

Package ‘vartools’

January 3, 2015

Type Package

Title Variant Association Tools R-package

Version 1.1.0

Date 2015-01-03

Author I.Y. Zhbannikov

Maintainer I.Y. Zhbannikov <ilyaz@uidaho.edu>

Description

Computes statistics of rare variants using CMC, KBAC, VT and other association methods

License GPL

R topics documented:

vartools-package	2
asum	3
calpha	4
cmc	5
kbac	6
rarecover	7
rbt	9
skat	10
vt	11
wss	12
Index	14

vartools-package	<i>Variant Association tools</i>
------------------	----------------------------------

Description

vartools is an R-package for the analysis of rare variants and currently has the following methods: CMC, KBAC, RVE, VT, RareCover, RBT, C-alpha.

Installation

git clone <https://github.com/izhbannikov/vartools.git>
R CMD INSTALL vartools

Details

Package:	vartools
Type:	Package
Version:	1.0
Date:	2014-12-10
License:	GPL (>= 2)
Project URL:	https://github.com/izhbannikov/vartools

Author(s)

Ilya Y. Zhbannikov

Maintainer: Ilya Y. Zhbannikov | i.zhbannikov@mail.ru

Examples

```
#CMC test
# Load the package
library(vartools)
?cmc

pgdata <- as.matrix(read.table(system.file("extdata", "phengen.dat", package="vartools"), as.is=T, skip = 1))
cmc.pvalue <- cmc(table=pgdata)
print(cmc.pvalue)

#KBAC test
?kbac

alpha <- 0.05
num.permutation <- 3000
quiet <- 1
alternative <- 1
maf.upper.bound <- 0.05
casectrl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
kbac.pvalue <- kbac(table=casectrl.dat, alpha, num.permutation, quiet, maf.upper.bound, alternative)
print(kbac.pvalue)
```

asum	<i>aSum statistical test</i>
------	------------------------------

Description

aSum, Adaptive Sum Test (Han and Pan, 2010) is a statistical test which utilizes the difference in direction of effects (protective or deleterious) of rare variants within the same genetic region analyzed by a rare variant association test.

Usage

```
asum(table, perm=100)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
perm	positive integer that defines the number of permutations. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

asum.stat	aSum statistic
perm.pval	permuted p-value; only when permutations is used
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Fang Han and Wei Pan (2010) A Data-Adaptive Sum Test for Disease Association with Multiple Common or Rare Variants. Human Heredity doi:10.1159/000288704. <http://www.karger.com/doi/10.1159/000288704>

Examples

```
# Load the package

library(vartools)
?asum

casectl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
asum.stat <- asum(casectl.dat)
asum.stat
```

calpha

*C(alpha) statistical test***Description**

C(alpha) (Neale et al., 2011) is a statistical test, developed for disease traits, in order to test for the hypothesis of rare variants disease association. This test has an assumption that rare variants, which were observed in cases and controls and are a mix of various types of variants: phenotypically deleterious, protective and neutral variants.

Usage

```
calpha(table, permutations=NULL)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
permutations	positive integer that defines the number of permutations. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

calpha.stat	C(alpha) statistic
asym.pval	asymptotic p-value, distributed as Chi-square, with parameter degrees of freedom df=1
perm.pval	permuted p-value; only when perm is used
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Benjamin M. Neale, Manuel A. Rivas, Benjamin F. Voight, David Altshuler, Bernie Devlin, Marju Orho-Melander, Sekar Kathiresan, Shaun M. Purcell, Kathryn Roeder and Mark J. Daly (2011), Testing for an Unusual Distribution of Rare Variants. PLoS Genetics doi:10.1371/journal.pgen.1001322. <http://dx.plos.org/10.1371/journal.pgen.1001322>

Examples

```
# Load the package

library(vartools)
?calpha

casectl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
calpha.stat <- calpha(casectl.dat)
calpha.stat
```

cmc

CMC statistic implementation

Description

This program implements the CMC statistic in [Liu and Leal 2008]. It carries out case-control association testing for rare variants for whole exome association studies. Briefly, consider a gene of length n which harbors m rare variants. Genotype on the m variant sites & the disease status (case/control) are known for each individual. The program takes as input the m -site genotype and disease status (case/control) data files, and computes a p-value indicating the significance of association. In order to speed up permutation testing we use an "adaptive" approach to obtain p-values.

Usage

```
cmc(table, method = "fisher")
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'. DO NOT allow for missing data.
method	statistical method, Fisher test used by default

Details

....

Value

pvalue	the p-value of test.
--------	----------------------

Author(s)

Ilya Y. Zhbannikov | i.zhbannikov@mail.ru

References

Liu DJ, Leal SM, 2008 Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. The American Journal of Human Genetics, 83, DOI10.1016/j.ajhg.2008.06.024

Examples

```
# Load the package
library(vartools)
?cmc

pgdata <- as.matrix(read.table(system.file("extdata", "phengen.dat", package="vartools"), as.is=T, skip = 1))
cmc.pvalue <- cmc(table=pgdata)
print(cmc.pvalue)
```

kbac

KBAC statistic implementation

Description

This program implements the KBAC statistic in [Liu and Leal 2010]. It carries out case-control association testing for rare variants for whole exome association studies. Briefly, consider a gene of length n which harbors m rare variants. Genotype on the m variant sites & the disease status (case/control) are known for each individual. The program takes as input the m -site genotype and disease status (case/control) data files, and computes a p-value indicating the significance of association. In order to speed up permutation testing we use an "adaptive" approach to obtain p-values.

Usage

```
kbac(table, alpha = NULL, num.permutation = NULL, quiet = T, maf.upper.bound = 1.0, alternative = 1)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'. DO NOT allow for missing data.
alpha	size of test, or the significant level. This feature will be useful in adaptive p-value calculation. If you do not want to use adaptive p-value, set alpha = 999 (or any number greater than 1.0).
num.permutation	number of permutations for p-value calculation. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.
quiet	when quiet = 0 the screen output would contain a summary of the KBAC test; otherwise only the p-value will be printed on screen.
maf.upper.bound	The upper bound of the MAF to be included in analysis. MAF is calculated based on observed sample. This can be arbitrary although it is usually defined as 0.01 for analysis of rare variants.
alternative	Set alternative = 1 for test of deleterious variants, = 2 for test of both deleterious and protective variants. Please note that this is different from the "one/two-sided" definition in the KBAC paper.

Details

....

Value

pvalue the p-value of test.

Author(s)

Gao Wang | wangow@gmail.com

References

Liu DJ, Leal SM, 2010 A Novel Adaptive Method for the Analysis of Next-Generation Sequencing Data to Detect Complex Trait Associations with Rare Variants Due to Gene Main Effects and Interactions. PLoS Genet 6(10): e1001156. doi:10.1371/journal.pgen.1001156

Examples

```
# Load the package
library(vartools)
?kbac

# Set parameters and use the kbac() function to obtain p-value
alpha <- 0.05
num.permutation <- 3000
quiet <- 1
alternative <- 1
maf.upper.bound <- 0.05
casectrl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
kbac.pvalue <- kbac(table=casectrl.dat, alpha, num.permutation, quiet, maf.upper.bound, alternative)
print(kbac.pvalue)

# To evaluate test at small alpha we need huge number of permutations. Adaptive approach is thus necessary.
casectrl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
kbac.pvalue <- kbac(table=casectrl.dat, 0.00001, 1000000, quiet, maf.upper.bound, alternative)
print(kbac.pvalue)

# Not using adaptive p-value calculation; will take longer time
casectrl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
kbac.pvalue <- kbac(table=casectrl.dat, 9, 1000000, quiet, maf.upper.bound, alternative)
print(kbac.pvalue)
```

rarecover

*RareCover statistical test***Description**

Testing for rare variants with RareCover algorithm. This algorithm is similar to CMC, meaning that it follows its collapsing strategy, but uses greedy algorithm to find an optimized combination of variants in a loci for which its association signal is strongest.

Usage

```
rarecover(table, maf=0.05, dif=0.5, perm=250)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
maf	numeric value indicating the minor allele frequency threshold for rare variants
dif	numeric value between 0 and 1 as a threshold for the decision criterion in the RareCover algorithm
perm	positive integer that defines the number of permutations. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

rc.stat	RareCover statistic
perm.pval	permuted p-value
set	set of selected variants
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Bhatia G, Bansal V, Harismendy O, Schork NJ, Topol EJ, Frazer K, Bafna V (2010) A Covering Method for Detecting Genetic Associations between Rare Variants and Common Phenotypes. *PLoS Computational Biology*, **6(10)**: e1000954

Examples

```
# Load the package

library(vartools)
?rarecover

casectl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
rarecover.stat <- rarecover(casectl.dat)
rarecover.stat
```

rbt	<i>Replication-Based statistical test</i>
-----	---

Description

Implementation of the replication base test (RBT) proposed by Ionita-Laza et al., 2011. Replication means that it computes evidences to reject each of the two hypothesis: (1) Deleterious rare variants are enriched in cases; (2) Protective rare variants are enriched in controls.

Usage

```
rbt(table, perm=150)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
perm	positive integer that defines the number of permutations, 150 by default. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

rbt.stat	RBT statistic
perm.pval	permuted p-value
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Ionita-Laza I, Buxbaum JD, Laird NM, Lange C (2011) A New Testing Strategy to Identify Rare Variants with Either risk or Protective Effects on Disease. *PLoS Genetics*, **7(2)**: e1001289

Examples

```
# Load the package

library(vartools)
?rbt

casectl1.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
rbt.stat <- rbt(casectl1.dat)
rbt.stat
```

skat	<i>SKAT statistical test</i>
------	------------------------------

Description

SKAT (Wu et al., 2010, 2011) is a statistical test, developed for disease traits, in order to test for the hypothesis of rare variants disease association. This is a score-based variance-component test.

Usage

```
skat(table, kernel="linear", weights=NULL, a=1, b=25, permutations=NULL)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
kernel	Kernel function used. By default, a linear function is used. Other available kernels are: "wlinear", "quadratic", "IBS", "wIBS", "twowayx"
weights	a set of numeric weights for genetic variants.
a	a positive numeric value for the parameter a in the Beta distribution (a=1 by default)
b	a positive numeric value for the parameter b in the Beta distribution (b=1 by default)
permutations	positive integer that defines the number of permutations. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

skat.stat	SKAT statistic
asym.pval	asymptotic p-value, distributed as Chi-square, with parameter degrees of freedom df=1
perm.pval	permuted p-value; only when permutations is used
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X (2010) Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies. *The American Journal of Human Genetics*, **86**: 929-942

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X (2011) Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *The American Journal of Human Genetics*, **89**: 82-93

Examples

```
# Load the package

library(vartools)
?skat

casectl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
skat.stat <- skat(casectl.dat)
skat.stat
```

vt	<i>Variable Threshold statistical test</i>
----	--

Description

This is implementation for the Variable Threshold statistical test. It assigns different weights to rare variants base on their frequencies.

Usage

```
vt(table, maf=0.05, perm=50)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
maf	numeric value indicating the minor allele frequency threshold for rare variants (must be a positive number between 0 and 1, maf=0.05 by default).
perm	positive integer that defines the number of permutations, perm=50 by default. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

vt.stat	VT statistic
perm.pval	permuted p-value
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Price AL, Kryukov GV, de Bakker PIW, Purcell SM, Staples J, Wei LJ, Sunyaev SR (2010) Pooled Association Tests for Rare Variants in Exon-Sequencing Studies. *The American Journal of Human Genetics*, **86**: 832-838

Examples

```
# Load the package
library(vartools)
?vt

casectl.dat <- read.table(system.file("extdata", "phengen2.dat", package="vartools"), skip = 1)
vt.stat <- vt(casectl.dat)
vt.stat
```

WSS	WSS statistical test
-----	----------------------

Description

WSS (Madsen and Browning, 2009) is a statistical test, developed for disease traits, in order to test for the hypothesis of rare variants disease association. WSS test introduces the method of assigning "weights" to rare variants found in a genetic region before they are collapsed.

Usage

```
wss(table, perm=100)
```

Arguments

table	a numeric matrix with first column having disease status '0' or '1' and the rest columns codes the locus genotype as '0', '1', and '2'.
perm	positive integer that defines the number of permutations. In permutation test, the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points.

Details

....

Value

A list with the following elements:

wss.stat	WSS statistic
perm.pval	permuted p-value; only when permutations is used
args	descriptive information with number of controls, cases, variants, and permutations
name	name of the statistic

References

Bo Eskerod Madsen and Sharon R. Browning (2009) A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. PLoS Genetics doi:10.1371/journal.pgen.1000384. <http://dx.plos.org/10.1371/journal.pgen.1000384>

Examples

```
# Load the package

library(vartools)
?wss

casectl.dat <- read.table(system.file("extdata", "phengen.dat", package="vartools"), skip = 1)
wss.stat <- wss(casectl.dat)
wss.stat
```

Index

*Topic **package**

vartools-package, [2](#)

asum, [3](#)

calpha, [4](#)

cmc, [5](#)

kbac, [6](#)

rarecover, [7](#)

rbt, [9](#)

skat, [10](#)

vartools (vartools-package), [2](#)

vartools-package, [2](#)

vt, [11](#)

wss, [12](#)