

Package ‘corTest’

October 12, 2022

Type Package

Title Robust Tests for Equal Correlation

Version 1.0.7

Author Danyang Yu, Weiliang Qiu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kellan Tantisira, Scott T. Weiss

Maintainer Danyang Yu <dyu33@jhu.edu>

Description There are 6 novel robust tests for equal correlation. They are all based on logistic regressions. The score statistic U is proportion to difference of two correlations based on different types of correlation in 6 methods. The ST1() is based on Pearson correlation. ST2() improved ST1() by using median absolute deviation. ST3() utilized type M correlation and ST4() used Spearman correlation. ST5() and ST6() used two different ways to combine ST3() and ST4(). We highly recommend ST5() according to the article titled "New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs" published in Scientific Reports. Please see the reference: Yu et al. (2019) <[doi:10.1038/s41598-019-40167-8](https://doi.org/10.1038/s41598-019-40167-8)>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports MASS, graphics, stats,igraph,clusterGeneration,Matrix,Biobase, methods, ggplot2

NeedsCompilation no

RoxygenNote 7.1.1

Repository CRAN

Date/Publication 2020-09-25 15:30:02 UTC

R topics documented:

construct_network	2
es	4
fisher_transfer_test	5

generate_data	6
genEset	7
ghdist	8
plotDiffCor	9
st1	11
st2	12
st3	13
st4	14
st5	15
st6	16

Index	17
--------------	-----------

construct_network	<i>Construct Differential Correlation Network</i>
-------------------	---

Description

Construct differential correlation network with expressionSet, st5 is recommend for testing equal correlation.

Usage

```
construct_network(es,
                 cor_method = "st5",
                 var.grp,
                 pseudo_adjust_cutoff = FALSE,
                 pAdjMethod = 'fdr',
                 cutoff = 0.05,
                 nPseudo = 25)
```

Arguments

es	an ExpressionSet object of microRNA dataset
cor_method	a string represents the method for equal correlation, 'st5' is recommend.
var.grp	character. phenotype variable name indicating case-control status, 0 as control, 1 as case.
pseudo_adjust_cutoff	if the value is TRUE, pseudo probes will be used for setting the cutoff of p-value for differential correlation test. Otherwise, adjusted p-value will be compared with cutoff.
pAdjMethod	if pAdjMethod='none', the function will not do multiple testing adjustment. If pAdjMethod="fdr"/"BH"/"BY"/"holm"/"hochberg"/"hommel"/"bonferroni"/"BH"/"BY", the specific method will be used for adjusting p-value. pAdjMethod will only be used when pseudo_adjust_cutoff=FALSE
cutoff	if p value is smaller than the cutoff, there will be an edge between the two nodes. cutoff will only be used when pseudo_adjust_cutoff=FALSE.

nPseudo if pseudo_adjust_cutoff=TRUE, then nPseudo genes will be randomly chosen to produce pseudo genes (i.e., non-differentially correlated with other genes between cases and controls). We will use the minimum of the p-values of these pseudo genes as the cutoff of p-values for differential correlation analysis.

Value

A list with 6 elements:

my_graph	obtained network as igraph object
my_dat	obtained network as data frame with 3 columns: edge_id, node_id1, node_id2
pvalMat	raw p-values for testing differential correlation for each pair of genes
pAdjMat	adjusted p-values for testing differential correlation for each pair of genes
pvalPseudo	p-values for testing differential correlation between pseudo genes and other genes
alpha1	cutoff for p-values for testing differential correlation

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
set.seed(1234567)
res = generate_data(n1 = 50, n2 = 60, p1 = 5, p2 = 50)
es = res$es
print(es)

covCtrl = res$covCtrl
covCase = res$covCase

# we expect cov for 1st 5 genes are different between cases and controls
print(round(covCtrl[1:5, 1:5], 2))
print(round(covCase[1:5, 1:5], 2))

# we expect cov for other genes are same between cases and controls
print(round(covCtrl[6:10, 6:10], 2))
print(round(covCase[6:10, 6:10], 2))

res2 = construct_network(es = es,
                        cor_method = "st5",
                        pseudo_adjust_cutoff = FALSE,
                        var.grp = "grp",
```

```
pAdjMethod = 'fdr',  
cutoff = 0.05,  
nPseudo = 25)  
  
print(res2$graph)  
print(res2$network_dat)
```

es

A BioConductor ExpressionSet Object Storing Gene Expression Data

Description

A BioConductor ExpressionSet object storing gene expression data.

Usage

