Package ‘EFA.MRFA’

October 12, 2022

Type Package

Title Dimensionality Assessment Using Minimum Rank Factor Analysis

Version 1.1.2

Date 2021-06-15

Author David Navarro-Gonzalez, Urbano Lorenzo-Seva

Maintainer David Navarro-Gonzalez <david.navarro@urv.cat>

Description Performs parallel analysis (Timmerman & Lorenzo-Seva, 2011 <doi:10.1037/a0023353>) and hull method (Lorenzo-Seva, Timmerman, & Kiers, 2011 <doi:10.1080/00273171.2011.564527>) for assessing the dimensionality of a set of variables using minimum rank factor analysis (see ten Berge & Kiers, 1991 <doi:10.1007/BF02294464> for more information). The package also includes the option to compute minimum rank factor analysis by itself, as well as the greater lower bound calculation.

Depends R (>= 2.10)

Imports stats, psych, scales, PCovR, ggplot2, reshape2

License GPL-3

Encoding UTF-8

NeedsCompilation no

LazyData true

RoxygenNote 6.0.1

Repository CRAN

Date/Publication 2021-06-16 08:00:02 UTC

R topics documented:

`EFA.MRFA-package` *Dimensionality Assesment using Minimum Rank Factor Analysis (MRFA)*

Description

Package for performing Parallel Analysis using Minimum Rank Factor Analysis (MRFA) . It also include a function to perform the MRFA only and another function to compute the Greater Lower Bound step for estimating the variables communalities.

Details

For more information about the methods used in each function, please go to each main page.

Value

`parallelMRFA`	Performs Parallel Analysis using Minimum Rank Factor Analysis (MRFA).
`hullEFA`	Performs Hull analysis for assessing the number of factors to retain.
`mrfa`	Performs Minimum Rank Factor Analysis (MRFA) procedure.
`GreaterLowerBound`	Estimates the communalities of the variables from a factor model.

Author(s)

David Navarro-Gonzalez

Urbano Lorenzo-Seva

References

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association, 76, 354-362. doi: 10.1080/01621459.1981.10477654

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull Method for Selecting the Number of Common Factors. Multivariate Behavioral Research, 46(2), 340-364. doi: 10.1080/00273171.2011.564527

ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. Psychometrika, 56(2), 309-315. doi: 10.1007/BF02294464

Ten Berge, J.M.F., Snijders, T.A.B. & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. Psychometrika, 46, 201-213.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods, 16(2), 209-220. doi: 10.1037/a0023353

Examples

```
## Example 1:

## perform a Parallel Analysis using an example Database with only 5 random data sets and
## using the 90th percentile of distribution of the random data
parallelMRFA(IDAQ, Ndatsets=5, percent=90)

## For speeding purposes, the number of datasets have been largely reduced. For a proper
## use of parallelMRFA, we recommend to use the default Ndatsets value (Ndatsets=500)

#Example 2:

## Perform the Hull method defining the maximum number of dimensions to be tested by the
## Parallel Analysis + 1 rule, with Maximum Likelihood factor extraction method and CAF
## as Hull index.
hullEFA(IDAQ, extr = "ML")
```

GreaterLowerBound   *Greater Lower Bound step (glb)*

Description

Estimates the communalities of the variables from a factor model where the number of factors is the number with positive eigenvalues.

Usage

```
GreaterLowerBound(C, conv = 0.000001, T, pwarnings = FALSE)
```

Arguments

C	Covariance/correlation matrix to be used in the analysis.
conv	Convergence criterion for glb step. The default convergence criterion will be conv=0.000001 . If the user determine a specific value, this will prevail.
T	Random matrix for start (can be omitted). If provided, it has to be the same size than the matrix provided in the C argument.
pwarnings	Determines if the possible warnings occurred during the computation will be printed in the console.

Details

Code adapted from a MATLAB function by Jos Ten Berge based on Ten Berge, Snijders & Zegers (1981) and Ten Berge & Kiers (1991).

Value

gam	Optimal communalities for each variable

Author(s)

David Navarro-Gonzalez

Urbano Lorenzo-Seva

References

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. Psychometrika, 56, 309-315.

Ten Berge, J.M.F., Snijders, T.A.B. & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. Psychometrika, 46, 201-213.

