



Simply Smart Software

SigmaXL[®] Version 8.0 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

Table of Contents

SigmaXL® What's New in Version 8.0, System Requirements and Getting Help	1
What's New in Version 8.0.....	2
Activating SigmaXL.....	6
Error Messages	8
SigmaXL® Defaults and Menu Options	9
Clear Saved Defaults	9
Data Selection Default	9
Menu Options (Classical or DMAIC).....	10
SigmaXL® System Requirements	11
Getting Help and Product Registration.....	12
Introduction to SigmaXL® Data Format and Tools Summary	13
Introduction	15
The Y=f(X) Model	15
Data Types: Continuous Versus Discrete	16
Stacked Data Column Format versus Unstacked Multiple Column Format	17
Summary of Graphical Tools.....	18
Summary of Statistical Tools.....	19
SigmaXL: Measure Phase Tools	21
Part A - Basic Data Manipulation	23
Introduction to Basic Data Manipulation	23
Category Subset	23
Random Subset	24
Numerical Subset	25
Date Subset	25
Transpose Data	26
Stack Subgroups Across Rows.....	26
Stack Columns	28
Random Data	29
Box-Cox Transformation	29
Standardize Data	29

Data Preparation – Remove Blank Rows and Columns	30
Data Preparation – Change Text Data Format to Numeric.....	32
Recall SigmaXL Dialog	33
Activate Last Worksheet	33
Worksheet Manager	34
Part B – Templates & Calculators	35
Introduction to Templates & Calculators	35
SigmaXL Templates & Calculators.....	36
Team/Project Charter	38
SIPOC Diagram	39
Data Measurement Plan	39
Cause & Effect (Fishbone) Diagram	40
Cause & Effect (Fishbone) Template.....	41
Cause & Effect (XY) Matrix Example	42
Failure Mode & Effects Analysis (FMEA) Example.....	43
Quality Function Deployment (QFD) Example.....	46
Pugh Concept Selection Matrix Example.....	47
Control Plan.....	48
Lean – Takt Time Calculator Example	49
Lean – Value Analysis/Process Load Balance.....	50
Lean – Value Stream Mapping.....	51
Basic Graphical Templates – Pareto Example.....	53
Basic Graphical Templates – Histogram Example.....	54
Basic Graphical Templates – Run Chart Example	56
Basic Statistical Templates – Sample Size – Discrete.....	58
Basic Statistical Templates – Sample Size – Continuous	60
Basic Statistical Templates – Minimum Sample Size for Robust t-Tests and ANOVA	61
Basic Statistical Templates – 1 Sample Z-Test and Confidence Interval for Mean.....	65
Basic Statistical Templates – 1 Sample t-Test and Confidence Interval for Mean	66
Basic Statistical Templates – 2 Sample t-Test and Confidence Interval (Compare 2 Means).....	67
Basic Statistical Templates – 1 Sample Equivalence Test for Mean	68
Basic Statistical Templates – 2 Sample Equivalence Test (Compare 2 Means)	69

Basic Statistical Templates – 1 Sample Chi-Square Test and CI for Standard Deviation	70
Basic Statistical Templates – 2 Sample F-Test and CI (Compare 2 Standard Deviations).....	71
Basic Statistical Templates – 1 Proportion Test and Confidence Interval	72
Basic Statistical Templates – 2 Proportions Test and Confidence Interval	74
Basic Statistical Templates – 2 Proportions Equivalence Test.....	76
Basic Statistical Templates – 1 Poisson Rate Test and Confidence Interval	78
Basic Statistical Templates – 2 Poisson Rates Test and Confidence Interval	80
Basic Statistical Templates – 2 Poisson Rates Equivalence Test.....	82
Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test.....	84
Basic Statistical Templates – One-Way Chi-Square Goodness-of-Fit Test - Exact	85
Probability Distribution Calculators – Normal	88
Basic MSA Templates – Type 1 Gage Study	89
Basic MSA Templates – Gage Bias and Linearity Study	91
Basic MSA Templates – Gage R&R Study (MSA)	93
Basic MSA Templates - Gage R&R: Multi-Vari & X-bar R Charts.....	95
Basic MSA Templates – Attribute Gage R&R (MSA)	96
Basic Process Capability Templates – Process Sigma Level – Discrete Data	98
Basic Process Capability Templates – Process Sigma Level – Continuous Data	99
Basic Process Capability Templates – Process Capability Indices.....	100
Basic Process Capability Templates – Process Capability & Confidence Intervals	101
Basic Process Capability Templates – Tolerance Interval Calculator (Normal Exact).....	102
Basic Control Chart Templates – Individuals Chart.....	103
Part C – Descriptive/Summary Statistics	106
Descriptive Statistics	106
Descriptive Statistics - Options	108
Part D – Histograms	115
Basic Histogram Template	115
Single (Basic) Histogram.....	115
Multiple Histograms.....	116
Part E – Dotplots	118
Dotplots.....	118
Part F – Boxplots	119

Boxplots	119
Part G – Normal Probability Plots	123
Normal Probability Plots	123
Part H– Run Charts	127
Basic Run Chart Template	127
Run Charts	127
Overlay Run Charts	130
Part I – Measurement Systems Analysis	131
Type 1 Gage Study Template	131
Gage Bias and Linearity Study Template	131
Gage R&R Study (MSA) Template	131
Attribute Gage R&R (MSA) Template	131
Create Gage R&R (Crossed) Worksheet.....	132
Analyze Gage R&R (Crossed).....	134
Attribute MSA (Binary).....	138
Attribute MSA (Ordinal).....	149
Attribute MSA (Nominal)	155
Part J – Process Capability	157
Process Capability Templates	157
Histograms and Process Capability	158
Capability Combination Report (Individuals).....	160
Capability Combination Report (Subgroups)	162
Capability Combination Report (Individuals Nonnormal)	164
Distribution Fitting	172
Box-Cox Transformation	176
Part K – Reliability/Weibull Analysis	178
Reliability/Weibull Analysis.....	178
SigmaXL: Analyze Phase Tools	183
Part A – Stratification with Pareto	185
Basic Pareto Chart Template	186
Basic (Single) Pareto Charts	186
Advanced (Multiple) Pareto Charts	188

Part B - EZ-Pivot	191
Example of Three X's, No Response Y's	191
Example of Three X's and One Y	194
Example of 3 X's and 3 Y's.....	195
Part C – Confidence Intervals.....	196
One Sample Confidence Interval Templates	196
Confidence Intervals – Descriptive Statistics	196
Part D – Hypothesis Testing – One Sample Z and t-Test	199
One Sample Z, t and Equivalence Test Templates.....	199
One Sample t-Test.....	200
Part E – Power and Sample Size.....	203
Power and Sample Size – One Sample t-Test – Customer Data	203
Power and Sample Size – One Sample t-Test – Graphing the Relationships between Power, Sample Size, and Difference	207
Part F – One Sample Nonparametric Tests	209
Introduction to Nonparametric Tests	209
Exact Nonparametric Tests	209
One Sample Sign Test.....	211
One Sample Wilcoxon Signed Rank Test.....	212
One Sample Wilcoxon Signed Rank Test - Exact.....	213
Part G – Two Sample t-Test	219
Two Sample t and Equivalence Test Templates	219
Two Sample t-Test	220
Paired t-Test.....	224
Unpaired 2 Sample t-Test vs. Paired t-Test	226
Power & Sample Size for 2 Sample T-Test	227
Part H – Two Sample F and Comparison Test	229
Two Sample F-Test Template.....	229
Two Sample Comparison Test.....	230
Part I – Two Sample Nonparametric Test: Mann-Whitney.....	232
Two Sample Mann-Whitney Test.....	232
Two Sample Mann-Whitney Test - Exact.....	233

Part J – One-Way ANOVA & Means Matrix	238
One-Way ANOVA & Means Matrix	238
One-Way ANOVA & Means Matrix - Options	241
Power & Sample Size for One-Way ANOVA.....	247
Part K – Two-Way ANOVA.....	249
Two-Way ANOVA	249
Part L – Tests for Equal Variance & Welch’s ANOVA	253
Bartlett’s Test.....	253
Levene’s Test.....	257
Welch’s ANOVA Test (Assume Unequal Variance)	261
Part M – Nonparametric Multiple Comparison	268
Kruskal-Wallis.....	268
Kruskal-Wallis – Exact	271
Mood’s Median Test	275
Mood’s Median Test - Exact	277
Part N – Nonparametric Runs Test	283
Nonparametric Runs Test for Randomness	283
Nonparametric Runs Test for Randomness - Examples.....	284
Nonparametric Runs Test for Randomness - Exact	287
Part O – Attribute/Discrete Data Tests.....	292
1 Proportion Test and Confidence Interval Template	292
2 Proportions Test and Confidence Interval Template.....	292
2 Proportions Equivalence Test Template	292
One-Way Chi-Square Goodness-of-Fit Template.....	292
One-Way Chi-Square Goodness-of-Fit Template - Exact.....	293
1 Poisson Rate Test and Confidence Interval Template	293
2 Poisson Rates Test and Confidence Interval Template.....	293
2 Poisson Rates Equivalence Test Template	293
2 Proportions Test and Confidence Interval Template Example	294
Chi-Square Test – Two-Way Table Data.....	295
Chi-Square Test – Two-Way Table Data: Advanced Tests and Measures of Association – Nominal Categories	300

Computing Odds Ratio and Confidence Interval for 2x2 Table	304
Chi-Square Test – Two-Way Table Data: Advanced Tests and Measures of Association – Ordinal Categories	305
Chi-Square Test (Stacked Column Format Data)	310
Chi-Square Test – Fisher’s Exact	315
Chi-Square Test – Two Way Table Data – Fisher’s Exact	319
References for Fisher’s Exact	323
Power & Sample Size for One Proportion Test	324
Power & Sample Size for Two Proportions Test	326
Part P – Analysis of Means (ANOM) Charts	328
Introduction – What is ANOM	328
ANOM Normal One-Way	332
ANOM Normal Two-Way (with Main Effects and Slice Charts)	334
ANOM Binomial Proportions One-Way	342
ANOM Binomial Proportions Two-Way	344
ANOM Poisson Rates One-Way	346
ANOM Poisson Rates Two-Way	348
ANOM Nonparametric Transformed Ranks	350
ANOM Variances	354
ANOM Levene Robust Variances	356
Part Q – Multi-Vari Charts	358
Multi-Vari Charts	358
Part R – Scatter Plots	362
Scatter Plots	362
Scatter Plot Matrix	364
Part S – Correlation Matrix	365
Correlation Matrix	365
Reference	366
Part T – Multiple Regression	367
Multiple Regression	367
Multiple Regression with Continuous and Categorical Predictors	371
Part U – Logistic Regression	373

Binary Logistic Regression.....	373
Ordinal Logistic Regression	379
SigmaXL: Improve Phase Tools: Design of Experiments (DOE).....	383
Part A – Overview of Basic Design of Experiments (DOE) Templates	385
Part B – Three Factor Full Factorial Example Using DOE Template.....	386
Part C – Design and Analysis of Catapult Full Factorial Experiment.....	395
Part D – Design and Analysis of Response Surface Experiment – Cake Bake	406
SigmaXL: Control Phase Tools: Statistical Process Control (SPC) Charts.....	413
Part A - Individuals Charts.....	415
Individuals Chart Template.....	415
Individuals Chart	416
Individuals Charts: Advanced Limit Options – Historical Groups	424
Individuals & Moving Range Charts	426
Individuals Charts for Nonnormal Data (Box-Cox Transformation).....	427
Part B - X-Bar & Range/StDev Charts	431
X-Bar & R Charts.....	431
X-Bar & R Charts – Exclude Subgroups	437
X-Bar & S Charts	439
Part C – Attribute Charts (P, NP, C, U)	441
P-Charts.....	441
P-Charts: Advanced Limit Options – Historical Groups	442
NP-Charts	444
C-Charts.....	445
U-Charts	446
Part D – P' & U' Charts (Laney)	447
P'-Charts.....	447
U'-Charts	449
Part E – Control Chart Selection Tool	451
Data Types and Definitions.....	451
Control Chart Selection Tool – Individuals Chart.....	452
Control Chart Selection Tool – X-Bar & R Chart.....	453
Control Chart Selection Tool – P-Chart.....	454

Part F – Advanced Charts: I-MR-R/S	455
I-MR-R Charts	455
SigmaXL Appendix	459
SigmaXL Version 8.0 Feature List Summary.....	460
Descriptive Statistics Options and Tolerance Interval Calculator (Normal Exact).....	464
Shapiro-Wilk (SW) and Kolmogorov-Smirnov-Lilliefors (KSL) Normality Test	464
Doornik-Hansen (DH) Normality Test	464
Confidence and Tolerance Intervals	464
Tolerance Interval Calculator (Normal Exact).....	465
Percentile (Nonparametric) Confidence and Tolerance Intervals.....	466
References for Descriptive Statistics Options and Tolerance Interval Calculator (Normal Exact).....	467
Statistical Details for Nonnormal Distributions and Transformations	468
Maximum Likelihood Estimation (MLE).....	468
Distributions	469
Transformations	481
Automatic Best Fit.....	484
Process Capability Indices (Nonnormal)	485
Control Charts (Nonnormal)	486
Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests	487
Introduction	487
Full Enumeration.....	487
Exact.....	488
Monte Carlo	489
Two Tailed (Sided) Tests	489
Sample Sizes for Exact.....	490
Power	490
Validation of SigmaXL Exact P-Values.....	491
1 Sample Tests	491
2 Sample Tests	491
K Sample Tests	491
Contingency Tests	492

References for Exact and Monte Carlo P-Values	492
Hypothesis Test Assumptions Report	494
Normality	494
Robustness	494
Outliers (Boxplot Rules)	497
Randomness (Independence)	498
Equal Variance	499
Attribute Measurement Systems Analysis.....	501
Percent Confidence Intervals (Exact Versus Wilson Score)	501
Kappa	501
Concordance	502
Kendall's Coefficient of Concordance	502
Kendall's Correlation Coefficient	503
References for Attribute Measurement Systems Analysis	505
Chi-Square Tests and (Contingency) Table Associations	506
Chi-Square Tests for Nominal Categories	506
Measures of Association for Nominal Categories	507
Tests for Ordinal Categories	509
Measures of Association for Ordinal Categories	510
References for Chi-Square Tests and (Contingency) Table Associations.....	512
Multiple Comparison of Means and Variances (a.k.a. Post-Hoc Tests).....	514
One-Way ANOVA	514
Welch's ANOVA (Unequal Variance).....	515
Bartlett's Test for Equal Variance	515
Levene's Test for Equal Variance	515
References for Multiple Comparisons of Means and Variances	516
Analysis of Means (ANOM) Charts.....	517
ANOM Normal One-Way	517
ANOM Normal Two-Way	518
ANOM Binomial Proportions One-Way	519
ANOM Binomial Proportions Two-Way	520
ANOM Poisson Rates One-Way	521

ANOM Poisson Rates Two-Way	521
ANOM Nonparametric Transformed Ranks.....	522
ANOM Variances	522
ANOM Levene Robust Variances	524
References for Analysis of Means (ANOM) Charts	524
About SigmaXL, Inc.	527

SigmaXL® What's New in Version 8.0, System Requirements and Getting Help

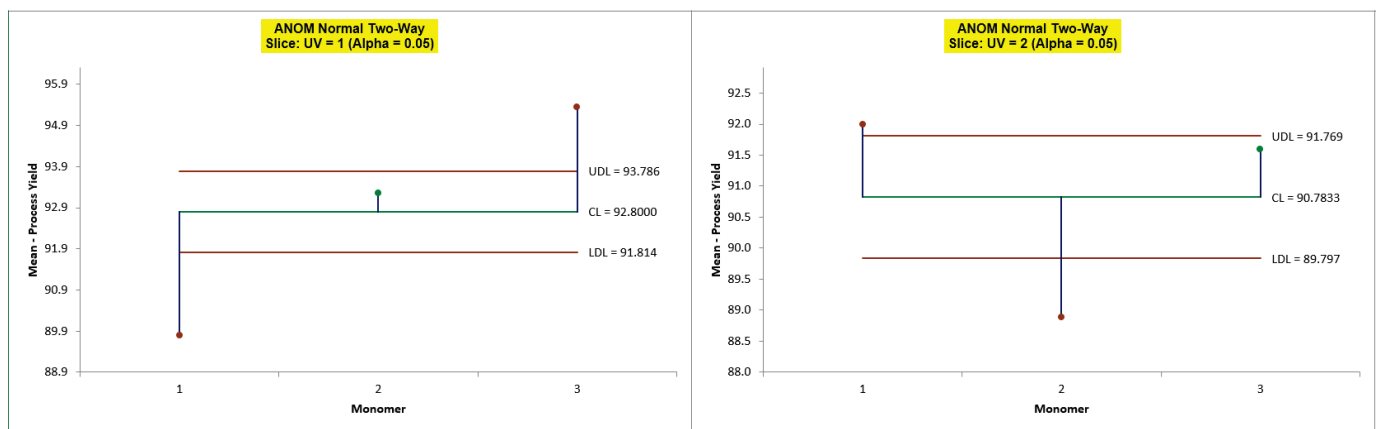
What's New in Version 8.0

See [Appendix: SigmaXL Version 8.0 Feature List Summary](#) for a complete feature list summary.

New features in SigmaXL Version 8.0 include:

- **Templates and Calculators**
 - 1 Sample Z-test and Confidence Interval for Mean
 - Tolerance Interval Calculator (Normal Exact)
 - Equivalence Tests - Two One-Sided Tests (TOST)
 - 1 Sample Equivalence Test for Mean
 - 2 Sample Equivalence Test (Compare 2 Means)
 - 2 Proportions Equivalence Test
 - 2 Poisson Rates Equivalence Test
 - Type 1 Gage Study
 - Gage Bias and Linearity Study
- **Descriptive Statistics**
 - New features are available by clicking the dialog Options button
 - Percentile Report
 - 27 values from 0.135 to 99.865
 - Percentile Ranges
 - 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)
 - Percentile Confidence and Tolerance Intervals
 - Interpolated or Exact
 - Minimum sample size reported if unable to compute CI or TI
 - Quartile Confidence Intervals (25, 50, 75)
 - Percentile Confidence Intervals (27 values from 0.135 to 99.865)
 - Percentile Tolerance Intervals (50%, 80%, 90%, 95%, 98%, 99%, 99.73%)
 - Additional Descriptive Statistics
 - 5% Trimmed Mean
 - Standard Error of Mean
 - Variance
 - Coefficient of Variation
 - Short Term StDev (MR-bar/d2)

- Additional Normality Tests
 - Shapiro-Wilk ($n \leq 5000$) and Kolmogorov-Smirnov-Lilliefors ($n > 5000$)
 - Doornik-Hansen
 - Univariate omnibus test based on Skewness and Kurtosis
 - Best for data with ties ("chunky" data)
 - Outlier Tests
 - Boxplot: Potential 1.5(IQR), Likely 2.2(IQR), Extreme 3.0(IQR)
 - Grubbs
 - Randomness Test
 - Nonparametric Runs Test (Exact)
 - Normality, Outlier and Randomness Tests display the same Green, Yellow, Red highlight used in Version 7 hypothesis tests.
- **Analysis of Means (ANOM) Charts**
 - All charts support balanced & unbalanced data
 - Normal One-Way
 - Exact critical values via table lookup are used rather than approximation formulas
 - Normal Two-Way with Main Effects and Wludyka 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.
 - Yellow highlight automatically recommends Main Effects (if interaction is not significant) or Slice Chart (if interaction is significant).
 - Interaction P-Value automatically determined from ANOVA
 - Option to specify correction to alpha for multiple chart family-wise error rate (Bonferroni)



- Binomial Proportions One-Way
 - Warning is given if normal approximation to binomial is not applicable ($np, nq \leq 5$)

- Binomial Proportions Two-Way with Main Effects and Wludyka & Noguera Slice Charts
 - In collaboration, Peter Wludyka and John Noguera of SigmaXL extended the Slice Charts to Binomial and Poisson.
 - Interaction P-Value automatically determined from Logistic regression
- Poisson Rates One-Way
 - Warning is given if normal approximation to Poisson is not applicable ($\mu \leq 5$)
- Poisson Rates Two-Way with Main Effects and Wludyka & Noguera Slice Charts
 - Interaction P-Value automatically determined from Poisson regression
- Nonparametric Transformed Ranks
- Variances
- Levene Robust Variances
- **One-Way ANOVA**
 - New features are available by clicking the dialog Options button
 - Multiple Comparison of Means Probability Methods (a.k.a. Post-Hoc)
 - Fisher
 - Tukey
 - Dunnett with Control
 - Display ANOM Normal One-Way Chart
 - Confidence Level
 - Residual Plots

Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.8117	-0.2476
2		0	0.5641
3			0

Tukey Probabilities	1	2	3
1		0.0000	0.3777
2			0.0044
3			

- **Welch ANOVA (Assume Unequal Variance)**
 - New features are available by clicking the dialog Options button
 - Multiple Comparison of Means Probability Methods
 - Welch Pairwise
 - Games-Howell
 - Confidence Level
 - Residual Plots
- **Two-Way ANOVA**
 - Display ANOM Normal Two-Way Chart

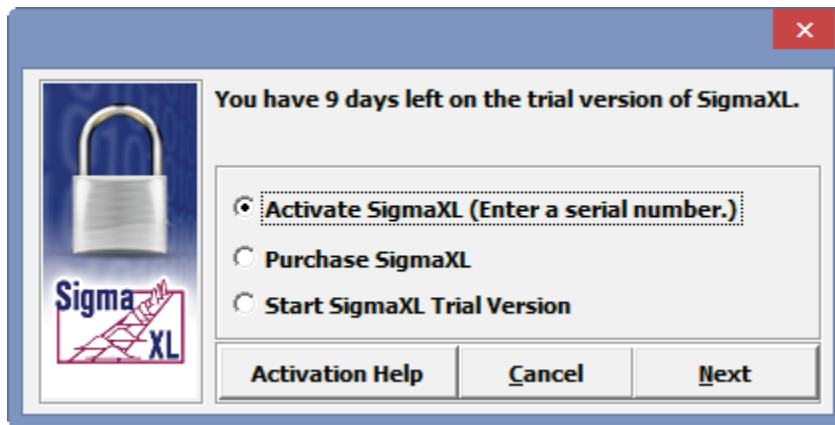
- **Bartlett Equal Variance**
 - New features are available by clicking the dialog Options button
 - Multiple Comparison of Variances Probability Methods
 - F-Test Pairwise
 - F-Test with Bonferroni Correction
 - Display ANOM Variances Chart
 - Confidence Level
- **Levene Equal Variance**
 - New features are available by clicking the dialog Options button
 - Multiple Comparison of Variances Probability Methods
 - Levene Pairwise
 - Tukey ADM (Absolute Deviations from Median)
 - Display ANOM Levene Robust Variances Chart
 - Confidence Level
- **Chi-Square Tests & Table Associations**
 - Improved dialog labels for stacked data (Rows, Cols, Frequency)
 - Adjusted Residuals
 - 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
 - Confidence Level
 - Additional Chi-Square Tests (optional):
 - Likelihood Ratio
 - McNemar-Bowker Symmetry (Square Table)
 - Measures of Association for Nominal Categories (optional)
 - Pearson's Phi
 - Cramer's V
 - Contingency Coefficient
 - Cohen's Kappa (Agreement - Square Table)
 - Goodman-Kruskal Lambda (Cols & Rows Dependent, Symmetric)
 - Goodman-Kruskal Tau (Cols & Rows Dependent)
 - Theil's Uncertainty (Cols & Rows Dependent, Symmetric)
 - Tests of Association for Ordinal Categories (optional)
 - Concordant - Discordant
 - Spearman Rank Correlation
 - Measures of Association for Ordinal Categories (optional)
 - Spearman Rank Correlation
 - Kendall's Tau-B (Square Table)
 - Kendall-Stuart Tau-C (Rectangular Table)
 - Goodman-Kruskal Gamma
 - Somers' D (Cols & Rows Dependent, Symmetric)

Activating SigmaXL

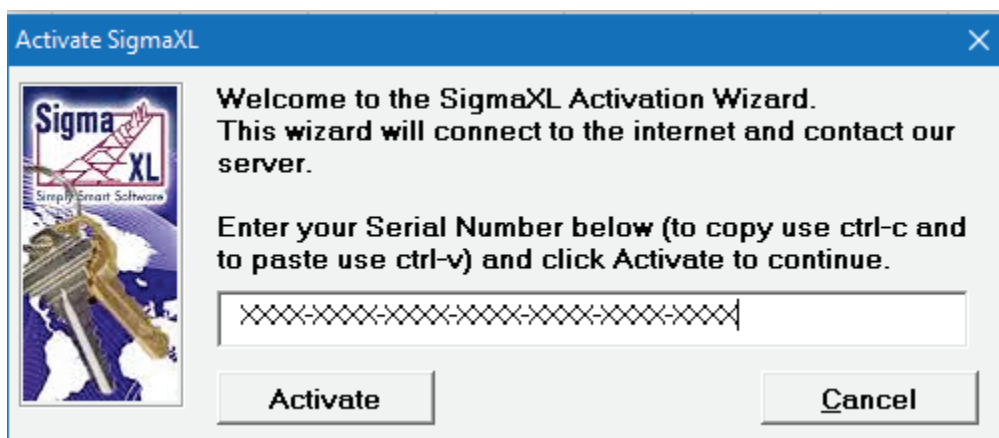
Please proceed with the following steps if you have a valid serial number and your computer is connected to the Internet. If you do not have an Internet connection, please email Support@SigmaXL.com from a different system or call 1-888-SigmaXL (1-888-744-6295) or 1-519-579-5877. If you purchased SigmaXL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

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.

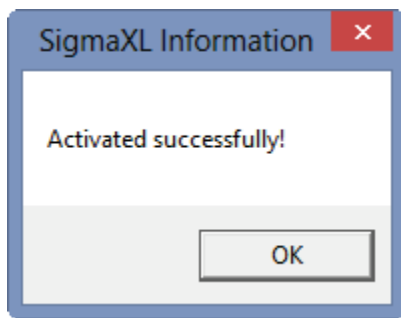
1. In the **Activation Wizard** box select **Activate SigmaXL (Enter a serial number.)**



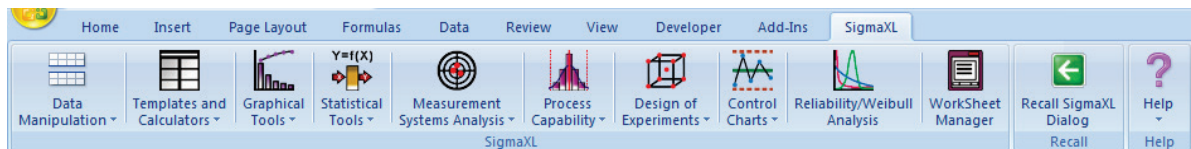
2. Click **Next**. Enter your serial number. We recommend that you copy and paste the serial number into the serial number field to avoid typos.



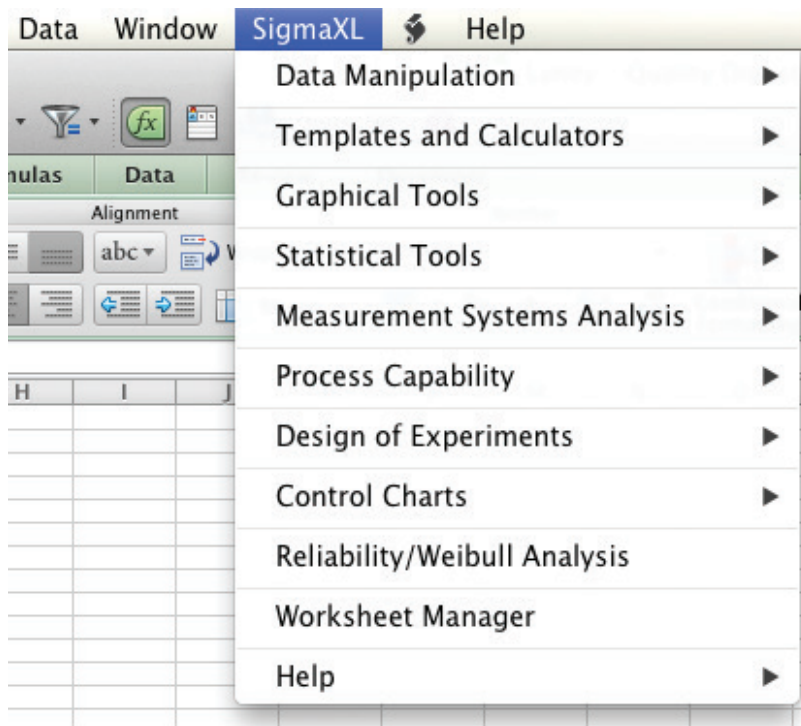
3. Click **Next**. The activation process continues and is confirmed as shown:



4. Click **OK**. In Excel 2007 and newer, SigmaXL appears as a new Ribbon:

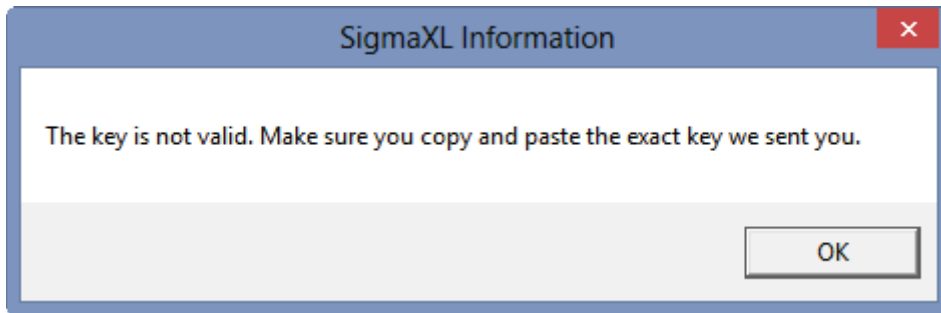


5. In Excel 2011 for Mac, SigmaXL is added to Excel's menu:



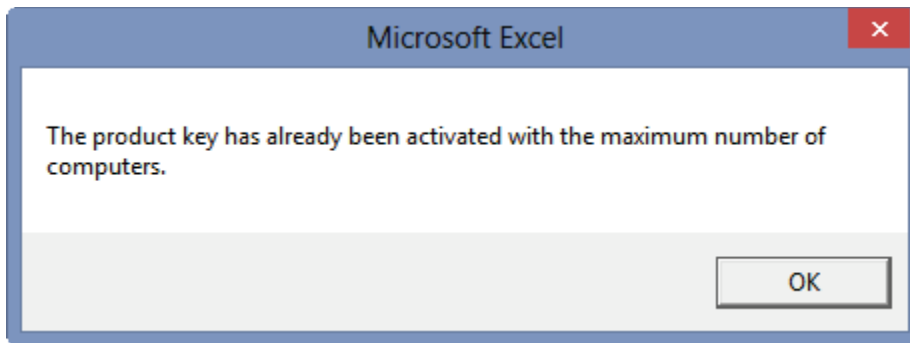
Error Messages

1. Incorrect Serial Number:



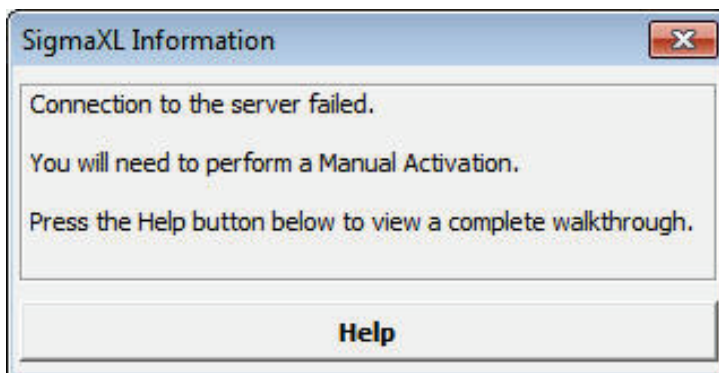
This is due to an incorrect serial number entry in the previous registration window – click **OK** and please reenter your serial number.

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:



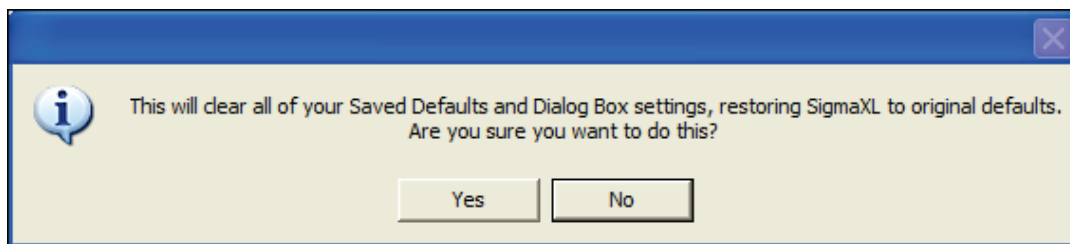
If you do not have an Internet connection, please email Support@SigmaXL.com from a different system or call 1-888-SigmaXL (1-888-744-6295) or 1-519-579-5877.

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.

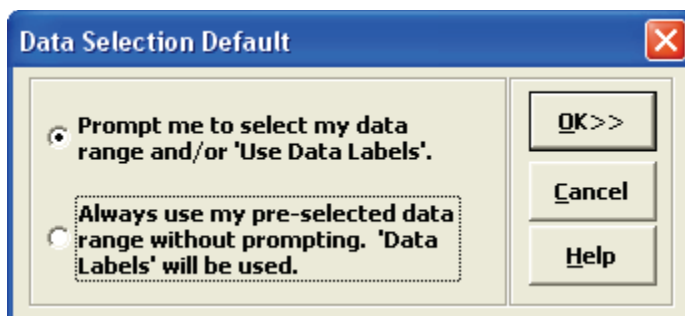


Data Selection Default

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.

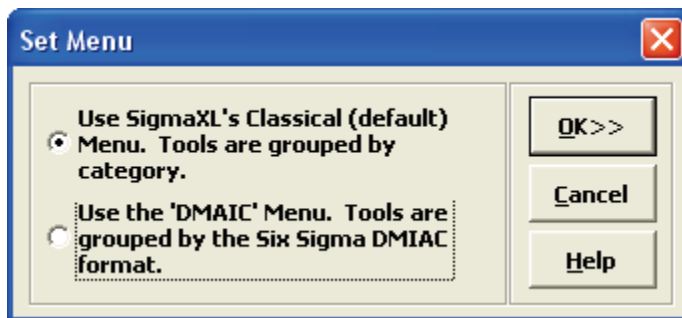
Click **SigmaXL > Help > SigmaXL Defaults > Data Selection Default** to make this change. This will apply permanently unless you revert back to the **Prompt me** setting or click **Clear Saved Defaults** shown above.



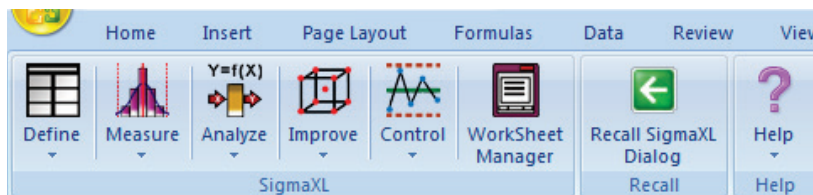
Menu Options (Classical or DMAIC)

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 (Excel 2007 and newer):



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 gigahertz (GHz) processor or higher.

Memory: 2 GB of RAM or greater.

Hard disk: 200 MB of available hard-disk space.

Drive: CD-ROM or DVD drive.

Display: 1024x768 or higher resolution monitor.

Windows Operating System: Microsoft Windows 7 with current Service Packs, or later operating system.

Microsoft Windows Excel version: Excel 2007 and newer with latest service packs installed.

Mac Operating System: OS X Snow Leopard (10.6) or later

Mac Excel Version: Excel 2011 or later 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

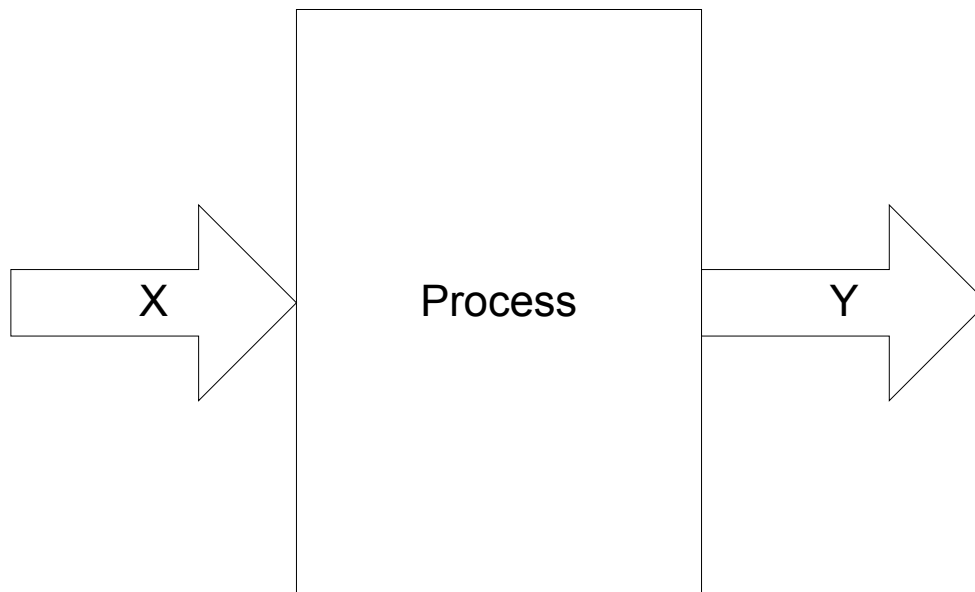
Introduction

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. 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.	<ul style="list-style-type: none"> ➤ 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
Histogram	Visual display of one variable showing data center, spread, shape and outliers.	Y=Continuous X=Discrete (Category)	<ol style="list-style-type: none"> 1. Summarize large amounts of data 2. To get a 'feel for the data' 3. To compare actual description to customer specifications 	<ul style="list-style-type: none"> ➤ 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)	<ol style="list-style-type: none"> 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
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
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	<ul style="list-style-type: none"> ➤ 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.	<ul style="list-style-type: none"> ➤ SigmaXL > Graphical Tools > Scatter Plot ➤ SigmaXL > Graphical Tools > Scatter Plot Matrix
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 >

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)	<ol style="list-style-type: none"> Test if mean = specified value Test if 2 sample means are equal Paired t. to reduce variation when comparing two sample means 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) SigmaXL > Statistical Tools > Two-Way ANOVA
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)	<ol style="list-style-type: none"> Test if median = specified value: 1 Sample Sign Test or Wilcoxon Test if 2 sample medians are equal: 2 Sample Mann-Whitney 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)	<ol style="list-style-type: none"> Determine the confidence interval for a single standard deviation. Test if 2 sample variances (standard deviations) are equal. 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)	<ol style="list-style-type: none"> Determine the confidence interval for a single proportion. 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)	<ol style="list-style-type: none"> Determine the confidence interval for a single rate. 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)
X ² 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)
Anderson-Darling Normality Test	Determine if the data is normally distributed.	Y=Continuous	Test if the sample data is normally distributed.	SigmaXL > Statistical Tools > Chi-Square Tests – Exact (for small sample) SigmaXL > Graphical Tools > Histograms & 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 > Descriptive Statistics SigmaXL > Statistical Tools > Correlation Matrix
Regression (Simple Linear & Multiple)	Summarizes, describes, predicts and quantifies relationships.	Y=Continuous X=Continuous	<ol style="list-style-type: none"> Determine if there is evidence of a relationship between Xs and Ys. Model data to develop a mathematical equation to quantify the relationship. Identify root causes. Make predictions using the model. 	SigmaXL > Statistical Tools > Regression > Multiple Regression
Logistic Regression	Summarizes, describes, predicts and quantifies relationships.	Y=Discrete (Binary or Ordinal) X=Continuous		SigmaXL > Statistical Tools > Regression > Binary Logistic Regression Ordinal Logistic Regression
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

SigmaXL: Measure Phase Tools

Part A - Basic Data Manipulation

Introduction to Basic Data Manipulation

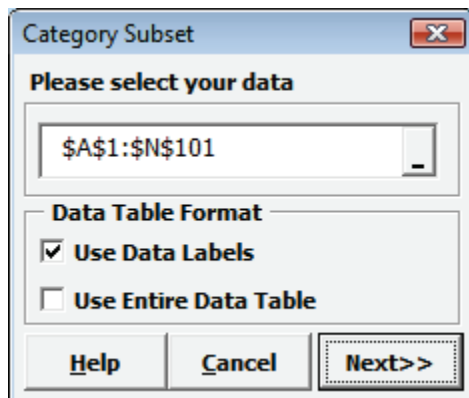
Open **Customer Data.xlsx** (to access, click **SigmaXL > Help > Sample Data** or **Start > Programs > SigmaXL > Sample Data**). This data is in **stacked** column format. This format is highly recommended for use with SigmaXL. Note that all pertinent information is provided in each record (row). Also 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	3.86	Small	Not-available	Consumer
6	1/5/2001	2	52.2	55	4	4.12	4.51	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.

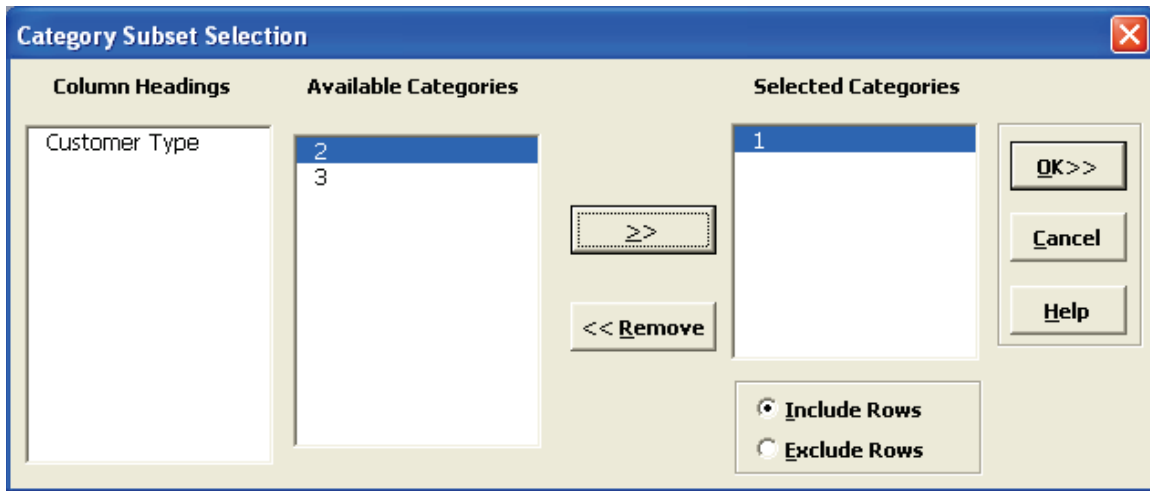
Category Subset

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:



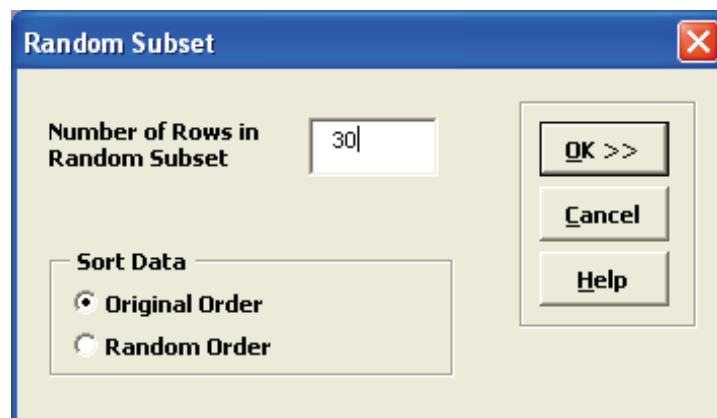
5. Click **OK**.

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

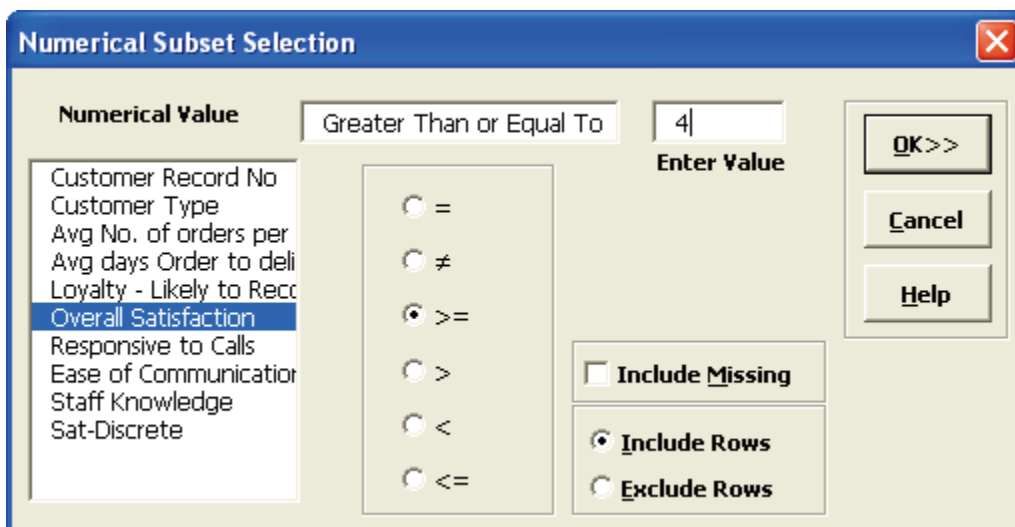
1. Click **Sheet 1** Tab of **Customer Data.xlsx**.
2. Click **SigmaXL > Data Manipulation > Random Subset**.
3. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
4. Enter **Number of Rows in Random Subset** as 30. The default **Sort Data** selection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



- 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

- Click **Sheet 1** Tab of **Customer Data.xlsx**.
- Click **SigmaXL > Data Manipulation > Numerical Subset**.
- Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
- Select *Overall Satisfaction*, **>=**, **Enter Value** as 4.



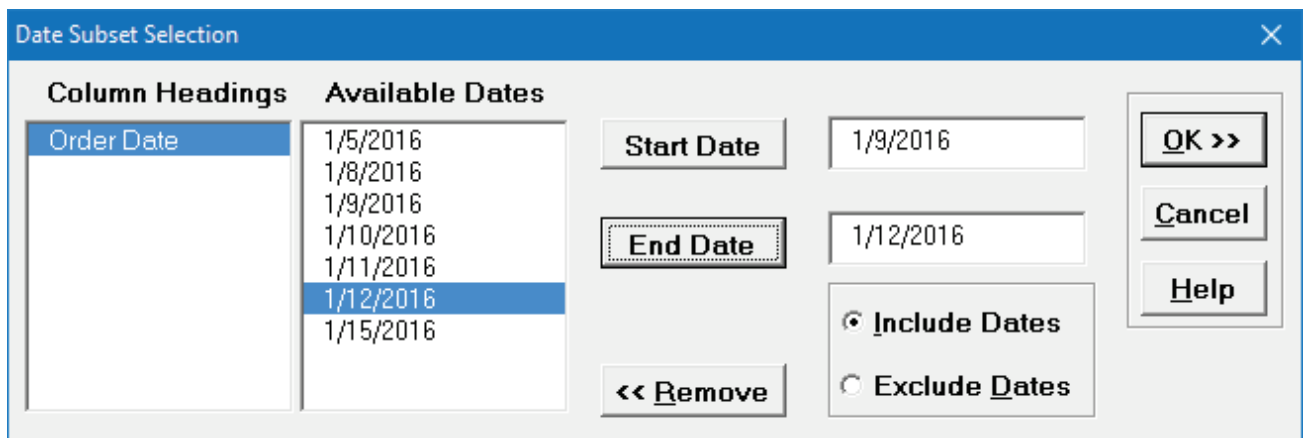
- Click **OK**.

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

Date Subset

- Click **Sheet 1** Tab of **Customer Data.xlsx**.
- Click **SigmaXL > Data Manipulation > Date Subset**.
- 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**.



The dialog box titled "Date Subset Selection" has a blue header bar with a close button (X) on the right. It contains two main sections: "Column Headings" and "Available Dates".

Column Headings	Available Dates
Order Date	1/5/2016
	1/8/2016
	1/9/2016
	1/10/2016
	1/11/2016
	1/12/2016
	1/15/2016

Below the "Available Dates" list are three buttons: "Start Date", "End Date", and "<< Remove". To the right of these buttons are two date input fields: the first contains "1/9/2016" and the second contains "1/12/2016". Below these fields are two radio buttons: "Include Dates" (which is selected) and "Exclude Dates". On the far right, there are three buttons: "OK >>", "Cancel", and "Help".

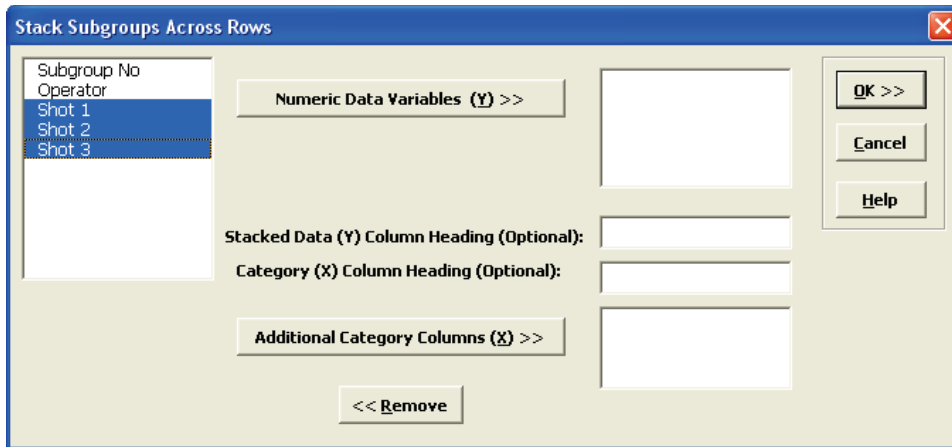
5. Click **OK**. A new subset worksheet is created containing only those rows with Order Date between 1/9/2016 to 1/12/2016.

Transpose Data

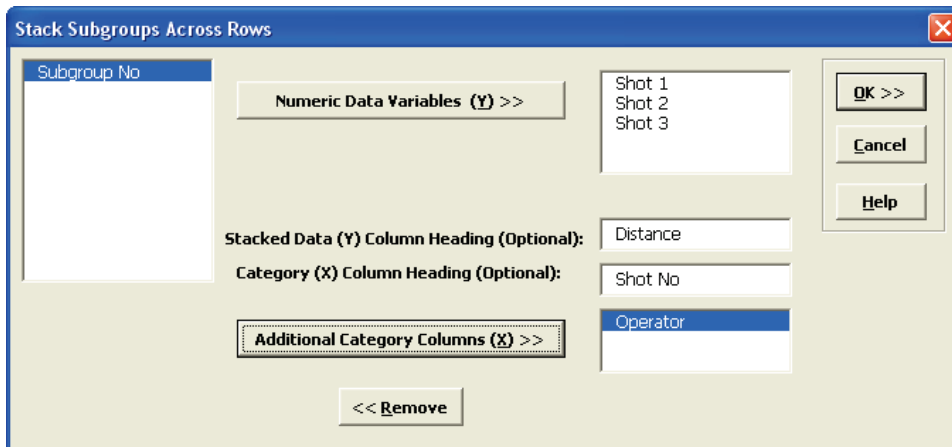
1. Open **Catapult Data Row Format.xlsx**.
2. Manually select the entire data table if the data is not already selected.
3. Click **SigmaXL > Data Manipulation > Transpose Data**.
4. This will transpose rows to columns or columns to rows. It is equivalent to **Copy, Paste Special, Transpose**.

Stack Subgroups Across Rows

5. 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.
6. Click **SigmaXL > Data Manipulation > Stack Subgroups Across Rows**.
7. Check **Use Entire Data Table**, click **Next**.
8. Click on *Shot 1*. Shift Click on *Shot 3* to highlight the three columns of interest.



9. 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.

10. 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

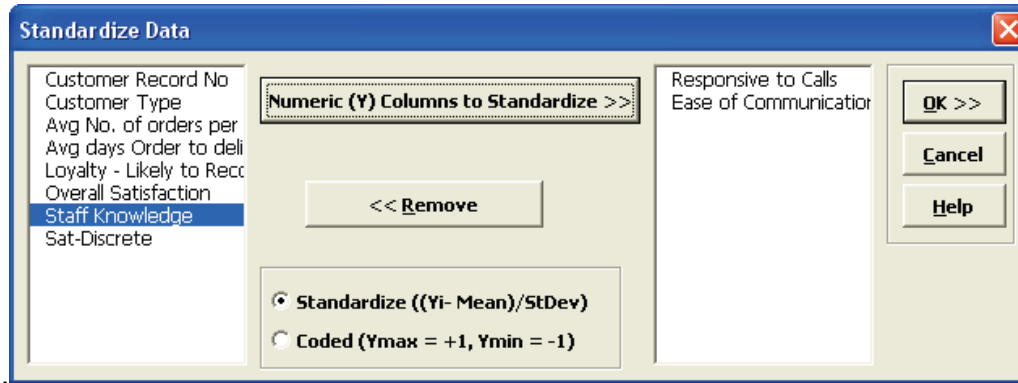
This tool is used to convert nonnormal data to normal by applying a power transformation. Examples of use are given in **Measure Phase Tools, Part J – Process Capability for Nonnormal Data** and **Control Phase Tools, Part A – Individuals Charts for Nonnormal Data**.

Standardize Data

This tool is used to Standardize $((Y_i - \text{Mean})/\text{StDev})$ or Code ($Y_{\text{max}} = +1, Y_{\text{min}} = -1$) 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

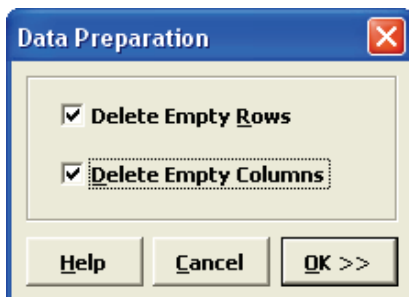
Data Preparation – Remove Blank Rows and Columns

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

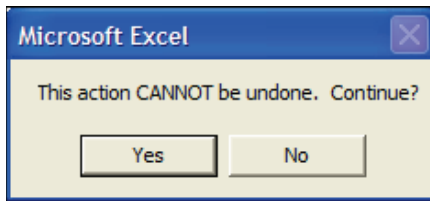
1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab.
2. Insert a new column in B; Click **Column B heading**, **Right-Click > Insert > Columns**.
3. Insert a new row in row 2. Click **Row 2 label**, **Right-Click > Insert > Rows** as shown:

	A	B	C	D
1	Customer Record No		Order Date	Customer Type
2				
3		1	1/5/2001	2
4		2	1/5/2001	3
5		3	1/5/2001	2
6		4	1/5/2001	2
7		5	1/5/2001	3
8		6	1/5/2001	2
9		7	1/5/2001	1
10		8	1/5/2001	2
11		9	1/5/2001	2
12		10	1/5/2001	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 2007 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 2011). This feature can also be accessed by clicking **SigmaXL > Help > Hot Keys > Recall SigmaXL Dialog**.

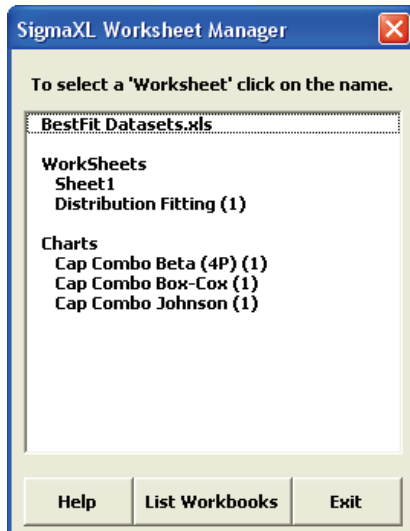
Note that **Recall SigmaXL Dialog** may not be available for all functions.

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 2011). This feature can also be accessed by clicking **SigmaXL > Help > Hot Keys > Activate Last Sheet**.

Worksheet Manager

The **Worksheet Manager** is a powerful utility that allows you to list all open Excel workbooks, display all worksheets and chart sheets in a selected workbook and quickly select a worksheet or chart sheet of interest. To access, click **SigmaXL > Worksheet Manager**.



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 **Tools > Options**, select **Calculation, Automatic**, and click **OK** (Excel 2007: **Office Button | Excel Options | Formulas**, select **Automatic**, click **OK**; Excel 2010: **File | Options...**). Templates and Calculators are protected worksheets by default, but this may be modified by clicking **SigmaXL > Help > Unprotect Worksheet**. Alternatively, you can click Excel **Tools > Protection > Unprotect Sheet** (Excel 2007 and newer: **Review | Changes | Unprotect Sheet**). Gage R&R is password protected, so it cannot be unprotected.

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 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 Control Chart Templates:**

- Individuals
- C-Chart

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:				Team:		
Task/Phase	Start Date	End Date	Actual End	Name:	Role:	Commitment (%)

SIPOC Diagram

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

SIPOC DIAGRAM

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

Suppliers	Inputs		Process	Outputs		Customers
Provider	Input Description	Input Requirements (optional)	<div style="border: 1px solid black; padding: 10px; display: inline-block;"> See High Level Process Steps Below </div>	Output Description	Output Requirements (optional)	Recipient of Output

Start Boundary:

Step 1

→

Step 2

→

Step 3

→

Step 4

→

Step 5

End Boundary:

Data Measurement Plan

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

DATA MEASUREMENT PLAN

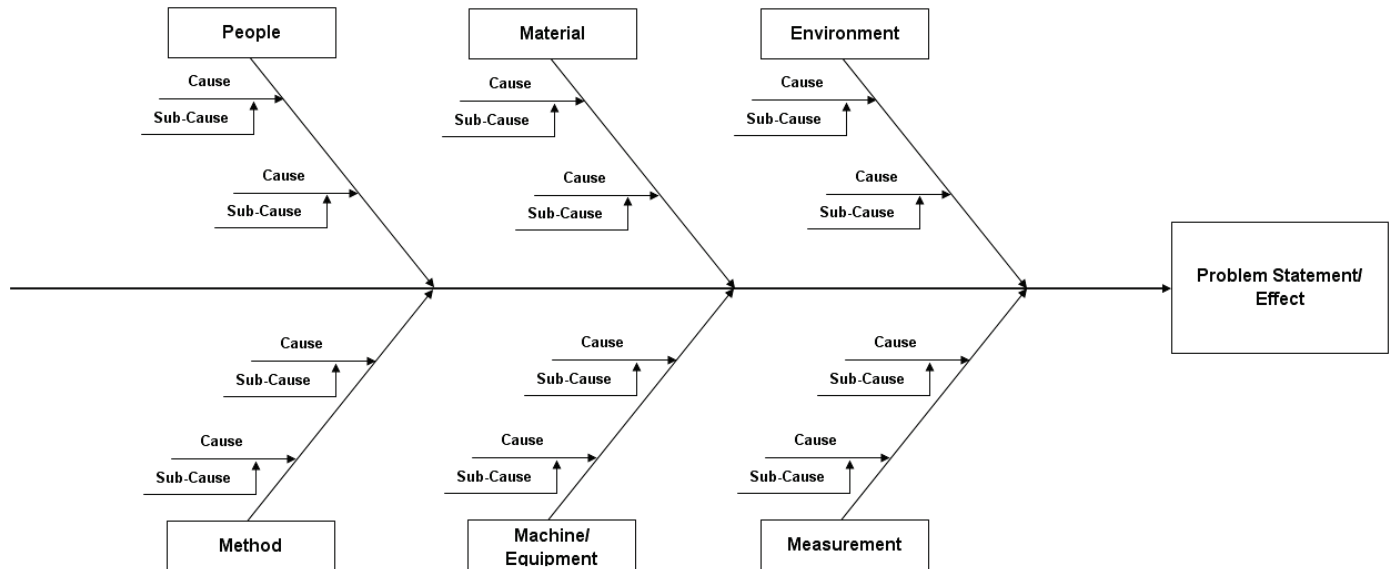
[illegible]

Cause & Effect (Fishbone) Diagram

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

CAUSE & EFFECT (FISHBONE) DIAGRAM

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



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:	

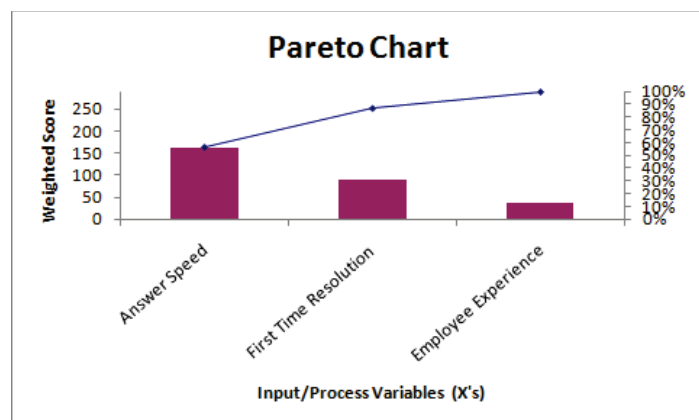
	Call Abandon Rate	Customer Satisfaction	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Pareto Chart
Output Variables (Y's):											
Importance Score (1-10):	8	10									

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
											n

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	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)
<div>Sort</div>									<div>Sort</div>
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:

Severity (1-10)	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
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

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
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

Bad
↓
Good

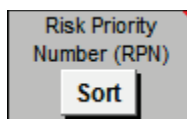
Score	Occurrence Guidelines	
	Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma
10	Very High: ≥ 100 per thousand; ≥ 1 in 10	More than once per day > 30%
9	High: 50 per thousand; 1 in 20	Once every 3-4 days < 30%
8	High: 20 per thousand; 1 in 50	Once every week < 5%
7	High: 10 per thousand; 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

Bad
↓
Good

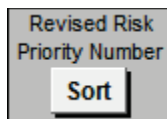
Score	Detection Guidelines	
	Automotive Industry Action Group AIAG (PFMEA 4th ed.)	Six Sigma
10	Almost Impossible. No detection opportunity: No current process control; cannot detect or is not analyzed.	Defect caused by failure is not detectable
9	Very Remote. Not likely to detect at any stage: Failure 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

Bad
↓
Good

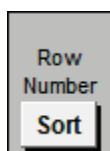
Click **Risk Priority Number (RPN) Sort** to sort the rows by RPN in descending order.



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.

		Functional Requirements (How):										Satisfaction Ratings:						
		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 Requirements (What):	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	0						
Rank:		2	3	4	1	5	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:

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	0	
Sum of Negatives (-):	0	0	2	0	0	0	0	0	0	0	0	
Sum of Sames (S):	1	3	3	0	0	0	0	0	0	0	0	
Positives - Negatives:	4	2	-2	0	0	0	0	0	0	0	0	
Weighted Sum of Positives (+):	16	8	0	0	0	0	0	0	0	0	0	
Weighted Sum of Negatives (-):	0	0	7	0	0	0	0	0	0	0	0	
Weighted Sum of Sames (S):	3	11	12	0	0	0	0	0	0	0	0	
Weighted Positives - Weighted Negatives:	16	8	-7	0	0	0	0	0	0	0	0	

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	<input type="text" value="22"/>
Scheduled Work:	hours per shift	<input type="text" value="8"/>
Shifts per Day:		<input type="text" value="2"/>
Lunch:	minutes per shift	<input type="text" value="30"/>
Breaks:	minutes per shift	<input type="text" value="30"/>
Planned Downtime:	minutes per shift	<input type="text" value="30"/>
Staff/Operator Cycle Time:	minutes per unit	<input type="text" value="226"/>
Available Time:	minutes per day	<input type="text" value="780.0"/>
Takt Time:	minutes per unit	<input type="text" value="35.5"/>
Required Number of Staff/Operators:		<input type="text" value="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:

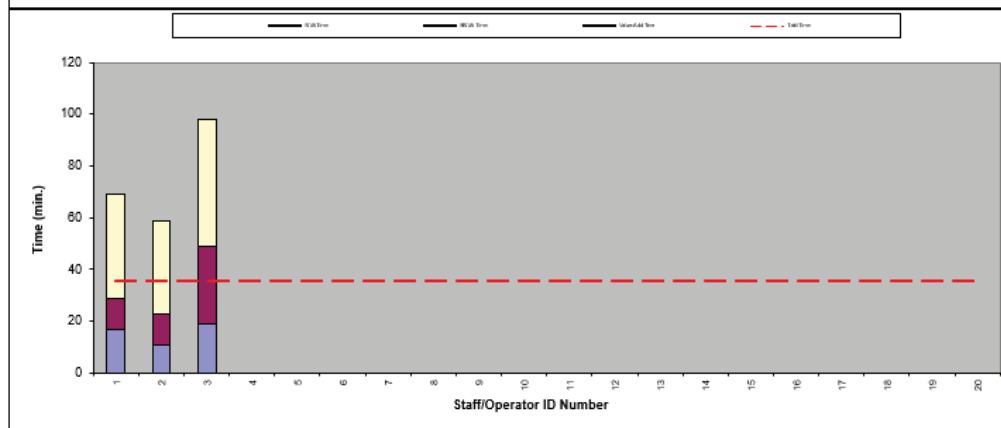
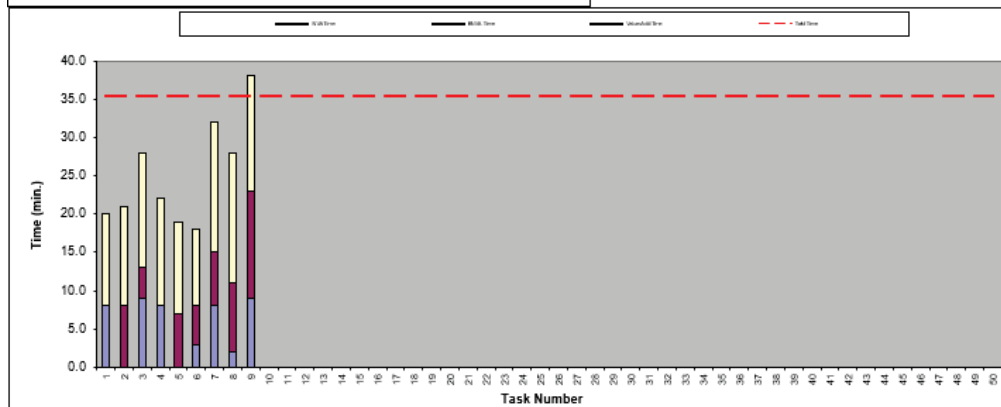
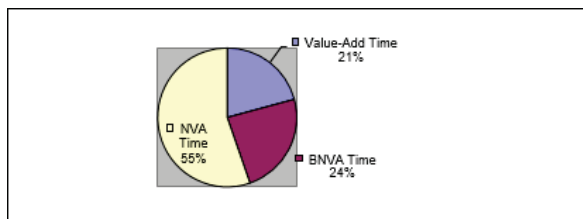
SigmaXL Lean Templates: Value Analysis/Process Load Balance

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

Information Symbols	
Production Control	Central production scheduling or control
Production Kanban	Triggers production of a pre-defined number of units.
Information Type	Information type includes: Monthly Forecast, Weekly Orders or Schedule, Daily Orders or Schedule
Manual Information Flow	Manual information flow from memos, reports, or verbal. Add information type icon above as necessary.
Electronic Information Flow	Information flow via computer network. Add information type icon as necessary.
MRP/ERP	Schedule using MRP/ERP system
Go See	Gather information visually
Withdrawal Kanban	Instruction to obtain items from the supermarket
Signal Kanban	Initiate a batch operation
Kanban Post	A physical location for kanban cards
Load Leveling	Batch kanbans to load level production
Sequenced Pull	Kanban signal, without using a supermarket

SigmaXL Inc., Copyright © 2011-2014

Portions courtesy and copyright of Strategos, Inc. www.strategosinc.com

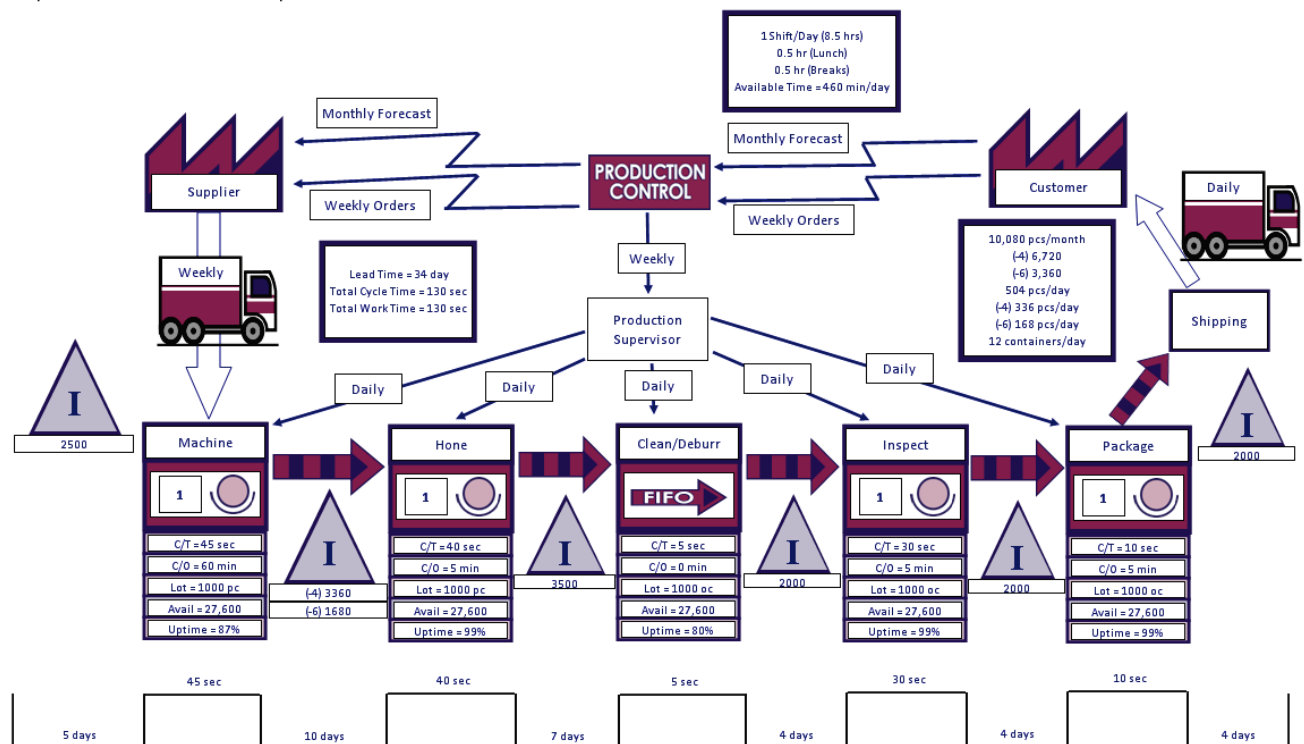
Notes:

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.

Open the file **Template & Calculator Examples.xlsx**. Click on the worksheet named **VSM Example**. 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

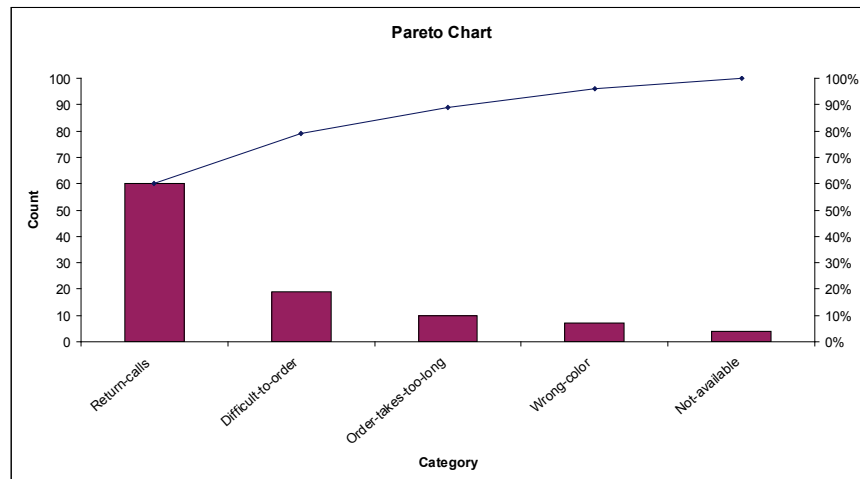
Click **SigmaXL > Templates & Calculators > Basic Graphical Templates > Pareto Chart**. Enter data as shown:

A	B	C	D	E
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.

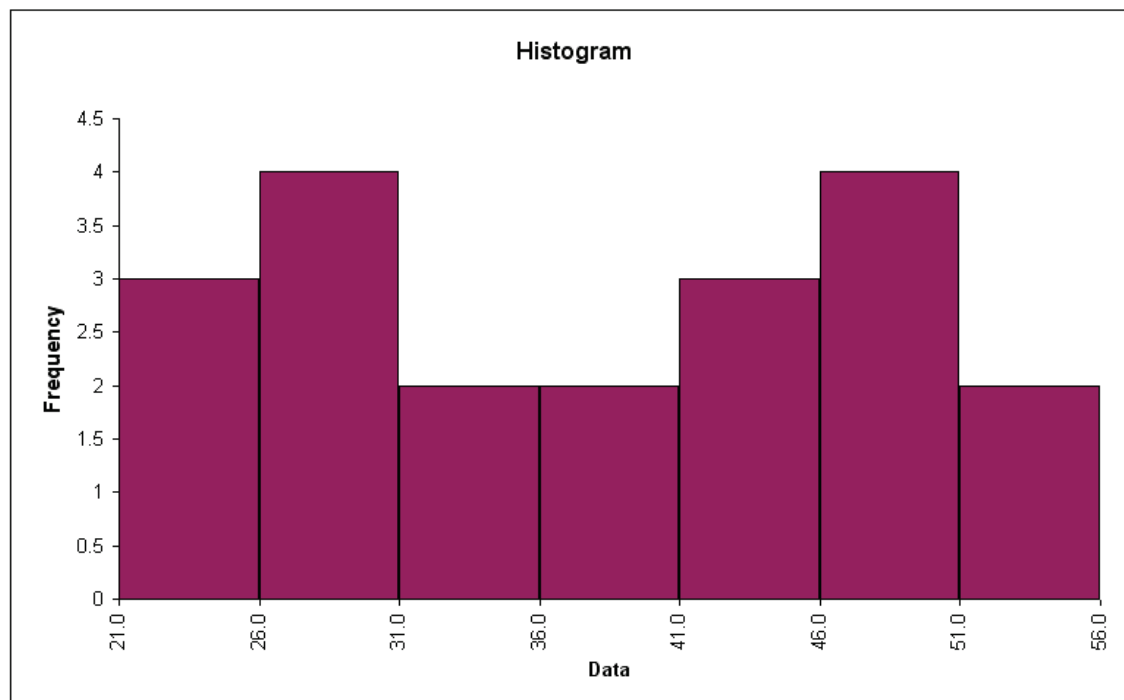
Basic Graphical Templates – Histogram Example

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	

Histogram

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.

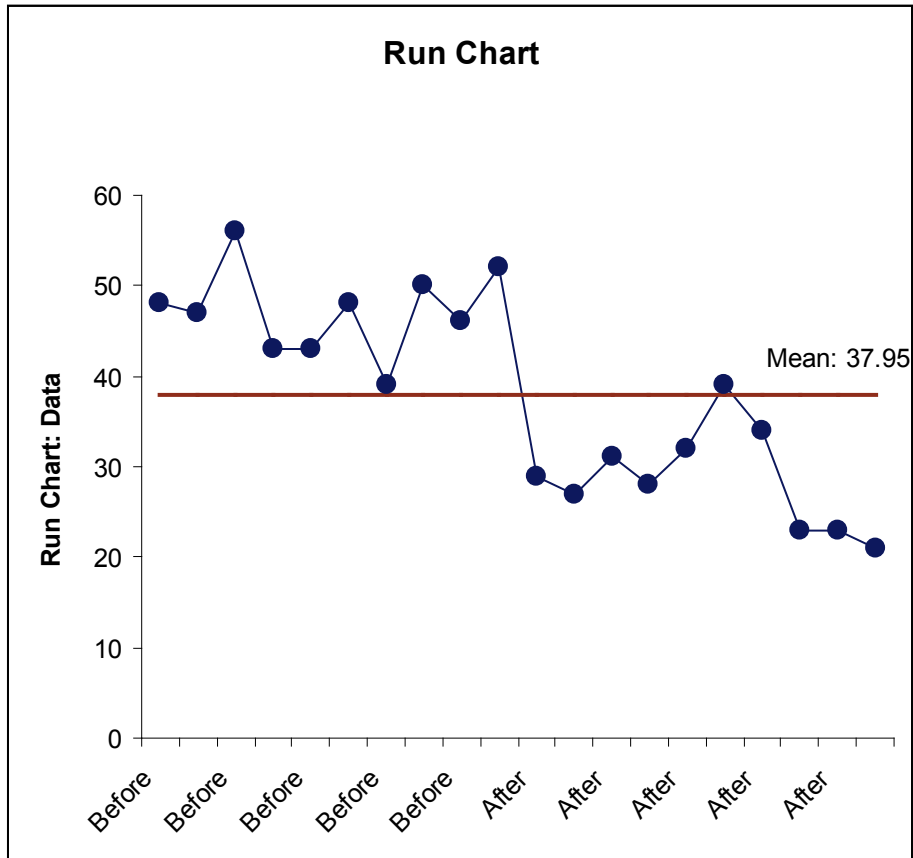
Basic Graphical Templates – Run Chart Example

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 ($1 - \text{Beta}$) 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 ($1 - \text{Beta}$) 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% ($\alpha = 0.1$), 95% ($\alpha = .05$) or 99% ($\alpha = .01$) 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 - Skew² 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 Rhie 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% ($\alpha=0.1$); 2=95% ($\alpha=0.05$); 3=99% ($\alpha=0.01$)
1	1	Skew ²
-0.48	-1.48	Kurt_delta = Kurt - Skew ² . Minimum = -1.48.

