

Package: Taba (via r-universe)

September 11, 2024

Type Package

Title Taba Robust Correlations

Version 1.0.0

Imports robustbase, stats

Description Calculates the robust Taba linear, Taba rank (monotonic), TabWil, and TabWil rank correlations. Test statistics as well as one sided or two sided p-values are provided for all correlations. Multiple correlations and p-values can be calculated simultaneously across multiple variables. In addition, users will have the option to use the partial, semipartial, and generalized partial correlations; where the partial and semipartial correlations use linear, logistic, or Poisson regression to modify the specified variable.

License GPL-3

Encoding UTF-8

LazyData false

RoxygenNote 7.1.1

Suggests testthat

NeedsCompilation no

Author Mohammad Tabatabai [aut], Derek Wilus [aut, cre]

Maintainer Derek Wilus <dwilus@mmc.edu>

Repository CRAN

Date/Publication 2021-03-31 19:50:02 UTC

Contents

taba	2
taba.gpartial	3
taba.matrix	5
taba.partial	7
taba.test	8

Index	11
--------------	-----------

taba

Robust Correlation

Description

Returns the Taba robust linear, Taba rank (monotonic), TabWil, or TabWil rank correlation coefficient between two numeric vectors.

Usage

```
taba(x, y, method = c("taba", "tabarank", "tabwil", "tabwilrank"), omega)
```

Arguments

x	A numeric vector of length greater than 2 must be same length as y
y	A numeric vector of length greater than 2 must be same length as x
method	A character string of "taba", "tabarank", "tabwil", or "tabwilrank" determining if one wants to calculate Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation, respectively. If no method is specified, the function will output Taba Linear correlation.
omega	Numeric allowing the user to alter the tuning constant. If one is not specified, the function will default to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1.

Details

This function can be used to compare two non-empty numeric vectors of length greater than two, or two columns of a data frame or matrix composed of more than two numeric elements. Missing values in either x or y are deleted row-wise. The default method is Taba Linear correlation, with the tuning constant omega.

Value

This function returns a the robust linear or monotonic association between two numeric vectors as a numeric.

References

Tabatabai, M., Bailey, S., Bursac, Z. et al. An introduction to new robust linear and monotonic correlation coefficients. *BMC Bioinformatics* 22, 170 (2021). <https://doi.org/10.1186/s12859-021-04098-4>

doi: [10.1186/s12859021040984](https://doi.org/10.1186/s12859021040984)

See Also

[taba.test](#) for testing Taba linear or Taba rank (monotonic) correlations
[taba.partial](#) for partial and semipartial correlations
[taba.gpartial](#) for generalized partial correlations
[taba.matrix](#) for calculating correlation, p-value, and distance matrices

Examples

```
x = rnorm(100)
y = rnorm(100)
taba(x, y)
taba(x, y, method = "tabarank", omega = 0.4)
taba(x, y, method = "tabwil", omega = 0.22)
```

taba.gpartial *Generalized Taba Partial and Taba Rank Partial Correlation*

Description

Calculates a generalized partial correlation using one of the specified robust methods Taba linear or Taba rank correlation.

Usage

```
taba.gpartial(x, y, xcov, ycov, regress.x, regress.y,
             method = c("taba", "tabarank", "tabwil", "tabwilrank"),
             alternative = c("less", "greater", "two.sided"),
             omega)
```

Arguments

- x A numeric vector of length greater than 2 must be same length as y and covariates listed in x and ycov
- y A numeric vector of length greater than 2 must be same length as x and covariates listed in y and xcov
- xcov A data frame, matrix, or numeric vectors combined columnwise used as covariates for x, which have length equal to x
- ycov A data frame, matrix, or numeric vectors combined columnwise used as covariates for y, which have length equal to y
- regress.x A string variable "linear" for linear regression, "logistic" for binary logistic regression, and "poisson" for Poisson regression
- regress.y A string variable "linear" for linear regression, "logistic" for binary logistic regression, and "poisson" for Poisson regression

method	A character string of "taba", "tabarank", "tabwil", or "tabwilrank" determining if one wants to calculate Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation, respectively. If no method is specified, the function will output Taba Linear correlation.
alternative	Character string specifying the alternative hypothesis must be one of "less" for negative association, "greater" for positive association, or "two.sided" for difference in association. If the alternative is not specified, the function will default to a two sided test.
omega	Numeric allowing the user to alter the tuning constant. If one is not specified, the function will default to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1.

