Monte Carlo Simulation (General Simulation Models)

statgraphics 18® centurion

Revised: 10/11/2017

Summary

Monte Carlo simulation is used to estimate the distribution of variables when it is impossible or impractical to determine that distribution theoretically. It is used in many areas, including engineering, finance, and DFSS (Design for Six Sigma). A typical Monte Carlo simulation includes:

1. One or more input variables X, some of which usually follow a probability distribution.
2. One or more output variables Y, whose distribution is desired.
3. A mathematical model coupling the X's and the Y's.

This document considers some examples.

Example #1

The first example comes from the book titled Design for Six Sigma in Technology and Product Development by Clyde M. Creveling, Jeff Slutsky, and Dave Antis (Prentice Hall, 2002). They describe the use of a Monte Carlo simulation to estimate the distribution of the total time required to complete the development of a DFSS Phase 1 concept. Development of the concept involves 12 tasks, each of which has an uncertain duration. In their example, there are 12 inputs:

X_i = duration of task i in days for i = 1, 2, …, 12.

The output variable is

Y = total time required to complete all 12 tasks.

The model linking the input and outputs is simply

$$Y = X_1 + X_2 + \ldots + X_{12} \quad (1)$$

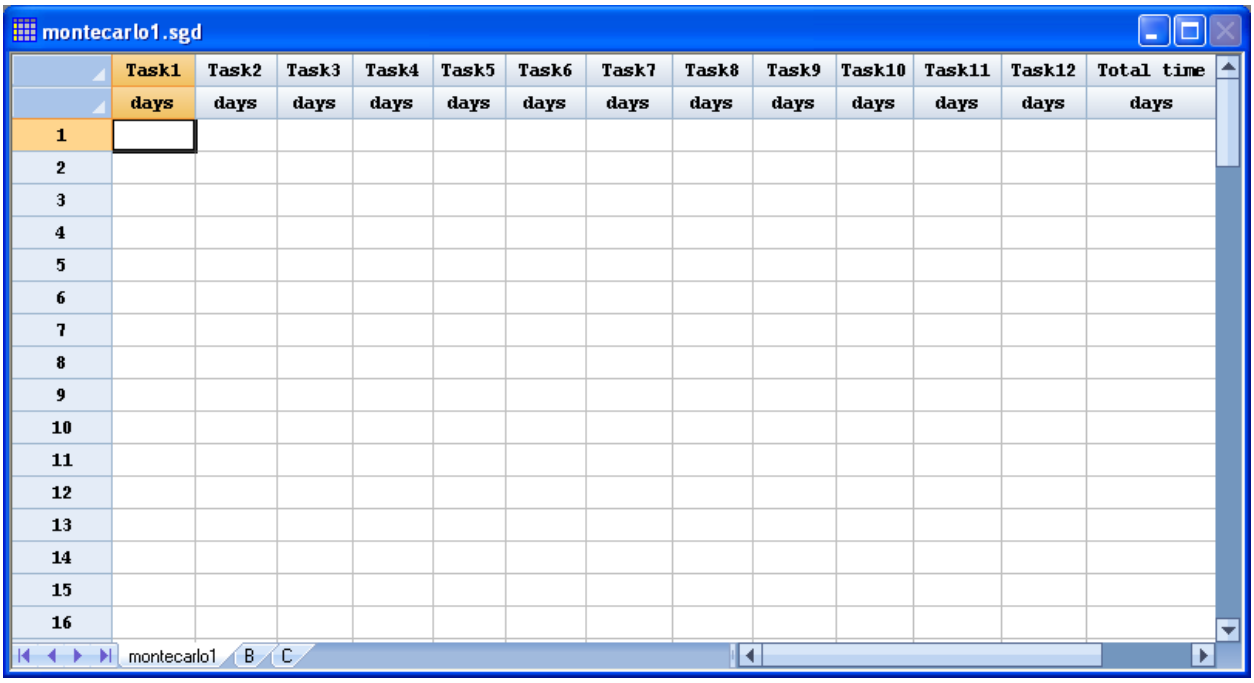
Each of the input variables is assumed to follow a triangular distribution with the following parameters:

Task	Lower limit	Median	Upper limit
1	7	10	13
2	1.5	2.83	4
3	4	5	6
4	7	8.33	10
5	12	14	16
6	1	1.83	2.5
7	1	1.67	2
8	20	25	30
9	4	5.67	7
10	2	3	4
11	54	60	66
12	18	20	22

While the mean and variance of the total time *Y* could be determined theoretically, they were also interested in determining percentiles of the distribution from which specification limits for product development cycle times could be established.

