Package ‘cdmTools’

January 8, 2024

Type  Package
Title  Useful Tools for Cognitive Diagnosis Modeling
Version  1.0.5
Date  2024-01-04

Description  Provides useful tools for empirical Q-matrix estimation and validation, such as the Hull method (Nájera, Sorrel, de la Torre, & Abad, 2021, <doi:10.1111/bmsp.12228>) and the discrete factor loading method (Wang, Song, & Ding, 2018, <doi:10.1007/978-3-319-77249-3_29>). It also contains dimensionality assessment procedures for CDM, including parallel analysis and automated fit comparison as explored in Nájera, Abad, and Sorrel (2021, <doi:10.3389/fpsyg.2021.614470>). Other relevant methods and features for CDM applications, such as the restricted DINA model (Nájera et al., 2023; <doi:10.3102/10769986231158829>), the general nonparametric classification method (Chiu et al., 2018; <doi:10.1007/s11336-017-9595-4>), and corrected estimation of the classification accuracy via multiple imputation (Kreitchmann et al., 2022; <doi:10.3758/s13428-022-01967-5>) are also available. Lastly, the package provides some useful functions for CDM simulation studies, such as random Q-matrix generation and detection of complete/identified Q-matrices.

License  GPL-3
Depends  R (>= 3.6.0)
Imports  GDINA (>= 2.8.0), ggplot2 (>= 3.3.0), psych (>= 1.9.12), sirt (>= 3.9-4), parallel (>= 3.6.3), stats (>= 3.6.3), GPArotation (>= 2014.11-1), combinat (>= 0.0-8), fungible, foreach, doSNOW

URL  https://github.com/pablo-najera/cdmTools
BugReports  https://github.com/pablo-najera/cdmTools/issues
RoxygenNote  7.2.3
Encoding  UTF-8

Author  Pablo Nájera [aut, cre, cph], Miguel A. Sorrel [aut, cph], Francisco J. Abad [aut, cph], Rodrigo S. Kreitchmann [ctb], Kevin Santos [ctb]
CA.MI

Calculate corrected classification accuracy with multiple imputation

Description

This function calculates the test-, pattern-, and attribute-level classification accuracy indices based on integrated posterior probabilities from multiple imputed item parameters (Kreitchmann et al., 2022). The classification accuracy indices are the ones developed by Iaconangelo (2017) and Wang et al. (2015). It is only applicable to dichotomous attributes. The function is built upon the CA function from the GDINA package (Ma & de la Torre, 2020).

Usage

CA.MI(fit, what = "EAP", R = 500, n.cores = 1, verbose = TRUE, seed = NULL)

Arguments

- **fit**: An object of class RDINA or GDINA (Ma & de la Torre, 2020).
- **what**: What attribute estimates are used? The default is "EAP".
- **R**: Number of bootstrap samples and imputations. The default is 500.
- **n.cores**: Number of processors to use to speed up multiple imputation. The default is 2.
- **verbose**: Show progress. The default is TRUE.
- **seed**: A seed for obtaining consistent results. If NULL, no seed is used. The default is NULL.
**Value**

`CA.MI` returns an object of class `CA`, with a list of elements:

- `tau` Estimated test-level classification accuracy, see Iaconangelo (2017, Eq 2.2) (vector).
- `CCM` Conditional classification matrix, see Iaconangelo (2017, p. 13) (matrix).

**References**


**Examples**

```r
library(GDINA)
dat <- sim10GDINA$simdat[1:100,]
Q <- sim10GDINA$simQ
fit <- GDINA(dat = dat, Q = Q, model = "GDINA")
ca.mi <- CA.MI(fit)
ca.mi
```

**Description**

Empirical Q-matrix estimation based on the *discrete factor loading* method (Wang, Song, & Ding, 2018) as used in Nájera, Abad, and Sorrel (2021). Apart from the conventional dichotomization criteria, the procedure based on loading differences described in Garcia-Garzon, Abad, and Garrido (2018) is also available. Furthermore, the bagging bootstrap implementation (Xu & Shang, 2018) can be applied; it is recommended when working with small sample sizes. The `psych` package (Revelle, 2020) is used for estimating the required exploratory factor analysis (EFA).
Usage

\texttt{estQ(}
\begin{itemize}
  \item \texttt{r},
  \item \texttt{K},
  \item \texttt{n.obs = NULL},
  \item \texttt{criterion = "row"},
  \item \texttt{boot = FALSE},
  \item \texttt{efa.args = list(cor = "tet", rotation = "oblimin", fm = "uls")},
  \item \texttt{boot.args = list(N = 0.8, R = 100, verbose = TRUE, seed = NULL)}
\end{itemize}
\texttt{)}

Arguments

- **r**
  A correlation matrix or raw data (\texttt{matrix} or \texttt{data.frame}). If a correlation matrix is used, it must have dimensions $J$ items $\times J$ items. Please note that tetrachoric or polychoric correlations should be used when working with dichotomous or polytomous items, respectively. If raw data is used, it must have dimensions $N$ individuals $\times J$ items. Missing values need to be coded as \texttt{NA}.