Details

This function generalizes the partial correlation. In the event that the controlling variables for x and y are identical, it reduces to Taba, Taba rank, TabWil, and TabWil rank partial correlation. Covariates used to control for x should be represented columnwise in a matrix or data frame as `xcov`. Similarly, covariates used to control for y should be represented columnwise in a matrix or data frame as `ycov`. When controlling an outcome variable with one covariate, a vector will suffice. Because x and y refer to the outcome variables, names of covariates (or control variables) must not be named "x" or "y". The user has the option of using different regression methods when controlling each outcome variable. Missing values in x , y , or any of the covariates are deleted row-wise. All categorical variables must be converted to type factor prior to using this function.

The default for this function is a two sided test using generalized partial Taba correlation using a linear regression to obtain residuals, with the tuning constant ω equal to 0.45.

Value

This function returns the robust association between two numeric vectors, adjusting for specified covariates. In addition, this function can provide the semipartial correlation, if specified.

References

Tabatabai, M., Bailey, S., Bursac, Z. et al. An introduction to new robust linear and monotonic correlation coefficients. *BMC Bioinformatics* 22, 170 (2021). <https://doi.org/10.1186/s12859-021-04098-4>

doi: [10.1186/s12859021040984](https://doi.org/10.1186/s12859021040984)

See Also

[taba](#) for calculating Taba linear or Taba rank (monotonic) correlations
[taba.test](#) for testing Taba linear or Taba rank Monotonic correlations
[taba.partial](#) for partial and semipartial correlations
[taba.matrix](#) for calculating correlation, p-value, and distance matrices

Examples

```
x = rnorm(100)
```

```

y = rnorm(100)
z1 = rnorm(100)
z2 = rnorm(100)
z3 = rnorm(100)
w = sample(c(0,1), replace=TRUE, size=100)
taba.gpartial(x, y, xcov = cbind(z1, z2), ycov = cbind(z1, z3), method = "tabarank")
taba.gpartial(x, y, z2, ycov = cbind(z1, z2), alternative = "less")
taba.gpartial(w, y, z1, cbind(z2, z3), regress.x = "logistic")

```

taba.matrix

Robust Correlation Matrix

Description

Calculates a correlation, distance, and p-value matrix using one of the specified robust methods Taba linear or Taba rank correlation.

Usage

```

taba.matrix(x, y = NULL, ..., method = c("taba", "tabarank", "tabwil", "tabwilrank"),
           alternative = c("less", "greater", "two.sided"),
           omega)

```

Arguments

x	A numeric vector of length greater than 2 must be same length as all other vectors.
y	A numeric vector of length greater than 2 must be same length as all other vectors.
...	Numeric vector(s) of length equal to x and y. May be of class matrix or data.frame, whose columns will be compared and whose column's length must be of equal length to x and y. Not one vector or column name can be "x" or "y."
method	A character string of "taba", "tabarank", "tabwil", or "tabwilrank" determining if one wants to calculate Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation, respectively. If no method is specified, the function will output Taba Linear correlation.
alternative	Character string specifying the alternative hypothesis must be one of "less" for negative association, "greater" for positive association, or "two.sided" for difference in association. If the alternative is not specified, the function will default to a two sided test.
omega	Numeric allowing the user to alter the tuning constant. If one is not specified, the function will default to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1.

