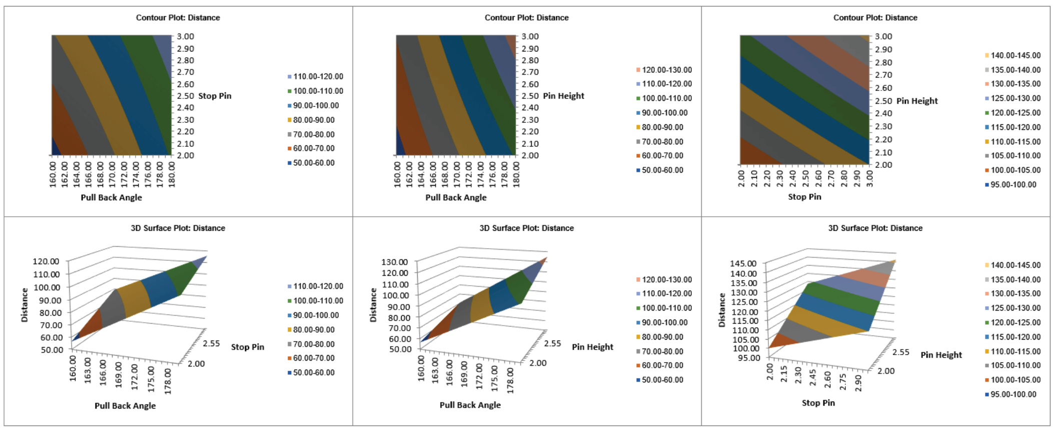
Predicted Response Calculator								
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Pull Back Angle	179.5238095	100	0.734622848	98.33816766	101.6618323	96.94450226	103.0554977
Stop Pin	2						
Pin Height	2						

This now matches the Solver solution obtained in the previous analysis. Note however that the SE for the original SigmaXL solution is lower than the Solver solution. This is by design, when multiple valid solutions are available, SigmaXL selects the one with the lowest prediction SE.

19. Next, we will create a Contour/Surface Plot. Click the **Contour/Surface Plots** button. Note that *Stop Pin* and *Pin Height* are not constrained to be integers in these plots.

20. A new sheet is created, **MReg1 – Contour** that displays the plots:

Predictors	Hold Values
Pull Back Angle	179.5238095
Stop Pin	2
Pin Height	2



SigmaXL automatically creates a Contour/Surface Plot for each pairwise combination of continuous predictors. The plots on the left match those specified in the previous analysis.

Note that the table with the **Hold Values** gives the values used to hold a predictor constant if it is not in the plot.

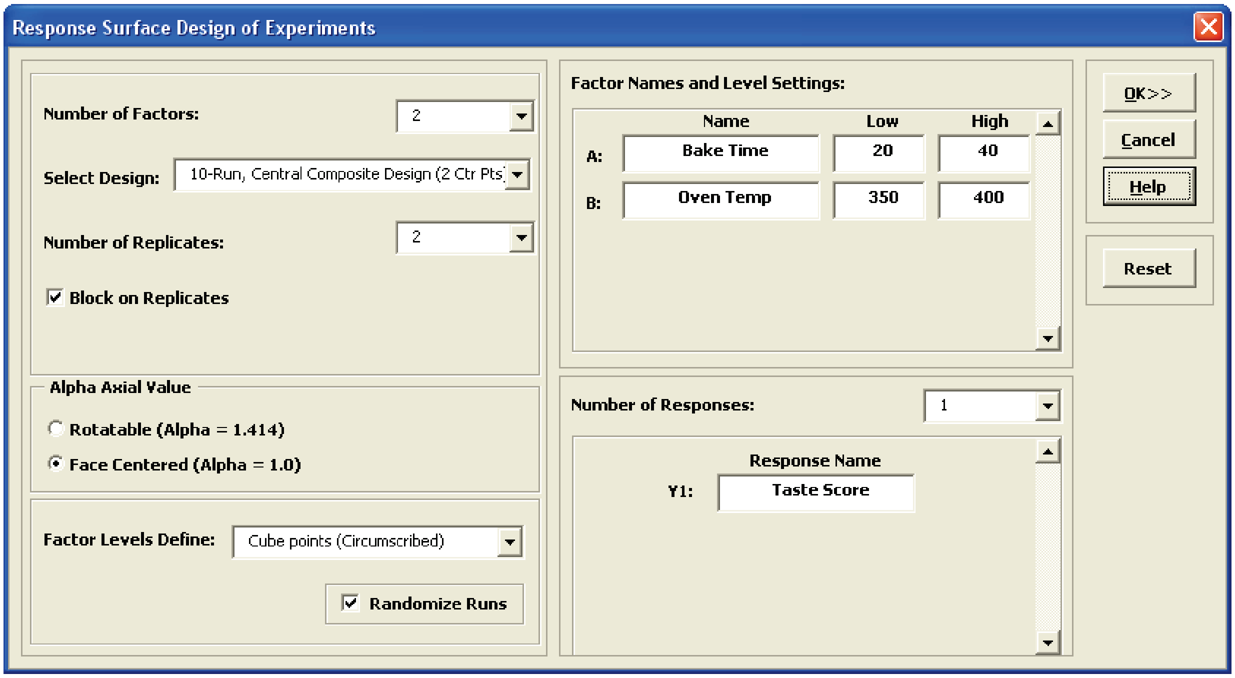
Tip: The hold values are obtained from the Predicted Response Calculator settings, so if you wish to use different **Hold Values**, simply select the Model sheet, change the **Enter Settings** values and recreate the plots.

Tip: Use the contour/surface plots in conjunction with the predicted response calculator to determine optimal settings.

Part D – Design and Analysis of Response Surface Experiment – Cake Bake

1. We will illustrate the use of response surface methods using a layer cake baking experiment. The response variable is Taste Score (on a scale of 1-7 where 1 is "awful" and 7 is "delicious"). Average scores for a panel of tasters have been recorded. The X Factors are A: Bake Time (20 to 40 minutes) and B: Oven Temperature (350 to 400 F). The experiment goal is to find the settings that maximize taste score. Other factors such as pan size need to be taken into consideration. In this experiment, these are held constant.
2. Click **SigmaXL > Design of Experiments > Response Surface > Response Surface Designs**.
3. The Number of X Factors can be 2 to 5. Select **Number of Factors** = 2.
4. The available designs are sorted by number of runs. Increasing the number of runs allows for uniform precision or blocking, but we will select the design with fewest runs, the **10-Run, Central Composite Design (2 Ctr Pts)**.
5. As discussed earlier with the catapult experiment, it is a good practice to replicate an experiment if affordable to do so. Here we will select **Number of Replicates** = 2 and check **Block on Replicates**. Blocking on replicates allows us to perform the experiment over a two week period, with each block corresponding to week number.
6. Change the **Alpha Axial Value** option to **Face Centered (Alpha = 1.0)**. This simplifies the design to a 3-level design, rather than a 5-level design with alpha = 1.414. (The trade-off is that we lose the desirable statistical property of rotatability for prediction).

   Tip: This alpha is the distance from the center point to the end of the axial (star) point. Do not confuse this with **Alpha for Pareto Chart** which is the P-Value used to determine statistical significance. Unfortunately, the term “alpha” has been chosen by statisticians to define two completely different things.
7. Enter **Factor Names and Level Settings** and **Response Name** as shown:



8. Click **OK**. The following worksheet is produced:

Response Surface Methods Design Worksheet

Title:	
Date:	
Name of Experimenter:	
Notes:	

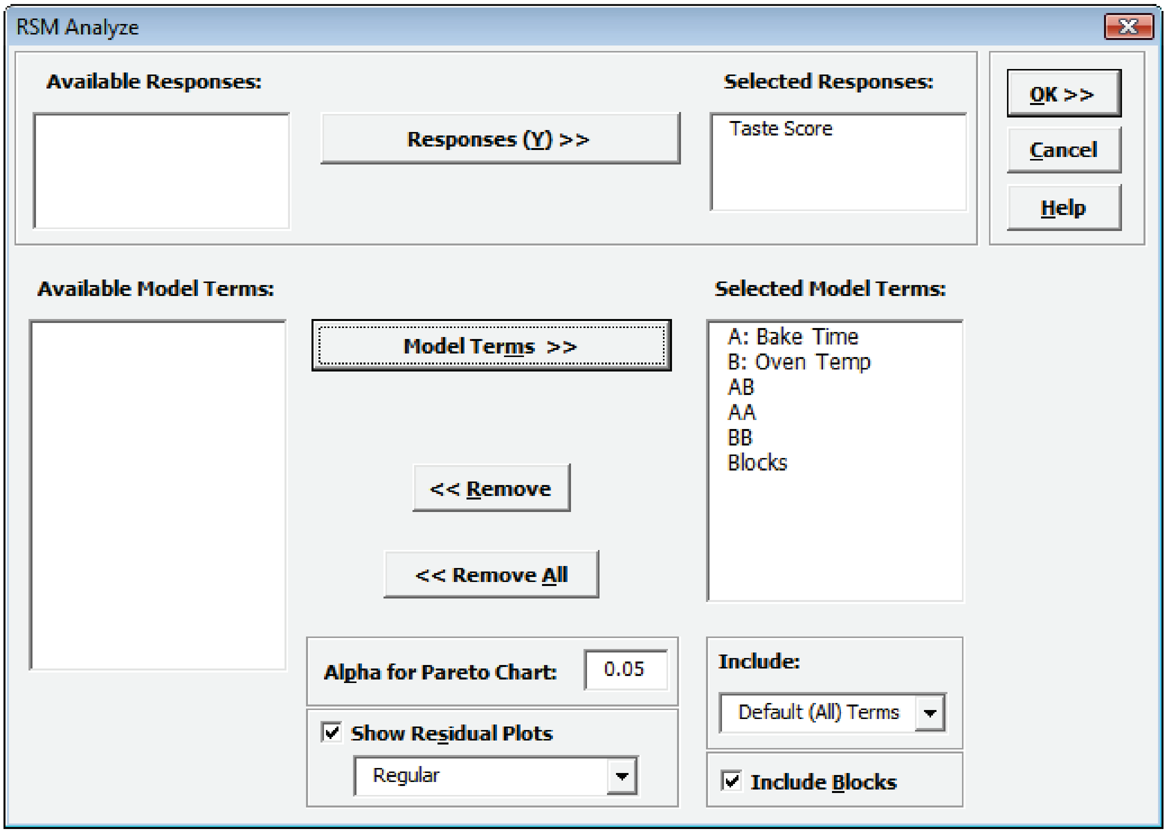
Design Type:	2 Factor, 10-Run, Central Composite Design (2 Ctr Pts)
Alpha Axial Value:	Face Centered (Alpha = 1.0)
Factor Levels Define:	Cube points (Circumscribed)
Number of Replicates:	2
Block on Replicates:	Yes
Number of Responses:	1

Run Order	Std. Order	Center Points	Blocks	A: Bake Time	B: Oven Temp	Taste Score
1	9	0	1	30	375	
2	1	1	1	20	350	
3	7	1	1	30	350	
4	10	0	1	30	375	
5	5	1	1	20	375	
6	3	1	1	20	400	
7	2	1	1	40	350	
8	6	1	1	40	375	
9	4	1	1	40	400	
10	8	1	1	30	400	
11	14	1	2	40	400	
12	20	0	2	30	375	
13	12	1	2	40	350	
14	13	1	2	20	400	
15	16	1	2	40	375	
16	19	0	2	30	375	
17	18	1	2	30	400	
18	15	1	2	20	375	
19	17	1	2	30	350	
20	11	1	2	20	350	

Do not add or delete rows in this worksheet.

9. Open the file **RSM Example – Cake Bake** to obtain response values.

10. Click **SigmaXL > Design of Experiments > Response Surface > Analyze Response Surface Design**.



11. We will use the default analyze settings (all terms in the model, including the block term) to start. Click **OK**. The resulting Analysis report is shown:

Response Surface Methods Analysis

RSM Regression Model: Taste Score = (6.537142857) + (-1.3) * A: Bake Time + (-0.966666667) * B: Oven Temp + (-0.4625) * AB + (-0.8892

Title:	Cake Bake RSM Example
Date:	
Name of Experimenter:	
Notes:	

Design Type:	2 Factor, 10-Run, Central Composite Design (2 Ctr Pts)
Alpha Axial Value:	Face Centered (Alpha = 1.0)
Factor Levels Define:	Cube points (Circumscribed)
Number of Replicates:	2 - Center Points are excluded from the model
Block on Replicates:	Yes
Response	Taste Score

Model Summary:

R-Square	99.00%
R-Square Adjusted	98.54%
S (Root Mean Square Error)	0.186698

Parameter Estimates (Coded Units):

Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	6.537142857	0.089258824	73.238	0.0000		
A: Bake Time	-1.3	0.053895135	-24.121	0.0000	1	1
B: Oven Temp	-0.966666667	0.053895135	-17.936	0.0000	1	1
AB	-0.4625	0.06600779	-7.007	0.0000	1	1
AA	-0.889285714	0.086424485	-10.290	0.0000	1.028571429	0.972222222
BB	-1.139285714	0.086424485	-13.182	0.0000	1.028571429	0.972222222
Blocks_2	-0.11	0.083493984	-1.317	0.2104	1	1

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Blocks	1	0.0605	0.0605	1.736	0.2104
Model	5	44.852	8.970	257.35	0.0000
Error	13	0.453131	0.03485623		
Lack of Fit	11	0.373131	0.033921	0.848025	0.6564
Pure Error	2	0.08	0.04		
Total (Model + Error)	19	45.366	2.388		

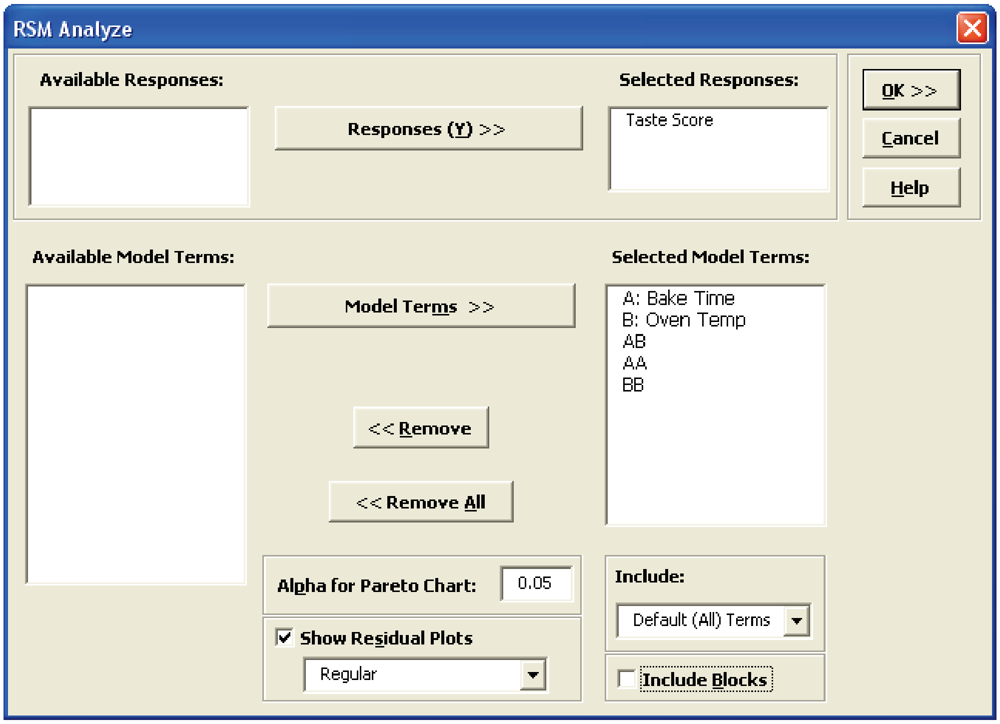
Durbin-Watson Test for Autocorrelation in Residuals:

DW Statistic	2.720
P-Value Positive Autocorrelation	0.9407
P-Value Negative Autocorrelation	0.0315

12. The model looks very good with an R-Square value of 99.0%. The standard deviation (experimental error) is only 0.19 on a 1 to 7 taste scale. All of the model terms are statistically significant ($P < .05$), but the Block term is not, so it should be removed from the model. Note that AA and BB denote the quadratic model terms.

13. Click **Recall Last Dialog** (or press **F3**).

14. Uncheck **Include Blocks** as shown:



15. Click **OK**. The revised report is shown below:

Model Summary:

R-Square	98.87%
R-Square Adjusted	98.46%
S (Root Mean Square Error)	0.191541

Parameter Estimates (Coded Units):

Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	6.482142857	0.080940813	80.085	0.0000		
A: Bake Time	-1.3	0.055293102	-23.511	0.0000	1	1
B: Oven Temp	-0.966666667	0.055293102	-17.483	0.0000	1	1
AB	-0.4625	0.067719943	-6.830	0.0000	1	1
AA	-0.889285714	0.088666218	-10.029589	0.0000	1.028571429	0.972222222
BB	-1.139285714	0.088666218	-12.849	0.0000	1.028571429	0.972222222

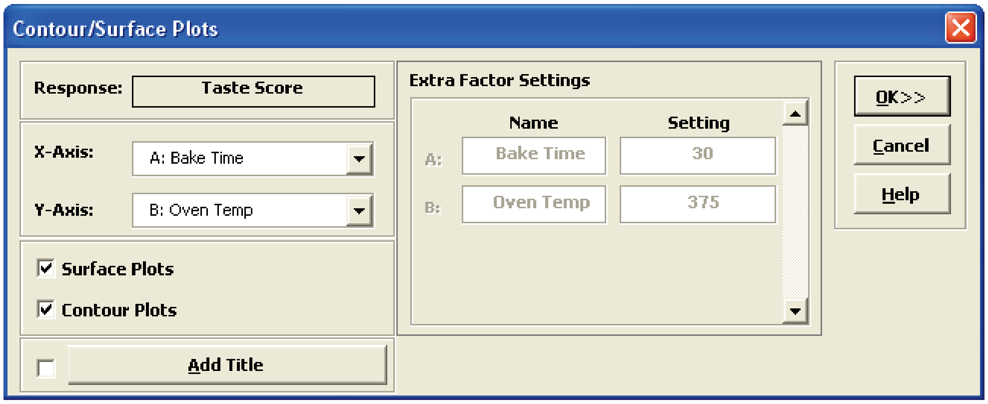
Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	5	44.852	8.970	244.50	0.0000
Error	14	0.513631	0.03668793		
Lack of Fit	3	0.248631	0.08287698	3.440	0.0555
Pure Error	11	0.265000	0.02409091		
Total (Model + Error)	19	45.366	2.388		

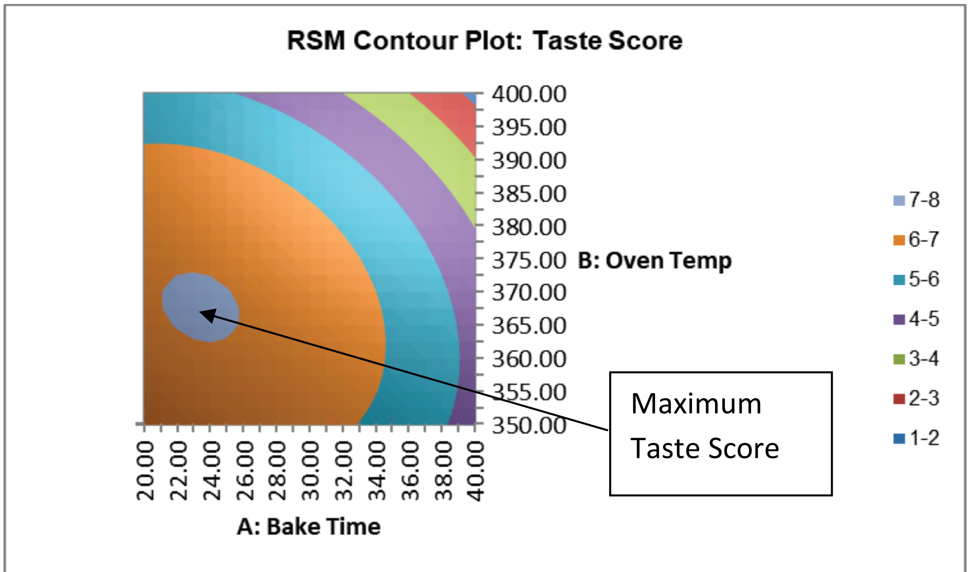
Durbin-Watson Test for Autocorrelation in Residuals:

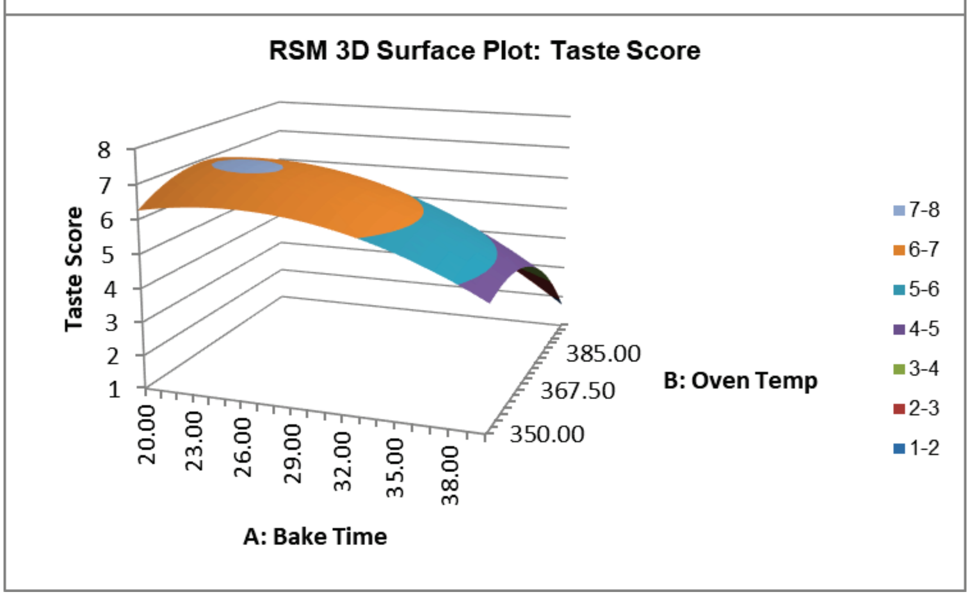
DW Statistic	2.687
P-Value Positive Autocorrelation	0.9597
P-Value Negative Autocorrelation	0.0798

16. To create a contour and surface plot, click **SigmaXL > Design of Experiments > Response Surface > Contour/Surface Plots**.



17. Click **OK**. The following Contour and Surface Plots are displayed.





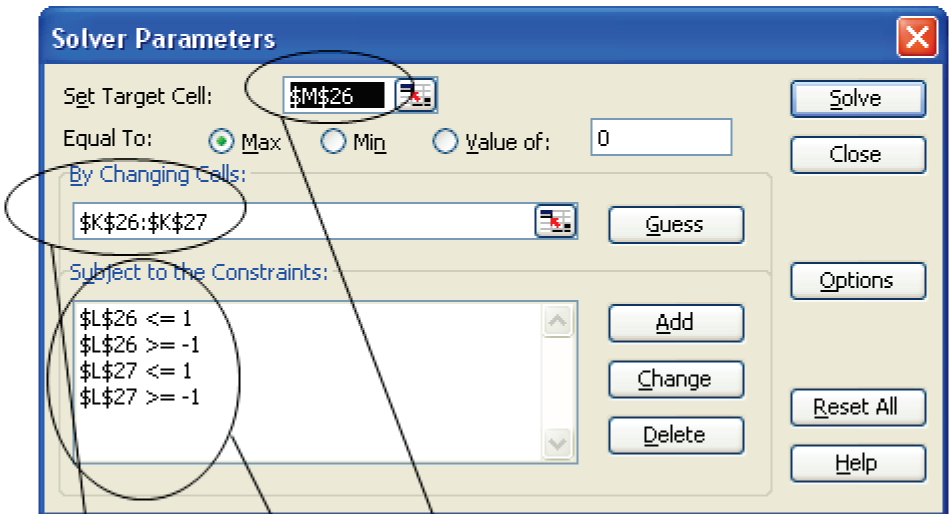
18. Go to **Analyze – 2 Factor RSM** tab. Scroll across to the **Predicted Response Calculator**. Enter the predicted values shown. These initial settings were determined from the contour plot as an estimate to yield a maximum taste score.

Predicted Response Calculator:

Predictors	Enter Actual Settings:	Coded Settings	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A:	23	-0.7	7.047089524	6.883297028	7.21088202	6.604826626	7.489352422
B:	368	-0.28					

Note: The **RSM Regression Model: Taste Score** equation given above uses the coded coefficients, so if this equation is used to do predictions, the inputs must first be coded as done in the Predicted Response Calculator.

19. Excel's **Solver** may also be used to get a more exact solution:



Predicted Response Calculator:

Predictors	Enter Actual Settings:	Coded Settings	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A:	23.44815482	-0.655184518	7.048786514	6.883297028	7.21088202	6.604826626	7.489352422
B:	367.7184597	-0.29126161					

20. Although the model is predicting values that exceed the maximum taste score of 7, this is expected to give us the best possible settings for cook time and bake temperature. To quote the eminent statistician George Box, "All models are wrong, some are useful". Additional experimental runs carried out at Time = 23.4 minutes and Temperature = 367.7 F. confirm that these are ideal settings.

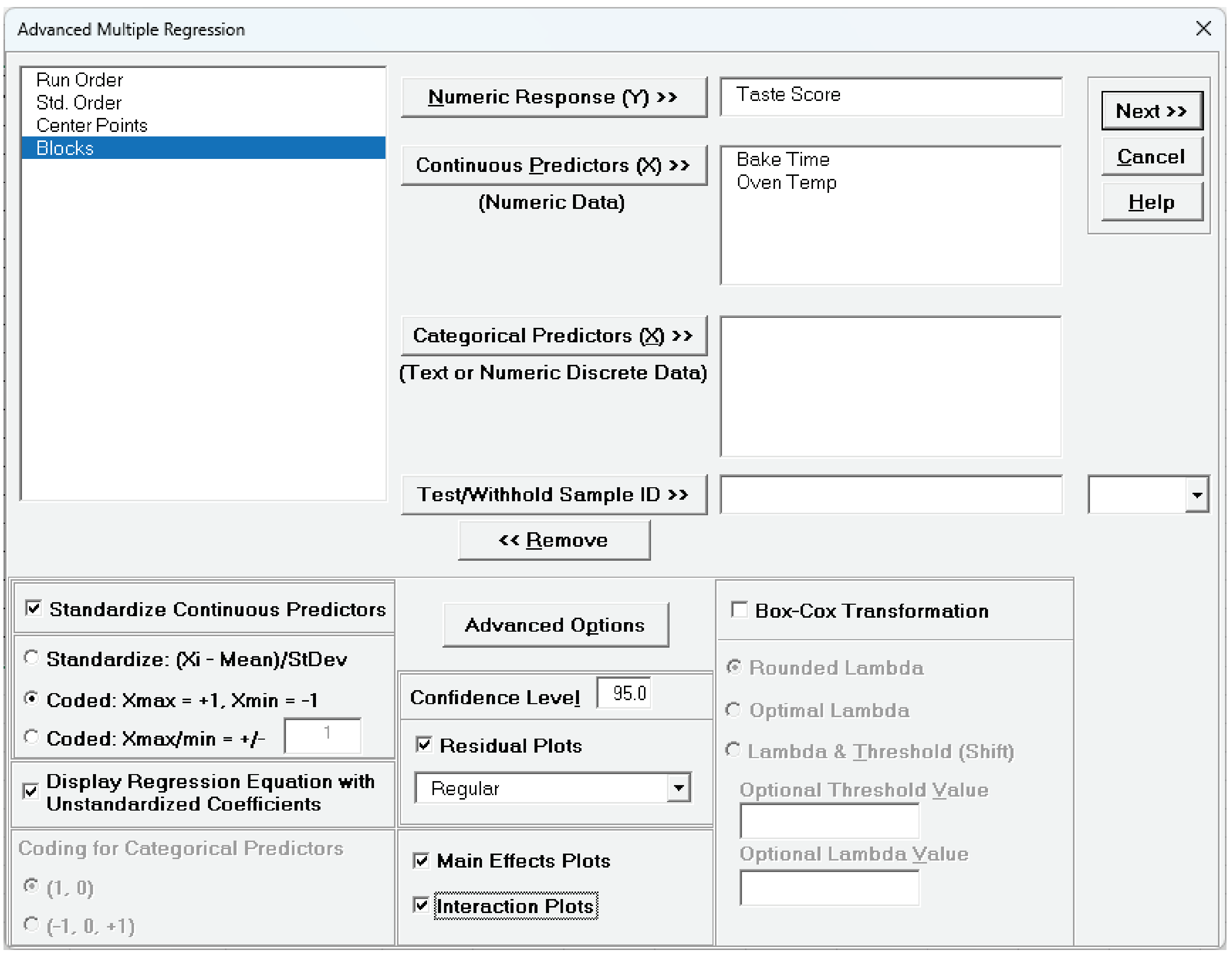
Analysis of Response Surface Experiment – Cake Bake with Advanced Multiple Regression

We will now redo the above analysis and optimization using Advanced Multiple Regression.

1. Open the file **RSM Example – Cake Bake Data for Adv MReg.xlsx**. This is the RSM Example - Cake Bake data copied into a workbook with *A:* and *B:* removed from the Factor Names as they are not needed for Advanced Multiple Regression**.**

2. Click **Sheet 1** Tab. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**. If necessary, click **Use Entire Data Table**, click **Next**.

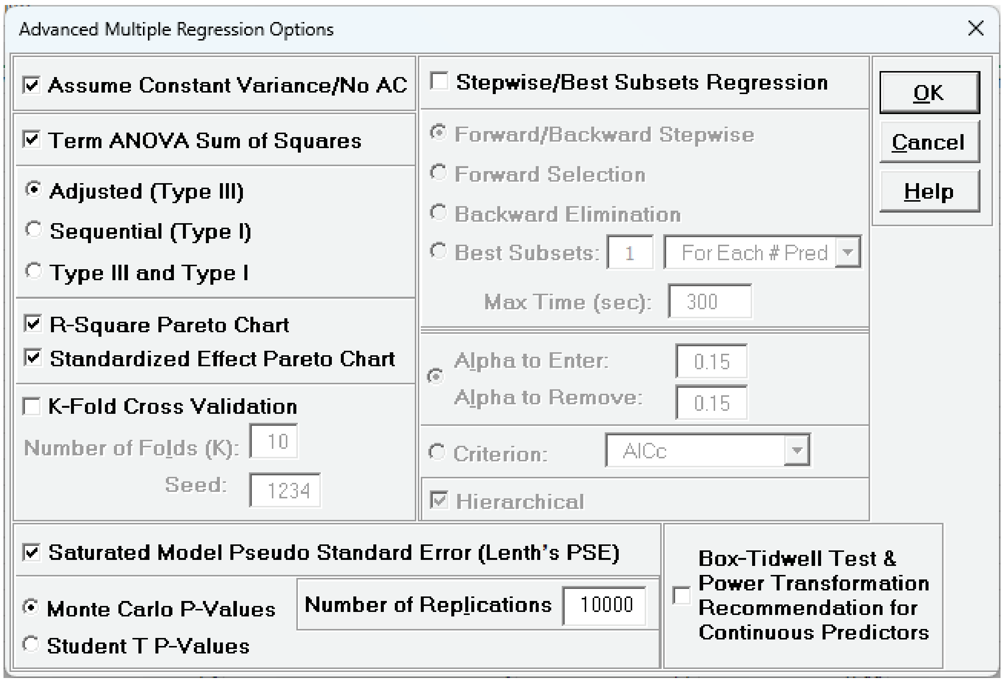
3. Select *Taste Score*, click **Numeric Response (Y) >>**; select *Bake Time* and *Oven Temp*; click **Continuous Predictors (X) >>**. Check **Standardize Continuous Predictors** with option **Coded: Xmax = +1, Xmin = -1**. Check **Display Regression Equation with Unstandardized Coefficients**. Use the default **Confidence Level** = 95.0%. Regular **Residual Plots** are checked by default. Check **Main Effects Plots** and **Interaction Plots.** Leave **Box-Cox Transformation** unchecked.



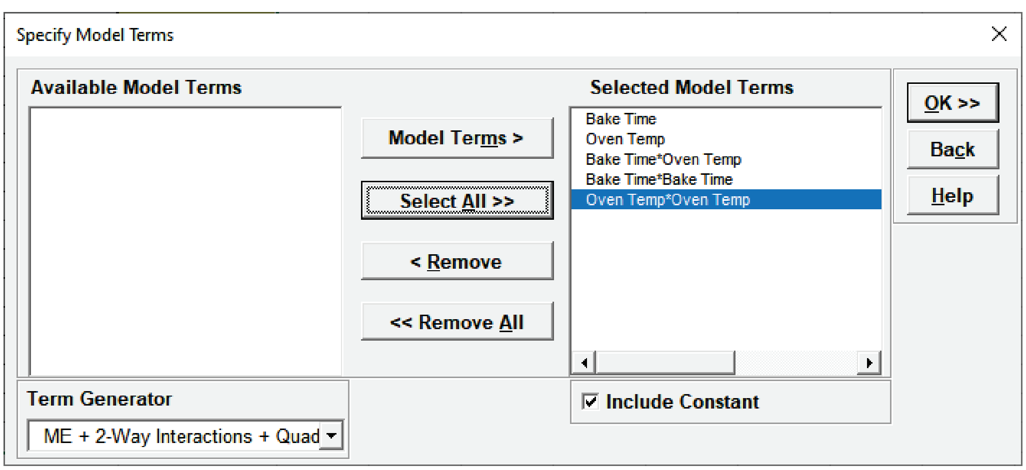
- **Standardize Continuous Predictors** with **Coded: Xmax = +1, Xmin = -1** scales the continuous predictors so that Xmax is set to +1 and Xmin is set to -1. This is particularly useful for analyzing data from a 3-Level Face-Centered RSM design of experiments with Alpha axial = 1.0. If the design was a 5-Level Rotatable with Alpha axial = 1.414, we would have selected **Standardize Continuous Predictors** with **Coded: Xmax/Xmin = +/-** 1.414**.**

- **Display Regression Equation with Unstandardized Coefficients** displays the prediction equation with unstandardized/uncoded coefficients but the Parameter Estimates table will still show the standardized coefficients. This format is easier to interpret since there is only one coefficient value for each predictor.

4. Click **Advanced Options**. We will use the defaults as shown. Ensure that **Stepwise/Best Subsets Regression** is unchecked.



5. Click **OK**. Using **Term Generator**, select *ME + 2-Way Interactions + Quadratic*. Click **Select All >>**. **Include Constant** is checked by default.



This matches the final model used in the original analysis for Taste Score.

6. Click **OK**. The Advanced Multiple Regression report for Taste Score is given:

Multiple Regression Model (Uncoded): Taste Score = (-260.273) + (1.09732)*Bake_Time + (1.38398)*Oven_Temp + (-0.00185)*Bake_Time*Oven_Temp + (-0.00889286)*Bake_Time*Bake_Time + (-0.00

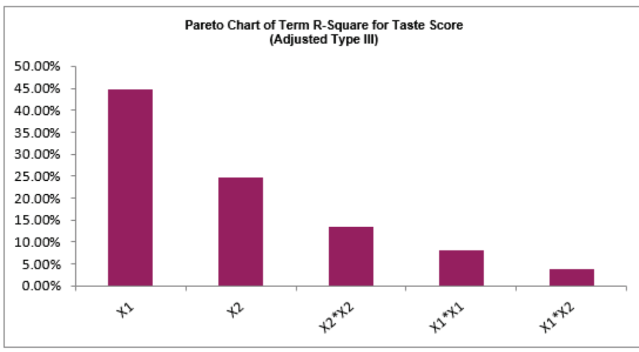
Model Summary	
R-Square	98.87%
R-Square Adjusted	98.46%
R-Square Predicted	97.89%
S (Root Mean Square Error)	0.1915

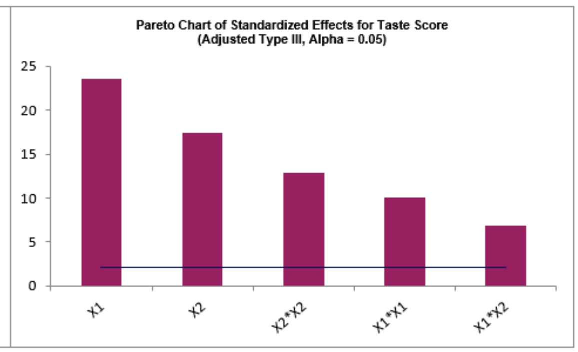
Model Information	
Continuous Predictor Standardization/Coding	Xmax = +1, Xmin = -1
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates - Coded						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	6.482142857	0.080940813	80.0850	0.0000		
Bake Time	-1.3	0.055293102	-23.5111	0.0000	1.0000	1.0000
Oven Temp	-0.966666667	0.055293102	-17.4826	0.0000	1.0000	1.0000
Bake Time*Oven Temp	-0.4625	0.067719943	-6.8296	0.0000	1.0000	1.0000
Bake Time*Bake Time	-0.889285714	0.088666218	-10.0296	0.0000	1.0286	0.9722
Oven Temp*Oven Temp	-1.139285714	0.088666218	-12.8492	0.0000	1.0286	0.9722

Analysis of Variance for Model					
Source	DF	SS	MS	F	P
Model	5	44.85186905	8.97037381	244.5048	0.0000
Error	14	0.513630952	0.036687925		
Error: Lack-of-Fit	3	0.248630952	0.082876984	3.4402	0.0555
Error: Pure Error	11	0.265	0.024090909		
Total (Model + Error)	19	45.3655	2.387657895		

Analysis of Variance for Predictors (Adjusted Type III)							
Predictor Term	DF	SS	MS	F	P	R-Square	Std. Effect (T)
Bake Time	1	20.28	20.28	552.7704	0.0000	44.70%	23.5111
Oven Temp	1	11.21333333	11.21333333	305.6410	0.0000	24.72%	17.4826
Bake Time*Oven Temp	1	1.71125	1.71125	46.6434	0.0000	3.77%	6.8296
Bake Time*Bake Time	1	3.690535714	3.690535714	100.5927	0.0000	8.14%	10.0296
Oven Temp*Oven Temp	1	6.057202381	6.057202381	165.1007	0.0000	13.35%	12.8492





Pareto Legend
X1 = Bake Time
X2 = Oven Temp

Note, the prediction equation is uncoded so the coefficients do not match the coded coefficients given in the Parameter Estimates table. If consistency is desired, one can always rerun the analysis with **Display Regression Equation with Unstandardized Coefficients** unchecked. Blanks and special characters in the predictor names of the equation are converted to the underscore character "_".

The model summary statistics match the previous analysis. R-Square Predicted = 97.89%, also known as Leave-One-Out Cross-Validation, indicates how well a regression model predicts responses for new observations and is typically less than R-Square Adjusted. This is also very good.

The Parameter Estimates and ANOVA match the previous analysis. The Pareto Chart of Standardized Effects for Taste Score is similar to the Pareto Chart of Abs(Coefficient) but is based on the term T statistic.

Since this is an orthogonal design, Adjusted (Type III) Sum-of-Squares are the same as Sequential (Type I) Sum-of-Squares (not shown), so the Term R-Square Pareto shows the percent contribution to variability in the Taste Score and sums to R-Square = 98.87%.

7. The Durbin-Watson Test for Autocorrelation in Residuals table is:

Durbin-Watson Test for Autocorrelation in Residuals	
DW Statistic	2.6874
P-Value Positive Autocorrelation	0.9597
P-Value Negative Autocorrelation	0.0798

The Durbin Watson (DW) test is used to detect the presence of positive or negative autocorrelation in the residuals at Lag 1. If either P-Value is < .05, then there is significant autocorrelation. Here, there is no significant autocorrelation in the residuals, which is what we would expect in a randomized design of experiments.

8. The Breusch-Pagan Test for Constant Variance is:

Breusch-Pagan Test for Constant Variance (Normal)			
H0: Variance is constant; Ha: Variance is not constant.			
Predictor Term	Chi-Square	DF	P-Value
All Terms	4.42181	5	0.4904
Bake Time	0.0383644	1	0.8447
Oven Temp	1.61007	1	0.2045
Bake Time*Oven Temp	0.0553441	1	0.8140
Bake Time*Bake Time	1.12188	1	0.2895
Oven Temp*Oven Temp	2.02278	1	0.1550

There are two versions of the Breusch-Pagan (BP) test for Constant Variance: Normal and Koenker Studentized – Robust. SigmaXL applies an Anderson-Darling Normality test to the residuals in order to automatically select which version to use. If the AD P-Value < 0.05, Koenker Studentized – Robust is used.

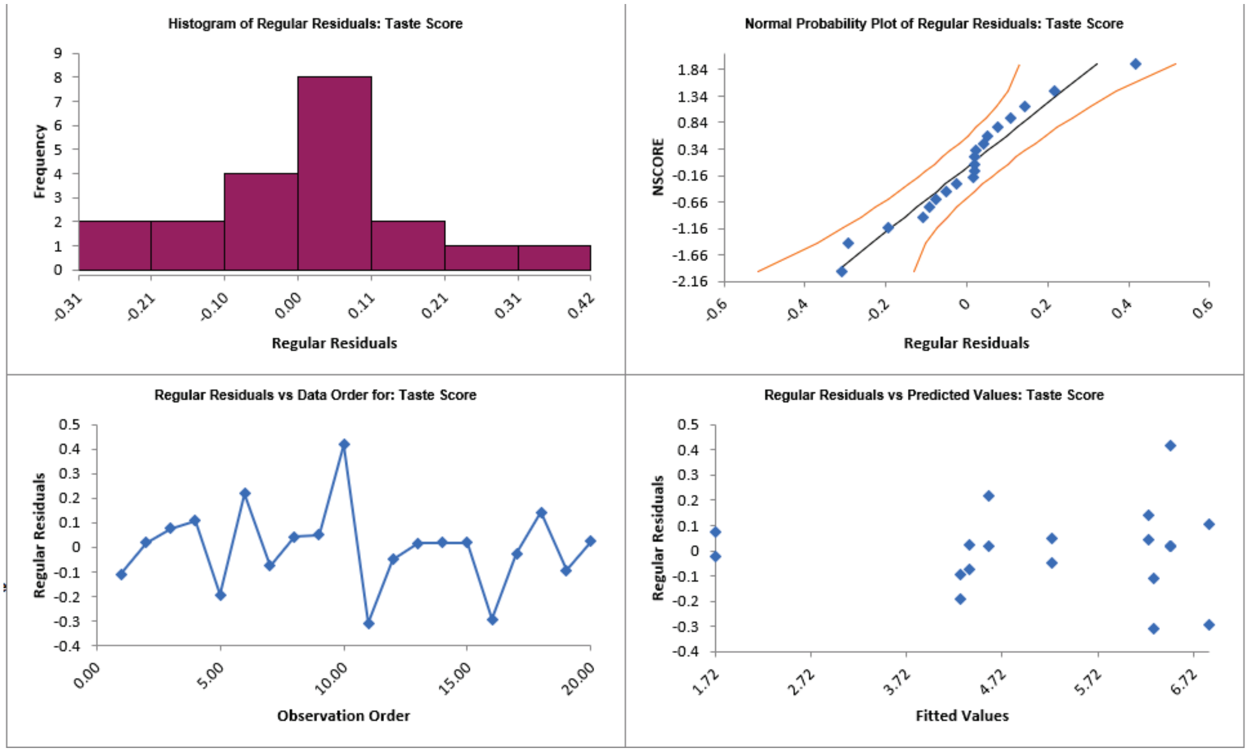
The report includes the test for *All Terms* and for individual predictors. *All Terms* denotes that all terms are in the model. This should be used to decide whether or not to take corrective action. The individual predictor terms are evaluated one-at-a-time and provide supplementary information for diagnostic purposes. Note, this should always be used in conjunction with an examination of the residual plots.

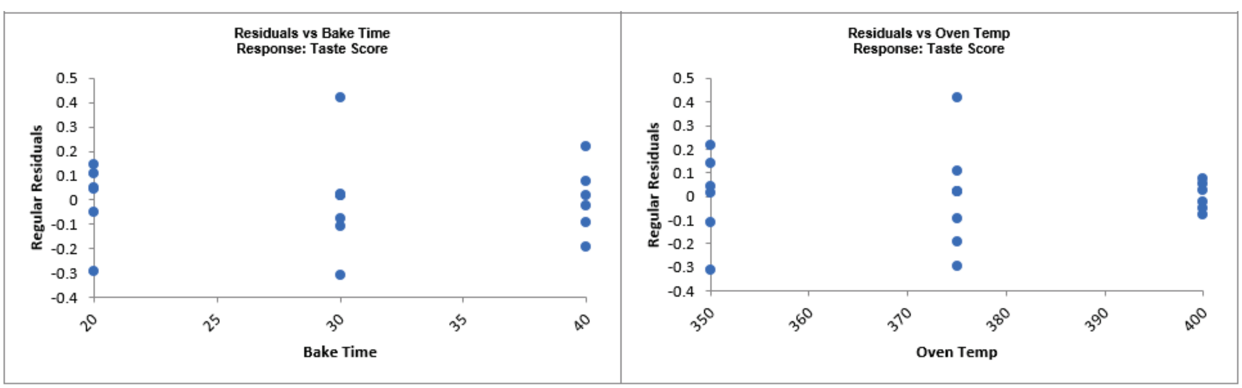
Here we see that the *All Terms* test is not significant, so we conclude that the variance is constant.

Tip: If the *All Terms* test is significant after model refinement, try a Box-Cox transformation. If that does not work, refit the model using **Recall Last Dialog,** click **Advanced Options** in the **Advanced Multiple Regression** dialog, and uncheck **Assume Constant Variance/No AC**. SigmaXL will apply the White robust standard errors for non-constant variance. For details, see the Appendix: **Advanced Multiple Regression**.

Tip: Lack of Constant Variance (a.k.a. Heteroskedasticity) is a nuisance for regression modelling but is also an opportunity. Examining the residual plots and individual predictors may yield process knowledge that identifies variance reduction opportunities.

9. Click on Sheet **MReg1 – Residuals** to view the Residual Plots. Note, Sheet MReg# will increment every time a model is refitted.



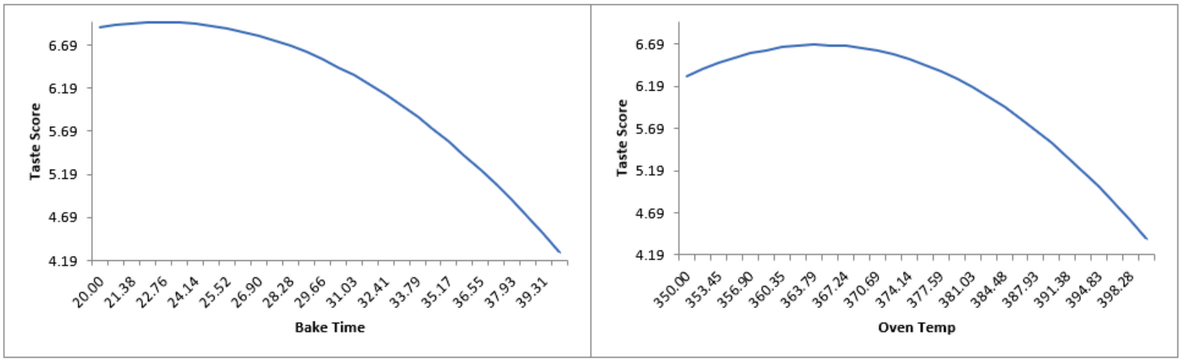


The residuals are approximately normal. There is a possible pattern in the Residuals vs Oven Temp with Oven Temp = 400 showing less variance, but the Breusch-Pagan test for constant variance for Oven Temp is not significant, so here we cannot conclude that a higher Oven Temp would result in a lower Taste Score variance.

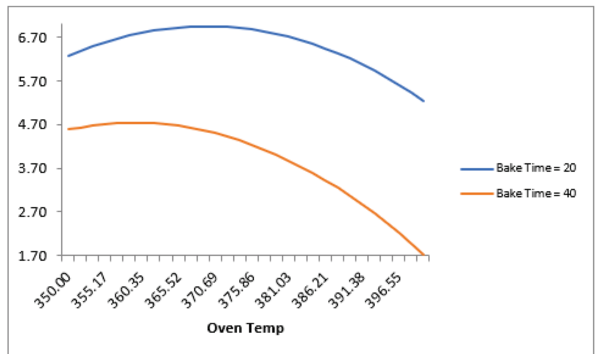
Note: Residuals versus interaction or quadratic terms are not plotted, but they can be manually created using the model design matrix to the right of the Residual Plots (use **SigmaXL > Graphical Tools > Scatter Plots**).

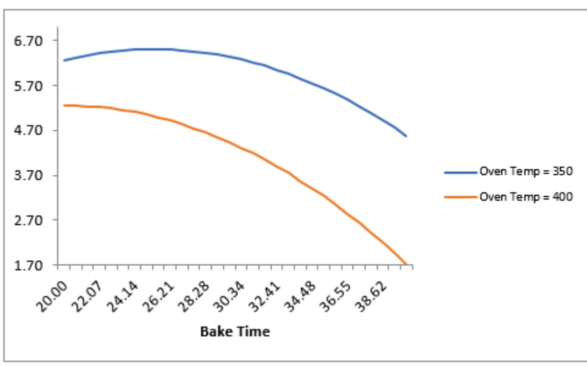
10. Click on Sheet **MReg1 – Plots**. The Main Effects Plots and Interaction Plots for Taste Score are shown.

Main Effects Plots for Taste Score - Fitted Means:



Interaction Plots for Taste Score - Fitted Means:





These are based on Fitted Means as predicted by the model, not Data Means as used in the previous analysis. Main Effects Plots with Fitted Means use the predicted value for the response versus input predictor value, while holding all other variables at their respective means. Similarly for Interaction Plots, all predictors not being plotted are held at their respective means.

The curvature is due to the quadratic terms in the model. These plots give us an initial look at the Fitted Means for Taste Score, but Contour and Surface Plots will be more useful for this Response Surface Design.

11. Click on Sheet **MReg1 – Model**. Scroll to the Predicted Response Calculator. Enter *Bake Time* = 23, *Oven Temp* = 368 to predict Taste Score with the 95% confidence interval for the long term mean and 95% prediction interval for individual values:

Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Bake Time	23	7.047089524	0.076367732	6.883297028	7.21088202	6.604826626	7.489352422
Oven Temp	368						

Note the formula at cell **L14** is an Excel formula. This matches the previous initial settings. Here the full predictor names are used making it easier to use and interpret. The Coded Settings are calculated as part of the Excel formula. Also, the prediction standard error **SE** is given.

12. Next, we will use SigmaXL's built in Optimizer. Scroll to view the Optimize Options:

Optimize Options			
Continuous Predictors	Lower Bound	Upper Bound	Integer
Bake Time	20	40	0
Oven Temp	350	400	0

Here we can constrain the lower and upper bounds of the continuous predictors, but we will leave the default settings as is, which are obtained from the minimum and maximum of the predictor values.

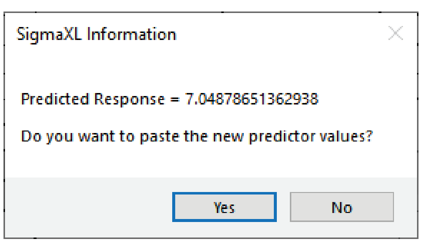
13. Scroll back to view the Goal setting and Optimize button. Specify Goal = *Maximize*.

Goal:	Maximize
Target:	

Optimize   Contour/Surface Plots

The optimizer uses Multistart Nelder-Mead Simplex to solve for the desired response goal with given constraints. For more information see the Appendix: Single Response Optimization.

14. Click **Optimize**. The response solution and prompt to paste values into the Input Settings of the Predicted Response Calculator is given:



15. Click **Yes** to paste the values.

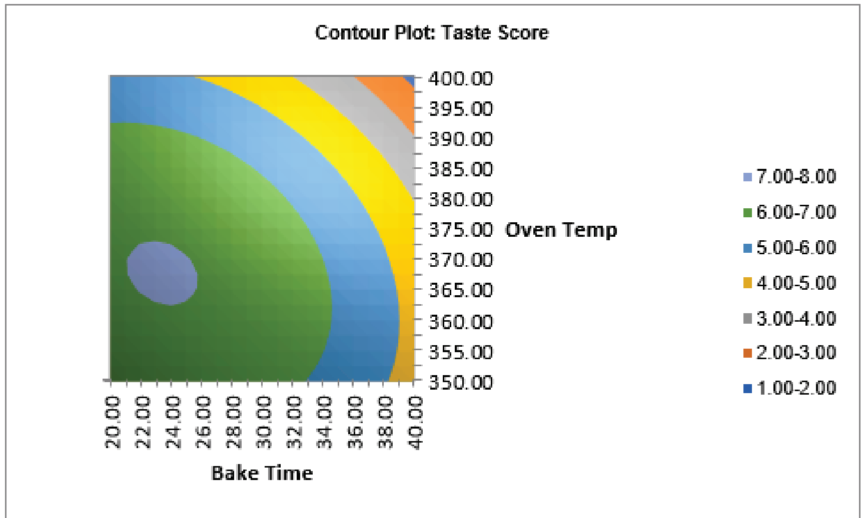
Predicted Response Calculator							
Predictors	Enter Settings:	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
Bake Time	23.4481406	7.048786514	0.075221866	6.887451657	7.21012137	6.607427901	7.490145126
Oven Temp	367.7186467						

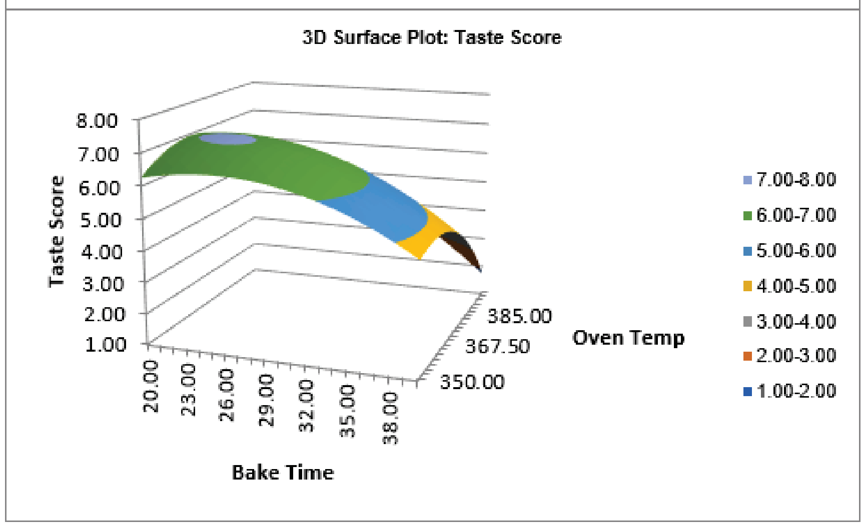
This matches the solution that was obtained using Solver in the previous analysis.

16. Next, we will create a Contour/Surface Plot. Click the **Contour/Surface Plots** button.

17. A new sheet is created, **MReg1 – Contour** that displays the plots:

Predictors	Hold Values
Bake Time	23.4481406
Oven Temp	367.7186467





This matches the Contour and Surface Plots given in the previous analysis.

The table with the **Hold Values**, gives the values used to hold a predictor constant if it is not in the plot, so is not applicable here with only one plot based on the two continuous predictors.

Tip: Use the contour/surface plots in conjunction with the predicted response calculator to determine optimal settings.

Part E – Basic Taguchi DOE Templates

Introduction – Taguchi Methods

Dr. Genichi Taguchi was a Japanese engineer and quality consultant who codified his consulting techniques into a formal methodology. This generic engineering tool, called the Taguchi method, is used for system studies. The purpose of these studies is to reduce system variability while simultaneously decreasing costs and increasing productivity.

The features of the Taguchi method include:

- extensive use of experimental design
- separation of factors by role (control and noise)
- use of measures of variability as responses
- dual objectives of process centering and noise minimization
- use of loss functions for economic justification of each application

The following steps outline the Taguchi Design of Experiments method:

1. Identify what are the control factors and what are the noise factors. Noise factors are costly and difficult to control, while an ideal control factor is easy to control precisely. Noise factors can include environmental factors, component deterioration, and process variation.
2. Identify the levels for the control and noise factors.
3. Identify the responses of interest and determine a quality criterion for each response. Does the system require that the response match a specified value, be as large as possible, or approach zero?
4. Construct a design for the control factors (inner array) and a design for the noise factors (outer array). The inner-array is selected based on the number of control factors. Typically, interactions are assumed to be negligible, but if they are to be included, then a larger design may be required. The outer array simulates the various conditions that the noise factors would produce in reality. For each run in the inner array, all runs in the outer array are carried out.
5. Conduct the design of experiments. Typically, you complete all the runs in the outer array before proceeding to the next run of the inner array. The runs for the inner array and outer array are (ideally) randomized separately.
6. Evaluate the performance statistic for each run of the inner array. These measures become the responses for the inner array. The performance statistics include Mean(Y), StDev(Y) or Ln(StDev), and Taguchi's signal-to-noise ratios. Some practitioners prefer to use Ln(StDev) as a variance stabilizing transformation. Taguchi signal-to-noise ratios include Nominal is Best, Larger is Better and Smaller is Better. Signal-to-noise ratios are always maximized.

7. From the Pareto of Deltas (Main Effects), Pareto of ANOVA SS (Sum-of-Squares) % Contribution, Main Effects & Interaction Plots and Prediction Equation model, determine the new set points for the control factors. Taguchi's two step optimization first finds control settings that maximize the SN Ratio (and/or minimize the StDev), then if available, factors that move the mean to Target without affecting the SN Ratio.
8. Confirm that the new settings meet target, dispersion and loss criterion with follow-up experimental runs.

Taguchi Orthogonal Arrays are a cookbook of designs that are similar to Full & Fractional-Factorial and Plackett-Burman designs. For example, the L4 design is 2-level, 4 runs with up to 3 factors; L8 is 2-level, 8 runs with up to 7 factors; L9 is 3-level, 9 runs with up to 4 factors. Taguchi designs use 2-level coded values of 1, 2 instead of the orthogonal coding -1, +1 and 3-level coded values of 1, 2, 3 instead of -1, 0, +1.

Interactions are typically assumed to be negligible compared to main effects, but some designs permit the analysis of all interactions or aliased interactions. Selection of aliased interactions is more economical than all interactions, but they should be used with caution. Process knowledge, engineering or theory are used to make the selection and assume that the chosen interaction is dominant and the others are negligible. Aliased interactions are often associated with the largest main effects. Confirmation runs should always be used to validate the model.

For further reading, see:

Basic:

Fowlkes, W.Y.; Creveling, C.M. (2006) *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*, Prentice Hall.

Ross, P.J. (1996) *Taguchi Techniques for Quality Engineering*, 2nd Edition, McGraw-Hill, New York, NY.

Roy, R.K. (2010) *A Primer on the Taguchi Method*, 2nd Edition, Society of Manufacturing Engineers, Dearborn, MI.

Advanced:

Taguchi, G.; Chowdhury, S.; Wu, Y. (2005) *Taguchi's Quality Engineering Handbook*, John Wiley, Hoboken, NJ.

Overview of Taguchi DOE Templates

The Taguchi DOE templates are similar to the other SigmaXL templates: simply enter the inputs and resulting outputs are produced immediately.

Click **SigmaXL > Design of Experiments > Basic Taguchi DOE Templates** or **SigmaXL > Templates and Calculators > Basic Taguchi DOE Templates** to access these templates:

- Taguchi L4 (2 Level)
    - Two-Factor (with Two-Way Interaction)
    - Three-Factor

- Taguchi L8 (2 Level)
    - Three-Factor (with Two-Way Interactions)
    - Four to Six-Factor (with Aliased Two-Way Interactions)
    - Seven-Factor

- Taguchi L9 (3 Level)
    - Two-Factor (with Two-Way Interaction)
    - Four-Factor

- Taguchi L12 (2 Level): Eleven Factor

- Taguchi L16 (2 Level)
    - Five-Factor (with Two-Way Interactions)
    - Eight to Fourteen-Factor (with Aliased Two-Way Interactions)
    - Fifteen-Factor

- Taguchi L18 (2/3 Level)
    - Three-Factor (with Two-Way Interactions)
    - Eight-Factor (with A*B Interaction)

- Taguchi L27 (3 Level)
    - Three-Factor (with Two-Way Interactions)
    - Thirteen-Factor

Note: SigmaXL Basic Taguchi DOE templates do not include L32 (2 Level), L36 (2/3 Level), L54 (2/3 Level), designs with more than 3 Levels, or a signal factor for dynamic characteristics.

Template Features:

- Levels are discrete categorical so may be numeric or text
- Fill in the blanks template, charts automatically update
- Predicted Response Calculator and Charts for Mean, Standard Deviation (or Ln Standard Deviation) and Signal-to-Noise Ratio
- Available Signal-to-Noise Ratios:
    - Nominal is Best
    - Nominal is Best (Variance Only)
    - Nominal is Best (Mean Square Deviation with Target)
    - Larger is Better
    - Smaller is Better

Signal-to-Noise Ratios	Formula
SN: Nominal is Best	10*Log10(Ybar^2/S^2)
SN: Nominal is Best (Variance Only)	-10*Log10(S^2)
SN: Nominal is Best (MSD with Target)	-10*Log10(Sum((Y-T)^2)/n)
SN: Larger is Better	-10*Log10(Sum(1/Y^2)/n)
SN: Smaller is Better	-10*Log10(Sum(Y^2)/n)

- Up to 27 Replications for Outer Array (i.e., support up to L27 Outer Array)
- Pareto of Deltas (Effects) and ANOVA SS (Sum-of-Squares) % Contribution (for Main Effects and Two-Way Interactions)
- Main Effects Plot and Interaction Plots (if applicable)
- For designs with aliased interactions a drop-down list of available aliased interactions is provided. This is much easier to use than Linear Graphs.
- Column assignments to Orthogonal Array are optimized to ensure maximum design resolution.

Template Notes:

1. Select desired Signal-to-Noise Ratio to maximize. The SN formula is displayed.
2. For Nominal is Best, "SN: Nominal is Best" is recommended for non-negative data. Use "SN: Nominal is Best (Variance Only)" if the data has a mixture of negative and positive values.
3. If selection is "SN: Nominal is Best MSD with Target", enter Target value (MSD denotes Mean Square Deviation); if Target = 0, this is equivalent to Smaller Is Better. For use of MSD in SN Ratio, see Ranjit Roy (2010) "A Primer on the Taguchi Method," Second Edition, Society of Manufacturing Engineers.
4. Larger is Better requires positive data. Smaller is Better requires non-negative data (target is zero).
5. Select desired Standard Deviation Response: StDev(Y) or Ln(StDev(Y)).
6. Enter Factor Names and Factor Levels. Levels are discrete categorical, so may be numeric or text.
7. If applicable, select Aliased Interactions. Enter Run Number (if runs have been randomized) and Outer Array Response values.
8. Selection of Aliased Interaction assumes that the chosen interaction is dominant and the others are negligible. It is often associated with the largest main effects. Confirmation runs should be used to validate model. Please refresh selection if message appears.
9. For Taguchi L8 Five to Six Factors and L16 Nine to Fourteen Factors, not all possible interactions are available in the drop-down list (they are aliased with main effects). If an interaction of interest is not available in any of the drop-down lists, please modify the Factor Names so that they are assigned to columns shown in the list and then select that interaction.
10. The typical default display of Yrep1 to Yrep9 accommodates an outer array up to L9. Unhide columns to display Yrep10 to Yrep27, which permits an outer array up to L27.
11. Delta is the difference between the maximum and minimum average response for the levels of each factor in the inner array. Two-way interactions are not included. ANOVA SS (Sum-of-Squares) % Contribution includes two-way interactions.
12. To compute a predicted response, enter Actual Factor Setting using the drop-down selection. Factors are treated as categorical. If Factor Setting Coded is "FALSE", please refresh selection to match level settings. Note, Excel Solver cannot be used with this calculator.
13. The Interaction Plot X-axis/Legend factors may be switched by clicking on the chart, then select Excel Design > Switch Row/Column.
14. DOE Templates are protected worksheets by default, but this may be modified by clicking **SigmaXL > Help > Unprotect Worksheet**. Alternatively, you can click Excel **File > Info > Unprotect** or **Home > Format > Unprotect Sheet**.

15. Chart Y-axis scaling is automatic. To modify, double-click on the Y axis and adjust Minimum and Maximum.
16. The Orthogonal Array or Dummy Coding may be analyzed using Multiple Regression Analysis, adding and removing terms from the model as necessary. Terms that are removed will be used to estimate error for p-values. In particular, regression analysis of 3 Level designs with interactions (L9 Two Factor, L18 Three Factor and L27 Three Factor) should use Dummy Coding.
17. Dummy Coding uses Level 1 as the reference value, so Level 1 does not appear in the array. The Predicted Output calculator formula coefficients and ANOVA Sum-of-Squares (SS) % Contribution are computed using Dummy Coding regression. This is consistent with SigmaXL's use of Dummy Coding regression for categorical factors in other tools.
18. Do not modify any other part of the template.
19. Caution: The use of Cell Autocomplete with Flash Fill in Excel 2016 or higher may result in a crash (this is a known Microsoft issue). If this occurs, please turn off Automatic Flash Fill (File > Options > Advanced, uncheck "Automatically Flash Fill" under "Enable Autocomplete for cell value").
20. If there are further stability issues, please turn off Automatic Recalculate (Formulas > Calculation Options, Select Manual). Use "Calculate Now" to refresh calculations after entry of Factor Names, Levels, Outer Array data or as needed.

Taguchi L8 (2 Level) Three Factor – Robust Cake Example

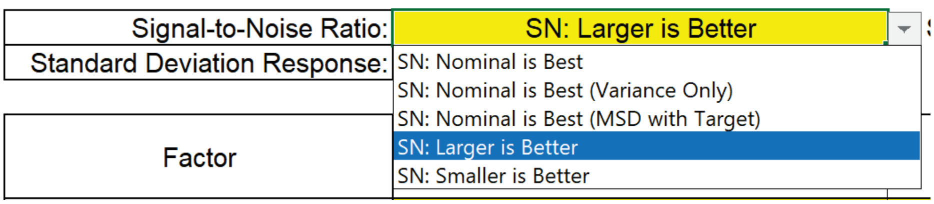
This is the same Robust Cake Experiment adapted from the Video *Designing Industrial Experiments*, by Box, Bisgaard and Fung, analyzed earlier using a Three Factor Full-Factorial. Here we will use the Taguchi L8 Three Factor.

1. Click **SigmaXL > Help > Template Examples > Taguchi Examples > L8 Three Factor – Robust Cake** to open the example file. If you wish to start with a blank template and populate the values, click **SigmaXL > Design of Experiments > Basic Taguchi DOE Templates > Taguchi L8 (2 Level) > Three Factor (with Two-Way Interactions)**.

2. The response is Taste Score (on a scale of 1-7 where 1 is "awful" and 7 is "delicious").

3. The five **Outer Array** values (**Yrep1** to **Yrep5**) have different Cooking Time and Temperature Conditions. Note, the outer array was a two-factor, full-factorial plus center point, hence 5 replications, so was not a typical Taguchi outer array.

4. The goal is to maximize the Taste Score (with minimum variation), so we will maximize the Signal-to-Noise Ratio **SN:Larger is Better**, selected from the list as shown:



Signal-to-Noise Ratio:	SN: Larger is Better	SN Formula: -10*Log10(Sum(1/Y^2)/n)
Standard Deviation Response:	StDev (Y)	

5. The **Inner Array** controllable factors are Flour, Butter and Egg. Actual low and high settings are not given in the video, so we will use coded 1 and 2 values for **Level 1** and **Level 2**. Note that the Taguchi L8 standard sort order (left to right) is different from Full-Factorial (right to left), so we will enter **Factor Name** as Egg, Butter, Flour to adjust for this difference. (In the Full-Factorial it was Flour, Butter, Egg).

Factor	Factor Name	Level 1	Level 2
A	Egg	1	2
B	Butter	1	2
C	Flour	1	2

6. We are looking for a combination of Egg, Butter, and Flour that will not only taste good, but consistently taste good over a wide range of Cooking Time and Temperature conditions.

7. The runs were randomized and order recorded in the column **Run No.** The **Outer Array** Taste Score values are entered as shown:

Inner Array:

Std. Order	Run No.	Egg	Butter	Flour
1	6	1	1	1
2	2	1	1	2
3	7	1	2	1
4	1	1	2	2
5	5	2	1	1
6	8	2	1	2
7	3	2	2	1
8	4	2	2	2

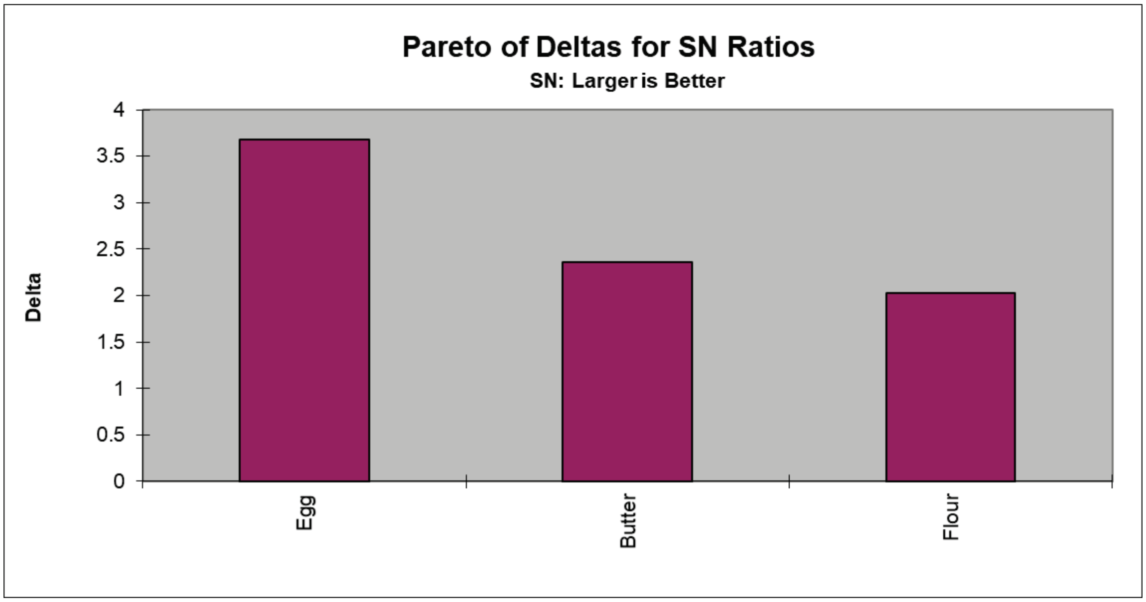
Outer Array:

Yrep1	Yrep2	Yrep3	Yrep4	Yrep5
3.1	1.1	5.7	6.4	1.3
3.2	3.8	4.9	4.3	1.3
5.3	3.7	5.1	6.7	2.9
4.1	4.5	6.4	5.8	5.2
5.9	4.2	6.8	6.5	3.5
6.9	5	6	5.9	5.7
3	3.1	6.3	6.4	3
4.5	3.9	5.5	5	5.4

Scroll to the right to view the calculated statistics:

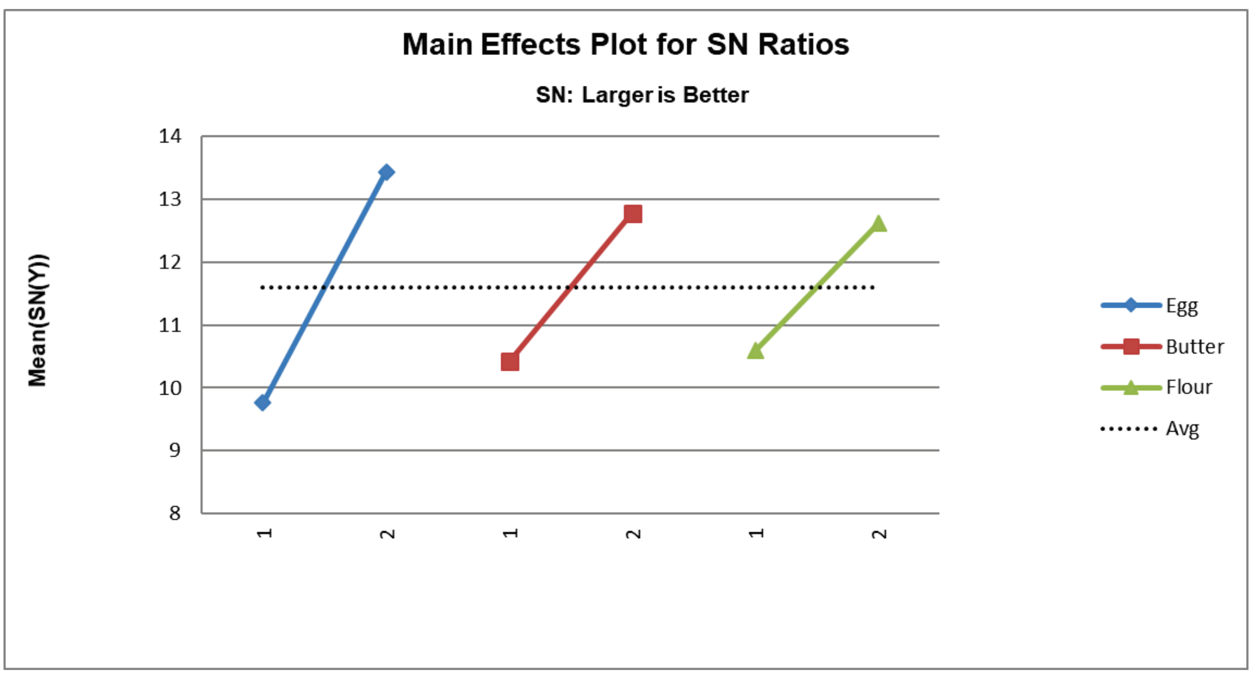
Mean (Y)	StDev (Y)	SN: Larger is Better
3.52	2.449897957	5.010245078
3.5	1.380217374	7.673306706
4.74	1.479188967	12.39162469
5.2	0.935414347	13.97953162
5.38	1.454991409	13.71929378
5.9	0.681909085	15.27864445
4.36	1.817415748	11.23909439
4.86	0.665582452	13.51678365

8. We will focus our analysis on the Signal-To-Noise Ratio and review the Mean and StDev afterward as a check on the results. Scroll right and down to view the **Pareto of Deltas for SN Ratios**:

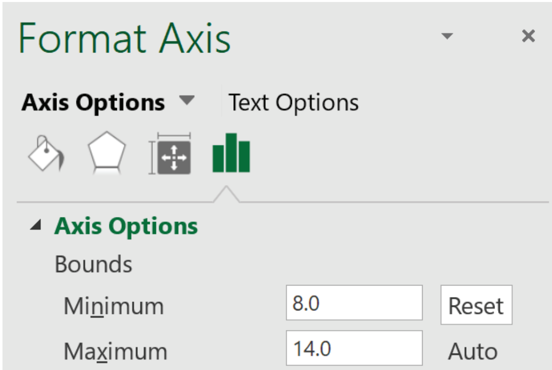


Delta is the difference between the maximum and minimum average response for the levels of each factor in the inner array. It appears that Egg is the dominant factor followed by Butter.

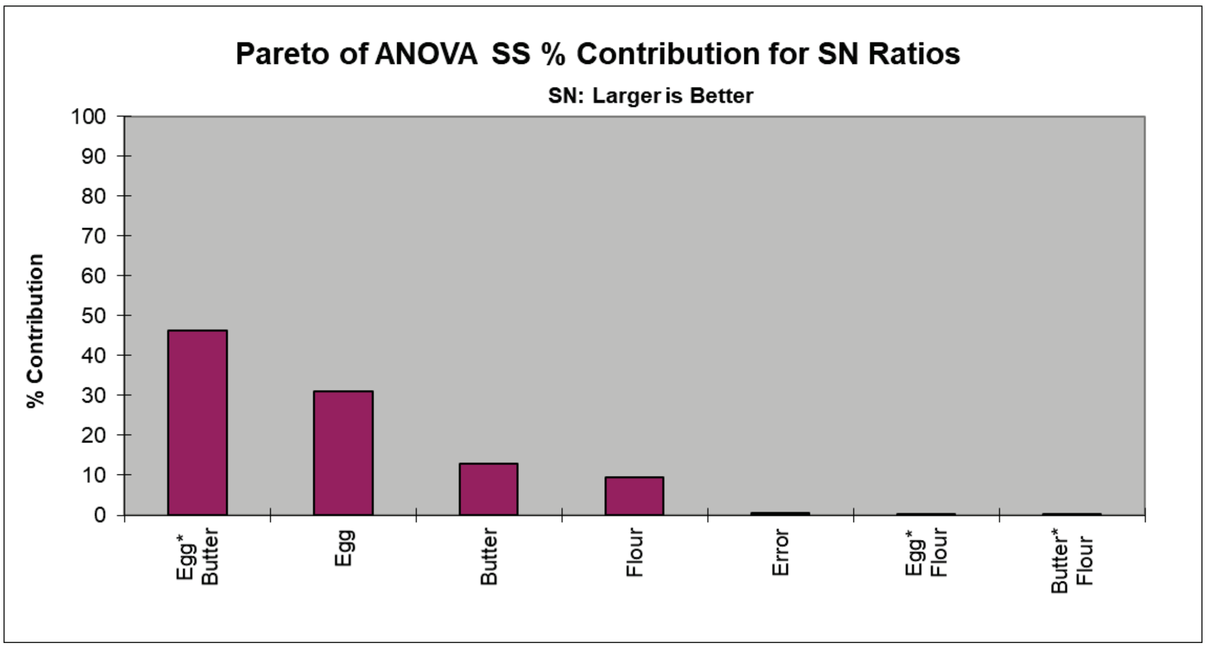
9. Scroll down to look at the **Main Effects Plot for SN Ratios**:



It initially appears that the maximum Taste Score occurs with Egg = 2, Butter = 2 and Flour = 2. Note: The Y Axis has been adjusted to get a better scale (double click on the Y-Axis to open Excel Format Axis dialog, click **Axis Options** and set **Minimum** = 8 as shown).



10. However, this design permits the analysis of interactions, so we will now look at the **Pareto of ANOVA SS (Sum-of-Squares) % Contribution for SN Ratios** which includes Main Effects and Interactions:

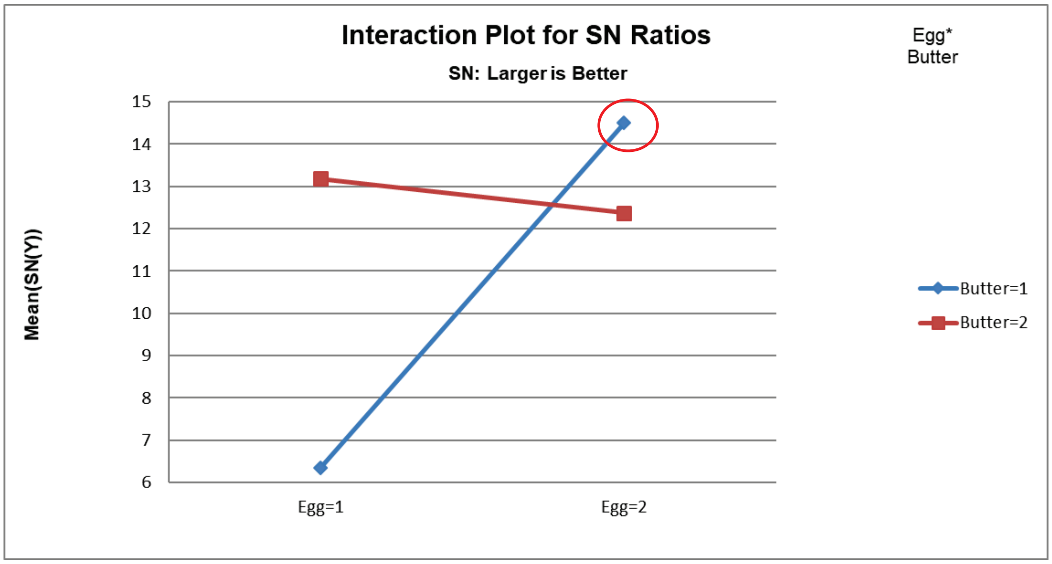


11. Now we see that the Egg*Butter interaction is clearly the dominant factor (in agreement with the analysis done earlier using the Three Factor Full-Factorial).

    Note 1: The "Error" term in this model is the three-way interaction which is assumed to be negligible, so is not included in the model.

    Note 2: SigmaXL does not compute a significance line in the Taguchi Templates, as done in the Three Factor Full-Factorial Template. Taguchi recommends pooling of effects to estimate experimental error which will be demonstrated later.

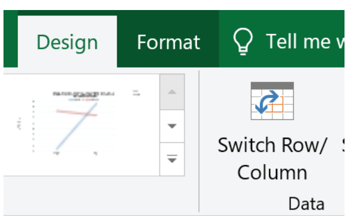
12. Scroll down to view the **Interaction Plot for SN Ratios** for the Egg*Butter Interaction:

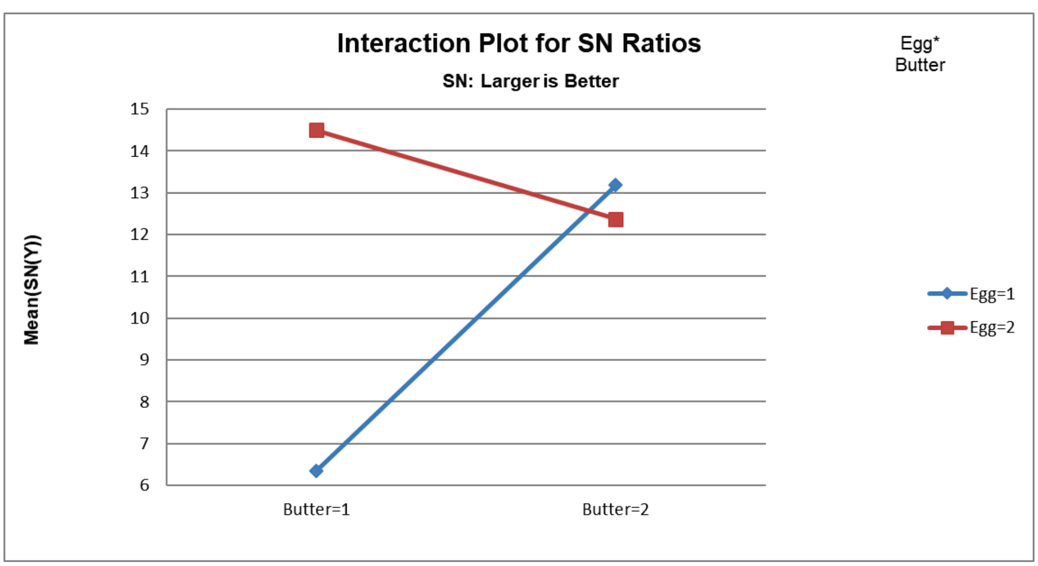


Now we can clearly see the large effect due to the two-way interaction. The effect that Egg has on Taste depends on what the Butter setting is. The maximum SN Ratio for Taste Score is achieved with Egg = 2 and Butter = 1.

Note 1: The Y Axis has been adjusted to improve visibility (**Minimum** = 6).

Note 2: The Interaction Plot X-axis/Legend factors may be switched by clicking on the chart, then selecting Excel **Design > Switch Row/Column** as shown:





13. From the above plots, we conclude that the optimum setting to produce a maximum Taste Score SN Ratio is: Egg = 2, Butter = 1 and Flour = 2. (If we had only considered the Main Effects plot, we would have incorrectly set Butter = 2).

14. Scroll up and left to enter these values in the Predicted Output Calculator. Factor settings can be manually entered or selected from the drop-down list as shown:

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded
Egg	2	2
Butter	1	1
Flour	2	2

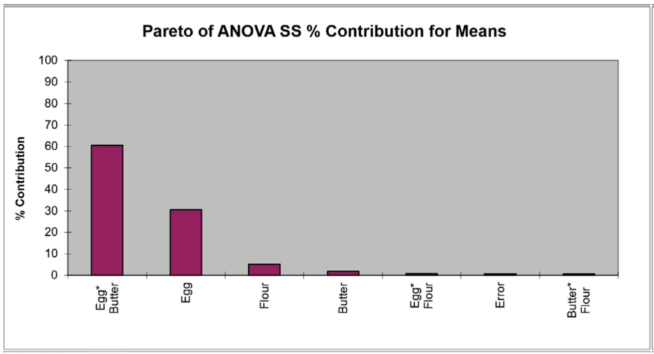
Predicted Output for Y:

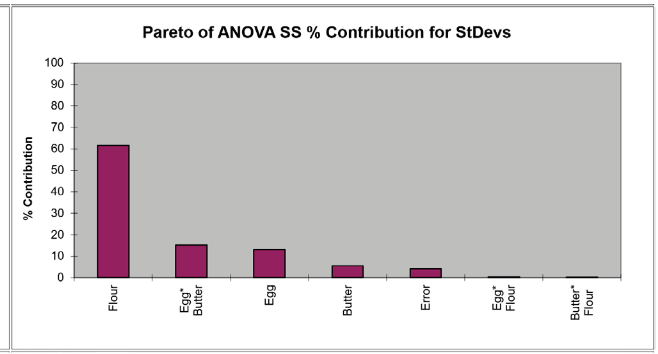
Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded	$\hat{Y}$	$\hat{S}$	SN: Larger is Better
Egg	2	2	5.8375	0.568826968	15.50283111
Butter	1	1			
Flour	2	2			

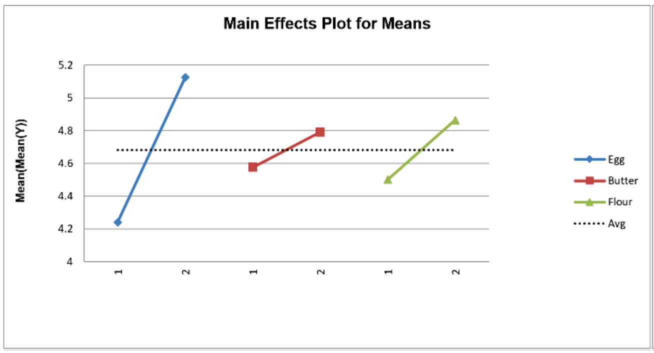
This gives the predicted Mean, Standard Deviation and SN Ratio for Taste Score and agrees with the results obtained in Part B, the analysis of the Three-Factor Full Factorial. (The predicted values are slightly different because the Full Factorial includes the three-way interaction term in the model, whereas this is considered as an error term in the Taguchi Template).

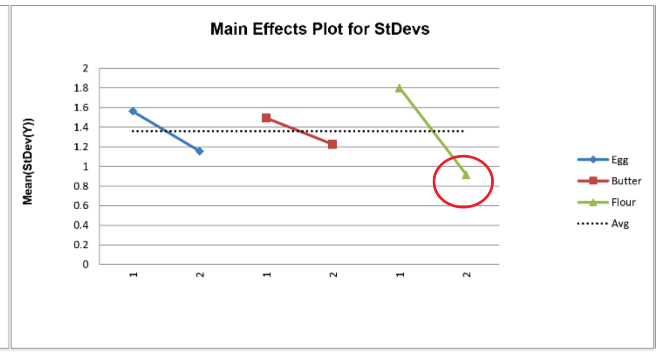
Note: This calculator treats the factors as categorical, so levels can be numeric or text. The drop-down list is a convenient way to select the factor levels, but unfortunately Excel Solver cannot be used with this calculator to perform optimization. We will demonstrate the use of Solver with Multiple Regression later.

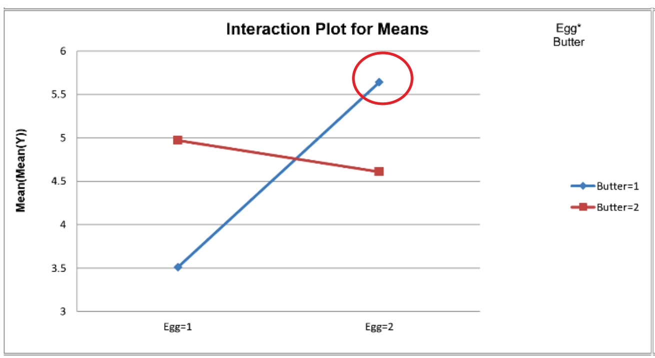
15. As a validation of the SN Ratio analysis, we will review the Plots for Mean and Standard Deviation:

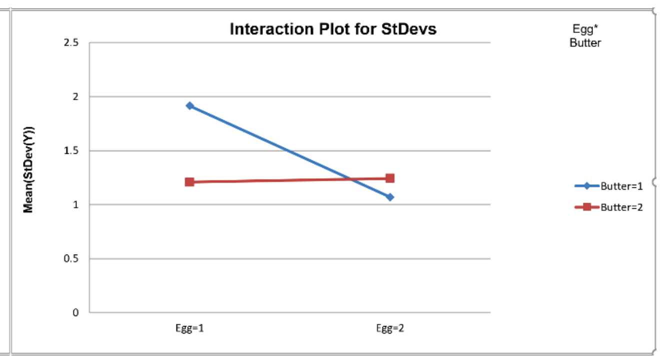












From the **Pareto of ANOVA SS % Contribution for Means**, we see that the Egg*Butter interaction is the dominant contributor to Mean Taste Score. Looking at the Egg*Butter **Interaction Plot for Means**, the optimum setting to maximize Mean Taste Score is Egg = 2 and Butter = 1. (Note: The Y Axis has been adjusted).

Looking at the **Pareto of ANOVA SS % Contribution for StDevs**, Flour is the dominant factor for Standard Deviation. The **Main Effects Plot for StDevs** shows that the optimum setting to minimize the StDev Taste Score is Flour = 2.

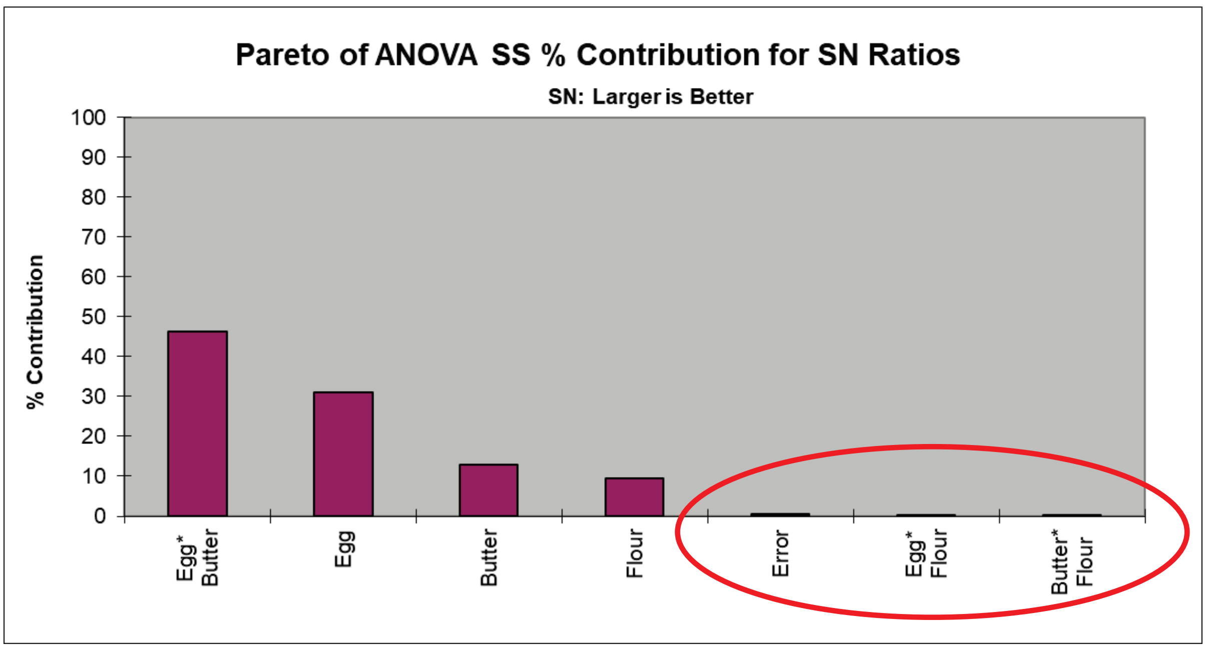
This is in agreement with the settings determined from the SN Ratio. Taguchi advocates the use of SN Ratio as it is simpler, looking at only one variable, but some practitioners prefer to analyze the Mean and Standard Deviation separately.

Multiple Regression and Excel Solver (Advanced Topics):

16. The Taguchi L8 Orthogonal Array (or L8 Dummy Coding) may be analyzed using Multiple Regression Analysis, adding and removing terms from the model as necessary. Terms that are removed will be used to estimate error for p-values. As noted earlier, SigmaXL does not compute a significance line in the Taguchi Template Pareto Charts. Taguchi recommends pooling of effects to estimate experimental error and this will be demonstrated here. Also, Solver cannot be used with the Predicted Output calculator in the Taguchi Template, but will be used in the Multiple Regression report.

17. We will be analyzing the SN: Larger is Better Ratio as the response in this example. The analysis may be repeated separately for the Mean and StDev.

18. From the **Pareto of ANOVA SS % Contribution for SN Ratios**, we see that the Error term (three-way interaction), Egg*Flour and Butter*Flour interactions are negligible compared to the other factors (each less than 1% contribution). As an initial attempt at pooling we will remove these terms from the regression model.



19. Highlight the L8 Orthogonal Array cells (**AL18:AV26**) as shown.

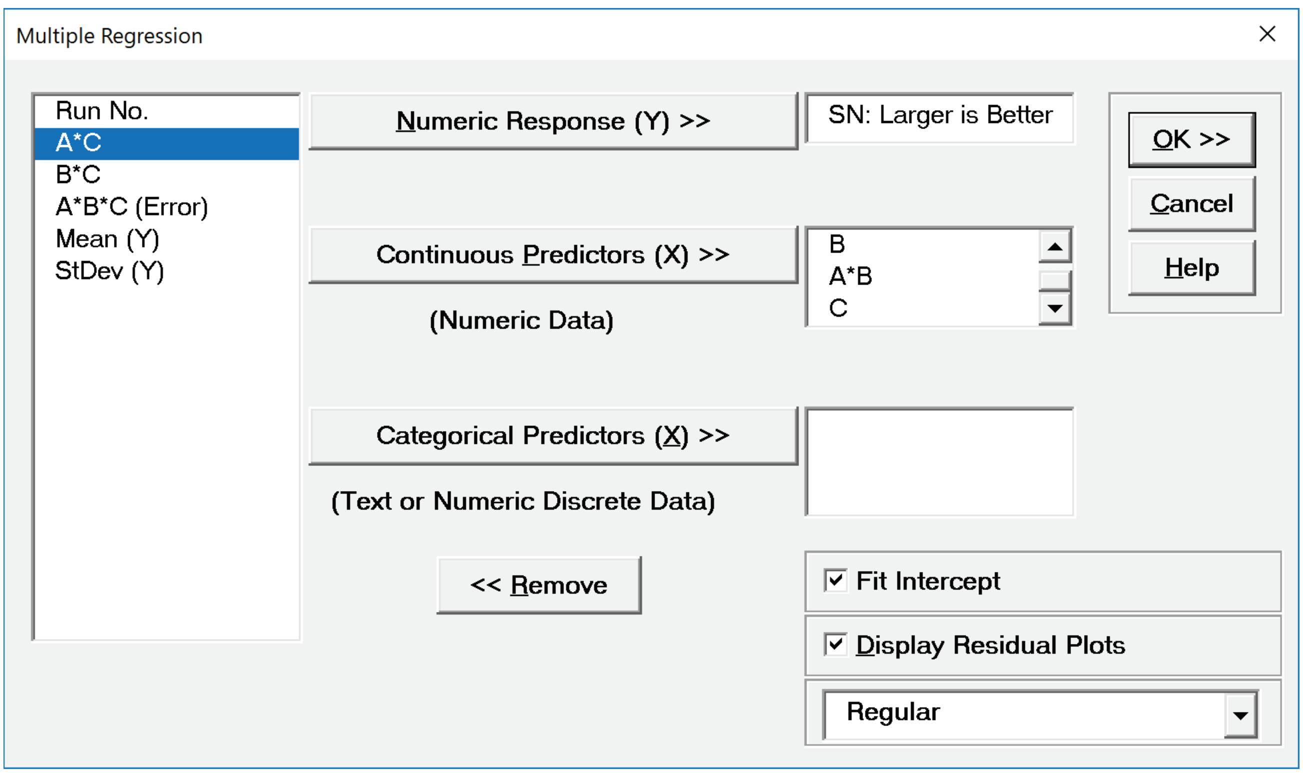
L8 Orthogonal Array:

	L8_1	L8_2	L8_3	L8_4	L8_5	L8_6	L8_7			
Run No.	A	B	A*B	C	A*C	B*C	A*B*C (Error)	Mean (Y)	StDev (Y)	SN: Larger is Better
6	1	1	1	1	1	1	1	3.52	2.449897957	5.010245078
2	1	1	1	2	2	2	2	3.5	1.380217374	7.673306706
7	1	2	2	1	1	2	2	4.74	1.479188967	12.39162469
1	1	2	2	2	2	1	1	5.2	0.935414347	13.97953162
5	2	1	2	1	2	1	2	5.38	1.454991409	13.71929378
8	2	1	2	2	1	2	1	5.9	0.681909085	15.27864445
3	2	2	1	1	2	2	1	4.36	1.817415748	11.23909439
4	2	2	1	2	1	1	2	4.86	0.665582452	13.51678365

The Orthogonal Array displays the factor letters, not the names, so recall that A is Egg, B is Butter and C is Flour.

20. Click **SigmaXL > Statistical Tools > Regression > Multiple Regression**. Click **Next**.

21. Select *SN: Larger is Better*, click **Numeric Response (Y) >>**; holding the CTRL key, select *A, B, A*B, and C*, click **Continuous Predictors (X)** >> as shown:



22. Click **OK**. The resulting regression report is shown:

Multiple Regression Model: SN: Larger is Better = (-7.210) + (3.675) * A + (2.361) * B + (4.482) * A*B + (2.022) * C

Model Summary:

R-Square	99.49%
R-Square Adjusted	98.82%
S (Root Mean Square Error)	0.382714

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	-7.210	0.823057	-8.760	0.0031		
A	3.675	0.270619	13.579	0.0009	1	1.00000
B	2.361	0.270619	8.726	0.0032	1	1.00000
A*B	4.482	0.270619	16.564	0.0005	1	1.00000
C	2.022	0.270619	7.472	0.0050	1	1.00000

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	4	86.521	21.630	147.68	0.0009
Error	3	0.439409	0.146470		
Total (Model + Error)	7	86.961	12.423		

Durbin-Watson Test for Autocorrelation in Residuals:

DW Statistic	2.506
P-Value Positive Autocorrelation	0.5149
P-Value Negative Autocorrelation	0.0000

All terms are significant (using the pooled estimate of error) so we will use the model as is.

23. Now we will use the **Predicted Response Calculator** to determine the factor settings that maximize the SN Ratio:

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A						
B						
A*B						
C						

24. All entries must be in coded units of 1 to 2 as given in the Orthogonal Array. While the calculator in the Taguchi Template uses strictly categorical factors, here we can use either categorical or continuous.

25. Since the Taguchi Orthogonal Array is coded 1, 2 rather than -1, +1, we cannot simply multiply Factor A * Factor B to compute the A*B interaction.

26. Turn on the Excel Row and Column headers in the Multiple Regression sheet: **File > Options > Advanced**. Scroll down to **Display options for the worksheet**. Check **Show row and column headers**. Click **OK**.

27. In cell A*B (**K14)**, enter the formula:

    =1.5 - 2*(K12-1.5)*(K13-1.5)

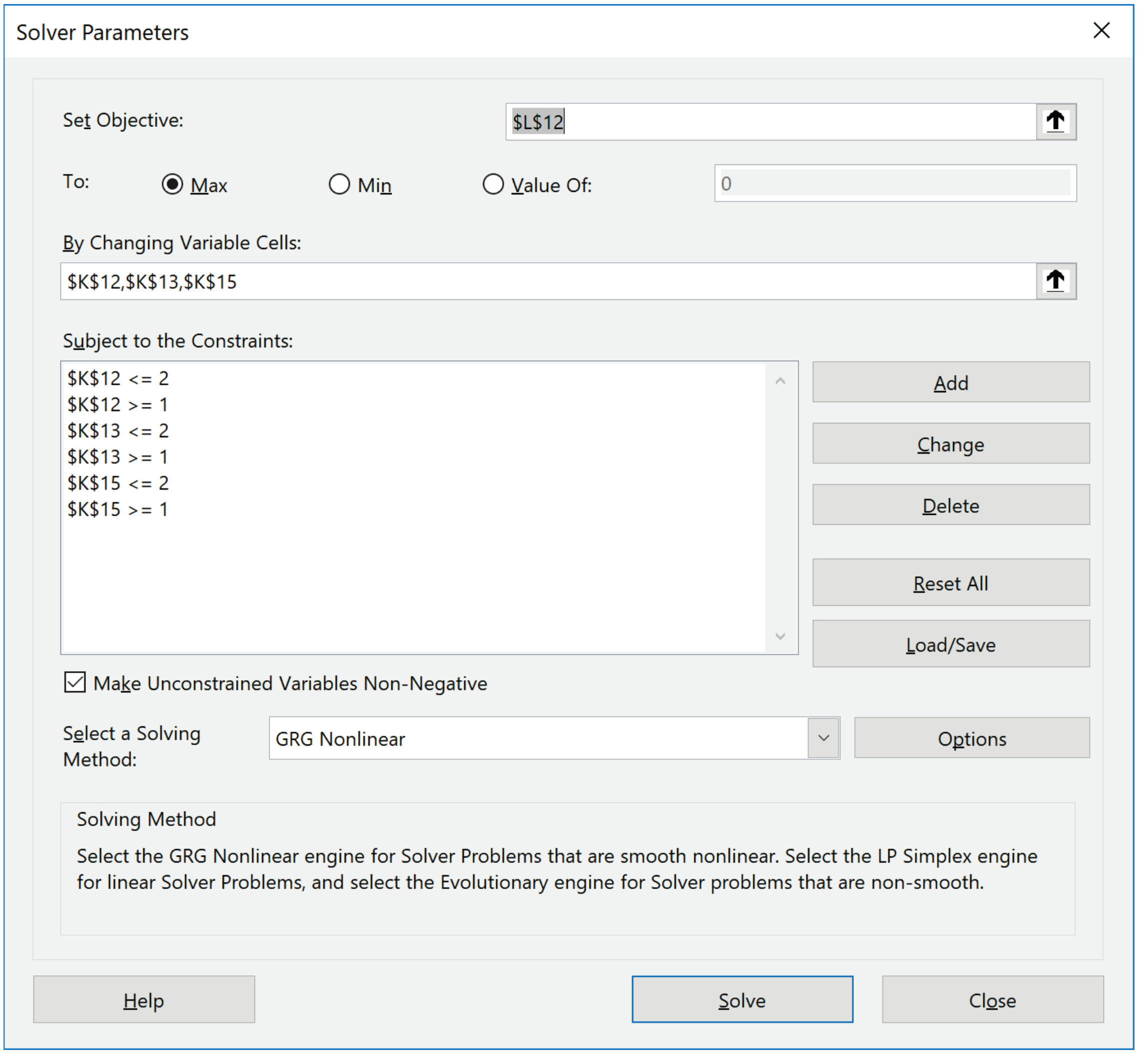
    This formula centers and scales the A (**K12**) and B (**K13**) values so that the A*B (**K14**) value will match the values in the Taguchi array. Enter A = 1, B = 1, C =1 (be careful to not overwrite the formula – you may wish to change the **K14** cell color as shown). The A*B value is 1.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A	1	5.330774831	4.367888571	6.29366109	3.778167389	6.883382272
B	1					
A*B	1					
C	1					

28. Optionally, test the formula entry, enter A=1, B=2, confirming that A*B = 2. Repeat for A=2, B=1 resulting in A*B = 2 and A=2, B=2 resulting in A*B = 1. Now reset back to A=1 and B=1.

29. Click Excel **Data > Solver** to open the Solver dialog (if the Solver menu option does not appear, click **File > Options > Add-Ins**. Adjacent to **Manage Excel Add-ins**, click **Go**. Check **Solver Add-in**. Click **OK**). Set the **Solver Parameters** as shown:

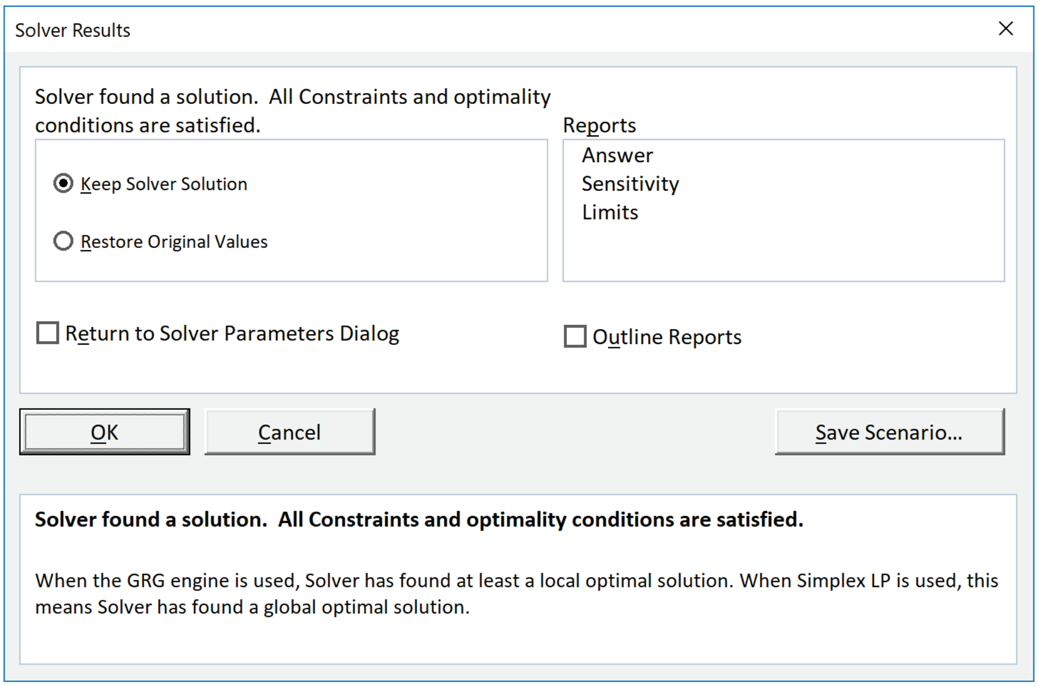


Cell **L12** is the predicted response (SN Ratio). Solver will try to maximize this value. Cells **K12**, **K13**, and **K15** are the A, B, and C Coded Factor Settings to be changed, with the constraints that the values are >=1 and <= 2. GRG Nonlinear is used here. (Note that if any of the factors are modeled as categorical, Evolutionary is recommended). Click **Solve**.

30. The Solver solution that maximizes SN Ratio is given in the **Predicted Response Calculator** as A (Egg) = 2, B (Butter) = 1 and C (Flour) = 2. These are in agreement with our earlier manual analysis. Confidence Intervals for the Mean SN Ratio and Prediction Intervals for Individual SN Ratio values are also reported.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A	2	15.50997018	14.54708392	16.47285643	13.95736273	17.06257762
B	1					
A*B	2					
C	2					



31. Click **OK** to keep the Solver solution.

32. This analysis may be repeated using the Mean and Standard Deviation as responses. A note of caution when using the Prediction Interval for the Mean, the individual values are averages across the Outer Array, not true individual values.

Taguchi L8 (2 Level) Four Factor – Catapult Example

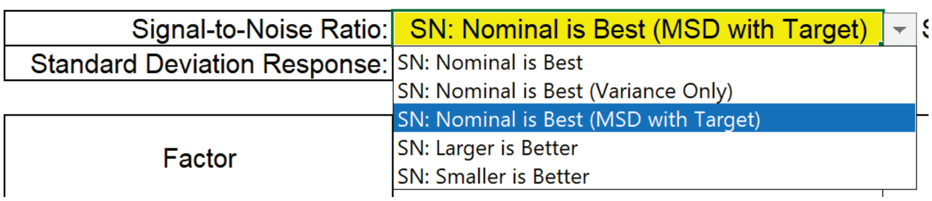
This Catapult experiment is adapted from the article "THE CATAPULT PROBLEM: ENHANCED ENGINEERING MODELING USING EXPERIMENTAL DESIGN", by Schubert et al., Quality Engineering, 4(4), 463-473 (1992).

1. Click **SigmaXL > Help > Template Examples > Taguchi Examples > L8 Four Factor – Catapult** to open the example file. If you wish to start with a blank template and populate the values, click **SigmaXL > Design of Experiments > Basic Taguchi DOE Templates > Taguchi L8 (2 Level) > Four Factor (with Aliased Two-Way Interactions)**.

2. The response is Distance in Inches with Target = 50".

3. The three **Outer Array** (**Yrep1** to **Yrep3**) values are repeat shots fired.

4. The goal is to hit a specific target of 50 inches with minimum variation, so we will maximize the Signal-to-Noise Ratio **SN: Nominal is Best (MSD with Target)**, selected from the list as shown:



Signal-to-Noise Ratio:	SN: Nominal is Best (MSD with Target)	SN Formula: -10*Log10(Sum((Y-T)^2)/n)
Standard Deviation Response:	StDev (Y)	

5. Enter the Target value 50 in cell **H8**:

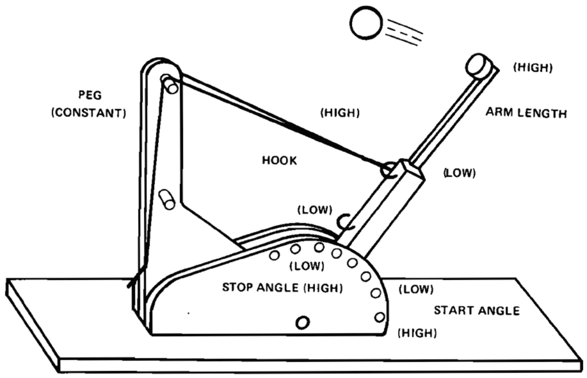
MSD Target:	50

Note: If Target = 0, Nominal is Best (MSD with Target) is equivalent to Smaller Is Better. For use of MSD in SN Ratio, see Ranjit Roy (2010) *A Primer on the Taguchi Method*, Second Edition, Society of Manufacturing Engineers.

6. The **Inner Array** controllable factors used in this study are: Hook (Bottom/Top), Arm Length (Short/Long), Start Angle (150/177) and Stop Angle (Hole 4/Hole 5).

Factor	Factor Name	Level 1	Level 2
A	Hook	Bottom	Top
B	Arm Length	Short	Long
C	Start Angle	150	177
D	Stop Angle	Hole 4	Hole 5

Note that these are different factors and factor levels than those used earlier in the Full Factorial Catapult experiment (Part C), so the results cannot be compared. In that experiment we used Pull Back Angle, Stop Pin and Pin Height. Here, Pin Height is kept fixed so not included, but Hook position and Arm Length are used. (In typical Catapult designs used in training today, Arm Length is adjusted by varying the Cup position).



7. Schubert used a Three Factor 8 Run Resolution IV design. We will use the Taguchi L8 Four Factor design. The column assignments in the SigmaXL Taguchi Templates are chosen to maximize the design resolution, so the L8 Four Factor Template is also Resolution IV, with main effects free and clear of two-way interactions and two-way interactions aliased with each other. The sort order used in this paper matches the Taguchi L8 standard sort order (left to right) so factor assignments remain the same. The paper uses the conventional -1 to +1 coding, whereas we will use the Taguchi 1, 2 coding.

8. Select Aliased Interactions using the drop-down list. Process knowledge, engineering or theory are used to make the selection and assume that the chosen interaction is dominant and the others are negligible. Aliased interactions are often associated with the largest main effects. Confirmation runs should always be used to validate the model. In this study, engineering knowledge was used to select all interactions that included the Stop Angle:

Aliased Interaction 1	Start Angle*Stop Angle
Aliased Interaction 2	Hook*Arm Length
Aliased Interaction 3	Start Angle*Stop Angle

Aliased Interaction 1	Start Angle*Stop Angle
Aliased Interaction 2	Arm Length*Stop Angle
Aliased Interaction 3	Hook*Stop Angle

Tip: For Taguchi L8 Five to Six Factors and L16 Nine to Fourteen Factors, not all possible interactions are available in the drop-down list (they are aliased with main effects). If an interaction of interest is not available in any of the drop-down lists, please modify the Factor Names so that they are assigned to columns shown in the list and then select that interaction.

9. A randomized run order was not given in the paper. The **Outer Array** Distance values are entered as shown:

Inner Array:

Std. Order	Run No.	Hook	Arm Length	Start Angle	Stop Angle
1	1	Bottom	Short	150	Hole 4
2	2	Bottom	Short	177	Hole 5
3	3	Bottom	Long	150	Hole 5
4	4	Bottom	Long	177	Hole 4
5	5	Top	Short	150	Hole 5
6	6	Top	Short	177	Hole 4
7	7	Top	Long	150	Hole 4
8	8	Top	Long	177	Hole 5

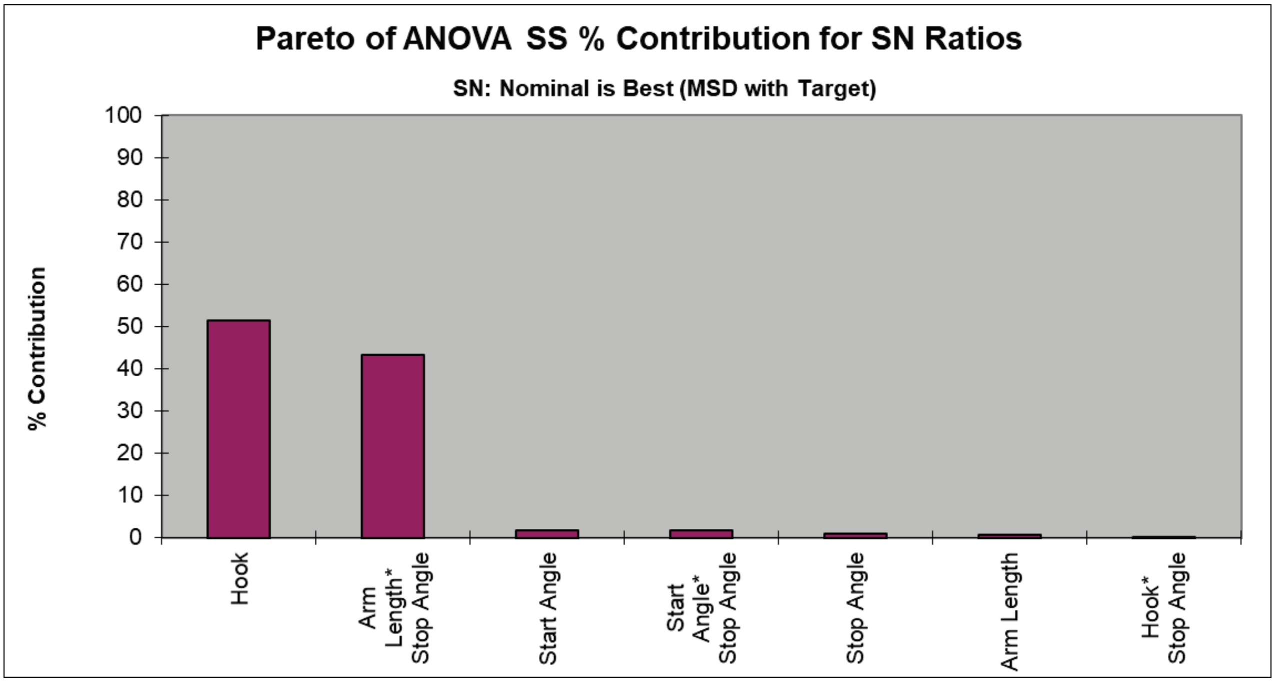
Outer Array:

Yrep1	Yrep2	Yrep3
28	27.1	26.2
46.3	43.5	46.5
21.9	21	20.1
52.9	53.7	52
75	73.1	74.3
127.7	126.9	128.7
86.2	86.5	87
195	195.9	195.7

Scroll to the right to view the calculated statistics:

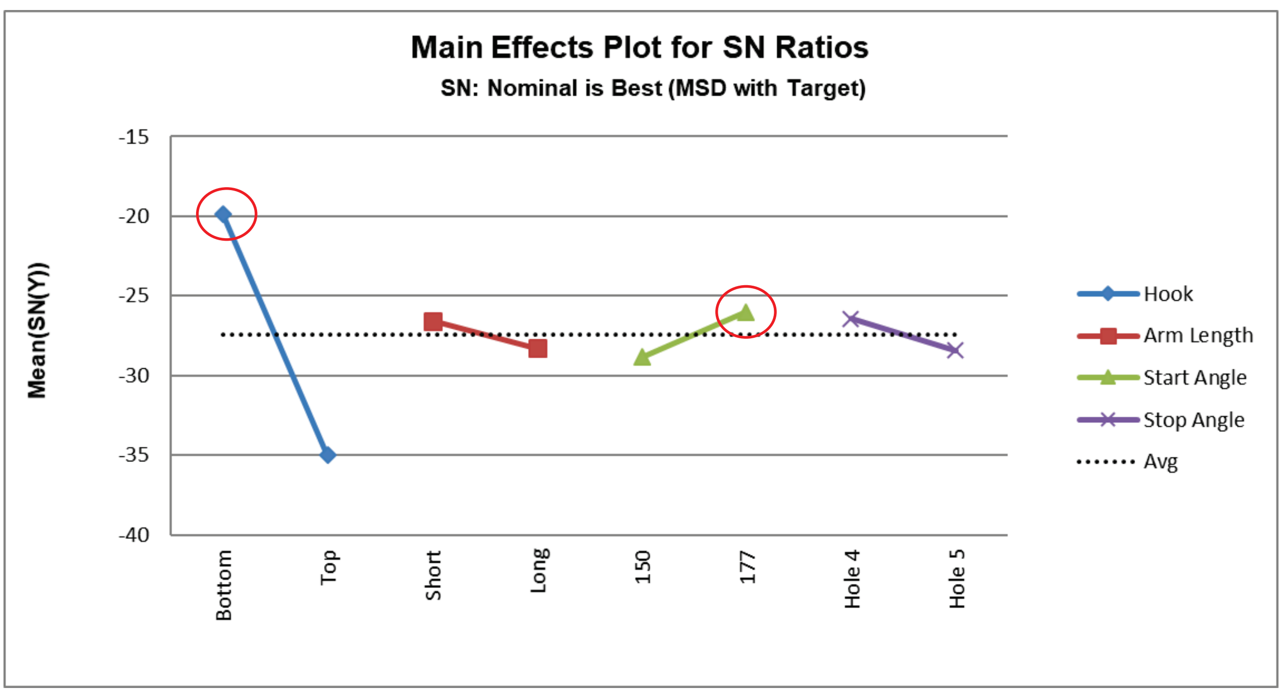
Mean (Y)	StDev (Y)	SN: Nominal is Best (MSD with Target)
27.1	0.9	-27.2011794
45.43333333	1.677299417	-13.56599436
21	0.9	-29.25074764
52.86666667	0.850490055	-9.395192526
74.13333333	0.960902354	-27.65693385
127.7666667	0.901849951	-37.81625904
86.56666667	0.404145188	-31.26206111
195.5333333	0.472581563	-43.25928006

10. We will focus our analysis on the Signal-To-Noise Ratio: Nominal is Best (MSD with Target). Since this design includes interactions, we will look at the **Pareto of ANOVA SS (Sum-of-Squares) % Contribution for SN Ratios**. Scroll right and down to view:



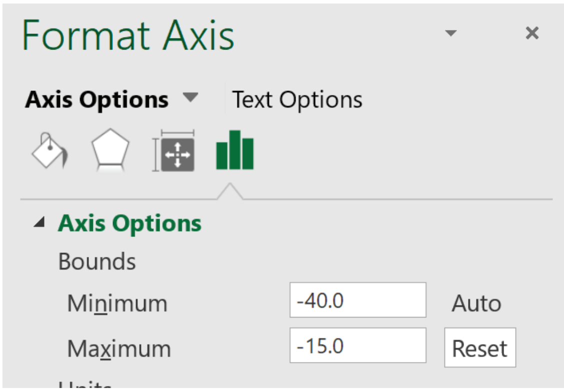
11. We see that the Hook position and Arm Length* Stop Angle interaction are the dominant factors.

12. Scroll down to view the **Main Effects Plot for SN Ratios**:

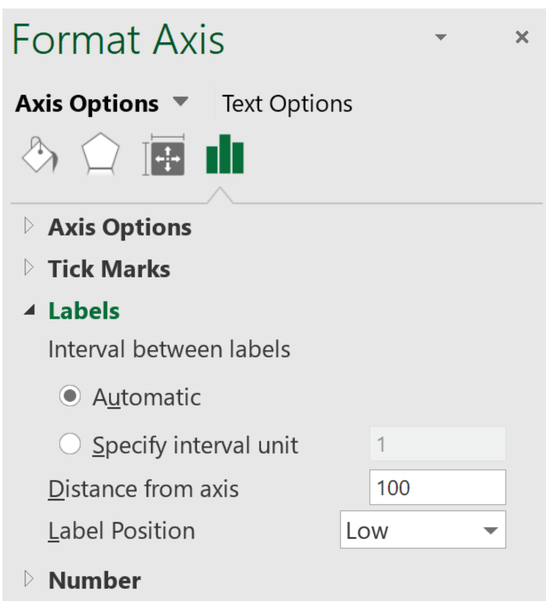


Setting Hook = Bottom and Start Angle = 177 maximizes the SN Ratio. (We will use the interaction plot to determine the optimum settings for Arm Length and Stop Angle).

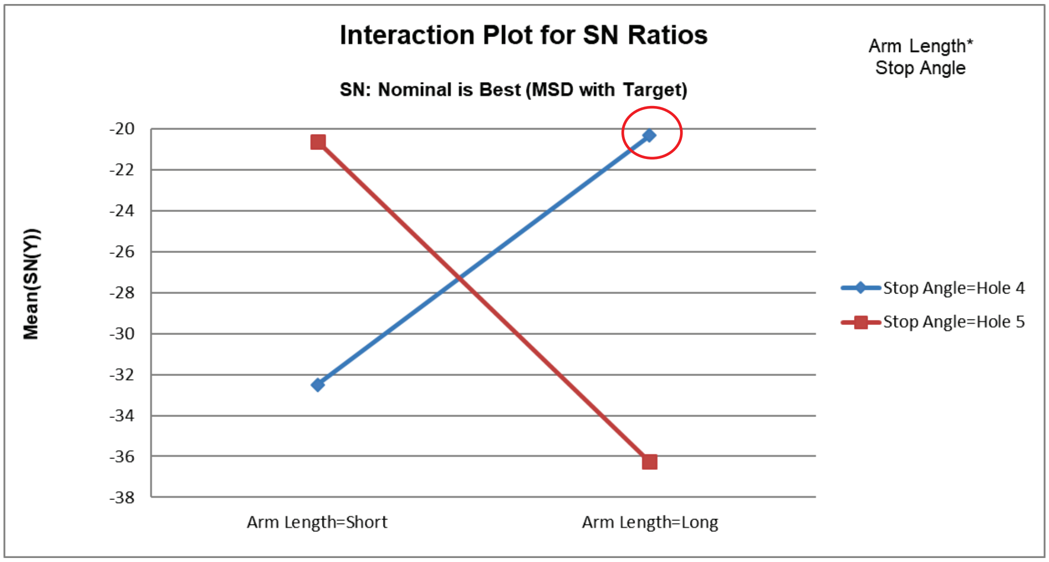
Note: The Y Axis has been adjusted to get a better scale (double click in the Y-Axis to open Excel Format Axis dialog, click **Axis Options** and set **Maximum** = -15 as shown).



With the negative SN values, the default X-axis labels appear as on top. To change the X-Axis location to bottom, double click on the X-Axis to open Excel Format Axis dialog, click **Axis Options**, **Labels** and set **Label Position** to Low as shown)



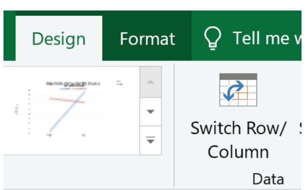
13. Scroll down to view the **Interaction Plot for SN Ratios** for the Arm Length * Stop Angle Interaction:

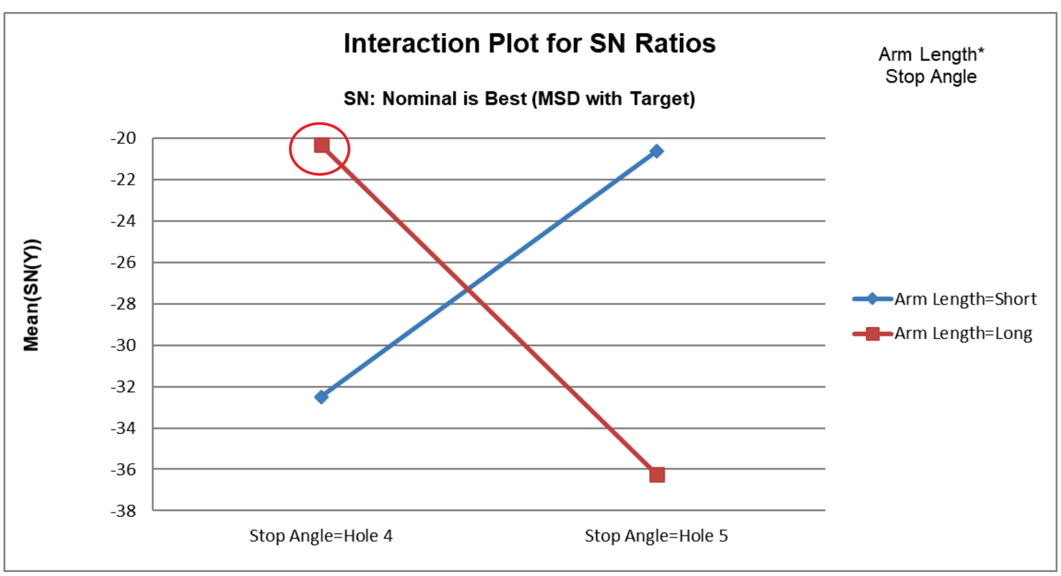


Here we can clearly see the large effect due to the two-way interaction. The effect that Arm Length has on Distance depends on what the Stop Angle is. The maximum SN Ratio is achieved with Arm Length = Long and Stop Angle = Hole 4.

Note 1: The Y-Axis has been adjusted to improve visibility (Maximum = -20).

Note 2: The Interaction Plot X-axis/Legend factors may be switched by clicking on the chart, then select Excel **Design > Switch Row/Column**:





14. From the above plots, we conclude that the optimum setting to produce a maximum SN Ratio for Distance = 50” is: Hook = Bottom, Arm Length = Long, Start Angle = 177 and Stop Angle = Hole 4.

15. Scroll up and left to enter these values in the **Predicted Output** Calculator. Factor settings can be selected from the drop-down list as shown:

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded
Hook	Bottom	1
Arm Length	Bottom	2
Start Angle	Top	2
Stop Angle	Hole 4	1

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded	$\hat{Y}$	$\hat{S}$	SN: Nominal is Best (MSD with Target)
Hook	Bottom	1	52.86666667	0.850490055	-9.395192526
Arm Length	Long	2			
Start Angle	177	2			
Stop Angle	Hole 4	1			

This gives the predicted Mean, Standard Deviation and SN Ratio for Distance and agrees with the results given in the paper. The predicted value of 52.9 inches is slightly off of the Target of 50 inches due to the discrete settings of the categorical factors.

Excel Solver cannot be used with this calculator to perform optimization, so we will use Multiple Regression and Solver, treating the Start Angle as continuous rather than categorical. This will allow interpolation in order to obtain a predicted Mean response of 50 inches.

Multiple Regression and Excel Solver (Advanced Topics):

16. The Taguchi L8 Orthogonal Array (or L8 Dummy Coding) may be analyzed using Multiple Regression Analysis, adding and removing terms from the model as necessary. However, for this example, we will not remove any terms from the model, following the analysis as performed in the paper.

17. We will be analyzing the Mean Distance to refine the settings determined above from **SN: Nominal is Best (MSD with Target)**.

18. Highlight the **L8 Orthogonal Array** cells (**AM23:AW31**) as shown.

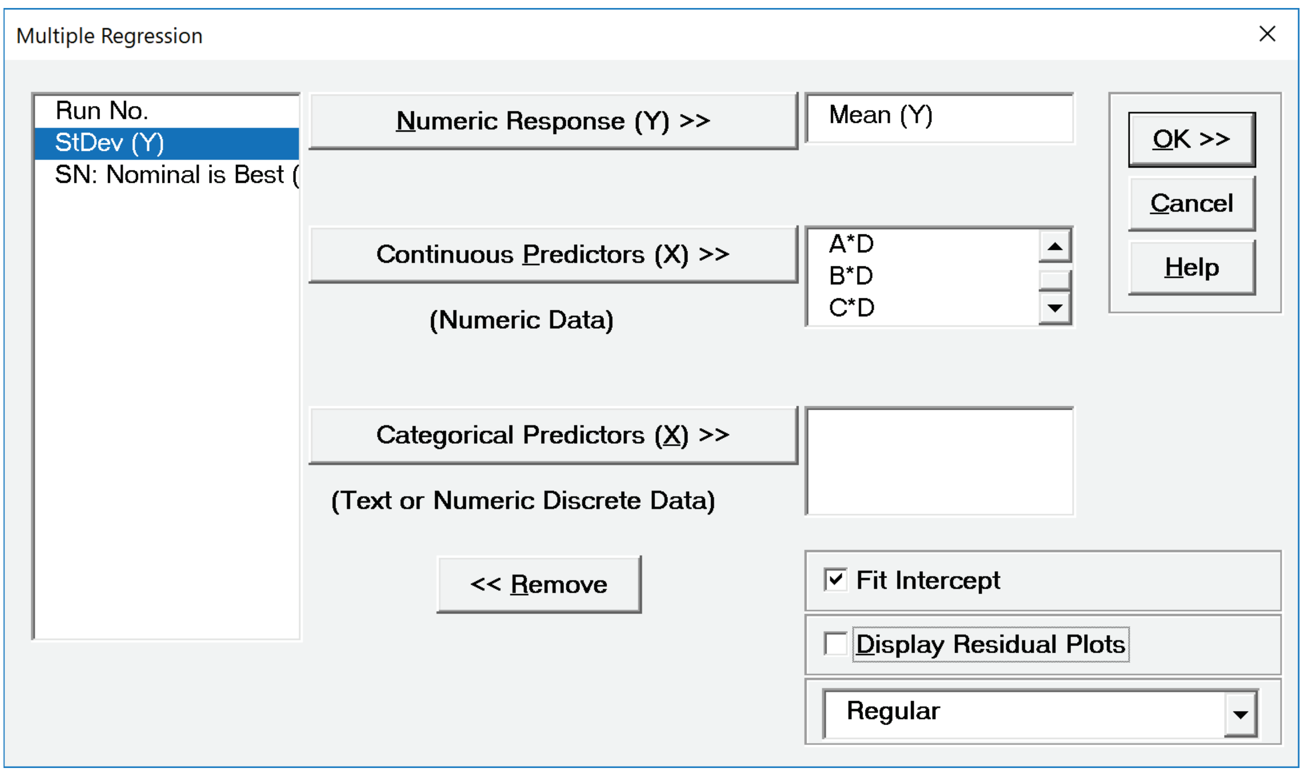
L8 Orthogonal Array:

	L8_1	L8_2	L8_3	L8_4	L8_5	L8_6	L8_7			
Run No.	A	B	C*D	C	B*D	A*D	D	Mean (Y)	StDev (Y)	SN: Nominal is Best (MSD with Target)
1	1	1	1	1	1	1	1	27.1	0.9	-27.2011794
2	1	1	1	2	2	2	2	45.43333333	1.677299417	-13.56599436
3	1	2	2	1	1	2	2	21	0.9	-29.25074764
4	1	2	2	2	2	1	1	52.86666667	0.850490055	-9.395192526
5	2	1	2	1	2	1	2	74.13333333	0.960902354	-27.65693385
6	2	1	2	2	1	2	1	127.7666667	0.901849951	-37.81625904
7	2	2	1	1	2	2	1	86.56666667	0.404145188	-31.26206111
8	2	2	1	2	1	1	2	195.5333333	0.472581563	-43.25928006

The Orthogonal Array displays the factor letters, not the names, so recall that A is Hook, B is Arm Length, C is Start Angle and D is Stop Angle.

19. Click **SigmaXL > Statistical Tools > Regression > Multiple Regression**. Click **Next**.

20. Select *Mean (Y)*, click **Numeric Response (Y) >>**; now, one at a time carefully select the factors in the sequence: *A, B, C, D, A*D, B*D and C*D*, clicking **Continuous Predictors (X)** >> each time. Using this order will make subsequent analysis easier. Uncheck **Display Residual Plots**.



21. Click **OK**. The resulting regression report is shown:

Multiple Regression Model: Mean (Y) = (-76.300) + (84.400) * A + (20.383) * B + (53.200) * C + (10.450) * D + (-17.

Model Summary:	
R-Square	100.00%
R-Square Adjusted	0.00%
S (Root Mean Square Error)	

Parameter Estimates:

Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	-76.300					
A	84.400				1	1.00000
B	20.383				1	1.00000
C	53.200				1	1.00000
D	10.450				1	1.00000
A*D	-17.217				1	1.00000
B*D	-28.100				1	1.00000
C*D	-19.717				1	1.00000

Analysis of Variance for Model:

Source	DF	SS	MS	F	P
Model	7	23906	3415.2		
Error	0	2.38026E-23			
Total (Model + Error)	7	23906	3415.2		

ANOVA and P-Values are not computed since there are no degrees of freedom for the error term.

22. Now we will use the **Predicted Response Calculator** to determine the factor settings that achieve the Target value of 50 inches.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response
A		
B		
C		
D		
A*D		
B*D		
C*D		

23. All entries must be in coded units of 1 to 2 as given in the Orthogonal Array. While the calculator in the Taguchi Template uses strictly categorical factors, here we can use either categorical or continuous. We will use the categorical settings obtained from the SN Ratio analysis for: A = 1 (Hook = Bottom), B = 2 (Arm Length = Long), and D = 1 (Stop Angle = Hole 4). Factor C (Start Angle) is continuous and will be determined using Solver in order to achieve the Target Distance = 50 inches. The initial setting will be C = 2 (Start Angle = 177).

24. Since the Taguchi Orthogonal Array is coded 1, 2 rather than -1, +1, we cannot simply multiply Factor A * Factor D to compute the A*D interaction.

25. Turn on the Excel Row and Column headers in the Multiple Regression sheet: **File > Options > Advanced**. Scroll down to **Display options for the worksheet**. Check **Show row and column headers**. Click **OK**.

26. In cell A*D (**K16)**, enter the formula:

    =1.5 - 2*(K12-1.5)*(K15-1.5)

    This formula centers and scales the A (**K12**) and D (**K15**) values so that the A*D (**K16**) value will match the values in the Taguchi array.

27. In cell B*D (**K17)**, enter the formula:

    =1.5 - 2*(K13-1.5)*(K15-1.5)

28. In cell C*D (**K18)**, enter the formula:

    =1.5 - 2*(K14-1.5)*(K15-1.5)

29. Enter A = 1, B = 1, C =1, D=1 (be careful to not overwrite the formulas – you may wish to change the **K16** to **K18** cell color as shown). The cells should show A*D = 1, B*D = 1 and C*D = 1.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response
A	1	27.1
B	1	
C	1	
D	1	
A*D	1	
B*D	1	
C*D	1	

Confidence and Prediction Intervals are not given because the model does not include an error term.

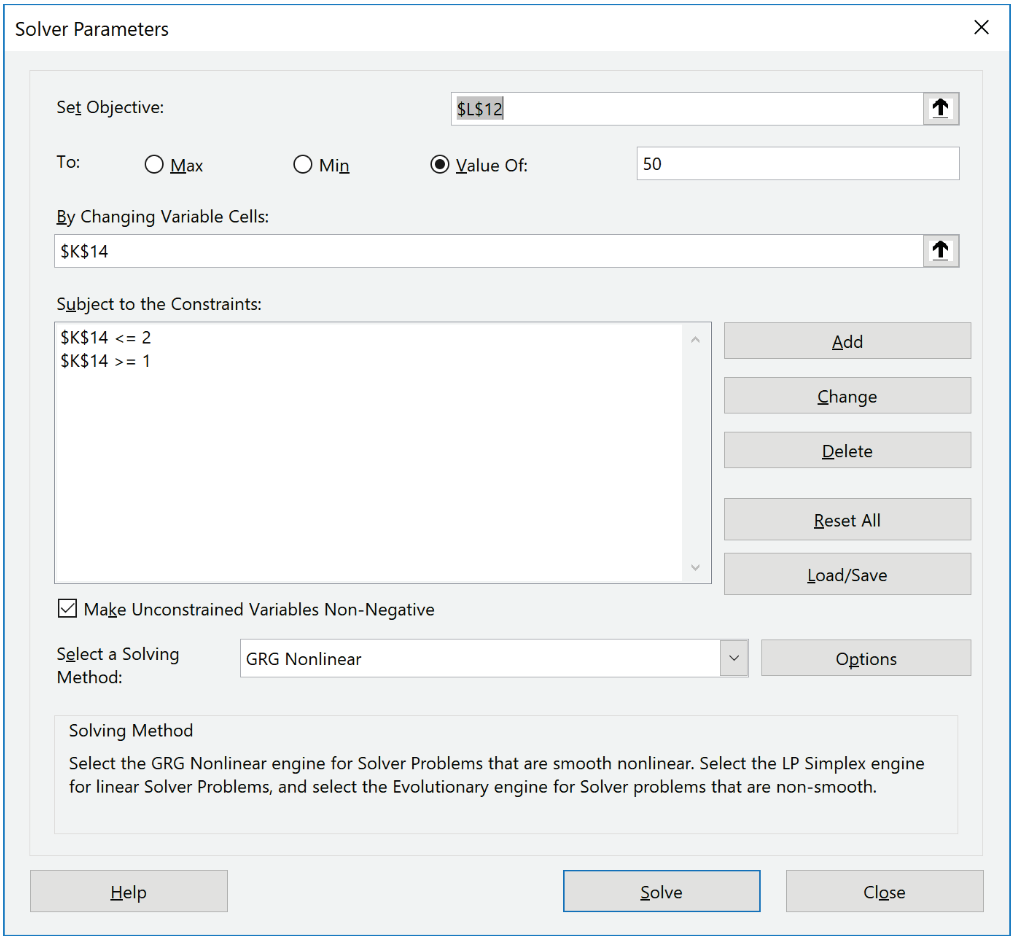
30. As mentioned above, we will use the categorical settings obtained from the SN Ratio analysis, but Factor C (Start Angle) is continuous and will be determined using Solver in order to achieve the Target Distance = 50 inches. Enter the values as shown:

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response
A	1	52.86666667
B	2	
C	2	
D	1	
A*D	1	
B*D	2	
C*D	2	

The initial Predicted Response matches the Mean in the Predicted Output given in the Taguchi Template.

31. Click Excel **Data > Solver** to open the Solver dialog (if the Solver menu option does not appear, click **File > Options > Add-Ins**. Adjacent to **Manage Excel Add-ins**, click **Go**. Check **Solver Add-in**. Click **OK**). Set the **Solver Parameters** as shown:

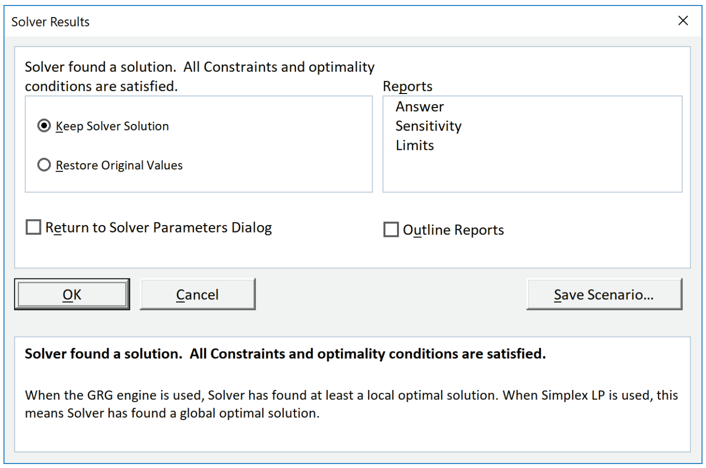


Cell **L12** is the predicted response (Mean Distance). Solver will try to set this to the Target value of 50 inches. Cell **K14** is the C Coded Factor Setting to be changed, with the constraint that the value is >=1 and <= 2. GRG Nonlinear is used here. (Note that if any of the Solver factors are modeled as categorical, Evolutionary is recommended).

32. Click **Solve**.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response
A	1	50
B	2	
C	1.914385266	
D	1	
A*D	1	
B*D	2	
C*D	1.914385266	



33. The Predicted Response is 50 inches. Click **OK** to keep the Solver solution.

34. Now we have to convert the coded C value to uncoded units. In cell **L14**, enter the formula:

    =27*K14+123

    This converts the coded 1 to 2 back to the uncoded 150 to 177 for the Start Angle.

Predicted Response Calculator:

Predictors	Enter Settings:	Predicted Response
A	1	50
B	2	
C	1.914385266	174.6884022
D	1	
A*D	1	
B*D	2	
C*D	1.914385266	

35. The factor setting of 174.7 degrees to obtain a Mean Distance of 50 inches matches the value given in the paper. Confirmation runs were performed and validated the model prediction, resulting in greater precision than a theoretical model requiring 200 person-hours to develop!

Taguchi L9 (3 Level) Four Factor – Paper Airplane Example

This Paper Airplane Experiment is adapted from the article "Teaching Taguchi's Approach to Parameter Design", by Sanjiv Sarin, Quality Progress, May 1997.

1. Click **SigmaXL > Help > Template Examples > Taguchi Examples > L9 Four Factor – Paper Airplane** to open the example file. If you wish to start with a blank template and populate the values, click **SigmaXL > Design of Experiments > Basic Taguchi DOE Templates > Taguchi L9 (3 Level) > Four Factor**.

2. The response is Distance of Flight in Inches. The goal is to maximize Distance.

3. We will maximize the Signal-to-Noise Ratio **SN: Larger is Better**, selected from the list as shown:

Signal-to-Noise Ratio:	SN: Larger is Better
Standard Deviation Response:	SN: Nominal is Best SN: Nominal is Best (Variance Only) SN: Nominal is Best (MSD with Target) SN: Larger is Better SN: Smaller is Better
Factor	

Signal-to-Noise Ratio:	SN: Larger is Better
Standard Deviation Response:	StDev (Y)

SN Formula: -10*Log10(Sum(1/Y^2)/n)

4. The **Inner Array** Controllable Input factors used in this study are: Paper Weight (1 Sheet, 2 Sheets, 3 Sheets), Design (Design 1, Design 2, Design 3), Paper Width (4 inches, 6 inches, 8 inches) and Paper Length (6 inches, 8 inches, 10 inches).

Factor	Factor Name	Level 1	Level 2	Level 3
A	Paper Weight	1	2	3
B	Design	1	2	3
C	Paper Width	4	6	8
D	Paper Length	6	8	10

5. The Uncontrollable Factors are Launch Height, Launch Angle and Ground Surface. The **Outer Array** was created using an L4 Three Factor design. Scroll right to view the screen capture:

Outer Array created using Taguchi L4 Three Factor Template:

Factor	Factor Name	Level 1	Level 2
A	Launch Height	Ground	Chair
B	Launch Angle	Horizontal	45 Degrees
C	Ground Surface	Concrete	Wood

Run No.	Launch Height	Launch Angle	Ground Surface
1	Ground	Horizontal	Concrete
2	Ground	45 Degrees	Wood
3	Chair	Horizontal	Wood
4	Chair	45 Degrees	Concrete

6. The distance values are entered into the **Outer Array** of the L9 as **Yrep1** to **Yrep4**:

Inner Array:

Std. Order	Run No.	Paper Weight	Design	Paper Width	Paper Length
1	1	1	1	4	6
2	2	1	2	6	8
3	3	1	3	8	10
4	4	2	1	6	10
5	5	2	2	8	6
6	6	2	3	4	8
7	7	3	1	8	8
8	8	3	2	4	10
9	9	3	3	6	6

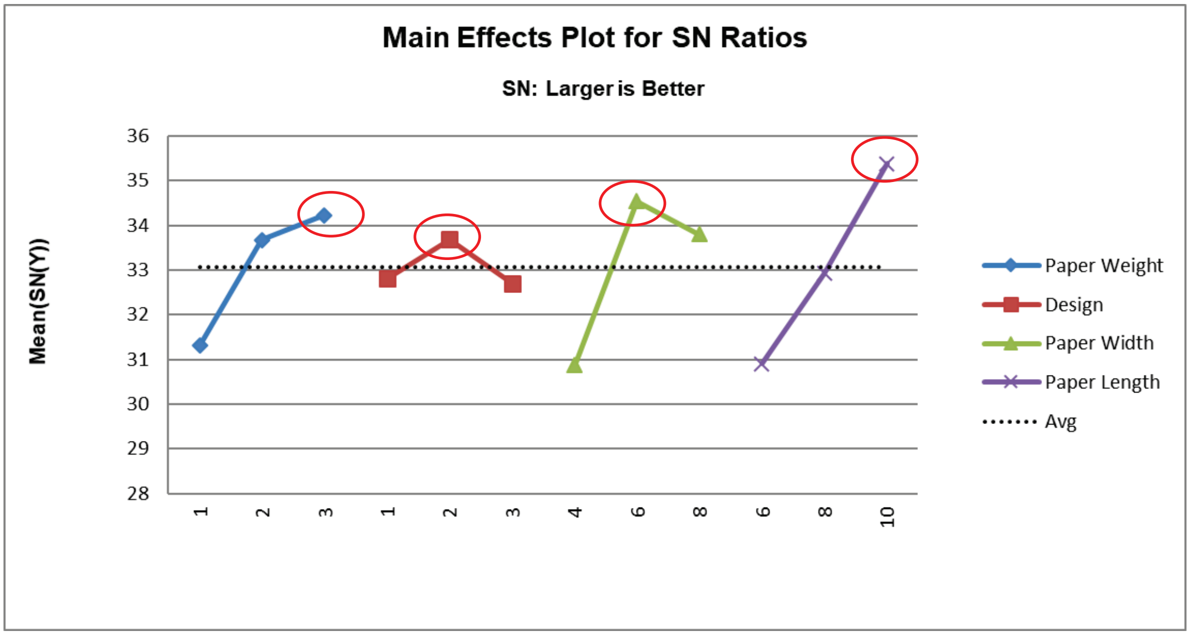
Outer Array:

Yrep1	Yrep2	Yrep3	Yrep4
49	44	12	38
91	42	44	38
59	48	39	67
116	89	48	88
32	38	56	108
24	55	39	46
42	76	122	41
50	73	47	65
76	34	80	37

Scroll to the right to view the calculated statistics:

Mean (Y)	StDev (Y)	SN: Larger is Better
35.75	16.45954637	26.69079814
53.75	24.95829855	33.24061006
53.25	12.28481447	33.97524775
85.25	28.01636427	37.1912413
58.5	34.53983208	32.85315172
41	13.08943594	30.97352457
70.25	38.14336989	34.55282905
58.75	12.33896268	34.95858081
56.75	24.6221445	33.17653559

7. Since there are no interactions, we will focus our analysis on the Main Effects Plots for Signal-To-Noise Ratio and Mean Distance as done in the paper. Scroll right and down to view the **Main Effects Plot for SN Ratios**:



The optimum settings to maximize SN Ratio are: Paper Weight = 3, Design = 2, Paper Width = 6 and Paper Length = 10.

8. Scroll left to look at the **Main Effects Plot for Means**:

The optimum settings to maximize the Mean are: Paper Weight = 3, Design = 1, Paper Width = 6 and Paper Length = 10. (Note, The Y-Axis minimum has been adjusted to 40 to improve the scale).

9. Commenting on this difference, Sarin states:

   > *Factor B.* For the factor of design, a plane built using design 1 was superior based on the $\overline{\overline{X}}$ value. Based on the average S/N ratio, however, the difference between design 1 and design 2 was marginal. When there is disagreement between the level that optimizes the $\overline{\overline{X}}$ value and the one that maximizes the average S/N ratio, it is usually recommended that the average S/N ratio be given priority. This is due to the fact that the Taguchi approach focuses first on selecting levels of controllable factors that reduce variability; bringing the process on target is the secondary objective. Thus, according to the Taguchi approach, design 2 should be recommended. But as a matter of practicality, design 1 was chosen since it results in significant improvement in the $\overline{\overline{X}}$ value while sacrificing the S/N value to a relatively lesser degree.

10. We will compare the two settings using the Predicted Output Calculator. Scroll up and enter (or select from the drop-down list) as shown:

Predicted Output for Y:

Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded	$\hat{Y}$	$\hat{S}$	SN: Larger is Better
Paper Weight	3	3	78.83333333	24.24191679	38.62040862
Design	2	2			
Paper Width	6	2			
Paper Length	10	3			

Now change the Design to 1 as shown:

Predicted Output for Y:

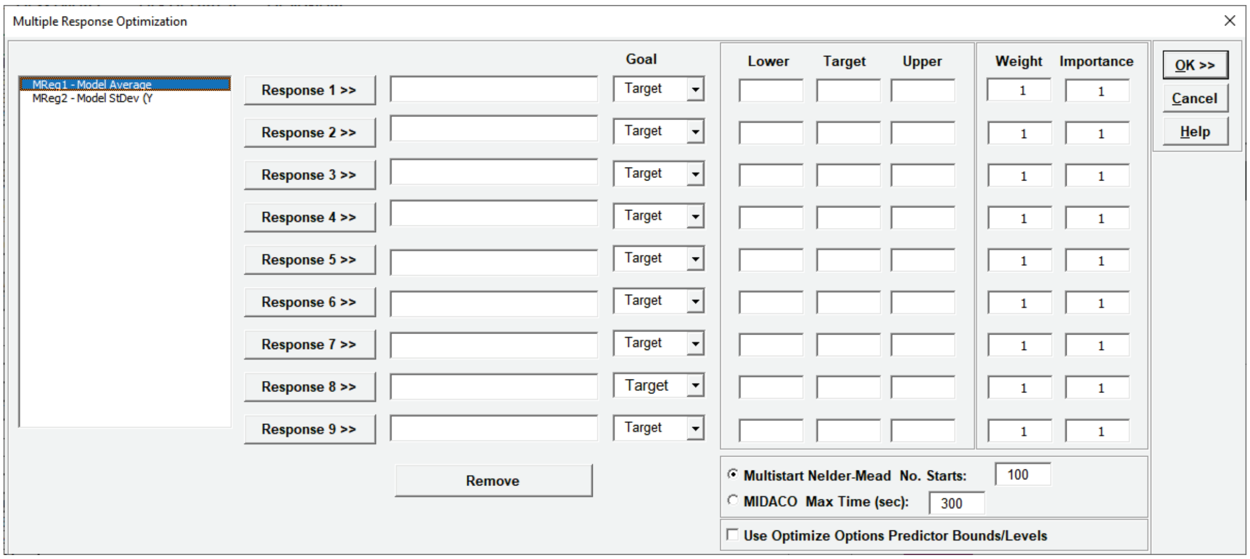
Factor Name	Enter Actual Factor Setting - Uncoded	Factor Setting Coded	$\hat{Y}$	$\hat{S}$	SN: Larger is Better
Paper Weight	3	3	85.58333333	27.83597919	37.74791725
Design	1	1			
Paper Width	6	2			
Paper Length	10	3			

11. As noted by the author, depending on the priority given to Mean or Signal-To-Noise Ratio, one would select Design 1 or 2. In either case, confirmation runs need to be performed to validate the experimental results.

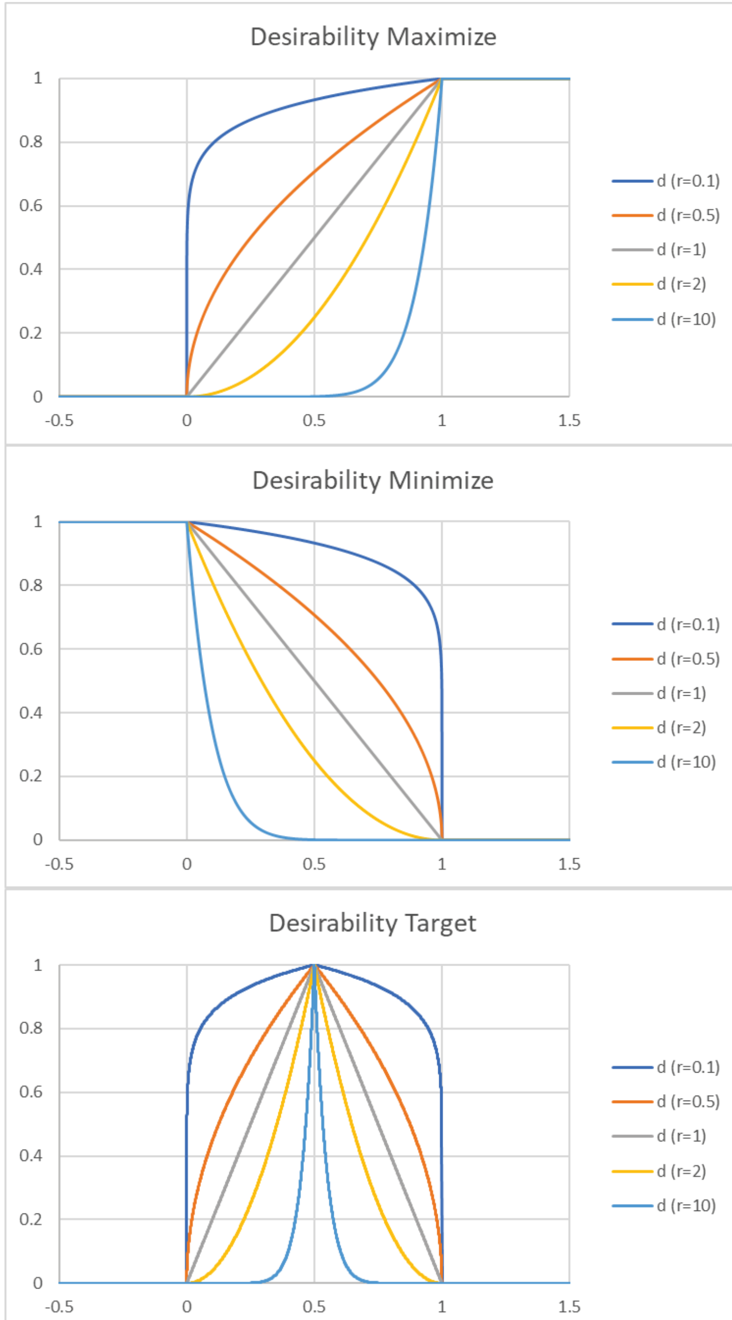
Part F – Multiple Response Optimization with Advanced Multiple Regression

Multiple Response Optimization (MRO) in Advanced Multiple Regression maximizes the Desirability function (with values 0 to 1). For details on the statistical methods, formulas and references, see the Appendix: Multiple Response Optimization.

Multiple Response Optimization Dialog



- Select Responses to use in MRO from the available models in the workbook that have been created using Advanced Multiple Regression.
- If the response **Goal** is *Target*, a **Target** value must be specified. It must be a value greater than **Lower** and less than **Upper**.
- If the response **Goal** is *Maximize* or *Minimize*, Target is not used.
- **Lower** and **Upper** limits must always be specified. They can be specification limits or practical limits for maximization/minimization. For maximization, a practical upper limit would be a value beyond which an increase in desirability is not beneficial. For minimization, a practical lower limit is a value below which an increase in desirability is not beneficial. The initial values for Lower and Upper that are given in the dialog are the minimum and maximum response values respectively.
- **Importance** allows for weighting of the responses according to relative importance. Use a value between 0.1 and 10 (default = 1).
- **Weight** (r) is a shape factor. For weight = 1 (default), the desirability function increases linearly. For weight < 1, the function is concave, so low priority on hitting the target or goal. For weight > 1, the function is convex, so high priority on hitting the target or goal. Use a value between 0.1 and 10.



- **Multistart Nelder-Mead:** Nelder-Mead Simplex is derivative free so can accommodate the non-smooth desirability response with both continuous and categorical predictors. It is very fast but gives a local solution. Multistart helps to improve the chances of finding a global solution. If there are a number of equally valid best solutions, the one with the smallest predicted response standard error is chosen.

- **MIDACO:** Mixed Integer Distributed Ant Colony Optimization (MIDACO) is a global optimizer so would be more suitable for a complex problem with multiple local solutions, but is slower than Nelder-Mead.

- **Use Optimize Options Predictor Bounds/Levels:** For each specified response model sheet that is used in MRO, continuous predictors can optionally be constrained to a range or specified as integers and categorical predictors can be held to a specified level. If this option is checked, MRO will apply these constraints. If there is a conflict in the constraints across models for a given continuous predictor, the lower bound used in MRO is the maximum of the lower bounds and upper bound is the minimum of the upper bounds. If any of the models flag a continuous predictor as an integer, then it is assumed to be integer continuous for MRO. For categorical levels, if they are inconsistent across models, then MRO will use all levels. In order to constrain MRO to use a specific categorical level, all models must have that same level specified.

Multiple Response Optimization Example: Robust Cake

We will re-analyze the Robust Cake DOE data from Part B – Three Factor Full Factorial Example Using DOE Template and apply Multiple Response Optimization to the Average and Standard Deviation of Taste Score.

1. Open the file **DOE Example - Robust Cake.xlsx**. This is a Robust Cake Experiment adapted from the Video "Designing Industrial Experiments" by Box, Bisgaard and Fung.

2. The response is Taste Score (on a scale of 1-7 where 1 is "awful" and 7 is "delicious").

3. The five Outer Array Reps have different Cooking Time and Temperature Conditions. The outer array was a two-factor, full-factorial plus center point, hence 5 replications.

4. The goal is to Maximize Mean and Minimize StDev of the Taste Score.

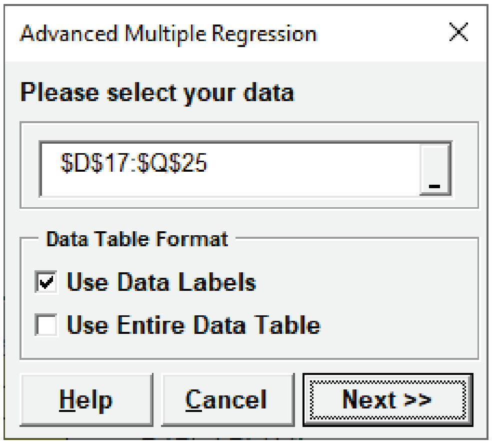
5. The X factors are Flour, Butter, and Egg. Actual low and high settings are not given in the video, so we will use coded -1 and +1 values. We are looking for a combination of Flour, Butter, and Egg that will not only taste good, but consistently taste good over a wide range of Cooking Time and Temperature conditions.

6. In the Experimental Worksheet, highlight the columns Flour to StDev (Y) (cells **D17:Q25**) as shown:

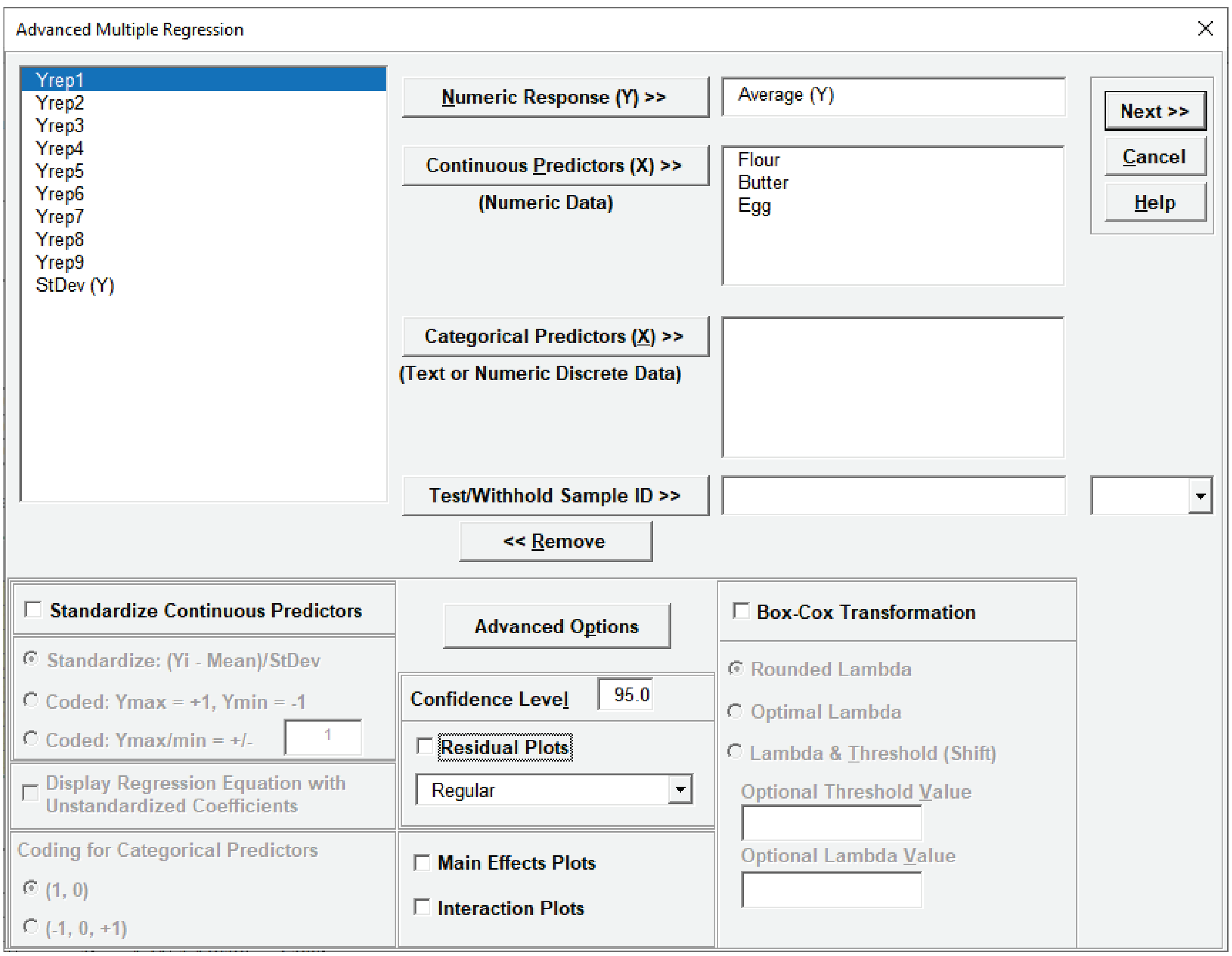
Experimental Worksheet: **Outer Array or Response Replicates:**

Standard Run Order	Actual Run Order	Flour	Butter	Egg	Yrep1	Yrep2	Yrep3	Yrep4	Yrep5	Yrep6	Yrep7	Yrep8	Yrep9	Average (Y)	StDev (Y)
1	6	-1	-1	-1	3.1	1.1	5.7	6.4	1.3					3.52	2.4499
2	2	1	-1	-1	3.2	3.8	4.9	4.3	1.3					3.5	1.38022
3	7	-1	1	-1	5.3	3.7	5.1	6.7	2.9					4.74	1.47919
4	1	1	1	-1	4.1	4.5	6.4	5.8	5.2					5.2	0.93541
5	5	-1	-1	1	5.9	4.2	6.8	6.5	3.5					5.38	1.45499
6	8	1	-1	1	6.9	5	6	5.9	5.7					5.9	0.68191
7	3	-1	1	1	3	3.1	6.3	6.4	3					4.36	1.81742
8	4	1	1	1	4.5	3.9	5.5	5	5.4					4.86	0.66558

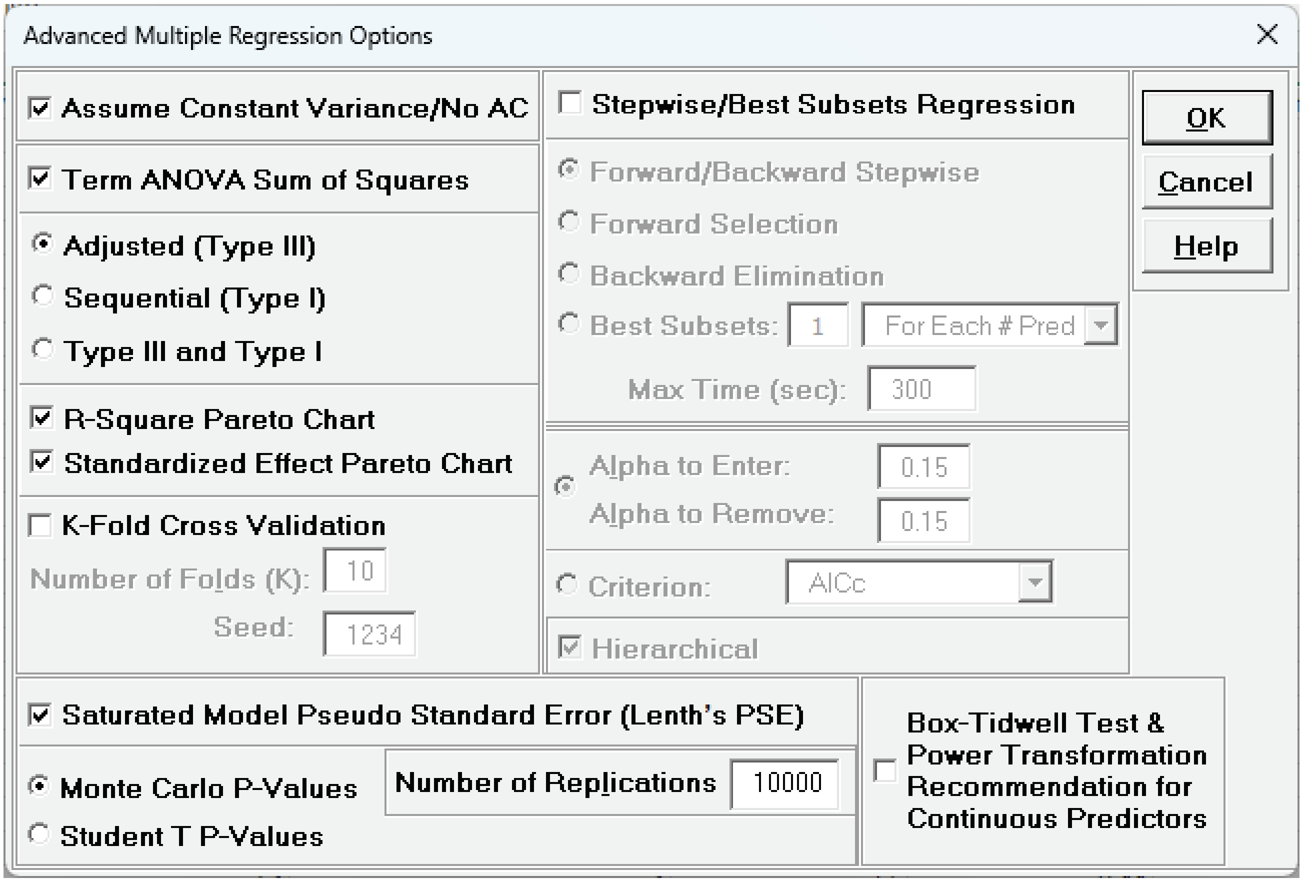
7. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Fit Multiple Regression Model**.



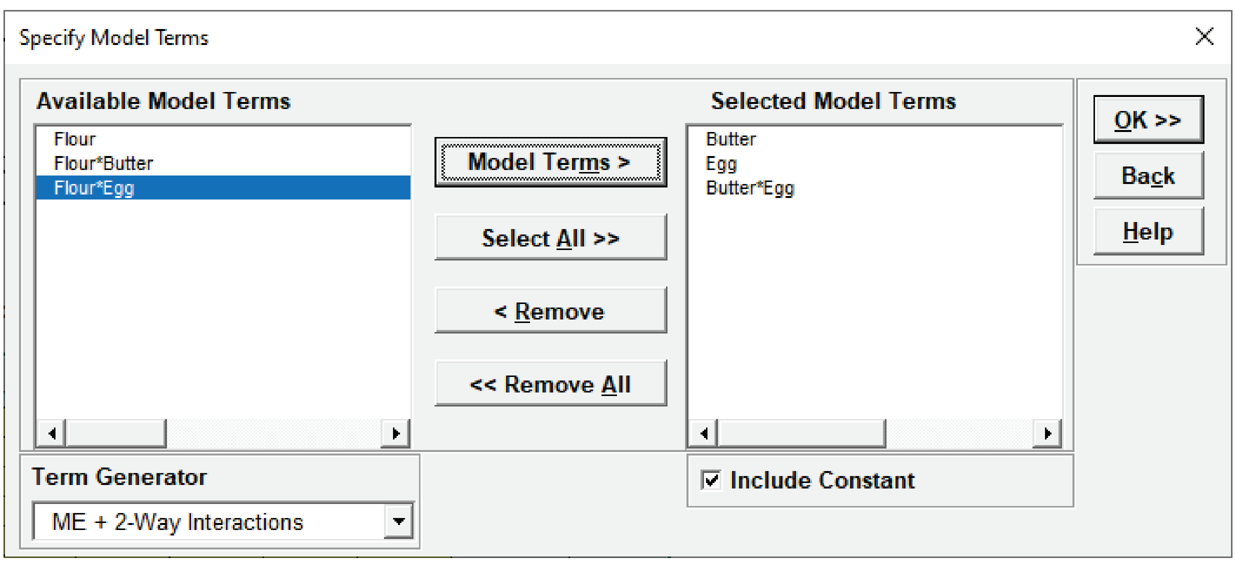
8. Ensure that the data selection is **$D$17:$Q$25**. Click **Next >>**. Select *Average (Y)*, click **Numeric Response (Y) >>**; select *Flour*, *Butter*, and *Egg*; click **Continuous Predictors (X) >>**. Uncheck **Residual Plots**, **Main Effects Plots** and **Interaction Plots** as we will not be referring to them in this demonstration.



9. Click **Advanced Options**. We will use the defaults as shown:



10. Click **OK**. Click **Next >>**. Select *ME + 2-Way Interactions*. Holding the CTRL key, select *Butter*, *Egg*, and *Butter*Egg*, click **Model Terms >**.



This matches the final model used in the original analysis of Average (Y).

11. Click **OK**. The Advanced Multiple Regression report for Average (Y) is given:

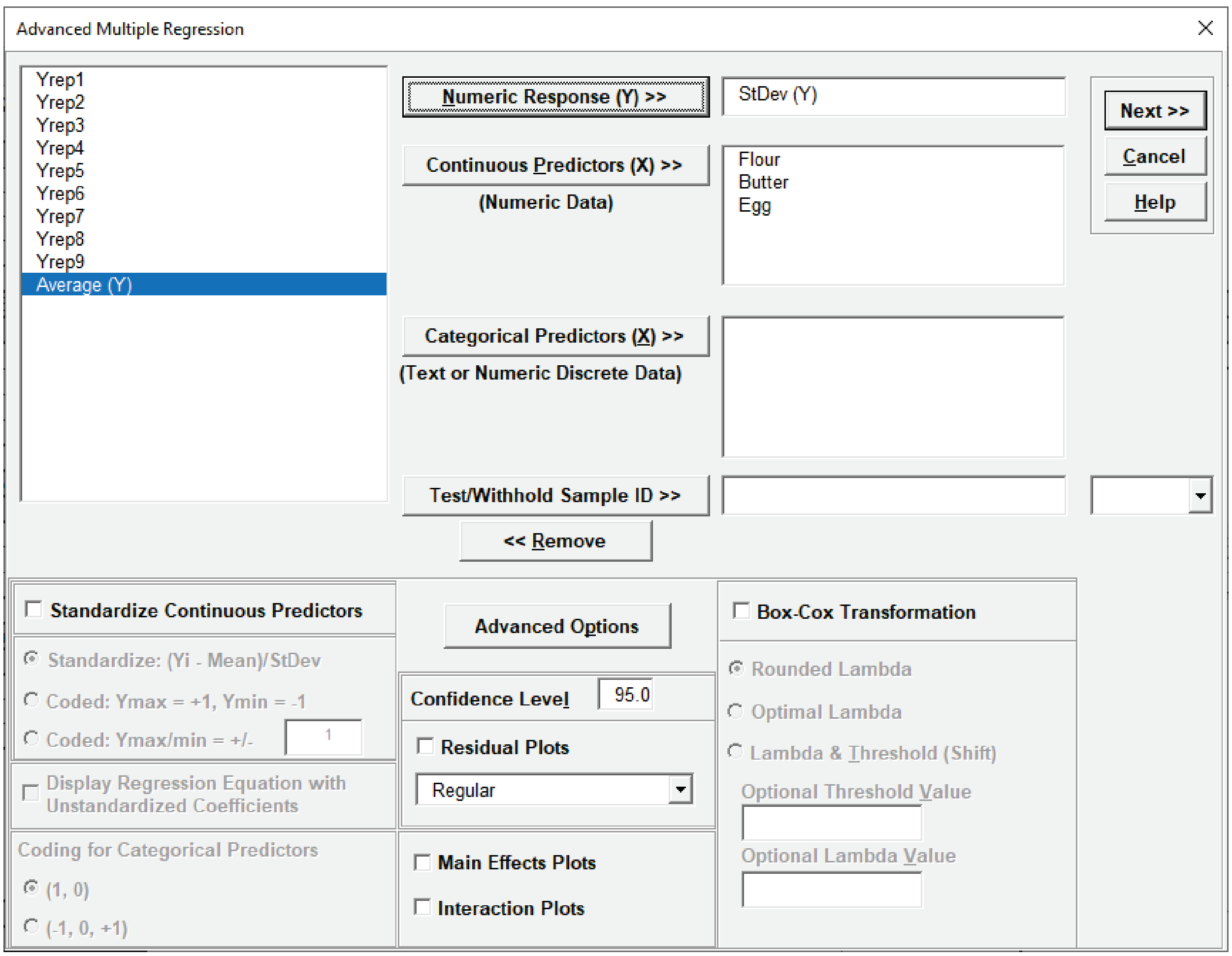
Multiple Regression Model: Average (Y) = (4.6825) + (0.1075)*Butter + (0.4425)*Egg + (-0.6225)*Butter*Egg

Model Summary	
R-Square	92.85%
R-Square Adjusted	87.50%
R-Square Predicted	71.42%
S (Root Mean Square Error)	0.3026

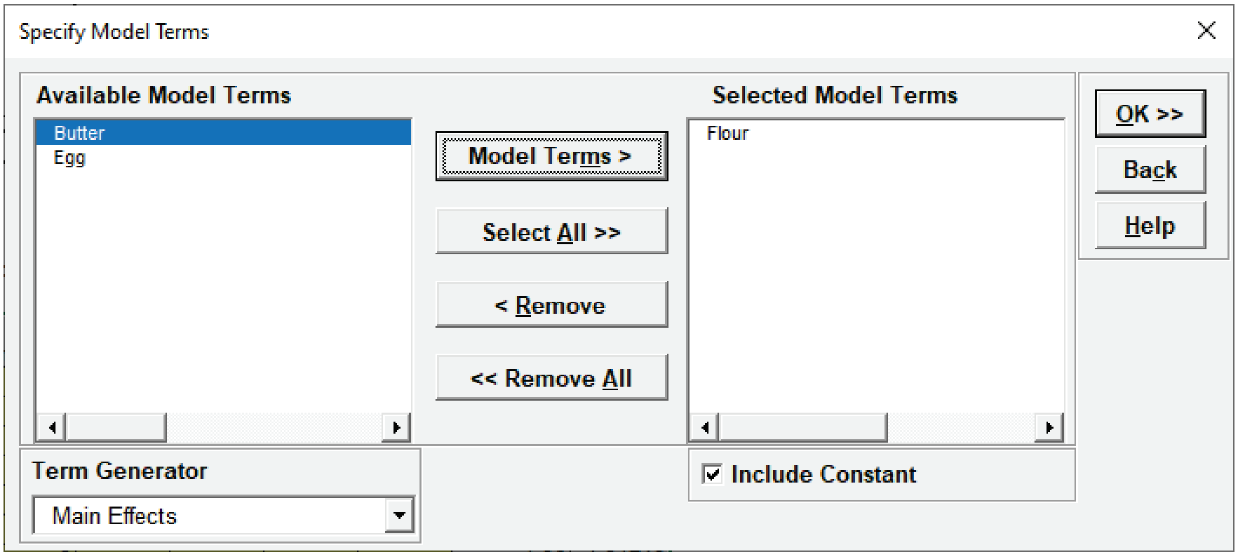
Model Information	
Continuous Predictor Standardization/Coding	N/A
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	4.6825	0.106975464	43.7717	0.0000		
Butter	0.1075	0.106975464	1.0049	0.3718	1.0000	1.0000
Egg	0.4425	0.106975464	4.1365	0.0144	1.0000	1.0000
Butter*Egg	-0.6225	0.106975464	-5.8191	0.0043	1.0000	1.0000

12. Next, we will create a regression model for StDev (Y). Click **Recall Last Dialog** (or press **F3**). Select *StDev (Y)*, click **Numeric Response (Y) >>**.



13. Click **Next >>**. Select *Flour*. Click **Model Terms >**.



This matches the final model used in the original analysis of StDev (Y).

14. Click **OK**. The Advanced Multiple Regression report for StDev (Y) is given:

Multiple Regression Model: StDev (Y) = (1.35808) + (-0.442296)*Flour

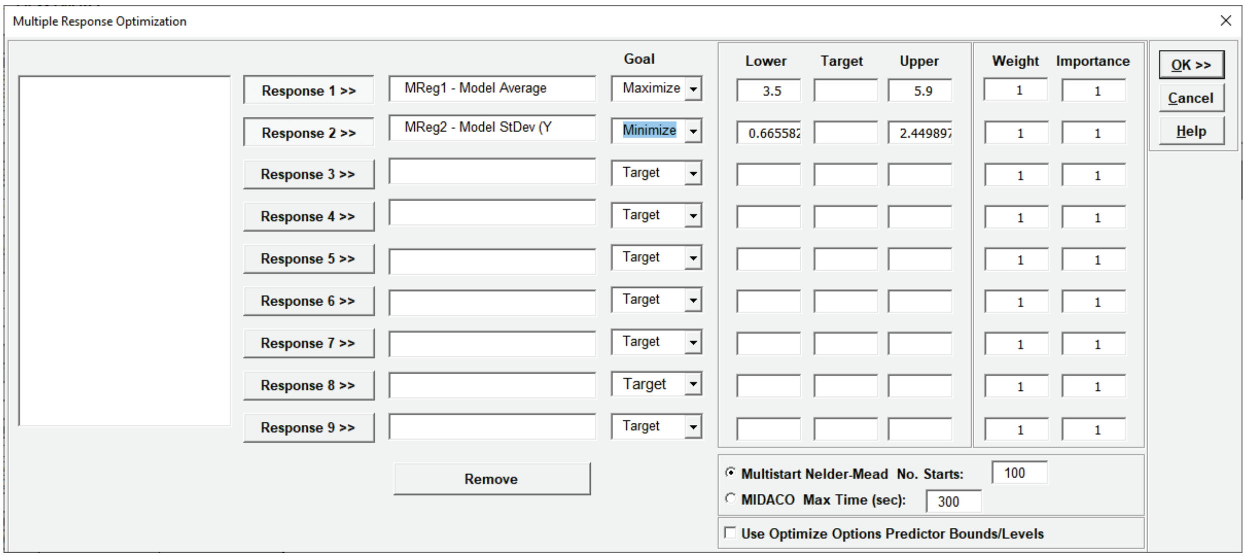
Model Summary	
R-Square	61.54%
R-Square Adjusted	55.13%
R-Square Predicted	31.63%
S (Root Mean Square Error)	0.4037

Model Information	
Continuous Predictor Standardization/Coding	N/A
Categorical Predictor Coding	N/A
Box-Cox Transformation Lambda/Threshold	N/A
Stepwise Method	N/A

Parameter Estimates						
Predictor Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	1.358077167	0.142741225	9.5143	0.0001		
Flour	-0.44229635	0.142741225	-3.0986	0.0212	1.0000	1.0000

15. Now we will simultaneously maximize the Average (Y) and minimize the StDev (Y) using Multiple Response Optimization. Click **SigmaXL > Statistical Tools > Advanced Multiple Regression > Multiple Response Optimization**.

16. Select *MReg1 – Model Average*, click **Response 1 >>**. Set **Goal** = *Maximize*. Select *MReg2 – Model StDev*, click **Response 2 >>**. Set **Goal** = *Minimize*. We will use the default **Lower** and **Upper** values which are derived from the response minimum and maximum, and the default **Weight** = 1 and **Importance** = 1 for both responses.



Note, Target is not required for Goal = Maximize or Minimize.

17. Click **OK >>**. The Multiple Response Optimization report is given:

Predictors	Settings:
Butter	-1
Egg	1
Flour	1

Response	Goal	Predicted Response	SE	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI	Desirability
Average (Y)	Maximize	5.64	0.213950929	5.045976991	6.234023009	4.611121967	6.668878033	0.891666667
StDev (Y)	Minimize	0.915780814	0.201866576	0.421831096	1.409730532	-0.188724333	2.020285962	0.859779079

Composite Desirability
0.875577721

Optimization: Maximize Desirability using Multistart Nelder-Mead.

The Predictor Settings match those used in the original analysis, with *Butter* = -1, *Egg* = +1, and *Flour* = +1. Note, predictor names in this report are sorted alphanumerically.

The Predicted Response values are slightly different than those in the previous analysis because here we are using the refined models, whereas the DOE Template Predicted Output uses the model with all terms included.

MRO maximizes the Composite Desirability score which can be a 0 to 1 value. For more information on the individual and composite desirability scores, see the Appendix: [Multiple Response Optimization](#).

