Package ‘MAVTgsa’

February 19, 2015

Type Package

Title Three methods to identify differentially expressed gene sets, ordinary least square test, Multivariate Analysis Of Variance test with n contrasts and Random forest.

Version 1.3

Date 2014-05-27

Author Chih-Yi Chien, Chen-An Tsai, Ching-Wei Chang, and James J. Chen

Maintainer Chih-Yi Chien <92354503@nccu.edu.tw>

Depends R (>= 2.13.2), corpcor, foreach, multcomp, randomForest, MASS

Description This package is a gene set analysis function for one-sided test (OLS), two-sided test (multivariate analysis of variance).
If the experimental conditions are equal to 2, the p-value for Hotelling’s $t^2$ test is calculated.
If the experimental conditions are greater than 2, the p-value for Wilks’ Lambda is determined and post-hoc test is reported too.
Three multiple comparison procedures, Dunnett, Tukey, and sequential pairwise comparison, are implemented.
The program computes the p-values and FDR (false discovery rate) q-values for all gene sets.
The p-values for individual genes in a significant gene set are also listed.
MAVTgsa generates two visualization output: a p-value plot of gene sets (GSA plot) and a GST-plot of the empirical distribution function of the ranked test statistics of a given gene set.
A Random Forests-based procedure is to identify gene sets that can accurately predict samples from different experimental conditions or are associated with the continuous phenotypes.

License GPL-2

LazyData Yes

Repository CRAN

NeedsCompilation no

Date/Publication 2014-07-02 13:48:35

R topics documented:

MAVTgsa-package .......................................................... 2
data .................................................................................. 3
description

A gene set analysis function for one-sided test (OLS) and two-sided test (multivariate analysis of variance). Three multiple comparison procedures, Dunnett, Tukey, and sequential pairwise comparison, are implemented. MAVTgsa computes the p-values and FDR (false discovery rate) q-values for all gene sets. The p-values for individual genes in a significant gene set are also listed. MAVTgsa generates a GST-plot of the empirical distribution function of the ranked test statistics of a given gene set.

details

Package: MAVTgsa
Type: Package
Version: 1.0
Date: 2012-10-04
License: What license is it under?
LazyLoad: yes

author(s)

Chih-Yi Chien, Chen-An Tsai, Ching-Wei Chang, and James J. Chen
Maintainer: Chih-Yi Chien <chihyi.chien@fda.hhs.gov>

references


**Example data for MAVTn**

**Description**

A data matrix with 15 samples in column and one class type and 10 gene expression data in row.

**Usage**

data(data)

**Format**

The format is: num [1:16, 1:10] 1 -0.0853 1.0868 -0.4995 0.9477 ... - attr(*, "dimnames")=List of 2

\[
\begin{array}{c}
\text{..$: chr [1:16] "cl" "" "" "" ...} \\
\text{..$ : NULL}
\end{array}
\]

**Design matrix**

**Description**

To construct a design matrix of the clinical outcome of the samples.

**Usage**

design.matrix(factors)

**Arguments**

factors The clinical outcome of the samples.

**Value**

A design matrix is returned.
Example data for MAVTn

Description

A data matrix with 15 samples in column and one class type and 10 gene expression data in row.

Usage

data(GS)

Format

The format is: num [1:10, 1:4] 1 1 1 0 0 1 1 1 1 ...
References


---

**Hott2**

**Hottelling’s T square**

---

**Description**

To compute Hotelling’s T square statistic for multivariate analysis of variance using Shrinkage covariance matrix estimates.

**Usage**

```r
Hott2(x, y, var.equal = TRUE)
```

**Arguments**

- `x`  
  Data matrix; row is sample; each column is variable (gene)

- `y`  
  Vector defining two-group of the samples.

- `var.equal`  
  Logical.

**Value**

Hotelling’s T square statistic is calculated.

**Note**

R > 2.13.2

**Author(s)**

Chen-An Tsai, James J. Chen, Ching-Wei Chang, and Chih-Yi Chien

**References**


ma.estimate  

*Estimate of the coefficients*

**Description**

To calculate the ordinary least square estimate of the coefficients.

**Usage**

```r
ma.estimate(Y, X)
```

**Arguments**

- **Y**: Data matrix; row is sample; each column is variable (gene).
- **X**: Design matrix.

**Value**

The estimate of the coefficients is returned.

---

Mavtn  

*OLS, Hottelling's T2 and MANOVA with n contrasts*

**Description**

A gene set analysis functions for computing the p-values for one-sided test (OLS) and two-sided test (multivariate analysis of variance). If the experimental conditions are equal to 2, the p-value for Hotelling’s t square test is calculated. If the experimental conditions are greater than 2, the p-value for Wilks’ Lambda is determined and post-hoc test is reported too. The p-value for individual gene test of significant gene sets are also listed.

