Package 'vartools'

December 10, 2014

Type Package

Title Variant Association Tools R-package

Version 1.0

Date 2014-12-10

Author I.Y. Zhbannikov

Maintainer I.Y. Zhbannikov, i.zhbannikov@mail.ru, ilyaz@uidaho.edu

Description

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

License GPL

R topics documented:

`vartools-package` *Variant Association tools*

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)
```

---

calpha *C(alpha) statistical test*

Description

C(alpha) (Neale at 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, perm=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'.
perm	positive integer that defines the number of permutations

Details

....

Value

A list with the following elements:

calpha.stat	C(alpha) statistic

asym.pval

asymptotic p-value

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

Author(s)

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

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

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

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

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

`rbt` *Replication-Based statistical test*

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.

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

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

---

rve *RVE statistic implementation*

Description

This program implements the RVE statistic in [Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, et al. (2004)]. The RVE was first introduced in the analysis of seuence data from Dallas Heart Study. Performs a one-sided Wilcoxon's two-sample test.

Usage

```
rve(x,y)
```

Arguments

x	numeric vector of data values, DO NOT allow for missing data (will be omitted).
y	an optional numeric vector of data values, missing data will be omitted.

Details

....

Value

A list with class "htest" containing the following components:

statistic	the value of the test statistic with a name describing it.
parameter	the parameter(s) for the exact distribution of the test statistic.
p.value	the p-value for the test.
null.value	the location parameter mu.
alternative	a character string describing the alternative hypothesis.
method	the type of test applied.
data.name	a character string giving the names of the data.
conf.int	a confidence interval for the location parameter. (Only present if argument conf.int = TRUE.)
estimate	an estimate of the location parameter. (Only present if argument conf.int = TRUE.)

Author(s)

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

References

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, et al. (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305: 869-872.

Examples

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

x <- c(0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46)
y <- c(1.15, 0.88, 0.90, 0.74, 1.21)
rve.stat <- rve(x, y)
rve.stat
```

vt *Variable Threshold statistical test*

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.

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

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

Index