Package 'mazeinda'

May 9, 2022

Title Monotonic Association on Zero-Inflated Data

Version 0.0.2

Author Alice Albasi [aut, cre]

Maintainer Alice Albasi <albasialice@gmail.com>

Description Methods for calculating and testing the significance of pairwise monotonic association from and based on the work of Pimentel (2009) <doi:10.4135/9781412985291.n2>. Computation of association of vectors from one or multiple sets can be performed in parallel thanks to the packages 'foreach' and 'doMC'.

Depends R (>= 3.3.0)

Imports foreach

License GPL-3

Encoding UTF-8

RoxygenNote 6.0.1

Suggests doMC, gamlss.dist, knitr, testthat, R.rsp, rmarkdown

VignetteBuilder R.rsp, knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-09 07:10:07 UTC

R topics documented:

`associate` *Associate pairwise vectors form one or two sets*

Description

Given two matrices m_1 and m_2, computes all pairwise correlations of each vector in m_1 with each vector in m_2. Thanks to the package foreach, computation can be done in parallel using the desired number of cores.

Usage

```
associate(m1, m2, parallel = FALSE, n_cor = 1, estimator = "values", d1,
  d2, p11 = 0, p01 = 0, p10 = 0)
```

Arguments

`m1, m2`	matrices whose columns are to be correlated. If no estimation calculations are needed, default is NA.
`parallel`	should the computations for associating the matrices be done in parallel? Default is FALSE
`n_cor`	number of cores to be used if the computation is run in parallel. Default is 1
`estimator`	string indicating how the parameters p_{11}, p_{01}, p_{10}, p_{00} are to be estimated. The default is 'values', which indicates that they are estimated based on the entries of x and y. If estimates=='mean', each p_{ij} is estimated as the mean of all pairs of column vectors in m_1, and of m_2 if needed. If estimates=='own', the p_{ij}'s must be given as arguments.
`d1, d2`	sets of vectors used to estimate p_{ij} parameters. If just one set is needed set d_1=d_2.
`p11`	probability that a bivariate observation is of the type (m,n), where m,n>0.
`p01`	probability that a bivariate observation is of the type (0,n), where n>0.
`p10`	probability that a bivariate observation is of the type (n,0), where n>0.

Details

To find pairwise monotonic associations of vectors within one set m, run associate(m,m). Note that the values on the diagonal will not be necessarely 1 if the vectors contain 0's, as it can be seen by the formula $p_{11}^2 t_{11} + 2 * (p_{00}p_{11} - p_{01}p_{10})$

Value

matrix of correlation values.

Examples

```
v1=c(0,0,10,0,0,12,2,1,0,0,0,0,0,1)
v2=c(0,1,1,0,0,0,1,1,64,3,4,2,32,0)
associate(v1,v2)
m1=matrix(c(0,0,10,0,0,12,2,1,0,0,0,0,0,1,1,64,3,4,2,32,0,0,43,54,3,0,0,3,20,1),6)
associate(m1,m1)
m2=matrix(c(0,1,1,0,0,0,1,1,64,3,4,2,32,0,0,43,54,3,0,0,3,20,10,0,0,12,2,1,0,0),6)
associate(m1,m2)
```

`combine` *combine*

Description

Designed to combine the matrix of correlation values with the matrix of p-values so that in the cases when the null hypothesis cannot be rejected with a level of confidence indicated by the significance, the correlation is set to zero. Thanks to the package foreach, computation can be done in parallel using the desired number of cores.

Usage

```
combine(m1, m2, sl = 0.05, parallel = FALSE, n_cor = 1,
  estimator = "values", d1, d2, p11 = 0, p01 = 0, p10 = 0)
```

Arguments

`m1, m2`	matrices whose columns are to be correlated. If no estimation calculations are needed, default is NA.
`sl`	level of significance for testing the null hypothesis. Default is 0.05.
`parallel`	should the computations for associating the matrices be done in parallel? Default is FALSE
`n_cor`	number of cores to be used if the computation is run in parallel. Default is 1
`estimator`	string indicating how the parameters p_{11}, p_{01}, p_{10}, p_{00} are to be estimated. The default is 'values', which indicates that they are estimated based on the entries of x and y. If estimates=='mean', each p_{ij} is estimated as the mean of all pairs of column vectors in m_1, and of m_2 if needed. If estimates=='own', the p_{ij}'s must be given as arguments.
`d1, d2`	sets of vectors used to estimate p_{ij} parameters. If just one set is needed set d_1=d_2.
`p11`	probability that a bivariate observation is of the type (m,n), where m,n>0.
`p01`	probability that a bivariate observation is of the type (0,n), where n>0.
`p10`	probability that a bivariate observation is of the type (n,0), where n>0.

