Package ‘GEEaSPU’

August 4, 2016

Type Package

Title Adaptive Association Tests for Multiple Phenotypes using Generalized Estimating Equations (GEE)

Version 1.0.2

Date 2016-08-03

Author Junghi Kim and Wei Pan

Maintainer Junghi Kim <junghikim0@gmail.com>

Description Provides adaptive association tests for SNP level, gene level and pathway level analyses.

License GNU General Public License (>= 3)

Depends R (>= 2.10), gee, stats

Imports Rcpp (>= 0.12.6)

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 5.0.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-08-04 11:47:48

R topics documented:

GEEaSPU-package ................................................................. 2
GEEaSPU ................................................................. 2
GEEaSPUpath ................................................................. 4
GEEaSPUset ................................................................. 7

Index 10
Description

Provides adaptive association tests for SNP level, gene level and pathway level analyses.

Details

Functions for adaptive association tests including GEEaSPU, GEEaSPUset and GEEaSPUpath. These methods use a weighting scheme for testing associations with multiple phenotypes. GEEaSPU function computes a SNP level p value; GEEaSPUset function can be used for the SNP-set (gene) level association test, while GEEaSPUpath function is for the gene-set (pathway) level analysis.

Author(s)

Junghi Kim and Wei Pan

References


Description

Tests whether a SNP is associated with multivariate outcomes; provides a series of p-values of GEE-SPU(pow) and GEE-aSPU tests.

Usage

GEEaSPU(pheno, geno, Z = NULL, model = "gaussian", corstr = "independence",
        pow = c(1, Inf), n.perm = 1000, null.type = "perm", score.test = FALSE)
Arguments

- **pheno**: a numeric phenotype matrix with each row as a different individual and each column as a separate phenotype.
- **geno**: a numeric vector with each element for an individual genotype.
- **Z**: a numeric covariate matrix with each row as a different individual and each column as a covaried to be adjusted.
- **model**: a character string specifying the model of the phenotypes. Models supported are "gaussian" for a quantitative trait and "binomial" for a binary trait (default = "gaussian").
- **corstr**: a character string specifying the correlation structure of phenotypes. The following are permitted: "independence", "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M" and "unstructured" (default = "independence").
- **pow**: a vector of the power weight to be used at a trait level (default = c(1:8, Inf)).
- **n.perm**: a numeric value of number of null statistics (default = 1000).
- **null.type**: a character string specifying how to generate null statistics; "perm" is used when null statistics are generated using permutations and "sim" is used when null statistics are generated using simulations (default = "perm").
- **score.test**: a logical value indicating whether to include GEEaSPU-Score test along with GEE-Score test (default = FALSE). If TRUE, it computes p-values of GEEaSPU-Score and GEE-Score as well as GEEaSPU test.

Details
Adaptive association tests for single SNP and multiple phenotypes using GEE.

Value
a vector of p-values from GEE-SPU(pow) tests and GEE-aSPU test.

Note
When large SNP-set (namely large gene size) or large number of phenotypes are included, the permutation based test (null.type = "perm") is recommended.
An option "binomial" model only supports the option, null.type = "sim".

Author(s)
Junghi Kim and Wei Pan

References
Examples

# -- simulating phenotypes
# -- n.subjects: number of subjects
# -- n.traits: number of phenotypes
# -- Sigma: covariance matrix of phenotypes (e.g. AR(1))

set.seed(136)
n.subjects <- 100
n.traits <- 3
sigma <- 2; rho <- 0.5
Sigma0 <- diag(n.traits);
Sigma <- sigma * rho^abs(row(Sigma) - col(Sigma))
eS <- eigen(Sigma, symmetric = TRUE)
ev <- eS$values
X <- matrix(rnorm(n.traits * n.subjects), n.subjects)
pheno <- X * diag(sqrt(pmax(ev, 0)), ncol(Sigma))

# -- simulating genotype

# -- Computing the p-value of GEEaSPU test with the permutation based method
Pv1 <- GEEaSPU(pheno = pheno, geno = geno, Z = NULL, pow = c(1,2,4,Inf),
n.perm = 1000, null.type = "perm", score.test = FALSE)

# -- Each element of Pv1 is a p value of GEE-SPU(pow) in order
# -- The last element of Pv1 is a p value of GEE-aSPU test

Pv1

# > Pv1
#  SPU.1  SPU.2  SPU.4  SPU.Inf  aSPU
# 0.1890000 0.4070000 0.3520000  0.3040000 0.2917083

# > Pv1[length(Pv1)]
# aSPU
# 0.2917083

GEEaSPUpath

An adaptive sum of powered score (SPU) test for gene-set (pathway) and multiple phenotypes

Description

Tests whether gene-set (pathway) is associated with multivariate outcomes; provides a series of p-values of GEE-SPU(pow, pow2, pow3) and GEEaSPUpath tests.
**Usage**

```
GEEaSPUpath(phen, geno, nSNPs, Z = NULL, corstr = "independence",
  pow = c(1,2,4,8), pow2 = c(1,2,4,8), pow3 = c(1,2,4,8), n.perm = 1000)
```

**Arguments**

- **phen**
  a numeric phenotype matrix with each row as a different individual and each column as a separate phenotype.

