DiscoverSim™ Version 2.2 Workbook

Contact Information:
Technical Support: 1-866-475-2124 (Toll Free in North America) or
1-416-236-5877
Sales: 1-888-SigmaXL (888-744-6295)
E-mail: support@SigmaXL.com Web: www.SigmaXL.com

Published: July 28 2023

Table of Contents

DiscoverSim™ Feature List Summary, Installation Notes, System Requirements and Getting Help

DiscoverSim™ Version 2.2 Feature List Summary
Key Features (bold denotes new in Versions 2.0, 2.1 and 2.2):

- Excel add-in for Monte Carlo Simulation and Optimization
- Uses Six Sigma Language: Specify Inputs (X's) and Outputs (Y's)
- Monte Carlo simulation with 53 Continuous and 10 Discrete Distributions
- Batch Distribution Fitting for all Continuous and Discrete Distributions (excluding custom)
- **Nonnormal Process Capability for all Continuous and Discrete Distributions (excluding custom)**
- **Percentiles to Parameters Calculator**
- Specify Input Correlations
- Excel Formula Interpreter for Accelerated Calculations
- Sensitivity Analysis based on correlation or stepwise regression
- Optimization using Input Controls:
    - **Mixed Integer Distributed Ant Colony Global Optimization (MIDACO)**
    - Genetic Algorithm (GA) Global Optimization
    - Sequential Quadratic Programming (SQP) for fast local optimization
    - **Nelder-Mead (NM) simplex optimization for fast local non-smooth problems**
    - **Exhaustive Discrete optimization – all combinations for small discrete problems**
    - **Powerful hybrid of the above methods:**
        1. All discrete controls: MIDACO or Exhaustive Discrete if applicable
        2. All continuous controls: MIDACO, GA, followed by SQP or NM
        3. Mixed continuous/discrete controls: MIDACO 1 (Course), MIDACO 2 (Fine), followed by SQP or NM
    - Specify linear and nonlinear constraints
    - Stochastic Optimization – minimize dpm or maximize Ppk
    - **Insert DSIM functions such as Percentiles or Capability Metrics. These can then be used in constraints.**
    - Multiple Response Optimization using Desirability, Weighted Sum or Deviation from Target
- **Recall DiscoverSim Dialog button.**
- DiscoverSim is also bundled with SigmaXL **Version 10.0** for statistical and graphical analysis.

Common Continuous Distributions:

1. Normal
2. Triangular
3. Uniform
4. Pearson Family (specify Mean, StdDev, Skewness, Kurtosis)
5. Log Normal**
6. Exponential**
7. Weibull**
8. PERT
9. Custom Continuous (**Weighted and Unweighted**)

Advanced Continuous Distributions:

1. Beta
2. Beta (4 Parameter)
3. Box-Cox**
4. Burr**
5. Cauchy
6. Chi-Squared**
7. Chi-Squared with Scale (3 Parameter)**
8. Error Function (ERF)
9. F**
10. F with Scale (4 Parameter)**
11. Fatigue Life**
12. Fisk**
13. Folded Normal
14. Frechet**
15. Gamma**
16. Generalized Error
17. Generalized Gamma**
18. Generalized Logistic
19. Generalized Pareto
20. Half Normal**
21. Inverse Gamma**
22. Inverse Gaussian
23. Johnson SB
24. Johnson SL
25. Johnson SU
26. Laplace
27. Largest Extreme Value
28. Levy
29. Logistic
30. Log Gamma**
31. Log Logistic**
32. Maxwell
33. Non-Central Chi-Squared**
34. Non-Central F**
35. Non-Central T
36. Pareto
37. Power
38. Rayleigh
39. Reciprocal
40. Skew Normal
41. Smallest Extreme Value
42. Student's T
43. Student's T with Location and Scale (3 Parameter)
44. Von Mises

** denotes with/without Threshold

Discrete Distributions:

1. Bernoulli (Yes/No)
2. Binomial
3. Geometric
4. Hypergeometric
5. Logarithmic
6. Negative Binomial
7. Poisson
8. Step
9. Uniform (Discrete)
10. Custom Discrete (**Weighted and Unweighted**)

Stochastic Information Packet (SIP):

- **Import Stochastic Information Packet (SIP).**
- A SIP is a standard library of data (see ProbabilityManagement.org for more details on this standard).

What's New in Version 2.2

New features in DiscoverSim Version 2.2 include:

- DiscoverSim is bundled with SigmaXL Version 10.0 for statistical and graphical analysis. See "What's New in SigmaXL Version 10.0" in the SigmaXL workbook.
- **Recall DiscoverSim Dialog** button. Recall is used to activate the last data worksheet and recall the last dialog, making it very easy to do repetitive analysis.
- Accelerated Simulation and Optimization support for new Excel 2019 functions: IFS, MAXIFS, MINIFS and SWITCH.

What's New in Version 2.1

New features in DiscoverSim Version 2.1 include:

- **Percentiles to Parameters Calculator**. If one does not have sufficient data to do distribution fitting but does know some percentile values of a specified distribution, these can be converted to parameter values for use in simulation.
- The performance of **DSIM_IsAllDifferent** has been dramatically improved within the **Genetic Algorithm** and **Discrete Exhaustive** optimization. The function is now recognized as an internal constraint so that only solutions that have all different integer values are considered. Case Study 7 has been added to demonstrate this feature using the Travelling Salesperson Problem.
- **MIDACO Solver Version 6.0** with improved algorithmic performance, internal local solver (backtracking line search) and AllDifferent constraint support.
- Speed improvement in DiscoverSim's **Copy/Paste/Clear Cells**.
- Accelerated mode support for Excel **TEXT** functions or IF(A1="Yes",…) and **OFFSET** function referencing DiscoverSim inputs. Note, however, that simulation and optimization speed will be faster if these can be avoided in the model.

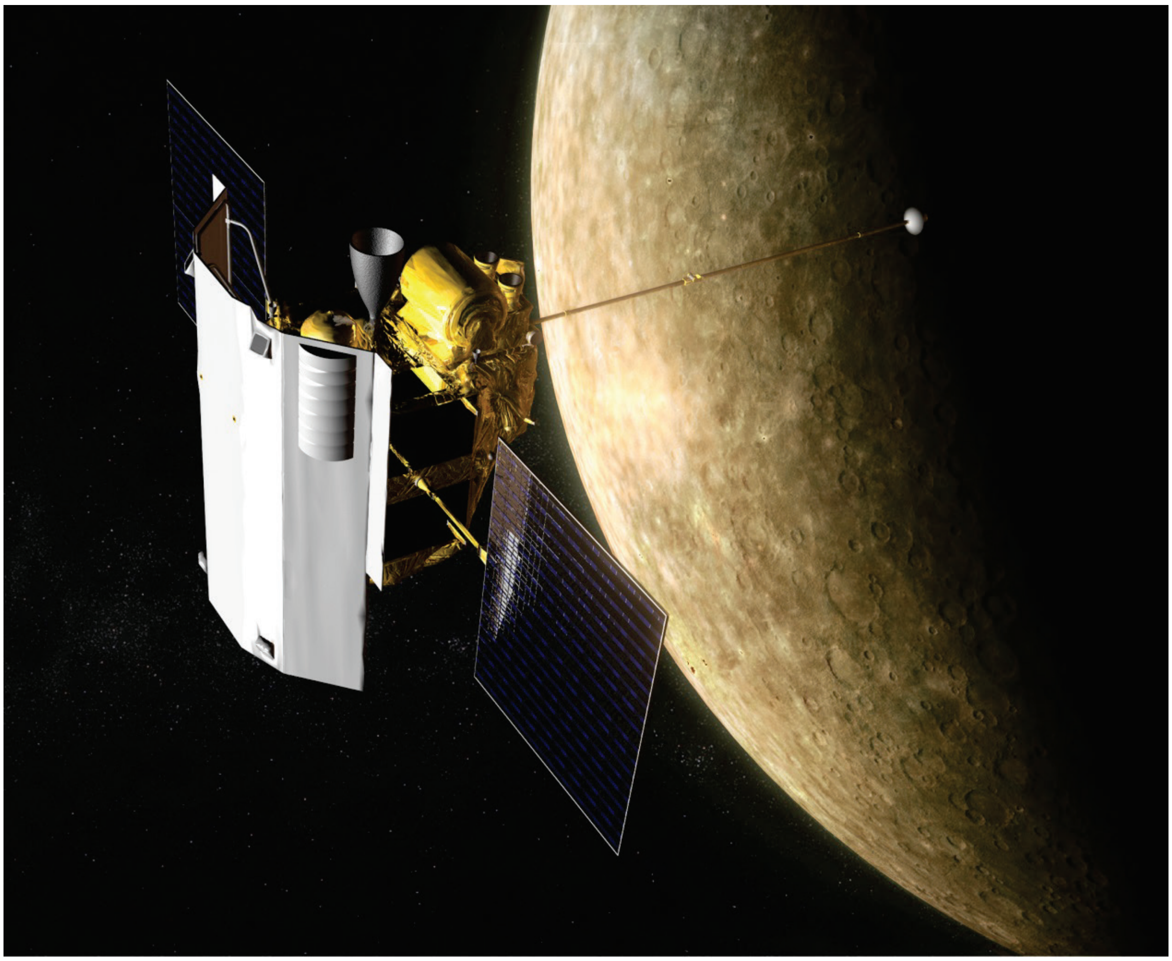
Case Study 7 - Travelling Salesperson Problem

What's New in Version 2.0

DiscoverSim and MIDACO:

DiscoverSim is now bundled with MIDACO, one of the world's strongest evolutionary solvers for mixed discrete, continuous, constrained global optimization.

MIDACO stands for **M**ixed **I**nteger **D**istributed **A**nt **C**olony **O**ptimization. It is suitable for problems with up to several hundreds to some thousands of optimization variables. It was developed in collaboration with the European Space Agency (ESA) and represents the state-of-the-art in interplanetary space trajectory optimization. MIDACO holds the benchmark world record best solution to Full Messenger (Mission to Mercury), which is considered the most difficult space trajectory problem in the ESA Global Trajectory Optimization Database.



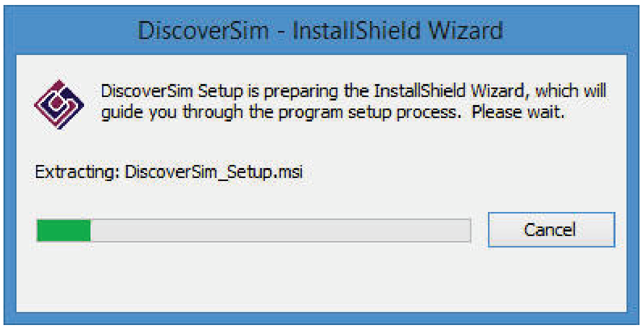
Other new features include:

- **Nelder-Mead** simplex optimization for fast local non-smooth problems.
- **Exhaustive Discrete** optimization – all combinations for small discrete problems.
- **Genetic Algorithm** improved constraint handling using MIDACO's Oracle method.
- **Nonnormal Process Capability** for all 53 continuous and 10 discrete distributions.
- Improved robustness on parameter estimation for distribution fitting using the methods of BFGS and Nelder-Mead.
- Threshold distributions solved using a hybrid of Maximum Likelihood and Iterative Bias Reduction. Added option to exclude threshold distributions from distribution fitting.
- Anderson Darling P-Values for all continuous distributions via table lookup. Tables with critical values computed using extensive Monte Carlo simulations.
- **Discrete Control** now includes a **Step** size, so is no longer limited to integer increments.
- Insert **DSIM function** such as Percentile or Capability Metric. This can then be used in constraints.
- New **Percentile** statistic for optimization.
- **Weighted** and **Unweighted** Custom Distributions
- Import Stochastic Information Packet (**SIP**). A SIP is a standard library of data (see ProbabilityManagement.org for more details on this standard). DiscoverSim treats this as a custom distribution with unweighted data and is sampled sequentially.
- **Correlations > Reset with Blanks** allows you to specify correlations between inputs without the constraint of requiring independence on the other inputs.

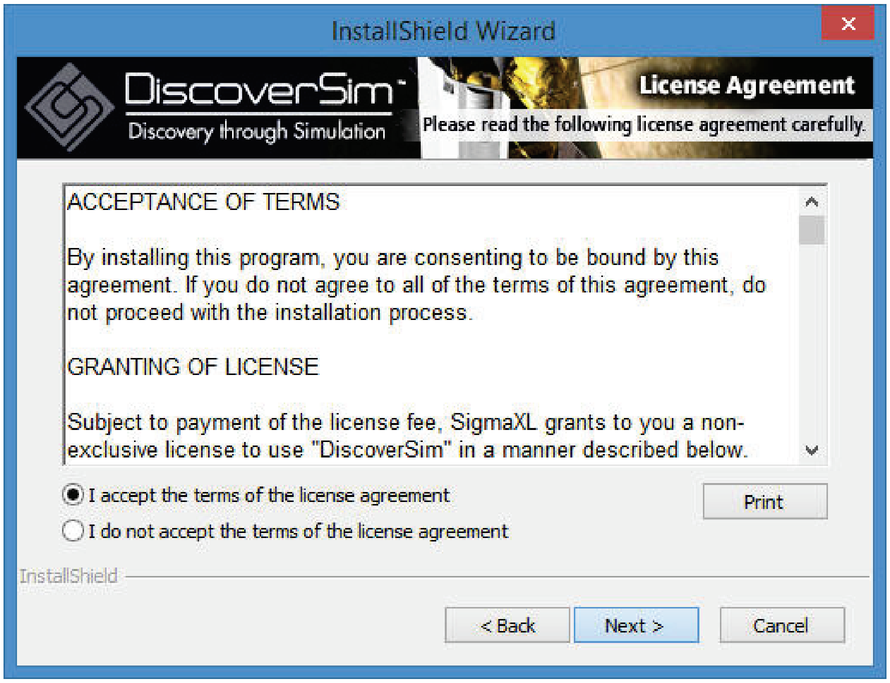
Installation Notes

1. **This installation procedure requires that you have Administrator rights to install the software on your computer. Also, please ensure that you have the latest Microsoft Office service pack by using Windows Update before installing DiscoverSim.**

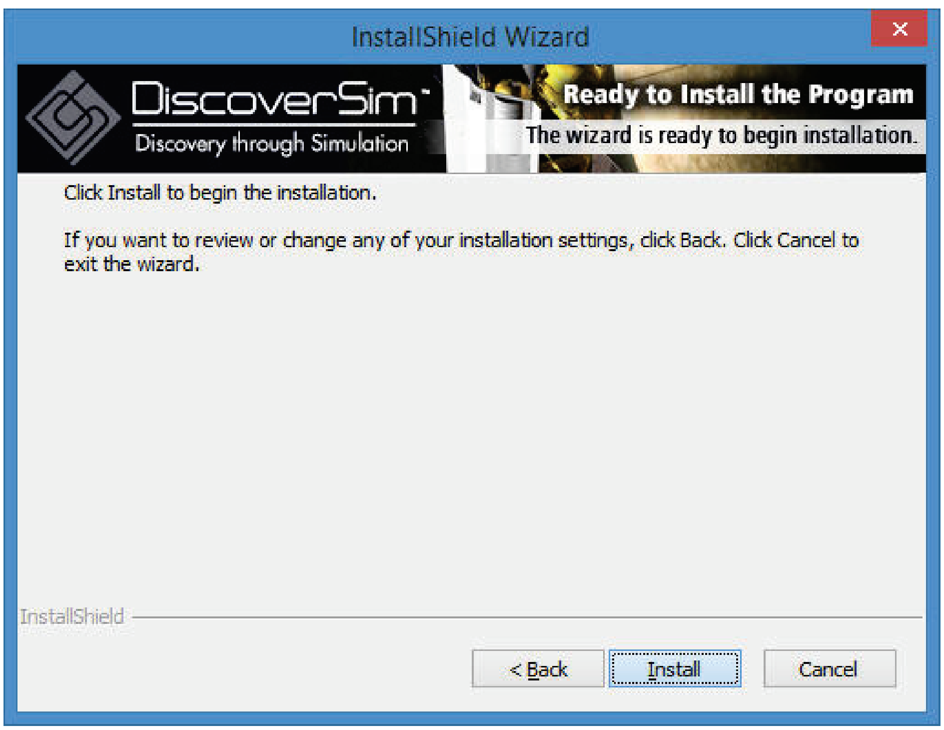
2. Download DiscoverSim from the link that is provided. Double-click on the file **DiscoverSim_Setup.msi**. The setup file will launch the installation of DiscoverSim. Click on the "Next" button to begin the installation.



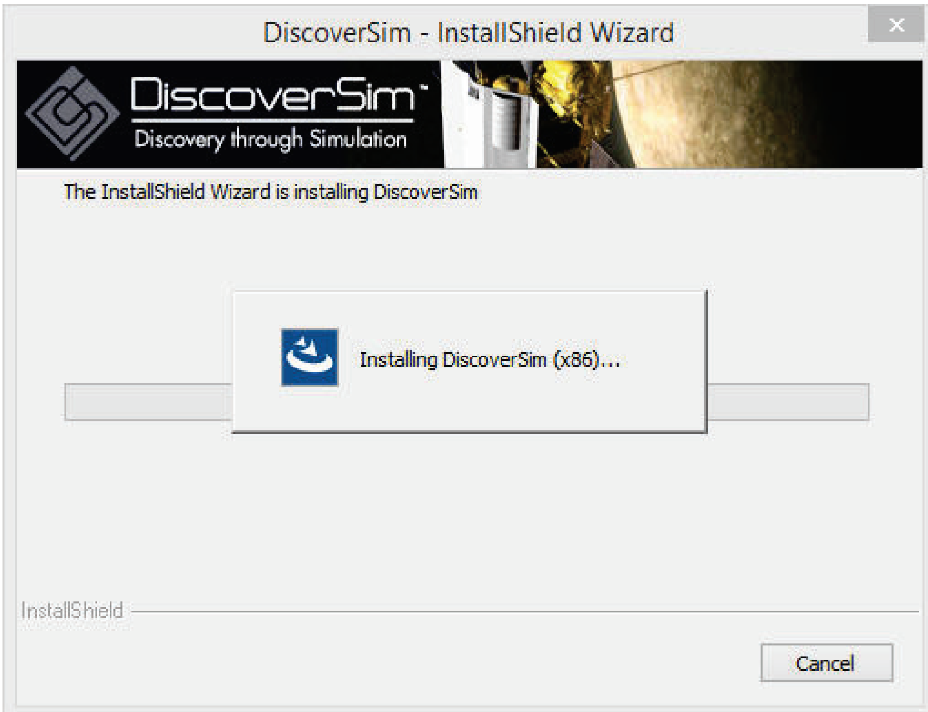
3. Read the License Agreement and select the option "I accept the terms of the license agreement" to continue.



4. Press the “Install” button.



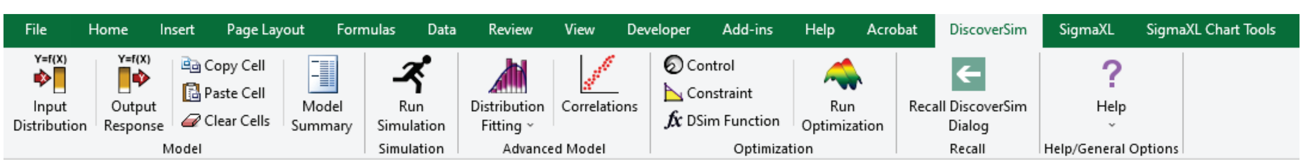
The wizard will then install the version of DiscoverSim which is required for your system. A progress bar window will appear as the files are transferred.



5. Press the “Finish” button.

6. To start DiscoverSim, double-click on the DiscoverSim desktop icon or click **Start > Programs >All Programs > SigmaXL > DiscoverSim**.

7. The DiscoverSim Ribbon appears as shown below:



Note: SigmaXL is bundled with DiscoverSim, so the SigmaXL menu tab is also available. See the SigmaXL workbook for further details.

DiscoverSim™ System Requirements

Minimum System Requirements:

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

Memory: 4 GB of RAM or greater.

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

Display: 1280x768 or higher resolution monitor.

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

Microsoft Windows Excel version: Excel 2016+, Office 365 with latest service packs installed. 64 bit versions of Excel are highly recommended for large simulation or optimization models.

Administrative Rights to install software.

Getting Help and Product Registration

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

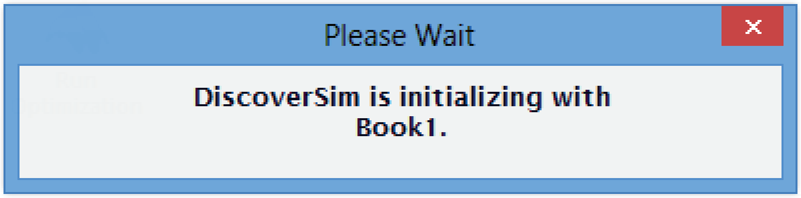
Technical support is available by phone at 1-866-475-2124 (toll-free in North America) or 1-519-579-5877 or e-mail support@sigmaxl.com.

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

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

Initializing DiscoverSim Workbook

When you create a new workbook, and click DiscoverSim Input Distribution or Input Control, the following message appears:



Initialization is only done once in a workbook. This creates the following hidden worksheets:

- Options
- XLDta
- DSim_Correlations

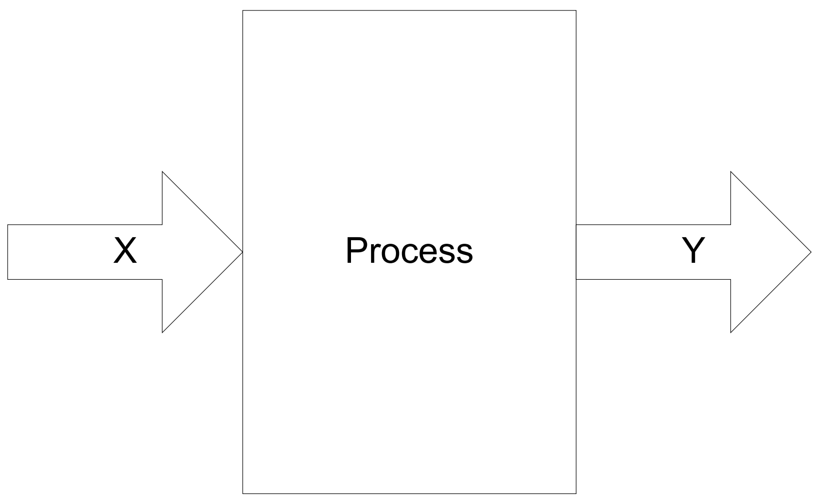
These hidden sheets should not be modified. Modifications will likely corrupt the DiscoverSim model.
If you create a DiscoverSim model, and then wish to completely delete it, you can delete all of the above hidden worksheets using **DiscoverSim > Help > Clear DiscoverSim Data**.

Introduction to DiscoverSim™, Getting Started Tutorial and DiscoverSim™ Menu & Dialogs

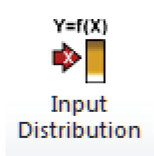
Introduction to Monte Carlo Simulation and Optimization with DiscoverSim™

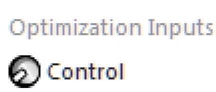
The Y=f(X) Model

DiscoverSim utilizes the "Y=f(X)" model, where Y denotes a key process output response and X denotes a key process input. This is shown pictorially as:



The DiscoverSim menu buttons to specify inputs are:





An input distribution is stochastic and includes random variation. It can be continuous or discrete. The most common continuous distribution is the normal distribution with parameters mean and standard deviation. DiscoverSim includes 53 continuous and 10 discrete distributions.

An input control is set to a constant value, but it can be varied like the temperature control knob of a thermostat and is used for optimization. An input control can be continuous or discrete.

The DiscoverSim menu button to specify an output response is:

A DiscoverSim model requires at least one input and one output. The output cell must contain an Excel formula that is a function of the inputs and references the cell address of the inputs or intermediate calculations including the inputs. The formula can be linear or non-linear and include Excel functions. The formula can reference other sheets within the workbook, but should not reference other workbooks.

The DiscoverSim Excel Formula Interpreter is used to dramatically accelerate the speed of calculations during simulation or optimization.

The Y = f(X) equation should be based on theory, process knowledge, or the prediction formula of a designed experiment or regression analysis. In Design for Six Sigma (DFSS), this is referred to as the "Transfer Function." This prediction equation should be validated prior to use in DiscoverSim. As the eminent statistician George Box said, "All models are wrong, some are useful," so while the model does not have to be exact, it should be a reasonable approximation of reality. The results of a DiscoverSim analysis should also be validated with further experimentation or use of prototypes.

Monte Carlo Simulation

The Y = f(X) model gives us a starting point in the relationship between X and Y. After the Y = f(X) relationship has been validated, an important question that then needs to be answered is: "What does the distribution of Y look like when I cannot hold X constant, but have some uncertainty in X?" In other words, "How can I quantify my risk?"

Take an example of a sales forecast. Y is the predicted monthly sales, which is a sum of forecasts from five different product line sales managers. If each manager simply reports a single "most likely" estimate, the total predicted sales will also be a single "most likely" estimate. The problem with this approach is obvious; it does not take the uncertainty into account. So what then should be done? In addition to the most likely estimates, each sales manager could include best and worst case estimates. Alternatively, an estimate of the mean and standard deviation could be reported.

The challenge then becomes how to best take the uncertainty into account. One could simply sum the best and worst case values (or the +/-3 standard deviation values), and thus report the total range. This approach, however, is unsatisfactory because it does not take into account the very low likelihood that each line will simultaneously have a best case or worst case value. This would be akin to rolling 5 dice and getting 5 ones or 5 sixes.

If each of the sales managers assume a normal distribution, the distribution for total sales will also be normal with:

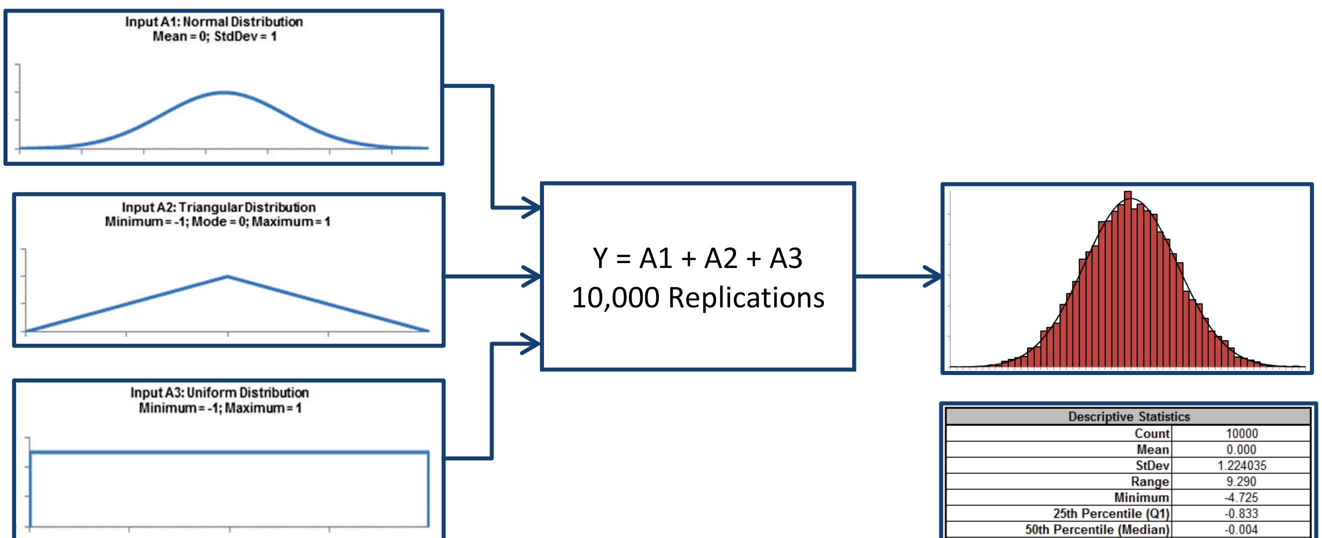
Total Mean = $\text{Mean}_1 + \ldots + \text{Mean}_5$

Total StdDev = $\text{SQRT}((\text{StdDev}_1)^2 + \ldots + (\text{StdDev}_5)^2)$.

If, however, some of the product lines have correlated sales, i.e., an increase in sales for product A also means an increase in sales for product B, then "things get complicated." The StdDev formula given above does not hold because it assumes independence. Further complications arise if the sales managers need to use different distributions or if the total sales involve more than a simple sum. This is where Monte-Carlo simulation comes in to solve the complex problem of dealing with uncertainty by "brute force" using computational power.

The Monte Carlo method was coined in the 1940s by John von Neumann, Stanislaw Ulam and Nicholas Metropolis, while they were working on nuclear weapon projects in the Los Alamos National Laboratory. It was named in homage to Monte Carlo Casino, a famous casino, where Ulam's uncle would often gamble away his money
(Ref: http://en.wikipedia.org/wiki/Monte_Carlo_method).

The following diagram illustrates a simple Monte Carlo simulation using DiscoverSim with three different input distributions (X's also known as "Assumptions"):



Descriptive Statistics	
Count	10000
Mean	0.000
StDev	1.224035
Range	9.290
Minimum	-4.725
25th Percentile (Q1)	-0.833
50th Percentile (Median)	-0.004
75th Percentile (Q3)	0.836
Maximum	4.565

Normality Tests	
Anderson-Darling Normality Test	0.234853
p-value (A-D Test)	0.7926
Skewness	0.014145774
p-value (Skewness)	0.5634
Kurtosis	-0.075178812
p-value (Kurtosis)	0.1159

A random draw is performed from each input distribution, Y is calculated, and the process is repeated 10,000 times. The histogram and descriptive statistics show the simulation results. While not shown here, requirements (specification limits) can easily be added to Y to obtain probability of nonconformance.

Notice that even though input distributions A2 and A3 are not normal, the distribution for the total is normal. This is due to the central limit theorem. Of course, this will not always be the case when performing a Monte-Carlo simulation!

The areas of application for Monte-Carlo simulation are very wide, including Design for Six Sigma (DFSS), Tolerance Design, Project Management and Risk Management with common use in Engineering, Finance, Telecommunications and Oil & Gas Exploration.

The following books are recommended for further reading:

Savage, Sam (2009), *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty,* Hoboken, NJ, Wiley.

Sleeper, Andrew (2006), *Design for Six Sigma Statistics: 59 Tools for Diagnosing and Solving Problems in DFSS Initiatives,* NY, McGraw-Hill.

Components of Uncertainty

There are two components to uncertainty: one is the actual process variation, and the second is measurement error or error due to lack of knowledge. The contribution of the latter can be incorporated into a model, but for simplicity in our discussions and case studies, we will assume that the measurement error is negligible.

Selecting a Distribution

Selecting the correct distribution is a critical step towards building a useful model.

The best choice for a distribution is one based on known theory. For example, the use of a Weibull Distribution for reliability modeling.

A common distribution choice is the Normal Distribution, but this assumption should be verified with data that passes a normality test with a minimum sample size of 30; preferably 100.

If data is available and the distribution is not normal, use DiscoverSim's Distribution Fitting tool to find a best fit distribution. Alternatively, the Pearson Family Distribution allows you to simply specify Mean, StdDev, Skewness and Kurtosis.

In the absence of data or theory, commonly used distributions are: Uniform, Triangular and PERT. Uniform requires a Minimum and Maximum value, and assumes an equal probability over the range. This is commonly used in tolerance design. Triangular and PERT require a Minimum, Most Likely (Mode) and Maximum. PERT is similar to Triangular, but it adds a "bell shape" and is popular in project management.

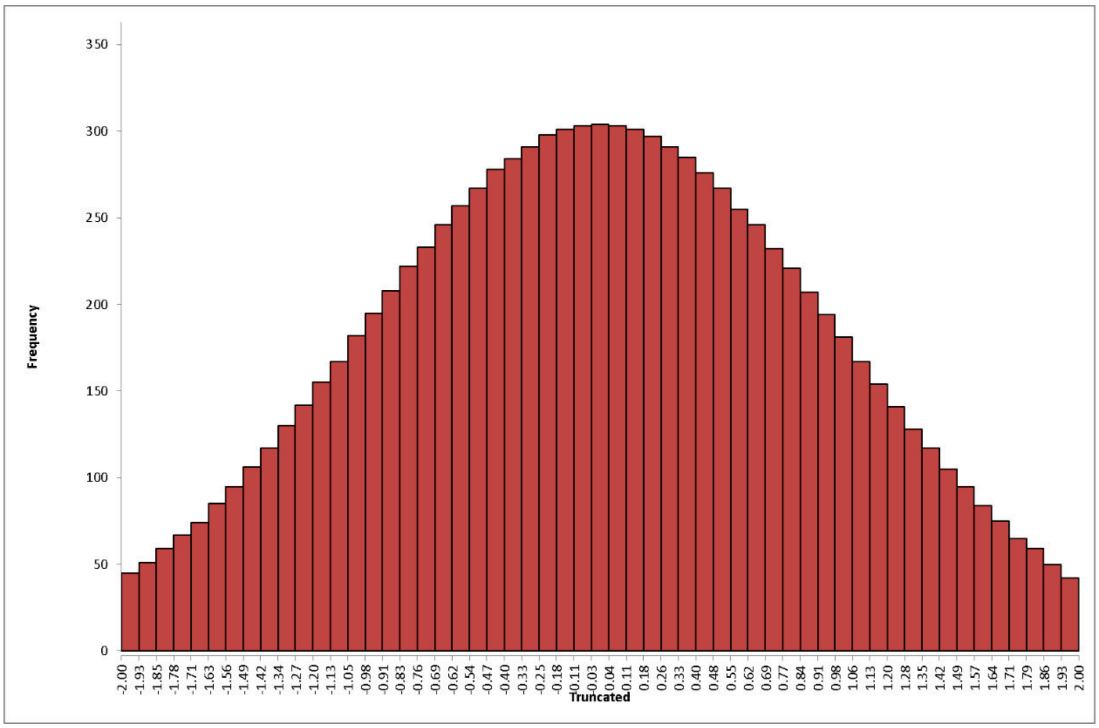
The following table is a summary of all distributions in DiscoverSim (** denotes with/without Threshold).

Common Continuous	**Advanced Continuous**	**Discrete**
• Normal • Triangular • Uniform • Pearson Family (specify Mean, StdDev, Skewness, Kurtosis) • Log Normal** • Exponential** • Weibull** • PERT • Custom Continuous	• Beta • Beta (4 Parameter) • Box-Cox** • Burr** • Cauchy • Chi-Squared** • Chi-Squared with Scale (3 Parameter)** • Error Function (ERF) • F** • F with Scale (4 Parameter)** • Fatigue Life** • Fisk** • Folded Normal • Frechet**	• Bernoulli (Yes/No) • Binomial • Geometric • Hypergeometric • Logarithmic • Negative Binomial • Poisson • Step • Uniform (Discrete) • Custom Discrete • SIP (Stochastic Information

Common Continuous	**Advanced Continuous**	**Discrete**
	• Gamma** • Generalized Error • Generalized Gamma** • Generalized Logistic • Generalized Pareto • Half Normal** • Inverse Gamma** • Inverse Gaussian • Johnson SB • Johnson SL • Johnson SU • Laplace • Largest Extreme Value • Levy • Logistic • Log Gamma** • Log Logistic** • Maxwell • Non-Central Chi-Squared** • Non-Central F** • Non-Central T • Pareto • Power • Rayleigh • Reciprocal • Skew Normal • Smallest Extreme Value • Student's T • Student's T with Location and Scale (3 Parameter) • Von Mises	Packet – Unweighted Sequential Custom Discrete)

Specifying Truncation Values

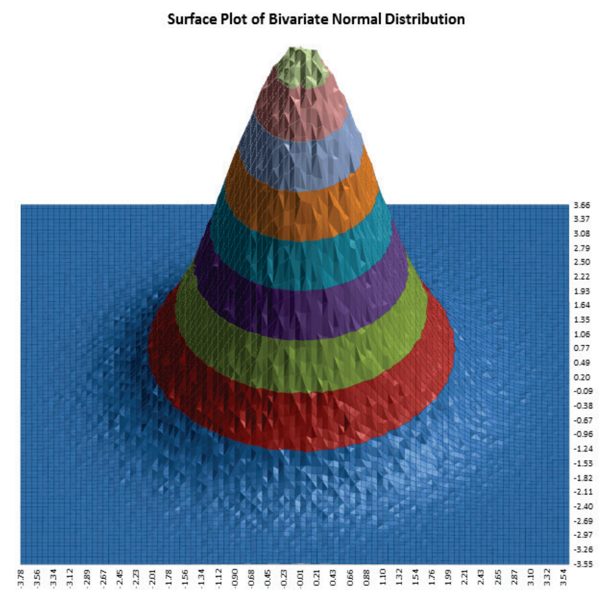
DiscoverSim allows you to specify truncation values for any input distribution. For example, a truncated normal distribution would be appropriate if you have a component supplier that starts with a normal distribution, then rejects the parts that fail the lower and upper specification, and ships you the good parts.
The following histogram illustrates a truncated normal distribution (Mean = 0, StdDev = 1) with Minimum = -2 and Maximum = 2.

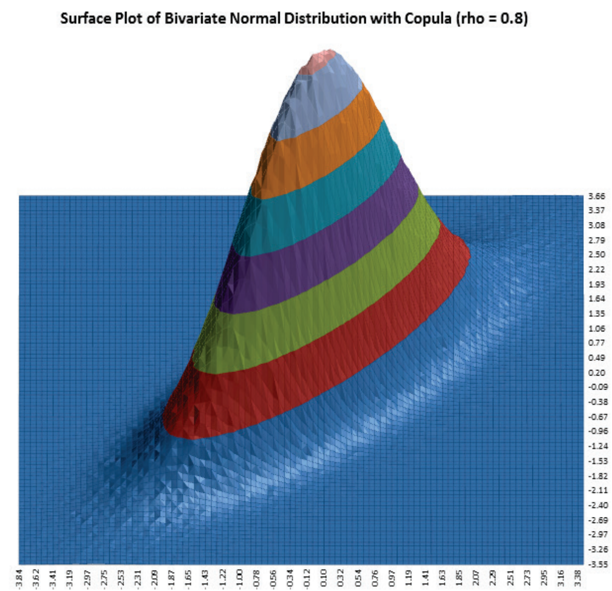


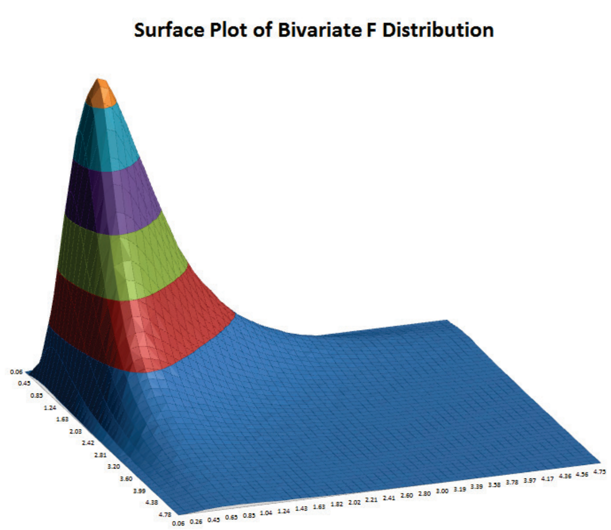
Specifying Correlations

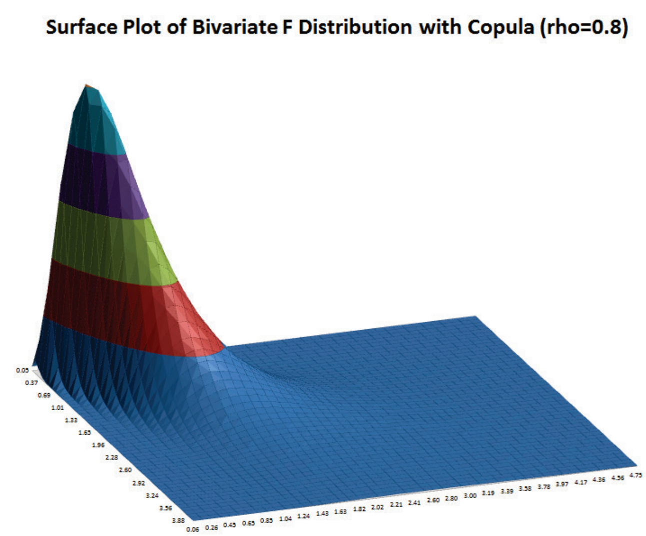
DiscoverSim allows you to specify correlations between any inputs. DiscoverSim utilizes correlation copulas to achieve the desired Spearman Rank correlation values.

The following surface plots illustrate how a correlation copula results in a change in the shape of a bivariate (2 input) distribution:







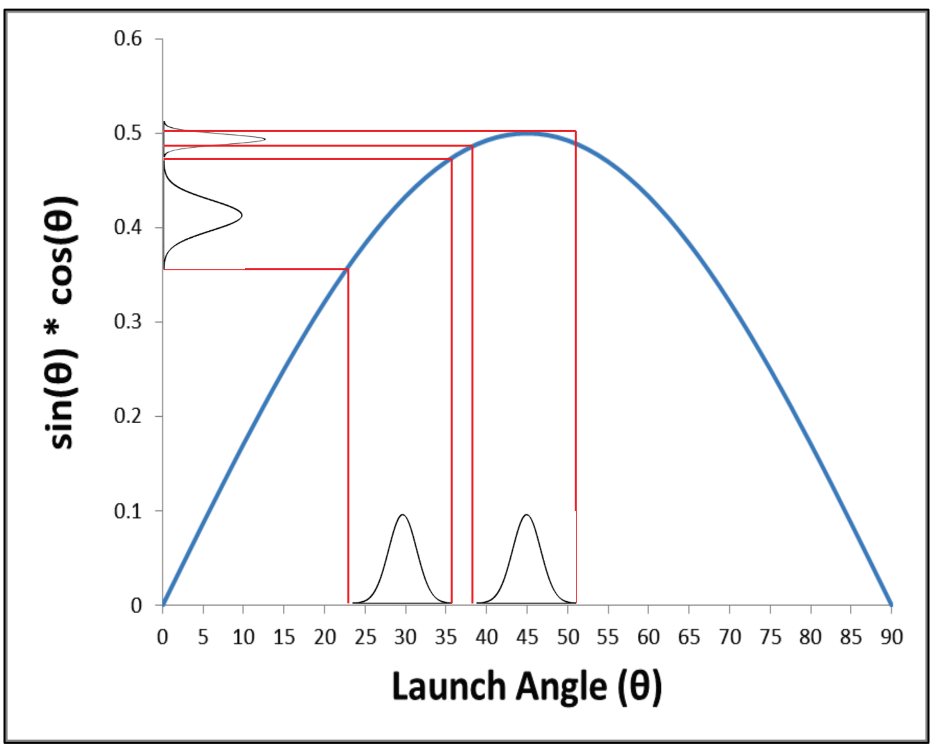


Optimization: Stochastic Versus Deterministic

Monte-Carlo simulation enables you to quantify risk, whereas stochastic optimization enables you to minimize risk.

Deterministic optimization is a commonly used tool (e.g. with Excel Solver) but has the drawback that it does not take uncertainty of the inputs into account. Stochastic optimization will not only find the optimum X settings that result in the best mean Y response value, it will also look for a solution that will reduce the standard deviation of Y.

Deterministic optimization will look for a minimum or maximum, whereas stochastic optimization looks for a minimum or maximum that is robust to variation in X, thus reducing the transmitted variation in Y. This is referred to as “Robust Parameter Design” in Design For Six Sigma (DFSS). In Case Study 4, Catapult Variation Reduction, the transmitted variation of Catapult distance is significantly reduced by shifting the launch angle mean from 30 degrees to 45 degrees, as illustrated in the diagram below:



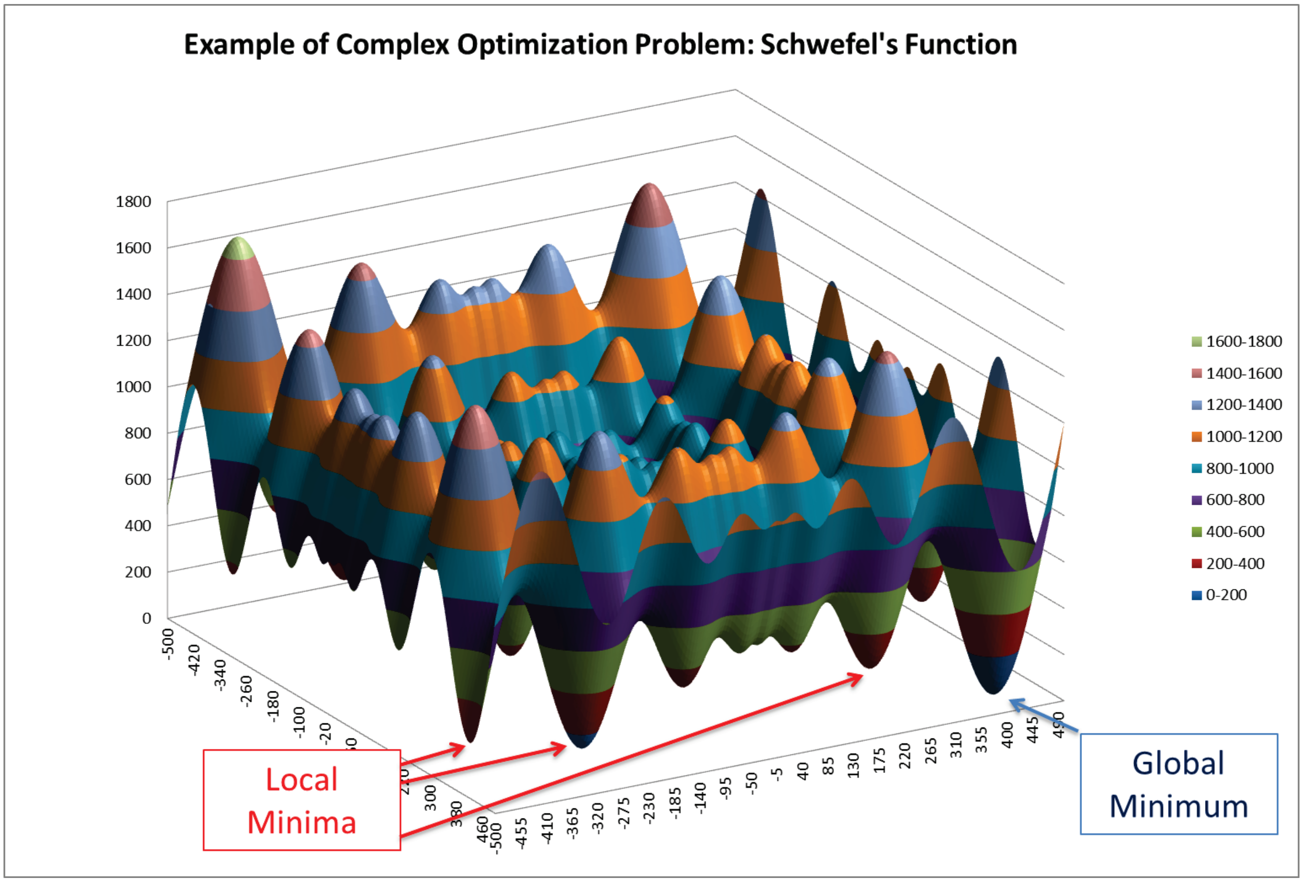
Optimization: Local Versus Global

Local optimization methods are good at finding local minima and have fast convergence. DiscoverSim uses Sequential Quadratic Programming (SQP) which takes derivatives of the objective function to find the path of greatest improvement for fast optimization. However, it requires a smooth response so will not work with discontinuous functions. Nelder-Mead simplex optimization is used for local non-smooth problems.

Global optimization finds the global minimum, and is derivative free, so will work with discontinuous functions. However, because of the larger design space, convergence is much slower than that of local optimization. DiscoverSim uses Mixed Integer Distributed Ant Colony Optimization (MIDACO) and Genetic Algorithm (GA) for global optimization.

A hybrid of the above methodologies is also available in DiscoverSim, and in the case of all continuous input variables, starts with MIDACO to do a thorough initial search, followed by GA, and then fine tuning with SQP or Nelder-Mead (if SQP fails).

The following surface plot illustrates a function with local minima and a global minimum.



Note: The Schwefel function is a standard *n* dimensional global optimization benchmark problem. Function details are widely available. The plot was created using Excel's 3D Surface Plot.

Components of Optimization

Output: The output will be a function of the model equation, intermediate cell calculations, as well as input distributions and input controls.

Input Control: The permissible range for the control is specified and, unlike an input distribution, has no statistical variation. Think of this as the temperature control knob of a thermostat. This is also known as a “Decision Variable.” An input control can be referenced by an input distribution parameter, constraint and/or an output function. It is possible to have a model that consists solely of controls with no input distributions. (In this case, the optimization is deterministic, so the number of replications, *n*, should be set to the minimum of 3 to save computation time.) An input control can be continuous or discrete.

Constraint: A constraint can be applied to an Input Control or DSIM Function but cannot directly reference a stochastic (statistical) Input Distribution or stochastic Output Response. Constraints can be simple linear or complex nonlinear. Each constraint will contain a function of Input Controls on the Left Hand Side (**LHS**), and a constant on the Right Hand Side (**RHS**). A constraint formula or cell reference is specified in the **LHS**.

Constraints are managed automatically within SQP and Discrete Exhaustive; for MIDACO, GA and NM, the Oracle Penalty Function is applied.

DSIM Function: Insert a function such as **DSim_Mean**, **DSim_Pcntle** (Percentile) or **Dsim_Ppk** (Capability Metric). The returned function value is a single number computed using all of the replications within a simulation run or optimization function evaluation. These computed values can then be referenced in a constraint **LHS**.

Summary of Optimization Features

DiscoverSim optimization features include:

- Mixed Integer Distributed Ant Colony Global Optimization (MIDACO)
- Genetic Algorithm (GA) Global Optimization
- Sequential Quadratic Programming (SQP) for fast local optimization
- Nelder-Mead (NM) simplex optimization for fast local non-smooth problems
- Exhaustive Discrete optimization – all combinations for small discrete problems
- Powerful hybrid of the above methods:
  - All discrete controls: MIDACO or Exhaustive Discrete if applicable
  - All continuous controls: MIDACO, GA, followed by SQP or NM
  - Mixed continuous/discrete controls: MIDACO 1 (Course), MIDACO 2 (Fine), followed by SQP or NM
- Specify linear and nonlinear constraints
- Stochastic Optimization – minimize dpm or maximize Ppk
- Insert DSIM functions such as Percentiles or Capability Metrics. These can then be used in constraints.
- Multiple Response Optimization using Desirability, Weighted Sum or Deviation from Target

Getting Started Tutorial

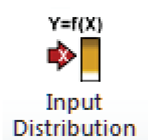
1. Open the workbook **DMAIC Project Duration**. This is an example model of a Six Sigma project with five phases: Define, Measure, Analyze, Improve and Control. We are interested in estimating the total project duration in days. Management requires that the project be completed in 120 days. We will model duration simply as:

   Total Duration = Define + Measure + Analyze + Improve + Control

	A	B	C	D	E	F	G	H
1		**DiscoverSim Tutorial: Six Sigma DMAIC Project Expected Duration**						
2						Triangular Distribution		
3		Project Phase:	DiscoverSim Input:			Minimum	Mode (Most Likely)	Maximum
4		Define				5	10	15
5		Measure				10	20	30
6		Analyze				5	10	15
7		Improve				10	20	30
8		Control				15	30	45
9								
10			DiscoverSim Output:					
11		Total Duration (days):	0			45	90	135
12								
13		Management Requirement (days):	120					

2. We will model each phase using a Triangular Distribution, which is popular when data or theory is not available to estimate or identify the distribution type and parameter values. (Another similar distribution that is commonly used for project management is PERT). The Minimum, Mode (Most Likely) and Maximum durations were estimated by the Six Sigma project team based on their experience.

3. Click on cell **C4** to specify the Input Distribution for Define. Select **DiscoverSim > Input Distribution:**



4. Select **Triangular Distribution**. A brief description of the Triangular Distribution is given in the dialog.

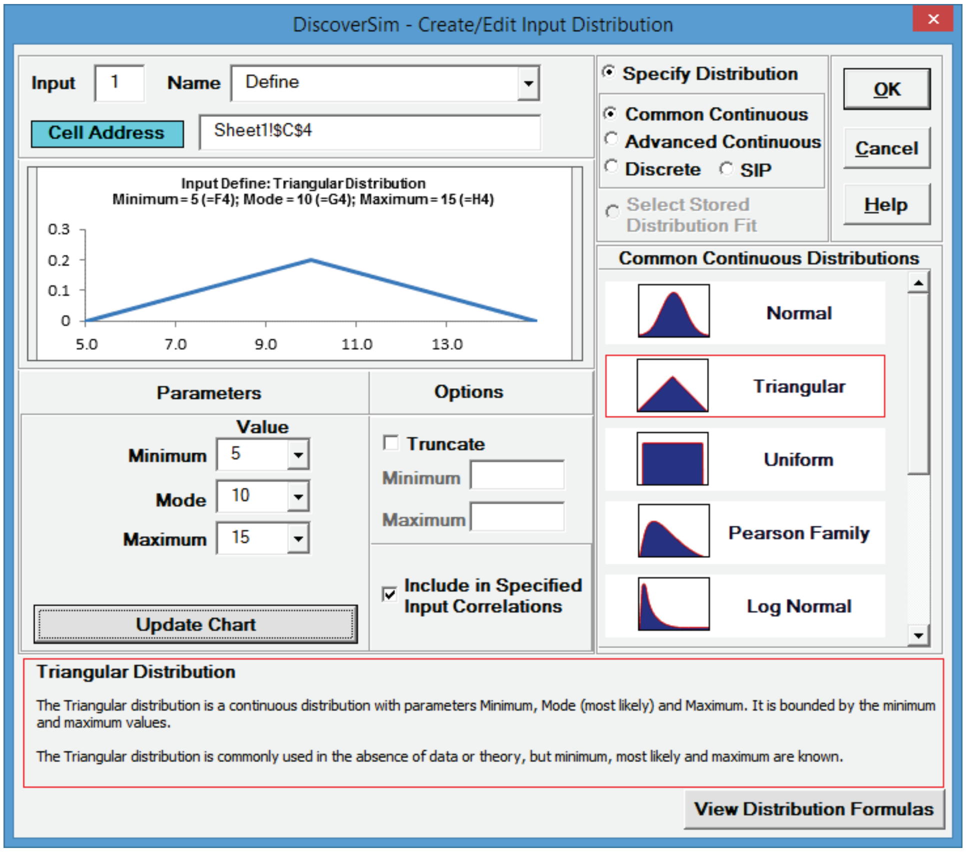
5. Click input **Name** cell reference and specify cell **B4** containing the input name "Define." After specifying a cell reference, the dropdown symbol changes from to .

6. Click the **Minimum** parameter cell reference and specify cell **F4** containing the minimum parameter value = 5.

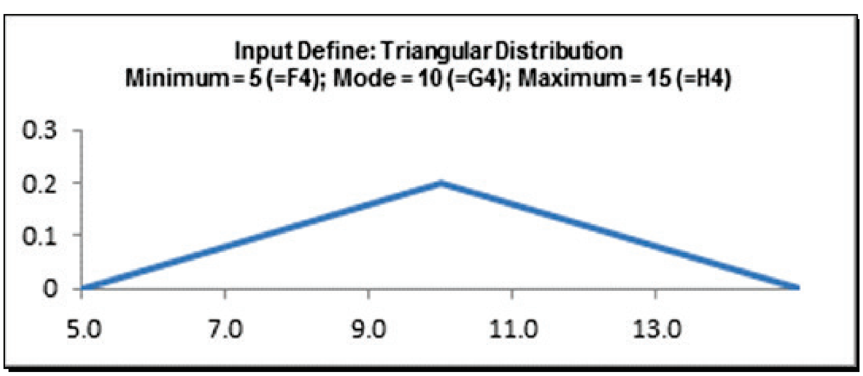
7. Click the **Mode** parameter cell reference and specify cell **G4** containing the mode (most likely) parameter value = 10.

8. Click the **Maximum** parameter cell reference and specify cell **H4** containing the maximum parameter value = 15.

9. Click **Update Chart** to view the triangular distribution as shown:



10. Click **OK**. Hover the cursor on cell **C4** to view the DiscoverSim graphical comment showing the distribution and parameter values:

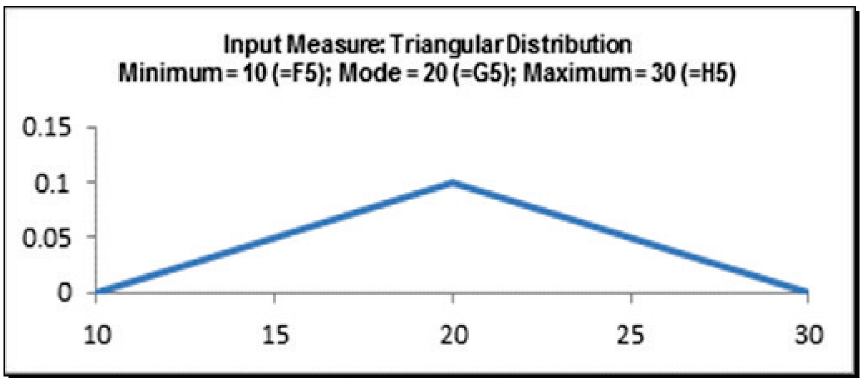


11. Click on cell **C4**. Click the **DiscoverSim Copy Cell** menu button Copy Cell (**Do not use Excel's Copy – it will not work!**).

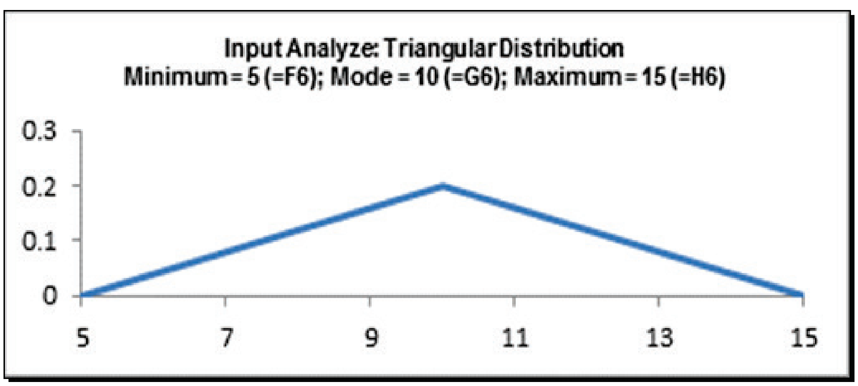
12. Select cells **C5:C8**. Click the **DiscoverSim Paste Cell** menu button Paste Cell (**Do not use Excel's Paste – it will not work!**).

13. Verify that the input comments appear as shown:

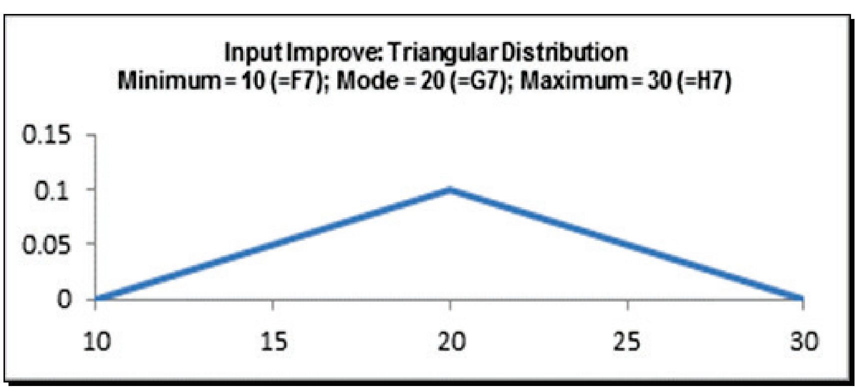
Cell **C5**:



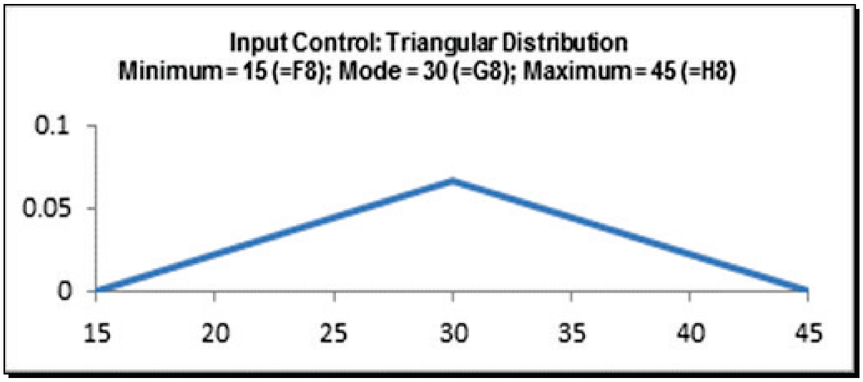
Cell **C6**:



Cell **C7**:

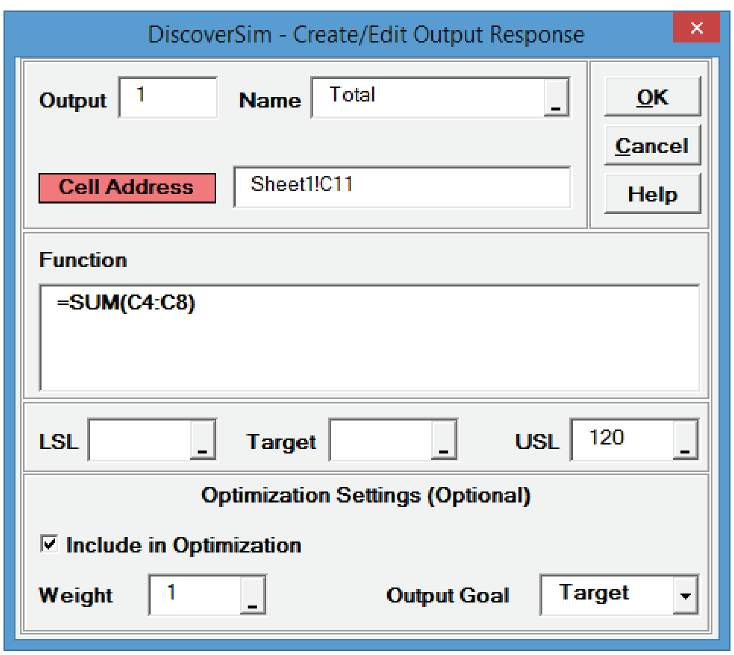


Cell **C8**:



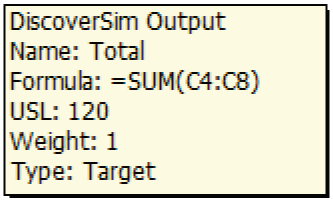
14. Click on cell **C11**. Note that the Excel formula =SUM (C4:C8) has already been entered. Select **DiscoverSim > Output Response**:

15. Enter the output **Name** as "Total." Enter the Upper Specification Limit (**USL**) as 120 (or specify cell reference **C13**). **Optimization Settings** are not used in this example since this is strictly a simulation.



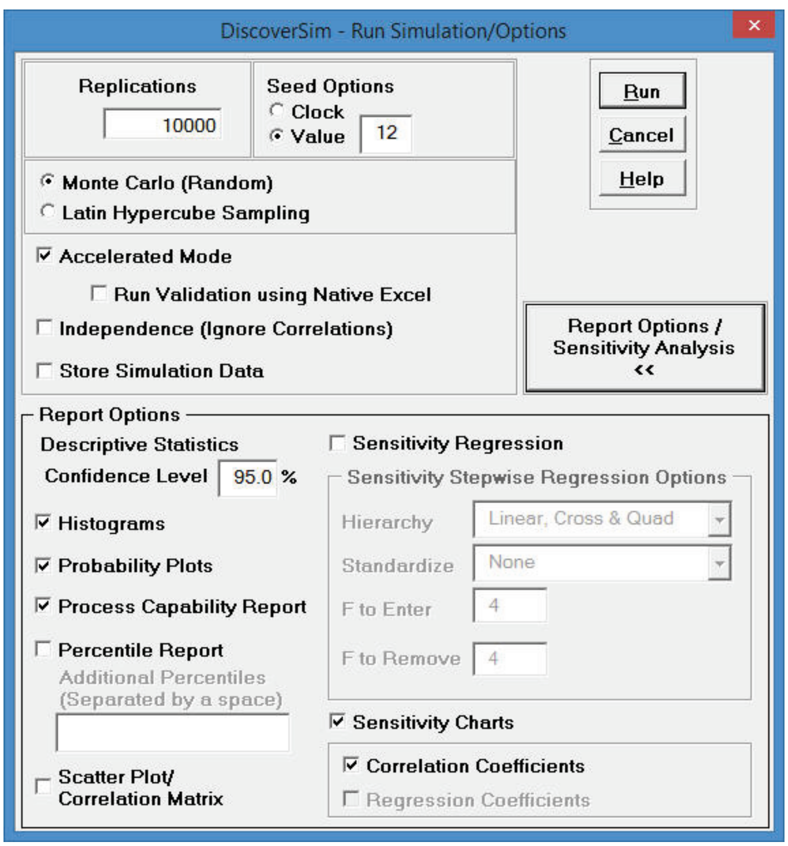
Click **OK**.

16. Hover the cursor on cell **C11** to view the DiscoverSim Output information.



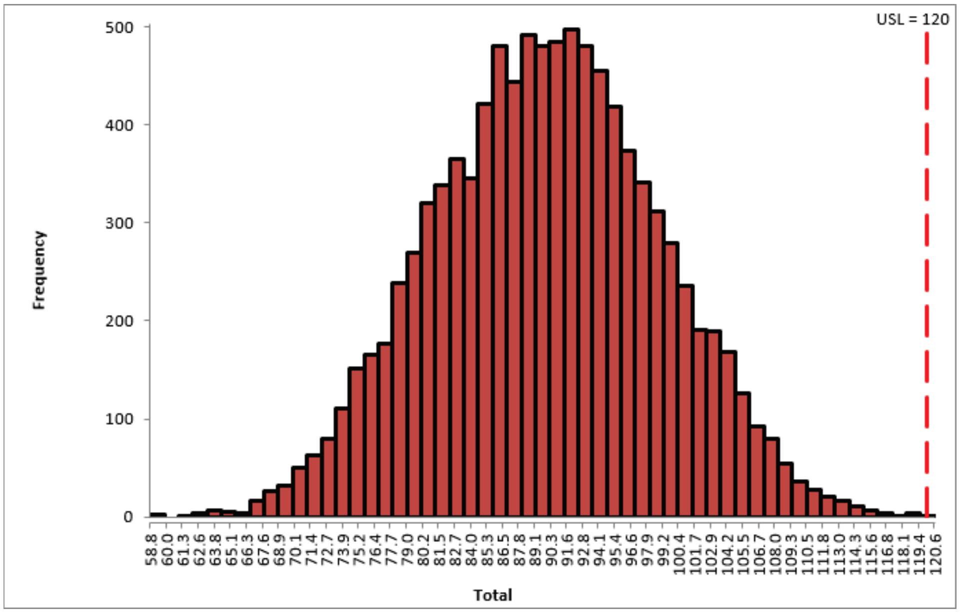
17. Select **DiscoverSim** > **Run Simulation**:

18. Click **Report Options/Sensitivity Analysis**. Check **Sensitivity Charts** and **Correlation Coefficients**. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below (note that 64 bit DiscoverSim will show slightly different results).



Click **Run**.

19. The DiscoverSim output report shows a histogram, descriptive statistics and process capability indices:



Output 1 Model Information	
Output Name (or Cell Address)	Total
Function	=SUM(C4:C8)
USL	120
Target	
LSL	
Weight	1
Output Goal	Target

Descriptive Statistics	
Count	10000
Mean	90.023
StDev	8.925891
Range	61.842
Minimum	58.770
25th Percentile (Q1)	83.784
50th Percentile (Median)	90.119
75th Percentile (Q3)	96.123
Maximum	120.612
95.0% CI Mean	89.84819 to 90.19812
95.0% CI Median	89.84714 to 90.30615
95.0% CI StDev	8.803885 to 9.05135

Normality Tests	
Anderson-Darling Normality Test	1.548445
p-value (A-D Test)	0.0006
Skewness	0.008590875
p-value (Skewness)	0.7256
Kurtosis	-0.238108
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	1.12
Ppl	
Ppk	1.12
Cpm	

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	391.95
ppm < LSL	
ppm Total	391.95
% > USL	0.04
% < LSL	
% Total (out of spec.)	0.04
% Total (within spec.)	99.96

Actual Performance (Empirical)	
ppm > USL	100.00
ppm < LSL	
ppm Total	100.00
% > USL	0.01
% < LSL	
% Total (out of spec.)	0.01
% Total (within spec.)	99.99

From the histogram and capability report we see that the Total Project duration should easily meet the requirement of 120 days, assuming that our model is valid. The likelihood of failure (based on the actual simulation performance) is approximately .01%.

Note: If **Seed** is set to **Clock**, there will be slight differences in the reported values with every simulation run due to a different starting seed value derived from the system clock.

20. A summary of the model input distributions, parameters and output response is also given in the simulation report.

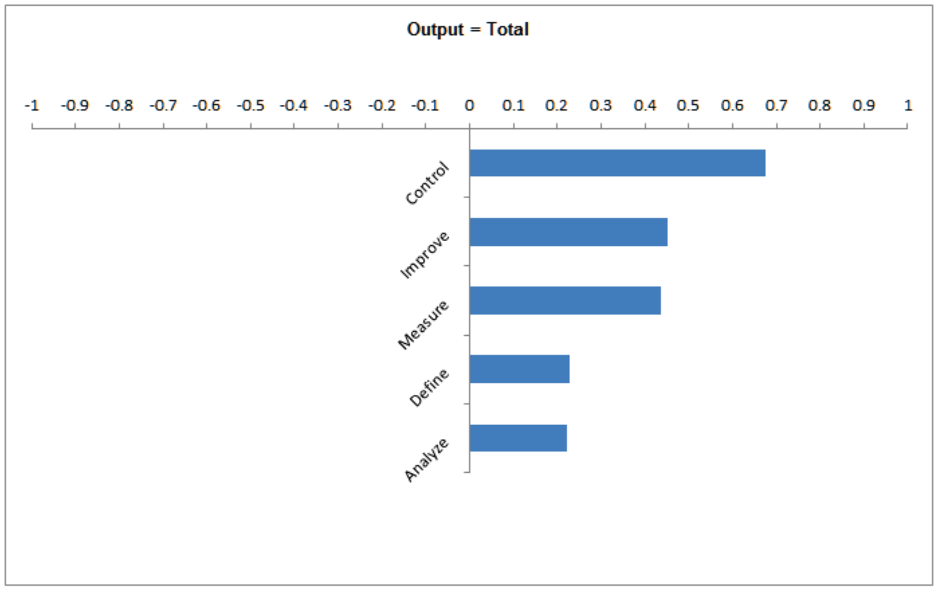
Input Name	Input Cell Address	Distribution	Distribution Type	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value	Parameter 3 Name	Parameter 3 Value
Define	C4	Triangular	Continuous	Minimum	5 (=Sheet1!$F$4)	Mode	10 (=Sheet1!$G$4)	Maximum	15 (=Sheet1!$H$4)
Measure	C5	Triangular	Continuous	Minimum	10 (=Sheet1!$F$5)	Mode	20 (=Sheet1!$G$5)	Maximum	30 (=Sheet1!$H$5)
Analyze	C6	Triangular	Continuous	Minimum	5 (=Sheet1!$F$6)	Mode	10 (=Sheet1!$G$6)	Maximum	15 (=Sheet1!$H$6)
Improve	C7	Triangular	Continuous	Minimum	10 (=Sheet1!$F$7)	Mode	20 (=Sheet1!$G$7)	Maximum	30 (=Sheet1!$H$7)
Control	C8	Triangular	Continuous	Minimum	15 (=Sheet1!$F$8)	Mode	30 (=Sheet1!$G$8)	Maximum	45 (=Sheet1!$H$8)

Output Name	Output Cell Address	LSL	Target	USL	Weight	Output Goal	Function
Total	C11			120	1	Target	=SUM(C4:C8)

Tip: Use this summary to keep track of simulation runs with modified parameter values.

If we needed to improve the performance, the sensitivity charts would guide us where to focus our efforts. Click on the **Sensitivity Correlations** sheet:

Inputs	Spearman Rank
Control	0.675972435
Improve	0.451123789
Measure	0.436461129
Define	0.22775111
Analyze	0.222844794



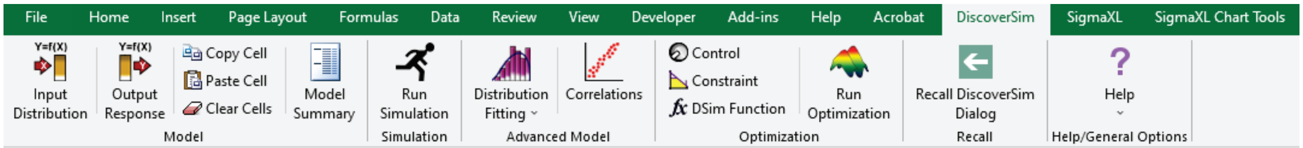
The control phase is the dominant input factor affecting Total Duration, followed by Measure and Improve.

Tip: This Sensitivity Chart uses the Spearman Rank Correlation, and the results may be positive or negative. If you wish to view R-Squared percent contribution to variation, rerun the simulation with **Sensitivity Regression Analysis** and **Sensitivity Charts**, **Regression Coefficients** checked.

The results in this simple example are intuitively obvious from the input values, but keep in mind that a real model will likely be considerably more complex with sensitivity analysis results identifying important factors that one might not have expected.

Overview of DiscoverSim™ Menu and Dialogs

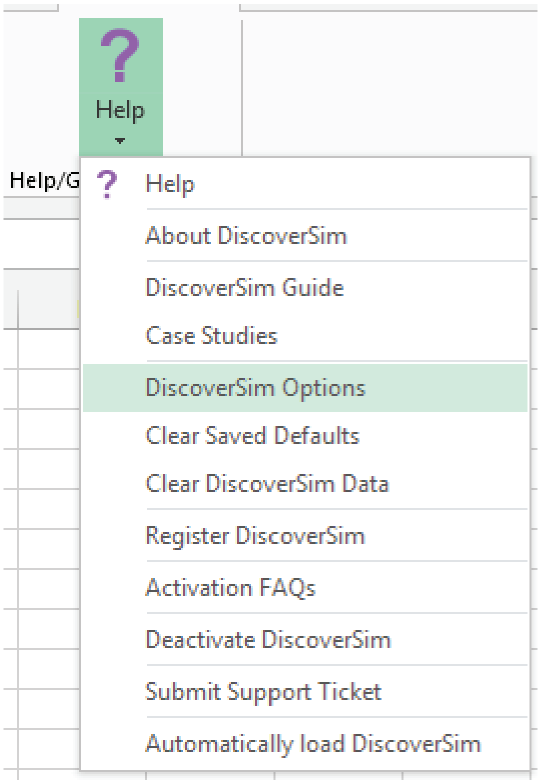
DiscoverSim Menu

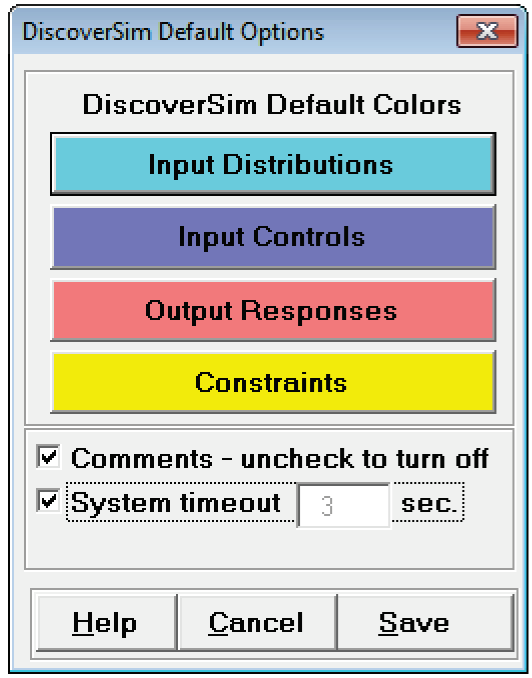


Recall DiscoverSim Dialog

Recall is used to activate the last data worksheet and recall the last dialog, making it very easy to do repetitive analysis.

DiscoverSim Options





1. Default colors for Input Distributions, Input Controls, Output Responses and Constraints can be modified by clicking on the Color button.

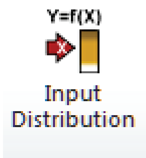
2. Check **Comments** to display DiscoverSim comments in Input/Output cells.

