CMH Test over Several Strata

Author: Autumn Laughbaum, Golden Helix, Inc.

Overview

This script is designed to calculate Cochran-Mantel-Haenszel statistics, given several different spreadsheets corresponding to data from several different strata. Each spreadsheet must have one binary dependent column and several genotypic columns. Individual genotype associations test are computed for each strata, then the appropriate cell counts are calculated based on the genetic model selected.

Recommended Directory Location

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

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 go to **Tools > Open Folder > UserScripts Folder**. If saved to the proper folder, this script will be accessible from the project navigator's **Tools** menu.

Internal Calculations

For each marker, 2x2 tables are computed for every strata. The tables are constructed based on the genetic model selected; Basic Allelic, Dominant, or Recessive. The following tables demonstrate the difference.

Basic Allelic:

Marker X	D	d
Case	a	b
Control	c	d

Dominant:

Marker X	DD + Dd	dd
Case	a	b
Control	c	d

Recessive:

Marker X	DD	Dd + dd
Case	a	b
Control	c	d

Where **d** is the major allele and **D** is the minor allele computed through the genotype association tests. See the [Genotype Association Tests](#) section in the [SVS 7 Manual](#) for more information regarding these tests. Note: Although other genetic models are available with Genotype Association Tests, only the specified models have corresponding 2x2 tables and thus can be used for CMH tests.

For each marker, a set of 2x2 tables are formed. Then a CMH Chi-Squared statistic is calculated as follows with continuity correction:

$$\chi^2_{MH} = \frac{\{|\sum[a - (a + b)(a + c)/n]| - 0.5\}^2}{\sum(a + b)(a + c)(b + d)(c + d)/(n^3 - n^2)}$$

and without the continuity correction:

$$\chi^2_{MH} = \frac{\{|\sum[a - (a + b)(a + c)/n]|\}^2}{\sum(a + b)(a + c)(b + d)(c + d)/(n^3 - n^2)}$$

A p-value is calculated using this statistic and a Chi-Square distribution with 1 degree of freedom. The following statistics are also calculated using the p-value.

CMH -log10(p) = -log$_{10}$(p-value)

CMH Bonf P = N * p-value

Where N is the number of tests performed, or the number of intersecting markers (i.e. markers that are found in every strata).

The Mantel-Haenszel estimate of the common odds ratio (**Estimate OR**) is computed as:

$$OR_{MH} = \frac{\sum \frac{ad}{n}}{\sum \frac{bc}{n}}$$

where $n = a + b + c + d$

The variance of the **log(Estimate OR)** is computed as:

$$\text{Var}(\ln(OR_{MH})) = \frac{\sum \frac{ad(a + d)}{n^2}}{2\left[\sum \frac{ad}{n}\right]^2} + \frac{\sum \frac{bc(a + d) + ad(b + c)}{n^2}}{2\left[\sum \frac{ad}{n}\right]\left[\sum \frac{bc}{n}\right]} + \frac{\sum \frac{bc(b + c)}{n^2}}{2\left[\sum \frac{bc}{n}\right]^2}$$

Therefore, the 95% confidence interval for the estimate of the common odds ratio is

$$\left(OR_{MH} * \exp\left(-1.95\sqrt{Var(\ln(OR_{MH}))}\right), OR_{MH} * \exp\left(1.95\sqrt{Var(\ln(OR_{MH}))}\right) \right)$$

The OR Test Statistic is calculated by dividing the log of the common OR estimate by the standard error (or square root of the variance) of the log of the OR estimate.

$$z^2 = \frac{OR_{MH}}{\sqrt{Var(\ln(OR_{MH}))}}$$

The p-value is based on a log normal distribution.

Using the Script

1. Open a project that contains several spreadsheets with data from several strata. Go through each spreadsheet and designate the appropriate binary column as dependent (usually Case/Control).

2. From the project navigator, choose **Tools > CMH Test for Several Strata**.

3. Choose the appropriate spreadsheets and select a genetic model. Optionally choose to use the continuity correction. Click **OK**.

Separate genotype association tests will be performed for each spreadsheet. The appropriate cell counts from these spreadsheets will be used to calculate the CMH statistics. The newly created sheet, **Cochran-Mantel-Haenzsel Test Results**, contains CMH statistics for all markers that were found in every strata.