Package ‘difR’

May 26, 2025

Type Package

Title Collection of Methods to Detect Dichotomous and Polytomous Differential Item Functioning (DIF)

Version 6.0.0

Date 2025-05-24

Maintainer Sebastien Beland <sebastien.beland@umontreal.ca>

Depends R (>= 3.0.0)

Imports mirt, ltm, lme4, deltaPlotR, DescTools, VGAM, glmnet

Description Methods to detect differential item functioning (DIF) in dichotomous and polytomous items, using both classical and modern approaches. These include Mantel-Haenszel procedures, logistic regression (including ordinal models), and regularization-based methods such as LASSO. Uniform and non-uniform DIF effects can be detected, and some methods support multiple focal groups. The package also provides tools for anchor purification, rest score matching, effect size estimation, and DIF simulation. See Magis, Beland, Tuerlinckx, and De Boeck (2010, Behavior Research Methods, 42, 847–862, <doi:10.3758/BRM.42.3.847>) for a general overview.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

URL https://github.com/343Babou/difR

BugReports https://github.com/343Babou/difR/issues

NeedsCompilation no

Author David Magis [aut] (IQVIA Belux),
Sebastien Beland [aut, cre] (Universite de Montreal),
Carl F. Falk [aut] (McGill University),
Gilles Raiche [aut] (UQAM)

Repository CRAN

Date/Publication 2025-05-26 07:30:02 UTC

Contents

`difR-package` *Collection of methods to detect dichotomous and polytomous differential item functioning (DIF) in psychometrics*

Description

The difR package contains several methods to detect DIF in dichotomous and polytomously scored items. Both uniform and non-uniform DIF effects can be detected, using approaches that either rely on item response theory models or not. Some methods can handle more than one focal group. Missing data, however, are not analyzed and should be removed or imputed beforehand.

Methods currently available are:

1. Transformed Item Difficulties (TID) method (Angoff and Ford, 1973)
2. Breslow-Day statistics (Breslow and Day, 1980)
3. Mantel-Haenszel for dichotomlous item (Holland and Thayer, 1988)
4. Mantel for polytomous item (Mantel, 1963)
5. Generalized Mantel-Haenszel (Penfield, 2001)
6. Standardization (Dorans and Kullick, 1986)
7. Breslow-Day (Aguerri et al., 2009; Penfield, 2003)
8. Logistic regression for dichotomlous item (Swaminathan and Rogers, 1990)
9. Logistic regression for polytomous item (Zumbo, 1999)
10. Generalized logistic regression (Magis, Raiche, Beland and Gerard, 2011)
11. Lasso regression (Magis, Tuerlinckx and De Boeck, 2015)
12. SIBTEST (Shealy and Stout) and Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996)
13. Lord's chi-square test (Lord, 1980)
14. Raju's area (Raju, 1990)
15. Likelihood-ratio test (Thissen, Steinberg and Wainer, 1988)
16. Common cumulative odds ratio (Liu and Agresti, 1996)
17. Indices based on pairwise comparisons of ordinal items (Woods, 1996)
18. Generalized Lord's chi-square test (Kim, Cohen and Park, 1995).

The difR package is further described in Magis, Beland, Tuerlinckx and De Boeck (2010).

Details

Package:	difR
Type:	Package
Version:	6.0.0
Date:	2025-05-12
Depends:	R (>= 3.0.0)
Imports:	mirt, ltm, lme4, deltaPlotR
License:	GPL

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Carl F. Falk
Department of Psychology
McGill University (Canada)
<carl.falk@mcgill.ca>, https://www.mcgill.ca/psychology/carl-f-falk
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity, 43*, 35-44. doi:10.1007/s1113500791302

Angoff, W. H., and Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 2*, 95-106. doi:10.1111/j.17453984.1973.tb00787.x

Chalmers, R. P. (2018). Improving the Crossing-SIBTEST statistic for detecting non-uniform DIF. *Psychometrika, 83*(2), 376–386. doi:10.1007/s1133601795838

Dorans, N. J. and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368. doi:10.1111/j.17453984.1986.tb00255.x

Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Dirs.), *Test validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Li, H.-H., and Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*, 647–677. doi:10.1007/BF02294041

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education, 14*, 235-259. doi:10.1207/S15324818AME1403_3

Penfield, R.D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research, 49*, 231-243.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. doi:10.1177/014662169001400208

Shealy, R. and Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194. doi:10.1007/BF02294572

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See Also

Other useful packages can be found in the *R Psychometric* task view.

`breslowDay` *Breslow-Day DIF statistic*

Description

Computes Breslow-Day statistics for DIF detection.

Usage

```
breslowDay(data, member, match = "score", anchor = 1:ncol(data),
     BDstat = "BD")
```

Arguments

- `data` — numeric: the data matrix (one row per subject, one column per item).
- `member` — numeric: the vector of group membership with zero and one entries only. See **Details**.
- `match` — specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `data`. See **Details**.
- `anchor` — a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.
- `BDstat` — character specifying the DIF statistic to be used. Possible values are `"BD"` (default) and `"trend"`. See **Details**.

Details

`breslowDay` computes one of the Breslow-Day statistics (1980) in the specific framework of differential item functioning. It forms the basic command of `difBD` and is specifically designed for this call.

The data are supplied by the `data` argument, with one row per subject and one column per item. Missing values are allowed but must be coded as `NA` values. They are discarded from sum-score computation.

The vector of group membership, specified by the `member` argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `breslowDay` function. This is specified by the `match` argument. By default, it takes the value `"score"` and the test score (i.e. raw score) is computed. The second option is to assign to `match` a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the `data` matrix.

Option `anchor` sets the items which are considered as anchor items for computing Breslow-Day DIF statistics. Items other than the anchor items and the tested item are discarded. `anchor` must hold integer values specifying the column numbers of the corresponding anchor items. It is primarily designed to perform item purification.

Two test statistics are available: the usual Breslow-Day statistic for testing homogeneous association (Aguerri, Galibert, Attorresi and Maranon, 2009) and the trend test statistic for assessing some monotonic trend in the odss ratios (Penfield, 2003). The DIF statistic is supplied by the `BDstat` argument, with values `"BD"` (default) for the usual statistic and `"trend"` for the trend test statistic.

Value

A list with three arguments:

- `res` — A matrix with one row per item and three columns: the first one contains the Breslow-Day statistic values, the second column indicates the degrees of freedom, and the last column displays the asymptotic p-values.
- `BDstat` — the value of the `BDstat` argument.
- `match` — a character string, either `"score"` or `"matching variable"` depending on the `match` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity, 43*, 35-44. doi:10.1007/s1113500791302

Breslow, N.E. and Day, N.E. (1980). *Statistical methods in cancer research, vol. I: The analysis of case-control studies*. Scientific Publication No 32. International Agency for Research on Cancer, Lyon, France.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R.D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research, 49*, 231-243.

See Also

difBD, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # With all items as anchor items
 breslowDay(verbal[,1:24], verbal[,26])

 # With all items as anchor items and trend
 # test statistic
 breslowDay(verbal[,1:24], verbal[,26], BDstat = "trend")

 # Removing item 3 from the set of anchor items
 breslowDay(verbal[,1:24], verbal[,26], anchor = c(1:5, 7:24))

## End(Not run)
```

`contrastMatrix` *Contrast matrix for computing generalized Lord's chi-squared DIF statistic*

Description

This command sets the appropriate contrast matrix C for computing the generalized Lord's chi-squared statistics in the framework of DIF detection among multiple groups.

Usage

```
contrastMatrix(nrFocal, model)
```

Arguments

`nrFocal`	numeric: the number of focal groups.
`model`	character: the logistic model to be fitted (either `"1PL"`, `"2PL"`, `"3PL"` or `"3PLc"`). See **Details**.

Details

The contrast matrix C is necessary to calculate the generalized Lord's chi-squared statistic. It is designed to perform accurate tests of equality of item parameters accross the groups of examinees (see Kim, Cohen and Park, 1995). This is a subroutine for the command `genLordChi2` which returns the DIF statistics.

The number of focal groups has to be specified by the argument `nrFocal`. Moreover, four logistic IRT models can be considered: the 1PL, 2PL and 3PL models can be set by using their acronyms (e.g. `"1PL"` for 1PL model, and so on). It is also possible to consider the constrained 3PL model, where all pseudo-guessing values are equal across the groups of examinees and take some predefined values which do not need to be supplied here. This model is specified by the value `"3PLc"` for argument `model`.

Value

A contrast matrix designed to test equality of item parameter estimates from the specified `model` and with `nrFocal` focal groups. The output matrix has a number of rows equal to `nrFocal` times the number of tested parameters (one for 1PL model, two for 2PL and constrained 3PL models, three for 3PL model). The number of columns is equal to (`nrFocal+1`) times the number of tested parameters. See Kim, Cohen and Park (1995) for further details.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland

Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

genLordChi2, difGenLord

Examples

```
## Not run:

 # Contrast matrices with 1PL model and several focal groups
 contrastMatrix(2, "1PL")
 contrastMatrix(3, "1PL")
 contrastMatrix(4, "1PL")

 # Contrast matrices with 2PL, constrained and unconstrained 3PL models and three
 # focal groups
 contrastMatrix(3, "2PL")
 contrastMatrix(3, "3PLc")
 contrastMatrix(3, "3PL")

## End(Not run)
```

---

dichoDif — *Comparison of DIF detection methods*

Description

This function compares the specified DIF detection methods with respect to the detected items and can only be used with dichotomous items.

Usage

```
dichoDif(Data, group, focal.name, method, anchor = NULL, props = NULL,
  thrTID = 1.5, alpha = 0.05, MHstat = "MHChisq", correct = TRUE,
  exact = FALSE, stdWeight = "focal", thrSTD = 0.1, BDstat = "BD",
  member.type = "group", match = "score", type = "both", criterion = "LRT",
  model = "2PL", c = NULL, engine = "ltm", discr = 1, irtParam = NULL,
  same.scale = TRUE, signed = FALSE, purify = FALSE, purType = "IPP1",
  nrIter = 10, extreme = "constraint", const.range = c(0.001, 0.999),
  nrAdd = 1, p.adjust.method = NULL, save.output = FALSE,
  output = c("out", "default"))
## S3 method for class 'dichoDif'
print(x, ...)
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within `Data`) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level of `group` which corresponds to the focal group.
`method`	character: the name of the selected method. Possible values are `"TID"`, `"MH"`, `"Std"`, `"Logistic"`, `"BD"`, `"SIBTEST"`, `"Lord"`, `"Raju"` and `"LRT"`. See **Details**.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`props`	either `NULL` (default) or a two-column matrix with proportions of success in the reference group and the focal group. See **Details**.
`thrTID`	numeric: the threshold for detecting DIF items with TID method (default is 1.5).
`alpha`	numeric: significance level (default is 0.05).
`MHstat`	character: specifies the DIF statistic to be used for DIF identification. Possible values are `"MHChisq"` (default) and `"logOR"`. See **Details**.
`correct`	logical: should the Mantel-Haenszel continuity correction be used? (default is TRUE).
`exact`	logical: should an exact test be computed? (default is `FALSE`).
`stdWeight`	character: the type of weights used for the standardized P-DIF statistic. Possible values are `"focal"` (default), `"reference"` and `"total"`. See **Details**.
`thrSTD`	numeric: the threshold (cut-score) for standardized P-DIF statistic (default is 0.10).
`BDstat`	character specifying the DIF statistic to be used. Possible values are `"BD"` (default) and `"trend"`. See **Details**.
`member.type`	character: either `"group"` (default) to specify that group membership is made of two groups, or `"cont"` to indicate that group membership is based on a continuous criterion. See **Details**.

match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "both" (default), "udif" and "nudif". See **Details**.
criterion	a character string specifying which DIF statistic is computed. Possible values are "LRT" (default) or "Wald". See **Details**.
model	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL"). Default is "2PL".
c	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
discr	either NULL or a real positive value for the common discrimination parameter (default is 1). Used onlky if model is "1PL" and engine is "ltm". See **Details**.
irtParam	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
same.scale	logical: are the item parameters of the irtParam matrix on the same scale? (default is "TRUE"). See **Details**.
signed	logical: should the Raju's statistics be computed using the signed (TRUE) or unsigned (FALSE, default) area? See **Details**.
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
purType	character: the type of purification process to be run. Possible values are "IPP1" (default), "IPP2" and "IPP3". Ignored if purify is FALSE or method does not supply the "TID" method.
nrIter	numeric: the maximal number of iterations in the item purification process (default is 10).
extreme	character: the method used to modify the extreme proportions. Possible values are "constraint" (default) or "add". Ignored if method is not "TID".
const.range	numeric: a vector of two constraining proportions. Default values are 0.001 and 0.999. Ignored if method is not "TID" or if extreme is "add".
nrAdd	integer: the number of successes and the number of failures to add to the data in order to adjust the proportions. Default value is 1. Ignored if method is not "TID" or if extreme is "constraint".
p.adjust.method	either NULL (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.
x	result from a dichoDif class object.
...	other generic parameters for the print function.

Details

dichoDif is a generic function which calls one or several DIF detection methods and summarize their output. The possible methods are:

1. "TID" for Transformed Item Difficulties (TID) method (Angoff and Ford, 1973),
2. "MH" for mantel-Haenszel (Holland and Thayer, 1988),
3. "Std" for standardization (Dorans and Kulick, 1986),
4. "BD" for Breslow-Day method (Penfield, 2003),
5. "Logistic" for logistic regression (Swaminathan and Rogers, 1990),
6. "SIBTEST" for SIBTEST (Shealy and Stout) and Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996) methods,
7. "Lord" for Lord's chi-square test (Lord, 1980),
8. "Raju" for Raju's area method (Raju, 1990), and
9. "LRT" for likelihood-ratio test method (Thissen, Steinberg and Wainer, 1988).

If method has a single component, the output of dichoDif is exactly the one provided by the method itself. Otherwise, the main output is a matrix with one row per item and one column per method. For each specified method and related arguments, items detected as DIF and non-DIF are respectively encoded as "DIF" and "NoDIF". When printing the output an additional column is added, counting the number of times each item was detected as functioning differently (Note: this is just an informative summary, since the methods are obviously not independent for the detection of DIF items).

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from either the computation of the sum-scores, the fitting of the logistic models or the IRT models (according to the method).

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument focal.name.

For "MH", "Std", "Logistic" and "BD" methods, the matching criterion can be either the test score or any other continuous or discrete variable to be passed in the Logistik function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

For Lord and Raju methods, one can specify either the IRT model to be fitted (by means of model, c, engine and discr arguments), or the item parameter estimates with arguments irtParam and same.scale. See difLord and difRaju for further details.

The threshold for detecting DIF items depends on the method. For standardization it has to be fully specified (with the thr argument), as well as for the TID method (through the thrTID argument). For the other methods it is depending on the significance level set by alpha.

For Mantel-Haenszel method, the DIF statistic can be either the Mantel-Haenszel chi-square statistic or the log odds-ratio statistic. The method is specified by the argument `MHstat`, and the default value is `"MHChisq"` for the chi-square statistic. Moreover, the option `correct` specifies whether the continuity correction has to be applied to Mantel-Haenszel statistic. See `difMH` for further details.

By default, the asymptotic Mantel-Haenszel statistic is computed. However, the exact statistics and related P-values can be obtained by specifying the logical argument `exact` to `TRUE`. See Agresti (1990, 1992) for further details about exact inference.

The weights for computing the standardized P-DIF statistics are defined through the argument `stdWeight`, with possible values `"focal"` (default value), `"reference"` and `"total"`. See `stdPDIF` for further details.

For Breslow-Day method, two test statistics are available: the usual Breslow-Day statistic for testing homogeneous association (Aguerri, Galibert, Attorresi and Maranon, 2009) and the trend test statistic for assessing some monotonic trend in the odss ratios (Penfield, 2003). The DIF statistic is supplied by the `BDstat` argument, with values `"BD"` (default) for the usual statistic and `"trend"` for the trend test statistic.

For logistic regression, the argument `type` permits to test either both uniform and nonuniform effects simultaneously (`type="both"`), only uniform DIF effect (`type="udif"`) or only nonuniform DIF effect (`type="nudif"`). The `criterion` argument specifies the DIF statistic to be computed, either the likelihood ratio test statistic (by setting `criterion="LRT"`) or the Wald test (by setting `criterion="Wald"`). Moreover, the group membership can be either a vector of two distinct values, one for the reference group and one for the focal group, or a continuous or discrete variable that acts as the "group" membership variable. In the former case, the `member.type` argument is set to `"group"` and the `focal.name` defines which value in the `group` variable stands for the focal group. In the latter case, `member.type` is set to `"cont"`, `focal.name` is ignored and each value of the `group` represents one "group" of data (that is, the DIF effects are investigated among participants relying on different values of some discrete or continuous trait). See `Logistik` for further details.

The SIBTEST method (Shealy and Stout, 1993) and its modified version, the Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996) are returned by the `difSIBTEST` function. SIBTEST method is returned when `type` argument is set to `"udif"`, while Crossing-SIBTEST is set with `"nudif"` value for the `type` argument. Note that `type` takes the by-default value `"both"` which is not allowed within the `difSIBTEST` function; however, within this fucntion, keeping the by-default value yields selection of Crossing-SIBTEST.

The `difSIBTEST` function is a wrapper to the `SIBTEST` function from the **mirt** package (Chalmers, 2012) to fit within the `difR` framework (Magis et al., 2010). Therefore, if you are using this function for publication purposes please cite Chalmers (2018; 2012) and Magis et al. (2010).

For Raju's method, the type of area (signed or unsigned) is fixed by the logical `signed` argument, with default value `FALSE` (i.e. unsigned areas). See `RajuZ` for further details.

Item purification can be requested by specifying `purify` option to `TRUE`. Recall that item purification process is slightly different for IRT and for non-IRT based methods. See the corresponding methods for further information.

Adjustment for multiple comparisons is possible with the argument `p.adjust.method`. See the corresponding methods for further information.

