Package 'TANOVA'

July 2, 2014

Version 1.0.0

Date 2010-05-05

Title Time Course Analysis of Variance for Microarray

Depends R (>= 2.3.0), MASS, splines

Author Baiyu Zhou <zhouby98@stanford.edu> and Weihong Xu <weihongx@stanford.edu>

Maintainer Weihong Xu <weihongx@stanford.edu>

Description Functions for performing analysis of variance on time course microarray data

License LGPL

URL Wing Wong's lab (http://www.stanford.edu/group/wonglab/) &
Stanford Genome Technology Center (http://gluegrant1.stanford.edu/~DIC/)

Repository CRAN

Date/Publication 2012-10-29 08:57:44

NeedsCompilation no

R topics documented:

`data.form` *Convert input data into appropriate format for TANOVA*

Description

This is an internal function to be called by `tanova` to format input data into appropriate format. Users should call `tanova` whenever is possible.

Usage

```
data.form(data, f1, f2, tp)
```

Arguments

`data`	data matrix (gene * array). Each row is a gene. Each column is an array. If data is longitudinal (e.g., time course measurements from patients), arrays from same experimental units (e.g. patient) should be next to each other (but not necessary in time course order.
`f1`	a vector of length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be numeric 1,2,3...
`f2`	a vector of length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be numeric 1,2,3... If there's only one factor, let f2=0.
`tp`	a vector of length equal to the number of arrays. Each entry indicates the time point for the corresponding array. The values of tp should be numeric 1,2,3 ... For non-time course data, let tp=0.

Value

The output is a list object.

`d`	data matrix. nrow=#genes*#time points, ncol=factor levels (#factor1*#factor2). Rows are ordered by gene name and then by time points.
`fc1`	a vector of length equal to the number of arrays divided by the number of time points. Each entry indicates the level of the first factor for the group of arrays with the same combination of factor levels. The values of fc1 should be numeric 1,2,3...
`fc2`	a vector of length equal to the number of arrays divided by the number of time points. Each entry indicates the level of the first factor for the group of arrays with the same combination of factor levels. The values of fc1 should be numeric 1,2,3...

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

Examples

```
##f1=rep(1:2, each=8)
##f2=rep(c(1,2,1,2), each=4)
##tp=rep(1:4, 4)
##data=matrix(rnorm(16*1000), nrow=1000, ncol=16)
##formated.data=data.form(data,f1,f2,tp)
```

---

design.matrix *Generate design matrix for two-way factorial analysis*

Description

This is an internal function to be called by ls.estimate to generate design matrix. Users should call ls.estimate whenever is possible.

Usage

```
design.matrix(f1,f2)
```

Arguments

f1	a vector of length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be 1,2,3,...
f2	a vector of length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be 1,2,3,... If the experimental has only one factor, let f2=0.

Value

The output is a list object.

X0	???
Xa	???
Xb	???

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

ls.estimate

F.stat *Compute F-statistics for ANOVA model*

Description

This is an internal function to be called by tanova to generate design matrix. Users should call tanova whenever is possible.

Usage

```
F.stat(data,f1,f2,type,equal.size=FALSE,trim, eb=FALSE)
F.stat2(data,f1,f2,tp,type,trim,eb=FALSE)
```

Arguments

data	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively
f1	a vector containing the levels of a factor in each array
f2	a vector containing the levels of a factor in each array
tp	a vector with length equal to the number of arrays. Each entry indicates the time point for the corresponding array. tp takes values 1,2,3 .... For non-time course data, let tp=0.
type	type of test the null F-statistics is for, 1 for, 2 for, 3 for, 4 for
trim	the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
equal.size	a logical indicator of whether the number of replicates under each biological condition is equal. Default is FALSE.
eb	whether to use Bayesian prior

Value

F	observed F-statistics

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

Examples

```
##data=matrix(rnorm(10000,mean=6, sd=1),nrow=500, ncol=20)
##f1=rep(c(1,2), each=10)
##f2=rep(c(1,2), 10)
##F.stat.null(data,f1,f2,type=1,trim=0,equal.size=FALSE,eb=FALSE)
```

`F.stat.null` *Generation of null F-statistics by bootstrap method*

Description

This is an internal function to be called by `tanova` to generate null distribution of F-statistics. Users should call `tanova` whenever is possible.

