Package ‘MPAT’

October 19, 2016

Type Package
Title For multiple phenotype genetic association studies
Version 1.0
Date 2015-09-10
Author Zhonghua Liu
Maintainer Zhonghua Liu <zliu@mail.harvard.edu>
Description Conduct multiple phenotype genetic association studies based on summary statistics.
License GPL (>= 2)
NeedsCompilation yes
Imports mvtnorm

R topics documented:

MPAT-package .......................................................... 2
DSUM ................................................................. 2
lipids ................................................................. 3
MinP ................................................................. 4
mixAda .............................................................. 5
mixFisher .......................................................... 6
mixSD ............................................................. 7
mixTippett ........................................................ 8
mixVar ............................................................ 9
PC ................................................................. 10
PCAZ .............................................................. 11
PCFisher .......................................................... 12
PCLC ............................................................... 12
PCMnP ........................................................... 13
PCO ............................................................... 14
SigmaOEestimate .................................................. 15
SigmaXEestimate ................................................ 15
SUM ............................................................. 16
VC ............................................................... 17
Wald ............................................................. 18
WI ............................................................... 18

Index 20
Description

This package is for conducting multiple phenotype genetic association testings by combining univariate summary statistics for each phenotype in GWAS. It contains fourteen methods that can be used in a variety of situations. An overall p-value will be returned for assessing the significance of associations between a SNP and multiple phenotypes. It contains the following functions: PC, PCMinP, PCFisher, PCLC, PCSS, MinP, WI, VC, DSUM, SUM, Wald, mixAda, mixFisher, mixTippett, mixVar, mixSD.

Details

Package: MPAT
Type: Package
Version: 1.0
Date: 2014-09-11
License: License: GPL(>= 2)

Author(s)
Zhonghua Liu <zliu@mail.harvard.edu>
Maintainer: Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

DSUM

Direct summation of Z-statistics

Description

For SNP, the Z testing statistics for multiple phenotypes are summed together as an overall testing statistic.

Usage

DSUM(Z.vec, Sigma)

Arguments

Z.vec  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma  Sigma is the correlation matrix among the Z testing statistics that can be estimated by its sample version over the SNPs.
**Value**

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

**Author(s)**

Zhonghua Liu <zliu@mail.harvard.edu>

**References**

MPAT: an R package for multivariate phenotype association studies.

**Examples**

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
DSUM(Z.vec=lipids_zscore[1,],Sigma) ## p-value for the first SNP

---

**lipids**

An example data set that contains summary statistics from GWAS studies of four lipids levels.

**Description**

This is an example data set that contains summary statistics (Z-scores) for 2000 genetic variants and four lipids levels.

**Usage**

data(lipids)

**Format**

A data frame with 2000 observations on the following 15 variables.

- **MarkerName** SNP names
- **Allele1** This the allele used as the effect allele
- **Allele2** This is the "other" allele
- **W.HDL** The sum of the individual study weights (typically, N) for this marker
- **Zscore.HDL** Z scores for HDL
- **Pvalue.HDL** P-values for HDL
- **W.LDL** The sum of the individual study weights (typically, N) for this marker
- **Zscore.LDL** Z scores for LDL
- **Pvalue.LDL** P-values for LDL
- **W.TC** The sum of the individual study weights (typically, N) for this marker
- **Zscore.TC** Z scores for TC
- **Pvalue.TC** P-values for TC
The sum of the individual study weights (typically, N) for this marker

Zscore.TG  Z scores for TG

Pvalue.TG  P-values for TG

Source

http://www.sph.umich.edu/csg/abecasis/public/lipids2010/

References

http://www.nature.com/nature/journal/v466/n7307/full/nature09270.html

Examples

```r
data(lipids)
head(lipids)
```
Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
minP(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP

mixAda

Adaptively select optimal combination coefficient of two independent score statistics based on linear mixed effects model.

Description

mixAda combines the Z testing statistics for multiple phenotypes at a genetic variant in an adaptive fashion to maximize the power.

Usage

mixAda(Z.vec, Sigma)

Arguments

Z.vec Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma Sigma is the correlation matrix among the Z testing statistics.

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
mixAda(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP
mixFisher

Combine Z-statistics for multiple phenotypes using Fisher’s method in a linear mixed effects model.

Description

Z-statistics for multiple phenotypes are modeled using a linear mixed effects model. Two independent score testing statistics are combined using Fisher’s method.

Usage

mixFisher(Z.vec, Sigma, method)

Arguments

Z.vec
  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma
  Sigma is the correlation matrix among the Z testing statistics.
method
  The method to compute the p-value which takes three values: davies, liu and liumod.

Value

p_group
  p-value for the group effects of a genetic variant on the multiple phenotypes
p_individual
  p-value for the individual effects of a genetic variant on the multiple phenotypes
p_overall
  An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
imxFisher(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP
Description

Z-statistics for multiple phenotypes are modeled using a linear mixed effects model. Two independent score testing statistics are combined using inverse standard deviation weighting method.