```
data("es")
```

Format

The expression data are stored in a matrix, which contains 55 row genes and 110 column subjects.

The phenotype data contains two variables: `sid` (subject ID) and `grp` (case-control status). `grp=1` indicates the subject is a case. `grp=0` indicates the subject is a control.

The feature data contains two variables: `geneid` (gene ID) and `memGenes`. `memGenes=1` indicates the gene is differentially correlated between cases and controls with at least one other gene. `memGenes=0` indicates the gene is not differentially correlated with other genes.

Details

The data contains 50 cases and 60 controls. The first 5 genes are differentially correlated with at least one other genes, while the remaining 50 genes are non-differentially correlated with other genes.

The covariance matrix of the first 5 genes in cases were generated by the function `rcorrmatrix` in R package `clusterGeneration`. The covariance matrix of the first 5 genes in controls are identical matrix.

The covariance matrix of the remaining 50 genes in cases were generated by the function `rcorrmatrix`. The covariance matrix of the 50 genes in controls were the same as that in cases.

The first 50 subjects are cases. The remaining 60 subjects are controls.

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu(corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
data(es)

print(es)
```

fisher_transfer_test *Test for equal correlation*

Description

Compute p-value with Fisher's Z-transformation test. If biasCorrection is true, the corrected correlation is used. The formula is $\text{rho.corrected} = \text{rho} - \text{rho}/(2*(n-1))$.

Usage

```
fisher_transfer_test(x1, z1, x0, z0, biasCorrection = TRUE)
```

Arguments

x1	a numeric vector
z1	a numeric vector with same length as x1
x0	a numeric vector
z0	a numeric vector with same length as x0
biasCorrection	a boolean value

Value

p-value of test for testing if correlation between x1 and z1 is the same as that between x0 and z0

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu(corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = fisher_transfer_test(x1, z1, x0, z0)
print(p)
```

 generate_data

 Generate Data for Control Subjects and Case Subjects

Description

The function is to generate expression level matrixes of control subjects and case subjects. X matrix is for control subjects with the default sample size $n_1=50$. Z matrix is for case subjects with the default sample size $n_2=60$. X is generated from multivariate normal distribution $N(0, \text{SigmaX})$, where SigmaX is a block matrix $((\text{SigmaP1}, 0), (0, \text{SigmaP2}))$, sigmaP1 is the $p_1 \times p_1$ matrix and SigmaP2 is the $p_2 \times p_2$ matrix. Z is generated from multivariate normal distribution $N(0, \text{SigmaZ})$, where SigmaZ is a block matrix $((E_P1, 0), (0, \text{SigmaP2}))$ and E_P1 is $p_1 \times p_1$ identity matrix.

Usage

```
generate_data(n1, n2, p1, p2)
```

Arguments

n1	a numeric value representing the sample size of control subjects
n2	a numeric value representing the sample size of case subjects
p1	a numeric value representing the dimension of sigma_1
p2	a numeric value representing the dimension of sigma_0

Value

A list with 3 elements:

es	An ExpressionSet object contains gene expression data.
covCase	Covariance matrix of all genes for case subjects.
covCtrl	Covariance matrix of all genes for control subjects.

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
library(Biobase)
set.seed(1234567)
res = generate_data(n1 = 50, n2 = 60, p1 = 5, p2 = 50)
es = res$es
print(es)

# gene expression data
dat = exprs(es)
print(dim(dat))
print(dat[1:2,1:3])

# phenotype data
pDat = pData(es)
print(dim(pDat))
print(pDat[1:2,])
print(table(pDat$grp, useNA = "ifany"))

# feature data
fDat = fData(es)
print(dim(fDat))
print(fDat[1:2,])
print(table(fDat$memGenes, useNA = "ifany"))

covCtrl = res$covCtrl
covCase = res$covCase

# we expect cov for 1st 5 genes are different between cases and controls
print(round(covCtrl[1:5, 1:5], 2))
print(round(covCase[1:5, 1:5], 2))

# we expect cov for other genes are same between cases and controls
print(round(covCtrl[6:10, 6:10], 2))
print(round(covCase[6:10, 6:10], 2))
```

genEset

Generate an ExpressionSet Object

Description

Generate an ExpressionSet object.

Usage

```
genEset(ex, pDat, fDat = NULL, annotation = "")
```

Arguments

ex	a data matrix stores gene expression data. Rows are genes and columns are subjects.
pDat	a data frame stores phenotype data. Rows are subjects and columns are variables describing subjects. Row names of pDat must be identical to the column names of ex.
fDat	a data frame stores feature data. Rows are genes and columns are variables describing genes. Row names of fDat must be identical to the row names of ex.
annotation	character. indicates R Bioconductor annotation package that provides annotation for genes.