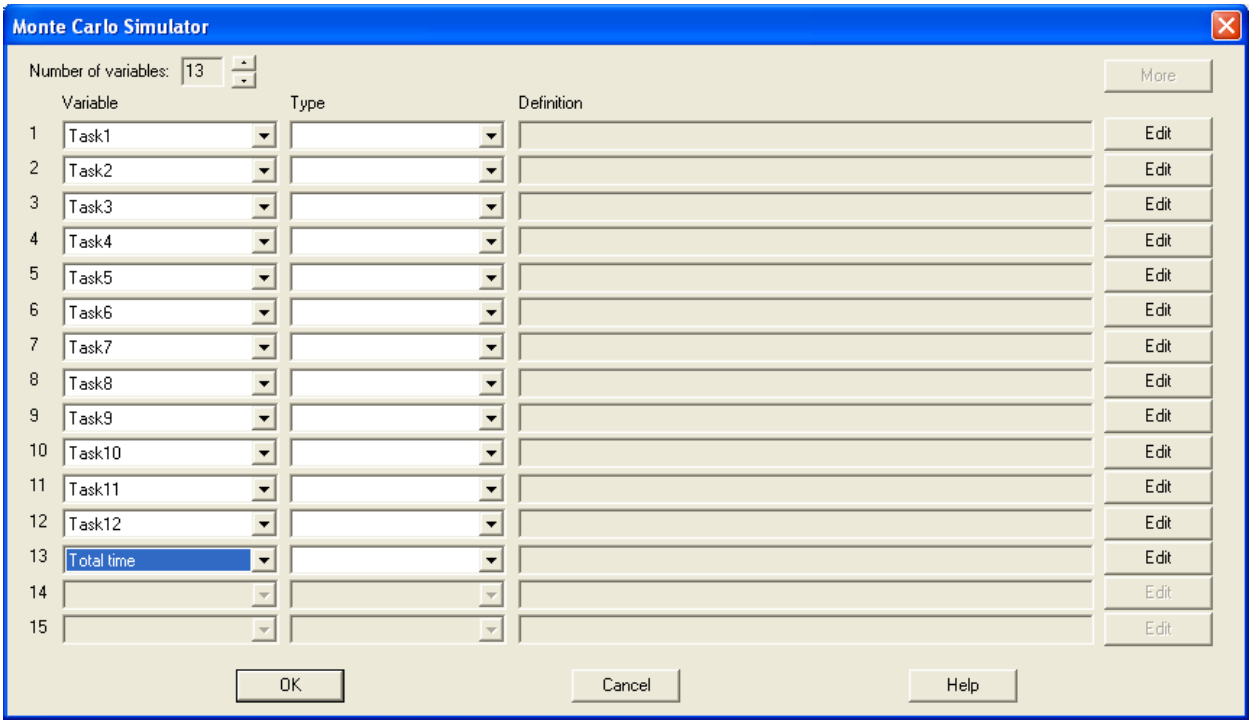
To build a simulation model for this problem, the following steps are required:

Step 1: Create a datasheet with columns for each of the input and output variables. In this case, that requires a datasheet with the following 13 numeric columns:



This sheet has been saved as *montecarlo1.sgd*.

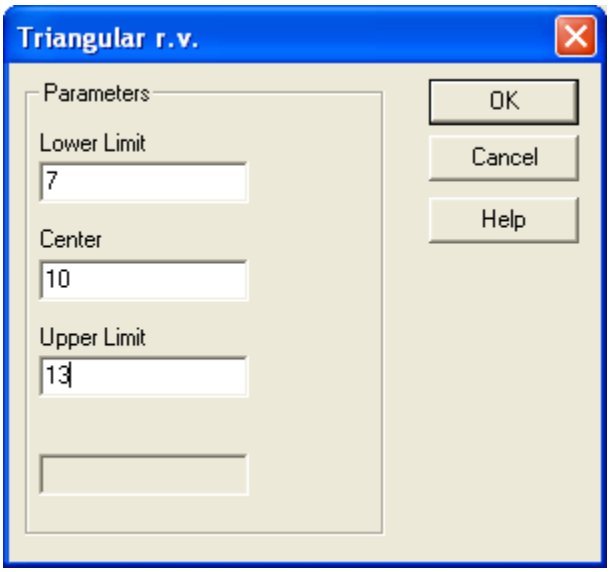
Step 2: Specify the input and output variables. This is done by selecting *Monte Carlo Simulation – General Simulation Model* from the *Tools* menu. When a new analysis window is created, press the leftmost button on the analysis toolbar (the *Input dialog* button) to display the dialog box shown below:



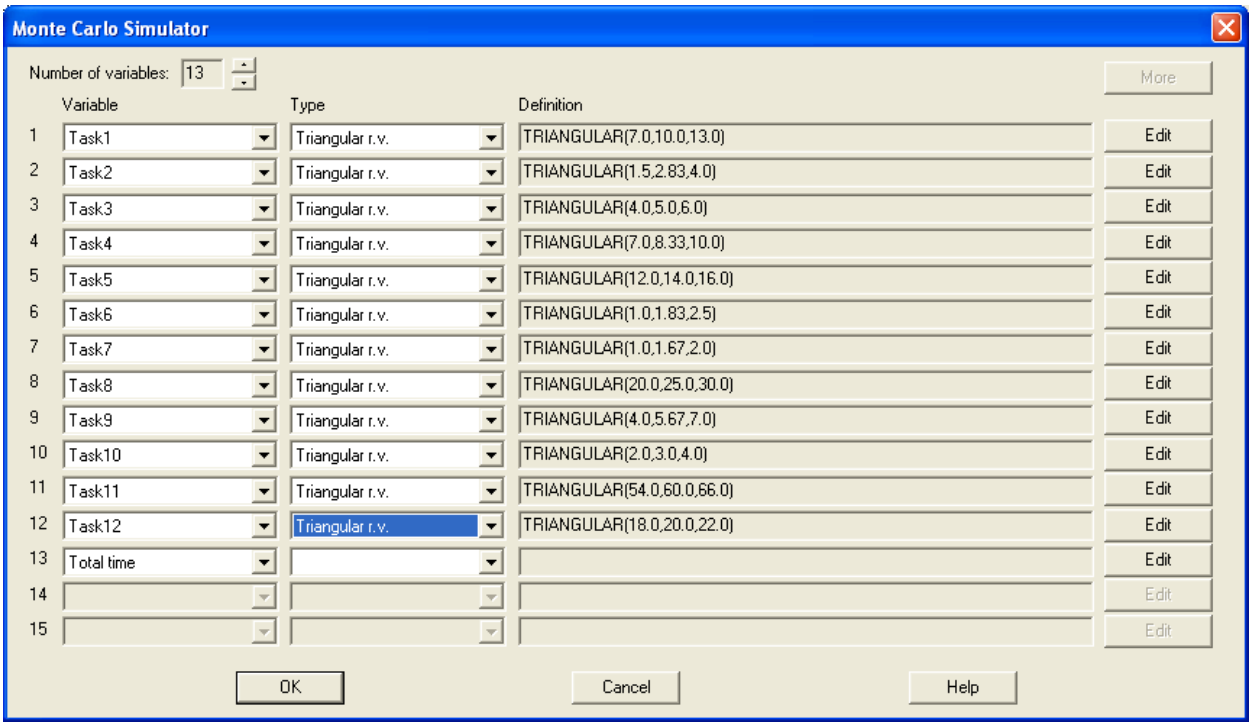
First set the number of variables to 13. Then use the pulldown variable lists to select each input X variable and then the output Y variables.