A pre-specified set of anchor items can be provided through the `anchor` argument. For non-IRT methods, anchor items are used to compute the test score (as matching criterion). For IRT methods, anchor items are used to rescale the item parameters on a common metric. See the corresponding methods for further information. Note that `anchor` argument is not working with `"LRT"` method.

The output of the dichoDif function can be stored in a text file by fixing save.output and output appropriately. See the help file of selectDif function (or any other DIF method) for further information.

Value

Either the output of one of the DIF detection methods, or a list of class "dichoDif" with the following arguments:

DIF	a character matrix with one row per item and whose columns refer to the different specified detection methods. See **Details**.
props	the value of the props argument.
thrTID	the value of the thrTID argument.
correct	the value of correct argument.
exact	the value of exact argument.
alpha	the significance level alpha.
MHstat	the value of the MHstat argument.
stdWeight	the value of the stdWeight argument.
thrSTD	the value of thrSTD argument.
BDstat	the value of the BDstat argument.
member.type	the value of the member.type argument.
match	the value of the match argument.
type	the value of the type argument.
criterion	the value of the criterion argument.
model	the value of model argument.
c	the value of c argument.
engine	The value of the engine argument.
discr	the value of the discr argument.
irtParam	the value of irtParam argument.
same.scale	the value of same.scale argument.
p.adjust.method	the value of the p.adjust.method argument.
purification	the value of purify argument.
nrPur	an integer vector (of length equal to the number of methods) with the number of iterations in the purification process. Returned only if purify is TRUE.
convergence	a logical vector (of length equal to the number of methods) indicating whether the iterative purification process converged. Returned only if purify is TRUE.
anchor.names	the value of the anchor argument.
save.output	the value of the save.output argument.
output	the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7*, 131-177. doi:10.1214/ss/1177011454

Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity, 43*, 35-44. doi:10.1007/s1113500791302

Angoff, W. H., and Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 2*, 95-106. doi:10.1111/j.17453984.1973.tb00787.x

Chalmers, R. P. (2012). mirt: A Multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48(6)*, 1-29. doi:10.18637/jss.v048.i06

Chalmers, R. P. (2018). Improving the Crossing-SIBTEST statistic for detecting non-uniform DIF. *Psychometrika, 83*(2), 376–386. doi:10.1007/s1133601795838

Dorans, N. J. and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368. doi:10.1111/j.17453984.1986.tb00255.x

Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Dirs.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Li, H.-H., and Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*, 647–677. doi:10.1007/BF02294041

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R.D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research, 49*, 231-243.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. doi:10.1177/014662169001400208

Shealy, R. and Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194. doi:10.1007/BF02294572

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See Also

`difTID`, `difMH`, `difStd`, `difBD`, `difLogistic`, `difSIBTEST`, `difLord`, `difRaju`, `difLRT`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal)!="Anger"]

 # Comparing TID, Mantel-Haenszel, standardization; logistic regression and SIBTEST
 # TID threshold 1.0
 # Standardization threshold 0.08
 # no continuity correction,
 # with item purification
 # both types of DIF effect for logistic regression
 # CSIBTEST method
 dichoDif(verbal, group = 25, focal.name = 1, method = c("TID", "MH", "Std",
        "Logistic", "SIBTEST"), correct = FALSE, thrSTD = 0.08, thrTID = 1, purify = TRUE)

 # Same analysis, but using items 1 to 5 as anchor and saving the output into
 # the 'dicho' file
 dichoDif(verbal, group = 25, focal.name = 1, method = c("TID", "MH", "Std",
          "Logistic"), correct = FALSE, thrSTD = 0.08, thrTID = 1, purify = TRUE,
          anchor = 1:5,save.output = TRUE, output = c("dicho", "default"))

 # Comparing Lord and Raju results with 2PL model and
 # with item purification
 dichoDif(verbal, group = 25, focal.name = 1, method = c("Lord", "Raju"),
          model = "2PL", purify = TRUE)

## End(Not run)
```

difBD *Breslow-Day DIF method*

Description

Performs DIF detection using Breslow-Day method.

Usage

```
difBD(Data, group, focal.name, anchor = NULL, match = "score", BDstat = "BD",
   alpha = 0.05, purify = FALSE, nrIter = 10, p.adjust.method = NULL,
   save.output = FALSE, output = c("out", "default"))
## S3 method for class 'BD'
print(x, ...)
## S3 method for class 'BD'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within Data) of group membership. See **Details**.
focal.name	numeric or character indicating the level of group which corresponds to the focal group.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
BDstat	character specifying the DIF statistic to be used. Possible values are "BD" (default) and "trend". See **Details**.
alpha	numeric: significance level (default is 0.05).
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
nrIter	numeric: the maximal number of iterations in the item purification process (default is 10).
p.adjust.method	either NULL (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
save.output	logical: should the output be saved into a text file? (Default is FALSE).

output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.
x	the result from a BD class object.
pch, col	type of usual pch and col graphical options.
number	logical: should the item number identification be printed (default is TRUE).
save.plot	logical: should the plot be saved into a separate file? (default is FALSE).
save.options	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or "default" (default value), and the third component is the file extension, either "pdf" (default) or "jpeg". See **Details**.
...	other generic parameters for the plot or the print functions.

Details

The method of Breslow-Day (1980) allows for detecting non-uniform differential item functioning without requiring an item response model approach.

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from sum-score computation.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument focal.name.

Two test statistics are available: the usual Breslow-Day statistic for testing homogeneous association (Aguerri, Galibert, Attorresi and Maranon, 2009) and the trend test statistic for assessing some monotonic trend in the odds ratios (Penfield, 2003). The DIF statistic is supplied by the BDstat argument, with values "BD" (default) for the usual statistic and "trend" for the trend test statistic.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the breslowDay function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha, and the degrees of freedom depend on the DIF statistic. With the usual Breslow-Day statistic (BDstat=="BD"), it is the number of partial tables taken into account (Aguerri *et al.*, 2009). With the trend test statistic, the degrees of freedom are always equal to one (Penfield, 2003).

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item was detected as functioning differently at the first step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument `p.adjust.method`. The latter must be an acronym of one of the available adjustment methods of the `p.adjust` function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by `"Holm"` and `"BH"`) perform best for DIF purposes. See `p.adjust` function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the `anchor` argument. It must be a vector of either item names (which must match exactly the column names of `Data` argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if `purify` is set to `TRUE`). By default it is `NULL` so that no anchor item is specified.

The output of the `difBD`, as displayed by the `print.BD` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The `plot.BD` function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the colour are fixed by the usual `pch` and `col` arguments. Option `number` permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "BD" with the following arguments:

- `BD`: a matrix with one row per item and three columns: the first one contains the Breslow-Day statistic value, the second column indicates the degrees of freedom, and the last column displays the asymptotic *p*-values.
- `p.value`: the vector of p-values for the BD statistics.
- `alpha`: the significance level for DIF detection.
- `DIFitems`: either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
- `BDstat`: the value of the `BDstat` argument.
- `match`: a character string, either `"score"` or `"matching variable"` depending on the `match` argument.
- `p.adjust.method`: the value of the `p.adjust.method` argument.

`adjusted.p` either `NULL` or the vector of adjusted p-values for multiple comparisons.

`purification` the value of `purify` option.

`nrPur` the number of iterations in the item purification process. Returned only if `purify` is `TRUE`.

`difPur` a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if `purify` is `TRUE`.

`convergence` logical indicating whether the iterative item purification process stopped before the maximal number `nrIter` of allowed iterations. Returned only if `purify` is `TRUE`.

`names` the names of the items.

`anchor.names` the value of the `anchor` argument.

`save.output` the value of the `save.output` argument.

`output` the value of the `output` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity, 43*, 35-44. doi:10.1007/s1113500791302

Breslow, N.E. and Day, N.E. (1980). *Statistical methods in cancer research, vol. I: The analysis of case-control studies*. Scientific Publication No 32. International Agency for Research on Cancer, Lyon.

Clauser, B.E. and Mazor, K.M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R.D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research, 49*, 231-243.

See Also

breslowDay, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 verbal<-verbal[colnames(verbal) != "Anger"]

 # Three equivalent settings of the data matrix and the group membership
 difBD(verbal, group = 25, focal.name = 1)
 difBD(verbal, group = "Gender", focal.name = 1)
 difBD(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # With the BD trend test statistic
 difBD(verbal, group = 25, focal.name = 1, BDstat = "trend")

 # Multiple comparisons adjustment using Benjamini-Hochberg method
 difBD(verbal, group = 25, focal.name = 1, p.adjust.method = "BH")

 # With item purification
 difBD(verbal, group = "Gender", focal.name = 1, purify = TRUE)
 difBD(verbal, group = "Gender", focal.name = 1, purify = TRUE, nrIter = 5)

 # With items 1 to 5 set as anchor items
 difBD(verbal, group = "Gender", focal.name = 1, anchor = 1:5)
 difBD(verbal, group = "Gender", focal.name = 1, anchor = 1:5, purify = TRUE)

 # Saving the output into the "BDresults.txt" file (and default path)
 r <- difBD(verbal, group = 25, focal.name = 1, save.output = TRUE,
            output = c("BDresults","default"))

 # Graphical devices
 plot(r)

 # Plotting results and saving it in a PDF figure
 plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))
```

```
## End(Not run)
```

---

difGenLogistic *Generalized logistic regression DIF method*

Description

Performs DIF detection among multiple groups using generalized logistic regression method.

Usage

```
difGenLogistic(Data, group, focal.names, anchor = NULL, match = "score",
  type = "both", criterion = "LRT", alpha = 0.05, purify = FALSE, nrIter = 10,
  p.adjust.method = NULL, save.output = FALSE, output = c("out", "default"))
## S3 method for class 'genLogistic'
print(x, ...)
## S3 method for class 'genLogistic'
plot(x, plot = "lrStat", item = 1, itemFit = "best",pch = 8, number = TRUE,
   col = "red", colIC = rep("black", length(x$focal.names)+1),
   ltyIC = 1:(length(x$focal.names)+1), title = NULL, save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ref.name = NULL, ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
focal.names	numeric or character vector indicating the levels of group which correspond to the focal groups.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. Ignored if match is not "score". See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "both" (default), "udif" and "nudif". See **Details**.
criterion	character: the type of test statistic used to detect DIF items. Possible values are "LRT" (default) and "Wald". See **Details**.
alpha	numeric: significance level (default is 0.05).
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).

`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
`x`	the result from a `Logistik` class object.
`plot`	character: the type of plot, either `"lrStat"` or `"itemCurve"`. See **Details**.
`item`	numeric or character: either the number or the name of the item for which logistic curves are plotted. Use only when `plot="itemCurve"`.
`itemFit`	character: the model to be selected for drawing the item curves. Possible values are `"best"` (default) for drawing from the best of the two models, and `"null"` for using fitted parameters of the null model M_0. Not used if `"plot"` is `"lrStat"`. See **Details**.
`pch, col`	type of usual `pch` and `col` graphical options.
`number`	logical: should the item number identification be printed (default is `TRUE`).
`colIC, ltyIC`	vectors of elements of the usual `col` and `lty` arguments for logistic curves. Used only when `plot="itemCurve"`.
`title`	either a character string with the title of the plot, or `NULL` (default), for which a specific title is automatically displayed.
`save.plot`	logical: should the plot be saved into a separate file? (default is `FALSE`).
`save.options`	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
`ref.name`	either `NULL`(default) or a character string for the name of the reference group (to be used instead of "Reference" in the legend). Ignored if `plot` is `"lrStat"`.
`...`	other generic parameters for the `plot` or the `print` functions.

Details

The generalized logistic regression method (Magis, Raiche, Beland and Gerard, 2011) allows for detecting both uniform and non-uniform differential item functioning among multiple groups without requiring an item response model approach. It consists in fitting a logistic model with the matching criterion, the group membership and an interaction between both as covariates. The statistical significance of the parameters related to group membership and the group-score interaction is then evaluated by means of the usual likelihood-ratio test. The argument `type` permits to test either both uniform and nonuniform effects simultaneously (`type="both"`), only uniform DIF effect (`type="udif"`) or only nonuniform DIF effect (`type="nudif"`). The identification of DIF items can be performed with either the Wald test or the likelihood ratio test, by setting the `criterion` argument to `"Wald"` or `"LRT"` respectively. See `genLogistik` for further details.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the genLogistik function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from the fitting of the logistic models (see glm for further details).

The vector of group membership must hold at least three values, either as numeric or character. The focal groups are defined by the values of the argument focal.names. If there is a unique focal group, then difGenLogistic returns the output of difLogistic.

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha and with *J* (if type="udif" or type="nudif") or *2J* (if type="both") degrees of freedom (*J* is the number of focal groups).

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item is detected as functioning differently at the first step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. By default it is NULL so that no anchor item is specified. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). Moreover, if the match argument is not set to "score", anchor items will not be taken into account even if anchor is not NULL.

The measures of effect size are provided by the difference ΔR^2 between the R^2 coefficients of the two nested models (Nagelkerke, 1991; Gomez-Benito, Dolores Hidalgo and Padilla, 2009). The effect sizes are classified as "negligible", "moderate" or "large". Two scales are available, one from Zumbo and Thomas (1997) and one from Jodoin and Gierl (2001). The output displays the ΔR^2 measures, together with the two classifications.

The output of the difGenLogistic, as displayed by the print.genLogistic function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of output. The default value is "default", meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

Two types of plots are available. The first one is obtained by setting plot="lrStat" and it is the default option. The likelihood ratio statistics are displayed on the Y axis, for each item. The detection threshold is displayed by a horizontal line, and items flagged as DIF are printed with the color defined by argument col. By default, items are spotted with their number identification (number=TRUE); otherwise they are simply drawn as dots whose form is given by the option pch.

The other type of plot is obtained by setting plot="itemCurve". In this case, the fitted logistic curves are displayed for one specific item set by the argument item. The latter argument can hold either the name of the item or its number identification. If the argument itemFit takes the value "best", the curves are drawn according to the output of the best model among M_0 and M_1. That is, two curves are drawn if the item is flagged as DIF, and only one if the item is flagged as non-DIF. If itemFit takes the value "null", then the two curves are drawn from the fitted parameters of the null model M_0. See genLogistik for further details on the models. The colors and types of traits for these curves are defined by means of the arguments colIC and ltyIC respectively. These are set as vectors of length $J + 1$, the first element for the reference group and the others for the focal groups. Finally, the ref.name argument permits to display the name if the reference group (instead of "Reference") in the legend.

Both types of plots can be stored in a figure file, either in PDF or JPEG format. Fixing save.plot to TRUE allows this process. The figure is defined through the components of save.options. The first two components perform similarly as those of the output argument. The third component is the figure format, with allowed values "pdf" (default) for PDF file and "jpeg" for JPEG file.

Value

A list of class "genLogistic" with the following arguments:

genLogistik	the values of the generalized logistic regression statistics.
p.value	the vector of p-values for the generalized logistic regression statistics.
logitPar	a matrix with one row per item and $2 + J * 2$ columns, holding the fitted parameters of the best model (among the two tested models) for each item.
parM0	the matrix of fitted parameters of the null model M_0, as returned by the Logistik command.
covMat	a 3-dimensional matrix of size *p* x *p* x *K*, where *p* is the number of estimated parameters and *K* is the number of items, holding the *p* x *p* covariance matrices of the estimated parameters (one matrix for each tested item).
deltaR2	the differences in Nagelkerke's R^2 coefficients. See **Details**.
alpha	the value of alpha argument.
thr	the threshold (cut-score) for DIF detection.
DIFitems	either the column indicators for the items which were detected as DIF items, or "No DIF item detected".
type	the value of type argument.

p.adjust.method
: the value of the p.adjust.method argument.

adjusted.p
: either NULL or the vector of adjusted p-values for multiple comparisons.

purification
: the value of purify option.

nrPur
: the number of iterations in the item purification process. Returned only if purify is TRUE.

difPur
: a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if purify is TRUE.

convergence
: logical indicating whether the iterative item purification process stopped before the maximal number of nrItem allowed iterations. Returned only if purify is TRUE.

names
: the names of the items.

anchor.names
: the value of the anchor argument.

focal.names
: the value of focal.names argument.

criterion
: the value of the criterion argument.

save.output
: the value of the save.output argument.

output
: the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Clauser, B.E. and Mazor, K.M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Gomez-Benito, J., Dolores Hidalgo, M. and Padilla, J.-L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology, 5*, 18-25. doi:10.1027/1614-2241.5.1.18

Hidalgo, M. D. and Lopez-Pina, J.A. (2004). Differential item functioning detection and effect size: a comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64*, 903-915. doi:10.1177/0013164403261769

Jodoin, M. G. and Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349. doi:10.1207/S15324818AME1404_2

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692. doi:10.1093/biomet/78.3.691

Zumbo, B. D. and Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.

See Also

`genLogistik`, `genDichoDif`, `subtestLogistic`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and
 # trait anger score ("Low" or "High")
 group <- rep("WomanLow", nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)

 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Testing both types of DIF effects
 # Three equivalent settings of the data matrix and the group membership
 r <- difGenLogistic(Verbal, group = 25, focal.names = names)
 difGenLogistic(Verbal, group = "group", focal.name = names)
 difGenLogistic(Verbal[,1:24], group = Verbal[,25], focal.names = names)
```

```
 # Using the Wald test
 difGenLogistic(Verbal, group = 25, focal.names = names, criterion = "Wald")

 # Multiple comparisons adjustment using Benjamini-Hochberg method
difGenLogistic(Verbal, group = 25, focal.names = names, p.adjust.method = "BH")

 # With item purification
 difGenLogistic(Verbal, group = 25, focal.names = names, purify = TRUE)
 difGenLogistic(Verbal, group = 25, focal.names = names, purify = TRUE,
   nrIter = 5)

 # With items 1 to 5 set as anchor items
 difGenLogistic(Verbal, group = 25, focal.name = names, anchor = 1:5)

 # Testing for nonuniform DIF effect
 difGenLogistic(Verbal, group = 25, focal.names = names, type = "nudif")

 # Testing for uniform DIF effect
 difGenLogistic(Verbal, group = 25, focal.names = names, type = "udif")

 # User anger trait score as matching criterion
 anger <- verbal[,25]
 difGenLogistic(Verbal, group = 25, focal.names = names, match = anger)

 # Saving the output into the "GLresults.txt" file (and default path)
 r <- difGenLogistic(Verbal, group = 25, focal.name = names,
                save.output = TRUE, output = c("GLresults","default"))

 # Graphical devices
 plot(r)
 plot(r, plot = "itemCurve", item = 1)
 plot(r, plot = "itemCurve", item = 1, itemFit = "best")
 plot(r, plot = "itemCurve", item = 6)
 plot(r, plot = "itemCurve", item = 6, itemFit = "best")

 # Plotting results and saving it in a PDF figure
 plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

---

difGenLord *Generalized Lord's chi-squared DIF method*

Description

Performs DIF detection among multiple groups using generalized Lord's chi-squared method.

