Monte Carlo Simulation (Random Number Generation)

statgraphics 18®
centurion

Revised: 10/11/2017

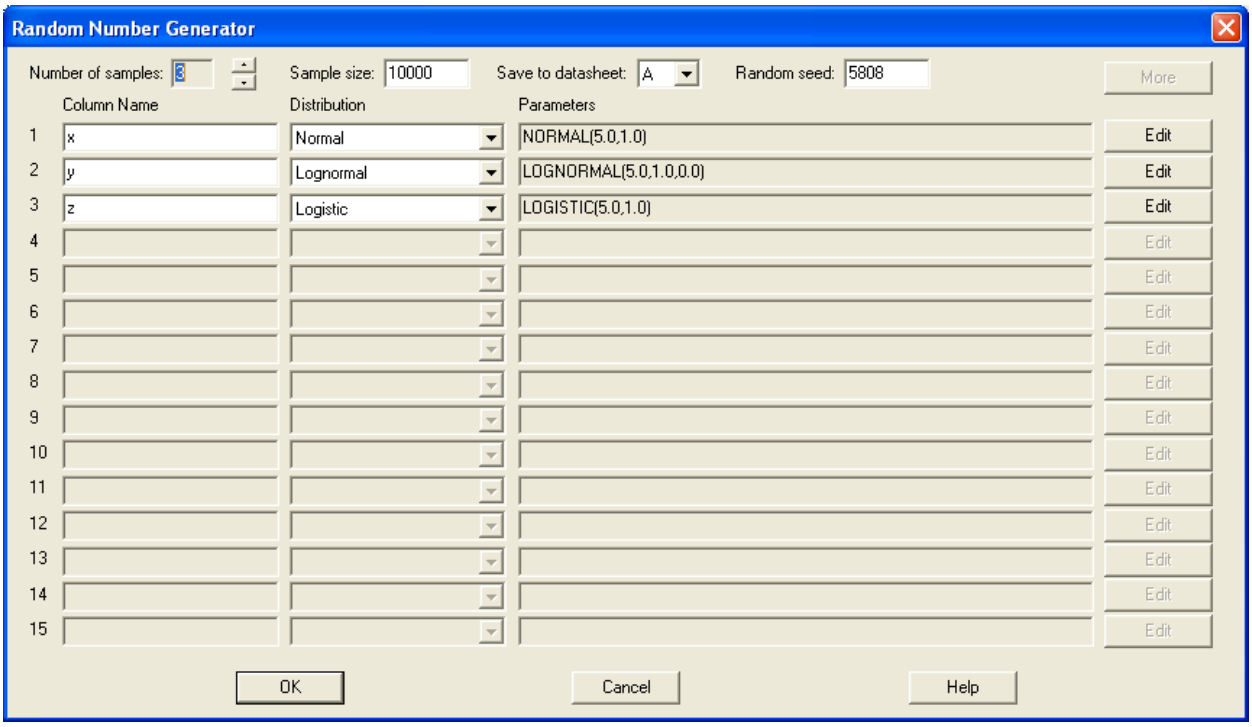
Summary 1
Data Input 1
Analysis Options 6
Summary Statistics 6
Box-and-Whisker Plots 7
Percentiles 9
Quantile Plots 10

Summary

It is frequently necessary to generate random numbers from different probability distributions. This procedure simplifies the process of creating multiple samples of random numbers.

Data Input

To create a sample of pseudo-random numbers, select *Monte Carlo Simulation –Random Number Generation* from the *Tools* menu. The following data input dialog box will be displayed:



- **Number of samples:** the number of samples that will be generated. Different distributions may be specified for each sample.

