Package 'HardyWeinberg'

August 11, 2014

Type Package

Title Graphical tests for Hardy-Weinberg equilibrium

Version 1.5.4

Date 2014-08-11

Author Jan Graffelman <jan.graffelman@upc.edu>

Maintainer Jan Graffelman <jan.graffelman@upc.edu>

Depends R (>= 1.8.0), mice

Description Package HardyWeinberg is a package for exploring
bi-allelic marker data. It focuses on the graphical
representation of the results of tests for Hardy-Weinberg equilibrium
in a ternary plot. Routines for several tests for Hardy-Weinberg equilibrium
are included in the package. Procedures for handling missing genotype data are included.

License GPL (>= 2)

URL http://www.r-project.org, http://www-eio.upc.edu/~jan

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-11 21:44:30

R topics documented:

`HardyWeinberg-package` *Graphical tests for Hardy-Weinberg equilibrium*

Description

The HardyWeinberg-package provides graphical tests for Hardy-Weinberg equilibrium (HWE) based on the ternary plot (de Finetti diagram). The package constructs ternary plots for genotypic compositions for bi-allelic marker data. The acceptance region for several statistical tests of HWE (Chisquare test, Chisquare test with continuity correction, Haldane's exact test) can be depicted inside the ternary plot with the routines of the package. Large numbers of bi-allelic markers (e.g. SNPs) can be represented in a single ternary diagram and the statistical (non)significance of a test for HWE can be inferred from the position of the marker in the plot.

Details

Package:	HardyWeinberg
Type:	Package
Version:	1.5.3
Date:	2014-04-13
License:	GPL Version 2 or later.

The most important function of the package is `HWTernaryPlot` that can be used to create ternary plots with acceptance regions for HWE. Other routines implement statistical tests for HWE such as `HWChisq` and `HWLratio`.

Author(s)

Jan Graffelman

Maintainer: Jan Graffelman <jan.graffelman@upc.edu>

References

Weir, B.S. (1996) *Genetic Data Analysis II*. Sinauer Associates, Massachusetts.

Graffelman, J. and Morales, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* 65(2):77-84.

Examples

```
library(HardyWeinberg)

# make random compositions that are in HWE

set.seed(123)

m <- 100 # number of markers
n <- 100 # sample size

out <- HWData(n,m)
Xc <- out$Xc
out <- HWTernaryPlot(Xc,100,region=1,vertex.cex=2,signifcolour=TRUE)
```

---

af *Function to compute allele frequencies*

Description

Function `af` computes the allele frequencies for a matrix or a vector containing genotypic compositions.

Usage

```
af(x)
```

Arguments

x	a vector or matrix with compositions

Value

a vector with allele frequencies

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

maf

Examples

```
X <- as.vector(rmultinom(1,100,c(0.5,0.4,0.1)))
X <- X/sum(X)
print(X)
print(af(X))
```

Alzheimer *Genotype frequencies for 70 SNPs related to Alzheimer's disease*

Description

The dataframe contains the genotype frequencies MM, Mm and mm for the 70 SNPs for both cases and controls. The data are taken from table 7.11 in Laird & Lange.

Usage

```
data(Alzheimer)
```

Format

A data frame containing 70 observations.

Source

Laird, N. M. and Lange, C. Table 7.11, p. 124

References

Laird, N. M. and Lange, C. (2011) The fundamentals of modern statistical genetics. Springer.

fisherz *Fisher's z transformation*

Description

Calculates Fisher's z transformation for a correlation coefficient

Usage

```
fisherz(r)
```

Arguments

r	a correlation coefficient

Value

a real number

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

cor

Examples

```
r <- 0.5
print(fisherz(r))
```

GenerateSamples *Generate genotypic compositions*

Description

GenerateSamples generates all possible genotypic compositions (AA,AB,BB) for a given sample size n.

Usage

```
GenerateSamples(n = 5)
```

Arguments

n	the desired sample size

Value

returns a matrix with in each row a possible genotypic compostion for the given sample size.

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

Examples

```
GenerateSamples(5)
```

HapMapCHBChr1 *Genotype frequencies for 225 SNPs on chromosome 1 of the CHB population.*

Description

The dataframe contains the genotype frequencies in generic notation, AA, AB and BB the first 225 polymorphic SNPs without missing data on chromosome 1 of the Han Chinese in Beijing. The data are compiled from the HapMap project, phase 3.2, containing genotype information of 84 individuals.

Usage

```
data(HapMapCHBChr1)
```

Format

A matrix containing 225 rows and 3 columns (AA, AB, BB).

Source

http://hapmap.ncbi.nlm.nih.gov/

References

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs Nature 449, pp. 851–861.

HWAlltests *Perform all tests for Hardy-Weinberg equilibrium*

Description

HWAlltests performs all classical frequentists tests for Hardy-Weinberg equilibrium and lists their p-values.

Usage

```
HWAlltests(x, verbose = FALSE)
```

Arguments

x	a vector with a set of genotype counts (AA, AB, BB)
verbose	print output if set to TRUE

Value

A dataframe with test statistics and p-values.

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

HWLratio, HWChisq, HWExact

Examples

```
x <- c(298,489,213)
names(x) <- c("MM","MN","NN")
HWAlltests(x,verbose=TRUE)
```

HWAlr *Compute additive log-ratio transformation*

Description

HWAlr computes the additive log-ratio transformation for genotype counts of bi-allelic genetic markers.

