

Package ‘MetaP’

Christina Hartnett, Matthew Brulhardt, Joshua Foote, Natalie Murawski

May 7th 2019

OVERVIEW

This package provides a method for pooling p-values of biomarkers using Fisher, Stouffer, Min or Max tests for the given multiple (2 to 5) data frames.

```
library(MetaP)
```

Workflow

Project

Description

This function uses different tests (Fisher, Stouffer, Min, Max) to pool data from different data frames.

Usage

```
Project(x1, x2, x3, x4, x5, test)
```

Arguments

x1, x2, x3, x4, x5 are data frames with each data frame consisting of identical number of columns. Must input at least 2 but at most 5 data frames.

Test takes values of:

- "fisher" for Fisher
- "stouffer" for Stouffer
- "min" for Minimum P-value
- "max" for Maximum P-value

Value

Returns a vector of p-values for all Biomarkers being pooled from our data frames.

GroupDifference

Description

This function determines our p-values based on different factors of the data frame. The test determines if the data is normally distributed then based on amount of groups will perform the proper test. For a data frame with 2 groups we will perform either a two sample t-test or Wilcoxon rank sum test. For a data frame with more than 2 groups we will perform ANOVA or Kruskal Wallis test.

Usage

GroupDifference(x)

Arguments

x is a data frame, can be normally distributed or not normally distributed. For 2 groups we will perform a two sample t-test or Wilcoxon rank sum test based on Normality. For more than 2 groups we will perform ANOVA or Kruskal Wallis test based on Normality.

Value

Returns a list of p-values

check.input

Description

This checks our inputs for our main function.

This function checks that:

- All inputs are lists/ Data Frames
- There are the same number of columns in each Data Frame
- That row 1 is group and rows 2 to p are biomarkers (p is length on data frame)
- The group membership column will take at least two unique values

Usage

check.input(frames)

Arguments

frames is a list of 2 to 5 Data Frames

pool.fisher

Description

This function sums the log-transformed p-values, following a chi-squared distributions with 2k degrees of freedom.

$$\chi^2_{\text{Fisher}} = -2 \sum_{i=1}^k \log(P_i)$$

Usage

pool.fisher(pvalues)

Arguments

pvalues is a list of p-values

Value

Returns a pooled p-value for the list of p-values given.

pool.stouffer

Description

This function sums the inverse normal p-values, following a standard normal distribution.

$$T_{\text{Stouffer}} = \sum_{i=1}^k z_i / \sqrt{k} (z_i \Phi^{-1}(p_i)),$$

Usage

pool.stouffer(pvalues)

Arguments

pvalues is a list of p-values.

Value

Returns a pooled p-value for the list of p-values given.

pool.min

Description

This function follows a beta distribution with degrees of freedom $\alpha = 1$ and $\beta = k$

Usage

`pool.min(pvalues)`

Arguments

pvalues is a list of p-values.

Value

Returns a pooled p-value for the list of p-values given.

pool.max

Description

This function follows a beta distribution with degrees of freedom $\alpha = k$ and $\beta = 1$

Usage

`pool.max(pvalues)`

Arguments

pvalues is a list of p-values.

Value

Returns a pooled p-value for the list of p-values given.

REFERENCES

Lun-Ching Chang, Hui-Min Lin, et al. “Meta-Analysis Methods for Combining Multiple Expression Profiles: Comparisons, Statistical Characterization and an Application Guideline. *BMC Bioinformatics*, BioMed Central, 21 Dec. 2013, bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-368.