NOTE: When the simulation is performed, the variables will be created from the top down. If any variables depend on others, be sure that the dependent variables are listed below the variables that they depend on.

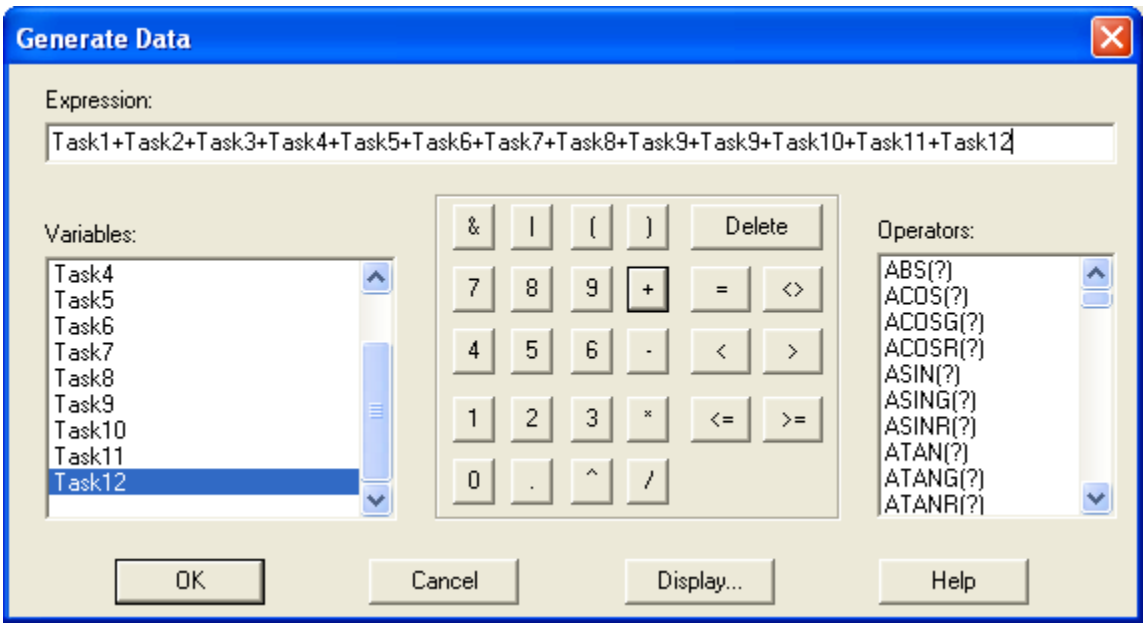
Step 3: For each input variable, specify its probability distribution. This is done by selecting a distribution from the *Type* pulldown list. In this case, the length of each input task is assumed to be a triangular random variable. When *Triangular r.v.* is selected, a dialog box will be displayed on which to enter the parameters of the selected distribution:



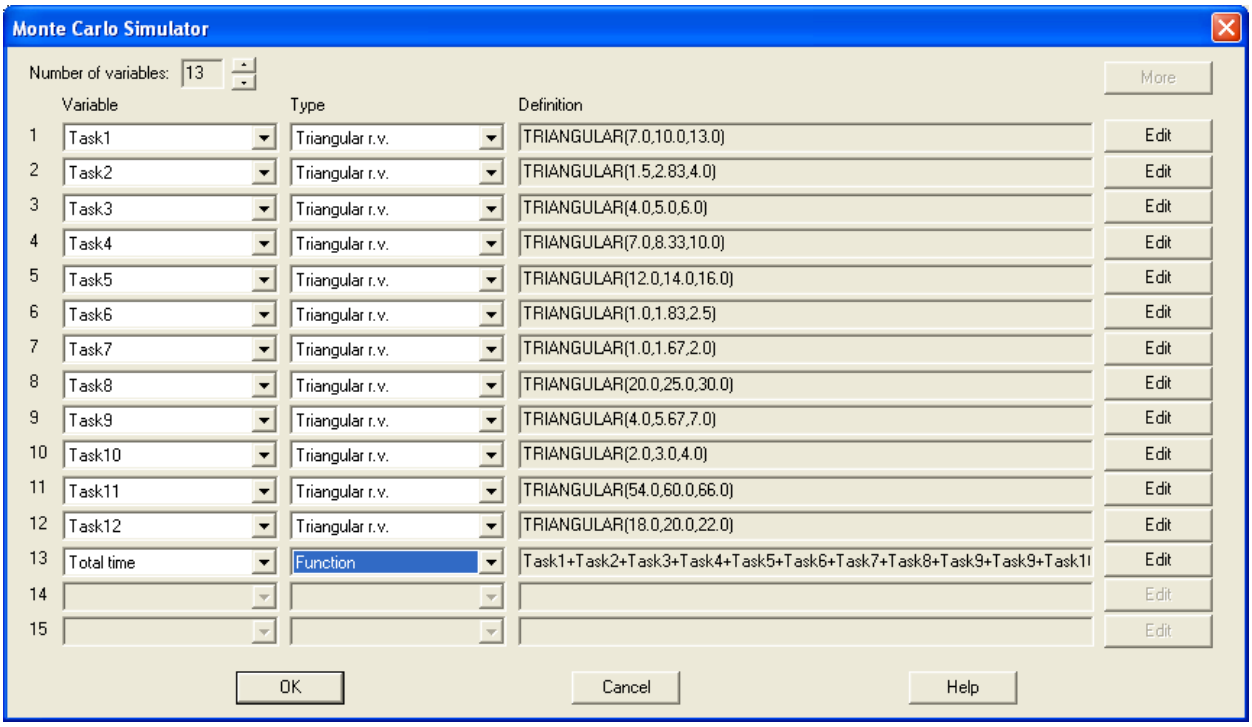
After all distributions are specified, the main dialog box will appear as shown below:



Step 4: For each output variable, specify its dependence on the input variables. This is down by selecting *Function* from the *Type* pulldown list, which displays the following dialog box:



Any valid STATGRAPHICS expression can be used. After pressing OK, the input dialog box will appear as shown below:



Use the *Edit* buttons to correct any problems. Then press *OK*. The *Monte Carlo Simulation* analysis window will summarize the variables you have entered:

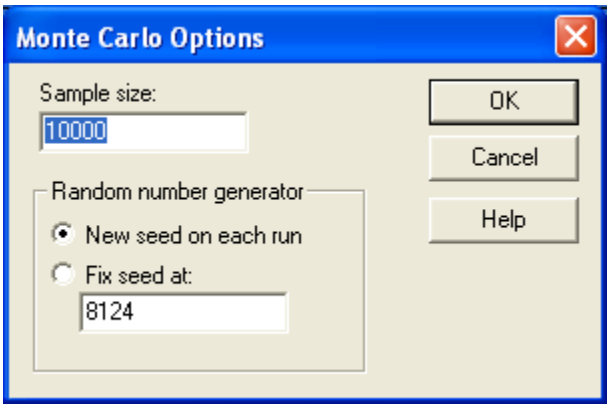
Monte Carlo Simulation

Sample size: 10000

Seed for random number generator: 8124

Variable	*Type*	*Definition*
Task1	Triangular r.v.	TRIANGULAR(7.0,10.0,13.0)
Task2	Triangular r.v.	TRIANGULAR(1.5,2.83,4.0)
Task3	Triangular r.v.	TRIANGULAR(4.0,5.0,6.0)
Task4	Triangular r.v.	TRIANGULAR(7.0,8.33,10.0)
Task5	Triangular r.v.	TRIANGULAR(12.0,14.0,16.0)
Task6	Triangular r.v.	TRIANGULAR(1.0,1.83,2.5)
Task7	Triangular r.v.	TRIANGULAR(1.0,1.67,2.0)
Task8	Triangular r.v.	TRIANGULAR(20.0,25.0,30.0)
Task9	Triangular r.v.	TRIANGULAR(4.0,5.67,7.0)
Task10	Triangular r.v.	TRIANGULAR(2.0,3.0,4.0)
Task11	Triangular r.v.	TRIANGULAR(54.0,60.0,66.0)
Task12	Triangular r.v.	TRIANGULAR(18.0,20.0,22.0)
Total time	Function	Task1+Task2+Task3+Task4+Task5+Task6+Task7+Task8+Task9+Task9+Task10+Task11+Task12

Step 5: To specify the parameters of the simulation, press the *Analysis Options* button on the analysis toolbar to display the following dialog box:



Sample size: the number of values to be generated for each variable.

Random number generator: whether to generate a new seed for the random number generator when the simulation is run, or if the seed displayed on the dialog box should be used. If you use a fixed seed, you can replicate the results of a previous simulation.

