

hweStrata
(version 1.0)

Software for Exact Stratified Test for
Hardy-Weinberg Equilibrium

DJ Schaid, GD Jenkins, AJ Batzler, JP Sinnwell
Mayo Clinic, Rochester MN USA
Contact: schaid@mayo.edu

May 3, 2011

Description

hweStrata calculates an exact stratified test for HWE for diallelic markers, such as single nucleotide polymorphisms (SNPs), and an exact test for homogeneity of Hardy Weinberg disequilibrium. In addition, exact tests for HWE are calculated for each stratum. For more details on the tests, see Schaid et al. [1].

New to this version is an update to the homogeneity exact test. The exact test for homogeneity of Hardy-Weinberg disequilibrium sometimes cannot be calculated; in those instances hweStrata switches to an asymptotic test defined by Olson and Foley [2].

The exact test cannot be calculated when the log of the denominator of equation (3) in Schaid et al. [1] cannot be represented by a double data type in C, which is a function of the possible number of heterozygotes over all strata, as described in Appendix A of Schaid et al. [1]. Because the limit is

system-dependent, we cannot provide an exact upper limit, but on our CentOS 64-bit Linux machines the limit occurs when the number of possible heterozygotes exceeds 300. The asymptotic test has proven to perform well in most scenarios when that count is reached, as shown in Table 1 of Olson and Foley [2]. We demonstrate the use of this update in our example below.

Operating System and Installation

hweStrata is written in the C programming language, with commands from the user given at the command-line. Because the user interface is simplistic, the program is easy to install on multiple computer platforms. Currently, we distribute the source code and executables for Linux and Windows operating systems.

Data Input Format

A single data file is required as input to hweStrata. This file has a row for each marker. Each row must have a marker label followed by counts of the three possible genotypes for each of the strata. For K strata, there will be 3*K genotype counts. The genotype counts for each stratum are expected to be in the order of the genotypes AA, AB, and BB, where A and B are the two alleles of a diallelic marker. It does not matter which allele is labeled A and which is labeled B, but what does matter is that the heterozygote is in the middle, and the genotype counts are arranged in the same manner across all strata. This input format is illustrated below for three markers and three strata. We provide headers in red font above the line to describe the input, but only the content below the line is in the input file. The input file can be tab- or space-delimited.

Stratum	1			2			3		
Genotype	AA	AB	BB	AA	AB	BB	AA	AB	BB
Marker1	4	13	1	3	4	5	3	4	1
Marker2	16	73	224	18	78	242	21	72	245
Marker3	5	5	13	19	4	1	10	6	7

Running hweStrata

For users who are unfamiliar with running an executable on command line on either Windows or unix systems, we recommend copying all the contents in *hweStrata/bin* to the directory containing the input genotype file.

In Windows, the MS-DOS command shell can be accessed via *Start- >Run*, then enter *cmd* at the prompt. Within the DOS command shell, use *cd* to change your current directory to where the genotype and hweStrata executable files are stored.

We refer to the executable as *hweStrata*, but note users to run the system-appropriate version from your command shell. You can run *hweStrata* from the command line with three possible input parameters. Typing *hweStrata -h* gives this help menu below describing them:

```
Command line options for hweStrata, version 1.0
-h                : print this help menu
-geno <file>: input genotype file
                 (e.g., -geno geno.dat) default = geno.dat
-pstop : stop computing pval for HWE if p > pstop
        (e.g., -pstop 0.5) default = 1.100000
```

To analyze an input file called *test.dat*, simply type

```
hweStrata -geno test.dat
```

which will display results on the computer screen. To direct output to a file, say *test.out*, simply type

```
hweStrata -geno test.dat > test.out
```

The optional parameter *-pstop* quits computing the exact stratified test if the p-value is larger than the value following *-pstop*. The default is not to stop (by setting *-pstop 1.1*), but to stop at say 0.05, simply add to the command line *-pstop 0.05*.

A test file, called *test.dat* is distributed with *hweStrata*, along with its output in the file *save.out*. When first using *hweStrata*, it is a good idea to compare the output from *test.dat* with the distributed file *save.out* to be sure that *hweStrata* is properly working.

Output

The output from *hweStrata* for the *test.dat* example is illustrated below, where the second marker has high genotype counts that require use of the asymptotic test for homogeneity.

```
**** Program: hweStrata
**** Version: 1.0
**** Authors: DJ Schaid, AJ Batzler, GD Jenkins, JP Sinnwell
```

locus	homog	hweAll	hwe1	hwe2	hwe3
Marker1	1.046541e-01	1.390454e-01	7.382614e-02	2.892718e-01	1.000000e+00

Marker2	6.587511e-01**	8.453578e-08	7.084521e-03	2.838904e-03	4.815379e-05
Marker3	1.000000e+00	2.321356e-03	1.911668e-02	2.980373e-01	3.256184e-02

Note: ** asymptotic test

- **locus:** marker name as defined in the input file
- **homog:** exact p-value for test of homogeneity of Hardy Weinberg disequilibrium across strata; asymptotic p-value used when ** indicates
- **hweAll:** exact stratified test of HWE across all strata
- **hwe1 ... hwe3:** exact tests of HWE within each stratum

License and Warranty

License and Warranty

Copyright 2003 Mayo Foundation for Medical Education and Research.

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to indent Free Software Foundation, Inc.

59 Temple Place, Suite 330
Boston, MA 02111-1307 USA

For other licensing arrangements, please contact Daniel J. Schaid.

Daniel J. Schaid, Ph.D.
Division of Biomedicatl Statistics and Informatics
Harwick Building - Room 775

Mayo Clinic
200 First St., SW
Rochester, MN 55905
phone: 507-284-0639
fax: 507-284-9542
email: schaid@mayo.edu

Acknowledgements

This work was supported by the U.S. Public Health Service, National Institutes of Health, contract grant numbers GM065450 and GM61388 (The Pharmacogenetics Research Network).

References

- [1] Schaid D, Batzler A, Jenkins G, Hildebrandt M (2006) Exact Tests of Hardy Weinberg Equilibrium and Homogeneity of Disequilibrium Across Strata. *Am J Hum Genet* 79(6):1071-80.
- [2] Olson J, Foley M (1996) Testing for Homogeneity of Hardy-Weinberg Disequilibrium Using Data Sampled from Several Populations. *Biometrics* 52:971-9.