Package 'HybridMTest'

November 14, 2025

Type Package

Title Hybrid Multiple Testing

Version 1.54.0

Date 2011-06-14

Author Stan Pounds <stanley.pounds@stjude.org>, Demba Fofana <demba.fofana@stjude.org>

Maintainer Demba Fofana <demba.fofana@stjude.org>

Depends R (>= 2.9.0), Biobase, fdrtool, MASS, survival

Description Performs hybrid multiple testing that incorporates method selection and assumption evaluations into the analysis using empirical Bayes probability (EBP) estimates obtained by Grenander density estimation. For instance, for 3-group comparison analysis, Hybrid Multiple testing considers EBPs as weighted EBPs between F-test and H-test with EBPs from Shapiro Wilk test of normality as weigth. Instead of just using EBPs from F-test only or using H-test only, this methodology combines both types of EBPs through EBPs from Shapiro Wilk test of normality. This methodology uses then the law of total EBPs.

License GPL Version 2 or later

Imports stats

LazyLoad yes

biocViews GeneExpression, Genetics, Microarray

git_url https://git.bioconductor.org/packages/HybridMTest

git_branch RELEASE_3_22

git_last_commit 4dada28

git_last_commit_date 2025-10-29

Repository Bioconductor 3.22

Date/Publication 2025-11-13

Contents

HybridMTest-package 2
correlation.data 3
grenander.ebp 4

GroupComp.data 5
hybrid.test 6
row.kgrp.shapiro 7
row.kruskal.wallis 8
row.oneway.anova 9
row.pearson 10
row.slr.shapiro 11
row.spearman 12

Index **14**

`HybridMTest-package` *A powerful tool in gene expression hypothesis testing.*

Description

This package enables users to generalize the assumption adequacy averaging (AAA) procedure proposed by Pounds and Rai (2009). AAA uses empirical Bayes methodology (Efron et al 2001) to simultaneously evaluate assumptions for each hypothesis test, select the best hypothesis testing procedure for each hypothesis test, and adjust for multiple testing.

Details

Package:	HybridMTest
Type:	Package
Version:	1.0
Date:	2010-07-24
License:	GPL (>=2)
LazyLoad:	yes

The main function is hybrid.test. The users may use existing row.test functions (such as row.T.test) or utilize their own row.test functions with similar input and output structures.

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>
Maintainer: Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<dfofana@yahoo.com>>

References

Pounds SB, Rai SN. (2009) Assumption Adequacy Averaging as a Concept for Developing More Robust Methods for Differential Gene Expression Analysis. Computational Statistics and Data Analysis, 53: 1604-1612 .

B. Efron, R. Tibshirani, J.D. Storey and V. Tusher, Empirical Bayes analysis of a microarray experiment. Journal of American Statistical Association, 96 (2001), pp. 1151-1160.

Examples

```
####################Correlation Data##############
# load data
data(correlation.data)
# Read the expression values
Y<-exprs(correlation.data)
head(Y)
# Read the phenotype
x<-pData(correlation.data)
####################Three group comparison Data####
# load data
data(GroupComp.data)
# Read the expression values
brain.express.set <- exprs(GroupComp.data)
head(brain.express.set)
# Read the phenotype
brain.pheno.data <- pData(GroupComp.data)
```

`correlation.data` *Sample ExpressionSet object of correlation.data*

Description

An *ExpressionSet* object of correlation.data.

Usage

```
data(correlation.data)
```

Details

correlation.data includes expression data for 100 randomly selected genes for the subjects of the AML97 clinical trial in Lamba et al (2010). The phenotype is the log-transformed baseline DNA synthesis rate. The data set is included for an example of exploring the correlation of expression with a quantitative phenotype.

Value

`expr(correlation.data)`
: A matrix with 100 rows and 83 columns with rows representing probe-sets and cloumns of human sample IDs.

`pData(correlation.data)`
: A data frame with 100 rows and 2 columns. Each row represents one human sample. Column *id* is the column without title and *x* is the phenotype observations transformed by log function.

References

Lamba JK et al. Identification of predictive markers of cytarabine response in acute myeloid leukemia by integrative analysis of gene-expression profiles with multiple phenotypes. Pharmacogenomics, 12: 327-39.

See Also

ExpressionSet-class

hybrid.test; GroupComp.data

Examples

```
data(correlation.data)
tumor.express.set <- exprs(correlation.data)
tumor.pheno.data <- pData(correlation.data)
```

---

grenander.ebp *Grenander EBP.*

Description

computes Grenander EBP

Usage

```
grenander.ebp(p)
```

Arguments

p	is a vector of p-values.