Step 6: Press *OK* to run the simulation. The specified number of observations will be generated for each variable, working from the top down. The results will then be added to the datasheet, as shown below:

montecarlo1.sgd

	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9	Task10	Task11	Task12	Total time
	days	days	days	days	days	days	days	days	days	days	days	days	days
1	8.22415	2.56458	5.26572	9.3078	13.6678	2.02984	1.57681	27.4244	5.74361	3.51548	64.1278	21.4864	170.678
2	8.0048	3.03756	4.95496	7.66279	15.6069	2.16825	1.48409	26.8182	5.48043	3.49388	57.5951	19.8508	161.638
3	9.73434	2.81759	5.1654	8.96058	14.0313	1.24248	1.51932	27.7113	6.38841	2.39524	60.1727	19.362	165.889
4	10.7337	3.13045	5.8092	8.63797	14.737	1.84546	1.4885	21.1043	5.35246	2.45872	60.8351	20.9017	162.387
5	10.5851	2.03841	4.54132	7.65202	13.1117	1.59874	1.05218	21.5811	4.98032	3.28561	58.3515	20.79	154.548
6	8.95031	2.57584	5.46832	7.48576	15.185	1.57493	1.60145	22.139	5.33315	3.51161	64.9032	20.1598	164.222
7	10.29	3.02325	5.2068	7.83116	13.8456	1.53094	1.2104	26.2665	6.00974	3.55993	61.6366	18.3662	164.787
8	10.8179	2.33489	4.82969	9.38313	13.5726	1.55335	1.68399	28.3144	4.91046	3.0636	58.524	18.3987	162.297
9	9.7285	2.74608	4.95642	8.2942	14.1132	2.3373	1.64475	26.7524	5.77288	3.51496	56.5799	19.621	161.834
10	8.3666	2.55553	5.01427	9.23444	12.6811	1.80826	1.85451	22.7482	5.49876	2.04385	60.626	20.2363	158.167
11	11.3204	3.0456	4.96631	8.77245	14.1937	1.77595	1.20339	24.7641	6.47886	3.21541	62.6669	19.7189	168.601
12	8.90391	3.49247	5.42954	8.17106	14.4065	1.85181	1.32524	26.6919	6.40571	3.50292	61.5038	19.2574	167.348
13	12.294	3.53811	5.354	8.30654	15.3901	1.88959	1.23897	24.8428	4.1581	2.38982	59.4081	20.4837	163.452
14	10.3328	2.87105	5.28159	8.60503	15.349	1.82323	1.52891	23.1482	5.54758	2.6874	59.9651	19.9929	162.68
15	9.4198	1.78929	5.43197	8.16325	12.7243	1.52321	1.65803	24.5527	5.51414	3.41914	58.1734	20.8005	158.684
16	10.7549	2.09606	5.17716	8.04735	14.1074	1.37638	1.38285	25.8503	5.31368	3.74121	59.5388	19.4305	162.13

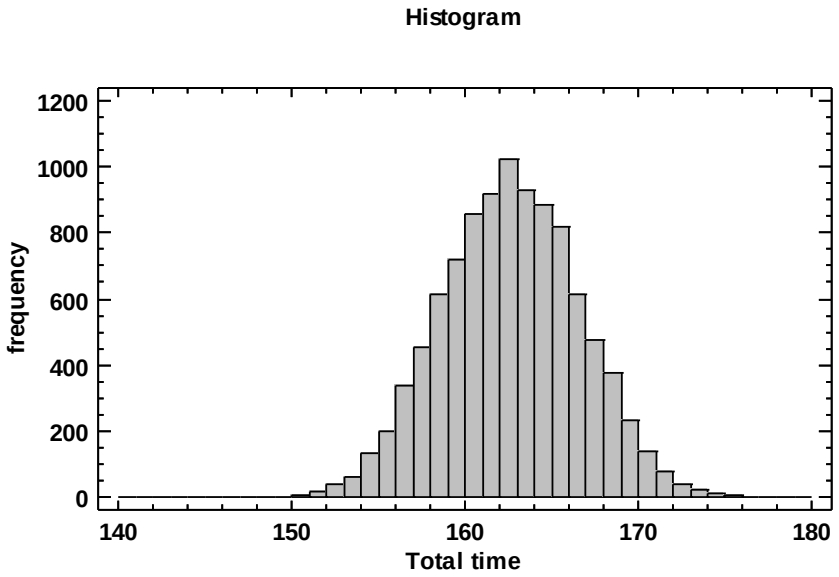
montecarlo1 | B | C

Step 7: To display statistics for the generated data, press the *Tables and Graphs* button on the analysis toolbar to display any of the following:

1. Summary Statistics
2. Confidence Intervals
3. Percentiles
4. Frequency Tabulation

5. Frequency Histogram
6. Box-and-Whisker Plot
7. Quantile Plot
8. Sensitivity Tornado Plot

Each of these tables and graphs is described in the pdf document titled *One Variable Analysis*. For example, the following histogram shows the generated values for *Total Time*:



The *Percentiles* table is also shown below:

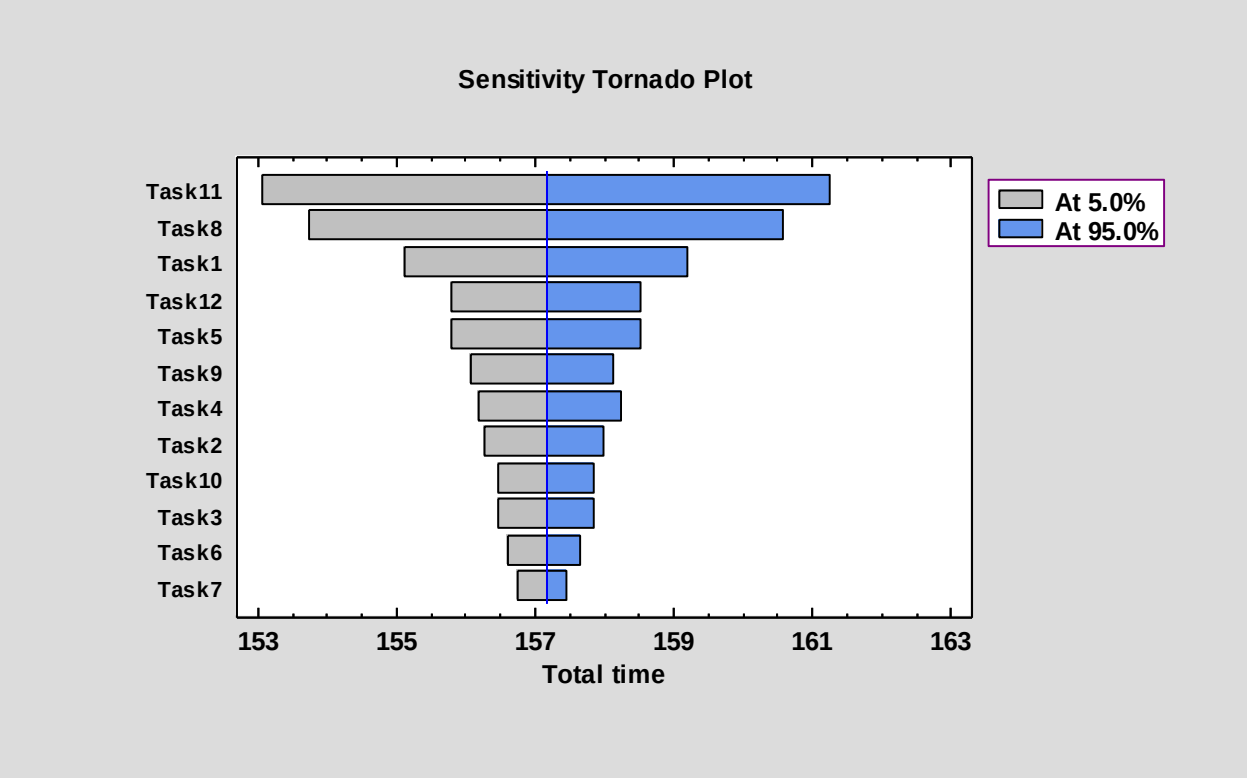
Percentiles

Percentage	*Task1*	*Task2*	*Task3*	*Task4*	*Task5*	*Task6*	*Task7*	*Task8*	*Task9*	*Task10*	*Task11*
0.5%	7.27797	1.65071	4.08967	7.14401	12.1977	1.07348	1.05868	20.4875	4.15759	2.1003	54.5876
5.0%	7.94879	1.9206	4.30941	7.44423	12.6303	1.245	1.18018	21.5862	4.50821	2.31955	55.8663
10.0%	8.34462	2.09321	4.44226	7.62505	12.8887	1.34992	1.25874	22.2251	4.72264	2.44623	56.6326
25.0%	9.11429	2.41901	4.70768	7.99624	13.4096	1.55267	1.40716	23.5412	5.12175	2.7088	58.2207
50.0%	9.98932	2.79413	4.99743	8.41716	13.995	1.7876	1.57634	24.9819	5.57716	3.00201	60.0133
75.0%	10.8946	3.15733	5.29213	8.8652	14.5808	2.00214	1.71326	26.4357	6.00805	3.29392	61.7501
90.0%	11.6667	3.47345	5.54721	9.28147	15.1054	2.18686	1.82158	27.7268	6.37606	3.55461	63.313
95.0%	12.0824	3.63313	5.68803	9.50463	15.3639	2.27822	1.87457	28.3706	6.55702	3.67682	64.0846
99.5%	12.7253	3.89126	5.89146	9.83956	15.8101	2.43375	1.95894	29.4661	6.85088	3.90243	65.3877

Percentage	*Task12*	*Total time*
0.5%	18.2041	152.776
5.0%	18.6331	156.154
10.0%	18.8843	157.541
25.0%	19.4143	159.897
50.0%	20.0009	162.623
75.0%	20.5791	165.368
90.0%	21.108	167.766
95.0%	21.3668	169.101
99.5%	21.7935	172.655

Note that 99% of all total times were between 152.776 days and 172.655 days. These values could be used to set the desired specification limits.

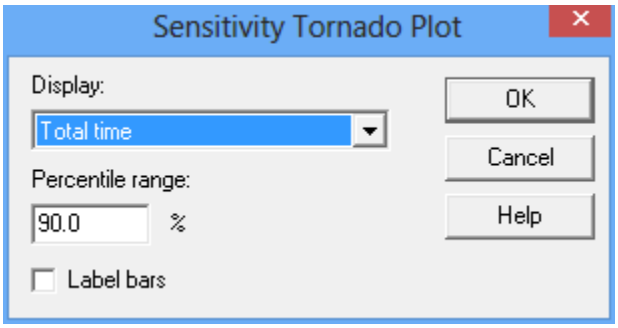
Of particular interest is the *Sensitivity Tornado Plot*, shown below:



This plot illustrates the effect of each input variable, sorted from top to bottom in decreasing order of importance. To judge the importance of each variable, each variable is set equal to its median value and the value of *Total time* is calculated. This value forms the baseline of the plot and its location is shown with a vertical line. Each input variable is then set equal to a lower and upper percentile spanning *p*% of its distribution, where *p* equals a value such as 90. The value of *Total time* is calculated at those 2 percentiles, with all other input variables held at their median values. Bars are plotted showing the effect on *Total time* by changing that input variable. The bars are then sorted according to the difference between the response at the upper and lower percentiles.

In the above plot, it can be seen that Task #11 has the greatest effect on *Total time*, followed by Task #8. Task #7 has the smallest effect.