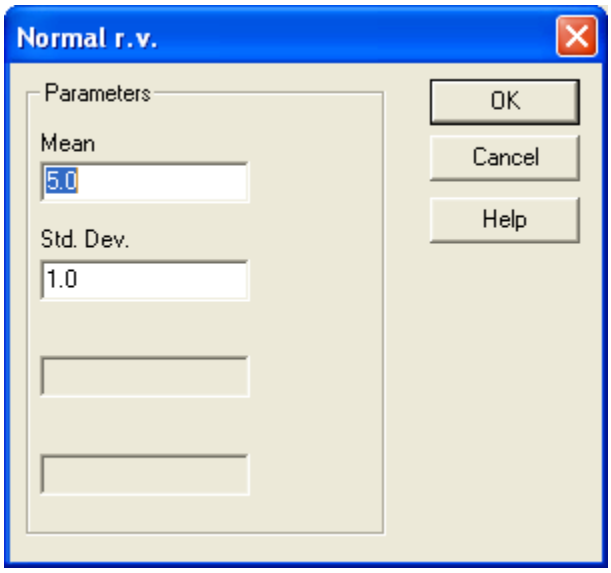
- **Sample size**: the number of values to be generated for each sample.

- **Save to datasheet**: the letter corresponding to the datasheet in which the values will be saved.

- **Random seed:** the seed for the random number generator. The initial default value is set based on the time of day. If you use the same seed more than once, you will get the same results.

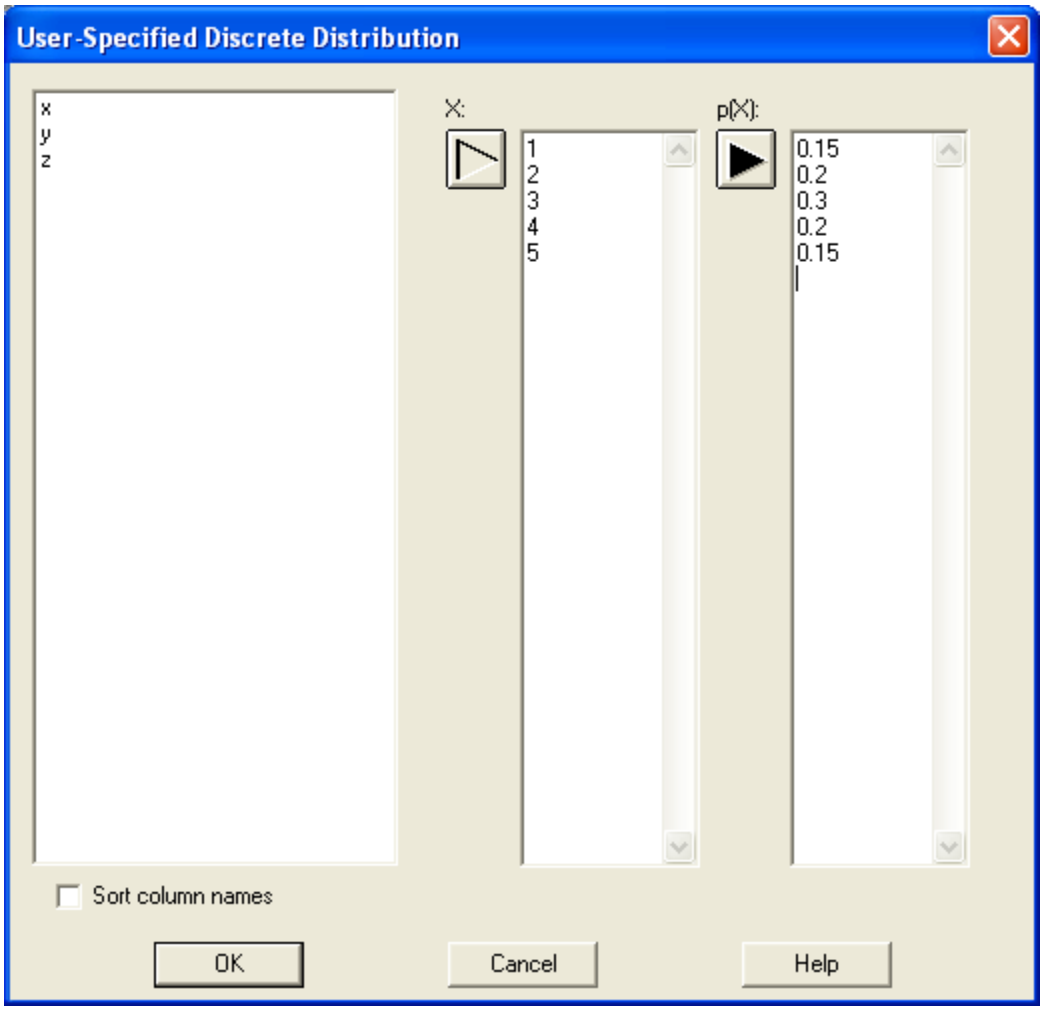
- **Column Name**: the names of the columns in which the random values will be saved.