Usage

```
difGenLord(Data, group, focal.names, model, c = NULL, engine = "ltm",
  discr = 1, irtParam = NULL, nrFocal = 2, same.scale = TRUE, anchor = NULL,
  alpha = 0.05, purify = FALSE, nrIter = 10, p.adjust.method = NULL,
  save.output = FALSE,  output = c("out", "default"))
## S3 method for class 'GenLord'
print(x, ...)
## S3 method for class 'GenLord'
plot(x, plot = "lordStat", item = 1, pch = 8,
   number = TRUE, col = "red", colIC = rep("black",
   length(x$focal.names)+1), ltyIC = 1:(length(x$focal.names)
   + 1), save.plot = FALSE, save.options = c("plot", "default", "pdf"),
      ref.name = NULL, ...)
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within `Data`) of group membership. See **Details**.
`focal.names`	numeric or character vector indicating the levels of `group` which correspond to the focal groups.
`model`	character: the IRT model to be fitted (either `"1PL"`, `"2PL"` or `"3PL"`).
`c`	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
`engine`	character: the engine for estimating the 1PL model, either `"ltm"` (default) or `"lme4"`.
`discr`	either `NULL` or a real positive value for the common discrimination parameter (default is 1). Used onlky if `model` is `"1PL"` and `engine` is `"ltm"`. See **Details**.
`irtParam`	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
`nrFocal`	numeric: the number of focal groups (default is 2).
`same.scale`	logical: are the item parameters of the `irtParam` matrix on the same scale? (default is `TRUE`). See **Details**.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`alpha`	numeric: significance level (default is 0.05).
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.

save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.
x	the result from a GenLord class object.
plot	character: the type of plot, either "lordStat" or "itemCurve". See **Details**.
item	numeric or character: either the number or the name of the item for which ICC curves are plotted. Used only when plot="itemCurve".
pch, col	type of usual pch and col graphical options.
number	logical: should the item number identification be printed (default is TRUE).
colIC, ltyIC	vectors of elements of the usual col and lty arguments for ICC curves. Used only when plot="itemCurve".
save.plot	logical: should the plot be saved into a separate file? (default is FALSE).
save.options	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or "default" (default value), and the third component is the file extension, either "pdf" (default) or "jpeg". See **Details**.
ref.name	either NULL(default) or a character string for the name of the reference group (to be used instead of "Reference" in the legend). Ignored if plot is "lordStat".
...	other generic parameters for the plot or the print functions.

Details

The generalized Lord's chi-squared method (Kim, Cohen and Park, 1995), also referred to as *Qj* statistic, allows for detecting uniform or non-uniform differential item functioning among multiple groups by setting an appropriate item response model. The input can be of two kinds: either by displaying the full data, the group membership, the focal groups and the model, or by giving the item parameter estimates (with the option irtParam). Both can be supplied, but in this case only the parameters in irtParam are used for computing generalized Lord's chi-squared statistic.

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded for item parameter estimation.

The vector of group membership must hold at least three different values, either as numeric or character. The focal groups are defined by the values of the argument focal.names.

If the model is not the 1PL model, or if engine is equal to "ltm", the selected IRT model is fitted using marginal maximum likelihood by means of the functions from the ltm package (Rizopoulos, 2006). Otherwise, the 1PL model is fitted as a generalized linear mixed model, by means of the glmer function of the lme4 package (Bates and Maechler, 2009).

With the "1PL" model and the "ltm" engine, the common discrimination parameter is set equal to 1 by default. It is possible to fix another value through the argumentdiscr. Alternatively, this common discrimination parameter can be estimated (though not returned) by fixing discr to NULL.

The 3PL model can be fitted either unconstrained (by setting c to NULL) or by fixing the pseudo-guessing values. In the latter case, the argument c is either a numeric vector of same length of the number of items, with one value per item pseudo-guessing parameter, or a single value which is duplicated for all the items. If c is different from NULL then the 3PL model is always fitted (whatever the value of model).

The irtParam matrix has a number of rows equal to the number of groups (reference and focal ones) times the number of items J. The first J rows refer to the item parameter estimates in the reference group, while the next sets of J rows correspond to the same items in each of the focal groups. The number of columns depends on the selected IRT model: 2 for the 1PL model, 5 for the 2PL model, 6 for the constrained 3PL model and 9 for the unconstrained 3PL model. The columns of irtParam have to follow the same structure as the output of itemParEst command (the latter can actually be used to create the irtParam matrix). The number of focal groups has to be specified with argument nrFocal (default value is 2).

In addition to the matrix of parameter estimates, one has to specify whether items in the focal groups were rescaled to those of the reference group. If not, rescaling is performed by equal means anchoring (Cook and Eignor, 1991). Argument same.scale is used for this choice (default option is TRUE and assumes therefore that the parameters are already placed on a same scale).

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha and p degrees of freedom. The value of p is the product of the number of focal groups by the number of item parameters to be tested (1 for the 1PL model, 2 for the 2PL model or the constrained 3PL model, and 3 for the unconstrained 3PL model).

Item purification can be performed by setting purify to TRUE. In this case, the purification occurs in the equal means anchoring process: items detected as DIF are iteratively removed from the set of items used for equal means anchoring, and the procedure is repeated until either the same items are identified twice as functioning differently, or when nrIter iterations have been performed. In the latter case a warning message is printed. See Candell and Drasgow (1988) for further details.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to rescale the item parameters on a common metric. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). By default it is NULL so that no anchor item is specified. If item parameters are provided thorugh the irtParam argument and if they are on the same scale (i.e. if same.scale is TRUE), then anchor items are not used (even if they are specified).

The output of the difGenLord, as displayed by the print.GenLord function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

Two types of plots are available. The first one is obtained by setting `plot="lordStat"` and it is the default option. The chi-squared statistics are displayed on the Y axis, for each item. The detection threshold is displayed by a horizontal line, and items flagged as DIF are printed with the color defined by argument `col`. By default, items are spotted with their number identification (`number=TRUE`); otherwise they are simply drawn as dots whose form is given by the option `pch`.

The other type of plot is obtained by setting `plot="itemCurve"`. In this case, the fitted ICC curves are displayed for one specific item set by the argument `item`. The latter argument can hold either the name of the item or its number identification. The item parameters are extracted from the `itemParFinal` matrix if the output argument `purification` is `TRUE`, otherwise from the `itemParInit` matrix and after a rescaling of the item parameters using the `itemRescale` command. A legend is displayed in the upper left corner of the plot. The colors and types of traits for these curves are defined by means of the arguments `colIC` and `ltyIC` respectively. These are set as vectors of length 2, the first element for the reference group and the second for the focal group. Finally, the `ref.name` argument permits to display the name if the reference group (instead of "Reference") in the legend.

Both types of plots can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "GenLord" with the following arguments:

`genLordChi`	the values of the generalized Lord's chi-squared statistics.
`p.value`	the vector of p-values for the generalized Lord's chi-square statistics.
`alpha`	the value of `alpha` argument.
`thr`	the threshold (cut-score) for DIF detection.
`df`	the degrees of freedom of the asymptotic null distribution of the statistics.
`DIFitems`	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
`p.adjust.method`	the value of the `p.adjust.method` argument.
`adjusted.p`	either `NULL` or the vector of adjusted p-values for multiple comparisons.
`purification`	the value of `purify` option.
`nrPur`	the number of iterations in the item purification process. Returned only if `purify` is `TRUE`.
`difPur`	a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the i-th row refer to items which were classified respectively as non-DIF and DIF items at the $(i-1)$-th step. The first row corresponds to the initial classification of the items. Returned only if `purify` is `TRUE`.

convergence	logical indicating whether the iterative item purification process stopped before the maximal number nrIterof allowed iterations. Returned only if purify is TRUE.
model	the value of model argument.
c	The value of the c argument.
engine	The value of the engine argument.
discr	the value of the discr argument.
itemParInit	the matrix of initial parameter estimates, with the same format as irtParam either provided by the user (through irtParam) or estimated from the data (and displayed after rescaling).
itemParFinal	the matrix of final parameter estimates, with the same format as irtParam, obtained after item purification. Returned only if purify is TRUE.
estPar	a logical value indicating whether the item parameters were estimated (TRUE) or provided by the user (FALSE).
names	the names of the items.
anchor.names	the value of the anchor argument.
focal.names	the value of the focal.names argument.
save.output	the value of the save.output argument.
output	the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Candell, G.L. and Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253-260.

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

itemParEst

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and trait
 # anger score ("Low" or "High")
 group <- rep("WomanLow",nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)

 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Three equivalent settings of the data matrix and the group membership
 # 1PL model, "ltm" engine
 r <- difGenLord(Verbal, group = 25, focal.names = names, model = "1PL")
 difGenLord(Verbal, group = "group", focal.name = names, model = "1PL")
 difGenLord(Verbal[,1:24], group = Verbal[,25], focal.names = names, model = "1PL")

 # 1PL model, "ltm" engine, estimated common discrimination
 r <- difGenLord(Verbal, group = 25, focal.names = names, model = "1PL", discr = NULL)

 # 1PL model, "lme4" engine
 difGenLord(Verbal, group = "group", focal.name = names, model = "1PL", engine = "lme4")

 # With items 1 to 5 set as anchor items
 difGenLord(Verbal, group = 25, focal.names = names, model = "1PL", anchor = 1:5)

 # Multiple comparisons adjustment using Benjamini-Hochberg method
difGenLord(Verbal, group = 25, focal.names = names, model = "1PL", p.adjust.method = "BH")
```

```
# With item purification
difGenLord(Verbal, group = 25, focal.names = names, model = "1PL", purify = TRUE)

# Saving the output into the "GLresults.txt" file (and default path)
r <- difGenLord(Verbal, group = 25, focal.names = names, model = "1PL",
        save.output = TRUE, output = c("GLresults", "default"))

# Splitting the data into the four subsets according to "group"
data0<-data1<-data2<-data3<-NULL
for (i in 1:nrow(verbal)){
 if (group[i]=="WomanLow") data0<-rbind(data0,as.numeric(verbal[i,1:24]))
 if (group[i]=="WomanHigh") data1<-rbind(data1,as.numeric(verbal[i,1:24]))
 if (group[i]=="ManLow") data2<-rbind(data2,as.numeric(verbal[i,1:24]))
 if (group[i]=="ManHigh") data3<-rbind(data3,as.numeric(verbal[i,1:24]))
 }

# Estimation of the item parameters (1PL model)
m0.1PL<-itemParEst(data0, model = "1PL")
m1.1PL<-itemParEst(data1, model = "1PL")
m2.1PL<-itemParEst(data2, model = "1PL")
m3.1PL<-itemParEst(data3, model = "1PL")

# Merging the item parameters WITHOUT rescaling
irt.noscale<-rbind(m0.1PL,m1.1PL,m2.1PL,m3.1PL)
rownames(irt.noscale)<-rep(colnames(verbal[,1:24]),4)

# Merging the item parameters WITH rescaling
irt.scale<-rbind(m0.1PL, itemRescale(m0.1PL,m1.1PL),
itemRescale(m0.1PL,m2.1PL) ,itemRescale(m0.1PL,m3.1PL))
rownames(irt.scale)<-rep(colnames(verbal[,1:24]),4)

# Equivalent calculations
difGenLord(irtParam = irt.noscale, nrFocal = 3, same.scale = FALSE)
difGenLord(irtParam = irt.scale, nrFocal = 3, same.scale = TRUE)

# With item purification
difGenLord(irtParam = irt.noscale, nrFocal = 3, same.scale = FALSE, purify = TRUE)

# Graphical devices
plot(r)
plot(r, plot = "itemCurve", item = 1)
plot(r, plot = "itemCurve", item = 6)
plot(r, plot = "itemCurve", item = 6, ref.name = "WomanHigh")

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

# Changing the path, JPEG figure
path <- "c:/Program Files/"
plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

difGMH *Generalized Mantel-Haenszel DIF method*

Description

Performs DIF detection among multiple groups using the generalized Mantel-Haenszel method.

Usage

```
difGMH(Data, group, focal.names, anchor = NULL, match = "score", alpha = 0.05,
   purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
   output = c("out", "default"))
## S3 method for class 'GMH'
print(x, ...)
## S3 method for class 'GMH'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within Data) of group membership. See **Details**.
focal.names	numeric or character vector indicating the levels of group which correspond to the focal groups.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
alpha	numeric: significance level (default is 0.05).
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
nrIter	numeric: the maximal number of iterations in the item purification process (default is 10).
p.adjust.method	either NULL (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.

x	the result from a GMH class object.
pch, col	type of usual pch and col graphical options.
number	logical: should the item number identification be printed (default is TRUE).
save.plot	logical: should the plot be saved into a separate file? (default is FALSE).
save.options	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or "default" (default value), and the third component is the file extension, either "pdf" (default) or "jpeg". See **Details**.
...	other generic parameters for the plot or the print functions.

Details

The generalized Mantel-Haenszel statistic (Somes, 1986) can be used to detect uniform differential item functioning among multiple groups, without requiring an item response model approach (Penfield, 2001).

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from sum-score computation.

The vector of group membership must hold at least three value, either as numeric or character. The focal groups are defined by the values of the argument focal.names. If there is a unique focal group, then difGMH returns the output of difMH (**without** continuity correction).

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha and with as many degrees of freedom as the number of focal groups.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the genMantelHaenszel function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item detected as functioning differently at the first step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). By default it is NULL so that no anchor item is specified.

The output of the difGMH, as displayed by the print.GMH function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of output. The default value is "default", meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The plot.GMH function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the colour are fixed by the usual pch and col arguments. Option number permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing save.plot to TRUE allows this process. The figure is defined through the components of save.options. The first two components perform similarly as those of the output argument. The third component is the figure format, with allowed values "pdf" (default) for PDF file and "jpeg" for JPEG file.

Value

A list of class "GMH" with the following arguments:

GMH the values of the generalized Mantel-Haenszel statistics.

p.value the vector of p-values for the generalized Mantel-Haenszel statistics.

alpha the value of alpha argument.

thr the threshold (cut-score) for DIF detection.

DIFitems either the items which were detected as DIF items, or "No DIF item detected".

match a character string, either "score" or "matching variable" depending on the match argument.

p.adjust.method
the value of the p.adjust.method argument.

adjusted.p either NULL or the vector of adjusted p-values for multiple comparisons.

purification the value of purify option.

nrPur the number of iterations in the item purification process. Returned only if purify is TRUE.

difPur a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if purify is TRUE.

`convergence`	logical indicating whether the iterative item purification process stopped before the maximal number `nrIter` of allowed iterations. Returned only if `purify` is `TRUE`.
`names`	the names of the items.
`anchor.names`	the value of the `anchor` argument.
`focal.names`	the value of `focal.names` argument.
`save.output`	the value of the `save.output` argument.
`output`	the value of the `output` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Clauser, B. E. and Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education, 14*, 235-259. doi:10.1207/S15324818AME1403_3

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician, 40*, 106-108. doi:10.2307/2684866

See Also

`difGMH`, `difMH`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and
 # trait anger score ("Low" or "High")
 group <- rep("WomanLow",nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)

 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Three equivalent settings of the data matrix and the group membership
 difGMH(Verbal, group = 25, focal.names = names)
 difGMH(Verbal, group = "group", focal.name = names)
 difGMH(Verbal[,1:24], group = Verbal[,25], focal.names = names)

 # Multiple comparisons adjustment using Benjamini-Hochberg method
 difGMH(Verbal, group = 25, focal.names = names, p.adjust.method = "BH")

 # With item purification
 difGMH(Verbal, group = 25, focal.names = names, purify = TRUE)
 difGMH(Verbal, group = 25, focal.names = names, purify = TRUE, nrIter = 5)

 # With items 1 to 5 set as anchor items
 difMH(Verbal, group = 25, focal.name = names, anchor = 1:5)
 difMH(Verbal, group = 25, focal.name = names, anchor = 1:5, purify = TRUE)


 # Saving the output into the "GMHresults.txt" file (and default path)
 r <- difGMH(Verbal, group = 25, focal.name = names, save.output = TRUE,
           output = c("GMHresults","default"))

 # Graphical devices
 plot(r)

 # Plotting results and saving it in a PDF figure
 plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

difLogistic *Logistic regression DIF method*

Description

Performs DIF detection using logistic regression method.