Usage

```
HWAlr(X, zeroadj = 0.5, denominator = 2)
```

Arguments

X	A matrix of genotype counts (columns AA, AB and BB)
zeroadj	A zero adjustment parameter (0.5 by default)
denominator	The genotype count put in the denominator of the log-ratio (1=AA, 2=AB, 3=BB)

Value

A matrix or vector of log-ratio coordinates

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a non-parametric compositional approach. In: Vera Pawlowsky-Glahn and Antonella Buccianti (eds.) Compositional Data Analysis: Theory and Applications, John Wiley \& Sons, Ltd, pp. 207-215

See Also

HWClr,HWIlr

Examples

```
X <- HWData(100,100)$Xt
Y <- HWAlr(X)
```

HWAlrPlot *Plot genetic markers in additive log-ratio coordinates*

Description

HWAlrPlot creates a scatter plot of the log-ratio coordinates of bi-allelic genetic markers. Hardy-Weinberg equilibrium is indicated by a straigh line in the plot.

Usage

```
HWAlrPlot(X, zeroadj = 0.5)
```

Arguments

X	A matrix of genotype counts (columns AA, AB, BB)
zeroadj	Zero-adjustment parameter. Zero counts in the count matrix are substituted by zeroadj which is 0.5 by default.

Value

NULL

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a non-parametric compositional approach. In: Vera Pawlowsky-Glahn and Antonella Buccianti (eds.) Compositional Data Analysis: Theory and Applications, John Wiley \& Sons, Ltd, pp. 207-215

See Also

HWClrPlot,HWIlrPlot

Examples

```
X <- HWData(100,100)$Xt
HWAlrPlot(X)
```

HWChisq *Chi square tests for Hardy Weinberg equilibrium*

Description

HWChisq performs the chi-square test for Hardy Weinberg equilibrium with or without continuity correction.

Usage

```
HWChisq(X, cc = 0.5, alpha = 0.05, verbose = FALSE)
```

Arguments

X	X a vector containg the genotypic counts (AA,AB,BB).
cc	cc the continuity correction parameter (default cc = 0.5).
alpha	significance level (0.05 by default).
verbose	verbose = 1 prints results, verbose = 0 is silent.

Details

HWChisq does a chi-square test for Hardy-Weinberg equilibrium, and by default applies a continuity correction. For extreme allele frequencies, the continuity correction can lead to excessive type 1 error rates, and is better turned off in that case. The continuity correction can be turned off by specifying cc=0.

Value

HWChisq returns a list with the components:

chisq	value of the chi-square statistic. NA is returned if the marker is monomorphic.
pval	p-value of the chi-square test for Hardy-Weinberg equilibrium.
D	Half the deviation from Hardy-Weinberg equilibrium for the AB genotype.
p	the allele frequency of A.
f	the inbreeding coefficient.

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter3.

See Also

HWLratio

Examples

```
x <- c(MM=298,MN=489,NN=213)
HW.test <- HWChisq(x,verbose=TRUE)
```

HWChisqMat *Matrix version of HWChisq*

Description

HWChisqMat executes the Chisquare test for HWE for each row in a matrix.

Usage

```
HWChisqMat(X, ...)
```

Arguments

X	A n times 3 matrix of genotypic counts (AA,AB,BB)
...	extra arguments that are passed on to HWChisq

Value

pvalvec	Vector with the p-values of each test
chisqvec	Vector with the chi-square statistics
Dvec	Vector with deviations from independence

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

HWChisq

Examples

```
X <- HWData(100,10)$Xt
colnames(X) <- c("MM","MN","NN")
Results <- HWChisqMat(X)
Output <- cbind(X,Results$chisqvec,Results$pvalvec)
print(Output)
```

HWClr *Compute the centred log-ratio transformation*

Description

HWClr computes the centred log-ratio transformation for genotype counts of bi-allelic genetic markers.

Usage

```
HWClr(X, zeroadj = 0.5)
```

Arguments

X	A matrix of genotype counts (columns AA, AB and BB)
zeroadj	A zero adjustment parameter (0.5 by default)

Value

A matrix or vector of log-ratio coordinates

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a non-parametric compositional approach. In: Vera Pawlowsky-Glahn and Antonella Buccianti (eds.) Compositional Data Analysis: Theory and Applications, John Wiley \& Sons, Ltd, pp. 207-215

See Also

HWAlr,HWIlr

Examples

```
X <- HWData(100,100)$Xt
Y <- HWClr(X)
```

HWClrPlot *Plot genetic markers in centred log-ratio coordinates*

Description

HWClrPlot creates a scatter plot of the centred log-ratio coordinates of bi-allelic genetic markers. Hardy-Weinberg equilibrium is indicated by a straigh line in the plot.

Usage

```
HWClrPlot(X, zeroadj = 0.5)
```

Arguments

X	A matrix of genotype counts (columns AA, AB, BB)
zeroadj	Zero-adjustment parameter. Zero counts in the count matrix are substituted by zeroadj which is 0.5 by default.

Value

NULL

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a non-parametric compositional approach. In: Vera Pawlowsky-Glahn and Antonella Buccianti (eds.) Compositional Data Analysis: Theory and Applications, John Wiley \& Sons, Ltd, pp. 207-215

See Also

HWAlrPlot,HWIlrPlot

Examples

```
X <- HWData(100,100)$Xt
HWClrPlot(X)
```

HWCondProbAB *Compute probability of a genotypic sample*

Description

Computes the probability of a particular genotypic sample given the allele count, sample size and number of heterozygotes.