SigmaXL: Control Phase Tools: Statistical Process Control (SPC) Charts and Time Series Forecasting

Part A - Individuals Charts

Tip: See Part E – Control Chart Selection Guide. The Control Chart Selection Guide makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

Individuals Chart Template

Click **SigmaXL > Templates and Calculators** > **Control Chart Templates > Basic > Individuals.**

This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Basic > Individuals.**

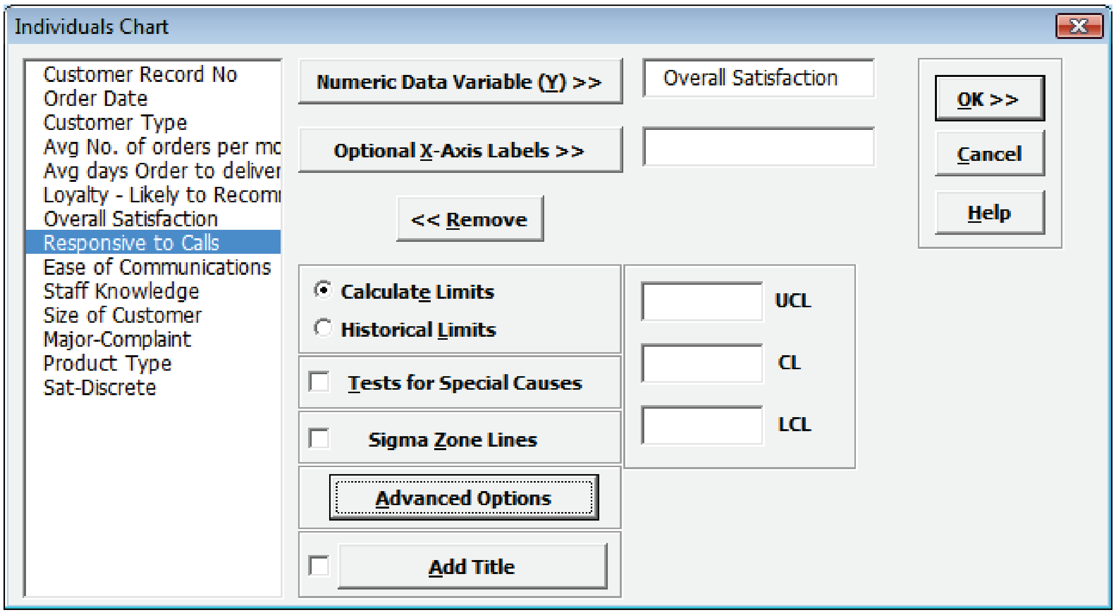
See **Measure Phase Part B – Templates and Calculators** for Individuals Chart template example:

Basic Control Chart Templates – Individuals Chart

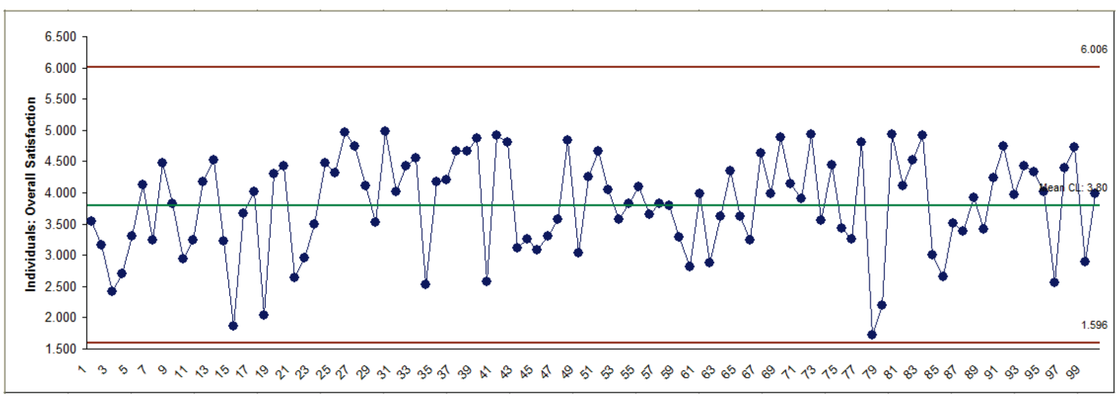
Individuals Chart

1. Open **Customer Data.xlsx**, click on **Sheet 1**. Click **SigmaXL>Control Charts>Individuals**. Ensure that entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>.** Ensure that **Calculate Limits** is selected.

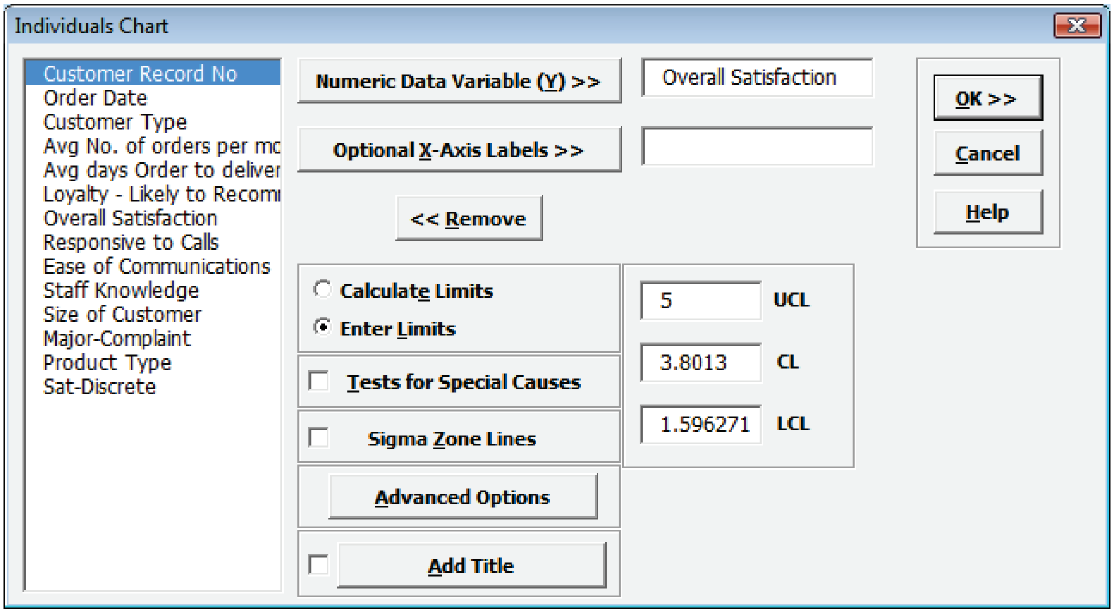


3. Click **OK**. Resulting Individuals control chart:

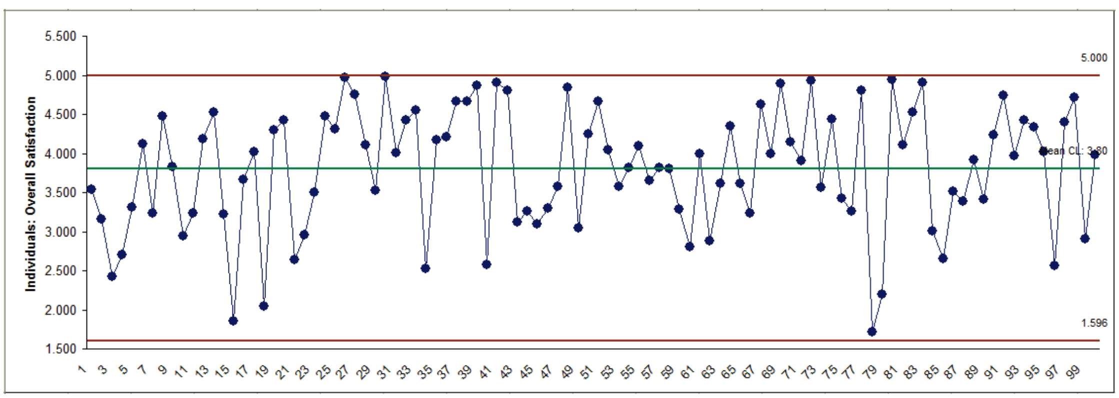


4. We have seen this data earlier as a run chart. The Control Chart adds calculated control limits. Note that the Upper Control Limit exceeds the survey upper limit of 5. Here it would be appropriate to change the UCL to 5.0. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.

5. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**, change **UCL** to 5.



6. Click **OK**. Resulting Individuals chart with modified UCL:



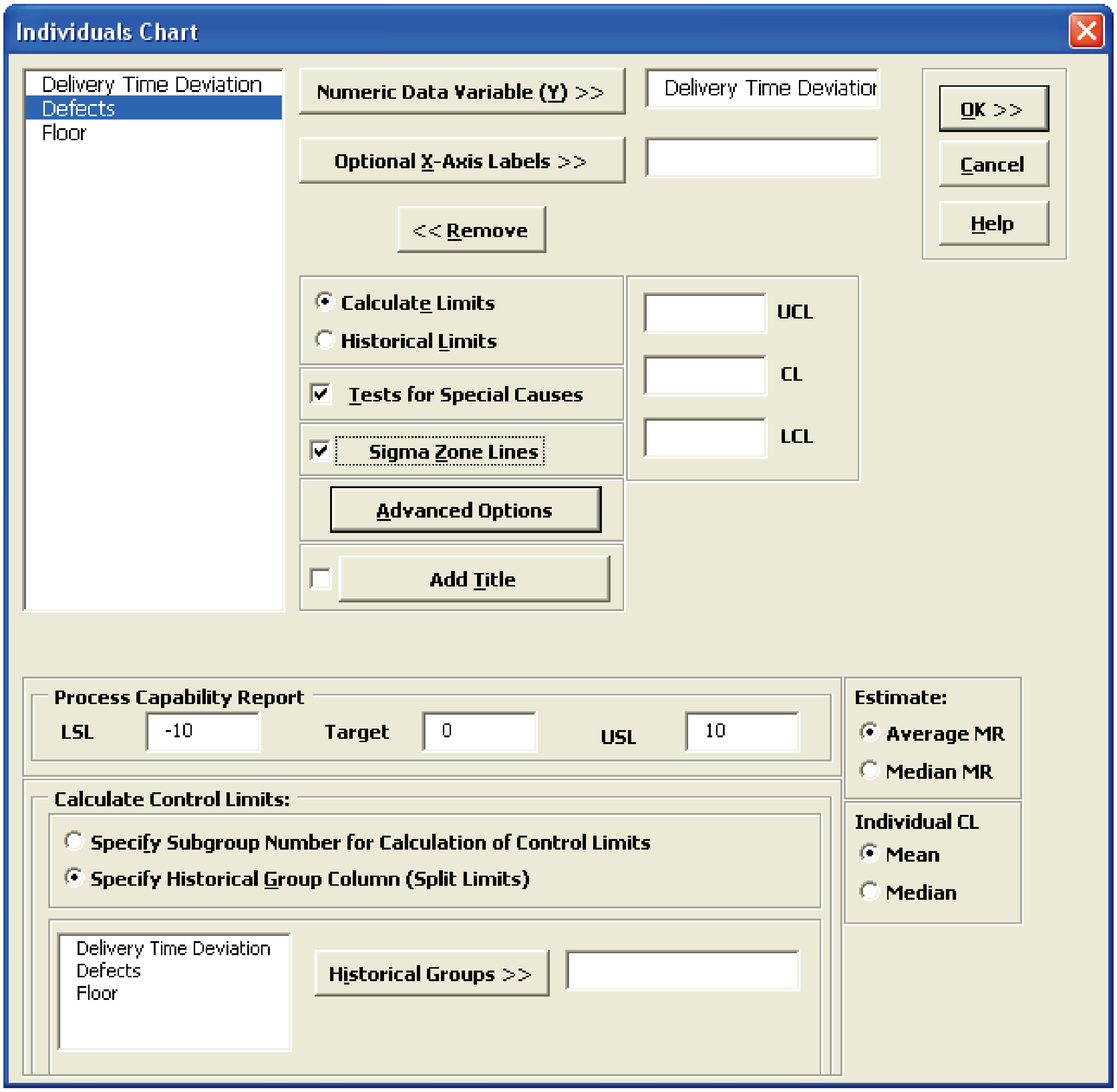
Tip: You should not change the calculated control limits unless you have a legitimate reason such as a boundary condition. Do not replace the control limits with specification limits – the control chart will lose its statistical ability to detect assignable causes. We will redo the Individuals chart for Overall Satisfaction later using the Individuals Nonnormal tool.

7. Open **Delivery Times.xlsx.** Click **Sheet 1** Tab. This data set contains room service delivery time deviations in minutes. The Critical Customer Requirement is target time +/- 10 minutes.

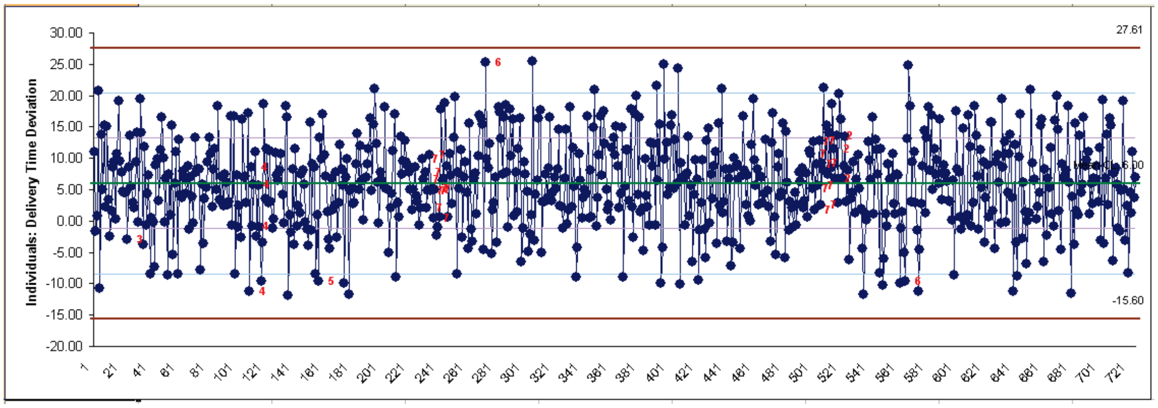
8. Click **SigmaXL > Control Charts > Individuals**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

9. Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**, check **Tests for Special Causes**, check **Sigma Zone Lines**. Ensure that **Calculate Limits** is selected.

10. Select **Advanced Options**, specify **LSL** = -10, **Target** = 0, **USL** = 10 as shown:


Individuals Chart
Delivery Time Deviation
Defects
Floor
Numeric Data Variable (Y) >>
Delivery Time Deviatior
Optional X-Axis Labels >>
<< Remove
OK >>
Cancel
Help
Calculate Limits
Historical Limits
UCL
CL
LCL
Tests for Special Causes
Sigma Zone Lines
Advanced Options
Add Title
Process Capability Report
LSL
-10
Target
0
USL
10
Estimate:
Average MR
Median MR
Calculate Control Limits:
Specify Subgroup Number for Calculation of Control Limits
Specify Historical Group Column (Split Limits)
Individual CL
Mean
Median
Delivery Time Deviation
Defects
Floor
Historical Groups >>

11. Click **OK**. The resulting chart is shown:



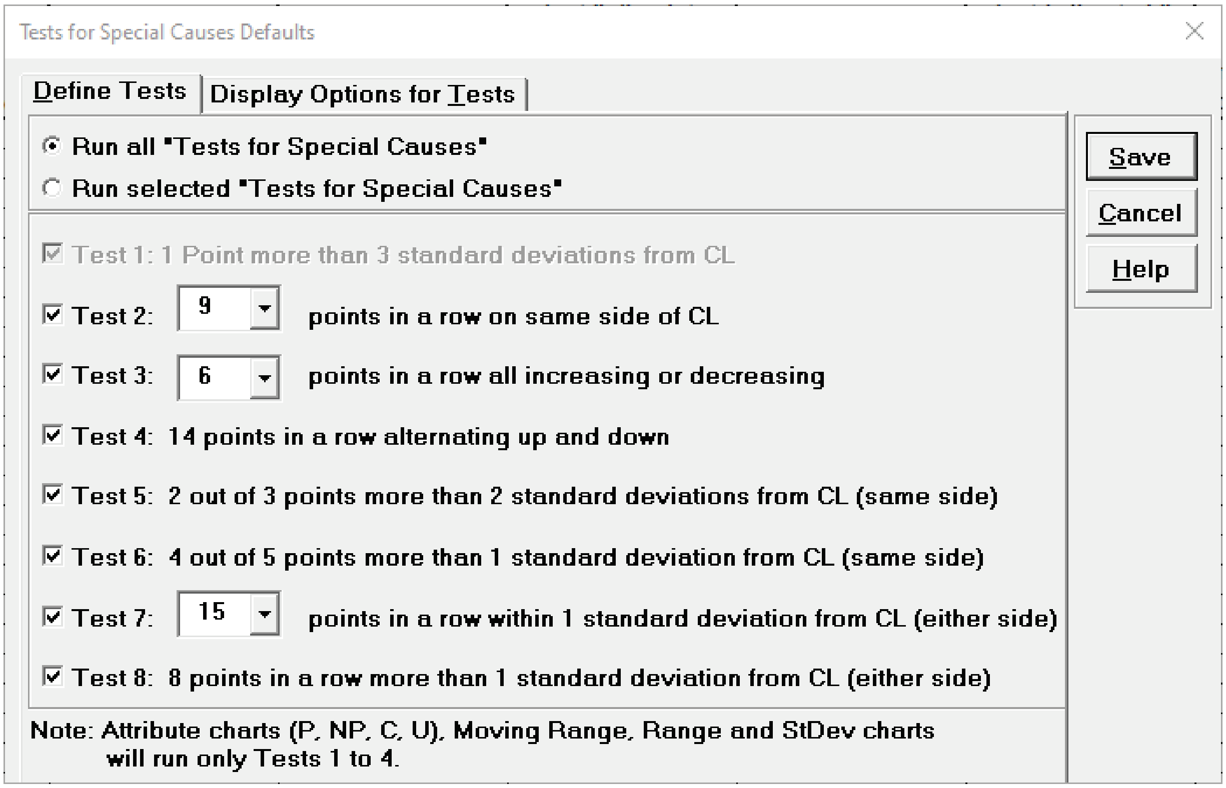
Some tests for special causes are indicated on the chart. If more than one test fails, the number corresponds to the first failed test.

12. There are no points that exceed the +/- 3 sigma limits on this chart, but we see some indication of instability with tests for special causes. The **Tests for Special Causes** report below the chart provides detailed information about each observation identified as a special cause. Note that the control chart also shows the +/- 1 sigma and +/- 2 sigma lines to aid in viewing these tests.

Tests for Special Causes - Individuals: Delivery Time Deviation
Number of Data Points Failing Tests = 30

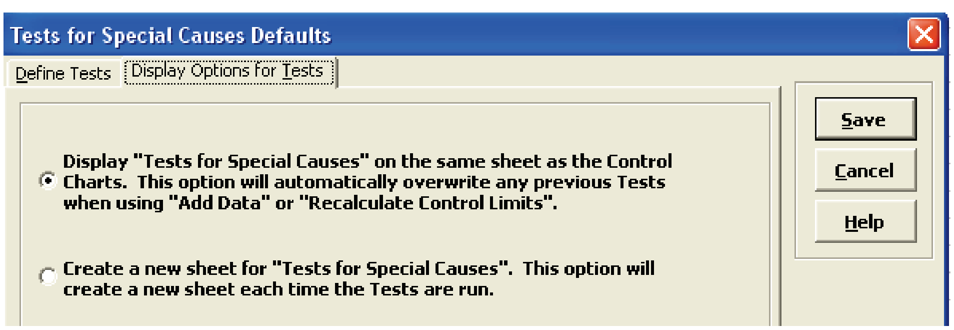
Observation No.	Test 1: 1 point more than 3 Stdev from CL	Test 2: 9 points in a row on same side of CL	Test 3: 6 points in a row all increasing or all decreasing	Test 4: 14 points in a row alternating up and down	Test 5: 2 out of 3 points more than 2 Stdev from CL (same side)	Test 6: 4 out of 5 points more than 1 Stdev from CL (same side)	Test 7: 15 points in a row within 1 Stdev from CL (either side)	Test 8: 8 points in a row more than 1 Stdev from CL (either side)
24			x					
109				x				
110				x				
111				x				
112				x				
157					x			
229							x	
230							x	
231							x	
232							x	
233							x	
234							x	
235							x	
236							x	
237							x	
238							x	
273						x		
499							x	
500							x	
501							x	
502							x	
503							x	
504							x	
505							x	
506							x	
507							x	
515		x						
516		x						
517		x						
565						x		

13. These tests for special causes can have defaults set to apply any or all of Tests 1-8. Test 2 can be set to 7, 8, or 9 points in a row on same side of CL. Test 3 can be set to 6 or 7 points in a row all increasing or decreasing. Test 7 can be set to 14 or 15 points in a row within 1 standard deviation from CL. Click **Sheet 1** Tab. Click **SigmaXL > Control Charts > "Tests for Special Causes" Defaults** to run selected tests for special causes:



Note that these defaults will apply to Individuals and X-bar charts. Test 1 to 4 settings will be applied to Attribute Charts and (if checked) Moving Range, Range, StDev and Attribute Charts.

14. Click on the Tab **Display Options for Tests**.



15. If you prefer to create a separate worksheet for each **Tests for Special Causes** report, choose **Create a new sheet** option. The default is to display **Tests for Special Causes** on the same sheet as the Control Chart. Note that this report will be overwritten when you add data or recalculate control limits.

16. Click **Save**.

17. Click on **Indiv Proc Cap** Tab to view the Process Capability Report, which includes potential (short term) capability indices Cp and Cpk:

Process Capability Report: Delivery Time Deviation	
Count	725
Mean	6.004
StDev (Overall, Long Term)	7.162
StDev (Within, Short Term)	7.202
USL	10
Target	0
LSL	-10

Capability Indices using Overall StDev	
Pp	0.47
Ppu	0.19
Ppl	0.74
Ppk	0.19
Cpm	0.36

Potential Capability Indices using Within StDev	
Cp	0.46
Cpu	0.18
Cpl	0.74
Cpk	0.18

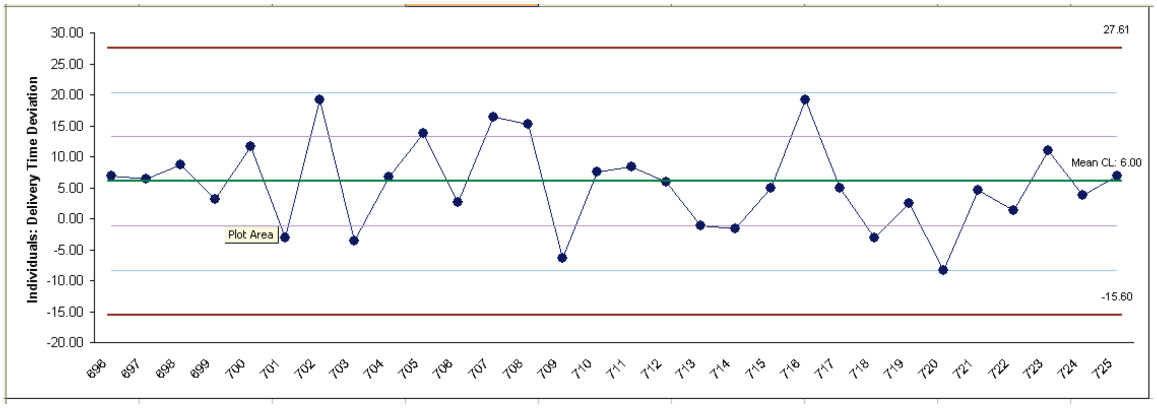
Expected Overall Performance	
ppm > USL	288409
ppm < LSL	12720.5
ppm Total	301129.8
% > USL	28.84%
% < LSL	1.27%
% Total	30.11%

Actual (Empirical) Performance	
% > USL	26.90%
% < LSL	1.38%
% Total	28.28%

Anderson-Darling Normality Test	
A-Squared	0.708616
P-Value	0.064143

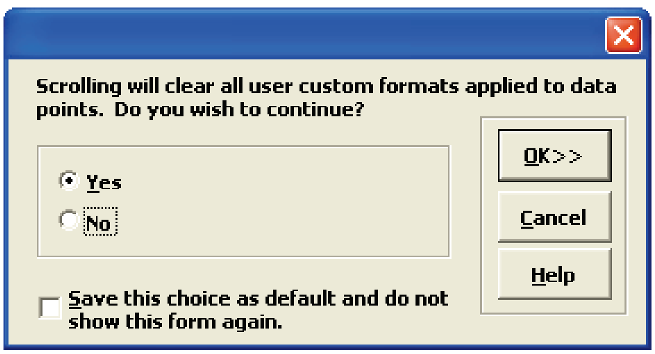
18. While this process demonstrated some slight instability on the control charts, the bigger issue was being late 6 minutes on average and having a Standard Deviation of 7.2 minutes! One improvement implemented was rescheduling the service elevators so that Room Service and Maintenance were not both trying to use them during peak times.

19. Click on the **Indiv** sheet. With 725 data points, you may want to have a closer look at the most recent data. To do this, click **SigmaXL Chart Tools > Show Last 30 Points**. (If this menu item does not appear, click on any cell adjacent to the chart.) The resulting chart is shown:



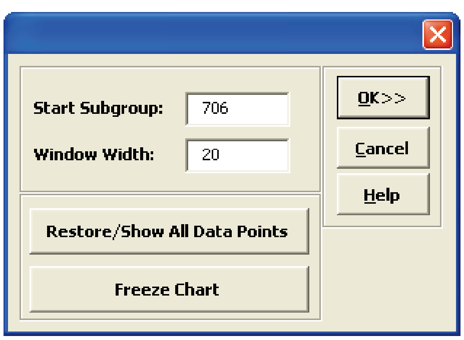
20. To reset the chart, click **SigmaXL Chart Tools > Show All Data Points**.

21. To enable scrolling, click **SigmaXL Chart Tools > Enable Scrolling**. A warning message is given:



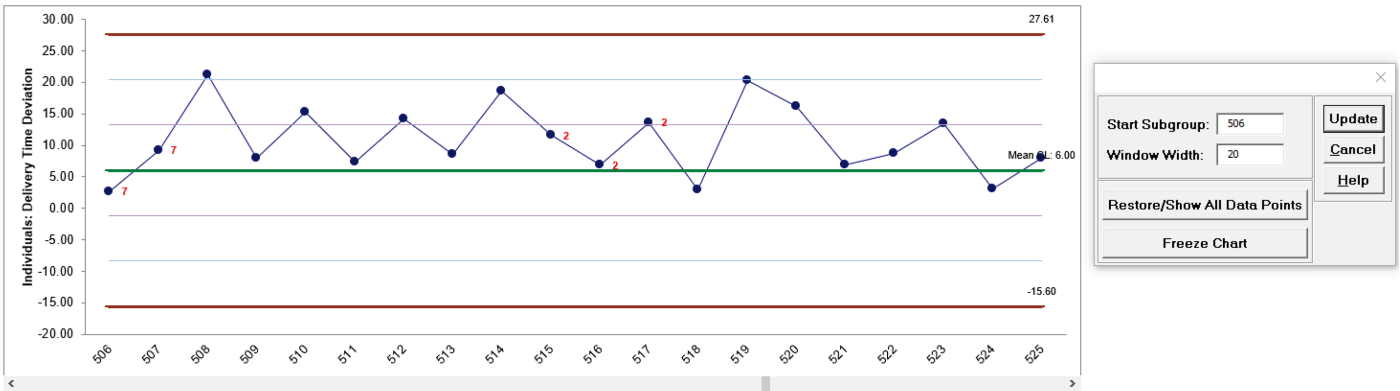
Scrolling will clear user custom formats, but does not affect Tests for Special Causes.

22. Click **OK**. A scroll dialog appears allowing you to specify the **Start Subgroup** and **Window Width**:



23. At any point, you can click **Restore/Show All Data Points** or **Freeze Chart**. Freezing the chart will remove the scroll and unload the dialog. The scroll dialog will also unload if you change worksheets. To restore the dialog, click **SigmaXL Chart Tools > Enable Scrolling**.

24. Click **OK**. The control chart appears with the scroll bar beneath it. You can also change the **Start Subgroup** and **Window Width** and **Update**.



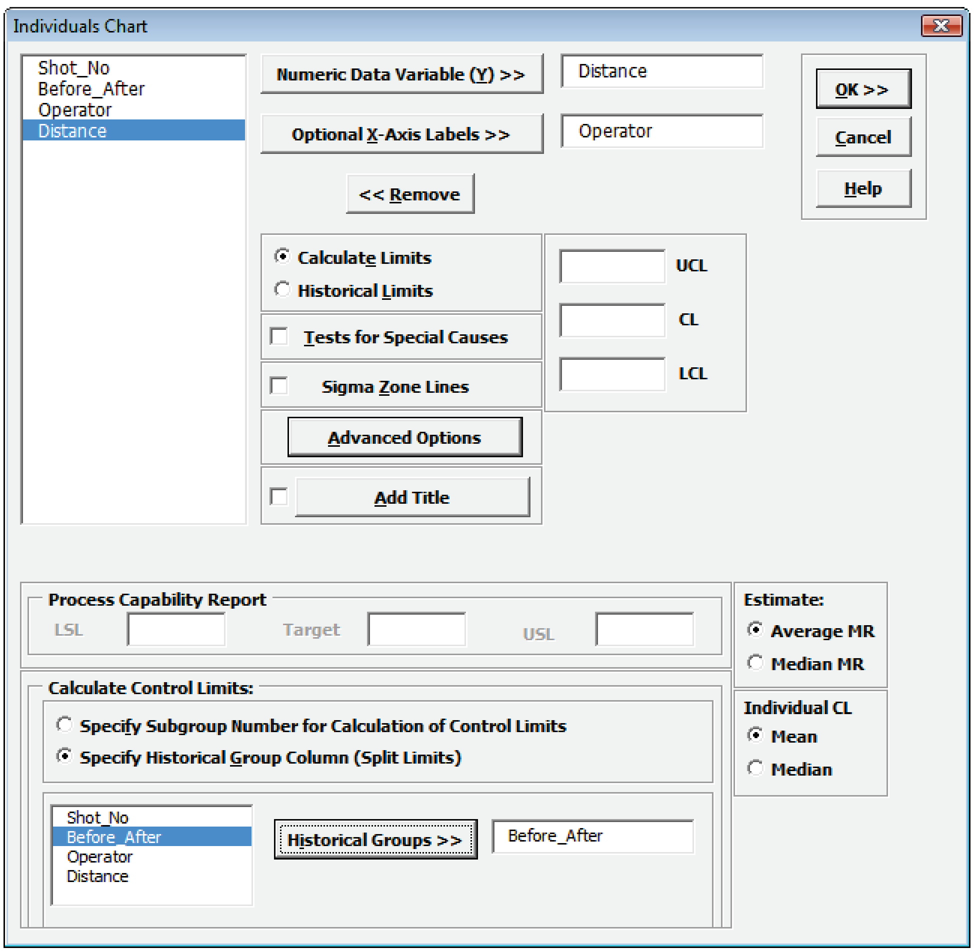
Individuals Charts: Advanced Limit Options – Historical Groups

1. Open **Catapult Data - Before After Improvement.xlsx**. Click **Sheet 1 Tab**. This data set contains Catapult firing distances. Before_After denotes before improvement and after improvement. The target distance is 100 inches, with the goal being to hit the target and minimize variation about the target. We would like to use an individuals control chart with historical groups to split the limits demonstrating the before versus after improvement.

2. Click **SigmaXL > Control Charts > Individuals**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

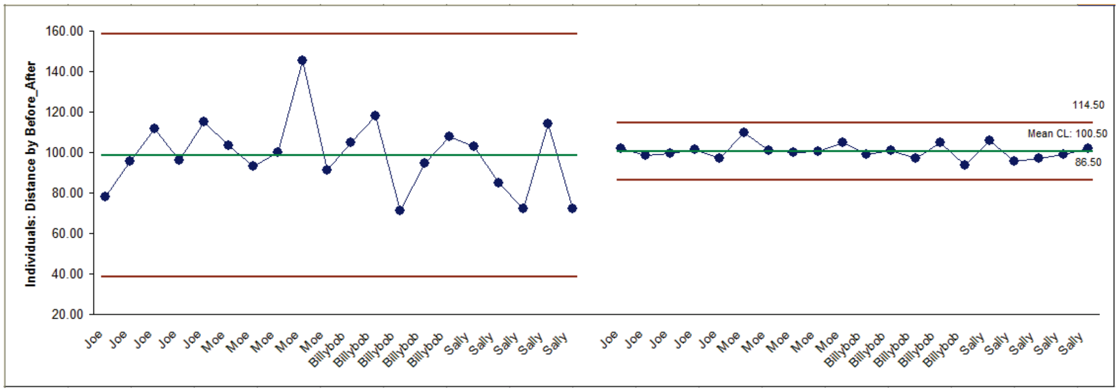
3. Select *Distance*, click **Numeric Data Variable (Y) >>**; select *Operator*, click **Optional X-Axis Labels >>**.

4. Click **Advanced Options**. Select **Specify Historical Groups**. Select *Before_After*, click **Historical Groups >>**.



Note: Process Capability analysis is not permitted when Historical Groups are used.

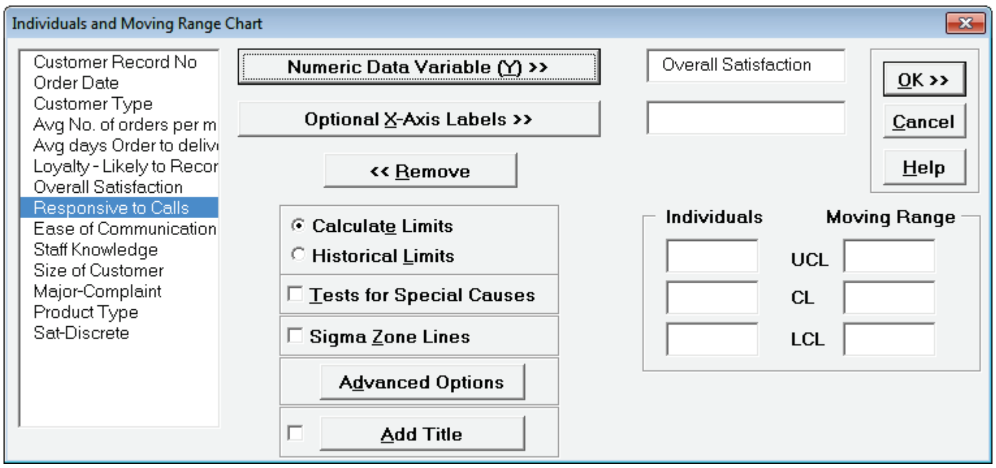
5. Click **OK**. The resulting Individuals Control Chart with split limits based on historical groups is shown, demonstrating a clear process improvement:



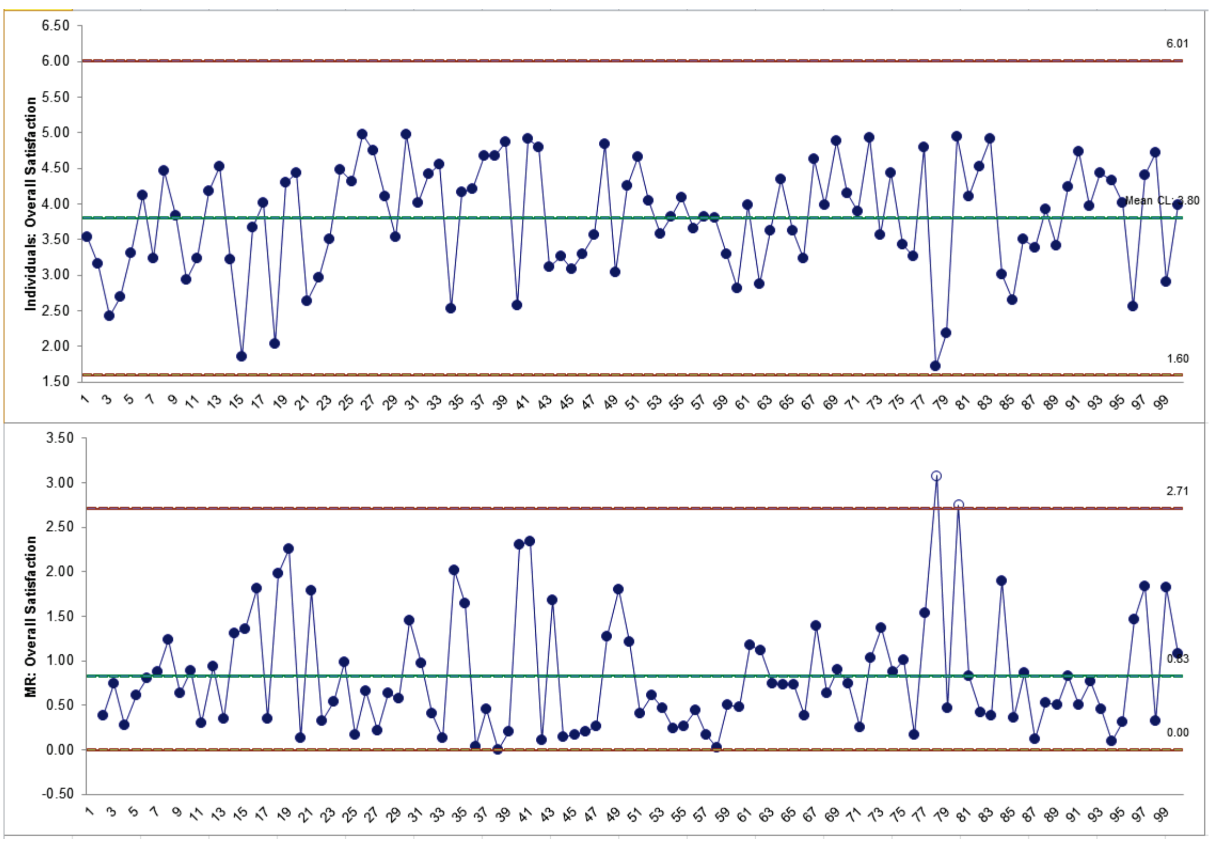
Individuals & Moving Range Charts

1. Open **Customer Data.xlsx**, click on **Sheet 1**. Click **SigmaXL>Control Charts>Individuals & Moving Range**. Ensure that entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.

2. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>.** Ensure that **Calculate Limits** is selected.



3. Click **OK**. Resulting Individuals & Moving Range control chart:



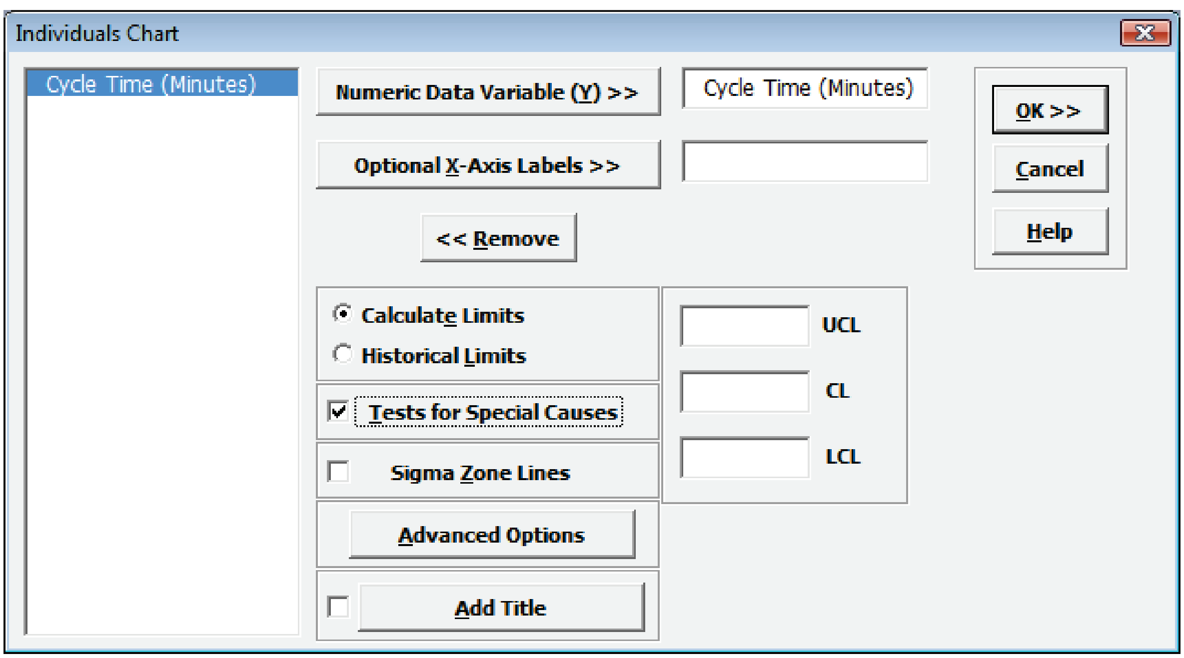
Individuals Charts for Nonnormal Data (Box-Cox Transformation)

An important assumption for Individuals Charts is that the data be normally distributed (unlike the X-Bar Chart which is robust to nonnormality due to the central limit theorem). If the data is inherently nonnormal (i.e., the nonnormality is not due to an outlier or assignable cause), the Box-Cox Transformation tool can be used to convert nonnormal data to normal by applying a power transformation. The Johnson transformation and other distributions may also be used with Automatic Best Fit (see **Measure Phase Tools, Part J – Process Capability, Capability Combination Report Individuals Nonnormal**).

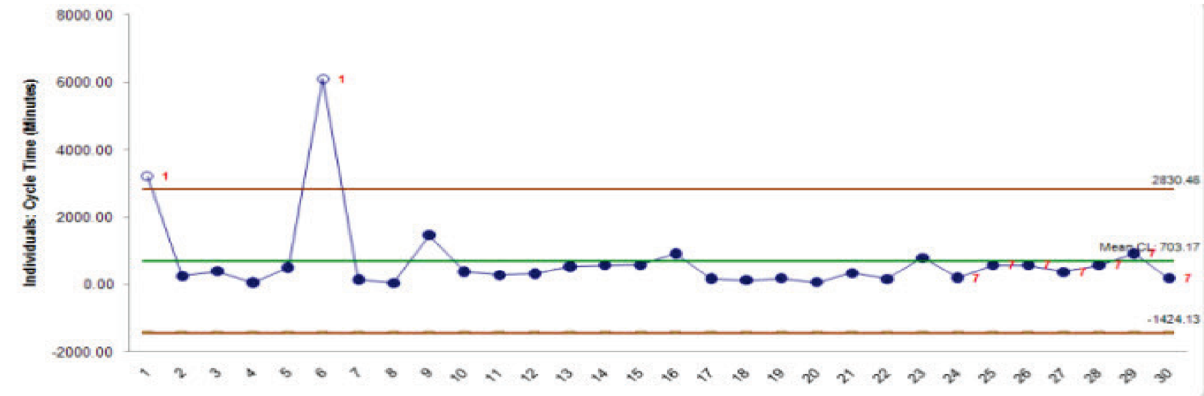
1. Open the file **Nonnormal Cycle Time2.xlsx**. This contains continuous nonnormal data of process cycle times. We performed a Process Capability study with this data earlier in the Measure Phase, Part H.

2. Initially, we will ignore the nonnormality in the data and construct an Individuals Chart. Click **SigmaXL > Control Charts > Individuals**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>**. Select **Calculate Limits**. Check **Tests for Special Causes**. Click **OK**.



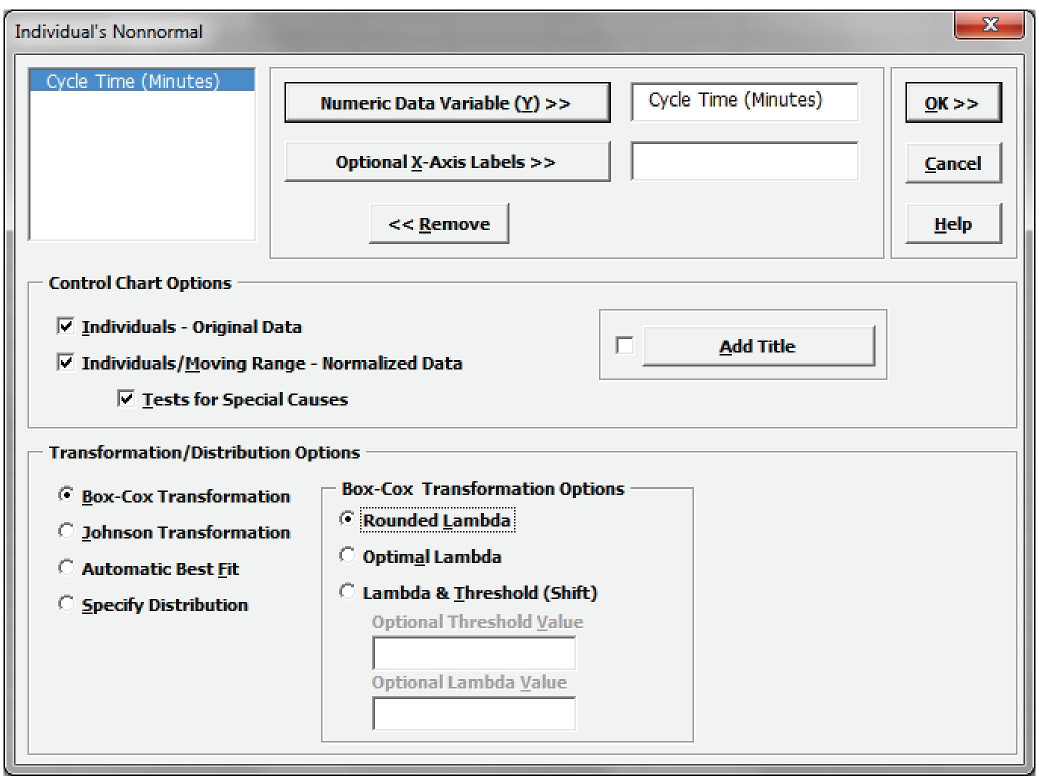
4. The resulting Individuals Chart is shown:



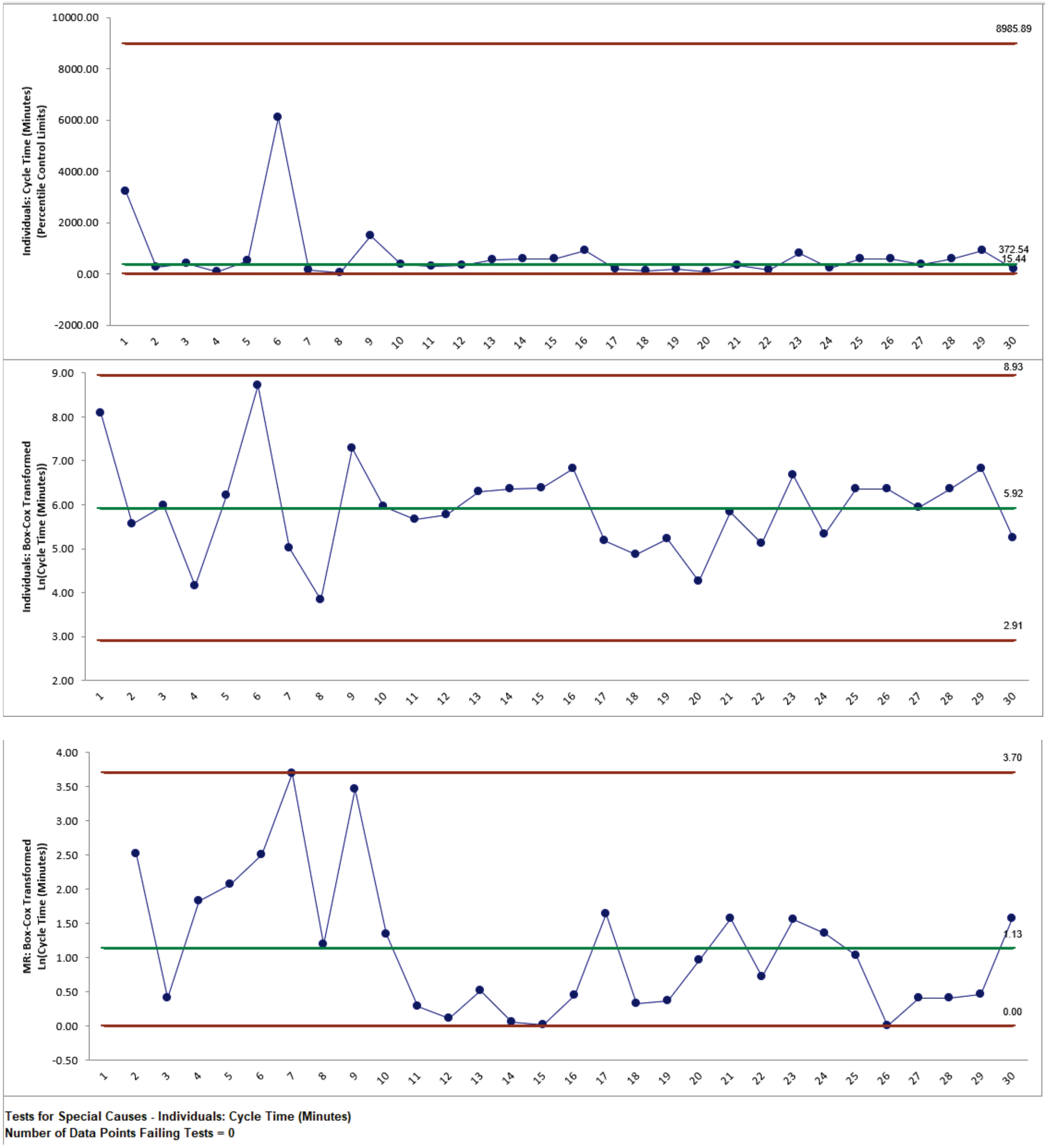
This chart clearly shows that the process is "out-of-control". But is it really? Nonnormality can cause serious errors in the calculation of Individuals Chart control limits, triggering false alarms (Type I errors) or misses (Type II errors).

5. We will now construct Individuals Control Charts for nonnormal data. Select **Sheet 1** Tab (or press **F4**). Click **SigmaXL > Control Charts > Nonnormal > Individuals Nonnormal.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

6. Select *Cycle Time (Minutes)*, click **Numeric Data Variable (Y) >>.** We will use the default selection for **Transformation/Distribution Options: Box-Cox Transformation** with **Rounded Lambda**. Check **Tests for Special Causes** as shown:



7. Click **OK**. The resulting control charts are shown below:



Note that there are no out-of-control signals on the control charts, so the signals observed earlier when normality was assumed, were false alarms.

The **Individuals – Original Data** chart displays the untransformed data with control limits calculated as:

UCL = 99.865 percentile

CL = 50th percentile

LCL = 0.135 percentile

The benefit of displaying this chart is that one can observe the original untransformed data. Since the control limits are based on percentiles, this represents the overall, long term variation rather than the typical short term variation. The limits will likely be nonsymmetrical.

The **Individuals/Moving Range – Normalized Data** chart displays the transformed z-values with control limits calculated using the standard Shewhart formulas for Individuals and Moving Range charts. The benefit of using this chart is that tests for special causes can be applied and the control limits are based on short term variation. The disadvantage is that one is observing transformed data on the chart rather than the original data.

Part B - X-Bar & Range/StDev Charts

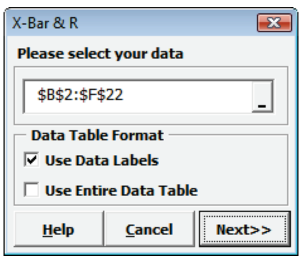
Tip: See Part E – Control Chart Selection Guide. The Control Chart Selection Guide makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

X-Bar & R Charts

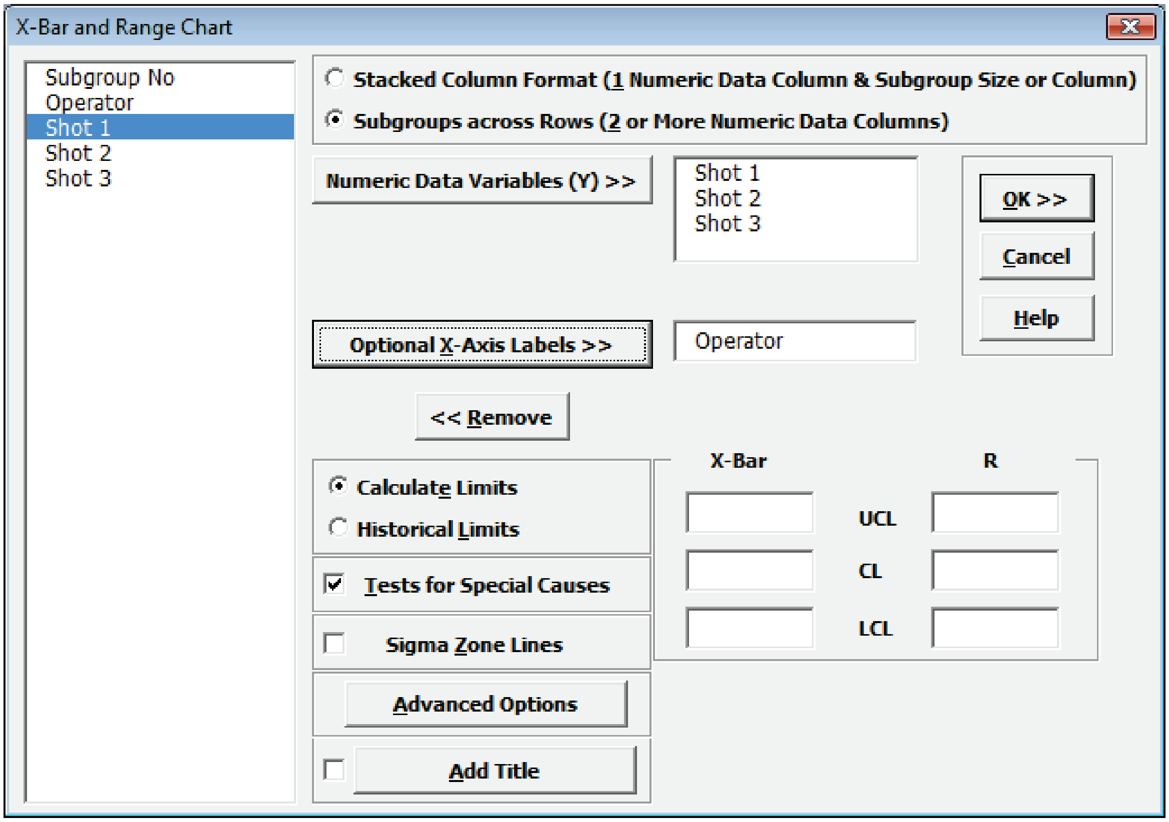
1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the ball 3 times. The target distance is 100 inches. The Upper Specification Limit (USL) is 108 inches. The Lower Specification Limit (LSL) is 92 inches.

2. Select **B2:F22** (if not already selected); here, we will only use the first 20 subgroups to determine the control limits.

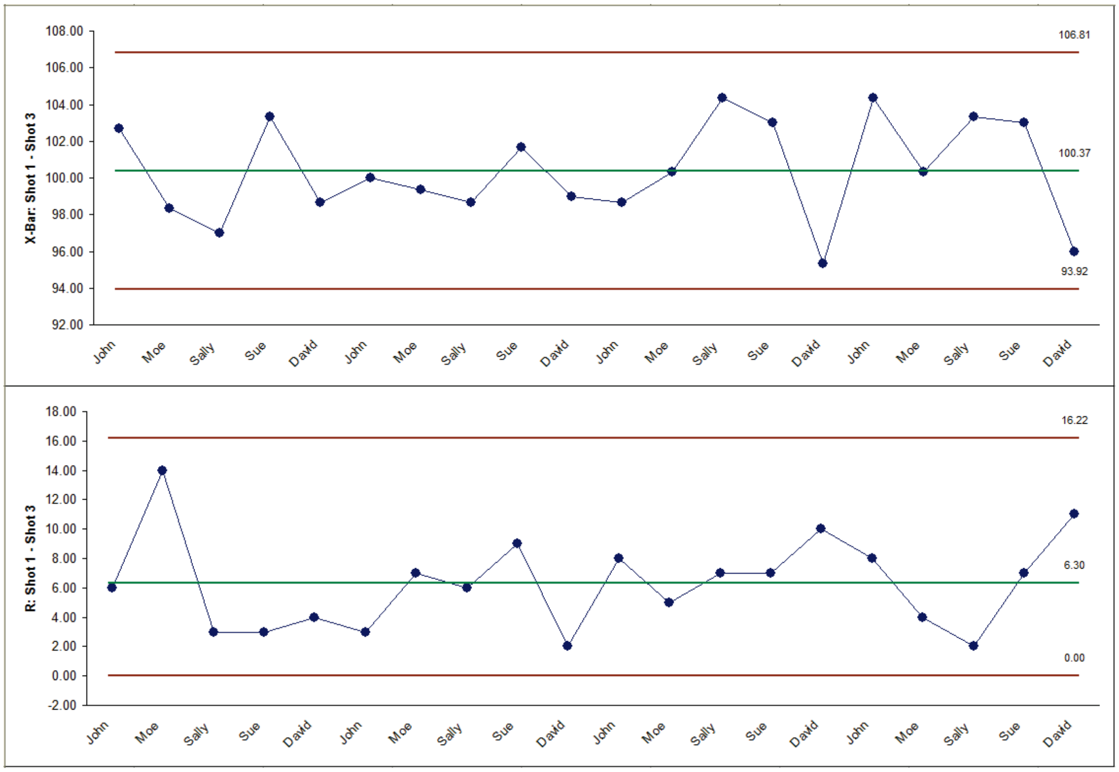
3. Select **SigmaXL > Control Charts > X-Bar & R**. Do not check **Use Entire Data Table**!



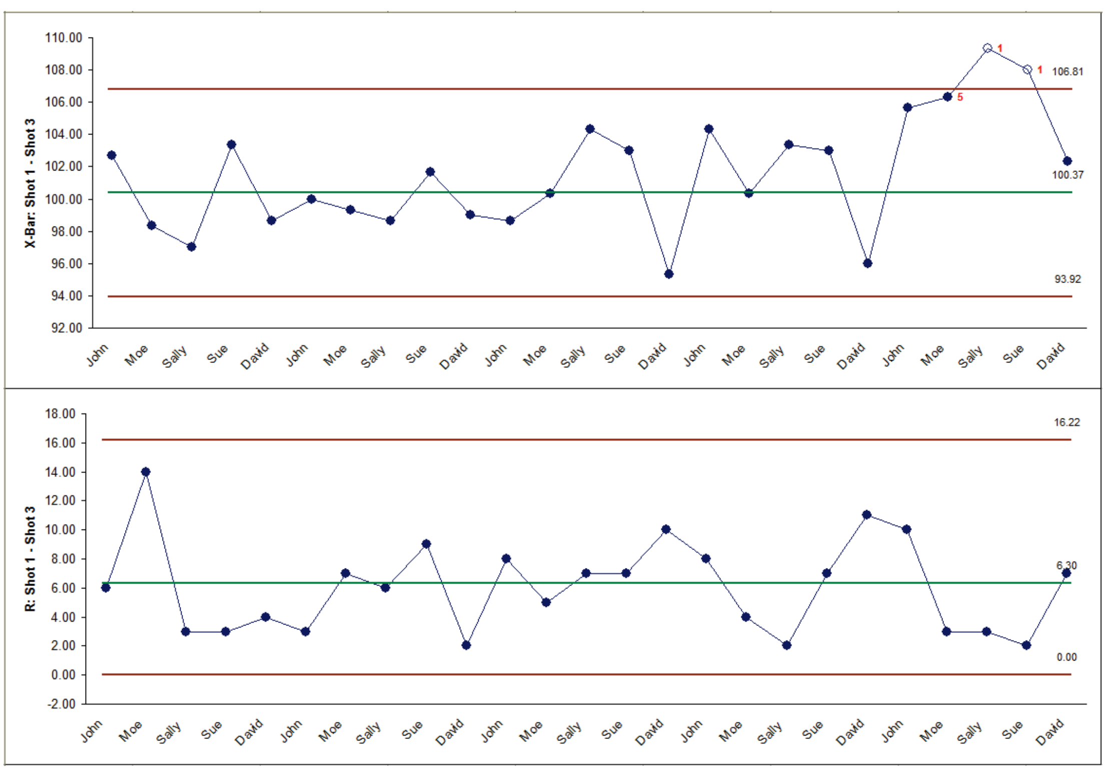
4. Click **Next**. Select **Subgroups across Rows**, select *Shot 1*, *Shot 2*, *Shot 3*, click **Numeric Data Variables (Y) >>**; select *Operator*, click **Optional X-Axis Labels >>**. Check **Tests for Special Causes** as shown:


X-Bar and Range Chart
Subgroup No
Operator
Shot 1
Shot 2
Shot 3
Stacked Column Format (1 Numeric Data Column & Subgroup Size or Column)
Subgroups across Rows (2 or More Numeric Data Columns)
Numeric Data Variables (Y) >>
Shot 1
Shot 2
Shot 3
OK >>
Cancel
Help
Optional X-Axis Labels >>
Operator
<< Remove
X-Bar
R
Calculate Limits
Historical Limits
UCL
CL
LCL
Tests for Special Causes
Sigma Zone Lines
Advanced Options
Add Title

5. Click **OK**. Resulting X-bar & R charts:



6. This is currently a stable catapult process. Subgroups 21 to 25 were added afterwards. To add the additional data to this chart, click **SigmaXL Chart Tools > Add Data to this Control Chart** Add Data to this Control Chart as shown:



7. Note that the Add Data button does NOT recalculate the control limits. Once control limits are established, they should only be recalculated when a deliberate process change or improvement is introduced.

8. The **Tests for Special Causes** report gives us more detail on the recent instability:

Tests for Special Causes - X-Bar: Shot 1 - Shot 3
Number of Data Points Failing Tests = 3

Observation No.	Test 1: 1 point more than 3 StDev from CL	Test 2: 9 points in a row on same side of CL	Test 3: 6 points in a row all increasing or all decreasing	Test 4: 14 points in a row alternating up and down	Test 5: 2 out of 3 points more than 2 StDev from CL (same side)	Test 6: 4 out of 5 points more than 1 StDev from CL (same side)	Test 7: 15 points in a row within 1 StDev from CL (either side)	Test 8: 8 points in a row more than 1 StDev from CL (either side)
22					x	x		
23	x				x	x		
24	x				x	x		

The X-bar chart and **Tests for Special Causes** report clearly shows that this process is now out of control with an unstable mean. The process must be stopped, and the Out-of-Control Action Plan must be followed to determine and fix the root cause. In this case, the assignable cause was a change of rubber band requiring a reset of the pull back angle. The use of tests for special causes gave us an early warning of this at observation number 22.

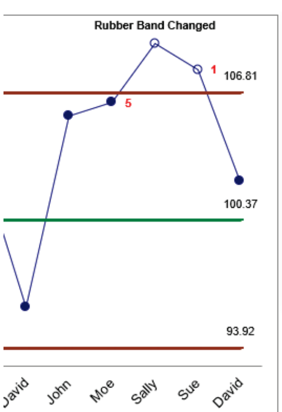
Note that the Range chart is in-control even though the X-Bar chart is out-of-control:

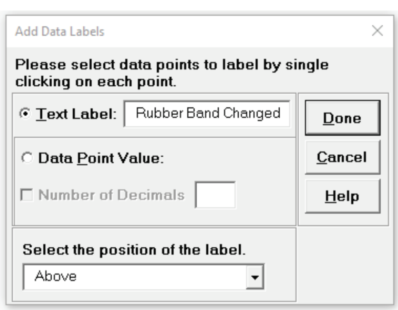
Tests for Special Causes - Range: Shot 1 - Shot 3
Number of Data Points Failing Tests = 0

Observation No.	Test 1: 1 point more than 3 StDev from CL	Test 2: 9 points in a row on same side of CL	Test 3: 6 points in a row all increasing or all decreasing	Test 4: 14 points in a row alternating up and down

9. The tests for special causes can have defaults set to apply any or all of Tests 1-8. Test 2 can be set to 7, 8, or 9 points in a row on same side of CL. Click **SigmaXL > Control Charts > "Tests for Special Causes" Defaults** to run selected tests for special causes. Note that these defaults will apply to Individuals and X-bar charts. Test 1 to 4 settings will be applied to Attribute Charts and (if checked) Moving Range, Range, StDev. and Attribute Charts.

10. To add a comment to a data point, select **SigmaXL Chart Tools > Add Data Label.** Select **Text Label.** Enter a comment as shown. Click on the data point to add the comment. Click **Done**.



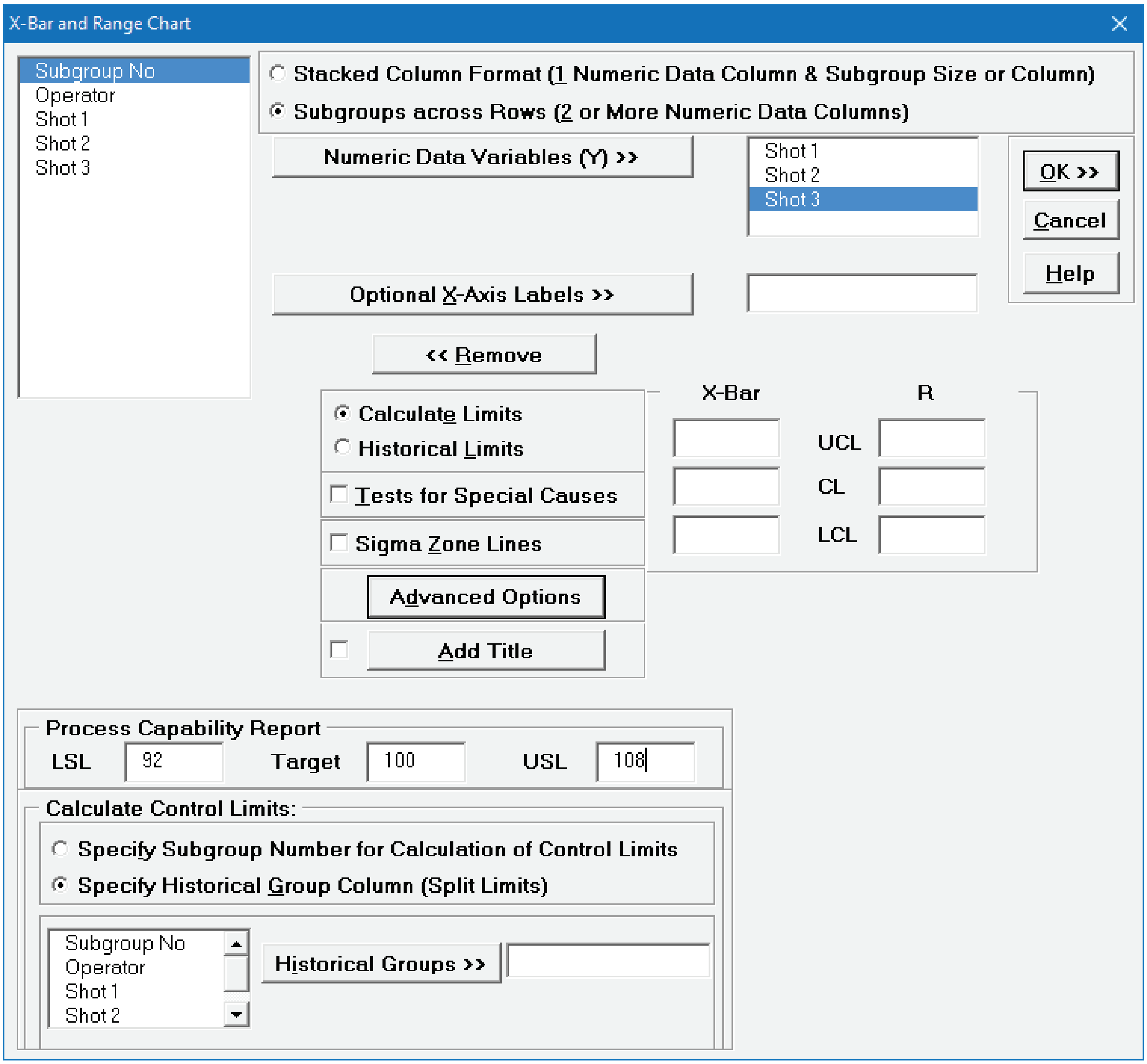


11. Now we will look at Process Capability Indices for this process. Click on **Sheet 1** (or press **F4** to activate last worksheet). Click **SigmaXL > Control Charts > X-Bar & R**. Check **Use Entire Data Table**. Click **Next**. (Alternatively select **B2:F27**, press **F3**.)

12. Select *Shots 1-3*, click **Numeric Data Variables (Y) >>**.

13. Click **Advanced Options**. Enter **LSL** = 92, **Target** = 100, **USL** = 108.

14. The resulting dialog box settings are shown:



15. Click **OK**. Click **X-Bar & R – Proc Cap** sheet for the Process Capability report:

Process Capability Report: X-Bar: Shot 1 - Shot 3	
Count	75
Mean	101.56
StDev (Overall, Long Term)	4.616
StDev (Within, Short Term)	3.568
USL	108
Target	100
LSL	92

Capability Indices using Overall StDev	
Pp	0.58
Ppu	0.47
Ppl	0.69
Ppk	0.47
Cpm	0.55

Potential Capability Indices using Within StDev	
Cp	0.75
Cpu	0.60
Cpl	0.89
Cpk	0.60

Expected Overall Performance	
ppm > USL	81468
ppm < LSL	19168.4
ppm Total	100636.0
% > USL	8.15%
% < LSL	1.92%
% Total	10.06%

Actual (Empirical) Performance	
% > USL	5.33%
% < LSL	4.00%
% Total	9.33%

Anderson-Darling Normality Test	
A-Squared	0.486247
P-Value	0.2192

Note the difference between Pp and Cp; Ppk and Cpk. This is due to the process instability. If the process was stable, the actual performance indices Pp and Ppk would be closer to the Cp and Cpk values.

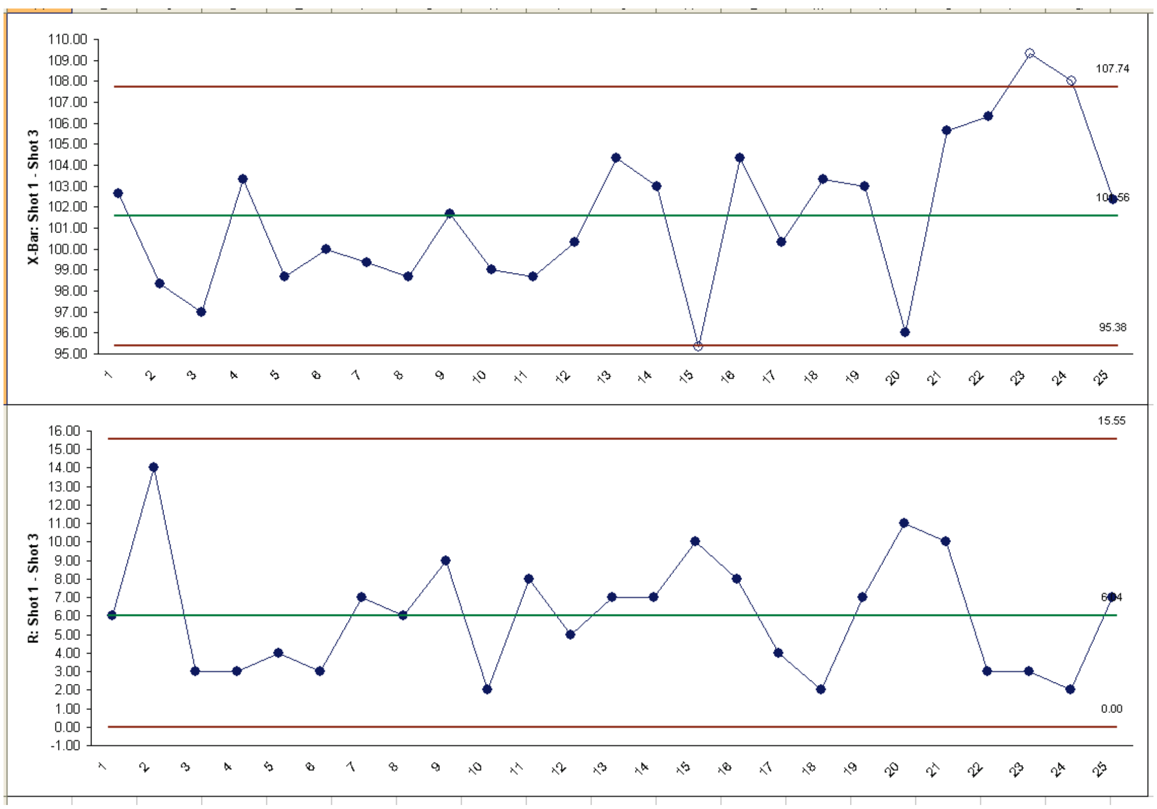
X-Bar & R Charts – Exclude Subgroups

After creating a control chart, you can specify subgroups (or rows) to exclude by using the **Exclude Data** tool**.**

16. Click on **Sheet 1** (or press **F4** to activate last worksheet). Click **SigmaXL > Control Charts > X-Bar & R**. Check **Use Entire Data Table**. Click **Next**.

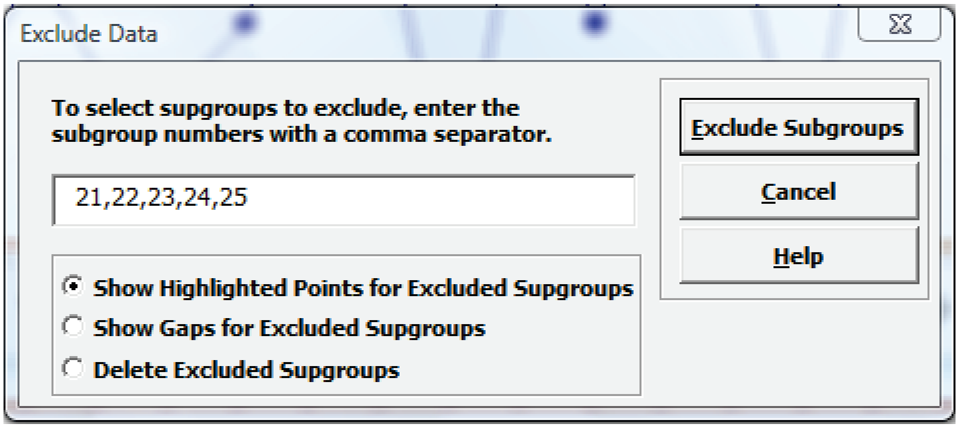
17. Select *Shots 1-3*, click **Numeric Data Variables (Y) >>**. Ensure that **Calculate Limits** is selected. Click **OK**.

18. The resulting X-bar & R charts are displayed:

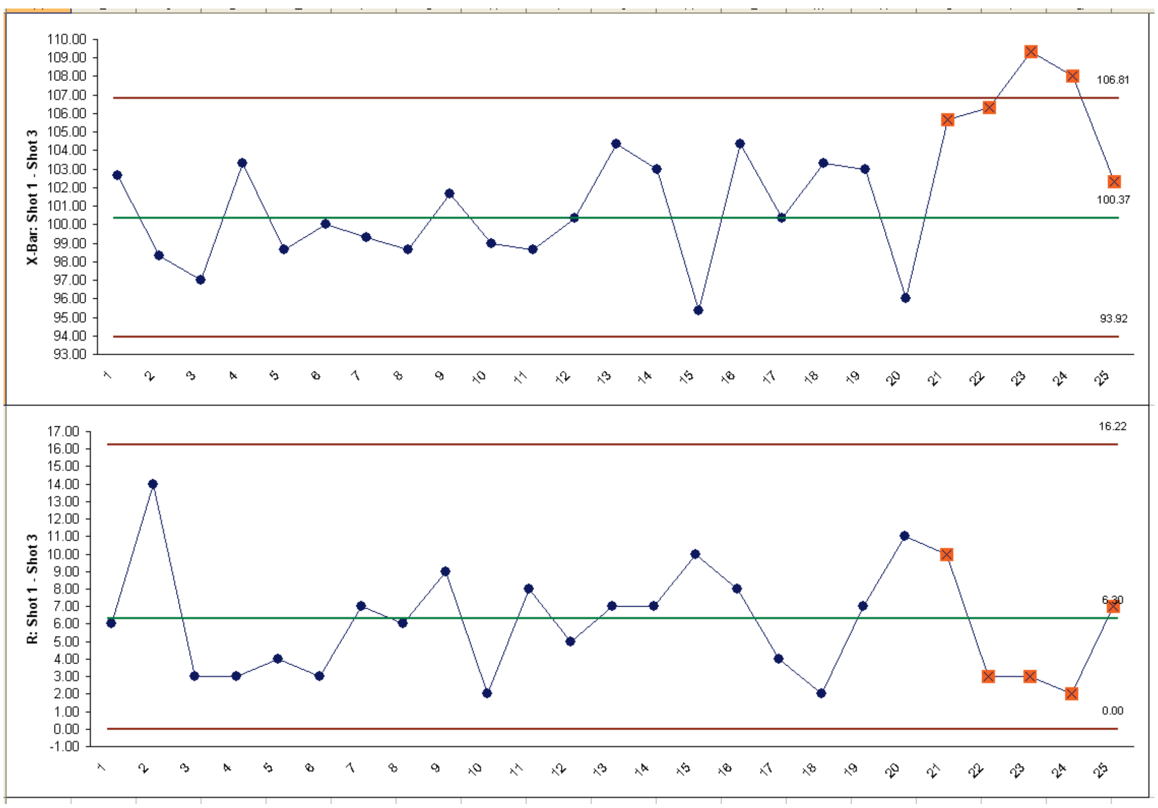


The control limits here were calculated including subgroups 21 to 25 which have a known assignable cause.

19. To calculate the control limits excluding subgroups 21 to 25, click **SigmaXL Chart Tools > Exclude Subgroups**. Select **Show Highlighted Points for Excluded Subgroups**. Enter 21,22,23,24,25 as shown:



20. Click **Exclude Subgroups.** The control chart limits are recalculated and the excluded points are highlighted:



Tip: You can also choose to show gaps for excluded subgroups or delete excluded subgroups from the charts.

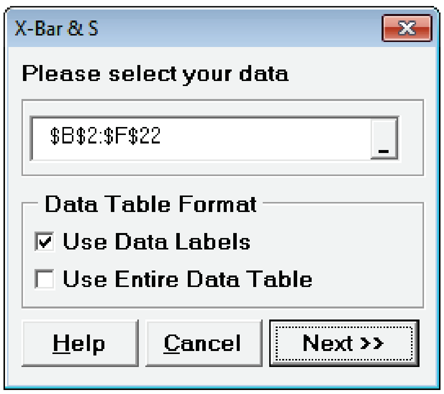
X-Bar & S Charts

1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the ball 3 times. The target distance is 100 inches. The Upper Specification Limit (USL) is 108 inches. The Lower Specification Limit (LSL) is 92 inches.

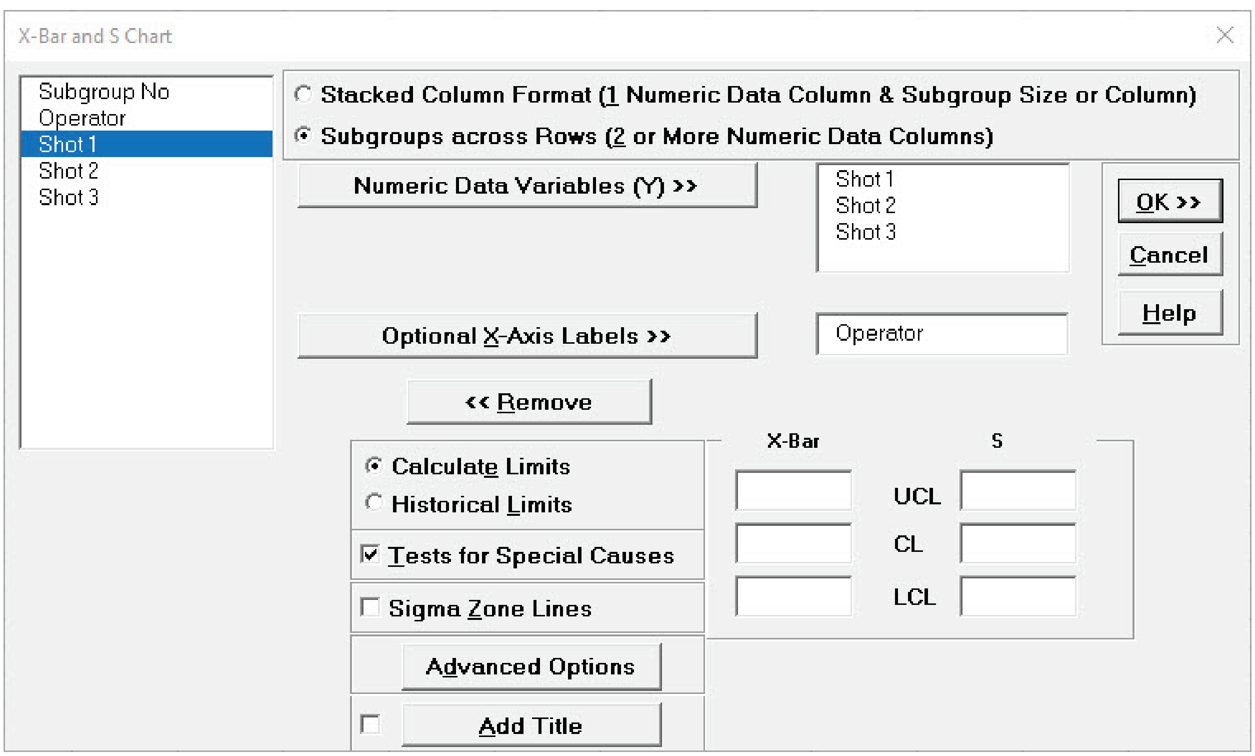
2. Select **B2:F22**; here, we will only use the first 20 subgroups to determine the control limits.

3. Select **SigmaXL > Control Charts > X-Bar & S**.

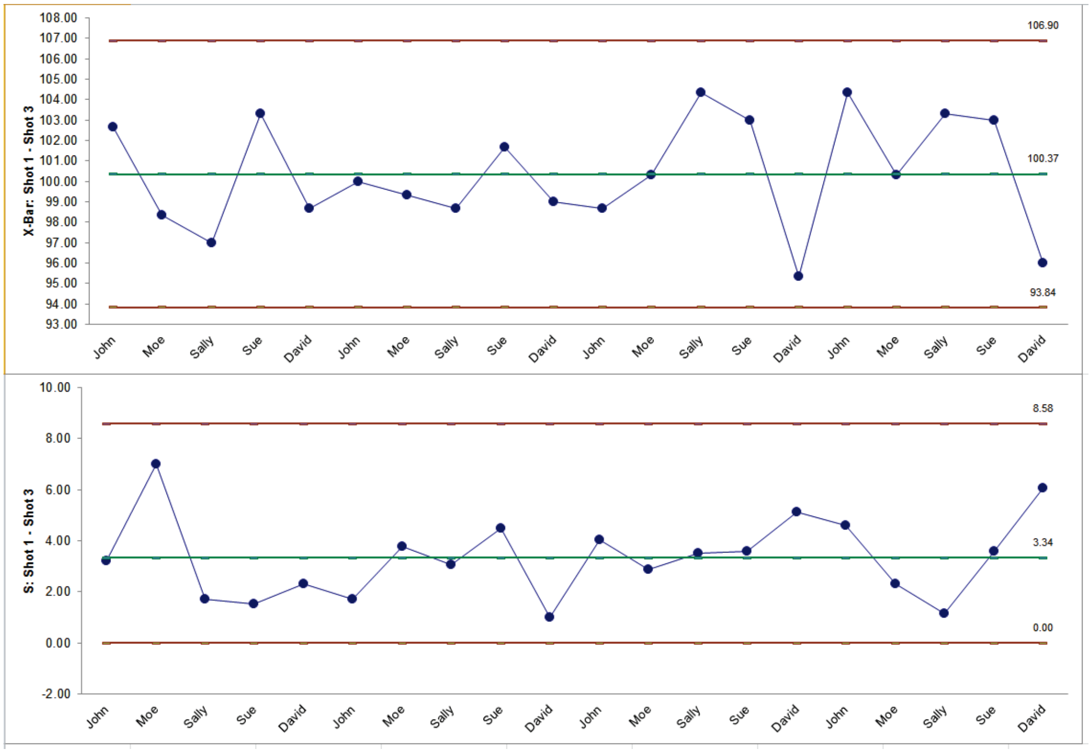
4. Do not check **Use Entire Data Table**!



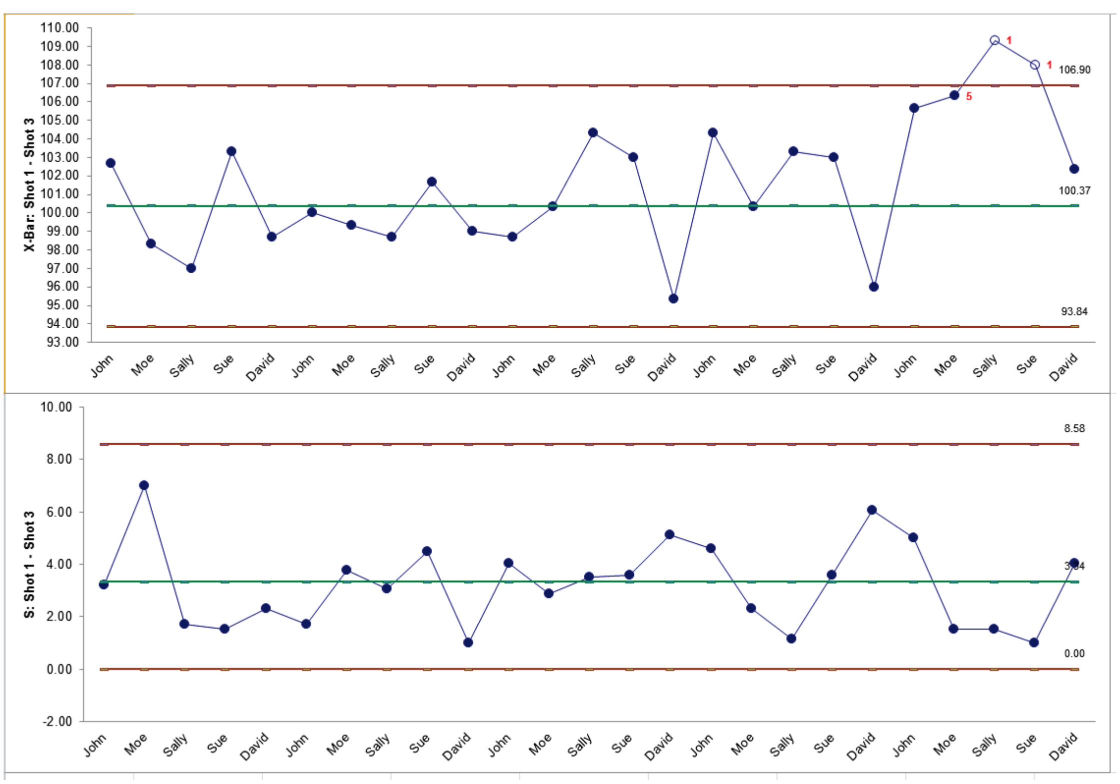
5. Click **Next**. Select **Subgroups across Rows**, select *Shot 1*, *Shot 2*, *Shot 3*, click **Numeric Data Variables (Y) >>**; select *Operator*, click **Optional X-Axis Labels >>**. Check **Tests for Special Causes** as shown:



6. Click **OK**. Resulting X-bar & S charts:



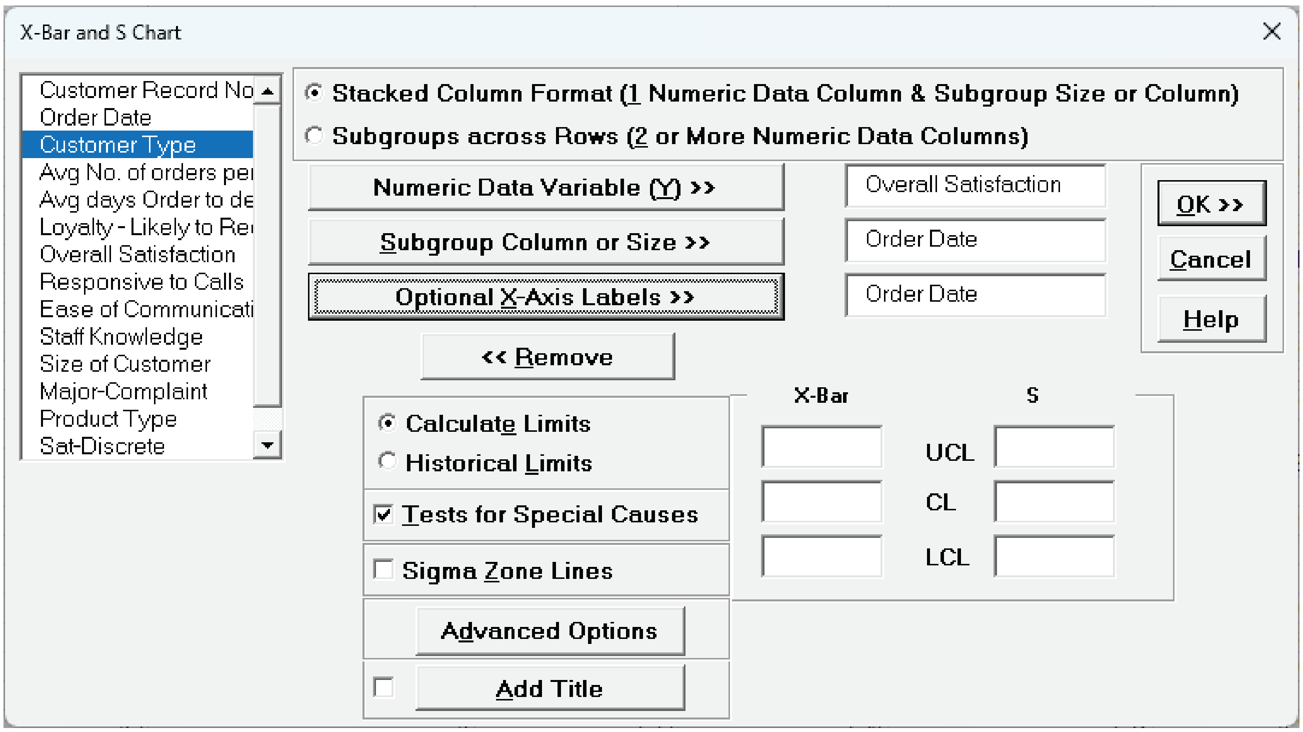
7. This is currently a stable catapult process. Subgroups 21 to 25 were added afterwards. To add the additional data to this chart, click **SigmaXL Chart Tools > Add Data to this Control Chart** Add Data to this Control Chart as shown:



X-Bar & S Charts – Stacked Column Format

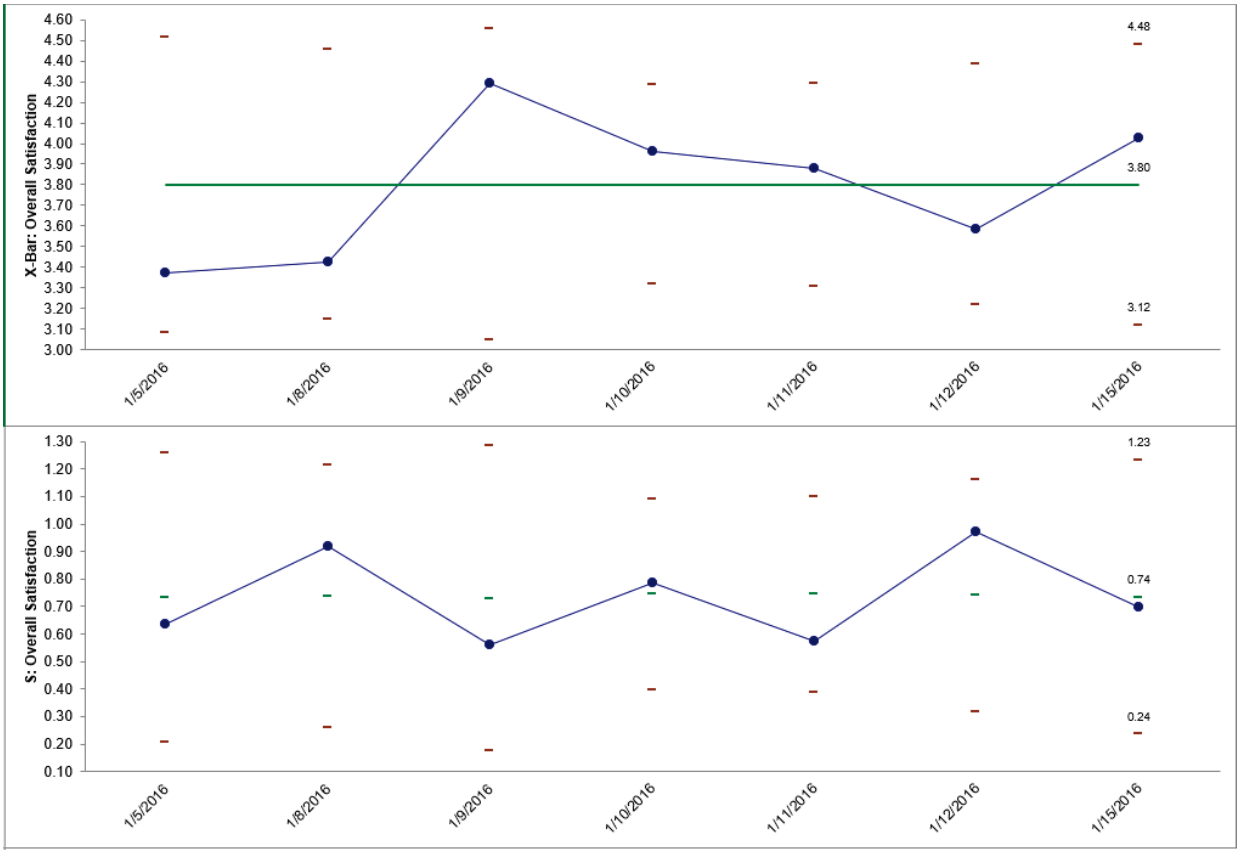
1. Open **Customer Data.xlsx**, click on **Sheet 1**. Click **SigmaXL > Control Charts > X-Bar & S**. Ensure that entire data table is selected. If not, check Use **Entire Data Table**. Click **Next**.

2. Select **Stacked Column Format**, select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Order Date*, click **Subgroup Column or Size >>**; select *Order Date*, click **Optional X-Axis Labels >>**. Ensure that **Calculate Limits** is selected. Check **Tests for Special Causes** as shown:



Note: Subgroups are always displayed in the same order as given in the rows. They are not sorted by Order Date.

3. Click **OK**. The resulting X-bar & S charts are shown:



The moving limits are due to the varying subgroup sizes. This appears to be a stable process, but typically one would want a minimum of 20 subgroups when creating a control chart. This example is for demonstration purposes only.

Part C – Attribute Charts (P, NP, C, U)

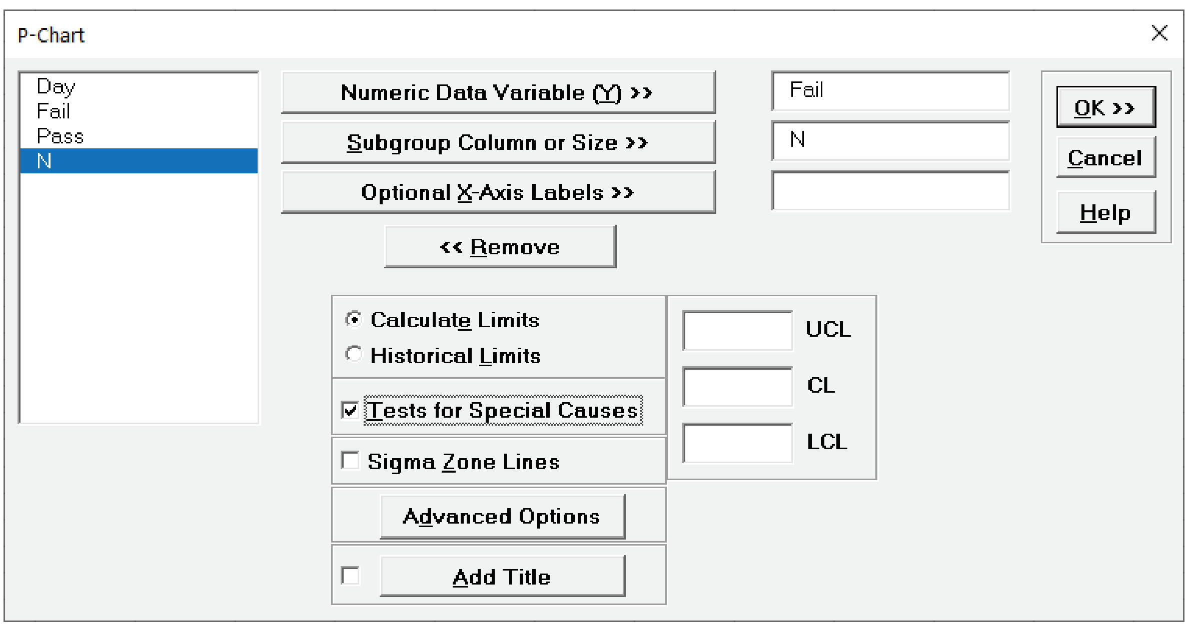
Tip: See Part E – Control Chart Selection Guide. The Control Chart Selection Guide makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

P-Charts

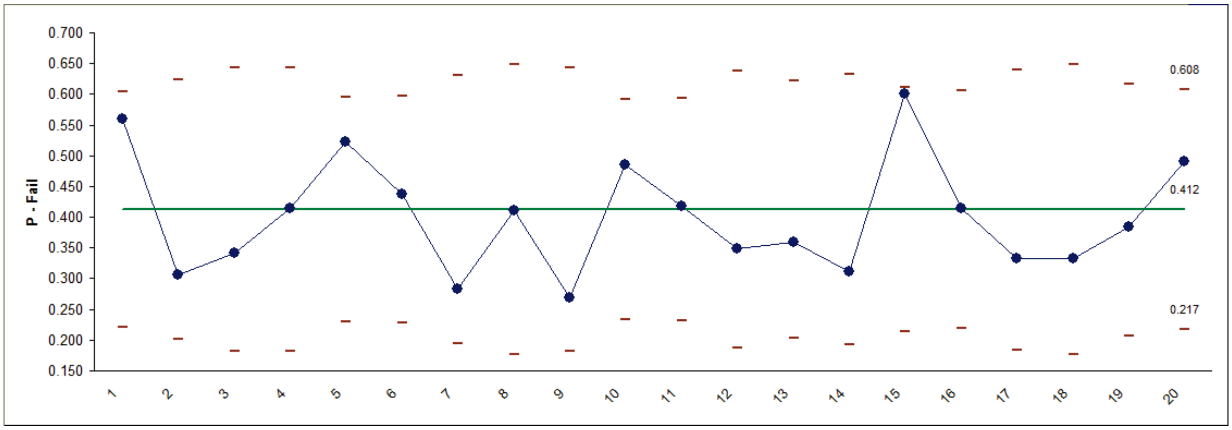
1. Open **New York Daily Cycle Time – Discrete.xlsx**. This is data from the Sigma Savings and Loans Company, New York location. Each day, the cycle time (in days) for completed loans and leases was recorded. *N* indicates the number of loans counted. A *Fail* was recorded if the cycle time exceeded the critical customer requirement of 8 days. Note that we are not recommending that continuous data be converted to discrete data in this manner, but rather using this data to illustrate the use of P charts for Discrete or Attribute data. P Charts (for Defectives) can have fixed or varying subgroup sizes.

2. Select **SigmaXL > Control Charts > Attribute Charts > P**. Ensure that **B3:E23** are selected, click **Next**.

3. Select *Fail*, click **Numeric Data Variable (Y) >>**; select *N*, click **Subgroup Column or Size >>**. If we had a fixed subgroup size, the numerical value of the subgroup size could be entered instead of Column N. Check **Tests for Special Causes**.



4. Click **OK**. The resulting P-Chart is shown:



The moving limits are due to the varying subgroup sizes. While this P-chart shows stability, a much bigger concern is the average 41% failure rate to deliver the loans/leases in 8 days or less!

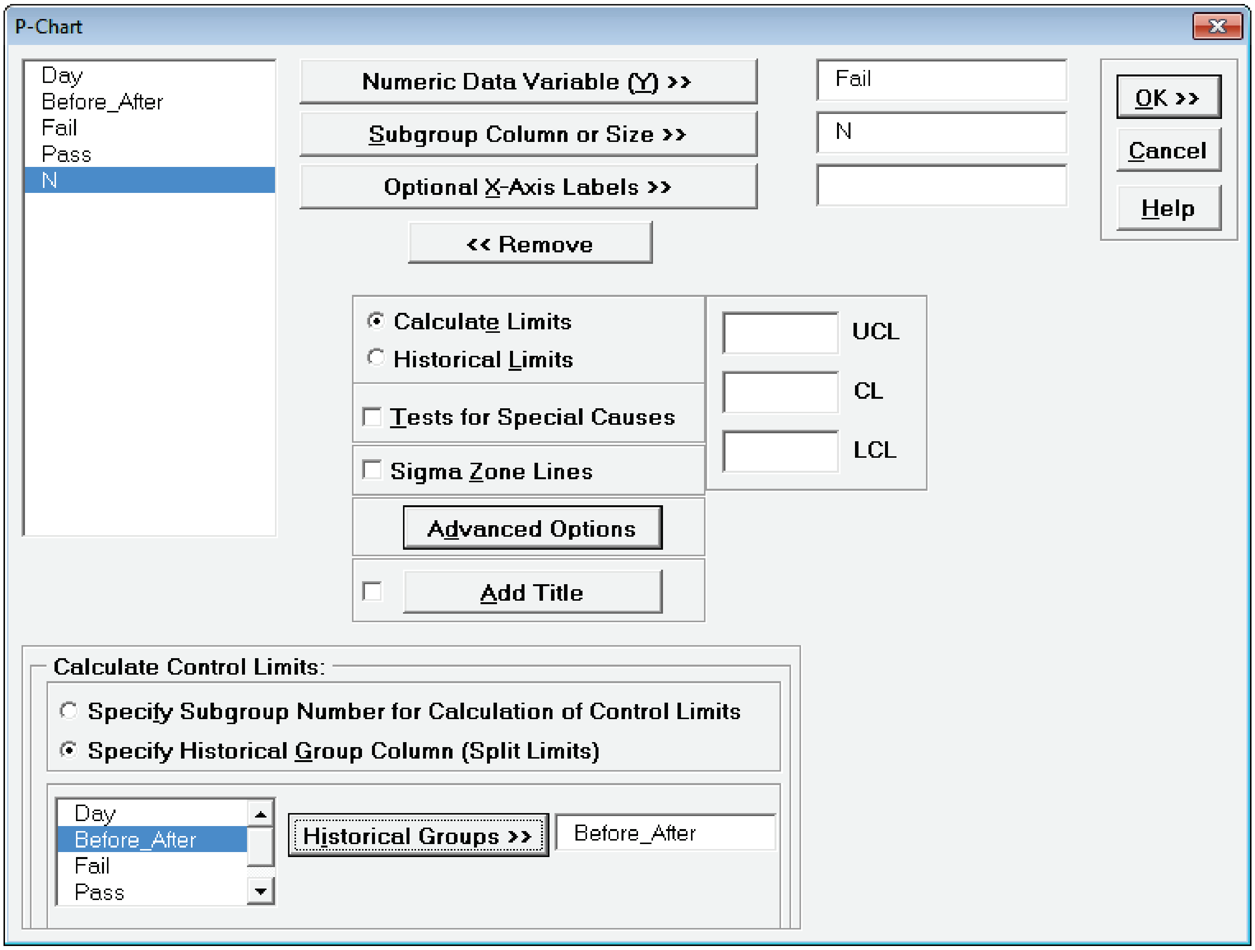
P-Charts: Advanced Limit Options – Historical Groups

1. Open **New York Daily Cycle Time Discrete – Before After Improvement.xlsx**. *Before_After* denotes before improvement and after improvement. We would like to use a P-Chart with historical groups to split the limits demonstrating the before versus after improvement.

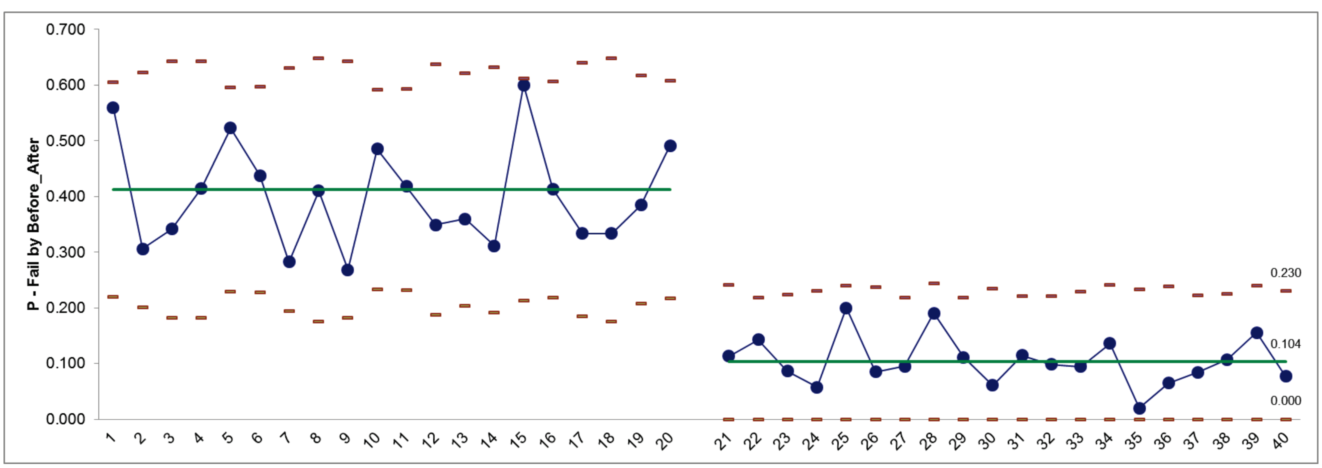
2. Click **SigmaXL > Control Charts > Attribute Charts > P**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Fail*, click **Numeric Data Variable (Y) >>**; select *N*, click **Subgroup Column or Size >>**.

4. Click **Advanced Options**. Select **Specify Historical Group Column**. Select *Before_After*, click **Historical Groups >>**.



5. Click **OK**. The resulting P-Chart with split limits based on historical groups is shown, demonstrating a clear process improvement:

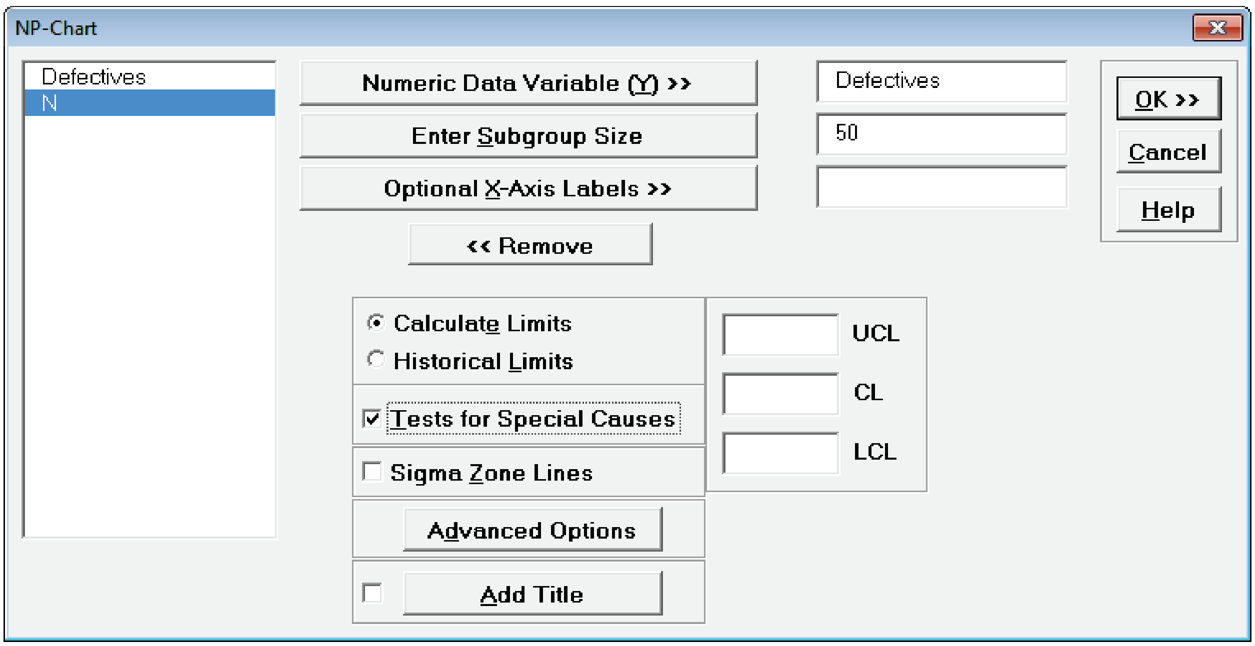


NP-Charts

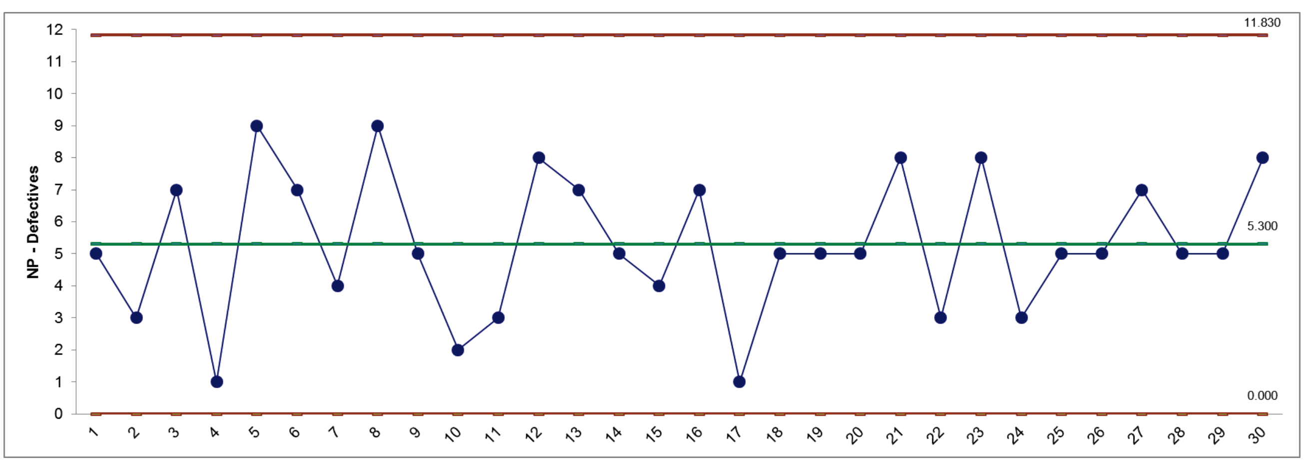
1. Open **Attribute Data NP Chart Defectives.xlsx**. The subgroup size is constant at 50. NP Charts (for Defectives) require a fixed subgroup size.

2. Select **SigmaXL > Control Charts > Attribute Charts > NP**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Defectives*, click **Numeric Data Variable (Y) >>**; enter 50 for **Enter Subgroup Size**. Check **Tests for Special Causes**.



4. Click **OK**. The resulting NP-Chart is shown:



This is a stable process, with no tests for special causes flagged.

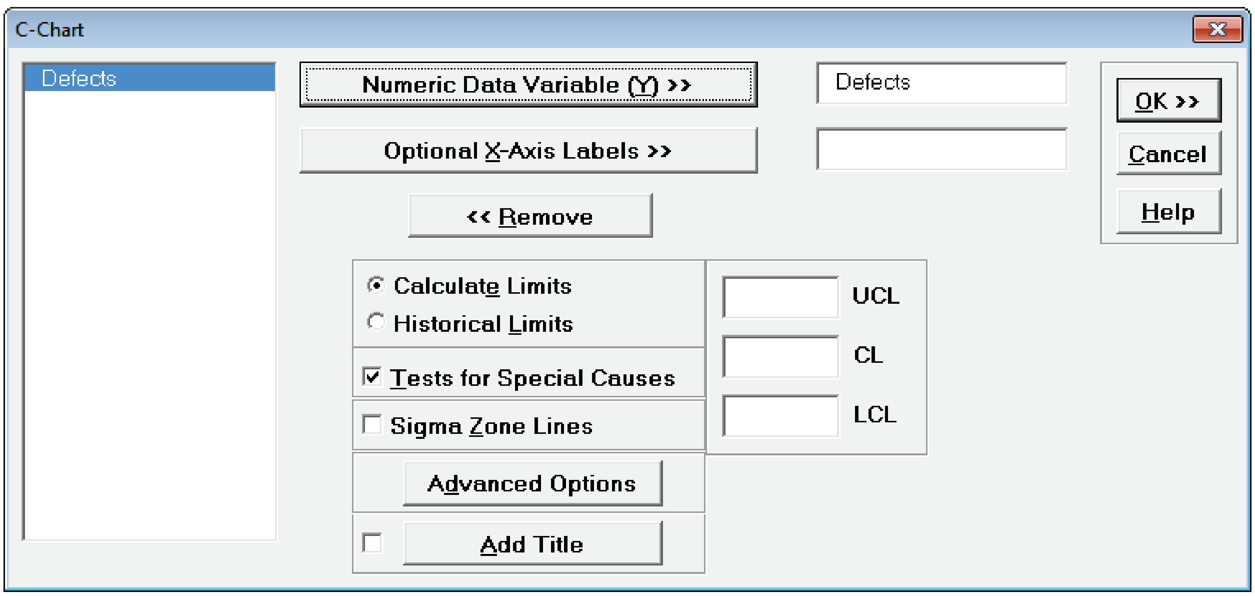
C Chart Template

Click **SigmaXL > Templates and Calculators** > **Control Chart Templates > Basic > C (Count).**

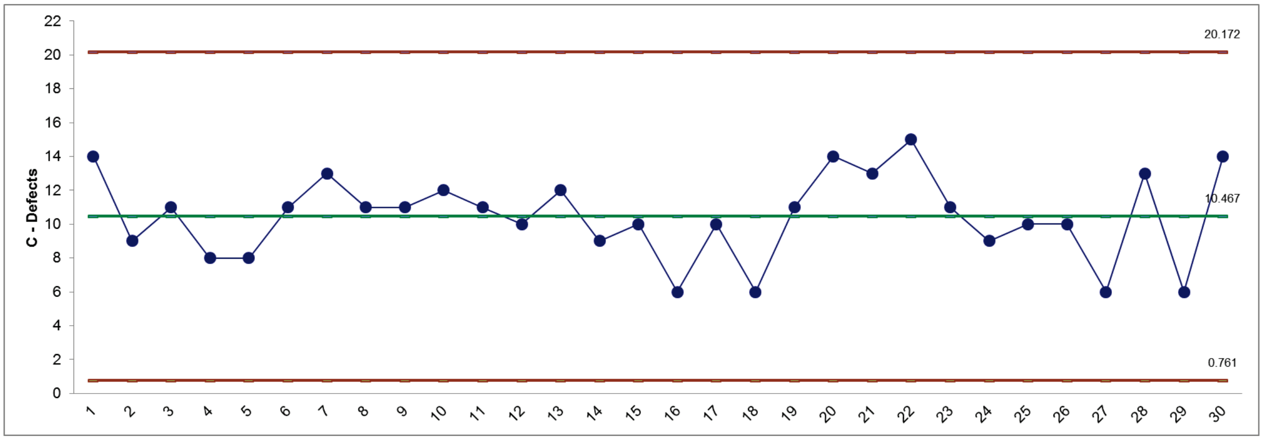
This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Basic > C (Count).**

C-Charts

1. Open **Attribute Data C Chart Defects.xlsx**. C Charts (for Defects) assume a fixed subgroup size.
2. Select **SigmaXL > Control Charts > Attribute Charts > C**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
3. Select *Defects*, click **Numeric Data Variable (Y) >>**. Check **Tests for Special Causes**.



4. Click **OK**. The resulting C-Chart is shown:



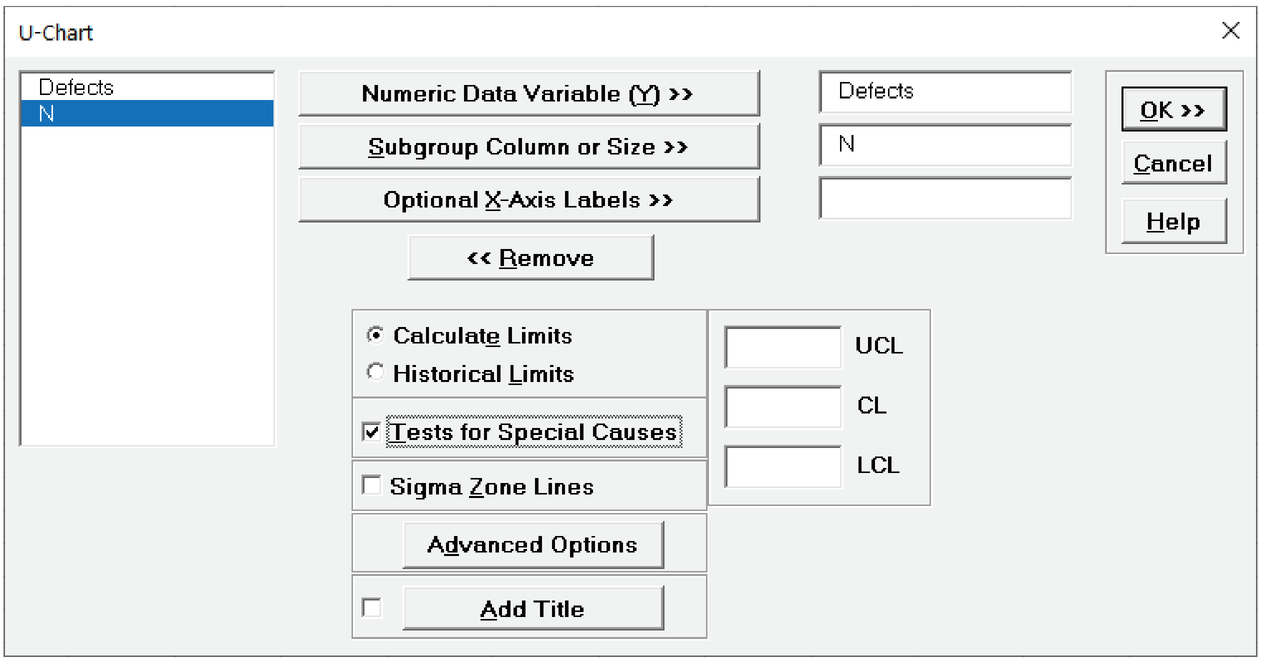
This is a stable process, with no tests for special causes flagged.

U-Charts

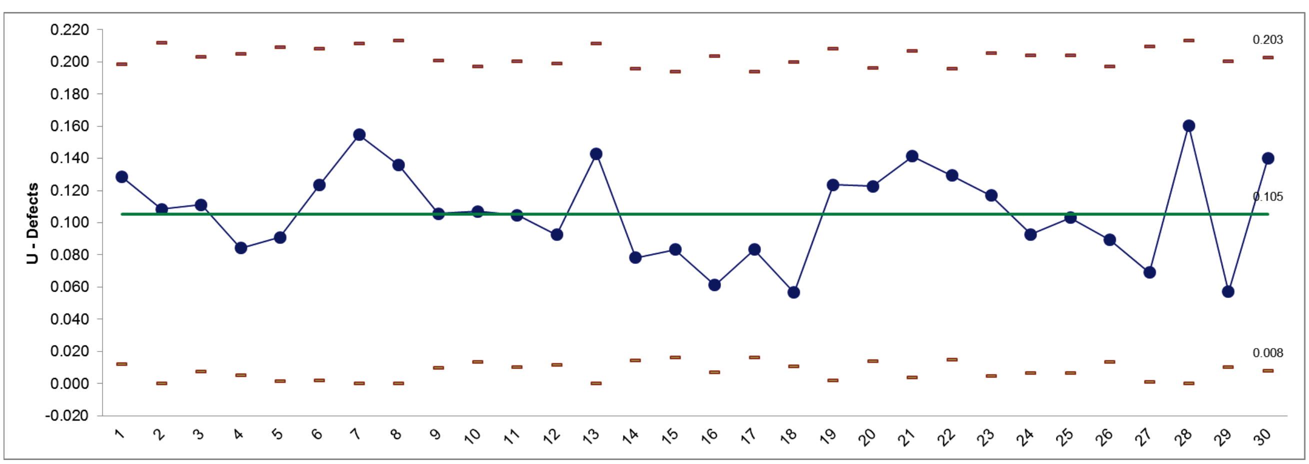
1. Open **Attribute Data U Chart Defects.xlsx**. U Charts (for Defects) can have fixed or varying subgroup sizes.

2. Select **SigmaXL > Control Charts > Attribute Charts > U**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Defects*, click **Numeric Data Variable (Y) >>**. Select *N*, click **Subgroup Column or Size >>**. Check **Tests for Special Causes**.



4. Click **OK**. The resulting U-Chart is shown:



This is a stable process, with no tests for special causes flagged.

Part D – P' & U' Charts (Laney)

P' and U' (Laney) Control Charts are attribute control charts that should be used when the subgroup/sample size is large and assumptions are not met. Typically, you will see that the control limits do not "look right," being very tight with many data points appearing to be out-of-control. This problem is also referred to as *overdispersion*. This occurs when the assumption of a Binomial distribution for defectives or Poisson distribution for defects is not valid. Individuals charts are often recommended in these cases, but Laney's P' and U' charts are a preferred alternative.

Tip: See Part E – Control Chart Selection Guide. The Control Chart Selection Guide makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

References:

[1] Crossley, Mark L.(2007), *The Desk Reference of Statistical Quality Methods*, Second Edition, Milwaukee, WI, ASQ Quality Press, pp. 345 – 356.

[2] Laney, David B., "P-Charts and U-Charts Work (But Only Sometimes)," Quality Digest, http://www.qualitydigest.com/sept07/departments/what_works.shtml.

[3] Laney, David B. (2002), "Improved Control Charts for Attribute Data," Quality Engineering 14:531–7.

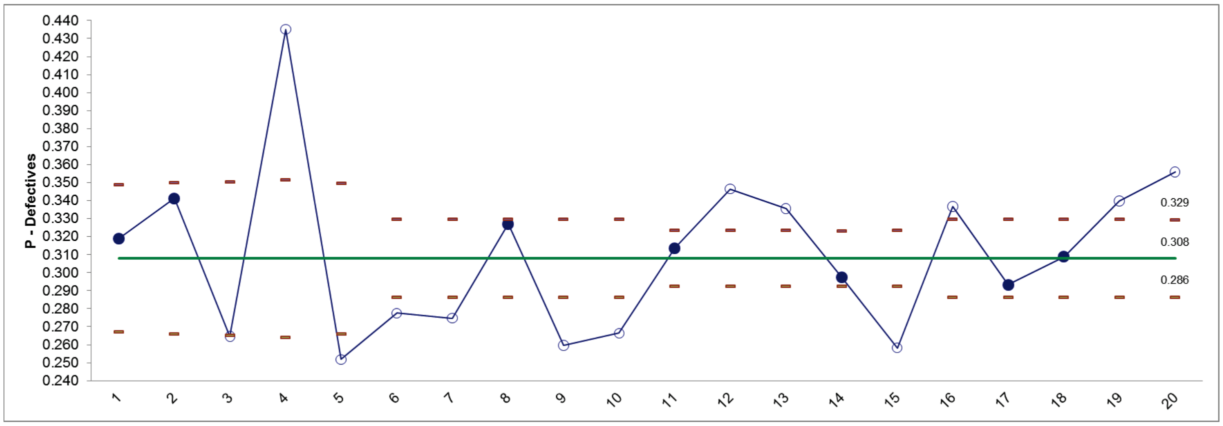
[4] M. A. Mohammed and D. Laney (2006), "Overdispersion in health care performance data: Laney's approach," Qual. Saf. Health Care 15; 383-384.

P'-Charts

1. Open **Laney – Quality Digest – Defectives.xlsx**. This data is used with permission from David Laney and was used in the example given in reference [2].

2. We will begin with the creation of a regular P-Chart for this data. Select **SigmaXL > Control Charts > Attribute Charts > P**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

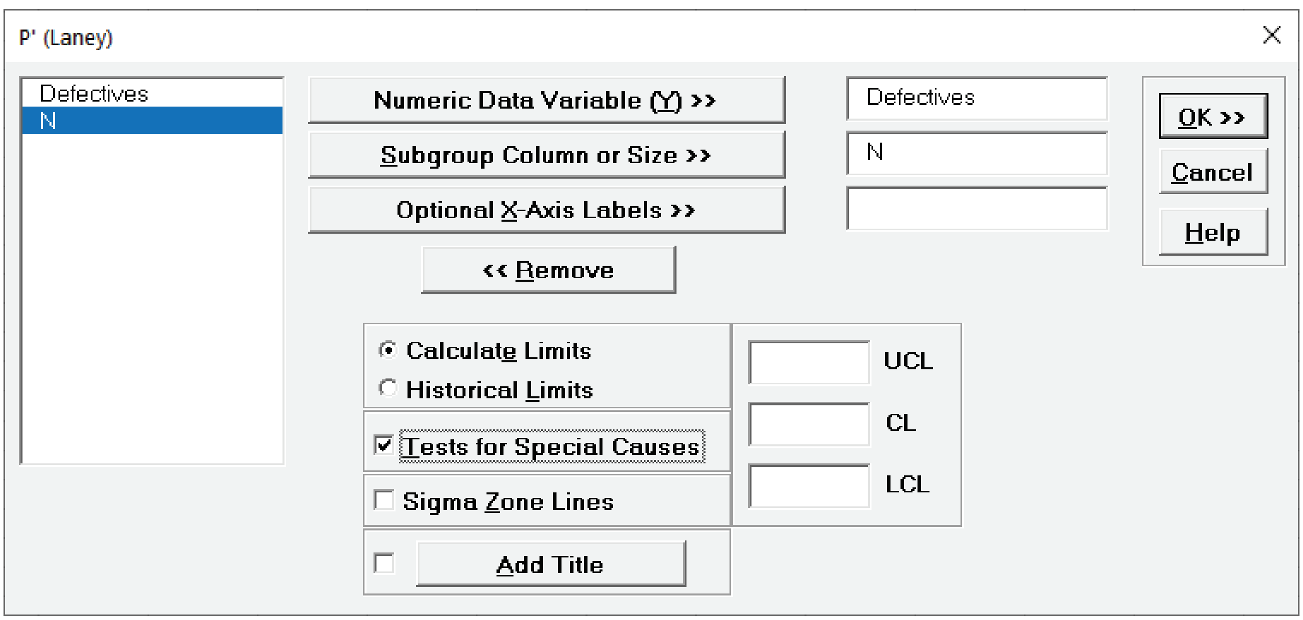
3. Select *Defectives* as the **Numeric Data Variable (Y)**, *N* as the **Subgroup Column or (Size)**. Click **OK**. The resulting P-Chart is shown:



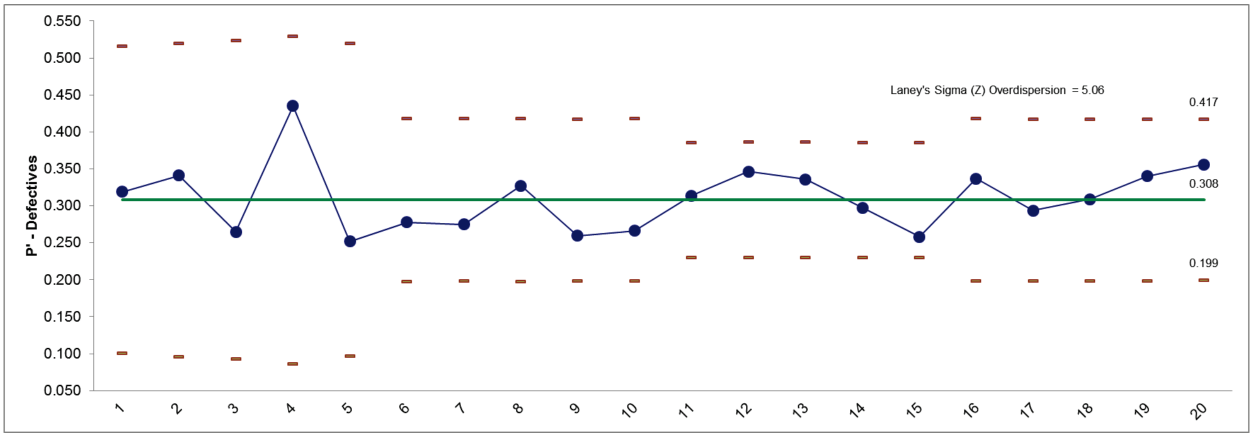
At first glance, this chart suggests that the process is out of control. The problem here is actually due to the Binomial assumption not being valid (also known as *overdispersion*). This problem becomes evident when the sample sizes are large.

4. Select Sheet **Defectives** (or press **F4**). **Click SigmaXL > Control Charts > Attribute Charts > P' (Laney)**. Click **Next**.

5. Select *Defectives*, click **Numeric Data Variable (Y) >>**; select *N*, click **Subgroup Column or Size >>**. (If we had a fixed subgroup size, the numerical value of the subgroup size could be entered instead of Column N.) Check **Tests for Special Causes**.



6. Click **OK**. The resulting P'-Chart is shown:



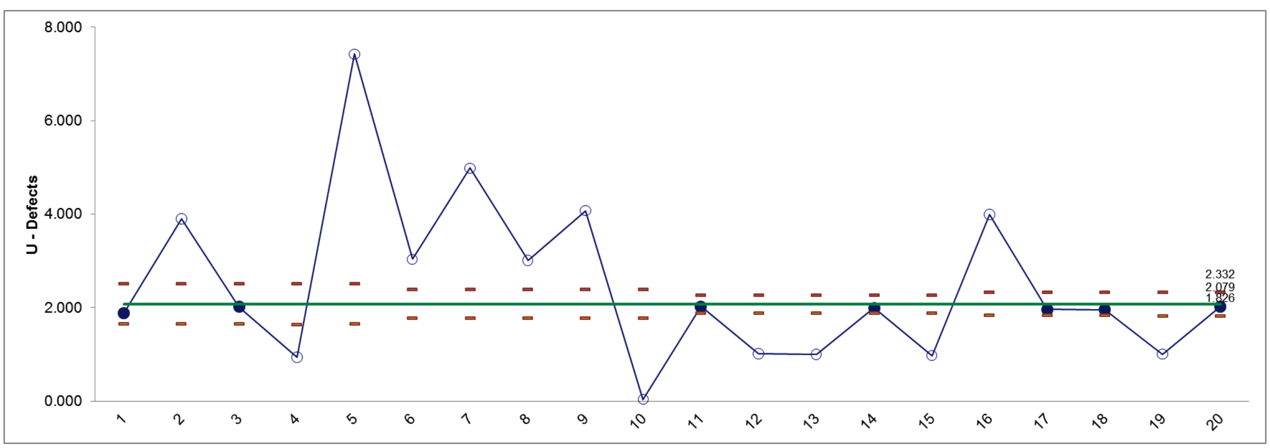
7. Now we see that the process is actually "in-control." Laney's Sigma (Z) is a measure of the *overdispersion*. See referenced articles for further details.

U'-Charts

1. Open **Attribute Data – U' Defects - Crossley.xlsx**. This data is used with permission from Mark Crossley and was used in the example given in reference [1]. Note that there are multiple defects per unit.

2. We will begin with the creation of a regular U-Chart for this data. Select **SigmaXL > Control Charts > Attribute Charts > U**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

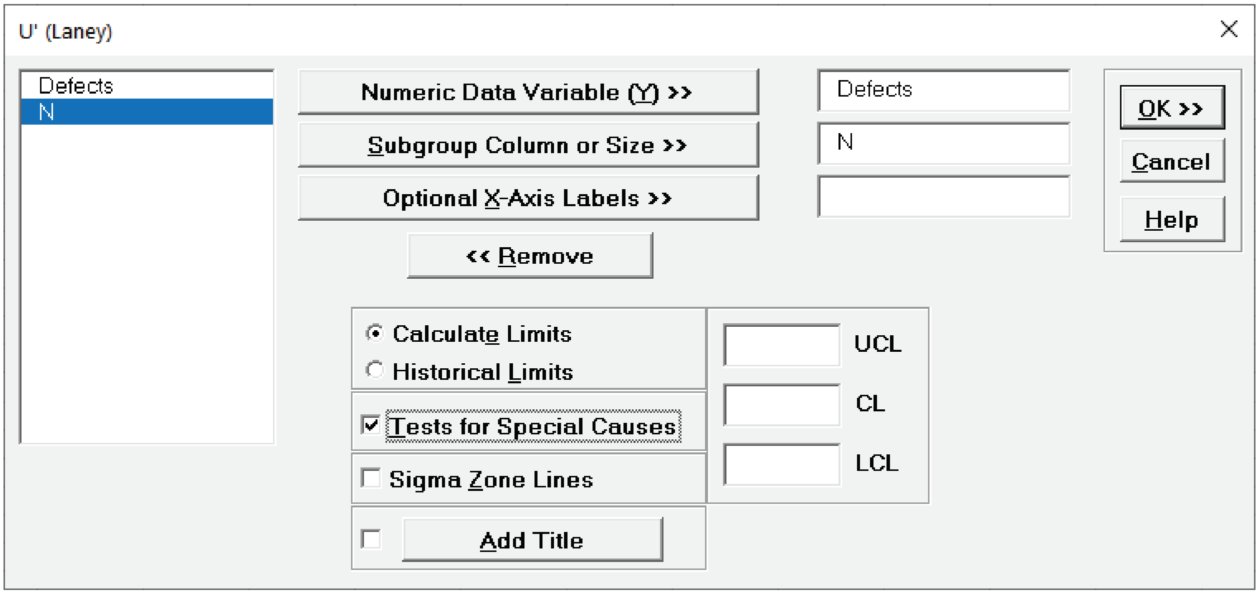
3. Select *Defects* as the **Numeric Data Variable (Y)**, *N* as the **Subgroup Column (Size)**. Click **OK**. The resulting U-Chart is shown:



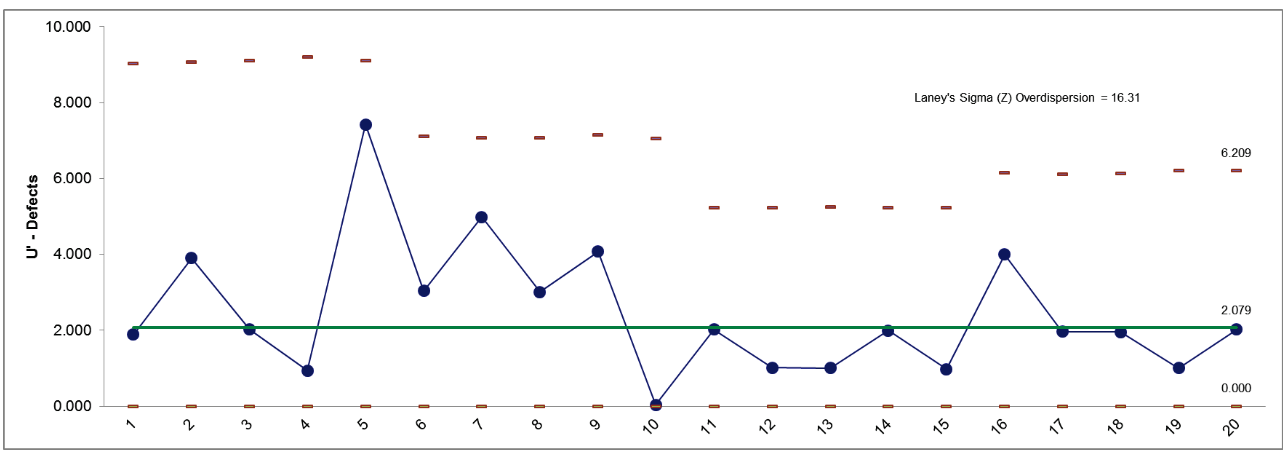
This chart suggests that the process is out of control, but the problem here is actually due to the Poisson assumption not being valid (also known as *overdispersion*). This problem becomes evident when the sample sizes are large.

4. Select Sheet **U' Chart Defects - Crossley** (or press **F4**). **Click SigmaXL > Control Charts > Attribute Charts > U' (Laney)**. Click **Next**.

5. Select *Defects*, click **Numeric Data Variable (Y) >>**; select *N*, click **Subgroup Column or Size >>**. (If we had a fixed subgroup size, the numerical value of the subgroup size could be entered instead of Column N.) Check **Tests for Special Causes**.



6. Click **OK**. The resulting U'-Chart is shown:



7. Now we see that the process is actually "in-control." Laney's Sigma (Z) is a measure of the *overdispersion*. See reference [1] for further details.

Part E – Control Chart Selection Guide

The Control Chart Selection Guide makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

Data Types and Definitions

Continuous/Variable: Data that is measured on a continuous scale where a mid-point (or other subdivision) has meaning. For example, when measuring cycle time, 2.5 days has meaning. Other examples include distance, weight, thickness, length and cost. Customer Satisfaction on a 1 to 5 scale can be considered as continuous in that a satisfaction score of 3.5 has meaning. Continuous data is always in numeric format.

Discrete/Attribute: Data that is categorical in nature. If we have defect types 1, 2, and 3, defect type 1.5 has no meaning. Other examples of discrete data would be customer complaints and reasons for product return. Discrete data can be text or integer numeric format.

Defective: An entire unit that is nonconforming to customer requirements. A unit may be defective because of one or more defects. For example, an application form is good only if all critical entry fields are correct. Any error in a critical field is a defect, resulting in a defective form. A single form can have more than one defect.

Defect: Any specific nonconformity to customer requirements. There can be more than one defect per unit or area of opportunity, such as the entry errors described above.

Subgroup/Sample: Data for a subgroup are usually collected within a short period of time to ensure homogeneous conditions within the subgroup (common cause variation), in order to detect differences between subgroups (special cause variation).

Subgroup/Sample Size: The number of observations within your sample, not the number of samples. Subgroup sizes of 3 to 5 are common for continuous measures in parts manufacturing, while individual measurements are common in chemical processes (temperature, pH) and transactional areas (financial). Subgroup size for discrete data should be a minimum of 50.

Subgroup/Sample Size is constant: The number of observations within your sample remains fixed over time.

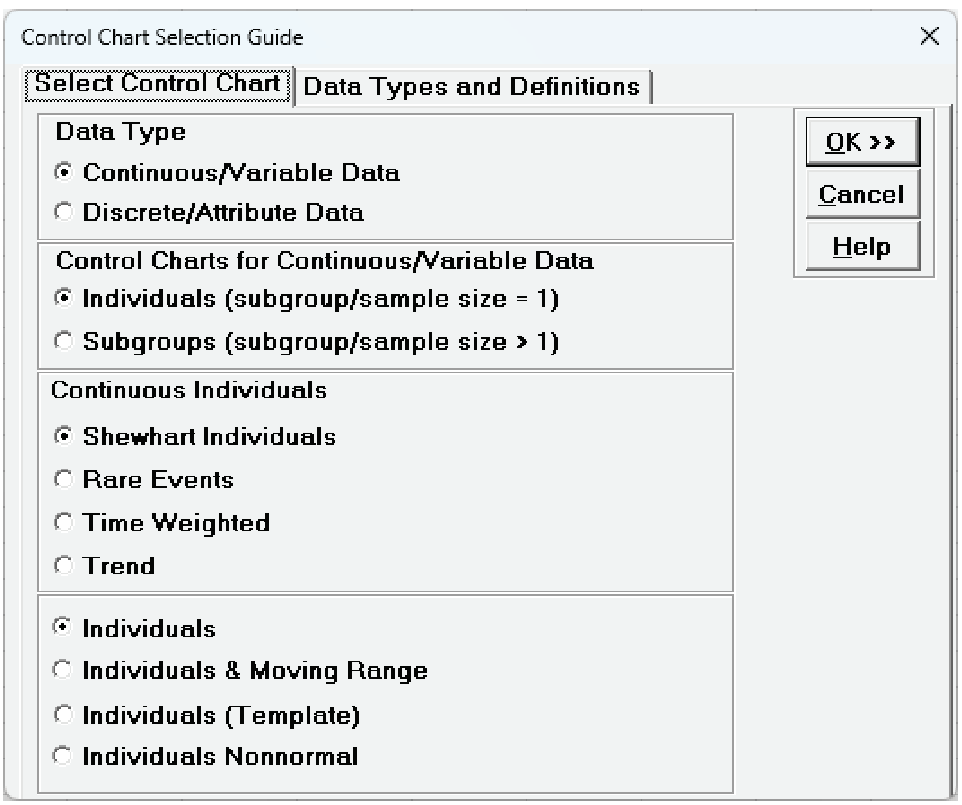
Subgroup/Sample Size varies: The number of observations within your sample varies over time.

Subgroup/Sample Size is very large and assumptions not met: This applies to discrete data when the subgroup sizes are approximately 5,000 or higher and the control limits do not "look right," being very tight with many data points appearing to be out-of-control. This problem is also referred to as "overdispersion." This occurs when the assumption of a Binomial distribution for defectives or Poisson distribution for defects is not valid. (Note: If the problem of overdispersion is apparent with your continuous data, use **SigmaXL > Control Charts > I-MR-R** or **I-MR-S**).

Control Chart Selection Guide – Individuals Chart

1. Open **Customer Data.xlsx**, click on **Sheet 1**. Click **SigmaXL > Control Charts > Control Chart Selection Guide**.

2. We would like to create a control chart of the Overall Satisfaction data. Since this can be considered as continuous data, the data type is **Continuous/Variable Data**. The subgroup/sample size is 1 (i.e., there is no subgrouping), so select **Individuals (subgroup/sample size = 1)**. At this point, we can choose **Individuals** or **Individuals and Moving Range**. We will keep the simpler **Individuals** selection as shown. (Note that the above data types and definitions can be viewed by clicking the **Data Types and Definitions** tab):



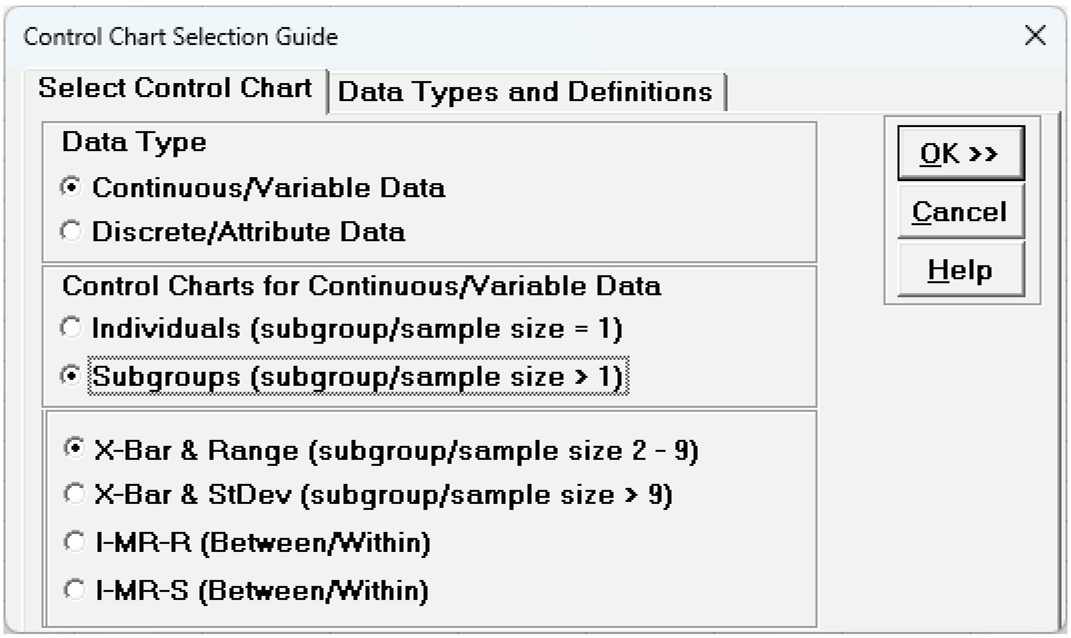
3. Click **OK**. This starts up the Individuals Chart dialog (see Part A – Individuals Charts for continuation).

Control Chart Selection Guide – X-Bar & R Chart

1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the ball 3 times. The target distance is 100 inches. Select **B2:F22**; here we will only use the first 20 subgroups to determine the control limits.

2. Click **SigmaXL > Control Charts > Control Chart Selection Guide**.

3. Since catapult shot distance measurement is continuous, we keep the default selection **Continuous/Variable Data**. The catapult shot data are in subgroups, so select **Subgroups (subgroup/sample size > 1)**. The subgroup/sample size is small (3), so we will use the **X-Bar & Range Chart** as shown:



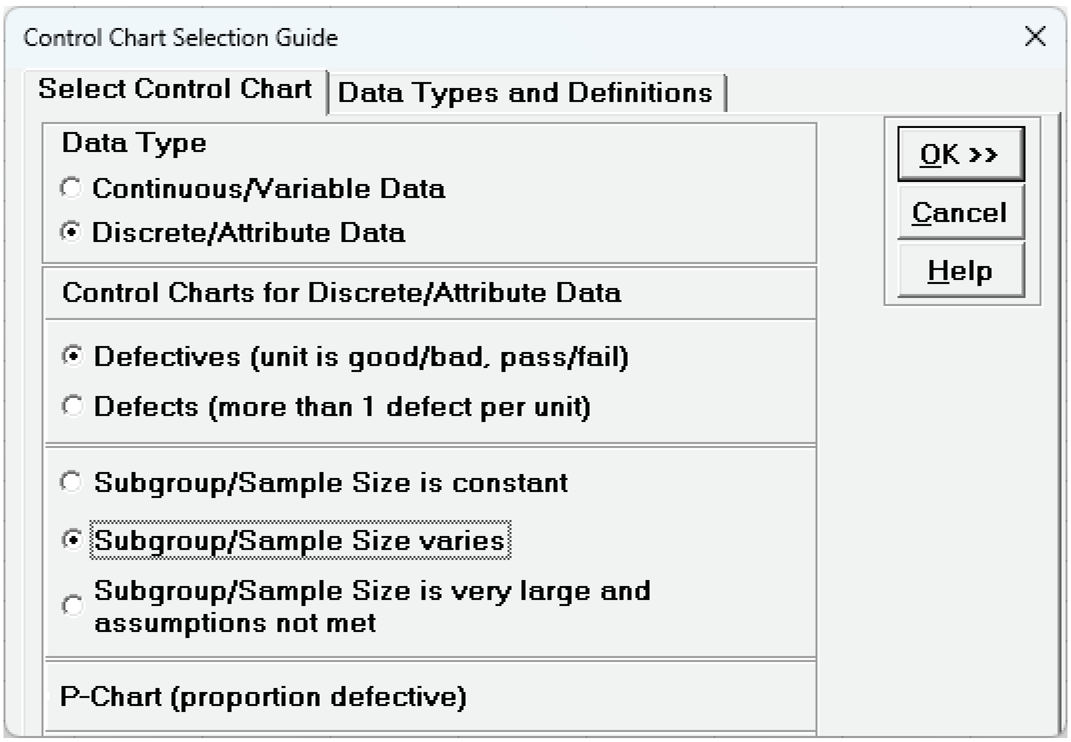
4. Click **OK**. This starts up the **X-Bar & Range** dialog (see Part B – X-Bar & Range Charts for continuation).

Control Chart Selection Guide – P-Chart

1. Open **New York Daily Cycle Time – Discrete.xlsx**. This is data from the Sigma Savings and Loans Company, New York location. Each day, the cycle time (in days) for completed loans and leases was recorded. *N* indicates the number of loans counted. A *Fail* was recorded if the cycle time exceeded the critical customer requirement of 8 days.

2. Click **SigmaXL > Control Charts > Control Chart Selection Guide**.

3. Since this data is discrete, select **Discrete/Attribute Data**. We are looking at Defectives data since each loan cycle time is a pass or fail, so select **Defectives (unit is good/bad, pass/fail)**. The subgroup/sample size varies day to day so **Subgroup/Sample Size varies** is selected as shown. The recommended chart is the **P-Chart (proportion defective)**:



4. Click **OK**. This starts up the P-Chart dialog (see Part C – P-Charts for continuation).

Part F – I-MR-R/S Charts

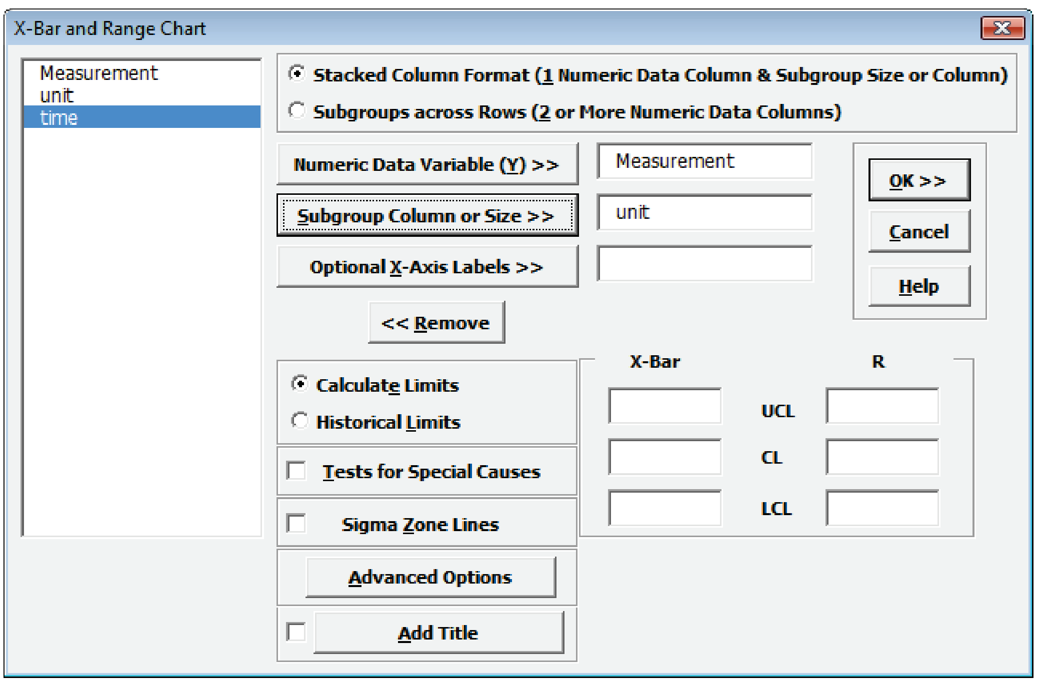
I-MR-R Chart

If the within-subgroup variability is much smaller than between subgroup, the classical X-bar & R (or S) chart will not work, producing numerous (false) alarms. The correct chart to use, in this case, is the I-MR-R (or S) chart. The subgroup averages are treated as individual values (I-MR) and the within subgroup ranges are plotted on the Range chart.

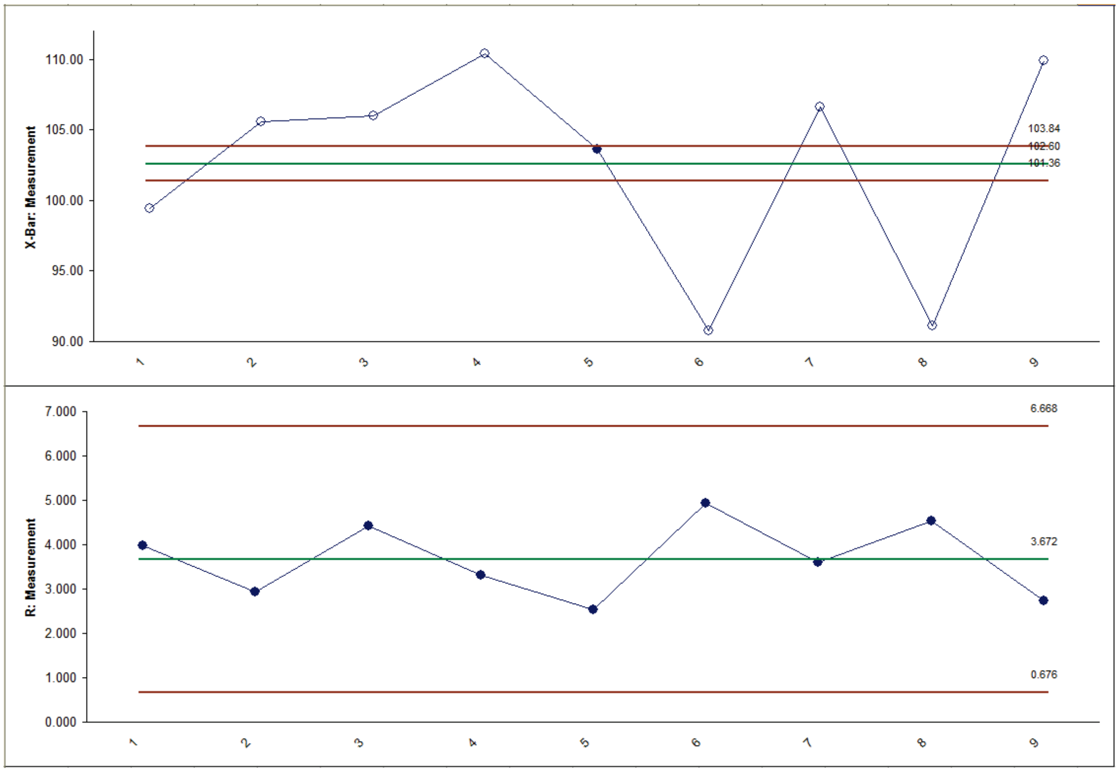
1. Open **Multi-Vari Data.xlsx**. Select Sheet **Between.** We saw this data previously using Multi-Vari charts. First, we will incorrectly use the X-bar & R chart, and then apply the correct I-MR-R chart.

2. Click **SigmaXL > Control Charts > X-bar & R**. Check **Use Entire Data Table**.

3. Click **Next**. Select **Stacked Column Format**. Select *Measurement*, click **Numeric Data Variable (Y) >>**; select *unit*, click **Subgroup Column or Size >>**.



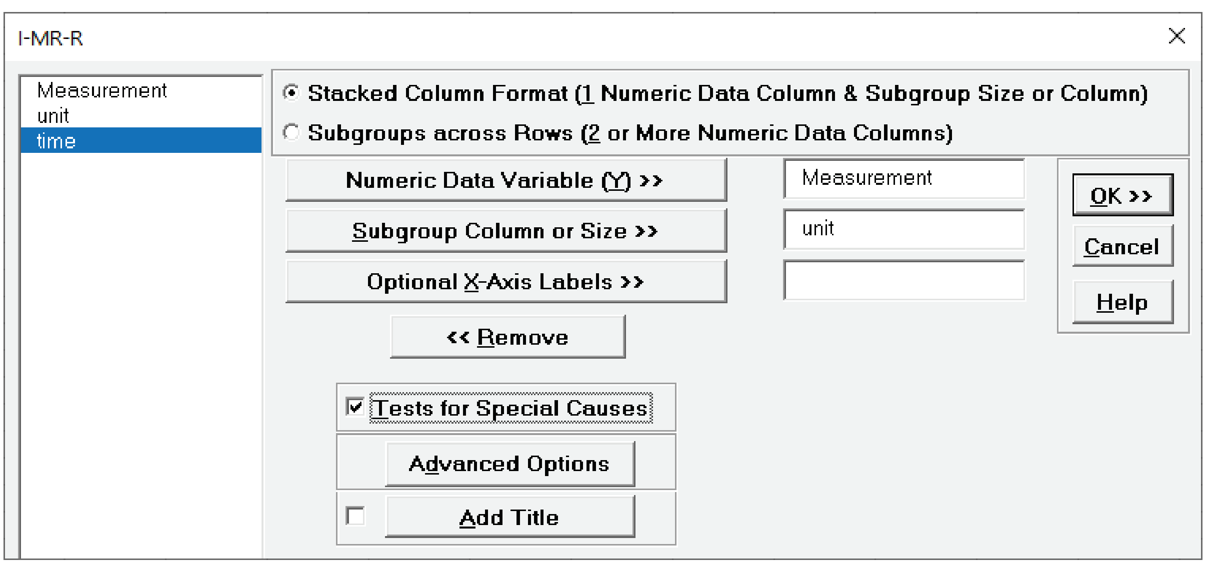
4. Click **OK**. Resulting X-bar & R chart:



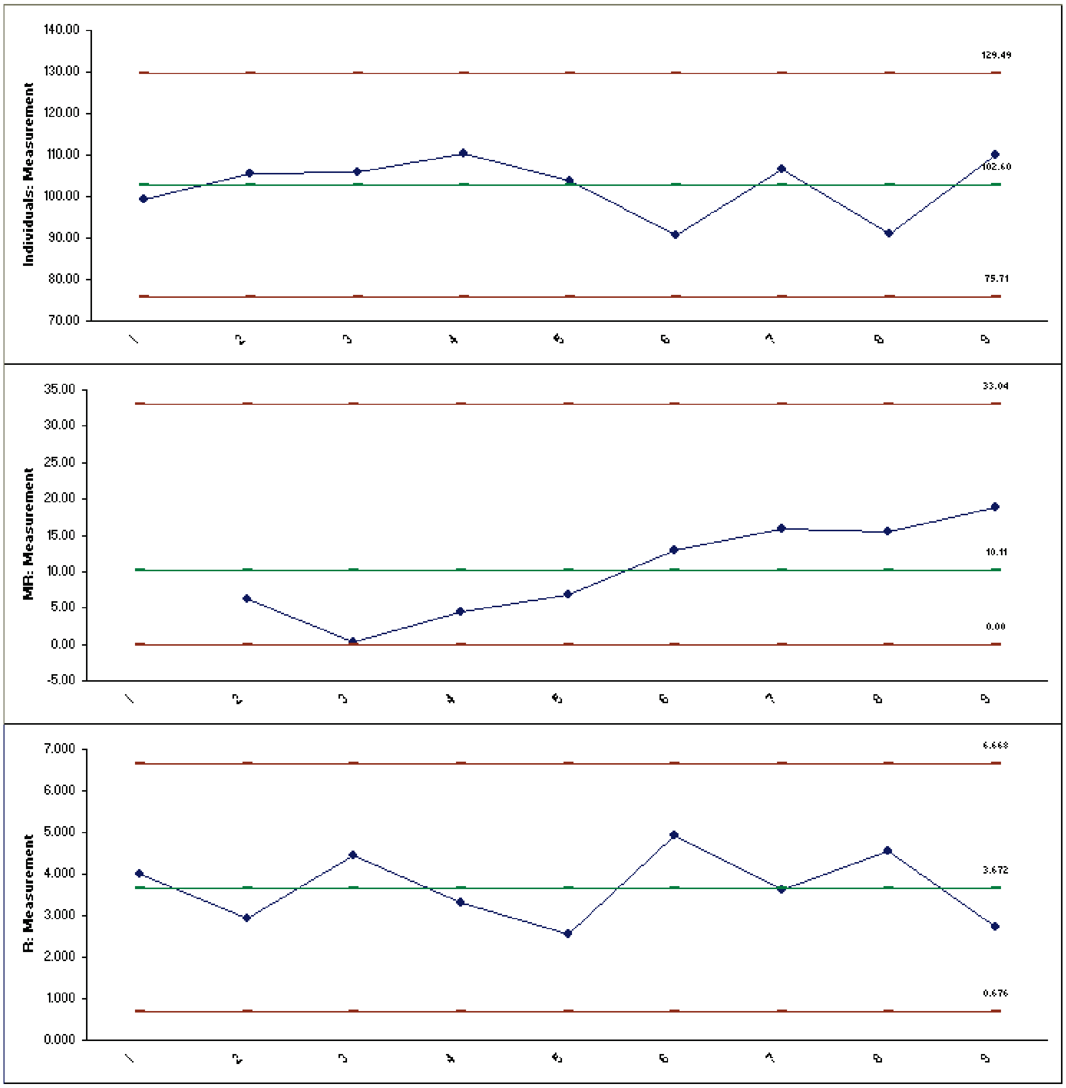
Almost all of the data points in the X-bar chart are out-of-control! This is due to the small within-subgroup variability (the control limits are calculated from the within-subgroup variability).

5. Click Sheet **Between**. Select **SigmaXL > Control Charts > I-MR-R**.

6. Click **Next**. Select **Stacked Column Format**. Select *Measurement*, click **Numeric Data Variable (Y) >>**; select *unit*, click **Subgroup Column or Size >>**. Check **Tests for Special Causes**.



7. Click **OK**. Resulting I-MR-R chart:



This chart is much cleaner, showing a stable Individuals and Range chart. The MR chart may be trending up, but we would want to collect more data before making this conclusion. Typically, you want at least 20 (30 preferred) subgroups before calculating final control limits.

Part G – Control Chart Templates: Rare Events

Rare Events T Chart - Introduction

The Rare Events T Chart (or time-between chart) is an alternative to a standard attribute chart when the adverse event of interest is relatively rare and a measurement of time between each occurrence can be obtained. If the rate of occurrence follows a Poisson distribution (the usual assumption for a C or U chart), then the times between occurrences will have an exponential distribution. The exponential distribution can be transformed to a symmetric Weibull distribution by raising the time measure to the (1/3.6) power. A Shewhart Individuals chart is calculated on the transformed data and then an inverse transformation is applied to the control limits in order to get back to the original time between units. This is an approximate model to the exponential and will result in asymmetric control limits. Note that an "out-of-control" signal above the UCL is desirable, indicating a significant increase in time between adverse rare events. See Provost and Murray, 2011 and Nelson, 1994. The Provost and Murray book is popular in health care, so SigmaXL uses the control limit calculations as given in the book:

$\mathrm{t} = \text{time between incidents}$

$\mathrm{z} = \text{transformed time}\ \left[\mathrm{z} = \mathrm{t}^{(1/3.6)}\right]$

Construct a Shewhart individuals control chart of z values:

$\overline{MR}_z = \text{average moving range of z's}$

Remove moving range outliers (i.e., exceed UCL for moving range) for added robustness:

$\overline{MR}'_z = \text{average moving range of z's with any MRs} > (3.27 * \overline{MR}_z)\ \text{removed}$

$CL_z = \bar{Z}$ (average of transformed time)

$$UCL_z \quad = \bar{Z} + 3 * \left(\frac{\overline{MR}'_z}{1.128}\right)$$

$$LCL_z \quad = \bar{Z} - 3 * \left(\frac{\overline{MR}'_z}{1.128}\right)$$

Transform the center line and the limits back to time scale by raising them to the 3.6 power:

$$CL_t = {CL_z}^{3.6}, UCL_t = {UCL_z}^{3.6}, LCL_t = {LCL_z}^{3.6}$$

Tip: Alternatively, a Rare Events T Chart may be created using an Exponential or Weibull Control Chart on the Days Between data. Click **SigmaXL > Control Charts > Nonnormal > Individuals Nonnormal**. Select **Specify Distribution**, **Exponential (1 Parameter)** or **Weibull (2 Parameter).** To select which distribution is the best fit, use distribution fitting: **SigmaXL > Control Charts > Nonnormal > Distribution Fitting**. Select **All Transformations & Distributions** and compare.

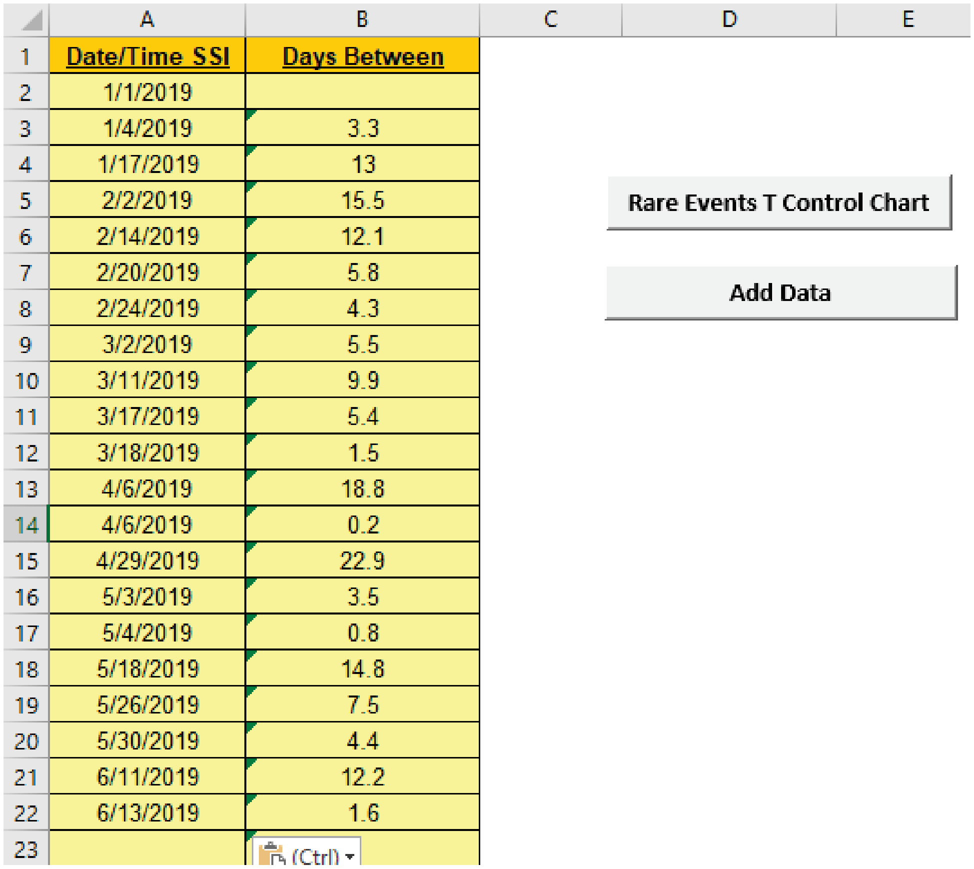
Rare Events T Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Rare Events > Rare Events T**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Rare Events > Rare Events T.**

2. Open **Days Between Surgical Site Infections.xlsx** (**Sheet 1** tab). Days between surgical site infections are an important health care metric. Since they are rare events, days between (or time between) events are tracked rather than events per day. The event date/times 1/1/2019 to 6/13/2019 are "Before Improvement" and the days between will be used to calculate the control limits. Date/times 6/14/2019 to 2/26/2020 are "After Improvement" and the days between will be added to the control chart. The goal is to decrease the occurrence of Surgical Site Infections which will increase the days between, with an "out-of-control" signal above the UCL being desirable and confirmation of the improvement effort.

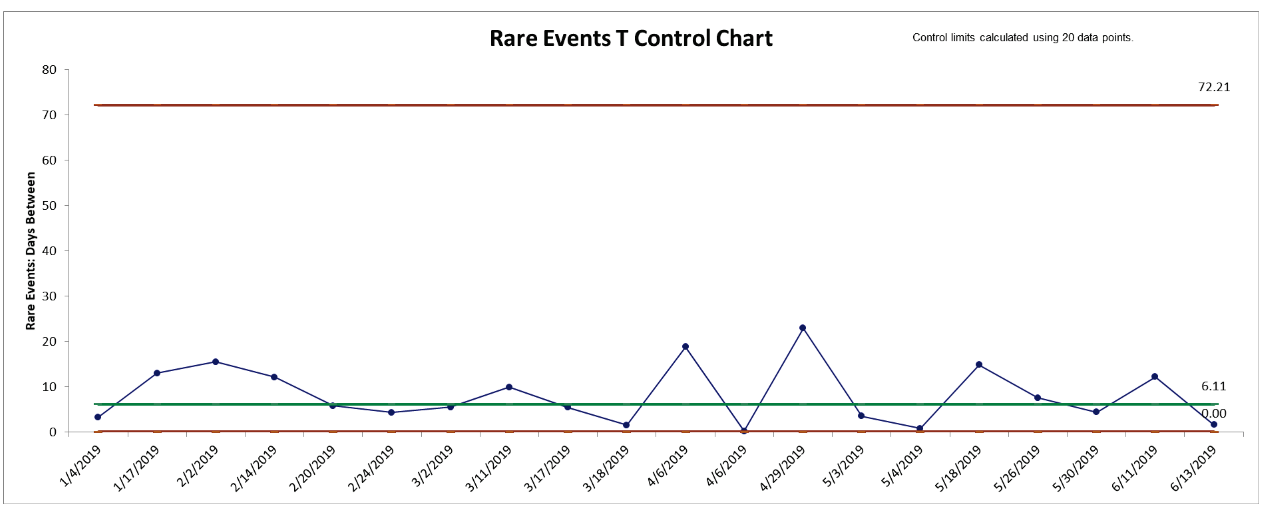
	A	B
1	Date/Time SSI	Before/After Improvement
2	1/1/2019	Before
3	1/4/2019	Before
4	1/17/2019	Before
5	2/2/2019	Before
6	2/14/2019	Before
7	2/20/2019	Before
8	2/24/2019	Before
9	3/2/2019	Before
10	3/11/2019	Before
11	3/17/2019	Before
12	3/18/2019	Before
13	4/6/2019	Before
14	4/6/2019	Before
15	4/29/2019	Before
16	5/3/2019	Before
17	5/4/2019	Before
18	5/18/2019	Before
19	5/26/2019	Before
20	5/30/2019	Before
21	6/11/2019	Before
22	6/13/2019	Before
23	6/14/2019	After
24	7/26/2019	After
25	8/3/2019	After
26	8/6/2019	After
27	8/25/2019	After
28	9/8/2019	After
29	11/13/2019	After
30	2/15/2020	After
31	2/16/2020	After
32	2/26/2020	After

3. Copy the “Before Improvement” date/times in cells **A1:A22** and Paste Values to the template at **A1**. **Days Between** are automatically calculated using the Excel formula =A3-A2, =A4-A3, etc. Cell **B2** is not used.



	A	B
1	Date/Time SSI	Days Between
2	1/1/2019	
3	1/4/2019	3.3
4	1/17/2019	13
5	2/2/2019	15.5
6	2/14/2019	12.1
7	2/20/2019	5.8
8	2/24/2019	4.3
9	3/2/2019	5.5
10	3/11/2019	9.9
11	3/17/2019	5.4
12	3/18/2019	1.5
13	4/6/2019	18.8
14	4/6/2019	0.2
15	4/29/2019	22.9
16	5/3/2019	3.5
17	5/4/2019	0.8
18	5/18/2019	14.8
19	5/26/2019	7.5
20	5/30/2019	4.4
21	6/11/2019	12.2
22	6/13/2019	1.6
23		

4. Click the **Rare Events T Control Chart** button to create the Rare Events T Control Chart:



5. This confirms that the process is “in-control”. As discussed above the control limits are asymmetrical.

6. Although the process is "in-control", as a critical health care metric, efforts were made to improve the process and this data will be now added to the chart. Switch back to **Days Between Surgical Site Infections.xlsx** (**Sheet 1** tab). Select and copy the "After Improvement" date/times in cells **A23: A32** as shown.

	A	B
1	Date/Time SSI	Before/After Improvement
2	1/1/2019	Before
3	1/4/2019	Before
4	1/17/2019	Before
5	2/2/2019	Before
6	2/14/2019	Before
7	2/20/2019	Before
8	2/24/2019	Before
9	3/2/2019	Before
10	3/11/2019	Before
11	3/17/2019	Before
12	3/18/2019	Before
13	4/6/2019	Before
14	4/6/2019	Before
15	4/29/2019	Before
16	5/3/2019	Before
17	5/4/2019	Before
18	5/18/2019	Before
19	5/26/2019	Before
20	5/30/2019	Before
21	6/11/2019	Before
22	6/13/2019	Before
23	6/14/2019	After
24	7/26/2019	After
25	8/3/2019	After
26	8/6/2019	After
27	8/25/2019	After
28	9/8/2019	After
29	11/13/2019	After
30	2/15/2020	After
31	2/16/2020	After
32	2/26/2020	After

7. Switch to the Rare Events T Control Chart template and Paste Values to cell **A23** as shown.

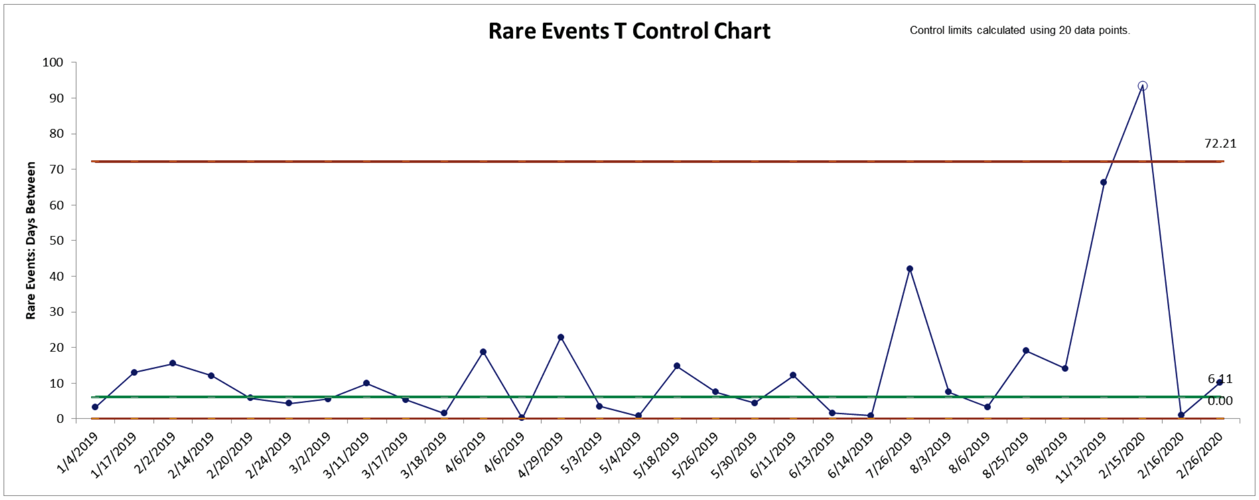
	A	B
1	Date/Time SSI	Days Between
2	1/1/2019	
3	1/4/2019	3.3
4	1/17/2019	13
5	2/2/2019	15.5
6	2/14/2019	12.1
7	2/20/2019	5.8
8	2/24/2019	4.3
9	3/2/2019	5.5
10	3/11/2019	9.9
11	3/17/2019	5.4
12	3/18/2019	1.5
13	4/6/2019	18.8
14	4/6/2019	0.2
15	4/29/2019	22.9
16	5/3/2019	3.5
17	5/4/2019	0.8
18	5/18/2019	14.8
19	5/26/2019	7.5
20	5/30/2019	4.4
21	6/11/2019	12.2
22	6/13/2019	1.6
23	6/14/2019	0.9
24	7/26/2019	42.1
25	8/3/2019	7.5
26	8/6/2019	3.3
27	8/25/2019	19.1
28	9/8/2019	14.1
29	11/13/2019	66.3
30	2/15/2020	93.6
31	2/16/2020	1
32	2/26/2020	10.2
33		

Rare Events T Control Chart

Add Data

Notes:
1. This Rare Events T Con
2. You can replace the Da
3. Enter date or date/time
4. Alternatively, you may
5. Click the **Rare Events**
6. After the control chart h
7. An "out-of-control" sign

8. Click the **Add Data** button to add the new **Days Between** data to the Rare Events T Control Chart:



9. This confirms that the process is now “out-of-control”, so the improvement efforts to decrease the occurrence of Surgical Site Infections and increase the Days Between have been successful. Note that the control limits should be recalculated with just the “After Improvement” data when 20 data points are available.

Template Notes and References:

1. This Rare Events T Control Chart template should be used with days or time between (typically adverse) rare events.
2. You can replace the **Date/Time** and **Days Between** column headings with any headings that you wish.
3. Enter date or date/time in the **Date/Time** column. Days between are automatically calculated and entered into **Days Between** column. Cell **B2** is not used in this case.
4. Alternatively, you may manually enter data in **Days Between**. Note, this will overwrite the cell formula.
5. Click the **Rare Events T Control Chart** button to create a control chart. This will overwrite any existing control chart.
6. After the control chart has been created and additional new **Date/Time** or **Days Between** data entered, click the **Add Data** button to add the data to the existing chart.
7. An "out-of-control" signal above the UCL is desirable, indicating a significant increase in time between adverse rare events.
8. Be careful to not have any zeros in the data.
9. This chart uses the control limit formulas given in Provost and Murray, 2011.
10. The data are transformed with Y^(1/3.6), which transforms an exponential distribution to a symmetric Weibull. See Nelson, 1994.
11. A Shewhart Individuals Chart is created on the transformed data and the final control limits are then calculated as UCL^3.6, CL^3.6, LCL^3.6.

12. Alternatively, a Rare Events T Chart may be created using a Weibull or Exponential Control Chart. Click **SigmaXL > Control Charts > Nonnormal > Individuals Nonnormal**. Select **Specify Distribution**, **Exponential (1 Parameter)** or **Weibull (2 Parameter).**
13. References:
Nelson, L.S. (1994), “A Control Chart for Parts-Per-Million Nonconforming Items”, *Journal of Quality Technology*, 26:3, pp. 239-240.
Provost L, Murray S. (2011), *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco: Jossey-Bass, pp. 230-231.

Rare Events G Chart - Introduction

The G chart (or Geometric chart) is an alternative to a standard attribute chart when the adverse event of interest is rare and discrete opportunities between events are counted (e.g., number of units or days between).

The calculation of control limits is an approximation based on the geometric distribution. An "out-of-control" signal above the UCL is desirable, indicating a significant increase in units/opportunities or days between adverse rare events. See Provost and Murray, 2011. The Provost and Murray book is popular in health care, so SigmaXL uses the control limit calculations as given in the book:

g = number of opportunities or units between incidents

$\bar{g}$ = average of g's

$CL = 0.693 * \bar{g}$ (CL is the theoretical median for a geometric distribution)

$$UCL \quad = \bar{g} + 3 * \sqrt{\bar{g}(1 + \bar{g})}$$

$$LCL \quad = 0$$

For more accurate probability-based limits, see the Rare Events Prob G Control Chart.

Rare Events G Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Rare Events > Rare Events G**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Rare Events > Rare Events G.**

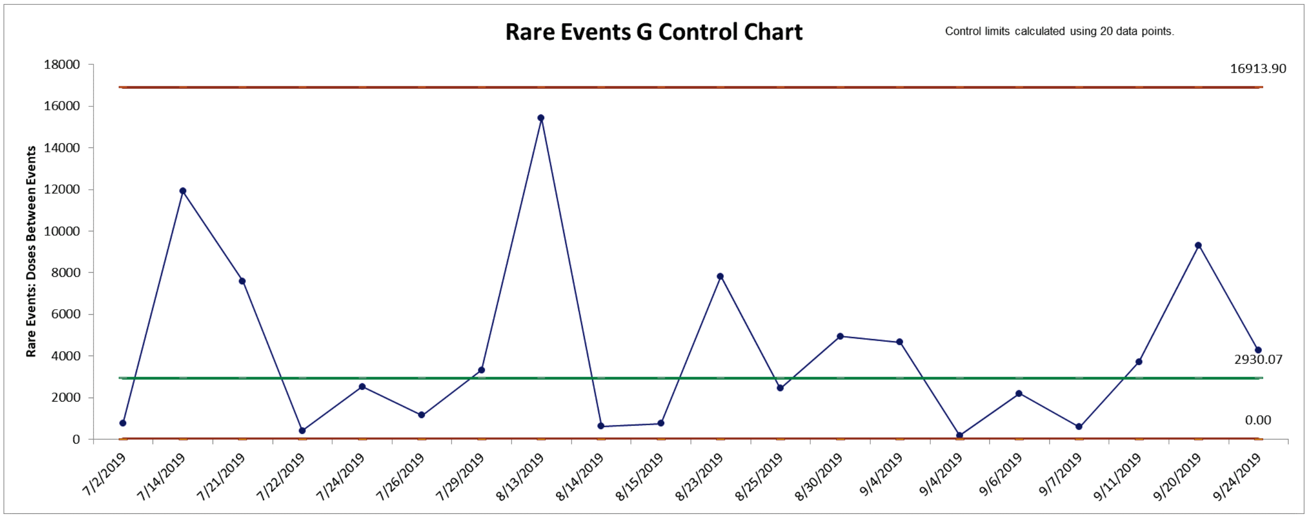
2. Open **Doses Dispensed Between Adverse Drug Events.xlsx** (**Sheet 1** tab). Adverse Drug Events (ADE) are rare, so Doses Dispensed Between ADEs is used as a health care metric. The data is discrete and distributed geometrically, so the G chart is appropriate to use here. The event dates 7/2/2019 to 9/24/2019 are "Before Improvement" and the Doses Between ADEs will be used to calculate the control limits. Dates 9/28/2019 to 1/17/2020 are "After Improvement" and the Doses Between ADEs will be added to the control chart. The goal is to decrease the occurrence of ADEs which will increase the doses between, with an "out-of-control" signal above the UCL being desirable and confirmation of the improvement effort.

	A	B	C
1	Date ADE	Doses Between Events	Before/After Improvement
2	7/2/2019	758	Before
3	7/14/2019	11911	Before
4	7/21/2019	7593	Before
5	7/22/2019	398	Before
6	7/24/2019	2516	Before
7	7/26/2019	1147	Before
8	7/29/2019	3322	Before
9	8/13/2019	15414	Before
10	8/14/2019	626	Before
11	8/15/2019	752	Before
12	8/23/2019	7819	Before
13	8/25/2019	2445	Before
14	8/30/2019	4937	Before
15	9/4/2019	4663	Before
16	9/4/2019	175	Before
17	9/6/2019	2187	Before
18	9/7/2019	592	Before
19	9/11/2019	3721	Before
20	9/20/2019	9317	Before
21	9/24/2019	4269	Before
22	9/28/2019	3058	After
23	10/22/2019	24638	After
24	10/27/2019	4579	After
25	10/28/2019	1497	After
26	11/1/2019	3761	After
27	11/19/2019	17947	After
28	12/1/2019	11790	After
29	12/3/2019	1725	After
30	1/4/2020	32255	After
31	1/17/2020	12793	After

3. Copy the "Before Improvement" data in cells **A1:B21** and Paste Values to the template at **A1**. This will overwrite the formulas in Column B. Adjust the template column widths as necessary.