Usage

```
difLogistic(Data, group, focal.name, anchor = NULL, member.type = "group",
  match = "score", type = "both", criterion = "LRT", alpha = 0.05,
  all.cov = FALSE, purify = FALSE, nrIter = 10, p.adjust.method = NULL,
  save.output = FALSE, output = c("out", "default"))
## S3 method for class 'Logistic'
print(x, ...)
## S3 method for class 'Logistic'
plot(x, plot="lrStat", item = 1, itemFit = "best", pch = 8, number = TRUE,
  col = "red", colIC = rep("black", 2), ltyIC = c(1, 2), save.plot = FALSE,
  save.options = c("plot", "default", "pdf"), group.names = NULL, ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
focal.name	numeric or character indicating the level of group which corresponds to the focal group. Ignored if member.type is not "group".
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. Ignored if match is not "score". See **Details**.
member.type	character: either "group" (default) to specify that group membership is made of two groups, or "cont" to indicate that group membership is based on a continuous criterion. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "both" (default), "udif" and "nudif". See **Details**.
criterion	a character string specifying which DIF statistic is computed. Possible values are "LRT" (default) or "Wald". See **Details**.
alpha	numeric: significance level (default is 0.05).
all.cov	logical: should *all* covariance matrices of model parameter estimates be returned (as lists) for both nested models and all items? (default is FALSE.

purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE). Ignored if `match` is not `"score"`.
nrIter	numeric: the maximal number of iterations in the item purification process. (default is 10).
p.adjust.method	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
save.output	logical: should the output be saved into a text file? (Default is `FALSE`).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
x	the result from a `Logistik` class object.
plot	character: the type of plot, either `"lrStat"` (default) or `"itemCurve"`. See **Details**.
item	numeric or character: either the number or the name of the item for which logistic curves are plotted. Used only when `plot="itemCurve"`.
itemFit	character: the model to be selected for drawing the item curves. Possible values are `"best"` (default) for drawing from the best of the two models, and `"null"` for using fitted parameters of the null model M_0. Not used if `"plot"` is `"lrStat"`. See **Details**.
pch, col	type of usual `pch` and `col` graphical options.
number	logical: should the item number identification be printed (default is `TRUE`).
colIC, ltyIC	vectors of two elements of the usual `col` and `lty` arguments for logistic curves. Used only when `plot="itemCurve"`.
save.plot	logical: should the plot be saved into a separate file? (default is `FALSE`).
save.options	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
group.names	either `NULL` (default) or a vector of two character strings giving the names of the reference group and the focal group (in this order) for display in the legend. Ignored if `plot` is `"lrStat"`.
...	other generic parameters for the `plot` or the `print` functions.

Details

The logistic regression method (Swaminathan and Rogers, 1990) allows for detecting both uniform and non-uniform differential item functioning without requiring an item response model approach. It consists in fitting a logistic model with the matching criterion, the group membership and an interaction between both as covariates. The statistical significance of the parameters related to group membership and the group-score interaction is then evaluated by means of either the likelihood-ratio test or the Wald test. The argument `type` permits to test either both uniform and nonuniform effects simultaneously (`type="both"`), only uniform DIF effect (`type="udif"`) or only nonuniform DIF effect (`type="nudif"`). The argument `criterion` permits to select either the likelihood ratio test (`criterion=="LRT"`) or the Wald test (`criterion=="Wald"`). See `Logistik` for further details.

The group membership can be either a vector of two distinct values, one for the reference group and one for the focal group, or a continuous or discrete variable that acts as the "group" membership variable. In the former case, the member.type argument is set to "group" and the focal.name defines which value in the group variable stands for the focal group. In the latter case, member.type is set to "cont", focal.name is ignored and each value of the group represents one "group" of data (that is, the DIF effects are investigated among participants relying on different values of some discrete or continuous trait). See Logistik for further details.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the Logistik function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from the fitting of the logistic models (see glm for further details).

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha and with one (if type="udif" or type="nudif") or two (if type="both") degrees of freedom.

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item is detected as functioning differently at the first step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed. Note that purification is possible only if the test score is considered as the matching criterion. Thus, purify is ignored when match is not "score".

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. By default it is NULL so that no anchor item is specified. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). Moreover, if the match argument is not set to "score", anchor items will not be taken into account even if anchor is not NULL.

The measures of effect size are provided by the difference ΔR^2 between the R^2 coefficients of the two nested models (Nagelkerke, 1991; Gomez-Benito, Dolores Hidalgo and Padilla, 2009). The effect sizes are classified as "negligible", "moderate" or "large". Two scales are available, one from Zumbo and Thomas (1997) and one from Jodoin and Gierl (2001). The output displays the ΔR^2 measures, together with the two classifications.

The output of the `difLogistic`, as displayed by the `print.Logistic` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

Two types of plots are available. The first one is obtained by setting `plot="lrStat"` and it is the default option. The likelihood ratio statistics are displayed on the Y axis, for each item. The detection threshold is displayed by a horizontal line, and items flagged as DIF are printed with the color defined by argument `col`. By default, items are spotted with their number identification (`number=TRUE`); otherwise they are simply drawn as dots whose form is given by the option `pch`.

The other type of plot is obtained by setting `plot="itemCurve"`. In this case, the fitted logistic curves are displayed for one specific item set by the argument `item`. The latter argument can hold either the name of the item or its number identification. If the argument `itemFit` takes the value `"best"`, the curves are drawn according to the output of the best model among M_0 and M_1. That is, two curves are drawn if the item is flagged as DIF, and only one if the item is flagged as non-DIF. If `itemFit` takes the value `"null"`, then the two curves are drawn from the fitted parameters of the null model M_0. See `Logistik` for further details on the models. The colors and types of traits for these curves are defined by means of the arguments `colIC` and `ltyIC` respectively. These are set as vectors of length 2, the first element for the reference group and the second for the focal group. Finally, the argument `group.names` permits to display the names of the reference and focal groups (instead of "Reference" and "Focal") in the legend.

Both types of plots can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "Logistic" with the following arguments:

- `Logistik` — the values of the logistic regression statistics.
- `p.value` — the vector of p-values for the logistic regression statistics.
- `logitPar` — a matrix with one row per item and four columns, holding the fitted parameters of the best model (among the two tested models) for each item.
- `logitSe` — a matrix with one row per item and four columns, holding the standard errors of the fitted parameters of the best model (among the two tested models) for each item.
- `parM0` — the matrix of fitted parameters of the null model M_0, as returned by the `Logistik` command.
- `seM0` — the matrix of standard error of fitted parameters of the null model M_0, as returned by the `Logistik` command.

cov.M0 — either NULL (if all.cov argument is FALSE) or a list of covariance matrices of parameter estimates of the "full" model (M_0) for each item (if all.cov argument is TRUE).

cov.M1 — either NULL (if all.cov argument is FALSE) or a list of covariance matrices of parameter estimates of the "reduced" model (M_1) for each item (if all.cov argument is TRUE).

deltaR2 — the differences in Nagelkerke's R^2 coefficients. See **Details**.

alpha — the value of alpha argument.

thr — the threshold (cut-score) for DIF detection.

DIFitems — either the column indicators for the items which were detected as DIF items, or "No DIF item detected".

member.type — the value of the member.type argument.

match — a character string, either "score" or "matching variable" depending on the match argument.

type — the value of type argument.

p.adjust.method — the value of the p.adjust.method argument.

adjusted.p — either NULL or the vector of adjusted p-values for multiple comparisons.

purification — the value of purify option.

nrPur — the number of iterations in the item purification process. Returned only if purify is TRUE.

difPur — a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if purify is TRUE.

convergence — logical indicating whether the iterative item purification process stopped before the maximal number of nrItem allowed iterations. Returned only if purify is TRUE.

names — the names of the items.

anchor.names — the value of the anchor argument.

criterion — the value of the criterion argument.

save.output — the value of the save.output argument.

output — the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)

<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Clauser, B.E. and Mazor, K.M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Finch, W.H. and French, B. (2007). Detection of crossing differential item functioning: a comparison of four methods. *Educational and Psychological Measurement, 67*, 565-582. doi:10.1177/0013164406296975

Gomez-Benito, J., Dolores Hidalgo, M. and Padilla, J.-L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology, 5*, 18-25. doi:10.1027/1614-2241.5.1.18

Hidalgo, M. D. and Lopez-Pina, J.A. (2004). Differential item functioning detection and effect size: a comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64*, 903-915. doi:10.1177/0013164403261769

Jodoin, M. G. and Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349. doi:10.1207/S15324818AME1404_2

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692. doi:10.1093/biomet/78.3.691

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

Zumbo, B.D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): logistic regression modelling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, B. D. and Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.

See Also

Logistik, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 anger <- verbal[,colnames(verbal)=="Anger"]
 verbal <- verbal[,colnames(verbal)!="Anger"]

 # Testing both DIF effects simultaneously
 # Three equivalent settings of the data matrix and the group membership
 r <- difLogistic(verbal, group=25, focal.name = 1)
 difLogistic(verbal, group = "Gender", focal.name = 1)
 difLogistic(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # Returning all covariance matrices of model parameters
 difLogistic(verbal, group=25, focal.name = 1, all.cov = TRUE)

 # Testing both DIF effects with the Wald test
 r2 <- difLogistic(verbal, group = 25, focal.name = 1, criterion = "Wald")

 # Testing nonuniform DIF effect
 difLogistic(verbal, group = 25, focal.name = 1, type = "nudif")

 # Testing uniform DIF effect
 difLogistic(verbal, group = 25, focal.name = 1, type = "udif")

 # Multiple comparisons adjustment using Benjamini-Hochberg method
 difLogistic(verbal, group=25, focal.name = 1, p.adjust.method = "BH")

 # With item purification
 difLogistic(verbal, group = "Gender", focal.name = 1, purify = TRUE)
 difLogistic(verbal, group = "Gender", focal.name = 1, purify = TRUE, nrIter = 5)

 # With items 1 to 5 set as anchor items
 difLogistic(verbal, group = 25, focal.name = 1, anchor = 1:5)

 # Using anger trait score as the matching criterion
 difLogistic(verbal,group = 25, focal.name = 1,match = anger)

 # Using trait anger score as the group variable (i.e. testing
 # for DIF with respect to trait anger score)
 difLogistic(verbal[,1:24],group = anger,member.type = "cont")

 # Saving the output into the "Lresults.txt" file (and default path)
 r <- difLogistic(verbal, group = 25, focal.name = 1, save.output = TRUE,
           output = c("Lresults", "default"))

 # Graphical devices
 plot(r)
 plot(r2)
```

```
 plot(r, plot = "itemCurve", item = 1)
 plot(r, plot = "itemCurve", item = 1, itemFit = "null")
 plot(r, plot = "itemCurve", item = 6)
 plot(r, plot = "itemCurve", item = 6, itemFit = "null")

 # Plotting results and saving it in a PDF figure
 plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

---

difLogReg *General logistic regression DIF method*

Description

Performs DIF detection using logistic regression method with either two groups, more than two groups, or a continuous group variable.

Usage

```
difLogReg(Data, group, focal.name, anchor = NULL, group.type = "group",
  match = "score", type = "both", criterion = "LRT", alpha = 0.05,
  purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
  output = c("out", "default"))
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within `data`) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level(s) of `group` which corresponds to the focal group(s). Ignored if `group.type` is not `"group"`.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`group.type`	character: either `"group"` (default) to specify that group membership is made of two (or more than two) groups, or `"cont"` to indicate that group membership is based on a continuous criterion. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `Data`. See **Details**.

`type`	a character string specifying which DIF effects must be tested. Possible values are `"both"` (default), `"udif"` and `"nudif"`. See **Details**.
`criterion`	a character string specifying which DIF statistic is computed. Possible values are `"LRT"` (default) or `"Wald"`. See **Details**.
`alpha`	numeric: significance level (default is 0.05).
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE). Ignored if `match` is not `"score"`.
`nrIter`	numeric: the maximal number of iterations in the item purification process. (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.

Details

The `difLogReg` function is a meta-function for logistic regression DIF analysis. It encompasses all possible cases that are currently implemented in difR and makes appropriate calls to the function `difLogistic` or `difGenLogistic`.

Three situations are embedded in this function.

1. The group membership is defined by two distinct groups. In this case, `group.type` must be `"group"` and `focal.name` must be a single value, referring to the name or label of the focal group.
2. The group membership is defined by a finite, yet larger than two, number of groups. In this case, `group.type` must be `"group"` and `focal.name` must be a vector with the names or labels of all focal groups.
3. The group membership is a continuous or discrete (but treated as continuous) variable. In this case, DIF is tested with respect to this "membership" variable. Furthermore, `group.type` must be `"cont"` and `focal.name` is ignored (though some value must be specified, for instance `NULL`).

The specification of the data, the options for item purification, DIF statistic selection, and output saving, are identical to the options arising from the `difLogistic` and `difGenLogistic` functions.

Value

A list of class "Logistic" (if `group.type` is `"cont"` or with the length of `focal.name` is one) or "genLogistic", with related arguments (see `difLogistic` and `difGenLogistic`).

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

See Also

difLogistic, difGenLogistic, dichoDif, genDichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Few examples
 difLogReg(Data=verbal[,1:24], group=verbal[,26], focal.name=1)
 difLogReg(Data = verbal[,1:24], group = verbal[,26], focal.name = 1, match = verbal[,25])
 difLogReg(Data = verbal[,1:24], group = verbal[,25], focal.name = 1, group.type = "cont")

 group<-rep("WomanLow",nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 difLogReg(Data = verbal[,1:24], group = group, focal.name = names)

## End(Not run)
```

difLord *Lord's chi-squared DIF method*

Description

Performs DIF detection using Lord's chi-squared method.

Usage

```
difLord(Data, group, focal.name, model, c = NULL, engine = "ltm", discr = 1,
  irtParam = NULL, same.scale = TRUE, anchor = NULL, alpha = 0.05,
  purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
  output = c("out", "default"))
## S3 method for class 'Lord'
print(x, ...)
## S3 method for class 'Lord'
plot(x, plot = "lordStat", item = 1, pch = 8, number = TRUE, col = "red",
   colIC = rep("black", 2), ltyIC = c(1, 2), save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), group.names = NULL, ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
focal.name	numeric or character indicating the level of group which corresponds to the focal group.
model	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL").
c	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
discr	either NULL or a real positive value for the common discrimination parameter (default is 1). Used onlky if model is "1PL" and engine is "ltm". See **Details**.
irtParam	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
same.scale	logical: are the item parameters of the irtParam matrix on the same scale? (default is "TRUE"). See **Details**.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
alpha	numeric: significance level (default is 0.05).

`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
`x`	the result from a `Lord` class object.
`plot`	character: the type of plot, either `"lordStat"` or `"itemCurve"`. See **Details**.
`item`	numeric or character: either the number or the name of the item for which ICC curves are plotted. Used only when `plot="itemCurve"`.
`pch, col`	type of usual `pch` and `col` graphical options.
`number`	logical: should the item number identification be printed (default is `TRUE`).
`colIC, ltyIC`	vectors of two elements of the usual `col` and `lty` arguments for ICC curves. Used only when `plot="itemCurve"`.
`save.plot`	logical: should the plot be saved into a separate file? (default is `FALSE`).
`save.options`	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
`group.names`	either `NULL` (default) or a vector of two character strings giving the names of the reference group and the focal group (in this order) for display in the legend. Ignored if `plot` is `"lordStat"`.
`...`	other generic parameters for the `plot` or the `print` functions.

Details

Lord's chi-squared method (Lord, 1980) allows for detecting uniform or non-uniform differential item functioning by setting an appropriate item response model. The input can be of two kinds: either by displaying the full data, the group membership and the model, or by giving the item parameter estimates (through the option `irtParam`). Both can be supplied, but in this case only the parameters in `irtParam` are used for computing Lord's chi-squared statistic.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded for item parameter estimation.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument `focal.name`.

If the model is not the 1PL model, or if `engine` is equal to `"ltm"`, the selected IRT model is fitted using marginal maximum likelihood by means of the functions from the `ltm` package (Rizopoulos, 2006). Otherwise, the 1PL model is fitted as a generalized linear mixed model, by means of the `glmer` function of the `lme4` package (Bates and Maechler, 2009).

With the `"1PL"` model and the `"ltm"` engine, the common discrimination parameter is set equal to 1 by default. It is possible to fix another value through the argument`discr`. Alternatively, this common discrimination parameter can be estimated (though not returned) by fixing `discr` to `NULL`.

The 3PL model can be fitted either unconstrained (by setting `c` to `NULL`) or by fixing the pseudo-guessing values. In the latter case, the argument `c` holds either a numeric vector of same length of the number of items, with one value per item pseudo-guessing parameter, or a single value which is duplicated for all the items. If `c` is different from `NULL` then the 3PL model is always fitted (whatever the value of `model`).

The `irtParam` matrix has a number of rows equal to twice the number of items in the data set. The first J rows refer to the item parameter estimates in the reference group, while the last J ones correspond to the same items in the focal group. The number of columns depends on the selected IRT model: 2 for the 1PL model, 5 for the 2PL model, 6 for the constrained 3PL model and 9 for the unconstrained 3PL model. The columns of `irtParam` have to follow the same structure as the output of `itemParEst` command (the latter can actually be used to create the `irtParam` matrix).

In addition to the matrix of parameter estimates, one has to specify whether items in the focal group were rescaled to those of the reference group. If not, rescaling is performed by equal means anchoring (Cook and Eignor, 1991). Argument `same.scale` is used for this choice (default option is `TRUE` and assumes therefore that the parameters are already placed on the same scale).

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus `alpha` and p degrees of freedom ($p=1$ for the 1PL model, $p=2$ for the 2PL model or the 3PL model with constrained pseudo-guessing parameters, and $p=3$ for the unconstrained 3PL model).

Item purification can be performed by setting `purify` to `TRUE`. In this case, the purification occurs in the equal means anchoring process. Items detected as DIF are iteratively removed from the set of items used for equal means anchoring, and the procedure is repeated until either the same items are identified twice as functioning differently, or when `nrIter` iterations have been performed. In the latter case a warning message is printed. See Candell and Drasgow (1988) for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

Adjustment for multiple comparisons is possible with the argument `p.adjust.method`. The latter must be an acronym of one of the available adjustment methods of the `p.adjust` function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by `"Holm"` and `"BH"`) perform best for DIF purposes. See `p.adjust` function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the `anchor` argument. It must be a vector of either item names (which must match exactly the column names of `Data` argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to rescale the item parameters on a common metric. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if `purify` is set to `TRUE`). By default it is `NULL` so that no anchor item is specified.

If item parameters are provided thorugh the `irtParam` argument and if they are on the same scale (i.e. if `same.scale` is `TRUE`), then anchor items are not used (even if they are specified).

Under the 1PL model, the displayed output also proposes an effect size measure, which is -2.35 times the difference between item difficulties of the reference group and the focal group (Penfield and Camilli, 2007, p. 138). This effect size is similar Mantel-Haenszel's Δ_{MH} effect size, and the ETS delta scale is used to classify the effect sizes (Holland and Thayer, 1985).