The model predictor terms are Skewness² 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 -; $\alpha=0.1$	1 Sample t; Ha = LT +, GT -; $\alpha=0.05$	1 Sample t; Ha = LT +, GT -; $\alpha=0.01$	1 Sample t; Ha = NE; $\alpha=0.1$	1 Sample t; Ha = NE; $\alpha=0.05$

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%)	$100*(1-\alpha)\%$	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) \neq 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 H_a : Mean (μ) \neq 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)	$H_0: \text{Mean } (\mu) =$	0.5
Alternative Hypothesis	$H_a: \text{Mean } (\mu)$	Not Equal To
Confidence Level (enter .95 for 95%)	$100*(1-\alpha)\%$	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 $H_0: \text{Mean } (\mu) = 0.5$		Reject
Alternative Hypothesis $H_a: \text{Mean } (\mu) \neq 0.5$		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: $\text{Mean } (\mu) = 0.5$ and conclude that the alternative hypothesis $H_a: \text{Mean } (\mu) \neq 0.5$ 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 : μ ₁ - μ ₂ =	0.0	
Alternative Hypothesis	H _a : μ ₁ - μ 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 (μ ₁ - μ ₂) = 0		Reject	
Alternative Hypothesis H _a : Mean Difference (μ ₁ - μ ₂) ≠ 0		Conclude true at 95.0% confidence level	
Interpretation of P-Value and Confidence Intervals			
<p>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 H_a: Mean Difference (μ₁ - μ₂) ≠ 0 is true at the confidence level of 95.0%.</p> <p>We are 95% confident that the true population mean difference lies within the interval (-1.9396 to -0.0604).</p>			

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.

SigmaXL 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 \geq UEL; Null hypothesis for P2: Difference \leq LEL. Both null hypothesis must be rejected to conclude that equivalence is true.
3. LEL and UEL establish the zone or region of equivalence and are determined by what size difference is considered practically significant.
4. 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		-0.0684	
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 \geq UEL; Null hypothesis for P2: Mean Difference \leq LEL. Both null hypothesis must be rejected to conclude that equivalence is true.
4. LEL and UEL establish the zone or region of equivalence and are determined by what size mean difference is considered practically significant.
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 – 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 ($\mu_1 - \mu_2$) = 0		Reject
Alternative Hypothesis H _a : Mean Difference ($\mu_1 - \mu_2$) \neq 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 H _a : Sigma (σ) \neq 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.

Sigma XL			
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)	$H_0: \sigma_1/\sigma_2 =$	1.0	
Alternative Hypothesis	$H_a: \sigma_1/\sigma_2$	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 $H_0: \sigma_1/\sigma_2 = 1$	Fail to Reject
Alternative Hypothesis $H_a: \sigma_1/\sigma_2 \neq 1$	


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 $H_0: \sigma_1/\sigma_2 = 1$.
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%)	$100*(1-\alpha)\%$	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) \neq 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) \neq 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 $np \geq 5$ and $nq \geq 5$, 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%)	$100 \cdot (1 - \alpha)\%$	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 $H_0: P_1 - P_2 = 0$.			
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 $H_0: P_1 - P_2 = 0$.
- Use Normal Approximation only when minimum expected cell value, shown above, is ≥ 5 . Unpooled is used if $H_0: P_1 - P_2 \neq 0$. Use pooled if $H_0: P_1 - P_2 = 0$. (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.

Sigma
XL

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:

- 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.
- 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 \geq UEL; Null hypothesis for P2: Proportion Difference \leq LEL. Both null hypothesis must be rejected to conclude that equivalence is true. The P-Values are based on the normal approximation unpooled method.
 4. LEL and UEL 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. 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. 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%)	$100 \cdot (1 - \alpha)\%$	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) \neq 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)	$H_0: \text{Rate}_1 - \text{Rate}_2 (\lambda_1 - \lambda_2) =$	0	
Alternative Hypothesis	$H_a: \text{Rate}_1 - \text{Rate}_2 (\lambda_1 - \lambda_2)$	Not Equal To	
Confidence Level (enter .95 for 95%)	$100 \cdot (1 - \alpha)\%$	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 $H_0: \lambda_1 - \lambda_2 = 0$.
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 $H_0: \lambda_1 - \lambda_2 = 0$.
- Use Normal Approximation only when the number of occurrences for each sample is greater than 10. Unpooled is used if $H_0: \lambda_1 - \lambda_2 \neq 0$. Use pooled if $H_0: \lambda_1 - \lambda_2 = 0$.

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.

Sigma
XL

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 \geq UEL; Null hypothesis for P2: Rate Difference \leq LEL. Both null hypothesis 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. 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 – 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.

Sample Data (user inputs):			Results:	
Category ID	Observed Count	Specify Historical Count (optional)	Percentage of cells with expected count < 5	40.0%
1	1	1	Sum of Observed Counts	26
2	2	2	Chi-Square	3.4231
3	3	3	df	4
4	10	4	p-value	0.4897
5	10	5		

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			Results:		Show/Hide Calculations
Sample Data (user inputs):					
Category ID	Observed Count	Specify Historical Count (optional)	Percentage of cells with expected count < 5		
1	1	1	40.0%		Exact P-Value
2	2	2	Sum of Observed Counts		
3	3	3	Chi-Square		
4	10	4	DF		
5	10	5	Chi-Square P-Value		
6			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%.

Monte-Carlo P-Value

Number of Replications: 10000

Confidence Level for P-Value: 99 %

OK >> Cancel Help

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	

Show/Hide Calculations

Exact P-Value

Monte Carlo P-Value

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

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

Exact P-Value

Time Limit for Exact Computation:

60

(Seconds)

OK >>

Cancel

Help

- 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

Select the Probability of Interest from the following probabilities:				
Areas:				
between X1 and X2		<table><tr><td>Prob($1 \leq X \leq 2$)</td><td>0.135905122</td></tr></table>	Prob($1 \leq X \leq 2$)	0.135905122
Prob($1 \leq X \leq 2$)	0.135905122			
outside X1 and X2		<table><tr><td>Prob($X \leq 1$ AND $X \geq 2$)</td><td>0.864094878</td></tr></table>	Prob($X \leq 1$ AND $X \geq 2$)	0.864094878
Prob($X \leq 1$ AND $X \geq 2$)	0.864094878			
below X1		<table><tr><td>Prob($X \leq 1$)</td><td>0.841344746</td></tr></table>	Prob($X \leq 1$)	0.841344746
Prob($X \leq 1$)	0.841344746			
above X1		<table><tr><td>Prob($X \geq 1$)</td><td>0.158655254</td></tr></table>	Prob($X \geq 1$)	0.158655254
Prob($X \geq 1$)	0.158655254			
below X2		<table><tr><td>Prob($X \leq 2$)</td><td>0.977249868</td></tr></table>	Prob($X \leq 2$)	0.977249868
Prob($X \leq 2$)	0.977249868			
above X2		<table><tr><td>Prob($X \geq 2$)</td><td>0.022750132</td></tr></table>	Prob($X \geq 2$)	0.022750132
Prob($X \geq 2$)	0.022750132			

Basic MSA Templates – Type 1 Gage 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 > Type 1 Gage Study**.

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

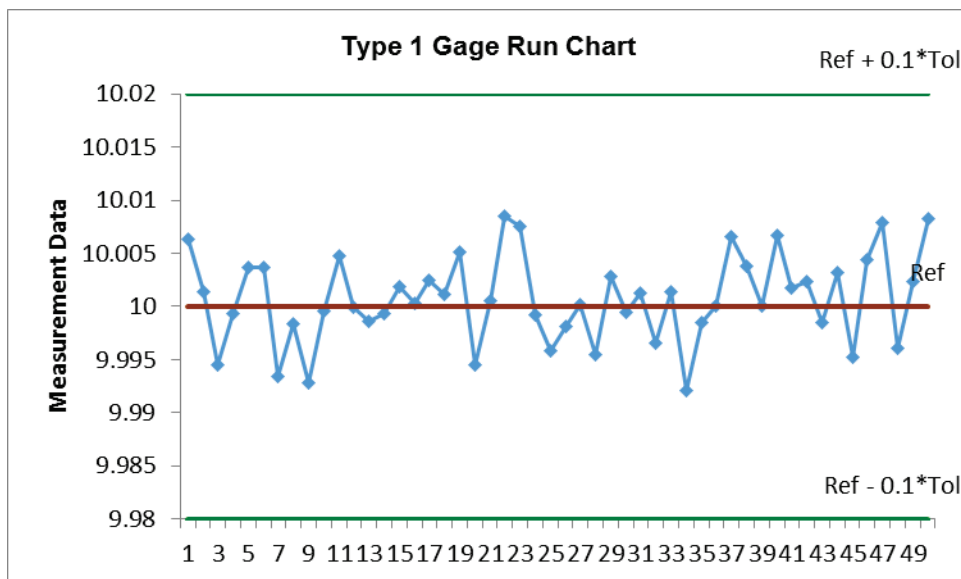
Tolerance (Tol, USL - LSL):	0.2
Part Reference Value (Ref):	10
Gage Resolution:	0.0001
StDev Multiplier:	6.00
Percent of Tolerance for Cg/Cgk:	20

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 $\leq 5\%$, 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 = $\text{Ref} + 0.1 * \text{Tol}$ and lower limit = $\text{Ref} - 0.1 * \text{Tol}$ (assuming 20% of Tolerance).
11. This calculator requires Excel 2010 or higher.

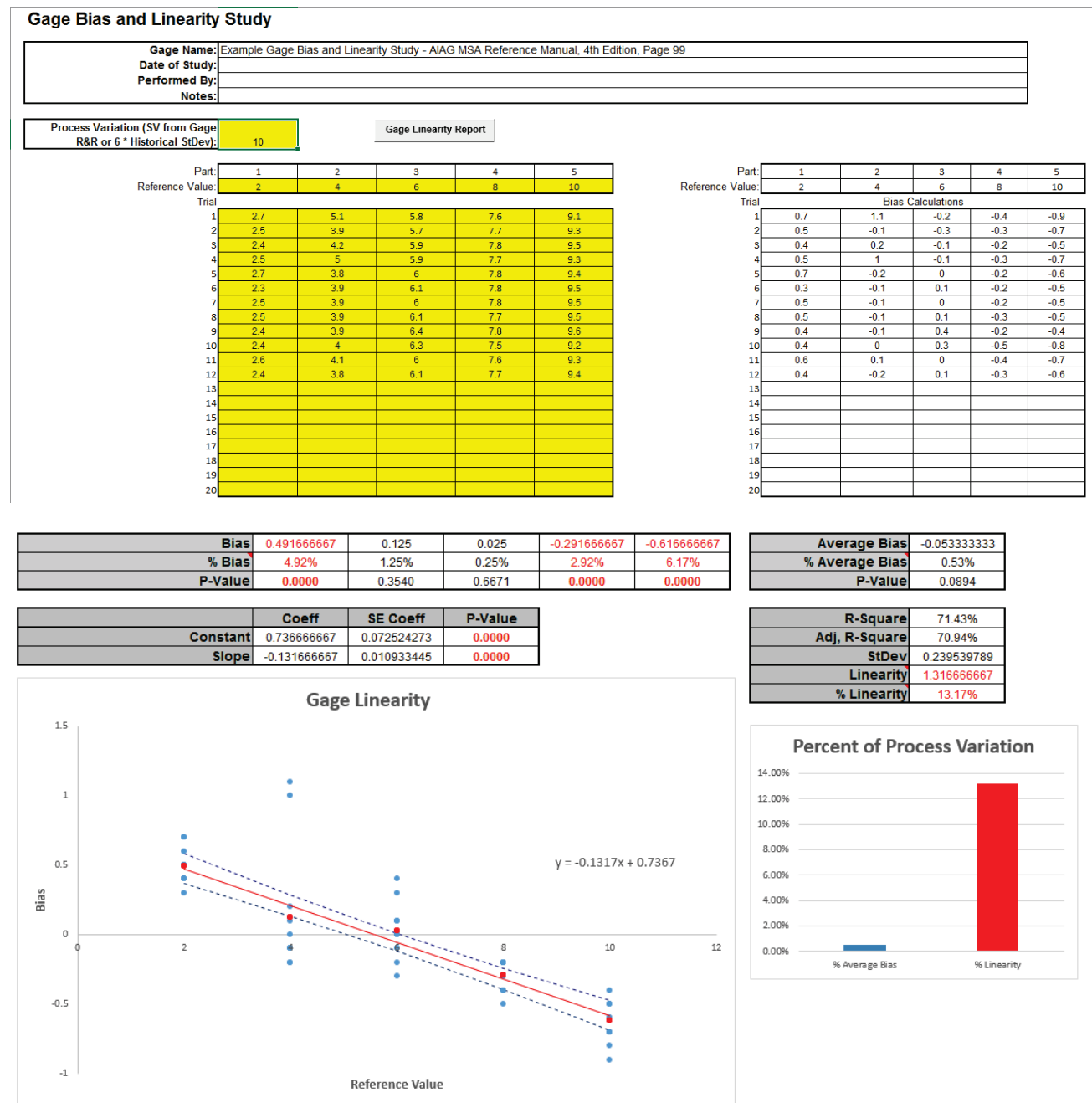
REFERENCES

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**.



Notes:

1. Enter **Process Variation** (SV from Gage R&R or $6 * \text{Historical StDev}$). 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

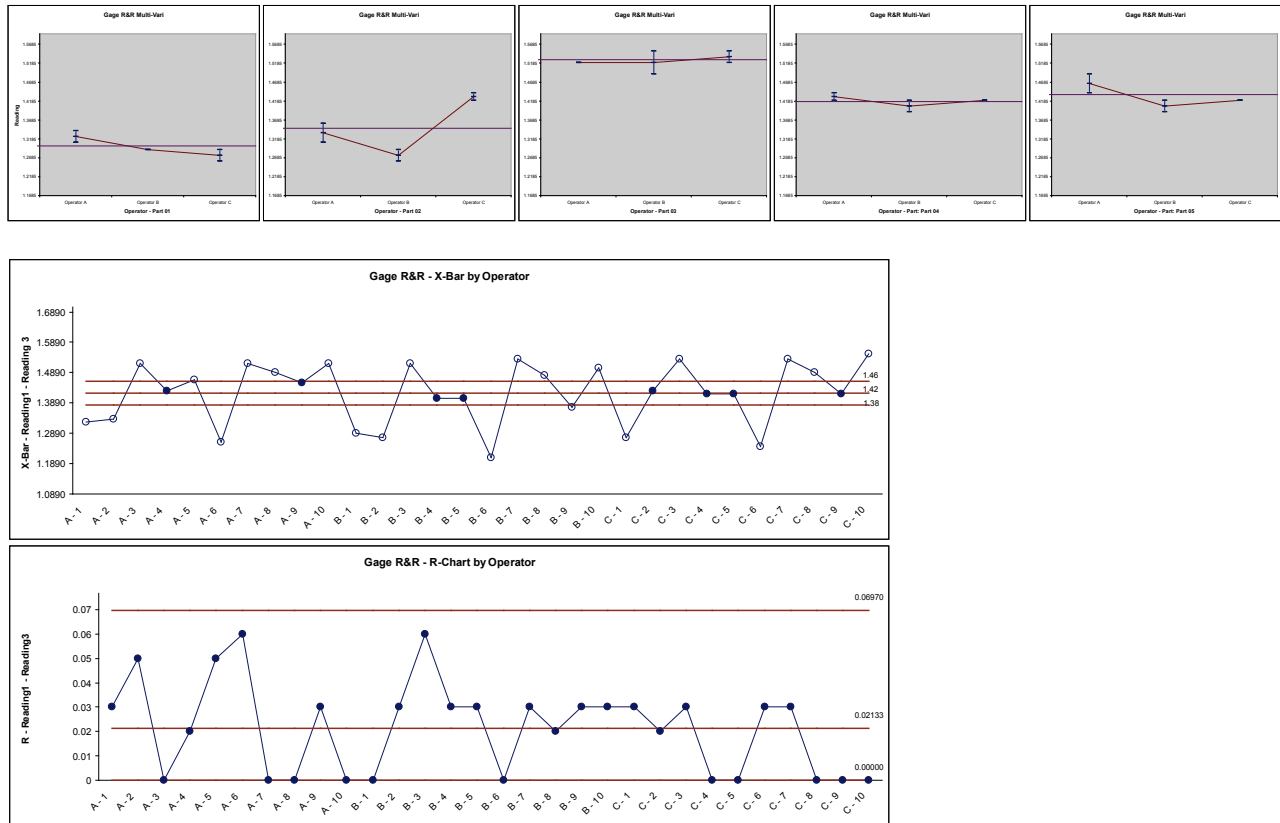
Notes for use of the Gage R&R Template:

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
1	G
2	G
3	NG
4	G
5	NG
6	NG
7	G
8	NG
9	G
10	NG
11	G
12	NG
13	NG
14	G
15	NG
16	G
17	G
18	NG
19	NG
20	G

Appraiser A		Appraiser B		Appraiser C	
Trial # 1	Trial #2	Trial # 1	Trial #2	Trial # 1	Trial #2
G	G	NG	G	G	G
NG	G	G	G	G	G
NG	NG	NG	NG	NG	NG
G	G	G	NG	G	G
NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
G	G	NG	G	G	G
NG	NG	NG	NG	NG	NG
G	G	G	G	G	G
NG	NG	NG	NG	NG	NG
G	G	NG	G	G	G
NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
NG	G	NG	G	G	G
NG	NG	NG	G	NG	NG
G	NG	G	G	G	NG
G	G	G	G	G	G
NG	NG	NG	G	NG	NG
NG	NG	NG	NG	NG	NG
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 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 for long term data)		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 for long term data)		1.5
Results:		
Expected dpm > USL		1349.9
Expected % > USL		0.13%
Expected dpm < LSL		1349.9
Expected % < LSL		0.13%
Expected dpm (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 for long term data)	1.5				
Process Sigma Level Results:		Process Capability Results:			
Expected dpm > USL	22,750.1	Cp, Pp	1.00		
Expected % > USL	2.28%	Lower Limit Cp, Pp	0.74		
Expected dpm < LSL	31.7	Upper Limit Cp, Pp	1.26		
Expected % < LSL	0.00%	Cpu, Ppu	0.67		
Expected dpm (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	0.87		

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 > Basic Control Chart Templates > Individuals Chart**.

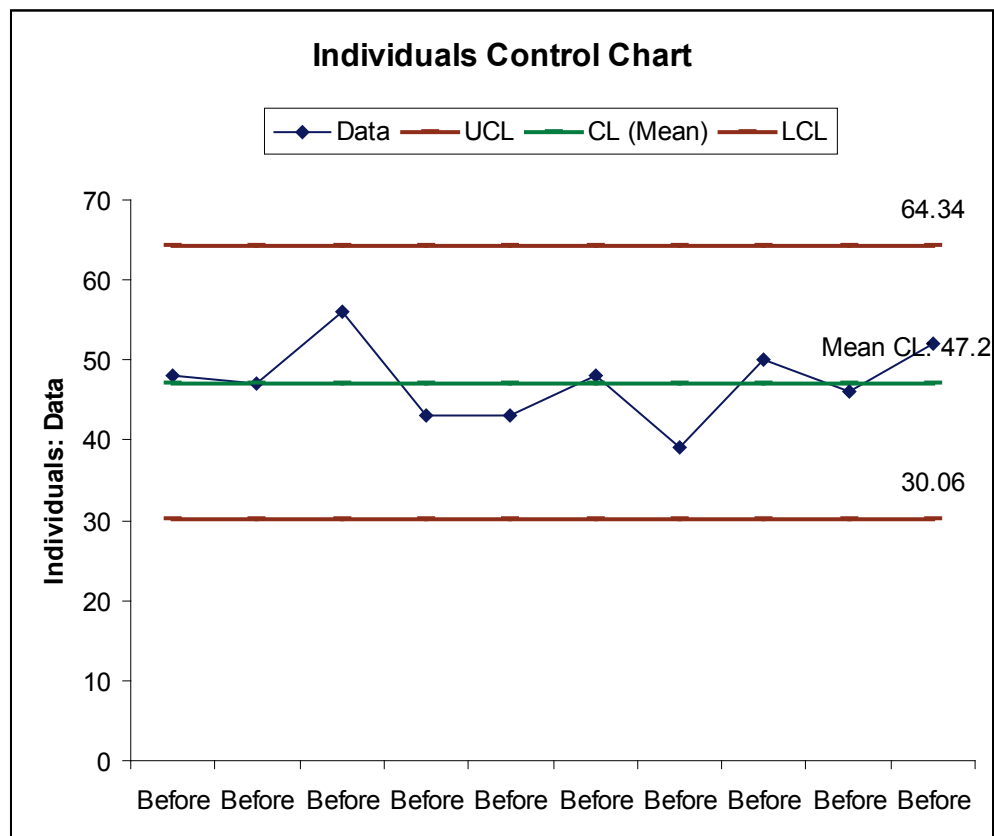
Enter the data as shown (or copy and paste the “Before” data from the histogram data above):

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

Individuals Control Chart

Add Data

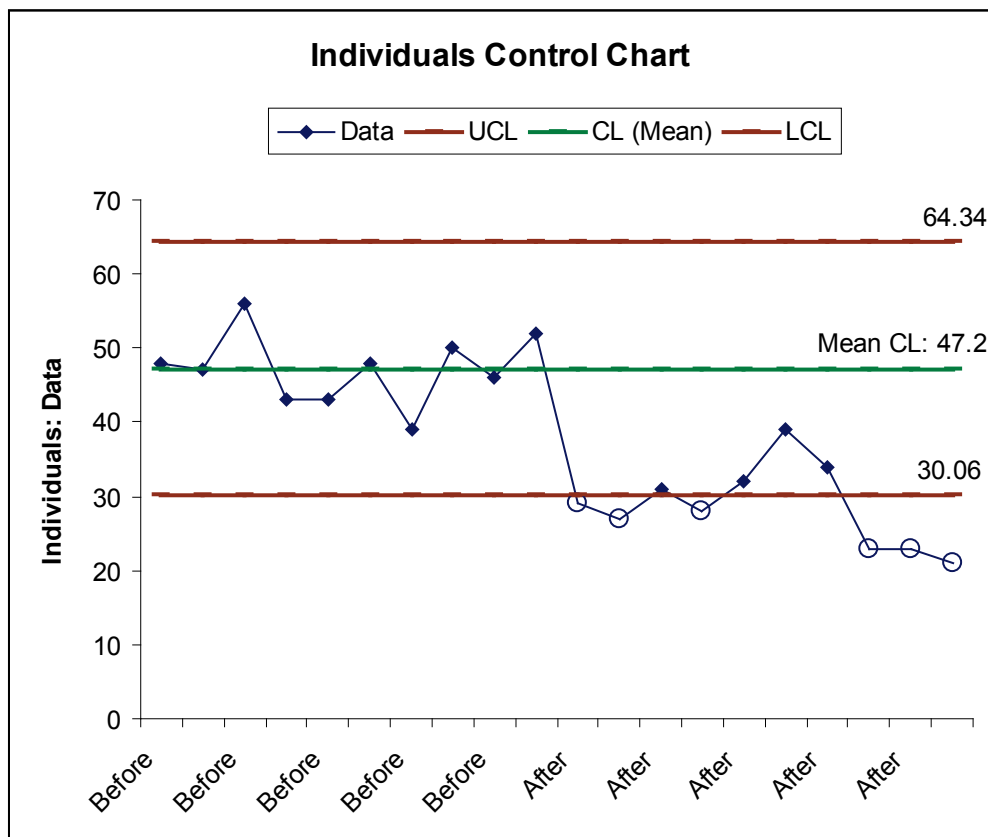
Click the Individuals Control Chart button to produce the Individuals Chart (note that in actual application, one should use a minimum of 20 to 30 data points to compute the initial control limits).



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.



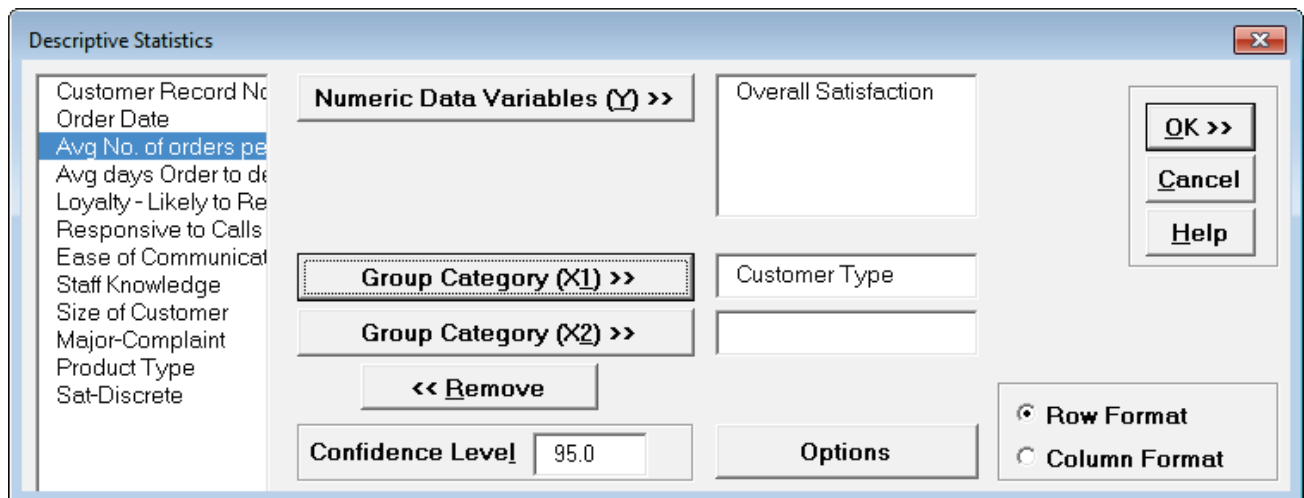
Notes:

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:



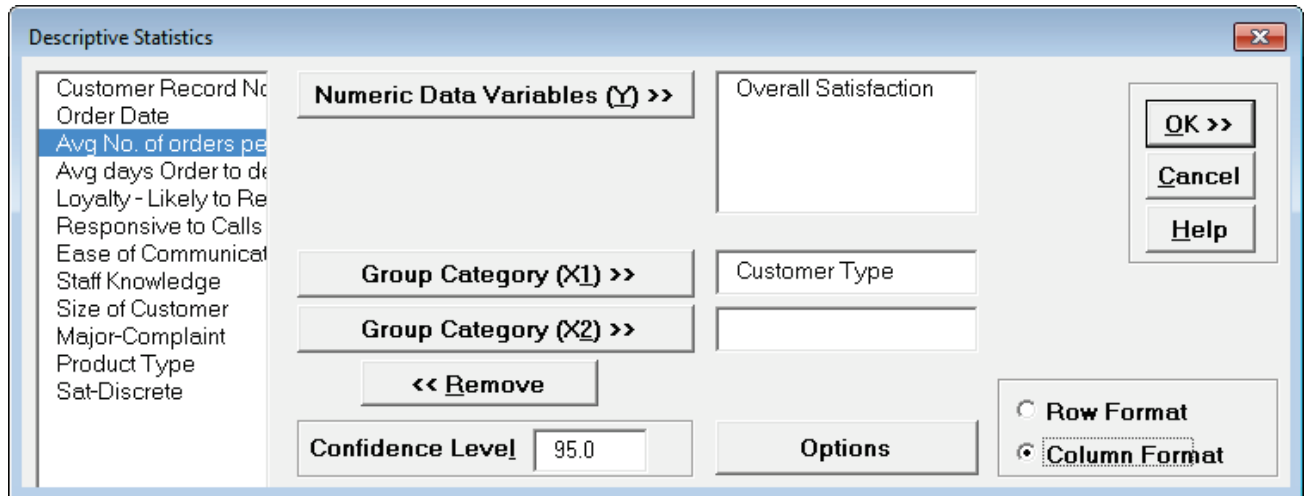
5. Click **OK**.
6. Descriptive Statistics are given for Customer Satisfaction grouped by Customer Type:

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.

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



- 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)
Customer Type = 1	31	3.394	0.824886	3.088	1.720	2.810	3.560	4.020	4.800	3.091 to 3.698	0.659012 to 1.102328	0.312776	0.5308
Customer Type = 2	42	4.205	0.621206	2.568	2.420	3.828	4.340	4.725	4.980	4.012 to 4.398	0.511126 to 0.792132	0.826259	0.0380
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

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.

The screenshot shows the 'Descriptive Statistics' dialog box. On the left, a list of variables includes 'Customer Record No.', 'Order Date', 'Avg No. of orders per', 'Avg days Order to de', 'Loyalty - Likely to Re', 'Responsive to Calls', 'Ease of Communicat', 'Staff Knowledge', 'Size of Customer', 'Major-Complaint', 'Product Type', and 'Sat-Discrete'. The 'Avg No. of orders per' variable is selected. The 'Numeric Data Variables (Y) >>' button is active. The 'Group Category (X1) >>' button is also active, and 'Customer Type' is entered in the adjacent field. The 'Group Category (X2) >>' button is disabled. The 'Confidence Level' is set to 95.0. The 'Options' button is active. The 'Row Format' radio button is selected, and the 'Column Format' radio button is disabled. The 'Select All' checkbox is checked. The 'Percentile Confidence Intervals' section has 'Interpolated' and 'Exact' radio buttons disabled, and 'Percentile' is selected. The 'Percentile Tolerance Intervals' section has 'Interpolated' and 'Exact' radio buttons disabled.

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 ($100 * \text{StDev}/\text{Mean}$)
- Short Term StDev ($\text{MR-bar}/d2$)

4. The **Additional Normality Tests** are:

- Shapiro-Wilk ($n \leq 5000$) 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	2.441	2.19
2.5	1.72	2.42	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	2.55925	2.74
99 - 1 (98%)	3.08	2.539	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.
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.
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:

Descriptive Statistics

Customer Record No.
Order Date
Avg No. of orders per month
Avg days Order to delivery
Loyalty - Likely to Repeat
Responsive to Calls
Ease of Communication
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

Numeric Data Variables (Y) >>
Overall Satisfaction

Group Category (X1) >>
Customer Type

Group Category (X2) >>

<< Remove

Confidence Level 95.0

Options

☐ Select All

☐ Additional Descriptive Statistics

☐ Additional Normality Tests

☐ Outlier and Randomness Tests

☐ Percentile Report

☒ Percentile Confidence Intervals

☐ Interpolated
☒ Exact

☐ Quartile
☒ Percentile

☒ Percentile Tolerance Intervals

☐ Interpolated
☒ Exact

☒ Row Format
☐ Column Format

OK >>
Cancel
Help

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.

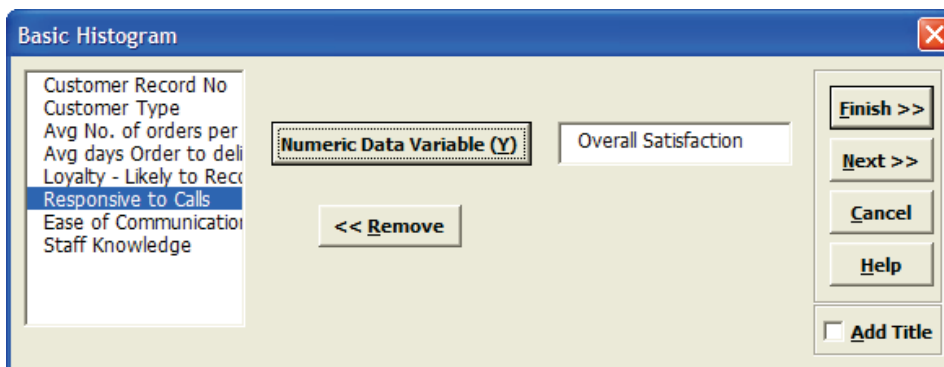
Part D – Histograms

Basic Histogram Template

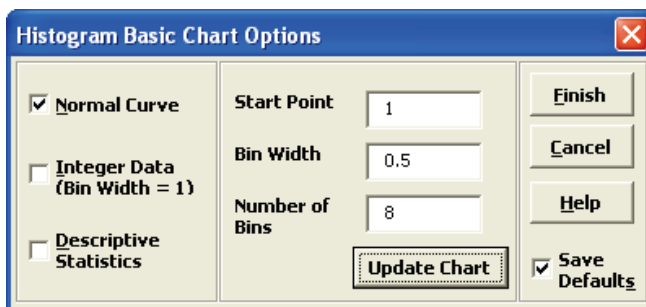
Click **SigmaXL > Templates and Calculators > Basic Graphical Templates > Histogram** or **SigmaXL > Graphical Tools > Basic Graphical Templates > Histogram**. See **Part B – Templates and Calculators** for a Histogram Template example.

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:

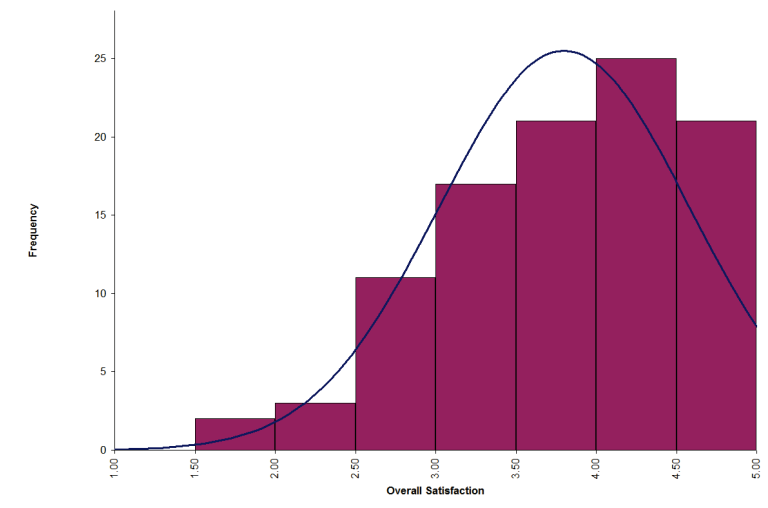


5. Click **Next**. Ensure that **Normal Curve** is checked. Set **Start Point** = 1. Change the **Bin Width** to 0.5, and the **Number of Bins** to 8. Click **Update Chart** to view the histogram. (If the survey satisfaction data was pure integer format we would have checked the **Integer Data** option).



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.

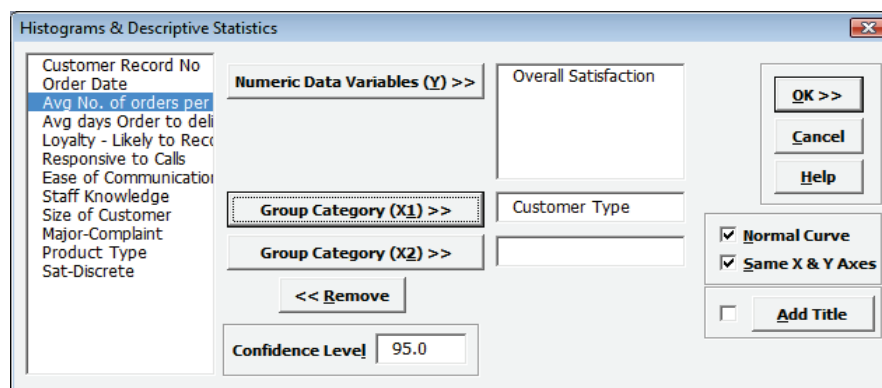


7. Note that bin one is 1 to < 1.5, bin 2 is 1.5 to < 2, etc.

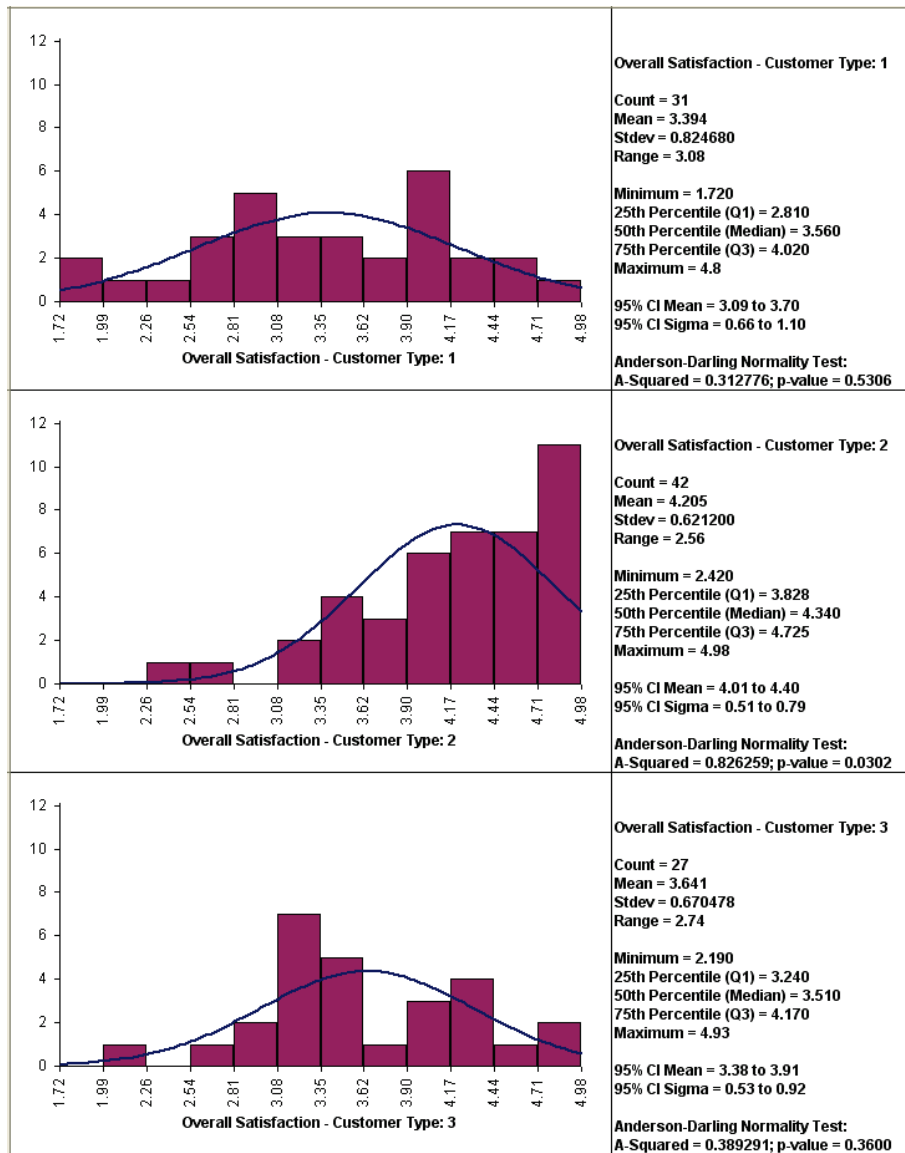
Tip: Any graph produced by SigmaXL can be Copied/Pasted into Word. It can also be enlarged by clicking on the graph and dragging the corner. The number of decimal places displayed can be modified by clicking on the Axis Label and selecting the **Number** tab to adjust. The text label alignment can also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



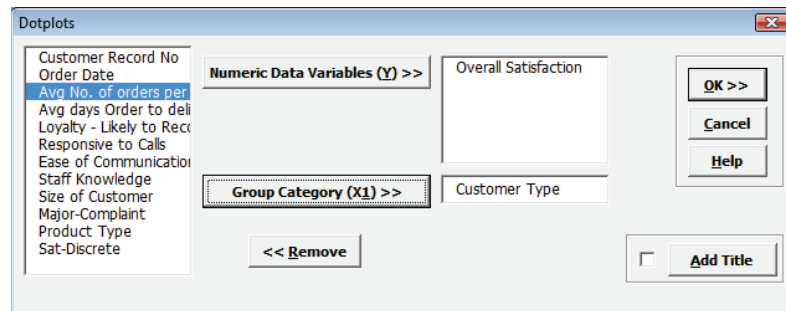
Clearly Customer Type 2 shows a higher level of overall satisfaction, with the data skewed left. Note that Customer Type 1 and 3 have data that is normally distributed, but this is not desirable when the response is a satisfaction score!

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.

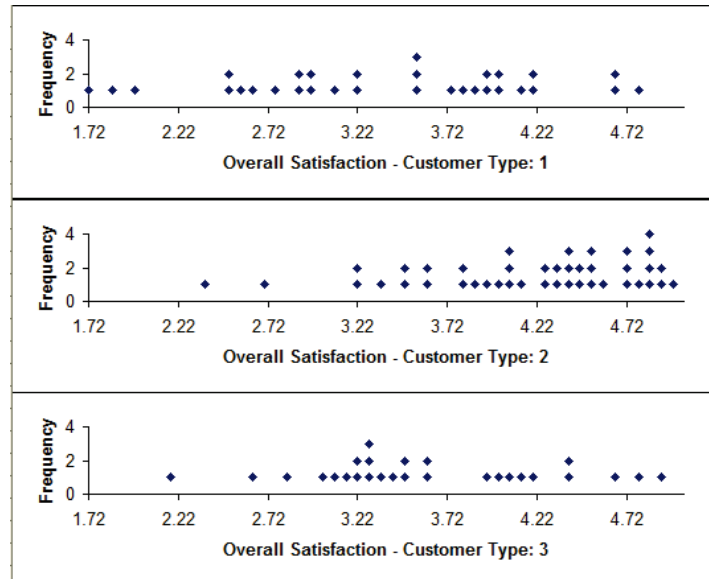
Part E – Dotplots

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:

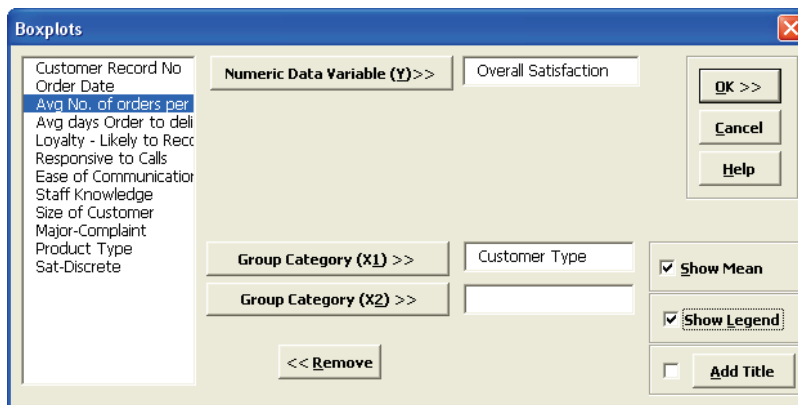


5. Since all data points are shown, dotplots are a useful alternative to Histograms, particularly when the group sample sizes are small ($n < 30$).

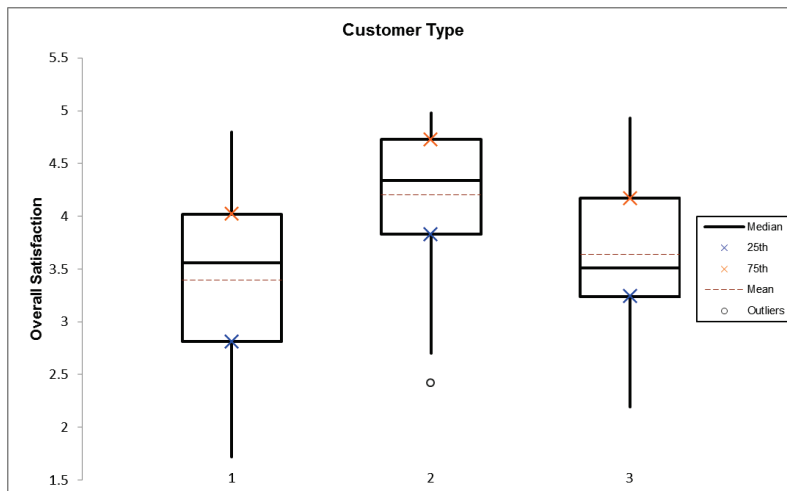
Part F – Boxplots

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**:

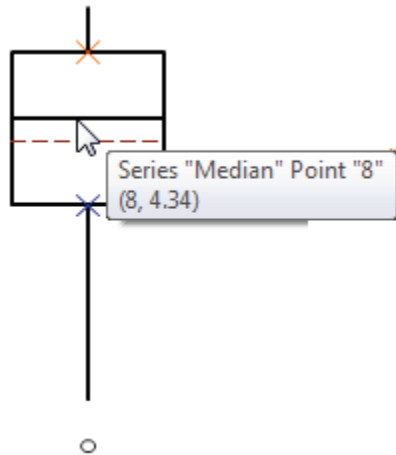


5. Click **OK**. A boxplot of Customer Satisfaction By Customer Type is produced:



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.



- 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:

Format Axis ▼ ×

AXIS OPTIONS ▼ | TEXT OPTIONS

AXIS OPTIONS

Bounds

Minimum Reset

Maximum Reset

Units

Major Reset

Minor Reset

Horizontal axis crosses

☐ Automatic

☒ Axis value

☐ Maximum axis value

Display units None ▼

☐ Show display units label on chart

☐ Logarithmic scale Base

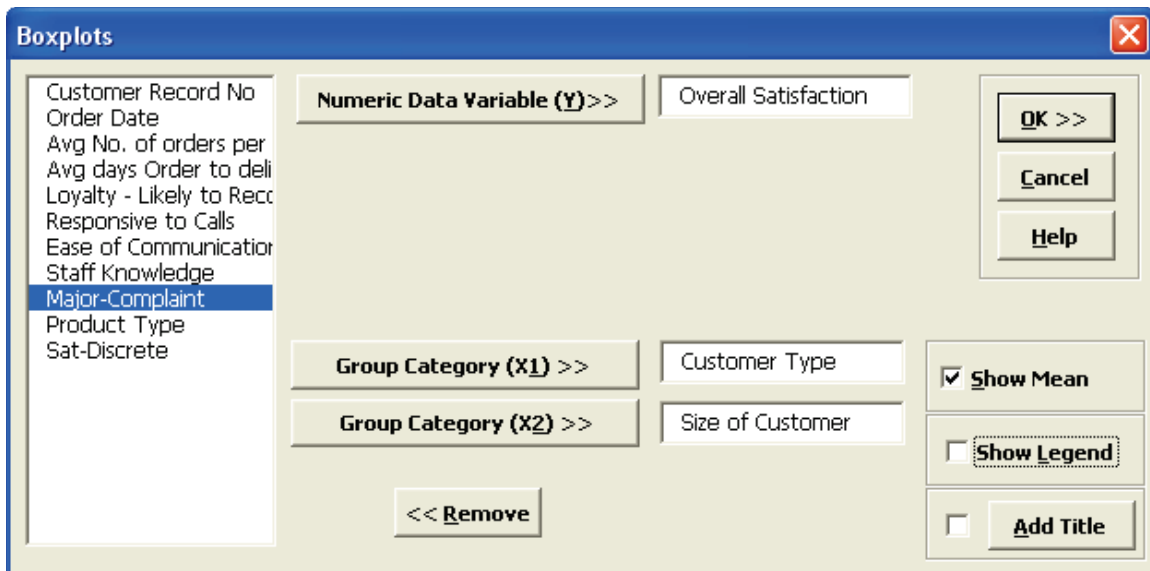
☐ Values in reverse order

8. The Boxplot axis is modified as shown below:

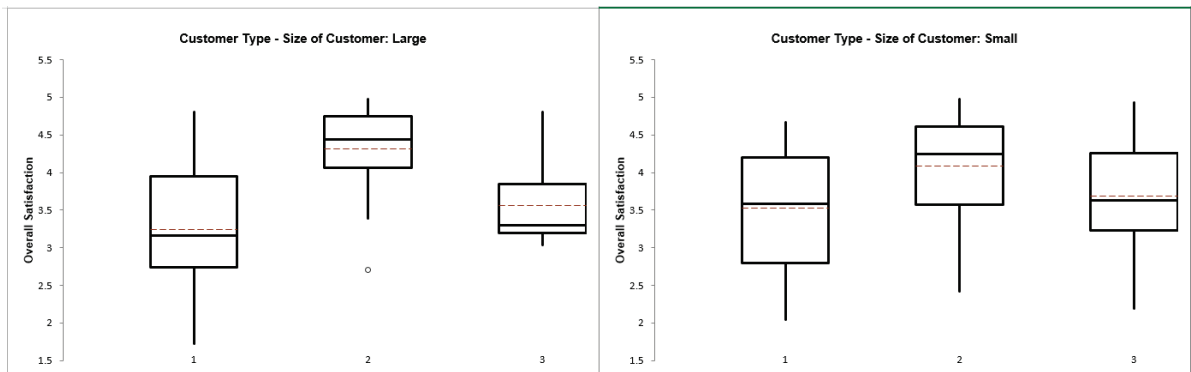


9. Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog.

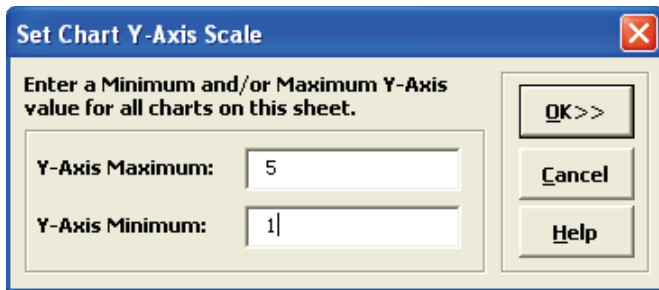
10. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**; select *Customer Type*, click **Group Category (X1) >>**; select *Size of Customer*, click **Group Category (X2) >>**; check **Show Mean**; uncheck **Show Legend**:



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**.

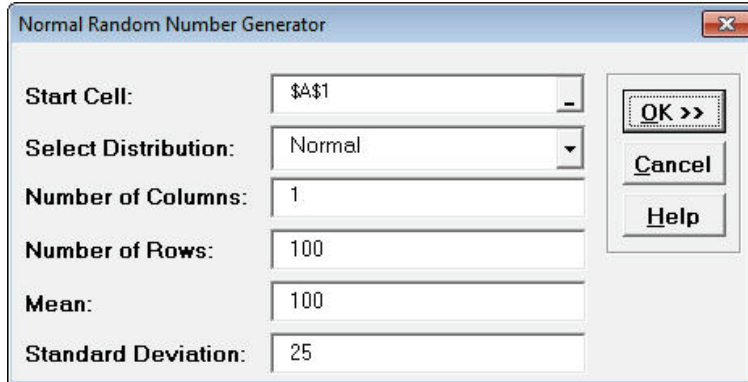


13. Click **OK**. The Y-axis scale maximum and minimum are now modified for both charts.

Part G – Normal Probability Plots

Normal Probability Plots

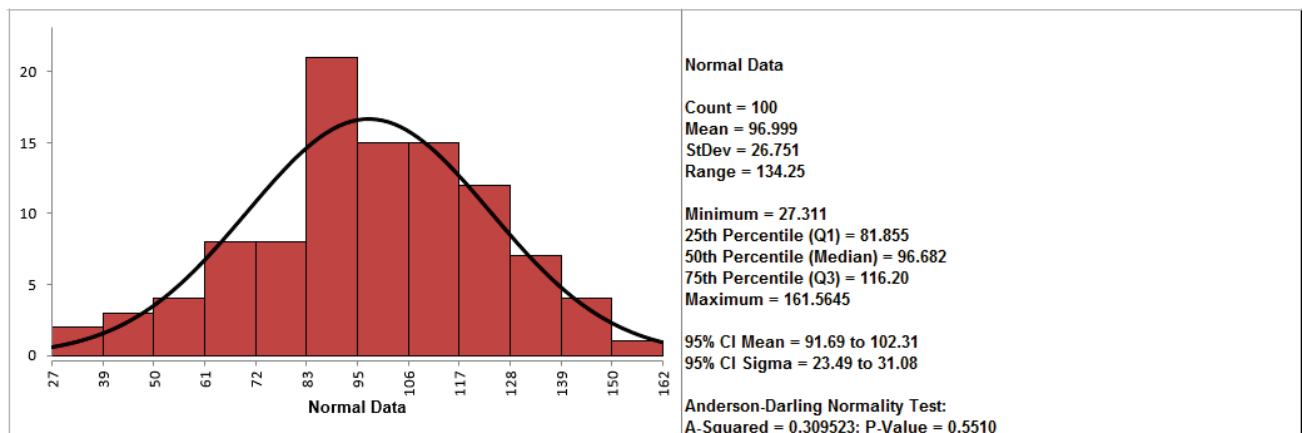
1. Create 100 random normal values as follows: Click **SigmaXL > Data Manipulation > Random Data > Normal**. Specify 1 Column, 100 Rows, Mean of 100 and Standard Deviation of 25 as shown below:



The dialog box titled "Normal Random Number Generator" contains the following fields and buttons:

- Start Cell:** \$A\$1
- Select Distribution:** Normal
- Number of Columns:** 1
- Number of Rows:** 100
- Mean:** 100
- Standard Deviation:** 25
- Buttons:** OK >>, Cancel, Help

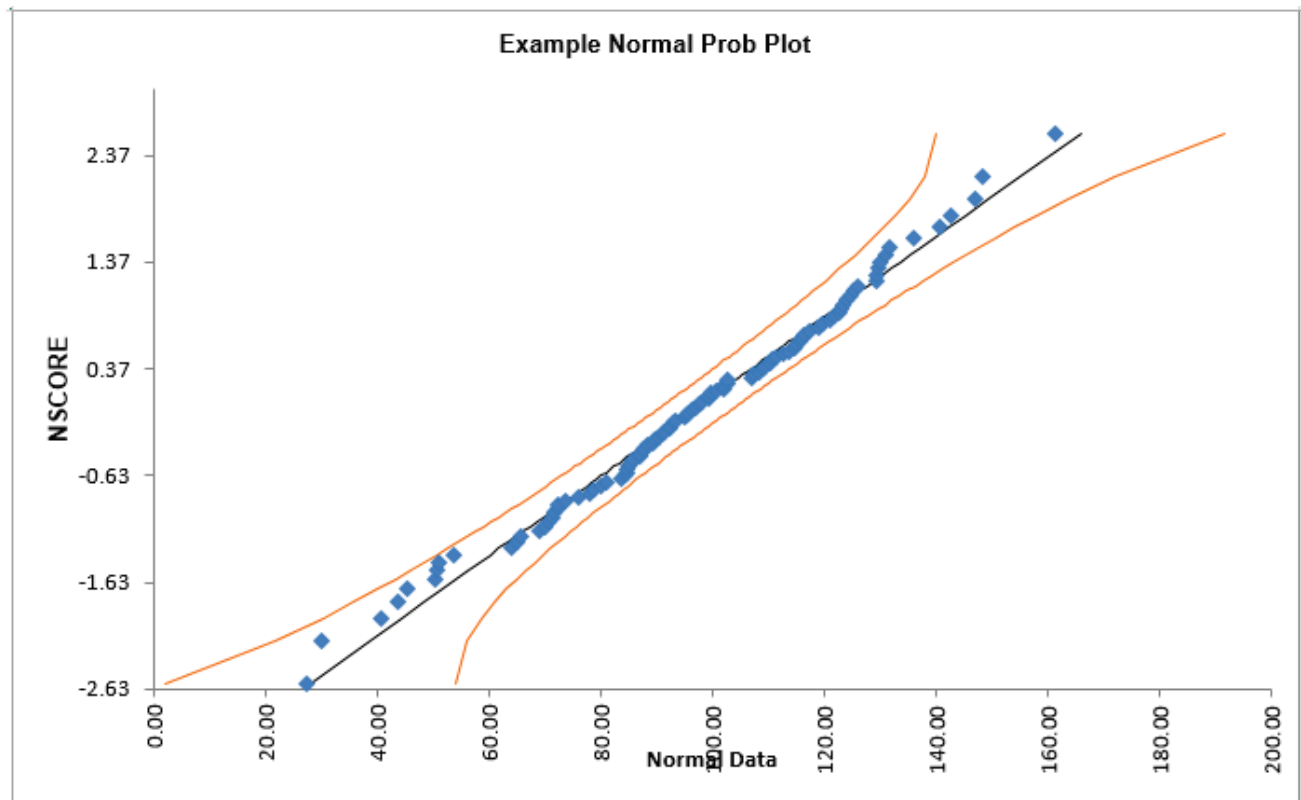
2. Click **OK**. Change Column heading to *Normal Data*.
3. Create a Histogram & Descriptive Statistics for this data. Your data will be slightly different due to the random number generation:



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**.

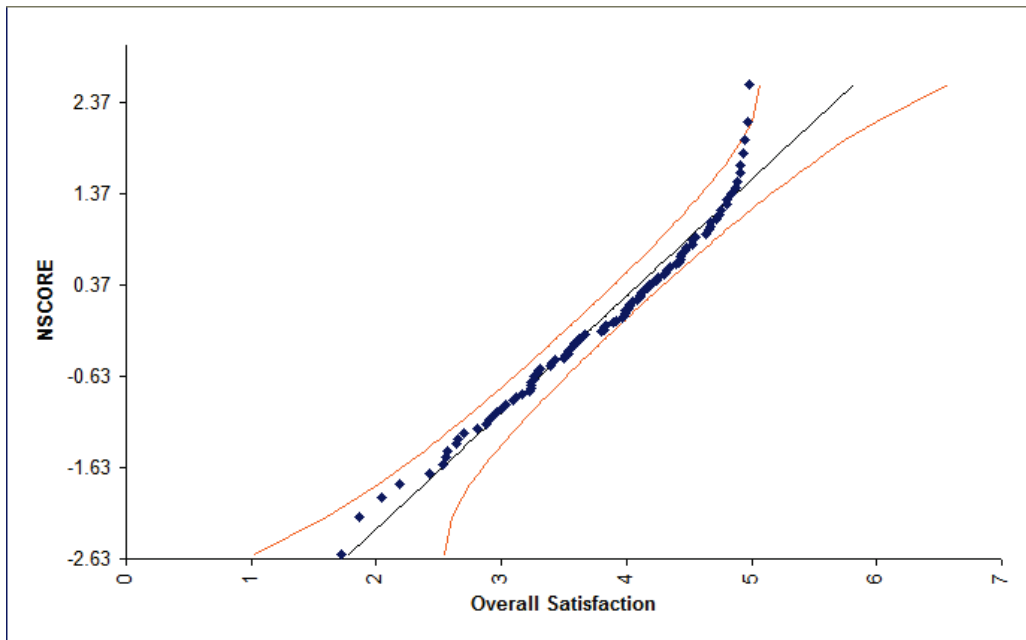
- Click **OK**. A Normal Probability Plot of simulated random data is produced (again, your plot will be slightly different due to the random number generation):



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).

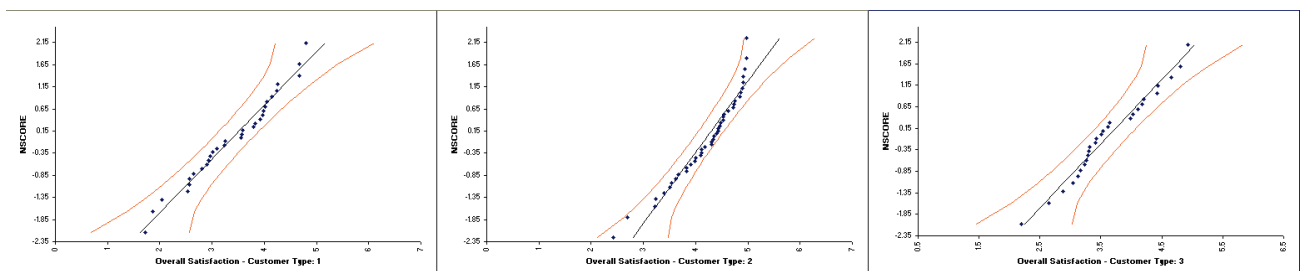
- Click **Sheet 1** Tab of **Customer Data.xlsx**.
- Click **SigmaXL > Graphical Tools > Normal Probability Plots**.
- Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.

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 $\text{LN}(Y)$ or $\text{SQRT}(Y)$ 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.

Part H– Run Charts

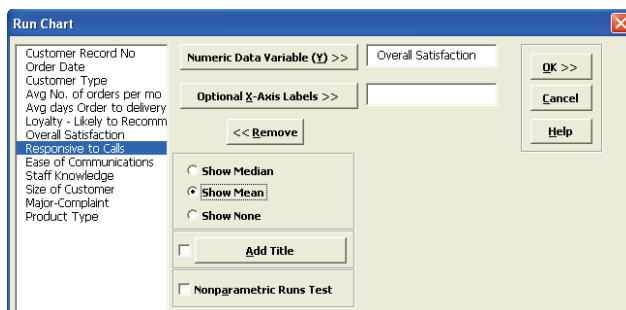
Run charts, also known as trend charts and time series plots, add the dimension of time to the graphical tools. They allow us to see trends and outliers in the data. Run Charts are a precursor to control charts, which add calculated control limits. Note that Run Charts should be used only on unsorted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

Basic Run Chart Template

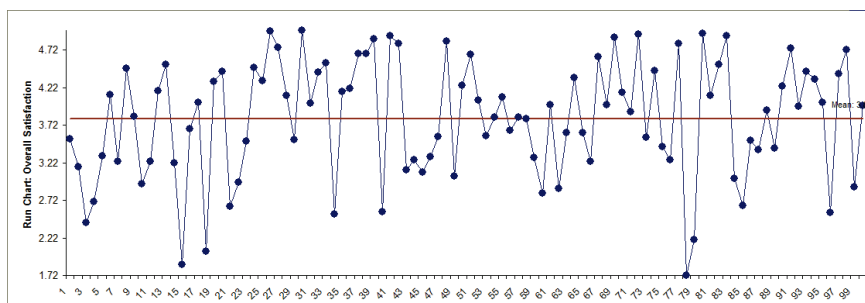
Click **SigmaXL > Templates and Calculators > Basic Graphical Templates > Run Chart** or **SigmaXL > Graphical Tools > Basic Graphical Templates > Run Chart**. See **Part B – Templates and Calculators** for a Run Chart Template example.

Run Charts

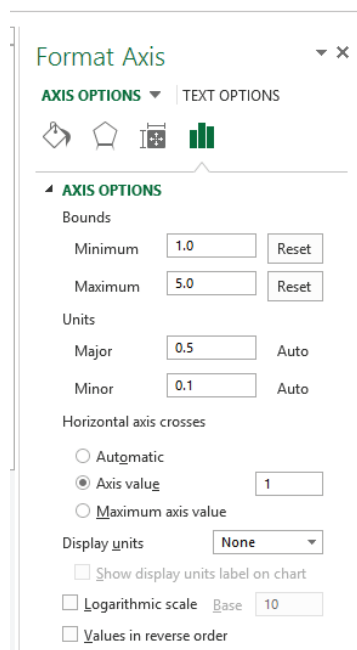
1. Click **Sheet 1** Tab of **Customer Data.xlsx** (or press **F4** to activate last worksheet). Click **SigmaXL > Graphical Tools > Run Chart**. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
2. Select *Overall Satisfaction*, click **Numeric Data Variable (Y) >>**. Select **Show Mean**. Uncheck **Nonparametric Runs Test** (to be discussed later in Part N of Analyze Phase).



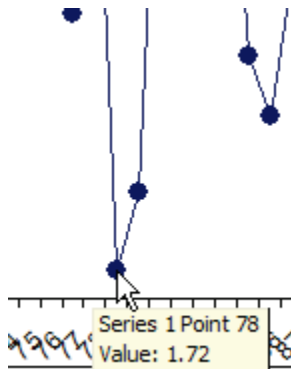
3. Click **OK**. A Run Chart of Overall Satisfaction with Mean center line is produced.



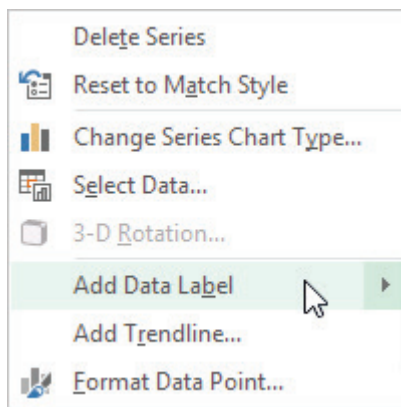
4. Select the Y axis, Right Click, **Format Axis**, to activate the **Format Axis** dialog. Change **Minimum** to 1, **Maximum** to 5, **Horizontal axis crosses > Axis value** to 1:



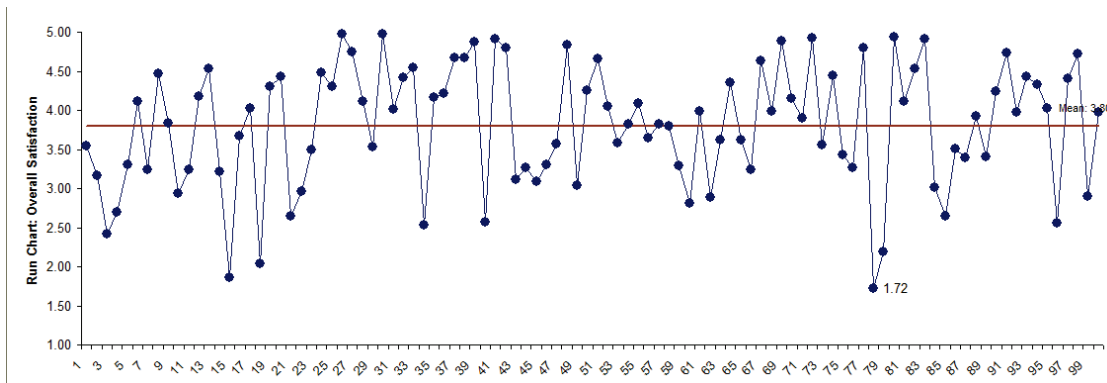
5. The chart is automatically updated.
6. Are there any obvious trends? Some possible cycling, but nothing clearly stands out. It may be interesting to look more closely at a specific data point. Any data point value can be identified by simply moving the cursor over it:



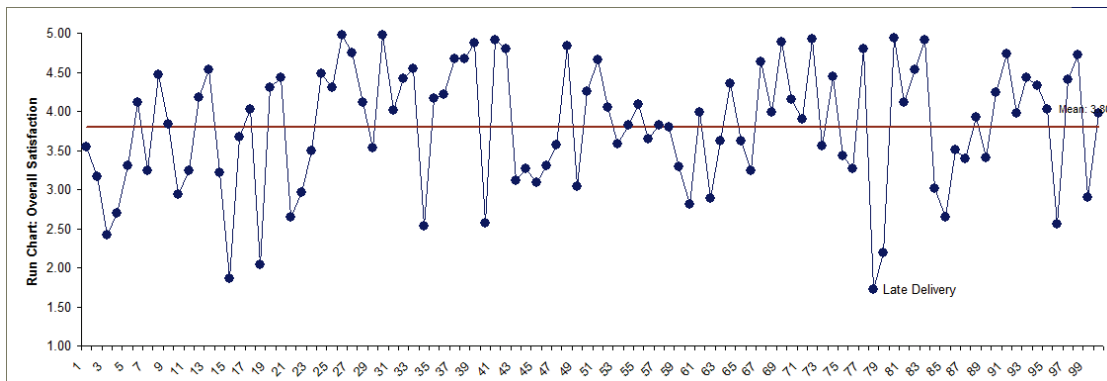
7. A label can be added to a data point by one single-click on the data point, followed by a Right mouse click, and select **Add Data Label**. See also **SigmaXL Chart Tools > Add Data Label** in Control Phase Tools, Part B - X-Bar & Range Charts.



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

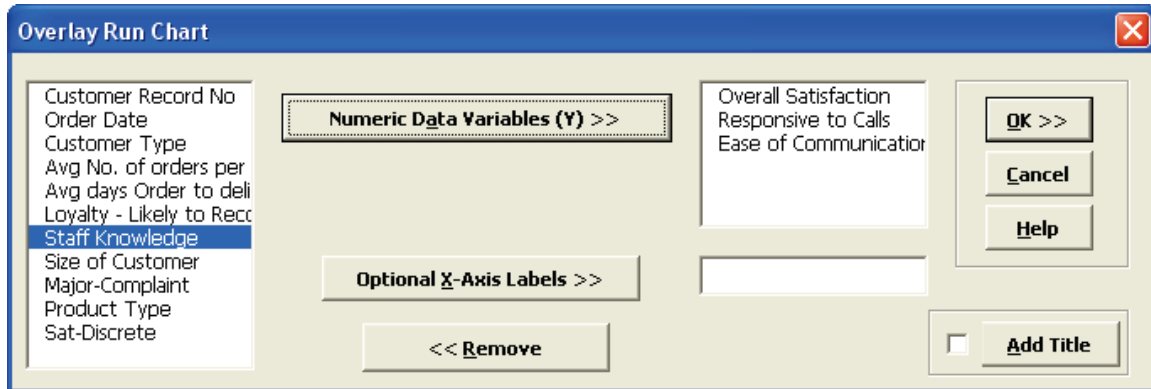


9. This label can be changed to a text comment. Single-click three times on the label and type in a comment as shown:

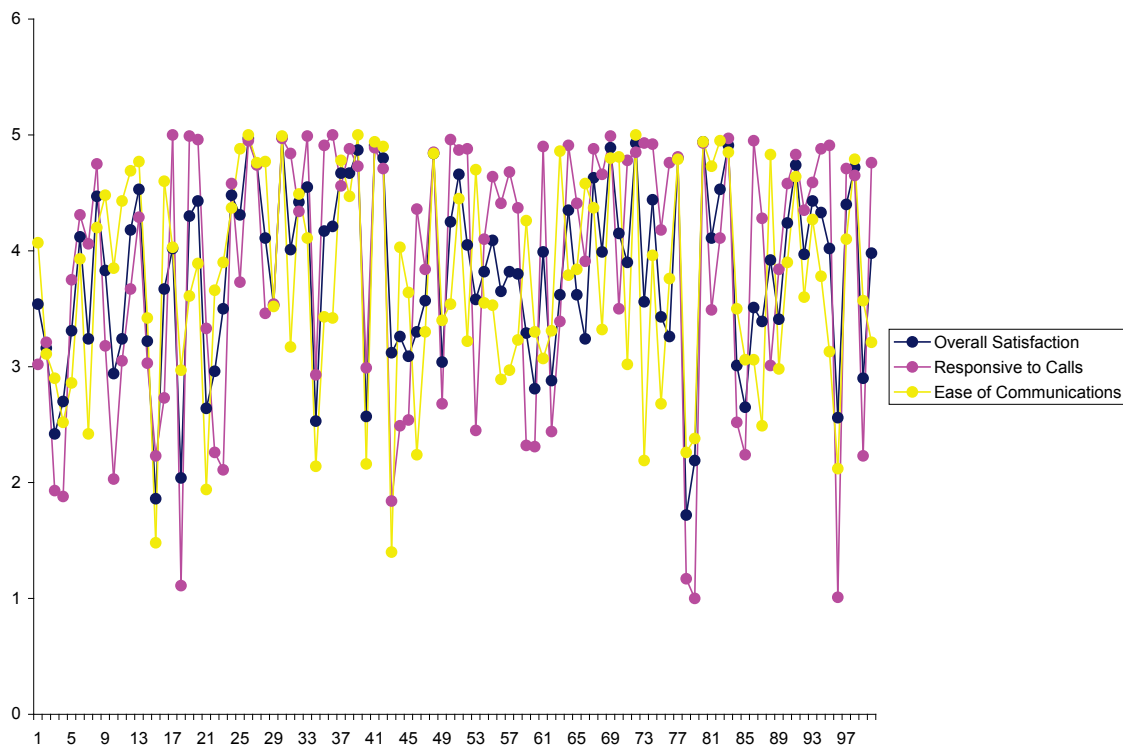


Overlay Run Charts

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.

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.

Create Gage R&R (Crossed) Worksheet

Number of Parts/Samples: 10

Number of Operators/Appraisers: 3

Number of Replicates/Trials: 3

☒ Randomize Parts/Samples

☒ Randomize Operators/Appraisers

Part/Sample Names:

1: Part 01

2: Part 02

3: Part 03

Operator/Appraiser Names:

1: Operator A

2: Operator B

3: Operator C

Buttons: OK>>, Cancel, Help, Reset

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**.