	A	B	C	D	E
1	Date ADE	Doses Between Events			
2	7/2/2019	758			
3	7/14/2019	11911			
4	7/21/2019	7593			
5	7/22/2019	398		Rare Events G Control Chart	
6	7/24/2019	2516			
7	7/26/2019	1147			
8	7/29/2019	3322		Add Data	
9	8/13/2019	15414			
10	8/14/2019	626			
11	8/15/2019	752			
12	8/23/2019	7819			
13	8/25/2019	2445			
14	8/30/2019	4937			
15	9/4/2019	4663			
16	9/4/2019	175			
17	9/6/2019	2187			
18	9/7/2019	592			
19	9/11/2019	3721			
20	9/20/2019	9317			
21	9/24/2019	4269			
22					

4. Click the **Rare Events G Control Chart** button to create the Rare Events G Control Chart:

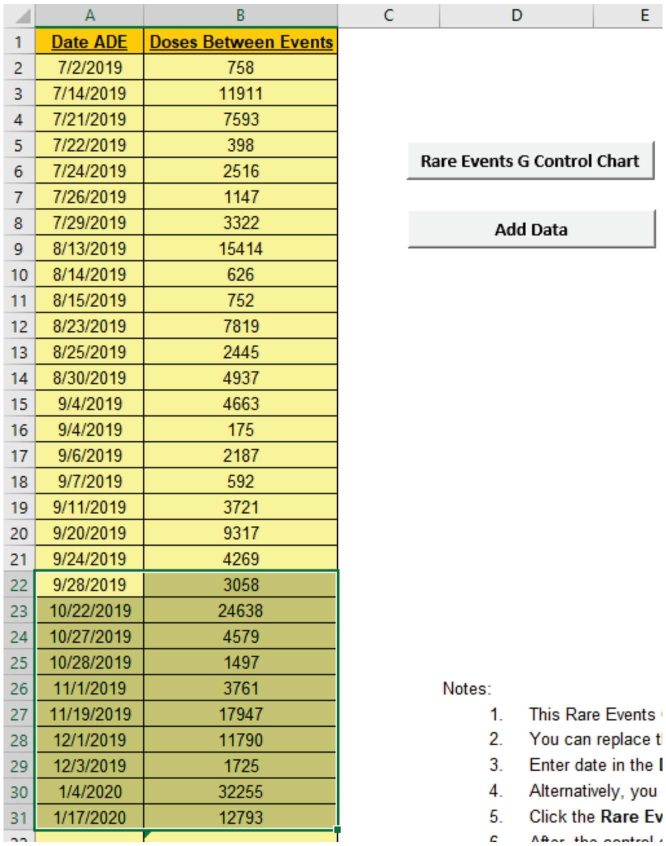


5. This confirms that the process is "in-control".

6. Although the process is "in-control", as a critical health care metric, efforts were made to improve the process and this data will be now added to the chart. Switch back to **Doses Dispensed Between Adverse Drug Events.xlsx** (**Sheet 1** tab). Select and copy the "After Improvement" date/times in cells **A22: B31** as shown.

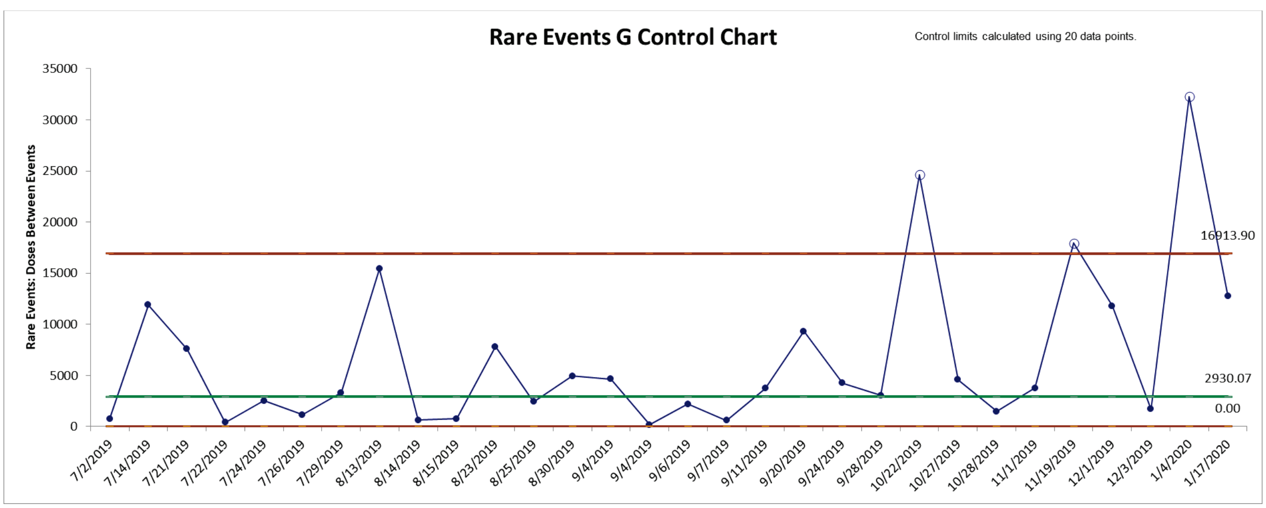
	A	B	C
1	Date ADE	Doses Between Events	Before/After Improvement
2	7/2/2019	758	Before
3	7/14/2019	11911	Before
4	7/21/2019	7593	Before
5	7/22/2019	398	Before
6	7/24/2019	2516	Before
7	7/26/2019	1147	Before
8	7/29/2019	3322	Before
9	8/13/2019	15414	Before
10	8/14/2019	626	Before
11	8/15/2019	752	Before
12	8/23/2019	7819	Before
13	8/25/2019	2445	Before
14	8/30/2019	4937	Before
15	9/4/2019	4663	Before
16	9/4/2019	175	Before
17	9/6/2019	2187	Before
18	9/7/2019	592	Before
19	9/11/2019	3721	Before
20	9/20/2019	9317	Before
21	9/24/2019	4269	Before
22	9/28/2019	3058	After
23	10/22/2019	24638	After
24	10/27/2019	4579	After
25	10/28/2019	1497	After
26	11/1/2019	3761	After
27	11/19/2019	17947	After
28	12/1/2019	11790	After
29	12/3/2019	1725	After
30	1/4/2020	32255	After
31	1/17/2020	12793	After

7. Switch to the Rare Events G Control Chart template and Paste Values to cell **A22** as shown. This will overwrite the formulas in Column B. Adjust the template column widths as necessary.



	A	B
1	Date ADE	Doses Between Events
2	7/2/2019	758
3	7/14/2019	11911
4	7/21/2019	7593
5	7/22/2019	398
6	7/24/2019	2516
7	7/26/2019	1147
8	7/29/2019	3322
9	8/13/2019	15414
10	8/14/2019	626
11	8/15/2019	752
12	8/23/2019	7819
13	8/25/2019	2445
14	8/30/2019	4937
15	9/4/2019	4663
16	9/4/2019	175
17	9/6/2019	2187
18	9/7/2019	592
19	9/11/2019	3721
20	9/20/2019	9317
21	9/24/2019	4269
22	9/28/2019	3058
23	10/22/2019	24638
24	10/27/2019	4579
25	10/28/2019	1497
26	11/1/2019	3761
27	11/19/2019	17947
28	12/1/2019	11790
29	12/3/2019	1725
30	1/4/2020	32255
31	1/17/2020	12793

8. Click the **Add Data** button to add the new **Doses Between Events** data to the Rare Events G Control Chart:



9. This confirms that the process is now "out-of-control", so the improvement efforts to decrease the occurrence of ADEs and increase the Doses Between have been successful. Note that the control limits should be recalculated with just the "After Improvement" data when 20 data points are available.

Template Notes and Reference:

1. This Rare Events G Control Chart template should be used with days or units/opportunities between (typically adverse) rare events.
2. You can replace the **Date/Time** and **Days Between** column headings with any headings that you wish.
3. Enter date in the **Date** column. Days between are automatically calculated and entered into **Days Between** column. Cell **B2** is not used in this case.
4. Alternatively, you may manually enter data in **Days/Units Between**. Note, this will overwrite the cell formula.
5. Click the **Rare Events G Control Chart** button to create a control chart. This will overwrite any existing control chart.
6. After the control chart has been created and additional new **Date** or **Days/Units Between** data entered, click the **Add Data** button to add the data to the existing chart.
7. An "out-of-control" signal above the UCL is desirable, indicating a significant increase in days or units/opportunities between adverse rare events.
8. The calculation of control limits is based on the geometric distribution. The center line is the theoretical median for a geometric distribution = 0.693 * mean.
9. Reference: Provost L, Murray S., *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco: Jossey-Bass, 2011, pp. 228-229.

Rare Events Prob G Chart - Introduction

The Rare Events Prob (Probability-Based) G chart (or Geometric chart) is an alternative to a standard attribute chart when the adverse event of interest is rare and discrete opportunities between events are counted (e.g., number of units or days between). The use of probability-based control limits is recommended in order to properly control the Type I (false alarm) error rate.

The calculation of control limits is based on the geometric distribution. Event probability and alpha are used to compute the non-symmetrical limits. An "out-of-control" signal above the UCL is desirable, indicating a significant increase in units/opportunities or days between adverse rare events. See Benneyan, 2001.

$\mathrm{g} =$ number of opportunities or units between incidents

$\bar{g} =$ average of g's

The event probability is user specified or estimated as:

$$\hat{p} = \left(\frac{1}{\bar{\mathrm{g}} + 1}\right)\left(\frac{N - 1}{N}\right)$$

This is the minimum variances unbiased estimator.

The probability-based control limits are calculated as:

$$UCL = \frac{\ln(\alpha_{\mathrm{UCL}})}{\ln(1 - p)} - 1$$

$$CL = \frac{\ln(0.5)}{\ln(1 - p)} - 1$$

$$LCL = \max\left(0, \frac{\ln(1 - \alpha_{\mathrm{UCL}})}{\ln(1 - p)} - 1\right)$$

where α_{UCL} is user specified.

SigmaXL adjusts the Benneyan CL and LCL by subtracting 1 as done in the UCL and does not round the limit values. Opportunities between events are counted as 0, 1, 2, 3, … ($a = 0$ in Benneyan Table 3).

Rare Events Prob G Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Rare Events > Rare Events Prob G**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Rare Events > Rare Events Prob G.**

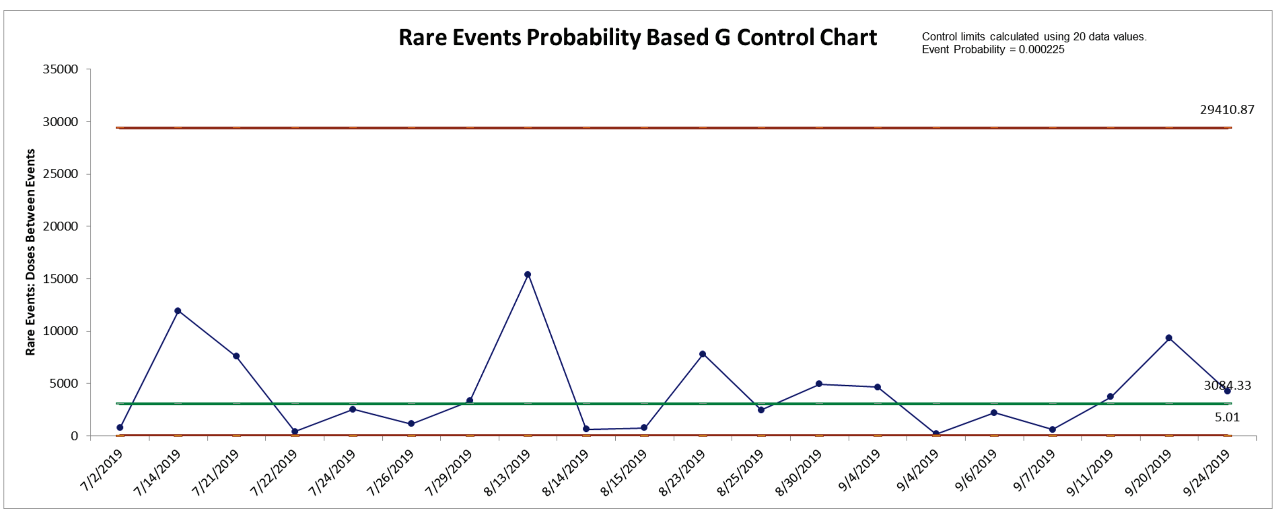
2. Open **Doses Dispensed Between Adverse Drug Events.xlsx** (**Sheet 1** tab). Adverse Drug Events (ADE) are rare, so Doses Dispensed Between ADEs is used as a health care metric. The data is discrete and distributed geometrically, so the G chart is appropriate to use here. The event dates 7/2/2019 to 9/24/2019 are "Before Improvement" and the Doses Between ADEs will be used to calculate the control limits. Dates 9/28/2019 to 1/17/2020 are "After Improvement" and the Doses Between ADEs will be added to the control chart. The goal is to decrease the occurrence of ADEs which will increase the doses between, with an "out-of-control" signal above the UCL being desirable and confirmation of the improvement effort.

	A	B	C
1	Date ADE	Doses Between Events	Before/After Improvement
2	7/2/2019	758	Before
3	7/14/2019	11911	Before
4	7/21/2019	7593	Before
5	7/22/2019	398	Before
6	7/24/2019	2516	Before
7	7/26/2019	1147	Before
8	7/29/2019	3322	Before
9	8/13/2019	15414	Before
10	8/14/2019	626	Before
11	8/15/2019	752	Before
12	8/23/2019	7819	Before
13	8/25/2019	2445	Before
14	8/30/2019	4937	Before
15	9/4/2019	4663	Before
16	9/4/2019	175	Before
17	9/6/2019	2187	Before
18	9/7/2019	592	Before
19	9/11/2019	3721	Before
20	9/20/2019	9317	Before
21	9/24/2019	4269	Before
22	9/28/2019	3058	After
23	10/22/2019	24638	After
24	10/27/2019	4579	After
25	10/28/2019	1497	After
26	11/1/2019	3761	After
27	11/19/2019	17947	After
28	12/1/2019	11790	After
29	12/3/2019	1725	After
30	1/4/2020	32255	After
31	1/17/2020	12793	After

3. Copy the "Before Improvement" data in cells **A1:B21** and Paste Values to the template at **A1**. This will overwrite the formulas in Column B. Adjust the template column widths as necessary.

	A	B	C	D
1	Date ADE	Doses Between Events	Alpha UCL:	0.00135
2	7/2/2019	758	Event Probability:	
3	7/14/2019	11911		
4	7/21/2019	7593	Rare Events Prob G Chart	
5	7/22/2019	398		
6	7/24/2019	2516	Add Data	
7	7/26/2019	1147		
8	7/29/2019	3322		
9	8/13/2019	15414		
10	8/14/2019	626		
11	8/15/2019	752		
12	8/23/2019	7819		
13	8/25/2019	2445		
14	8/30/2019	4937		
15	9/4/2019	4663		
16	9/4/2019	175		
17	9/6/2019	2187		
18	9/7/2019	592		
19	9/11/2019	3721		
20	9/20/2019	9317		
21	9/24/2019	4269		

4. We will use the default **Alpha UCL** = 0.00135 (this corresponds to the one-sided probability for UCL in a classical Shewhart control chart with 3 sigma limits). Click the **Rare Events Prob G Chart** button to create the Rare Events Probability Based G Control Chart:



5. This confirms that the process is "in-control". Note that the probability based UCL is much larger than the one calculated above for the regular G chart.

6. **Tip:** If the UCL is considered to be "too large" for practical use, adjust the **Alpha UCL** to a larger value, for example, .01 instead of .00135 and recreate the control chart. Note however that this will increase the Type I (false alarm) rate.

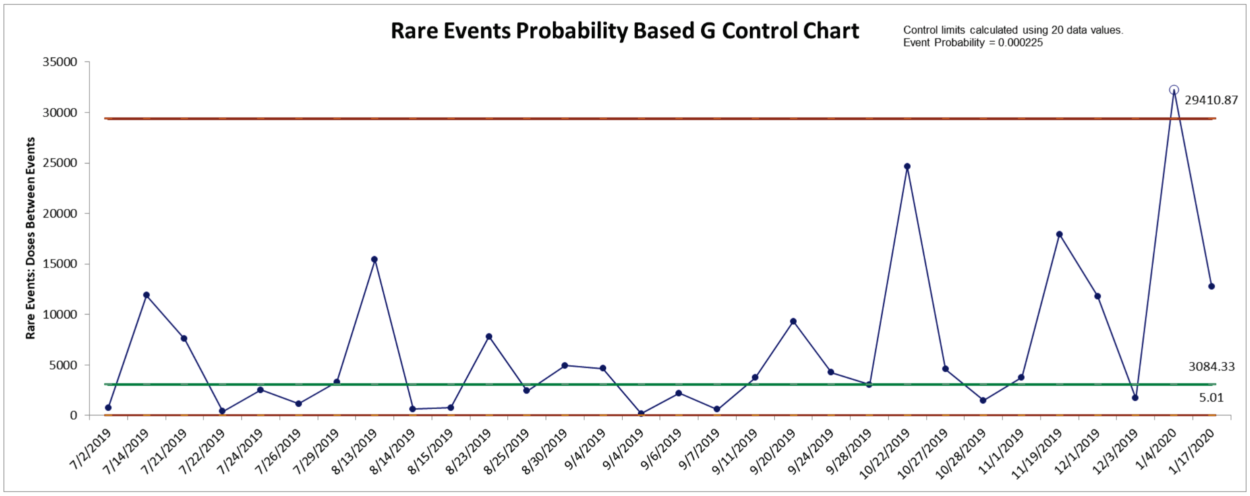
7. Although the process is "in-control", as a critical health care metric, efforts were made to improve the process and this data will be now added to the chart. Switch back to **Doses Dispensed Between Adverse Drug Events.xlsx** (**Sheet 1** tab). Select and copy the "After Improvement" date/times in cells **A22: B31** as shown.

	A	B	C
1	Date ADE	Doses Between Events	Before/After Improvement
2	7/2/2019	758	Before
3	7/14/2019	11911	Before
4	7/21/2019	7593	Before
5	7/22/2019	398	Before
6	7/24/2019	2516	Before
7	7/26/2019	1147	Before
8	7/29/2019	3322	Before
9	8/13/2019	15414	Before
10	8/14/2019	626	Before
11	8/15/2019	752	Before
12	8/23/2019	7819	Before
13	8/25/2019	2445	Before
14	8/30/2019	4937	Before
15	9/4/2019	4663	Before
16	9/4/2019	175	Before
17	9/6/2019	2187	Before
18	9/7/2019	592	Before
19	9/11/2019	3721	Before
20	9/20/2019	9317	Before
21	9/24/2019	4269	Before
22	9/28/2019	3058	After
23	10/22/2019	24638	After
24	10/27/2019	4579	After
25	10/28/2019	1497	After
26	11/1/2019	3761	After
27	11/19/2019	17947	After
28	12/1/2019	11790	After
29	12/3/2019	1725	After
30	1/4/2020	32255	After
31	1/17/2020	12793	After

8. Switch to the Rare Events Prob G Chart template and Paste Values to cell **A22** as shown. This will overwrite the formulas in Column B. Adjust the template column widths as necessary.

	A	B	C	D
1	Date ADE	Doses Between Events	Alpha UCL:	0.00135
2	7/2/2019	758	Event Probability:	
3	7/14/2019	11911		
4	7/21/2019	7593	Rare Events Prob G Chart	
5	7/22/2019	398		
6	7/24/2019	2516	Add Data	
7	7/26/2019	1147		
8	7/29/2019	3322		
9	8/13/2019	15414		
10	8/14/2019	626		
11	8/15/2019	752		
12	8/23/2019	7819		
13	8/25/2019	2445		
14	8/30/2019	4937		
15	9/4/2019	4663		
16	9/4/2019	175		
17	9/6/2019	2187		
18	9/7/2019	592		
19	9/11/2019	3721		
20	9/20/2019	9317		
21	9/24/2019	4269		
22	9/28/2019	3058		
23	10/22/2019	24638		
24	10/27/2019	4579		
25	10/28/2019	1497		
26	11/1/2019	3761		Notes:
27	11/19/2019	17947		1. This Rare I
28	12/1/2019	11790		2. You can re
29	12/3/2019	1725		3. Enter date
30	1/4/2020	32255		4. Alternative
31	1/17/2020	12793		5. Enter **Alph**

9. Click the **Add Data** button to add the new **Doses Between Events** data to the Rare Events Probability Based G Control Chart:



10. This confirms that the process is now “out-of-control”, so the improvement efforts to decrease the occurrence of ADEs and increase the Doses Between have been successful. Note that the control limits should be recalculated with just the “After Improvement” data when 20 data points are available.

Template Notes and Reference:

1. This Rare Events Prob (Probability-Based) G Control Chart template should be used with days or units/opportunities between (typically adverse) rare events.
2. You can replace the **Date/Time** and **Days Between** column headings with any headings that you wish.
3. Enter date in the **Date** column. Days between are automatically calculated and entered into **Days Between** column. Cell **B2** is not used in this case.
4. Alternatively, you may manually enter data in **Days/Units Between**. Note, this will overwrite the cell formulas.
5. Enter **Alpha UCL**, typically 0.00135, corresponding to the one-sided probability for UCL in a classical Shewhart control chart with 3 sigma limits. This will also be applied to the LCL if greater than 0.
6. The calculation of control limits is based on the geometric distribution. Event probability and alpha are used to compute the non-symmetrical limits. Calculated event probability is noted on the control chart.
7. Optionally enter the historical **Event Probability.**
8. Click the **Rare Events Prob G Chart** button to create a control chart. This will overwrite any existing control chart.

9. After the control chart has been created and additional new **Date** or **Days/Units Between** data entered, click the **Add Data** button to add the data to the existing chart. Control limits will be calculated using the original chart event probability or specified event probability.

10. **Add Data** should only be used if there are at least 20 observations in the original chart or known historical Event Probability has been specified.

11. An "out-of-control" signal above the UCL is desirable, indicating a significant increase in days or units/opportunities between adverse rare events.

12. Reference: Benneyan, J.C. (2001), "Performance of Number-Between g-Type Statistical Control Charts for Monitoring Adverse Events", *Health Care Management Science*, 4, pp. 319–336, Table 3.

Part H – Control Chart Templates: Time-Weighted

Exponentially Weighted Moving Average (EWMA) Chart - Introduction

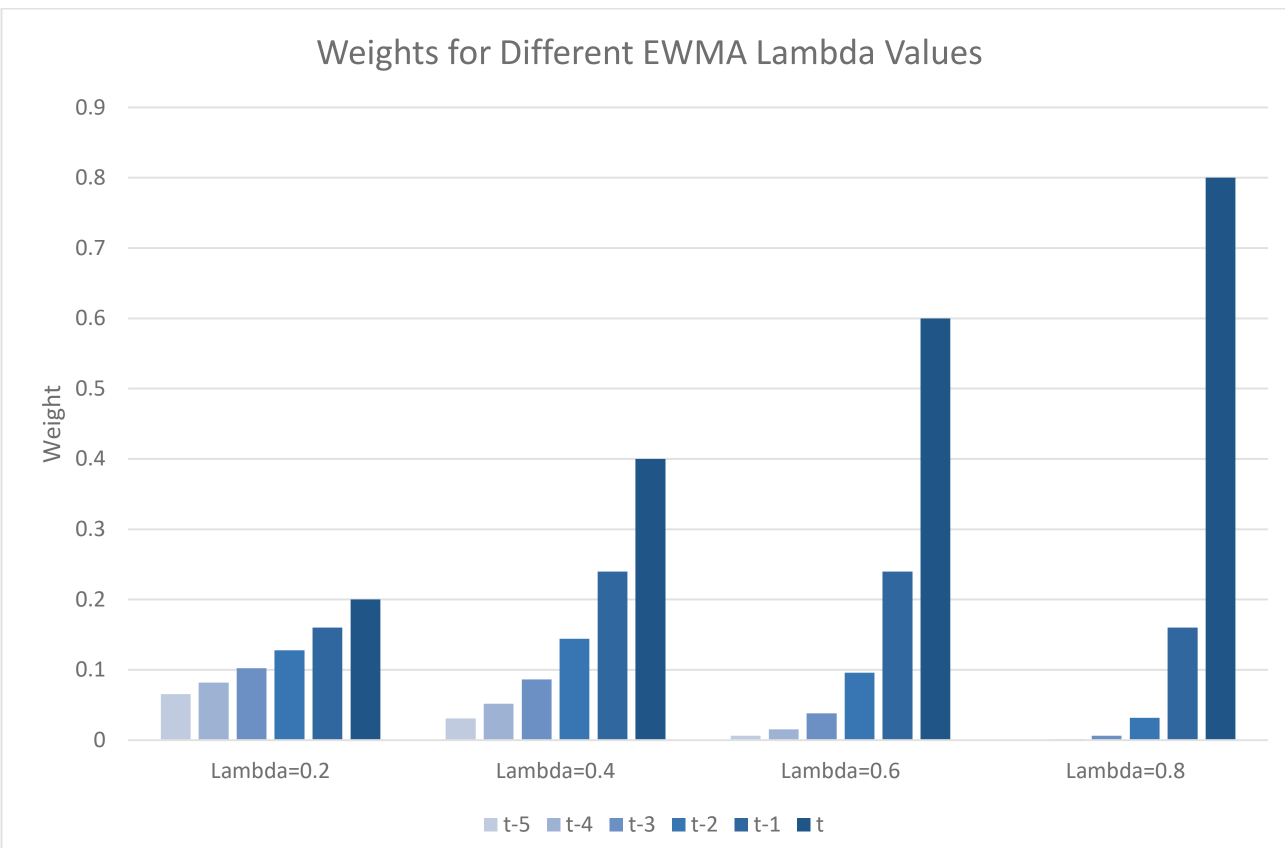
The EWMA control chart uses weighted averages, where the weights decrease exponentially as observations come from further in the past with the smallest weights associated with the oldest observations.

The formula used for the EWMA statistic is given as:

$$\mathrm{EWMA}_t = \lambda X_t + (1 - \lambda)\mathrm{EWMA}_{t-1}$$

with the starting value EWMA_0 estimated as the data mean and the smoothing parameter λ specified according to desired average run length characteristics (see Average Run Length (ARL) Calculators).

Weights for different λ values are shown graphically:



The control limits for an EWMA chart are calculated as:

$$\mathrm{UCL} = \mu_0 + K\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2t}]}$$

$$\mathrm{CL} = \mu_0$$

$$\mathrm{LCL} = \mu_0 - K\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2t}]}$$

where μ_0 is specified as the historical mean or estimated as:

$$\hat{\mu}_0 = \bar{x}$$

and σ is specified as the historical standard deviation or estimated as:

$$\widehat{\sigma} = \overline{MR}/d_2\,.$$

Note, the *K* multiplier is called *L* in Montgomery (2013).

Exponentially Weighted Moving Average (EWMA) Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Time Weighted > EWMA**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Time Weighted > EWMA.**

2. Open **Montgomery Table 9.1.xlsx** (**Sheet 1** tab). This is simulated data from Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley, pp. 415. Table 9.1. Samples 1 to 20 are drawn from a random normal distribution with population mean µ = 10 and population standard deviation σ = 1. Samples 21 to 30 have a mean = 11 and standard deviation = 1, so the process has shifted by 1 sigma and is "out-of-control". While this is a small shift in process mean, it is something that we want to be able to detect and correct as quickly as possible.

	A	B	C	D
1	Sample	X	Process Mean	Process StDev
2	1	9.45	10	1
3	2	7.99	10	1
4	3	9.29	10	1
5	4	11.66	10	1
6	5	12.16	10	1
7	6	10.18	10	1
8	7	8.04	10	1
9	8	11.46	10	1
10	9	9.2	10	1
11	10	10.34	10	1
12	11	9.03	10	1
13	12	11.47	10	1
14	13	10.51	10	1
15	14	9.4	10	1
16	15	10.08	10	1
17	16	9.37	10	1
18	17	10.62	10	1
19	18	10.31	10	1
20	19	8.52	10	1
21	20	10.84	10	1
22	21	10.9	11	1
23	22	9.33	11	1
24	23	12.29	11	1
25	24	11.5	11	1
26	25	10.6	11	1
27	26	11.08	11	1
28	27	10.38	11	1
29	28	11.62	11	1
30	29	11.31	11	1
31	30	10.52	11	1

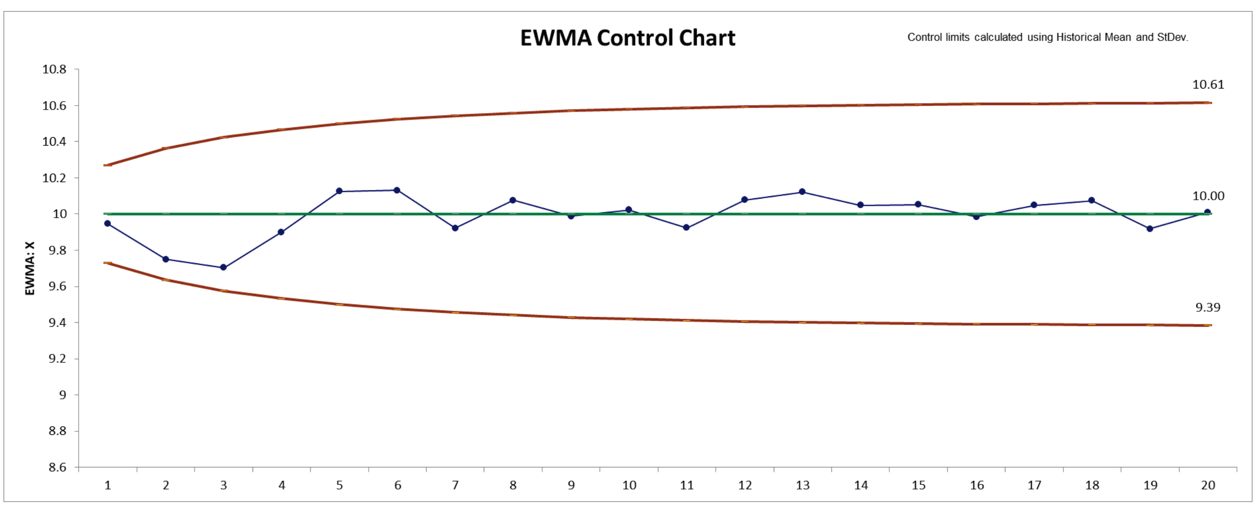
3. Copy cells **A1:B21** and Paste Values to the template at **A1**. Use the default **Weight (Lambda)** = 0.1, **K** = 2.7. Specify **Historical Mean** = 10 and **Historical StDev** = 1 as shown:

	A	B	C	D
1	Sample	X	Weight (Lambda):	0.1
2	1	9.45	K:	2.7
3	2	7.99	Historical Mean:	10
4	3	9.29	Historical StDev:	1
5	4	11.66		
6	5	12.16		
7	6	10.18		
8	7	8.04		
9	8	11.46		
10	9	9.2		
11	10	10.34		
12	11	9.03		
13	12	11.47		
14	13	10.51		
15	14	9.4		
16	15	10.08		
17	16	9.37		
18	17	10.62		
19	18	10.31		
20	19	8.52		
21	20	10.84		
22				

EWMA Control Chart

Add Data

4. Click the **EWMA Control Chart** button to create the EWMA Control Chart:



5. This confirms that the process is “in-control”.

6. Switch back to **Montgomery Table 9.1.xlsx** (**Sheet 1** tab). Select and copy cells **A22: B31** as shown. This is the data with shifted mean.

	A	B	C	D
1	Sample	X	Process Mean	Process StDev
2	1	9.45	10	1
3	2	7.99	10	1
4	3	9.29	10	1
5	4	11.66	10	1
6	5	12.16	10	1
7	6	10.18	10	1
8	7	8.04	10	1
9	8	11.46	10	1
10	9	9.2	10	1
11	10	10.34	10	1
12	11	9.03	10	1
13	12	11.47	10	1
14	13	10.51	10	1
15	14	9.4	10	1
16	15	10.08	10	1
17	16	9.37	10	1
18	17	10.62	10	1
19	18	10.31	10	1
20	19	8.52	10	1
21	20	10.84	10	1
22	21	10.9	11	1
23	22	9.33	11	1
24	23	12.29	11	1
25	24	11.5	11	1
26	25	10.6	11	1
27	26	11.08	11	1
28	27	10.38	11	1
29	28	11.62	11	1
30	29	11.31	11	1
31	30	10.52	11	1

7. Switch to the EWMA template and Paste Values to cell **A22** as shown.

	A	B	C	D
1	Sample	X	Weight (Lambda):	0.1
2	1	9.45	K:	2.7
3	2	7.99	Historical Mean:	10
4	3	9.29	Historical StDev:	1
5	4	11.66		
6	5	12.16		
7	6	10.18		
8	7	8.04		
9	8	11.46		
10	9	9.2		
11	10	10.34		
12	11	9.03		
13	12	11.47		
14	13	10.51		
15	14	9.4		
16	15	10.08		
17	16	9.37		
18	17	10.62		
19	18	10.31		
20	19	8.52		
21	20	10.84		
22	21	10.9		
23	22	9.33		
24	23	12.29		
25	24	11.5		
26	25	10.6		
27	26	11.08		
28	27	10.38		
29	28	11.62		
30	29	11.31		
31	30	10.52		

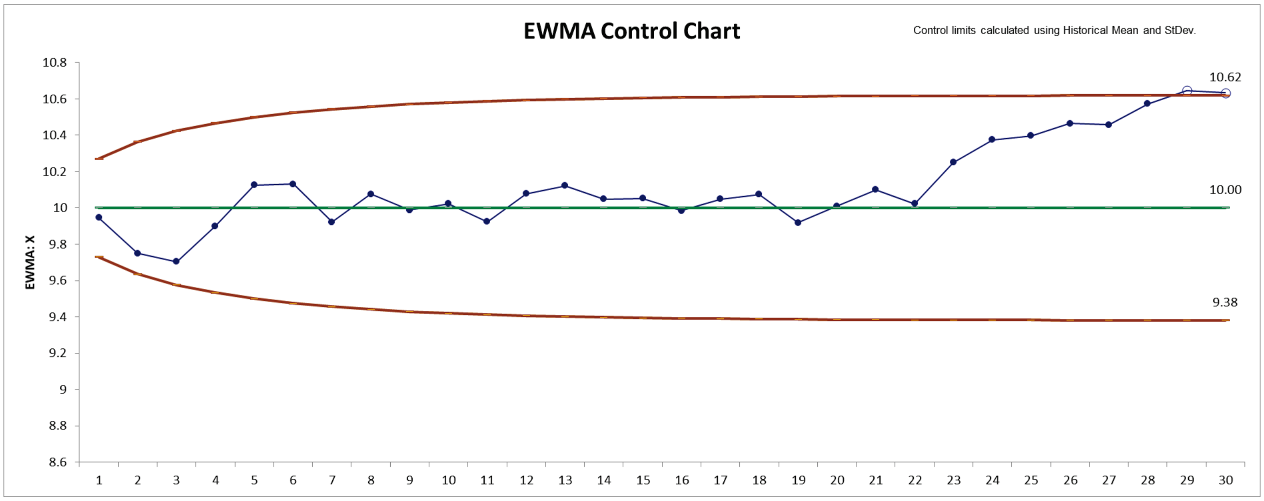
EWMA Control Chart

Add Data

Notes:

1. This Exponer
2. You can repl:
3. Enter labels i
4. Enter the EW
5. Enter the EW

8. Click the **Add Data** button to add this data to the EWMA Control Chart:



9. This confirms that the process is now “out-of-control” with signals at samples 29 and 30. This matches Figure 9.7 in Montgomery.

Template Notes and Reference

1. This Exponentially Weighted Moving Average (EWMA) Control Chart template should be used with continuous data. The data must be in chronological time-sequence order.

2. You can replace the **X-Axis Label** and **Data** column headings with any headings that you wish. Enter your data in the **Data** column.

3. Enter labels in **X-Axis Label** column. Labels can be Date, Time, Name, or other text information. These labels are optional and will appear on the horizontal X-Axis of the EWMA Control Chart.

4. Enter the EWMA **Weight (Lambda)** in cell D1. This is a value between 0 and 1 and controls the amount of influence that previous observations have on the current EWMA statistic. A value near 1 puts almost all weight on the current observation, making it resemble a Shewhart chart. For values near 0, a small weight is applied to almost all of the past observations, so the EWMA chart performance is similar to that of a CUSUM chart.

5. Enter the EWMA **K** Sigma multiplier in cell D2. This is a value typically between 2 and 4. It is also referred to as L, but SigmaXL uses K to avoid confusion with Lambda.

6. **Weight (Lambda)** and **K** values affect the Average Run Length (ARL) characteristics. To determine optimal EWMA parameter values and calculate ARL, click **SigmaXL > Templates and Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > EWMA ARL**.

7. **Historical Mean** (D3) and **Historical StDev** (D4) are optional. Enter values if process mean and standard deviation are known.

8. Click the **EWMA Control Chart** button to create an EWMA Control Chart.

9. After the control chart has been created and additional new data entered into the **Data** column, click the **Add Data** button to add the data to the existing chart. Control limits will be calculated using the original chart mean and stdev or specified Historical Mean/StDev.

10. **Add Data** should only be used if there are at least 20 observations in the original chart, or Historical Mean and StDev have been specified.

11. **Weight (Lambda)** and **K** parameters are dynamic. If they are modified, the chart will automatically update with the new parameters. However, they should be selected prior to creating the EWMA chart. Data values and out-of-control formatting are refreshed only when the buttons are used.

12. Reference: Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley, pp. 433-438.

Tabular Cumulative Sum (CUSUM) Chart - Introduction

The CUSUM chart plots the cumulative sums of deviations of sample values from a target value. Because they combine information from several samples, cumulative sum charts are more effective than Shewhart charts for detecting small process shifts. There are two ways to represent CUSUMs: the tabular (or algorithmic) CUSUM, and the V-mask form of the CUSUM (see Montgomery, 2013). SigmaXL utilizes the Tabular CUSUM.

The formulas used for the Tabular CUSUM statistics are given as:

$$C_1^+ = \max[0, x_t - (\mu_0 + k\sigma) + FIR\sigma]$$
$$C_1^- = \min[0, x_t - (\mu_0 - k\sigma) + FIR\sigma]$$

$$C_t^+ = \max[0, x_t - (\mu_0 + k\sigma) + C_{t-1}^+]$$
$$C_t^- = \min[0, x_t - (\mu_0 - k\sigma) + C_{t-1}^-]$$

where μ_0 is specified as the Target or estimated as:

$$\hat{\mu}_0 = \bar{x}$$

and σ is specified as the historical standard deviation or estimated as:

$$\widehat{\sigma} = \overline{MR}/d_2$$

and k and FIR are specified.

k is the reference (or slack) value, typically set to 0.5. It sets the size of mean shift ($2k\sigma$) that you would like to detect quickly, so 0.5 denotes rapid detection of a shift in mean = 1 σ.

FIR is the optional fast initial response (or headstart) value. This sets the initial CUSUM statistic so that it improves the sensitivity to a mean shift at startup.

The Tabular CUSUM control limits are calculated as:

$$\mathrm{UCL} = h\sigma$$

$$\mathrm{CL} = 0$$

$$\mathrm{LCL} = -h\sigma$$

where h is the decision interval, typically set to 4 or 5.

If FIR is used, it is typically set to h /2.

Tabular Cumulative Sum (CUSUM) Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Time Weighted > CUSUM**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Time Weighted > CUSUM.**

2. Open **Montgomery Table 9.1.xlsx** (**Sheet 1** tab). This is simulated data from Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley, pp. 415. Table 9.1. Samples 1 to 20 are drawn from a random normal distribution with population mean µ = 10 and population standard deviation σ = 1. Samples 21 to 30 have a mean = 11 and standard deviation = 1, so the process has shifted by 1 sigma and is “out-of-control”. While this is a small shift in process mean, it is something that we want to be able to detect and correct as quickly as possible.

	A	B	C	D
1	Sample	X	Process Mean	Process StDev
2	1	9.45	10	1
3	2	7.99	10	1
4	3	9.29	10	1
5	4	11.66	10	1
6	5	12.16	10	1
7	6	10.18	10	1
8	7	8.04	10	1
9	8	11.46	10	1
10	9	9.2	10	1
11	10	10.34	10	1
12	11	9.03	10	1
13	12	11.47	10	1
14	13	10.51	10	1
15	14	9.4	10	1
16	15	10.08	10	1
17	16	9.37	10	1
18	17	10.62	10	1
19	18	10.31	10	1
20	19	8.52	10	1
21	20	10.84	10	1
22	21	10.9	11	1
23	22	9.33	11	1
24	23	12.29	11	1
25	24	11.5	11	1
26	25	10.6	11	1
27	26	11.08	11	1
28	27	10.38	11	1
29	28	11.62	11	1
30	29	11.31	11	1
31	30	10.52	11	1

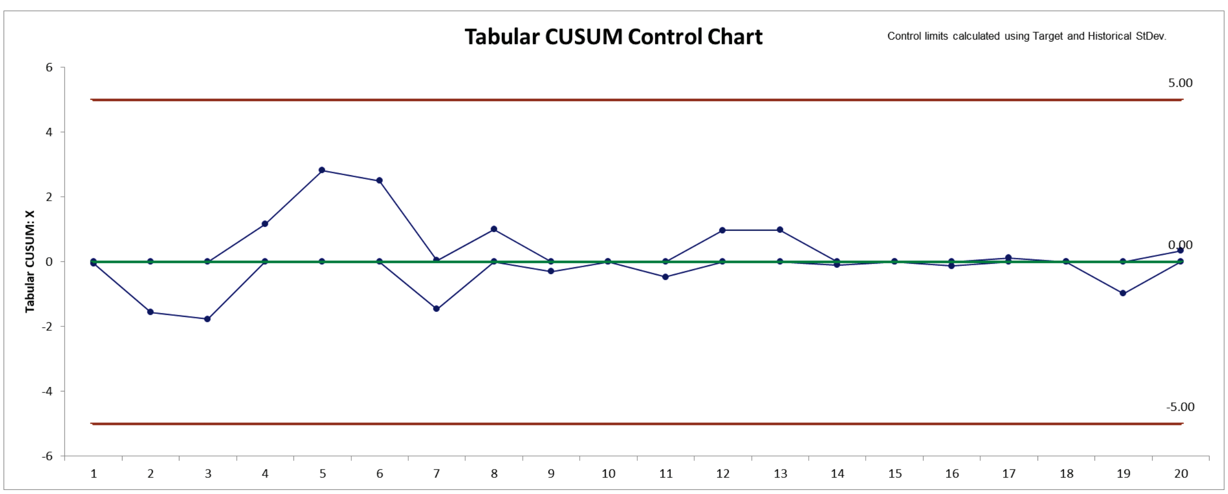
3. Copy cells **A1:B21** and Paste Values to the template at **A1**. Use the default **k** = 0.5, **h** = 5, **FIR** = 0. Specify **Target** = 10 and **Historical StDev** = 1 as shown:

	A	B	C	D
1	Sample	X	k:	0.5
2	1	9.45	h:	5
3	2	7.99	FIR:	0
4	3	9.29	Target:	10
5	4	11.66	Historical StDev:	1
6	5	12.16		
7	6	10.18		
8	7	8.04		
9	8	11.46		
10	9	9.2		
11	10	10.34		
12	11	9.03		
13	12	11.47		
14	13	10.51		
15	14	9.4		
16	15	10.08		
17	16	9.37		
18	17	10.62		
19	18	10.31		
20	19	8.52		
21	20	10.84		

Tabular CUSUM Chart

Add Data

4. Click the **Tabular CUSUM Chart** button to create the Tabular CUSUM Control Chart:



5. This confirms that the process is “in-control”.

6. Switch back to **Montgomery Table 9.1.xlsx** (**Sheet 1** tab). Select and copy cells **A22: B31** as shown. This is the data with shifted mean.

	A	B	C	D
1	Sample	X	Process Mean	Process StDev
2	1	9.45	10	1
3	2	7.99	10	1
4	3	9.29	10	1
5	4	11.66	10	1
6	5	12.16	10	1
7	6	10.18	10	1
8	7	8.04	10	1
9	8	11.46	10	1
10	9	9.2	10	1
11	10	10.34	10	1
12	11	9.03	10	1
13	12	11.47	10	1
14	13	10.51	10	1
15	14	9.4	10	1
16	15	10.08	10	1
17	16	9.37	10	1
18	17	10.62	10	1
19	18	10.31	10	1
20	19	8.52	10	1
21	20	10.84	10	1
22	21	10.9	11	1
23	22	9.33	11	1
24	23	12.29	11	1
25	24	11.5	11	1
26	25	10.6	11	1
27	26	11.08	11	1
28	27	10.38	11	1
29	28	11.62	11	1
30	29	11.31	11	1
31	30	10.52	11	1

7. Switch to the Tabular CUSUM template and Paste Values to cell **A22** as shown.

	A	B	C	D
1	Sample	X	k:	0.5
2	1	9.45	h:	5
3	2	7.99	FIR:	0
4	3	9.29	Target:	10
5	4	11.66	Historical StDev:	1
6	5	12.16		
7	6	10.18		
8	7	8.04		
9	8	11.46		
10	9	9.2		
11	10	10.34		
12	11	9.03		
13	12	11.47		
14	13	10.51		
15	14	9.4		
16	15	10.08		
17	16	9.37		
18	17	10.62		
19	18	10.31		
20	19	8.52		
21	20	10.84		
22	21	10.9		
23	22	9.33		
24	23	12.29		
25	24	11.5		
26	25	10.6		
27	26	11.08		
28	27	10.38		
29	28	11.62		
30	29	11.31		
31	30	10.52		

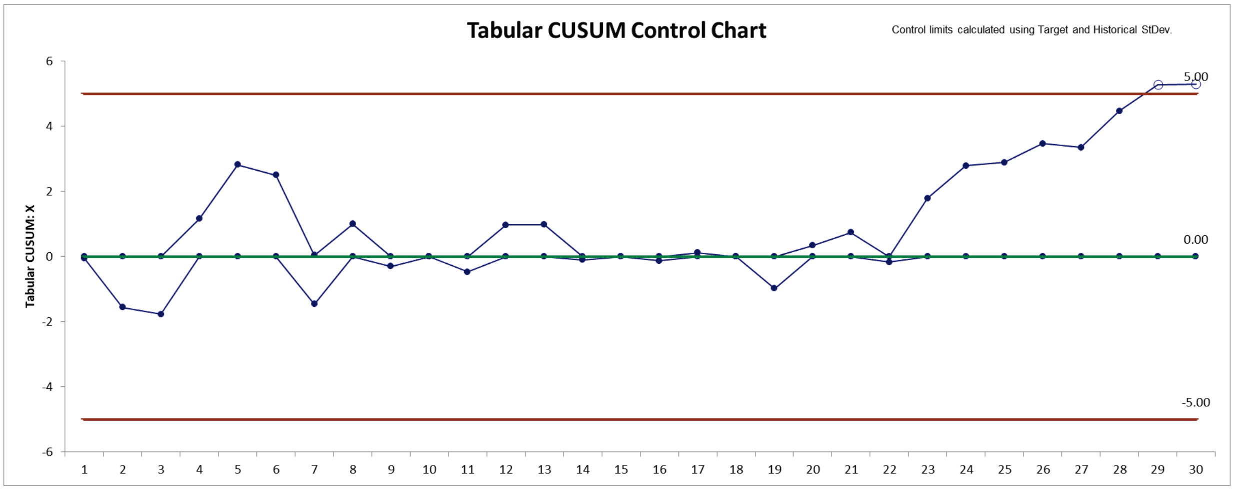
Tabular CUSUM Chart

Add Data

Notes:

1. This Tabula
2. You can re
3. Enter label
4. Enter the C
5. Enter the C

8. Click the **Add Data** button to add this data to the Tabular CUSUM Control Chart:



9. This confirms that the process is now "out-of-control" with signals at samples 29 and 30.

Template Notes and Reference:

1. This Tabular Cumulative Sum (CUSUM) Control Chart template should be used with continuous data. The data must be in chronological time-sequence order.
2. You can replace the **X-Axis Label** and **Data** column headings with any headings that you wish. Enter your data in the **Data** column.
3. Enter labels in the **X-Axis Label** column. Labels can be Date, Time, Name, or other text information. These labels are optional and will appear on the horizontal X-Axis of the Tabular CUSUM Control Chart.
4. Enter the CUSUM **k** parameter in cell D1. This is the reference (or slack) value, typically set to 0.5. It sets the size of mean shift (2k sigma) that you would like to detect quickly, so 0.5 denotes rapid detection of a shift in mean = 1 sigma.
5. Enter the CUSUM **h** parameter in cell D2. This is the decision interval, typically set to 4 or 5. The upper and lower control limits = +/- h*StDev (MR-bar/d2). The center line = 0.
6. Enter the CUSUM **FIR** parameter in cell D3. This is the fast initial response (or headstart) value, typically set to h/2 if used, 0 otherwise. This sets the initial CUSUM statistic so that it improves the sensitivity to a mean shift at startup.
7. Optionally enter the CUSUM **Target** in cell D4. This is your process target value, typically the midpoint of your specification limits or historical mean. If you do not specify a Target, the data average will be used.
8. Optionally enter the CUSUM **Historical StDev** in cell D5. If you do not specify a Historical StDev, it will be estimated using MR-bar/d2.
9. The **h**, **k** and **FIR** parameters affect the Average Run Length (ARL) characteristics. For example, h=4 will detect a small shift more quickly than h=5, but has a shorter ARL(0) run length (higher false alarm rate).
10. To determine optimal CUSUM parameter values and calculate ARL, click **SigmaXL > Templates and Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > CUSUM ARL.**

11. Click the **Tabular CUSUM Chart** button to create a Tabular CUSUM Control Chart.
12. After the control chart has been created and additional new data entered into the **Data** column, click the **Add Data** button to add the data to the existing chart. Control limits will be calculated using the original chart mean and stdev or specified Target and Historical StDev.
13. **Add Data** should only be used if there are at least 20 observations in the original chart, or Target and Historical StDev have been specified.
14. The **k**, **h** and **FIR** parameters are dynamic. If they are modified, the chart will automatically update with the new parameters. However, they should be selected prior to creating the Tabular CUSUM chart. Data values and out-of-control formatting are refreshed only when the buttons are used.
15. Reference: Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley, pp. 418-427.

Part I – Control Chart Templates: Trend

Trend Chart - Introduction

The Trend Control Chart template should be used with continuous data. The data must be in chronological time-sequence order and have a consistent positive or negative linear trend which is inherent to the process. This is also known as a Toolwear Control Chart.

Center line (CL) is the linear regression equation. Stdev is estimated using MR-bar/1.128 of regression residuals.

Note that the regression model estimation error is not included in the calculation of the control limits. R-Square should be at least 50%, preferably greater than 80%.

Trend control limit formulas are given in Provost L, Murray S., *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco: Jossey-Bass, 2011, p. 254.

Alternatively, a regular Individuals Chart may be constructed on the regression model residuals.

Trend Chart - Example

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Trend**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Trend.**

2. Open **Trend Chart Example.xlsx** (**Sheet 1** tab). We will use Sample 1 to 20 to construct the Trend Control Chart and then add Sample 21 to 30 to this chart.

Sample	X Trend
1	10.45
2	9.99
3	12.29
4	15.66
5	17.16
6	16.18
7	15.04
8	19.46
9	18.2
10	20.34
11	20.03
12	23.47
13	23.51
14	23.4
15	25.08
16	25.37
17	27.62
18	28.31
19	27.52
20	30.84
21	34.4
22	33.83
23	37.79
24	38
25	38.1
26	39.58
27	39.88
28	42.12
29	42.81
30	43.02

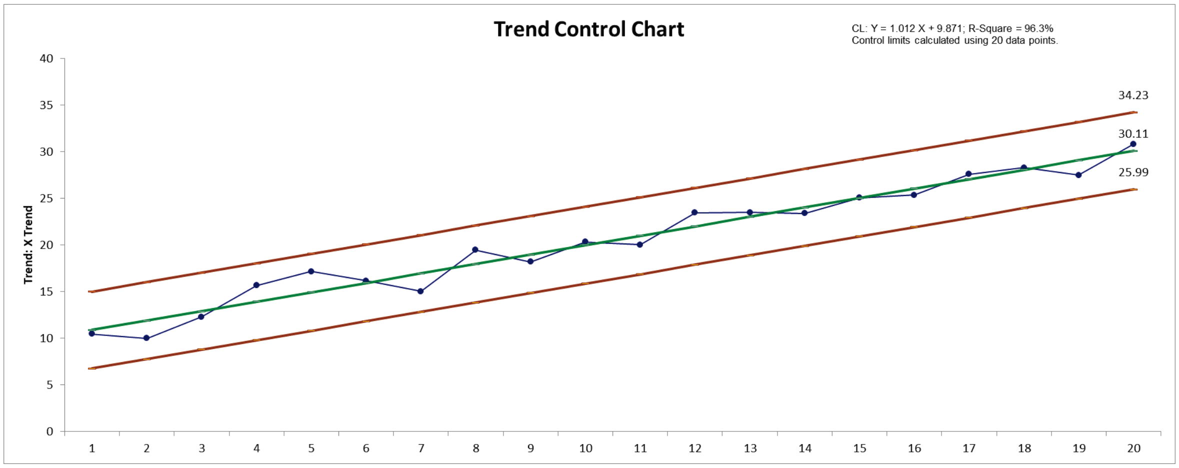
3. Copy the data in cells **A1:B21** and Paste Values to the template at **A1**.

Sample	X Trend
1	10.45
2	9.99
3	12.29
4	15.66
5	17.16
6	16.18
7	15.04
8	19.46
9	18.2
10	20.34
11	20.03
12	23.47
13	23.51
14	23.4
15	25.08
16	25.37
17	27.62
18	28.31
19	27.52
20	30.84

Trend Control Chart

Add Data

4. Click the **Trend Control Chart** button to create the Trend Control:

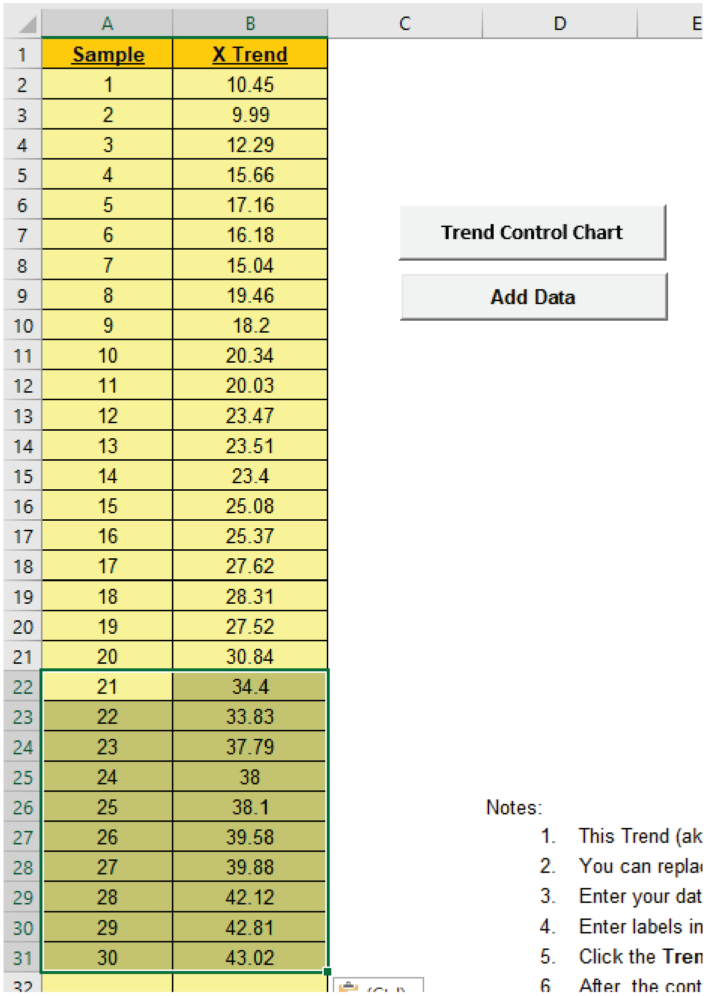


5. This confirms that the process is "in-control" after taking the trend into account. The regression equation and R-Square are reported, with R-Square = 96.3% well above the recommended 80%.

6. Switch back to **Trend Chart Example** (**Sheet 1** tab). Select and copy the data **A22: B31** as shown.

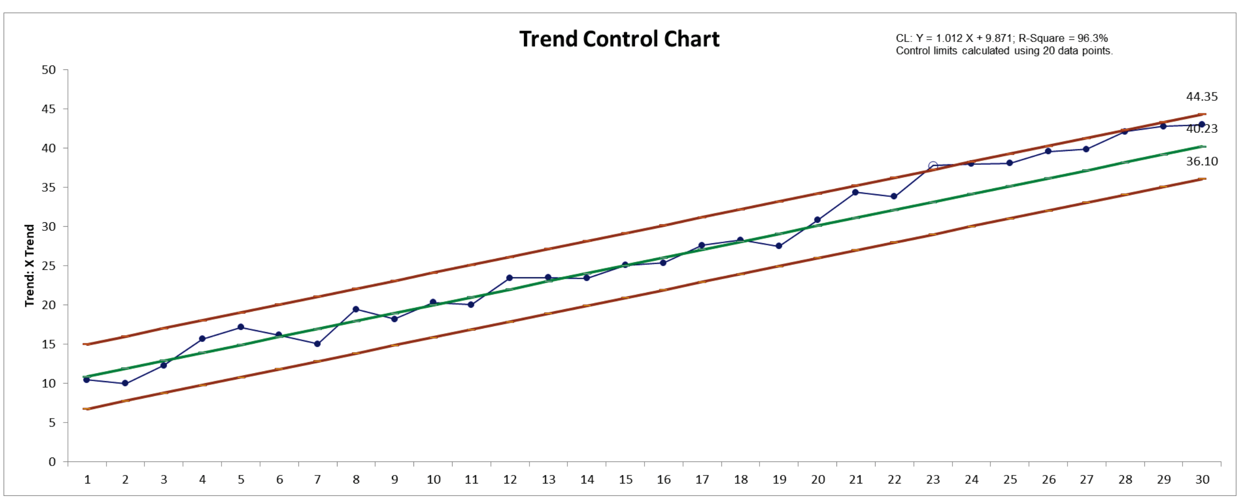
	A	B
1	Sample	X Trend
2	1	10.45
3	2	9.99
4	3	12.29
5	4	15.66
6	5	17.16
7	6	16.18
8	7	15.04
9	8	19.46
10	9	18.2
11	10	20.34
12	11	20.03
13	12	23.47
14	13	23.51
15	14	23.4
16	15	25.08
17	16	25.37
18	17	27.62
19	18	28.31
20	19	27.52
21	20	30.84
22	21	34.4
23	22	33.83
24	23	37.79
25	24	38
26	25	38.1
27	26	39.58
28	27	39.88
29	28	42.12
30	29	42.81
31	30	43.02

7. Switch to the Trend Chart template and Paste Values to cell **A23** as shown.



	A	B
1	Sample	X Trend
2	1	10.45
3	2	9.99
4	3	12.29
5	4	15.66
6	5	17.16
7	6	16.18
8	7	15.04
9	8	19.46
10	9	18.2
11	10	20.34
12	11	20.03
13	12	23.47
14	13	23.51
15	14	23.4
16	15	25.08
17	16	25.37
18	17	27.62
19	18	28.31
20	19	27.52
21	20	30.84
22	21	34.4
23	22	33.83
24	23	37.79
25	24	38
26	25	38.1
27	26	39.58
28	27	39.88
29	28	42.12
30	29	42.81
31	30	43.02
32		

8. Click the **Add Data** button to add the new data to the Trend Control Chart:



9. This confirms that the process is now "out-of-control", so corrective action to find and remove the assignable cause is needed.

Template Notes and Reference:

1. This Trend (aka Toolwear) Control Chart template should be used with continuous data. The data must be in chronological time-sequence order and has a consistent positive or negative linear trend which is inherent to the process.
2. You can replace the **X-Axis Label** and **Data** column headings with any headings that you wish.
3. Enter your data in the **Data** column.
4. Enter labels in **X-Axis Label** column. Labels can be Date, Time, Name, or other text information. These labels are optional and will appear on the horizontal X-Axis of the Trend Control Chart.
5. Click the **Trend Control Chart** button to create a Trend Control Chart.
6. After the control chart has been created and additional new data entered into the **Data** column, click the **Add Data** button to add the data to the existing chart. Control limits will be calculated using the original chart slope, intercept and stdev.
7. **Add Data** should only be used if there are at least 20 observations in the original chart.
8. Center line (CL) is the linear regression equation. Stdev. is estimated using MR-bar/1.128 of regression residuals.
9. Note that the regression model estimation error is not included in the calculation of the control limits. R-Square should be at least 50%, preferably greater than 80%.
10. Reference: Trend control limit formulas are given in Provost L, Murray S., *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco: Jossey-Bass, 2011, p. 254.

Part J – Control Chart Templates: Average Run Length (ARL) Calculators

Average Run Length (ARL) - Introduction

Average Run Length (ARL) characteristics are very useful to compare the performance of control charts and determine optimal parameter settings for EWMA and CUSUM time weighted control charts.

The ARL value for a shift in mean = 0 sigma is the "in-control" average run length and is denoted as ARL_0. ARL_0 is $1/\alpha$, where α is the type I false alarm probability, so this should be as large as possible minimizing the likelihood that an out-of-control signal is a false alarm. In a Shewhart Individuals Control Chart, $ARL_0 = 1/\alpha = 1/(0.00135*2) = 370.4$. On average, we will see a false alarm once every 370 observations. Note that this is a mean of a geometric distribution, so in practice the actual ARL_0 will vary widely with the standard deviation approximately equal to the mean value.

When we have a sustained shift in mean > 0, the ARL value is the "out-of-control" run length and is denoted as ARL_1. $ARL_1 = 1/(1-\beta)$, where β is the type II miss probability and $(1-\beta)$ is the power to detect. This should be as small as possible so that a shift in process mean is quickly detected.

The calculations for ARL are quite complex, involving either Exact, Markov Chain approximation or Monte Carlo simulation to solve. The SigmaXL ARL Templates take care of these calculations and are easy to use. If Monte Carlo simulation is used, additional Run Length standard deviation and percentile statistics are reported. Monte Carlo results will take some time and vary slightly every time they are run. With 1000 (1e3) replications it will be fast, approx. 10 seconds, but will have an ARL_0 error of approximately +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL_0 error of +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL_0 error = +/- 1%. More than 1e5 replications will be very time consuming and may run into memory limitations, so is not recommended.

When using EWMA or CUSUM charts, we typically set the parameters to minimize the ARL_1 to give rapid detection for a small shift in mean of 1 sigma. Shewhart charts are typically used when trying to rapidly detect a large shift in mean of >= 3 sigma. Tests for special causes may be used with Shewhart to improve the small shift performance, but they give poor ARL_0 performance resulting in frequent false alarms.

Subgroups improve the small shift performance of a Shewhart chart without impacting the ARL_0 rate and, if possible, should be used. The ARL for subgroup averages is adjusted by using the sigma of averages, sigma/$\sqrt{n}$. For example, with a subgroup size of 4, the ARL_1 values at shift in mean = 1 will match the ARL performance of an Individuals chart with shift in mean = 2 sigma.

Note that subgroups for CUSUM and EWMA are not available in SigmaXL.

The problem of robustness to nonnormality can also be considered by using the Pearson family to specify any value of skewness and kurtosis and estimate the ARLs. For further details, see Appendix: Pearson Family of Distributions.

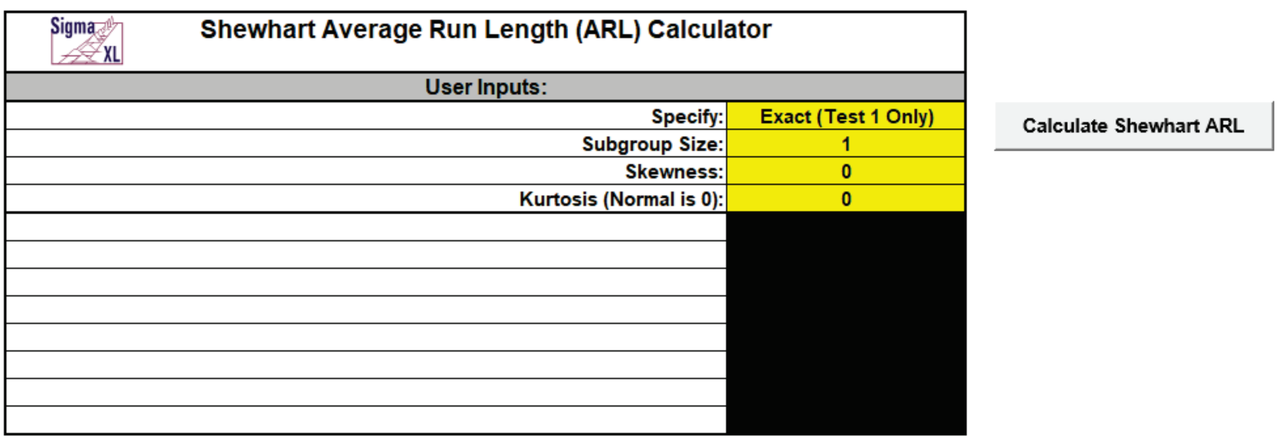
The average run length calculators are for two-sided charts with zero-state, i.e., the shift is assumed to occur at the start. The parameters (mean, standard deviation and proportion) are also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.

Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for the ARL templates to function. Do not add or delete rows or columns in these templates.

Shewhart ARL

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > Shewhart ARL**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Average Run Length (ARL) Calculators > Shewhart ARL**.

2. The default template settings are **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 1, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0.



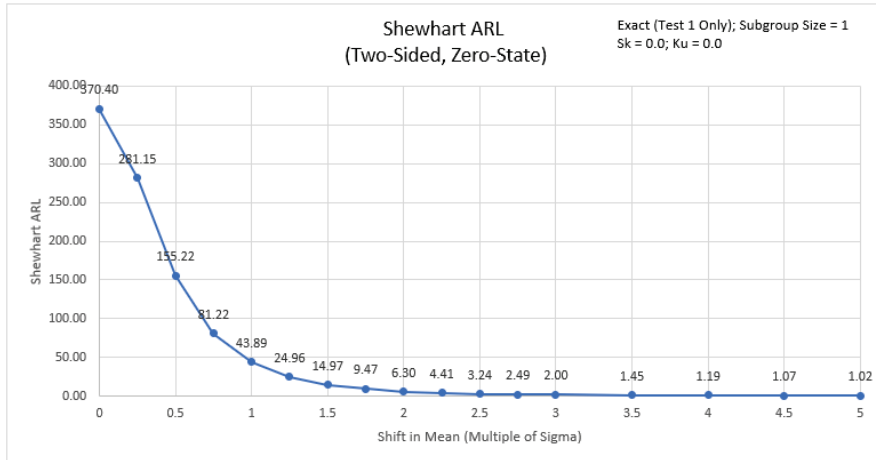
Shewhart Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Exact (Test 1 Only)
Subgroup Size:	1
Skewness:	0
Kurtosis (Normal is 0):	0

Notes: Specify Exact (Test 1 Only) or Monte Carlo using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden. Exact uses the normal or Pearson cumulative distribution function and is fast. Monte Carlo simulation allows you to assess the ARL performance of all 8 Tests for Special Causes. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied. Monte Carlo simulation also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view). Both methods allow you to assess robustness to nonnormality.

All ARL calculations for Shewhart use a standardized in-control mean=0 and sigma=1.

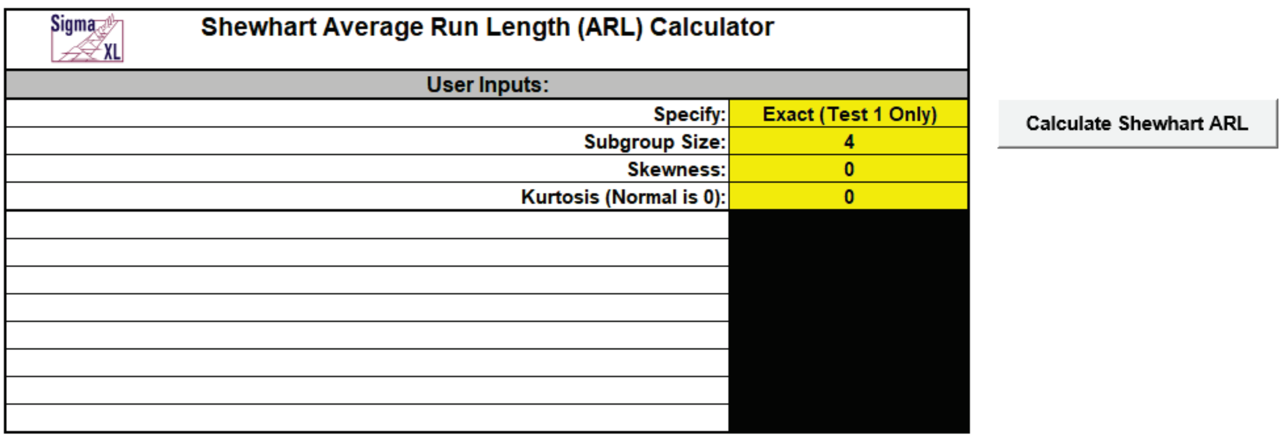
3. Click the **Calculate Shewhart ARL** button to reproduce the ARL table and chart.

Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	370.40
0.25	281.15
0.5	155.22
0.75	81.22
1	43.89
1.25	24.96
1.5	14.97
1.75	9.47
2	6.30
2.25	4.41
2.5	3.24
2.75	2.49
3	2.00
3.5	1.45
4	1.19
4.5	1.07
5	1.02



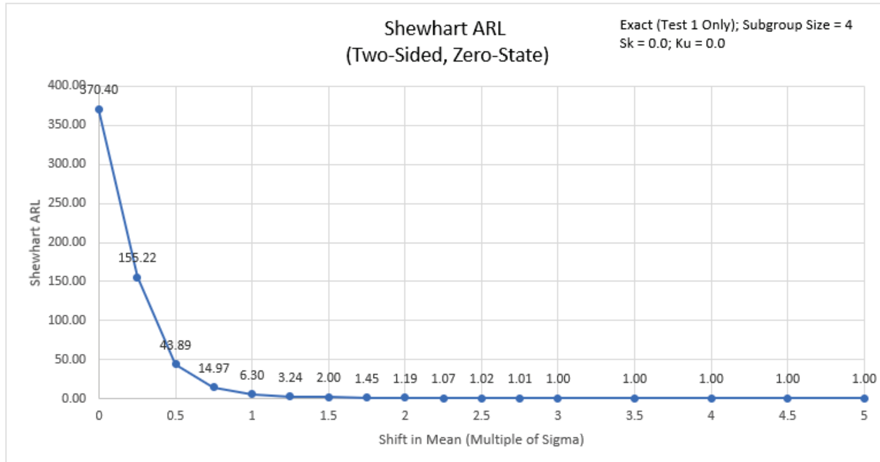
As discussed in the introduction, the ARL_0 (in-control ARL with 0 shift in mean) for the Shewhart chart is 370.4. The ARL_1 for a small 1 sigma shift in mean is 43.89, so is slow to detect. On the other hand, a large 3 sigma shift in mean has an ARL = 2.0, so is detected rapidly.

4. We will now assess ARL for a Shewhart X-bar chart. Select **Specify** = *Exact (Test 1 Only)*. Enter **Subgroup Size** = 4, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0.



5. Click the **Calculate Shewhart ARL** button to produce the ARL table and chart for these settings:

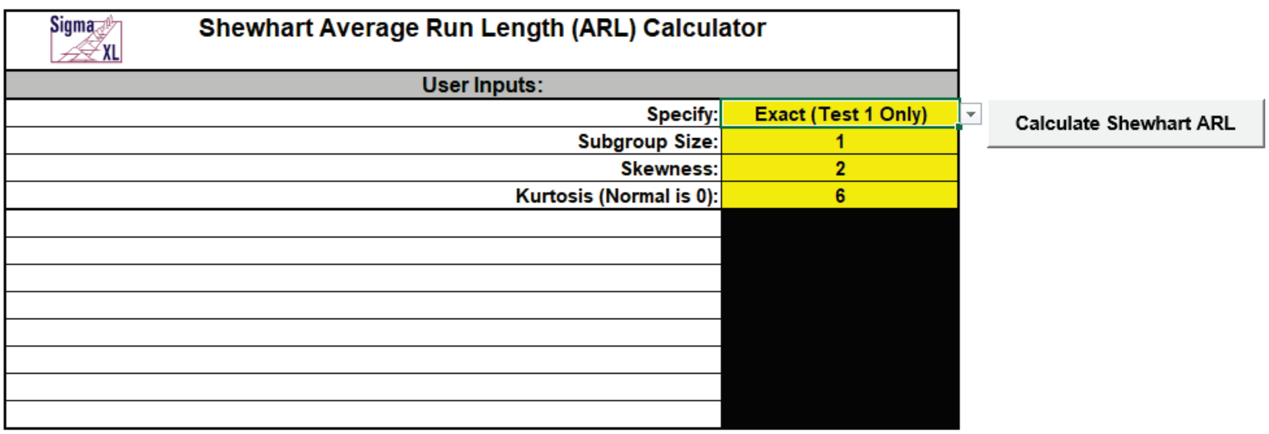
Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	370.40
0.25	155.22
0.5	43.89
0.75	14.97
1	6.30
1.25	3.24
1.5	2.00
1.75	1.45
2	1.19
2.25	1.07
2.5	1.02
2.75	1.01
3	1.00
3.5	1.00
4	1.00
4.5	1.00
5	1.00



The ARL_0 for the Shewhart x-bar chart is the same as the Individuals chart, 370.4. The ARL_1 for a small 1 sigma shift in mean is 6.3, so is much more rapid to detect than the Individuals ARL_1 of 43.89, so if possible, subgrouping should always be used.

Note: The ARL for subgroup averages is adjusted by using the sigma of averages, sigma/$\sqrt{n}$. For example, with a subgroup size of 4, the ARL_1 values at shift in mean = 1 will match the ARL performance of an Individuals chart with shift in mean = 2 sigma.

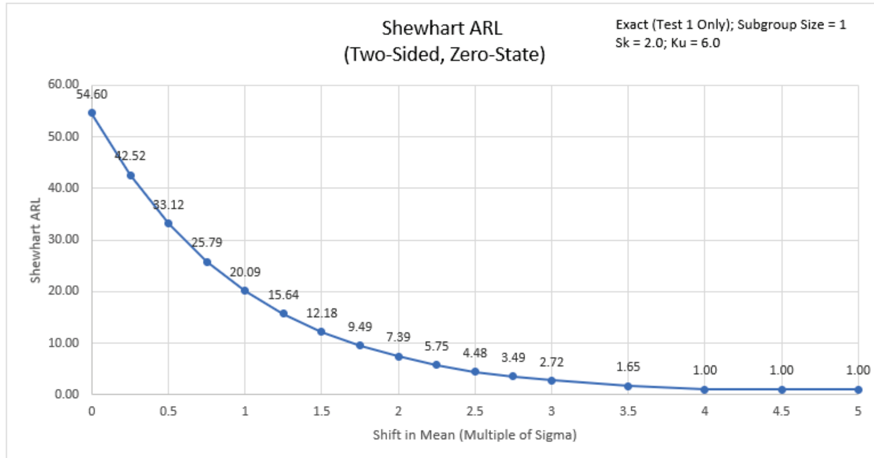
6. Now we will assess robustness to nonnormality. Enter **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 1, **Skewness** = 2, **Kurtosis (Normal is 0)** = 6.



Note: We are specifying a severe degree of Skewness (Skewness = 0.5 is mild, 1 is moderate, 2 is severe, and > 2 is very severe). The Pearson family is used to create a distribution that matches the specified Skewness and Kurtosis. Skewness = 2 and Kurtosis = 6 corresponds to an Exponential distribution or Gamma distribution with *Shape* = 1 and *Scale* = 1 (for Gamma, Skewness = $2/\sqrt{Shape}$ and Kurtosis = $6/Shape$).

7. Click the **Calculate Shewhart ARL** button to produce the ARL table and chart for these settings:

Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	54.60
0.25	42.52
0.5	33.12
0.75	25.79
1	20.09
1.25	15.64
1.5	12.18
1.75	9.49
2	7.39
2.25	5.75
2.5	4.48
2.75	3.49
3	2.72
3.5	1.65
4	1.00
4.5	1.00
5	1.00



ARL$_0$ with these settings is 54.6. This is a very poor performance with a 6.8 x increase (370.4/54.6) in false alarms compared to normal data. The Shewhart Individuals chart is not robust to severe skewness. A Box-Cox Transformation or other Individuals Nonnormal chart should be used (see **SigmaXL > Control Charts > Nonnormal > Individuals Nonnormal**).

Note: ARL$_0$ = 54.6 matches the result given in Montgomery [2], Table 9.12 for Gam(1,1).

8. Next, we will assess robustness to nonnormality for a Shewhart X-bar chart. Enter **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 4, **Skewness** = 2, **Kurtosis (Normal is 0)** = 6.

Shewhart Average Run Length (ARL) Calculator

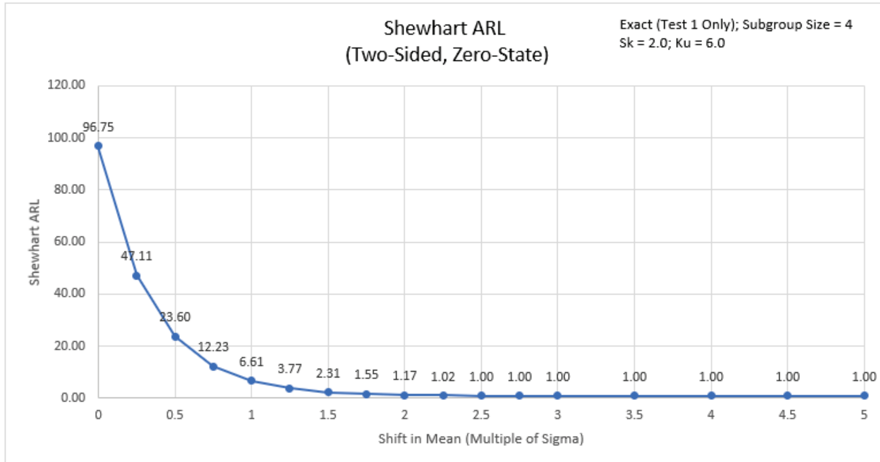
User Inputs:	
Specify:	Exact (Test 1 Only)
Subgroup Size:	4
Skewness:	2
Kurtosis (Normal is 0):	6

Calculate Shewhart ARL

Note: Skewness of averages = $Skewness/\sqrt{n}$. Kurtosis of averages = $Kurtosis/n$. For a subgroup size of 4, the skewness of averages is 1, so is reduced from severe to moderate. Kurtosis of averages is 1.5 (corresponding to a Gamma distribution with *Shape* = 4).

9. Click the **Calculate Shewhart ARL** button to produce the ARL table and chart for these settings:

Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	96.75
0.25	47.11
0.5	23.60
0.75	12.23
1	6.61
1.25	3.77
1.5	2.31
1.75	1.55
2	1.17
2.25	1.02
2.5	1.00
2.75	1.00
3	1.00
3.5	1.00
4	1.00
4.5	1.00
5	1.00



ARL_0 with these settings is 96.75. This is an improvement over the Individuals 54.6, but is still a 3.8 x increase (370.4/96.75) in false alarms compared to normal data.

Note: ARL_0 = 96.75 matches the results given in Schilling & Nelson [3] (Table 1, Gamma, shape = 1, n=4), =1/.01034. In Table 2, they point out that a subgroup size of 166 would be required to achieve robustness for this severe skewness.

10. Now we will use Monte Carlo simulation to obtain approximate Run Length standard deviation and percentiles for an Individuals Shewhart chart with normal data. Enter **Specify** = *Monte Carlo*, **Subgroup Size** = 1, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0, **Number of Replications** = 1e4, and **Test 2** to **Test 8** = *N/A*.

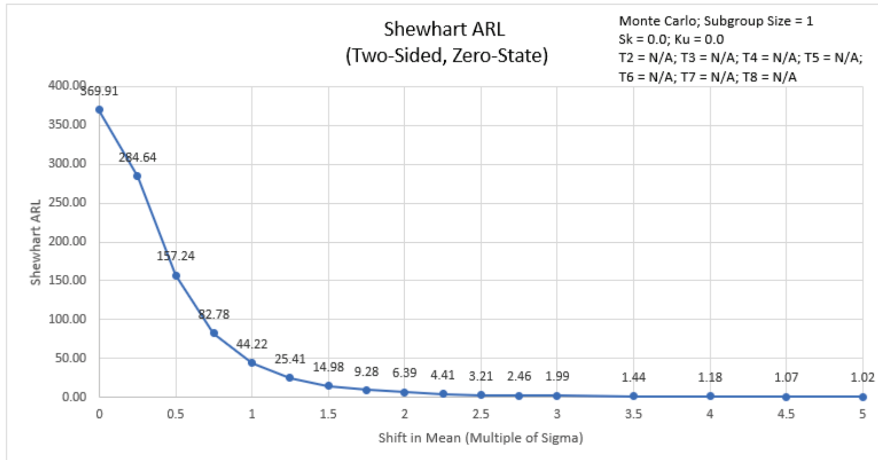
Shewhart Average Run Length (ARL) Calculator

User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	1
Skewness:	0
Kurtosis (Normal is 0):	0
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	N/A
Test 3 - points in a row all increasing or decreasing:	N/A
Test 4 - points in a row alternating up and down:	N/A
Test 5 - points more than 2 standard deviations from CL (same side):	N/A
Test 6 - points more than 1 standard deviation from CL (same side):	N/A
Test 7 - points in a row within 1 standard deviation from CL (either side):	N/A
Test 8 - points in a row more than 1 standard deviation from CL (either side):	N/A

Calculate Shewhart ARL

11. Click the **Calculate Shewhart ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table (scroll right to view). Monte Carlo simulation with 10,000 (1e4) replications will take about a minute to run.

Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	369.91
0.25	284.64
0.5	157.24
0.75	82.78
1	44.22
1.25	25.41
1.5	14.98
1.75	9.28
2	6.39
2.25	4.41
2.5	3.21
2.75	2.46
3	1.99
3.5	1.44
4	1.18
4.5	1.07
5	1.02



	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	370.23	5	19	37	105	257	515	863	1106	1738
0.25	284.79	4	16	30	82	199	391	664	851	1274
0.5	157.51	2	9	17	45	111	216	357	469	737
0.75	83.00	1	4	9	24	58	113	190	247	393
1	43.44	1	3	5	13	31	61	101	132	202
1.25	24.89	1	2	3	8	18	35	58	75	114
1.5	14.36	1	1	2	5	11	20	34	44	67
1.75	8.86	1	1	1	3	7	13	20	27	41
2	5.70	1	1	1	2	5	9	14	18	26
2.25	3.97	1	1	1	2	3	6	10	12	19
2.5	2.67	1	1	1	1	2	4	7	8	13
2.75	1.89	1	1	1	1	2	3	5	6	9
3	1.40	1	1	1	1	1	2	4	5	7
3.5	0.79	1	1	1	1	1	2	2	3	4
4	0.47	1	1	1	1	1	1	2	2	3
4.5	0.29	1	1	1	1	1	1	1	2	2
5	0.16	1	1	1	1	1	1	1	1	2

The additional run length statistics show the large variation of run length values. The median MRL_0 = 257 (in-control median run length with 0 shift in process mean). The run length percentiles approximately match those given in Chakraborti [4] (Table 1, Standards Known, Shift 0.0).

Note: The results will vary slightly since this is Monte Carlo simulation.

12. Now we will use Monte Carlo simulation to assess the Shewhart Individuals chart with all 8 Tests for Special Causes applied. Enter **Specify** = *Monte Carlo*, **Subgroup Size** = 1, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0, **Number of Replications** = 1e4, **Test 2** = 9, **Test 3** = 6, **Test 4** = 14, **Test 5** = 2 out of 3, **Test 6** = 4 out of 5, **Test 7** = 15 and **Test 8** = 8.

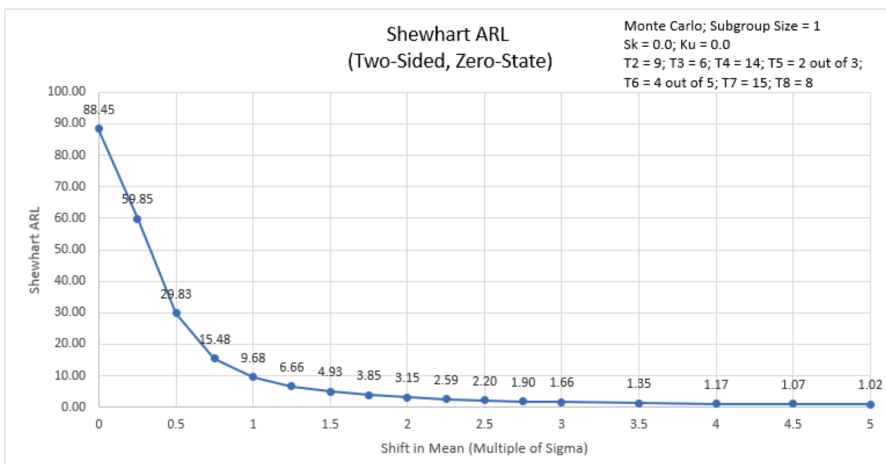
Shewhart Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	1
Skewness:	0
Kurtosis (Normal is 0):	0
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	9
Test 3 - points in a row all increasing or decreasing:	6
Test 4 - points in a row alternating up and down:	14
Test 5 - points more than 2 standard deviations from CL (same side):	2 out of 3
Test 6 - points more than 1 standard deviation from CL (same side):	4 out of 5
Test 7 - points in a row within 1 standard deviation from CL (either side):	15
Test 8 - points in a row more than 1 standard deviation from CL (either side):	8

Calculate Shewhart ARL

Note: These are the test settings used as defaults in **SigmaXL > Control Charts > 'Tests for Special Causes' Defaults**. Test 1 is always applied.

13. Click the **Calculate Shewhart ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table.

Shift in Mean (Multiple of Sigma)	Shewhart ARL (Two-Sided, Zero-State)
0	88.45
0.25	59.85
0.5	29.83
0.75	15.48
1	9.68
1.25	6.66
1.5	4.93
1.75	3.85
2	3.15
2.25	2.59
2.5	2.20
2.75	1.90
3	1.66
3.5	1.35
4	1.17
4.5	1.07
5	1.02



		Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)								
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	83.72	3	8	14	29	63	120	197	254	393
0.25	55.52	3	7	10	21	43	82	132	168	260
0.5	26.21	2	5	6	11	22	40	64	82	124
0.75	12.10	1	3	4	7	12	20	31	39	58
1	6.65	1	2	3	5	8	12	18	23	33
1.25	4.21	1	2	2	4	5	9	12	15	21
1.5	2.85	1	1	2	3	4	6	9	10	14
1.75	2.09	1	1	1	2	4	5	6	8	10
2	1.67	1	1	1	2	3	4	5	6	9
2.25	1.35	1	1	1	2	2	3	4	5	7
2.5	1.12	1	1	1	1	2	3	4	4	5
2.75	0.94	1	1	1	1	2	2	3	4	5
3	0.80	1	1	1	1	1	2	3	3	4
3.5	0.56	1	1	1	1	1	2	2	2	3
4	0.40	1	1	1	1	1	1	2	2	2
4.5	0.26	1	1	1	1	1	1	1	2	2
5	0.15	1	1	1	1	1	1	1	1	2

ARL_0 with all 8 tests for special causes is approx. 88.5. This is a poor performance with a 4.2 x increase (370.4/88.5) in false alarms compared to Test 1 only. MRL_0 is approx. 63. On the other hand, ARL_1 for a small 1 sigma shift in mean is approx. 9.7, so is much faster to detect than the Exact Test 1 only ARL_1 of 43.89.

If small shifts are to be detected quickly and subgrouping is not possible, then an EWMA or CUSUM chart is recommended.

Template Notes:

1. Specify **Exact (Test 1 Only)** or **Monte Carlo** using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden.
2. Exact uses the cumulative distribution function and is fast. Monte Carlo simulation allows you to assess the ARL performance of all 8 Tests for Special Causes and also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view). Both methods allow you to assess robustness to nonnormality.
3. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied.
4. The Pearson Family of distributions is used to match the specified Skewness and Kurtosis.
5. Enter the **Subgroup Size.** Subgroup size = 1 denotes a Shewhart Individuals chart. Subgroup size > 1 is an X-Bar chart.
6. Enter **Skewness**. Skewness = 0 is symmetric.
7. Enter **Kurtosis (Normal is 0)**. Also known as Excess Kurtosis, it must be >= Skewness^2 - 1.48. This is required to keep the distribution unimodal. If Skewness=0 and Kurtosis = 0, the distribution is normal.
8. If applicable, enter **Number of Replications**. 1000 (1e3) replications will be fast, approx. 10 seconds, but will have an ARL0 error approx. = +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL0 error = +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL0 error = +/- 1%.
9. If applicable, select values for Tests 2 to 8 using the drop-down list. "N/A" indicates that the test is not applied. Tests 2, 3 and 7 provide options that match those provided in SigmaXL's '**Tests for Special Causes' Defaults** dialog.
10. Click the **Calculate Shewhart ARL** button to produce the ARL table and chart. If Monte Carlo was selected, the table of Run Length Standard Deviation and Percentiles will also be produced.
11. The Shewhart ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The mean and standard deviation are also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.
12. Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for this template to function. Do not add or delete rows or columns in this template.

REFERENCES:

[1] Champ, C.W. and Woodall, W.H. (1987), "Exact results for Shewhart control charts with supplementary runs rules", *Technometrics* 29, 393-399.

[2] Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley.

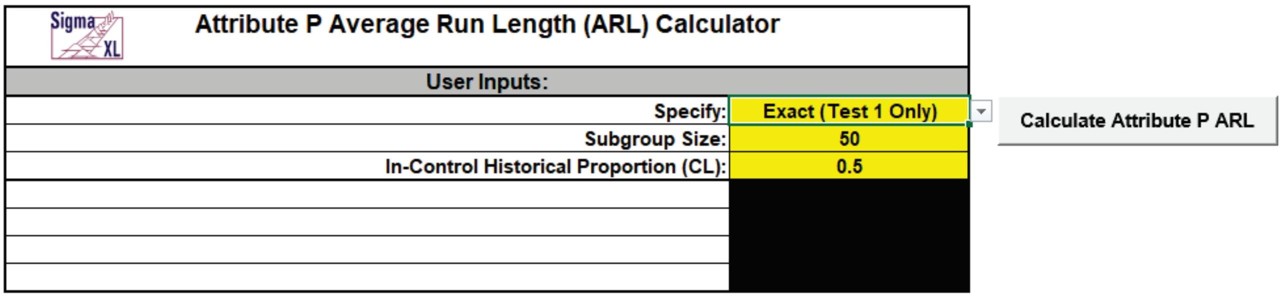
[3] Schilling, E. G., and P. R. Nelson (1976), “The Effect of Nonnormality on the Control Limits of $\bar{X}$ Charts,” *Journal of Quality Technology*, Vol. 8(4), pp. 183–188.

[4] Chakraborti, S. (2007), “Run Length Distribution and Percentiles: The Shewhart Chart with Unknown Parameters”, *Quality Engineering* 19, 119–127.

Attribute P ARL

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > Attribute P ARL**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Average Run Length (ARL) Calculators > Attribute P ARL**.

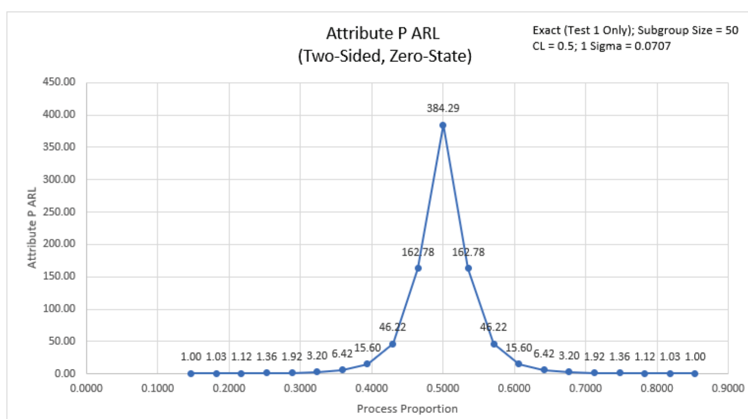
2. The default template settings are **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 50, **In-Control Historical Proportion (CL)** = 0.5.



Note: Specify Exact (Test 1 Only) or Monte Carlo using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden. Exact uses the binomial cumulative distribution function. Monte Carlo simulation uses binomial random data with specified proportion and allows you to assess the ARL performance of all 4 Tests for Special Causes. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied. Monte Carlo simulation also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).

3. Click the **Calculate Attribute P ARL** button to reproduce the ARL table and chart.

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0707	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-5.0	0.1464	1.00
-4.5	0.1818	1.03
-4.0	0.2172	1.12
-3.5	0.2525	1.36
-3.0	0.2879	1.92
-2.5	0.3232	3.20
-2.0	0.3586	6.42
-1.5	0.3939	15.60
-1.0	0.4293	46.22
-0.5	0.4646	162.78
0.0	0.5000	384.29
0.5	0.5354	162.78
1.0	0.5707	46.22
1.5	0.6061	15.60
2.0	0.6414	6.42
2.5	0.6768	3.20
3.0	0.7121	1.92
3.5	0.7475	1.36
4.0	0.7828	1.12
4.5	0.8182	1.03
5.0	0.8536	1.00



The ARL_0 (in-control ARL with 0 sigma shift in proportion) for the Attribute P chart is 384.29. The ARL_1 for a small 1 sigma shift is 46.22, so is slow to detect. On the other hand, a large 3 sigma shift in proportion has an ARL = 1.92, so is detected rapidly.

Note: The Process Proportion values are displayed on the ARL chart, but the shift in proportion as a multiple of sigma is also included in the table.

4. Now we will use Monte Carlo simulation to obtain approximate Run Length standard deviation and percentiles for the Attribute P chart. Select **Specify** = *Monte Carlo*. Enter **Subgroup Size** = 50, **In-Control Historical Proportion (CL)** = 0.5, **Number of Replications** = 1e4, **Test 2** = *N/A*, **Test 3** = *N/A*, **Test 4** = *N/A*.

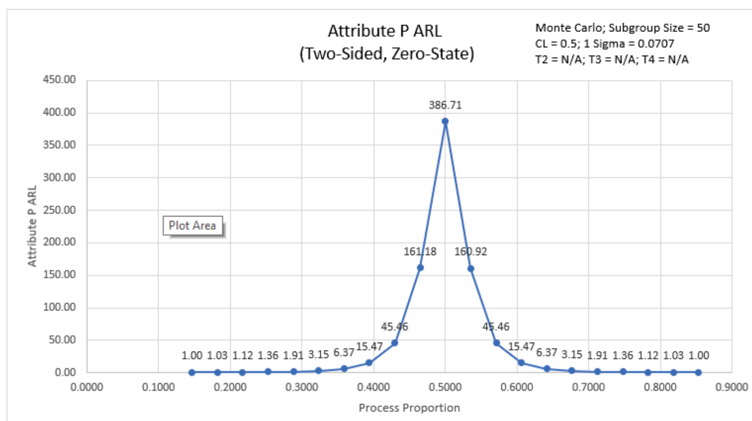
Attribute P Average Run Length (ARL) Calculator

User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	50
In-Control Historical Proportion (CL):	0.5
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	N/A
Test 3 - points in a row all increasing or decreasing:	N/A
Test 4 - points in a row alternating up and down:	N/A

Calculate Attribute P ARL

5. Click the **Calculate Attribute P ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table (scroll right to view). Monte Carlo simulation with 10,000 (1e4) replications will take about a minute to run.

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0707	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-5.0	0.1464	1.00
-4.5	0.1818	1.03
-4.0	0.2172	1.12
-3.5	0.2525	1.36
-3.0	0.2879	1.91
-2.5	0.3232	3.15
-2.0	0.3586	6.37
-1.5	0.3939	15.47
-1.0	0.4293	45.46
-0.5	0.4646	161.18
0.0	0.5000	386.71
0.5	0.5354	160.92
1.0	0.5707	45.46
1.5	0.6061	15.47
2.0	0.6414	6.37
2.5	0.6768	3.15
3.0	0.7121	1.91
3.5	0.7475	1.36
4.0	0.7828	1.12
4.5	0.8182	1.03
5.0	0.8536	1.00



Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0707	Process Proportion	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
		Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
-5.0	0.1464	0.07	1	1	1	1	1	1	1	1	1
-4.5	0.1818	0.18	1	1	1	1	1	1	1	1	2
-4.0	0.2172	0.36	1	1	1	1	1	1	2	2	3
-3.5	0.2525	0.69	1	1	1	1	1	2	2	3	4
-3.0	0.2879	1.32	1	1	1	1	1	2	4	5	7
-2.5	0.3232	2.61	1	1	1	1	2	4	6	8	12
-2.0	0.3586	5.86	1	1	1	2	5	9	14	18	28
-1.5	0.3939	14.85	1	1	2	5	11	21	35	45	70
-1.0	0.4293	45.38	1	3	5	13	31	63	105	138	205
-0.5	0.4646	161.29	2	8	17	46	113	224	372	480	746
0.0	0.5000	384.33	4	19	41	113	267	540	899	1165	1744
0.5	0.5354	162.28	2	8	17	46	111	222	368	480	760
1.0	0.5707	45.38	1	3	5	13	31	63	105	138	205
1.5	0.6061	14.85	1	1	2	5	11	21	35	45	70
2.0	0.6414	5.86	1	1	1	2	5	9	14	18	28
2.5	0.6768	2.61	1	1	1	1	2	4	6	8	12
3.0	0.7121	1.32	1	1	1	1	1	2	4	5	7
3.5	0.7475	0.69	1	1	1	1	1	2	2	3	4
4.0	0.7828	0.36	1	1	1	1	1	1	2	2	3
4.5	0.8182	0.18	1	1	1	1	1	1	1	1	2
5.0	0.8536	0.07	1	1	1	1	1	1	1	1	1

The additional run length statistics show the large variation of run length values. The MRL_0 = 267 (in-control median run length with 0 sigma shift in process proportion).

Note: The results will vary slightly since this is Monte Carlo simulation.

6. Now we will use Monte Carlo simulation to assess the Attribute P chart with all 4 Tests for Special Causes applied. Enter **Specify** = *Monte Carlo*, **Subgroup Size** = 50, **In-Control Historical Proportion (CL)** = 0.5, **Number of Replications** = 1e4, **Test 2** = 9, **Test 3** = 6, and **Test 4** = 14.

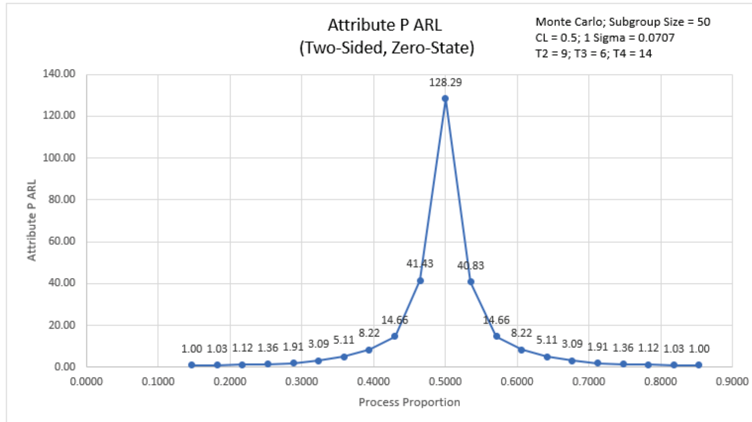
Attribute P Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	50
In-Control Historical Proportion (CL):	0.5
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	9
Test 3 - points in a row all increasing or decreasing:	6
Test 4 - points in a row alternating up and down:	14

Calculate Attribute P ARL

Note: These are the test settings used as defaults in **SigmaXL > Control Charts > 'Tests for Special Causes' Defaults**. Test 1 is always applied.

7. Click the **Calculate Attribute P ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table:

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0707	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-5.0	0.1464	1.00
-4.5	0.1818	1.03
-4.0	0.2172	1.12
-3.5	0.2525	1.36
-3.0	0.2879	1.91
-2.5	0.3232	3.09
-2.0	0.3586	5.11
-1.5	0.3939	8.22
-1.0	0.4293	14.66
-0.5	0.4646	41.43
0.0	0.5000	128.29
0.5	0.5354	40.83
1.0	0.5707	14.66
1.5	0.6061	8.22
2.0	0.6414	5.11
2.5	0.6768	3.09
3.0	0.7121	1.91
3.5	0.7475	1.36
4.0	0.7828	1.12
4.5	0.8182	1.03
5.0	0.8536	1.00



		Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0707	Process Proportion	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
-5.0	0.1464	0.06	1	1	1	1	1	1	1	1	1
-4.5	0.1818	0.18	1	1	1	1	1	1	1	1	2
-4.0	0.2172	0.37	1	1	1	1	1	1	2	2	3
-3.5	0.2525	0.71	1	1	1	1	1	2	2	3	4
-3.0	0.2879	1.29	1	1	1	1	1	2	4	4	7
-2.5	0.3232	2.27	1	1	1	1	2	4	7	8	9
-2.0	0.3586	3.25	1	1	1	2	5	9	9	9	13
-1.5	0.3939	4.62	1	1	2	5	9	9	14	17	23
-1.0	0.4293	10.00	1	3	5	9	11	18	27	34	50
-0.5	0.4646	35.49	2	9	9	16	31	56	88	112	168
0.0	0.5000	122.77	4	12	19	41	91	176	287	376	568
0.5	0.5354	34.95	2	8	9	16	30	55	87	111	164
1.0	0.5707	10.00	1	3	5	9	11	18	27	34	50
1.5	0.6061	4.62	1	1	2	5	9	9	14	17	23
2.0	0.6414	3.25	1	1	1	2	5	9	9	9	13
2.5	0.6768	2.27	1	1	1	1	2	4	7	8	9
3.0	0.7121	1.29	1	1	1	1	1	2	4	4	7
3.5	0.7475	0.71	1	1	1	1	1	2	2	3	4
4.0	0.7828	0.37	1	1	1	1	1	1	2	2	3
4.5	0.8182	0.18	1	1	1	1	1	1	1	1	2
5.0	0.8536	0.06	1	1	1	1	1	1	1	1	1

ARL_0 with all 4 tests for special causes is approx. 128.3. This is a poor performance with a 3 x increase (384.3/128.3) in false alarms compared to Exact Test 1 only. MRL_0 is approx. 91. On the other hand, ARL_1 for a small 1 sigma shift in proportion is approx. 14.7, so is much faster to detect than the Exact Test 1 only ARL_1 of 46.22.

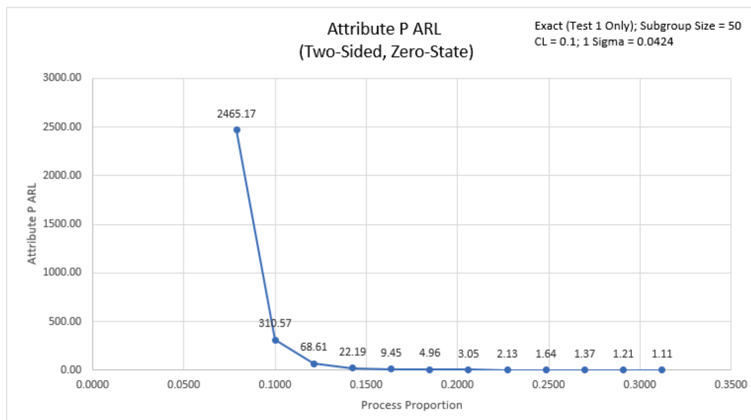
8. We will now assess ARL for a low in-control proportion value. Enter **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 50, **In-Control Historical Proportion (CL)** = 0.1.

Attribute P Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Exact (Test 1 Only)
Subgroup Size:	50
In-Control Historical Proportion (CL):	0.1

Calculate Attribute P ARL

9. Click the **Calculate Attribute P ARL** button to produce the ARL table and chart for these settings:

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0424	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-0.5	0.0788	2465.17
0.0	0.1000	310.57
0.5	0.1212	68.61
1.0	0.1424	22.19
1.5	0.1636	9.45
2.0	0.1849	4.96
2.5	0.2061	3.05
3.0	0.2273	2.13
3.5	0.2485	1.64
4.0	0.2697	1.37
4.5	0.2909	1.21
5.0	0.3121	1.11



The ARL_0 for the Attribute P chart is 310.57. The ARL_1 for a small positive 1 sigma shift in proportion is 22.19, so is slow to detect. On the other hand, a large positive 3 sigma shift in proportion has an ARL = 2.13, so is detected rapidly. The ARL for a small negative 0.5 sigma shift in proportion is 2465.7 so negative shifts cannot be detected with these settings.

Note: The P Chart LCL is 0 for this CL value.

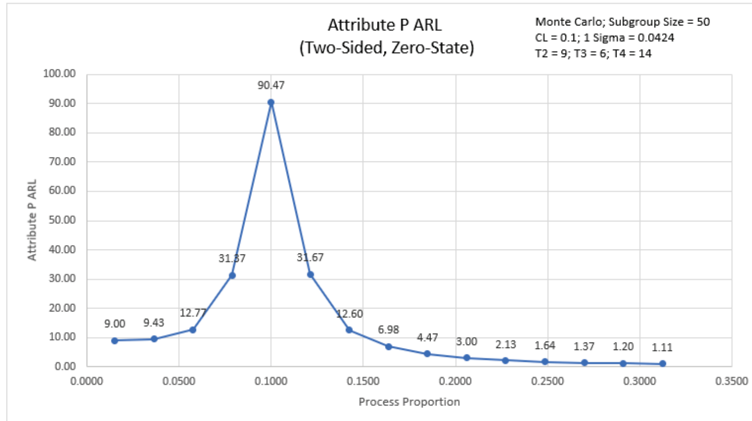
10. Now we will use Monte Carlo simulation to assess ARL for a low in-control proportion value but with all 4 Tests for Special Causes applied. Enter **Specify** = *Monte Carlo*, **Subgroup Size** = 50, **In-Control Historical Proportion (CL)** = 0.1, **Number of Replications** = 1e4, **Test 2** = 9, **Test 3** = 6, and **Test 4** = 14.

Attribute P Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	50
In-Control Historical Proportion (CL):	0.1
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	9
Test 3 - points in a row all increasing or decreasing:	6
Test 4 - points in a row alternating up and down:	14

Calculate Attribute P ARL

11. Click the **Calculate Attribute P ARL** button to produce the ARL table and chart for these settings.

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.0424	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-2.0	0.0151	9.00
-1.5	0.0364	9.43
-1.0	0.0576	12.77
-0.5	0.0788	31.37
0.0	0.1000	90.47
0.5	0.1212	31.67
1.0	0.1424	12.60
1.5	0.1636	6.98
2.0	0.1849	4.47
2.5	0.2061	3.00
3.0	0.2273	2.13
3.5	0.2485	1.64
4.0	0.2697	1.37
4.5	0.2909	1.20
5.0	0.3121	1.11



ARL_0 with all 4 tests for special causes is approx. 90.5. This is a poor performance with a 3.4 x increase (310.6/90.5) in false alarms compared to Exact Test 1 only. On the other hand, ARL_1 for a small 1 sigma shift in proportion is approx. 12.6, so is much faster to detect than the Exact Test 1 only ARL_1 of 22.19. Also, it is now possible to detect a negative one or two sigma shift in proportion.

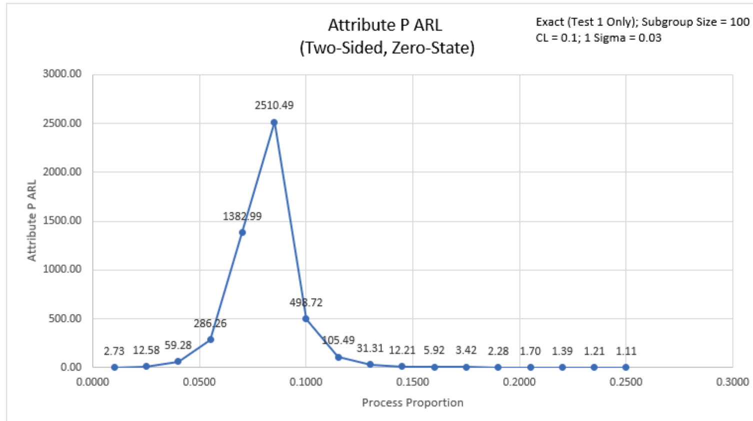
12. We will now assess ARL for a low in-control proportion value but with larger subgroup size. Enter **Specify** = *Exact (Test 1 Only)*, **Subgroup Size** = 100, **In-Control Historical Proportion (CL)** = 0.1.

Attribute P Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	**Exact (Test 1 Only)**
Subgroup Size:	**100**
In-Control Historical Proportion (CL):	**0.1**

Calculate Attribute P ARL

13. Click the **Calculate Attribute P ARL** button to produce the ARL table and chart for these settings:

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.03	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-3.0	0.0100	2.73
-2.5	0.0250	12.58
-2.0	0.0400	59.28
-1.5	0.0550	286.26
-1.0	0.0700	1382.99
-0.5	0.0850	2510.49
0.0	0.1000	498.72
0.5	0.1150	105.49
1.0	0.1300	31.31
1.5	0.1450	12.21
2.0	0.1600	5.92
2.5	0.1750	3.42
3.0	0.1900	2.28
3.5	0.2050	1.70
4.0	0.2200	1.39
4.5	0.2350	1.21
5.0	0.2500	1.11



The ARL_0 for the Attribute P chart is 498.72. The ARL_1 for a small positive 1 sigma shift in mean is 31.31, so is slow to detect. On the other hand, a large positive 3 sigma shift in proportion has an ARL = 2.28, so is detected rapidly. Small negative shifts still cannot be detected, but a large negative 3 sigma shift in proportion can be detected with ARL = 2.73.

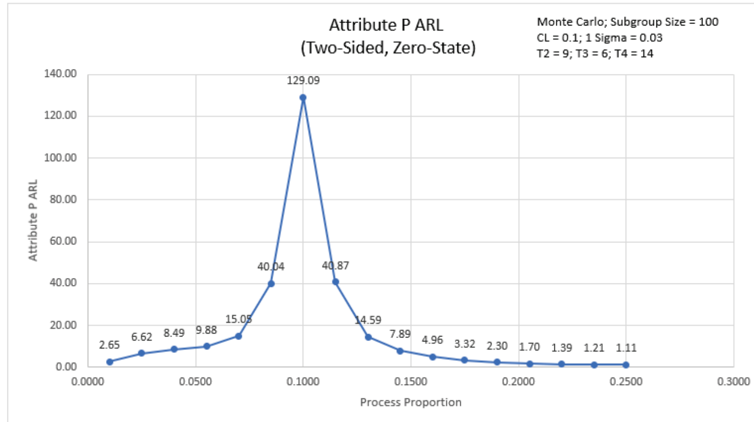
14. Now we will use Monte Carlo simulation to assess ARL for a low in-control proportion value with subgroup size = 100, but with all 4 Tests for Special Causes applied. Enter **Specify** = *Monte Carlo*, **Subgroup Size** = 100, **In-Control Historical Proportion (CL)** = 0.1. **Number of Replications** = 1e4, **Test 2** = 9, **Test 3** = 6, and **Test 4** = 14.

Attribute P Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
Subgroup Size:	100
In-Control Historical Proportion (CL):	0.1
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	9
Test 3 - points in a row all increasing or decreasing:	6
Test 4 - points in a row alternating up and down:	14

Calculate Attribute P ARL

15. Click the **Calculate Attribute P ARL** button to produce the ARL table and chart for these settings:

Shift in Proportion (Multiple of Sigma) 1 Sigma = 0.03	Process Proportion	Attribute P ARL (Two-Sided, Zero-State)
-3.0	0.0100	2.65
-2.5	0.0250	6.62
-2.0	0.0400	8.49
-1.5	0.0550	9.88
-1.0	0.0700	15.05
-0.5	0.0850	40.04
0.0	0.1000	129.09
0.5	0.1150	40.87
1.0	0.1300	14.59
1.5	0.1450	7.89
2.0	0.1600	4.96
2.5	0.1750	3.32
3.0	0.1900	2.30
3.5	0.2050	1.70
4.0	0.2200	1.39
4.5	0.2350	1.21
5.0	0.2500	1.11



ARL_0 with all 4 tests for special causes is approx. 129.1. This is a poor performance with a 3.9 x increase (498.7/129.1) in false alarms compared to Exact Test 1 only. On the other hand, ARL_1 for a small 1 sigma shift in proportion is approx. 14.6, so is much faster to detect than the Exact Test 1 only ARL_1 of 31.31. Also, it is now possible to detect a negative one, two or three sigma shift in proportion.

Template Notes:

1. Specify **Exact (Test 1 Only)** or **Monte Carlo** using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden.
2. Exact uses the binomial cumulative distribution function. Monte Carlo simulation uses binomial random data with specified proportion and allows you to assess the ARL performance of all 4 Tests for Special Causes. It also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).
3. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied.
4. Enter the **Subgroup Size.**
5. Enter the **In-Control Historical Proportion (CL)**. This is a value between 0 and 1.
6. If applicable, enter **Number of Replications**. 1000 (1e3) replications will be fast, approx. 10 seconds, but will have an ARL0 error approx. = +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL0 error = +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL0 error = +/- 1%.
7. If applicable, select values for Tests 2 to 4 using the drop-down list. "N/A" indicates that the test is not applied. Tests 2 and 3 provide options that match those provided in SigmaXL's '**Tests for Special Causes' Defaults** dialog.
8. Click the **Calculate Attribute P ARL** button to produce the ARL table and chart. If Monte Carlo was selected, the table of Run Length Standard Deviation and Percentiles will also be produced.
9. The Attribute P ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The proportion is also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.
10. The ARL Chart is similar to an Operating Characteristic (OC) Curve, except that the Y axis is ARL rather than Beta probability. Note that the P chart may be ARL biased, with maximum ARL occurring above or below the CL.
11. Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for this template to function. Do not add or delete rows or columns in this template.

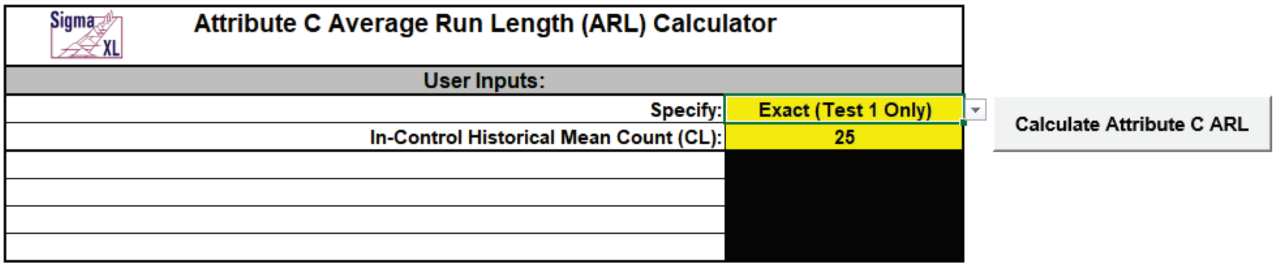
REFERENCE:

[1] Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley.

Attribute C ARL

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > Attribute C ARL**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Average Run Length (ARL) Calculators > Attribute C ARL**.

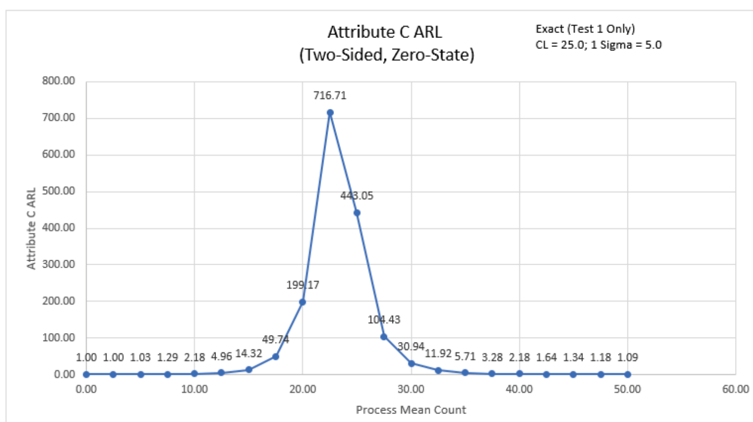
2. The default template settings are **Specify** = *Exact (Test 1 Only)*, **In-Control Historical Mean Count (CL)** = 25.



Note: Specify Exact (Test 1 Only) or Monte Carlo using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden. Exact uses the Poisson cumulative distribution function. Monte Carlo simulation uses Poisson random data with specified proportion and allows you to assess the ARL performance of all 4 Tests for Special Causes. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied. Monte Carlo simulation also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).

3. Click the **Calculate Attribute C ARL** button to reproduce the ARL table and chart.

Shift in Mean (Multiple of Sigma) 1 Sigma = 5.0	Process Mean Count	Attribute C ARL (Two-Sided, Zero-State)
-5.0	0.00	1.00
-4.5	2.50	1.00
-4.0	5.00	1.03
-3.5	7.50	1.29
-3.0	10.00	2.18
-2.5	12.50	4.96
-2.0	15.00	14.32
-1.5	17.50	49.74
-1.0	20.00	199.17
-0.5	22.50	716.71
0.0	25.00	443.05
0.5	27.50	104.43
1.0	30.00	30.94
1.5	32.50	11.92
2.0	35.00	5.71
2.5	37.50	3.28
3.0	40.00	2.18
3.5	42.50	1.64
4.0	45.00	1.34
4.5	47.50	1.18
5.0	50.00	1.09



The ARL_0 (in-control ARL with 0 shift in mean) for the Attribute C chart is 443.05. The ARL_1 for a small positive 1 sigma shift in mean count is 30.94, so is slow to detect. On the other hand, a large 3 sigma shift in mean count has an ARL = 2.18, so is detected rapidly. Small negative shifts of one-half or one sigma cannot be detected.

Note: The Process Mean Count values are displayed on the ARL chart, but the shift in mean count as a multiple of sigma is also included in the table.

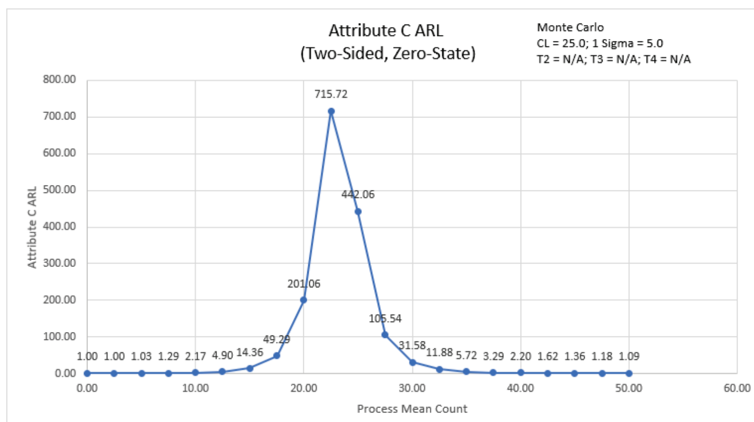
4. Now we will use Monte Carlo simulation to obtain approximate Run Length standard deviation and percentiles for the Attribute C chart. Select **Specify** = *Monte Carlo*. Enter **In-Control Historical Mean Count (CL)** = 25, **Number of Replications** = 1e4, **Test 2** = *N/A*, **Test 3** = *N/A*, **Test 4** = *N/A*.

Attribute C Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
In-Control Historical Mean Count (CL):	25
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	N/A
Test 3 - points in a row all increasing or decreasing:	N/A
Test 4 - points in a row alternating up and down:	N/A

Calculate Attribute C ARL

5. Click the **Calculate Attribute C ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table (scroll right to view). Monte Carlo simulation with 10,000 (1e4) replications will take about a minute to run.

Shift in Mean (Multiple of Sigma) 1 Sigma = 5.0	Process Mean Count	Attribute C ARL (Two-Sided, Zero-State)
-5.0	0.00	1.00
-4.5	2.50	1.00
-4.0	5.00	1.03
-3.5	7.50	1.29
-3.0	10.00	2.17
-2.5	12.50	4.90
-2.0	15.00	14.36
-1.5	17.50	49.29
-1.0	20.00	201.06
-0.5	22.50	715.72
0.0	25.00	442.06
0.5	27.50	105.54
1.0	30.00	31.58
1.5	32.50	11.88
2.0	35.00	5.72
2.5	37.50	3.29
3.0	40.00	2.20
3.5	42.50	1.62
4.0	45.00	1.36
4.5	47.50	1.18
5.0	50.00	1.09



Shift in Mean (Multiple of Sigma) 1 Sigma = 5.0	Process Mean Count	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
		Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
-5.0	0.00	0.00	1	1	1	1	1	1	1	1	1
-4.5	2.50	0.01	1	1	1	1	1	1	1	1	1
-4.0	5.00	0.19	1	1	1	1	1	1	1	1	2
-3.5	7.50	0.61	1	1	1	1	1	1	2	2	4
-3.0	10.00	1.63	1	1	1	1	2	3	4	5	8
-2.5	12.50	4.32	1	1	1	2	4	7	11	14	20
-2.0	15.00	13.98	1	1	2	4	10	20	32	43	66
-1.5	17.50	49.70	1	3	6	14	34	68	113	147	228
-1.0	20.00	200.60	3	12	23	59	140	279	460	588	942
-0.5	22.50	720.75	8	37	75	201	491	989	1659	2149	3415
0.0	25.00	438.88	5	24	47	128	308	609	1013	1320	2053
0.5	27.50	105.07	1	6	11	31	73	147	241	312	488
1.0	30.00	30.83	1	2	4	10	22	44	71	92	147
1.5	32.50	11.30	1	1	2	4	8	16	27	35	52
2.0	35.00	5.24	1	1	1	2	4	8	13	16	24
2.5	37.50	2.73	1	1	1	1	2	4	7	9	13
3.0	40.00	1.60	1	1	1	1	2	3	4	5	8
3.5	42.50	1.01	1	1	1	1	1	2	3	4	5
4.0	45.00	0.71	1	1	1	1	1	2	2	3	4
4.5	47.50	0.46	1	1	1	1	1	1	2	2	3
5.0	50.00	0.31	1	1	1	1	1	1	1	2	2

The additional run length statistics show the large variation of run length values. The MRL_0 = 308 (in-control median run length with 0 sigma shift in process mean count).

Note: The results will vary slightly since this is Monte Carlo simulation

6. Now we will use Monte Carlo simulation to assess the Attribute C chart with all 4 Tests for Special Causes applied. Enter **Specify** = *Monte Carlo*, **In-Control Historical Mean Count (CL)** = 25. **Number of Replications** = 1e4, **Test 2** = 9, **Test 3** = 6, and **Test 4** = 14.

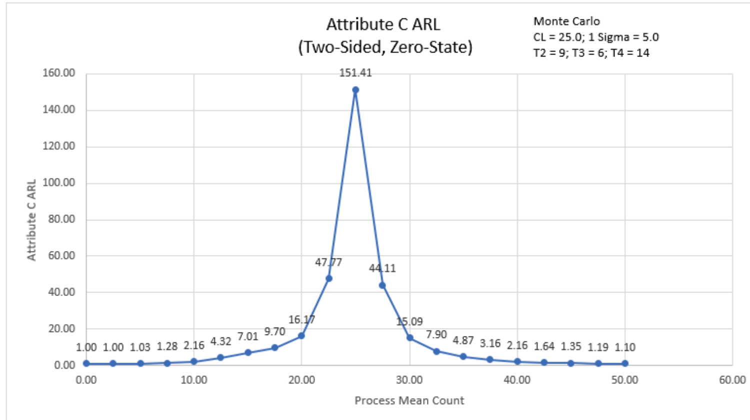
Attribute C Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Monte Carlo
In-Control Historical Mean Count (CL):	25
Number of Replications:	1.00E+04
Test 2 - points in a row on same side of CL:	9
Test 3 - points in a row all increasing or decreasing:	6
Test 4 - points in a row alternating up and down:	14

Calculate Attribute C ARL

Note: These are the test settings used as defaults in **SigmaXL > Control Charts > 'Tests for Special Causes' Defaults**. Test 1 is always applied.

7. Click the **Calculate Attribute C ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table.

Shift in Mean (Multiple of Sigma) 1 Sigma = 5.0	Process Mean Count	Attribute C ARL (Two-Sided, Zero-State)
-5.0	0.00	1.00
-4.5	2.50	1.00
-4.0	5.00	1.03
-3.5	7.50	1.28
-3.0	10.00	2.16
-2.5	12.50	4.32
-2.0	15.00	7.01
-1.5	17.50	9.70
-1.0	20.00	16.17
-0.5	22.50	47.77
0.0	25.00	151.41
0.5	27.50	44.11
1.0	30.00	15.09
1.5	32.50	7.90
2.0	35.00	4.87
2.5	37.50	3.16
3.0	40.00	2.16
3.5	42.50	1.64
4.0	45.00	1.35
4.5	47.50	1.19
5.0	50.00	1.10



		Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma) 1 Sigma = 5.0	Process Mean Count	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
-5.0	0.00	0.00	1	1	1	1	1	1	1	1	1
-4.5	2.50	0.02	1	1	1	1	1	1	1	1	1
-4.0	5.00	0.17	1	1	1	1	1	1	1	1	2
-3.5	7.50	0.61	1	1	1	1	1	1	2	2	4
-3.0	10.00	1.53	1	1	1	1	2	3	4	5	8
-2.5	12.50	2.85	1	1	1	2	4	7	9	9	9
-2.0	15.00	3.07	1	1	2	5	9	9	9	9	15
-1.5	17.50	3.91	1	3	6	9	9	9	15	17	24
-1.0	20.00	10.36	2	9	9	9	13	20	29	37	54
-0.5	22.50	41.12	7	9	10	18	35	64	101	131	194
0.0	25.00	146.52	5	13	21	48	106	206	344	448	674
0.5	27.50	39.66	1	6	9	16	32	59	95	122	186
1.0	30.00	11.18	1	2	4	9	12	20	30	37	54
1.5	32.50	5.22	1	1	2	4	8	9	15	18	25
2.0	35.00	3.42	1	1	1	2	4	8	9	10	16
2.5	37.50	2.33	1	1	1	1	2	4	7	9	9
3.0	40.00	1.56	1	1	1	1	2	3	4	5	8
3.5	42.50	1.02	1	1	1	1	1	2	3	4	5
4.0	45.00	0.68	1	1	1	1	1	2	2	3	4
4.5	47.50	0.48	1	1	1	1	1	1	2	2	3
5.0	50.00	0.32	1	1	1	1	1	1	1	2	2

ARL_0 with all 4 tests for special causes is approx. 151.4. This is a poor performance with a 2.9 x increase (443.1/151.4) in false alarms compared to Exact Test 1 only. MRL_0 is approx. 106. On the other hand, ARL_1 for a small 1 sigma shift in proportion is approx. 15.1, so is much faster to detect than the Exact Test 1 only ARL_1 of 30.94. It is now possible to detect a small negative one sigma shift in mean count.

Template Notes:

1. Specify **Exact (Test 1 Only)** or **Monte Carlo** using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden.
2. Exact uses the Poisson cumulative distribution function. Monte Carlo simulation uses Poisson random data with specified proportion and allows you to assess the ARL performance of all 4 Tests for Special Causes. It also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).
3. Test 1 - 1 point more than 3 standard deviations from the center line (CL) is always applied.
4. Enter the **Subgroup Size.**
5. Enter the **In-Control Historical Mean Count (CL)**.
6. If applicable, enter **Number of Replications**. 1000 (1e3) replications will be fast, approx. 10 seconds, but will have an ARL0 error approx. = +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL0 error = +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL0 error = +/- 1%.
7. If applicable, select values for Tests 2 to 4 using the drop-down list. "N/A" indicates that the test is not applied. Tests 2 and 3 provide options that match those provided in SigmaXL's '**Tests for Special Causes' Defaults** dialog.
8. Click the **Calculate Attribute C ARL** button to produce the ARL table and chart. If Monte Carlo was selected, the table of Run Length Standard Deviation and Percentiles will also be produced.
9. The Attribute C ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The proportion is also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.
10. The ARL Chart is similar to an Operating Characteristic (OC) Curve, except that the Y axis is ARL rather than Beta probability. Note that the C chart is ARL biased for small mean counts (<40), with maximum ARL occurring below the CL.
11. Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for this template to function. Do not add or delete rows or columns in this template.

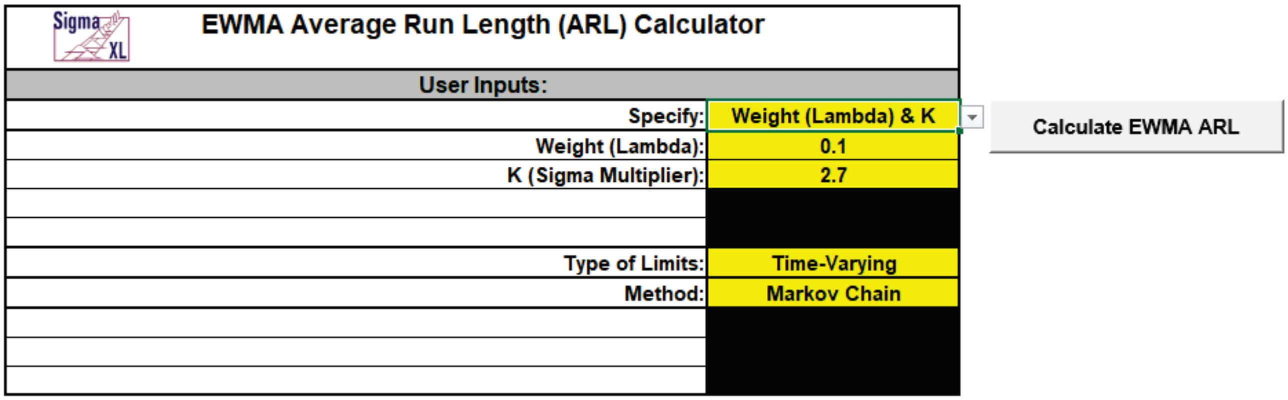
REFERENCE:

[1] Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley.

EWMA ARL

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > EWMA ARL**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Average Run Length (ARL) Calculators > EWMA ARL**.

2. The default template settings are **Specify** = *Weight (Lambda) & K*, **Weight (Lambda)** = 0.1, **K (Sigma Multiplier)** = 2.7, **Type of Limits** = *Time-Varying*, **Method** = *Markov Chain.*



Notes: Parameters to be specified will be shown in yellow highlight, otherwise they are hidden. The EWMA parameter Weight (Lambda) is a value between 0 and 1 and controls the amount of influence that previous observations have on the current EWMA statistic. A value near 1 puts almost all weight on the current observation, making it resemble a Shewhart chart. For values near 0, a small weight is applied to almost all of the past observations, so the EWMA chart performance is similar to that of a CUSUM chart. The EWMA parameter K (Sigma Multiplier) is a value typically between 2 and 4. It is also referred to as L, but SigmaXL uses K to avoid confusion with Lambda.

The EWMA control chart template has hard coded the Type of Limits as *Time-Varying*, since they improve the sensitivity of the EWMA to detect early changes in the process mean. *Fixed* is included as an option here for comparison of ARL results to published papers. Also, if the process is in control when the EWMA is started but shifts out of control after the control limits have stabilized, the more appropriate ARL for such a case would be Type of Limits = *Fixed*.

The Markov Chain approximation is fast and accurate to compute ARLs. Monte Carlo simulation allows you to assess robustness to nonnormality and also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).

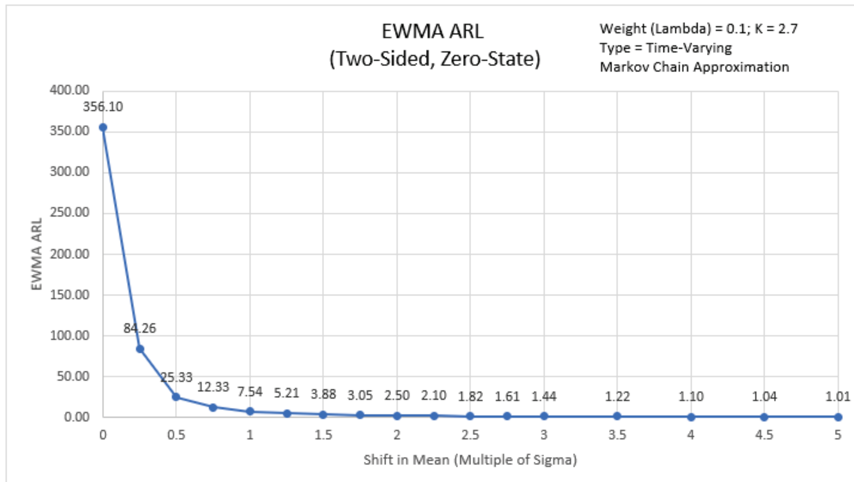
For further details on the Markov Chain approximation see Lucas [1] for fixed and Steiner [3] for time-varying. Monte Carlo simulation uses the Pearson Family of distributions to match the specified skewness and kurtosis.

The EWMA ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The mean and standard deviation are also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.

All ARL calculations for EWMA use a standardized in-control mean=0 and sigma=1.

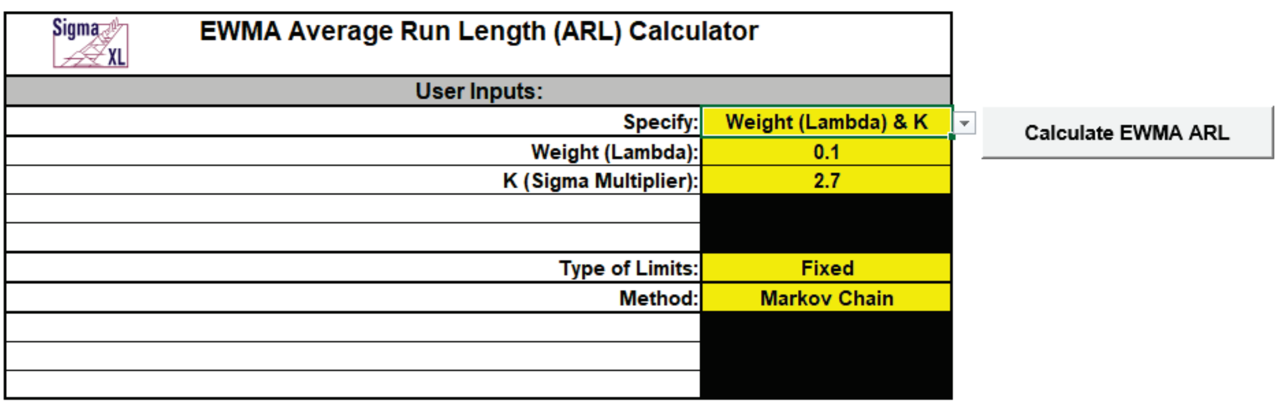
3. Click the **Calculate EWMA ARL** button to reproduce the ARL table and chart.

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	356.10
0.25	84.26
0.5	25.33
0.75	12.33
1	7.54
1.25	5.21
1.5	3.88
1.75	3.05
2	2.50
2.25	2.10
2.5	1.82
2.75	1.61
3	1.44
3.5	1.22
4	1.10
4.5	1.04
5	1.01



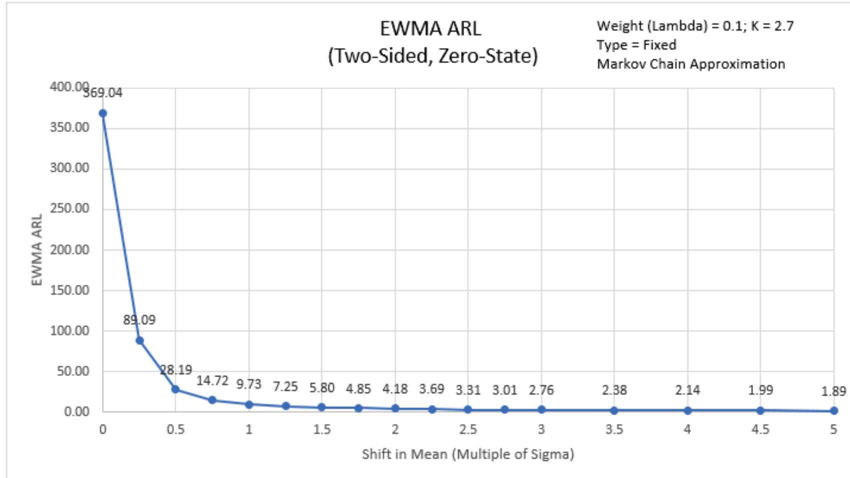
The ARL_0 (in-control ARL with 0 shift in mean) for the EWMA chart with these settings is 356.1, which is close to the Shewhart ARL_0 of 370.4. The ARL_1 for a small 1 sigma shift in mean is 7.54, so is much faster to detect than the ARL_1 of 43.89 for **Shewhart Individuals** and faster to detect than the Monte Carlo ARL_1 of 9.7 for the **Shewhart Individuals with 8 tests for special causes**.

4. Now we will compare time-varying to fixed limits. Select **Specify** = *Weight (Lambda) & K*. Enter **Weight (Lambda)** = 0.1, **K (Sigma Multiplier)** = 2.7, **Type of Limits** = *Fixed*, **Method** = *Markov Chain.*



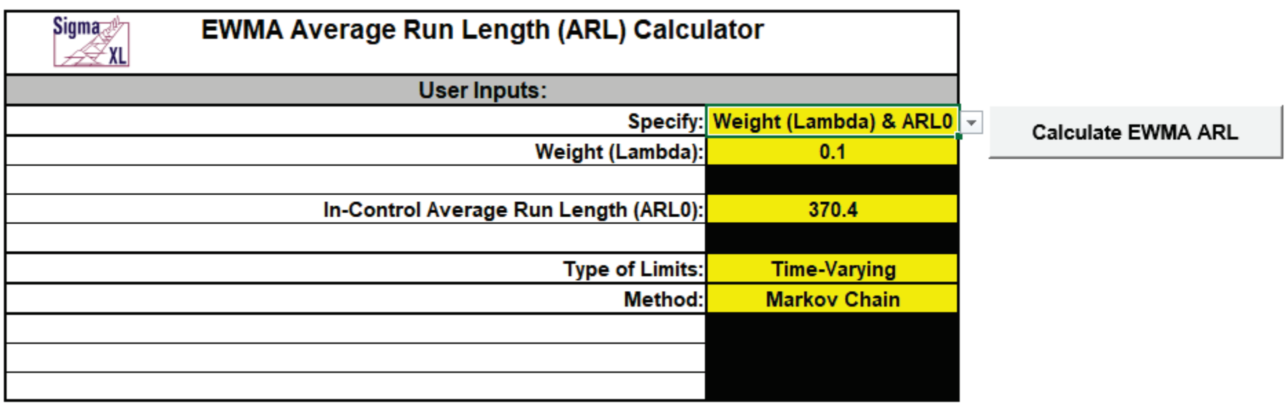
5. Click the **Calculate EWMA ARL** button to produce the ARL table and chart for these settings:

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	369.04
0.25	89.09
0.5	28.19
0.75	14.72
1	9.73
1.25	7.25
1.5	5.80
1.75	4.85
2	4.18
2.25	3.69
2.5	3.31
2.75	3.01
3	2.76
3.5	2.38
4	2.14
4.5	1.99
5	1.89



The ARL_0 for the EWMA chart with these settings is 369.04, which is close to the Shewhart ARL_0 of 370.4. The ARL_1 for a small 1 sigma shift in mean with fixed limits is 9.73, which is slower than the ARL_1 with time-varying limits of 7.54.

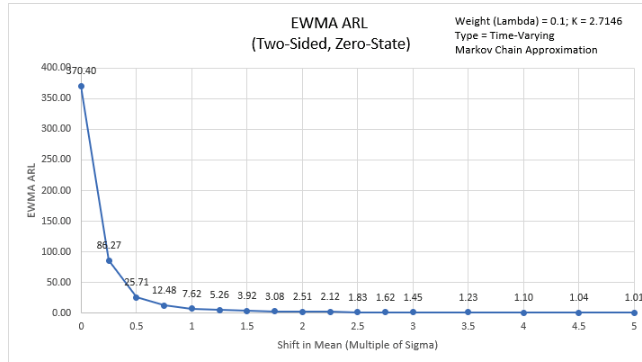
6. Next, we will specify Weight (Lambda) = 0.1, the Shewhart ARL_0 value of 370.4 and solve for the K (Sigma Multiplier). Enter **Specify** = *Weight (Lambda) & ARL0*, **Weight (Lambda)** = 0.1, **In-Control Average Run Length (ARL0)** = 370.4, **Type of Limits** = *Time-Varying*, **Method** = *Markov Chain.*



7. Click the **Calculate EWMA ARL** button to produce the updated EWMA Parameters, ARL table and chart for these settings:

EWMA Parameters:	
Weight (Lambda)	0.1
K (Sigma Multiplier)	2.7146
Type of Limits	Time-Varying
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	370.40
0.25	86.27
0.5	25.71
0.75	12.48
1	7.62
1.25	5.26
1.5	3.92
1.75	3.08
2	2.51
2.25	2.12
2.5	1.83
2.75	1.62
3	1.45
3.5	1.23
4	1.10
4.5	1.04
5	1.01



The ARL_0 for the EWMA chart with these settings is 370.4 as specified. The K (Sigma Multiplier) solved to obtain this ARL_0 value is 2.7146. The ARL_1 for a small 1 sigma shift in mean is 7.62.

8. Now we will specify Weight (Lambda) = 0.05, the Shewhart ARL_0 value of 370.4 and solve for the K (Sigma Multiplier). Enter **Specify** = *Weight (Lambda) & ARL0*, **Weight (Lambda)** = 0.05, **In-Control Average Run Length (ARL0)** = 370.4, **Type of Limits** = *Time-Varying*, **Method** = *Markov Chain.*

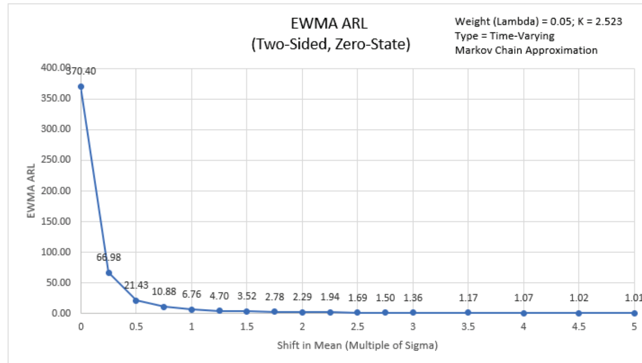
EWMA Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Weight (Lambda) & ARL0
Weight (Lambda):	0.05
In-Control Average Run Length (ARL0):	370.4
Type of Limits:	Time-Varying
Method:	Markov Chain

Calculate EWMA ARL

9. Click the **Calculate EWMA ARL** button to produce the updated EWMA Parameters, ARL table and chart for these settings:

EWMA Parameters:	
Weight (Lambda)	0.05
K (Sigma Multiplier)	2.523
Type of Limits	Time-Varying
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	370.40
0.25	66.98
0.5	21.43
0.75	10.88
1	6.76
1.25	4.70
1.5	3.52
1.75	2.78
2	2.29
2.25	1.94
2.5	1.69
2.75	1.50
3	1.36
3.5	1.17
4	1.07
4.5	1.02
5	1.01



The ARL_0 for the EWMA chart with these settings is 370.4 as specified. The K (Sigma Multiplier) solved to obtain this ARL_0 value is 2.523. The ARL_1 for a small 1 sigma shift in mean is 6.76.

10. Next, we will specify an ARL_0 value of 500 with a desired optimization to detect a 1 sigma shift in mean. The calculator will solve for the optimal Weight (Lambda) and K (Sigma Multiplier). Enter **Specify** = *ARL0 & Shift*, **In-Control Average Run Length (ARL0)** = 500, **Shift in Mean to Detect (Multiple of Sigma)** = 1, **Type of Limits** = *Time-Varying*, **Method** = *Markov Chain.*

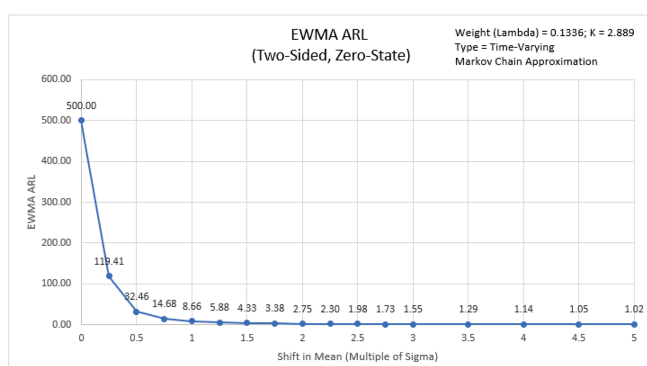
EWMA Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	ARL0 & Shift
In-Control Average Run Length (ARL0):	500
Shift in Mean to Detect (Multiple of Sigma):	1
Type of Limits:	Time-Varying
Method:	Markov Chain

Calculate EWMA ARL

11. Click the **Calculate EWMA ARL** button to produce the EWMA Parameters, ARL table and chart for these settings:

EWMA Parameters:	
Weight (Lambda)	0.1336
K (Sigma Multiplier)	2.889
Type of Limits	Time-Varying
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	500.00
0.25	119.41
0.5	32.46
0.75	14.68
1	8.66
1.25	5.88
1.5	4.33
1.75	3.38
2	2.75
2.25	2.30
2.5	1.98
2.75	1.73
3	1.55
3.5	1.29
4	1.14
4.5	1.05
5	1.02



The ARL_0 for the EWMA chart with these settings is 500.0 as specified. The ARL_1 for a small 1 sigma shift in mean is 8.66. The solved parameters are Weight (Lambda) = 0.1336 and K (Sigma Multiplier) = 2.889.

Note: Weight (Lambda) is first solved using fixed limits. This value is then used to solve for K using time-varying limits.

12. Now we will use Monte Carlo simulation to obtain approximate Run Length standard deviation and percentiles. Enter **Specify** = *Weight (Lambda) & K*, **Weight (Lambda)** = 0.1, **K (Sigma Multiplier)** = 2.7, **Type of Limits** = *Time-Varying*, **Method** = *Monte Carlo*, **Number of Replications** = 1e4, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0.

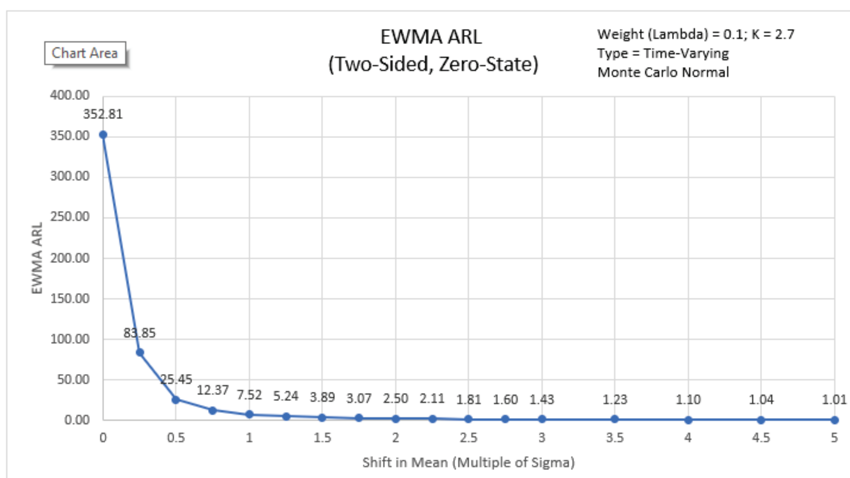
EWMA Average Run Length (ARL) Calculator

User Inputs:	
Specify:	Weight (Lambda) & K
Weight (Lambda):	0.1
K (Sigma Multiplier):	2.7
Type of Limits:	Time-Varying
Method:	Monte Carlo
Number of Replications:	1.00E+04
Skewness:	0
Kurtosis (Normal is 0):	0

Calculate EWMA ARL

13. Click the **Calculate EWMA ARL** button to produce the Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table. Monte Carlo simulation with 10,000 (1e4) replications will take about a minute to run.

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	352.81
0.25	83.85
0.5	25.45
0.75	12.37
1	7.52
1.25	5.24
1.5	3.89
1.75	3.07
2	2.50
2.25	2.11
2.5	1.81
2.75	1.60
3	1.43
3.5	1.23
4	1.10
4.5	1.04
5	1.01



	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	361.02	2	14	33	100	238	485	830	1083	1651
0.25	79.98	2	7	12	28	60	114	185	243	388
0.5	20.40	1	3	5	11	20	34	52	65	96
0.75	8.65	1	2	3	6	10	16	24	29	40
1	4.88	1	2	2	4	7	10	14	17	24
1.25	3.19	1	1	2	3	5	7	10	11	15
1.5	2.25	1	1	1	2	3	5	7	8	11
1.75	1.69	1	1	1	2	3	4	5	6	8
2	1.32	1	1	1	2	2	3	4	5	7
2.25	1.05	1	1	1	1	2	3	3	4	5
2.5	0.85	1	1	1	1	2	2	3	3	4
2.75	0.72	1	1	1	1	1	2	3	3	4
3	0.61	1	1	1	1	1	2	2	3	3
3.5	0.45	1	1	1	1	1	1	2	2	3
4	0.31	1	1	1	1	1	1	1	2	2
4.5	0.19	1	1	1	1	1	1	1	1	2
5	0.11	1	1	1	1	1	1	1	1	2

The additional run length statistics show the large variation of run length values. The MRL_0 = 238 (in-control median run length with 0 shift in process mean).

Note: The results will vary slightly since this is Monte Carlo simulation.

14. We will now assess robustness to nonnormality using Monte Carlo simulation with Weight (Lambda) = 0.1 and a specified ARL_0 of 370.4 for comparison to Shewhart. Enter **Specify** = *Weight (Lambda) & ARL0*, **Weight (Lambda)** = 0.1, **In-Control Average Run Length (ARL0)** = 370.4, **Type of Limits** = *Time-Varying*, **Method** = *Monte Carlo*, **Number of Replications** = 1e4, **Skewness** = 2, **Kurtosis (Normal is 0)** = 6.

EWMA Average Run Length (ARL) Calculator

User Inputs:	
Specify:	Weight (Lambda) & ARL0
Weight (Lambda):	0.1
In-Control Average Run Length (ARL0):	370.4
Type of Limits:	Time-Varying
Method:	Monte Carlo
Number of Replications:	1.00E+04
Skewness:	2
Kurtosis (Normal is 0):	6

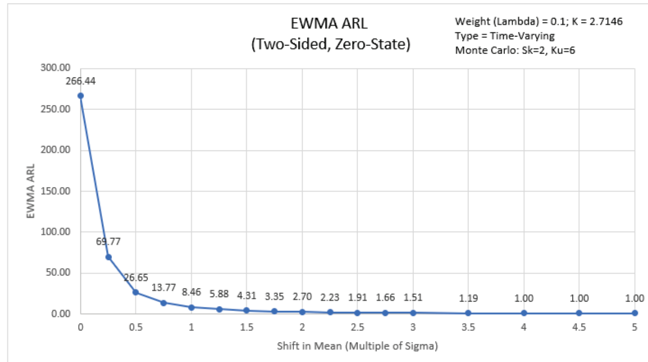
Calculate EWMA ARL

Note: The EWMA parameter K (Sigma Multiplier) will be solved using Markov-Chain approximation and assume a Normal distribution, so will match the value previously calculated above (6).

15. Click the **Calculate EWMA ARL** button to produce the updated EWMA Parameters, Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table:

EWMA Parameters:	
Weight (Lambda)	0.1
K (Sigma Multiplier)	2.7146
Type of Limits	Time-Varying
Method	Monte Carlo: Sk=2, Ku=6

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	266.44
0.25	69.77
0.5	26.65
0.75	13.77
1	8.46
1.25	5.88
1.5	4.31
1.75	3.35
2	2.70
2.25	2.23
2.5	1.91
2.75	1.66
3	1.51
3.5	1.19
4	1.00
4.5	1.00
5	1.00



	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	276.09	1	5	18	68	183	372	618	804	1298
0.25	68.39	1	2	7	21	50	97	157	205	322
0.5	22.63	1	2	4	10	21	36	57	71	103
0.75	10.24	1	1	3	6	12	19	27	33	48
1	5.42	1	1	2	4	8	12	16	19	25
1.25	3.38	1	1	2	3	6	8	10	12	16
1.5	2.24	1	1	1	3	4	6	7	8	10
1.75	1.59	1	1	1	2	3	4	5	6	7
2	1.17	1	1	1	2	3	4	4	5	5
2.25	0.89	1	1	1	2	2	3	3	4	4
2.5	0.71	1	1	1	1	2	2	3	3	3
2.75	0.55	1	1	1	1	2	2	2	2	3
3	0.50	1	1	1	1	2	2	2	2	2
3.5	0.39	1	1	1	1	1	1	2	2	2
4	0.00	1	1	1	1	1	1	1	1	1
4.5	0.00	1	1	1	1	1	1	1	1	1
5	0.00	1	1	1	1	1	1	1	1	1

ARL_0 is approximately 266.4 which is a 1.4 x increase (370.4/266.4) in false alarms compared to Normal but is a much better performance than the ARL_0 = 55 result for **Shewhart Individuals**. MRL_0 is approx. 183.

16. Next, we will assess robustness to nonnormality using Monte Carlo simulation with a lower Weight (Lambda) = 0.05 and a specified ARL_0 of 370.4. Enter **Specify** = *Weight (Lambda) & ARL0*, **Weight (Lambda)** = 0.05, **In-Control Average Run Length (ARL0)** = 370.4, **Type of Limits** = *Time-Varying*, **Method** = *Monte Carlo*, **Number of Replications** = 1e4, **Skewness** = 2, **Kurtosis (Normal is 0)** = 6.

EWMA Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	Weight (Lambda) & ARL0
Weight (Lambda):	0.05
In-Control Average Run Length (ARL0):	370.4
Type of Limits:	Time-Varying
Method:	Monte Carlo
Number of Replications:	1.00E+04
Skewness:	2
Kurtosis (Normal is 0):	6

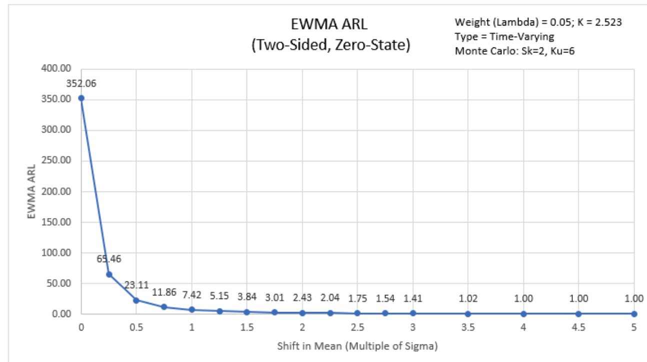
Calculate EWMA ARL

Note: Montgomery [2] (Table 9.12) and Borror, Montgomery & Runger [4] point out that the EWMA becomes more robust with lower values of Lambda.

17. Click the **Calculate EWMA ARL** button to produce the updated EWMA Parameters, Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table:

EWMA Parameters:	
Weight (Lambda)	0.05
K (Sigma Multiplier)	2.523
Type of Limits	Time-Varying
Method	Monte Carlo: Sk=2, Ku=6

Shift in Mean (Multiple of Sigma)	EWMA ARL (Two-Sided, Zero-State)
0	352.06
0.25	65.46
0.5	23.11
0.75	11.86
1	7.42
1.25	5.15
1.5	3.84
1.75	3.01
2	2.43
2.25	2.04
2.5	1.75
2.75	1.54
3	1.41
3.5	1.02
4	1.00
4.5	1.00
5	1.00



		Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)								
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	370.14	1	3	15	83	239	499	842	1101	1625
0.25	62.29	1	2	5	20	49	91	146	193	286
0.5	18.80	1	1	3	9	19	33	48	58	86
0.75	8.38	1	1	2	5	10	16	23	28	37
1	4.76	1	1	2	4	7	10	14	16	21
1.25	2.96	1	1	1	3	5	7	9	11	13
1.5	2.01	1	1	1	2	4	5	7	7	9
1.75	1.44	1	1	1	2	3	4	5	5	6
2	1.05	1	1	1	2	2	3	4	4	5
2.25	0.81	1	1	1	1	2	3	3	3	4
2.5	0.63	1	1	1	1	2	2	3	3	3
2.75	0.50	1	1	1	1	2	2	2	2	2
3	0.49	1	1	1	1	1	2	2	2	2
3.5	0.15	1	1	1	1	1	1	1	1	2
4	0.00	1	1	1	1	1	1	1	1	1
4.5	0.00	1	1	1	1	1	1	1	1	1
5	0.00	1	1	1	1	1	1	1	1	1

ARL_0 is approximately 352 which is close to the original specified 370.4. The MRL_0 is approx. 239. If robustness to non-normality is a concern then a Weight (Lambda) = 0.05 is recommended.

Template Notes:

1. Specify EWMA parameters: **Weight (Lambda) & K, Weight (Lambda) & ARL0** or **ARL0 & Shift** using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden.
2. If applicable, enter the EWMA parameter **Weight (Lambda)**. This is a value between 0 and 1 and controls the amount of influence that previous observations have on the current EWMA statistic. A value near 1 puts almost all weight on the current observation, making it resemble a Shewhart chart. For values near 0, a small weight is applied to almost all of the past observations, so the EWMA chart performance is similar to that of a CUSUM chart.
3. If applicable, enter the EWMA parameter **K (Sigma Multiplier)**. This is a value typically between 2 and 4. It is also referred to as L, but SigmaXL uses K to avoid confusion with Lambda.
4. If applicable, enter the desired **In-Control Average Run Length (ARL0)**. This will be the target ARL for mean shift = 0 and should be a large value to minimize false alarms, typically 370 to 500. The K (Sigma Multiplier) will be solved to achieve this ARL0, given a specified Weight (Lambda) value.
5. If applicable, enter the desired **Shift in Mean to Detect (Multiple of Sigma)**. The Weight (Lambda) value that minimizes ARL for the specified shift will be solved.
6. Select Type of Limits: **Time-Varying** or **Fixed** using the drop-down list.
7. The EWMA control chart template uses time-varying control limits since they improve the sensitivity of the EWMA to detect early changes in the process mean. Published ARL tables typically use fixed limits, so providing both allows comparison between the two types.
8. Select Method: **Markov Chain** or **Monte Carlo** using the drop-down list. Markov Chain approximation is fast and accurate to compute ARLs. Monte Carlo simulation allows you to assess robustness to non-normality and also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).
9. For further details on the Markov Chain approximation see Lucas [1] for fixed and Steiner [3] for time-varying. Monte Carlo simulation uses the Pearson Family of distributions to match the specified skewness and kurtosis.
10. If applicable, enter **Number of Replications**. 1000 (1e3) replications will be fast, approx. 10 seconds, but will have an ARL0 error approx. = +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL0 error = +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL0 error = +/- 1%.
11. If applicable, enter **Skewness**. Skewness must be >= 0. Skewness = 0 is symmetric.
12. If applicable, enter **Kurtosis (Normal is 0)**. Also known as Excess Kurtosis, it must be >= Skewness^2 - 1.48. This is required to keep the distribution unimodal. If Skewness=0 and Kurtosis = 0, the distribution is normal.
13. Click the **Calculate EWMA ARL** button to produce the ARL table and chart. If Monte Carlo was selected, the table of Run Length Standard Deviation and Percentiles will also be produced.
14. The EWMA ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The mean and standard deviation are also assumed to be known.

This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.

15. Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for this template to function. Do not add or delete rows or columns in this template.

REFERENCES:

[1] Lucas J.M. and Saccucci M.S. (1990), “Exponentially weighted moving average control schemes: Properties and enhancements”, *Technometrics* 32, 1-12.

[2] Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley.

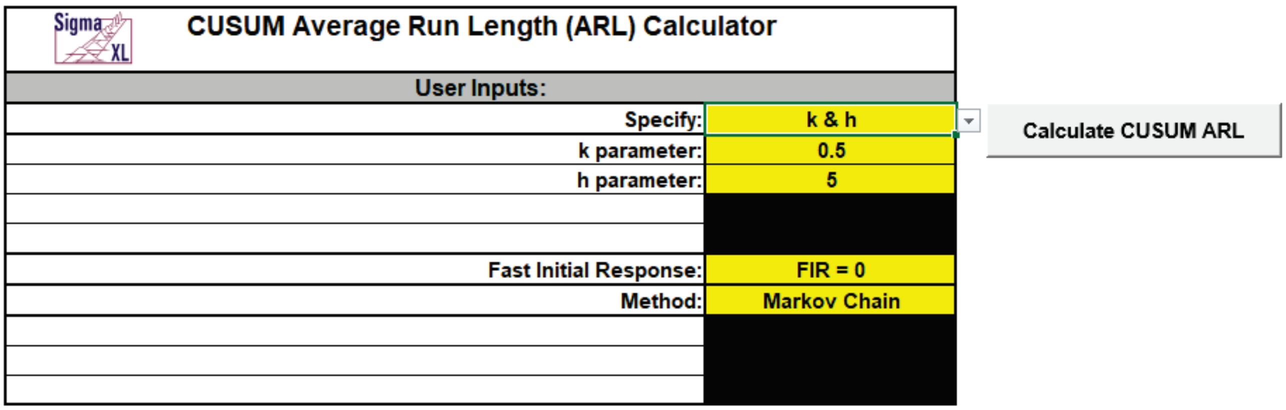
[3] Steiner, S. H. (1999), "EWMA control charts with time-varying control limits and fast initial response", *Journal of Quality Technology* 31(1), 75-86.

[4] Borror, C. M., Montgomery, D.C. and Runger G. C. (1999). “Robustness of the EWMA Control Chart to Nonnormality,” *Journal of Quality Technology*, 31(3), 309–316.

CUSUM ARL

1. Click **SigmaXL > Templates & Calculators > Control Chart Templates > Average Run Length (ARL) Calculators > CUSUM ARL**. This template is also located at **SigmaXL > Control Charts > Control Chart Templates > Average Run Length (ARL) Calculators > CUSUM ARL**.

2. The default template settings are **Specify** = *k & h*, **k parameter** = 0.5, **h parameter** = 5, **Fast Initial Response** *FIR = 0*, **Method** = *Markov Chain.*



Notes: Parameters to be specified will be shown in yellow highlight, otherwise they are hidden. The CUSUM parameter k is the reference (or slack) value, typically set to 0.5. It sets the size of mean shift (2k sigma) that you would like to detect quickly, so 0.5 denotes rapid detection of a shift in mean = 1 sigma. Alternatively, Woodall & Faltin [4] recommend larger k values (e.g., k = 0.9) to avoid false alarms and detect shifts of practical significance.

The CUSUM parameter h is the decision interval, typically set to 4 or 5.

FIR sets the initial CUSUM statistic so that it improves the sensitivity to a mean shift at startup. Note that if the process is in control when the CUSUM is started but shifts out of control later, the more appropriate ARL for such a case would be FIR=0. See Montgomery [3], pages 426-427.

Markov Chain approximation is fast and accurate to compute ARLs. Monte Carlo simulation allows you to assess robustness to nonnormality and also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).

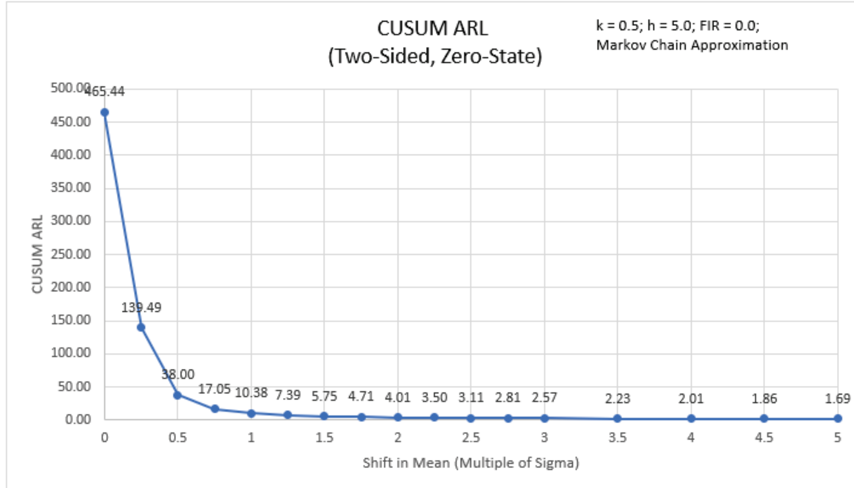
For further details on the Markov Chain approximation see Hawkins [1] and Lucas [2]. Monte Carlo simulation uses the Pearson Family of distributions to match the specified skewness and kurtosis.

The CUSUM ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The mean and standard deviation are also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.

All ARL calculations for CUSUM use a standardized in-control mean=0 and sigma=1.

3. Click the **Calculate CUSUM ARL** button to reproduce the ARL table and chart.

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	465.44
0.25	139.49
0.5	38.00
0.75	17.05
1	10.38
1.25	7.39
1.5	5.75
1.75	4.71
2	4.01
2.25	3.50
2.5	3.11
2.75	2.81
3	2.57
3.5	2.23
4	2.01
4.5	1.86
5	1.69



The ARL_0 (in-control ARL with 0 shift in mean) for the CUSUM chart with these settings is 465.44. The ARL_1 for a small 1 sigma shift in mean is 10.38.

4. Now we will evaluate CUSUM with the same parameters, but use the Fast Initial Response option. Select **Specify** = *k & h*. Enter **k parameter** = 0.5, **h parameter** = 5, **Fast Initial Response** *FIR = h/2*, **Method** = *Markov Chain.*

CUSUM Average Run Length (ARL) Calculator

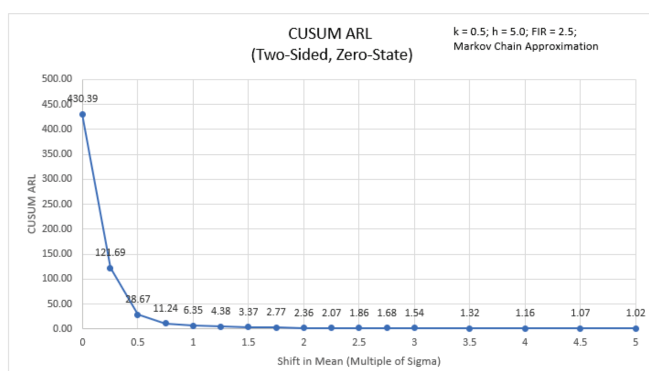
User Inputs:	
Specify:	k & h
k parameter:	0.5
h parameter:	5
Fast Initial Response:	FIR = h/2
Method:	Markov Chain

Calculate CUSUM ARL

5. Click the **Calculate CUSUM ARL** button to produce the ARL table and chart for these settings.

Specified CUSUM Parameters:	
k	0.5
h	5.0
FIR	2.5
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	430.39
0.25	121.69
0.5	28.67
0.75	11.24
1	6.35
1.25	4.38
1.5	3.37
1.75	2.77
2	2.36
2.25	2.07
2.5	1.86
2.75	1.68
3	1.54
3.5	1.32
4	1.16
4.5	1.07
5	1.02



The ARL_0 for the CUSUM chart with these settings is 430.39 which is slightly lower than the *FIR = 0* setting, so slightly higher false alarm rate. The ARL_1 for a small 1 sigma shift in mean is 6.35 which is faster than the 10.38 for *FIR = 0*.

6. Now we will specify the CUSUM k parameter = 0.5 with a Shewhart ARL_0 value of 370.4 and solve for the h parameter. Enter **Specify** = *k & ARL0*, **k parameter** = 0.5, **In-Control Average Run Length (ARL0)** = 370.4, **Fast Initial Response** *FIR = 0*, **Method** = *Markov Chain.*

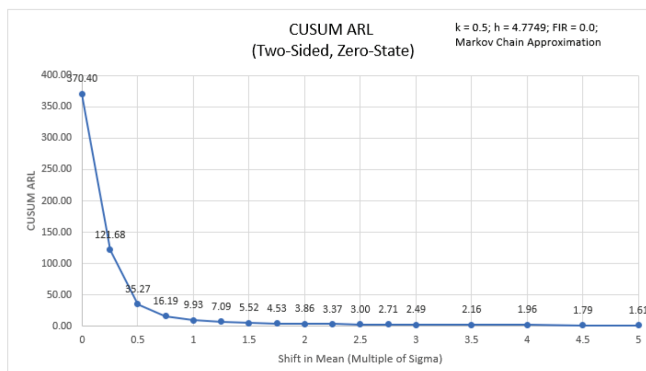
CUSUM Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	k & ARL0
k parameter:	0.5
In-Control Average Run Length (ARL0):	370.4
Fast Initial Response:	FIR = 0
Method:	Markov Chain

Calculate CUSUM ARL

7. Click the **Calculate CUSUM ARL** button to produce the CUSUM Parameters, ARL table and chart for these settings.

CUSUM Parameters:	
k	0.5
h	4.7749
FIR	0.0
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	370.40
0.25	121.68
0.5	35.27
0.75	16.19
1	9.93
1.25	7.09
1.5	5.52
1.75	4.53
2	3.86
2.25	3.37
2.5	3.00
2.75	2.71
3	2.49
3.5	2.16
4	1.96
4.5	1.79
5	1.61



The ARL_0 for the CUSUM chart with these settings is 370.4 as specified. The h parameter solved to obtain this ARL_0 value is 4.7749. The ARL_1 for a small 1 sigma shift in mean is 9.93 so is much faster to detect than the ARL_1 of 43.89 for **Shewhart Individuals** and close to the Monte Carlo ARL_1 of 9.7 for **Shewhart Individuals with 8 tests for special causes**.

8. Next, we will specify a Shewhart ARL_0 value of 370.4, with a desired optimization to detect a 1 sigma shift in mean and use Fast Initial Response. The calculator will solve for the optimal k and h parameters. Enter **Specify** = *ARL0 & Shift*, **In-Control Average Run Length (ARL0)** = 370.4, **Shift in Mean to Detect (Multiple of Sigma)** = 1, **Fast Initial Response** *FIR = h/2*, **Method** = *Markov Chain.*

CUSUM Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	ARL0 & Shift
In-Control Average Run Length (ARL0):	370.4
Shift in Mean to Detect (Multiple of Sigma):	1
Fast Initial Response:	FIR = h/2
Method:	Markov Chain

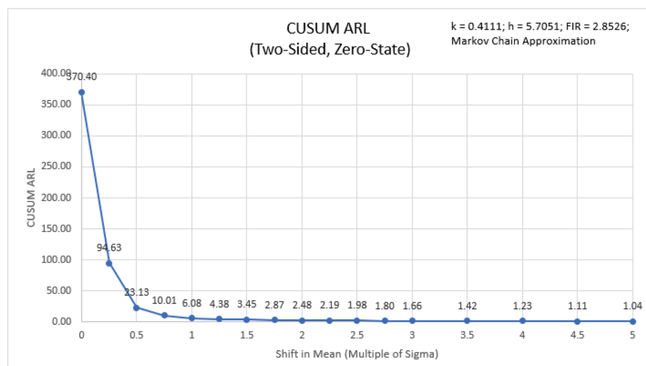
Calculate CUSUM ARL

Note: Since both k and h are solved, this takes about 20-30 seconds to compute.

9. Click the **Calculate CUSUM ARL** button to produce the CUSUM Parameters, ARL table and chart for these settings.

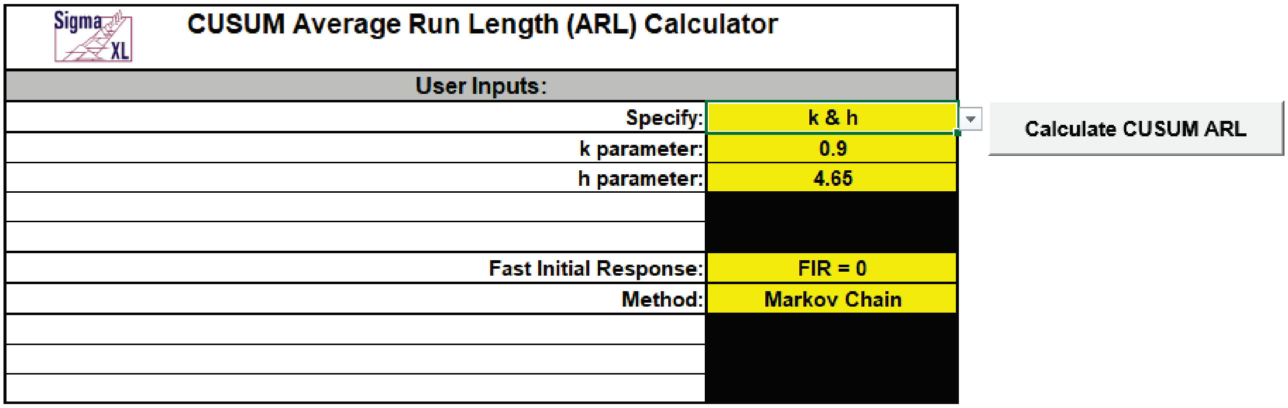
CUSUM Parameters:	
k	0.4111
h	5.7051
FIR	2.8526
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	370.40
0.25	94.63
0.5	23.13
0.75	10.01
1	6.08
1.25	4.38
1.5	3.45
1.75	2.87
2	2.48
2.25	2.19
2.5	1.98
2.75	1.80
3	1.66
3.5	1.42
4	1.23
4.5	1.11
5	1.04



The ARL_0 for the CUSUM chart with these settings is 370.4 as specified. ARL_1 for a small 1 sigma shift in mean is 6.08 so is much faster to detect than the ARL_1 of 43.89 for **Shewhart Individuals** and faster than the Monte Carlo ARL_1 of 9.7 for **Shewhart Individuals with 8 tests for special causes**. It is slightly faster to detect than the ARL_1 of 6.76 for **Time-Varying EWMA with Weight(Lambda) = .05**. The solved parameters are k = 0.4111, h = 5.7051 and FIR = 2.8526.

10. As an alternative to using the CUSUM to rapidly detect small shifts in mean, Woodall & Faltin [4] recommend larger k values to avoid false alarms and detect shifts of practical significance. Enter **Specify** = *k & h*, **k parameter** = 0.9, **h parameter** = 4.65, **Fast Initial Response** *FIR = 0*, **Method** = *Markov Chain.*



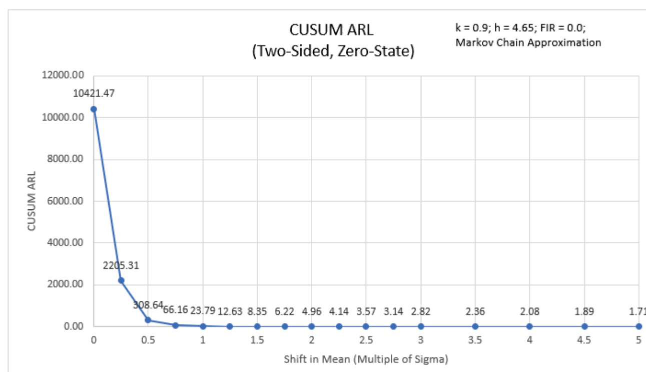
CUSUM Average Run Length (ARL) Calculator	
User Inputs:	
Specify:	k & h
k parameter:	0.9
h parameter:	4.65
Fast Initial Response:	FIR = 0
Method:	Markov Chain

Calculate CUSUM ARL

11. Click the **Calculate CUSUM ARL** button to produce the CUSUM Parameters, ARL table and chart for these settings.

Specified CUSUM Parameters:	
k	0.9
h	4.65
FIR	0.0
Method	Markov Chain Approximation

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	10421.47
0.25	2205.31
0.5	308.64
0.75	66.16
1	23.79
1.25	12.63
1.5	8.35
1.75	6.22
2	4.96
2.25	4.14
2.5	3.57
2.75	3.14
3	2.82
3.5	2.36
4	2.08
4.5	1.89
5	1.71



This gives large ARL values for shift in mean <= 1 sigma and small ARL values for a shift in mean >= 1.5 sigma.

12. Now we will use Monte Carlo simulation to obtain approximate Run Length standard deviation and percentiles using the CUSUM k parameter = 0.5 with a Shewhart ARL_0 value of 370.4. Enter **Specify** = *k & ARL0*, **k parameter** = 0.5, **In-Control Average Run Length (ARL0)** = 370.4, **Fast Initial Response** *FIR = 0*, **Method** = *Monte Carlo*, **Number of Replications** = 1e4, **Skewness** = 0, **Kurtosis (Normal is 0)** = 0.

CUSUM Average Run Length (ARL) Calculator

User Inputs:	
Specify:	k & ARL0
k parameter:	0.5
In-Control Average Run Length (ARL0):	370.4
Fast Initial Response:	FIR = 0
Method:	Monte Carlo
Number of Replications:	1.00E+04
Skewness:	0
Kurtosis (Normal is 0):	0

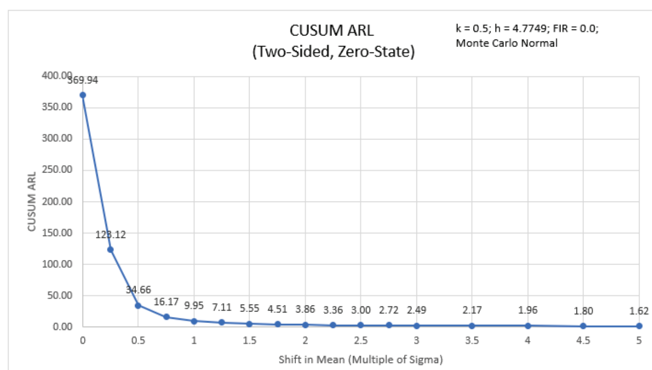
Calculate CUSUM ARL

Note: The CUSUM h parameter will be solved first using the Markov Chain approximation and assumes a Normal distribution, so will match the 4.7749 value previously calculated above (6).

13. Click the **Calculate CUSUM ARL** button to produce the CUSUM Parameters, Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table. Monte Carlo simulation with 10,000 (1e4) replications will take about a minute to run.

CUSUM Parameters:	
k	0.5
h	4.7749
FIR	0.0
Method	Monte Carlo Normal

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	369.94
0.25	123.12
0.5	34.66
0.75	16.17
1	9.95
1.25	7.11
1.5	5.55
1.75	4.51
2	3.86
2.25	3.36
2.5	3.00
2.75	2.72
3	2.49
3.5	2.17
4	1.96
4.5	1.80
5	1.62



	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	366.35	9	24	45	110	255	513	860	1098	1662
0.25	115.82	6	12	18	40	89	167	276	359	541
0.5	28.17	4	7	9	15	26	45	72	90	136
0.75	10.50	3	5	6	9	13	21	30	37	53
1	5.30	3	4	5	6	9	12	17	20	28
1.25	3.22	2	3	4	5	6	9	11	13	18
1.5	2.22	2	3	3	4	5	7	8	10	13
1.75	1.60	2	2	3	3	4	5	7	8	9
2	1.27	2	2	2	3	4	5	5	6	8
2.25	1.03	2	2	2	3	3	4	5	5	6
2.5	0.87	2	2	2	2	3	3	4	5	5
2.75	0.73	2	2	2	2	3	3	4	4	5
3	0.64	1	2	2	2	2	3	3	4	4
3.5	0.48	1	2	2	2	2	2	3	3	3
4	0.40	1	1	2	2	2	2	2	3	3
4.5	0.43	1	1	1	2	2	2	2	2	3
5	0.49	1	1	1	1	2	2	2	2	2

The additional run length statistics show the large variation of run length values. The MRL_0 = 255 (in-control median run length with 0 shift in process mean).

Note: The results will vary slightly since this is Monte Carlo simulation.

14. We will now assess robustness to nonnormality using Monte Carlo simulation and compare to Shewhart and EWMA charts. Enter **Specify** = *k & ARL0*, **k parameter** = 0.5, **In-Control Average Run Length (ARL0)** = 370.4, **Fast Initial Response** *FIR = 0*, **Method** = *Monte Carlo*, **Number of Replications** = 1e4, **Skewness** = 2, **Kurtosis (Normal is 0)** = 6.

CUSUM Average Run Length (ARL) Calculator

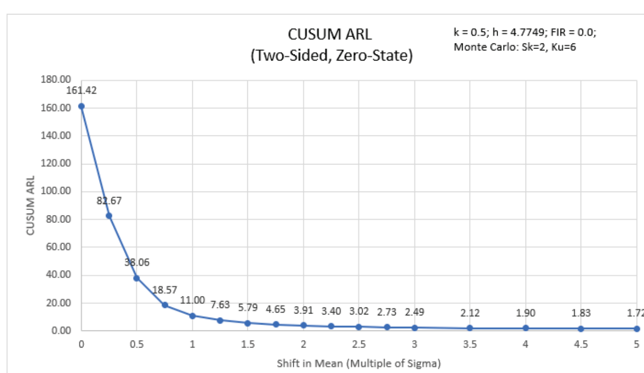
User Inputs:	
Specify:	k & ARL0
k parameter:	0.5
In-Control Average Run Length (ARL0):	370.4
Fast Initial Response:	FIR = 0
Method:	Monte Carlo
Number of Replications:	1.00E+04
Skewness:	2
Kurtosis (Normal is 0):	6

Calculate CUSUM ARL

15. Click the **Calculate CUSUM ARL** button to produce the CUSUM Parameters, Monte Carlo approximate ARL table, ARL chart and Run Length Standard Deviation and Percentiles table:

CUSUM Parameters:	
k	0.5
h	4.7749
FIR	0.0
Method	Monte Carlo: Sk=2, Ku=6

Shift in Mean (Multiple of Sigma)	CUSUM ARL (Two-Sided, Zero-State)
0	161.42
0.25	82.67
0.5	38.06
0.75	18.57
1	11.00
1.25	7.63
1.5	5.79
1.75	4.65
2	3.91
2.25	3.40
2.5	3.02
2.75	2.73
3	2.49
3.5	2.12
4	1.90
4.5	1.83
5	1.72



	Monte Carlo Simulation Run Length Standard Deviation and Percentiles (Two-Sided, Zero-State)									
Shift in Mean (Multiple of Sigma)	Standard Deviation	1	5th	10th	25th	50th (Median)	75th	90th	95th	99th
0	161.72	3	10	18	47	112	223	363	480	757
0.25	80.11	3	7	12	26	57	113	187	243	372
0.5	33.56	2	5	8	15	28	50	82	106	157
0.75	13.19	2	4	6	9	15	24	36	44	63
1	5.96	2	3	4	7	10	14	19	22	30
1.25	3.37	2	3	4	5	7	10	12	14	17
1.5	2.18	2	3	3	4	6	7	9	10	11
1.75	1.54	1	2	3	4	5	6	7	7	8
2	1.16	1	2	2	3	4	5	5	6	6
2.25	0.93	1	2	2	3	3	4	5	5	5
2.5	0.77	1	2	2	3	3	4	4	4	4
2.75	0.66	1	2	2	2	3	3	3	4	4
3	0.58	1	2	2	2	3	3	3	3	3
3.5	0.49	1	1	2	2	2	2	3	3	3
4	0.31	1	1	1	2	2	2	2	2	2
4.5	0.38	1	1	1	2	2	2	2	2	2
5	0.45	1	1	1	1	2	2	2	2	2

ARL_0 is approximately 161.4 which is a 2.3 x increase (370.4/161.4) in false alarms compared to Normal but is a much better performance than the ARL_0 = 54.6 result for **Shewhart Individuals**.

It is, however, less robust to nonnormality than **EWMA chart with Weight (Lambda) = 0.1** which had a Monte Carlo ARL_0 = 266.4.

MRL_0 is approximately 112.

Stoumbos and Reynolds [5] recommend setting h=6.148 (with k=0.5) as a way to improve the CUSUM robustness to non-normality.