The output of the `difLord`, as displayed by the `print.Lord` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

Two types of plots are available. The first one is obtained by setting `plot="lordStat"` and it is the default option. The chi-squared statistics are displayed on the Y axis, for each item. The detection threshold is displayed by a horizontal line, and items flagged as DIF are printed with the color defined by argument `col`. By default, items are spotted with their number identification (`number=TRUE`); otherwise they are simply drawn as dots whose form is given by the option `pch`.

The other type of plot is obtained by setting `plot="itemCurve"`. In this case, the fitted ICC curves are displayed for one specific item set by the argument `item`. The latter argument can hold either the name of the item or its number identification. The item parameters are extracted from the `itemParFinal` matrix if the output argument `purification` is `TRUE`, otherwise from the `itemParInit` matrix and after a rescaling of the item parameters using the `itemRescale` command. A legend is displayed in the upper left corner of the plot. The colors and types of traits for these curves are defined by means of the arguments `colIC` and `ltyIC` respectively. These are set as vectors of length 2, the first element for the reference group and the second for the focal group. Finally, the argument `group.names` permits to display the names of the reference and focal groups (instead of "Reference" and "Focal") in the legend.

Both types of plots can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "Lord" with the following arguments:

`LordChi`	the values of the Lord's chi-square statistics.
`p.value`	the vector of p-values for the Lord's chi-square statistics.
`alpha`	the value of `alpha` argument.
`thr`	the threshold (cut-score) for DIF detection.
`DIFitems`	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
`purification`	the value of `purify` option.
`nrPur`	the number of iterations in the item purification process. Returned only if `purify` is `TRUE`.

`difPur` a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if `purify` is `TRUE`.

`convergence` logical indicating whether the iterative item purification process stopped before the maximal number `nrIter`of allowed iterations. Returned only if `purify` is `TRUE`.

`model` the value of `model` argument.

`c` The value of the `c` argument.

`engine` The value of the `engine` argument.

`discr` the value of the `discr` argument.

`p.adjust.method` the value of the `p.adjust.method` argument.

`adjusted.p` either `NULL` or the vector of adjusted p-values for multiple comparisons.

`itemParInit` the matrix of initial parameter estimates,with the same format as `irtParam` either provided by the user (through `irtParam`) or estimated from the data (and displayed without rescaling).

`itemParFinal` the matrix of final parameter estimates, with the same format as `irtParam`, obtained after item purification. Returned only if `purify` is `TRUE`.

`estPar` a logical value indicating whether the item parameters were estimated (`TRUE`) or provided by the user (`FALSE`).

`names` the names of the items.

`anchor.names` the value of the `anchor` argument.

`save.output` the value of the `save.output` argument.

`output` the value of the `output` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Candell, G.L. and Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253–260. doi:10.1177/014662168801200304

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37–45.

Holland, P. W. and Thayer, D. T. (1985). An alternative definition of the ETS delta scale of item difficulty. *Research Report RR-85-43*. Princeton, New-Jersey: Educational Testing Service.

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847–862. doi:10.3758/BRM.42.3.847

Penfield, R. D., and Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao and S. Sinharray (Eds.), *Handbook of Statistics 26: Psychometrics* (pp. 125-167). Amsterdam, The Netherlands: Elsevier.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1-25. doi:10.18637/jss.v017.i05

See Also

`itemParEst`, `dichoDif`, `p.adjust`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal)!="Anger"]

 # Three equivalent settings of the data matrix and the group membership
 # (1PL model, "ltm" engine)
 r <- difLord(verbal, group = 25, focal.name = 1, model = "1PL")
 difLord(verbal, group = "Gender", focal.name = 1, model = "1PL")
 difLord(verbal[,1:24], group = verbal[,25], focal.name = 1, model = "1PL")

 # With items 1 to 5 set as anchor items
 difLord(verbal, group = 25, focal.name = 1, model = "1PL", anchor = 1:5)

 # Multiple comparisons adjustment of p-values with Benjamini-Hochberg method
difLord(verbal, group = 25, focal.name = 1, model = "1PL", anchor = 1:5, p.adjust.method = "BH")
```

```
# 1PL model, "lme4" engine
difLord(verbal, group = 25, focal.name = 1, model = "1PL", engine = "lme4")

# 2PL model
difLord(verbal, group = "Gender", focal.name = 1, model = "2PL")

# 3PL model with all pseudo-guessing parameters constrained to 0.05
difLord(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05)

# Same models, with item purification
difLord(verbal, group = 25, focal.name = 1, model = "1PL", purify = TRUE)
difLord(verbal, group = "Gender", focal.name = 1, model = "2PL", purify = TRUE)
difLord(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05,
purify = TRUE)

# Saving the output into the "LordResults.txt" file (and default path)
r <- difLord(verbal, group = 25, focal.name = 1, model = "1PL",
     save.output = TRUE, output = c("LordResults","default"))

# Splitting the data into reference and focal groups
nF<-sum(Gender)
nR<-nrow(verbal)-nF
data.ref<-verbal[,1:24][order(Gender),][1:nR,]
data.focal<-verbal[,1:24][order(Gender),][(nR+1):(nR+nF),]

## Pre-estimation of the item parameters (1PL model, "ltm" engine)
item.1PL<-rbind(itemParEst(data.ref, model = "1PL"),
itemParEst(data.focal, model = "1PL"))
difLord(irtParam = item.1PL, same.scale = FALSE)

## Pre-estimation of the item parameters (1PL model, "lme4" engine)
item.1PL<-rbind(itemParEst(data.ref, model = "1PL", engine = "lme4"),
itemParEst(data.focal, model = "1PL", engine = "lme4"))
difLord(irtParam = item.1PL, same.scale = FALSE)

## Pre-estimation of the item parameters (2PL model)
item.2PL<-rbind(itemParEst(data.ref, model = "2PL"),
itemParEst(data.focal, model = "2PL"))
difLord(irtParam = item.2PL, same.scale = FALSE)

## Pre-estimation of the item parameters (constrained 3PL model)
item.3PL<-rbind(itemParEst(data.ref, model = "3PL", c = 0.05),
itemParEst(data.focal, model = "3PL", c = 0.05))
difLord(irtParam = item.3PL, same.scale = FALSE)

# Graphical devices
plot(r)
plot(r, plot = "itemCurve", item = 1)
plot(r, plot = "itemCurve", item = 6)

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))
```

```
 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

---

difLRT *Likelihood-Ratio Test DIF method*

Description

Performs DIF detection using Likelihood Ratio Test (LRT) method.

Usage

```
difLRT(Data, group, focal.name, alpha = 0.05, purify = FALSE, nrIter = 10,
  p.adjust.method = NULL, save.output = FALSE, output = c("out", "default"))
## S3 method for class 'LRT'
print(x, ...)
## S3 method for class 'LRT'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
  save.options = c("plot", "default", "pdf"), ...)
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within `data`) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level of `group` which corresponds to the focal group.
`alpha`	numeric: significance level (default is 0.05).
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.

x	the result from a LRT class object.
pch, col	type of usual pch and col graphical options.
number	logical: should the item number identification be printed (default is TRUE).
save.plot	logical: should the plot be saved into a separate file? (default is FALSE).
save.options	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or "default" (default value), and the third component is the file extension, either "pdf" (default) or "jpeg". See **Details**.
...	other generic parameters for the plot or the print functions.

Details

The likelihood-ratio test method (Thissen, Steinberg and Wainer, 1988) allows for detecting uniform differential item functioning by fitting a closed-form Rasch model and by testing for extra interactions between group membership and item response. Currently only the Rasch model can be used, so only uniform DIF can be detected. Moreover, items are tested one by one and the other items act as anchor items.

The Data is a matrix whose rows correspond to the subjects and columns to the items. Missing values are allowed but must be coded as NA values. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument focal.name.

The function glmer from package lme4 (Bates and Maechler, 2009) is used to fit the closed-form Rasch model. More precisely, the probability that response Y_{ijg} of subject *i* from group *g* (focal or reference) to item *j* is modeled as

$$logit(Pr(Y_{ijg} = 1) = \theta_{ig} + \gamma_g - \beta_j$$

where θ_i is subject's ability, β_j is the item difficulty and γ_g is the difference mean ability level between the focal and the reference groups. Subject abilities are treated as random effects, while item difficulties and γ_g are treated as fixed effects. Each item is tested by incorporating an interaction term, δ_{gj}, and by testing its statistical significance using the traditional likelihood-ratio test.

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-squared distribution with lower-tail probability of one minus alpha and one degree of freedom.

Item purification can be performed by setting purify to TRUE. In this case, items detected as DIF are iteratively removed from the set of tested items, and the procedure is repeated (using the remaining items) until no additional item is identified as functioning differently. The process stops when either there is no new item detected as DIF, or when nrIter iterations are run and new DIF items are nevertheless detected. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details.

Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

The output of the difLRT, as displayed by the print.LRT function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of output. The default value is "default", meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The plot.LRT function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the color are fixed by the usual pch and col arguments. Option number permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing save.plot to TRUE allows this process. The figure is defined through the components of save.options. The first two components perform similarly as those of the output argument. The third component is the figure format, with allowed values "pdf" (default) for PDF file and "jpeg" for JPEG file.

Value

A list of class "LRT" with the following arguments:

- LRT: the values of the likelihood-ratio statistics.
- p.value: the vector of p-values for the likelihood-ratio statistics.
- alpha: the value of alpha argument.
- thr: the threshold (cut-score) for DIF detection.
- DIFitems: either the items which were detected as DIF items, or "No DIF item detected".
- p.adjust.method: the value of the p.adjust.method argument.
- adjusted.p: either NULL or the vector of adjusted p-values for multiple comparisons.
- purification: the value of purify option.
- nrPur: the number of iterations in the item purification process. Returned only if purify is TRUE.
- convergence: logical indicating whether the iterative item purification process stopped before the maximal number of allowed iterations (10 by default). Returned only if purify is TRUE.
- names: the names of the items.
- save.output: the value of the save.output argument.
- output: the value of the output argument.

Note

Because of the fitting of the modified Rasch model with glmer, the process can be very time consuming.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See Also

LRT, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal)!="Anger"]

 # Keeping the first 5 items and the first 50 subjects
 # (this is an artificial simplification to reduce the computational time)
 verbal <- verbal[1:50, c(1:5, 25)]

 # Three equivalent settings of the data matrix and the group membership
 r <- difLRT(verbal, group = 6, focal.name = 1)
 difLRT(verbal, group = "Gender", focal.name = 1)
```

```
difLRT(verbal[,1:5], group = verbal[,6], focal.name = 1)

# Multiple comparisons adjustment using Benjamini-Hochberg method
difLRT(verbal, group = 6, focal.name = 1, p.adjust.method = "BH")

# With item purification
difLRT(verbal, group = 6, focal.name = 1, purify = TRUE)

# Saving the output into the "LRTresults.txt" file (and default path)
r <- difLRT(verbal, group = 6, focal.name = 1, save.output = TRUE,
           output = c("LRTresults", "default"))

# Graphical devices
plot(r)

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

# Changing the path, JPEG figure
path <- "c:/Program Files/"
plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

# WARNING: do not trust the results above since they are based on a selected
# subset of the verbal data set!

## End(Not run)
```

`difMantel.poly` *Mantel Differential Item Functionning Detection for Polytomous Items*

Description

Implements the Mantel (1963) test for detecting DIF in polytomous items.

Usage

```
difMantel.poly(data, group, focal.name, ref.name,
             match = "score", sig.level = 0.05,
             purify = FALSE, max.iter = 10)
```

Arguments

`data`	A matrix or data frame of polytomous item responses (one row per subject, one column per item).
`group`	A vector indicating group membership (same length as number of rows in `data`).
`focal.name`	The value in `group` corresponding to the focal group.
`ref.name`	The value in `group` corresponding to the reference group.

match	Specifies the matching variable. Can be "score" (default) for total score or "restscore" to exclude the item being tested from the matching score.
sig.level	Significance level for the DIF test (default = 0.05).
purify	Logical. If TRUE, performs iterative purification to exclude DIF items from the anchor set. Ignored if match = "restscore".
max.iter	Maximum number of purification iterations (default = 10).

Details

Chi-square statistic computed for each item using the generalized Mantel (1963) procedure for ordinal responses. This test evaluates whether the distribution of item responses differs significantly between the reference and focal groups, conditioning on the matching score (either total score or rest score). The statistic asymptotically follows a chi-square distribution with 1 degree of freedom under the null hypothesis of no DIF.

If match = "score", the total test score is used as the matching criterion. If match = "restscore", the item under evaluation is excluded from the score, reducing contamination and improving DIF detection accuracy.

When purify = TRUE, anchor items are iteratively refined: items flagged as DIF (p < sig.level) are excluded from the matching score in subsequent iterations. The process stops when the anchor set stabilizes or after max.iter iterations. If no items remain, the last computed statistics are retained.

For each item, the Mantel statistic is computed. Additionally, Liu–Agresti cumulative odds ratios (Psi_hat, Alpha_hat) and their standard errors (SE_log_Psi) are reported when possible. The logical flag LA.valid indicates whether these estimates could be computed.

Note: All response categories must be observed in both groups for Liu–Agresti estimates to be valid. Missing data should be removed prior to analysis.

Value

A data.frame with one column per item and the following columns:

Stat	The Mantel test statistic.
p.value	Associated p-value for the DIF test.
p.adj	p-value adjusted for multiple comparisons using Holm's method.
Psi_hat	Liu-Agresti's estimate of the odds ratio.
Alpha_hat	Estimated difficulty ratio.
SE_log_Psi	Standard error of the log-odds ratio.
rho.spear	Spearman correlation between item score and matching score.
LA.valid	Logical indicator of whether Liu-Agresti estimates were valid for each item.

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Liu, I., & Agresti, A. (1996). Mantel–Haenszel–Type Inference for Cumulative Odds Ratios with a Stratified Ordinal Response. *Biometrics*, 52(4), 1223–1234.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58, 690–700.

Examples

```
## Not run:
# Real data example
data(SCS)
# Without purification
difMantel.poly(data = SCS[, 1:10], group = SCS$Gender, focal.name = 1,
ref.name = 2, purify = FALSE)

# Without purification and restscore
difMantel.poly(data = SCS[, 1:10], group = SCS$Gender, focal.name = 1,
ref.name = 2, purify = TRUE,match = "restscore")

# With purification
difMantel.poly(data = SCS[, 1:10], group = SCS$Gender, focal.name = 1,
ref.name = 2, purify = TRUE)

# With simulated data

set.seed(1234)

# original item parameters
a <- rlnorm(10, -0.5)  # slopes
b <- runif(10, -2, 2)  # difficulty
d <- list()
d[[1]] <- c(0, 2, .5, -.15, -1.1)
d[[2]] <- c(0, 2, .25, -.45, -.75)
d[[3]] <- c(0, 1, .5, -.65, -1)
d[[4]] <- c(0, 2, .5, -.85, -2)
d[[5]] <- c(0, 1, .25, -.05, -1)
d[[6]] <- c(0, 2, .5, -.95, -1)
d[[7]] <- c(0, 1, .25, -.35, -2)
d[[8]] <- c(0, 2, .5, -.15, -1)
d[[9]] <- c(0, 1, .25, -.25, -2)
d[[10]] <- c(0, 2, .5, -.35, -1)

# Uniform DIF
It <- 10; NR <- 1000; NF <- 1000
ItDIFa <- NULL; Ga <- NULL
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2)

Out.Unif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                       ncat = 5, Ga = Ga, Gb = Gb)
Out.Unif$ipars
```

```
Data <- Out.Unif$data

# Without purification and rest score
difMantel.poly(data = Data[, 1:10], group = Data$group, focal.name = "G1",
ref.name = "G2", purify = FALSE,match = "restscore")

# With purification
difMantel.poly(data = Data[, 1:10], group = Data$group, focal.name = "G1",
ref.name = "G2", purify = TRUE)

# We implemented a specific S3 plot method: plot.Logistic. It can be used as follows:
res <- difMantel.poly(data = Data[, 1:10], group = Data$group, focal.name = "G1",
ref.name = "G2", purify = FALSE)
plot.MHPoly(res)

## End(Not run)
```

---

difMH *Mantel-Haenszel DIF method*

Description

Performs DIF detection using Mantel-Haenszel method.

