Confidence & Prediction Intervals

These intervals are usually specified as an interval above and below the point estimate (eg the mean). These are the set of parameter values which could reasonably be expected to have given rise to the observed sample (or sample estimate), usually specified by the confidence coefficient, typically 95% or 99%. These values refer to the probability content of the specified interval, ie the likelihood of the true population parameter value actually belonging to the confidence interval. For discrete distributions, it is customary to refer to confidence intervals which are at least 95% etc.

The confidence interval level for the mean of the sample population is specified as 90%, 95% or 99% in both this and the Preferences dialog box. This level will be used in the Prediction intervals calculation also. The population under consideration is assumed to be at least approximately normally distributed, and a correction is included using the t-distribution. For example, if n is the number of data points, then the t-distribution value for n-1 degrees of freedom (denoted t(n-1)) at the 0.975 (95%) probability is used in conjunction with the calculated mean (xbar), standard deviation (s) and n in the formula:

$$C.I. = [\ \bar{x} - t_{(n-1)}\, s/\sqrt{n};\quad \bar{x} + t_{(n-1)}\, s/\sqrt{n}\ ]$$

As n increases the width of the confidence interval decreases in magnitude, ie. the accuracy with which the true population mean is being estimated improves. In addition, as n increases the sample standard deviation also approaches the true population standard deviation.

The Prediction intervals allow you to estimate the likely range of values into which eg 95% of the values in the distribution given by your sample data would be expected to fall, whereas the confidence intervals give the likely range of values for the sample mean estimate (with 95% confidence at being correct).