Value

pval.pdf	vector of Grenander PDF estimates corresponding to the input vector of p-values
ebp.null	vector of Grenander EBP estimates corresponding to the input vector of p-values

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Langaas, M., Lindqvist, B., 2005. Estimating the proportion of true null hypotheses, with application to DNA microarray data. J.R. Statist. Soc. B67, part4, 555-572. Strimmer, K. 2008. A unified approach to false discovery rate estimation. BMC Bioinformatics 9: 303. Strimmer, K. 2008. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics 24: 1461-1462.

Examples

```
####################Grenander estimation ####################
# load data  and compute grenander
p<-rbeta(1000,0.8,1) # Grenander example p-values
gren.res<-grenander.ebp(p) # Compute grenander results
```

GroupComp.data *Sample ExpressionSet of GroupComp.data*

Description

An *ExpressionSet* object of GroupComp.data

Usage

```
data(GroupComp.data)
```

Details

GroupComp.data is an *ExpressionSet* object with the expression data from Johnson et al (2010) for 100 randomly selected genes. The expression data was collected from 83 subjects with ependymoma defined by anatomic subclass (grp1 = posterior fossa, grp2 = supratentorial, grp3 = spinal). The data set is used for an example of using the package for a comparison of expression across 3 groups.

Value

`expr(GroupComp.data)`
: A matrix with 100 rows and 83 columns with rows representing probe-sets and cloumns of human sample IDs.

`pData(GroupComp.data)`
: A data frame with 83 rows and 2 columns. Each row represents one human sample. Column *id* is the human sample ID and *sppfst.grps* is the assigned sample group label.

References

R. Johnson et al.(2010) Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature*, 466: 632-6.

See Also

ExpressionSet-class

`hybrid.test`; `correlation.data`

Examples

```
data(GroupComp.data)
brain.express.set <- exprs(GroupComp.data)
brain.pheno.data <- pData(GroupComp.data)
```

hybrid.test *Hybrid Multiple Testing of Gene Expression Data*

Description

This function allows users to generalize the assumption adequacy averaging method of Pounds and Rai (2009). Multiple hypothesis testing methods are applied to the data of each gene. Additionally, for each gene, assumptions of those methods are formally evaluated with a statistical test. For each gene, the results of the statistical evaluation of model assumptions (e.g. normality) are used to define weights to combine or select the results of the tests of primary interest (e.g. differential expression). For example, the Shapiro-Wilk (1965) test may be applied to residuals to test the assumption of normality and used to select the t-test or rank-sum test for differential expression analysis of each gene.

Usage

```
hybrid.test(express.set, test.specs, ebp.def=NULL)
```

Arguments

express.set	A Bioconductor ExpressionSet (add a link) object. The AssayData component contains the normalized log-expression data and the phenoData component contains the phenotype, biological condition, or treatment data.
test.specs	A data.frame that describes the statistical tests to be performed. Each row gives details about one statistical test to be applied to the data. For each test, provide a label, the name of the R function that performs the test, the x argument of the function, and the opts (options). All entries are to be character strings.
ebp.def	A data.frame that describes how the statistical test results are to be combined to give the final hybrid test result. It has a column 'wght' and a column 'mthd'. Each row defines one term in the sum of probabilities for the final hybrid test result. The EBP will be the sum of the product of the wght EBP and the mthd EBP.

Value

Returns a data.frame, each row giving results for one gene. The columns include the test statistic, p-value, and EBP for each test applied to the data, the final hybrid.test EBP defined by the weighted average (wgt.mean.ebp), the p-value from the empirically best test (best.pval), and the EBP computed from best.pval.

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>> or <<dfofana@yahoo.com>>

References

Pounds, S., Rai, S., 2009. Assumption adequacy averaging as a concept for developing more robust methods for differential gene expression analysis. Computational Statistics and Data Analysis 53, 1604-1612. S.S. Shapiro and M.B. Wilk, An analysis of variance test for normality (complete samples). Biometrika, 52 (1965), pp. 591-611.