Usage

```
difMH(Data, group, focal.name , anchor = NULL, match = "score", MHstat = "MHChisq",
   correct = TRUE, exact = FALSE, alpha = 0.05, purify = FALSE, nrIter = 10,
   p.adjust.method = NULL, save.output = FALSE, output = c("out", "default"))
## S3 method for class 'MH'
print(x, ...)
## S3 method for class 'MH'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
focal.name	numeric or character indicating the level of group which corresponds to the focal group.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.

`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `Data`. See **Details**.
`MHstat`	character: specifies the DIF statistic to be used for DIF identification. Possible values are `"MHChisq"` (default) and `"logOR"`. See **Details** .
`correct`	logical: should the continuity correction be used? (default is `TRUE`)
`exact`	logical: should an exact test be computed? (default is `FALSE`).
`alpha`	numeric: significance level (default is 0.05).
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
`x`	the result from a `MH` class object.
`pch, col`	type of usual `pch` and `col` graphical options.
`number`	logical: should the item number identification be printed (default is `TRUE`).
`save.plot`	logical: should the plot be saved into a separate file? (default is `FALSE`).
`save.options`	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
`...`	other generic parameters for the `plot` or the `print` functions.

Details

The method of Mantel-Haenszel (1959) allows for detecting uniform differential item functioning without requiring an item response model approach.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded from sum-score computation.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument `focal.name`.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `mantelHaenszel` function. This is specified by the `match` argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

The DIF statistic is specified by the MHstat argument. By default, MHstat takes the value "MHChisq" and the Mantel-Haenszel chi-square statistic is used. The other optional value is "logOR", and the log odds-ratio statistic (that is, the log of alphaMH divided by the square root of varLambda) is used. See Penfield and Camilli (2007), Philips and Holland (1987) and mantelHaenszel help file.

By default, the asymptotic Mantel-Haenszel statistic is computed. However, the exact statistics and related P-values can be obtained by specifying the logical argument exact to TRUE. See Agresti (1990, 1992) for further details about exact inference.

The threshold (or cut-score) for classifying items as DIF depends on the DIF statistic. With the Mantel-Haenszel chi-squared statistic (MHstat=="MHChisq"), it is computed as the quantile of the chi-square distribution with lower-tail probability of one minus alpha and with one degree of freedom. With the log odds-ratio statistic (MHstat=="logOR"), it is computed as the quantile of the standard normal distribution with lower-tail probability of 1-alpha/2. With exact inference, it is simply the alpha level since exact P-values are returned.

By default, the continuity correction factor -0.5 is used (Holland and Thayer, 1988). One can nevertheless remove it by specifying correct=FALSE.

In addition, the Mantel-Haenszel estimates of the common odds ratios α_{MH} are used to measure the effect sizes of the items. These are obtained by $\Delta_{MH} = -2.35 \log \alpha_{MH}$ (Holland and Thayer, 1985). According to the ETS delta scale, the effect size of an item is classified as negligible if $|\Delta_{MH}| \leq 1$, moderate if $1 \leq |\Delta_{MH}| \leq 1.5$, and large if $|\Delta_{MH}| \geq 1.5$. The values of the effect sizes, together with the ETS classification, are printed with the output. Note that this is returned only for asymptotic tests, i.e. when exact is FALSE.

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item was detected as functioning differently at some step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). By default it is NULL so that no anchor item is specified.

The output of the `difMH`, as displayed by the `print.MH` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The `plot.MH` function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the color are fixed by the usual `pch` and `col` arguments. Option `number` permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file. Note that no plot is returned for exact inference.

Value

A list of class "MH" with the following arguments:

- `MH` the values of the Mantel-Haenszel DIF statistics (either exact or asymptotic).
- `p.value` the p-values for the Mantel-Haenszel statistics (either exact or asymptotic).
- `alphaMH` the values of the mantel-Haenszel estimates of common odds ratios. Returned only if `exact` is `FALSE`.
- `varLambda` the values of the variances of the log odds-ratio statistics. Returned only if `exact` is `FALSE`.
- `MHstat` the value of the `MHstat` argument. Returned only if `exact` is `FALSE`.
- `alpha` the value of `alpha` argument.
- `thr` the threshold (cut-score) for DIF detection. Returned only if `exact` is `FALSE`.
- `DIFitems` either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
- `correct` the value of `correct` option.
- `exact` the value of `exact` option.
- `match` a character string, either `"score"` or `"matching variable"` depending on the `match` argument.
- `p.adjust.method` the value of the `p.adjust.method` argument.
- `adjusted.p` either `NULL` or the vector of adjusted p-values for multiple comparisons.
- `purification` the value of `purify` option.
- `nrPur` the number of iterations in the item purification process. Returned only if `purify` is `TRUE`.
- `difPur` a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if `purify` is `TRUE`.

convergence	logical indicating whether the iterative item purification process stopped before the maximal number nrIter of allowed iterations. Returned only if purify is TRUE.
names	the names of the items.
anchor.names	the value of the anchor argument.
save.output	the value of the save.output argument.
output	the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7*, 131-177. doi:10.1214/ss/1177011454

Holland, P. W. and Thayer, D. T. (1985). An alternative definition of the ETS delta scale of item difficulty. *Research Report RR-85-43*. Princeton, NJ: Educational Testing Service.

Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Ed.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Penfield, R. D., and Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao and S. Sinharray (Eds.), *Handbook of Statistics 26: Psychometrics* (pp. 125-167). Amsterdam, The Netherlands: Elsevier.

Philips, A., and Holland, P. W. (1987). Estimators of the Mantel-Haenszel log odds-ratio estimate. *Biometrics, 43*, 425-431. doi:10.2307/2531824

Raju, N. S., Bode, R. K. and Larsen, V. S. (1989). An empirical assessment of the Mantel-Haenszel statistic to detect differential item functioning. *Applied Measurement in Education, 2*, 1-13. doi:10.1207/s15324818ame0201_1

Uttaro, T. and Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement, 18*, 15-25. doi:10.1177/014662169401800102

See Also

mantelHaenszel, dichoDif, p.adjust

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal) != "Anger"]

 # Three equivalent settings of the data matrix and the group membership
 r <- difMH(verbal, group = 25, focal.name = 1)
 difMH(verbal, group = "Gender", focal.name = 1)
 difMH(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # With log odds-ratio statistic
 r2 <- difMH(verbal, group = 25, focal.name = 1, MHstat = "logOR")

 # With exact inference
 difMH(verbal, group = 25, focal.name = 1, exact = TRUE)

# Multiple comparisons adjustment using Benjamini-Hochberg method
 difMH(verbal, group = 25, focal.name = 1, p.adjust.method = "BH")

 # With item purification
 difMH(verbal, group = "Gender", focal.name = 1, purify = TRUE)
 difMH(verbal, group = "Gender", focal.name = 1, purify = TRUE, nrIter = 5)

 # Without continuity correction and with 0.01 significance level
 difMH(verbal, group = "Gender", focal.name = 1, alpha = 0.01, correct = FALSE)

 # With items 1 to 5 set as anchor items
 difMH(verbal, group = "Gender", focal.name = 1, anchor = 1:5)
 difMH(verbal, group = "Gender", focal.name = 1, anchor = 1:5, purify = TRUE)

 # Saving the output into the "MHresults.txt" file (and default path)
 r <- difMH(verbal, group = 25, focal.name = 1, save.output = TRUE,
           output = c("MHresults","default"))

 # Graphical devices
 plot(r)
```

```
 plot(r2)

 # Plotting results and saving it in a PDF figure
 plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

`difPolyLogistic` *Logistic regression DIF statistics for polytomous (ordinal) items*

Description

Computes DIF detection using logistic regression models for ordinal (polytomous) items.

Usage

```
difPolyLogistic(Data, group, focal.name, anchor = NULL, member.type = "group",
match = "score", type = "both", criterion = "LRT", alpha = 0.05, all.cov=FALSE,
purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
output = c("out", "default"))
```

Arguments

`Data`	a data frame or matrix: one row per respondent, one column per item. Items must be coded as ordinal variables.
`group`	a vector or column index/name from Data: specifies group membership.
`focal.name`	the label identifying the focal group (ignored if `member.type` = `"cont"`).
`anchor`	a vector of column indices or names specifying anchor (non-DIF) items. If `NULL` and `purify` = `FALSE`, all items are used as anchors. Ignored if `match` is not `"score"`.
`member.type`	`"group"` (default) if `group` is categorical; `"cont"` if `group` is continuous.
`match`	matching criterion. Use `"score"` for test score, `"restscore"` for item-excluded score, or provide an external continuous/discrete vector.
`type`	DIF type to test: `"both"` (default), `"udif"` (uniform DIF only), or `"nudif"` (non-uniform DIF only).
`criterion`	test statistic: `"LRT"` (default) for likelihood ratio test or `"Wald"` for Wald test.
`alpha`	significance level for DIF detection (default = 0.05).
`all.cov`	logical: whether to return full covariance matrices of the parameter estimates. Default is `FALSE`.

`purify`	logical: whether to apply iterative purification to refine anchor items (default = `FALSE`). Requires `match = "score"`.
`nrIter`	maximum number of iterations for purification (default = 10).
`p.adjust.method`	method for p-value adjustment across items (e.g., `"BH"`, `"bonferroni"`). Default = `NULL`.
`save.output`	logical: if `TRUE`, saves output to a text file.
`output`	character vector: `output\[1]` is the filename (without extension), `output\[2]` is the directory path (or `"default"` for working directory).

Details

The function fits cumulative ordinal logistic regression models (via `VGAM::vglm`) to detect DIF in polytomous items.

Three nested models are fit per item and compared to assess DIF:

- M_0: No DIF (only match effect)
- M_1: Uniform DIF (match + group)
- M_2: Uniform + non-uniform DIF (match + group + interaction)

M_0, M_1, and M_2 are compared using either likelihood-ratio or Wald tests, depending on the `criterion` argument.

When `match = "restscore"`, the matching variable is defined as the sum score excluding the item being tested.

When `purify = TRUE`, the algorithm iteratively refines the anchor set by excluding detected DIF items and updating scores.

This function handles both group-based DIF (`member.type = "group"`) and DIF based on continuous moderators (`member.type = "cont"`).

For each item, the DIF analysis is performed using only complete cases. Respondents with missing data on the item being tested, the matching variable, or the group variable are excluded from the estimation for that item.

For very strong predictors (e.g., matching variables that nearly perfectly separate response categories), the underlying ordinal regression models may become numerically unstable. This can result in extreme coefficients, saturation warnings, and possibly negative pseudo-R^2 values. These cases reflect data properties rather than programming errors.

Value

Returns an object of class `"Logistic"`, a list with elements:

`LogistikPoly`	numeric vector of DIF test statistics for each item.
`p.value`	p-values associated with each test statistic.
`logitPar`	matrix of estimated parameters for best-fitting models (per item).
`logitSe`	matrix of standard errors for `logitPar`.
`parM0`, `seM0`	parameter estimates and SEs for the null model (no DIF).

`cov.M0`, `cov.M1` covariance matrices for null and full models (if `all.cov = TRUE`).

`deltaR2` effect sizes (McKelvey & Zavoina R^2) per item.

`alpha`, `thr` alpha value and corresponding statistical threshold.

`DIFitems` indices of items detected as DIF (or `"No DIF item detected"`).

`type`, `criterion`, `match`, `member.type`
echoed inputs.

`p.adjust.method`, `adjusted.p`
if adjustment requested, adjusted p-values and method used.

`purification`, `nrPur`, `difPur`, `convergence`
details of the purification process.

`names`, `anchor.names`
item names and anchor items used.

`save.output`, `output`
output options echoed.

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
`<sebastien.beland@umontreal.ca>`

References

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Department of National Defense.

Zumbo, B. D. & Thomas, D. R. (1997). A measure of effect size for a model-based approach for studying DIF. *Educational and Psychological Measurement, 57*(4), 679-688.

See Also

`LogistikPoly`, `VGAM::vglm`

Examples

```
## Not run:

# With real data

data(SCS)

# Without item purification
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=FALSE)

# Without item purification and the rest score
```

```
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=FALSE,, match = "restscore")

# With item purification
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=TRUE)

# With item purification
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=TRUE)

# With item purification with LRT criterion
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=TRUE, criterion = "LRT")

# With item purification with LRT criterion and alpha = 0.01
difPolyLogistic(SCS[,1:10], group=SCS[,11],
focal.name = "1", purify=TRUE, criterion = "LRT", alpha = 0.01)

# With simulated data

set.seed(1234)

# original item parameters
a <- rlnorm(10, -0.5)  # slopes
b <- runif(10, -2, 2)  # difficulty
d <- list()
d[[1]] <- c(0, 2, .5, -.15, -1.1)
d[[2]] <- c(0, 2, .25, -.45, -.75)
d[[3]] <- c(0, 1, .5, -.65, -1)
d[[4]] <- c(0, 2, .5, -.85, -2)
d[[5]] <- c(0, 1, .25, -.05, -1)
d[[6]] <- c(0, 2, .5, -.95, -1)
d[[7]] <- c(0, 1, .25, -.35, -2)
d[[8]] <- c(0, 2, .5, -.15, -1)
d[[9]] <- c(0, 1, .25, -.25, -2)
d[[10]] <- c(0, 2, .5, -.35, -1)

# Uniform DIF
It <- 10; NR <- 1000; NF <- 1000
ItDIFa <- NULL; Ga <- NULL
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2)

Out.Unif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                       ncat = 5, Ga = Ga, Gb = Gb)
Out.Unif$ipars
Data <- Out.Unif$data

# Without item purification
difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=FALSE)
```

```
# Without item purification and restscore
difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=FALSE, match = "restscore")

# With item purification
difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=TRUE)

# With item purification with LRT criterion
difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=TRUE, criterion = "LRT")

# With item purification with LRT criterion and alpha = 0.01
difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=TRUE, criterion = "LRT", alpha = 0.01)

# We implemented a specific S3 plot method: plot.Logistic. It can be used as follows

res <- difPolyLogistic(Out.Unif$data[,1:10], group=Out.Unif$data[,11],
focal.name = "G1", purify=FALSE)
plot.Logistic(res)


## End(Not run)
```

`difQuade` *Detection of Differential Item Functioning Using Quade-Type Association Indices for Polytomous (Ordinal) Item*

Description

This function detects DIF in ordinal items using association indices based on pairwise comparisons, as proposed by Quade (1974) and extended in Woods (2009). It supports various ordinal measures of association to identify uniform DIF only.

Usage

```
difQuade(Data, group, focal.name = NULL, anchor = NULL,
         match = "score", type = c("ta", "e", "dxy", "dyx", "gamma"),
         alpha = 0.05, purify = FALSE, nrIter = 10,
         save.output = FALSE, output = c("out", "default"))
```

Arguments

`Data`	A data frame or matrix of ordinal item responses.
`group`	A vector indicating group membership.
`focal.name`	Value in `group` identifying the focal group.

`anchor`	Optional vector of anchor item indices. If `NULL`, all items are used.
`match`	Type of matching score: `"score"` (total test score) or `"restscore"` (excluding item).
`type`	Type of ordinal association index: `"ta"` (Kendall's tau-a), `"e"` (Wilson's e), `"gamma"` (Goodman & Kruskal's gamma), `"dyx"` (Somers' dyx), or `"dxy"` (Somers' dxy).
`alpha`	Significance level for DIF detection.
`purify`	Logical: should purification be applied?
`nrIter`	Number of iterations for purification.
`save.output`	Logical: should the results be saved to a text file?
`output`	Name of the output file (or `"out"` to use default).

Details

The function implements the ordinal association approach introduced by Quade (1974), where pairwise comparisons are made between respondents' item responses and total scores. Five indices are supported:

- `"ta"`: Kendall's tau-a, considers all pair types including ties.
- `"e"`: Wilson's e index, accounts for ties in both variables.
- `"gamma"`: Goodman & Kruskal's gamma, ignores tied pairs.
- `"dyx"`: Somers' dyx, conditions on the matching score.
- `"dxy"`: Somers' dxy, conditions on the item response.

These indices follow the methodology validated in Woods (2009), who confirmed through simulation their robustness across various ordinal DIF contexts.

Value

An object of class `"difQuade"` with components:

- `stat`, `se`, `zstat`, `p.value`: test statistics, standard errors, z-values, and p-values for each item.
- `DIFitems`: Logical vector indicating flagged items.
- `match`, `type`, `anchor`, `purification`: arguments used.

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Quade, D. (1974). Nonparametric tests for the comparison of two groups of multivariate observations. *The Annals of Statistics*, 2(5), 949–960.

Woods, C. M. (2009). Testing for differential item functioning with measures of partial association. *Applied Psychological Measurement*, 33(7), 538–554.

See Also

`plot.difQuade`, `print.difQuade`

Examples

```
## Not run:
# With real data
# DIF detection using tau-a and purification
data(SCS)
Data <- SCS[, 1:10]
group <- SCS$Gender

# Using ta and purification
res1 <- difQuade(Data, group, focal.name = 2,
type = "ta", purify = TRUE)
print(res1)
# Here is a function thta plot the results
plot(res1)

# Using Goodman & Kruskal's gamma with restscore matching
res2 <- difQuade(Data, group, focal.name = 2,
type = "gamma", match = "restscore")
print(res2)

# Using Wilson's e index (recommended for tied ordinal data)
res3 <- difQuade(Data, group, focal.name = 2,
type = "e")
print(res3)

# Somers' dyx index with no purification
res4 <- difQuade(Data, group, focal.name = 2,
type = "dyx", purify = FALSE)
print(res4)

# With simulated data

set.seed(1234)

# original item parameters
a <- rlnorm(10, -0.5)  # slopes
b <- runif(10, -2, 2)  # difficulty
d <- list()
d[[1]] <- c(0, 2, .5, -.15, -1.1)
d[[2]] <- c(0, 2, .25, -.45, -.75)
```

```
d[[3]] <- c(0, 1, .5, -.65, -1)
d[[4]] <- c(0, 2, .5, -.85, -2)
d[[5]] <- c(0, 1, .25, -.05, -1)
d[[6]] <- c(0, 2, .5, -.95, -1)
d[[7]] <- c(0, 1, .25, -.35, -2)
d[[8]] <- c(0, 2, .5, -.15, -1)
d[[9]] <- c(0, 1, .25, -.25, -2)
d[[10]] <- c(0, 2, .5, -.35, -1)

# Uniform DIF
It <- 10; NR <- 1000; NF <- 1000
ItDIFa <- NULL; Ga <- NULL
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2)

Out.Unif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                       ncat = 5, Ga = Ga, Gb = Gb)
Out.Unif$ipars
Data <- Out.Unif$data

# Using ta and purification
res5 <- difQuade(Data = Data[, 1:10], group = Data$group,
focal.name = "G1", type = "ta", purify = TRUE)
print(res5)
# Here is a function thta plot the results
plot(res5)

# Using Goodman & Kruskal's gamma with restscore matching
res6 <- difQuade(Data = Data[, 1:10], group = Data$group,
focal.name = "G1", type = "gamma", match = "restscore")
print(res6)

# Using Wilson's e index (recommended for tied ordinal data)
res7 <- difQuade(Data = Data[, 1:10], group = Data$group,
focal.name = "G1", type = "e")
print(res7)

# Somers' dyx index with no purification
res8 <- difQuade(Data = Data[, 1:10], group = Data$group,
focal.name = "G1", type = "dyx", purify = FALSE)
print(res8)

## End(Not run)
```

---

difRaju *Raju's area DIF method*

Description

Performs DIF detection using Raju's area method.

Usage

```
difRaju(Data, group, focal.name, model, c = NULL, engine = "ltm", discr = 1,
  irtParam = NULL,  same.scale = TRUE, anchor = NULL, alpha = 0.05,
  signed = FALSE, purify = FALSE, nrIter = 10, p.adjust.method = NULL,
  save.output = FALSE, output = c("out","default"))
## S3 method for class 'Raj'
print(x, ...)
## S3 method for class 'Raj'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
  save.options = c("plot","default","pdf"), ...)
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within `data`) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level of `group` which corresponds to the focal group.
`model`	character: the IRT model to be fitted (either `"1PL"`, `"2PL"` or `"3PL"`).
`c`	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
`engine`	character: the engine for estimating the 1PL model, either `"ltm"` (default) or `"lme4"`.
`discr`	either `NULL` or a real positive value for the common discrimination parameter (default is 1). Used only if `model` is `"1PL"` and `engine` is `"ltm"`. See **Details**.
`irtParam`	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
`same.scale`	logical: are the item parameters of the `irtParam` matrix on the same scale? (default is "TRUE"). See **Details**.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`alpha`	numeric: significance level (default is 0.05).
`signed`	logical: should the Raju's statistics be computed using the signed (`TRUE`) or unsigned (`FALSE`, default) area? See **Details**.
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).