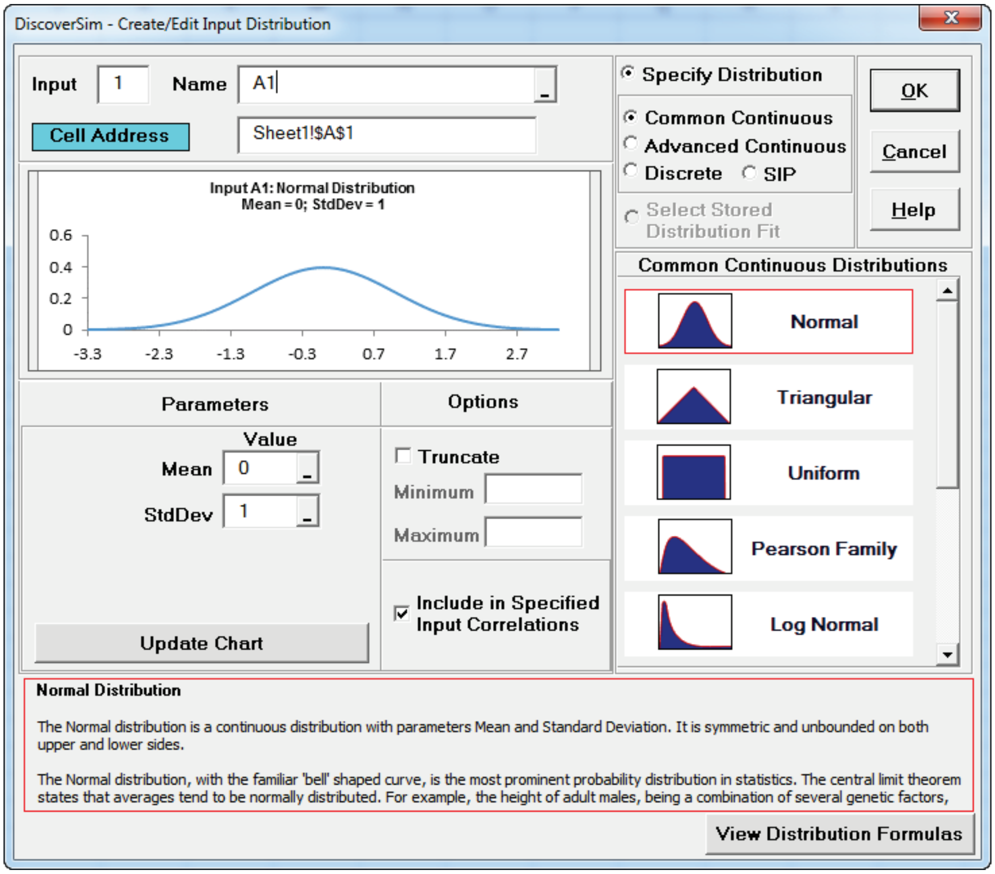
3. **System timeout** (seconds) is the maximum time for a messaging instance (communication between Excel and Windows) in optimization.

4. Click **Save** to save option default changes.

5. To restore the options to the original default values, click **Help > Clear Saved Defaults**.

Create/Edit Input Distribution





1. **Input** Number is provided for reference purposes only. (To switch to a different input cell, click cancel, select the input cell and click **Input Distribution**.)

2. Input **Name** is the cell address by default. Enter a name to describe the input or click the cell reference and specify the cell containing the input name. After specifying a cell reference, the dropdown symbol changes from to .

3. Input cell color can be modified by clicking on **Cell Address** (default colors can be changed in **Help > DiscoverSim Options**).

4. Select from **Common Continuous**, **Advanced Continuous**, **Discrete** or **SIP** Distributions. A brief description of the selected distribution is given in the dialog. Click **View Distribution Formulas** for distribution formula details: Probability Density Function (PDF) and Cumulative Distribution Function (CDF). These distribution formulas are also given in the Appendix.

5. If distribution fitting has been performed, the **Select Stored Distribution Fit** option will become available for selection of variable name(s) and stored distribution(s). The parameter values will automatically be populated.

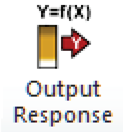
6. Parameter values can be manually entered or click the cell reference and specify the cell containing the parameter value.

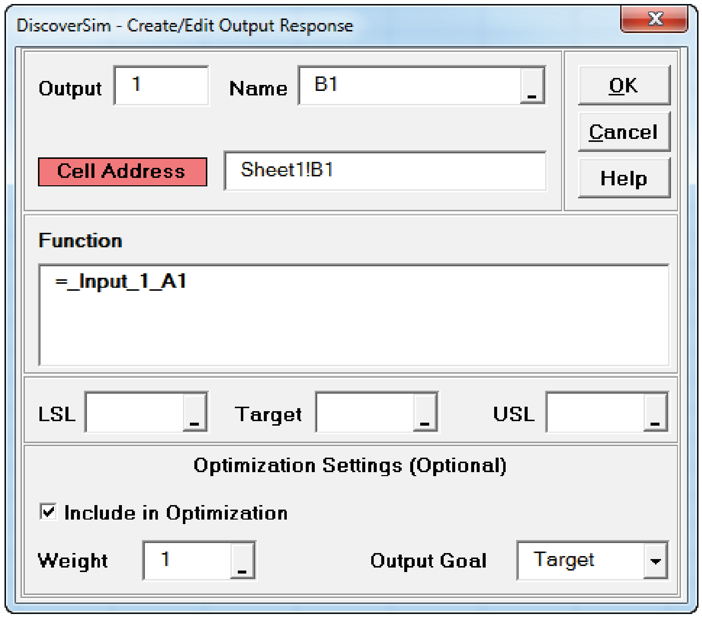
7. Click **Update Chart** after changing parameter values to view distribution.

8. Check **Truncate** and enter truncation minimum and/or maximum values to produce a truncated distribution. Blank entries will be treated as negative infinity for minimum and positive infinity for maximum.

9. Check **Include in Specified Input Correlations** (checked by default) to allow correlations to be specified for this input.

Create/Edit Output Response





1. **Output** Number is provided for reference purposes. (To switch to a different output cell, click cancel, select the output cell and click **Output Response**.)

2. Output **Name** is the cell address by default. Enter a name to describe the output or click the cell reference and specify the cell containing the output name. After specifying a cell reference, the dropdown symbol changes from to .

3. Output cell color can be modified by clicking on **Cell Address**. (Default colors can be changed in **Help > DiscoverSim Options**).

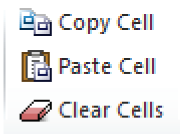
4. **Function** displays the output cell equation. The equation can also be changed in this field.

5. **LSL** is the Lower Specification Limit; **USL** is the Upper Specification Limit for the output. Entry is optional for simulation but required to produce Process Capability Indices. LSL and USL are also used as the lower and upper bounds for the desirability function in multiple response optimization. Specification limits can be manually entered or click the cell reference and specify the cell containing the specification value.

6. **Include in Optimization**, **Weight** and **Output Goal** are not used for simulation - they are optional output settings used for multiple response optimization. If **Include in Optimization** is unchecked this output will not be considered for optimization. Note that the **Output Goal** (Target, Maximize or Minimize) is specific to each output. For example, if Output 1 is production rate, the goal would be set to maximize, whereas Output 2, cost, would have a minimize goal. For further details, see Run Optimization.

Copy Cell, Paste Cell, Clear Cells

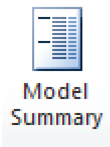
DiscoverSim **Copy Cell** and **Paste Cell** are used to quickly create duplicate input distributions, input controls, or output responses. **Clear Cells** deletes the input or output from the DiscoverSim model. A cell address reference is relative (**A1**) by default, but can be changed to absolute (**$A$1**) by using the **F4** key.

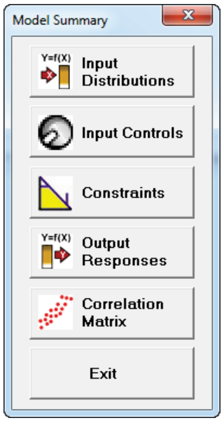


Note that Excel's copy, paste and delete commands cannot be used since DiscoverSim needs to update hidden model worksheets for any additions or deletions.

Model Summary

Click **Model Summary** and select a summary of Input Distributions, Input Controls, Constraints, Output Responses, or Correlation Matrix.





Tip: Summaries of the Input Distributions, Input Controls and Output Responses are also included in the Simulation Report.

The **Input Distributions** summary table displays all of the Input Distributions with the selected distributions, parameter values and settings. The **Input Controls** summary table displays all of the Input Controls, type, and optimization boundaries. The **Constraints** summary table displays all of the constraints. The **Output Responses** summary table displays all of the Output Responses, specification limits and optimization options. The **Correlation Matrix** summary table displays the entire specified Correlation Matrix.

Clicking a button scrolls the selected summary table to the top and is useful to view a summary for a large model. Click **Exit** to remove the **Model Summary** dialog.

Summary of DiscoverSim Input Distributions

Input Name	Input Cell Address	Distribution	Distribution Type	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value
A1	A1	Normal	Continuous	Mean	0	StdDev	1
A2	A2	Normal	Continuous	Mean	0	StdDev	1
A3	A3	Normal	Continuous	Mean	0	StdDev	1

Summary of DiscoverSim Input Controls

Control Name	Control Cell Address	Control Type	Start Value	Lower Bound	Upper Bound
C1	C1	Continuous	0	0	1
C2	C2	Continuous	0	0	1
C3	C3	Continuous	0	0	1

Summary of DiscoverSim Constraints

Constraint Name	Constraint Cell Address	LHS	Comparison	RHS
C5	C5	C1	<	0.5

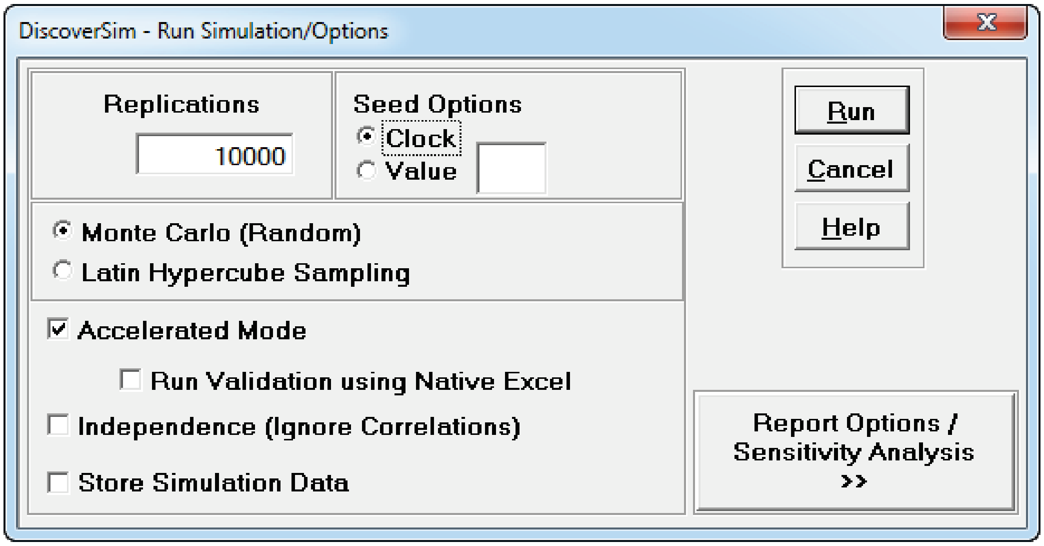
Summary of DiscoverSim Output Responses

Output Name	Output Cell Address	LSL	Target	USL	Weight	Output Goal	Function
E1	E1	-3	0	3	1	Target	=SUM(A1:A3)+SUM(C1:C3)

Summary of DiscoverSim Specified Input Correlations

	A1	A2	A3
A1	1		
A2	0	1	
A3	0	0	1

Run Simulation



1. **Replications** value sets the number of simulation replications.

2. **Seed** is set to **Clock** by default so that the starting seed of random number generation will be different with each run. If you want the simulation results to match every time (for example in a classroom setting where you want all students to obtain the same results), select **Value** and enter an integer number. Note that the results of a fixed seed for 32-bit Excel will be slightly different for 64-bit Excel.

3. Select **Monte Carlo (Random)** for full randomization. **Latin Hypercube Sampling** is less random than Monte Carlo but enables more accurate simulations with fewer replications. For further details see Appendix: Input Distribution Sampling.

4. For details on Random Number Generation (RNG) in DiscoverSim, see Appendix: Input Distribution Random Number Generation.

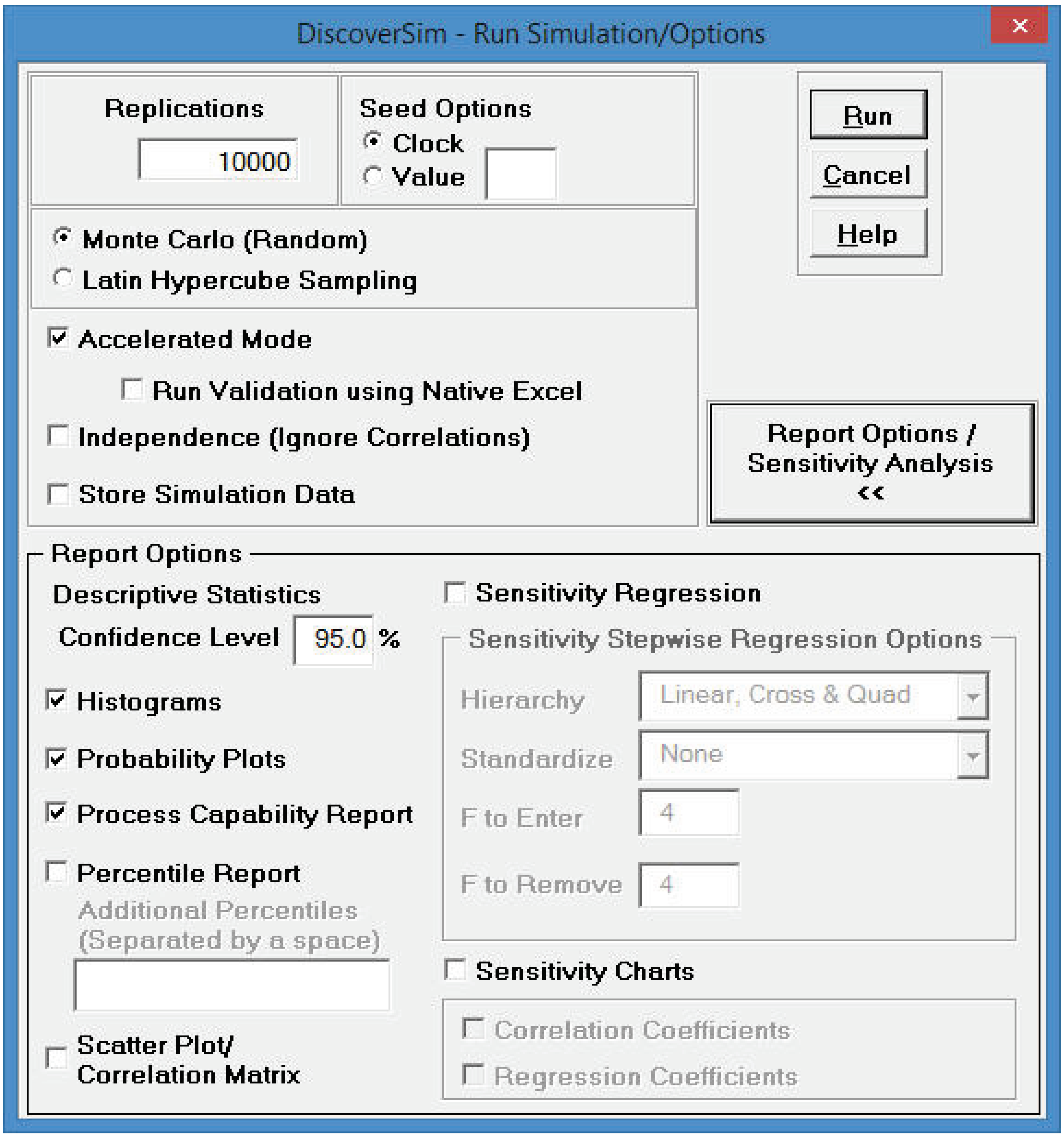
5. **Accelerated Mode** uses DiscoverSim's Excel Formula Interpreter to dramatically increase the speed of calculations for rapid simulation. If unchecked, the calculations are performed using native Excel. The interpreter supports the majority of all Excel numeric functions (for more details see Appendix: DiscoverSim Engine and Excel Formula Interpreter). If the DiscoverSim interpreter sees a function that it does not support, you will be prompted to use Excel's Native mode.

6. **Run Validation using Native Excel** runs a validation test to compare Accelerated Mode versus Native Excel. Each output is assessed by comparing the simulation means. If the worst case relative difference is less than or equal to 1e-10%, the test passes and the status is "Success." If the relative difference is between 1e-10% to 1e-4%, the status is "Good," and if greater than or equal to 1e-4%, the test status is "Poor."

7. Check **Independence (Ignore Correlations)** to run the simulation with all inputs independent of each other (zero correlation). This is recommended if you are running a **Sensitivity Regression Analysis**.

8. Check **Store Simulation Data in Worksheet** to store simulation input and output data in a worksheet. The number of simulation replicates or runs must not exceed the maximum number of rows permitted in the version of Excel that you are using and the total number of inputs and outputs cannot exceed the maximum number of columns permitted by Excel.

9. Click **Report Options/Sensitivity Analysis** to display all report options.



10. **Histograms**, **Probability Plots**, and **Process Capability Report** are checked by default for display in the simulation report. These graphs and tables apply only to the output response data. A report will be produced for each output.

    - **Descriptive Statistics** are always reported. They include confidence intervals for Mean, StdDev, and Median with a default 95% level. The **Confidence Level** can be modified. Descriptive statistics also include the normality tests: Anderson-Darling, Skewness and Kurtosis.
    - **Histograms** will display a normal curve and specification limits (if applicable).
    - **Probability Plots** are normal probability plots, but only display a small sample (500) of the data to speed up the report display.
    - **Process Capability Report** indices require at least one specification limit.

11. Check **Percentile Report** to view a table with the following percentiles:
    0.135, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99, 99.865.
    Additional Percentiles can be added to the report by entering the desired values, separated by a space.
    A **Percentile Process Capability** report is also produced. These indices should be used if the output response is not normally distributed:

    Percentile Pp = (USL – LSL)/(99.865 percentile – 0.135 percentile)
    Percentile Ppu = (USL – 50th percentile)/(99.865 percentile – 50th percentile)
    Percentile Ppl = (50th percentile – LSL)/(50th percentile – 0.135 percentile)
    Percentile Ppk = min(Ppu, Ppl)

    Tip: Percentile Process Capability indices require a minimum of 10,000 replications in order to be reliable.

12. Check **Scatter Plot/Correlation Matrix** to view a Scatter Plot of outputs versus outputs, outputs versus inputs, and inputs versus inputs. The correlation matrix includes both Pearson and Spearman Rank correlation coefficients. Spearman Rank is recommended for non-normal data.

13. Check **Sensitivity/Regression Analysis** to perform a backwards stepwise regression analysis. This will automatically be run on all outputs.
    Hierarchy options are:

    - Linear (Main Effects)
    - Linear & Quad (Main Effects and Quadratic Terms)
    - Linear & Cross (Main Effects and Cross-Product 2-Way Interactions)
    - Linear, Cross & Quad (Main Effects, Cross-Product 2-Way Interactions and Quadratic Terms)

If selected, Cross-Product and Quadratic terms are calculated and included in the initial stepwise regression model along with the Main Effects. Following **Hierarchy** means that the second order terms are removed from the model before the first order terms. If a Cross-Product or Quadratic term remains in the model, the associated Main Effects will also remain in the model, even if the Main Effect term is not significant.
Standardize options are:

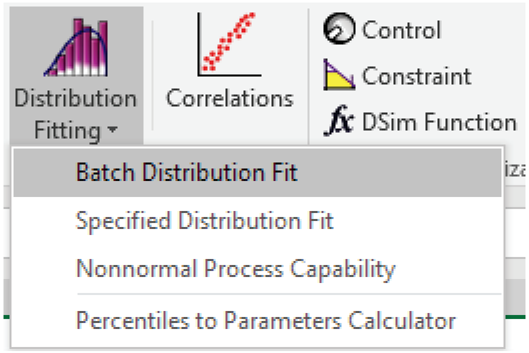
- None
- Std: Mean=0, StdDev=1 ((Xi – Mean)/StdDev)
- Range -1 to +1 (Coded Xmin = -1; Xmax = +1)

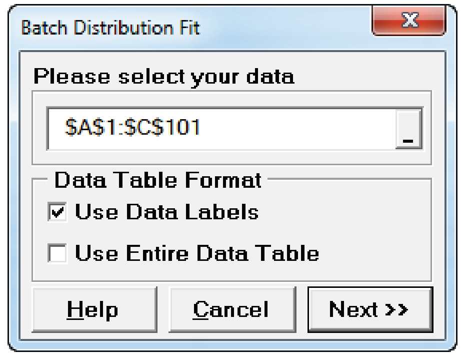
Use one of the standardize options if you have a second order model and wish to avoid additional collinearity that is introduced due to the multiplication of different units. The disadvantage of this is that the model coefficients are then standardized or coded so will not be as easy to interpret.
F to Enter allows you to set the value of F to enter a new variable in the model (default is 4).
F to Remove allows you to set the value of F to remove a variable from the model (default is 4).
Note: As mentioned above, it is highly recommended that you check **Independence (Ignore Correlations)** when performing a **Sensitivity/Regression Analysis** to avoid problems with multi-collinearity.

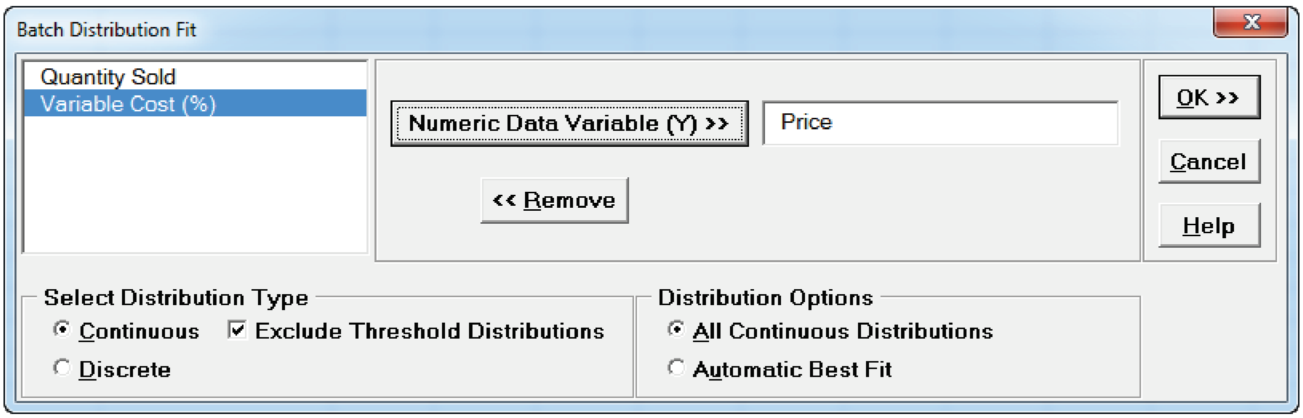
14. Check **Sensitivity Charts** to view vertical bar chart(s) of:

    - **Correlation Coefficients** – Sorted Spearman Rank correlation for output versus each input. The sort is by magnitude, with the chart x-axis range being -1 to +1.
    - **Regression Coefficients** – Sorted R-Square values for output versus each input, Cross-Product and Quadratic terms that remain in the Stepwise Regression model. R-Square for each final model term is obtained by sequentially removing and replacing each term from the model, with the decrease in model R-Square used as the term value.

Distribution Fitting: Batch Distribution Fit







1. Select **Continuous** or **Discrete**. Continuous distributions will be sorted in ascending order using the Anderson Darling goodness of fit statistic. Discrete distributions will be sorted with the Chi-Square goodness of fit statistic. P-Values and distribution parameter values are also reported for each distribution.

   - See Appendix: Distribution Fitting - Maximum Likelihood Estimation of Parameters for further details on parameter estimation.
   - See Appendix: Distribution Fitting – Goodness of Fit for further details on goodness of fit estimation.

2. Check **Exclude Threshold Distributions** to run the batch distribution fit without estimation of distributions that have a Threshold parameter. The default is checked because it is much faster. Note, however, that if the data contains negative values then Threshold distributions will be included in the batch fit.

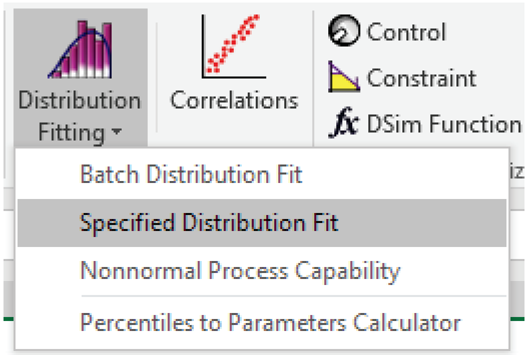
   Tip: If Threshold distributions are included, and two candidate distributions with similar Anderson Darling statistic values are being considered, one with a Threshold and one without, choose the simpler model without the Threshold. For further details on Threshold distributions and estimation, see Appendix: Distribution Fitting – Threshold Distributions.

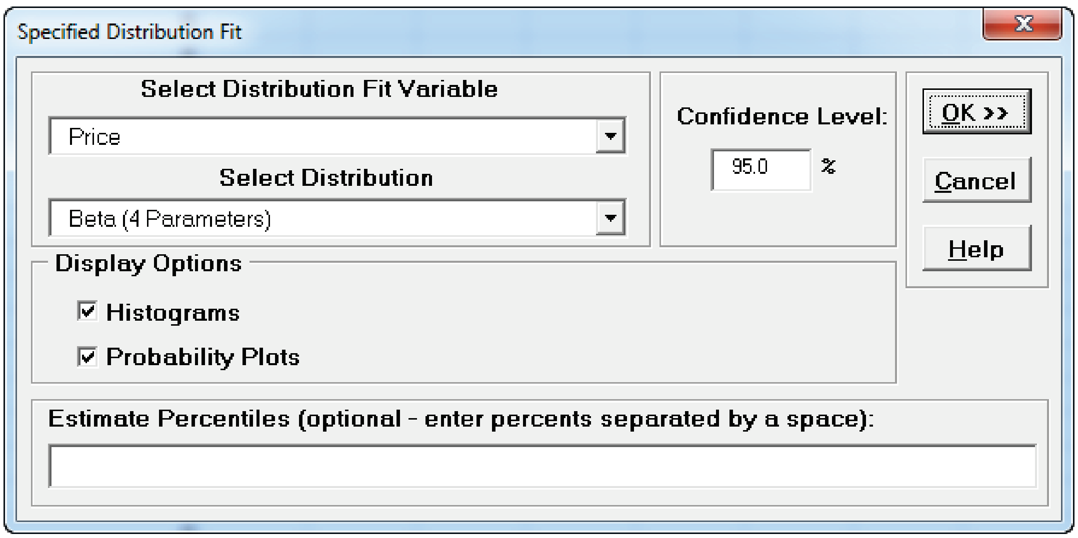
3. Select **All Distributions** or **Automatic Best Fit**. Automatic Best Fit will return the distribution with the lowest Anderson-Darling or Chi-Square statistic.

4. If distribution fitting has been performed, the **Select Stored Distribution Fit** option will become available in the **Input Distribution** dialog for selection of variable name(s) and stored distribution(s). The input distribution parameter values will automatically be populated.

5. See Case Study 1 – Basic Profit Simulation for examples of **Batch Distribution Fit**.

Distribution Fitting: Specified Distribution Fit





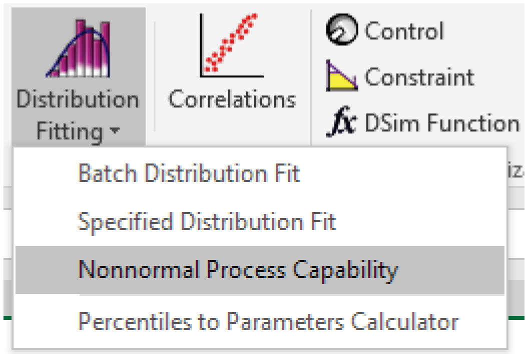
1. **Specified Distribution Fit** provides a more detailed distribution fit report for a selected variable and specified distribution. **Select Distribution Fit Variable** and **Select Distribution** provide a drop down list with available variable/distribution options from the Batch Distribution Fit (similar to **Select Stored Distribution Fit** in the **Create/Edit Input Distribution** dialog).

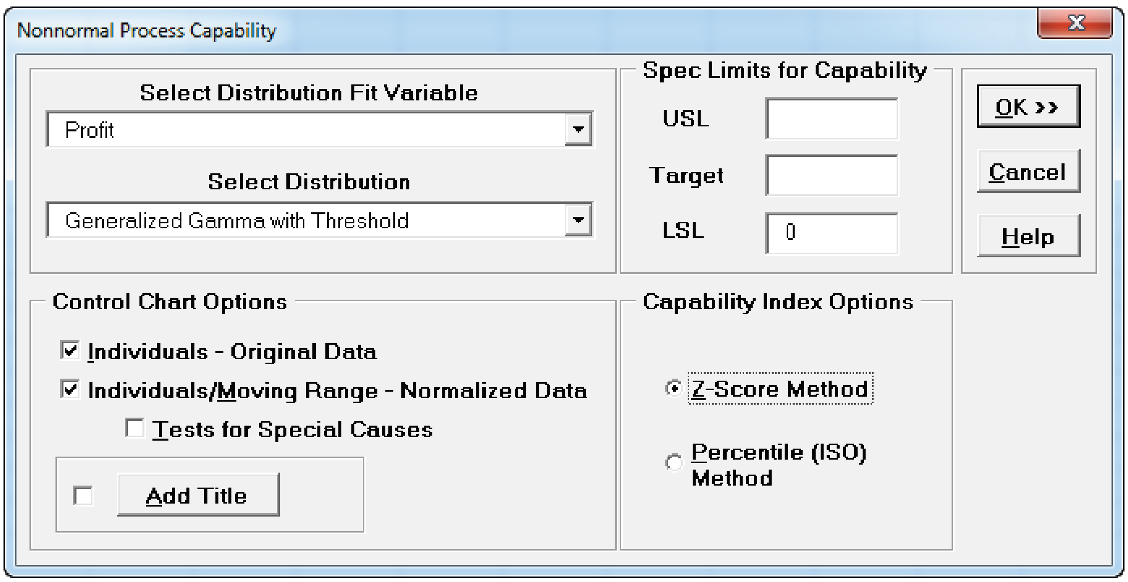
2. The **Specified Distribution Fit** report includes:
   - **Histogram** and **Probability Plot** (optional with check box).
   - Parameter estimates, standard errors (SE Estimate) and confidence intervals for parameter estimates.
     - The **Confidence Level** can be set in the dialog above with default value = 95%.
     - In some cases standard errors and confidence intervals may not be available.
   - Percentile Report with the following percentiles: 0.135, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99, 99.865.
     - Other values can be added by entering in the **Estimate Percentiles** box.

    - Note that the percentiles are computed using the distribution and estimated parameters, not the empirical data as is the case with the percentiles in the simulation report.
- Standard errors and confidence intervals for the percentile values, at the **Confidence Level** specified above (95% default).
    - In some cases standard errors and confidence intervals may not be available.
- Model Summary and Goodness-of-Fit statistics.

3. See Case Study 1 – Basic Profit Simulation for examples of **Specified Distribution Fit**.

Distribution Fitting: Nonnormal Process Capability





1. **Nonnormal Process Capability** provides a nonnormal process capability report with capability indices computed from the specified distribution with estimated parameters.
   - Empirical Performance is also reported for comparison.
   - Capability indices are for individual observations (subgrouping is not available).

2. **Select Distribution Fit Variable** and **Select Distribution** provide a drop down list with available variable/distribution options from the Batch Distribution Fit (similar to **Select Stored Distribution Fit** in the **Create/Edit Input Distribution** dialog).

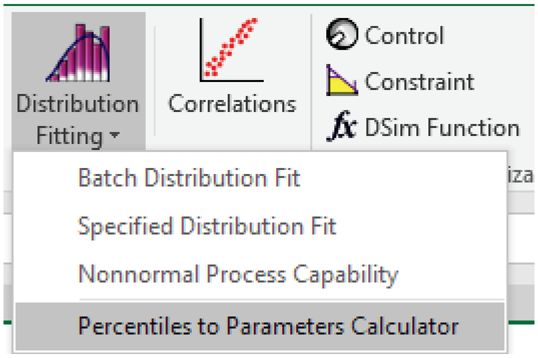
3. Enter Upper Specification Limit (**USL**) and/or Lower Specification Limit (**LSL**). Target is optional (used in the Cpm calculation).

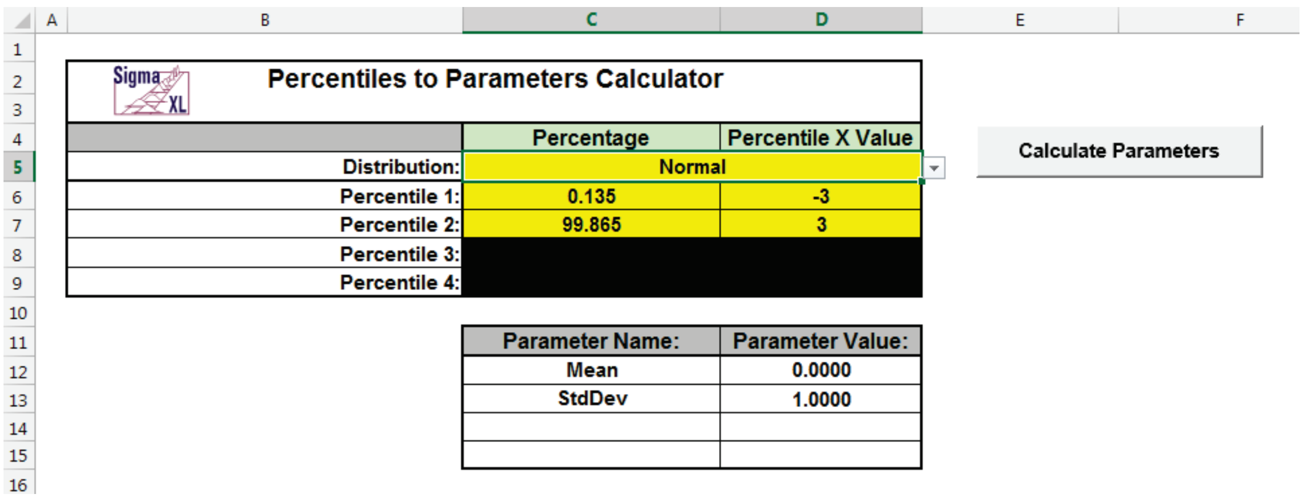
4. For details on Z-Score versus ISO methods and Control Charts see Appendix: Distribution Fitting – Nonnormal Process Capability Indices and Distribution Fitting – Nonnormal Control Charts.

5. Standard formulas for Process Capability Indices are given in the Appendix: Formulas for Quality and Process Capability Indices, but note that indices here use the specified distribution, not the normal distribution.

6. See Case Study 1 – Basic Profit Simulation for an example of **Nonnormal Process Capability**.

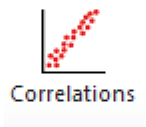
Distribution Fitting: Percentiles to Parameters Calculator

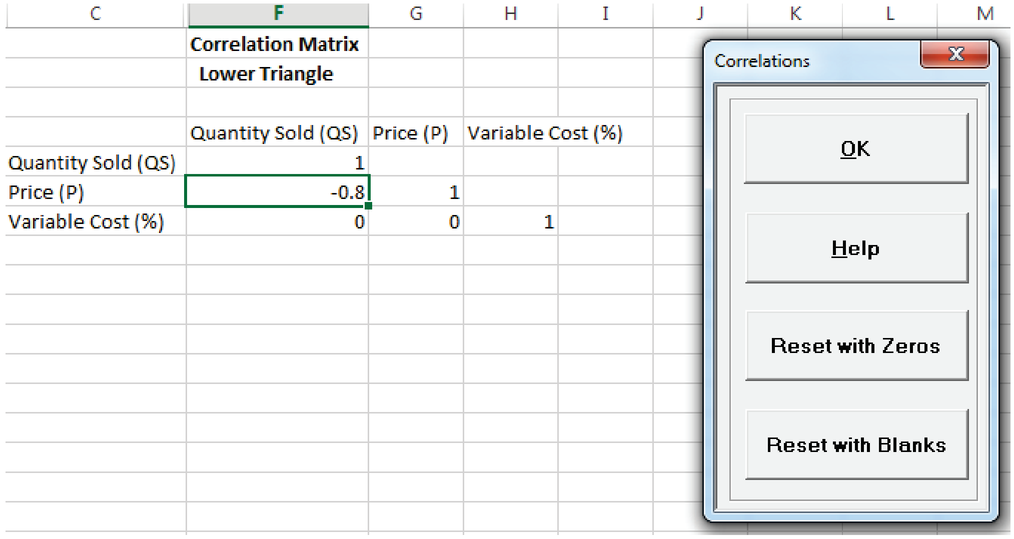




1. If one does not have sufficient data to do distribution fitting but does know some percentile values of a specified distribution, these can be converted to parameter values for use in simulation.
2. Select Distribution using drop down. Select Percentage values using drop down. Enter Percentile X Values.
3. Percentiles not used by a distribution are blacked out.
4. Click **Calculate Parameters**. Parameter Name and Values are shown in cells C12 to D15. **If the Parameter Value cells are blank, please ensure that the Percentile X Values are viable for the chosen distribution.**
5. Due to the complexity of calculations, this template uses VBA macros rather than Excel formulas. DiscoverSim must be initialized and appear on the menu in order for this template to function.
6. For further details on calculation methods, see Appendix: Distribution Fitting – Percentiles to Parameters Calculator.

Correlations





1. Specify **Correlations** for Input Distributions. Enter the desired Spearman Rank Correlation (-1 to 1) value in the lower triangle matrix.
   - Ensure that you press enter after entering a correlation value in a cell, otherwise the correlation will remain unchanged.
   - It is not necessary to enter a correlation value in the upper triangle.
   - The lower triangle specifies the correlations so any value entered in the upper will be ignored.
   - The diagonal of 1’s should not be altered.

2. Only those input distributions with “Include in Correlations” checked will be displayed here. Input Controls are not included in the correlation matrix.

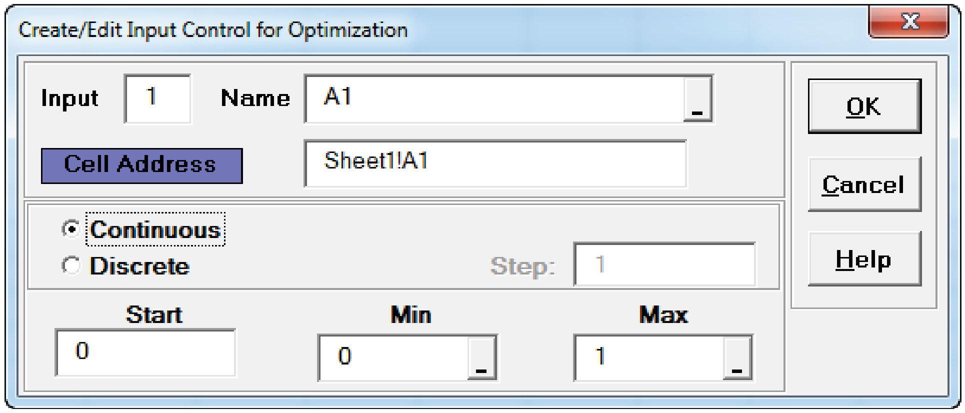
3. **Reset with Zeros** clears any specified correlations in the lower triangle and replaces them with zeros.

4. **Reset with Blanks** clears any correlations in the lower triangle and replaces them with blanks. This is useful if you wish to then specify correlations between inputs without the constraint of requiring independence on the other inputs.

5. Click **OK** to close the **Correlation** dialog and revert back to the model sheet.

6. See Case Study 1 – Basic Profit Simulation for an example of specifying Correlations.

Optimization: Control



1. The **Input Control**, unlike an input distribution, has no statistical variation. Think of this as the temperature control knob of a thermostat. This is also known as a "Decision Variable."

2. An **Input Control** can be referenced by an input distribution parameter, constraint and/or an output function.

3. It is possible to have a model that consists solely of controls with no input distributions. (In this case, the optimization is deterministic, so the number of optimization replications should be set to the minimum of 3 to save computation time.)

4. **Input Control** Number is provided for reference purposes only.

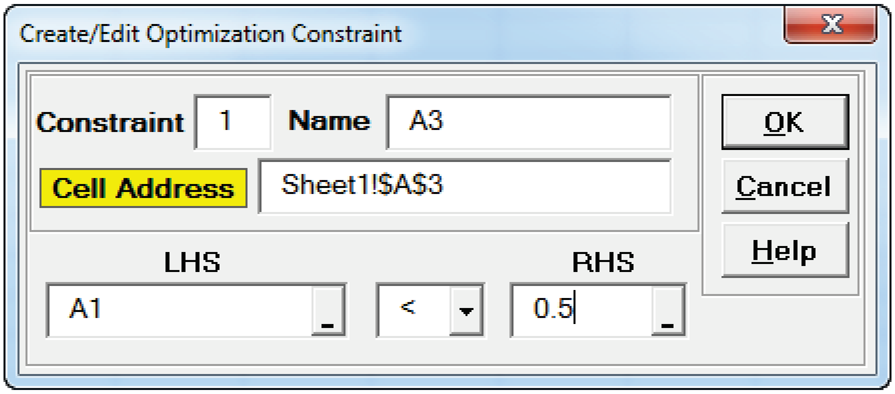
5. Input Control **Name** is the cell address by default. Enter a name to describe the input control or click the cell reference and specify the cell containing the input control name. Note that DiscoverSim copy/paste uses relative addressing (see Copy Cell, Paste Cell, Clear Cells).

6. Input Control cell color can be modified by clicking on **Cell Address** (default colors can be changed in **Help > DiscoverSim Options**).

7. Specify the **Start Value**, Minimum (**Min**) and Maximum (**Max**) boundaries for the Input Control. You can also click the cell reference and specify a cell containing the min or max value. After specifying a cell reference, the dropdown symbol changes from to .

8. An input control can be **Continuous** or **Discrete** with **Step** (default = 1, i.e. integer increment).

Optimization: Constraint



1. A constraint can be applied to an Input Control or DSIM Function but cannot directly reference a stochastic (statistical) Input Distribution or stochastic Output Response.

2. Each constraint will contain a function of Input Controls on the Left Hand Side (**LHS**), and a constant on the Right Hand Side (**RHS**). The **LHS** can be simple linear or complex nonlinear.

3. A constraint formula or cell reference is specified in the **LHS**. You can manually enter a simple formula in the **LHS** box, or click the cell reference and specify a cell containing the constraint formula. After specifying a cell reference, the dropdown symbol changes from to

4. You can manually enter a constant in the **RHS** box, or click the cell reference and specify a cell containing the constant. Note that the **RHS** value must be a fixed constant. After specifying a cell reference, the dropdown symbol changes from to

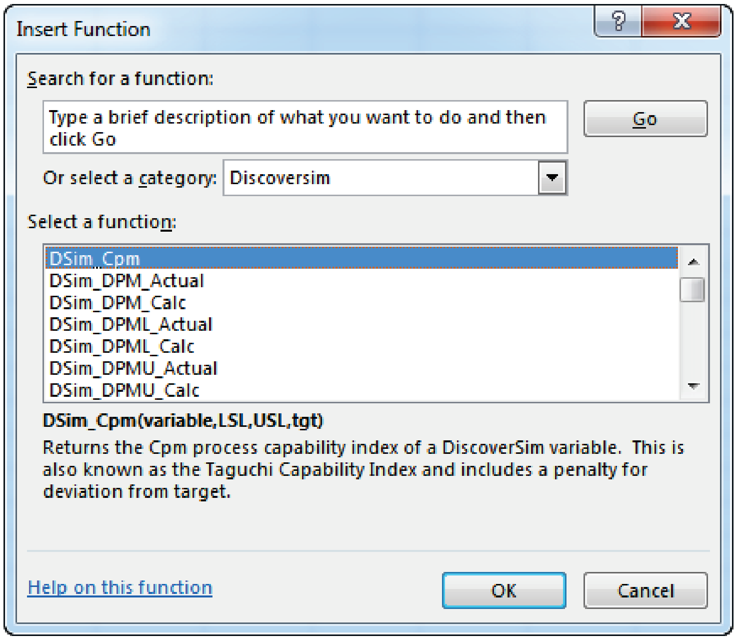
5. Select the comparison operator: **<**, **<=**, **=**, **>=**, >.
   - Note that optimization is much harder with equality constraints than with inequality constraints.
   - The equality tolerance can be set in **Optimization > Advanced Options** (default = 0.001, i.e. the equality constraint is satisfied if **LHS** is within +/- 0.001 of **RHS**).

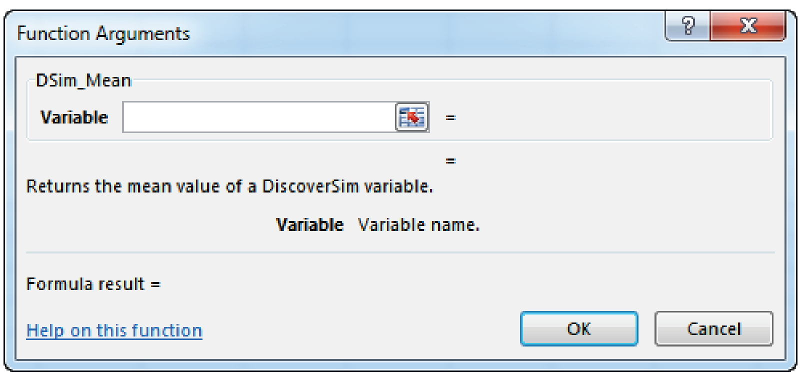
6. **Constraint** Number is provided for reference purposes only.

7. **Constraint Name** is the cell address by default. Enter a name to describe the constraint.

8. Constraint cell color can be modified by clicking on **Cell Address** (default colors can be changed in **Help > DiscoverSim Options**).

Optimization: DSIM Function



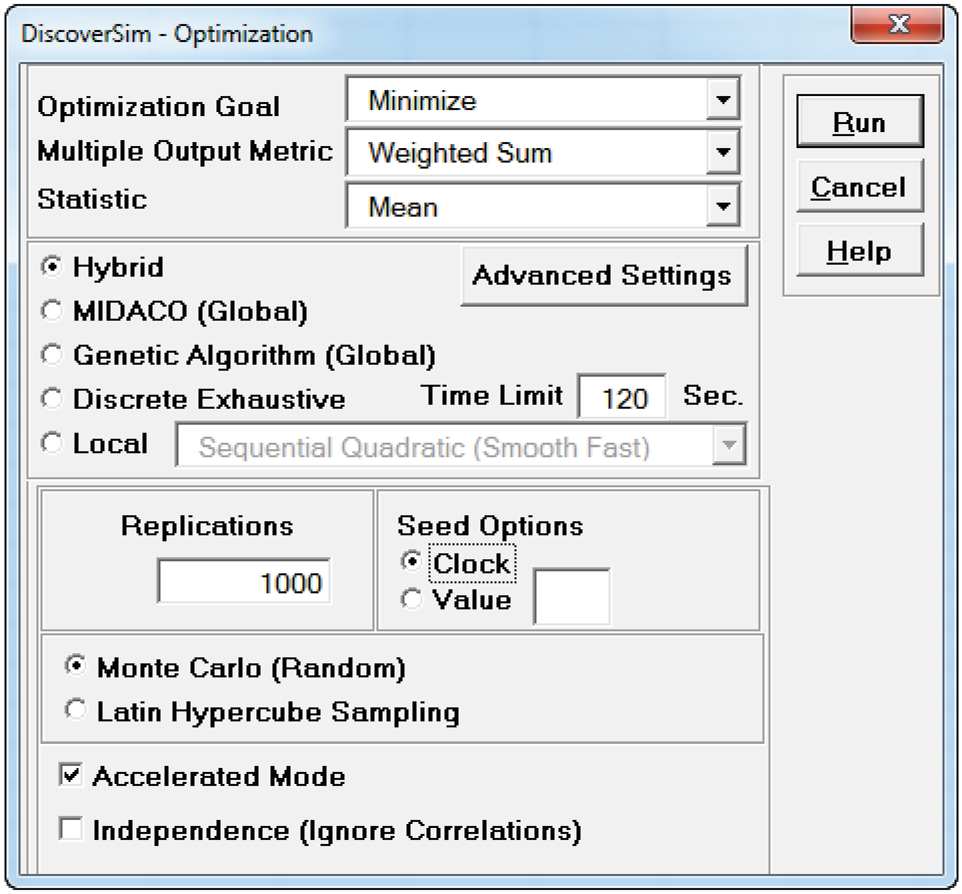


1. Insert a function in a cell such as **DSim_Mean**, **DSim_Pcntle** (Percentile) or **Dsim_Ppk** (Capability Metric). The returned function value is a single number computed using all of the replications within a simulation run or optimization function evaluation. These computed values can then be referenced in a constraint **LHS**.

2. See Appendix: Table of DSIM Functions for all **DSIM Function** descriptions and formulas.

3. See Case Study 3 – Six Sigma DMAIC Project Portfolio Selection for an example of **DSIM Function** used in optimization.

Run Optimization



1. **Optimization Goal/Multiple Output Metric: Minimize** or **Maximize Weighted Sum:** minimize or maximize the *weighted sum of statistic*. Weights are specified for each output in the **Output Response** dialog. For example, if the selected statistic was the Mean, and there were two outputs, the objective function would be:

   $$\text{Weight}_1\text{Mean}_1 + \text{Weight}_2\text{Mean}_2$$

   Note, if there is a single output in the model, this simplifies to minimize or maximize the statistic value.

   Note that the **Output Goal** specified in the **Output Response** dialog is not used here. It is only used in **Maximize Desirability**.

2. **Optimization Goal/Multiple Output Metric: Minimize Deviation from Target:** minimize the square root of *weighted sum of deviations* squared. A target must be specified for each output in the **Output Response** dialog. The only statistic available for this option is the mean. This is also known as the Taguchi or Quadratic Loss Function. If there were two outputs, the objective function would be:

$$\text{SQRT}(\ \text{Weight}_1(\text{Mean}_1 - \text{Target}_1)^2 + \text{Weight}_2(\text{Mean}_2 - \text{Target}_2)^2)$$

3. **Optimization Goal/Multiple Output Metric:** Maximize the weighted linear sum of the Desirability Function (Derringer and Suich, 1980). Each output must specify:

    - **Weight** (default = 1). This is also referred to as "Importance". (Note, another factor, the desirability shape is sometimes called "weight." In DiscoverSim, the desirability shape factor is fixed at 1.)
    - **Output Goal** (Target, Maximize or Minimize) – this is specific to an output. For example, if Output 1 is production rate, the goal would be set to maximize, and Output 2, cost, would have a minimize goal. However, the specified overall objective function goal is to maximize desirability.
        - If the output goal is Target, then **LSL**, **Target**, and **USL** are required. LSL and USL are the lower and upper specification limits used for process capability and dpm calculations, but are also used as the lower and upper bounds for desirability.
        - If the output goal is Minimize, then **Target** and **USL** are required.
        - If the output goal is Maximize, then **LSL** and **Target** are required.

    The only statistic available for Desirability is the mean.

4. For further details on Multiple Output Metrics see Appendix: Formulas for Multiple Output Metrics.

5. **Statistic**: The following tables give the **Statistic** options available depending on the selected **Optimization Goal** and **Multiple Output Metric**:

Optimization Goal:	Minimize	
Multiple Output Metric:	Weighted Sum	Deviation from Target
Statistic:	Mean Median 1st quartile 3rd quartile Percentile (%) Minimum Maximum Standard Deviation Variance Mean Squared Error (requires Target) Skewness Kurtosis Range IQR (75-25) Span (95-5) Actual DPM (defects per million – requires LSL/USL) Actual DPMU (Upper – requires USL) Actual DPML (Lower – requires LSL) Calculated DPM (defects per million assuming normal distribution – requires LSL/USL) Calculated DPMU (Upper – requires USL) Calculated DPML (Lower – requires LSL)	Mean (requires Target)

Optimization Goal:	Maximize	
Multiple Output Metric:	Weighted Sum	Desirability
Statistic:	Mean Median 1st quartile 3rd quartile Percentile (%) Minimum Maximum Standard Deviation Variance Mean Squared Error (requires Target) Skewness Kurtosis Range IQR (75-25) Span (95-5) Pp (requires LSL/USL) PpU (Upper – requires USL) PpL (Lower – requires LSL) Ppk (requires LSL/USL) Cpm (requires Target/LSL/USL) %Pp (Percentile Pp – requires LSL/USL) %PpU (Percentile PpU – requires USL) %PpL (Percentile PpL – requires LSL) %Ppk (Percentile Ppk – requires LSL/USL)	Mean (requires Target and LSL, USL or LSL/USL)

6. Formulas for general statistical measures are given in Appendix: Formulas for Statistical Measures.

7. Formulas for Quality and Process Capability Indices are given in Appendix: Formulas for Quality and Process Capability Indices.

8. DiscoverSim includes the following Global and Local optimization methods:

   - Mixed Integer Distributed Ant Colony Global Optimization (MIDACO)
     - See Appendix: MIDACO Mixed Integer Distributed Ant Colony Global Optimization for more details, including **Advanced Settings** for MIDACO**.**

   - Genetic Algorithm (GA) Global Optimization
     - See Appendix: Genetic Algorithm (GA) Global Optimization for more details, including **Advanced Settings** for GA.

   - Discrete Exhaustive optimization – all combinations for small discrete problems
     - See Appendix: Discrete Exhaustive Global Optimization for more details, including **Advanced Settings** for Discrete Exhaustive**.**

   - Sequential Quadratic Programming (SQP) for fast local smooth optimization
     - See Appendix: Sequential Quadratic Programming (SQP) Local Optimization for more details, including **Advanced Settings** for SQP**.**

   - Nelder-Mead (NM) simplex optimization for fast local non-smooth problems
     - See Appendix: Nelder-Mead (NM) Local Optimization for more details, including **Advanced Settings** for NM**.**

   - Powerful **Hybrid** of the above methods:
     - All discrete controls: MIDACO or Exhaustive Discrete if applicable
     - All continuous controls: MIDACO, GA, followed by SQP or NM
     - Mixed continuous/discrete controls: MIDACO 1 (Course), MIDACO 2 (Fine), followed by SQP or NM
     - See Appendix: HYBRID Optimization for more details, including **Advanced Settings** for Hybrid**.**

   - Note that the **Advanced Settings** button will open a dialog with settings specific to the selected method.

9. **Replications** value sets the number of replications used in optimization to obtain the **Statistic**. The default value is 1000.

10. **Seed** is set to **Clock** by default so that the starting seed of random number generation will be different with each run. If you want the optimization results to match every time (for example in a classroom setting where you want all students to obtain the same results),

select **Value** and enter an integer number. Stochastic optimization requires a fixed seed in order to avoid “chatter” that would result in inconsistent comparisons. If the **Seed** is set to **Clock**, the initial seed value will be obtained from the system clock and then kept fixed throughout the optimization. Note that the results of a fixed seed for 32-bit Excel will be slightly different for 64-bit Excel.

11. Select **Monte Carlo (Random)** for full randomization. **Latin Hypercube Sampling** is less random than Monte Carlo but enables more accurate simulations with fewer replications.

12. **Accelerated Mode** uses DiscoverSim’s Excel Formula Interpreter to dramatically increase the speed of calculations for rapid optimization. If unchecked, the calculations are performed using native Excel. The interpreter supports the majority of all Excel numeric functions (for more details see Appendix: DiscoverSim Engine and Excel Formula Interpreter). If the DiscoverSim interpreter sees a function that it does not support, you will be prompted to use the Excel Native mode.

13. Check **Independence (Ignore Correlations)** to run the optimization with all inputs independent of each other (zero correlation).

14. Optimization in process may be interrupted/paused or stopped. Upon completion, the user can paste the optimum input control values in order to perform further simulation studies.

DiscoverSim™: Case Studies

Case Study 1 – Basic Profit Simulation

Introduction: Profit Simulation

This is an example of how DiscoverSim Monte Carlo simulation can be used to determine probability of daily profit using a basic profit model for a small retail business. We will apply distribution fitting to historical data and specify input correlations to define the model in a way that closely matches our real world business.

The profit (Pr) requirement is Pr > 0 dollars (i.e., the lower specification limit LSL = 0)

The profit equation, or "Y = f(X) transfer function," is calculated as follows:

Total Revenue, TR = Quantity Sold * Price

Total Cost, TC = Quantity Sold * Variable Cost + Fixed Cost

Profit, Pr = TR – TC

In this study we will use DiscoverSim to help us answer the following questions:

1. What is the predicted probability of daily profit?
2. What are the key X variables that influence profit Y? Can we reduce the variation in profit by reducing the variation of the important input variables?

Summary of DiscoverSim Features Demonstrated in Case Study 1:

- Distribution Fitting – Discrete Batch Fit
- Distribution Fitting – Continuous Batch Fit
- Distribution Fitting – Specified Distribution Fit
- Create Input Distributions with Stored Distribution Fit
- Specify Input Correlations
- Run Simulation and display
  - Histograms, Descriptive Statistics, Process Capability Report
  - Percentile Report
  - Scatter Plot/Correlation Matrix
  - Sensitivity Chart of Correlation Coefficients
  - Sensitivity Chart of Regression Coefficients
- Distribution Fitting – Nonnormal Process Capability

Profit Simulation with DiscoverSim

1. Open the workbook **Profit Simulation**. DiscoverSim Input Distributions will simulate the variability in Quantity Sold, Price and Variable Cost. We will use distribution fitting with historical data to determine which distributions to use and what parameter values to enter. Input Distributions will then be specified in cells **C11**, **C12** and **C13**. The output, Profit, will be specified at cell **C21** using the formula given above.

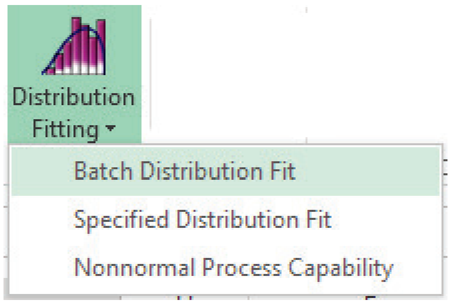
	A	B	C
9			
10		**Input**	**DiscoverSim Input Distribution**
11		Quantity Sold (QS)	97
12		Price (P)	$10.05
13		Variable Cost (%)	39.00%
14			
15		Variable Cost (VC)	$3.92
16		Fixed Cost (FC)	$200.00
17			
18			**Output**
19		Total Revenue (TR)	$974.85
20		Total Cost (TC)	$580.19
21		Profit (Pr)	$394.66

2. Select the **Historical_Data** sheet. This gives historical data for Quantity Sold, Price and Variable Cost.

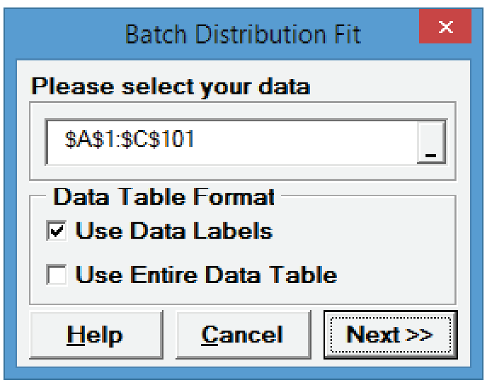
	A	B	C	D	E	F
1	Quantity Sold	Price	Variable Cost (%)			Data Average
2	80	10.16	28.09%		Quantity Sold	97.24
3	95	10.17	49.09%		Price	10.05
4	128	9.62	25.70%		Variable Cost (%)	39.37%
5	118	9.58	28.31%			
6	91	9.65	47.65%			

Tip: At this point use the SigmaXL tool (bundled with DiscoverSim) to obtain descriptive statistics to test for normality (**SigmaXL > Statistical Tools > Descriptive Statistics**). *Quantity Sold* has an Anderson Darling Normality test p-value = 0.83, so may be considered a Normal Distribution, but since it is a count we will apply DiscoverSim's distribution fitting using discrete distributions. *Price* and *Variable Cost* are non-normal with AD p-values less than .05, so we will use DiscoverSim's distribution fitting for continuous data. SigmaXL's correlation matrix should also be used to evaluate correlations (**SigmaXL > Statistical Tools > Correlation Matrix**). The Spearman Rank correlation for *Quantity Sold* versus *Price* is -0.8 (**Note:** DiscoverSim uses the more robust Spearman Rank correlation rather than Pearson's correlation). SigmaXL's graphical tools such as Histograms and Scatterplots should also be used to view the historical data.

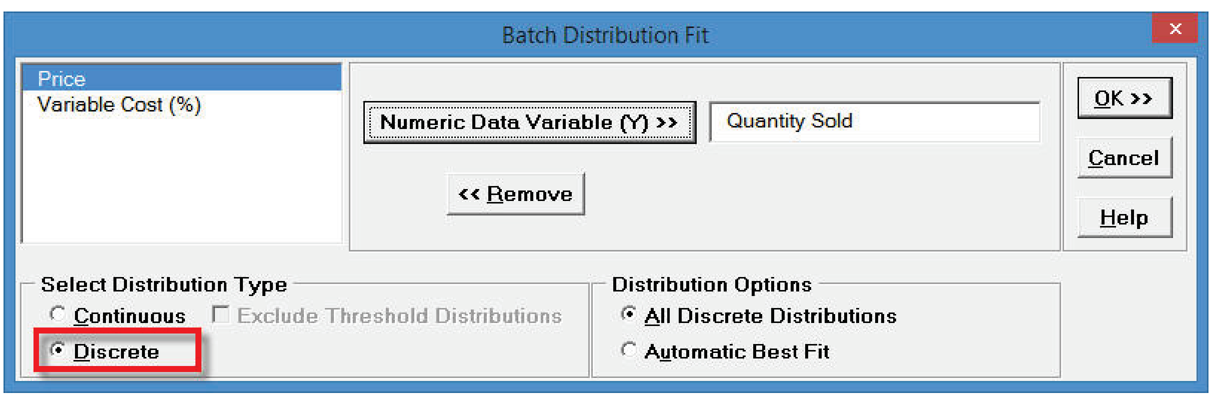
3. Select **DiscoverSim** > **Distribution Fitting > Batch Distribution Fit:**



4. The data range has been preselected and appears in the dialog. **Note:** if a different range is required, click on  to change it or select **Use Entire Data Table** to automatically select the data.



5. Click **Next**. Since *Quantity Sold* is count data, i.e. a discrete variable, select **Discrete.** Select *Quantity Sold* as the **Numeric Data Variable (Y)**. For **Distribution Options** use the default **All Discrete Distributions** as shown:



6. Click **OK**. The resulting discrete distribution report is shown below:

Quantity Sold						
Distribution Name	Chi-Square	Chi-Square P-Value	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value
Negative Binomial	35.42727331	0.9916	Number of Required Events	33	Event Probability	0.253378342
Uniform (Discrete)	78.81411111	0.0359	Minimum	53	Maximum	146
Geometric	406.0388791	< .001	Event Probability	0.010179153		0
Logarithmic	1282.806021	< .001	Event Probability	0.998436851		0
Poisson	9821.096431	< .001	Event Mean	97.24		0

Since these are discrete distributions, Chi-Square is the statistic used to determine goodness-of-fit. The distributions are sorted by Chi-Square in ascending order. The best fit distribution is **Negative Binomial** with Chi-Square = 35.4 and p-value = 0.99. The parameters are **Number of Required Events** = 33 and **Event Probability** = 0.2534. (See Appendix for further details on distributions and distribution fitting).

Tip: If the best fit discrete distribution has a p-value less than .05, indicating a poor goodness-of-fit, none of the discrete distributions are adequate for use in Monte-Carlo simulation. In this case you should redo the distribution fit using the **Continuous** option (Note that you can use the continuous option for discrete data, but you cannot use discrete distributions for continuous data). After creating a DiscoverSim input distribution with the best fit (or normal if applicable), use Excel's ROUND(*number*, 0) function to obtain integer values from the continuous distribution.

7. Now we will apply distribution fitting to *Price* and *Variable Cost*. Select the **Historical_Data** sheet and repeat the above steps. Both variables are continuous so use the default **Select Distribution Type Continuous**. The resulting distribution reports are shown below.

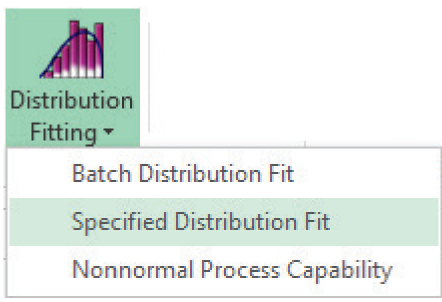
Price										
Distribution Name	AD Stat	AD P-Value	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value	Parameter 3 Name	Parameter 3 Value	Parameter 4 Name	Parameter 4 Value
Beta (4 Parameters)	0.339933952	0.4719	Shape1	1.495422468	Shape2	3.969596105	Minimum	9.160926022	Maximum	12.4075503
Pert	0.354110661	0.4907	Minimum	9.16361092	Likeliest (Mode)	9.627163326	Maximum	12.59599593		0
Johnson SB	0.361319293	0.2935	Location	9.124296559	Scale	2.848916618	Shape1	0.822108658	Shape2	0.907605997
Skew Normal	0.416899627	0.2736	Location (Mean)	9.267950841	Scale (StdDev)	0.967845341	Shape	11.24460024		0
Johnson SL	0.747753248	0.0177	Location	8.743985377	Scale	1	Shape1	-0.378269313	Shape2	2.233494986
Johnson SU	0.747757691	0.0019	Location	8.744034466	Scale	0.0038276	Shape1	-14.35628072	Shape2	2.233403203

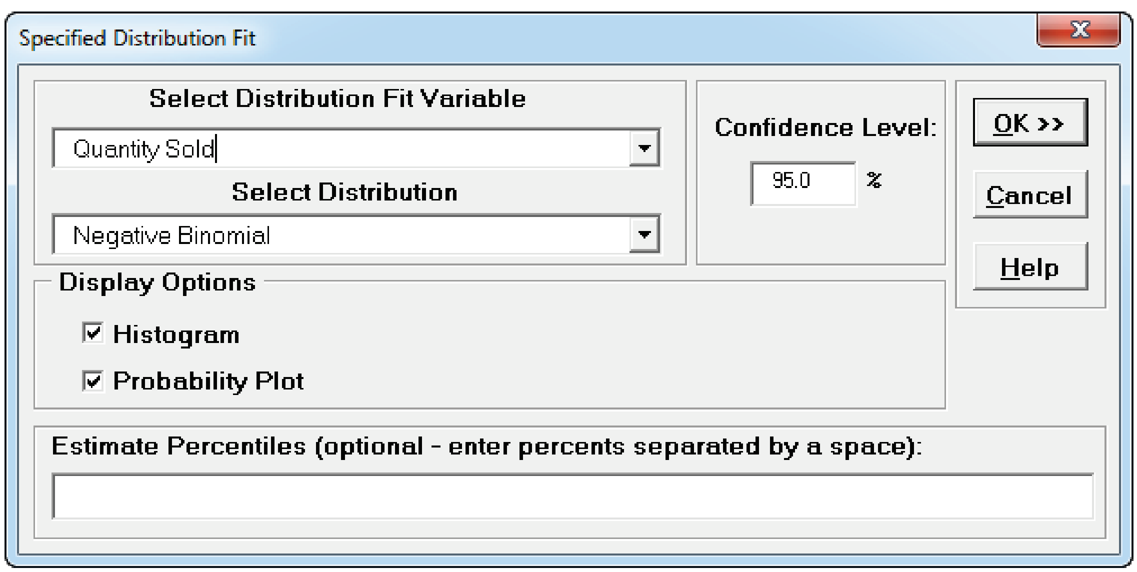
Variable Cost (%)										
Distribution Name	AD Stat	AD P-Value	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value	Parameter 3 Name	Parameter 3 Value	Parameter 4 Name	Parameter 4 Value
Johnson SB	0.314814976	0.4105	Location	0.11890641	Scale	0.682672928	Shape1	0.435815207	Shape2	0.899475645
Skew Normal	0.401492538	0.2964	Location (Mean)	0.195594254	Scale (StdDev)	0.24842092	Shape	5.579766123		0
Fatigue Life	0.546750336	0.177	Scale	0.364358015	Shape	0.401003911		0		0
Inverse Gaussian	0.560346239	0.2074	Location (Mean)	0.393696986	Shape	2.353681252		0		0
Beta (4 Parameters)	0.576750219	0.1932	Shape1	1.244759764	Shape2	1.933521754	Minimum	0.1369863	Maximum	0.786190154
Lognormal	0.58525293	0.1311	Location: Mean(Ln(X))	-1.007464511	Scale: StdDev(Ln(X))	0.395054274		0		0

8. DiscoverSim uses the Anderson Darling statistic to determine goodness-of-fit for continuous distributions. The distributions are sorted by the AD statistic in ascending order. The best fit distribution for *Price* is **Beta (4 Parameters)** with AD Stat = .34 and AD p-value = 0.47. The best fit distribution for *Variable Cost (%)* is **Johnson SB** with AD Stat = .31 and AD p-value = 0.41. (See Appendix for further details on distributions and distribution fitting).

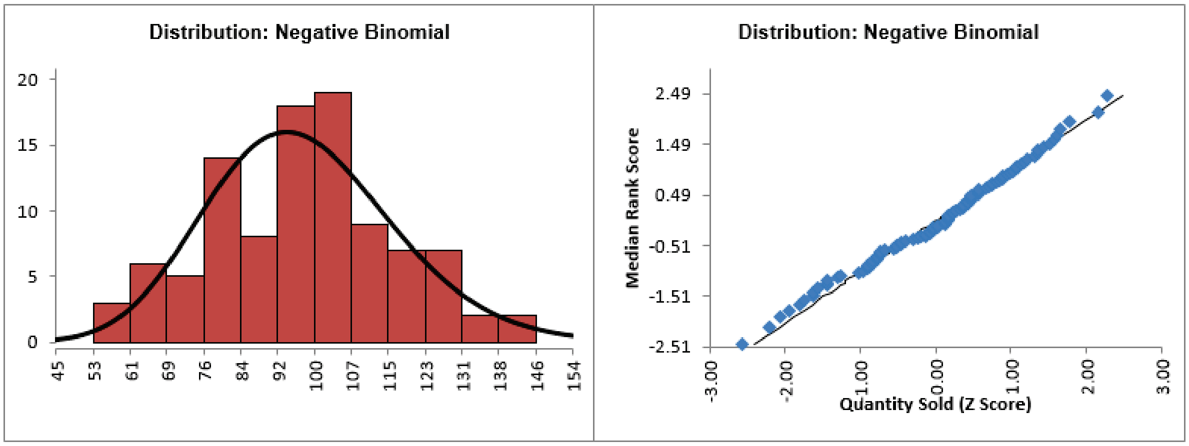
9. **Optional Specified Distribution Fit** analysis: This is a detailed view of the distribution fit for a specified distribution with:
    - **Histogram** and **Probability Plot**.
    - Parameter estimates, standard errors (SE Estimate) and confidence intervals for parameter estimates.
    - Percentile Report with the following percentiles: 0.135, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99, 99.865.
        - Note that the percentiles are computed using the distribution and estimated parameters, not the empirical data.
    - Standard errors and confidence intervals for the percentile values
    - Model Summary and Goodness-of-Fit statistics.

    Select **DiscoverSim** > **Distribution Fitting > Specified Distribution Fit:**





Using the default settings as shown, click **OK**. This produces a detailed distribution fit report for *Quantity Sold* using the *Negative Binomial* distribution:



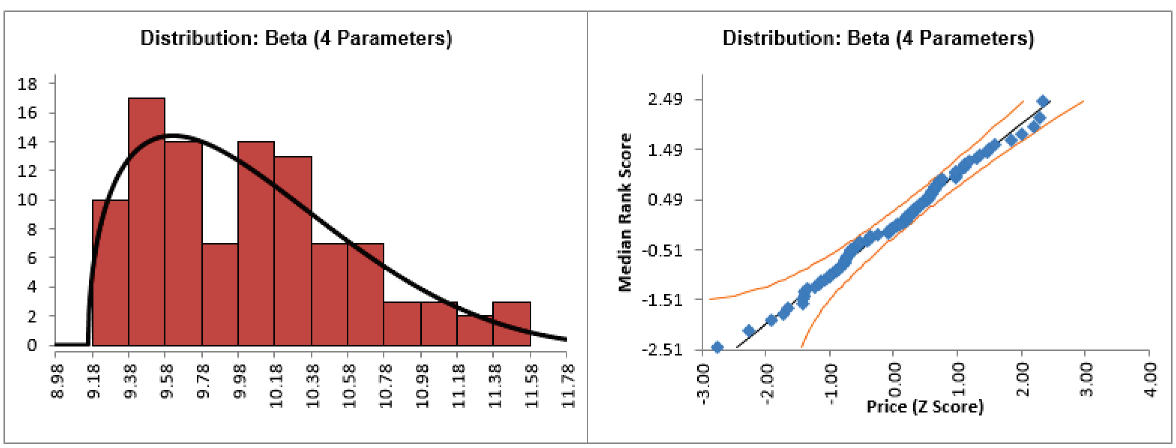
Distribution Fitting Report: Quantity Sold	
Distribution: Negative Binomial	
Sample Count	100
Sample Mean	97.240

Model Summary and Goodness-of-Fit	
Method of Estimation	Maximum Likelihood
Log-Likelihood	-440.042
Chi-Square Negative Binomial	35.427268
Chi-Square Negative Binomial P-Value	0.9916

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Number of Required Events	33			
Event Probability	0.253378	0.003811239	0.245908493	0.260848275

Percentile Report				
Percentage	Quantity Sold	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	48			
1	57			
5	67			

10. Now we will apply specified distribution fitting to *Price* and *Variable Cost*. Repeat the above steps using the default best fit distributions to produce a detailed distribution fit report:

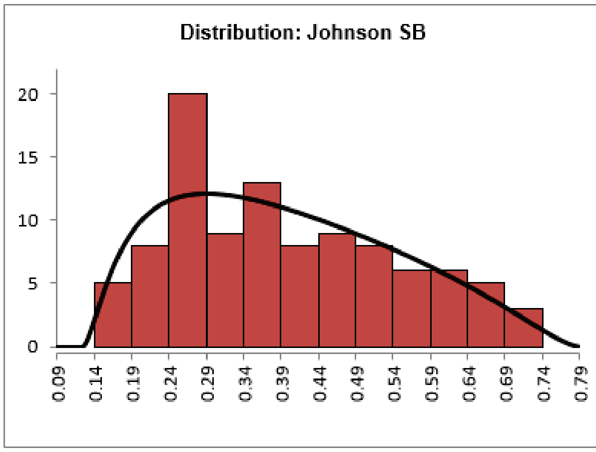


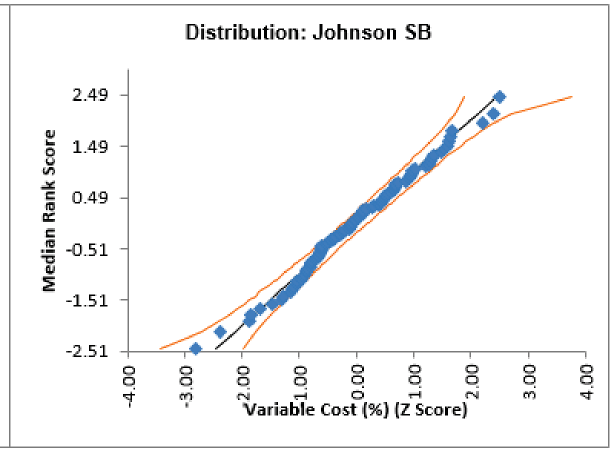
Distribution Fitting Report: Price	
Distribution: Beta (4 Parameters)	
Sample Count	100
Sample Mean	10.05

Model Summary and Goodness-of-Fit	
Method of Estimation	Maximum Likelihood with Min. & Max.
Log-Likelihood	-74.664
AD Beta (4 Parameters)	0.339934
AD Beta (4 Parameters) P-Value	0.4719

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Shape1	1.495	0.193588188	1.115996591	1.874848345
Shape2	3.970	0.568546244	2.855265943	5.083926268
Minimum	9.161	0.070710677	9.022335642	9.299516402
Maximum	12.408	0.070710677	12.26895992	12.54614068

Percentile Report				
Percentage	Price	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	9.172	0.070706462	9.033621767	9.310786006
1	9.204470352	0.07142439	9.06448112	9.344459585
5	9.292921687	0.074072541	9.147742174	9.4381012





Distribution Fitting Report: Variable Cost (%)	
Distribution: Johnson SB	
Sample Count	100
Sample Mean	0.393697

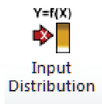
Model Summary and Goodness-of-Fit	
Method of Estimation	Maximum Likelihood
Log-Likelihood	56.474
AD Johnson SB	0.314815
AD Johnson SB P-Value	0.4105

Parameter Estimates				
Parameter	Estimate	SE Estimate	Lower 95.0 CI	Upper 95.0 CI
Location	0.118906	0.015142736	0.089227192	0.148585627
Scale	0.682673	0.061939875	0.561273005	0.804072851
Shape1	0.435815	0.198844371	0.046087402	0.825543013
Shape2	0.899476	0.137469	0.630041357	1.168909934

Percentile Report				
Percentage	Variable Cost (%)	SE Percentile	Lower 95.0 CI	Upper 95.0 CI
0.135	0.133558	0.00995281	0.114050835	0.153065133
1	0.149166171	0.008641743	0.132228665	0.166103677
5	0.180371185	0.010325094	0.160134373	0.200607996

11. The detailed distribution analysis clearly shows that we have a good fit for all 3 variables. Now we will use the stored distribution fits to create a DiscoverSim model.