Usage

```
F.stat.null(data,f1,f2,type,trim=0,B=100,equal.size=FALSE,eb=FALSE)
F.stat.null2(data,f1,f2,tp,type,B=100,trim=trim,eb=FALSE)
```

Arguments

`data`	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively
`f1`	a vector containing the levels of a factor in each array
`f2`	a vector containing the levels of a factor in each array
`tp`	a vector with length equal to the number of arrays. Each entry indicates the time point for the corresponding array. tp takes values 1,2,3 .... For non-time course data, let tp=0.
`type`	type of test the null F-statistics is for, 1 for, 2 for, 3 for, 4 for
`trim`	the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
`B`	number of bootstrap resampling
`equal.size`	a logical indicator of whether the number of replicates under each biological condition is equal. Default is FALSE.
`eb`	whether to use Bayesian prior

Value

`F.null`	null F-statistics, each column is a bootstrap sampling.

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

Examples

```
##data=matrix(rnorm(10000,mean=6, sd=1),nrow=500, ncol=20)
##f1=rep(c(1,2), each=10)
##f2=rep(c(1,2), 10)
##F.stat.null(data,f1,f2,type=1,trim=0,B=100,equal.size=FALSE,eb=FALSE)
```

---

fdr.table *TANOVA False Discovery Table*

Description

It's an internal function. The function compare ANOVA F-statistics to null distribution generated by Bootstrap resampling to estimate False Discovery Rate

Usage

```
fdr.table(obj)
```

Arguments

obj	List object containing at least two components: one is the F-statistics for ANOVA, the other is the bootstrap generated null distribution of F-statistics

Value

table	a matrix contains estimated false discovery rate at various cutoffs

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

group.ix *This is an internal function*

Description

This is an internal function.

Usage

```
group.ix(f1,f2)
```

Arguments

f1	a vector of length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be 1,2,3,...
f2	a vector of length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be 1,2,3,... If the experimental has only one factor, let f2=0.

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

ls.estimate *Least square estimation*

Description

This is an internal function.

Usage

```
ls.estimate(data,f1,f2)
```

Arguments

data	data matrix (gene * array). Each row is a gene. Each column is an array. If data are longitudinal (for example, time course measurements from patients), arrays from same experimental units (e.g. patient) should be adjacent to each other.
f1	a vector of length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be 1,2,3,...
f2	a vector of length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be 1,2,3,... If the experimental has only one factor, let f2=0.

Value

The output is a list object.

Mab	???
Ma	???
Mb	???

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongxu@gmail.com>

NANOVA.test *Non-parametric analysis of variance (NANOVA)*

Description

This are an internal functions to provide non-parametric ANOVA function.

Usage

```
NANOVA.test(data,f1,f2,type=2,B=100, robustify=FALSE,equal.size=FALSE,eb=FALSE)
NANOVA.test2(data,f1,f2,type,time.course,equal.size=FALSE,B=100,robustify=FALSE,eb=FALSE,df=0)
NANOVA.test3(data,f1,f2,tp,type=2,B=100,robustify=FALSE,eb=FALSE)
```

Arguments

data	data matrix (gene * array). Each row is a gene. Each column is an array. If data are longitudinal (for example, time course measurements from patients), arrays from same experimental units (e.g. patient) should be adjacent to each other.
f1	a vector with length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be 1,2,3,...
f2	a vector with length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be 1,2,3,... If the experimental has only one factor, let f2=0.
tp	a vector with length equal to the number of arrays. Each entry indicates the time point for the corresponding array. tp takes values 1,2,3 .... For non-time course data, let tp=0.
B	the number of bootstrap resampling. Default is 100. Large B lead to more accurate inference, but need more running time.
robustify	a logical indicator of whether a robust test statistic should be used. Default is FALSE.
equal.size	a logical indicator of whether the number of replicates under each biological condition is equal. Default is FALSE.

type	an indicator of TANOVA test type. 0: classifies genes into gene sets C1,C2, C3,C4 and C5 (constant genes). 1: test for interaction effect. 2: one-way NANOVA test. 3: test main effect f1. 4: test main effect f2.
eb	a logical indicator of whether Empirical Bayesian method should be used in the estimation of significance
df	degree of freedom
time.course	the number of time points we sampled

Value

Return list contains

gene.order	A numeric vector indicating the positions in which the genes are called significant for the test
F	observed F-statistics
F.null	bootstrap generated null F-statistics
pvalue	a numeric vector of the corresponding p-value of NANOVA.
delta	a numeric vector of summary statistic for NANOVA

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

---

prior.sigma *Compute the prior of covariance matrix*

Description

This is an internal function.

