Package ‘metaGroup8’

May 6, 2019

Type Package

Title Pooled p-value Calculator

Version 1.0

Date 2019-05-06

Author

Ma Zhiyan

Sun Jiashuo

Sun Oliver

Zhao Qiang

Depends None

Description Pooled p-value calculator will check for normality of the dataframe by group, perform the appopriate statistical test and return a vector of p-values, one for each biomarker in the dataframe.

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-05-06 17:00:00 EST

beststatest *Select the Best Statistical Test*

Description

beststatest will check for normality of the dataframe by group, perform the appopriate statistical test and return a vector of p-values, one for each biomarker in the dataframe.

Usage

beststatest(onedataframe)

Arguments

onedataframe - the dataframe which contains group membership as the first column and biomarker measurements as the remaining columns

Details

If the dataframe is normal and there are > 2 groups it will perform one way ANOVA to determine if there is group difference. If the biomarker is not normally distributed, it will use the Kruskal Wallis test (this is the non-parametric version of one way ANOVA). If there are = 2 groups it will perform two sample t-test to determine if there is group difference (also checking for variance). If the biomarker is not normally distributed, use the Wilcoxon rank sum test. Returns a vector of p-values, one for each biomarker in the dataframe.

Authors

Ma Zhiyan

Sun Jiashuo

Sun Oliver

Zhao Qiang

Examples

```
## Create a dataframe

set.seed(123)

p <- 100

data1 <- data.frame(group=sample(1:3,200,replace=TRUE),

matrix(rnorm(p*200),ncol=p))

## pass in data1 dataframe arg into beststatest and assign to p.values

p.values = beststatest(data1)
```

poolp *Pooled p-values for each biomarker across K Data*

Description

Implements Fisher, Stouffer, Minimum p-value (minP), Maximum p-value (maxP) methods for pooling p-values

Usage

poolp(p, method = c("fisher", "stouffer", "minP", "maxP"))

Arguments

p - the vector of p-values to be pooled

method - selection of which method to use when pooling p-values

Authors

Ma Zhiyan

Sun Jiashuo

Sun Oliver

Zhao Qiang

Examples

```
## generate vector of p-values
p = runif(100)

## call poolp with p-values vector and method selection
poolp(p, "fisher")
```

colcheck *Check for valid columns*

Description

Input data and columns are checked to see if they are valid. Return 1 if valid else return 0.

Usage

colcheck(a, b, c = NULL, d = NULL, e = NULL)

Arguments

a – first dataframe

b – second dataframe

c – third dataframe (optional)

d – fourth dataframe (optional)

e – fifth dataframe (optional)

Details

Checks if the number of columns in the data frames are all equal to one another.

Authors

Ma Zhiyan

Sun Jiashuo

Sun Oliver

Zhao Qiang

Examples

```
## generate two dataframes
set.seed(123)
p <- 100
## dataframe 1
data1 <- data.frame(group=sample(1:3,200,replace=TRUE), matrix(rnorm(p*200),ncol=p))
## dataframe 2
data2 <- data.frame(group=sample(1:2,150,replace=TRUE), matrix(rnorm(p*150),ncol=p))
## check if the columns are valid
colcheck(data1, data2)
```

metaGroup8 *Main Function*

Description

Takes in the dataframes and method of pooling p-values as arguments and returns pooled p-values

Usage

metaGroup8(a, b, c = NULL, d = NULL, e = NULL, method = c("fisher", "stouffer", "minP", "maxP"))

Arguments

a – first dataframe

b – second dataframe

c – third dataframe (optional)

d – fourth dataframe (optional)

e – fifth dataframe (optional)

method - Choose between fisher, stouffer, minP, and maxP methods of pooling p-values

Details

metaGroup8 first performs a column check by calling colcheck(). If the columns are valid then it will perform the best statistical test to investigate if there is any group difference for each biomarker within each data frame by calling beststatest(). Lastly the p-values for each biomarker across the different data frames are pooled and returned to the user.

Authors

Ma Zhiyan

Sun Jiashuo

Sun Oliver

Zhao Qiang

Examples

```
## generate two dataframes
set.seed(123)
p <- 100
## dataframe 1
data1 <- data.frame(group=sample(1:3,200,replace=TRUE), matrix(rnorm(p*200),ncol=p))
## dataframe 2
data2 <- data.frame(group=sample(1:2,150,replace=TRUE), matrix(rnorm(p*150),ncol=p))

## get the pooled p-values
metaGroup8(data1, data2, method="stouffer")
```