

## Chi-Squared Contingency Table

Author: Greta Peterson, Golden Helix, Inc.

### Overview

This script computes the Pearson's Chi-Squared Statistic for a contingency table with  $m$  groups and  $n$  observations ( $m$  rows and  $n$  columns). The  $p$ -values,  $-\log_{10}$   $p$ -values, Bonferroni  $p$ -values,  $-\log_{10}$  Bonferroni  $p$ -values, and FDR values are also computed.

### Recommended Directory Location

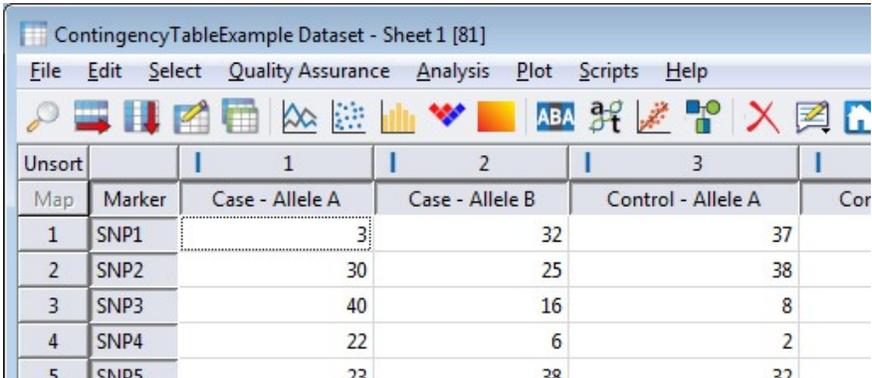
Save the script to the following directory:

\*..\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 operating systems. The easiest way to locate this directory on your computer is to open SVS and go to **Tools > Open > Open UserScripts Folder** and save the script in the **\Spreadsheet\Numeric** folder. If saved to the proper folder, this script will be accessible from the spreadsheet **Numeric** menu.

### Format of the Spreadsheet

This script assumes that the cell counts for the  $m$ -by- $n$  contingency tables are contained in one row with columns specified for each group and outcome combination. A spreadsheet with  $r$  rows will compute the chi-squared statistic for  $r$  contingency tables.



Map	Marker	Case - Allele A	Case - Allele B	Control - Allele A	Control - Allele B
1	SNP1	33	32	37	38
2	SNP2	30	25	38	37
3	SNP3	40	16	8	37
4	SNP4	22	6	2	37
5	SNP5	22	28	27	37

Figure 1: Example of spreadsheet containing contingency table data

Only integer columns can be selected for input columns for the script.

### Using the Script

1. Open the spreadsheet with the columns containing the cell counts for the contingency table, like Figure 1.

- Go to **Numeric > Chi-Squared Contingency Table**.
- Select the number of groups and outcomes, in this case, there are two groups (Case and Control) and two outcomes (Allele A and Allele B). See Figure 2.

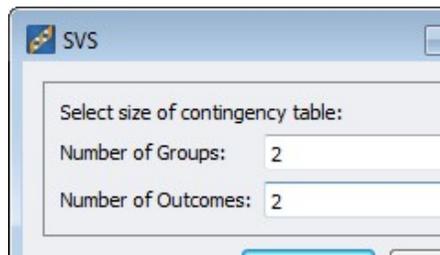


Figure 2: Specify the dimensions of the contingency table

- Next, from the drop down menus listing all integer columns in the spreadsheet, select the appropriate columns for each group and outcome combination. See Figure 3

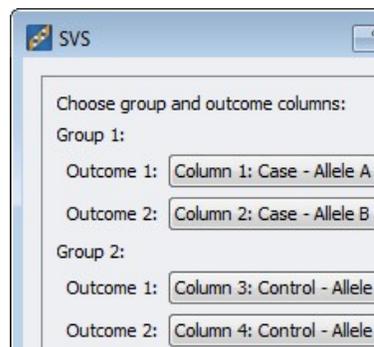


Figure 3: Specification of the count columns

- The output will contain values for  $X^2$  Stat, df (degrees of freedom),  $X^2$  P-value,  $-\log_{10} P$  ( $-\log_{10}(X^2 \text{ P-value})$ ), Bonf. P (p-value \* number of tests),  $-\log_{10}$  Bonf. P ( $-\log_{10}(\text{Bonf. P})$ ), and FDR. For other table sizes there will only be values for  $X^2$  Stat and df. See Figure 4.

Chi-Squared Values [121]							
File Edit Select Quality Assurance Analysis Plot Scripts Help							
Unsort		R	I	R	R	R	R
Map	Marker	1	2	3	4	5	6
		$X^2$ Stat	DF	$X^2$ P-value	$-\log_{10} P$	Bonf. P	$-\log_{10}$ Bonf. P
1	SNP1	20.9913460348029	1	4.61362625886768e-006	5.33595758976413	2.30681312943384e-005	4.636987!
2	SNP2	11.5061669829222	1	0.000693656670505577	3.15885543300234	0.00346828335252788	2.459885!
3	SNP3	27.1732460426031	1	1.86015044749816e-007	6.73045192896707	9.30075223749082e-007	6.031481!
4	SNP4	35.0134615384615	1	3.2743381478133e-009	8.48487647209999	1.63716907390665e-008	7.785906!
5	SNP5	24.1579479415545	1	8.87491372071237e-007	6.05183586033491	4.43745686035618e-006	5.35286!

Figure 4: Results from the  $X^2$  Contingency Table calculations