`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
`x`	the result from a `Raj` class object.
`pch, col`	type of usual `pch` and `col` graphical options.
`number`	logical: should the item number identification be printed (default is `TRUE`).
`save.plot`	logical: should the plot be saved into a separate file? (default is `FALSE`).
`save.options`	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
`...`	other generic parameters for the `plot` or the `print` functions.

Details

Raju's area method (Raju, 1988, 1990) allows for detecting uniform or non-uniform differential item functioning by setting an appropriate item response model. The input can be of two kinds: either by displaying the full data, the group membership and the model, or by giving the item parameter estimates (with the option `irtParam`). Both can be supplied, but in this case only the parameters in `irtParam` are used for computing Raju's statistic.

By default, the Raju's Z statistics are obtained by using the *unsigned* areas between the ICCs. However, these statistics can also be computed using the *signed* areas, by setting the argument `signed` to `TRUE` (default value is `FALSE`). See `RajuZ` for further details.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded for item parameter estimation.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument `focal.name`.

If the model is not the 1PL model, or if `engine` is equal to `"ltm"`, the selected IRT model is fitted using marginal maximum likelihood by means of the functions from the `ltm` package (Rizopoulos, 2006). Otherwise, the 1PL model is fitted as a generalized linear mixed model, by means of the `glmer` function of the `lme4` package (Bates and Maechler, 2009).

With the `"1PL"` model and the `"ltm"` engine, the common discrimination parameter is set equal to 1 by default. It is possible to fix another value through the argument`discr`. Alternatively, this common discrimination parameter can be estimated (though not returned) by fixing `discr` to `NULL`.

The 3PL model can be fitted either unconstrained (by setting `c` to `NULL`) or by fixing the pseudo-guessing values. In the latter case, the argument `c` holds either a numeric vector of same length of the number of items, with one value per item pseudo-guessing parameter, or a single value which is duplicated for all the items. If `c` is different from `NULL` then the 3PL model is always fitted (whatever the value of `model`).

The `irtParam` matrix has a number of rows equal to twice the number of items in the data set. The first J rows refer to the item parameter estimates in the reference group, while the last J ones correspond to the same items in the focal group. The number of columns depends on the selected IRT model: 2 for the 1PL model, 5 for the 2PL model, 6 for the constrained 3PL model and 9 for the unconstrained 3PL model. The columns of irtParam have to follow the same structure as the output of itemParEst command (the latter can actually be used to create the irtParam matrix).

In addition to the matrix of parameter estimates, one has to specify whether items in the focal group were rescaled to those of the reference group. If not, rescaling is performed by equal means anchoring (Cook and Eignor, 1991). Argument same.scale is used for this choice (default option is TRUE and assumes therefore that the parameters are already placed on the same scale).

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the standard normal distribution with lower-tail probability of 1-alpha/2.

Item purification can be performed by setting purify to TRUE. In this case, the purification occurs in the equal means anchoring process. Items detected as DIF are iteratively removed from the set of items used for equal means anchoring, and the procedure is repeated until either the same items are identified twice as functioning differently, or when nrIter iterations have been performed. In the latter case a warning message is printed. See Candell and Drasgow (1988) for further details.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to rescale the item parameters on a common metric. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). By default it is NULL so that no anchor item is specified. If item parameters are provided thorugh the irtParam argument and if they are on the same scale (i.e. if same.scale is TRUE), then anchor items are not used (even if they are specified).

Under the 1PL model, the displayed output also proposes an effect size measure, which is -2.35 times the difference between item difficulties of the reference group and the focal group (Penfield and Camilli, 2007, p. 138). This effect size is similar Mantel-Haenszel's Δ_{MH} effect size, and the ETS delta scale is used to classify the effect sizes (Holland and Thayer, 1985).

The output of the difRaju, as displayed by the print.Raj function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of output. The default value is "default", meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The plot.Raj function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the color are fixed by the usual pch and col arguments. Option number permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing save.plot to TRUE allows this process. The figure is defined through the components of save.options. The first two components perform similarly as those of the output argument. The third component is the figure format, with allowed values "pdf" (default) for PDF file and "jpeg" for JPEG file.

Value

A list of class "Raj" with the following arguments:

RajuZ	the values of the Raju's statistics.
p.value	the p-values for the Raju's statistics.
alpha	the value of alpha argument.
thr	the threshold (cut-score) for DIF detection.
DIFitems	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
signed	the value of the signed argument.
p.adjust.method	the value of the p.adjust.method argument.
adjusted.p	either NULL or the vector of adjusted p-values for multiple comparisons.
purification	the value of purify option.
nrPur	the number of iterations in the item purification process. Returned only if purify is TRUE.
difPur	a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if purify is TRUE.
convergence	logical indicating whether the iterative item purification process stopped before the maximal number nrIterof allowed iterations. Returned only if purify is TRUE.
model	the value of model argument.
c	The value of the c argument.
engine	The value of the engine argument.
discr	the value of the discr argument.
itemParInit	the matrix of initial parameter estimates,with the same format as irtParam either provided by the user (through irtParam) or estimated from the data (and displayed without rescaling).
itemParFinal	the matrix of final parameter estimates, with the same format as irtParam, obtained after item purification. Returned only if purify is TRUE.
estPar	a logical value indicating whether the item parameters were estimated (TRUE) or provided by the user (FALSE).
names	the names of the items.
anchor.names	the value of the anchor argument.
save.output	the value of the save.output argument.
output	the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Candell, G.L. and Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253–260. doi:10.1177/014662168801200304

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Holland, P. W. and Thayer, D. T. (1985). An alternative definition of the ETS delta scale of item difficulty. *Research Report RR-85-43*. Princeton, NJ: Educational Testing Service.

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R. D., and Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao and S. Sinharray (Eds.), *Handbook of Statistics 26: Psychometrics* (pp. 125-167). Amsterdam, The Netherlands: Elsevier.

Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495-502. doi:10.1007/BF02294403

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. doi:10.1177/014662169001400208

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1-25. doi:10.18637/jss.v017.i05

See Also

`RajuZ`, `itemParEst`, `dichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal<-verbal[colnames(verbal)!="Anger"]

 # Three equivalent settings of the data matrix and the group membership
 # (1PL model, "ltm" engine)
 difRaju(verbal, group = 25, focal.name = 1, model = "1PL")
 difRaju(verbal, group = "Gender", focal.name = 1, model = "1PL")
 difRaju(verbal[,1:24], group = verbal[,25], focal.name = 1, model = "1PL")

 # Multiple comparisons adjustment using Benjamini-Hochberg method
 difRaju(verbal, group = 25, focal.name = 1, model = "1PL", p.adjust.method = "BH")

 # With signed areas
 difRaju(verbal, group = 25, focal.name = 1, model = "1PL", signed = TRUE)

 # With items 1 to 5 set as anchor items
 difRaju(verbal, group = 25, focal.name = 1, model = "1PL", anchor = 1:5)

 # (1PL model, "lme4" engine)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "1PL",
 engine = "lme4")

 # 2PL model, signed and unsigned areas
 difRaju(verbal, group = "Gender", focal.name = 1, model = "2PL")
 difRaju(verbal, group = "Gender", focal.name = 1, model = "2PL", signed = TRUE)

 # 3PL model with all pseudo-guessing parameters constrained to 0.05
 # Signed and unsigned areas
 difRaju(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05,
   signed = TRUE)

 # Same models, with item purification
 difRaju(verbal, group = "Gender", focal.name = 1, model = "1PL", purify = TRUE)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "2PL", purify = TRUE)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05,
 purify = TRUE)

 # With signed areas
 difRaju(verbal, group = "Gender", focal.name = 1, model = "1PL", purify = TRUE,
   signed = TRUE)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "2PL", purify = TRUE,
   signed = TRUE)
 difRaju(verbal, group = "Gender", focal.name = 1, model = "3PL", c = 0.05,
 purify = TRUE, signed = TRUE)
```

```
## Splitting the data into reference and focal groups
nF<-sum(Gender)
nR<-nrow(verbal)-nF
data.ref<-verbal[,1:24][order(Gender),][1:nR,]
data.focal<-verbal[,1:24][order(Gender),][(nR+1):(nR+nF),]

## Pre-estimation of the item parameters (1PL model, "ltm" engine)
item.1PL<-rbind(itemParEst(data.ref,model = "1PL"),
itemParEst(data.focal,model = "1PL"))
difRaju(irtParam = item.1PL,same.scale = FALSE)

## Pre-estimation of the item parameters (1PL model, "lme4" engine)
item.1PL<-rbind(itemParEst(data.ref, model = "1PL", engine = "lme4"),
itemParEst(data.focal, model = "1PL", engine = "lme4"))
difRaju(irtParam = item.1PL, same.scale = FALSE)

## Pre-estimation of the item parameters (2PL model)
item.2PL<-rbind(itemParEst(data.ref, model = "2PL"),
itemParEst(data.focal, model = "2PL"))
difRaju(irtParam = item.2PL, same.scale = FALSE)

## Pre-estimation of the item parameters (constrained 3PL model)
item.3PL<-rbind(itemParEst(data.ref, model = "3PL", c = 0.05),
itemParEst(data.focal, model = "3PL", c = 0.05))
difRaju(irtParam = item.3PL, same.scale = FALSE)

# Saving the output into the "RAJUresults.txt" file (and default path)
r <- difRaju(verbal, group = 25, focal.name = 1, model = "1PL",
         save.output = TRUE, output = c("RAJUresults","default"))

# Graphical devices
plot(r)

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

# Changing the path, JPEG figure
path <- "c:/Program Files/"
plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

---

difSIBTEST *SIBTEST and Crossing-SIBTEST DIF method*

Description

Performs DIF detection using SIBTEST (Shealy and Stout, 1993) or the modified Crossing-SIBTEST method (Chalmers, 2018).

Usage

```
difSIBTEST(Data, group, focal.name, type = "udif", anchor = NULL, alpha = 0.05,
   purify = FALSE, nrIter = 10, p.adjust.method = NULL,
   save.output = FALSE, output = c("out", "default"))
## S3 method for class 'SIBTEST'
print(x, ...)
## S3 method for class 'SIBTEST'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ...)
```

Arguments

`Data` numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.

`group` numeric or character: either the vector of group membership or the column indicator (within `data`) of group membership. See **Details**.

`focal.name` numeric or character indicating the level of `group` which corresponds to the focal group.

`type` character: the type of DIF effect to test. Possible values are `"udif"` (default) for uniform DIF using SIBTEST, or `"nudif"` for nonuniform DIF using Crossing-SIBTEST (CSIBTEST).

`anchor` either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.

`alpha` numeric: significance level (default is 0.05).

`purify` logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).

`nrIter` numeric: the maximal number of iterations in the item purification process (default is 10).

`p.adjust.method` either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.

`save.output` logical: should the output be saved into a text file? (Default is `FALSE`).

`output` character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.

`x` the result from a `SIBTEST` class object.

`pch, col` type of usual `pch` and `col` graphical options.

`number` logical: should the item number identification be printed (default is `TRUE`).

`save.plot` logical: should the plot be saved into a separate file? (default is `FALSE`).

`save.options` character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.

`...` other generic parameters for the `plot` or the `print` functions.

Details

The SIBTEST method (Shealy and Stout, 1993) allows for detecting uniform differential item functioning without requiring an item response model approach. Its modified version, the Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996), focuses on nonuniform DIF instead. This function provides a wrapper to the SIBTEST function from the **mirt** package (Chalmers, 2012) to fit within the difR framework (Magis et al., 2010). Therefore, if you are using this function for publication purposes please cite Chalmers (2018; 2012) and Magis et al. (2010).

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument focal.name.

The type of DIF effect, uniform through SIBTEST or nonuniform through Crossing-SIBTEST, is determined by the type argument. By default it is "udif" for uniform DIF, and may take the value "nudif" for nonuniform DIF.

The threshold (or cut-score) for classifying items as DIF is computed as the quantile of the chi-square distribution with lower-tail probability of one minus alpha and with one degree of freedom. Note that the degrees of freedom are also returned by the df argument.

Item purification can be performed by setting purify to TRUE. Purification works as follows: if at least one item was detected as functioning differently at some step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when nrIter iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. The latter must be an acronym of one of the available adjustment methods of the p.adjust function. According to Kim and Oshima (2013), Holm and Benjamini-Hochberg adjustments (set respectively by "Holm" and "BH") perform best for DIF purposes. See p.adjust function for further details. Note that item purification is performed on original statistics and p-values; in case of adjustment for multiple comparisons this is performed *after* item purification.

A pre-specified set of anchor items can be provided through the anchor argument. It must be a vector of either item names (which must match exactly the column names of Data argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if purify is set to TRUE). By default it is NULL so that no anchor item is specified.

The output of the difSIBTEST, as displayed by the print.SIBTEST function, can be stored in a text file provided that save.output is set to TRUE (the default value FALSE does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the output argument (default name is "out"), and the path for saving the text file can be given through the second component of output. The default value is "default", meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The plot.SIBTEST function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the color are fixed by the usual pch and col arguments. Option number permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing save.plot to TRUE allows this process. The figure is defined through the components of save.options. The first two components perform similarly as those of the output argument. The third component is the figure format, with allowed values "pdf" (default) for PDF file and "jpeg" for JPEG file. Note that no plot is returned for exact inference.

Value

A list of class "SIBTEST" with the following arguments:

- `Beta` the values of the SIBTEST Beta values.
- `SE` the standard errors of the Beta values.
- `X2` the values of the SIBTEST or Crossing-SITBTEST chi-square statistics.
- `df` the degrees of freedom for X2 statistics.
- `p.value` the p-values for the SIBTEST or Crossing-SIBTEST statistics.
- `type` the value of the type argument.
- `alpha` the value of alpha argument.
- `DIFitems` either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
- `p.adjust.method` the value of the p.adjust.method argument.
- `adjusted.p` either NULL or the vector of adjusted p-values for multiple comparisons.
- `purification` the value of purify option.
- `nrPur` the number of iterations in the item purification process. Returned only if purify is TRUE.
- `difPur` a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if purify is TRUE.
- `convergence` logical indicating whether the iterative item purification process stopped before the maximal number nrIter of allowed iterations. Returned only if purify is TRUE.
- `names` the names of the items or NULL if the items have no name.
- `anchor.names` the value of the anchor argument.
- `save.output` the value of the save.output argument.
- `output` the value of the output argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium

References

Chalmers, R. P. (2012). mirt: A Multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48(6)*, 1-29. doi:10.18637/jss.v048.i06

Chalmers, R. P. (2018). Improving the Crossing-SIBTEST statistic for detecting non-uniform DIF. *Psychometrika, 83*(2), 376–386. doi:10.1007/s1133601795838

Kim, J., and Oshima, T. C. (2013). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement, 73*, 458–470. doi:10.1177/0013164412467033

Li, H.-H., and Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*, 647–677. doi:10.1007/BF02294041

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Shealy, R. and Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194. doi:10.1007/BF02294572

See Also

`sibTest`, `dichoDif`, `p.adjust`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal) != "Anger"]

 # Three equivalent settings of the data matrix and the group membership
 r <- difSIBTEST(verbal, group = 25, focal.name = 1)
 difSIBTEST(verbal, group = "Gender", focal.name = 1)
 difSIBTEST(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # Test for nonuniform DIF
 difSIBTEST(verbal, group = 25, focal.name = 1, type = "nudif")

 # Multiple comparisons adjustment using Benjamini-Hochberg method
 difSIBTEST(verbal, group = 25, focal.name = 1, p.adjust.method = "BH")
```

```
# With item purification
difSIBTEST(verbal, group = 25, focal.name = 1, purify = TRUE)
r2 <- difSIBTEST(verbal, group = 25, focal.name = 1, purify = TRUE, nrIter = 5)

# With items 1 to 5 set as anchor items
difSIBTEST(verbal, group = "Gender", focal.name = 1, anchor = 1:5)
difSIBTEST(verbal, group = "Gender", focal.name = 1, anchor = 1:5, purify = TRUE)

# Saving the output into the "SIBresults.txt" file (and default path)
r <- difSIBTEST(verbal, group = 25, focal.name = 1, save.output = TRUE,
           output = c("SIBresults","default"))

# Graphical devices
plot(r)
plot(r2)

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

# Changing the path, JPEG figure
path <- "c:/Program Files/"
plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

difStd *Standardization DIF method*

Description

Performs DIF detection using standardization method.

Usage

```
difStd(Data, group, focal.name, anchor = NULL, match = "score",
   stdWeight = "focal", thrSTD = 0.1, purify = FALSE, nrIter = 10,
   save.output = FALSE, output = c("out", "default"))
## S3 method for class 'PDIF'
print(x, ...)
## S3 method for class 'PDIF'
plot(x, pch = 8, number = TRUE, col = "red", save.plot = FALSE,
   save.options = c("plot", "default", "pdf"), ...)
```

Arguments

`Data` numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.