Pane Options



Display: select a response variable to display on the plot.

Percentile range: the percentage covered by the difference between the lower and upper percentiles.

Label bars: display the values next to each bar.

Example #2

The second example deals with sales forecasting. Suppose that we wish to predict the monthly profit from selling a software product. Let:

S = number of units sold during a month

P = average price per unit

C = cost per unit (cost of goods only)

E = other expenses for the month

We propose the following simple model for profit during a month:

profit = S * (P – C) – E

To predict the profits during a month, we need to make assumptions about the input variables:

(1) Assume that S follows a normal distribution with a mean of 2,000 and a standard deviation of 250.

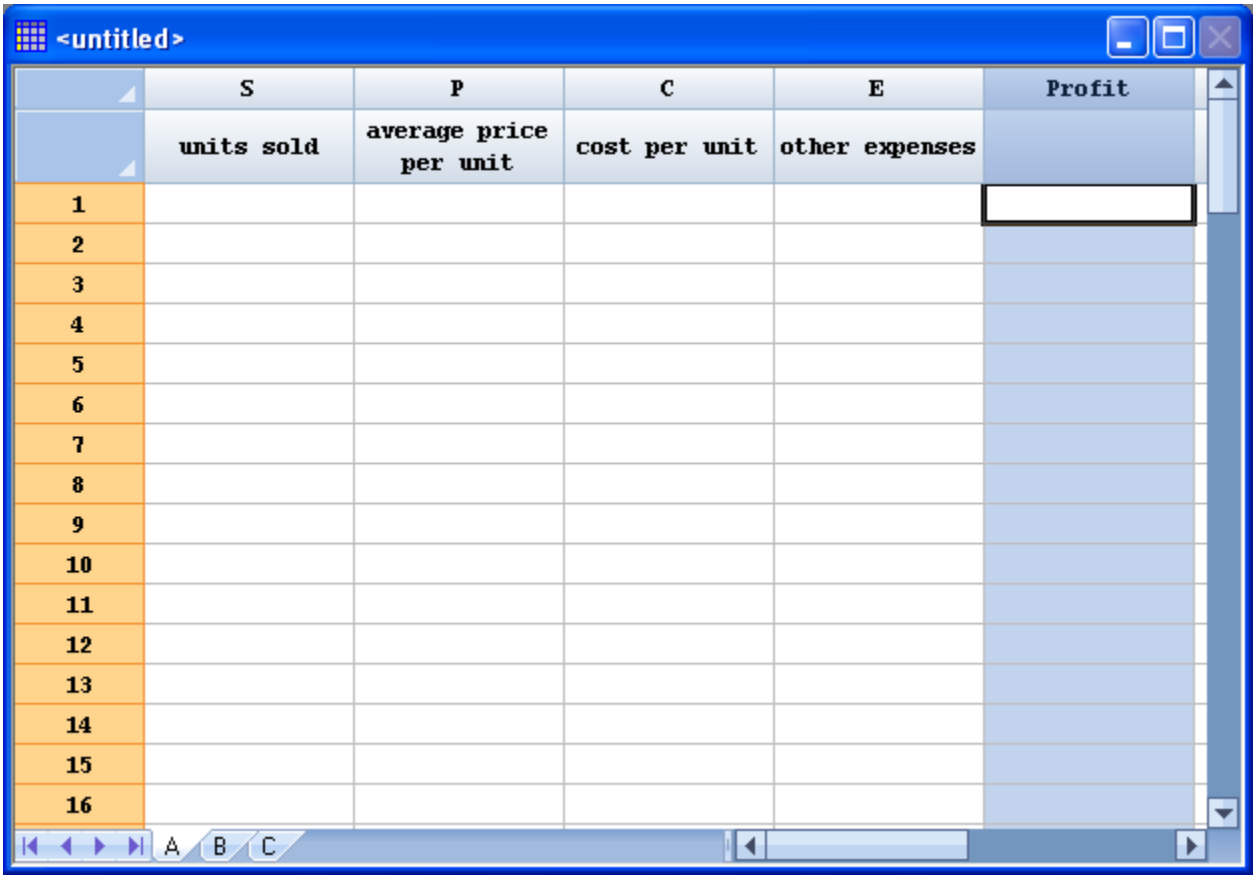
(2) Assume that P follows a normal distribution with a mean of $500 and a standard deviation of $10.

(3) Assume that E follows a Weibull distribution with a shape parameter equal to 2 and a scale parameter equal to 500,000.

(4) Assume that C is known to equal $30.

