

## ANOVA with Phenotype and SNPs

**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 numeric phenotype column and several genotype columns which provide the grouping structure in each test. Including missings as a group is optional.

### Recommended Directory Location

Save the script to the following directory:

**\*..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Genotype**

**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 **Genotype** 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, choose **Genotype > ANOVA with Phenotype and SNPs**.
2. Choose the numeric phenotype column and the appropriate test; either the Anova F-test or Kruskal-Wallis H-Test (Nonparametric version).
3. Optionally you can include missing genotypes as a group and you may also choose to output the Bonferroni adjusted p-values and the  $-\log_{10}$  p-values. If you do not want to run the script over every genotype column, you can select a subset of columns to use.
4. 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  $-\log_{10}(P)$  and Bonf-P columns. There are also columns containing the major and minor allele, the mean, standard deviation and sample size for each group and unique allele. 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](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>