- **geno**
  a numeric genotype matrix with each row as a different individual and each column as a SNP; the SNPs (with the number stored in nSNPs) from one gene are stored consecutively from the first gene.

- **nSNPs**
  A numeric vector, whose length matches to the total number of genes; each element of vector indicate the number of SNPs in each gene.

- **Z**
  a numeric covariate matrix with each row as a different individual and each column as a covariated to be adjusted.

- **corstr**
  a character string specifying the correlation structure of phenotypes. The following are permitted: "independence", "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M" and "unstructured" (default = "independence").

- **pow**
  a vector of the power weight to be used at a SNP level (default = c(1,2,4,8)).

- **pow2**
  a vector of the power weight to be used at a trait level (default = c(1,2,4,8)).

- **pow3**
  a vector of the power weight to be used at a gene level (default = c(1,2,4,8)).

- **n.perm**
  a numeric value of number of null statistics (default = 1000).

**Details**

Adaptive association tests for gene-set (pathway) and multiple phenotypes using GEE.

**Value**

a vector of p-values from GEE-SPU(pow, pow2, pow3) tests and GEE-aSPUpath test.

**Note**

GEEaSPUpath function only supports a case for a quantitative trait (model = "gaussian") and a permutation based test (null.type = "perm").

**Author(s)**

Junghi Kim and Wei Pan

**References**

Examples

# -- simulating phenotypes
# -- n.subjects: number of subjects
# -- n.traits: number of phenotypes
# -- Sigma: covariance matrix of phenotypes (e.g. AR(1))

set.seed(136)
n.subjects <- 100
n.traits <- 3
sigma <- diag(n.traits)
Sigma <- sigma * rho^abs(row(sigmaPI - col(sigmaPII
eS <- eigen(Sigma, symmetric = TRUE)
ev <- eS$values
X <- matrix(rnorm(n.subjects * n.traits), n.subjects)
pheno <- X %*% diag(sqrt(pmax(evL PIIL ncol(Sigma)) %*% eS$vectors

# -- simulating genotype
# -- Assume we have two genes each of which has 3 and 5 SNPs respectively.
# -- n.genos: number of SNPs included in the gene1
# -- n.genos2: number of SNPs included in the gene2
# -- nSNPs <- c(3,5)

n.genos1 <- 3
n.genos2 <- 5
maf1 <- 0.2
maf2 <- 0.4
gene1 <- matrix(rbinom(n = n.subjects*n.genos1, size = 2, prob = maf1), ncol = n.genos1)
gene2 <- matrix(rbinom(n = n.subjects*n.genos2, size = 2, prob = maf2), ncol = n.genos2)
geno <- cbind(gene1, gene2)

# -- Computing the p-value of GEEaSPUpath test
Pvl <- GEEaSPUpath(pheno = pheno, geno = geno, nSNPs = c(3,5), Z = NULL,
                     corstr = "independence", pow = c(1,4,8), pow2 = c(1,4,8),
                     pow3 = c(1,4,8), n.perm = 1000)

# -- Each element of Pvl is a p value of GEE-SPU(pow,pow2,pow3) in order
# -- The last element of Pvl is a p value of GEE-aSPUpath test
Pvl
Pvl[length(Pvl)]

> Pvl
# SPU.1.1.1  SPU.1.1.4  SPU.1.1.8  SPU.1.4.1  SPU.1.4.4  SPU.1.4.8  SPU.1.8.1  SPU.1.8.4
#  0.0090000  0.0560000  0.0700000  0.0620000  0.0830000  0.1120000  0.0610000  0.0820000
# SPU.1.8.8  SPU.4.1.1  SPU.4.1.4  SPU.4.1.8  SPU.4.4.1  SPU.4.4.4  SPU.4.4.8  SPU.4.8.1
#  0.1060000  0.5810000  0.5430000  0.4920000  0.6240000  0.6400000  0.6270000  0.6490000
# SPU.4.8.4  SPU.4.8.8  SPU.8.1.1  SPU.8.1.4  SPU.8.1.8  SPU.8.4.1  SPU.8.4.4  SPU.8.4.8
#  0.6710000  0.6750000  0.5830000  0.5370000  0.4810000  0.6340000  0.6460000  0.6380000
GEEaSPUset

An adaptive sum of powered score (SPU) test for SNP-set (gene) and multiple phenotypes

Description

Tests whether SNP-set (gene) is associated with multivariate outcomes; provides a series of p-values of GEE-SPU(pow, pow2) and GEEaSPUset tests.