Usage

```
HWCondProbAB(n, nA, nAB)
```

Arguments

n	n is the total sample size (total number of individuals)
nA	nA is the number of A alleles in the sample
nAB	nAB is the number of heterozygotes in the sample

Value

p	probability of the particular sample

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

HWExact

Examples

```
x <- c(298,489,213)
names(x) <- c("MM","MN","NN")
n <- sum(x)
nM <- 2*x[1]+x[2]
nMN <- x[2]
p <- HWCondProbAB(n,nM,nMN)
```

HWD *Compute disequilibrium statistic D*

Description

Function HWD computes Weir's disequilibrium coefficient D.

Usage

```
HWD(X)
```

Arguments

X	a vector of genotype counts (AA, AB, BB)

Value

Returns the disequilibrium coefficient

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter3.

See Also

HWf HWChisq

Examples

```
x <- c(MM=298,MN=489,NN=213)
D <- HWD(x)
cat("Disequilibrium coefficient: ",D,"\n")
```

HWData *Generate genetic marker data in or out of Hardy-Weinberg Equilibrium*

Description

HWData generates samples of genotypic counts under various schemes. It mainly uses sampling from the multinomial distribution given Hardy-Weinberg allele frequencies.

Usage

```
HWData(nm = 100, n = rep(100, nm), f = rep(0, nm), p = runif(nm),
       pfixed = FALSE, exactequilibrium = FALSE, pdist = "runif", ...)
```

Arguments

nm	the number of markers (or samples).
n	the sample size.
f	the inbreeding coefficient
p	a vector of allele frequencies
pfixed	if TRUE Haldane's distribution is used for sampling, if FALSE a multinomial distribution is used
exactequilibrium	generates data in exact HWE if set to TRUE
pdist	take a random allele frequency from a uniform or beta distribution of pfixed = FALSE and p is not given.
...	specific parameters for the uniform or beta

Value

Xt	the genotypic counts.
Xc	the genotypic compositions.

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

HWTernaryPlot

Examples

```
nm <- 100
n <- 100
out <- HWData(nm,n)
```

HWExact *Exact test for Hardy-Weinberg equilibrium*

Description

HWExact performs an exact test for Hardy-Weinberg equilibrium

Usage

```
HWExact(X, alternative = "two.sided", pvaluetype = "selome", verbose = FALSE)
```

Arguments

X	vector with the genotype counts AA, AB, BB
alternative	two.sided (default) will perform a two-sided test where both an excess and a dearth of heterozygotes count as evidence against HWE. less is a one-sided test where only dearth of heterozygotes counts a evidence against HWE, greater is a one-sided test where only excess of heterozygotes counts as evidence against HWE.
pvaluetype	if pvaluetype is set to dost then the p-value of a two-sided test is computed as twice the tail area of a one-sided test. When set to selome, the p-value is computed as the sum of the probabilities of all samples less or equally likely as the current sample. When set to midp, the p-value is computed as half the probability of the current sample + the probabilities of all samples that are more extreme.
verbose	print results or not.

Details

HWExact uses the recursion equations described by Wigginton et. al.

For large samples, HWExact may give the error message: "evaluation nested too deeply: infinite recursion". This can usually be resolved by increasing R's limit on nested expressions with options(expressions=10000) or a higher limit. With higher limits, the error message "protect(): protection stack overflow" can occur. This error can usually be resolved by increasing R's protection stack with the command line option --max-ppsize 100000 or higer values. However, with such large samples the exact test will give virtually the same result as a chi-square test, and it may be easier to use HWChisq in these circumstances.

Value

pval	p-value of the exact test
prob	probabilities of all possible samples with the same sample size and minor allele count
pofthesample	probability of the observed sample

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter3.

Wigginton, J.E., Cutler, D.J. and Abecasis, G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium, American Journal of Human Genetics (76) pp. 887-893.

See Also

HWLratio, HWChisq

Examples

```
x <- c(298,489,213)
names(x) <- c("MM","MN","NN")
HW.test <- HWExact(x)
print(HW.test)
```

HWExactMat *Matrix version of HWExact*

Description

HWExactMat executes a fast Exact test for HWE for each row in a matrix.

Usage

```
HWExactMat(X, ...)
```

Arguments

X	A n times 3 matrix of genotypic counts (AA,AB,BB)
...	extra arguments that are passed on to HWExact

Value

pvalvec	Vector with the p-values of each test

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

HWExact

Examples

```
X <- HWData(100,10)$Xt
colnames(X) <- c("MM","MN","NN")
Results <- HWExactMat(X)
Output <- cbind(X,Results$pvalvec)
print(Output)
```

HWf *Computation of inbreeding coefficient*

Description

HWf computes the inbreeding coefficient for a sample of genotypes.

Usage

```
HWf(X)
```

Arguments

X	a vector of genotype counts (AA, AB, BB)

Details

For monomorphic markers a warning is issued, and the estimate for the inbreeding coefficient is set to zero.

Value

Returns the inbreeding coefficient (intraclass correlation coefficient)

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Crow, J. F. and Kimura, M. (1970) An introduction to population genetics theory. Harper \& Row, publishers, New York

See Also

HWChisq

Examples

```
x <- c(MM=298,MN=489,NN=213)
f <- HWf(x)
cat("Inbreeding coefficient: ",f,"\n")
```

HWGenotypePlot *Scatter plot of the genotype frequencies*

Description

HWGenotypePlot makes a scatterplots of the AB or BB frequency versus the AA frequency and represents a blue curve indicating the Hardy-Weinberg equilibrium condition.

