ANOVA on Numeric Columns

Author: Autumn Laughbaum, Golden Helix, Inc.

Overview

This function makes use of the scipy package, specifically the *scipy.stats.f_oneway* and *scipy.stats.kruskal* functions. This requires a categorical dependent column that provides the grouping structure and several numeric columns.

Recommended Directory Location

Save the script to the following directory:
*..**\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Numeric**

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Vista. The easiest way to locate this directory on your computer is to open SVS and select **Tools > Open Folder > User Scripts Folder**. If saved to the proper folder, this script will be accessible from the spreadsheet **Numeric** menu.

Preparing to use the Script

This script should be run from a spreadsheet containing a categorical dependent column and several active numeric columns.

1. From an appropriate spreadsheet, specify a categorical column as dependent by clicking once on the column header, turning the column magenta. Then choose **Numeric > ANOVA on Numeric Columns**.
2. Choose the appropriate test; either the Anova F-test or Kruskal-Wallis H-Test (Nonparametric version). You may also choose to output the Bonferroni adjusted p-values and the –log10 p-values.
3. The resulting spreadsheet is named One-Way Anova Results or Kruskal-Wallis H-Test Results. It has a column containing the test statistic for each active numeric column in the original spreadsheet, a p-value column and optional –log10(P) and Bonf-P columns. If a marker map was applied to the columns of the original spreadsheet, it is reapplied to the rows of the Results spreadsheet.

For more information about the internal scipy functions see:

http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html#scipy.stats.f_oneway

and

http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html#scipy.stats.kruskal