Template Notes:

1. Specify CUSUM parameters: **k & h**, **k & ARL0** or **ARL0 & Shift** using the drop-down list. Parameters to be specified will be shown in yellow highlight, otherwise they are hidden.
2. If applicable, enter the CUSUM parameter **k**. This is the reference (or slack) value, typically set to 0.5. It sets the size of mean shift (2k sigma) that you would like to detect quickly, so 0.5 denotes rapid detection of a shift in mean = 1 sigma. Alternatively, Woodall & Faltin [4] recommend larger k values (e.g., k = 0.9) to avoid false alarms and detect shifts of practical significance.
3. If applicable, enter the CUSUM parameter **h**. This is the decision interval, typically set to 4 or 5.
4. If applicable, enter the desired **In-Control Average Run Length (ARL0)**. This will be the target ARL for mean shift = 0 and should be a large value to minimize false alarms, typically 370 to 500. The h parameter will be solved to achieve this ARL0, given a specified k value.
5. If applicable, enter the desired **Shift in Mean to Detect (Multiple of Sigma)**. This will minimize ARL for the given shift. If FIR=0, then k will be shift/2. If FIR=h/2, then k will be optimized, requiring about 20-30 seconds to compute.
6. Select **FIR=0** or **FIR=h/2** using the drop-down list. This is the fast initial response (or headstart) value.
7. FIR sets the initial CUSUM statistic so that it improves the sensitivity to a mean shift at startup. Note that if the process is in control when the CUSUM is started but shifts out of control later, the more appropriate ARL for such a case would be FIR=0. See Montgomery [3], pages 426-427.
8. Select Method: **Markov Chain** or **Monte Carlo** using the drop-down list. Markov Chain approximation is fast and accurate to compute ARLs. Monte Carlo simulation allows you to assess robustness to nonnormality and also produces the table of Run Length Standard Deviation and Percentiles (scroll right to view).
9. For further details on the Markov Chain approximation see Hawkins [1] and Lucas [2]. Monte Carlo simulation uses the Pearson Family of distributions to match the specified skewness and kurtosis.
10. If applicable, enter **Number of Replications**. 1000 (1e3) replications will be fast, approx. 10 seconds, but will have an ARL0 error approx. = +/- 10%; 10,000 (1e4) replications will take about a minute, with an ARL0 error = +/- 3.2%; 100,000 (1e5) replications will take about ten minutes, with an ARL0 error = +/- 1%.
11. If applicable, enter **Skewness**. Skewness must be >= 0. Skewness = 0 is symmetric.
12. If applicable, enter **Kurtosis (Normal is 0)**. Also known as Excess Kurtosis, it must be >= Skewness^2 - 1.48. This is required to keep the distribution unimodal. If Skewness=0 and Kurtosis = 0, the distribution is normal.
13. Click the **Calculate CUSUM ARL** button to produce the ARL table and chart. If Monte Carlo was selected, the table of Run Length Standard Deviation and Percentiles will also be produced.
14. The CUSUM ARL is for a two-sided chart with zero-state, i.e., the shift is assumed to occur at the start. The mean and standard deviation are also assumed to be known. This will not likely be the case in use, but is still useful for determining parameter settings and comparison of ARL across chart types.

15. Due to the complexity of calculations, SigmaXL must be loaded and appear on the menu in order for this template to function. Do not add or delete rows or columns in this template.

REFERENCES:

[1] Hawkins, D. M. and Olwell, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement (Information Science and Statistics)*, Springer, New York.
[2] Lucas, J.M. and Crosier R.B. (1982), “Fast Initial Response for CUSUM Quality-Control Schemes: Give Your CUSUM A Headstart”, *Technometrics* 24, 199-205.
[3] Montgomery, D.C. (2013), *Introduction to Statistical Quality Control*, Seventh Ed., Wiley.
[4] Woodall, W. H. and Faltin, F.W. (2019), "Rethinking control chart design and evaluation", *Quality Engineering* 31, 596-605.
[5] Stoumbos, Z. G. and Reynolds, M.R. Jr. (2004), “The Robustness and Performance of CUSUM Control Charts Based on the Double-Exponential and Normal Distributions”, In: Lenz, H. J., Wilrich, P. T. (eds) *Frontiers in Statistical Quality Control 7,* Physica, Heidelberg, 79-100.

Part K – Time Series Forecasting and Control Charts for Autocorrelated Data

Introduction – Time Series Forecasting

From https://en.wikipedia.org/wiki/Time_series:

> A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time.
>
> Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

The free online book by Rob Hyndman and George Athanasopoulos, "Forecasting: Principles and Practice," (https://otexts.com/fpp2/) is an excellent introductory resource to time series analysis (further denoted as "fpp2"). The book references the R forecast package (https://cran.r-project.org/web/packages/forecast/index.html), but the results given by SigmaXL will closely match those in the book. Note that SigmaXL is independent of R. Slight differences in results are due to minor differences in algorithm options and optimization methods.

The following Wikipedia articles are also helpful resources:

https://en.wikipedia.org/wiki/Exponential_smoothing

https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

Box and Jenkins, et al., 2016, is the classic book on ARIMA modeling (further denoted as "Box and Jenkins"):

> Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016). *Time Series Analysis, Forecasting and Control*, 5th edition, Wiley.

Datasets to be used in the examples include:

1. Chemical Process Concentration – Series A (Box and Jenkins)
2. Monthly Airline Passengers – Series G (Box and Jenkins)
3. Monthly Airline Passengers - Missing Values
4. Monthly Airline Passengers - Modified for Control Charts
5. Daily Electricity Demand with Predictors – ElecDaily (fpp2)
6. Sales with Indicator – Modified Series M (Box and Jenkins)
7. Half-Hourly Multiple Seasonal Electricity Demand – Taylor (R forecast)

SigmaXL provides the following tools for exploratory data analysis of time series data:

- Run Chart
- Autocorrelation Function (ACF)/Partial Autocorrelation (PACF) Plots
- Cross Correlation (CCF) Plots with Pre-Whiten Data option
- Seasonal Trend Decomposition Plots
- Spectral Density Plot with Detection of Seasonal Frequency

SigmaXL provides the following methods for time series forecasting:

- Exponential Smoothing
- Exponential Smoothing – Multiple Seasonal Decomposition (MSD)
- ARIMA – Box-Jenkins Autoregressive Integrated Moving Average
- ARIMA with Predictors
- ARIMA – MSD

Typically, either Exponential Smoothing or ARIMA may be used. It may be useful to try both to see which one gives a better model or use the average of the forecast from both methods. If the data has negative autocorrelation, ARIMA is recommended. If the data includes continuous or categorical predictors, use ARIMA with Predictors.

ARIMA assumes that the time series is stationary, i.e., it has the property that the mean, variance and autocorrelation structure do not change over time. If a time series mean is not stationary (e.g. trending), this can be corrected by differencing, computing the differences between consecutive observations for nonseasonal and between consecutive periods for seasonal data (e.g., Jan 2019 – Jan 2018, etc.). If the variance changes over time, a Box-Cox transformation may be applied to achieve constant variance. Exponential Smoothing does not require stationarity.

If the data are seasonal (i.e., influenced by seasonal factors), SigmaXL requires that the seasonal frequency be specified. Frequency is the number of observations per "cycle" unit of time, so monthly sales would be specified as seasonal frequency = 12 (observations per year). Quarterly revenue would be specified as seasonal frequency = 4. Hourly data would be 24 (observations per day). Note that "seasonal frequency" is also referred to as "seasonal period", but SigmaXL follows the convention as used in Hyndman and Athanasopoulos (fpp2).

Exponential Smoothing is limited to a maximum seasonal frequency of 24. For higher frequencies use Exponential Smoothing – Multiple Seasonal Decomposition (MSD). In MSD the seasonal component is first removed through decomposition, a nonseasonal exponential smooth model fitted to the remainder (+trend), and then the seasonal component is added back in. For forecasting, a naïve seasonal forecast is used on the seasonal component. As the name implies, Multiple Seasonal Decomposition (MSD) also accommodates multiple seasonality.

ARIMA does not have a theoretical frequency limit, but for computational efficiency and to minimize the potential loss of observations through differencing, we recommend using ARIMA – MSD for seasonal frequency greater than 52 (or with multiple frequencies). Note, ARIMA with Predictors – MSD is not available.

Details of the time series forecast methods and formulas used in SigmaXL are given in the Appendix: **Time Series Forecasting and Control Charts for Autocorrelated Data**. References are given at: **References for Time Series Forecasting and Control Charts for Autocorrelated Data**.

Introduction – Control Charts for Autocorrelated Data

Statistical process control for autocorrelated processes have typically used the EWMA (Exponentially Weighted Moving Average) one-step-ahead forecast model. The time series model forecasts the motion in the mean and an Individuals control chart is plotted of the residuals to detect assignable causes. SigmaXL extends this concept to include all of the forecast methods mentioned above.

Woodall and Faltin give some helpful guidelines on dealing with autocorrelation [20]:

- If possible, one should first attempt to remove the source of the autocorrelation.
- If the source of autocorrelation cannot be removed directly, then it may be possible to model the autocorrelation and use a feedback control scheme to reduce variability about a specified target value.
- If the source of the autocorrelation cannot be removed directly, and feedback control is not a viable option, then it is important to monitor the process with control charts which do not repeatedly give signals due to presence of the autocorrelation.

Failure to account for positive autocorrelation will produce limits that are too narrow resulting in excessive false alarms. Failure to account for negative autocorrelation will produce limits that are too wide resulting in misses. An accurate forecast for your time series means that the residuals will most often have the right properties to correctly apply a control chart, thus leading to an improved control chart with reduced false alarms and misses.

In addition to creating an Individuals control chart for residuals, a Moving Limits chart is included, which uses the one step prediction as the center line, so the control limits move with the center line.

The popular “Add Data”, “Show Last 30” and “Scroll” features in SigmaXL Chart Tools are available for these control charts. For “Add Data”, the time series models are not refitted, but used to compute the residual values for the new data.

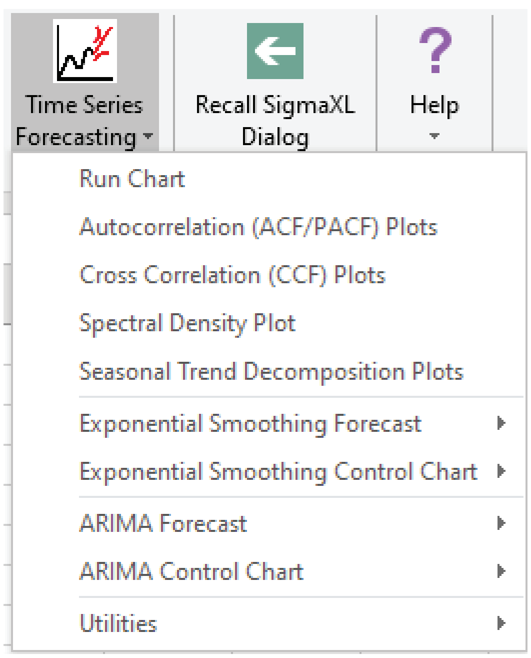
For further details and references, see the Appendix: **Control Charts for Autocorrelated Data**.

Summary of Features in Time Series Forecasting and Control Charts for Autocorrelated Data

- Run Chart
- Autocorrelation Function (ACF)/Partial Autocorrelation (PACF) Plots
- Cross Correlation (CCF) Plots with Pre-Whiten Data option
- Seasonal Trend Decomposition Plots
- Spectral Density Plot with Detection of Seasonal Frequency
- Exponential Smoothing:
    - Forecast with Prediction Intervals
    - Exponential Smoothing models use Rob Hyndman's taxonomy:
        - Additive/Multiplicative Error
        - Additive/Additive Damped Trend
        - Additive/Multiplicative Seasonal
        - This includes all of the classical exponential smoothing models such as Simple/Single/EWMA, Double and Holt-Winters
    - Multiple Seasonal Decomposition (MSD) option
        - Useful for high frequency and/or multiple frequency data, such as Monthly with frequency = 12, Daily with frequency = 7 and Hourly with frequency = 24
- Exponential Smoothing Residuals Control Chart for autocorrelated data:
    - Individuals and Moving Limits (with One-Step Ahead Forecast) Charts
    - Add Data, Show Last 30 Data Points, Enable Scroll options
    - MSD option
- Autoregressive Integrated Moving Average (ARIMA):
    - Forecast with Prediction Intervals
    - ARIMA Forecast with Predictors (Continuous and/or Categorical)
    - MSD option
- ARIMA Residuals Control Chart for autocorrelated data:
    - Individuals and Moving Limits (with One-Step Ahead Forecast) Charts
    - ARIMA Control Chart with Predictors
    - Add Data, Show Last 30 Data Points, Enable Scroll options
    - MSD option
- Utilities:
    - Difference Data
    - Lag Data
    - Interpolate Missing Values (seasonally adjusted linear interpolation)
- Time Series Forecasting Model Features:
    - ARIMA and Exponential Smoothing models are fully automatic or user specified
    - Utilizes modern State Space and Kalman Filter models for accurate parameter estimation
    - ARIMA estimates missing values with Kalman Filter; Exponential Smoothing uses seasonally adjusted linear interpolation

    - Automatic Box-Cox Transformation
    - Automatic seasonal frequency detection
- Model Diagnostics:
    - ACF/PACF Plots
    - Ljung-Box p-values
    - Log-Likelihood, AIC, AICc, BIC, Residual StDev
    - Residual plots (histogram, normal probability, residual versus fits, residuals versus order)
- Forecast Accuracy:
    - In-Sample (Estimation) one-step-ahead forecast errors (RMSE, MAE, MASE, MAPE)
    - Out-of-Sample (Withhold) one-step-ahead forecast errors
    - Out-of-Sample (Withhold) multi-step-ahead forecast errors
    - Evaluated using the benchmark standard M4 forecast competition data, a total of 100,000 data sets with Yearly, Quarterly, Monthly, Weekly, Daily and Hourly data. Using a hybrid average of automatic Exponential Smoothing and ARIMA, SigmaXL (unofficially) ranked 10th out of 60 in the Overall Weighted Average forecast accuracy score, ahead of three well known commercial forecast software packages.

Time Series Forecasting Menu



Run Chart

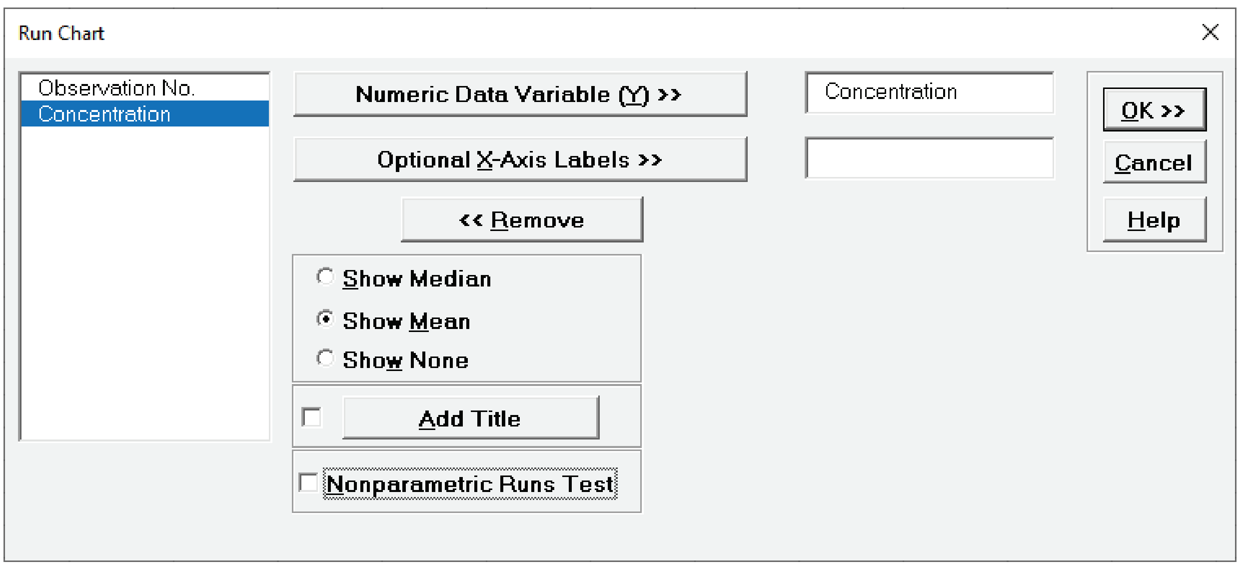
The Run Chart in Time Series Forecasting is the same tool as used in Graphical Tools > Run Chart (see Part H – Run Charts). We will demonstrate the use of Run Charts for initial exploratory data analysis of the time series examples, as well as the use of the difference and interpolate utilities, and use of SigmaXL graphical tools for data with predictors.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

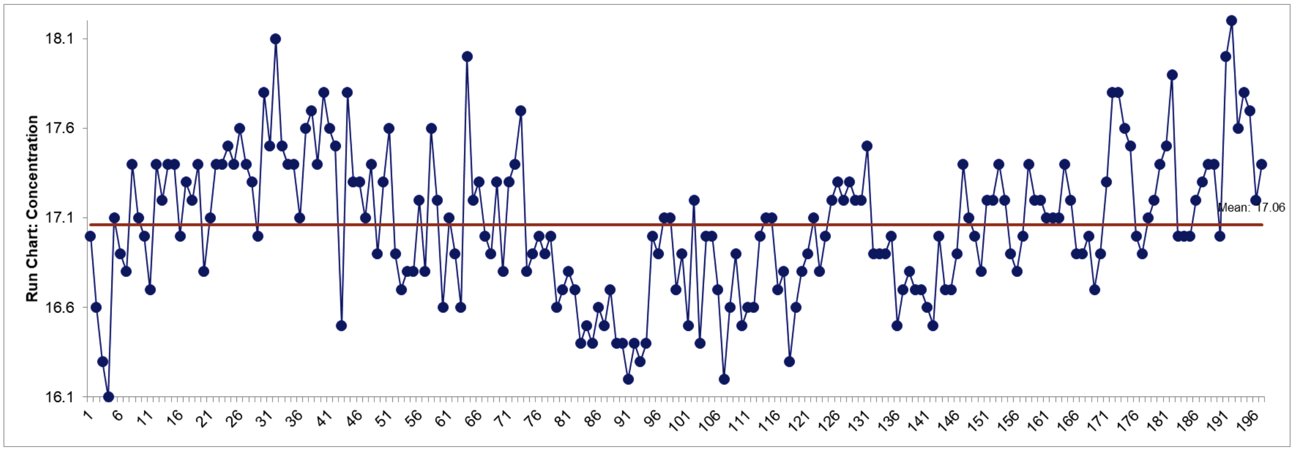
2. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.



Tip: **Show Mean** is selected for consistency with the use of the mean in forecasting. If you wish to use the **Nonparametric Runs Test**, then use the default **Show Median** to display the Median rather than the Mean.

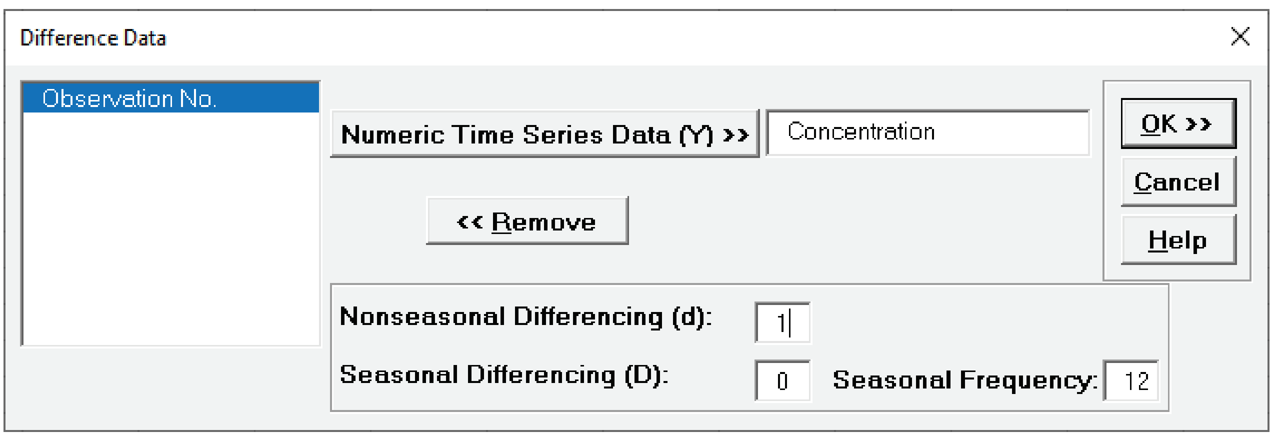
4. Click **OK**. A Run Chart of Concentration with Mean center line is produced.



5. Clearly, this is a non-stationary process with a “wandering mean”. Exponential smoothing can model this, but ARIMA requires a stationary process. We will now demonstrate how differencing can be used to make the process stationary. (Note that ARIMA Forecast will automatically take care of this).

6. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

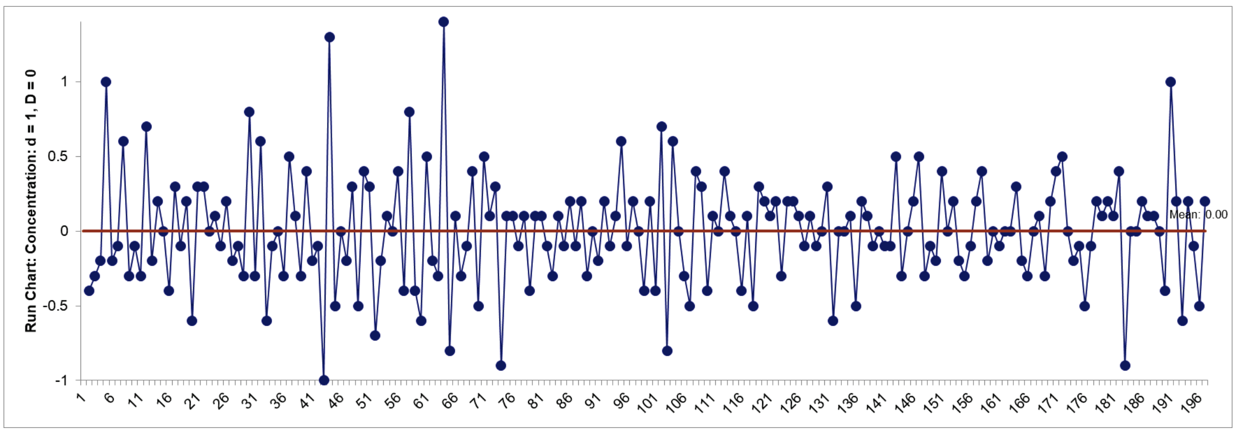
7. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Enter 1 for **Nonseasonal Differencing (d)**.



8. Click **OK**. A new sheet is created with the order d = 1 differenced data (Y_2 - Y_1, Y_3-Y_2, …).

Concentration	Concentration: d = 1, D = 0
17	
16.6	-0.4
16.3	-0.3
16.1	-0.2
17.1	1
16.9	-0.2
16.8	-0.1
17.4	0.6
17.1	-0.3
17	-0.1
16.7	-0.3
17.4	0.7

9. Redo the Run Chart for the differenced data:



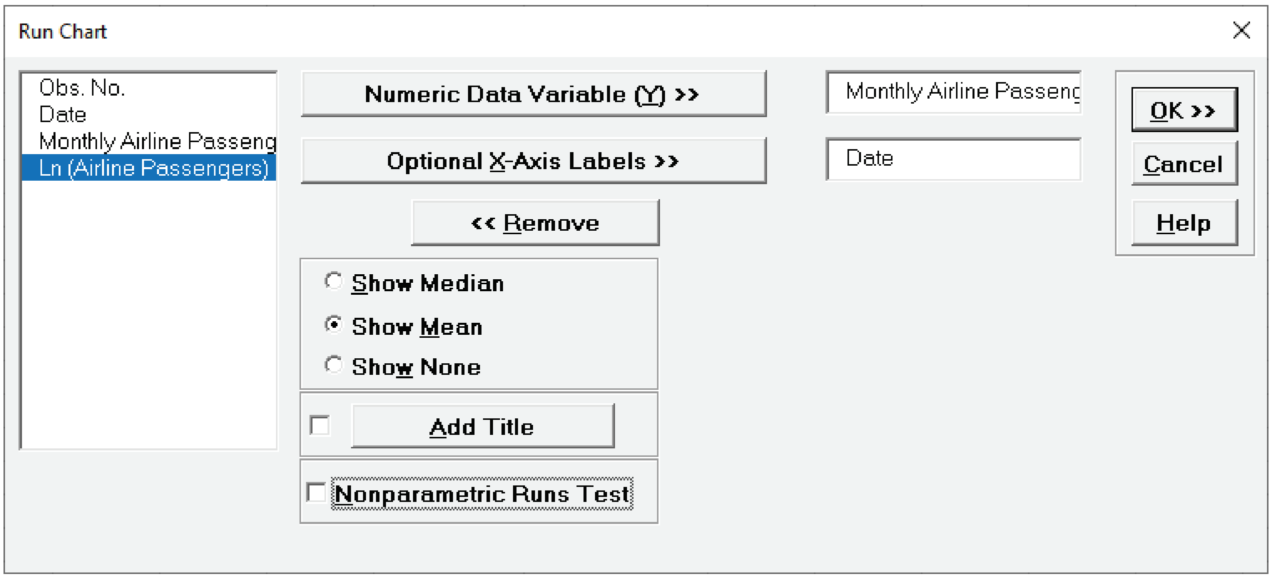
This now appears to be a stationary process and can be modeled using ARIMA. Save the workbook with the differenced data for use with Autocorrelation Plots.

Monthly Airline Passengers - Series G

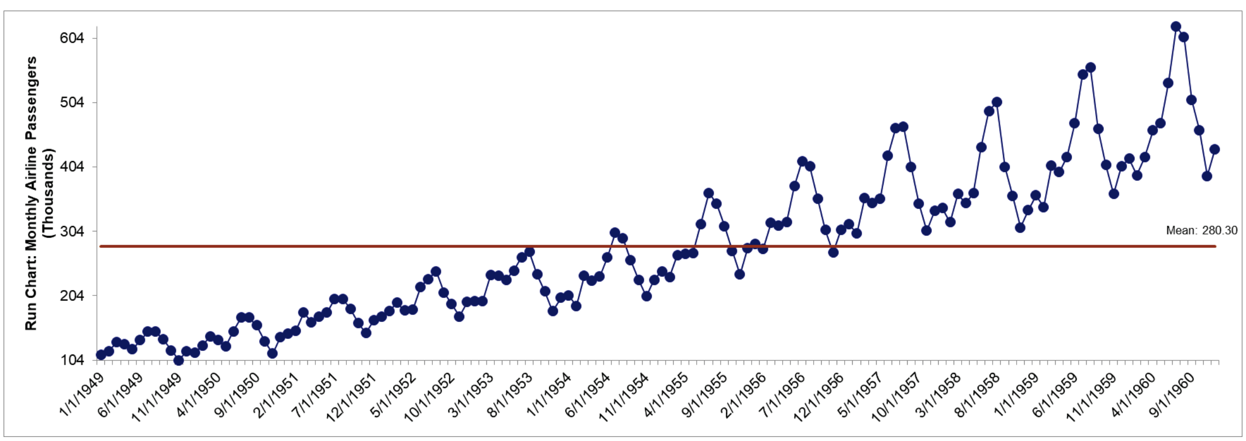
10. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960 and is one of the most popular datasets used in introductory time series forecasting.

11. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

12. Select *Monthly Airline Passengers (Thousands)*, click **Numeric Data Variable (Y) >>**; select *Date*, click **Optional X-Axis Labels >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.

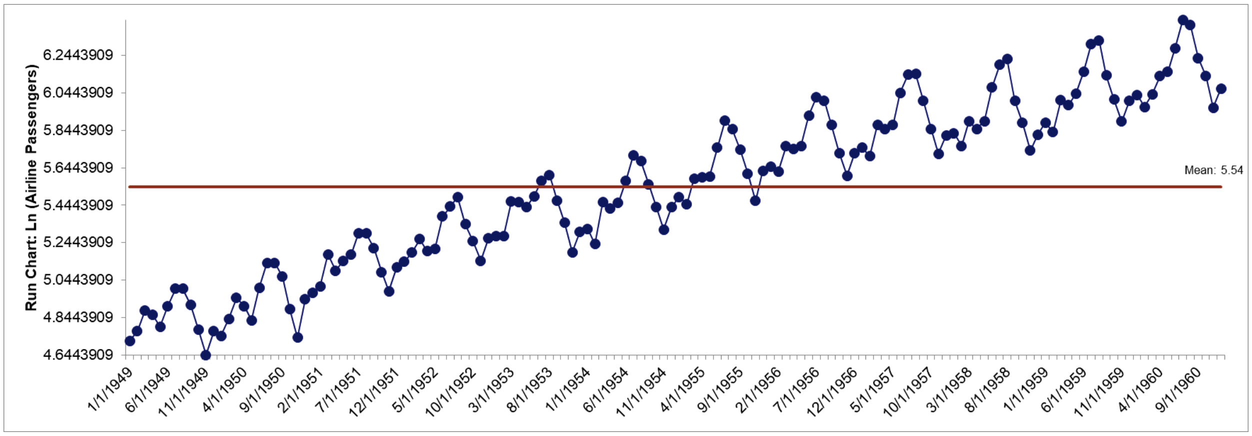


13. Click **OK**. A Run Chart of Monthly Airline Passengers with Mean center line is produced.



14. Here we see a strong positive trend as well as a monthly seasonal effect (yearly cycle of 12 months), with an increase in the variance over time. This non-stationarity in the variance can be addressed by using a Ln transformation of the data. Later, we will demonstrate the use of a Box-Cox Transformation to automatically determine the best transformation, but for now we will use the Ln(Airline Passenger) data.

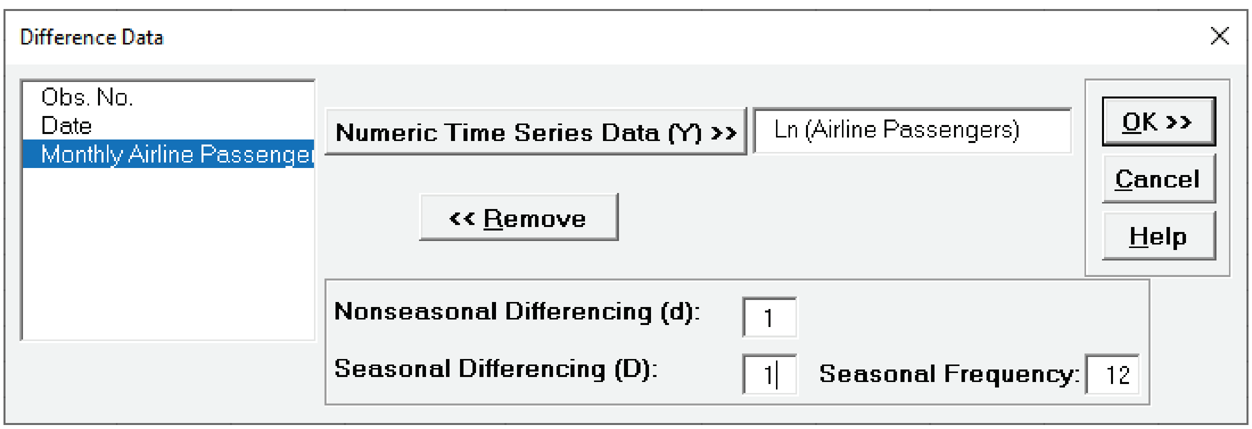
15. Click the **Sheet 1** tab. Redo the Run Chart for *Ln (Airline Passengers)*:



16. This addresses the change in variance over time, but still needs to be differenced in order to make it stationary for ARIMA. (Note that ARIMA Forecast will automatically take care of this).

17. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

18. Select *Ln(Airline Passengers)*, click **Numeric Time Series Data (Y) >>**. Enter 1 for **Nonseasonal Differencing (d)**; enter 1 for **Seasonal Differencing (D)**; **Seasonal Frequency** is specified as 12.

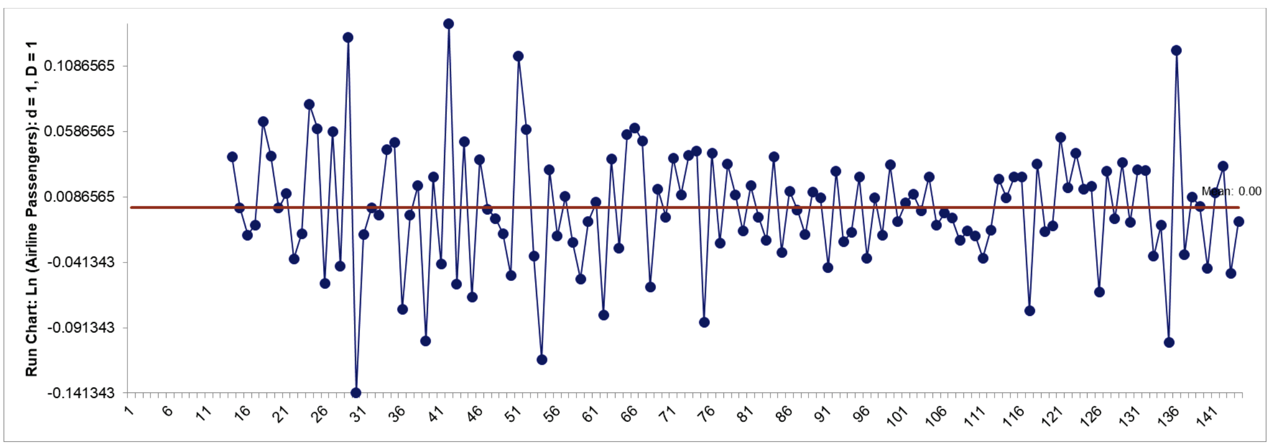


19. Click **OK**. A new sheet is created with the seasonal order D = 1 differenced data (Y_{13} - Y_1, Y_{14}-Y_2, ...) and nonseasonal order d=1 ($D_2 – D_1$, $D_3 – D_2$, ...).

Ln (Airline Passengers)	Ln (Airline Passengers): d = 1, D = 1
4.718498871	
4.770684624	
4.882801923	
4.859812404	
4.795790546	
4.905274778	
4.997212274	
4.997212274	
4.912654886	
4.779123493	
4.644390899	
4.770684624	
4.744932128	
4.836281907	0.039164025
4.94875989	0.000360685
4.905274778	-0.020495594
4.828313737	-0.012939182

Tip: For an example of manual seasonal and nonseasonal differencing in Excel see https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/Class10notes.htm.

20. Redo the Run Chart for the differenced data:



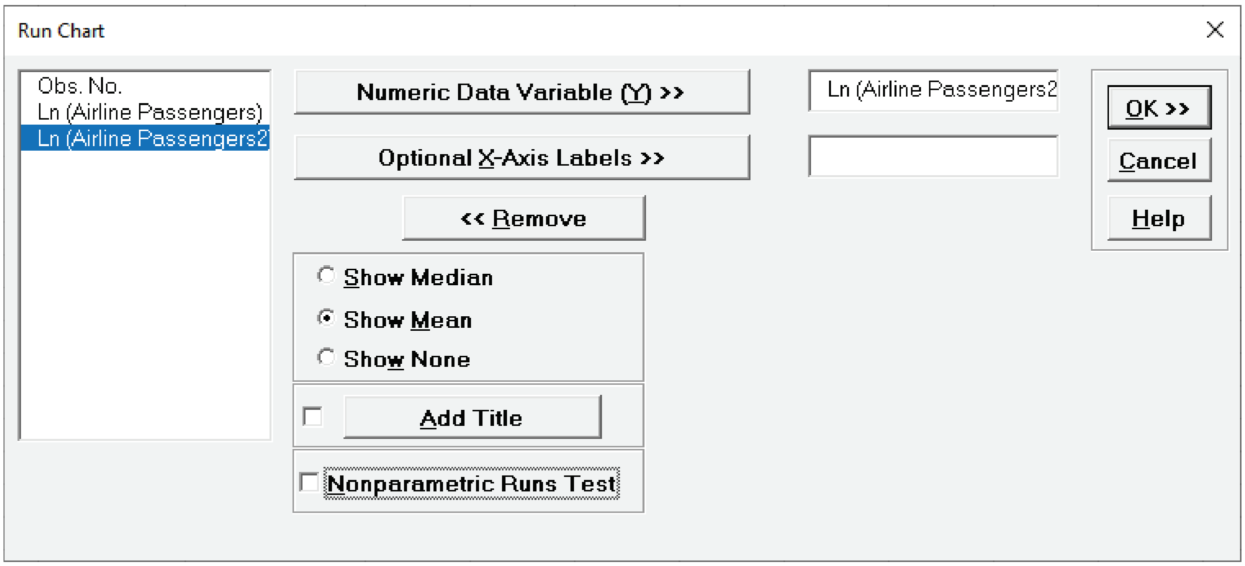
This now appears to be a stationary process and can be modeled using ARIMA. Save the workbook with the differenced data for use with Autocorrelation Plots.

Monthly Airline Passengers – Missing Values

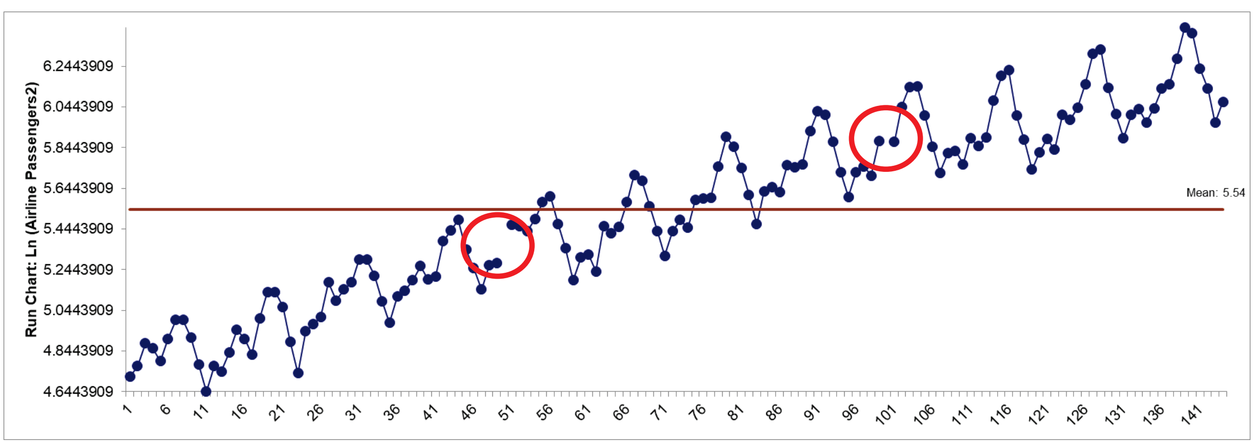
21. Open **Monthly Airline Passengers – Missing Values.xlsx** (**Sheet 1** tab). Ln(Airline Passengers2) have missing values at observations 50 and 100.

22. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

23. Select *Ln(Airline Passengers2)*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.



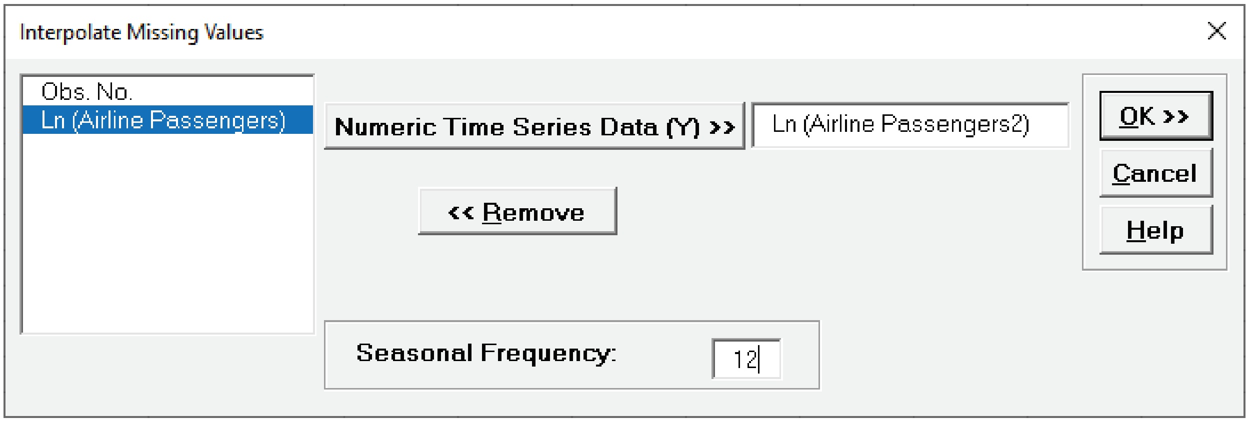
24. Click **OK**. A Run Chart of Ln(Airline Passengers2) with Mean center line is produced.



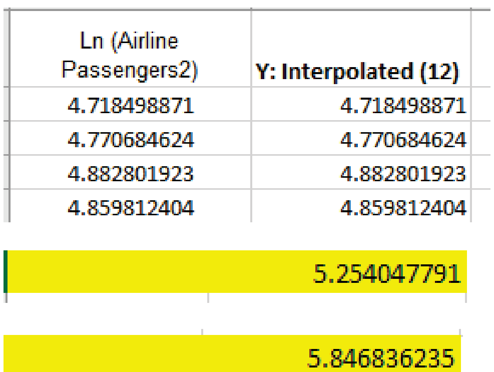
25. The missing values result in a broken line on the run chart.

26. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Interpolate Missing Values.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

27. Select *Ln(Airline Passengers2)*, click **Numeric Time Series Data (Y) >>**. **Seasonal Frequency** is specified as 12.



28. Click **OK**. A new sheet is created with the seasonally adjusted linear interpolated values highlighted in yellow.



Ln (Airline Passengers2)	**Y: Interpolated (12)**
4.718498871	4.718498871
4.770684624	4.770684624
4.882801923	4.882801923
4.859812404	4.859812404
	5.254047791
	5.846836235

Y50 is estimated as 5.254 (versus original value of 5.278). Y100 is estimated as 5.847 (versus original value of 5.852).

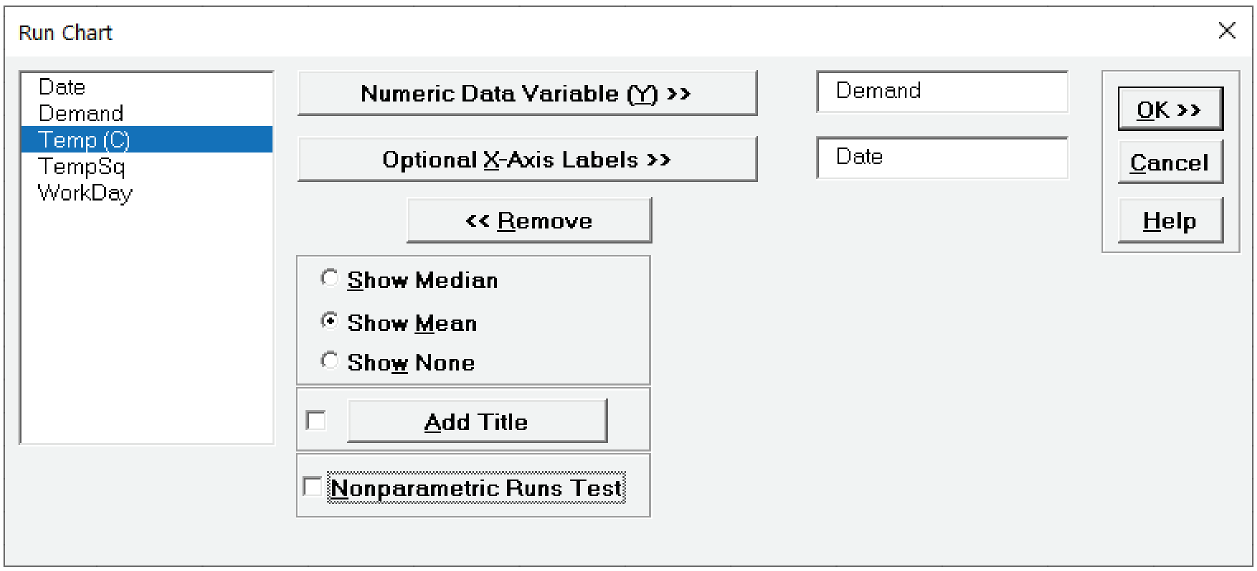
Note, while there is robustness against some missing values, if the number of missing values is large, then model estimation and forecast accuracy will be degraded.

Daily Electricity Demand with Predictors – ElecDaily

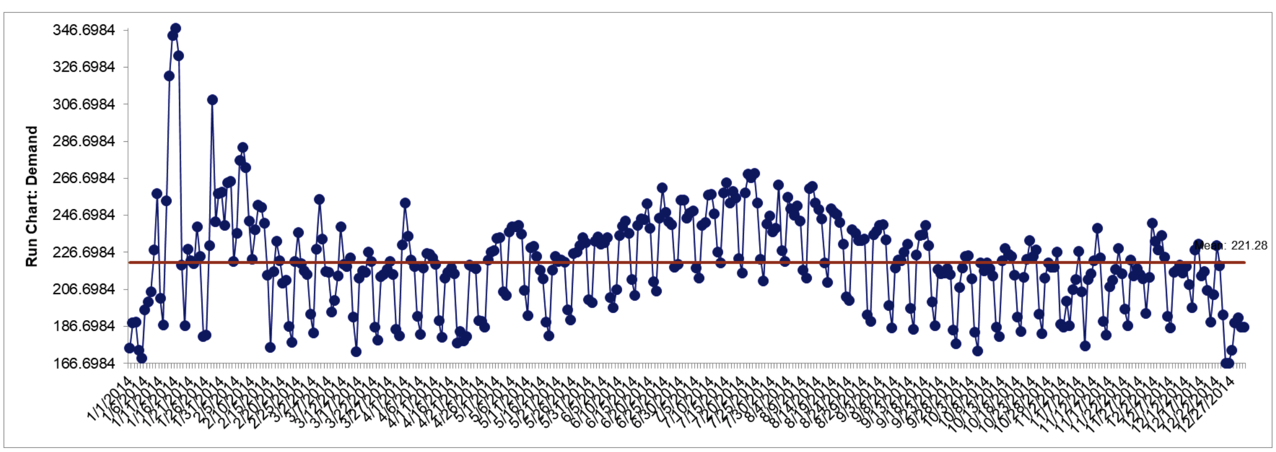
29. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (Hyndman, fpp2). Temp (C) is the maximum daily temperature in degrees Celsius for the city of Melbourne. TempSq is Temperature squared. WorkDay takes on the value 1 on work days and 0 otherwise. This data has a seasonal frequency = 7 (observations per week).

30. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

31. Select *Demand*, click **Numeric Data Variable (Y) >>**; select *Date*, click **Optional X-Axis Labels >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.

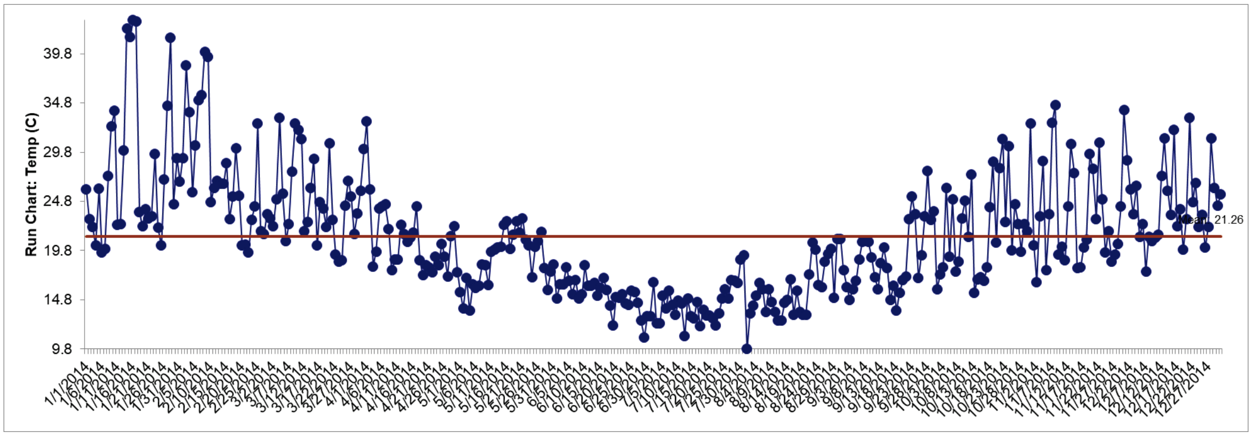


32. Click **OK**. A Run Chart of Electricity Demand with Mean center line is produced.



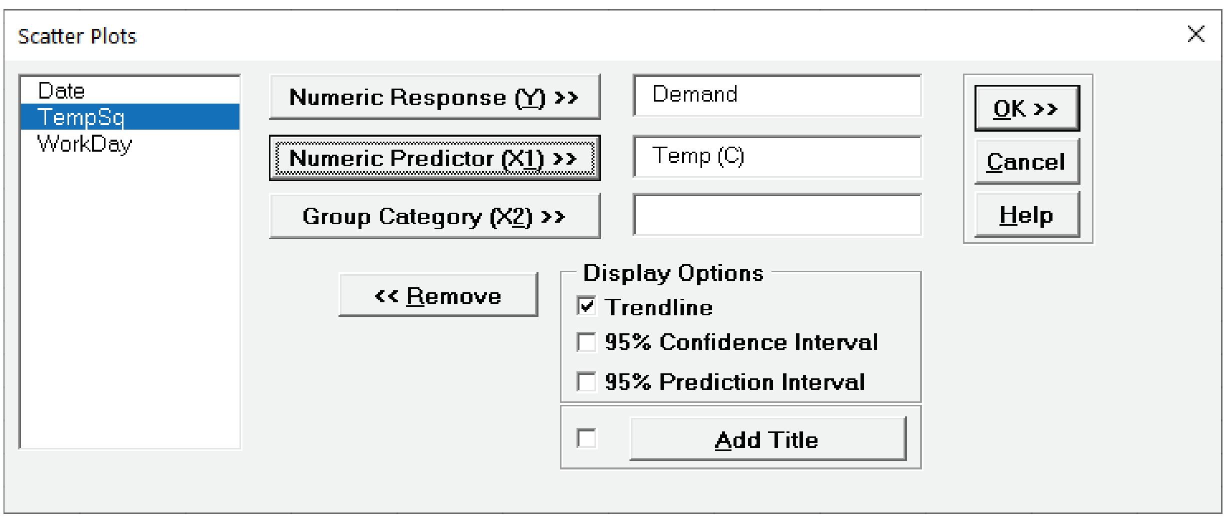
33. We can see the daily seasonality with 7-day cycle, but there is also a summer winter effect, with the peak demand occurring mid-January, which is summer in Australia. Later we will use a Spectral Density Plot to confirm the seasonal frequency value of 7.

34. Click the **Sheet 1** tab. Redo the Run Chart for Temp (C):

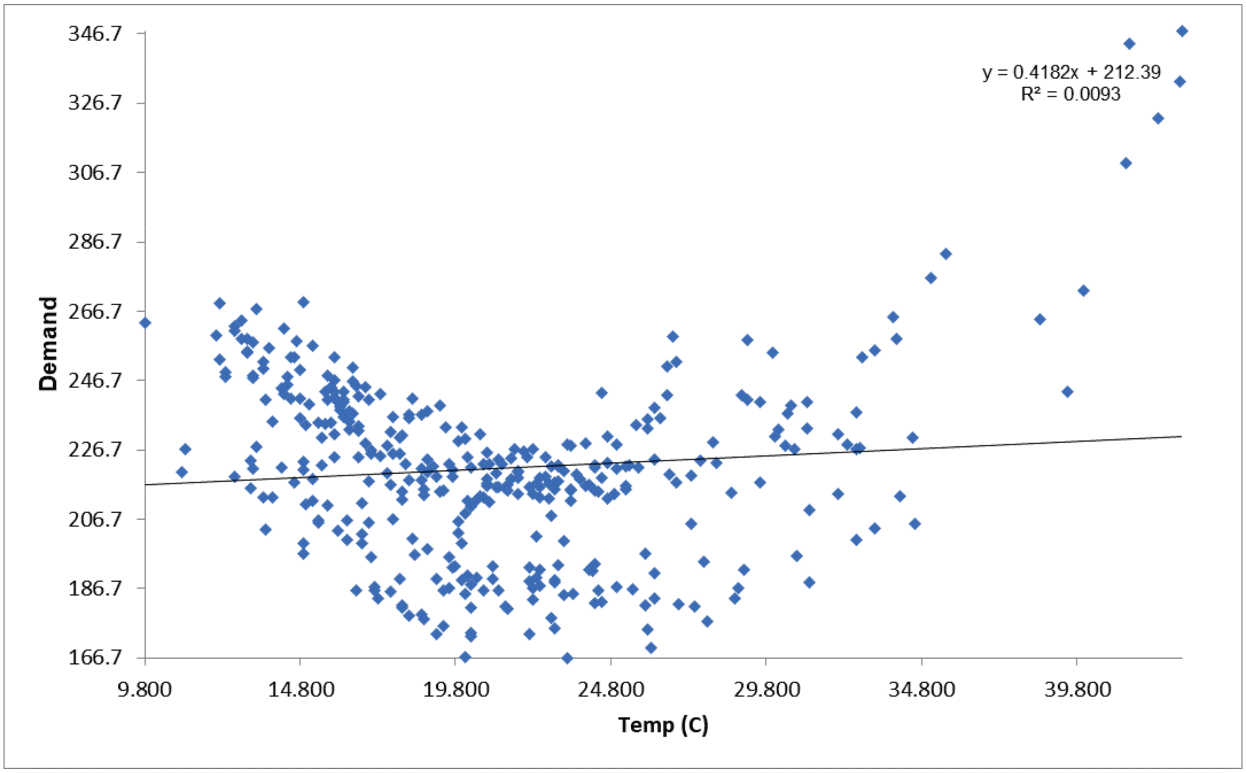


35. Now we will explore the relationship between Demand and Temperature. Click the **Sheet 1** tab. Click **SigmaXL > Graphical Tools > Scatterplots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

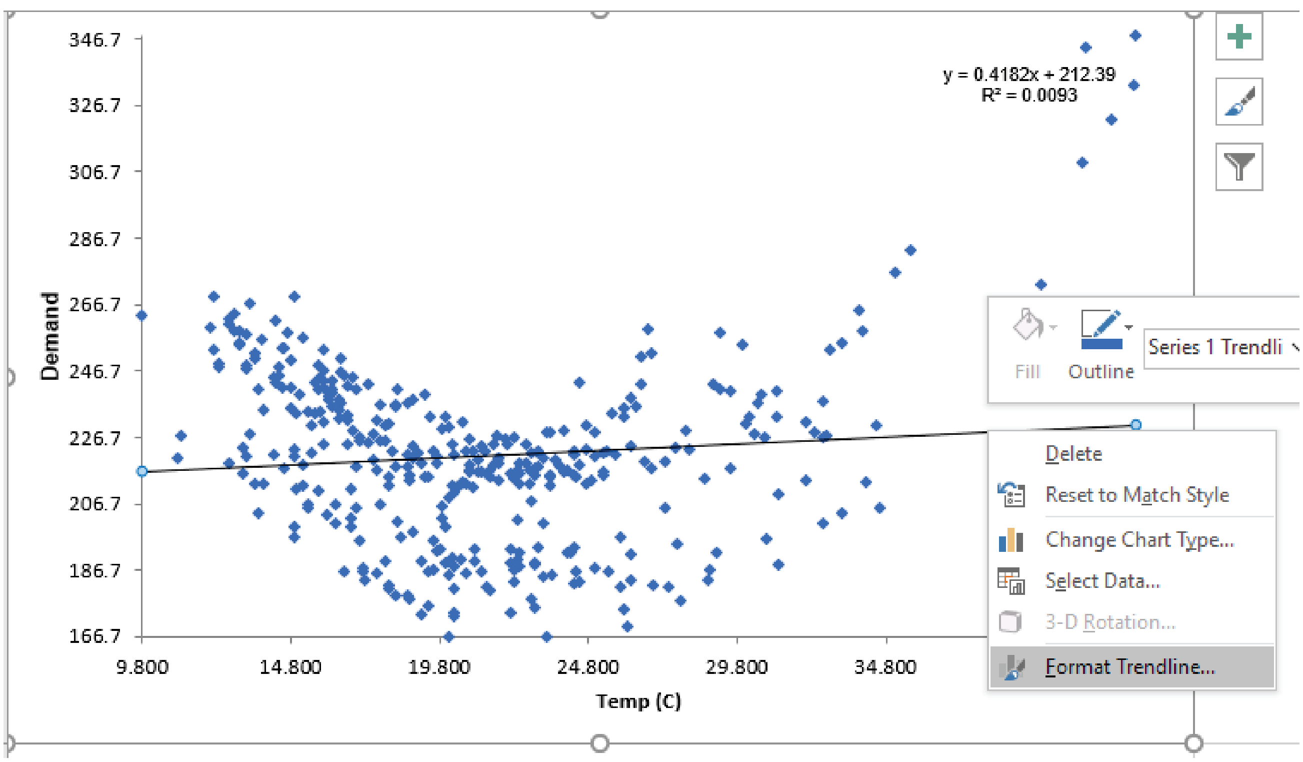
36. Select *Demand*, click **Numeric Response (Y) >>**; select *Temp (C)*, click **Numeric Predictor (X1) >>**. Check **Trendline**.



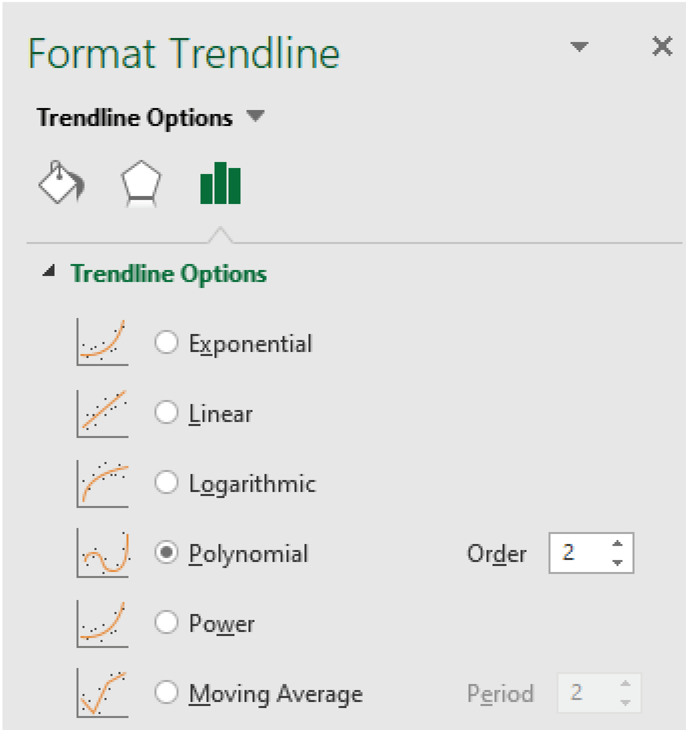
37. Click **OK**. A Scatterplot of Electricity Demand versus Temperature is produced.



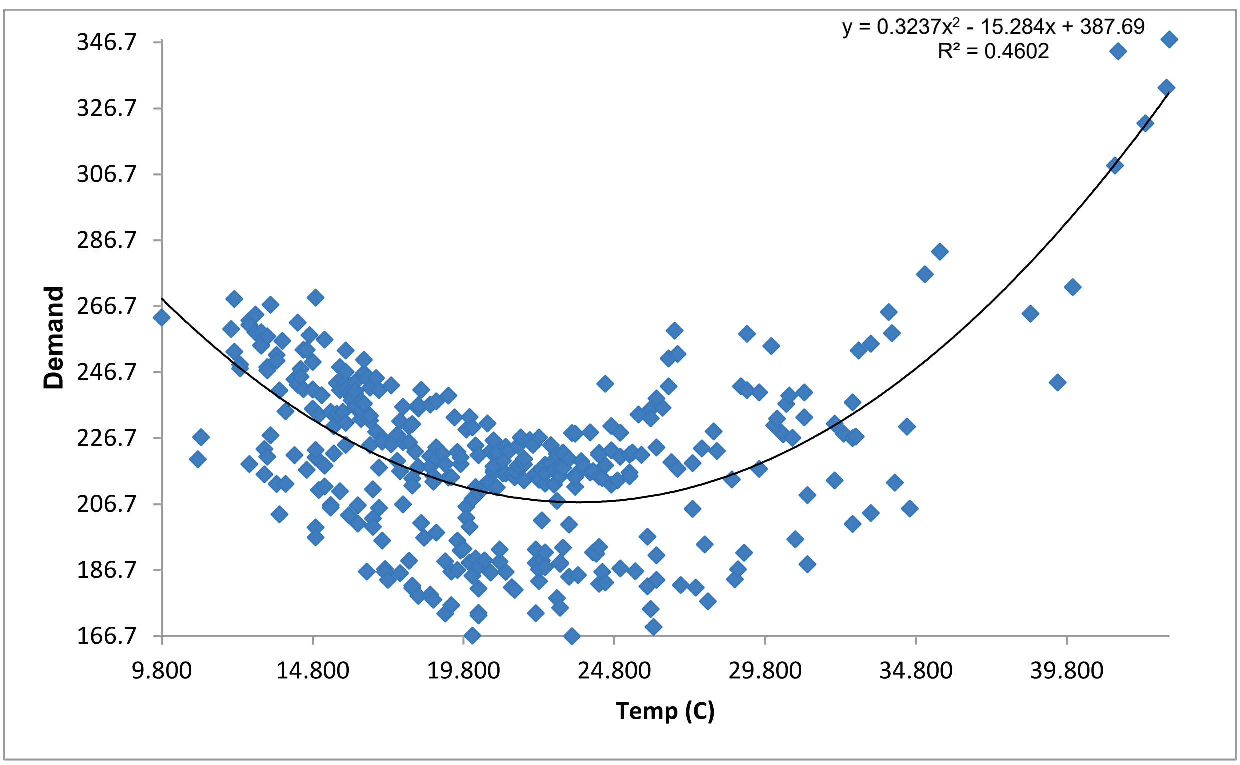
38. This shows a quadratic relationship: high temperatures in the summer cause electricity demand for air conditioning, low temperatures in the winter cause demand for heating. We will modify the trendline in Excel to a quadratic fit. Click on the Trendline, right click and select Format Trendline as shown:



The **Format Trendline** options are given. Select **Polynomial** with **Order** 2 as shown.



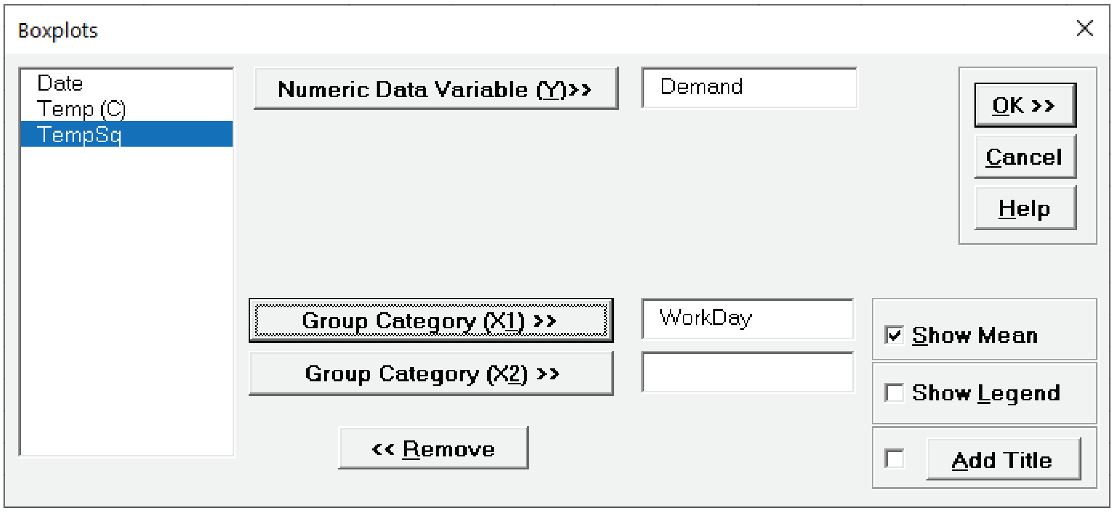
39. The Trend Line is now a quadratic function as shown:



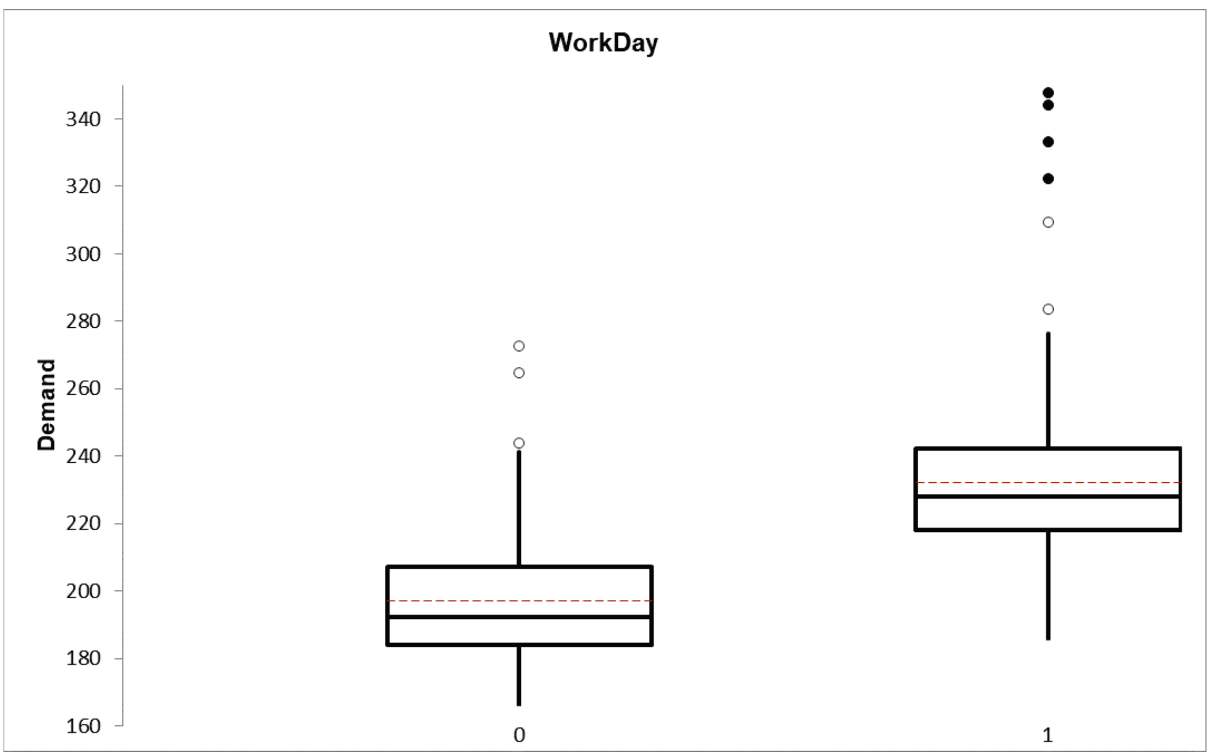
40. TempSq (Temperature Squared) has been added to the data so that it may be included when we do ARIMA forecast with predictors.

41. Now we will explore the relationship between Demand and WorkDay. Click the **Sheet 1** tab. Click **SigmaXL > Graphical Tools > Boxplots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

42. Select *Demand*, click **Numeric Data Variable (Y) >>**; select *WorkDay*, click **Group Category (X1) >>**. Check **Show Mean**.



43. Click **OK**. A Boxplot of Electricity Demand versus Work Day is produced.



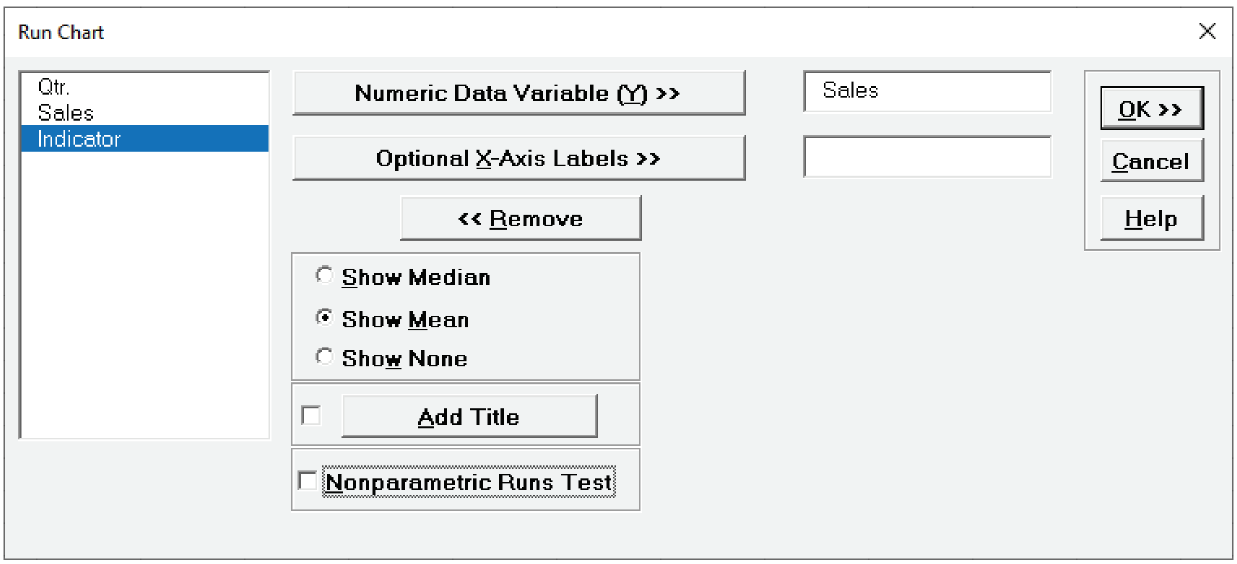
As expected, a work day (1) has higher electricity demand that a non work day (0). Later, Temp (C), TempSq and WorkDay will be included as predictors in the ARIMA forecast model.

Sales with Indicator - Modified Series M

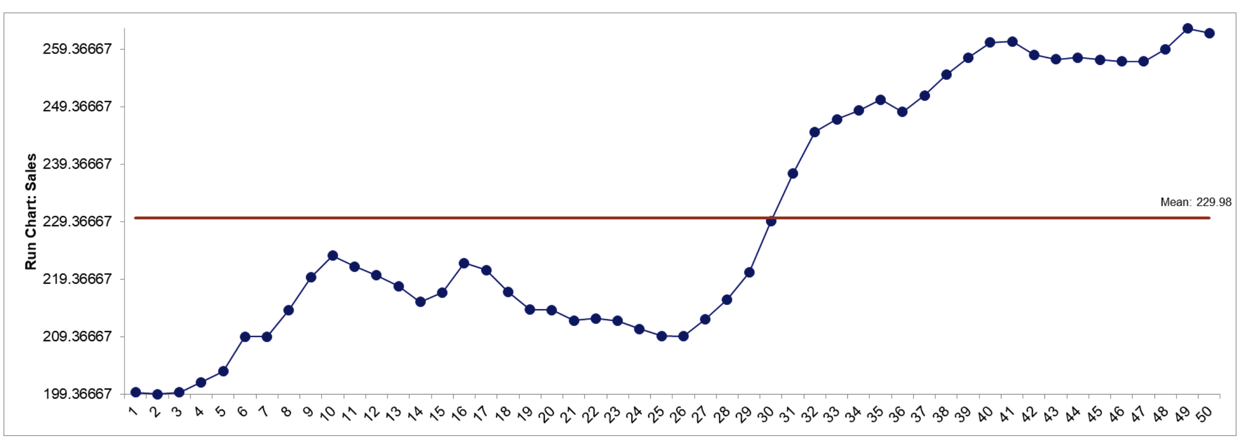
44. Open **Sales with Indicator - Modified Series M.xlsx**. (**Sheet 1** tab). This is modified Series M data from Box and Jenkins. Originally, a set of 150 monthly corporate sales values along with a leading indicator, the data was modified by converting it to quarterly values by averaging every three months, so 50 quarters. This was done in order to simplify the analysis of the leading indicator. Although the data was monthly and summarized to quarterly, it will be treated as nonseasonal, as done in Box and Jenkins.

45. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

46. Select *Sales*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.

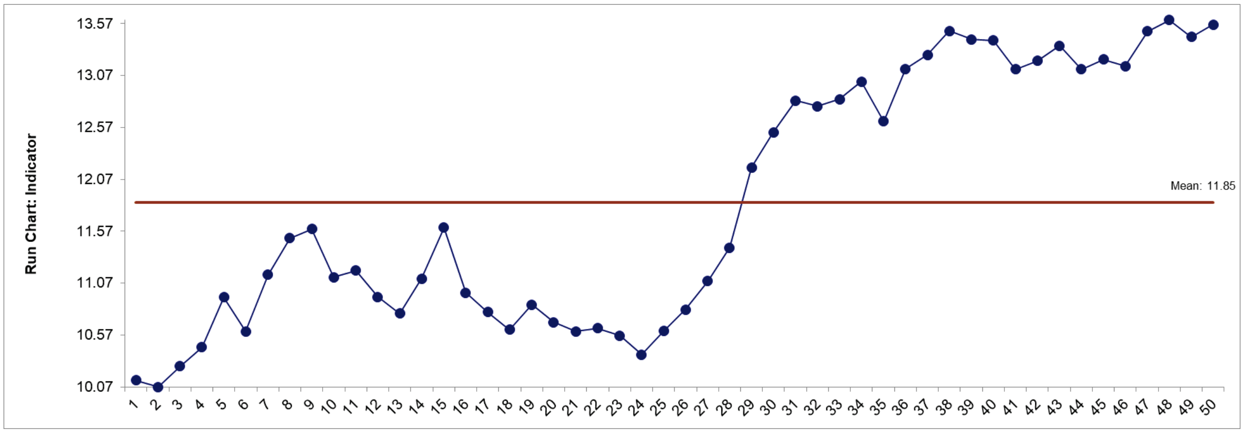


47. Click **OK**. A Run Chart of Sales with Mean center line is produced.



As discussed above, there is no obvious seasonality in the data. Later we will use a Spectral Density Plot to confirm the seasonal frequency value of 1.

48. Click the **Sheet 1** tab. Redo the Run Chart for Indicator:



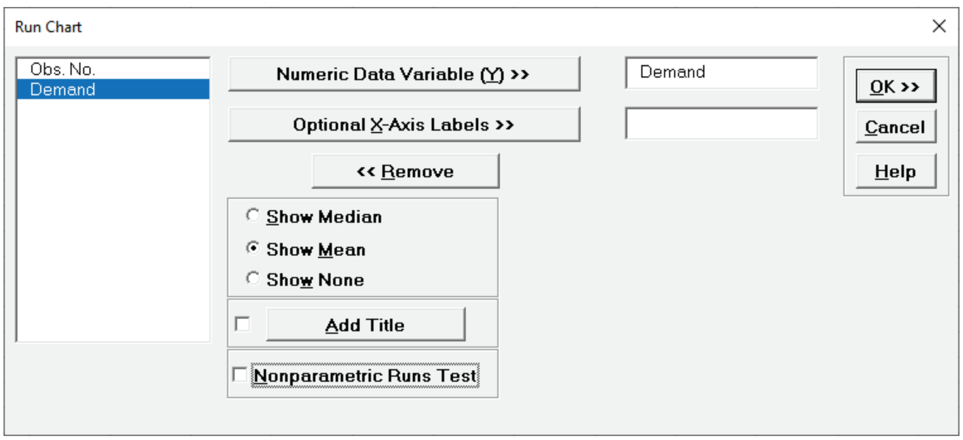
Clearly there is a close relationship between Sales and Indicator. We will analyze this relationship later with the Cross Correlation Function (CCF) Plot and with an ARIMA forecast model that includes Indicator as a predictor.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

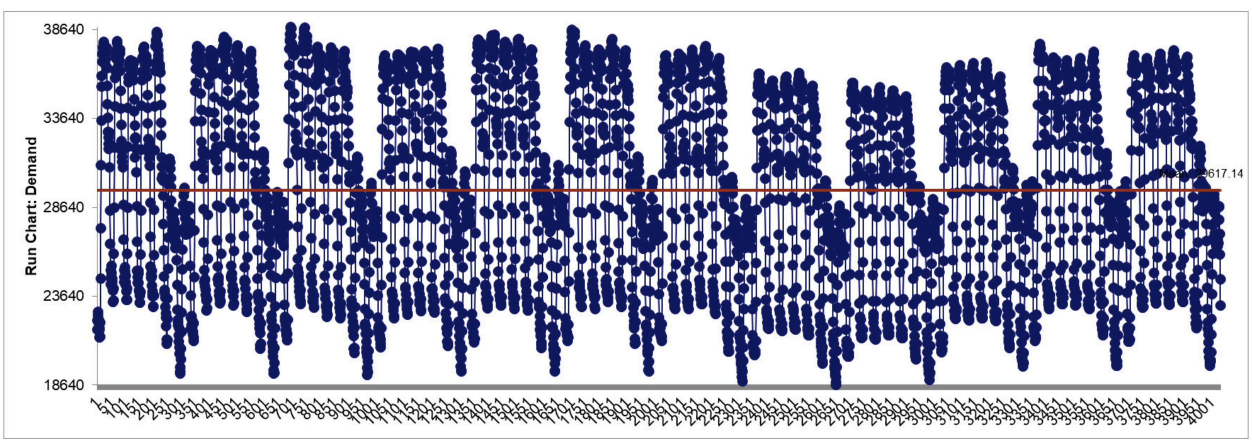
49. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations. Predictors such as temperature and work day are not available for this data. Reference: Taylor, J.W. (2003) Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, **54**, 799-805.

50. Click **SigmaXL > Time Series Forecasting > Run Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

51. Select *Demand*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test**.



52. Click **OK**. A Run Chart of Electricity Demand with Mean center line is produced.



Here we see the strong half-hourly seasonal effect with a daily cycle of 48 observations and a weekly cycle of 48*7 = 336 observations. Since this data has multiple frequencies (and frequencies > 24), multiple seasonal decomposition (MSD) will be required to model it. Later, we will use a Spectral Density Plot to confirm the multiple seasonal frequency values of 48 and 336

Autocorrelation (ACF/PACF) Plots

Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between lagged values of data. A plot of the data vs. the same data at lag k may show a positive or negative trend. If the slope is positive, the autocorrelation is positive; if there is a negative slope, the autocorrelation is negative. We will use the lag utility and scatterplots to demonstrate this for 3 lags using the Series A data.

ACF denotes the AutoCorrelation Function plot (sometimes called a Correlogram). PACF denotes the Partial AutoCorrelation Function plot. For further details and formulas, see Appendix: **Autocorrelation (ACF), Partial Autocorrelation (PACF) and Cross Correlation (CCF).**

We will use the ACF and PACF to get a general idea of what models should be used, but let the automatic algorithms do the heavy lifting of determining which model is optimal. Later, ACF on model residuals are used to assist in determining how well a model fits the autocorrelation structure, ideally with the residuals showing no statistically significant autocorrelation.

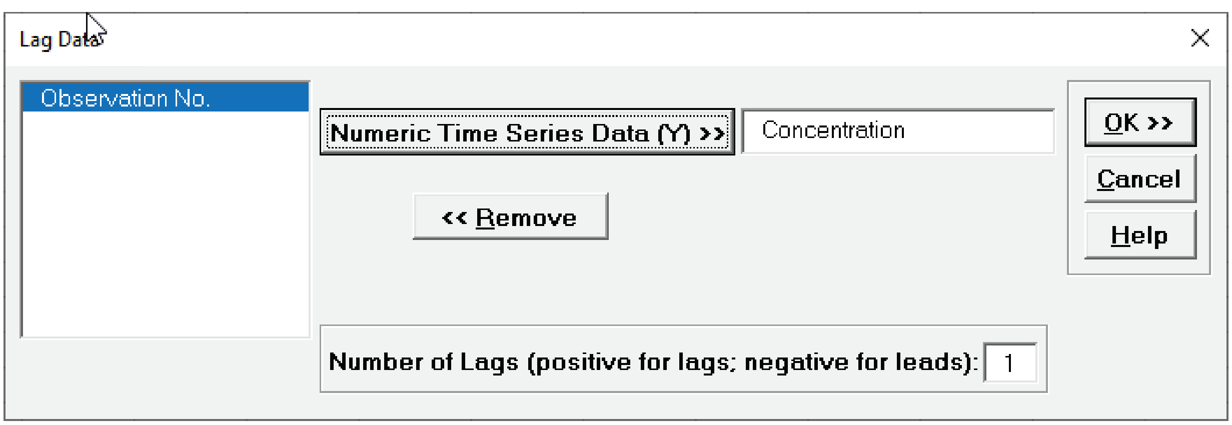
We will also examine how differencing results in stationarity with the ACF and PACF plots.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

2. Click **SigmaXL > Time Series Forecasting > Utilities > Lag Data**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Use the default **Number of Lags** = 1.



4. Click **OK**. A new sheet is created with the Lag 1 data.

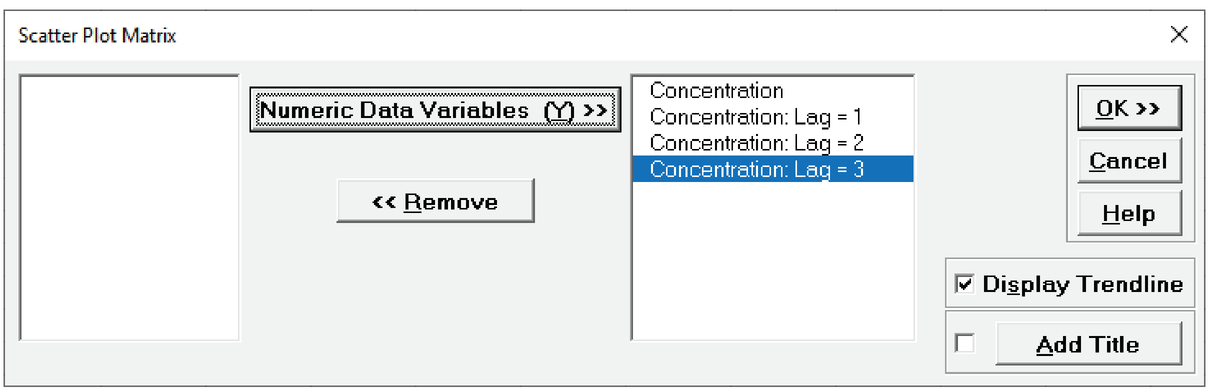
Concentration	Concentration: Lag = 1
17	
16.6	17
16.3	16.6
16.1	16.3
17.1	16.1
16.9	17.1

5. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Enter **Lag** = 2. Click **OK**. Repeat for **Lag** = 3. The Lag columns are appended as shown:

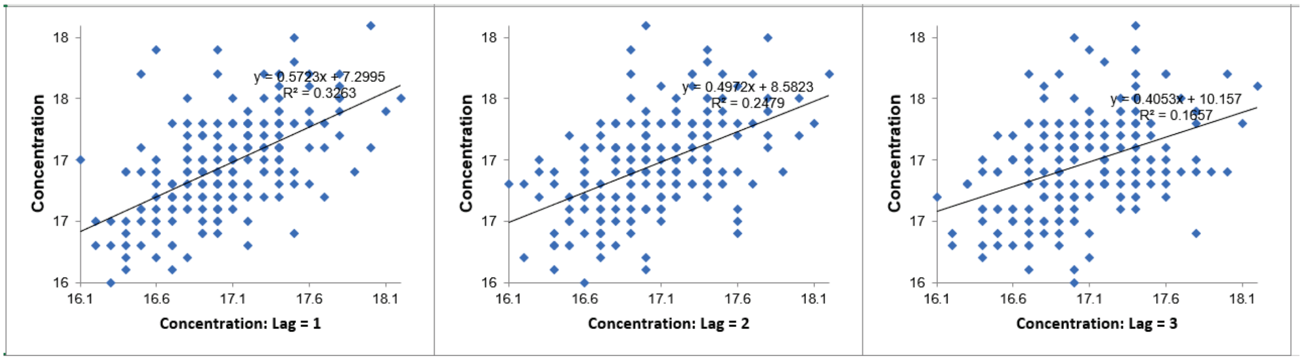
Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
17			
16.6	17		
16.3	16.6	17	
16.1	16.3	16.6	17
17.1	16.1	16.3	16.6
16.9	17.1	16.1	16.3
16.8	16.9	17.1	16.1

6. Click **SigmaXL > Graphical Tools > Scatter Plot Matrix**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

7. Select all four columns, click **Numeric Data Variables (Y) >>**. Check **Display Trendline**.

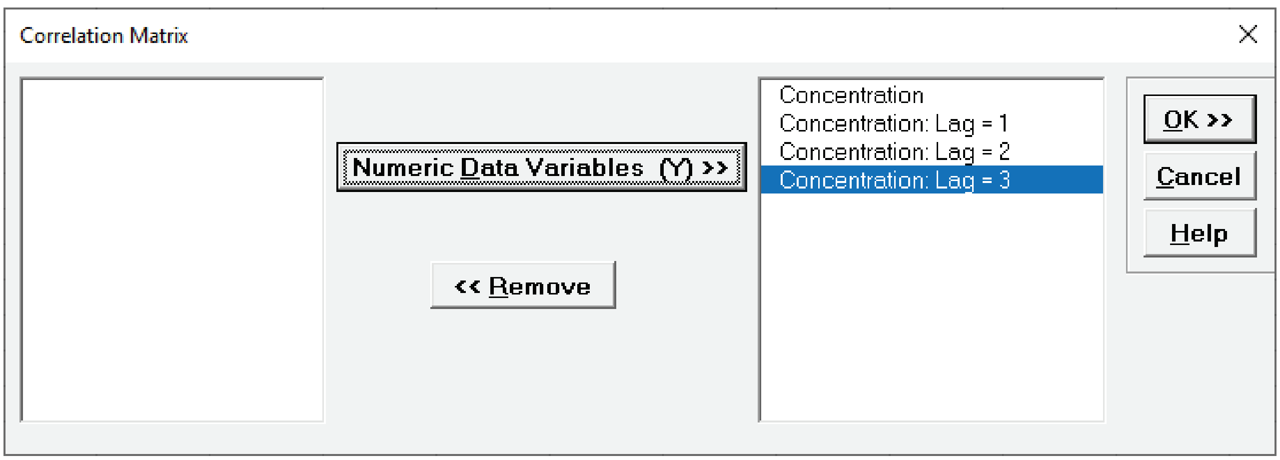


8. Click **OK**. The Scatterplots for Lag 1 to 3 are produced, showing a positive trend, so positive autocorrelation.



9. Click the **Lag Data** sheet tab. Click **SigmaXL > Statistical Tools > Correlation Matrix**. Ensure that the entire data table is selected. If not, check Use Entire Data Table. Click Next.

10. Select all four columns, click **Numeric Data Variables (Y) >>**.



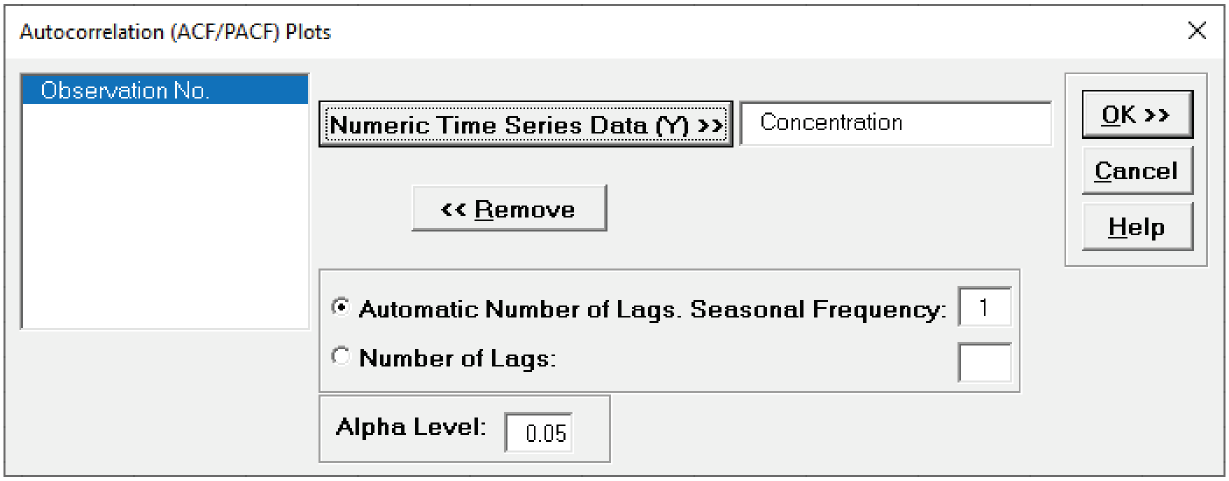
11. Click **OK**. The Correlation Coefficient Matrix is produced.

Pearson Correlations	Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
Concentration	1	0.5713	0.4979	0.4070
Concentration: Lag = 1		1	0.5710	0.4952
Concentration: Lag = 2			1	0.5721
Concentration: Lag = 3				1

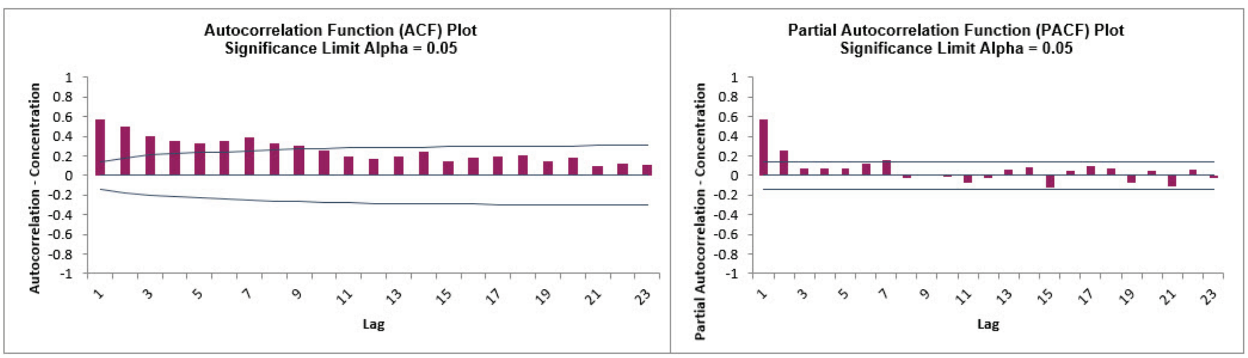
Lag 1 correlation is .57, Lag 2 is .5, and Lag 3 is .41. The correlations in red (and bolded P-Values) indicates that the correlations are significant.

12. Now we will create the ACF/PACF plots. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

13. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Use the default **Automatic Number of Lags. Seasonal Frequency** = 1. **Alpha Level** = 0.05.



14. Click **OK**. The ACF and PACF plots are produced.

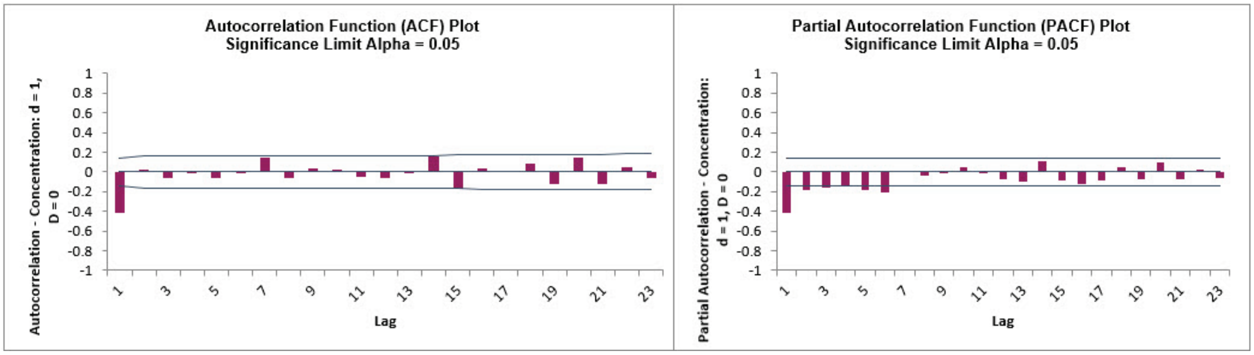


Hover the mouse cursor on the Lag 1 ACF bar to view the correlation value = .57, Lag 2 = 0.5, and Lag 3 = .4. The values are approximately the same as those obtained manually (with some minor differences in how the correlations are calculated).

The lags that exceed the significance limit lines are statistically significant at alpha = 0.05, so there is an autocorrelation structure in the data which can be modeled using Exponential Smoothing or ARIMA.

15. The slow decrease in the ACF as the lags increase is due to the non-stationarity, so we will now compare to the differenced data. Select the **Difference Data** sheet (or recreate using **SigmaXL > Time Series Forecasting > Utilities > Difference Data**, with **Nonseasonal Differencing (d)** = 1).

16. Redo the ACF/PACF Plots for the differenced data:



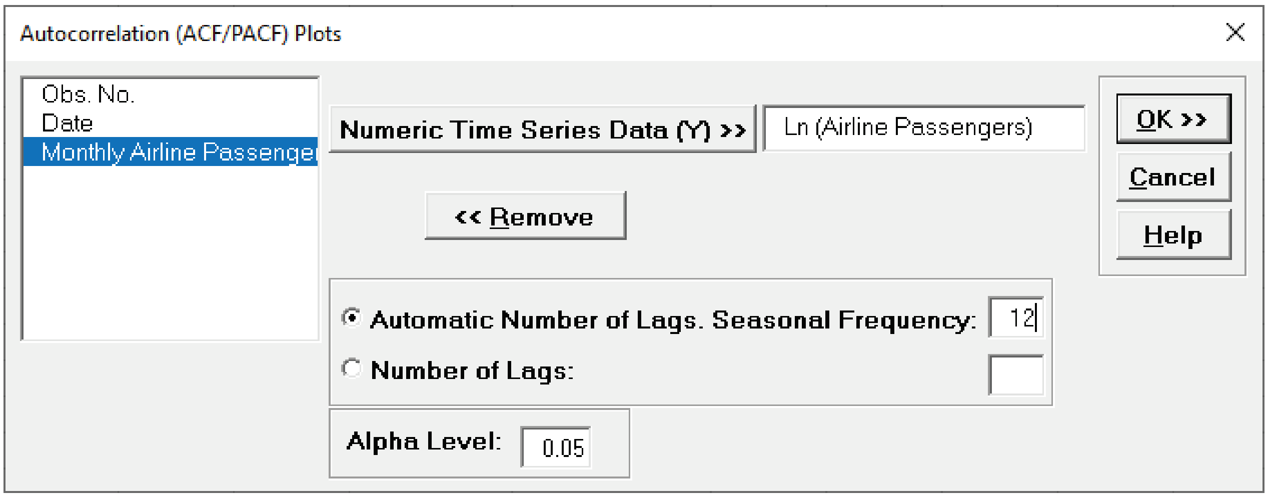
After differencing Lag 1 now shows a negative autocorrelation, but Lag 2 and following have insignificant autocorrelations and small partial autocorrelations. This is in agreement with the results in Box and Jenkins (2016, pp.185-186) who also suggest that, after differencing, the model might be a moving average of order 1. We will examine this later in ARIMA Forecasting.

Monthly Airline Passengers - Series G

17. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960.

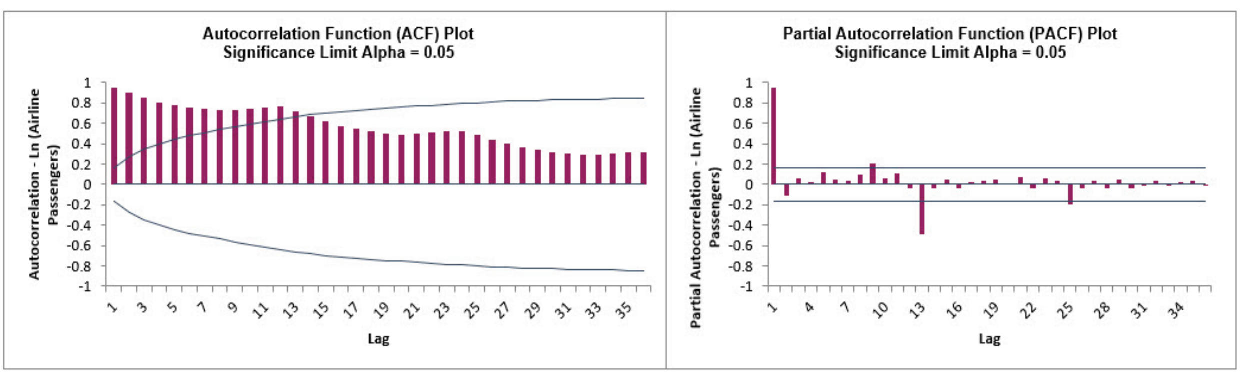
18. Click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

19. Select *Ln(Airline Passengers)*, click **Numeric Time Series Data (Y) >>**. Select **Automatic Number of Lags.** Specify **Seasonal Frequency** = 12 and **Alpha Level** = 0.05.



The automatic number of lags will be (a minimum of) Seasonal Frequency * 3 = 36 allowing us to see seasonal patterns in the ACF.

20. Click **OK**. The ACF and PACF plots are produced.

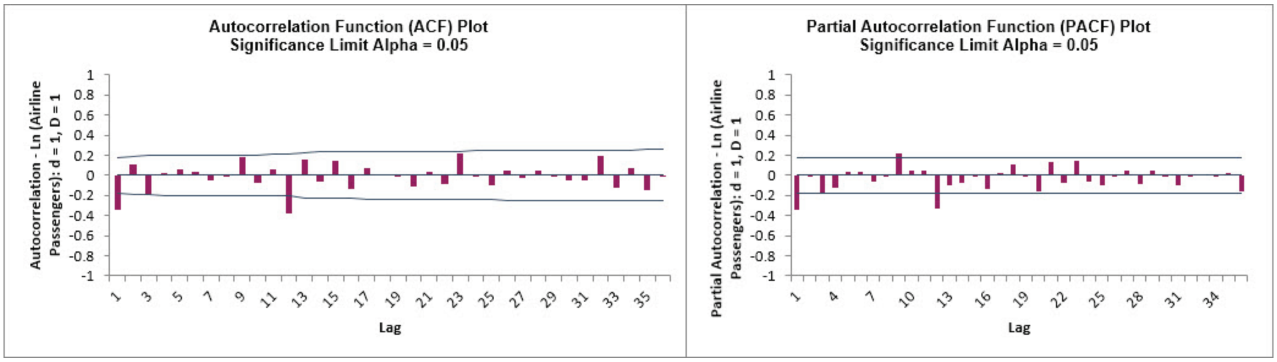


The slow decrease in the ACF as the lags increase is due to the trend, while the “scalloped” shape is due to the seasonality.

21. We will now compare to the differenced data for Ln(Airline Passengers). Select the **Difference Data** sheet (or recreate using **SigmaXL > Time Series Forecasting > Utilities > Difference Data**,

with **Nonseasonal Differencing (d)** = 1 and **Seasonal Differencing (D)** = 1 with **Seasonal Frequency** = 12).

22. Redo the ACF/PACF Plots for the differenced data:



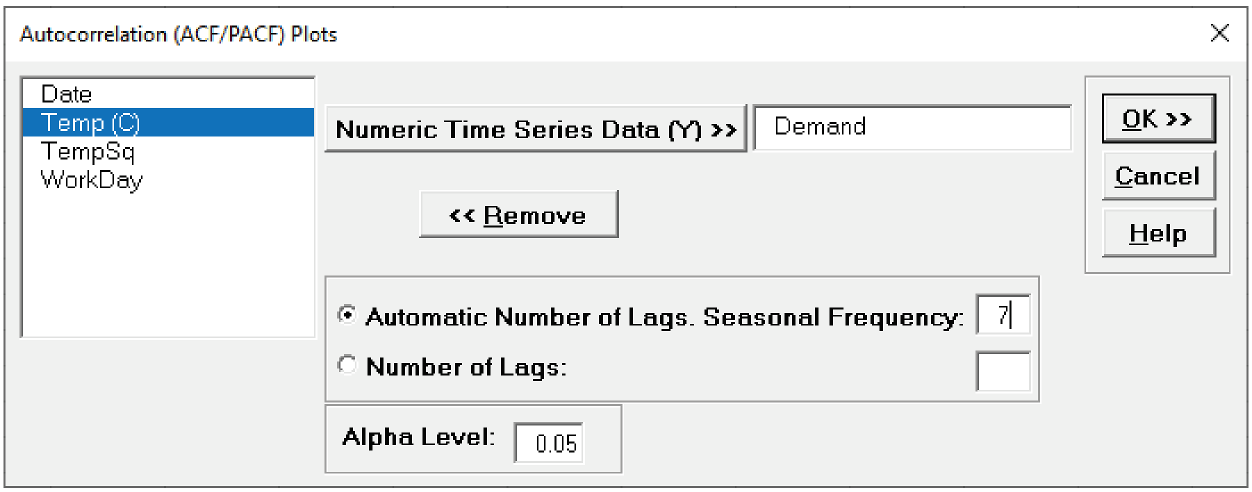
After nonseasonal and seasonal differencing, Lag 1 now shows a negative autocorrelation, but Lag 2 and following have mostly insignificant autocorrelations and partial autocorrelations. Lag 12 is due to the seasonality. This is in agreement with the results in Box and Jenkins (2016, Chapter 9, "Analysis of Seasonal Time Series", p. 319) who also suggest that, after differencing, the model might be a moving average of order 1 and seasonal moving average of order 1. We will examine this later in ARIMA Forecasting.

Daily Electricity Demand with Predictors – ElecDaily

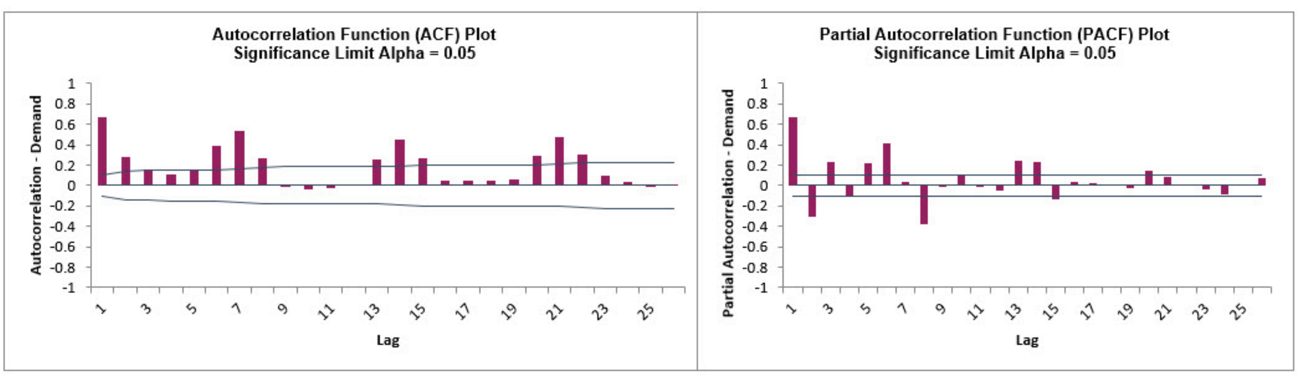
23. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (Hyndman, fpp2). This data has a seasonal frequency = 7 (observations per week).

24. Click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

25. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Select **Automatic Number of Lags.** Specify **Seasonal Frequency** = 7 and **Alpha Level** = 0.05.



26. Click **OK**. The ACF and PACF plots are produced.



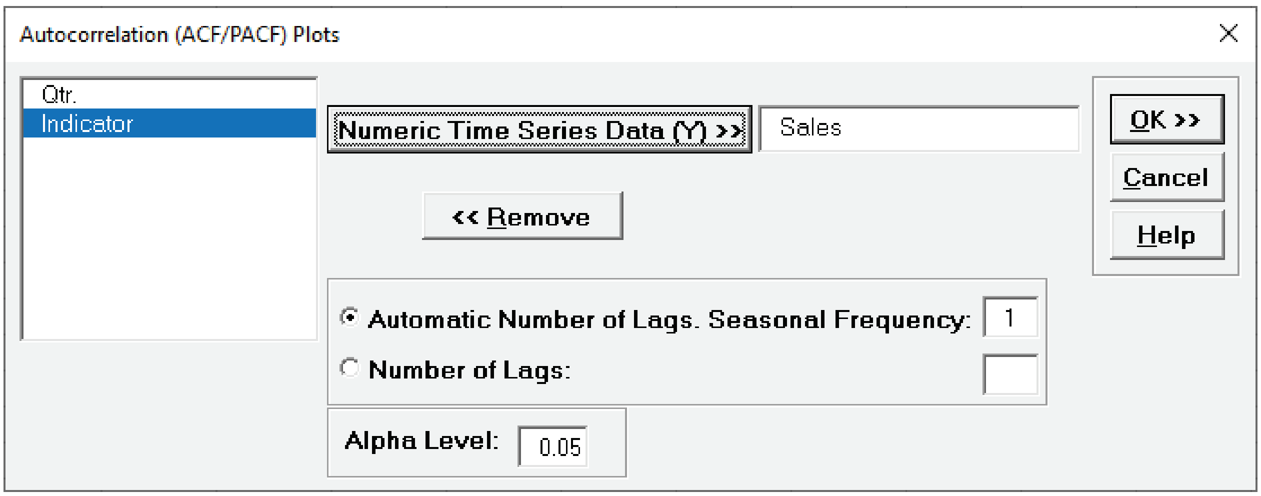
Here we can see the 7-day seasonal pattern. We will not do manual differencing for this example, but the automatic ARIMA algorithm will do a nonseasonal difference (d) = 1 and leave the seasonal differencing (D) = 0.

Sales with Indicator - Modified Series M

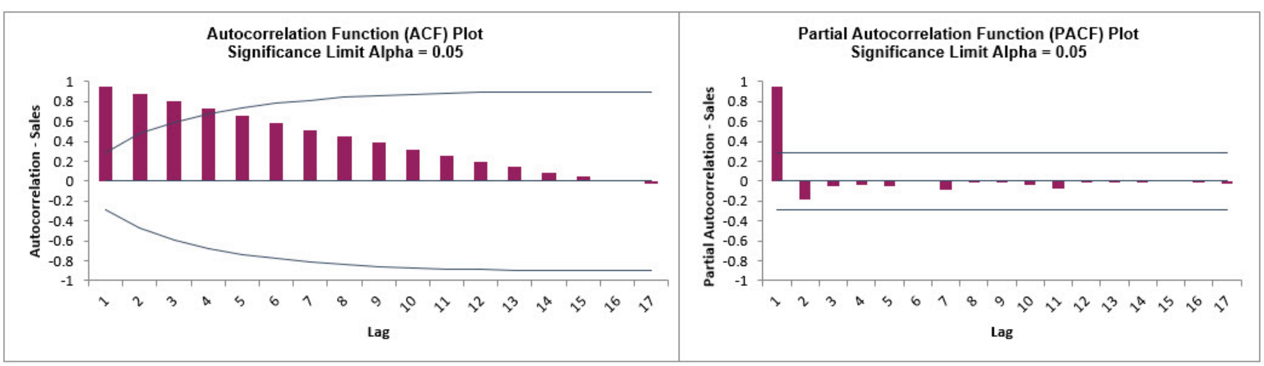
27. Open **Sales with Indicator - Modified Series M.xlsx**. (**Sheet 1** tab). This is modified Series M data from Box and Jenkins, with 50 quarters of corporate sales values along with a leading indicator. The data is treated as nonseasonal, as done in Box and Jenkins.

28. Click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

29. Select *Sales*, click **Numeric Time Series Data (Y) >>**. Select **Automatic Number of Lags.** Specify **Seasonal Frequency** = 1 and **Alpha Level** = 0.05.

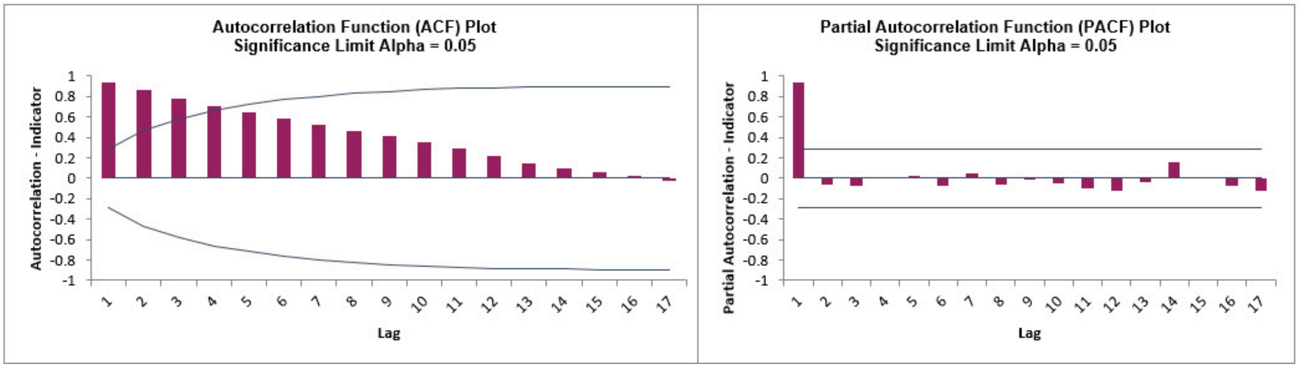


30. Click **OK**. The ACF and PACF plots are produced.



Here we can see the strong autocorrelation with slow decay due to the trend.

31. Click the **Sheet 1** tab. Redo the ACF/PACF Plots for Indicator:



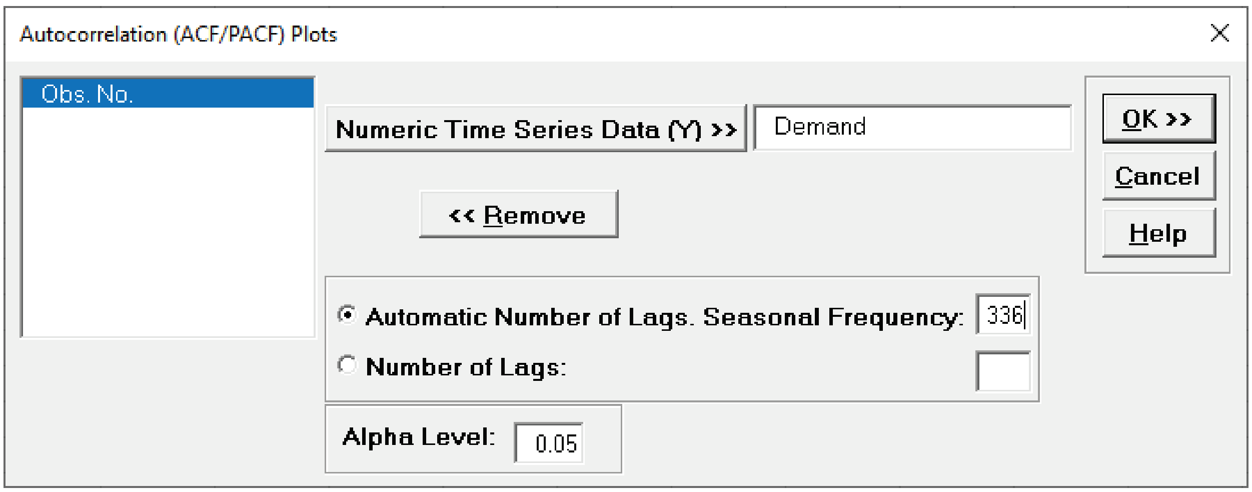
As expected, the autocorrelation for Indicator is very similar to Sales. This will be addressed when Cross Correlation (CCF) Plots are created.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

32. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations.

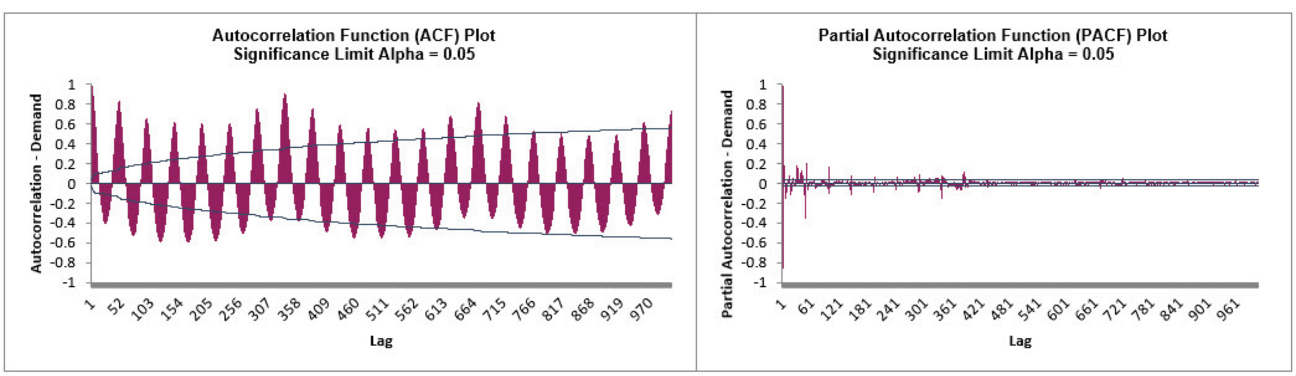
33. Click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

34. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Select **Automatic Number of Lags.** Specify **Seasonal Frequency** = 336 and **Alpha Level** = 0.05.



The automatic number of lags will be (a minimum of) Seasonal Frequency * 3 = 1008 allowing us to see seasonal patterns in the ACF.

35. Click **OK**. The ACF and PACF plots are produced.



Here we can see the half-hourly seasonal patterns of 48 per day and 336 per week.

Cross Correlation (CCF) Plots

Cross Correlation is similar to autocorrelation, but the correlations are computed on two related time series variables, typically a process input and output. A plot of the X data vs. the Y data at lag k may show a positive or negative trend. If the slope is positive, the cross correlation is positive; if there is a negative slope, the cross correlation is negative. This helps to identify important lags (or leads) in the process and is useful for application when there are predictors in an ARIMA model.

If the X input is autocorrelated, the CCF is affected by its time series structure and any "in common" trends the X and Y series may have over time. Pre-whitening solves this problem by removing the autocorrelation and trends.

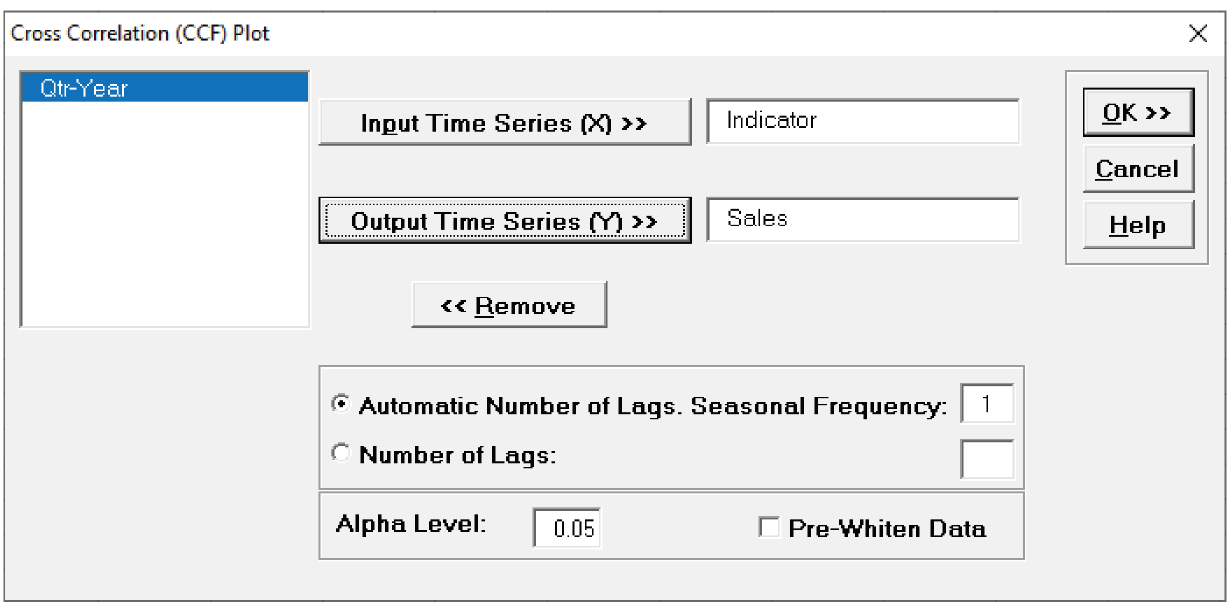
For further details and formulas, see Appendix: Autocorrelation (ACF), Partial Autocorrelation (PACF) and Cross Correlation (CCF).

Sales with Indicator - Modified Series M

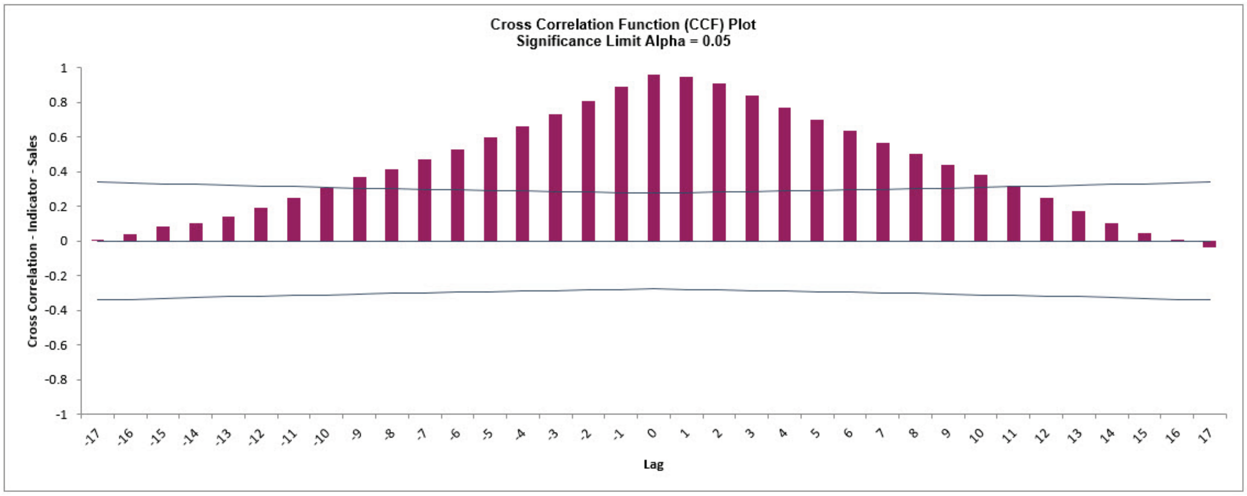
1. Open **Sales with Indicator - Modified Series M.xlsx**. (**Sheet 1** tab). This is modified Series M data from Box and Jenkins, with 50 quarters of corporate sales values along with a leading indicator. The data is treated as nonseasonal, as done in Box and Jenkins.

2. Click **SigmaXL > Time Series Forecasting > Cross Correlation (CCF) Plot**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Indicator*, click **Input Time Series (X) >>**. Select *Sales*, click **Output Time Series (Y) >>**. Use the default **Automatic Number of Lags. Seasonal Frequency =** 1 and **Alpha Level** = 0.05.

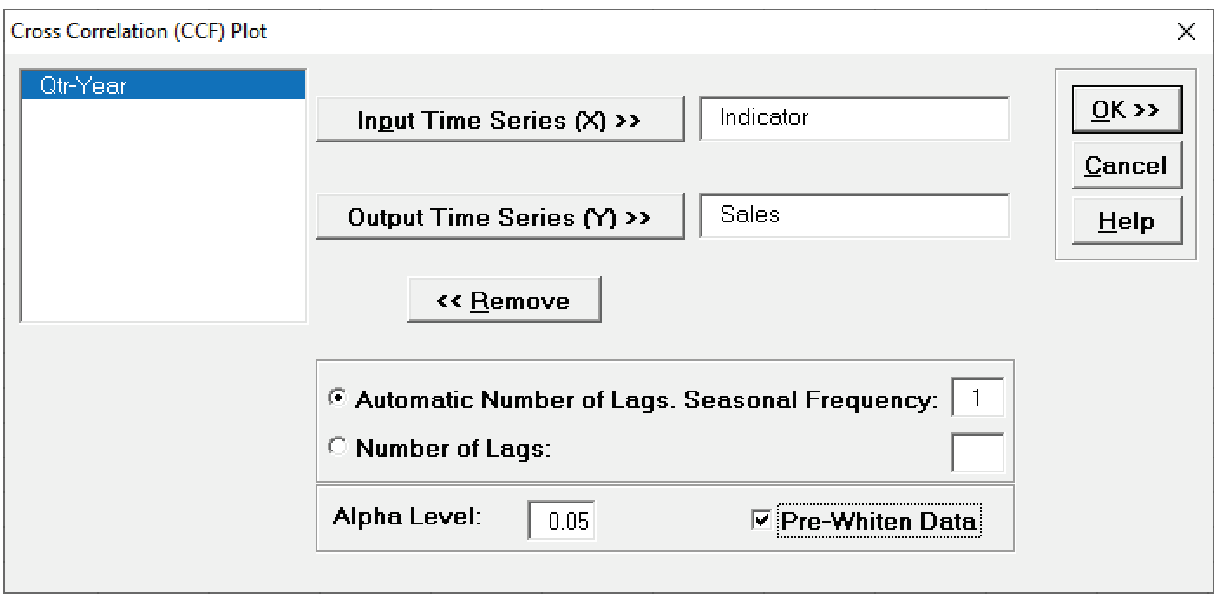


4. Click **OK**. The CCF plot is produced.

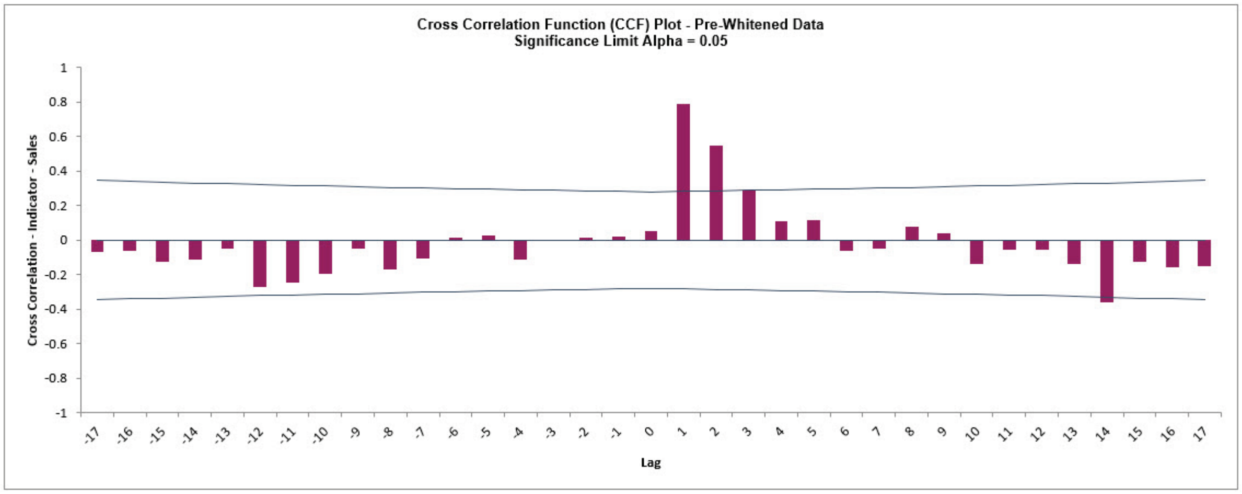


This CCF plot shows significant cross correlation from lag = -9 to +9, with a peak at 0, however the autocorrelation in X and Y data is masking the true nature of the cross correlation.

5. We will redo the CCF Plot with the Pre-Whiten Data option. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Check **Pre-Whiten Data**.



6. Click **OK**. The CCF plot is produced.



Pre-Whitening the data has dramatically altered the CCF plot, allowing us to see the underlying cross correlation pattern. Lags 1 and 2 are significantly positive, and Lag 3 is just on the significance line.

Note that while X is called a leading indicator, i.e., X comes before Y in time, the positive lag means that the X variable is lagging the Y variable in terms of correlation structure. SigmaXL uses this convention as given in Box and Jenkins (2016, pp. 437-440).

This CCF plot will be useful later when we model the Sales data using ARIMA Forecast with Predictors.

Spectral Density Plot

The Spectral Density Plot is used to identify the dominant integer seasonal frequency in time series data using spectral analysis with fast Fourier transforms. The algorithm used here is the same as used in the forecast model option to automatically detect seasonal frequency. If there is multiple seasonality, up to three integer frequencies will be identified. If the peak frequency is not an integer, it is rounded.

The Y axis is Spectral Density, the X axis is Seasonal Frequency. The Spectral Density Plot is also known as a Periodogram. Note that SigmaXL's use of the term "seasonal frequency" is the inverse of what is typically used in Fourier transforms "seasonal period", as discussed in the Introduction.

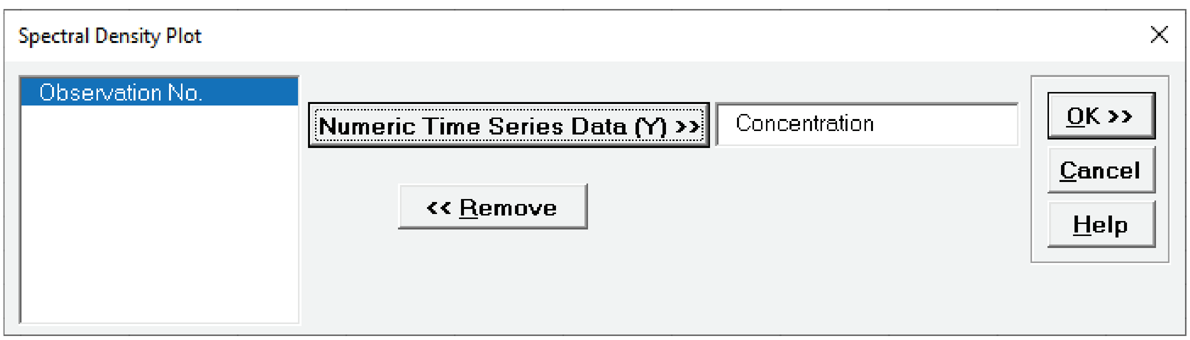
For further details and formulas, see Appendix: Spectral Density and Automatic Detection of Seasonal Frequency.

Chemical Process Concentration - Series A

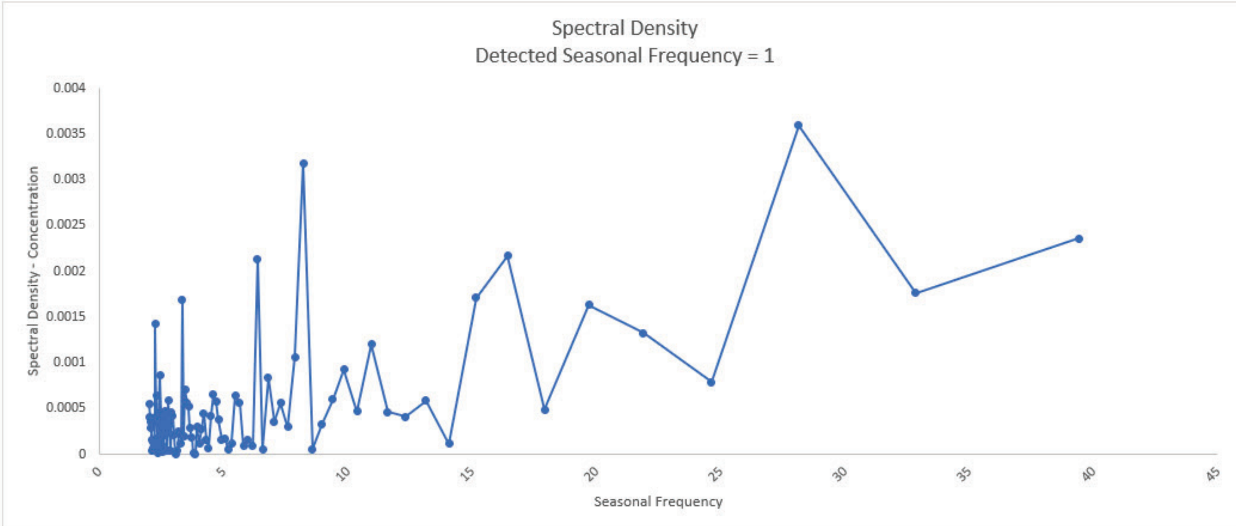
1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

2. Click **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**.



4. Click **OK**. A Spectral Density Plot for Concentration is produced.



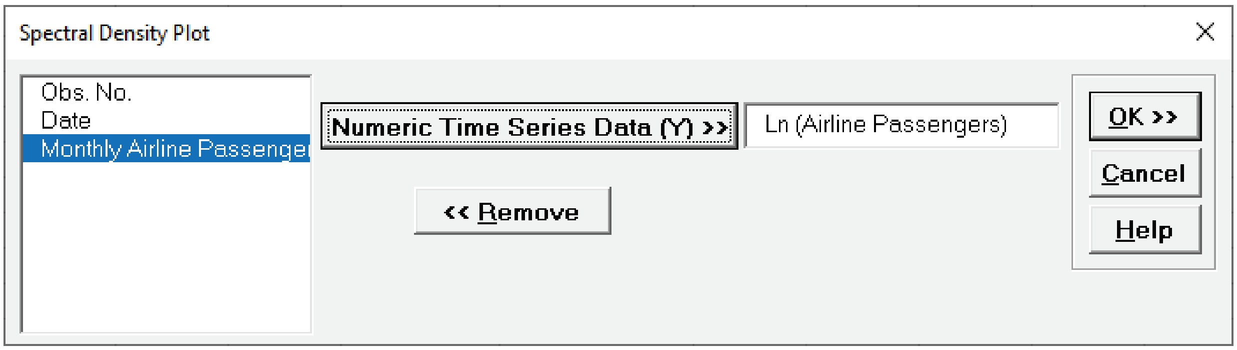
The detected seasonal frequency is 1, which means that it is a nonseasonal process. The peak at 28 does not have enough seasonal "strength" to be considered for use as seasonal frequency in a time series model.

Monthly Airline Passengers - Series G

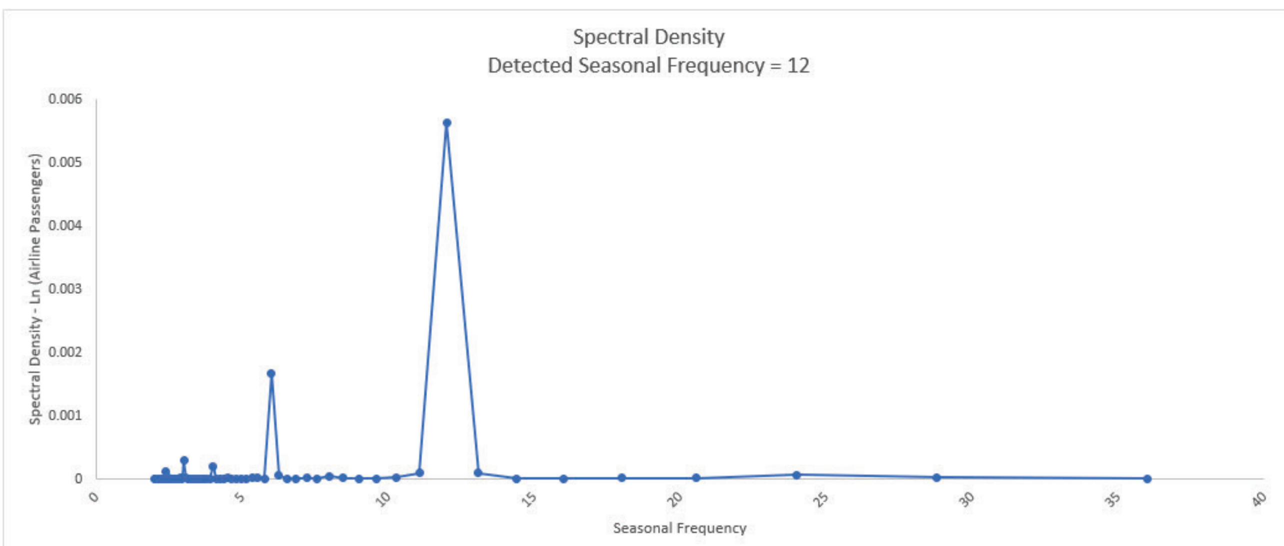
5. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960.

6. Click **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

7. Select *Ln(Airline Passengers)*, click **Numeric Time Series Data (Y) >>**.



8. Click **OK**. A Spectral Density Plot for Ln(Airline Passengers) is produced.



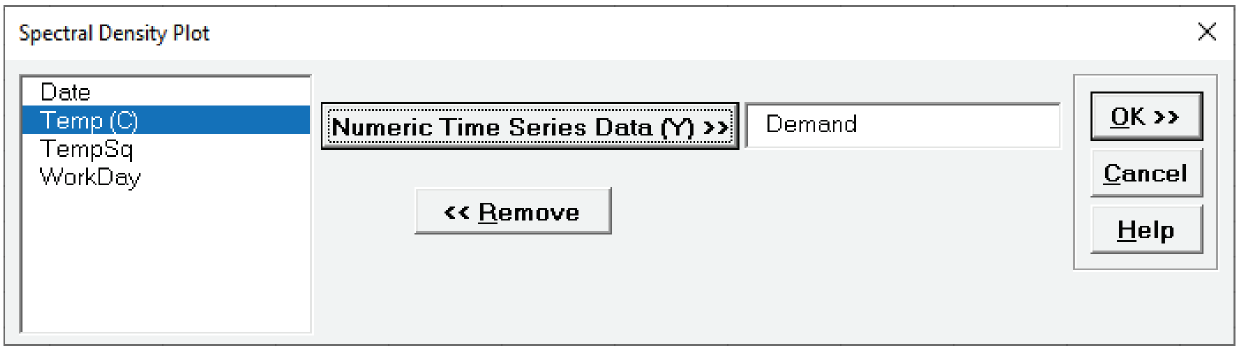
As expected, the detected seasonal frequency for the monthly data is 12. The peak at 6 does not have enough seasonal “strength” to be considered as a second seasonal frequency.

Daily Electricity Demand with Predictors – ElecDaily

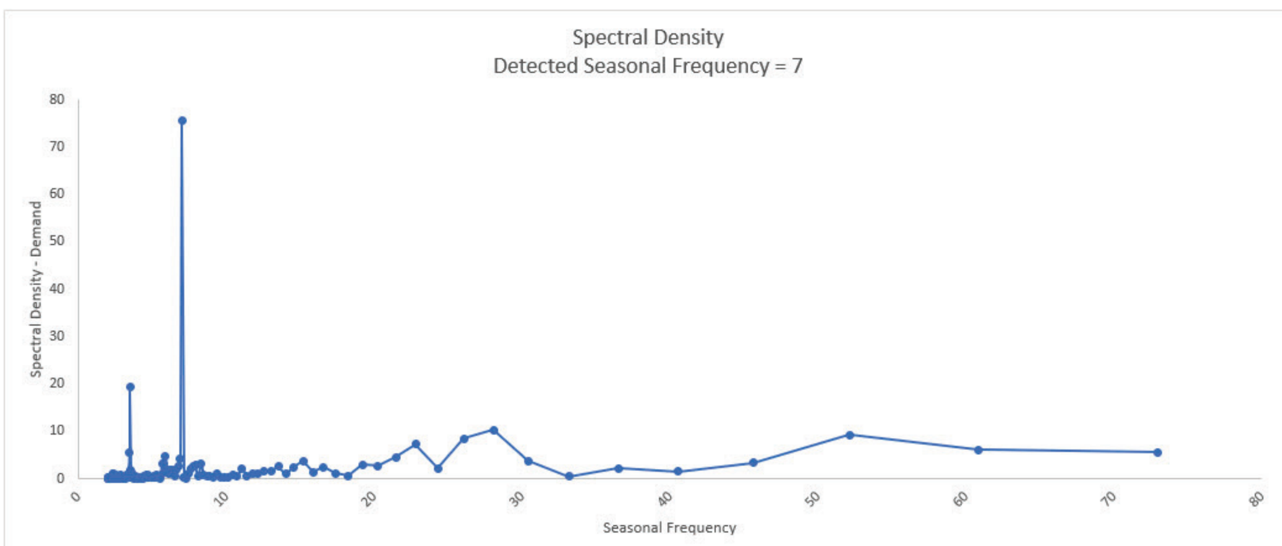
9. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (Hyndman, fpp2). This data has a seasonal frequency = 7 (observations per week).

10. Click **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

11. Select *Demand*, click **Numeric Time Series Data (Y) >>**.



12. Click **OK**. A Spectral Density Plot for Demand is produced.



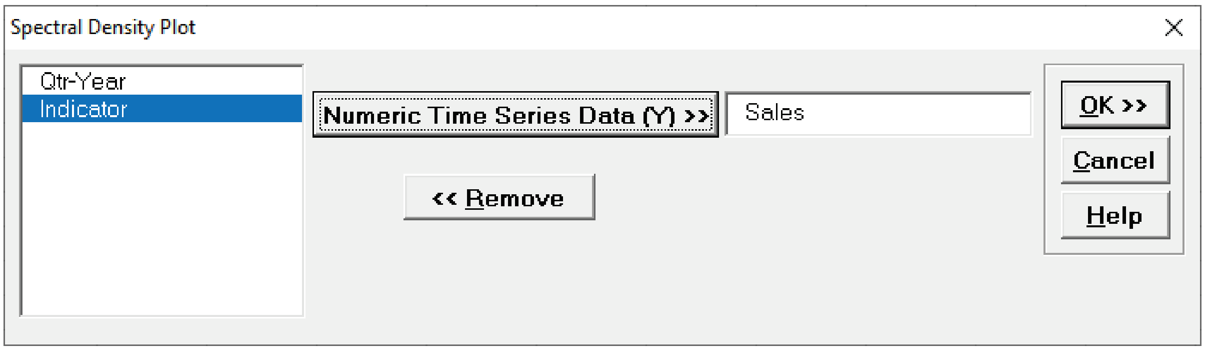
As expected, the detected seasonal frequency for the daily data is 7.

Sales with Indicator - Modified Series M

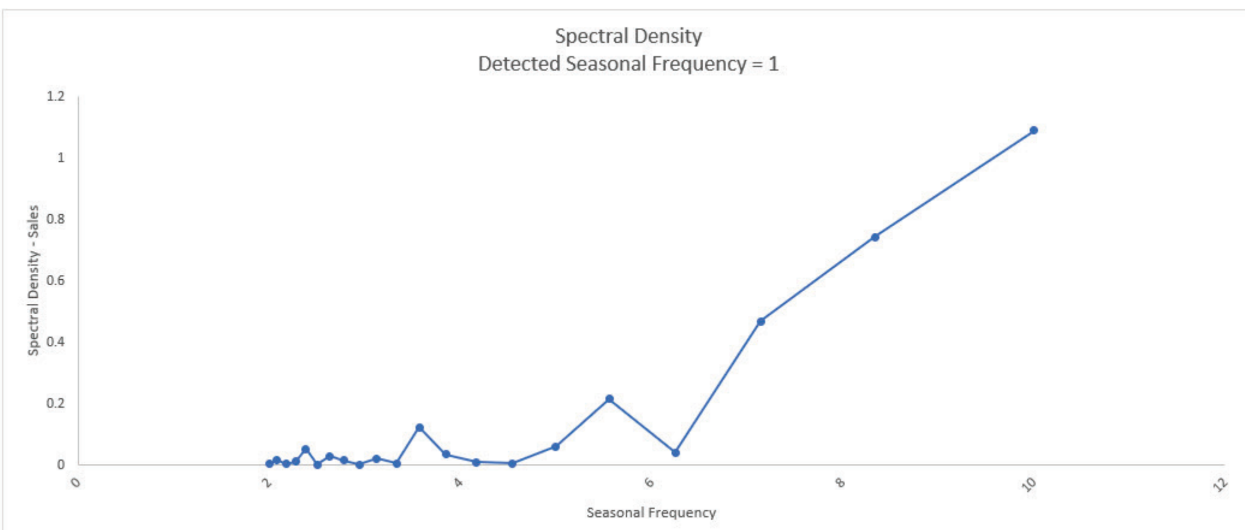
13. Open **Sales with Indicator - Modified Series M.xlsx**. (**Sheet 1** tab). This is modified Series M data from Box and Jenkins, with 50 quarters of corporate sales values along with a leading indicator. The data is treated as nonseasonal, as done in Box and Jenkins.

14. Click **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

15. Select *Sales*, click **Numeric Time Series Data (Y) >>**.



16. Click **OK**. A Spectral Density Plot for Sales is produced.



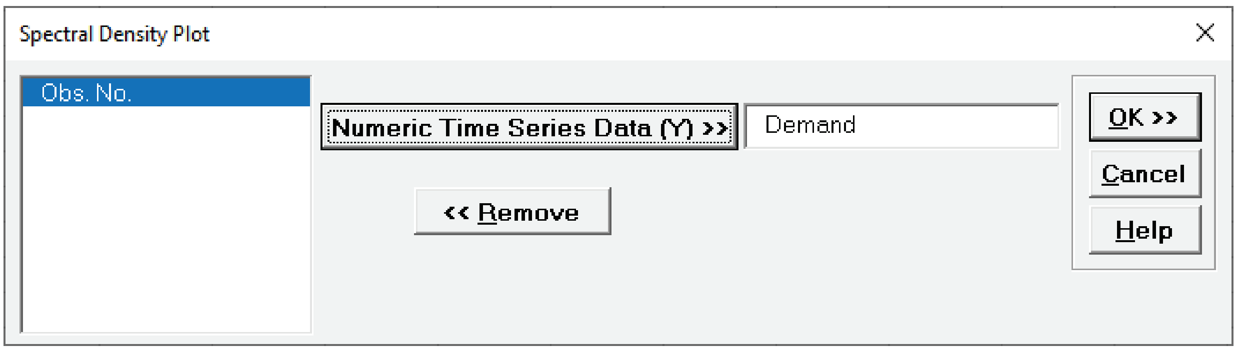
The detected seasonal frequency is 1, which means that it is confirmed to be a nonseasonal process. The peak at 10 does not have enough seasonal "strength" to be considered for use as seasonal frequency in a time series model.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

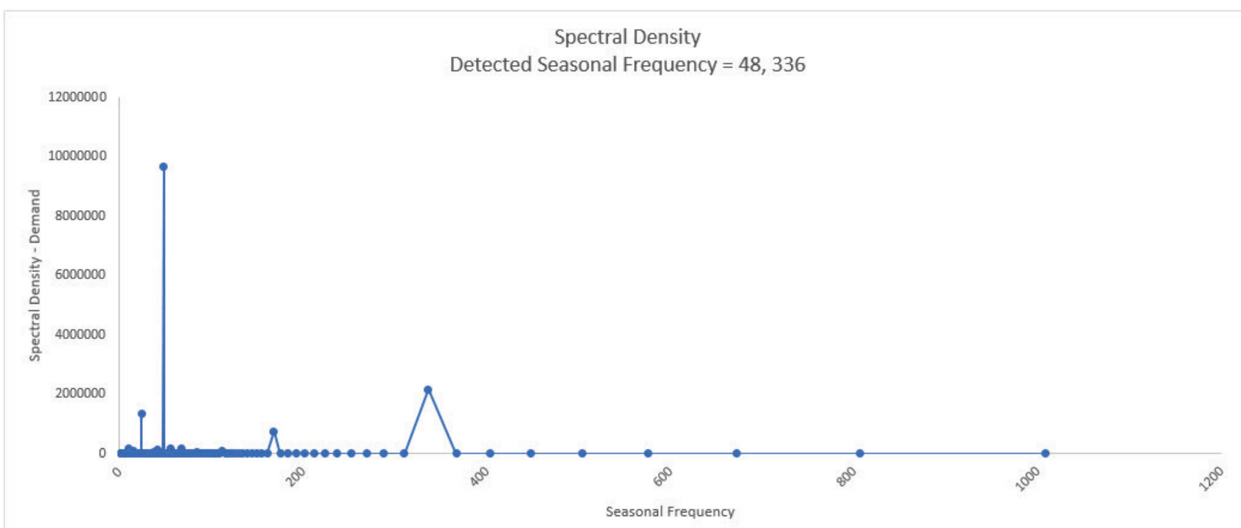
17. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations.

18. Click **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

19. Select *Demand*, click **Numeric Time Series Data (Y) >>**.



20. Click **OK**. A Spectral Density Plot for Demand is produced.



The detected multiple seasonal frequency is confirmed as 48, 336.

Seasonal Trend Decomposition Plots

The Seasonal Trend Decomposition Plots are useful to visually distinguish trend and seasonal components in the time series data. If the Seasonal Frequency is unchecked, a Trend Decomposition Plot is produced as the first chart, showing the raw data and the trend. The trend component uses data smoothing, rather than a linear trend so that it may display either a linear trend or cyclical patterns. If a single seasonal frequency is specified, a Seasonal Trend Decomposition Plot is produced, showing the data, smoothed trend and seasonal component. If a multiple seasonal frequency is specified, a Multiple Seasonal Trend Decomposition Plot is produced, showing the data, smoothed trend and multiple seasonal component.

The second chart shows just the smoothed trend; the third chart (if applicable) shows just the seasonal or multiple seasonal component. The final chart is the remainder component.

This is an additive decomposition model, so the sum of the trend value + seasonal value(s) + remainder value gives the original data value. A multiplicative equivalent may be obtained by specifying the Box-Cox Transformation with Lambda = 0, which is a Ln transformation, but the charts will display the transformed data to maintain an additive model. Rounded or Optimal Lambda may also be used, but will only consider the range of values 0 to 1 (this conservative approach is used in time series forecasting, unlike regular Box-Cox in SigmaXL which uses a range of -5 to 5). See Appendix: **Box-Cox Transformation.**

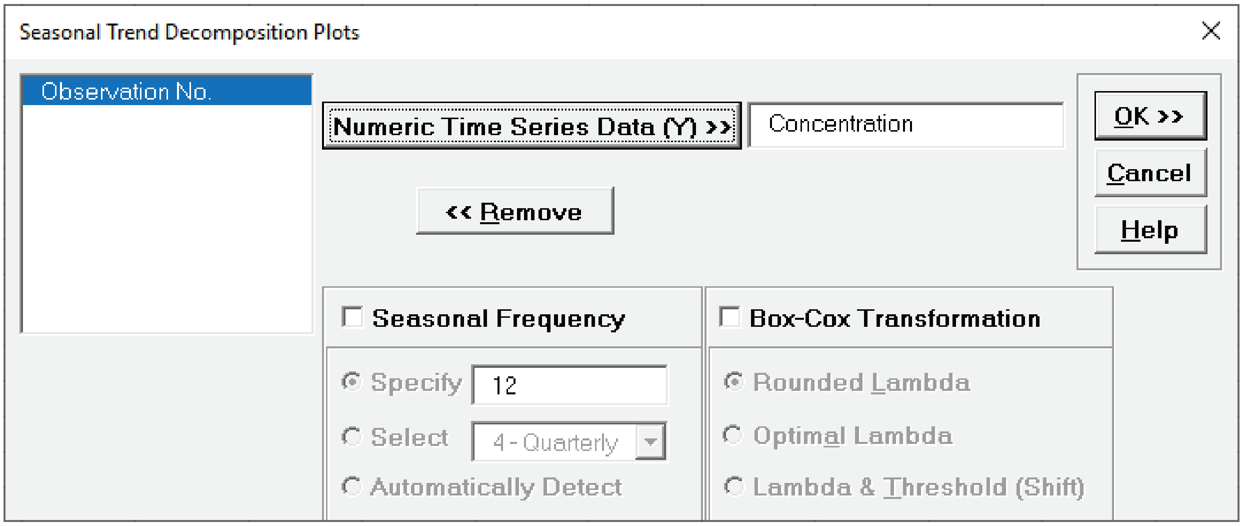
The decomposition algorithms used here are the same as used in Exponential Smoothing – Multiple Seasonal Decomposition (MSD), and ARIMA – MSD. The seasonal component is first removed through decomposition, a nonseasonal exponential smooth model fitted to the remainder + trend, and then the seasonal component is added back in. This is mainly used for high seasonal frequency and/or multiple seasonal frequency time series. For further details and formulas, see Appendix: **Seasonal Trend Decomposition.**

Chemical Process Concentration – Series A

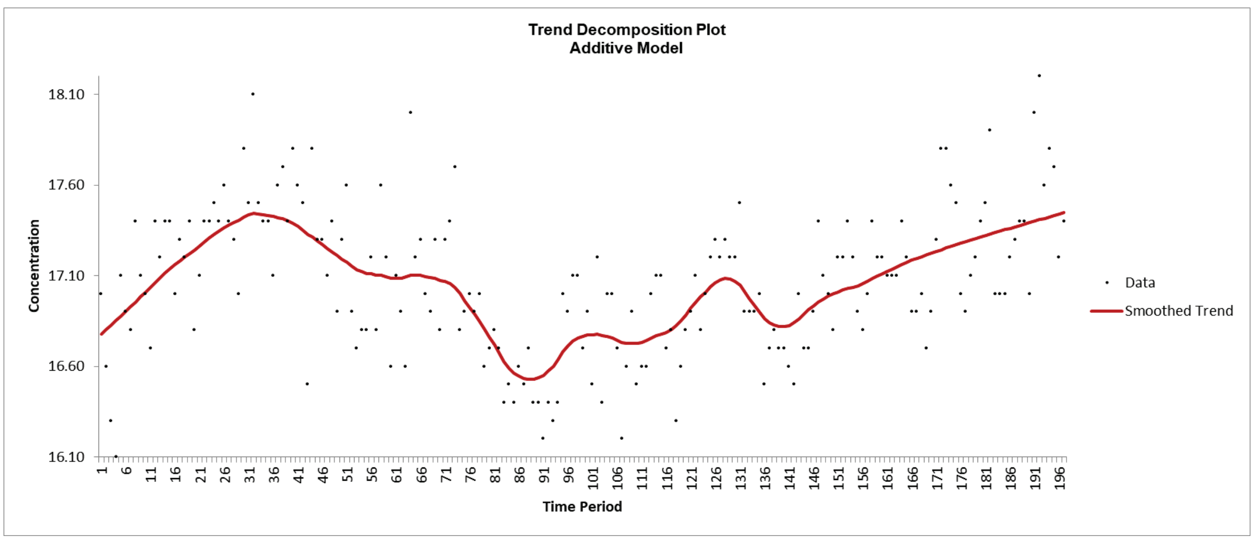
1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

2. Click **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. **Seasonal Frequency** and **Box-Cox Transformation** should be unchecked as shown.



4. Click **OK**. A Trend Decomposition Plot for Concentration is produced.



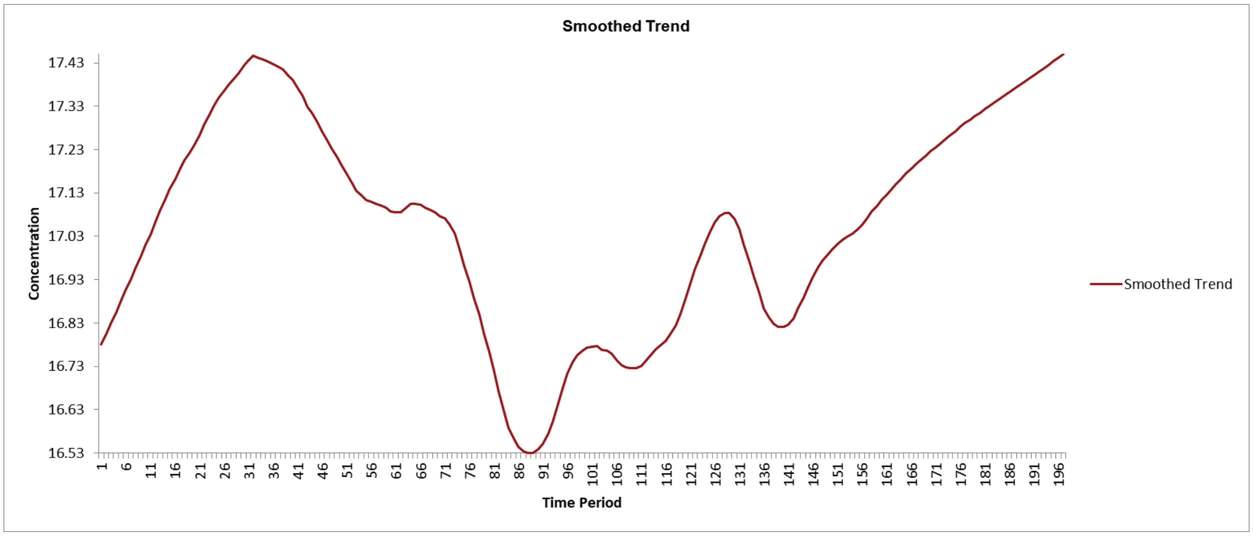
We can clearly see the "wandering mean" in this process. As discussed earlier, this can be modeled with exponential smoothing or with ARIMA after differencing.

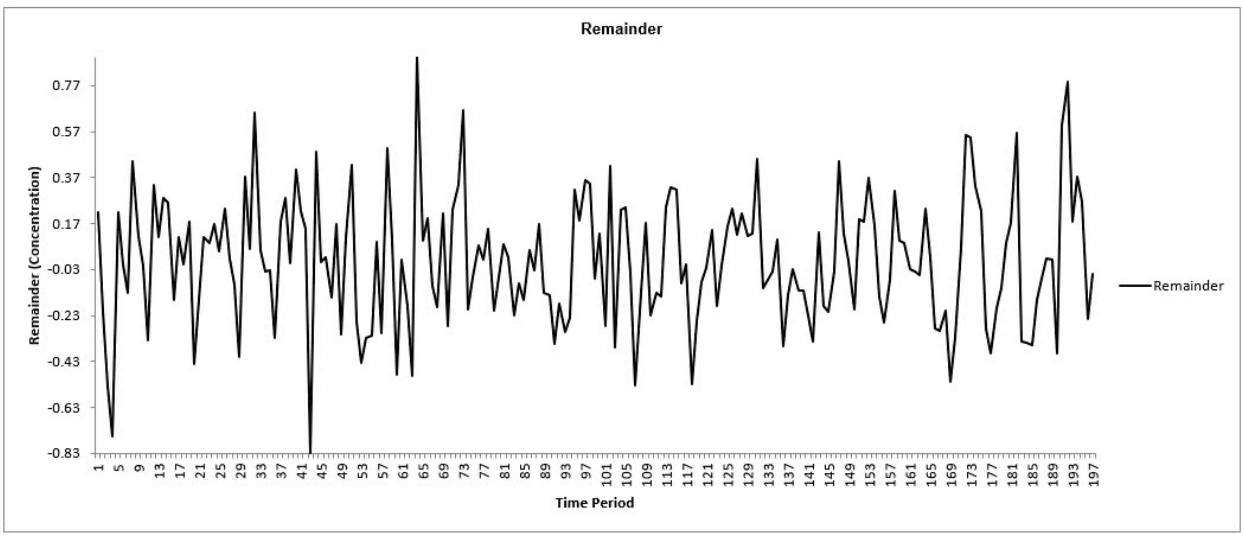
5. A Decomposition Summary report is also included to the right of the plot, indicating seasonal frequency and Box-Cox parameters (if applicable):

Decomposition Summary	
Seasonal Frequency	1
Box-Cox Transformation	N/A
Lambda	
Threshold	

Seasonal Frequency = 1 denotes a nonseasonal process.

6. The Smoothed Trend and Remainder Plots are also produced as shown:



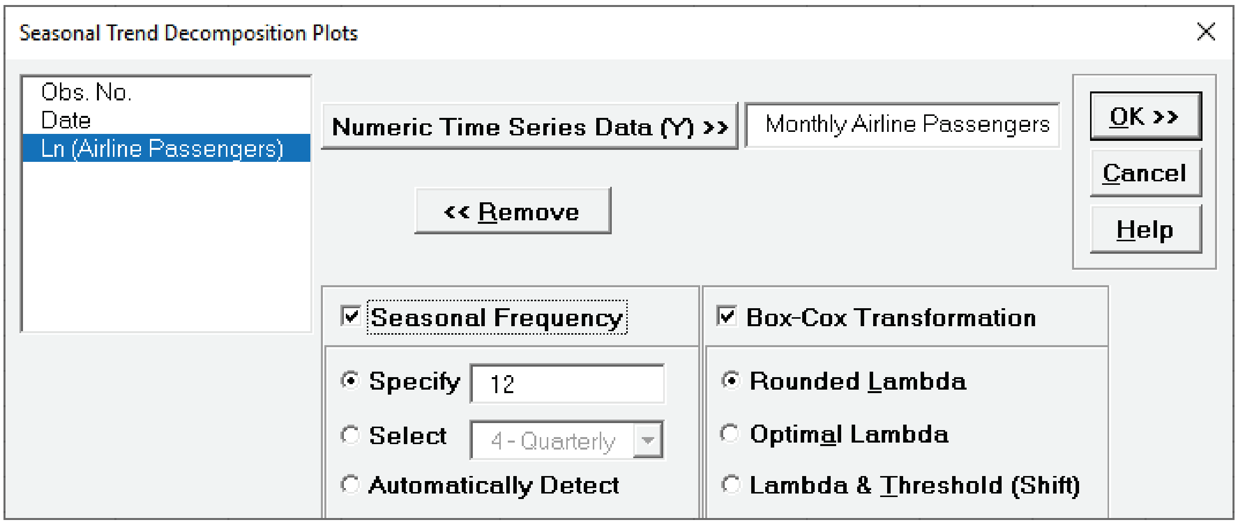


Monthly Airline Passengers - Series G

7. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960.

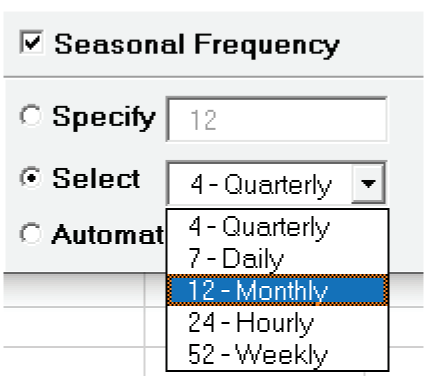
8. Click **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

9. Select *Monthly Airline Passengers*, click **Numeric Time Series Data (Y) >>**. Check **Seasonal Frequency** with **Specify** = 12**.** Check **Box-Cox Transformation** with **Rounded Lambda** (selected because the Run Chart showed an increase in the seasonal variance over time).



- **Seasonal Frequency Specify** also permits the entry of multiple frequencies

- **Seasonal Frequency Select** gives a drop-down list of commonly used seasonal frequencies:



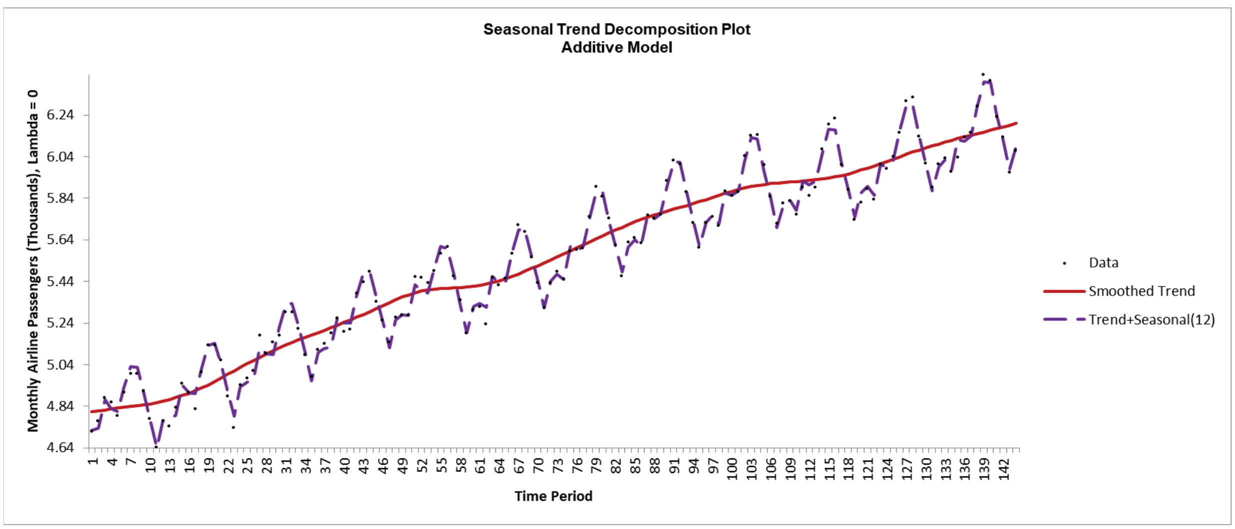
- **Seasonal Frequency Automatically Detect** should be used if uncertain what the seasonal frequency value is (or do a Spectral Density Plot prior to the Seasonal Trend Decomposition Plots).

- **Box-Cox Transformation** with **Rounded Lambda** will select Lambda = 0 (Ln), 0.5 (SQRT) or 1 (Untransformed). Threshold (Shift) is computed automatically if the time series data includes 0 or negative values, otherwise it is 0.

- **Box-Cox Transformation** with **Optimal Lambda** uses the range of 0 to 1 for Lambda. Threshold is computed automatically if the time series data includes 0 or negative values.

- **Box-Cox Transformation** with **Lambda & Threshold (Shift)** if left blank, will compute optimal lambda and threshold. The user may also specify Lambda and Threshold. Lambda may be specified outside of the 0 to 1 range, but practically for time series analysis, should be limited to -1 to 2. Threshold is typically 0, but if the time series data includes 0 or negative values, a negative threshold value should be entered that is smaller than the minimum data value. This value will be subtracted from the data resulting in positive time series values.

10. Click **OK**. A Seasonal Trend Decomposition Plot for Monthly Airline Passengers is produced.

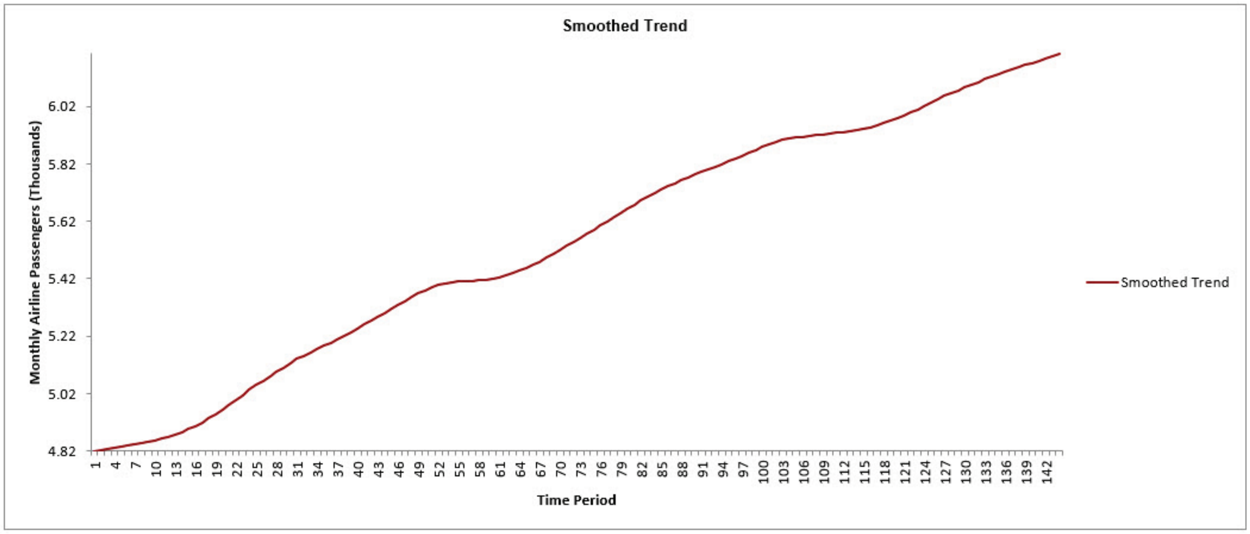


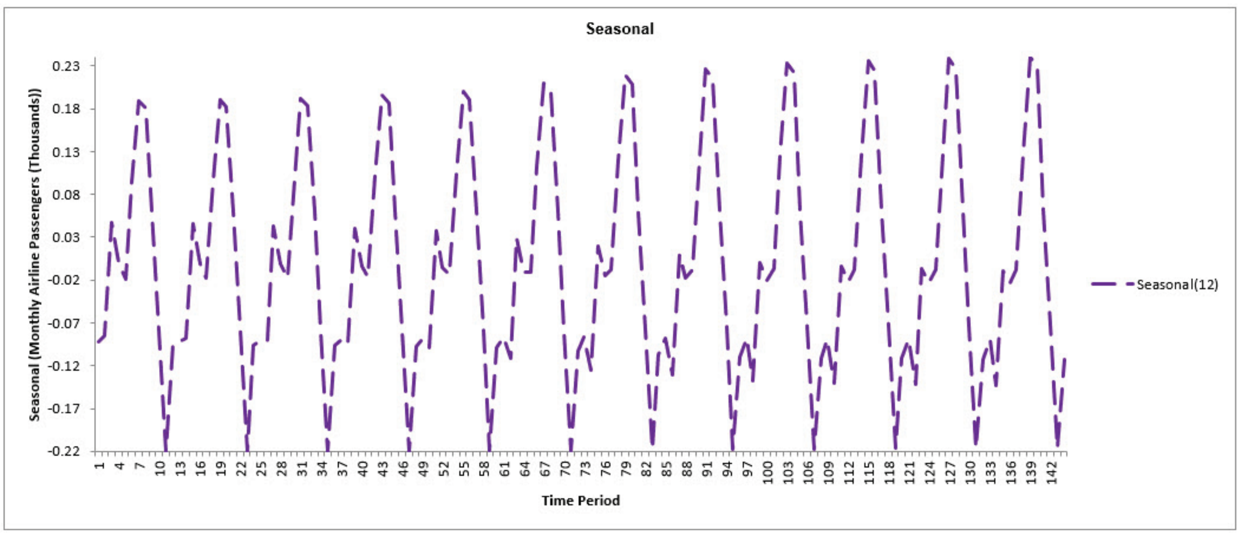
Here we see a strong positive trend as well as the monthly seasonal effect. Note that this is the Lambda = 0 (Ln transformed) data. The Box-Cox transformation information is given in the Decomposition Summary report to the right of the plot:

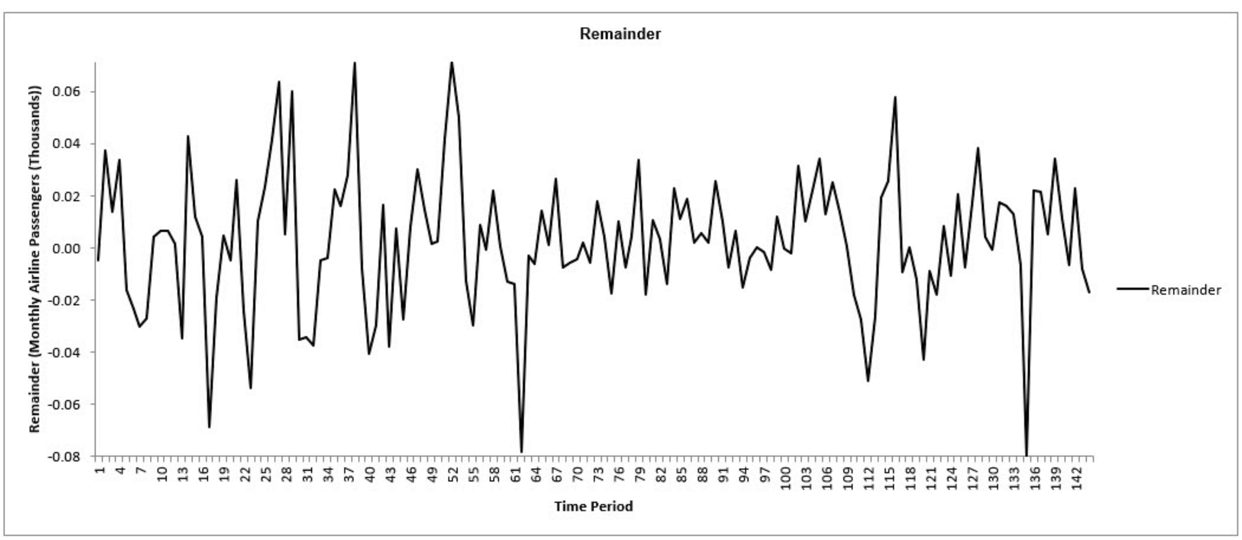
Decomposition Summary	
Seasonal Frequency	12
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Seasonal Frequency is 12 and Lambda = 0. The Ln transformed data are displayed to maintain an additive model, which is easier to interpret than a multiplicative model.

11. The Smoothed Trend, Seasonal and Remainder Plots are also produced as shown:





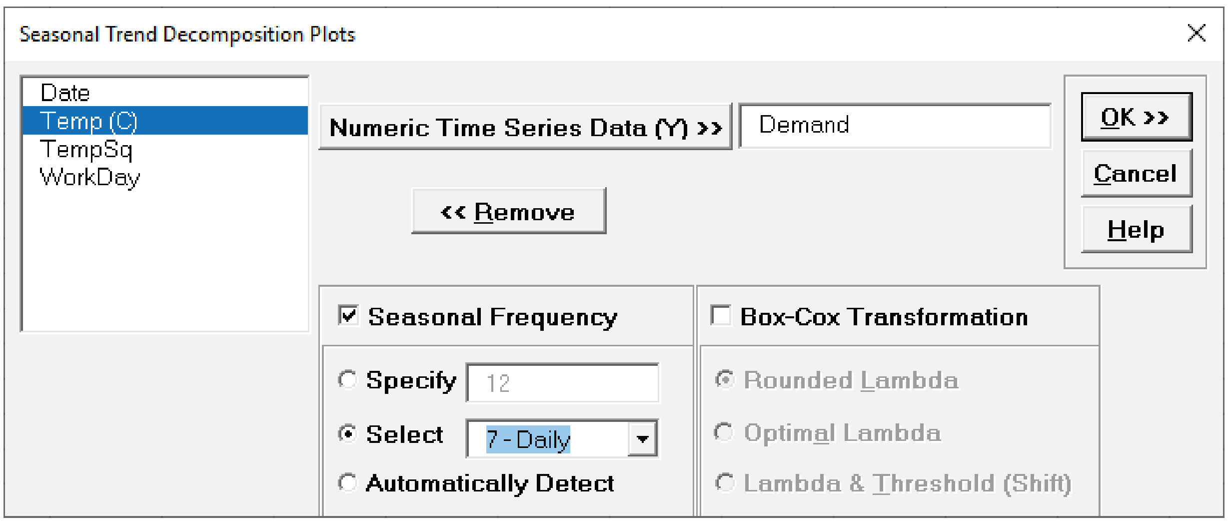


Daily Electricity Demand with Predictors – ElecDaily

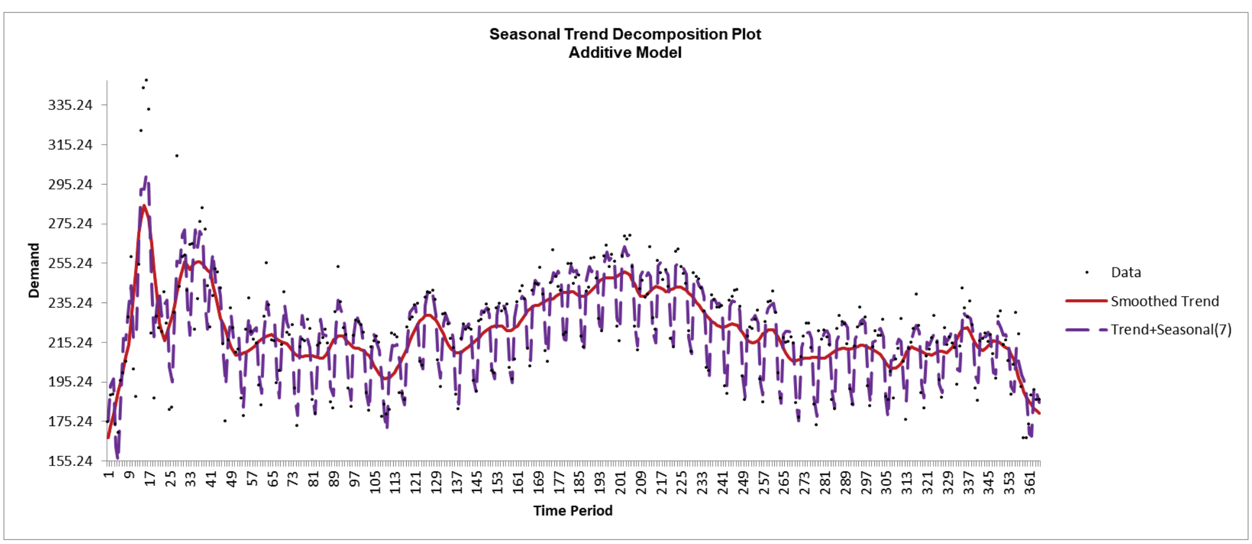
12. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (Hyndman, fpp2). This data has a seasonal frequency = 7 (observations per week).

13. Click **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

14. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Check **Seasonal Frequency** with **Select** = *7-Daily* selected from the drop-down list**.** Uncheck **Box-Cox Transformation**.

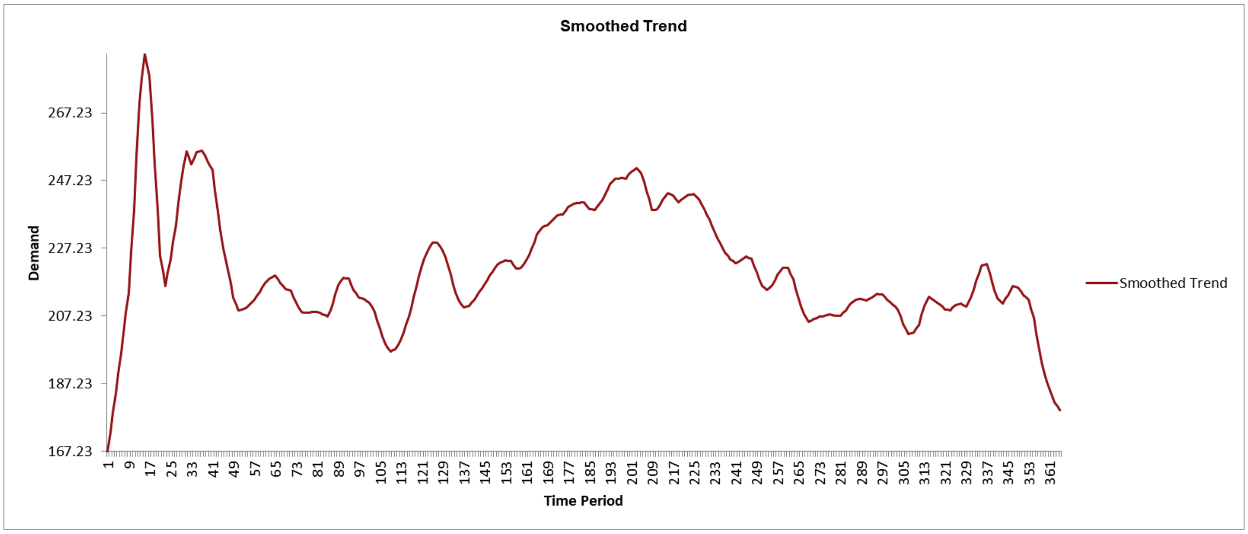


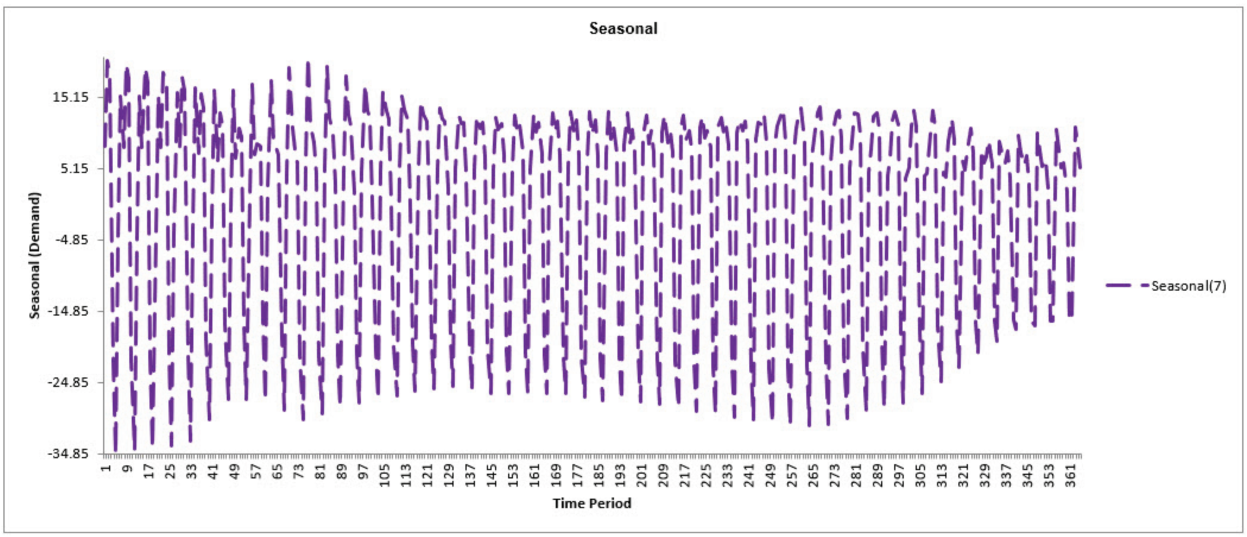
15. Click **OK**. A Seasonal Trend Decomposition Plot for Demand is produced.

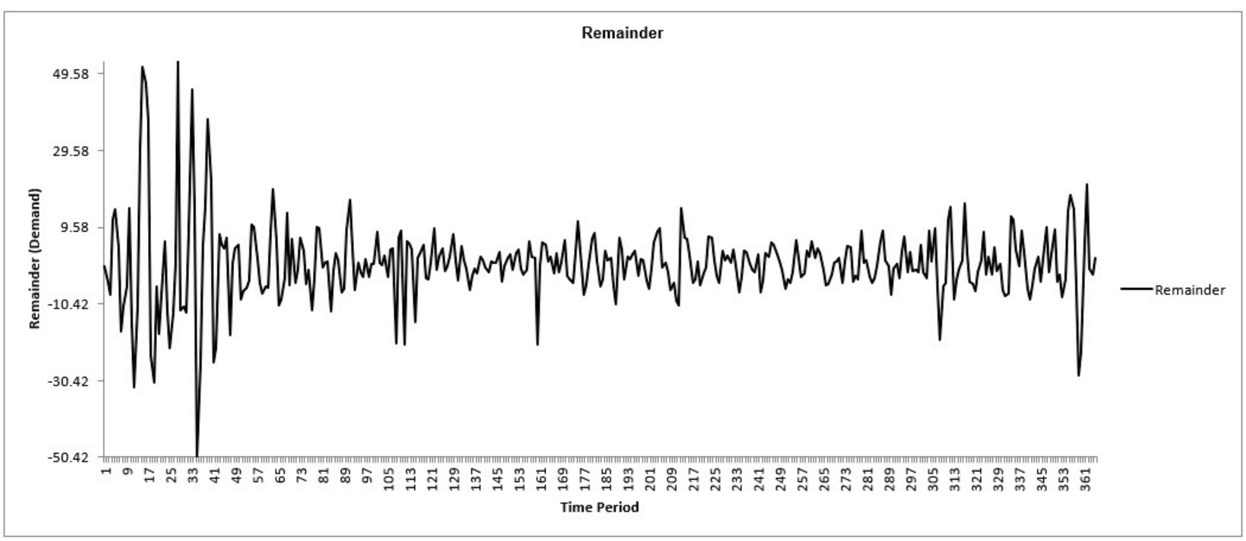


Here we see the daily seasonal effect. As discussed earlier, some of the trend patterns can be explained by the predictors Temp (C), TempSq and WorkDay.

16. The Smoothed Trend, Seasonal and Remainder Plots are also produced as shown:





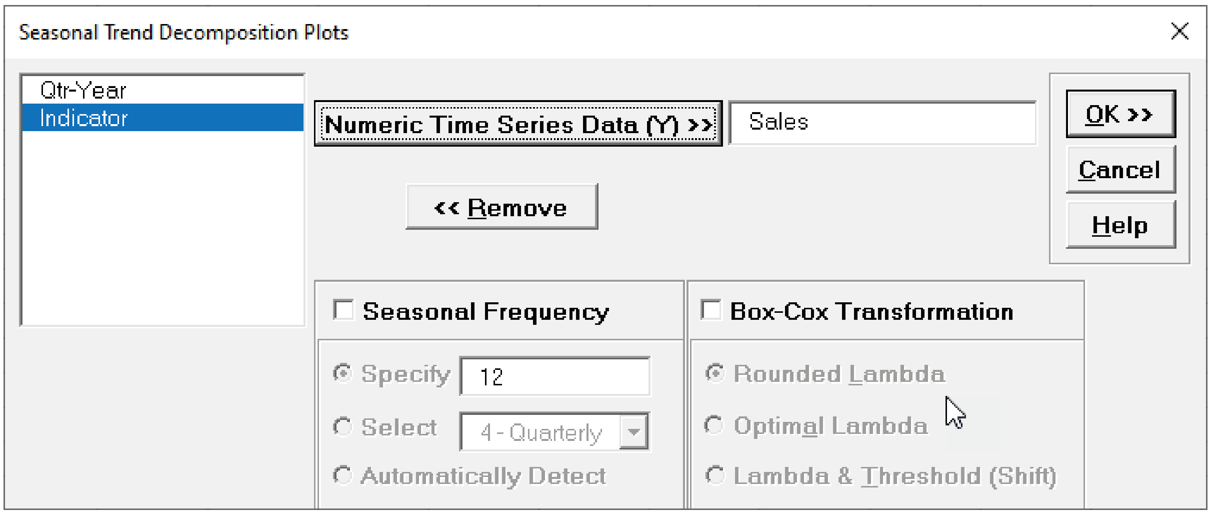


Sales with Indicator - Modified Series M

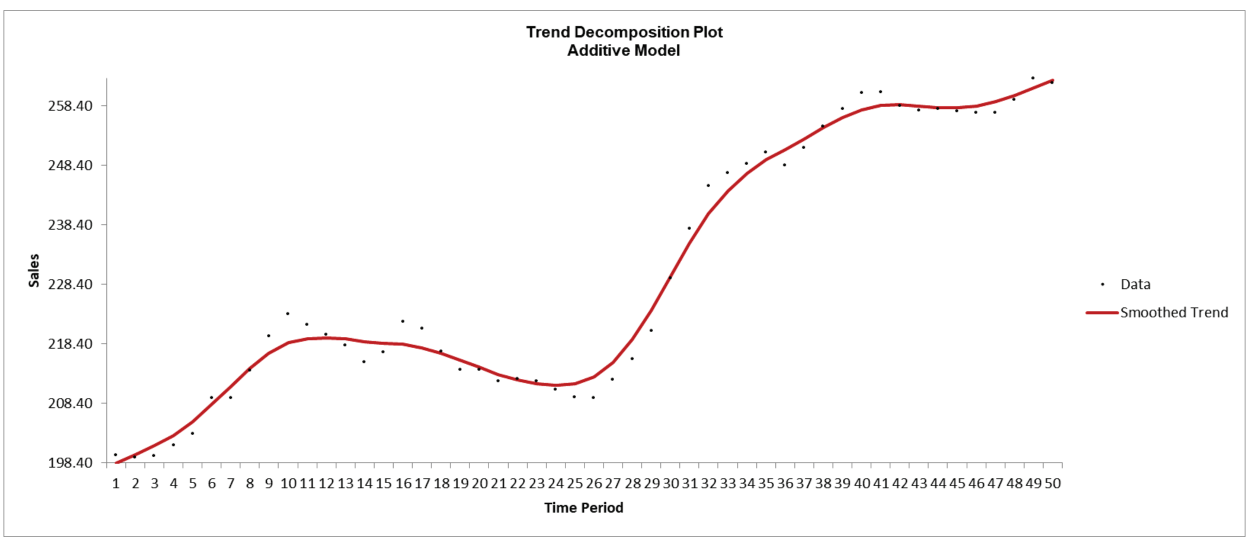
17. Open **Sales with Indicator - Modified Series M.xlsx**. (**Sheet 1** tab). This is modified Series M data from Box and Jenkins, with 50 quarters of corporate sales values along with a leading indicator. The data is treated as nonseasonal, as done in Box and Jenkins.