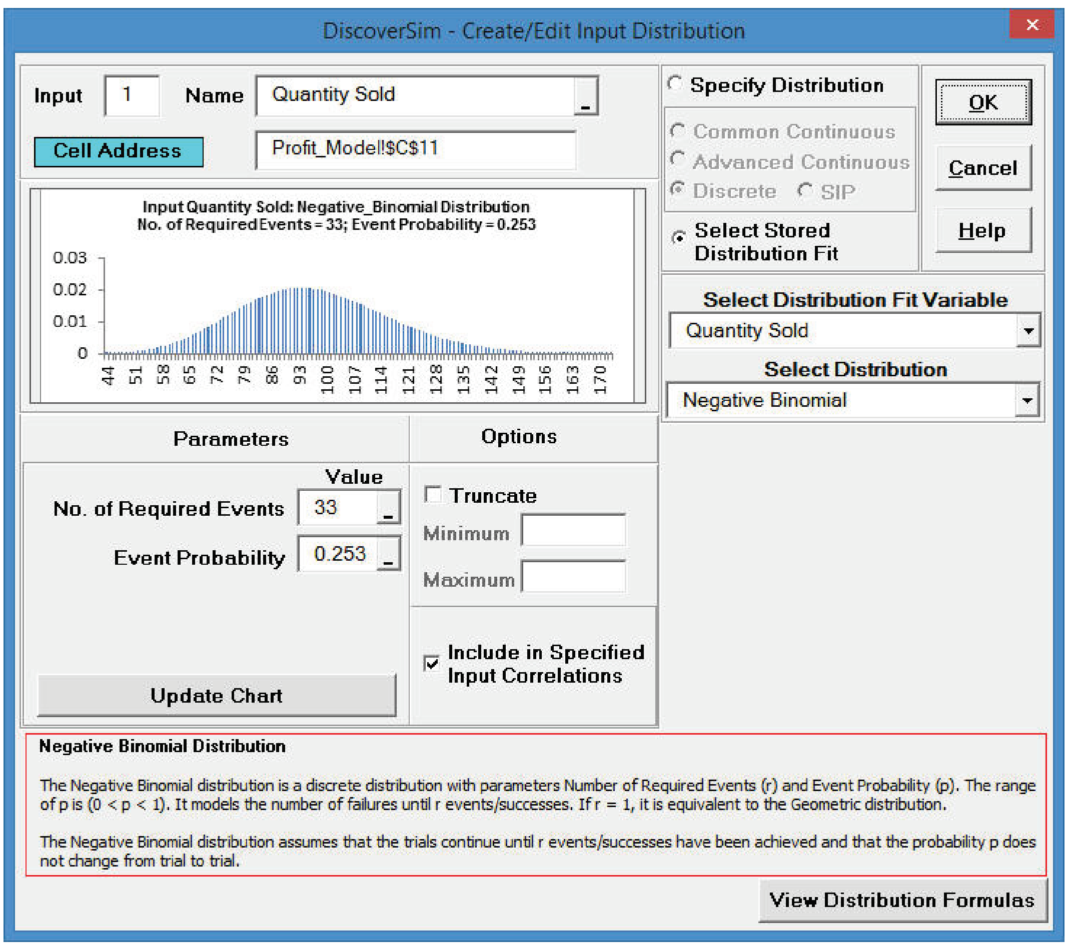
12. Select **Profit_Model** sheet. Click on cell **C11** to specify the Input Distribution for *Quantity Sold*. Select **DiscoverSim > Input Distribution**:



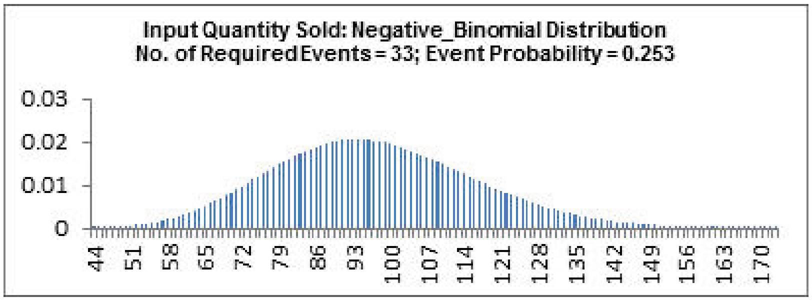
13. Click **Select Stored Distribution Fit**. We will use the default variable *Quantity Sold* for **Select Distribution Fit Variable** and *Negative Binomial* for **Select Distribution**. The parameter values for **Number of Required Events** and **Event Probability** are automatically populated from the distribution fit results for *Quantity Sold*.

14. Enter "Quantity Sold" as the Input **Name**.

15. Click **Update Chart**. The completed Create/Edit Input Distribution dialog is shown below:



16. Click **OK**. Hover the cursor on cell **C11** in the **Profit_Model** sheet to view the DiscoverSim graphical comment showing the distribution and parameter values:

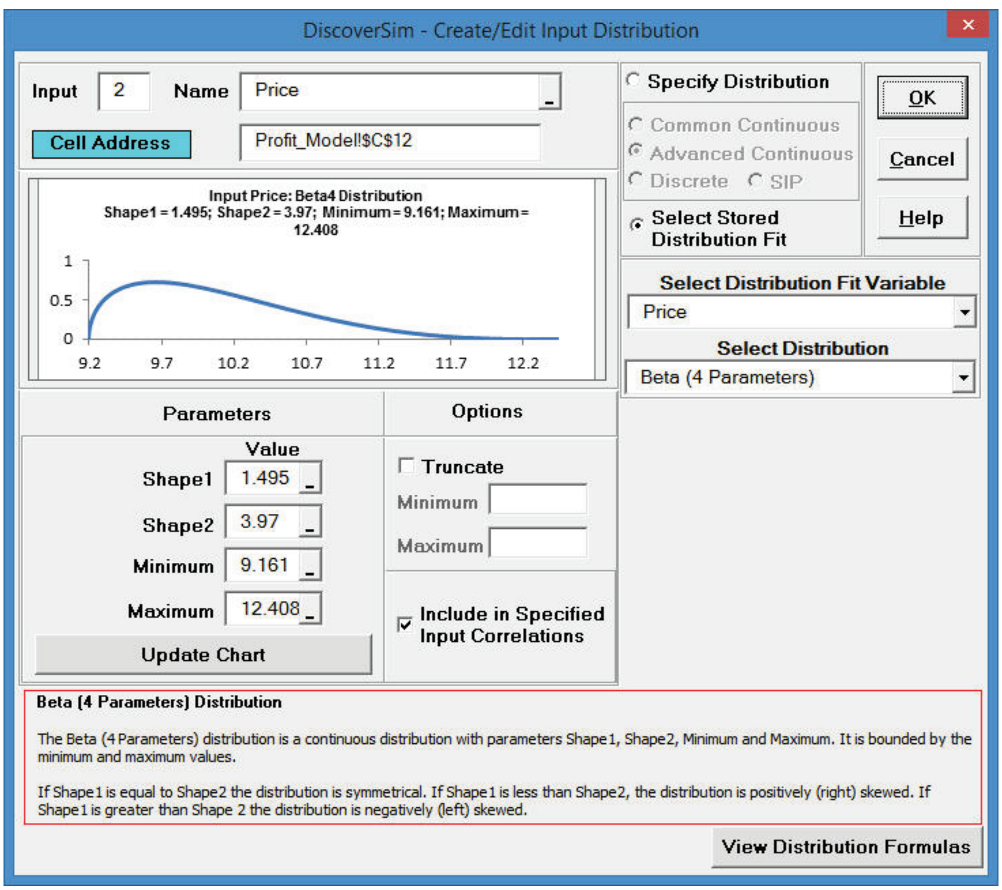


17. Click on cell **C12** to specify the Input Distribution for *Price*. Select **DiscoverSim > Input Distribution.**

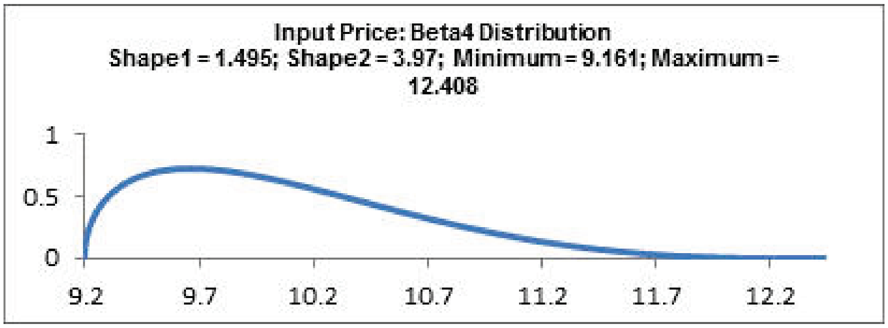
18. Click **Select Stored Distribution Fit**. Select the variable *Price* for **Select Distribution Fit Variable** and the default *Beta (4 Parameters)* for **Select Distribution**. The parameter values for this distribution are automatically populated from the distribution fit results for *Price*.

19. Enter "Price" as the Input **Name**.

20. Click **Update Chart**. The completed Create/Edit Input Distribution dialog is shown below:



21. Click **OK**. Hover the cursor on cell **C12** in the **Profit_Model** sheet to view the DiscoverSim graphical comment showing the distribution and parameter values:

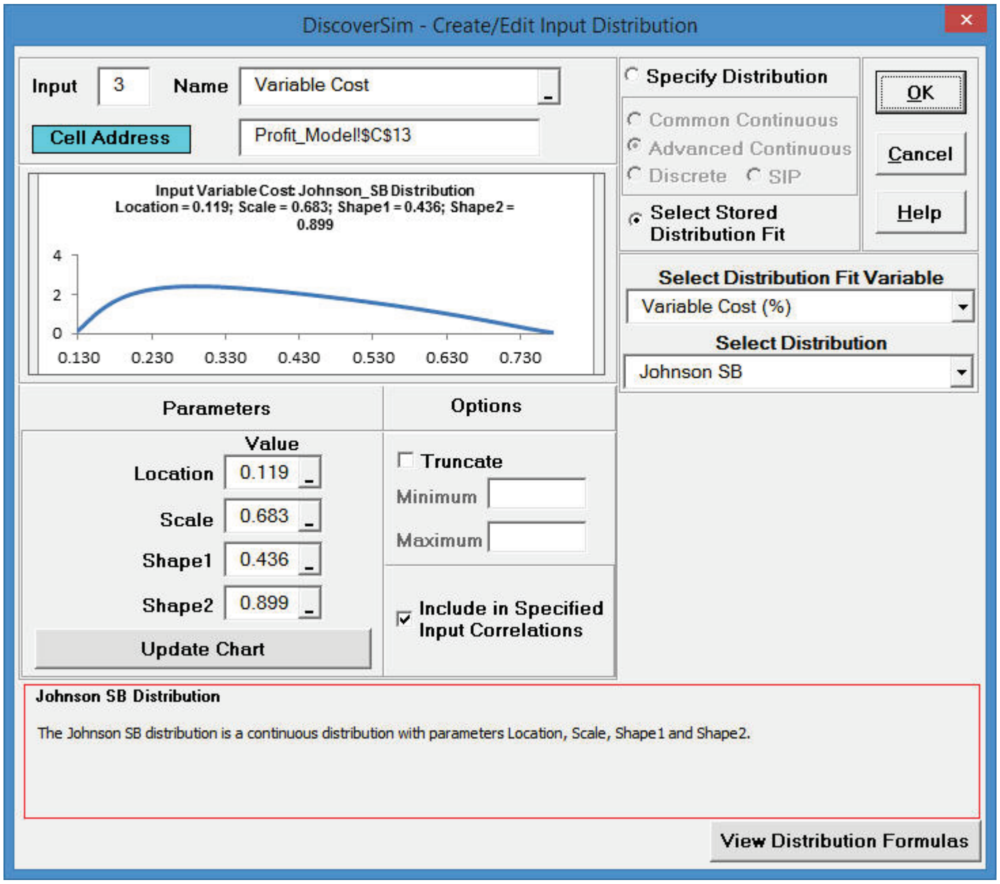


22. Click on cell **C13** to specify the Input Distribution for *Variable Cost*. Select **DiscoverSim > Input Distribution.**

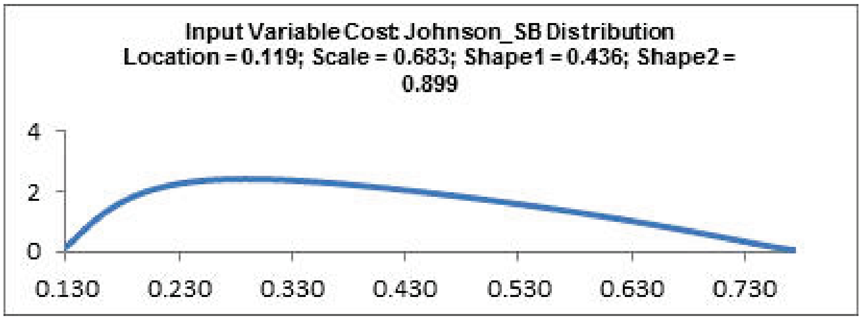
23. Click **Select Stored Distribution Fit**. Select *Variable Cost* for **Select Distribution Fit Variable** and the default *Johnson SB* for **Select Distribution**. The parameter values for this distribution are automatically populated from the distribution fit results for *Variable Cost*.

24. Enter “Variable Cost” as the Input **Name**.

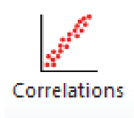
25. Click **Update Chart**. The completed Create/Edit Input Distribution dialog is shown below:


DiscoverSim - Create/Edit Input Distribution
Input 3 Name Variable Cost
Cell Address Profit_Model!$C$13
Input Variable Cost: Johnson_SB Distribution
Location = 0.119; Scale = 0.683; Shape1 = 0.436; Shape2 = 0.899
Specify Distribution
Common Continuous
Advanced Continuous
Discrete
SIP
Select Stored Distribution Fit
OK
Cancel
Help
Select Distribution Fit Variable
Variable Cost (%)
Select Distribution
Johnson SB
Parameters
Options
Value
Location 0.119
Scale 0.683
Shape1 0.436
Shape2 0.899
Truncate
Minimum
Maximum
Include in Specified Input Correlations
Update Chart
Johnson SB Distribution
The Johnson SB distribution is a continuous distribution with parameters Location, Scale, Shape1 and Shape2.
View Distribution Formulas

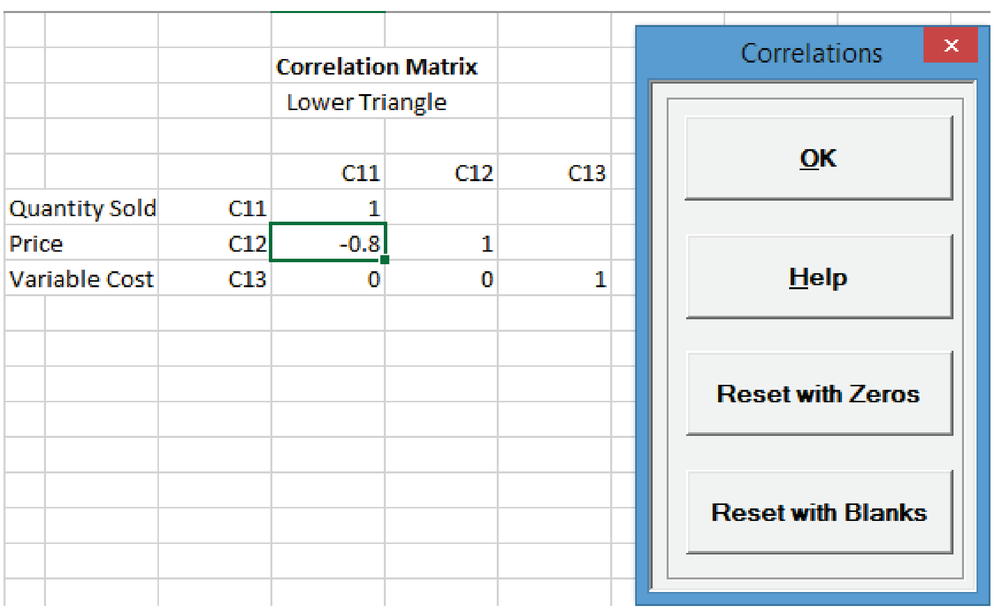
26. Click **OK**. Hover the cursor on cell **C13** in the **Profit_Model** sheet to view the DiscoverSim graphical comment showing the distribution and parameter values:



27. Now we will specify the correlation between the inputs *Quantity Sold* and *Price*. Select **DiscoverSim > Correlations**:



28. As discussed above, the Spearman Rank correlation between Quantity and Price is -.8. This negative correlation is expected: as order quantity increases, unit price decreases. Enter the value -.8 in the column *Price*, row *Quantity* and press Enter.

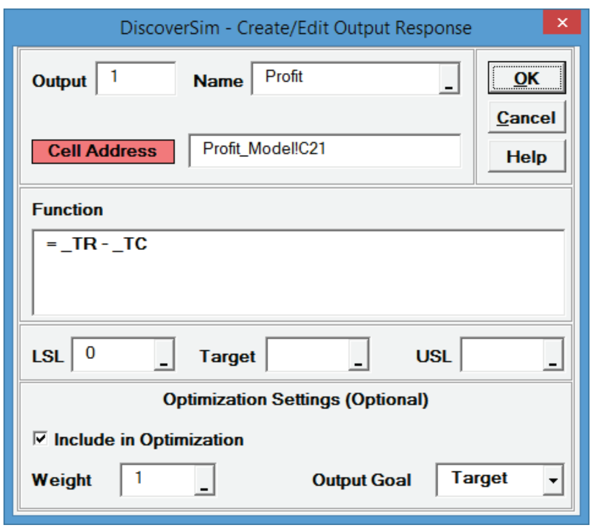


Tips: It is not necessary to enter a correlation value in the upper triangle. The lower triangle specifies the correlations so any value entered in the upper will be ignored. The diagonal of 1's should not be altered. **Reset with Zeros** clears any specified correlations in the lower triangle and replaces them with zeros. **Reset with Blanks** clears any correlations in the lower triangle and replaces them with blanks. This is useful if you wish to then specify correlations between inputs without the constraint of requiring independence on the other inputs.

Click **OK**.

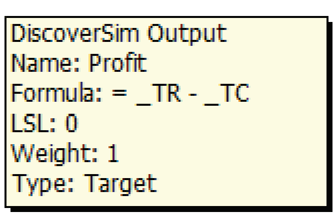
29. Now we will specify the Profit model output. Click on cell **C21**. Select **DiscoverSim > Output Response**:

30. Enter the output **Name** as "Profit". Enter the Lower Specification Limit (**LSL**) as 0. The **Include in Optimization**, **Weight** and **Output Goal** settings are used only for multiple response optimization, so do not need to be modified in this example.



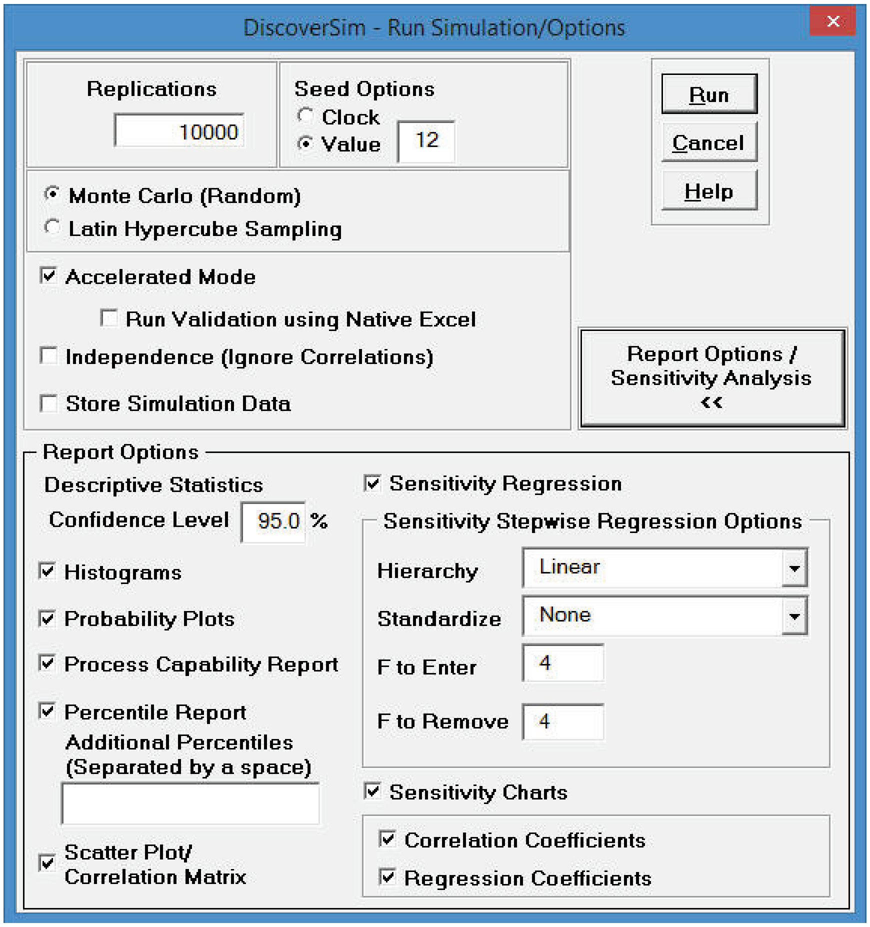
Click **OK**.

31. Hover the cursor on cell **C21** to view the DiscoverSim Output information.



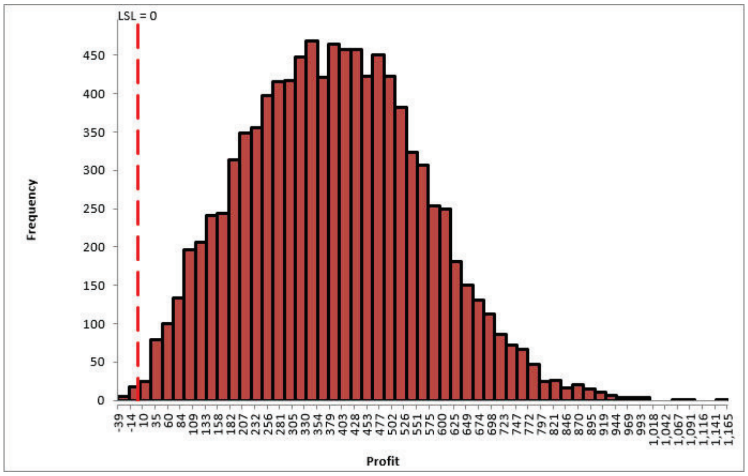
32. Select **DiscoverSim** > **Run Simulation**: 

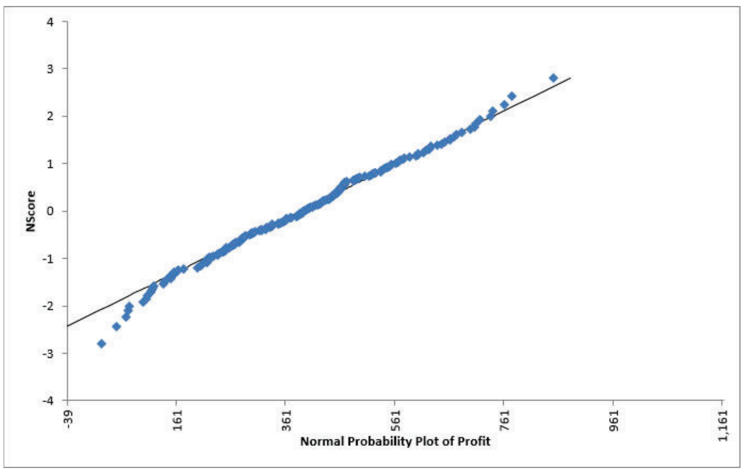
33. Click **Report Options/Sensitivity Analysis**. Check **Percentile Report**, **Scatter Plot/Correlation Matrix**, **Sensitivity Regression Analysis**, **Sensitivity Charts** - **Correlation Coefficients** and **Regression Coefficients**. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below (note that 64 bit DiscoverSim will show slightly different results).



Click **Run**.

34. The DiscoverSim Output Report shows a histogram, normal probability plot, descriptive statistics, process capability indices, a percentile report and a percentile process capability report:





Output 1 Model Information	
Output Name (or Cell Address)	Profit
Function	= _TR - _TC
USL	
Target	
LSL	0
Weight	1
Output Goal	Target

Descriptive Statistics	
Count	10000
Mean	389.019
StDev	176.088351
Range	1203.804
Minimum	-38.716
25th Percentile (Q1)	258.580
50th Percentile (Median)	385.030
75th Percentile (Q3)	508.629
Maximum	1165.088
95.0% CI Mean	385.5674 to 392.4708
95.0% CI Median	381.0643 to 389.943
95.0% CI StDev	173.6814 to 178.5634

Normality Tests	
Anderson-Darling Normality Test	9.782538
P-Value (A-D Test)	0.0000
Skewness	0.241483
P-Value (Skewness)	0.0000
Kurtosis	-0.226779
P-Value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	
Ppl	0.74
Ppk	0.74
Cpm	

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	
ppm < LSL	13579.44
ppm Total	13579.44
% > USL	
% < LSL	1.36
% Total (out of spec.)	1.36
% Total (within spec.)	98.64

Actual Performance (Empirical)	
ppm > USL	
ppm < LSL	2000.00
ppm Total	2000.00
% > USL	
% < LSL	0.20
% Total (out of spec.)	0.20
% Total (within spec.)	99.80

Percentile Report	
0.135	-5.073507
1	42.2017
5	108.3167011
10	158.0384194
15	198.2251881
20	229.9919661
25	258.5798253
30	286.1330656
35	311.9712224
40	335.9178522
45	360.442413
50	385.0300771
55	409.9013885
60	433.0351041
65	459.2568274
70	483.0836531
75	508.6285654
80	537.1547395
85	572.8964055
90	616.5602599
95	688.4845711
99	821.6063331
99.865	947.063505

Percentile Process Capability Indices	
Percentile Pp	
Percentile Ppu	
Percentile Ppl	0.99
Percentile Ppk	0.99

From the histogram and detailed report we see that typically we should expect a positive daily profit, but the variation is large. The likelihood of profit loss is approximately 0.20% (see **Actual Performance (Empirical): %Total (out of spec)**). Note that the "expected" loss of 1.36% assumes a normal distribution, so that is not applicable here because the output distribution is not normal (Anderson-Darling p-value is much less than .05).

Note: If **Seed** is set to **Clock**, there will be slight differences in the reported values with every simulation run due to a different starting seed value derived from the system clock.

The Percentile Process Capability report computes capability indices for non-normal data as follows:

Percentile Pp = (USL – LSL)/(99.865 percentile – 0.135 percentile)

Percentile Ppu = (USL – 50th percentile)/(99.865 percentile – 50th percentile)

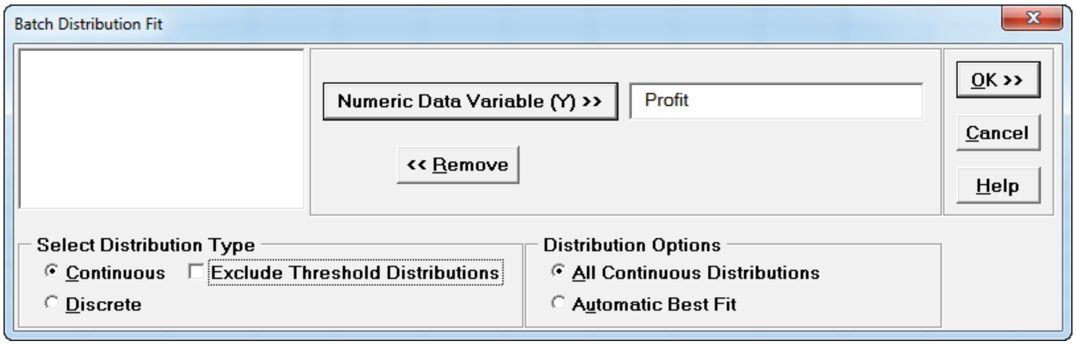
Percentile Ppl = (50th percentile – LSL)/(50th percentile – 0.135 percentile)

Percentile Ppk = min(Ppu, Ppl)

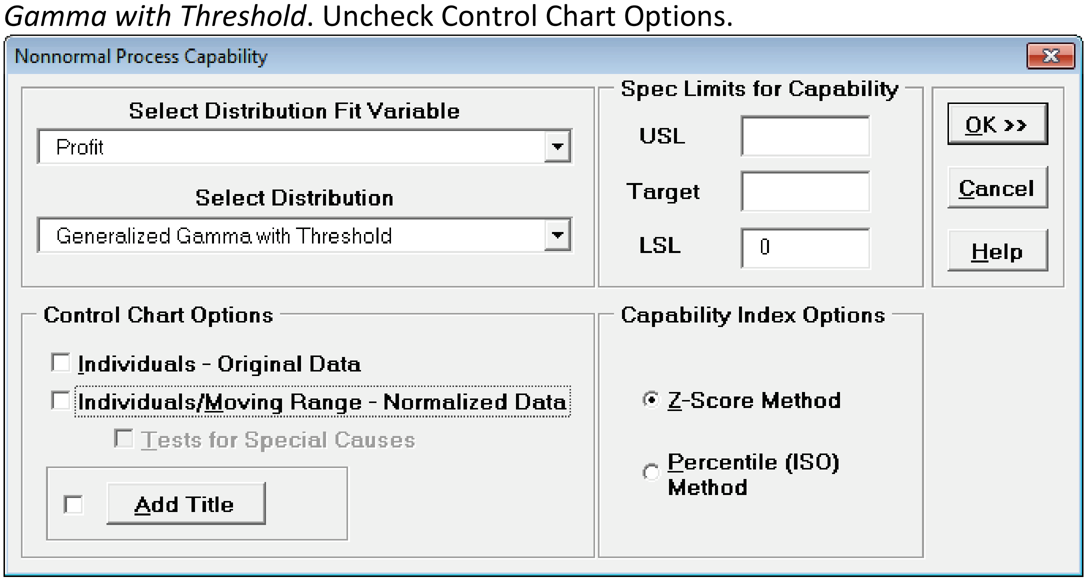
Since we only have a lower specification limit (LSL = 0), Percentile Ppl is calculated as:

Percentile Ppl = (385.03 - 0) / (385.03 - (-5.07)) = 0.99

35. **Optional Nonnormal Process Capability Analysis:** DiscoverSim Version 2.1 now includes **Nonnormal Process Capability** with **Distribution Fitting**. To utilize this stand-alone feature for the Profit data, perform the following steps:
    - Click **Run Simulation**, select **Store Simulation Data**, click **Run**.
    - Select **DSim Data** sheet, select *Profit* column (D1:D10001).
    - Select **Distribution Fitting > Batch Distribution Fit**, click **Next,** select *Profit*, uncheck **Exclude Threshold Distributions:**

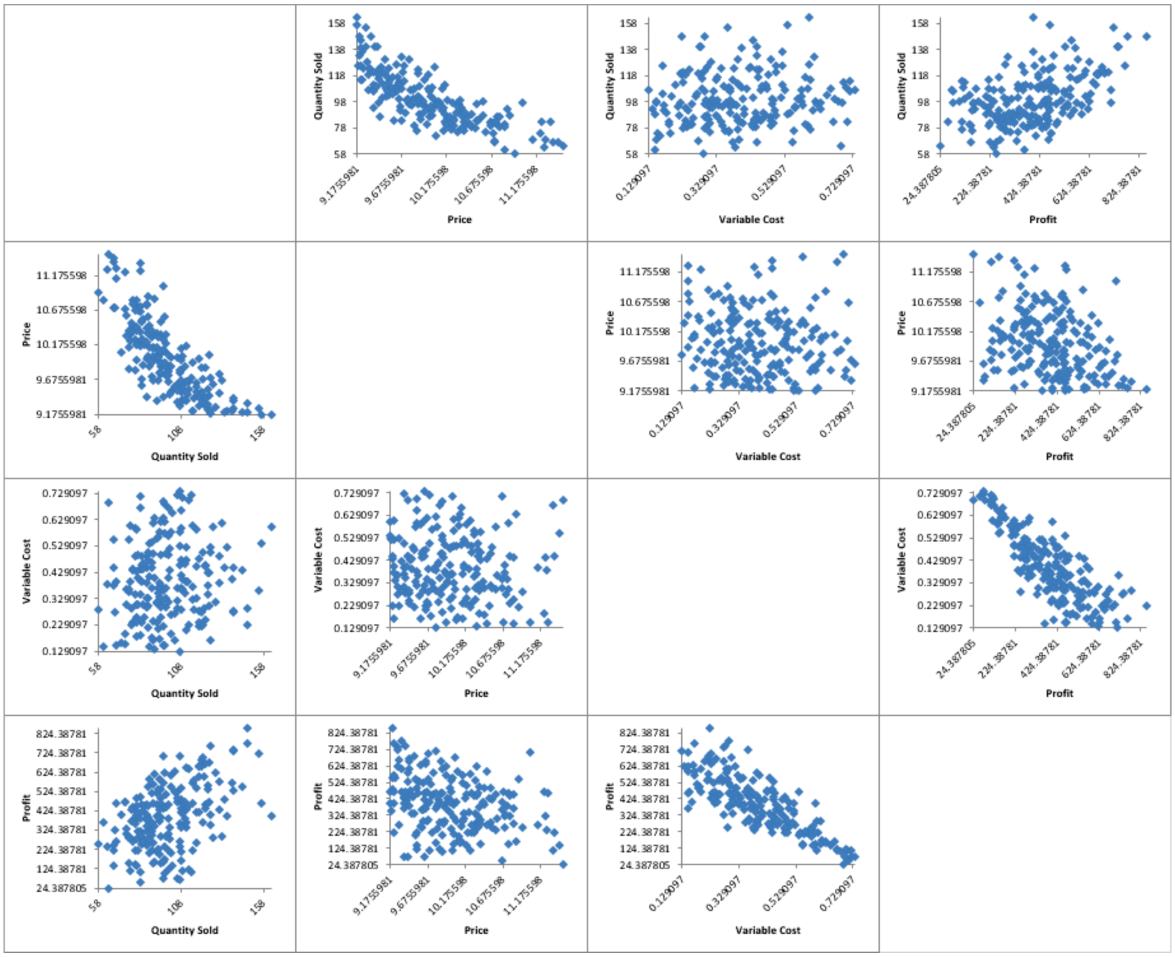


    - Click **OK.** The batch fit will take approximately 1-2 minutes due to the large dataset and inclusion of Threshold distributions.
    - The best fit distribution is Generalized Gamma with Threshold. Unfortunately, it is not a good fit to the data with the AD P-Value < .001, but we will proceed for demonstration purposes. Select **Distribution Fitting** > **Nonnormal Process Capability**.
    - Select *Profit*. Enter **LSL** = 0. The best fit distribution is selected as *Generalized Gamma with Threshold*. Uncheck Control Chart Options.



    - Click **OK**. The Ppk using Generalized Gamma is 0.95 which, even though a poor fit, is closer to the above Percentile Ppk value of 0.99 than the normal distribution Ppk of .74.

36. Click on the **Scatter Plot Matrix** sheet to view the Input-Output relationships graphically:



Here we see the negative correlation that was specified between the inputs “Price” and “Quantity Sold,” as well as the strong negative correlation between “Variable Cost” and “Profit.”

37. Click on the **Correlation Matrix** sheet to view the Input-Output correlations numerically:

Pearson Correlations	Quantity_Sold	Price	Variable_Cost	Profit
Quantity_Sold	1.0000	-0.7862	-0.0042	0.5361
Price		1.0000	-0.0038	-0.3544
Variable_Cost			1.0000	-0.8288
Profit				1.0000

Pearson Probabilities	Quantity_Sold	Price	Variable_Cost	Profit
Quantity_Sold		0.0000	0.6776	0.0000
Price			0.7055	0.0000
Variable_Cost				0.0000
Profit				

Spearman Rank Correlations	Quantity_Sold	Price	Variable_Cost	Profit
Quantity_Sold	1.0000	-0.8027	-0.0017	0.4939
Price		1.0000	-0.0067	-0.3360
Variable_Cost			1.0000	-0.8366
Profit				1.0000

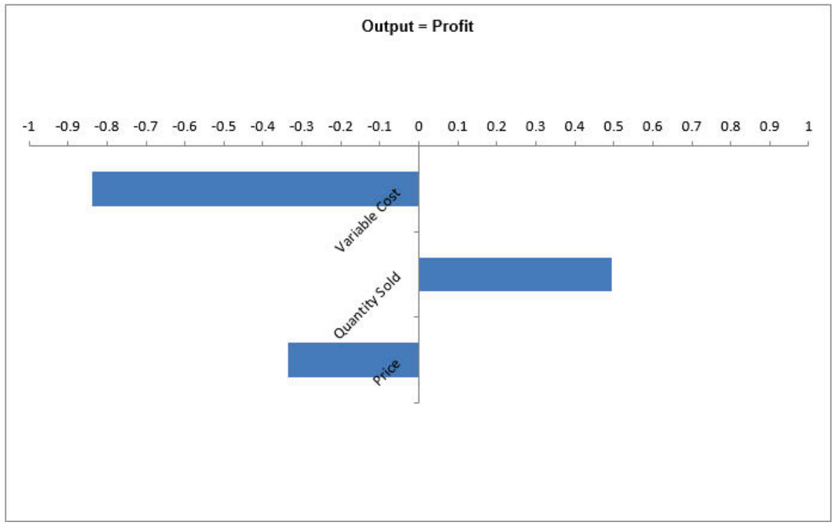
Spearman Rank Probabilities	Quantity_Sold	Price	Variable_Cost	Profit
Quantity_Sold		0.0000	0.8650	0.0000
Price			0.5029	0.0000
Variable_Cost				0.0000
Profit				

DiscoverSim uses the robust Spearman Rank correlation but the Pearson is also reported here for reference purposes.

The correlations highlighted in red are statistically significant (p-value < .05).

38. In order to increase profit (and reduce the variation), we need to understand what is driving profit, i.e., the key "X" factor. To do this we will look at the sensitivity charts. Click on the **Sensitivity Correlations** sheet:

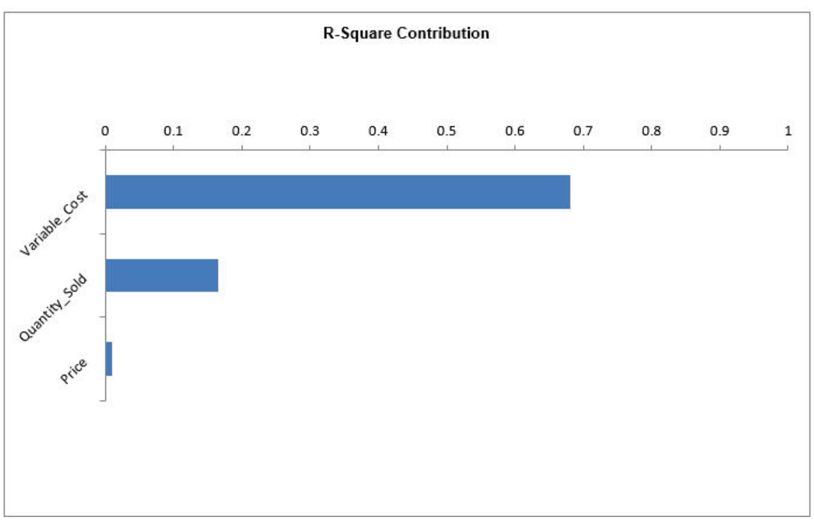
Inputs	Spearman Rank
Variable Cost	-0.836519281
Quantity Sold	0.493928293
Price	-0.33590763



Here we see that "Variable Cost" is the dominant input factor affecting "Profit" with a negative correlation (lower variable cost means higher profit). The next step would then be to find ways to minimize the variable cost (and reduce the variation of variable cost). "Quantity Sold" is the second important input factor. It is interesting to note that "Price" is the least important factor in this Profit simulation model.

39. To view R-Squared percent contribution to variation, click on the **Sensitivity Regression** sheet:

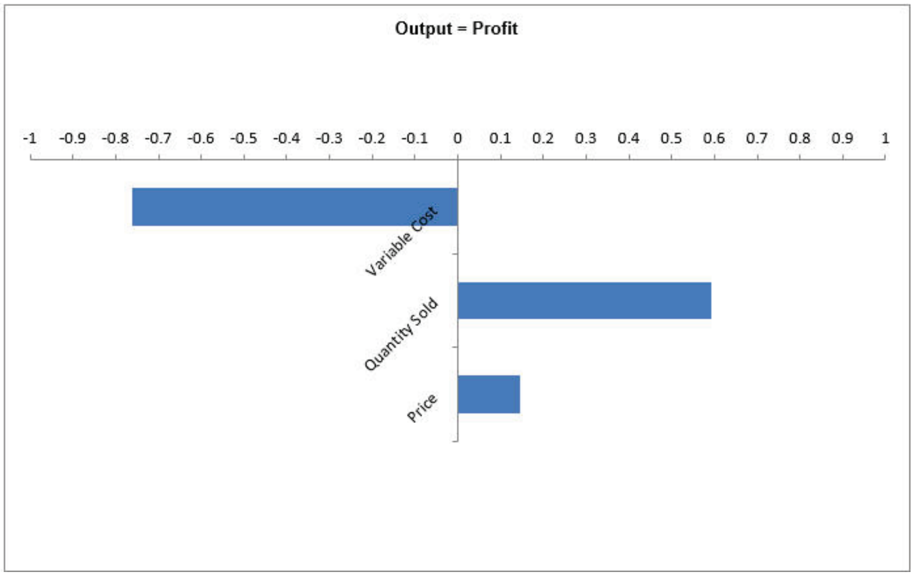
Model Term	Coefficient	R-Square
Constant	-298.9808946	
Quantity_Sold	5.899012197	0.165686787
Price	49.31937913	0.009815191
Variable_Cost	-972.7559521	0.681356677



Variable Cost contributes to 68% of the variation in Profit (R-Square = 0.68).

40. When strong correlations are present in the inputs, Sensitivity Correlation Analysis and Sensitivity Regression Analysis may be misleading, so it is recommended that the simulation be rerun with **Independence (Ignore Correlations)** checked to validate the sensitivity results:

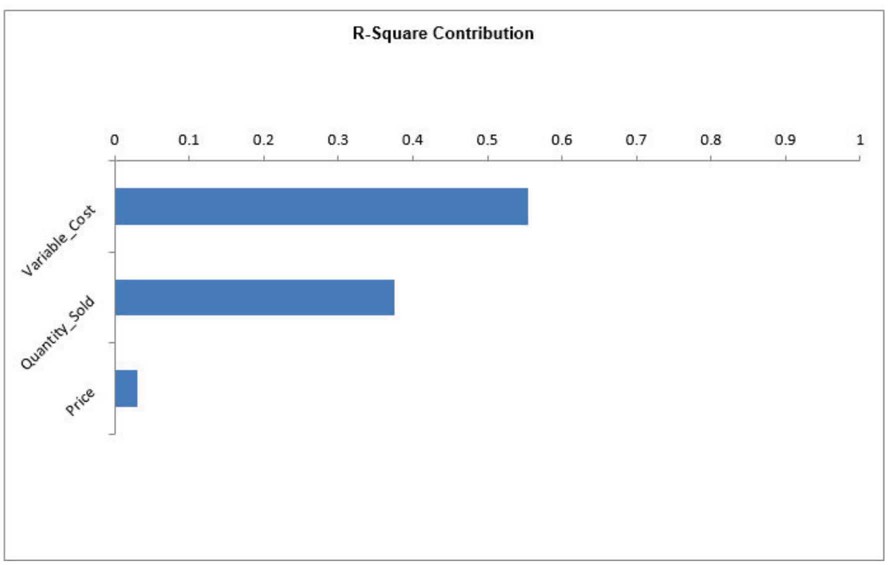
Inputs	Spearman Rank
Variable Cost	-0.760628156
Quantity Sold	0.591725764
Price	0.144948633



In this case the input factor prioritization remains the same, but "Price" shows a small positive correlation rather than negative.

41. The revised R-Square Percent Contribution report is shown below:

Model Term	Coefficient	R-Square
Constant	-413.0607097	
Quantity_Sold	6.087253016	0.37607235
Price	59.34236653	0.029897543
Variable_Cost	-971.1009674	0.554581007



Case Study 2 –Magazine Production Optimization

Introduction: Simulation and Optimization to Determine Optimal Magazine Production Quantity to Maximize Profit

This is an example of DiscoverSim simulation and optimization to determine optimal magazine production quantity to maximize net profit.

Monthly Demand is modeled using a discrete custom distribution.

A DiscoverSim Input Control is used to set the number of magazines produced (in steps of 10,000).

Production and disposal costs are variable costs per unit. (**Note:** content and fixed printing costs are covered by advertising and subscriptions).

Revenue = MIN(Monthly_Demand,Monthly_Demand,Magazines_Produced)*Selling_Price_to_Retailers

Total Production Cost = Magazines_Produced*Production_Cost

Total Disposal Cost = IF((Magazines_Produced-Monthly_Demand)>0,Magazines_Produced-Monthly_Demand,0)*Disposal_Cost

Net Profit = Revenue - (Total_Production_Cost + Total_Disposal_Cost)

Summary of DiscoverSim Features Demonstrated in Case Study 2:

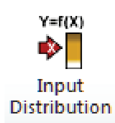
- Create Input Distributions with Custom Discrete
- Create Discrete Input Controls
- Run Simulation
- Run Optimization

Optimization of Magazine Production Quantity with DiscoverSim

1. Open the workbook **Magazine Production – Optimization**. A custom discrete input distribution will simulate the variability in monthly demand and is specified in cell **C15**. The demand quantity data for this custom distribution is given in cells **F15: F24** with weights specified in cells **G15:G24**. The weights are converted to calculated probabilities in cells **H15:H24**. A DiscoverSim Input Control (Discrete) will be used to specify the magazine production quantity (10,000 to 100,000) at cell **C17**. The output, Net Profit will be specified at cell **C27**.

	A	B	C	D	E	F	G	H
12								
13						Data for DiscoverSim Custom Distribution - Discrete		
14						Qty	Weight	Calculated Probability
15		Monthly Demand	50000	DiscoverSim Input Distribution - Custom Discrete		10000	1	0.0513
16						20000	1.5	0.0769
17		Magazines Produced	100000	DiscoverSim Input Control - Discrete (10,000 to 100,000 with Step Size of 10,000)		30000	2	0.1026
18						40000	3	0.1538
19		Production Cost (per unit)	$ 0.50			50000	4	0.2051
20		Selling Price to Retailers (per unit)	$ 2.00			60000	3	0.1538
21		Disposal Cost (per unit)	$ 0.20			70000	2	0.1026
22						80000	1.5	0.0769
23		Revenue	$100,000.00			90000	1	0.0513
24		Total Production Cost	$ 50,000.00			100000	0.5	0.0256
25		Total Disposal Cost	$ 10,000.00					
26								
27		Net Profit	$ 40,000.00	DiscoverSim Output				

2. Click on cell **C15** to specify the "Monthly Demand" discrete custom input distribution. Select **DiscoverSim > Input Distribution**:



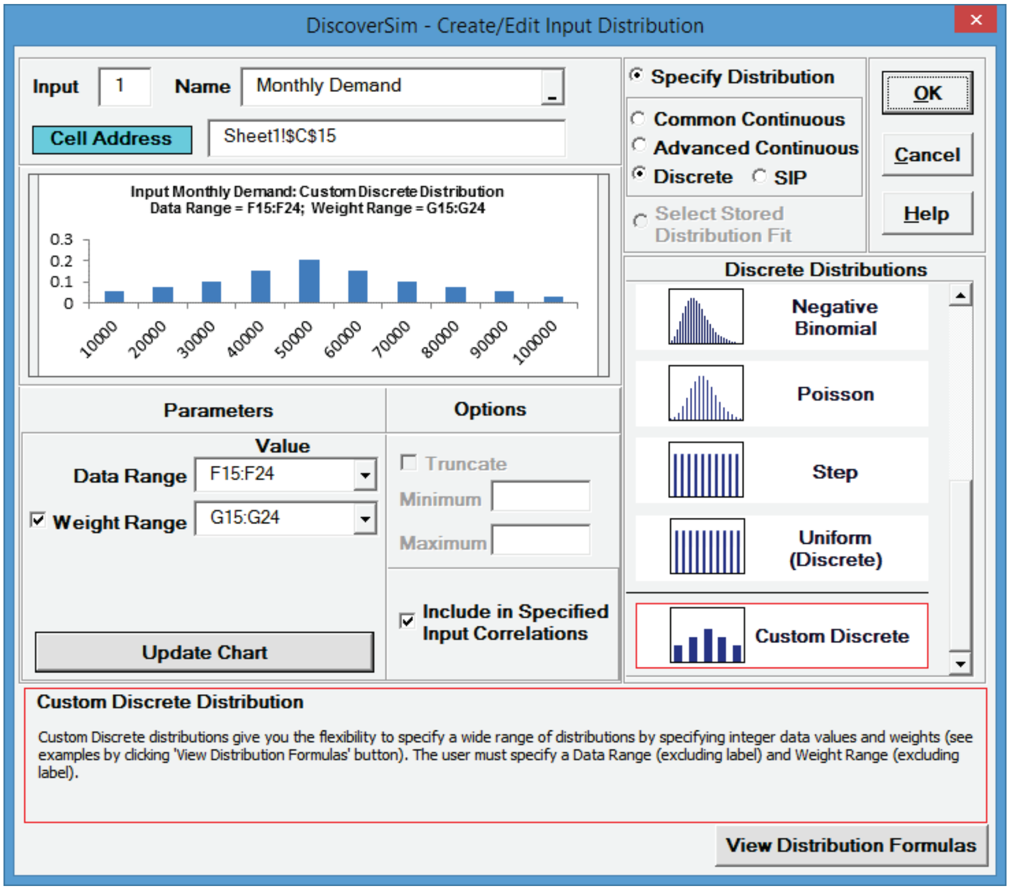
3. Select **Discrete** > **Custom Discrete Distribution**.

4. Click the input name cell reference  and specify cell **B15** containing the input **Name** "Monthly Demand" (**Note:** spaces are replaced with the underscore "_" character for all input names). After specifying a cell reference, the dropdown symbol changes from  to  .

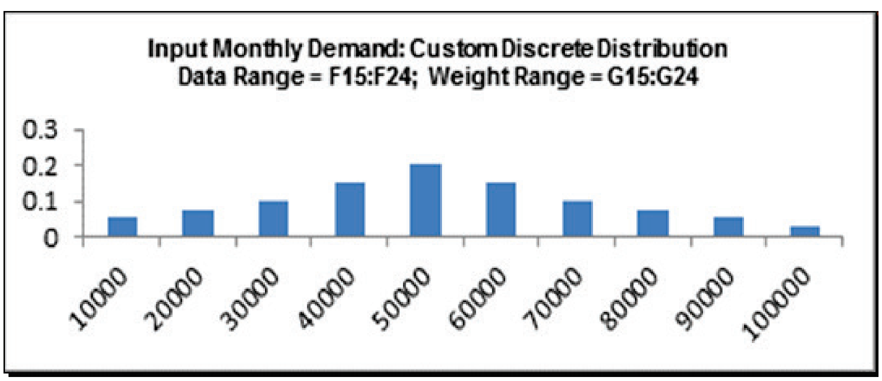
5. Click the **Data Range** cell reference  and specify cells **F15:F24** as the data range for the custom discrete distribution.

6. Click the **checkbox** to enable **Weight Range** followed by the cell reference  and specify cells **G15:G24** as the weight range for the custom discrete distribution. These weights are used to determine the distribution probabilities as displayed in cells **H15:H24**.

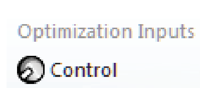
7. Click **Update Chart** to view the custom discrete distribution as shown:



8. Click **OK**. Hover the cursor on cell **C15** to view the DiscoverSim graphical comment showing the distribution and parameter values:

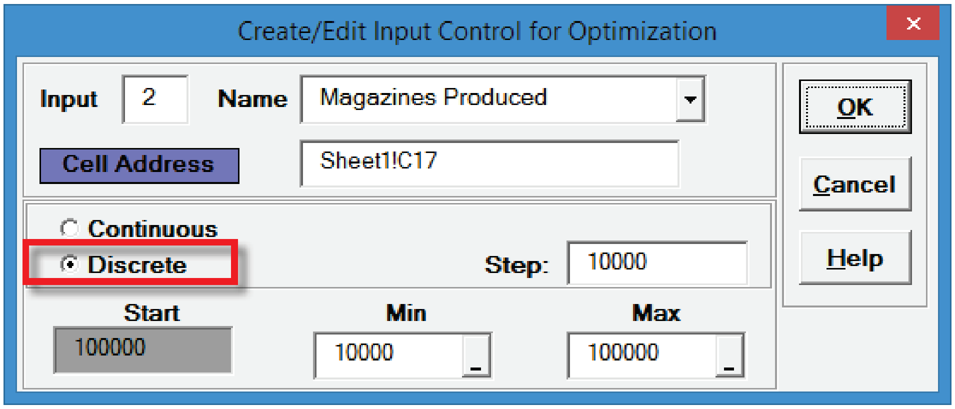


9. Now we will add the Discrete Input Control (also referred to as a “Decision Variable”). This will be used to set the magazine production quantity. DiscoverSim will use Stochastic Global Optimization to determine what quantity will maximize the mean net profit.

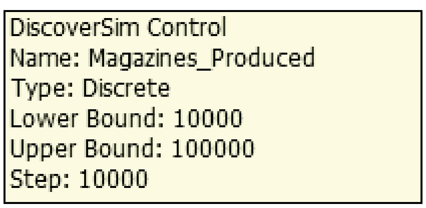
10. Click on cell **C17**. Select **Control**: 

11. Click input **Name** cell reference and specify cell **B17** containing the Input Control name "Magazines Produced".

12. Select **Discrete** (this Input Control setting will consider only discrete values). The **Start** value is set to 100000 from cell **C17** (current production volume per month). Set the **Min** value to 10000, **Max** value to 100000 and **Step** to 10000 as shown:



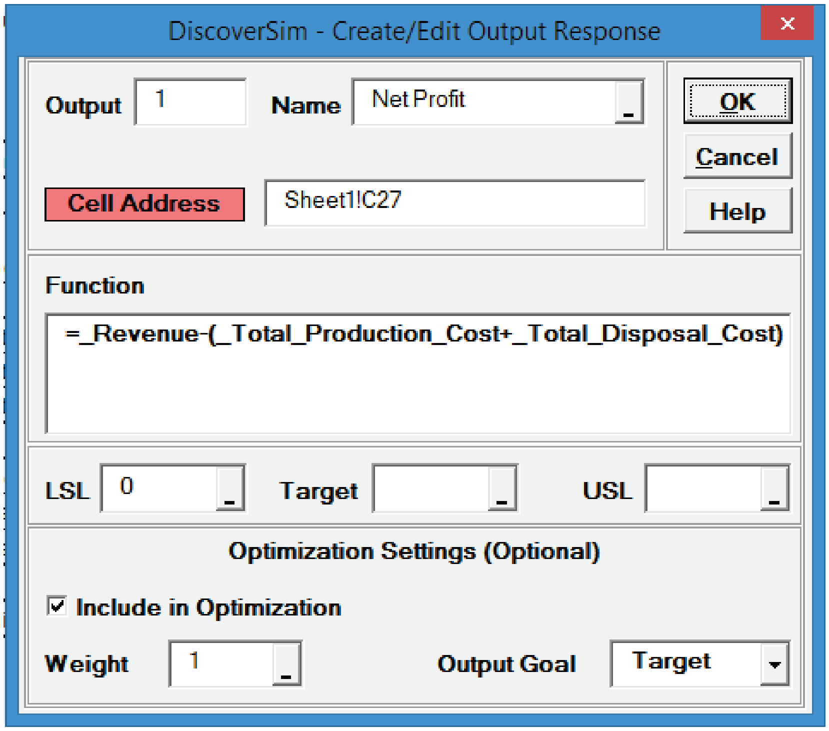
13. Click **OK**. Hover the cursor on cell **C17** to view the comment displaying the Input Control settings:



14. Now we will specify the output. Click on cell **C27**. The cell contains the Excel formula for the Net Profit: = _Revenue - (_Total_Production_Cost + _Total_Disposal_Cost).

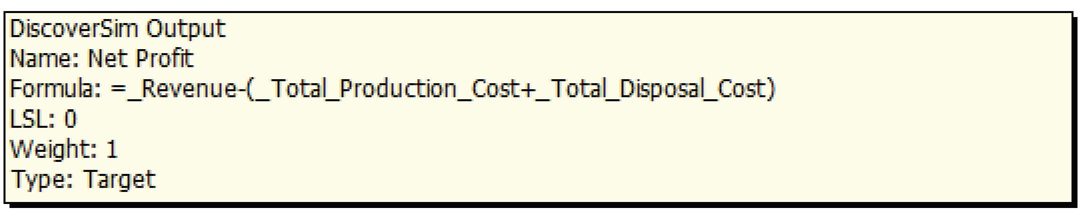
15. Select **DiscoverSim** > **Output Response**:

16. Click output **Name** cell reference and specify cell **B27** containing the Output name "Net Profit." Enter the Lower Specification Limit (**LSL**) as 0. The **Weight** and **Output Goal** settings are used only for multiple response optimization, so do not need to be modified in this example.



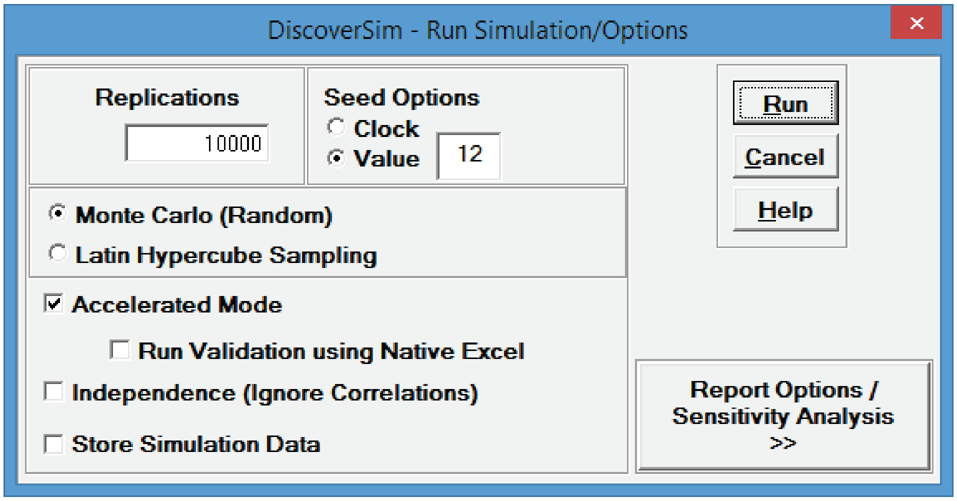
Click **OK**.

17. Hover the cursor on cell **C27** to view the DiscoverSim Output information.



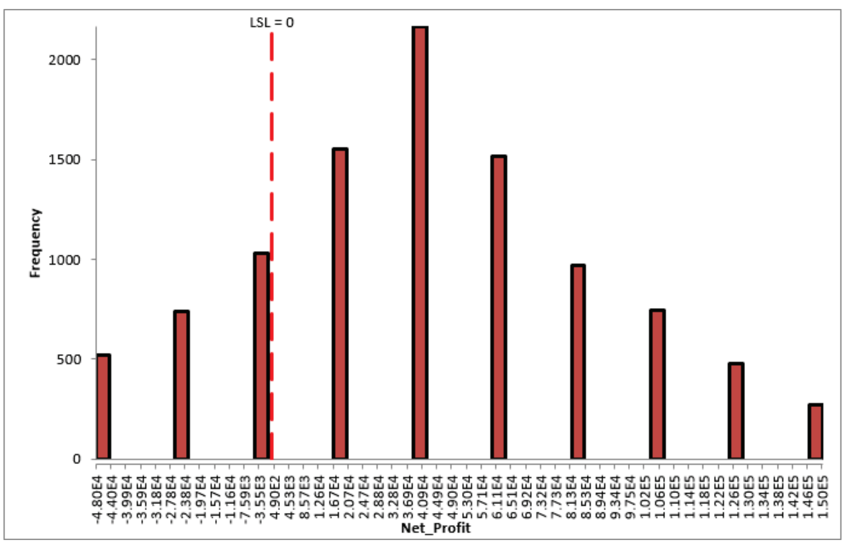
18. Select **DiscoverSim** > **Run Simulation**:

19. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below (note that 64 bit DiscoverSim will show slightly different results).



Click **Run**.

20. The DiscoverSim Output Report displays a histogram, descriptive statistics and process capability indices:



Descriptive Statistics	
Count	10000
Mean	42327.600
StDev	47803.719498
Range	198000.000
Minimum	-48000.000
25th Percentile (Q1)	18000.000
50th Percentile (Median)	40000.000
75th Percentile (Q3)	62000.000
Maximum	150000.000
95.0% CI Mean	41390.55 to 43264.65
95.0% CI Median	40000 to 40000
95.0% CI StDev	47150.3 to 48475.63

Normality Tests	
Anderson-Darling Normality Test	102.804646
p-value (A-D Test)	0.0000
Skewness	0.156488
p-value (Skewness)	0.0000
Kurtosis	-0.429082
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	
Ppl	0.30
Ppk	0.30
Cpm	

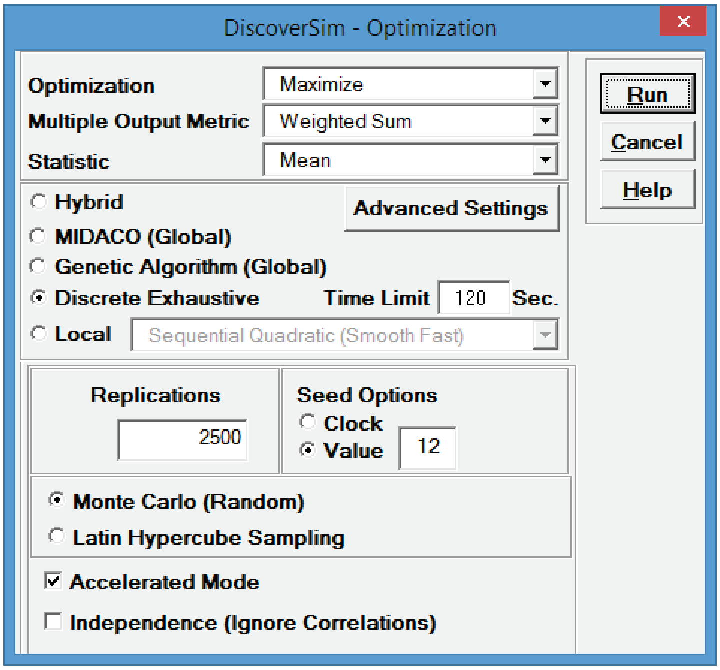
Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	
ppm < LSL	187958.13
ppm Total	187958.13
% > USL	
% < LSL	18.80
% Total (out of spec.)	18.80
% Total (within spec.)	81.20

Actual Performance (Empirical)	
ppm > USL	
ppm < LSL	229200.00
ppm Total	229200.00
% > USL	
% < LSL	22.92
% Total (out of spec.)	22.92
% Total (within spec.)	77.08

The Mean Net Profit is approximately $42,300, which is good, but we also see from the Actual Performance report, we have a 22.9% chance of monthly loss. The standard deviation is also quite high at $47,800.

21. Now we are ready to perform the discrete optimization to maximize the mean of net profit. Select **DiscoverSim > Run Optimization**:

22. Select “Maximize” for **Optimization Goal**, “Weighted Sum” for **Multiple Output Metric** and “Mean” for **Statistic**. Select **Discrete Exhaustive**. Set **Seed Value** to 12 and **Replications** to 2500 in order to replicate the optimization results given below. All other settings will be the defaults as shown:



Tip: Here we are using **Discrete Exhaustive** because this particular problem is not complex, given that there is only one input control variable with 10 discrete levels. The **Time Limit** determines how long **Discrete Exhaustive** will run. A single simulation run is first performed and the computation time is measured. If the total number of discrete combinations * single run time exceed the **Time Limit**, an error message is produced, so you can either increase the Time Limit, or use a different optimization method.

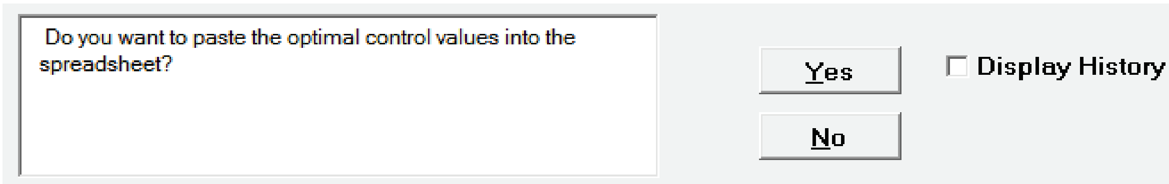
23. Click **Run**.

24. The final optimal parameter value is given as:

```
*****************************************************************
                     Optimal Parameters
*****************************************************************

   Maximize Weighted Sum (Mean) :        59076.800

   Parameter Values:

   Magazines_Produced: Value        60000.0000
```

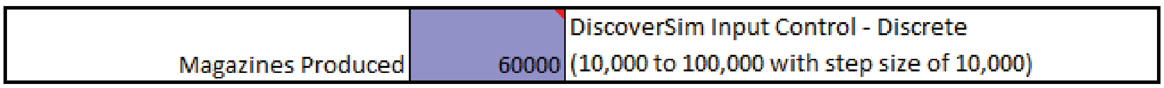
The optimum number of magazines to produce per month is 60,000, resulting in an expected mean profit of $59,077.

25. You are prompted to paste the optimal value into the spreadsheet:



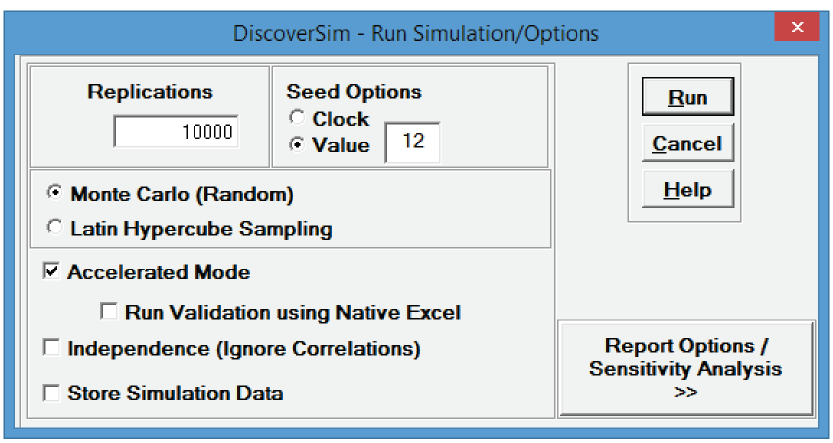
Tip: If you check **Display History**, the complete history of optimization iterations will be displayed in a separate window after clicking **Yes** or **No** to the paste prompt.

Click **Yes**. This sets the input control value to the optimum value of 6 as shown:



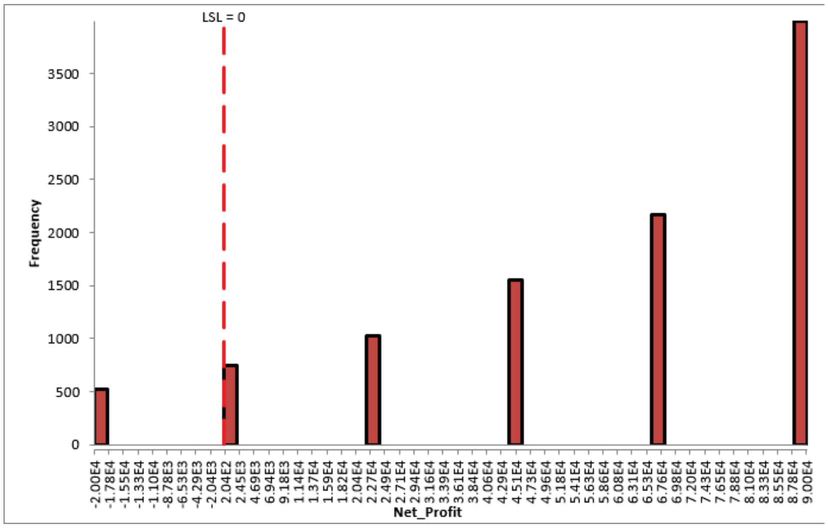
26. Select **DiscoverSim** > **Run Simulation**:

27. Select **Seed Value** and enter "12" as shown, in order to replicate the simulation results given below.



Click **Run**.

28. The DiscoverSim Output Report displays a histogram, descriptive statistics and process capability indices for the optimized setting:



Descriptive Statistics	
Count	10000
Mean	59338.600
StDev	33465.548618
Range	110000.000
Minimum	-20000.000
25th Percentile (Q1)	46000.000
50th Percentile (Median)	68000.000
75th Percentile (Q3)	90000.000
Maximum	90000.000
95.0% CI Mean	58682.61 to 59994.59
95.0% CI Median	68000 to 68000
95.0% CI StDev	33008.12 to 33935.93

Normality Tests	
Anderson-Darling Normality Test	612.940987
p-value (A-D Test)	0.0000
Skewness	-0.889894
p-value (Skewness)	0.0000
Kurtosis	-0.295289
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	
Ppl	0.59
Ppk	0.59
Cpm	

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	
ppm < LSL	38104.00
ppm Total	38104.00
% > USL	
% < LSL	3.81
% Total (out of spec.)	3.81
% Total (within spec.)	96.19

Actual Performance (Empirical)	
ppm > USL	
ppm < LSL	52300.00
ppm Total	52300.00
% > USL	
% < LSL	5.23
% Total (out of spec.)	5.23
% Total (within spec.)	94.77

From the histogram and capability report we see that the expected mean net profit has increased from $42,300 to $59,300. The standard deviation has decreased from $47,800 to $33,500. Looking at the Actual Performance report we see a decrease in the likelihood of a profit loss from 22.9% to 5.2%.

29. Obviously the quality of this prediction will depend on the validity of the custom distribution and does not take into account hidden soft costs like customer dissatisfaction due to lack of availability. If necessary, the model can be refined. Nevertheless, the tool is very useful to help us to determine what our production rate should be in order to maximize net profit.

Case Study 3 – Six Sigma DMAIC Project Portfolio Selection

Introduction: Optimizing DMAIC Project Portfolio to Maximize Cost Savings

This is an example of DiscoverSim optimization used to select Six Sigma DMAIC projects that maximize expected cost savings.

This example model was contributed by:

Dr. Harry Shah, Business Excellence Consulting.

The goal is to select projects that maximize Total Expected Savings subject to a constraint of Total Resources <= 20. Management requires a minimum total project savings of $1 Million (Lower Specification Limit, **LSL** = 1000 $K). We will explore the optimal selection of projects that maximize the Mean Savings and also consider maximizing the Process Capability Index Ppl.

Cost Savings are specified with a Triangular Distribution.

The Probability of Success is modeled using the Bernoulli (Yes/No) Distribution.

"Resources Required" are specified for each project. DiscoverSim Input Controls (Discrete) are used to select the project (0,1).

Expected Project Savings are calculated as the sum of: Cost Savings * Probability of Success (if project is selected) or 0 (if project is not selected).

Summary of DiscoverSim Features Demonstrated in Case Study 3:
• Create Input Distributions (Continuous Triangular) • DiscoverSim Copy/Paste Cell • Create Input Distributions (Discrete Bernoulli) • Create Discrete Input Controls • Specify Constraints • Run Optimization • Run Simulation • Insert DSIM Function

Optimization of Project Portfolio with DiscoverSim

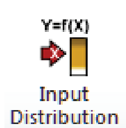
1. Open the workbook **DMAIC Project Selection - Optimization**. DiscoverSim Input Distributions will simulate the variability in the cost savings and probability of success for each project. We will specify cost savings as a Triangular Distribution for each project in cells **G14** to **G22**. The probability of success will be specified at cells **K14 to K22** using the Bernoulli (Yes/No) Distribution.

	B	C	D	E	F	G	I	J	K
13	DMAIC Project Name	Project Cost Savings - Minimum $K	Project Cost Savings - Most Likely (Mode) $K	Project Cost Savings - Maximum $K	Cost Savings Input Distribution Name	DiscoverSim Input Distribution - Cost Savings	Probability of Success	Probability of Success Input Distribution Name	DiscoverSim Input Distribution - Success (0/1)
14	Project_A	100	125	150	Project_A_CS		0.9	Project_A_Pr	
15	Project_B	50	100	150	Project_B_CS		0.8	Project_B_Pr	
16	Project_C	500	600	700	Project_C_CS		0.8	Project_C_Pr	
17	Project_D	100	150	200	Project_D_CS		0.9	Project_D_Pr	
18	Project_E	1500	2000	2500	Project_E_CS		0.5	Project_E_Pr	
19	Project_F	200	250	300	Project_F_CS		0.8	Project_F_Pr	
20	Project_G	300	400	500	Project_G_CS		0.8	Project_G_Pr	
21	Project_H	700	750	800	Project_H_CS		0.7	Project_H_Pr	
22	Project_I	300	350	400	Project_I_CS		0.8	Project_I_Pr	

2. DiscoverSim Input Controls (Discrete) will be used to "select" the project (0,1) and specified in cells **O14** to **O22**.

	M	N	O
13	Resources Required	Input Control Name	DiscoverSim Input Control - Project Selected (0/1)
14	1	Project_A_Select	
15	2	Project_B_Select	
16	5	Project_C_Select	
17	2	Project_D_Select	
18	10	Project_E_Select	
19	2	Project_F_Select	
20	4	Project_G_Select	
21	5	Project_H_Select	
22	3	Project_I_Select	

3. Click on cell **G14** to specify the "Cost Savings" Input Distribution for Project_A. Select **DiscoverSim > Input Distribution**:



4. Select **Triangular Distribution**.

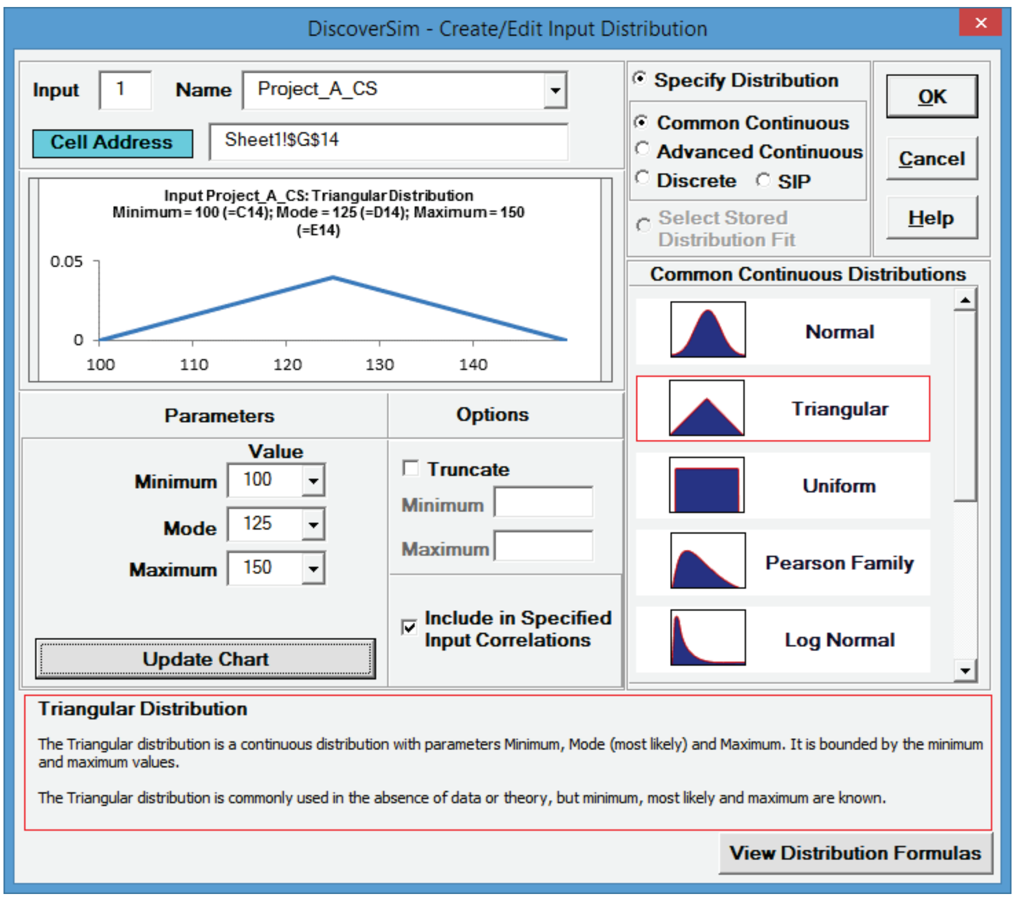
5. Click the input name cell reference and specify cell **F14** containing the input **Name** "Project_A_CS" (denoting "Project A Cost Savings"). After specifying a cell reference, the dropdown symbol changes from to .