Examples

```
## perform glb using the correlation matrix of the IDAQ dataset, and using severe convergence
## criterion.
GreaterLowerBound(cor(IDAQ), conv=0.000001)
```

`hullEFA` *Hull method for selecting the number of common factors*

Description

Performs the Hull method (Lorenzo-Seva, Timmerman, & Kiers, 2011), which aims to find a model with an optimal balance between model fit and number of parameters.

Usage

```
hullEFA(X, maxQ, extr = "ULS", index_hull = "CAF", display = TRUE, graph = TRUE,
        details = TRUE)
```

Arguments

`X`	Raw sample scores.
`maxQ`	Maximum of dimensions to be tested. By default it will be determined by the Parallel Analysis advised dimensions, but the user can define it manually.
`extr`	Extraction method, the two options available being: "ULS" (Unweigthed Least Squares, by default) and "ML" (Maximum Likelihood).
`index_hull`	The index that will be used for determining the number of dimensions. The available options are the following: "CAF", "CFI""RMSEA", being "CAF" by default.
`display`	Determines if the output will be displayed in the console, TRUE by default. If it is TRUE, the output is returned silently and if it is FALSE, the output is returned in the console.
`graph`	Request a plot representing the Hull curve.
`details`	If detailed table will be displayed, containing the factors outside the convex Hull.

Details

`hullEFA` is based on the procedure proposed by Lorenzo-Seva, Timmerman, & Kiers (2011) which is designed for assessing the dimensionality of a variable set. The hull heuristic was originally proposed by Ceulemans & Kiers (2006) in the context of model selection in multiway data analysis.

The hull analysis is performed in four main steps:

1. The range of factors to be considered is determined.
2. The goodness-of-fit of a series of factor solutions is assessed.
3. The degrees of freedom of the series of factor solutions is computed.
4. The elbow is located in the higher boundary of the convex hull of the hull plot.

The number of factors extracted in the solution associated with the elbow is considered the optimal number of common factors.

In the Lorenzo-Seva, Timmerman, & Kiers (2011) simulation study, the Hull method outperformed the other selected methods in recovering the corrrect number of major factors.

Value

`Matrix`	Matrix containing the results of the Hull method using for the selected index.
`n_factors`	Number of advised dimensions by the selected index.

Author(s)

David Navarro-Gonzalez

Urbano Lorenzo-Seva

References

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull Method for Selecting the Number of Common Factors. Multivariate Behavioral Research, 46(2), 340-364. doi: 10.1080/00273171.2011.564527

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. British Journal of Mathematical and Statistical Psychology, 59: 133–150. doi: 10.1348/000711005X64817

Examples

```
## Perform the Hull method defining the maximum number of dimensions to be tested by the
## Parallel Analysis + 1 rule, with Maximum Likelihood factor extraction method and CAF
## as Hull index.
hullEFA(IDAQ, extr = "ML")
```

`IDAQ` *IDAQ database*

Description

A database to be used as example in the functions included on `DA.MRFA` package. It contains the answers of 100 participants to IDAQ questionnaire (Ruiz-Pamies, Lorenzo-Seva, Morales-Vives, Cosi, Vigil-Colet, 2014), which was developed for assessing Physical, Verbal and Indirect aggression. The original questionnaire contains 27 Likert-items, ranging from 1 to 5.

Usage

```
data("IDAQ")
```

Format

A data frame with 100 observations and 23 variables measuring 3 different types of aggression (Physical, Verbal and Indirect).

Details

The original sample contains 27 items, because includes 4 Social Desirability markers, but for the purpose of the `DA.MRFA` functions, they had been removed. Also, the original sample contains 750 participants, and the following database only contains 100 for speeding purposes.

Source

More information about the questionnaire can be found at:

`http://psico.fcep.urv.cat/tests/idaq/en/descripcion.html`

References

Ruiz-Pamies, M., Lorenzo-Seva, U., Morales-Vives, F., Cosi, S., & Vigil-Colet, A. (2014). I-DAQ: a new test to assess direct and indirect aggression free of response bias. The Spanish Journal of Psychology, 17, E41. doi: 10.1017/sjp.2014.43

Examples

```
data(IDAQ)
```

`mrfa` *Minimum Rank Factor Analysis function*

Description

Performs Minimum Rank Factor Analysis (MRFA) procedure, proposed by Ten Berge & Kiers (1991).