18. Click **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

19. Select *Sales*, click **Numeric Time Series Data (Y) >>**. **Seasonal Frequency** and **Box-Cox Transformation** should be unchecked as shown.

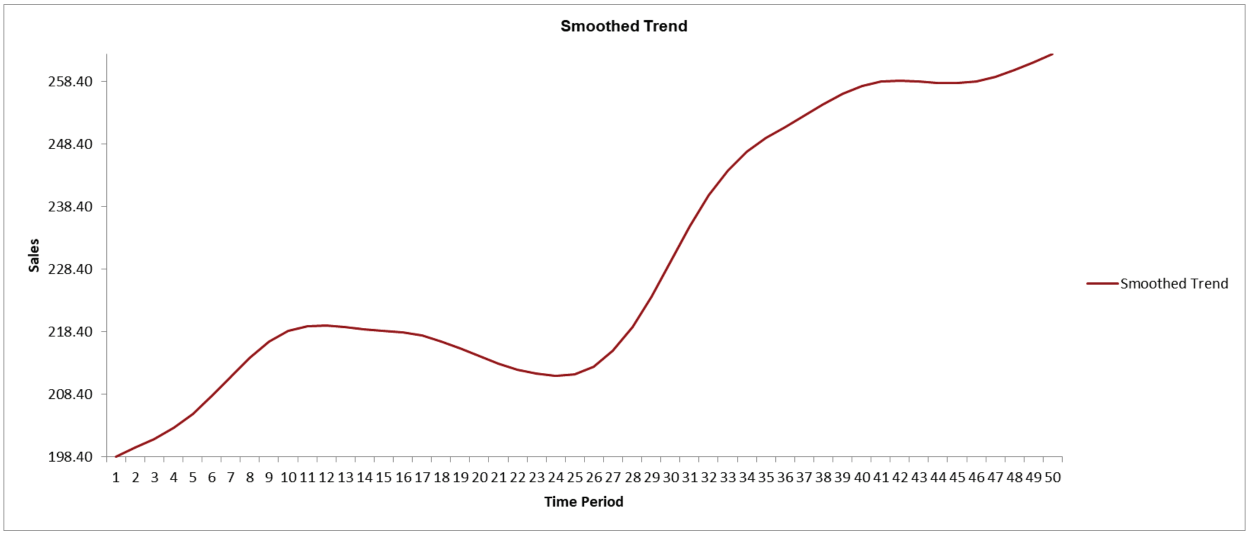


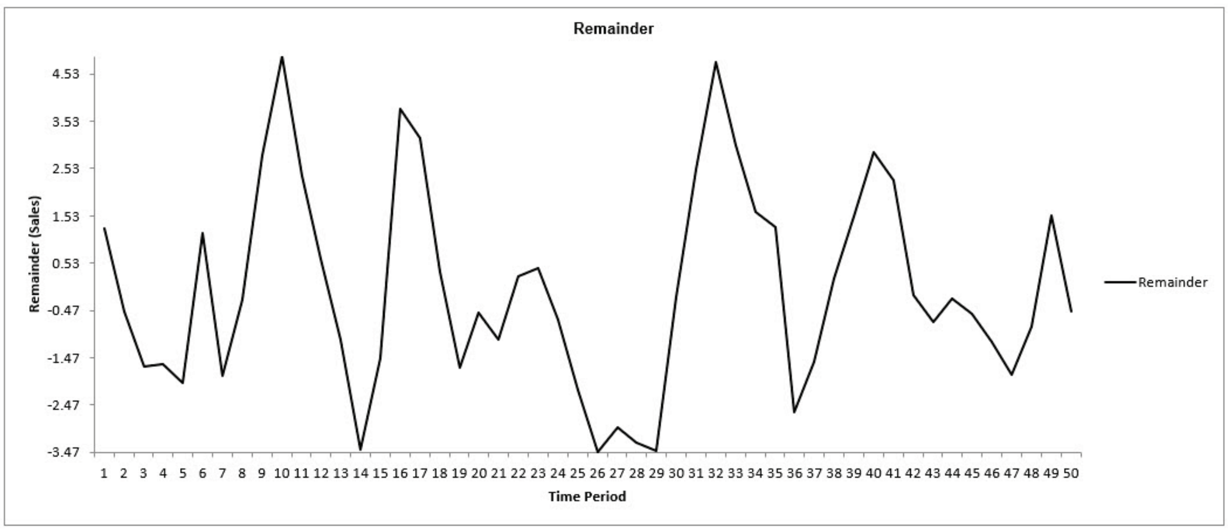
20. Click **OK**. A Trend Decomposition Plot for Concentration is produced.



Here we see an overall positive trend over the 50 quarters.

21. The Smoothed Trend and Remainder Plots are also produced as shown:



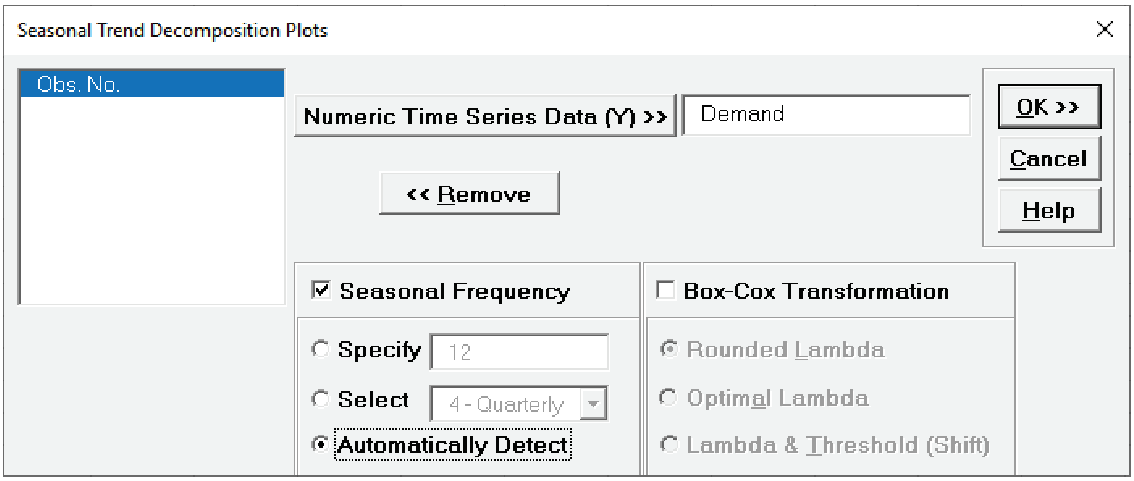


Half-Hourly Multiple Seasonal Electricity Demand – Taylor

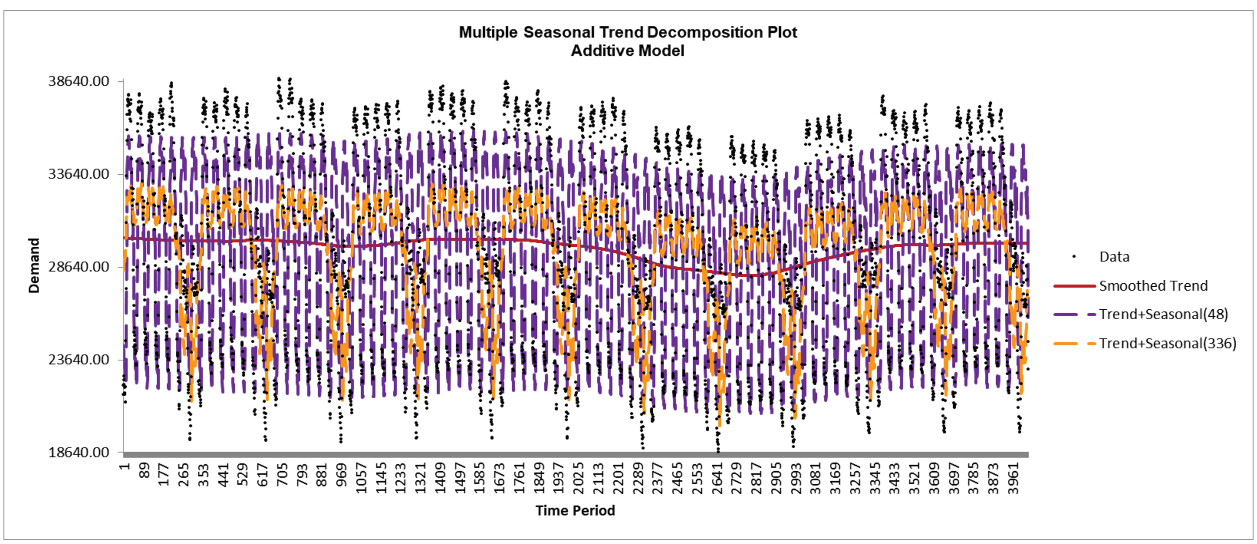
22. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations.

23. Click **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

24. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Check **Seasonal Frequency** with **Automatically Detect.** Uncheck **Box-Cox Transformation**.

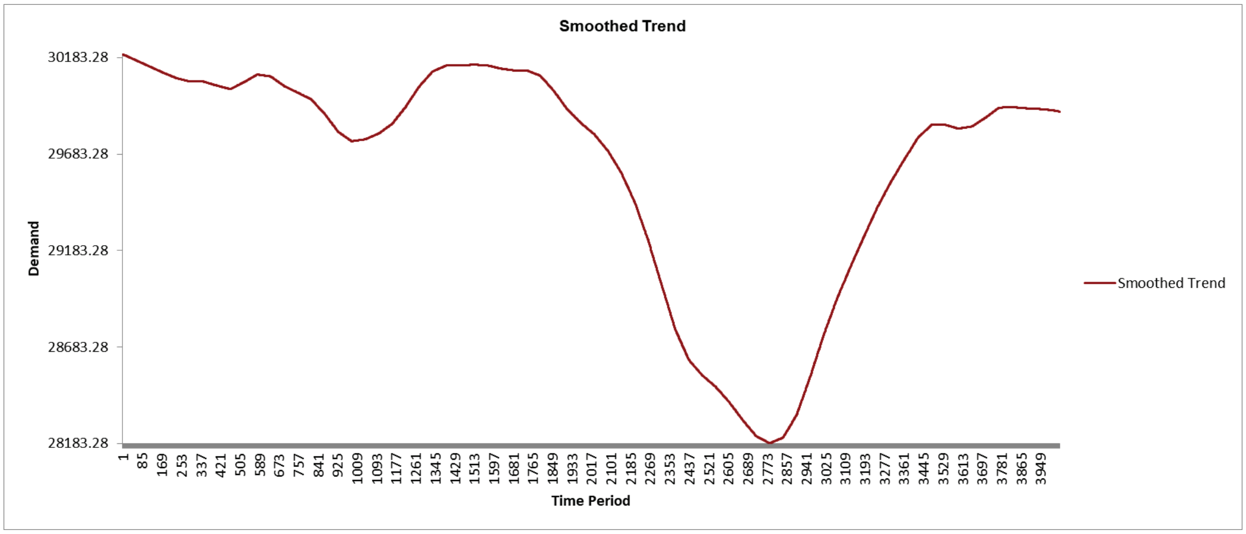


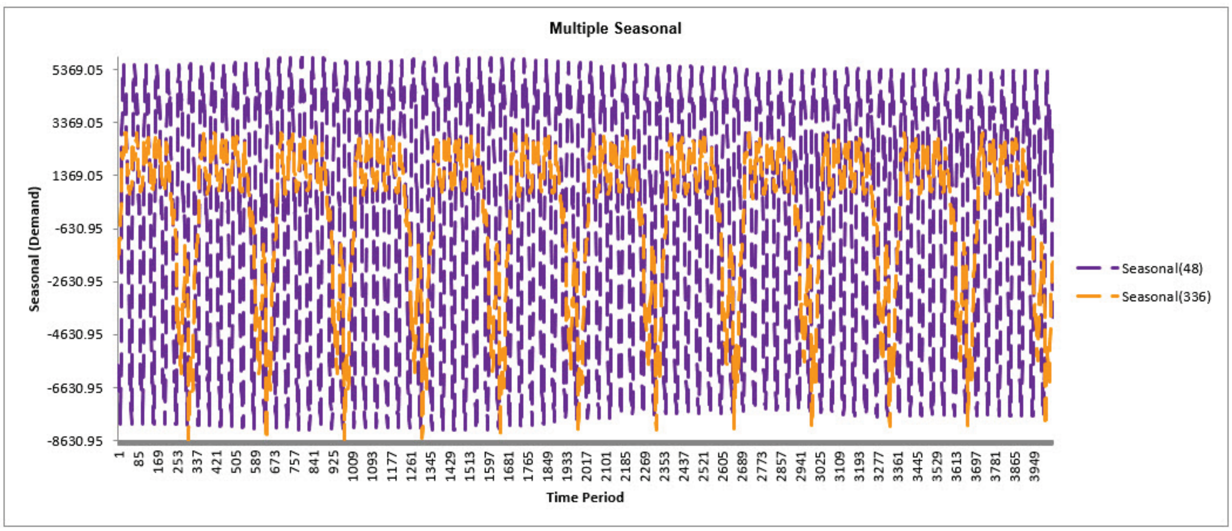
25. Click **OK**. A Seasonal Trend Decomposition Plot for Demand is produced.

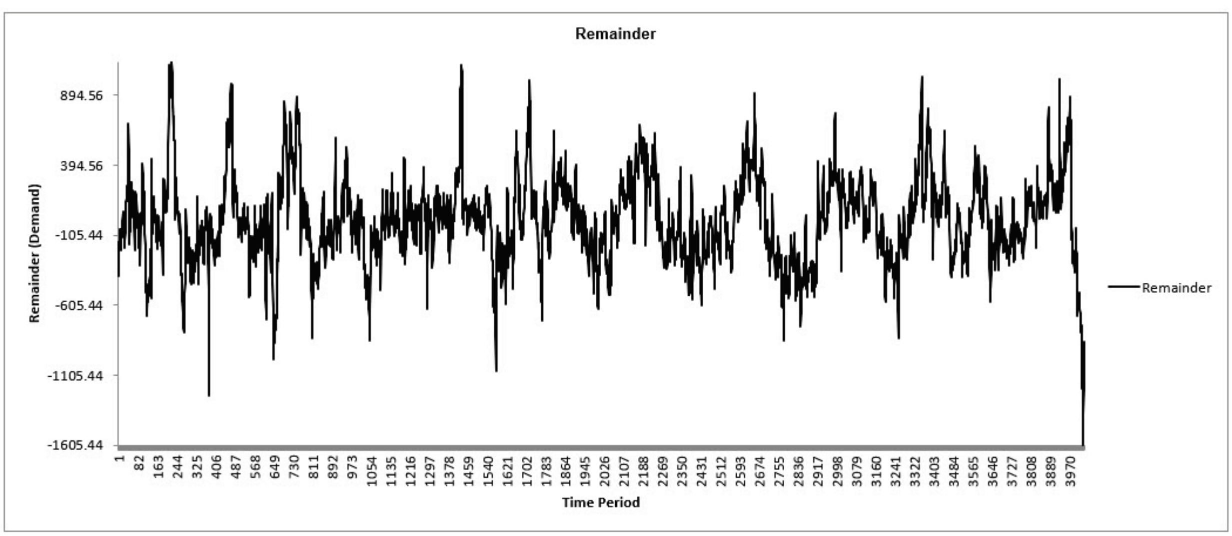


Automatic detection of seasonal frequency gives 48, 336, as obtained earlier with the Spectral Density Plot. Here we see the half-hourly multiple seasonal effect with 48 observations per day and 336 observations per week.

26. The Smoothed Trend, Seasonal and Remainder Plots are also produced as shown:







Seasonal Interaction Plots

Bisgaard and Kulahchi (2011) give a novel use of two-way interaction plots to view trends and seasonal effects in data. We will use Two-Way ANOVA to reproduce the interaction charts given in the book.

Note, in order to produce these charts, the data must be balanced, e.g., every year must have 12 months of data.

Reference:

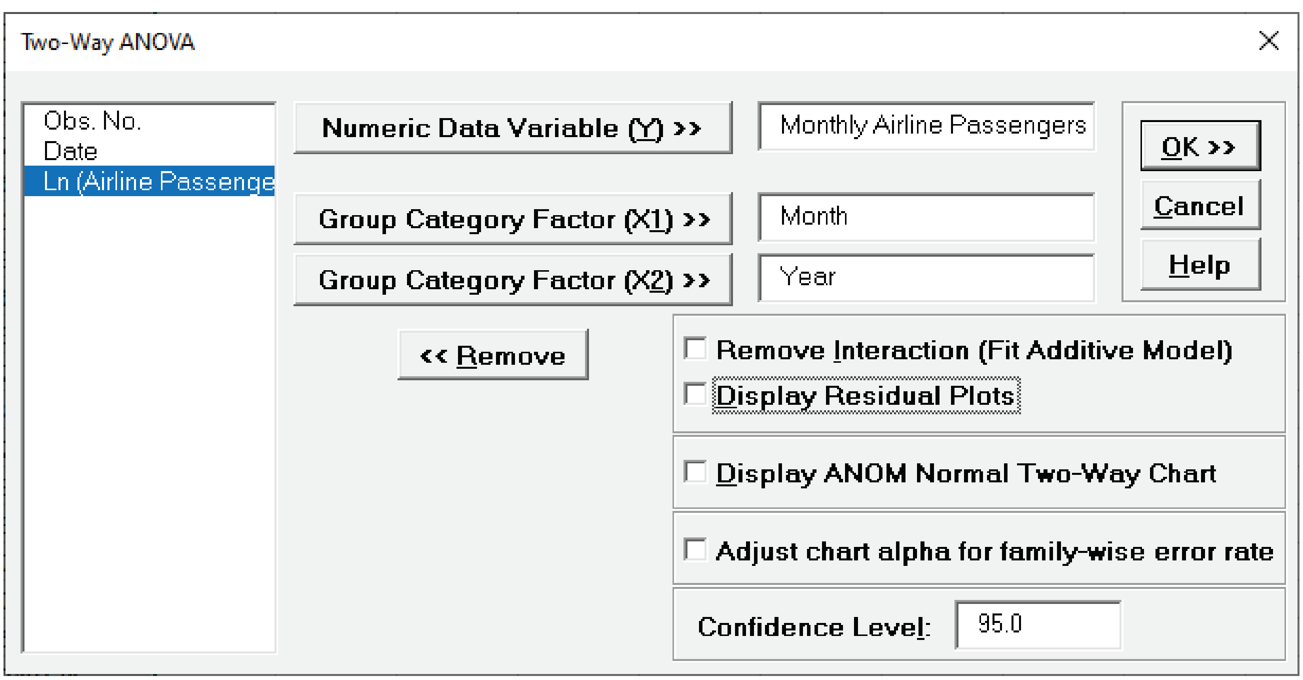
Bisgaard, S. and Kulahchi, M. (2011), *Time Series Analysis and Forecasting by Example*, Wiley, pp. 111-115.

Monthly Airline Passengers - Series G

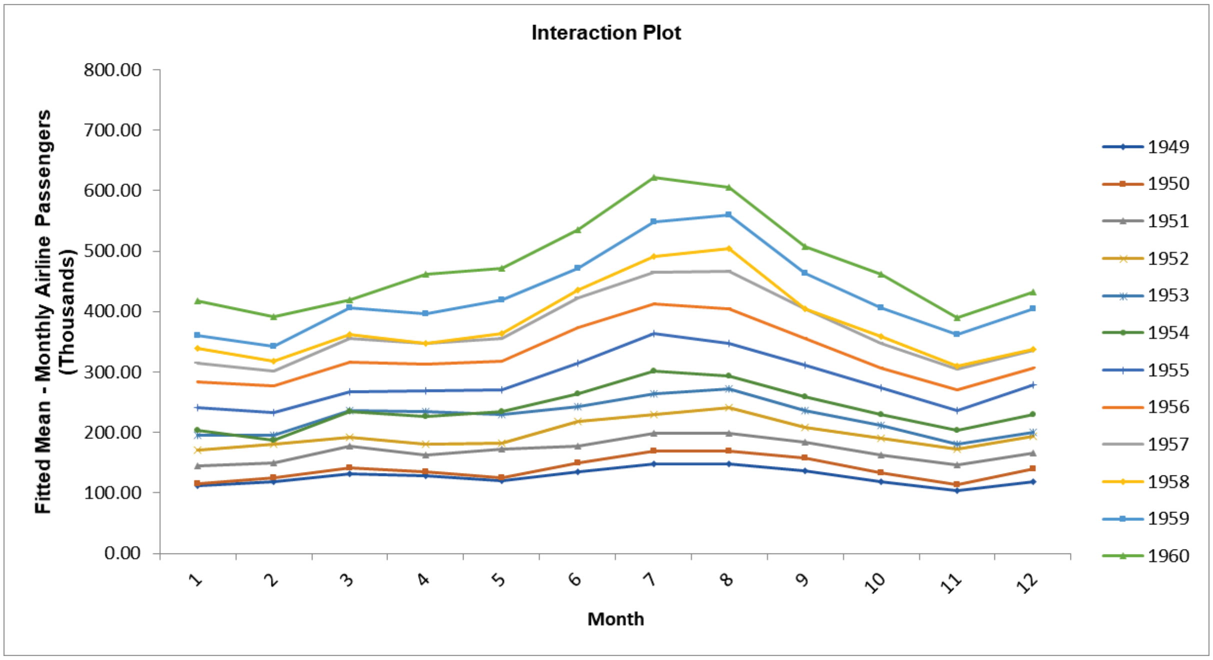
1. Open **Monthly Airline Passengers - Series G.xlsx.** Click the **Month Year for Interaction Plot** tab. This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. Month and Year columns have been added and calculated using the Excel date functions =MONTH() and =YEAR().

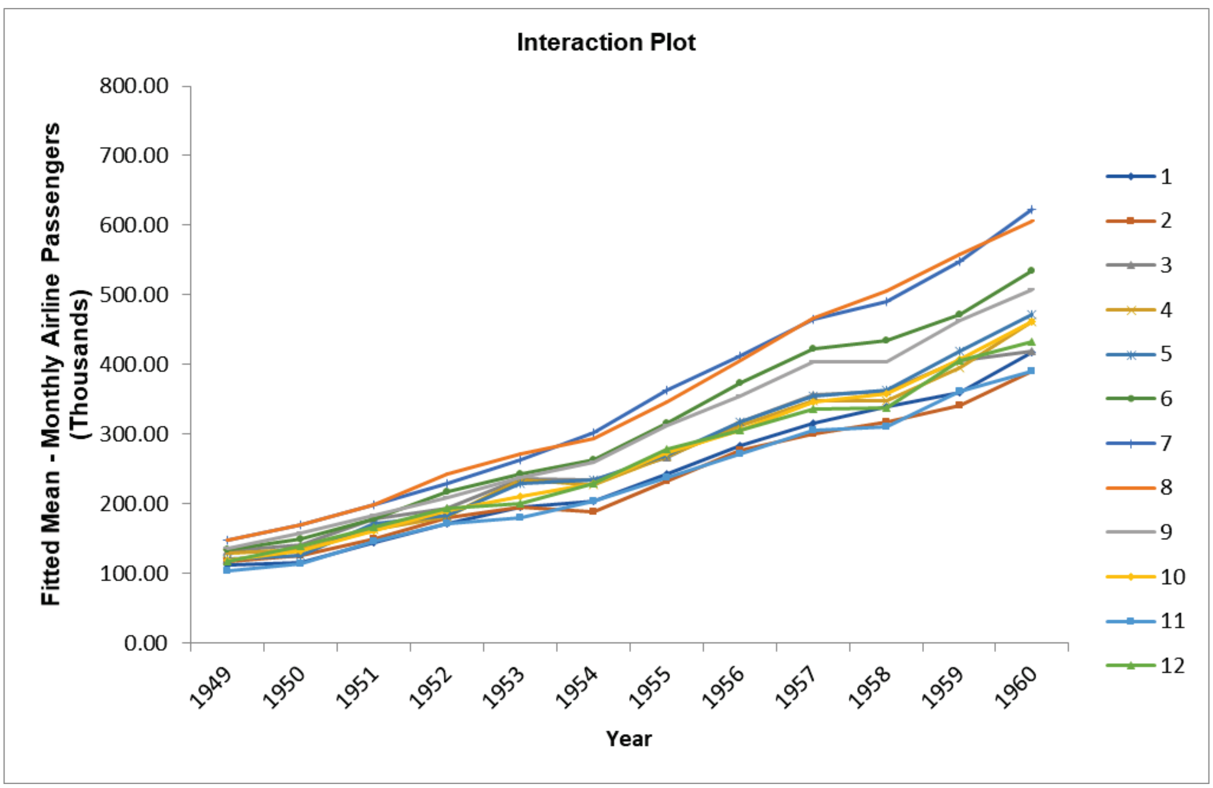
2. Click **SigmaXL > Statistical Tools > Two-Way ANOVA.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Monthly Airline Passengers*, click **Numeric Data Variable (Y) >>**. Select *Month* for **Group Category Factor (X1) >>** and Year for **Group Category Factor (X2) >>.** Uncheck all options: **Remove Interaction (Fit Additive Model)**, **Display Residual Plots**, **Display ANOM Normal Two-Way Chart** and **Adjust chart alpha for family-wise error rate.** Use the default **Confidence Level** = 95%.



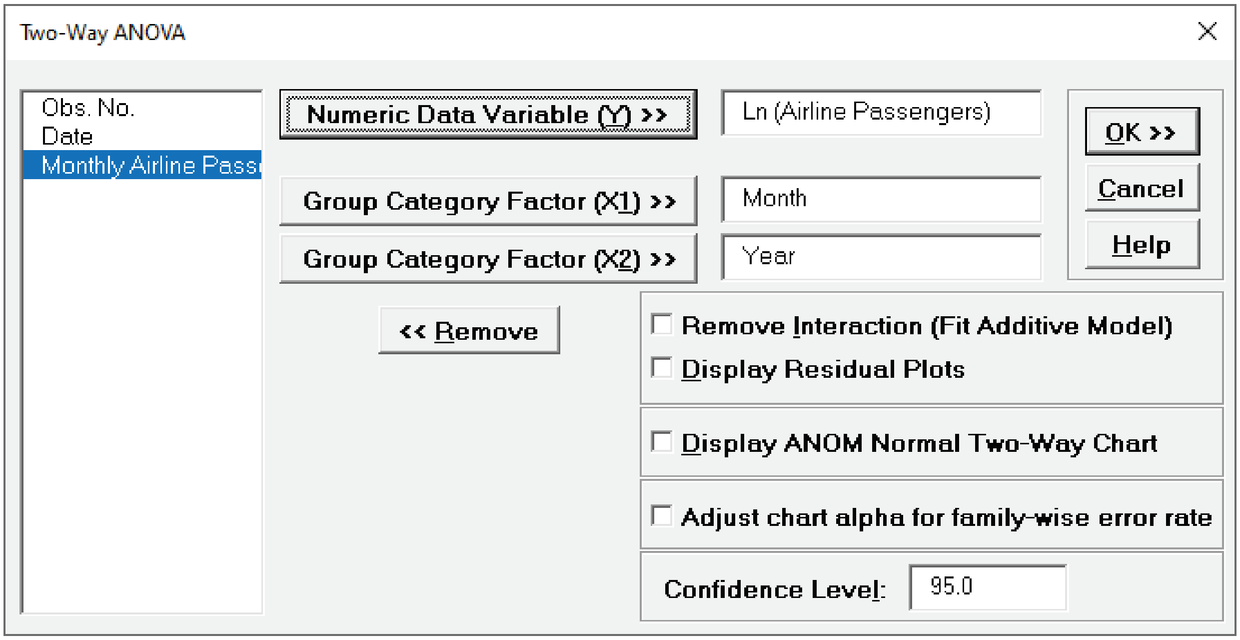
4. Click **OK**. We will not use the ANOVA report. Scroll down to the Interaction Plots. Resize to view the full legend, double click on each Y axis and set Minimum to 0.



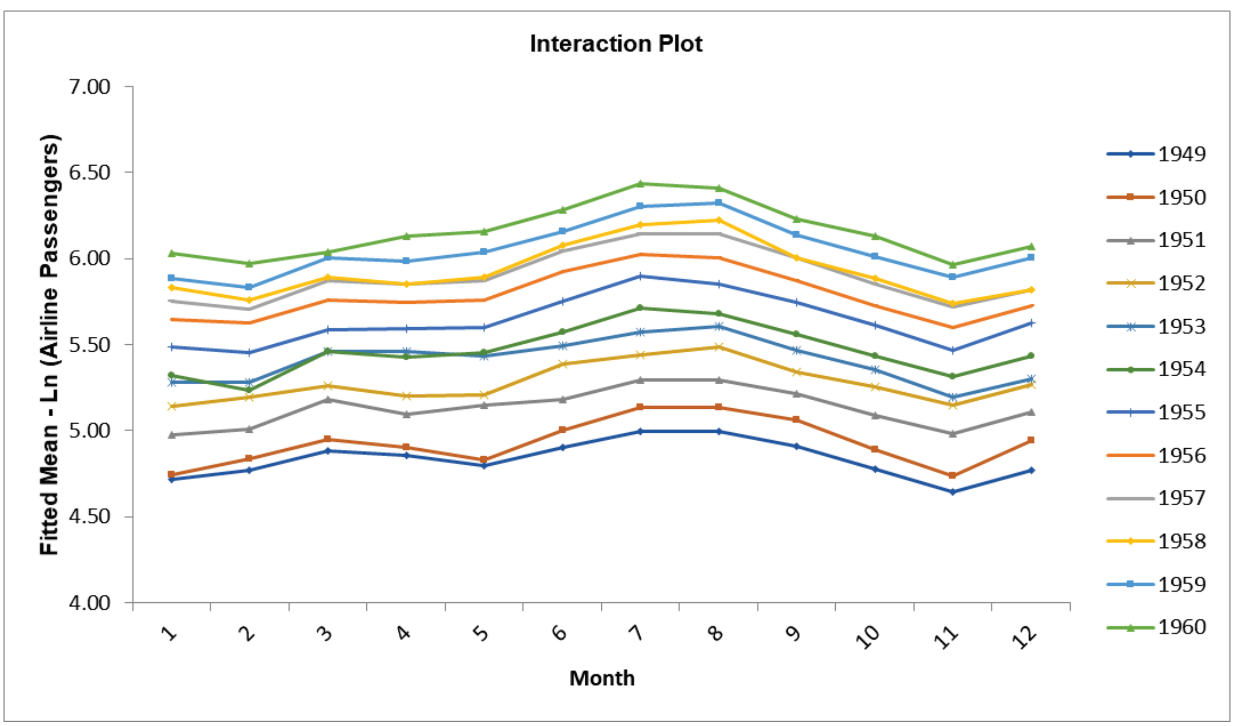


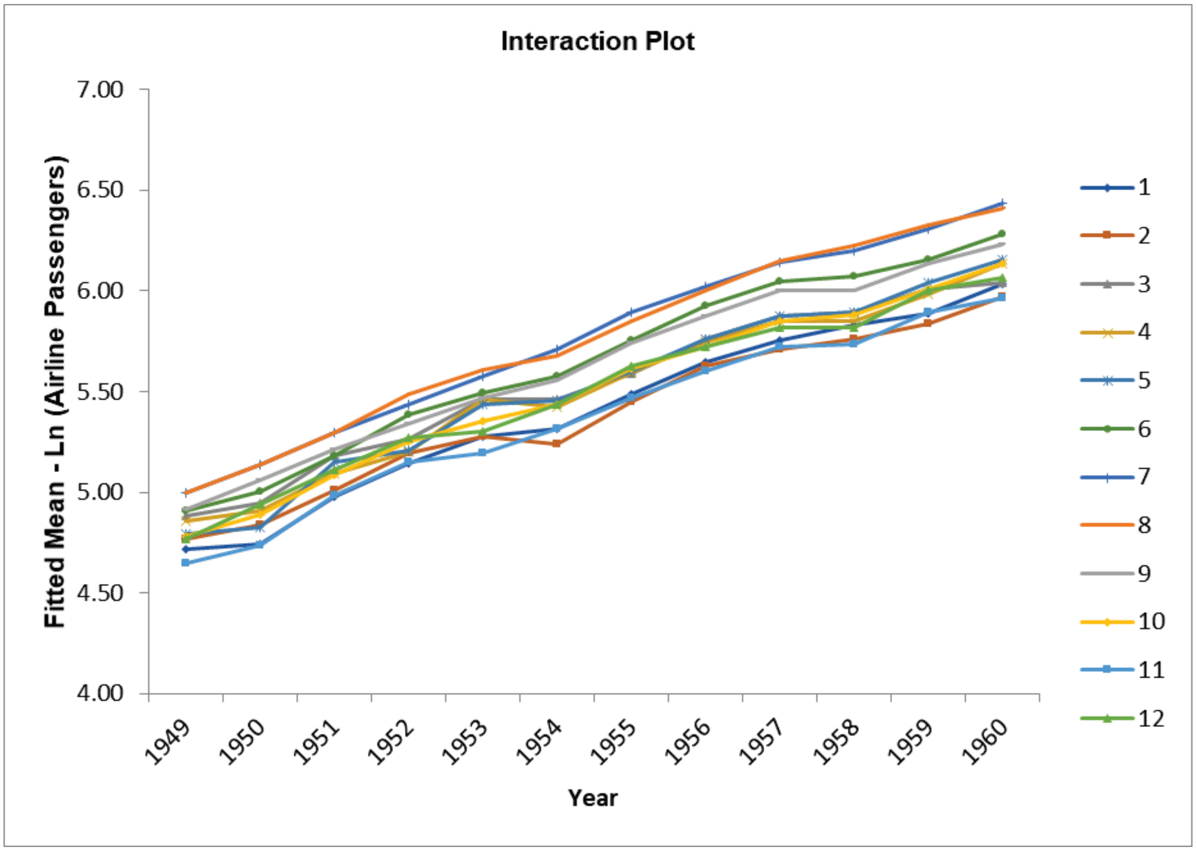
In the first interaction plot (with Month on the X axis) we can see the monthly seasonal effect and how it gets stronger by year. The second interaction plot (with Year on the X axis) shows the same increasing seasonal effect but we can also clearly see the strong positive trend by year.

5. Now we will create Seasonal Interaction Plots for Ln (Airline Passengers). Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select *Ln (Airline Passengers)* and click **Numeric Data Variable (Y) >>.**



6. Click **OK**. Scroll down to the Interaction Plots. Resize to view the full legend, double click on each Y axis and set Minimum to 4 and Maximum to 7.





In these interaction plots with Ln (Airline Passengers), we can see that the variability due to monthly seasonal effect is consistent and the yearly trend is more linear. Bisgaard and Kulahchi point out that using a traditional interpretation of interaction plots, the similar slopes indicate that the Ln transformation has effectively removed the month by year interaction, so the month and year effect is now additive.

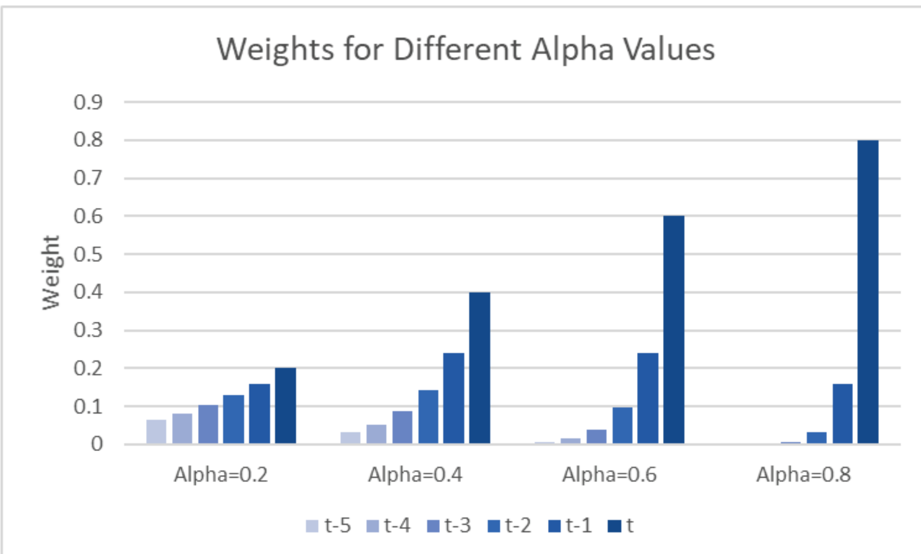
Exponential Smoothing Forecast

Simple Exponential Smoothing

Simple Exponential Smoothing forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past with the smallest weights associated with the oldest observations:

$$\hat{y}_{t+1} = \alpha\ y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \cdots$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter. This is also known as the Exponentially Weighted Moving Average (EWMA). Weights for different α values are shown graphically:



An equivalent formulation for simple exponential smoothing is:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

with the starting forecast value (initial level) $\hat{y}_1$ to be estimated and denoted as level l_0.

The smoothing parameter α and initial level l_0 are determined by minimizing the sum-of-square forecast errors (residuals):

$$\text{SSE} = \sum_{t=1}^{T}\ (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}\ e_t^2.$$

Error, Trend, Seasonal (ETS) Models

Error, Trend, Seasonal (ETS) models expand on simple exponential smoothing to accommodate trend and seasonal components as well as additive or multiplicative errors. Simple Exponential Smoothing is an Error Model. Error, Trend model is Holt's Linear, also known as double exponential smoothing. Error, Trend, Seasonal model is Holt-Winters, also known as triple exponential smoothing.

Rob Hyndman has developed a complete taxonomy that describes all of the combinations of exponential smooth models in a consistent manner (see fpp2):

- Error:
    - Additive or Multiplicative. Multiplicative error is a relative error: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$
    - The point forecasts produced by the models are identical if they use the same smoothing parameter values. Multiplicative will, however, generate different prediction intervals to accommodate change in variance.
    - An alternative to multiplicative is to use the Ln transformation (Box-Cox transformation with Lambda = 0).
    - Error models include the smoothing parameter α and initial level value.

- Trend:
    - None, Additive, Additive Damped
    - Multiplicative Trend is not recommended as it tends to produce poor forecasts, so is not included in SigmaXL
    - Trend models add a smoothing parameter β and initial trend value.
    - Damped trend models add a smoothing parameter ϕ that "dampens" the trend to a flat line some time in the future.

- Seasonal:
    - None, Additive, Multiplicative
        - The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.
    - Seasonal models add a smoothing parameter γ and initial seasonal values.
        - Maximum seasonal frequency is 24. For higher frequencies use Exponential Smoothing – Multiple Seasonal Decomposition (MSD)
        - # of initial values estimated = seasonal frequency – 1
        - constrained to sum to 0 for additive or seasonal frequency *m* for multiplicative

For further details, see Appendix: **Exponential Smoothing - ETS.**

Summary of ETS Models in SigmaXL:

Short hand (Error, Trend, Seasonal)	Method
(A, N, N)	Simple Exponential Smoothing with Additive Errors - Exponentially Weighted Moving Average (EWMA)
(M, N, N)	Simple Exponential Smoothing with Multiplicative Errors
(A, A, N)	Additive Trend Method with Additive Errors (Holt's Linear)
(M, A, N)	Additive Trend Method with Multiplicative Errors (Holt's Linear)
(A, A, A)	Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters)
(M, A, A)	Additive Trend, Additive Seasonal Method with Multiplicative Errors (Holt-Winters)
(A, N, A)	Additive Seasonal Method with Additive Errors
(M, N, A)	Additive Seasonal Method with Multiplicative Errors
(A, Ad, N)	Additive Damped Trend Method with Additive Errors
(M, Ad, N)	Additive Damped Trend Method with Multiplicative Errors
(A, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Additive Errors
(M, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Multiplicative Errors
(M, A, M)	Additive Trend, Multiplicative Seasonal Method with Multiplicative Errors (Holt-Winters)
(M, N, M)	Multiplicative Seasonal Method with Multiplicative Errors
(M, Ad, M)	Additive Damped Trend, Multiplicative Seasonal Method with Multiplicative Errors (Holt-Winters)

Exponential Smoothing Parameter Estimation, Model Statistics and Information Criteria for Model Comparison

Model parameters are solved using nonlinear maximization of the Log-Likelihood function. Log-Likelihood is related to -Ln(Sum-of-Squares Error). Information Criteria AICc, AIC and BIC are calculated using -2*Log-Likelihood and incorporate a penalty for the number of terms in the model, so smaller is better. These are used in automatic model selection. AICc is the default Information Criterion, based on forecast error performance with competition data. For further details, see Appendix: **Information Criteria for Model Comparison**.

Missing Values

Exponential Smoothing handles missing values with seasonally adjusted linear interpolation. While there is robustness against some missing values, if the number of missing values is large then model estimation and forecast accuracy will be degraded. Upon selection of a time series with missing values, you will see a pop-up "Warning: Missing values detected. Seasonally adjusted linear interpolation will be used." See Appendix: **Seasonally Adjusted Linear Interpolation of Missing Values**.

Forecast Accuracy Metrics

SigmaXL uses the following forecast accuracy metrics:

Root mean squared error: RMSE $= \sqrt{\text{mean}(e_t^2)}$

Mean absolute error: MAE $= \text{mean}(|e_t|)$

Mean absolute percentage error: MAPE $= \text{mean}\left(\left|\frac{100e_t}{y_t}\right|\right)$

Mean absolute scaled error: MASE $= \text{mean}(|e_t|)/scale$

where *scale* is the MAE of the in-sample naïve or seasonal naïve forecast (set all forecasts to be the value of the last observation/period). Note that scale for nonseasonal is identical to MR-bar of an Individuals Moving Range chart. A scaled error is less than one if it arises from a better forecast than the average naïve/seasonal naïve forecast. Conversely, it is greater than one if the forecast is worse than the average naïve forecast. MASE is recommended over the popular MAPE, because MAPE becomes infinite if any y_t=0. For further details, see Appendix: **Forecast Accuracy**.

In-Sample is also referred to as the "Train" data. The same metrics may be applied to Out-of-Sample (Withhold), also referred to as the "Test" data and may be One-Step-Ahead or Multi-Step-Ahead. This will be demonstrated later. Out-of-Sample (Withhold) data is not used in the model parameter estimation, so is a much better indicator of true forecast accuracy. In-sample accuracy metrics can be biased due to overfitting. Scale is always computed using the in-sample data.

Demo of Simple Exponential Smoothing – Concentration

1. Open **Demo of Simple Exponential Smoothing – Concentration.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals in Column B.

	A	B	C	D
1	Obs. No.	Concentration (y)	One-step-ahead forecast (y-hat)	Residuals (e)
2	1	17	17	0
3	2	16.6	17	-0.4
4	3	16.3	16.96	-0.66
5	4	16.1	16.894	-0.794
6	5	17.1	16.8146	0.2854
7	6	16.9	16.84314	0.05686
8	7	16.8	16.848826	-0.048826
9	8	17.4	16.8439434	0.5560566
10	9	17.1	16.89954906	0.20045094
11	10	17	16.91959415	0.080405846

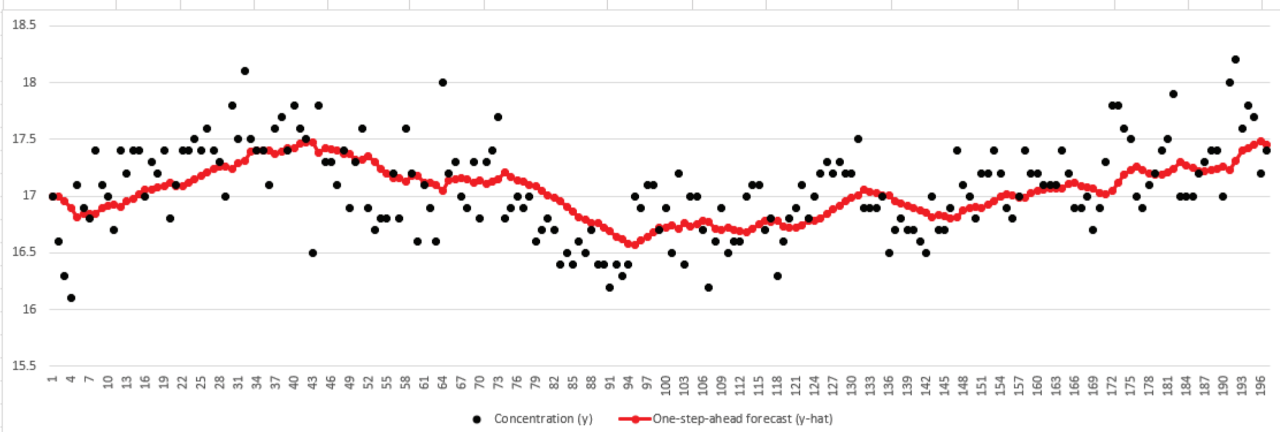
Exponential Smoothing Model Parameters	
alpha	0.1
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	21.8341

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-303.72228
AICc	613.56892
AIC	613.44457
BIC	623.29418

In-Sample Forecast Accuracy Metrics	
RMSE	0.332916231
MAE	0.271930445
MAPE	1.592107249
Scale (MR-bar)	0.275510204
MASE	0.9870068



Cell **G3** is the alpha smoothing parameter value, cell **G4** is the initial level. These may be manually entered or solved with Solver. The One-step-ahead forecast is calculated as follows: C2 is the initial level value; C3 =alpha*B2+(1-alpha)*C2; C4 =alpha*B3+(1-alpha)*C3, etc.

Cell **G7** is the Residuals Sum-of-Squares error and is the value to be minimized, in order to produce the most accurate one-step-ahead forecast. The exponential smooth model statistics are also given for later comparison to the SigmaXL report.

The raw Concentration data are the black dots in the graph; the one-step-ahead forecast values are the red dots. Model Statistics and Forecast Accuracy Metrics will be discussed later.

2. Before we use Solver to determine the optimal alpha and initial level values, we will manually enter some values to see how the graph changes. Enter alpha = 0.01 as shown:

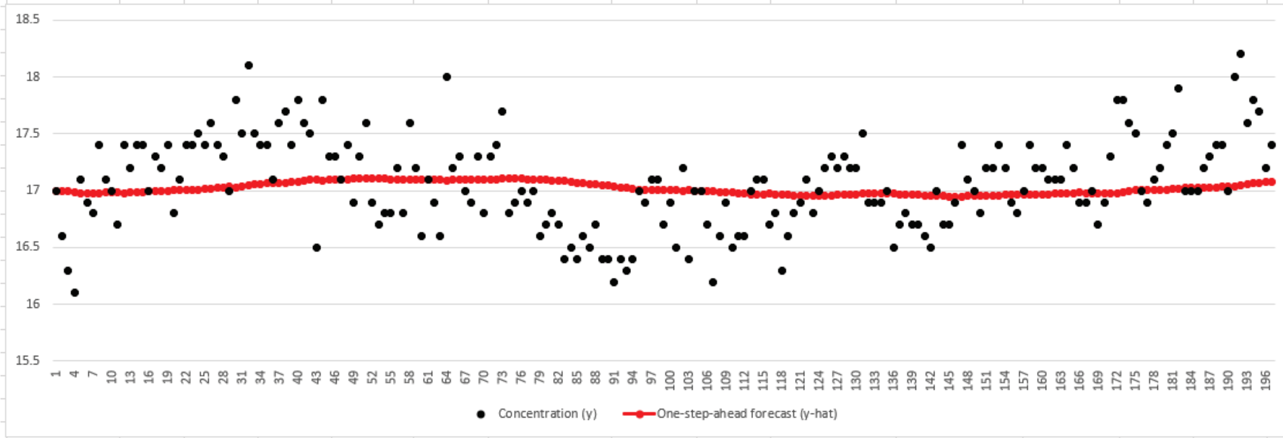
Exponential Smoothing Model Parameters	
alpha	0.01
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	31.022

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-338.31756
AICc	682.75947
AIC	682.63511
BIC	692.48472

In-Sample Forecast Accuracy Metrics	
RMSE	0.396827397
MAE	0.318831251
MAPE	1.864660975
Scale (MR-bar)	0.275510204
MASE	1.157239356



This small value of alpha produces a smooth fit with SSE = 31.02 that is larger than what we started with (SSE = 21.83).

3. Now enter alpha = 0.99 as shown:

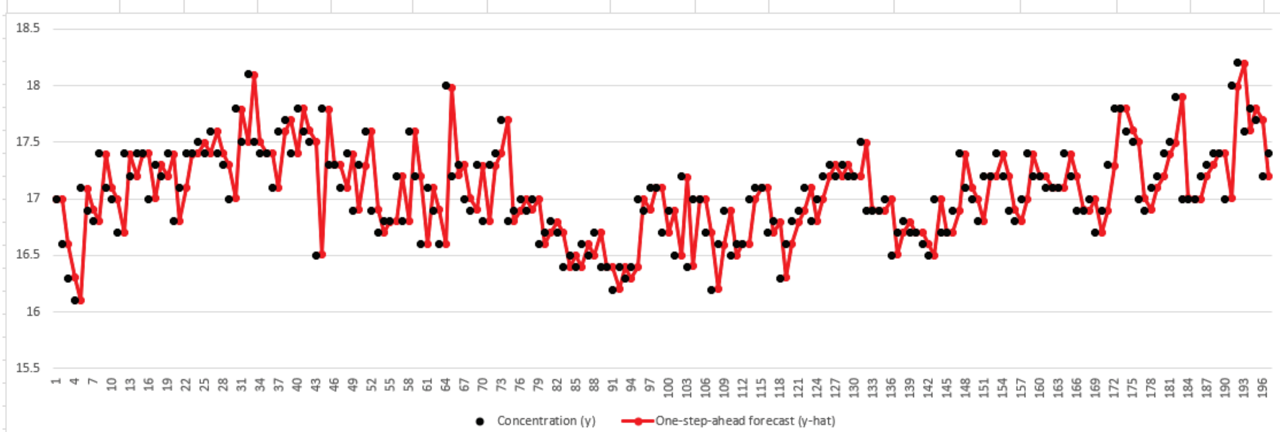
Exponential Smoothing Model Parameters	
alpha	0.99
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	26.5219

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-322.88029
AICc	651.88493
AIC	651.76058
BIC	661.61019

In-Sample Forecast Accuracy Metrics	
RMSE	0.366918467
MAE	0.273380342
MAPE	1.598165452
Scale (MR-bar)	0.275510204
MASE	0.992269389



This large value of alpha results approximately in a naïve forecast, with the one-step-ahead forecast value equal to the previous actual. SSE = 26.52 is larger than what we started with.

4. Enter alpha = 0.1 and initial level = 20 as shown:

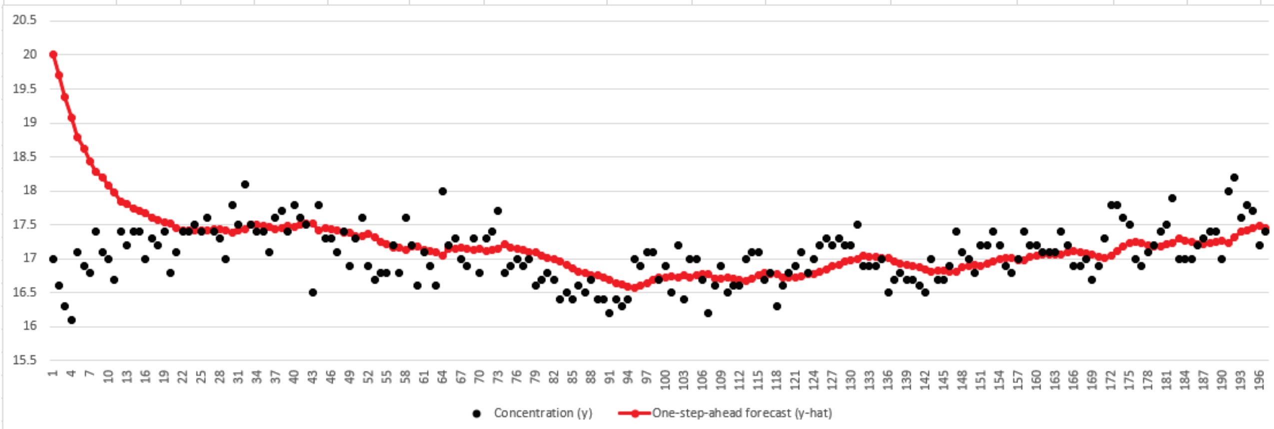
Exponential Smoothing Model Parameters	
alpha	0.1
initial level	20

Use Solver to obtain parameter values that minimize SSE:

SSE	71.0199

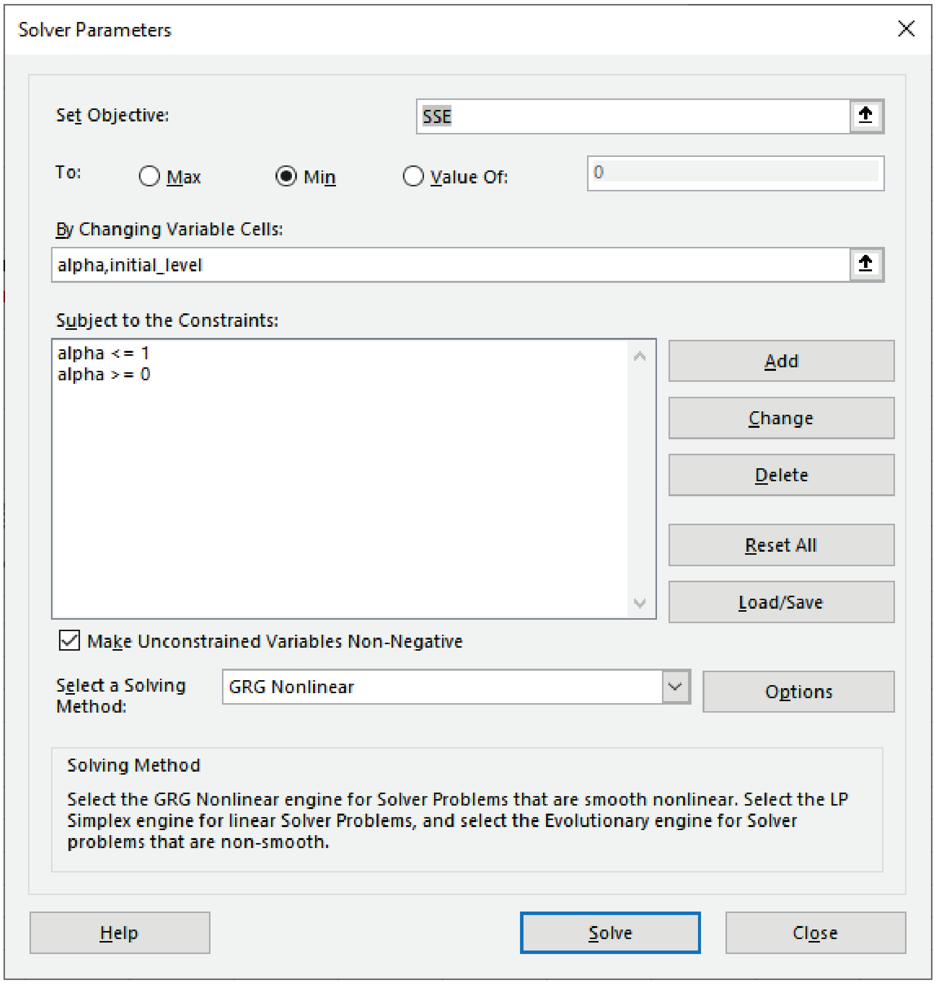
Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-419.90158
AICc	845.92751
AIC	845.80316
BIC	855.65277

In-Sample Forecast Accuracy Metrics	
RMSE	0.60042246
MAE	0.365989697
MAPE	2.156801395
Scale (MR-bar)	0.275510204
MASE	1.32840705



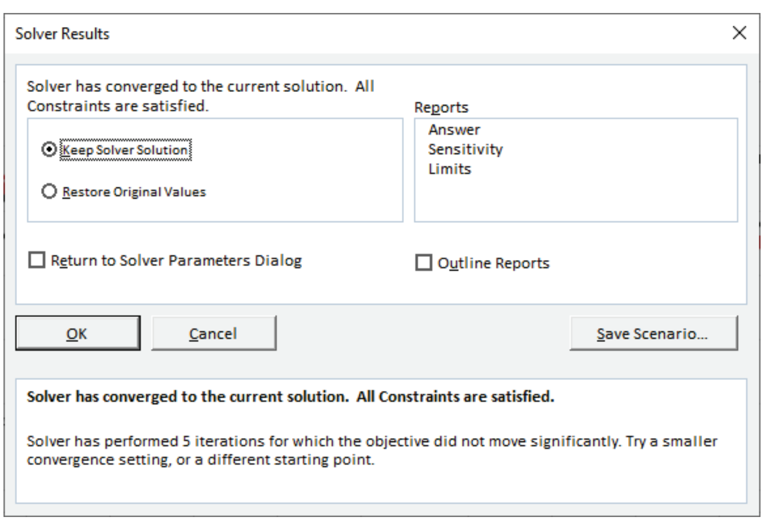
This initial level is a poor estimate of the start value and results in SSE = 71.02. After about 30 observations, the influence of this poor start value is negligible. The duration for which the initial level has influence on the one-step-ahead forecast depends on alpha, a smaller alpha would be a longer duration; a larger alpha would be a shorter duration.

5. Now we will use Excel's Solver to find the optimal values for alpha and initial level. Click **Data > Solver**. (If the Solver menu is not available, click **File > Options > Add-Ins**. Click **Go**… Manage Excel Add-ins, check **Solver Add-in**. Click **OK**. Click **Data > Solver**.)



The cell addresses, constraints and settings for Solver have been stored in the workbook, so no changes are necessary. Solver will minimize cell **G7** (SSE), by varying cells **G3** (alpha) and **G4** (initial_level). The solving method is GRG Nonlinear; Evolutionary could also be used but it is slower.

6. Click **Solve**. The Solver Results dialog shows that a solution has been found.



7. Click **OK**. The SSE is now 19.74; alpha value = 0.295 and initial level = 16.732.

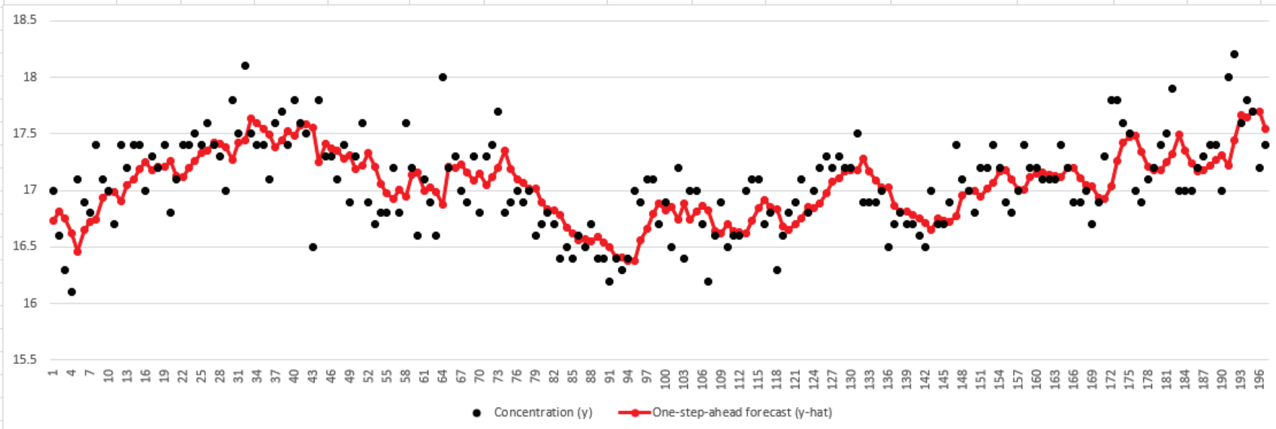
Exponential Smoothing Model Parameters	
alpha	0.29491
initial level	16.7319

Use Solver to obtain parameter values that minimize SSE:

SSE	19.7426

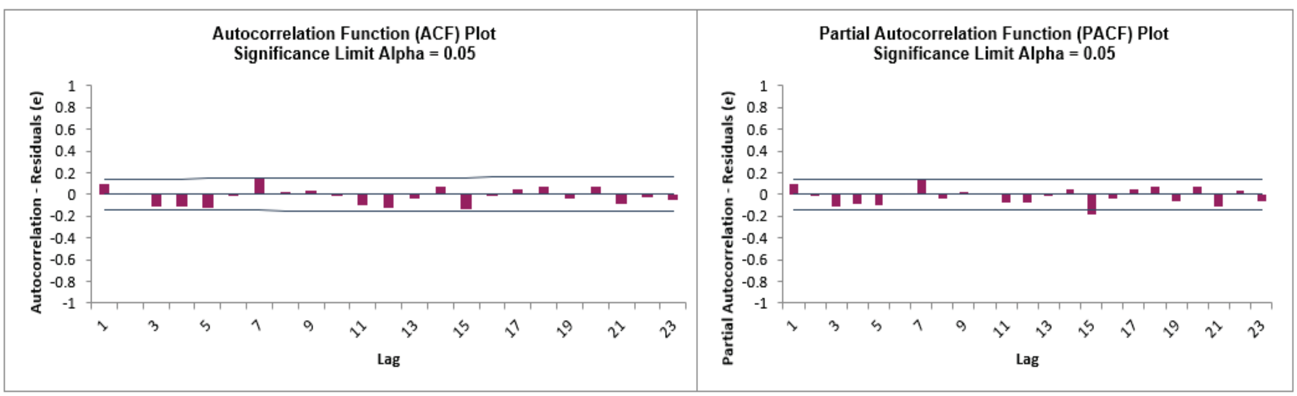
Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-293.8036
AICc	593.73155
AIC	593.6072
BIC	603.45681

In-Sample Forecast Accuracy Metrics	
RMSE	0.316569323
MAE	0.247320186
MAPE	1.446469166
Scale (MR-bar)	0.275510204
MASE	0.897680675



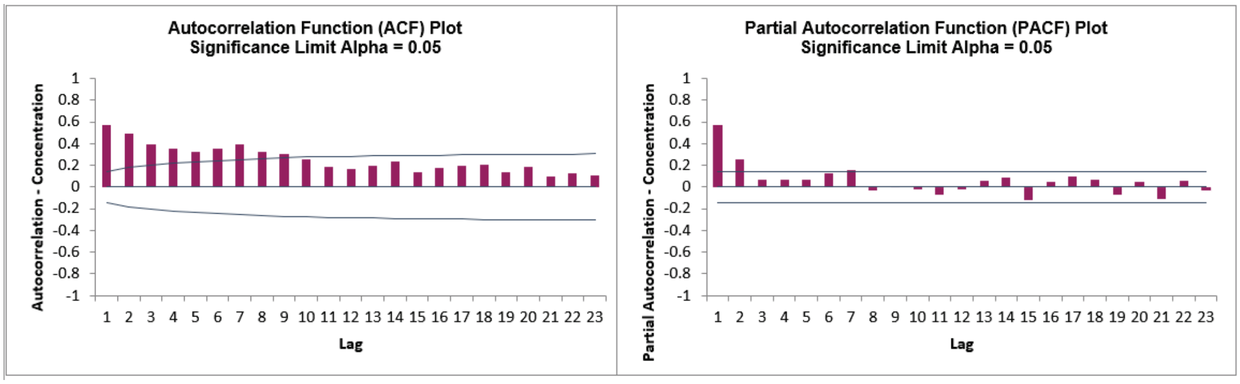
By minimizing the residual sum-of-squares error (SSE), we have our best estimate of the parameters to produce one-step-ahead forecast values.

8. We will now produce ACF/PACF plots for the Residuals. Select cells **D1:D198** and click **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Click **Next**, Select *Residuals (e)*, Click **OK**.

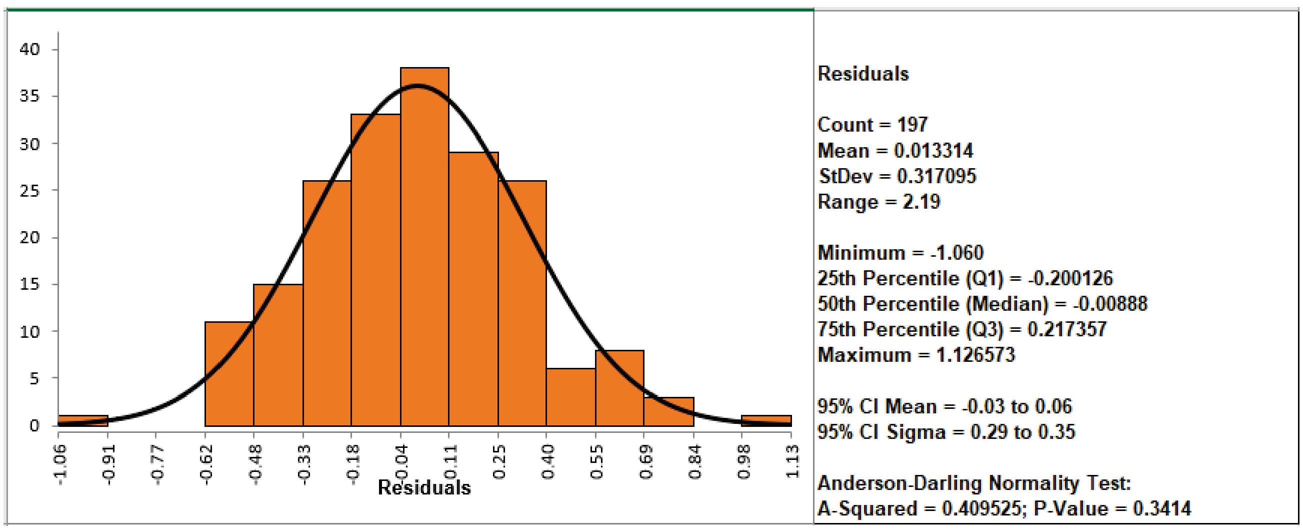


We can see that all of the autocorrelation has been removed by the simple exponential smoothing model (with the exception of lag 15 in the PACF), so this is a good fit to the time series data.

9. By way of comparison, these are the ACF/PACF plots for the original Concentration data that we produced earlier.



10. We will now have a quick look at the Residuals using a Histogram. Select the Residuals data (D1:D198). Click **SigmaXL > Graphical Tools > Histograms and Descriptive Statistics**. **Next**, Select *Residuals (e)*, Click **OK**.



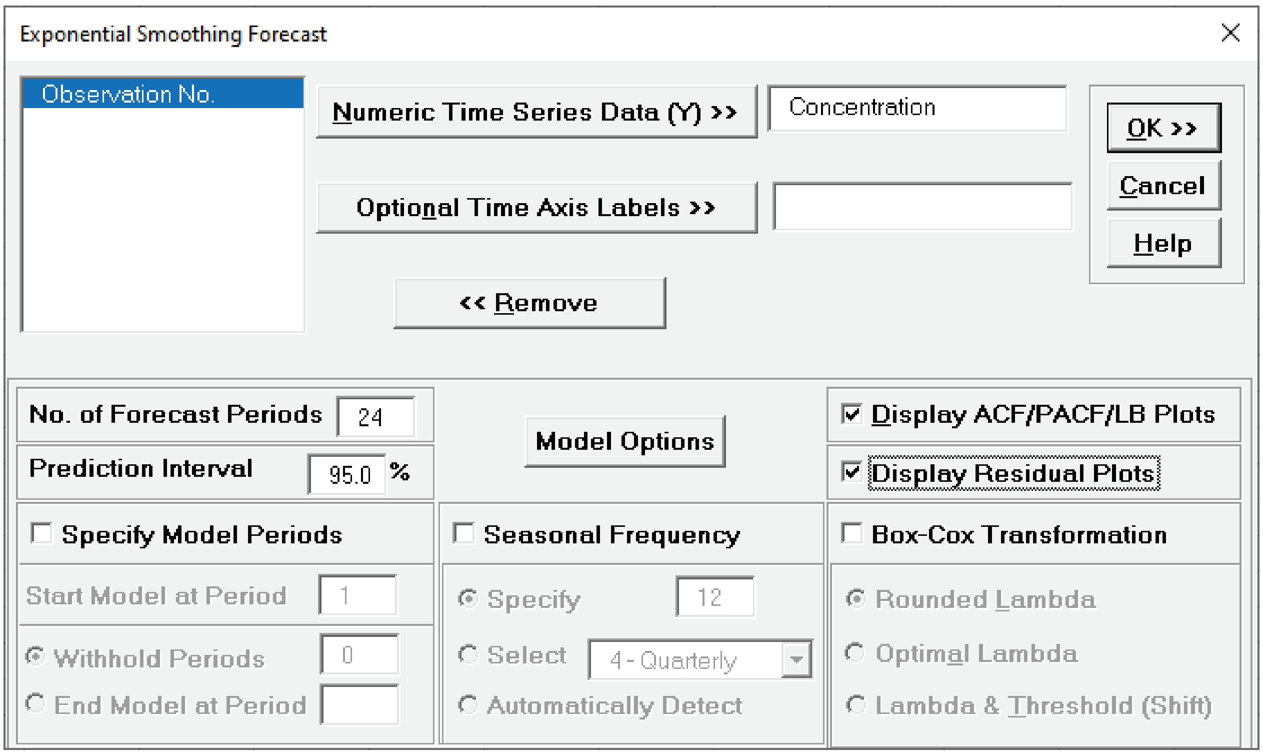
The residuals are normally distributed with no obvious extreme outliers.

Chemical Process Concentration - Series A

11. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

12. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

13. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Leave **Specify Model Periods, Seasonal Frequency** and **Box-Cox Transformation** unchecked. We will use the default **No. of Forecast Periods** = 24 and **Prediction Interval** = 95.0 %.



- **Optional Time Axis Labels** will be displayed on the forecast chart time axis. If used, dates for the forecast periods should also be included, otherwise the time axis will be blank for the forecast periods.

- **No. of Forecast Periods** are the number of time series values to be predicted (forecast horizon). The most accurate forecast will be for the first predicted value (one-step-ahead).

- **Prediction Interval** % is the confidence level for the individual predictions. For example, a 95% prediction interval contains a range of values which should include the actual future value with 95% probability. The interval will get larger the further out you predict.

- **Model Options** opens another dialog which allows you to set automatic options or to specify a model.

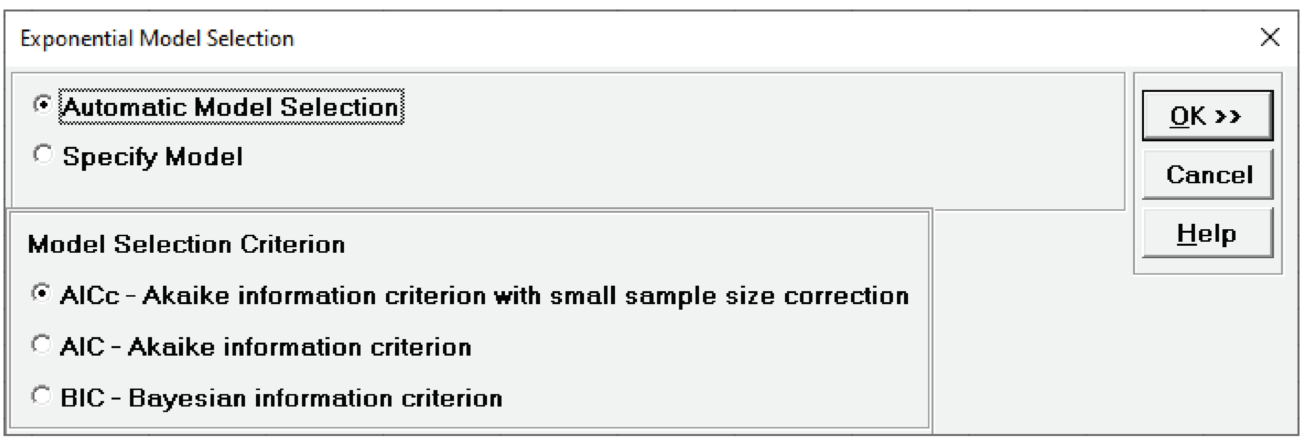
- **Display ACF/PACF/LB** option will produce ACF and PACF plots for the raw data as well as for the model residuals. The LB plot is a plot of Ljung-Box test P-Values for various lags and is used to determine if a group of autocorrelations are significant, (i.e., the autocorrelations do not come from a white noise series). For further details, see **Ljung-Box Test**.

- **Display Residual Plots** will produce a table of model residuals and the usual model residual plots: histogram, normal probability plot, residuals versus data order, and residuals vs forecast value. Note that if a Box-Cox transformation is applied, the residuals are transformed so will not be equal to forecast - actual.

- **Specify Model Periods** are used to specify a start period, end period or withhold sample. The withheld data is not used in model estimation, so this is very useful for model validation and comparison. This will be used in a later example.

- **Seasonal Frequency** and **Box-Cox Transformation** will be used in a later example.

14. Click **Model Options**.

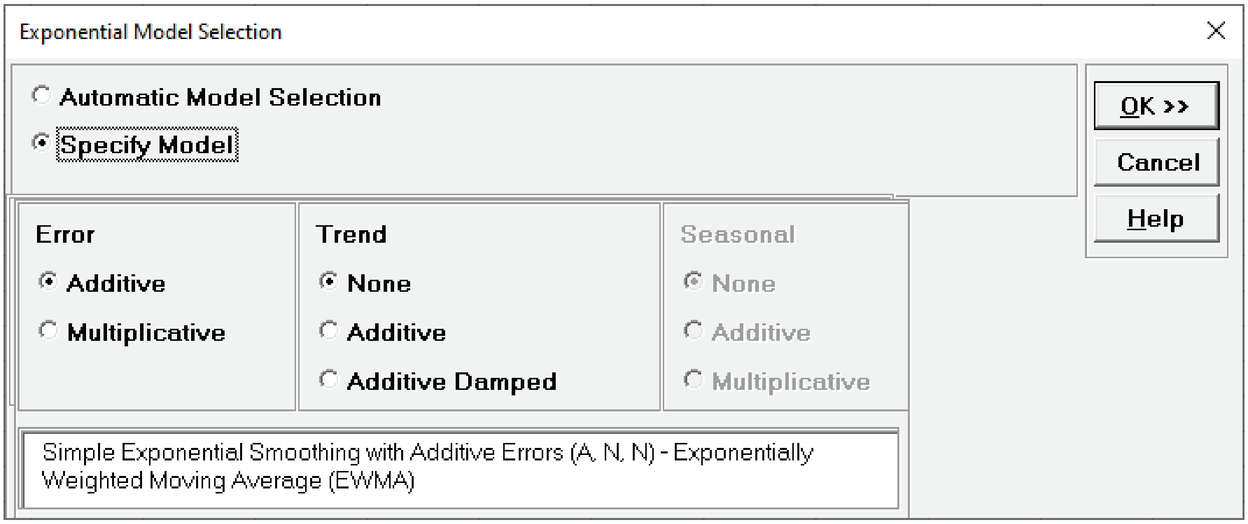


- **Automatic Model Selection** will be used later. It is the default selection.

- **Model Selection Criterion** is the information criterion metric to be used in automatic model selection. AICc is the default selection.

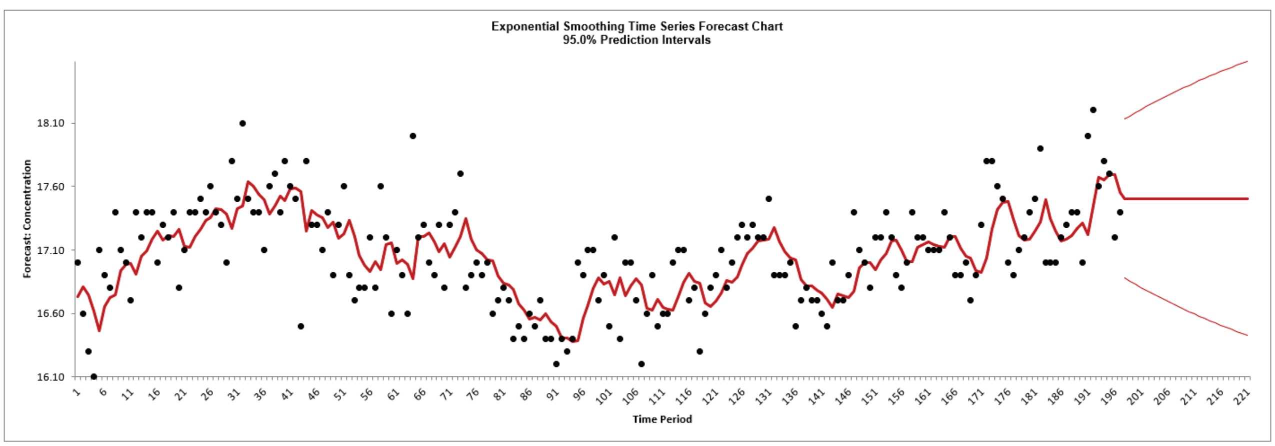
- Clicking **OK** accepts the settings and returns you to the previous dialog. Clicking **Cancel** will cancel any changes and return you to the previous dialog.

15. Select **Specify Model**.



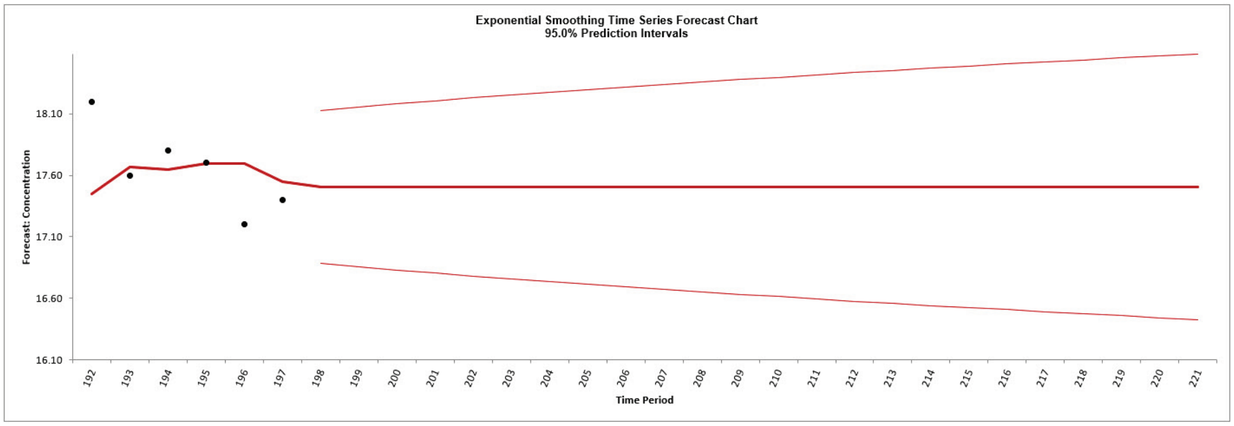
- **Specify Model** allows you to manually specify the Error-Trend-Seasonal model. Seasonal options are greyed out because the **Seasonal Frequency** option in the main dialog is unchecked. A summary description of the model is also given. This will be included in the forecast report.

16. We will use the default **Error: Additive** and **Trend: None**, which is a simple exponential smoothing model, the same as was demonstrated previously. Click **OK** to return to the Exponential Smoothing Forecast dialog. Click **OK**. The exponential smoothing forecast chart is given:



This is very similar to the exponential smooth plot demonstrated above, showing the raw Concentration data (black) and one-step-ahead forecast values (red), but with the addition of a 24-period forecast and the 95% prediction interval.

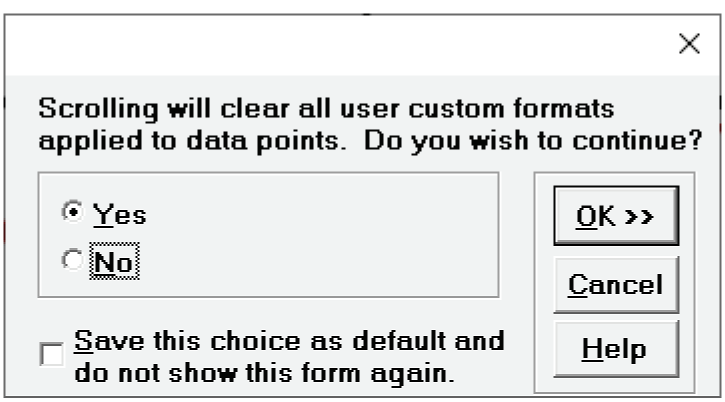
17. You can zoom in to view the last 30 points on the Forecast Chart. Click **SigmaXL Chart Tools > Show Last 30 Data Points**.



18. To restore the chart, click **SigmaXL Chart Tools > Show All Data Points**.

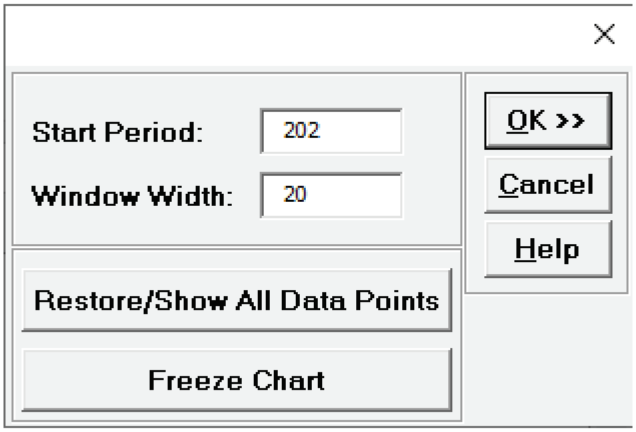
19. You can also scroll through the data. Click **SigmaXL Chart Tools > Enable Scrolling**

You are prompted with a warning message that custom formatting on the chart will be cleared:



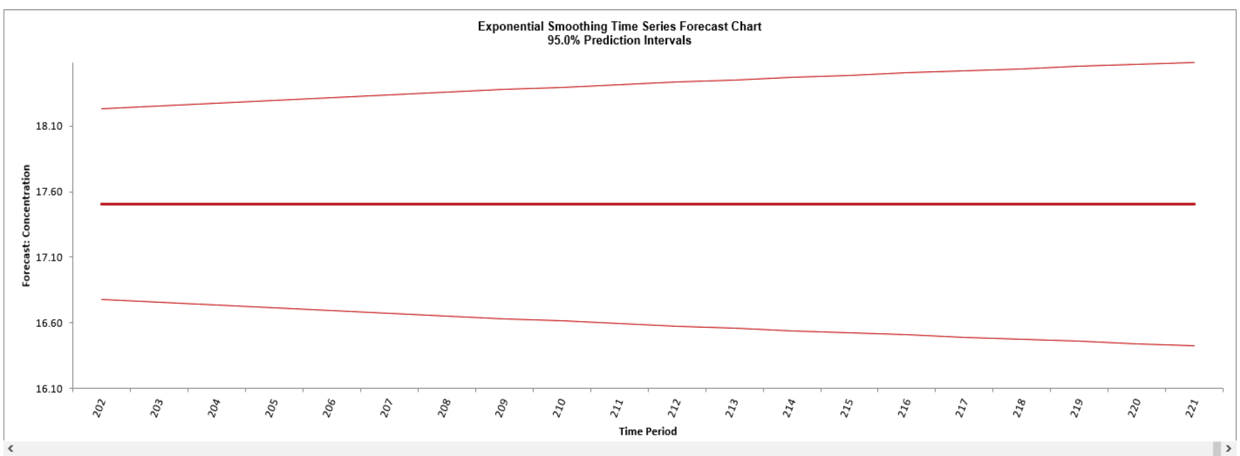
You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

20. Click **OK.** The scroll dialog appears allowing you to specify the **Start Period** and **Window Width**:



At any point, you can click **Restore/Show All Data Points** or **Freeze Chart**. Freezing the chart will remove the scroll and unload the dialog. The scroll dialog will also unload if you change worksheets. To restore the dialog, click **SigmaXL Chart Tools > Enable Scrolling**.

21. Click **OK.** A scroll bar appears beneath the forecast chart. You can also change the **Start Subgroup** and **Window Width** and **Update**.



You can scroll through by clicking to the right or left, with the specified window width of 20. Click left once to view the chart as shown.

22. Click **Cancel** to exit the scroll dialog.

23. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Forecast: Simple Exponential Smoothing with Additive Errors (A, N, N) - Exponentially Weighted Moving Average (EWMA) - User Specified Model
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

The model **Simple Exponential Smoothing with Additive Errors (A, N, N) – Exponentially Weighted Moving Average (EWMA)** is the user specified model. This is the same model information that was displayed in the Model Selection dialog.

If we had checked Specify Model Periods in the main dialog, the start, end or withhold selection would be summarized here as well.

24. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	1
Model Selection Criterion	Specified
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of model information with Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = "Specified" because the model was user specified; and Box-Cox Transformation = "N/A" because Box-Cox Transformation was unchecked.

25. The Parameter Estimates are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.294785988
l (initial level)	16.73121246

The parameter estimates closely match the values obtained earlier in the demonstration using Solver:

Exponential Smoothing Model Parameters	
alpha	0.29491
initial level	16.7319

The slight differences in parameter results are due to differences in optimization method.

26. The Exponential Smoothing Model Statistics are:

Exponential Smoothing Model Statistics	
No. of Observations	197
DF	195
StDev	0.31819
Variance	0.10124
Log-Likelihood	-293.804
AICc	593.732
AIC	593.607
BIC	603.457

Degrees of freedom (DF) = *n* – 2 (terms in the model). See Exponential Smoothing Parameter Estimation, Model Statistics and Information Criteria for Model Comparison.

The results match those given in the demonstration using Solver:

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-293.8036
AICc	593.73155
AIC	593.6072
BIC	603.45681

The equations may be viewed by clicking on the Demo cells **M5**, **M6**, **M7** or **M8**.

27. The In-Sample Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	197	
RMSE	0.316569334	
MAE	0.247329038	
MAPE	1.446520183	
MASE	0.897712804	

MASE is less than one, so it is a better forecast than would be obtained from a naïve forecast (set all forecasts to be the value of the last observation). See Forecast Accuracy Metrics. The results closely match those given in the demonstration using Solver:

In-Sample Forecast Accuracy Metrics	
RMSE	0.316569323
MAE	0.247320195
MAPE	1.446469239
Scale (MR-bar)	0.275510204
MASE	0.897680708

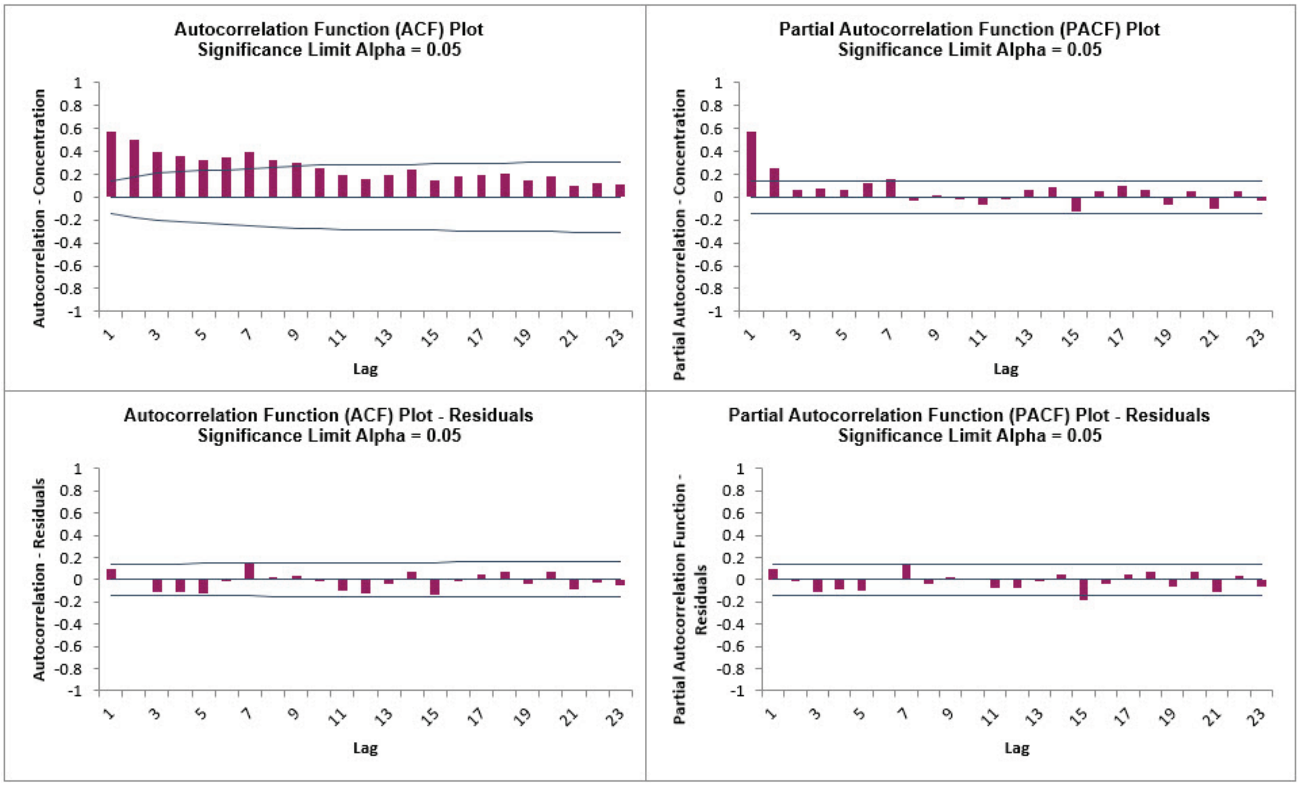
The equations may be viewed by clicking on the Demo cells **P3** to **P7**.

28. The Forecast Table is given as:

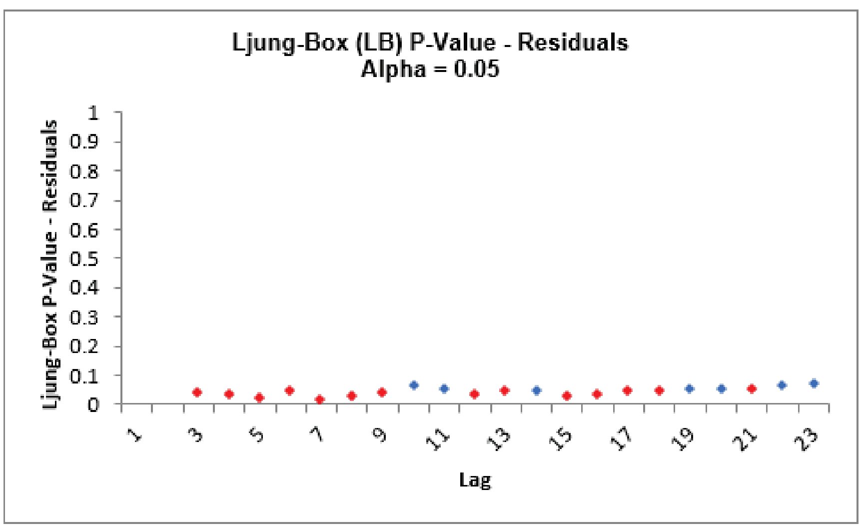
Forecast Table				
Period	Withhold Data	Forecast	Lower 95.0% PI	Upper 95.0% PI
198		17.50540234	16.8817641	18.12904059
199		17.50540234	16.8552318	18.15557289
200		17.50540234	16.82974058	18.1810641
201		17.50540234	16.80517674	18.20562795
202		17.50540234	16.78144587	18.22935881
203		17.50540234	16.75846858	18.25233611
204		17.50540234	16.73617733	18.27462736
205		17.50540234	16.7145141	18.29629058
206		17.50540234	16.69342865	18.31737604
207		17.50540234	16.67287705	18.33792764
208		17.50540234	16.65282071	18.35798398
209		17.50540234	16.63322546	18.37757923

These are the same forecast and prediction interval values displayed in the Forecast Chart but provided for further analysis or charting. If Withhold Periods are specified, the Withhold Data will be displayed as well.

29. Click on the **Exp. Smooth. ACF PACF LB** sheet to view the ACF/PACF/LB Plots:



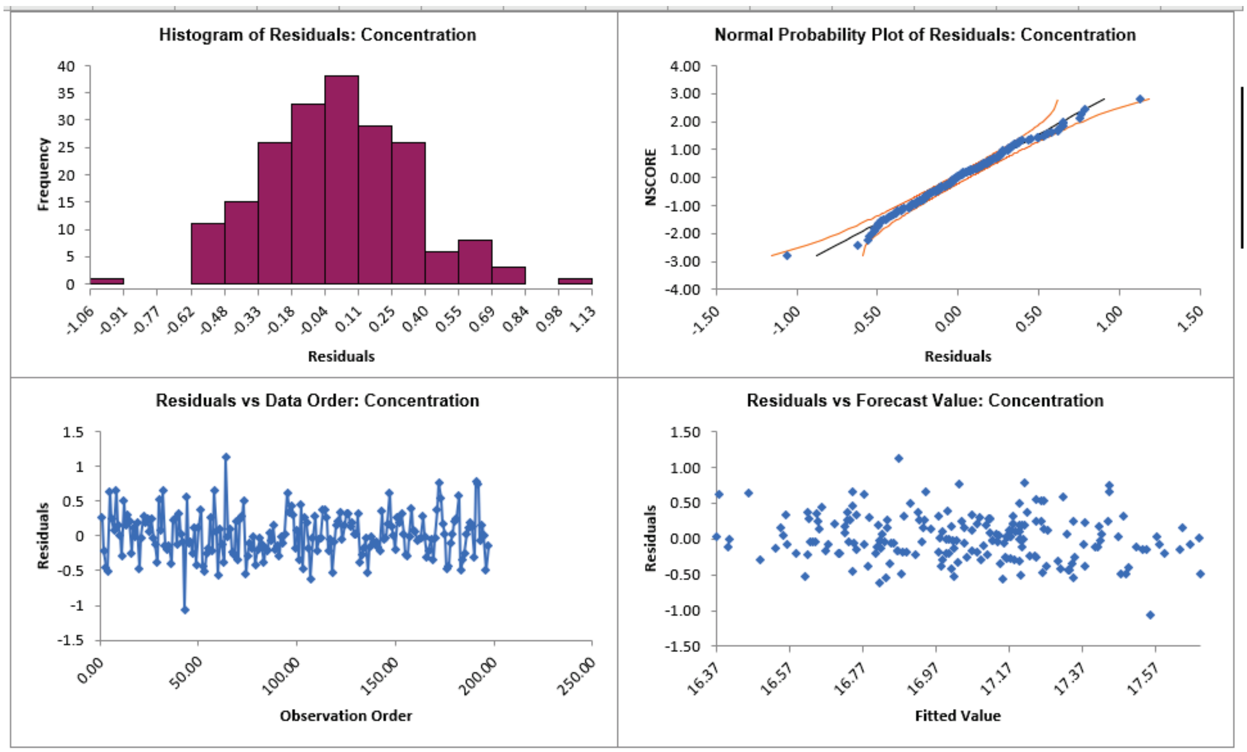
These match the plots that we obtained previously in the Demo. We can see that all of the autocorrelation has been removed by the exponential smoothing model (with the exception of lag 15 in the PACF), so this is a good fit to the time series data.



The LB plot is a plot of Ljung-Box test P-Values for various lags and is used to determine if a group of autocorrelations are significant, (i.e., the autocorrelations do not come from a white noise series). For further details, see **Ljung-Box Test**.

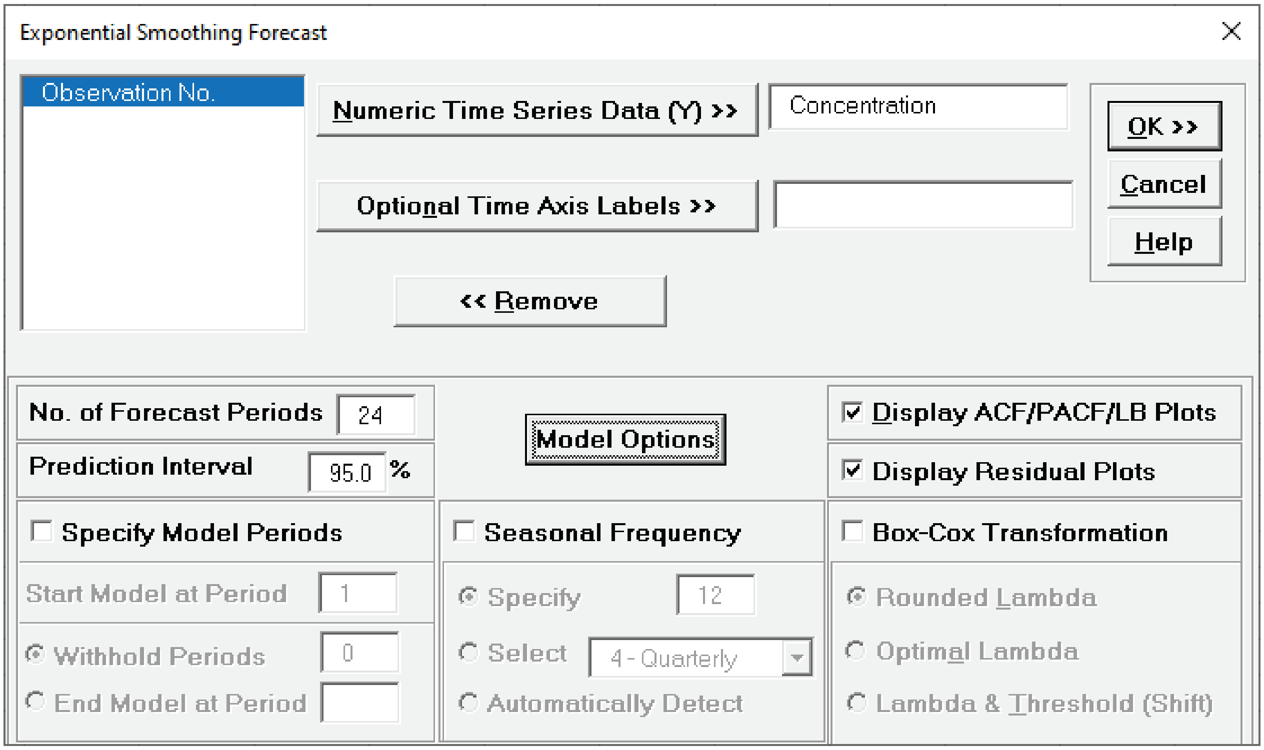
The red P-Values are significant (alpha=.05) and the blue P-Values are not significant. It is desirable that all P-Values be blue. The ACF/PACF plots indicated that almost all of the correlation has been accounted for in the model, but the Ljung-Box plot shows that some significant autocorrelation still remains - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

30. Click on the **Exp. Smoothing Residuals** sheet to view the Residual Plots:

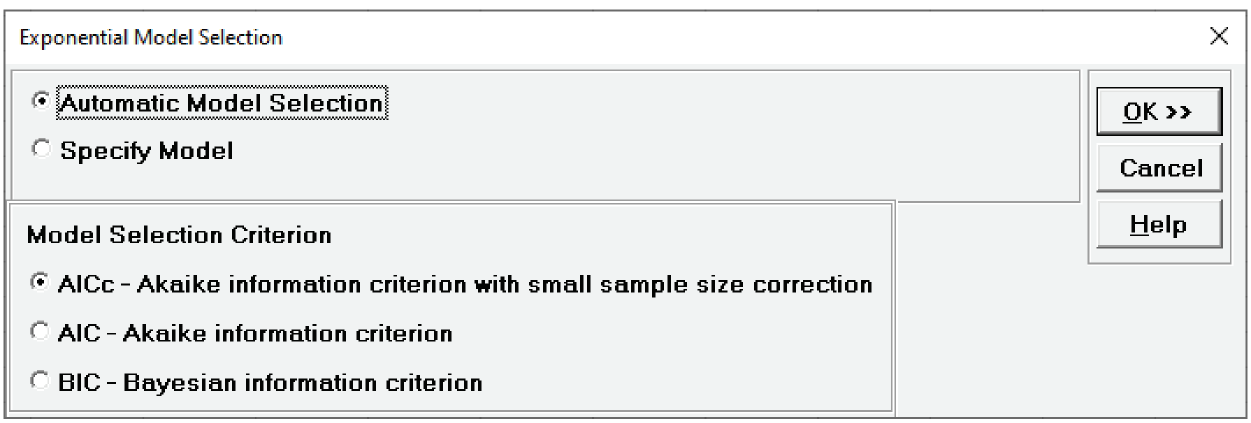


These residual plots are the same as used in SigmaXL's Multiple Regression. The histogram matches what we obtained in the Demo (the normal curve is not applied here). The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts. Later, we will apply a control chart to the residuals to formally test for significant outliers or assignable causes.

31. Now we will rerun Exponential Smoothing on the Concentration data but use Automatic Model Selection. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.



32. Click **Model Options**. Select **Automatic Model Selection**. We will use the default **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**.



33. Click **OK** to return to the Exponential Smoothing Forecast dialog. Click **OK**. The exponential smoothing forecast report is given.

34. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Forecast: Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

The model **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** was selected as the best fit for the Concentration data based on the AICc criterion. The point forecasts produced by the Multiplicative and Additive models are identical if they use the same smoothing parameter values. Multiplicative will, however, generate different prediction intervals to accommodate change in variance.

35. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	1
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of model information with Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “AICc” and Box-Cox Transformation = “N/A” because Box-Cox Transformation was unchecked.

36. The Parameter Estimates are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.295846234
l (initial level)	16.73860868

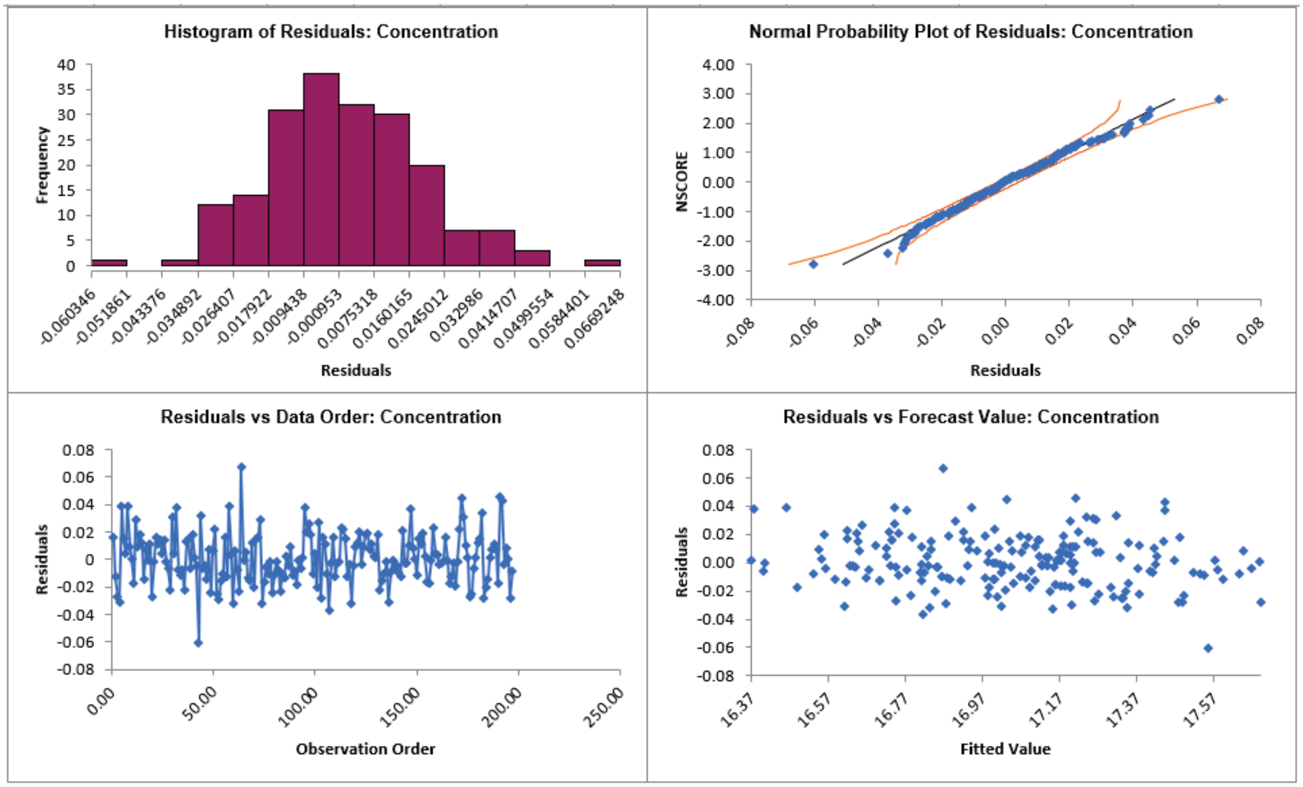
These are fairly close to the parameter estimates obtained above using the additive model.

37. The Exponential Smoothing Model Statistics are:

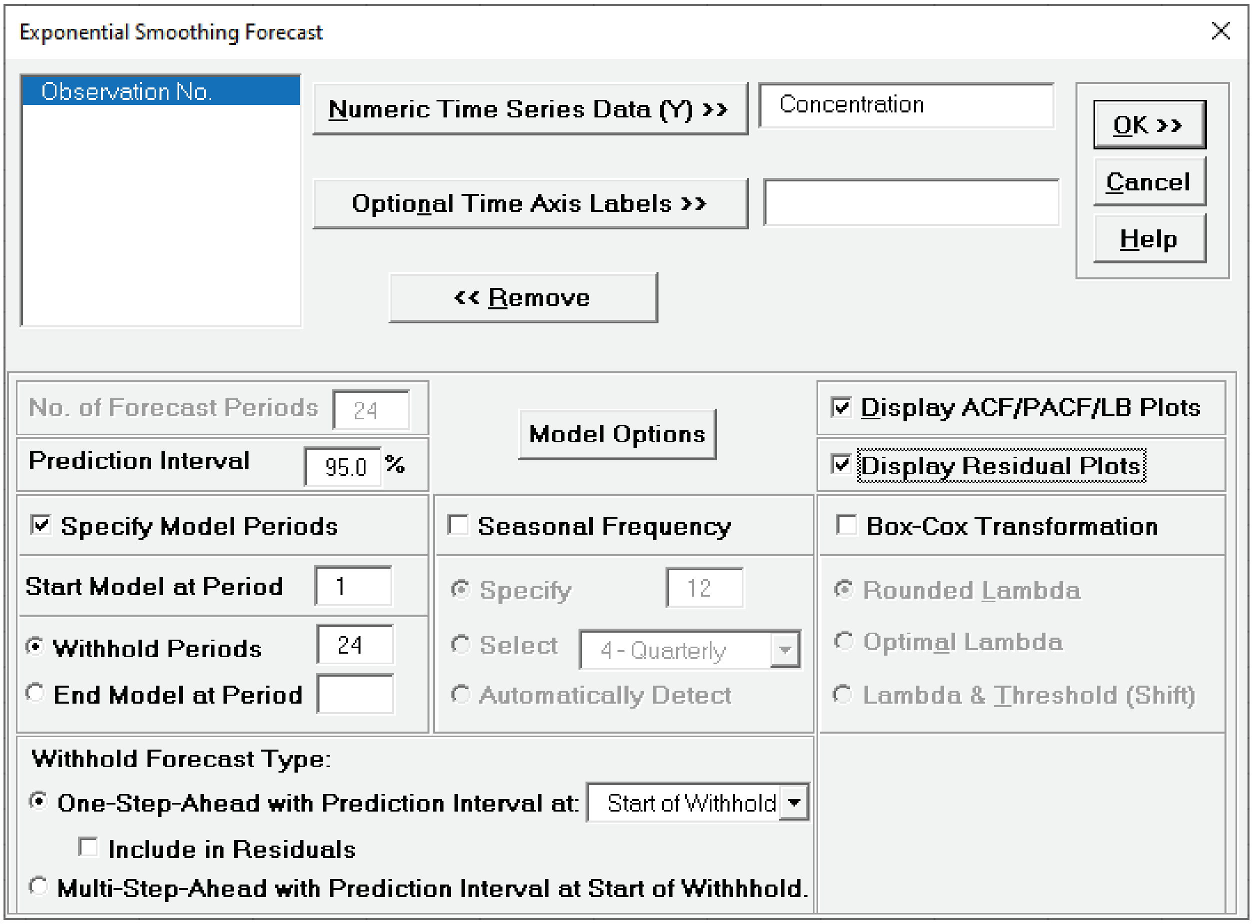
Exponential Smoothing Model Statistics	
No. of Observations	197
DF	195
StDev	0.01865
Variance	0.00035
Log-Likelihood	-293.601
AICc	593.326
AIC	593.201
BIC	603.051

The Log-Likelihood, AICc, AIC and BIC are close to the values obtained above using the additive model, but the Log-Likelihood is slightly higher giving a lower AICc, so multiplicative was selected as the best model. Note however that the StDev and Variance are very different. This difference is due to the multiplicative residuals being relative errors: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$.

38. The In-Sample Forecast Accuracy metrics, Forecast Table and ACF/PACF/LB Plots for multiplicative are very close to the additive model. The multiplicative residual plots look the same, but note the different scale due to the relative errors.



39. Now we will rerun Exponential Smoothing on the Concentration data with Automatic Model Selection, but will use Specify Withhold Periods. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Check **Specify Model Periods**. Set **Withhold Periods** = 24 (i.e., we will forecast 24 periods and compare against the withheld actual). We will use the default **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold*.



- **Specify Model Periods** option allows you to specify the start and end periods used in automatic model identification and parameter estimation. Typically, **Start Model at Period** is kept = 1 and **Withhold Periods** specifies the number of periods to be withheld for out-of-sample testing. **End Model at Period** specifies the end period, so the withhold sample size would be: total number of observations – end period.

- **Withhold Forecast Type: One-Step-Ahead** will exclude the withhold sample from automatic model identification and parameter estimation, but uses the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you only care about the short-term one-step ahead prediction.

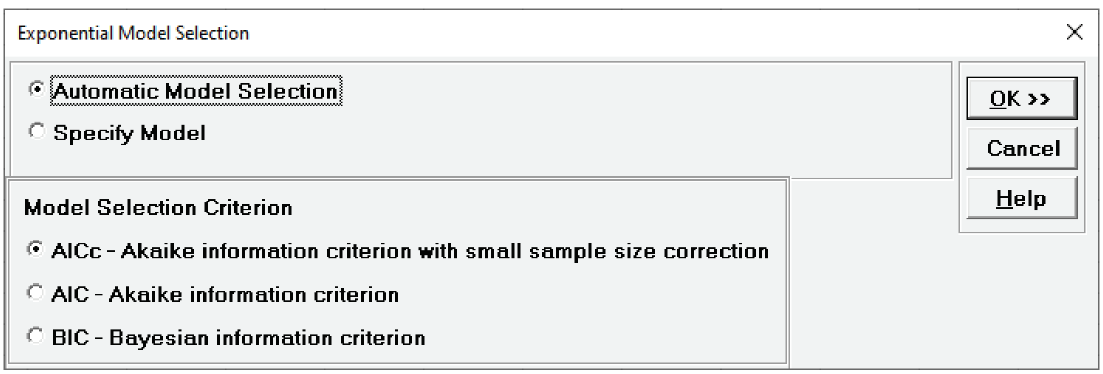
- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold* will display the prediction interval for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods.**

- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *End of Withhold* will display the prediction interval at the end of the withhold sample. The length of the prediction interval is determined by the specified **No. of Forecast Periods.**

- **Include in Residuals** will treat the one-step-ahead forecast errors as residuals (even though they were not part of the model estimation process) and will be included in the ACF/PACF/LB Residual Plots along with the Residuals report and graphs. Typically, this is kept unchecked.

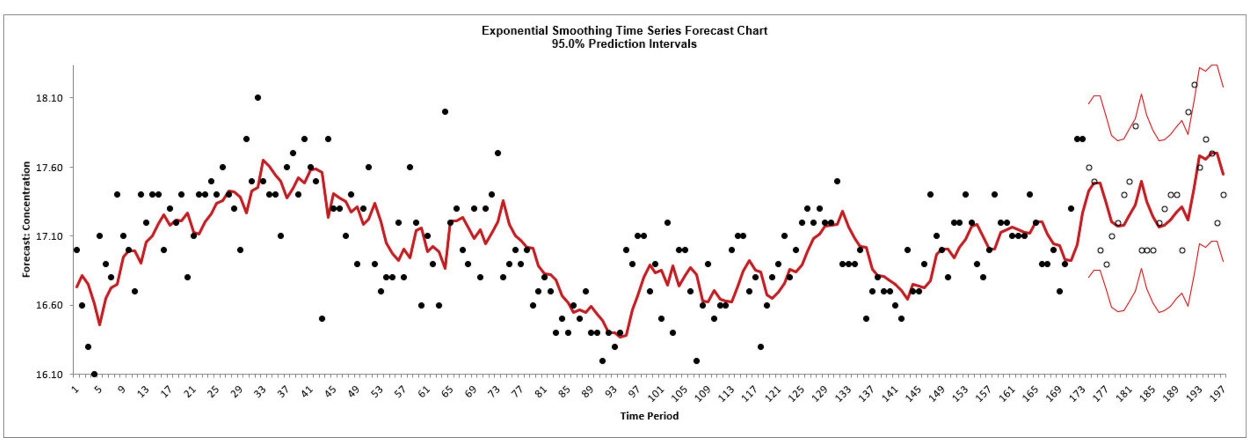
- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** will exclude the withhold sample from automatic model identification and parameter estimation and does not use the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you are interested in a long-term forecast window (horizon). The prediction interval will be displayed for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods**. These forecast errors are not included in ACF/PACF/LB Residual Plots or the Residuals report and graphs.

40. Click **Model Options**. Select **Automatic Model Selection**. We will use the default **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**.



Tip: When using **Recall SigmaXL Dialog**, and if there are no changes to the **Model Option** settings, the previous settings will be used. It is not necessary to repeat this step.

41. Click **OK** to return to the Exponential Smoothing Forecast dialog. Click **OK**. The exponential smoothing forecast report is given:



The blank dots are the data values in the withhold sample with a one-step-ahead forecast and prediction intervals displayed at the start of the withhold sample.

42. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Forecast: Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

As with the complete data, the model **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** was selected as the best fit for the Concentration data based on the AICc criterion. The header also includes the number of specified withhold periods.

43. The Parameter Estimates are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.303967286
l (initial level)	16.73554259

These are close to the parameter estimates obtained above (which used all of the data with the multiplicative model).

44. The Exponential Smoothing Model Statistics are:

Exponential Smoothing Model Statistics	
No. of Observations	173
DF	171
StDev	0.01841
Variance	0.00034
Log-Likelihood	-243.805
AICc	493.752
AIC	493.61
BIC	503.07

These are fairly close to the model statistics obtained above (which used all of the data with the multiplicative model). Here we are using only 173 of the 197 observations.

45. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	173	24
RMSE	0.31133921	0.35209899
MAE	0.243032918	0.273389284
MAPE	1.425751706	1.567334268
MASE	0.878186174	0.987877246

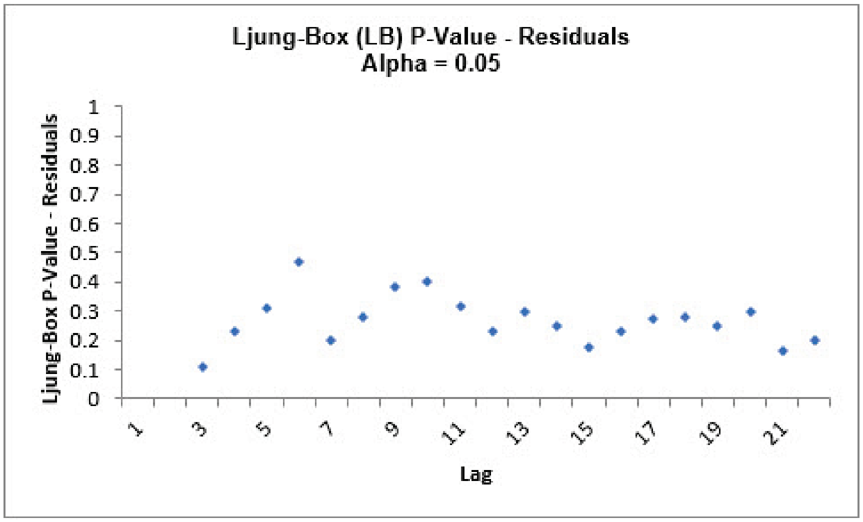
As expected, the **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors.

46. The Forecast Table is given as:

Forecast Table				
Period	Withhold Data	One-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
174	17.6	17.43034504	16.80141557	18.05927452
175	17.5	17.4819146	16.85112437	18.11270483
176	17	17.48741197	16.85642338	18.11840056
177	16.9	17.33925468	16.71361197	17.96489738
178	17.1	17.20573562	16.58491061	17.82656064
179	17.2	17.17359545	16.55393014	17.79326077
180	17.4	17.18162157	16.56166665	17.80157649
181	17.5	17.24800147	16.6256514	17.87035154
182	17.9	17.32460078	16.69948682	17.94971474
183	17	17.49950332	16.86807845	18.13092819
184	17	17.34767065	16.72172427	17.97361703
185	17	17.24199015	16.61985698	17.86412331
186	17.2	17.16843306	16.54895401	17.78791211
187	17.3	17.17802838	16.55820311	17.79785365
188	17.4	17.21510376	16.59394072	17.8362668
189	17.4	17.27130617	16.64811521	17.89449713
190	17	17.31042488	16.68582242	17.93502734
191	18	17.21606587	16.59486812	17.83726363
192	18.2	17.4543562	16.82456035	18.08415206
193	17.6	17.68100752	17.04303354	18.31898151
194	17.8	17.65638389	17.01929838	18.29346939
195	17.7	17.70003849	17.06137782	18.33869916
196	17.2	17.70002679	17.06136654	18.33868704
197	17.4	17.548035	16.91485899	18.18121102

These are the same forecast and prediction interval values displayed in the Forecast Chart, but provided for further analysis or charting. The **Withhold Data** is also displayed.

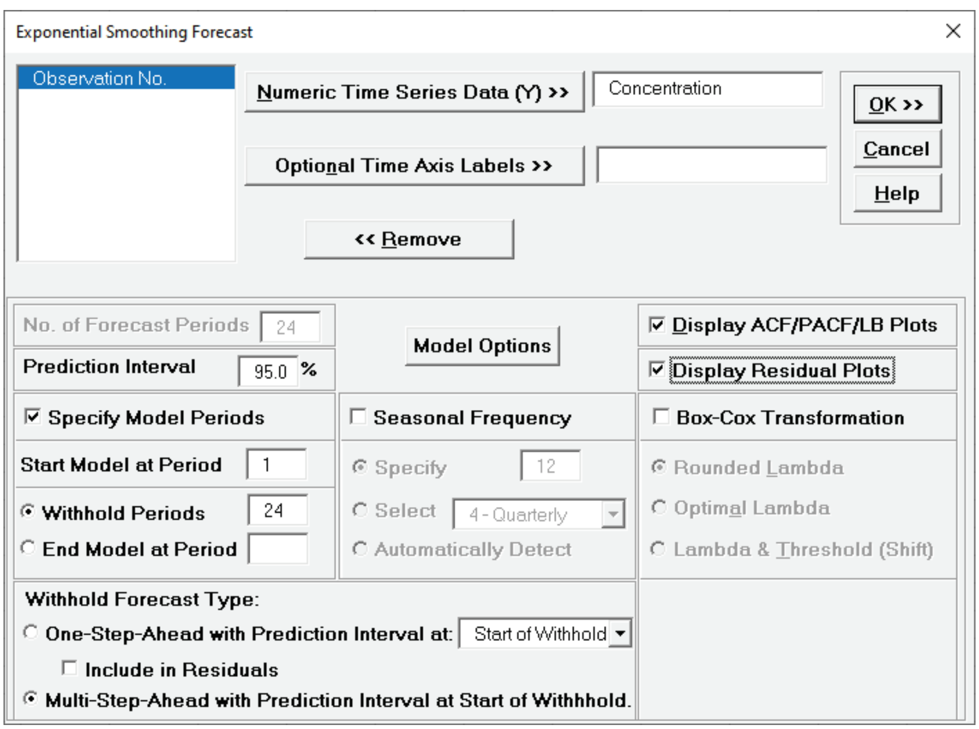
47. The ACF/PACF/LB Residual Plots and Residual Plots are based on the in-sample data. The plots look similar to the complete data above, except for the Ljung-Box P-Values:



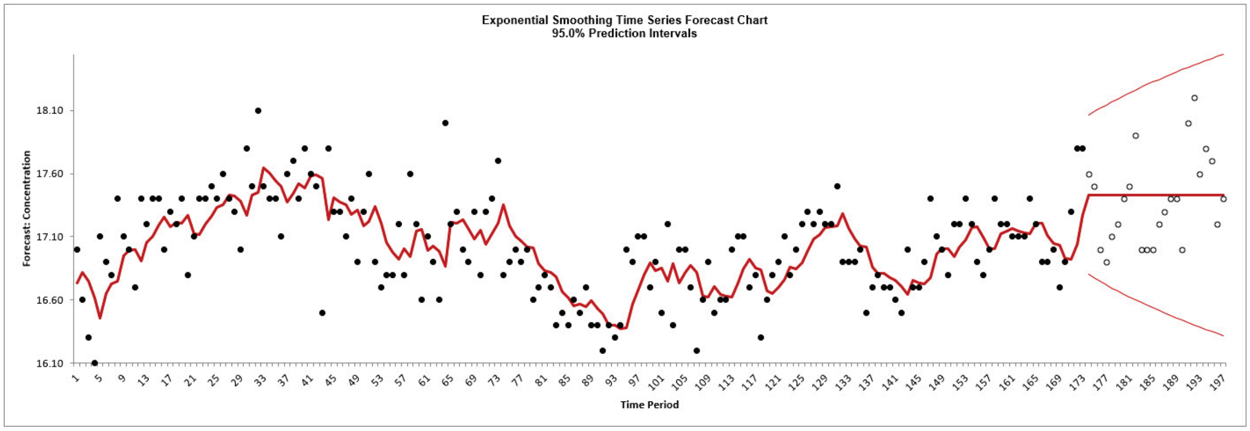
The **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** model is a better fit to the subset than the complete data, with all P-Values being blue (> .05).

48. If **Include in Residuals** was checked then the residuals would also include the **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors.

49. Now we will rerun Exponential Smoothing on the Concentration data, but use Multi-Step-Ahead for the Withhold Forecast. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**.



50. Click **OK**. The exponential smoothing forecast report is given:



The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample.

51. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	173	24
RMSE	0.31133921	0.348228059
MAE	0.243032918	0.286752928
MAPE	1.425751706	1.647391419
MASE	0.878186174	1.036166041

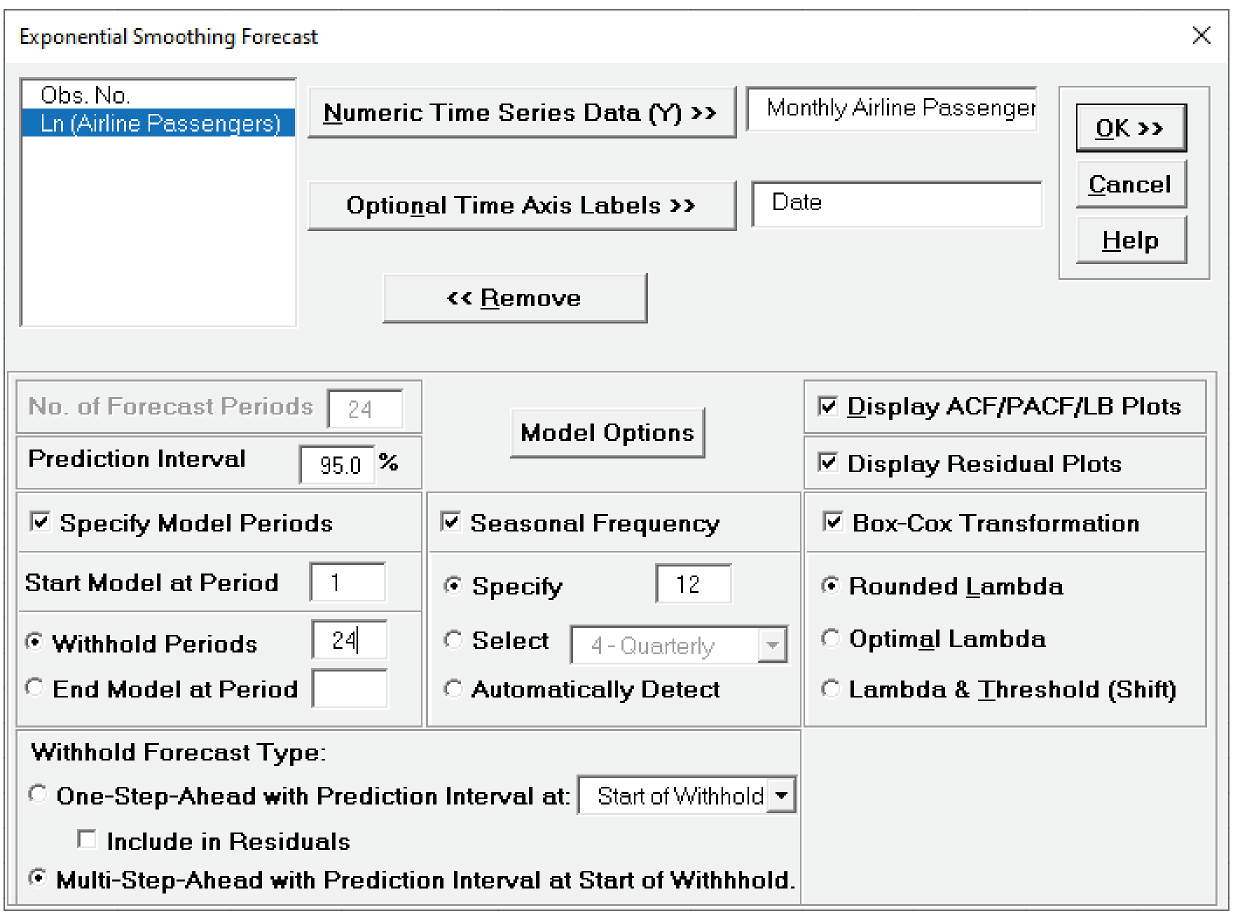
As expected, the **Out-of-Sample (Withhold) Multi-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors and the **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors above.

Monthly Airline Passengers - Series G

52. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

53. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

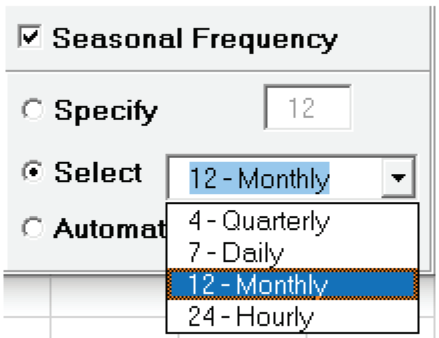
54. Select *Monthly Airline Passengers*, click **Numeric Time Series Data (Y) >>**. Select *Date*, click **Optional Time Axis Labels >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 24 (i.e., we will forecast 24 months and compare to withheld actual). Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Check **Seasonal Frequency** with **Specify** = 12. Check **Box-Cox Transformation** and select **Rounded Lambda** (selected because the Run Chart showed an increase in the seasonal variance over time). We will use the default **Prediction Interval** = 95.0 %.



- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** will exclude the withhold sample from automatic model identification and parameter estimation and does not use the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you are interested in a long-term forecast window (horizon). The prediction interval will be displayed for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods.**

- **Seasonal Frequency Specify** is used to specify the seasonal frequency. Note that Exponential Smoothing is limited to a maximum seasonal frequency of 24. For higher frequencies use Exponential Smoothing – Multiple Seasonal Decomposition (MSD)

- **Seasonal Frequency Select** gives a drop-down list of commonly used seasonal frequencies:



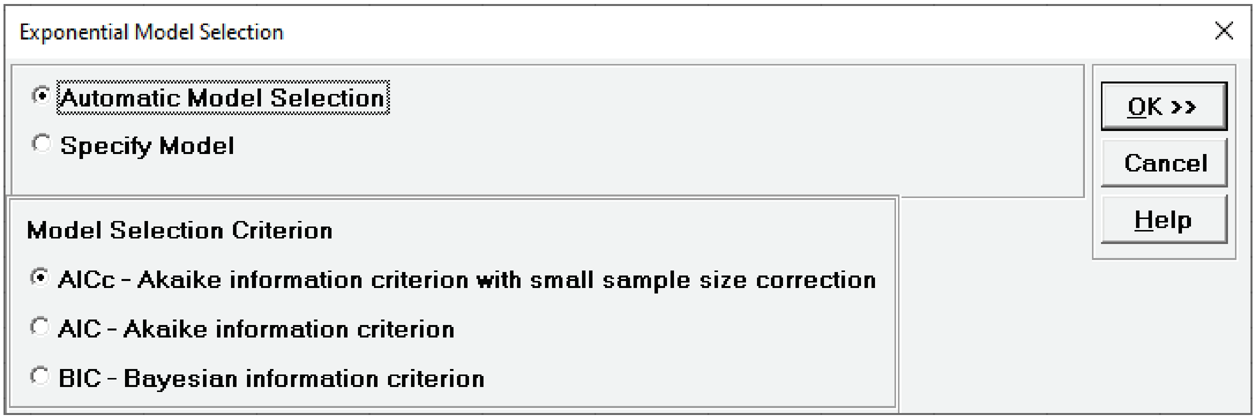
- **Seasonal Frequency Automatically Detect** should be used if uncertain what the seasonal frequency value is (or do a Spectral Density Plot prior to the Seasonal Trend Decomposition Plots). If detected frequency is > 24, 1 is returned. If that occurs, please use Exponential Smoothing – Multiple Seasonal Decomposition (MSD).

- **Box-Cox Transformation** with **Rounded Lambda** will select Lambda = 0 (Ln), 0.5 (SQRT) or 1 (Untransformed). Threshold (Shift) is computed automatically if the time series data includes 0 or negative values, otherwise it is 0.

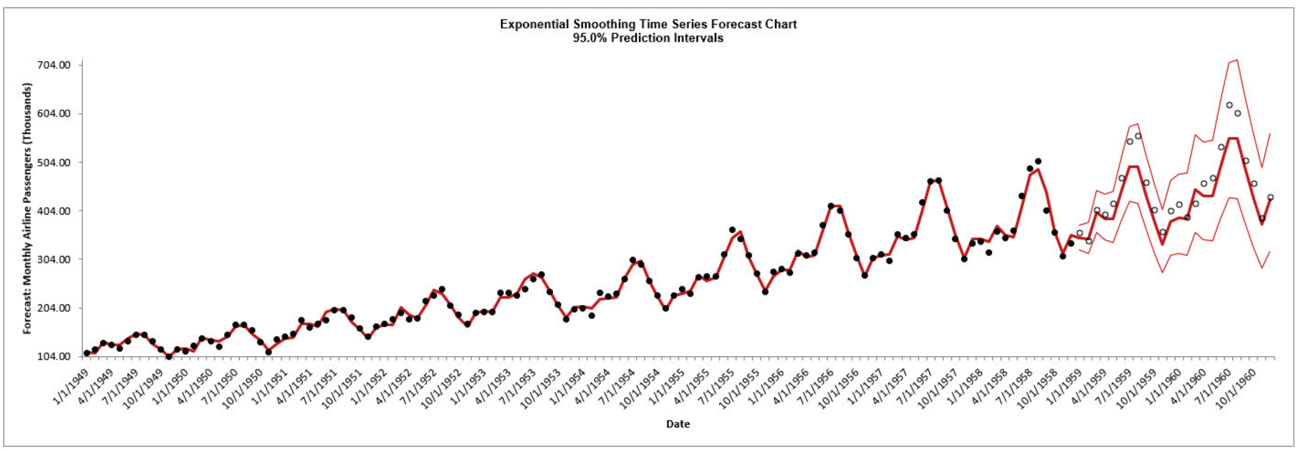
- **Box-Cox Transformation** with **Optimal Lambda** uses the range of 0 to 1 for Lambda. Threshold is computed automatically if the time series data includes 0 or negative values.

- **Box-Cox Transformation** with **Lambda & Threshold (Shift)** if left blank, will compute optimal lambda and threshold. The user may also specify Lambda and Threshold. Lambda may be specified outside of the 0 to 1 range, but practically for time series analysis, should be limited to -1 to 2. Threshold is typically 0, but if the time series data includes 0 or negative values, a negative threshold value should be entered that is smaller than the minimum data value. This value will be subtracted from the data resulting in positive time series values.

55. Click **Model Options**.



56. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing Forecast dialog. Click **OK**. The exponential smoothing forecast report is given:



The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample. Note that the data has been transformed using Box-Cox, but the inverse transformation is applied to produce this chart with the original units.

57. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

The model **Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A)** was automatically selected as the best fit for the Box-Cox transformed Airline Passenger data based on the AICc criterion. Note that multiplicative models are not considered when **Box-Cox Transformation** is checked. If a multiplicative model was desired, then one could manually transform the data and model Ln(Airline Passengers) with **Box-Cox Transformation** unchecked.

The header also includes the number of specified withhold periods.

58. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	12
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

This is a summary of model information with Seasonal Frequency = 12; Model Selection Criterion = "AICc" and Box-Cox Transformation = "Rounded Lambda" with Lambda = 0 (i.e., a Ln transformation).

59. The Parameter Estimates are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.770415301
beta (trend smoothing)	0.0001
gamma (seasonal smoothing)	0.0001
l (initial level)	4.80925762
b (initial trend)	0.00935261
s1 (initial seasonal)	-0.099989196
s2 (initial seasonal)	-0.217929483
s3 (initial seasonal)	-0.076652841
s4 (initial seasonal)	0.063787774
s5 (initial seasonal)	0.198157807
s6 (initial seasonal)	0.205968013
s7 (initial seasonal)	0.108472035
s8 (initial seasonal)	-0.016634733
s9 (initial seasonal)	-0.008278553
s10 (initial seasonal)	0.033611999
s11 (initial seasonal)	-0.101347496
s12 (initial seasonal)	-0.089165324

- Error includes the smoothing parameter *alpha* and initial level value (*l*). The error is additive, but on the Ln transformed data.
- Trend adds a smoothing parameter (*beta*) and initial trend value (*b*).
- Seasonal adds a smoothing parameter (*gamma*) and initial seasonal values (*s1* to *s12*), constrained to sum to zero for additive.

60. The Exponential Smoothing Model Statistics are:

Exponential Smoothing Model Statistics	
No. of Observations	120
DF	104
StDev	0.037144
Variance	0.00138
Log-Likelihood	116.4916
AICc	-192.983
AIC	-198.983
BIC	-151.596

- The number of observations, *n* = 144 – 24 (*withhold*) = 120
- Degrees of freedom (DF) = 120 (*n*) – 16 (terms in the model, excluding *s12*) = 104
- Note that the model statistics are based on the Ln transformed data, not the original data.

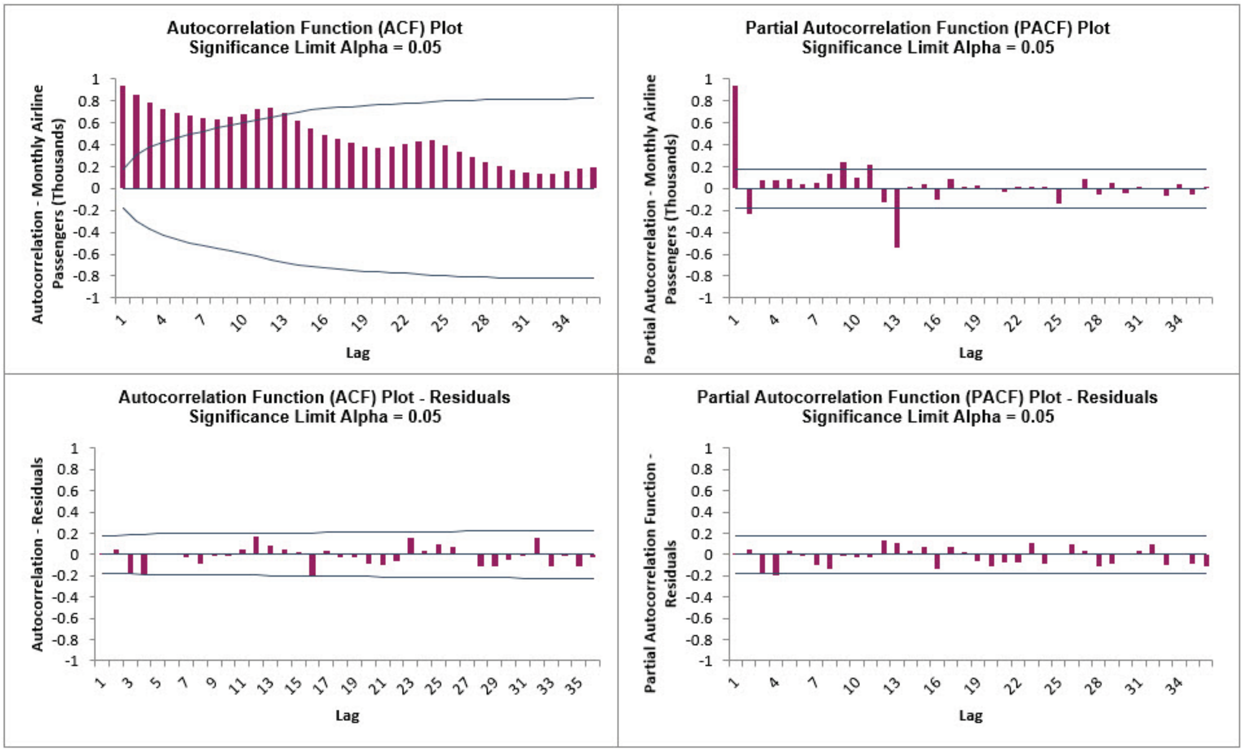
61. The Forecast Accuracy metrics are:

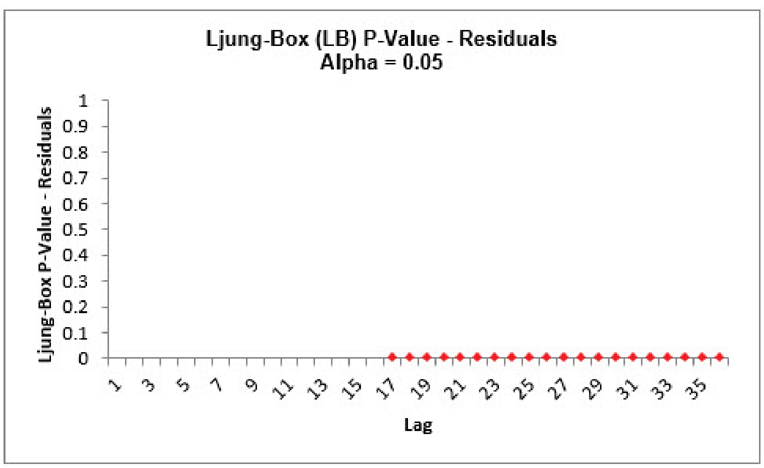
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	120	24
RMSE	8.67099231	33.06342473
MAE	6.481327147	27.83270693
MAPE	2.730421504	5.80549634
MASE	0.226825448	0.974054552

As expected, the **Out-of-Sample (Withhold) Multi-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors. Note, if we were primarily interested in a short term one-step ahead forecast, then we would have selected **Withhold Forecast Type: One-Step-Ahead** and this table would show **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors.

Forecast Accuracy metrics are calculated using the actual raw data versus inverse transformed forecast as displayed in the Forecast Chart and Table, so allows comparison across all model types and transformations.

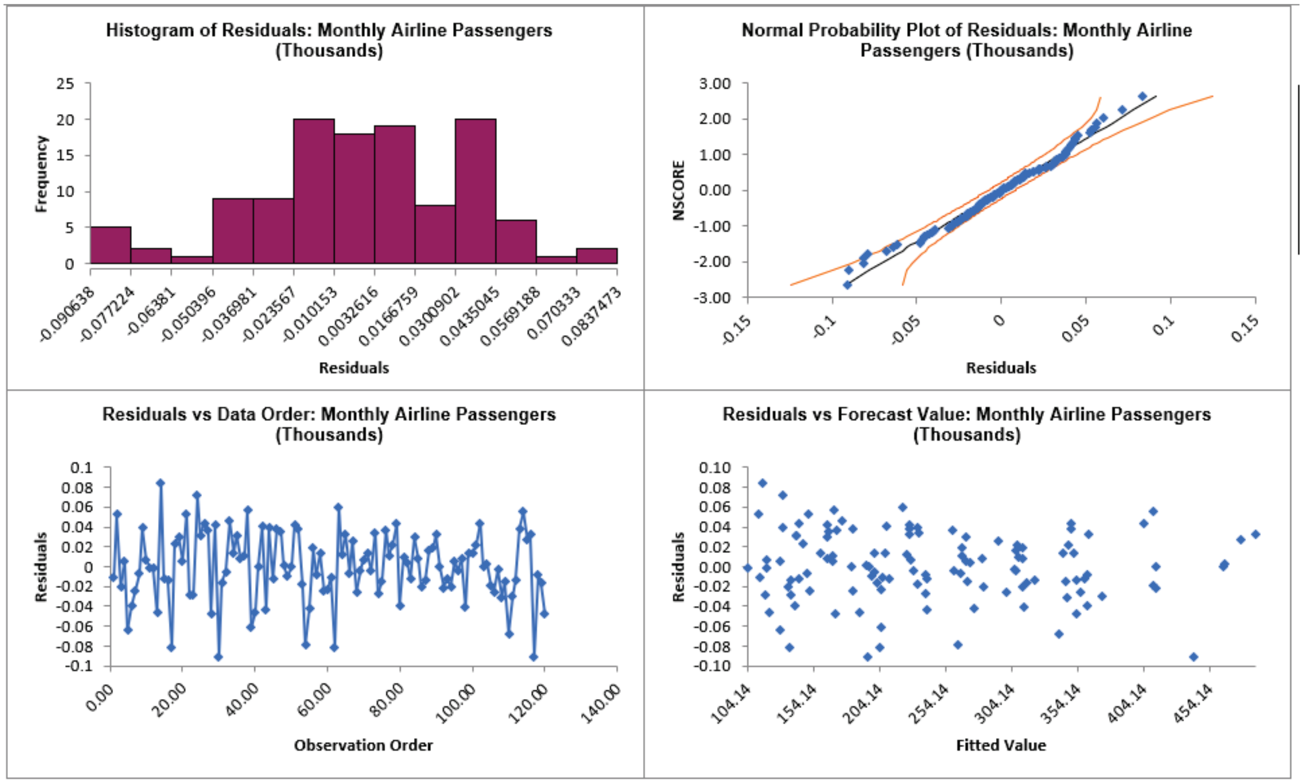
62. Click on the **Exp. Smooth. ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ACF/PACF Residuals Plots indicate that almost all of the autocorrelation has been accounted for in the model, but the Ljung-Box plot shows that significant autocorrelation still remains (the red P-Values are significant at alpha=.05) - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

63. Click on the **Exp. Smoothing Residuals** sheet to view the Residual Plots:



The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts.

Note that the residuals are based on the Ln transformed data, not the original data. The **Exponential Smoothing Model Information** to the right of the plots shows the Box-Cox Transformation information.

Exponential Smoothing Model Information	
Seasonal Frequency	12
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Exponential Smoothing – Multiple Seasonal Decomposition (MSD) Forecast

Exponential Smoothing is limited to a maximum seasonal frequency of 24. For higher frequencies use Exponential Smoothing – Multiple Seasonal Decomposition (MSD). The seasonal component is first removed through decomposition, a nonseasonal exponential smooth model fitted to the remainder (+trend), and then the seasonal component is added back in. For forecasting, a naïve seasonal forecast is used on the seasonal component. Note that the prediction intervals are derived from the exponential smoothing model and do not include uncertainty in the seasonal component.

As the name implies, Multiple Seasonal Decomposition (MSD) also accommodates multiple seasonality, for example the half-hourly data with a seasonal frequency of 48 observations per day and 336 observations per week. When using MSD, it is recommended to limit the forecast period to 2*dominant seasonal frequency.

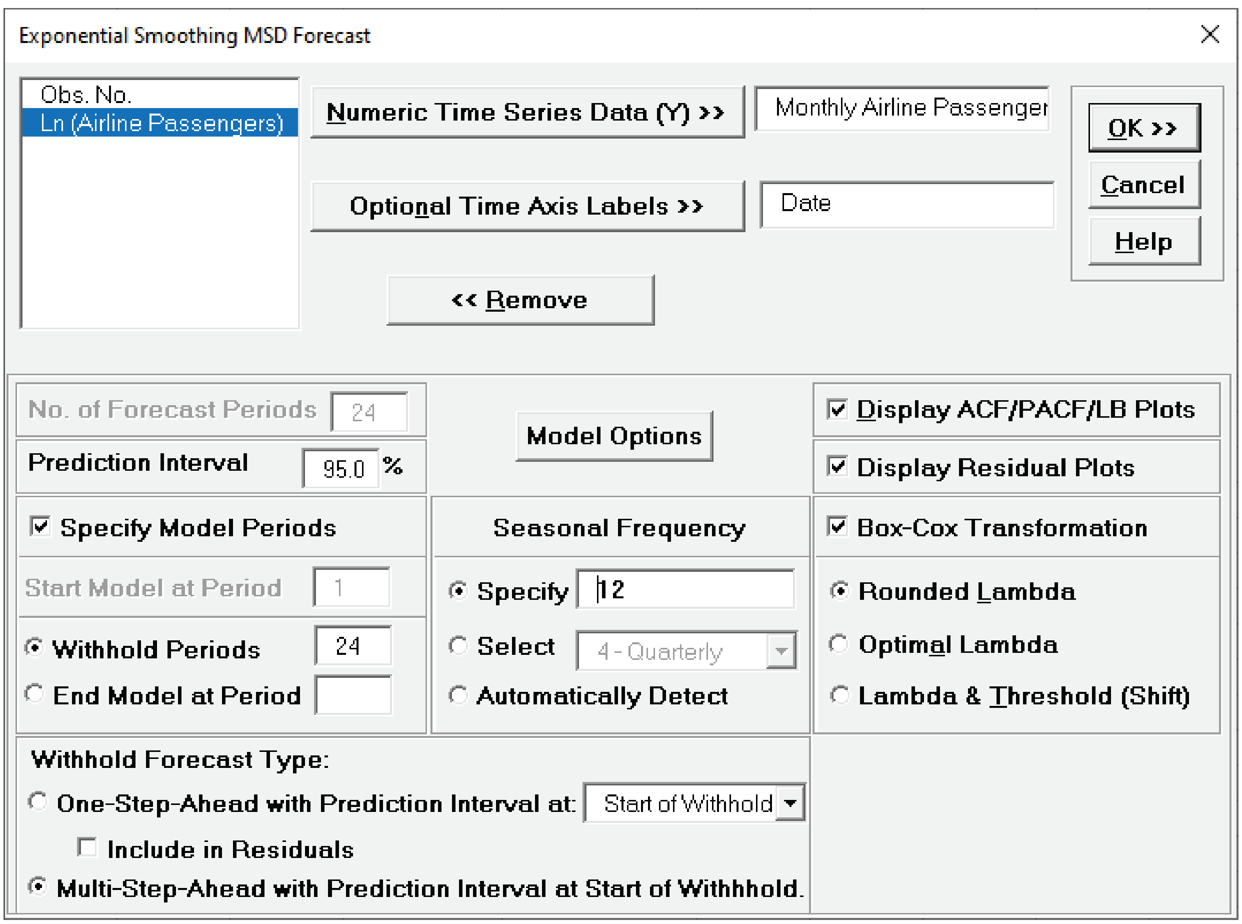
For further details and formulas, see Appendix: **Seasonal Trend Decomposition.**

Monthly Airline Passengers - Series G

1. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data. The Multiple Seasonal Decomposition (MSD) option is not necessary for this data, but by way of introduction, we will use this to compare to the previous analysis.

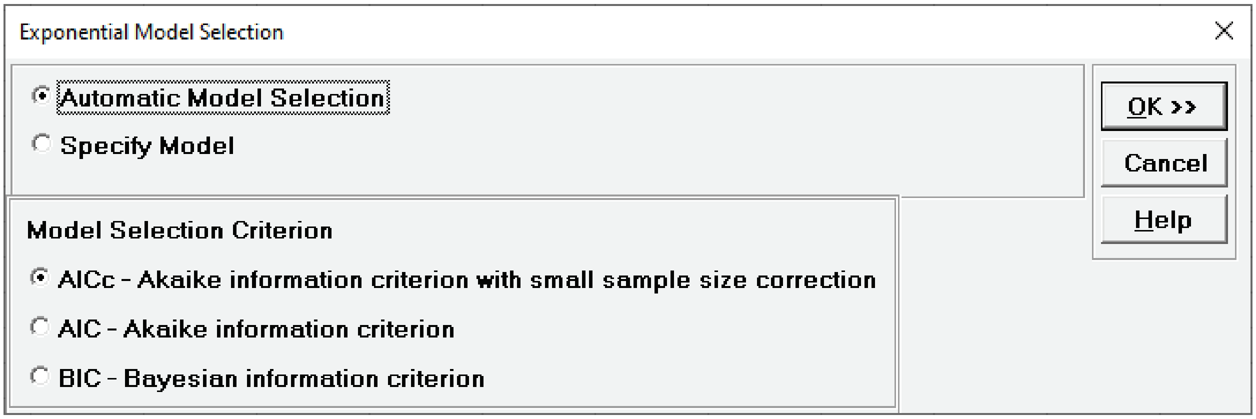
2. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Multiple Seasonal Decomposition Forecast**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Monthly Airline Passengers*, click **Numeric Time Series Data (Y) >>**. Select *Date*, click **Optional Time Axis Labels >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 24. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Select **Seasonal Frequency Specify** and enter 12**.** Check **Box-Cox Transformation** and select **Rounded Lambda**. We will use the default **Prediction Interval** = 95.0 %.

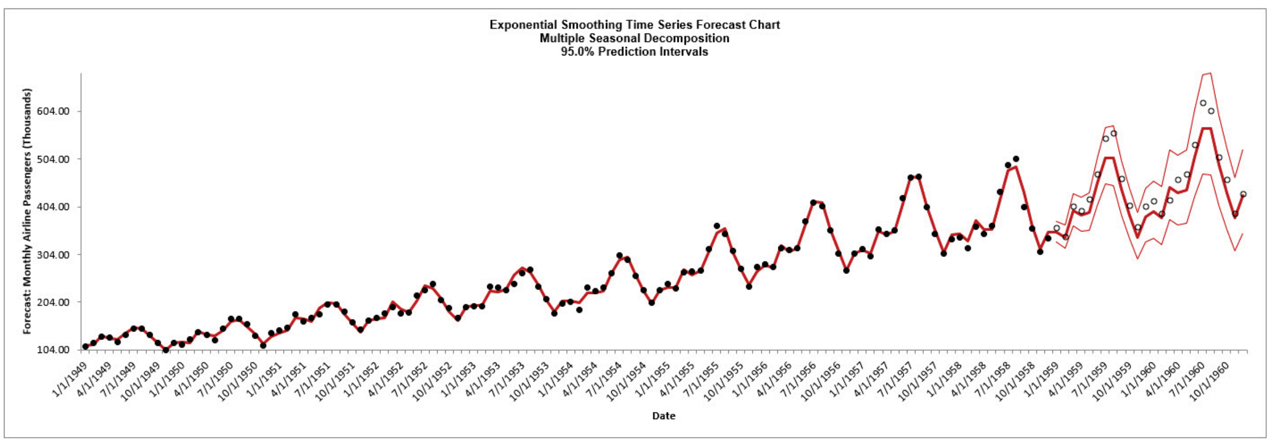


- **Seasonal Frequency** can have multiple entries however, we recommend no more than 3 values.

4. Click **Model Options**.



5. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing MSD Forecast dialog. Click **OK**. The exponential smoothing forecast report is given:



6. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model (Multiple Seasonal Decomposition): Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

After **Multiple Seasonal Decomposition**, the model is **Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N)**.

(A, A, N) was automatically selected as the best fit for the deseasonalized Airline Passenger data based on the AICc criterion.

The header also includes the number of specified withhold periods.

7. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	12 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

   This is a summary of model information with Seasonal Frequency = 12 using Decomposition and Model Selection Criterion = "AICc". The Box-Cox Transformation is "Rounded Lambda" with Lambda = 0 (Ln transformation).

8. The Parameter Estimates for the deseasonalized Airline Passenger data are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.674923407
beta (trend smoothing)	0.0001
l (initial level)	4.806932238
b (initial trend)	0.009425139

   - Error includes the smoothing parameter *alpha* and initial level value (*l*). The error is additive, but on the Ln transformed data.
   - Trend adds a smoothing parameter (*beta*) and initial trend value (*b*).
   - Seasonal smoothing parameter (*gamma*) and initial seasonal values are not computed because the data has been deseasonalized.

9. The Exponential Smoothing Model Statistics are:

Exponential Smoothing Model Statistics	
No. of Observations	120
DF	116
StDev	0.030549511
Variance	0.000933273
Log-Likelihood	133.3933791
AICc	-256.2604423
AIC	-256.7867581
BIC	-242.8492994

   - The number of observations, *n* = 144 – 24 (*withhold*) = 120
   - Degrees of freedom (DF) = 120 (*n*) – 4 (terms in the model) = 116
   - Note that the model statistics are based on the Ln transformed data, not the original data.

- The model statistics are better than the previous analysis: lower StDev and Variance, higher Log-Likelihood and lower AICc, AIC and BIC. This is to be expected because the data has been deseasonalized so the seasonal error component is not included.

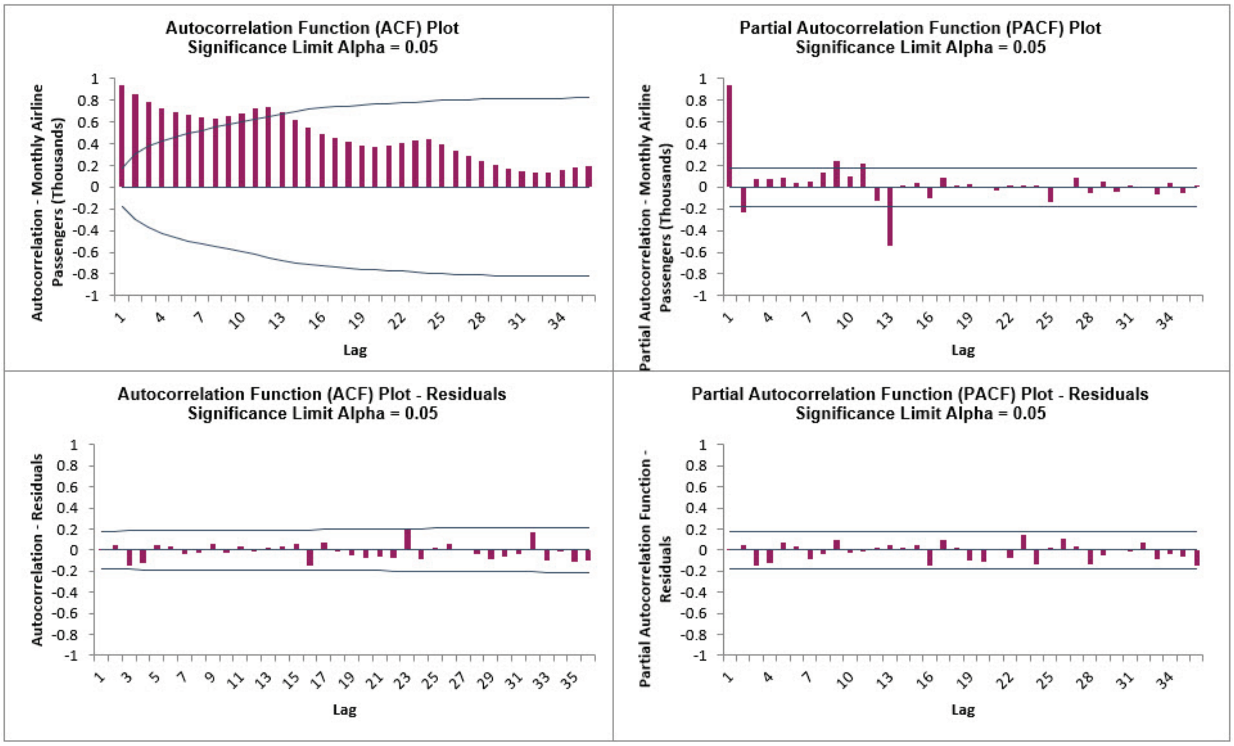
10. The Forecast Accuracy metrics are:

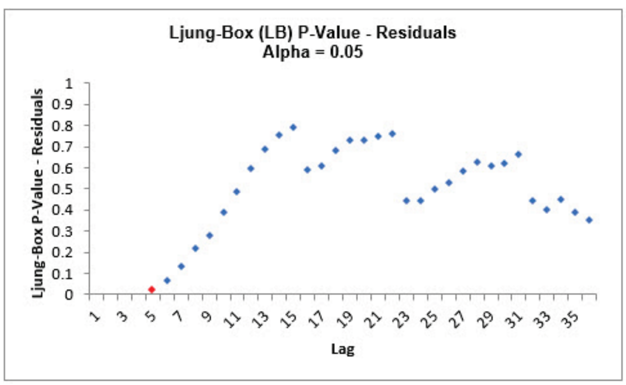
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	120	24
RMSE	7.115702971	26.70412899
MAE	5.455681009	22.99181598
MAPE	2.349324345	4.842245097
MASE	0.190931156	0.804639056

Comparing to our earlier analysis, both the **In-Sample (Estimation) One-Step-Ahead Forecast** errors and **Out-of Sample (Withhold) Multi-Step-Ahead Forecast** errors are slightly smaller. (This was not expected, typically a Seasonal Exponential Smoothing or ARIMA model would give a more accurate forecast, based on comparison of methods using forecast competition data). However, given that we are forecasting out for two years, both models look very good.

Forecast Accuracy metrics are calculated using the actual raw data versus inverse transformed forecast as displayed in the Forecast Chart and Table, so allow comparison across all model types and transformations.

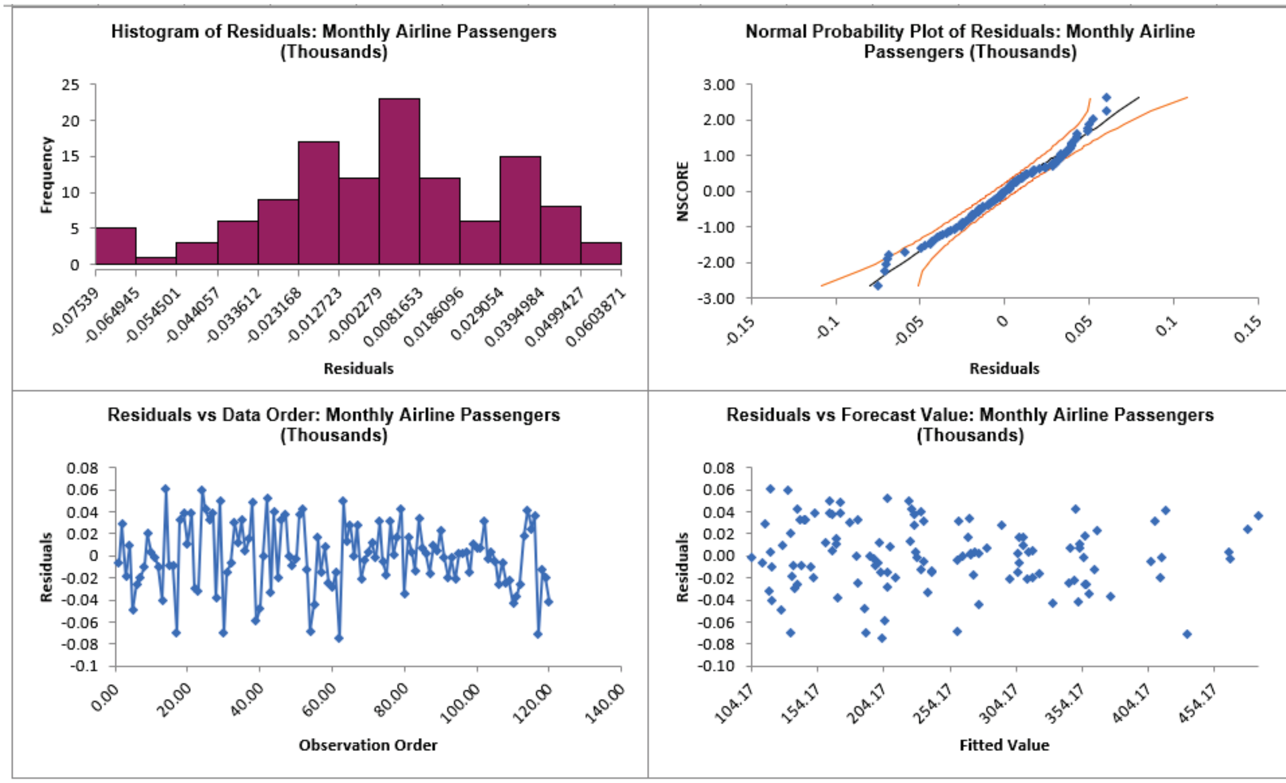
11. Click on the **Exp Smooth MSD ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ACF/PACF Residuals Plots are similar to the previous analysis and indicate that almost all of the autocorrelation has been accounted for in the model, however the Ljung-Box plot confirms that this is a better fit, with most P-Values being blue (> .05).

12. Click on the **Exp Smooth MSD Residuals** sheet to view the Residual Plots:



The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts.

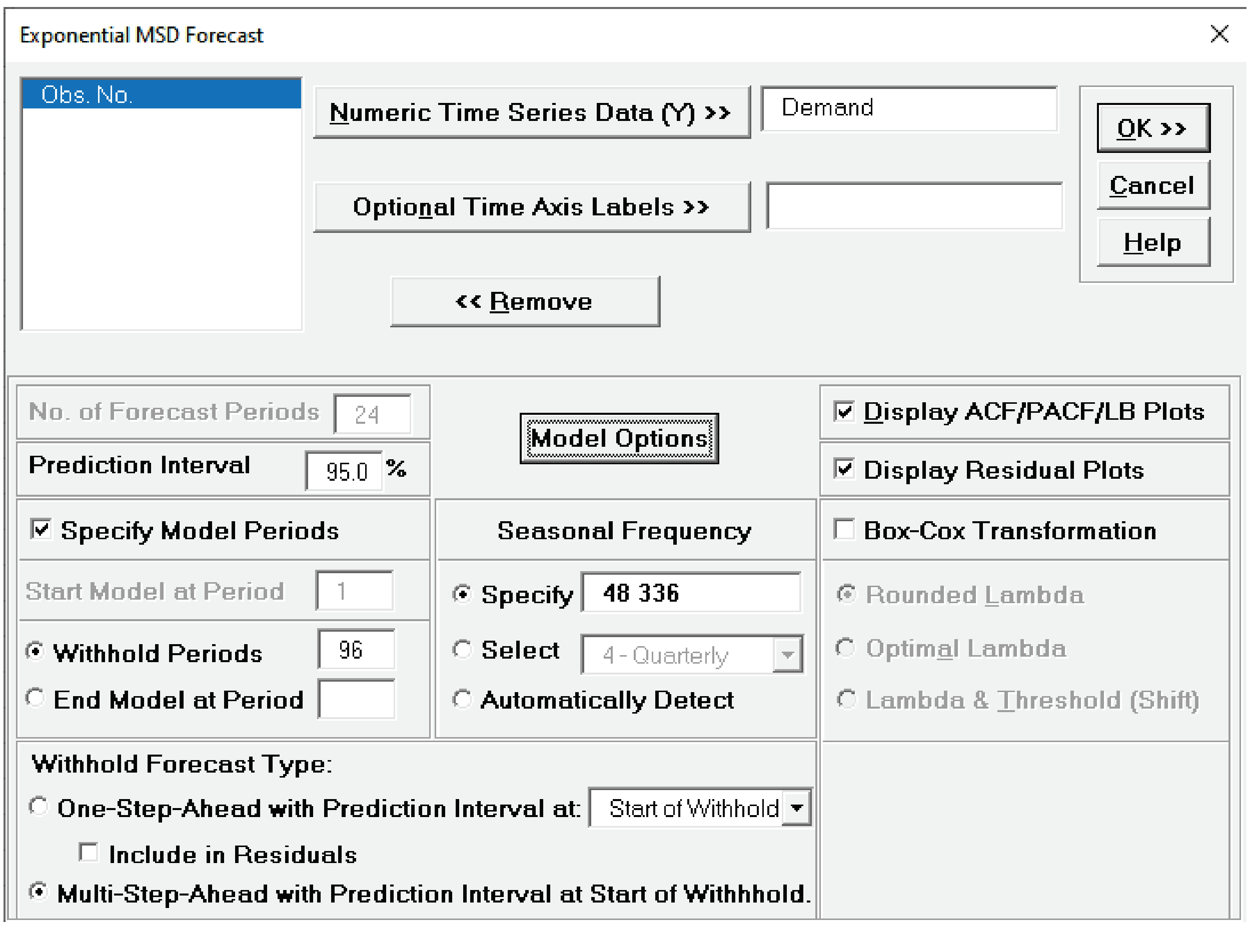
Note that Residuals for MSD are the final observed -predicted values, so there are no scaling differences if the model uses additive or multiplicative error. Since a Box-Cox transformation was used, the residuals are in Ln transformed units.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

13. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

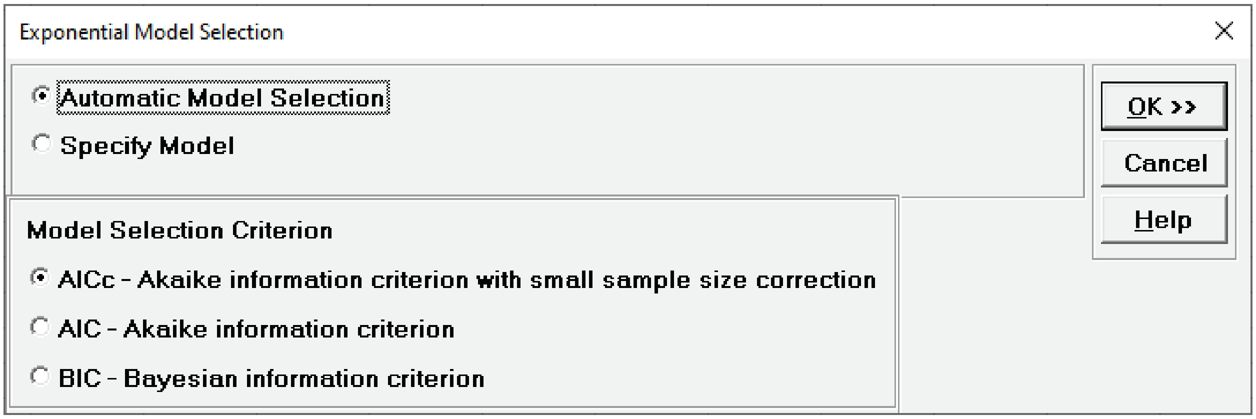
14. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Multiple Seasonal Decomposition Forecast**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

15. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 96. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Check **Seasonal Frequency** with **Specify** = 48 336**.** Leave **Box-Cox Transformation** unchecked. We will use the default **Prediction Interval** = 95.0 %.

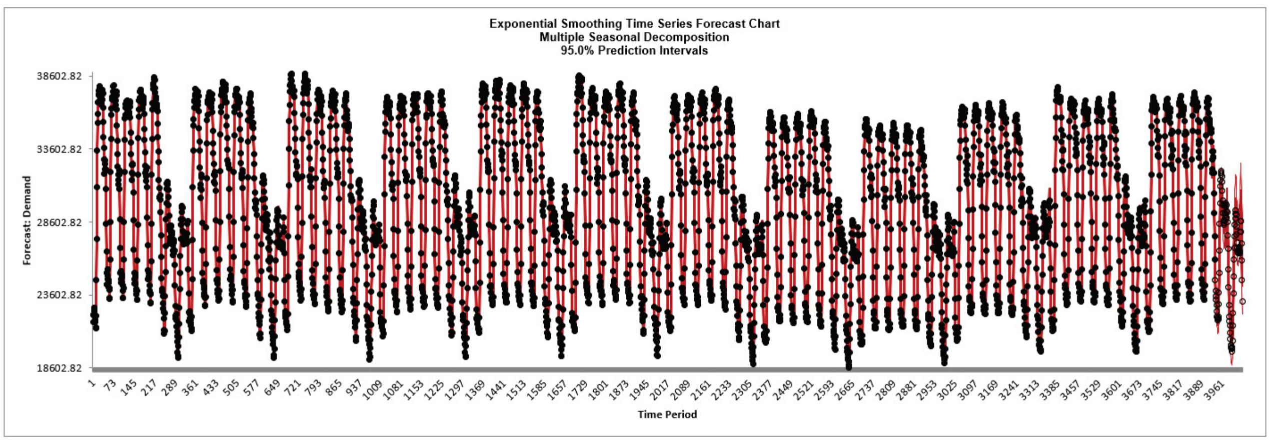


Withhold Periods is 2*dominant seasonal frequency (48). Dominant frequency is obtained from the Spectral Density Plot. Start Model at Period = 1 is always greyed out for MSD.

16. Click **Model Options**.



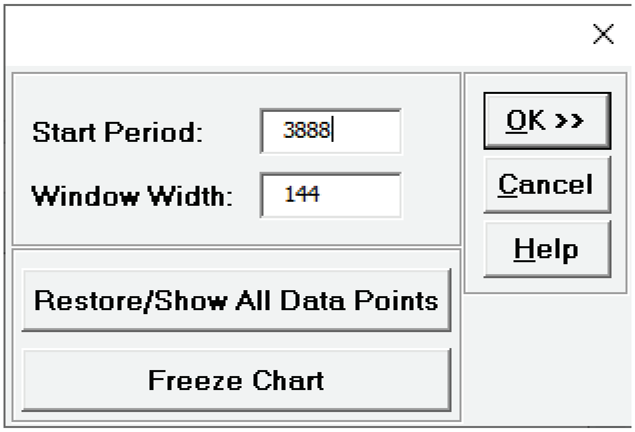
17. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing MSD Forecast dialog. Click **OK**. The exponential smoothing forecast report is given:



18. We will want to zoom in on the last 3 days, i.e., 144 half-hourly time periods, using chart scrolling. Click **SigmaXL Chart Tools > Enable Scrolling**

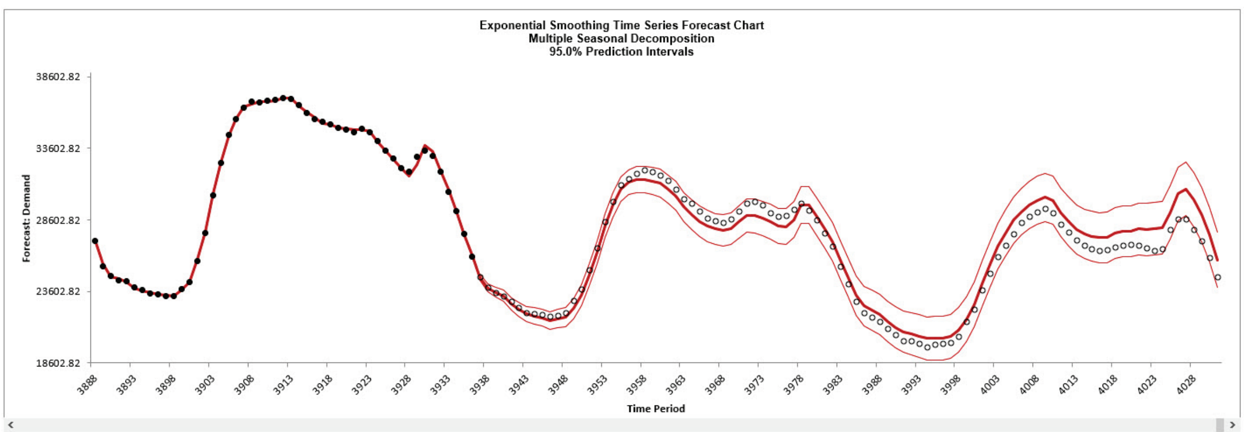
You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

19. Click **OK.** The scroll dialog appears allowing you to specify the **Start Period** and **Window Width**. Enter **Start Period** = 3888 and **Window Width** = 144:



At any point, you can click **Restore/Show All Data Points** or **Freeze Chart**. Freezing the chart will remove the scroll and unload the dialog. The scroll dialog will also unload if you change worksheets. To restore the dialog, click **SigmaXL Chart Tools > Enable Scrolling**.

20. Click **OK.** A scroll bar appears beneath the forecast chart. You can also change the **Start Subgroup** and **Window Width** and **Update**.



You can scroll through by clicking to the right or left, with the specified window width of 144.

The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample. The model does quite well at predicting the withhold 96 half-hour demand values. Note that the prediction error increases the further out we predict.

Click **Cancel** to exit the scroll dialog

21. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model (Multiple Seasonal Decomposition): Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 96 withhold periods.

After **Multiple Seasonal Decomposition**, the model is **Simple Exponential Smoothing with Multiplicative Errors (M, N, N).**

(M, N, N) was automatically selected as the best fit for the deseasonalized Demand data based on the AICc criterion.

The header also includes the number of specified withhold periods.

22. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	48, 336 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of model information with Seasonal Frequency = 48, 336 using decomposition and Model Selection Criterion = “AICc”.

23. The Parameter Estimates for the deseasonalized Demand data are:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.9999
l (initial level)	29798.08498

- Error includes the smoothing parameter *alpha* and initial level value (*l*). With *alpha* = 0.9999, this is effectively a naïve forecast, with multiplicative errors.
- Seasonal smoothing parameter (*gamma*) is not computed because the data has been deseasonalized.

24. The Exponential Smoothing Model Statistics are:

Exponential Smoothing Model Statistics	
No. of Observations	3936
DF	3934
StDev	0.003336007
Variance	1.11289E-05
Log-Likelihood	-34366.34649
AICc	68738.69908
AIC	68738.69298
BIC	68757.52674

- The number of observations, n = 4032 – 96 (*withhold*) = 3936
- Degrees of freedom (DF) = 3936 (n) – 2 (terms in the model) = 3934
- Note that the model statistics are calculated using the deseasonalized data and residuals are multiplicative (relative) errors: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$.

25. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	3936	96
RMSE	101.5029509	776.7232991
MAE	73.73426784	649.5237662
MAPE	0.253138484	2.45436482
MASE	0.040157098	0.353743116

As expected, the **Out-of Sample (Withhold) Multi-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors. Note, if we were primarily interested in a short term one-step ahead forecast, then we would have selected **Withhold Forecast Type: One-Step-Ahead** and the above table would show **Out-of Sample (Withhold) One-Step-Ahead Forecast** errors.

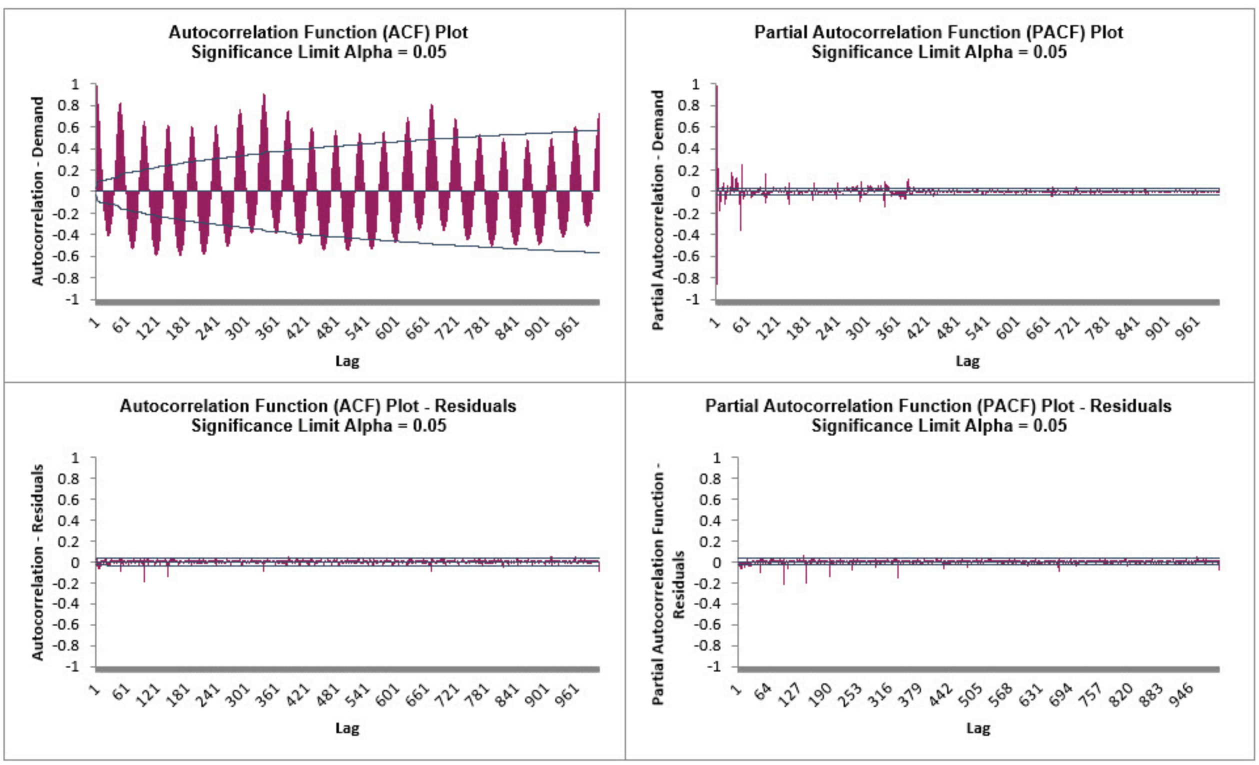
Forecast Accuracy metrics are calculated using the actual raw data versus forecast as displayed in the Forecast Chart and Table so, unlike the model statistics above, allows comparison across all forecast model types and transformations.

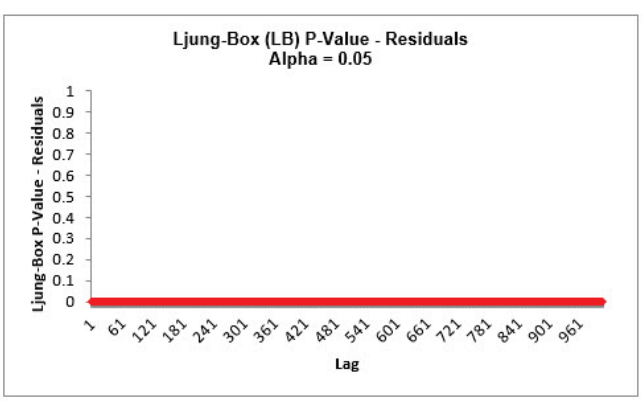
26. The Forecast Table is given as:

Forecast Table				
Period	Withhold Data	Multi-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
3937	24653	24595.68006	24399.19061	24792.1695
3938	23879	23839.01162	23561.1467	24116.87654
3939	23508	23546.05076	23205.74185	23886.35968
3940	23275	23300.43291	22907.48021	23693.38561
3941	22890	22735.89622	22296.56272	23175.22972
3942	22462	22349.79507	21868.5296	22831.06054

These are the same forecast and prediction interval values displayed in the Forecast Chart, but provided for further analysis or charting (e.g., a run chart of the forecast errors). The **Withhold Data** is also displayed.

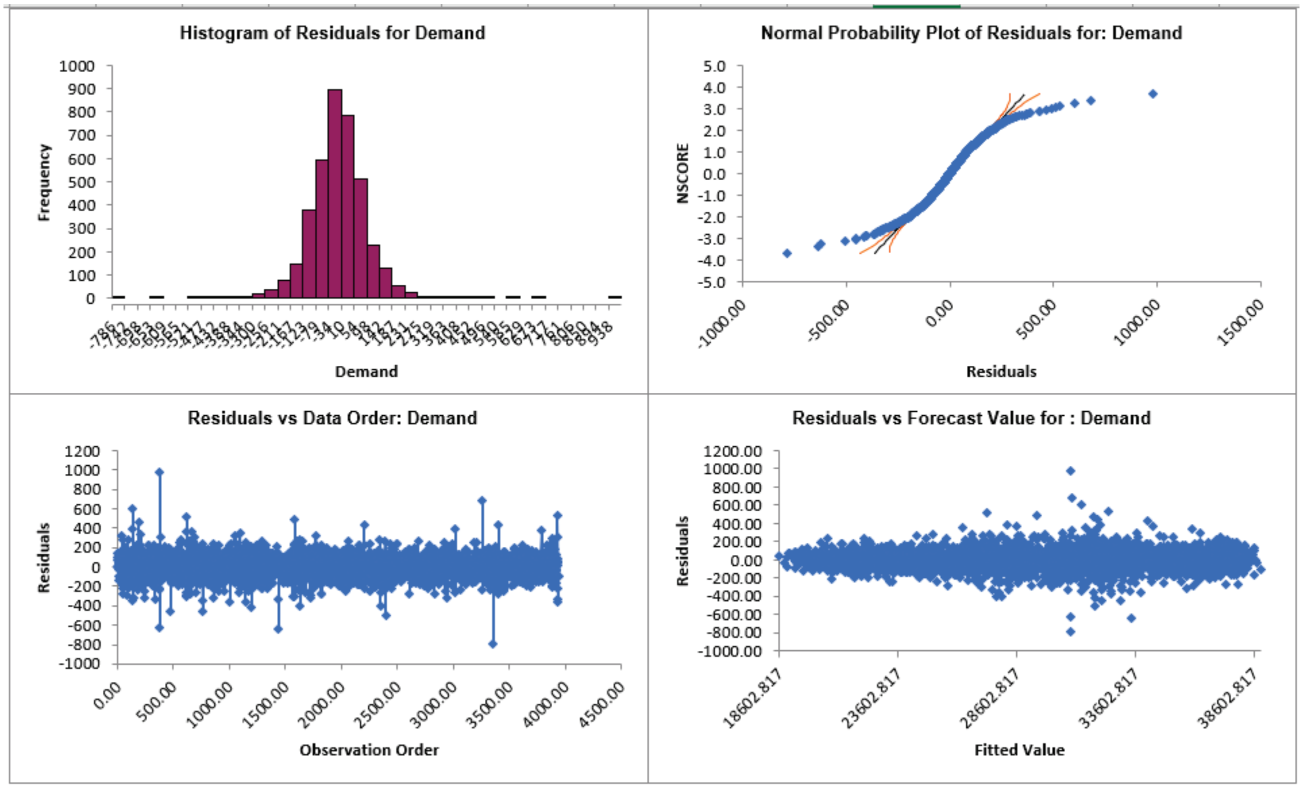
27. Click on the **Exp Smooth MSD ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ACF/PACF Residuals Plots indicate that much of the autocorrelation has been accounted for in the model, but the Ljung-Box plot shows that some significant autocorrelation still remains (the red P-Values are significant at alpha=.05) - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

28. Click on the **Exp Smooth MSD Residuals** sheet to view the Residual Plots:



The residuals are not normally distributed and there are extreme outliers. These should be investigated with a control chart on the residuals. Outliers in Electricity Demand are often explained by Temperature and this is something we will look at later with a different data set (Daily Electricity Demand with Predictors – ElecDaily).