Analyze Gage R&R (Crossed)

Run Order
Std. Order

Part >> Part

Operator >> Operator

Measurement >> Measurement

<< Remove

OK>>
Cancel
Help

Standard Deviation Multiplier: 6

Alpha to Remove Interaction: 0.1

Confidence Level: 90.0 %

☒ Display Multi-Vari & X-Bar R Charts

☒ Report Information (Optional)

☒ Tolerance/Historical StDev (Optional)

Gage Name: AIAG Example, MSA Reference

Performed By:

Date:

Notes: Parts were measured in random

☒ Upper-Lower Spec: 8

☐ Upper Spec:

☐ Lower Spec:

Historical Process Standard Deviation:

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		

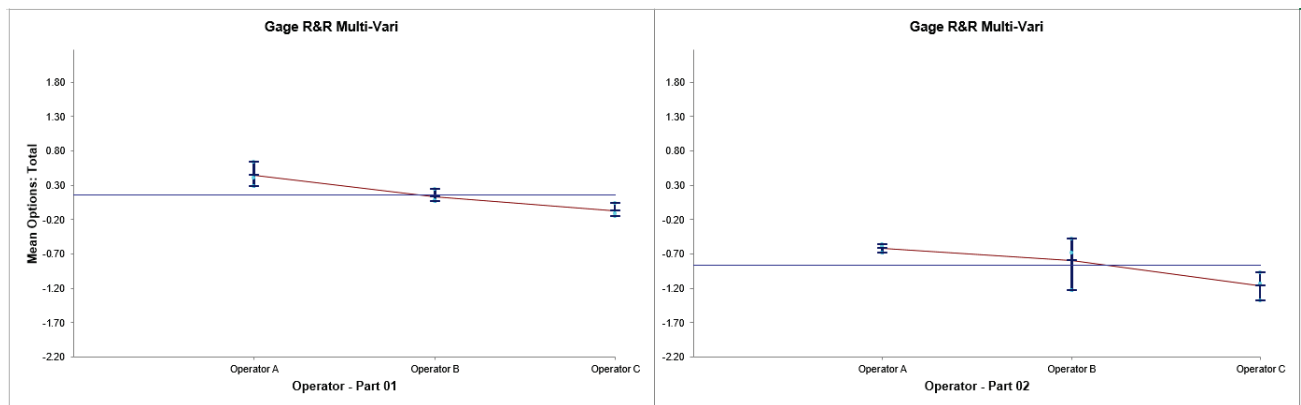
Gage R&R Metrics	StDev	StDev Lower 90% CI	StDev Upper 90% CI	6 * StDev	% Total Variation (TV)	% TV Lower 90% CI	% TV Upper 90% CI
Gage R&R:	0.302372	0.235108	1.03	1.8142	27.86	15.46	70.93
Operator (AV Appraiser Variation):	0.226838	0.127545	1.01	1.3610	20.90		
Part * Operator (INT Interaction):	0	0	0	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

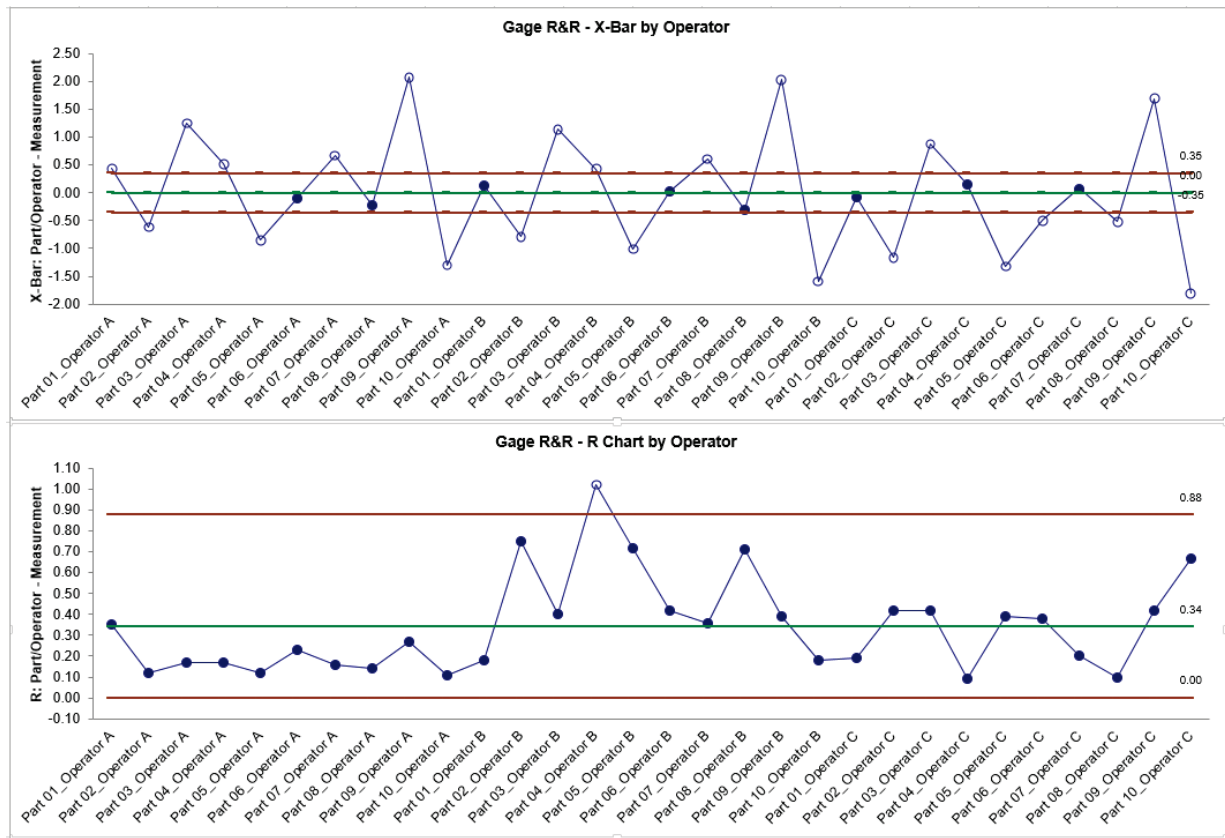
Gage R&R Metrics	NDC	NDC Lower 90% CI	NDC Upper 90% CI
Number of Distinct Categories (Signal-to-Noise Ratio: 1.41 * PV/R&R):	4.9	1.4	9.0

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 :

The screenshot shows the 'Attribute MSA (Binary)' dialog box. It has a title bar with a close button. The main area contains several input fields and buttons. On the left is a large empty box. To its right are four rows of input fields with corresponding buttons: 'Part/Sample >>' with a text box containing 'Part', 'Appraiser >>' with a text box containing 'Appraiser', 'Assessed Result >>' with a text box containing 'Assessed Result', and 'True Standard (Optional) >>' with a text box containing 'Reference'. Below these is a '<< Remove' button. To the right of these fields are three buttons: 'OK >>', 'Cancel', and 'Help'. Below the main input area is a 'Good Level >>' section with a list box showing '0' and '1', where '1' is selected, and a text box containing '1'. At the bottom, there is a section for 'Report Information (Optional)' which is checked. It includes a 'Product/Unit Name' field with 'AIAG Attribute MSA Binary', and empty fields for 'Performed By', 'Date', and 'Notes'. To the right of this is a 'Confidence Level' field with '95.0' and a 'Percent Confidence Interval Type' section with two radio buttons: 'Wilson Score' (unselected) and 'Exact' (selected).

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

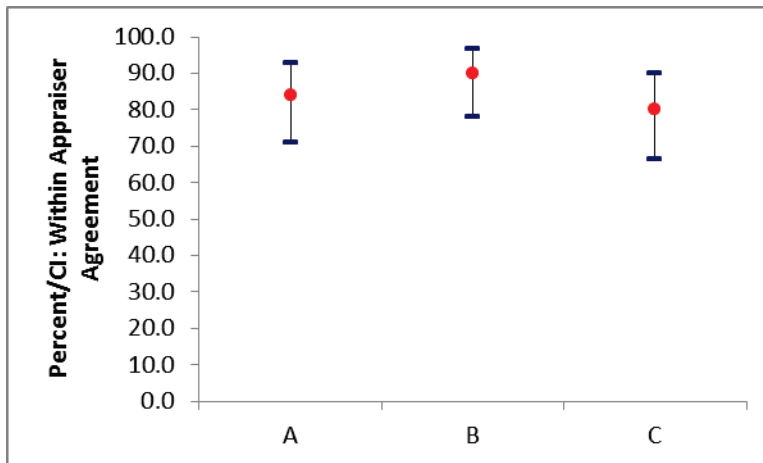
probability that matches the specified confidence interval. Since the intervals are narrower and thereby more powerful, Wilson Score is recommended for use in attribute MSA studies due to the small sample sizes typically used. **Exact** is selected in this example for continuity with the results from 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.

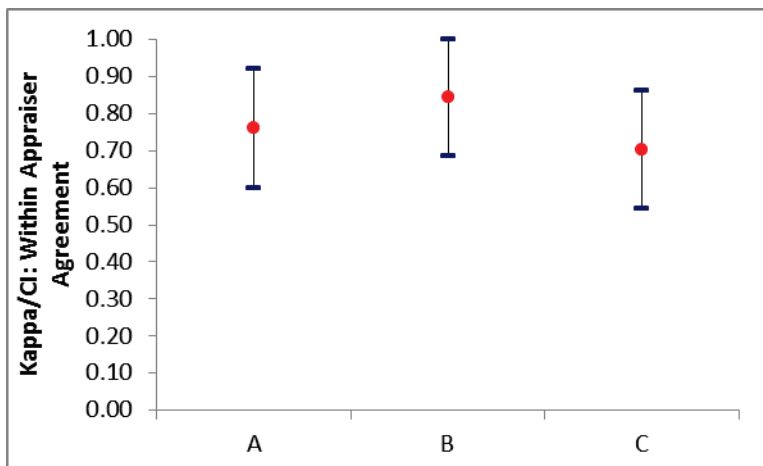
Tip: While this report is quite extensive, a quick assessment of the attribute measurement system can be made by viewing the Kappa color highlights: **Green** - very good agreement (Kappa ≥ 0.9); **Yellow** - marginally acceptable, improvement should be considered (Kappa 0.7 to < 0.9); **Red** - unacceptable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

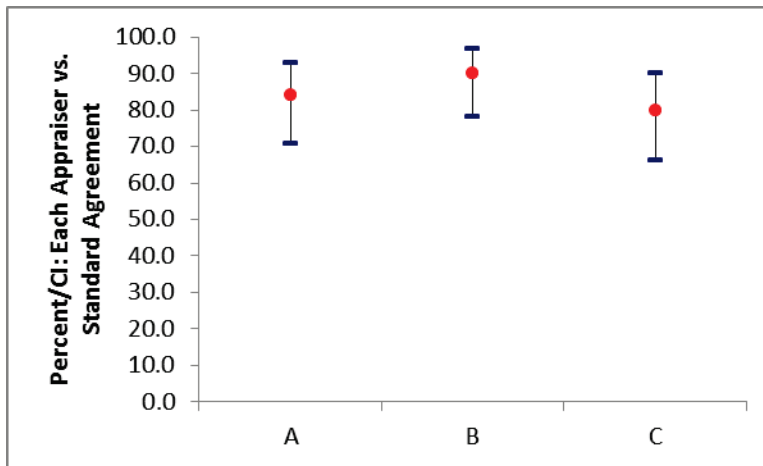
Fleiss' Kappa P-Value: H0: Kappa = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

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

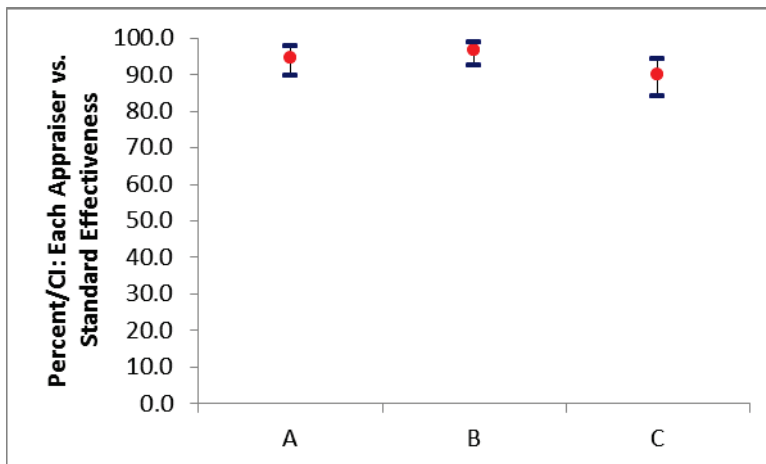
Kappa is interpreted as above: **>= 0.9 very good agreement (green); 0.7 to < 0.9 marginally acceptable (yellow); < 0.7 unacceptable (red).**

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.

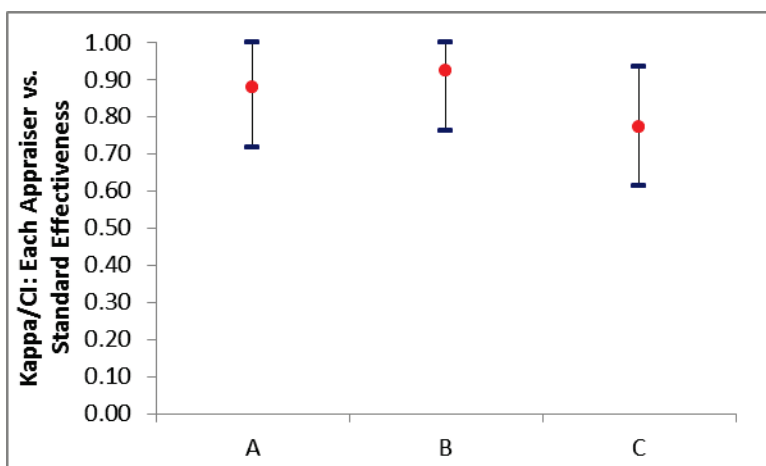
The **Attribute Effectiveness Report** is similar to the **Attribute Agreement Report**, but treats each trial as an opportunity. Consistency across trials or appraisers is not considered. This has the benefit of providing a Percent measure that is unaffected by the number of trials or appraisers. Also, 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	1	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**.

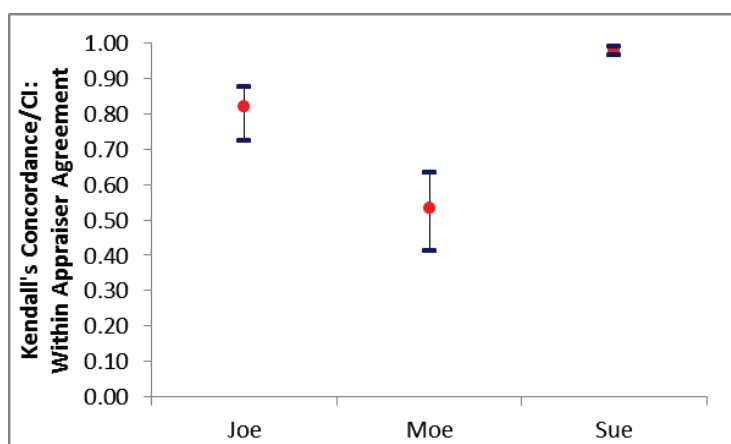
3. Select *Sample No.*, *Appraiser*, *Assessed Result* and *Expert Reference* as shown. Check **Report Information** and enter *Attribute MSA Ordinal* for **Product/Unit Name**. Select **Percent Confidence Interval Type – Wilson Score**:

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	0.9794	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	0.0495	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.

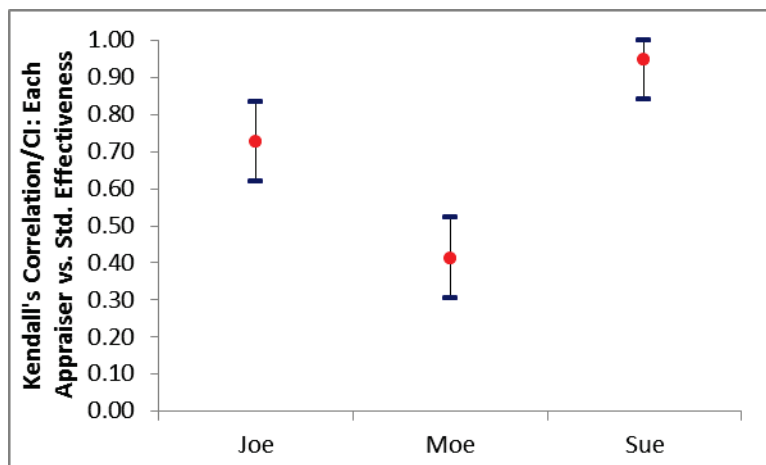
Kendall's Concordance P-Value: H_0 : Kendall's Coefficient of Concordance = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H_0 and conclude that agreement is not the same as would be expected by chance. Significant P-Values are highlighted in red.

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: H_0 : Kendall's Correlation Coefficient = 0. If P-Value < alpha (.05 for specified 95% confidence level), reject H_0 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	5	3	0	4	4
2	2	0	0	2	0	2	2	2	2	2
3	5	5	5	5	5	4	4	5	5	5
4	4	0	0	0	0	0	4	0	0	4
5	3	4	0	0	2	1	5	0	0	0
6	1	0	1	0	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	0	2	5	5	5	5
10	1	1	2	0	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	0.9118	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: H_0 : Kappa = 0. If P-Value $< \alpha$ (.05 for specified 95% confidence level), reject H_0 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](#)

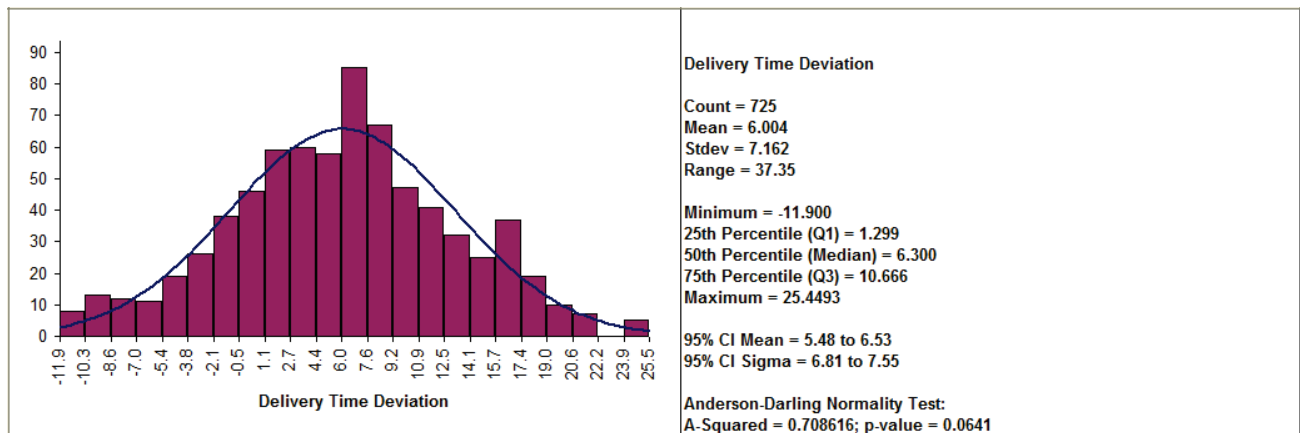
[Basic Process Capability Templates – Process Capability Indices](#)

[Basic Process Capability Templates – Process Capability & Confidence Intervals](#)

[Basic Process Capability Templates – Tolerance Interval Calculator \(Normal Exact\)](#)

Histograms and Process Capability

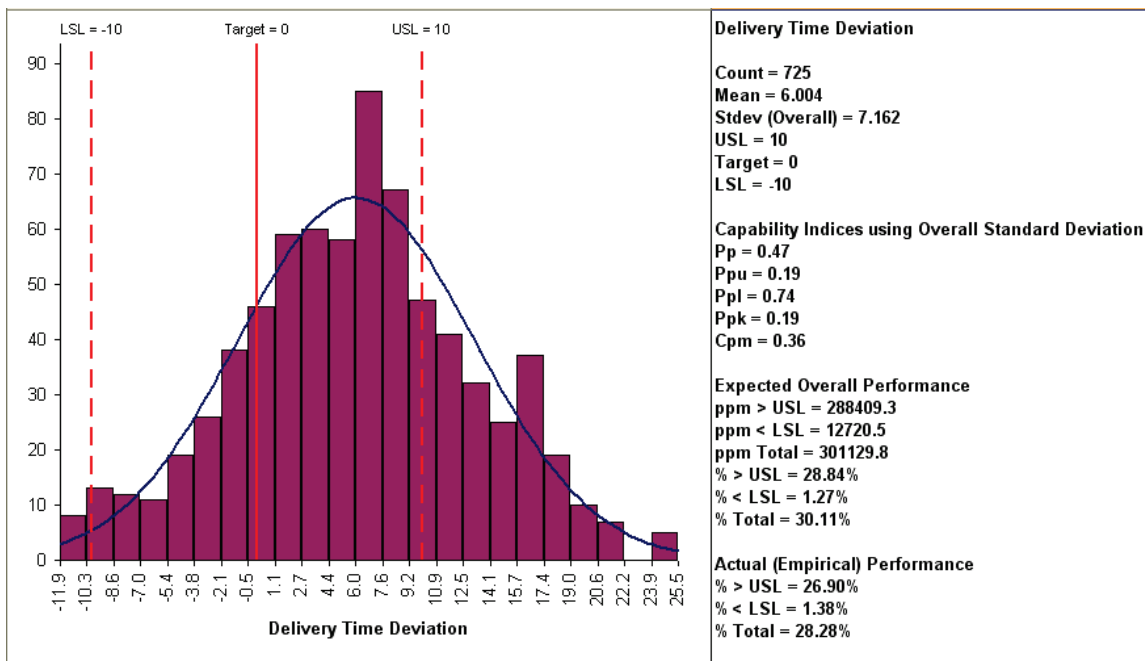
1. Open the file **Delivery Times.xlsx**. This contains continuous data of hotel breakfast delivery times. Deviation Time is the deviation around targeted delivery time in minutes. The Critical Customer Requirements (CCR's) are as follows: USL = 10 minutes late, LSL = -10 minutes (early).
2. Let's begin 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.

- Click **Sheet 1** Tab. Click **SigmaXL > Process Capability > Histograms & Process Capability** (or **SigmaXL > Graphical Tools > Histograms & Process Capability**).
- Select *Delivery Time Deviation*, click **Numeric Data Variable (Y) >>**. Enter **USL = 10**, **Target = 0**, **LSL = -10**, check **Normal Curve** as shown below:

- Click **OK**. Resulting Histogram & Process Capability report is shown below:



With the Process Performance indices P_p and $P_{pk} < 1$, this process is clearly in need of improvement. Note that the difference between P_p and P_{pk} is due to the off-center process mean. (C_p and C_{pk} 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:

Capability Combination Report (Individuals)

Delivery Time Deviation
Defects
Floor

Numeric Data Variable (Y) >> Delivery Time Deviation

Optional X-Axis Labels >>

<< Remove

OK >>
Cancel
Help

Individuals/Moving Range Control Chart Options

☒ Calculate Control Limits
☐ Historical Control Limits

☐ Tests for Special Causes

☐ Add Title

Individuals

UCL
CL
LCL

Moving Range

UCL
CL
LCL

Estimate:

☒ Average MR
☐ Median MR

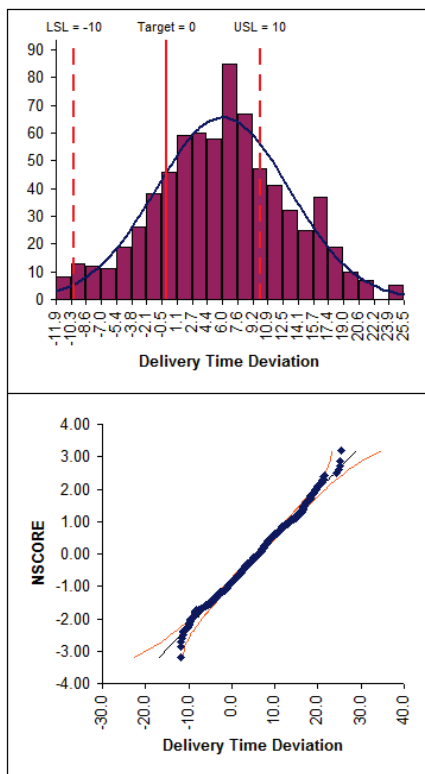
Individual CL

☒ Mean
☐ Median

Spec Limits for Capability

USL 10
Target 0
LSL -10

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

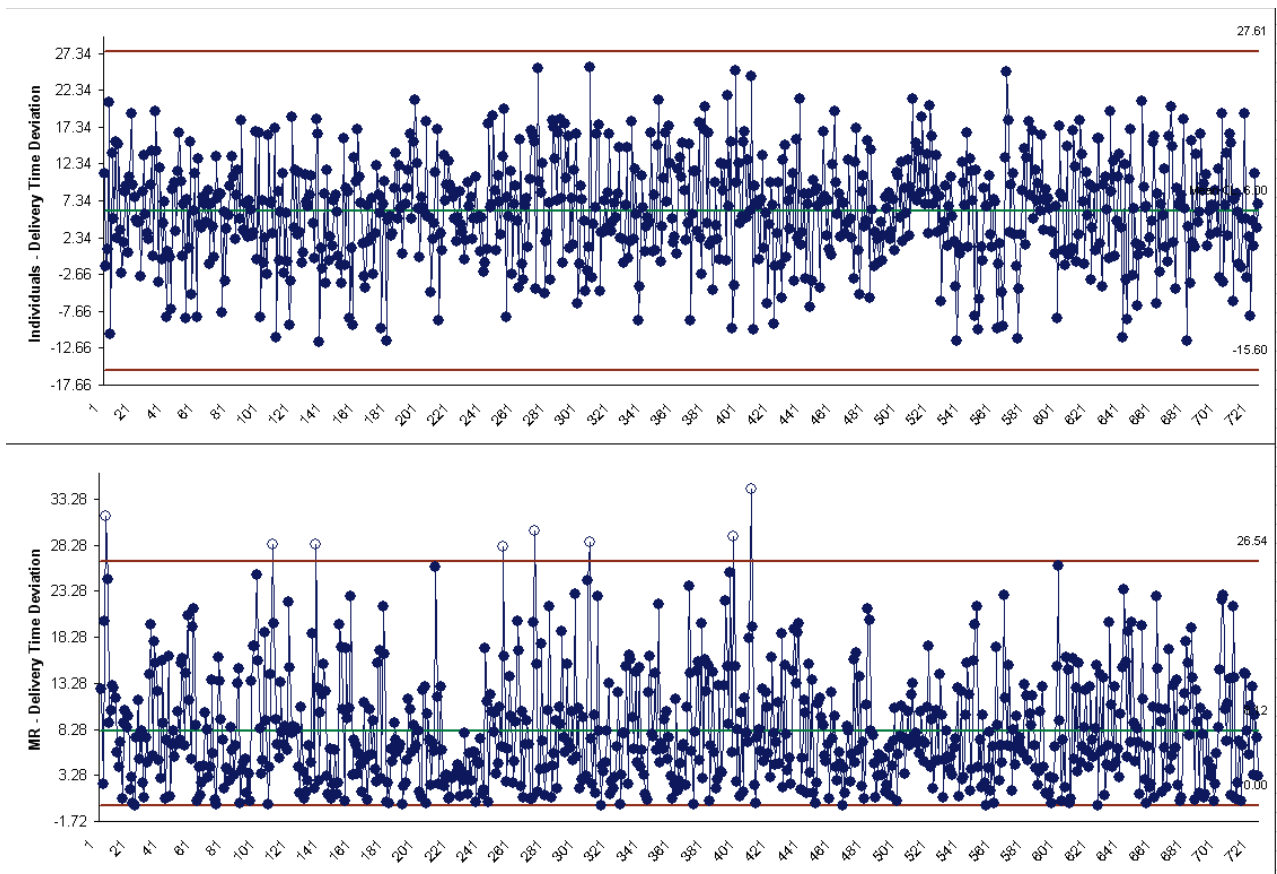
Capability Indices using Overall StDev		
Pp		0.47
Ppu		0.19
Ppl		0.74
Ppk		0.19
Cpm		0.36

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		301129.8
% > USL		28.84%
% < LSL		1.27%
% Total		30.11%

Actual (Empirical) Performance		
% > USL		26.90%
% < LSL		1.38%
% Total		28.28%

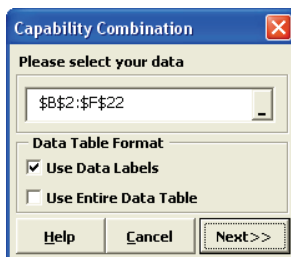
Anderson-Darling Normality Test		
A-Squared		0.708616
p-value		0.0641



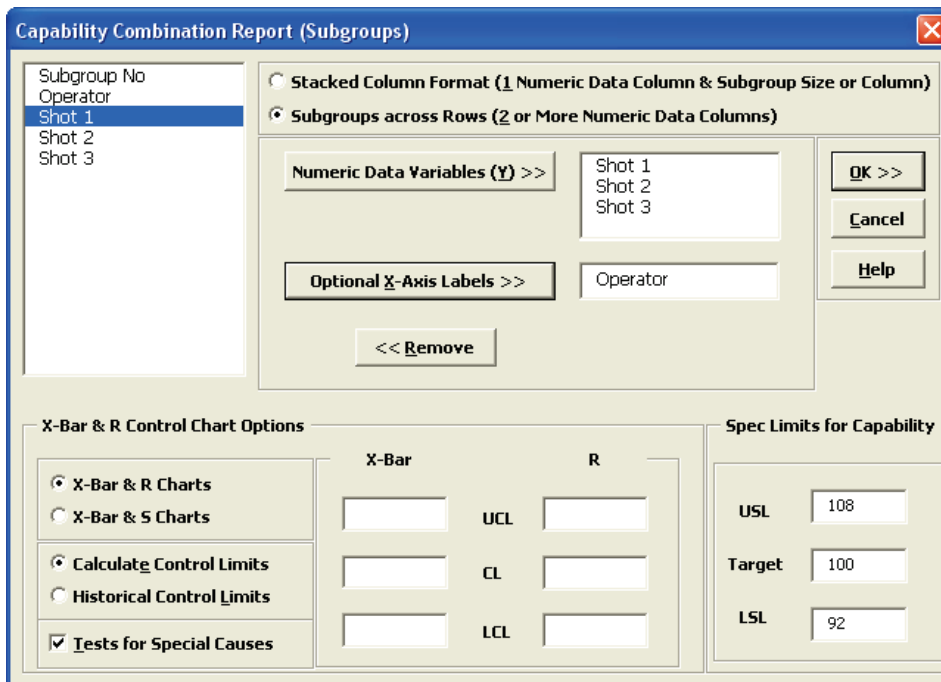
5. To print this report, click **SigmaXL Chart Tools > Print Charts**. Do not use Excel's print function.
6. Cp is the "best case" potential capability index. Ppk is the "worst case" actual performance. If Cp is excellent (say $C_p = 2$) and Ppk is poor (say $P_{pk} < 1$), this indicates that the process 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 $P_{pk} = C_p$, 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.
7. Cpm is a similar statistic but it incorporates a penalty for deviation from target.
8. 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)

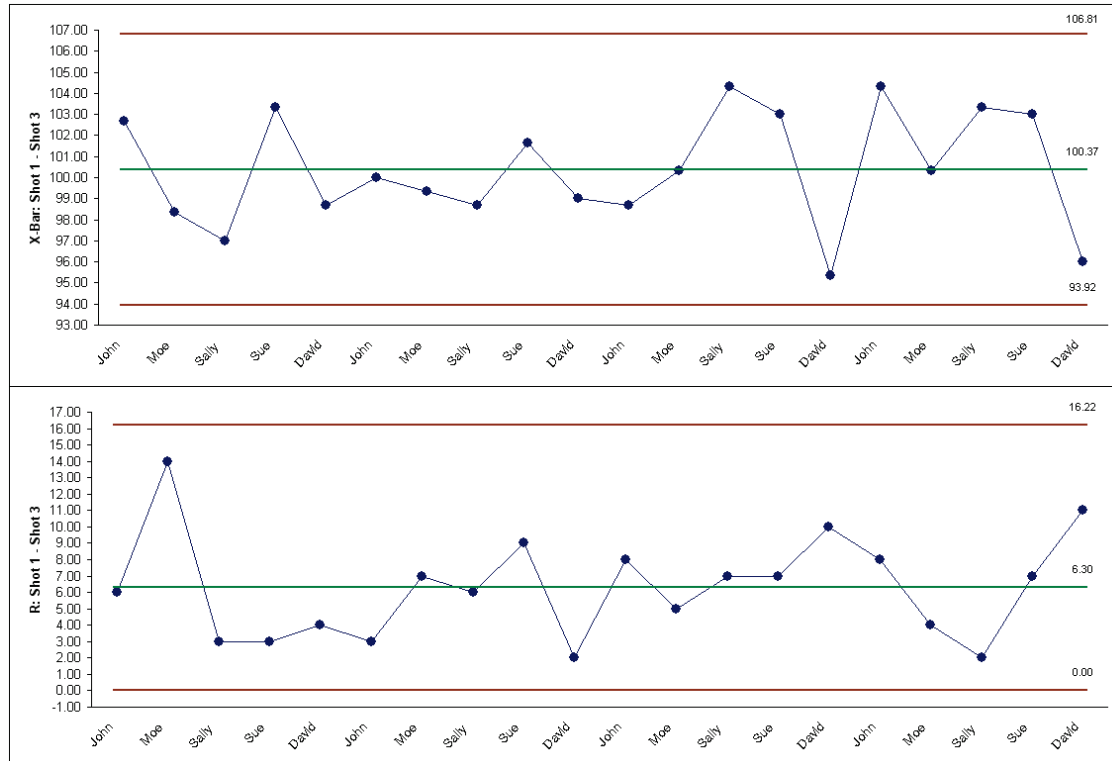
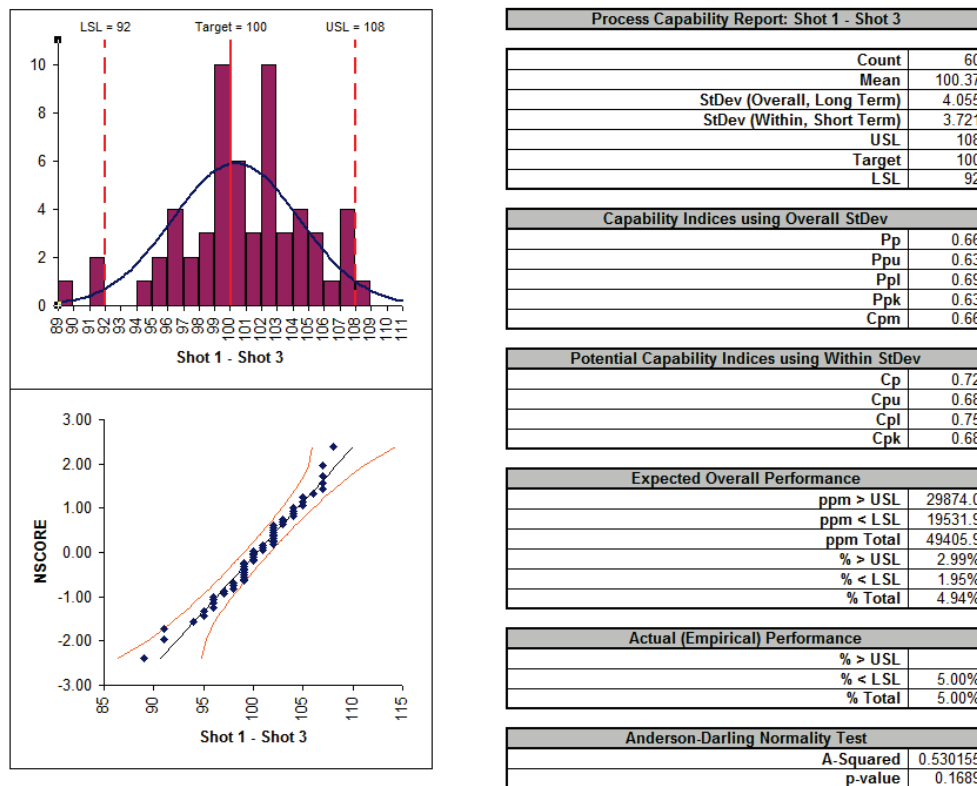
1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the catapult ball 3 times. The target distance is 100 inches. The Upper Specification Limit (USL) is 108 inches. The Lower Specification Limit (LSL) is 92 inches.
2. Select **B2:F22**; here, we will only use the first 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:



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

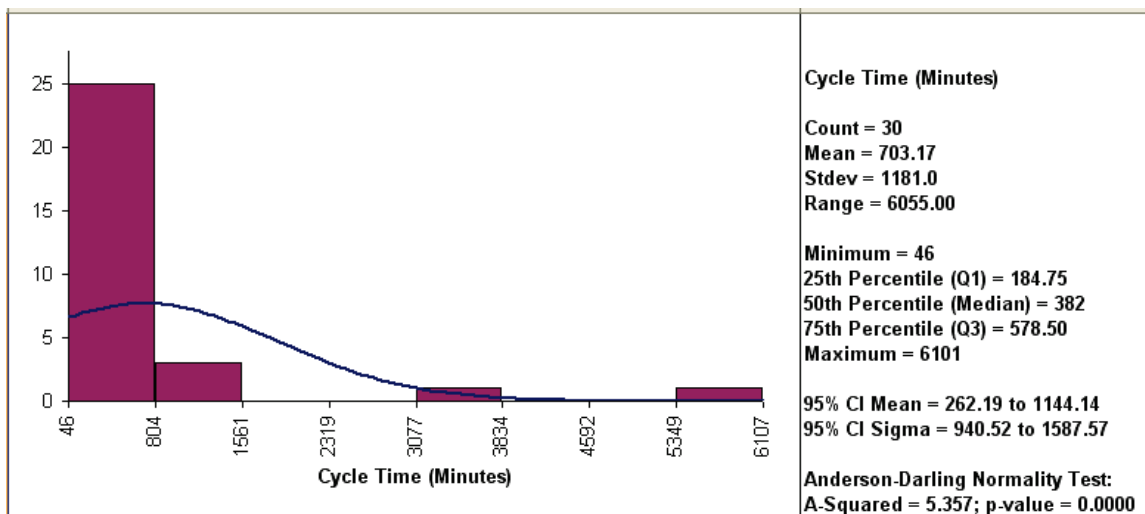
For technical details, see Appendix: [**Statistical Details for Nonnormal Distributions and Transformations**](#). See also Andrew Sleeper, *Six Sigma Distribution Modeling*, for further information on these methods.

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^λ , where λ varies from -5 to +5. You may select rounded or optimal λ . 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 λ & Threshold. SigmaXL will solve for an optimal threshold which is a shift factor on the data so that all of the values are positive.

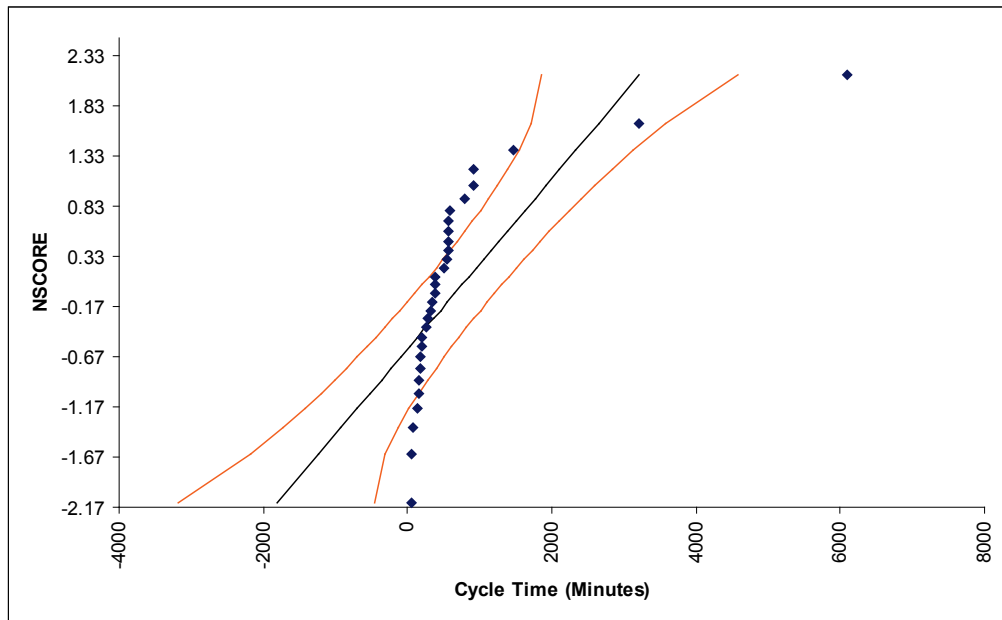
1. Open the file **Nonnormal Cycle Time2.xlsx**. This contains continuous data of process cycle times. The Critical Customer Requirement is: USL = 1000 minutes.
2. Let's begin with a view of the data using Histograms and Descriptive Statistics. Click **SigmaXL > Graphical Tools > Histograms & Descriptive 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.

Capability Combination Report (Individuals)

Cycle Time (Minutes)

Numeric Data Variable (Y) >> Cycle Time (Minutes)

Optional X-Axis Labels >>

<< Remove

OK >>

Cancel

Help

Individuals/Moving Range Control Chart Options

☒ Calculate Control Limits
☐ Historical Control Limits

☐ Tests for Special Causes

☐ Add Title

Individuals

UCL

CL

LCL

Moving Range

Estimate:

☒ Average MR
☐ Median MR

Individual CL

☒ Mean
☐ Median

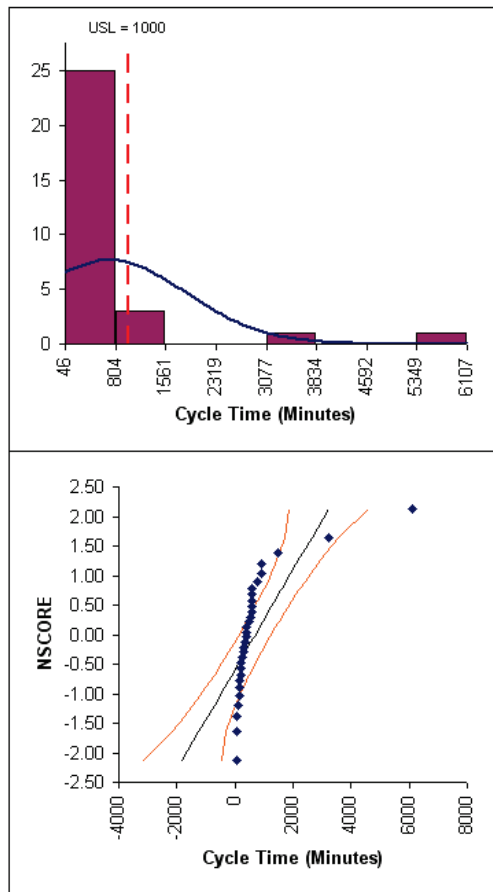
Spec Limits for Capability

USL 1000

Target

LSL

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	

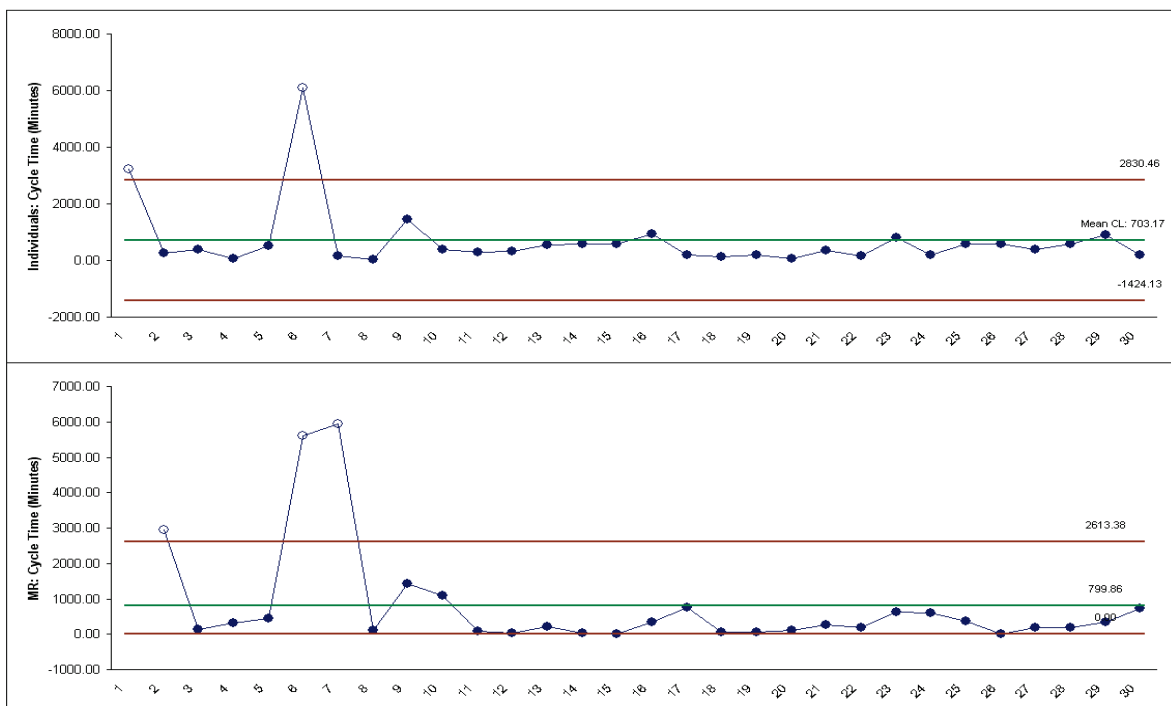
Capability Indices using Overall StDev	
Pp	
Ppu	0.08
Ppl	
Ppk	0.08
Cpm	

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	400771.3
% > 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:

Capability Combination Report (Individuals - Nonnormal)

Cycle Time (Minutes)

Numeric Data Variable (Y) >> Cycle Time (Minutes)

Optional X-Axis Labels >>

<< Remove

OK >> Cancel Help

Control Chart Options

- ☒ Individuals - Original Data
- ☒ Individuals/Moving Range - Normalized Data
- ☒ Tests for Special Causes
- ☐ Add Title

Spec Limits for Capability

USL 1000

Target

LSL

Transformation/Distribution Options

- ☒ Box-Cox Transformation
- ☐ Johnson Transformation
- ☐ Automatic Best Fit
- ☐ Specify Distribution

Box-Cox Transformation Options

- ☒ Rounded Lambda
- ☐ Optimal Lambda
- ☐ Lambda & Threshold (Shift)

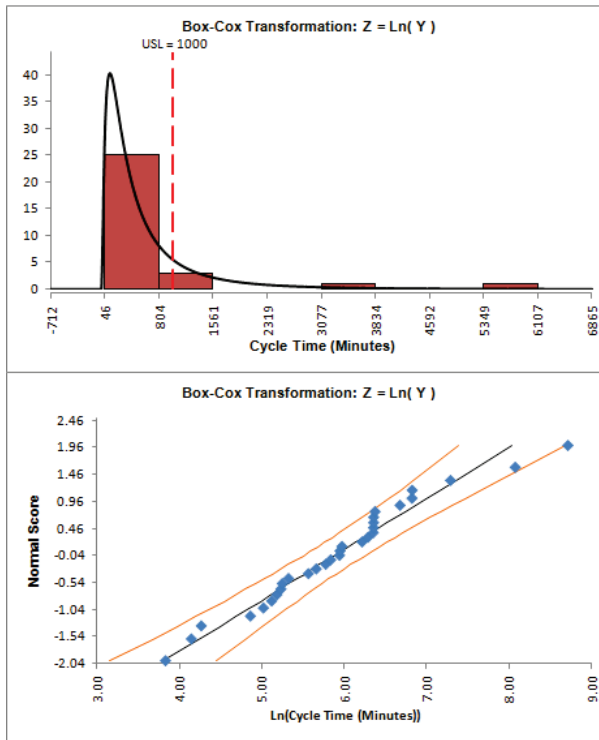
Optional Threshold Value

Optional Lambda Value

Capability Index Options

- ☒ Z-Score Method
- ☐ Percentile (ISO) Method

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)	

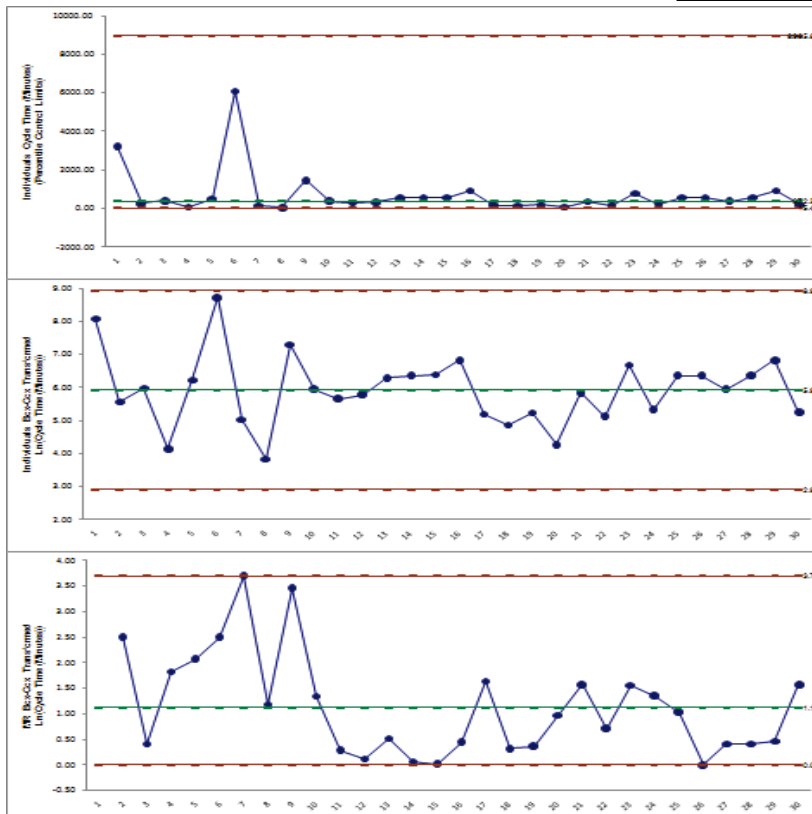
Capability Indices using Transformed Overall StDev	
Pp	
Ppu	0.31
Ppl	
Ppk	0.31

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:

Capability Combination Report (Individuals - Nonnormal)

Cycle Time (Minutes)

Numeric Data Variable (Y) >> Cycle Time (Minutes) OK >>

Optional X-Axis Labels >>

<< Remove Cancel Help

Control Chart Options

☒ Individuals - Original Data

☒ Individuals/Moving Range - Normalized Data

☒ Tests for Special Causes

☐ Add Title

Spec Limits for Capability

USL 1000

Target

LSL

Capability Index Options

☒ Z-Score Method

☒ Percentile (ISO) Method

Transformation/Distribution Options

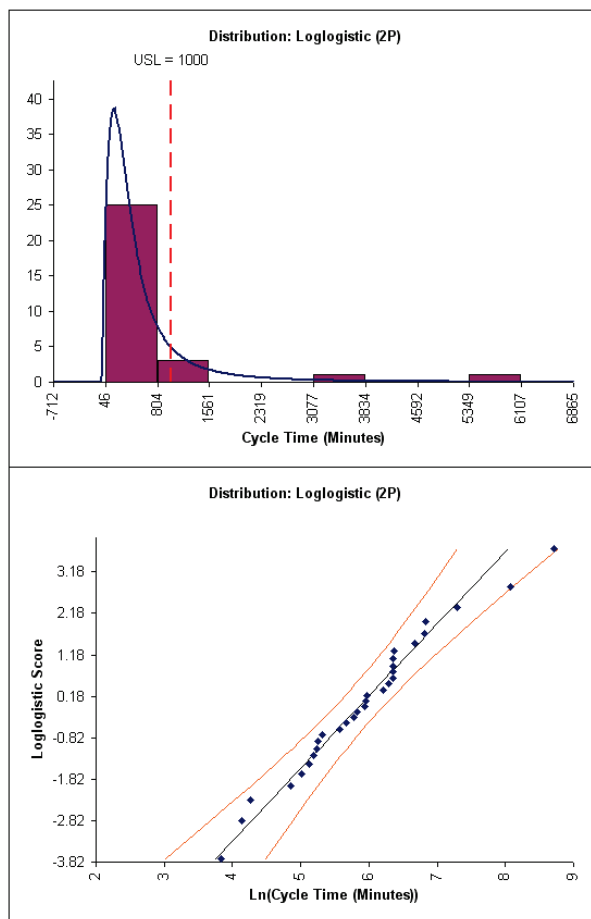
☐ Box-Cox Transformation

☐ Johnson Transformation

☒ Automatic Best Fit

☐ Specify Distribution

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	0.571470906
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	145274.9
ppm < LSL	
ppm Total	145274.9
% > 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

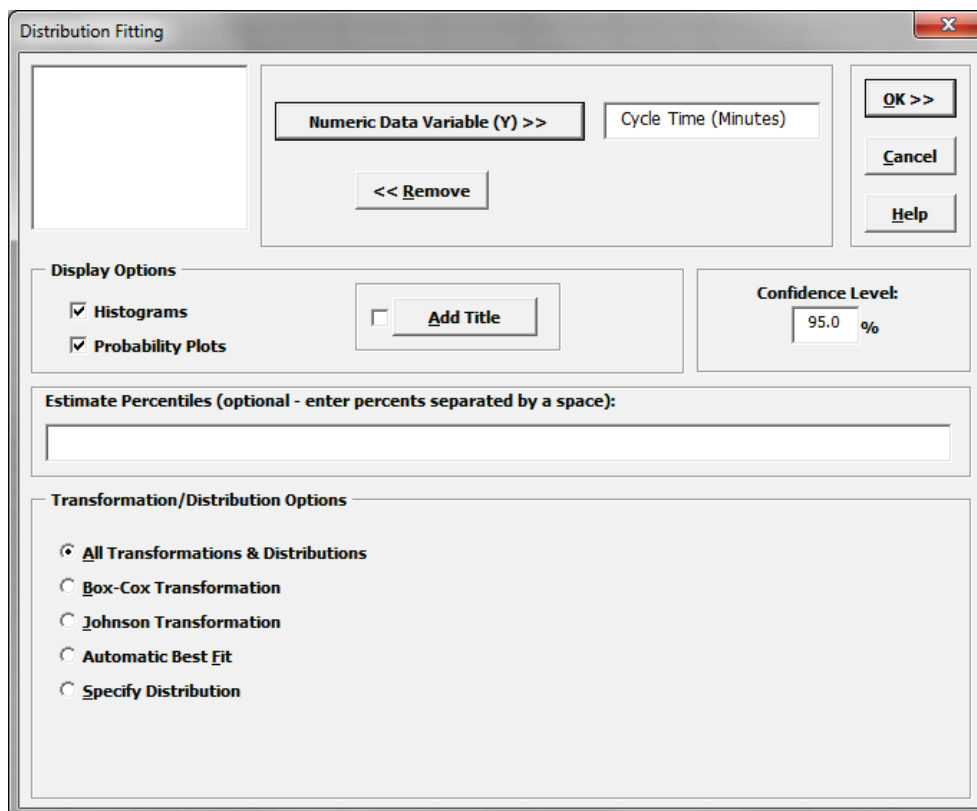
The 2 Parameter Loglogistic distribution was selected as the best fit distribution. For details on how this selection was made, see Appendix: [Statistical Details for Nonnormal Distributions and Transformations](#).

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).

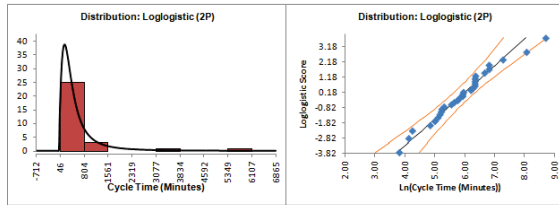
Distribution Fitting

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:



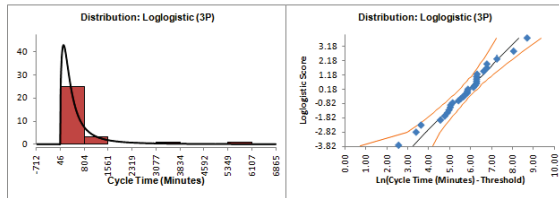
3. Click **OK**. The resulting Distribution Fitting report is shown below. Please note that due to the extensive computations required, this could take up to 1 minute (or longer for large datasets):



Distribution Fitting Report: Cycle Time (Minutes)	
Distribution: Loglogistic (2P)	
Sample Count	30
Sample Mean	703.17
Parameter Estimates	
Parameter	Estimate
Location	5.895
Scale	0.571471

Model Summary and Goodness-of-Fit	
Log-Likelihood	-221.011
AD Loglogistic (2P)	0.245264
AD Loglogistic (2P) p-value	> 0.2500

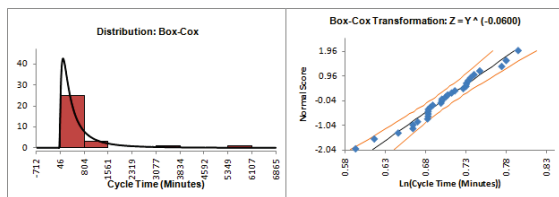
Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	0.329	5.067	2.520	27.442
50	363.22	65.277	255.39	516.59
99.865	15841	9631.1	4811.0	52156



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.)

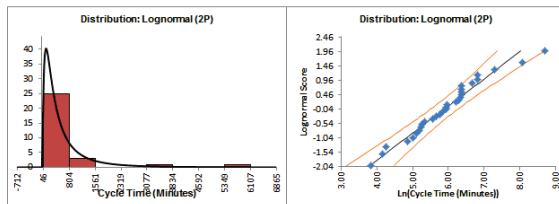
Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.771	0.232556	5.315	6.227
Scale	0.613791	0.151455	0.413179	1.045070
Threshold (ML Estimate)	33.025	23.549	-13.129	79.180
Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	35.816	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: $Z = Y^{\lambda} (-0.0600)$				
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

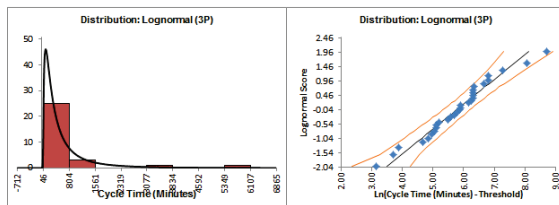
Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Lambda (Optimal)	-0.06	0.073963	-0.573963	0.423205
Mean (Transformed Data)	0.702382	0.007937578	0.688025	0.717639
StDev (Transformed Data)	0.044219135	0.00570855	0.034333861	0.056956541
Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	11810	20.605	36159	4138.7
50	360.67	67.889	523.88	250.34
99.865	20.179	25811	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

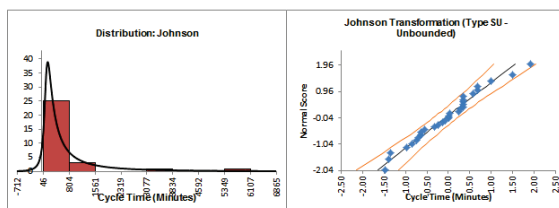
Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	5.920	0.193717	5.541	6.300
Scale	1.061030703	0.140537	0.818434	1.376
Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	15.445	7.168	6.221	38.348
50	372.54	72.167	254.85	544.59
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	23.764	-23.706	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	95.419
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)				
$Z = \text{Shape1} + \text{Shape2} * \text{Asinh}((Y - \text{Location}) / \text{Scale})$				
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	0.144403	-0.330234	0.236931
StDev (Transformed Data)	0.894576	0.103876532	0.624699	1.036246788
Percentile Report				
Percentage	Cycle Time (Minutes)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	-380.45		-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):

Capability Combination Report (Individuals - Nonnormal)

Cycle Time (Minutes)

Numeric Data Variable (Y) >> Cycle Time (Minutes) **OK >>**

Optional X-Axis Labels >> **Cancel**

<< Remove **Help**

Control Chart Options

☒ Individuals - Original Data

☒ Individuals/Moving Range - Normalized Data

☐ Tests for Special Causes

☐ Add Title

Spec Limits for Capability

USL 1000

Target

LSL

Capability Index Options

☒ Z-Score Method

☐ Percentile (ISO) Method

Transformation/Distribution Options

☐ Box-Cox Transformation

☐ Johnson Transformation

☐ Automatic Best Fit

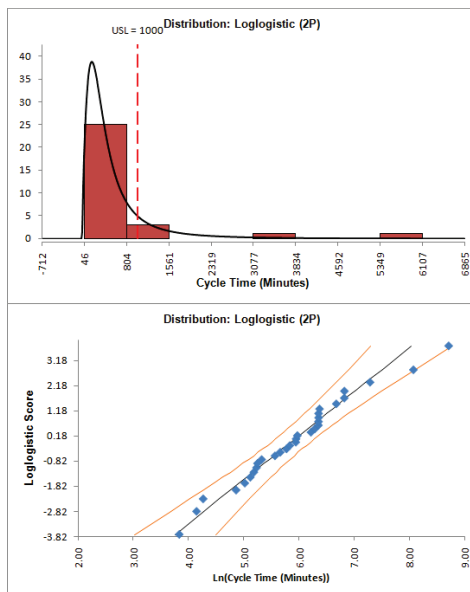
☒ Specify Distribution

Selected Distribution

Gamma with Threshold (3 Parameter)

Logistic (2 Parameter)

Loglogistic (2 Parameter)



Process Capability Report (Nonnormal): Cycle Time (Minutes)	
Distribution: Loglogistic (2P)	
Sample Count	30
Sample Mean	703.17
Location	5.895
Scale	0.571470925
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	145274.9
ppm < LSL	
ppm Total	145274.9
% > USL	14.53%
% < LSL	
% Total	14.53%

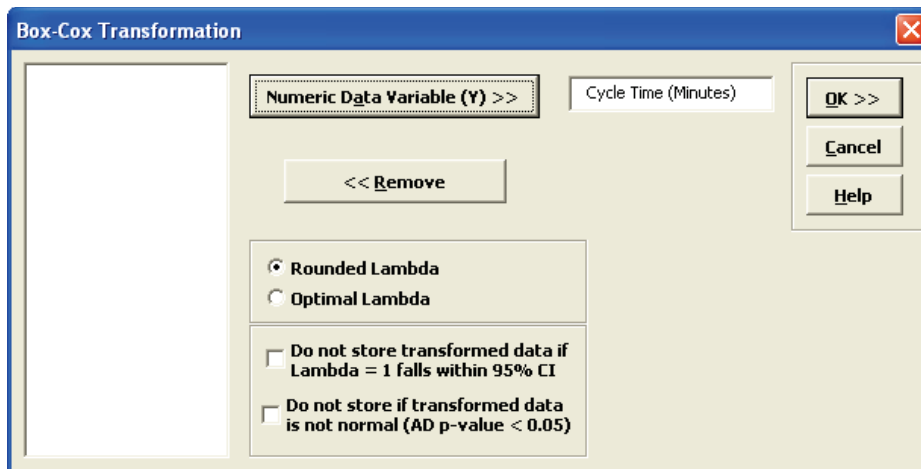
Actual (Empirical) Performance	
% > USL	10.00%
% < LSL	
% Total	10.00%

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

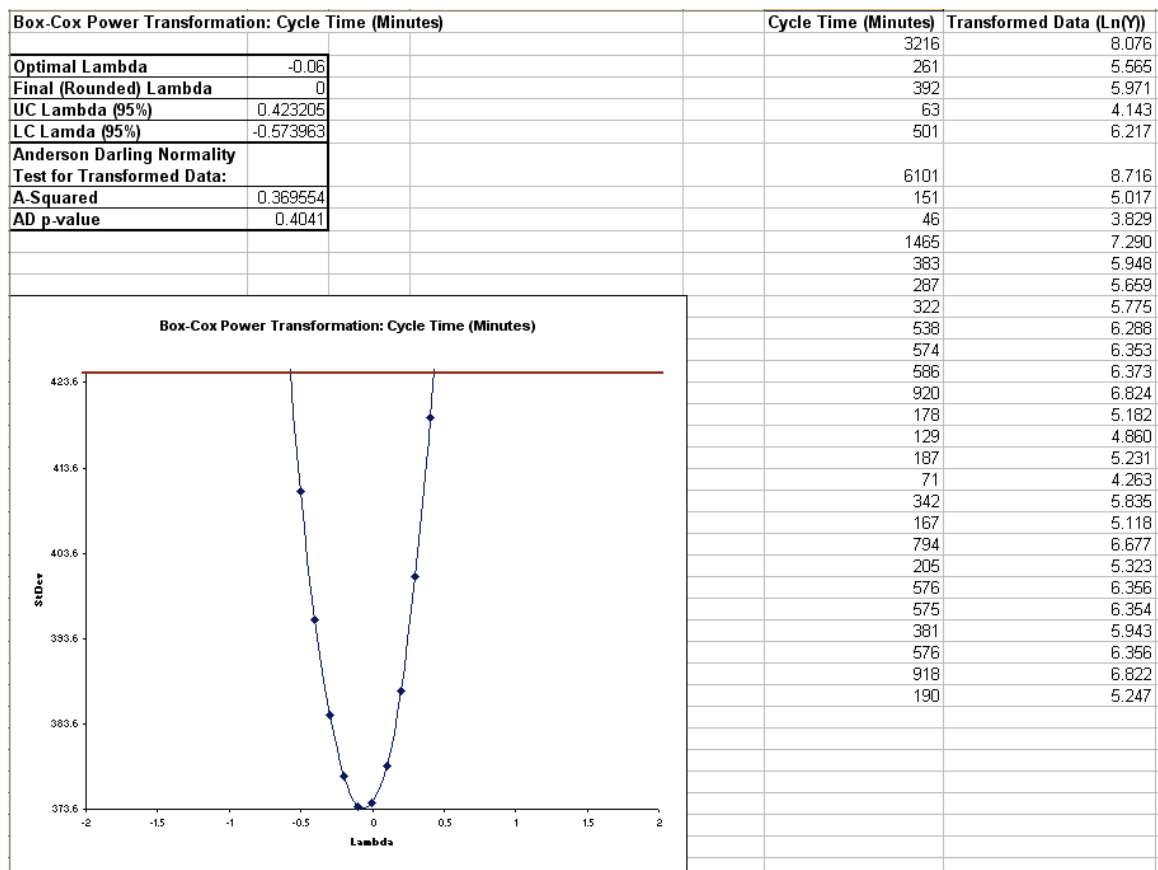
This is a standalone tool that allows you to visually see how the Box-Cox transformation selects a rounded or optimal lambda value.

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:



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.

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:

Reliability/Weibull Analysis

Censor Description

Numeric Response (Y) >> Time-to-Fail
(Time to Event)

Right Censor Column >> Censor Code
(Optional: Text or Numeric)

<< Remove

Censor Value >> 0 1

Method of Estimation:
☒ Maximum Likelihood
☐ Least Squares Regression (X on Y)

Threshold: 0.0
Confidence Level: 90.0 %

Estimate Additional Time Percentiles (enter percents separated by a space):

Estimate Survival Probabilities (enter time values separated by a space):
 10 20 30 40 50 60 70 80 90 100 150 200 250 300

OK >>
Cancel
Help

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:

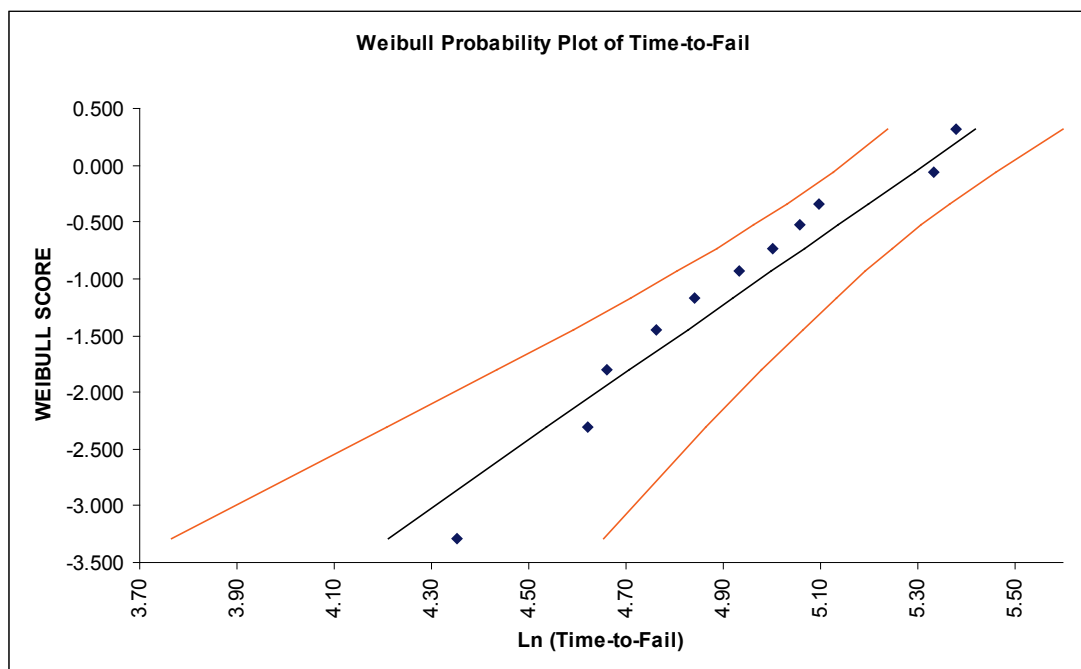
	Estimate	SE Estimate	Lower 90% CI	Upper 90% CI
Mean (MTTF)	181.47	18.420	153.57	214.44
Standard Deviation	66.463	16.172	44.541	99.174

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:

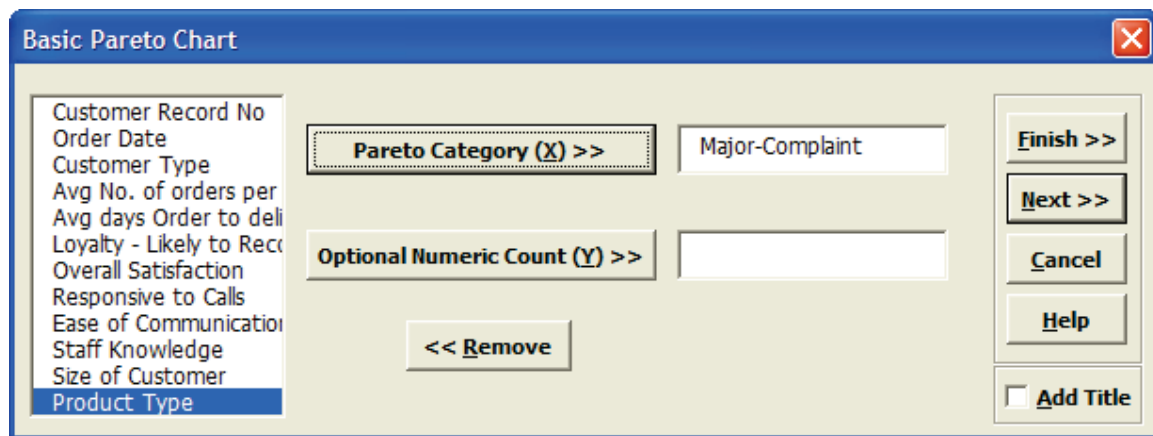
- Your Pareto analysis will only be as good as the quality of the data collected. Ensure that you have the right data and that the data is correct. Use other graphs such as run charts to apply a sanity check to your data.
- Check process stability using appropriate control charts. If the process is not 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

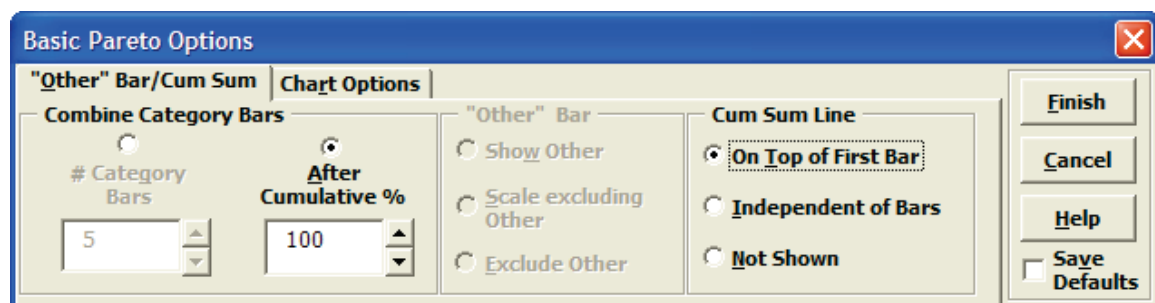
1. Open the file **Customer Data.xlsx**. Click **SigmaXL>Graphical Tools >Basic Pareto Chart**.
2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
3. Select *Major-Complaint*, click **Pareto Category (X) >>**.

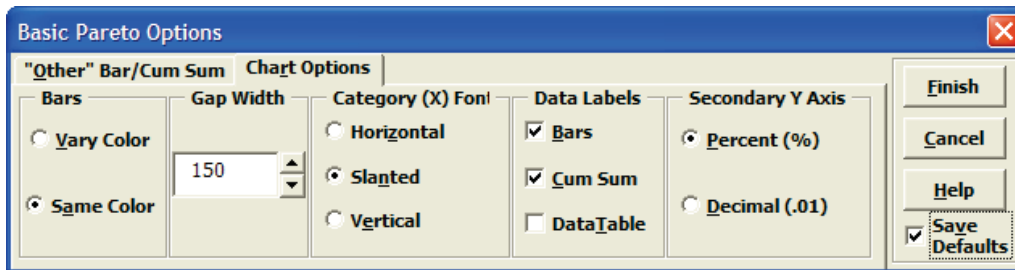


Tip: SigmaXL will automatically count the number of unique items in the Pareto Category. If we had a separate column with a count (or cost), this count column would be selected as the Optional Numeric Count (Y).

4. Click **Next**. Set **Basic Chart Options** as follows:

Tab “Other” Bar/Cum Sum:
Cum Sum Line – On Top of First Bar

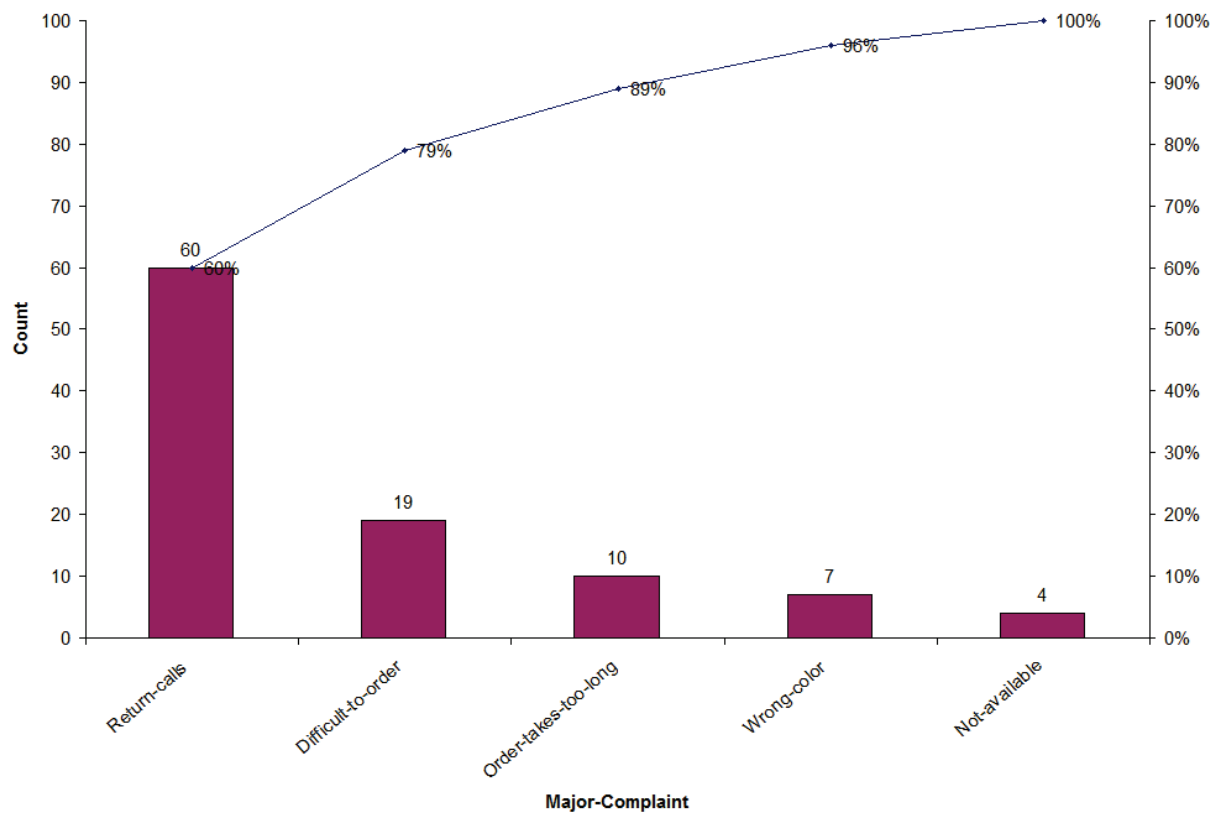


Tab **Chart Options**:**Category (X) Font – Slanted****Data Labels – Check Bars, Cum Sum**Check **Save Defaults**.

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.

"Other" Bar/Cum Sum | Chart Options

Combine Category Bars <input checked="" type="checkbox"/> Unlimited Categories # of Categories: 20	"Other" Bar <input type="radio"/> Show Other <input type="radio"/> Exclude Other	Order of Bars <input type="radio"/> Descending <input checked="" type="radio"/> Same Order	Cum Sum Line <input type="radio"/> On Top of First Bar <input checked="" type="radio"/> Independent of Bars <input type="radio"/> Not Shown	Finish Cancel Help <input checked="" type="checkbox"/> Save Defaults
---	---	---	---	---

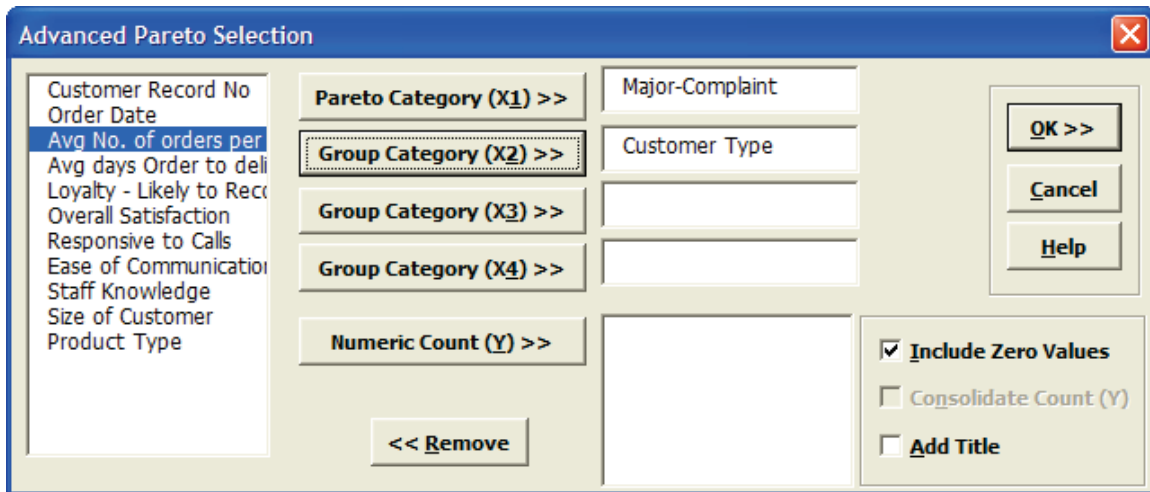
4. Click **Chart Options** tab. Set according to choice – in this case we have selected **Data Labels** for the **Bars** but not for the **Cum Sum** line.