6. Click the **Minimum** parameter cell reference and specify cell **C14** containing the minimum parameter value = 100 ($K).

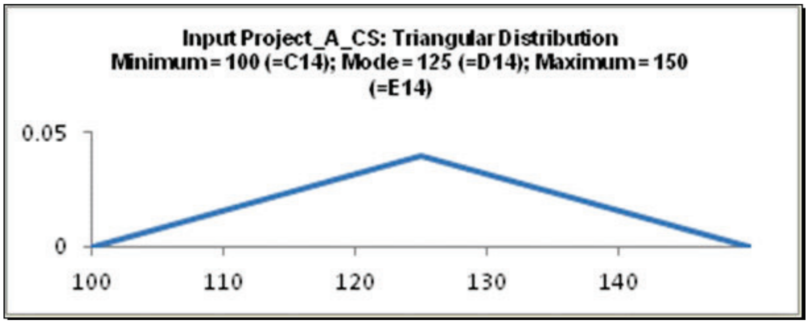
7. Click the **Mode** parameter cell reference and specify cell **D14** containing the mode (most likely) parameter value = 125 ($K).

8. Click the **Maximum** parameter cell reference and specify cell **E14** containing the maximum parameter value = 150 ($K).

9. Click **Update Chart** to view the **Triangular Distribution** as shown:



10. Click **OK**. Hover the cursor on cell **G14** to view the DiscoverSim graphical comment showing the distribution and parameter values:

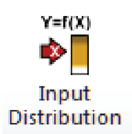


11. Click on cell **G14**. Click the **DiscoverSim Copy Cell** menu button  (**Do not use Excel's Copy – it will not work!**).

12. Select cells **G15:G22**. Now click the **DiscoverSim Paste Cell** menu button  (**Do not use Excel's Paste – it will not work!**).

13. Review the input comments in cells **G15** to **G22**.

14. Click on cell **K14** to specify the "Probability of Success" Input Distribution for Project_A. Select **DiscoverSim > Input Distribution**:

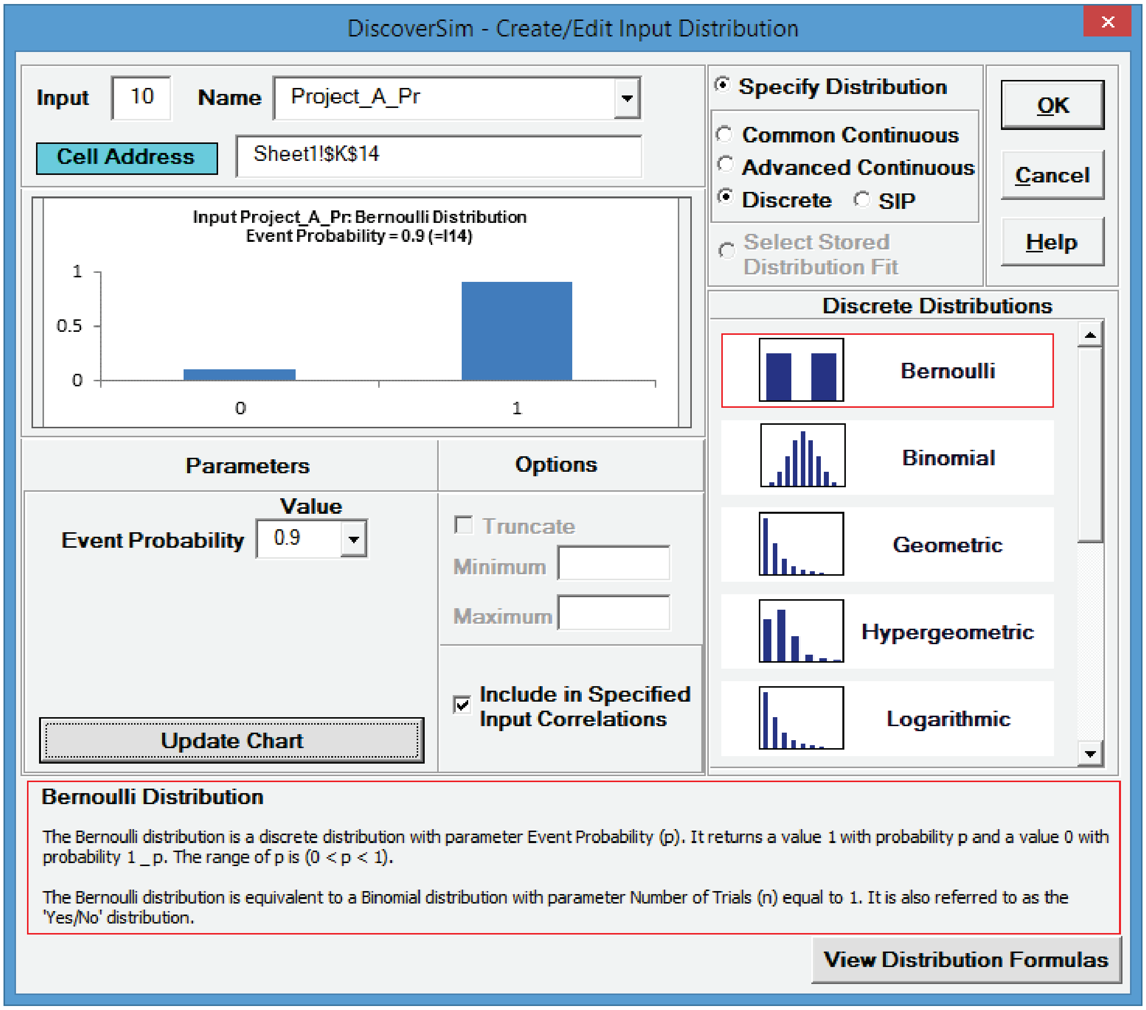


15. Select **Discrete** and **Bernoulli Distribution**. Bernoulli is also sometimes referred to as the "Yes/No" distribution.

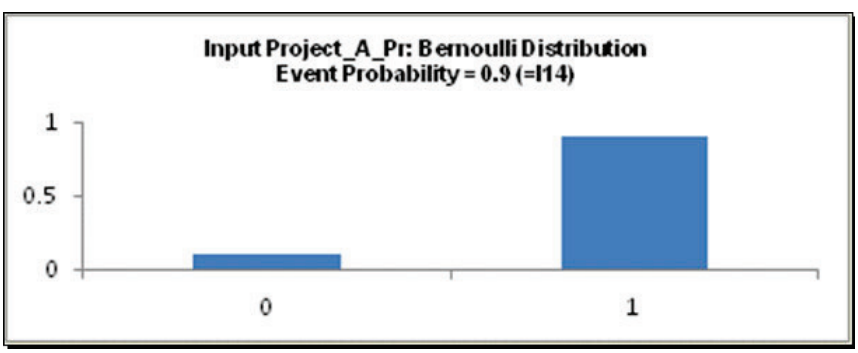
16. Click the input **Name** cell reference  and specify cell **J14** containing the input name "Project_A_Pr" (denoting "Project A Probability of Success").

17. Click the **Event Probability** parameter cell reference  and specify cell **I14** containing the probability parameter value = 0.9.

18. Click **Update Chart** to view the Bernoulli distribution as shown:



19. Click **OK**. Hover the cursor on cell **K14** to view the DiscoverSim graphical comment showing the distribution and parameter values:



20. Click on cell **K14**. Click the **DiscoverSim Copy Cell** menu button 

 (**Do not use Excel's Copy** – **it will not work!**).

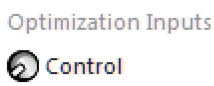
21. Select cells **K15:K22**. Now click the **DiscoverSim Paste Cell** menu button 

 (**Do not use Excel's Paste** – **it will not work!**).

22. Review the input comments in cells **K15** to **K22**.

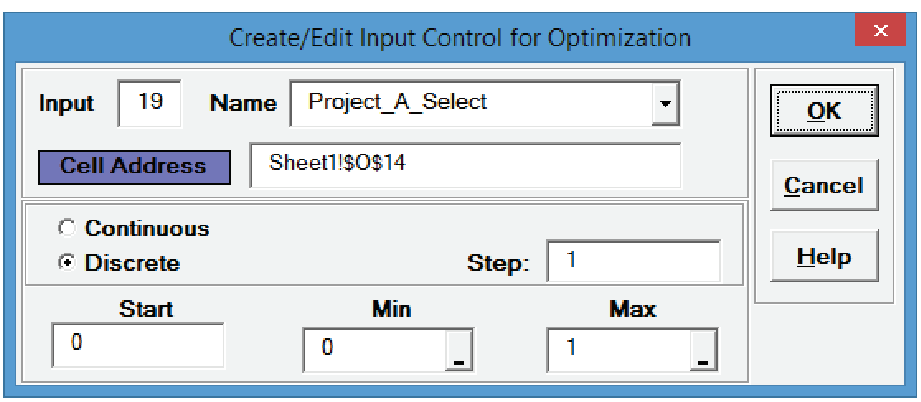
23. Now we will add Discrete Input Controls (also referred to as "Decision Variables"). These will be used to "select" the project (0 denotes project not selected, 1 denotes project selected). DiscoverSim will use stochastic global optimization to determine which projects to select and which not to select in order to maximize total cost savings. A constraint will also be added to ensure that the total number of resources is <= 20.

24. Click on cell **O14**. Select **Control**:

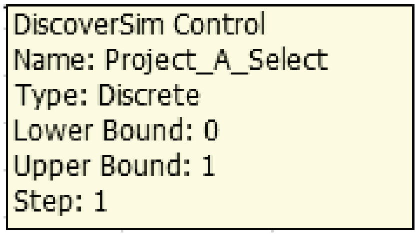


25. Click input **Name** cell reference  and specify cell **N14** containing the input control name "Project_A_Select."

26. Select **Discrete** (this input control setting will consider only integer values). Use the default values as shown:



27. Click **OK**. Hover the cursor on cell **O14** to view the comment displaying the Input Control settings:

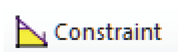


28. Click on cell **O14**. Click the **DiscoverSim Copy Cell** menu button  (**Do not use Excel's Copy – it will not work!**).

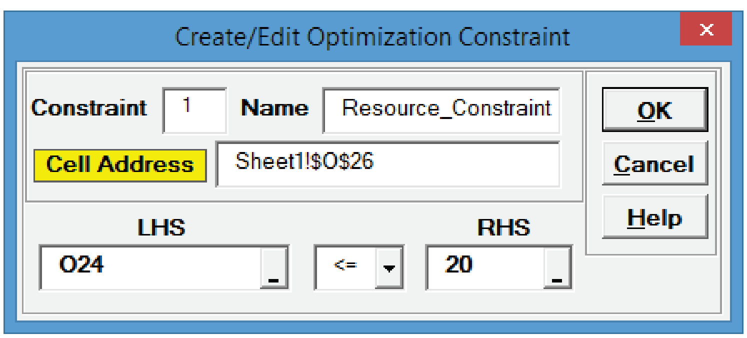
29. Select cells **O15:O22**. Click the **DiscoverSim Paste Cell** menu button  (**Do not use Excel's Paste – it will not work!**).

30. Review the input control comments in cells **O15** to **O22**.

31. Now we will add the constraint. Click on cell **O26**. Select **Constraint**:

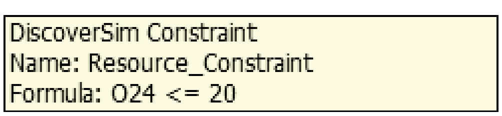


32. Enter the Constraint **Name** as "Resource_Constraint." Enter **O24** in the "Left Hand Side" (**LHS**) or click the **LHS** cell reference and select **O24**. Select **<=** . Enter 20 in the "Right Hand Side" (**RHS**).



Note: Cell **O24** contains the Excel formula**:** =SUMPRODUCT(O14:O22,M14:M22). This sums the product of resources*selected(0,1). In other words, the sum of selected project resources.

33. Click **OK**. Review the comment at cell **O26**.

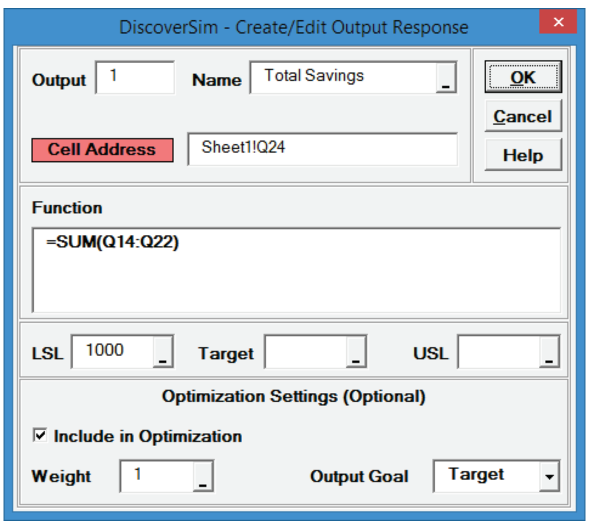


The cell display is "TRUE" since the initial value for **O24** is zero which obviously satisfies the constraint.

34. Now we will specify the output. Click on cell **Q24**. The cell contains the Excel formula for the sum of expected savings as the sum of: Cost Savings * Probability of Success (if project is selected) or 0 (if project is not selected).

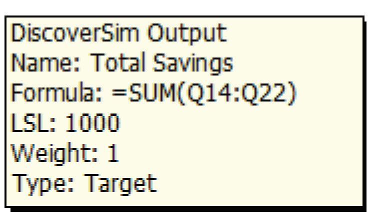
35. Select **DiscoverSim** > **Output Response**:

36. Enter the output **Name** as "Total Savings." Enter the Lower Specification Limit (**LSL**) as 1000. The **Weight** and **Output Goal** settings are used only for multiple response optimization, so do not need to be modified in this example.



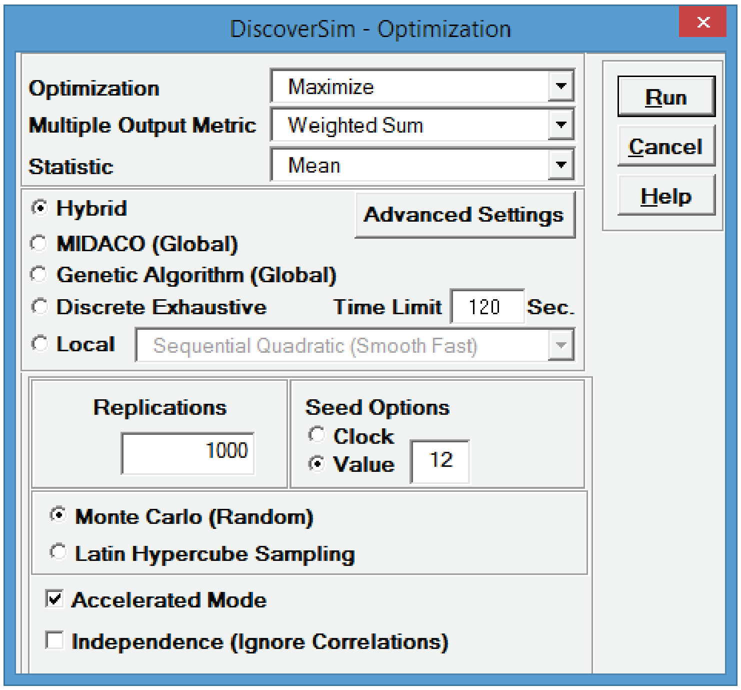
Click **OK**.

37. Hover the cursor on cell **Q24** to view the DiscoverSim Output information.



38. Now we are ready to perform the optimization to select projects. Select **DiscoverSim > Run Optimization**:

39. Select “Maximize” for **Optimization Goal** and “Mean” for **Statistic**. Select **Seed Value** and enter “12” in order to replicate the optimization results given below. Change **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:



Note: Maximize Mean may not produce the lowest possible standard deviation, so we will consider Maximize Ppl later.

We will use the **Hybrid** optimization method which may require more time to compute, but is powerful to solve complex optimization problems. Since this is a small discrete problem, Discrete Exhaustive will be used. If there were more projects, say 20, then MIDACO would have been used.

40. Click **Run**.

41. The final optimal parameter values are given as:

```
******************************************************************
                     Optimal Parameters
******************************************************************

   Maximize Weighted Sum (Mean) :        2048.2941

   Parameter Values:

   Project_A_Select: Value            0.0000
   Project_B_Select: Value            0.0000
   Project_C_Select: Value            0.0000
   Project_D_Select: Value            0.0000
   Project_E_Select: Value            1.0000
   Project_F_Select: Value            1.0000
   Project_G_Select: Value            0.0000
   Project_H_Select: Value            1.0000
   Project_I_Select: Value            1.0000
```

Projects E, F, H and I have been selected giving a predicted mean cost savings of 2004 ($K). Intuitively, these would be the projects one would select given that they have the highest expected values per resource. However, a simple "pick the winner" approach fails to take into account the variability and likelihood of meeting the minimum total cost saving requirement of 1000 $K.

42. The constraint equation evaluated at this optimum has been satisfied (amount violated = 0):

```
 Constraint Equations evaluated at the optimum:

    Constraint       Type         Amount Violated

       1              le              0.0000
```

Project E requires 10 resources, F requires 2, H requires 5 and I requires 3 for a total of 20, satisfying our constraint that the total resources be less than or equal to 20.

43. You are prompted to paste the optimal value into the spreadsheet:

Do you want to paste the optimal control values into the spreadsheet?

Yes

No

☐ Display History

Tip: If you check **Display History**, the complete history of optimization iterations will be displayed in a separate window after clicking **Yes** or **No** to the paste prompt.

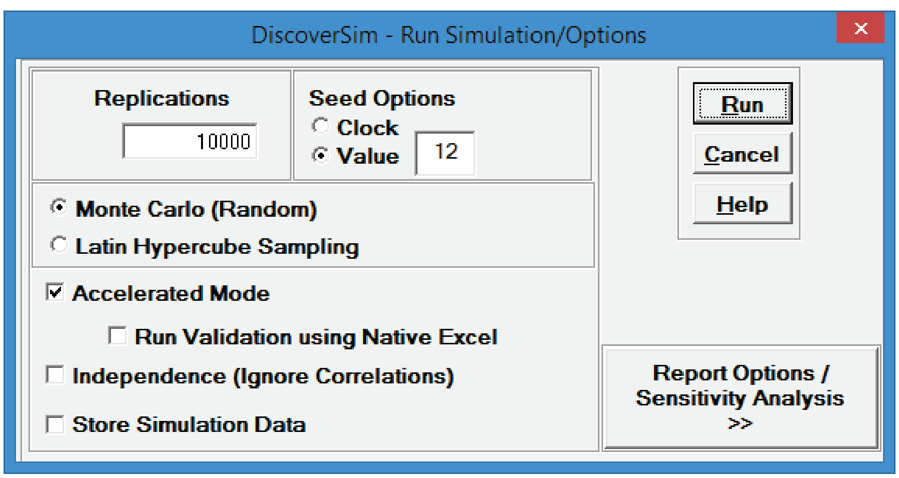
Click **Yes**. This sets the Input Control values to the optimum values as shown:

Resources Required	Input Control Name	DiscoverSim Input Control - Project Selected (0/1)	Expected Project Savings
1	Project_A_Select	0	0
2	Project_B_Select	0	0
5	Project_C_Select	0	0
2	Project_D_Select	0	0
10	Project_E_Select	1	1000
2	Project_F_Select	1	200
4	Project_G_Select	0	0
5	Project_H_Select	1	525
3	Project_I_Select	1	280

Total Resources for Selected Projects:	20	2005	DiscoverSim Output: Total Expected Savings

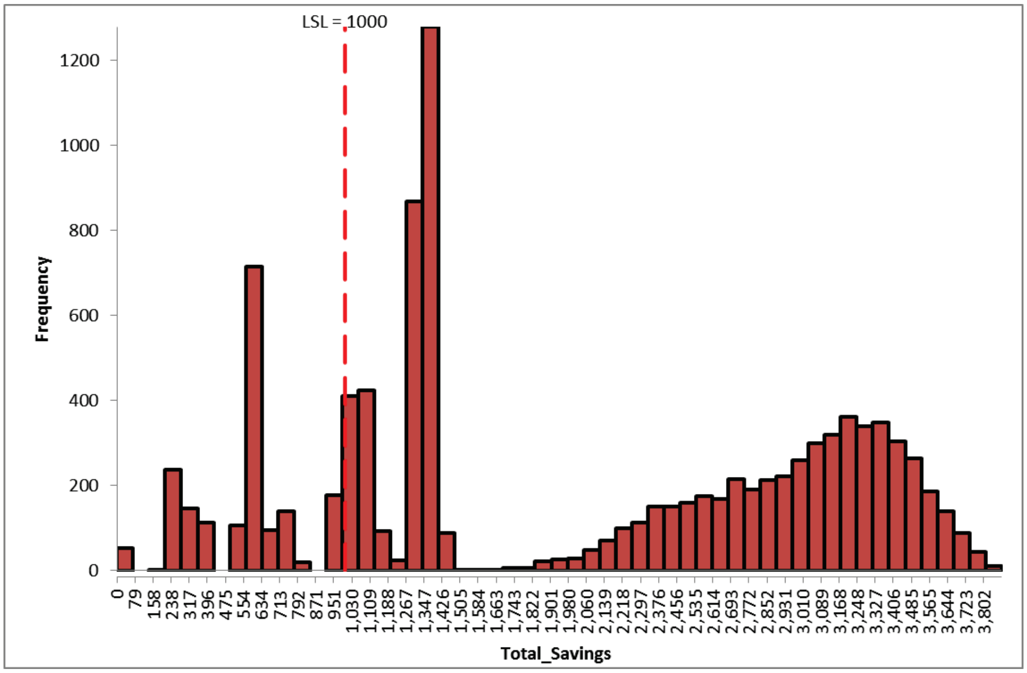
44. Select **DiscoverSim** > **Run Simulation**:

45. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below.



Click **Run**.

46. The DiscoverSim Output Report displays a histogram, descriptive statistics and process capability indices:



Descriptive Statistics	
Count	10000
Mean	1997.361
StDev	1084.518900
Range	3893.202
Minimum	0.000
25th Percentile (Q1)	1095.990
50th Percentile (Median)	1438.503
75th Percentile (Q3)	3076.770
Maximum	3893.202
95.0% CI Mean	1976.102 to 2018.62
95.0% CI Median	1408.719 to 1999.282
95.0% CI StDev	1069.695 to 1099.763

Normality Tests	
Anderson-Darling Normality Test	369.457322
p-value (A-D Test)	0.0000
Skewness	0.024130384
p-value (Skewness)	0.3243
Kurtosis	-1.472
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	
Ppl	0.31
Ppk	0.31
Cpm	

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	
ppm < LSL	178881.87
ppm Total	178881.87
% > USL	
% < LSL	17.89
% Total (out of spec.)	17.89
% Total (within spec.)	82.11

Actual Performance (Empirical)	
ppm > USL	
ppm < LSL	195000.00
ppm Total	195000.00
% > USL	
% < LSL	19.50
% Total (out of spec.)	19.50
% Total (within spec.)	80.50

From the histogram and capability report we see that the expected mean total savings are good (1997 $K), but the standard deviation is very large (1085 $K). Looking at the Actual Performance report we see that there is a 19.5% chance of failing to meet the management requirement of 1000 $K.

47. Now we will rerun the optimization to select projects, but this time, we will maximize the process capability index Ppl. (Ppl is used since we only have a lower specification limit). This will simultaneously maximize the mean and minimize the standard deviation:

    Ppl = (Mean – LSL)/(3*StdDev).

    An equivalent alternative would be to minimize Calculated DPM.

    Note: A nonparametric alternative statistic would be %Ppl (percentile Ppl):

    %Ppl = (Median – LSL)/(Median - 0.135 Percentile).

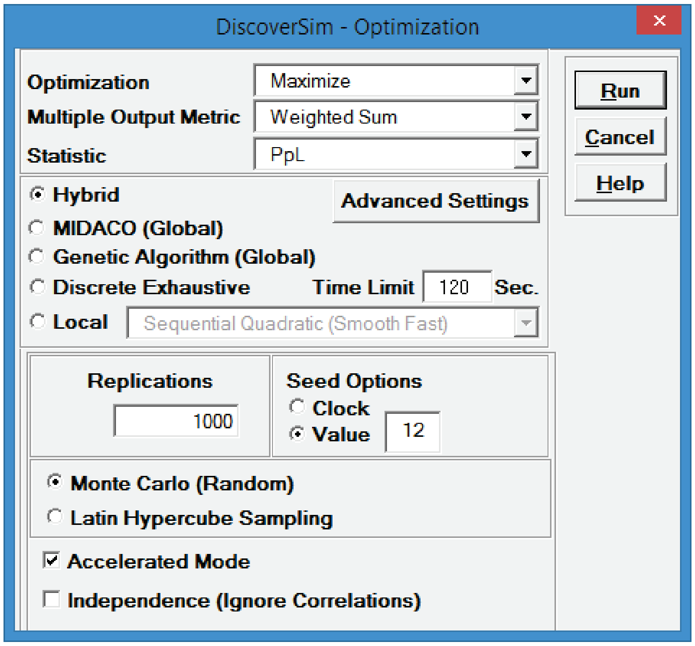
    This is robust to the assumption of normality. Another robust statistic would be Actual DPM. With either of these we would recommend increasing the number of replications to 10,000 to improve the precision of the statistic, but this would require significantly more computation time for optimization.

48. Reset the project selection by changing the Input Control values to 0:

Resources Required	Input Control Name	DiscoverSim Input Control - Project Selected (0/1)
1	Project_A_Select	0
2	Project_B_Select	0
5	Project_C_Select	0
2	Project_D_Select	0
10	Project_E_Select	0
2	Project_F_Select	0
4	Project_G_Select	0
5	Project_H_Select	0
3	Project_I_Select	0

49. Select **DiscoverSim** > **Run Optimization**:

50. Select “Maximize” for **Optimization Goal**, “Weighted Sum” for **Multiple Output Metric** and “PpL” for **Statistic**. Select **Seed Value** and enter “12,” in order to replicate the optimization results given below. Set **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:



Click **Run**.

51. The final optimal parameter values are given as:

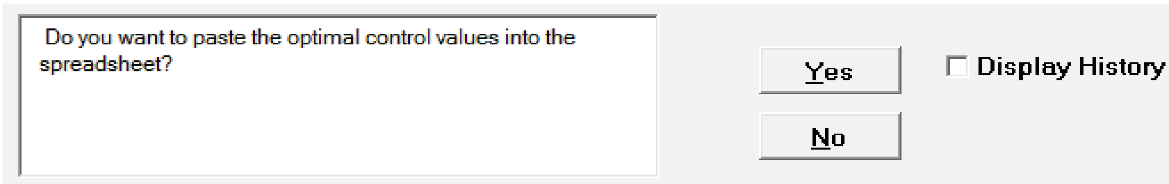
```
****************************************************************
                     Optimal Parameters
****************************************************************

   Maximize Weighted Sum (PpL) :        0.61581359

   Parameter Values:

   Project_A_Select: Value            1.0000
   Project_B_Select: Value            0.0000
   Project_C_Select: Value            1.0000
   Project_D_Select: Value            0.0000
   Project_E_Select: Value            0.0000
   Project_F_Select: Value            1.0000
   Project_G_Select: Value            1.0000
   Project_H_Select: Value            1.0000
   Project_I_Select: Value            1.0000
```

Projects A, C, F, G, H, and I have been selected giving a predicted Ppl of 0.62.

Note: Project E, previously selected to maximize the mean, was not selected when the goal was to maximize the process capability, Ppl.

The constraint requirement that the total number of resources be <= 20 is satisfied with this selection.

52. You are prompted to paste the optimal values into the spreadsheet:



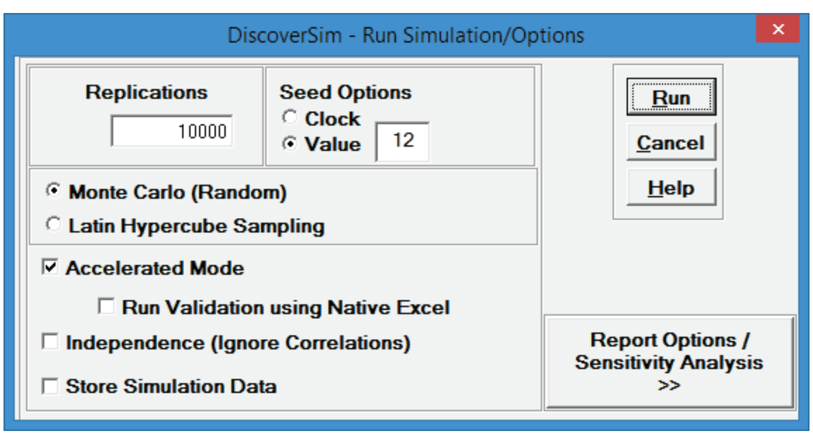
Click **Yes**. This sets the Input Control values to the optimum values as shown:

Resources Required	Input Control Name	DiscoverSim Input Control - Project Selected (0/1)	Expected Project Savings
1	Project_A_Select	1	112.5
2	Project_B_Select	0	0
5	Project_C_Select	1	480
2	Project_D_Select	0	0
10	Project_E_Select	0	0
2	Project_F_Select	1	200
4	Project_G_Select	1	320
5	Project_H_Select	1	525
3	Project_I_Select	1	280

Total Resources for Selected Projects:	20	1917.5	DiscoverSim Output: Total Expected Savings

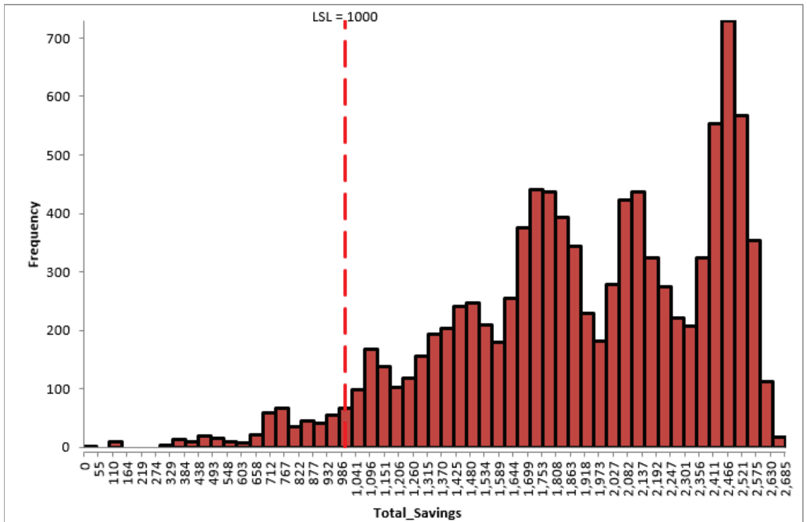
53. Select **DiscoverSim** > **Run Simulation**:

54. Select **Seed Value** and enter "12" as shown, in order to replicate the simulation results given below.



Click **Run**.

55. The DiscoverSim Output Report displays a histogram, descriptive statistics and process capability indices:



Descriptive Statistics	
Count	10000
Mean	1915.177
StDev	489.531624
Range	2685.020
Minimum	0.000
25th Percentile (Q1)	1605.459
50th Percentile (Median)	1955.173
75th Percentile (Q3)	2373.182
Maximum	2685.020
95.0% CI Mean	1905.581 to 1924.773
95.0% CI Median	1928.605 to 1983.519
95.0% CI StDev	482.8403 to 496.4123

Normality Tests	
Anderson-Darling Normality Test	122.423168
p-value (A-D Test)	0.0000
Skewness	-0.612190
p-value (Skewness)	0.0000
Kurtosis	-0.203514
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	
Ppu	
Ppl	0.62
Ppk	0.62
Cpm	

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	
ppm < LSL	30776.96
ppm Total	30776.96
% > USL	
% < LSL	3.08
% Total (out of spec.)	3.08
% Total (within spec.)	96.92

Actual Performance (Empirical)	
ppm > USL	
ppm < LSL	43500.00
ppm Total	43500.00
% > USL	
% < LSL	4.35
% Total (out of spec.)	4.35
% Total (within spec.)	95.65

Note that the mean total savings have been reduced somewhat from 2011 to 1915 $K, however, the standard deviation has been reduced from 1080 to 490 $K and the chance of failing to meet the management requirement of 1000 $K has been reduced from 18.9% to 4.35% (actual performance).

Considering the two selection scenarios, one can present the risks to management and make the case for either option: take the risk and maximize the mean with potential larger savings, or choose the safer projects that result in less variation and significantly improve the chance of meeting a minimum savings of 1 million.

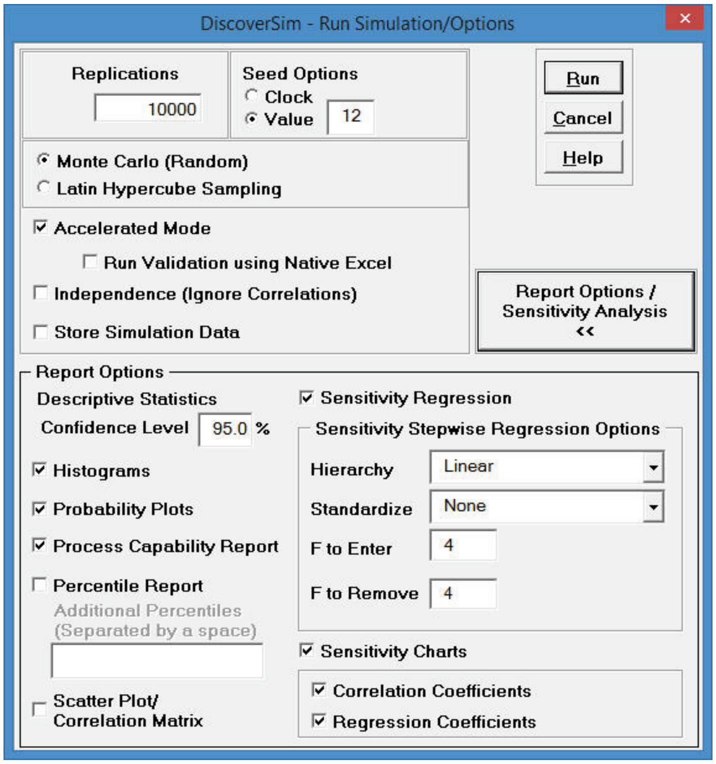
56. Clearly, Project E is the key project to consider in this analysis. The impact of this project can be confirmed with a sensitivity analysis.

57. Set all of the Input Control values to 1 as shown:

Resources Required	Input Control Name	DiscoverSim Input Control - Project Selected (0/1)
1	Project_A_Select	1
2	Project_B_Select	1
5	Project_C_Select	1
2	Project_D_Select	1
10	Project_E_Select	1
2	Project_F_Select	1
4	Project_G_Select	1
5	Project_H_Select	1
3	Project_I_Select	1

58. Select **DiscoverSim** > **Run Simulation**:

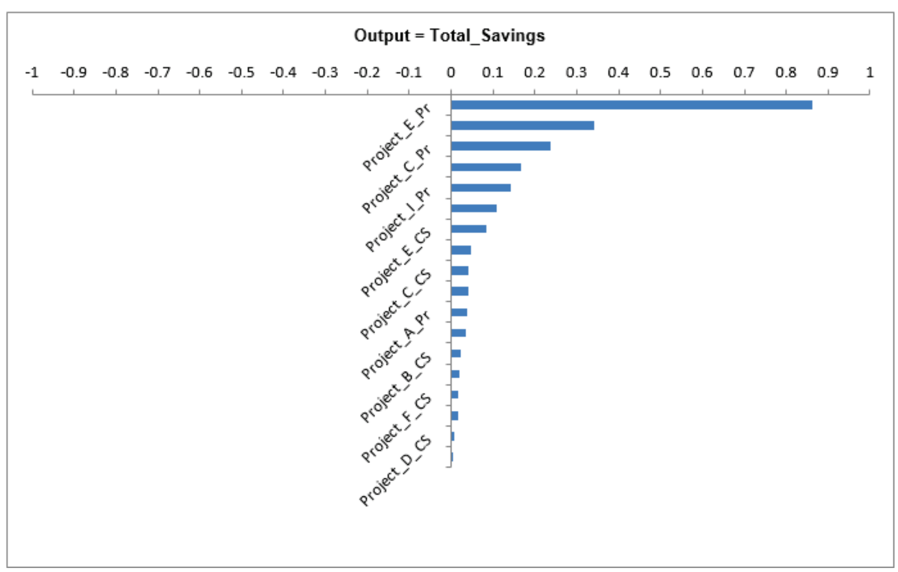
59. Click **Report Options/Sensitivity Analysis**. Check **Sensitivity Regression Analysis**, **Sensitivity Charts** - **Correlation Coefficients** and **Regression Coefficients**. Select **Seed Value** and enter "12" as shown, in order to replicate the simulation results given below.



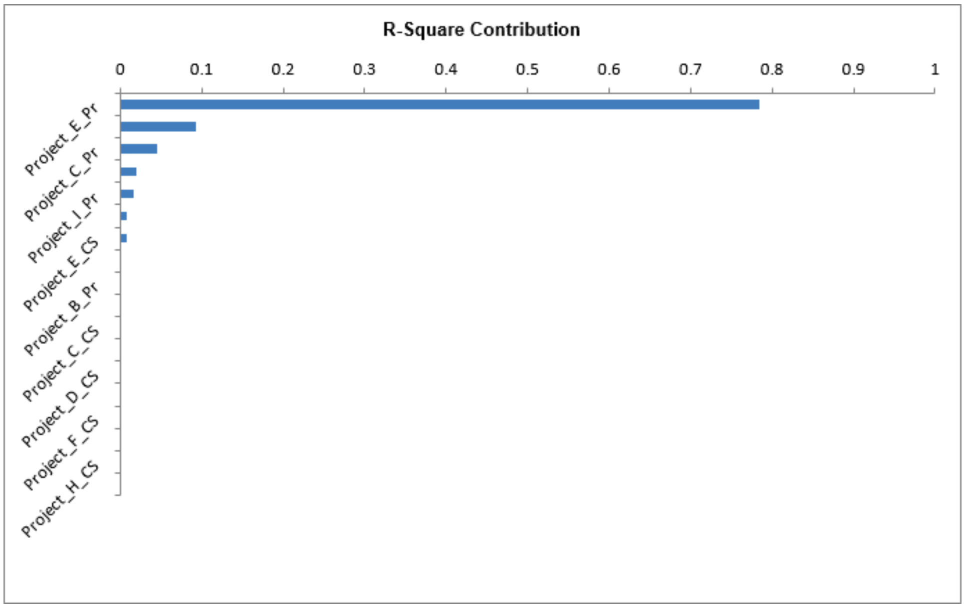
Click **Run**.

60. Click on the **Sensitivity Correlations** sheet:

Inputs	Spearman Rank
Project_E_Pr	0.861383953
Project_H_Pr	0.34134467
Project_C_Pr	0.236070465
Project_G_Pr	0.168302066
Project_I_Pr	0.142211752
Project_F_Pr	0.108907076
Project_E_CS	0.083449953
Project_D_Pr	0.046059181
Project_C_CS	0.042220236
Project_B_Pr	0.040413659
Project_A_Pr	0.039612761
Project_G_CS	0.036003476
Project_B_CS	0.022871417
Project_H_CS	0.020640952
Project_F_CS	0.017517096
Project_I_CS	0.016135851
Project_D_CS	0.007707907
Project_A_CS	0.000741024

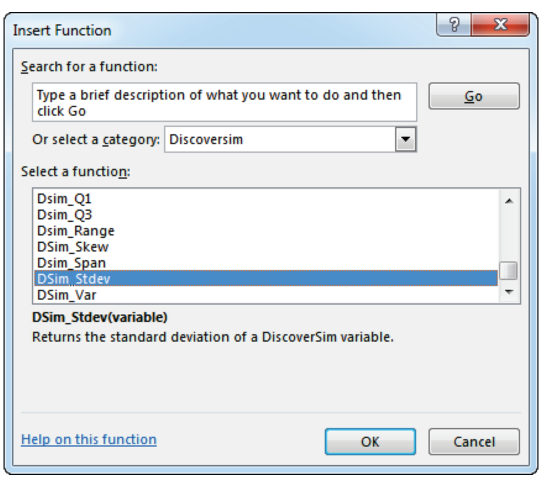


61. Click on the **Sensitivity Regression** sheet to view the R-Square percent contribution to variation:



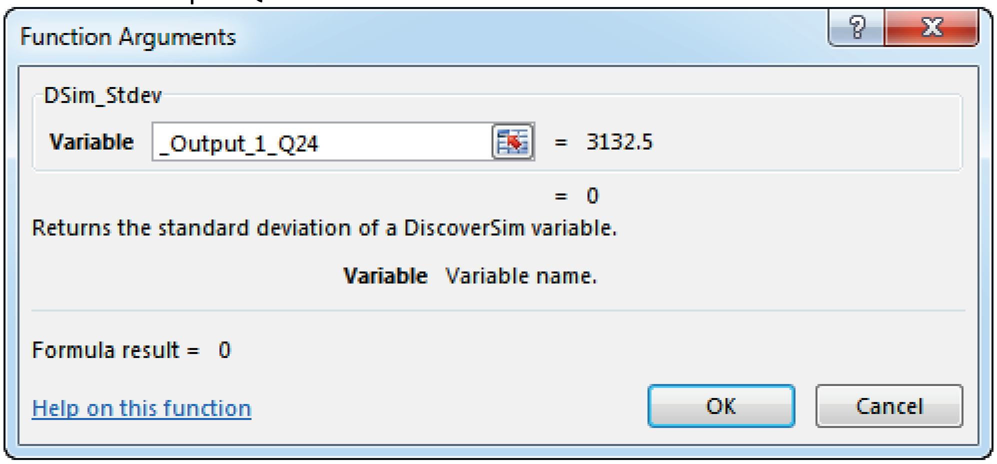
The sensitivity charts confirm that Project E is the dominant "X" factor. This is a high impact project, but is also a difficult project with a probability of success of only 0.5 and a large variation (min = 2000, max = 2500). If possible, this project might be split into smaller, lower risk (higher probability of success) projects with less variation.

62. **Optional Insert DSIM Function:** As an alternative to Maximize CpL as shown above, we could also Maximize Mean Savings subject to an added constraint that the Standard Deviation of Savings be < 500. Click on cell **Q27** and select **DiscoverSim > DSIM Function**:



Scroll down to select *DSIM_Stdev*. Click **OK**.

63. Select the Output **Q24** as the DSIM Variable.



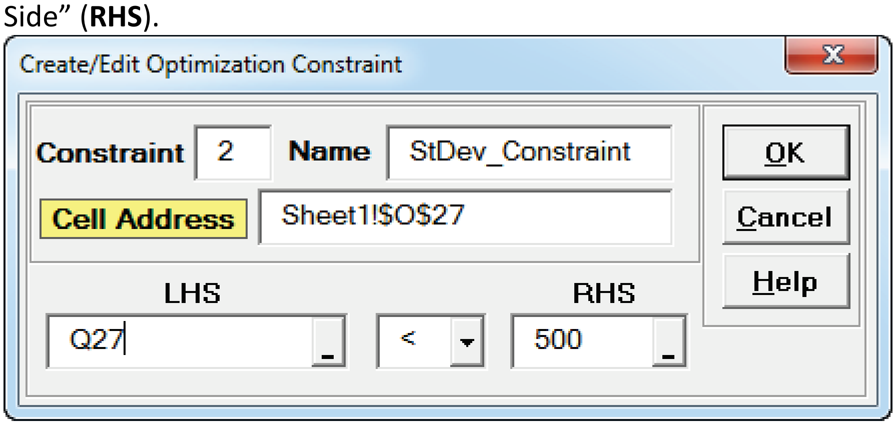
The DiscoverSim Range Name for **Q24** is _Output_1_Q24. Click **OK**.

64. The DSIM function formula is now in cell **Q24**:

    =DSim_Stdev(_Output_1_Q24)

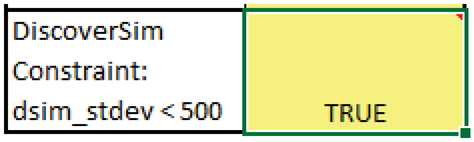
65. Now we will add a constraint at cell **O27** that references the DSIM function. Click on cell **O27**. Select **Constraint**:

66. Enter the Constraint **Name** as "StDev_Constraint." Enter **Q27** in the "Left Hand Side" (**LHS**) or click the **LHS** cell reference and select **Q27**. Select **<** . Enter 500 in the "Right Hand Side" (**RHS**).



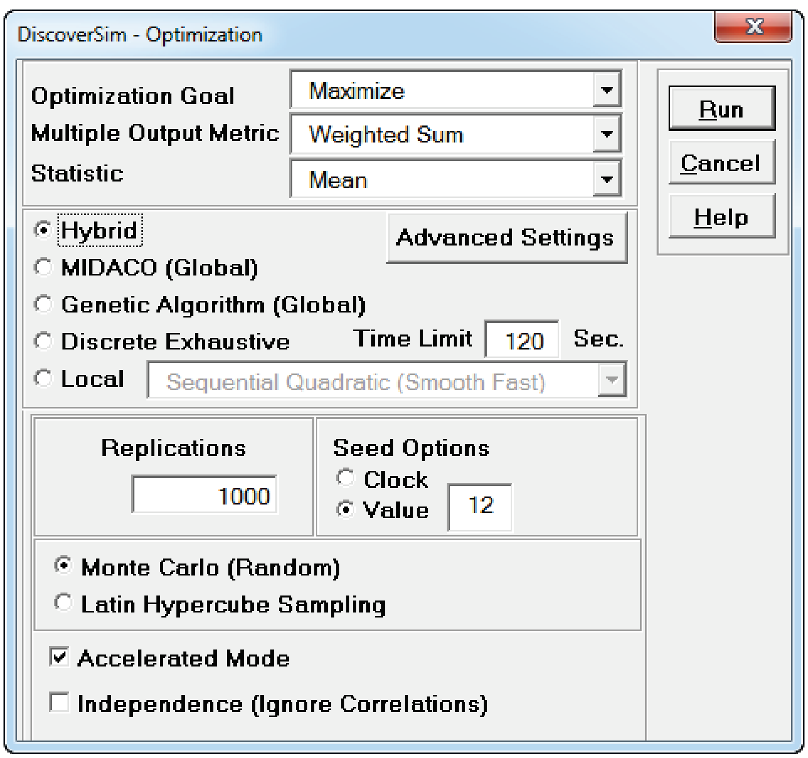
67. Click **OK**. Review the comment at cell **O27:**

The cell display is "TRUE" since the initial value for **Q27** is zero which obviously satisfies the constraint.



68. With all of the Input Control values set to 1, select **DiscoverSim** > **Run Optimization**:

69. Select "Maximize" for **Optimization Goal**, "Weighted Sum" for **Multiple Output Metric** and "Mean" for **Statistic**. Select **Seed Value** and enter "12," in order to replicate the optimization results given below. Set **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:



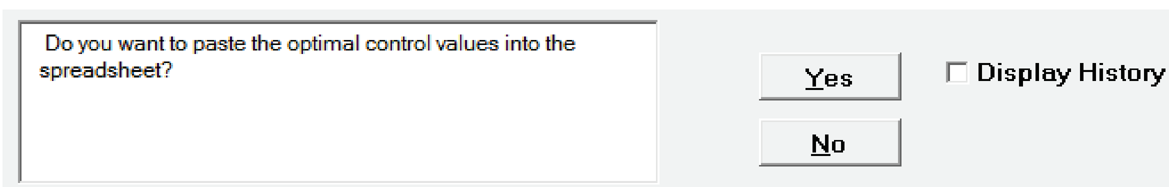
Click **Run**.

70. The final optimal parameter values are given as:

```
****************************************************************
                   Optimal Parameters
****************************************************************

   Maximize Weighted Sum (Mean) :         1899.9301

   Parameter Values:

   Project_A_Select: Value          1.0000
   Project_B_Select: Value          0.0000
   Project_C_Select: Value          1.0000
   Project_D_Select: Value          0.0000
   Project_E_Select: Value          0.0000
   Project_F_Select: Value          1.0000
   Project_G_Select: Value          1.0000
   Project_H_Select: Value          1.0000
   Project_I_Select: Value          1.0000

   Constraint Equations evaluated at the optimum
   Constraint Methodology:  Evaluate

      Constraint       Type        Amount Violated

         1              le            0.0000
         2              lt            0.0000
```

Projects A, C, F, G, H, and I have been selected giving a predicted Mean of 1899.9. These are exactly the same projects that were selected above with maximize PpL.

71. You are prompted to paste the optimal values into the spreadsheet:



Since we have already analyzed this project selection, we will not pursue this further. Click **No**.

72. The use of **Insert DSIM Function** as shown provides flexibility in specifying the optimization objective function. For example in a multiple output scenario, you can Maximize the Mean of Output 1, with a constraint that Output 2 Ppk process capability must be > 1.

Case Study 4 – Catapult Variation Reduction

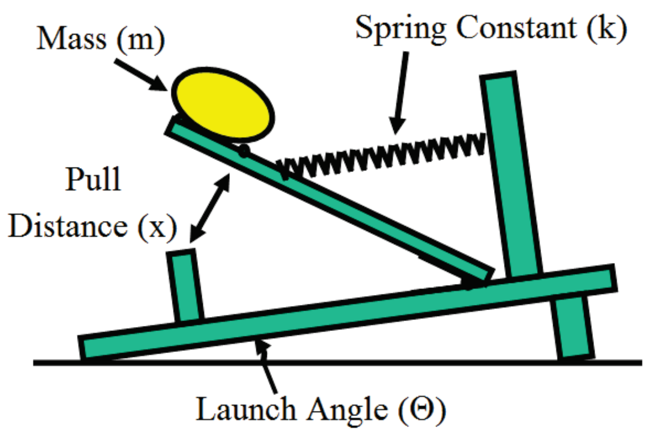
Introduction: Optimizing Catapult Distance Firing Process

This is an example of DiscoverSim stochastic optimization for catapult distance variation reduction, adapted from:

John O'Neill, Sigma Quality Management.

This example is used with permission of the author.

The figure below is a simplified view of the catapult:



The target catapult firing distance, Y = 50 +/- 2.5 feet.

To keep the model simple, only four factors (X's) are included:

1. **Spring Constant** (k) – The spring constant is the force/foot required to pull back the catapult arm (in lbf/ft.). The initial parameters for this factor: normally distributed, mean of 47.3, standard deviation of 0.1.

2. **Pull Distance** (x) – This is the distance in feet the arm is pulled back to launch the mass. The initial parameters for this factor: normally distributed, mean of 6.0, standard deviation of 0.1.

3. **Mass** (m) – This is the mass of the object in slugs (lbf/ft/sec^2). The initial parameters for this factor: normally distributed, mean of 0.5, standard deviation of 0.01.

4. **Launch Angle** (θ) – This is the angle in degrees to the horizontal at which the mass leaves the catapult. The initial parameters for this factor: normally distributed, mean of 35, standard deviation of 3.

   Note: The degrees must be converted into radians for input to the Excel sine and cosine functions.

The Y = f(X) relationship is derived from conservation of energy and basic time, distance, acceleration relationships. Note that air resistance and other “real world” factors have been ignored to keep the model simple. Standard Gravity (g) is 32.174 ft/sec^2:

$$y = \frac{kx^2}{mg} \sin\theta \cos\theta$$

In this study we will use DiscoverSim to help us answer the following questions:

- What is the predicted process capability with these nominal settings?
- What are the key X variables that influence catapult firing distance Y?
- Can we adjust the nominal settings of X to reduce the transmitted variation in Y, thereby making the distance response robust to the variation in inputs?
- Can we further reduce the variation of the key input X’s in order to achieve an acceptable process capability?

Summary of DiscoverSim Features Demonstrated in Case Study 4:
• Create Input Distributions (Continuous Normal) • DiscoverSim Copy/Paste Cell • Run Simulation with Sensitivity Chart of Correlation Coefficients • Create Continuous Input Controls for Parameter Optimization • Run Optimization

Catapult Simulation and Optimization with DiscoverSim

1. Open the workbook **Catapult Model**. DiscoverSim Input Distributions will simulate the variability in the catapult input factor X's. We will specify the Normal Distribution for each X in cells **D30** to **D33**. The output, Distance, will be specified at cell **D36** using the formula given above.

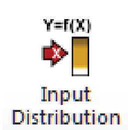
Factor	Units	DiscoverSim Input Distribution
Spring Constant (k)	lb/ft	47.3
Pull Distance (x)	feet	6
Mass (m)	lbf/ft/sec^2	0.5
Launch Angle (θ)	degrees	35

	Output
Distance, Y	49.73297056

2. The Input Distribution parameters are given in the following table:

DiscoverSim Input Distribution Parameters					
DiscoverSim Input Distribution Name	**Distribution**	**Parameter 1 Name**	**Parameter 1 Value**	**Parameter 2 Name**	**Parameter 2 Value**
Spring_Constant	Normal	Mean	47.3	Standard Deviation	0.1
Pull_Distance	Normal	Mean	6	Standard Deviation	0.1
Mass	Normal	Mean	0.5	Standard Deviation	0.01
Launch_Angle	Normal	Mean	35	Standard Deviation	3

3. Click on cell **D30** to specify the Input Distribution for X1, Spring Constant. Select **DiscoverSim > Input Distribution:**



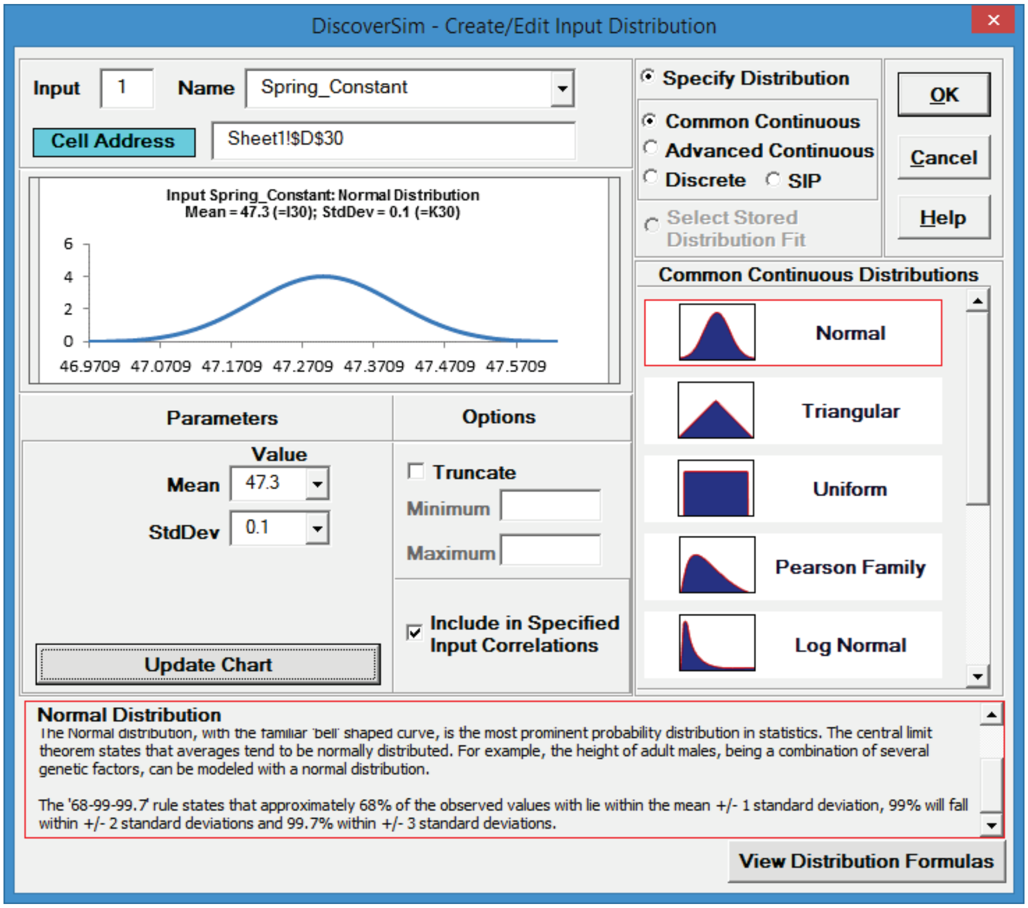
4. We will use the selected default **Normal Distribution**.

5. Click input **Name** cell reference and specify cell **F30** containing the input name "Spring_Constant". After specifying a cell reference, the dropdown symbol changes from to .

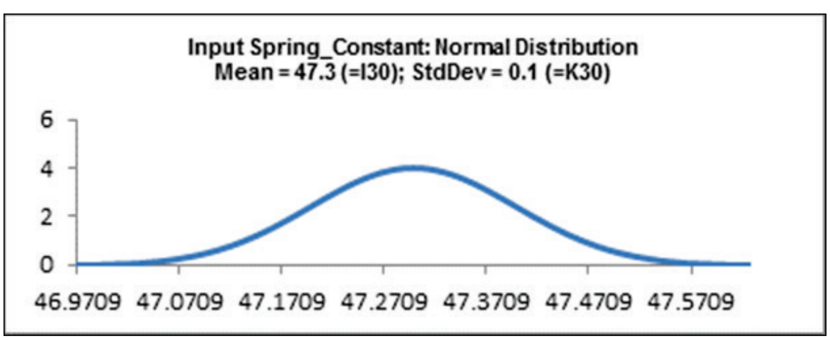
6. Click the **Mean** parameter cell reference and specify cell **I30** containing the mean parameter value = 47.3.

7. Click the **StdDev** parameter cell reference and specify cell **K30** containing the Standard Deviation parameter value = 0.1.

8. Click **Update Chart** to view the Normal Distribution as shown:



9. Click **OK**. Hover the cursor on cell **D30** to view the DiscoverSim graphical comment showing the distribution and parameter values:



10. Click on cell **D30**. Click the **DiscoverSim Copy Cell** menu button 

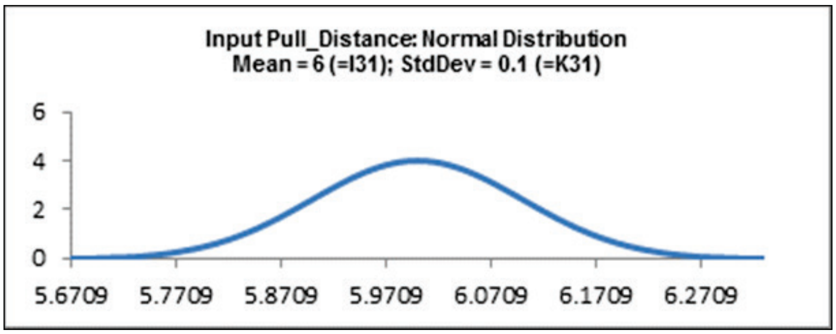
(**Do not use Excel's Copy – it will not work!**).

11. Select cells **D31:D33**. Now click the **DiscoverSim Paste Cell** menu button 

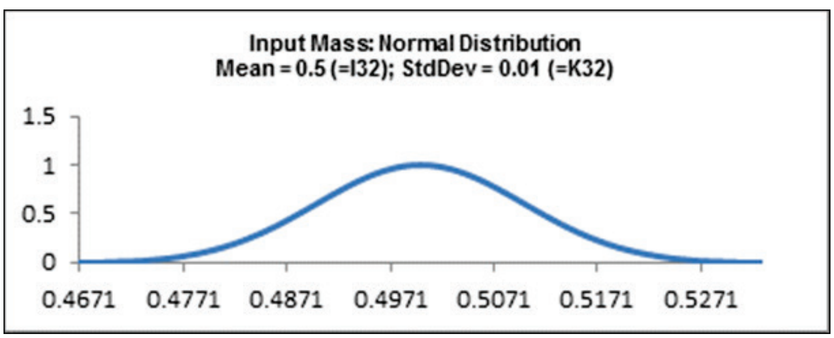
(**Do not use Excel's Paste – it will not work!**).

12. Review the input comments in cells **D31** to **D33**:

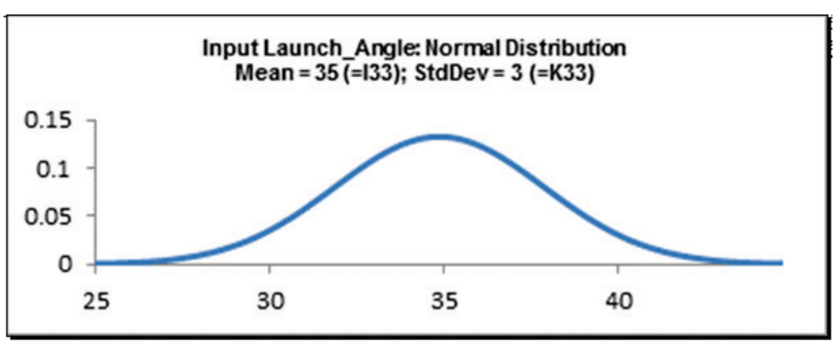
Cell **D31**:



Cell **D32**:



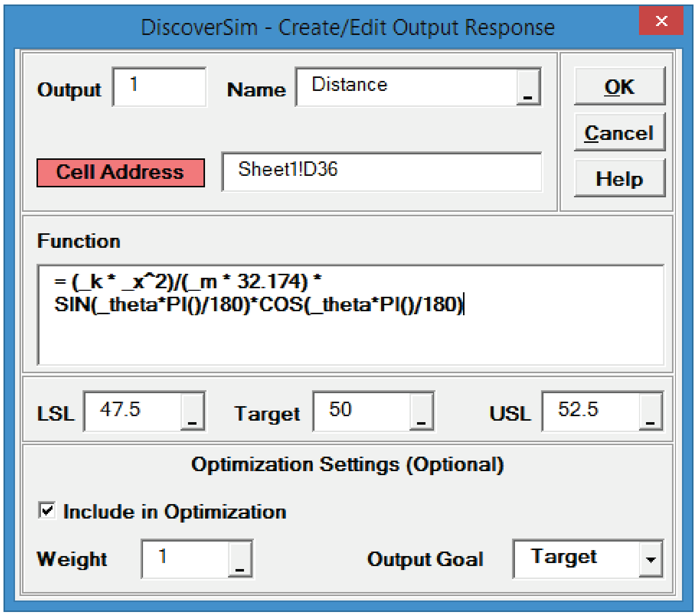
Cell **D33**:



13. Click on cell **D36**. Note that the cell contains the Excel formula for Catapult Firing Distance. Excel range names are used rather than cell addresses to simplify interpretation.

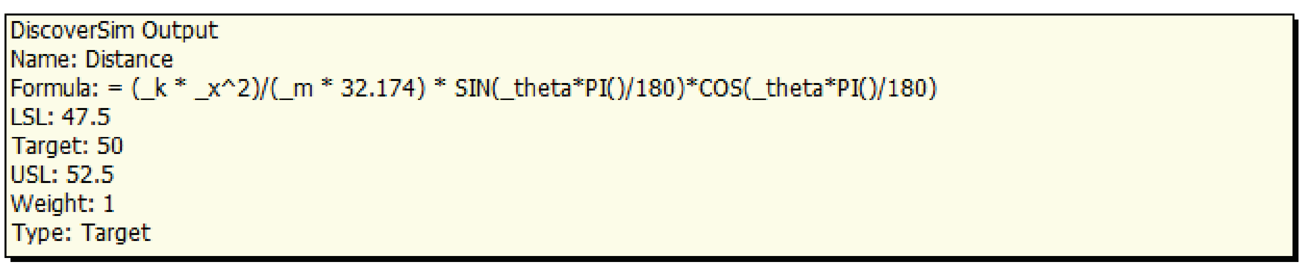
14. Select **DiscoverSim** > **Output Response**:

15. Enter the output **Name** as “Distance.” Enter the Lower Specification Limit (**LSL**) as 47.5, **Target** as 50, and Upper Specification Limit (**USL**) as 52.5.



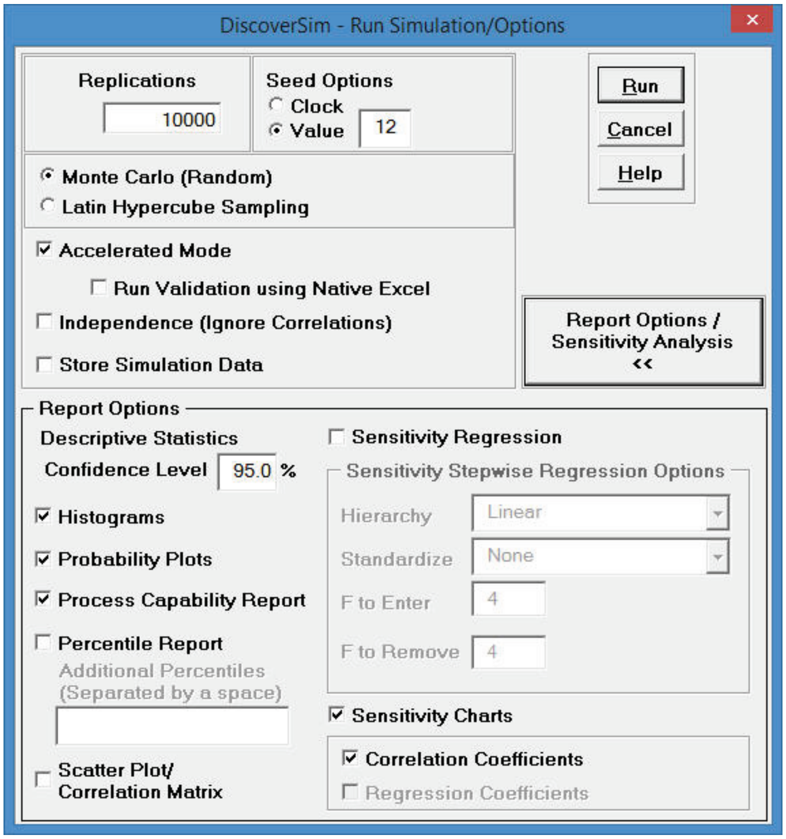
Click **OK**.

16. Hover the cursor on cell **D36** to view the DiscoverSim Output information.



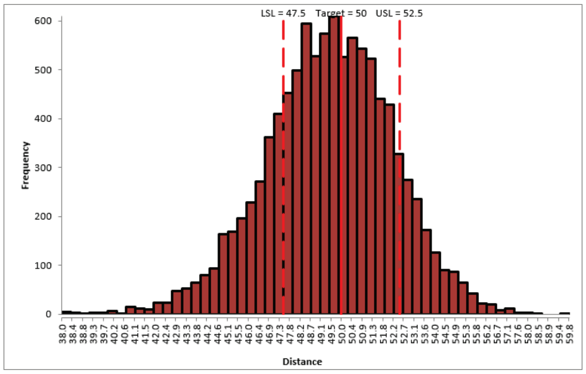
17. Select **DiscoverSim** > **Run Simulation**:

18. Click **Report Options/Sensitivity Analysis**. Check **Sensitivity Charts** and **Correlation Coefficients**. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below.



Click **Run**.

19. The DiscoverSim Output Report shows a histogram, descriptive statistics and process capability indices:



Descriptive Statistics	
Count	10000
Mean	49.486
StDev	2.732499
Range	21.854
Minimum	37.953
25th Percentile (Q1)	47.725
50th Percentile (Median)	49.560
75th Percentile (Q3)	51.353
Maximum	59.807
95.0% CI Mean	49.43219 to 49.53931
95.0% CI Median	49.49627 to 49.62208
95.0% CI StDev	2.695149 to 2.770906

Normality Tests	
Anderson-Darling Normality Test	4.674192
p-value (A-D Test)	0.0000
Skewness	-0.214078
p-value (Skewness)	0.0000
Kurtosis	0.223449
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	0.30
Ppu	0.37
Ppl	0.24
Ppk	0.24
Cpm	0.30

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	134989.43
ppm < LSL	233700.08
ppm Total	368689.51
% > USL	13.50
% < LSL	23.37
% Total (out of spec.)	36.87
% Total (within spec.)	63.13

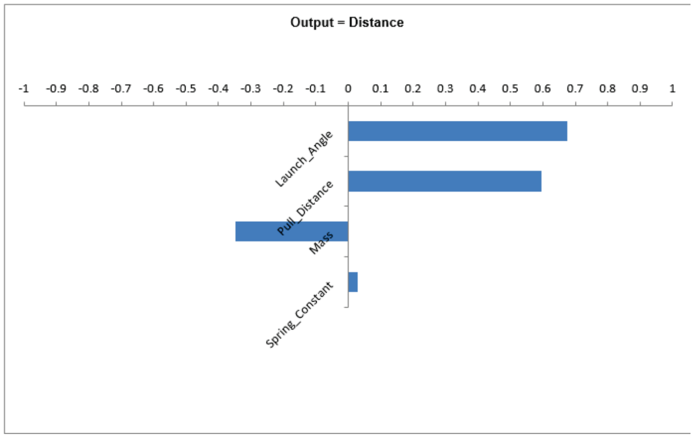
Actual Performance (Empirical)	
ppm > USL	129700.00
ppm < LSL	224200.00
ppm Total	353900.00
% > USL	12.97
% < LSL	22.42
% Total (out of spec.)	35.39
% Total (within spec.)	64.61

From the histogram and capability report we see that the catapult firing distance is not capable of meeting the specification requirements. The variation is unacceptably large. Approximately 35% of the shots fired would fail, so we must improve the process to reduce the variation.

Note: The output distribution is not normal (skewed left), even though the inputs are normal. This is due to the non-linear transfer function.

20. Click on the **Sensitivity Correlations** sheet:

Inputs	Spearman Rank
Launch_Angle	0.676950276
Pull_Distance	0.597038188
Mass	-0.346953706
Spring_Constant	0.030285229



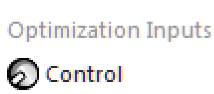
Launch Angle is the dominant input factor affecting Distance, followed by Pull Distance. At this point we could put procedures in place to reduce the standard deviation of these input factors, but before we do that we will run optimization to determine if the process can be made robust to the input variation.

21. In order to perform parameter optimization, we will need to add Input Controls in cells **I30** to **I33**. Select **Sheet1**. The Input Control parameters (min/max) are also shown below:

DiscoverSim Input Distribution Parameters						DiscoverSim Parameter Optimization		
DiscoverSim Input Distribution Name	Distribution	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value	DiscoverSim Input Control Name	Mean Min	Mean Max
Spring_Constant	Normal	Mean	47.3	Standard Deviation	0.1	Spring_Constant_Control	30	70
Pull_Distance	Normal	Mean	6	Standard Deviation	0.1	Pull_Distance_Control	4	8
Mass	Normal	Mean	0.5	Standard Deviation	0.01	Mass_Control	0.5	0.5
Launch_Angle	Normal	Mean	35	Standard Deviation	3	Launch_Angle_Control	20	50

These controls will vary the nominal mean values for Spring Constant, Pull Distance and Launch Angle. Mass is kept fixed at 0.5. The standard deviation values will also be kept fixed for this optimization.

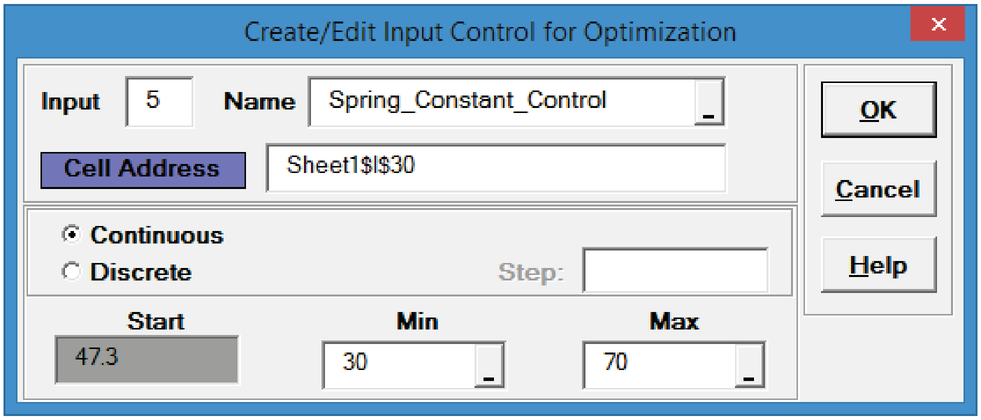
22. Click on cell **I30**. Select **Control**:



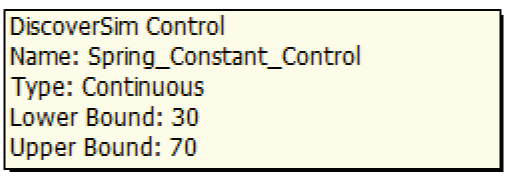
23. Click Input **Name** cell reference and specify cell **M30** containing the input control name "Spring_Constant_Control".

24. Click the **Min** value cell reference and specify cell **N30** containing the minimum optimization boundary value = 30.

25. Click the **Max** value cell reference and specify cell **O30** containing the maximum optimization boundary value = 70:



26. Click **OK**. Hover the cursor on cell **I30** to view the comment displaying the input control settings:



27. Click on cell **I30**. Click the **DiscoverSim Copy Cell** menu button Copy Cell (**Do not use Excel's Copy – it will not work!**).

28. Select cells **I31:I33**. Click the **DiscoverSim Paste Cell** menu button Paste Cell (**Do not use Excel's Paste – it will not work!**).

29. Review the Input Control comments in cells **I31** to **I33**.

30. The completed model is shown below:

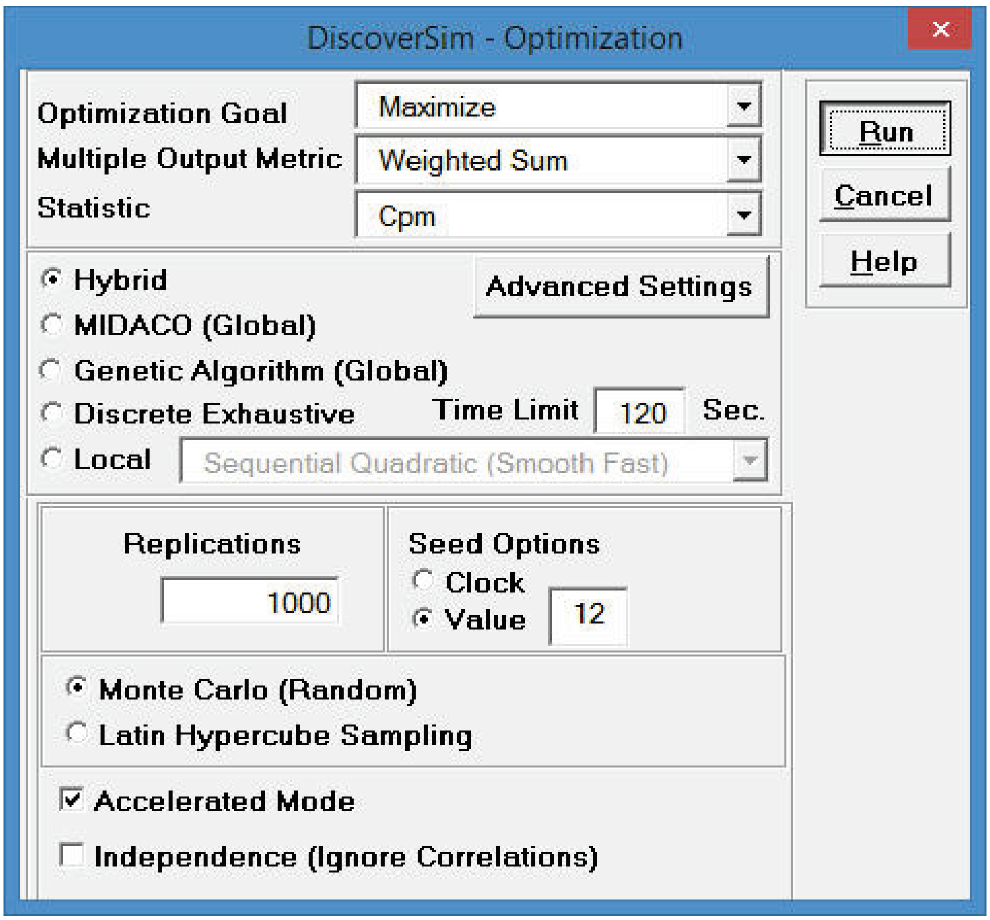
Factor	Units	DiscoverSim Input Distribution
Spring Constant (k)	lb/ft	47.3
Pull Distance (x)	feet	6
Mass (m)	lbf/ft/sec^2	0.5
Launch Angle (θ)	degrees	35

	Output
Distance, Y	49.73297056

DiscoverSim Input Distribution Parameters					
DiscoverSim Input Distribution Name	Distribution	Parameter 1 Name	Parameter 1 Value	Parameter 2 Name	Parameter 2 Value
Spring_Constant	Normal	Mean	47.3	Standard Deviation	0.1
Pull_Distance	Normal	Mean	6	Standard Deviation	0.1
Mass	Normal	Mean	0.5	Standard Deviation	0.01
Launch_Angle	Normal	Mean	35	Standard Deviation	3

31. Now we are ready to perform the optimization. Select **DiscoverSim > Run Optimization**:

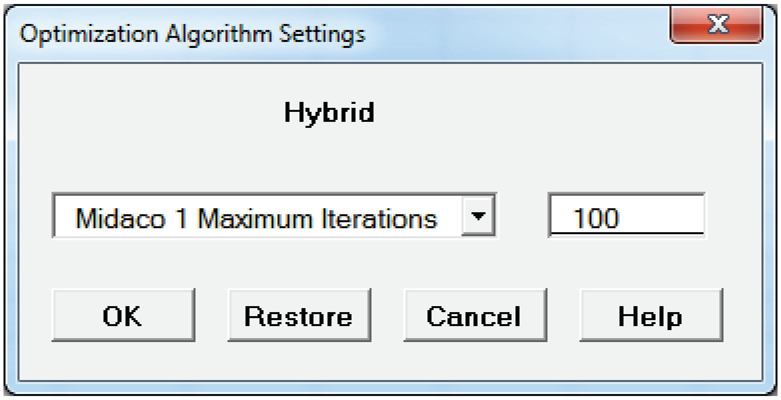
32. Select “Maximize” for **Optimization Goal**, “Weighted Sum” for **Multiple Output Metric** and “Cpm” for **Statistic**. Set **Seed Value** to 12 and **Replications** to 1000 in order to replicate the optimization results given below. All other settings will be the defaults as shown:



Maximize Cpm is used here because it incorporates a penalty for mean deviation from target. We want the distance mean to be on target with minimal variation.