Details

This function uses Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation to calculate the association of two or more numeric vectors. Numeric vectors under `...` are combined column-wise with `x` and `y`. When inserting a single matrix `x`, the function will calculate the correlation matrix using the columns of matrix `x`.

Matrices or data frames with numeric cells can be inserted in `...`, whereby each column in the matrix or data frame will be treated as a different vector for comparison. Columns must all have different names from each other. No vector or column should be named "x" or "y," as these refer to the first two vectors respectively, if inserted as a vector or matrix with no name. Missing values in any of the vectors are deleted row-wise.

The default for this function is a two sided test using Taba linear partial correlation, with the tuning constant ω equal to 0.45.

Value

This function returns the robust association between two or more numeric vectors, as a matrix; the distance matrix, as type `dist`; and a p-value matrix corresponding to the correlation matrix.

References

Tabatabai, M., Bailey, S., Bursac, Z. et al. An introduction to new robust linear and monotonic correlation coefficients. *BMC Bioinformatics* 22, 170 (2021). <https://doi.org/10.1186/s12859-021-04098-4>

doi: [10.1186/s12859021040984](https://doi.org/10.1186/s12859021040984)

See Also

[taba](#) for calculating Taba linear or Taba rank (monotonic) correlations
[taba.test](#) for testing Taba linear or Taba rank (monotonic) correlations
[taba.gpartial](#) for generalized partial correlations
[taba.partial](#) for partial and semipartial correlations

Examples

```
x = rnorm(100)
y = rnorm(100)
z1 = rnorm(100)
z2 = rnorm(100)
z3 = rnorm(100)
Z = cbind(z1,z3)
colnames(Z) = c("A","B")
taba.matrix(x, y, z1, z2, z3, method = "tabarank")
taba.matrix(x, y, z2, Z, alternative = "less", omega = 0.4)
taba.matrix(Z, method = "tabarank")
```

taba.partial *Robust Partial and Semipartial Correlation*

Description

Calculates a partial or semipartial correlation using one of the specified robust methods Taba linear or Taba rank correlation.

Usage

```
taba.partial(x, y, ..., regress, method = c("taba", "tabarank", "tabwil", "tabwilrank"),
            alternative = c("less", "greater", "two.sided"),
            semi = c("none", "x", "y"), omega)
```

Arguments

x	A numeric vector of length greater than 2 must be same length as y and covariates listed in ...
y	A numeric vector of length greater than 2 must be same length as x and covariates listed in ...
...	Numeric vectors used as covariates of length equal to x and y
regress	A string variable "linear" for linear regression, "logistic" for binary logistic regression, and "poisson" for Poisson regression
method	A character string of "taba", "tabarank", "tabwil", or "tabwilrank" determining if one wants to calculate Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation, respectively. If no method is specified, the function will output Taba Linear correlation.
alternative	Character string specifying the alternative hypothesis must be one of "less" for negative association, "greater" for positive association, or "two.sided" for difference in association. If the alternative is not specified, the function will default to a two sided test.
semi	A character string specifying which variable (x or y) should be adjusted.
omega	Numeric allowing the user to alter the tuning constant. If one is not specified, the function will default to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1.

Details

This function calculates the partial or semipartial association of two numeric vectors, or columns of a matrix or data frame composed of more than two numeric elements, adjusting for covariates of length equal to x and y. Covariates are combined column-wise and can be numeric vectors, matrices, or data frames with numeric cells. Each column in the matrix or data frame will be treated as a different covariate, and must have different names from x and y. Missing values in x, y, or any of the covariates are deleted row-wise. The default for this function is a two sided test using Taba linear partial correlation, with the tuning constant omega equal to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1. The variable you are not controlling must be continuous when using semipartial correlation.

Value

This function returns the robust association between two numeric vectors, adjusting for specified covariates. In addition, this function can provide the semipartial correlation, if specified.