Examples

```
####################Correlation Study#####################
# load data
data(correlation.data)
# Read the expression values
Y<-exprs(correlation.data)
# Read the phenotype
x<-pData(correlation.data)
# Create test Spectrum
 test.specs<-cbind.data.frame(label=c("pearson","spearman","shapiro"),
                              func.name=c("row.pearson","row.spearman","row.slr.shapiro"),
                               x=rep("x",3),
                               opts=rep("",3))
# Specify the tests
ebp.def<-cbind.data.frame(wght=c("shapiro.ebp","(1-shapiro.ebp)"),
                          mthd=c("pearson.ebp","spearman.ebp"))
# Perform the Hybrid test
corr.res<-hybrid.test(correlation.data,test.specs,ebp.def)
head(corr.res)

####################Three group comparison###################
# load data
data(GroupComp.data)
# Read the expression values
brain.express.set <- exprs(GroupComp.data)
head(brain.express.set)
# Read the phenotype
brain.pheno.data <- pData(GroupComp.data)
brain.pheno.data
# Specify the tests
test.specs<-cbind.data.frame(label=c("anova","kw","shapiro"),
                   func.name=c("row.oneway.anova","row.kruskal.wallis","row.kgrp.shapiro"),
                             x=rep("grp",3),
                             opts=rep("",3))

# Define the final ebp
ebp.def<-cbind.data.frame(wght=c("shapiro.ebp","(1-shapiro.ebp)"),
                          mthd=c("anova.ebp","kw.ebp"))

#Perform the Hybrid test
Kgrp.res<-hybrid.test(GroupComp.data,test.specs,ebp.def)
```

`row.kgrp.shapiro` *Shapiro Wilk test of normality.*

Description

For each row of the expression matrix Y, use Shapiro-Wilks test to determine whether the residuals of one-way ANOVA (with groups defined by x) are normally distributed.

Usage

```
row.kgrp.shapiro(Y, x)
```

Arguments

Y	the data matrix with variables in rows and observations (subjects) in columns
x	x is vector of group labels

Value

A data.frame with three columns

stat	a vector with the Shapiro-Wilk test statistic for each row of Y
pval	a vector with the Shapiro-Wilk p-value for each row of Y
ebp	a vector with the estimated empirical Bayes probability of normality for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Patrick Royston (1982) An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics, 31, 115-124. Patric Royston (1992) Algorithm As 181: The W test for Normality. Applied Statistics, 31, 176-180. Patric Royston (1995) Remarks As R94: A remark on Algorithm AS 181:The W test for normality. Applied Statistics, 44, 547-551

Examples

```
####################Three group comparison###################
# load data
data(GroupComp.data)
# Read the expression values
brain.express.set <- exprs(GroupComp.data)
head(brain.express.set)
# Read the phenotype
brain.pheno.data <- pData(GroupComp.data)
brain.pheno.data[,1]
#Shapiro Test of Normality
row.kgrp.shapiro(brain.express.set,brain.pheno.data[,1] )
```

---

row.kruskal.wallis *Apply the Kruskal-Wallis test many times*

Description

For each row of Y, use the Kruskal-Wallis test to compare medians across groups defined in grplbl

Usage

```
row.kruskal.wallis(Y, grplbl)
```

Arguments

Y	data matrix with variables in rows and observations (subjects) in columns
grplbl	vector of group labels

Details

The alternative hypothesis is that, for each gene, there are at least two groups of different median. The null hypothesis is that all groups have the same median for each gene studied.

Value

A data.frame with three columns

stat	a vector with the test-statistic for each row of Y
pval	a vector with the p-value for each row of Y
ebp	a vector with the empirical Bayes probability of equal medians for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Kruskal, W.H. and W.A. Wallis (1952) Use of ranks in one-criterion variance analysis. J. Amer. Stat. Assoc. 47:583-621.

Examples

```
####################Three group comparison###################
# load data
data(GroupComp.data)
# Read the expression values
brain.express.set <- exprs(GroupComp.data)
head(brain.express.set)
# Read the phenotype
brain.pheno.data <- pData(GroupComp.data)
brain.pheno.data[,1]
row.kruskal.wallis(brain.express.set,brain.pheno.data[,1])
```

---

row.oneway.anova *Perform one-way ANOVA for many variables.*

Description

For each row of Y, use one-way ANOVA to compare means across groups defined by grplbl.

Usage

```
row.oneway.anova(Y, grplbl)
```

Arguments

Y	data matrix with variables in rows and subjects in columns
grplbl	vector of group labels for the subjects

Details

The alternative hypothesis is that, for each gene, there are at least two groups of different mean. The null hypothesis is that all groups have the same mean for each gene studied.