We will use the **Hybrid** optimization method which requires more time to compute, but is powerful to solve complex optimization problems.

Tip: To speed up this optimization, change the maximum number of iterations in Hybrid Midaco 1 (first MIDACO run) from 1000 to 100. Click **Advanced Settings**. Select Midaco 1 Maximum Iterations. Enter 100 as shown. Click **OK**.



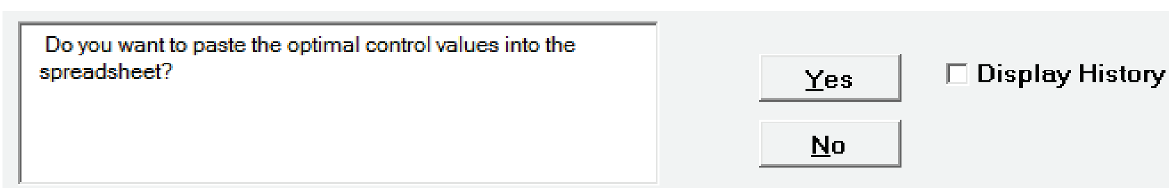
33. Click **Run**.

34. The final optimal parameter values are given as:

```
****************************************************************
                    Optimal Parameters
****************************************************************

   Maximize Weighted Sum (Cpm) : 0.48094734

   Parameter Values:

   Spring_Constant_Control: Value           30.0000
   Pull_Distance_Control: Value              7.3385
   Launch_Angle_Control: Value              45.1925
```

The predicted Cpm value is 0.48, which is an improvement over the baseline value of 0.3.

35. You are prompted to paste the optimal values into the spreadsheet:



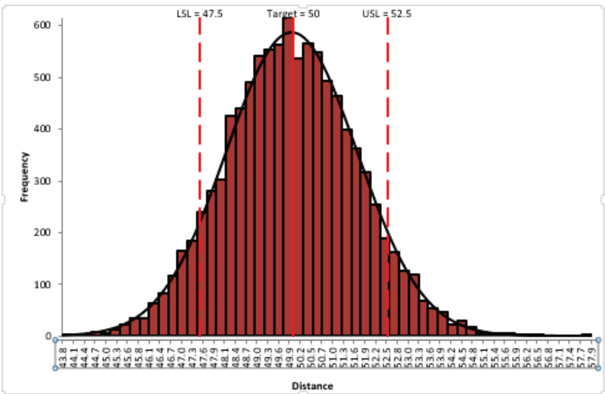
Click **Yes**. This replaces the nominal Input Control values to the optimum values.

DiscoverSim Input Distribution Parameters					
DiscoverSim Input Distribution Name	**Distribution**	**Parameter 1 Name**	**Parameter 1 Value**	**Parameter 2 Name**	**Parameter 2 Value**
Spring_Constant	Normal	Mean	30	Standard Deviation	0.1
Pull_Distance	Normal	Mean	7.33849503	Standard Deviation	0.1
Mass	Normal	Mean	0.5	Standard Deviation	0.01
Launch_Angle	Normal	Mean	45.19248422	Standard Deviation	3

Note that the expected values displayed in the Input Distributions are not automatically updated, but the new referenced Mean values will be used in simulation. To manually update the expected values and chart comments, click on each input cell, select **Input Distribution** and click **OK**.

36. Select **Run Simulation**. Click **Report Options/Sensitivity Analysis**. Check **Sensitivity Charts** and **Correlation Coefficients**. Select **Seed Value** and enter “12,” in order to replicate the simulation results given below. Click **Run**.

37. The resulting simulation report confirms the predicted process improvement:



Descriptive Statistics	
Count	10000
Mean	49.938
StDev	1.750266
Range	14.122
Minimum	43.826
25th Percentile (Q1)	48.743
50th Percentile (Median)	49.903
75th Percentile (Q3)	51.112
Maximum	57.948
95.0% CI Mean	49.90404 to 49.97266
95.0% CI Median	49.86539 to 49.95094
95.0% CI StDev	1.726342 to 1.774868

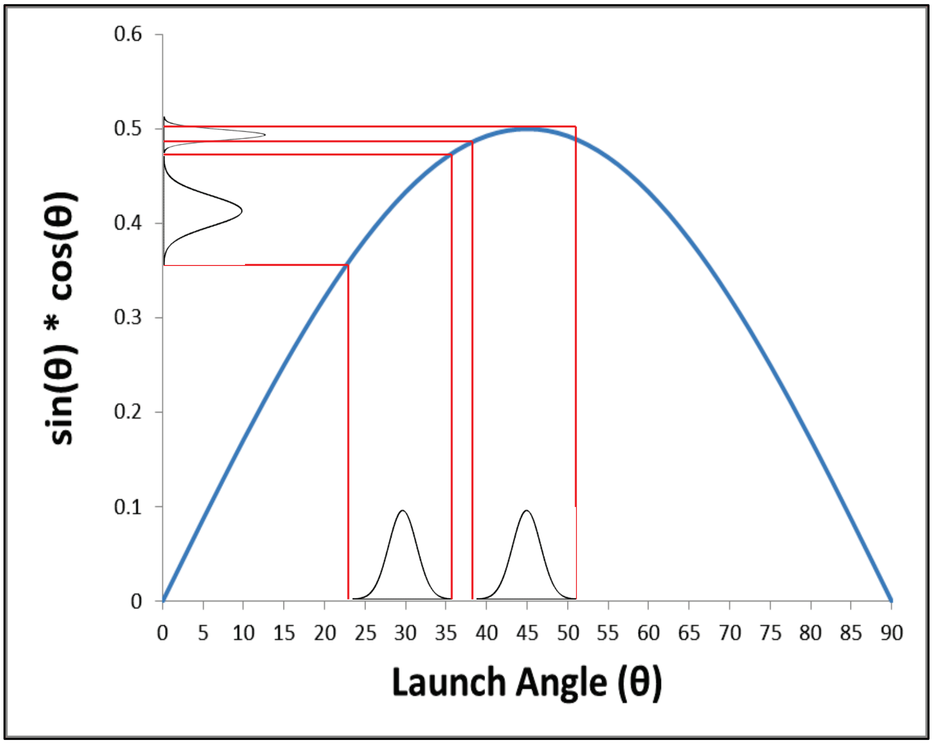
Normality Tests	
Anderson-Darling Normality Test	0.608906
p-value (A-D Test)	0.1135
Skewness	0.073972315
p-value (Skewness)	0.0025
Kurtosis	0.033235273
p-value (Kurtosis)	0.4886

Process Capability Indices	
Pp	0.48
Ppu	0.49
Ppl	0.46
Ppk	0.46
Cpm	0.48

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	71654.67
ppm < LSL	81790.23
ppm Total	153444.90
% > USL	7.17
% < LSL	8.18
% Total (out of spec.)	15.34
% Total (within spec.)	84.66

Actual Performance (Empirical)	
ppm > USL	72300.00
ppm < LSL	81300.00
ppm Total	153600.00
% > USL	7.23
% < LSL	8.13
% Total (out of spec.)	15.36
% Total (within spec.)	84.64

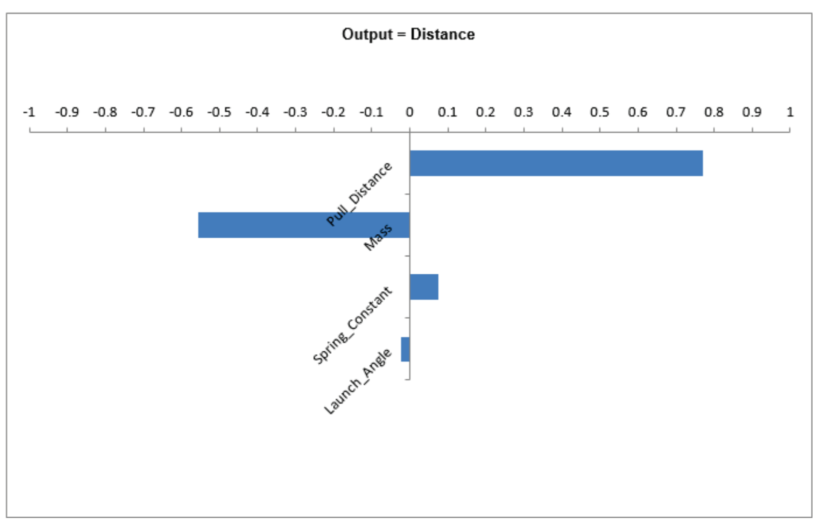
The standard deviation has been reduced from 2.73 to 1.75. The % Total (out-of-spec) has been decreased from 35.0% to 15.3%. This was achieved simply by shifting the mean input values so that the catapult transmitted variation is reduced. The key variable here is the launch angle mean shifting from 30 degrees to 45 degrees as illustrated in the simplified diagram below:



Note: the illustrated improvement here is more dramatic than what was actually realized in the optimization due to the influence of the other input variables.

38. We will continue our improvement efforts, now focusing on reducing the variation of the inputs. Click on the **Sensitivity Correlations (2)** sheet:

Inputs	Spearman Rank
Pull_Distance	0.771702616
Mass	-0.557546948
Spring_Constant	0.075728618
Launch_Angle	-0.024442193



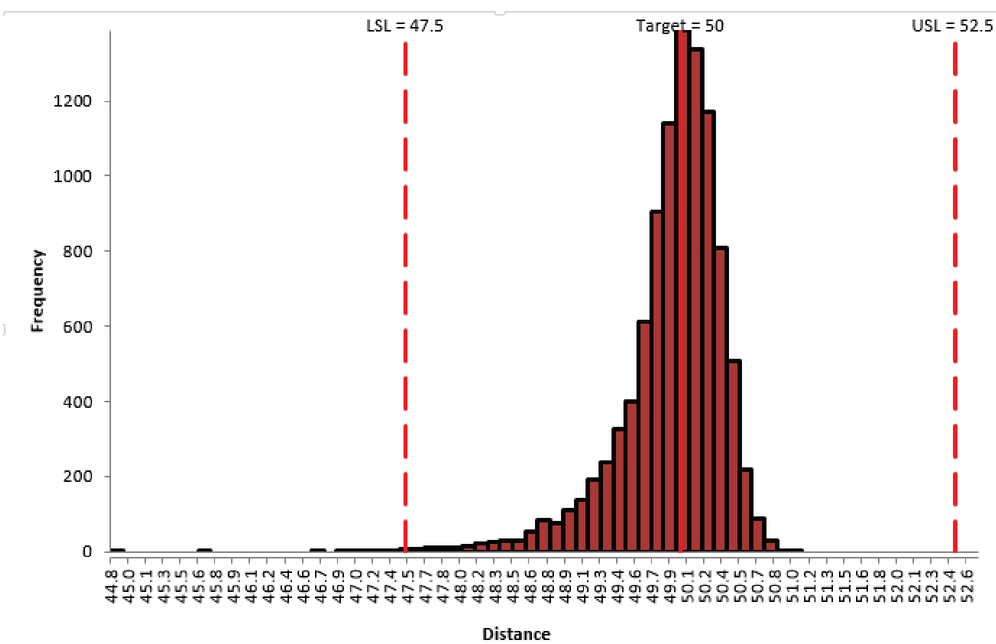
Pull Distance and Mass are now the dominant input factors affecting Distance.
Note: Launch Angle now has the least influence due to the minimized transmitted variation at 45 degrees.

39. Select **Sheet1**, and change the standard deviation for Pull Distance from 0.1 to 0.01 (**K31**), and standard deviation for Mass from .01 to .001 (**K32**) as shown.

	E	F	G	H	I	J	K
27							
28		**DiscoverSim Input Distribution Parameters**					
29		**DiscoverSim Input Distribution Name**	**Distribution**	**Parameter 1 Name**	**Parameter 1 Value**	**Parameter 2 Name**	**Parameter 2 Value**
30		Spring_Constant	Normal	Mean	30	Standard Deviation	0.1
31		Pull_Distance	Normal	Mean	7.3378	Standard Deviation	0.01
32		Mass	Normal	Mean	0.5	Standard Deviation	0.001
33		Launch_Angle	Normal	Mean	45.3656	Standard Deviation	3

Of course it is one thing to type in new standard deviation values to see what the results will be in simulation, but it is another to implement such a change. This requires updates to standard operating procedures, training of operators, tightening of component tolerances and likely increased cost. Robust parameter optimization, as demonstrated above, is easy and cheap. Reducing the variation of inputs can be difficult and expensive.

40. Select **Run Simulation**. Click **Run**. The resulting simulation report shows the further process improvement:



Descriptive Statistics	
Count	10000
Mean	49.928
StDev	0.463031
Range	6.280
Minimum	44.821
25th Percentile (Q1)	49.751
50th Percentile (Median)	50.009
75th Percentile (Q3)	50.218
Maximum	51.101
95.0% CI Mean	49.91843 to 49.93659
95.0% CI Median	50.00277 to 50.01851
95.0% CI StDev	0.456702 to 0.469539

Normality Tests	
Anderson-Darling Normality Test	233.487731
p-value (A-D Test)	0.0000
Skewness	-1.775
p-value (Skewness)	0.0000
Kurtosis	6.571
p-value (Kurtosis)	0.0000

Process Capability Indices	
Pp	1.80
Ppu	1.85
Ppl	1.75
Ppk	1.75
Cpm	1.78

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	0.01
ppm < LSL	0.08
ppm Total	0.09
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	1700.00
ppm Total	1700.00
% > USL	0.00
% < LSL	0.17
% Total (out of spec.)	0.17
% Total (within spec.)	99.83

The standard deviation has been further reduced from 1.73 to 0.46. The % Total (out-of-spec) has decreased from 15.3% to 0.2%

Note: The distribution is not normal (skewed left) due to the non-linearity in the transfer function.

This is not yet a Six Sigma capable process, but it has been dramatically improved. Additional process capability improvement could be realized by optimizing %Ppk (Percentile based Ppk is best for non-normal distributions) rather than Cpm. This would shift the mean to the right of the target, to compensate for the skew left distribution, but the trade-off of mean being on target versus reduction in defects below the lower specification would need to be considered.

Case Study 5 – Robust New Product Design

Introduction: Optimizing Shutoff Valve Spring Force

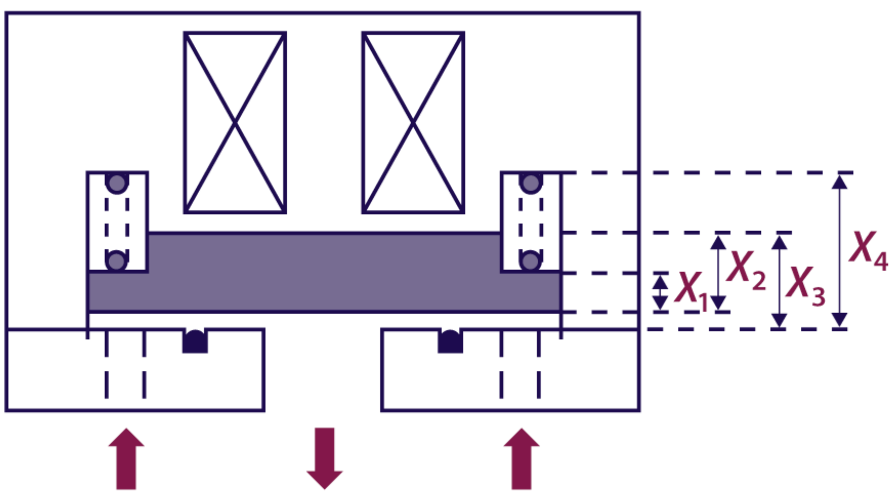
This is an example of DiscoverSim stochastic optimization for robust new product design, adapted from:

Sleeper, Andrew (2006), Design for Six Sigma Statistics: 59 Tools for Diagnosing and Solving Problems in DFSS Initiatives, NY, McGraw-Hill, pp. 782-789.

Sleeper, Andrew, “Accelerating Product Development with Simulation and Stochastic Optimization,” E-mail: AndrewDSleeper at gmail.com.

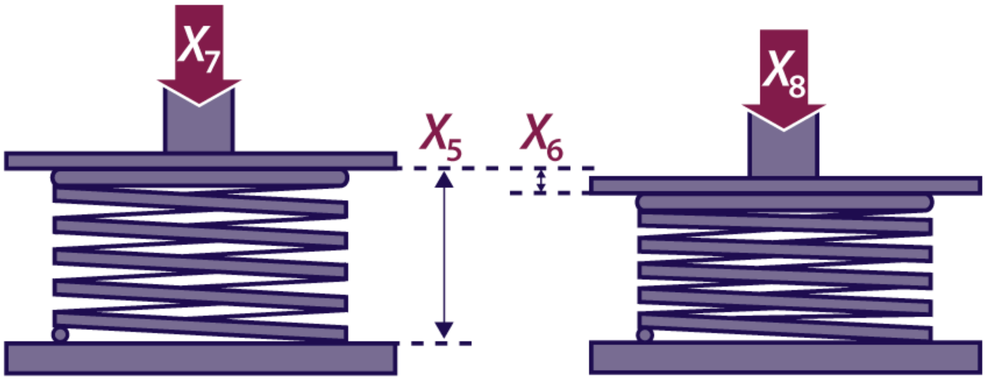
This example is used with permission of the author.

The figure below is a simplified cross-sectional view of a solenoid-operated gas shutoff valve:



The arrows at the bottom of the figure indicate the direction of gas flow. A solenoid holds the plate (shaded) open when energized. When the solenoid is not energized, the spring pushes the plate down to shut off gas flow. If the spring force is too high, the valve will not open or stay open. If the spring force is too low, the valve can be opened by the inlet gas pressure.

The method of specifying and testing the spring is shown below:



The spring force requirement is 22 +/- 2 Newtons.

The spring force equation, or the Y = f(X) transfer function, is calculated as follows:

Spring Length, L = -X1 + X2 - X3 + X4

Spring Rate, R = (X8 - X7)/X6

Spring Force, Y = X7 + R * (X5 - L)

The shut off valve features, tolerances and nominal settings are given as:

Variable	Part	Feature	Units	Feature Tolerance (+/-)	DiscoverSim Input Control (Nominal)
X1	Plate	Lip Height	mm	0.1	3
X2	Plate	Height	mm	0.1	5
X3	Stator	Well Depth	mm	0.1	6.1
X4	Stator	Spring Gap Depth	mm	0.25	10.5
X5	Spring	Initial Compression	mm	0.5	6.2
X6	Spring	Incremental Compression	mm	0.02	0.2
X7	Spring	Force 1	N	0.2	22.3
X8	Spring	Force 2	N	0.2	23

In this study we will use DiscoverSim to help us answer the following questions:

- What is the predicted process capability with these feature tolerances and nominal settings?

- What are the key X variables that influence spring force Y? Can we improve the process capability by tightening the tolerance of the important variables? Can this be done economically?

- Can we adjust the nominal settings of X to reduce the transmitted variation in Y, thereby making the Spring Force robust to the variation due to feature tolerances?

Summary of DiscoverSim Features Demonstrated in Case Study 5:
• Create Input Distributions (Continuous Uniform) • DiscoverSim Copy/Paste Cell • Run Simulation with Sensitivity Chart of Correlation Coefficients • Create Continuous Input Controls • Specify Constraints • Run Optimization

Spring Force Simulation and Optimization with DiscoverSim

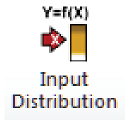
1. Open the workbook **Shutoff Valve Spring Force**. DiscoverSim Input Distributions will simulate the variability in feature tolerance. We will use the Uniform Distribution, since we do not have historical data to estimate a best fit distribution, and will assume an equal probability for a feature to have a minimum tolerance or maximum tolerance value. Input Distributions with Uniform Distributions (Random) will be specified in cells **I15** to **I22**. Nominal (midpoint) values are specified in cells **H15** to **H22** with initial values determined by engineering best estimate. Input Controls will be applied here for optimization. The sum of Nominal + Random is used to calculate the X1 to X8 variables and given in cells **J15** to **J22**. The output, Spring Rate, will be specified at cell **J27** using the formula given in the introduction.

	H	I	J
13			
14	DiscoverSim Input Control (Nominal)	DiscoverSim Random Input Distribution	Nominal + Random
15	3		3
16	5		5
17	6.1		6.1
18	10.5		10.5
19	6.2		6.2
20	0.2		0.2
21	22.3		22.3
22	23		23
23			
24			**Output**
25		Spring Length, L	6.4
26		Spring Rate, R	3.5
27		Spring Force, Y	21.6

2. The Input Distribution parameters are given in the following table:

	L	M	N
12			
13	**DiscoverSim Input Distribution Parameters**		
14	Input Distribution Name	Uniform Minimum	Uniform Maximum
15	X1_Dist	-0.1	0.1
16	X2_Dist	-0.1	0.1
17	X3_Dist	-0.1	0.1
18	X4_Dist	-0.25	0.25
19	X5_Dist	-0.5	0.5
20	X6_Dist	-0.02	0.02
21	X7_Dist	-0.2	0.2
22	X8_Dist	-0.2	0.2

3. Click on cell **I15** to specify the **Input Distribution** for X1, Lip Height.
   Select **DiscoverSim** > **Input Distribution**:



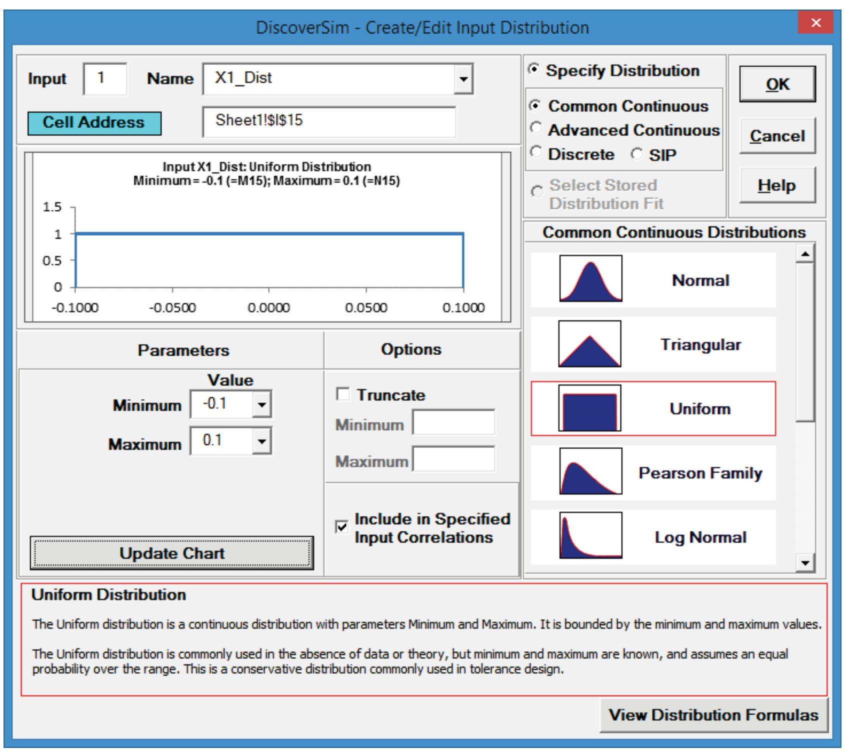
4. Select **Uniform Distribution**.

5. Click input **Name** cell reference and specify cell **L15** containing the input name "X1_Dist." After specifying a cell reference, the dropdown symbol changes from to .

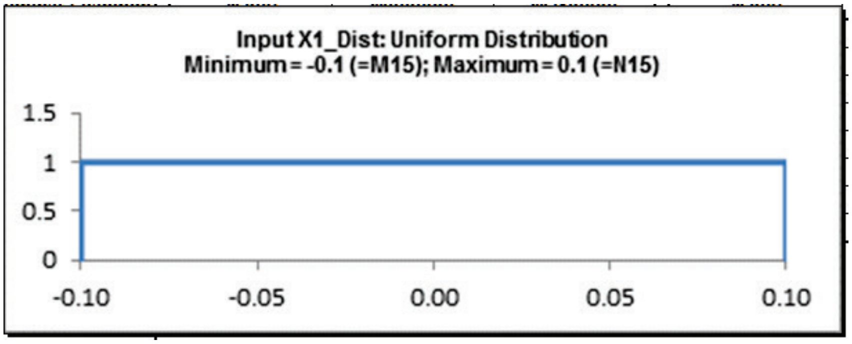
6. Click the **Minimum** parameter cell reference and specify cell **M15** containing the minimum parameter value = -0.1.

7. Click the **Maximum** parameter cell reference and specify cell **N15** containing the maximum parameter value = 0.1.

8. Click **Update Chart** to view the **Uniform Distribution** as shown:



9. Click **OK**. Hover the cursor on cell **I15** to view the DiscoverSim graphical comment showing the distribution and parameter values:

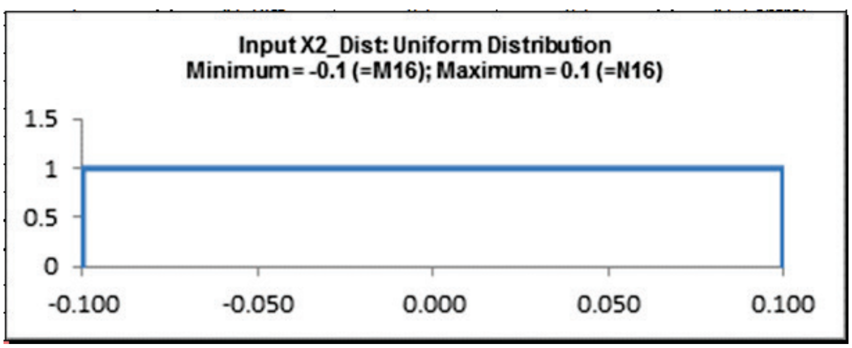


10. Click on cell **I15**. Click the **DiscoverSim Copy Cell** menu button  (**Do not use Excel's Copy – it will not work!**).

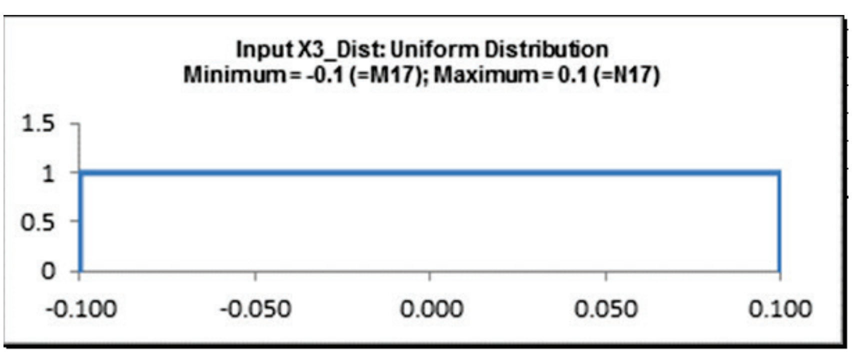
11. Select cells **I16:I22**. Click the **DiscoverSim Paste Cell** menu button  (**Do not use Excel's Paste – it will not work!**).

12. Review the input comments in cells **I16** to **I22**:

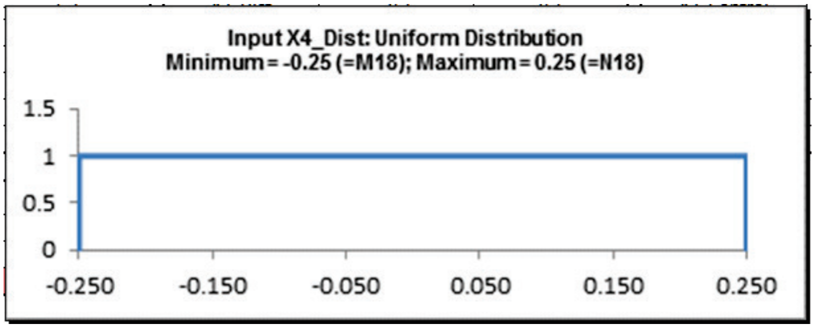
Cell **I16**:



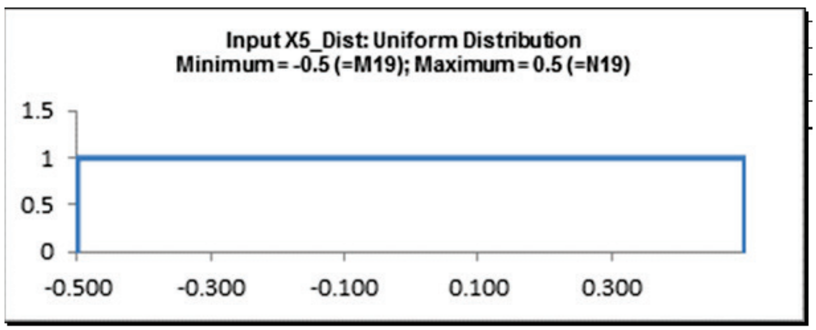
Cell **I17**:



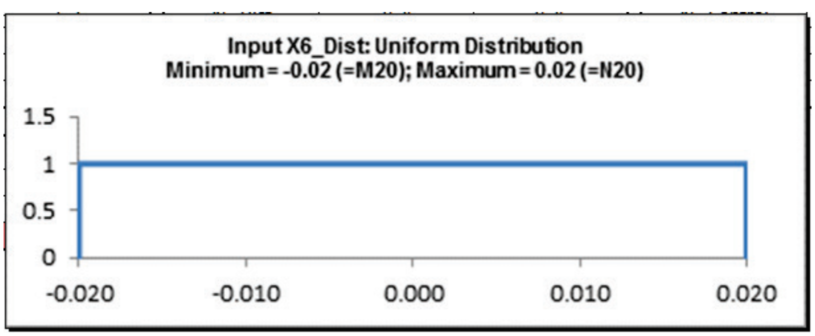
Cell **I18**:



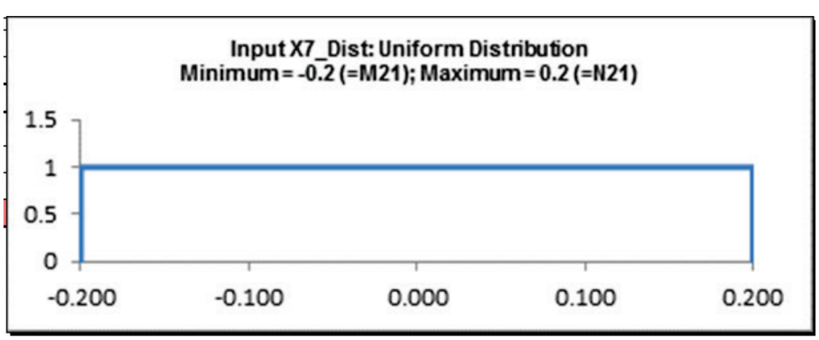
Cell **I19**:



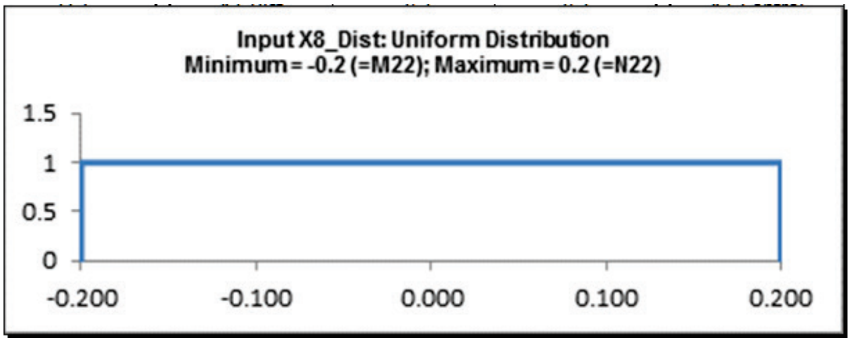
Cell **I20**:



Cell **I21**:



Cell **I22**:

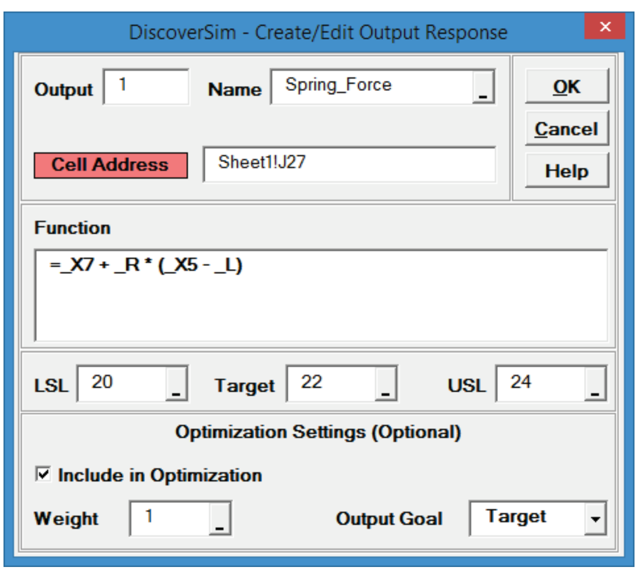


13. Click on cell **J27**.

Note: The cell contains the Excel formula for Spring Force:
=_X7 + _R * (_X5 - _L). Excel range names _X7, _R, _X5 and _L are used rather than cell addresses to simplify interpretation.

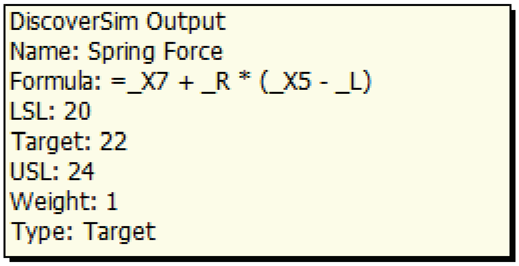
14. Select **DiscoverSim** > **Output Response**:

15. Enter the output **Name** as "Spring_Force." Enter the Lower Specification Limit (**LSL**) as 20, **Target** as 22, and Upper Specification Limit (**USL**) as 24.



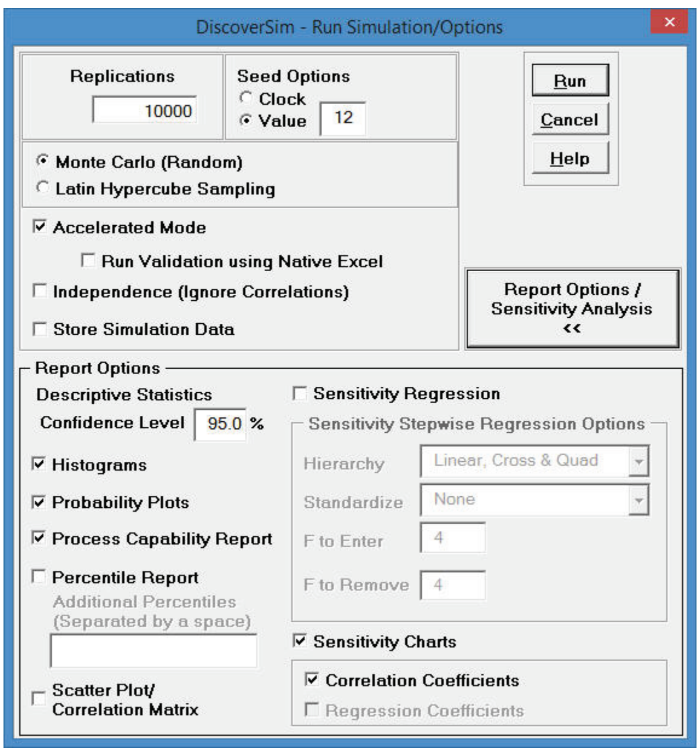
Click **OK**.

16. Hover the cursor on cell **J27** to view the DiscoverSim Output information.



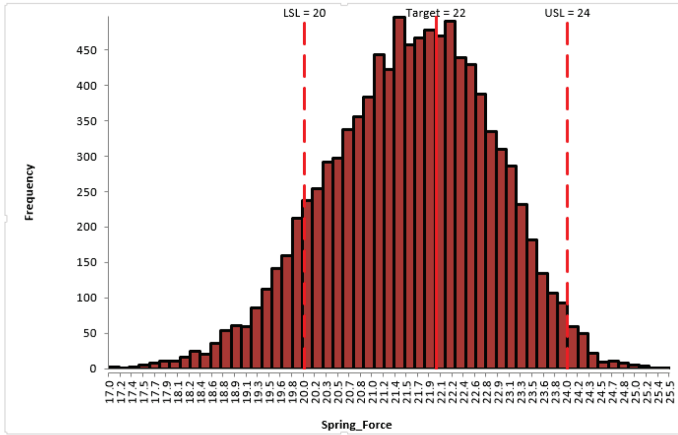
17. Select **DiscoverSim** > **Run Simulation**:

18. Click **Report Options/Sensitivity Analysis**. Check Sensitivity Charts and Correlation Coefficients. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below:



Click **Run**.

19. The DiscoverSim Output Report shows a histogram, descriptive statistics and process capability indices:



Descriptive Statistics	
Count	10000
Mean	21.614
StDev	1.245238
Range	8.525
Minimum	17.020
25th Percentile (Q1)	20.774
50th Percentile (Median)	21.678
75th Percentile (Q3)	22.508
Maximum	25.545
95.0% CI Mean	21.58978 to 21.6386
95.0% CI Median	21.64499 to 21.70796
95.0% CI StDev	1.228217 to 1.262741

Normality Tests	
Anderson-Darling Normality Test	11.263789
p-value (A-D Test)	0.0000
Skewness	-0.246478
p-value (Skewness)	0.0000
Kurtosis	-0.163759
p-value (Kurtosis)	0.0003

Process Capability Indices	
Pp	0.54
Ppu	0.64
Ppl	0.43
Ppk	0.43
Cpm	0.51

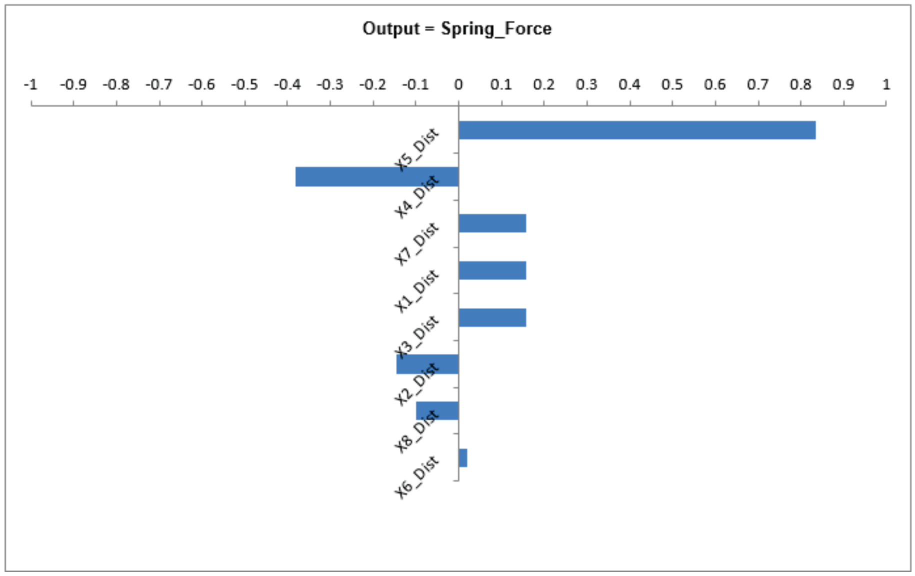
Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	27685.85
ppm < LSL	97438.08
ppm Total	125123.93
% > USL	2.77
% < LSL	9.74
% Total (out of spec.)	12.51
% Total (within spec.)	87.49

Actual Performance (Empirical)	
ppm > USL	17100.00
ppm < LSL	106300.00
ppm Total	123400.00
% > USL	1.71
% < LSL	10.63
% Total (out of spec.)	12.34
% Total (within spec.)	87.66

From the histogram and capability report we see that the Spring Force is not capable of meeting the specification requirements. The process mean is off target and the variation due to feature tolerance is unacceptably large. Approximately 12.3% of the shutoff valves would fail, so we must improve this design to center the mean and reduce the variation.

20. Click on the **Sensitivity Correlations** sheet:

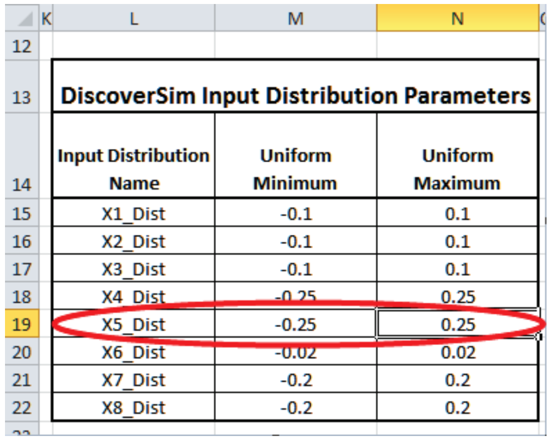
Inputs	Spearman Rank
X5_Dist	0.835145971
X4_Dist	-0.38008182
X7_Dist	0.158555638
X1_Dist	0.158329552
X3_Dist	0.156395986
X2_Dist	-0.146289558
X8_Dist	-0.101055421
X6_Dist	0.020896162



X5 (Initial Compression) is the dominant input factor affecting Spring Force, followed by X4 (Spring Gap Depth). The tolerance on Initial Compression is +/- 0.5. Discussing this with our supplier it turns out that the tolerance for this spring feature can be tightened to +/- 0.25 without increasing the cost.

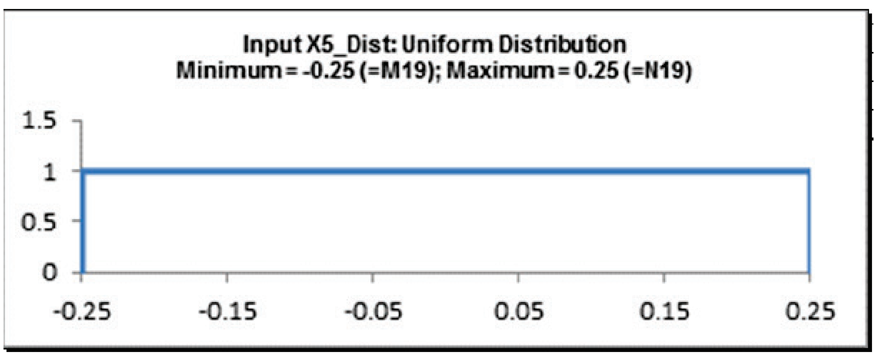
This Sensitivity Chart uses the Spearman Rank Correlation, and the results may be positive or negative. If you wish to view R-Squared percent contribution to variation, rerun the simulation with **Sensitivity Regression Analysis** and **Sensitivity Charts**, **Regression Coefficients** checked.

21. Select **Sheet1**, click on cell **M19**. Change the **Uniform Minimum** value from -0.5 to -0.25. Click on cell **N19**. Change the **Uniform Maximum** value from 0.5 to 0.25.

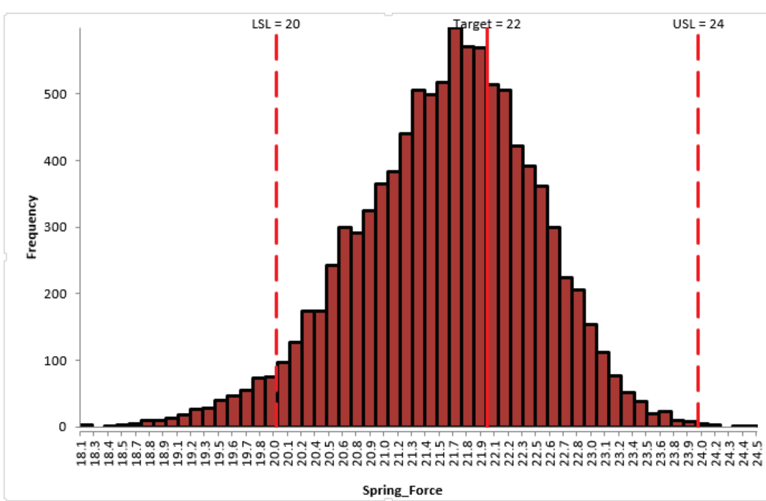


DiscoverSim Input Distribution Parameters		
Input Distribution Name	Uniform Minimum	Uniform Maximum
X1_Dist	-0.1	0.1
X2_Dist	-0.1	0.1
X3_Dist	-0.1	0.1
X4_Dist	-0.25	0.25
X5_Dist	-0.25	0.25
X6_Dist	-0.02	0.02
X7_Dist	-0.2	0.2
X8_Dist	-0.2	0.2

22. DiscoverSim references to cells are dynamic, so this change will take effect on the next simulation run, but the input comment will need to be refreshed manually. Click on cell **I19** (Input **X5_Dist**). Select **Input Distribution**. Click **OK**. Review the revised input comment in cell **I19**:



23. Rerun the simulation without sensitivity charts:



Descriptive Statistics	
Count	10000
Mean	21.610
StDev	0.857187
Range	6.401
Minimum	18.146
25th Percentile (Q1)	21.061
50th Percentile (Median)	21.674
75th Percentile (Q3)	22.203
Maximum	24.547
95.0% CI Mean	21.59289 to 21.6265
95.0% CI Median	21.6546 to 21.69309
95.0% CI StDev	0.845471 to 0.869236

Normality Tests	
Anderson-Darling Normality Test	15.859589
p-value (A-D Test)	0.0000
Skewness	-0.334410
p-value (Skewness)	0.0000
Kurtosis	0.158057
p-value (Kurtosis)	0.0025

Process Capability Indices	
Pp	0.78
Ppu	0.93
Ppl	0.63
Ppk	0.63
Cpm	0.71

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	2647.27
ppm < LSL	30198.90
ppm Total	32846.16
% > USL	0.26
% < LSL	3.02
% Total (out of spec.)	3.28
% Total (within spec.)	96.72

Actual Performance (Empirical)	
ppm > USL	900.00
ppm < LSL	39200.00
ppm Total	40100.00
% > USL	0.09
% < LSL	3.92
% Total (out of spec.)	4.01
% Total (within spec.)	95.99

Tightening the tolerance of X5, Initial Compression resulted in an improvement with the Spring Force standard deviation reduced from 1.24 to 0.86, actual percent out of specification from 12.5 to 4.0% and Ppk increased from .43 to .63. However a Six Sigma design should have a minimum Ppk value of 1.5.

At this point one could use Excel's Solver to further improve the process by finding the nominal values that center the mean, but that still would not achieve the desired quality level.

DiscoverSim's Stochastic Global Optimization will not only find the optimum X settings that result in the best mean spring force value, it will also look for a solution that will reduce the standard deviation. Stochastic optimization looks for a minimum or maximum that is robust to variation in X, thus reducing the transmitted variation in Y. This is referred to as "Robust Parameter Design" in Design For Six Sigma (DFSS).

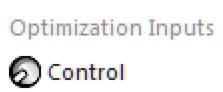
24. In order to run optimization, we will need to add Input Controls (also known as Decision Variables). Select **Sheet1**. The Input Control parameters are given in the table:

	O	P	Q	R	S	T	U
12							
13		DiscoverSim Input Control Parameters				DiscoverSim Constraints	
14		Input Control Name	Input Control Min	Input Control Max		Constraint Description	DiscoverSim Constraint
15		X1_Control	2	5			
16		X2_Control	4	7		X2_Control - X1_Control > 0.2	
17		X3_Control	5	7			
18		X4_Control	8	20		X4_Control - X3_Control > 0.35	
19		X5_Control	5	15			
20		X6_Control	0.1	3			
21		X7_Control	15	35			
22		X8_Control	15	35		X8_Control - X7_Control > 0.5	

The controls will vary the nominal X values in **H15** to **H22**. The nominal values are added to the random input distribution (**J15** to **J22**) as discussed above in Step 1.

Constraints will also be added as shown so that there can be no overlap due to tolerance. For example if X2, Height, was at the low end of its tolerance and X1, Lip Height, was at the high end of its tolerance, it is possible that an overlap would leave no place for the spring to be seated.

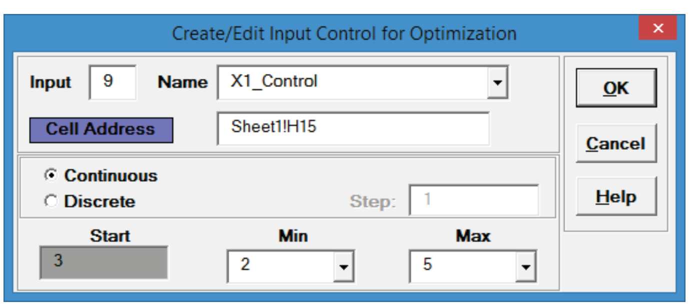
25. Click on cell **H15**. Select **Control**:



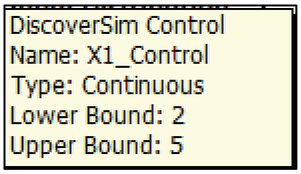
26. Click input **Name** cell reference and specify cell **P15** containing the input control name "X1_Control."

27. Click the **Min** value cell reference and specify cell **Q15** containing the minimum optimization boundary value = 2.

28. Click the **Max** value cell reference and specify cell **R15** containing the maximum optimization boundary value = 5:



29. Click **OK**. Hover the cursor on cell **H15** to view the comment displaying the input control settings:

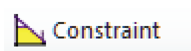


30. Click on cell **H15**. Click the **DiscoverSim Copy Cell** menu button  (**Do not use Excel's Copy** – **it will not work!**).

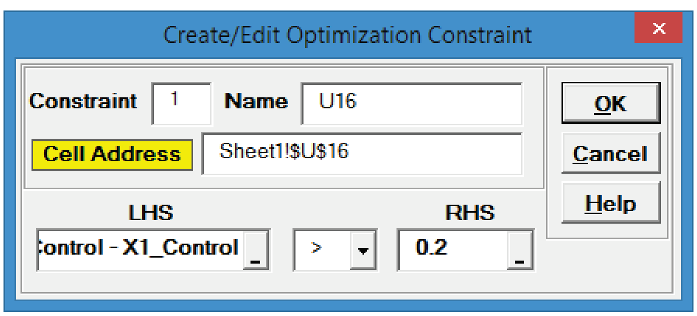
31. Select cells **H16:H22**. Click the **DiscoverSim Paste Cell** menu button  (**Do not use Excel's Paste** – **it will not work!**).

32. Review the input control comments in cells **H16** to **H22**.

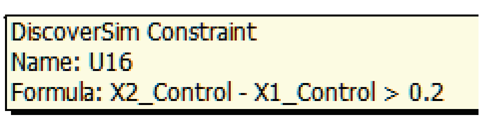
33. Now we will add the Constraints. Click on cell **U16**. Select **Constraint**:



34. Type **X2_Control – X1_Control** in the "Left Hand Side" (**LHS**). Select "Greater Than **>**." Enter 0.2 in the "Right Hand Side" (**RHS**).



35. Click **OK**. Review the comment at cell **U16**.



The cell display is "TRUE" since the initial value for X2_Control (5) is greater than the initial value for X1_Control (3).

36. Repeat steps 33 to 35 with constraints at cells **U18** and **U22** as shown:

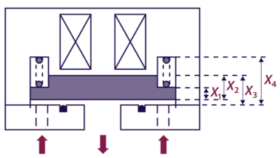
DiscoverSim Constraint
Name: U18
Formula: X4_Control - X3_Control > 0.35

DiscoverSim Constraint
Name: U22
Formula: X8_Control - X7_Control > 0.5

37. The completed model is shown below:

Variable	Part	Feature	Units	Feature Tolerance (+/-)	DiscoverSim Input Control (Nominal)	DiscoverSim Random Input Distribution	0
X1	Plate	Lip Height	mm	0.1	3	0	3
X2	Plate	Height	mm	0.1	5	0	5
X3	Stator	Well Depth	mm	0.1	6.1	0	6.1
X4	Stator	Spring Gap Depth	mm	0.25	10.5	0	10.5
X5	Spring	Initial Compression	mm	0.5	6.2	0	6.2
X6	Spring	Incremental Compression	mm	0.02	0.2	0	0.2
X7	Spring	Force 1	N	0.2	22.3	0	22.3
X8	Spring	Force 2	N	0.2	23	0	23

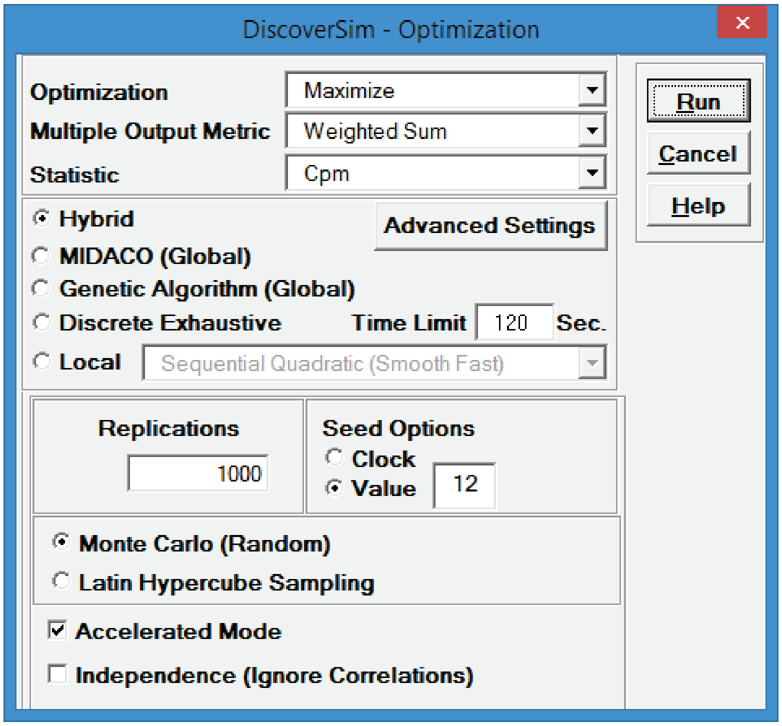
DiscoverSim Input Distribution Parameters			DiscoverSim Input Control Parameters			DiscoverSim Constraints	
Input Distribution Name	Uniform Minimum	Uniform Maximum	Input Control Name	Input Control Min	Input Control Max	Constraint Description	DiscoverSim Constraint
X1_Dist	-0.1	0.1	X1_Control	2	5		
X2_Dist	-0.1	0.1	X2_Control	4	7	X2_Control - X1_Control > 0.2	TRUE
X3_Dist	-0.1	0.1	X3_Control	5	7		
X4_Dist	-0.25	0.25	X4_Control	8	20	X4_Control - X3_Control > 0.35	TRUE
X5_Dist	-0.25	0.25	X5_Control	5	15		
X6_Dist	-0.02	0.02	X6_Control	0.1	3		
X7_Dist	-0.2	0.2	X7_Control	15	35		
X8_Dist	-0.2	0.2	X8_Control	15	35	X8_Control - X7_Control > 0.5	TRUE



	Output
Spring Length, L	6.4
Spring Rate, R	3.5
Spring Force, Y	21.6

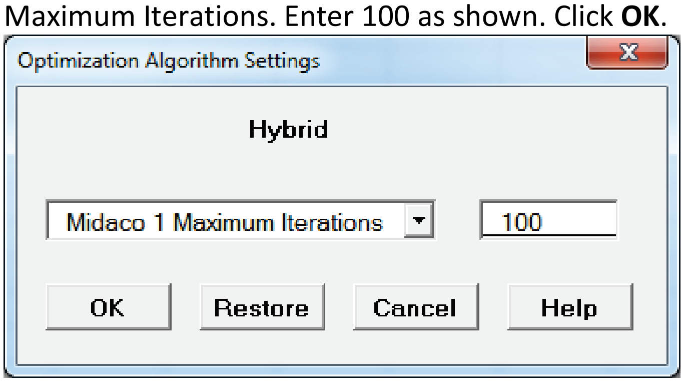
38. Now we are ready to perform the optimization. Select **DiscoverSim > Run Optimization**:

39. Select “Maximize” for **Optimization Goal**, “Weighted Sum” for **Multiple Output Metric** and “Cpm” for **Statistic**. Select **Seed Value** and enter “12,” in order to replicate the optimization results given below. Set **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:



Maximize Cpm is used here, rather than Ppk, because it incorporates a penalty for mean deviation from target. We want the spring force mean to be on target with minimal variation. Note, however, that this solution may not produce the lowest possible standard deviation.
We will use the **Hybrid** optimization method which requires more time to compute, but is powerful to solve complex optimization problems.

Tip: To speed up this optimization, change the maximum number of iterations in Hybrid Midaco 1 (first MIDACO run) from 1000 to 100. Click **Advanced Settings**. Select Midaco 1 Maximum Iterations. Enter 100 as shown. Click **OK**.



40. Click **Run**.

41. Scroll through the optimization report window to view the various optimization methods. The initial method in Hybrid, MIDACO global optimization, found a solution that resulted in a world class Cpm value of 7.4!:

```
******************************************************************
       Input Control Optimization - Method = MIDACO
******************************************************************

 Convergence achieved after     100.00000  iterations.
 Objective value:               7.3975319
 Number of observations:        1000.0000
 Execution Time:                105.63300  secs.
 Return Code:                   3.0000000
 Number of function calls:      100053.00

     Input Control       Value
    ----------------------------
       Control_1        2.994082
       Control_2        4.609373
       Control_3        6.090986
       Control_4        8.063186
       Control_5        5.135034
       Control_6        3.000000
       Control_7        21.744164
       Control_8        22.244165

******************************************************************
                     Optimal Parameters
******************************************************************

   Maximize Weighted Sum (Cpm) : 7.3975319

   Parameter Values:

   X1_Control: Value            2.9941
   X2_Control: Value            4.6094
   X3_Control: Value            6.0910
   X4_Control: Value            8.0632
   X5_Control: Value            5.1350
   X6_Control: Value            3.0000
   X7_Control: Value           21.7442
   X8_Control: Value           22.2442
```

42. The Genetic Algorithm uses the above optimal values as a starting point, but is unable to further improve on MIDACO's Cpm value of 7.4

```
******************************************************************
       Input Control Optimization - Method = GenAlg
******************************************************************

 Convergence achieved after     103.00000  iterations.
 Objective value:               7.3975322
 Number of observations:        1000.0000
 Execution Time:                11.699000  secs.
 Number of function calls:      9329.0000

     Input Control       Value
    ----------------------------
       Control_1        2.994082
       Control_2        4.610004
       Control_3        6.091742
       Control_4        8.063186
       Control_5        5.134761
       Control_6        3.000000
       Control_7        21.744164
       Control_8        22.244165

******************************************************************
                     Optimal Parameters
******************************************************************

   Maximize Weighted Sum (Cpm) : 7.3975322

   Parameter Values:

   X1_Control: Value            2.9941
   X2_Control: Value            4.6100
   X3_Control: Value            6.0917
   X4_Control: Value            8.0632
   X5_Control: Value            5.1348
   X6_Control: Value            3.0000
   X7_Control: Value           21.7442
   X8_Control: Value           22.2442
```

43. The final method in this Hybrid, Sequential Quadratic Programming local optimization, uses the Genetic Algorithm optimal values as a starting point but is unable to further improve on the Cpm value of 7.4.

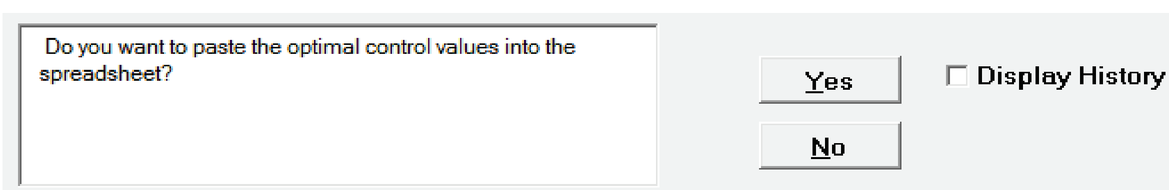
```
*****************************************************************
        Input Control Optimization - Method = SQP
*****************************************************************

 Convergence achieved after     2.0000000  iterations.
 Objective value:               7.3975324
 Number of observations:        1000.0000
 Execution Time:                0.16500000  secs.
 Return Code:                   0.00000000
 Number of function calls:      128.00000

     Input Control        Value
    ---------------------------
       Control_1        2.994068
       Control_2        4.610014
       Control_3        6.091727
       Control_4        8.063194
       Control_5        5.134746
       Control_6        3.000000
       Control_7        21.744175
       Control_8        22.244176

*****************************************************************
                     Optimal Parameters
*****************************************************************

   Maximize Weighted Sum (Cpm) : 7.3975324

   Parameter Values:

   X1_Control: Value             2.9941
   X2_Control: Value             4.6100
   X3_Control: Value             6.0917
   X4_Control: Value             8.0632
   X5_Control: Value             5.1347
   X6_Control: Value             3.0000
   X7_Control: Value            21.7442
   X8_Control: Value            22.2442
```

44. The constraint equations evaluated at this optimum are satisfied:

```
Constraint Equations evaluated at the optimum
Constraint Methodology:  Native

   Constraint        Type        Amount Violated

      1               gt             0.0000
      2               gt             0.0000
      3               gt             0.0000
```

Note: The constraint cell values in Excel will all appear as TRUE at the optimum.

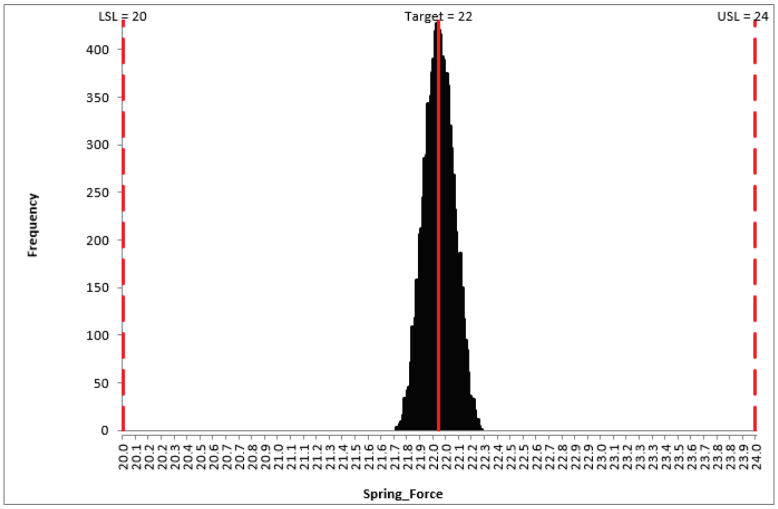
45. You are prompted to paste the optimal values into the spreadsheet:



Click **Yes**. This replaces the nominal input control values to the optimum values.

46. Select **Run Simulation**. Click **Run**.

47. The resulting simulation report confirms the predicted dramatic improvement:



Descriptive Statistics	
Count	10000
Mean	22.002
StDev	0.090694
Range	0.538
Minimum	21.732
25th Percentile (Q1)	21.938
50th Percentile (Median)	22.001
75th Percentile (Q3)	22.065
Maximum	22.270
95.0% CI Mean	22.00031 to 22.00387
95.0% CI Median	21.99891 to 22.00336
95.0% CI StDev	0.089455 to 0.091969

Normality Tests	
Anderson-Darling Normality Test	3.159880
p-value (A-D Test)	0.0000
Skewness	0.022600081
p-value (Skewness)	0.3560
Kurtosis	-0.379043
p-value (Kurtosis)	0.0000

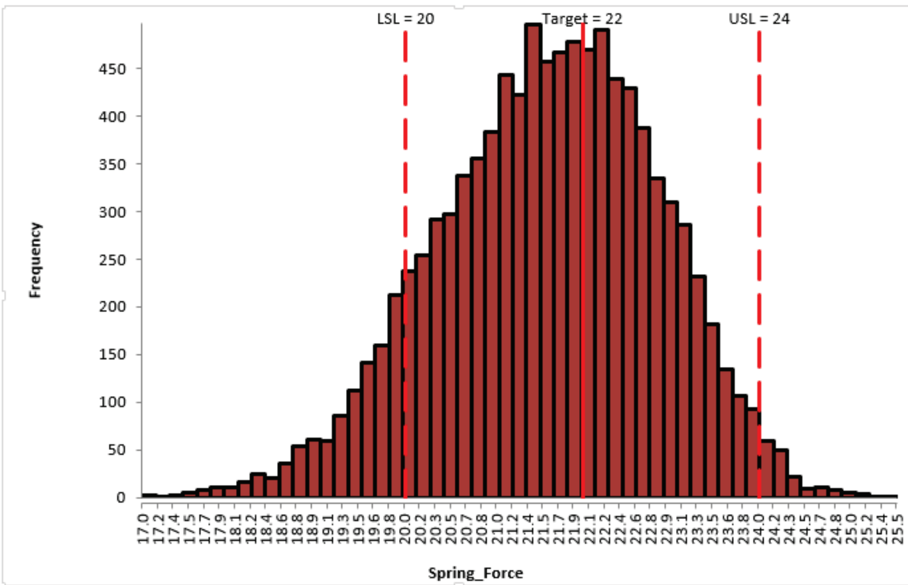
Process Capability Indices	
Pp	7.35
Ppu	7.34
Ppl	7.36
Ppk	7.34
Cpm	7.35

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

The slight discrepancy in predicted Cpm (Optimization 7.4 versus Simulation 7.35 is due to the difference in number of replications for Optimization (1000) versus Simulation (10,000).

48. Recall the histogram for the initial simulation run:



49. In summary, DiscoverSim was used to dramatically improve the Spring Force performance as follows:

    a. Mean centered from 21.6 to 22.0
    b. Standard Deviation reduced from 1.24 to 0.09, more than a ten-fold reduction!
    c. Ppk increased from 0.43 to 7.34, Cpm increased from 0.51 to 7.35!
    d. Actual % Total (out of spec) reduced from 12.3% to 0%!

50. The benefits do not stop here. Since the design is now so robust, we can review the input tolerances to see if there is a cost saving opportunity by widening the feature tolerances, and re-running the simulation to study the impact.

51. Finally the results predicted here should be validated with physical prototypes before proceeding to finalize the design parameters. Remember "All models are wrong, some are useful" (George Box)!

Case Study 6 – Multiple Response Optimization and Practical Tolerance Design

Introduction: Optimizing Low Pass RC Filter

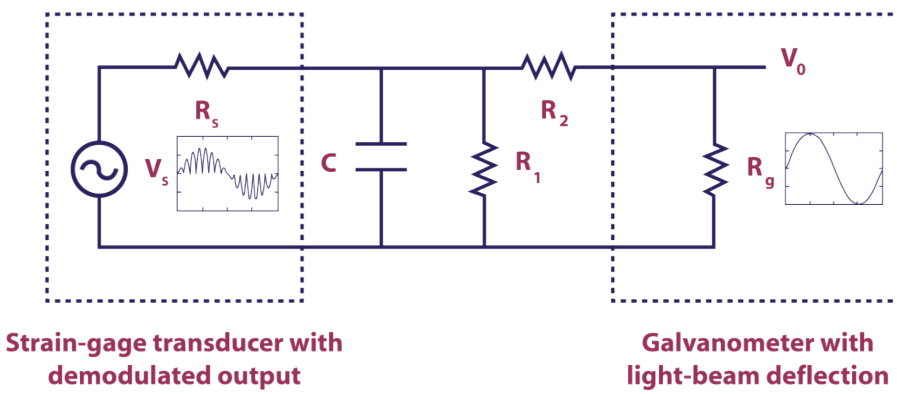
The passive low pass RC (Resistor Capacitor) filter network shown below has been analyzed in several papers (see Case Study 6 References) and will be used to demonstrate how DiscoverSim allows you to easily solve the difficult problems of multiple response optimization for parameter design and practical tolerance design.

The design approaches presented in the reference papers and books include: Axiomatic Design (Suh, El-Haik), Taguchi Methods (Filippone), Analytical Design for Six Sigma (Ferryanto), and Integrated Parameter and Tolerance Design using Factorial Design of Experiments and Loss Function (Feng).

Acknowledgements and thanks to Dr. Eric Maass and Andrew Sleeper for their valuable feedback on this case study.

The RC filter network to be designed (C, R1 and R2) suppresses the high-frequency carrier signal and passes only the desired low-frequency displacement signal from the strain-gage transducer to the recorder with a galvanometer/light-beam deflection indicator (El-Haik, Feng). Light beam galvanometers have been replaced with digital oscilloscopes, but they still have practical application today in laser scanning systems.

Passive RC Filter Network



V_s is the demodulated transducer voltage and R_s is the transducer resistance (impedance). R_g is the galvanometer terminal resistance. V_o is the galvanometer output voltage which produces the deflection displacement. One of the output responses of interest, galvanometer full scale deflection D, is a function of V_o, R_g, and the galvanometer sensitivity G_s. The values for V_s, R_s, R_g, and G_s are determined by the transducer and galvanometer vendors and are therefore considered as noise factors. We will use the nominal and tolerance values given by Feng and Filippone for the noise factors: V_s = 15 mV; R_s = 120 ohms; R_g = 84 ohms; G_s = 0.56364 mA/inch. All tolerance values are ± 0.15%. We will model these noise factors using a normal distribution with mean set to the nominal value and standard deviation calculated as nominal*(tolerance/100)/3 = nominal*0.0015/3.

Capacitor C, and resistors R_1, R_2 are the components of the passive filter network. These are the design control factors of interest in this study. The mean values for C, R_1, and R_2 will be varied using parameter design optimization to meet the required output specifications with minimal variation. Tolerances on these components are also noise factors that contribute to variability in the output but we can select those tolerance values that are required to give us the desired performance. The starting nominal and tolerance values will be those given in Feng and Filippone: C = 257 uF, R1 = 1060 ohms, and R2 = 464 ohms with tolerance = 5% for each. As with the noise factors, we will model the control factors using a normal distribution with mean initially set to the nominal value and standard deviation initially set to nominal*(tolerance/100)/3 = nominal*0.05/3.

The output responses of interest are full scale deflection D and filter cut-off frequency F. The equations (derived from Kirchhoff's Current Law) are given as:

$$D = \frac{V_s R_g R_1}{G_s\left[\left(R_g + R_2\right)\left(R_s + R_1\right) + R_s R_1\right]}$$

$$F = \frac{\left(R_g + R_2\right)\left(R_s + R_1\right) + R_s R_1}{2\pi\left(R_g + R_2\right) R_s R_1 C \bullet 10^{-6}}$$

Note: Resistors R_1 and R_2 affect both output responses, and capacitor C only affects the cut-off frequency F.

The target value for Deflection D is 3.0 inches. We will use ± 1% as the tolerance for D so the lower specification limit LSL_D = 2.97 and upper specification limit USL_D = 3.03. The target value for Filter Cut-Off Frequency F is 6.84 Hz. We will use ± 5% as the tolerance for F so LSL_F = 6.498 and USL_F = 7.182. Process Capability indices will be used to measure the quality of each output response, but optimization will be performed by minimizing the multiple output metric “Weighted Sum of Mean Squared Error (MSE),” also known as the Taguchi loss function:

MSE (Loss) = $Weight_D[(Mean_D - Target_D)^2 + StdDev_D^2] + Weight_F[(Mean_F - Target_F)^2 + StdDev_F^2]$

The control and noise factor settings are summarized in the table below:

				Input Distribution Parameters		
Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	257	4.283333333	5
Control	R1	Filter Resistor 1 (ohms)	1060	1060	17.66666667	5
Control	R2	Filter Resistor 2 (ohms)	464	464	7.733333333	5
Noise	Vs	Transducer Voltage (mV)	15	15	0.0075	0.15
Noise	Rs	Transducer Resistance (ohms)	120	120	0.06	0.15
Noise	Rg	Galvanometer Resistance (ohms)	84	84	0.042	0.15
Noise	Gs	Galvanometer Sensitivity (mA/inch)	0.56364	0.56364	0.00028182	0.15

The minimum and maximum mean values used for parameter optimization of the control factors are the values from Filippone for the Taguchi Inner Array:

Input Control Parameters (Mean)		
Name	Min	Max
C_Control	231	1400
R1_Control	20	100000
R2_Control	0.01	525

Note that if we did not have the above nominal mean values for control factors available as starting values, we would use the mid-point of the minimum/maximum to start the optimization. The final optimization result would be the same in either case because the optimizer is global not local.

We will also consider standard component availability and cost in this case study (source: electronic component supplier www.mouser.com – note prices are subject to change). One practical concern will be the capacitor. Capacitors larger than 230 uF will typically be electrolytic with a 20% tolerance or more, so specifying a tolerance of 5% or less will significantly increase the component cost.

In this study we will use DiscoverSim to help us answer the following questions:

- What are the predicted process capability indices with the published nominal settings?
- What are the predicted process capability indices after parameter optimization? How do they compare to the baseline results above?
- What are the key control factors (Xs) that influence Deflection D and Frequency F?
- What standard components are available that are close to the optimal values obtained from parameter design?
- What tolerances are available for those standard components? Can we minimize the component cost and achieve a Six Sigma quality design?

Summary of DiscoverSim Features Demonstrated in Case Study 6:
• Create Input Distributions (Continuous Normal) • DiscoverSim Copy/Paste Cell • Run Simulation with Sensitivity Chart of Correlation Coefficients • Create Continuous Input Controls for Parameter Optimization • Run Optimization for Multiple Outputs

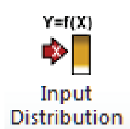
Multiple Response Optimization and Practical Tolerance Design of Low Pass RC Filter with DiscoverSim

1. Open the workbook **Passive RC Filter Design**. DiscoverSim Input Distributions will simulate the variability in control and noise component tolerances. The Normal Distribution (as used in the reference papers) will be specified in cells **E17** to **E23**. The mean values are specified in cells **F17** to **F23**. The standard deviation values are calculated in cells **G17** to **G23** as discussed above to model the component tolerances, specified in cells **H17** to **H23**. Input Controls will be added to cells **F17** to **F19** which will vary the mean for the control factors C, R1 and R2 in optimization. The outputs, Deflection and Frequency, will be specified in cells **E26** and **E27** using the formulas given above.

Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Input Distribution Parameters: Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	257	4.283333333	5
Control	R1	Filter Resistor 1 (ohms)	1060	1060	17.66666667	5
Control	R2	Filter Resistor 2 (ohms)	464	464	7.733333333	5
Noise	Vs	Transducer Voltage (mV)	15	15	0.0075	0.15
Noise	Rs	Transducer Resistance (ohms)	120	120	0.06	0.15
Noise	Rg	Galvanometer Resistance (ohms)	84	84	0.042	0.15
Noise	Gs	Galvanometer Sensitivity (mA/inch)	0.56364	0.56364	0.00028182	0.15

Output Name	DiscoverSim Output Responses	Specification
Deflection	3.062128625	3.0 inches +/- 1%
Frequency	6.874965344	6.84 Hz +/- 5%

2. Click on cell **E17** to specify the Input Distribution for the Capacitor C. Select **DiscoverSim > Input Distribution:**



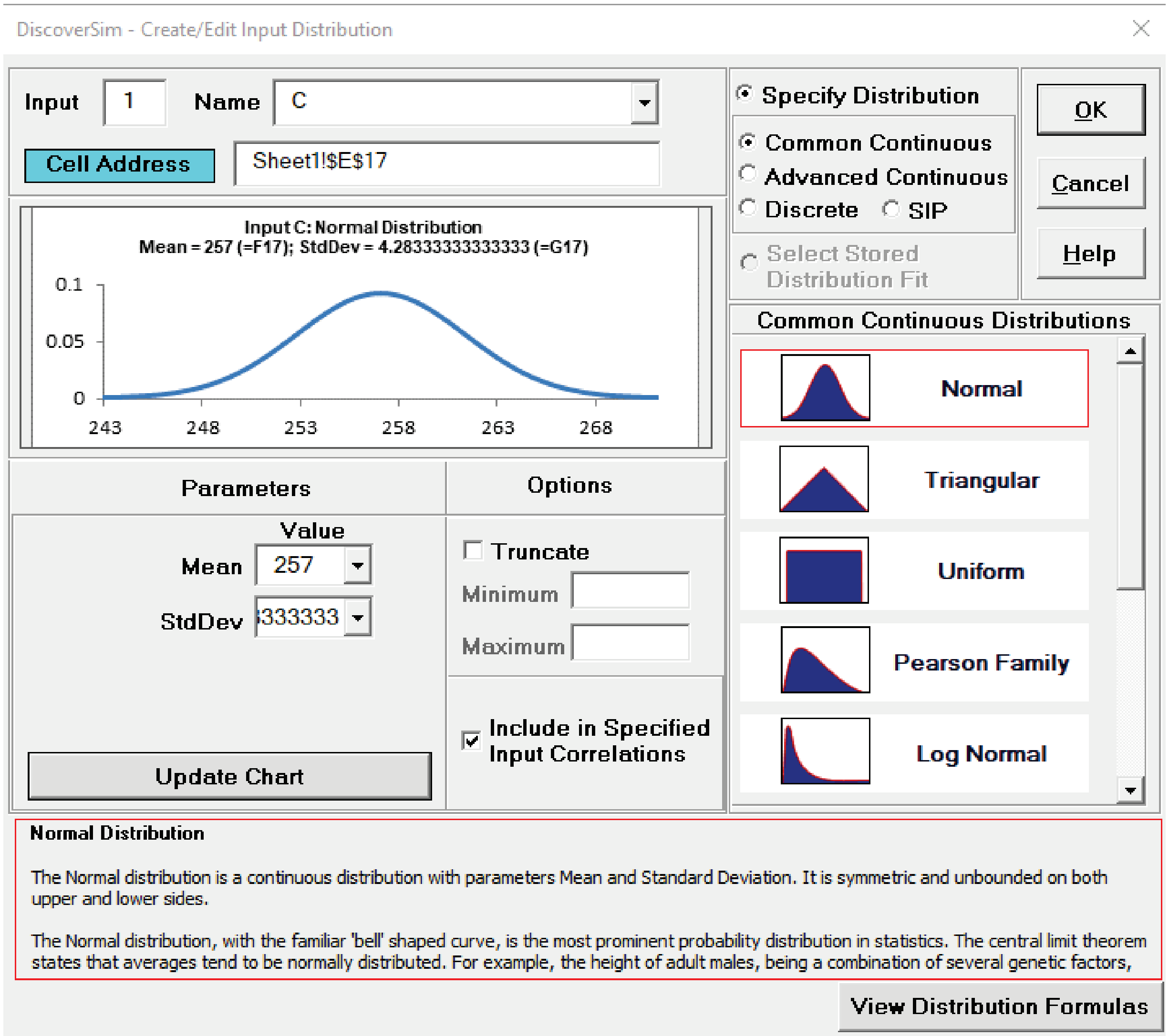
3. We will use the default **Normal Distribution**.

4. Click input **Name** cell reference and specify cell **C17** containing the input name "C." After specifying a cell reference, the dropdown symbol changes from to .

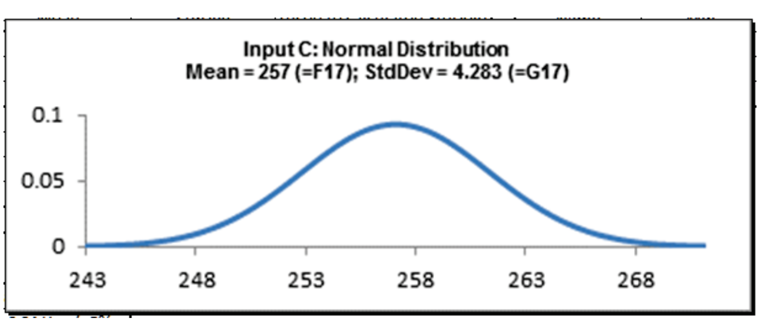
5. Click the **Mean** parameter cell reference and specify cell **F17** containing the mean parameter value = 257.

6. Click the **StdDev** parameter cell reference and specify cell **G17** containing the standard deviation parameter value = Mean * (Tolerance/100)/3 = (257 * .05)/3 = 4.283.

7. Click **Update Chart** to view the normal distribution as shown:



8. Click **OK**. Hover the cursor on cell **E17** to view the DiscoverSim graphical comment showing the distribution and parameter values:



9. Click on cell **E17**. Click the **DiscoverSim Copy Cell** menu button 

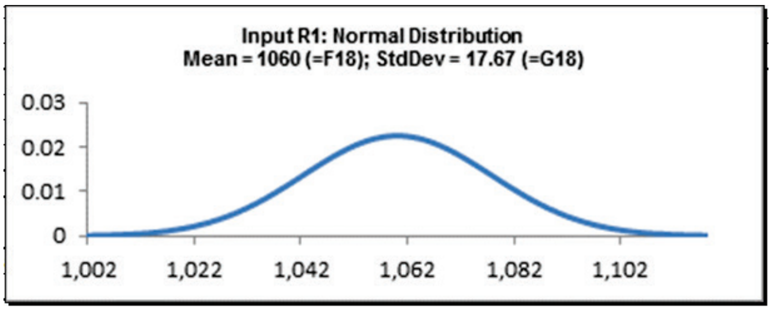
 (**Do not use Excel's Copy – it will not work!**).

10. Select cells **E18:E23**. Click the **DiscoverSim Paste Cell** menu button 

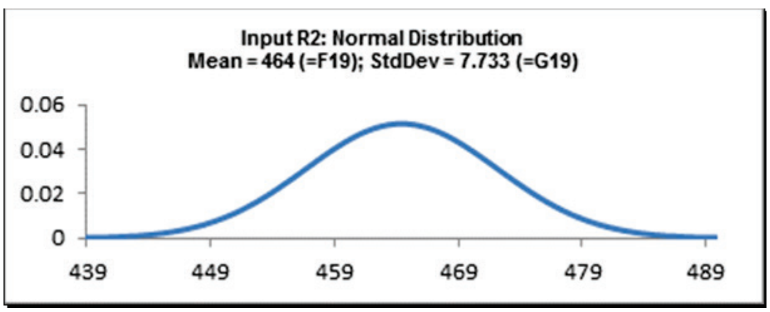
(**Do not use Excel's Paste – it will not work!**).

11. Review the input comments in cells **E18** to **E23**:

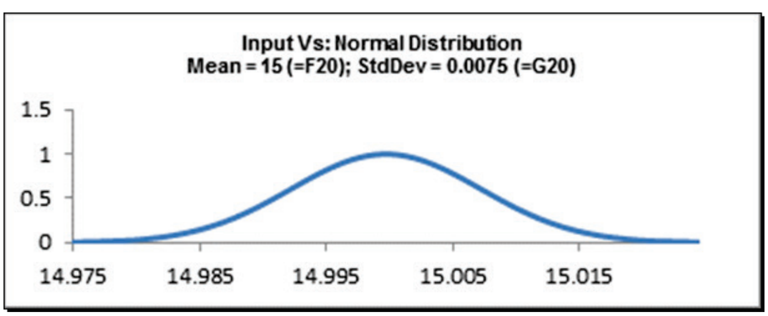
    Cell **E18**:



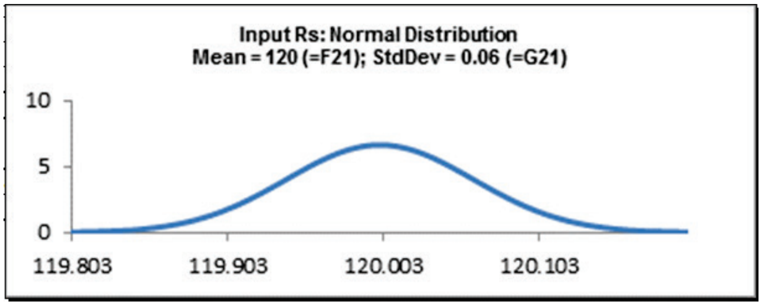
    Cell **E19**:



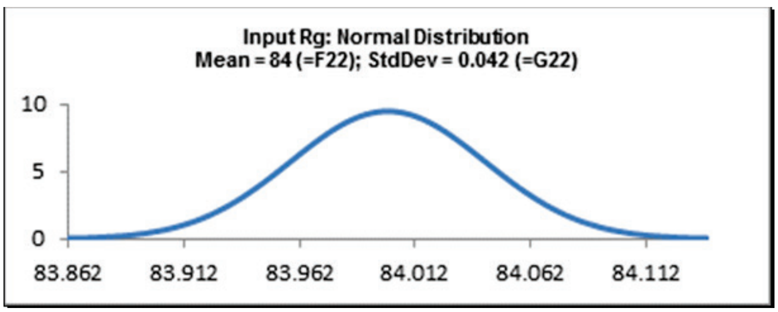
    Cell **E20**:



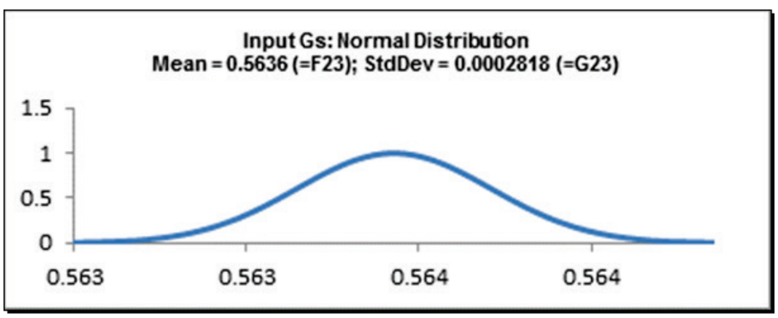
Cell **E21**:



Cell **E22**:



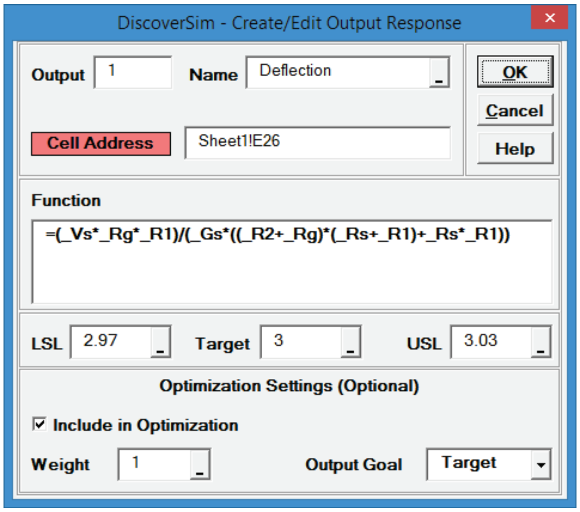
Cell **E23**:



12. Now we will specify the outputs. Click on cell **E26**. Note that the cell contains the Excel formula for Deflection:
    =(_Vs*_Rg*_R1)/(_Gs*((_R2+_Rg)*(_Rs+_R1)+_Rs*_R1)).
    Excel range names are used rather than cell addresses to simplify interpretation.

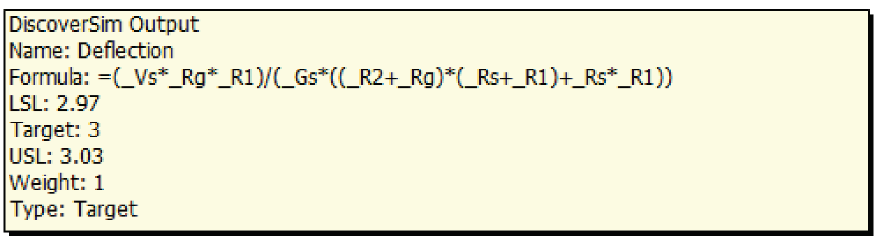
13. Select **DiscoverSim** > **Output Response**:

14. Enter the output **Name** as "Deflection." Enter the Lower Specification Limit (**LSL**) as 2.97, **Target** as 3, and Upper Specification Limit (**USL**) as 3.03. **Weight** = 1 and **Output Goal** should be "Target."



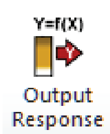
Click **OK**.

15. Hover the cursor on cell **E26** to view the DiscoverSim Output information.

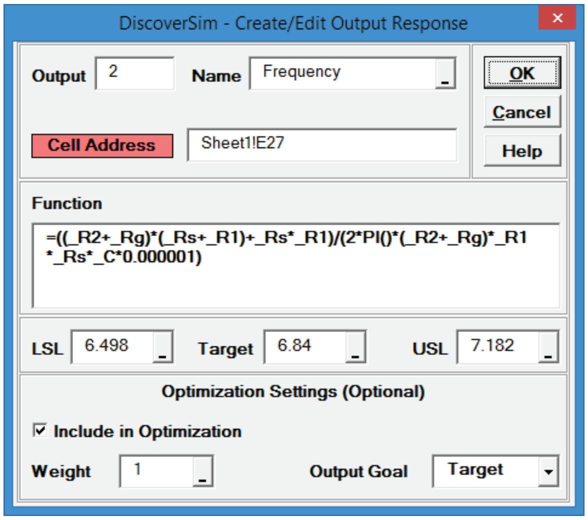

DiscoverSim Output
Name: Deflection
Formula: =(_Vs*_Rg*_R1)/(_Gs*((_R2+_Rg)*(_Rs+_R1)+_Rs*_R1))
LSL: 2.97
Target: 3
USL: 3.03
Weight: 1
Type: Target

16. Click on cell **E27**. Note that the cell contains the Excel formula for Frequency:
=((_R2+_Rg)*(_Rs+_R1)+_Rs*_R1)/(2*PI()*(_R2+_Rg)*_R1*_Rs*_C*0.000001)

17. Select **DiscoverSim** > **Output Response**:

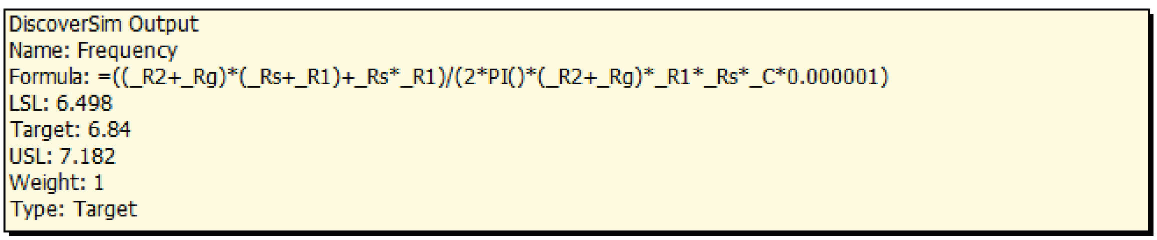


18. Enter the output **Name** as "Frequency." Enter the Lower Specification Limit (**LSL**) as 6.498, **Target** as 6.84, and Upper Specification Limit (**USL**) as 7.182. **Weight** = 1 and **Output Goal** should be "Target."



Click **OK**.

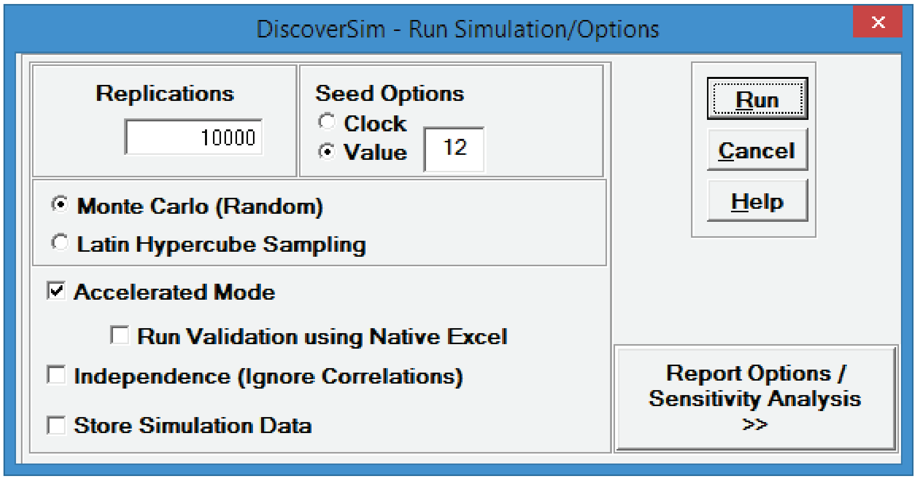
19. Hover the cursor on cell **E27** to view the DiscoverSim Output information.



20. Now we are ready to perform a simulation run. This is our baseline, after which we will run parameter optimization and compare results. We will also perform a Sensitivity Analysis after optimization.

21. Select **DiscoverSim** > **Run Simulation**:

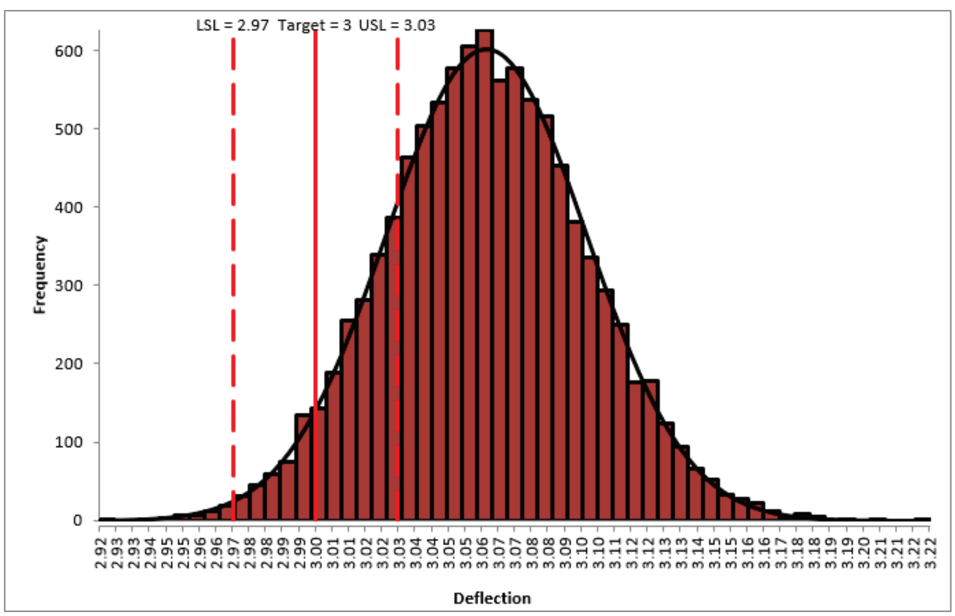
22. Select **Seed Value** and enter “12” as shown, in order to replicate the simulation results given below.



Click **Run**.

23. The DiscoverSim Output Report shows a histogram, descriptive statistics and process capability indices for each output.

Deflection Report:



Descriptive Statistics	
Count	10000
Mean	3.063
StDev	0.036576
Range	0.303
Minimum	2.921
25th Percentile (Q1)	3.038
50th Percentile (Median)	3.062
75th Percentile (Q3)	3.087
Maximum	3.224
95.0% CI Mean	3.061809 to 3.063243
95.0% CI Median	3.061007 to 3.062858
95.0% CI StDev	0.036076 to 0.03709

Normality Tests	
Anderson-Darling Normality Test	0.575252
p-value (A-D Test)	0.1352
Skewness	0.07038036
p-value (Skewness)	0.0041
Kurtosis	-0.011721293
p-value (Kurtosis)	0.8286

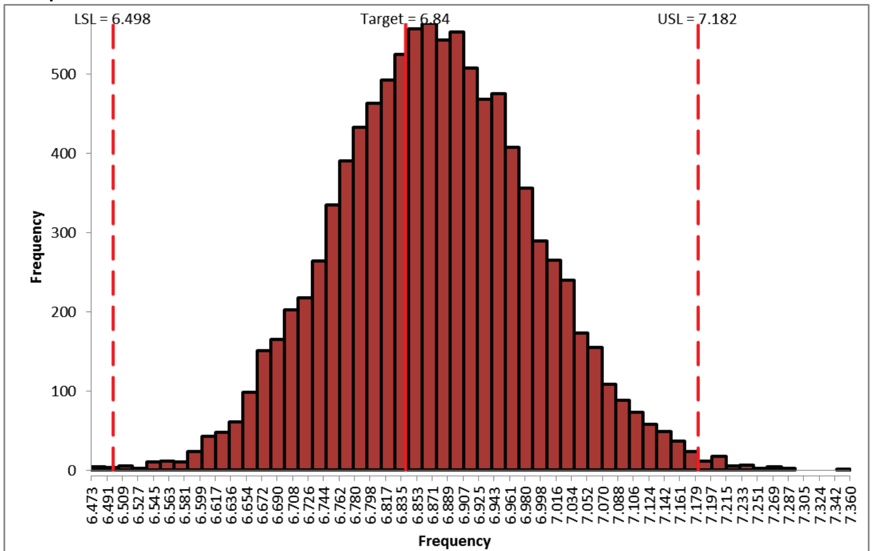
Process Capability Indices	
Pp	0.27
Ppu	-0.30
Ppl	0.84
Ppk	-0.30
Cpm	0.14

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	813072.90
ppm < LSL	5707.91
ppm Total	818780.81
% > USL	81.31
% < LSL	0.57
% Total (out of spec.)	81.88
% Total (within spec.)	18.12

Actual Performance (Empirical)	
ppm > USL	812500.00
ppm < LSL	4200.00
ppm Total	816700.00
% > USL	81.25
% < LSL	0.42
% Total (out of spec.)	81.67
% Total (within spec.)	18.33

From the histogram and capability report we see that the output Deflection is not capable of meeting the specification requirements using the baseline values. The process mean is off target and the variation due to component tolerance is unacceptably large. We expect that 81.9% of the units would fail this requirement, so we must improve the design to center the mean and reduce the variation of Deflection, while simultaneously satisfying the Frequency requirement.

Frequency Report:



Descriptive Statistics	
Count	10000
Mean	6.880
StDev	0.116066
Range	0.834
Minimum	6.472
25th Percentile (Q1)	6.800
50th Percentile (Median)	6.878
75th Percentile (Q3)	6.957
Maximum	7.306
95.0% CI Mean	6.877653 to 6.882203
95.0% CI Median	6.875338 to 6.881085
95.0% CI StDev	0.11448 to 0.117697

Normality Tests	
Anderson-Darling Normality Test	1.657236
p-value (A-D Test)	0.0003
Skewness	0.129396
p-value (Skewness)	0.0000
Kurtosis	0.000829548
p-value (Kurtosis)	0.9670

Process Capability Indices	
Pp	0.98
Ppu	0.87
Ppl	1.10
Ppk	0.87
Cpm	0.93

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	4626.20
ppm < LSL	499.85
ppm Total	5126.05
% > USL	0.46
% < LSL	0.05
% Total (out of spec.)	0.51
% Total (within spec.)	99.49

Actual Performance (Empirical)	
ppm > USL	6900.00
ppm < LSL	200.00
ppm Total	7100.00
% > USL	0.69
% < LSL	0.02
% Total (out of spec.)	0.71
% Total (within spec.)	99.29

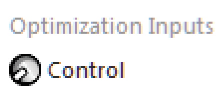
From the histogram and capability report we see that the output Frequency is slightly off target and just capable of meeting the specification requirements using the baseline values. We expect that 0.71% of the units would fail this requirement, so if possible we should improve the design to reduce the variation of Frequency while simultaneously improving Deflection.

24. In order to run optimization, we will need to add the Input Controls in cells **F17** to **F19**. Select **Sheet1**. The Input Control parameters (min/max) are shown below (from Filippone):

	E	F	G	H	I	J	K	L
14								
15		Input Distribution Parameters				Input Control Parameters (Mean)		
16	DiscoverSim Input Distributions	Mean	StdDev	% Tolerance (used to calculated StdDev)		Name	Min	Max
17	257	257	4.283333333	5		C_Control	231	1400
18	1060	1060	17.66666667	5		R1_Control	20	100000
19	464	464	7.733333333	5		R2_Control	0.01	525

These controls will vary the nominal mean values for C, R1 and R2. The standard deviation will be recalculated with each optimization iteration, so that a 5% tolerance is preserved (see above for formula).

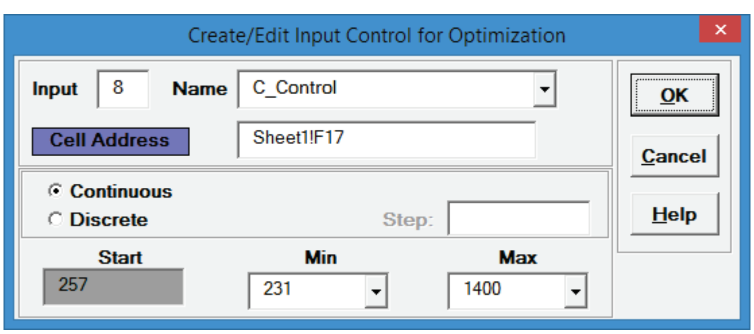
25. Click on cell **F17**. Select **Control**:



26. Click Input **Name** cell reference and specify cell **J17** containing the input control name "C_Control."

27. Click the **Min** value cell reference and specify cell **K17** containing the minimum optimization boundary value = 231.

28. Click the **Max** value cell reference and specify cell **L17** containing the maximum optimization boundary value = 1400:



29. Click **OK**. Hover the cursor on cell **F17** to view the comment displaying the input control settings:

DiscoverSim Control
Name: C_Control
Type: Continuous
Lower Bound: 231
Upper Bound: 1400

30. Click on cell **F17**. Click the **DiscoverSim Copy Cell** menu button Copy Cell (**Do not use Excel's Copy** – **it will not work!**).

31. Select cells **F18:F19**. Click the **DiscoverSim Paste Cell** menu button Paste Cell (**Do not use Excel's Paste** – **it will not work!**).

32. Review the Input Control comments in cells **F18** and **F19**.

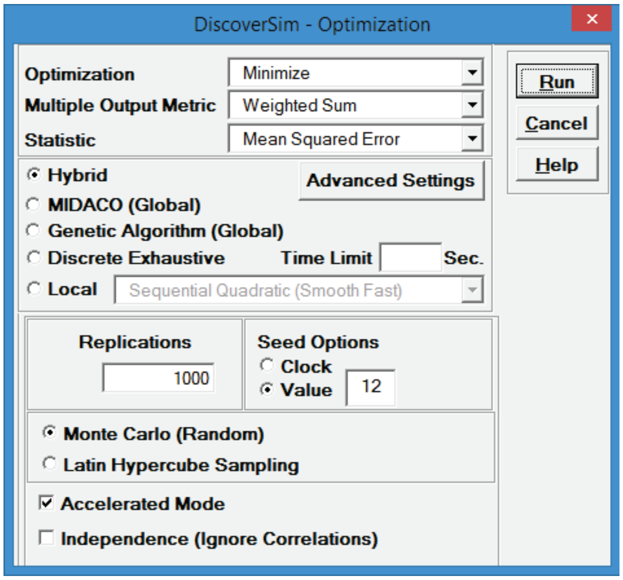
33. The completed model is shown below:

				Input Distribution Parameters		
Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	257	4.283333333	5
Control	R1	Filter Resistor 1 (ohms)	1060	1060	17.66666667	5
Control	R2	Filter Resistor 2 (ohms)	464	464	7.733333333	5
Noise	Vs	Transducer Voltage (mV)	15	15	0.0075	0.15
Noise	Rs	Transducer Resistance (ohms)	120	120	0.06	0.15
Noise	Rg	Galvanometer Resistance (ohms)	84	84	0.042	0.15
Noise	Gs	Galvanometer Sensitivity (mA/inch)	0.56364	0.56364	0.00028182	0.15

Output Name	DiscoverSim Output Responses	Specification
Deflection	3.062128625	3.0 inches +/- 1%
Frequency	6.874965344	6.84 Hz +/- 5%

34. Now we are ready to perform the optimization. Select **DiscoverSim > Run Optimization**:

35. Select “Minimize” for **Optimization Goal**, “Weighted Sum” for **Multiple Output Metric** and “Mean Squared Error” for **Statistic**. Select **Seed Value** and enter “12,” in order to replicate the optimization results given below. Set **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:



Mean Squared Error is also known as the Taguchi Loss Function, and is calculated as:

MSE (Loss) = $\text{Weight}_D[(\text{Mean}_D - \text{Target}_D)^2 + \text{StdDev}_D^2] + \text{Weight}_F[(\text{Mean}_F - \text{Target}_F)^2 + \text{StdDev}_F^2]$

We want both the Deflection and Frequency outputs to be on target with minimal variation. The weights were set when the outputs were created and are the same for each output (= 1).

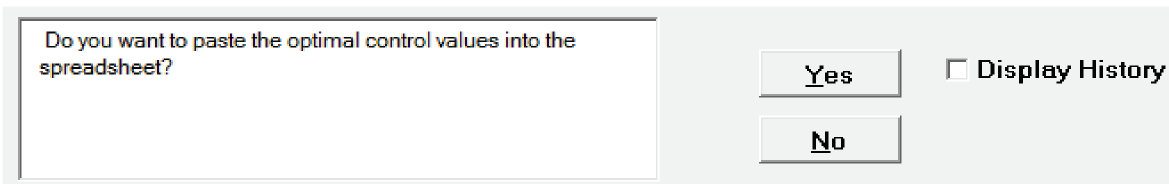
We will use the **Hybrid** optimization method which requires more time to compute, but is powerful to solve complex optimization problems.

36. Click **Run**. This hybrid optimization will take approximately 2 minutes.

37. The final optimal parameter values are given as:

```
*************************************************************************
                          Optimal Parameters
*************************************************************************

   Minimize Weighted Sum (Mean Squared Error) : 0.0077298695

   Parameter Values:

   C_Control: Value           249.9256
   R1_Control: Value         1499.8844
   R2_Control: Value          494.9780
```

38. You are prompted to paste the optimal values into the spreadsheet:

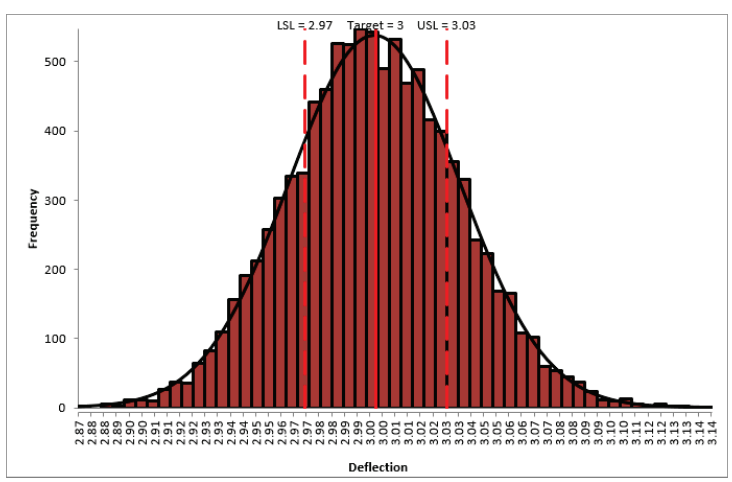


Click **Yes**. This replaces the nominal input control values with the new optimum values. Note that the displayed values for the input distributions at cells **E17** to **E19** are not overwritten, but when the simulation is run, the referenced mean and standard deviation given in **F17** to **G19** will be applied.

39. Select **Run Simulation**. Click **Report Options/Sensitivity Analysis**. Check **Sensitivity Charts** and **Correlation Coefficients**. Select **Seed Value** and enter “12,” in order to replicate the simulation results given below. Click **Run**.

40. The DiscoverSim Output Report shows a histogram, descriptive statistics and process capability indices for each output.

Deflection Report (after optimization):



Descriptive Statistics	
Count	10000
Mean	2.999
StDev	0.036069
Range	0.268
Minimum	2.874
25th Percentile (Q1)	2.975
50th Percentile (Median)	2.999
75th Percentile (Q3)	3.024
Maximum	3.142
95.0% CI Mean	2.998728 to 3.000142
95.0% CI Median	2.997842 to 2.999798
95.0% CI StDev	0.035576 to 0.036576

Normality Tests	
Anderson-Darling Normality Test	0.629654
p-value (A-D Test)	0.1009
Skewness	0.076103359
p-value (Skewness)	0.0019
Kurtosis	0.006561173
p-value (Kurtosis)	0.8744

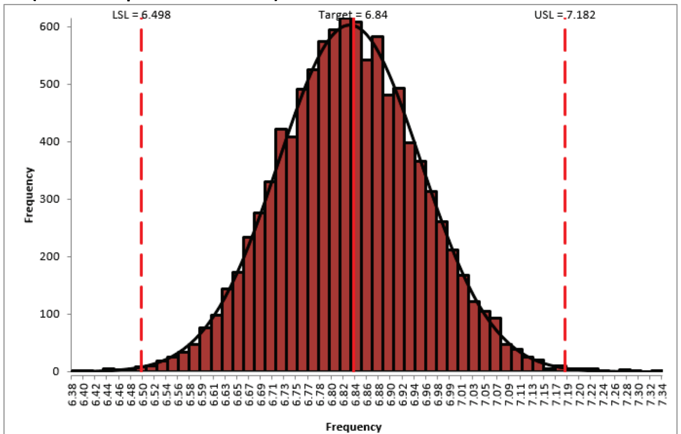
Process Capability Indices	
Pp	0.28
Ppu	0.28
Ppl	0.27
Ppk	0.27
Cpm	0.28

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	198383.10
ppm < LSL	207233.16
ppm Total	405616.26
% > USL	19.84
% < LSL	20.72
% Total (out of spec.)	40.56
% Total (within spec.)	59.44

Actual Performance (Empirical)	
ppm > USL	198000.00
ppm < LSL	207200.00
ppm Total	405200.00
% > USL	19.80
% < LSL	20.72
% Total (out of spec.)	40.52
% Total (within spec.)	59.48

From the histogram and capability report we see that the output Deflection is still not capable of meeting the specification requirements after optimization, however the mean is now on target (mean shifted from 3.062 to 2.999). The expected performance failure rate was reduced from 81.5% to 40.56% due to centering the mean. Since we only saw a slight reduction in the standard deviation (0.0366 to 0.0361), the component tolerances will need to be tightened. The sensitivity analysis will help us to focus on which components are critical and require tighter tolerances.

Frequency Report (after optimization):



Descriptive Statistics	
Count	10000
Mean	6.834
StDev	0.114806
Range	0.954
Minimum	6.384
25th Percentile (Q1)	6.756
50th Percentile (Median)	6.833
75th Percentile (Q3)	6.911
Maximum	7.338
95.0% CI Mean	6.831938 to 6.836439
95.0% CI Median	6.82985 to 6.835512
95.0% CI StDev	0.113236 to 0.116419

Normality Tests	
Anderson-Darling Normality Test	0.620111
p-value (A-D Test)	0.1065
Skewness	0.072815441
p-value (Skewness)	0.0030
Kurtosis	0.051440866
p-value (Kurtosis)	0.2940

Process Capability Indices	
Pp	0.99
Ppu	1.01
Ppl	0.98
Ppk	0.98
Cpm	0.99

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	1224.51
ppm < LSL	1703.96
ppm Total	2928.47
% > USL	0.12
% < LSL	0.17
% Total (out of spec.)	0.29
% Total (within spec.)	99.71

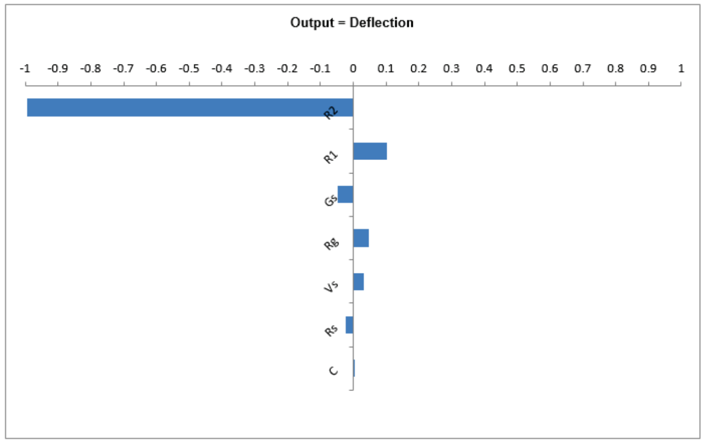
Actual Performance (Empirical)	
ppm > USL	2000.00
ppm < LSL	1200.00
ppm Total	3200.00
% > USL	0.20
% < LSL	0.12
% Total (out of spec.)	0.32
% Total (within spec.)	99.68

From the histogram and capability report we see that the output Frequency after optimization is now on target (mean shifted from 6.875 to 6.834) but still just capable of meeting the specification. The actual performance failure rate was reduced from 0.71% to 0.32% due to centering the mean. Since we did not see a reduction in the standard deviation, the component tolerances will need to be tightened. As with Deflection, the sensitivity analysis will help us to focus on which components are critical and require tighter tolerances.

41. Click on the **Sensitivity Correlations** sheet to view the sensitivity charts.

Deflection Sensitivity Chart:

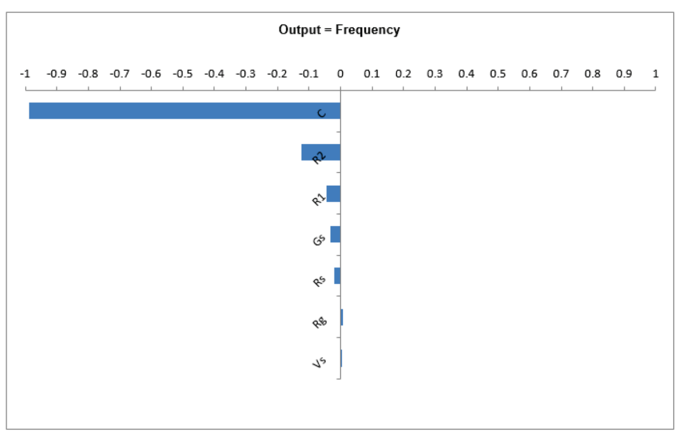
Inputs	Spearman Rank
R2	-0.993019529
R1	0.104066934
Gs	-0.048204823
Rg	0.046000726
Vs	0.031710276
Rs	-0.023935356
C	0.003716622



Resistor R2 is the dominant component affecting Deflection. A search at the electronic component supplier www.mouser.com, shows that 510 ohm resistors (¼ watt axial) with a 5% tolerance or 1% tolerance, cost approximately $0.04 each (qty. 100 pricing); resistors with a 0.1% tolerance are $0.29 to $1.00 each. Since R2 has such a strong influence on the variability we will use 0.1 % tolerance, even with the higher cost. Afterward, we will consider using 1% tolerance for R2 as a possible cost saving measure. Even though resistor R1 has negligible influence (on Deflection or Frequency below), we will use a 1% tolerance since the price is approximately the same as that of 5% tolerance resistors (5K ohms value).

42. Frequency Sensitivity Chart:

Inputs	Spearman Rank
C	-0.987509704
R2	-0.123303693
R1	-0.044722458
Gs	-0.031799641
Rs	-0.020881087
Rg	0.008667638
Vs	0.001092623



Capacitor C is the dominant component affecting Frequency, followed by Resistor R2. Unfortunately, capacitors larger than 230 uF will typically be electrolytic with a 20% tolerance or more. Specifying a tolerance of 5% or less will significantly increase the component cost. Another issue is that the number of available capacitor standard values is very limited compared to that of resistors. We need to reduce the variability of Frequency so we will specify a 2% tolerance (electrolytic > 10V axial). The lowest cost option available to us in the 230 to 300 uF range is a Vishay 280 uF @2%, at $11.79 each (qty. 101). If this cost is unacceptable then we would need to use a 20% tolerance and relax the output specifications for Frequency or accept the lower process capability/higher defect rate. The cost for a 250uF @20% is $1.84, qty. 100.

43. Now we will modify the control factor settings and rerun the optimization for practical tolerance design, starting with the fixed capacitor value of 280 uF @2% tolerance. The capacitor is selected first because of the limited selection for available capacitor standard values.

44. Enter the value 280 in control cell **F17**. Click on cell **H17** and change the tolerance value to 2. This will result in a reduced standard deviation value in cell **G17** for capacitor C.

Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Input Distribution Parameters: Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	280	1.866666667	2
Control	R1	Filter Resistor 1 (ohms)	1060	1499.884436	24.99807393	5
Control	R2	Filter Resistor 2 (ohms)	464	494.9779687	8.249632812	5
Noise	Vs	Transducer Voltage (mV)	15	15	0.0075	0.15
Noise	Rs	Transducer Resistance (ohms)	120	120	0.06	0.15
Noise	Rg	Galvanometer Resistance (ohms)	84	84	0.042	0.15
Noise	Gs	Galvanometer Sensitivity (mA/inch)	0.56364	0.56364	0.00028182	0.15

Input Control Parameters (Mean): Name	Min	Max
C_Control	231	1400
R1_Control	20	100000
R2_Control	0.01	525

Output Name	DiscoverSim Output Responses	Specification
Deflection	3.062128625	3.0 inches +/- 1%
Frequency	6.874965344	6.84 Hz +/- 5%

Note: The original **E17** cell value of 257 for input distribution C is not overwritten in the sheet display, but the referenced control value in **F17**, 280, will be used in simulation and further optimization.

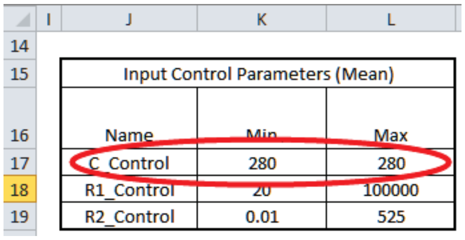
45. Click on cell **H18** and change the tolerance value to 1. This will result in a reduced standard deviation value in cell **G18** for resistor R1.

Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Input Distribution Parameters: Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	280	1.866666667	2
Control	R1	Filter Resistor 1 (ohms)	1060	1499.884436	4.999614787	1
Control	R2	Filter Resistor 2 (ohms)	464	494.9779687	8.249632812	5

46. Click on cell **H19** and change the tolerance value to 0.1. This will result in a reduced standard deviation value in cell **G19** for resistor R2.

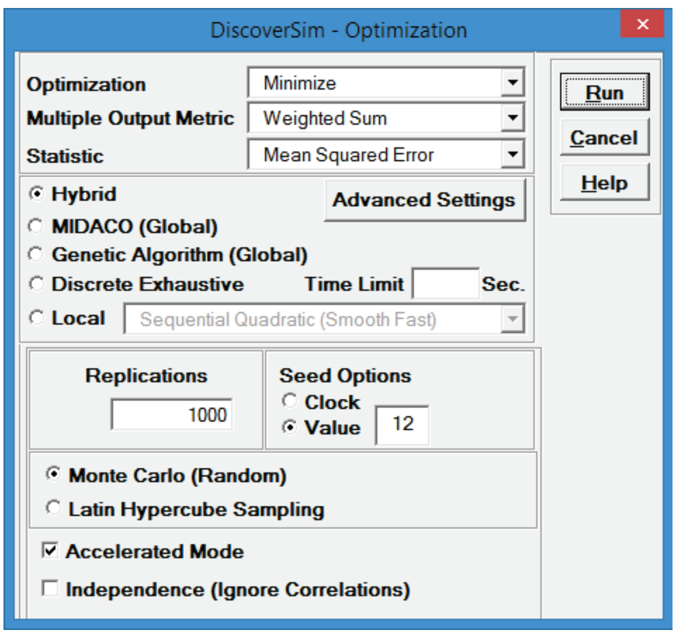
Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Input Distribution Parameters: Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	257	280	1.866666667	2
Control	R1	Filter Resistor 1 (ohms)	1060	1499.884436	4.999614787	1
Control	R2	Filter Resistor 2 (ohms)	464	494.9779687	0.164992656	0.1

47. Enter the Input Control min/max values for C in cells **K17** and **L17** as 280. This fixes the C value so that it will not vary in optimization.



48. Now we are ready to perform the optimization. Select **DiscoverSim > Run Optimization**:

49. Select "Minimize" for **Optimization Goal**, "Weighted Sum" for **Multiple Output Metric** and "Mean Squared Error" for **Statistic**. Select **Seed Value** and enter "12," in order to replicate the optimization results given below. Set **Replications** to 1000 to reduce the optimization time. All other settings will be the defaults as shown:

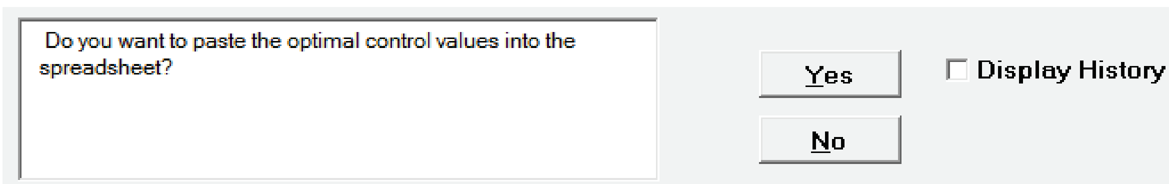


50. Click **Run**. This hybrid optimization will take approximately 1 minute.

51. The final optimal parameter values are given as:

```
***************************************************************
                     Optimal Parameters
***************************************************************

   Minimize Weighted Sum (Mean Squared Error) : 0.0011252746

   Parameter Values:

   R1_Control: Value          568.5673
   R2_Control: Value          432.2261
```

52. You are prompted to paste the optimal values into the spreadsheet:



Click **Yes**. This replaces the nominal input control values with the new optimum values.

53. For R2 we will use a ¼ watt axial 432 ohm 0.1% tolerance resistor, available at a cost of 0.64 each (qty 100). Later we will consider using a 432 ohm 1% tolerance resistor as a possible cost saving measure since they are only .04 each (qty 100).

54. For R1, we will use a ¼ watt axial 565 ohm 1% tolerance resistor for R1, which is the closest available standard value to 566.8 ohms.

55. Enter the value 565 in control cell **F18** for R1. The 1.0% tolerance specified in cell **H18** does not change. Enter the value 432 in control cell **F19** for R2. The 0.1% tolerance specified in cell **H19** does not change.

	A	B	C	D	E	F	G	H
14								
15						Input Distribution Parameters		
16		Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Mean	StdDev	% Tolerance (used to calculated StdDev)
17		Control	C	Filter Capacitor (uF)	257	280	1.866666667	2
18		Control	R1	Filter Resistor 1 (ohms)	1060	565	1.883333333	1
19		Control	R2	Filter Resistor 2 (ohms)	464	432	0.144	0.1

56. It is not necessary to "freeze" the R1 and R2 min/max values since we are no longer running optimization.

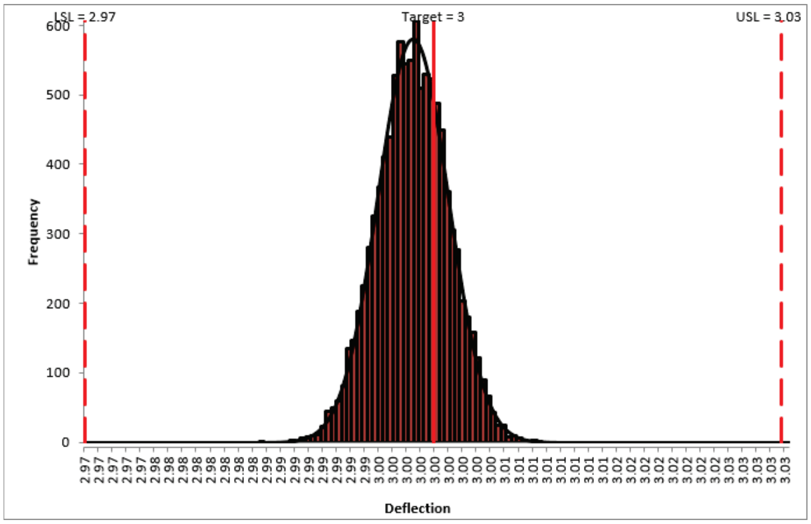
57. Now type in the C, R1 and R2 values in the Input Distributions cells **E17** to **E19**. Note that this is not required to run the simulation, but allows us to quickly check what the expected values will be for the outputs Deflection (at **E26**) and Frequency (at **E27**).

Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Input Distribution Parameters: Mean	StdDev	% Tolerance (used to calculated StdDev)
Control	C	Filter Capacitor (uF)	280	280	1.866666667	2
Control	R1	Filter Resistor 1 (ohms)	565	565	1.883333333	1
Control	R2	Filter Resistor 2 (ohms)	432	432	0.144	0.1
Noise	Vs	Transducer Voltage (mV)	15	15	0.0075	0.15
Noise	Rs	Transducer Resistance (ohms)	120	120	0.06	0.15
Noise	Rg	Galvanometer Resistance (ohms)	84	84	0.042	0.15
Noise	Gs	Galvanometer Sensitivity (mA/inch)	0.56364	0.56364	0.00028182	0.15

Output Name	DiscoverSim Output Responses	Specification
Deflection	2.998243931	3.0 inches +/- 1%
Frequency	6.844361319	6.84 Hz +/- 5%

58. Now we are ready to perform the simulation. Select **Run Simulation**. Click **Run**.

59. The DiscoverSim Output Report shows a histogram, descriptive statistics and process capability indices for each output and confirms that the RC design has dramatically improved:



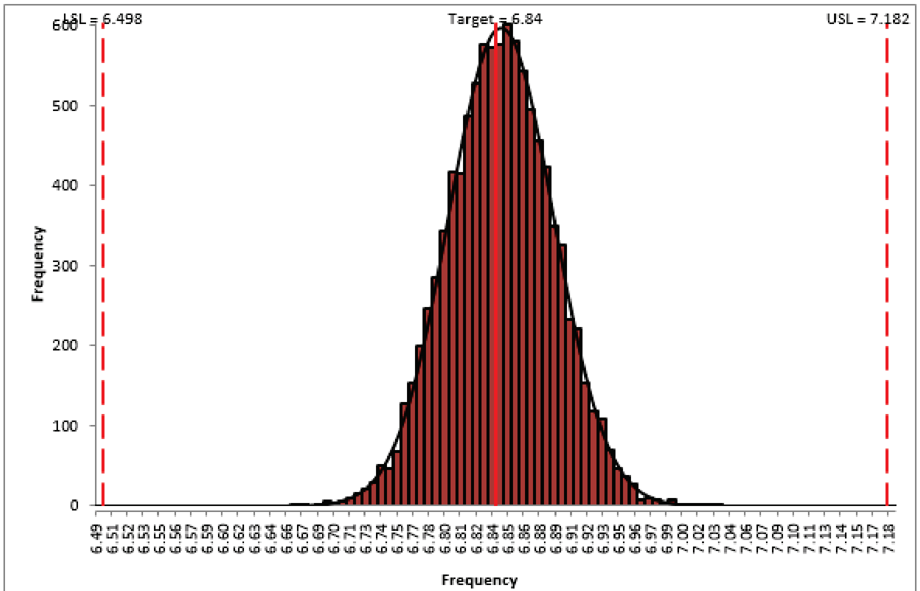
Descriptive Statistics	
Count	10000
Mean	2.998
StDev	0.003048
Range	0.024
Minimum	2.985
25th Percentile (Q1)	2.996
50th Percentile (Median)	2.998
75th Percentile (Q3)	3.000
Maximum	3.009
95.0% CI Mean	2.998192 to 2.998311
95.0% CI Median	2.998181 to 2.998344
95.0% CI StDev	0.003007 to 0.003091

Normality Tests	
Anderson-Darling Normality Test	0.492879
p-value (A-D Test)	0.2172
Skewness	-0.025593011
p-value (Skewness)	0.2959
Kurtosis	-0.103022058
p-value (Kurtosis)	0.0281

Process Capability Indices	
Pp	3.28
Ppu	3.47
Ppl	3.09
Ppk	3.09
Cpm	2.85

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00



Descriptive Statistics	
Count	10000
Mean	6.845
StDev	0.045646
Range	0.375
Minimum	6.663
25th Percentile (Q1)	6.814
50th Percentile (Median)	6.845
75th Percentile (Q3)	6.876
Maximum	7.038
95.0% CI Mean	6.844128 to 6.845918
95.0% CI Median	6.843852 to 6.846207
95.0% CI StDev	0.045022 to 0.046287

Normality Tests	
Anderson-Darling Normality Test	0.218980
p-value (A-D Test)	0.8380
Skewness	0.01934351
p-value (Skewness)	0.4295
Kurtosis	0.028661819
p-value (Kurtosis)	0.5472

Process Capability Indices	
Pp	2.50
Ppu	2.46
Ppl	2.53
Ppk	2.46
Cpm	2.48

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

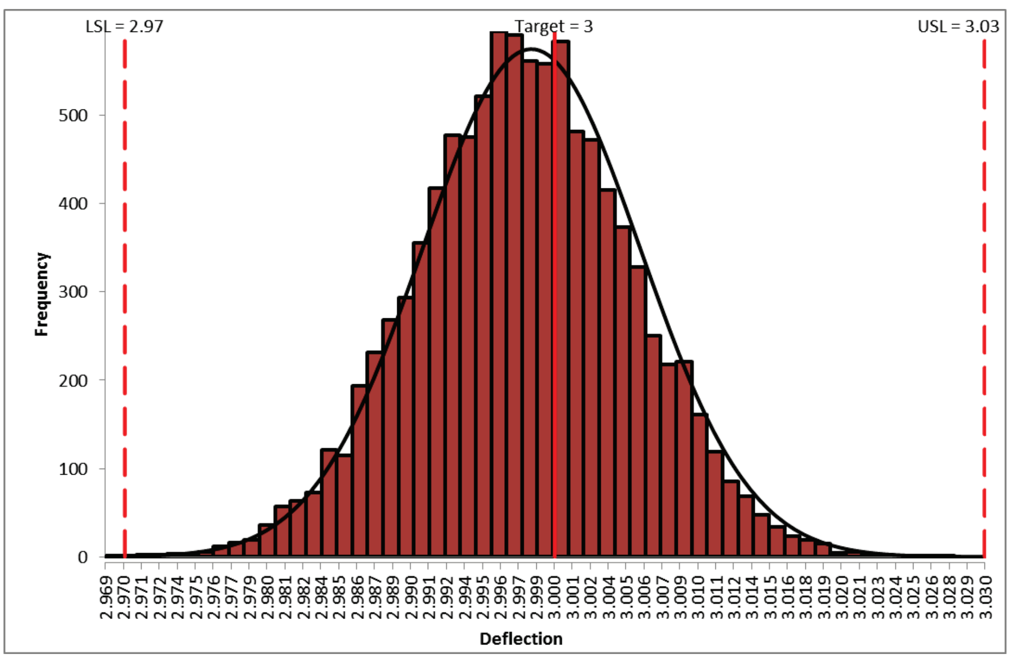
60. Now that we have achieved beyond Six Sigma quality for this design (Ppk > 1.5 for each output), we need to consider if we can relax the tolerance for R2 from 0.1% to 1%, thus reducing the component cost (from 0.64 each to .04 each in qty. 100).

61. Click on cell **H19** and change the tolerance value from 0.1 to 1, resulting in an increased standard deviation value in cell **G19** for resistor R2:

	A	B	C	D	E	F	G	H
14								
15						Input Distribution Parameters		
16		Factor Type	Factor Name	Factor Description	DiscoverSim Input Distributions	Mean	StdDev	% Tolerance (used to calculated StdDev)
17		Control	C	Filter Capacitor (uF)	257	280	1.866666667	2
18		Control	R1	Filter Resistor 1 (ohms)	1060	565	1.883333333	1
19		Control	R2	Filter Resistor 2 (ohms)	464	432	1.44	1

62. Now re-run the simulation. Select **Run Simulation**. Click **Run**.

63. The resulting simulation report for both outputs is shown below:



Descriptive Statistics	
Count	10000
Mean	2.998
StDev	0.007650
Range	0.058
Minimum	2.970
25th Percentile (Q1)	2.993
50th Percentile (Median)	2.998
75th Percentile (Q3)	3.003
Maximum	3.028
95.0% CI Mean	2.998052 to 2.998352
95.0% CI Median	2.997961 to 2.998299
95.0% CI StDev	0.007546 to 0.007758

Normality Tests	
Anderson-Darling Normality Test	0.190866
p-value (A-D Test)	0.8979
Skewness	0.009729621
p-value (Skewness)	0.6910
Kurtosis	0.018520007
p-value (Kurtosis)	0.6896

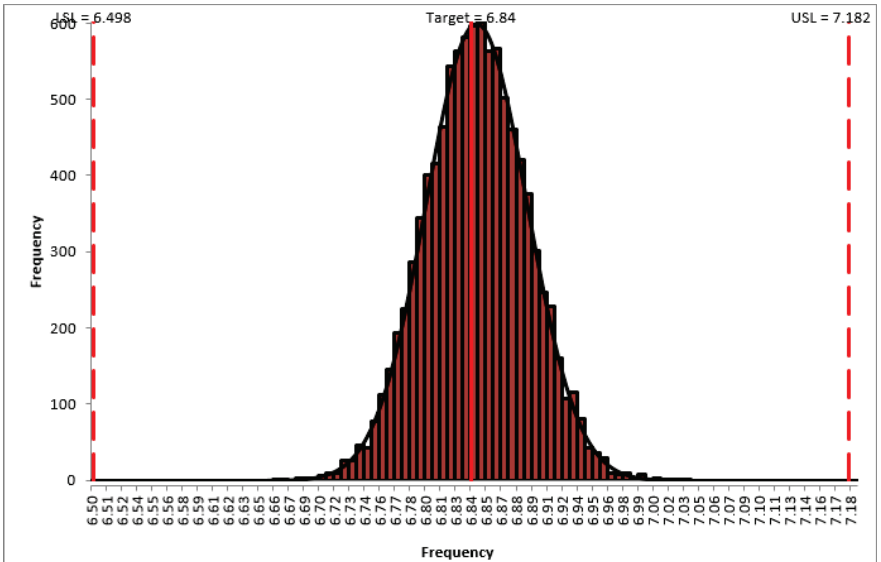
Process Capability Indices	
Pp	1.31
Ppu	1.39
Ppl	1.23
Ppk	1.23
Cpm	1.27

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	16.15
ppm < LSL	113.69
ppm Total	129.84
% > USL	0.00
% < LSL	0.01
% Total (out of spec.)	0.01
% Total (within spec.)	99.99

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	100.00
ppm Total	100.00
% > USL	0.00
% < LSL	0.01
% Total (out of spec.)	0.01
% Total (within spec.)	99.99

Sum of PPM Performance for All Outputs	
ppm (Expected assuming Normal)	129.8
ppm Weighted (Expected assuming Normal)	129.8
ppm (Actual)	100.0

As expected from our sensitivity analysis, we see that Ppk for Deflection has dropped from 3.11 to 1.23, and the expected ppm failure rate has increased from 0.0 to 129.84. (Here we also display the Sum of PPM report which includes the sum of ppm values for both Deflection and Frequency).



Descriptive Statistics	
Count	10000
Mean	6.845
StDev	0.045737
Range	0.378
Minimum	6.661
25th Percentile (Q1)	6.814
50th Percentile (Median)	6.845
75th Percentile (Q3)	6.876
Maximum	7.038
95.0% CI Mean	6.844106 to 6.845899
95.0% CI Median	6.844024 to 6.846276
95.0% CI StDev	0.045112 to 0.04638

Normality Tests	
Anderson-Darling Normality Test	0.198861
p-value (A-D Test)	0.8859
Skewness	0.016813131
p-value (Skewness)	0.4922
Kurtosis	0.033619685
p-value (Kurtosis)	0.4838

Process Capability Indices	
Pp	2.49
Ppu	2.46
Ppl	2.53
Ppk	2.46
Cpm	2.48

Expected Overall Performance (Assuming Normal Distribution)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

Actual Performance (Empirical)	
ppm > USL	0.00
ppm < LSL	0.00
ppm Total	0.00
% > USL	0.00
% < LSL	0.00
% Total (out of spec.)	0.00
% Total (within spec.)	100.00

We see that Ppk for Deflection has not changed from 2.43, so increasing the tolerance for R2 from 0.1% to 1% does not have any effect on Deflection.

64. Now that we have run both simulations, we see the trade-off between a difference in predicted ppm failure rate of 129 and a cost difference of 0.60 per unit, thus enabling us to make an informed decision of which tolerance to use for R2.

65. A similar analysis can be performed in the tolerance versus cost consideration for the capacitor.

Case Study 6 References

El-Haik, B.S. (2005) Axiomatic Quality: Integrating Axiomatic Design with Six-Sigma, Reliability, and Quality Engineering, Hoboken, N.J.: Wiley.

Feng, Q. (2008) “An integrated method of parameter design and tolerance design for multiple criteria systems,” *Int. J. Computer Applications in Technology*, Vol. 32, No. 2, pp.118–127.

Ferryanto, L. (2007) “Analytical DFSS for multiple response products,” *Int. J. Six Sigma and Competitive Advantage*, Vol. 3, No. 1, pp.13–32.

Filippone, S.F. (1989) “Using Taguchi methods to apply the axioms of design,” *Robotics and Computer-Integrated Manufacturing*, Vol. 6, pp.133–142.

Rinderle, J.R. (1982) *Measures of Functional Coupling in Design*, Thesis (Ph.D.)--Massachusetts Institute of Technology, Dept. of Mechanical Engineering.

Suh, N.P. (1990) *The Principles of Design*. New York, NY: Oxford Press.

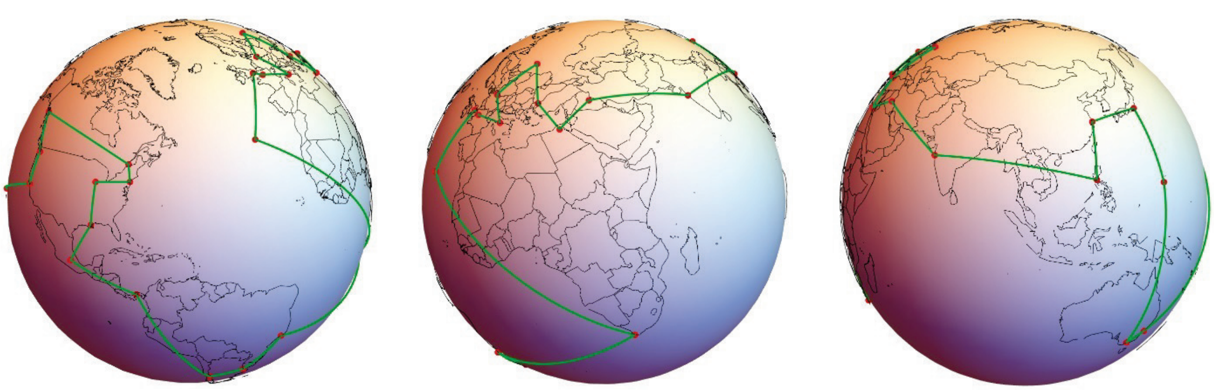
Case Study 7 – The Travelling Salesperson Problem (TSP)

Introduction: Optimizing Travel to 30 International Cities

The **travelling salesperson problem** (**TSP**) asks the following question: *"Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?"* The problem was first formulated in 1930 and is one of the most intensively studied problems in optimization. It is used as a benchmark for many optimization methods. The TSP has several applications such as planning, logistics, and the manufacture of microchips. Slightly modified, it appears as a sub-problem in many areas, such as DNA sequencing. In these applications, the concept city represents, for example, customers, soldering points, or DNA fragments, and the concept distance represents travelling times or cost, or a similarity measure between DNA fragments. The TSP also appears in astronomy, as astronomers observing many sources will want to minimize the time spent moving the telescope between the sources. In many applications, additional constraints such as limited resources or time windows may be imposed. (Ref.: https://en.wikipedia.org/wiki/Travelling_salesman_problem).

The books by Applegate and Cook (see Case Study 7 References) and the University of Waterloo TSP web site (http://www.math.uwaterloo.ca/tsp/index.html) are excellent resources for further study.

We will now consider a 30 City TSP problem. The data is from city distance dataset HA30 which describes the airline distances, in hundreds of miles, between 30 international cities: https://people.sc.fsu.edu/~jburkardt/datasets/cities/cities.html. A 30 city problem has 29!/2 = 4.42e30 possible solutions, so obviously this is not something that can be solved using the Discrete Exhaustive option! The goal is to achieve the best known solution minimum of 503 (50,300 miles), as given in Ouyang (2013):



The 3D plots were created using Mathematica with code adapted from Dr. Ben Nolting: (http://datavoreconsulting.com/programming-tips/a-traveling-salesman-on-a-sphere-pitbulls-arctic-adventure/).

In DiscoverSim Version 2.1, the **Genetic Algorithm** has improved handling of the **DSIM_IsAllDifferent** function, so we will use the Genetic Algorithm to solve this TSP problem. The DSIM_IsAllDifferent function will be used in a constraint in order to ensure that the only solutions considered are unique city sequences.

Since this is a deterministic optimization problem, we technically only need a single replication, but DiscoverSim requires a minimum of 2 replications for internal consistency on stochastic problems. If the model included Input Distributions to represent uncertainty (for example if we were using travel times rather than distance), then we would typically use 1000 replications in the optimization settings.

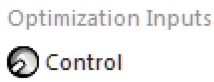
Summary of DiscoverSim Features Demonstrated in Case Study 7:
• Create Discrete Input Controls • DiscoverSim Copy/Paste Cell • Insert DSIM_IsAllDifferent function • Create a Constraint that references the DSIM_IsAllDifferent function • Run Optimization with Genetic Algorithm

Optimizing Travel to 30 International Cities with DiscoverSim

1. Open the workbook **30 International Cities TSP**. This is a matrix showing the distance between cities in hundreds of airline miles. The city names are shown in column B and row 3. DiscoverSim Discrete Input Controls will be created in cells **C36** to **AE36**. The output, Total Distance Travelled, will be specified in cell **B40**. A constraint will be added at cell **B42** that references a DSIM_IsAllDifferent function at cell **B41**. Excel's OFFSET function is used to extract the city names and distances for the selected Input Control values. Row 37 shows the names of the tour cities in the order visited. Row 38 shows the distances between the tour cities.

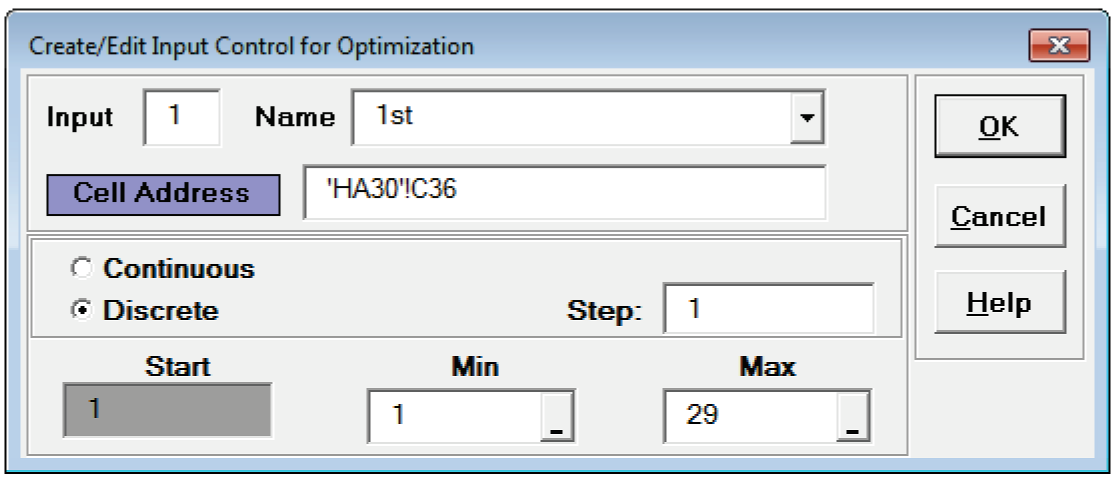
	A	B	C	D	E	F	G	H	I
1	30 International Cities (HA30) - Airline Distance Matrix in Hundreds of Miles								
2									
3			Azores	Baghdad	Berlin	Bombay	Buenos Aires	Cairo	Capetown
4		Azores	0	39	22	59	54	33	57
5		Baghdad	39	0	20	20	81	8	49
6		Berlin	22	20	0	39	74	18	60
7		Bombay	59	20	39	0	93	27	51
8		Buenos Aires	54	81	74	93	0	73	43
9		Cairo	33	8	18	27	73	0	45
10		Capetown	57	49	60	51	43	45	0
11		Chicago	32	64	44	81	56	61	85
12		Guam	89	63	71	48	104	71	88
13		Honolulu	73	84	73	80	76	88	115
14		Istanbul	29	10	11	30	76	8	52
15		Juneau	46	61	46	69	77	63	103
16		London	16	25	6	45	69	22	60
17		Manila	83	49	61	32	111	57	75
18		Melbourne	120	81	99	61	72	87	64
19		Mexico City	45	81	61	97	46	77	85
20		Montreal	24	58	37	75	56	54	79
21		Moscow	32	16	10	32	84	18	63
22		New Orleans	36	72	51	89	49	68	83
23		New York	25	60	40	78	53	56	78
24		Panama City	38	78	59	97	33	71	70
25		Paris	16	24	5	44	69	20	58
26		Rio de Janeiro	43	69	62	83	12	61	38
27		Rome	21	18	7	38	69	13	52
28		San Francisco	50	75	57	84	64	75	103
29		Santiago	57	88	78	100	7	80	49
30		Seattle	46	68	51	77	69	68	102
31		Shanghai	72	44	51	31	122	52	81
32		Sydney	121	83	100	63	73	90	69
33		Tokyo	73	52	56	42	114	60	92
34									
35		Start	1st	2nd	3rd	4th	5th	6th	7th
36	DiscoverSim Input Controls (Discrete 1 to 29):		1	2	3	4	5	6	7
37	Name of City in Tour:	Azores	Baghdad	Berlin	Bombay	Buenos Aires	Cairo	Capetown	Chicago
38	Distance between Cities:		39	20	39	93	73	45	85
39									
40	Total Distance Travelled:	1616							
41	Insert DSIM_IsAllDifferent:								
42	DiscoverSim Constraint:								

2. Click on cell **C36**. Select **Control**:

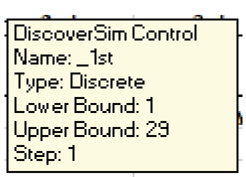


3. Click Input **Name** cell reference  and specify cell **C35** containing the input control name "1st".

4. Select **Discrete**. Set **Step** = 1, the **Min** value = 1 and the **Max** value = 29 as shown. There are 30 cities but the first city is defined as Azores, so only 29 need to be solved.



5. Click **OK**. Hover the cursor on cell **C36** to view the comment displaying the input control settings:



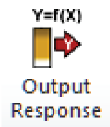
6. With cell **C36** selected, click the **DiscoverSim Copy Cell** menu button  (**Do not use Excel's Copy – it will not work!**).

7. Select cells **D36:AE36** (do not include the empty cell **AF36**). Click the **DiscoverSim Paste Cell** menu button  (**Do not use Excel's Paste – it will not work!**).

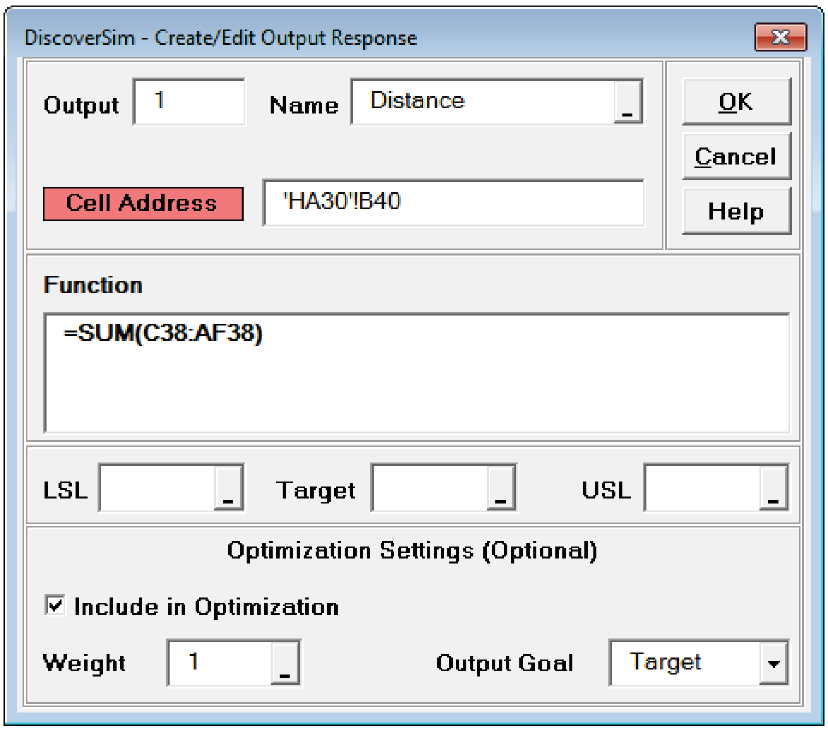
8. Review the Input Control comments in cells **D36** to **AE36**.

9. Now we will specify the output. Click on cell **B40**. Note that the cell contains the Excel formula for Total Distance Travelled:
 =SUM(C38:AF38).

10. Select **DiscoverSim** > **Output Response**:

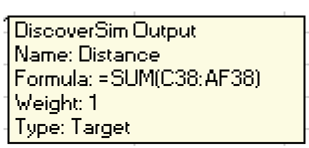


11. Enter the output **Name** as “Distance.” We will not enter any specifications.

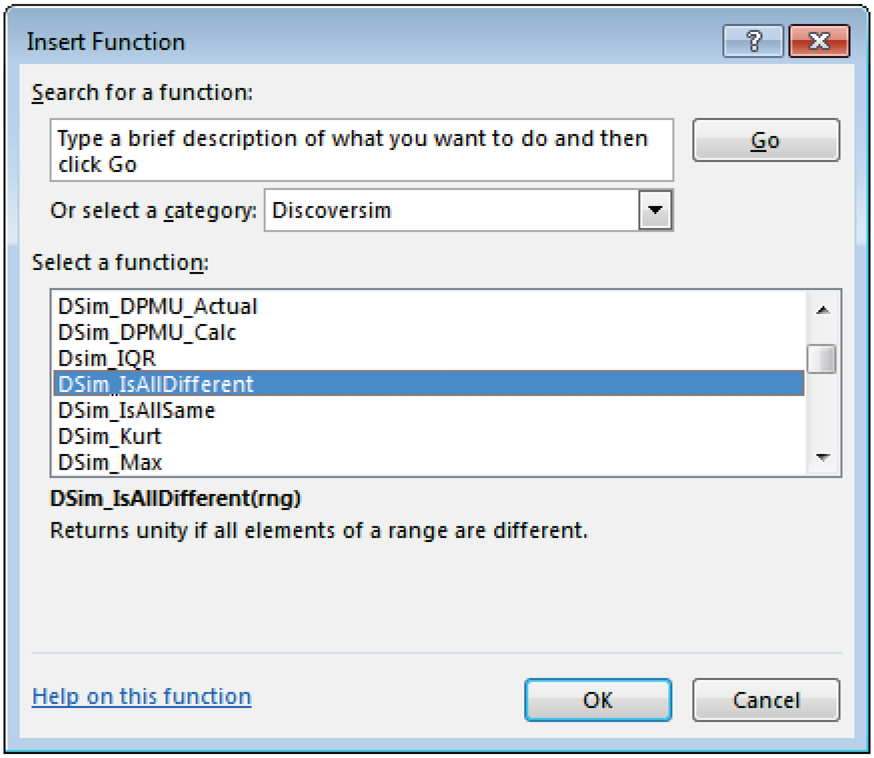


Click **OK**.

12. Hover the cursor on cell **B40** to view the DiscoverSim Output information.

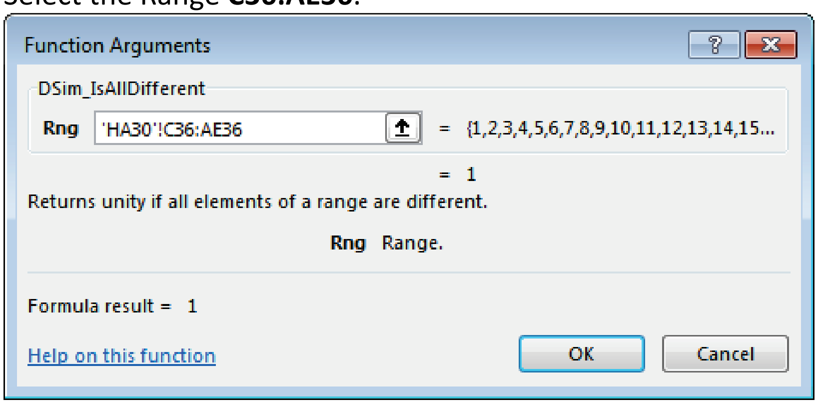


13. Click on cell **B41** and select **DiscoverSim > DSIM Function**:



Scroll down to select *DSIM_IsAllDifferent*. Click **OK**.

