One-Minute Spotlight

DISTRIBUTION FITTING

Often, when working on a project or problem, you may have the option of using real data. There are two types of historical data: data that exhibits some form of trend and/or seasonality over time (referred to as time-series data) and data that varies randomly within some bounds. Examples of this second form of data are the cost of procuring a machine part (for cost estimating), the performance of a sales person compared to quota, or even plane arrival times.

To model time-series data, you should use CB Predictor. However, if you have random historical data available, Crystal Ball's distribution fitting feature can substantially simplify the process of selecting a probability distribution. Not only is the process simple, but the resulting distribution more accurately reflects the nature of your data than if you estimated the shape and parameters of the distribution. This example uses spring coil diameter data to fit to a distribution.

HOW DISTRIBUTION FITTING WORKS

In distribution fitting, Crystal Ball automatically matches your data against the continuous and/or discrete probability distributions. You can choose to fit to some or all of the available distributions, with the exception of the Yes No and Custom distributions. Crystal Ball performs a mathematical fit to determine the set of parameters for each distribution that best describes the characteristics of your data.

Crystal Ball judges the quality, or *goodness*, of each fit using one of several standard **goodness-of-fit** tests. It then chooses the distribution with the highest-ranking fit to represent your data.

Distribution Fitting, continued

REVIEWING YOUR FIT RESULTS

You can review the distributions sorted in order of their fit tests using a comparison chart. This chart shows the fitted distributions superimposed over your data, letting you to visually check the quality of the fits.
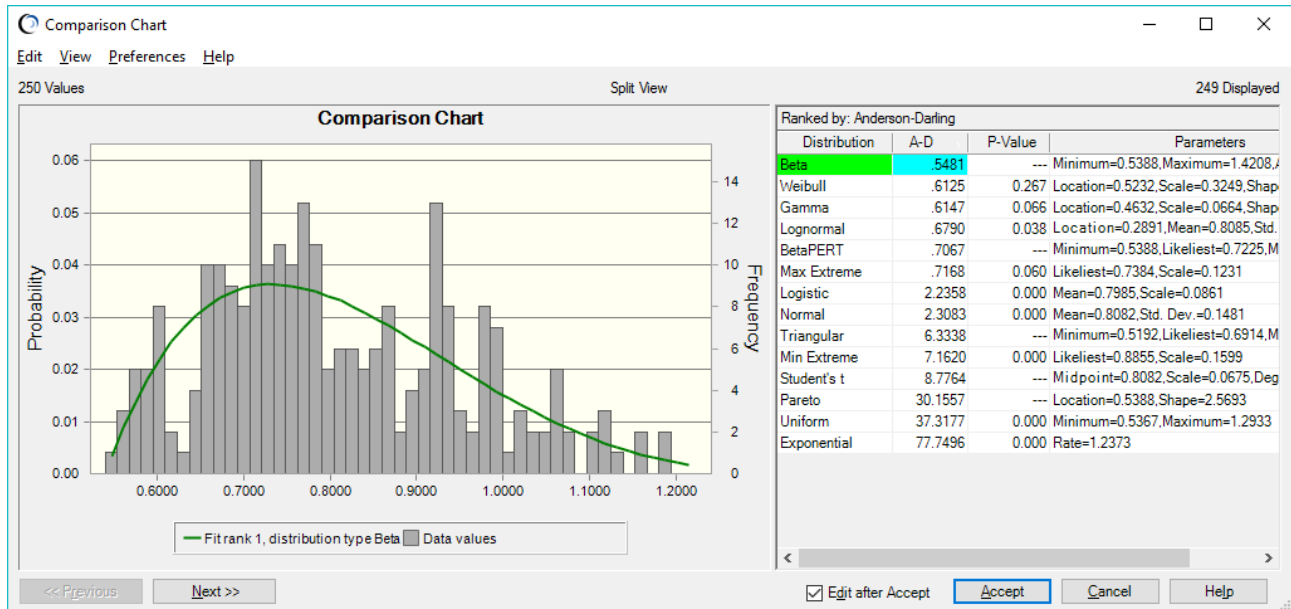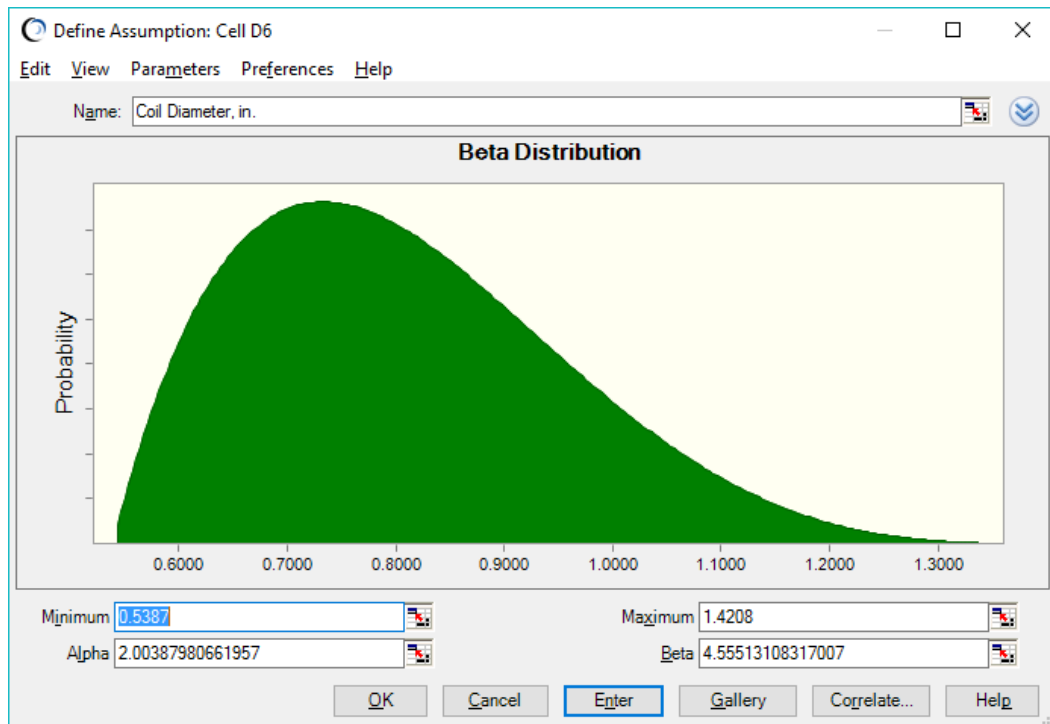

A summary table can show some or all of the goodness-of-fit statistics, including P-values for certain distributions. You can override the highest-ranking probability distribution with another one of your choice.

Each goodness-of-fit test is calculated for every distribution, but only the selected test determines how the distributions are ranked. The Chi-Square test is selected for discrete distributions, and the default for continuous distributions is the Anderson-Darling test.

CREATING A DISTRIBUTION

The final step in fitting is to select a probability distribution with an appropriate fit. Crystal Ball then applies the calculated parameters to the appropriate distribution and displays the distribution for you to review and select.

Distribution Fitting, continued

For more information or to contact us, browse to
http://www.oracle.com/technetwork/middleware/crystalball/overview/index.html.