"Other" Bar/Cum Sum | Chart Options

Bars <input type="radio"/> Vary by Color <input checked="" type="radio"/> Same color	Gap Width 150	Category (X) Font <input type="radio"/> Horizontal <input checked="" type="radio"/> Slanted <input type="radio"/> Vertical	Data Labels <input checked="" type="checkbox"/> Bars <input type="checkbox"/> Cum Sum <input type="checkbox"/> Data Table	Secondary Y Axis <input checked="" type="radio"/> Percent (%) <input type="radio"/> Decimal (.01)	Finish Cancel Help <input checked="" type="checkbox"/> Save Defaults
---	-------------------------	--	---	--	---

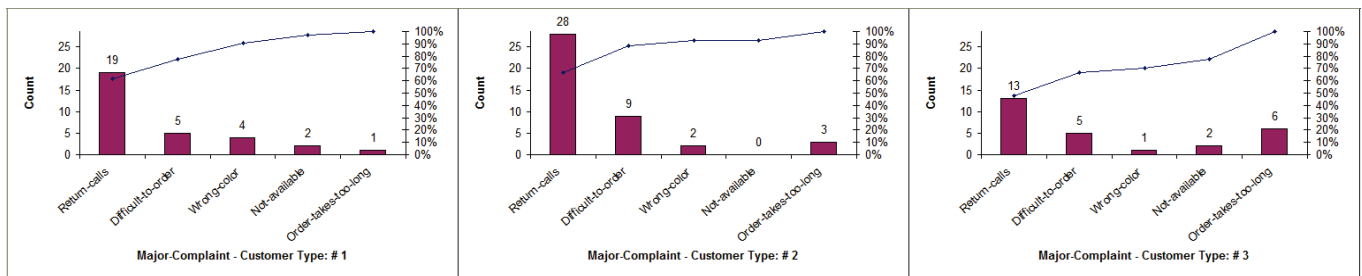
5. Ensure that **Save Defaults** is checked. Note that these options will be saved and applied to all Advanced Pareto Charts. Click **Finish**
6. SigmaXL automatically takes you to the next step of Chart Generation (This is equivalent to clicking **SigmaXL > Graphical Tools > Advanced Pareto Charts**). 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) >>**.



The dialog box titled "Advanced Pareto Selection" contains a list of variables on the left: Customer Record No, Order Date, Avg No. of orders per, Avg days Order to deli, Loyalty - Likely to Rec, Overall Satisfaction, Responsive to Calls, Ease of Communication, Staff Knowledge, Size of Customer, and Product Type. The "Avg No. of orders per" variable is selected. On the right, there are four input fields for Pareto Category (X1), Group Category (X2), Group Category (X3), and Group Category (X4). The first field contains "Major-Complaint" and the second contains "Customer Type". Below these is a "Numeric Count (Y) >>" field. At the bottom left is a "<< Remove" button. On the right side of the dialog are buttons for "OK >>", "Cancel", and "Help". At the bottom right, there are three checkboxes: "Include Zero Values" (checked), "Consolidate Count (Y)" (unchecked), and "Add Title" (unchecked).

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).

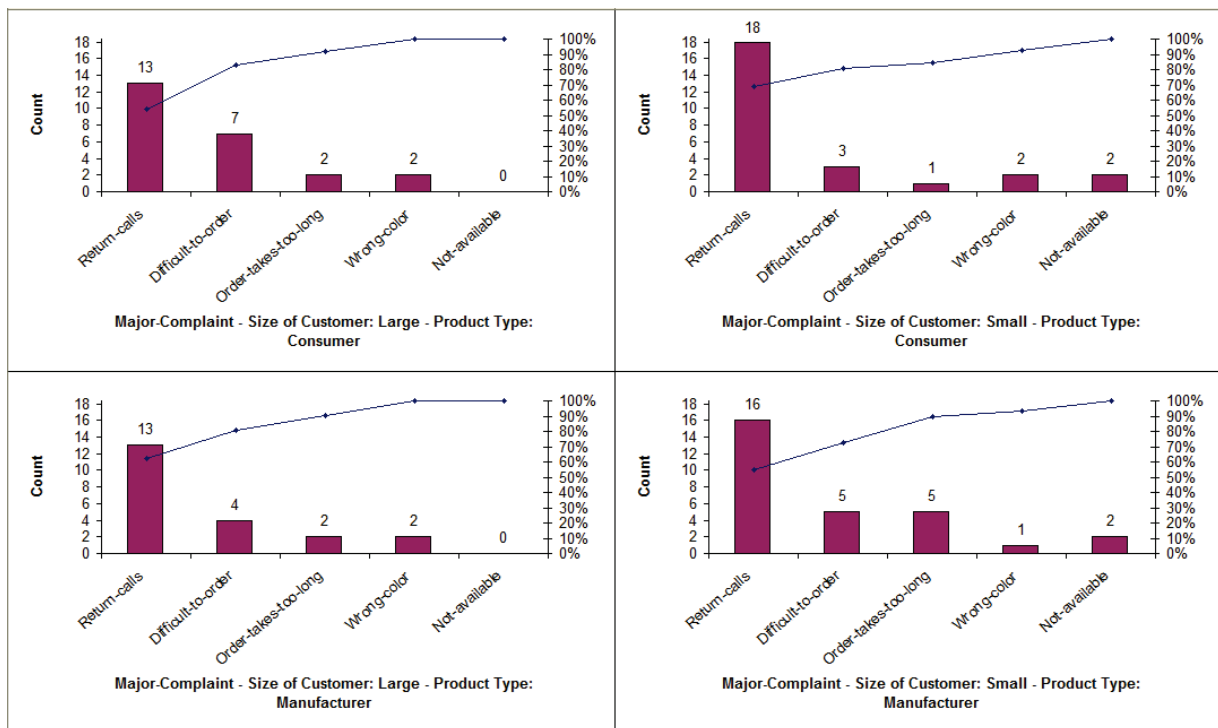
12. Select *Major Complaint*, click **Pareto Category (X1) >>**; select *Size of Customer*, click **Group Category (X2) >>**; select *Product Type*, click **Group Category (X3) >>**.

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:

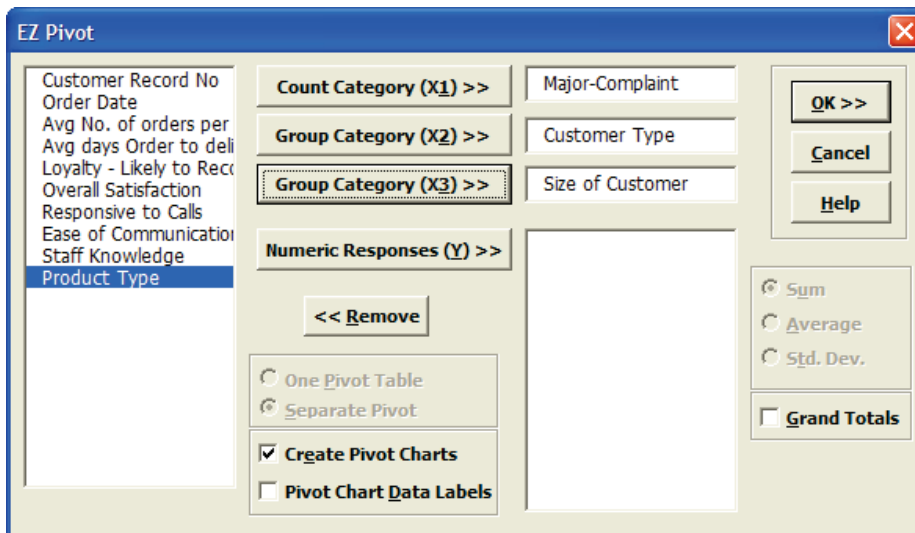


Part B - EZ-Pivot

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

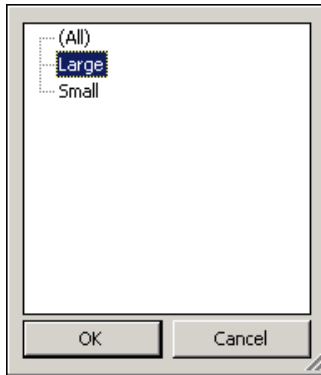
1. Open **Customer Data.xlsx**, click **Sheet 1** (or press **F4** to activate last worksheet). Select **SigmaXL > Graphical Tools > EZ-Pivot/Pivot Charts**.
2. Ensure that entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
3. Select *Major Complaint*, click **Count 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**.



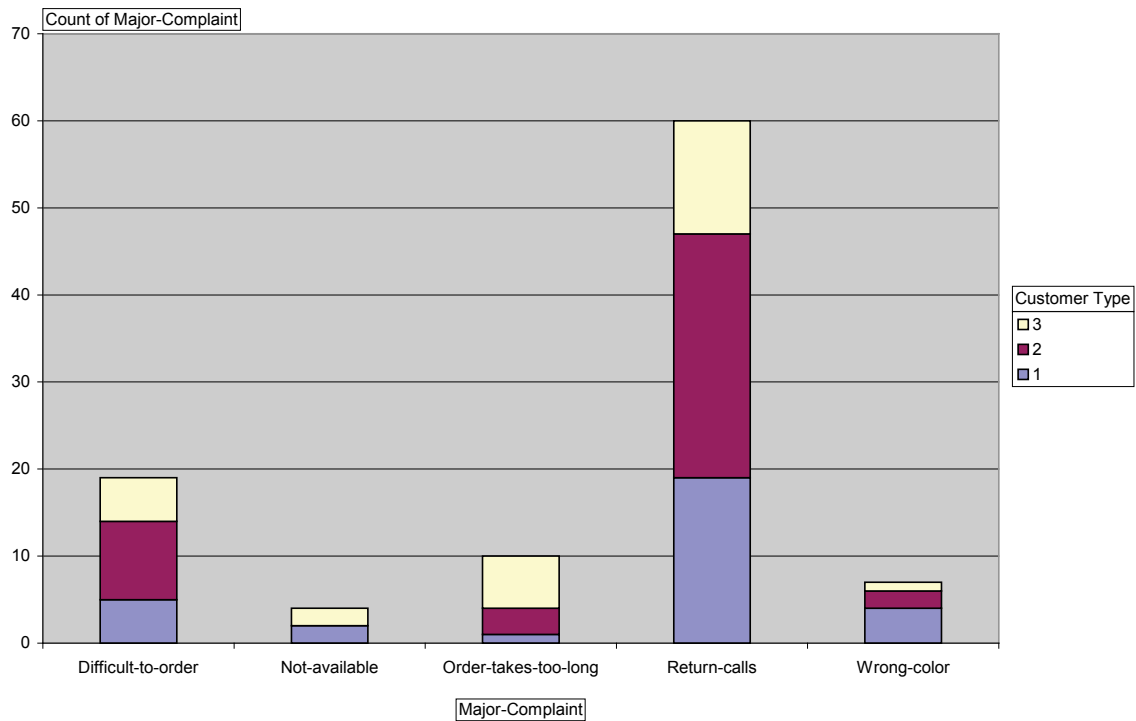
8. Click **OK**. Resulting Pivot Table is:

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

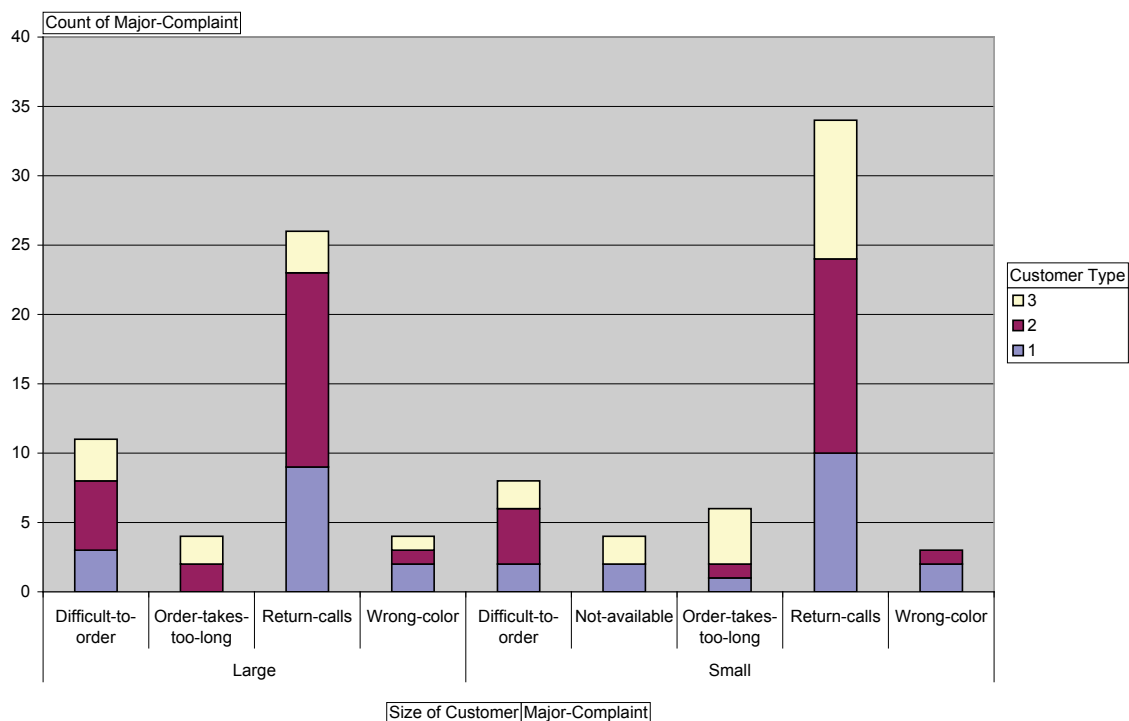
Note that the Major Complaint “Not-Available” is not shown. Pivot tables only show rows where there is at least a count of one.

9. The Pivot Chart can be seen by clicking the **EZ Pivot Chart (1)** tab; reset **Size of Customer** to **All** as shown below:

Size of Customer (All)

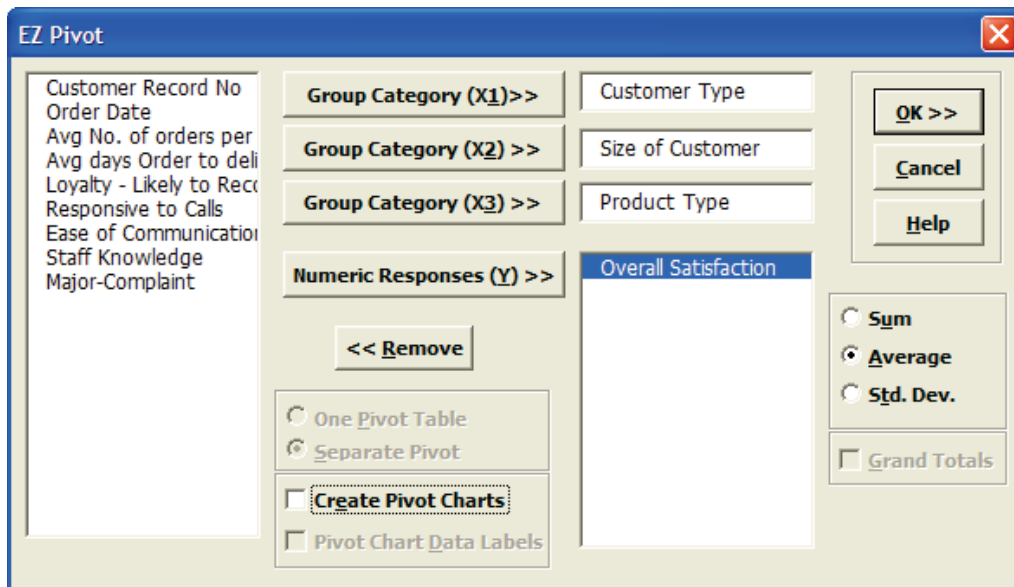


10. Drag **Size of Customer** from **Report Filter** to **Axis Fields** (shown under **PivotChart Fields** to the right of the chart) and Excel will automatically split the Pivot Chart showing both Large and Small Customers.



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**).



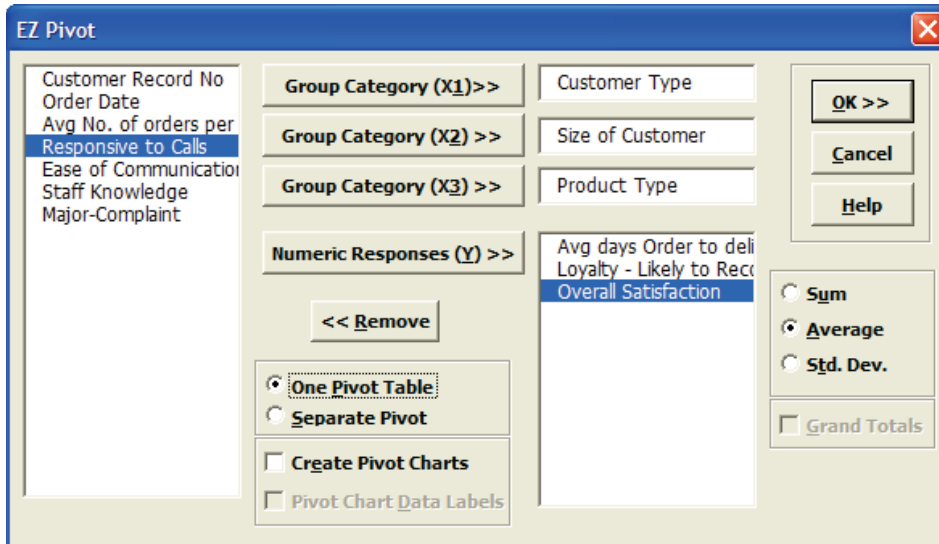
4. Click **OK**. The resulting Pivot Table is:

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.

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

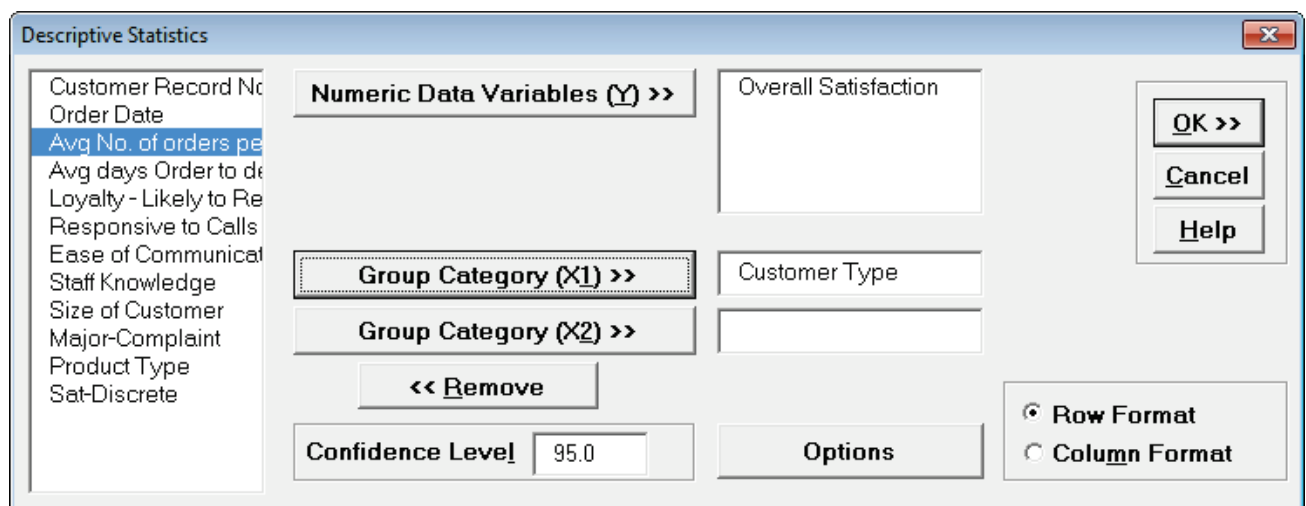
1 Poisson Rate Test and Confidence Interval.

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).

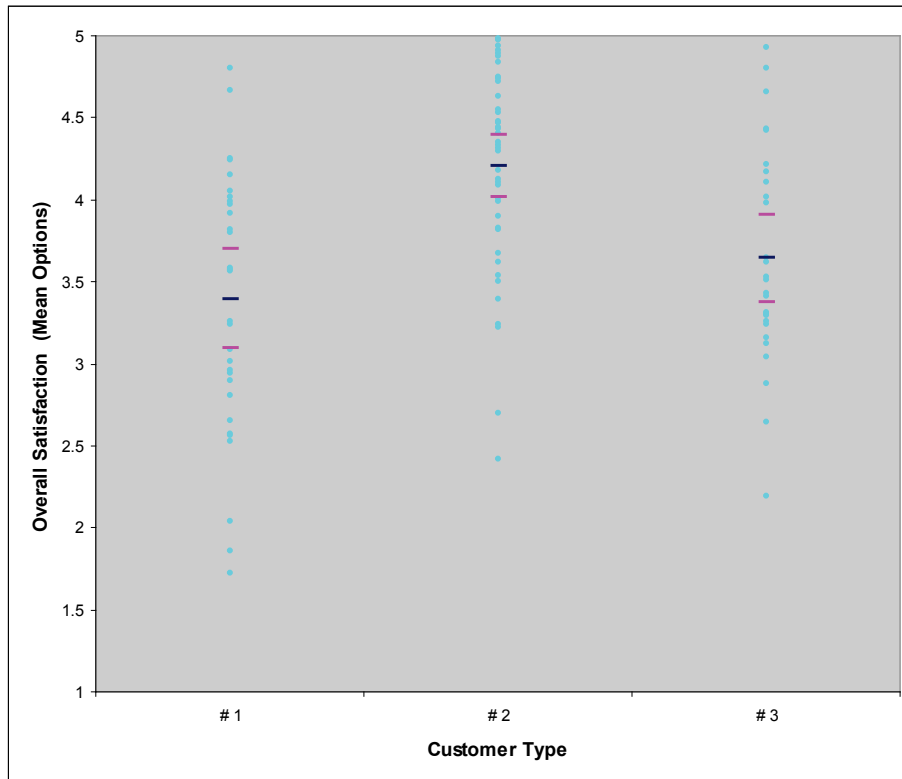
These confidence intervals are very important in understanding our data and making decisions from the data. How often are we driven by sample estimates only and fail to consider the confidence interval or margin of error? For example, newspapers will often fail to take into account the confidence interval 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, and Multi-Vari Charts.

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 now be used to compare the mean satisfaction scores more precisely and determine statistical significance for the results.

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 $H_0: \mu=3.5$; Alternative Hypothesis $H_a: \mu \neq 3.5$
4. Enter 3.5 for the **Null Hypothesis H_0 : Mean** value. Set **H_a** as *Not Equal To*, **Confidence Level** = 95.0%, and check **Display Test Assumptions Report**.

1 Sample t-Test

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ **Stacked Column Format (1 Numeric Data Column & 1 Optional Group Category)**
☐ **Unstacked Column Format (1 or more Numeric Data Columns)**

Numeric Data Variable (Y) >> Overall Satisfaction

Optional Group Category (X) >> Customer Type

<< Remove

H0: Mean = 3.5

Ha: Not Equal To

Confidence Level: 95.0

☒ **Display Test Assumptions Report**

OK >>
Cancel
Help

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 \geq .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 ($> Q3 + 1.5 \cdot IQR$ or $< Q1 - 1.5 \cdot IQR$); Likely: Tukey's Boxplot $2.2 \cdot IQR$; Extreme: Tukey's Boxplot $3 \cdot IQR$. 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 $H_0: \mu=3.5$ ”. A failure to reject H_0 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 $H_a: \mu \neq 3.5$ was true, we would have rejected H_0 and accepted H_a ?” 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 $\geq .99$ (Beta Risk is $\leq .01$) is considered Very High Power

Power $\geq .95$ and $< .99$ (Beta Risk is $\leq .05$) is High Power

Power $\geq .8$ and $< .95$ (Beta Risk is $\leq .2$) 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 $\geq .5$ and $< .8$ (Beta Risk is $\leq .5$) 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 **H_a : 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 **H_a** at the default values as shown:

Power and Sample Size: 1 Sample t-Test Calculator

Solve For

Power (1-Beta) ☐

Sample Size (N) ☐ 27

Difference (Mean - Ref) ☐ 0.1411

Standard Deviation .6705

Significance Level (Alpha) 0.05

Ha: Not Equal To

OK >> Cancel Help

3. Click **OK**. The resulting report is shown:

Power and Sample Size: 1 Sample t Test				
H0: Mean (μ) = Reference				
Ha: Mean (μ) \neq 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:

Power and Sample Size: 1 Sample t-Test Calculator

Solve For

Power (1-Beta) 0.9 ☐

Sample Size (N) 27 ☒

Difference (Mean - Ref) .1411 ☐

Standard Deviation .6705

Significance Level (Alpha) 0.05

Ha: Not Equal To

OK >> Cancel Help

6. Click **OK**. The resulting report is shown:

Power and Sample Size: 1 Sample t Test					
H0: Mean (μ) = Reference					
Ha: Mean (μ) \neq 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:

Power and Sample Size: 1 Sample t-Test Calculator

Solve For

Power (1-Beta) 0.9 ☐

Sample Size (N) 27 ☐

Difference (Mean - Ref) .1411 ☒

Standard Deviation .6705

Significance Level (Alpha) 0.05

Ha: Not Equal To

OK >> Cancel Help

9. Click **OK**. The resulting report is shown:

Power and Sample Size: 1 Sample t Test				
H0: Mean (μ) = Reference				
Ha: Mean (μ) \neq 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 ($1 - \text{Beta}$), 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 ($1 - \text{Beta}$)** 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:

The dialog box is titled "Power and Sample Size: 1 Sample t-Test". It features a "Select Column or Enter Value" section on the left and a "Solve For" section on the right. In the "Solve For" section, "Power (1-Beta) >>" is selected with a radio button. In the "Select Column or Enter Value" section, "Sample Size (N) >>" and "Difference (Mean - Ref) >>" are selected. Below these, the "Standard Deviation >>" field contains the value "1", and the "Significance Level (Alpha) >>" field contains ".05". The "Ha:" dropdown menu is set to "Not Equal To". At the bottom left is a "<< Remove" button. On the right side, there are three buttons: "OK >>", "Cancel", and "Help".

3. Click **OK**. The output report is shown below:

Power and Sample Size: 1 Sample t Test				
H0: Mean (μ) = Reference				
Ha: Mean (μ) \neq 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

4. 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**.
5. 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*:

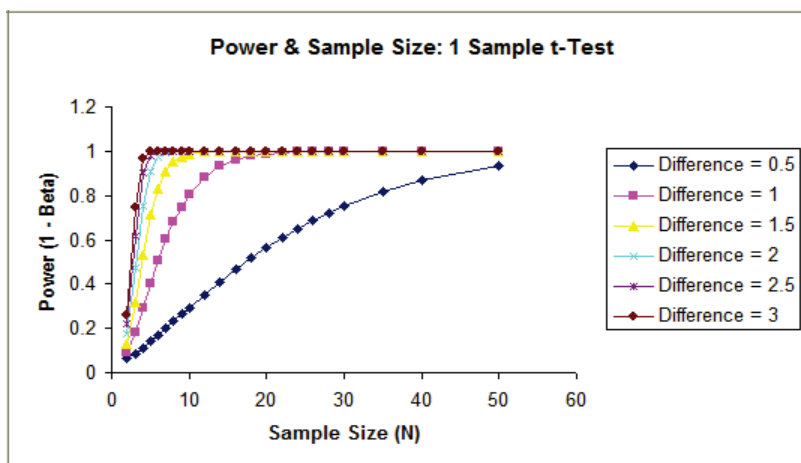
Power & Sample Size Chart

Standard Deviation: 1
Significance Level (Alpha): 0.05

Y Axis (Y): Power (1 - Beta)
X Axis (X1): Sample Size (N)
Group Category (X2): Difference

Buttons: OK >>, Cancel, Help, << Remove, Modify Title (checked)

6. Click **OK**. Click **OK**. The resulting Power & Sample Size Chart is displayed:



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: $N \leq 50$

Wilcoxon Signed Rank: $N \leq 15$

Mann-Whitney: Each sample $N \leq 10$

Kruskal-Wallis: Each sample $N \leq 5$

Mood's Median: Each sample $N \leq 10$

Runs Test (Above/Below) or Runs Test (Up/Down) Test: $N \leq 50$

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.

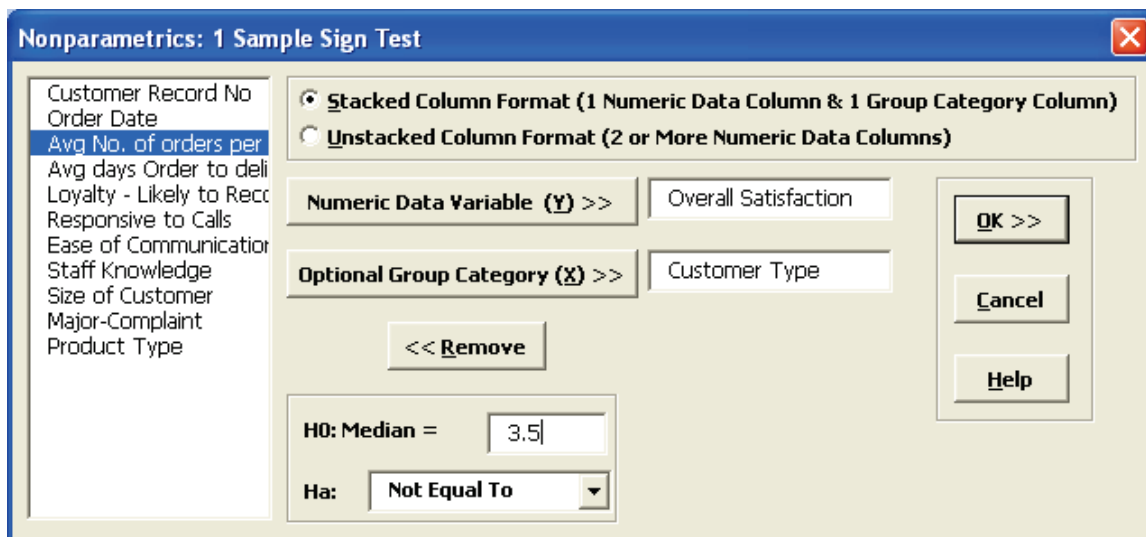
For further details and references refer to the Appendix [Exact and Monte Carlo P-Values for Nonparametric and Contingency Tests.](#)

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 (H_0 : Median=3.5, H_a : Median \neq 3.5, $\alpha = 0.05$).

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 H_0 : Median** value. Set **H_a** 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.

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, H_a: 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 **H_a** as *Not Equal To*.

1 Sample Wilcoxon

Customer Record No
Order Date
Avg No. of orders per
Avg days Order to deli
Loyalty - Likely to Recd
Responsive to Calls
Ease of Communication
Staff Knowledge
Size of Customer
Major-Complaint
Product Type

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Overall Satisfaction

Optional Group Category (X) >> Customer Type

<< Remove

H0: Median = 3.5

Ha: Not Equal To

OK >>
Cancel
Help

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 ($N \leq 15$ 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:

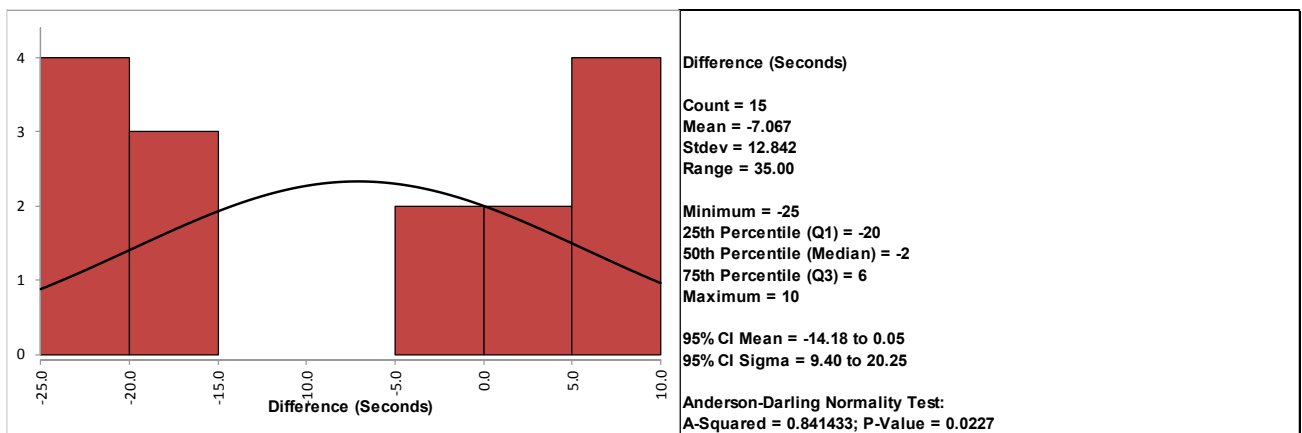
1 Sample Wilcoxon Test - Exact: Overall Satisfaction				
Test Information				
H ₀ : Median = 3.5				
H _a : Median Not Equal To 3.5				
Customer Type	1	2	3	
Count (N)	31	42	27	
Count for Test	31	41	27	
Median	3.56	4.34	3.51	
Wilcoxon Statistic	217.50	802.50	222.00	
Exact P-Value (2-sided)	0.5584	0.0000	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 $H_0: \mu = 0$; and the Alternative Hypothesis $H_a: \mu < 0$. Enter 0 for the **H0: Mean** value. Select **Ha** as *Less Than*, **Confidence Level** = 95.0%, and check **Display Test Assumptions Report**.

The image shows the '1 Sample t-Test' dialog box. The 'Stacked Column Format (1 Numeric Data Column & 1 Optional Group Category)' radio button is selected. The 'Numeric Data Variable (Y) >>' field contains 'Difference (Seconds)'. The 'Optional Group Category (X) >>' field is empty. The 'H0: Mean =' field contains '0'. The 'Ha:' dropdown menu is set to 'Less Than'. The 'Confidence Level:' field contains '95.0'. The 'Display Test Assumptions Report' checkbox is checked. Buttons for 'OK >>', 'Cancel', and 'Help' are on the right. A '<< Remove' button is below the variable fields.

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 ($N \leq 15$).

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*.

1 Sample Wilcoxon

☒ Stacked Column Format (1 Numeric Data Column & 1 Optional Group Category)
☐ Unstacked Column Format (1 or more Numeric Data Columns)

Numeric Data Variable (Y) >> Difference (Seconds)

Optional Group Category (X) >>
 << Remove

H0: Median = 0
 Ha: Less Than

OK >>
 Cancel
 Help

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 \(Compare 2 Means\)](#)

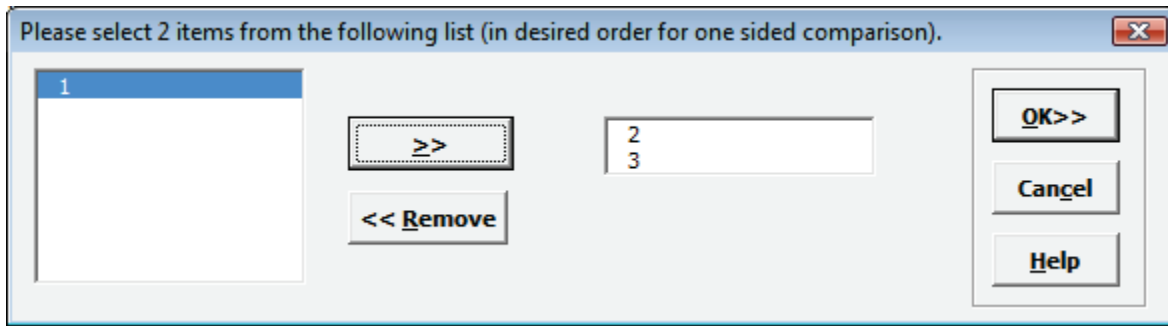
[Basic Statistical Templates – 2 Sample Equivalence Test \(Compare 2 Means\)](#)

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. $H_0: \mu_2 = \mu_3$, $H_a: \mu_2 \neq \mu_3$
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:

The screenshot shows the '2 Sample t-Test' dialog box. On the left is a list of variables: Customer Record No, Order Date, Avg No. of orders per mo (highlighted), Avg days Order to delivery, Loyalty - Likely to Recomm, Responsive to Calls, Ease of Communications, Staff Knowledge, Size of Customer, Major-Complaint, Product Type, and Sat-Discrete. The main area has two radio buttons: 'Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)' (selected) and 'Unstacked Column Format (2 Numeric Data Columns)'. Below these are two input fields: 'Numeric Data Variable (Y) >>' containing 'Overall Satisfaction' and 'Group Category (X) >>' containing 'Customer Type'. There is a '<< Remove' button. Below the input fields are three more fields: 'H0: Mean Diff =' with a value of '0', 'Ha:' with a dropdown menu showing 'Not Equal To', and 'Confidence Level:' with a value of '95.0'. At the bottom, there are two checked checkboxes: 'Assume Equal Variances' and 'Display Test Assumptions Report'. On the right side of the dialog are three buttons: 'OK >>', 'Cancel', and 'Help'.

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 ($> Q3 + 1.5 \cdot IQR$ or $< Q1 - 1.5 \cdot IQR$); Likely: Tukey's Boxplot $2.2 \cdot IQR$; Extreme: Tukey's Boxplot $3 \cdot IQR$. 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.

The Paired t-Test dialog box shows the following configuration:

- Variable List:** TASTER, Difference (selected)
- Numeric Data Variable 1 >>:** After
- Numeric Data Variable 2 >>:** Before
- H0: Mean Diff =:** 0
- Ha:** Less Than
- Confidence Level:** 95.0
- Display Test Assumptions Report:** ☒
- Buttons:** OK >>, Cancel, Help, << Remove

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:

The screenshot shows the 'Power and Sample Size: 2 Sample t-Test Calculator' dialog box. The 'Solve For' section has 'Power (1-Beta)' selected. The 'Sample Size (N) for Each Group' is 30, and the 'Difference (Mean1 - Mean2)' is 1. The 'Standard Deviation' is 1.0, and the 'Significance Level (Alpha)' is 0.05. The 'Ha' dropdown is set to 'Not Equal To'. Buttons for 'OK >>', 'Cancel', and 'Help' are on the right.

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.



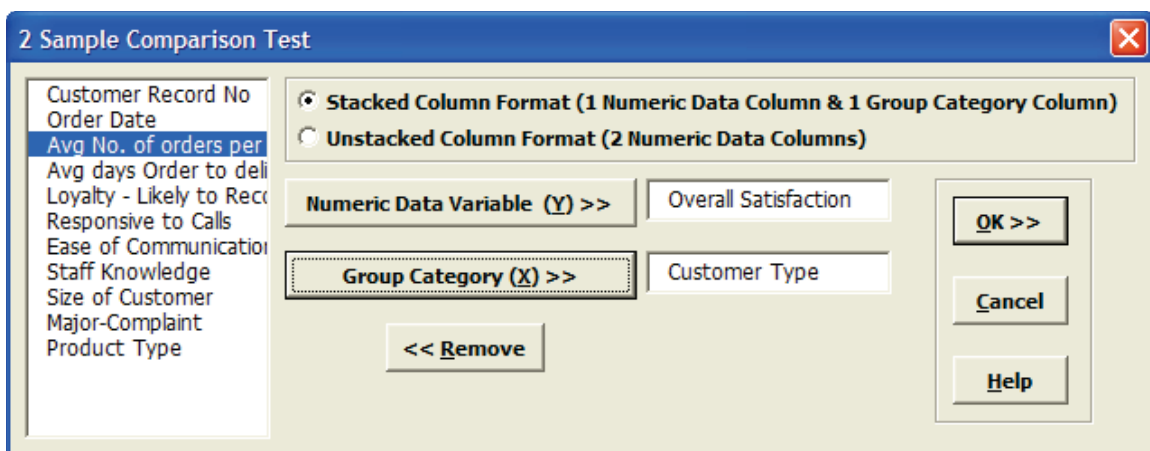
2 Sample F-Test and Ratio Confidence Interval (Compare 2 Standard Deviations)

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)	$H_0: \sigma_1/\sigma_2 =$	1.0	
Alternative Hypothesis	$H_a: \sigma_1/\sigma_2$	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	

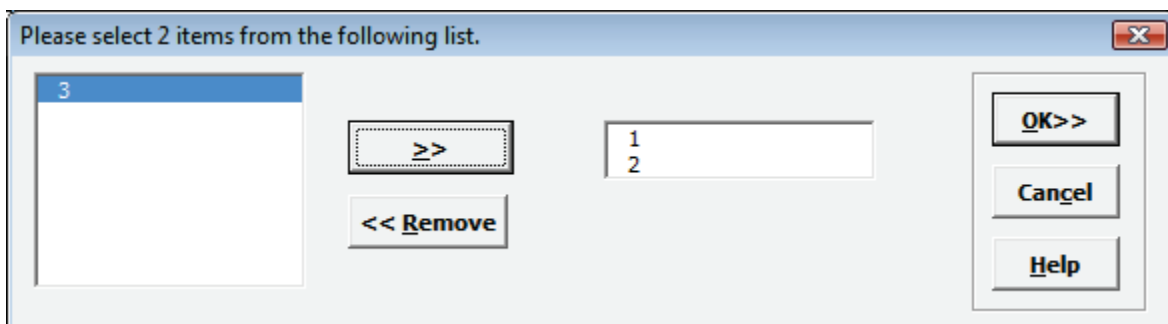
Two Sample Comparison Test

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: Median₁ = Median₂
Ha: Median₁ \neq Median₂

←

←

←

←

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.

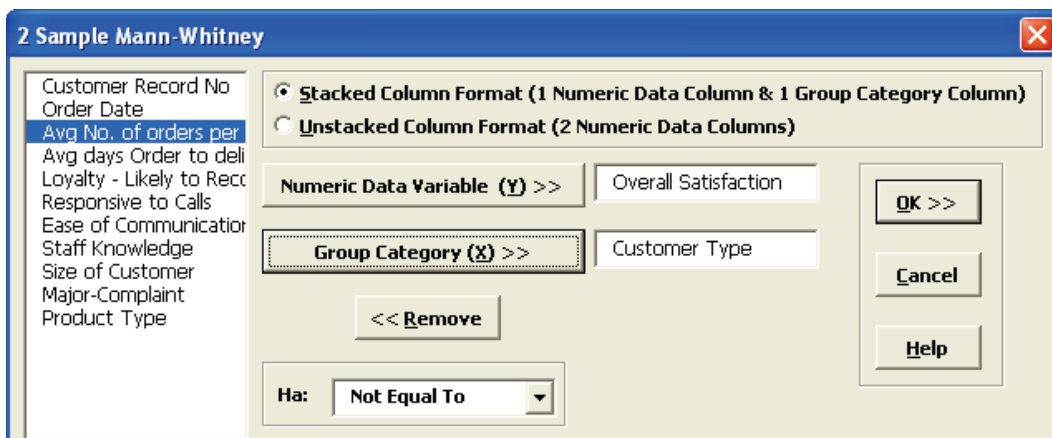
Clearly the next step would be to determine a root cause or best practices to reduce the variability in overall satisfaction and increase the mean for all customer types.

Part I – Two Sample Nonparametric Test: Mann-Whitney

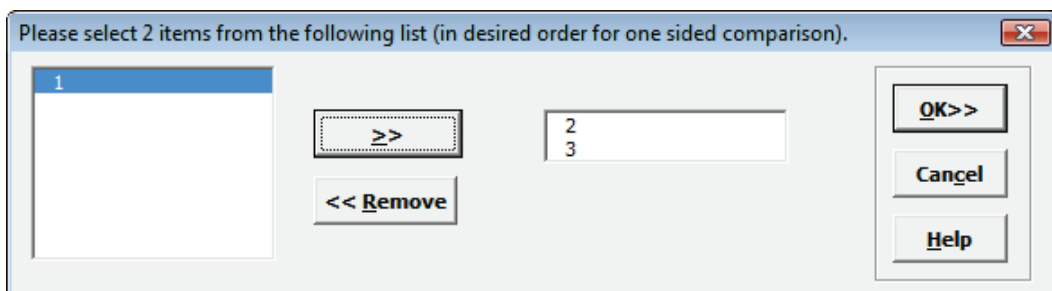
Two Sample Mann-Whitney Test

We will look at comparing medians of Customer Satisfaction by Customer Type, using the Two Sample Mann-Whitney test. (H_0 : Median2 = Median3, H_a : Median2 \neq Median3.) The Two Sample Mann-Whitney Test is the nonparametric equivalent to the parametric Two Sample t-Test. 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 **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**:



4. Click **OK**. Select *Customer Type 2* and 3.



5. Click **OK**. Resulting output:

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.

Two Sample Mann-Whitney 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 $N \leq 10$ 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%**.

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.

4. Click **OK**. Select *Customer Type* 2 and 3. Click **OK**. Results:

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 Mann Whitney Statistic 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. 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.

6. 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**.
7. 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.

2 Sample Mann-Whitney - Exact

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 Numeric Data Columns)

Numeric Data Variable (Y) >> Reaction Time
 Group Category (X) >> Stimulant
 << Remove

Ha: Less Than

☒ Exact
 Time Limit for Exact Computation: 60 (Seconds)
☐ Monte-Carlo Exact
 Number of Replications: 10000
 Confidence Level for P-Value: 99 %

OK >> Cancel Help

8. Click **OK**. Select *Stimulant 1* and *2*.

Select 2 items from the list (in desired order for one sided comparison).

>> 1
 << Remove 2

OK >> Cancel Help

This sets the order for the one sided test, so the alternative hypothesis H_a is Median 1 < Median 2.

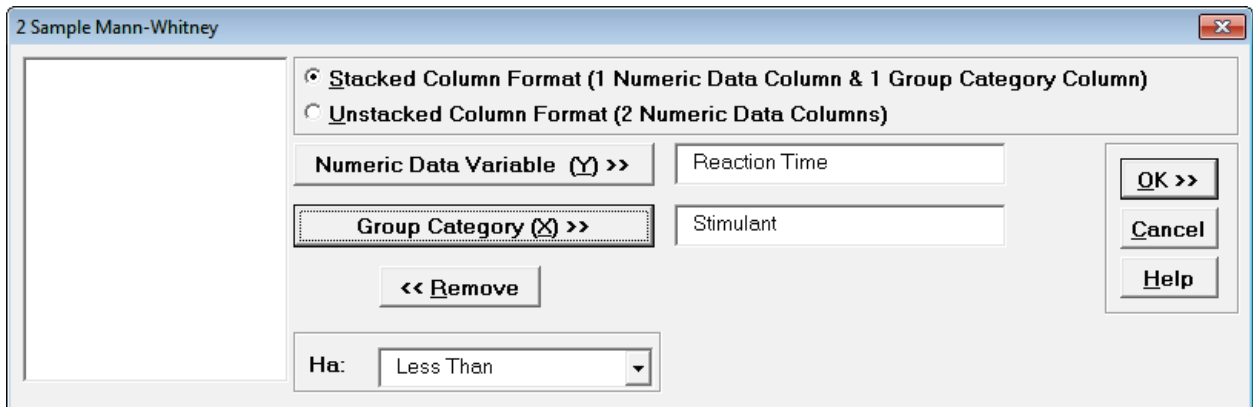
9. Click OK. Results:

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 H_0 , 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.

10. 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**.
11. With **Stacked Column Format** checked, select *Reaction Time*, click **Numeric Data Variable (Y) >>**; select *Stimulant*, click **Group Category (X) >>**; and **Ha: Less Than**.



12. 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** H_0 .

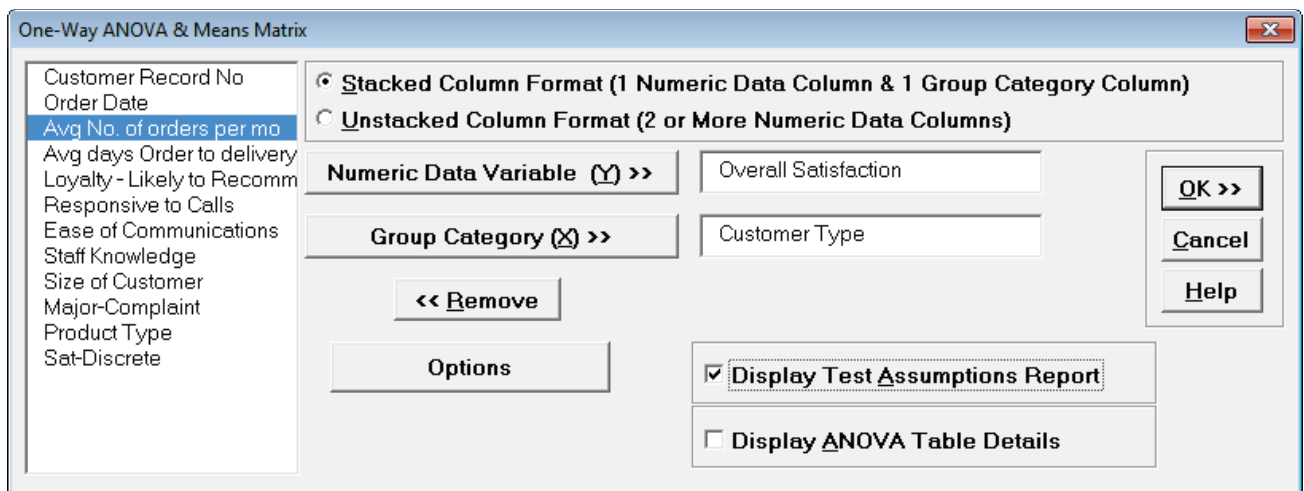
The difference between exact and large sample P-Value is small but it was enough to lead us to falsely conclude that stimulant 1 resulted in a reduced median reaction time.

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 $H_0: \mu_1 = \mu_2 = \mu_3$; H_a : 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**.

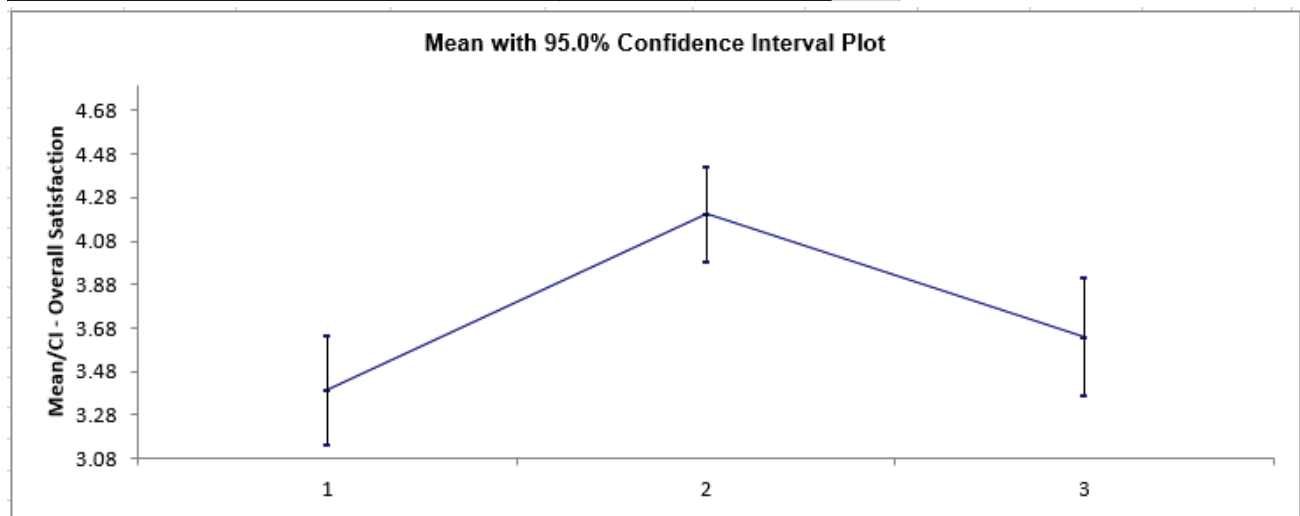


5. Click **OK**. The results are shown below:

One-Way ANOVA & Means Matrix: Overall Satisfaction

H₀: Mean 1 = Mean 2 = ... = Mean kH_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:

$$\alpha = (100 - \text{Confidence Level})/100$$

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 ($> Q3 + 1.5 \cdot IQR$ or $< Q1 - 1.5 \cdot IQR$); Likely: Tukey's Boxplot $2.2 \cdot IQR$; Extreme: Tukey's Boxplot $3 \cdot IQR$. 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 $\alpha = (100 - \text{Confidence Level})/100$.
 - This chart is also available at **SigmaXL > Graphical Tools > Analysis of Means (ANOM)**. See [Part P – Analysis of Means \(ANOM\) Charts](#).

- 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:

One-Way ANOVA & Means Matrix

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or More Numeric Data Columns)

Numeric Data Variable (Y) >> Overall Satisfaction

Group Category (X) >> Customer Type

<< Remove

Options

Means Matrix Probability Method (Multiple Comparison of Means)
☐ Fisher
☒ Tukey
☐ Dunnett

Confidence Level: 95.0

☐ Display Test Assumptions Report
☐ Display ANOVA Table Details
☒ Display Residual Charts
☒ Display ANOM One-Way Normal Chart

OK >> Cancel Help

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%.

- 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):

One-Way ANOVA & Means Matrix

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or More Numeric Data Columns)

Numeric Data Variable (Y) >> Overall Satisfaction

Group Category (X) >> Customer Type

<< Remove

Options

Means Matrix Probability Method (Multiple Comparison of Means)
☐ Fisher
☐ Tukey
☒ Dunnett
2

Confidence Level: 95.0

☐ Display Test Assumptions Report
☐ Display ANOVA Table Details
☐ Display Residual Charts
☐ Display ANOM One-Way Normal Chart

OK >>
Cancel
Help

5. Click **OK**. The Pairwise Means Difference (Means Matrix) and Dunnett Probability results are:

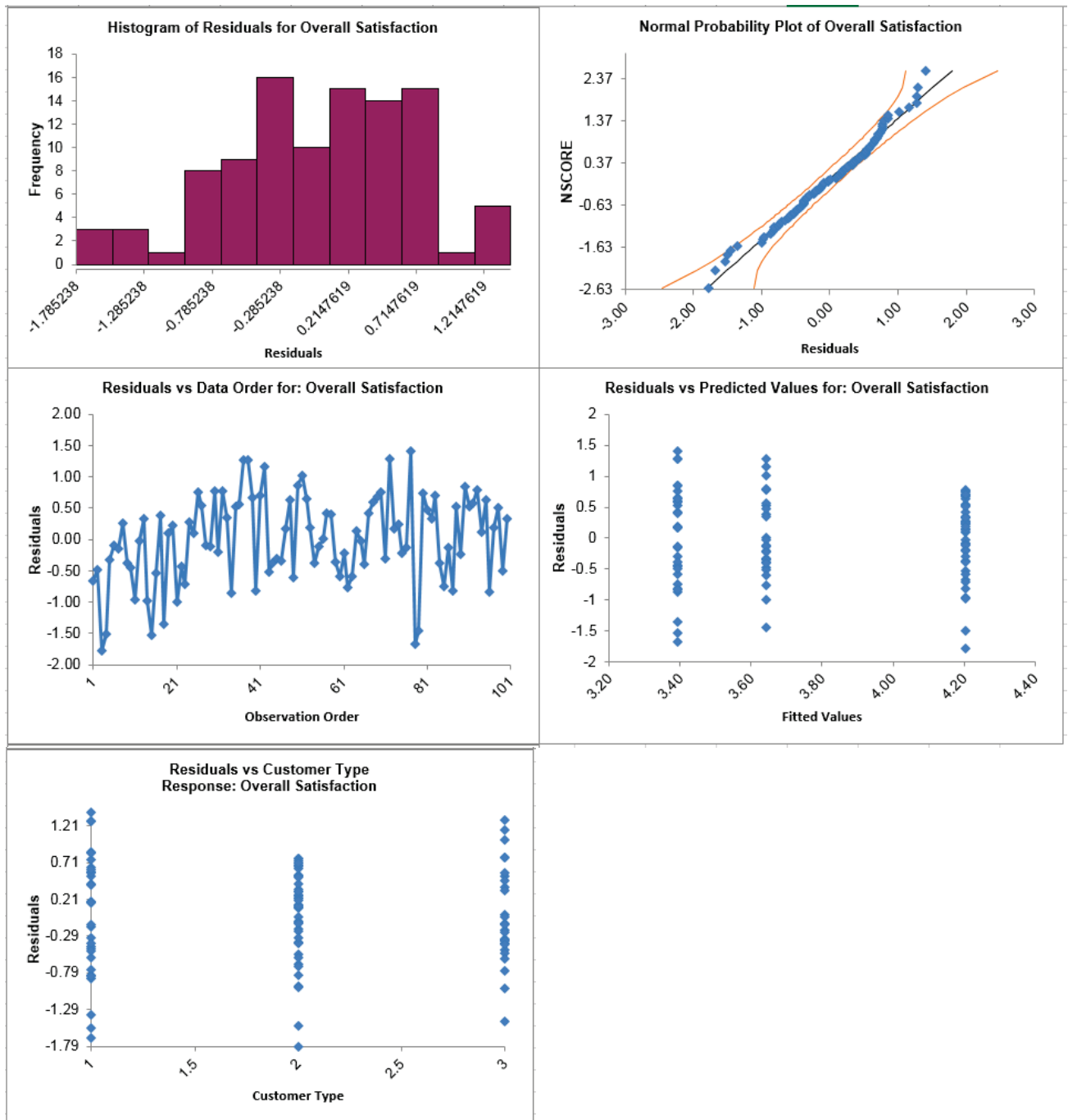
Pairwise Mean Difference (row - column)		1	2	3
1		0	-0.811690	-0.247563
2			0	0.564127
3				0

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).

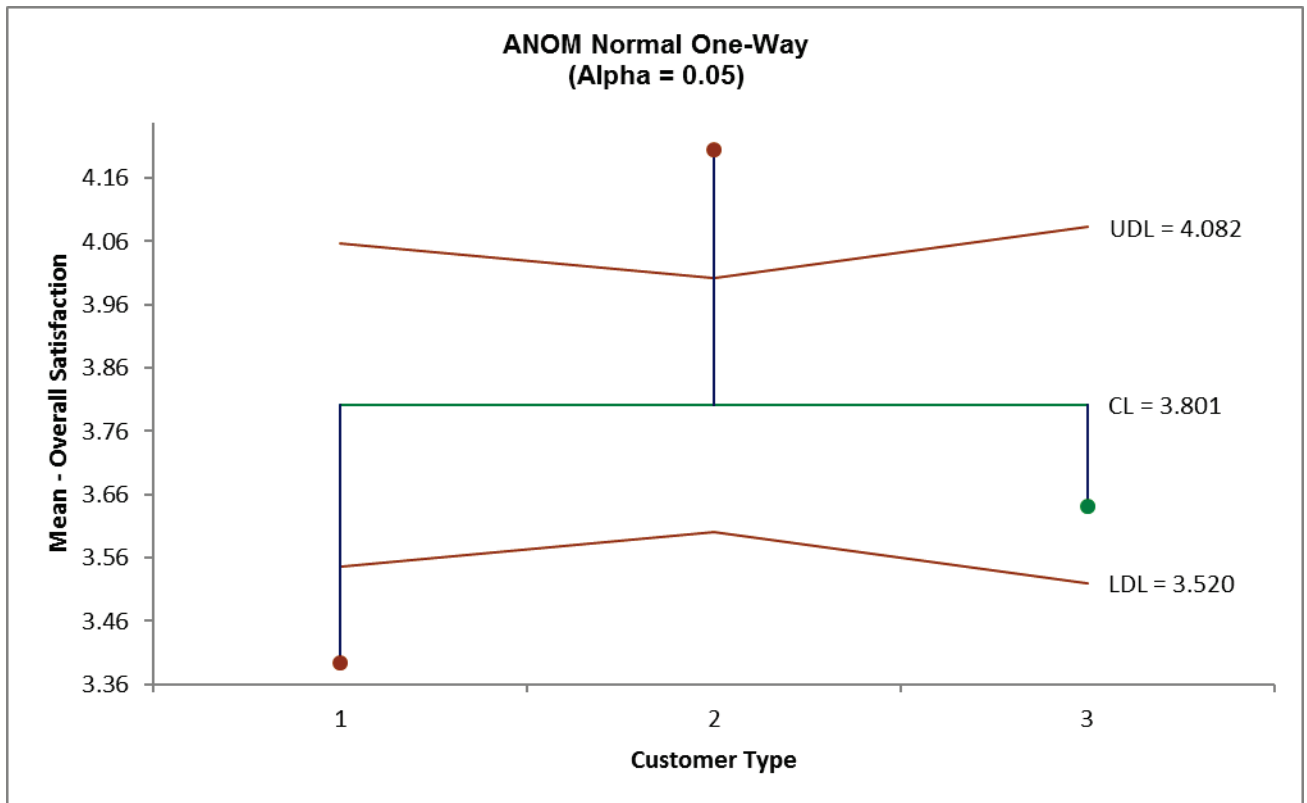
Note that the Dunnett 2 – 3 P-Value is .003 which is smaller than the Tukey .004 but larger than the Fisher .002. Dunnett is more powerful than Tukey (due to having fewer tests), but less powerful than Fisher. It does, however, have an advantage over Fisher because it protects the family-wise error rate 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 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:

Power and Sample Size: One-Way Anova Calculator

Solve For

Power (1-Beta) ☒

Sample Size (N) for Each Group 30 ☐

Maximum Difference 1 ☐

Groups (Levels) 3

Standard Deviation 1.0

Significance Level (Alpha) .05

OK >> **Cancel** **Help**

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 H_a : 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): $\mu_1 = \mu_2 = \dots = \mu_k$

Ha (Factor X1): at least one pairwise set of means are not equal ($\mu_i \neq \mu_j$);

H0 (Factor X2): $\mu_1 = \mu_2 = \dots = \mu_k$

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 $(100 - \text{Confidence Level})/100$.

Two-Way ANOVA

Customer Record No
Order Date
Avg days Order to de
Loyalty - Likely to Re
Overall Satisfaction
Responsive to Calls
Ease of Communicat
Staff Knowledge
Major-Complaint
Product Type
Sat-Discrete

Numeric Data Variable (Y) >> Avg No. of orders per mo

Group Category Factor (X1) >> Customer Type

Group Category Factor (X2) >> Size of Customer

<< Remove

☐ Remove Interaction (Fit Additive Model)

☒ Display Residual Plots

☐ Display ANOM Normal Two-Way Chart

☐ Adjust chart alpha for family-wise error

Confidence Level: 95.0

OK >>
Cancel
Help

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

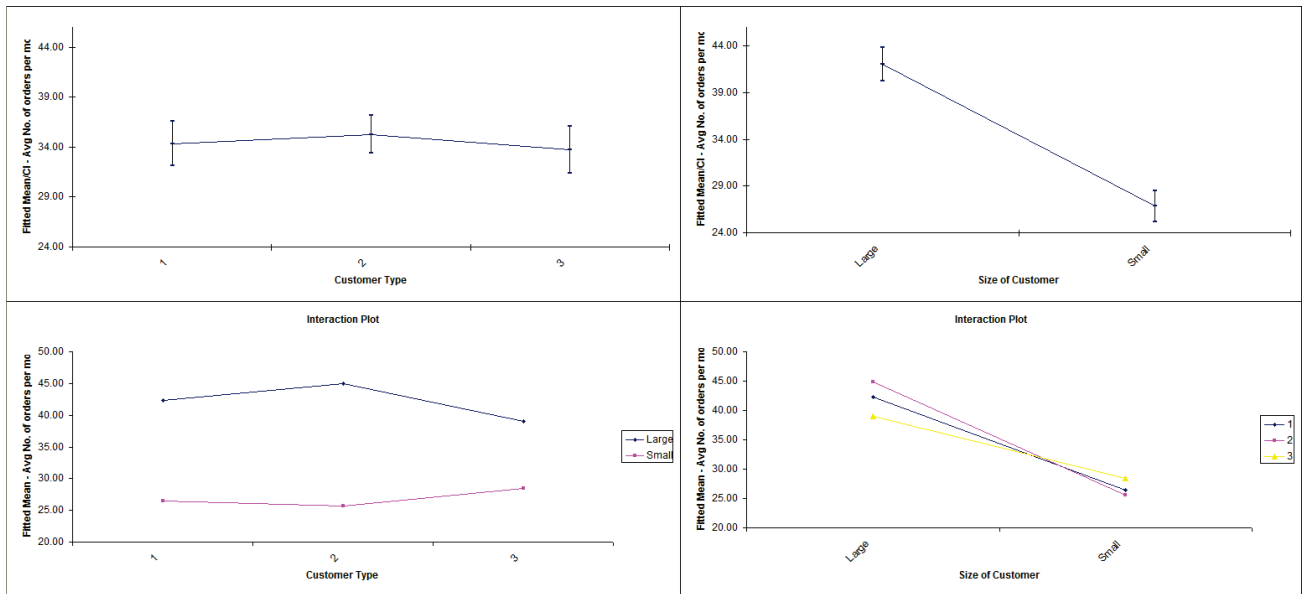
Factor and Model Summary:

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		

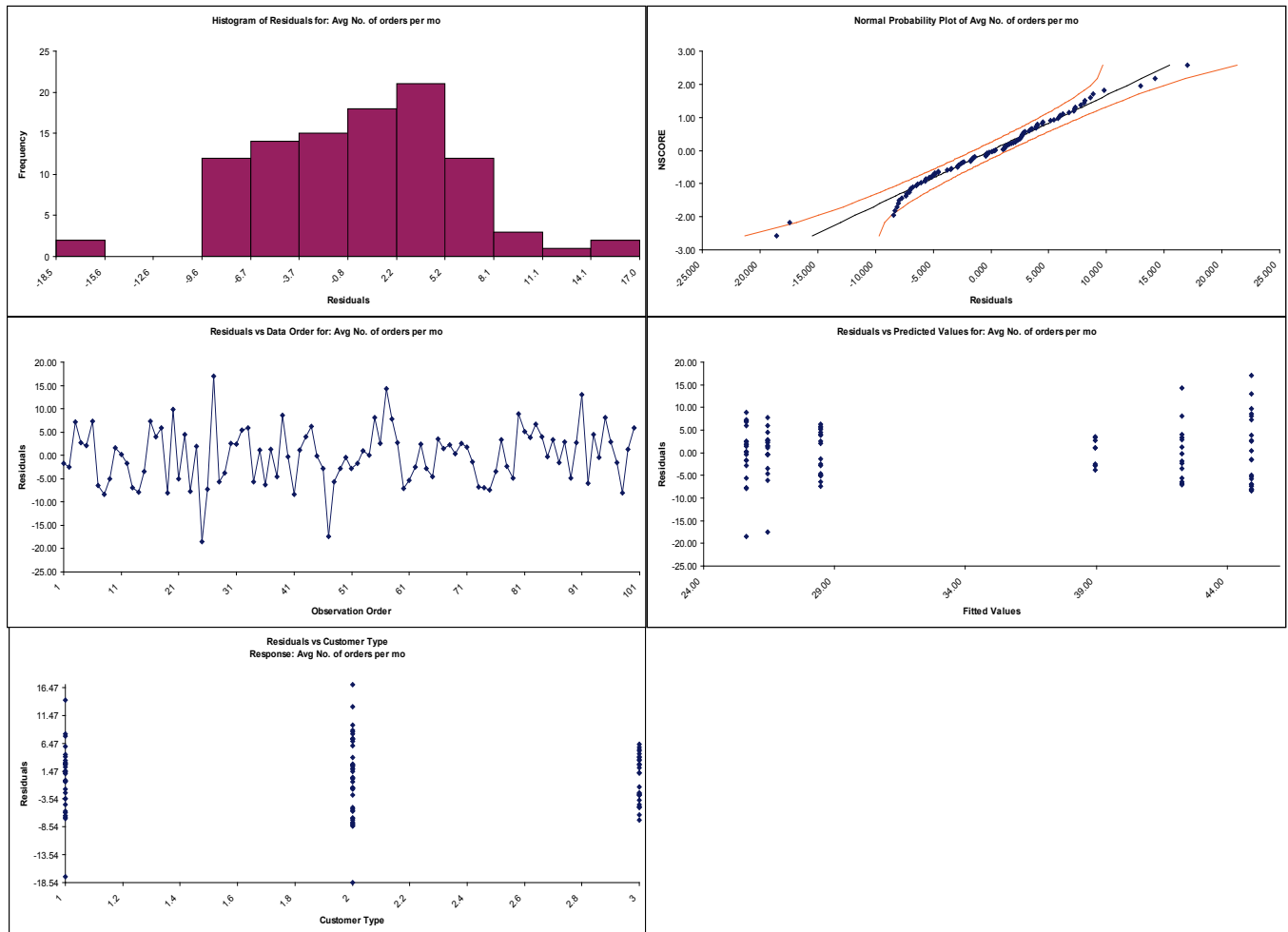
5. Scroll down to view the Main Effects (with confidence intervals) and Interaction Plots:



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

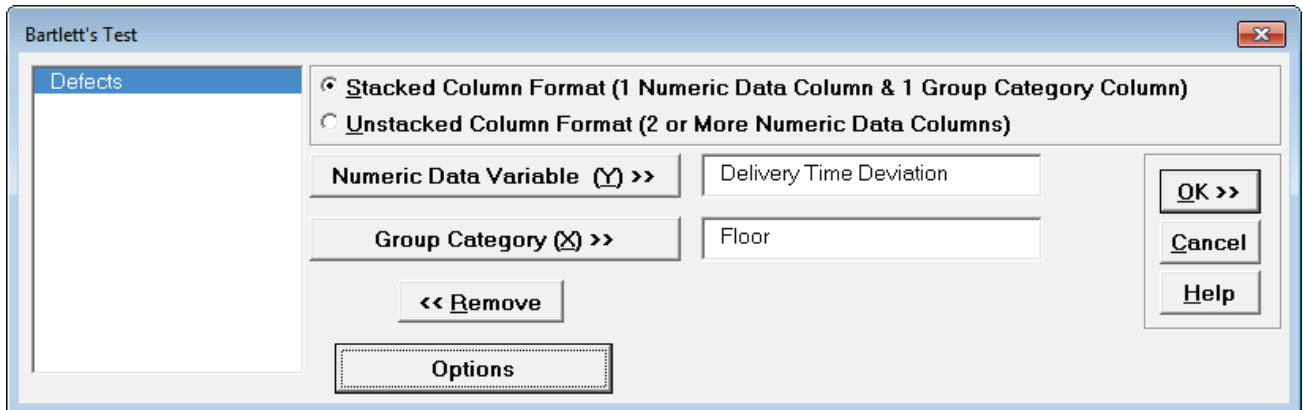
Bartlett's Test

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
 - $m = \text{number of pairwise comparisons } k(k-1)/2$
- 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 = $(100 - \text{Confidence Level})/100$.
 - 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										
H _a : 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**.

Bartlett's Test

Defects

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Delivery Time Deviation

Group Category (X) >> Floor

<< Remove

Options

Multiple Comparison of Variances Probability Method

☒ **F-Test Pairwise**
☐ **F-Test with Bonferroni Correction**

Confidence Level: 95.0

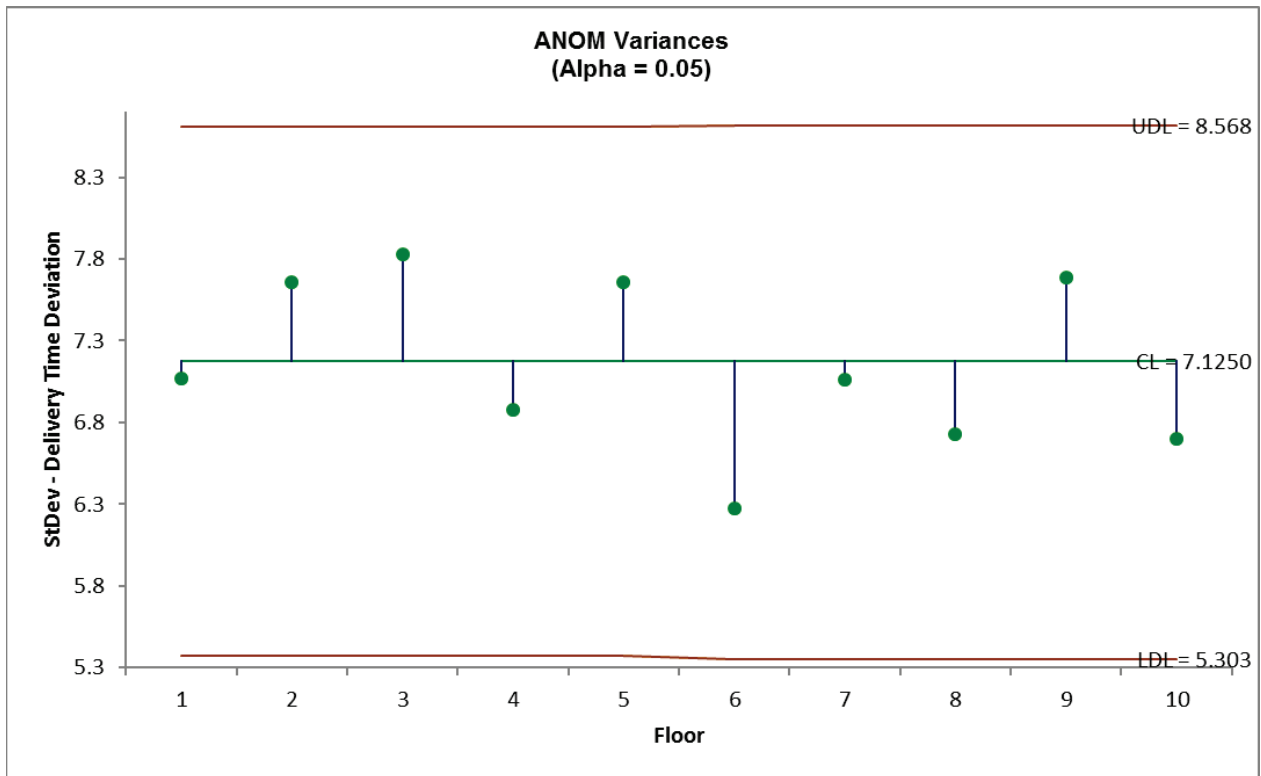
☒ **Display ANOM Variances Chart**

OK >> Cancel Help

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_Variiances** sheet tab to display the ANOM Chart:



The ANOM Variance 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 Variance 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 Variance chart in **SigmaXL > Graphical Tools > Analysis of Means (ANOM) > ANOM Variance** has an option to display Variances.

Levene's Test

Levene's Test for multiple group comparison of variances is less powerful than 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 $\alpha = (100 - \text{Confidence Level})/100$.
 - 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) >>**.

Levene's Test

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Overall Satisfaction
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Responsive to Calls

Group Category (X) >> Customer Type

<< Remove

Options

OK >>
Cancel
Help

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			
H _a : 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**.