- **K**
  Number of attributes to use.

- **n.obs**
  Number of individuals if \texttt{r} is a correlation matrix. If \texttt{n.obs} is provided, \texttt{r} will be treated as a correlation matrix. Use \texttt{NULL} if \texttt{r} is raw data. The default is \texttt{NULL}.

- **criterion**
  Dichotomization criterion to transform the factor loading matrix into the Q-matrix. The possible options include "row" (for row means), "col" (for column means), "loaddiff" (for the procedure based on loading differences), or a value between 0 and 1 (for a specific threshold). The default is "row".

- **boot**
  Apply the bagging bootstrap implementation? Only available if \texttt{r} is raw data. If \texttt{FALSE}, the EFA will be applied once using the whole sample size. If \texttt{TRUE}, several EFAs will be applied with different subsamples; the estimated Q-matrix will be dichotomized from the bootstrapped Q-matrix, but the EFA fit indices, factor loadings, and communalities will be computed from the EFA with the whole sample size. The default is \texttt{FALSE}.

- **efa.args**
  A list of arguments for the EFA estimation:

  - **cor**
    Type of correlations to use. It includes "cor" (for Pearson correlations) and "tet" (for tetrachoric/polychoric correlations), among others. See \texttt{fa} function from the \texttt{psych} R package for additional details. The default is "tet".

  - **rotation**
    Rotation procedure to use. It includes "oblimin", "varimax", and "promax", among others. An oblique rotation procedure is usually recommended. See \texttt{fa} function from the \texttt{psych} R package for additional details. The default is "oblimin".

  - **fm**
    Factoring method to use. It includes "uls" (for unweighted least squares), "ml" (for maximum likelihood), and "wls" (for weighted least squares), among others. See \texttt{fa} function from the \texttt{psych} R package for additional details. The default is "uls".

- **boot.args**
  A list of arguments for the bagging bootstrap implementation (ignored if \texttt{boot = FALSE}):
estQ returns an object of class estQ.

est.Q Estimated Q-matrix (matrix).

efa.loads Factor loading matrix (matrix).

efa.comm EFA communalities (vector).

efa.fit EFA model fit indices (vector).

boot.Q Bagging bootstrap Q-matrix before dichotomization. Only if boot = TRUE (matrix).

is.Qid Q-matrix identifiability information (list).

specifications Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas

References


Examples

library(GDINA)
dat <- sim30GDINA$simdat
Q <- sim30GDINA$simQ

#-----------------------------
# Using default specifications
genQ <- estQ(r = dat, K = 5)  # Estimate Q-matrix
sugQ1$est.Q <- orderQ(sugQ1$est.Q, Q)$order.Q  # Reorder Q-matrix attributes
mean(sugQ1$est.Q == Q)  # Check similarity with the generating Q-matrix

# Using the bagging bootstrap method
#------------------------------------
# In boot.args argument, R >= 100 is recommended (R = 20 is here used for illustration purposes)
sugQ2 <- estQ(r = dat, K = 5, boot = TRUE, boot.args = list(R = 20, seed = 123))  # Estimate Q-matrix
sugQ2$est.Q <- orderQ(sugQ2$est.Q, Q)$order.Q  # Reorder Q-matrix attributes
sugQ2$boot.Q  # Proportion of replicas a q-entry was specified in the estimated Q-matrix
mean(sugQ2$est.Q == Q)  # Check similarity with the generating Q-matrix

---

**genQ**

*Generate Q-matrix*

**Description**

Generates a Q-matrix. The criteria from Chen, Liu, Xu, & Ying (2015) and Xu & Shang (2018) can be used to generate identifiable Q-matrices. Only binary Q-matrix are supported so far. Useful for simulation studies.

