

Evaluating Normality

There are several ways of checking for the normality assumption. Normal probability plots of the residuals and histograms of the residuals are graphical methods to evaluate the normality of the data and are available as part of the ODS Graphics output for PROC GLM. Normal probability plots graph the distribution of the residuals against how the residuals would be distributed if they were normally distributed. If the residuals are normally distributed, the plot should be a straight line with a slope of one. Histograms can be constructed with normal curves superimposed on them. This enables you to visually compare the distribution of the residuals against normally distributed data with the same mean and variance.

PROC UNIVARIATE can be used to determine whether your data is normal. This procedure generates a variety of summary statistics, such as the mean and median, as well as numerical representations of properties such as skewness and kurtosis. If the population from which the data is obtained is normal, the mean and median should be equal or nearly equal. The skewness coefficient, which is a measure of symmetry, should be near zero. Positive values for the skewness coefficient indicate that the data is right skewed, and negative values indicate that that data is left skewed. The kurtosis coefficient, which is a measure of spread, should also be near zero. Positive values for the kurtosis coefficient indicate that the distribution of the data is steeper than a normal distribution, and negative values for kurtosis indicate that the distribution of the data is flatter than normal distribution.

The NORMAL option in PROC UNIVARIATE produces a table with tests for normality. In general, if the p -values are less than 0.05, then the data should be considered non-normally distributed. It should be noted that although the kurtosis statistic is usually equal to three for the normal distribution, it was rescaled in PROC UNIVARIATE to be zero for normally distributed data.