Usage

mixSD(Z.vec, Sigma, method)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z.vec</td>
<td>Z.vec is a vector representing the Z testing statistics for multiple phenotypes.</td>
</tr>
<tr>
<td>Sigma</td>
<td>Sigma is the correlation matrix among the Z testing statistics.</td>
</tr>
<tr>
<td>method</td>
<td>The method to compute the p-value which takes three values: davies, liu and liumod.</td>
</tr>
</tbody>
</table>

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
mixSD(Z.vec=lipids_zscore[1,],Sigma) ## p-value for the first SNP
Combine Z-statistics for multiple phenotypes using Tippett’s method in a linear mixed effects model.

Description

Z-statistics for multiple phenotypes are modeled using a linear mixed effects model. Two independent score testing statistics are combined using Tippett’s method.

Usage

mixTippett(Z.vec, Sigma, method)

Arguments

- **Z.vec**: Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
- **Sigma**: Sigma is the correlation matrix among the Z testing statistics.
- **method**: The method to compute the p-value which takes three values: davies, liu and liumod.

Value

- **p_group**: p-value for the group effects of a genetic variant on the multiple phenotypes.
- **p_individual**: p-value for the individual effects of a genetic variant on the multiple phenotypes.
- **p_overall**: An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
mixTippett(Z.vec=lipids_zscore[1,],Sigma) ## p-value for the first SNP
mixVar

Combine Z-statistics for multiple phenotypes using inverse variance weighting method in a linear mixed effects model.

Description

Z-statistics for multiple phenotypes are modeled using a linear mixed effects model. Two independent score testing statistics are combined using inverse variance weighting method.

Usage

mixVar(Z.vec, Sigma, method)

Arguments

Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma is the correlation matrix among the Z testing statistics.
The method to compute the p-value which takes three values: davies, liu and liumod.

Value

An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])Sigma = cor(lipids_zscore)
mixVar(Z.vec=lipids_zscore[1,],Sigma) # p-value for the first SNP
PC is a function used to combine correlated GWAS summary statistics for multiple phenotypes at a genetic locus.

Description

By combining univariate summary statistics for multiple phenotypes, we can obtain an overall association p-value that assess the association between a genetic variant and multiple phenotypes as a whole. PC allows users to conduct dimension reduction by specifying which principal component to use to represent the original multiple summary statistics.

Usage

PC(Z.vec, Sigma, PCorder)

Arguments

Z.vec is a column vector representing the Wald-type Z testing statistics for multiple phenotypes.

Sigma is the correlation matrix of the multiple Z testing statistics.

PCorder specifies which principal component to use, it takes integer values. For example, PCorder=1 means we choose to use the first principal component.

Value

An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
PC(Z.vec=lipids_zscore[1,],Sigma,PCorder=1) # p-value for the first SNP using PC1
**PCAQ**

*Use PCAQ method to combine Z testing statistics.*

---

**Description**

Use PCAQ method to combine Z testing statistics.

**Usage**

```r
PCAQ(Z.vec, Sigma, SigmaX)
```

**Arguments**

- `Z.vec`: Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
- `Sigma`: Sigma is the correlation matrix among the Z testing statistics.
- `SigmaX`: The correlation among X which can be estimated by simulation using SigmaX-Estimate.

**Value**

- `p-value`: An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

**Author(s)**

Zhonghua Liu <zliu@mail.harvard.edu>

**References**

MPAT: an R package for multivariate phenotype association studies.

**Examples**

```r
data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
SigmaX = SigmaXEstimate(Sigma, simNum=1000)
PCAQ(Z.vec=lipids_zscore[,1], Sigma=Sigma, SigmaX=SigmaX) # p-value for the first SNP
```
PCLC

Fisher’s method to combine principal components based on Z testing statistics

Description

Z testing statistics for multiple phenotype are orthogonally transformed onto principal component axes and combined using Fisher’s method.

Usage

PCFisher(Z.vec, Sigma)

Arguments

Z.vec Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma Sigma is the correlation matrix among the Z testing statistics.

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
PCFisher(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP

PCLC

Linear combination of the principal components of the Z testing statistics.

Description

Z testing statistics for multiple phenotype are orthogonally transformed onto principal component axes and combined in a linear fashion.

Usage

PCLC(Z.vec, Sigma)
Arguments

Z.vec  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma  Sigma is the correlation matrix among the Z testing statistics.

Value

p-value  An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
PCCLC(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP

Description

The p-values using principal components of the Z testing statistics are computed, and the minimum one is taken as the testing statistic.

Usage

PCMinP(Z.vec, Sigma)

Arguments

Z.vec  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma  Sigma is the correlation matrix among the Z testing statistics.

