Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

Given any sample it is useful to generate a summary of the all the parameters that can be estimated together with the ability to plot the data in alternative ways.

Summary statistics

For example, from the main SIMFIT menu choose [Statistics] then [Data exploration] and read the default vector test file `normal.tf1` into the procedure called exhaustive analysis of a vector. Here you can obtain the usual summary statistics as in this table, including the range, hinges (i.e. quartiles), mean $\bar{x}$, standard deviation s, coefficient of variation CV% ($100s/\bar{x}$, i.e. the reciprocal of the signal to noise ratio), and the normalized sample moments s_3 (coefficient of skewness), and s_4 (coefficient of kurtosis).

Exhaustive analysis of a vector
Data: Test file normal.tf1: 50 random numbers

Sample size	50
Minimum, Maximum values	-2.20820, 1.61750
Lower and Upper Hinges	-0.85502, 0.78597
Coefficient of skewness	-0.01669
Coefficient of kurtosis	-0.76840
Median value	-0.09736
Sample mean	-0.02579
Sample standard deviation	1.00553: CV% = 3899%
Standard error of the mean	0.14220
Upper 2.5% t-value	2.00958
Lower 95% confidence limit for mean	-0.31156
Upper 95% confidence limit for mean	0.25998
Variance of the sample	1.01109
Lower 95% confidence limit for variance	0.70552
Upper 95% con limit for variance	1.57006
Shapiro-Wilks W statistic	0.96270
Significance level for W	0.1153 *Tentatively accept normality*

Testing for a normal distribution

The normalized sample moments shown in this table are useful for seeing how far a sample departs from a normal distribution and are defined in a sample of size n by the following equations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$s_3 = \frac{n}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

$$s_4 = \frac{(n+1)n}{(n-1)(n-2)(n-3)}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

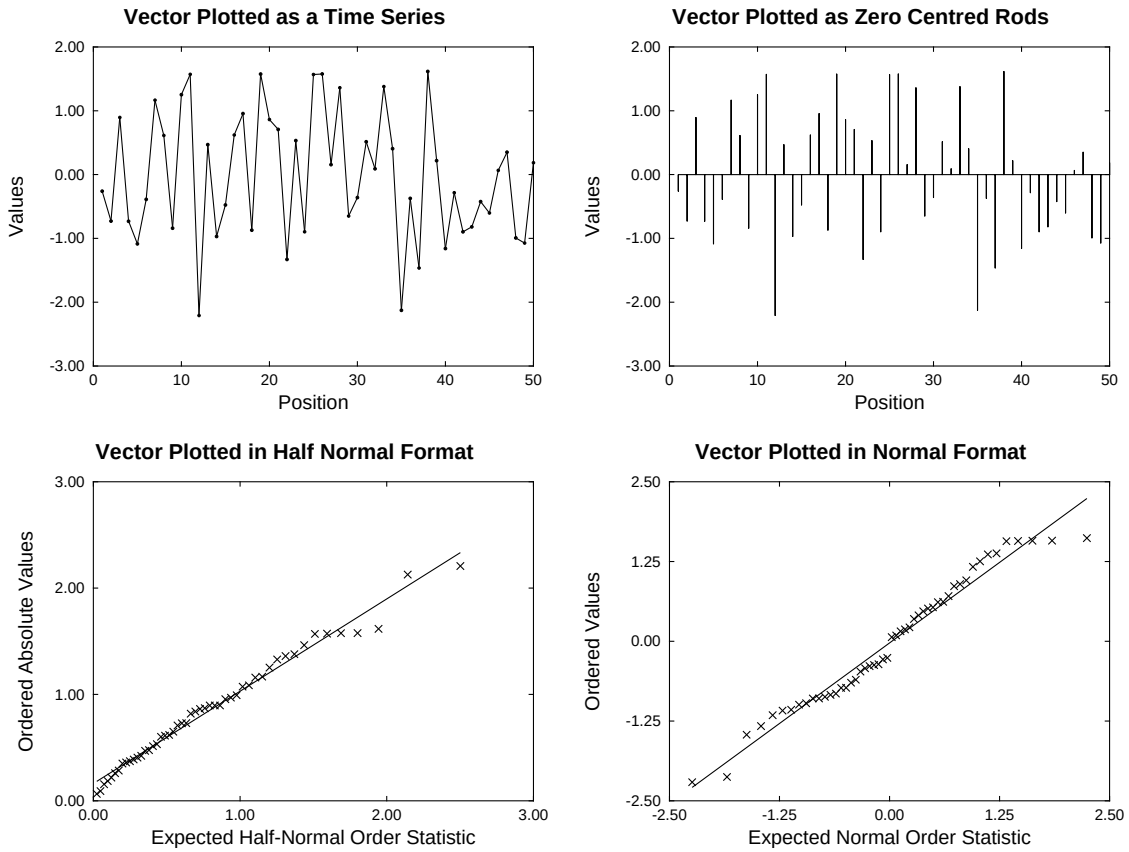
The coefficient of skewness or symmetry indicates the extent to which the sample suggest deviation from a symmetrical distribution. Values less than zero indicate skew to the left with a mean less than the median, while values greater than zero indicate skew to the right with mean greater than the median. The coefficient of kurtosis indicates the amount of peakedness in the distribution. Values less than zero indicate a platykurtic distribution which is more humped than a normal distribution, while values greater than zero indicate a leptokurtic distribution which is more peaked than a normal distribution. A normal distribution is said to be mesokurtic with both coefficients equal to zero.

As it is often wished to see how closely a sample resembles a normal distribution several options are provided for this purpose. You can perform a Shapiro-Wilks test for normality (only on demand since this will, of course, not always be appropriate) or create a histogram, pie chart, cumulative distribution plot or appropriate curve-fitting files. This option is a very valuable way to explore any single sample before considering other tests.

Graph plotting options

Since vectors have only one coordinate, graphical display requires a further coordinate. In the case of histograms the extra coordinate is provided by the choice of bins, which dictates the shape, but in the case of cumulative distributions it is automatically created as steps and therefore of unique shape. Pie chart segments are calculated in proportion to the sample values, which means that this is only appropriate for positive samples, e.g., counts.

The other techniques illustrated in this next figure may require further explanation.



If the sample values have been measured in some sequence of time or space, then the y values could be the sample values while the x values would be successive integers, as in the time series plot. Sometimes it is

useful to see the variation in the sample with respect to some fixed reference value, as in the zero centered rods plot. The data can be centered automatically about zero by subtracting the sample mean if this is required.

The half normal and normal plots are particularly useful when testing for a normal distribution with residuals, which should be approximately normally distributed if the correct model is fitted.

In the half normal plot, the absolute values of a sample of size n are first ordered then plotted as $y_i, i = 1, \ldots, n$, while the half normal order statistics are approximated by

$$x_i = \Phi^{-1}\left(\frac{n + i + \frac{1}{2}}{2n + \frac{9}{8}}\right),\ i = 1, \ldots, n$$

which is valuable for detecting outliers in regression.

The normal scores plot simply uses the ordered sample as y and the normal order statistics are approximated by

$$x_i = \Phi^{-1}\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right),\ i = 1, \ldots, n$$

which makes it easy to visualize departures from normality. Best fit lines, correlation coefficients, and significance values are also calculated for half normal and normal plots.

Note that elsewhere in SimFIT a more accurate calculation for expected values of normal order statistics is employed for a normal scores plot and also the Shapiro-Wilks test is just one of several tests for normality available.