**Usage**

```r
Mavtn(DATA, GS, MCP = 1, alpha = 0.01, nbPerm = 5000)
```

**Arguments**

- **DATA**: an (m+1) x n gene expression data matrix with n samples in columns. The first row contains the information of experimental condition of each sample. The genes are expressed in the rest m rows.
- **GS**: an m x k binary matrix with code (0, 1), where k is the number of gene sets. Each column represents a pre-defined gene set.
- **MCP**: the choice for one of three multiple comparison methods, Dunnett = 1, Tuckey = 2, Sequential pairwise = 3.
- **alpha**: the significant level
- **nbPerm**: the number of permutation specified
**Value**

The p-values of OLS and MANOVA test are returned. If there is any significant gene set, the p values for individual genes in the gene set will be reported.

**Note**

R > 2.13.2

**Author(s)**

Chih-Yi Chien, Chen-An Tsai, Ching-Wei Chang, and James J. Chen

**References**


**Examples**

```r
# -------------- simulate data matrix --------------#

data(data)
data(GS)

MAVTp(data, GS, MCP=1, nbPerm = 100)
```

**Description**

A Random Forests-based procedure is to identify gene sets that can accurately predict samples from different experimental conditions or are associated with the continuous phenotypes.

**Usage**

```r
MAVTp(DATA, GS, nbPerm = 5000, numoftree = 500, type = c("cont", "cate"), impt = TRUE)
```
Arguments

**DATA**

a gene expression data matrix with samples in columns. The first row contains the information of the experimental condition of each sample. The remaining rows contain gene expression.

**GS**

an $m \times k$ binary matrix with code (0, 1), where $k$ is the number of gene sets. Each column represents a pre-defined gene set.

**nbPerm**

the number of permutation specified

**numoftree**

the number of trees to grow

**type**

This can be one of "cont" (continuous phenotypes) and "cate" (categorical phenotypes).

**impt**

If TRUE (default), the importance measurement will be output.

Value

A list of the p-values of random forests for GSA. The importance measurement of individual genes for those significant gene sets will also be output when impt is set TRUE.

Note

R > 2.14.0

Author(s)

Chih-Yi Chien, Chen-An Tsai, Ching-Wei Chang, and James J. Chen

References


Examples

data(data)
data(GS)
a=proc.time()
MAVTp(data,Gs, nbPerm = 50, numoftree = 500, type = "cate", impt = TRUE)
proc.time()-a
**minp**

**P-values adjustment in permutation**

**Description**

Returns the p-values in each permutation.

**Usage**

\[
\text{minp}(p, \text{rank}, n.\text{GeneSets}, \text{nbPerm})
\]

**Arguments**

- **p**: input p-values.
- **rank**: the rank of the p-values.
- **n.\text{GeneSets}**: the number of genes in a given gene set.
- **nbPerm**: the number of permutation times.

**Value**

a permutation p-value matrix.

---

**Tols**

**Ordinary Least Square test**

**Description**

To compute OLS statistic for one-sided test.

**Usage**

\[
\text{Tols}(x, y)
\]

**Arguments**

- **x**: Data matrix; row is sample; each column is variable(gene).
- **y**: Vector defining the clinical outcome of the samples.

**Value**

Returns OLS test statistic for gene set analysis.

**Author(s)**

Chih-Yi Chien, Chen-An Tsai, Ching-Wei Chang, and James J. Chen
Wilksn

Wilk’s Lambda for n-group multiple comparisons

Description
To compute Wilk’s Lambda statistic for multivariate analysis of variance and multiple comparisons.

Usage
Wilksn(y, class, type = c("Tukey", "Dunnett", "Sequence"), base = 1)

Arguments
Y Data matrix; row is sample; each column is variable(gene).
class Vector defining the clinical outcome of the samples.
type Type of contrast
base An integer to denote which group is considered the baseline group for Dunnett contrasts.

Value
Wilk’s Lambdas for MANOVA and multiple comparisons are returned.

Note
R > 2.13.2

Author(s)
Chen-An Tsai, James J. Chen, Ching-Wei Chang, and Chih-Yi Chien

References
Frank Bretz, Torsten Hothorn and Peter Westfall (2010), Multiple Comparison Using R, CRC Press, Boca Raton
Index

*Topic MANOVA test
   MAVTn, 6
*Topic OLS
   Tols, 9
*Topic Random forests
   MAVTp, 7
*Topic Wilks’ Lambda
   Wilksn, 10
*Topic one-sided test, MANOVA test, multiple comparisons, OLS test, gene set analysis, GST-plot
   MAVTgsa-package, 2

data, 3
design.matrix, 3

GS, 4
GSTplot, 4

Hott2, 5

ma.estimates, 6
MAVTgsa (MAVTgsa-package), 2
MAVTgsa-package, 2
MAVTn, 6
MAVTp, 7
minp, 9

Tols, 9

Wilksn, 10