**Usage**

```r
genQ(J, K, Kj, I = 2, min.JK = 3, max.Kcor = 1, Qid = "none", seed = NULL)
```

**Arguments**

- **J** Number of items.
- **K** Number of attributes.
- **Kj** A vector specifying the number (or proportion, if summing up to 1) of items measuring 1, 2, 3, ..., attributes. The first element of the vector determines the number (or proportion) of items measuring 1 attribute, and so on. See Examples.
- **I** Number of identity matrices to include in the Q-matrix (up to column permutation). The default is 2.
- **min.JK** Minimum number of items measuring each attribute. It can be overwritten by I, if I is higher than min.JK. The default is 3.
- **max.Kcor** Maximum allowed tetrachoric correlation among the columns to avoid overlapping (Nájera, Sorrel, de la Torre, & Abad, 2020). The default is 1.
- **Qid** Assure that the generated Q-matrix is generically identifiable. It includes "none" (for no identifiability assurance), "DINA", "DINO", or "others" (for other CDMs identifiability). The default is "none".
- **seed** A seed for obtaining consistent results. If NULL, no seed is used. The default is NULL.
### Value

`genQ` returns an object of class `genQ`.

- `gen.Q` The generated Q-matrix (matrix).
- `JK` Number of items measuring each attribute (vector).
- `Kcor` Tetrachoric correlations among the columns (matrix).
- `is.Qid` Q-matrix identifiability information (list).
- `specifications` Function call specifications (list).

### Author(s)

Pablo Nájera, Universidad Pontificia Comillas

### References


### Examples

```r
Kj <- c(15, 10, 0, 5) # 15 one-att, 10 2-atts, 0 3-atts, and 5 four-atts items
Q <- genQ(J = 30, K = 4, Kj = Kj, Qid = "others", seed = 123)
```

### Description

Attribute profile estimation using the *general nonparametric classification method* (GNPC; Chiu, Sun, & Bian, 2018). The GNPC can be considered as a robust alternative to the parametric G-DINA model with low sample sizes. The `AlphaNP` function from the `NPCD` package (Zheng & Chiu, 2019; Chiu, Sun, & Bian, 2018) using weighted Hamming distances is used to initiate the procedure.

### Usage

```r
GNPC(
  dat,
  Q,
  initiate = "AND",
  min.change = 0.001,
  maxitr = 1000,
  verbose = TRUE
)
```
Arguments

dat  A N individuals x J items (matrix or data.frame). Missing values need to be coded as NA. Caution is advised if missing data are present.

Q  A J items x K attributes Q-matrix (matrix or data.frame).

initiate  Should the conjunctive ("AND") or disjunctive ("OR") NPC be used to initiate the procedure? Default is "AND".

min.change  Minimum proportion of modified attribute profiles to use as a stopping criterion. Default is .001.

maxitr  Maximum number of iterations. Default is 1000.

verbose  Print information after each iteration. Default is TRUE.

Value

GNPC returns an object of class GNPC.

alpha.est  Estimated attribute profiles (matrix).

loss.matrix  The distances between the weighted ideal responses from each latent class (rows) and examinees’ observed responses (columns) (matrix).

eta.w  The weighted ideal responses for each latent class (rows) on each item (columns) (matrix).

w  The estimated weights, used to compute the weighted ideal responses (matrix).

n.ite  Number of iterations required to achieve convergence (double).

hist.change  Proportion of modified attribute profiles in each iteration (vector).

specifications  Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas

References


Examples

```r
library(GDINA)
Q <- sim30GDINA$simQ # Q-matrix
K <- ncol(Q)
J <- nrow(Q)
set.seed(123)
```
is.Qid <- data.frame(guessing = rep(0.1, J), slip = rep(0.1, J))
sim <- simGDINA(200, Q, GS)
simdat <- sim$dat # Simulated data
simatt <- sim$attribute # Generating attributes
fit.GNPC <- GNPC(simdat, Q) # Apply the GNPC method
ClassRate(fit.GNPC$alpha.est, simatt) # Check classification accuracy

is.Qid

Check whether a Q-matrix is identifiable

Description
Checks whether a Q-matrix fulfills the conditions for strict and generic identifiability according to Gu & Xu (2021).

Usage
is.Qid(Q, model)

Arguments

Q A J items x K attributes Q-matrix (matrix or data.frame).
model CDM to be considered. It includes "DINA", "DINO", or "others" (for other CDMs: e.g., G-DINA, A-CDM).

Value
is.Qid returns an object of class is.Qid.

strict Is the Q-matrix strictly identifiable? (logical).
generic Is the Q-matrix generically identifiable? (logical).
conditions Identifiability criteria and whether they are fulfilled or not (vector).
specifications Function call specifications (list).

Author(s)
Pablo Nájera, Universidad Pontificia Comillas
Miguel A. Sorrel, Universidad Autónoma de Madrid

References

Examples
Kj <- c(15, 10, 0, 5)
Q <- genQ(J = 30, K = 4, Kj = Kj, Qid = "others", seed = 123)$gen.Q
idQ <- is.Qid(Q, model = "DINA")
**Description**

Introduces random misspecifications in a Q-matrix. Only binary Q-matrix are supported so far. Useful for simulation studies.