14. Select the Range **C36:AE36**.



Click **OK**.

15. The DSIM function formula is now in cell **B41**:

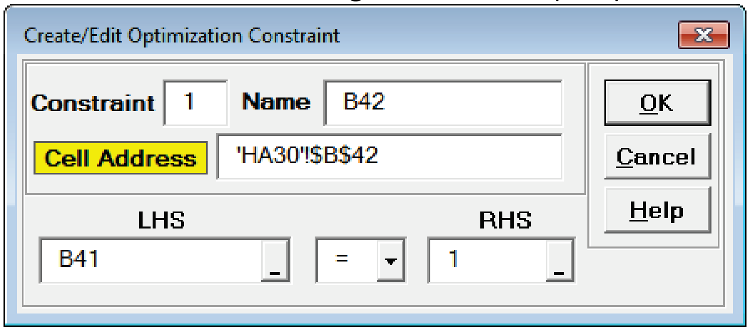
=DSim_IsAllDifferent('HA30'!C36:AE36)

which returns a 1 since the default input control values are all different.

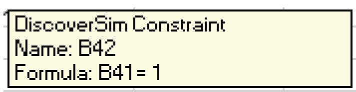
Note: All Discrete Input Controls should be included in the DSIM_IsAllDifferent range. DiscoverSim does not support a mixture of having some Discrete Input Controls constrained by DSIM_IsAllDifferent and some not constrained. (However, it is possible to have a mixture with Continuous Input Controls).

16. Now we will add a constraint that references the DSIM function. Click on cell **B42**. Select **Constraint**:

17. Enter **B41** in the “Left Hand Side” (**LHS**) or click the **LHS** cell reference and select **B41**. Select **=** . Enter 1 in the “Right Hand Side” (**RHS**).



18. Click **OK**. Review cell **B41:**

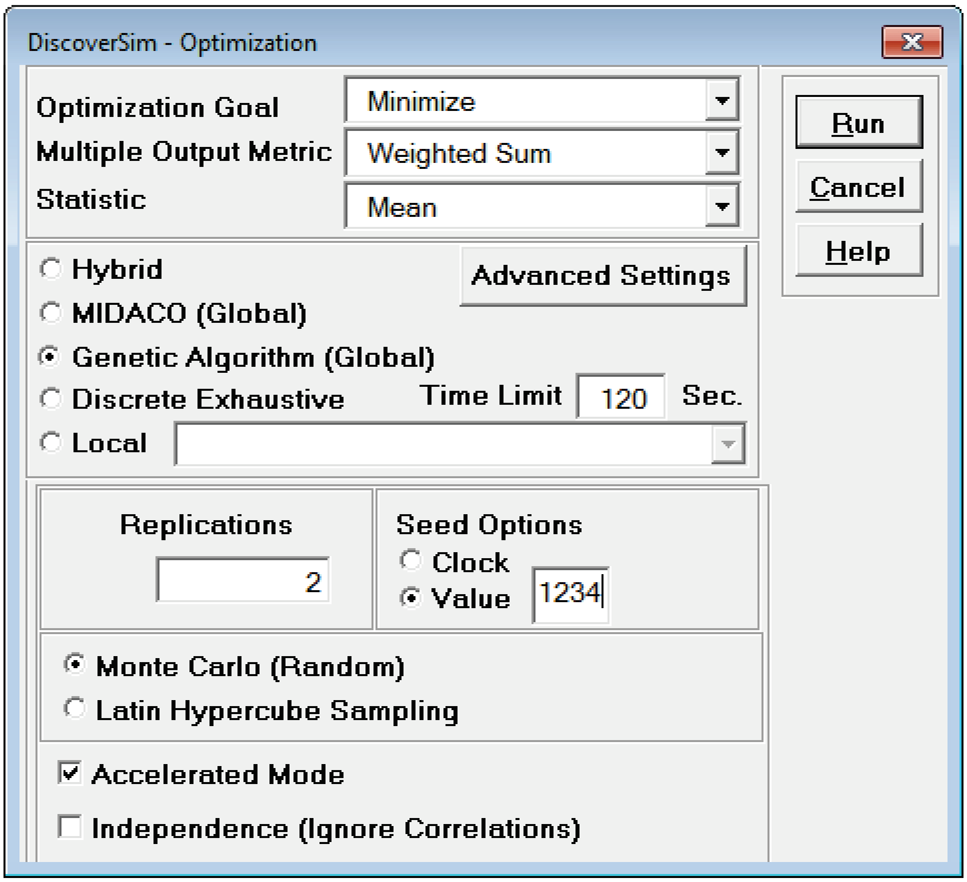


19. The completed model is shown below:

	A	B	C	D	E	F	G	H	I
1	30 International Cities (HA30) - Airline Distance Matrix in Hundreds of Miles								
2									
3			Azores	Baghdad	Berlin	Bombay	Buenos Aires	Cairo	Capetown
4		Azores	0	39	22	59	54	33	57
5		Baghdad	39	0	20	20	81	8	49
6		Berlin	22	20	0	39	74	18	60
7		Bombay	59	20	39	0	93	27	51
8		Buenos Aires	54	81	74	93	0	73	43
9		Cairo	33	8	18	27	73	0	45
10		Capetown	57	49	60	51	43	45	0
11		Chicago	32	64	44	81	56	61	85
12		Guam	89	63	71	48	104	71	88
13		Honolulu	73	84	73	80	76	88	115
14		Istanbul	29	10	11	30	76	8	52
15		Juneau	46	61	46	69	77	63	103
16		London	16	25	6	45	69	22	60
17		Manila	83	49	61	32	111	57	75
18		Melbourne	120	81	99	61	72	87	64
19		Mexico City	45	81	61	97	46	77	85
20		Montreal	24	58	37	75	56	54	79
21		Moscow	32	16	10	32	84	18	63
22		New Orleans	36	72	51	89	49	68	83
23		New York	25	60	40	78	53	56	78
24		Panama City	38	78	59	97	33	71	70
25		Paris	16	24	5	44	69	20	58
26		Rio de Janeiro	43	69	62	83	12	61	38
27		Rome	21	18	7	38	69	13	52
28		San Francisco	50	75	57	84	64	75	103
29		Santiago	57	88	78	100	7	80	49
30		Seattle	46	68	51	77	69	68	102
31		Shanghai	72	44	51	31	122	52	81
32		Sydney	121	83	100	63	73	90	69
33		Tokyo	73	52	56	42	114	60	92
34									
35		Start	1st	2nd	3rd	4th	5th	6th	7th
36	DiscoverSim Input Controls (Discrete 1 to 29):		1	2	3	4	5	6	7
37	Name of City in Tour:	Azores	Baghdad	Berlin	Bombay	Buenos Aires	Cairo	Capetown	Chicago
38	Distance between Cities:		39	20	39	93	73	45	85
39									
40	Total Distance Travelled:	1616							
41	Insert DSIM_IsAllDifferent:	1							
42	DiscoverSim Constraint:	TRUE							

20. Now we are ready to perform the optimization. Select **DiscoverSim > Run Optimization**:

21. Select "Minimize" for **Optimization Goal**, "Weighted Sum" for **Multiple Output Metric** and "Mean" for **Statistic**. Select **Genetic Algorithm.** Set **Replications** to 2. Select **Seed Value** and enter "1234" in order to replicate the optimization results given below. All other settings will be the defaults as shown:



Note: As mentioned above, since this is a deterministic problem, we technically only need a single replication, but DiscoverSim requires a minimum of 2 replications for internal consistency on stochastic problems.

22. Click **Run**. This optimization will take approximately 4 minutes. To save time, when the solution reaches the best known value of 503, click the **Interrupt** button.

23. The final optimal parameter values are given as:

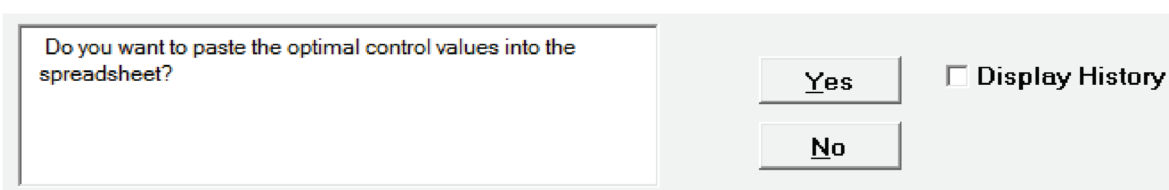
```
*************************************************************
                 Optimal Parameters
*************************************************************

  Minimize Weighted Sum (Mean) :        503.00000

  Parameter Values:

  _1st: Value           12.0000
  _2nd: Value           21.0000
  _3rd: Value           23.0000
  _4th: Value            2.0000
  _5th: Value           17.0000
  _6th: Value           10.0000
  _7th: Value            5.0000
  _8th: Value            1.0000
  _9th: Value            3.0000
  _10th: Value          13.0000
  _11th: Value          27.0000
  _12th: Value          29.0000
  _13th: Value           8.0000
  _14th: Value          14.0000
  _15th: Value          28.0000
  _16th: Value           9.0000
  _17th: Value          24.0000
  _18th: Value          26.0000
  _19th: Value          11.0000
  _20th: Value          16.0000
  _21st: Value          19.0000
  _22nd: Value           7.0000
  _23rd: Value          18.0000
  _24th: Value          15.0000
  _25th: Value          20.0000
  _26th: Value          25.0000
  _27th: Value           4.0000
  _28th: Value          22.0000
  _29th: Value           6.0000
```

As discussed in the introduction, the final solution of 503 (50,300 miles) is the best known solution for this 30 International City problem.

24. We can see that all of the values are all different satisfying the constraint using DSIM_IsAllDifferent and the "Amount Violated" is 0.

```
Constraint Equations evaluated at the optimum
Constraint Methodology:  Oracle

   Constraint       Type        Amount Violated

      1              eq             0.0000
```

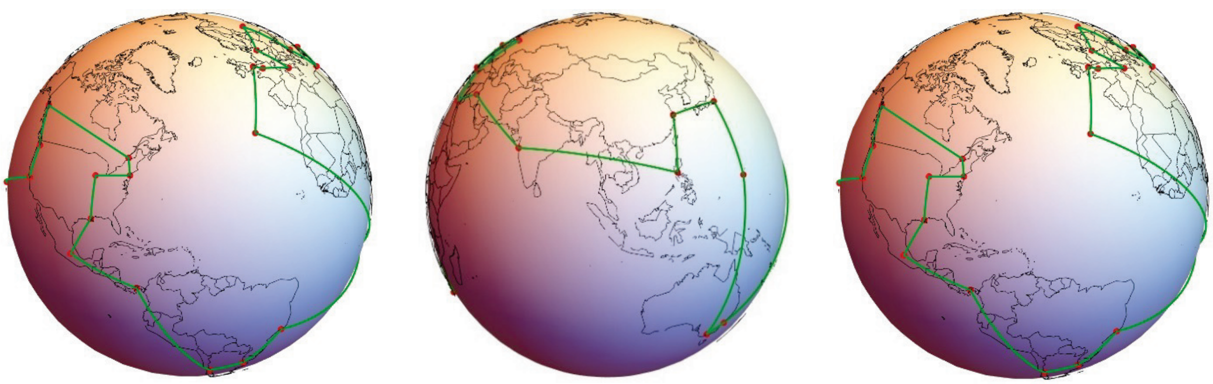
25. You are prompted to paste the optimal values into the spreadsheet:



26. Click **Yes**. This replaces the input control values in cells **C36** to **AE36**.with the optimum values.

27. The final city tour is given in cells **B37** to **AF37:**
Azores → London →Paris → Rome → Berlin → Moscow → Istanbul → Cairo → Baghdad → Bombay → Manila → Shanghai → Tokyo → Guam → Melbourne → Sydney → Honolulu → San Francisco → Seattle → Juneau → Montreal → New York → Chicago → New Orleans → Mexico City → Panama City → Santiago → Buenos Aires → Rio de Janeiro → Capetown → Azores

28. This matches the known best tour given in Ouyang (2013):



29. If one wanted to evaluate flight times with uncertainty, rather than distances, then input distributions could be added to create a stochastic model.

30. Another example where DSIM_IsAllDifferent would be used in a constraint is Stochastic Job Scheduling with Due Dates

Case Study 7 References

Applegate, D., Bixby, R. E., Chvátal, V., & Cook, W. (2006). *The traveling salesman problem: A computational study*. Princeton: Princeton University Press.

Cook, W. (2011). *In pursuit of the salesman: Mathematics at the limits of computation*. Princeton: Princeton University Press.

X. Ouyang, Y. Zhou, Q. Luo, and H. Chen, "A novel discrete cuckoo search algorithm for spherical traveling salesman problem," *Applied Mathematics & Information Sciences*, vol. 7, no. 2, pp. 777–784, 2013, http://www.naturalspublishing.com/files/published/xk232153a2f4jw.pdf.

DiscoverSim™ Appendix: Statistical Details for Distributions and Optimization Methods

Statistical Details for Distributions and Optimization Methods

DiscoverSim™ Engine and Excel Formula Interpreter

DiscoverSim uses Excel for the user interface but random number generation, distribution fitting parameter estimation, accelerated simulation and optimization are performed using the Gauss Mathematical and Statistical Engine by Aptech (www.aptech.com). GAUSS is a fast matrix programming language widely used by econometricians and financial analysts since 1984. Communication between Excel and the GAUSS Engine is performed with Mercury by Econotron Software (www.econotron.com). Econotron Software has developed interprocess communication tools for a number of applications based on the Gauss Engine.

Accelerated mode simulation and optimization use DiscoverSim's Excel Formula Interpreter. This is based on a Common Object Interface (COI) by Econotron Software to implement an Excel spreadsheet in GAUSS with speed increases up to 40 times.

The interpreter supports the majority of all Excel numerical functions including those in the Analysis Toolpak, but currently does not support the following:

- Excel 2007 CUBE functions
- Excel 2010 AGGREGATE, NETWORKDAYS.INTL, WORKDAYS.INTL
- Excel 2013 ARABIC, ENCODEURL, FILTERXML, FORMULATEXT, IFNA, ISFORMULA, SHEET, SHEETS, UNICHAR, UNICODE, WEBSERVICE
- Excel 2016 FORECAST
- CONVERT, GETPIVOTDATA, HYPERLINK
- ISTEXT, ISNONTEXT, ISLOGICAL, ISREF
- PowerPivot Data Analysis Expressions (DAX) Functions such as ISO.CEILING
- Third party vba functions

Functions supported but with limitations:

- Excel functions that return an array or non-single value (e.g. LINEST, LOGEST, MINVERSE, MDETERM, MMULT, MUNIT, STANDARDIZE, TRANSPOSE and TREND) are limited to intermediate calculations. The final output result must be a scalar.
- Excel functions that return a string value are limited to intermediate calculations. The final result must be a scalar.

If the DiscoverSim interpreter sees a function that it does not support, you will be prompted to use Excel's Native mode, where the transfer function, Y = f(X), calculations are performed in Excel. Note that in this native mode, DiscoverSim's random input distributions are still generated with GAUSS.

The Excel Formula Interpreter can be validated using **Accelerated Mode - Run Validation using Native Excel.** This runs a validation test to compare Accelerated Mode versus Native Excel. Each output is assessed by comparing the simulation means. If the worst case relative difference is less than or equal to 1e-10%, the test passes and the status is “Success.” If the relative difference is between 1e-10% to 1e-4%, the status is “Good,” and if greater than or equal to 1e-4%, the test status is “Poor.”

DiscoverSim can quickly perform millions of simulation replications, but is limited by system memory. If the user chooses to store the simulation data in Excel, then Excel’s row and column limitations will also apply (1,048,576 rows and 16,384 columns). For large simulation and/or optimization models, 64 bit Excel is highly recommended.

Input Distribution Random Number Generation

A random number generator (RNG) is required when using Monte Carlo simulations in order to draw random samples from the specified distribution. On a computer, the creation of a random number involves an algorithm that can automatically create long runs of numbers with good random properties. These algorithms are called pseudo-random number generators. The state of the RNG after each iteration is used as an input for the generation of the next random number. The initial state is called a seed; this seed can be derived using the system clock, or it can be specified by the user. In the latter case, the same sequence of numbers will be created each time the same seed is specified, which can be useful, for example, in a classroom setting where you want all students to obtain the same results. Stochastic optimization requires a fixed seed in order to avoid “chatter” that would result in inconsistent comparisons.

A good RNG will produce a long run of numbers which are independent, so that there is no correlation between successive numbers. Eventually, the sequence of numbers will repeat, since eventually the seed will repeat. The RNG used in DiscoverSim is based on a recur-with-carry “KISS+Monster” random number generator developed by George Marsaglia. This algorithm produces random integers between 0 and $2^{32} - 1$ and has a period of 10^{8859}. This implementation of the KISS+Monster algorithm has been tested and passes all of the Diehard tests (Marsaglia, 2003).

Marsaglia, G. (2003). “Random Number Generators,” *Journal Of Modern Applied Statistical Methods*, Vol 2(1), pp. 2-13.

Input Distribution Correlations

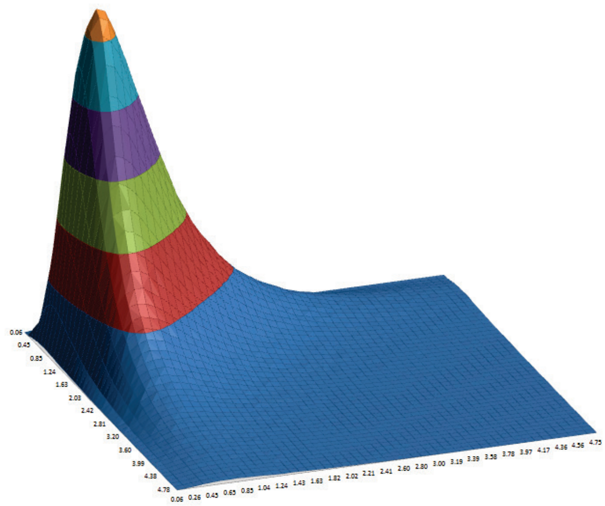
If input distribution correlations are specified, DiscoverSim utilizes correlation copulas to achieve the desired Spearman Rank correlation values.

Copulas are functions that describe dependencies among variables, and provide a way to create distributions to model correlated multivariate data. Using a copula a data analyst can construct a multivariate distribution by specifying marginal univariate distributions, and choosing a particular copula to provide a correlation structure between variables. There are a number of different families of copulas; in this context we use a Gaussian copula.

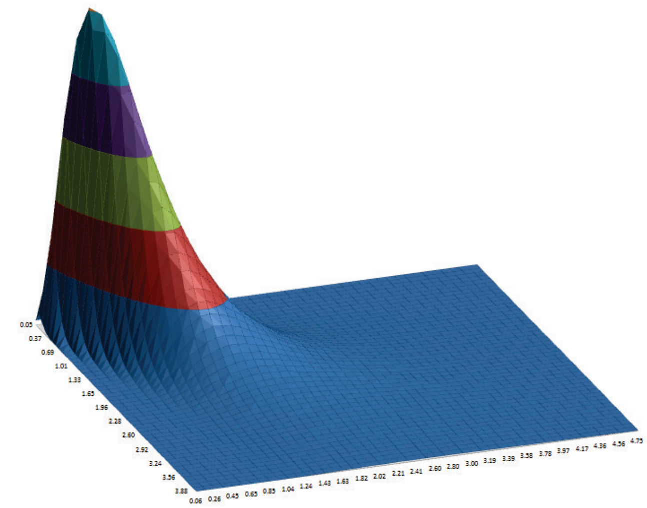
To take the simplest example, we would first generate pairs of values from a bivariate normal distribution - this is easily done by multiplying two independent normal draws using the Cholesky matrix. The copula is then generated simply by taking the CDF of each of the vectors of the bivariate draw, so that each is uniformly distributed U(0,1). Finally, applying the inverse CDF of any distribution, e.g., F, to a U(0,1) random variable results in a random variate whose distribution is exactly F, with the specified correlation structure. This is known as the Inverse method. For further details, see http://en.wikipedia.org/wiki/Copula_(statistics).

The following surface plots illustrate how a correlation copula changes the shape of a bivariate distribution (these plots were created using Excel's 3D Surface Plot):

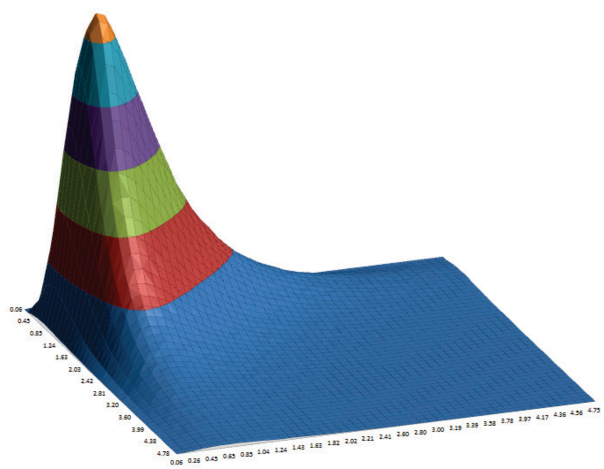
Surface Plot of Bivariate F Distribution



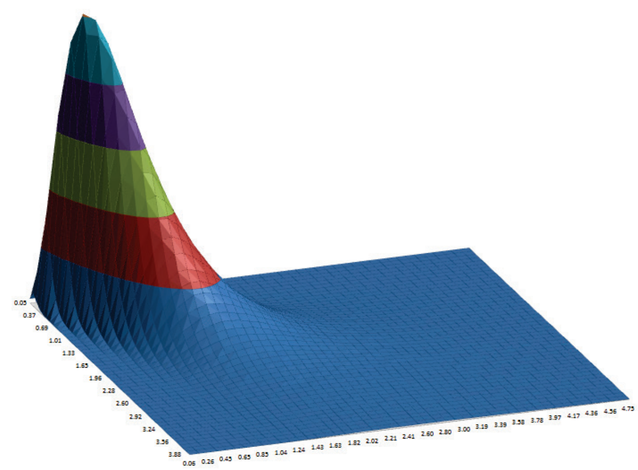
Surface Plot of Bivariate F Distribution with Copula (rho=0.8)



Surface Plot of Bivariate F Distribution



Surface Plot of Bivariate F Distribution with Copula (rho=0.8)



Input Distribution Sampling

DiscoverSim supports two sampling methodologies: Monte Carlo and Latin Hypercube.

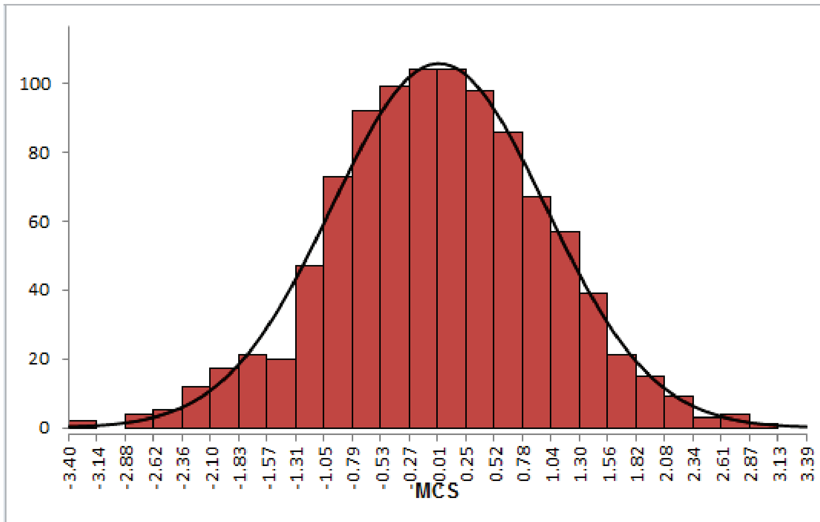
Monte Carlo Sampling (MCS) involves the repeated drawing of random vectors from randomly distributed variates with the specified distribution. New sample points are generated without taking into account the previously generated sample points. When a correlation is specified between the draws, the sampling occurs from the individual copulas.

Latin hypercube sampling (LHS) uses a technique known as "stratified sampling without replacement." The probability distribution is split into n intervals of equal probability, where n is the number of samples that are to be performed on the model. For a simulation with k variates, the hypercube will contain n rows with k columns. As the simulation progresses each of the n intervals in each of the k columns is sampled once. Once the sample is generated, the uniform sample from a column can be transformed to any distribution by using the quantile functions. Different columns can have different distributions. For further details, see: http://en.wikipedia.org/wiki/Latin_hypercube_sampling.

When an identity matrix is specified for the correlation matrix, the basic LHS process is adjusted such that the sampling from each column is nearly orthogonal. This requires that all the random samples over the k variates be generated simultaneously. A similar situation arises when correlation is specified between the sampled variates.

Unlike simple random sampling, this method ensures a full coverage of the range of each variable by maximally stratifying each marginal distribution. Thus LHS has the advantage of generating a set of samples that more precisely reflect the shape of a sampled distribution than pure random MCS and therefore can perform well with fewer replications (n = 1000 is recommended).

The following Histograms and Descriptive Statistics illustrate MCS and LHS for a normal distribution N(0,1) with n = 1000:



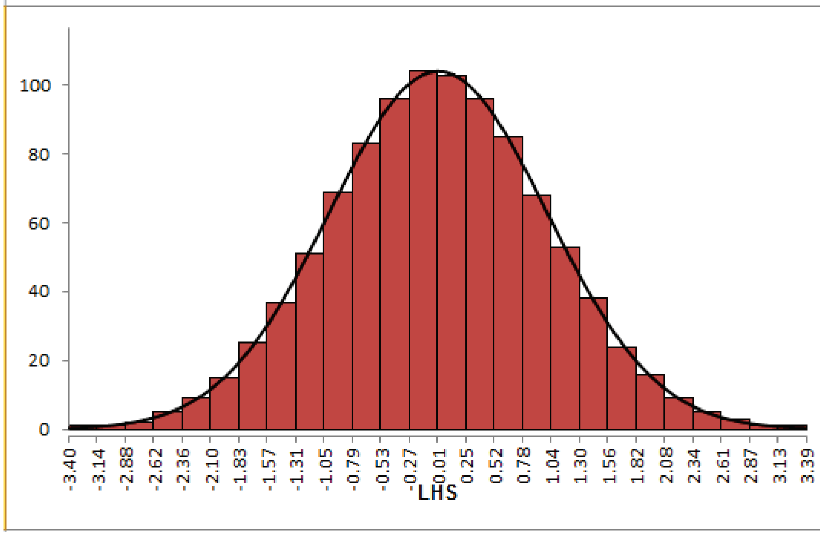
MCS

Count = 1000
Mean = 0.005714
Stdev = 0.983757
Range = 6.23

Minimum = -3.333
25th Percentile (Q1) = -0.631130
50th Percentile (Median) = 0.007624
75th Percentile (Q3) = 0.658016
Maximum = 2.902422

95% CI Mean = -0.06 to 0.07
95% CI Sigma = 0.94 to 1.03

Anderson-Darling Normality Test:
A-Squared = 0.436560; p-value = 0.2967



LHS

Count = 1000
Mean = 4.85E-05
Stdev = 1.000866
Range = 6.78

Minimum = -3.401
25th Percentile (Q1) = -0.675267
50th Percentile (Median) = 0.000241
75th Percentile (Q3) = 0.674767
Maximum = 3.38169

95% CI Mean = -0.06 to 0.06
95% CI Sigma = 0.96 to 1.05

Anderson-Darling Normality Test:
A-Squared = 0.003108; p-value = 1.0000

Note the smoother shape for LHS as well as the very low Anderson Darling A Squared value.

Input Distribution Truncation

A truncated distribution is a conditional distribution derived by restricting the domain of the original distribution. Consider a distribution, $f(x)$, and we wish to know what is the distribution of x if x is to be restricted to the domain $a < x \leq b$. The truncated distribution is then defined as:

$$f(x|a < x \leq b) = \frac{g(x)}{F(b) - F(a)}$$

where $F(x)$ is the CDF of $f(x)$, and $g(x) = f(x)$ for $a < x \leq b$, and zero everywhere else.

The following histogram illustrates a truncated normal distribution (Mean = 0, StdDev = 1) with Minimum = -2 and Maximum = 2.

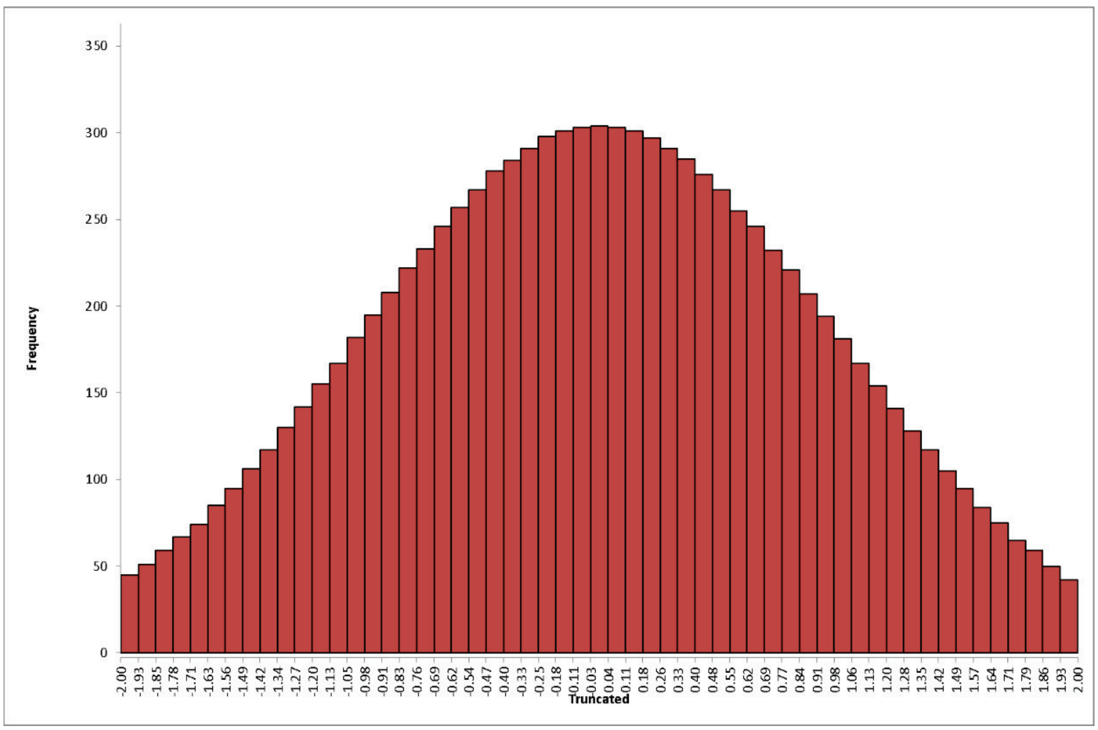


Table of DSIM Functions

DSIM Function (all arguments are required)	Description	Formula
DSim_Cpm (variable, LSL, USL, tgt)	Returns the Cpm process capability index of a DiscoverSim variable. This is also known as the Taguchi Capability Index and includes a penalty for deviation from target. Requires lower spec. limit (LSL), upper spec. limit (USL) and target.	$$C_{pm} = \frac{min(USL - T, T - LSL)}{3\sqrt{s^2 + (\bar{x} - T)^2}}$$ where: LSL is the lower specification limit. USL is the upper specification limit. $\bar{x}$ is the sample mean. T is the target value. s is the overall (long-term) sample standard deviation.
DSim_DPM_Actual (variable, LSL, USL)	Returns the Actual DPM (Defects per Million) of a DiscoverSim variable (also known as Actual PPM - Parts per Million). It is based on an empirical count of defects. A defect is an observation with a value that is less than the lower spec. limit (LSL) or greater than the upper spec. limit (USL).	$$DPM = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i < LSL) + \sum_{i=1}^{n}(x_i > USL)\right]$$ where: LSL is the lower specification limit. USL is the upper specification limit. x_i is the i^{th} observation in the sample. n is the number of observations in the sample.
DSim_DPM_Calc (variable, LSL, USL)	Returns the Calculated DPM (Defects per Million) of a DiscoverSim variable (also known as Expected PPM - Parts per Million). It is based on the sample mean, standard deviation, lower specification limit (LSL), upper	$$DPM = 10^6\left[(1 - \Phi(\frac{\bar{x} - LSL}{s})) + (1 - \Phi(\frac{USL - \bar{x}}{s}))\right]$$ where: LSL is the lower specification limit. USL is the upper specification limit. $\bar{x}$ is the sample mean. s is the sample standard deviation. Φ is the cdf of the standard normal distribution.

DSIM Function (all arguments are required)	Description	Formula
	specification limit (USL) and assumes a normal distribution.	
DSim_DPML_Actual (variable, LSL)	Returns the Actual DPML (Defects per Million - Lower) of a DiscoverSim variable. It is based on an empirical count of defects. A defect is an observation with a value that is less than the lower spec. limit (LSL).	$$DPML = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i < LSL)\right]$$ where: LSL is the lower specification limit. x_i is the i^{th} observation in the sample. n is the number of observations in the sample.
DSim_DPML_Calc (variable, LSL)	Returns the Calculated DPML (Defects per Million - Lower) of a DiscoverSim variable. It is based on the sample mean, standard deviation, lower specification limit (LSL) and assumes a normal distribution.	$$DPML = 10^6\left[(1 - \Phi(\frac{\bar{x} - LSL}{s}))\right]$$ where: LSL is the lower specification limit. $\bar{x}$ is the sample mean. s is the sample standard deviation. Φ is the cdf of the standard normal distribution.
DSim_DPMU_Actual (variable, USL)	Returns the Actual DPMU (Defects per Million - Upper) of a DiscoverSim variable. It is based on an empirical count of defects. A defect is an observation with a value that is greater than the upper spec. limit (USL).	$$DPMU = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i > USL)\right]$$ where: USL is the upper specification limit. x_i is the i^{th} observation in the sample. n is the number of observations in the sample.

DSIM Function (all arguments are required)	Description	Formula
DSim_DPMU_Calc (variable, USL)	Returns the Calculated DPMU (Defects per Million - Upper) of a DiscoverSim variable. It is based on the sample mean, standard deviation, upper specification limit (USL) and assumes a normal distribution.	$DPMU = 10^6 \left[(1 - \Phi(\frac{USL - \bar{x}}{s}))\right]$ where: USL is the upper specification limit. $\bar{x}$ is the sample mean. s is the sample standard deviation. Φ is the cdf of the standard normal distribution.
DSim_IQR (variable)	Returns the interquartile range (Q3 - Q1) of a DiscoverSim variable.	$IQR = Q3 - Q1$ where: Q1 is the first quartile. Q3 is the third quartile.
DSim_IsAllDifferent (rng)	Returns unity if all elements of a range are different.	
DSim_IsAllSame	Returns unity if all elements of a range are identical.	
DSim_Kurt(variable)	Returns the kurtosis (g2) of a DiscoverSim variable. Kurtosis is a measure of the peakedness of the data set. A normal distribution has a kurtosis of zero.	$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} - 3$ where: x_i is the i^{th} element in the sample. $\bar{x}$ is the sample mean.
DSim_Max(variable)	Returns the maximum value of a DiscoverSim variable.	$x_{max} = x_n$ where: x_n is the n^{th} element in the ordered sample. n is the number of observations in the sample.

DSIM Function (all arguments are required)	Description	Formula
DSim_Mean(variable)	Returns the mean value of a DiscoverSim variable.	$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ where: x_i is the i^{th} element in the sample. n is the number of observations in the sample.
DSim_Median (variable)	Returns the median of a DiscoverSim variable. The median is the second quartile, or 50th percentile - the point that divides the ordered data into two equal groups.	$Pr[x_i < \tilde{x}] \leq 0.5$ where: x_i is the i^{th} element in the ordered sample. If the number of observations in a data set is odd, the median is the value in the middle. If the number of observations in a data set is even, the median is the average of the values of the two middle terms.
DSim_Min(variable)	Returns the minimum value of a DiscoverSim variable.	$x_{min} = x_1$ where: x_1 is the 1$^{\text{st}}$ element in the ordered sample.
DSim_MSE (variable, tgt)	Returns the mean squared error of a DiscoverSim variable. It is the expected value of the quadratic loss (also known as the Taguchi Loss Function) and incorporates both the variance of the estimator and its bias. Requires a target.	$mse = (\bar{x} - T)^2 + s^2$ where: $\bar{x}$ is the sample mean. T is the specified target. s^2 is the sample variance.
DSim_Pcntile (variable,k_prob)	Returns the specified percentile of a DiscoverSim variable. k_prob is the percentile probability (0 < k_prob < 1), so a value of 0.95 is the 95$^{\text{th}}$ percentile.	

DSIM Function (all arguments are required)	Description	Formula
DSim_Pp (variable,LSL,USL)	Returns the Pp process capability index of a DiscoverSim variable. It assumes a normal distribution.	$P_p = \frac{USL - LSL}{6s}$ where: USL is the upper specification limit. LSL is the lower specification limit. s is the overall (long-term) sample standard deviation. Note that C_p, which uses the within-subgroup (short term) standard deviation, as applied in Six Sigma and Statistical Process Control, is not included in DiscoverSim. A typical Monte Carlo simulation run will use a stable mean and standard deviation, so there would be no statistically significant difference between C_p and P_p or short-term versus long-term standard deviation. This note also applies to Ppu, Ppl, Ppk, and Cpm.
DSim_Pp_Pcntile (variable,LSL,USL)	Returns the Percentile Pp process capability index of a DiscoverSim variable. This index uses the 99.865 and 0.135 percentiles, so is useful when the data are not normally distributed.	$\%P_p = \frac{USL - LSL}{x_{0.99865} - x_{.00135}}$ where: USL is the upper specification limit. LSL is the lower specification limit. x_q = is the q^{th} percentile of x.
DSim_Ppk (variable,LSL,USL)	Returns the Ppk process capability index of a DiscoverSim variable. It assumes a normal distribution.	$P_{pk} = \frac{min(USL - \bar{x}, \bar{x} - LSL)}{3s}$ where: USL is the upper specification limit. LSL is the lower specification limit. $\bar{x}$ is the sample mean. s is the overall (long-term) sample standard deviation.
DSim_Ppk_Pcntile (variable,LSL,USL)	Returns the Percentile Ppk process capability index of a DiscoverSim variable. This index	$\%P_{pk} = min(\%P_{pu}, \%P_{pl})$ $\%P_{pu} = \frac{USL - \tilde{x}}{x_{.99865} - \tilde{x}}$

DSIM Function (all arguments are required)	Description	Formula
	uses percentiles, so is useful when the data are not normally distributed.	$$\%P_{pl} = \frac{\tilde{x} - LSL}{\tilde{x} - x_{.00135}}$$ where: USL is the upper specification limit. LSL is the lower specification limit. $\tilde{x}$ is the sample median. x_q = is the q^{th} percentile of x.
DSim_Ppl (variable,LSL)	Returns the Ppl (lower) process capability index of a DiscoverSim variable. It assumes a normal distribution.	$$P_{pl} = \frac{\bar{x} - LSL}{3s}$$ where: LSL is the lower specification limit. $\bar{x}$ is the sample mean. s is the overall (long-term) sample standard deviation.
DSim_Ppl_Pcntile (variable,LSL)	Returns the Percentile Ppl (lower) process capability index of a DiscoverSim variable. This index uses the 0.135 percentile and median, so is useful when the data are not normally distributed.	$$\%P_{pl} = \frac{\tilde{x} - LSL}{\tilde{x} - x_{.00135}}$$ where: LSL is the lower specification limit. $\tilde{x}$ is the sample median. x_q = is the q^{th} percentile of x.
DSim_Ppu (variable,USL)	Returns the Ppu (upper) process capability index of a DiscoverSim variable. It assumes a normal distribution.	$$P_{pu} = \frac{USL - \bar{x}}{3s}$$ where: USL is the upper specification limit. $\bar{x}$ is the sample mean. s is the overall (long-term) sample standard deviation.
DSim_Ppu_Pcntile (variable,USL)	Returns the Percentile Ppu (upper) process capability index of a DiscoverSim variable. This index uses the 99.865 percentile and median, so is useful	$$\%P_{pu} = \frac{USL - \tilde{x}}{x_{.99865} - \tilde{x}}$$ where: $USL$is the upper specification limit. $\tilde{x}$ is the sample median. x_q = is the q^{th} percentile of x.

DSIM Function (all arguments are required)	Description	Formula
	when the data are not normally distributed.	
DSim_Q1(variable)	Returns the first quartile (25th percentile) of a DiscoverSim variable. This is the point that marks the first boundary when the ordered data is divided into four equal groups.	$Pr[x_i < Q1] \leq 0.25$ where: x_i is the i^{th} element in the ordered sample. If $k = (n + 1)/4$ is an integer, the first quartile is taken as the x_k element of the ordered observations; otherwise the first quartile is taken as the sum of the x_k element and the fractional component applied to $x_{k+1} - x_k$.
DSim_Q3(variable)	Returns the third quartile (75th percentile) of a DiscoverSim variable. This is the point that marks the third boundary when the ordered data is divided into four equal groups.	$Pr[x_i < Q3] \leq 0.75$ where: x_i is the i^{th} element in the ordered sample. If $k = 3(n + 1)/4$is an integer, the third quartile is taken as the x_k element of the ordered observations; otherwise the third quartile is taken as the sum of the x_k element and the fractional component applied to $x_{k+1} - x_k$.
DSim_Range (variable)	Returns the range (maximum – minimum) of a DiscoverSim variable.	$R_x = x_{max} - x_{min}$ where: x_{min} is the 1st element in the ordered sample. x_{max} is the n^{th} element in the ordered sample. *n* is the number of observations in the sample.
DSim_Skew(variable)	Returns the skewness (g1) of a DiscoverSim variable. This is a measure of the asymmetry of the data set. A negative value implies a long tail to the left of the mean, while a positive value implies a long tail to the right of the mean. A	$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{3/2}}$ where: x_i is the i^{th} element in the sample. $\bar{x}$ is the sample mean.

DSIM Function (all arguments are required)	Description	Formula
	normal distribution has a skew of zero.	
DSim_Span(variable)	Returns the span (0.95 – 0.05) of a DiscoverSim variable. This is less sensitive to outliers than the range.	$S_x = x_{.95} - x_{.05}$ where: x_q = is the q^{th} percentile of x.
DSim_Stdev(variable)	Returns the sample standard deviation of a DiscoverSim variable.	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}}$ where: x_i is the i^{th} element in the sample. $\bar{x}$ is the sample mean. $n$is the number of observations in the sample.
DSim_Var(variable)	Returns the sample variance of a DiscoverSim variable.	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$ where: x_i is the i^{th} element in the sample. $\bar{x}$ is the sample mean. n is the number of observations in the sample.

Specifying the Optimization Objective Function

The following tables give the options available to specify an optimization objective function in the **Run Optimization** dialog:

Optimization Goal:	Minimize	
Multiple Output Metric:	Weighted Sum	Deviation from Target
Statistic:	Mean Median 1st quartile 3rd quartile Percentile (%) Minimum Maximum Standard Deviation Variance Mean Squared Error (requires Target) Skewness Kurtosis Range IQR (75-25) Span (95-5) Actual DPM (defects per million – requires LSL/USL) Actual DPMU (Upper – requires USL) Actual DPML (Lower – requires LSL) Calculated DPM (defects per million assuming normal distribution – requires LSL/USL) Calculated DPMU (Upper – requires USL) Calculated DPML (Lower – requires LSL)	Mean (requires Target)

Optimization Goal:	Maximize	
Multiple Output Metric:	Weighted Sum	Desirability
Statistic:	Mean Median 1st quartile 3rd quartile Percentile (%) Minimum Maximum Standard Deviation Variance Mean Squared Error (requires Target) Skewness Kurtosis Range IQR (75-25) Span (95-5) Pp (requires LSL/USL) PpU (Upper – requires USL) PpL (Lower – requires LSL) Ppk (requires LSL/USL) Cpm (requires Target/LSL/USL) %Pp (Percentile Pp – requires LSL/USL) %PpU (Percentile PpU – requires USL) %PpL (Percentile PpL – requires LSL) %Ppk (Percentile Ppk – requires LSL/USL)	Mean (requires Target and LSL, USL or LSL/USL)

Formulas for Multiple Output Metrics

1. **Minimize/Maximize Weighted Sum:** minimize or maximize the weighted sum of selected statistic. Weights are specified for each output in the **Output Response** dialog. For example, if the selected statistic was the Mean, and there were two outputs, the objective function would be:

   $\text{Weight}_1\text{Mean}_1 + \text{Weight}_2\text{Mean}_2$

   If there is a single output in the model, this simplifies to minimize or maximize the statistic value.
   Note that the **Output Goal** specified in the **Output Response** dialog is not used here. It is only used in **Maximize Desirability**.

2. **Minimize Deviation from Target:** minimize the square root of weighted sum of deviations squared. A target must be specified for each output in the **Output Response** dialog. The only statistic available for this option is the mean. This is also known as the Taguchi or Quadratic Loss Function. If there were two outputs, the objective function would be:

   $\text{SQRT}(\ \text{Weight}_1(\text{Mean}_1 - \text{Target}_1)^2 + \text{Weight}_2(\text{Mean}_2 - \text{Target}_2)^2)$.

3. **Maximize Desirability:** Maximize the weighted linear sum of the Derringer and Suich desirability measure (Derringer and Suich, 1980). Each output must specify:

   - **Weight** (default = 1). This is also referred to as "Importance." (Note, another factor, the desirability shape is sometimes called "weight." In DiscoverSim, the desirability shape factor is fixed at 1.)
   - **Output Goal** (Target, Maximize or Minimize) – this is specific to an output. For example, if Output 1 is production rate, the goal would be set to maximize, and Output 2, cost, would have a minimize goal. However, the specified overall Optimization Goal is to maximize desirability.
     - If the output goal is Target, then **LSL**, **Target**, and **USL** are required. LSL and USL are the lower and upper specification limits used for process capability and dpm calculations, but are also used as the lower and upper bounds for desirability. The objective function is:

$$F(x) = \begin{cases} 0 & if\ x < x_{lb} \\ \left(\dfrac{x - x_{lb}}{x_{tgt} - x_{lb}}\right) & if\ x_{lb} \le x \le x_{tgt} \\ \left(\dfrac{x - x_{ub}}{x_{tgt} - x_{ub}}\right) & if\ x_{tgt} \le x \le x_{ub} \\ 0 & if\ x > x_{ub} \end{cases}$$

- If the output goal is Minimize, then **Target** and **USL** are required. The objective function is:

$$F(x) = \begin{cases} 1 & if\ x\ <\ x_{\,tgt} \\ \left(\dfrac{x - x_{ub}}{x_{tgt} - x_{ub}}\right) & if\ x_{\,tgt}\ \leq\ x\ \leq\ x_{\,ub} \\ 0 & if\ x\ >\ x_{\,ub} \end{cases}$$

- If the output goal is Maximize, then **LSL** and **Target** are required. The objective function is:

$$F(x) = \begin{cases} 0 & if\ x\ <\ x_{\,lb} \\ \left(\dfrac{x - x_{lb}}{x_{tgt} - x_{lb}}\right) & if\ x_{\,lb}\ \leq\ x\ \leq\ x_{\,tgt} \\ 1 & if\ x\ >\ x_{\,tgt} \end{cases}$$

The only statistic available for the desirability option is the mean.

4. Stochastic optimization requires a fixed seed in order to avoid “chatter” that would result in inconsistent comparisons. If the **Seed** is set to **Clock**, the initial seed value will be obtained from the system clock and then kept fixed throughout the optimization.

Reference for Multiple Output Metrics

Derringer, G., and R. Suich (1980), “Simultaneous Optimization of Several Response Variables,” *Journal of Quality Technology*, 12(4), pp. 214-219.

Formulas for Statistical Measures

Mean $\bar{x}$: The average value of the elements in the vector x. It is the sample estimate of the population mean μ.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where:

x_i is the i^{th} element in the sample.
n is the number of observations in the sample.

Median $\tilde{x}$:The median is the second quartile, or 50^{th} percentile - the point that divides the ordered data into two equal groups.

$$Pr[x_i < \tilde{x}] \leq 0.5$$

where:

x_i is the i^{th} element in the ordered sample.
If the number of observations in a data set is odd, the median is the value in the middle. If the number of observations in a data set is even, the median is the average of the values of the two middle terms.

1st Quartile $Q1$: The first quartile, or 25^{th} percentile - the point that marks the first boundary when the ordered data is divided into four equal groups.

$$Pr[x_i < Q1] \leq 0.25$$

where:

x_i is the i^{th} element in the ordered sample.
If $k = (n + 1)/4$ is an integer, the first quartile is taken as the x_k element of the ordered observations; otherwise the first quartile is taken as the sum of the x_k element and the fractional component applied to $x_{k+1} - x_k$.

3rd Quartile $Q3$: The third quartile, or 75^{th} percentile - the point that marks the third boundary when the ordered data is divided into four equal groups.

$$Pr[x_i < Q3] \leq 0.75$$

where:

x_i is the i^{th} element in the ordered sample.
If $k = 3(n + 1)/4$is an integer, the third quartile is taken as the x_k element of the ordered observations; otherwise the third quartile is taken as the sum of the x_k element and the fractional component applied to $x_{k+1} - x_k$.

Percentile: The value below which a given percentage of observations in a group of observations fall. For example, the default 95th percentile is the value below which 95 percent of the observations may be found.

Minimum x_{min}: The smallest elements in the vector x.

$$x_{min} = x_1$$

where:

x_1 is the 1st element in the ordered sample.

Maximum x_{max}: The largest elements in the vector x.

$$x_{max} = x_n$$

where:

x_n is the nth element in the ordered sample.
n is the number of observations in the sample.

Standard. Deviation s: The sample standard deviation is a measure of the dispersion from the mean in the vector x. It is the sample estimate of the population standard deviation σ.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

where:

x_i is the ith element in the sample.
$\bar{x}$ is the sample mean.
$n$is the number of observations in the sample.

Variance s^2: The sample variance is the expected value of the square of the sample deviations from the mean in the vector x It is the sample estimate of the population variance σ^2.

$$s^2 = \frac{\sum_{i=1}^{n} \quad (x_i - \overline{x})^2}{n-1}$$

where:

x_i is the ith element in the sample.
$\bar{x}$ is the sample mean.
n is the number of observations in the sample.

Mean Squared Error *mse*: The mean squared error is the expected value of the quadratic loss and incorporates both the variance of the estimator and its bias. This is also known as the Taguchi Loss Function.

$$mse = (\bar{x} - T)^2 + s^2$$

where:

$\bar{x}$ is the sample mean.
T is the specified target.
s^2 is the sample variance.

Skewness g_1: Skewness is a measure of the asymmetry of the data set. A negative value implies a long tail to the left of the mean, while a positive value implies a long tail to the right of the mean. A normal distribution has a skew of zero.

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{3/2}}$$

where:

x_i is the i^{th} element in the sample.
$\bar{x}$ is the sample mean.

Kurtosis g_2: Kurtosis is a measure of the peakedness of the data set. A normal distribution has a kurtosis of zero.

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{2}} - 3$$

where:

x_i is the i^{th} element in the sample.
$\bar{x}$ is the sample mean.

Range R_x: The difference between the largest and smallest elements in the vector *x*.

$$R_x = x_{max} - x_{min}$$

where:

x_{min} is the 1st element in the ordered sample.
x_{max} is the n^{th} element in the ordered sample.
n is the number of observations in the sample.

IQR The interquartile range, a measure of dispersion, is the difference between the third and first quartiles.

$$IQR = Q3 - Q1$$

where:

Q1 is the first quartile.
Q3 is the third quartile.

Span S_x: The range of the elements in the vector *x*, after discarding the lower and upper 5% of the sample. This is less sensitive to outliers than the range.

$$S_x = x_{.95} - x_{.05}$$

where:

x_q = is the q^{th} percentile of x.

Formulas for Quality and Process Capability Indices

Actual DPM: Actual DPM (Defects per Million), also known as Actual PPM (Parts per Million), is based on an empirical count of defects. A defect is an observation with a value that is less than the lower spec. limit (LSL) or greater than the upper spec. limit (USL). Requires LSL and USL.

$$DPM = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i < LSL) + \sum_{i=1}^{n}(x_i > USL)\right]$$

where:

LSL is the lower specification limit.
USL is the upper specification limit.
x_i is the i^{th} observation in the sample.
n is the number of observations in the sample.

Actual DPMU: Actual DPMU (Defects per Million - Upper), also known as Actual PPMU (Parts per Million - Upper), is based on an empirical count of defects. A defect is an observation with a value that is greater than the upper spec. limit (USL). Requires USL.

$$DPMU = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i > USL)\right]$$

where:

USL is the upper specification limit.
x_i is the i^{th} observation in the sample.
n is the number of observations in the sample.

Actual DPML: Actual DPML (Defects per Million - Lower), also known as Actual PPML (Parts per Million - Lower), is based on an empirical count of defects. A defect is an observation with a value that is less than the lower spec. limit (LSL). Requires LSL.

$$DPML = \frac{10^6}{n}\left[\sum_{i=1}^{n}(x_i < LSL)\right]$$

where:

LSL is the lower specification limit.
x_i is the i^{th} observation in the sample.
n is the number of observations in the sample.

Calculated DPM: Calculated DPM (Defects per Million), also known as Expected PPM (Parts per Million), is based on the sample mean, standard deviation, lower specification limit (LSL), upper specification limit (USL) and assumes a normal distribution. Requires LSL and USL.

$$DPM = 10^6 \left[(1 - \Phi(\frac{\bar{x} - LSL}{s})) + (1 - \Phi(\frac{USL - \bar{x}}{s})) \right]$$

where:

LSL is the lower specification limit.
USL is the upper specification limit.
$\bar{x}$ is the sample mean.
s is the sample standard deviation.
Φ is the cdf of the standard normal distribution.

Calculated DPMU: Calculated DPMU (Defects per Million - Upper), also known as Expected PPMU (Parts per Million - Upper), is based on the sample mean, standard deviation, upper specification limit (USL) and assumes a normal distribution. Requires USL.

$$DPMU = \frac{10^6}{n} \left[\sum_{i=1}^{n} (x_i > USL) \right]$$

where:

USL is the upper specification limit.
x_i is the i^{th} observation in the sample.
n is the number of observations in the sample.

Calculated DPML: Calculated DPML (Defects per Million - Lower), also known as Expected PPML (Parts per Million - Lower), is based on the sample mean, standard deviation, lower specification limit (LSL) and assumes a normal distribution. Requires LSL.

$$DPML = 10^6 \left[(1 - \Phi(\frac{\bar{x} - LSL}{s})) \right]$$

where:

LSL is the lower specification limit.
$\bar{x}$ is the sample mean.
s is the sample standard deviation.
Φ is the cdf of the standard normal distribution.

Pp: Process Capability (or Performance) Index. It assumes a normal distribution. Requires LSL and USL.

$$P_p = \frac{USL - LSL}{6s}$$

where:

USL is the upper specification limit.
LSL is the lower specification limit.
s is the overall (long-term) sample standard deviation.
Note that C_p, which uses the within-subgroup (short term) standard deviation, as applied in Six Sigma and Statistical Process Control, is not included in DiscoverSim. A typical Monte Carlo simulation run will use a stable mean and standard deviation, so there would be no

statistically significant difference between C_p and P_p or short-term versus long-term standard deviation. This note also applies to PpU, PpL, Ppk, and Cpm.

PpU: Process Capability (or Performance) Index. It assumes a normal distribution. Requires USL.

$$P_{pu} = \frac{USL - \bar{x}}{3s}$$

where:

USL is the upper specification limit.
$\bar{x}$ is the sample mean.
s is the overall (long-term) sample standard deviation.

PpL: Process Capability (or Performance) Index. It assumes a normal distribution. Requires LSL.

$$P_{pl} = \frac{\bar{x} - LSL}{3s}$$

where:

LSL is the lower specification limit.
$\bar{x}$ is the sample mean.
s is the overall (long-term) sample standard deviation.

Ppk: Process Capability (or Performance) Index. It assumes a normal distribution. Requires LSL and USL.

$$P_{pk} = \frac{min(USL - \bar{x}, \bar{x} - LSL)}{3s}$$

where:

LSL is the lower specification limit.
USL is the upper specification limit.
$\bar{x}$ is the sample mean.
s is the overall (long-term) sample standard deviation.

Cpm: Process Capability Index, is also known as the Taguchi Capability Index and it includes a penalty for deviation from target. It assumes a normal distribution. Requires LSL and USL.

$$C_{pm} = \frac{min(USL - T, T - LSL)}{3\sqrt{s^2 + (\bar{x} - T)^2}}$$

where:

LSL is the lower specification limit.
USL is the upper specification limit.
$\bar{x}$ is the sample mean.
T is the target value.
s is the overall (long-term) sample standard deviation.

%Pp: Percentile Process Capability Index. This index uses the 99.865 and .135 percentiles, so is useful when the data are not normally distributed. Requires LSL and USL.

$$\%P_p = \frac{USL - LSL}{x_{0.99865} - x_{.00135}}$$

where:

> USL is the upper specification limit.
> LSL is the lower specification limit.
> x_q = is the q^{th} percentile of x.

%PpU: Percentile Process Capability Index. This index uses the 99.865 percentile and median, so is useful when the data are not normally distributed. Requires USL.

$$\%P_{pu} = \frac{USL - \tilde{x}}{x_{.99865} - \tilde{x}}$$

where:

> USL is the upper specification limit.
> $\tilde{x}$ is the sample median.
> x_q = is the q^{th} percentile of x.

%PpL: Percentile Process Capability Index. This index uses the 0.135 percentile and median, so is useful when the data are not normally distributed. Requires LSL.

$$\%P_{pl} = \frac{\tilde{x} - LSL}{\tilde{x} - x_{.00135}}$$

where:

> LSL is the lower specification limit.
> $\tilde{x}$ is the sample median.
> x_q = is the q^{th} percentile of x.

%Ppk: Percentile Process Capability Index. This index uses percentiles, so is useful when the data are not normally distributed. Requires LSL and USL.

$$\%P_{pk} = min(\%P_{pu}, \%P_{pl})$$

MIDACO Mixed Integer Distributed Ant Colony Global Optimization

From http://www.midaco-solver.com/index.php/about. Used with permission from Martin Schlueter, MIDACO-SOLVER.com.

MIDACO is a general-purpose software for solving mathematical optimization problems. Initially developed for Mixed Integer Nonlinear Programming (MINLP) problems arising from demanding space applications at the European Space Agency [3], [7] and Astrium (EADS) [7], the software can be applied to a wide range of optimization problems. MIDACO handles problems where the objective function ***f(x)*** depends on continuous, integer or both types (mixed integer case) of variables ***x***. The problem might be further restricted to equality and/or inequality constraints ***g(x)***.

Optimization Problem

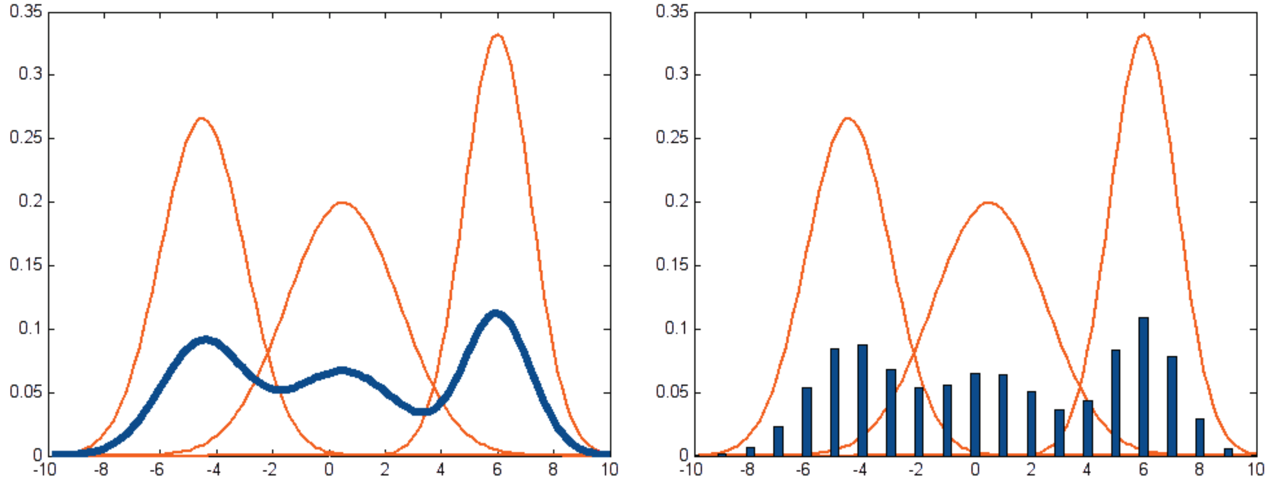
$$
\begin{array}{llll}
\text{Minimize} & f(x) & & \text{(Objective Function)} \\
\\
\text{Subject to:} & g(x) & = 0 & \text{(Equality Constraints)} \\
 & g(x) & \geq 0 & \text{(Inequality Constraints)}
\end{array}
$$

As a **black-box** optimizer, MIDACO does not require the objective and constraint functions to hold any particular property like convexity or differentiability. MIDACO can robustly solve problems with critical function properties like high non-convexity, non-differentiability, flat spots and even stochastic noise. Extensively tested on over a hundred benchmark problems [5] and several real world applications, the software represents the state of the art for evolutionary computing on MINLP. Comparisons with established deterministic algorithms (see [5], [6]) revealed that MIDACO is fully competitive in respect to the number of global optimal solutions and cpu-runtime behaviour.

MIDACO holds the benchmark world record best solution to Full Messenger (Mission to Mercury), which is considered the most difficult space trajectory problem in the ESA Global Trajectory Optimization Database. See http://www.midaco-solver.com/index.php/about/benchmarks/esa-gtop for record solution(s).

Note that the use of MIDACO with DiscoverSim, as a user friendly Excel add-in, while extremely powerful for the environment, does add significant overhead resulting in performance degradation compared to the published benchmarks. Users requiring the full speed capabilities of MIDACO including parallelization with high performance computing would need to purchase stand-alone MIDACO SOLVER (see http://www.midaco-solver.com/index.php/contact).

The MIDACO algorithm is based on a mixed integer extension of the Ant Colony Optimization (ACO) metaheuristic in combination with the recently developed Oracle Penalty Method [4] for constraint handling. The key element of MIDACO's ACO algorithm are so called “multi-kernel gaussian probability density functions" (Gauss PDF's), which are used to stochastically sample solution candidates (also called "ants" or "iterates") for the optimization problem. In case of integer variables, a discretized version of the Gauss PDF is applied instead of the continuous version.



Continuous (left) and discretized (right) multi-kernel gaussian probability density functions (Gauss PDF's**)**

MIDACO References:

[1] Schlueter M., Egea J. A., Banga J. R.: Extended Ant Colony Optimization for non-convex Mixed Integer Nonlinear Programming, Computers & Operations Research (Elsevier), Vol. 36, Issue 7, Pages 2217-2229 (2009). Preprint: http://www.midaco-solver.com/data/pub/Extended_ACO_for_MINLP.pdf.

[2] Schlueter M., Egea J. A., Antelo L.T., Alonso A.A., Banga J. R.: An extended Ant Colony Optimization Algorithm for Integrated Process and Control System Design, Industrial & Engineering Chemistry Res. (ACS), Vol. 48, Issue 14, Pages 6723-6738 (2009). Preprint: http://www.midaco-solver.com/data/pub/Extended_ACO_for_Control_System_Design.pdf.

[3] Schlueter M., Rückmann J.J., Gerdts M.: Non-linear Mixed-Integer-based Optimization Technique for Space Applications, ESA Conference - International NPI Day, European Space Agency, Noordwijk, Netherlands (2010). Poster: http://www.midaco-solver.com/data/pub/ESA_NPI_Poster_2010.jpg.

[4] Schlueter M., Gerdts, M.: The Oracle Penalty Method, Journal of Global Optimization (Springer), Vol. 47, Issue 2, Pages 293-325 (2010). Preprint: http://www.midaco-solver.com/data/pub/The_Oracle_Penalty_Method.pdf.

[5] Schlueter M., Gerdts M., Rückmann J.J.: A Numerical Study of MIDACO on 100 MINLP Benchmarks, Optimization (Taylor & Francis), Vol 61, Issue 7, Pages 873-900 (2012). Preprint: http://www.midaco-solver.com/data/pub/MIDACO_on_100_MINLP_Benchmarks.pdf.

[6] Schlueter M.: Nonlinear Mixed Integer based Optimization Technique for Space Applications, Ph.D. Thesis, School of Mathematics, University of Birmingham (UK) (2012). Thesis: http://www.midaco-solver.com/data/pub/Thesis_Schlueter.pdf.

[7] Schlueter M., Erb S., Gerdts M., Kemble S., Rückmann J.J.: MIDACO on MINLP Space Applications, Advances in Space Research (Elsevier), Vol 51, Issue 7, Pages 1116–1131 (2013). Preprint: http://www.midaco-solver.com/data/pub/MIDACO_on_Space_Applications.pdf.

[8] Schlueter M., Munetomo M.: MIDACO User Guide, Technical Report, Hokkaido University, Japan (HUSCAP) (2013). PDF: http://www.midaco-solver.com/data/other/MIDACO_User_Guide.pdf.

[9] Schlueter M., Munetomo M.: Parallelization Strategies for Evolutionary Algorithms for MINLP, Proceedings of the IEEE Congress on Evolutionary Computation (IEEE-CEC), pp.635-641 (2013). Preprint: http://www.midaco-solver.com/data/pub/MIDACO_CEC2013.pdf.

[10] Schlueter M., Munetomo M.: Midaco-Solver, Global Optimization Software for Mixed Integer Nonlinear Programming, SC13 - Super Computing Conference, Exhibitor Booth Nr. 320, Denver, USA (2013). Poster: http://www.midaco-solver.com/data/other/SC13_MIDACO.pdf.

[11] Schlueter M.: Global Optimization of MINLP by Evolutionary Algorithms, APMonitor Webinar (AIChE), Symposium on Modeling and Optimization, USA (2014). PDF: http://www.midaco-solver.com/data/pub/MIDACO_APMonitor.pdf. Video: http://www.youtube.com/watch?v=oCVib-sRbjY&list=UU2GuY-AxnNxIJFAVfEW0QFA&feature=c4-overview.

[12] Schlueter M.: Recent Advances in Optimization Software for Space Applications, Invited Presentation, Japanese Space Exploration Agency (ISAS/JAXA), Japan (2014). PDF: http://www.midaco-solver.com/data/pub/MIDACO_JAXA_2014.pdf.

[13] Schlueter M. Munetomo M.: Parallelization for Space Trajectory Optimization, Proceedings of the IEEE World Congress on Computational Intelligence (IEEE-WCCI), pp. 832 - 839 (2014). Preprint: http://www.midaco-solver.com/data/pub/CEC2014_Schlueter_Munetomo.pdf.

[14] Schlueter M., MIDACO Software Performance on Interplanetary Trajectory Benchmarks, Advances in Space Research (Elsevier), Vol 54, Issue 4, Pages 744 - 754 (2014). Preprint: http://www.midaco-solver.com/data/pub/MIDACO_on_ESA_GTOP_Benchmarks.pdf.

ORACLE PENALTY METHOD

From http://www.midaco-solver.com/index.php/more/oracle-penalty-method. Used with permission from Martin Schlueter, MIDACO-SOLVER.com.

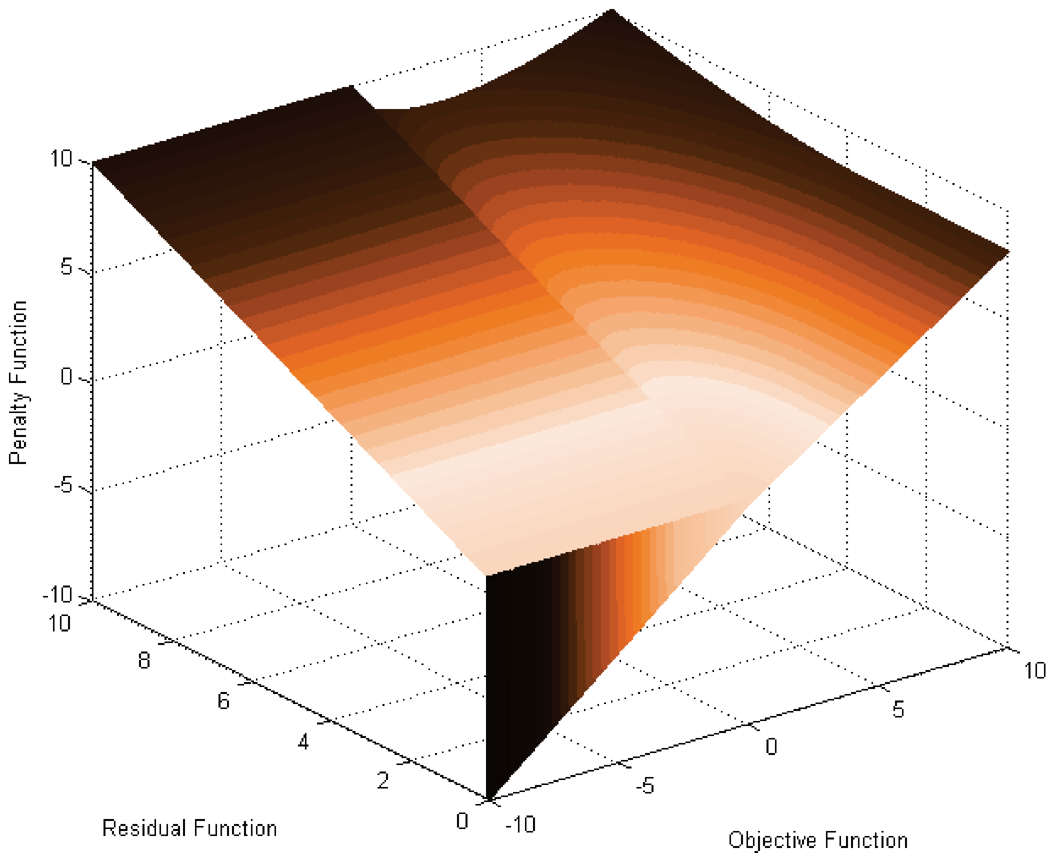
The oracle penalty method is an advanced approach for constraint handling within evolutionary optimization algorithms. The method is based on a single parameter **Omega** (**Ω**), which is called an oracle due to its predictive nature. This parameter directly corresponds to the global optimal objective function value f(x). In case of real-world applications where the user has some (rough) guess about a possible global optimal f(x) value, this information can be exploited as oracle information to improve the convergence speed of the optimization algorithm. In case no such user guided information is available, an (automated) update rule can be applied for Ω, starting with a value of infinity (∞) and updating Ω after individual optimization runs, when better (feasible) solutions have been found.

Algorithms implementing the oracle penalty method with automated update are classified as **self-tuning**. The oracle penalty method has been applied for, example, in Ant Colony Optimization (ACO), see [1], [6], [7]; Genetic Algorithms (GA), see [2], [10]; Particle Swarm Optimization (PSO), see [3] and Differential Evolution (DE), see [4], [5], [8]. The most detailed introduction to the method can be found in [1]; a brief motivation of the key idea of the method can be found in this PDF presentation (slide 11-15). The oracle penalty method is one of the major reasons why MIDACO is capable to solve problems with up to hundreds of (non-linear) constraints (see Benchmarks).

Mathematical Formulation of the Oracle Penalty Function

$$p(x) \;=\; \begin{cases} \alpha \cdot |f(x)-\Omega| + (1-\alpha)\cdot res(x) & ,\ \text{if}\quad f(x) > \Omega \quad \text{or} \quad res(x) > 0 \\ -|f(x)-\Omega| & ,\ \text{if}\quad f(x) \le \Omega \quad \text{and} \quad res(x) = 0 \end{cases}$$

$$\alpha \;=\; \begin{cases} \dfrac{|f(x)-\Omega|\cdot\frac{6\sqrt{3}-2}{6\sqrt{3}} - res(x)}{|f(x)-\Omega| - res(x)} & ,\ \text{if}\quad f(x) > \Omega \quad \text{and} \quad res(x) < \frac{|f(x)-\Omega|}{3} \\ 1 - \dfrac{1}{2\sqrt{\frac{|f(x)-\Omega|}{res(x)}}} & ,\ \text{if}\quad f(x) > \Omega \quad \text{and} \quad \frac{|f(x)-\Omega|}{3} \le res(x) \le |f(x)-\Omega| \\ \frac{1}{2}\sqrt{\frac{|f(x)-\Omega|}{res(x)}} & ,\ \text{if}\quad f(x) > \Omega \quad \text{and} \quad res(x) > |f(x)-\Omega| \\ 0 & ,\ \text{if}\quad f(x) \le \Omega \end{cases}$$



(Graphical illustration of the extended oracle penalty function for Omega = 0)

ORACLE References:

[1] Schlueter, Gerdts: The Oracle Penalty Method, Journal of Global Optimization (Springer), Vol. 47, Issue 2, Pages 293-325 (2010). Preprint: http://www.midaco-solver.com/data/pub/The_Oracle_Penalty_Method.pdf.

[2] Munawar et al.: Advanced genetic algorithm to solve MINLP problems over GPU, IEEE Congress on Evolutionary Computation (CEC), 318 - 325 (2011).

[3] Dong, Cheng, Niu: A Constrained Particle Swarm Optimization Algorithm with Oracle Penalty Method, Applied Mechanics and Material (AMM) 303-306, 1519 (2013).

[4] Dong, Wang, Cheng, Jiang: Composite differential evolution with modified oracle penalty method for constrained optimization problems, Mathematical Problems in Engineering (2014). Preprint: http://downloads.hindawi.com/journals/mpe/aip/617905.pdf.

[5] Dong, Cheng, Niu: Adaptive Constrained Differential Evolution Algorithms based on Oracle Penalty Function, Computer Applications and Software (Chinese), DOI:10.3969/j.issn.1000-386x.2014.01.078 (2014).

[6] Gebreslassie, Diwekar: Efficient ant colony optimization for computer aided molecular design: case study solvent selection problem, Computers & Chemical Engineering (Elsevier) Volume 78, Pages 1–9 (2015). Preprint: http://www.midaco-solver.com/data/other/EACO.pdf.

[7] Gebreslassie, Diwekar: Efficient ant colony optimization (EACO) algorithm for deterministic optimization, AIChE Annual Meeting Atlanta, GA (2014). Preprint: http://www.midaco-solver.com/data/other/EACO.docx.

[8] Chen, Liang, An Optimization Approach of SRM Sphere Slot Grain Design Based on Improved Differential Evolution Algorithm, Advanced Materials Research, Vol 971-973, Pages 1072-1075 (2014).

[9] Gago-Vargas et al.: Nonlinear integer problems with applications to engineering design, Computational Optimization and Applications (Springer), http://dx.doi.org/10.1007/s10589-015-9739-3 (2015). Preprint: http://www.midaco-solver.com/data/other/COAP-2015.pdf.

[10] Sato, Watanabe, Igarashi: Multimaterial Topology Optimization of Electric Machines Based on Normalized Gaussian Network, IEEE Transaction on Magnetics, Vol. 51, No. 3, Art. 7202604 (2015).