Levene's Test

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Overall Satisfaction
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or More Numeric Data Columns)

Numeric Data Variable (Y) >> Responsive to Calls

Group Category (X) >> Customer Type

<< Remove

Options

Multiple Comparison of Variances
Probability Method

☐ Levene Pairwise
☒ Tukey ADM (Absolute Deviations from the Median)

Confidence Level: 95.0

☒ Display ANOM Levene Robust Variances Chart

OK >>
Cancel
Help

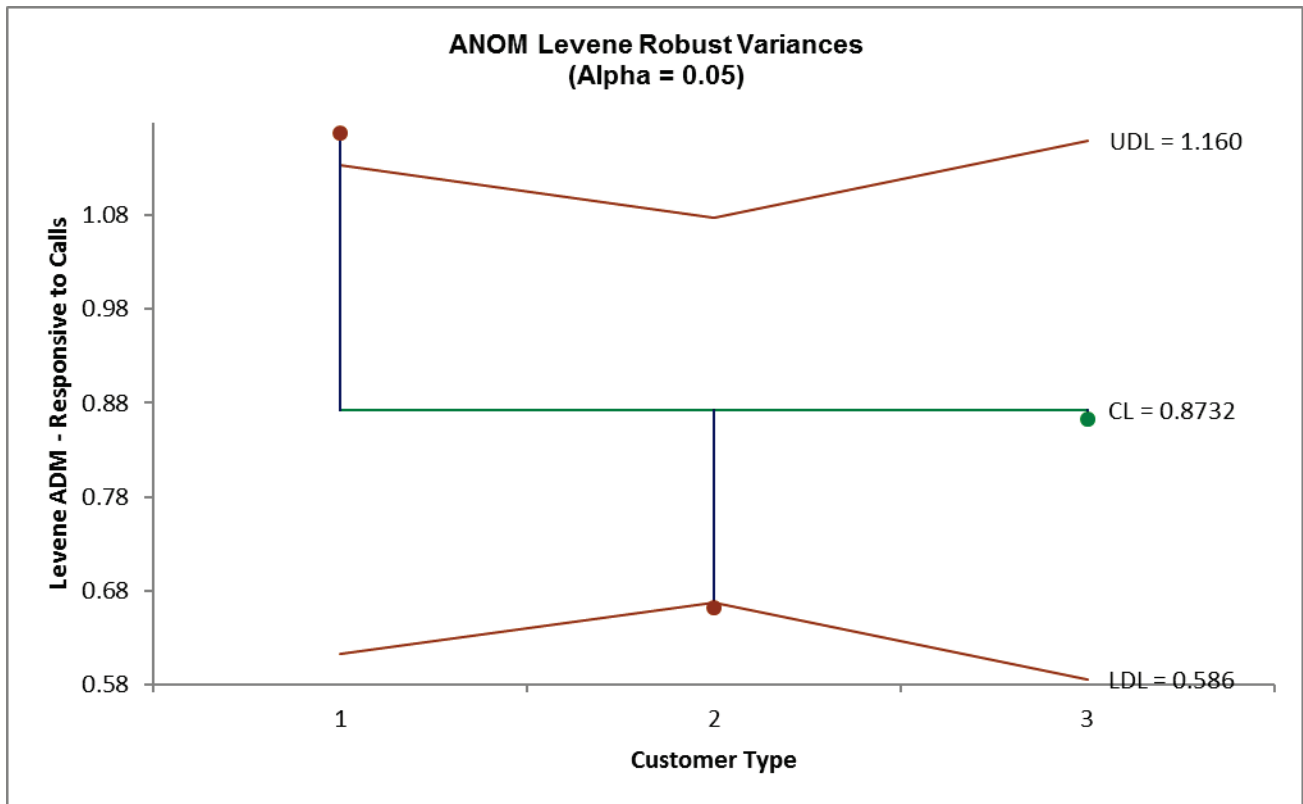
Note: The **Confidence Level** determines the alpha level ($\alpha = (100 - CI)/100$) used to highlight the P-Values and the alpha level for the ANOM chart. However, the alpha level used to highlight P-Values in the Anderson-Darling Normality Test is always 0.05.

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**.

Welch's ANOVA

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Overall Satisfaction
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Responsive to Calls

Group Category (X) >> Customer Type

<< Remove

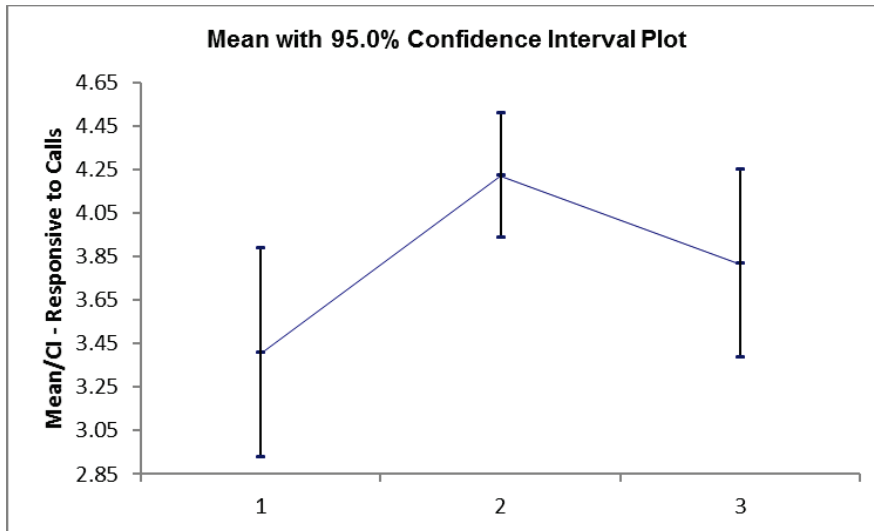
Options

☒ **Display Test Assumptions Report**

OK >>
Cancel
Help

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			
H _a : 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 H_0 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 ($> Q3 + 1.5 \cdot IQR$ or $< Q1 - 1.5 \cdot IQR$); Likely: Tukey's Boxplot $2.2 \cdot IQR$; Extreme: Tukey's Boxplot $3 \cdot IQR$. 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:

Welch's ANOVA

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Overall Satisfaction
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Responsive to Calls
Group Category (X) >> Customer Type

<< Remove

Options

☐ **Display Test Assumptions Report**

Means Matrix Probability Method (Multiple Comparison of Means)
☐ Welch Pairwise
☒ **Games-Howell**

Confidence Level: 95.0
☒ **Display Residual Plots**

OK >>
Cancel
Help

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:

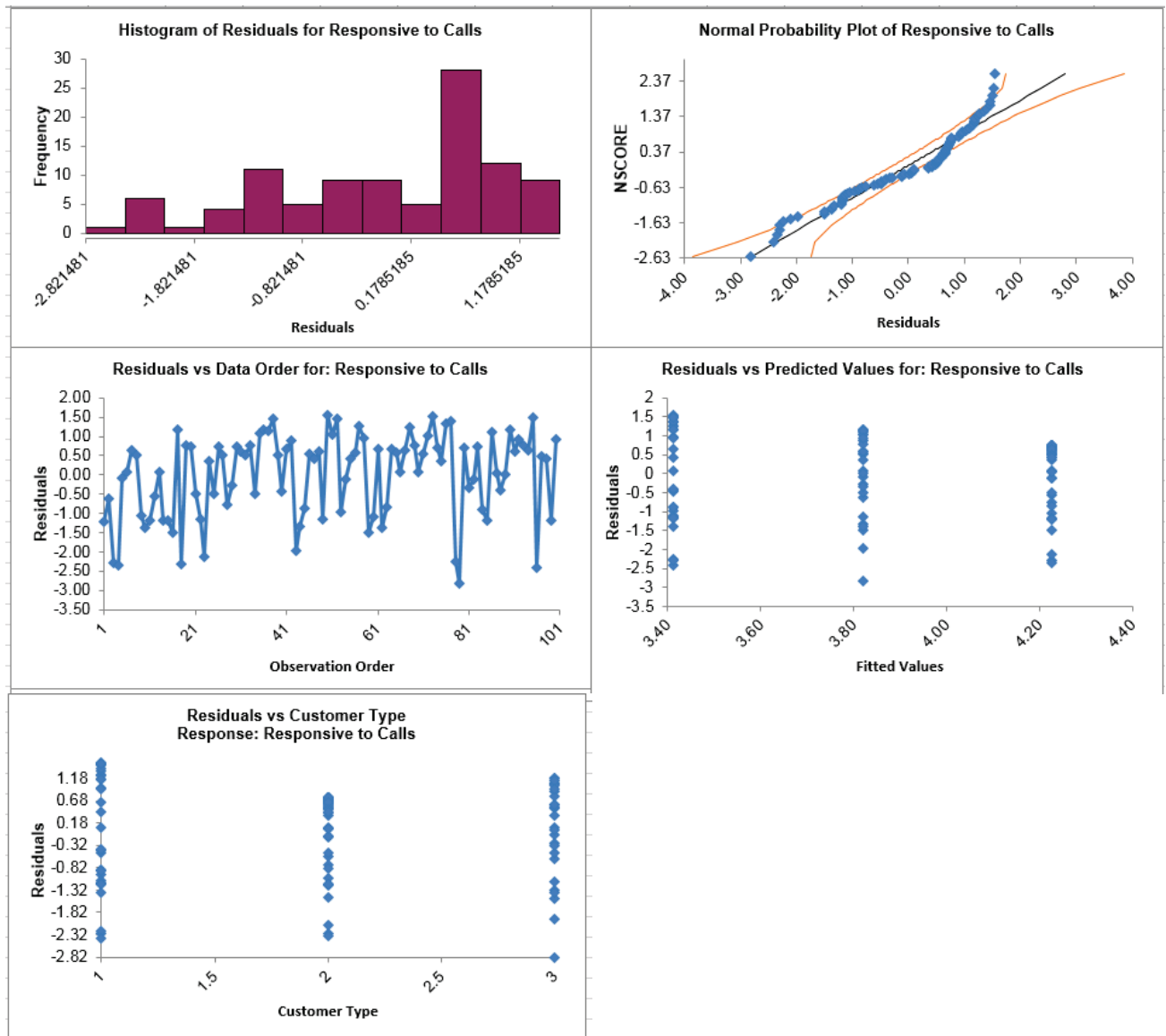
Pairwise Mean Difference (row - column)	1	2	3
1	0	-0.814578	-0.409869
2		0	0.404709
3			0
Games-Howell Probabilities	1	2	3
1		0.0123	0.3998
2			0.2572
3			

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:



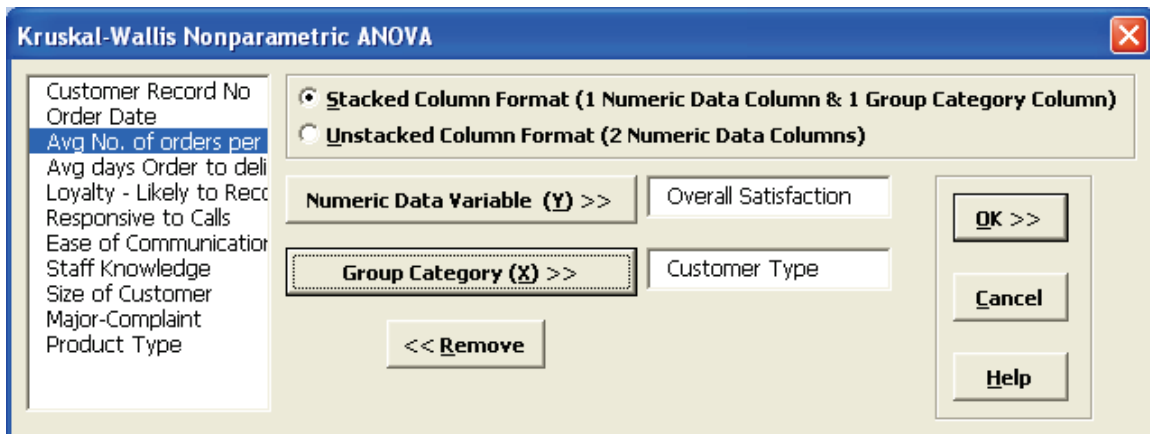
Residuals are the unexplained variation from the ANOVA model (Actual – Predicted or Fitted values). Note that the residuals are not normally distributed - as expected from the assumptions report - but like the regular ANOVA, Welch's ANOVA is quite robust to the assumption of normality. Also, as expected 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

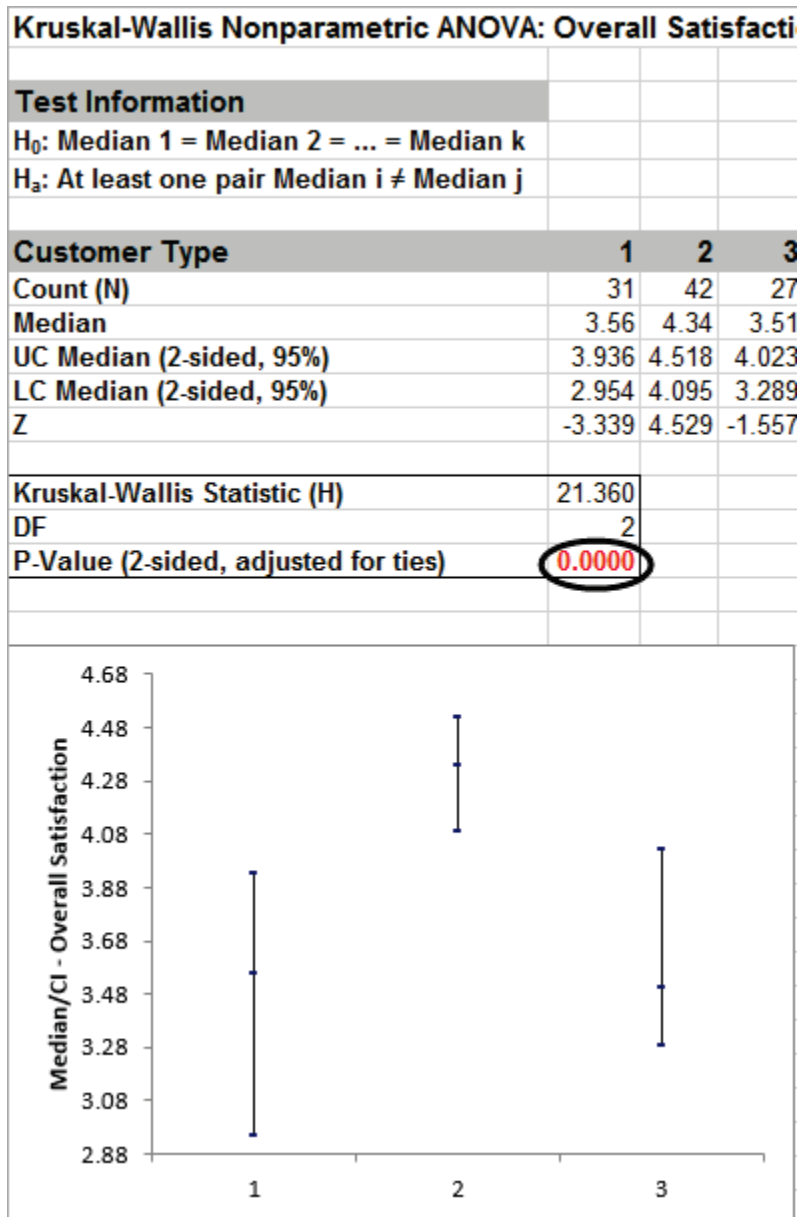
Kruskal-Wallis

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. The Null Hypothesis is: $H_0: \text{Median}_1 = \text{Median}_2 = \dots = \text{Median}_K$. H_a : 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:



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 (2.3×10^{-5}) tells us that we reject H_0 . 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 $N \leq 5$ 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:

$$(31+42+27)! / (31! * 42! * 27!) = 7.42 \text{ E}44$$

(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.

Kruskal-Wallis Nonparametric ANOVA - Exact

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or more Numeric Data Columns)

Numeric Data Variable (Y) >> Overall Satisfaction

Group Category (X) >> Customer Type

<< Remove

OK >>
Cancel
Help

☐ Exact
Time Limit for Exact Computation: 60 (Seconds)

☒ Monte-Carlo Exact
Number of Replications: 1e6
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**.

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 \neq 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 H_0 : 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 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.

Kruskal-Wallis Nonparametric ANOVA - Exact

☐ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☒ Unstacked Column Format (2 or more Numeric Data Columns)

Numeric Data Variables (Y) >>

 Device A
 Device B
 Device C

<< Remove

☒ Exact
 Time Limit for Exact Computation: 60 (Seconds)

☐ Monte-Carlo Exact
 Number of Replications: 1e6
 Confidence Level for P-Value: 99 %

OK >> Cancel Help

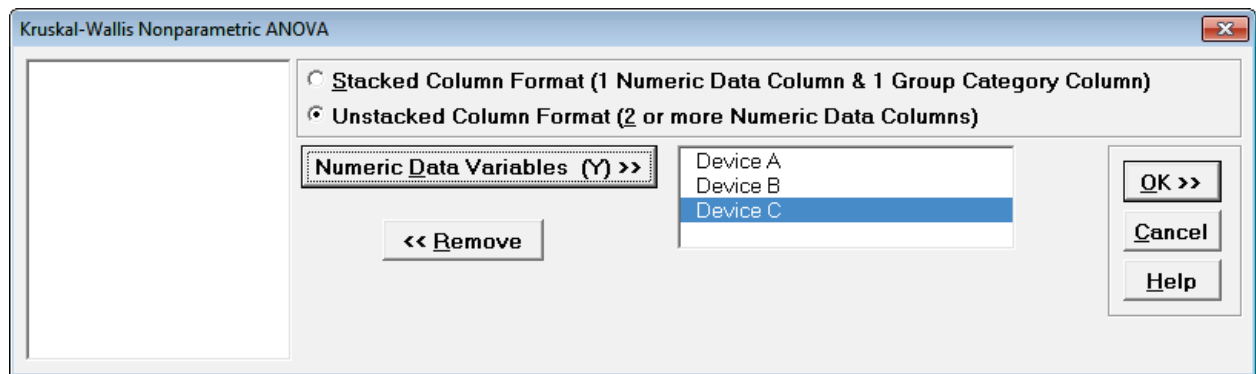
8. Click **OK**. Results:

Kruskal-Wallis Nonparametric ANOVA - Exact			
Test Information			
H_0 : Median 1 = Median 2 = ... = Median k			
H_a : At least one pair Median i \neq 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 H_0 , 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 \neq 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 H_0 (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.

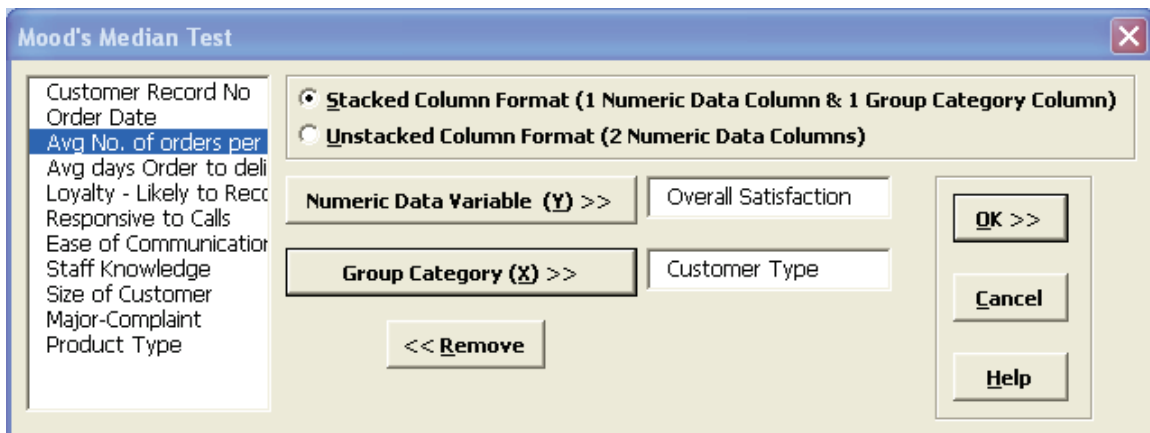
In conclusion, whenever you have a small sample size and are performing a nonparametric test, always use the Exact option.

Mood's Median Test

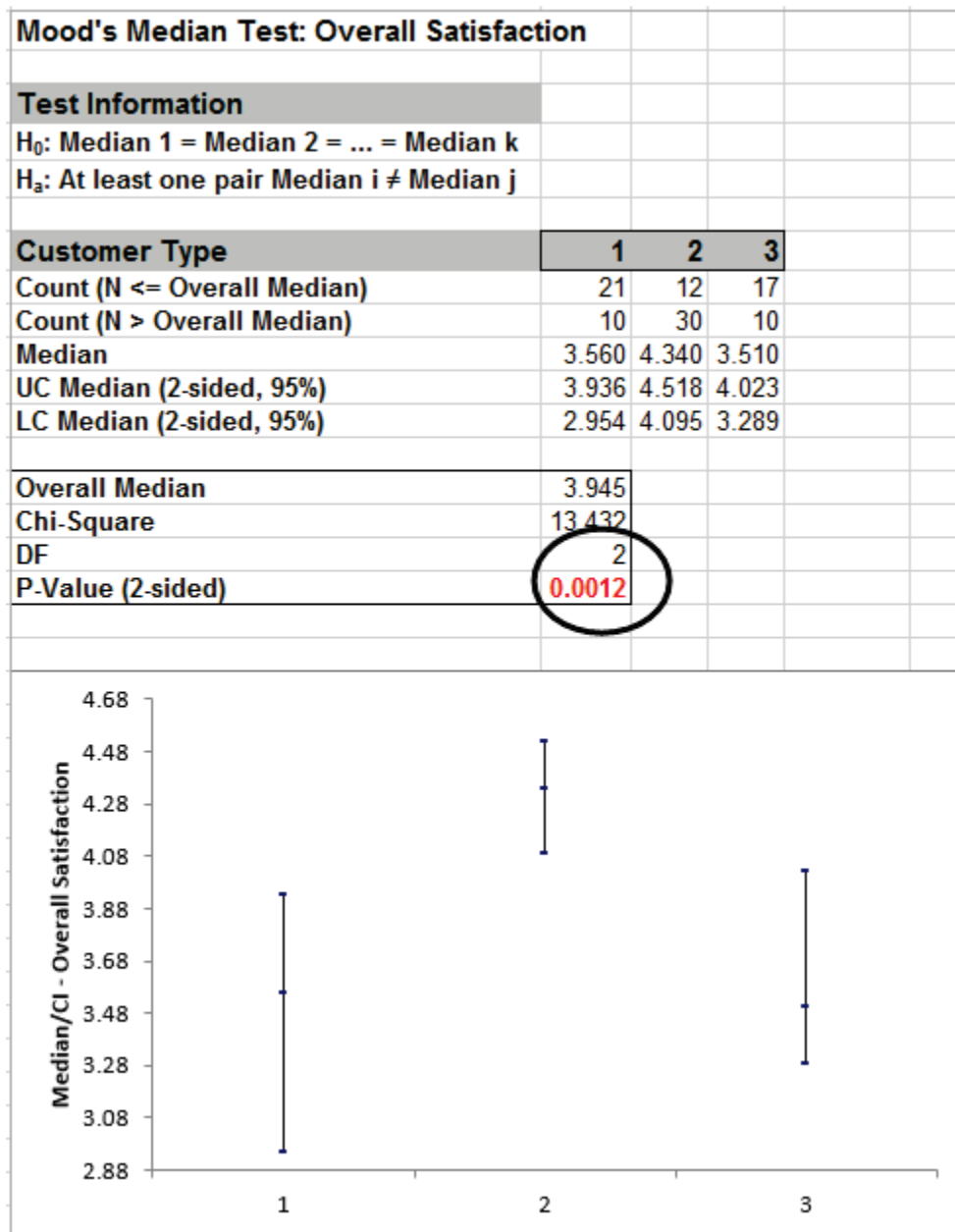
Mood's Median Test is an extension of the One Sample Sign Test, using Chi-Square as the test statistic. Like the Kruskal-Wallis test, Mood's median test can be used to test the equality of medians from multiple samples. It provides a nonparametric alternative to the one-way analysis of variance. The Null Hypothesis is: $H_0: \text{Median}_1 = \text{Median}_2 = \dots = \text{Median}_K$. H_a : 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:



The P-Value of 0.0012 tells us that we reject H_0 . 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 $N \leq 10$ 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**.

The screenshot shows the 'Mood's Median Test - Exact' dialog box. On the left is a list of variables: Customer Record No, Order Date, Avg No. of orders per mo, Avg days Order to delivery, Loyalty - Likely to Recomm, Responsive to Calls, Ease of Communications, Staff Knowledge, Size of Customer, Major-Complaint, Product Type, and Sat-Discrete. The 'Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)' is selected. 'Overall Satisfaction' is entered in the 'Numeric Data Variable (Y) >>' field, and 'Customer Type' is entered in the 'Group Category (X) >>' field. A '<< Remove' button is below these fields. In the 'Values Equal to Overall Median:' section, 'Counted as Below' is selected. The 'Exact' test is selected, with a 'Time Limit for Exact Computation' of 60 (Seconds). The 'Monte-Carlo Exact' section is also visible, showing 'Number of Replications' as 10000 and 'Confidence Level for P-Value' as 99 %.

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 H_0 . 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 \neq Median j			
Customer Type	1	2	3
Count (N \leq 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 H_0 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.

Mood's Median Test - Exact

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or more Numeric Data Columns)

Numeric Data Variable (Y) >> Overall Satisfaction

Group Category (X) >> Customer Type

<< Remove

Values Equal to Overall Median:

☒ Counted as Below
☐ Counted as Above
☐ Not Counted

☐ Exact
Time Limit for Exact Computation: 60 (Seconds)

☒ Monte-Carlo Exact
Number of Replications: 1e6
Confidence Level for P-Value: 99 %

OK >>
Cancel
Help

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 \neq Median j			
Customer Type	1	2	3
Count (N \leq 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 H_0 : at least one pairwise set of medians are not equal.

- 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 amount of snoring on a 100-point scale while using a device.

- 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**.
- 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.

The screenshot shows the 'Mood's Median Test - Exact' dialog box. It has a title bar with a close button. The main area contains several sections: 1. Format selection: 'Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)' is unselected, and 'Unstacked Column Format (2 or more Numeric Data Columns)' is selected. 2. Variable selection: A box labeled 'Numeric Data Variables (Y) >>' contains a list with 'Device A', 'Device B', and 'Device C'. 'Device C' is highlighted. To the right of this list are buttons for 'OK >>', 'Cancel', and 'Help'. Below the list is a '<< Remove' button. 3. Median handling: A section titled 'Values Equal to Overall Median:' has three radio buttons: 'Counted as Below' (selected), 'Counted as Above', and 'Not Counted'. 4. Test selection: A section with two radio buttons: 'Exact' (selected) and 'Monte-Carlo Exact'. 5. Monte-Carlo settings: Under 'Monte-Carlo Exact', there are two input fields: 'Number of Replications' with the value '1e6' and 'Confidence Level for P-Value' with the value '99 %'.

- 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 \neq Median j			
Results:	Device A	Device B	Device C
Count (N \leq 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 H_0 , 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.

- Press **F3** or click **Recall SigmaXL Dialog** to recall last dialog. Set **Values Equal to Overall Median:** to **Counted as Above**.

Mood's Median Test - Exact

☐ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☒ Unstacked Column Format (2 or more Numeric Data Columns)

Numeric Data Variables (Y) >>

<< Remove

Device A
Device B
Device C

OK >>
Cancel
Help

Values Equal to Overall Median:

☐ Counted as Below
☒ Counted as Above
☐ Not Counted

☒ Exact
 Time Limit for Exact Computation: 60 (Seconds)
☐ Monte-Carlo Exact
 Number of Replications: 1e6
 Confidence Level for P-Value: 99 %

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 \neq 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 H_0 , 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.

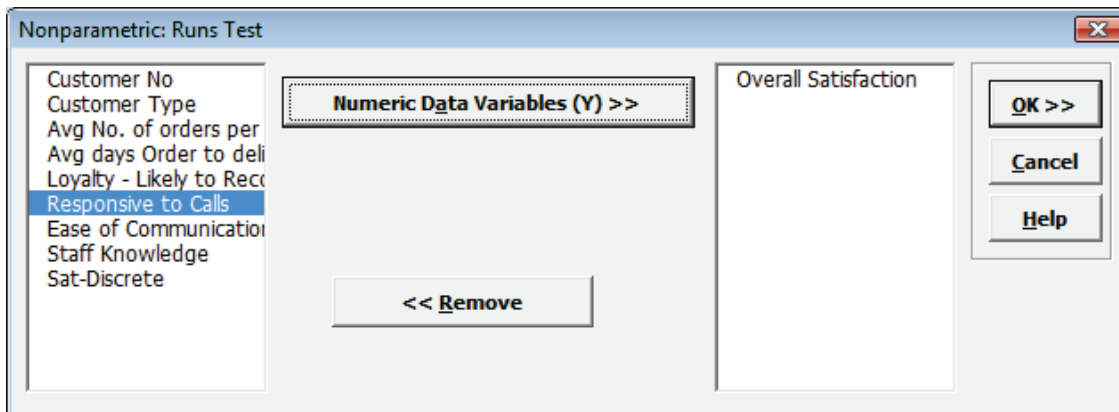
In conclusion, when using Mood's Median Exact, always try **Counted as Below** and **Counted as Above** to ensure that the P-Values agree with each other.

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 H_0 : The data is random (or independent). The alternative hypothesis is H_a : 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 $\alpha = 0.01$), 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:

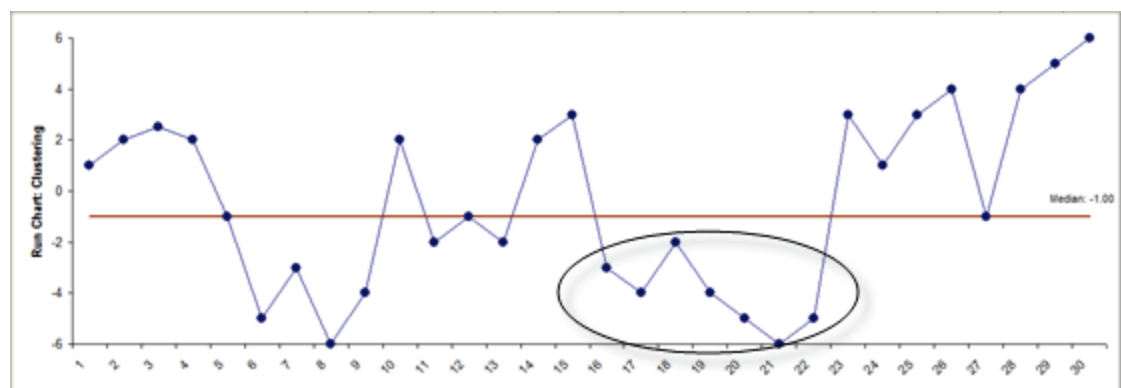
Nonparametric Runs Test	
Test Information	Overall Satisfaction
Number of Runs about Median:	44
Expected Number of Runs about Median:	51
Number of Points above Median:	50
Number of Points equal to or below Median:	50
P-Value for Clustering:	0.0797
P-Value for Mixtures:	0.9203
P-Value for Lack of Randomness (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 H_0 , 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

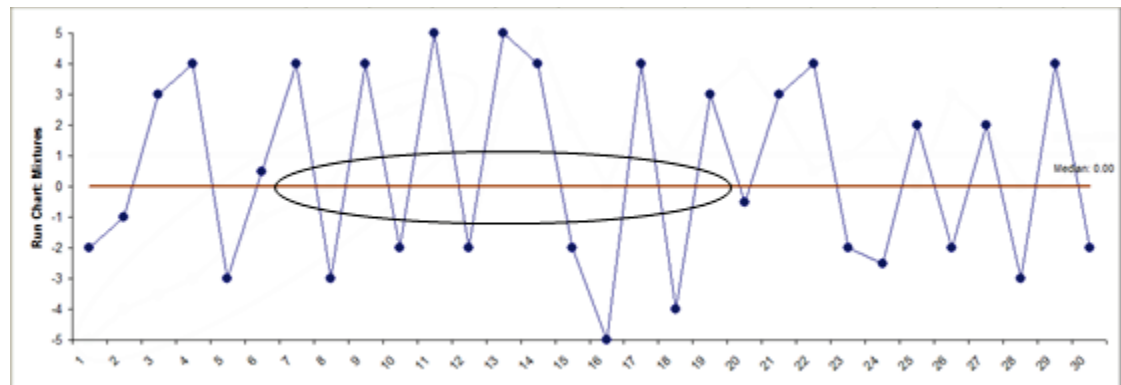
Examples of Clustering, Mixtures, Trends, and Oscillations are given below using the Run Chart to illustrate. The data for these examples are given in the file **Runs Test Example Data.xlsx**. (Use **SigmaXL > Graphical Tools > Run Charts** to create the run charts).

- 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:	
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.1168
P-Value for Oscillation:	0.8832

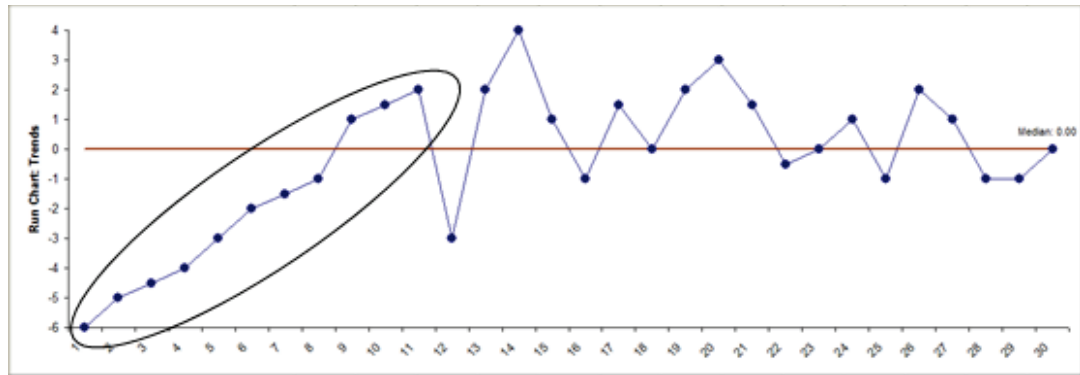
- b. **Mixtures** appear as an absence of data points near the center line. A mixture may indicate a bimodal distribution due to a regular change of shift, machinery, or raw materials.



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:	
Expected Number of Runs Up or Down:	19.667
P-Value for Trends:	0.8514
P-Value for Oscillation:	0.1486

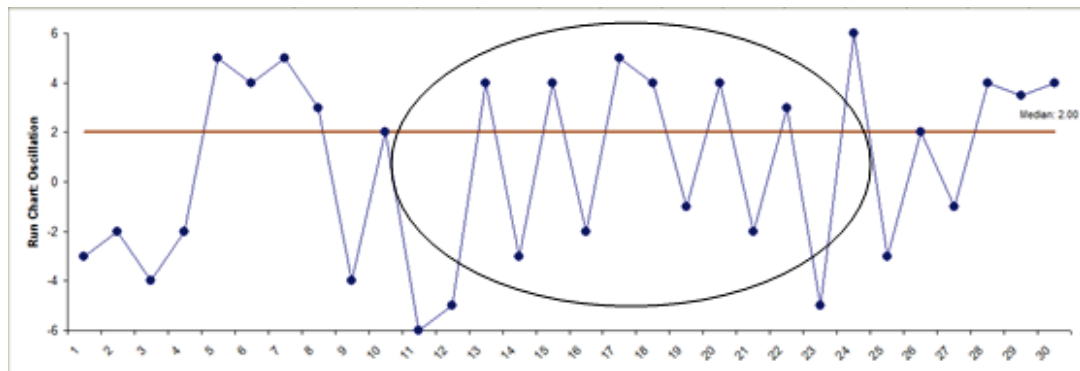
Note that the P-Value for Mixtures = 1 – P-Value for Clustering. They are mutually exclusive. The P-Value for Lack of Randomness = 2 * minimum of (P-Value Clustering, P-Value Mixtures).

- 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 ($N \leq 50$), 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)

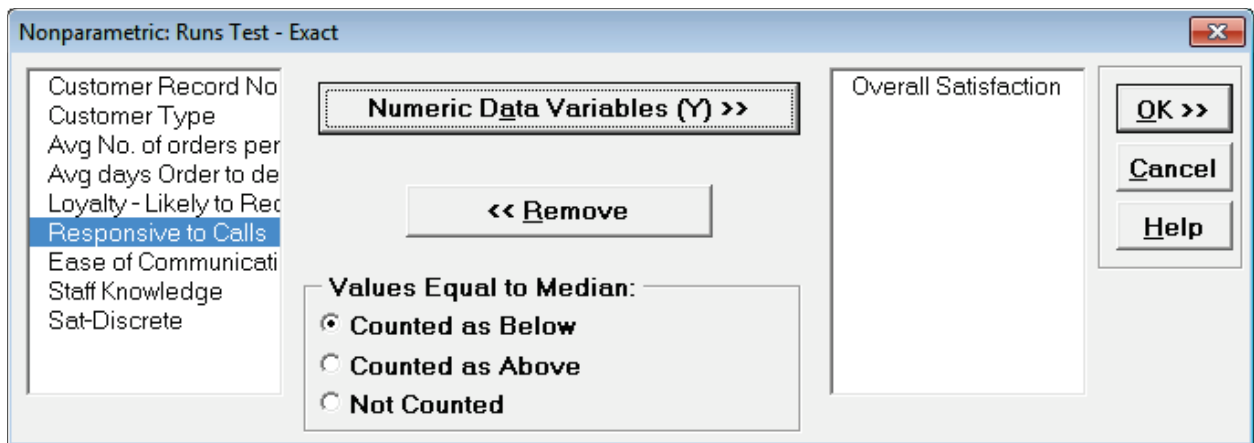
Exact P-Values are derived from published tables, given the sample size and the number of up/down runs. The exact tables apply to $N \leq 50$. If $N > 50$, a continuity-corrected normal approximation is used.

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 ($N \leq 50$), 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:

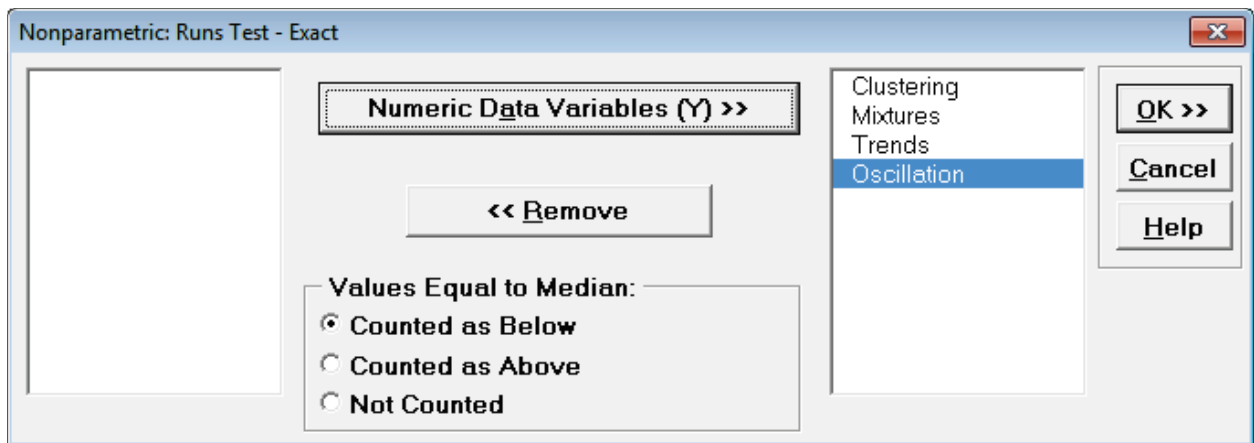
Nonparametric Runs Test - Exact	
Test Information	Overall Satisfaction
Number of Runs about Median:	44
Expected Number of Runs about Median:	51
Number of Points equal to or below Median:	50
Number of Points above Median:	50
Exact P-Value for Clustering:	0.0957
Exact P-Value for Mixtures:	0.9345
Exact 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 H_0 , 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.

- Now we will consider the small sample examples. Open **Runs Test Example Data.xlsx**, click **Runs Test Example Data** tab.
- Click **SigmaXL > Statistical Tools > Nonparametric Tests – Exact > Runs Test – Exact**. If necessary, check **Use Entire Data Table**, click **Next**.
- Select *Clustering* and shift click to *Oscillation*, click **Numeric Data Variables (Y) >>**. Set **Values Equal to Median:** to **Counted as Below**.



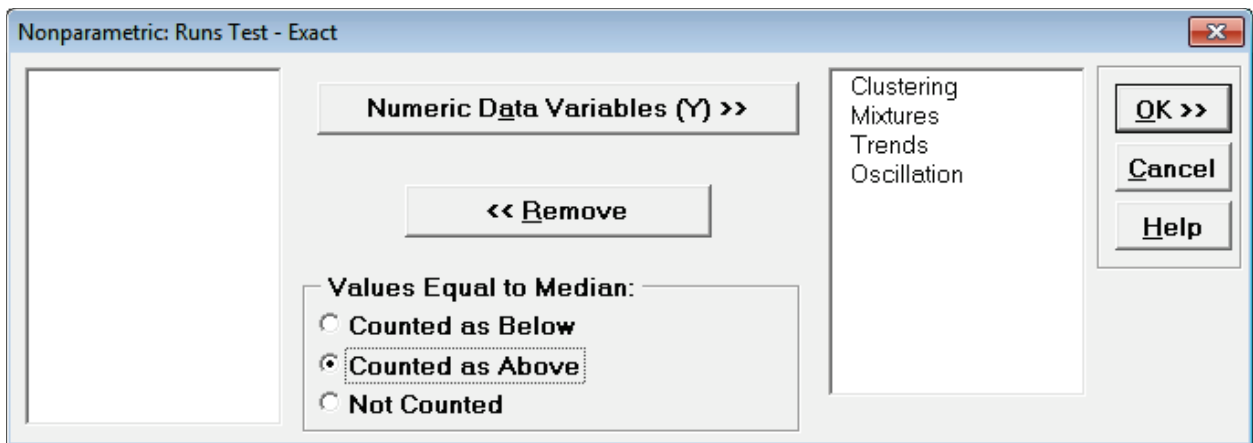
8. Click **OK**. Resulting output:

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 H_0 . 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, H_0 : Proportion P_1 = Proportion P_2 ; the alternative hypothesis, H_a is: $P_1 \neq P_2$. 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%)	$100*(1-\alpha)\%$	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 H_0 . 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 H_0 .

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**.

Chi-Square Test & Association - Two Way Table Data

Please select your data

\$B\$2:\$D\$4

Note:
 1. Two-Way (pivot) table must be based on a single Y or count variable.
 2. Selection should include only one row and only one column label.
 3. Selection should exclude Grand Totals.
 See Help for example.

Advanced Tests and Measures of Association

☐ Nominal Categories

☐ Ordinal Categories

Confidence Level: 95.0

Help Cancel Next >>

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.

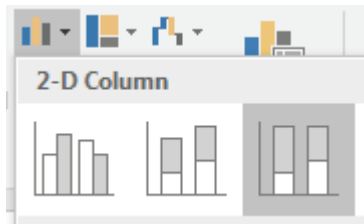
- Click the **Example 4** Sheet tab. Here we have doubled the sample size to 200 per supplier. Note that the percentages are identical to example 3. Redoing the above analysis yields the following:

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		

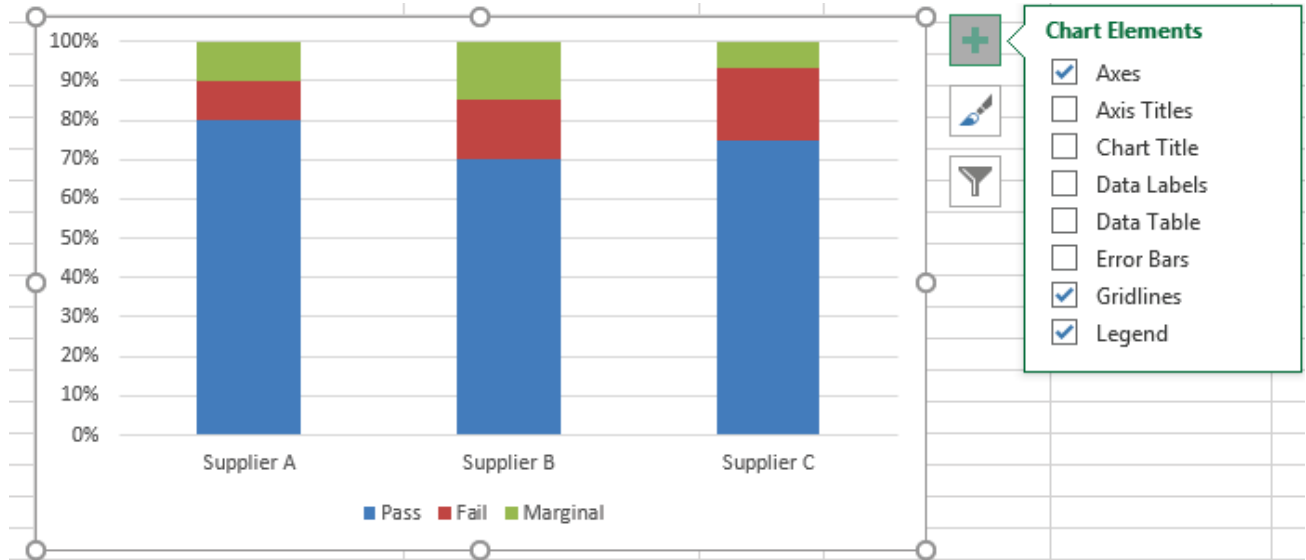
With the P-Value < .05 we now conclude that there is a significant difference across suppliers. Examining the Std. (Standardized) Residuals tells us that Supplier A has fewer failures than

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.

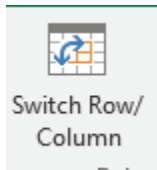
7. The table row and column cell percentages can be visualized using Excel's 100% Stacked Column Chart. Select cells **A3:D6** of the **Chi-Square** sheet. Click Excel's **Insert > Insert Column or Bar Chart** and select **100% Stacked Column** as shown.

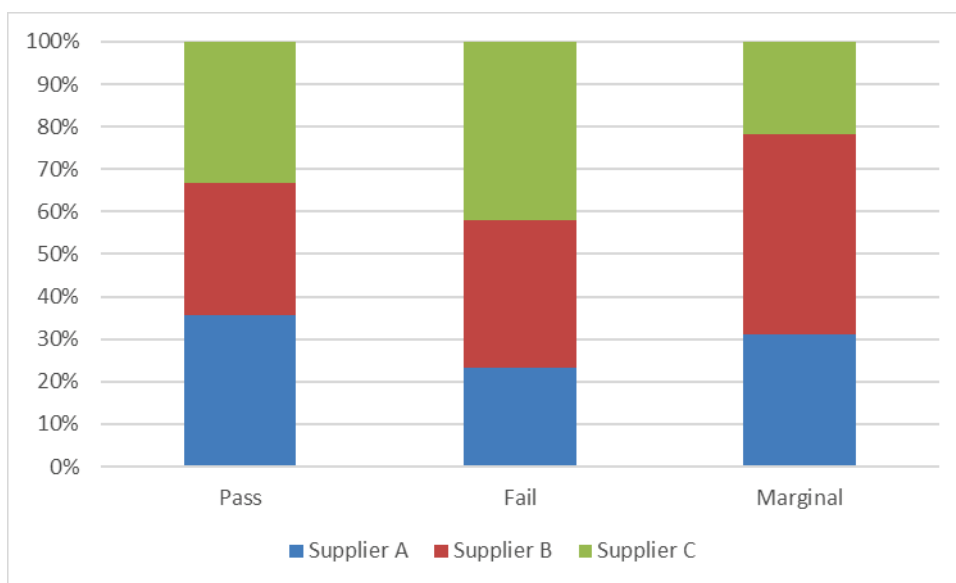


8. Click to create the 100% stacked column chart. Uncheck the Chart Title as shown.



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:

Chi-Square Test & Association - Two Way Table Data

Please select your data

\$B\$2:\$E\$5

Note:
 1. Two-Way (pivot) table must be based on a single Y or count variable.
 2. Selection should include only one row and only one column label.
 3. Selection should exclude Grand Totals.
 See Help for example.

Advanced Tests and Measures of Association

☒ Nominal Categories

☐ Ordinal Categories

Confidence Level: 95.0

Help Cancel Next >>

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	0.66666667	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 be 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**:

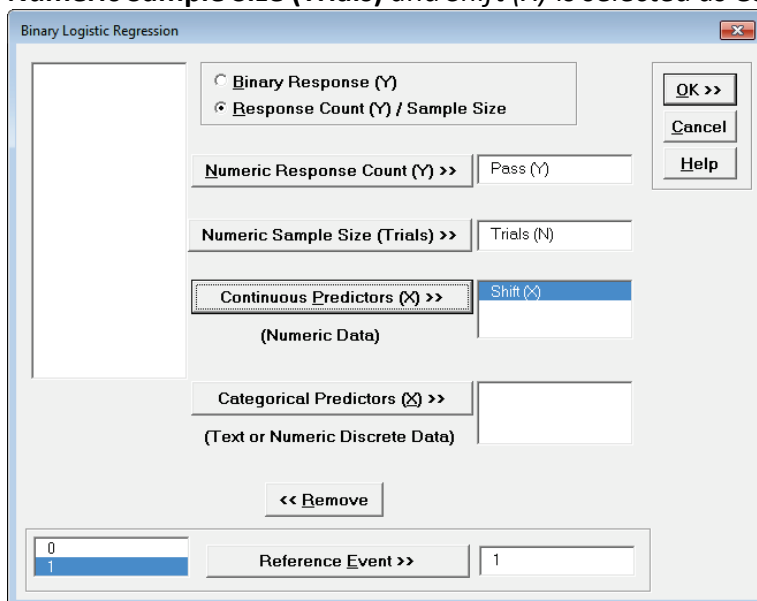
	Day Shift	Night Shift
Pass	160	140
Fail	40	60

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 \neq 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:

Chi-Square Test & Association - Two Way Table Data

Please select your data

\$A\$1:\$E\$5

Note:
1. Two-Way (pivot) table must be based on a single Y or count variable.
2. Selection should include only one row and only one column label.
3. Selection should exclude Grand Totals.
See Help for example.

Advanced Tests and Measures of Association

☒ Nominal Categories

☒ Ordinal Categories

Confidence Level: 95.0

Help Cancel Next >>

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).

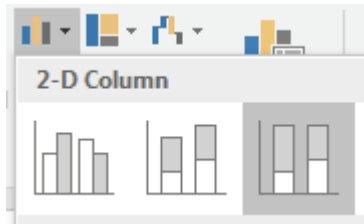
Since the Chi-Square P-Value is not significant, the **Adjusted Residuals** are not highlighted, even though some values are greater than 1.96 (and less than -1.96). This follows the concept used in ANOVA called "Fisher Protected" where one considers the significance of post-hoc tests only when the overall 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

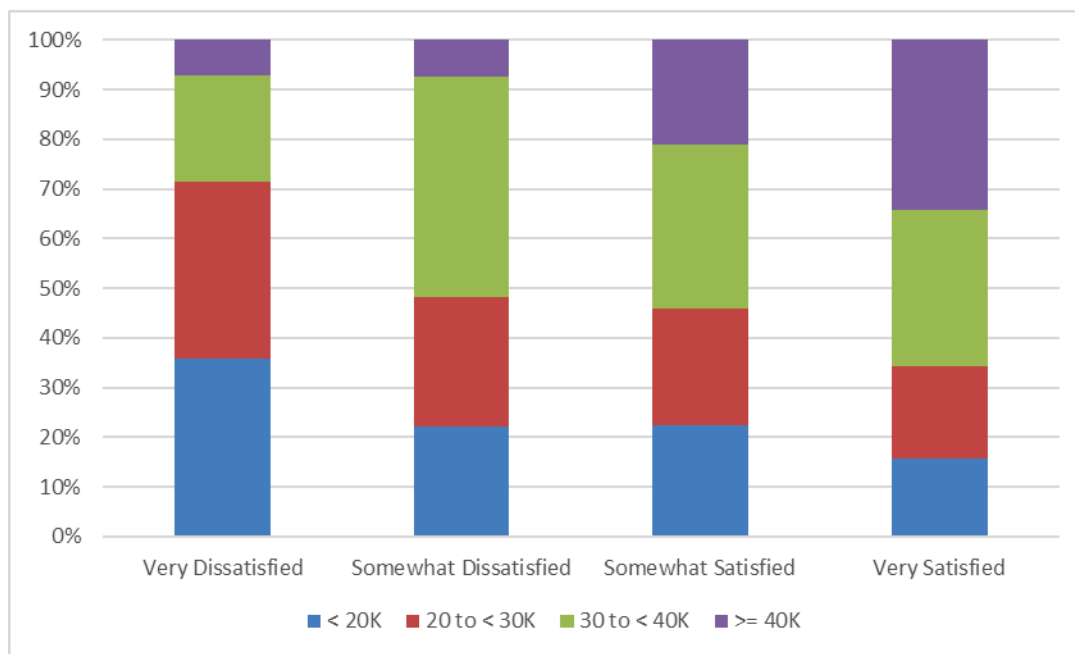
later, but for this example the Fisher's Monte-Carlo Exact P-Value = 0.095 so does not change the interpretation of the results for the above Chi-Square analysis.

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.

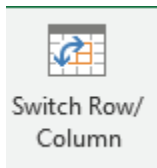
- The table row and column cell percentages can be visualized using Excel's 100% Stacked Column Chart. Select cells **A3:E7** of the **Chi-Square** sheet. Click Excel's **Insert > Insert Column or Bar Chart** and select **100% Stacked Column** as shown.

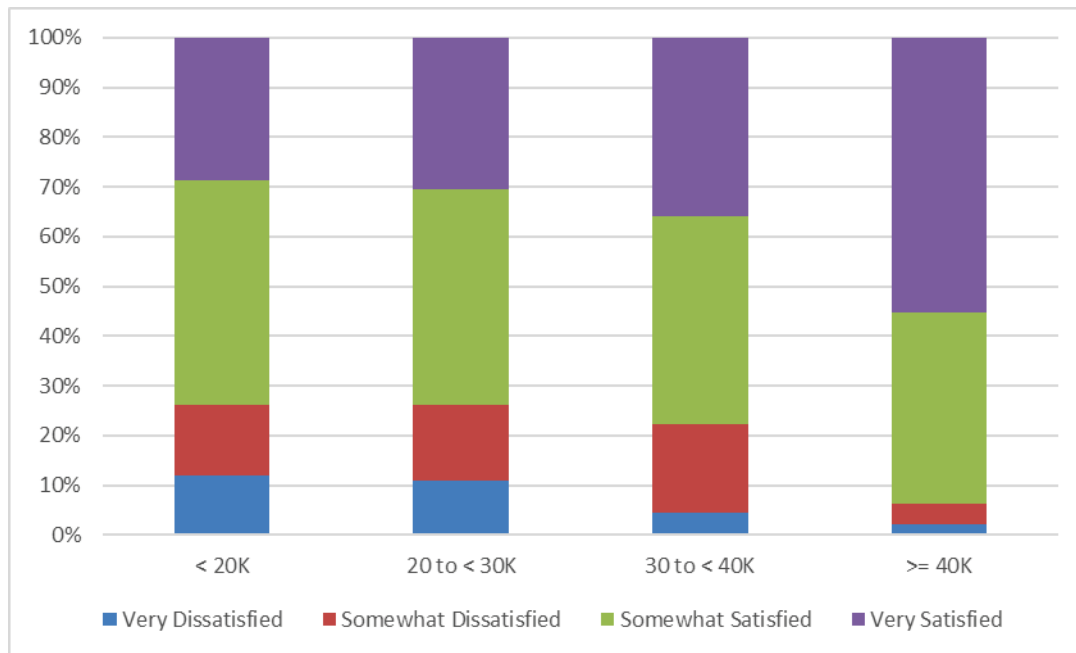


- Click to create the 100% stacked column chart (uncheck the Chart Title):



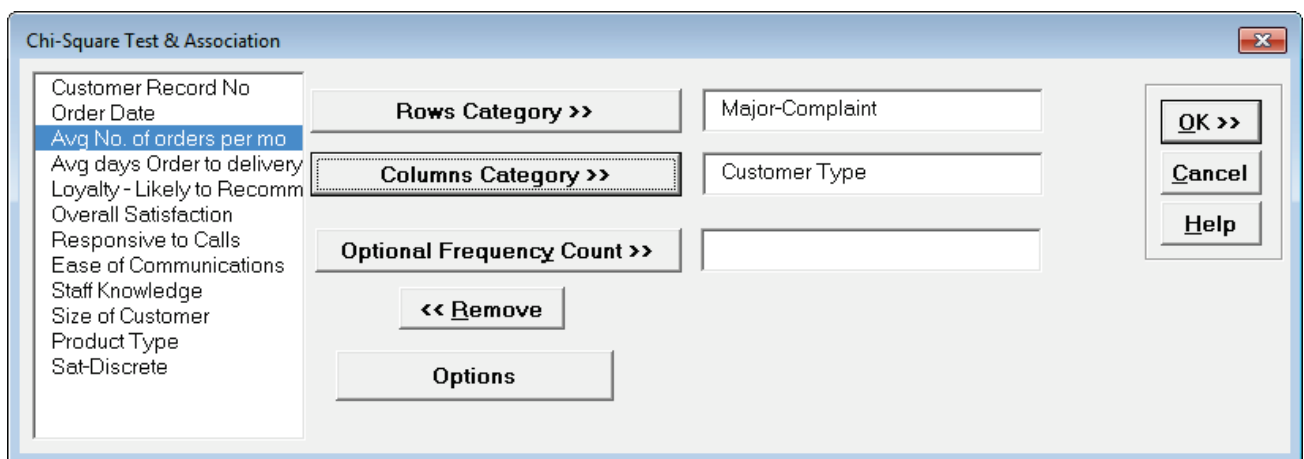
- The rows and columns can easily be switched by clicking **Design > Switch Row/Column**





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 H_0 , 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).

- 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**.

Chi-Square Test & Association

Customer Record No
Order Date
Customer Type
Avg No. of orders per mo
Avg days Order to delivery
Overall Satisfaction
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type

Rows Category >> Loyalty - Likely to Recommend

Columns Category >> Sat-Discrete

Optional Frequency Count >>

<< Remove

Options

Confidence Level: 95.0

Advanced Tests and Measures of Association

☐ Nominal Categories

☒ Ordinal Categories

OK >>
Cancel
Help

Sat-Discrete is derived from Overall Satisfaction where a score ≥ 3.5 is considered a 1, and scores < 3.5 are considered a 0.

- 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.

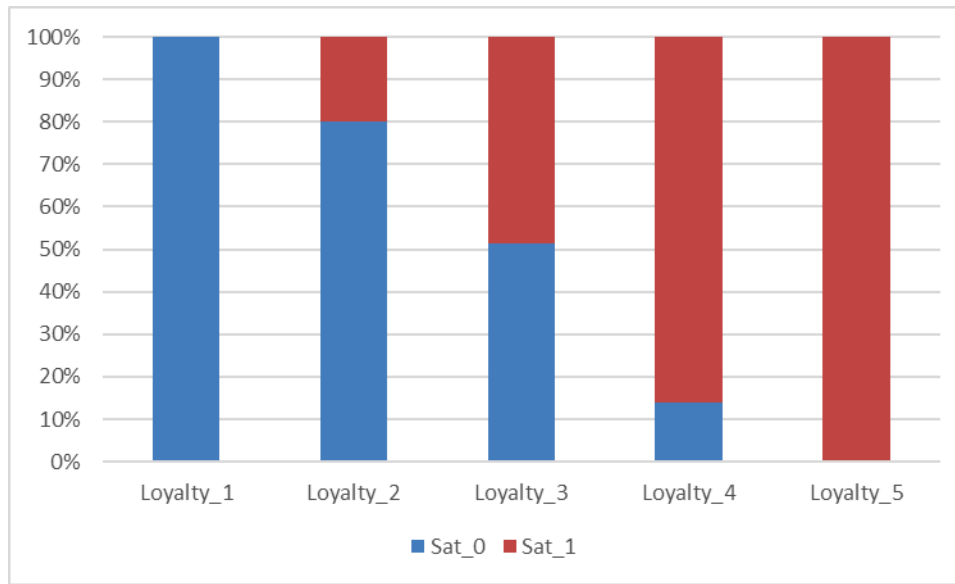
- In order to visualize the row and column percentages with Excel's 100% Stacked Column Chart, we will need to modify the numeric row and column labels as shown, converting to text as shown:

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

- 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.

The screenshot shows the 'Chi-Square Test - Fisher's Exact' dialog box. On the left is a list of variables: Customer Record No, Order Date, Avg No. of orders per mo (highlighted), Avg days Order to delivery, Loyalty - Likely to Recomm, Overall Satisfaction, Responsive to Calls, Ease of Communications, Staff Knowledge, Size of Customer, Product Type, and Sat-Discrete. In the center, there are three buttons: 'Rows Category >>', 'Columns Category >>', and 'Optional Frequency Count >>'. To the right of these buttons are text boxes containing 'Major-Complaint', 'Customer Type', and an empty box. Below these is a '<< Remove' button. On the far right are 'OK >>', 'Cancel', and 'Help' buttons. At the bottom right, there are radio buttons for 'Exact' (selected) and 'Monte Carlo Exact'. Below the 'Exact' radio button is a 'Time Limit for Exact Computation' field set to '60 (Seconds)'. Below the 'Monte Carlo Exact' radio button are fields for 'Number of Replications' set to '10000' and 'Confidence Level for P-Value' set to '99 %'.

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 H_0 , 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%.

Chi-Square Test - Fisher's Exact

Customer Record No
Order Date
Avg No. of orders per mo
Avg days Order to delivery
Loyalty - Likely to Recomm
Overall Satisfaction
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Product Type
Sat-Discrete

Rows Category >> Major-Complaint

Columns Category >> Customer Type

Optional Frequency Count >>

<< Remove

☐ Exact
Time Limit for Exact Computation: 60 (Seconds)

☒ Monte Carlo Exact
Number of Replications: 10000
Confidence Level for P-Value: 99 %

OK >>
Cancel
Help

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.

Chi-Square 2 Way Table Statistics - Fisher's Exact

Please select your data

\$A\$1:\$D\$10

Note:
 1. Pivot table must be based on a single Y or count variable.
 2. Selection should include only one row and only one column label.
 3. Selection should exclude Grand Totals.
 See Help for example.

☒ Exact
 Time Limit for Exact Computation: 60 (Seconds)

☐ Monte-Carlo Exact
 Number of Replications: 10000
 Confidence Level for P-Value: 99 %

Help Cancel Next >>

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 H_0 , 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%.

Chi-Square 2 Way Table Statistics - Fisher's Exact

Please select your data

\$A\$1:\$D\$10

Note:
 1. Pivot table must be based on a single Y or count variable.
 2. Selection should include only one row and only one column label.
 3. Selection should exclude Grand Totals.
 See Help for example.

☐ Exact
 Time Limit for Exact Computation: 60 (Seconds)

☒ Monte-Carlo Exact
 Number of Replications: 10000
 Confidence Level for P-Value: 99 %

Help Cancel Next >>

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” H_0 and lower is a “reject” H_0), 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, <http://www.cytel.com/Papers/sxpaper.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:

The screenshot shows the 'Power and Sample Size: 1 Proportion Test Calculator' dialog box. The 'Solve For' section has 'Power (1-Beta)' selected with a radio button. The 'Sample Size (N)' is entered as 300. The 'Alternative Proportion 1 (P1)' is entered as .6. The 'Hypothesized Proportion (P0)' is 0.5. The 'Significance Level (Alpha)' is 0.05. The 'Ha:' dropdown is set to 'Not Equal To'. Buttons for 'OK >>', 'Cancel', and 'Help' are on the right.

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:

Power and Sample Size: 2 Proportions Test Calculator

Solve For

Power (1-Beta) ☒

Sample Size (N) for Each Group 300 ☐

Proportion 1 (P1) .6 ☐

Proportion 2 (P2) 0.5

Significance Level (Alpha) 0.05

Ha: Not Equal To

OK >> **Cancel** **Help**

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 H_a : 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>:

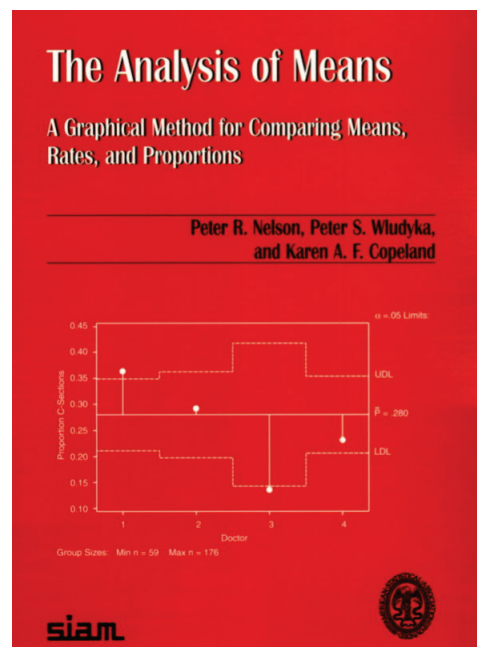
A statistical procedure for troubleshooting industrial processes and analyzing the results of experimental designs with factors at fixed levels. It provides a graphical display of data. Ellis R. Ott developed the procedure in 1967 because he observed that nonstatisticians had difficulty understanding 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)

$$\begin{aligned}\text{UDL} &= \bar{y}_{..} + h(\alpha; I, N - I) \sqrt{MS_e} \sqrt{\frac{I-1}{N}} \\ \text{LDL} &= \bar{y}_{..} - h(\alpha; I, N - I) \sqrt{MS_e} \sqrt{\frac{I-1}{N}}\end{aligned}$$

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

$\sqrt{MS_e}$ 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)

$$\begin{aligned}\text{UDL} &= \bar{p} + h(\alpha; I, \infty) \sqrt{\bar{p}(1 - \bar{p})} \sqrt{\frac{I-1}{N}} \\ \text{LDL} &= \bar{p} - h(\alpha; I, \infty) \sqrt{\bar{p}(1 - \bar{p})} \sqrt{\frac{I-1}{N}}\end{aligned}$$

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)

$$\begin{aligned}\text{UDL} &= U(\alpha; I, n - 1) \text{IMS}_e \\ \text{CL} &= \text{MS}_e \\ \text{LDL} &= L(\alpha; I, n - 1) \text{IMS}_e\end{aligned}$$

where U, L are critical values from Table B.4 ($n \leq 35$). 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:

$$y_{ij} = |x_{ij} - \tilde{x}_i|$$

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.

ANOM Normal One-Way

Customer Record No
Order Date
Avg No. of orders per m
Avg days Order to deliv
Loyalty - Likely to Recor
Responsive to Calls
Ease of Communication
Staff Knowledge
Size of Customer
Major-Complaint

☒ **Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)**
☐ **Unstacked Column Format (2 or More Numeric Data Columns)**

Numeric Data Variable (Y) >> Overall Satisfaction

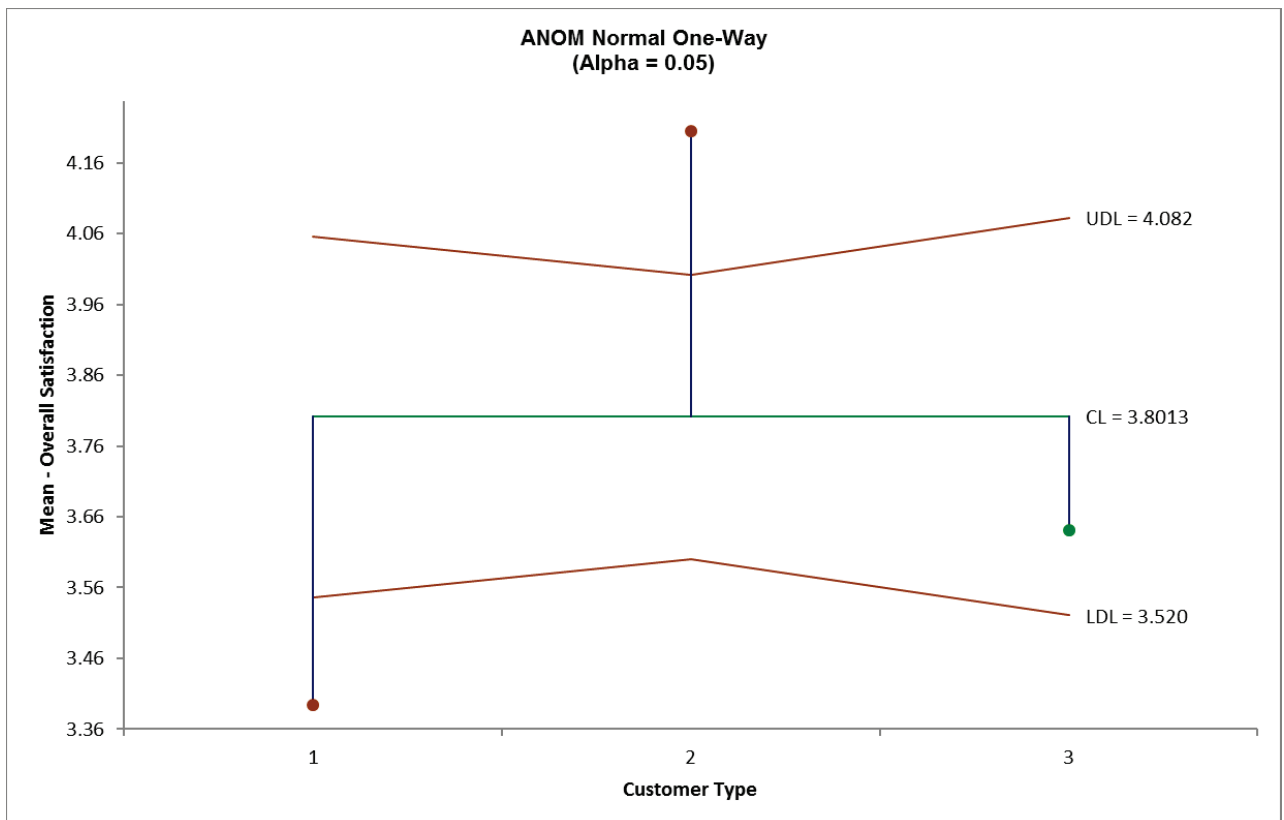
Group Category (X) >> Customer Type

<< Remove

Alpha Level: 0.05

OK >>
Cancel
Help

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:

ANOM Normal Two-Way

Numeric Data Variable (Y) >> Hemoglobin Level

Group Category Factor (X1) >> Therapy

Group Category Factor (X2) >> Drug

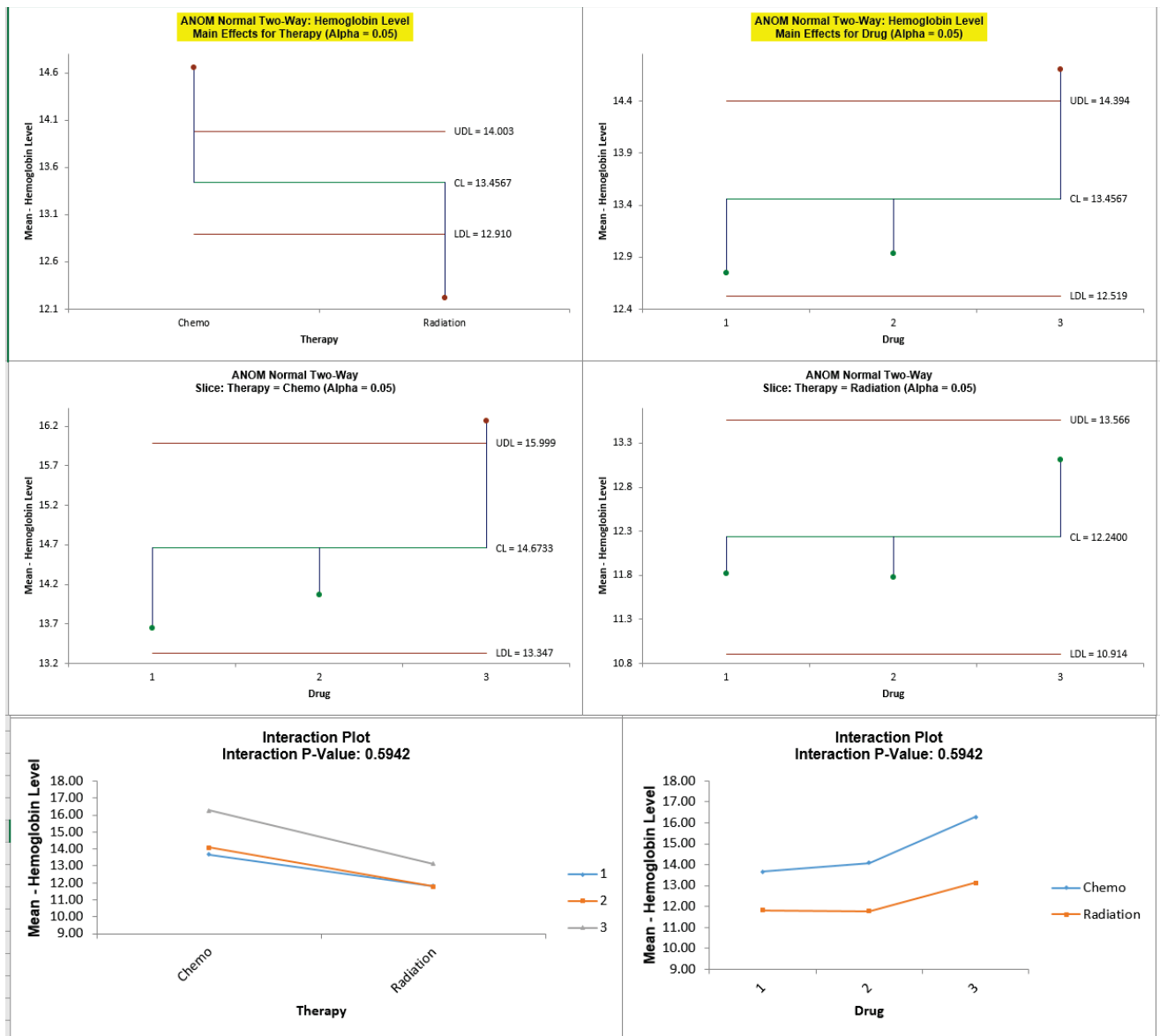
<< Remove

Alpha Level: 0.05

☐ Adjust chart alpha for family-wise error

OK >> Cancel Help

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 ($P\text{-Value} \geq 0.1$), 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 ($\alpha = 0.05$) 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 $P\text{-Value} = 0.0001$, Drug is also significant with a $P\text{-Value} = 0.0105$ and the Interaction is not significant, with a $P\text{-Value} = 0.5942$.

- Click **Recall SigmaXL Dialog** menu or press **F3** to recall last dialog. Check **Adjust chart alpha for family-wise error rate**.