- **Distribution**: the type of probability distribution from which the values will be generated. If you select a predefined distribution such as the normal distribution, a dialog box will be displayed on which to enter the parameters of the selected distribution such as:



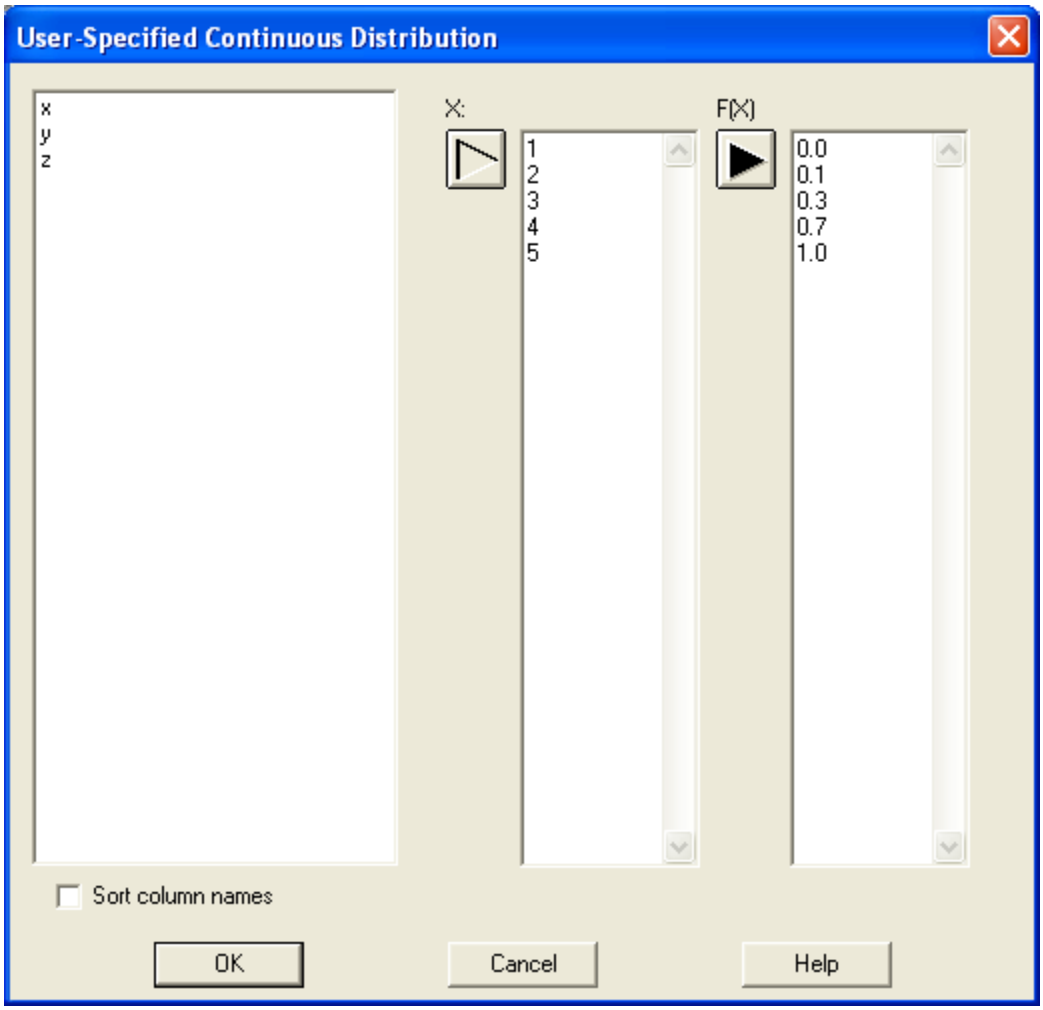
The parameters of each available distribution and the method by which random numbers are generated are described in the pdf document titled *Probability Distributions*.