Usage

```
mrfa(SIGMA, dimensionality = 1, random = 10, conv1, conv2, display = TRUE,
    pwarnings = FALSE)
```

Arguments

`SIGMA`	Covariance/correlation matrix to be used in the analysis.
`dimensionality`	Common factors used to find communality estimates. The value has to be between 0 and the number of items minus 1, being the default option: 1 dimension to be retained. If 0 is selected, a more strict convergence criterion will be used.
`random`	Number of random starts.
`conv1`	Convergence criterion for MRFA. The default convergence criterion will be conv1=0.0001 . If the user determine a specific value, this will prevail.
`conv2`	Convergence criterion for glb step. The default convergence criterion will be conv2=0.001 . If the user determine a specific value, this will prevail.
`display`	Determines if the output will be displayed in the console, TRUE by default. If it is TRUE, the output is returned silently and if it is FALSE, the output is returned in the console.
`pwarnings`	Determines if the possible warnings occurred during the computation will be printed in the console.

Value

`A`	Factor loading matrix
`Matrix`	Covariance/Correlation matrix with optimal communalities in the diagonal
`gam`	Optimal communalities for each variable

Author(s)

David Navarro-Gonzalez

Urbano Lorenzo-Seva

References

ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. Psychometrika, 56(2), 309-315. doi: 10.1007/BF02294464

Examples

```
## perform MRFA using the correlation matrix of the IDAQ dataset, and using the default
## convergence criterion for MRFA and glb step.
mrfa(cor(IDAQ), dimensionality=3)
```

---

parallelMRFA — *Parallel Analysis using Minimum Rank Factor Analysis (MRFA)*

Description

Performs Parallel Analysis using Minimum Rank Factor Analysis (MRFA).

Usage

```
parallelMRFA(X, Ndatsets = 500, percent = 95, corr= "Pearson", display = TRUE,
    graph = TRUE)
```

Arguments

`X`	Raw sample scores.
`Ndatsets`	Number of random datasets used to compute the random distribution of eigenvalues.
`percent`	Desired percentile of distribution of random eigenvalues (for example 95 for the 95th percentile) to be used as threshold.
`corr`	Determine if Pearson or Polychoric matrix will be used "Pearson": Computes Pearson correlation matrix "Polychoric": Computes Polychoric/Tetrachoric correlation matrix (heavy time consuming).
`display`	Determines if the output will be displayed in the console, TRUE by default. If it is TRUE, the output is returned silently and if it is FALSE, the output is returned in the console.
`graph`	Request a plot representing the percentage of explained variance by the real data, by the mean of the random data and using the percentile of distribution of random eigenvalues, defined in the percent argument.

Details

`parallelMRFA` is based on the procedure proposed by Timmerman and Lorenzo-Seva (2011) which is designed for assessing the dimensionality of a variable set. The principal advantage of using MRFA (Ten Berge & Kiers, 1991) instead the usual PCA extraction process is that the eigenvalues obtained from MRFA can be used to estimate the explained common variance per factor.

The eigenvalue sampling distribution is obtaining using a nonparametric approach: a permutation of the raw data (Buja & Eyuboglu, 1992). This approach is recommended for PA especially in cases where the observed data ditribution clearly deviates from normality.

If the matrix to analyze is not positive-defined, a smoothering procedure will be applied (Devlin, Gnanadesikan & Kettenring, 1981).

Value

- `Real_Data` A vector containing the percentage of explained variance by the real data for each factor
- `Mean_random` A vector containing the percentage of explained variance by the mean of random data for each factor
- `Percentile_random` A vector containing the percentage of explained variance by the percentile of distribution of random data for each factor
- `Number_factors_mean` The number of factors to be retained suggested comparing the real data with the mean of the random data
- `Number_factors_percentiles` The number of factors to be retained suggested comparing the real data with the percentile of distribution of the random data

Author(s)

David Navarro-Gonzalez

Urbano Lorenzo-Seva

References

Buja, A., & Eyuboglu, N. (1992). Remarks on Parallel Analysis. Multivariate Behavioral Research, 27(4), 509-540. doi: 10.1207/s15327906mbr2704_2

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association, 76, 354-362. doi: 10.1080/01621459.1981.10477654

ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. Psychometrika, 56(2), 309–315. doi: 10.1007/BF02294464

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods, 16(2), 209-220. doi: 10.1037/a0023353

Examples

```
## perform a Parallel Analysis using an example Database with only 10 random data sets and
## using the 90th percentile of distribution of the random data
parallelMRFA(IDAQ, Ndatsets=10, percent=90)

## For speeding purposes, the number of datasets have been largely reduced. For a proper
## use of parallelMRFA, we recommend to use the default Ndatsets value (Ndatsets=500)
```

Index