Usage

```
HWGenotypePlot(X, plottype = 1, xlab = expression(f[AA]), ylab =
ifelse(plottype == 1, expression(f[AB]), expression(f[BB])), asp = 1,
pch = 19, xlim = c(0, 1), ylim = c(0, 1), cex = 1, cex.axis = 2, cex.lab = 2, ...)
```

Arguments

X	A matrix of genotype counts or frequencies with three columns (AA, AB, BB)
plottype	plottype=1 produces a plot of AB versus AA, plottype=2 produced a plot of BB versus AA.
xlab	A label for the x axis
ylab	A label for the y axis
asp	Aspec ratio (1 by default)
pch	Plotting charachter (19 by default)
xlim	Limits for the x axis (0-1 by default)
ylim	Limits for the y axis (0-1 by default)
cex	Character expansion factor (1 by default)
cex.axis	Character expansion factor for the axes (2 by default)
cex.lab	Character expansion factor for labels of axis (2 by default)
...	Additional arguments for the plot function

Value

NULL

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

HWTernaryPlot

Examples

```
n <- 100 # sample size
m <- 100 # number of markers
out <- HWData(n,m)
Xc <- out$Xc
HWGenotypePlot(Xc,plottype=1,main="Heterozygote-homozygote scatterplot")
```

HWIlr *Compute isometric log ratio coordinates.*

Description

HWIlr computes isometric log ratio coordinates for genotypic compositions (AA, AB, BB)

Usage

```
HWIlr(X, zeroadj = 0.5)
```

Arguments

X	A matrix of genotype counts, markers in rows, counts for AA, AB and BB in three columns
zeroadj	Adjustment for zeros (0.5 by defaults)

Value

A matrix of log ratio coordinates.

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barcelo-Vidal, C. (2003) Isometric Logratio Transformations for Compositional Data Analysis. Mathematical Geology 35(3), pp. 279-300.

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a non-parametric compositional approach. In: Vera Pawlowsky-Glahn and Antonella Buccianti (eds.) Compositional Data Analysis: Theory and Applications, John Wiley \& Sons, Ltd, pp. 207-215

See Also

HWAlr,HWClr

Examples

```
X <- HWData(100,100)$Xt
Y <- HWIlr(X)
```

HWIlrPlot *Plot bi-allelic genetic markers in isometric log ratio coordinates*

Description

HWIlrPlot makes a scatter plot of the isometric log ratio coordinates for bia-llelic markers.

Usage

```
HWIlrPlot(X, zeroadj = 0.5, ...)
```

Arguments

X	Matrix of genotype counts, one marker per row, AA, AB and BB in three columns
zeroadj	Adjustment for zero values (0.5 by default)
...	Additional arguments for function plot

Value

A matrix of log ratio coordinates.

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a nonparametric compositional approach. In Pawlowsky-Glahn, V. and Buccianti A., editors, Compositional Data Analysis: Theory and Applications, pages 208-217, John Wiley & Sons, Ltd.

See Also

HWAlrPlot,HWClrPlot

Examples

```
X <- HWData(100,100)$Xt
HWIlrPlot(X)
```

HWLratio *Likelihood ratio test for Hardy Weinberg equilibrium*

Description

HWLratio performs the Likelihood ratio test for Hardy Weinberg equilibrium.

Usage

```
HWLratio(X, verbose = FALSE)
```

Arguments

X	X a vector containing the genotypic counts (AA,AB,BB).
verbose	verbose = 1 prints results, verbose = 0 is silent.

Value

HWLratio returns a list with the components:

Lambda	the likelihood ratio
G2	-2*log(Lambda)
pval	the p-value

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter 3.

See Also

HWChisq

Examples

```
x <- c(298,489,213)
names(x) <- c("MM","MN","NN")
HW.test <- HWLratio(x,verbose=TRUE)
```

HWMissing *Test a bi-allelic marker for Hardy-Weinberg equilibrium in the presence of missing genotype information.*

Description

Function HWMissing imputes missing genotype data with a multinomial logit model that uses information from allele intensities and/or neighbouring markers. Multiple imputation algorithms implemented in the Mice package are used to obtain imputed data sets. Inference for HWE is carried out by estimating the inbreeding coefficient for each imputed data set, and by combining all estimates using Rubin's pooling rules.

Usage

```
HWMissing(X,imputecolumn=1,m=50,verbose=FALSE,alpha=0.05,
          varest="oneovern",return.imputed.sets=FALSE,...)
```

Arguments

X	An input data frame. By default, the first column should contain the SNP with missing values.
imputecolumn	Indicates which column of the supplied data frame is to be imputed (by default, the first colum, imputecolumn=1
m	The number of imputations (50 by default)
verbose	verbose = TRUE prints results, verbose = FALSE is silent.
alpha	significance level (0.05 by default) used when computing confidence intervals
varest	Estimator for the variance of the inbreeding coefficient. varest="oneovern" is the default and sets the variance under the null (1/n). varest="bailey" uses an approximation (see details).
return.imputed.sets	If return.imputed.sets=TRUE then the imputed datasets are returned as a list object.
...	additional options for function mice of the Mice package

Details

The function HWMissing tests one genetic marker (e.g. a SNP) with missings for HWE. By default, this marker is supposed to be the first column of dataframe X. The other columns of X contain covariates to be used in the imputation model. Covariates will typically be other, correlated markers or allele intensities of the SNP to be imputed. Covariate markers should be coded as factor variables whereas allele intensities should be numerical variables. By default, a polytomous regression model will be used to impute the missings. If the covariates also contain missings, an imputation method for each column of X can be specified by using the method of mice (see example below).