Usage

```
prior.sigma(Q,f1,f2,tp=0)
prior.SIGMA(data,f1,f2,time.course)
```

Arguments

Q	internal parameter
data	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively
f1	a vector containing the levels of a factor in each array
f2	a vector containing the levels of a factor in each array
tp	a vector containing the sampling time point of each array
time.course	the number of time points sampled

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

---

proj.data *Projection of Raw Data*

Description

Project the time course data onto the optimum direction

Usage

```
## S3 method for class 'data'
proj(data,time.course,a,...)
```

Arguments

data	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively. Columns are ordered by first by subject, then by time
time.course	the number of time points per subject
a	projection matrix
...	other arguments to be passed on

Value

d	projected data. Row and column represent gene (probe set) and subject respectively.

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

proj.dir *projection direction*

Description

This function is used to estimate the direction that maximize the ANOVA structure of interest.i.e., the direction that contains most useful signal

Usage

```
## S3 method for class 'dir'
proj(data,f1,f2,time.course,type,eb=FALSE,df=0,...)
## S3 method for class 'dir2'
proj(data, f1, f2, tp,type=2,trim=0,...)
```

Arguments

data	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively
f1	a vector containing the levels of a factor in each array
f2	a vector containing the levels of a factor in each array
time.course	the number of time points for which we took sample
tp	a vector containing the sampling time point of each array
type	an indicator of TANOVA test type. 0: classifies genes into gene sets C1,C2, C3,C4 and C5 (constant genes). 1: test for interaction effect. 2: one-way NANOVA test. 3: test main effect f1. 4: test main effect f2.
trim	the percentage to be trimmed from each end of the data
eb	a logical indicator of whether Empirical Bayesian method should be used in the estimation of significance
df	degree of freedom
...	other arguments to be passed on

Value

a	Estimated optimum projection direction (ANOVA direction)

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

sig.number *The number of significant genes in the FDR table at specified quantiles*

Description

Estimate the number of significant genes in the FDR table using user specified cutoff.

Usage

```
sig.number(fdr.table,FDR=0.05,qt=-1)
```

Arguments

fdr.table	a table containing the estimated FDR for each gene
FDR	the false discovery cutoff, the default is 0.05
qt	a vector of quantiles for which FDR is estimated by averaging over B times of number of false genes. If qt=0.25 (0.5, 0.75, 0.9), FDR is estimated by 25% (50%, 75%, 90%) quantiles of false genes over B times.Default is -1, which means all genes are used.

Value

The function returns a number showing the number of significant genes at specified FDR cutoff.

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

sigma.hat *Estimation of Covariance Matrix*

Description

Internal function

Usage

```
sigma.hat(y,f1,f2)
```

Arguments

y	a data matrix containing expression values. Row and column represent gene (probe set) and array respectively
f1	a vector containing the levels of a factor in each array
f2	a vector containing the levels of a factor in each array

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

---

tanova *Classification of genes by time course analysis of variance(TANOVA)*

Description

The method is useful to capture gene specific response during a time course and their dependency on multiple experimental factors. It is based on non-parametric ANOVA technique. It's applicable to both longitudinal and cross-sectional data. gene.classifier is the major function users should call. gene.classifier1/gene.classifier2/gene.classifier3 are internal functions called by gene.classifier for specific analysis situation.