29. Note that Residuals for MSD are the final observed - predicted values, so there are no scaling differences if the model uses additive or multiplicative error. If a Box-Cox transformation is used, then the residuals are in transformed units.

Exponential Smoothing Control Chart

Statistical Process Control (SPC) For Autocorrelated Data

An Individuals control chart is created using the residuals of the Exponential Smoothing forecast model.

The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits. If the residuals are exponential smoothing multiplicative, the control limits are approximate and out-of-control signals may not exactly match the Individuals Chart. If that occurs, the Individuals Chart should be used to determine what points are out-of-control.

The popular “Add Data”, “Show Last 30” and “Scroll” features in SigmaXL Chart Tools are available for these control charts. For “Add Data”, the time series models are not refitted, but used to compute the residual values for the new data.

For further details and references, see the Appendix: **Control Charts for Autocorrelated Data**.

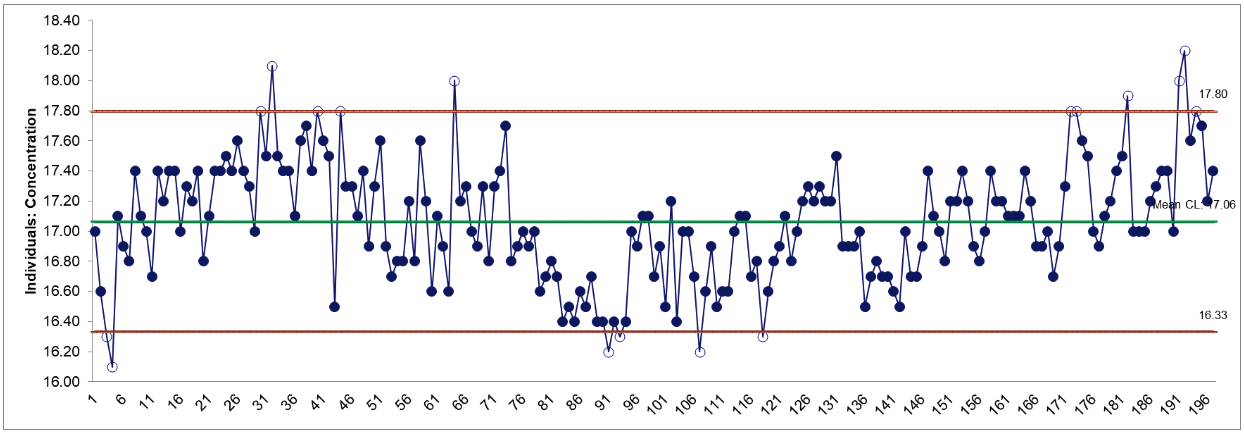
Note that a Moving Range Chart and Tests for Special Causes are not available here, but the user can store and select Residuals, then create with **SigmaXL > Control Charts > Individuals & Moving Range**.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

2. Earlier we saw that this process has significant autocorrelation. In order to see the impact on a control chart, we will construct an Individuals chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

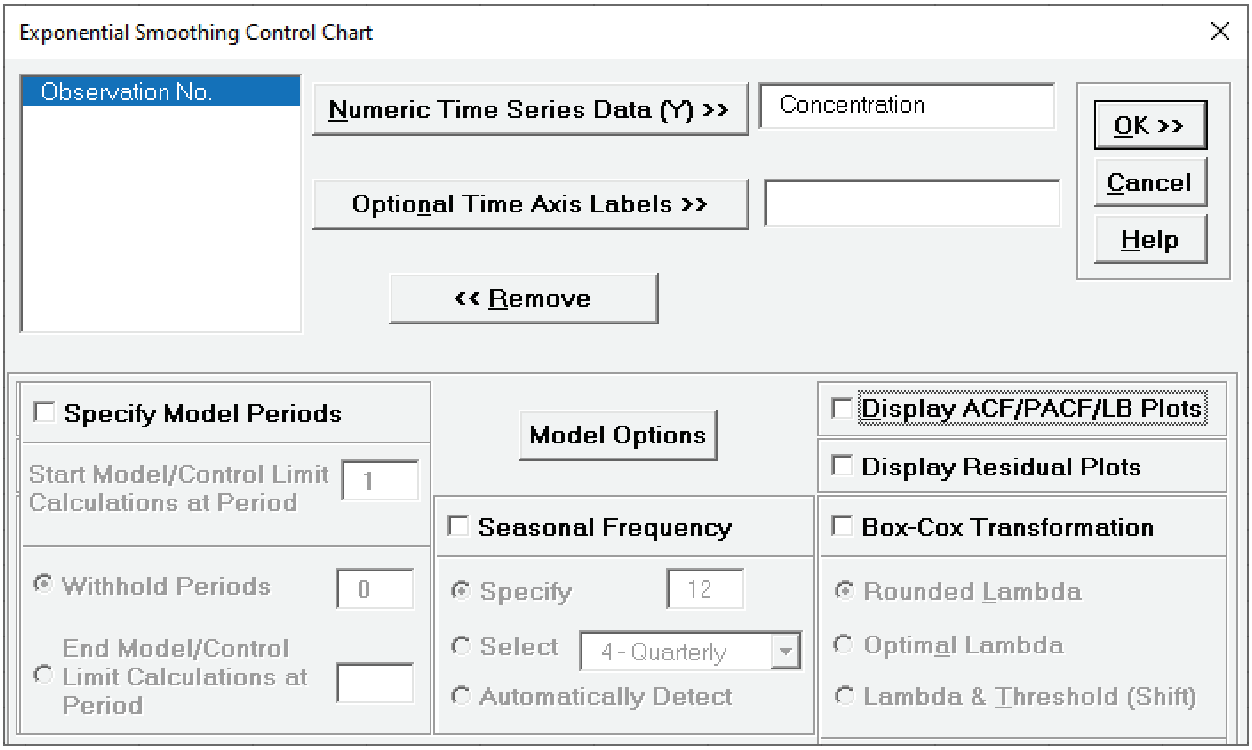
3. Select *Concentration*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



   There are 17 out-of-control data points, largely due to the autocorrelation. Searching for assignable causes using this chart as is would be futile.

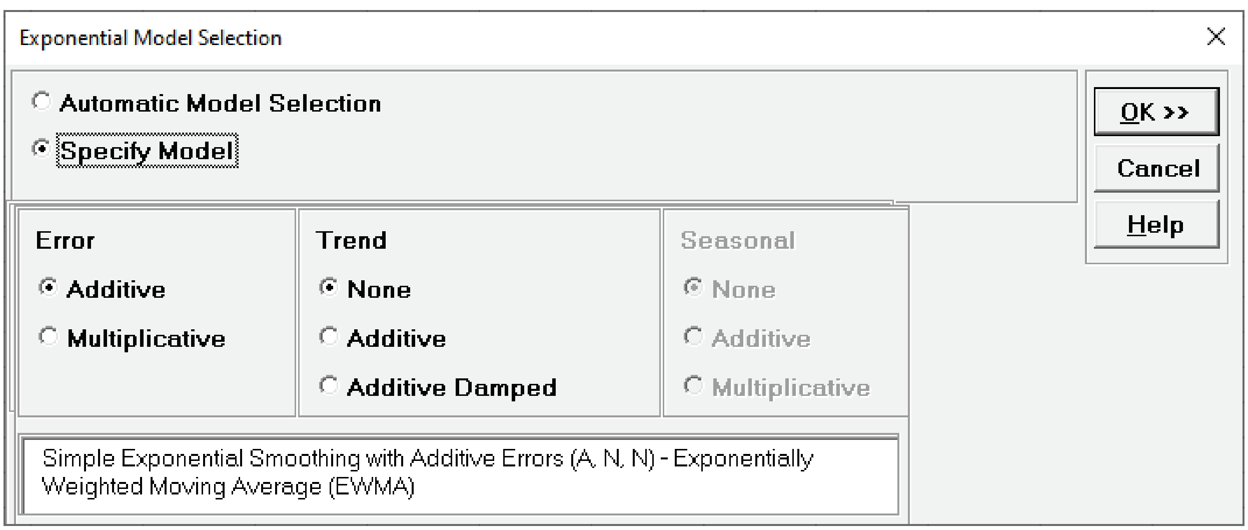
4. Now click **Sheet 1** tab and **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

5. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots.** Leave **Display Residual Plots**, **Specify Model Periods, Seasonal Frequency** and **Box-Cox Transformation** unchecked.

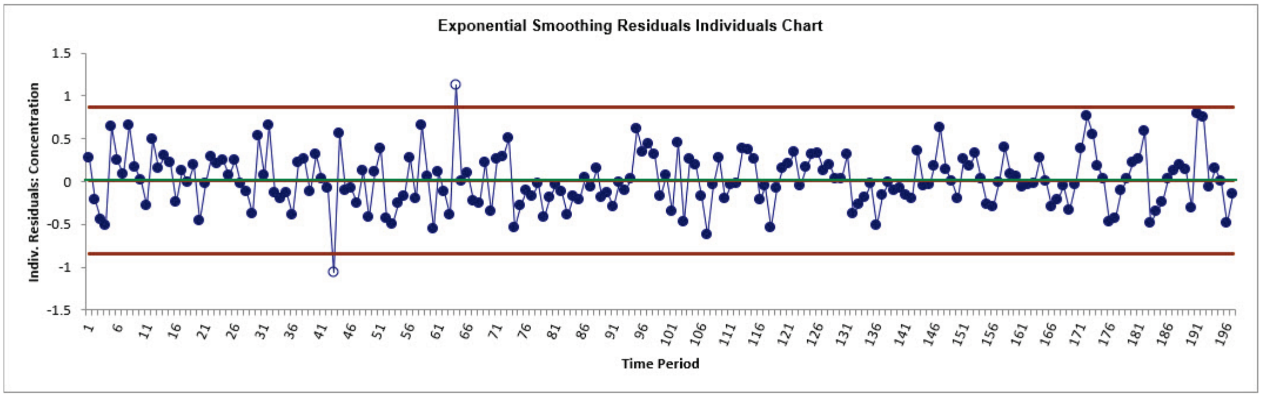


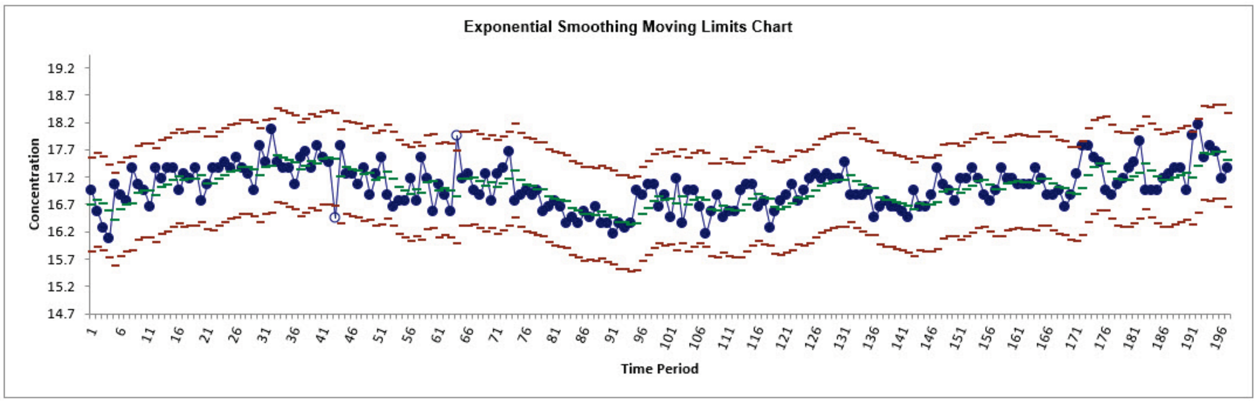
Since we will be running the same (A, N, N) model as used earlier, we will not need the ACF/PACF/LB and Residuals Plots.

6. Click **Model Options**. Select **Specify Model**.



7. We will use the default **Error: Additive** and **Trend: None**, which is a simple exponential smoothing model, or Exponentially Weighted Moving Average (EWMA). Click **OK** to return to the Exponential Smoothing Control Chart dialog. Click **OK**. The exponential smoothing control charts are produced:



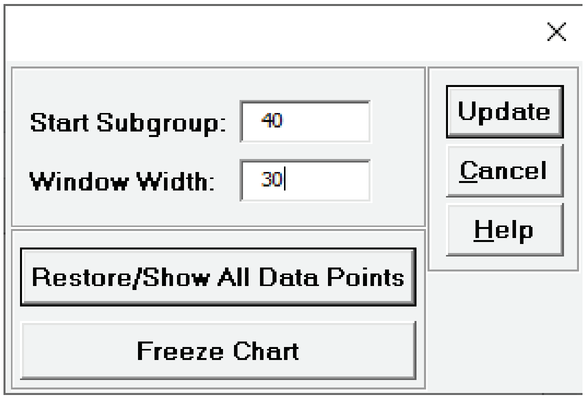


Now we only have two out-of-control data points on the Individuals chart to investigate. The Moving Limits chart uses the one step prediction as the center line, so the control limits move with the center line. The reduction in out-of-control data points is due to the removal of the autocorrelation as noted earlier in the ACF/PACF/LB Plots for this model.

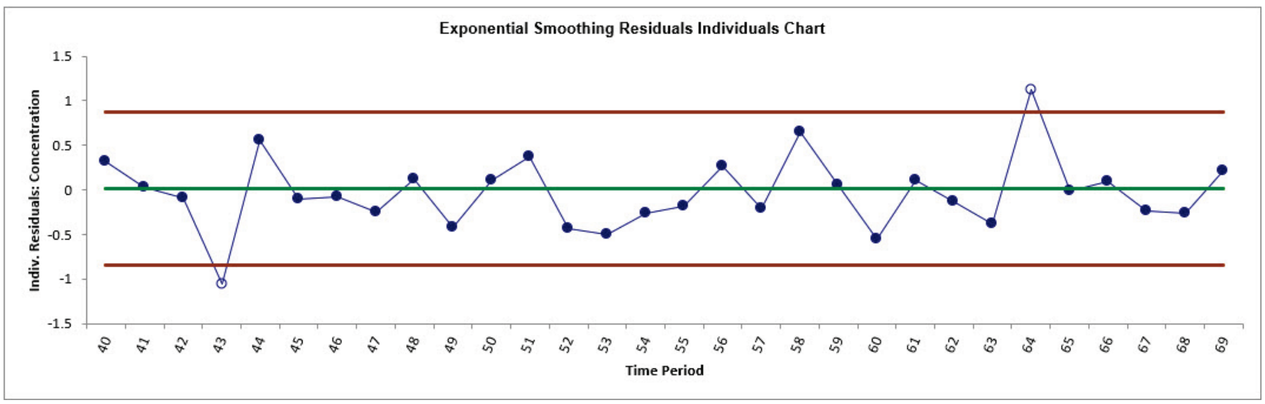
8. You can scroll through the chart data points. Click **SigmaXL Chart Tools > Enable Scrolling**

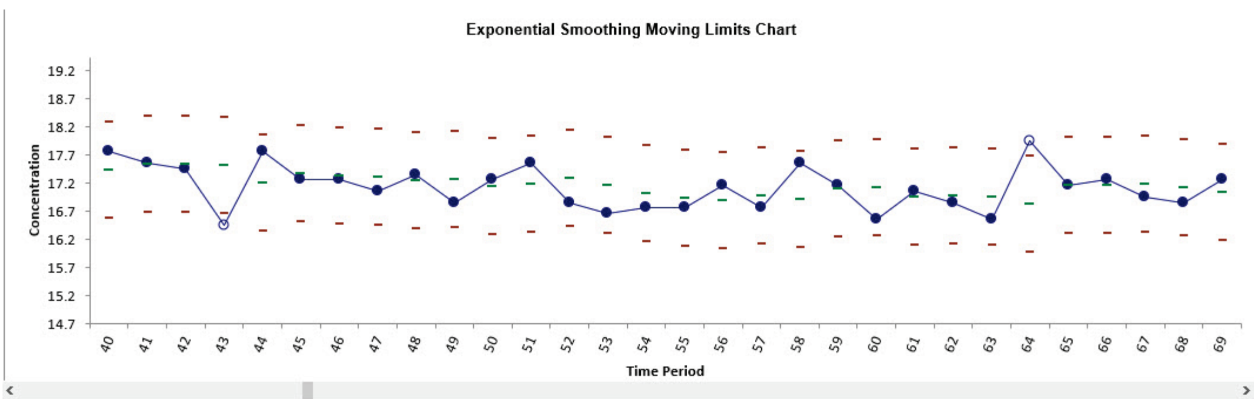
You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

9. Click **OK.** The scroll dialog appears allowing you to specify the **Start Subgroup** and **Window Width**. Enter **Start Subgroup** = 40 and **Window Width** = 30 to view the two out-of-control data points.



10. Click **OK.** This allows us to zoom in on the out-of-control points at 43 and 64.





Observation **43** is lower than expected from the exponential smoothing forecast model. Observation **64** is higher than expected.

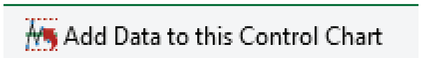
11. Click **Cancel** to exit the scroll dialog.

12. Now we will add a new data point to the Series A Concentration Data. The residuals will be computed using the same model as above without re-estimation of the model parameters or recalculation of the control limits. This is also known as the “Phase II” application of a Control Chart, where an out-of-control signal should lead to an investigation into the assignable cause and corrective action or process adjustment applied. Click Sheet1, enter the value 19 as shown in cell **B199** (and optionally Observation number 198 in cell **A199**).

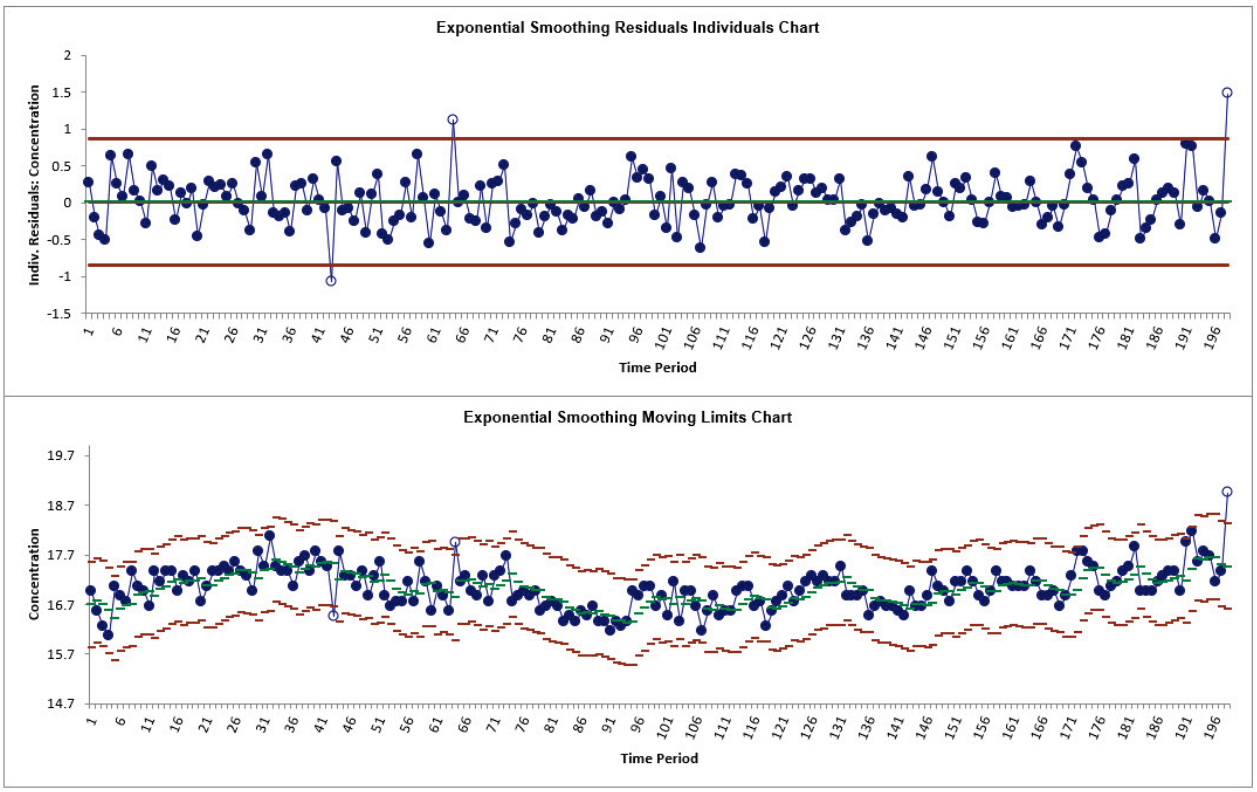
196	195	17.7
197	196	17.2
198	197	17.4
199	198	19

13. Click **Exp. Smoothing Control Charts** tab (if more than one control chart sheet exists in the workbook, please select the chart where the data will be added).

14. Click **SigmaXL Chart Tools > Add Data to this Control Chart**



15. The Residuals Individuals Control Chart and Moving Limits Charts are now updated with the new data, showing this as an out-of-control data point:



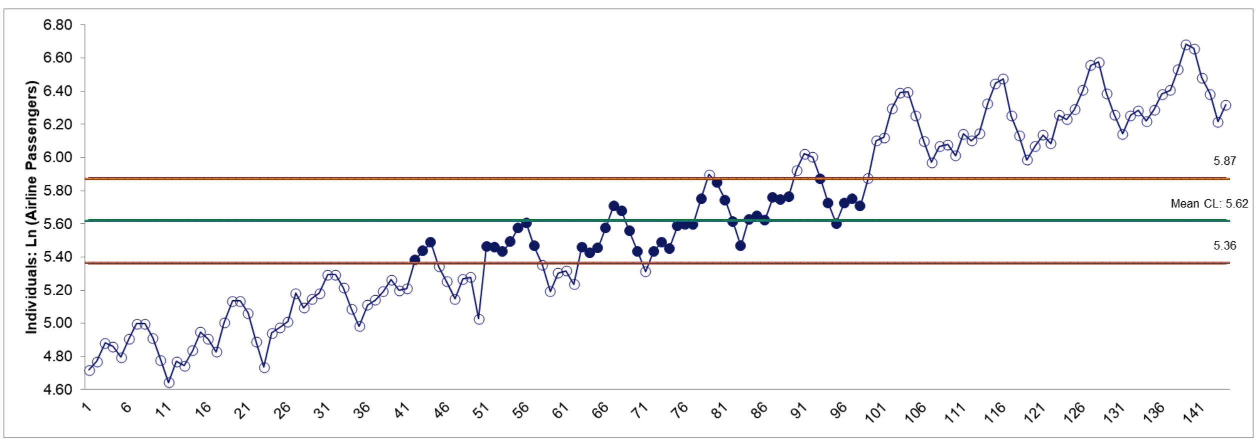
16. We recommend renaming the workbook to **Chemical Process Concentration – Series A_AddData.xlsx**, so that later use of the Concentration data does not include the added data point.

Monthly Airline Passengers – Modified for Control Charts

17. Open **Monthly Airline Passengers – Modified for Control Charts.xlsx** (**Sheet 1** tab). This is based on the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. A Ln transformation is applied (avoiding the need for a Box-Cox transformation), a negative outlier is added at 50 (-.25) and a level shift applied (+.25), starting at 100. Coded variables were added to help distinguish an outlier versus a shift, but they will be analyzed later using ARIMA Forecast with Predictors. Exponential Smoothing does not support predictors.

18. Earlier we saw that this process has significant autocorrelation with a strong trend and seasonality. In order to see the impact on a control chart, we will construct an Individuals chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

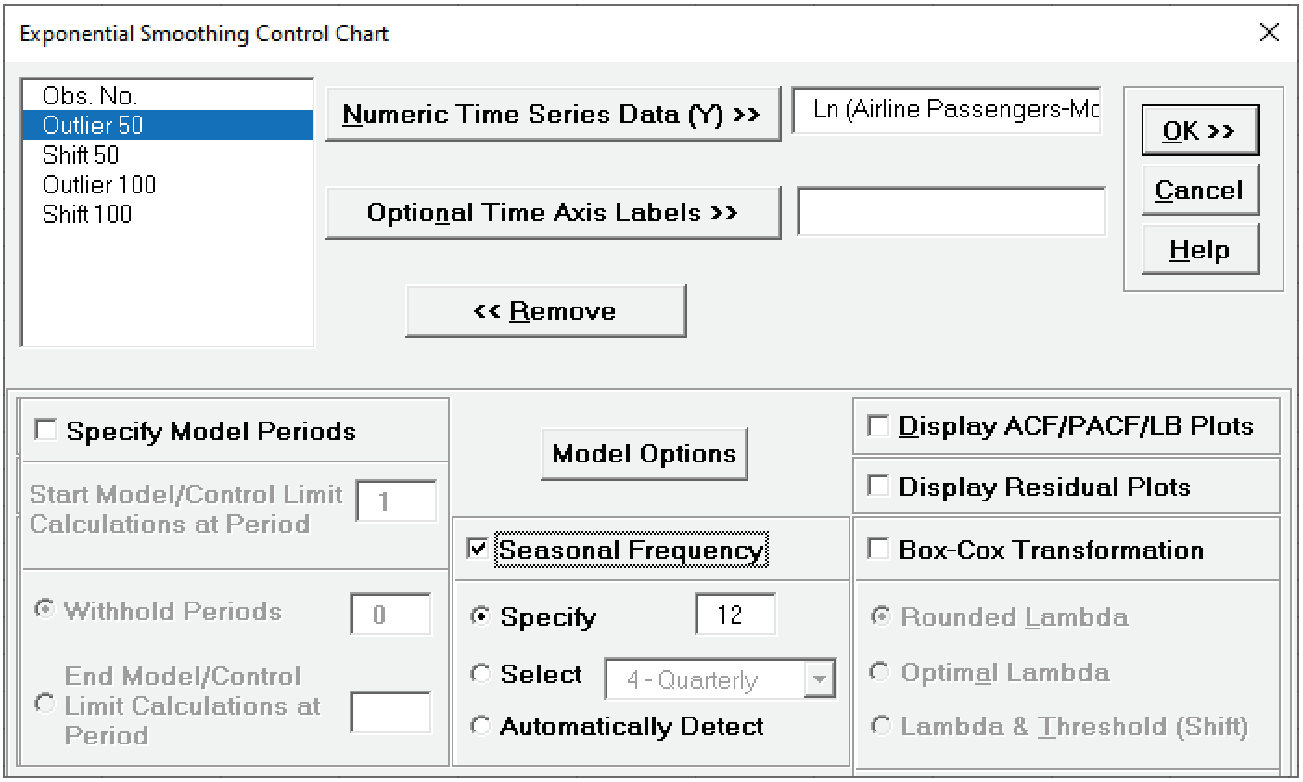
19. Select *Ln (Airline Passengers-Modified)*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



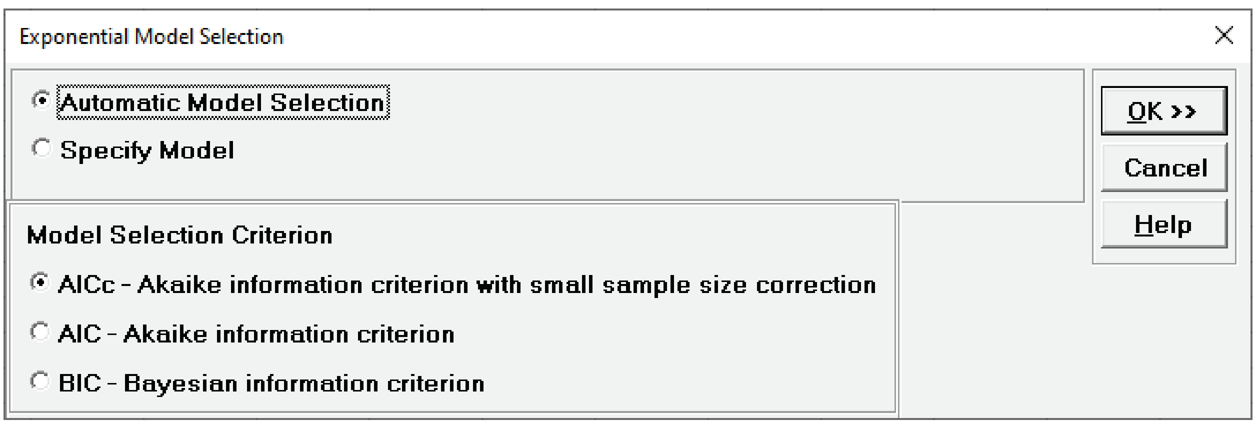
With strong trend, seasonality and positive autocorrelation, this control chart is meaningless.

20. Now click **Sheet 1** tab and **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

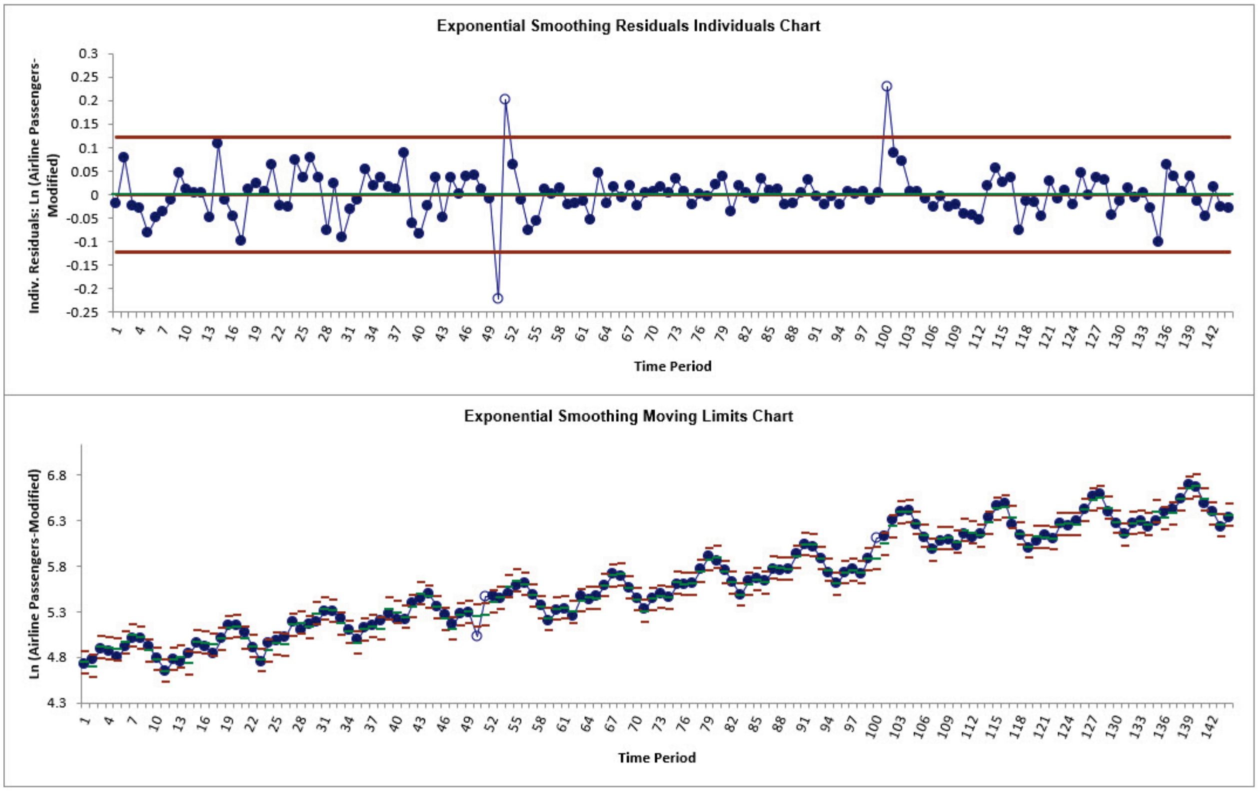
21. Select *Ln(Airline Passengers-Modified)*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots.** Check **Seasonal Frequency** with **Specify** = 12. Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



22. Click **Model Options**.



23. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing Control Chart dialog. Click **OK**. The exponential smoothing control charts are produced:

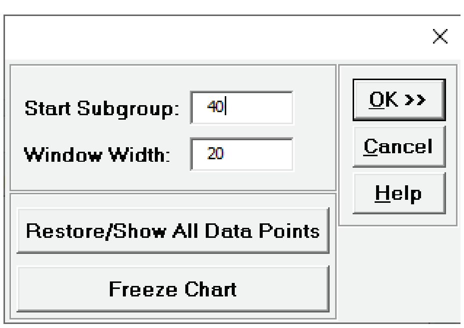


Now we can clearly see the out-of-control data points at 50, 51 and 100 on the Individuals chart. In order to view the points on the Moving Limits chart we will use scrolling.

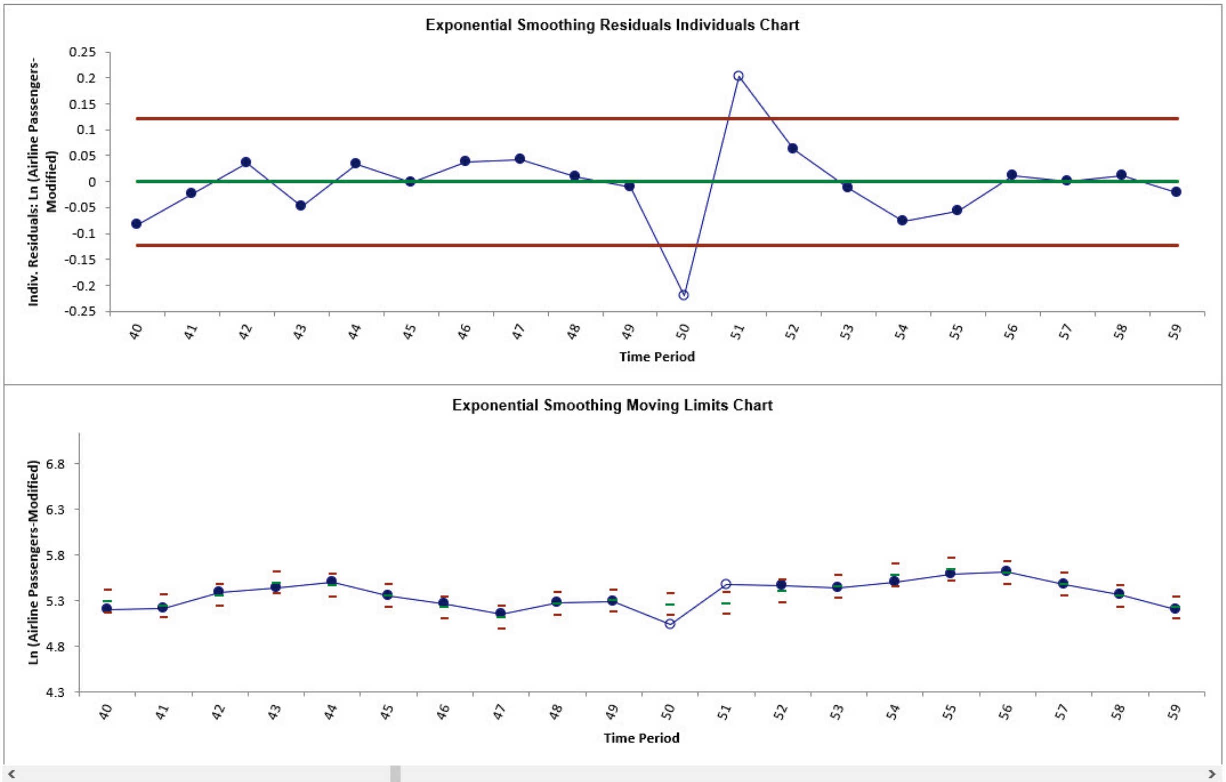
24. Click **SigmaXL Chart Tools > Enable Scrolling**

You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

25. Click **OK.** The scroll dialog appears allowing you to specify the **Start Subgroup** and **Window Width**. Enter **Start Subgroup** = 40 and **Window Width** = 20 to view the first two out-of-control data points.



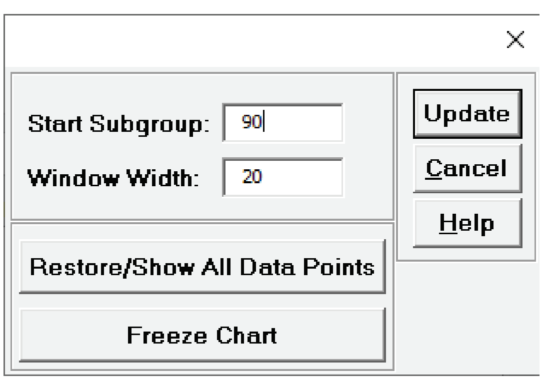
26. Click **OK.** This allows us to zoom in on the out-of-control points at 50 and 51.



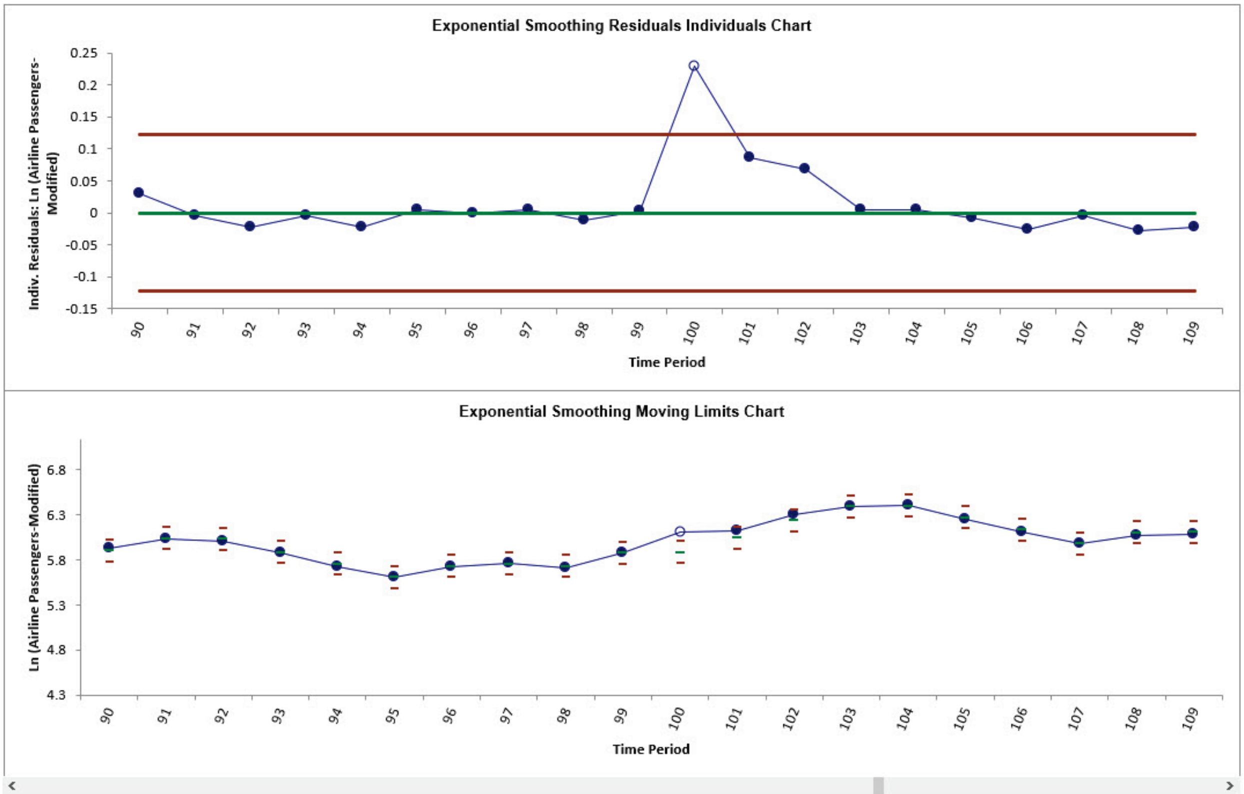
Observation **50** is lower than expected from the exponential smoothing forecast model. Observation **51** is higher than expected. Later investigation will reveal that this is a single negative outlier.

Tip: Scrolling keeps the original Y axis minimum and maximum setting. You may wish to change this to auto by clicking on the Y axis, right click **Format Axis**, click Bounds Minimum **Reset** and Bounds Maximum **Reset.** This changes the axis settings to Auto so when you scroll or Update the Y axis will automatically adjust as well.

27. Now enter **Start Subgroup** = 90 and **Window Width** = 20 to view the third out-of-control data point.



28. Click **Update**.



Observation **100** is higher than expected from the exponential smoothing forecast model. Later investigation will reveal that this is a shift in the mean.

29. Click **Cancel** to exit the scroll dialog.

30. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

The model **Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A)** was automatically selected as the best fit for the Modified Ln Airline Passenger data based on the AICc criterion.

31. The Parameter Estimates and Exponential Smoothing Model Statistics are slightly different than our earlier analysis because we have introduced an outlier and a shift, as well here we are using all of the data, i.e., there are no withhold periods. Note that earlier we used a Box-Cox Transformation with Lambda=0 and here we are using Ln of the data.

32. The Forecast Accuracy metrics are given as:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	144	
RMSE	0.049508351	
MAE	0.033506668	
MAPE	0.609538124	
MASE	0.230072112	

Note that these forecast errors are very different than our earlier analysis where the forecast errors were calculated on the raw data versus final predicted values, but here we are using Ln of the Airline Passenger data.

Exponential Smoothing Multiple Seasonal Decomposition (MSD) Control Chart

Exponential Smoothing is limited to a maximum seasonal frequency of 24. For higher frequencies use Exponential Smoothing – Multiple Seasonal Decomposition (MSD). The seasonal component is first removed through decomposition, a nonseasonal exponential smooth model fitted to the remainder (+trend), and then the seasonal component is added back in.

As the name implies, Multiple Seasonal Decomposition (MSD) also accommodates multiple seasonality, for example the half-hourly data with a seasonal frequency of 48 observations per day and 336 observations per week.

An Individuals control chart of the residuals is created for this forecast method. The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits. If the residuals are exponential smoothing multiplicative, the moving control limits are approximate and out-of-control signals may not exactly match the Individuals Chart. If that occurs, the Individuals Chart should be used to determine what points are out-of-control.

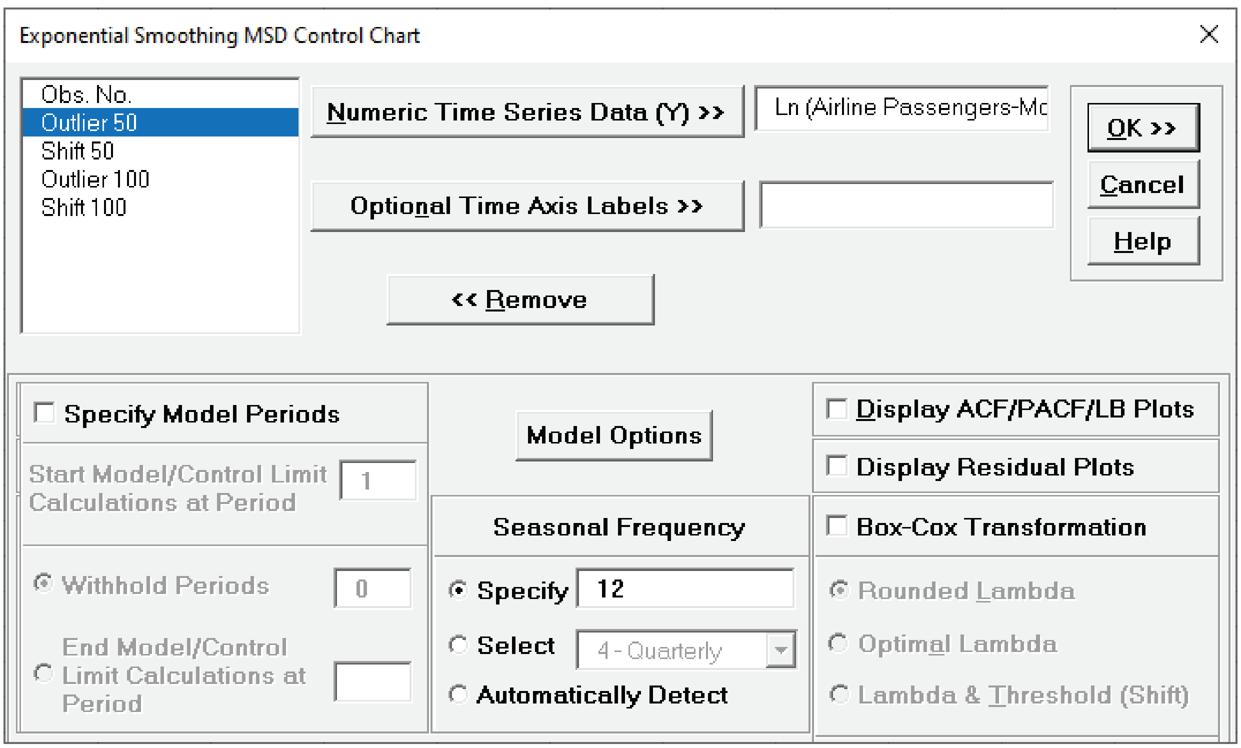
The popular “Add Data”, “Show Last 30” and “Scroll” features in SigmaXL Chart Tools are available for these control charts. For “Add Data”, the time series models are not refitted, but used to compute the residual values for the new data.

Monthly Airline Passengers – Modified for Control Charts

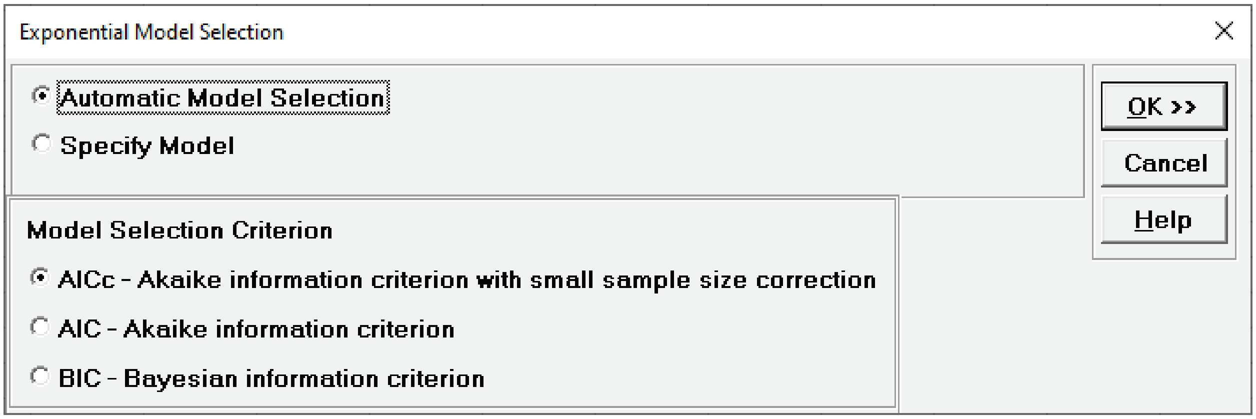
1. Open **Monthly Airline Passengers – Modified for Control Charts.xlsx** (**Sheet 1** tab). This is based on the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. A Ln transformation is applied (avoiding the need for a Box-Cox transformation), a negative outlier is added at 50 (-.25) and a level shift applied (+.25), starting at 100. The Multiple Seasonal Decomposition (MSD) option is not necessary for this data, but by way of introduction, we will use this to compare to the previous analysis.

2. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Multiple Seasonal Decomposition Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

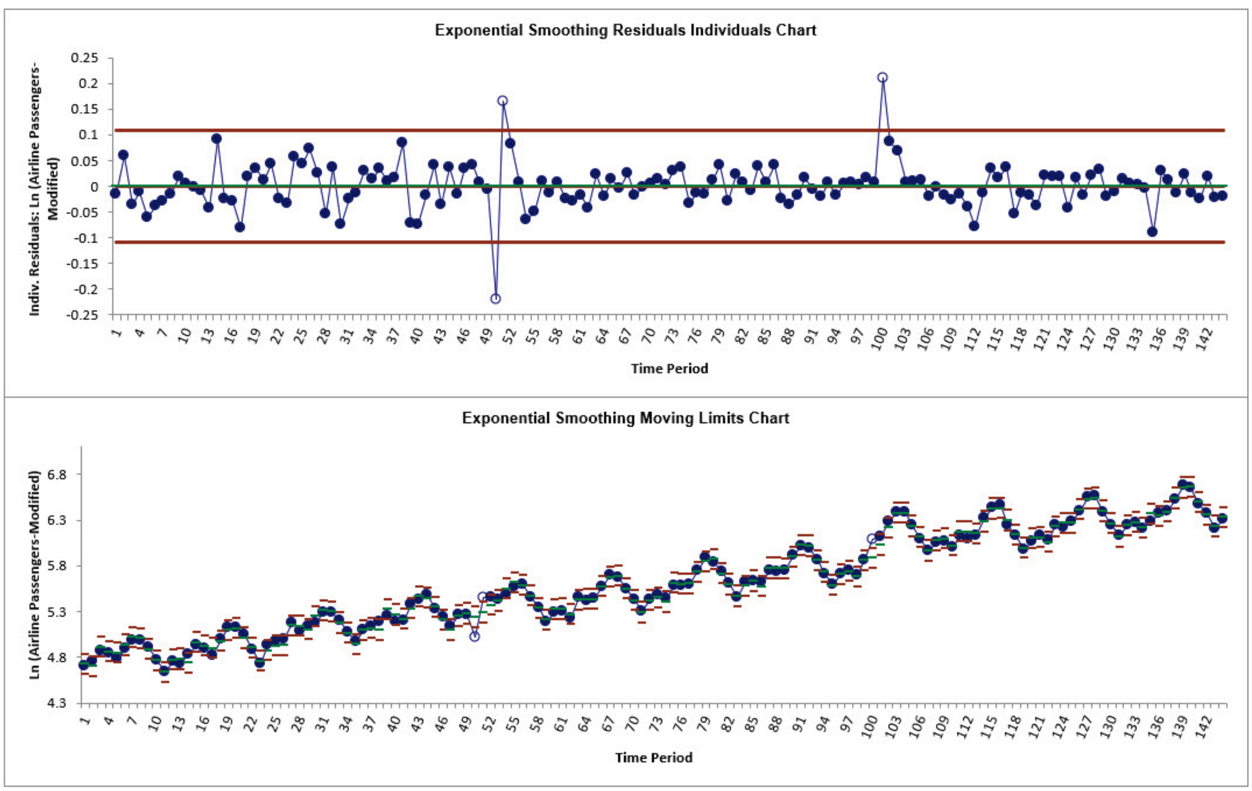
3. Select *Ln(Airline Passengers-Modified)*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots.** Check **Seasonal Frequency** with **Specify** = 12**.** Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



4. Click **Model Options**.



5. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing Control Chart dialog. Click **OK**. The exponential smoothing (MSD) control charts are produced:



We can clearly see the out-of-control data points at 50, 51 and 100 on the Residuals Individuals chart. This matches what we observed previously with regular Exponential Smoothing Control Charts.

6. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model (Multiple Seasonal Decomposition): Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

The model **Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N)** was automatically selected as the best fit for the deseasonalized Modified Ln Airline Passenger data based on the AICc criterion. There are no withhold periods.

7. The Exponential Smoothing Model Information is given as:

Exponential Smoothing Model Information	
Seasonal Frequency	12 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of model information with Seasonal Frequency = 12 using Decomposition and Model Selection Criterion = "AICc". The Box-Cox Transformation is "N/A".

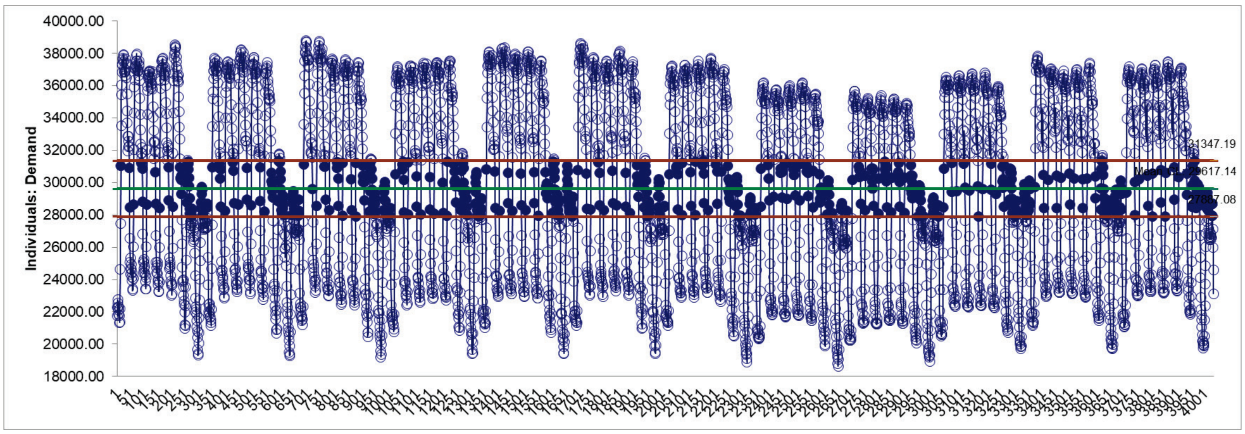
8. We will not review the Parameter Estimates, Model Statistics, and Forecast Accuracy as they are close to the MSD values given earlier, although note that slight differences are due to the introduction of an outlier and a shift, as well now we are using all of the data, i.e., there are no withhold periods. Earlier we used a Box-Cox Transformation with Lambda=0 and here we are using Ln of the data.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

9. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

10. We will first construct a classical Individuals Control Chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

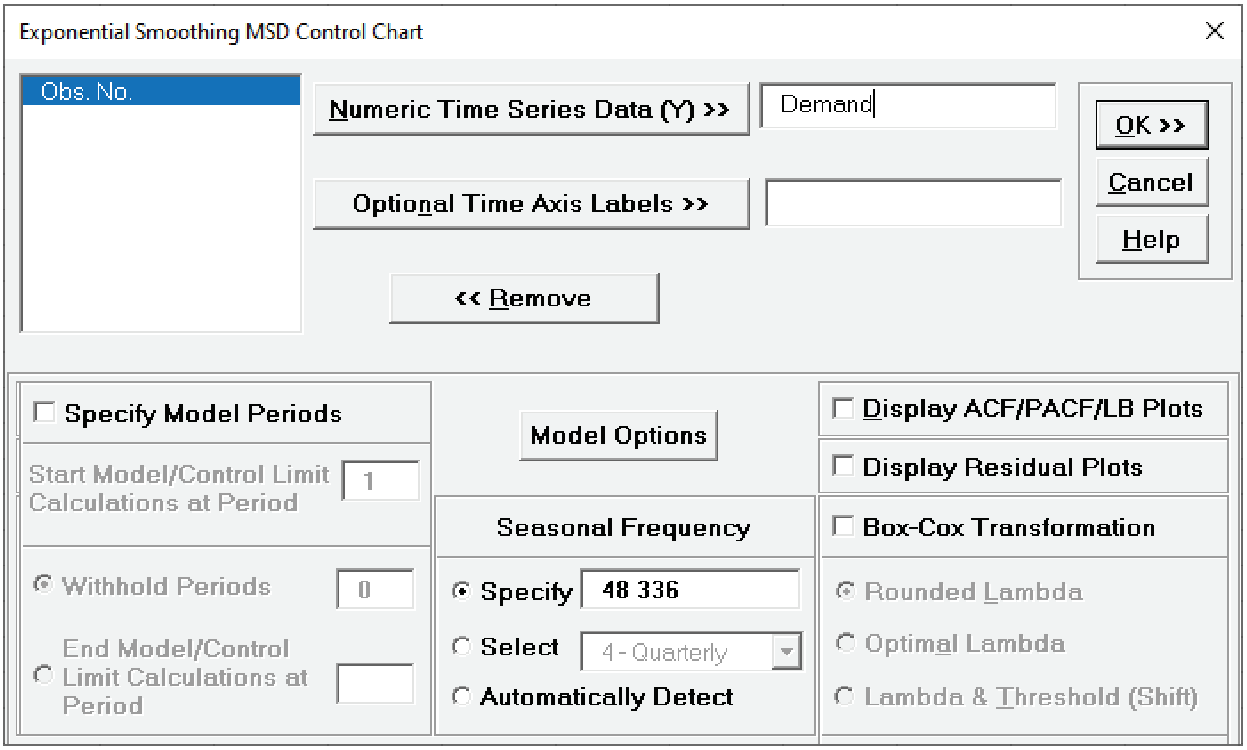
11. Select *Demand*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



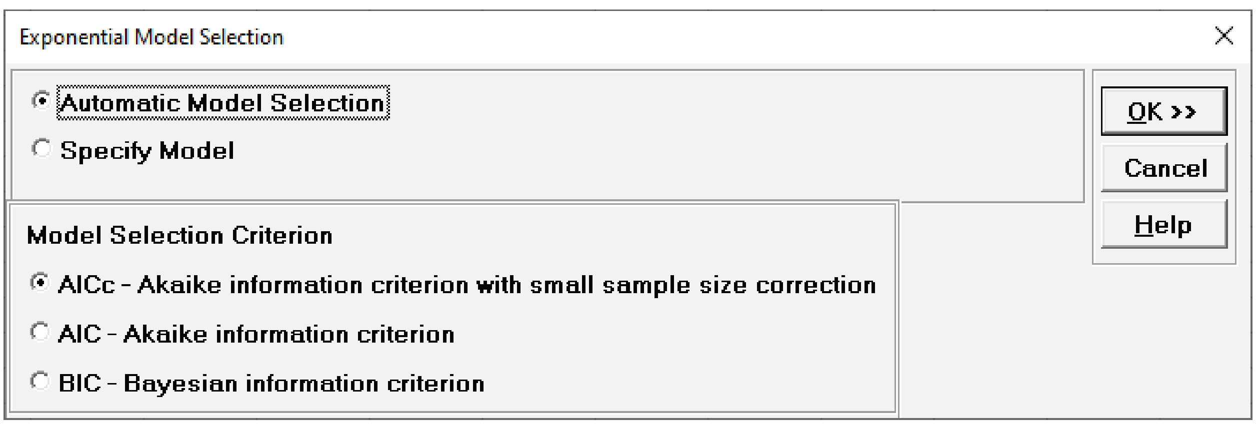
With the high frequency seasonality, this control chart is meaningless.

12. Click **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Multiple Seasonal Decomposition Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

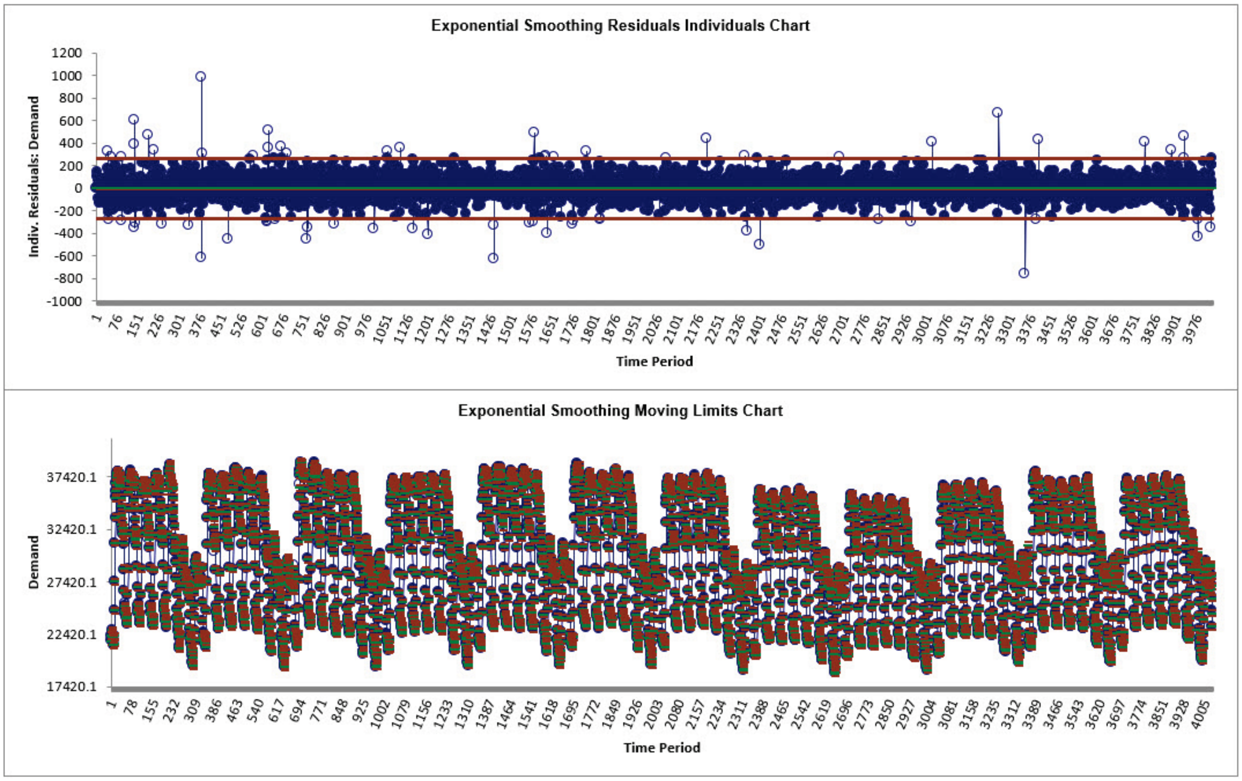
13. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Seasonal Frequency** with **Specify** = 48 336**.** Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



14. Click **Model Options**.

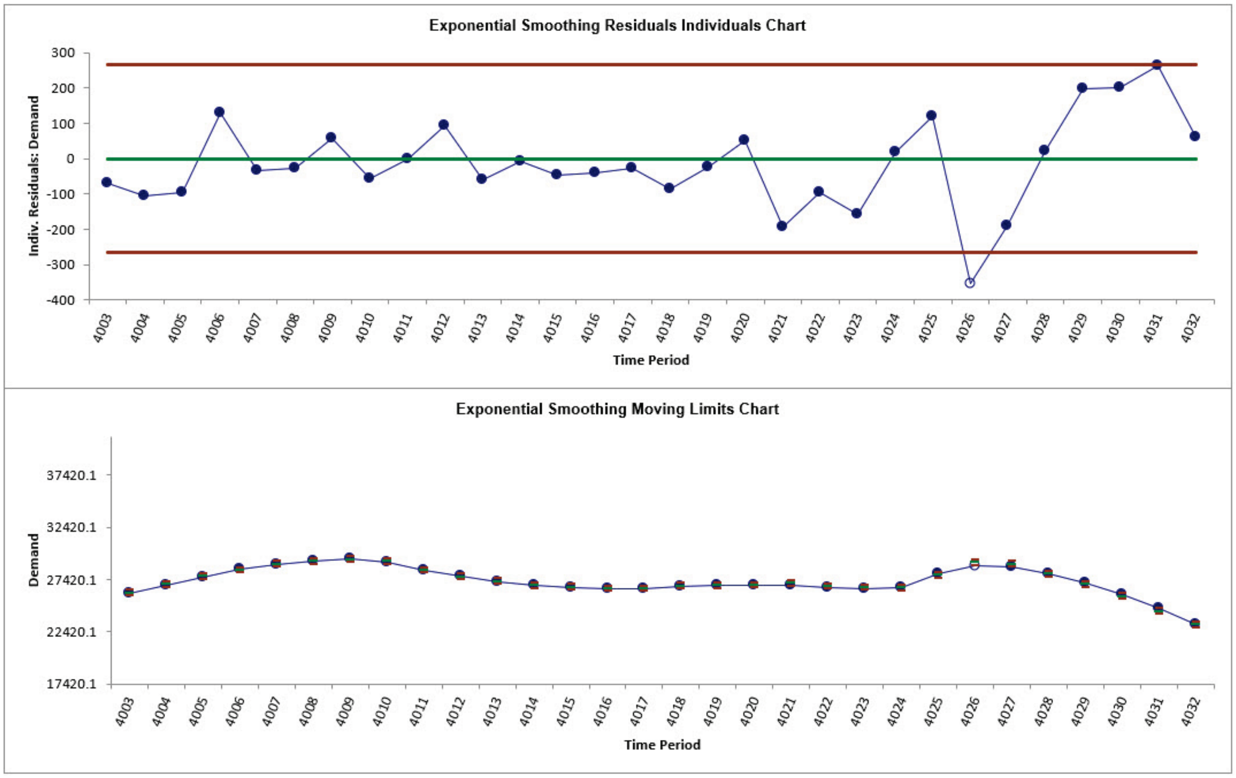


15. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the Exponential Smoothing MSD Control Chart dialog. Click **OK**. The exponential smoothing MSD control charts are produced:



Now we have a chart that can be used to identify assignable causes. The number of out-of-control signals have been dramatically reduced.

16. You can enable scrolling or zoom in to view the last 30 points on the Control Charts. Here we will do the latter. Click **SigmaXL Chart Tools > Show Last 30 Data Points**.



The out-of-control data point at 4026 is now clearly visible.

Tip: The Y axis scaling for the Moving Limits chart makes the same point difficult to see. You may wish to change the maximum and minimum values, right click **Format Axis**, and change the Bounds Minimum Maximum values.

17. To restore the chart, click **SigmaXL Chart Tools > Show All Data Points**.

18. Scroll down to view the Exponential Smoothing Model header:

Exponential Smoothing Model (Multiple Seasonal Decomposition): Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

The model **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** was automatically selected as the best fit for the deseasonalized Demand data based on the AICc criterion. There are no withhold periods.

This is the same model used previously in Exponential Smoothing MSD Forecast, but that used a withhold sample of 96.

ARIMA Forecast

The Autoregressive Integrated Moving Average (ARIMA) model was developed by Box and Jenkins. The default automatic determination of the best model order in SigmaXL uses the stepwise method of Hyndman and Khandakar (see fpp2).

Stationarity, Differencing and Constant

ARIMA assumes that the time series is stationary, i.e., it has the property that the mean, variance and autocorrelation structure do not change over time. If a time series mean is not stationary (e.g. trending), this can be corrected by differencing, computing the differences between consecutive observations for nonseasonal and between consecutive periods for seasonal data (e.g., Jan 2019 – Jan 2018, etc.). For nonseasonal, this will typically involve 1 or 2 orders of differencing. This order is the Integrated term *d.* For seasonal, this will typically involve 1 order of differencing. This order is the Seasonal Integrated term *D*. A constant term *c* is optional:

- If $d + D = 0$, a constant term in the model is the mean.
- If $d + D = 1$, a constant term in the model is a trend (drift).
- If $d + D > 1$, a constant term would be a quadratic or higher trend, so constant should not be included.
- It is recommended that $d + D$ should not be > 3.
- If the variance is not stationary, use a Box-Cox transformation.

Autoregressive (AR) Model

In an autoregressive model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregressive indicates that it is a regression of the variable against itself:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where ε_t is white noise. This is like a multiple regression but with lagged values of y_t as predictors. We refer to this as an AR(p) model, an autoregressive model of order p (fpp2).

Moving Average (MA) Model

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

We refer to this as an MA(q) model, a moving average model of order q (fpp2).

Autoregressive Integrated Moving Average (ARIMA) Model

If we combine differencing with autoregression and a moving average model, we obtain a nonseasonal ARIMA model:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where y'_t is the differenced series. This is the ARIMA (p, d, q) model, where p is the number of autoregressive terms, d is the degree of differencing and q is the number of moving average terms.

Seasonal ARIMA

For seasonal, the model consists of terms that are similar to the nonseasonal components of the model, but include the seasonal components. The seasonal model is ARIMA (P, D, Q) and combined we have ARIMA (p, d, q) (P, D, Q).

ARIMA Model Order

Model order may be automatically determined or user specified. The **Stepwise Procedure** utilizes the stepwise method of Hyndman and Khandakar (see Appendix: **Automatic Model Selection**):

- Seasonal D (0 or 1) is determined using a Seasonal Strength test, or user specified.
- Nonseasonal d (0, 1, or 2) is determined using a modified KPSS unit root test, or user specified.
- AR (p), MA (q), with orders from 0 to a maximum of 5.
- Seasonal, SAR (P), SMA (Q), with orders from 0 to a maximum of 2.
- Constant (included or not included) if $d + D \leq 2$; not included if $d + D > 2$.

An **Extended Model Search** will search over the same range of order values as Stepwise but do so using all combinations, subject to the following constraints for consistency and computational efficiency:

- d and D are determined using the same methods as Stepwise, or specified by the user.
- Constant (included or not included) if $d + D \leq 2$; not included if $d + D > 2$.
- $p + q + P + Q \leq 7$.
- Computation time limit is specified by user.

In general, the default **Stepwise Procedure** is recommended over the **Extended Model Search**, as it is much faster and usually finds the best ARIMA model, or a simpler one that is close to the best ARIMA model.

Model Parameter Estimation and Missing Values

Model parameters are solved using nonlinear maximization of the Log-Likelihood function. Two general models are available - the conditional sum of squares (CSS) and the state space Kalman maximum likelihood. The CSS is always used for initial estimates and is used if $n > 500$ or seasonal frequency > 12 for computational efficiency. Kalman Filters permit exact calculations and can handle missing values. For CSS, if missing values are encountered, the largest contiguous range is used. For further details and formulas, see Appendix: **Autoregressive Integrated Moving Average - ARIMA.**

ARIMA Model Statistics and Information Criteria for Model Comparison

The ARIMA model statistics are similar to those used in Exponential Smoothing. Log-Likelihood is related to -Ln(Sum-of-Squares Error), so is maximized. Information Criteria AICc, AIC and BIC are calculated using -2*Log-Likelihood and incorporate a penalty for the number of terms in the model, so smaller is better. These are used in automatic model selection. AICc is the default Information Criterion, based on forecast error performance with competition data. For further details, see **Information Criteria for Model Comparison**.

Forecast Accuracy Metrics

The Forecast Accuracy Metrics for ARIMA are the same as those used in Exponential Smoothing:

Root mean squared error: RMSE $= \sqrt{\text{mean}(e_t^2)}$

Mean absolute error: MAE $= \text{mean}(|e_t|)$

Mean absolute percentage error: MAPE $= \text{mean}\left(\left|\frac{100e_t}{y_t}\right|\right)$

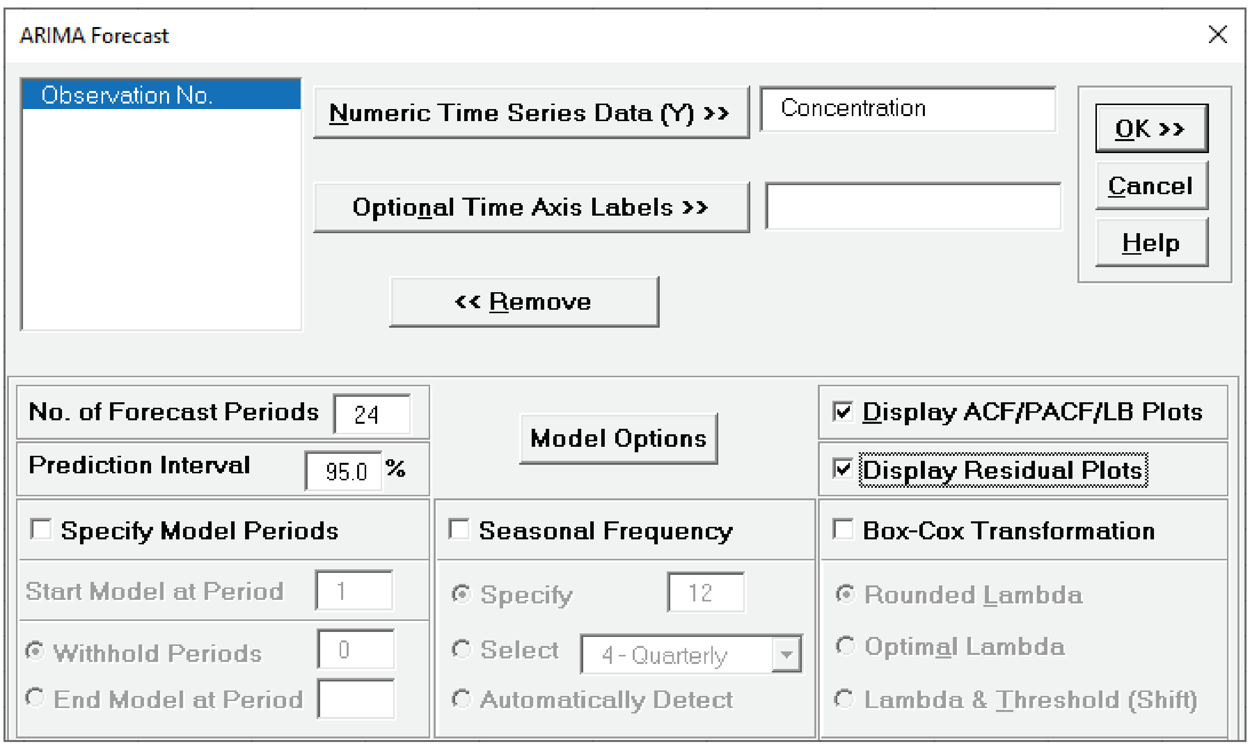
Mean absolute scaled error: MASE $= \text{mean}(|e_t|)/scale$

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

2. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Leave **Specify Model Periods, Seasonal Frequency** and **Box-Cox Transformation** unchecked. We will use the default **No. of Forecast Periods** = 24 and **Prediction Interval** = 95.0 %.



- **Optional Time Axis Labels** will be displayed on the forecast chart time axis. If used, dates for the forecast periods should also be included, otherwise the time axis will be blank for the forecast periods.

- **No. of Forecast Periods** are the number of time series values to be predicted (forecast horizon). The most accurate forecast will be for the first predicted value (one-step-ahead).

- **Prediction Interval** % is the confidence level for the individual predictions. For example, a 95% prediction interval contains a range of values which should include the actual future value with 95% probability. The interval will get larger the further out you predict.

- **Model Options** opens another dialog which allows you to set automatic options or to specify a model.

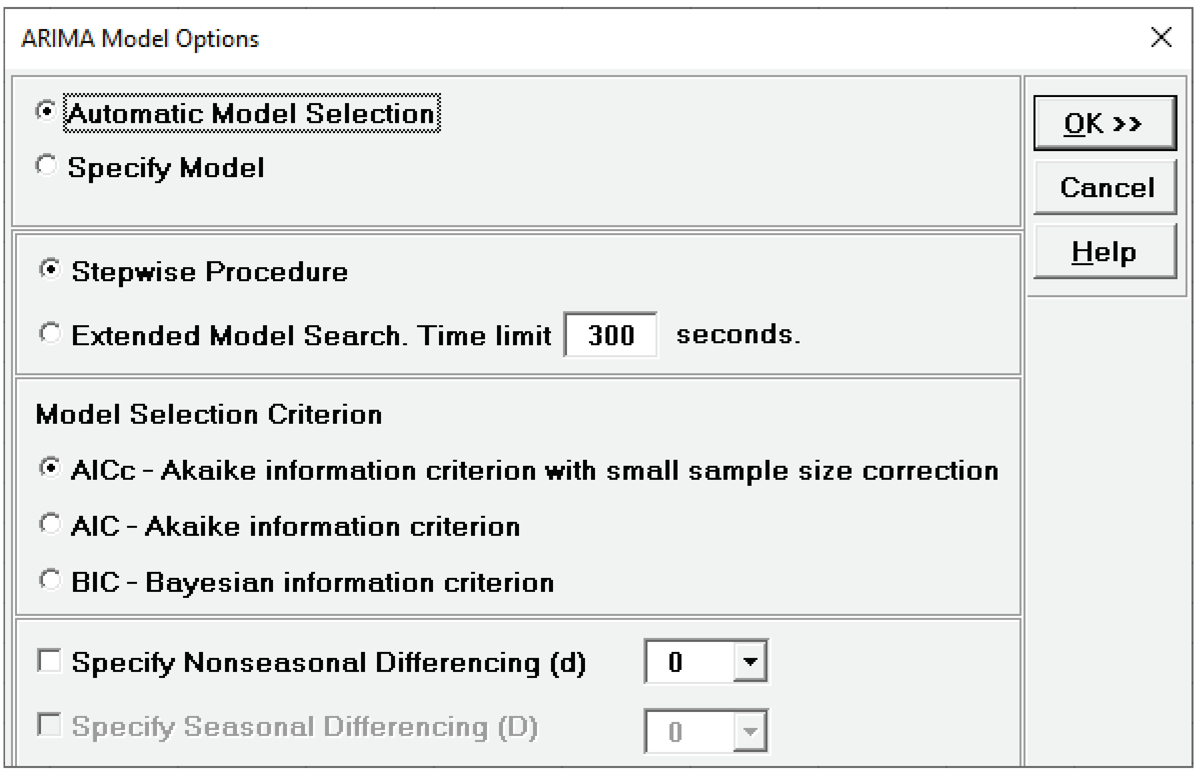
- **Display ACF/PACF/LB** option will produce ACF and PACF plots for the raw data as well as for the model residuals. The LB plot is a plot of Ljung-Box test P-Values for various lags and is used to determine if a group of autocorrelations are significant, (i.e., the autocorrelations do not come from a white noise series). For further details, see **Ljung-Box Test**.

- **Display Residual Plots** will produce a table of model residuals and the usual model residual plots: histogram, normal probability plot, residuals versus data order, and residuals vs forecast value. Note that if a Box-Cox transformation is applied, the residuals are transformed so will not be equal to forecast - actual.

- **Specify Model Periods** are used to specify a start period, end period or withhold sample. The withheld data is not used in model estimation, so this is very useful for model validation and comparison. This will be used in a later example.

- **Seasonal Frequency** and **Box-Cox Transformation** will be used in a later example.

4. Click **Model Options**.



- **Automatic Model Selection** will be used later. It is the default selection.

- **Stepwise Procedure** selects the use of the stepwise method.

- **Extended Model Search** is described above. The **Time limit** may need to be increased for seasonal models with high seasonal frequency and/or large number of observations.

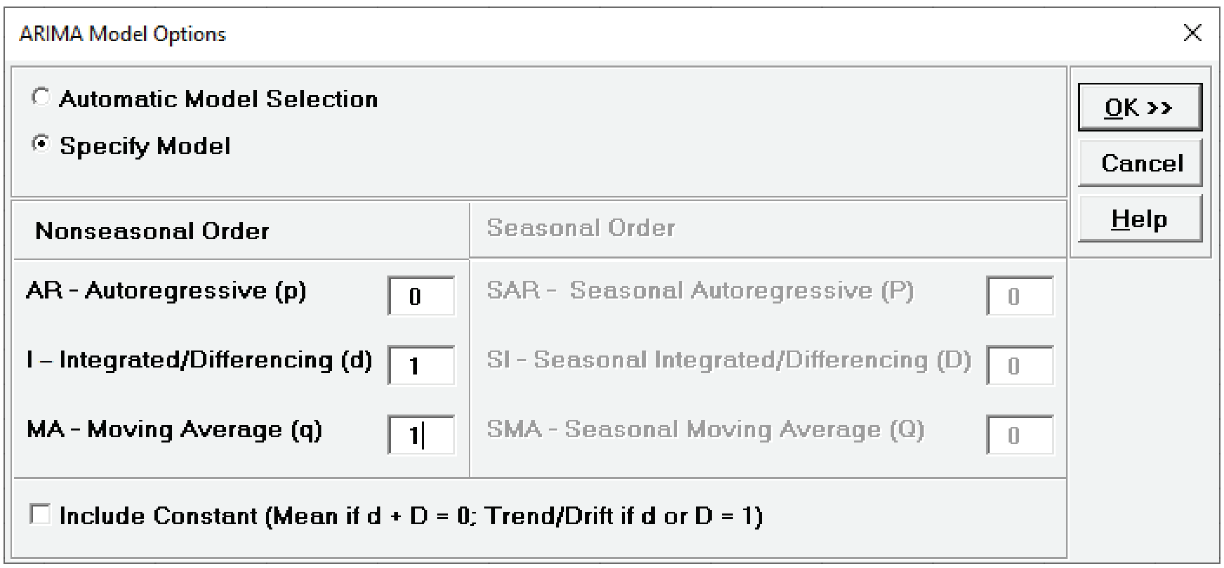
- **Model Selection Criterion** is the information criterion metric to be used in automatic model selection. AICc is the default selection.

- **Specify Nonseasonal Differencing (d)** = 0, 1, or 2, overrides the automatic nonseasonal differencing. This is useful to compare models for borderline cases that are "nearly nonstationary" (see Box and Jenkins).

- **Specify Seasonal Differencing (D)** = 0 or 1, overrides the automatic seasonal differencing. It is greyed out here because **Seasonal Frequency** was unchecked in the previous dialog.

- Clicking **OK** accepts the settings and returns you to the previous dialog. Clicking **Cancel** will cancel any changes and return you to the previous dialog.

5. Select **Specify Model**. Specify **Nonseasonal Order I – Integrated/Differencing (d)** = 1 and **MA – Moving Average (q)** = 1. Leave **Include Constant** unchecked.

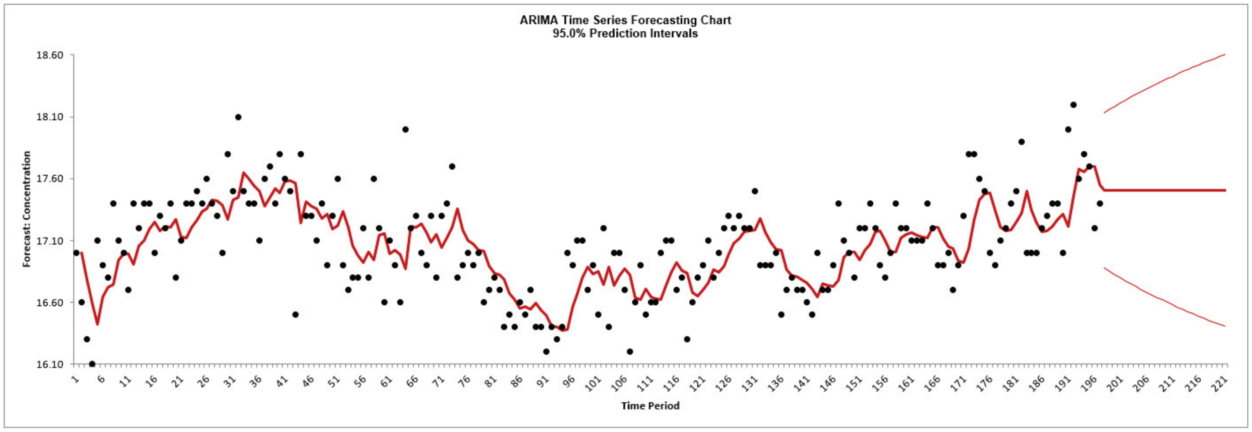


- **Specify Model** allows you to manually specify the **Nonseasonal Order** and **Seasonal Order** values and option for **Include Constant (Mean if d + D = 0; Trend/Drift if d or D = 1)**. If d + D = 0, then the constant term is the mean; if d or D = 1, then the constant term is a Trend; if d + D > 1, then the constant term is quadratic or higher – this is not recommended.

- **Seasonal Order** is greyed out because **Seasonal Frequency** was unchecked in the previous dialog.

The specified ARIMA (0,1,1) is equivalent to a simple exponential smoothing model (with slight differences due to estimation of the initial value).

6. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. The ARIMA forecast report is given:



As expected, this is very similar to the exponential smoothing forecast chart that was produced using the **Simple Exponential Smoothing with Additive Errors (A, N, N) – Exponentially Weighted Moving Average (EWMA)** model. The initial in-sample predicted value for ARIMA is slightly different and starts at the second time period due to differencing.

7. Scroll down to view the ARIMA Model header:

ARIMA Model: Concentration - User Specified Model
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

If we had checked Specify Model Periods in the main dialog, the start, end or withhold selection would be summarized here as well.

8. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	0
No. of Predictors	0
Model Selection Criterion	Specified
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of model information: ARIMA (0,1,1) with no constant and no predictors. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “Specified” because the model was user specified; and Box-Cox Transformation = “N/A” because Box-Cox Transformation was unchecked.

9. The Parameter Estimates are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.699386992	0.064505199	10.8423	0.0000

The MA_1 parameter coefficient value is approximately equal to 1 – alpha = 1 - 0.2948 = 0.7052 in Exponential Smoothing Parameter Estimates. The slight difference is due to estimation of the initial value.

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. This may be useful for model refinement with multiple predictors (and will be demonstrated later). Note that for AR/MA model order selection, minimum AICc should be used, rather than significance tests (see Kostenko, A.V. and Hyndman, R.J.).

10. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	197
DF	195
StDev	0.317609332
Variance	0.100875688
Log-Likelihood	-53.5087905
AICc	111.0797573
AIC	111.0175811
BIC	117.5738104

Degrees of freedom (DF) = *n* – 2 (1 term in the model, 1 order of differencing). See ARIMA Model Statistics and Information Criteria for Model Comparison.

Comparing to the Exponential Smoothing Model Statistics, we see that the StDev and Variance are approximately equal, but the Log-Likelihood, AICc, AIC and BIC are very different. This is due to different formulas being used in the Likelihood function. You cannot use Information Criteria to compare ARIMA and Exponential Smooth models to determine which model has the best fit.

11. The In-Sample Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	196	
RMSE	0.318324153	
MAE	0.248181659	
MAPE	1.45163544	
MASE	0.900807504	

MASE is less than one, so it is a better forecast than would be obtained from a naïve forecast (set all forecasts to be the value of the last observation). See Forecast Accuracy Metrics.

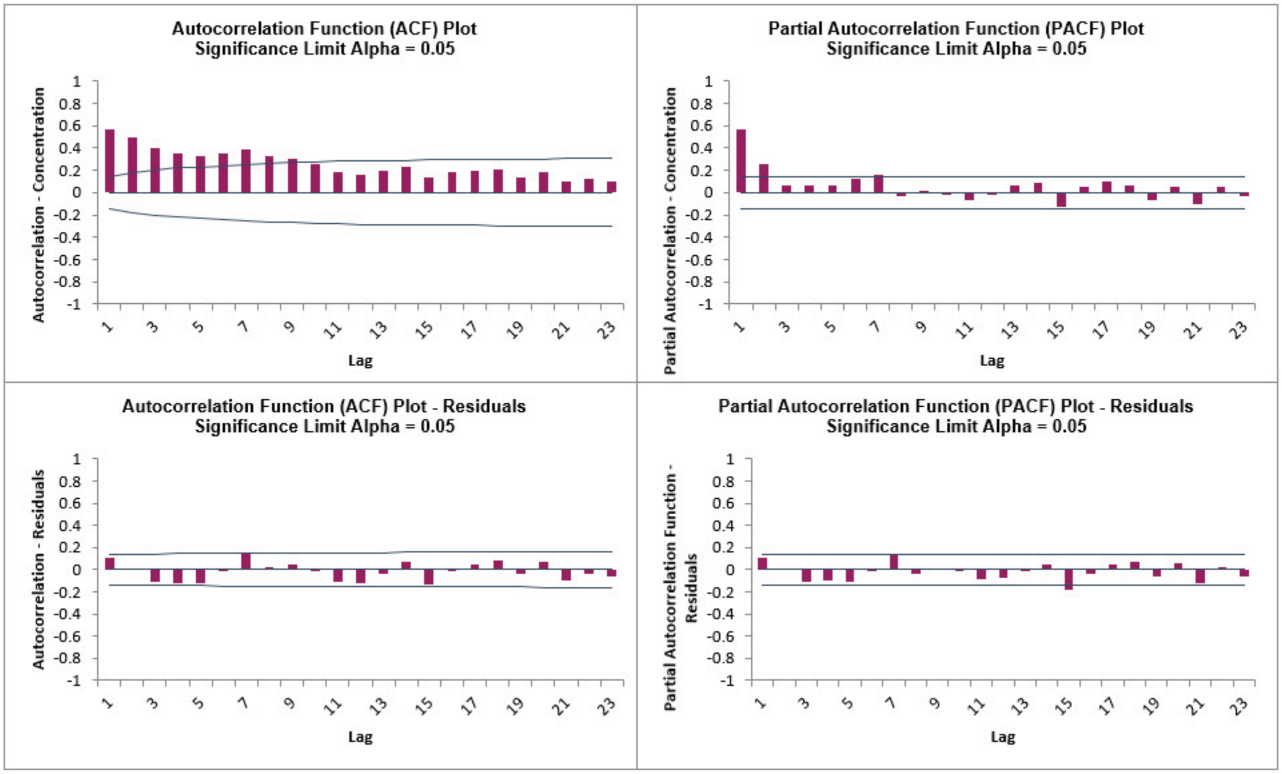
Comparing to the Exponential Smoothing Exponential Smoothing In-Sample Forecast Accuracy metrics, as expected, we see the forecast errors are approximately the same. Note that we lose one observation (N=196) since we do not have a predicted value at time period = 1 due to differencing.

12. The Forecast Table is given as:

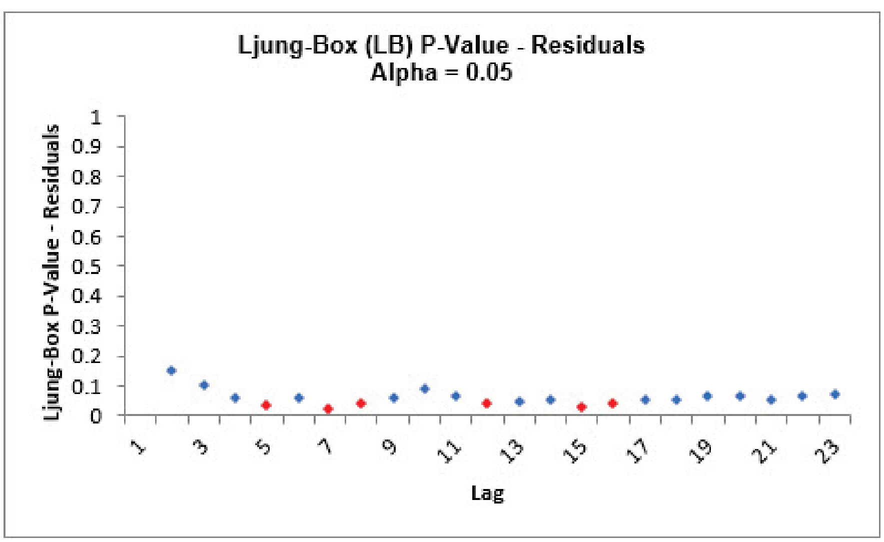
Forecast Table				
Period	Withhold Data	Forecast	Lower 95.0% PI	Upper 95.0% PI
198		17.50392441	16.87755395	18.13029487
199		17.50392441	16.84986402	18.1579848
200		17.50392441	16.82329967	18.18454915
201		17.50392441	16.79773386	18.21011496
202		17.50392441	16.77306181	18.23478701
203		17.50392441	16.74919586	18.25865296
204		17.50392441	16.72606181	18.28178701
205		17.50392441	16.70359618	18.30425264
206		17.50392441	16.68174418	18.32610464
207		17.50392441	16.66045813	18.34739069
208		17.50392441	16.63969619	18.36815263
209		17.50392441	16.61942146	18.38842736

These are the same forecast and prediction interval values displayed in the Forecast Chart but provided for further analysis or charting. If Withhold Periods are specified, the Withhold Data will be displayed as well.

13. Click on the **ARIMA ACF PACF LB** sheet to view the ACF/PACF/LB Plots:



These are approximately the same as what we obtained previously with Exponential Smoothing ACF/PACF/LB Plots. We can see that all of the autocorrelation has been removed by the exponential smoothing model (with the exception of lag 15 in the PACF), so this is a good fit to the time series data.

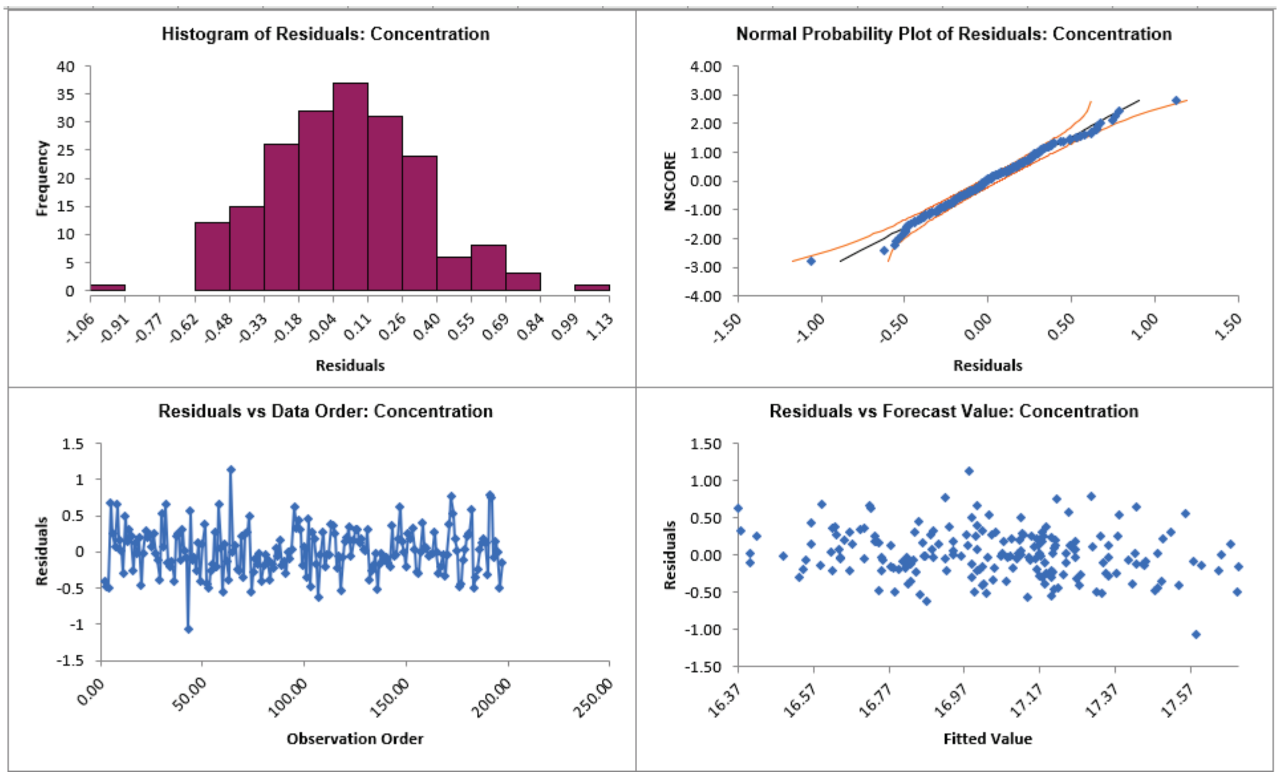


The LB plot is a plot of Ljung-Box test P-Values for various lags and is used to determine if a group of autocorrelations are significant, (i.e., the autocorrelations do not come from a white noise series). For further details, see **Ljung-Box Test**.

The red P-Values are significant (alpha=.05) and the blue P-Values are not significant. It is desirable that all P-Values be blue. The ACF/PACF plots indicated that almost all of the correlation has been accounted for in the model, but the Ljung-Box plot shows that some significant autocorrelation still remains - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

There does appear to be fewer significant P-Values than we obtained previously with the Exponential Smoothing LB plot, but this may not be a practical difference, given the similarity of all the other statistics.

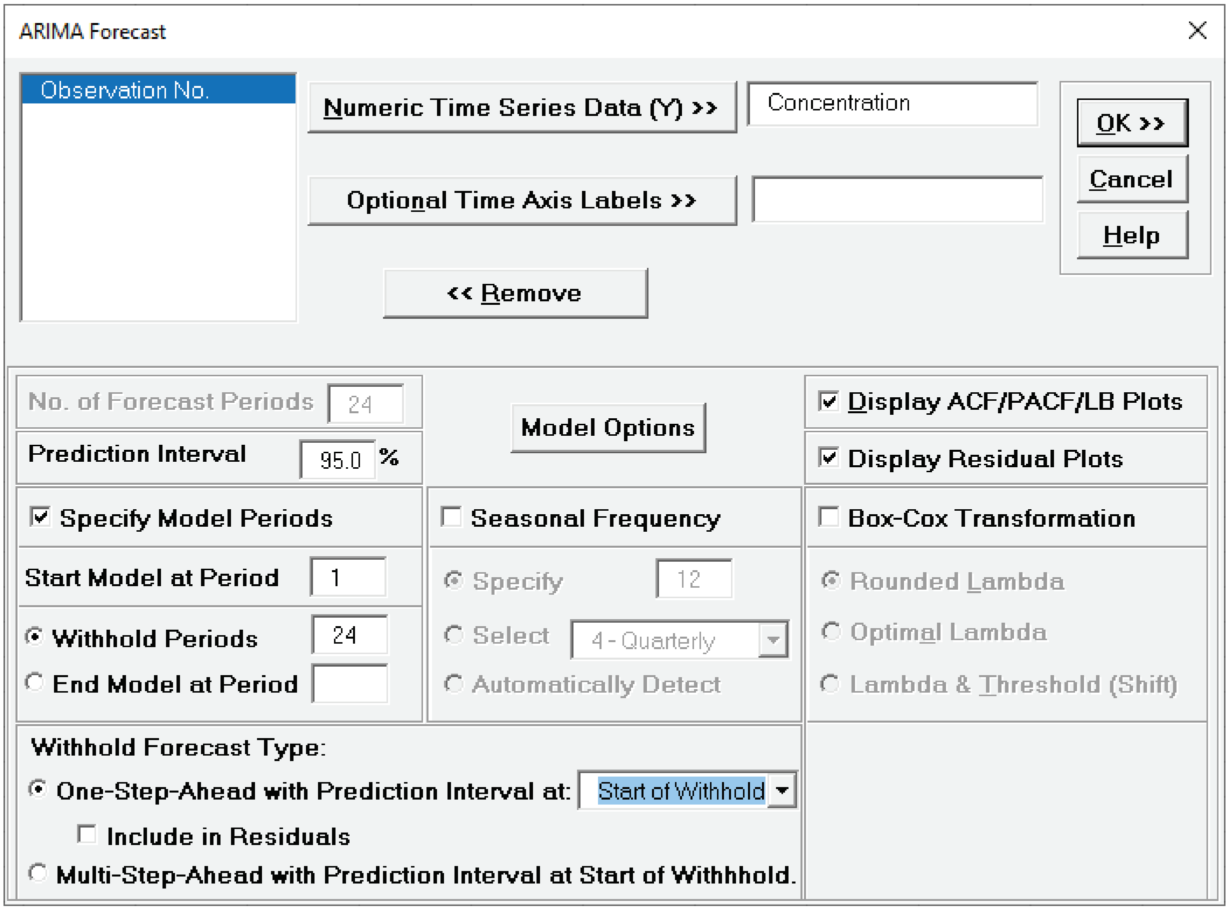
14. Click on the **ARIMA Residuals** sheet to view the Residual Plots:



The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts. Later, we will apply a control chart to the residuals to formally test for significant outliers or assignable causes.

These are approximately the same as what we obtained previously with Exponential Smoothing Residual Plots.

15. Now we will rerun ARIMA Forecast on the Concentration data with Automatic Model Selection and Specify Withhold Periods. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Check **Specify Model Periods**. Set **Withhold Periods** = 24. We will use the default **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold*.



- **Specify Model Periods** option allows you to specify the start and end periods used in automatic model identification and parameter estimation. Typically, **Start Model at Period** is kept = 1 and **Withhold Periods** specifies the number of periods to be withheld for out-of-sample testing. **End Model at Period** specifies the end period, so the withhold sample size would be: total number of observations – end period.

- **Withhold Forecast Type: One-Step-Ahead** will exclude the withhold sample from automatic model identification and parameter estimation, but uses the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you only care about the short-term one-step ahead prediction.

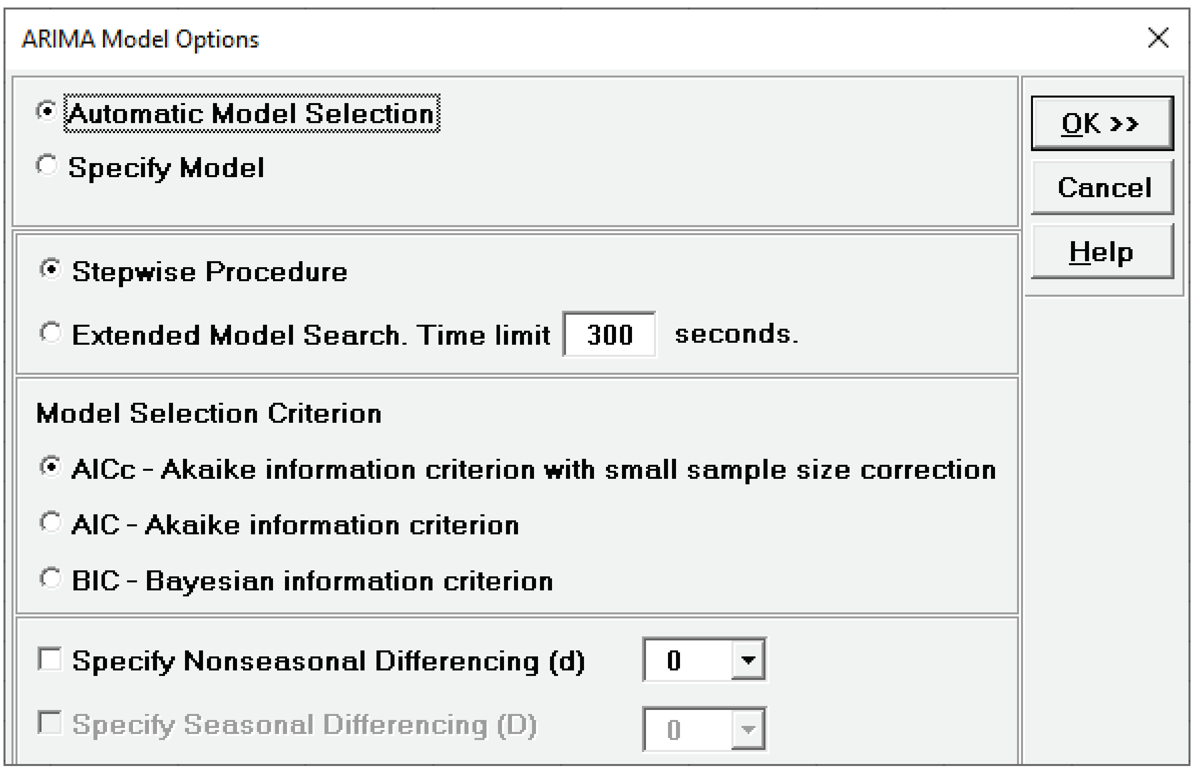
- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold* will display the prediction interval for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods.**

- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *End of Withhold* will display the prediction interval at the end of the withhold sample. The length of the prediction interval is determined by the specified **No. of Forecast Periods.**

- **Include in Residuals** will treat the one-step-ahead forecast errors as residuals (even though they were not part of the model estimation process) and will be included in the ACF/PACF/LB Residual Plots along with the Residuals report and graphs. Typically, this is kept unchecked.

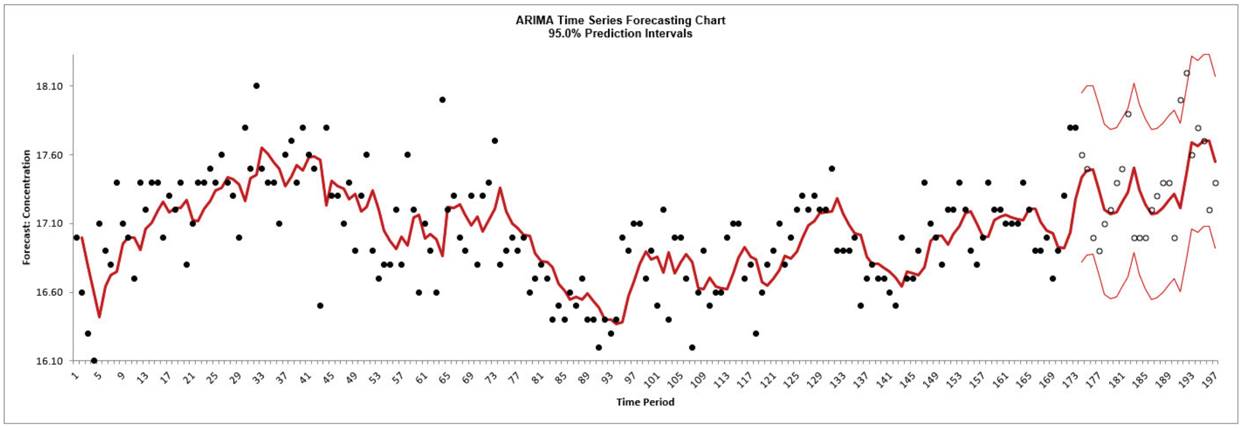
- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** will exclude the withhold sample from automatic model identification and parameter estimation and does not use the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you are interested in a long-term forecast window (horizon). The prediction interval will be displayed for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods.** These forecast errors are not included in ACF/PACF/LB Residual Plots or the Residuals report and graphs.

16. Click **Model Options**. Select **Automatic Model Selection**. We will use the defaults: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, leave **Specify Nonseasonal Differencing (d)** unchecked.



Tip: When using **Recall SigmaXL Dialog** and if there are no changes to the **Model Option** settings**,** the previous settings will be used. It is not necessary to repeat this step.

17. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. The ARIMA forecast report is given:



The blank dots are the data values in the withhold sample with a one-step-ahead forecast and prediction intervals displayed at the start of the withhold sample.

This is very similar to the exponential smoothing forecast chart that was produced using the **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** model. The initial in-sample predicted value for ARIMA is slightly different and starts at the second time period due to differencing.

18. Scroll down to view the ARIMA Model header:

ARIMA Model: Concentration - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

19. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

The ARIMA (0,1,1) model that we manually specified above, was also automatically selected based on the AICc criterion.

20. The Parameter Estimates are:

Parameter Estimates				
Term	Coefficient	SE Coefficient	T	P
MA_1	0.690028724	0.065305747	10.56613	0.0000

This is close to the parameter estimate obtained above (which used all of the data).

21. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	173
DF	171
StDev	0.312472686
Variance	0.09763918
Log-Likelihood	-44.18071519
AICc	92.4324363
AIC	92.36143039
BIC	98.65641934

These are fairly close to the model statistics obtained above using all of the data. Here we are using only 173 of the 197 observations.

22. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	172	24
RMSE	0.313295746	0.352257753
MAE	0.243952039	0.273585123
MAPE	1.431266547	1.568577996
MASE	0.881507368	0.988584899

As expected, the **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors.

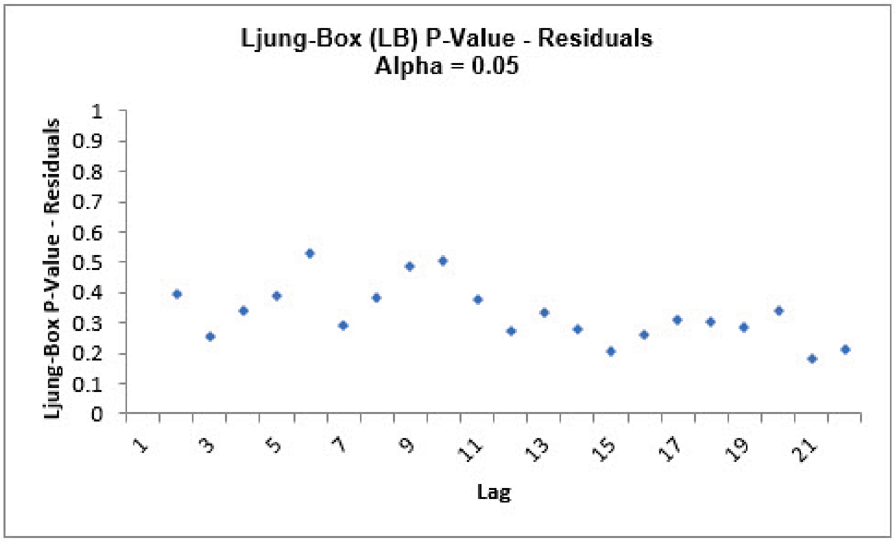
These are very similar to the Exponential Smooth Forecast Accuracy Metrics that were produced using the **Simple Exponential Smoothing with Multiplicative Errors (M, N, N)** model. Note that we lose one observation on the In-Sample (N=172) since we do not have a predicted value at time period = 1 due to differencing.

23. The Forecast Table is given as:

Forecast Table				
Period	Withhold Data	One-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
174	17.6	17.43699636	16.82022147	18.05377125
175	17.5	17.48752281	16.87207955	18.10296607
176	17	17.49139038	16.87775042	18.10503034
177	16.9	17.33907348	16.72282244	17.95532451
178	17.1	17.20297331	16.58502829	17.82091833
179	17.2	17.17105454	16.55470128	17.78740781
180	17.4	17.1800268	16.56542588	17.79462772
181	17.5	17.24821218	16.63450051	17.86192384
182	17.9	17.32625917	16.71316165	17.93935668
183	17	17.50410235	16.88693756	18.12126714
184	17	17.3478451	16.72798565	17.96770455
185	17	17.24002311	16.61980167	17.86024455
186	17.2	17.16562284	16.54613085	17.78511483
187	17.3	17.17627877	16.55847435	17.79408319
188	17.4	17.2146288	16.59825692	17.83100068
189	17.4	17.27208855	16.6568128	17.8873643
190	17	17.31173742	16.69785289	17.92562196
191	18	17.21510778	16.60123859	17.82897696
192	18.2	17.45840182	16.83590796	18.08089568
193	17.6	17.68827595	17.05843876	18.31811315
194	17.8	17.66091294	17.03262123	18.28920466
195	17.7	17.70402594	17.07708046	18.33097141
196	17.2	17.70277801	17.07747894	18.32807709
197	17.4	17.54693127	16.91921585	18.17464669

These are the same forecast and prediction interval values displayed in the Forecast Chart, but provided for further analysis or charting. The **Withhold Data** is also displayed.

24. The ACF/PACF/LB Residual Plots and Residual Plots are based on the in-sample data. The plots look similar to the complete data above, except for the Ljung-Box P-Values:



As was the case with exponential smoothing, the ARIMA (0,1,1) model is a better fit to the subset than the complete data, with all P-Values being blue (> .05).

25. If **Include in Residuals** was checked then the residuals would also include the **Out-of-Sample (Withhold) One-Step-Ahead Forecast** errors.

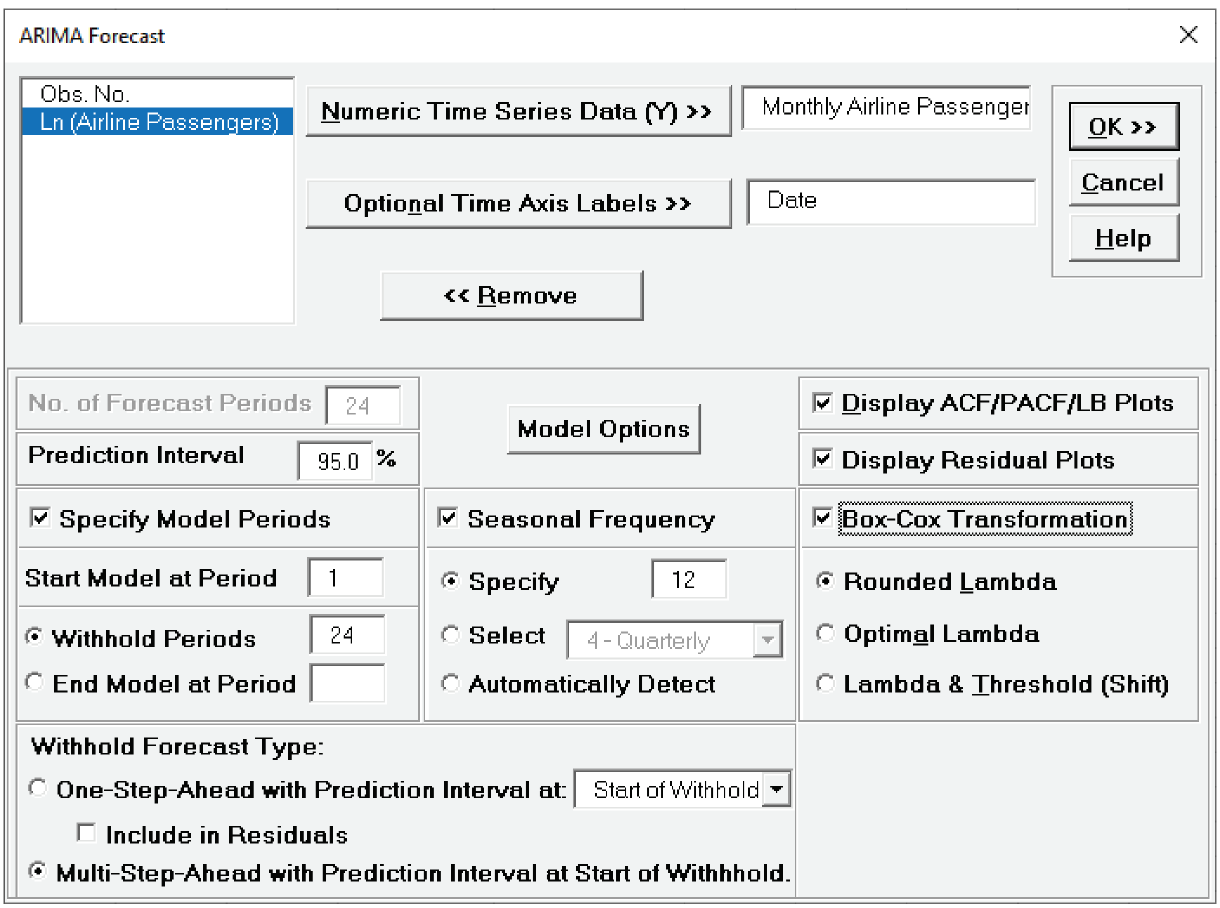
26. The above analysis can be rerun using **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** (click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog), but we will not do so here. The results would be very similar to those obtained with exponential smoothing.

Monthly Airline Passengers - Series G

27. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

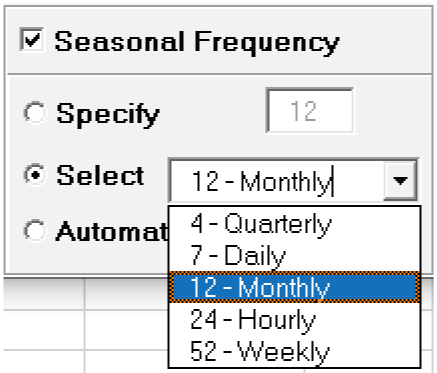
28. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

29. Select *Monthly Airline Passengers*, click **Numeric Time Series Data (Y) >>**. Select *Date*, click **Optional Time Axis Labels >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 24. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Check **Seasonal Frequency** with **Specify** = 12**.** Check **Box-Cox Transformation** and select **Rounded Lambda** (selected because the Run Chart showed an increase in the seasonal variance over time). We will use the default **Prediction Interval** = 95.0 %.



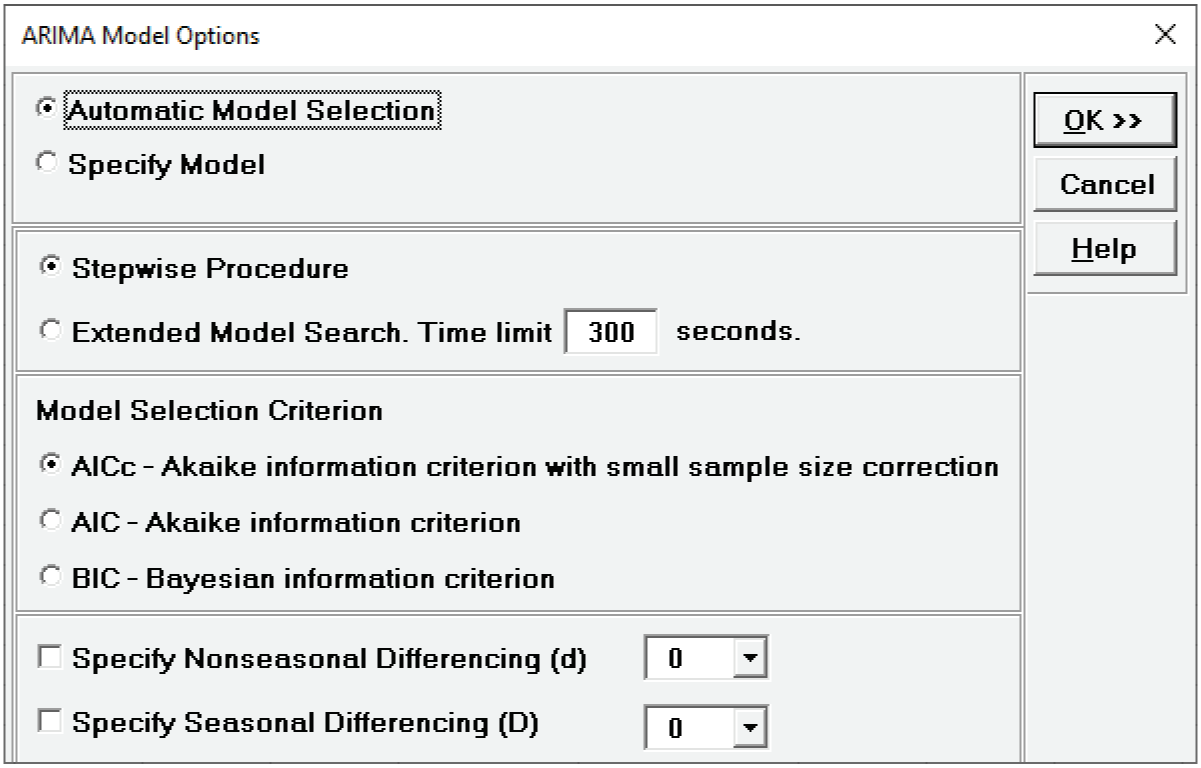
- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** will exclude the withhold sample from automatic model identification and parameter estimation and does not use the withhold data to update the predicted one-step ahead forecast. This is useful to assess forecast error when you are interested in a long-term forecast window (horizon). The prediction interval will be displayed for the duration of the withhold sample. Note that the length of the prediction interval is determined by the number of withhold periods, so overrides the specified **No. of Forecast Periods.**

- **Seasonal Frequency Specify** is used to specify the seasonal frequency.
- **Seasonal Frequency Select** gives a drop-down list of commonly used seasonal frequencies:

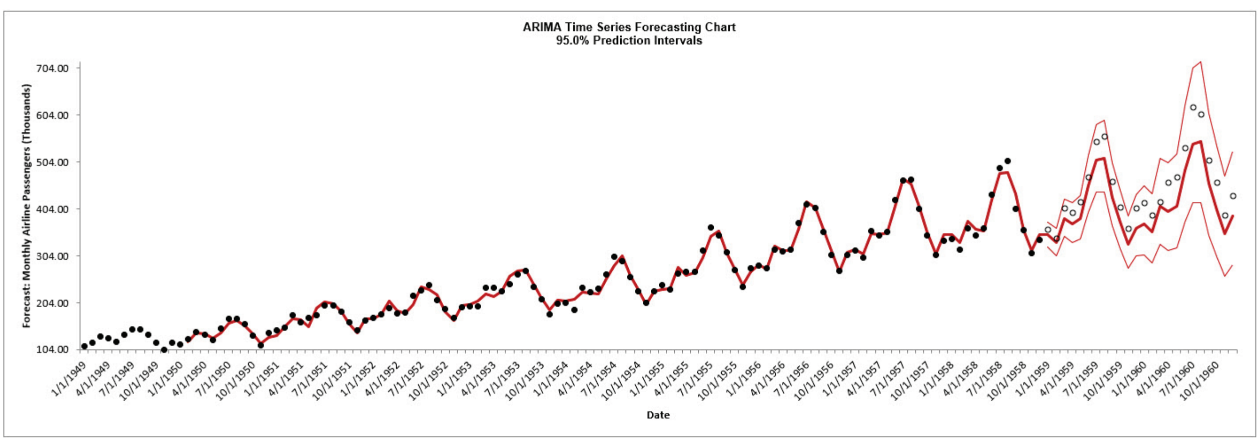


- **Seasonal Frequency Automatically Detect** should be used if uncertain what the seasonal frequency value is (or do a Spectral Density Plot prior to the Seasonal Trend Decomposition Plots).
- **Box-Cox Transformation** with **Rounded Lambda** will select Lambda = 0 (Ln), 0.5 (SQRT) or 1 (Untransformed). Threshold (Shift) is computed automatically if the time series data includes 0 or negative values, otherwise it is 0.
- **Box-Cox Transformation** with **Optimal Lambda** uses the range of 0 to 1 for Lambda. Threshold is computed automatically if the time series data includes 0 or negative values.
- **Box-Cox Transformation** with **Lambda & Threshold (Shift)** if left blank, will compute optimal lambda and threshold. The user may also specify Lambda and Threshold. Lambda may be specified outside of the 0 to 1 range, but practically for time series analysis, should be limited to -1 to 2. Threshold is typically 0, but if the time series data includes 0 or negative values, a negative threshold value should be entered that is smaller than the minimum data value. This value will be subtracted from the data resulting in positive time series values.

30. Click **Model Options**. Select **Automatic Model Selection**. We will use the defaults: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, leave **Specify Nonseasonal Differencing (d)** and **Specify Seasonal Differencing (D)** unchecked.



31. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. The ARIMA forecast report is given:



The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample. Note that the data has been transformed using Box-Cox, but the inverse transformation is applied to produce this chart with the original units.

This is similar to the exponential smoothing forecast chart that was produced using the **Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A)** model. The initial in-sample predicted value for ARIMA starts at the 14th time period due to nonseasonal and seasonal differencing.

32. Scroll down to view the ARIMA Model header:

ARIMA Model: Monthly Airline Passengers (Thousands) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

33. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

The ARIMA (0,1,1) (0,1,1) model (with no constant) was automatically selected based on the AICc criterion. This agrees with the results shown in the ACF/PACF Plots of the manually differenced data and is the model used in Box and Jenkins, Chapter 9., “Analysis of Seasonal Time Series”.

There are no predictors. Seasonal Frequency = 12; Model Selection Criterion = “AICc” and Box-Cox Transformation = “Rounded Lambda” with Lambda = 0 (i.e., a Ln transformation).

34. The Parameter Estimates are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.342313249	0.100902427	3.392517	**0.0010**
SMA_1	0.540469465	0.087677292	6.164304	**0.0000**

Unlike the previous example with nonseasonal and nontrended Concentration data, we cannot compare these parameters to the exponential smooth model.

35. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	120
DF	105
StDev	0.037414431
Variance	0.00139984
Log-Likelihood	197.5047754
AICc	-388.776541
AIC	-389.009551
BIC	-380.991064

- The number of observations, *n* = 144 – 24 (*withhold*) = 120
- Degrees of freedom (DF) = 120 (*n*) – 15 (2 terms in the model, 1 nonseasonal difference, 12 seasonal differences) = 105
- Note that the model statistics are based on the Ln transformed data, not the original data.

Comparing to the Exponential Smoothing Model Statistics, we see that the StDev and Variance are approximately equal, but the Log-Likelihood, AICc, AIC and BIC are very different. This is due to different formulas being used in the Likelihood function. You cannot use Information Criteria to compare ARIMA and Exponential Smooth models to determine which model has the best fit.

36. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step-Ahead Forecast**
N	107	24
RMSE	9.44337388	43.18833723
MAE	7.384036891	39.45185993
MAPE	3.001604734	8.51734062
MASE	0.258417364	1.380687256

Comparing to Exponential Smoothing Forecast Accuracy, we have higher Out-of-Sample forecast errors. MASE is greater than one, so it is a worse forecast than would be obtained from a seasonal naïve forecast. However, it is important to keep in mind that we are forecasting out for 24 months, so this is still quite reasonable. Selecting a 12 month Multi-Step Ahead Withhold Sample (or a One-Step-Ahead Withhold Sample) would dramatically improve the forecast.

As expected, the **Out-of Sample (Withhold) Multi-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors. Note, if we were primarily interested in a short term one-step ahead forecast, then we would have selected **Withhold Forecast Type: One-Step-Ahead** and this table would show **Out-of Sample (Withhold) One-Step-Ahead Forecast** errors.

Note that we lose 13 observations on the In-Sample (N=107) since we do not have predicted values at time period = 1 to 13, due to differencing.

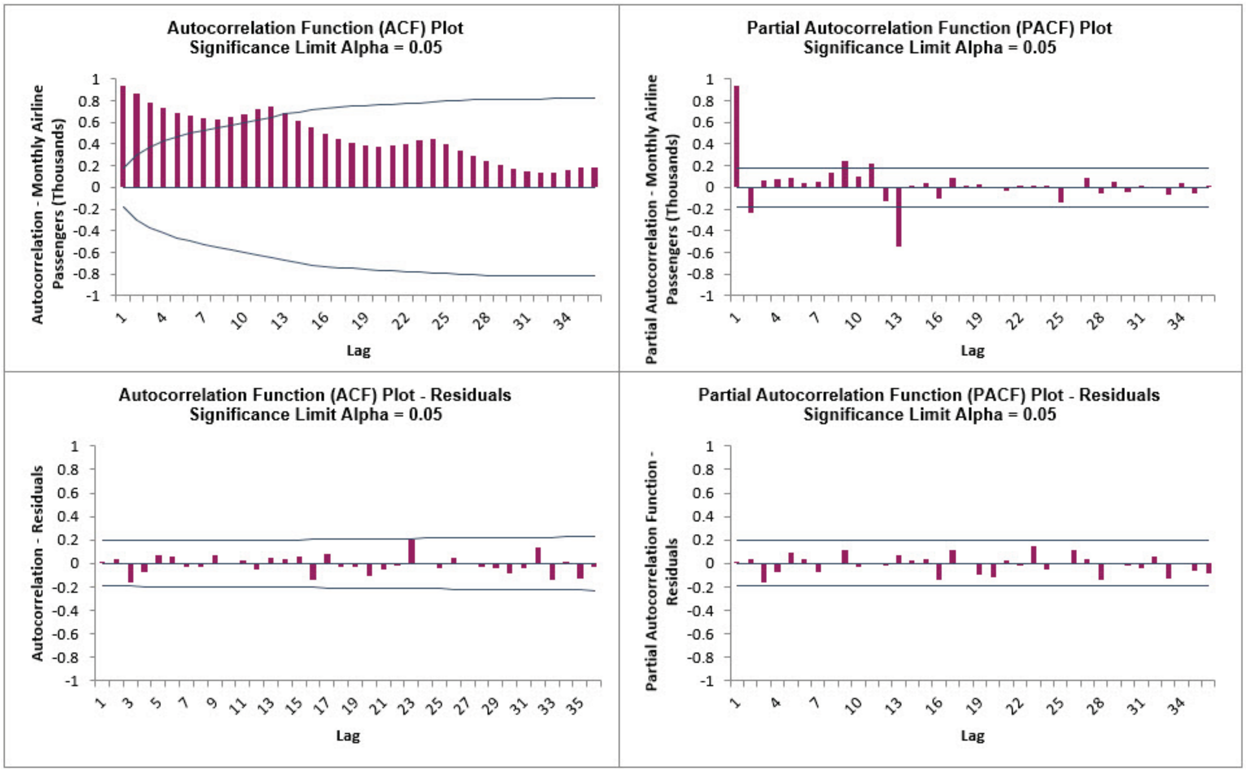
Forecast Accuracy metrics are calculated using the actual raw data versus inverse transformed forecast as displayed in the Forecast Chart and Table, so allow comparison across all model types and transformations.

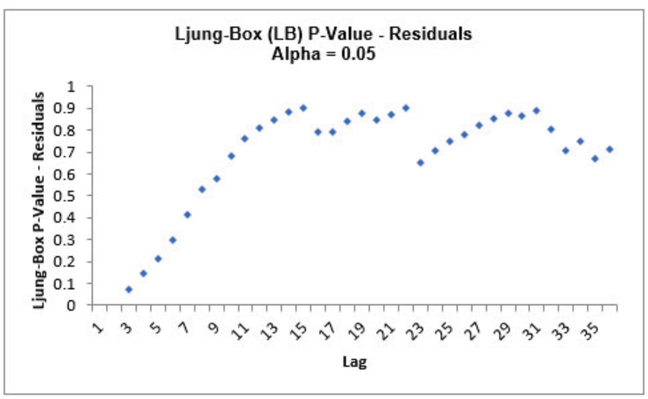
37. The Forecast Table is given as:

Forecast Table				
Period	Withhold Data	Multi-Step Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
121	360	348.5828062	323.6916935	375.3879856
122	342	331.5740082	303.4388404	362.3178982
123	406	383.1608225	346.2907533	423.9565003
124	396	372.606915	333.0281012	416.8894835
125	420	382.9486611	338.8267505	432.8161127
126	472	453.1350644	397.199074	516.9483013
127	548	507.198003	440.7280155	583.6929018
128	559	511.2341986	440.6029023	593.1881166
129	463	429.3989001	367.2058264	502.1255171
130	407	376.442152	319.5422535	443.4740391
131	362	328.1517697	276.5830729	389.3354095
132	405	362.9392672	303.8283644	433.5504089
133	417	372.7865814	306.4056918	453.5484783
134	391	354.596781	287.8353252	436.8431047
135	419	409.7655151	328.7189301	510.7943655
136	461	398.4787992	316.1075385	502.3143522
137	472	409.5386223	321.4319641	521.7959068
138	535	484.5984041	376.4753418	623.7742214
139	622	542.4151917	417.2724015	705.0891434
140	606	546.7316396	416.6289379	717.4621313
141	508	459.2141241	346.749965	608.1546736
142	461	402.5803351	301.3041689	537.8980544
143	390	350.9369201	260.4037852	472.9452062
144	432	388.1398803	285.6122259	527.4724013

These are the same forecast and prediction interval values displayed in the Forecast Chart but provided for further analysis or charting. Note that this table gives the Forecast Period number whereas the Chart displays the Date.

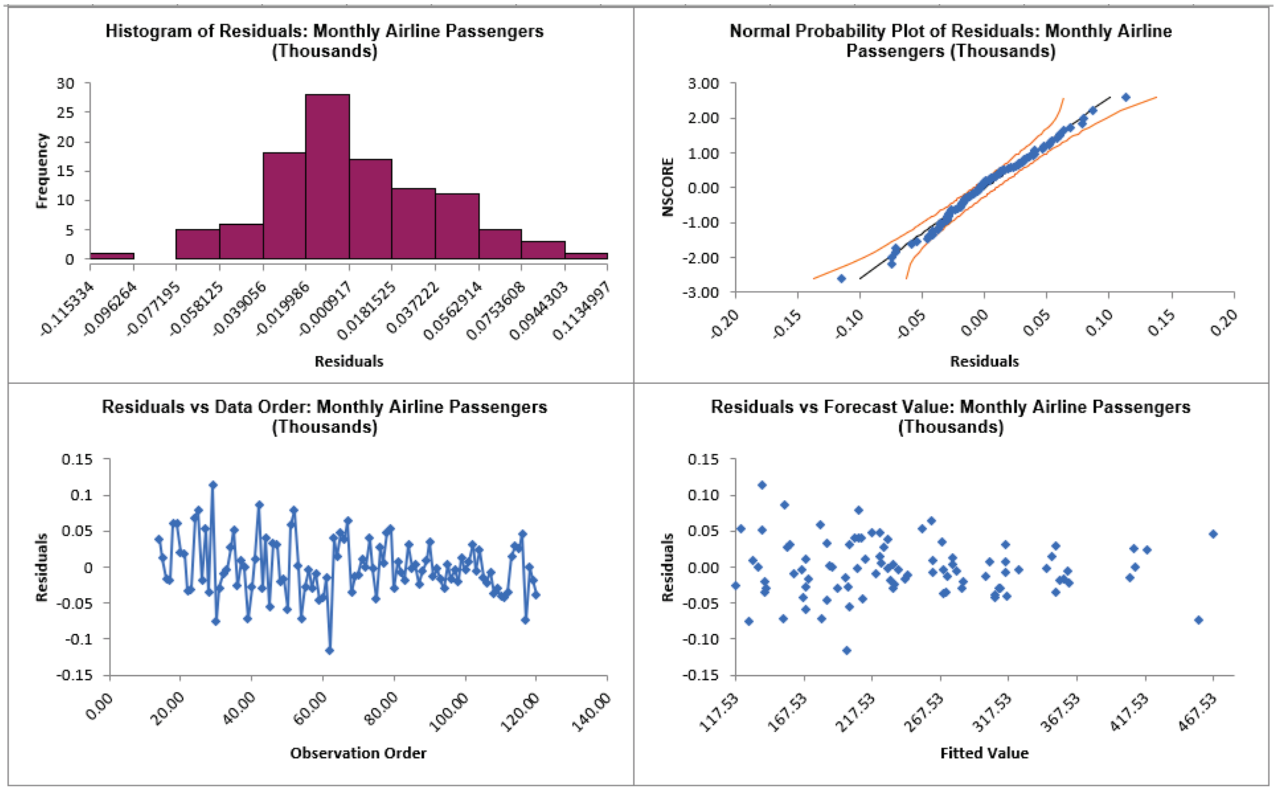
38. Click on the **ARIMA ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ACF/PACF Residuals Plots indicate that the autocorrelation has been accounted for in the model. The Ljung-Box plot confirms that no significant autocorrelation remains (all blue P-Values > .05), so this is a good fit to the time series data (as compared to the Exponential Smoothing LB Plot).

39. Click on the **ARIMA Residuals** sheet to view the Residual Plots:



The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts.

Note that the residuals are based on the Ln transformed data, not the original data. The **ARIMA Model Information** to the right of the plots shows the Box-Cox Transformation information.

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

ARIMA Forecast with Predictors

The ARIMA model supports continuous or categorical predictors, similar to multiple regression in SigmaXL. For further details see Appendix: ARIMA with Covariates (Predictors).

In order to provide a forecast, additional predictor (X) values must be added to the dataset prior to running the analysis. The number of forecast periods will be equal to the number of additional predictor rows. Alternatively, the predictor values from a withhold sample may be used. If neither are provided, SigmaXL will use the last row of predictor values and compute one forecast period.

Continuous predictors are numeric, categorical predictors can be text or numeric. All predictors must have the same number of rows. A predictor cannot have all values the same.

As with multiple linear regression, predictors should not be strongly correlated. SigmaXL will automatically remove terms with very high variance inflation factors (VIF > 100) and give a warning message “Multicollinearity detected in predictors. The following predictors were removed…”.

Also as done in multiple linear regression, categorical predictors are automatically “dummy coded”. Predictors will have a level append to their term names in the Parameter Estimates table. The first level (sorted alphanumerically) is a hidden reference level. If there are three levels, only two will appear in the Parameter Estimates table.

Missing values, while permitted in the Time Series Data (Y), are not permitted in the predictors. A warning message will be given, “Missing values detected in predictors. The following predictors were removed…”. If all predictors have missing values an error message is returned. An exception is made for missing values in the first rows to accommodate indicators with lags.

Sometimes the impact of a predictor will not be simple and immediate. For example, an advertising campaign may impact sales for some time beyond the end of the campaign, and sales in one month will depend on the advertising expenditure in each of the past few months. In these situations, we need to allow for lagged effects of the predictor (Hyndman, fpp2, 9.6 Lagged Predictors). A Pre-Whitened CCF Plot will show which lags need to be included in the model. SigmaXL can accommodate lagged predictors, but they must be manually created using **SigmaXL > Utilities > Lag Data** (or simply: copy, shift down one row, paste new column, delete extra row at bottom, add column header for the predictor data, and include as a predictor in the ARIMA model). Note, Box and Jenkins, Ch. 11, describe a more complex model method called a Transfer Function, but this is not provided in SigmaXL.

Daily Electricity Demand with Predictors – ElecDaily

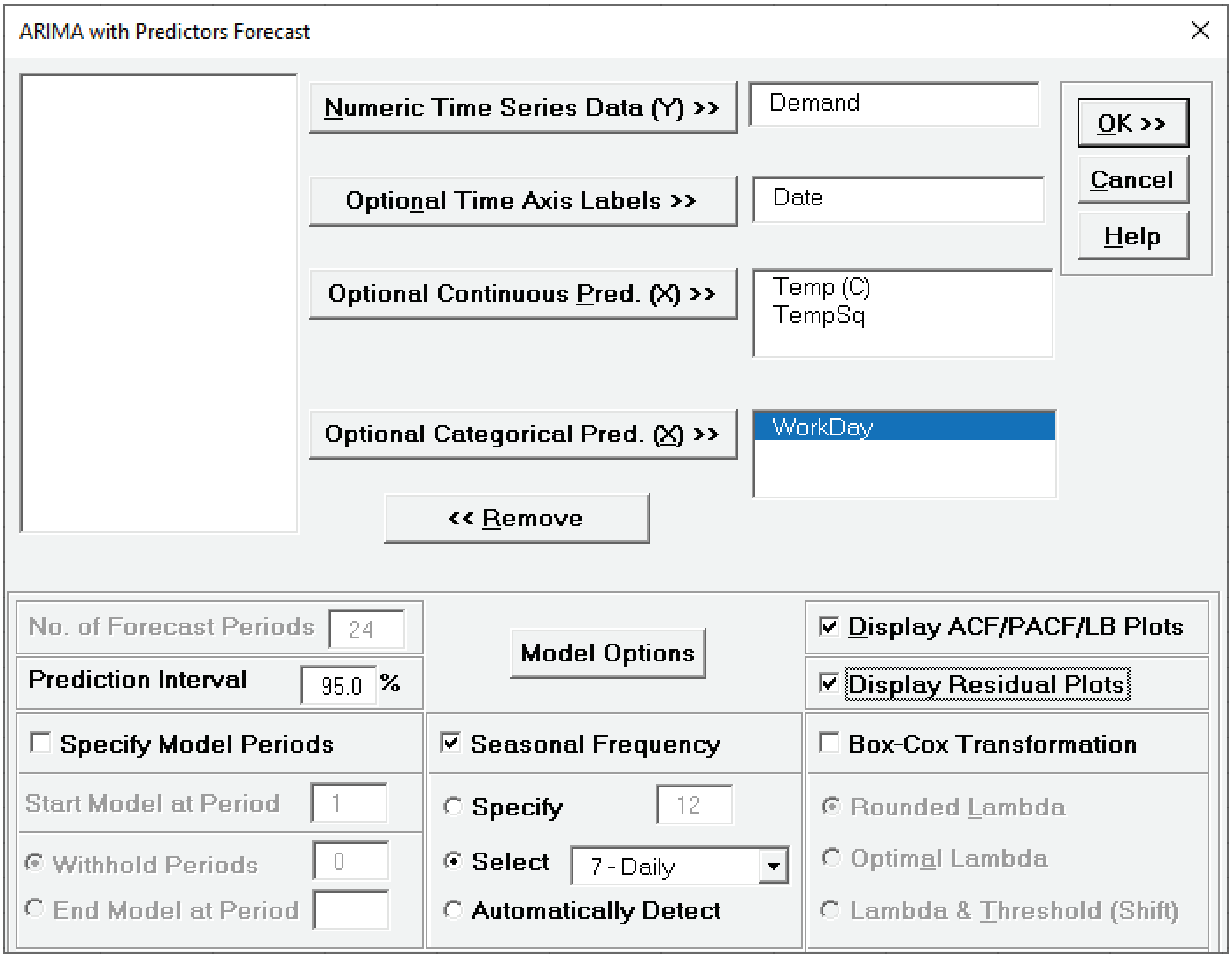
1. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx**. This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (see Hyndman, fpp2, Section 9.3, **Example: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html). Temp (C) is the maximum daily temperature in degrees Celsius for the city of Melbourne. TempSq is Temperature squared. WorkDay takes on the value 1 on work days and 0 otherwise. This data has a seasonal frequency = 7 (observations per week). See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

2. Click the **Forecast 2 Weeks** Sheet tab. Following the referenced example, we will use the ARIMA model to forecast 14 days ahead starting from January 1, 2015 (a non-work-day being a public holiday for New Year's Day). We could obtain weather forecasts for those 14 days, but for the sake of illustration, we will set the temperature for the 14 days to a constant 26 degrees and TempSq to 676. Scroll down to view the added data. Note that Date, Temp (C), TempSq and WorkDay are added for these 14 days, but the Demand values are blank:

1	Date	Demand	Temp (C)	TempSq	WorkDay
366	12/31/2014	186.370181	25.5	650.25	1
367	1/1/2015		26	676	0
368	1/2/2015		26	676	1
369	1/3/2015		26	676	0
370	1/4/2015		26	676	0
371	1/5/2015		26	676	1
372	1/6/2015		26	676	1
373	1/7/2015		26	676	1
374	1/8/2015		26	676	1
375	1/9/2015		26	676	1
376	1/10/2015		26	676	0
377	1/11/2015		26	676	0
378	1/12/2015		26	676	1
379	1/13/2015		26	676	1
380	1/14/2015		26	676	1

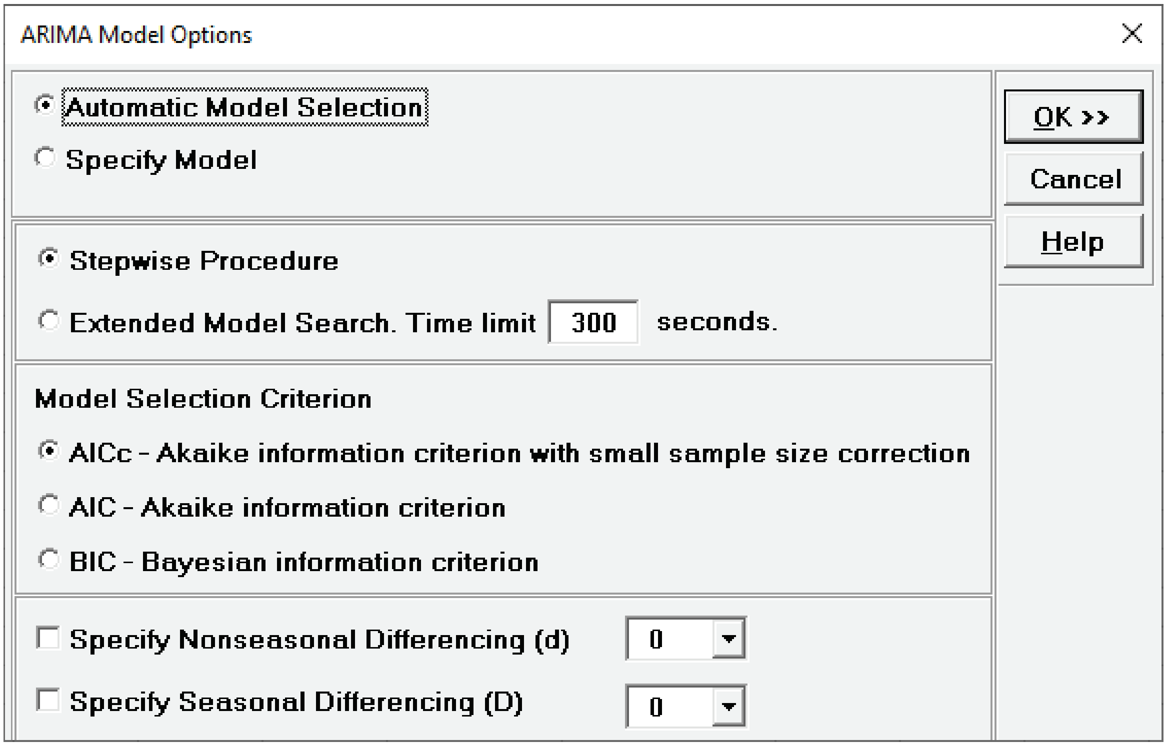
3. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

4. Select *Demand*, click **Numeric Time Series Data (Y) >>**; select *Date*, click **Optional X-Axis Labels >>**; select *Temp (C)* and *TempSq*, click **Optional Continuous Pred. >>**; select *WorkDay*, click **Optional Categorical Pred. >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Seasonal Frequency** with **Select** = 7 - Daily (or **Specify** = 7). Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked. We will use the default **Prediction Interval** = 95.0 %.

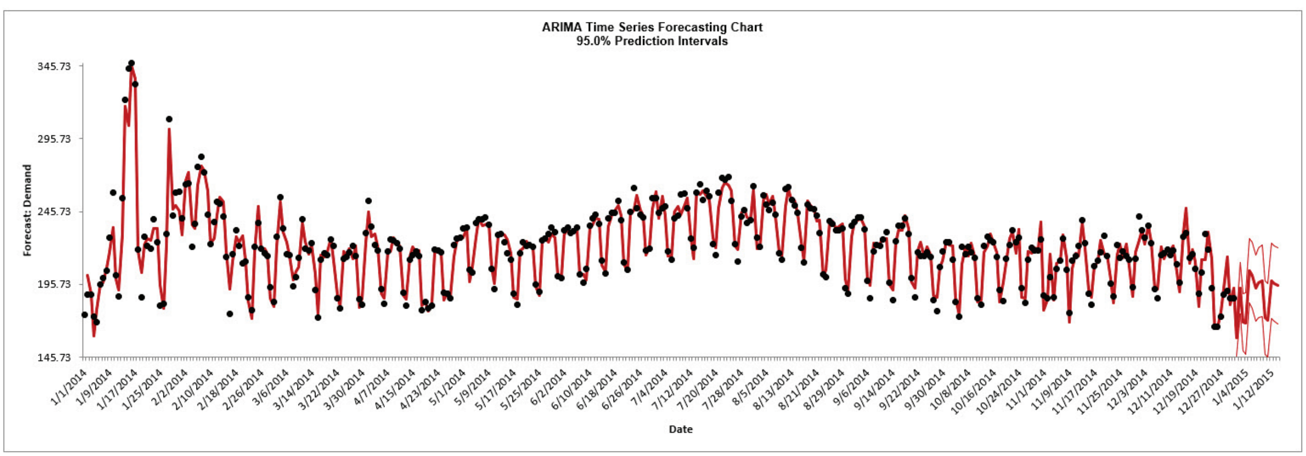


- **No. of Forecast Periods** is greyed out because they are determined by the number of additional predictor rows that are provided, in this example, 14.
- **Optional Continuous Pred. (X)** are continuous predictors.
- **Optional Categorical Pred. (X)** are categorical predictors. In this example WorkDay could be either continuous or categorical since it is coded as 0,1.

5. Click **Model Options**. Select **Automatic Model Selection**. We will use the defaults: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, leave **Specify Nonseasonal Differencing (d)** and **Specify Seasonal Differencing (D)** unchecked.



6. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. This is a complex model, so computation time will be approximately one to two minutes. The ARIMA forecast report is given:



This agrees with Figure 9.8 given in Hyndman, fpp2, Section 9.3, **Example: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html. There are slight differences in the initial predicted values.

7. Scroll down to view the ARIMA Model header:

ARIMA Model: Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

If we had checked Specify Model Periods in the main dialog, the start, end or withhold selection would be summarized here as well.

8. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	2
SAR Order (P)	2
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	7
Include Constant	0
No. of Predictors	3
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information: ARIMA (2,1,2) (2,0,0) with no constant and 3 predictors. Seasonal Frequency = 7; Model Selection Criterion = "AICc" and Box-Cox Transformation = "N/A" because Box-Cox Transformation was unchecked.

This agrees with the model given in Hyndman, fpp2, Section 9.3, **Example: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html.

9. The Parameter Estimates are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
AR_1	-0.063223451	0.075658448	0.83564	0.4039
AR_2	0.673128346	0.067270503	10.0063	**0.0000**
MA_1	0.022660844	0.043288704	0.52348	0.6010
MA_2	0.929862871	0.039474102	23.5563	**0.0000**
SAR_1	0.200902989	0.053912363	3.72647	**0.0002**
SAR_2	0.402632085	0.05676416	7.09307	**0.0000**
Temp (C)	-7.501559029	0.446098708	16.8159	**0.0000**
TempSq	0.17890261	0.008530253	20.9727	**0.0000**
WorkDay_1	30.56943168	1.295720007	23.5926	**0.0000**

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. All of the predictors are significant. The AR_1 and MA_1 terms are not significant but they must remain in the model due to hierarchy. Note that for AR/MA model order selection, minimum AICc should be used, rather than significance tests (see Kostenko, A.V. and Hyndman, R.J.).

The categorical predictor *WorkDay_1* has 0 as the hidden reference level. If there were three category levels, only two would appear in the table.

Note: SE Coefficients, T and P-Values may be slightly different due to machine precision.

10. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	365
DF	355
StDev	6.565744215
Variance	43.1089971
Log-Likelihood	-1200.73256
AICc	2422.088353
AIC	2421.465123
BIC	2460.436662

Degrees of freedom (DF) = *n* – 10 (9 terms in the model, 1 order of differencing). See ARIMA Model Statistics and Information Criteria for Model Comparison.

11. The In-Sample Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	364	
RMSE	6.615605184	
MAE	4.768089714	
MAPE	2.162224989	
MASE	0.327382095	

MASE is less than one, so it is a better forecast than would be obtained from a naïve forecast (set all forecasts to be the value of the last observation). See Forecast Accuracy Metrics.

The analysis can be rerun (using Recall SigmaXL Dialog) with a withhold sample to obtain Out-of-Sample One-Step-Ahead or Multi-Step-Ahead Forecast errors, but we will not do so here.

12. The Forecast Table is given as:

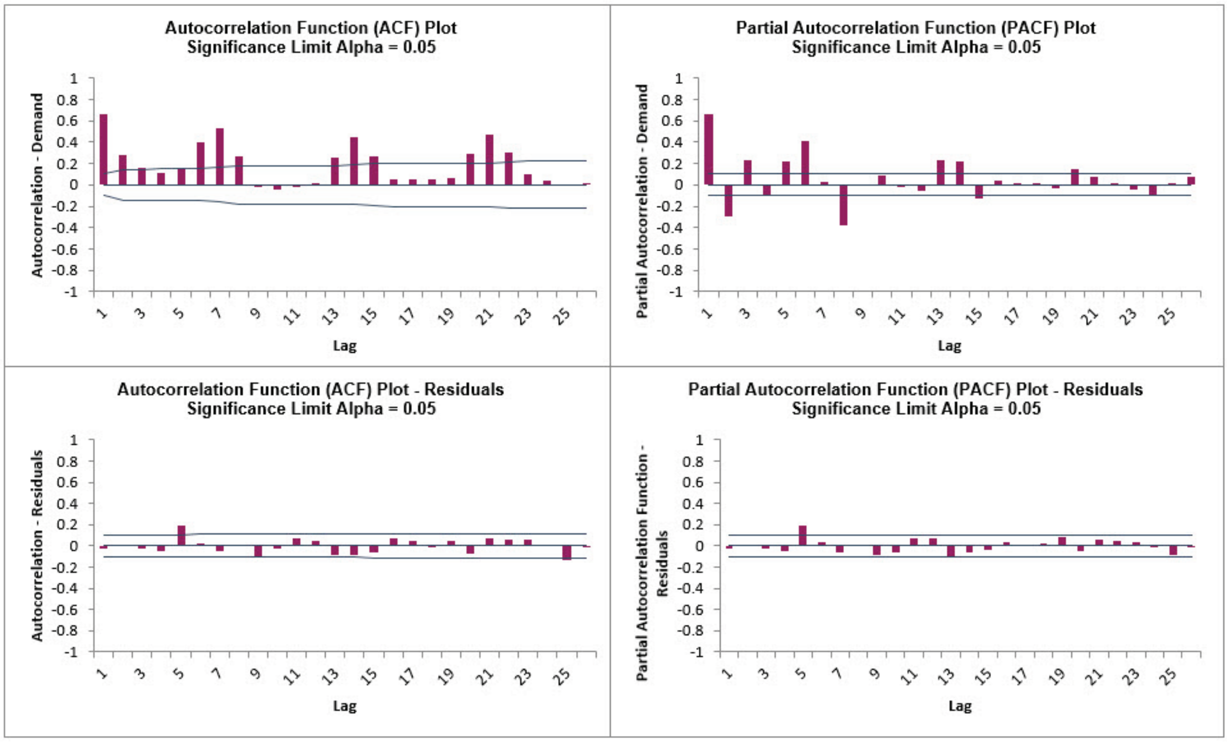
Forecast Table				
Period	Withhold Data	Forecast	Lower 95.0% PI	Upper 95.0% PI
366		159.2517864	146.3402335	172.1633392
367		193.2139395	175.7207745	210.7071045
368		169.8531097	150.3787979	189.3274216
369		168.9969534	147.9371527	190.0567541
370		205.1573872	183.2956912	227.0190833
371		201.9616978	179.3846259	224.5387697
372		193.1293236	170.1638907	216.0947565
373		197.1190445	173.1860174	221.0520716
374		198.0338842	173.4650501	222.6027184
375		173.0122062	147.9556892	198.0687231
376		171.1330507	145.7330385	196.5330629
377		198.1491778	172.4687919	223.8295637
378		196.3934701	170.5020251	222.2849151
379		194.6921179	168.6187184	220.7655175

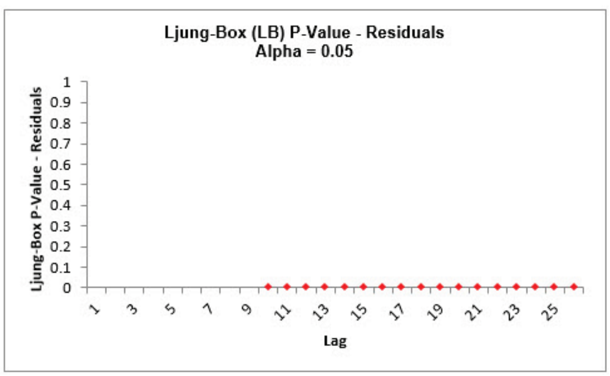
These are the same forecast and prediction interval values displayed in the Forecast Chart but provided for further analysis or charting. If Withhold Periods are specified, the Withhold Data will be displayed as well.

13. The Predictor Values for Forecast are the additional predictor (X) values used to obtain the 14 period forecast in the Forecast Table:

Predictor Values for Forecast		
Temp (C)	TempSq	WorkDay_1
26	676	0
26	676	1
26	676	0
26	676	0
26	676	1
26	676	1
26	676	1
26	676	1
26	676	1
26	676	0
26	676	0
26	676	1
26	676	1
26	676	1

14. Click on the **ARIMA ACF PACF LB** sheet to view the ACF/PACF/LB Plots:

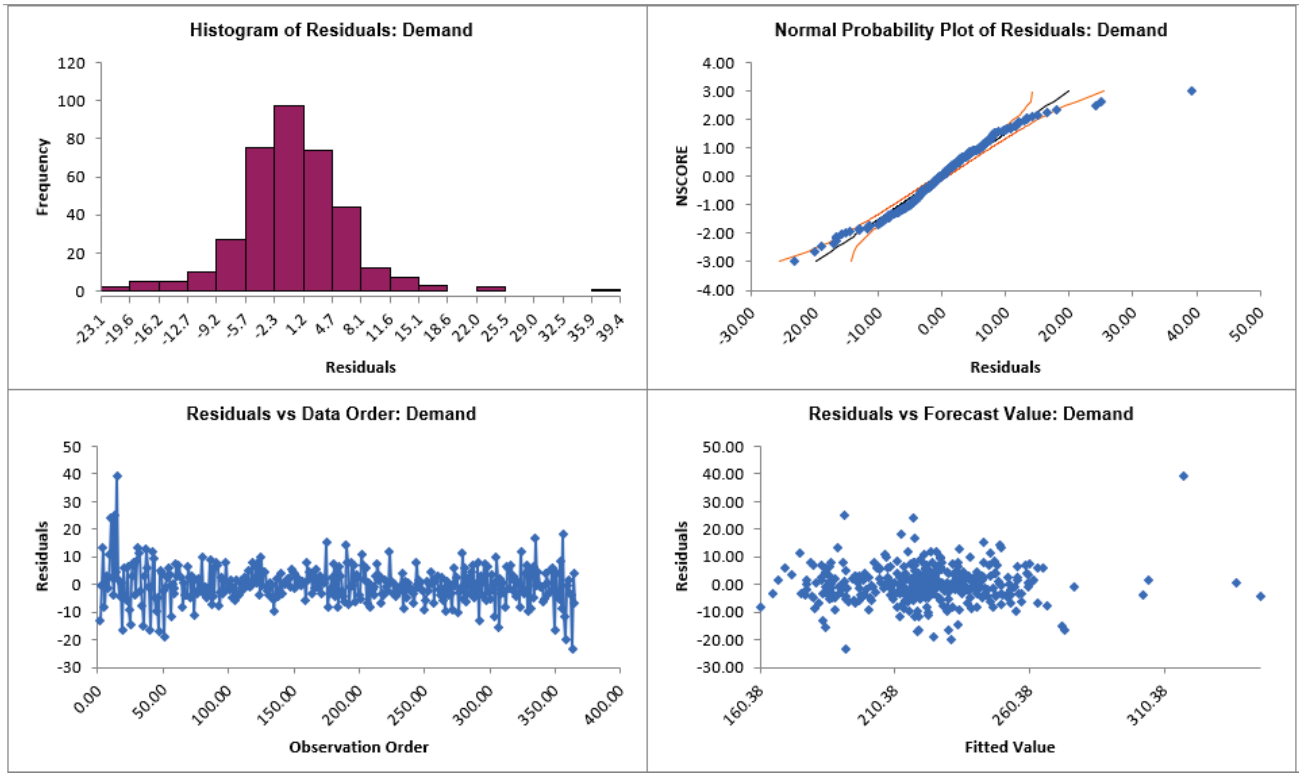


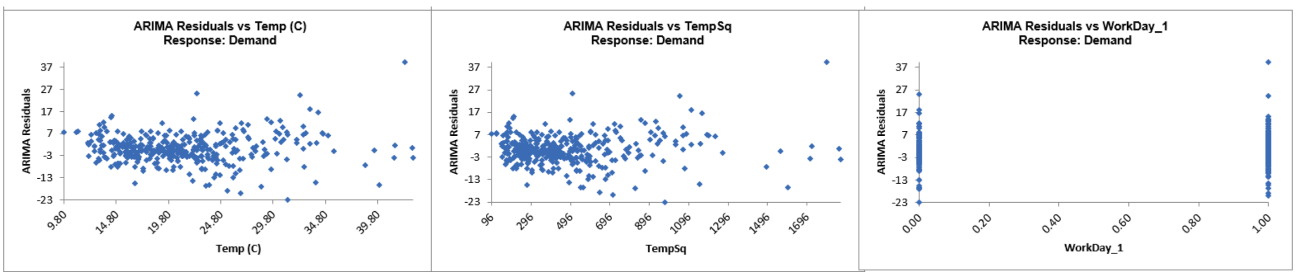


We can see that much of the autocorrelation has been removed by the ARIMA with Predictors model (with the exception of lag 5 and 25 in the ACF, lag 5 in the PACF).

The Ljung-Box plot shows that some significant autocorrelation still remains (the red P-Values are significant at alpha=.05) - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

15. Click on the **ARIMA Residuals** sheet to view the Residual Plots:





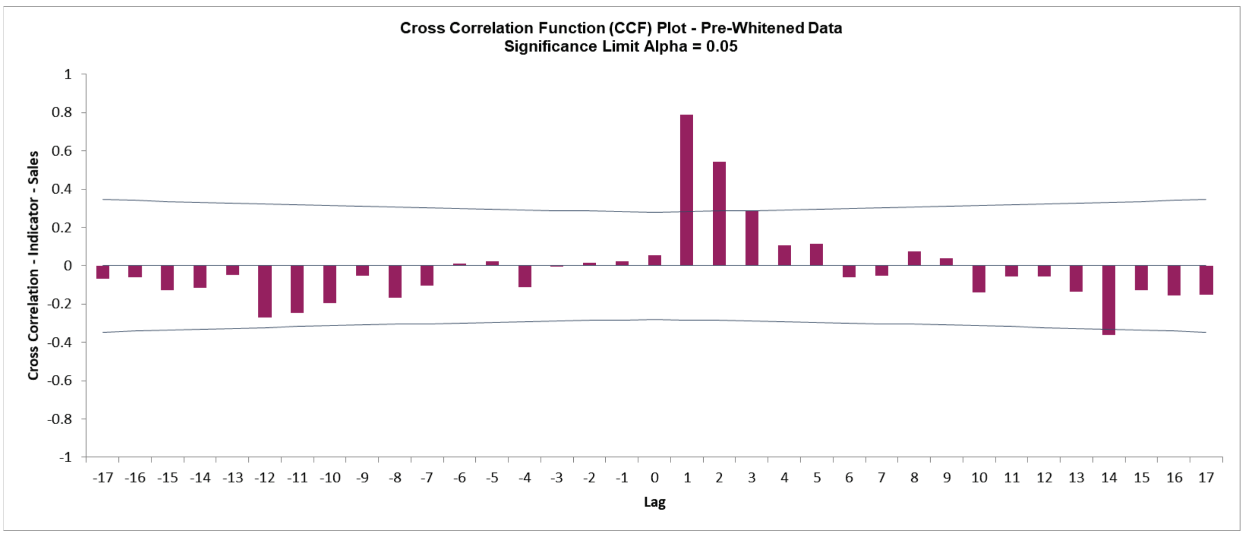
Looking at the histogram and normal probability plot, there are some outliers (but smaller than would have been the case if we did not include the predictors). Later, these will be investigated with a control chart on the residuals.

Note that additional residual plots are provided for each predictor.

Sales with Indicator - Modified Series M

16. Open **Sales with Indicator - Modified Series M.xlsx**. This is modified Series M data from Box and Jenkins. Originally, a set of 150 monthly corporate sales values along with a leading indicator, the data was modified by converting it to quarterly values by averaging every three months, so 50 quarters, labelled as Q1-Y1, Q2-Y1, etc. This was done in order to simplify the analysis of the leading indicator. Although the data was monthly and summarized to quarterly, it will be treated as nonseasonal, as done in Box and Jenkins. See the Run Chart, ACF/PACF Plots, CCF Plot, Spectral Density Plot and Trend Decomposition Plots for this data.

17. Recall that the Pre-Whitened CCF Plot for this data is:

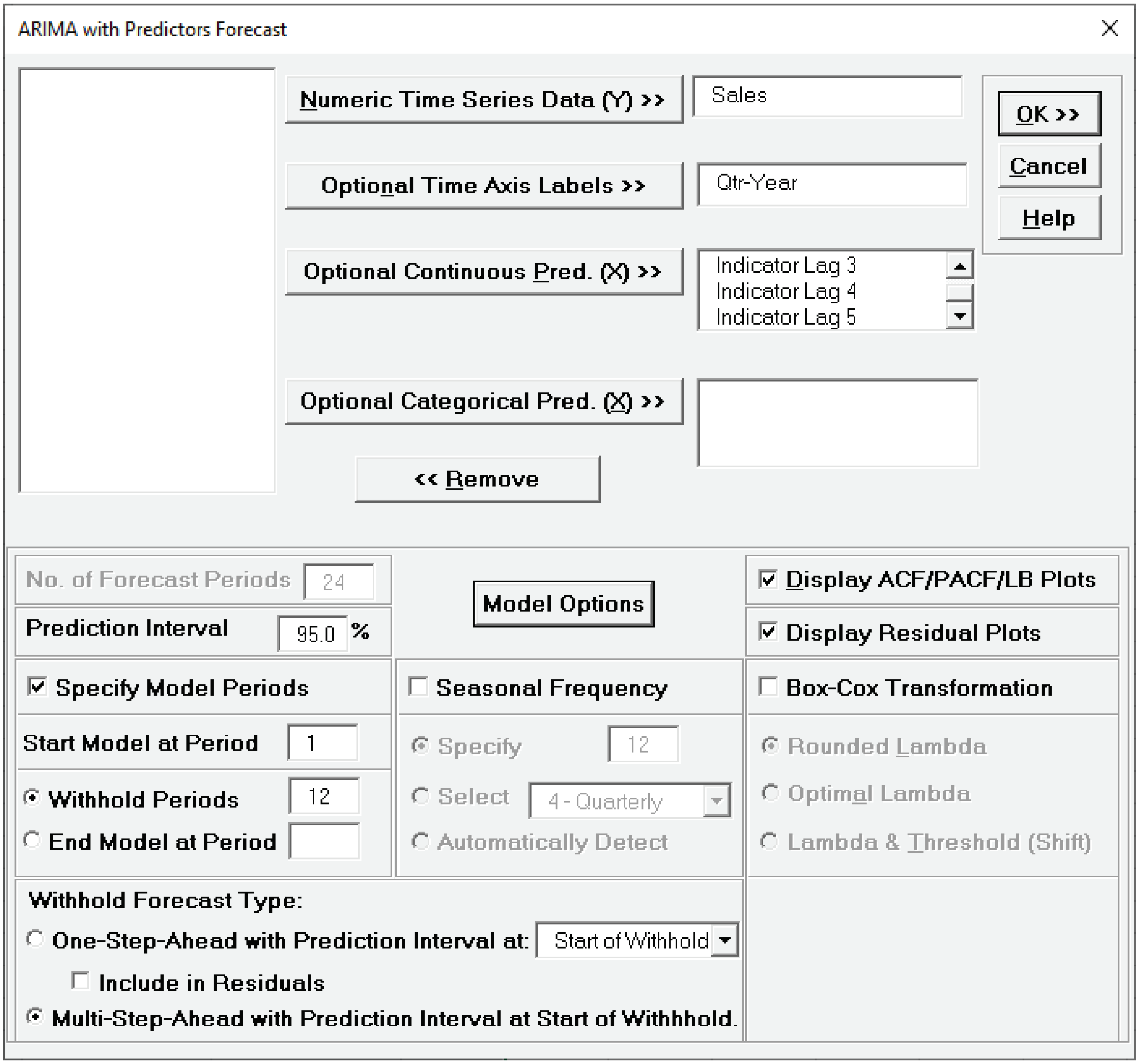


    Pre-Whitening the data has dramatically altered the CCF plot, allowing us to see the underlying cross correlation pattern. Lags 1 and 2 are significantly positive, and Lag 3 is just on the significance line. Use this as a guide to assist with what lags to include in the model but it is possible that some additional lags may be significant. In this example, we will initially include up to Lag 5.

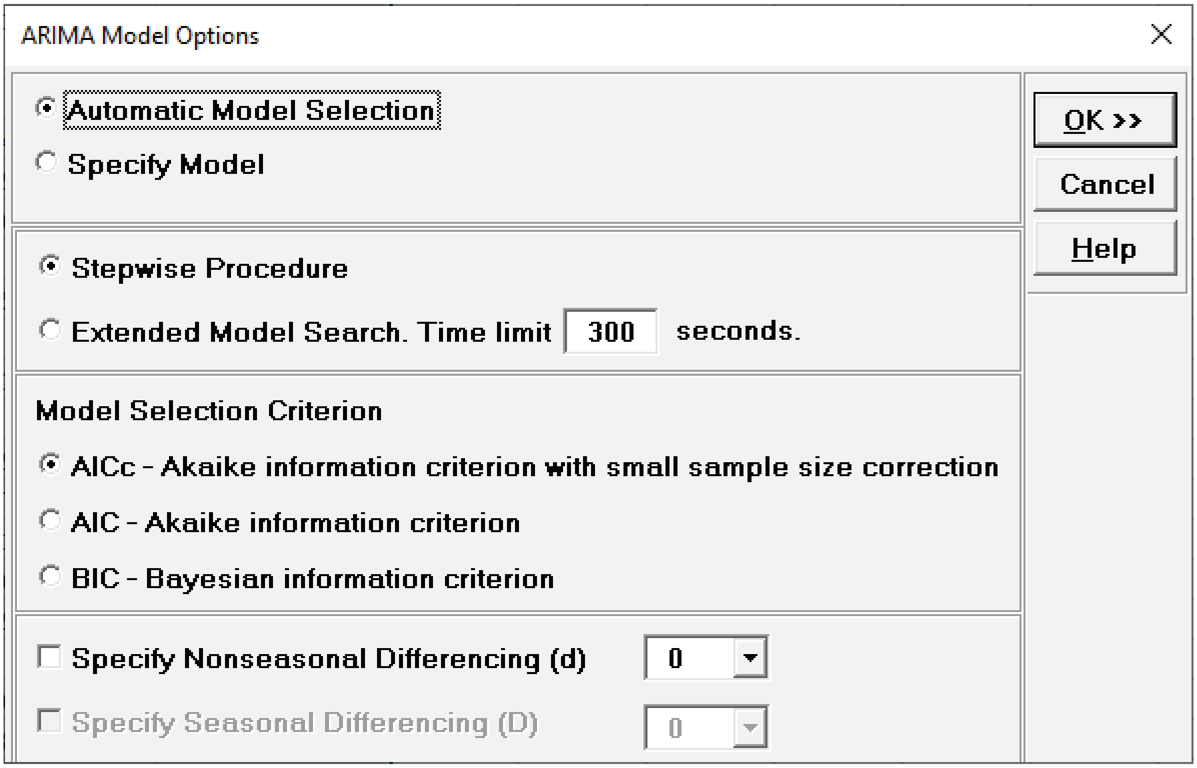
    Note that while X is called a leading indicator, i.e., X comes before Y in time, the positive lag means that the X variable is lagging the Y variable in terms of correlation structure. SigmaXL uses this convention as given in Box and Jenkins (2016, pp. 437-440).

18. Click the **Indicator with Lags** Sheet tab. **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

19. Select *Sales*, click **Numeric Time Series Data (Y) >>**; select *Qtr-Year*, click **Optional Time Axis Labels >>**; select *Indicator* to *Indicator Lag 5*, click **Optional Continuous Pred. >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 12 (i.e., 3 years). Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Leave **Specify Model Periods**, **Seasonal Frequency** and **Box-Cox Transformation** unchecked. We will use the default **Prediction Interval** = 95.0 %.

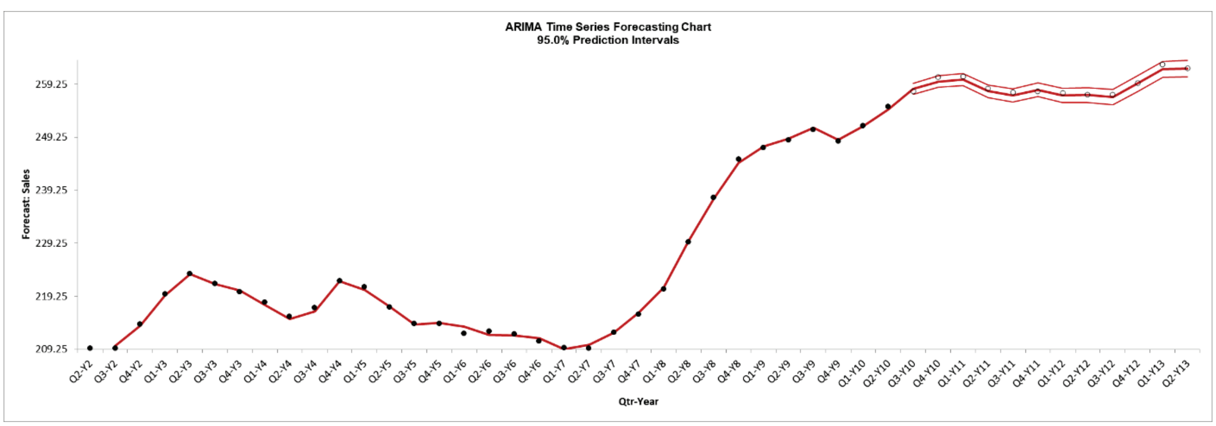


20. Click **Model Options**. Select **Automatic Model Selection**. We will use the defaults: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, leave **Specify Nonseasonal Differencing (d)** and **Specify Seasonal Differencing (D)** unchecked.



Tip: When using **Recall SigmaXL Dialog** and if there are no changes to the **Model Option** settings**,** the previous settings will be used. It is not necessary to repeat this step.

21. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. The ARIMA with Predictors forecast report is given:



The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample. The model uses the Indicator Lag values in the withhold sample and does quite well at predicting the 12 quarters of Sales values.

Note that the chart starts at Q2-Y2. This is due to the first 5 rows of X and Y data being deleted since there are missing values in Indicator Lags 1 to 5.

22. Scroll down to view the ARIMA Model header:

ARIMA Model: Sales - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 12 withhold periods. Missing values in initial rows of predictors. Deleted first 5 rows of X and Y data.

23. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	1
No. of Predictors	6
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information: ARIMA (0,1,1) with no constant and 6 predictors. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = "AICc" and Box-Cox Transformation = "N/A" because Box-Cox Transformation was unchecked.

24. The Parameter Estimates are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.661826535	0.171947916	3.84899	**0.0008**
Const:Trend	0.107385184	0.033507981	3.20476	**0.0038**
Indicator	-0.193804762	0.279261631	0.69399	0.4944
Indicator Lag 1	8.199265436	0.304329926	26.942	**0.0000**
Indicator Lag 2	5.392904003	0.2952757	18.264	**0.0000**
Indicator Lag 3	2.481788513	0.301027513	8.24439	**0.0000**
Indicator Lag 4	0.885376557	0.324714746	2.72663	**0.0118**
Indicator Lag 5	0.154053202	0.264999213	0.58133	0.5664

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. The CCF Plot suggested that only Lags 1 and 2, possibly 3, would be significant, but here we see that Lag 4 is also significant (hence why CCF is just an approximate guide). Later we will rerun the model and remove Indicator & Indicator Lag 5 and manually compare AICc values. Note: SE Coefficients, T and P-Values may be slightly different due to machine precision.

25. The ARIMA with Predictors Model Statistics are:

ARIMA Model Statistics	
No. of Observations	33
DF	24
StDev	0.505833251
Variance	0.255867277
Log-Likelihood	-19.8991901
AICc	65.98019828
AIC	57.7983801
BIC	70.99000323

- The number of observations, *n* = 50 – 12 (*withhold*) – 5 (*deleted rows*) = 33
- Degrees of freedom (DF) = 33 (*n*) – 9 (8 terms in the model, 1 nonseasonal difference) = 24

26. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Multi-Step-Ahead Forecast**
N	32	12
RMSE	0.441579854	0.519299269
MAE	0.355486014	0.433503476
MAPE	0.160638285	0.167034359
MASE	0.123200929	0.150239472

The Out-of-Sample forecast errors are only slightly larger than the In-Sample, so this is a good prediction.

MASE is less than one, so it is a better forecast than would be obtained from a naïve forecast (set all forecasts to be the value of the last observation). See Forecast Accuracy Metrics.

27. The Forecast Table is given as:

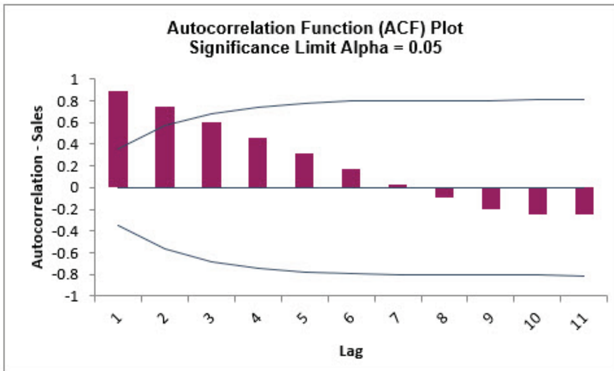
Forecast Table				
Period	**Withhold Data**	**Multi-Step Ahead Forecast**	**Lower 95.0% PI**	**Upper 95.0% PI**
34	257.9333333	258.3217379	257.2913893	259.3520865
35	260.5666667	259.7168718	258.6292016	260.8045419
36	260.7333333	260.1163414	258.974223	261.2584597
37	258.4333333	257.9037045	256.7096181	259.097791
38	257.6666667	257.0834025	255.8395172	258.3272877
39	257.9333333	258.1603883	256.8686227	259.4521539
40	257.5666667	257.1341925	255.7962588	258.4721261
41	257.2666667	257.1889296	255.8063687	258.5714904
42	257.3	256.7947588	255.3689669	258.2205507
43	259.4	259.3538829	257.8861328	260.8216331
44	263	262.0730127	260.5644708	263.5815546
45	262.2333333	262.2149879	260.6667286	263.7632471

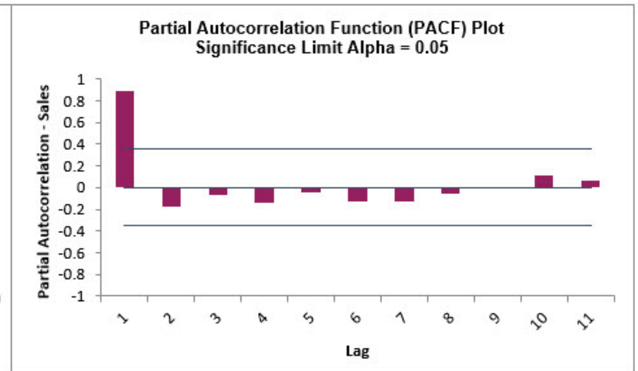
These are the same forecast and prediction interval values displayed in the Forecast Chart but provided for further analysis or charting. Note that this table gives the Forecast Period number whereas the Chart displays the Quarter-Year.

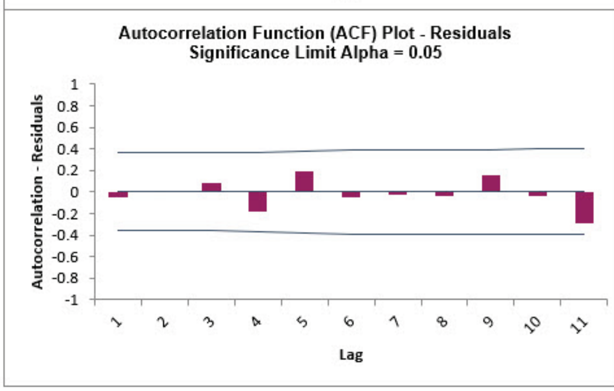
28. The Predictor Values for Forecast are the additional predictor (X) values used to obtain the 12 period forecast in the Forecast Table:

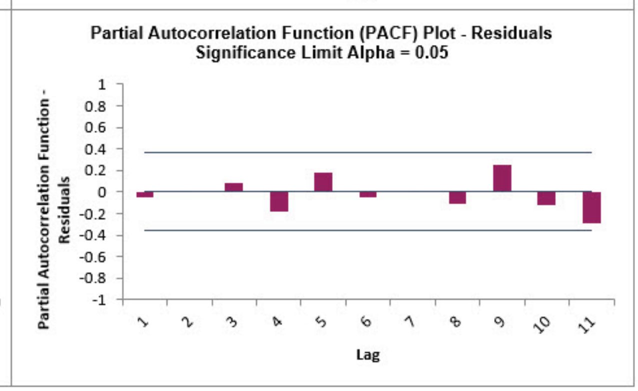
Predictor Values for Forecast					
Indicator	**Indicator Lag 1**	**Indicator Lag 2**	**Indicator Lag 3**	**Indicator Lag 4**	**Indicator Lag 5**
13.4166667	13.5	13.27	13.13333333	12.62666667	13.01
13.4066667	13.41666667	13.5	13.27	13.13333333	12.62666667
13.13	13.40666667	13.41666667	13.5	13.27	13.13333333
13.21	13.13	13.40666667	13.41666667	13.5	13.27
13.3566667	13.21	13.13	13.40666667	13.41666667	13.5
13.13	13.35666667	13.21	13.13	13.40666667	13.41666667
13.2233333	13.13	13.35666667	13.21	13.13	13.40666667
13.16	13.22333333	13.13	13.35666667	13.21	13.13
13.4966667	13.16	13.22333333	13.13	13.35666667	13.21
13.6033333	13.49666667	13.16	13.22333333	13.13	13.35666667
13.4433333	13.60333333	13.49666667	13.16	13.22333333	13.13
13.56	13.44333333	13.60333333	13.49666667	13.16	13.22333333

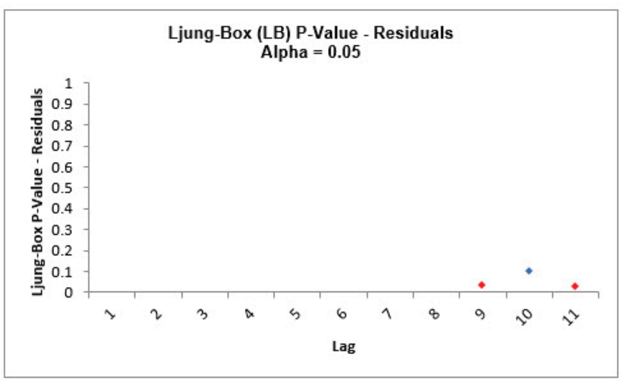
29. Click on the **ARIMA ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





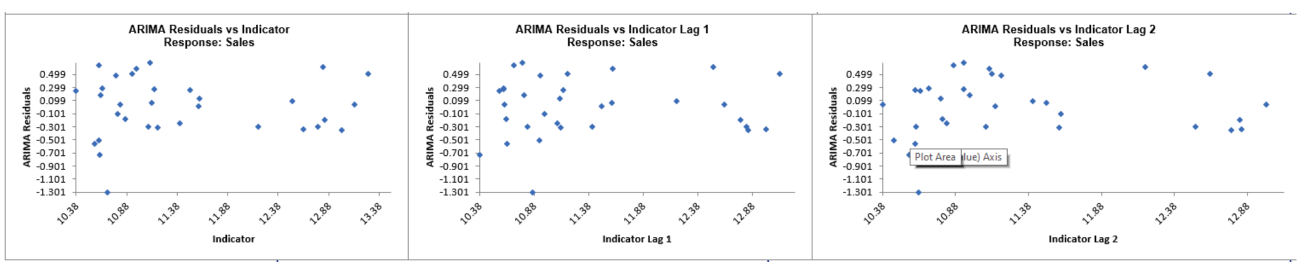






We can see that much of the autocorrelation has been removed by the ARIMA with Predictors model. The Ljung-Box plot shows that some significant autocorrelation still remains (the red P-Values are significant at alpha = .05) - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

30. Click on the **ARIMA Residuals** sheet to view the Residual Plots:

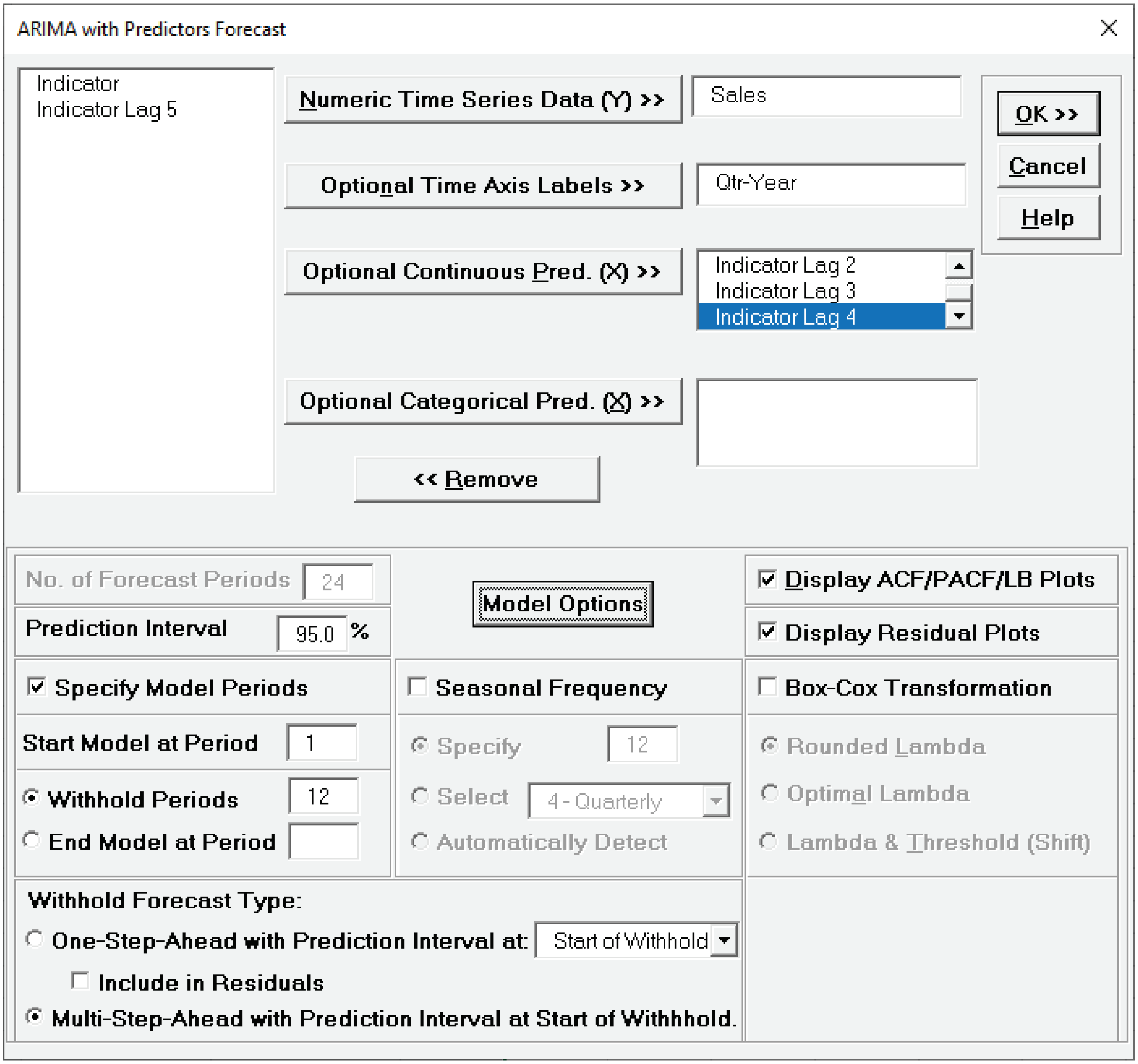


The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts. The histogram might suggest a left skew but the sample size is quite small. (Normality tests may also be applied on the residuals using SigmaXL > Statistical Tools > Descriptive Statistics: Options, Additional Normality Tests, and these would all show the residuals to be normal).