ANOM Normal Two-Way

Numeric Data Variable (Y) >> Hemoglobin Level

Group Category Factor (X1) >> Therapy

Group Category Factor (X2) >> Drug

<< Remove

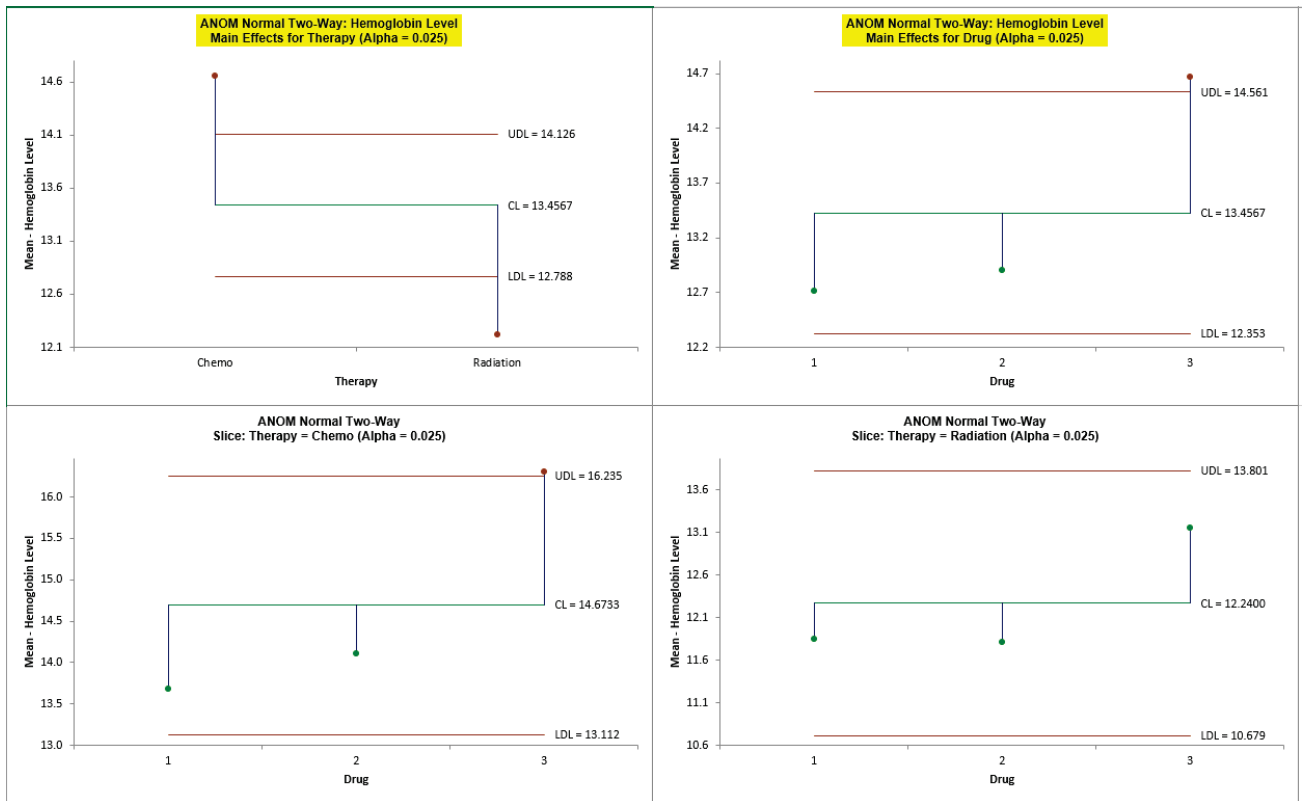
Alpha Level: 0.05

☒ Adjust chart alpha for family-wise error

OK >> Cancel Help

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:

ANOM Normal Two-Way

Numeric Data Variable (Y) >> Process Yield

Group Category Factor (X1) >> UV

Group Category Factor (X2) >> Monomer

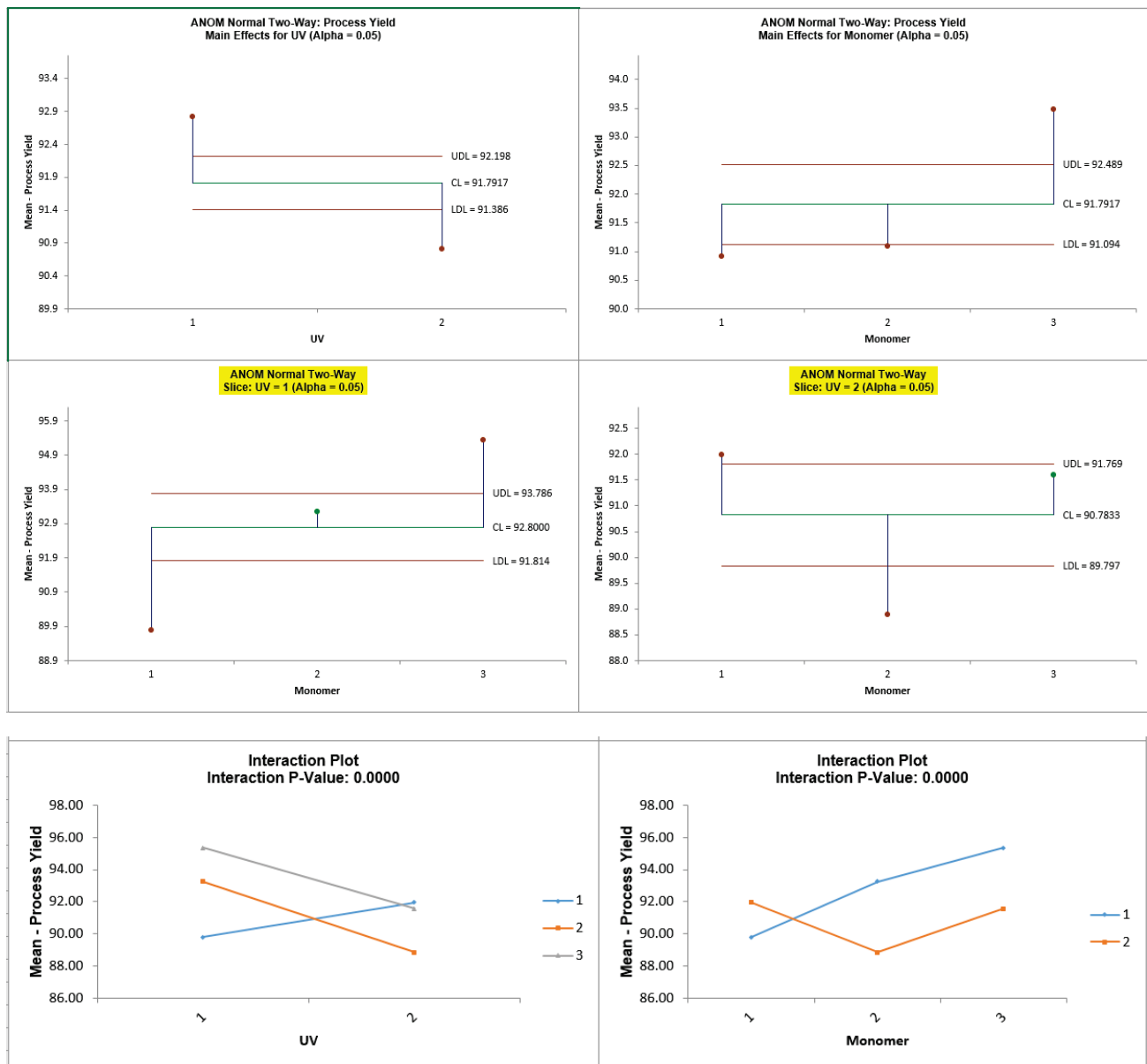
<< Remove

Alpha Level: 0.05

☐ Adjust chart alpha for family-wise error

OK >> Cancel Help

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:

ANOM Normal Two-Way

Numeric Data Variable (Y) >> Process Yield

Group Category Factor (X1) >> UV

Group Category Factor (X2) >> Monomer

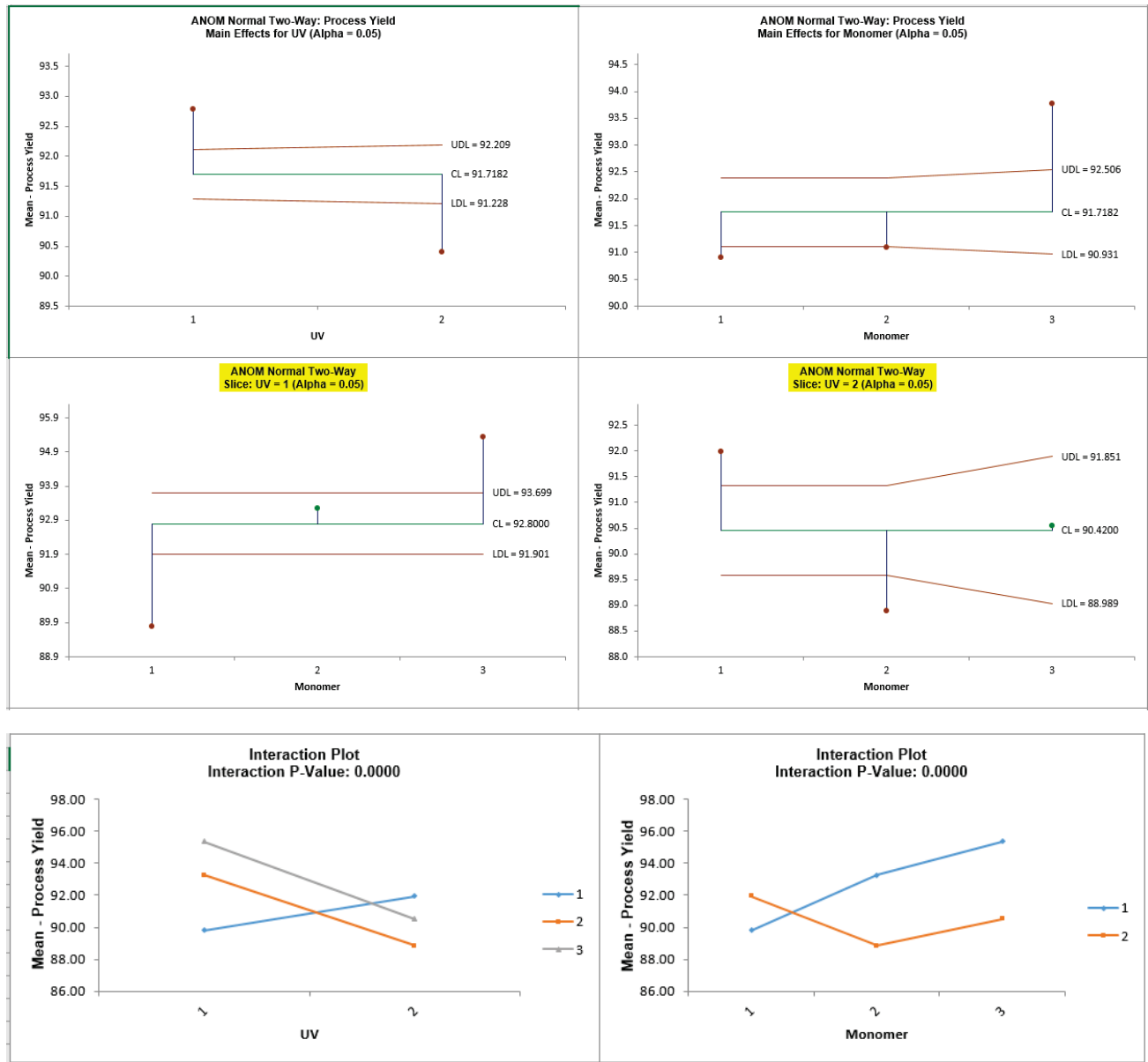
<< Remove

Alpha Level: 0.05

☐ Adjust chart alpha for family-wise error

OK >> Cancel Help

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.

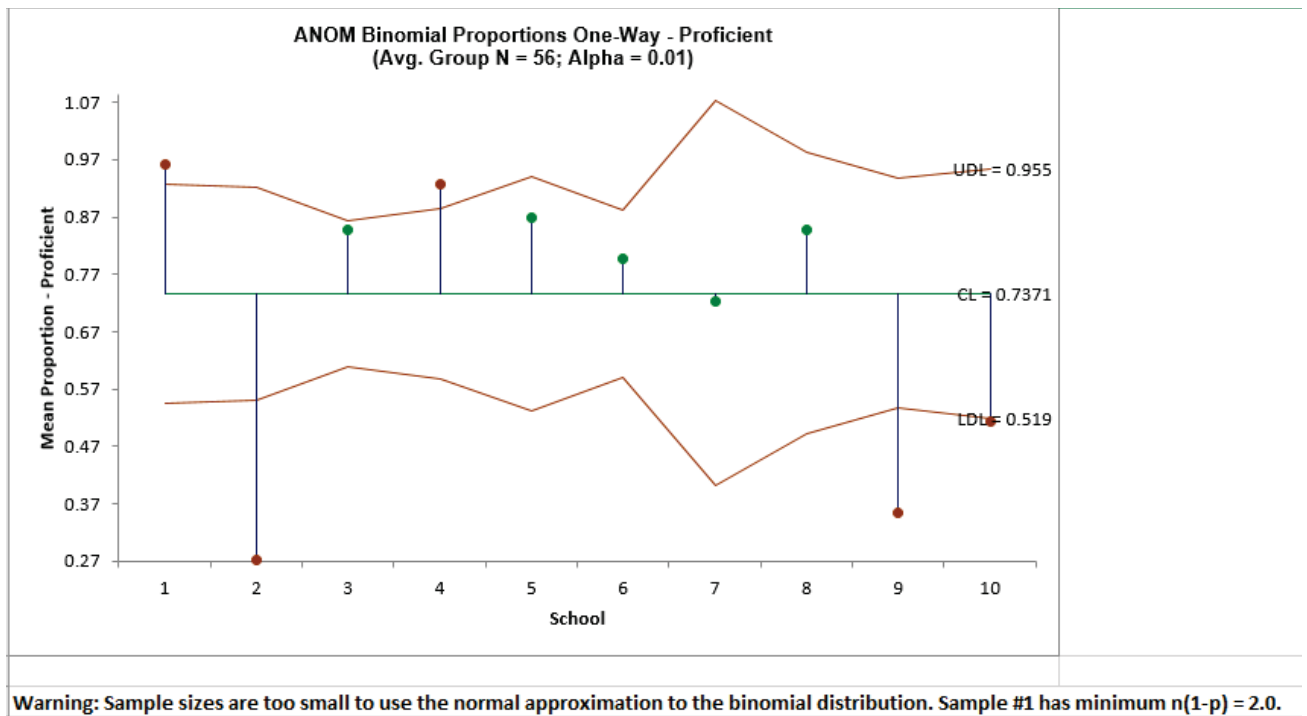
School	Proficient	Enrollment	Proportion
1	45	52	0.8654
2	15	55	0.2727
3	89	105	0.8476
4	75	81	0.9259
5	40	46	0.8696
6	67	84	0.7976
7	13	18	0.7222
8	28	33	0.8485
9	17	48	0.3542
10	21	41	0.5122

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:

ANOM Binomial Proportions One-Way

School	Numeric Data Variable (Y) >>	Proficient	<input type="button" value="OK >>"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/>
Proficient	Subgroup Column or Size >>	Enrollment	
Enrollment	Optional Group Category (X) >>	School	
Proportion	<< Remove	Alpha Level: 0.05	

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:

ANOM Binomial Proportions Two-Way

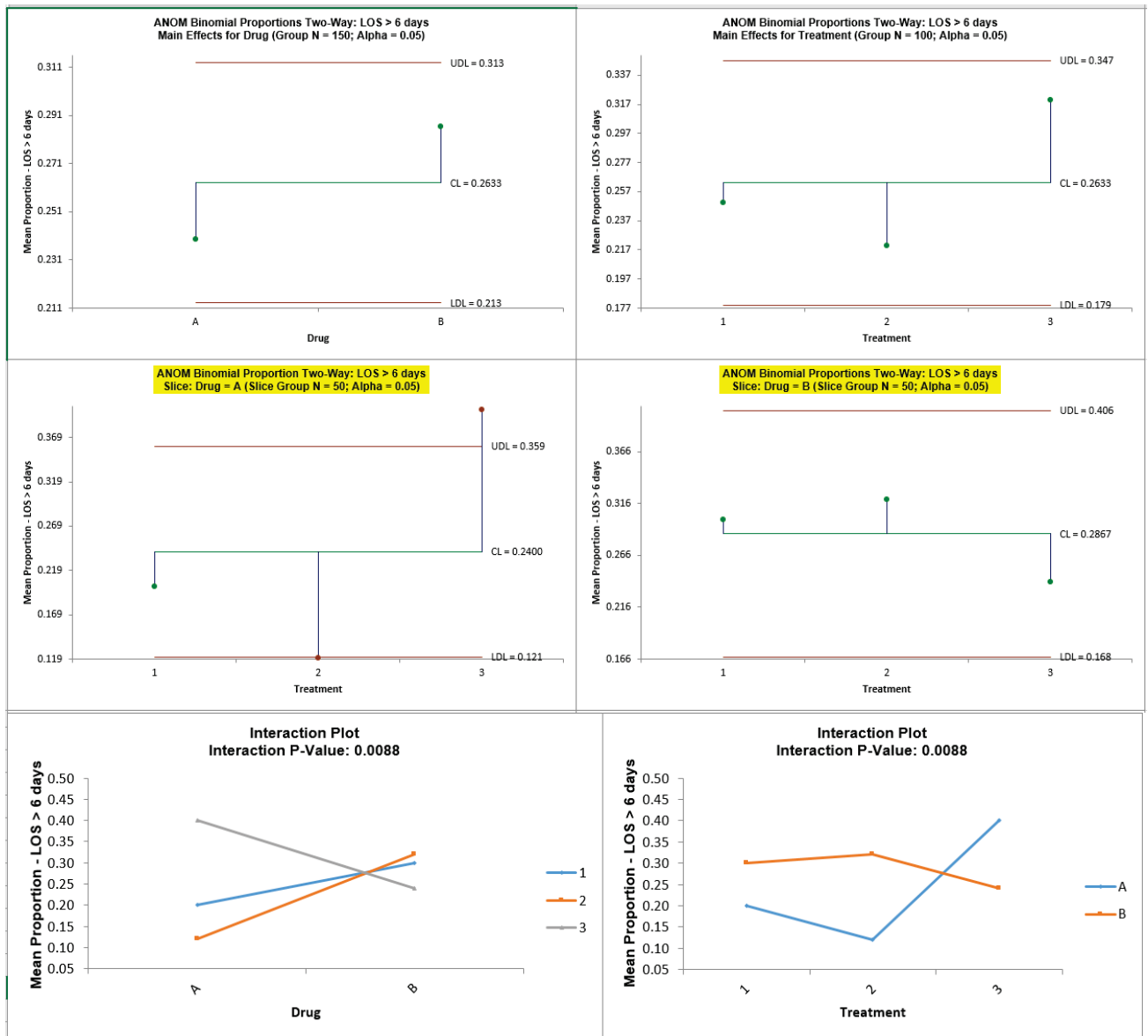
Left Pane (Selected): Drug, Treatment, **LOS > 6 days**, Sample Size

Right Pane:

- Numeric Data Variable (Y) >>: LOS > 6 days
- Subgroup Column or Size >>: Sample Size
- Group Category Factor (X1) >>: Drug
- Group Category Factor (X2) >>: Treatment
- Alpha Level: 0.05
- ☐ Adjust chart alpha for family-wise error rate

Buttons: OK >>, Cancel, Help, << Remove

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.

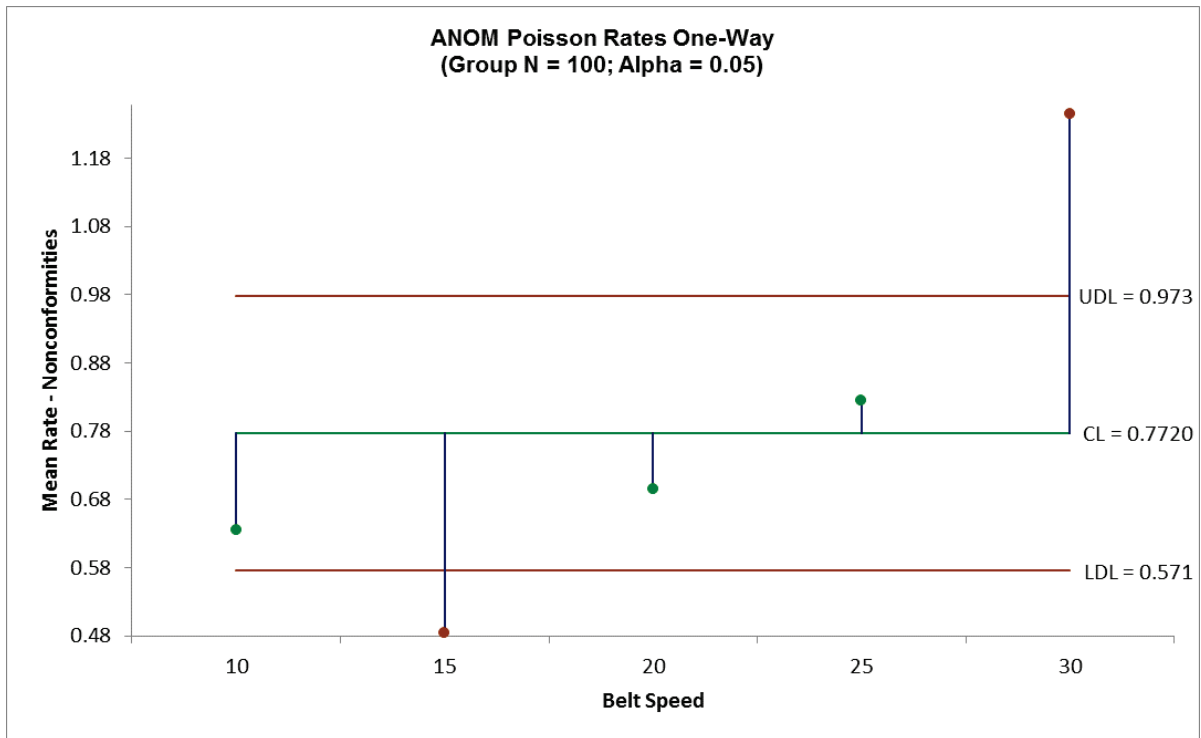
ANOM Poisson Rates One-Way

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:

ANOM Poisson Rates Two-Way

Intervention
Payor
ER Visits
Sample Size

Numeric Data Variable (Y) >> ER Visits

Subgroup Column or Size >> Sample Size

Group Category Factor (X1) >> Intervention

Group Category Factor (X2) >> Payor

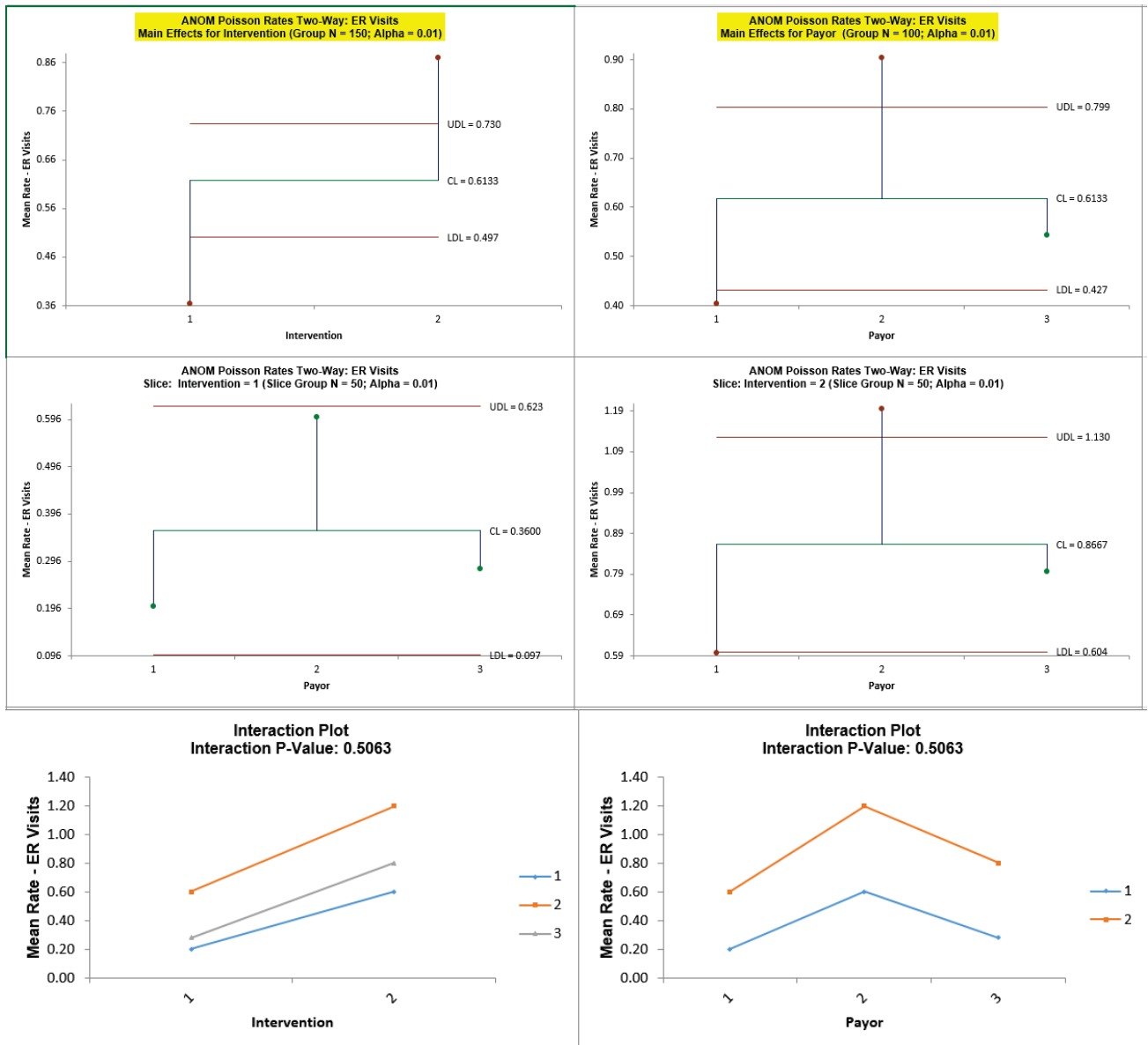
<< Remove

Alpha Level: 0.01

☐ Adjust chart alpha for family-wise error rate

OK >>
Cancel
Help

4. Click **OK**. The ANOM Poisson Rates Two-Way Main Effects, Slice charts and Interaction plots are shown below:

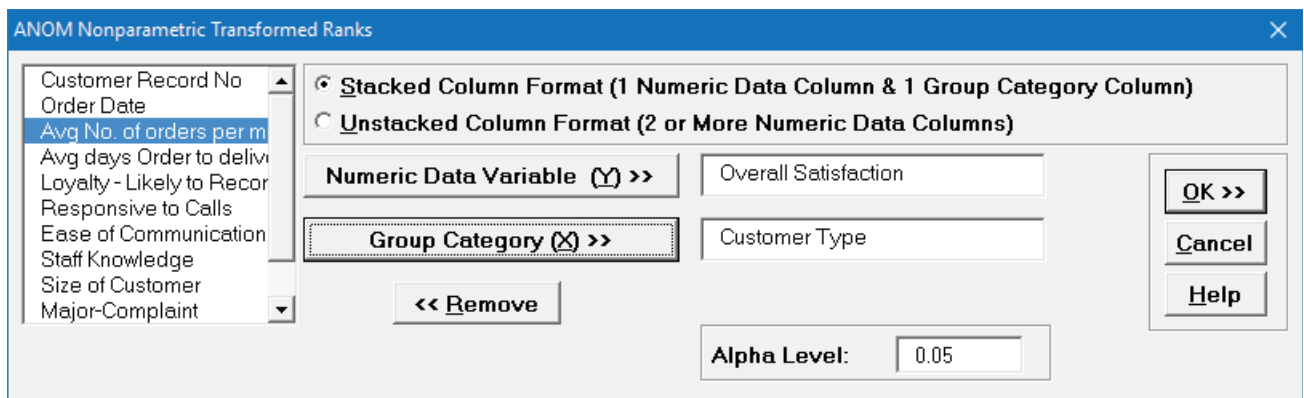


5. Since the Interaction is not significant (P-Value ≥ 0.1 , automatically determined from Poisson regression), the Main Effects charts are highlighted. Both factors Intervention and Payor are significant at $\alpha = 0.01$. "Home nurse assigned" results in a significantly lower rate of ER visits versus "no home nurse assigned." Payor Type 1 has a significantly lower rate of ER visits and Type 2 has a significantly higher rate.

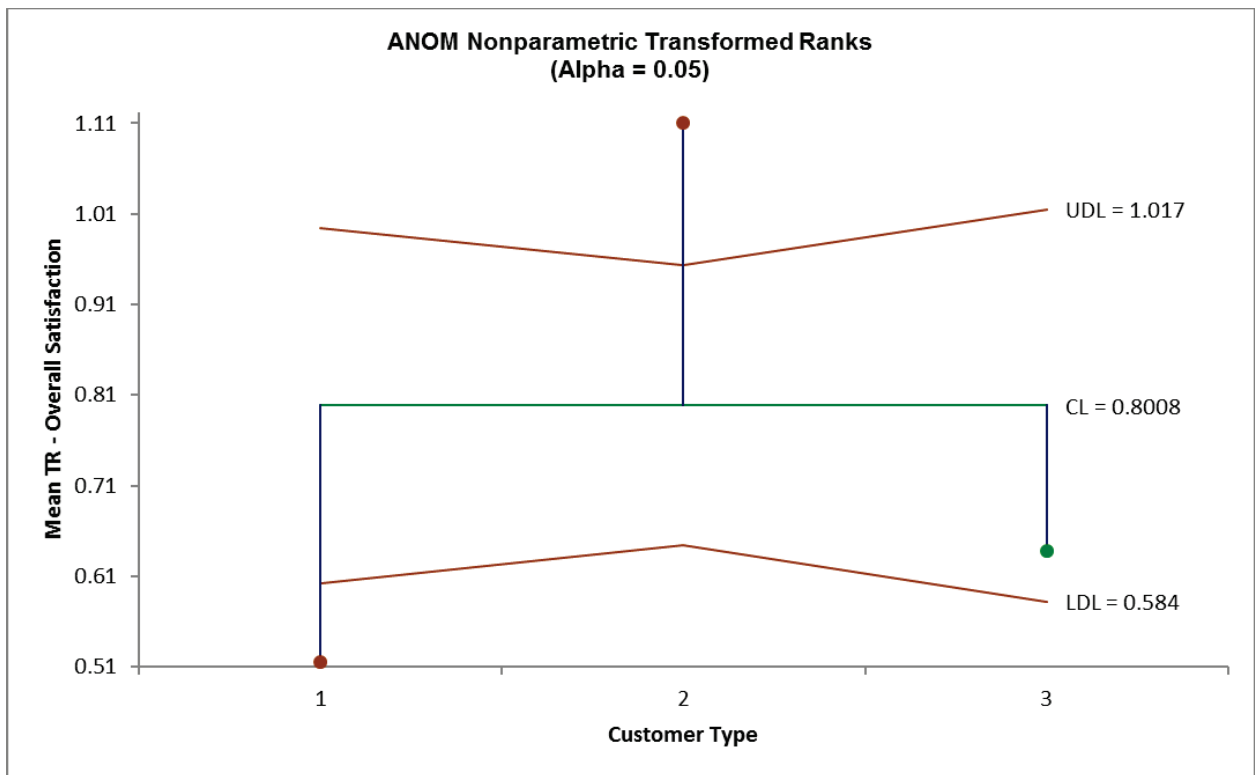
ANOM Nonparametric Transformed Ranks

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.

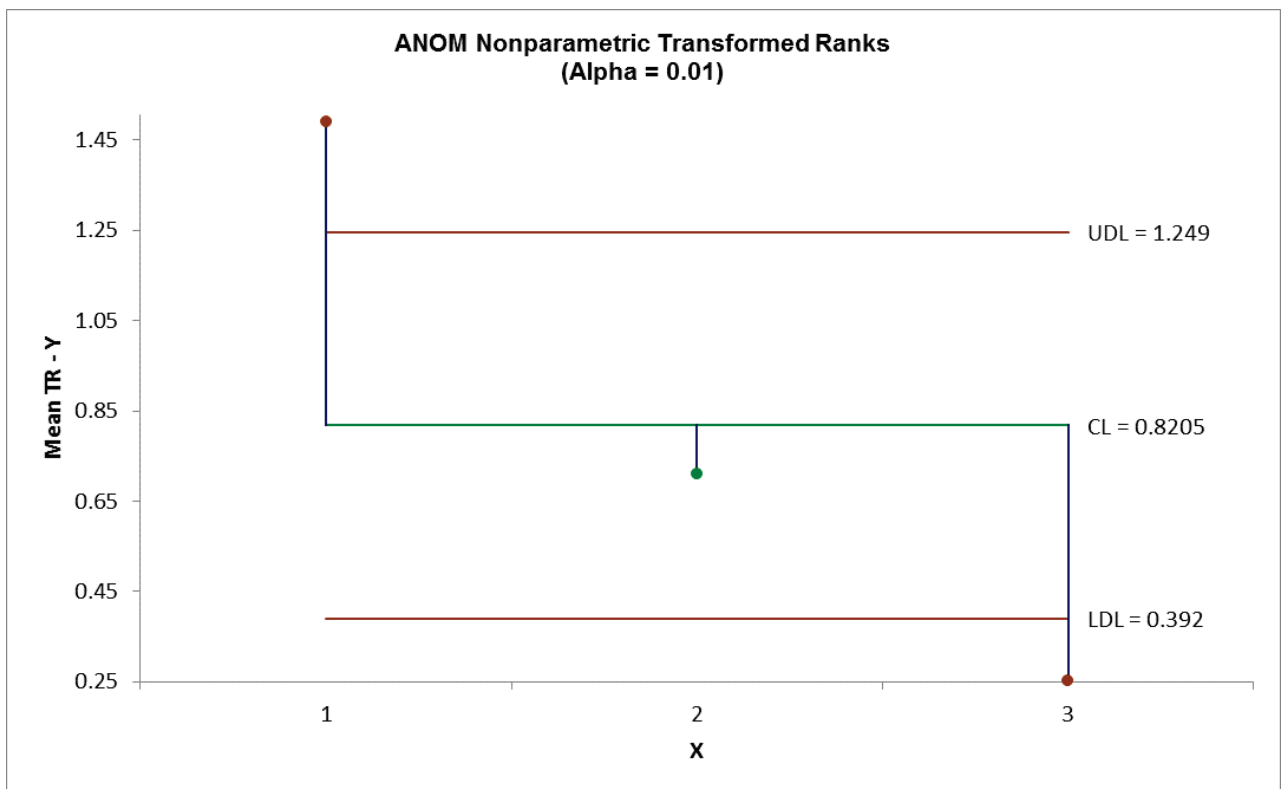
ANOM Nonparametric Transformed Ranks

☒ Stacked Column Format (1 Numeric Data Column & 1 Group Category Column)
☐ Unstacked Column Format (2 or More Numeric Data Columns)

Numeric Data Variable (Y) >>

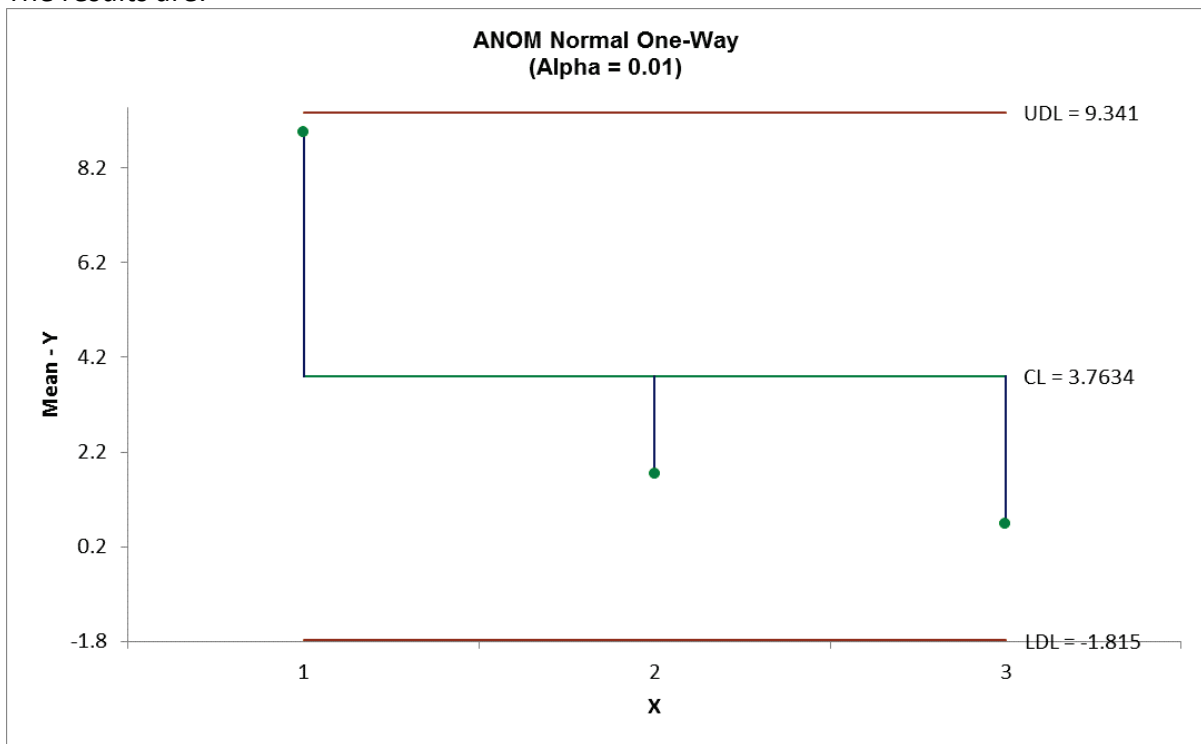
Group Category (X) >>

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:

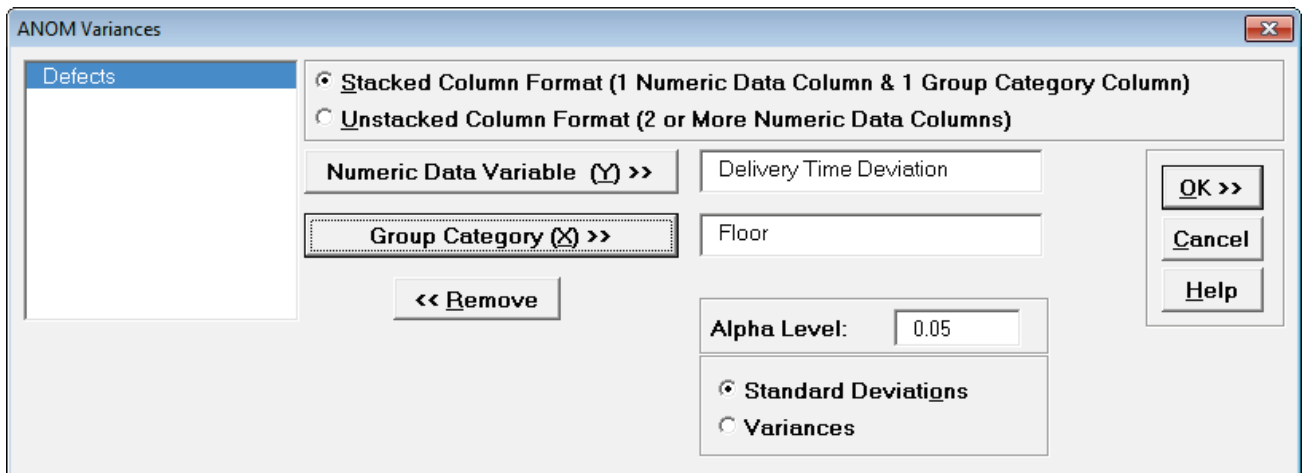


13. ANOM Normal fails to detect the significant differences in mean noted above.

ANOM Variances

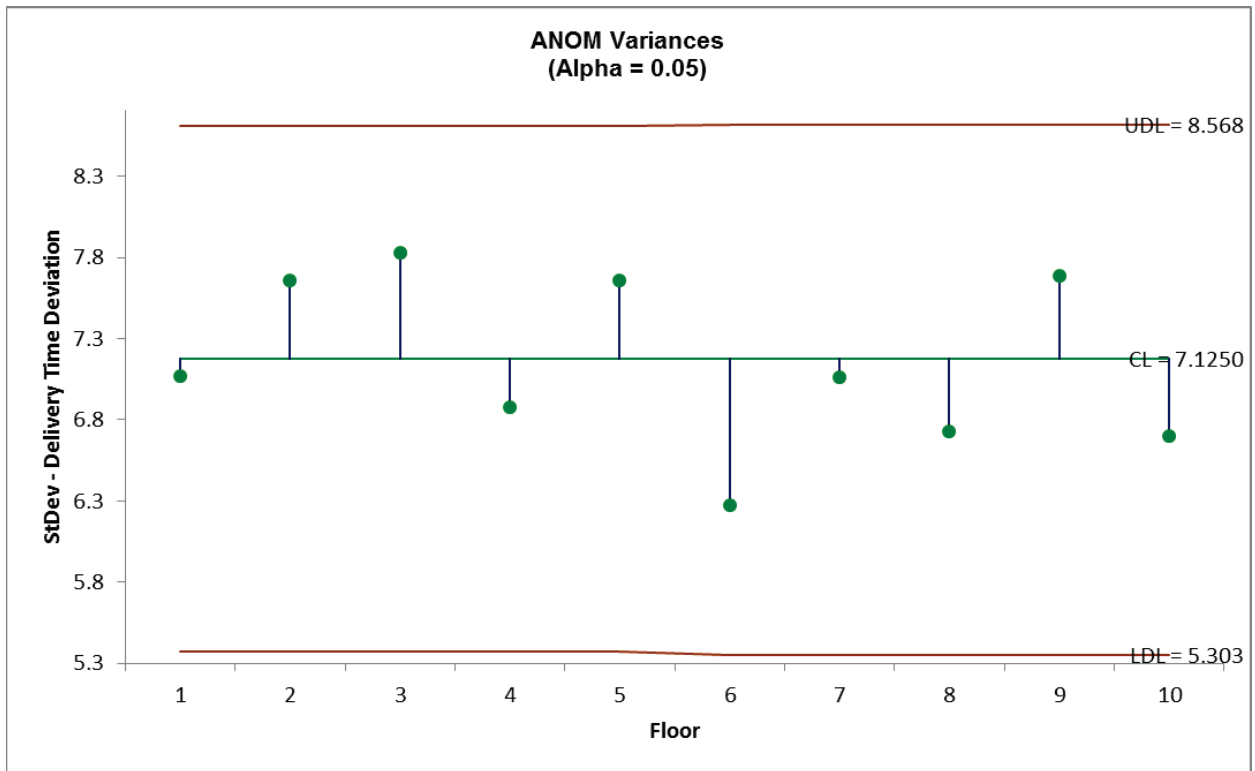
The ANOM Variances chart is complementary to Bartlett's Test for Equal Variance and is also available as an option in the Bartlett's Test dialog. It assumes that the data are normally distributed.

1. Open **Delivery Times.xlsx**, click on **Sheet 1** tab.
2. 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**.
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) >>**; 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](#).

4. Click **OK**. The ANOM Variances chart is shown below:

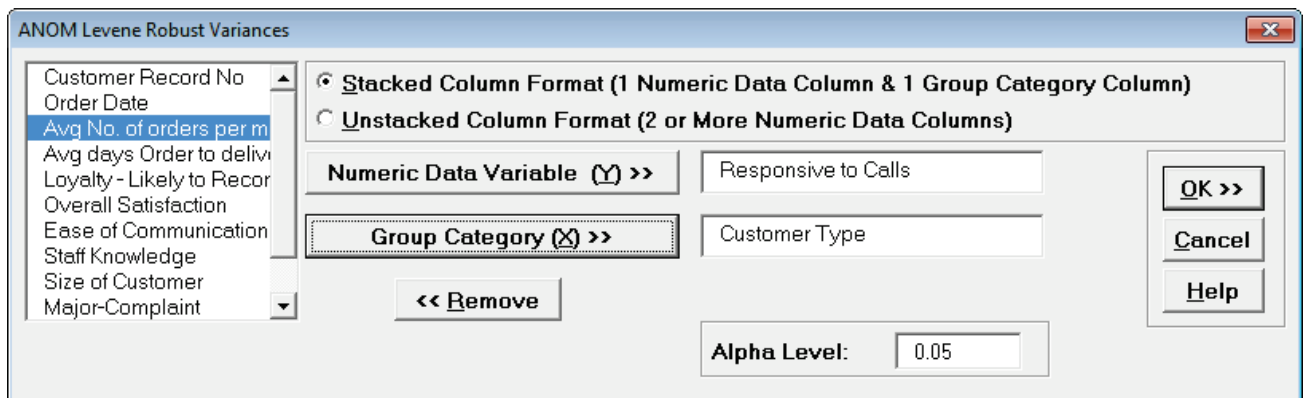


5. The ANOM Variances chart visually shows that none of the group standard deviations are significantly different from the grand mean of all the standard deviations. It is called an ANOM Variances Chart but displays Standard Deviations for ease of interpretation (similar to a Standard Deviation S Control 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

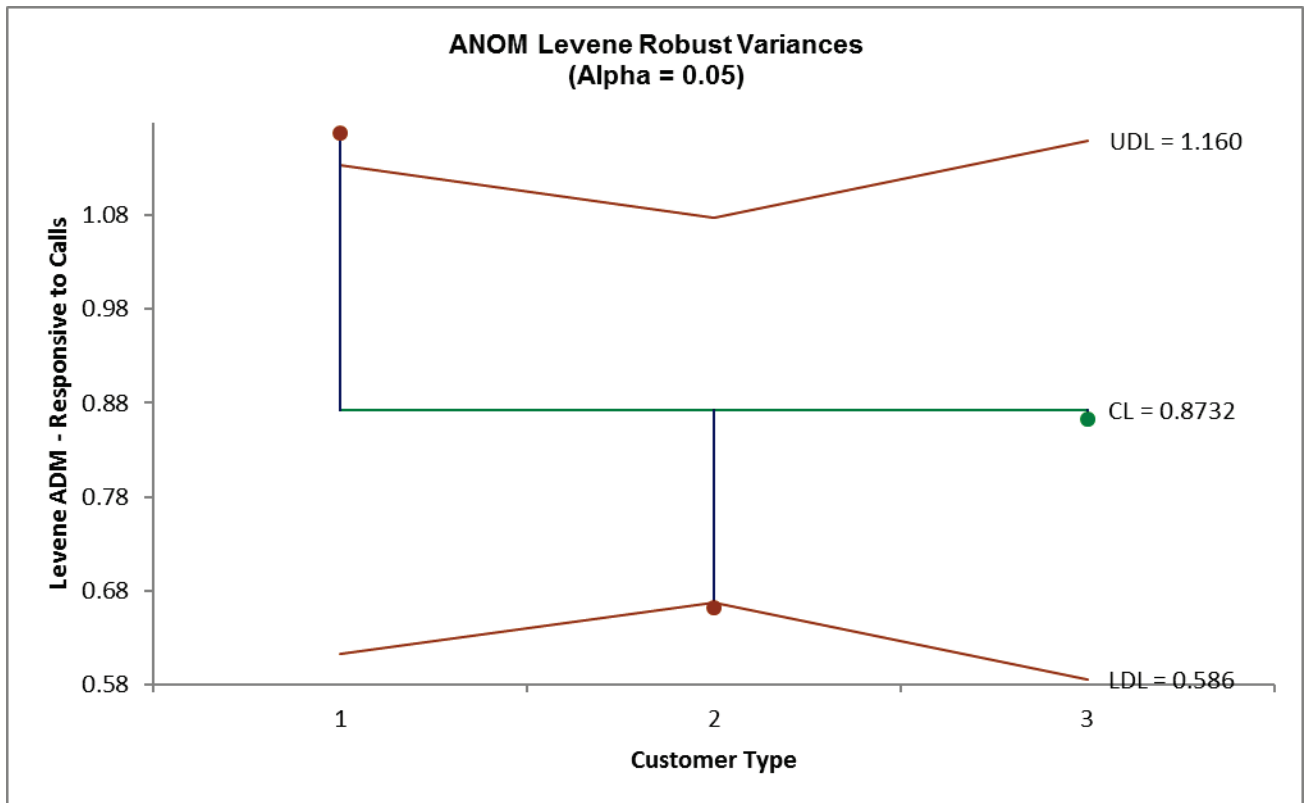
The ANOM Levene Robust Variances chart is complementary to Levene's Test for Equal Variance and is also available as an option in the Levene's Test dialog. Since it uses absolute deviations from the median, it is robust to the assumption of normality.

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.

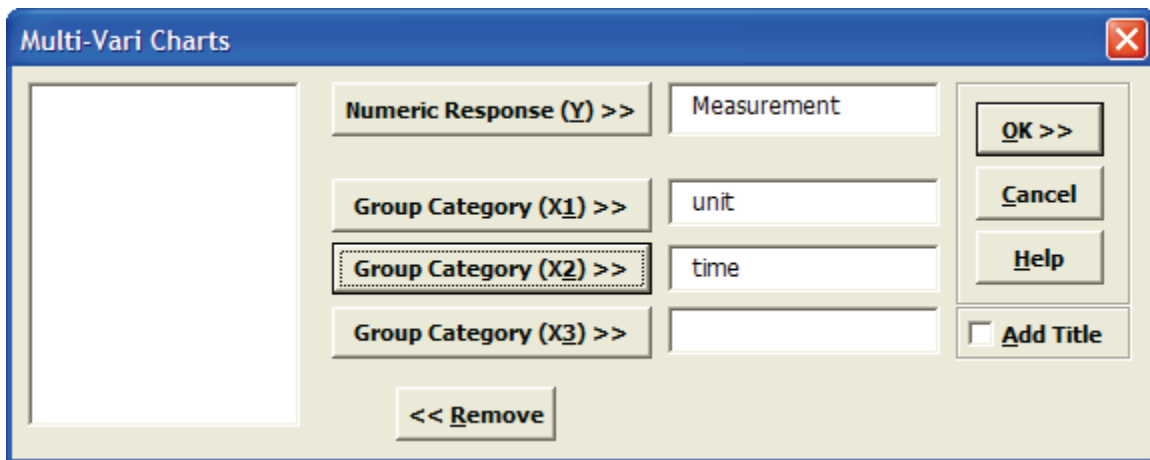
1. Open **Multi-Vari Data.xlsx**, click Sheet **Within**. Select **SigmaXL > Graphical Tools > Multi-Vari Options**.
2. The charts shown will be updated as options are selected. Note that they are for demonstration purposes and are not based on the **Multi-Vari Data.xlsx** data (sample charts are not 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).

General Options	Mean Options	Median Options	Buttons
<input checked="" type="checkbox"/> Range Line	<input checked="" type="radio"/> Mean Options	<input type="radio"/> Median Options	<input type="button" value="Finish >>"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/> <input checked="" type="checkbox"/> Save Defaults
<input checked="" type="checkbox"/> Individual Data Points	<input checked="" type="checkbox"/> Show Means	<input type="checkbox"/> Show Medians	
<input checked="" type="checkbox"/> Min and Max	<input type="checkbox"/> 95% Confidence Interval	<input type="checkbox"/> 25th, 75th Percentile	
<input checked="" type="checkbox"/> Standard Deviation Chart	<input checked="" type="checkbox"/> Connect Means	<input type="checkbox"/> Connect Medians	
	<input checked="" type="checkbox"/> Group Means	<input type="checkbox"/> Group Medians	
	<input type="checkbox"/> Grand Mean	<input type="checkbox"/> Grand Median	

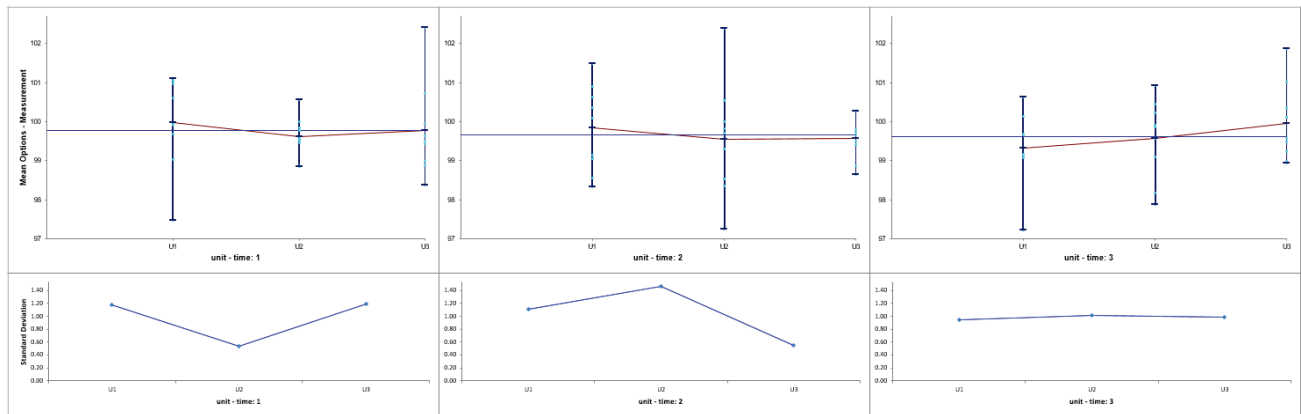
Tip: Multi-Vari Charts can be used to display Confidence Intervals (as we did earlier in Part C). To do this, check the 95% Confidence Interval.

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.

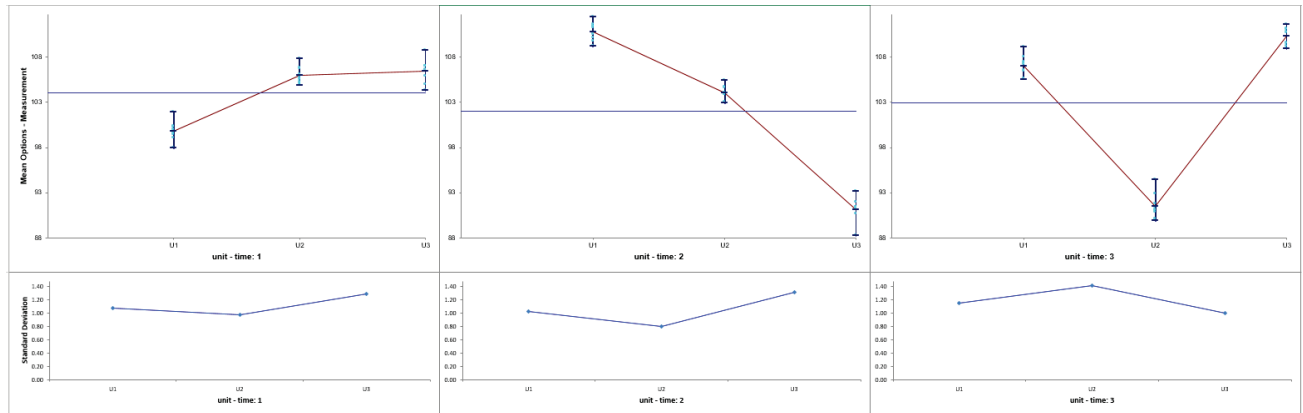
6. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1) >>**; select *time*, click **Group Category (X2) >>**.



7. Click **OK**. Resulting Multi-Vari Chart illustrating dominant “Within Unit” Source of Variation:



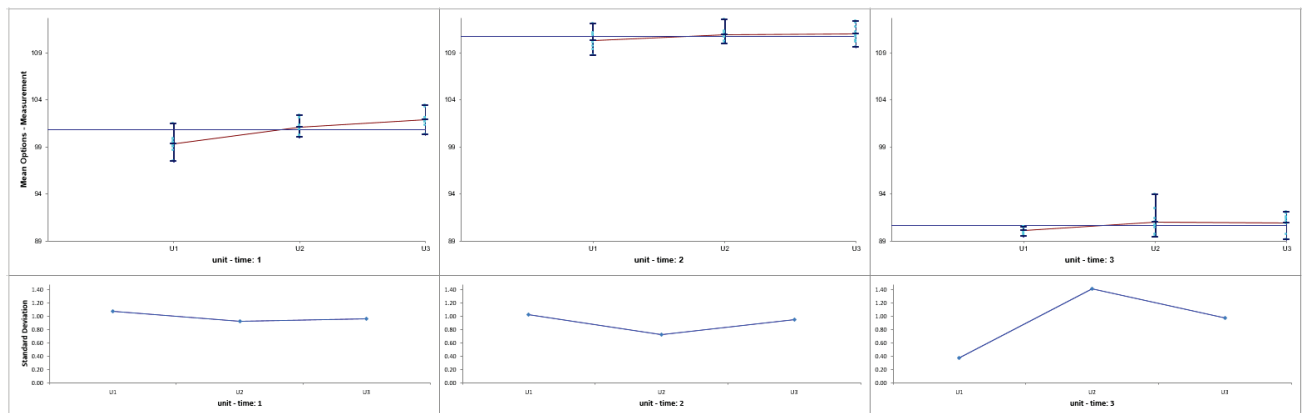
8. Using **Multi-Vari Data.xlsx**, click Sheet **Between**. Select **SigmaXL > Graphical Tools > Multi-Vari Charts**. Check **Use Entire Data Table**. Click **Next**.
9. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1) >>**; select *time*, click **Group 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**.

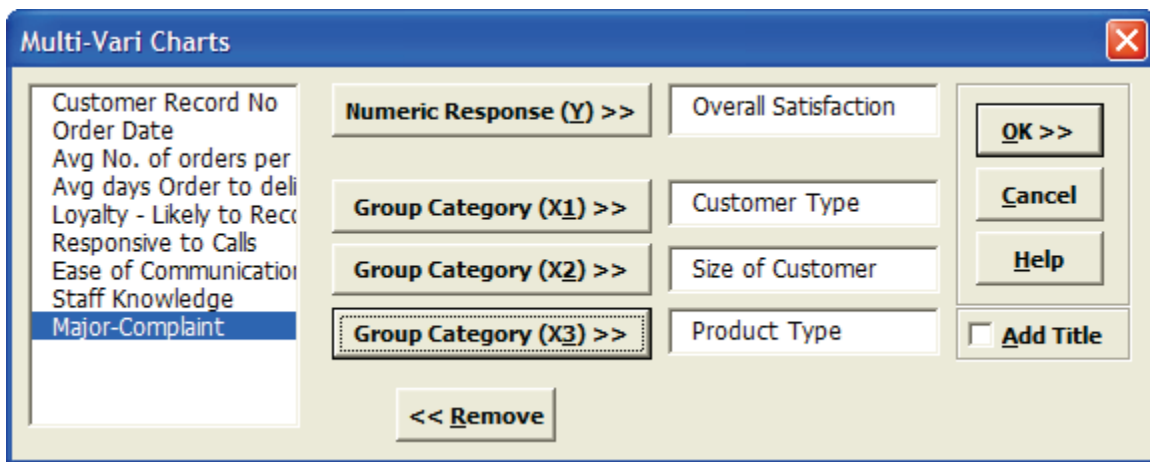
12. Select *Measurement*, click **Numeric Response (Y) >>**; select *unit*, click **Group Category (X1) >>**; select *time*, click **Group Category (X2) >>**.

13. Click **OK**. Resulting Multi-Vari Chart illustrating dominant “Over Time” Source of Variation:

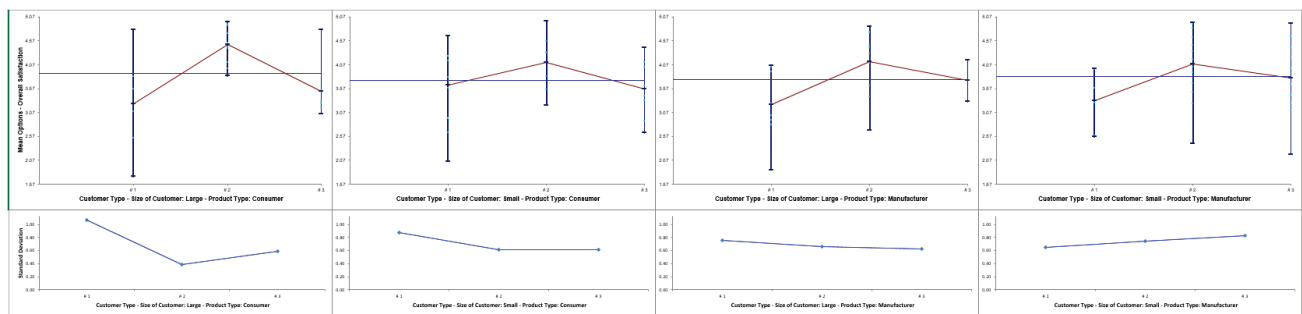


14. Open **Customer Data.xlsx**, click **Sheet 1**. Select **SigmaXL > Graphical Tools > Multi-Vari Charts**.

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:



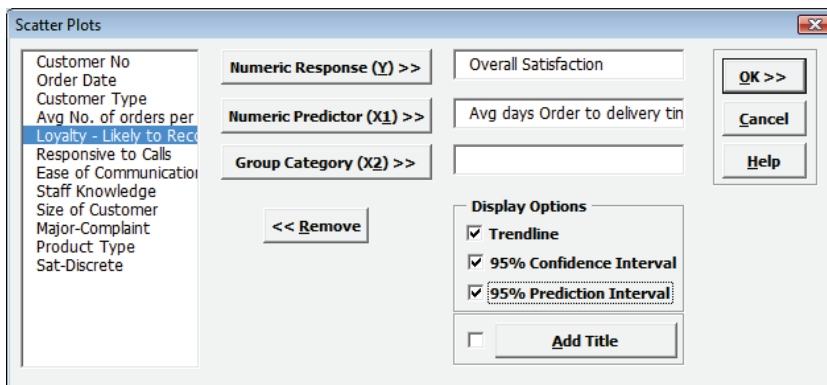
Examining this Multi-Vari chart reveals that the dominant Source of Variation is “within” Customer Type, followed by “between” Customer Type. Furthermore, it would be worthwhile to examine the combination of Customer Type 2, Customer Size Large, and Product Type Consumer.

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.

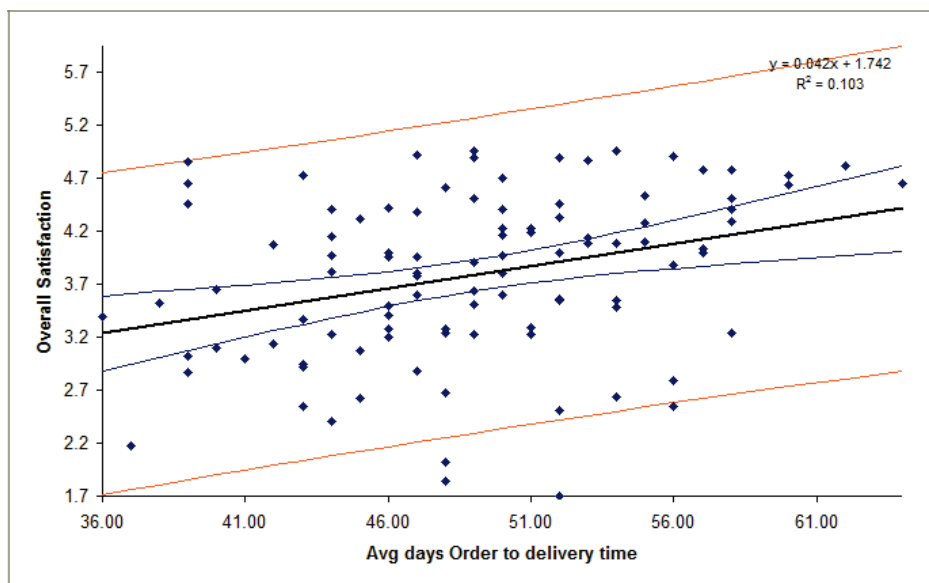
Part R – Scatter Plots

Scatter Plots

1. Open **Customer Data.xlsx**. Click **Sheet 1** Tab. Click **SigmaXL > Graphical Tools > Scatter Plots**; if necessary, click **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:

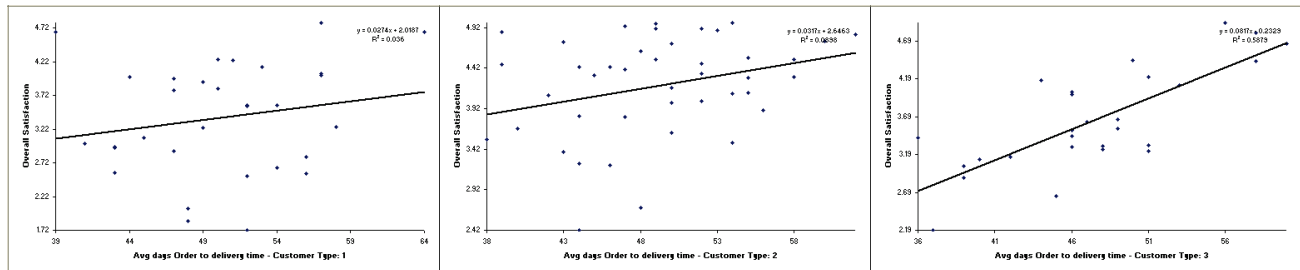


3. Click **OK**. The resulting Scatter Plot is shown with equation, trendline, 95% confidence interval (blue lines – for a given X value this is the 95% confidence interval for predicted mean Overall Satisfaction) and 95% prediction interval (red lines – for a given X value this is the 95% confidence 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**:

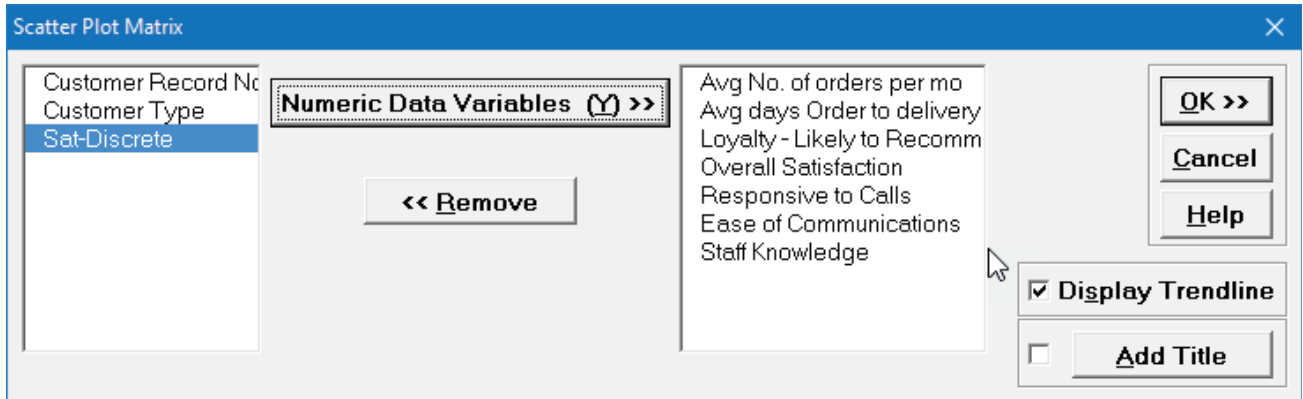


Clearly, according to the analysis, Customer Type 3 is happier if orders take longer! But, does this make sense? Of course not! Customer Sat scores should not increase with Order to Delivery time. What is happening here? This is a coincidental situation. Something else is driving customer satisfaction. We will now look at the Scatter Plot Matrix to help us investigate other factors influencing Customer Satisfaction.

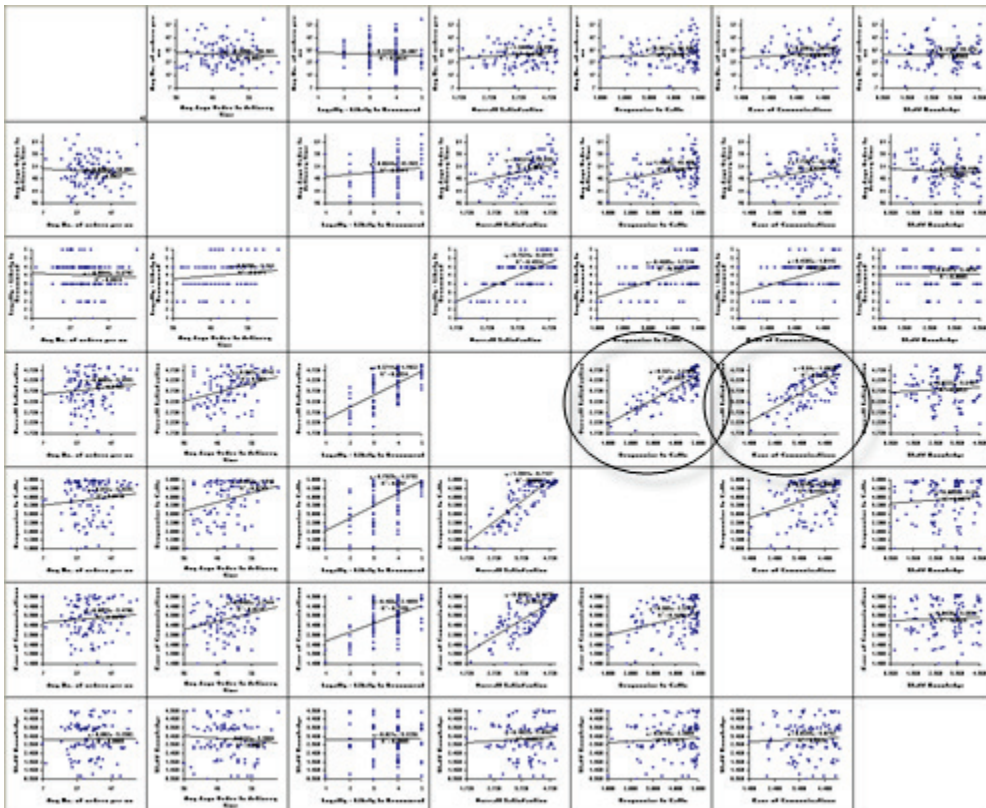
Tip: Be careful when interpreting scatter plots (and other statistical tools): Y versus X correlation or statistical significance does not always mean that we have a causal relationship. Umbrella sales are highly correlated to traffic accidents, but we cannot reduce the rate of traffic accidents by purchasing fewer umbrellas! The best way to validate a relationship is to perform a Design of Experiments (see Improve Phase).

Scatter Plot 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.

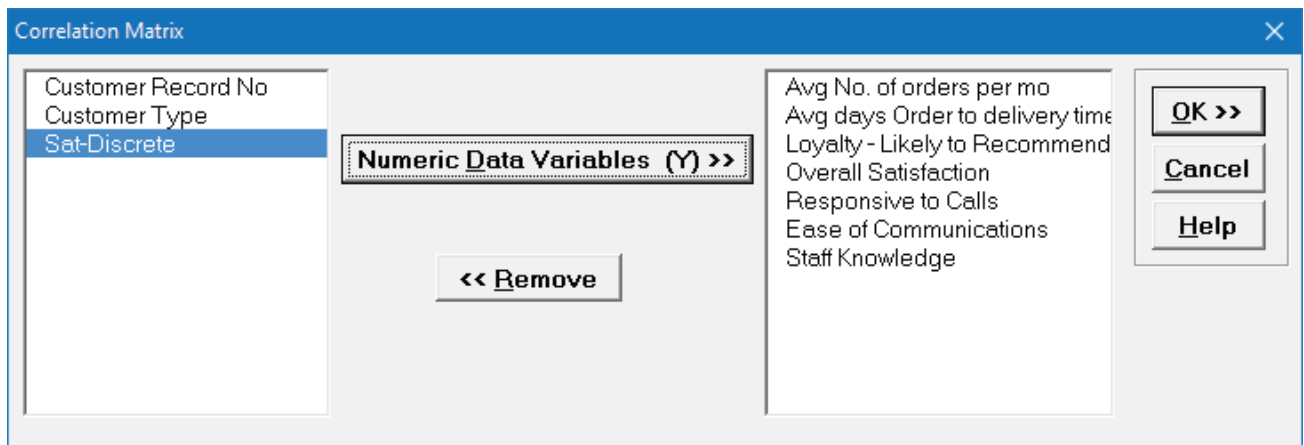
Part S – Correlation Matrix

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
$0.9 \leq r \leq 1$	> 80 %	Strong
$0.7 \leq r < 0.9$	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	0.3218	0.2725	0.2681	-0.0781
Loyalty - Likely to Recommend			1	0.6598	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	0.0051	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	0.2513	-0.0828
Loyalty - Likely to Recommend			1	0.6167	0.5503	0.4071	-0.0190
Overall Satisfaction				1	0.7782	0.7009	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:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n$$

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 is useful when you have strong *a priori* reasons to believe that $Y = 0$ when the X or X's are equal to 0 and the relationship is linear. An example would be $Y = \text{automobile fuel consumption}$, $X = \text{automobile weight}$.

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 we need to focus on these two X factors to improve customer satisfaction.

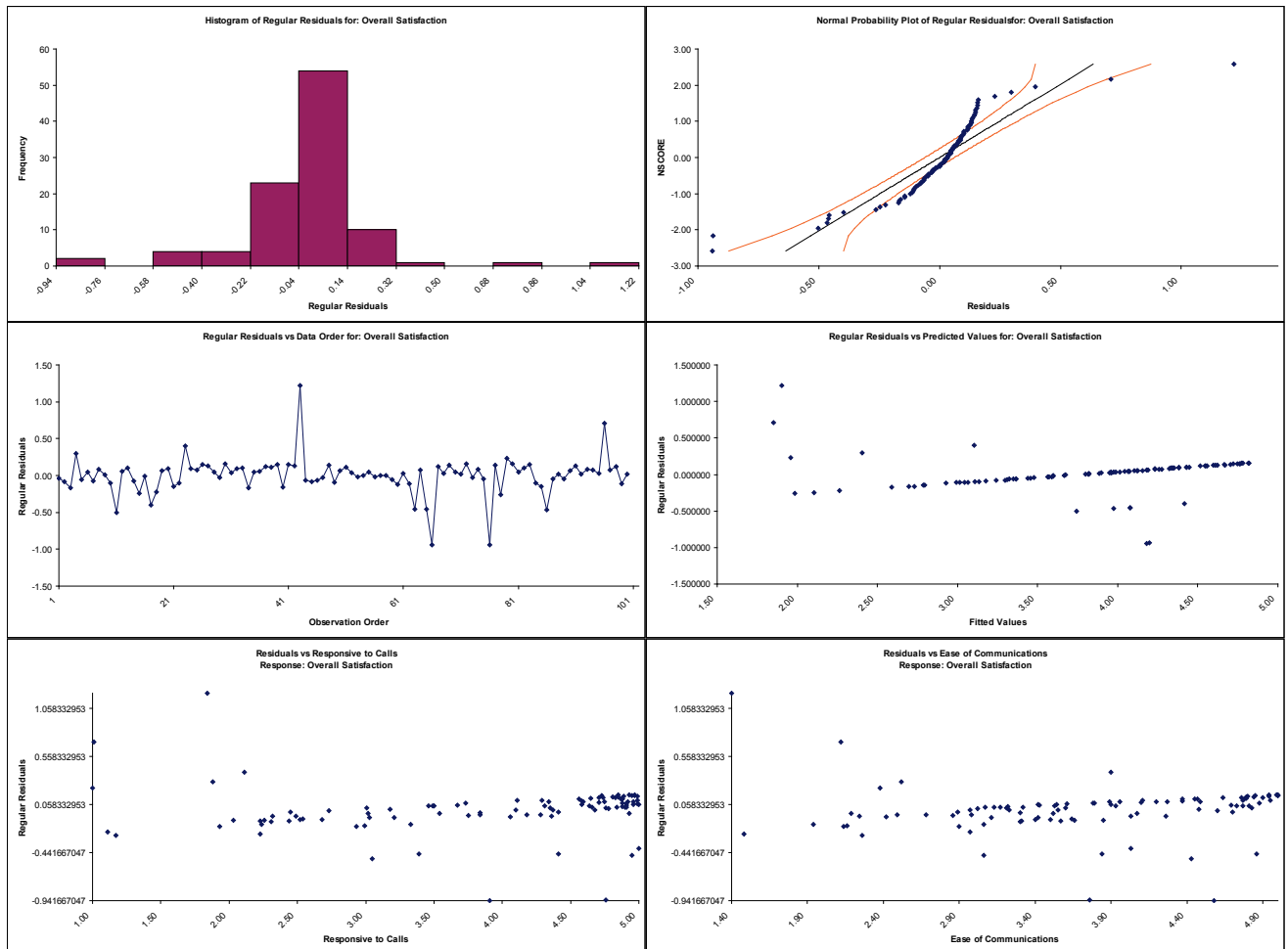
- The variance inflation factor (VIF) and Tolerance scores are used to measure multicollinearity or correlation among predictors. VIF = 1 indicates no relation among predictors (which is highly desirable – you will see examples of this in Design of Experiments); VIF > 1 indicates that the predictors are correlated; VIF > 5 indicates that the regression coefficients are strongly correlated and this will lead to poor estimates of the coefficients. Tolerance = 1/VIF.
- The Durbin-Watson Test is used to determine if the residuals are 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.
- 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:

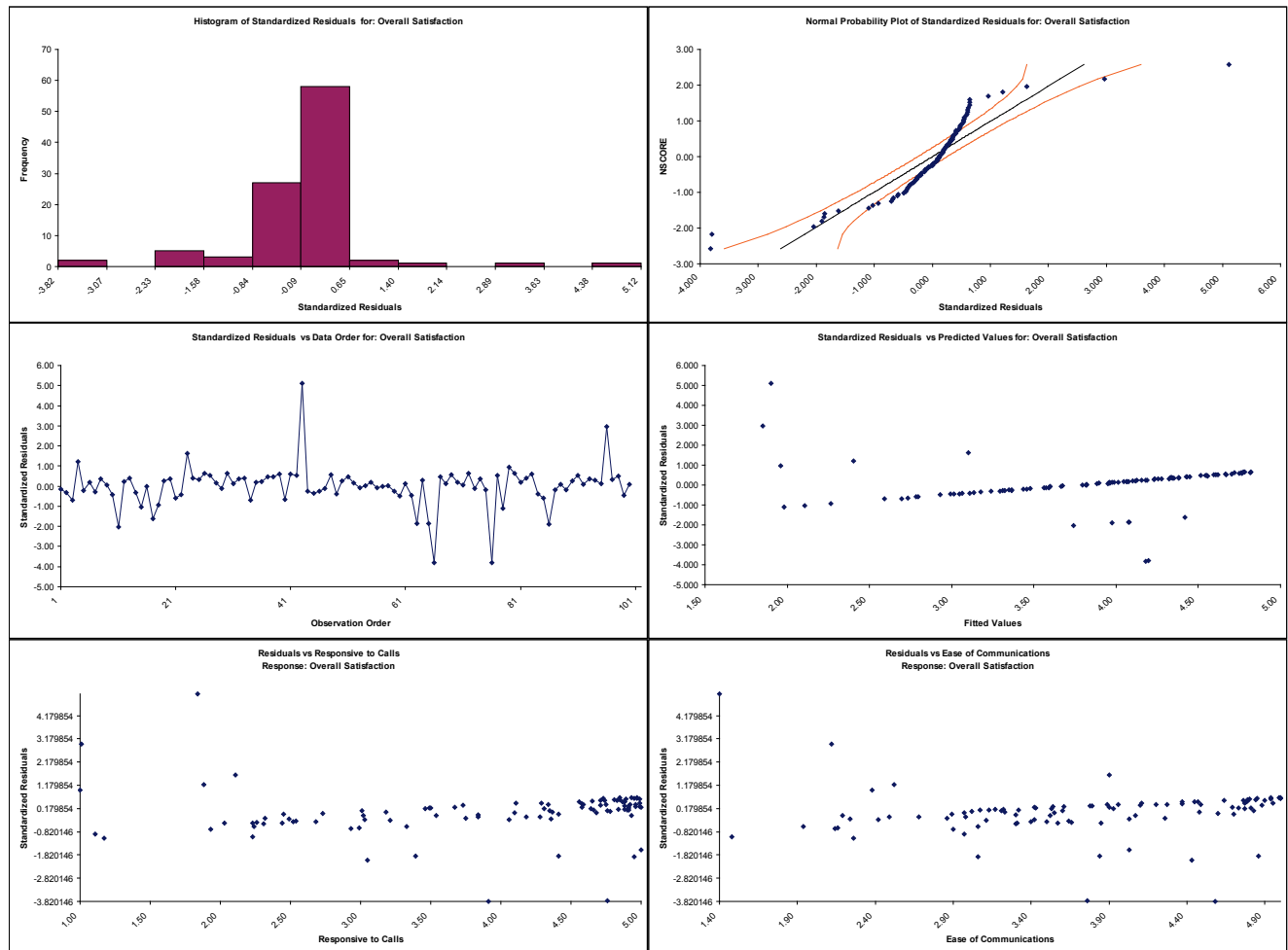


Residuals are the unexplained variation from the regression model ($Y - \hat{Y}$). We expect to see the residuals normally distributed with no obvious patterns in the above graphs. Clearly this is not the case here, with the Residuals versus Predicted Values indicating there is likely some other X factor 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; 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 Predictor (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)**:

Multiple Regression

Customer Record No
Order Date
Avg No. of orders per
Avg days Order to deli
Loyalty - Likely to Rec
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

Numeric Response (Y) >> Overall Satisfaction

Continuous Predictors (X) >>
(Numeric Data) Responsive to Calls
Ease of Communications

Categorical Predictors (X) >>
(Text or Numeric Discrete Data) Customer Type

<< Remove

☒ Fit Intercept

☐ Display Residual Plots

Regular

OK >>
Cancel
Help

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	2.492000684	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” (i.e. Customer Type 1 corresponds to Coded Variable 1 = 0 and Coded Variable 2 = 0; Customer Type 2 has Coded Variable 1 = 1 and Coded Variable 2 = 0; and Customer Type 3 has Coded Variable 1 = 0 and Coded Variable 2=1). Hence Customer Type 1 becomes the “hidden” or reference value. Other statistical software packages may use a different coding scheme based on -1,0,+1 instead of 0,1. 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 for 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 .08, 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 criteria 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 – 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:

$$\ln(P_Y/(1-P_Y)) = b_0 + b_1*X_1 + b_2*X_2 + \dots + b_n*X_n$$

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 *Sat-Discrete*, click **Binary Response (Y) >>**; select *Responsive to Calls* and *Ease of Communications*, click **Continuous Predictors (X) >>**; select *Customer Type*, click **Categorical Predictors (X) >>**.