References

Tabatabai, M., Bailey, S., Bursac, Z. et al. An introduction to new robust linear and monotonic correlation coefficients. *BMC Bioinformatics* 22, 170 (2021). <https://doi.org/10.1186/s12859-021-04098-4>

doi: [10.1186/s12859021040984](https://doi.org/10.1186/s12859021040984)

See Also

[taba](#) for calculating Taba linear or Taba rank (monotonic) correlations
[taba.test](#) for testing Taba linear or Taba rank (monotonic) correlations
[taba.gpartial](#) for generalized partial correlations
[taba.matrix](#) for calculating correlation, p-value, and distance matrices

Examples

```
x = rnorm(100)
y = rnorm(100)
z1 = rnorm(100)
z2 = rnorm(100)
z3 = rnorm(100)
taba.partial(x, y, z1, z2, z3, method = "tabwilrank")
taba.partial(x, y, z2, alternative = "less", semi = "x")
```

taba.test

Robust Correlation Test

Description

Tests the association between two numeric vectors using Taba robust linear, Taba rank (monotonic), TabWil, or TabWil rank correlation coefficient.

Usage

```
taba.test(x, y, method = c("taba", "tabarank", "tabwil", "tabwilrank"),
          alternative = c("less", "greater", "two.sided"),
          omega, alpha = 0.05)
```

Arguments

x	A numeric vector of length greater than 2 must be same length as y
y	A numeric vector of length greater than 2 must be same length as x
method	A character string of "taba", "tabarank", "tabwil", or "tabwilrank" determining if one wants to calculate Taba linear, Taba rank (monotonic), TabWil, or TabWil rank correlation, respectively. If no method is specified, the function will output Taba Linear correlation.
alternative	Character string specifying the alternative hypothesis must be one of "less" for negative association, "greater" for positive association, or "two.sided" for difference in association. If the alternative is not specified, the function will default to a two sided test.
omega	Numeric allowing the user to alter the tuning constant. If one is not specified, the function will default to 0.45 for Taba and Taba rank, and 0.1 for TabWil and TabWil rank. Range is between 0 and 1.
alpha	Type I error rate. Numeric must be between 0 and 1. Default set to 0.05.

Details

This function tests the association of two non-empty numeric vectors of length greater than two, or two columns of a data frame or matrix composed of more than two numeric elements. Covariates are combined column-wise and can be numeric vectors, matrices, or data frames with numeric cells. Each column in the matrix or data frame will be treated as a different covariate, and must have different names. Missing values in either x or y are deleted row-wise. The two sided test with the null hypothesis correlation is equal to zero. The default is a two sided test using Taba Linear correlation, with tuning constant omega.

Value

This function returns the robust linear or monotonic association between two numeric vectors, along with it's respective test statistic, and p-value.

References

Tabatabai, M., Bailey, S., Bursac, Z. et al. An introduction to new robust linear and monotonic correlation coefficients. BMC Bioinformatics 22, 170 (2021). <https://doi.org/10.1186/s12859-021-04098-4>

doi: [10.1186/s12859021040984](https://doi.org/10.1186/s12859021040984)

See Also

- [taba](#) for calculating Taba linear or Taba rank (monotonic) correlations
- [taba.partial](#) for partial and semipartial correlations
- [taba.gpartial](#) for generalized partial correlations
- [taba.matrix](#) for calculating correlation, p-value, and distance matrices

Examples

```
x = rnorm(10)
y = rnorm(10)
taba.test(x, y)
taba.test(x, y, method = "tabarank", alternative = "less")$p.value
taba.test(x, y, method = "tabwil", omega = .1)
```

Index

`taba`, 2, 4, 6, 8, 9

`taba.gpartial`, 3, 3, 6, 8, 9

`taba.matrix`, 3, 4, 5, 8, 9

`taba.partial`, 3, 4, 6, 7, 9

`taba.test`, 3, 4, 6, 8, 8