If there are no covariates, missings can be imputed under the MCAR assumption. In that case, missings are imputed by taking a random sample from the observed data. This is what `HWMissing` will do if no covariates are supplied, `X` being a single factor variable.

Several estimators for the variance of the inbreeding coefficient have been described in the literature. The asymptotic variance of the inbreeding coefficient under the null hypothesis is 1/n, and is used if `varest = "oneovern"` is used. This is the recommended option. Alternatively, the approximation described in Weir (p. 66) can be used with `varest = "bailey"`.

Value

`Res`	A vector with the inbreeding coefficient, a confidence interval for the inbreeding coefficient, a p-value for a HWE test and missing data statistics.
`Xmat`	A matrix with the genotypic composition of each of the `m` imputed data sets.
`ImputedSets`	A list object with all `m` imputed data sets. Only returned if `return.imputed.sets=TRUE`

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Little, R. J. A. and Rubin, D. B. (2002) Statistical analysis with missing data. Second edition, New York, John Wiley & sons.

Weir, B. S. (1996) Genetic Data Analysis II, Sinauer Associates, Massachusetts

See Also

`HWChisq`

Examples

```
data(Markers)
## Not run:
set.seed(123)
Results <- HWMissing(Markers[,1],m=50,verbose=TRUE)$Res # no covariates, imputation assuming MCAR.
set.seed(123)
Results <- HWMissing(Markers[,1:3],m=50,verbose=TRUE)$Res # impute with two allele intensities.
set.seed(123)
Results <- HWMissing(Markers[,c(1,4,5)],m=50,verbose=TRUE)$Res # impute with two covariate SNPs

## End(Not run)
```

HWPerm *Permutation test for Hardy-Weinberg equilibrium*

Description

Function HWPerm does a permutation test for Hardy-Weinberg equilibrium using a user-supplied test statistic.

Usage

```
HWPerm(x, nperm = 17000, verbose = TRUE, FUN = Chisquare, ...)
```

Arguments

x	A vector of genotype counts (AA,AB,BB)
nperm	The number of permutations
verbose	verbose = TRUE will print results, verbose = FALSE is silent.
FUN	An function call for calculating the test statistic for HWE (see examples below)
...	Additional parameters for the function call argument FUN

Details

The set of alleles for the observed sample is permuted. Consequently, the test is conditional on allele frequency.

Value

Returns the p-value of the test

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Ziegler, A. & K\"onig, I.R. (2006) A statistical approach to genetic epidemiology. Wiley.

See Also

HWChisq,HWExact,HWLratio

Examples

```
x <- c(MM=298,MN=489,NN=213)
## Not run:
HW.test <- HWPerm(x,nperm=10000,verbose=TRUE) # uses default chi-square statistic
HW.test <- HWPerm(x,nperm=10000,verbose=TRUE,function(z)
HWChisq(z)$chisq,cc=0.5) # uses chi-square statistic with continuity correction.
HW.test <- HWPerm(x,nperm=10000,verbose=TRUE,function(y) HWLratio(y)$G2)
# uses likelihood ratio statistic.
HWPerm(x,nperm=10000,verbose=TRUE,function(y) 1-HWExact(y)$pval) # uses
exact test p-value

## End(Not run)
```

HWPower *Compute the power of a test for Hardy-Weinberg equilibrium.*

Description

HWPower is a function that computes the power of a test for Hardy-Weinberg equilibrium.

Usage

```
HWPower(n = 100, nA = 100, pA = 0.5, y = c(AA=25,AB=50,BB=25),
alpha = 0.05, theta = 4, f = NULL, test = "exact",
alternative = "two.sided", pvaluetype = "selome", cc = 0.5)
```

Arguments

n	The sample size
nA	The minor allele count
pA	The minor allele frequency
y	A sample of genotype counts (AA,AB,BB)
alpha	The significance level (0.05 by default)
theta	The degree of disequilibrium (theta = 4 is equilibrium, theta > 4 is heterozygote excess, theta < 4 is heterozygote dearth)
f	The inbreeding coefficient. Overrules theta if specified.
test	The type of test for which power is to be computed. Can be "exact" (default) or "chisq" (chi-square)
alternative	The nature of the alternative hypothesis ("two.sided" (default), "greater" or "less")
pvaluetype	The type of p-value used in an exact test ("selome", "dost" or "midp")
cc	Continuity correction parameter for the chi-square test (0.5 by default)

Details

HWPower uses the Levene-Haldane distribution (distribution of the number of heterzygotes given the minor allele count) for computing power.

HWPower can be used in three different way. In principle, the power is calcuted on the basis of the sample size (n) and the minor allele count (nA). Alternatively, the user may specifiy sample size (n) and minor allele frequency (pA). Finally, power can also be calculated directly from a sample of genotype counts. In that case the calculated power is the power for a sample of the given sample size and minor allele count. The three ways to use HWPower are illustrated in the example section.

Value

if test = "exact" the power of the exact test is computed for the given significance level and minor allele count.

if test = "chisq" the power of the chi-square test is computed for the given significance level and minor allele count.