Value

p-value  An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>
References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)

lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])

Sigma = cor(lipids_zscore)

PCOminpHzNvec=lipids_zscore[1,],Sigma) ## p-value for the first SNP

Description

Use PCO method to combine Z testing statistics.

Usage

PCO(Z.vec, Sigma,Sigma0)

Arguments

Z.vec  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma  Sigma is the correlation matrix among the Z testing statistics.
Sigma0  The correlation among X which can be estimated by simulation using Sigma0Estimate.

Value

p-value  An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)

lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])

Sigma = cor(lipids_zscore)

Sigma0 = Sigma0Estimate(Sigma,simNum=1000)

PCO(Z.vec=lipids_zscore[1,],Sigma=Sigma,Sigma0=Sigma0) ## p-value for the first SNP
SigmaOEstimate

To estimate the correlation matrix of among X used for PCO

Description

Compute SigmaX which is required for PCO function.

Usage

SigmaOEstimate(Sigma, simNum)

Arguments

Sigma

Sigma is the correlation matrix among the Z testing statistics.

simNum

The number of simulations to be performed to estimate SigmaX.

Value

sigma0

A correlation matrix among X.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
SigmaOEstimate(Sigma, simNum=1000)

SigmaXestimate

To estimate the correlation matrix of among X used for PCAQ

Description

Compute SigmaX which is required for PCAQ function.

Usage

SigmaXEstimate(Sigma, simNum)

Arguments

Sigma

Sigma is the correlation matrix among the Z testing statistics.

simNum

The number of simulations to be performed to estimate SigmaX.
Value

SigmaX A correlation matrix among X.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
SigmaEstimate(Sigma,simNum=1000)

Description

The Z testing statistics for multiple phenotypes are combined using SUM method with correlation structures explicitly taken into account.

Usage

SUM(Z.vec, Sigma)

Arguments

Z.vec Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma Sigma is the correlation matrix among the Z testing statistics.

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.
Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
SUM(Z.vec=lipids_zscore[1,],Sigma) ## p-value for the first SNP

VC

Use variance component test (VC) to combine Z testing statistics

Description

Use variance component test (VC) to combine Z testing statistics of multiple phenotypes.

Usage

VC(Z.vec, Sigma, method)

Arguments

Z.vec Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma Sigma is the correlation matrix among the Z testing statistics.
method The method to compute the p-value which takes three values: davies.liu and liumod.

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
VC(Z.vec=lipids_zscore[1,],Sigma) ## p-value for the first SNP


**Wald**

*Traditional Wald type combination of Z testing statistics*

**Description**

The Z testing statistics for multiple phenotypes are combined using Wald type method.

**Usage**

\[ \text{Wald}(Z.\text{vec}, \Sigma) \]

**Arguments**

- **Z.vec**: Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
- **Sigma**: Sigma is the correlation matrix among the Z testing statistics.

**Value**

- **p-value**: An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

**Author(s)**

Zhonghua Liu <zliu@mail.harvard.edu>

**References**

MPAT: an R package for multivariate phenotype association studies.

**Examples**

data(lipids)

lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])

Sigma = cor(lipids_zscore)

Wald(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP

---

**WI**

*Use WI method to combine Z testing statistics.*

**Description**

Use WI method to combine Z testing statistics.

**Usage**

\[ \text{WI}(Z.\text{vec}, \Sigma, \text{method}) \]

**Examples**

data(lipids)

lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])

Sigma = cor(lipids_zscore)

WI(Z.vec=lipids_zscore[,1],Sigma) ## p-value for the first SNP
Arguments

Z.vec  Z.vec is a vector representing the Z testing statistics for multiple phenotypes.
Sigma  Sigma is the correlation matrix among the Z testing statistics.
method The method to compute the p-value which takes three values: davies, liu and liumod.

Value

p-value An overall p-value assessing the association between a genetic variant and multiple phenotypes is returned.

Author(s)

Zhonghua Liu <zliu@mail.harvard.edu>

References

MPAT: an R package for multivariate phenotype association studies.

Examples

data(lipids)
lipids = as.data.frame(lipids)
lipids_zscore = as.matrix(lipids[,c("Zscore.HDL","Zscore.LDL","Zscore.TG","Zscore.TC")])
Sigma = cor(lipids_zscore)
WI(Z.vec=lipids_zscore[1,],Sigma,method="liu") # p-value for the first SNP
Index

*Topic datasets*
  - lipids, 3

DSUM, 2

lipids, 3

MinP, 4
mixAda, 5
mixFisher, 6
mixSD, 7
mixTippett, 8
mixVar, 9
MPAT (MPAT-package), 2
MPAT-package, 2

PC, 10
PCAQ, 11
PCFisher, 12
PCLC, 12
PCMinP, 13
PCQ, 14
PCSS (Wald), 18

SigmaOEstimate, 15
SigmaXEstimate, 15
SUM, 16

VC, 17

Wald, 18
WI, 18