Binary Logistic Regression

Customer Record No
Order Date
Avg No. of orders per
Avg days Order to deli
Loyalty - Likely to Recc
Overall Satisfaction
Staff Knowledge
Size of Customer
Major-Complaint
Product Type

☒ **Binary Response (Y)**
☐ Response Count (Y) / Sample Size (Trials)

OK >>
Cancel
Help

Binary Response (Y) >> Sat-Discrete

Continuous Predictors (X) >> Responsive to Calls
Ease of Communication
(Numeric Data)

Categorical Predictors (X) >> Customer Type
(Text or Numeric Discrete Data)

<< Remove

0
1

Reference Event >> 1

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: $\ln(P/(1-P)) = (-16.647) + (2.454) * \text{Responsive to Calls} + (2.366) * \text{Ease of Communications} + (0.546089) * \text{Customer Type}_2$
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		Row Total
	$\hat{Y} = 0$	$\hat{Y} = 1$	
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).

Binary Logistic Regression

Customer Record No
Order Date
Avg No. of orders per
Avg days Order to deli
Loyalty - Likely to Recd
Overall Satisfaction
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Customer Type

☒ Binary Response (Y)
☐ Response Count (Y) / Sample Size (Trials)

Binary Response (Y) >> Sat-Discrete

Continuous Predictors (X) >>
(Numeric Data)

Categorical Predictors (X) >>
(Text or Numeric Discrete Data)

<< Remove

Reference Event >> 0 1

OK >> Cancel Help

8. Click **OK**. The resulting Binary Logistic report is shown:

SigmaXL: Analyze Phase Tools

Binary Logistic Regression Model: $\ln(P_y/(1-P_y)) = [-17.703] + [2.462] * \text{Responsive to Calls} + [2.589] * \text{Ease of Communications}$
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:

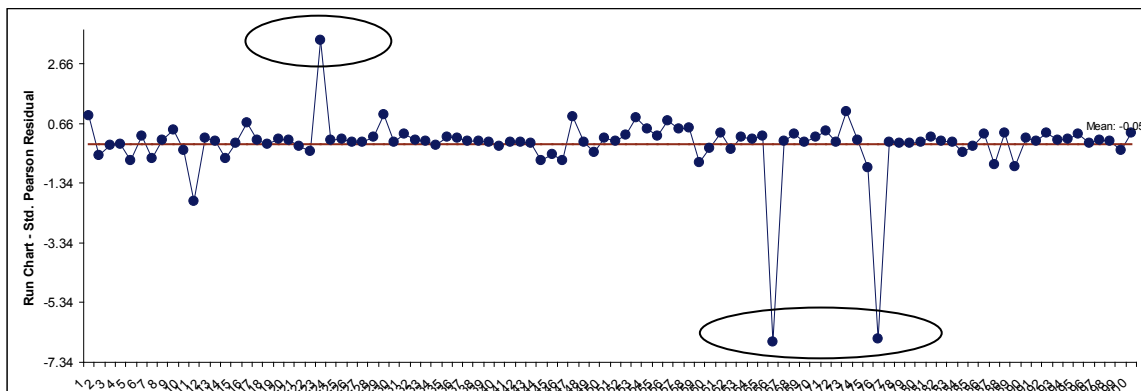
Log-Likelihood	-21.275
Test that all slope coefficients are equal to zero:	
Likelihood Ratio Chi-Square (G)	84.285
DF	2
P-Value	0.0000
McFadden's Pseudo R-Square	66.45%
Goodness-of-Fit Tests (P-Value < .05 indicates Lack-of-Fit):	
Pearson Residuals Chi-Square	112.63
DF	97
P-Value	0.1325
Deviance Residuals 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		Row Total
	$\hat{y} = 0$	$\hat{y} = 1$	
Y = 0	30	3	33
Y = 1	1	66	67
Column Total	31	69	100
Percent Correctly Predicted:	96.00%		

- 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.
- 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	

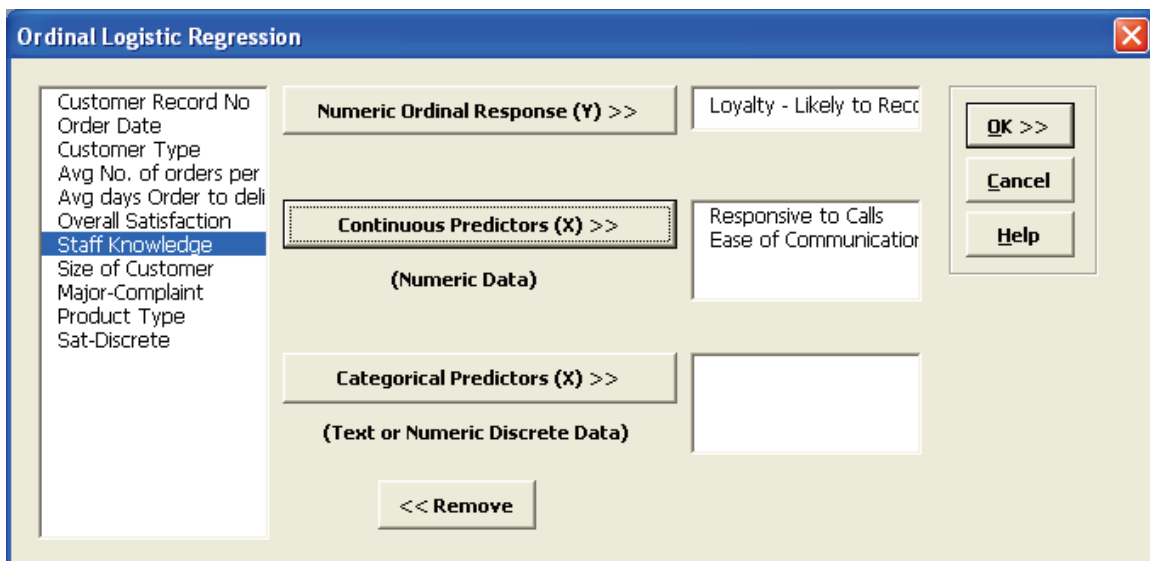
If we wanted to select Customer Type 1, enter a 0 for both **Customer Types 2 and 3**. Customer Type 1 is the hidden reference value.

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: $\ln(\text{Py}/(1-\text{Py})) = \text{Constant} + (-1.011305417) * \text{Responsive to Calls} + (-0.790025) * \text{Ease of Communications}$
 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					Row Total
	Y = 1	Y = 2	Y = 3	Y = 4	Y = 5	
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%					

- 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.
- 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.
- 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%.
- 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 5, we would expect typical loyalty scores of 4 (53 %) with some at 3 (34%) and 5 (9%).

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

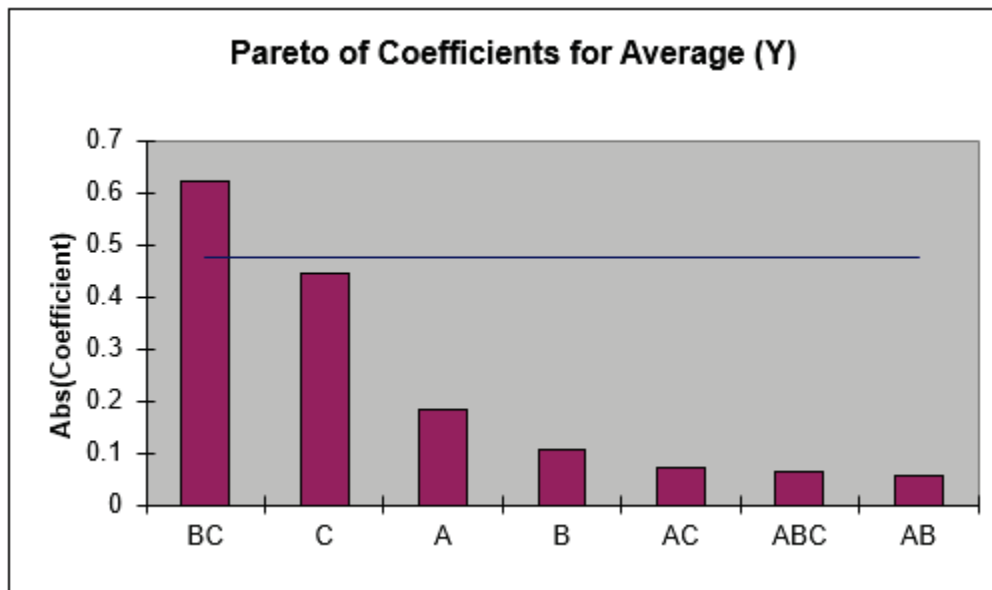
After entering the template data, main effects and interaction plots may be created by clicking **SigmaXL > Basic DOE Templates > Main Effects & Interaction Plots**. The DOE template must be the active worksheet.

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.
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.

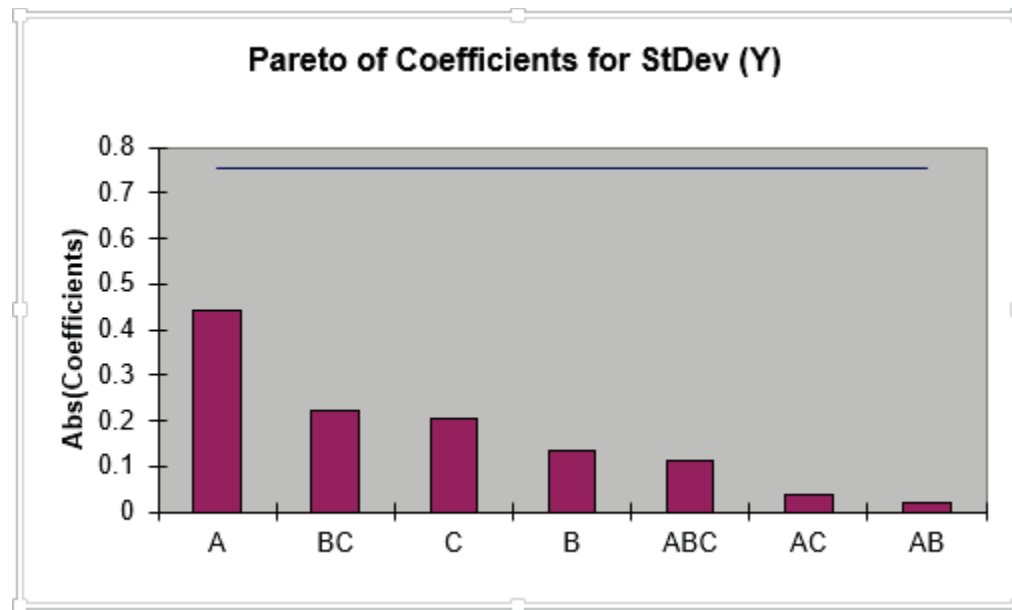
10. If the Responses are replicated, SigmaXL draws the blue line on the Pareto Chart using an estimate of experimental error from the replicates. If there are no replicates, an estimate called Lenth's Pseudo Standard Error is used.
11. If the 95% Confidence line for coefficients were to be drawn using Lenth's method, the value would be 0.409 as given in the table:

Lenth's Pseudo Standard Error (PSE) Analysis for Unreplicated Data:

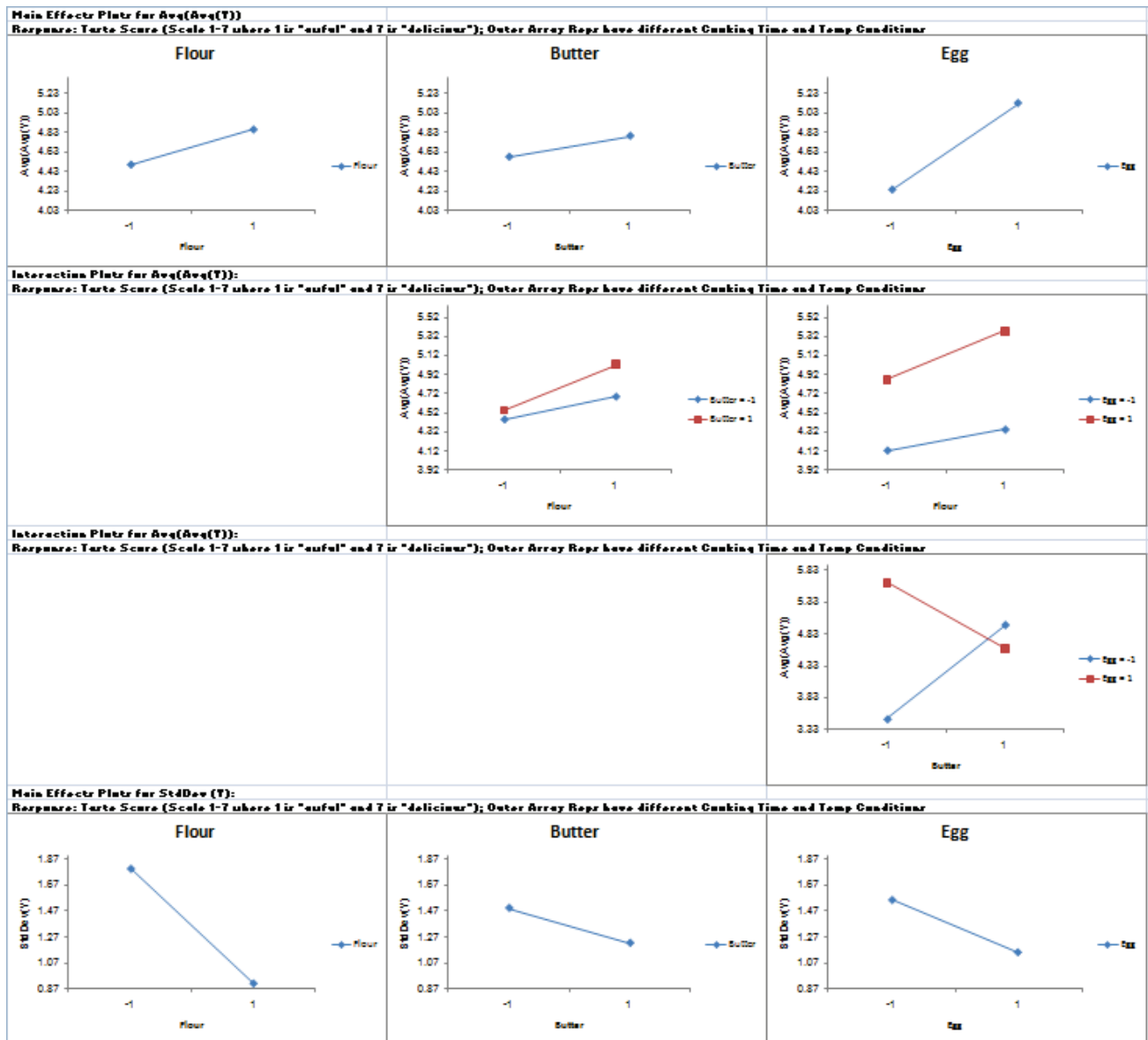
Lenth's PSE for Coefficients:	0.10875
Lenth's Margin of Error for Coefficients (95% Conf. Level):	0.40935
Lenth's Margin of Error for Effects (95% Conf. Level):	0.8187

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. The significance of this result will be demonstrated using Regression analysis.

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: $\text{Average (Y)} = (4.683) + (0.1075) * B + (0.442500) * C + (-0.622500) * BC$

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:

Multiple Regression

Numeric Response (Y) >> StDev (Y) **OK >>**

Continuous Predictors (X) >> A **Cancel**

(Numeric Data)

Categorical Predictors (X) >>

(Text or Numeric Discrete Data)

<< Remove

☒ Fit Intercept

☒ Display Residual Plots

Regular

26. Click **OK**. The resulting regression report is shown:

Multiple Regression Model: $\text{StDev (Y)} = (1.358) + (-0.442296) * A$

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.

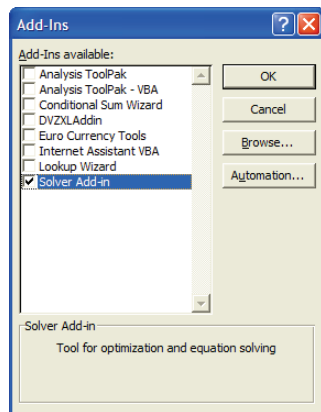
28. Click on the Sheet **Three-Factor 8-Run DOE**. At the **Predicted Output for Y**, enter 1 for *Flour*. We are setting this as a constraint, because Flour = +1 minimizes the Standard Deviation. Reset the *Butter* and *Egg* to 0 as shown:

Factor	Factor Name	Low	High
A	Flour	-1	1
B	Butter	-1	1
C	Egg	-1	1

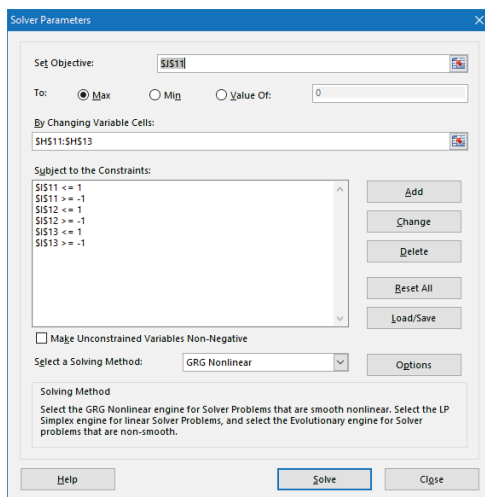
Predicted Output for Y:

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 2007 click **Office Button | Excel Options | Add-Ins > Manage Excel Add-Ins**, click **Go**; in Excel 2010: **File | Options**. 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 | Analysis | 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: $I11=1$; $I12 \geq -1$; $I12 \leq 1$; $I13 \geq -1$; $I13 \leq 1$.

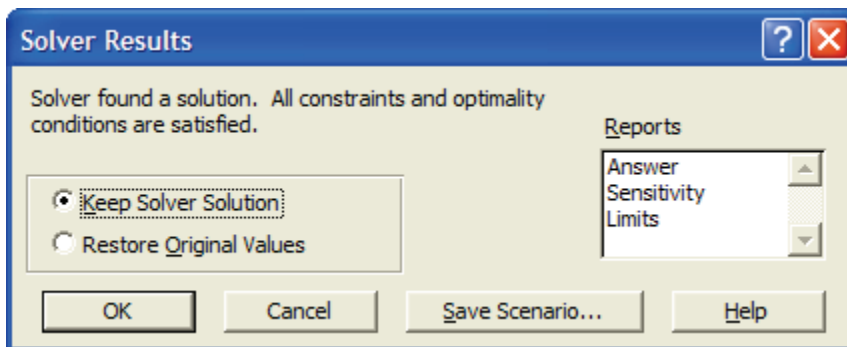
31. Click **Solve**. The solver results are given in the **Predicted Output for Y** as *Butter* = -1 and *Egg* = 1.

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.

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.

2-Level Factorial/Screening Design of Experiments

Number of Factors: 3

Select Design: 8-Run, 2**3, Full-Factorial

Number of Replicates: 1

Power Information (based on # of runs and replicates):

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).

Number of Blocks: 1

Number of Center Points per Block: 0

☒ Randomize Runs

Factor Names and Level Settings:

	Name	Low	High
A:	A	-1	1
B:	B	-1	1
C:	C	-1	1

Number of Responses: 1

Response Name

Y1: Y1

OK>> Cancel Help Reset

5. Next we will review the Power Information to help us determine the number of replicates. 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).

Note: the power calculations require an estimate of experimental error, hence an assumption of 3 center points is used.

We would like to have a reasonable chance (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:

2-Level Factorial/Screening Design of Experiments

Number of Factors: 3

Select Design: 8-Run, 2**3, Full-Factorial

Number of Replicates: 2

Power Information (based on # of runs and replicates):
 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).

Number of Blocks: 1

Number of Center Points per Block: 0

☒ Randomize Runs

Factor Names and Level Settings:

	Name	Low	High
A:	Pull Back Angle	160	180
B:	Stop Pin	2	3
C:	Pin Height	2	3

Number of Responses: 1

Y1: Response Name
Distance

Buttons: OK>>, Cancel, Help, Reset

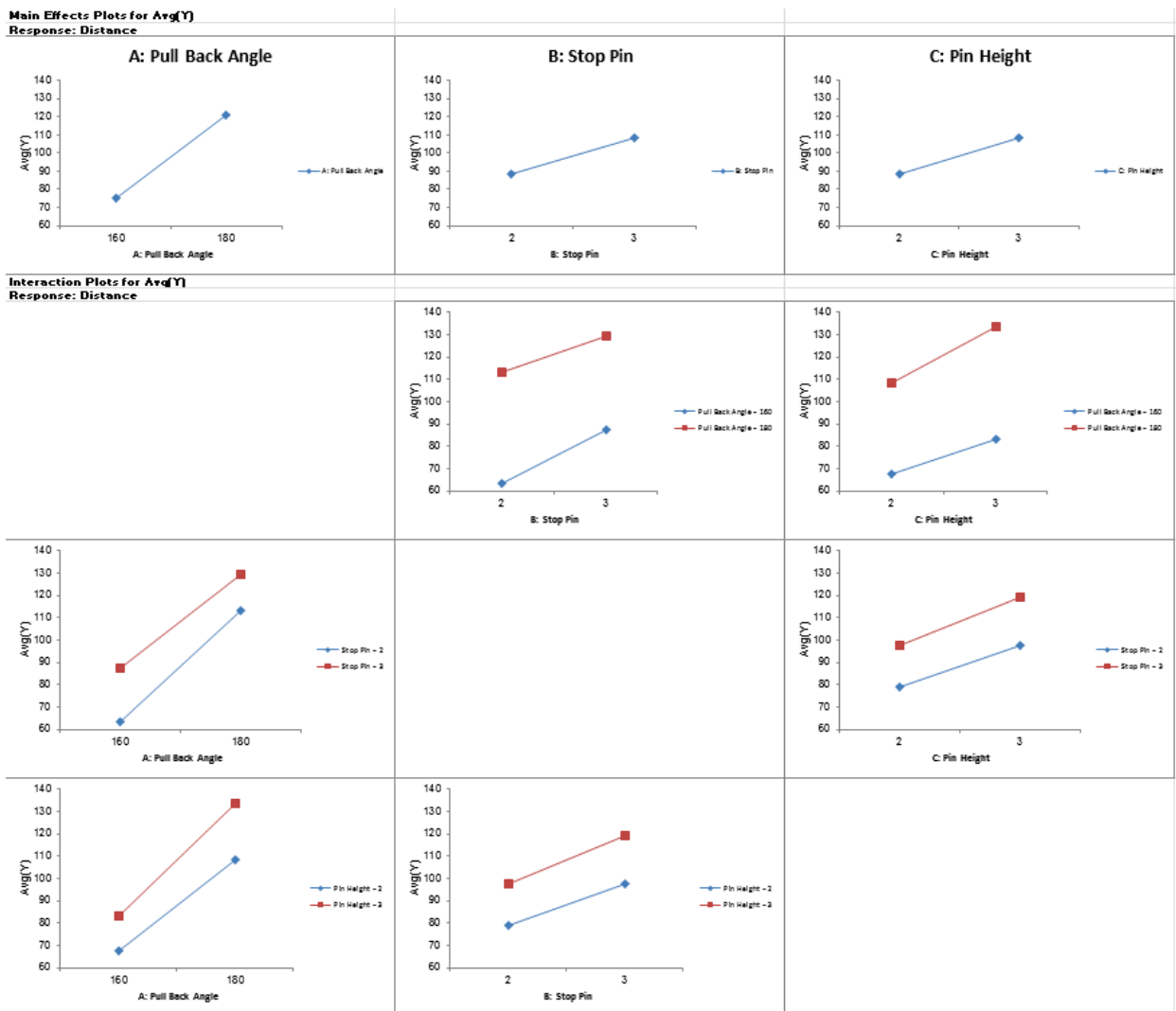
7. Click **OK**. The following worksheet is produced:

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**.

Analyze 2-Level Factorial/Screening Design

Available Responses:

Selected Responses: Distance

Responses (Y) >>

OK >> **Cancel** **Help**

Available Model Terms:

Selected Model Terms: A: Pull Back Angle
B: Stop Pin
C: Pin Height
AB
AC
BC
ABC

Model Terms >>

<< Remove

<< Remove All

Alpha for Pareto Chart: 0.05

☒ **Show Residual Plots**

Regular

Include: Default (All) Terms

☐ **Include Blocks**

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):

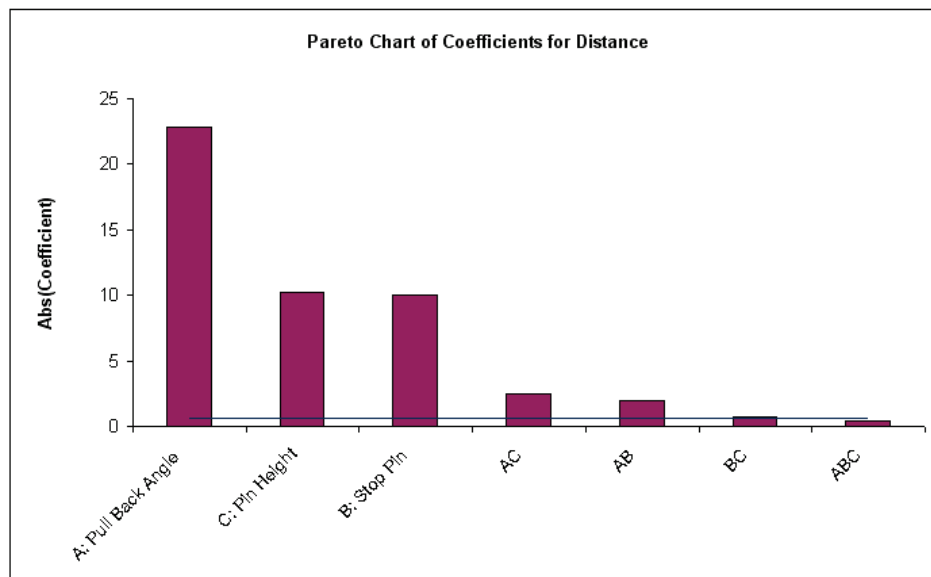
Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	98.3125	0.257694102	381.51	0.0000		
A: Pull Back Angle	22.8125	0.257694102	88.526	0.0000	1	1
B: Stop Pin	10.0625	0.257694102	39.048	0.0000	1	1
C: Pin Height	10.1875	0.257694102	39.533	0.0000	1	1
AB	-1.9375	0.257694102	-7.519	0.0001	1	1
AC	2.4375	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:

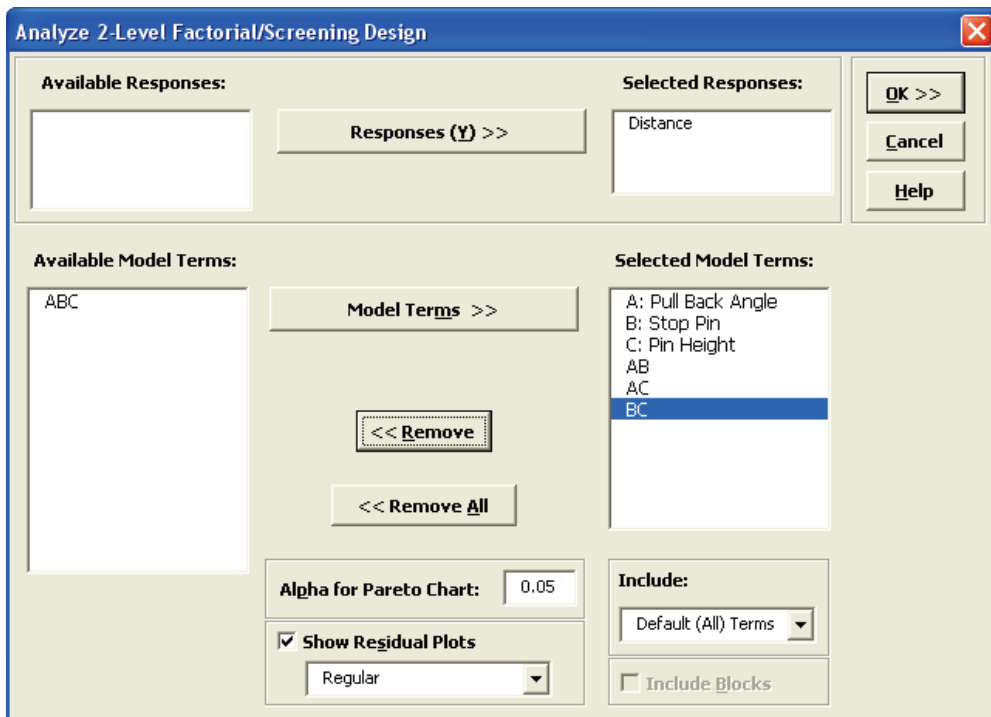
Source	DF	SS	MS	F	P
Model	7	11773	1681.8	1582.9	0.0000
Error	8	8.500	1.0625		
Pure Error	8	8.500	1.0625		
Total (Model + Error)	15	11781	785.43		

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):

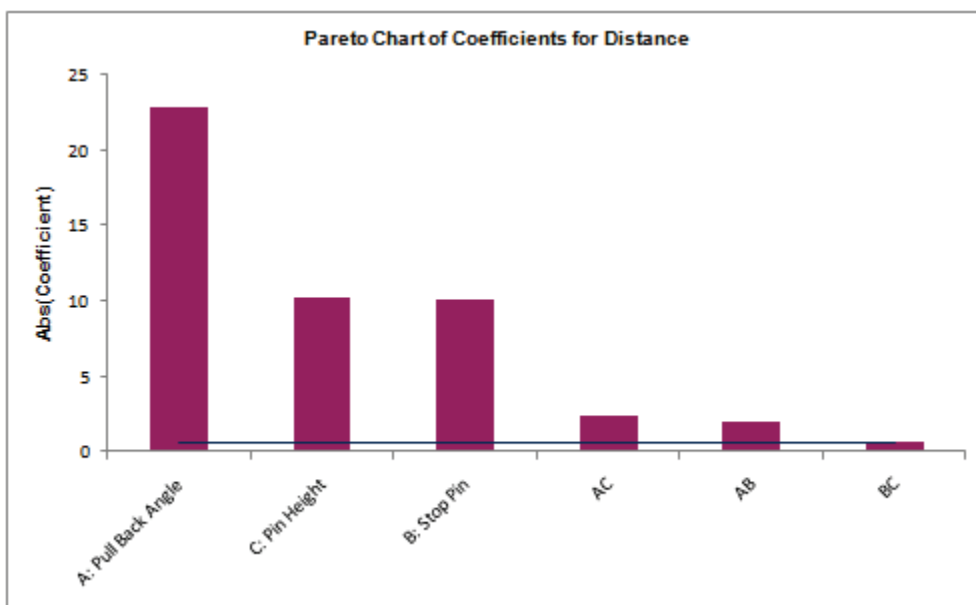
Term	Coefficient	SE Coefficient	T	P	VIF	Tolerance
Constant	98.3125	0.283363969	346.95	0.0000		
A: Pull Back Angle	22.8125	0.283363969	80.506	0.0000	1	1
B: Stop Pin	10.0625	0.283363969	35.511	0.0000	1	1
C: Pin Height	10.1875	0.283363969	35.952	0.0000	1	1
AB	-1.9375	0.283363969	-6.837	0.0001	1	1
AC	2.4375	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		

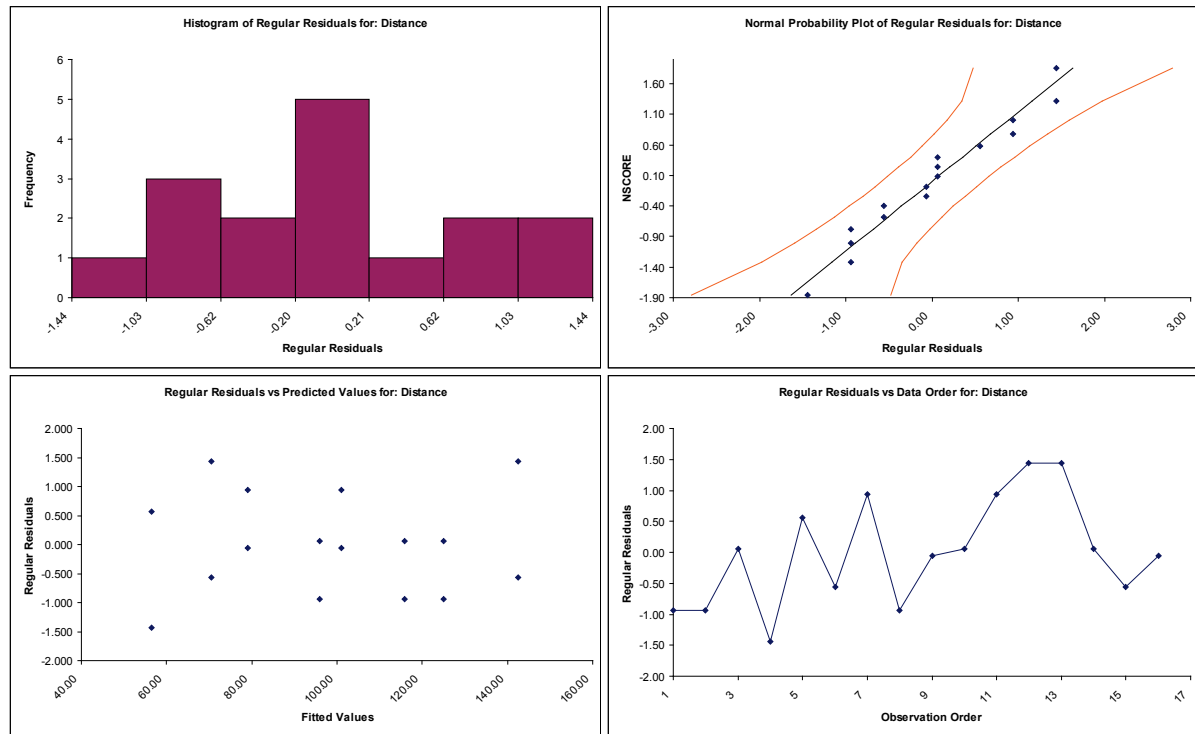
Durbin-Watson Test for Autocorrelation in Residuals:

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. The Residual Plots also look very good:



19. Scroll up to the **Predicted Response Calculator**. Enter the predicted values shown. These initial settings were determined by trial and error.

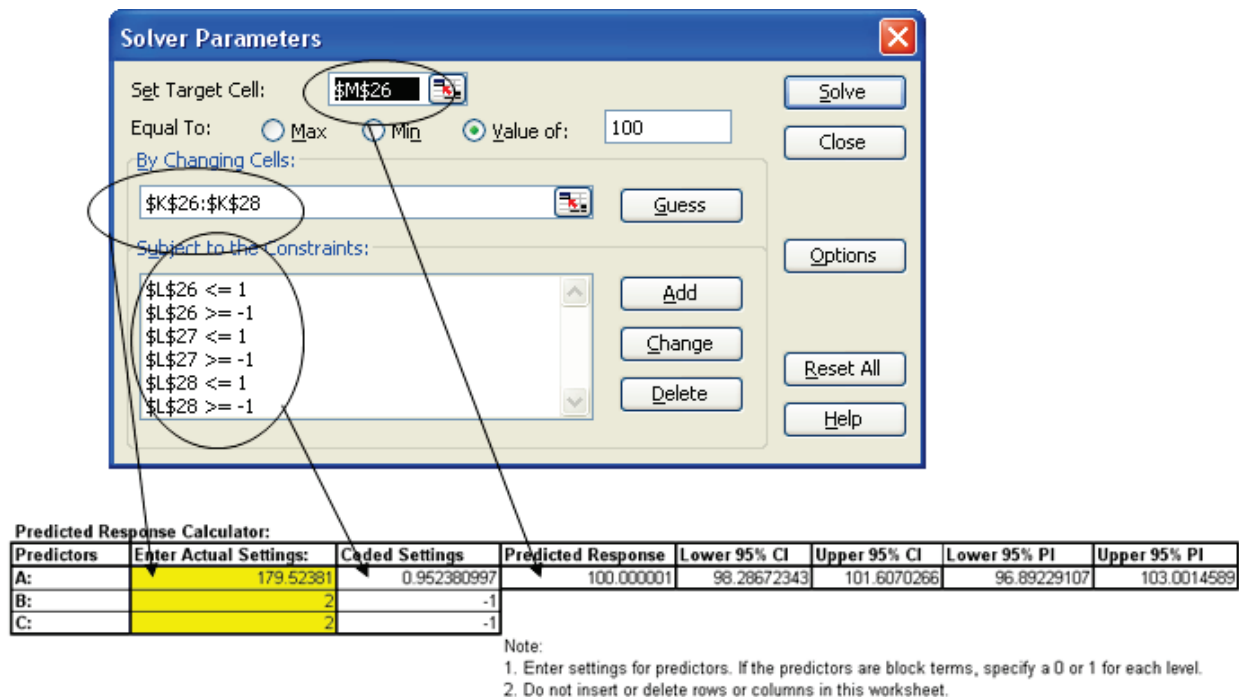
Predicted Response Calculator:

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.

20. Excel's **Solver** may also be used to get a more exact solution:



Solver Parameters

Set Target Cell: **\$M\$26**

Equal To: ☐ Max ☐ Min ☒ Value of: 100

By Changing Cells: **\$K\$26:\$K\$28**

Subject to the Constraints:

- \$L\$26 <= 1
- \$L\$26 >= -1
- \$L\$27 <= 1
- \$L\$27 >= -1
- \$L\$28 <= 1
- \$L\$28 >= -1

Predicted Response Calculator:

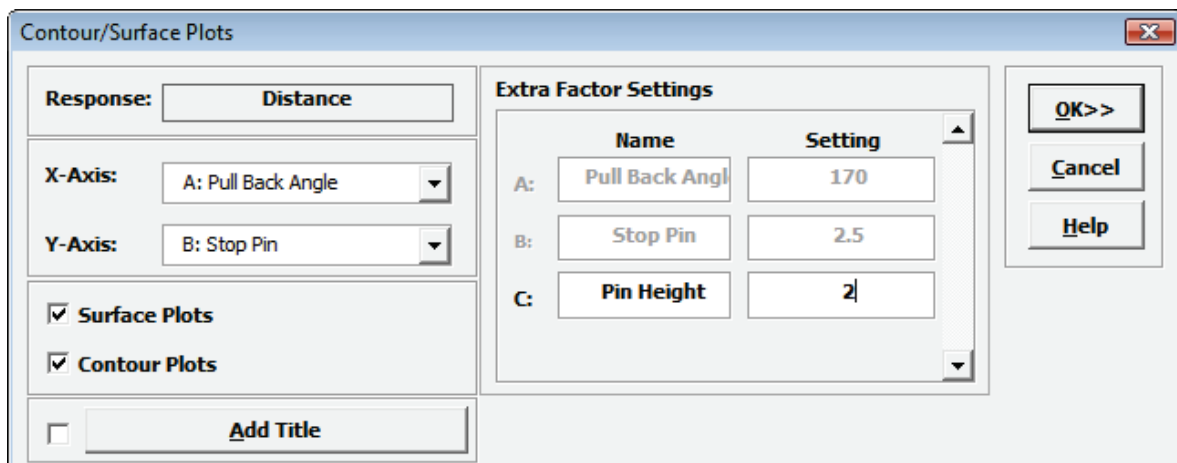
Predictors	Enter Actual Settings:	Coded Settings	Predicted Response	Lower 95% CI	Upper 95% CI	Lower 95% PI	Upper 95% PI
A:	179.52381	0.952380997	100.000001	98.28672343	101.6070266	96.89229107	103.0014589
B:	2	-1					
C:	2	-1					

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.

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) :



Contour/Surface Plots

Response: **Distance**

X-Axis: **A: Pull Back Angle**

Y-Axis: **B: Stop Pin**

☒ Surface Plots

☒ Contour Plots

☐ Add Title

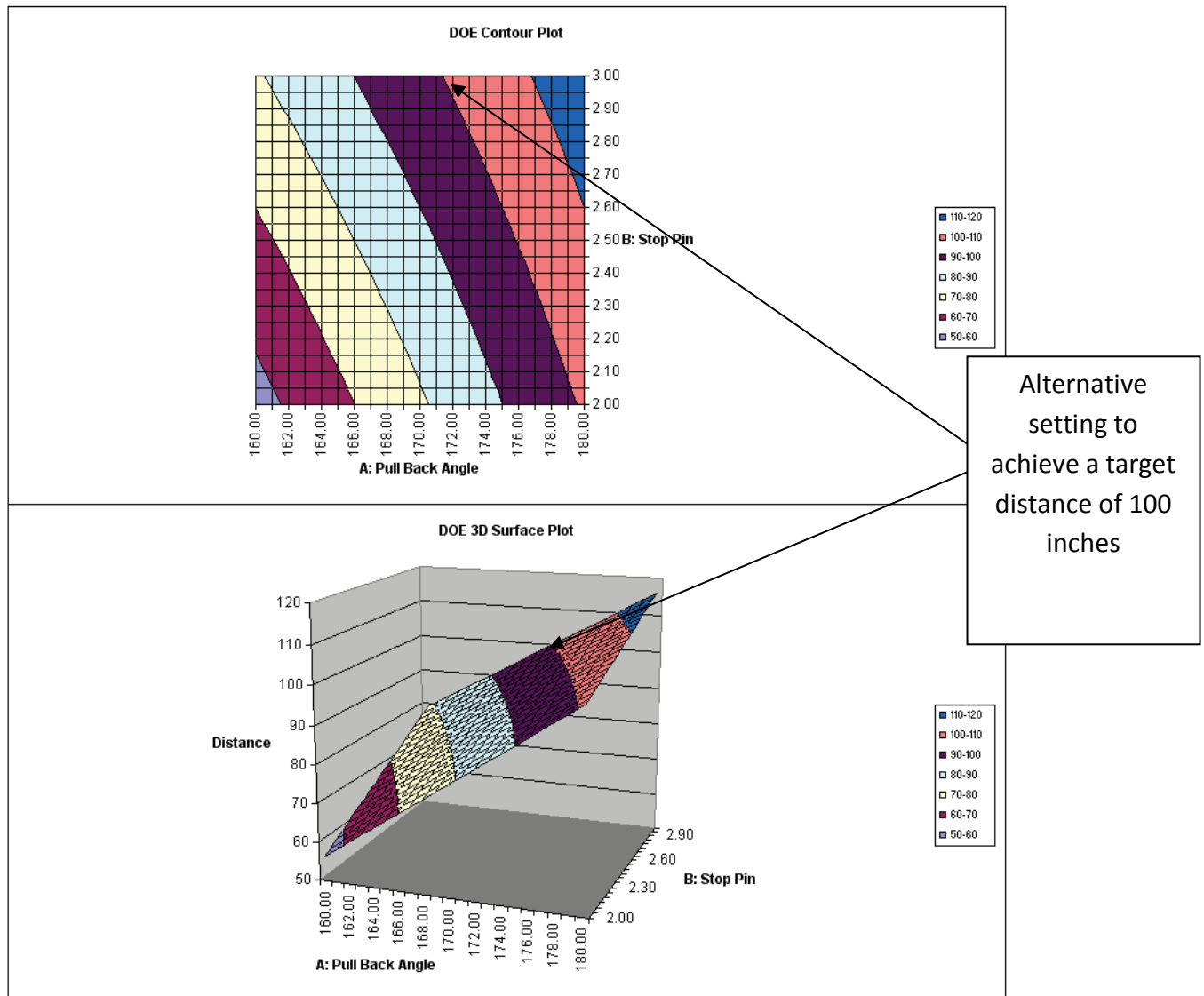
Extra Factor Settings

Name	Setting
A: Pull Back Angl	170
B: Stop Pin	2.5
C: Pin Height	2

OK>> Cancel Help

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 = 171.4, 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.

DOE Multiple Regression Model: Distance = (98.3125) + (22.8125) * A: Pull Back Angle + (10.0625) * B: Stop Pin + (10.1875) * C: Pin Height + (-1.9375) * AB + (2.4375) * AC + (0.6875) * BC



Tip: Use the contour/surface 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:

Response Surface Design of Experiments

Number of Factors: 2

Select Design: 10-Run, Central Composite Design (2 Ctr Pts)

Number of Replicates: 2

☒ Block on Replicates

Alpha Axial Value

☐ Rotatable (Alpha = 1.414)

☒ Face Centered (Alpha = 1.0)

Factor Levels Define: Cube points (Circumscribed)

☒ Randomize Runs

Factor Names and Level Settings:

	Name	Low	High
A:	Bake Time	20	40
B:	Oven Temp	350	400

Number of Responses: 1

Response Name

Y1: Taste Score

OK>> Cancel Help Reset

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**.

RSM Analyze

Available Responses:

Selected Responses: Taste Score

Responses (Y) >>

Available Model Terms:

Selected Model Terms: A: Bake Time, B: Oven Temp, AB, AA, BB, Blocks

Model Terms >>

<< Remove

<< Remove All

Alpha for Pareto Chart: 0.05

Include: Default (All) Terms

☒ Show Residual Plots

Regular

☒ Include Blocks

OK >> Cancel Help

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 + (-0.889285714) * AA + (-1.139285714) * BB + (-0.11) * Blocks_2

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
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 99.09%. 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:

RSM Analyze

Available Responses:

Selected Responses: Taste Score

Responses (Y) >>

OK >>
Cancel
Help

Available Model Terms:

Selected Model Terms: A: Bake Time
B: Oven Temp
AB
AA
BB

Model Terms >>

<< Remove

<< Remove All

Alpha for Pareto Chart: 0.05

Include: Default (All) Terms

☒ **Show Residual Plots**

Regular

☐ **Include Blocks**

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

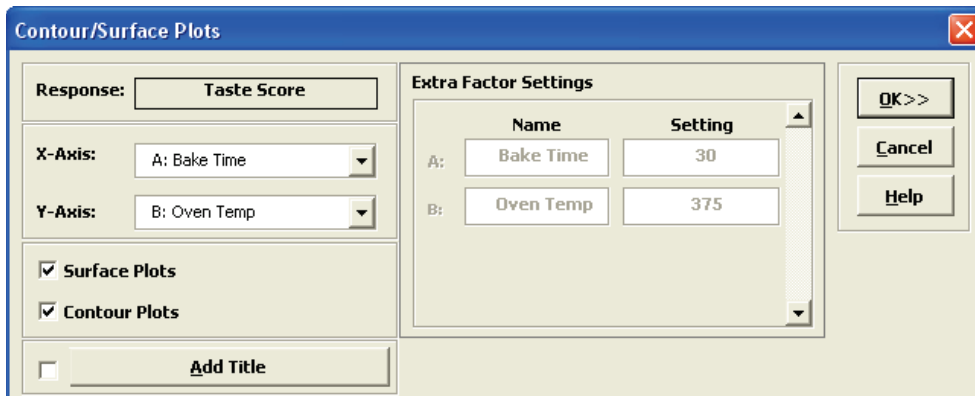
Analysis of Variance for Model:

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		

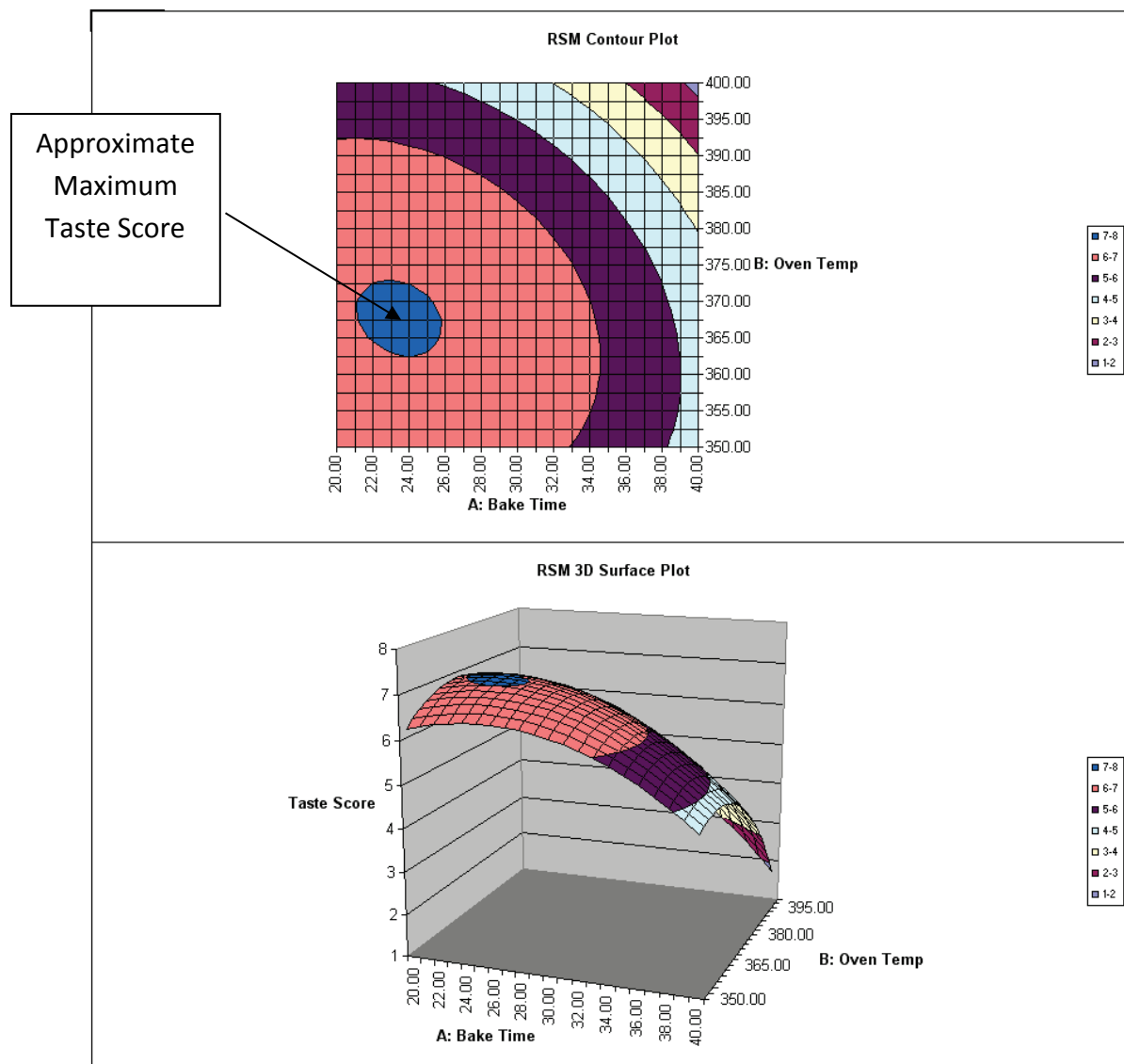
Durbin-Watson Test for Autocorrelation in Residuals:

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.

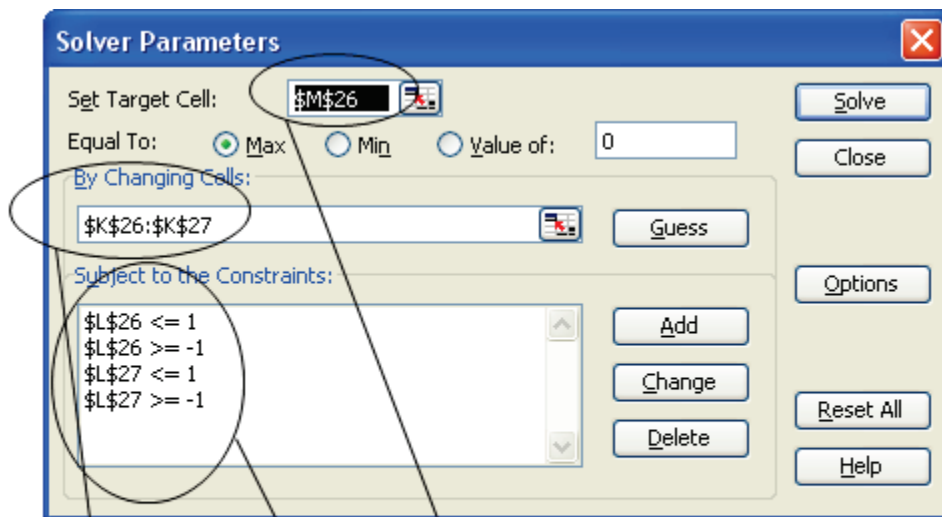


18. Go to **Analyze – 2 Factor RSM** tab. Scroll across to the **Predicted Response Calculator**. Enter the predicted values shown. These initial settings were determined from the contour plot as an estimate to yield a maximum taste score.

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					

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.

SigmaXL: Control Phase Tools: Statistical Process Control (SPC) Charts

Part A - Individuals Charts

Tip: See [Part E – Control Chart Selection Tool](#). The Control Chart Selection Tool 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 > Basic Control Chart Templates > Individuals Chart**.

This template is also located at **SigmaXL > Control Charts > Basic Control Chart Templates > Individuals Chart**.

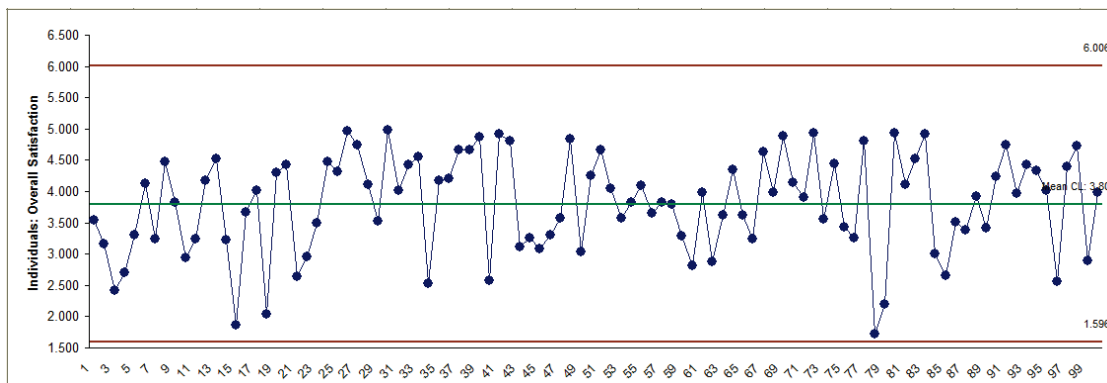
See **Measure Phase Part B – Templates and Calculators** for Individuals Chart template example:

[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.

Individuals Chart

Customer Record No
Order Date
Customer Type
Avg No. of orders per mo
Avg days Order to deliver
Loyalty - Likely to Recom
Overall Satisfaction
Responsive to Calls
Ease of Communications
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

Numeric Data Variable (Y) >> Overall Satisfaction

Optional X-Axis Labels >>

<< Remove

☐ Calculate Limits
☒ Enter Limits

☐ Tests for Special Causes

☐ Sigma Zone Lines

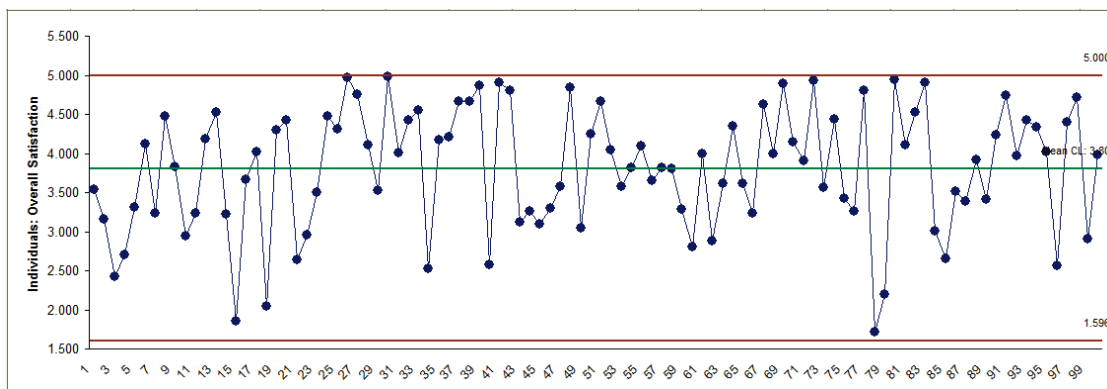
Advanced Options

☐ Add Title

5 UCL
3.8013 CL
1.596271 LCL

OK >>
Cancel
Help

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 Deviation

Optional X-Axis Labels >>

<< Remove

☒ Calculate Limits

☐ Historical Limits

☒ Tests for Special Causes

☒ Sigma Zone Lines

Advanced Options

☐ Add Title

UCL

CL

LCL

OK >>

Cancel

Help

Process Capability Report

LSL

-10

Target

0

USL

10

Calculate Control Limits:

☐ Specify Subgroup Number for Calculation of Control Limits

☒ Specify Historical Group Column (Split Limits)

Delivery Time Deviation

Defects

Floor

Historical Groups >>

Estimate:

☒ Average MR

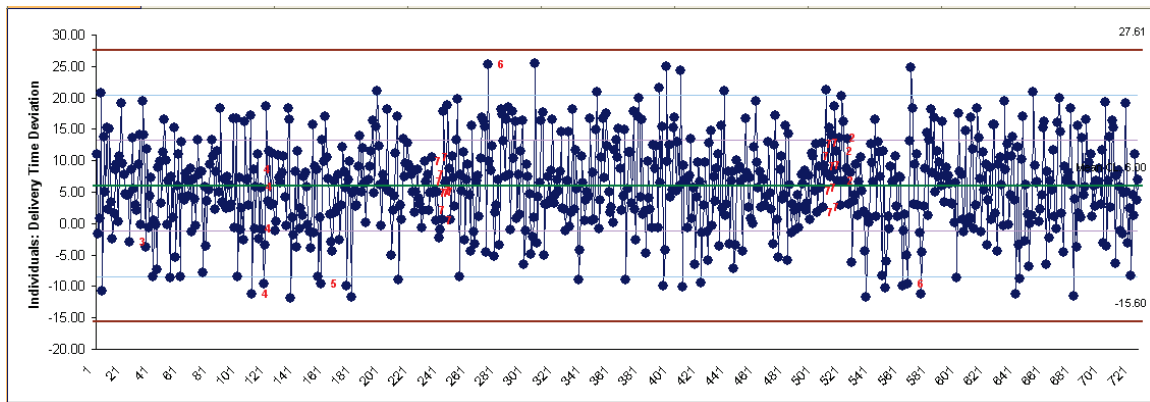
☐ Median MR

Individual CL

☒ Mean

☐ Median

11. Click **OK**. The resulting chart is shown:

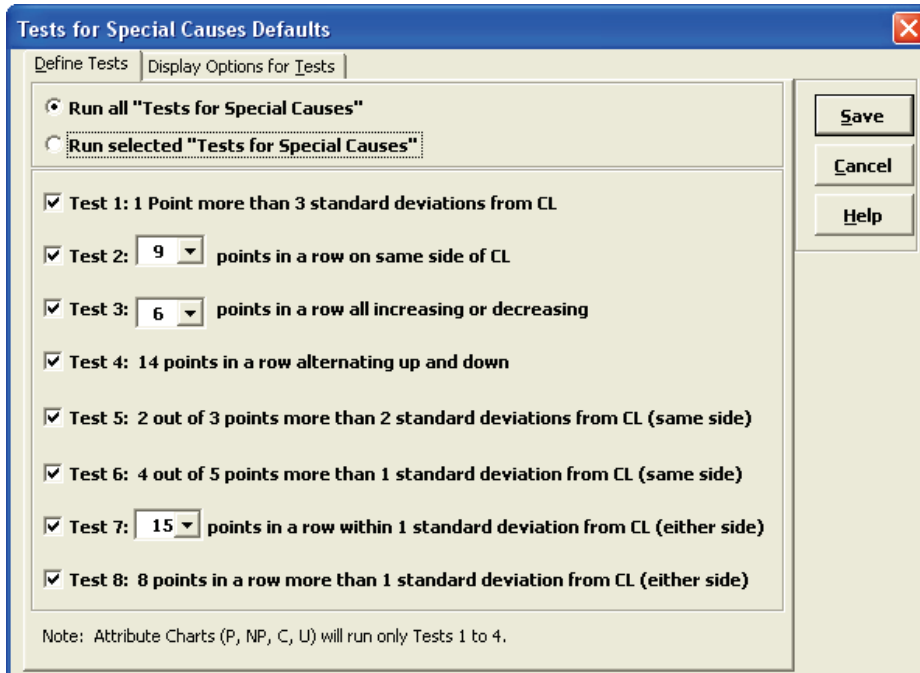


Some tests for special causes are indicated on the chart. If more than one test fails, the number corresponds to the first failed test.

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.

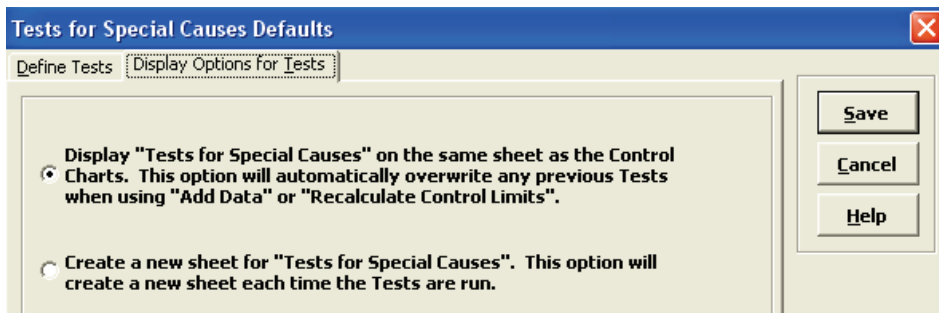
Tests for Special Causes - Individuals: Delivery Time Deviation								
Number of Data Points Failing Tests = 30								
Observation No.	Test 1: 1 point more than 3 Stdev from CL	Test 2: 9 points in a row on same side of CL	Test 3: 6 points in a row all increasing or all decreasing	Test 4: 14 points 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.

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

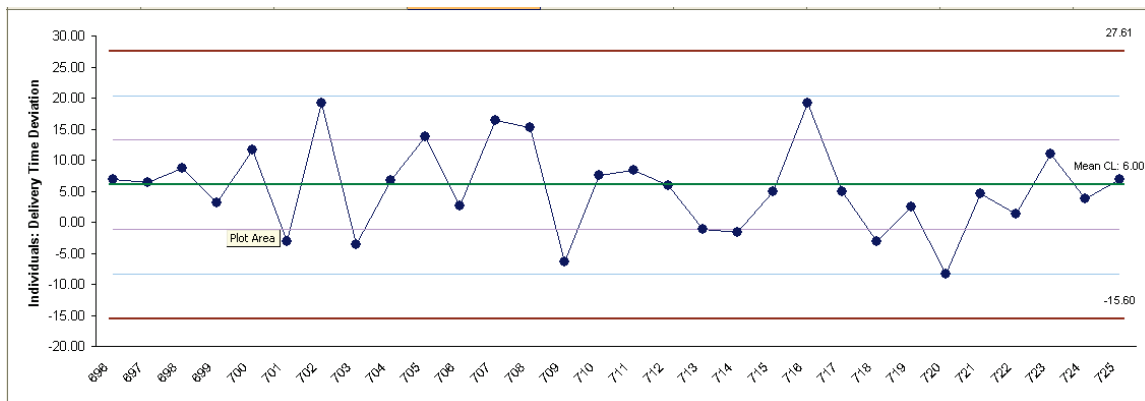
Potential Capability Indices using Within StDev		
Cp		0.46
Cpu		0.18
Cpl		0.74
Cpk		0.18

Expected Overall Performance		
ppm > USL		288409
ppm < LSL		12720.5
ppm Total		301129.8
% > USL		28.84%
% < LSL		1.27%
% Total		30.11%

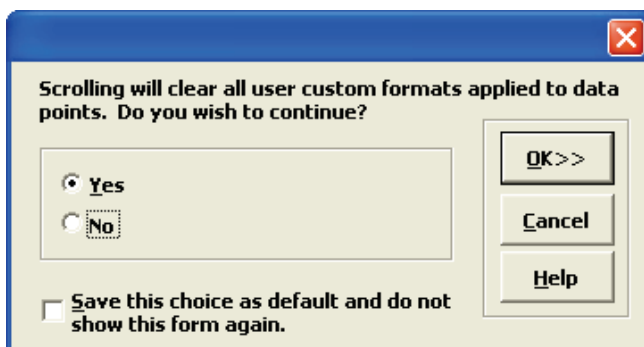
Actual (Empirical) Performance		
% > USL		26.90%
% < LSL		1.38%
% Total		28.28%

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:

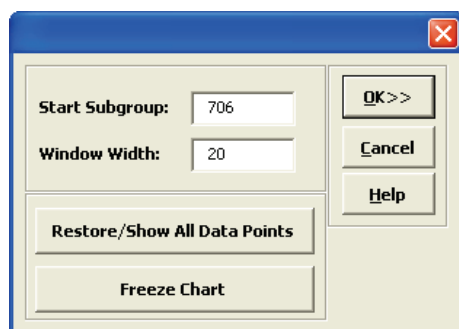


20. To reset the chart, click **SigmaXL Chart Tools > Show All Data Points**.
21. To enable scrolling, click **SigmaXL Chart Tools > Enable Scrolling**. A warning message is given:



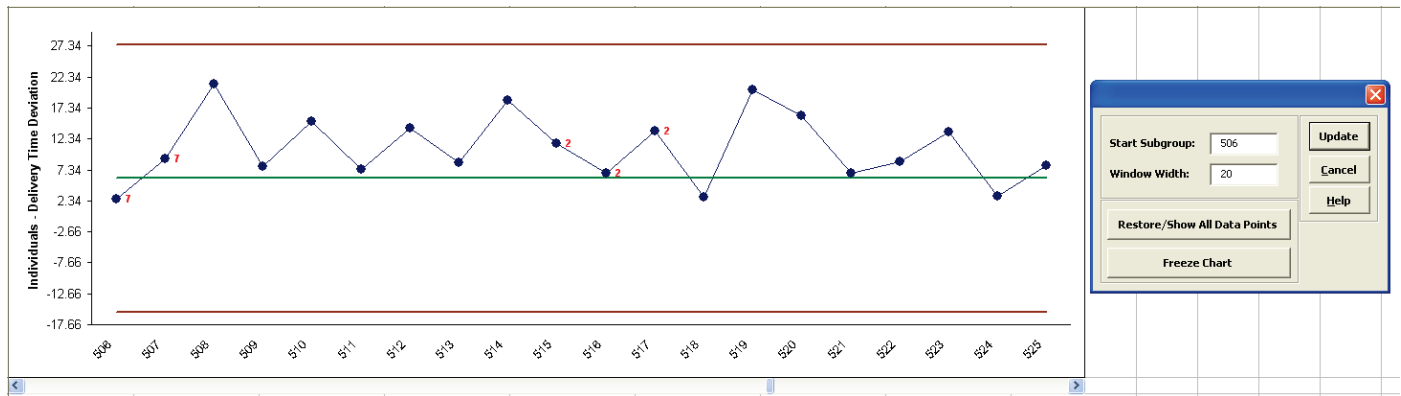
Scrolling will clear user custom formats, but does not affect Tests for Special Causes.

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**.



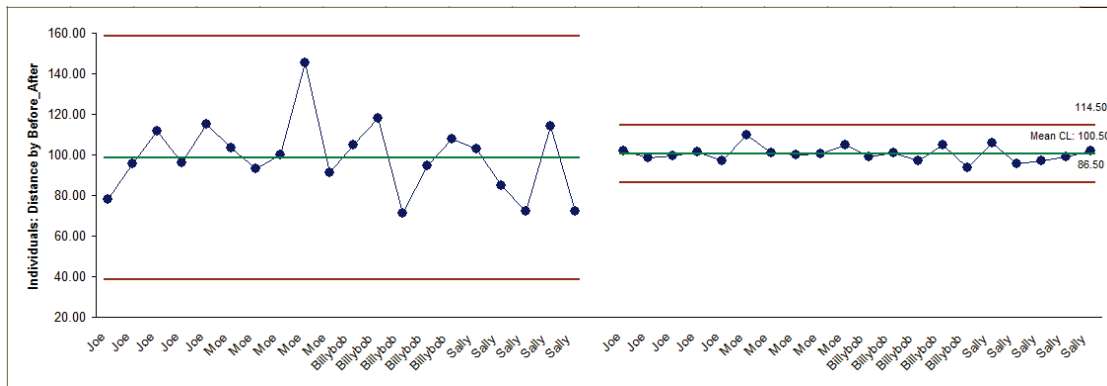
Individuals Charts: Advanced Limit Options – Historical Groups

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 >>**.

The screenshot shows the 'Individuals Chart' dialog box in SigmaXL. The 'Numeric Data Variable (Y) >>' field is set to 'Distance'. The 'Optional X-Axis Labels >>' field is set to 'Operator'. The 'Calculate Limits' section has 'Calculate Limits' selected. The 'Advanced Options' button is highlighted. The 'Process Capability Report' section shows LSL, Target, and USL fields. The 'Calculate Control Limits' section has 'Specify Historical Group Column (Split Limits)' selected. The 'Historical Groups >>' field is set to 'Before_After'. The 'Estimate' section has 'Average MR' selected. The 'Individual CL' section has 'Mean' selected.

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.

Individuals and Moving Range Chart

Customer Record No
Order Date
Customer Type
Avg No. of orders per m
Avg days Order to deliv
Loyalty - Likely to Recor
Overall Satisfaction
Responsive to Calls
Ease of Communication
Staff Knowledge
Size of Customer
Major-Complaint
Product Type
Sat-Discrete

Numeric Data Variable (Y) >> Overall Satisfaction

Optional X-Axis Labels >>

<< Remove

☒ Calculate Limits
☐ Historical Limits

☐ Tests for Special Causes
☐ Sigma Zone Lines

Advanced Options

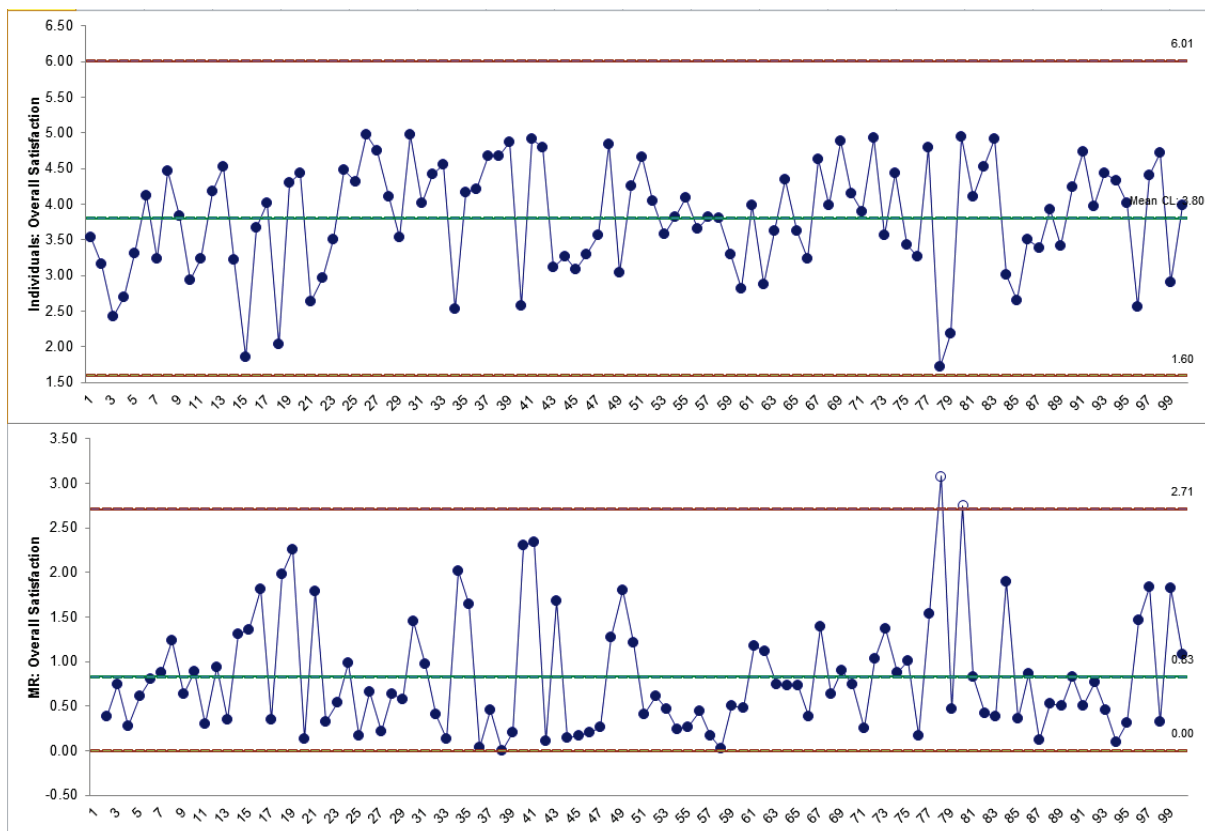
☐ Add Title

Individuals Moving Range

UCL CL LCL

OK >> Cancel Help

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 nonnormal, 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**.

