

hweStrata
(version 0.1)

***Software for Exact Stratified Test for Hardy Weinberg
Equilibrium***

Daniel J. Schaid, Anthony J. Batzler, Gregory D. Jenkins

Mayo Clinic, Rochester MN USA

Email contact: schaid@mayo.edu

November 2, 2006

Brief 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, see Schaid et al. (Schaid et al. 2006).

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 Solaris 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 two markers and three strata. The red italic font is not part of the input file, but is used to describe the input.

	<i>Stratum-1</i>			<i>Stratum-2</i>			<i>Stratum-3</i>		
	<i>AA</i>	<i>AB</i>	<i>BB</i>	<i>AA</i>	<i>AB</i>	<i>BB</i>	<i>AA</i>	<i>AB</i>	<i>BB</i>
Marker1	4	13	1	3	4	5	3	4	1
Marker2	5	5	13	19	4	1	10	6	7

Running hweStrata

hweStrata is command-line driven, with three possible input parameters. Typing hweStrata -h gives the “help” menu below:

Command line options for hweStrata, version 0.1

```
-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
```

So, 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.

locus	homog	hweAll	hwe1	hwe2	hwe3
Marker1	1.04654137e-01	1.390454e-01	7.382614e-02	2.892718e-01	1.000000e+00
Marker2	1.00000000e+00	2.321356e-03	1.911668e-02	2.980373e-01	3.256184e-02

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

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 Biostatistics
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

Schaid D, Batzler A, Jenkins G, Hildebrandt M (2006) Exact tests of Hardy Weinberg equilibrium, and homogeneity of disequilibrium, across strata. To appear in Am J Hum Genet