Usage

```
tanova(data, f1, f2, tp, B=100, FDR=0.05,robustify=FALSE, equal.size=FALSE, qt=-1, longitudinal=TRUE,
gene.classifier1(data, f1, f2, B=100,FDR=0.05, robustify=FALSE, equal.size=FALSE, eb=FALSE,qt=-1)
gene.classifier2(data, f1, f2, B=100,FDR=0.05, robustify=FALSE, equal.size=FALSE, time.course, qt=-1,
gene.classifier3(data, f1, f2, tp, B=100,FDR=0.05, qt=-1,robustify=FALSE,eb=FALSE)
```

Arguments

data	data matrix (gene * array). Each row is a gene. Each column is an array. If data are longitudinal (for example, time course measurements from patients), arrays from same experimental units (e.g. patient) should be adjacent to each other.
f1	a vector with length equal to the number of arrays. Each entry indicates the level of the first factor for corresponding array. The values of f1 should be 1,2,3,...
f2	a vector with length equal to the number of arrays. Each entry indicates the level of the second factor for the corresponding array. The values of f2 should be 1,2,3,... If the experimental has only one factor, let f2=0.
tp	a vector with length equal to the number of arrays. Each entry indicates the time point for the corresponding array. tp takes values 1,2,3 .... For non-time course data, let tp=0.

B	the number of bootstrap resampling. Default is 100. Large B lead to more accurate inference, but need more running time.
FDR	false discovery rate (FDR) for each test. Default is 0.05.
robustify	a logical indicator of whether a robust test statistic should be used. Default is FALSE.
equal.size	a logical indicator of whether the number of replicates under each biological condition is equal. Default is FALSE.
qt	a vector of quantiles for which FDR is estimated by averaging over B times of number of false genes. If qt=0.25 (0.5, 0.75, 0.9), FDR is estimated by 25% (50%, 75%, 90%) quantiles of false genes over B times.Default is -1, which means all genes are used.
longitudinal	a logical indicator of whether the data is longitudinal. The default is TRUE
eb	a logical indicator of whether Empirical Bayesian method should be used in the estimation of significance
test.type	an indicator of TANOVA test type. 0: classifies genes into gene sets C1,C2, C3,C4 and C5 (constant genes). 1: test for interaction effect. 2: one-way NANOVA test. 3: test main effect f1. 4: test main effect f2.
df	degree of freedom
time.course	the number of time points we sampled

Value

Depends on the test.type user specified, the output will be different. If test.type=0, the output is a list object containing four classes of genes that response to factors differently determined by tanova. They are described below:

`C1, C1.delta, C1.pvalue, a1`
: the gene index, test statistic, p-value and projection vector of genes that show significant interaction effect of two treatment factors (f1*f2)

`C2, C2.delta, C2.pvalue, a2`
: the gene index, test statistic, p-value and projection vector of genes that show significant additive effect of two treatment factors (f1+f2)

`C3, C3.delta, C3.pvalue, a3`
: the gene index, test statistic, p-value and projection vector of genes that show significant effect of treatment factor #1 (f1)

`C4, C4.delta, C4.pvalue, a4`
: the gene index, test statistic, p-value and projection vector of genes that show significant effect of treatment factor #2 (f2)

If test.type=1,2,3,4. Only one of the above class will be returned. The items in the output list are described as follows:

genes	A numeric vector indicating the positions in which the genes are called significant for the test
pvalue	a numeric vector of the corresponding p-value of TANOVA.
delta	a numeric vector of summary statistic for non-parametric ANOVA

a	a matrix containing projection direction (gene by time point)
dir	a list object returned by proj.dir
obj	a list object returned by NANOVA.test

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

NANOVA.test

Examples

```
##f1=rep(1:2, each=8)
##f2=rep(c(1,2,1,2), each=4)
##tp=rep(1:4, 4)
##data=matrix(rnorm(16*1000), nrow=1000, ncol=16)
##result=gene.classifier(data,f1,f2,tp)
```

---

trigammaInverse *Trigamma Inverse Function*

Description

This is an internal function.

Usage

```
trigammaInverse(x)
```

Arguments

x	internal object

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

tanova

`z.score` *Z Score*

Description

This is an internal function for calculating Z-score by comparing observed F-statistics to bootstrap generated null F-statistics.

Usage

```
z.score(F, F.null)
```

Arguments

`F`	Observed F statistics
`F.null`	Bootstrap generated null F-statistics

Author(s)

Baiyu Zhou <zhouby98@stanford.edu> & Weihong xu <weihongx@stanford.edu>

See Also

`tanova`

Index