Details

To test pairwise monotonic associations of vectors within one set m, run combine(m,m). Note that the values on the diagonal will not be necessarily significant if the vectors contain 0's, as it can be seen by the formula $p_{11}^2 t_{11} + 2 * (p_{00}p_{11} - p_{01}p_{10})$. The formula for the variance of the estimator proposed by Pimentel(2009) does not apply in case p_{11}, p_{01},p_{10}, p_{00} attain the values 0 or 1. In these cases the R function cor.test is used. Note that while independence implies that the estimator is 0, if the estimator is 0, it does not imply that the vectors are independent.

Value

matrix of combined association values and p-values.

`test_associations` *test_associations*

Description

To test pairwise monotonic associations of vectors within one set m, run test_associations(m,m). Note that the values on the diagonal will not be necessarily significant if the vectors contain 0's, as it can be seen by the formula $p_{11}^2 t_{11} + 2 * (p_{00}p_{11} - p_{01}p_{10})$. The formula for the variance of the estimator proposed by Pimentel(2009) does not apply in case p_{11}, p_{00}, p_{01}, p_{10} attain the values 0 or 1. In these cases the R function cor.test is used. Note that while independence implies that the estimator is 0, the estimator being 0 does not imply that the vectors are independent.

Usage

```
test_associations(m1, m2, parallel = FALSE, n_cor = 1,
  estimator = "values", d1, d2, p11 = 0, p01 = 0, p10 = 0)
```

Arguments

`m1, m2`	matrices whose columns are used to estimate the p_{ij} parameters. If no estimation calculations are needed, default is NA. Both are necessary if cross-correlating pairwise the vectors from two datasets.
`parallel`	should the computations for combiing the matrices be done in parallel? Default is FALSE.
`n_cor`	number of cores to be used if the computation is run in parallel. Default is 1.
`estimator`	string indicating how the parameters p_{11}, p_{01}, p_{10}, p_{00} are to be estimated. The default is 'values', which indicates that they are estimated based on the entries of x and y. If estimates=='mean', each p_{ij} is estimated as the mean of all pairs of column vectors in m_1, and of m_2 if needed. If estimates=='own', the p_{ij}'s must be given as arguments.
`d1, d2`	sets of vectors used to estimate p_{ij} parameters. If just one set is needed set d_1=d_2.
`p11`	probability that a bivariate observation is of the type (m,n), where m,n>0
`p01`	probability that a bivariate observation is of the type (0,n), where n>0.
`p10`	probability that a bivariate observation is of the type (n,0), where n>0.

Details

Given two matrices m_1 and m_2, computes all pairwise correlations of each vector in m_1 with each vector in m_2. Thanks to the package foreach, computation can be done in parallel using the desired number of cores.

Value

matrix of p-values of association.

Examples

```
v1=c(0,0,10,0,0,12,2,1,0,0,0,0,0,1)
v2=c(0,1,1,0,0,0,1,1,64,3,4,2,32,0)
test_associations(v1,v2)
m1=matrix(c(0,0,10,0,0,12,2,1,0,0,0,0,0,1,1,64,3,4,2,32,0,0,43,54,3,0,0,3,20,1),6)
test_associations(m1,m1)
m2=matrix(c(0,1,1,0,0,0,1,1,64,3,4,2,32,0,0,43,54,3,0,0,3,20,10,0,0,12,2,1,0,0),6)
test_associations(m1,m2)
m3= matrix(abs(rnorm(36)),6)
m4= matrix(abs(rnorm(36)),6)
test_associations(m3,m4)
```

Index