**Usage**

```r
missQ(Q, qjk, retainJ = 0, Qid = "none", seed = NULL)
```

**Arguments**

- `Q`: A $J$ items x $K$ attributes Q-matrix (matrix or data.frame).
- `qjk`: Number (or proportion, if lower than 1) of q-entries to modify in the Q-matrix.
- `retainJ`: Number of items to retain (i.e., not modify) in the Q-matrix. It will retain the first `retainJ` items. It is useful for assuring the completeness of the misspecified Q-matrix if the first items conform one or more identity matrices. The default is 0.
- `Qid`: Assure that the generated Q-matrix is generically identifiable. It includes "none" (for no identifiability assurance), "DINA", "DINO", or "others" (for other CDMs identifiability). The default is "none".
- `seed`: A seed for obtaining consistent results. If NULL, no seed is used. The default is NULL.

**Value**

`missQ` returns an object of class `missQ`.

- `Q`: The input (true) Q-matrix (matrix).
- `JK`: Number of items measuring each attribute (vector).
- `Kcor`: Tetrachoric correlations among the columns (matrix).
- `is.Qid`: Q-matrix identifiability information (list).
- `specifications`: Function call specifications (list).

**Author(s)**

Pablo Nájera, Universidad Pontificia Comillas

**References**

Examples

Kj <- c(15, 10, 0, 5) # 15 one-att, 10 2-atts, 0 3-atts, and 5 four-atts items
Q <- genQ(J = 30, K = 4, Kj = Kj, Qid = "others", seed = 123)
miss.Q <- missQ(Q = Q$gen.Q, qjk = .20, retainJ = 4, seed = 123)

Description

A procedure for determining the number of attributes underlying CDM using model fit comparison. For each number of attributes under exploration, a Q-matrix is estimated from the data using the discrete factor loading method (Wang, Song, & Ding, 2018), which can be further validated using the Hull method (Nájera, Sorrel, de la Torre, & Abad, 2020). Then, a CDM is fitted to the data using the resulting Q-matrix, and several fit indices are computed. After the desired range of number of attributes has been explored, the fit indices are compared. A suggested number of attributes is given for each fit index. The AIC index should be preferred among the other fit indices. For further details, see Nájera, Abad, & Sorrel (2021). This function can be also used by directly providing different Q-matrices (instead of estimating them from the data) in order to compare their fit and select the most appropriate Q-matrix. Note that, if Q-matrices are provided, this function will no longer serve as a dimensionality assessment method, but just as an automated model comparison procedure.

Usage

modelcompK(
  dat,
  exploreK = 1:7,
  Qs = NULL,
  stop = "none",
  val.Q = TRUE,
  estQ.args = list(criterion = "row", cor = "tet", rotation = "oblimin", fm = "uls"),
  valQ.args = list(index = "PVAF", iterative = "test.att", maxitr = 5, CDMconv = 0.01),
  verbose = TRUE
)

Arguments

dat A N individuals x J items (matrix or data.frame). Missing values need to be coded as NA.
exploreK Number of attributes to explore. The default is from 1 to 7 attributes.
Qs A list of Q-matrices to compare in terms of fit. If Qs is used, exploreK is ignored.
stop A fit index to use for stopping the procedure if a model leads to worse fit than a simpler one. This can be useful for saving time without exploring the whole exploreK when it is probable that the correct dimensionality has been already visited. It includes "AIC", "BIC", "CAIC", "SABIC", "M2", "SRMSR", "RMSEA2", 


or "sig.item.pairs". The latter represents the number of items that show bad fit with at least another item based on the transformed correlations (see itemfit function in the GDINA package; Ma & de la Torre, 2020). It can be also "none", which means that the whole exploreK will be examined. The default is "none".

`val.Q` Validate the estimated Q-matrices using the Hull method? Note that validating the Q-matrix is expected to increase its quality, but the computation time will increase. The default is TRUE.