Value

A data.frame with three columns:

stat	a vector with the ANOVA F-statistic for each row of Y
pval	a vector with the ANOVA p-value for each row of Y
ebp	a vector with the empirical Bayes probability of equal means for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Chambers, J. M. and Hastie, T. J. (1992) Statistical Models in S, Wadsworth & Brooks/Cole.

Examples

```
####################Three group comparison###################
# load data
data(GroupComp.data)
# Read the expression values
brain.express.set <- exprs(GroupComp.data)
head(brain.express.set)
# Read the phenotype
brain.pheno.data <- pData(GroupComp.data)
brain.pheno.data[,1]
# ANOVA test
row.oneway.anova(brain.express.set,brain.pheno.data[,1])
```

---

row.pearson *Compute the Pearson correlation of a variable x with many variables in a matrix Y*

Description

For each row of a data matrix Y, compute the Pearson correlation with the variable x.

Usage

```
row.pearson(Y, x)
```

Arguments

Y	A data matrix with rows for variables and columns for subjects.
x	a vector with the variable to be correlated with each variable of Y

Value

A data.frame with three columns:

stat	a vector with the Pearson correlation for each row of Y
pval	a vector with the p-value for each row of Y
ebp	a vector with the empirical Bayesian probability that the correlation is zero for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) The New S Language. Wadsworth & Brooks/Cole.

Examples

```
####################Correlation Study#####################
# load data
data(correlation.data)
# Read the expression values
Y<-exprs(correlation.data)
# Read the phenotype
x<-pData(correlation.data)
x[,1]
#Pearson Test
row.pearson(Y,x[,1])
```

---

row.slr.shapiro *Test normality of residuals for many variables.*

Description

For each row of the data matrix Y, use the Shapiro-Wilk test to determine if the residuals of simple linear regression on x are normally distributed.

Usage

```
row.slr.shapiro(Y, x)
```

Arguments

Y	a data matrix with rows for variables and columns for subjects
x	a vector with values of the independent variables for regression of each row of Y.

Value

A data.frame with three columns:

stat	A vector with the Shapiro-Wilk test statistic for each row of Y
pval	A vector with the Shapiro-Wilk p-value for each row of Y
ebp	A vector with the estimated empirical Bayes probability of normally distributed residuals for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Patrick Royston (1982) Algorithm AS 181: The W test for Normality. Applied Statistics, 31, 176-180.

Examples

```
####################Correlation Study#####################
# load data
data(correlation.data)
# Read the expression values
Y<-exprs(correlation.data)
# Read the phenotype
x<-pData(correlation.data)
x[,1]
#Shapiro Test
row.slr.shapiro(Y,x[,1])
```

`row.spearman` *Compute Spearmans rank-based correlation of many variables with a variable x*

Description

For each row of the data matrix Y, compute its Spearman correlation with x.

Usage

```
row.spearman(Y, x)
```

Arguments

Y	a data matrix with rows for variables and columns for subjects
x	a vector of the variable to be associated with each row of Y

Value

A data.frame with three components:

stat	a vector with the Spearman correlation for each row of Y
pval	a vector with the p-value for each row of Y
ebp	a vector with the estimated empirical Bayes probability of zero correlation for each row of Y

Author(s)

Stan Pounds <<stanley.pounds@stjude.org>>; Demba Fofana <<demba.fofana@stjude.org>>

References

Spearman, C. (1904) The proof and measurement of association between two things. Amer. J. Psychol. 15:72-101.

Examples

```
####################Correlation Study#####################
# load data
data(correlation.data)
# Read the expression values
Y<-exprs(correlation.data)
# Read the phenotype
x<-pData(correlation.data)
x[,1]
#Spearman Test
row.spearman(Y,x[,1])
```

Index

∗ **Grenander estimation**
grenander.ebp, 4
∗ **Hybrid Multiple testing**
hybrid.test, 6
∗ **Multiple comparison**
row.kruskal.wallis, 8
row.oneway.anova, 9
row.pearson, 10
row.spearman, 12
∗ **Test for normality**
row.kgrp.shapiro, 7
row.slr.shapiro, 11
∗ **package**
HybridMTest-package, 2
∗
row.kruskal.wallis, 8
row.oneway.anova, 9

correlation.data, 3, *5*

ExpressionSet, *4, 5*

grenander.ebp, 4
GroupComp.data, *4*, 5

hybrid.test, *4, 5*, 6
HybridMTest-package, 2

row.kgrp.shapiro, 7
row.kruskal.wallis, 8
row.oneway.anova, 9
row.pearson, 10
row.slr.shapiro, 11
row.spearman, 12