Value

A BioConductor ExpressionSet object.

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

ghdist

generating variables from g-and-h distribution

Description

generating variables from g-and-h distribution

Usage

```
ghdist(n, g = 0, h = 0)
```

Arguments

n	the number of the variables you want to generate
g	the parameter g of g-and-h distribution
h	the parameter h of g-and-h distribution

Value

n variables generated from g-and-h distribution

Examples

```
x = ghdist(50, 0.2, 0.2)
print(x)
```

`plotDiffCor`*Scatter Plot of 2 Genes for Cases and Controls*

Description

Scatter plot of 2 genes for cases and controls, superimposed with linear regression lines.

Usage

```
plotDiffCor(x1,  
            z1,  
            x0,  
            z0,  
            pval = NULL,  
            xlab = "gene1",  
            ylab = "gene2",  
            title = "scatter plots")
```

Arguments

<code>x1</code>	numeric. vector of gene expression for gene 1 for cases.
<code>z1</code>	numeric. vector of gene expression for gene 2 for cases.
<code>x0</code>	numeric. vector of gene expression for gene 1 for controls.
<code>z0</code>	numeric. vector of gene expression for gene 2 for controls.
<code>pval</code>	numeric. p-value for testing differential correlation of the 2 genes between cases and controls.
<code>xlab</code>	character. label for x-axis.
<code>ylab</code>	character. label for y-axis.
<code>title</code>	character. plot title.

Value

A list with 4 elements:

<code>g</code>	A ggplot2 object.
<code>dat</code>	a data frame with 3 variables: x, z, and grp.
<code>coef1</code>	a vector of length two giving the intercept and slope of linear regression for cases.
<code>coef0</code>	a vector of length two giving the intercept and slope of linear regression for controls.

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu(corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
library(Biobase)
set.seed(1234567)
res = generate_data(n1 = 50, n2 = 60, p1 = 5, p2 = 50)
es = res$es
print(es)

# gene expression data
dat = exprs(es)
print(dim(dat))
print(dat[1:2,1:3])

# 3rd gene
x = dat[3,]
# 5th gene
z = dat[5,]

# for cases
x1 = x[which(es$grp == 1)]
z1 = z[which(es$grp == 1)]

# for controls
x0 = x[which(es$grp == 0)]
z0 = z[which(es$grp == 0)]

# st5
res2 = st5(x1 = x1, z1 = z1, x0 = x0, z0 = z0)
pval = res2$pval

plotDiffCor(x1 = x1,
            z1 = z1,
            x0 = x0,
            z0 = z0,
            pval = pval,
            xlab = "gene3",
            ylab = "gene5",
            title = "scatter plots"
            )
```

`st1`*Test Differential Correlation Using st1 Method*

Description

Compute p-value for the equal correlation test with Pearson correlation based on a logistic regression model corresponding to two independent groups

Usage

```
st1(x1, z1, x0, z0)
```

Arguments

<code>x1</code>	a numeric vector
<code>z1</code>	a numeric vector with same length as <code>x1</code>
<code>x0</code>	a numeric vector
<code>z0</code>	a numeric vector with same length as <code>x0</code>

Value

<code>stat</code>	Test statistic for testing if correlation between <code>x1</code> and <code>z1</code> is the same as that between <code>x0</code> and <code>z0</code>
<code>p-value</code>	p-value of test
<code>signedStat</code>	Signed test statistic

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu(corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = st1(x1, z1, x0, z0)
print(p)
```

`st2`*Test Differential Correlation Using st2 Method*

Description

Compute p-value for the equal correlation test with mad-replacing-Pearson correlation based on a logistic regression model corresponding to two independent groups

Usage

```
st2(x1, z1, x0, z0)
```

Arguments

<code>x1</code>	a numeric vector
<code>z1</code>	a numeric vector with same length as <code>x1</code>
<code>x0</code>	a numeric vector
<code>z0</code>	a numeric vector with same length as <code>x0</code>

Value

<code>stat</code>	Test statistic for testing if correlation between <code>x1</code> and <code>z1</code> is the same as that between <code>x0</code> and <code>z0</code>
<code>p-value</code>	p-value of test
<code>signedStat</code>	Signed test statistic