The following MIDACO settings can be modified in DiscoverSim by clicking **Run Optimization > Advanced Settings**:

- **Maximum iterations**. Default = 10,000. Note that 1 iteration is equivalent to 1000 function evaluations so default is maximum of 1e7 function evaluations.
- **Maximum time**. Default 100,000 seconds.
- **Print frequency**. Results are displayed with specified frequency. Default = 1 (at every iteration, result is displayed).
- **Convergence criteria - iterations [0 - auto]**. The maximum number of iterations where the objective function has not changed (by more than 1e-9) to declare convergence. If set to 0 (auto), the convergence criteria varies with number of input controls as follows:
    - Number of input controls <= 4: Convergence criteria = 1000 (as noted above, this is 1e6 function evaluations)
    - Number of input controls = 5 - 9: Convergence criteria = 500
    - Number of input controls >= 9: Convergence criteria = 100

  The rationale for these automatic settings is the simpler the model (with fewer controls) the less computation time is expected to be required. You can also set the convergence criteria – iterations to any specified value.
- **Autostop [0 – disable]**. Automatic stopping criteria. Default = 0 (disabled).
  Examples of AUTOSTOP values and their possible impact on MIDACO (from MIDACO User Guide):
    - 1 Fastest runtime but very low chance of global optimality

  - 5 Fast runtime but low chance of global optimality
  - 50 Medium long runtime and medium chance of global optimality
  - 500 Long runtime but high chance of global optimality
  - 5000 Very long runtime but also very high chance of global optimality.
- **Focus [0 – auto]**. From MIDACO User Guide:
  - This parameter forces MIDACO to focus its search process around the current best solution. For many problems, tuning this parameter is useful and will result in a faster convergence speed (e.g., for convex and semi-convex problems). This parameter is also especially useful for refining solutions (e.g. to improve the precision of their objective function value or constraint violation).
  - If FOCUS is not equal to zero, MIDACO will apply an upper bound for the standard deviation of its Gauss PDF's. The upper bound for the standard deviation for continuous variables is given by (XU(i)-XL(i))/FOCUS, whereas the upper bound for the standard deviation for integer variables is given by MAX((XU(i)-XL(i))/FOCUS,1/SQRT(FOCUS)).
  - In other words, the larger the value of FOCUS, the closer MIDACO will concentrate its search around its current best solution.
  - The value for FOCUS must be an integer. Smaller values for FOCUS (e.g. 10 or 100) are recommended for first test runs (without a specific starting point). Larger values for FOCUS (e.g. 10000 or 100000) are normally only useful for refinement runs (where a specific solution is used as a starting point).
- **Ants [0 – auto]**. From MIDACO User Guide:
  - This parameter allows the user to fix the number of ants (solution candidates) which MIDACO generates within one generation (major iteration of the evolutionary ACO algorithm).
  - This parameter must be used in combination with KERNEL. Using the ANTS and KERNEL parameters can be promising for some problems (in esp. large scale problems or cpu-time intensive applications). However, tuning these parameters might also significantly reduce the MIDACO performance.
  - If ANTS parameter is equal to zero, MIDACO will dynamically change the number of ants per generation.
- **Kernel [0 – auto]**. From MIDACO User Guide:
  - This parameter allows the user to fix the number of kernels within MIDACO's multi-kernel Gauss PDF's. The kernel size corresponds also to the number of solutions stored in MIDACO's solution archive.
  - On rather convex problems it can be observed, that a lower kernel number will result in faster convergence while a larger kernel number will result in lower convergence. On the contrary, a lower kernel number will increase the risk of MIDACO getting stuck in a local optimum, while a larger kernel number increases the chance of reaching the global optimum.
  - The KERNEL parameter must be used in combination with the ANTS parameter. Some examples of possible ants/kernel settings are given and explained below.

Setting 1		Setting 2		Setting 3		Setting 4	
ANTS	2	ANTS	30	ANTS	500	ANTS	100
KERNEL	2	KERNEL	5	KERNEL	10	KERNEL	50

  - The 1st setting is the smallest possible one. This setting might be useful for very cpu-time expensive problems where only some hundreds of function evaluation are possible or for problems with a specific structure (e.g. convexity). The 2nd setting might also be used for cpu-time expensive problems, as a relatively low number of ANTS is considered. The 3rd and 4th setting would only be promising for problems with a fast evaluation time. As tuning the ants and kernel parameters is highly problem dependent, the user needs to experiment with those values.
  - Note that the maximum kernel number for MIDACO is fixed to 100.

- **Oracle eq. constraint tolerance.** Oracle equality constraint tolerance. Default = 0.001. From MIDACO User Guide:
  - This parameter defines the accuracy tolerance for the equality constraints G(X). An equality constraint is considered feasible by MIDACO, if |G(X)| <= Oracle equality constraint tolerance.
  - This parameter has strong influence on the MIDACO performance on equality constraint problems. For problems with difficult constraints, it is recommended to start with some test runs using a less precise tolerance (e.g. 0.05 or even 0.1) and to apply some refinement runs with a higher precision afterwards (e.g. 1e-4 or 1e-6).
- **Oracle constraint penalty [0 - auto]**. From MIDACO User Guide:
  - This parameter specifies a user given oracle parameter to the penalty function within MIDACO. This parameter is only relevant for constrained problems. If this is equal to zero, MIDACO will use an Oracle constraint penalty of 1e9, otherwise the specified value is used.
  - This option can be especially useful for constrained problems where some background knowledge on the problem exists. For example: It is known that a given application has a feasible solution X corresponding to F(X)=1000 (e.g. plant operating cost in Dollars). Therefore it might be reasonable to submit an oracle value of 800 or 600 to MIDACO, as this cost region might hold a new feasible solution (to operate the plant at this cost value). Whereas an oracle value of more than 1000 would be uninteresting to the user, while a too low value (e.g. 200) would be unreasonable.
  - See Oracle Penalty Method above for further information.

Genetic Algorithm (GA) Global Optimization

The Genetic Algorithm (GA) is a derivative free global optimization method that attempts to find a global optimum by simulating an evolutionary process. Each member of a population (a chromosome or solution) has a set of genes (parameter values). The members breed (crossover), and the genes can mutate. Successful chromosomes – i.e., those that are most fit, as evaluated by the optimization process - survive, while the rest die. The parameters associated with the successful genes are returned. See http://en.wikipedia.org/wiki/Genetic_algorithm for further details. An online interactive tutorial is available at: http://www.obitko.com/tutorials/genetic-algorithms/.

GA is a useful complement to MIDACO and with some problems such as all continuous input control variables, may be faster to find a global optimum.

The MIDACO Oracle Penalty Method for constraints described above has been implemented in DiscoverSim's GA.

The following GA settings can be modified by clicking **Run Optimization > Advanced Settings**:

- **Maximum iterations**. Default = 10,000.
- **Maximum time**. Default 100,000 seconds.
- **Print frequency**. Results are displayed with specified frequency. Default = 1 (at every iteration, result is displayed).
- **Convergence criteria - iterations [0 - auto]**. The maximum number of iterations where the objective function has not changed (by more than 1e-9) to declare convergence. If set to 0 (auto), the convergence criteria varies with number of input controls as follows:
    - Number of input controls <= 4: Convergence criteria = 1000
    - Number of input controls = 5 - 9: Convergence criteria = 500
    - Number of input controls >= 9: Convergence criteria = 100

  The rationale for these automatic settings is the simpler the model (with fewer controls) the less computation time is expected to be required. You can also set the convergence criteria – iterations to any specified value.
- **Convergence criteria - fit-stdev [0 - disable]**. Standard deviation of chromosome set of objective function must be less than this for 10 iterations to declare convergence. Default = .00001.
- **Chromosomes [0 – auto]**. Number of chromosomes (population size). Each chromosome (solution member) of the population has a set of genes which correspond to parameter values. The larger the population, the greater the variability which increases the chances of finding a global optimum, but the longer the estimation time. If set to default = 0 (auto), the number of chromosomes vary with the number of input controls as follows:
    - 10 * number of input control variables
    - Minimum = 40

  If the problem is highly non-linear with multiple minima (like the Schwefel function) we recommend 40* number of input control variables.

- **Matings**. Number of matings (crossovers). Multiple matings result in an increased population, but only the original population size is maintained by culling the least fit. Default = 4.
- **Prob. of mutation [0 – auto]**. Genes can mutate on a chance basis, specified by this probability. Typically a low probability of mutation works well for global optimization with continuous variables. If set to default = 0 (auto), the probability of mutation = 1/(10*number of input control variables).
- **Degree of mutation**. Should mutation occur, this specifies the degree of mutation. Large values make for a faster solution, but can result in poorer optimization near the end. Default = 0.5.
- **Oracle eq. constraint tolerance.** Oracle equality constraint tolerance. Default = 0.001. See MIDACO Oracle Penalty Method above for further details.

Discrete Exhaustive Global Optimization

Discrete Exhaustive optimization is applicable for small problems where all of the input control variables are discrete. All combinations are evaluated and the best solution that satisfies all constraints (if applicable) is selected.

Obviously if the number of combinations (computed as # levels ^ # variables – assuming same # levels for each variable) is large then this method will not be feasible. A user defined **Time Limit** (default = 120 seconds) sets the maximum computation time. However, you won't have to wait for this time to be exceeded if the problem is too large. After 50 trials, the average time per run is estimated and if this time, multiplied by the total number of combinations, exceeds the **Time Limit**, an error message is given and the optimization is stopped.

Practically this method will work well with binary discrete (2 levels) and up to 13 or 14 variables.

The following Discrete Exhaustive settings can be modified by clicking **Run Optimization > Advanced Settings**:

- **Maximum iterations**. Default = 10,000.
- **Print frequency**. Results are displayed with specified frequency. Default = 1 (at every iteration, result is displayed).

Sequential Quadratic Programming (SQP) Local Optimization

SQP is a very fast and robust algorithm for smooth local nonlinear continuous optimization. The objective function and the constraints must be twice continuously differentiable. The method is based on solving a series of subproblems designed to minimize a quadratic model of the objective subject to a linearization of the constraints, using the Simplex algorithm. If the problem is unconstrained, then the method reduces to Newton's method for finding a point where the gradient of the objective vanishes. For further details see http://en.wikipedia.org/wiki/Sequential_quadratic_programming.

The following SQP settings can be modified by clicking **Run Optimization > Advanced Settings**:

- **Maximum iterations**. Default = 10,000.
- **Maximum time**. Default 100,000 seconds.
- **Print frequency**. Results are displayed with specified frequency. Default = 1 (at every iteration, result is displayed).
- **Convergence criteria - parameter slope**. Slope of each parameter must be less than this for convergence. Default = 0.00001.

Nelder-Mead (NM) Local Optimization

The Nelder Mead algorithm is a fast downhill simplex method that provides a local optimum for an objective function when the gradients are not known (non-differentiable), so is suitable for non-smooth problems.

NM creates a simplex – an n+1 space – where each point (vertex) in the space represents an initial starting value. Each vertex is evaluated in terms of the criteria functions, the worst vertex is eliminated, and another vertex is added. The process is continued until no further reduction in the criteria function occurs.

The creation of the new vertex involves reflection, expansion, contraction and shrinkage – the simplex changes shape and size at each iteration. The parameters allow one to influence how this process occurs.

The MIDACO Oracle Penalty Method for constraints described above has been implemented in DiscoverSim's NM.

For further details see:

https://en.wikipedia.org/wiki/Nelder%E2%80%93Mead_method

http://www.scholarpedia.org/article/Nelder-Mead_algorithm

Nelder, John A.; R. Mead (1965). "A simplex method for function minimization." *Computer Journal* **7**: 308–313.doi:10.1093/comjnl/7.4.308.

The following NM settings can be modified by clicking **Run Optimization > Advanced Settings**:

- **Maximum iterations**. Default = 10,000.
- **Maximum time**. Default 100,000 seconds.
- **Print frequency**. Results are displayed with specified frequency. Default = 1 (at every iteration, result is displayed).
- **Convergence criteria - iterations [0 - auto]**. The maximum number of iterations where the objective function has not changed (by more than 1e-9) to declare convergence. If set to 0 (auto), the convergence criteria varies with number of input controls as follows:
    - Number of input controls <= 4: Convergence criteria = 1000
    - Number of input controls = 5 - 9: Convergence criteria = 500
    - Number of input controls >= 9: Convergence criteria = 100

  The rationale for these automatic settings is the simpler the model (with fewer controls) the less computation time is expected to be required. You can also set the convergence criteria – iterations to any specified value.
- **Reflection**. (Reflection > 0), Default = 1.
- **Contraction**. (0 < Contraction < 1), Default = 0.5.

- **Expansion**. (Expansion > 0), Default = 2.
- Shrinkage is hard coded = 0.5.
- **Initial Perturbation**. (Initial Simplex Perturbation > 0), Default = 0.1.
- **Oracle eq. constraint tolerance.** Oracle equality constraint tolerance. Default = 0.001. See MIDACO Oracle Penalty Method above for further details.
- **Oracle update frequency.** Default = 20. See MIDACO Oracle Penalty Method above for further details.

HYBRID Optimization

HYBRID optimization is a powerful hybrid of the above 5 methods that takes advantage of the strengths of each method:

1. All discrete controls: MIDACO or Exhaustive Discrete if applicable
2. All continuous controls: MIDACO, GA, followed by SQP or NM (if SQP fails due to non-smooth objective function). The solution from MIDACO provides starting values for GA. The solution for GA provides starting values for SQP/NM.
3. Mixed continuous/discrete controls: MIDACO 1 (Default settings), MIDACO 2 (Fine FOCUS = 100000), followed by SQP or NM (if SQP fails due to non-smooth objective function). The solution from MIDACO 1 provides starting values for MIDACO 2. The solution for MIDACO 2 provides starting values for SQP/NM.

Note that using Hybrid will be slower than using a specific method, so if a quick solution is required, it is recommended to select one of the specific methods.

The following Hybrid settings can be modified by clicking **Run Optimization > Advanced Settings**:

- **GA Maximum Iterations**. Default = 1000.
- **SQP Maximum Iterations**. Default = 1000.
- **NM Maximum Iterations**. Default = 1000
- **MIDACO 1 Maximum Iterations**. Default = 1000 (1e6 function evaluations)
- **MIDACO 2 Maximum Iterations**. Default = 1000 (1e6 function evaluations)

To change other settings for a specific method, select the method of interest and modify the respective settings. Going back to Hybrid, those modified settings will be used.

Distribution Fitting - Maximum Likelihood Estimation of Parameters

Maximum likelihood (ML) 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 observation.

Initial parameter values are derived from method of moments estimates or, if applicable, ordinary least squares.

The maximum likelihood parameter estimates are then calculated using the method of Broyden, Fletcher, Goldfarb, and Shanno (BFGS). The BFGS method approximates Newton's method, a class of hill-climbing optimization techniques that seeks a stationary point of a (preferably twice continuously differentiable) function. These methods use both the first and second derivatives of the function. However, BFGS has proven to have good performance even for non-smooth optimizations. For further details see:

https://en.wikipedia.org/wiki/Broyden%E2%80%93Fletcher%E2%80%93Goldfarb%E2%80%93Shanno_algorithm

If BFGS fails to converge, then the robust method of Nelder-Mead (described above) is employed.

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.

Distribution Fitting – Percentiles to Parameters Calculator

If one does not have sufficient data to do maximum likelihood distribution fitting but does know some percentile values of a specified distribution, these can be converted to parameter values for use in simulation.

Consider a distribution in which there are k parameters. Given only k observations, and a knowledge of the order (percentile or quantile) of each observation, one can estimate the k parameters.

As an example, consider estimating the scale (b) and shape (c) parameters for a gamma distribution. One has two observations, x_1 and x_2, which correspond to quantiles p_1 and p_2. Thus for $p_1 = .25$, and $p_2 = .75$, x_1 and x_2 correspond to the first and third quartile. Define $\Phi^{-1}(p, b, c)$ as the inverse gamma distribution function. We then have:

$$\Phi^{-1}(p_1, b, c) = x_1$$
$$\Phi^{-1}(p_2, b, c) = x_2$$

These parameters can then be solved, either analytically, or numerically. The numerical solution involves writing the equations in the canonical form ($f(v) = 0$), where v is the vector of parameters, specifying a starting value for v, and producing a sequence of estimates for v using a Newton algorithm that converge towards the root as a limit.

Reference for Percentiles to Parameters Calculator:

Cook, J.D. (2010). “Determining distribution parameters from quantiles,” http://www.johndcook.com/quantiles_parameters.pdf, University of Texas.

Distribution Fitting – Threshold Distributions

If the threshold parameter is not known, maximum likelihood is used to estimate the parameters. For some data sets, the likelihood function for threshold models is unbounded, and the maximum likelihood methodology fails (infinite likelihood). In other cases multiple modes are possible, leading to unstable results.

An alternative method, the iterative bias reduction procedure by Lockhart (1994), and discussed in William (2011), is a robust alternative to maximum likelihood. 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 (using cdf inverse), conditional on the current values of the remaining parameters (solved using BFGS maximum likelihood).

The probability for the minimum used in the cdf inverse has several possible choices as used in probability plotting positions, including the most common:

- (i-0.5)/N
- i/N
- i/(N+1)
- (i-0.3)/(N+0.4)

The minimum is specified with i=1. Using Monte Carlo small sample simulation we determined that cdf inverse((1-0.5)/N) produced the best overall result for threshold estimation (minimum bias against known). The exception to this was the 2 Parameter Exponential where the predicted minimum value = scale/N gave minimum bias for small samples.

Initial estimates of Threshold use bias reduction with method of moments (or least squares if applicable) on the other parameters.

In some cases maximum likelihood will produce the best parameter estimates (e.g. 3 Parameter Weibull with shape > 2), in others the iterative bias procedure is the best or only solution (e.g. 3 Parameter Weibull with shape < 1). In the case of 3 Parameter Weibull, one can use likelihood profiling to select which method to use, see Lockhart (1994), but this “switch” is not generally applicable to all threshold distributions. DiscoverSim will attempt to solve the threshold distribution using BFGS maximum likelihood as well as with iterative bias reduction. If maximum likelihood fails, then iterative bias reduction is used. If both successfully converge, the result producing the lowest Anderson Darling statistic is selected.

There are good alternative methods to the iterative bias reduction, such as Maximum Product Spacing (see Akram, 2012), but in the context of batch distribution fitting, we consider the above hybrid to be the best overall approach.

References for Threshold Distributions:

Akram, M. and Hayat, A. (2012) “Comparison of estimators of the Weibull Distribution” Financial Econometrics Series, SWP 2012/04, Deakin University Australia, http://www.deakin.edu.au/buslaw/aef/workingpapers/fin-econometrics/2012_04.pdf.

David, H.A. (1981). *Order Statistics*, John Wiley & Sons, New York.

Giesbrecht, F. and Kempthorne, A.H. (1966). “Maximum Likelihood Estimation in the Three-parameter Lognormal Distribution,” *Journal of the Royal Statistical Society*, B 38, pp. 257-264.

Lockhart, R.A. and Stephens, M.A. (1994). “Estimation and Tests of Fit for the Three-parameter Weibull Distribution,” *Journal of the Royal Statistical Society*, Vol.56(3), pp. 491-500.

William, V. H., Ted, G. E. & Thomas, R. J. (2011) “Concerns About Maximum Likelihood Estimation for the Three-Parameter Weibull Distribution: Case Study of Statistical Software,” *The American Statistician*, 65:1, 44-54, DOI: 10.1198/tast.2011.09103.

Distribution Fitting – Goodness of Fit

The Anderson Darling statistic measures how well the data fits a particular continuous distribution - the smaller the AD value, the better the fit. In general, when comparing several distributions, the distribution with the smallest AD statistic has the best fit to the data.

Anderson Darling p-values are obtained from tabular interpolation. These tables were produced using extensive Monte Carlo simulations, varying sample size and the shape parameter(s) - if applicable. A minimum of 50,000 replications were used to obtain the critical values.

For those distribution cases where published critical values exist, see D'Agostino & Stephens (1986), the simulation results were compared and found to be a close match (within 1% error for large sample asymptotic, 2.5% for small sample).

In batch mode, all of the distributions are sorted by the AD statistic, smallest to largest. If the data set is large (N > 500), then a random subset of 500 is selected and AD is estimated. The 10 distributions with the lowest AD statistic are then recomputed using the entire data set and resorted. This process dramatically speeds up the batch process for large data sets.

The Chi-Square goodness of fit test and associated p-value are used when distribution fitting with discrete distributions.

Reference for Anderson Darling Goodness of Fit:

D'Agostino, R.B. and M.A. Stephens (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

Distribution Fitting – Nonnormal Process Capability Indices

Z-Score Method (Default)

The Z-Score method for computing process capability indices utilizes 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 distribution and all nonnormal distributions, and short term capability indices Cp and Cpk can be estimated using the standard deviation from control chart methods on the transformed z-values. The Z-Score method was initially developed by Davis Bothe and expanded on by Andrew Sleeper. For further details, see Sleeper, *Six Sigma Distribution Modeling*.

Percentile (ISO) Method

The Percentile method to calculate process capability indices uses the following formulas:

Ppu = (USL – 50^{th} percentile)/(99.865 percentile – 50^{th} percentile)

Ppl = (50^{th} percentile – LSL)/(50^{th} percentile – 0.135 percentile)

Ppk = min(Ppu, Ppl)

Pp = (USL – LSL)/(99.865 percentile – 0.135 percentile)

Distribution Fitting – Nonnormal Control Charts

Individuals – Original Data

The **Individuals – Original Data** chart displays the untransformed data with control limits calculated as:

UCL = 99.865 percentile

CL = 50^{th} 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 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.

References for Nonnormal Process Capability Indices and Control Charts:

Bothe, D.R. (1997). *Measuring Process Capability*, McGraw-Hill, New York.

Chou,Y., A.M. Polansky, and R.L. Mason (1998). “Transforming Non-Normal Data to Normality in Statistical Process Control,” *Journal of Quality Technology*, Vol. 30(2), pp. 133-141.

Sleeper, A. (2006). *Six Sigma Distribution Modeling*, McGraw-Hill, New York.

Summary of Distributions

The following table is a summary of all distributions in DiscoverSim.

Common Continuous	Advanced Continuous	Discrete
• Normal • Triangular • Uniform • Pearson Family (specify Mean, StdDev, Skewness, Kurtosis) • Log Normal** • Exponential** • Weibull** • PERT • Custom Continuous	• Beta • Beta (4 Parameter) • Box-Cox** • Burr** • Cauchy • Chi-Squared** • Chi-Squared with Scale (3 Parameter)** • Error Function (ERF) • F** • F with Scale (4 Parameter)** • Fatigue Life** • Fisk** • Folded Normal • Frechet** • Gamma** • Generalized Error • Generalized Gamma** • Generalized Logistic • Generalized Pareto • Half Normal** • Inverse Gamma** • Inverse Gaussian • Johnson SB • Johnson SL • Johnson SU • Laplace • Largest Extreme Value • Levy • Logistic • Log Gamma**	• Bernoulli (Yes/No) • Binomial • Geometric • Hypergeometric • Logarithmic • Negative Binomial • Poisson • Step • Uniform (Discrete) • Custom Discrete • SIP (Stochastic Information Packet – Unweighted Sequential Custom Discrete)

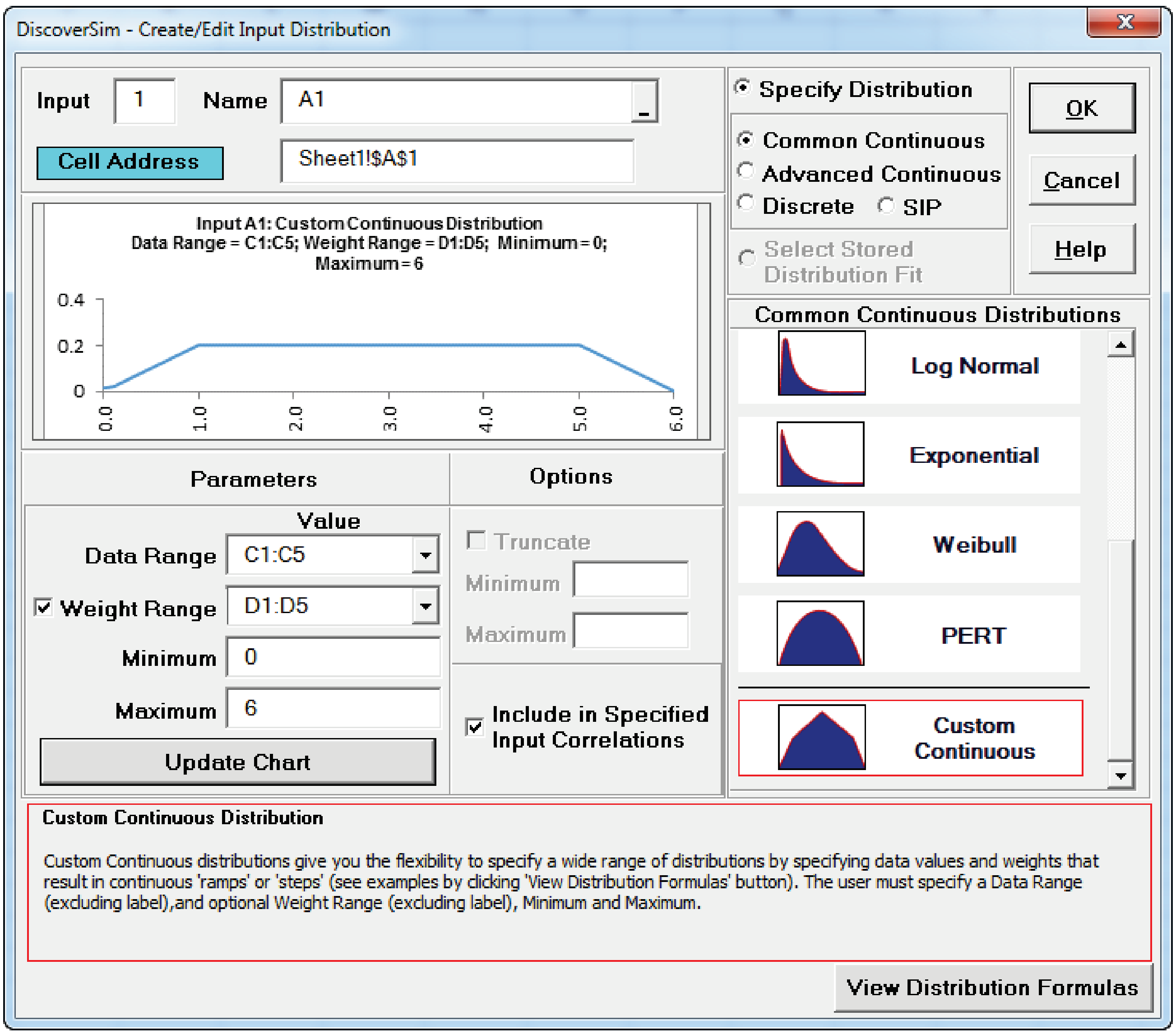
Common Continuous	Advanced Continuous	Discrete
	• Log Logistic** • Maxwell • Non-Central Chi-Squared** • Non-Central F** • Non-Central T • Pareto • Power • Rayleigh • Reciprocal • Skew Normal • Smallest Extreme Value • Student's T • Student's T with Location and Scale (3 Parameter) • Von Mises	

** denotes with/without Threshold

Custom Distributions

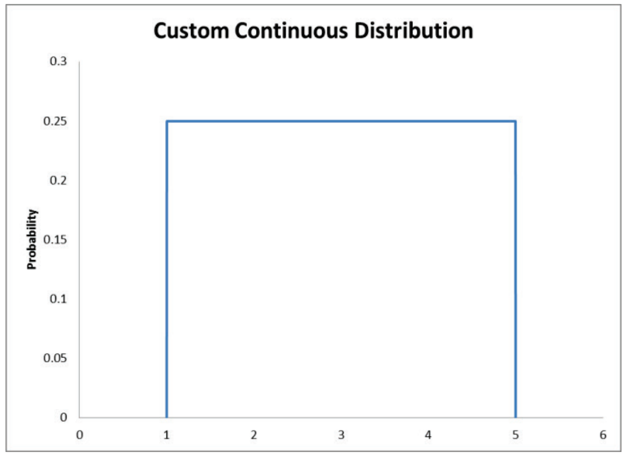
Custom Continuous

Custom continuous distributions give you the flexibility to specify a wide range of distributions by specifying data values and weights that result in continuous "ramps" or "steps" as shown in the following examples. The user must specify a Data Range (excluding label), and optional Weight Range (excluding label), Minimum and Maximum. If Weight Range is not specified a default weight of 1 is used for each data value. Upon selection of the Data Range, the default Minimum and Maximum are obtained from the data.



Example 1:

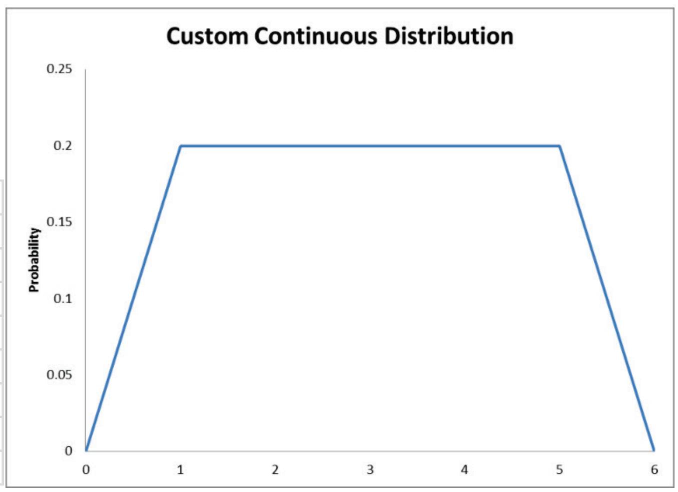
	Data	Weight
	1	1
	2	1
	3	1
	4	1
	5	1
Minimum	1	
Maximum	5	



In this example, the same results would be obtained with Weight Range unchecked.

Example 2:

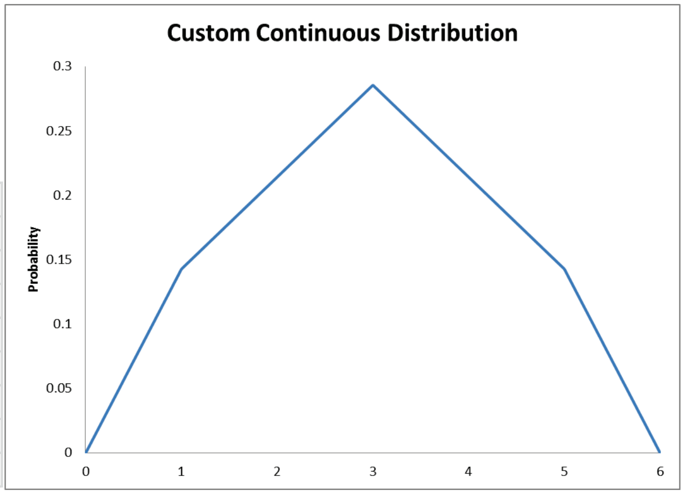
	Data	Weight
	1	1
	2	1
	3	1
	4	1
	5	1
Minimum	0	
Maximum	6	



In this example, the same results would be obtained with Weight Range unchecked.

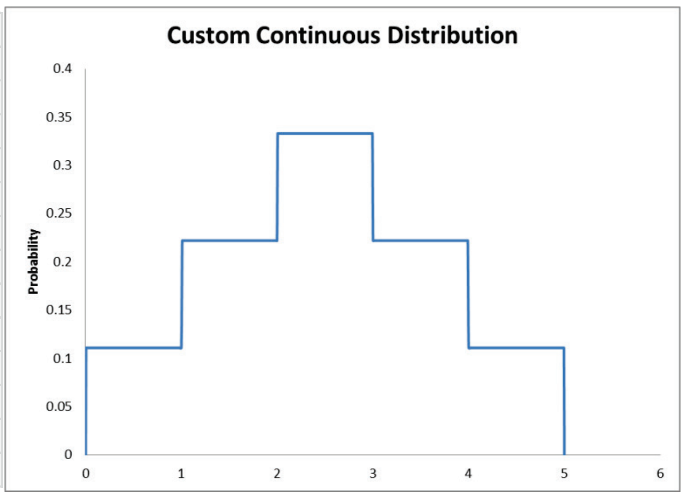
Example 3:

	Data	Weight
	1	2
	2	3
	3	4
	4	3
	5	2
Minimum	0	
Maximum	6	



Example 4:

	Data	Weight
	0	1
	1	1
	1	2
	2	2
	2	3
	3	3
	3	2
	4	2
	4	1
	5	1
Minimum	0	
Maximum	5	



Formulas for Custom Continuous Distribution

Probability Density Function (PDF):

$$p_i + (\frac{x - x_i}{2(x_{i+1} - x_i)})(p_{i+1} - p_i) \quad x_i \leq x \leq x_{i+1}$$

Cumulative Distribution Function (CDF):

$$F(x_i) + (x - x_i)(p_i + \frac{(x - x_i)(p_{i+1} - p_i)}{2(x_{i+1} - x_i)}) \quad x_i \leq x \leq x_{i+1}$$

where $p_i = w_i / \sum_{j=1}^{N} (w_j)$ and $F(x_i) = \sum_{j=1}^{i} (p_j)$.

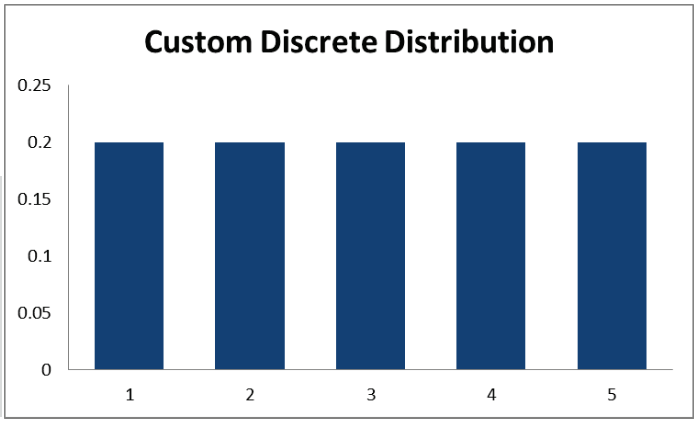
N ordered data values $x_{min} < x < x_{max}$.
Probability weights $w > 0$.
Minimum value x_{min} ($w_{x_{min}} = 0$)
Maximum value x_{max} ($w_{x_{max}} = 0$)

Custom Discrete

Custom discrete distributions give you the flexibility to specify a wide range of distributions by specifying integer data values and weights as shown in the following examples. The user must specify a Data Range (excluding label) and Weight Range (excluding label).

Example 1:

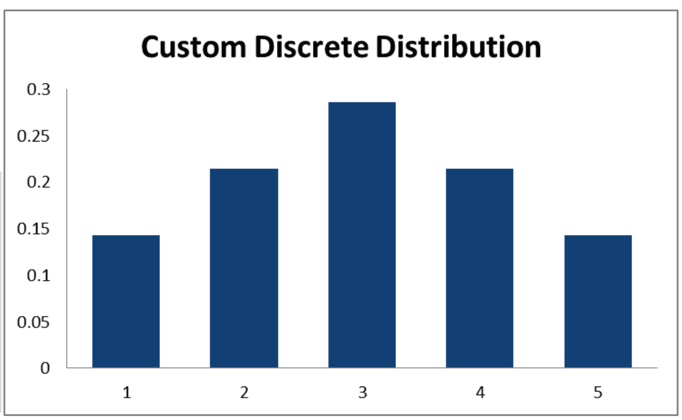
Data	Weight
1	1
2	1
3	1
4	1
5	1



In this example, the same results would be obtained with Weight Range unchecked.

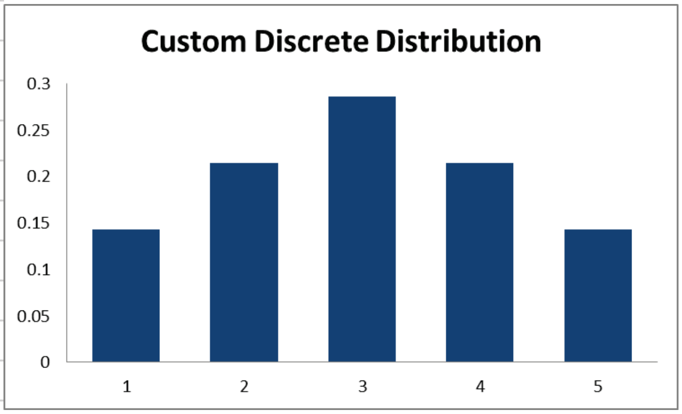
Example 2a (Weighted):

Data	Weight
1	2
2	3
3	4
4	3
5	2



Example 2b (Unweighted):

Data
1
1
2
2
2
3
3
3
3
4
4
4
5
5



Formulas for Custom Discrete Distribution

Probability Density Function (PDF):

$$p_i \quad x = x_i$$

Cumulative Distribution Function (CDF):

$$\sum_{j=1}^{i} (p_j) \quad x = x_i$$

where $p_i = w_i / \sum_{j=1}^{N} (w_j)$.

N data values x.
Probability weights $w > 0$.

SIP Import

A SIP is a Stochastic Information Packet that is a standard library of data (see ProbabilityManagement.org for more details on this standard). DiscoverSim treats this as a Custom Discrete distribution with unweighted data and is sampled sequentially (unrandomized).

A SIP cannot be included in Specified Input Correlations, nor can it be truncated. The number of rows (nrows) in the SIP data should match the number of replicates (nreps), or nreps/nrows should be evenly divisible.

Formulas for Continuous Distributions

For x to be distributed by a continuous distribution, x must be continuous and smooth over the specified range. The formulas provided are the Probability Density Function (PDF) and Cumulative Distribution Function (CDF).

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 (see Functions below).

Range $0 \leq x \leq 1$.
Shape1 parameter $\alpha > 0$.
Shape2 parameter $\beta > 0$.

Beta Distribution with Minimum and Maximum

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 $\theta 1 \leq x < \theta 2$.
Shape1 parameter $\alpha > 0$.
Shape2 parameter $\beta > 0$.
Minimum (lower threshold) θ_1
Maximum (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.

BoxCox Distribution

PDF

$$\frac{x^{\lambda-1}}{\sigma\sqrt{2\pi}}e^{-((x^{\lambda}-1)/\lambda-\mu)^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{(x^{\lambda}-1)/\lambda}\frac{1}{\sigma\sqrt{2\pi}}e^{-(t-\mu)^2/2\sigma^2}dt$$

Range $0 < x < \infty$.
Location parameter, μ, the mean.
Scale parameter, $\sigma > 0$, the standard deviation.
Shape parameter λ.

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.

BoxCox 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.

Burr Distribution

PDF

$$\frac{ck(x/\beta)^{c-1}}{\beta(1+(x/\beta)^c)^{k+1}}$$

CDF

$$1-(1+(x/\beta)^c)^{-k}$$

Range $0 \leq x < \infty$.
Scale parameter $\beta > 0$.
Shape parameter, $c > 0$.
Shape parameter, $k > 0$.

Burr Distribution with Threshold

PDF

$$\frac{ck((x-\theta)/\beta)^{c-1}}{\beta(1+((x-\theta)/\beta)^c)^{k+1}}$$

CDF

$$1-(1+((x-\theta)/\beta)^c)^{-k}$$

Range $0 \leq x - \theta < \infty$.
Scale parameter $\beta > 0$.
Shape parameter, $c > 0$.
Shape parameter, $k > 0$.
Threshold parameter $\theta < min(x)$.

Cauchy Distribution

PDF

$$\left(\pi\beta\left[1+\left(\frac{x-\alpha}{\beta}\right)^2\right]\right)^{-1}$$

CDF

$$0.5+\frac{1}{\pi}\tan^{-1}\left(\frac{x-\alpha}{\beta}\right)$$

Range $-\infty < x < \infty$.
Location parameter α, the median.
Scale parameter $\beta > 0$.

Chi-Squared Distribution

PDF

$$\frac{x^{v/2}\exp(-x/2)}{2^{v/2}\Gamma(v/2)}$$

CDF

$$\frac{\gamma(v/2, x/2)}{\Gamma(v/2)}$$

where $\Gamma(k)$ is the Gamma function, and $\gamma(k, z)$ is the lower incomplete Gamma function. See Functions below.

Range $0 \leq x \leq \infty$.
Shape parameter $v > 0$, the degrees of freedom.

Chi-Squared Distribution with Threshold

PDF

$$\frac{(x-\theta)^{v/2}\exp(-(x-\theta)/2)}{2^{v/2}\Gamma(v/2)}$$

CDF

$$\frac{\gamma(v/2, (x-\theta)/2)}{\Gamma(v/2)}$$

where $\Gamma(k)$ is the Gamma function, and $\gamma(k, z)$ is the lower incomplete Gamma function.

Range $0 \leq x - \theta \leq \infty$.
Shape parameter $v > 0$, the degrees of freedom.
Threshold parameter $\theta < min(x)$.

Chisq3: Chi-Squared Distribution with Scale

PDF

$$\frac{(x/\beta)^{.5\nu-1}e^{-0.5x/\beta}}{\beta \ \ 2^{.5\nu}\Gamma(.5\nu)}$$

CDF

$$\int_0^{.5x/\beta} \frac{e^{-t}t^{.5\nu-1}}{\Gamma(.5\nu)}dt$$

where Γ is the gamma function.

Range $0 \leq x < \infty$.
Shape parameter $\nu > 0$.
Scale parameter $\beta > 0$.

Chisq3: Chi-Squared Distribution with Scale and Threshold

PDF

$$\frac{((x-\theta)/\beta)^{.5\nu-1}e^{-0.5(x-\theta)/\beta}}{\beta\ 2^{.5\nu}\Gamma(.5\nu)}$$

CDF

$$\int_0^{.5(x-\theta)/\beta}\frac{e^{-t}t^{.5\nu-1}}{\Gamma(.5\nu)}dt$$

where Γ is the gamma function.

Range $0 \leq x-\theta < \infty$.
Shape parameter $\nu > 0$.
Scale parameter $\beta > 0$.
Threshold parameter $\theta < min(x)$.

Custom Continuous Distribution

PDF

$$p_i + (\frac{x-x_i}{2(x_{i+1}-x_i)})(p_{i+1}-p_i) \quad x_i \leq x \leq x_{i+1}$$

CDF

$$F(x_i) + (x-x_i)(p_i + \frac{(x-x_i)(p_{i+1}-p_i)}{2(x_{i+1}-x_i)}) \quad x_i \leq x \leq x_{i+1}$$

where $p_i = w_i / \sum_{j=1}^{N}(w_j)$ and $F(x_i) = \sum_{j=1}^{i}(p_j)$.

N ordered data values $x_{min} < x < x_{max}$.
Probability weights $w > 0$.
Minimum value x_{min} ($w_{x_{min}} = 0$)
Maximum value x_{max} ($w_{x_{max}} = 0$)

ERF Distribution

PDF

$$\frac{\beta e^{-(\beta x)^2}}{\sqrt{\pi}}$$

CDF

$$\int_{-\infty}^{x}\frac{e^{-(\beta t)^2}}{\sqrt{\pi}}dt$$

Range $-\infty < x < \infty$.
Scale parameter $\beta > 0$.

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 $0 \leq x - \theta < \infty$.
Scale parameter, $\alpha > 0$, the mean.
Threshold parameter $\theta < min(x)$.

F Distribution

PDF

$$\frac{(v/w)^{v/2} x^{(v/2-1)}}{B(v/2, w/2)[1 + (v/w)x]^{(v+w)/2)}}$$

CDF

$$I_{\frac{vx}{vx+w}}(v/2, w/2)$$

where $B(a, b)$ is the Beta function, and where $I_x(a, b)$ is the regularized incomplete Beta function.

Range $0 \leq x \leq \infty$.
Shape1 parameter $v > 0$, integer, first degrees of freedom.
Shape2 parameter $w > 0$, integer, second degrees of freedom.

F Distribution with Threshold

PDF

$$\frac{(v/w)^{v/2}(x-\theta)^{(v/2-1)}}{B(v/2,w/2)[1+(v/w)(x-\theta)]^{(v+w)/2)}}$$

CDF

$$I_{\frac{vx}{v(x-\theta)+w}}(v/2,w/2)$$

where B is the Beta function, and where $I_x(a,b)$ is the regularized incomplete Beta function.

Range $0 \le x-\theta < \infty$.
Shape1 parameter $v > 0$, integer, first degrees of freedom.
Shape2 parameter $w > 0$, integer, second degrees of freedom.
Threshold parameter $\theta < min(x)$.

F4: F Distribution with Scale

PDF

$$\frac{1}{x\,\mathrm{B}(.5\nu,.5\omega)}\sqrt{\frac{(\nu\,x/\alpha)^{\nu}\;\;\omega^{\omega}}{(\nu\,x/\alpha+\omega)^{\nu+\omega}}}$$

CDF

$$\int_{-\infty}^{z}\frac{1}{B(.5\,\nu,.5\,\omega)}t^{.5\nu-1}(1-t)^{.5\omega-1}dt$$

where

$$z=\frac{(\nu\,x)}{(\nu\,x+\alpha\omega)}$$

and where B is the Beta function.

Range $0 \le x < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\nu > 0$.
Shape parameter $\omega > 0$.

F4: F Distribution with Scale and Threshold

PDF

$$\frac{1}{(x-\theta)\,\mathrm{B}(.5\nu, .5\omega)}\sqrt{\frac{(\nu\,(x-\theta)/\alpha)^{\nu}\ \ \omega^{\omega}}{(\nu\,(x-\theta)/\alpha+\omega)^{\nu+\omega}}}$$

CDF

$$\int_{-\infty}^{z}\frac{1}{B(.5\,\nu, .5\,\omega)}t^{.5\nu-1}(1-t)^{.5\omega-1}dt$$

where

$$z=\frac{(\nu\,(x-\theta))}{(\nu\,x+\alpha\omega)}$$

and where B is the Beta function.

Range $0 \leq x-\theta < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\nu > 0$.
Shape parameter $\omega > 0$. Threshold parameter $\theta < min(x)$.

Fatigue Life Distribution

PDF

$$\frac{\sqrt{\frac{x}{\beta}}+\sqrt{\frac{\beta}{x}}}{2\gamma x)}\phi\left(\frac{\sqrt{\frac{x}{\beta}}-\sqrt{\frac{\beta}{x}}}{\gamma}\right)$$

CDF

$$\Phi\left(\frac{\sqrt{x}-\sqrt{\frac{1}{x}}}{\gamma}\right)$$

where $\phi(x)$ and $\Phi(x)$ are respectively the PDF and CDF of the standard normal distribution.

Range $0 < x < \infty$.
Scale parameter, $\beta > 0$.
Shape parameter $\gamma > 0$.

Notes: This is also known as the Birnbaum Saunders distribution.

Fatigue Life Distribution with Threshold

PDF

$$\frac{\sqrt{\frac{x-\theta}{\beta}}+\sqrt{\frac{\beta}{x-\theta}}}{2\gamma(x-\theta)}\phi\left(\frac{\sqrt{\frac{x-\theta}{\beta}}-\sqrt{\frac{\beta}{x-\theta}}}{\gamma}\right)$$

CDF

$$\Phi\left(\frac{\sqrt{x-\theta}-\sqrt{\frac{1}{x-\theta}}}{\gamma}\right)$$

where $\phi(x)$ and $\Phi(x)$ are respectively the PDF and CDF of the standard normal distribution.

Range $0 < x-\theta < \infty$.
Scale parameter, $\beta > 0$.
Shape parameter $\gamma > 0$.
Threshold parameter $\theta < min(x)$.

Notes: This is also known as the Birnbaum Saunders distribution.

Fisk Distribution

PDF

$$\frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{[1+(x/\alpha)^{\beta}]^2}$$

CDF

$$11+(x/\alpha)^{-\beta}$$

Range $0 < x < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter, $\beta > 0$.

Fisk Distribution with Threshold

PDF

$$\frac{(\beta/\alpha)((x-\theta)/\alpha)^{\beta-1}}{[1+((x-\theta)/\alpha)^{\beta}]^{2}}$$

CDF

$$11+((x-\theta)/\alpha)^{-\beta}$$

Range $0 < x - \theta < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter, $\beta > 0$.
Threshold parameter $\theta < min(x)$.

Folded Normal Distribution

PDF

$$\frac{\sqrt{2\pi}}{\sigma}\cosh(\mu/\sigma^2)e^{-(x^2+\mu^2)/2\sigma^2}$$

CDF

$$\Phi\left(\frac{x-\mu}{\sigma}\right)-\Phi\left(\frac{-x-\mu}{\sigma}\right)$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

Range $0 \leq x < \infty$.
Location parameter, μ, the mean.
Scale parameter, $\sigma > 0$, the standard deviation.

Frechet Distribution

PDF

$$(\beta/\alpha)(\alpha/x)^{1+\beta}e^{-(\alpha/x)^{\beta}}$$

CDF

$$e^{-(\alpha/x)^{\beta}}$$

Range $0 \leq x < \infty$.
Scale parameter, $\alpha > 0$.
Shape parameter $\beta > 0$.

Frechet Distribution with Threshold

PDF

$$(\beta/\alpha)(\alpha/(x-\theta))^{1+\beta}e^{-(\alpha/(x-\theta))^{\beta}}$$

CDF

$$e^{-(\alpha/(x-\theta))^{\beta}}$$

Range $0 \le x - \theta < \infty$.
Scale parameter, $\alpha > 0$.
Shape parameter $\beta > 0$.
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 the Gamma function.

Gamma Distribution with Threshold

PDF

$$\frac{[(x-\theta)/\alpha])^{\beta-1}e^{-(x-\theta)/\alpha}}{\alpha\Gamma(\beta)}$$

CDF

$$\int_0^{(x-\theta)/\alpha} \frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

where $\Gamma(\beta)$ is the the Gamma function:
Range $0 \le x - \theta < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.
Threshold parameter $\theta < min(x)$.

Generalized Error Distribution

PDF

$$\frac{\beta}{2\alpha\Gamma(1/\beta)} \; e^{-(|x-\mu|/\alpha)^{\beta}}$$

CDF

$$\frac{1}{2} + \operatorname{sign}(x-\mu)\frac{\gamma\left(1/\beta, \left(\frac{|x-\mu|}{\alpha}\right)^{\beta}\right)}{2\Gamma(1/\beta)}$$

where $\Gamma(\beta)$ is the the Gamma function and $\gamma(k, z)$ is the lower incomplete Gamma function.

Range $-\infty \le x < \infty$.
Location parameter μ. Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.

Notes: This is also known as the Exponential Power distribution or the Generalized Normal distribution.

Generalized Gamma Distribution

PDF

$$\frac{px^{(pk-1)}e^{(-x/\alpha^{p})}}{\alpha^{kp}\Gamma(k)}$$

CDF

$$\gamma(k, (x/\alpha)^{p})$$

where $\Gamma(\beta)$ is the the Gamma function and $\gamma(k, z)$ is the lower incomplete Gamma function.

Range $0 \le x < \infty$.
Scale parameter $\alpha > 0$.
Shape1 parameter $k > 0$.
Shape2 parameter $p > 0$.

Generalized Gamma Distribution with Threshold

PDF

$$\frac{p(x-\theta)^{(pk-1)}e^{(-(x-\theta)/\alpha^{p})}}{\alpha^{kp}\Gamma(k)}$$

CDF

$$\gamma(k,[(x-\theta)/\alpha]^{p})$$

where $\Gamma(\beta)$ is the the Gamma function and $\gamma(k,z)$ is the lower incomplete Gamma function.

Range $0 \leq x-\theta < \infty$.
Scale parameter $\alpha > 0$.
Shape1 parameter $k > 0$.
Shape2 parameter $p > 0$.
Threshold parameter $\theta < min(x)$.

Generalized Logistic Distribution

PDF

$$\frac{\alpha e^{-(x-\mu)/\sigma}}{\sigma(1+e^{-(x-\mu)/\sigma})^{1+\alpha}}$$

CDF

$$\frac{1}{(1+e^{-(x-\mu)/\sigma})^{\alpha}}$$

Range $-\infty < x < \infty$.
Location parameter, μ.
Scale parameter $\sigma > 0$.
Skew parameter α, < 1 for left skew, > 1 for right skew.

Notes: This is a Type I Generalized Logistic distribution; it is also known as the Skew-Logistic distribution.

Generalized Pareto Distribution

PDF

$$\frac{1}{\alpha}\left(1+\beta\ \frac{x-\mu}{\alpha}\right)^{-(1+1/\beta)}$$

CDF

$$1-\left(1+\beta\ \frac{x-\mu}{\alpha}\right)^{-1/\beta}$$

Range $0 < x < \infty$.
Location parameter μ.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.

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.

Half Normal Distribution with Threshold.

PDF

$$\frac{\sqrt{2/\pi}}{\sigma^2} e^{-(x-\theta)^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{x-\theta} \frac{\sqrt{2/\pi}}{\sigma} e^{-t^2/2\sigma^2} dt - 1$$

Range $0 \leq x - \theta < \infty$.
Scale parameter, $\sigma > 0$, the standard deviation.
Threshold parameter $\theta < min(x)$.

Inverse Gamma Distribution

PDF

$$\frac{\alpha^\beta}{\Gamma(\beta)} x^{-\beta-1} e^{-\alpha/x}$$

CDF

$$\frac{\gamma(\beta, \alpha/x)}{\Gamma(\beta)}$$

where $\Gamma(s)$ is the the gamma function, and $\gamma(s, x)$ is the lower incomplete gamma function.

Range $0 \leq x < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.

Inverse Gamma Distribution with Threshold

PDF

$$\frac{\alpha^{\beta}}{\Gamma(\beta)}(x-\theta)^{-\beta-1}e^{-\alpha/(x-\theta)}$$

CDF

$$\frac{\gamma(\beta,\alpha/(x-\theta))}{\Gamma(\beta)}$$

where $\Gamma(s)$ is the the gamma function, and $\gamma(s,x)$ is the incomplete gamma function.

Range $0 \leq x-\theta < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.
Threshold parameter $\theta < min(x)$.

Inverse Gaussian Distribution

PDF

$$\left[\frac{\lambda}{2\pi x^3}\right]^{1/2} e^{-\lambda(x-\mu)^2/(2\mu^2 x)}$$

CDF

$$\Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right)+e^{2\lambda/\mu}\,\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right)$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

Range $0 < x < \infty$.
Location parameter, μ, the mean.
Shape parameter $\lambda > 0$.

Johnson SB Distribution

PDF

$$\frac{\delta e^{-.5(\gamma+\delta\ln(z/(1-z)))^2}}{\lambda\sqrt{2\pi}z(1-z)}$$

CDF

$$\Phi\left[\gamma+\delta\ln\left(\frac{z}{1-z}\right)\right]$$

where $z=(x-\eta)/\lambda$ and $\Phi(x)$ is the CDF of the standard normal distribution.

Range $\eta<x<\eta+\lambda$.
Location parameter, η, the mean.
Scale parameter $\lambda>0$.
Shape1 parameter γ.
Shape2 parameter $\delta>0$.

Johnson SL Distribution

PDF

$$\frac{\delta\phi\left[\gamma+\delta\ln\left(\frac{x-\eta}{\lambda}\right)\right]}{(x-\eta)}$$

CDF

$$\Phi\left[\gamma+\delta\ln\left(\frac{x-\eta}{\lambda}\right)\right]$$

where $\phi(x)$ and $\Phi(x)$ are respectively the PDF and CDF of the standard normal distribution.

Range $-\eta<x<\infty$.
Location parameter, η, the mean.
Scale parameter $\lambda=1$.
Shape1 parameter γ.
Shape2 parameter δ.

Johnson SU Distribution

PDF

$$\frac{\delta e^{-.5(\gamma+\delta\sinh^{-1}(z))^2}}{\lambda\sqrt{2\pi(z^2+1)}}$$

CDF

$$\Phi(\gamma+\delta\sinh^{-1}(z))$$

where $z=(x-\eta)/\lambda$ and $\Phi(x)$ is the CDF of the standard normal distribution.

Range $-\infty < x < \infty$.
Location parameter, η, the mean.
Scale parameter $\lambda > 0$.
Shape1 parameter γ.
Shape2 parameter $\delta > 0$.

Laplace Distribution

PDF

$$\frac{1}{2\sigma}e^{-\frac{|x-\mu|}{\sigma}}$$

CDF

$$\frac{1}{2\sigma}e^{-\frac{\mu-x}{\sigma}} \quad if \quad x < \mu$$

$$\frac{1}{2\sigma}e^{-\frac{x-\mu}{\sigma}} \quad if \quad x \geq \mu$$

Range $-\infty < x < \infty$.
Location parameter, μ, the mean.
Scale parameter $\sigma > 0$.

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 $\sigma > 0$.

Notes: The Gumbel distribution is equivalent to the Largest Extreme Value.

Levy Distribution

PDF

$$\sqrt{\frac{\sigma}{2\pi}}\frac{e^{-\sigma/2x}}{x^{3/2}}$$

CDF

$$erfc\left(\sqrt{\sigma/2x}\right)$$

Range $0 < x < \infty$.
Scale parameter $\sigma > 0$.

Levy Distribution with Threshold

PDF

$$\sqrt{\frac{\sigma}{2\pi}}\frac{e^{-\sigma/2(x-\theta)}}{(x-\theta)^{3/2}}$$

CDF

$$erfc\left(\sqrt{\sigma/2(x-\theta)}\right)$$

Range $0 < x - \theta < \infty$.
Scale parameter $\sigma > 0$.
Threshold parameter $\theta < min(x)$.

Log Gamma Distribution

PDF

$$\frac{\ln(x)^{\beta-1}e^{-\ln(x)/\alpha}}{x\alpha^{\beta}\Gamma(\beta)}$$

CDF

$$\int_0^{\ln(x)/\alpha}\frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

where Γ is the gamma function.

Range $0 \le x < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.

Log Gamma Distribution with Threshold

PDF

$$\frac{\ln(x-\theta)^{\beta-1}e^{-\ln(x-\theta)/\alpha}}{(x-\theta)\alpha^{\beta}\Gamma(\beta)}$$

CDF

$$\int_0^{\ln(x-\theta)/\alpha}\frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

where Γ is the gamma function.

Range $0 < x-\theta < \infty$.
Scale parameter $\alpha > 0$.
Shape parameter $\beta > 0$.
Threshold parameter $\theta < min(x)$.

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 $\sigma > 0$.

Log Logistic 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 $\sigma > 0$.

Log Logistic 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 $0 < x - \theta < \infty$.
Location parameter, μ, the mean.
Scale parameter $\sigma > 0$.
Threshold parameter $\theta < min(x)$.

Log Normal 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 $0 < x < \infty$.
Location parameter, μ, the mean of $\ln(x)$.
Scale parameter, $\sigma > 0$, the standard deviation of $\ln(x)$.

Log Normal 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 $0 < x - \theta < \infty$.
Scale parameter, μ, the mean of $\ln(x)$.
Shape parameter, $\sigma > 0$, the standard deviation of $\ln(x)$.
Threshold parameter $\theta < min(x)$.

Maxwell Boltzmann Distribution

PDF

$$\sqrt{\frac{2}{\pi}}\frac{x^2 e^{-x^2/(2a^2)}}{a^3}$$

CDF

$$\gamma\left(1.5, \frac{x^2}{2a^2}\right)$$

where $\gamma(s, z)$ is the lower incomplete Gamma function.

Range $0 < x < \infty$.
Scale parameter $a > 0$.

Maxwell Boltzmann Distribution with Threshold

PDF

$$\sqrt{\frac{2}{\pi}}\frac{(x-\theta)^2 e^{-(x-\theta)^2/(2a^2)}}{a^3}$$

CDF

$$\gamma\left(1.5, \frac{(x-\theta)^2}{2a^2}\right)$$

where $\gamma(s, z)$ is the lower incomplete Gamma function.

Range $0 < x - \theta < \infty$.
Scale parameter $a > 0$.
Threshold parameter $\theta < min(x)$.

Non-Central Chi-Squared Distribution

PDF

$$.5e^{-(x+.5\lambda)}\left(\frac{x}{\lambda}\right)^{.25\nu-.5} I_{.5\nu-1}(\sqrt{\lambda x})$$

CDF

$$\sum_{j=0}^{\infty} e^{-.5\lambda}\frac{(.5\lambda)^j}{j!}\frac{\gamma(j+.5\nu, .5x)}{\Gamma(j+.5k)}$$

where $\Gamma(s)$ is the gamma function, $\gamma(s, x)$ is the lower incomplete gamma function. and I is the modified Bessel function of the first kind.

Range $0 \leq x < \infty$.
Shape parameter $\nu > 0$.
Non-centrality parameter $\lambda > 0$.

Non-Central Chi-Squared Distribution with Threshold

PDF

$$.5e^{-(x-\theta+.5\lambda)}\left(\frac{x-\theta}{\lambda}\right)^{.25\nu-.5} I_{.5\nu-1}(\sqrt{\lambda(x-\theta)})$$

CDF

$$\sum_{j=0}^{\infty} e^{-.5\lambda}\frac{(.5\lambda)^j}{j!}\frac{\gamma(j+.5\nu,.5(x-\theta))}{\Gamma(j+.5k)}$$

where $\Gamma(s)$ is the gamma function, $\gamma(s,x)$ is the lower incomplete gamma function. and I is the modified Bessel function of the first kind.

Range $0 \le x-\theta < \infty$.
Shape parameter $\nu > 0$.
Non-centrality parameter $\lambda > 0$.
Threshold parameter $\theta < min(x)$.

Non-Central F Distribution

PDF

$$\sum_{k=0}^{\infty} \frac{e^{-.5\lambda}(.5\lambda)^k}{B(.5\nu_2,.5\nu_1+k)k!}\left(\frac{\nu_1}{\nu_2}\right)^{.5\nu_1+k}\left(\frac{\nu_2}{\nu_2+\nu_1 x}\right)^{.5(\nu_1+\nu_2)+k} x^{.5\nu_1-1+k}$$

CDF

$$\sum_{j=0}^{\infty} \frac{(.5\lambda)^j\ e^{-.5\lambda}}{j!}B(z;\ .5\nu_1+j,.5\nu_2)$$

where

$$z=\frac{\nu_1\, x}{\nu_1 x+\nu_2}$$

where $B(a,b)$ is the beta function, and $B(z;\ a,b)$ is the incomplete beta function.

Range $0 \le x < \infty$.
Shape parameter $\nu_1 > 0$.
Shape parameter $\nu_2 > 0$.
Non-centrality parameter $\lambda > 0$.

Non-Central F Distribution with Threshold

PDF

$$\sum_{k=0}^{\infty} \frac{e^{-.5\lambda}(.5\lambda)^k}{B(.5\nu_2, .5\nu_1 + k)k!} \left(\frac{\nu_1}{\nu_2}\right)^{.5\nu_1+k} \left(\frac{\nu_2}{\nu_2 + \nu_1(x-\theta)}\right)^{.5(\nu_1+\nu_2)+k} (x-\theta)^{.5\nu_1-1+k}$$

CDF

$$\sum_{j=0}^{\infty} \frac{(.5\lambda)^j \; e^{-.5\lambda}}{j!} B(z;\; .5\nu_1 + j, .5\nu_2)$$

where

$$z = \frac{\nu_1(x-\theta)}{\nu_1(x-\theta) + \nu_2}$$

where $B(a, b)$ is the beta function, and $B(z;\; a, b)$ is the incomplete beta function.

Range $0 \le x - \theta < \infty$.
Shape parameter $\nu_1 > 0$.
Shape parameter $\nu_2 > 0$.
Non-centrality parameter $\lambda > 0$. Threshold parameter $\theta < min(x)$.

Non-Central T Distribution

PDF

$$\frac{\nu^{.5\nu} e^{-\nu\lambda^2/(2x^2+2\nu)}}{\sqrt{\pi}\Gamma(.5\nu)\, 2^{.5(\nu-1)}(x^2+\nu)^{.5(\nu+1)}} \int_0^{\infty} t^{\nu} e^{-.5(t-\lambda x/\sqrt{x^2+\nu}\,)^2}\, dt$$

CDF ($x \ge 0$)

$$\Phi(-\lambda) + \frac{1}{2}\sum_{j=0}^{\infty} \left[p_j I_z\left(j + \frac{1}{2}, \frac{\nu}{2}\right) + q_j \beta_z\left(j+1, \frac{\nu}{2}\right)\right]$$

where

$$z = \frac{x^2}{x^2 + \nu}$$

$$p_j = \frac{e^{-.5\lambda^2}}{j!}\left(\frac{\lambda^2}{2}\right)^j$$

$$q_j = \frac{\lambda e^{-.5\lambda^2}}{\sqrt{2}\Gamma(j+3/2)}\left(\frac{\lambda^2}{2}\right)^j$$

and where Φ is the standard normal CDF, Γ is the gamma function, and $I_z(a, b)$ is the regularized incomplete beta function.

Range $-\infty \le x < \infty$.
Shape parameter $\nu > 0$.
Non-centrality parameter $\lambda > 0$.

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.

Pareto Distribution

PDF

$$\frac{\alpha\, x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}}$$

CDF

$$1-\left(\frac{x_{\mathrm{m}}}{x}\right)^{\alpha}$$

Range $x_m < x < \infty$.
Location parameter $x_m > 0$, the minimum of x.
Shape parameter $\alpha > 0$.

Pearson Distribution

The Pearson family consists of eight distributions:
Type 0 Normal distribution.
Type 1 Beta distribution with minimum and maximum.
Type 2 Symmetric Beta distribution with minimum and maximum. The symmetry is imposed by requiring the two shape parameters to have the same value.
Type 3 Gamma distribution with threshold.
Type 4 Pearson IV distribution.
Type 5 Inverse Gamma distribution with threshold.
Type 6 F distribution with location and scale.
Type 7 Student's T distribution with location and scale
The Pearson methodology selects one member of the family using the methodology described in Johnson, Kotz and Balakrishnan.

Pearson Type IV Distribution

PDF

$$\frac{k(\nu,\omega)}{\lambda}\left(\frac{1}{1+z^2}\right)^{\omega} e^{-\nu\, atan(z)}$$

CDF

$$\int_{-\infty}^{z}\frac{k(\nu,\omega)}{\lambda}\left(\frac{1}{1+t^2}\right)^{\omega} e^{-\nu\, atan(t)}\, dt$$

where

$$b = 2(\omega-1)$$
$$a = \sqrt{\frac{b^3-b^2}{b^2+\nu^2}}$$
$$z = \frac{x-\mu}{a\lambda} - \frac{\nu}{b}$$

and $k(\nu,\omega)$ is a scaling constant.

Range $0 \le x < \infty$.
Location parameter μ
Scale parameter, $\lambda > 0$.
Shape parameter ν.
Shape parameter $\omega > 1.5$.

PERT 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.

Range $lb \le z \le ub$.
Parameter lb: $lb < z_{min}$, the lower bound of z.
Parameter ub: $z_{max} < ub$, the upper bound of z.
Parameter η: $lb < \eta < ub$, the mode of z.

Notes: The PERT argument z and the three parameters, lb, ub and η are transformed using the PERT transform; the resulting argument, x is distributed Beta, with shape parameters α and β.

Power Distribution

PDF

$$\frac{\nu(x-a)^{\nu-1}}{(b-a)^{\nu}}$$

CDF

$$\frac{(x-a)^{\nu}}{(b-a)^{\nu}}$$

Range $a \leq x \leq b$.
Lowerbound parameter $a > 0$.
Upperbound parameter b.
Shape parameter $\nu > 0$.

Rayleigh Distribution

PDF

$$\frac{x}{\alpha^2} e^{-x^2/(2\alpha^2)}$$

CDF

$$1 - e^{-x^2/(2\alpha^2)}$$

Range $0 \leq x < \infty$.
Scale parameter, $\alpha > 0$.

Reciprocal Distribution

PDF

$$\frac{1}{x(\ln(b) - \ln(a))}$$

CDF

$$\frac{\ln(x) - \ln(a)}{\ln(b) - \ln(a)}$$

Range $a \leq x \leq b$.
Lowerbound parameter $a > 0$.
Upperbound parameter b.

Skew Normal Distribution

PDF

$$\frac{2}{\sigma}\Phi\left(\frac{\alpha(x-\mu)}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right)$$

CDF

$$2\ \mathrm{cdfbvn}\left(\frac{x-\mu}{\sigma},0,\frac{-\alpha}{\sqrt{1+\alpha^2}}\right)$$

where $\phi(x)$ and $\Phi(x)$ are respectively the PDF and CDF of the standard normal distribution, and cdfbvn is the cumulative standardized bivariate normal distribution.

Range $-\infty < x < \infty$.
Location parameter, μ.
Scale parameter $\sigma > 0$.
Skew parameter α, negative for left skew, positive for right skew.

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 $\sigma > 0$.

Student's T Distribution

PDF

$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$$

CDF

$$\frac{1}{2}+x\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}$$

where ${}_2F_1$ is the hypergeometric function.

Range $-\infty < x < \infty$.
Shape parameter $\nu > 0$, degrees of freedom.

T3: Student's T Distribution with Location and Scale

PDF

$$\frac{1}{\alpha\sqrt{\pi\,\nu}}\frac{\Gamma(.5(\nu+1))}{\Gamma(.5\nu)}\left(\frac{\nu\alpha^2}{(x-\mu)^2+\nu\alpha^2}\right)^{.5(\nu+1)}$$

CDF

$$.5+.5\,I_z(.5,.5\nu) \quad x\geq 0$$

$$.5-.5\,I_z(.5,.5\nu) \quad x<0$$

where

$$z=\frac{(x-\mu)^2}{(x-\mu)^2+\nu\alpha^2}$$

and where Γ is the gamma function, and $I_z(a,b)$ is the regularized incomplete beta function.

Range $-\infty \leq x < \infty$.
Location parameter μ.
Scale parameter $\alpha > 0$.
Shape parameter $\nu > 0$.

Triangular Distribution

PDF

$$\frac{2(x-a)}{(b-a)(c-a)} \quad \text{for } a\leq x\leq c$$

$$\frac{2(b-x)}{(b-a)(b-c)} \quad \text{for } c\leq x\leq b$$

CDF

$$\frac{(x-a)^2}{(b-a)(c-a)} \quad \text{for } a\leq x\leq c$$

$$1-\frac{(b-x)^2}{(b-a)(b-c)} \quad \text{for } c\leq x\leq b$$

Range $a \leq x \leq b$.
Parameter a: $a \leq x_{min}$, the lower bound of x.
Parameter b: $x_{max} \leq b$, the upper bound of x.
Parameter c: $a < c < b$, the mode of x.

Uniform Distribution

PDF

$$\frac{1}{b-a}$$

CDF

$$\frac{x-a}{b-a}$$

Range $a \le x \le b$.
Parameter a: $a \le x_{min}$, the lower bound of x.
Parameter b: $x_{max} \le b$, the upper bound of x.

Von Mises Distribution

PDF

$$\frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

CDF

$$\frac{1}{2\pi}\left(x + \frac{2}{I_0(\kappa)}\sum_{j=1}^{\infty} I_j(\kappa)\frac{\sin[j(x-\mu)]}{j}\right)$$

where $I_j(x)$ is the modified Bessel function of order j.

Range $0 \le x < 2\pi$.
Location parameter, μ: $0 \le \mu \le 2\pi$.
Shape parameter $\kappa > 0$.

Weibull Distribution

PDF

$$\frac{\beta x^{\beta-1}}{\alpha^{\beta}} e^{-(x/\alpha)^{\beta}}$$

CDF

$$1 - e^{-(x/\alpha)^{\beta}}$$

Range $0 \le x < \infty$.
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 $0 \leq x - \theta < \infty$.
Scale parameter, $\alpha > 0$.
Shape parameter $\beta > 0$.
Threshold parameter $\theta < min(x)$.

Formulas for Discrete Distributions

For x to be distributed by a discrete distribution, x must only take discrete integer values over the specified range. The formulas provided are the Probability Density Function (PDF) and Cumulative Distribution Function (CDF).

Bernoulli Distribution

The Bernoulli distribution takes a value 1 with probability p and value 0 with probability $1-p$.

PDF

$$xp + (1-x)(1-p)$$

CDF

$$1 - p + px$$

Support $x \in \{0,1\}$.
Probability parameter, p: $0 \leq p \leq 1$.

Binomial Distribution

The binomial pdf is the probability of x successes in n independent trials, where p is the probability of success in any given trial.

PDF

$$\binom{n}{x} p^x (1-p)^{n-x}$$

CDF

$$I_{1-p}(n-x, x+1)$$

where $I_x(a,b)$ is the regularized incomplete Beta function.

Support $x \in \{0, \dots, n\}$.
Probability parameter, p: $0 \leq p \leq 1$.
Trials parameter, n: $n > 0$.

Custom Discrete Distribution

PDF

$$p_i \quad x = x_i$$

CDF

$$\sum_{j=1}^{i} (p_j) \quad x = x_i$$

where $p_i = w_i / \sum_{j=1}^{N} (w_j)$.

N data values x.
Probability weights $w > 0$.

Geometric Distribution

The geometric pdf is the probability of x failures before a success, where p is the probability of success in any given trial.

PDF

$$p(1-p)^x$$

CDF

$$1-(1-p)^{x+1}$$

Support $x \in \{1,2,3, \dots\}$.
Probability parameter, p: $0 \leq p \leq 1$.

Hypergeometric Distribution

The Hypergeometric pdf is the probability of drawing x successes in n draws, without replacement, from a population of size N which contains m successes.

PDF

$$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

CDF

$$\sum_{i=0}^{x}\left(\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}\right)$$

Support $x \in \{1,2,3, \dots\}$.
Population parameter, N: $N > 0$.
Success parameter, m: $m > 0$.
Sample parameter, n: $n > 0$.

Logarithmic Distribution

The Logarithmic pdf is a one parameter generalized power series distribution.

PDF

$$\frac{-p^x}{x\ln(1-p)}$$

CDF

$$1+\frac{B(p;x+1,0)}{\ln(1-p)}$$

where B(x; a,b) is the incomplete beta function.

Support $x \in \{1,2,3,\ldots\}$.
Probability parameter, p: $0 \le p \le 1$.

Negative Binomial Distribution

The Negative Binomial pdf is the probability of achieving r failures before the xth success, with p being the probability of a success.

PDF

$$\binom{x+r-1}{r-1}(1-p)^r p^x$$

CDF

$$1-I_p(x+1,r)$$

where $I_x(a,b)$ is the regularized incomplete Beta function.

Support $x \in \{1,2,3,\ldots\}$.
Probability parameter, p: $0 \le p \le 1$.
Failure parameter, r: $r > 0$.

Poisson Distribution

The Poisson pdf is the probability of x events occurring within a period, where λ is the expected number of events in that period.

PDF

$$\frac{\lambda^x}{x!}\,e^{-\lambda}$$

CDF

$$e^{-\lambda}\sum_{i=0}^{x}\left(\frac{\lambda^i}{i!}\right)$$

Support $x \in \{1,2,3,\dots\}$.
Event parameter, λ: $r > 0$, the mean of x.

Step Distribution

The step pdf is the same for each step.

PDF

$$\frac{s}{b-a+s}$$

CDF

$$\frac{x-a+s}{b-a+s}$$

Support $x \in \{a,\dots,b\}$.
Parameter a: $a \leq x_{min}$, the lower bound of x.
Parameter b: $x_{max} \leq b$, the upper bound of x.
Parameter s, the stepsize.

Uniform Distribution

The Uniform pdf is the same for each outcome.

PDF

$$\frac{1}{b-a+1}$$

CDF

$$\frac{x-a+1}{b-a+1}$$

Support $x \in \{a, \dots, b\}$.
Parameter a: $a \leq x_{min}$, the lower bound of x.
Parameter b: $x_{max} \leq b$, the upper bound of x.

Functions Used by Distributions

Beta Function

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Gamma Function

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

Incomplete Beta Function

$$B(x;\ a, b) = \int_0^x t^{a-1} \, (1 - t)^{b-1} \, dt$$

Incomplete Beta Function (regularized)

$$I_x(a, b) = \frac{B(x;\ a, b)}{B(a, b)}$$

Incomplete Gamma Function (lower)

$$\gamma(s, x) = \int_0^x t^{s-1} \, e^{-t} \, dt$$

Incomplete Gamma Function (regularized)

$$P(s, x) = \frac{\gamma(s, x)}{\Gamma(s)}$$

Modified Bessel Function

$$I_\alpha(x) = \sum_{m=0}^{\infty} \frac{1}{m! \, \Gamma(m + \alpha + 1)} \left(\frac{x}{2}\right)^{2m+\alpha}$$

References for Distributions

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.

Chou,Y., A.M. Polansky, and R.L. Mason (1998). “Transforming Non-Normal Data to Normality in Statistical Process Control,” *Journal of Quality Technology*, Vol. 30(2), pp. 133-141.

D'Agostino, R.B. and M.A. Stephens (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

David, H.A. (1981). *Order Statistics*, John Wiley & Sons, New York.

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.

Greene, W.H. *Econometric Analysis* 4th Ed Prentice Hall, New Jersey.

Johnson, N. L., and S. Kotz (1990). “Use of moments in deriving distributions and some characterizations,” *Mathematical Scientist*, Vol. 15, pp. 42-52.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1994) *Continuous Univariate Distributions*, Vol. 1, Wiley-Interscience.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1995) *Continuous Univariate Distributions*, Vol. 2, Wiley-Interscience.

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.

Sleeper, A. (2006). *Six Sigma Distribution Modeling*, McGraw-Hill, New York.

Tadikamalla,P.,R. (1980). “Notes and Comments: On Simulating Non-Normal Distributions,” *Psychometrika*, Vol. 45(2), pp. 273-279.

Wang, J.Z. (2005). “A note on Estimation in the Four Paramter Beta Distribution,” *Comm in Stats Simulation and computation*, Vol. 34 pp. 495-501.

Sigma XL
Simply Smart Software

About SigmaXL, Inc.

Established in 1998, SigmaXL Inc. is a leading provider of user friendly Excel Add-ins for Lean Six Sigma graphical and statistical tools and Monte Carlo simulation.

Our flagship product, SigmaXL®, was designed from the ground up to be a cost-effective, powerful, but easy to use tool that enables users to measure, analyze, improve and control their service, transactional, and manufacturing processes. As an add-in to the already familiar Microsoft Excel, SigmaXL® is ideal for Lean Six Sigma training and application, or use in a college statistics course.

DiscoverSim™ enables you to quantify your risk through Monte Carlo simulation and minimize your risk with global optimization. Business decisions are often based on assumptions with a single point value estimate or an average, resulting in unexpected outcomes. DiscoverSim™ allows you to model the uncertainty in your inputs so that you know what to expect in your outputs.

Contact Information:
Technical Support: 1-866-475-2124 (Toll Free in North America) or
1-519-579-5877
Sales: 1-888-SigmaXL (888-744-6295)
E-mail: support@SigmaXL.com Web: www.SigmaXL.com