You may also specify a distribution of your own, either discrete or continuous. If you select *User-specified discrete r.v.* in the distribution field, the following dialog box will be displayed:



Specify each possible value of the random variable in the *X* field and the probability of each value in the *P(X)* field. All probabilities must be between 0 and 1 and must sum to 1. You may type in the values of *X* and *P(X)*, or use the arrows to move those values from columns of a datasheet.

If you select *User-specified cont. r.v.* in the distribution field, the following dialog box will be displayed:



Specify the range of possible values of the random variable in the *X* field and the *cumulative* probability of each value in the *F(X)* field. The values of *X* must be unique and be in increasing order. The first value of *F(X)* must equal 0, while the last value must each 1. All other values of *F(X)* must be in increasing order. The cumulative distribution is assumed to increase linearly between each specified value of *X*. For example, the above dialog box specifies a distribution for *X* ranging between 1 and 5. The probability of obtaining a value between 1 and 2 is 0.1, the probability of obtaining a value between 2 and 3 is 0.2, etc.

After specifying all distributions, press OK. The random samples will then be created and added to the selected datasheet.

Analysis Options

The *Analysis Options* pane summarizes the samples that have been created:

Random Number Generation
Sample size: 10000
Seed for random number generator: 5590

Column Name	*Distribution*	*Parameters*
x	Normal	NORMAL(5.0,1.0)
y	Lognormal	LOGNORMAL(5.0,1.0,0.0)
z	Logistic	LOGISTIC(5.0,1.0)

As an example, samples of 10,000 values each have been generated from 3 distributions, each with the same mean (5) and standard deviation (1).

Summary Statistics

The *Summary Statistics* pane calculates a number of different statistics that are commonly used to summarize a sample of *n* observations:

Summary Statistics

	x	*y*	*z*
Count	10000	10000	10000
Average	5.0045	5.01011	4.9931
Standard deviation	1.00759	1.00703	1.00396
Coeff. of variation	20.1337%	20.0999%	20.1069%
Minimum	1.23009	2.36256	-0.249207
Maximum	9.24204	10.6265	10.551
Range	8.01195	8.26394	10.8002
Stnd. skewness	-0.123052	25.2443	-1.43823
Stnd. kurtosis	-0.360253	15.5069	25.07