`estQ.args` A list of arguments for the discrete factor loading empirical Q-matrix estimation method (see the estQ function):

- `criterion` Dichotomization criterion to transform the factor loading matrix into the Q-matrix. The possible options include "row" (for row means), "col" (for column means), "loadiff" (for the procedure based on loading differences), or a value between 0 and 1 (for a specific threshold). The default is "row".
- `cor` Type of correlations to use. It includes "cor" (for Pearson correlations) and "tet" (for tetrachoric/polychoric correlations), among others. See fa function from the psych R package for additional details. The default is "tet".
- `rotation` Rotation procedure to use. It includes "oblimin", "varimax", and "promax", among others. An oblique rotation procedure is usually recommended. See fa function from the psych R package for additional details. The default is "oblimin".
- `fm` Factoring method to use. It includes "uls" (for unweighted least squares), "ml" (for maximum likelihood), and "wls" (for weighted least squares), among others. See fa function from the psych R package for additional details. The default is "uls".

`valQ.args` A list of arguments for the Hull empirical Q-matrix validation method. Only applicable if `valQ = TRUE` (see the valQ function):

- `index` What index to use. It includes "PVAF" or "R2". The default is "PVAF".
- `iterative` (Iterative) implementation procedure. It includes "none" (for non-iterative), "test" (for test-level iterations), "test.att" (for test-level iterations modifying the least possible amount of q-entries in each iteration), and "item" (for item-level iterations). The default is "test.att".
- `maxitr` Maximum number of iterations if an iterative procedure has been selected. The default is 5.
- `CDMconv` Convergence criteria for the CDM estimations between iterations (only if an iterative procedure has been selected). The default is 0.01.
- `verbose` Show progress? The default is TRUE.

**Value**

`modelcompK` returns an object of class `modelcompK`.

- `sug.K` The suggested number of attributes for each fit index (vector). Only if `Qs = NULL`.
- `sel.Q` The suggested Q-matrix for each fit index (vector).
- `fit` The fit indices for each fitted model (matrix).
exp.exploreK Explored dimensionality (vector). It can be different from exploreK if stop has been used.

usedQ Q-matrices used to fit each model (list). They will be the estimated (and validated) Q-matrices if Qs = NULL. Otherwise, they will be Qs.

specifications Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas
Miguel A. Sorrel, Universidad Autónoma de Madrid
Francisco J. Abad, Universidad Autónoma de Madrid

References


Examples

```r
library(GDINA)
dat <- sim30GDINA$simdat
Q <- sim30GDINA$simQ

# Assess dimensionality from CDM data
mck <- modelcompK(dat = dat, exploreK = 4:7, stop = "AIC", val.Q = TRUE, verbose = TRUE)
mck$sug.K # Check suggested number of attributes by each fit index
mck$fit # Check fit indices for each K explored
sug.Q <- mck$usedQ[[paste0("K", mck$sug.K["AIC")]]] # Suggested Q-matrix by AIC
sug.Q <- orderQ(sug.Q, Q)$order.Q # Reorder Q-matrix attributes
mean(sug.Q == Q) # Check similarity with the generating Q-matrix

# Automatic fit comparison of competing Q-matrices
trueQ <- Q
missQ1 <- missQ(Q, .10, seed = 123)$miss.Q
missQ2 <- missQ(Q, .20, seed = 456)$miss.Q
missQ3 <- missQ(Q, .30, seed = 789)$miss.Q
```
Qs <- list(trueQ, missQ1, missQ2, missQ3)
mc <- modelcompK(dat = dat, Qs = Qs, verbose = TRUE)
mc$sel.Q # Best-fitting Q-matrix for each fit index
mc$fit # Check fit indices for each Q explored

---

**orderQ**  
*Reorder Q-matrix columns*

**Description**

Reorders Q-matrix columns according to a target matrix (e.g., another Q-matrix). Specifically, it provides a reordered Q-matrix which columns show the lowest possible average Tucker index congruent coefficient with the target columns. Reordering a Q-matrix is alike relabeling the attributes and it does not change the model. Useful for simulation studies (e.g., comparing a validated Q-matrix with the generating Q-matrix).

**Usage**

```
orderQ(Q, target)
```

**Arguments**

- **Q**: A J items x K attributes Q-matrix (matrix or data.frame). This is the Q-matrix that will be reordered.
- **target**: A J items x K attributes Q-matrix (matrix or data.frame). This could be the "true", generating Q-matrix.

**Value**

- **order.Q**: The reordered Q-matrix (matrix).
- **configs**: Comparison information between the different column configurations of the Q-matrix and the target Q-matrix, including the average absolute difference and the average Tucker index of factor congruence (matrix). The function will not look for all possible specifications if a perfect match is found.
- **specifications**: Function call specifications (list).