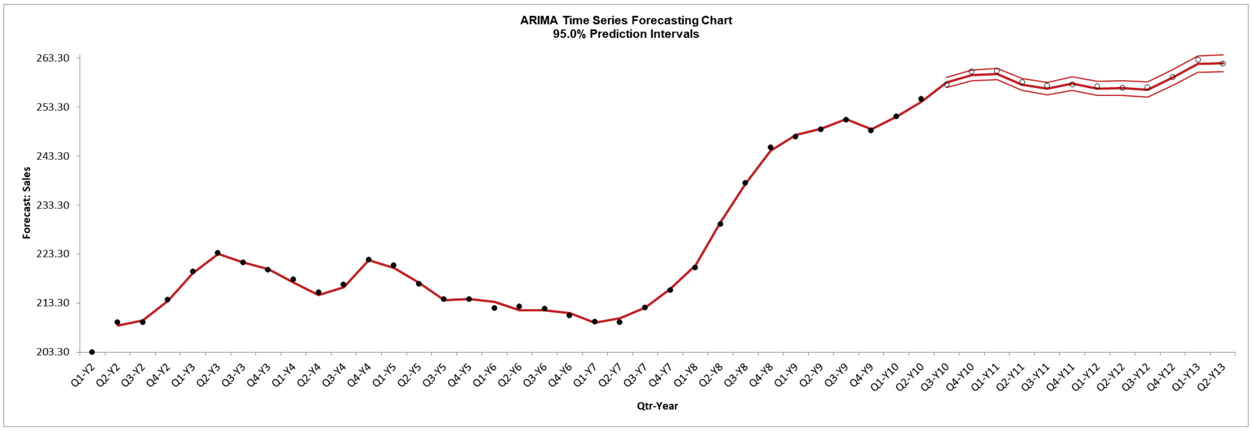
Note that additional residual plots are provided for each predictor.

31. Now we will rerun the analysis and remove the insignificant terms *Indicator* and *Indicator Lag 5*. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.

32. Select *Indicator,* click Remove; select *Indicator Lag 5*, click Remove:



33. We will not make any changes to the **Model Options**. Click **OK**. The revised ARIMA with Predictors forecast report is given:



Note that the chart now starts at Q1-Y2. This is due to the first 4 rows of X and Y data being deleted since there are missing values in Indicator Lags 1 to 4.

34. Scroll down to view the ARIMA Model header:

ARIMA Model: Sales - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 12 withhold periods. Missing values in initial rows of predictors. Deleted first 4 rows of X and Y data.

35. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	1
No. of Predictors	4
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information: ARIMA (0,1,1) without a constant has not changed. Now there are 4 predictors. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “AICc” and Box-Cox Transformation = “N/A” because Box-Cox Transformation was unchecked.

36. The Parameter Estimates are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.579458707	0.137596707	4.21128	**0.0003**
Const:Trend	0.110416608	0.039494253	2.79576	**0.0094**
Indicator Lag 1	8.14204032	0.260590425	31.2446	**0.0000**
Indicator Lag 2	5.381655285	0.288605873	18.6471	**0.0000**
Indicator Lag 3	2.521912726	0.293140527	8.60308	**0.0000**
Indicator Lag 4	0.932902504	0.262059255	3.55989	**0.0014**

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. The CCF Plot suggested that only Lags 1 and 2, possibly 3, would be significant, but here we see that Lag 4 is also significant (hence why CCF is just an approximate guide).

37. The ARIMA with Predictors Model Statistics are:

ARIMA Model Statistics	
No. of Observations	34
DF	27
StDev	0.491053186
Variance	0.241133231
Log-Likelihood	-20.6849129
AICc	59.84982585
AIC	55.36982585
BIC	65.84537878

- The number of observations, *n* = 50 – 12 (*withhold*) – 4 (*deleted rows*) = 34
- Degrees of freedom (DF) = 34 (*n*) – 7 (6 terms in the model, 1 nonseasonal difference) = 27
- AICc is lower than the previous model which included insignificant *Indicator* terms.

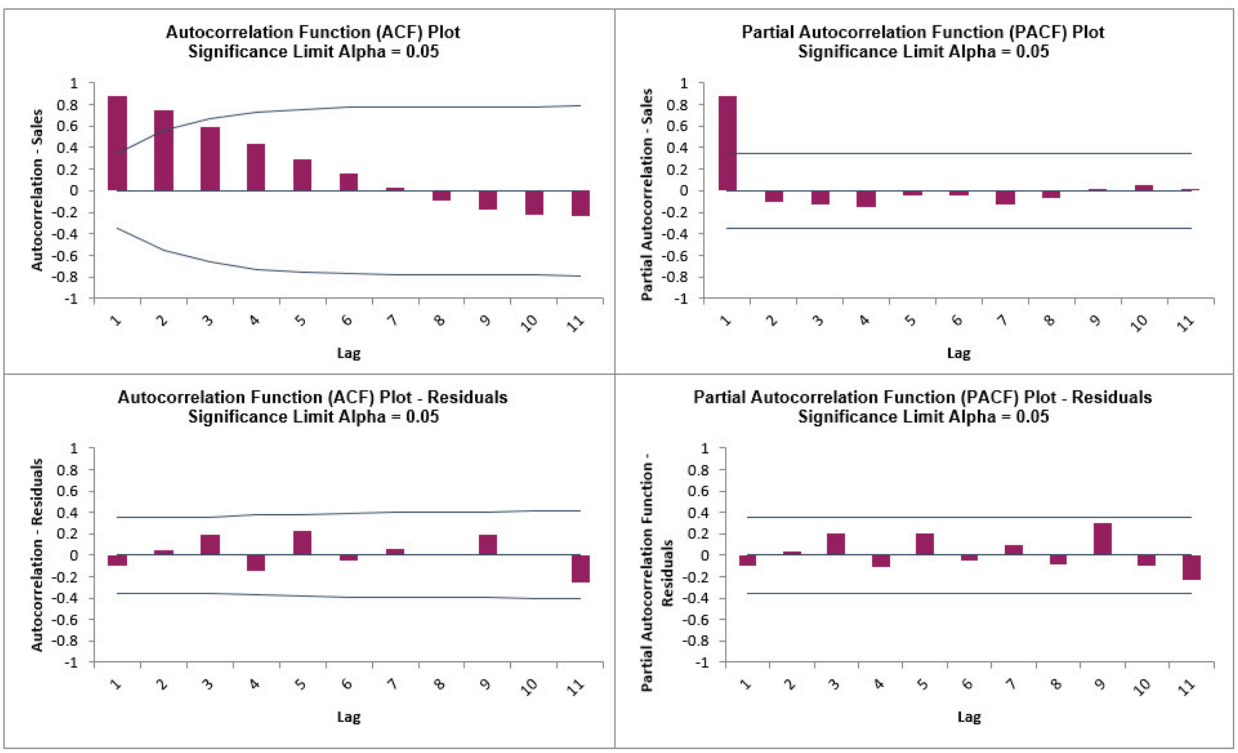
38. The Forecast Accuracy metrics are:

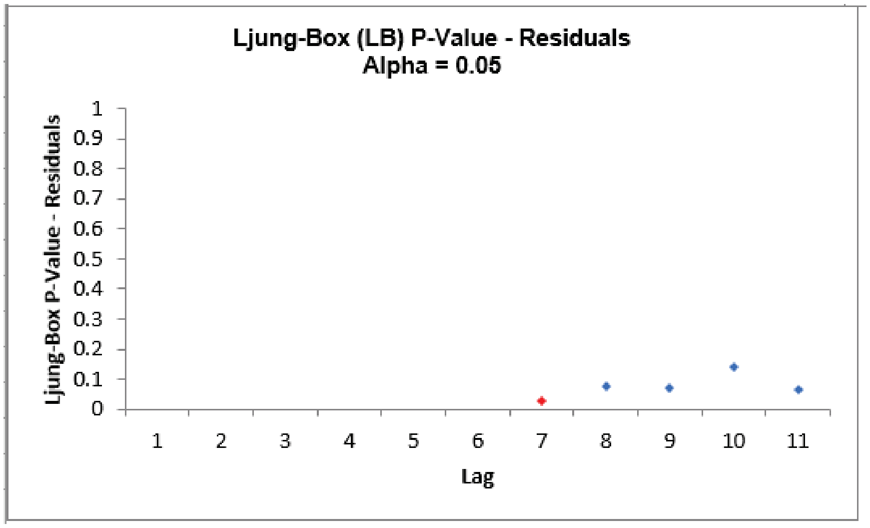
Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step Ahead Forecast**
N	33	12
RMSE	0.44993435	0.505510545
MAE	0.363143972	0.420790326
MAPE	0.164705945	0.162132467
MASE	0.121786088	0.141118707

Compared to the previous Forecast Accuracy metrics the Out-of-Sample forecast errors are slightly lower, so this is a good prediction with a simpler model.

MASE is less than one, so it is a better forecast than would be obtained from a naïve forecast (set all forecasts to be the value of the last observation). See Forecast Accuracy Metrics.

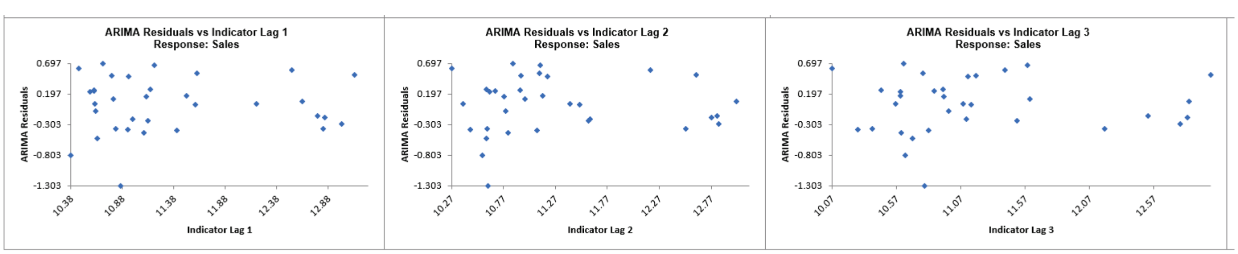
39. Click on the **ARIMA ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





We can see that much of the autocorrelation has been removed by the ARIMA with Predictors model. The Ljung-Box plot shows fewer significant P-Values than the previous model.

40. Click on the **ARIMA Residuals** sheet to view the Residual Plots:



The residuals look similar to those in the previous model, approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts. The histogram might suggest a left skew but the sample size is quite small. (Normality tests may also be applied on the residuals using SigmaXL > Statistical Tools > Descriptive Statistics: Options, Additional Normality Tests, and these would all show the residuals to be normal).

Note that additional residual plots are provided for each predictor, *Indicator Lag 1* to *4*.

ARIMA – Multiple Seasonal Decomposition (MSD) Forecast

ARIMA does not have a theoretical frequency limit, but for computational efficiency and to minimize the potential loss of observations through differencing, we recommend using ARIMA – Multiple Seasonal Decomposition (MSD) for seasonal frequency greater than 52 (or with multiple frequencies).

The seasonal component is first removed through decomposition, a nonseasonal ARIMA model fitted to the remainder (+trend), and then the seasonal component is added back in. For forecasting, a naïve seasonal forecast is used on the seasonal component. Note that the prediction intervals are derived from the ARIMA model and do not include uncertainty in the seasonal component.

As the name implies, Multiple Seasonal Decomposition (MSD) also accommodates multiple seasonality, for example the half-hourly data with a seasonal frequency of 48 observations per day and 336 observations per week. When using MSD, it is recommended to limit the forecast period to 2*dominant seasonal frequency.

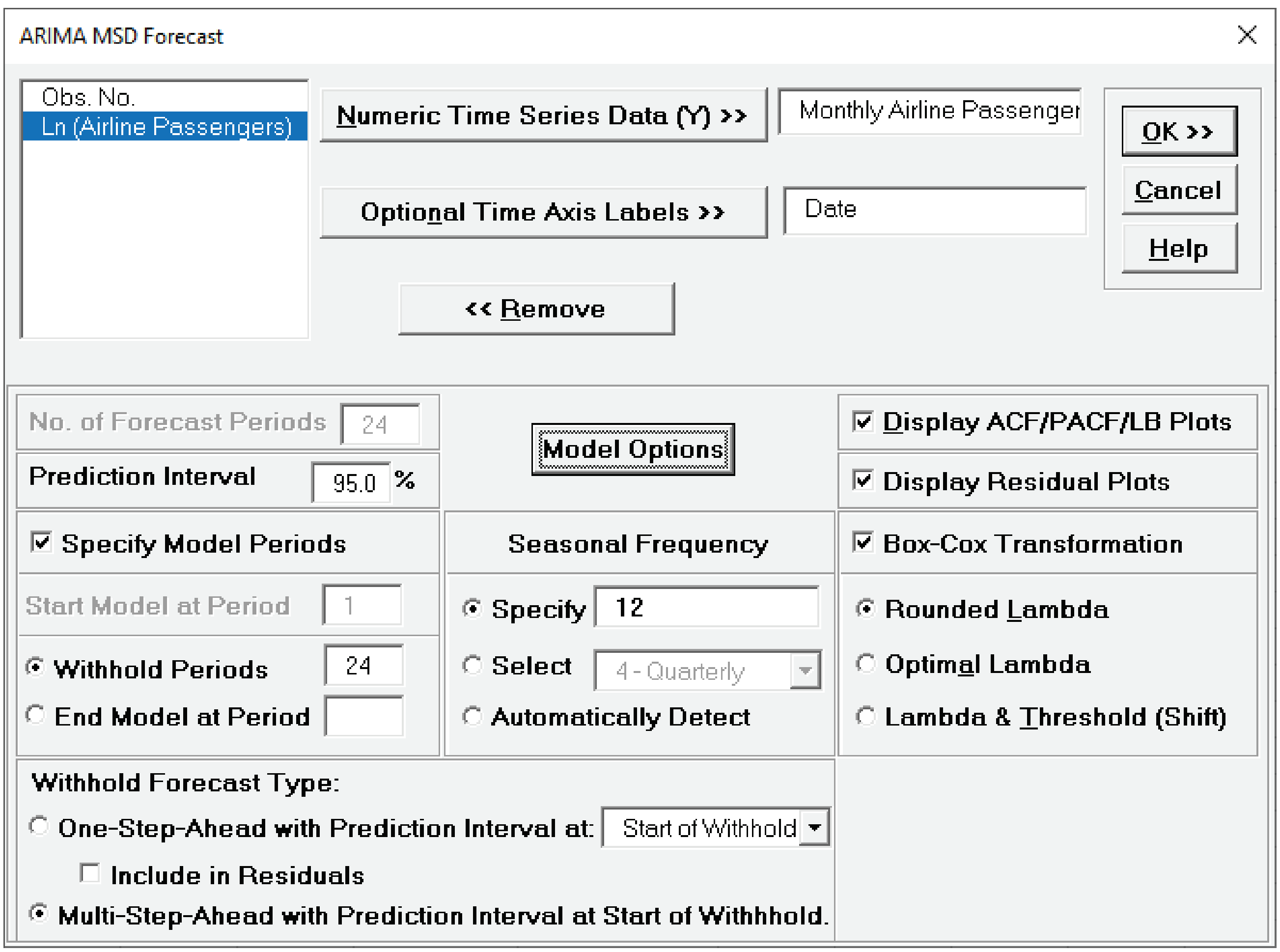
For further details and formulas, see Appendix: **Seasonal Trend Decomposition.**

Monthly Airline Passengers - Series G

1. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data. The Multiple Seasonal Decomposition (MSD) option is not necessary for this data, but by way of introduction, we will use this to compare to the previous ARIMA analysis.

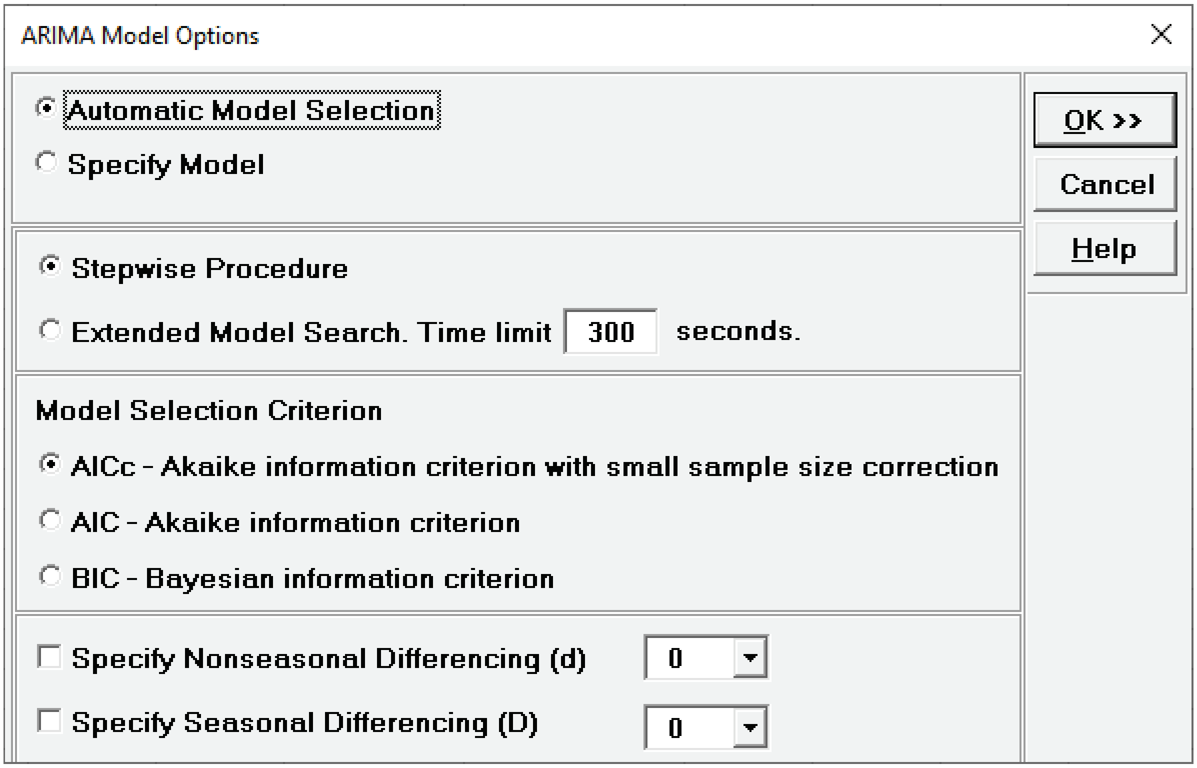
2. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Multiple Seasonal Decomposition Forecast**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Monthly Airline Passengers*, click **Numeric Time Series Data (Y) >>**. Select *Date*, click **Optional Time Axis Labels >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 24. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Select **Seasonal Frequency Specify** and enter 12**.** Check **Box-Cox Transformation** and select **Rounded Lambda**. We will use the default **Prediction Interval** = 95.0 %.

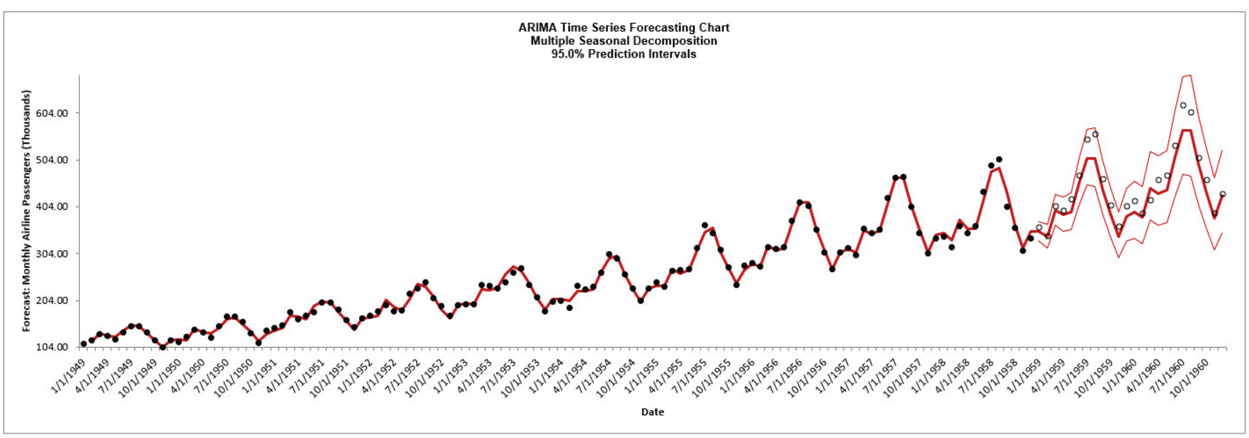


- **Seasonal Frequency** can have multiple entries however, we recommend no more than 3 values.

4. Click **Model Options**.



5. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the ARIMA MSD Forecast dialog. Click **OK**. The ARIMA forecast report is given:



6. Scroll down to view the ARIMA Model header:

ARIMA Model (Multiple Seasonal Decomposition): Monthly Airline Passengers (Thousands) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

7. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	12 Decomposition
Include Constant	1
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

This is a summary of the model information for the deseasonalized data: ARIMA (0,1,1) with a constant. Seasonal Frequency = 12 using Decomposition and Model Selection Criterion = "AICc". There are no seasonal terms in the model. The Box-Cox Transformation is "Rounded Lambda" with Lambda = 0 (Ln transformation).

8. The Parameter Estimates for the deseasonalized Airline Passenger data are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.322113898	0.095098865	3.38715	**0.0010**
Const:Trend	0.009427003	0.001881534	5.01027	**0.0000**

9. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	120
DF	117
StDev	0.030409792
Variance	0.000924755
Log-Likelihood	247.7313206
AICc	-489.253946
AIC	-489.462641
BIC	-481.125271

- The number of observations, *n* = 144 – 24 (*withhold*) = 120
- Degrees of freedom (DF) = 120 (*n*) – 3 (2 terms in the model, 1 nonseasonal difference) = 117
- Note that the model statistics are based on the Ln transformed data, not the original data.
- Comparing to the Exponential Smoothing MSD Model Statistics, we see that the StDev and Variance are approximately equal, but the Log-Likelihood, AICc, AIC and BIC are very different. This is due to different formulas being used in the Likelihood function. You cannot use Information Criteria to compare ARIMA and Exponential Smooth models to determine which model has the best fit.

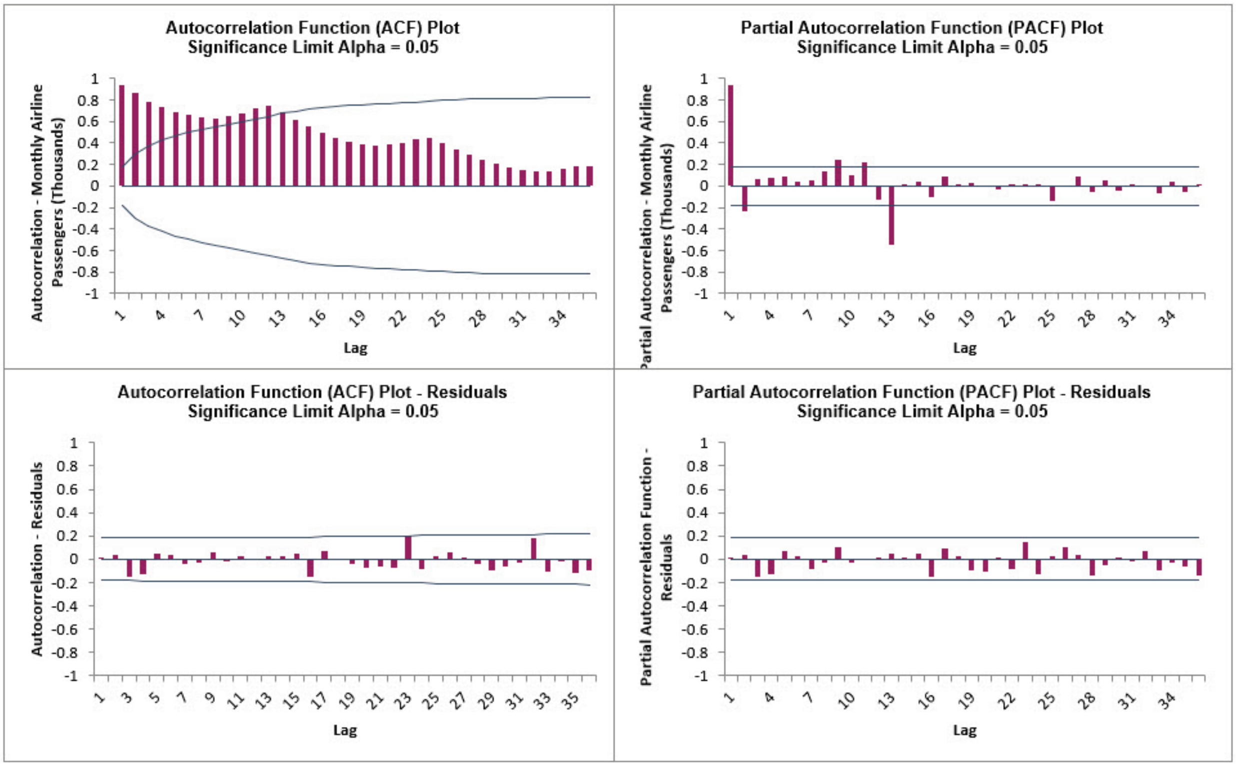
10. The Forecast Accuracy metrics are:

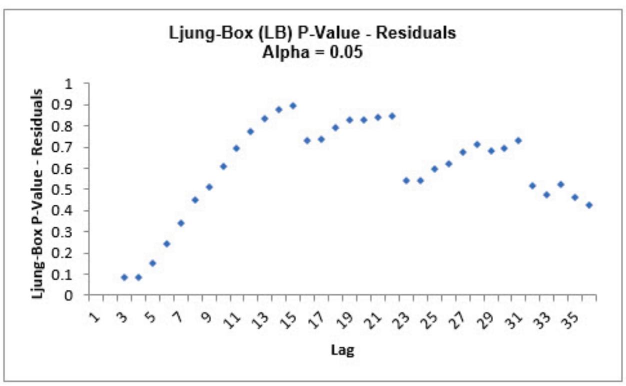
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step Ahead Forecast
N	119	24
RMSE	7.145744373	26.75002158
MAE	5.494721757	23.0424413
MAPE	2.364132406	4.85347701
MASE	0.192297456	0.806410778

Comparing to Exponential Smoothing MSD Forecast Accuracy metrics, we see that the ARIMA MSD results are approximately the same.

Forecast Accuracy metrics are calculated using the actual raw data versus inverse transformed forecast as displayed in the Forecast Chart and Table, so allow comparison across all model types and transformations.

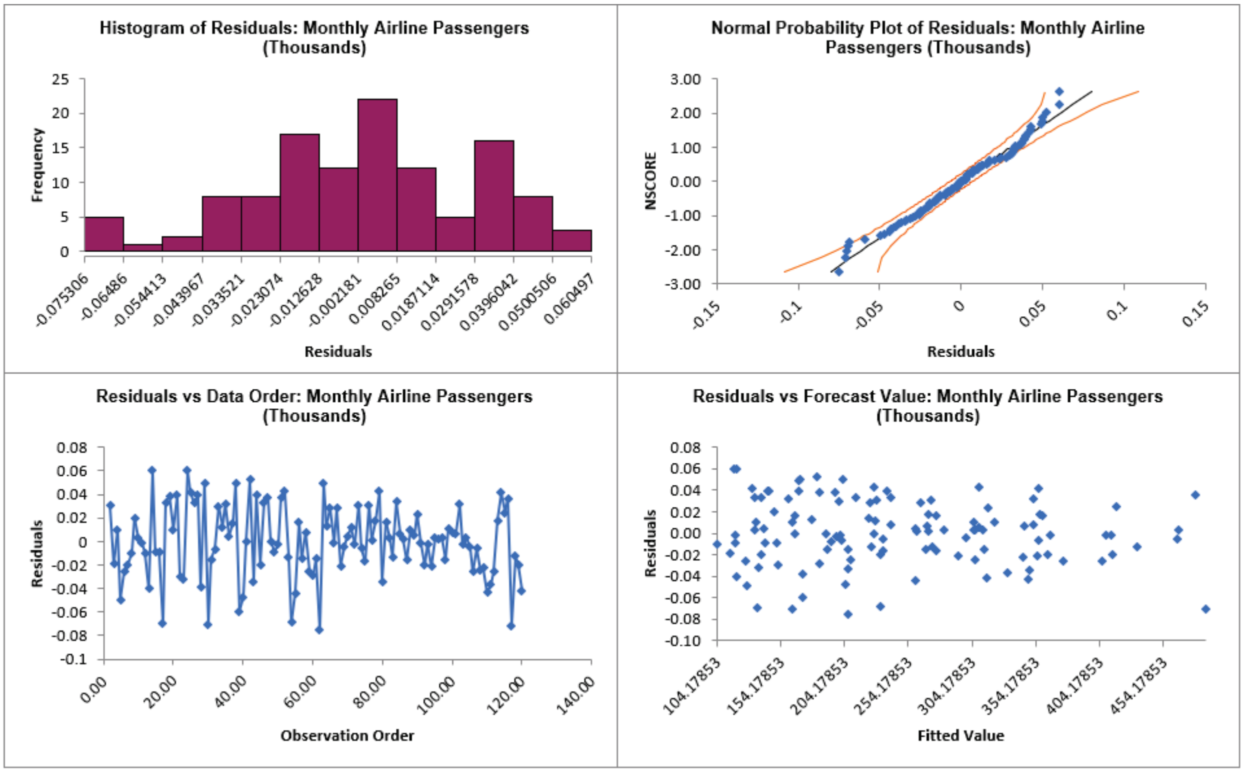
11. Click on the **ARIMA MSD ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ARIMA MSD ACF/PACF Residuals Plots are similar to the Exponential Smoothing MSD ACF/PACF/LB Plots and indicate that the autocorrelation has been accounted for in the model. All of the Ljung-Box P-Values are blue (i.e., not significant), with P-Values > .05.

12. Click on the **ARIMA MSD Residuals** sheet to view the Residual Plots:



The residuals are approximately normally distributed, with a roughly straight line on the normal probability plot. There are no obvious extreme outliers or patterns in the charts. The residual plots are approximately the same as those given previously in Exponential Smoothing MSD Residual Plots.

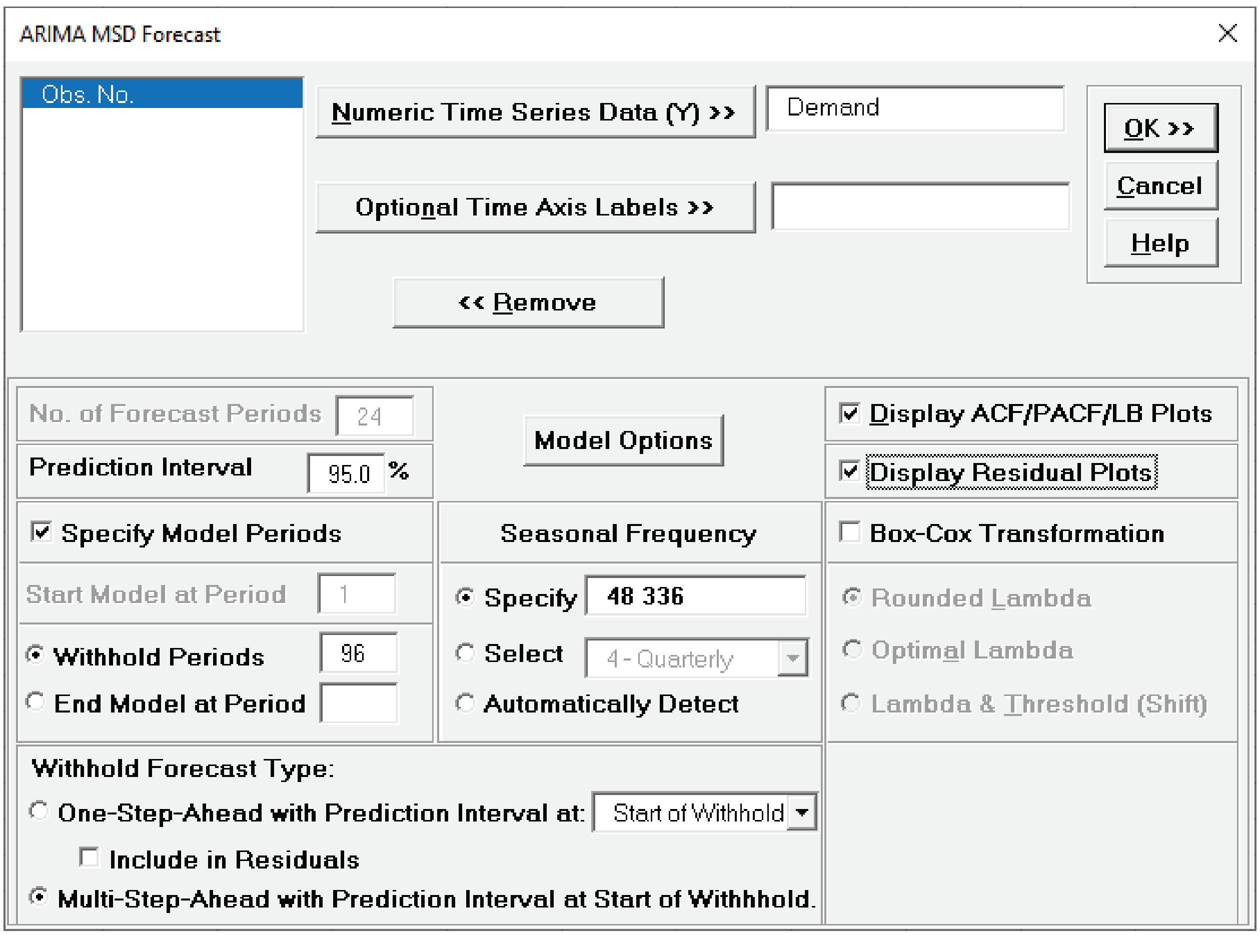
Since a Box-Cox transformation was used, the residuals are in Ln transformed units.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

13. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

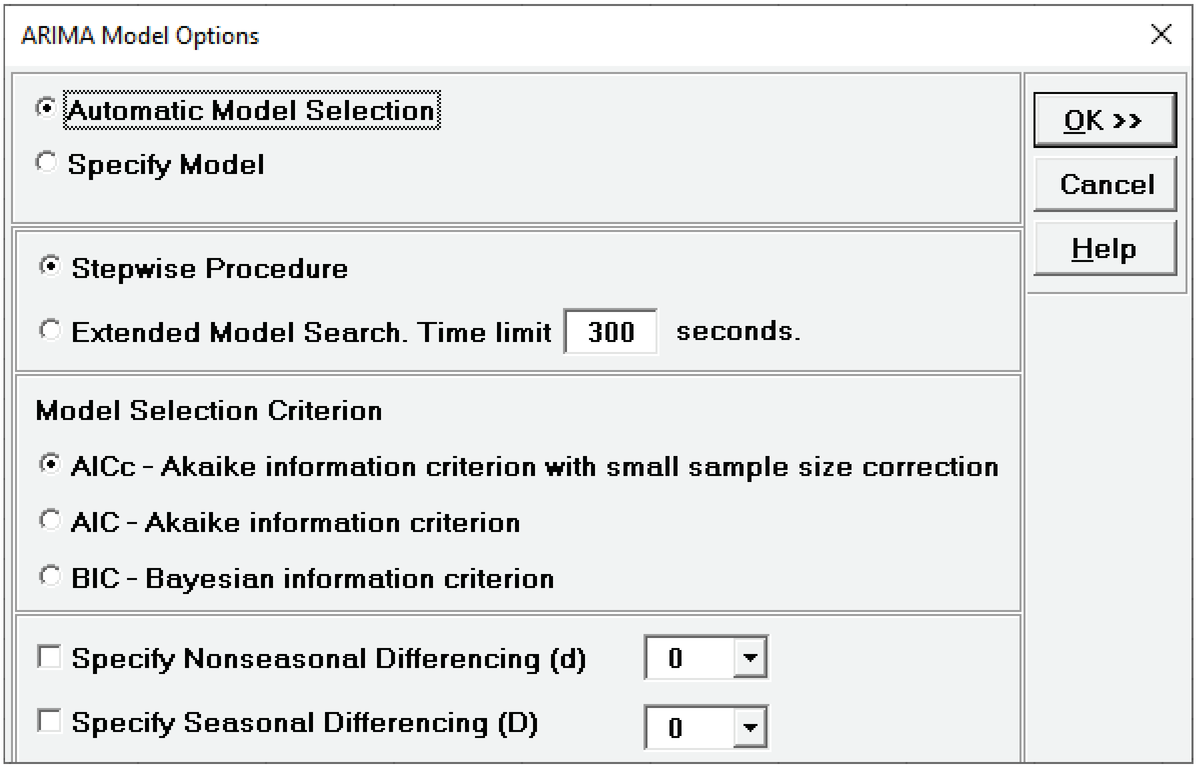
14. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Multiple Seasonal Decomposition Forecast**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

15. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Check **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Specify Model Periods**. Set **Withhold Periods** = 96. Select **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Check **Seasonal Frequency** with **Specify** = 48 336**.** Leave **Box-Cox Transformation** unchecked. We will use the default **Prediction Interval** = 95.0 %.

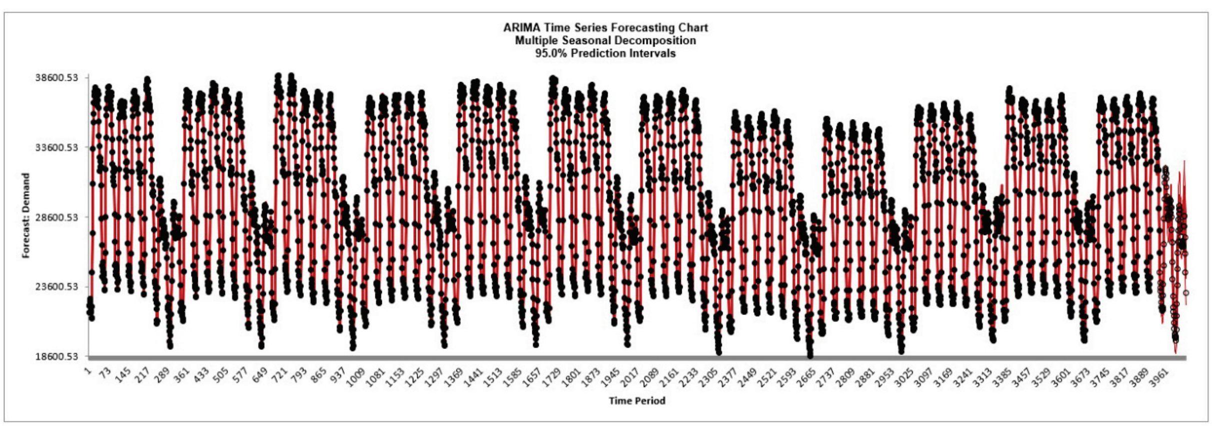


Withhold Periods is 2*dominant seasonal frequency (48). Dominant frequency is obtained from the Spectral Density Plot. Start Model at Period = 1 is always greyed out for MSD.

16. Click **Model Options**.



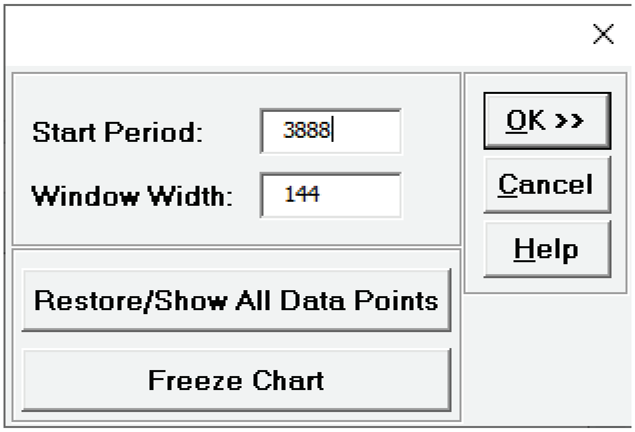
17. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the ARIMA MSD Forecast dialog. Click **OK**. The ARIMA forecast report is given:



18. We will want to zoom in on the last 3 days, i.e., 144 half-hourly time periods, using chart scrolling. Click **SigmaXL Chart Tools > Enable Scrolling**

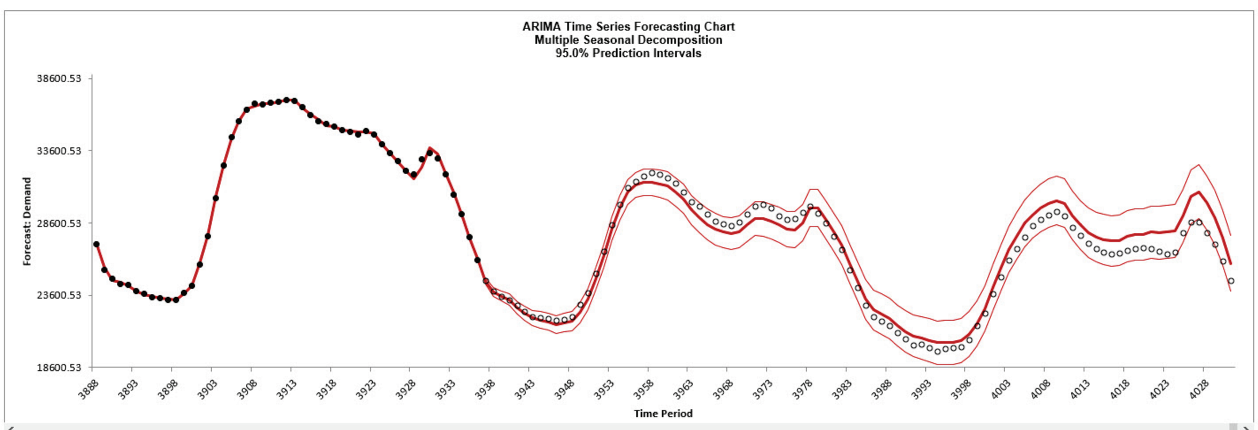
You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

19. Click **OK.** The scroll dialog appears allowing you to specify the **Start Period** and **Window Width**. Enter **Start Period** = 3888 and **Window Width** = 144:



At any point, you can click **Restore/Show All Data Points** or **Freeze Chart**. Freezing the chart will remove the scroll and unload the dialog. The scroll dialog will also unload if you change worksheets. To restore the dialog, click **SigmaXL Chart Tools > Enable Scrolling**.

20. Click **OK.** A scroll bar appears beneath the forecast chart. You can also change the **Start Subgroup** and **Window Width** and **Update**.



You can scroll through by clicking to the right or left, with the specified window width of 144.

The blank dots are the data values in the withhold sample with a multi-step forecast and prediction intervals displayed at the start of the withhold sample. The model does quite well at predicting the withhold 96 half-hour demand values. Note that the prediction error increases the further out we predict. The ARIMA MSD Forecast Chart looks approximately the same as the previous Exponential Smoothing MSD Forecast Chart

Click **Cancel** to exit the scroll dialog.

21. Scroll down to view the ARIMA Model header:

ARIMA Model (Multiple Seasonal Decomposition): Demand - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 96 withhold periods.

22. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	0
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	48, 336 Decomposition
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information for the deseasonalized data: ARIMA (2,1,0) without a constant. Seasonal Frequency = 48, 336 using Decomposition and Model Selection Criterion = "AICc". There are no seasonal terms in the model.

23. The Parameter Estimates for the deseasonalized Demand data are:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
AR_1	0.033677486	0.010694482	3.14905	**0.0017**
AR_2	-0.022798133	0.0160335	1.42191	0.1551

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. This may be useful for model refinement with multiple predictors but for AR/MA model order selection, minimum AICc should be used, rather than significance tests (see Kostenko, A.V. and Hyndman, R.J.).

24. The ARIMA Model Statistics are:

ARIMA Model Statistics	
No. of Observations	3936
DF	3933
StDev	98.8439103
Variance	9770.11861
Log-Likelihood	-23647.086
AICc	47315.0387
AIC	47315.0326
BIC	47333.8656

- The number of observations, *n* = 4032 – 96 (*withhold*) = 3936
- Degrees of freedom (DF) = 3936 (*n*) – 3 (2 terms in the model, 1 nonseasonal difference) = 3933
- Note that the model statistics are calculated using the deseasonalized data.

25. The Forecast Accuracy metrics are:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step Ahead Forecast
N	3935	96
RMSE	101.4595732	775.9894896
MAE	73.7919471	649.1824792
MAPE	0.253330267	2.452663125
MASE	0.040188511	0.353557244

As expected, the **Out-of Sample (Withhold) Multi-Step-Ahead Forecast** errors are larger than the **In-Sample (Estimation) One-Step-Ahead Forecast** errors. Note, if we were primarily interested in a short term one-step ahead forecast, then we would have selected **Withhold Forecast Type: One-Step-Ahead** and the above table would show **Out-of Sample (Withhold) One-Step-Ahead Forecast** errors.

Comparing to the previous Exponential Smoothing MSD Forecast Accuracy metrics, we see that the ARIMA MSD results are approximately the same.

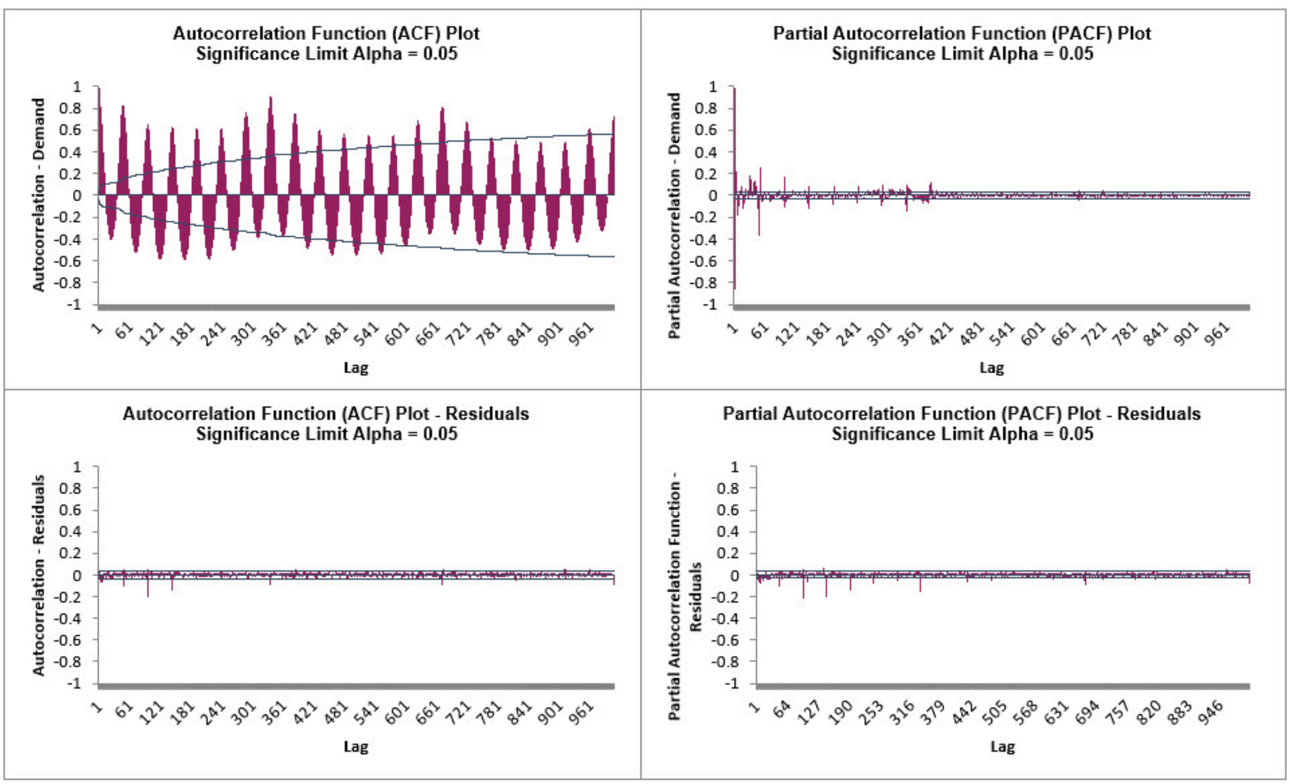
Forecast Accuracy metrics are calculated using the actual raw data versus forecast as displayed in the Forecast Chart and Table so, unlike the model statistics above, allows comparison across all forecast model types and transformations.

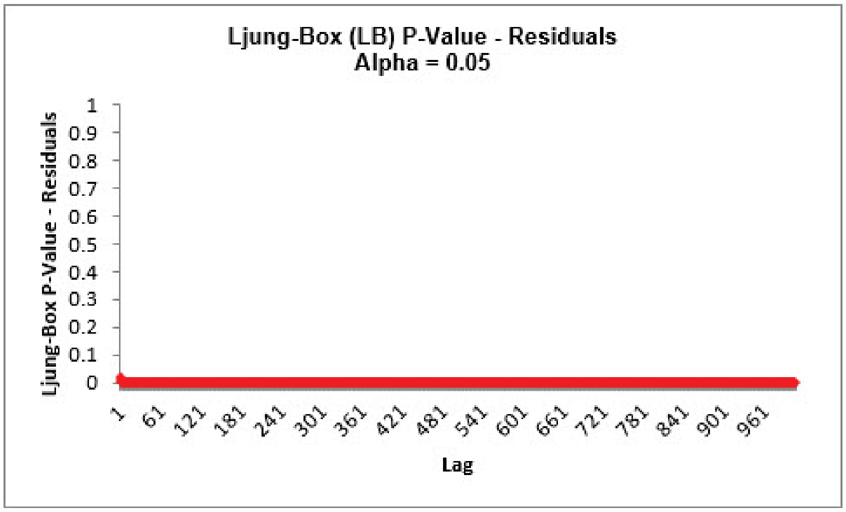
26. The Forecast Table is given as:

Forecast Table				
Period	**Withhold Data**	**Multi-Step Ahead Forecast**	**Lower 95.0% PI**	**Upper 95.0% PI**
3937	24653	24591.46494	24397.67483	24785.25505
3938	23879	23836.7087	23557.99505	24115.42235
3939	23508	23543.90813	23203.10958	23884.70668
3940	23275	23298.25208	22905.19703	23691.30713
3941	22890	22733.71045	22294.53559	23172.88531
3942	22462	22347.61	21866.71441	22828.5056

These are the same forecast and prediction interval values displayed in the Forecast Chart, but provided for further analysis or charting (e.g., a run chart of the forecast errors). The **Withhold Data** is also displayed.

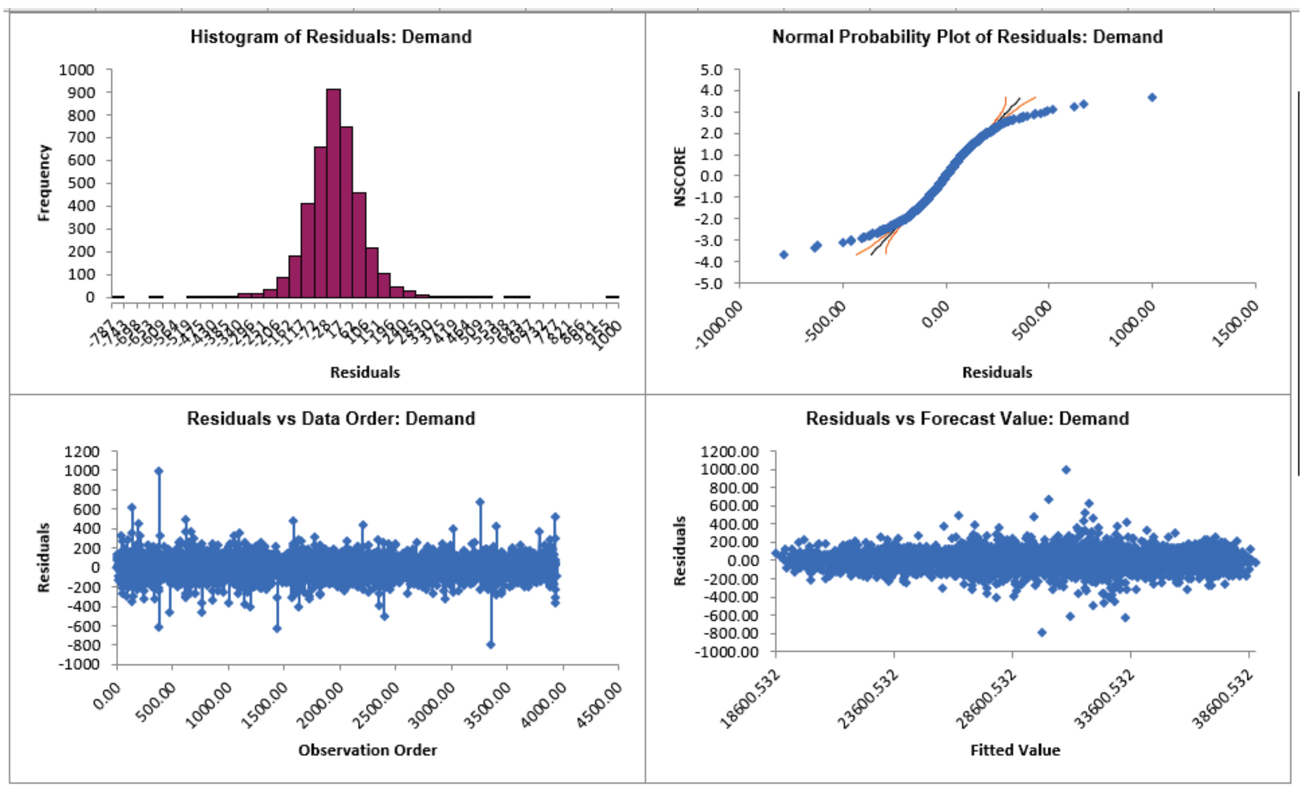
27. Click on the **ARIMA MSD ACF PACF LB** sheet to view the ACF/PACF/LB Plots:





The ACF/PACF Residuals Plots indicate that much of the autocorrelation has been accounted for in the model, but the Ljung-Box plot shows that some significant autocorrelation still remains (the red P-Values are significant at alpha=.05) - so the model can potentially be improved. This does not mean that the model is a bad model, it can still be very useful for prediction purposes, but the prediction intervals may not provide accurate coverage.

28. Click on the **ARIMA MSD Residuals** sheet to view the Residual Plots:



Similar to the Exponential Smoothing MSD Residual Plots, the residuals are not normally distributed and there are extreme outliers. These should be investigated with a control chart on the residuals. Outliers in Electricity Demand are often explained by Temperature.

ARIMA Control Chart

Statistical Process Control (SPC) For Autocorrelated Data

An Individuals control chart is created using the residuals of the ARIMA forecast model.

The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits.

The popular “Add Data”, “Show Last 30” and “Scroll” features in SigmaXL Chart Tools are available for these control charts. The time series models are not refitted, but used to compute the residual values for the new data.

For further details and references, see the Appendix: **Control Charts for Autocorrelated Data**.

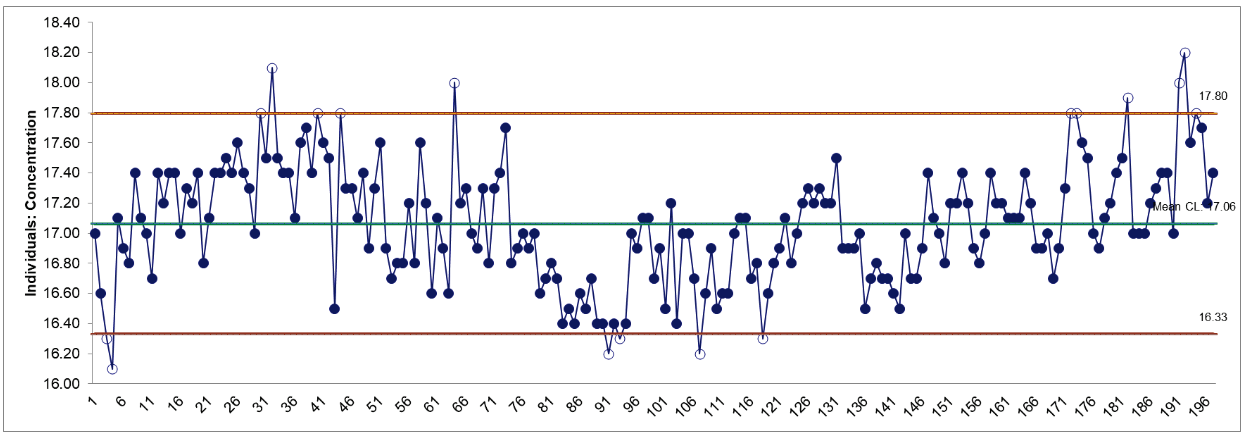
Note that a Moving Range Chart and Tests for Special Causes are not available here, but the user can store and select Residuals, then create with **SigmaXL > Control Charts > Individuals & Moving Range**.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

2. Earlier we saw that this process has significant autocorrelation. In order to see the impact on a control chart, we will construct an Individuals chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

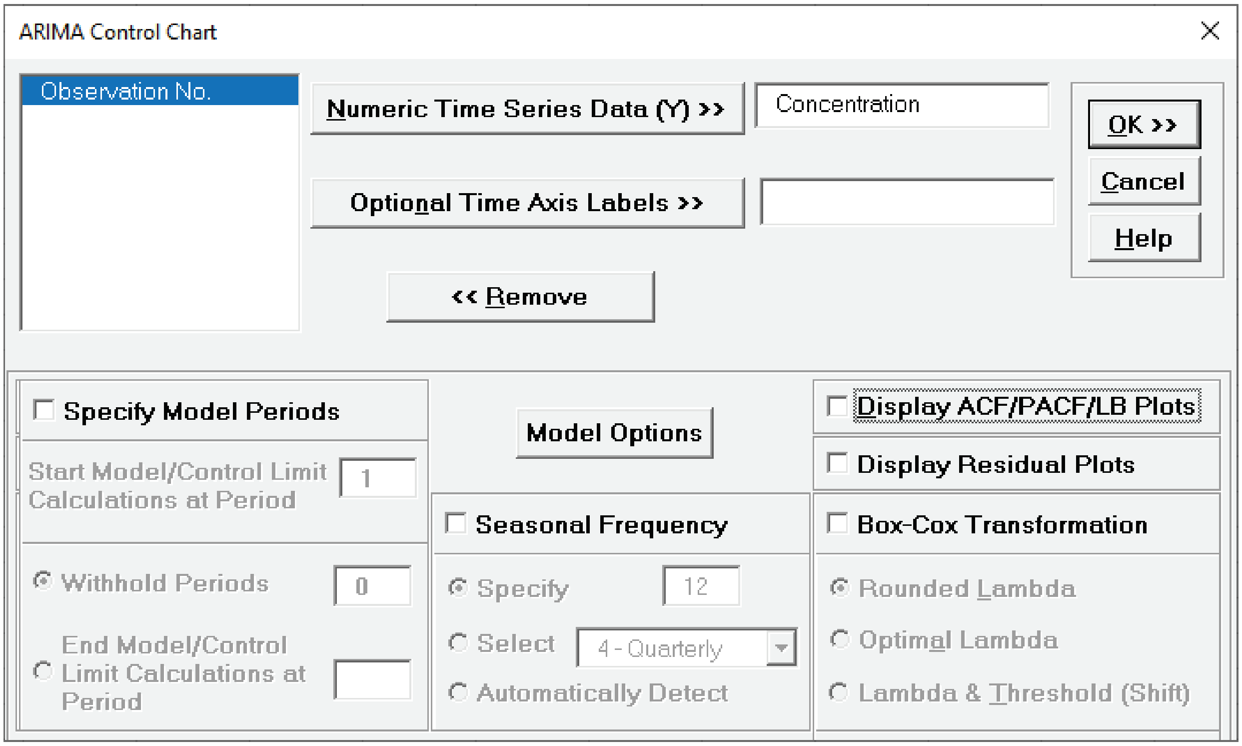
3. Select *Concentration*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



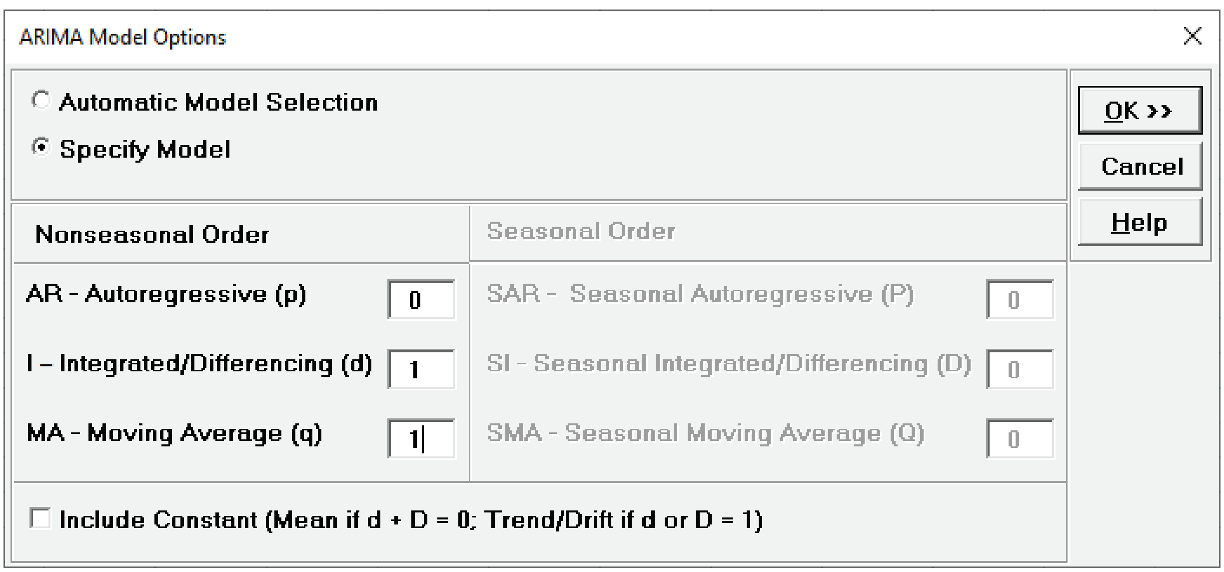
   There are 17 out-of-control data points, largely due to the autocorrelation. Searching for assignable causes using this chart as is, would be futile.

4. Now click **Sheet 1** tab and **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

5. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots.** Leave **Display Residual Plots**, **Specify Model Periods, Seasonal Frequency** and **Box-Cox Transformation** unchecked.

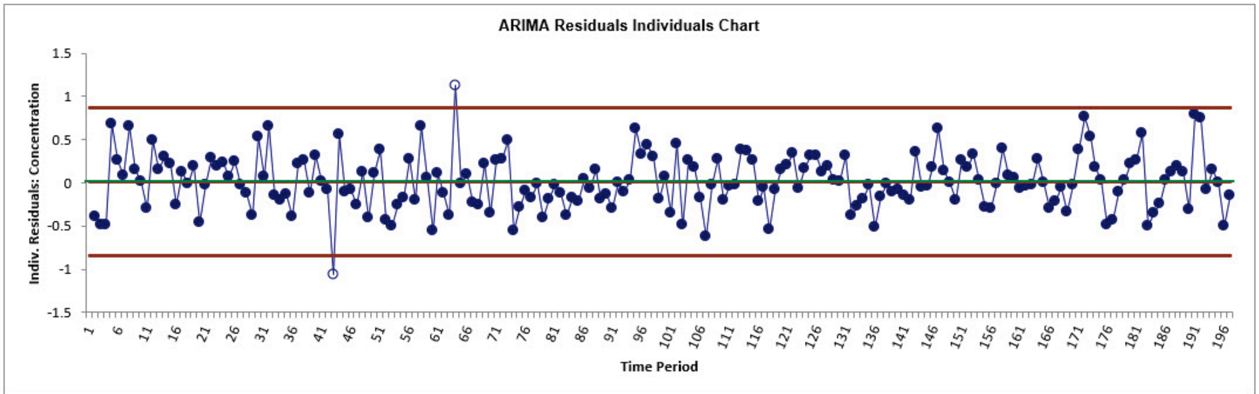


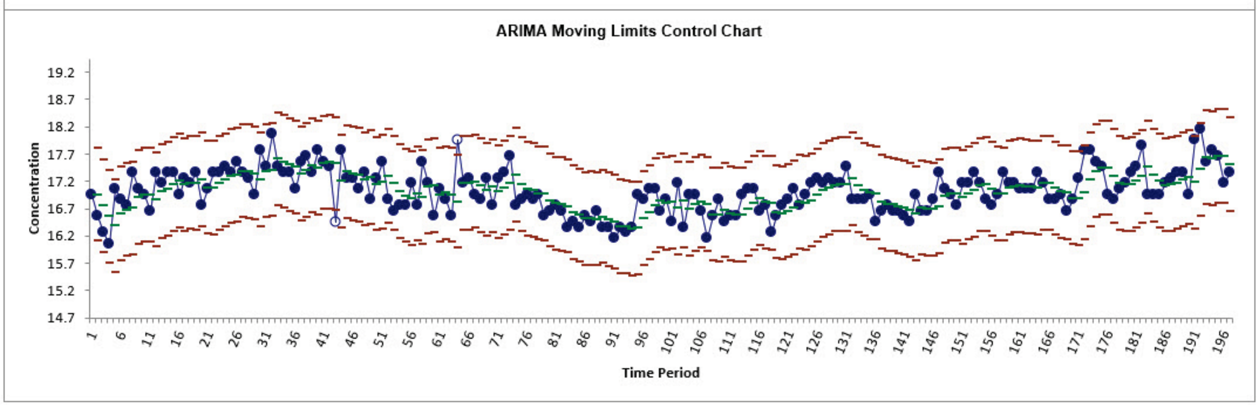
6. Click **Model Options**. Select **Specify Model**. Specify **Nonseasonal Order I – Integrated/Differencing (d)** = 1 and **MA – Moving Average (q)** = 1. Leave **Include Constant** unchecked.



Typically, one would use **Automatic Model Selection**, but we want to demonstrate the ARIMA control chart using the ARIMA (0,1,1) model as used above, and to compare to the equivalent Exponential Smoothing Control Charts.

7. Click **OK** to return to the ARIMA Control Chart dialog. Click **OK**. The ARIMA control charts are produced:





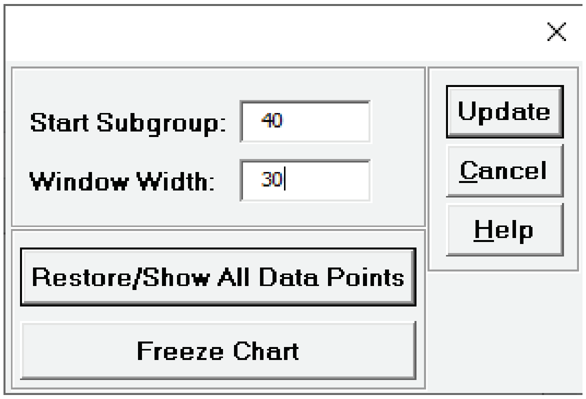
Now we only have two out-of-control data points on the Individuals chart to investigate. The Moving Limits chart uses the one step prediction as the center line, so the control limits move with the center line.

As expected, these are approximately the same as the Exponential Smoothing Control Charts, with slight differences on the initial values. The ARIMA Residuals Individuals Chart does not show a data point at Time Period = 1 due to the differencing.

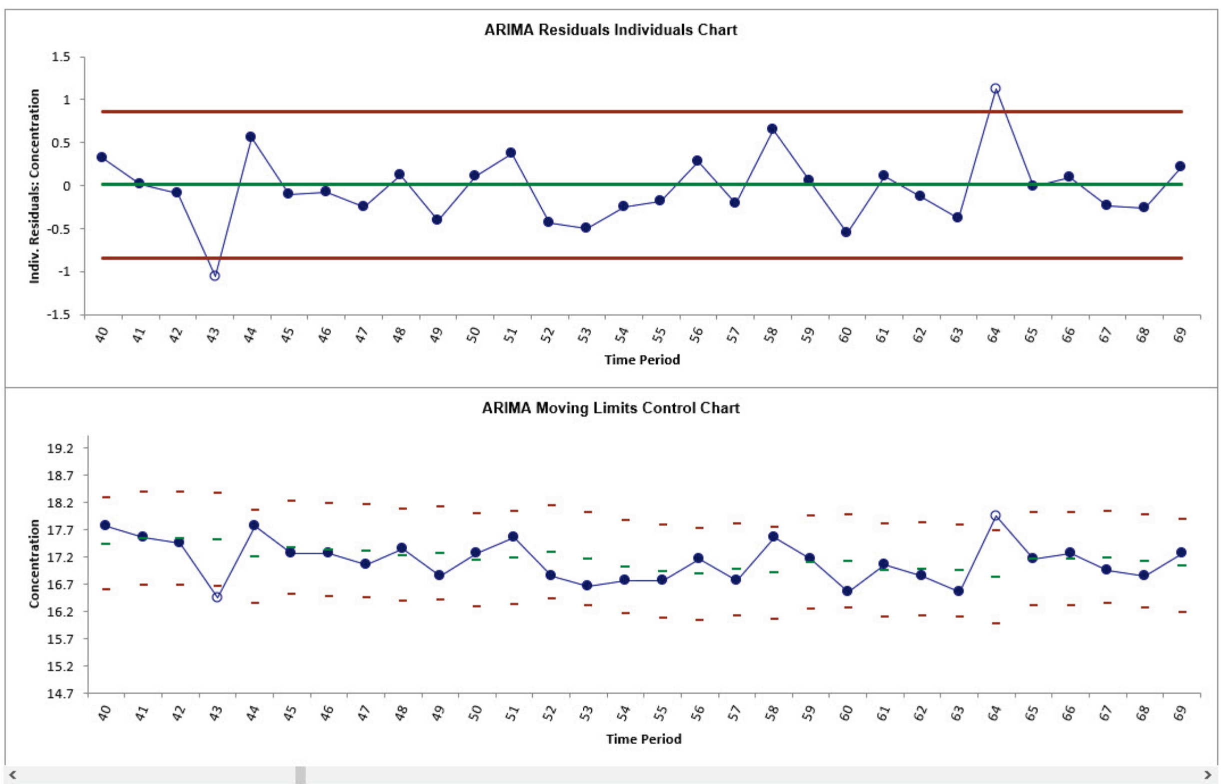
8. We will scroll through the chart data points as done in the Exponential Smoothing Control Charts. Click **SigmaXL Chart Tools > Enable Scrolling**

You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

9. Click **OK.** The scroll dialog appears allowing you to specify the **Start Subgroup** and **Window Width**. Enter **Start Subgroup** = 40 and **Window Width** = 30 to view the two out-of-control data points.



10. Click **OK.** This allows us to zoom in on the out-of-control points at 43 and 64.



Observation **43** is lower than expected from the ARIMA forecast model. Observation **64** is higher than expected.

11. Click **Cancel** to exit the scroll dialog.

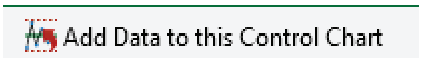
12. Now we will add a new data point to the Series A Concentration Data. The residuals will be computed using the same model as above without re-estimation of the model parameters or recalculation of the control limits. This is also known as the "Phase II" application of a Control Chart, where an out-of-control signal should lead to an investigation into the assignable cause

and corrective action or process adjustment applied. Click Sheet1, enter the value 19 as shown in cell **B199** (and optionally Observation number 198 in cell **A199**).

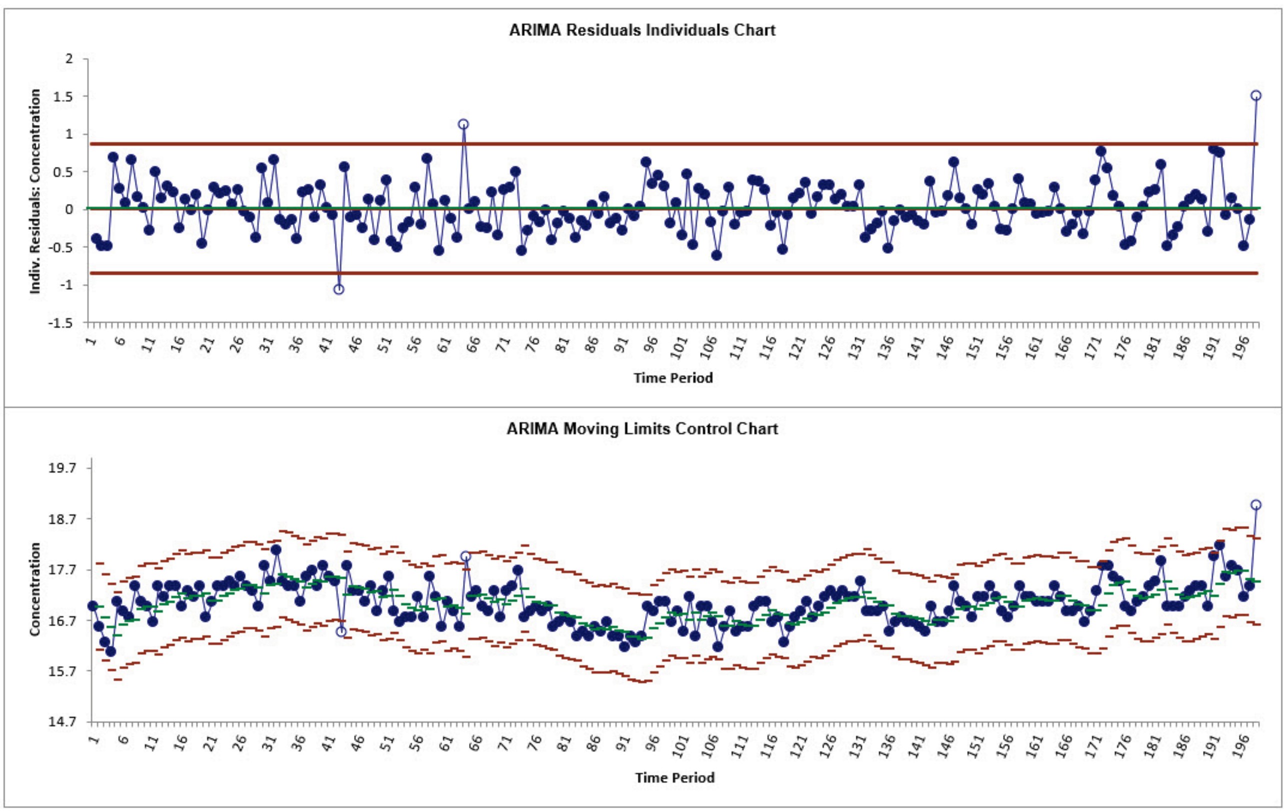
196	195	17.7
197	196	17.2
198	197	17.4
199	198	19

13. Click **ARIMA Control Charts** tab (if more than one control chart sheet exists in the workbook, please select the chart where the data will be added).

14. Click **SigmaXL Chart Tools > Add Data to this Control Chart**



15. The Residuals Individuals Control Chart and Moving Limits Charts are now updated with the new data, showing this as an out-of-control data point:



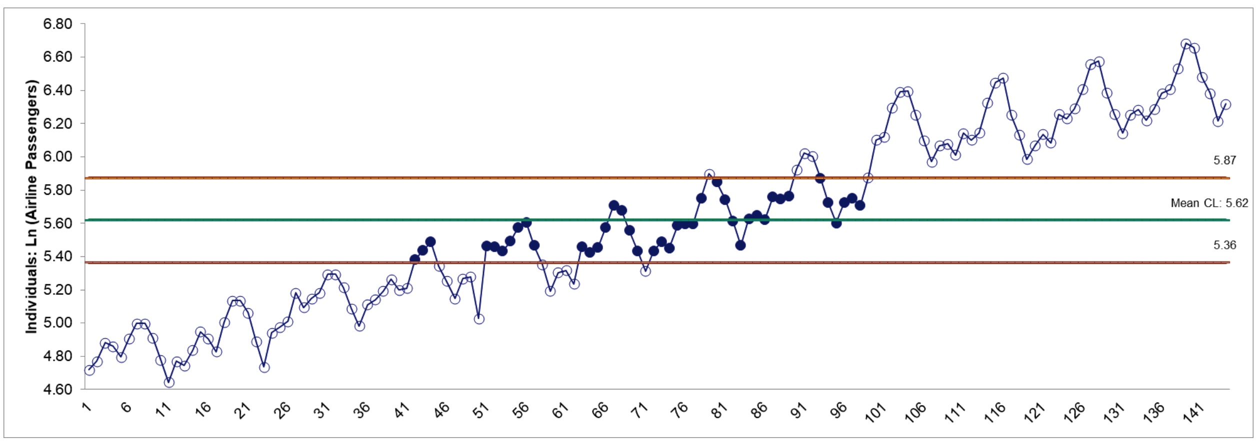
We recommend renaming the workbook to
Chemical Process Concentration – Series A_AddData2.xlsx, so that later use of the Concentration data does not include the added data point.

Monthly Airline Passengers – Modified for Control Charts

16. Open **Monthly Airline Passengers – Modified for Control Charts.xlsx** (**Sheet 1** tab). This is based on the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. A Ln transformation is applied (avoiding the need for a Box-Cox transformation), a negative outlier is added at 50 (-.25) and a level shift applied (+.25), starting at 100. Coded variables were added to help distinguish an outlier versus a shift and they will be analyzed later using ARIMA Forecast with Predictors.

17. Earlier we saw that this process has significant autocorrelation with a strong trend and seasonality. In order to see the impact on a control chart, we will construct an Individuals chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

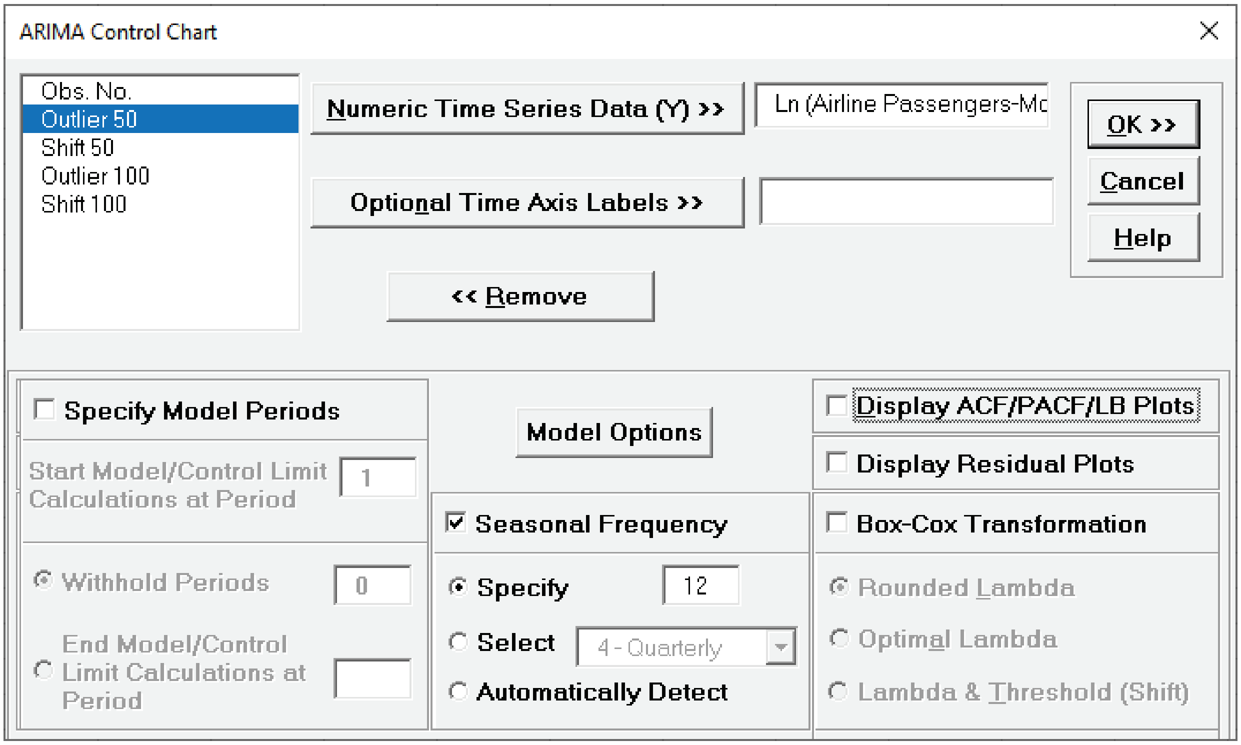
18. Select *Ln (Airline Passengers)*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



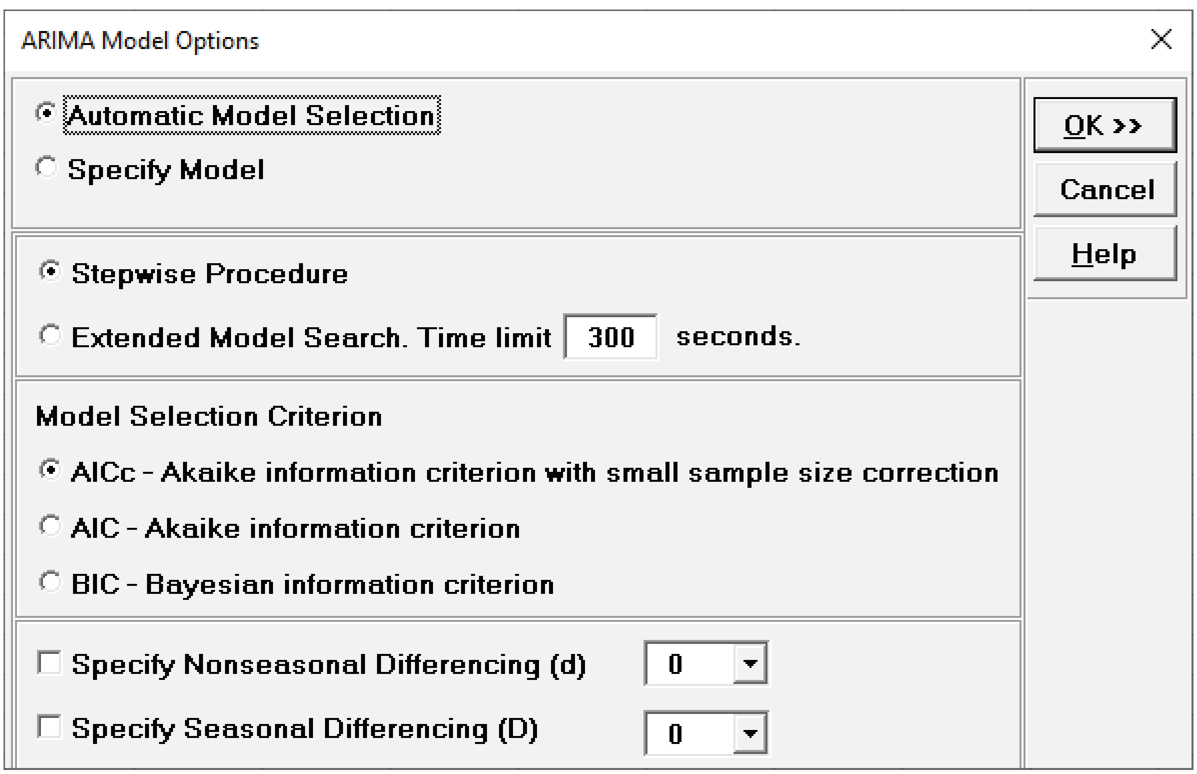
With strong trend, seasonality and positive autocorrelation, this control chart is meaningless.

19. Now click **Sheet 1** tab and **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

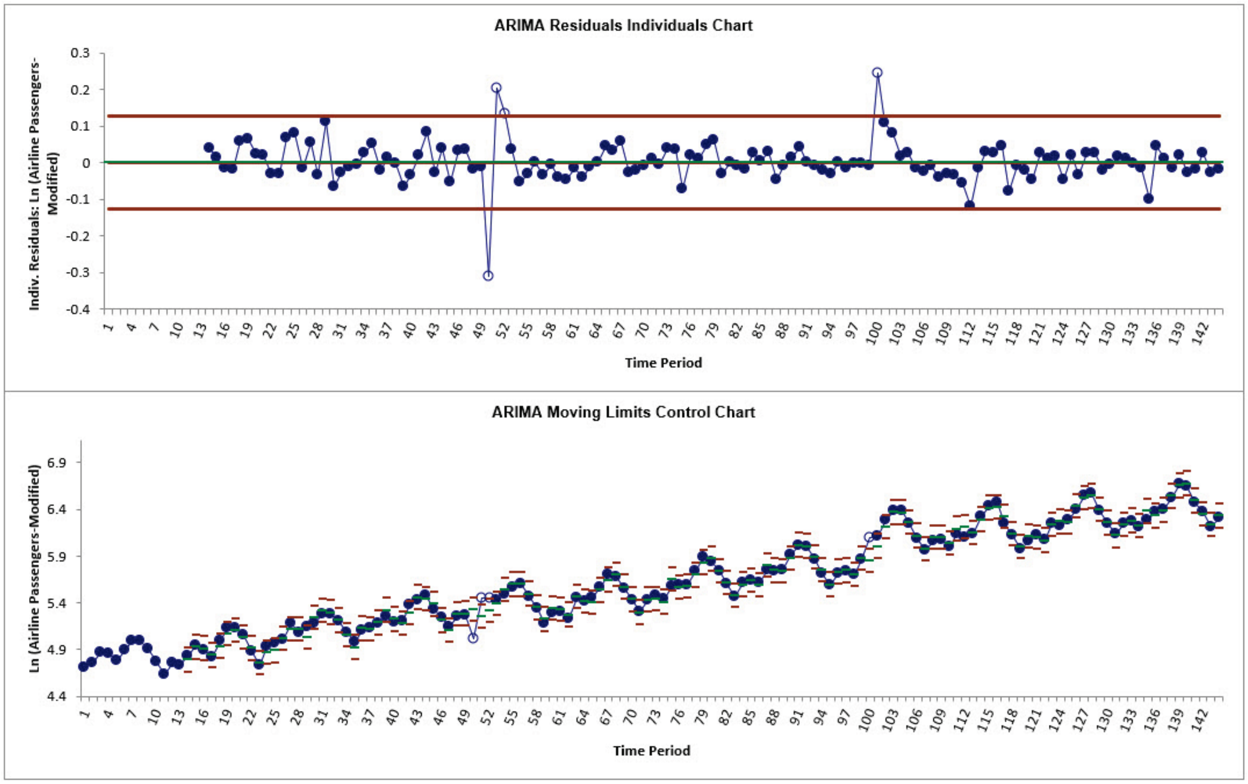
20. Select *Ln(Airline Passengers-Modified)*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots.** Check **Seasonal Frequency** with **Specify** = 12**.** Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



21. Click **Model Options**.



22. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the ARIMA Control Chart dialog. Click **OK**. The ARIMA control charts are produced:

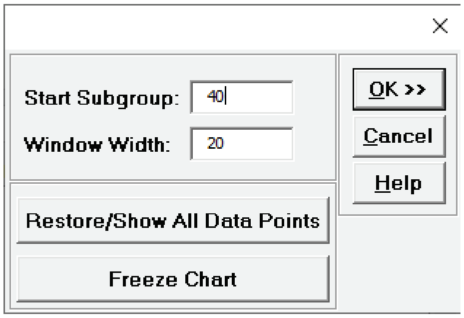


Now we can clearly see the out-of-control data points at 50, 51, 52 and 100 on the Individuals chart. These are approximately the same as those given in Exponential Smoothing Control Charts (52 was in-control). The ARIMA Residuals Individuals Chart does not show data points at Time Periods 1 to 13 due to the nonseasonal and seasonal differencing.

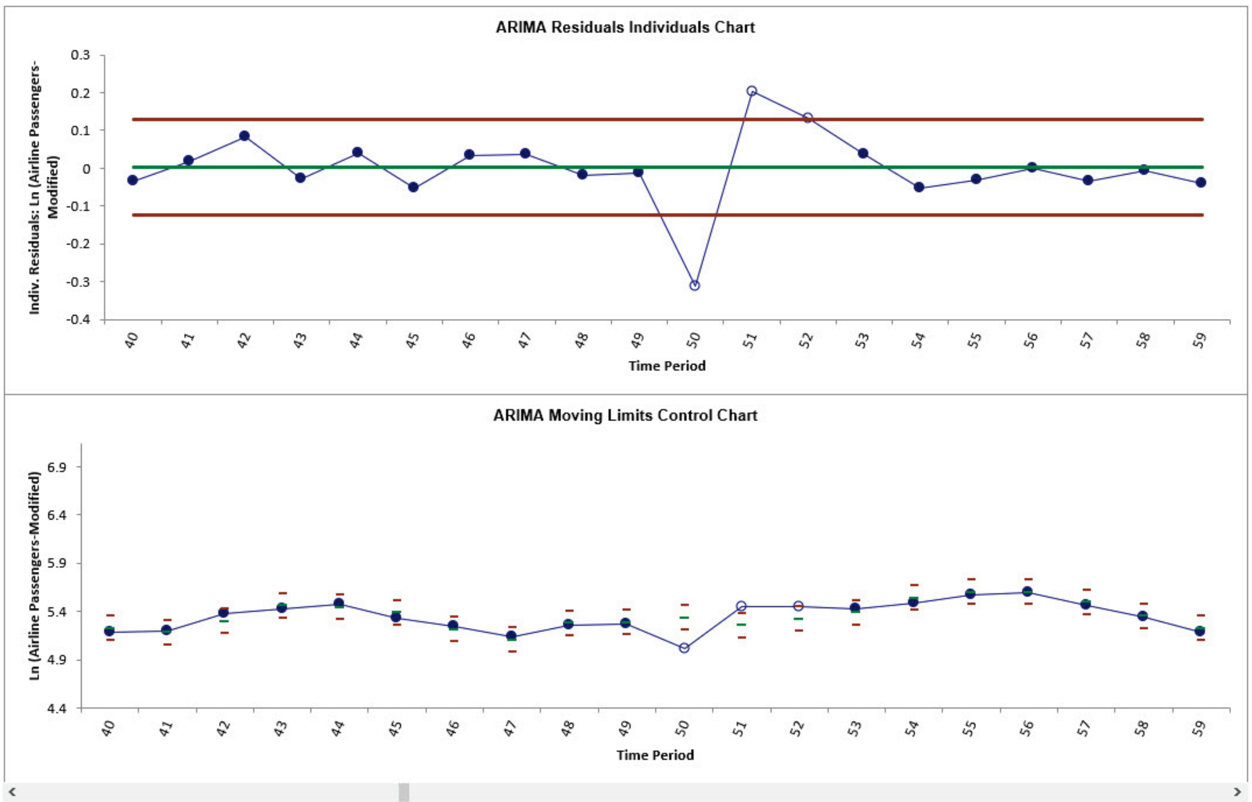
23. In order to view the points on the Moving Limits chart we will use scrolling. Click **SigmaXL Chart Tools > Enable Scrolling**

You may be prompted with a warning message that custom formatting on the chart will be cleared. You can avoid seeing this warning by checking **Save this choice as default and do not show this form again**.

24. Click **OK.** The scroll dialog appears allowing you to specify the **Start Subgroup** and **Window Width**. Enter **Start Subgroup** = 40 and **Window Width** = 20 to view the first two out-of-control data points.



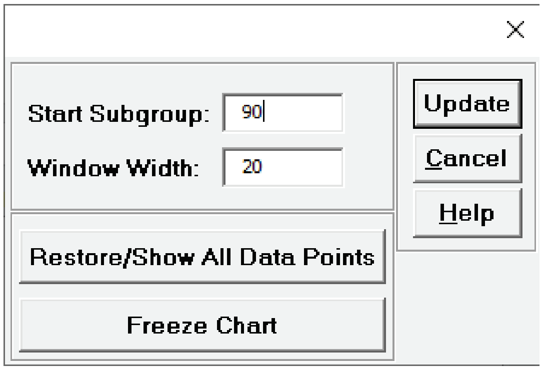
25. Click **OK.** This allows us to zoom in on the out-of-control points at 50, 51 and 52.



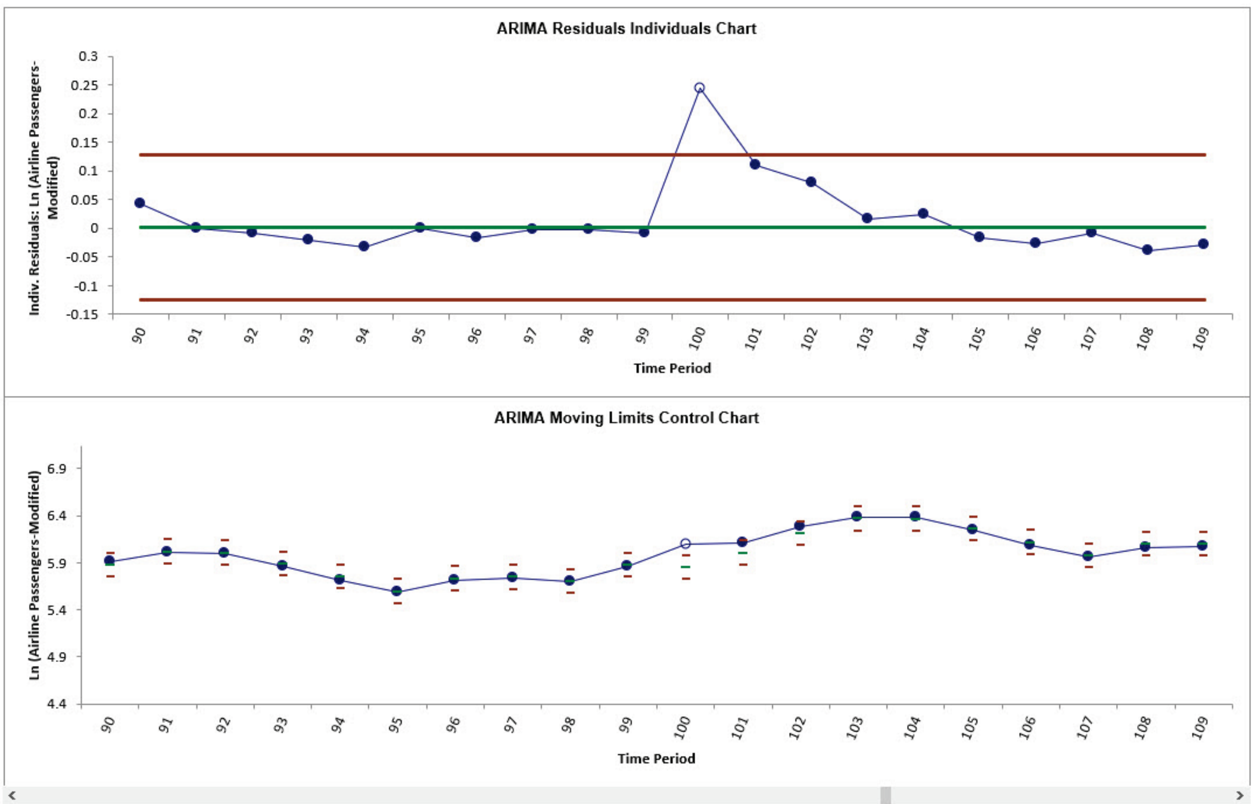
Observation **50** is lower than expected from the ARIMA forecast model. Observations **51** and **52** are higher than expected.

Tip: Scrolling keeps the original Y axis minimum and maximum setting. You may wish to change this to auto by clicking on the Y axis, right click **Format Axis**, click Bounds Minimum **Reset** and Bounds Maximum **Reset.** This changes the axis settings to Auto so when you scroll or Update the Y axis will automatically adjust as well.

26. Now enter **Start Subgroup** = 90 and **Window Width** = 20 to view the third out-of-control data point.



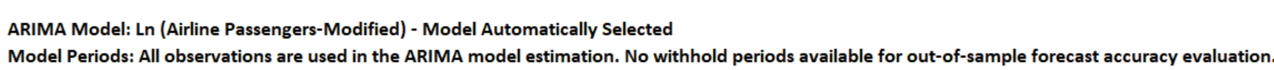
27. Click **Update**.



Observation **100** is higher than expected from the ARIMA forecast model. Later investigation will reveal that this is a shift in the mean.

28. Click **Cancel** to exit the scroll dialog.

29. Scroll down to view the ARIMA Model header:



ARIMA Model: Ln (Airline Passengers-Modified) - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

30. The ARIMA Model Summary is:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

The ARIMA (0,1,1) (0,1,1) model (with no constant) was automatically selected based on the AICc criterion. This is the same as we obtained earlier when using a withhold sample and agrees with the results shown in the ACF/PACF Plots of the manually differenced data and is the model used in Box and Jenkins, Chapter 9., "Analysis of Seasonal Time Series".

There are no predictors. Seasonal Frequency = 12; Model Selection Criterion = "AICc" and Box-Cox Transformation = "N/A".

31. The Parameter Estimates and ARIMA Model Statistics are slightly different than our earlier analysis because we have introduced an outlier and a shift, as well here we are using all of the data, i.e., there are no withhold periods. Note that earlier we used a Box-Cox Transformation with Lambda=0 and here we are using Ln of the data.

32. The Forecast Accuracy metrics are given as:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	131	
RMSE	0.055615558	
MAE	0.036285476	
MAPE	0.648152199	
MASE	0.249152676	

Note that these forecast errors are very different than our earlier analysis, where the forecast errors were calculated on the raw data versus final predicted values, but here we are using Ln of the Airline Passenger data.

33. Now we will reanalyze the data to determine whether the out-of-control signals are due to an outlier or process shift. Click the **Sheet 1** tab.

Obs. No.	Ln (Airline Passengers-Modified)	Outlier 50	Shift 50	Outlier 100	Shift 100
49	5.278114659	0	0	0	0
50	5.028114659	1	1	0	0
51	5.463831805	0	1	0	0
52	5.459585514	0	1	0	0
53	5.433722004	0	1	0	0
54	5.493061443	0	1	0	0
55	5.575949103	0	1	0	0
99	5.874930731	0	1	0	0
100	6.10220248	0	1	1	1
101	6.122117789	0	1	0	1
102	6.295005314	0	1	0	1
103	6.392037406	0	1	0	1
104	6.396329258	0	1	0	1
105	6.251414878	0	1	0	1

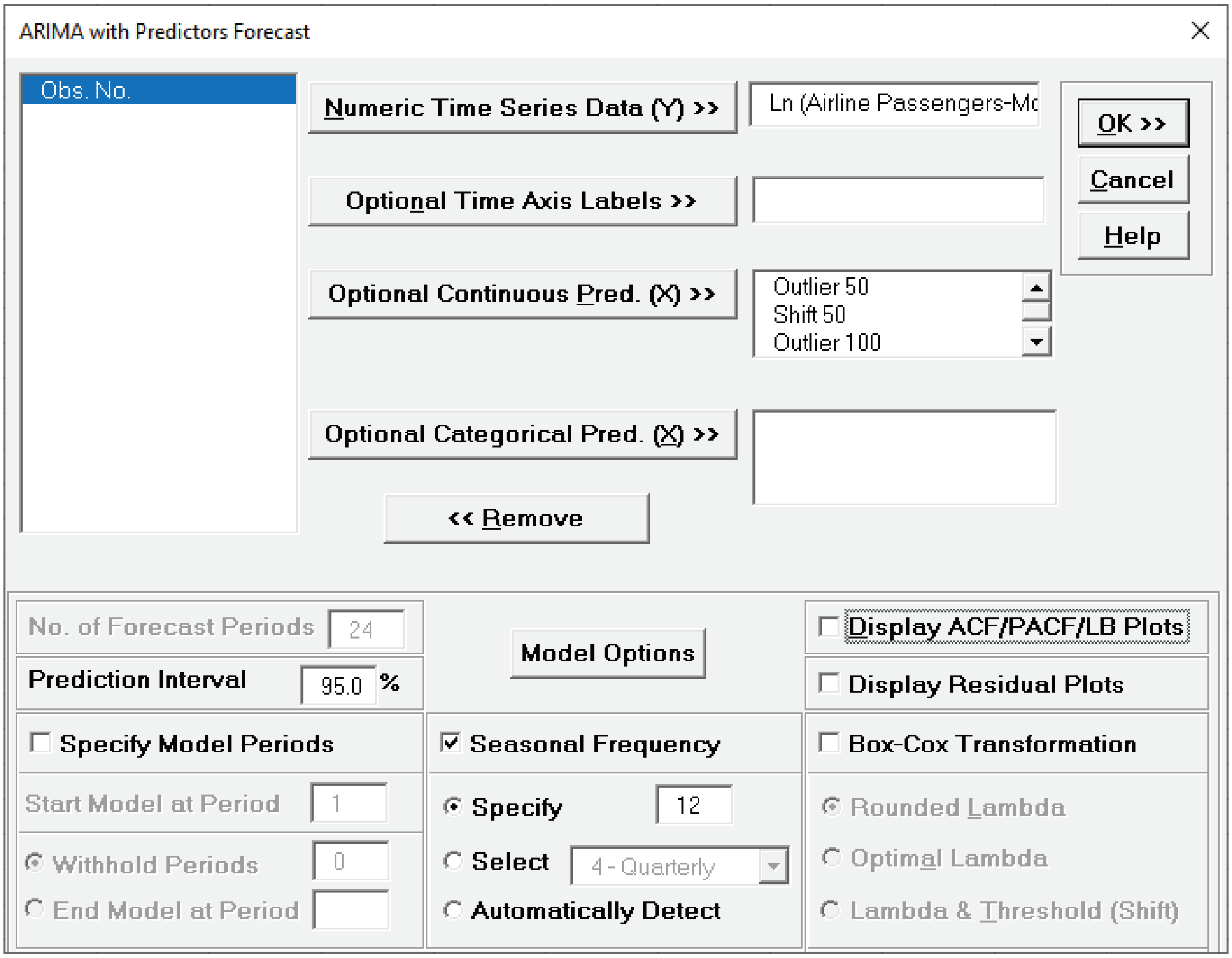
The first out-of-control signal occurred at Obs. No = 50, so two coded variables were created. *Outlier50* is coded so that Obs. No. 50 = 1 and all other values are 0. *Shift50* is coded so that Obs. No. 1 to 49 = 0, and Obs. No. 50 to 144 are 1.

To simplify the analysis, we will assume that the out-of-control signals at 51 and 52 are related to 50, so will ignore those. The next out-of-control signal occurs at 100, so similarly another two coded variables were created. *Outlier100* is coded so that Obs. No. 100 = 1 and all other values are 0. *Shift100* is coded so that Obs. No. 1 to 99 = 0, and Obs. No. 100 to 144 are 1.

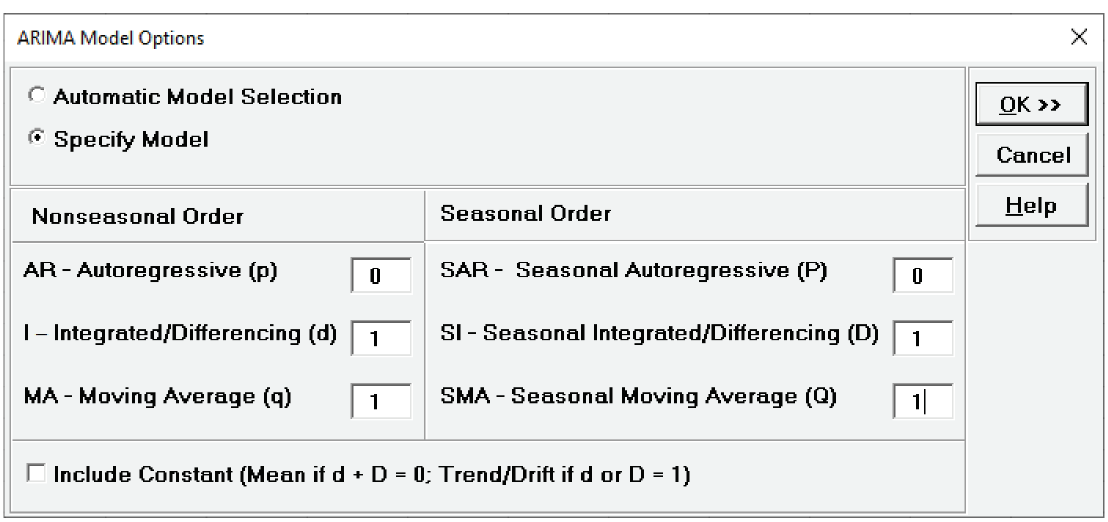
Since these are coded as 0, 1 we will treat them as Continuous.

34. Click **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

35. Select *Ln(Airline Passengers – Modified)*, click **Numeric Time Series Data (Y) >>**; select *Outlier50 to Shift100*, click **Optional Continuous Pred. >> (**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Seasonal Frequency** with **Specify** = 12. Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked. We will use the default **Prediction Interval** = 95.0 %.



36. Click **Model Options**. With the addition of the coded predictors, we want to ensure that the ARIMA model is the same as that used in the control chart. Select **Specify Model**. Specify **Nonseasonal Order I – Integrated/Differencing (d)** = 1 and **MA – Moving Average (q)** = 1. Specify **Seasonal Order SI – Seasonal Integrated/Differencing (D)** = 1 and **SMA – Seasonal Moving Average (Q)** = 1. Leave **Include Constant** unchecked.



37. Click **OK** to return to the ARIMA Forecast dialog. Click **OK**. The ARIMA forecast report is given.

38. Scroll down to view the Parameter Estimates.

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.395854258	0.086277876	4.588132	**0.0000**
SMA_1	0.556926317	0.073521579	7.575005	**0.0000**
Outlier 50	-0.296267312	0.032292885	9.174384	**0.0000**
Shift 50	0.044130277	0.035505653	1.242908	0.2162
Outlier 100	-0.000597974	0.031822469	0.018791	0.9850
Shift 100	0.249993132	0.035384072	7.065132	**0.0000**

Now we can see that *Outlier50* and *Shift100* are significant denoting Obs. No. 50 as an outlier and 100 as a shift. This is, of course, what we expected since that's how the Ln Airline Passenger data was modified.

This method to identify outlier versus shift is intended as a complement to process knowledge and the search for assignable causes used in classical SPC.

ARIMA Control Chart with Predictors

An Individuals control chart is created using the residuals of the ARIMA with Predictors forecast model.

The ARIMA with Predictors model supports continuous or categorical predictors. See **ARIMA Forecast with Predictors** for more information.

The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits.

The popular "Show Last 30" and "Scroll" features in SigmaXL Chart Tools are available for these control charts. Currently, the "Add Data" feature is not supported in ARIMA Control Chart with Predictors.

For further details and references, see the Appendix: **Control Charts for Autocorrelated Data**.

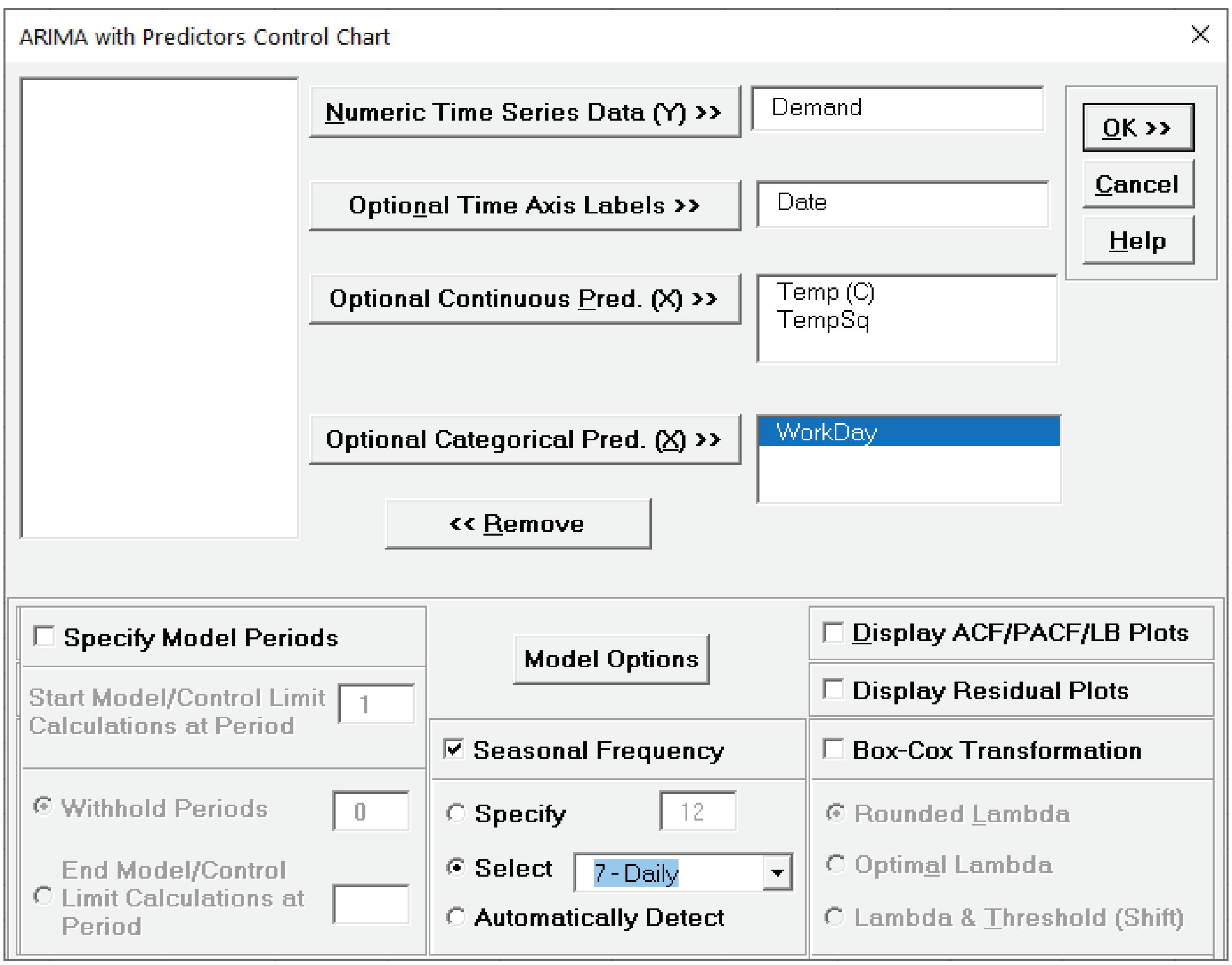
Note that a Moving Range Chart and Tests for Special Causes are not available here, but the user can store and select Residuals, then create with **SigmaXL > Control Charts > Individuals & Moving Range**.

Daily Electricity Demand with Predictors – ElecDaily

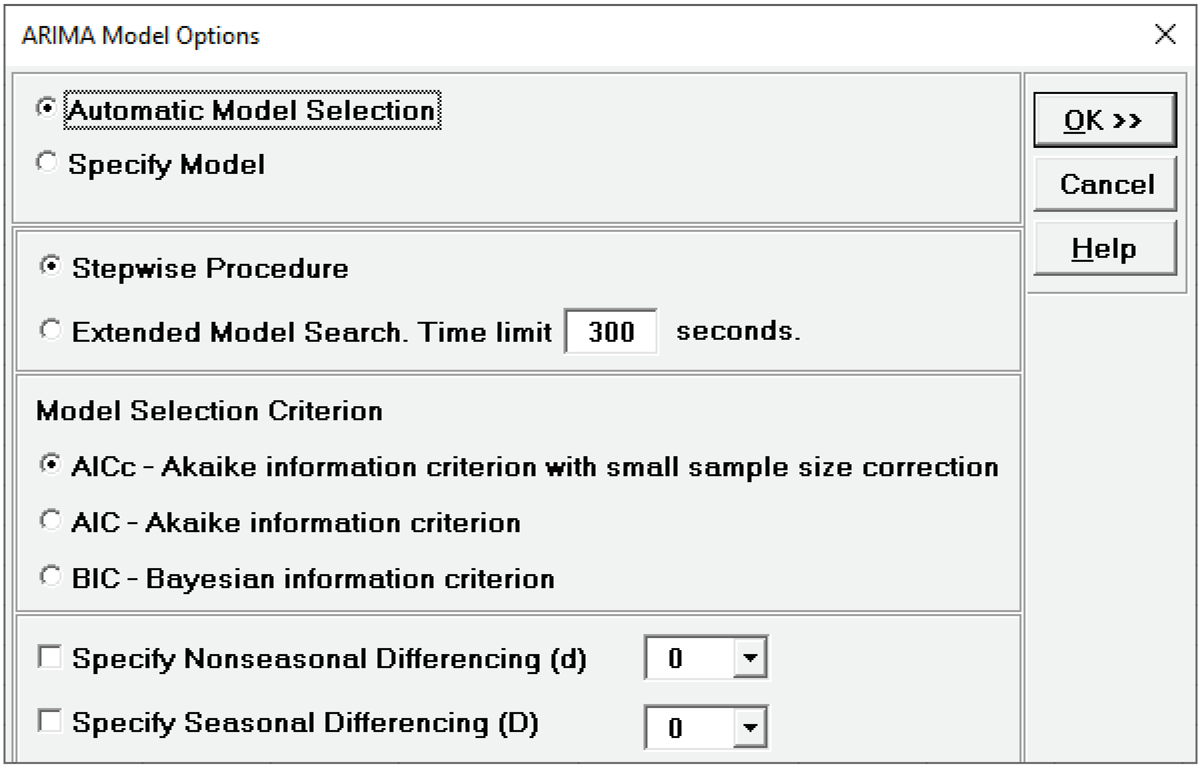
1. Open **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (**Sheet 1** tab). This is daily electricity demand (GW) for the state of Victoria, Australia, every day during 2014 (see Hyndman, fpp2, Section 9.3, **Example: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html). Temp (C) is the maximum daily temperature in degrees Celsius for the city of Melbourne. TempSq is Temperature squared. WorkDay takes on the value 1 on work days and 0 otherwise. This data has a seasonal frequency = 7 (observations per week). See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

2. Click **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart with Predictors**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

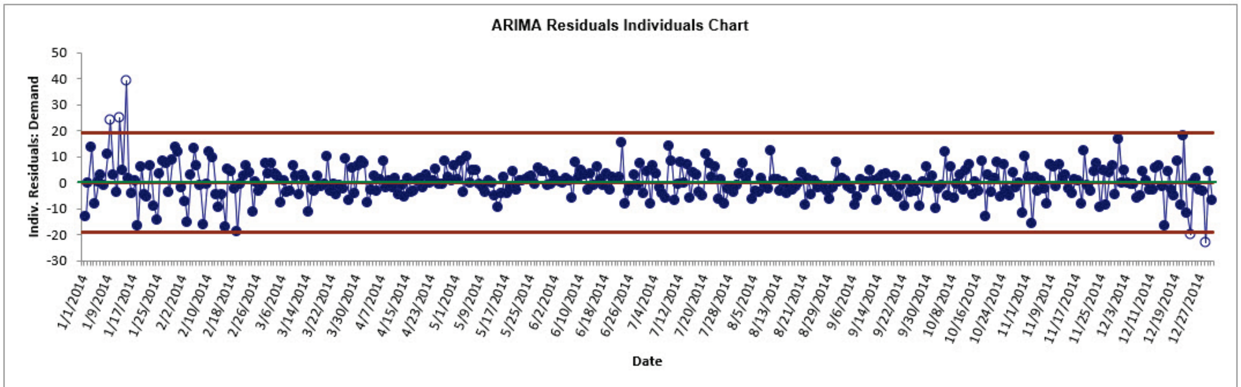
3. Select *Demand*, click **Numeric Time Series Data (Y) >>**; select *Date*, click **Optional X-Axis Labels >>**; select *Temp (C)* and *TempSq*, click **Optional Continuous Pred. >>**; select *WorkDay*, click **Optional Categorical Pred. >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Seasonal Frequency** with **Select** = 7 - Daily (or **Specify** = 7). Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.

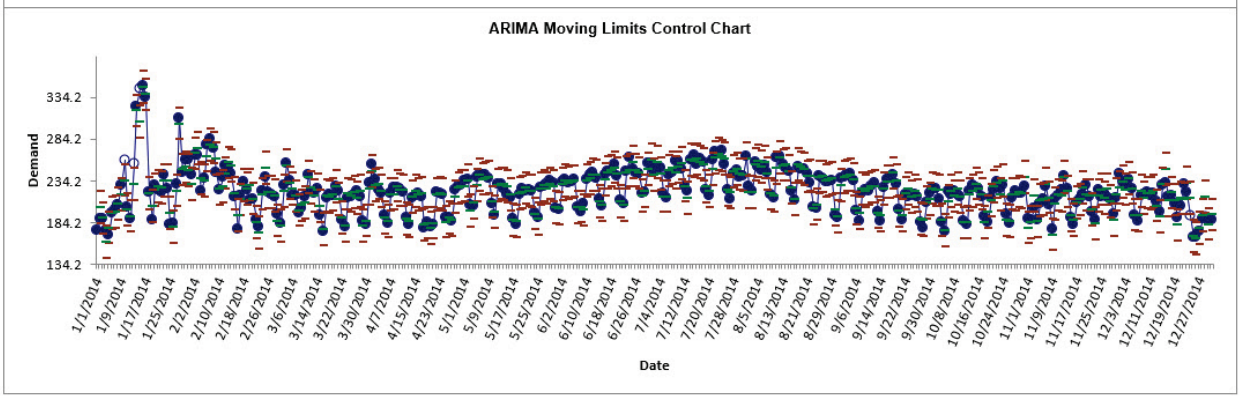


4. Click **Model Options**. Select **Automatic Model Selection**. We will use the defaults: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, leave **Specify Nonseasonal Differencing (d)** and **Specify Seasonal Differencing (D)** unchecked.



5. Click **OK** to return to the ARIMA with Predictors Control Chart dialog. Click **OK**. This is a complex model, so computation time will be approximately one to two minutes. The ARIMA with Predictors control charts are produced:





We can see out of control signals in mid-January with higher demand than the model predicted and late December with lower demand than the model predicted.

6. Scroll down to view the ARIMA Model header:

ARIMA Model: Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

7. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	2
SAR Order (P)	2
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	7
Include Constant	0
No. of Predictors	3
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information: ARIMA (2,1,2) (2,0,0) with no constant and 3 predictors. Seasonal Frequency = 7; Model Selection Criterion = “AICc” and Box-Cox Transformation = “N/A” because Box-Cox Transformation was unchecked.

This is the same model used previously in ARIMA MSD Forecast, and agrees with the model given in Hyndman, fpp2, Section 9.3, **Example: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html.

ARIMA Multiple Seasonal Decomposition (MSD) Control Chart

ARIMA does not have a theoretical frequency limit, but for computational efficiency and to minimize the potential loss of observations through differencing, we recommend using ARIMA – Multiple Seasonal Decomposition (MSD) for seasonal frequency greater than 52 (or with multiple frequencies). The seasonal component is first removed through decomposition, a nonseasonal ARIMA model fitted to the remainder (+trend), and then the seasonal component is added back in.

As the name implies, Multiple Seasonal Decomposition (MSD) also accommodates multiple seasonality, for example the half-hourly data with a seasonal frequency of 48 observations per day and 336 observations per week.

An Individuals control chart of the residuals is created for this forecast method. The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits.

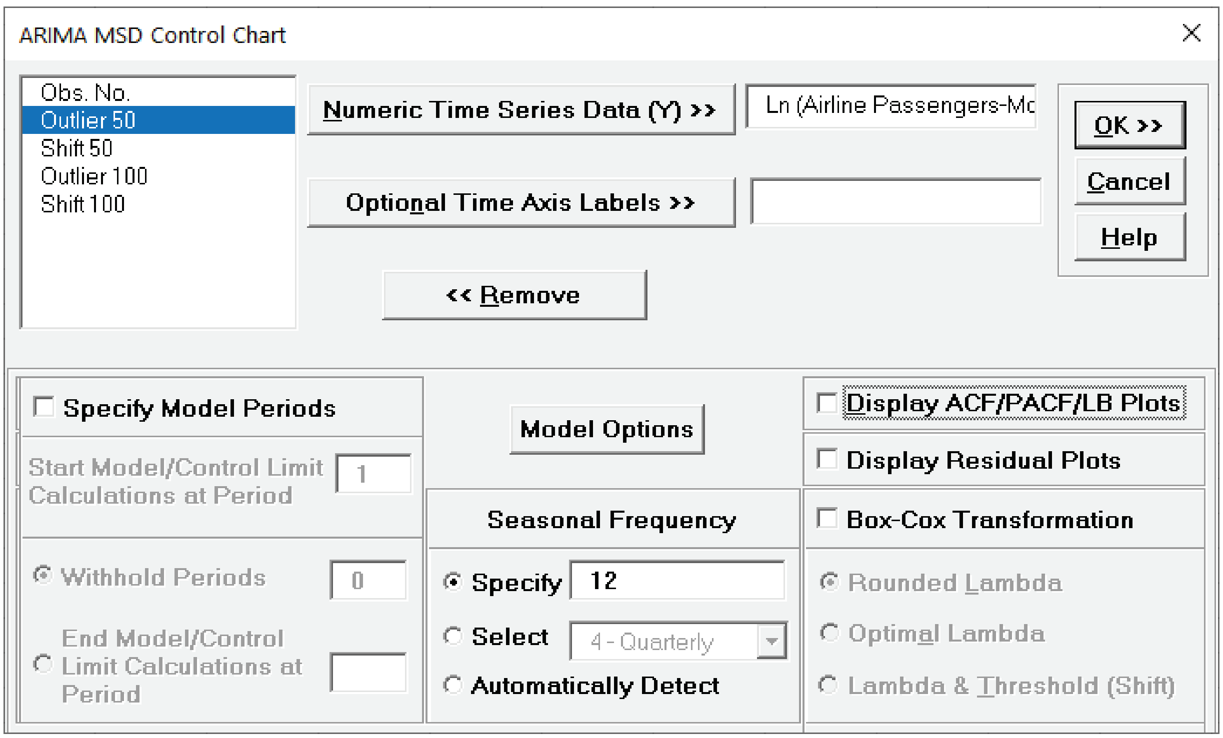
The popular “Add Data”, “Show Last 30” and “Scroll” features in SigmaXL Chart Tools are available for these control charts. For “Add Data”, the time series models are not refitted, but used to compute the residual values for the new data.

Monthly Airline Passengers – Modified for Control Charts

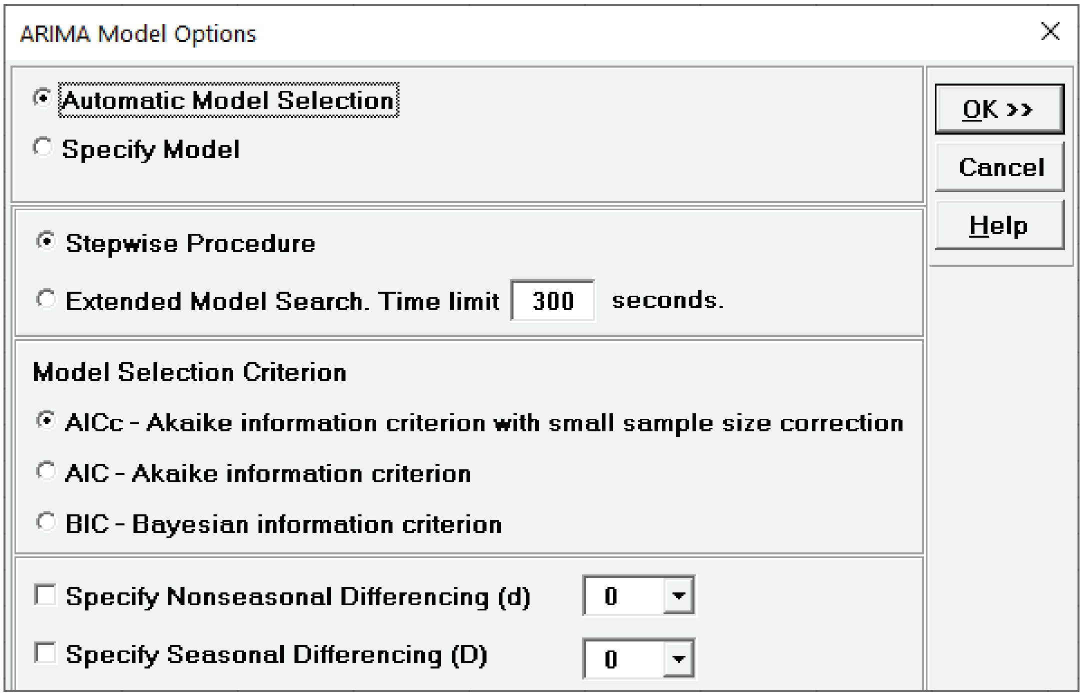
1. Open **Monthly Airline Passengers – Modified for Control Charts.xlsx** (**Sheet 1** tab). This is based on the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960. A Ln transformation is applied (avoiding the need for a Box-Cox transformation), a negative outlier is added at 50 (-.25) and a level shift applied (+.25), starting at 100. The Multiple Seasonal Decomposition (MSD) option is not necessary for this data, but by way of introduction, we will use this to compare to the previous analysis.

2. Click **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Multiple Seasonal Decomposition Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

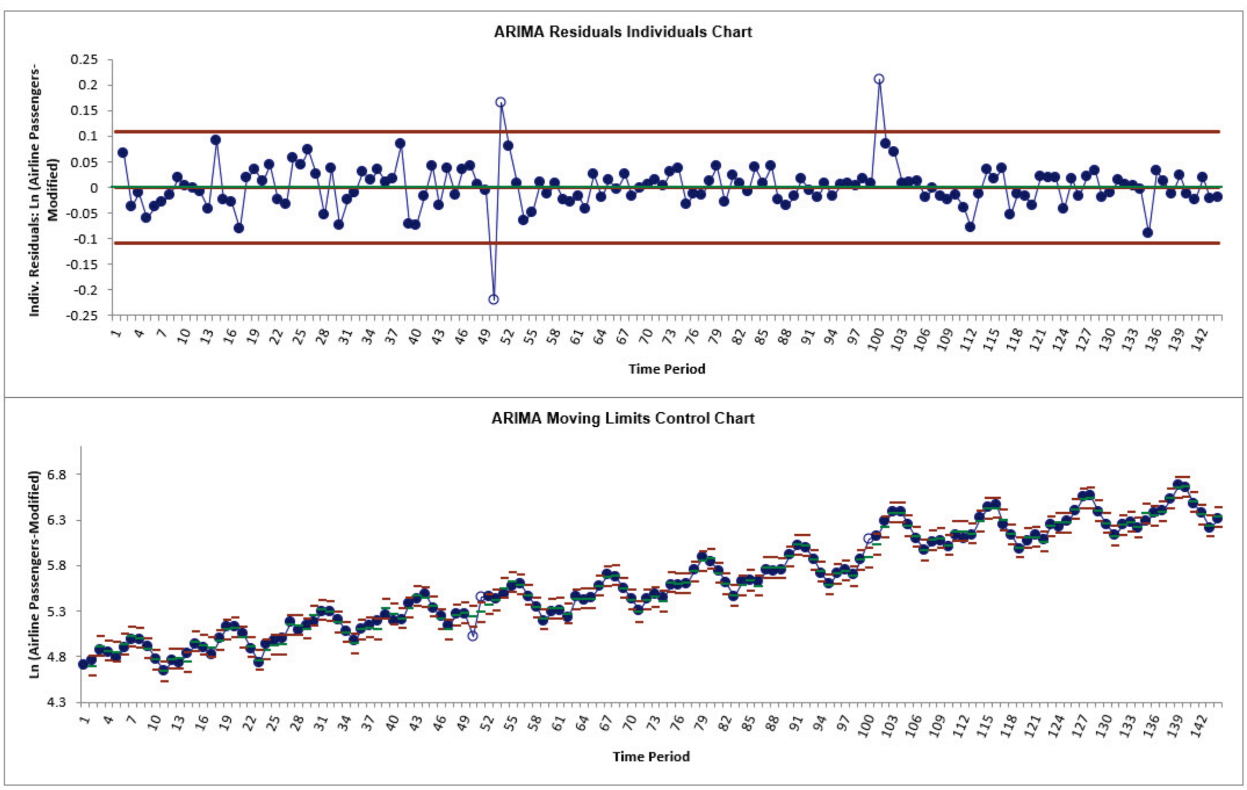
3. Select *Ln(Airline Passengers-Modified)*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots.** Check **Seasonal Frequency** with **Specify** = 12**.** Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



4. Click **Model Options**.



5. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the ARIMA Control Chart dialog. Click **OK**. The ARIMA (MSD) control charts are produced:



We can clearly see the out-of-control data points at 50, 51 and 100 on the Residuals Individuals chart. This is similar to what we observed previously with regular ARIMA Control Charts.

6. Scroll down to view the ARIMA MSD Model header:

ARIMA Model (Multiple Seasonal Decomposition): Ln (Airline Passengers-Modified) - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

7. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	12 Decomposition
Include Constant	1
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information for the deseasonalized data: ARIMA (0,1,1) with a constant. Seasonal Frequency = 12 using Decomposition and Model Selection Criterion = “AICc”. There are no seasonal terms in the model. The Box-Cox Transformation is “N/A”.

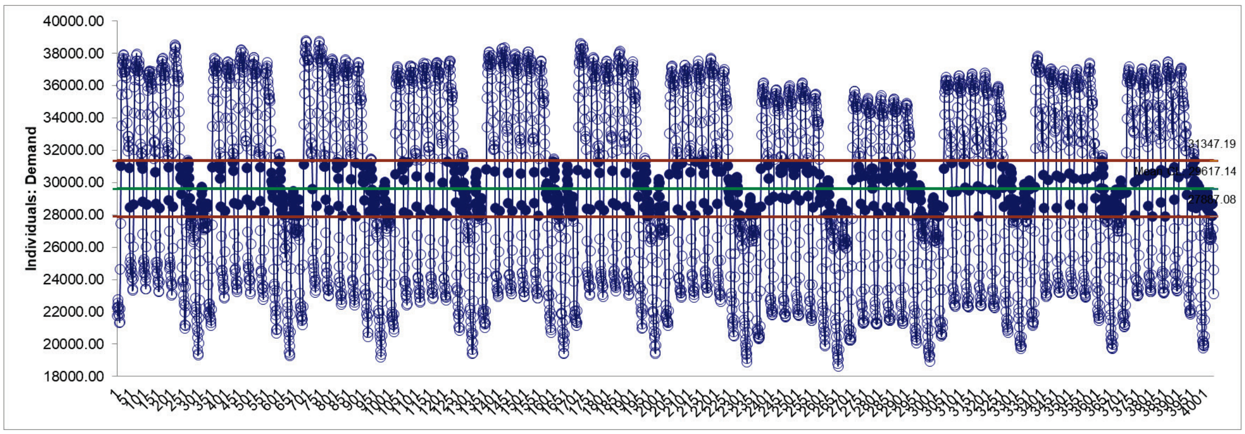
8. We will not review the Parameter Estimates, Model Statistics and Forecast Accuracy as they are close to the ARIMA MSD values given earlier, although note that slight differences are due to the introduction of an outlier and a shift, as well now we are using all of the data, i.e., there are no withhold periods. Earlier we used a Box-Cox Transformation with Lambda=0 and here we are using Ln of the data.

Half-Hourly Multiple Seasonal Electricity Demand – Taylor

9. Open **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (**Sheet 1** tab). This is half-hourly electricity demand (MW) in England and Wales from Monday, June 5, 2000 to Sunday, August 27, 2000 (taylor, R forecast). This data has multiple seasonality with frequency = 48 (observations per day) and 336 (observations per week), with a total of 4032 observations. See the Run Chart, ACF/PACF Plots, Spectral Density Plot and Seasonal Trend Decomposition Plots for this data.

10. We will first construct a classical Individuals Control Chart on the raw data. Click **SigmaXL > Control Charts > Individuals**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

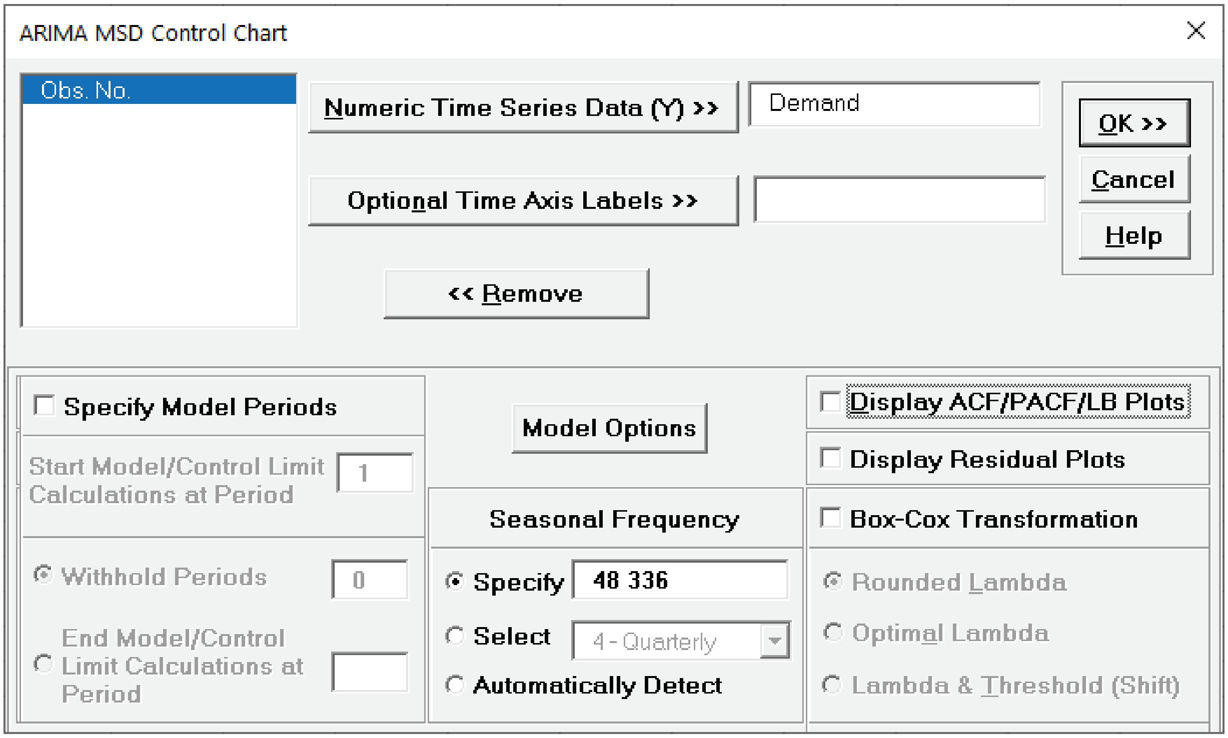
11. Select *Demand*, click **Numeric Data Variable (Y) >>**. Click **OK**. An Individuals Control Chart is produced:



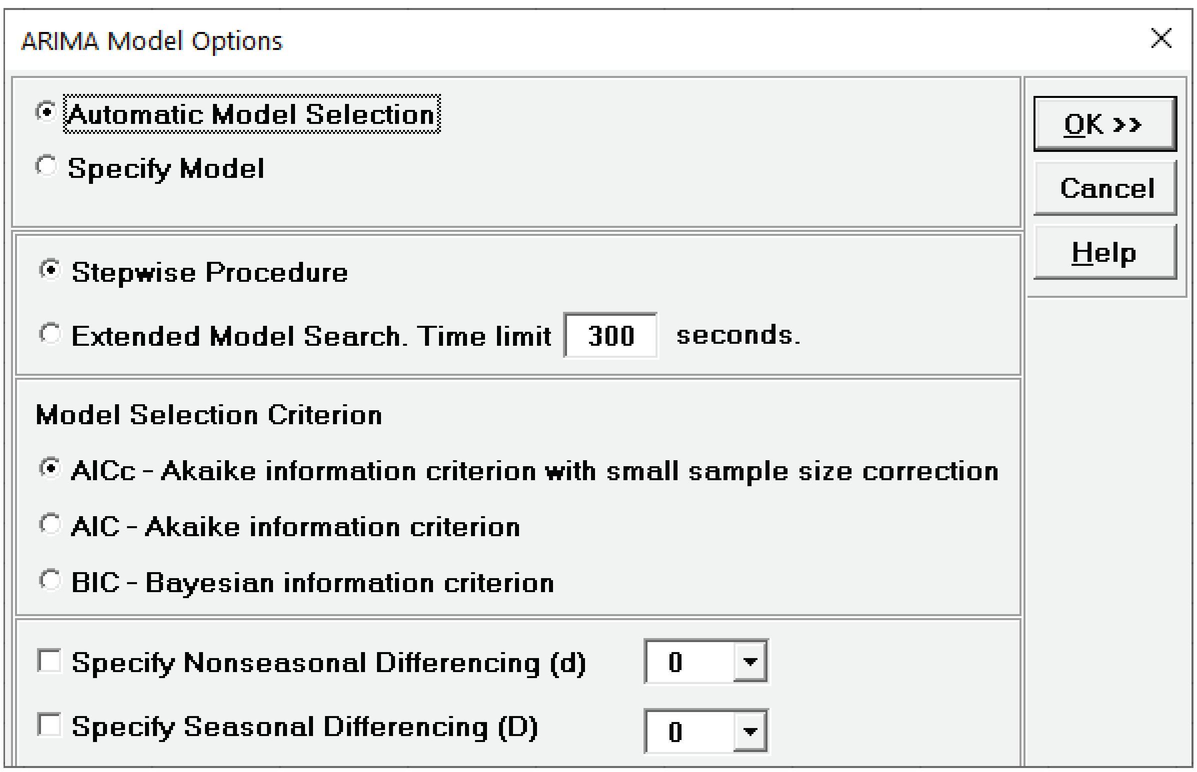
With the high frequency seasonality, this control chart is meaningless.

12. Click **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Multiple Seasonal Decomposition Control Chart.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

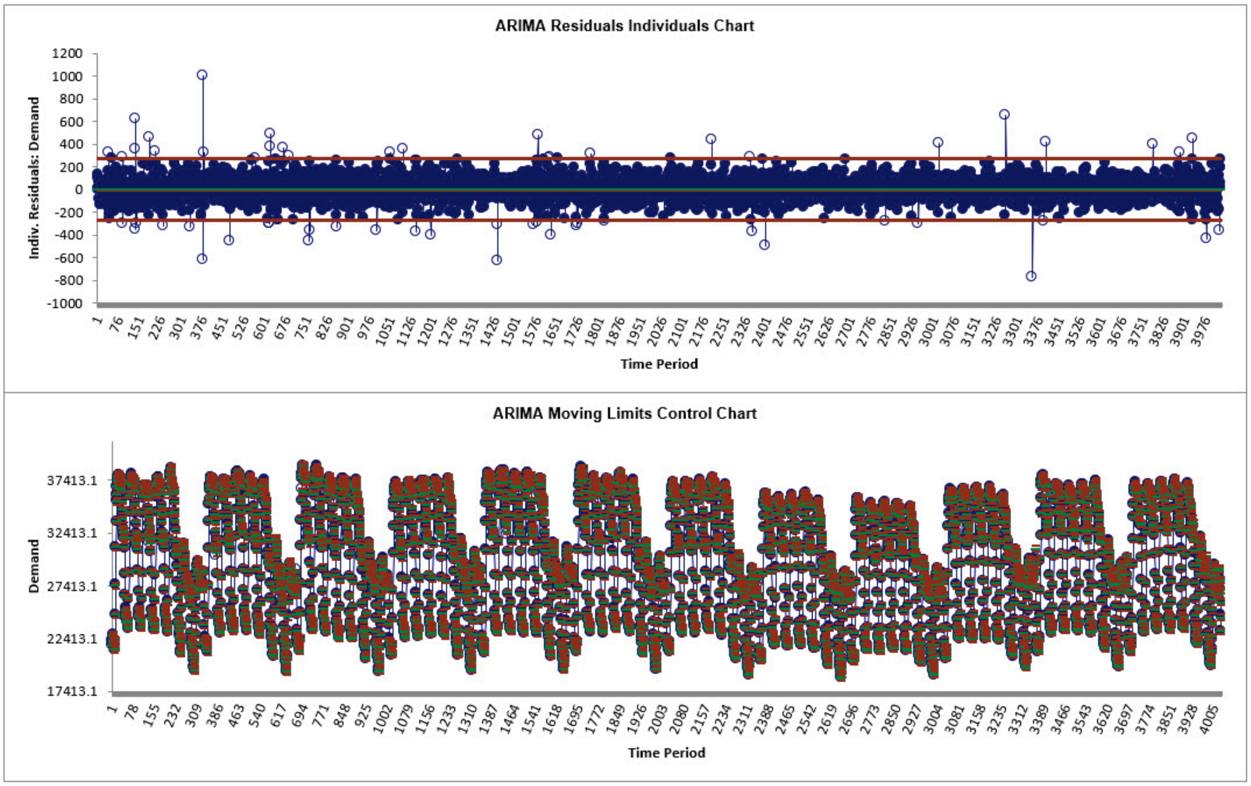
13. Select *Demand*, click **Numeric Time Series Data (Y) >>**. Uncheck **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Check **Seasonal Frequency** with **Specify** = 48 336**.** Leave **Specify Model Periods** and **Box-Cox Transformation** unchecked.



14. Click **Model Options**.



15. We will use the default **Automatic Model Selection** with **AICc** as the **Model Selection Criterion**. Click **OK** to return to the ARIMA MSD Control Chart dialog. Click **OK**. The ARIMA MSD control charts are produced:



Now we have a chart that can be used to identify assignable causes. The number of out-of-control signals have been dramatically reduced. This is similar to what we observed previously with Exponential Smoothing MSD Control Charts.

16. Scroll down to view the ARIMA MSD Model header:

ARIMA Model (Multiple Seasonal Decomposition): Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

17. The ARIMA Model Summary is given as:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	0
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	48, 336 Decomposition
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

This is a summary of the model information for the deseasonalized data: ARIMA (2,1,0) without a constant. Seasonal Frequency = 48, 336 using Decomposition and Model Selection Criterion = “AICc”. There are no seasonal terms in the model. This is the same model used previously in ARIMA MSD Forecast, but that used a withhold sample of 96.

18. We will not review the Parameter Estimates, Model Statistics and Forecast Accuracy as they are close to the ARIMA MSD values given earlier, although note that now we are using all of the data, i.e., there are no withhold periods.

Utilities – Difference Data

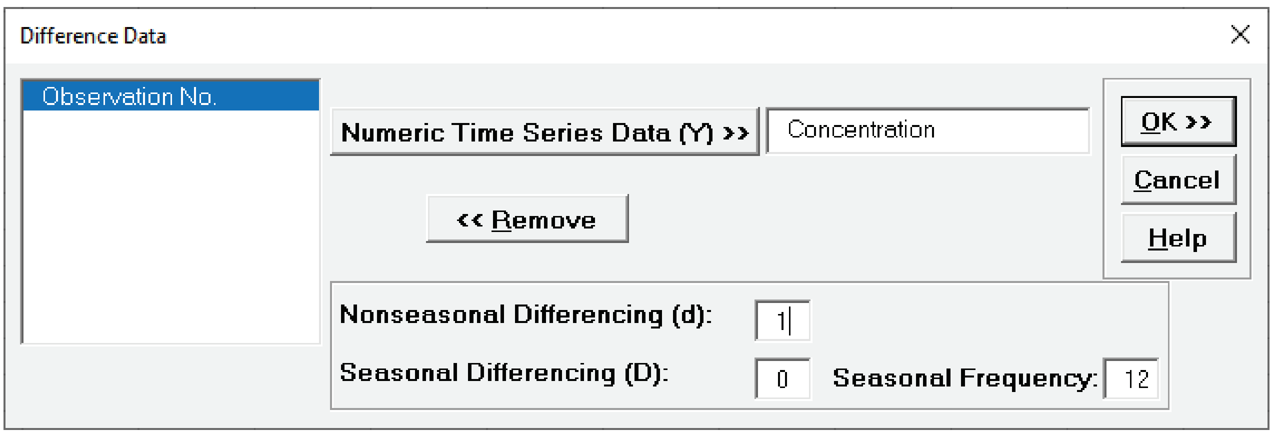
Nonseasonal and Seasonal Differencing are used to make a process stationary. This is done automatically in ARIMA, but this utility makes it easy to do manually. This was demonstrated previously in the section Run Chart.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

2. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Enter 1 for **Nonseasonal Differencing (d)**.



4. Click **OK**. A new sheet is created with the order d = 1 differenced data (Y_2 - Y_1, Y_3-Y_2, …).

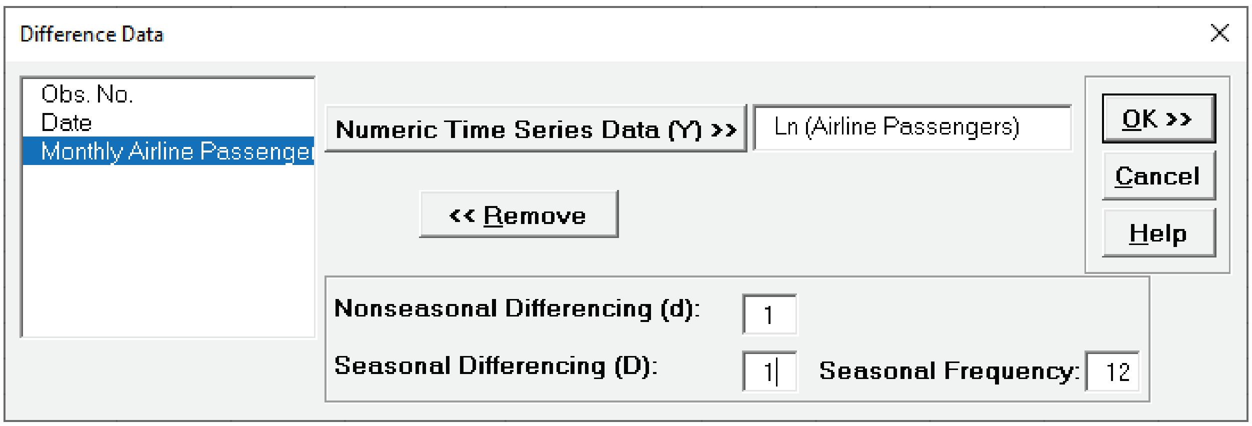
Concentration	Concentration: d = 1, D = 0
17	
16.6	-0.4
16.3	-0.3
16.1	-0.2
17.1	1
16.9	-0.2
16.8	-0.1
17.4	0.6
17.1	-0.3
17	-0.1
16.7	-0.3
17.4	0.7

Monthly Airline Passengers - Series G

5. Open **Monthly Airline Passengers - Series G.xlsx** (**Sheet 1** tab). This is the Series G data from Box and Jenkins, monthly total international airline passengers for January 1949 to December 1960 and is one of the most popular datasets used in introductory time series forecasting.

6. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

7. Select *Ln(Airline Passengers)*, click **Numeric Time Series Data (Y) >>**. Enter 1 for **Nonseasonal Differencing (d)**; enter 1 for **Seasonal Differencing (D)**; **Seasonal Frequency** is specified as 12.



8. Click **OK**. A new sheet is created with the seasonal order D = 1 differenced data (Y_{13} - Y_1, Y_{14}-Y_2, ...) and nonseasonal order d=1 ($D_2 – D_1$, $D_3 – D_2$, ...).

Ln (Airline Passengers)	Ln (Airline Passengers): d = 1, D = 1
4.718498871	
4.770684624	
4.882801923	
4.859812404	
4.795790546	
4.905274778	
4.997212274	
4.997212274	
4.912654886	
4.779123493	
4.644390899	
4.770684624	
4.744932128	
4.836281907	0.039164025
4.94875989	0.000360685
4.905274778	-0.020495594
4.828313737	-0.012939182

Utilities – Lag Data

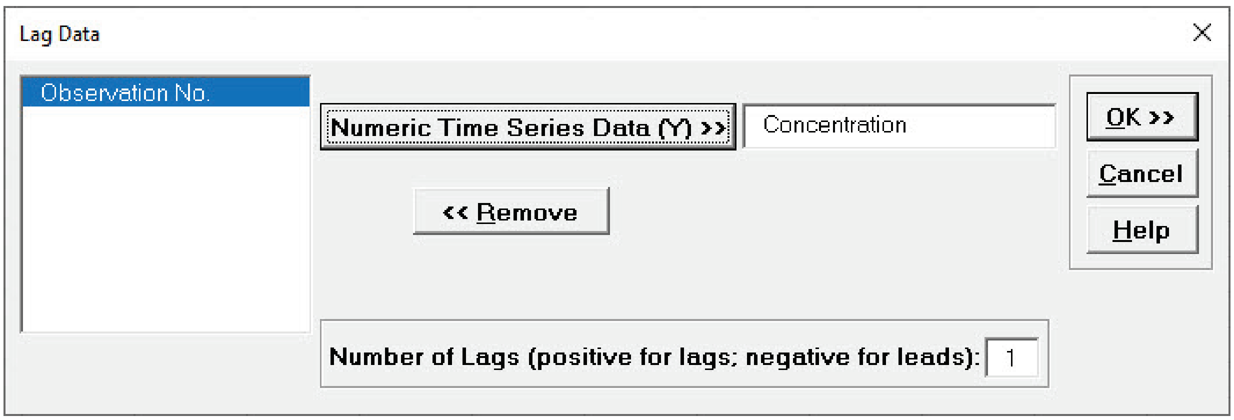
The Lag Data utility is a useful utility to create Lag (or Lead) columns of data. This was demonstrated previously in the section Autocorrelation (ACF/PACF) Plots.

Chemical Process Concentration - Series A

1. Open **Chemical Process Concentration – Series A.xlsx** (**Sheet 1** tab). This is the Series A data from Box and Jenkins, a set of 197 concentration values from a chemical process taken at two-hour intervals.

2. Click **SigmaXL > Time Series Forecasting > Utilities > Lag Data**. Ensure that the entire data table is selected. If not, check Use Entire Data Table. Click Next.

3. Select *Concentration*, click **Numeric Time Series Data (Y) >>**. Use the default **Number of Lags** = 1.



4. Click **OK**. A new sheet is created with the Lag 1 data.

Concentration	Concentration: Lag = 1
17	
16.6	17
16.3	16.6
16.1	16.3
17.1	16.1
16.9	17.1

5. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Enter **Lag** = 2. Click **OK**. Repeat for **Lag** = 3. The Lag columns are appended as shown:

Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
17			
16.6	17		
16.3	16.6	17	
16.1	16.3	16.6	17
17.1	16.1	16.3	16.6
16.9	17.1	16.1	16.3
16.8	16.9	17.1	16.1

Utilities – Interpolate Missing Values

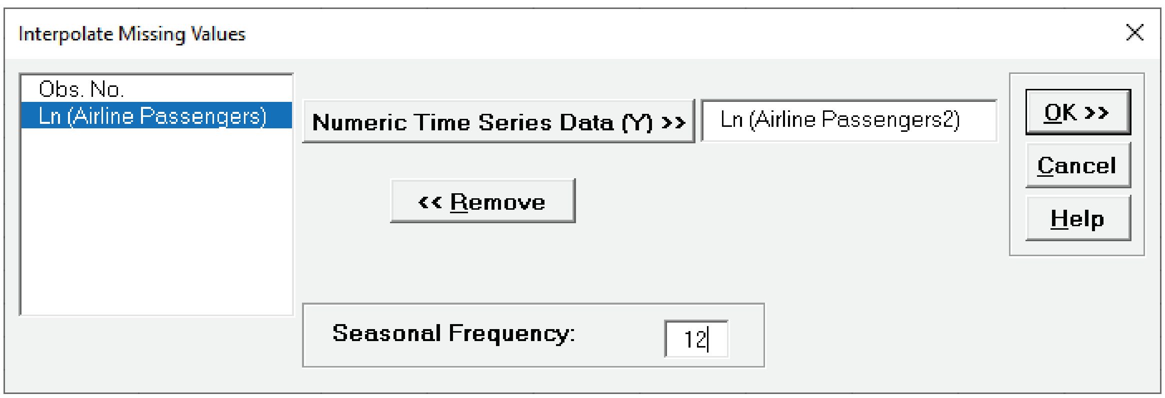
The Interpolate Missing Values utility was demonstrated previously. See the Appendix: **Seasonally Adjusted Linear Interpolation of Missing Values** for further details. It is used automatically in Exponential Smoothing when there are missing values. While there is robustness against some missing values, if the number of missing values is large, then model estimation and forecast accuracy will be degraded

Monthly Airline Passengers – Missing Values

1. Open **Monthly Airline Passengers – Missing Values.xlsx** (**Sheet 1** tab). Ln(Airline Passengers2) have missing values at observations 50 and 100.

2. Click the **Sheet 1** tab. Click **SigmaXL > Time Series Forecasting > Utilities > Interpolate Missing Values.** Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

3. Select *Ln(Airline Passengers2)*, click **Numeric Time Series Data (Y) >>**. **Seasonal Frequency** is specified as 12.



4. Click **OK**. A new sheet is created with the seasonally adjusted linear interpolated values highlighted in yellow.

Ln (Airline Passengers2)	Y: Interpolated (12)
4.718498871	4.718498871
4.770684624	4.770684624
4.882801923	4.882801923
4.859812404	4.859812404
	5.254047791
	5.846836235

Y50 is estimated as 5.254 (versus original value of 5.278). Y100 is estimated as 5.847 (versus original value of 5.852).

SigmaXL Appendix

SigmaXL Version 10.0 Feature List Summary

**** denotes new features in SigmaXL 10.0**

Compatible with Excel 2021 for Windows and Mac

Menu Layout: Classical or DMAIC

Recall Last Dialog

Data Manipulation:

- Subset by Category, Number, Date or Random
- Stack Subgroups Across Rows
- Stack and Unstack Columns
- Random Number Generator
    - Normal
    - Uniform (Continuous & Integer)
    - Lognormal
    - Weibull
    - Triangular
- Box-Cox Transformation
- Standardize Data
- Convert to Discrete
- Convert Raw Data to Frequency (Tally)**
- Convert Frequency to Raw Data**

Templates & Calculators:

- DMAIC & DFSS Templates
  - Team/Project Charter
  - SIPOC Diagram
  - Flowchart Toolbar
  - Data Measurement Plan
  - Cause & Effect (Fishbone) Diagram and Quick Template
  - Cause & Effect (XY) Matrix with Pareto
  - Failure Mode & Effects Analysis (FMEA) with RPN Sort
  - Quality Function Deployment (QFD)
  - Pugh Concept Selection Matrix
  - Control Plan
- Lean Templates
  - Takt Time Calculator
  - Value Analysis/Process Load Balance
  - Value Stream Mapping
- Graphical Templates
  - Pareto Chart, Histogram, Run Chart
- Statistical Templates
  - Sample Size – Discrete and Continuous
  - Minimum Sample Size for Robust t-Tests and ANOVA
  - 1 Sample Z-test & Confidence Interval for Mean
  - 1 Sample t-Test & Confidence Interval for Mean
  - 2 Sample t-Test & Confidence Interval (Compare 2 Means)
  - 1 Sample Equivalence Test for Mean
  - 2 Sample Equivalence Test (Compare 2 Means)
  - 1 Sample Chi-Square Test and CI for Standard Deviation
  - 2 Sample F-Test and CI (Compare 2 Standard Deviations)
  - 1 Proportion Test and Confidence Interval
  - 2 Proportions Test and Confidence Interval (with Fisher Exact P-Value)
  - 2 Proportions Equivalence Test
  - 1 Poisson Rate Test and Confidence Interval
  - 2 Poisson Rates Test and Confidence Interval
  - 2 Poisson Rates Equivalence Test
  - One-Way Chi-Square Goodness-of-Fit Test (with Exact and Monte Carlo P-Value)
- Probability Distribution Calculators
  - Normal, Lognormal, Exponential, Weibull
  - Binomial, Poisson, Hypergeometric
- Measurement Systems Analysis (MSA) Templates
  - Type 1 Gage Study
  - Gage Bias and Linearity Study
  - Gage R&R Study – with Multi-Vari Analysis
  - Attribute Gage R&R (Attribute Agreement Analysis)
  - GLM GageRR (Crossed) Metrics with/without Interaction; GageRR (Nested) Metrics; GLM GageRR (Expanded) Metrics**
  - Orthogonal (Deming) Regression**
- Process Sigma Level – Discrete and Continuous
- Process Capability & Confidence Intervals
- Tolerance Interval Calculator (Normal Exact)
- DOE Templates
  - 2 to 5 Factors
  - 2-Level Full and Fractional-Factorial designs
  - Main Effects & Interaction Plots
- Taguchi DOE Templates
  - L4, L8, L9, L12, L16, L18, L27
  - Signal-to-Noise Ratios: Nominal is Best, Nominal is Best (Variance Only), Nominal is Best (Mean Square Deviation with Target), Larger is Better, Smaller is Better
  - Pareto of Deltas (Effects) and ANOVA SS (Sum-of-Squares) % Contribution (for Main Effects and Two-Way Interactions)
  - Main Effects & Interaction Plots
- Control Chart Templates
  - Individuals, C (Count)
  - Rare Events: T, G and Probability Based G
  - Exponentially Weighted Moving Average (EWMA)
  - Tabular Cumulative Sum (CUSUM)
  - Trend/Tool Wear
  - Average Run Length (ARL) Calculators: Shewhart with Tests for Special Causes, Attribute C & P, EWMA & CUSUM

Graphical Tools:

- Graphical Tool Selection Guide**
- Basic and Advanced (Multiple) Pareto Charts
- EZ-Pivot/Pivot Charts: Easily create Pivot Tables and Charts
- Heatmap**
- Basic Histogram
- Multiple Histograms and Descriptive Statistics (includes Confidence Interval for Mean and StDev., and Anderson-Darling Normality Test)
- Multiple Histograms and Process Capability (Pp, Ppk, Cpm, ppm, %)
- Multiple Dotplots, Boxplots & Multiple X Boxplots
- Interval Plots and Multiple X Interval Plots**
- Multiple Normal Probability Plots (with 95% confidence intervals to ease interpretation of normality/non-normality)
- Empirical/Normal CDF Plots**
- Run Charts (with Nonparametric Runs Test allowing you to test for Clustering, Mixtures, Lack of Randomness, Trends and Oscillation)
- Overlay Run Charts
- Multi-Vari Charts
- Scatter Plots (with linear regression and optional 95% confidence intervals and prediction intervals)
- Scatter Plot Matrix
- XYZ Contour/Surface Plot**
    - Automatic Interpolation Method Selection and XY Standardization using Cross-Validation
    - Inverse Distance, Akima's Polynomial and Biharmonic Spline Interpolation
- Analysis of Means (ANOM) Charts
    - All charts support balanced & unbalanced data
    - Normal, Binomial Proportions and Poisson Rates: One-Way & Two-Way with Main Effects and Slice Charts
    - Slice Charts are a modified ANOM chart developed by Dr. Peter Wludyka that enables one to easily interpret the effects in the presence of an interaction. In collaboration, Peter Wludyka and John Noguera of SigmaXL extended the Slice Charts to Binomial and Poisson. Yellow highlight automatically recommends Main Effects (if interaction is not significant) or Slice Chart (if interaction is significant).
    - Nonparametric Transformed Ranks
    - Variances & Levene Robust Variances

Measurement Systems Analysis:

- Create Gage R&R (Crossed) Worksheet:
  - Generate worksheet with user specified number of parts, operators, replicates
- Analyze Gage R&R (Crossed)
  - ANOVA, %Total, %Tolerance (with upper and/or lower specifications), %Process, Variance Components, Number of Distinct Categories
  - Gage R&R Multi-Vari and X-bar R Charts
  - Confidence Intervals for %Total, %Tolerance, %Process and Standard Deviations
  - Handles unbalanced data
- Attribute MSA (Binary, Ordinal, Nominal)
  - Any number of samples, appraisers and replicates
  - Within Appraiser Agreement, Each Appraiser vs Standard Agreement, Each Appraiser vs Standard Disagreement, Between Appraiser Agreement, All Appraisers vs Standard Agreement; Fleiss' Kappa; Kendall's Coefficients (Ordinal Concordance and Correlation)
  - Confidence Interval Graphs for Percent Agreement, Fleiss' Kappa and Kendall's Coefficients
  - Kappa and Kendall color highlight to aid interpretation (Green/Yellow/Red)
  - Effectiveness Report (treats each appraisal trial as an opportunity) and Misclassification Summary

Process Capability:

- Multiple Histograms and Process Capability
- Capability Combination Report for Individuals/Subgroups:
  - Histogram, Normal Probability Plot and Normality Test
  - Capability Report (Cp, Cpk, Pp, Ppk, Cpm, ppm, %)
  - Control Charts
- Capability Combination Report for Nonnormal Data (Individuals)
  - Box-Cox Transformation (includes an automatic threshold option so that data with negative values can be transformed)
  - Johnson Transformation
  - Distributions supported: Half-Normal, Lognormal (2 & 3 parameter), Exponential (1 & 2), Weibull (2 & 3), Beta (2 & 4), Gamma (2 & 3), Logistic, Loglogistic (2 & 3), Largest Extreme Value, Smallest Extreme Value
  - Automatic Best Fit based on AD p-value
  - Nonnormal Process Capability Indices: Z-Score (Cp, Cpk, Pp, Ppk) and Percentile (ISO) Method (Pp, Ppk)
- Distribution Fitting Report
  - All valid distributions and transformations reported with histograms, curve fit and probability plots
  - Sorted by AD p-value

Statistical Tools:

- P-values turn red when results are significant: (p-value < alpha)
- Descriptive Statistics including Anderson-Darling Normality test, Skewness and Kurtosis with p-values
- Descriptive Statistics options:
  - Percentile Report and Percentile Ranges
  - Percentile Confidence and Tolerance Intervals
  - Additional Descriptive Statistics and Normality Tests (Shapiro-Wilk and Doornik-Hansen Normality)
  - Outlier (Boxplot and Grubbs) and Randomness (Nonparametric Runs) Tests
- Hypothesis Test Selection Guide**
- 1 Sample t-test and confidence intervals
- Paired t-test, 2 Sample t-test
- 2 Sample comparison tests:
  - Reports AD Normality, F-test and Levene's for variance, t-test assuming equal and unequal variance, Mann-Whitney test for medians.
  - Recommended tests are highlighted based on sample size, normality, and variance
- One-Way ANOVA and Means Matrix
  - Multiple Comparison of Means Probability Methods (Post-Hoc): Fisher, Tukey, Dunnett with Control
- Automatic Assumptions Check for One Sample, Two-Sample, Paired T-tests and One-Way ANOVA
  - Text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation)
  - Test each sample for Normality. If not, check minimum sample size for robustness of test
  - Check each sample for Outliers: Potential (Tukey's Boxplot 1.5*IQR); Likely (2.2*IQR); Extreme (3*IQR)
  - Randomness (Nonparametric Runs Test)
  - Equal Variance (for 2 or more samples)
- Two-Way ANOVA (Balanced and Unbalanced)
- Equal Variance Tests (Bartlett, Levene and Welch's ANOVA)
  - Multiple Comparison of Variances Probability Methods (Post-Hoc): F-Test (with Bonferroni Correction), Levene, Tukey ADM (Absolute Deviations from Median)
  - Welch Multiple Comparison of Means Probability Methods (Post-Hoc): Welch Pairwise, Games-Howell
- Correlation Matrix (Pearson and Spearman's Rank Correlation)
  - Automatic normality check for correlation utilizing the powerful Doornik-Hansen bivariate normality test
  - Yellow highlight to recommend significant Pearson or Spearman correlation - Pearson is highlighted if the data are bivariate normal, otherwise Spearman is highlighted
- Multiple Linear Regression:
  - Accepts continuous and/or categorical (discrete) predictors
  - Interactive Predicted Response Calculator with 95% Confidence Interval and 95% Prediction Interval
  - Residual Plots: histogram, normal probability plot, residuals vs. time, residuals vs. predicted and residuals vs. X factors
  - Residual types include Regular, Standardized, Studentized (Deleted t) and Cook's Distance (Influence), Leverage and DFITS
  - Highlight of significant outliers in residuals
  - Durbin-Watson Test for Autocorrelation in Residuals with p-value
  - ANOVA report for categorical predictors
  - Pure Error and Lack-of-Fit report
  - Collinearity Variance Inflation Factor (VIF) and Tolerance report
  - Fit Intercept is optional
- Binary and Ordinal Logistic Regression
  - Powerful and user-friendly logistic regression.
  - Report includes a calculator to predict the response event probability for a given set of input X values.
  - Categorical (discrete) predictors can be included in the model in addition to continuous predictors.
  - Model summary and goodness of fit tests include Likelihood Ratio Chi-Square, Pseudo R-Square, Pearson Residuals Chi-Square, Deviance Residuals Chi-Square, Observed and Predicted Outcomes – Percent Correctly Predicted.
  - Stored data includes Event Probabilities, Predicted Outcome, Observed-Predicted, Pearson Residuals, Standardized Pearson Residuals, and Deviance Residuals.
- Chi-Square Test (Stacked Column data and Two-Way Table data)
  - With Fisher Exact (utilizing permutations and fast network algorithms) and Monte Carlo P-Values
  - Options: Advanced Tests and Measures of Association for Nominal & Ordinal Categories
- Power and Sample Size Calculators for:
  - 1 and 2 Sample t-Test
  - One-Way ANOVA
  - 1 Proportion Test, 2 Proportions Test
  - The Power and Sample Size Calculators allow you to solve for Power (1 – Beta), Sample Size, or Difference (specify two, solve for the third).
- Power and Sample Size Chart. Quickly create a graph showing the relationship between Power, Sample Size and Difference.

Statistical Tools – Nonparametric Tests:

- 1 Sample Sign and 1 Sample Wilcoxon
- 2 Sample Mann-Whitney
- 2 Sample KS (option in Mann-Whitney)**
- Kruskal-Wallis and Mood's Median Test (with graph of Group Medians and 95% Median Confidence Intervals)
- Friedman Test**
- Runs Test
- With Exact and Monte Carlo P-Value

Statistical Tools - Advanced Multiple Regression:

- Standardization and coding of continuous predictors
- Option to display regression equation with unstandardized coefficients
- (1, 0) or (-1,0,+1) coding of categorical predictors
- Box-Cox Transformation
- Specify confidence level
- Residual Plots: Regular, Standardized and Studentized – Deleted t
- Diagnostic measures: Cook's Distance (Influence), Leverage and DFITS
- Storage of model design matrix
- Main Effects and Interaction Plots (Fitted Means)
- Contour and Surface Plots
- Optimization with optional constraints including integer continuous
- Automatic removal of extreme VIF or collinear terms (with alias and removal report)
- Specify interactions, quadratic and higher orders (all interactions or up to 3-Way)
- ANOVA Type I and/or Type III Sum-of-Squares with Pareto of Percent Contribution and Standardized Effects
- Lenth Pseudo Standard Error for Saturated Models (Orthogonal or Non-Orthogonal) with Monte Carlo or Student T P-Values
- Specify Test/Withhold Sample for R-square Test & StDev Test Validation
- R-Square Predicted (Leave-One-Out Cross Validation)
- R-Square K-Fold & StDev K-Fold (K-Fold Cross Validation)
- Test for Constant Variance: Breusch-Pagan. Anderson-Darling Normality test is applied to residuals in order to automatically select Normal or Koenker (Robust) version. Report includes the Overall test and Individual predictors as well.
- White robust standard errors for non-constant variance (Heteroskedasticity-Consistent)
- Durbin-Watson test for autocorrelation in residuals with P-Values
- Newey-West robust standard errors for non-constant variance with autocorrelation (Heteroskedasticity and Autocorrelation-Consistent)
- White or Newey-West automatically selected based on Durbin-Watson P-Values
- Stepwise/Best Subsets Regression:
    - Forward/Backward with alpha-to-enter, alpha-to-remove
    - Forward Selection with alpha-to-enter
    - Backward Elimination with alpha-to-remove
    - Forward, Backward Criterion: Minimize AICc, BIC; Maximize R-Square Adjusted, R-Square Predicted, R-Square K-Fold
    - Best Subsets utilizes the powerful MIDACO Solver (Mixed Integer Distributed Ant Colony Optimization) to solve best subsets with up to hundreds of continuous or categorical variables, including interactions and higher order terms. This feature gives SigmaXL a significant advantage over competitors with Best Subsets limited to 30 continuous variables.
    - Best Subsets Criterion: Minimize AICc, BIC; Maximize R-Square Adjusted
    - Hierarchical option
    - Detailed report with additional statistics such as Condition Number and Mallows' Cp.
- Box-Tidwell Test and Power Transformation Recommendation for Continuous Predictors (New in Version 10.02) **
- Multiple Response Optimization with Desirability
    - Multistart Nelder-Mead Simplex
    - MIDACO

Statistical Tools – General Linear Model**:

Extends Advanced Multiple Regression to include:

- Fixed and Random Factors
- Nested Factors
- Covariates (can be Nested)
- For Random or Mixed Random/Fixed Factors with a balanced design, the ANOVA and Variance Components (VC) report is given based on Expected Mean Squares. VC confidence intervals using Restricted Maximum Likelihood (REML) are included.
- If the design is unbalanced or model is non-hierarchical, REML is used to compute the VC values and confidence intervals. Fixed Effects Tests are based on Satterthwaite approximation degrees of freedom.
- Main Effects with Confidence Intervals and Interaction Plots of Fitted Means for Non-Nested Fixed Factors
- Tukey and Fisher Pairwise Comparison of Means for Non-Nested Fixed Factors
- Predicted Response Calculator
- Multiple Response Optimization for Nested or Non-Nested Fixed Factors

Design of Experiments:

- Generate 2-Level Factorial and Plackett-Burman Screening Designs
  - User-friendly dialog box
  - 2 to 19 Factors; 4,8,12,16,20 Runs
  - Unique “view power analysis as you design”
  - Randomization, Replication, Blocking and Center Points
- Basic DOE Templates
  - 2 to 5 Factors, 2-Level Full and Fractional-Factorial designs
  - Automatic update to Pareto of Coefficients
  - Easy to use, ideal for training
- Main Effects & Interaction Plots
- Analyze 2-Level Factorial and Plackett-Burman Screening Designs
  - Used in conjunction with Recall Last Dialog, it is very easy to iteratively remove terms from the model
  - Interactive Predicted Response Calculator with 95% Confidence Interval and 95% Prediction Interval.
  - ANOVA report for Blocks, Pure Error, Lack-of-Fit and Curvature
  - Collinearity Variance Inflation Factor (VIF) and Tolerance report
  - Residual plots: histogram, normal probability plot, residuals vs. time, residuals vs. predicted and residuals vs. X factors
  - Highlight of significant outliers in residuals
  - Durbin-Watson Test for Autocorrelation
- Contour & 3D Surface Plots
- Response Surface Designs
  - 2 to 5 Factors
  - Central Composite and Box-Behnken Designs
  - Easy to use design selection sorted by number of runs

Control Charts:

- Control Chart Selection Tool
- Individuals, Individuals & Moving Range
- X-Bar & R, X-Bar & S
- I-MR-R, I-MR-S (Between/Within)
- P, NP, C, U
- P' and U' (Laney) to handle overdispersion
- Control charts include a report on tests for special causes. Special causes are also labeled on the control chart data point. Set defaults to apply any or all of Tests 1-8.
- Tests for Special Causes: Support for Varying Subgroup Sizes (Moving Limits), Historical Groups and MR/Range/StDev Charts (Tests 1-4)
- Process Capability report (Pp, Ppk, Cp, Cpk) is available for I, I-MR, X-Bar & R, X-bar & S charts.
- Add data to existing charts for operator ease of use!
- Scroll through charts with user defined window size
- Advanced Control Limit options: Subgroup Start and End; Historical Groups (e.g. split control limits to demonstrate before and after improvement)
- Exclude data points for control limit calculation
- Add comment to data point for assignable cause
- ± 1, 2 Sigma Zone Lines
- Control charts for Nonnormal data (Individuals)
    - Box-Cox and Johnson Transformations
    - 16 Nonnormal distributions supported (see Process Capability)
    - Individuals chart of original data with percentile based control limits
    - Individuals/Moving Range chart for normalized data with optional tests for special causes
- Control chart templates:
    - Individuals, C (Count)
    - Rare Events: T, G and Probability Based G
    - Time Weighted: Exponentially Weighted Moving Average (EWMA) and Tabular Cumulative Sum (CUSUM)
    - Trend/Tool Wear
    - Average Run Length (ARL) Calculators: Shewhart with Tests for Special Causes, Attribute C & P, EWMA & CUSUM

Reliability/Weibull Analysis:

- Weibull Analysis
    - Complete and Right Censored data
    - Least Squares and Maximum Likelihood
    - Output includes percentiles with confidence intervals, survival probabilities, and Weibull probability plot.

Time Series Forecasting and Control Charts for Autocorrelated Data:

- Run Chart
- Autocorrelation Function (ACF)/Partial Autocorrelation (PACF) Plots
- Cross Correlation (CCF) Plots with Pre-Whiten Data option
- Seasonal Trend Decomposition Plots
- Spectral Density Plot with Detection of Seasonal Frequency
- Exponential Smoothing:
    - Forecast with Prediction Intervals
    - Exponential Smoothing models use Rob Hyndman's taxonomy:
    - Additive/Multiplicative Error
    - Additive/Additive Damped Trend
    - Additive/Multiplicative Seasonal
    - This includes all of the classical exponential smoothing models such as Simple/Single/EWMA, Double and Holt-Winters
    - Multiple Seasonal Decomposition (MSD) option: useful for high frequency and/or multiple frequency data, such as Monthly with frequency = 12, Daily with frequency = 7 and Hourly with frequency = 24
- Exponential Smoothing Residuals Control Chart for autocorrelated data:
    - Individuals and Moving Limits (with One-Step Ahead Forecast) Charts
    - Add Data, Show Last 30 Data Points, Enable Scroll options
    - MSD option
- Autoregressive Integrated Moving Average (ARIMA):
    - Forecast with Prediction Intervals
    - ARIMA Forecast with Predictors (Continuous and/or Categorical)
    - MSD option
- ARIMA Residuals Control Chart for autocorrelated data:
    - Individuals and Moving Limits (with One-Step Ahead Forecast) Charts
    - ARIMA Control Chart with Predictors
    - Add Data, Show Last 30 Data Points, Enable Scroll options
    - MSD option
- Utilities:
    - Difference Data
    - Lag Data
    - Interpolate Missing Values (seasonally adjusted linear interpolation)
- Time Series Forecasting Model Features:
    - ARIMA and Exponential Smoothing models are fully automatic or user specified
    - Utilizes modern State Space and Kalman Filter models for accurate parameter estimation
    - ARIMA estimates missing values with Kalman Filter; Exponential Smoothing uses seasonally adjusted linear interpolation
    - Automatic Box-Cox Transformation
    - Automatic seasonal frequency detection
- Model diagnostics:
    - ACF/PACF Plots
    - Ljung-Box p-values
    - Log-Likelihood, AIC, AICc, BIC, Residual StDev
    - Residual plots (histogram, normal probability, residual versus fits, residuals versus order)
- Forecast Accuracy:
    - In-Sample (Estimation) one-step-ahead forecast errors (RMSE, MAE, MASE, MAPE)
    - Out-of-Sample (Withhold) one-step-ahead forecast errors
    - Out-of-Sample (Withhold) multi-step-ahead forecast errors
    - Evaluated using the benchmark standard M4 forecast competition data, a total of 100,000 data sets with Yearly, Quarterly, Monthly, Weekly, Daily and Hourly data. Using a hybrid average of automatic Exponential Smoothing and ARIMA, SigmaXL (unofficially) ranked 10th out of 60 in the Overall Weighted Average forecast accuracy score, ahead of three well known commercial forecast software packages.

Descriptive Statistics Options and Tolerance Interval Calculator (Normal Exact)

Shapiro-Wilk (SW) and Kolmogorov-Smirnov-Lilliefors (KSL) Normality Test

Shapiro-Wilk [1] is a popular and powerful alternative to the Anderson Darling normality test. It was originally restricted to sample sizes less than 50. Royston [2] extended the sample size to 2000. Rahman and Govidarajulu [3] extended the sample size further to 5000 and this has been implemented in SigmaXL.

For sample sizes greater than 5000, Kolmogorov-Smirnov-Lilliefors (KSL) is used. Lilliefors [4] modified the Kolmogorov-Smirnov test for the case of unknown mean and standard deviation.

Doornik-Hansen (DH) Normality Test

The Doornik-Hansen Normality Test [5] is a univariate or multivariate variate omnibus test based on Skewness and Kurtosis. In descriptive statistics, the univariate version is used, which is based on the method of Shenton and Bowman [6]. The bivariate version of DH is used in the Correlation Matrix to test for bivariate normality.

This test is best suited for data with ties, i.e., “chunky” data. Anderson-Darling and Shapiro-Wilk are severely affected by ties in the data and will trigger a low P-Value even if the data are normal.

Confidence and Tolerance Intervals

A confidence interval is described by the percentile of the population and the confidence probability. Thus, the 95% confidence interval for the 80th percentile of the sample will consist of a lower bound and an upper bound, where the true population 80th percentile will lie 95% of the time.

A tolerance interval is the interval, described by a lower and upper bound, that a specified minimum proportion of the sample’s values will lie with a given confidence probability. Thus, as an example, the 95% tolerance interval for the range in which 80 percent of the sample will fall will consist of a lower bound and an upper bound.

Notation:

n	is the number of observations in the sample.
v	are the degrees of freedom in the sample $(n-1)$.
$\overline{X}$	is the mean of the sample.

s	is the standard deviation of the sample.
p	is the quantile (CI), or the percentage of the population in the interval (TI).
Z_p	is the pth quantile of the standard normal distribution.
α	is the alpha probability - typically 0.05.

Tolerance Interval Calculator (Normal Exact)

When the sample can be assumed to be derived from a population that has a normal distribution, then parametric assumptions can be used to ascertain the Tolerance Interval. For each of the following, the lower L and upper U bounds are given by:

$$L = \overline{X} - ks$$
$$U = \overline{X} + ks$$

The value of k is shown below:

Normal Tolerance Interval - one sided exact

$$k = \frac{t_{1-\alpha,\nu,\lambda}}{\sqrt{n}}$$

where $t_{1-\alpha,\nu,\lambda}$ is the $1-\alpha$ percentile of the non-central t distribution with ν degrees of freedom and non-centrality parameter $\lambda = \sqrt{n}Z_p$.

Normal Tolerance Interval - two sided exact

$$G(z,n,p,k) = 1 - F_v\left[\frac{v\chi^2_{p,1,\lambda}}{k^2}\right]$$

where: $\chi^2_{p,1,\lambda}$ is the p percentile of the non-central χ^2 distribution with 1 degree of freedom and non-centrality parameter $\lambda = z^2/n$, and where F_v is the cdf of the χ^2 distribution with v degrees of freedom. Then k is the root such that:

$$2\int_0^\infty G(z,n,p,k)\phi(z)dz = 1-\alpha$$

where $\phi(z)$ is the pdf of the standard normal distribution. SigmaXL uses a 40-point Gauss-Legendre quadrature to evaluate the integral. See Krishnamoorthy and Mathew [7]. The approximation method by Howe [8] is used to obtain initial values.