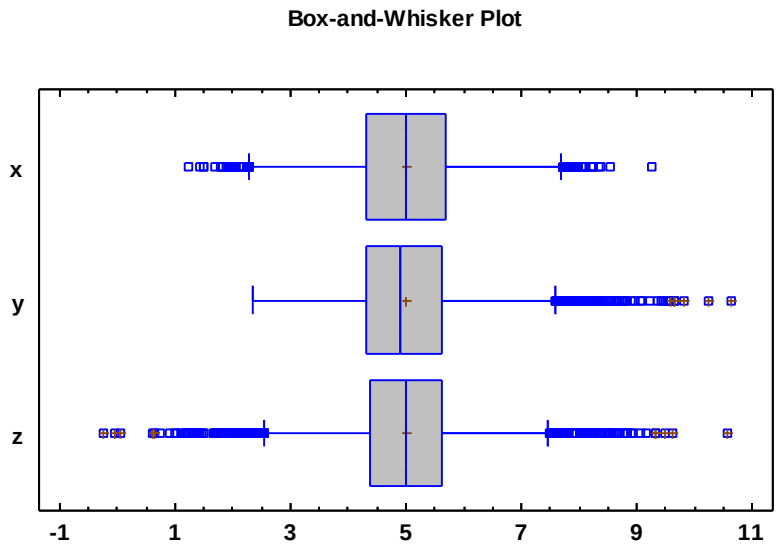
Most of the statistics fall into one of three categories:

1. measures of *central tendency* – statistics that characterize the "center" of the data.
2. measure of *dispersion* – statistics that measure the spread of the data.
3. measures of *shape* – statistics that measure the shape of the data relative to a normal distribution.

The statistics included in the table by default are controlled by the settings on the *Stats* pane of the *Preferences* dialog box. Within the procedure, the selection may be changed using *Pane Options*. The meaning of each statistic is described in the pdf document titled *One Variable Analysis*.

Box-and-Whisker Plots

This pane displays multiple box-and-whisker plots:

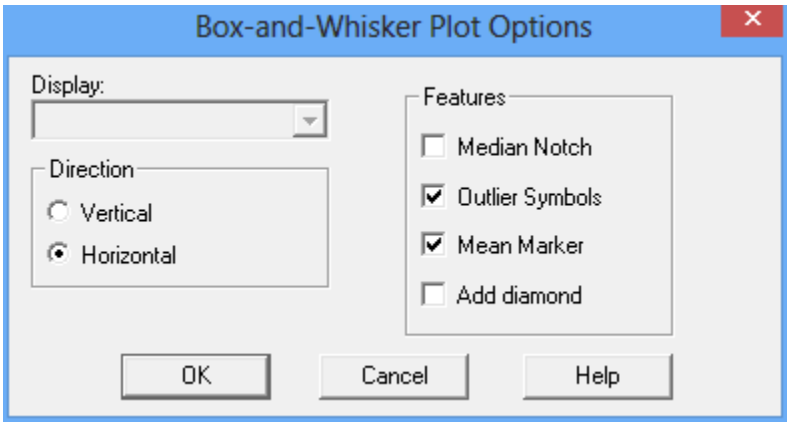


A box-and-whisker plot is constructed in the following manner:

- A box is drawn extending from the *lower quartile* of the sample to the *upper quartile*. This is the interval covered by the middle 50% of the data values when sorted from smallest to largest.
- A vertical line is drawn at the *median* (the middle value).
- If requested, a plus sign is placed at the location of the sample mean.
- Whiskers are drawn from the edges of the box to the largest and smallest data values, unless there are values unusually far away from the box (which Tukey calls *outside points*). Outside points, which are points more than 1.5 times the interquartile range (box width) above or below the box, are indicated by point symbols. Any points more than 3 times the interquartile range above or below the box are called *far outside points*, and are indicated by point symbols with plus signs superimposed on top of them. If outside points are present, the whiskers are drawn to the largest and smallest data values which are not outside points.

It can be seen from the above plot that the sample from the lognormal distribution is not as symmetric as the other 2 samples. In addition, the logistic distribution has a smaller interquartile range than the others but also has longer tails.

Pane Options



- **Direction**: the orientation of the plot, corresponding to the direction of the whiskers.
- **Median Notch**: if selected, a notch will be added to each plot showing an approximate 100(1-α)% confidence interval for the median at the default system confidence level (set on the *General* tab of the *Preferences* dialog box on the *Edit* menu).
- **Outlier Symbols**: if selected, indicates the location of outside points.
- **Mean Marker**: if selected, shows the location of the sample mean as well as the median.
- **Add diamond:** if selected, a diamond will be added to the plot showing a 100(1-α)% confidence interval for the mean at the default system confidence level.