**Author(s)**

Francisco J. Abad, Universidad Autónoma de Madrid
Pablo Nájera, Universidad Pontificia Comillas
Examples

```r
library(GDINA)
dat <- sim30GDINA$simdat
Q <- sim30GDINA$simQ
sugQ1 <- estQ(r = dat, K = 5) # Estimate Q-matrix
sugQ1$est.Q <- orderQ(sugQ1$est.Q, Q)$order.Q # Reorder Q-matrix attributes
mean(sugQ1$est.Q == Q) # Check similarity with the generating Q-matrix
```

**Description**

Parallel analysis with column permutation (i.e., resampling) as used in Nájera, Abad, & Sorrel (2021). It is recommended to use principal components, Pearson correlations, and mean criterion (Garrido, Abad, & Ponsoda, 2013; Nájera, Abad, & Sorrel, 2021). The parallel analysis based on principal axis factor analysis is conducted using the `fa.parallel` function of the `psych` R package (Revelle, 2020). The tetrachoric correlations are efficiently estimated using the `sirt` R package (Robitzsch, 2020). The graph is made with the `ggplot2` package (Wickham et al., 2020).

**Usage**

```r
paK(
  dat,
  R = 100,
  fa = "pc",
  cor = "both",
  cutoff = "mean",
  fm = "uls",
  plot = TRUE,
  verbose = TRUE,
  seed = NULL
)
```

**Arguments**

- `dat` A N individuals x J items (matrix or data.frame). Missing values need to be coded as `NA`.
- `R` Number of resampled datasets (i.e., replications) to generate. The default is 100.
- `fa` Extraction method to use. It includes "pc" (for principal components analysis), "fa" (for principal axis factor analysis), and "both". The default is "pc".
- `cor` What type of correlations to use. It includes "cor" (for Pearson correlations), "tet" (for tetrachoric/polychoric correlations), and "both". The default is "both".
- `cutoff` What criterion to use as the cutoff. It can be "mean" (for the average generated eigenvalues) or a value between 0 and 100 (for a percentile). A vector with several criteria can be used. The default is "mean".
Factoring method to use. It includes "uls" (for unweighted least squares), "ml" (for maximum likelihood), and "wls" (for weighted least squares), among others. The default is "uls".

Print the parallel analysis plot? Note that the plot might be messy if many variants are requested. The default is TRUE.

progress. The default is TRUE.

A seed for obtaining consistent results. If NULL, no seed is used. The default is NULL.

Value

paK returns an object of class paK.

sug.K The suggested number of attributes for each variant (vector).

e.values The sample and reference eigenvalues (matrix).

plot The parallel analysis plot. Only if plot = TRUE (plot).

specifications Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas
Miguel A. Sorrel, Universidad Autónoma de Madrid
Francisco J. Abad, Universidad Autónoma de Madrid

References


Examples

```r
library(GDINA)
dat <- sim30GDINA$simdat
Q <- sim30GDINA$simQ

# In paK, R = 100 is recommended (R = 30 is here used for illustration purposes)
pa.K <- paK(dat = dat, R = 30, fa = "pc", cutoff = c("mean", 95), plot = TRUE, seed = 123)
pa.K$sug.K # Check suggested number of attributes by each parallel analysis variant
pa.K$e.values # Check eigenvalues
pa.K$plot # Show parallel analysis plot
```
**personFit**

*Calculate standardized log-likelihood statistic (lZ) for person fit evaluation*

**Description**

This function calculates the standardized log-likelihood statistic (lZ; Cui & Li, 2015; Drasgow et al. 1985) and the proposals for correcting its distribution discussed in Santos et al. (2019).

**Usage**

```r
personFit(fit, att.est = "MLE", sig.level = 0.05, p.adjust.method = "BH")
```

**Arguments**

- **fit**
  - An object of class `RDINA` or `GDINA` (Ma & de la Torre, 2020).
- **att.est**
  - What attribute estimates are used? The default is "MLE".
- **sig.level**
  - Scalar numeric. Alpha level for decision. Default is 0.05.
- **p.adjust.method**
  - Scalar character. Correction method for p-values. Possible values include "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", and "none". See `p.adjust` function from the stats R package for additional details. Default is BH.

**Value**

`personFit` returns an object of class `personFit`, with a list of elements:

- **stat** Person fit statistics (`data.frame`).
- **p** p-values (two-sided test) for the person fit statistics (`data.frame`).
- **sigp** Scalar vectors denoting the examinees for which the person fit statistic is significant (p-value) (`list`).
- **sigadjp** Scalar vectors denoting the examinees for which the person fit statistic is significant (adjusted p-value) (`list`).

