

Package ‘RHT’

February 19, 2015

Type Package

Title Regularized Hotelling's T-square Test for Pathway (Gene Set) Analysis

Version 1.0

Date 2011-11-14

Author Lin S. Chen and Pei Wang

Maintainer Lin S. Chen <lchen11@uchicago.edu>

Description This package offers functions to perform regularized Hotelling's T-square test for pathway or gene set analysis. The package is tailored for but not limited to proteomics data, in which sample sizes are often small, a large proportion of the data are missing and/or correlations may be present.

License GPL

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-29 08:57:31

NeedsCompilation no

R topics documented:

RHT-package	2
RHT.2samp	3
RHT.fun	4
Index	6

RHT-package

Regularized Hotelling's T-square Test for Pathway (Gene Set) Analysis

Description

This package offers functions to perform regularized Hotelling's T-square test for pathway or gene set analysis. The package is tailored for but not limited to proteomics data, in which sample sizes are often small, and a large proportion of the data are missing and/or correlations may be present.

Details

Package: RHT
Type: Package
Version: 1.0
Date: 2011-11-14
License: GPL
LazyLoad: yes

Author(s)

Lin S. Chen and Pei Wang

Maintainer: Lin S. Chen <lchen11@uchicago.edu>

References

Chen LS, Paul D, Prentice RL and Wang P. (2011) A regularized Hotelling's T-square test for pathway analysis in proteomics studies. Journal of the American Statistical Association, in press.

See Also

[RHT.fun](#), [RHT.2samp](#)

Examples

```
## we simulate a data set with N=10 samples and p=50 proteins.  
## 20% of the data are missing.  
## Among the 50 proteins, we randomly assign 2 pathways, with 5 and 12 proteins, respectively.  
  
set.seed(1)  
X <- matrix(rnorm(500),nrow=10)  
X[sample(1:500, 0.2*500)] <- NA  
path.idx <- list()  
path.idx[[1]] <- 1:5  
path.idx[[2]] <- 13:24
```

```
names(path.idx) <- c("pathway A", "pathway B")

## The following function tests each pathway to see
## if any of the proteins in each pathway shows non-zero
## abundance/expression

pval <- RHT.fun(path.idx, dat=X)
```

RHT.2samp

Two-sample Regularized Hotelling's T-square Test

Description

This function tests if a pathway (or gene set) consists of any protein (or gene) that shows different mean abundance (or expression) between two groups of samples.

Usage

```
RHT.2samp(path.idx, datX, datY, nsim = 1000, seed = 123)
```

Arguments

path.idx	This is a LIST. Each element in the list contains the indice of proteins (or genes) for a pathway in the data set.
datX	An N1 by p matrix of protein abundance (or gene expression) from one group of samples. Each row represents one sample and each column represents a protein (or a gene).
datY	An N2 by p matrix of protein abundance (or gene expression) from another group of samples. Each row represents one sample and each column represents a protein (or a gene).
nsim	Number of resamples needed to calculate the p-value. By default, nsim=1000.
seed	A single integer that controls the random number generator in the resampling.

Value

The function returns the p-values for each pathway in the list path.idx.

Author(s)

Lin S. Chen and Pei Wang

References

Chen LS, Paul D, Prentice RL and Wang P. (2011) A regularized Hotelling's T-square test for pathway analysis in proteomics studies. *Journal of the American Statistical Association*, in press.

See Also

See Also [RHT.fun](#)

Examples

```
## We simulate a data set X with N=10 samples and p=50 proteins,
## and a second data set Y with N=8 sample and the same number of proteins.
## 20% of the data are missing.

set.seed(1)
X <- matrix(rnorm(500),nrow=10)
X[sample(1:500, 0.2*500)] <- NA

Y <- matrix(rnorm(400),nrow=8)
Y[sample(1:400, 0.2*400)] <- NA

## Among the 50 proteins, we randomly assign 2 pathways, with 5 and 12 proteins, respectively.
path.idx <- list()
path.idx[[1]] <- 1:5
path.idx[[2]] <- 13:24
names(path.idx) <- c("pathway A", "pathway B")

## The following function tests each pathway to see
## if any of the proteins in each pathway shows different
## abundance/expression between data X and Y.

pval <- RHT.2samp(path.idx, datX=X, datY=Y)
```

RHT.fun

One-sample Regularized Hotelling's T-square Test

Description

This function tests if a pathway (or gene set) consists of any protein (or gene) that shows non-zero abundance (or expression).

Usage

```
RHT.fun(path.idx, dat, nsim = 1000, seed = 123)
```

Arguments

path.idx	This is a LIST. Each element in the list contains the indice of proteins (or genes) for a pathway in the data set.
dat	An N by p matrix of protein abundance (or gene expression). Each row represents one sample and each column represents a protein (or a gene).
nsim	Number of resamples needed to calculate the p-value. By default, nsim=1000.
seed	A single integer that controls the random number generator in the resampling.

Value

The function returns the p-values for each pathway in the list `path.idx`.

Author(s)

Lin S Chen and Pei Wang

References

Chen LS, Paul D, Prentice RL and Wang P. (2011) A regularized Hotelling's T-square test for pathway analysis in proteomics studies. *Journal of the American Statistical Association*, in press.

See Also

See Also [RHT.2samp](#)

Examples

```
## we simulate a data set with N=10 samples and p=50 proteins.
## 20% of the data are missing.
## Among the 50 proteins, we randomly assign 2 pathways, with 5 and 12 proteins, respectively.

set.seed(1)
X <- matrix(rnorm(500),nrow=10)
X[sample(1:500, 0.2*500)] <- NA
path.idx <- list()
path.idx[[1]] <- 1:5
path.idx[[2]] <- 13:24
names(path.idx) <- c("pathway A", "pathway B")

## The following function tests each pathway to see
## if any of the proteins in each pathway shows non-zero
## abundance/expression

pval <- RHT.fun(path.idx, dat=X)
```

Index

*Topic **RHT**

RHT . 2samp, [3](#)

RHT . fun, [4](#)

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

RHT-package, [2](#)

RHT . 2samp, [2](#), [3](#), [5](#)

RHT . fun, [2](#), [4](#), [4](#)