Percentiles

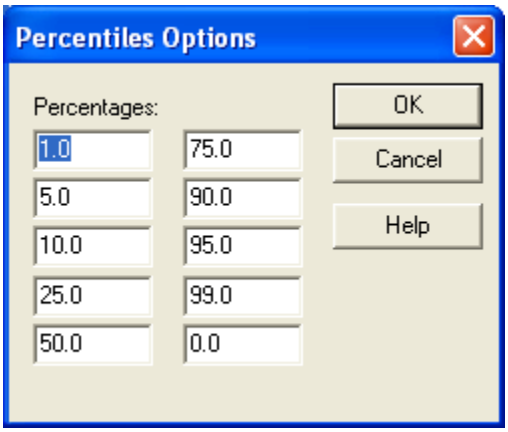
The *p-th percentile* of a continuous probability distribution is defined as that value of X for which the probability of being less than or equal to X equals p/100. For example, the 90-th percentile is that value below which lies 90% of the population. The *Percentiles* pane displays a table of selected percentiles based on the sample data.

Percentiles

Percentage	*x*	*y*	*z*
1.0%	2.68044	3.09314	2.4189
5.0%	3.34197	3.55172	3.36307
10.0%	3.70646	3.81097	3.77212
25.0%	4.3194	4.30013	4.38382
50.0%	4.99531	4.90467	4.9972
75.0%	5.67092	5.61482	5.60817
90.0%	6.30364	6.33632	6.20055
95.0%	6.65516	6.80164	6.59794
99.0%	7.34049	7.8422	7.51356

Note the long lower tail for sample z, which comes from the logistic distribution.

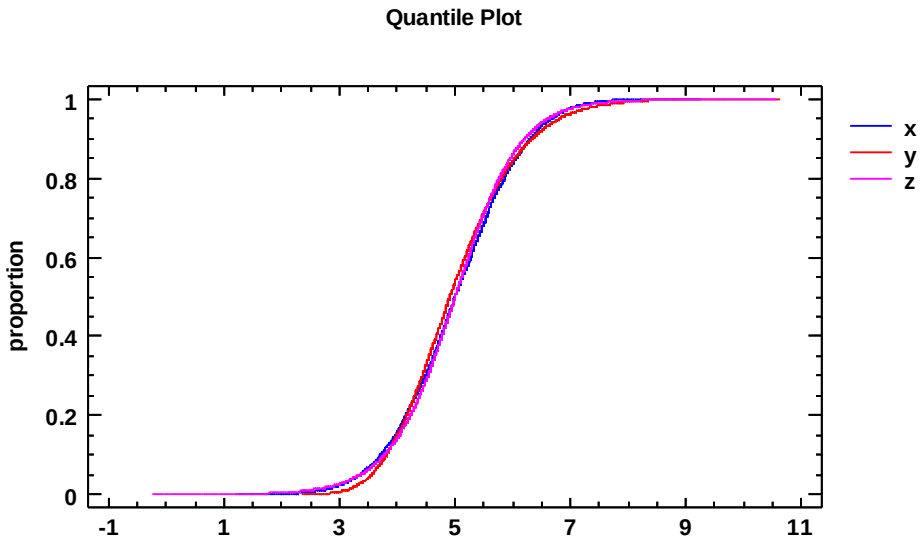
Pane Options



- **Percentages**: the percentages at which percentiles should be calculated. Set to 0 to suppress the calculation.

Quantile Plots

This pane plots the quantiles (percentiles) of the data.



For each sample, the data are sorted from smallest to largest and plotted at the coordinates

$$\left(x_{(j)}, \frac{j - 0.5}{n} \right) \tag{1}$$

for $j = 1, 2, \ldots, n$, where n is the number of observations in the sample.