**Author(s)**

Miguel A. Sorrel, Universidad Autónoma de Madrid,
Kevin Santos, University of the Philippines,
Pablo Nájera, Universidad Pontificia Comillas

**References**


Examples

```r
library(GDINA)
dat <- sim10GDINA$simdat[1:20, ]
Q <- sim10GDINA$simQ
fit <- GDINA(dat = dat, Q = Q, model = "GDINA")
res.personFit <- personFit(fit)
res.personFit
```

**RDINA**

*Restricted DINA model*

**Description**

Estimation of the restricted deterministic input, noisy "and" gate model (R-DINA; Nájera et al., 2023). In addition to the non-compensatory (i.e., conjunctive) condensation rule of the DINA model, the compensatory (i.e., disjunctive) rule of the DINO model can be also applied (i.e., R-DINO model). The R-DINA/R-DINO model should be only considered for applications involving very small sample sizes (N < 100; Nájera et al., 2023), and model fit evaluation and comparison with competing models (e.g., DINA/DINO, G-DINA) is highly recommended.

**Usage**

```r
RDINA(
  dat,
  Q,
  gate = "AND",
  att.prior = NULL,
  est = "Brent",
  EM.args = list(maxitr = 1000, conv.crit = 1e-04, init.phi = 0.2, verbose = TRUE),
  tau.alpha = "MAP",
  seed = NULL
)
```

**Arguments**

- `dat` A *N* individuals x *J* items (matrix or data.frame). Missing values need to be coded as NA. Caution is advised if missing data are present.

- `Q` A *J* items x *K* attributes Q-matrix (matrix or data.frame).

- `gate` Either a conjunctive ("AND") or disjunctive ("OR") condensation rule to estimate the RDINA or RDINO model, respectively. Default is "AND".

att.prior A $2^K$ attributes vector containing the prior distribution for each latent class. The sum of all elements does not have to be equal to 1, since the vector will be normalized. Default is NULL, which is a uniform prior distribution.

est Use the Brent’s method ("Brent") or the expectation-maximization algorithm ("EM") to estimate the model? Default is "Brent", since it is faster and both algorithms are virtually equivalent for the RDINA/RDINO model.

EM.args A list of arguments in case the EM algorithm is used to estimate the model:

- maxitr Maximum number of iterations. Default is 1000.
- conv.crit Convergence criterion regarding the maximum absolute change in either the phi parameter estimate or the marginal posterior probabilities of attribute mastery. Default is 0.0001.
- init.phi Initial value for the phi parameter. Default is 0.2.
- verbose Print information after each iteration. Default is TRUE.

tau.alpha Attribute profile estimator (either "MAP", "EAP", or "MLE") used to calculate the estimated classification accuracy as done with the CA function of the GDINA package (Ma & de la Torre, 2020).

seed Random number generation seed (e.g., to solve ties in case they occur with MLE or MAP estimation). Default is NULL, which means that no specific seed is used.

Value

RDINA returns an object of class RDINA.

- MLE Estimated attribute profiles with the MLE estimator (matrix).
- MAP Estimated attribute profiles with the MAP estimator (matrix).
- EAP Estimated attribute profiles with the EAP estimator (matrix).
- phi Phi parameter estimate (numeric).
- post.probs A (list) containing the estimates of the posterior probability of each examinee in each latent class (pp), marginal posterior probabilities of attribute mastery (mp), and posterior probability of each latent class (lp).
- likelihood A (list) containing the likelihood of each examinee in each latent class (lik_il) and the model log-likelihood (logLik).
- test.fit Relative model fit indices (list).
- class.accu A (list) containing the classification accuracy estimates at the test-level (tau), latent class-level (tau_l), and attribute-level (tau_k).
- specifications Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas
References


Examples

```r
library(GDINA)
Q <- sim30GDINA$simQ # Q-matrix
K <- ncol(Q)
J <- nrow(Q)
set.seed(123)
GS <- data.frame(guessing = rep(0.2, J), slip = rep(0.2, J))
sim <- simGDINA(20, Q, GS, model = "DINA")
simdat <- sim$dat # Simulated data
simatt <- sim$attribute # Generating attributes
fit.RDINA <- RDINA(simdat, Q) # Apply the GNPC method
ClassRate(fit.RDINA$EAP, simatt) # Check classification accuracy
```

---

**RDINA2GDINA**

*Translate RDINA object into GDINA object*

**Description**

This function translates an object of class RDINA to an object of class GDINA, so that the estimated R-DINA object is compatible with most of the functions in the GDINA package (Ma & de la Torre, 2020), including model fit, item fit, and Q-matrix validation.

**Usage**

```r
RDINA2GDINA(fit)
```

**Arguments**

- `fit` An object of class RDINA.

**Value**

RDINA2GDINA returns an object of class GDINA. See the GDINA package for more information.