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. *Scientific Reports* 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = st2(x1, z1, x0, z0)
print(p)
```

Description

Compute p-value for the equal correlation test with percentage bend correlation based on a logistic regression model corresponding to two independent groups

Usage

```
st3(x1, z1, x0, z0)
```

Arguments

x1	a numeric vector
z1	a numeric vector with same length as x1
x0	a numeric vector
z0	a numeric vector with same length as x0

Value

stat	Test statistic for testing if correlation between x1 and z1 is the same as that between x0 and z0
p-value	p-value of test
signedStat	Signed test statistic

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu(corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. Scientific Reports 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = st3(x1, z1, x0, z0)
print(p)
```

`st4`*Test Differential Correlation Using st4 Method*

Description

Compute p-value for the equal correlation test with Spearman correlation based on a logistic regression model corresponding to two independent groups

Usage

```
st4(x1, z1, x0, z0)
```

Arguments

<code>x1</code>	a numeric vector
<code>z1</code>	a numeric vector with same length as <code>x1</code>
<code>x0</code>	a numeric vector
<code>z0</code>	a numeric vector with same length as <code>x0</code>

Value

<code>stat</code>	Test statistic for testing if correlation between <code>x1</code> and <code>z1</code> is the same as that between <code>x0</code> and <code>z0</code>
<code>p-value</code>	p-value of test
<code>signedStat</code>	Signed test statistic

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. *Scientific Reports* 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = st4(x1, z1, x0, z0)
print(p)
```

`st5`*Test Differential Correlation Using st5 Method*

Description

```
# Compute p-value for the equal correlation test with combination of Spearman correlation and percentage bend correlation based on a logistic regression model corresponding to two independent groups
```

Usage

```
st5(x1, z1, x0, z0)
```

Arguments

<code>x1</code>	a numeric vector
<code>z1</code>	a numeric vector with same length as <code>x1</code>
<code>x0</code>	a numeric vector
<code>z0</code>	a numeric vector with same length as <code>x0</code>

Value

<code>stat</code>	Test statistic for testing if correlation between <code>x1</code> and <code>z1</code> is the same as that between <code>x0</code> and <code>z0</code>
<code>p-value</code>	p-value of test
<code>signedStat</code>	Signed test statistic

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. *Scientific Reports* 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
p = st5(x1, z1, x0, z0)
print(p)
```

st6

*Test Differential Correlation Using st6 Method***Description**

Compute p-value for the equal correlation test with another way to combine Spearman correlation and percentage bend correlation based on a multiple logistic regression model corresponding to two independent groups

Usage

```
st6(x1, z1, x0, z0)
```

Arguments

x1	a numeric vector
z1	a numeric vector with same length as x1
x0	a numeric vector
z0	a numeric vector with same length as x0

Value

A list with 3 elements:

u	the vector of Fisher score.
pval	p-value of test for testing if correlation between x1 and z1 is the same as that between x0 and z0
cov_u	covariance matrix for u.

Author(s)

Danyang Yu <dyu33@jhu.edu>, Weiliang Qiu <weiliang.qiu@gmail.com>

References

Danyang Yu, Zeyu Zhang, Kimberly Glass, Jessica Su, Dawn L. DeMeo, Kelan Tantisira, Scott T. Weiss, Weiliang Qiu (corresponding author). New Statistical Methods for Constructing Robust Differential Correlation Networks to characterize the interactions among microRNAs. *Scientific Reports* 9, Article number: 3499 (2019)

Examples

```
x1 = ghdist(n = 100, g = 0.2, h = 0.2)
x0 = ghdist(n = 100, g = 0.2, h = 0.2)
z1 = x1 + ghdist(n = 100, g = 0.2, h = 0.2)
z0 = x0 + ghdist(n = 100, g = 0.2, h = 0.2)
res = st6(x1, z1, x0, z0)
print(res)
```


Index

- * **correlation**
 - construct_network, 2
 - fisher_transfer_test, 5
 - generate_data, 6
 - st1, 11
 - st2, 12
 - st3, 13
 - st4, 14
 - st5, 15
 - st6, 16
- * **datasets**
 - es, 4
- * **distribution**
 - ghdist, 8
- * **generation**
 - ghdist, 8
- * **method**
 - genEset, 7
- * **test**
 - construct_network, 2
 - fisher_transfer_test, 5
 - generate_data, 6
 - st1, 11
 - st2, 12
 - st3, 13
 - st4, 14
 - st5, 15
 - st6, 16

construct_network, 2

es, 4

fisher_transfer_test, 5

generate_data, 6

genEset, 7

ghdist, 8

plotDiffCor, 9