Percentile (Nonparametric) Confidence and Tolerance Intervals

This methodology is used when no *a priori* assumptions can be made regarding the distribution of the data. Let x be the sorted data. The exact lower L and upper U bounds are given by:

$$L = x_r$$
$$U = x_s$$

The interpolated lower L and upper U bounds are given by:

$$L = \lambda_l x_{r+1} + (1-\lambda_l)x_r$$
$$U = \lambda_u x_{s+1} + (1-\lambda_u)x_s$$

Confidence Interval – two sided exact

$$r = B_{\alpha/2,n,p}$$
$$s = B_{1-\alpha/2,n,p} + 1$$

where $B_{\alpha,n,p}$ is the α percentile of the binomial distribution with parameters n and p.

Confidence Interval – two sided interpolated

Let r and s be the exact order statistics for confidence interval. Then

$$\lambda_l = \left(1 + \frac{r(1-p)(B_{n,p}(r) - \alpha/2)}{(n-r)p(\alpha/2 - B_{n,p}(r-1))}\right)^{-1}$$
$$\lambda_u = \left(1 + \frac{s(1-p)(B_{n,p}(s) - (1-\alpha/2))}{(n-s)p(1-\alpha/2 - B_{n,p}(s-1))}\right)^{-1}$$

where $B_{n,p}(k)$ is the cdf of the binomial distribution with parameters n and p. See Nyblom [9] which is an extension of the interpolation method of Hettmansperger and Sheather [10] for median using a double exponential distribution.

Tolerance Interval - two sided exact

$$k = B_{1-\alpha,n,p} + 1$$
$$r = \lfloor (n-k+1)/2 \rfloor$$
$$s = n - r + 1$$

where $B_{1-\alpha,n,p}$ is the $1-\alpha$ percentile of the binomial distribution with parameters n and p.

Tolerance Interval - two sided interpolated

Let k, r and s be the two sided exact order statistics for tolerance interval.

$$\lambda = \frac{1-\alpha/2 - B_{n,p}(k-2)}{B_{n,p}(k-1) - B_{n,p}(k-2)}$$

$$v_1 = \lambda x_r + (1-\lambda) x_{r+1}$$
$$v_2 = \lambda x_s + (1-\lambda) x_{s-1}$$

If $x_s - v_1 < v_2 - x_r$ then $L = v_1$ and $U = x_s$; else $L = x_r$ and $U = v_2$. See Young and Mathew [11].

References for Descriptive Statistics Options and Tolerance Interval Calculator (Normal Exact)

[1] "Shapiro-Wilk Test" in *Wikipedia: The Free Encyclopedia, Wikimedia Foundation Inc.,* https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test.

[2] Royston, Patrick (September 1992). "Approximating the Shapiro–Wilk W-test for non-normality", *Statistics and Computing*. 2 (3): 117–119.

[3] Rahman and Govidarajulu (1997). "A modification of the test of Shapiro and Wilk for normality", *Journal of Applied Statistics*. 24 (2): 219–236

[4] Lilliefors, H. (June 1967), "On the Kolmogorov–Smirnov test for normality with mean and variance unknown", *Journal of the American Statistical Association*, Vol. 62. pp. 399–402.

[5] Doornik, J.A. and Hansen, H. (2008). "An Omnibus test for univariate and multivariate normality", *Oxford Bulletin of Economics and Statistics*, 70, 927-939.

[6] Shenton, L. R. and Bowman, K. O. (1977). "A bivariate model for the distribution of √b1 and b2", *Journal of the American Statistical Association*, Vol. 72, pp. 206–211.

[7] Krishnamoorthy, K. and Mathew, T. (2009), *Statistical Tolerance Regions: Theory, Applications, and Computation*, Hoboken, NJ:Wiley.

[8] Howe, W. G. (1969), "Two-Sided Tolerance Limits for Normal Populations - Some Improvements", *Journal of the American Statistical Association* 64, pp. 610-620.

[9] Nyblom, J. (1992). "Note on interpolated order statistics", *Statistics and Probability Letters* 14, pp. 129-131.

[10] Hettmansperger, T.P. and S.J. Sheather (1986). "Confidence Intervals Based on Interpolated Order Statistics", *Statistics and Probability Letters* 4, pp. 75-79.

[11] Young, D.S. and Mathew, T. (2014). "Improved nonparametric tolerance intervals based on interpolated and extrapolated order statistics", *Journal of Nonparametric Statistics* 26 (3), pp. 415-432,

Statistical Details for Nonnormal Distributions and Transformations

Maximum Likelihood Estimation (MLE)

Maximum likelihood estimates of the parameters are calculated by maximizing the likelihood function with respect to the parameters. The likelihood function is simply the sum of the log of the probability density function (PDF) for each uncensored observation, and the log of the complement of the cumulative density function (CDF) for each right censored observation (in Reliability/Weibull Analysis). Initial estimates are derived using a branch and bound algorithm.

The maximum likelihood estimates are then calculated using the Newton-Raphson method. This is an iterative process that uses both the first and second derivatives to move to a point at which no further improvement in the likelihood is possible.

The standard errors of the parameter estimates are derived from the Hessian matrix. This matrix, which describes the curvature of a function, is the square matrix of second-order partial derivatives of the function.

For some data sets, the likelihood function for threshold models is unbounded, and the maximum likelihood methodology fails. In this context, a threshold is estimated using a bias correction method. This is an iterative methodology that evaluates the threshold based on the difference between the minimum value of the variate and the prediction for the minimum value, conditional on the current values of the parameters.

References for MLE and Distributions:

Greene, W.H. *Econometric Analysis* 4th Ed Prentice Hall, New Jersey.

Johnson, N. L., and Kotz, S. (1990). "Use of moments in deriving distributions and some characterizations", *Mathematical Scientist*, Vol. 15, pp. 42-52.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions-Volume 1*, Second Ed., Wiley, New York.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions-Volume 2*, Second Ed., Wiley, New York.

Nocedal, J. and Wright, S.J. (1999). *Numerical Optimization*, Springer-Verlag, New York.

Sleeper, A. (2006). *Six Sigma Distribution Modeling*, McGraw-Hill, New York.

Distributions

Beta Distribution

PDF

$$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

CDF

$$\int_{-\infty}^{x}\frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}dt$$

where *B* is the Beta function:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Range $0 \leq x \leq 1$.

Shape1 parameter α>0.

Shape2 parameter β>0.

Γ(β) is the Gamma function and is described below under Gamma distribution.

Beta Distribution with Lower/Upper Threshold

PDF

$$\frac{1}{B(\alpha,\beta)}\left(\frac{x-\theta_1}{\theta_2-\theta_1}\right)^{\alpha-1}\left(\frac{\theta_2-x}{\theta_2-\theta_1}\right)^{\beta-1}$$

CDF

$$\int_{-\infty}^{\frac{x-\theta_1}{\theta_2-\theta_1}}\frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}dt$$

where *B* is the Beta function.

Range $0 \le x \le 1$.

Shape1 parameter α>0.

Shape2 parameter β>0.

Lower threshold θ_1

Upper threshold θ_2

Notes: Estimation of the 4 parameter Beta distribution is undertaken in two parts. In the first part, initial parameter estimates are derived using the method of moments. The threshold parameters are then held using these values, and the shape parameters are then estimated using maximum likelihood.

Reference for Beta Distribution with Lower/Upper Threshold – Parameter Estimation:

Wang, J.Z. (2005). "A note on Estimation in the Four Parameter Beta Distribution", *Comm in Stats Simulation and computation*, Vol. 34 pp. 495- 501.

Box-Cox Distribution with Threshold

PDF

$$\frac{(x-\theta)^{\lambda-1}}{\sigma\sqrt{2\pi}}e^{-(((x-\theta)^{\lambda}-1)/\lambda-\mu)^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{[(x-\theta)^{\lambda}-1]/\lambda}\frac{1}{\sigma\sqrt{2\pi}}e^{-(t-\mu)^2/2\sigma^2}dt$$

Range $0 < x - \theta < \infty$.

Location parameter, μ, the mean.

Scale parameter, $\sigma > 0$, the standard deviation.

Shape parameter λ.

Threshold parameter $\theta < \min(x)$.

Notes: The concentrated likelihood is used in the ML estimation. This implies that the location and scale parameters are not estimated freely, but are derived as the mean and standard deviation of the BoxCox transformed variate. The estimated parameters λ and θ are then used in the Box-Cox (Power) transformation. See **Transformations** below.

Exponential Distribution

PDF

$$\frac{e^{-x/\alpha}}{\alpha}$$

CDF

$$1 - e^{-x/\alpha}$$

Range $0 \leq x < \infty$.

Scale parameter, $\alpha > 0$, the mean.

Exponential Distribution with Threshold

PDF

$$\frac{e^{-(x-\theta)/\alpha}}{\alpha}$$

CDF

$$1 - e^{-(x-\theta)/\alpha}$$

Range $0 \leq x < \infty$.

Scale parameter, $\alpha > 0$, the mean.

Threshold parameter $\theta < \min(x)$.

Gamma Distribution

PDF

$$\frac{(x/\alpha)^{\beta-1}e^{-x/\alpha}}{\alpha\Gamma(\beta)}$$

CDF

$$\int_0^{x/\alpha}\frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

where Γ(β) is the Gamma function:

$$\Gamma(\beta)=\int_0^{\infty}e^{-t}t^{\beta-1}dt$$

Range $0 \le x < \infty$.

Scale parameter $\alpha > 0$.

Shape parameter $\beta > 0$.

Gamma Distribution with Threshold

PDF

$$\frac{[(x-\theta)/\alpha])^{\beta-1}e^{-(x-\theta)/\alpha}}{\alpha\Gamma(\beta)}$$

CDF

$$\int_0^{(x-\theta)/\alpha} \frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

Range $0 \le x - \theta < \infty$.

Scale parameter $\alpha > 0$.

Shape parameter $\beta > 0$.

Threshold parameter $\theta < \min(x)$.

Half Normal Distribution

PDF

$$\frac{\sqrt{2/\pi}}{\sigma^2}e^{-x^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{x} \frac{\sqrt{2/\pi}}{\sigma}e^{-t^2/2\sigma^2}dt - 1$$

Range $0 \le x < \infty$.

Scale parameter, $\sigma > 0$, the standard deviation

Reference for Application of Half Normal Distribution:

Chou, C., & H. Liu, (1998). "Properties of the half-normal distribution and its application to quality control", *Journal of Industrial Technology* Vol. 14(3) pp. 4-7

Largest Extreme Value Distribution

PDF

$$\frac{1}{\sigma}e^{-(x-\mu)/\sigma}e^{-e^{-(x-\mu)/\sigma}}$$

CDF

$$e^{-e^{-(x-\mu)/\sigma}}$$

Range $-\infty < x < \infty$.

Location parameter, μ, the mode.

Scale parameter $\sigma > 0$.

Logistic Distribution

PDF

$$\frac{e^{(x-\mu)/\sigma}}{\sigma(1+e^{(x-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(x-\mu)/\sigma}}$$

Range $-\infty < x < \infty$.

Location parameter, μ, the mean.

Scale parameter $\sigma > 0$.

Loglogistic Distribution

PDF

$$\frac{e^{(\ln(x)-\mu)/\sigma}}{x\sigma(1+e^{(\ln(x)-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(\ln(x)-\mu)/\sigma}}$$

Range $0 < x < \infty$.

Location parameter, μ, the mean.

Scale parameter $\sigma > 0$.

Loglogistic Distribution with Threshold

PDF

$$\frac{e^{(\ln(x-\theta)-\mu)/\sigma}}{(x-\theta)\sigma(1+e^{(\ln(x-\theta)-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(\ln(x-\theta)-\mu)/\sigma}}$$

Range $0 < x < \infty$.

Location parameter, μ, the mean.

Scale parameter $\sigma > 0$.

Threshold parameter $\theta <$ min(x).

Lognormal Distribution

PDF

$$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-.5(\ln(x)-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x}\frac{1}{t\sigma\sqrt{2\pi}}e^{-.5(\ln(t)-\mu)^2/\sigma^2}dt$$

Range $0 < x < \infty$.

Scale parameter, μ, the mean of ln(x) .

Shape parameter, $\sigma > 0$, the standard deviation of ln(x) .

Lognormal Distribution with Threshold

PDF

$$\frac{1}{(x-\theta)\sqrt{2\pi\sigma^2}}e^{-.5(\ln(x-\theta))-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x-\theta} \frac{1}{t\sigma\sqrt{2\pi}} e^{-.5(\ln(t)-\mu)^2/\sigma^2} dt$$

Range $0 < x < \infty$.

Scale parameter, μ, the mean of ln(x) .

Shape parameter, $\sigma > 0$, the standard deviation of ln(x) .

Threshold parameter $\theta <$ min(x).

Reference for Lognormal Distribution with Threshold – Parameter Estimation:

Giesbrecht, F. and A.H. Kempthorne (1966). "Maximum Likelihood Estimation in the Three-parameter Lognormal Distribution", *Journal of the Royal Statistical Society*, B 38, pp. 257-264.

Normal Distribution

PDF

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5(x-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5(t-\mu)^2/\sigma^2} dt$$

Range $-\infty < x < \infty$.

Location parameter, μ, the mean.

Scale parameter, $\sigma > 0$, the standard deviation.

Note: For consistency with other reports in SigmaXL such as Descriptive Statistics, the standard deviation is estimated as the sample standard deviation using n-1 (rather than n).

Smallest Extreme Value Distribution

PDF

$$\frac{1}{\sigma}e^{(x-\mu)/\sigma}e^{-e^{(x-\mu)/\sigma}}$$

CDF

$$1-e^{-e^{(x-\mu)/\sigma}}$$

Range $-\infty < x < \infty$.

Location parameter, μ, the mode.

Scale parameter $\sigma > 0$.

Weibull Distribution

PDF

$$\frac{\beta x^{\beta-1}}{\alpha^{\beta}}e^{-(x/\alpha)^{\beta}}$$

CDF

$$1-e^{-(x/\alpha)^{\beta}}$$

Range $0 \leq x < \infty$.

Scale parameter, $\alpha > 0$, the characteristic life.

Shape parameter $\beta > 0$.

Weibull Distribution with Threshold

PDF

$$\frac{\beta x^{\beta-1}}{\alpha^{\beta}} e^{-[(x-\theta)/\alpha]^{\beta}}$$

CDF

$$1 - e^{-[x-\theta)/\alpha]^{\beta}}$$

Range $0 \leq x < \infty$.

Scale parameter, $\alpha > 0$, the characteristic life.

Shape parameter $\beta > 0$.

Threshold parameter $\theta < \min(x)$.

Reference for Weibull Distribution with Threshold – Parameter Estimation:

Lockhart, R.A. and M.A. Stephens (1994)."Estimation and Tests of Fit for the Three-parameter Weibull Distribution", *Journal of the Royal Statistical Society*, Vol.56(3), pp. 491-500.

Transformations

Box-Cox (Power) Transformation

$\lambda \neq 0$

$$\begin{aligned} z &= x^{\lambda} \\ x &= z^{1/\lambda} \end{aligned}$$

$\lambda = 0$

$$\begin{aligned} z &= \ln(x) \\ x &= e^{z} \end{aligned}$$

Range $0 < x < \infty$.

Shape parameter λ.

Note: The optimum shape parameter, λ, is derived using a grid search in which the criteria function is the standard deviation of the standardized transformed variable.

Box-Cox (Power) Transformation with Threshold

$\lambda \neq 0$

$$\begin{aligned} z &= (x-\theta)^{\lambda} \\ x &= z^{1/\lambda} + \theta \end{aligned}$$

$\lambda = 0$

$$\begin{aligned} z &= \ln(x-\theta) \\ x &= e^{z} + \theta \end{aligned}$$

Range $0 < x - \theta < \infty$.

Shape parameter λ.

Threshold parameter $\theta < \min(x)$.

Note: The parameters λ and θ are estimated using MLE as described above in **Box-Cox Distribution with Threshold.**

Johnson Transformation

The Johnson Transformation selects one of the three families of distribution: SB (bounded), SL (lognormal), and SU (unbounded) and the associated parameters so as to transform the data to be normally distributed. The methodology follows Chou et al (1998) and uses the Anderson Darling P-Value as the normality criteria.

Johnson Transformation - S_B

$$\begin{aligned} z &= \gamma + \eta \ln\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right) \\ x &= \epsilon + \frac{1}{1+e^{-(z-\gamma)/\eta}} \end{aligned}$$

z is $N(0, 1)$

Range $\varepsilon < x < \varepsilon + \lambda$.

Location parameter $\eta > 0$.

Scale parameter $\lambda > 0$.

Shape parameter γ unbounded.

Shape parameter ε unbounded.

Johnson Transformation - S_L

$$\begin{aligned} z &= \gamma + \eta \ln(x - \epsilon) \\ x &= \epsilon + e^{(z-\gamma)/\eta} \end{aligned}$$

z is N (0, 1)

Range $x > \varepsilon$

Location parameter $\eta > 0$.

Shape parameter γ unbounded.

Shape parameter ε unbounded.

Johnson Transformation - S_U

$$\begin{aligned} z &= \gamma + \eta \sinh^{-1}[(x - \epsilon)/\lambda] \\ x &= \epsilon + \lambda \sinh[(z - \gamma)/\eta] \end{aligned}$$

z is N (0, 1)

Range $-\infty < x < \infty$.

Location parameter $\eta > 0$.

Scale parameter $\lambda > 0$.

Shape parameter γ unbounded.

Shape parameter ε unbounded.

References for Johnson Transformation:

Chou,Y., A.M. Polansky, and R.L. Mason (1998). "Transforming Nonnormal Data to Normality in Statistical Process Control," *Journal of Quality Technology*, Vol. 30(2), pp. 133-141.

David, H.A. (1981). *Order Statistics*, John Wiley & Sons, New York

Tadikamalla,P.,R. (1980). "Notes and Comments: On Simulating Nonnormal Distributions", *Psychometrika,*, Vol. 45(2), pp. 273-279.

Monte Carlo Random Number Generation

A random number generator (RNG) is required when using Monte Carlo simulations in order to draw random samples from a specified distribution. The creation of a random number involves an algorithm that can automatically create long runs of numbers with good random properties. These algorithms are called pseudo-random number generators. The state of the RNG after each iteration is used as an input for the generation of the next random number. The initial state is called a seed; this seed is derived using the system clock.

A good RNG will produce a long run of numbers which are independent, so that there is no correlation between successive numbers. The RNG used in SigmaXL and DiscoverSim is based on a recur-with-carry “KISS+Monster” random number generator developed by George Marsaglia. This algorithm produces random integers between 0 and $2^{32} – 1$ and has a period of 10^{8859}. This implementation of the KISS+Monster algorithm has been tested and passes all of the Diehard tests (Marsaglia, 2003).

Pearson Family of Distributions

The Pearson family consists of eight distributions:
Type 0 Normal distribution.
Type 1 Beta distribution with minimum and maximum.
Type 2 Symmetric Beta distribution with minimum and maximum. The symmetry is imposed by requiring the two shape parameters to have the same value.
Type 3 Gamma distribution with threshold.
Type 4 Pearson IV distribution.
Type 5 Inverse Gamma distribution with threshold.
Type 6 F distribution with scale and threshold.
Type 7 Student's T distribution with location and scale
The Pearson methodology selects one member of the family using the methodology described in Johnson, Kotz and Balakrishnan, Volume 1, pp. 15-33. Specified Mean (0), Standard Deviation (1), Skewness and Excess Kurtosis are converted to the selected distribution’s parameters.

Types 0 to 3 have been described above.

Type 4 Pearson Type IV Distribution

PDF

$$\frac{k(\nu,\omega)}{\lambda}\left(\frac{1}{1+z^2}\right)^{\omega} e^{-\nu\, atan(z)}$$

CDF

$$\int_{-\infty}^{z}\frac{k(\nu,\omega)}{\lambda}\left(\frac{1}{1+t^2}\right)^{\omega} e^{-\nu\, atan(t)}\, dt$$

where

$$b = 2(\omega - 1)$$
$$a = \sqrt{\frac{b^3 - b^2}{b^2 + \nu^2}}$$
$$z = \frac{x-\mu}{a\lambda} - \frac{\nu}{b}$$

and $k(\nu,\omega)$ is a scaling constant.

Range $0 \le x < \infty$.

Location parameter μ

Scale parameter, $\lambda > 0$.

Shape parameter ν.

Shape parameter $\omega > 1.5$.

Type 5 Inverse Gamma Distribution with Threshold

PDF

$$\frac{\alpha^{\beta}}{\Gamma(\beta)}(x-\theta)^{-\beta-1}e^{-\alpha/(x-\theta)}$$

CDF

$$\frac{\gamma(\beta,\alpha/(x-\theta))}{\Gamma(\beta)}$$

where $\Gamma(s)$ is the the gamma function, and $\gamma(s,x)$ is the incomplete gamma function.

Range $0 \leq x - \theta < \infty$.

Scale parameter $\alpha > 0$.

Shape parameter $\beta > 0$.

Threshold parameter $\theta < min(x)$.

Type 6 F Distribution with Scale and Threshold

PDF

$$\frac{1}{(x-\theta)\ \mathrm{B}(.5\nu, .5\omega)} \sqrt{\frac{(\nu\ (x-\theta)/\alpha)^{\nu}\ \ \omega^{\omega}}{(\nu\ (x-\theta)/\alpha+\omega)^{\nu+\omega}}}$$

CDF

$$\int_{-\infty}^{z} \frac{1}{B(.5\ \nu, .5\ \omega)} t^{.5\nu-1}(1-t)^{.5\omega-1} dt$$

where

$$z = \frac{(\nu\ (x-\theta))}{(\nu\ x + \alpha\omega)}$$

and where B is the Beta function.

Range $0 \leq x - \theta < \infty$.

Scale parameter $\alpha > 0$.

Shape parameter $\nu > 0$.

Shape parameter $\omega > 0$. Threshold parameter $\theta < min(x)$.

Type 7 Student's T distribution with Location and Scale

PDF

$$\frac{1}{\alpha\sqrt{\pi\ \nu}} \frac{\Gamma(.5(\nu+1))}{\Gamma(.5\nu)} \left(\frac{\nu\alpha^2}{(x-\mu)^2+\nu\alpha^2}\right)^{.5(\nu+1)}$$

CDF

$$.5 + .5\ I_z(.5, .5\nu) \quad x \geq 0$$

$$.5 - .5\, I_z(.5, .5\nu) \quad x < 0$$

where

$$z = \frac{(x-\mu)^2}{(x-\mu)^2 + \nu\alpha^2}$$

and where Γ is the gamma function, and $I_z(a, b)$ is the regularized incomplete beta function.

Range $-\infty \le x < \infty$.

Location parameter μ.

Scale parameter $\alpha > 0$.

Shape parameter $\nu > 0$.

References for Pearson Family Monte Carlo Simulation:

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions-Volume 1*, Second Ed., Wiley, New York.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions-Volume 2*, Second Ed., Wiley, New York.

Marsaglia, G. (2003). "Random Number Generators," *Journal Of Modern Applied Statistical Methods*, Vol 2(1), pp. 2-13.

Automatic Best Fit

SigmaXL uses the Anderson Darling P-Value as the criteria to determine best fit. All distributions and transformations are considered and the model with the highest AD P-Value is initially selected (denoted as *adpvalmax*). A search is then carried out for models that are close, having an AD P-Value greater than *adpvalmax* - 0.1 (with an added criteria that AD P-Value be > 0.2), but having fewer parameters than the initial best fit model. If a simpler model is identified, then this is selected as the best fit.

Since AD P-Values are not available for distributions with thresholds (other than Weibull), an estimate is obtained by transforming the data to normality and then using a modified Anderson Darling Normality test on the transformed data. The transformed z-values are obtained by using the inverse cdf of the normal distribution on the cdf of the nonnormal distribution. The Anderson Darling Normality test assumes a mean = 0 and standard deviation = 1.

This approach, unique to SigmaXL, is an extension of the Chou methodology used in Johnson Transformations and allows a goodness of fit comparison across all distributions and transformations.

Another approach to comparing models is Akaike's information criterion (AIC) developed by Hirotsugu Akaike. AIC is the MLE log-likelihood with a penalty for the number of terms in the model (where the penalty factor also depends on the sample size, n). The AD P-Value method used by SigmaXL has the advantage that it is not limited to models with maximum likelihood parameter estimation.

References for Automatic Best Fit:

Chou,Y., A.M. Polansky, and R.L. Mason (1998). "Transforming Nonnormal Data to Normality in Statistical Process Control," *Journal of Quality Technology*, Vol. 30(2), pp. 133-141.

D'Agostino, R.B. and Stephens, M.A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

http://en.wikipedia.org/wiki/Akaike_information_criterion

Process Capability Indices (Nonnormal)

Z-Score Method (Default)

The Z-Score method for computing process capability indices utilizes the same z-value transformation to normality described above in **Automatic Best Fit.** The transformed z-values are obtained by using the inverse cdf of the normal distribution on the cdf of the nonnormal distribution. Normal based capability indices are then applied to the transformed z-values. This approach offers two key advantages: the relationship between the capability indices and calculated defects per million is consistent across the normal and all nonnormal distributions, and short term capability indices Cp and Cpk can be estimated using the standard deviation from control chart methods on the transformed z-values. The Z-Score method was initially developed by Davis Bothe and expanded on by Andrew Sleeper. For further details, see Sleeper, *Six Sigma Distribution Modeling*.

Percentile (ISO) Method

The Percentile method to calculate process capability indices uses the following formulas:

Ppu = (USL – 50^{th} percentile)/(99.865 percentile – 50^{th} percentile)

Ppl = (50^{th} percentile – LSL)/(50^{th} percentile – 0.135 percentile)

Ppk = min(Ppu, Ppl)

Pp = (USL – LSL)/(99.865 percentile – 0.135 percentile)

References for Process Capability Indices:

Bothe, D.R. (1997). *Measuring Process Capability*, McGraw-Hill, New York.

Sleeper, A. (2006). *Six Sigma Distribution Modeling*, McGraw-Hill, New York.

Control Charts (Nonnormal)

Individuals – Original Data

The **Individuals – Original Data** chart displays the untransformed data with control limits calculated as:

UCL = 99.865 percentile

CL = 50th percentile

LCL = 0.135 percentile

The benefit of displaying this chart is that one can observe the original untransformed data. Since the control limits are based on percentiles, this represents the overall, long term variation rather than the typical short term variation. The limits will likely be nonsymmetrical.

Individuals/Moving Range – Normalized Data

The **Individuals/Moving Range – Normalized Data** chart displays the transformed z-values with control limits calculated using the standard Shewhart formulas for Individuals and Moving Range charts. The benefit of using this chart is that tests for special causes can be applied and the control limits are based on short term variation. The disadvantage is that one is observing transformed data on the chart rather than the original data.

SigmaXL's default setting is to display the two charts: Individuals – Original Data, Individuals & Moving Range – Normalized Data (with tests for special causes unchecked).

Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests

Introduction

Nonparametric tests are popular because they do not assume that the sample data are normally distributed, but they do, however, assume that the test statistic follows a normal or chi-square distribution when computing the P-Value using "large sample" or "asymptotic" theory. Similarly, the Chi-Square test for independence used in r x c contingency tables assumes that the test statistic is distributed as chi-square. When the sample size is small or sparse, these assumptions may be invalid and can result in substantial error in the computed P-Value. For example, Cochrane's rule [5] for r x c contingency tables is a commonly used guideline: no more than 20% of the cells should have an expected value less than 5 and none should have an expected value less than 1.

The solution to the small sample problem is to compute exact P-Values using the true permutation distribution of the test statistic. However, explicit full enumeration of all possible outcomes quickly becomes computationally infeasible. For example, a 5x6 contingency table results in 1.6 billion possible outcomes.

Network algorithms by Mehta & Patel [6, 7] use techniques from operations research such as backward induction and forward probing to implicitly and efficiently solve the exact P-Value for contingency tables. These algorithms have also been extended to work with continuous data. For example, the nonparametric Mann-Whitney can be converted to a weighted singly ordered contingency table and the exact P-Value solved using the network algorithm.

Some data sets are too large even for the network algorithm to solve in a reasonable time period, but still do not satisfy Cochrane's rule. In these cases, Monte Carlo simulation is used to sample from the full enumeration of possible outcomes. 10,000 replications are typically used and 99% confidence intervals provided for the Monte Carlo P-Values. If the P-Value is low, 1 million replications are recommended.

Full Enumeration

To deal with a situation where the size of the data set does not justify the large sample asymptotic assumption, we will first consider full enumeration. To illustrate, consider the problem of evaluating the probability of getting a total of 3 or less when tossing two dice.

Evaluated for each of the n = 36 possible outcomes the sum of those situations (m) for which the sum or the two dice is 3 or less. Since m = 3 {1,1}, {1,2}, {2,1}, the probability is m/n = 3/36 =1/12.

Full enumeration is 100% accurate, and always produces the exact P-Value. It works fine provided *n* is not too large. However, for 20 dice, *n*, the number of possible outcomes is 3656 trillion, which makes full enumeration infeasible. As mentioned above, even a relatively small problem such as a 5x6 contingency table results in 1.6 billion possible outcomes, again making full enumeration infeasible.

Exact

Instead of evaluating each of the *n* possible outcomes, shortcuts can be derived which reduces the number of outcomes that need to be evaluated. Starting with the first dice showing unity, once the second dice exceeds 2 the remaining four outcomes can be counted as a block, since the statistic for each case will exceed the test statistic. Thus rather than evaluating each as one would with full enumeration, the block is evaluated. When the first dice shows two, and once the second dice exceeds 1, the remaining 5 outcomes can again be counted as a block. Once the first dice shows 3, all the remaining outcomes can be evaluated as a block, since again every possible outcome will then exceed 3. Hence instead of enumerating 36 outcomes, one enumerates 6 outcomes {1,1}, {1,2}, {1, 3+}. {2,1}, {2,2+}, {3+,1+}. This type of model is referred to as Enhanced Enumeration. It uses full enumeration of all possible outcomes node by node, and evaluates the statistic as a block if the largest statistic possible from the current position exceeds the critical value (in which case the probability of this path is added to the total probability), or if the smallest statistic possible from the current position is less than the critical value, implying no augmentation of the total probability [15].

A second methodology involves creating a network of a set of nodes in a number of stages. The distances between nodes are defined so that the total distance of a path through the network is the corresponding value of the test statistic. At each node, the algorithm computes the shortest and longest path distances for all the paths that pass through that node. This works easily for those statistics that can be evaluated as the linear sum of partial statistics evaluated at each node. This methodology is commonly referred to in the literature as the Network Model, after the pioneering work of Mehta and Patel [1, 6, 7].

The Network Model has advantages over enhanced enumeration in that the problem is broken down to nodes. For example, if we had 4 dice, then there would be 5 nodes. The advantage of the network model is that the length (and hence probability) at each node for the remaining nodes can be estimated using a backward step through the network. Since networks tend to be sparse, the set of all paths through the network which have the same probability up to the current node can be retained. Thus, for any set of paths which have the same length up to the current node, one can quickly read off the longest and shortest remaining sub-paths to the terminal node. Specifically, as in enhanced enumeration, if every possible outcome does not contribute to the P-Value, the entire set can be eliminated, while if every possible outcome contributes, then the probability of the entire set can be evaluated as a closed form.

Thus an exact evaluation of a probability is identical to full enumeration, but significantly faster. Again, as the size of the problem increases, the exact method also becomes infeasible since the number of outcomes that do have to be evaluated becomes too large. The methodology used to evaluate each of the exact nonparametric tests is described below.

Monte Carlo

Simulate the throw of the two dice *nrep* times, and count the number of times (m) that the sum of the two dice is 3 or less. The Monte Carlo P-Value is then defined as *m/nrep*; this is an estimate of the exact P-Value, and can be made more accurate by increasing *nrep*. Thus, with 20 dice, one could choose a value of *nrep* = 100,000, and carry out the simulation to ascertain an unbiased estimate of the P-Value. Thus the beauty of Monte Carlo is that it can be used on any problem, no matter how large the total possible outcomes; however it is only asymptotically exact.

Two Tailed (Sided) Tests

A two tailed test is a statistical test in which the critical area of a distribution is two sided and tests whether a sample is either greater than or less than a certain range of values. A one tailed test, on the other hand, tests a one sided hypothesis, such as whether a sample is less than a certain range of values. Thus a two tailed test will be the sum of a tail on the left hand side of a distribution, and the corresponding tail on the right hand side of the distribution. If a distribution is symmetric, a two tailed test will simply be twice the value of the one tailed test.

Formally, the two tailed probability is

Pr(|stat – mu| >= |statcrit – mu|)

summed over all possible combinations, where statcrit is the statistic evaluated for each possibility, and mu is the expected value of the statistic. For symmetric distributions, 2*1-Sided and 2-Sided will be identical. For non-symmetric distributions, such as the Runs Test and the Mann-Whitney Test, there will be a difference between 2*1-Sided and 2-Sided.

Sample Sizes for Exact

Nonparametric Tests

The following sample size guidelines are based on our recommendations and those given in Siegel [2] and Gibbons [3] for when exact nonparametric tests should be used rather than asymptotic (Normal or Chi-Square approximation):

Sign Test: N <= 50 (exact is automatically utilized in the regular Sign Test)

Wilcoxon Signed Rank: N <= 15

Mann-Whitney: Each sample N <= 10

Kruskal-Wallis: Each sample N <= 5

Mood's Median: Each sample N <= 10

Runs Test (Above/Below) or Runs Test (Up/Down) Test: N <= 50

These are minimum sample size guidelines. It is acceptable to use an exact test for larger sample sizes, but computation time can become an issue especially for tests with two or more samples. In those cases, one can always use a Monte Carlo P-Value with 99% confidence interval.

The maximum total sample size (total of all samples) that can be used for Nonparametric Exact or Monte Carlo is 1000. The Runs Test for Randomness and Wilcoxon Test automatically switch to large sample asymptotic if N > 1000.

One Way Chi-Square and Contingency Tables

Use Chi-Square Exact/Monte Carlo or Fisher's Exact/Monte Carlo if more than 20% of the cells have expected counts less than 5 or if any of the cells have an expected count less than 1 [4,5].

The maximum number of cells in a table is limited to 1000. For a One-Way Chi-Square the maximum sum of cell counts is 1000. For a Contingency Table, the maximum sum of cell counts is 5000.

Power

It is important to note that while exact P-Values are "correct," they do not increase (or decrease) the power of a small sample test, so they are not a solution to the problem of failure to detect a change due to inadequate sample size.

Validation of SigmaXL Exact P-Values

SigmaXL Exact P-Values are validated by comparison to textbook examples [2, 3], published paper examples [1, 6, 7, 8, 9] and other exact software such as StatXact, SPSS Exact, SAS and Matlab.

Monte Carlo P-Values are validated using 1e6 replications and compared against exact. Repeated simulations are used to validate the confidence intervals.

1 Sample Tests

Sign Test: The number of sign changes in one direction is distributed binomial, and so the binomial distribution function can be used directly to evaluate the tail probability [10].

Runs Test (Above/Below): Since the number of runs cannot exceed the length of the sample, it is easy to evaluate the probability of *r* runs when there is a binary variable with *n1* and *n2* elements [11].

Runs Test (Up/Down): Given the sample size and the number of up/down runs, the probability is provided by published tables [12].

Wilcoxon Test: Given the signed rank statistic, look at all combinations of ranks that could contribute to the observed value of the Wilcoxon critical value. For each of these cases, the number of combinations that actually contribute is evaluated. These counts are then converted to probabilities [13].

2 Sample Tests

Mann-Whitney Test: The continuous data is first converted to a weighted singly ordered contingency table and the exact P-Value is solved using the network algorithm.

Each possible combination of ranks is systematically evaluated, but the process is cut short for any given path if the largest statistic possible from the current position is less than the critical value (in which case all paths from the present are added to the feasible combinations), or if the smallest statistic possible from the current position is greater than the critical value (in which none of the paths from the present are added to the feasible combinations).

K Sample Tests

Kruskal-Wallis Test: Full enumeration.

Mood's Median Test: This test is evaluated using the Network Model. In each case, a chi-square statistic is evaluated. The number of combinations at the end of each path that has a chi-square statistic greater than the critical value is summed over all the enumerations, and is then converted to a probability.

Contingency Tests

One Way Chi-Square Test: For a one dimension vector of frequencies, and expected probabilities, the Chi Square test is evaluated using the Enhanced Enumeration model.

Two Way Fisher Exact Test: The exact probability for a two dimension *r* x *c* contingency table is evaluated using the Network model [1]. This is a one tailed test, and includes the probabilities of all the equivalent tables (those with same row and column counts) that have a lower or equal probability as the sample table. As in the one dimensional case, tests are made to see if the shortest path from the current position exceeds the path length of the observed table (in which case the probability of all paths from the current position are added to the probability, or whether the longest path from the current position is less than the path length of the current table, implying no augmentation of the probability.

References for Exact and Monte Carlo P-Values

[1] Cyrus R. Mehta & Nitin R. Patel. (1983). "A network algorithm for performing Fisher's exact test in **r** *x* c contingency tables." *Journal of the American Statistical Association,* Vol. 78, pp. 427-434.

[2] Siegel, S., & Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences* (2nd Ed.). New York, NY: McGraw-Hill.

[3] Gibbons, J.D. and Chakraborti, S. (2010). *Nonparametric Statistical Inference* (5th Edition). New York: Chapman & Hall.

[4] Yates, D., Moore, D., McCabe, G. (1999). *The Practice of Statistics* (1st Ed.). New York: W.H. Freeman.

[5] Cochran WG. "Some methods for strengthening the common [chi-squared] tests." *Biometrics* 1954; 10:417–451.

[6] Mehta, C. R.; Patel, N. R. (1997) "Exact inference for categorical data," unpublished preprint. See Table 7 (p. 33) for validation of Fisher's Exact P-Value.

[7] Mehta, C.R. ; Patel, N.R. (1998). "Exact Inference for Categorical Data." In P. Armitage and T. Colton, eds., *Encyclopedia of Biostatistics*, Chichester: John Wiley, pp. 1411–1422.

[8] Mehta, C.R. ; Patel, N.R. (1986). "A Hybrid Algorithm for Fisher's Exact Test in Unordered rxc Contingency Tables," *Communications in Statistics - Theory and Methods*, 15:2, 387-403.

[9] Narayanan, A. and Watts, D. "Exact Methods in the NPAR1WAY Procedure," SAS Institute Inc., Cary, NC.

[10] Myles Hollander and Douglas A. Wolfe (1973), *Nonparametric Statistical Methods.* New York: John Wiley & Sons.

[11] https://onlinecourses.science.psu.edu/stat414/node/330

[12] Eugene S. Edgington, (1961). "Probability Table for Number of Runs of Signs of First Differences in Ordered Series," *Journal of the American Statistical Association,* Vol. 56, pp. 156-159.

[13] David F. Bauer (1972), "Constructing confidence sets using rank statistics," *Journal of the American Statistical Association* Vol 67, pp. 687–690.

[14] Hongsuk Jorn & Jerome Klotz (2002). "Exact Distribution of the K Sample Mood and Brown Median Test," *Journal of Nonparametric Statistics*, 14:3, 249-257.

[15] Valz, P. D. and Thompson, M. E. (1994), "Exact Inference for Kendall's S and Spearman's Rho with Extensions to Fisher's Exact Test in rxc Contingency Tables," *Journal of Computational and Graphical Statistics*, Vol 3(4), pp. 459- 472.

Hypothesis Test Assumptions Report

The Hypothesis Test Assumptions Report is an optional text report for the 1 Sample t-Test, Paired t-Test, 2 Sample t-Test, One-Way ANOVA and Welch's ANOVA. The cells are color highlighted: Green (OK), Yellow (Warning) and Red (Serious Violation).

Normality

Each sample is tested for Normality using the Anderson-Darling (AD) test. If the AD P-Value is less than 0.05, the cell is highlighted as yellow (i.e., warning – proceed with caution). The Skewness and Kurtosis are reported and a note added, "See robustness and outliers." An example report is shown:

> Anderson Darling P-Value = 0.030. Reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is violated (at 95% confidence level). Skewness value = -0.9680 and Kurtosis value = 0.6796. See robustness and outliers.

If the AD P-Value is greater than or equal to 0.05, the cell is highlighted as green.

> Anderson Darling p-value = 0.531. Fail to reject null hypothesis: "data are sampled from a normal distribution," so conclude that the assumption of normality is not violated.

A red highlight is not used for the Normality report.

Robustness

A minimum sample size for robustness to nonnormality is determined using the regression equations in **Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA**. These were derived from extensive Monte Carlo simulations to determine a minimum sample size required for a test to be robust, given a specified sample Skewness and Kurtosis.

When there are 2 or more samples, a minimum sample size is calculated for each sample, and the largest value is (conservatively) considered as the minimum sample size required for the test.

If the AD Normality test for every sample is greater than or equal to 0.05, the following message is shown for robustness:

> **Not applicable for normal data.**

If each sample size is greater than or equal to the minimum for robustness, the minimum sample size value is reported and the test is considered to be robust to the degree of nonnormality present in the sample data:

> **Minimum sample size for a robust t-test = 22. Since the sample size is greater than this, the t-test is robust to nonnormality.**

If any sample size is less than the minimum for robustness, the minimum sample size value is reported and a suitable Nonparametric Test is recommended. The cell is highlighted in red:

> **Minimum sample size for a robust t-test = 16. It is recommended that the Exact Nonparametric One Sample Wilcoxon Test be used (SigmaXL > Statistical Tools > Nonparametric Tests - Exact > 1 Sample Wilcoxon - Exact).**

<u>Recommended Nonparametric Tests – One Sample</u>

The following Nonparametric Tests may be recommended for one sample:

1 Sample Sign

1 Sample Wilcoxon

1 Sample Wilcoxon – Exact

The 1 Sample Wilcoxon is more powerful that the 1 Sample Sign, but assumes that the data are symmetrical. Wilcoxon is therefore the default recommendation, but if the sample is moderately skewed then the Sign test is recommended.

Robustness to symmetry equations were developed in a manner similar to those used for robustness to nonnormality above. Monte Carlo simulations were run using sample sizes (N) of 20 to 2000. Observed alpha values were determined empirically from the P-Values of 100,000 replicate Wilcoxon tests. For a given sample size, nonnormal data with Skew = 0.01 to +1.0 was generated using the DiscoverSim Pearson Family function (see SigmaXL's DiscoverSim Workbook for details). The population test median (H0) used for the Wilcoxon test was estimated empirically by averaging the 100,000 sample medians. The test was considered robust when the simulated observed alpha was between .04 and .06, .i.e. within +/- 20% of the specified alpha of .05. "Maximum Skewness" occurred when the observed alpha exceeded .04 or .06.

Other alpha values were not considered. Power was not considered.

This simulation was performed for Ha = "Not Equal To" (NE), "Greater Than" (GT), and "Less Than" (LT). A regression analysis was performed on the Log10 (Maximum Skewness) versus Log10(N), resulting in the following equations (with R-Square reported):

1. log10(**Maximum Skewness NE**) = 0.5391 - 0.5028 log10(**N**) (R-Square = 99.5%)
2. log10(**Maximum Skewness GT**) = 0.2380 - 0.5725 log10(**N**) (R-Square = 98.1%)
3. log10(**Maximum Skewness LT**) = 0.2796 - 0.5459 log10(**N**) (R-Square = 99.7%)

Note that if Skewness is negative, equation 2 is used for LT and equation 3 is used for GT.

For example, given a sample size of 30 and Ha = "Not Equal To," solving for **Maximum Skewness NE** in equation 1, the value is 0.626. Therefore, if the sample Skewness is less than or equal to 0.626, Wilcoxon is recommended. If the sample Skewness is greater than 0.626, the Sign Test is recommended.

If the Sign Test is recommended, Exact is not explicitly recommended since it is already "built in" the Sign Test for N <= 50.

If the Wilcoxon Test is appropriate and N <= 15, then Wilcoxon – Exact is recommended (using the rules of thumb in **Sample Sizes for Exact**).

Recommended Nonparametric Tests – Two Samples

The following Nonparametric Tests may be recommended for two samples:

2 Sample Mann-Whitney

2 Sample Mann-Whitney – Exact

If each sample N <= 10, then Mann-Whitney – Exact is recommended (using the rules of thumb in **Sample Sizes for Exact**).

Recommended Nonparametric Tests – Three or More Samples (ANOVA)

The following Nonparametric Tests may be recommended for three or more samples:

Kruskal-Wallis

Kruskal-Wallis - Exact

Mood's Median Test

Mood's Median Test – Exact

Kruskal-Wallis is more powerful than Mood's Median Test, but Mood's Median is robust to outliers. Kruskal-Wallis is therefore the default recommendation, but if any sample has outliers that are Likely (2.2*IQR) or Extreme (3*IQR), then Mood's Median is recommended. See **Outliers (Boxplot Rules)** below for further details.

If Kruskal-Wallis is appropriate and each sample N <= 5, then Kruskal-Wallis – Exact is recommended (using the rules of thumb in **Sample Sizes for Exact**).

If Mood's Median is appropriate and each sample N <= 10, then Mood's Median – Exact is recommended.

Outliers (Boxplot Rules)

Each sample is tested for outliers using Tukey's Boxplot Rules: Potential (> Q3 + 1.5*IQR or < Q1 – 1.5*IQR); Likely: 2.2*IQR; Extreme: 3*IQR. If outliers are present, a warning is given and recommendation to review the data with a Boxplot and Normal Probability Plot and to consider using a Nonparametric Test.

If no outliers are found, the cell is highlighted as green:

No outliers found.

If a Potential or Likely outlier is found, the cell is highlighted as yellow:

Potential (1.5*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.

Likely (2*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.

Note that upper and lower outliers are distinguished.

If an Extreme outlier is found, the cell is highlighted as red:

> **Extreme (3*IQR) outlier lower count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test.**

The Anderson Darling normality test is applied to the sample data with outliers excluded. If this results in an AD P-Value that is greater than 0.1, a notice is given, "Excluding the outliers, data are inherently normal." The cell remains highlighted as yellow or red.

> **Extreme (3*IQR) outlier upper count = 1. It is recommended to review the data with graphical tools: Boxplot, Normal Probability Plot, Histogram and Run Chart / Control Chart. Consider using a Nonparametric Test. Note that if the outlier is removed, Anderson Darling P-Value = 0.961. Excluding the outlier, data are inherently normal.**

Randomness (Independence)

Each sample is tested for randomness (serial independence) using the Exact Nonparametric Runs Test. If the sample data is not random, a warning is given and recommendation to review the data with a Run Chart or Control Chart. See Nonparametric Runs Test for Randomness - Exact for further details. Note that for this report, the option **Values Equal to Median** is set to both **Counted as Below** and **Counted as Above**, with the reported P-Value being the larger of the two. This results in a more conservative test to minimize false alarms.

If the Exact Nonparametric Runs Test P-Value is greater than or equal to 0.05, the cell is highlighted as green:

> **Nonparametric Runs Test (Exact) P-Value = 0.396. Fail to reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is not violated.**

If the Exact Nonparametric Runs Test P-Value is less than 0.05, but greater than or equal to 0.01, the cell is highlighted as yellow:

Nonparametric Runs Test (Exact) P-Value = 0.017. Reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is violated (at 95% confidence level). It is recommended to review the data with a Run Chart or Control Chart.

If the Exact Nonparametric Runs Test P-Value is less than 0.01, the cell is highlighted as red:

Nonparametric Runs Test (Exact) P-Value = 0.004. Reject null hypothesis: "data are random," so conclude that the assumption of randomness (independence) is violated (at 99% confidence level). It is recommended to review the data with a Run Chart or Control Chart.

Equal Variance

The test for Equal Variances is applicable for two or more samples. If all sample data are normal, the F-Test (2 sample) or Bartlett's Test (3 or more samples) is utilized. If any samples are not normal, i.e., have an AD P-Value < .05, Levene's test is used. If the variances are unequal and the test being used is the equal variance option, then a warning is given and Unequal Variance (2 sample) or Welch's Test (3 or more samples) is recommended.

If the test for Equal Variances P-Value is >= .05, the cell is highlighted as green:

F-test for Equal Variances P-Value = 0.479. Fail to reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is not violated.

If the test for Equal Variances P-Value is >= .05, but the Assume Equal Variances is unchecked (2 sample) or Welch's ANOVA (3 or more samples) is used, the cell is highlighted as yellow:

F-test for Equal Variances P-Value = 0.479. Fail to reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is not violated. The Assume Equal Variances option should be checked.

If the test for Equal Variances P-Value is < .05, and the Assume Equal Variances is checked (2 sample) or regular One-Way ANOVA (3 or more samples) is used, the cell is highlighted as red:

Levene's Test for Equal Variances P-Value = 0.014. Reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is violated (at 95% confidence level). It is recommended that Welch's ANOVA be used (SigmaXL > Statistical Tools > Equal Variance Tests > Welch's ANOVA) or alternatively use a nonparametric test.

If the test for Equal Variances P-Value is < .05, and the Assume Equal Variances is unchecked (2 sample) or Welch's ANOVA (3 or more samples) is used, the cell is highlighted as green:

Levene's Test for Equal Variances P-Value = 0.014. Reject null hypothesis: "variances are equal," so conclude that the assumption of equal variances (or standard deviations) is violated (at 95% confidence level). Welch's ANOVA is appropriate.

Attribute Measurement Systems Analysis

Percent Confidence Intervals (Exact Versus Wilson Score)

Confidence intervals for binomial proportions have an "oscillation" phenomenon where the coverage probability varies with n and p. Exact (Clopper-Pearson) is strictly conservative and will guarantee the specified confidence level as a minimum coverage probability, but results in wide intervals. This is recommended only for applications requiring strictly conservative intervals. Wilson Score has mean coverage probability matching the specified confidence interval. Since the Wilson Score intervals are narrower and thereby more powerful, they are recommended for use in Attribute MSA studies due to the small sample sizes typically used [1, 2, 3].

Kappa

Kappa (κ) measures the degree of agreement between appraisers when they asses the same sample. Cohen's Kappa, used in the Attribute MSA template, is only defined for two raters. Fleiss' Kappa can be used for multiple assessors [4]. Thus Kappa provides a mechanism for evaluating the consistency of assessors.

Let $\overline{P}$ be the measure of agreement between the assessors, and $\overline{P}_e$ be the measure of agreement that would occur by chance. Then:

$$\kappa = \frac{\overline{P} - \overline{P}_e}{1 - \overline{P}_e}$$

The denominator provides the measure of agreement above chance, while the numerator provides the actual measure of agreement above chance. Thus κ is constrained between -1 and +1, with +1 implying complete consistency or perfect agreement between assessors, zero implying no more consistency between assessors than would be expected by chance and -1 implying perfect disagreement.

Rule-of-Thumb Interpretation for Kappa

Fleiss [4, p. 609] gives the following rule of thumb for interpretation of Kappa:

> Kappa: >= 0.75 or so signifies excellent agreement, for most purposes, and <= 0.40 or so signifies poor agreement.

AIAG recommends the Fleiss guidelines [5].

In Six Sigma process improvement applications, a more rigorous level of agreement is commonly used. Futrell [6] recommends:

> The lower limit for an acceptable Kappa value (or any other reliability coefficient) varies depending on many factors, but as a general rule, if it is lower than 0.7, the measurement system needs attention. The problems are almost always caused by either an ambiguous operational definition or a poorly trained rater. Reliability coefficients above 0.9 are considered excellent, and there is rarely a need to try to improve beyond this level.

SigmaXL uses the guidelines given by Futrell and color codes Kappa as follows: >= 0.9 is green, 0.7 to < 0.9 is yellow and < 0.7 is red. This is supported by the following relationship to Spearman Rank correlation and Percent Effectiveness/Agreement (applicable when the response is binary with an equal proportion of good and bad parts):

> Kappa = 0.7; Spearman Rank Correlation = 0.7; Percent Effectiveness = 85%; Percent Agreement = 85% (two trials)

> Kappa = 0.9; Spearman Rank Correlation = 0.9; Percent Effectiveness = 95%; Percent Agreement = 95% (two trials)

(Note that these relationships do not hold if there are more than two response levels or the reference proportion is different than 0.5).

Concordance

Let two assessors rank n subjects, such that x is the rankings of assessor 1, and y is the rankings assessor 2. If for any pair of observations (i, j), the ranking of the two assessors agree - that is $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$, then the pair of observations are said to be concordant. Pairs that are not concordant are said to be discordant. If x were a measurement of exercise, and y a measurement of fitness, then we would expect high levels of concordance.

Kendall's Coefficient of Concordance

Kendall's Coefficient of Concordance (Kendall's W) is a measure of association for discrete ordinal data, used for assessments that do not include a known reference standard. Kendall's coefficient of concordance ranges from 0 to 1: A coefficient value of 1 indicates perfect agreement. If the coefficient = 0, then agreement is random, i.e., the same as would be expected by chance.

There is a close relationship between ρ_s, Spearman's correlation coefficient and W, Kendall's coefficient of concordance. W can be calculated directly from the mean ($\overline{\rho}_s$) of the pairwise Spearman correlations using the relationship [7]:

$$W = \frac{(k-1)\overline{\rho}_s + 1}{k}$$

where k is the number of trials (within) or trials*appraisers (between) among which Spearman's correlation coefficients are computed.

Confidence limits for Kendall's Concordance cannot be solved analytically, so are estimated using bootstrapping. Ruscio [8] demonstrates the bootstrap for Spearman's correlation and we apply this method to Kendall's Concordance. The data are row wise randomly sampled with replacement to provide the bootstrap sample (N = 2000). W can be derived immediately from the mean value of the Spearman's correlation matrix from the bootstrap sample. The bias corrected values of α and $1-\alpha$ are then used to determine the confidence interval from the sorted bootstrap vector W. In Efron's bias-corrected and accelerated bootstrap, the value of α is adjusted to account for skew and/or bias. This has been determined to be a superior bootstrap method compared to the simple percentile method [9]. A fixed seed is used so that the report results are consistent.

Rule-of-Thumb Interpretation for Kendall's Coefficient of Concordance

SigmaXL uses the following "rule-of-thumb" interpretation guidelines: >= 0.9 very good agreement (color coded green); 0.7 to < 0.9 marginally acceptable, improvement should be considered (yellow); and < 0.7 unacceptable (red).

This is consistent with Kappa and is supported by the relationship to Spearman's correlation.

Note however that in the case of Within Appraiser agreement with only two trials, then the rules should be adjusted: very good agreement is >= 0.95 and unacceptable agreement is < 0.85.

Kendall's Correlation Coefficient

Kendall's Correlation Coefficient (Kendall's tau-b) is a measure of association for discrete ordinal data, used for assessments that include a known reference standard. Kendall's correlation coefficient ranges from -1 to 1: A coefficient value of 1 indicates perfect agreement. If coefficient = 0, then agreement is random, i.e., the same as would be expected by chance. A coefficient value of -1 indicates perfect disagreement. Kendall's Correlation Coefficient is a measure of rank correlation, similar to the Spearman rank coefficient, but uses concordant and discordant pairs and corrects for ties.

Let two assessors rank n subjects, such that x is the rankings of assessor 1, and y is the rankings of assessor 2. If for any pair of observations (i, j), the ranking of the two assessors agree - that is $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$, then the pair of observations are said to be concordant. Pairs that are not concordant are said to be discordant. If x were a measurement of exercise, and y a measurement of fitness, then we would expect high levels of concordance.

Define:

$$C_{ij} = \sum_{h<i}\sum_{k<j} x_{hk} + \sum_{h>i}\sum_{k>j} x_{hk}$$

$$D_{ij} = \sum_{h<i}\sum_{k>j} x_{hk} + \sum_{h>i}\sum_{k<j} x_{hk}$$

Then the number of concordant (n_c) and discordant (n_d) pairs are derived as:

$$n_c = 0.5P \text{ where } P = \sum_{i,j} x_{ij} C_{ij}$$

$$n_d = 0.5Q \text{ where } Q = \sum_{i,j} x_{ij} D_{ij}$$

and the Kendall Correlation τ_b is:

$$\tau_b = \frac{2(n_c - n_d)}{\sqrt{D_r D_c}}$$

where:

$$D_r = n^2 - \sum_i R_i^2$$

$$D_c = n^2 - \sum_j C_j^2$$

Rule-of-Thumb Interpretation for Kendall's Correlation Coefficient

SigmaXL uses the following "rule-of-thumb" interpretation guidelines:

>= 0.8 very good agreement (color coded green);

0.6 to < 0.8 marginally acceptable, improvement should be considered (yellow);

< 0.6 unacceptable (red).

These values were determined using Monte Carlo simulation with correlated integer uniform distributions. They correspond approximately to Spearman 0.7 and 0.9 when there are 5 ordinal response levels (1 to 5). With 3 response levels, the rule should be modified to 0.65 and 0.9.

References for Attribute Measurement Systems Analysis

[1] Agresti, A. and Coull, B.A.(1998). "Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions*." The American Statistician*, 52, 119–126.

[2] Clopper, C.J., and Pearson, E.S. (1934). "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." *Biometrika*, 26, 404–413.

[3] Newcombe, R. (1998a). "Two-sided confidence intervals for the single proportion: Comparison of seven methods." *Statistics in Medicine*,17, 857-872.

[4] Fleiss, J.L. (2003). *Statistical Methods for Rates and Proportions*, 3rd Edition, Wiley & Sons, NY.

[5] Automotive Industry Action Group AIAG (2010). *Measurement Systems Analysis MSA Reference Manual*, 4th Edition, p. 137.

[6] Futrell, D. (May 1995). "When Quality Is a Matter of Taste, Use Reliability Indexes," *Quality Progress*, 81-86.

[7] Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd Ed.). New York, NY: McGraw-Hill. p. 262.

[8] Ruscio, J. (2008). "Constructing Confidence Intervals for Spearman's Rank Correlation with Ordinal Data: A Simulation Study Comparing Analytic and Bootstrap Methods," *Journal of Modern Applied Statistical Methods*, Vol. 7, No. 2, 416-434.

[9] Bradley Efron, (1987). "Better bootstrap confidence intervals (with discussion)," *J. Amer. Statist. Assoc.* Vol. 82, 171-200.

[10] http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient

Chi-Square Tests and (Contingency) Table Associations

This section describes a number of tests and measures of association that are derived from a contingency table. These tables depict the observed counts of two variables, shown in r rows and c columns, with a row sum vector R and a column sum vector C. The total number of cells is n. In a contingency table, x_{ij} represents the number of elements that correspond to row attribute i and column attribute j.

The contingency table is described using the following summary statistics:

n_{ij}, the observed counts or number of elements in each cell.

n_i, the sum of the elements in each row.

n_j, the sum of the elements in each column.

$e_{ij} = n_i n_j / n$, the expected count or row-column product for that cell as a proportion of the total count.

$r_{ij} = n_{ij} - e_{ij}$, the raw residual or difference between the actual and expected cell count.

$s_{ij} = r_{ij} / \sqrt{e_{ij}}$, the standardized residual.

$a_{ij} = s_{ij} / \sqrt{(1 - n_i/n)(1 - n_j/n)}$, the adjusted (standardized) residual expressed as a z score.

s_{ij}^2, the cell's contribution to the chi-squared statistic.

Chi-Square Tests for Nominal Categories

Pearson Chi-squared

Pearson's chi-squared test [3] is applied to a contingency table to evaluate how likely it is that this observed table occurs by chance. Given the observed marginals (row and column counts), the expected cell observation $e_{ij} = R_i C_j / n$. The statistic:

$$\chi^2 = \sum_{i,j} (x_{ij} - e_{ij})^2 / e_{ij}$$

is distributed χ^2 with $df = (r-1)(c-1)$ degrees of freedom.

Likelihood Ratio Chi-squared

The likelihood ratio chi-squared test [3] is applied to a contingency table to evaluate the likelihood of the observed table under the null hypothesis relative to the maximum likelihood. The statistic:

$$\chi^2 = 2\sum_{i,j} e_{ij} \ln(x_{ij}/e_{ij})$$

is distributed χ^2 with df degrees of freedom.

McNemar-Bowker Symmetry

For square tables, the McNemar statistic [11] tests the hypothesis that the contingency table is symmetric. It is defined as:

$$M = \sum_{i<j} \frac{(x_{ij} - x_{ji})^2}{x_{ij} + x_{ji}}$$

for all $x_{ij} + x_{ji} > 0$. Under the null, M is distributed χ^2 with $n(n-1)/2$ degrees of freedom.

Measures of Association for Nominal Categories

Confidence bounds are evaluated as $t \pm z_\alpha s$, where z is the inverse of the standard normal cdf, t is the statistic and s is the asymptotic standard error (ASE_1) of the statistic. See [9,12] for ASE_1 formulas.

Perason's Phi

The Phi (ϕ) statistic [1,9] is defined as:

$$\phi = \sqrt{\chi^2/n}$$

where χ^2 is the Pearson Chi-squared statistic. For a 2x2 table, ϕ is a measure of association for two binary variables, and equivalent to Pearson Correlation, measures the percent of concentration of cases on the diagonal. See Cohen [15} for interpretation rules-of-thumb.

Cramer's V

The Cramer's V statistic [3, 9] measures the association between two variables as a percentage of their maximum possible variation. It has a range of $0 \le V \le 1$. It is defined as:

$$V = \sqrt{\chi^2/(n(m-1))}$$

where χ^2 is the Pearson Chi-squared statistic, and $m = \min(r, c)$

Contingency Coefficient

The contingency statistic [3] is defined as:

$$z = \sqrt{\chi^2/(n + \chi^2)}$$

where χ^2 is the Pearson Chi-squared statistic. Unlike the other statistics, it does not have a specified range.

Cohen's Kappa

For square tables, Cohen's κ statistic [3,5] tests the null hypothesis that there is no association between the row and column variables. It is calculated from the observed and expected frequencies on the diagonal of a square contingency table. It is frequently used to provide a mechanism for evaluating the consistency of assessors. It is defined as:

$$\kappa = \frac{n\sum_i x_{ii} - \sum_i R_i C_i}{n^2 - \sum_i R_i C_i}$$

A value of $\kappa = 0$ implies no association, while $\kappa = 1$ implies perfect association.

Goodman-Kruskal Lambda

Lambda [3] is defined as the proportional reduction in error of prediction of the dependent variable as a consequence of using the information in the independent variable.

Column as dependent, row as independent:

$$\lambda_r = \frac{\sum_i x_{im} - C_m}{n - C_m}$$

Row as dependent, column as independent:

$$\lambda_c = \frac{\sum_j x_{mj} - R_m}{n - R_m}$$

Symmetric version:

$$\lambda_s = \frac{\sum_i x_{im} + \sum_j x_{mj} - C_m - R_m}{2n - C_m - R_m}$$

where x_{im} and x_{mj} are the largest cell count in row i and column j respectively, and R_m and C_m are the largest row and column subtotals respectively.

Goodman-Kruskal Tau

Tau [9] is the same as lambda, except the probabilities are specified by marginal proportions.

Column as dependent, row as independent:

$$\tau_r = \frac{n\sum_{i,j} x_{ij}^2/R_i - \sum_j C_j^2}{n^2 - \sum_j C_j^2}$$

Row as dependent, column as independent:

$$\tau_c = \frac{n\sum_{i,j} x_{ij}^2/C_j - \sum_i R_i^2}{n^2 - \sum_i R_i^2}$$

Thiel's Uncertainty

Uncertainty [12] is defined as the proportional reduction in the uncertainty (entropy) of the dependent variable as a consequence of using the information in the independent variable.

Column as dependent, row as independent:

$$U_{c|r} = \frac{U_r + U_c - U_{rc}}{U_c}$$

Row as dependent, column as independent:

$$U_{r|c} = \frac{U_r + U_c - U_{rc}}{U_r}$$

Symmetric version:

$$U_s = \frac{2(U_r + U_c - U_{rc})}{U_r + U_c}$$

where:

$$U_r = -\sum_i R_i/n \ln(R_i/n)$$

$$U_c = -\sum_j C_j/n \ln(C_j/n)$$

$$U_{rc} = -\sum_{i,j} x_{ij}/n \ln(x_{ij}/n)$$

Tests for Ordinal Categories

Concordant - Discordant

Let two assessors rank n subjects, such that x is the rankings of assessor 1, and y is the rankings of assessor 2. If for any pair of observations (i, j), the ranking of the two assessors agree - that is $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$, then the pair of observations are said to be concordant [2]. Pairs that are not concordant are said to be discordant. If x were a measurement of exercise, and y a measurement of fitness, then we would expect high levels of concordance.

Define:

$$C_{ij} = \sum_{h<i}\sum_{k<j} x_{hk} + \sum_{h>i}\sum_{k>j} x_{hk}$$

$$D_{ij} = \sum_{h<i}\sum_{k>j} x_{hk} + \sum_{h>i}\sum_{k<j} x_{hk}$$

Then the number of concordant (n_c) and discordant (n_d) pairs are derived as:

$$n_c = 0.5P \text{ where } P = \sum_{i,j} x_{ij} C_{ij}$$

$$n_d = 0.5Q \text{ where } Q = \sum_{i,j} x_{ij} D_{ij}$$

The probability value returned for this test is given in **Kendall's Tau-b** below.

Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient (ρ) [3] provides a non-parametric measure of the association between the two variables. It has a range of $-1 \le \rho \le 1$. Let RI_i and CI_j be the row and column rank scores. Then ρ is evaluated as:

$$\rho = \frac{\sum_{i,j} x_{ij}(RI_i - n/2)(CI_j - n/2)}{\sqrt{(n^3 - \sum R^3)(n^3 - \sum C^3)}/12}$$

Under the null hypothesis that there is no correlation,

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}$$

has a t distribution with $n-2$ degrees of freedom.

Measures of Association for Ordinal Categories

Confidence bounds are evaluated as $t \pm z_\alpha s$, where z is the inverse of the standard normal cdf, t is the statistic and s is the asymptotic standard error (ASE_1) of the statistic. See [9,12] for ASE_1 formulas.

Spearman Rank Correlation Coefficient

Since the Spearman rank correlation coefficient is not distributed normally, the Fisher transform is used to evaluate the confidence bounds. See Bonnett & Wright [4] for details.

Kendall's Tau-b

The Kendall τ_b coefficient [10], like gamma, is a non-parametric measure of association using rank correlation, but makes a correction for ties. τ_b is defined as:

$$\tau_b = \frac{2(n_c - n_d)}{\sqrt{D_r D_c}}$$

where:

$$D_r = n^2 - \sum_i R_i^2$$

$$D_c = n^2 - \sum_j C_j^2$$

τ_b ranges from +1 to -1. A value of zero implies independence. Given *s*:

$$s = 2\sqrt{\frac{\sum_{i,j} x_{ij}(C_{ij} - D_{ij})^2 - (P-Q)^2/n}{D_r D_c}}$$ under the null hypothesis,

then τ_b/s is distributed standard normal. This is used to provide a P-Value for the **Concordant – Discordant test**.

Kendall-Stuart Tau-c

Stuart's τ_c coefficient [1] is a measure of association, but makes a correction for ties as well as table size. τ_c is defined as:

$$\tau_c = \frac{2m(n_c - n_d)}{n^2(m-1)}$$

where $m = \min(r, c)$. τ_c ranges from +1 to -1. A value of zero implies independence.

Goodman-Kruskal Gamma

The Goodman-Kruskal's gamma statistic [8] provides a measure of association using rank correlation, and is based on the number of concordant and discordant pairs. It is defined as:

$$\gamma = \frac{n_c - n_d}{n_c + n_d}$$

Somer's D

Somer's D statistic [1,14] is also a measure of association, similar to τ_b, but differs in respect to a consideration as to whether the row variable is considered the independent variable (and the column variable dependent), or *vice versa*.

Column as dependent, row as independent:

$$S_r = 2(n_c - n_d)/D_r$$

Row as dependent, column as independent:

$$S_c = 2(n_c - n_d)/D_c$$

Symmetric version:

$$S_s = 4(n_c - n_d)/(D_r + D_c)$$

References for Chi-Square Tests and (Contingency) Table Associations

[1] Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (Second Ed.), New York: John Wiley and Sons.

[2] Agresti, A. (2013). *Categorical Data Analysis* (Third Ed.), New York: John Wiley and Sons

[3] Bishop, Y, Fienberg, S and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: MIT Press.

[4] Bonett, D.G, and Wright, T.A. (2000). "Sample size requirements for estimating Pearson, Kendall and Spearman correlations", *Psychometrica* 65 (1),pp 23-28.

[5] Cohen, J. (1960). "A coefficient of agreement for nominal scale". *Educational and Psychological Measurement* 20 (1) pp. 37-46.

[6] Fisher, R.A. (1915). "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population". *Biometrika* 10 (4), pp. 507-521.

[7] Fisher, R.A. (1922). "On the interpretation of χ^2 from contingency tables, and the calculation of P". *Journal of the Royal Statistical Society* 85 (1), pp. 87-94.

[8] Goodman, L. A. and W. H. Kruskal, W.H. (1972). "Measures for association for cross-classification", *Journal of the American Statistical Association* 67. pp. 415-421.

[9] IBM SPSS Statistics 24 Algorithms, Crosstabs Algorithms, pp. 161 – 182. *ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/en/client/Manuals/IBM_SPSS_Statistics_Algorithms.pdf*

[10] Kendall, M. (1938). "A New Measure of Rank Correlation". *Biometrika* 30 (1–2) pp. 81-89.

[11] McNemar, Q.(1947)."Note on the sampling error of the difference between correlated proportions or percentages". *Psychometrika* 12 (2). pp. 153-157.

[12] SAS/STAT ® 14.1 Users Guide - Measures of Association, *http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_freq_details22.htm*

[13] Snedecor, G.W and Cochran, W.G. (1989), *Statistical Methods* 8th ed. Ames, Iowa. Blackwell Publishing.

[14] Somers, R. H. (1962). “A new asymmetric measure of association for ordinal variables”. *American Sociological Review* 27. pp. 799-811.

[15] Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (Second Ed.), Lawrence Erlbaum Associates, pp. 215 - 227.

Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests)

This set of tests is undertaken after an ANOVA or Equal Variance test, and are often referred to as "post hoc tests". After the null of equal means or variances has been rejected, the following tests are computed for each of the $.5k(k-1)$ sample pairs. Thus this permits the researcher to ascertain exactly which group(s) are responsible for the rejection of the null.

One-Way ANOVA

Fisher

Fisher's least significant difference test is a set of individual t-tests on each of the sample pairs, but uses the pooled standard deviation from all groups, which increases the power. The test statistic is:

$$t = \frac{\bar{y}_i - \bar{y}_j}{s\sqrt{1/n_i + 1/n_j}}$$

where s is the pooled standard error for all groups, and is assumed distributed Student t.

There is no adjustment for family-wise error rate, so should only be used for *k* = 3.

Tukey

Tukey's test is a set of individual t-tests on each of the sample pairs, but uses the pooled standard deviation from all group. The test statistic is:

$$t = \frac{\sqrt{2}(\bar{y}_i - \bar{y}_j)}{s\sqrt{1/n_i + 1/n_j}}$$

and is distributed as the Studentized range [1].

Dunnett

The Dunnett test compares each group to a given control group. For each group, a t statistic is calculated using the Fisher statistic. The distribution of this statistic is a $k-1$ multivariate t, with all $k-1$ arguments being the same Fisher t statistic, and with correlation matrix with off diagonal elements $\sqrt{2}$ for the balanced case, and weighted by sample size for the unbalanced case [1].

Welch's ANOVA (Unequal Variance)

Pairwise Welch

This is the pairwise two sample t-test for unequal variances. Like Fisher, there is no adjustment for family-wise error rate so should only be used for *k* = 3.

Games-Howell

The Games-Howell test does not assume equal variance for each group. The variance of each group is used. The test statistic is:

$$t = \frac{\sqrt{2}\,|\,\bar{y}_i - \bar{y}_j\,|}{\sqrt{s_i^2/n_i + s_j^2/n_j}}$$

And is distributed as the Studentized range [2].

Bartlett's Test for Equal Variance

F-Test

The standard F test statistic for testing the null of equal variances of two groups is:

$$F = \frac{s_i^2}{s_j^2}$$

where s_i^2 is the variance of group i, and is distributed as F_{n_i-1,n_j-1}.

There is no adjustment for family-wise error rate so should only be used for *k* = 3.

F-Test with Bonferroni Correction

The Bonferroni correction multiplies the F-Test probability by a factor of $.5k(k-1)$.

Levene's Test for Equal Variance

Pairwise Levene (a.k.a. Brown-Forsythe)

The pairwise Levene test statistic for testing the null of equal variances of two groups is a simple F statistic:

$$F = \frac{SSA/(k-1)}{SSE/(n-k)}$$

where $n = n_i + n_j$, and SSA and SSE are the sample sum of squares and the error sum of squares of the transformed variable respectively:

$$z = | x - median(x) |$$

The statistic is distributed as F with $k-1$ and $n-k$ degrees of freedom [3,4].

There is no adjustment for family-wise error rate, so should only be used for *k* = 3.

Tukey ADM (Absolute Deviations from Median)

Tukey's method is applied to the ADM (Absolute Deviations from Median). Conceptually, this is like an extension of the ANOMV-LEV chart, Analysis of Means version of the Levene Test. See [5] Nelson, Wludyka and Copeland (2005), page 65 for discussion on robustness.

References for Multiple Comparisons of Means and Variances

[1] J.C. Hsu, J.C. (1966). *Multiple Comparisons, Theory and methods*. Chapman & Hall.

[2] Games, P. A., and J. F. Howell. 1976. "Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study". *Journal of Educational Statistics*, 1, 113–125.

[3] Levene, H. (1960). In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

[4] Brown, M.B., and Forsythe, A.B. (1974). "Robust Tests for Equality of Variances." *Journal of the American Statistical Association* 69:364–367.

[5] Nelson, P.R., Wludyka, P.S. and Karen A. F. Copeland, K.A. (2005) *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions* ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

Analysis of Means (ANOM) Charts

A comparison of treatment means to determine which are significantly different from the overall average is commonly required. For example an educational professional might ask which schools in the system have a higher or lower rate of graduation compared to the average for all schools.

ANOM (Analysis of Means) addresses this problem for a fixed effects model. ANOM identifies the means that are significantly different from the overall mean. Analysis of means is often used in process control, since it has a graphical representation that is similar to a Shewhart chart. Typically, a central line represents the overall average. The treatment means, plotted as deviations from the overall average are compared with upper and lower decision limits to identify which are significantly different from the overall mean. The underlying assumptions are that the response belongs to a given distribution, and that the design is one way or two way.

The lower and upper decision levels (LDL, UDL) are used to ascertain whether the means at the factor level are different from the grand mean. A factor level mean that lies outside the UDL and LDL is statistically different from the grand mean at the specified confidence level.

Let $n = \sum_j n_j$ is the total number of observations, x_{ij} be the i^{th} observation at factor level j, and the mean for factor level j be:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

The grand mean is:

$$\bar{x} = \frac{\sum_{i,j} \bar{x}_{ij}}{n}$$

ANOM Normal One-Way

For a single factor with k factor levels, and with the response assumed normal, the LDL and UDL are given by:

$$LDL_j = \bar{x} - h_\alpha s_p w_j$$

$$UDL_j = \bar{x} + h_\alpha s_p w_j$$

where the pooled variance is:

$$s_p^2 = \sum_{j=1}^{k}(n_j - 1)s_j^2/(n-k)$$

where s_j^2 is the variance of factor j.

where the weighting factor for each group j is:

$$w_j = \sqrt{(n-n_j)/(nn_j)}$$

which reduces to $w_j = \sqrt{(k-1)/n}$ in the balanced case.

where h_α is the absolute value of the inverse CDF of the students t distribution at probability α^* with $df = n-k$ degrees of freedom, or is determined from lookup tables [4] (Table B.1 for balanced, Table B.3 for unbalanced) at the given probability level (α),
and where $\alpha^* = (1-(1-\alpha)^{1/k})/2$ if $k > 2$ and $\alpha^* = \alpha/2$ if $k = 2$.

ANOM Normal Two-Way

For a two factor model, with k_1 and k_2 factor levels respectively, and with the response assumed normal, the main effect is given by:

$$LDL_{fj} = \overline{x} - h_\alpha \sqrt{MSE\ w_{fj}}$$

$$UDL_{fj} = \overline{x} + h_\alpha \sqrt{MSE\ w_{fj}}$$

where LDL_{fj} and UDL_{fj} are the lower/upper bounds for factor f at level j respectively,
where MSE is derived from an ANOVA analysis with a two factor model with cross effects,
where the weighting factor for each group j is:

$$w_{fj} = (k_f - 1)/(k_f n_j)$$

which reduces to $w = (k-1)/n$ in the balanced case.

where h_α is the absolute value of the inverse CDF of the students t distribution at probability α^* with $df = k_1 k_2 (n-1)$ degrees of freedom, or is determined from Nelson's lookup tables at the given probability level (α),

and where $\alpha^* = (1-(1-\alpha)^{1/k_f})/2$ if $k > 2$ and $\alpha^* = \alpha/2$ if $k = 2$.

The significance of the interaction effect is evaluated using a standard F test:

$$f = \frac{(S_r - S_u)/(f_r - f_u)}{S_u/f_u}$$

where S_u is the unrestricted sum of squares,
where S_r is the restricted sum of squares,
and where f_u and f_r are the degrees of freedom for the unrestricted and restricted cases respectively.
f is distributed F with $f_r - f_u$ and f_u degrees of freedom.

BY method
A straight forward method of analyzing a two factor model with k_1 and k_2 groups for factor 1 and 2 respectively is to simply evaluate the one factor model with factor 1 for each value of factor 2. Thus a total of k_2 one way models runs are undertaken.

SLICE method
This method [Wludyka 2013, 2015] is the same as the BY method, but uses the pooled MSE and degrees of freedom from the full two factor model. The SLICE factor with minimum number of levels is automatically selected.

ANOM Binomial Proportions One-Way

A binomial distribution can be used for data that consists of proportions. Thus $p_j = \sum_i^{n_j} x_{ij}/n_j$.

For a single factor with k factor levels, and $N = \sum_j n_j$ observations, and with the response assumed binomial, the LDL and UDL are given by:

$$LDL_j = \overline{p} - h_\alpha s_p w_j$$

$$UDL_j = \overline{p} + h_\alpha s_p w_j$$

where the average proportion is given by:

$$\overline{p} = \frac{\sum_j n_j p_j}{N}$$

where the pooled variance is:

$$s_{p_j}^2 = \overline{p}(1-\overline{p})/n_j$$

where the weighting factor for each group j is:

$$w_j = \sqrt{(N - n_j)/(N)}$$

and where h_α is the absolute value of the inverse CDF of the normal distribution at probability $\alpha^* = 1 - \alpha/(2k)$ if $k > 2$, and $\alpha^* = 1 - \alpha/2$ if $k = 2$, or is determined from Nelson's lookup tables at the given probability level (α).

ANOM Binomial Proportions Two-Way

For a balanced two factor model, with k_1 and k_2 factor levels respectively, and with the response assumed binomial, the main effect is given by the one way factorial model, evaluated separately on each factor. Let k_1 be the number of levels of factor 1, with n observation per group.[1]

$$LDL = \overline{p} - h_\alpha s_p w$$

$$UDL = \overline{p} + h_\alpha s_p w$$

where LDL and UDL are the lower/upper bounds respectively,
where the average proportion is given by:

$$\overline{p} = \frac{\sum_j p_j}{k_1}$$

where the pooled variance is:

$$s_p^2 = \overline{p}(1 - \overline{p}/n$$

where the weighting factor is:

$$w = (k_1 - 1)/(k_1)$$

and where h_α is the absolute value of the inverse CDF of the normal distribution at probability $\alpha^* = 1 - \alpha/(2k)$ if $k > 2$, and $\alpha^* = 1 - \alpha/2$ if $k_1 = 2$, or is determined from Nelson's lookup tables at the given probability level (α).

The significance of the interaction effect is evaluated using a likelihood ratio test:

$$q = -2(\text{llf}_u - \text{llf}_r)$$

where llf_u is the unrestricted log likelihood,
where llf_r is the restricted log likelihood

[1] A similar analysis follows for factor 2.

and where the likelihood is specified as:

$$\text{llf} = \sum n\,(p\ln(\hat{p}) + (1-p)\ln(1-\hat{p}))$$

where n is the vector of sample size, p is the vector of probabilities, and $\hat{p}$ is the predicted probability given by $\exp(x\beta)/(1+\exp(x\beta))$. x is the factor matrix, and β is the maximum likelihood coefficient vector.

q is distributed χ^2 with degrees of freedom equal to the number of restrictions.

ANOM Poisson Rates One-Way

A Poisson distribution can be used for data that consists of counts. Thus the average count for group j is $c_j = \sum_i^{n_j} c_{ij}/n_j$.

For a single factor with k factor levels, and $N = \sum_j n_j$ observations, and with the response assumed distributed Poisson, the LDL and UDL are given by:

$$LDL_j = \bar{c} - h_\alpha s_p w_j$$

$$UDL_j = \bar{c} + h_\alpha s_p w_j$$

where the average count is given by:

$$\bar{c} = \frac{\sum_j n_j c_j}{N}$$

where the pooled variance is:

$$s_p^2 = \bar{c}$$

where the weighting factor for each group j is:

$$w_j = \sqrt{(N-n_j)/(Nn_j)}$$

and where h_α is the absolute value of the inverse CDF of the normal distribution at probability $\alpha^* = 1-\alpha/(2k)$ if $k > 2$, and $\alpha^* = 1-\alpha/2$ if $k = 2$, or is determined from Nelson's lookup tables at the given probability level (α).

ANOM Poisson Rates Two-Way

For a balanced two factor model, with k_1 and k_2 factor levels respectively, and with the response assumed Poisson, the main effect is given by the one way factorial model, evaluated separately on each factor.

The significance of the interaction effect is evaluated using a likelihood ratio test:

$$q = -2(\mathrm{llf}_u - \mathrm{llf}_r)$$

where llf_u is the unrestricted log likelihood,
where llf_r is the restricted log likelihood
and where the likelihood is specified as:

$$\mathrm{llf} = \sum y\ln(\lambda) - \lambda - \ln(y!)$$

where y is the vector of counts, $\lambda = \exp(x\beta + \ln(n))$, n is the vector of sample size, x is the factor matrix, and β is the maximum likelihood coefficient vector.
q is distributed χ^2 with degrees of freedom equal to the number of restrictions.

ANOM Nonparametric Transformed Ranks

When the distribution of a data set is not known, or cannot be assumed to be normal, Poisson or binomial, then a non-parametric methodology is used to compare means. Only a single factor (one way) is permitted.

The rank (r) of each element in the data set is derived. Then the rank is transformed using inverse normal scores:

$$x = \Phi^{-1}\left(\frac{0.5 + r}{2N - 1}\right)$$

where N is the number of observations, and Φ^{-1} is the inverse of the standard normal distribution function. x is then distributed normally, and the usual ANOM procedure can be applied.

ANOM Variances

ANOMV can be used to test whether the variance for any sample is significantly different from the pooled variance - it is thus a test of homogeneity. Typically, consistent output is required from a set of production lines, and ANOMV can be used to identify which line has variability significantly different from the average.

Let $N = \sum_j n_j$ be the total number of observations, x_{ij} be the i^{th} observation at factor level j, and the sample variance for factor level j be s_j^2.

One Way balanced

For a single factor with k factor levels, n observations for each group, and with the response assumed normal, the LDL, CL, and UDL are given by:

$$LDL = L_{\alpha,k,n-1} k \, MS_e$$
$$CL = MS_e$$
$$UDL = U_{\alpha,k,n-1} k \, MS_e$$

where the pooled MSE is:

$$MS_e = \sum_{j=1}^{k} s_j^2$$

and where $L_{\alpha,k,n-1}$ and $U_{\alpha,k,n-1}$ are determined from Nelson's lookup tables at the given probability level (α) and k and $n-1$ degrees of freedom.

One Way unbalanced

For a single factor with k factor levels, n_j observations for group j, and with the response assumed normal, the LDL, CL, and UDL are given by:

$$LDL_j = L^*_{\alpha,k,n_j} \frac{N-k}{n_j - 1} MS_e$$
$$CL_j = MS_e$$
$$UDL_j = U^*_{\alpha,k,n_j} \frac{N-k}{n_j - 1} MS_e$$

where the pooled MSE is:

$$MS_e = \frac{\sum_{j=1}^{k} (n_j - 1) s_j^2}{N-k}$$

where L^*_{α,k,n_j} is given by:

$$L^*_{\alpha,k,n_j} = \beta^{-1}(\alpha^*, \theta_j, \phi_j)$$

where L^*_{α,k,n_j} is given by:

$$U^*_{\alpha,k,n_j} = \beta^{-1}(1-\alpha^*, \theta_j, \phi_j)$$

and where $\alpha^* = (1-\alpha)/(2k)$, $\theta_j = (n_j - 1)/2$, $\phi_j = \sum_{i \neq j} \theta_i$, and β^{-1} is the inverse of the cdf of the beta function.

Large Sample - Balanced

For data sets with large samples - all groups have at least 35 observations, for a single factor with k factor levels, n observations for each group, and with the response assumed normal, the LDL, CL, and UDL are given by:

$$LDL = MS_e - h_{\alpha,k,\infty}\ MS_e\ w$$
$$CL = MS_e$$
$$UDL = MS_e + h_{\alpha,k,\infty}\ MS_e\ w$$

where the pooled MSE is:

$$MS_e = \sum_{j=1}^{k} s_j^2$$

where w is given by:

$$\sqrt{\frac{k-1}{k}\left(\frac{2}{n-1}+\frac{\gamma}{n}\right)}$$

where $h_{\alpha,k,\infty}$ is determined from Nelson's lookup tables at the given probability level α and k and ∞ degrees of freedom.

and where γ is an estimate of the common kurtosis - for samples distributed normal, this will be zero..

A similar formula is used for the large sample unbalanced case - see [4] for details.

ANOM Levene Robust Variances

When the distribution of a data set is not known, or cannot be assumed to be normal, then a non-parametric methodology is used to compare variances. The methodology is to first transform the data to the absolute deviation from the median for each group ADM_{ij}:

$$ADM_{ij} = | x_{ij} - md_j |$$

where md_j is the median for group j. Then, the ANOM Normal process is used on the ADM.

References for Analysis of Means (ANOM) Charts

[1] Bakir,S.T. (2010) "A Nonparametric Test For Homogeneity Of Variances: Application To GPAs Of Students Across Academic Majors" *American Journal of Business Education*, 3(3) pp. 47-54.

[2] Garrett, L and Nash, J.C. (2001). “Issues in Teaching the Comparison of Variability to Non-Statistics Students”, *Journal of Statistics Education* ((2), pp. 1-16.

[3] Nelson, L.S. (1983). "Exact Critical Values for Use with the Analysis of Means", Journal of Quality Technology, 15, pp. 40-44.

[4] Nelson, P.R., Wludyka, P.S. and Karen A. F. Copeland, K.A. (2005) *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions* ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

[5] Wludyka, P.S., (2013). “Analyzing Multiway Models with ANOM Slicing” *SESUG Working Paper SD-06* 21st Southeast SAS users Group Conference.

[6] Wludyka, P.S., (2015). “Using ANOM Slicing for Multi-Way Models with Significant Interaction”, *Journal of Quality Technology* 47(2), pp. 193-203.

[7] Wludyka, P.S. and Noguera, J.G. (2016, November). “Using Anom Slicing For Multiway Models With Binomial Or Poisson Data,” Poster session presented at INFORMS Annual Meeting, Nashville, TN.

[8] Wludyka, P.S., Bergen, A. & Liu, X. (2002) “A Monte Carlo Study of Two Normal Based Homogeneity of Variance Tests for Unbalanced Designs”, *Joint Statistical Meetings - Statistical Computing Section*, pp. 3758-3759.

[9] Wludyka, P.S. & Nelson, P.R. (1997). “An Analysis-of-Means-Type Test for Variances from Normal Populations”, *Technometrics*, 39(3), pp. 274-285

[10] Wludyka, P.S. & Nelson, P.R. (1999) “Two non-parametric, analysis-of-means-type tests for homogeneity of variances”, *Journal of Applied Statistics*, 26:2, pp. 243-256,

XYZ Contour/Surface Plot

Bivariate Interpolation for Scattered Data

A mesh is a grid of regularly spaced x- and y-values on which a 3D contour or surface plot is based. Given a set of N given distributed points, $\{x_i, y_i\}, i = 1, \dots, N$, along with data values z_i, construct a smooth function $F(x, y)$, such that $F(x_j, y_j)$ can be evaluated for points $\{x_j, y_j\}$ that are disjoint from the original given points. There are a number of scattered data methods available for interpolation, including weighted distance methods, cubic spline, and triangulation methods. The fitting function serves to provide values at points where it is impossible or expensive to obtain measurements, to provide approximations to derivatives or integrals of the underlying function, or to provide a visual representation of the data in the form of a surface plot or contour plot. For all methods, the input data consists of N points, where for each observation j, the $\{x_j, y_j, z_j\}$ values are provided. Interpolated values are values for unobserved z_k for a given $\{x_k, y_k\}$

Inverse Distance Weighting

Shepard's Inverse Distance Weighting Distance (IDW) [6] is given by:

$$z_k = \frac{\sum_{i=1}^{N} w_k z_i}{\sum_{i=1}^{N} w_k}$$

and

$$w_k = \frac{1}{\left(\sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}\right)^p}$$

where p is the power parameter. A weighted averaging of z values of points local to the interpolated point x_k, y_k is evaluated, with weights that decrease as the distance from the interpolated point increases. Increased values of p result in greater influence of points closer to the interpolated point; thus it acts as smoothing parameter.

Biharmonic Spline

Cubic splines are often used to define a smooth curve or surface that passes through a set of irregularly spaced data points. Normally one has a choice of specifying the curvature ranging from a tensor product of cubic splines for a high degree of curviness to a surface that is approximately bi-linear. An often used criteria is to require that the interpolating curve satisfies the biharmonic equation, and thus has a minimum curvature. SigmaXL uses Sandwell's algorithm [5], which employs a simpler algorithm for finding the minimum curvature surface that passes through the specified data points. Using this algorithm, the interpolated curve consists of a linear combination of Green functions of the biharmonic operator, centered at each data point, and with the amplitudes adjusted so that the interpolating function passes through the data points.

Delaunay Triangulation and Akima Polynomial

Consider making a triangular mesh of the given data points in a surface. Each data point is uniquely defined at the apex of a triangle, and each triangle is fully populated, and so consists of 3 data points. A Delaunay triangulation adds the requirement that the circumcircle of each triangle only contains the 3 data points. Delaunay triangulations maximize the minimum angle of all the angles of the triangles in the triangulation; they tend to avoid skinny triangles. There are many ways of interpolating a point within a triangle; For example, a common method is baycentric interpolation. Let the apex of the triangle be q_1, q_2 and q_3, where $q_i = \{x_i, y_i\}$, and let $q = x\{x, y\}$ be the interpolated point. Let weights a_i

$$a_1 + a_2 + a_3 = 1$$

Define:

$$a_1 = \frac{Area(\{q,q_2,q_3\})}{Area(\{q_1,q_2,q_3\})}$$

and:

$$q = a_1 q_1 + a_2 q_2 + a_3 q_3$$

Then the interpolated value is given by:

$$z = a_1 z_1 + a_2 z_2 + a_3 z_3$$

SigmaXL uses the Akima 761 algorithm [2] to undertake the interpolation. This requires a fifth degree polynomial in x and y, where the coefficients are derived from the estimation of the two first and three second degree partial differentials. The partial derivatives are derived from a cubic polynomial based on the 9 nearest points to the interpolation point. The interpolated value is then derived using the fifth degree polynomial.

"Algorithm 751: TRIPACK: A Constrained Two-dimensional Delaunay Triangulation Package", Renka, R.J., ACM Transactions on Mathematical Software, Vol. 22, No. 1, March 1996; Fortran source code is used with permission of ACM (Association for Computing Machinery) transactions on mathematical software; permission conveyed through Copyright Clearance Center, Inc., License Number 4598341298097 to SigmaXL, Inc., May 21, 2019.

“Algorithm 761: Scattered-Data Surface Fitting That Has the Accuracy of a Cubic Polynomial”, Akima, H., ACM Transactions on Mathematical Software, Vol. 22, No. 3, September 1996; Fortran source code is used with permission of ACM (Association for Computing Machinery); permission conveyed through Copyright Clearance Center, Inc., License Number 4598350123270 to SigmaXL, Inc., May 21, 2019.

References for XYZ Contour/Surface Plot

[1] Akima, H. (1975). *A Method of Bivariate Interpolation and Smooth Surface Fitting for Values given at Irregularly Distributed Points*. US Dept of Commerce. OT Report 75-70, Washington D.C.

[2] Akima, H. (1996). “Algorithm 761: Scattered-Data Surface Fitting that Has the Accuracy of a Cubic Polynomial”, *ACM Transactions on Mathematical Software*, Vol. 22(3), pp. 362–371.

[3] Renka, R.J. (1996). “Algorithm 751: TRIPACK: A Constrained Two-Dimensional Delaunay Triangulation Package", *ACM Transactions on Mathematical Software*, Vol. 22(1), pp. 1–8.

[4] Renka, R.J. and R. Brown (1999). “Algorithm 792: Accuracy Tests of ACM Algorithms for Interpolation of Scattered Data in the Plane”, *ACM Transactions on Mathematical Software*, Vol. 25(1), pp. 78-94.

[5] Sandwell, D.T. (1987). “Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data", *Geophysical Research Letters* Vol. 14(2), PP.139-142.

[6] Shepard, Donald (1968). “A two-dimensional interpolation function for irregularly-spaced data”, *Proceedings of the 1968 23rd ACM National Conference*, pp. 517–524.

Orthogonal (Deming) Regression

Introduction

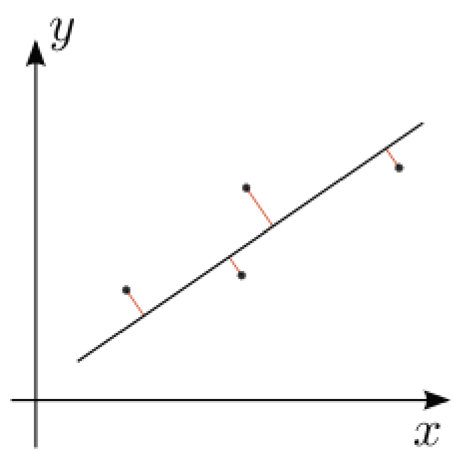
Deming Orthogonal Regression, also known as errors-in-variables regression, is a type of linear regression particularly useful when there are measurement errors in both the independent (predictor) and dependent (response) variables. For example, it can be used to compare the same parts on two measurement systems.

Ordinary Least Squares (OLS), the most commonly used form of regression, minimizes the sum of the squares of the residuals in the dependent variable only. However, in real-world scenarios, there can be considerable uncertainty in both dependent and independent variables. This is where Deming Orthogonal Regression becomes useful.

Unlike OLS, Deming Orthogonal Regression minimizes the sum of the squares of the residuals in both the dependent and independent variables. The residuals are calculated orthogonally as the perpendicular distances from the data points to the regression line, hence the term “Orthogonal Regression.”

The fundamental assumption of Deming Regression is that the ratio of the variances of the errors in the X and Y measurements is known. When this ratio is unknown, it needs to be estimated from the data, using separate Gage R&R studies or replicates.

From Wikipedia “Deming Regression”:



“The red lines show the error in both x and y. This is different from the traditional least squares method which measures error parallel to the y axis. The case shown, with deviations measured perpendicularly, arises when errors in x and y have equal variances” [1].

Formulas

An Orthogonal (Deming) Regression evaluates the simple regression between two $nx1$ variables, x and y, both of which have measurement errors. Given $\bar{x}$ and $\bar{y}$ as sample means and v_{xx}, v_{yy} and v_{xy} as the sample variance and covariance of x and y, and given the measurement error ratio $\lambda = V(\epsilon_y)/V(\epsilon_x)$ where ϵ is the respective measurement error [1,2,3]:

Slope coefficient:

$$b_1 = \frac{v_{yy} - \lambda v_{xx} + \sqrt{(v_{yy} - \lambda v_{xx})^2 + 4\lambda v_{xy}^2}}{2v_{xy}}$$

Constant (Intercept):

$$b_0 = \bar{y} - b_1\bar{x}$$

Error Variances:

$$Ve_x = \frac{v_{yy} + \lambda v_{xx} - \sqrt{(v_{yy} - \lambda v_{xx})^2 + 4\lambda v_{xy}^2}}{2\lambda}$$

$$Ve_y = \lambda Ve_x$$

Coefficient Variance:

$$V_{b_1} = \frac{\sigma_{vv}(\sigma_{xx} + Ve_x) - (b_1 Ve_x)^2}{(n-1)\sigma_{xx}^2}$$

$$V_{b_0} = \sigma_{vv}/n + \bar{x}^2 V_{b_1}$$

where

$$\sigma_{vv} = (n-1)(\lambda + b_1^2)Ve_x/(n-2)$$

$$\sigma_{xx} = \frac{\lambda v_{xx} - v_{yy} + \sqrt{(v_{yy} - \lambda v_{xx})^2 + 4\lambda v_{xy}^2}}{2\lambda}$$

Fitted values:

$$\hat{x} = x + \frac{b_1(y - b_0 - b_1x)}{b_1^2 + \lambda}$$

$$\hat{y} = b_0 + b_1\hat{x}$$

Predicted (Raw Y) Residuals:

$$\nu = y - b_0 - b_1x$$

Standardized (Raw Y) residuals:

$$\nu_s = \frac{\nu}{\sqrt{Ve_x(\lambda + b_1^2)}}$$

X Fit and Y Fit residuals:

$$\nu_{\hat{x}} = x - \hat{x}$$

$$\nu_{\hat{y}} = y - \hat{y}$$

Optimized residuals [1]:

$$\nu_o = sign(\nu)\sqrt{(x - \hat{x})^2 + (y - \hat{y})^2/\lambda}$$

Optimized residuals are orthogonal if $\lambda = 1$

Generalized R^2:

$$R_g^2 = \min(R_{gx}^2, R_{gy}^2)$$

where

$$R_{gx}^2 = 1 - \frac{\nu_{\hat{x}}'\nu_{\hat{x}} + \nu_{\hat{y}}'\nu_{\hat{y}}/\lambda}{(n-1)v_{xx}}$$

$$R_{gy}^2 = 1 - \frac{\lambda\nu_{\hat{x}}'\nu_{\hat{x}} + \nu_{\hat{y}}'\nu_{\hat{y}}}{(n-1)v_{yy}}$$

Acknowledgement

Special thanks to the authors of "Generalizing R^2 for deming regressions", Michael Bossé, Eric Marland, Gregory Rhoads, Jose Almer Sanqui & Zack BeMent, for providing R code used to validate Generalized R-Square.

References for Orthogonal (Deming) Regression

[1] Bossé, M., Marland, E., Rhoads, G., Sanqui, J.A. & BeMent, Z. (2022), "Generalizing R^2 for deming regressions", *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2022.2059678

[2] NCSS Procedures, "Deming Regression", https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Deming_Regression.pdf. Note that the error variance ratio is defined as Variance(X)/Variance(Y).

[3] Wikipedia "Deming Regression", https://en.wikipedia.org/wiki/Deming_regression.

Advanced Multiple Regression

Multiple Regression

Preprocessing of Rows with Missing Values

If a continuous or categorical predictor has a missing value, that row is automatically deleted. If a response value is missing, that row is automatically deleted, unless it is in the Test/Withhold Sample.

A general report is given: "Warning: # rows were deleted due to missing values in X or Y."

Preprocessing of Terms

The following preprocessing tests are performed prior to regression, in order to ensure that the data is sufficiently well-conditioned. These tests are undertaken on the data matrix of independent variables, and include:

- if a term causes the matrix to be rank deficient or a column has no variability
- perfect correlation (a.k.a. aliasing)

Also, for regular regression and backwards elimination (n/a for stepwise or forward selection):

- the matrix has fewer rows than columns
- severe multicollinearity, with variance inflation factor (VIF) > 1e10.

Terms are automically removed prior to doing the regression. In the case where there are multiple terms to remove, the order of precedence is last in the model, first out. A report of dropped terms is given. If terms are perfectly correlated, an alias report is also provided.

Stepwise/Best Subsets Regression

In regular regression, the independent variables, the regressors, are selected by the user. In stepwise, the user specifies the potential regressors, and the stepwise procedure, adds or removes these regressors one at a time based on the variable’s statistical significance. The stepwise process either adds the term which is most significant (largest F statistic), or removes the term that is least significant (smallest F statistic) and eventually produces a list of selected terms. It does not consider all possible regression models. The independent variables can be continuous and/or categorical.

The stepwise algorithm was first proposed by Efroymson [6]. Regressors are added or removed from the model based on the alpha-to-enter and alpha-to-remove. Terms are added or removed from the model until all variables that are not in the model have p-values greater than the specified in alpha-to-enter, and when all variables in the model have p-values that are less than (or equal to) the specified alpha-to-remove:

- Forward/Backward Stepwise: Standard stepwise using both alpha-to-enter and alpha-to-remove.
- Forward Selection: Terms are added based on alpha-to-enter, but no terms are removed.
- Backward Elimination: All terms given in the specified model are initially included, and then terms are sequentially removed based on alpha-to-remove; no terms are subsequently added.

Alternatively, criterion based selection may be used in Forward and Backward. Terms are added or removed using the F statistic, while at each stage the value of a measure, such as AICc or R-Square is monitored. If a minimum AICc is observed at step i, and this remains the minimum after 10 additional steps or the model includes all terms (forward) or 1 term (backward), then the model at the minimum AICc is selected. If a maximum R-Square is observed at step i, and this remains the maximum after 10 additional steps, then the model with the maximum R-Square is selected. Stepwise criterion options are: *AICc*, *BIC*, *R-Square Adjusted*, *R-Square Predicted* and *R-Square K-Fold*. See the Goodness-of-Fit statistics below.

Note that for R-Square K-Fold, the F-statistic to decide which term to enter or remove is based on all of the data. The K-Fold model is computed using the specified model, but a subset of the data is used as training data to estimate parameters and R-square is calculated using the out-of-sample validation data.

With Best Subsets, for any given model with p terms, there are $2^p - 1$ possible combinations (non-hierarchical models). A criterion such as AICc is specified, and the model which results in the minimum AICc is selected. If $k \leq 15$, all possible combinations are explored - this is called exhaustive. Otherwise, the best model is derived using discrete optimization with the powerful MIDACO Solver (Mixed Integer Distributed Ant Colony Optimization). See http://www.midaco-solver.com. Start values are obtained using forward selection with the AICc criterion. MIDACO does not guarantee a best solution as we have in exhaustive, but will be close to best, even for hundreds of terms!

Best Subsets criterion options are: *AICc*, *BIC* and *R-Square Adjusted*. *R-Square Predicted*, and *R-Square K-Fold* are not feasible as criterion here due to the computation times, but they are reported on the best selected models.

Best Subsets report options are: Best *For Each # of Predictors* or Best *Overall*.

Hierarchical

The Hierarchical option constrains the model so that all lower order terms that comprise the higher order terms are included in the model. This is checked by default. In Forward/Backward Stepwise and Forward Selection, a hierarchical model is required at each step, but extra terms can enter to maintain hierarchy. For Backward Elimination and Best Subsets extra terms are not permitted.

SWEEP Algorithm

Because the stepwise Ordinary Least Squares (OLS) procedure involves the addition or subtraction of a variable, it is much more efficient to augment the existing $X'X$ matrix, rather than recreate it from scratch. This process is called "Sweeping", and consists of sweeping in or out particular rows of the $X'X$ matrix [10]. SigmaXL uses it in the pre-processing test for severe multicollinearity and in forward selection prior to Best Subsets. SWEEP is also used in forward selection if the PRESS or K-Fold statistics are not computed, Box-Cox transformation is off, and there is a constant term in the model.

Coefficients and SE

The OLS coefficients are defined as:

$$\hat{\beta} = (X'X)^{-1}X'y$$

where X is the matrix of explanatory variables, and y is the vector of the dependent variable.

The vector of fitted values of the dependent variable is defined as:

$$\hat{y} = X\hat{\beta}$$

For a given observation of the independent variables, X_0:

$$\hat{y}_0 = X_0\hat{\beta}$$

$$se_{\hat{y}_0} = \sqrt{s^2(X_{0'}((X'X)^{-1}X_0)}$$

$$ci_{\hat{y}_0} = \hat{y}_0 \pm t_v \ se_{\hat{y}_0}$$

$$pi_{\hat{y}_0} = \hat{y}_0 \pm t_v \ \sqrt{s^2 + se_{\hat{y}_0}^2}$$

where $\hat{y}_0$ is the the fitted value of the dependent variable, $se_{\hat{y}_0}$ is its standard error, $ci_{\hat{y}_0}$ is its confidence interval and $pi_{\hat{y}_0}$ is its prediction interval, where s^2 is the mean square error and t_v is the quantile of the t distribution.

The residual sum of squares is defined as:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where n is the number of observations.

The standard error of the regression is an unbiased estimate of the standard deviation of the noise in the data, that is the variation in y that is not explained by the regression. It is defined as:

$$S = \sqrt{\frac{RSS}{n-p}}$$

where RSS is the residual sum of squares, n is the number of observations and p is the number of terms, including the constant (i.e., the number of columns in X).

The standard error of the coefficients provide an estimate of the variability in the estimate of the coefficient values because of the noise in the data. The coefficient variance-covariance matrix is defined as:

$$V_{\hat{\beta}} = S^2(X'X)^{-1}$$

where σ^2 is the variance of the regression. The standard error of the ith coefficient is $\sqrt{V_{ii}}$.

Residuals

The residuals are the difference between the actual value and fitted value of the dependent variable.

The raw residual for the ith observation is defined as

$$e_i = y_i - \hat{y}_i$$

Standardized residuals are divided by their standard deviation, such that they have zero mean and unit standard deviation. The ith standardized residual is defined as:

$$r_i = \frac{e_i}{S\sqrt{1-h_{ii}}}$$

where S is the standard error of the regression, and H is the "Hat" matrix $H = X(X'X)^{-1}X'$. The ith diagonal element of H, h_{ii} is called the leverage of the ith observation.

Studentized (Deleted t) residuals are the standardized residuals evaluated on the sample with the ith observation deleted. The ith studentized residual is defined as:

$$t_i = \frac{r_i\sqrt{n-p-1}}{\sqrt{n-p-r_i^2}}$$

where r_i is the standardized residual, n is the sample size and p is the number of terms, including the constant (i.e., the number of columns in X).

Diagnostic Meaures

Cook's Distance is a diagnostic measure that is used to indicate influential points in a regression and is a product of the standardized residual squared and the leverage factor $\sqrt{h_{ii}/(1-h_{ii})}$. Cook's distance for the ith observation is defined as:

$$d_i = \frac{r_i^2}{p}\frac{h_{ii}}{1-h_{ii}}$$

DFITS (Difference in Fits) is similar to Cook's distance and is used to indicate influential points in a regression. A cutoff value of $2\sqrt{p/n}$ is often used. It is defined as the product of the studentized residual and the leverage factor $\sqrt{h_{ii}/(1-h_{ii})}$. DFITS for the ith observation is defined as:

$$DFITS_i = t_i\sqrt{\frac{h_{ii}}{1-h_{ii}}}$$

Goodness-of-Fit Statistics

The ordinary least squares (OLS) Log Likelihood Function (LLF) is defined as:

$$LLF = -0.5n(1 + \ln(2\pi) + \ln(RSS/n))$$

where n is the sample size, and RSS is the residual sum of squares $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ (a.k.a. sum-of-squares error) from the current fitted model.

Given a set of candidate models for the data, the preferred model is the one with the minimum Information Criteria. The Information Criteria rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting, because increasing the number of parameters in the model almost always improves the goodness of the fit.

AICc (Akaike Information Criterion Corrected) is a measure of prediction error with a correction term for small sample size adjustment. It is defined as:

$$AICc = -2LLF + 2(p+1) + \frac{2(p+1)(p+2)}{(n-p-2)}$$

where LLF is the log likelihood from the current fitted model, n is the sample size, and p is the number of parameters in the model design matrix, including the constant. $AICc$ is generally recommended for model comparison, as it typically results in the model with the best out of sample prediction accuracy. Asymptotically, the Akaike Information Criterion is equivalent to Leave-One-Out-Cross-Validation (i.e., PRESS – see below) [18]. See also https://en.wikipedia.org/wiki/Akaike_information_criterion.

BIC (Bayesian Information Criterion) is defined as:

$$BIC = -2LLF + (p+1)\ln(n)$$

where LLF is the log likelihood from the current fitted model, n is the sample size, and p is the number of parameters in the model design matrix, including the constant. This is a stronger penalty for extra terms in the model than that of AICc, so will result in a parsimonious model that is more likely to be the “true” model.

R-Square is defined as:

$$R^2 = 1 - \frac{RSS}{TSS}$$

where RSS is the residual sum of squares (a.k.a. sum-of-squares error) $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ and TSS is the total sum of squares $\sum_{i=1}^{n} (y_i - \bar{y})^2$.

R-Square Adjusted includes a penalty for the number of terms in the model design matrix. It is defined as:

$$R^2_{adj} = 1 - \frac{RSS/(n-p)}{TSS/(n-1)} = 1 - \frac{MSE}{MST}$$

where n is the sample size, MSE is the mean square error, MST is the mean square total and p is the number of parameters in the model design matrix, including the constant.

A K-fold cross-validation model involves estimating the coefficients on a subset of the data - the training set - and then evaluating the statistical metrics on the remaining data - the validation set. This is repeated for each of the K-fold validation sets. The total k-fold RSS is the sum of the RSS for each of the response sets RSS_k. The R-Square K-fold is defined as:

$$R^2_{kfold} = 1 - \frac{\sum_{k=1}^{n_k} RSS_k}{TSS}$$

where TSS is the total sum of squares for all of the data.

The K-Fold Standard Deviation is evaluated using:

$$S_{kfold} = \sqrt{\frac{\sum_{k=1}^{n_k} RSS_k}{n}}$$

R-Square Test is calculated as:

$$R^2_{test} = 1 - \frac{RSS_{test}}{TSS_{test}}$$

The Test Standard Deviation is evaluated using:

$$S_{test} = \sqrt{\frac{RSS_{test}}{n_{test}}}$$

Note, R-Square Test is not available as a criterion to optimize because this would bias the test. It must be a “blind” test to be effective.

R-Square Predicted indicates how well a regression model predicts responses for new observations. A *fitted model* having been produced, each observation in turn is removed and the model is refitted using the remaining observations. The out-of-sample predicted value is calculated for the omitted observation in each case, and the PRESS (Predicted Residual Error Sum of Squares) statistic is calculated as the sum of the squares of all the resulting prediction errors:

$$PRESS = \sum_{i=1}^{n} \left(y_i - \hat{y}_{i(i)}\right)^2$$

However, given the model:

$$y = X\beta + \varepsilon$$

the leave one out cross validation can be efficiently computed using the “Hat” matrix $H = X(X'X)^{-1}X'$. The PRESS (Predicted Residual Error Sum of Squares) is calculated as:

$$\mathrm{PRESS} = \sum_{i=1}^{n} \left(\frac{\varepsilon_i}{1 - H_i}\right)^2$$

where H_i is the ith diagonal element of H (a.k.a. the leverage of the ith observation).

R-Square Predicted is given as:

$$R^2_{pred} = 1 - \frac{PRESS}{TSS}$$

Mallows' Cp is a measure that helps establish the number of predictors in a model. An unbiased model has the Mallows' Cp close to the number of parameters; it is defined as:

$$Mallows\ Cp = RSS/MSE_f - n + 2p$$

where RSS is residual sum of squares from the current fitted model, MSE_f is the mean square error for the full model with all terms, n is the sample size, and p is the number of parameters in the model design matrix, including the constant.

The Condition Number measures whether a model is well conditioned. An ill conditioned model will have a large change in coefficient values for a small change in the input data, such as occurs in multicollinearity. It is defined as:

$$CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

where λ are the characteristic roots evaluated on the $X'X$ matrix, where each column of X is scaled by the square root of its norm. A rule of thumb from Montgomery, Peck, and Vining [16] is that $CN > 100$ indicates moderate multicollinearity.

VIF is used to measure the effect of multicollinearity within the predictor variables. For each predictor, the VIF is calculated by regressing that predictor against all the other predictors and estimating the R-squared of that regression. The VIF for predictor i is given by

$$VIF_i = \frac{1}{1 - R^2_{x_i}}$$

A common rule of thumb is that $VIF > 5$ implies unacceptably high multicollinearity.

Breusch-Pagan Test for Constant Variance

The Breusch Pagan test evaluates whether the residuals in an estimated equation are homoskedastic by undertaking an auxiliary regression of the squared residuals against a set of explanatory variables, either with all terms in the model or individually included in the model. The BP statistic is distributed χ^2 with degrees of freedom equal to the number of parameters [1]. The Koenker studentized version is also supported, with a BP statistic equal to nR^2 [11]. The Anderson-Darling normality test is applied to the residuals in order to automatically select Normal or Koenker Robust (if AD P-Value < .05). See also https://en.wikipedia.org/wiki/Breusch%E2%80%93Pagan_test.

Box-Tidwell Test & Power Transformation Recommendation for Continuous Predictors

Multiple linear regression assumes that relationships between the predictors and the response variable are linear. If the relationship between a continuous predictor and the response variable is nonlinear, one remedy is to apply a power transformation to the predictor(s) so that the transformed predictors have a linear relationship with the response variable. This transformation can be crucial for improving the model fit and prediction accuracy.

The Box-Tidwell procedure [1] aims to find an optimal power transformation of the predictor variables to satisfy the linearity assumption. For a single predictor X:

$$X_{\text{transformed}} = X^{\gamma}$$

Where:

- γ is the transformation parameter, intended to linearize the relationship between X and the dependent variable Y.
- When γ approaches zero, the transformation shifts to a natural logarithmic scale:

$$\lim_{\gamma \to 0} X^{\gamma} = \ln{(X)}$$

The steps for the Box-Tidwell test and transformation for a single predictor are as follows:

- **Original Regression Model**:
    - Fit a linear regression model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
    - $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimated coefficients for the intercept and the slope of X, respectively.
- **Extended Regression Model with Interaction Term**:
    - Enhance the original model by adding an “interaction” term $X\ln{(X)}$: $\hat{Y} = \hat{\beta}_0^* + \hat{\beta}_1^* X + \hat{\beta}_2^*(X\ln{(X)})$
    - $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and $\hat{\beta}_2^*$ are the adjusted coefficients in the extended model, where $\hat{\beta}_2^*$ is specifically for the interaction term.
- **Statistical Testing**:
    - Perform a hypothesis test on $\hat{\beta}_2^*$.
    - The null hypothesis H_0: $\hat{\beta}_2^* = 0$ tests if the interaction term is necessary for improving model linearity.
- **Estimate γ**:

- If the null hypothesis is rejected, suggesting a transformation is beneficial, compute γ using: $\gamma_{\text{estimated}} = 1 + \frac{\hat{\beta}_2^*}{\hat{\beta}_1}$
- $\hat{\beta}_1$ from the original model is used to ensure that γ is calculated considering the primary impact of X.

- **Refit the Model and Iterate**:
  - Transform X using the estimated γ to create $X_{\text{transformed}}$.
  - Refit the model to assess improvements: $\hat{Y} = \hat{\beta}_0'' + \hat{\beta}_1'' X_{\text{transformed}}$
  - Iterate until the estimate of the transformation parameter stabilizes

- **Evaluation**:
  - Use residual analysis and R-Square Predicted values to evaluate the model fit. Box-Tidwell assumes that the error variance is constant (homoscedasticity).
  - Montgomery et al. [16] note that convergence problems may be encountered in cases where the error standard deviation σ is large or when the range of the regressor is very small compared to its mean. This situation implies that the data do not support the need for any transformation.

- **Inverse Transformation**:
  - To retrieve the original value of X from $X_{\text{transformed}}$, use the following inverse transformation formula:

    If γ is not zero, the inverse formula is:

    $$X = (X_{\text{transformed}})^{1/\gamma}$$

    If γ is zero, implying a logarithmic transformation was applied ($\ln(X)$), the the inverse transformation is:

    $$X = \exp(X_{\text{transformed}})$$

The steps for multiple predictors as outlined by Fox [8, pp. 326-329] are as follows:

1. **Initial Regression**: Conduct an initial regression of the response variable Y on the independent variables $X_1, X_2, \ldots, X_k$, obtaining coefficients $A, B_1, B_2, \ldots, B_k$.

2. **Construct Variables and Second Regression**: Perform a second regression of Y on both the original independent variables $X_1, X_2, \ldots, X_k$ and newly constructed variables $X_1 \ln X_1, X_2 \ln X_2, \ldots, X_k \ln X_k$, obtaining new coefficients $A', B_1', B_2', \ldots, B_k'$ and $D_1, D_2, \ldots, D_k$. The constructed variables are derived from a linear approximation (first-order

Taylor series) to the function $X_j^{\gamma_j}$ evaluated at $\gamma_j = 1$, where γ_j represents the power to which variable X_j is raised.

3. **Assessment for Need of Transformation**: The necessity for transforming a variable X_j is determined by testing the null hypothesis $H_0: \delta_j = 0$, where δ_j is the coefficient of the constructed variable $X_j \ln\ X_j$ in the regression model.

4. **Preliminary Estimate of Transformation Parameter**: A preliminary estimate of the transformation parameter γ_j, not the maximum likelihood estimate (MLE), is calculated as $\hat{\gamma}_j = 1 + \frac{D_j}{B_j}$, where B_j is the coefficient from the initial regression.

5. **Iteration for MLEs**: The procedure, from the initial regression through the calculation of the preliminary estimate of γ_j, may be iterated until the estimates of the transformation parameters stabilize. These final estimates approximate the MLEs, $\hat{\gamma}_j$.

Note, in SigmaXL:

- At least one continuous predictor with all positive data values must be included in the model.
- Do not use standardization or coding as this will introduce 0 or negative values in the predictors.
- Continuous predictors with values <= 0, categorical factors, interactions and higher order terms are included in the model but excluded from the Box-Tidwell (BT) test and transformation.
- The constant must be included in the model.
- Box-Tidwell cannot be performed with Error df = 0.
- Box-Tidwell power transformations are calculated only for significant continuous predictors detected by the BT Test. This improves the overall robustness of the procedure.
- If Box-Cox is used, Box-Tidwell power transformations are calculated using the Box-Cox transformed response. Box-Cox Lambda may not be optimal after refit.
- Optimal and rounded power values are reported. Rounded is recommended for ease of interpretation.
- Power values of -5 or +5 are limits and considered unstable, so rounded is set to 1.
- Sheet BoxTidwell contains the original data with new columns for the transformed continuous predictors using rounded power. The model should be refit with these transformed predictors. If optimal power is desired, please use the Excel formula "=X^(Power)"; if Power = 0, use "=LN(X)".

Durbin-Watson Test for Autocorrelation

The Durbin Watson test is used to detect the presence of positive or negative residual autocorrelation of lag 1. Specifically, if the residual $\epsilon_t = \rho\epsilon_{t-1} + \upsilon_t$, then the null hypothesis is that $\rho = 0$. The DW test statistic on a sample with T observations is:

$$DW = \frac{\sum_{t=2}^{T}(\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{T}\epsilon_t^2}$$

Under the assumption of normally distributed disturbances, the null distribution of the DW statistic is the distribution of a linear combination of chi-squared variables. The p-values for the sum of chi-squared variables is based on the Farebrother Applied Statistics Algorithm AS 153 [4][7]. See also https://en.wikipedia.org/wiki/Durbin%E2%80%93Watson_statistic.

White Robust Standard Errors for Non-Constant Variance (Heteroskedasticity-Consistent)

The standard errors used in the coefficient table, Anova F-tests and confidence/prediction intervals are predicated on the estimate of the parameter covariance matrix, Ω. X is the design matrix, ε_i is the ith estimated residual, σ^2 is the OLS estimated residual variance, $\Sigma = \sum_{i=1}^{n}(x_i x_i' \varepsilon_i^2)$ is the White corrected residual variance [8][18] :

OLS: $\Omega = (X'X)^{-1}\sigma^2$
White: $\Omega = (X'X)^{-1}\Sigma(X'X)^{-1}$

MacKinnon and White [14] improved on the small sample properties using a finite sample adjustment $n/(n-k)$ and leverage $H = 1 - D(X(X'X)^{-1}X')$, where $D(W)$ is the diagonal matrix of W:

HC1: $\Omega = n(X'X)^{-1}\Sigma(X'X)^{-1}/(n-k)$
HC2: $\Omega = (X'X)^{-1}\sum_{i=1}^{N}(x_i x_i' \varepsilon_i^2)/H_{ii}(X'X)^{-1}$
HC3: $\Omega = (X'X)^{-1}\sum_{i=1}^{N}(x_i x_i' (\varepsilon_i/H_{ii})^2)(X'X)^{-1}$

Long and Ervin [13] recommend HC3 based on Monte Carlo simulation and this is what is used in SigmaXL. For comparison and validation, this matches the hccm function in the Fox and Weisburg R package ‘car’ with default type = “hc3”. See also https://en.wikipedia.org/wiki/Heteroscedasticity-consistent_standard_errors.

Newey-West Robust Standard Errors for Non-Constant Variance with Autocorrelation (Heteroskedasticity and Autocorrelation-Consistent)

The finite sample adjusted Newey-West covariance matrix [8][17][21] is:

Newey-West: $\Omega = n(X'X)^{-1}\Sigma_{nw}(X'X)^{-1}/(n-k)$

where:

$$\Sigma_{nw} = \sum_{i=1}^{n}(x_i x_i' \varepsilon_i^2) + \sum_{i=1}^{h}(1 - i/(1+h))\sum_{j=i+1}^{n}\varepsilon_j\varepsilon_{j-1}(X_j, X_{j-1} + X_{j-1}{}'X_j),$$

and $h =$ lag number.

Note that for lag 0, Newey West is equivalent to White HC1. SigmaXL automatically uses the lag 1 Newey-West formula if either of the Durbin-Watson P-Values are < .05 (i.e., significant positive or negative autocorrelation).

For comparison and validation, this matches the NeweyWest function in the Zeileis R package 'sandwich' with lag=1, adjust=TRUE, and prewhite=FALSE. Currently, support for other options are not available in SigmaXL.

See also https://en.wikipedia.org/wiki/Newey%E2%80%93West_estimator.

Box-Cox Transformation

The classical Box-Cox procedure transforms the dependent variable vector y as follows:

$$y^* = \frac{(y-\tau)^{\lambda}-1}{\lambda}$$

where $\tau < y_{min}$ is a threshold parameter. The object of this procedure is to transform non-normal residuals data into data that is closer to normal.

The parameters λ and τ are estimated using the selected model. This is undertaken by maximizing the log-likelihood of the Box-Cox transform with respect to λ and τ, using Sequential Quadratic Programming (SQP), where τ is constrained such that $y_{min} + |\tau| > 0$, and where the log-likelihood is defined as:

$$LLF = -.5n\ln(\sigma^2) + (\lambda - 1)\sum \ln(y)$$

where $\sigma^2 = (\sum y_{bc}^2 - y_{bc}'x(x'x)^{-1}x'y_{bc})/n$, and x is the data for the selected model.

The minimum threshold value τ is specified as $y_{min} - \delta$, where $\delta = \min(0.001, 0.001(y_{max} - y_{min}))$.

After λ is solved (or specified by the user), SigmaXL uses a simplified transformation of the dependent variable using a power function:

$$y^* = (y - \tau)^{\lambda}$$

If λ is zero then $y^* = \ln(y)$. λ is constrained to be >= -5 and <= 5. If Rounded Lambda is selected, λ is rounded to the nearest value of: -5, -4, -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 5. The user may specify a λ value greater than +5 or less than -5, but this is not recommended.

In a stepwise context, if λ is not specified, it is estimated using an iterative process:

1. The stepwise process is undertaken without the Box-Cox transform
2. λ is estimated using the selected model
3. The stepwise process is undertaken with the Box-Cox transform using this λ on the entire model

4. Steps 2 and 3 are repeated until convergence or a maximum of 5 iterations is reached.

For Best Subsets, λ is solved using forward stepwise, and then held fixed for Best Subsets.
For prediction, the inverse transformation is applied:

$$y = (y^*)^{1/\lambda} + \tau$$

or, if $\lambda = 0$:

$$y = \exp(y^*) + \tau$$

Lenth Pseudo-Standard Error

For saturated models, Lenth's method [12] is used to identify inactive effects from which it constructs an estimate of the residual standard error, known as the Lenth Pseudo Standard Error (PSE). This can be used to estimate the standard error where contrasts are assumed independent. For each contrast, a t ratio is computed by dividing the contrast by the PSE. Since the distribution of the t ratio is not analytic, the probability is evaluated using Monte Carlo simulation [20]. Student T p-values are also available for comparison purposes. An enhancement by Edwards [5] draws from the multivariate distribution of the contrasts, using the covariance matrix.

Multiple Response Optimization

Multiple Response Optimization concerns mathematical optimization problems involving more than one response (objective function) to be optimized simultaneously. To provide a unique optimal solution, the problem is formulated as a single-objective optimization problem by maximizing the Derringer and Suich desirability index [3]. For each output i, a desirability is assigned within a range of zero (unacceptable) and unity (ideal). The individual desirabilities are combined using a geometric mean to create a composite desirability:

Let d_i be the ith individual desirability function, L_i, T_i and U_i be the lower, target and upper limits respectively, let r_i be a weight, and let m_i be an importance. Then, for a response y_i which is a maximum (larger the better):

$$d_i = 0,\ y_i < L_i$$
$$d_i = (\tfrac{y_i - L_i}{U_i - L_i})^{r_i}, L_i \le y_i \le U_i$$
$$d_i = 1,\ y_i > U_i$$

For a response y_i which is a minimum (smaller the better):

$$d_i = 1,\ y_i < L_i$$
$$d_i = (\tfrac{U_i - y_i}{U_i - L_i})^{r_i},\ L_i \le y_i \le U_i$$
$$d_i = 0,\ y_i > U_i$$

For a response y_i which is to be as close to the target (nominal is best):

$$d_i = 0,\ y_i < L_i$$
$$d_i = (\tfrac{y_i - L_i}{T_i - L_i})^{r_i},\ L_i \le y_i \le T_i$$
$$d_i = (\tfrac{U_i - y_i}{U_i - T_i})^{r_i},\ T_i \le y_i \le U_i$$
$$d_i = 0,\ y_i > U_i$$

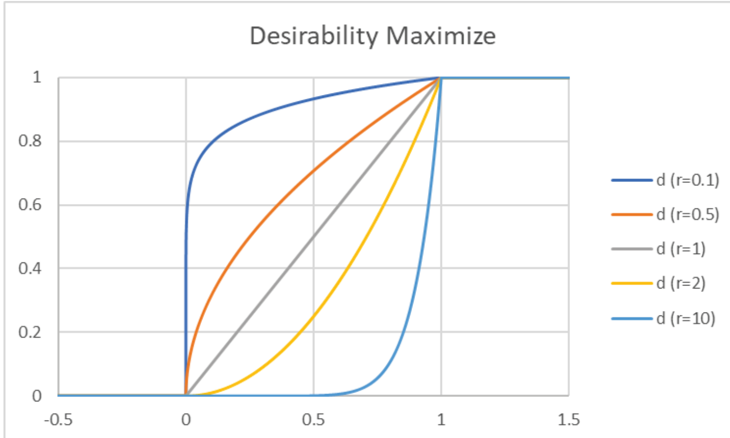
The overall desirability measure, D, is defined for n responses as the geometric mean:

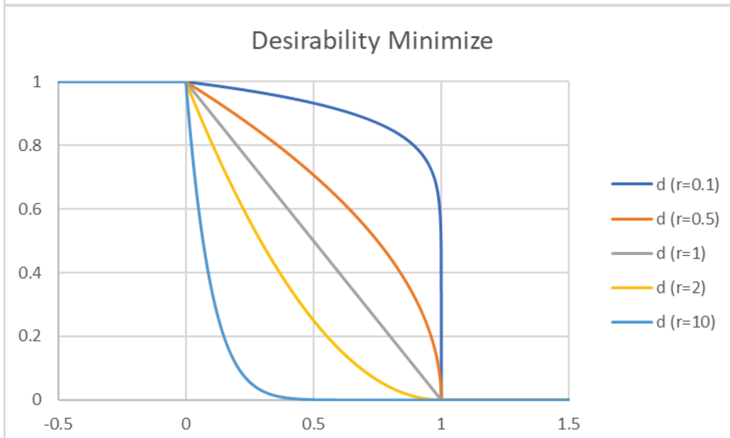
$$D_i = (\prod (d_i{}^{m_i}))^{1/\sum(m_i)}$$

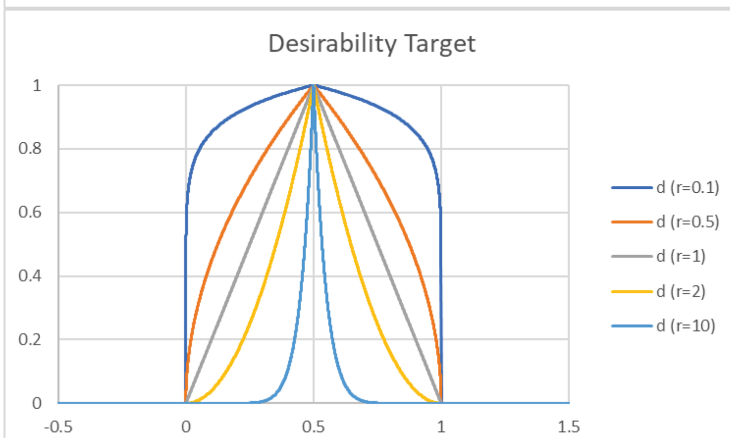
The upper and lower limits can be specification limits or practical limits for maximization/minimization. For maximization, a practical upper limit would be a value beyond which an increase in desirability is not beneficial. For minimization, a practical lower limit is a value below which an increase in desirability is not beneficial. Unlike other software packages, SigmaXL does not use Target in maximization or minimization. This is in agreement with the original one-sided formula given in Derringer and Suich.

Importance m_i allows for weighting of the responses according to relative importance. Use a value between 0.1 and 10.

Weight r_i is a shape factor. For $r_i = 1$ (default) the desirability function increases linearly. For $r_i < 1$, the function is concave, so low priority on hitting the target or goal. For $r_i > 1$, the function is convex, so high priority on hitting the target or goal. Use a value between 0.1 and 10.







SigmaXL provides two options for multiple response optimization: Multistart Nelder-Mead Simplex and MIDACO. Nelder-Mead is derivative free so can accommodate the non-smooth desirability response with both continuous and categorical predictors. It is very fast but gives a local solution. Multistart helps to improve the chances of finding a global solution. If there are a number of equally valid best solutions, the one with the smallest predicted response standard error is chosen.

MIDACO is a global optimizer so would be more suitable for a complex problem with multiple local solutions, but is slower than Nelder-Mead.

Single Response Optimization

For single response optimization, SigmaXL uses Multistart (100 starts) Nelder-Mead Simplex. If the goal is "Target", the metric to minimize is the deviation of response around the target squared (i.e., quadratic loss): $(y - T)^2$.

If the goal is "Maximize" or "Minimize", Nelder-Mead searches for the maximum or minimum response value.

Continuous predictors can optionally be constrained to a range or specified as integers. Categorical predictors can be held to a specified level.

If there are a number of equally valid best solutions, the one with the smallest prediction standard error is chosen.

References for Advanced Multiple Regression

[1] Box, G.E. P. and P.W. Tidwell (1962). "Transformation of the independent variables" *Technometrics* Vol. 4, pp. 531-550.

[2] Breusch, T. S. and A.R. Pagan, (1979). "A Simple Test for Heteroskedasticity and Random Coefficient Variation". *Econometrica*, Vol 47 (5): pp. 1287-1294.

[3] Derringer, G. and Suich, R. (1980) "Simultaneous Optimization of Several Response Variables". *Journal of Quality Technology*, Vol 12, pp. 214-219.

[4] Durbin, J. and G.S. Watson, (1950), "Testing for Serial Correlation in Least Squares Regression I." *Biometrika* Vol. 37, pp. 409-428.

[5] Edwards, D.J. and R.W. Mee (2008). "Empirically Determined p-Values for Lenth t-Statistics", *Journal of Quality Technology*, Vol. 4(4), pp. 368-380.

[6] Efroymson, M. A. (1960). "Multiple Regression Analysis", in A. Ralston and H.S. Wilf (Eds.), *Mathematical Methods for Digital Computers*, Wiley, New York, pp. 191-203.

[7] Farebrother, R.W. (1980), "Pan's Procedure for the Tail Probabilities of the Durbin-Watson Statistic", *Applied Statistics* Vol. 29, pp. 224-227.

[8] Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*, Third Edition. Sage.

[9] Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*, Third Edition, Sage. https://cran.r-project.org/web/packages/car/index.html.

[10] Goodnight, J. (1979), "A Tutorial on the SWEEP Operator", *The American Statistician* Vol. 33(3), pp.149-158.

[11] Koenker, R. (1981). "A Note on Studentizing a Test for Heteroscedasticity". *Journal of Econometrics*, Vol. 17: pp. 107-112.

[12] Lenth, R.V. (1989). "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, Vol 31, pp. 469-473.

[13] Long, J. S. and Ervin, L. H. (2000). "Using heteroscedasticity consistent standard errors in the linear regression model," *The American Statistician* Vol. 54, pp., 217–224.

[14] MacKinnon, J. G., and White, H. (1985). "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, Vol. 29, pp. 53-57.

[15] Miller, A. J. (1996). "The convergence of Efroymson's stepwise regression algorithm." *The American Statistician*, 50 (2), pp. 180-181.

[16] Montgomery, D.C. , E.A. Peck and G.G. Vining (2021). *Introduction to Linear Regression Analysis*, 6th Edition, John Wiley & Sons.

[17] Newey, W.K., and K.D. West, (1987). "Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix" *Econometrica*, Vol. 55(3), pp. 703-708.

[18] Stone M. (1977). "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion". *Journal of the Royal Statistical Society Series B*. 39, 44–7.

[19] White, H. (1980). "A heteroskedastic consistent covariance matrix estimator and a direct test of heteroskedasticity". *Econometrica* Vol. 48, pp. 817–838.

[20] Ye, K.Q. and Hamada, M. (2000). "Critical Values of the Lenth Method for Unreplicated Factorial Designs". *Journal of Quality Technology* Vol. 32, pp. 57–66.

[21] Zeileis A (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators."

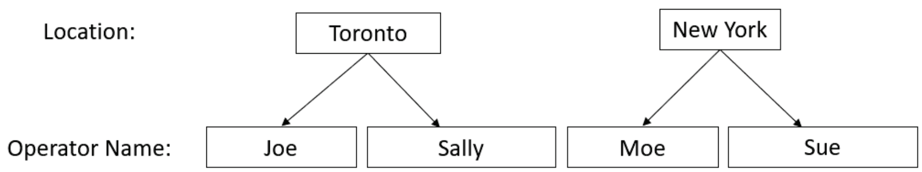
Journal of Statistical Software, 11(10), 1–17. https://www.jstatsoft.org/article/view/v011i10 and https://cran.r-project.org/web/packages/sandwich.

General Linear Model

Nesting

Check **Nesting** in the General Linear Model dialog to include nested terms in the model. The left-side combo dropdown selection can be a Factor or Covariate. The right-side dropdown selection is a Factor. This will create a term with the notation A(B) where A is nested in B.

A simple example of a nested term would be Operator(Location) where the operators in each plant location are different, operator is nested within location:



A maximum of three levels of nesting are permitted with up to four factors. In the model, this would appear as "D, C(D), B(C(D)), A(B(C(D)))" or "A, B(A), C(B(A)), D(C(B(A)))" depending on the nesting assignment. One should exercise caution when using three levels of nesting as the number of coefficients in the model is equal to the number of factor combinations = $levels_A*levels_B*levels_C*levels_D$.

A factor cannot be nested within itself, either directly or indirectly, so model "A(B), B(A)" or "A(B), B(C), C(A)" would be illegal.

Note, "B(A), C(A)" is legal, but "A(B), A(C)" is illegal. A workaround to this limitation would be to create a new factor "B_C" by concatenating the levels of B and C and using "A(B_C)".

If B is nested in A, or A nested in B, i.e., B(A) or A(B), the interaction term A*B is not permitted. SigmaXL automatically excludes these interactions in the Specify Model Terms dialog. Interactions involving A or B with other factors are permitted.

Term hierarchy in a model is recommended, but not required.

Nested Factors and Coding

The collected data from the simple Operator(Location) example are given as:

Location	Operator	Y
Toronto	Joe	1
Toronto	Sally	2
New York	Moe	3
New York	Sue	4

Coding for Categorical Factors (-1, 0, +1), also known as effects coding, estimates the difference between each factor level mean and the overall mean. The reference level is the last alpha-numerically sorted level and is hidden in the Parameter Estimates table.

The model/design matrix (stored in the Residuals sheet) is:

Constant	Location_New York	Operator(Location)_Moe_New York	Operator(Location)_Joe_Toronto
1	-1	0	1
1	-1	0	-1
1	1	1	0
1	1	-1	0

The Parameter Estimates are:

Predictor Term	Coefficient
Constant	2.5
Location_New York	1
Operator(Location)_Moe_New York	-0.5
Operator(Location)_Joe_Toronto	-0.5

Toronto is the hidden reference for Location. Sue is the hidden reference for Operator in New York and Sally is the hidden reference for Operator in Toronto.

The prediction equation (Y) as an Excel formula is:

```
= (2.5)
+ (1)*(IF(Location="New York",1,0)+IF(Location="Toronto",-1,0))
+ (-0.5)*(IF(Operator="Moe",1,0) + IF(Operator="Sally",-1,0) +IF(Operator="Sue",1,0))*IF(Location="New York",1,0)
+ (-0.5)*(IF(Operator="Joe",1,0) + IF(Operator="Sally",-1,0) + IF(Operator="Sue",-1,0))*IF(Location="Toronto",1,0)
```

Coding for Categorical Factors (1, 0), also known as dummy coding, has the hidden reference value as the first alpha-numerically sorted level.

The model/design matrix (stored in the Residuals sheet) is:

Constant	Location_Toronto	Operator(Location)_Sue_New York	Operator(Location)_Sally_Toronto
1	1	0	0
1	1	0	1
1	0	0	0
1	0	1	0

The Parameter Estimates are:

Predictor Term	Coefficient
Constant	3
Location_Toronto	-2
Operator(Location)_Sue_New York	1
Operator(Location)_Sally_Toronto	1

New York is the hidden reference for Location. Moe is the hidden reference for Operator in New York and Joe is the hidden reference for Operator in Toronto.

The prediction equation (Y) as an Excel formula is:

= (3)
+ (-2)*(IF(Location="Toronto",1,0))
+ (1)*(IF(Operator="Sue",1,0))*IF(Location="New York",1,0)
+ (1)*(IF(Operator="Sally",1,0))*IF(Location="Toronto",1,0)

Both coding schemes will have the same prediction values, but **(-1,0,1)** is preferred for GLM since It results in lower multicollinearity VIF scores than **(1, 0)** coding. **(1,0)** coding is popular in multiple regression because it is easier to interpret.

Nested Covariates

A simplified example of a nested covariate is X(Location) where the covariate (continuous predictor) X is nested within location.

The collected data are given as:

Location	X	Y
Toronto	1	1
Toronto	2	2
New York	2	3
New York	1	4

Using **Coding for Categorical Factors (-1, 0, +1)**, the model/design matrix (stored in the Residuals sheet) is:

Constant	X(Location)_New York	X(Location)_Toronto	Location_New York
1	0	1	-1
1	0	2	-1
1	2	0	1
1	1	0	1

The Parameter Estimates are:

Predictor Term	Coefficient
Constant	2.5
X(Location)_New York	-1
X(Location)_Toronto	1
Location_New York	2.5

Toronto is the hidden reference for Location. Note the different coefficient values (slopes) for X in New York versus X in Toronto.

The prediction equation (Y) as an Excel formula is:

= (2.5)
+ (-1)*(X)*IF(Location="New York",1,0)
+ (1)*(X)*IF(Location="Toronto",1,0)
+(2.5)*(IF(Location="New York",1,0)+IF(Location="Toronto",-1,0))

Random Factors

A random factor has levels that are randomly sampled from a larger number of possible levels but interest is in all possible levels.

In fitting a random effects or mixed (random & fixed effects) model, the contributions that different factors make to the overall variability of the data, as specified by their variance, are called the variance components (VC). These components can be estimated either from the ANOVA model with Expected Mean Squares (EMS), or by using Restricted Maximum Likelihood (REML). In SigmaXL, EMS is used if the data/model are:

- Balanced
- Hierarchical
- Nested terms are all categorical factors (no nested covariates)
- No covariate*random interaction terms

For the ANOVA model with EMS, the VCs are evaluated using the method of moments [2,4]. Each adjusted mean square has the expected value of $n\sigma_r^2 + \sigma_e^2$, where σ_r^2 is the VC for the rth term, σ_e^2 is the VC for the error term, and n is the number of samples taken for the rth term. Equating the actual and expected adjusted mean squares solves for σ_r^2, the VC for the rth term. VCs from the EMS model are unbiased for balanced data. Note that negative VCs are possible.

With REML, the VCs are estimated from the parameters derived from the maximization of the restricted maximum likelihood [2,4]. The mixed effect model consists of:

$$y = X\beta + \sum_i Z_i\mu_i + \epsilon$$

where X is the n x p design matrix of the fixed effect terms, Z_i is the design matrix for the ith random term, μ_i is an independent variable from $N(0, \sigma_i^2)$ and ϵ is an independent variable from $N(0, \sigma^2)$. The likelihood function is defined as:

$$LLF = .5[(n-p)\ln(2\pi\sigma^2) + \ln|H| + \ln|X'H^{-1}X| + (y - X\beta)'H^{-1}(y - X\beta)]$$

where $H = I_n + \sum (\theta_i Z_i Z'_i)$ and $VC_i = \sigma_i^2 = \theta_i\sigma^2$. The Expectation–Maximization (EM) algorithm [5] is used, whereby an iterative process is followed consisting of two steps - maximizing the likelihood conditional on β to obtain $\hat{\theta}_i$ and $\hat{\sigma}^2$, and then estimating $\hat{\beta} = (X'H^{-1}X)^{-1}X'H^{-1}y$. Starting values are derived by using the method of moments as in the ANOVA EMS model. This is an extremely computationally expensive process, so currently the maximimum sample size permitted for REML is 10,000. With N=10,000 the REML calculations will take approximately 1-2 hours. A future implemention will utilize sparse matrices to speed up the calculations and permit larger sample sizes.

Note that the GLM Regression Report treats random factors as fixed. ANOVA for Predictors, Pareto Charts and CI/PI for Predicted Response Calculator are not available in the regression report. Random effects best linear unbiased predictions (BLUPs) are not available.

The VC standard errors (s) are derived from the inverse of the square root of the diagonal elements of the information matrix in REML. The 95% confidence bounds are evaluated as $\exp(\ln(VC) \pm 1.96/z)$, where $z = VC/s$. P-Values for the variance components are also given but note that they are often underpowered [6].

If the VCs are computed from ANOVA EMS, then REML is also computed in order to provide confidence intervals, but only if the sample size is not greater than 1200. With N=1200, the REML calculations will take approximately one minute.

GLM ANOVA Table (EMS)

When there are random components which occur in more than one term of the model, the F-statistic must take this into account. If there are k random terms in the model, then there are $k1 = k + 1$ random terms in total, including the residual. Let A be the $k1$x1 vector of adjusted mean squares, E the $k1$x$k1$ matrix of expected mean squares augmented with rows of zeros and a column of ones to account for the residual random component, and S the $k1$x$k1$ matrix S, with the principal diagonal set to zero. Then the F statistic for the ith term is given by:

$$F_i = A_i/Q_i$$

where $Q = SE^{-1}A$. The degrees of freedom used in estimating the F-probability are derived using the Satterthwaite approximation [3].

Tests of Fixed Effects Table (REML)

The statistical significance of each fixed effects terms in the mixed effect model is tested using a Wald test. The test statistic W is defined as:

$$W = \hat{\beta}'L'(L\phi L')^{-1}L\hat{\beta}/k_1$$

where k_1 is the number of parameters in the tested term, and L is a zero square matrix of size β, with k_1 diagonal elements of unity for the appropriate elements of the tested term. Under the null hypothesis, W is distributed as F with k_1 and k_2 degrees of freedom, where k_2 is derived using the Satterthwaite approximation [1,3].

Pairwise Comparison of Means for Fixed Factors

Pairwise comparisons of means examines the difference between all combinations of the estimated means for each category of a factor, along with the standard error and confidence band for the difference. In estimating the difference between two categories of a factor, the values of the remaining factors are set to their average value; thus if a factor has 4 categories, the value for each dummy variable is set at 0.25. The mean value of the dependent variable for each category i of factor k, $\bar{y}_{i,k}$, and the standard error of this prediction, $\sigma_{i,k}$, are derived from the predicted value.

Thus the standard error of the difference is:

$$SE_d = SE(\bar{y}_{i,k} - \bar{y}_{j,k}) = \sqrt{\sigma_{i,k}^2 + \sigma_{j,k}^2}$$

The confidence band is:

$$\bar{y}_{i,k} - \bar{y}_{j,k} \pm qSE_d$$

where q is the inverse of the Student's t cdf for Fisher(individual confidence intervals), and $\sqrt{2}\ q$ is the inverse of the Studentized range cdf for Tukey (simultaneous confidence intervals).

This set of tests are referred to as "post hoc tests". These tests permits the researcher to assess the statistical significance of differences between means. Two tests are implemented:

Fisher's least significant difference test is a set of individual t-tests on each of the sample pairs, and uses the pooled standard deviation from both estimated means, which increases the power. The test statistic is:

$$t = \frac{\bar{y}_{i,k} - \bar{y}_{j,k}}{\sqrt{\sigma_{i,k}^2 + \sigma_{j,k}^2}}$$

and is assumed distributed Student t.

Tukey's test is a set of individual t-tests on each of the sample pairs, but uses the pooled standard deviation from both estimated means. The test statistic is:

$$t = \frac{\sqrt{2}(\bar{y}_{i,k} - \bar{y}_{j,k})}{\sqrt{\sigma_{i,k}^2 + \sigma_{j,k}^2}}$$

and is distributed as the Studentized range. This test provides protection against false positives due to multiple comparisons.

The maximum number of combinations permitted is 10,000 so there may be up to 141 category levels for each factor. Nested factors are not shown in the pairwise comparison report.

References for General Linear Model

[1] Fai, A. H. and Cornelius, P. L. (1996). " Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-Plot Experiments" *J. Statist. Comput. Simul.* Vol. 54, pp. 363-378.

[2] Montgomery, D.C. (2020). *Design and Analysis of Experiments*, 10th Edition, John Wiley & Sons.
[3] Satterthwaite F.E.(1946). "An approximate distribution of estimates of variance components". *Biom Bull*. Vol. 2, pp 110-114.
[4] Searle, S.R., G. Casella, and C.E. McCulloch, (1992), *Variance Components*, New York: Wiley.
[5] "Expectation–maximization algorithm" in Wikipedia: https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm.
[6] Dickey, D.A. (2020), "A Warning about Wald Tests". *SAS Global Forum 2020*, Paper 5088 - 2020: https://support.sas.com/resources/papers/proceedings20/5088-2020.pdf.

Time Series Forecasting and Control Charts for Autocorrelated Data

Autocorrelation (ACF), Partial Autocorrelation (PACF) and Cross Correlation (CCF)

Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between lagged values of data [9].

The autocorrelation coefficient (r_k) at lag k for a data set y is given by:

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{T} (y_t - \overline{y})^2}$$

where T is length of the time series (i.e., number of observations). This is known as the Autocorrelation Function (ACF) [9].

The maximum lag (k_{max}) may be user provided; the default value is the maximum of $10\log(T)$ and $3 * seasonal\ frequency$, but not exceeding $T/3$.

Partial autocorrelation is similar to autocorrelation but adjusts for correlation inherent in lags, e.g., y_t and y_{t-2} might be correlated simply because they are both connected to y_{t-1}, rather than because of any new information contained in y_{t-2} [9].

Partial autocorrelation is the autocorrelation between y_t and y_{t+k} after removing any linear dependence on $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$; thus it measures the correlation between y_t and y_{t+k} not accounted for by an Autoregressive AR(k-1) process. Each partial autocorrelation can be estimated as the last coefficient in an autoregressive model. Specifically, α_k, the kth partial autocorrelation coefficient, is equal to the estimate of ϕ_k in an AR(k) model [9].

The partial autocorrelation coefficient (ϕ_k) at lag $k, k > 1$ is given by:

$$\hat{y}_{t+k} = \beta_1 y_{t+k-1} + \beta_2 y_{t+k-2} + \cdots + \beta_{k-1} y_{t+1}$$
$$\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \cdots + \beta_{k-1} y_{t+k-1}$$
$$\phi_k = \Phi(y_{t+k} - \hat{y}_{t+k}, y_t - \hat{y}_t)$$

where Φ is the correlation function.

If there are missing values, the autocorrelation and partial autocorrelation coefficients are evaluated by skipping correlations that include missing values and adjusting the sample count.

Under the null, the sample autocorrelation and partial autocorrelation coefficients are distributed approximately normal, with a mean of zero, and a variance of $1/n$; thus, confidence intervals can be used to test the null hypothesis of no correlation. The bargraph of the sequence of autocorrelations with confidence intervals is called the Autocorrelation Function (ACF) Plot (sometimes also called a Correlogram). The bargraph of the sequence of partial autocorrelations is called the Partial Autocorrelation Function (PACF) Plot.

Cross Correlation is the standard method of estimating the degree to which two series are correlated. The bargraph of the sequence of cross correlations is called the Cross Correlation Function (CCF) Plot. The cross correlation coefficient ($r_{x,y}$) for data sets x and y is given by:

$$r_{x,y} = \frac{1}{n\sigma_x\sigma_y}\sum_{t=1}^{n} (x_t - \bar{x})(y_t - \bar{y})$$

were n is the number of observations, $\bar{x}$ and $\bar{y}$ are the means, and σ_x and σ_y are the standard errors of x and y respectively. The cross-correlation coefficient can also be estimated between lagged values of x and y; for example, between y_{t+k} and x_t. The cross-correlation coefficient ranges in value between +1 (perfect positive correlation), and -1 (perfect negative correlation).

If x is autocorrelated, the CCF is affected by its time series structure and any "in common" trends the x and y series may have over time. Pre-whitening solves this problem by removing the autocorrelation and trends. The Pre-Whiten Data option does the following:

1. Automatic ARIMA is applied to the x input; residuals are stored.
2. The y data is filtered using the same model and parameter values; residuals are stored.
3. A CCF is then created using the x and y residuals.

For further details, see "Pre-whitening as an Aid to Interpreting the CCF" https://newonlinecourses.science.psu.edu/stat510/lesson/9/9.1.

Ljung-Box Test

In addition to looking at the ACF plot, we can also do a more formal test for autocorrelation by considering a whole set of r_k values as a group, rather than treating each one separately. This is the Ljung-Box Q Test:

$$Q = T(T+2)\sum_{k=1}^{h} (T-k)^{-1} r_k^2,$$

where h is the maximum lag being considered and T is the number of observations.

If the autocorrelations did come from a white noise series, then Q would have a χ^2 distribution with $(h - k)$ degrees of freedom, where k is the number of parameters in the model [9]. This is also known as a Portmanteau test.

Box-Cox Transformation

The Box-Cox procedure transforms the vector x as follows:

$$y = \frac{(x - \tau)^{\lambda} - 1}{\lambda}$$

where $\tau < x_{min}$ is a threshold parameter, such that $x - \tau$ is always positive. The object of this procedure is to transform non-normal data into data that is closer to normal using Sequential Quadratic Programming (SQP) to maximize a likelihood function. The likelihood includes trend and seasonality dummy variables.

If λ is zero, then $y = \ln(x)$. λ is constrained to be >= 0 and <= 1. If Rounded Lambda is selected the choices for λ are $\ln(x)$ or $\sqrt{x}$. This conservative approach was found to give the best results with competition data. The user may also specify any value for λ and threshold.

If an invalid threshold is specified, a threshold value is specified as $x_{min} - \delta$, where $\delta = \min(0.001, 0.001(x_{max} - x_{min}))$.

If a threshold parameter is not specified, and x has negative values, then τ is estimated along with λ; these parameters are selected to maximize the constrained likelihood using SQP. The constraint is used to ensure that $x - \tau$ is always positive. Trend and seasonal dummy variables are not included in the likelihood for this case.

In both ARIMA and Exponential Smoothing, the user has the option of invoking a Box-Cox transform. Forecast values are specified using the inverse transform.

Seasonal Frequency

Seasonal Frequency is the number of observations per “cycle” unit of time. Some examples include:

- Quarterly data, seasonal frequency = 4 observations per year
- Daily data, seasonal frequency = 7 observations per week
- Monthly data, seasonal frequency = 12 observations per year
- Hourly data, seasonal frequency = 24 observations per day
- Half-hourly data, seasonal frequency = 48 observations per day
- Weekly data, seasonal frequency = 52 observations per year
- Hourly data, seasonal frequency = 168 observations per week
- Half-hourly data, seasonal frequency = 336 observations per week
- Daily data, seasonal frequency = 365 observations per year (excluding leap years).

Seasonal decomposition should be used for specified frequencies higher than 24.

SigmaXL follows the convention used in R time series objects (ts) and the R forecast package [12], but note that it is the opposite of the definition of frequency in physics, or in engineering Fourier analysis, where “period” is the length of the cycle, and “frequency” is the inverse of period (see https://robjhyndman.com/hyndsight/seasonal-periods/).

Seasonal Trend Decomposition

A time series can be decomposed into three components:

$$y_t = T_t + S_t + I_t$$

where T_t is the trend and cyclical component, S_t is the seasonal component, and I_t is the remainder, the irregular or noise component. STL is a versatile and robust method for decomposing time series into these three components. STL is an acronym for "Seasonal and Trend decomposition using Loess". Loess is locally weighted regression smoothing. The STL method was developed by Cleveland, et al. [3]. Note that it is strictly additive, but one can simply specify a Box-Cox Transformation with Lambda = 0 (Ln) to obtain an equivalent to multiplicative decomposition. Unlike the use of Box-Cox in forecasting, the results remain transformed, i.e., the inverse transformation is not applied to the decomposed data – the additive model is retained.

If seasonal frequency is not specified, the trend is computed using Freidman's "Supersmoother" [4], which uses local linear regression with adaptive bandwidths. The basic idea is to first estimate a number of fixed bandwidth smooths by local linear regression. The leave-one-out cross-validated residuals from each of those initial estimates are then smoothed using a constant bandwidth. A set of smoothed bandwidths are selected and the smoothed outcomes are linearly interpolated and smoothed with a fixed bandwidth.

Multiple Seasonal Decomposition (MSD) uses STL iteratively and was developed by Hyndman, et al. [12].

Spectral Density and Automatic Detection of Seasonal Frequency

Spectral Density determines the seasonal frequency of a time series using spectral analysis with Fast Fourier Transforms, using a specified uniform window (see https://en.wikipedia.org/wiki/Window_function).

The Fisher Kappa white noise test is automatically applied (see Fuller [5], Chapter 7) and, if significant (p-value < .05), returns the rounded frequency with the largest spectral power value. If not significant, a frequency of unity is returned (nonseasonal).

A further decomposition based seasonal strength test is applied to the rounded frequency (https://otexts.com/fpp2/seasonal-strength.html) and if the statistic is >= .5, the above frequency value is used, otherwise it is set to 1. This helps to distinguish between cyclical and seasonal time series.

For the Spectral Density Plot or MSD option, if the seasonal frequency is > 1, the above process is repeated two more times with the dominant frequency removed by seasonal decomposition, spectral analysis applied to the remainder and seasonal strength test applied to the new dominant frequency.

Seasonally Adjusted Linear Interpolation of Missing Values

Linear interpolation is carried out by first estimating missing values using a regression model with trend and dummy coding for seasonal. The time series is then decomposed into three components - trend, seasonal and remainder - using STL decomposition for seasonal data. A data point is estimated by linear interpolation of the nonseasonal components of the previous and subsequent values, and then adding the seasonal component for the observation in question [12, 17]. For nonseasonal data, linear interpolation is utilized.

Once a model has been specified, missing values are estimated using one step ahead forecasting.

Information Criteria for Model Comparison

Parameters are estimated by maximizing the Log-Likelihood function (which is similar to minimizing the residual sum-of-squares). Given a set of candidate models for the data, the preferred model is the one with the minimum Information Criteria value. The Information Criteria rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting, because increasing the number of parameters in the model almost always improves the goodness of the fit (see https://en.wikipedia.org/wiki/Akaike_information_criterion).

Akaike's Information Criterion (AIC) is:

$$\text{AIC} = -2\ln(L) + 2(k+1)$$

where L is the likelihood of the model and k is the total number of parameters (and initial states for exponential smoothing) that have been estimated.

The AIC corrected for small sample bias (AICc) is defined as:

$$\text{AIC}_c = \text{AIC} + \frac{2(k+1)(k+2)}{T-k-2}$$

The Bayesian Information Criterion (BIC) is:

$$\text{BIC} = \text{AIC} + (k+1)[\ln(T) - 2]$$

AICc is recommended as the default Information Criterion, based on forecast error performance with M3 competition data [12].

Notes on AIC from Rob Hyndman (see https://robjhyndman.com/hyndsight/aic/):

- You can compare very different models with AIC, however, make sure the likelihoods are computed on the same data. For example, you cannot compare an ARIMA model with differencing to an ARIMA model without differencing, because you lose one or more observations via differencing.
- For a similar reason, you cannot compare the AIC from an Exponential Smoothing model with the AIC from an ARIMA model. The likelihood function calculations are different and the two models treat initial values differently. For example, after differencing, an ARIMA model is computed on fewer observations, whereas an Exponential Smoothing model is always computed on the full set of data. Even when the models are equivalent (e.g., an ARIMA(0,1,1) and an ETS(A,N,N)), the AIC values will be different.

Forecast Accuracy

As noted above, Information Criteria AIC/AICc cannot be used to compare ARIMA and ETS models or ARIMA models with different d, D values, but forecast accuracy metrics can be compared. A forecast error e_t is the difference between an observed value and its forecast. It is not the same as a model residual because residuals may be in different units due to Box-Cox transformation or a Multiplicative model.

Types of forecast error include:

- In-Sample One-Step-Ahead Forecast. This is less useful because the model may be over-fitted.
- Out-of-Sample (Withhold) One-Step-Ahead. Model parameter estimates do not use any withhold data, but the forecast updates with every new withhold observation.
- Out-of-Sample (Withhold) Multi-Step-Ahead. This is important if one is assessing forecast accuracy over a horizon. This is used in forecast competitions.

Forecast accuracy measures used in SigmaXL include:

Root mean squared error: RMSE $= \sqrt{\text{mean}(e_t^2)}$

Mean absolute error: MAE $= \text{mean}(|e_t|)$

Mean absolute percentage error: MAPE $= \text{mean}\left(\left|\frac{100e_t}{y_t}\right|\right)$

Mean absolute scaled error: MASE $= \text{mean}(|e_t|)/scale$

where *scale* is the MAE of the in-sample naïve or seasonal naïve forecast (set all forecasts to be the value of the last observation/period). A scaled error is less than one if it arises from a better forecast than the average naïve/seasonal naïve forecast. Conversely, it is greater than one if the forecast is worse than the average naïve forecast [9].

Exponential Smoothing - ETS

Simple Exponential Smoothing:

Simple Exponential Smoothing forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past with the smallest weights associated with the oldest observations:

$$\hat{y}_{t+1} = \alpha\ y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \cdots$$

where $0 \le \alpha \le 1$ is the smoothing parameter.

An equivalent formulation for simple exponential smoothing is:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

with the starting forecast value (initial level) $\hat{y}_1$ to be estimated and denoted as level l_0.

Hyndman gives an alternative representation known as the component form [9]. For simple exponential smoothing, the only component included is the level ℓ_t. Other methods may also include a trend b_t and/or a seasonal component s_t.

Component form representations of exponential smoothing methods comprise a forecast equation and a smoothing equation for each of the components included in the method. The component form of simple exponential smoothing is given by:

$$\begin{array}{rll} \text{Forecast equation} & \hat{y}_{t+h|t} & = \ell_t \\ \text{Smoothing equation} & \ell_t & = \alpha y_t + (1-\alpha)\ell_{t-1}, \end{array}$$

where ℓ_t is the level (or the smoothed value) of the series at time t. $\hat{y}_{t+h|t}$ denotes an h multi-step forecast, taking account of all observations up to time t. Setting $h = 1$, a one step ahead forecast gives the in-sample fitted values.

The “error correction” form is:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

where $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ is the residual at time t. The equations of the additive error model can then be written as:

$$\begin{aligned} y_t &= \ell_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \alpha\varepsilon_t. \end{aligned}$$

Multiplicative errors are one-step-ahead relative errors:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

The equations of the multiplicative error model can be written as:

$$\begin{aligned} y_t &= \ell_{t-1}(1 + \varepsilon_t) \\ \ell_t &= \ell_{t-1}(1 + \alpha\varepsilon_t). \end{aligned}$$

Error, Trend, Seasonal (ETS) Models:

Error, Trend, Seasonal (ETS) models expand on simple exponential smoothing to accommodate trend and seasonal components as well as additive or multiplicative errors. Simple Exponential Smoothing is an Error Model. Error, Trend model is Holt's Linear, also known as double exponential smoothing. Error, Trend, Seasonal model is Holt-Winters, also known as triple exponential smoothing.

Rob Hyndman has developed a complete taxonomy that describes all of the combinations of exponential smooth models in a consistent manner [9]:

- Error:
  - Additive or Multiplicative
  - The point forecasts produced by the models are identical if they use the same smoothing parameter values. Multiplicative will, however, generate different prediction intervals to accommodate change in variance.
  - An alternative to multiplicative is to use the Ln transformation (Box-Cox transformation with Lambda = 0).
  - Error models include the smoothing parameter α and initial level value.

- Trend:
  - None, Additive, Additive Damped
  - Multiplicative Trend is not recommended as it tends to produce poor forecasts [9], so is not included in SigmaXL
  - Trend models add a smoothing parameter β and initial trend value.
  - Damped trend models add a smoothing parameter ϕ that "dampens" the trend to a flat line some time in the future.

- Seasonal:
  - None, Additive, Multiplicative
    - The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.
  - Seasonal models add a smoothing parameter γ and initial seasonal values.
    - # of initial values estimated = seasonal frequency – 1
    - constrained to sum to 0 for additive or seasonal frequency *m* for multiplicative

Summary of ETS Models in SigmaXL:

Short hand (Error, Trend, Seasonal)	Method
(A, N, N)	Simple Exponential Smoothing with Additive Errors - Exponentially Weighted Moving Average (EWMA)
(M, N, N)	Simple Exponential Smoothing with Multiplicative Errors
(A, A, N)	Additive Trend Method with Additive Errors (Holt's Linear)
(M, A, N)	Additive Trend Method with Multiplicative Errors (Holt's Linear)
(A, A, A)	Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters)
(M, A, A)	Additive Trend, Additive Seasonal Method with Multiplicative Errors (Holt-Winters)
(A, N, A)	Additive Seasonal Method with Additive Errors
(M, N, A)	Additive Seasonal Method with Multiplicative Errors
(A, Ad, N)	Additive Damped Trend Method with Additive Errors
(M, Ad, N)	Additive Damped Trend Method with Multiplicative Errors
(A, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Additive Errors
(M, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Multiplicative Errors
(M, A, M)	Additive Trend, Multiplicative Seasonal Method with Multiplicative Errors (Holt-Winters)
(M, N, M)	Multiplicative Seasonal Method with Multiplicative Errors
(M, Ad, M)	Additive Damped Trend, Multiplicative Seasonal Method with Multiplicative Errors (Holt-Winters)

ETS Model Equations:

The equations for each of the models (also known as the "innovations state space model") are given as [9]:

ADDITIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Automatic Model Selection (Identification):
The identification process is used to ascertain which ETS model to use. The best model is determined by estimation of all possible combinations of error, trend, seasonality and damping. AICc is recommended as the default Information Criteria, based on forecast error performance with M3 competition data.

Some of the model combinations lead to numerical instability and are not considered in the selection process: (A,N,M) (A,A,M) (A,Ad,M).

If a Box-Cox transformation is used, multiplicative models are not considered.

Estimation:
Given a specific ETS model, the values of the weights (α for error, β for trend, γ for season, and ϕ for damping) are estimated using exact constrained maximum likelihood in state space using recursion, employing the Nelder Mead simplex algorithm. Effectively, the objective is to minimize the residual sum of squares:

$$SSE = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$

The optimization returns the weights (α for error, β for trend, γ for seasonality, and ϕ for damping), as well as initial state values. There are a number of constraints on the weights (in addition to the weights each lying between 0 and 1) that are required for stability [8, 10]. If missing values are encountered, seasonally adjusted linear interpolation is used. A Box-Cox transform is permitted prior to estimation, although such a transformation should only be used for an additive ETS model.

Covariates cannot be estimated as part of an integrated ETS estimation (see https://robjhyndman.com/hyndsight/ets-regressors/ for a discussion on this issue).

A starting value of 0.5 is used for α, 0.05 for β, 0.025 for γ, and 0.978 for ϕ. For covariates, the initial values are taken from the OLS estimates of a regression of y against a constant and the scaled covariates.

Forecasting:
Once a model is specified, future values of the series can be evaluated easily. Given data $y_1, \cdots, y_n$. then the conditional mean of y_{n+1} is derived using the point forecast. The process is repeated dynamically.

See https://otexts.com/fpp2/ets-forecasting.html for a brief discussion of ETS model prediction intervals. Full formula details are quite complex and given in Hyndman et al. [11].

Autoregressive Integrated Moving Average - ARIMA

The Autoregressive Integrated Moving Average (ARIMA) model was developed by Box and Jenkins [2]. The automatic determination of the best model order in SigmaXL uses the method of Rob Hyndman [9,10,12].

Stationarity and Differencing:

ARIMA assumes that the time series is stationary, i.e., it has the property that the mean, variance and autocorrelation structure do not change over time. If a time series mean is not stationary (e.g. trending), this can be corrected by differencing, computing the differences between consecutive observations for nonseasonal and between consecutive periods (e.g. months) for seasonal data (Jan 2019 – Jan 2018, etc.). For nonseasonal, this will typically involve 1 or 2 orders of differencing. This order is the Integrated term *d.* For seasonal, this will typically involve 1 order of differencing. This order is the Seasonal Integrated term *D*. A constant term *c* is optional:

- If *d*+*D* = 0, a constant term in the model is the mean.
- If *d*+*D* = 1, a constant term in the model is a trend (drift).
- If *d*+*D* > 1, a constant term would be a quadratic trend, so constant should not be included.
- It is recommended that *d*+*D* should not be > 3.
- If the variance is not stationary, use a Box-Cox transformation.

Autoregressive (AR) Model:

In an autoregressive model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregressive indicates that it is a regression of the variable against itself:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where ε_t is white noise. This is like a multiple regression but with lagged values of y_t as predictors. We refer to this as an AR(p) model, an autoregressive model of order p [9].

Moving Average (MA) Model:

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model [9].

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

We refer to this as an MA(q) model, a moving average model of order q [9].

Autoregressive Integrated Moving Average (ARIMA) Model:

If we combine differencing with autoregression and a moving average model, we obtain a nonseasonal ARIMA model.

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where y'_t is the differenced series. This is the ARIMA (p, d, q) model, where p is the number of autoregressive terms, d is the degree of differencing and q is the number of moving average terms.

To form more complicated models, it is much easier to work with the backshift notation [9].

$$\underset{\text{AR}(p)}{\underset{\uparrow}{(1 - \phi_1 B - \cdots - \phi_p B^p)}} \quad \underset{d\text{ differences}}{\underset{\uparrow}{(1-B)^d y_t}} \quad = \quad c + \underset{\text{MA}(q)}{\underset{\uparrow}{(1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t}}$$

Seasonal ARIMA:

For seasonal, the model consists of terms that are similar to the nonseasonal components of the model, but involve backshifts of the seasonal period. The seasonal model is ARIMA (P, D, Q) and combined we have ARIMA (p, d, q) (P, D, Q).

For example, an ARIMA(1,1,1)(1,1,1) model for quarterly data (seasonal frequency = 4), can be written as [9]:

$$(1 - \phi_1 B)\ (1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)\ (1 + \Theta_1 B^4)\varepsilon_t.$$

Automatic Model Selection (Identification):

The identification process is required to determine the degree of differencing (d, D) necessary to generate a series that is stationary, and to then ascertain the order of the model (p, q, P, Q). This can be done manually using the correlogram (ACF) and partial autocorrelogram (PACF) both before and after differencing. This process is automated in SigmaXL.

First, the degree of seasonal differencing (D) is determined with the Seasonal Strength (SEAS) test [12, 19] which uses the ratio of the seasonal component to the detrended component and compares this with the SEAS critical value. If the ratio exceeds the critical value, then this implies the rejection of the no unit root null hypothesis, implying $D = 1$, and the data are seasonally differenced.

The degree of differencing for the nonseasonal component (d) is then determined using the KPSS unit root test [12, 13]. The data is tested for a unit root, and if found the data is differenced, and the process continues; it stops when no significant unit root is found or $d = 2$. SigmaXL uses the Schwert truncation lag parameter $= \text{floor}(12\text{*}(\text{n}/100)\text{^}.25)$ for $n \leq 70$ and Quadratic Spectral Kernal for $n > 70$ [21, 22]. If $d = 0$ and $n \leq 500$, a further test is applied comparing the standard deviation of the (0,0,0) with constant model against the standard deviation of (0,1,0) with constant.

If $d = 1$ produces a lower standard deviation value, then it is used as the order for nonseasonal differencing (see https://people.duke.edu/~rnau/411arim2.htm). While there is a potential risk of overdifferencing, this additional test results in dramatic improvement in forecast accuracy using M3 and M4 competition data. If the process is thought to be stationary and SigmaXL returns $d = 1$, the user should manually specify a $d = 0$ model and compare the two models.

Other tests for seasonal and nonseasonal stationarity (unit-roots) are available, but SEAS and KPSS are recommended by Hyndman based on forecast accuracy evaluation with M3 competition data (see https://robjhyndman.com/hyndsight/forecast83/ and https://robjhyndman.com/hyndsight/unit-root-tests/). Currently, SigmaXL does not include alternative unit root/stationarity tests. The test statistic values for SEAS and KPSS are used internally but not reported.

Having ascertained the degree of seasonal and nonseasonal differencing, model selection implies the choice of the AR and MA components - p, q, P, Q - as well as whether a constant should or should not be part of the model. One method of choosing the best model is to exhaustively estimate all possible combinations, subject to upper bounds on the AR/MA terms. A more efficient approach is to estimate a series of ARIMA models using a stepwise ordering suggested by Hyndman and Khandakar [10]. The preferred model is the one with the lowest fit metric, the default being AICc. Only models that are stationary and invertible are considered.

Estimation:

Given a specific (p, d, q, P, D, Q) ARIMA model, the values of the AR and MA terms are estimated. Covariates (which are scaled automatically) are permitted, as well as a constant if there is no differencing, and a trend if d and D equals { 0 1 } or { 1 0 }. Two general models are available - the conditional sum of squares (CSS) and the state space Kalman maximum likelihood. The likelihood for the CSS model is:

$$logL = \sum_i [\log(2\pi)/2 - 0.5\log(\sigma^2) - e_i^2/2\sigma^2]$$

The CSS is always used for initial estimates and is used if $n > 500$ or frequency > 12. The residuals used are derived from the differenced time series – for CSS, AR/MA filtering is used to generate the residuals, while for Kalman, which is a recursive estimator, the residuals are derived from a maximum likelihood AR/MA process. In each case, the maximum likelihood process uses the constrained Nelder Mead simplex algorithm. The constraints impose a maximum value of unity on AR/MA terms by using a tanh transformation, as well as using a unit root test to ensure that the AR terms are stationary, and the MA terms are invertible.

For CSS, if missing values are encountered, the largest contiguous range is used. For Kalman, missing values are evaluated by skipping the update phase in the state based model.

Confidence bounds on the coefficients are derived from the diagonal of the inverse of either the Hessian or the outer product of the gradient (OPG).

Covariates (predictors) are treated in the traditional ARIMAX model by adding them to the right-hand side before differencing. In order to estimate an ARIMA model with covariates, the covariate terms are subtracted from the data before evaluating the residuals in each iteration of the evaluation of the likelihood. The parameters of the covariate terms are estimated concurrently with all the other parameters. For prediction and forecasting, the reverse process occurs, and the covariate terms are added back to the predicted/forecast values. Thus, the effect of the exogenous variables is always additive, irrespective as to the ARIMA model.

A starting value of 0.1 is used for each parameter, with the following exceptions: For the constant term, the mean is used if there is no differencing, and the mean of the differenced data, adjusted by the number of AR terms, otherwise. For covariates, the initial values are taken from the OLS estimates of an ARX model with m AR terms, scaled covariates and trend, where m is the seasonal frequency.

ARIMA Parameter Estimates include significance tests; P-Values < .05 are significant and highlighted in red. This is useful for model refinement with multiple predictors. Note that for AR/MA model order selection, minimum AICc should be used rather than significance tests [25].

In-Sample Fitted (One-Step Prediction)

The predicted values for an arima process are calculated using the Kalman process on the de-trended, differenced data, y_d. For the initial data point, the Kalman process uses a predicted value y_p of zero; subsequent predicted values are derived from the Kalman process. The calculated residuals are derived as the difference between y_d and y_p, and the predicted data, which are one step ahead predictions, are derived from actual data plus the calculated residual.

The predicted for an arima process are calculated using either the Kalman process or the CSS process on the de-trended, differenced data, y_d. For the initial data points,the Kalman process uses a predicted value y_p of zero; subsequent predicted values are derived from the Kalman process; these initial values are excluded under CSS. The calculated residuals are derived as the difference between y_d and y_p, and the predicted data, after removing trend. These are one-step-ahead predictions, and are derived from actual data plus the calculated residual. If differencing occurs, the first $d + D$ predicted values are set to missing.

Forecasting:

The raw coefficients are transformed into the ϕ^* and θ^* vectors, which can be used on the original time series. A forecast is undertaken from the end of the series for the number of forecast periods specified. Under the default, the forecasts are evaluated dynamically, such that the predicted values at any point provide the basis for the next period’s prediction. A confidence band is produced for each forecast value, based on the standard deviation of the forecast error for observation $n + m$:

$$\sigma_{n+m} = \sqrt{\sigma_w^2 \sum_{j=0}^{m-1} \Psi_j^2}$$

where σ_w^2 is the residual variance derived from the n observations, and Ψ_j is the weight on the $t - j$ observation when the residual for the t observation is written as an infinite order MA model.

Assessing Forecast Accuracy with Forecast Competitions

The "M" competitions organized by Spyros Makridakis have had an enormous influence on the field of forecasting. They focused attention on what models produced good forecasts, rather than on the mathematical properties of those models.

The M3 competition involved 3003 time series with various seasonal frequencies and areas of application. The M3 data have continued to be used since 2000 for testing new time series forecasting methods. The recent M4 competition involved 100,000 time series! (See https://robjhyndman.com/hyndsight/m4comp/ and https://www.mcompetitions.unic.ac.cy/).

Using a hybrid average of automatic Exponential Smoothing and ARIMA, SigmaXL (unofficially) ranked 10^{th} out of 60 in the M4 Overall Weighted Average score, ahead of three well known commercial forecast software packages.

Control Charts for Autocorrelated Data

Statistical process control for autocorrelated processes typically use the EWMA (Exponentially Weighted Moving Average) one-step-ahead forecast model [1, 2, 7, 14, 15]. The formula used for EWMA is the same as simple exponential smoothing, but the smoothing parameter λ is typically used instead of α and X_t instead of y_t:

$$\mathrm{EWMA}_{t+1} = \lambda X_t + (1 - \lambda)\mathrm{EWMA}_t$$

with the starting forecast value EWMA_1 estimated as the data mean or target value. In the case of an EWMA control chart with independent data, the smoothing parameter λ is determined by desired average run length characteristics and is typically 0.2. For SPC for autocorrelated data, the smoothing parameter and initial level are determined by minimizing the sum-of-square forecast errors (residuals):

$$\mathrm{SSE} = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} e_t^2.$$

using non-linear minimization methods like Newton-Raphson or Nelder-Mead Simplex.

The time series model forecasts the motion in the mean and an Individuals control chart is plotted of the residuals to detect assignable causes. Failure to account for the autocorrelation will produce limits that are too narrow resulting in excessive false alarms, or limits that are too wide resulting in misses.

Woodall and Faltin give some helpful guidelines on dealing with autocorrelation [20]:

- If possible, one should first attempt to remove the source of the autocorrelation.
- If the source of autocorrelation cannot be removed directly, then it may be possible to model the autocorrelation and use a feedback control scheme to reduce variability about a

specified target value.
- If the source of the autocorrelation cannot be removed directly, and feedback control is not a viable option, then it is important to monitor the process with control charts which do not repeatedly give signals due to presence of the autocorrelation.

SigmaXL provides an option to create an Individuals and Moving Limits (with One-Step-Ahead Forecast) control chart for the residuals of any of the forecast models, including:
- Exponential Smoothing
- Multiple Seasonal Decomposition with Exponential Smoothing
- ARIMA
- ARIMA with Predictors
- Multiple Seasonal Decomposition with ARIMA

The Moving Limits Residuals chart uses the one step prediction as the center line, so the control limits will move with the center line. If a Box-Cox transformation is used then an inverse transformation is applied to calculate the control limits. If the residuals are exponential smoothing multiplicative, the control limits are approximate and out-of-control signals may not exactly match the Individuals Chart. If that occurs, the Individuals Chart should be used to determine what points are out-of-control.

Tests for Special Causes are not recommended to be used with these control charts. While the residuals are independent, any sustained shift in the process mean is quickly adjusted for in the time series model, so the additional runs rules would not work as designed in order to detect small shifts.

Distinguishing an assignable cause as an outlier or as a sustained shift can be done by adding coded predictors in an ARIMA model and comparing the regression coefficient p-values. A potential outlier would be designated with a “1” and all other observations are “0”. To test for a shift, all values up to the observation of interest would be a “0”, and the remaining are designated as a “1” [23].

A Moving Range Residuals Chart is not included as that has too many false alarms [24, pp. 114-115].

If Tests for Specials Causes or a Moving Range Chart are desired, simply create an Individuals and Moving Range Chart on the Forecast Residuals. Note that the “Add Data” feature with automatic update of residuals is not supported in this case.

References for Time Series Forecasting and Control Charts for Autocorrelated Data

[1] Alwan, L.C., and Roberts, H.V. (1988), "Time Series Modeling for Statistical Process Control", *Journal of Business and Economic Statistics*, 6, 87-95.

[2] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016). *Time Series Analysis, Forecasting and Control*, 5th edition, Wiley.

[3] Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I. (1990) "STL: A Seasonal-Trend Decomposition Procedure Based on Loess", *Journal of Official Statistics*, Vol. 6(1), pp. 3-73.

[4] Friedman, J. H. (1984). "A variable span smoother. Laboratory for Computational Statistics", *Technical Report(5)*, Department of Statistics, Stanford University.

[5] Fuller, W.A. (1996). *Introduction to Statistical Time Series*, 2nd ed. John Wiley and Sons, New York.

[6] Gardner, G, Harvey, A.C. and Phillips, G.D.A. (1980). "Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive - moving average models by means of Kalman filtering". *Applied Statistics*, Vol. 29, pp. 311 - 322.

[7] Hunter, J.S. (1986), "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, 18, 203-210.

[8] Hyndman, R.J., M. Akram, and B.C. Archibald, (2006). "The admissible parameter space for exponential smoothing models." *Annals of the Institute of Statistical Mathematics* Vol. 60. pp. 407-426.

[9] Hyndman, R.J., & Athanasopoulos,G. (2018). *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

[10] Hyndman, R.J. and Y. Khandakar, (2008). "Automatic Time Series Forecasting: The forecast Package for R." *Journal of Statistical Software*, Vol. 27(3), pp. 1-22.

[11] Hyndman, R.J., A.B. Koehler, J. Keith Ord and R.D. Snyder (2008). *Forecasting with Exponential Smoothing. The State Space Approach* Springer-Verlag, Berlin Heidelberg.

[12] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2019). _forecast: Forecasting functions for time series and linear models_. R package version 8.9.

[13] Kwiatowski, D., Phillips, P.C.B, Schmid, P. and Shin, T. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic series have a unit root", *Journal of Econometrics*, Vol. 54, pp. 159-178.

[14] Montgomery, D. C., and Mastrengelo, C.M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data", *Journal of Quality Technology*, 23, 179-204.

[15] Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*, 7th edition, Wiley.

[16] Montgomery, D. C., Jennings, C.L., and Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*, 2nd edition, Wiley.

[17] Moritz, S., Sarda, A. et al. (2015). "Comparison of different Methods for Univariate Time Series Imputation in R", https://arxiv.org/ftp/arxiv/papers/1510/1510.03924.pdf.

[18] *NIST/SEMATECH e-Handbook of Statistical Methods*,

https://www.itl.nist.gov/div898/handbook.

[19] Wang, X., Smith, K. and Hyndman, R. (2006) "Characteristic-Based Clustering for Time Series Data", *Data Mining and Knowledge Discovery*, Vol. 13, pp. 335-364.

[20] Woodall, W.H. and Faltin, F.W. "Autocorrelated Data and SPC", *ASQC Statistics Division Newsletter*, 13(4).

[21] Hobijn, B., P. H. Franses, and M. Ooms. 1998. "Generalizations of the KPSS-test for stationarity", *Econometric Institute Report 9802/A*, Econometric Institute, Erasmus University Rotterdam.

[22] Schwert, "G.W. Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 1989, 147-160.

[23] The ARIMA Procedure – Detecting Outliers, SAS/ETS(R) 9.3 User's Guide, http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_arima_sect045.htm.

[24] Fuentes, Jesús Cuéllar (2008), "Statistical Monitoring of a Process with Autocorrelated Output and Observable Autocorrelated Measurement Error" (Doctoral Dissertation), Baylor University. https://baylor-ir.tdl.org/handle/2104/5185.

[25] Kostenko, A.V. and Hyndman, R.J. (2008), "Forecasting without significance tests?", https://robjhyndman.com/papers/sst2.pdf

Sigma XL

Simply Smart Software

About SigmaXL, Inc.

Established in 1998, SigmaXL Inc. is a leading provider of user-friendly Excel Add-ins for Lean Six Sigma graphical and statistical tools and Monte Carlo simulation.

Our flagship product, SigmaXL®, was designed from the ground up to be a cost-effective, powerful, but easy to use tool that enables users to measure, analyze, improve and control their service, transactional, and manufacturing processes. As an add-in to the already familiar Microsoft Excel, SigmaXL® is ideal for Lean Six Sigma training and application, or use in a college statistics course.

DiscoverSim™ enables you to quantify your risk through Monte Carlo simulation and minimize your risk with global optimization. Business decisions are often based on assumptions with a single point value estimate or an average, resulting in unexpected outcomes. DiscoverSim™ allows you to model the uncertainty in your inputs so that you know what to expect in your outputs.

Contact Information:

Technical Support: 1-866-475-2124 (Toll Free in North America) or

1-519-579-5877

Sales: 1-888-SigmaXL (888-744-6295)

E-mail: support@SigmaXL.com Web: www.SigmaXL.com