**Author(s)**

Pablo Nájera, Universidad Pontificia Comillas
valQ

References

Examples

```r
library(GDINA)
dat <- sim3DINA$simdat
Q <- sim3DINA$simQ
fit1 <- RDINA(dat, Q)
fit2 <- RDINA2GDINA(fit1)
modelfit(fit2) # Model fit evaluation
itemfit(fit2) # Item fit evaluation
```

valQ

Empirical Q-matrix validation

Description

Empirical Q-matrix validation using the *Hull* method (Nájera, Sorrel, de la Torre, & Abad, 2020a). The procedure can be used either with the PVAF (de la Torre & Chiu, 2016) or McFadden’s pseudo R-squared (McFadden, 1974). The PVAF is recommended (Nájera, Sorrel, de la Torre, & Abad, 2020a). Note that the pseudo R-squared might not be computationally feasible for highly dimensional Q-matrices, say more than 10 attributes. Different iterative implementations are available, such as the test-level implementation (see Terzi & de la Torre, 2018), attribute-test-level implementation (Nájera, Sorrel, de la Torre, & Abad, 2020a), and item-level implementation (Nájera, Sorrel, de la Torre, & Abad, 2020b). If an iterative implementation is used, the GDINA R package (Ma & de la Torre, 2020) is used for the calibration of the CDMs.

Usage

```r
valQ(
  fit,
  index = "PVAF",
  iterative = "test.att",
  emptyatt = TRUE,
  maxitr = 100,
  CDMconv = 1e-04,
  verbose = TRUE
)
```

Arguments

- **fit** A G-DINA model fit object from the GDINA package (Ma & de la Torre, 2020).
- **index** What index to use. It includes "PVAF" or "R2". The default is "PVAF".
iterative

(Iterative) implementation procedure. It includes "none" (for non-iterative), "test" (for test-level iterations), "test.att" (for attribute-test-level), and "item" (for item-level iterations). The default is "test.att".

emptyatt

Is it possible for the suggested Q-matrix to have an empty attribute (i.e., an attribute not measured by any item)? Although rarely, it is possible for iterative procedures to provide a suggested Q-matrix in which one or more attributes are empty. This might indicate that the original Q-matrix had more attributes than necessary. If FALSE, then at least one item (i.e., the one that is most likely) will measure each attribute in the suggested Q-matrix. The default is TRUE.

maxitr

Maximum number of iterations if an iterative procedure has been selected. The default is 100.

CDMconv

Convergence criteria for the CDM estimations between iterations (only if an iterative procedure has been selected). The default is 0.0001.

verbose

Print information after each iteration if an iterative procedure is used. The default is TRUE.

Value

valQ returns an object of class valQ.

sug.Q Suggested Q-matrix (matrix).

Q Original Q-matrix (matrix).

sugQ.fit Several fit indices from the model obtained with the suggested Q-matrix (vector).

index PVAF or pseudo R-squared (depending on which one was used) for each item (matrix).

iter.Q Q-matrices used in each iteration (list). Provided only if an iterative procedure has been used.

iter.index PVAF or pseudo R-squared (depending on which one was used) for each item in each iteration (list). Provided only if an iterative procedure has been used.

n.iter Number of iterations used (double). Provided only if an iterative procedure has been used.

convergence Convergence information (double). It can be 1 (convergence), 2 (lack of convergence: maximum number of iterations achieved), 3 (lack of convergence: empty attribute obtained), and 4 (lack of convergence: loop Q-matrices). Provided only if an iterative procedure has been used.

time Initial and finish time (vector).

time.used Total computation time (difftime).

specifications Function call specifications (list).

Author(s)

Pablo Nájera, Universidad Pontificia Comillas
Miguel A. Sorrel, Universidad Autónoma de Madrid
Francisco J. Abad, Universidad Autónoma de Madrid
References


Examples

l library(GDINA)
dat <- sim30GDINA$simdat
Q <- sim30GDINA$simQ # Generating Q-matrix
miss.Q <- missQ(Q = Q, qjk = .30, retainJ = 5, seed = 123)$miss.Q # Misspecified Q-matrix
fit <- GDINA(dat, miss.Q) # GDINA object
sug.Q <- valQ(fit = fit, verbose = TRUE) # Hull method for Q-matrix validation
mean(sug.Q$sug.Q == Q) # Check similarity with the generating Q-matrix
Index

CA.MI, 2
estQ, 3
genQ, 6
GNPC, 7
is.Qid, 9
missQ, 10
modelcompK, 11
orderQ, 14
paK, 15
personFit, 17
RDINA, 18
RDINA2GDINA, 20
valQ, 21