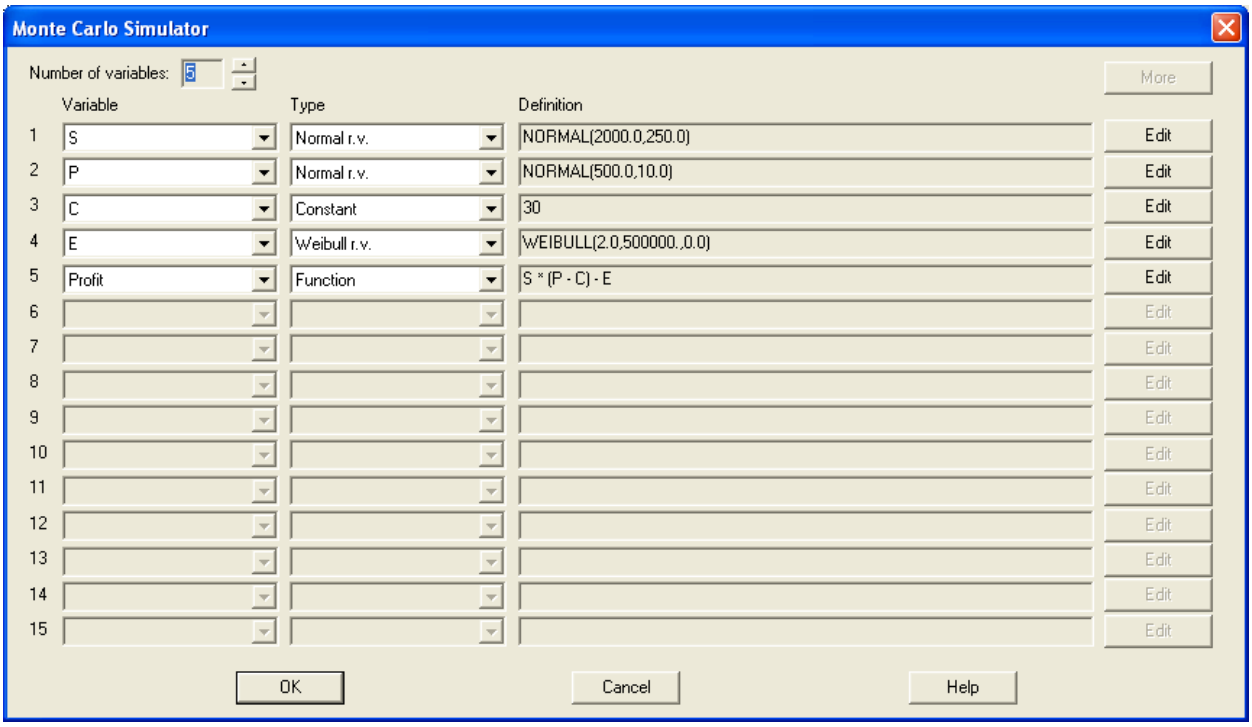
To build a simulation model for this problem, the following steps are required:

Step 1: Create a datasheet with columns for each of the input and output variables. In this case, that requires a datasheet with the following 5 numeric columns:



This sheet has been saved as *montecarlo2.sgd*.

Step 2: Specify the input and output variables by selecting *Monte Carlo Simulation – General Simulation Model* from the *Tools* menu. When a new analysis window is created, press the leftmost button on the analysis toolbar (the *Input dialog* button) to specify the names and types of the variables:



Note that S has been set to a constant value of 30.

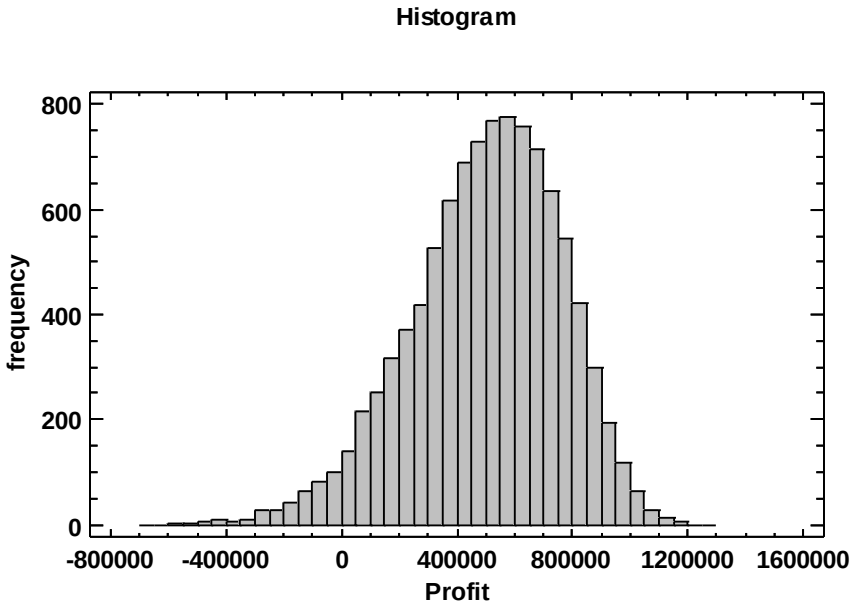
Step 3: To run the simulation, press the *Analysis Options* button on the analysis toolbar, set the desired parameters, and then press *OK*.

The summary statistics for the simulated results are shown below:

Summary Statistics

	S	*P*	*C*	*E*	*Profit*
Count	10000	10000	10000	10000	10000
Average	2004.26	500.049	30.0	441420.	500658.
Standard deviation	248.268	10.0147	0.0	230985.	258721.
Coeff. of variation	12.387%	2.00274%	0.0%	52.3278%	51.6762%
Minimum	979.188	459.339	30.0	6163.87	-556339.
Maximum	2936.64	548.351	30.0	1.45742E6	1.18733E6
Range	1957.45	89.012	0.0	1.45126E6	1.74367E6
Stnd. skewness	1.75782	-0.927263		25.3197	-18.8093
Stnd. kurtosis	0.409009	0.276884		3.6113	3.21139

A frequency histogram for monthly profit is shown below:



Notice that although the average profit is quite large, there is a chance of losing money in any given month. This is primarily due to the uncertainty about expenses, which can be displayed using *Pane Options* to change the histogram variable: