Package ‘RVtests’

February 19, 2015

Type Package
Title Rare Variant Tests
Version 1.2
Date 2013-05-27
Author C. Xu, and C. M. Greenwood
Maintainer C. Xu <changjiang.xu@mail.mcgill.ca>
Depends R (>= 2.12.1), glmnet, spls, pls
Description Use multiple regression methods to test rare variants association with disease traits.
License GPL (>= 2)
LazyLoad yes
NeedsCompilation no
Repository CRAN
Date/Publication 2013-05-28 07:40:45

R topics documented:

RVtests-package ......................................................... 2
count2geno ......................................................... 3
geno2count .......................................................... 3
LASSO ................................................................. 4
PCR ................................................................. 5
PLS ................................................................. 7
RR ................................................................. 8
sample.egeno ......................................................... 9
SPLS ................................................................. 10
VTWOD ............................................................. 11

Index 13
RVtests-package

**Rare Variants Tests**

**Description**

Use multiple regression methods to test rare variants association with disease traits.

**Details**

- **Package:** RVtests
- **Type:** Package
- **Version:** 1.2
- **Date:** 2013-05-27
- **License:** GLP 2.0 or greater
- **LazyLoad:** yes

An overview of how to use the package, including the most important functions

**Author(s)**

C. Xu, C. M. Greenwood, Maintainer: <changjiang.xu@mail.mcgill.ca>

**References**


**Examples**

```r
data(sample.cgeno)
str(sample.cgeno)
x=count2cgeno(sample.cgeno$cgeno)
dim(x)

set.seed(31018)
y=rowSums(x[,2:4]*rep(rnorm(3,1,0.1), each=nrow(x))) + 0.4*rnorm(nrow(x))

tmp<- proc.time();RR(x,y,lambda=0.5); proc.time()-tmp

tmp<- proc.time();RR(x,y,weights=c(rep(2,10), rep(1, ncol(x)-10)), lambda=0.5); proc.time()-tmp

tmp<- proc.time();RR(x,y,weights=c(rep(1,10), rep(0, ncol(x)-10)), lambda=0.5); proc.time()-tmp
```
**count2geno**

*Transforming genotype counts to genotype codes*

**Description**
Transform genotype counts data format to genotype codes format.

**Usage**
```r
count2geno(cgeno, indid)
```

**Arguments**
- `cgeno`: A matrix or data frame with 3 columns: indid (individual IDs), snpid (SNP IDs), and count
- `indid`: Individuals ID, including indid in cgeno

**Value**
A matrix of genotypes

**Author(s)**
C. Xu

**See Also**
- `geno2count`

---

**geno2count**

*Transforming genotype codes matrix to genotype counts*

**Description**
Genotype counts

**Usage**
```r
geno2count(genotype)
```

**Arguments**
- `genotype`: Genotype matrix or data frame with row and column names, each row as an individual and each column as a snp
Value
Data frame of genotype counts with 3 columns: indid (individual IDs), snpid (SNP IDs), and count

Author(s)
C. Xu

See Also
count2geno

Description
Use LASSO for selecting significant variants and testing the variants associated with disease traits.

Usage
LASSO(x, y, family = c("gaussian", "binomial", "poisson", "multinomial", "cox"),
alpha = 1, nlambda = 100, lambda.min.ratio, standardize = TRUE,
size.max, a = 2, npermutation = 0, npermutation.max, min.nonsignificant.counts)

Arguments
x Genotype matrix, each row as an individual and each column as a snp
y Phenotype vector
family Family: gaussian, binomial, poisson, multinomial, and cox
alpha alpha = 1 for LASSO, see glmnet
nlambda see glmnet
lambda.min.ratio see glmnet
standardize see glmnet
size.max Maximum number of variants included
a Penalty parameter for information criterion, a=2 for AIC.
npermutation Number of permutation, if less than 1, the permutation will not be run.
npermutation.max Maximum permutation
min.nonsignificant.counts Minimum nonsignificant counts
Details

Use glmnet package to implement LASSO and an information criterion (AIC, BIC, or GIC) to select a set of variants.

Value

nonsignificant.counts
Counts of permuted data that have a higher score than unpermuted data.

pvalue.empirical
Empirical pvalue via permutation

pvalue.nominal
Not available

vs
The selected variants

total.permutation
Total permutation

family
Family

Author(s)

C. Xu

References


See Also

SPLS, glmnet

Description

Use principal components for testing rare variants association with disease traits.

Usage

PCR(x, y, scale = FALSE, ncomp, varpercent, npermutation = 100, npermutation.max, min.nonsignificant.counts)
Arguments

- **x**: Genotype matrix
- **y**: Phenotype vector
- **scale**: If TRUE, scale x and y.
- **ncomp**: Number of components, which could be a vector containing a set of numbers.
- **varpercent**: Explained variance percentage
- **npermutation**: Number of permutation, if less than 1, the permutation will not be run.
- **npermutation.max**: Maximum permutation
- **min.nonsignificant.counts**: Minimum nonsignificant counts

Value

- **score**: Correlation between y and y_est
- **nonsignificant.counts**: Counts of permuted data that have a higher score than unpermuted data.
- **pvalue.empirical**: Empirical pvalue via permutation
- **pvalue.nominal**: Theoretical pvalue, not available now.
- **total.permutation**: Total permutation
- **ncomp.varp**: Number of components required for specified variance percentage

Author(s)

C. Xu

References


See Also

- PLS, RR
Partial Least Squares Regression for RV tests

**Description**

Use PLS components for testing rare variants association with disease traits.

**Usage**

```r
PLS(x, y, scale = FALSE, ncomp, varpercent, 
npermutation = 100, npermutation.max, min.nonsignificant.counts)
```

**Arguments**

- `x`: Genotype matrix
- `y`: Phenotype vector
- `scale`: If TRUE, scale x and y.
- `ncomp`: Number of components, which could be a vector containing a set of numbers.
- `varpercent`: Explained variance percentage
- `npermutation`: Number of permutation, if less than 1, the permutation will not be run.
- `npermutation.max`: Maximum permutation
- `min.nonsignificant.counts`: Minimum nonsignificant counts

**Value**

- `score`: Correlation between y and y_est
- `nonsignificant.counts`: Counts of permuted data that have a higher score than unpermuted data.
- `pvalue.empirical`: Empirical pvalue via permutation
- `pvalue.nominal`: Theoretical pvalue, not available now.
- `total.permutation`: Total permutation
- `ncomp.varp`: Number of components required for specified variance percentage

**Author(s)**

C. Xu
References


See Also

PCR, SPLS

Ridge Regression for RV Tests

Description

Use ridge regression for testing rare variants association with disease traits.

Usage

RR(x, y, z = NULL, scale = FALSE, weights = 1, lambda = 1, 
npermutation = 1000, npermutation.max, min.nonsignificant.counts = 100)

Arguments

x  Genotype matrix
y  Phenotype vector
z  Covariate matrix
scale  If TRUE, scale x and y.
weights  Genotype weights
lambda  Regularization parameter
npermutation  Number of permutation
npermutation.max  Maximum permutation
min.nonsignificant.counts  Minimum nonsignificant counts

Value

nonsignificant.counts  Counts of permuted data that have a higher score than unpermuted data.
total.permutation  Total permutation
score  Correlation between y and y_est if z=NULL.
pvalue.empirical  Empirical pvalue via permutation
pvalue.nominal  Theoretical pvalue, not available.
Author(s)

C. Xu

References


See Also

PCR, PLS

---

**sample.cgeno**

*Genotype Counts dataset*

**Description**

A list of genotype counts, phenotype, and polyphen weight

**Usage**

```r
data(sample.cgeno)
```

**Format**

The format is: List of 3

```r
$cgeno :'data.frame': 960 obs. of 3 variables: 
..$ indid: int [1:960] 16929 18167 28671 31308 33182 49716 53138 57548466 57548466 57548466 ... 
..$ snpid: int [1:960] 57548466 57548466 57548466 57548466 57548466 57548466 57548466 57548466 57548466 57548466 ... 
..$ count: int [1:960] 1 1 1 1 1 1 1 1 1 1 ... 

$ phen :'data.frame': 262 obs. of 2 variables: 
..$ indid: int [1:262] 32 90 101 109 129 133 225 236 253 349 ... 
..$ trait: num [1:262] 0.128 0.166 0.884 0.929 0.195 ... 

$ polyphen.weight :'data.frame': 71 obs. of 2 variables: 
..$ snpid: int [1:71] 57548364 57548466 57550649 57550666 57556205 57556220 57556236 57567762 57569339 57569466 ... 
..$ weight: num [1:71] 0.5 0.5 0.5 0.5 0.055 0.055 0.706 0.5 0.995 0.5 ... 
```

**Details**

The dataset was used in comparing VT and WOD methods.

**Examples**

```r
data(sample.cgeno)
str(sample.cgeno)
```
**SPLS**  
_Sparse PLS for RV Tests_

**Description**

Use SPLS for selecting significant variants and testing the variants associated with disease traits.

**Usage**

```r
SPLS(x, y, scale = TRUE, ncomp, eta.grid, size.max, a = 2, npermutation = 0, npermutation.max, min.nonsignificant.counts)
```

**Arguments**

- **x**  
  Genotype matrix, each row as an individual and each column as a SNP
- **y**  
  Phenotype vector
- **scale**  
  see `spls`
- **ncomp**  
  Number of components
- **eta.grid**  
  see `spls`
- **size.max**  
  Maximum number of variants included
- **a**  
  Penalty parameter for information criterion, a=2 for AIC.
- **npermutation**  
  Number of permutation, if less than 1, the permutation will not be run.
- **npermutation.max**  
  Maximum permutation
- **min.nonsignificant.counts**  
  Minimum nonsignificant counts

**Details**

Use `spls` package to implement SPLS and an information criterion (AIC, BIC, GIC) to select a set of variants.

**Value**

- **nonsignificant.counts**  
  Counts of permuted data that have a higher score than unpermuted data.
- **pvalue.empirical**  
  Empirical p-value via permutation
- **pvalue.nominal**  
  Not available
- **vs**  
  The selected variants
- **total.permutation**  
  Total permutation
VTWOD

Author(s)
C. Xu

References

See Also
spls, LASSO

VTWOD

VT and WOD for RV Tests

Description
Include methods: T1, T5, WE, VT, and WOD.

Usage
VTWOD(x, y, polyphen.weight, flipPhenotype = 0,
npermutation = 1000, npermutation.max, min.nonsignificant.counts)

Arguments
x Genotype matrix
y Phenotype vector
polyphen.weight Polyphen weight
flipPhenotype Logical, if TRUE, flip phenotype to opposite by multiplying -1
npermutation Number of permutation, if less than 1, the permutation will not be run.
npermutation.max Maximum permutation
min.nonsignificant.counts Minimum nonsignificant counts

Value
score Scores of T1, T5, WE, VT, and WOD
nonsignificant.counts Counts of permuted data that have a higher score than unpermuted data.
pvalue.empirical Empirical pvalue via permutation
pvalue.nominal
   Theoretical pvalue, not available now.
total.permutation
   Total permutation

Note
This R implementation by Adam Kiezun, based on reference implementation in C by Alkes Price. Added WOD tests to the program in 2011 by Celia Greenwood

Author(s)
C. Xu, Celia Greenwood

References

See Also
   RR, PCR, PLS
Index

*Topic **datasets**
  count2geno, 3
  geno2count, 3
  sample.cgeno, 9

*Topic **models**
  LASSO, 4
  PCR, 5
  PLS, 7
  RR, 8
  SPLS, 10
  VTWOD, 11

*Topic **multivariate**
  LASSO, 4
  PCR, 5
  PLS, 7
  RR, 8
  SPLS, 10
  VTWOD, 11

count2geno, 3, 4
geno2count, 3, 3
glmnet, 5
LASSO, 4, 11
PCR, 5, 8, 9, 12
PLS, 6, 7, 9, 12
RR, 6, 8, 12
RVtests (RVtests-package), 2
RVtests-package, 2
sample.cgeno, 9
SPLS, 5, 8, 10
spls, 11
VTWOD, 11