Usage

GEEaSPUset(pheno, geno, Z = NULL, model = "gaussian", corstr = "independence",
pow = c(1,2,4,8), pow2 = c(1,2,4,8), n.perm = 1000,
nullNtype = "perm", scoreNtest = FALSE)

Arguments

pheno a numeric phenotype matrix with each row as a different individual and each column as a separate phenotype.
geno a numeric genotype matrix with each row as a different individual and each column as a snp.
Z a numeric covariate matrix with each row as a different individual and each column as a covaried to be adjusted.
model a character string specifying the model of the phenotypes. Models supported are "gaussian" for a quantitative trait and "binomial" for a binary trait (default = "gaussian").
corstr a character string specifying the correlation structure of phenotypes. The following are permitted: "independence", "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M" and "unstructured" (default = "independence").
pow a vector of the power weight to be used at a SNP level (default = c(1,2,4,8)).
pow2 a vector of the power weight to be used at a trait level (default = c(1,2,4,8)).
nperm a numeric value of number of null statistics (default = 1000).
nulNtype a character string specifying how to generate null statistics; "perm" is used when null statistics are generated using permutations and "sim" is used when null statistics are generated using simulations (default = "perm").
scoreNtest a logical value indicating whether to include GEEaSPU-Score test along with GEE-Score test (default = FALSE). If TRUE, it computes p-values of GEEaSPU-Score and GEE-Score as well as GEEaSPU test.
Details

Adaptive association tests for SNP-set (gene) and multiple phenotypes using GEE.

Value

a vector of p-values from GEE-SPU(pow, pow2) tests and GEE-aSPUset test.

Note

When large SNP-set (namely large gene size) or large number of phenotypes are included, the permutation based test (null.type = "perm") is recommended.

An option "binomial" model only supports the option, null.type="sim".

Author(s)

Junghi Kim and Wei Pan

References


Examples

```r
set.seed(136)
n.subjects <- 100
n.traits <- 3
sigma <- 2; rho <- 0.5
Sigma0 <- diag(n.traits);
Sigma <- sigma * rho^abs(row(Sigma0) - col(Sigma0))
eS <- eigen(Sigma, symmetric = TRUE)
ev <- eS$values
X <- matrix(rnorm(n.subjects * n.traits), n.subjects)
pheno <- X %*% diag(sqrt(pmax(ev[, 0])), ncol(Sigma)) %*% eS$vectors

set.seed(136)
n.subjects <- 100
n.traits <- 3
sigma <- 2; rho <- 0.5
Sigma0 <- diag(n.traits);
Sigma <- sigma * rho^abs(row(Sigma0) - col(Sigma0))
eS <- eigen(Sigma, symmetric = TRUE)
ev <- eS$values
X <- matrix(rnorm(n.subjects * n.traits), n.subjects)
pheno <- X %*% diag(sqrt(pmax(ev[, 0])), ncol(Sigma)) %*% eS$vectors

# -- simulating genotype
# -- n.genoe: number of SNPs included in the SNP set/gene

n.genoe <- 3
maf <- 0.2
geno <- matrix(rbinom(n = n.subjects * n.genoe, size = 2, prob = maf), ncol = n.genoe)

# -- Computing the p-value of GEEaSPUset test with the permutation based method
Pv1 <- GEEaSPUset(pheno = pheno, geno = geno, Z = NULL, model = "gaussian",
```
corstr = "independence", pow = c(1,4,Inf), pow2 = c(1,4,Inf), n.perm = 1000,
null.type = "perm", score.test = FALSE)

# -- Each element of pvl is a p value of GEE-SPU(pow,pow2) in order
# -- The last element of pvl is a p value of GEE-aSPUset test
pvl
pvl[length(pvl)]

# > Pvl
# SPU.1.1   SPU.1.4   SPU.1.Inf  SPU.4.1   SPU.4.4   SPU.4.Inf   SPU.Inf.1
#  0.0140000  0.0880000  0.0720000  0.5300000  0.4100000  0.3210000  0.5510000
# SPU.Inf.4  SPU.Inf.Inf  aSPUset
#  0.4870000  0.4100000  0.04095904

# > Pvl[length(Pvl)]
#  aSPUset
#  0.04095904
# Index

<table>
<thead>
<tr>
<th>Topic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP based</td>
<td>GEEaSPU, 2</td>
</tr>
<tr>
<td>SNP-set/gene based</td>
<td>GEEaSPUset, 7</td>
</tr>
<tr>
<td>gene-set/pathway based</td>
<td>GEEaSPUpath, 4</td>
</tr>
<tr>
<td>multiple phenotypes</td>
<td>GEEaSPU, 2</td>
</tr>
<tr>
<td></td>
<td>GEEaSPUpath, 4</td>
</tr>
<tr>
<td></td>
<td>GEEaSPUset, 7</td>
</tr>
</tbody>
</table>

GEEaSPU, 2
GEEaSPU-package, 2
GEEaSPUpath, 4
GEEaSPUset, 7