Individuals Chart

Cycle Time (Minutes)

Numeric Data Variable (Y) >> Cycle Time (Minutes)

Optional X-Axis Labels >>

<< Remove

☒ Calculate Limits
☐ Historical Limits

☒ Tests for Special Causes

☐ Sigma Zone Lines

Advanced Options

☐ Add Title

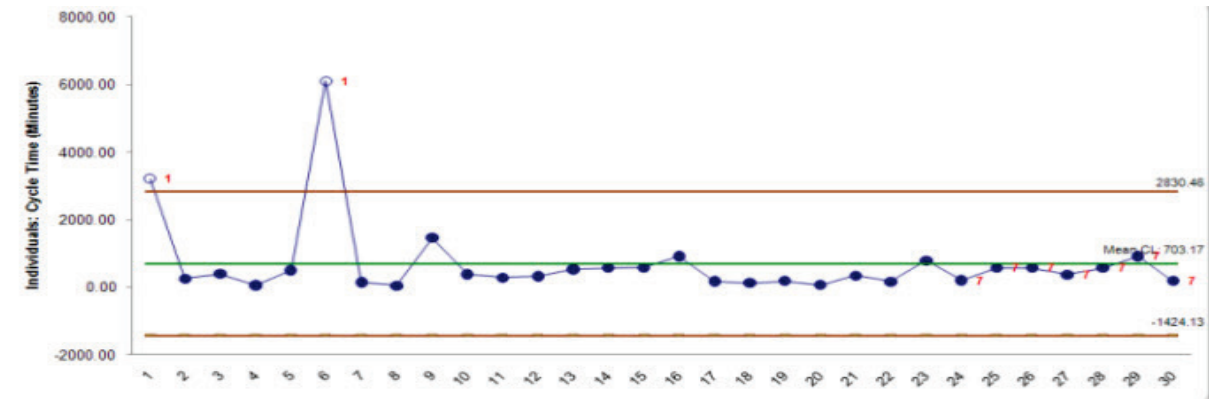
UCL

CL

LCL

OK >> Cancel Help

4. The resulting Individuals Chart is shown:

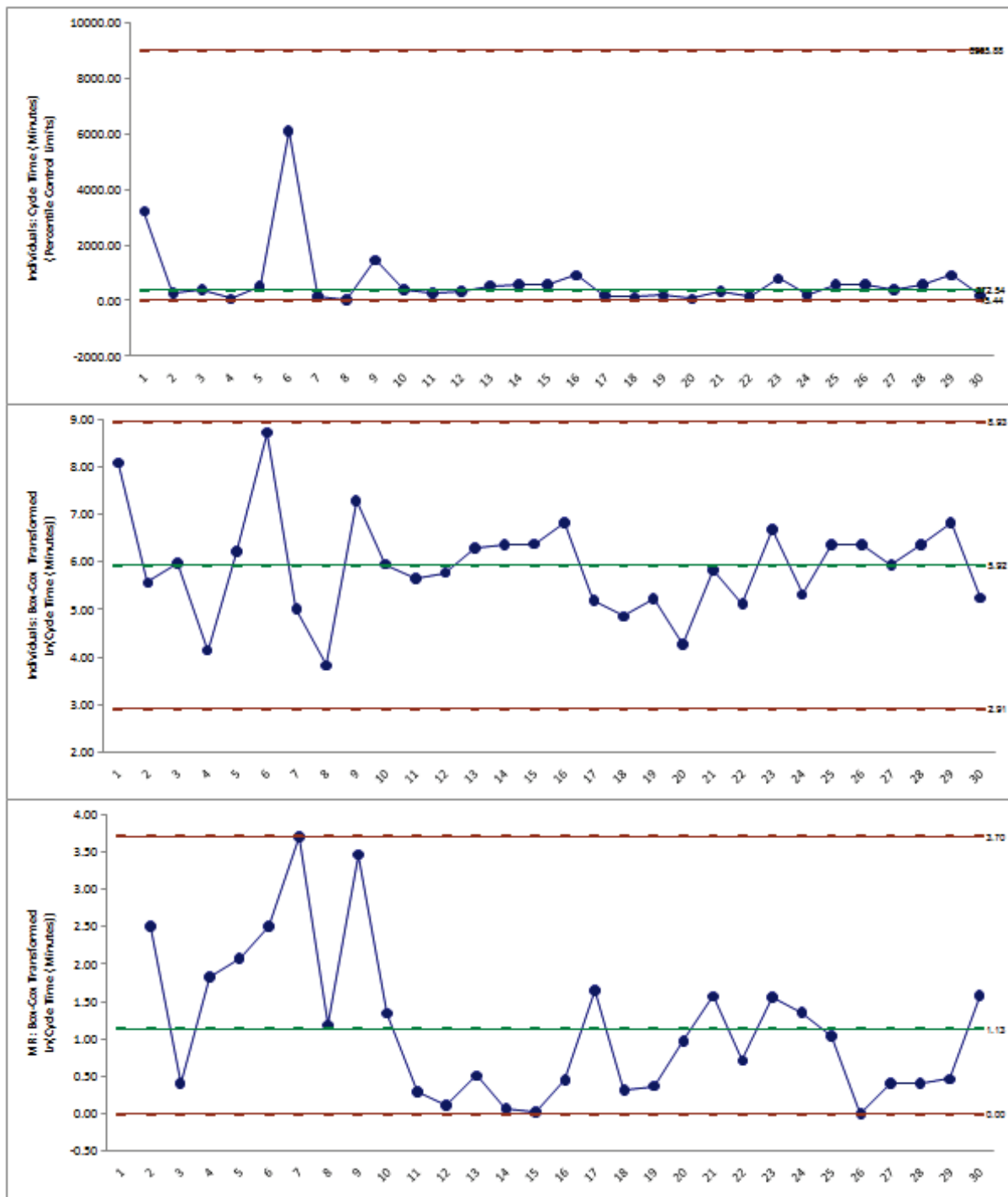


This chart clearly shows that the process is “out-of-control”. But is it really? Nonnormality can cause serious errors in the calculation of Individuals Chart control limits, triggering false alarms (Type I errors) or misses (Type II errors).

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:

The screenshot shows the 'Individual's Nonnormal' dialog box. The 'Cycle Time (Minutes)' variable is selected in the list. The 'Numeric Data Variable (Y) >>' button is highlighted. The 'Optional X-Axis Labels >>' button is also visible. The 'Control Chart Options' section has three checked options: 'Individuals - Original Data', 'Individuals/Moving Range - Normalized Data', and 'Tests for Special Causes'. The 'Add Title' checkbox is unchecked. The 'Transformation/Distribution Options' section has 'Box-Cox Transformation' selected. Under 'Box-Cox Transformation Options', 'Rounded Lambda' is selected. The 'Optimal Lambda' and 'Lambda & Threshold (Shift)' options are also visible but not selected. The 'Optional Threshold Value' and 'Optional Lambda Value' input fields are present but empty.

7. Click **OK**. The resulting control charts are shown below:



Tests for Special Causes - Indv: Cycle Time (Minutes)
Number of Data Points Failing Tests = 0

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.

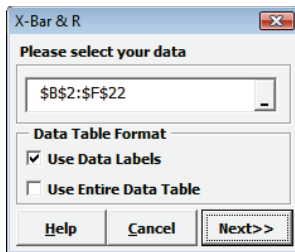
The **Individuals/Moving Range – Normalized Data** chart displays the transformed z-values with control limits calculated using the standard Shewhart formulas for Individuals and Moving Range charts. The benefit of using this chart is that tests for special causes can be applied and the control limits 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 Tool](#). The Control Chart Selection Tool 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

Optional X-Axis Labels >> Operator

<< Remove

☒ Calculate Limits
☐ Historical Limits

☒ Tests for Special Causes

☐ Sigma Zone Lines

Advanced Options

☐ Add Title


X-Bar R

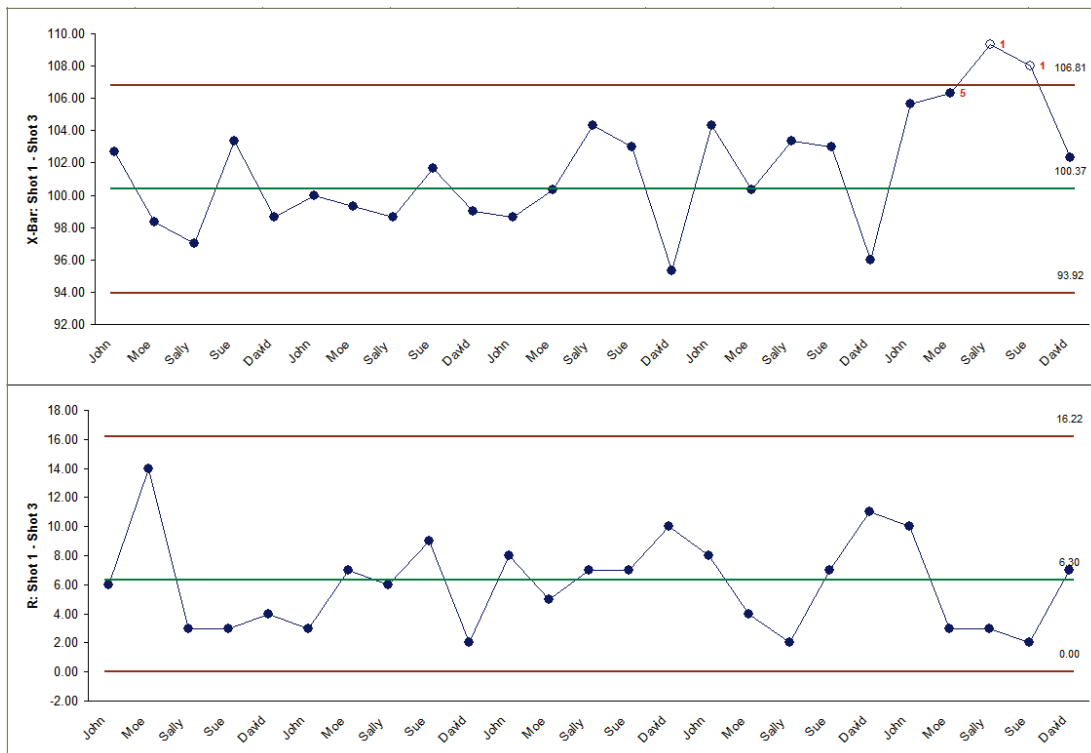
	UCL	
	CL	
	LCL	

OK >>
Cancel
Help

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** as shown. In Excel 2007+ click:  **Add Data to this Control Chart**



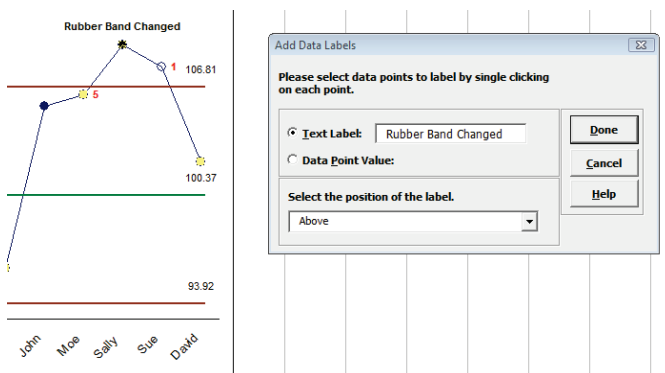
- Note that the Add Data button does NOT recalculate the control limits. Once control limits are established, they should only be recalculated when a deliberate process change or improvement is introduced. (Control Limits can be recalculated using **SigmaXL Chart Tools > Recalculate Control Limits**, but a warning message **Are you sure that you want to recalculate control limits?** is given.)
- The **Tests for Special Causes** report gives us more detail on the recent instability:

Tests for Special Causes - X-Bar - 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.

- 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.)
- 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**.



- 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**.)
- Select *Shots 1-3*, click **Numeric Data Variables (Y) >>**.
- Click **Advanced Options**. Enter **LSL = 92**, **Target = 100**, **USL = 108**.

14. The resulting dialog box settings are 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

Optional X-Axis Labels >>

<< Remove

☒ Calculate Limits
☐ Historical Limits

☐ Tests for Special Causes

☐ Sigma Zone Lines

Advanced Options

☐ Add Title

X-Bar		R	
	UCL		
	CL		
	LCL		

Process Capability Report
 LSL 92 Target 100 USL 108

Calculate Control Limits:
☐ Specify Subgroup Number for Calculation of Control Limits
☒ Specify Historical Group Column (Split Limits)

Subgroup No
Operator
Shot 1
Shot 2

Historical Groups >>

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%

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.

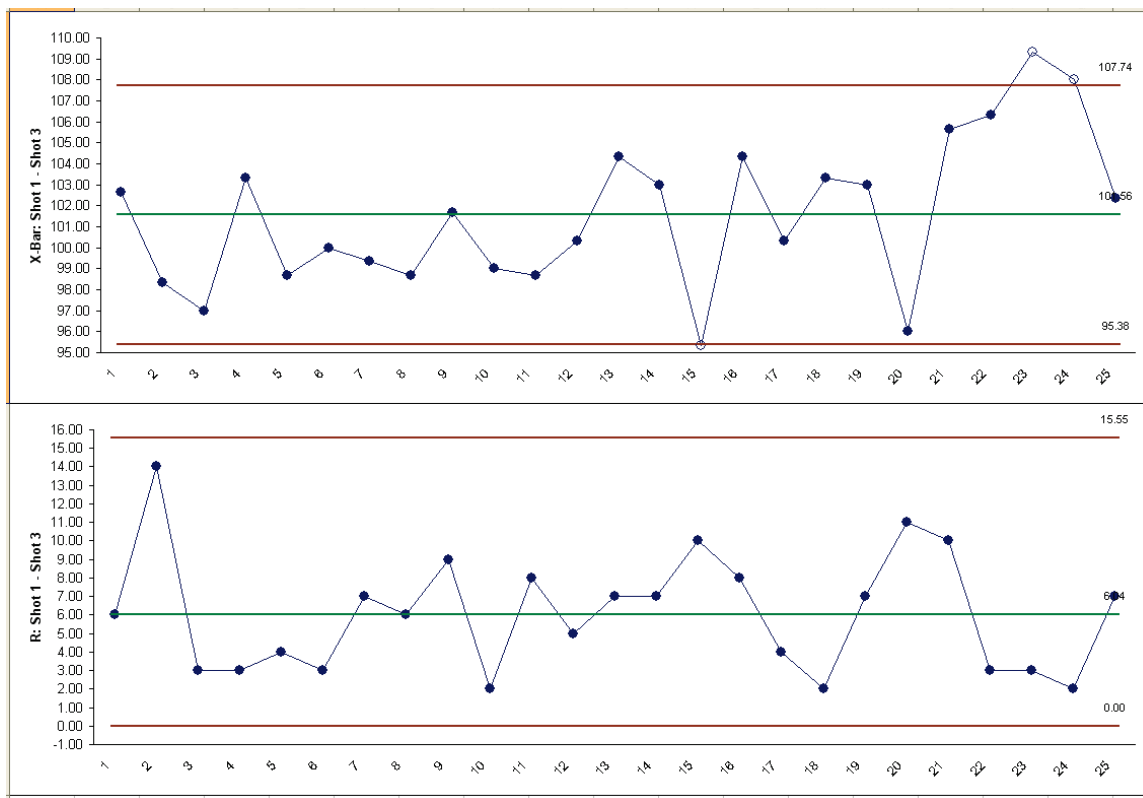
X-Bar & R Charts – Exclude Subgroups

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**.

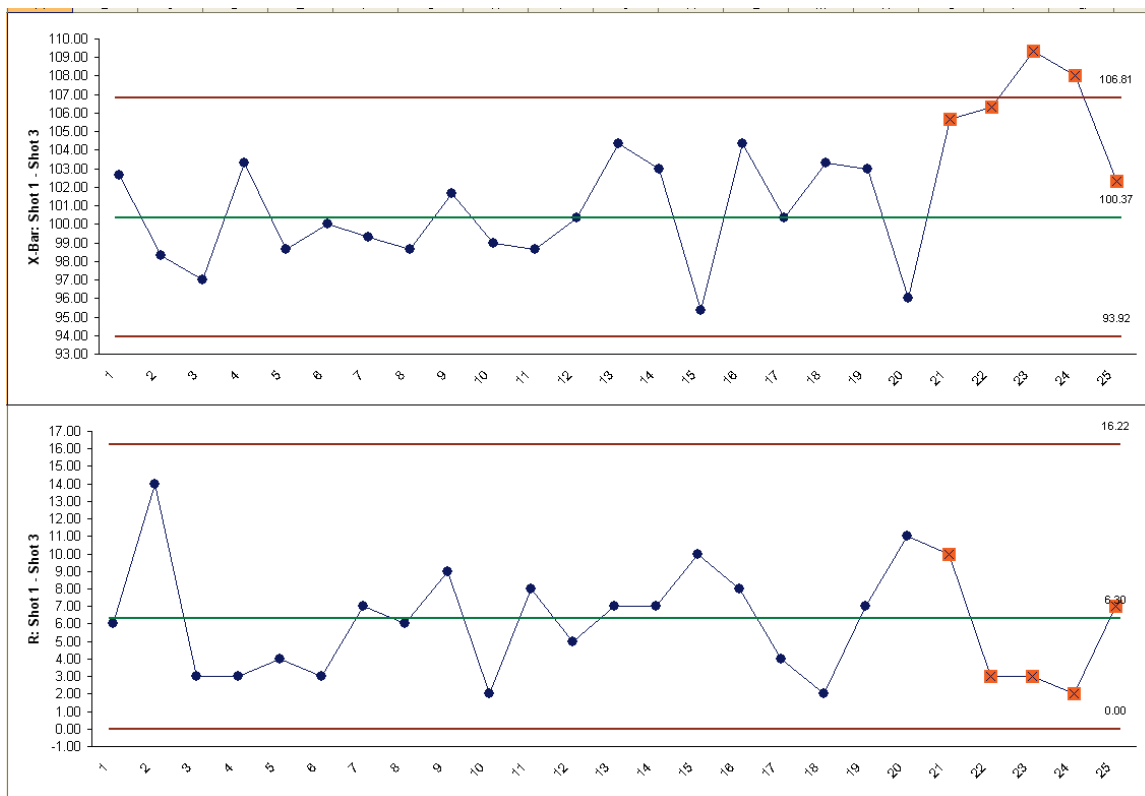
18. The resulting X-bar & R charts are displayed:



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:



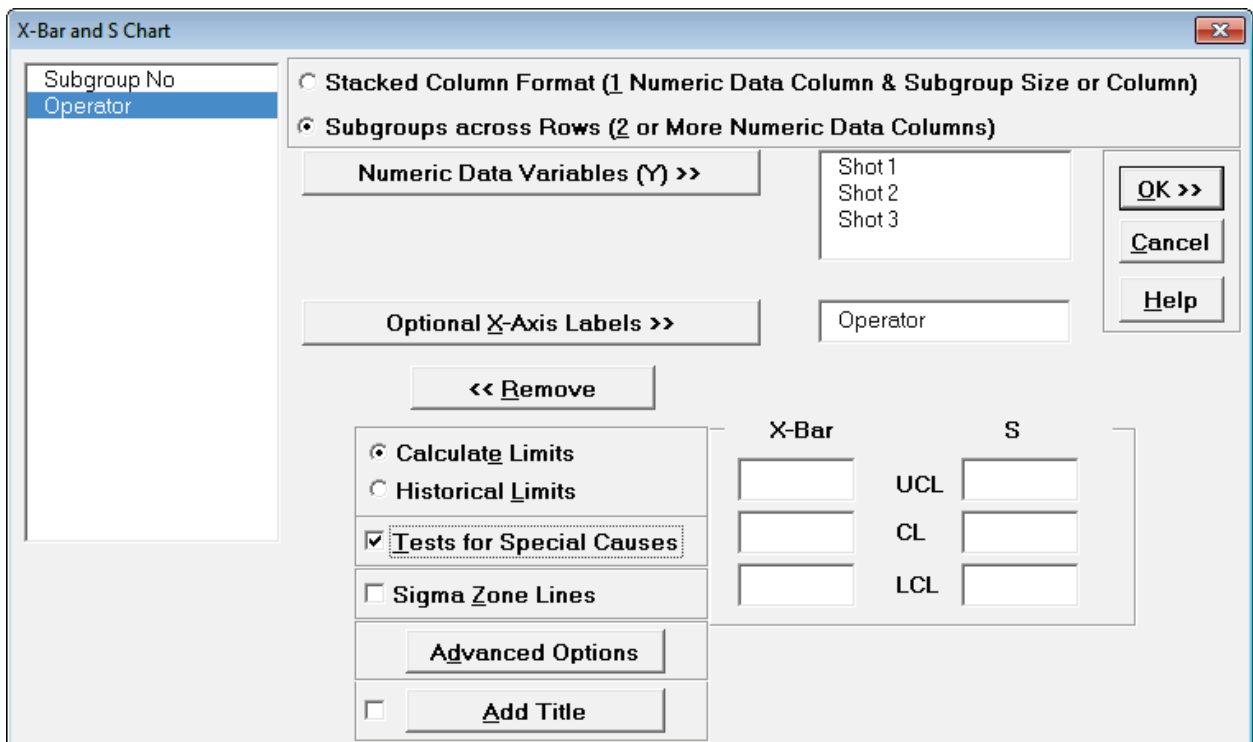
Tip: You can also choose to show gaps for excluded subgroups or delete excluded subgroups from the charts.

X-Bar & S Charts

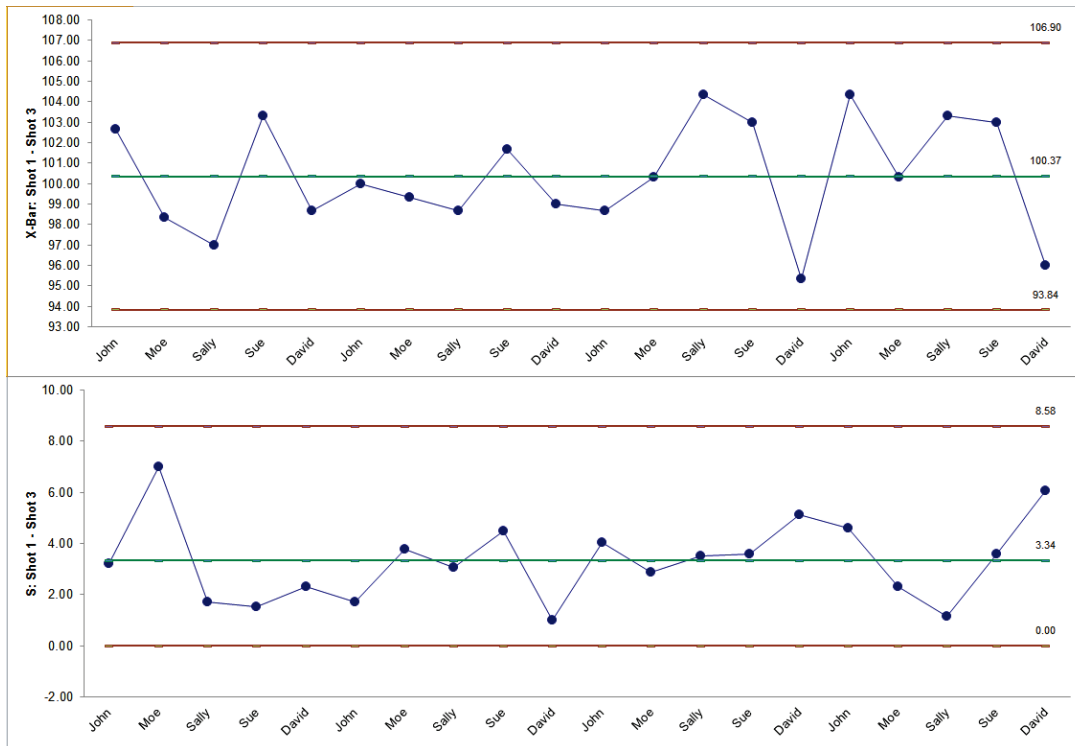
1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the ball 3 times. The target distance is 100 inches. The Upper Specification Limit (USL) is 108 inches. The Lower Specification Limit (LSL) is 92 inches.
2. Select **B2:F22**; here, we will only use the first 20 subgroups to determine the control limits.
3. Select **SigmaXL > Control Charts > X-Bar & S**.
4. Do not check **Use Entire Data Table**!




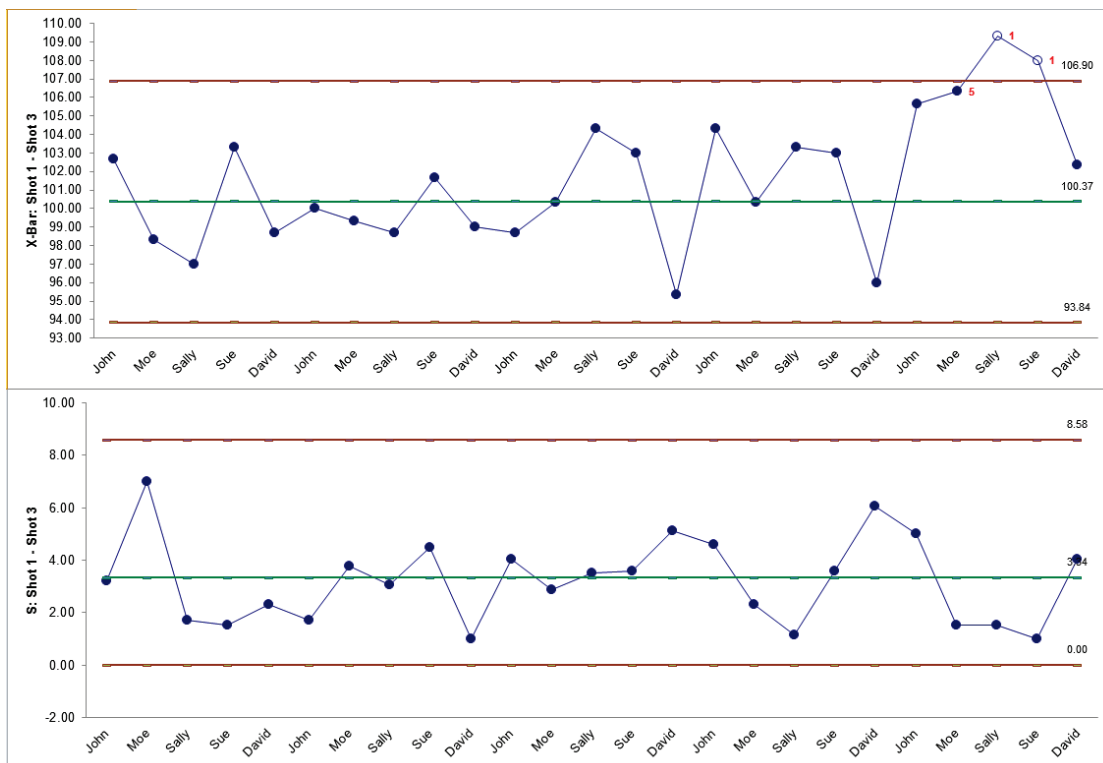
5. Click **Next**. Select **Subgroups across Rows**, select *Shot 1*, *Shot 2*, *Shot 3*, click **Numeric Data Variables (Y) >>**; select *Operator*, click **Optional X-Axis Labels >>**. Check **Tests for Special Causes** as shown:



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** as shown. In Excel 2007+ click:  **Add Data to this Control Chart**



Part C – Attribute Charts (P, NP, C, U)

Tip: See [Part E – Control Chart Selection Tool](#). The Control Chart Selection Tool makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

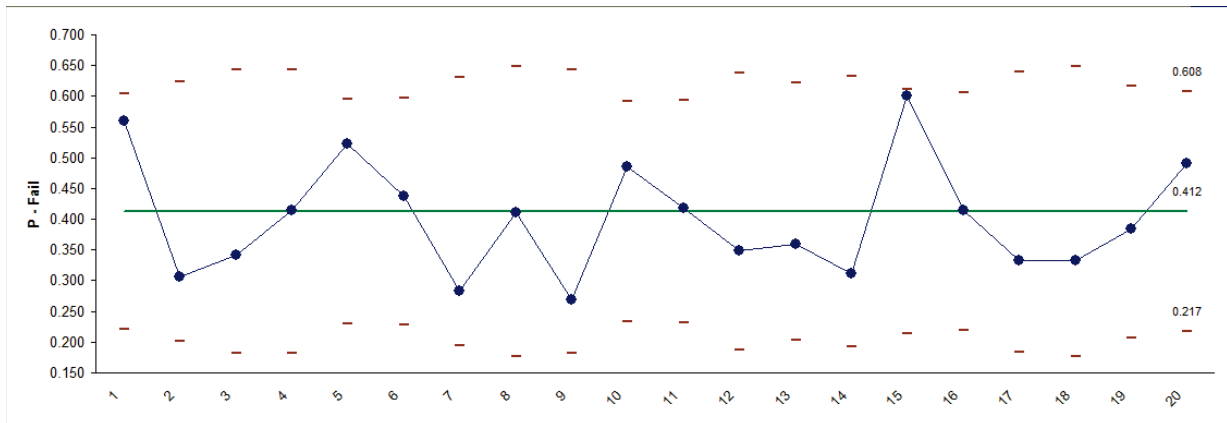
P-Charts

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. Ensure that **Tests for Special Causes** is unchecked.

The screenshot shows the 'P-Chart' dialog box. On the left, a list of variables includes 'Day', 'Fail', 'Pass', and 'N'. The 'Numeric Data Variable (Y) >>' field is set to 'Fail'. The 'Subgroup Column or Size >>' field is set to 'N'. The 'Optional X-Axis Labels >>' field is empty. Below these fields is a '<< Remove' button. Further down, there are radio buttons for 'Calculate Limits' (which is selected) and 'Historical Limits'. To the right of these are three input fields for 'UCL', 'CL', and 'LCL'. There are also checkboxes for 'Tests for Special Causes' (unchecked), 'Sigma Zone Lines' (unchecked), and 'Add Title' (unchecked). On the far right, there are buttons for 'OK >>', 'Cancel', and 'Help'.

Note: Tests for Special Causes cannot be used with varying subgroup sizes.

- 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

- 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.
- Click **SigmaXL > Control Charts > Attribute Charts > P**. Ensure that the entire data table is selected. If not, check **Use Entire Data Table**. Click **Next**.
- Select *Fail*, click **Numeric Data Variable (Y) >>**; select *N*, click **Subgroup Column or Size >>**. Ensure that **Tests for Special Causes** is unchecked.
- Click **Advanced Options**. Select **Specify Historical Group Column**. Select *Before_After*, click **Historical Groups >>**.

P-Chart

Day
Before_After
Fail
Pass
N

Numeric Data Variable (Y) >> Fail

Subgroup Column or Size >> N

Optional X-Axis Labels >>

<< Remove

☒ Calculate Limits
☐ Historical Limits

☐ Tests for Special Causes

☐ Sigma Zone Lines

Advanced Options

☐ Add Title

UCL
CL
LCL

Calculate Control Limits:

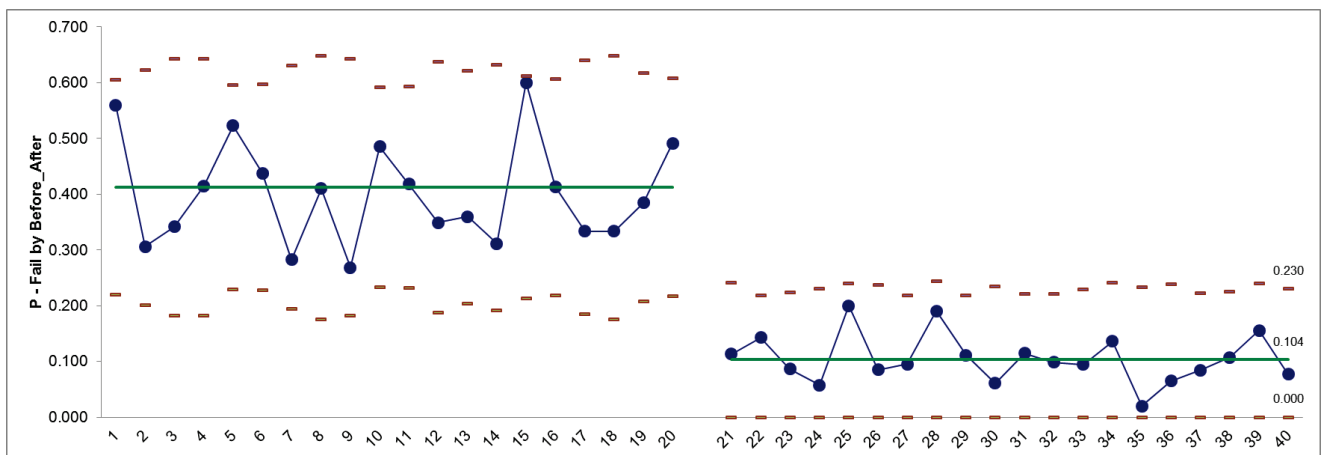
☐ Specify Subgroup Number for Calculation of Control Limits
☒ Specify Historical Group Column (Split Limits)

Day
Before_After
Fail
Pass

Historical Groups >> Before_After

OK >>
Cancel
Help

5. Click **OK**. The resulting P-Chart with split limits based on historical groups is shown, demonstrating a clear process improvement:

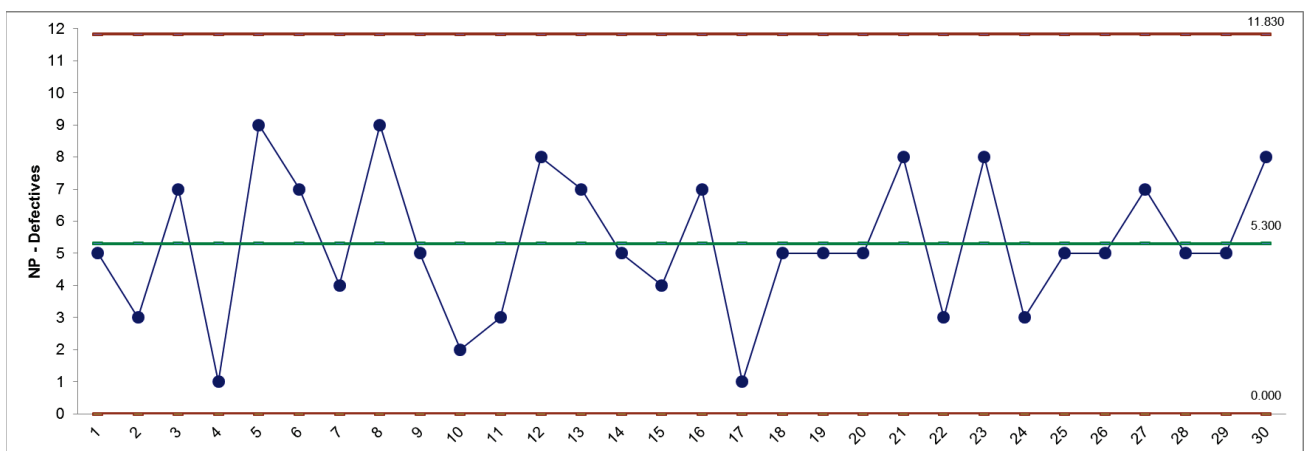


NP-Charts

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**.

Note: Tests for Special Causes can always be used with NP charts because the subgroup size is fixed, which results in constant control limits.

4. Click **OK**. The resulting NP-Chart is shown:



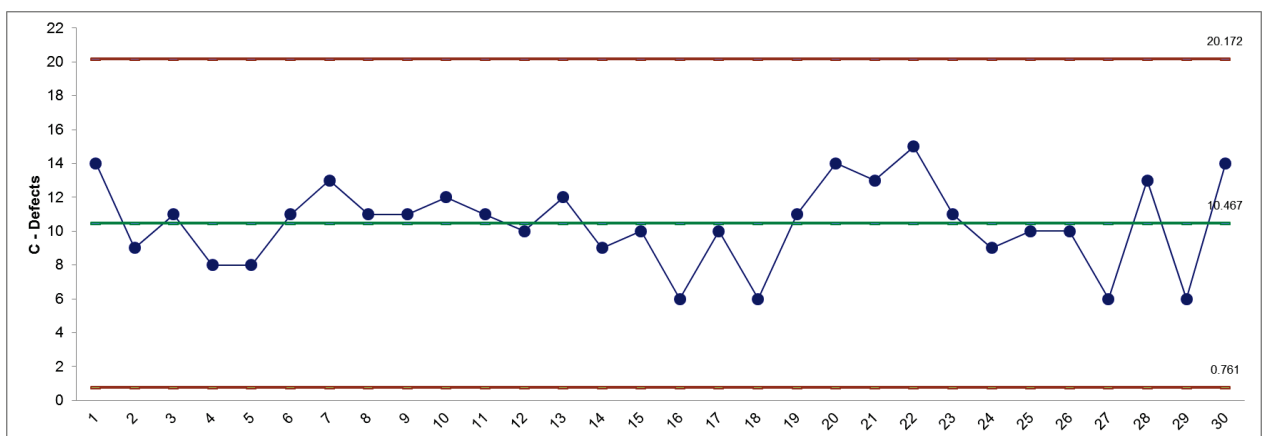
This is a stable process, with no tests for special causes flagged.

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**.

Note: Tests for Special Causes can always be used with C charts because the subgroup size is assumed to be fixed, which results in constant control limits.

4. Click **OK**. The resulting C-Chart is shown:



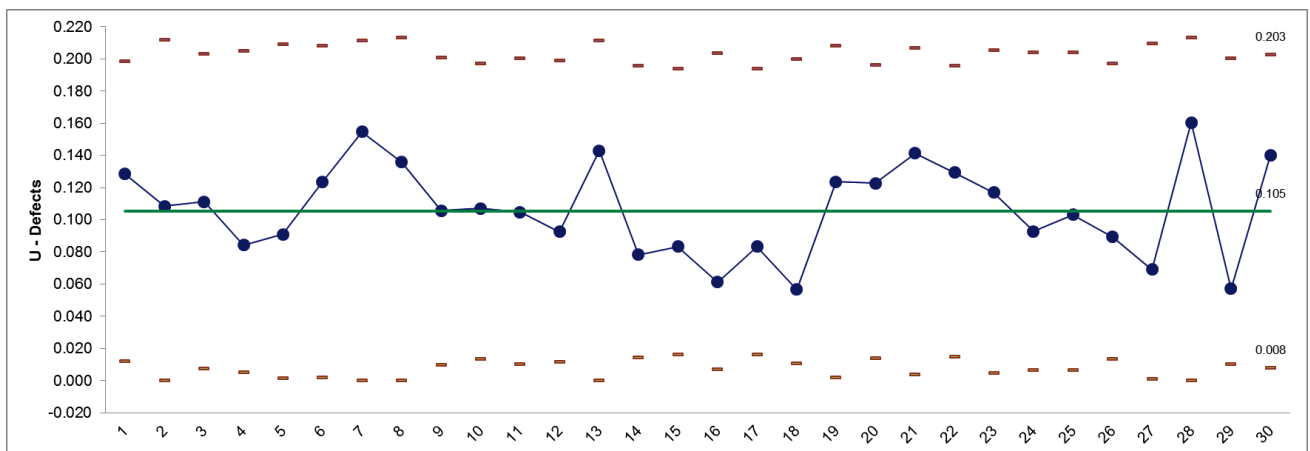
This is a stable process, with no tests for special causes flagged.

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 >>**. Ensure that **Tests for Special Causes** is unchecked (**Tests for Special Causes** cannot be used with varying subgroup sizes).

The screenshot shows the 'U-Chart' dialog box. On the left, a list of variables includes 'Defects' and 'N'. In the center, there are three buttons: 'Numeric Data Variable (Y) >>', 'Subgroup Column or Size >>', and 'Optional X-Axis Labels >>'. To the right of these buttons are three text boxes containing 'Defects', 'N', and an empty box. Below these is a '<< Remove' button. Further down, there are two radio buttons: 'Calculate Limits' (selected) and 'Historical Limits'. Below these are two checkboxes: 'Tests for Special Causes' (unchecked) and 'Sigma Zone Lines' (unchecked). To the right of these checkboxes are three text boxes for 'UCL', 'CL', and 'LCL'. At the bottom, there is an 'Advanced Options' button and an 'Add Title' checkbox.

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.

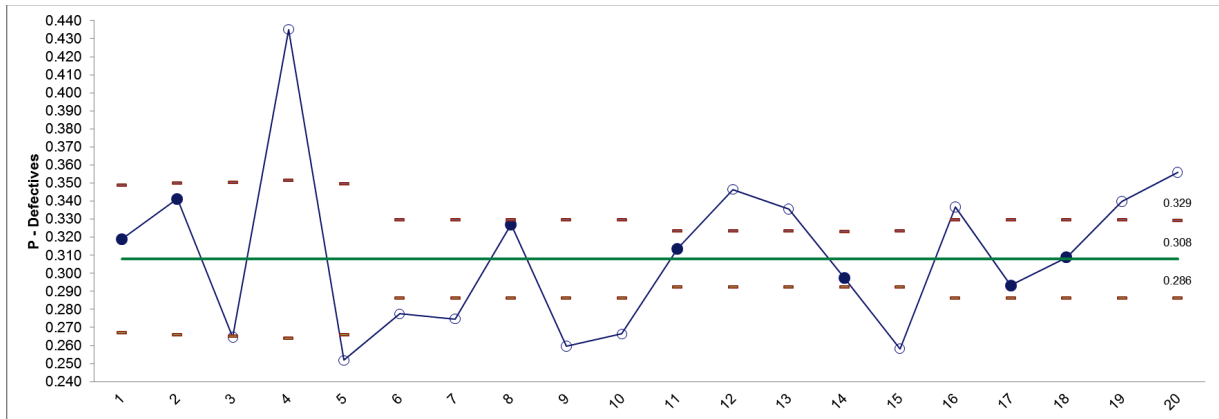
Tip: See [Part E – Control Chart Selection Tool](#). The Control Chart Selection Tool makes it easy for you to select the correct statistical process control chart depending on data type and subgroup/sample type and size.

References:

- [1] Crossley, Mark L.(2007), *The Desk Reference of Statistical Quality Methods*, Second Edition, Milwaukee, WI, ASQ Quality Press, pp. 352 – 355.
- [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

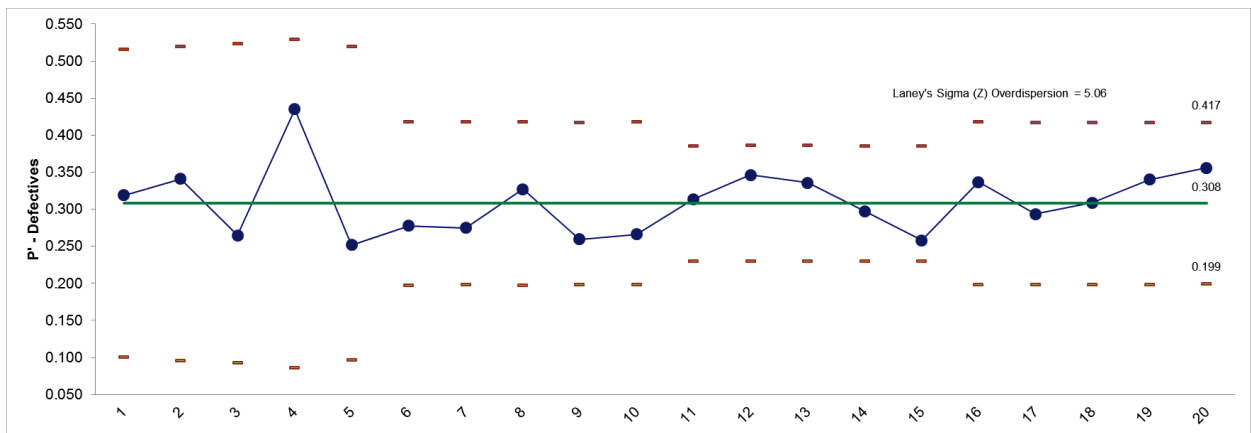
1. Open **Laney – Quality Digest – Defectives.xlsx**. This data is used with permission from David Laney and was used in the example given in reference [2].
2. We will begin with the creation of a regular P-Chart for this data. Select **SigmaXL > Control Charts > Attribute Charts > P**. Ensure that the 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)**. Ensure that **Tests for Special Causes** is unchecked (**Tests for Special Causes** cannot be used with varying subgroup sizes). 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.)

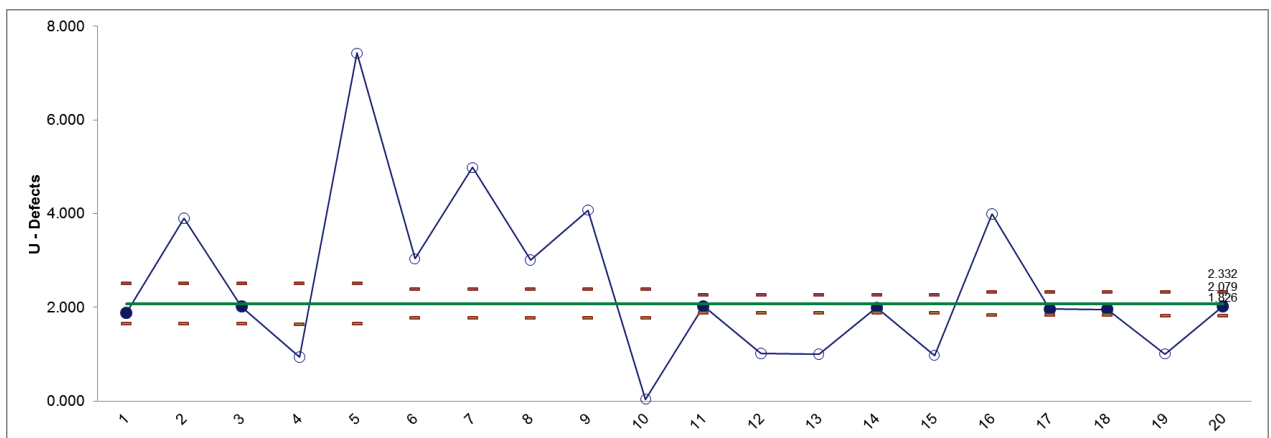
6. Click **OK**. The resulting P'-Chart is shown:



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:



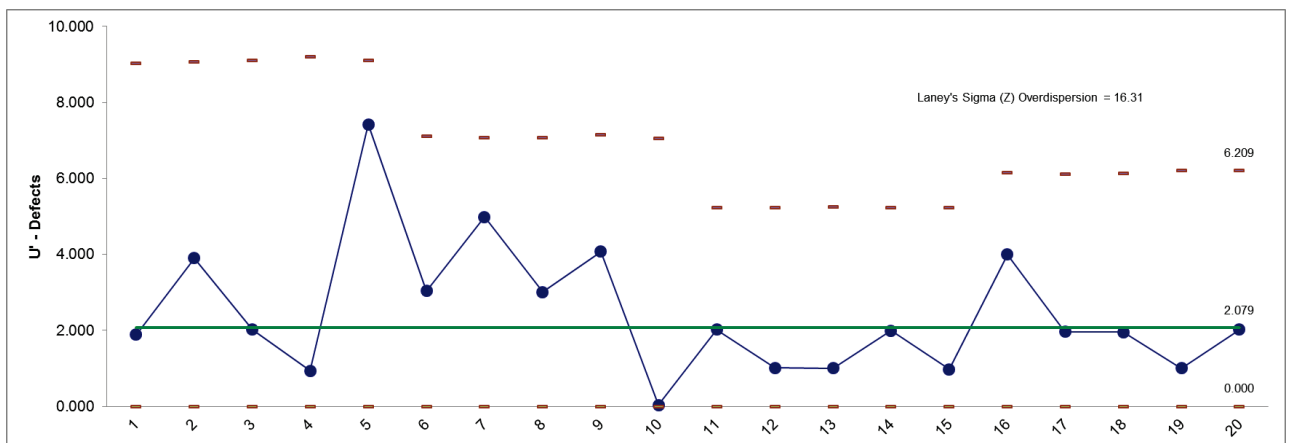
This chart suggests that the process is out of control, but the problem here is actually due to the Poisson assumption not being valid (also known as *overdispersion*). This problem becomes evident when the sample sizes are large.

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.) Ensure that **Tests for Special Causes** is unchecked (**Tests for Special Causes** cannot be used with varying subgroup sizes).

The screenshot shows the 'U' (Laney) dialog box. On the left, a list of variables includes 'Defects' and 'N', with 'N' selected. In the center, there are three buttons: 'Numeric Data Variable (Y) >>', 'Subgroup Column or Size >>', and 'Optional X-Axis Labels >>'. Below these is a '<< Remove' button. On the right, there are three input fields: the first contains 'Defects', the second contains 'N', and the third is empty. To the right of these fields are three buttons: 'OK >>', 'Cancel', and 'Help'. Below the buttons, there are three radio buttons: 'Calculate Limits' (selected), 'Historical Limits', and 'Tests for Special Causes' (unchecked). To the right of these radio buttons are three input fields for 'UCL', 'CL', and 'LCL', all of which are empty. Below the radio buttons are two checkboxes: 'Sigma Zone Lines' (unchecked) and 'Add Title' (unchecked).

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 Tool

The Control Chart Selection Tool 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.

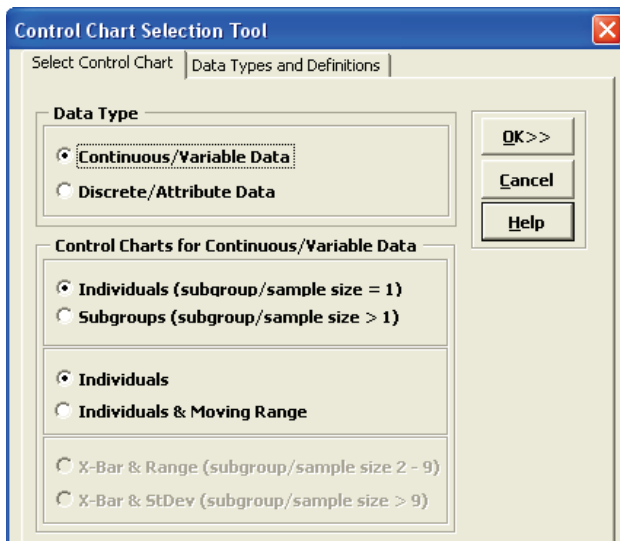
Subgroup/Sample Size is constant: The number of observations within your sample remains fixed over time.

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 > Advanced Charts > I-MR-R or I-MR-S**).

Control Chart Selection Tool – Individuals Chart

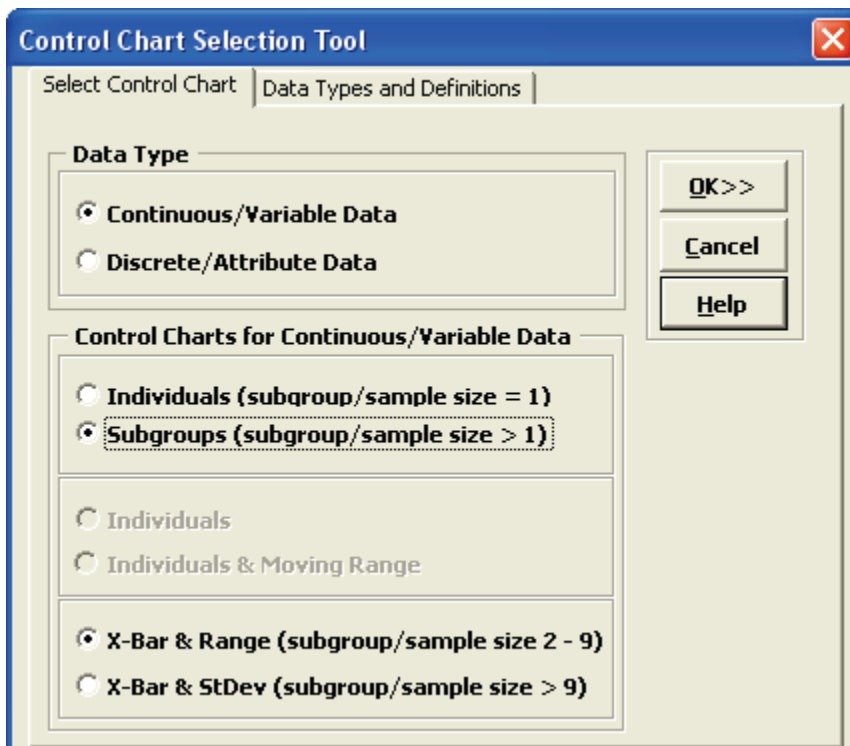
1. Open **Customer Data.xlsx**, click on **Sheet 1**. Click **SigmaXL > Control Charts > Control Chart Selection Tool**.
2. We would like to create a control chart of the Overall Satisfaction data. Since this can be considered as continuous data, the data type is **Continuous/Variable Data**. The subgroup/sample 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 Tool – X-Bar & R Chart

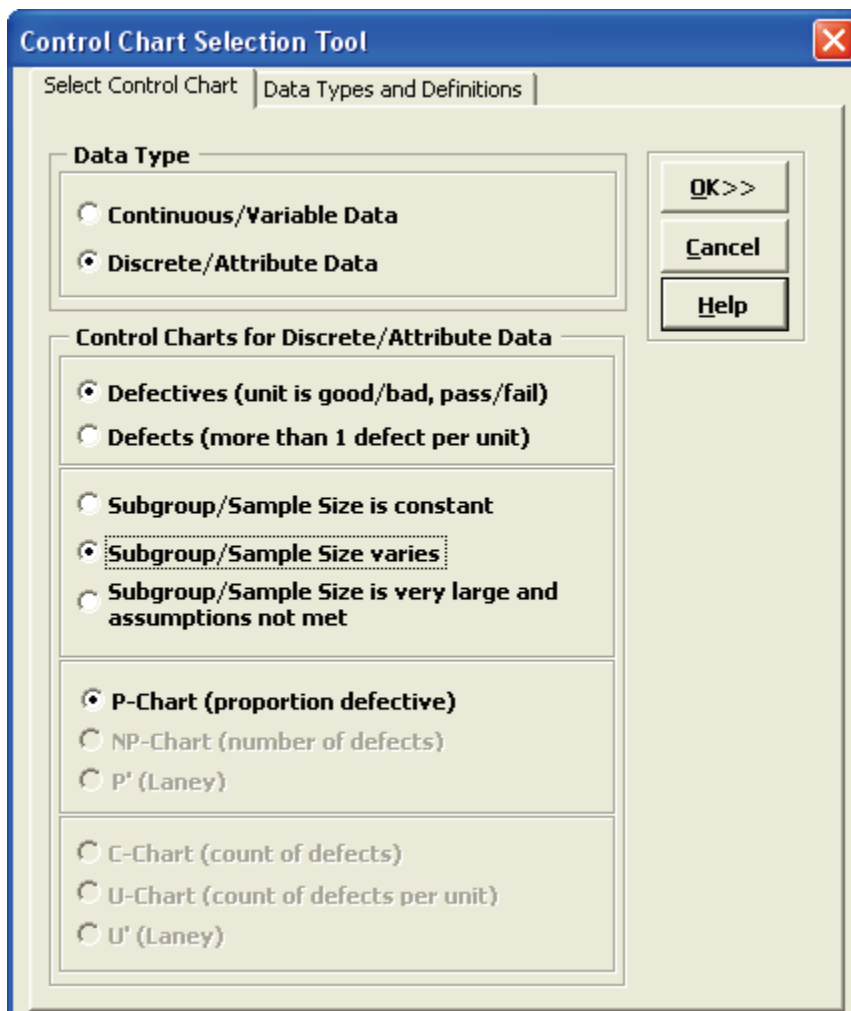
1. Open the file **Catapult Data – Xbar Control Charts.xlsx**. Each operator fires the ball 3 times. The target distance is 100 inches. Select **B2:F22**; here we will only use the first 20 subgroups to determine the control limits.
2. Click **SigmaXL > Control Charts > Control Chart Selection Tool**.
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 Tool – 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 Tool**.
3. Since this data is discrete, select **Discrete/Attribute Data**. We are looking at Defectives data since each loan 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 – Advanced Charts: I-MR-R/S

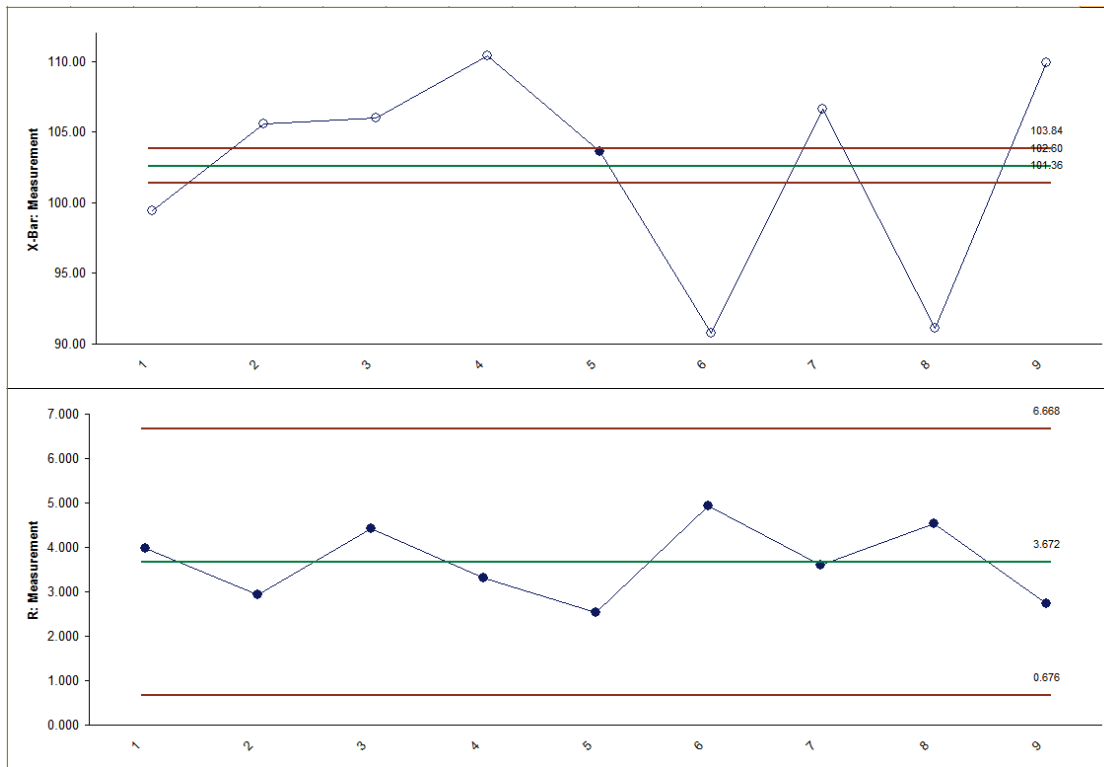
I-MR-R Charts

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 >>**.

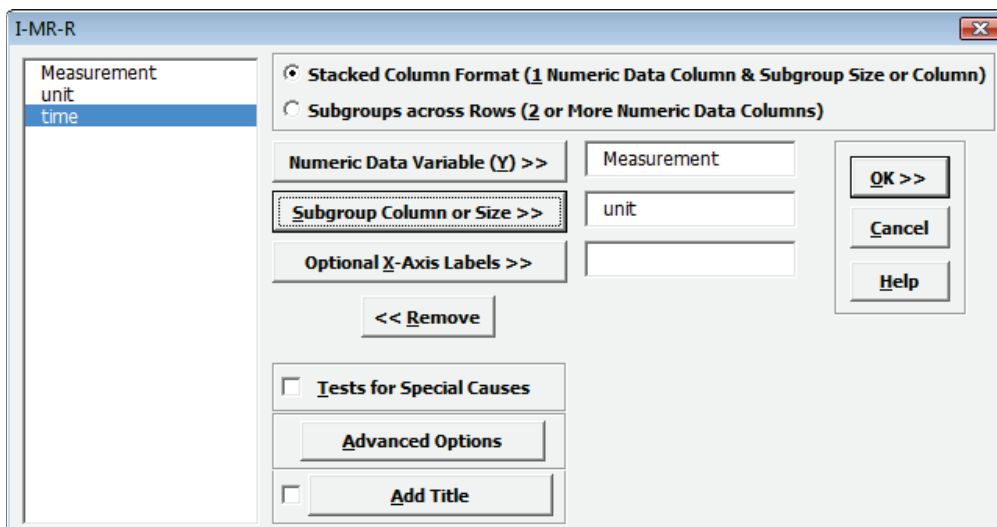
The screenshot shows the 'X-Bar and Range Chart' dialog box. On the left, a list of variables includes 'Measurement', 'unit', and 'time', with 'unit' selected. The main area has two radio buttons: 'Stacked Column Format (1 Numeric Data Column & Subgroup Size or Column)' (selected) and 'Subgroups across Rows (2 or More Numeric Data Columns)'. Below these are three buttons: 'Numeric Data Variable (Y) >>', 'Subgroup Column or Size >>', and 'Optional X-Axis Labels >>'. The first two buttons have text boxes next to them containing 'Measurement' and 'unit' respectively. To the right of these are 'OK >>', 'Cancel', and 'Help' buttons. Below the main area is a '<< Remove' button. At the bottom left, there are three checkboxes: 'Calculate Limits' (checked), 'Historical Limits' (unchecked), and 'Tests for Special Causes' (unchecked). Below these is a 'Sigma Zone Lines' checkbox (unchecked) and an 'Advanced Options' button. At the very bottom is an 'Add Title' checkbox (unchecked). On the right side of the dialog, there are two columns of input fields for control limits: 'X-Bar' and 'R'. The 'X-Bar' column has three fields for UCL, CL, and LCL. The 'R' column has three fields for UCL, CL, and LCL.

- Click **OK**. Resulting X-bar & R chart:

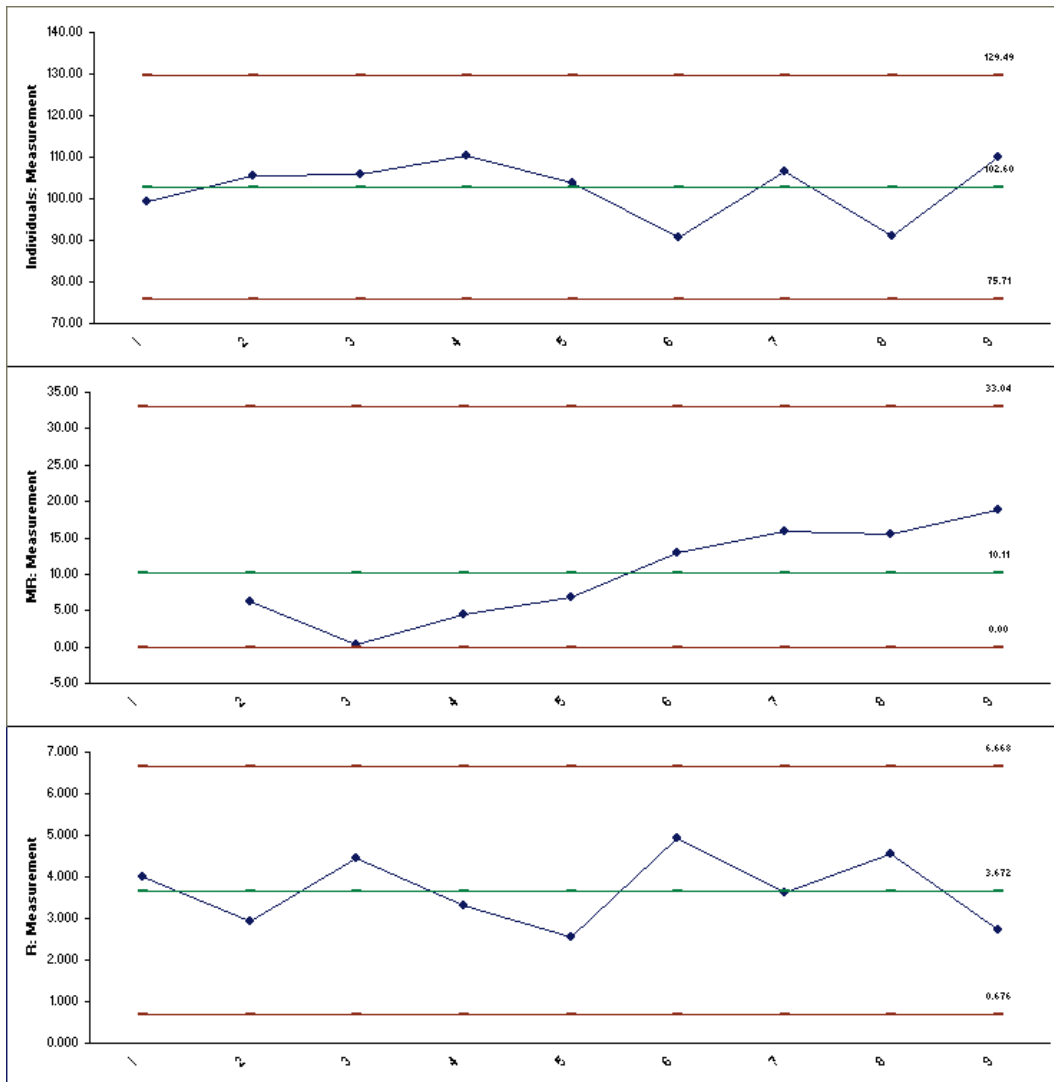


Almost all of the data points in the X-bar chart are out-of-control! This is due to the small within-subgroup variability (the control limits are calculated from the within-subgroup variability).

- Click Sheet **Between**. Select **SigmaXL > Control Charts > Advanced Charts > I-MR-R**.
- Click **Next**. Select **Stacked Column Format**. Select *Measurement*, click **Numeric Data Variable (Y) >>**; select *unit*, click **Subgroup Column or Size >>**.



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.

SigmaXL Appendix

SigmaXL® Version 8.0 Feature List Summary

Compatible with Excel 2016 for Windows and Mac: 32 & 64 bit

Menu Layout: Classical or DMAIC

Recall Last Dialog & Worksheet Manager

Data Manipulation:

- Subset by Category, Number, Date or Random
- Stack Subgroups Across Rows
- Stack and Unstack Columns
- Standardize Data
- Random Number Generator
 - Normal
 - Uniform (Continuous & Integer)
 - Lognormal
 - Weibull
 - Triangular
- Box-Cox Transformation

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)
- 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
- Control Chart Templates
 - Individuals, C-Chart

**** New features in SigmaXL® Version 8.0**

SigmaXL® Version 8.0 Feature List Summary

Graphical Tools:

- Basic and Advanced (Multiple) Pareto Charts
- EZ-Pivot/Pivot Charts: Easily create Pivot Tables and Charts
- 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 Boxplots, Dotplots
- Run Charts (with Nonparametric Runs Test allowing you to test for Clustering, Mixtures, Lack of Randomness, Trends and Oscillation)
- Overlay Run Chart
- Multiple Normal Probability Plots (with 95% confidence intervals to ease interpretation of normality/non-normality)
- Multi-Vari Charts
- Scatter Plots (with linear regression and optional 95% confidence intervals and prediction intervals)
- Scatter Plot Matrix
- 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

SigmaXL® Version 8.0 Feature List Summary

Statistical Tools:

- P-values turn red when results are significant: ($p\text{-value} < \alpha$)
- 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^{**}
- 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 $1.5 \times \text{IQR}$); Likely ($2.2 \times \text{IQR}$); Extreme ($3 \times \text{IQR}$)
 - 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^{**}
- Nonparametric Tests:
 - 1 Sample Sign and 1 Sample Wilcoxon
 - 2 Sample Mann-Whitney
 - Kruskal-Wallis and Mood's Median Test (with graph of Group Medians and 95% Median Confidence Intervals)
 - Runs Test
 - With Exact and Monte Carlo P-Value
- 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 ($1 - \text{Beta}$), 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.

SigmaXL® Version 8.0 Feature List Summary

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.
- 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
- $\pm 1, 2$ 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

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.

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 Govindarajulu [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	is the degrees of freedom in the sample $(n - 1)$.
\bar{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 p th 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:

$$L = \bar{X} - ks$$

$$U = \bar{X} + ks$$

The value of k is shown below:

Normal Tolerance Interval - one sided exact

$$k = \frac{t_{1-\alpha, v, \lambda}}{\sqrt{n}}$$

where $t_{1-\alpha, v, \lambda}$ is the $1 - \alpha$ percentile of the non-central t distribution with v degrees of freedom and non centrality parameter $\lambda = \sqrt{n}Z_p$.

Normal Tolerance Interval - two sided exact

$$G(z, n, p, k) = 1 - F_v \left[\frac{v \chi_{p, 1, \lambda}^2}{k^2} \right]$$

where: $\chi_{p, 1, \lambda}^2$ 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:

$$\begin{aligned} L &= x_r \\ U &= x_s \end{aligned}$$

The interpolated lower L and upper U bounds are given by:

$$\begin{aligned} L &= \lambda_l x_{r+1} + (1 - \lambda_l) x_r \\ U &= \lambda_u x_{s+1} + (1 - \lambda_u) x_s \end{aligned}$$

Confidence Interval – two sided exact

$$\begin{aligned} r &= B_{\alpha/2, n, p} \\ s &= B_{1-\alpha/2, n, p} + 1 \end{aligned}$$

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

$$\begin{aligned} \lambda_l &= \left(1 + \frac{r(1-p)(B_{n,p}(r) - \alpha/2)}{(n-r)p(\alpha/2 - B_{n,p}(r-1))} \right)^{-1} \\ \lambda_u &= \left(1 + \frac{s(1-p)(B_{n,p}(s) - (1-\alpha/2))}{(n-s)p(1-\alpha/2 - B_{n,p}(s-1))} \right)^{-1} \end{aligned}$$

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

$$\begin{aligned} k &= B_{1-\alpha, n, p} + 1 \\ r &= \lfloor (n - k + 1)/2 \rfloor \\ s &= n - r + 1 \end{aligned}$$

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 Govindarajulu (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 v_{b1} and b_2 ", *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 $\alpha > 0$.

Shape2 parameter $\beta > 0$.

$\Gamma(\beta)$ 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 $0 \leq x \leq 1$.

Shape1 parameter $\alpha > 0$.

Shape2 parameter $\beta > 0$.

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 $\Gamma(\beta)$ is the Gamma function:

$$\Gamma(\beta) = \int_0^{\infty} e^{-t} t^{\beta-1} dt$$

Range $0 \leq x < \infty$.

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 $0 \leq x - \theta < \infty$.

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 $0 \leq x < \infty$.

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

$$z = \gamma + \eta \ln \left(\frac{x - \epsilon}{\lambda + \epsilon - x} \right)$$
$$x = \epsilon + \frac{1}{1 + e^{-(z-\gamma)/\eta}}$$

z is $N(0, 1)$

Range $\epsilon < x < \epsilon + \lambda$.

Location parameter $\eta > 0$.

Scale parameter $\lambda > 0$.

Shape parameter γ unbounded.

Shape parameter ϵ unbounded.

Johnson Transformation - S_L

$$z = \gamma + \eta \ln(x - \epsilon)$$

$$x = \epsilon + e^{(z-\gamma)/\eta}$$

z is $N(0, 1)$

Range $x > \epsilon$

Location parameter $\eta > 0$.

Shape parameter γ unbounded.

Shape parameter ϵ unbounded.

Johnson Transformation - S_U

$$z = \gamma + \eta \sinh^{-1}[(x - \epsilon)/\lambda]$$

$$x = \epsilon + \lambda \sinh[(z - \gamma)/\eta]$$

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.

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 Hirotugu 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:

$$Ppu = (USL - 50^{\text{th}} \text{ percentile}) / (99.865 \text{ percentile} - 50^{\text{th}} \text{ percentile})$$

$$Ppl = (50^{\text{th}} \text{ percentile} - LSL) / (50^{\text{th}} \text{ percentile} - 0.135 \text{ percentile})$$

$$Ppk = \min(Ppu, Ppl)$$

$$Pp = (USL - LSL) / (99.865 \text{ percentile} - 0.135 \text{ percentile})$$

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 $r \times c$ 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, Cochran’s rule [5] for $r \times c$ 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 5×6 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 Cochran’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 of 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

$$\Pr(|stat - \mu| \geq |statcrit - \mu|)$$

summed over all possible combinations, where $statcrit$ is the statistic evaluated for each possibility, and μ 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: $N \leq 50$ (exact is automatically utilized in the regular Sign Test)

Wilcoxon Signed Rank: $N \leq 15$

Mann-Whitney: Each sample $N \leq 10$

Kruskal-Wallis: Each sample $N \leq 5$

Mood's Median: Each sample $N \leq 10$

Runs Test (Above/Below) or Runs Test (Up/Down) Test: $N \leq 50$

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 $r \times c$ 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 $r \times c$ 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 $r \times c$ 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).

Recommended Nonparametric Tests – One Sample

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 than 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 H_a = “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. $\log_{10}(\text{Maximum Skewness NE}) = 0.5391 - 0.5028 \log_{10}(N)$ (R-Square = 99.5%)
2. $\log_{10}(\text{Maximum Skewness GT}) = 0.2380 - 0.5725 \log_{10}(N)$ (R-Square = 98.1%)
3. $\log_{10}(\text{Maximum Skewness LT}) = 0.2796 - 0.5459 \log_{10}(N)$ (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 H_a = “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 $N \leq 50$.

If the Wilcoxon Test is appropriate and $N \leq 15$, 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 $N \leq 10$, 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 ($2.2 \times \text{IQR}$) or Extreme ($3 \times \text{IQR}$), then Mood's Median is recommended. See **Outliers (Boxplot Rules)** below for further details.

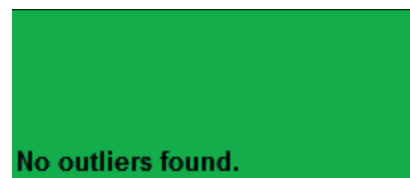
If Kruskal-Wallis is appropriate and each sample $N \leq 5$, then Kruskal-Wallis – Exact is recommended (using the rules of thumb in [Sample Sizes for Exact](#)).

If Mood's Median is appropriate and each sample $N \leq 10$, then Mood's Median – Exact is recommended.

Outliers (Boxplot Rules)

Each sample is tested for outliers using Tukey's Boxplot Rules: Potential ($> Q3 + 1.5 \times \text{IQR}$ or $< Q1 - 1.5 \times \text{IQR}$); Likely: $2.2 \times \text{IQR}$; Extreme: $3 \times \text{IQR}$. 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:



If a Potential or Likely outlier is found, the cell is highlighted as yellow:

Potential ($1.5 \times \text{IQR}$) 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 ($2 \times \text{IQR}$) 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 ($3 \times \text{IQR}$) 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 ($3 \times \text{IQR}$) 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 $\geq .05$, 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 $\geq .05$, 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 assess 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 \bar{P} be the measure of agreement between the assessors, and \bar{P}_e be the measure of agreement that would occur by chance. Then:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

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 ($\bar{\rho}_s$) of the pairwise Spearman correlations using the relationship [7]:

$$W = \frac{(k-1)\bar{\rho}_s + 1}{k}$$

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 (ASE₁) of the statistic. See [9,12] for ASE₁ formulas.

Pearson'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 $0 \leq V \leq 1$. 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 $-1 \leq \rho \leq 1$. 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 (ASE₁) of the statistic. See [9,12] for ASE₁ 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 :

$$s = 2 \sqrt{\frac{\sum_{i,j} x_{ij} (C_{ij} - D_{ij})^2 - (P - Q)^2 / n}{D_r D_c}} \text{ under the null hypothesis,}$$

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", *Psychometrika* 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:

$$z = |x - \text{median}(x)|$$

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:

$$w_j = \sqrt{(n - n_j) / (n n_j)}$$

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:

$$LDL_{fj} = \bar{x} - h_\alpha \sqrt{MSE w_{fj}}$$

$$UDL_{fj} = \bar{x} + h_\alpha \sqrt{MSE w_{fj}}$$

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 $p_j = \sum_i^n x_{ij}/n_j$.

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:

$$LDL_j = \bar{p} - h_\alpha s_p w_j$$

$$UDL_j = \bar{p} + h_\alpha s_p w_j$$

where the average proportion is given by:

$$\bar{p} = \frac{\sum_j n_j p_j}{N}$$

where the pooled variance is:

$$s_{p_j}^2 = \bar{p}(1 - \bar{p})/n_j$$

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

$$LDL = \bar{p} - h_\alpha s_p w$$

$$UDL = \bar{p} + h_\alpha s_p w$$

where LDL and UDL are the lower/upper bounds respectively,
where the average proportion is given by:

$$\bar{p} = \frac{\sum_j p_j}{k_1}$$

where the pooled variance is:

$$s_p^2 = \bar{p}(1 - \bar{p})/n$$

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

and where the likelihood is specified as:

$$\text{llf} = \sum n (p \ln(\hat{p}) + (1 - p) \ln(1 - \hat{p}))$$

¹ A similar analysis follows for factor 2.

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 $\bar{c}_j = \sum_i^j c_{ij}/n_j$.

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:

$$\bar{c} = \frac{\sum_j n_j \bar{c}_j}{N}$$

where the pooled variance is:

$$s_p^2 = \bar{c}$$

where the weighting factor for each group j is:

$$w_j = \sqrt{(N - n_j)/(N n_j)}$$

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:

$$q = -2(\text{llf}_u - \text{llf}_r)$$

where llf_u is the unrestricted log likelihood,

where llf_r is the restricted log likelihood

and where the likelihood is specified as:

$$\text{llf} = \sum y \ln(\lambda) - \lambda - \ln(y!)$$

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 U_{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,



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