Overview

This chapter describes operations and functions for statistical analysis together with some general guidelines for their use. This is not a statistics tutorial; for that you can consult one of the references at the end of this chapter or the references listed in the documentation of a particular operation or function. The material below assumes that you are familiar with techniques and methods of statistical analysis.

Most statistics operations and functions are named with the prefix "Stats". Naming exceptions include the random noise functions that have traditionally been named based on the distribution they represent.

There are six natural groups of statistics operations and functions. They include:

• Test operations

• Noise functions

• Probability distribution functions (PDFs)

• Cumulative distribution functions (CDFs)

• Inverse cumulative distribution functions

• General purpose statistics operations and functions

Statistics Test Operations

Test operations analyze the input data to examine the validity of a specific hypothesis. The common test involves a computation of some numeric value (also known as "test statistic") which is usually compared with a critical value in order to determine if you should accept or reject the test hypothesis (H_0). Most tests compute a critical value for the given significance *alpha* which has the default value 0.05 or a user-provided value via the /ALPH flag. Some tests directly compute the P value which you can compare to the desired significance value.

Critical values have been traditionally published in tables for various significance levels and tails of distributions. They are by far the most difficult technical aspect in implementing statistical tests. The critical values are usually obtained from the inverse of the CDF for the particular distribution, i.e., from solving the equation

$$cdf(criticalValue) = 1 - alpha,$$

where *alpha* is the significance. In some distributions (e.g., Friedman's) the calculation of the CDF is so computationally intensive that it is impractical (using desktop computers in 2006) to compute for very large parameters. Fortunately, large parameters usually imply that the distributions can be approximated using simpler expressions. Igor's tests provide whenever possible exact critical values as well as the common relevant approximations.

Comparison of critical values with published table values can sometimes be interesting as there does not appear to be a standard for determining the published critical value when the CDF takes a finite number of discrete values (step-like). In this case the CDF attains the value (1-*alpha*) in a vertical transition so one could use the X value for the vertical transition as a critical value or the X value of the subsequent vertical transition. Some tables reflect a "conservative" approach and print the X value of subsequent transitions.

Statistical test operations can print their results to the history area of the command window and save them in a wave in the current data folder. Result waves have a fixed name associated with the operation. Elements in the wave are designated by dimension labels. You can use the /T flag to display the results of the operation in a table with dimension labels. The argument for this flag determines what happens when you kill the table. You can use/Q in all test operations to prevent printing information in the history area and you can use the /Z flag to make sure that the operations do not report errors except by setting the V_Flag variable to -1.

Statistical test operations tend to include several variations of the named test. You can usually choose to execute one or more variations by specifying the appropriate flags. The following table can be used as a guide for identifying the operation associated with a given test name.