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

References

Graffelman, J. and Morales, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* 65(2):77-84.

See Also

HWExact

Examples

```
pw.chisq <- HWPower(n=100,nA=100,alpha=0.05,test="chisq",theta=16)
print(pw.chisq)
pw.exact <- HWPower(n=100,nA=100,alpha=0.05,test="exact",theta=16,pvaluetype="selome")
print(pw.exact)
pw.exact <- HWPower(n=100,nA=100)
print(pw.exact)
pw.exact <- HWPower(n=100,pA=0.5)
print(pw.exact)
pw.exact <- HWPower(y=c(AA=25,AB=50,BB=25))
print(pw.exact)
```

HWQqplot *A Q-Q plot for Hardy-Weinberg equilibrium*

Description

HWQqplot creates a Q-Q plot for the p-values obtained in an Exact test for Hardy-Weinberg equilibrium. Empirical p-values are plotted against multiple simulated quantiles of the theoretical p-value distribution.

Usage

```
HWQqplot(X, nsim = 100, fit = "curve", logplot = FALSE,
main = "Q-Q plot for HWE", mm = NULL, pvaluetype = "selome", ...)
```

Arguments

X	Data matrix with genotype counts, one row for each sample, 3 columns
nsim	Number of samples drawn from the null distribution (100 by default)
fit	If fit is set to "line" straight lines will be fitted to the simulated samples, if set to "curve", ascending curves will be shown.
logplot	If logplot is set to true, then the log10 of the p-values will be used in the plot. If not, untransformed p-values will be used.
main	Title for the plot
mm	Maximal value for x and y axis in the plot
pvaluetype	Type of p-value to be used in an exact test. Can be "selome" (default), "midp" or "dost".
...	Any additional arguments for the plot instruction

Details

HWQqplot constructs a Q-Q plot of the p-values of an exact test for Hardy-Weinberg equilibrium. Under the null, this p-value is not uniform. HWQqplot samples from the theoretical null distribution, taking into account that markers may vary in allele frequency and in sample size (due to missing values). For each simulated sample a grey curve or line is shown. A green reference line with intercept 0 and slope 1 is also shown in the plot.

Value

NULL

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Rohlfs, R.V. and Weir, B.S. (2008) Distributions of Hardy-Weinberg equilibrium test statistics. Genetics 180, pp. 1609-1616.

See Also

HWTernaryPlot HWExact qqplot

Examples

```
set.seed(1234)
n <- 200 # sample size
m <- 100 # number of markers
X <- HWData(n,m)$Xt
HWQqplot(X,logplot=TRUE,pvaluetype="selome",main="Q-Q Plot for HWE")
```

HWTernaryPlot *Ternary plot with the Hardy-Weinberg acceptance region*

Description

HWTernaryPlot is a routine that draws a ternary plot for three-way genotypic compositions (AA,AB,BB), and represents the acceptance region for different tests for Hardy-Weinberg equilibrium (HWE) in the plot. This allows for graphical testing of a large set of markers (e.g. SNPs) for HWE. The (non) significance of the test for HWE can be inferred from the position of the marker in the ternary plot. Different statistical tests for HWE can be done graphically with this routine: the ordinary chisquare test, the chisquare test with continuity correction and the Haldane's exact test.

Usage

```
HWTernaryPlot(X, n = NA, addmarkers = TRUE, newframe = TRUE, hwcurve = TRUE,
vbounds = TRUE, mafbounds = FALSE, mafvalue = 0.05, axis = 0, region = 1,
vertexlab = colnames(X), alpha = 0.05, vertex.cex = 1, pch = 19, cc = 0.5,
markercol = "black", markerbgcol = "black", cex = 0.75, axislab = "",
verbose = FALSE, markerlab = NULL, markerpos = NULL, mcex = 1, connect =
FALSE, curvecols = rep("black",5), signifcolour = TRUE, curtyp =
"solid", ssf = "max", pvaluetype = "dost", ...)
```

Arguments

X	a matrix of n genotypic compositions or counts. If it is a matrix of compositions, X should have (n rows that sum 1, and 3 columns, with the relative frequencies of AA, AB and BB respectively. Argument n should be supplied as well. If X is a matrix of raw genotypic counts, it should have 3 columns with the absolute counts of AA, AB and BB respectively. Argument n may be supplied and will be used for painting acceptance regions. If not supplied n is computed from the data in X.

n	the samples size (for a complete composition with no missing data).
addmarkers	represent markers by dots in the triangle (addmarkers=TRUE) or not (addmarkers=FALSE).
newframe	allows for plotting additional markers in an already existing ternary plot. Overplotting is achieved by setting newframe to FALSE. Setting newframe = TRUE (default) will create a new ternary plot.
hwcurve	draw the HW parabola in the plot (hwcurve=TRUE) or not (hwcurve=FALSE).
vbounds	indicate the area corresponding to expected counts > 5 (vbounds=TRUE) or not (vbounds=FALSE).
mafbounds	indicate the area corresponding to MAF < mafvalue.
mafvalue	a critical value for the minor allele frequency (MAF).
axis	draw a vertex axis 0 = no axis is drawn 1 = draw the AA axis 2 = draw the AB axis 3 = draw the BB axis
region	the type of acceptance region to be delimited in the triangle 0 = no acceptance region is drawn 1 = draw the acceptance region corresponding to a Chi-square test 2 = draw the acceptance region corresponding to a Chi-square test with continuity correction 3 = draw the acceptance region corresponding to a Chi-square test with continuity correction for D > 0 4 = draw the acceptance region corresponding to a Chi-square test with continuity correction for D < 0 5 = draw the acceptance regions for all preceding tests simultaneously 6 = draw the acceptance region corresponding to a Chi-square test with continuity correction with the upper limit for D > 0 and the lower limit for D < 0 7 = draw the acceptance region corresponding to a two-sided exact test
vertexlab	labels for the three vertices of the triangle
alpha	significance level (0.05 by default)
vertex.cex	character expansion factor for the labels of the vertices of the triangle.
pch	the plotting character used to represent the markers.
cc	value for the continuity correction parameter (0.5 by default).
markercol	vector with colours for the marker points in the triangle.
markerbgcol	vector with background colours for the marker points in the triangle.
cex	expansion factor for the marker points in the triangle.
axislab	a label to be put under the horizontal axis.
verbose	print information on the numerically found cut-points between curves of the acceptance region and the edges of the triangle.
markerlab	labels for the markers in the triangle.

markerpos	positions for the marker labels in the triangle (1,2,3 or 4).
mcex	character expansion factor for the labels of the markers in the ternary plot.
connect	connect the represented markers by a line in the ternary plot.
curvecols	a vector with four colour specifications for the different curves that can be used to delimit the HW acceptance region. E.g. curvecols=c("red", "green","blue","black","pu will paint the Hardy-Weinberg curve red, the limits of the acceptance region for an ordinary chi-square test for HWE green, the limits of the acceptance region for a chi-square test with continuity correction when D > 0 blue and the limits of the acceptance region for a chi-square test with continuity correction when D < 0 black, and the limits of the exact acceptance region purple.
signifcolour	colour the marker points automatically according to the result of a signifance test (green markers non-siginficant, red markers significant). signifcolour only takes effect if region is set to 1, 2 or 7.
curtyp	style of the drawn curves ("dashed","solid","dotted",...)
ssf	sample size function ("max","min","mean","median",...). Indicates how the sample size for drawing acceptance regions is determined from the matrix of counts.
pvaluetype	method to compute p-values in an exact test ("dost" or "selome")
...	other arguments passed on to the plot function (e.g. main for a main title).

Details

HWTernaryPlot automatically colours significant markers in red, and non-significant markers in green if region is set to 1, 2 or 7.

Value

minp	minimum allele frequency above which testing for HWE is appropriate (expected counts exceeding 5).
maxp	maximum allele frequency below which testing for HWE is appropriate.
inrange	number of markers in the appropriate range.
percinrange	percentage of markers in the appropriate.
nsignif	number of significant markers (only if region equals 1,2 or 7.)

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Graffelman, J. and Morales, J. (2008) Graphical Tests for Hardy-Weinberg Equilibrium Based on the Ternary Plot. *Human Heredity* 65(2):77-84.

Graffelman, J. (2012) Exploring Bi-Allelelic Genetic Markers: The HardyWeinberg Package. *Journal of Statistical Software*

See Also

HWChisq

Examples

```
n <- 100 # sample size
m <- 100 # number of markers

out <- HWData(n,m)
Xc <- out$Xc

HWTernaryPlot(Xc,100,region=1,hwcurve=TRUE,vbounds=FALSE,vertex.cex=2)
```

---

ifisherz *Inverse Fisher z transformation*

Description

Calculates the inverse of Fisher's z transformation

Usage

```
ifisherz(y)
```

Arguments

y	a real number

Value

a correlation coefficient in the range (-1,1)

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

cor

Examples

```
r <- 0.5
print(ifisherz(fisherz(r)))
```

mac *Compute the minor allele count.*

Description

mac computes the smallest allele count for a given vector of genotype counts.

Usage

```
mac(X)
```

Arguments

X	a vector or matrix with genotype counts (AA, AB, BB)

Value

a vector of the minor allele counts

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

See Also

maf

Examples

```
X <- as.vector(rmultinom(1,100,c(0.5,0.4,0.1)))
names(X) <- c("AA","AB","BB")
print(X)
print(mac(X))
```

maf *Function to compute minor allele frequencies*

Description

Function maf computes the minor allele frequency for a matrix or vector of compositions.

Usage

```
maf(x)
```

Arguments

x	a vector or matrix of genotypic compositions

Value

a vector of minor allele frequencies.

Author(s)

Jan Graffelman (jan.graffelman@upc.edu)

Examples

```
X <- as.vector(rmultinom(1,100,c(0.5,0.4,0.1)))
X <- X/sum(X)
print(X)
print(maf(X))
```

MakeCounts *Create genotype counts from bi-allelic marker data*

Description

MakeCounts creates a matrix of genotype counts, with one row for each bi-allelic marker, containing 4 columns with the counts AA, AB, BB and NA (missings) respectively

Usage

```
MakeCounts(X, alleles, pos1 = 1, pos2 = 3, coding = c(AA=0,AB=1,BB=2))
```

Arguments

X	A matrix or dataframe with bi-allelic genotyping information, markers in columns, individuals in rows
alleles	a vector of alleles for each marker (e.g. c("A/T","A/G",...)). Only relevanit if X is a matrix with text entries.
pos1	position of the first allele in the allele string (1 by default)
pos2	position of the second allele in the allele string (3 by default)
coding	indicates how homozygotes and heterozygote are coded as numbers. Only relevant if X is a matrix with numeric entries.

Details

MakeCounts is thought for bi-allelic marker data only. Missings are should be coded by NA. It produces the right input for HWTernaryPlot.

Value

A matrix of 4 columns

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

HWTernaryPlot

Examples

```
SNP1 <- c("GG","GG","GG","GG","GG","GG","GG","GG","GG")
SNP2 <- c("CG","GG","CC","GG","GG","CG","CG","CG","CG")
SNP3 <- c("AA","AA","AA","AG","AA","AG","AA","AA","AA")
SNP4 <- c("GG","GG","GG","GG","GG","GG","GG","GG","GG")
SNP5 <- c("CC","CC","CC","CC","CC","CC","CT","CT","CT")
X <- cbind(SNP1,SNP2,SNP3,SNP4,SNP5)
Y <- MakeCounts(X,c("A/G","C/G","A/G","A/G","C/T"))
print(Y)
W <- matrix(sample(c(0,1,2,NA),100,replace=TRUE),ncol=5)
Z <- MakeCounts(W,coding=c(0,1,2))
```

Markers *SNP data and intensities*

Description

The dataframe contains the genotypes of 3 SNPs and two allele intensities of 146 individuals. The first column is a GT polymorphism that has missing values for several individuals. The second and third column (iG and iG) are the allele intensities of this polymorphism. Column 4 and 5 are covariate SNPs (an AC and an AG polymorphism) that have no missing values.

Usage

```
data(Markers)
```

Format

A data frame containing 146 rows and 5 columns

References

Graffelman, J. (2013) Exploring bi-allelelic genetic markers: the HardyWeinberg package. Journal of Statistical Software.

Mourant — *Genotype frequencies for blood group locus MN*

Description

The dataframe contains the genotype frequencies MM, MN and NN for the MN blood group locus for 216 populations. The data are taken from table 2.5 in Mourant et al., using only entries with a sample size of at least 500.

Usage

```
data(Mourant)
```

Format

A data frame containing 216 observations.

Source

Mourant et al, Table 2.5

References

Mourant, A. E. and Kope\'c, A. C. and Domaniewska-Sobczak, K. (1976) The Distribution of the Human Blood Groups and other Polymorphisms. Second edition. Oxford University Press, London.

recode — *Recode genotype information*

Description

function `recode` recodes bi-allelic genetic marker information expressed as strings (e.g. "AA", "AB", "BB") into numerical form.

Usage

```
recode(X, alleles, values = c(0,1,2), pos1 = 1, pos2 = 3)
```

Arguments

`X`	A matrix or dataframe of bi-allelic markers, individuals in rows, markers in columns
`alleles`	a vector with the alleles for each marker (e.g. c("A/T", "A/G", etc))
`values`	a vector of numerica values for AA, AB and BB, (c(0,1,2) by default
`pos1`	position of the first allele in the allele string (1 by default)
`pos2`	position of the second allele in the allele string (3 by default)

Details

recode is written for bi-allelic marker data only.

Value

A numerical matrix, individuals in rows, markers in columns

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

MakeCounts

Examples

```
SNP1 <- c("GG","GG","GG","GG","GG","GG","GG","GG","GG")
SNP2 <- c("CG","GG","CC","GG","GG","CG","CG","CG","CG")
SNP3 <- c("AA","AA","AA","AG","AA","AG","AA","AA","AA")
SNP4 <- c("GG","GG","GG","GG","GG","GG","GG","GG","GG")
SNP5 <- c("CC","CC","CC","CC","CC","CC","CT","CT","CT")
X <- cbind(SNP1,SNP2,SNP3,SNP4,SNP5)
Y <- recode(X,c("A/G","C/G","A/G","A/G","C/T"))
print(Y)
```

UniqueGenotypeCounts *Extract unique genotypic compositions from a matrix*

Description

Function UniqueGenotypeCounts creates a matrix containing only the unique rows in the given matrix, together with their frequency of occurrence

Usage

```
UniqueGenotypeCounts(X)
```

Arguments

X	A n by 3 matrix with genotypic counts (AA,AB,BB)

Value

A matrix with 4 columns, AA, AB, BB, and frequency of occurrence

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

See Also

GenerateSamples

Examples

```
set.seed(123)
X <- HWData(n=100,nm=100)$Xt
print(nrow(X))
Y <- UniqueGenotypeCounts(X)
print(nrow(Y))
print(sum(Y$w))
```

---

vaf *Computes the sample variance of the minor allele frequencies*

Description

Function vaf computes the sample variance of the minor allele frequencies of a single sample or a matrix of samples.

Usage

```
vaf(X, hw = FALSE, minor = TRUE)
```

Arguments

X	vector or matrix with genotype counts (AA,AB,BB)
hw	assume Hardy-Weinberg proportions (hw=TRUE) or not (hw=FALSE)
minor	compute the variance for the minor allele minor=TRUE or major allele major=TRUE

Value

a numeric vector of variances.

Author(s)

Jan Graffelman <jan.graffelman@upc.edu>

References

Weir, B.S. (1996) Genetic data analysis II. Sinauer Associates, Massachusetts. See Chapter 2.

See Also

af,maf

Examples

```
x <- c(MM=298,MN=489,NN=213)
pA <- af(x)
vA <- vaf(x)
cat("allele frequency:",pA,"\n")
cat("sample variance allele frequency:",vA,"\n")
```

Index