`group`	numeric or character: either the vector of group membership or the column indicator (within `data`) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level of `group` which corresponds to the focal group.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `Data`. See **Details**.
`stdWeight`	character: the type of weights used for the standardized P-DIF statistic. Possible values are `"focal"` (default), `"reference"` and `"total"`. See **Details**.
`thrSTD`	numeric: the threshold (cut-score) for standardized P-DIF statistic (default is 0.10).
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is `FALSE`).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.
`x`	the result from a `PDIF` class object.
`pch, col`	type of usual `pch` and `col` graphical options.
`number`	logical: should the item number identification be printed (default is `TRUE`).
`save.plot`	logical: should the plot be saved into a separate file? (default is `FALSE`).
`save.options`	character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.
`...`	other generic parameters for the `plot` or the `print` functions.

Details

The method of standardization (Dorans and Kulick, 1986) allows for detecting uniform differential item functioning without requiring an item response model approach.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded from sum-score computation.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument `focal.name`.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `stdPDIF` function. This is specified by the `match` argument. By default, it takes the value `"score"` and the test score (i.e. raw score) is computed. The second option is to assign to `match` a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the `Data` matrix.

The threshold (or cut-score) for classifying items as DIF has to be set by the user by the argument `thrSTD`. Default value is 0.10 but Dorans (1989) also recommends value 0.05. For this reason it is not possible to provide asymptotic *p*-values.

The weights for computing the standardized P-DIF statistics are defined through the argument `stdWeight`, with possible values `"focal"` (default value), `"reference"` and `"total"`. See `stdPDIF` for further details.

In addition, two types of effect sizes are displayed. The first one is obtained from the standardized P-DIF statistic itself. According to Dorans, Schmitt and Bleistein (1992), the effect size of an item is classified as negligible if $|St-P-DIF| \leq 0.05$, moderate if $0.05 \leq |St-P-DIF| \leq 0.10$, and large if if $|St-P-DIF| \geq 0.10$. The second one is based on the transformation to the ETS Delta Scale (Holland and Thayer, 1985) of the standardized 'alpha' values (Dorans, 1989; Holland, 1985). The values of the effect sizes, together with the Dorans, Schmitt and Bleistein (DSB) and the ETS Delta scale (ETS) classification, are printed with the output.

Item purification can be performed by setting `purify` to `TRUE`. Purification works as follows: if at least one item was detected as functioning differently at some step of the process, then the data set of the next step consists in all items that are currently anchor (DIF free) items, plus the tested item (if necessary). The process stops when either two successive applications of the method yield the same classifications of the items (Clauser and Mazor, 1998), or when `nrIter` iterations are run without obtaining two successive identical classifications. In the latter case a warning message is printed.

A pre-specified set of anchor items can be provided through the `anchor` argument. It must be a vector of either item names (which must match exactly the column names of `Data` argument) or integer values (specifying the column numbers for item identification). In case anchor items are provided, they are used to compute the test score (matching criterion), including also the tested item. None of the anchor items are tested for DIF: the output separates anchor items and tested items and DIF results are returned only for the latter. Note also that item purification is not activated when anchor items are provided (even if `purify` is set to `TRUE`). By default it is `NULL` so that no anchor item is specified.

The output of the `difStd`, as displayed by the `print.PDIF` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

The `plot.PDIF` function displays the DIF statistics in a plot, with each item on the X axis. The type of point and the color are fixed by the usual `pch` and `col` arguments. Option `number` permits to display the item numbers instead. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "PDIF" with the following arguments:

- `PDIF` the values of the standardized P-DIF statistics.
- `stdAlpha` the values of the standardized alpha values (for effect sizes computation).
- `alpha` the value of `alpha` argument.
- `thr` the value of the `thrSTD` argument.
- `DIFitems` either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
- `match` a character string, either `"score"` or `"matching variable"` depending on the `match` argument.
- `purification` the value of `purify` option.
- `nrPur` the number of iterations in the item purification process. Returned only if `purify` is `TRUE`.
- `difPur` a binary matrix with one row per iteration in the item purification process and one column per item. Zeros and ones in the *i*-th row refer to items which were classified respectively as non-DIF and DIF items at the (*i*-1)-th step. The first row corresponds to the initial classification of the items. Returned only if `purify` is `TRUE`.
- `convergence` logical indicating whether the iterative item purification process stopped before the maximal number `nrIter` of allowed iterations. Returned only if `purify` is `TRUE`.
- `names` the names of the items.
- `anchor.names` the value of the `anchor` argument.
- `stdWeight` the value of the `stdWeight` argument.
- `save.output` the value of the `save.output` argument.
- `output` the value of the `output` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Clauser, B.E. and Mazor, K.M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning. Standardization and the Mantel-Haenszel method. *Applied Measurement in Education, 2*, 217-233. doi:10.1207/s15324818ame0203_3

Dorans, N. J. and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368. doi:10.1111/j.17453984.1986.tb00255.x

Dorans, N. J., Schmitt, A. P. and Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement, 29*, 309-319. doi:10.1111/j.17453984.1992.tb00379.x

Holland, P. W. (1985, October). *On the study of differential item performance without IRT*. Paper presented at the meeting of Military Testing Association, San Diego (CA).

Holland, P. W. and Thayer, D. T. (1985). An alternative definition of the ETS delta scale of item difficulty. *Research Report RR-85-43*. Princeton, NJ: Educational Testing Service.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

stdPDIF, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 verbal<-verbal[colnames(verbal) != "Anger"]

 # Three equivalent settings of the data matrix and the group membership
 difStd(verbal, group = 25, focal.name = 1)
 difStd(verbal, group = "Gender", focal.name = 1)
 difStd(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # With other weights
 difStd(verbal, group = "Gender", focal.name = 1, stdWeight = "reference")
 difStd(verbal, group = "Gender", focal.name = 1, stdWeight = "total")

 # With item purification
 difStd(verbal, group = "Gender", focal.name = 1, purify = TRUE)
 difStd(verbal, group = "Gender", focal.name = 1, purify = TRUE, nrIter = 5)

 # With items 1 to 5 set as anchor items
```

```
difStd(verbal, group = "Gender", focal.name = 1, anchor = 1:5)
difStd(verbal, group = "Gender", focal.name = 1, anchor = 1:5, purify = TRUE)


# With detection threshold of 0.05
difStd(verbal, group = "Gender", focal.name = 1, thrSTD = 0.05)

# Saving the output into the "STDresults.txt" file (and default path)
r <- difStd(verbal, group = 25, focal.name = 1, save.output  =  TRUE,
           output = c("STDresults","default"))

# Graphical devices
plot(r)

# Plotting results and saving it in a PDF figure
plot(r, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

# Changing the path, JPEG figure
path <- "c:/Program Files/"
plot(r, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

`difTID` *Transformed Item Difficulties (TID) DIF method*

Description

Performs DIF detection using Transformed Item Difficulties (TID) method.

Usage

```
difTID(Data, group, focal.name, thrTID = 1.5, purify = FALSE, purType = "IPP1",
   nrIter = 10, alpha = 0.05, extreme = "constraint",
   const.range = c(0.001, 0.999), nrAdd = 1, save.output = FALSE,
   output = c("out", "default"))
## S3 method for class 'TID'
print(x, only.final = TRUE, ...)
## S3 method for class 'TID'
plot(x, plot = "dist",pch = 2, pch.mult = 17, axis.draw = TRUE,
   thr.draw = FALSE, dif.draw = c(1, 3), print.corr = FALSE, xlim = NULL,
   ylim = NULL, xlab = NULL, ylab = NULL, main = NULL, col = "red",
   number = TRUE, save.plot = FALSE, save.options = c("plot",
   "default", "pdf"), ...)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within Data) of group membership. See **Details**.
focal.name	numeric or character indicating the level of group which corresponds to the focal group.
thrTID	either the threshold for detecting DIF items (default is 1.5) or "norm".
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
purType	character: the type of purification process to be run. Possible values are "IPP1" (default), "IPP2" and "IPP3". Ignored if purify is FALSE. See **Details**.
nrIter	numeric: the maximal number of iterations in the item purification process (default is 10).
alpha	numeric: the significance level for calculating the detection threshold (default is 0.05). Ignored if thrTID is numeric.
extreme	character: the method used to modify the extreme proportions. Possible values are "constraint" (default) or "add". See **Details**.
const.range	numeric: a vector of two constraining proportions. Default values are 0.001 and 0.999. Ignored if extreme is "add".
nrAdd	integer: the number of successes and the number of failures to add to the data in order to adjust the proportions. Default value is 1. Ignored if extreme is "constraint".
save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.
x	the result from a TID class object.
only.final	logical: should only the first and last steps of the purification process be printed? (default is TRUE. If FALSE all perpendicular distances, parameters of the major axis, and detection thresholds are printed additionally. Ignored if purify is FALSE.
plot	character: either "dist" (default) to display the perpendicular distances, or "delta" for the Delta plot. See **Details**.
pch	integer: the usual point character type for point display. Default value is 2, that is, Delta points are drawn as empty triangles.
pch.mult	integer: the type of point to be used for superposing onto Delta points that correspond to several items. Default value is 17, that is, full black traingles are drawn onto existing Delta plots wherein multiple items are located.
axis.draw	logical: should the major axis be drawn? (default is TRUE). If so, it will be drawn as a solid line.
thr.draw	logical: should the upper and lower bounds for DIF detection be drawn? (default is FALSE). If TRUE, they will be drawn as dashed lines.

`dif.draw` numeric: a vector of two integer values to specify how the DIF items should be displayed. The first component of `dif.draw` is the type of point (i.e. the usual `pch` argument) and the second component determines the point size (i.e. the usual `cex` argument). Default values are 1 and 3, meaning that empty circles of three times the usual size are drawn around the Delta points of items flagged as DIF.

`print.corr` logical: should the sample correlation of Delta scores be printed? (default is `FALSE`). If `TRUE`, it is printed in upper-left corner of the plot.

`xlim, ylim, xlab, ylab, main`
either the usual plot arguments `xlim`, `ylim`, `xlab`, `ylab` and `main`, or `NULL` (default value for all arguments). If `NULL`, the X and Y axis limits are computed from the range of Delta scores, the X and Y axis labels are `"Reference group"` and `"Focal group"` respectively, and no main title is produced.

`col` character: the color type for the items. Used only when `plot` is `"dist"`.

`number` logical: should the item number identification be printed (default is `TRUE`).

`save.plot` logical: should the plot be saved into a separate file? (default is `FALSE`).

`save.options` character: a vector of three components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value), and the third component is the file extension, either `"pdf"` (default) or `"jpeg"`. See **Details**.

`...` other generic parameters for the `plot` or the `print` functions.

Details

The Transformed Item Difficulties (TID) method, also known as Angoff's Delta method (Angoff, 1982; Angoff and Ford, 1973) allows for detecting uniform differential item functioning without requiring an item response model approach. The presnt implementation relies on the `deltaPlot` and `diagPlot` functions from package**deltaPlotR** (Magis and Facon, 2014).

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded from the computation of proportions of success.

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the value of the argument `focal.name`.

The threshold for flaging items as DIF can be of two types and is specified by the `thr` argument.

1. It can be fixed to some arbitrary positive value by the user, for instance 1.5 (Angoff and Ford, 1973). In this case, `thr` takes the required numeric threshold value.
2. Alternatively, it can be derived from the bivariate normal approximation of the Delta points (Magis and Facon, 2012). In this case, `thr` must be given the character value `"norm"` (which is the default value). This threshold equals

$$\Phi^{-1}(1-\alpha/2)\,\sqrt{\frac{b^2\,{s_0}^2-2\,b\,s_{01}+{s_1}^2}{b^2+1}}$$

where Φ is the density of the standard normal distribution, α is the significance level (set by the argument `alpha` with default value 0.05), b is the slope parameter of the major axis, s_0 and s_1 are the sample standard deviations of the Delta scores in the reference group and the focal group, respecively, and s_{01} is the sample covariance of the Delta scores (see Magis and Facon, 2012, for further details).

Item purification can be performed by setting the argument `purify` to `TRUE` (by default it is `FALSE` so no purification is performed). The item purification process (IPP) starts when at least one item was flagged as DIF after the first run of the Delta plot, and proceeds as follows.

1. The intercept and slope parameters of the major axis are re-calculated by removing all DIF that are currently flagged as DIF. This yields updated values a^*, b^*, s_0^*, s_1^* and s_{01}^* of the intercept and slope parameters, sample stanbdard deviations and sample covariance of the Delta scores.
2. Perpendicular distances (for all items) are updated with respect to the updated major axis.
3. Detection threshold is also updated. Three possible updates are possible: see below.
4. All items are now tested for the presence of DIF, given the updated perpendicular distances and major axis.
5. If the set of items flagged as DIF is the same as the one from the previous loop, stop the process. Otherwise go back to step 1.

Unlike traditional DIF methods, the detection threshold may also be updated since it depends on the sample estimates (when the normal approximation is considered). Three approaches are currently implemented and are specified by the `purType` argument.

1. Method 1 (`purType=="IPP1"`): the same threshold is used throughout the purification process, it is not iteratively updated. The threshold is the one obtained after the first run of the Delta plot.
2. Method 2 (`purType=="IPP2"`): only the slope parameter is updated in the threshold formula. By this way, one keeps the full data structure (i.e. neither the sample variances nor the sample covariance of the Delta scores are modified) but only the slope parameter is adjusted to lessen the impact of DIF items.
3. Method 3 (`purType=="IPP3"`): all adjusted parameters are plugged in the threshold formula. This approach completely discards the effect of items flagged as DIF from the computation of the threshold.

See Magis and Facon (2013) for further details. Note that purification can also be performed with fixed threshold (i.e. specified by the user), but then only IPP1 process is performed.

In order to avoid possible infinite loops in the purification process, a maximal number of iterations must be specified through the argument `maxIter`. The default maximal number of iterations is 10.

The output contains all input information, the Delta scores and perpendicular distances, the parameter of the major axis and the items flagged as DIF (if none, a character sentence is returned). In addition, the detection threshold and the type of threshold (fixed or normal approximation) is provided.

If item purification was run, several additional elements are returned: the number of iterations, a logical indicator whether the convergence was reached (or not, meaning that the process stopped because of reaching the maximal number of allowed iterations), a matrix with indicators of which items were flagged as DIF at each iteration, and the type of item purification process. Moreover, perpendicular distances are returned in a matrix format (one column per iteration), as well as successive major axis parameters (one row per iteration) and successive thresholds (as a vector).

The output is managed and printed in a more user-friendly way. When item purification is performed, only the first and last steps are displayed. Specifying the argument `only.final` to `FALSE` prints in addition all intermediate steps of the process (successive perpendicular distances, parameters of the major axis, and detection thresholds).

The output of the `difTID`, as displayed by the `print.TID` function, can be stored in a text file provided that `save.output` is set to `TRUE` (the default value `FALSE` does not execute the storage). In this case, the name of the text file must be given as a character string into the first component of the `output` argument (default name is `"out"`), and the path for saving the text file can be given through the second component of `output`. The default value is `"default"`, meaning that the file will be saved in the current working directory. Any other path can be specified as a character string: see the **Examples** section for an illustration.

Two types of plots are available through the `plot.TID` function. If the argument `plot` is set to `"dist"` (the default value), then the perpendicular distances are represented on the Y axis of a scatter plot, with each item on the X axis. If `plot` is set to `"delta"`, the Delta plot is returned. In the latter, all particular options can be found from the `diagPlot` function. Also, the plot can be stored in a figure file, either in PDF or JPEG format. Fixing `save.plot` to `TRUE` allows this process. The figure is defined through the components of `save.options`. The first two components perform similarly as those of the `output` argument. The third component is the figure format, with allowed values `"pdf"` (default) for PDF file and `"jpeg"` for JPEG file.

Value

A list of class "TID" with the following arguments:

`Props` the matrix of proportions of correct responses, or `NA` if `type` is `"delta"`.

`adjProps` the restricted proportions, in the same format as the output `Props` matrix, or `NA` if `type` is `"delta"`.

`Deltas` the matrix of Delta scores.

`Dist` a matrix with perpendicular distances, one row per item and one column per run of the Delta plot. If `purify` is `FALSE`, only a single column is returned.

`axis.par` a matrix with two columns, holding respectively the intercepts and the slope parameters of the major axis. Each row refers to one step of the purification process. If `purify` is `FALSE`, only a single row is returned.

`nrIter` the number of iterations invloved in the purification process. Returned only if `purify` is `TRUE`.

`maxIter` the value of the `maxIter` argument. Returned only if `purify` is `TRUE`.

`convergence` a logical value indicating whether convergence was reached in the purification process. Returned only if `purify` is `TRUE`.

`difPur` a matrix with one column per item and one row per iteration in the purification process, holding zeros and ones to indicate which items were flagged as DIF or not at each step of the process. Returned only if `purify` is `TRUE`.

`thr` a vector of successive threshold values used during the purification process. If `purify` is `FALSE`, a single value is returned.

`rule`	a character value indicating whether the threshold was `"fixed"` by the user (i.e. by setting `thr` to a numeric value) or whether it was computed by normal approximation (i.e. by setting `thr` to `"norm"`).
`purType`	the value of the `purType` argument. Returned only if `purify` is `TRUE`.
`DIFitems`	either `"No DIF item detected"` or an integer vector with the items that were flagged as DIF.
`adjust.extreme`	the value of the `extreme` argument.
`const.range`	the value of the `const.range` argument.
`nrAdd`	the value of the `nrAdd` argument.
`purify`	the value of the `purify` argument.
`alpha`	the value of the `alpha` argument.
`save.output`	the value of the `save.output` argument.
`output`	the value of the `output` argument.
`names`	either the names of the items (defined by the column names of the `Data` matrix) or the series of integers from one to the number of items.
`number`	a boolean value, being `TRUE` if the item names are simply their number in the `Data` matrix, or `FALSE` if real item names are available in the `names` element.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium

References

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berck (Ed.), *Handbook of methods for detecting item bias* (pp. 96-116). Baltimore, MD: Johns Hopkins University Press.

Angoff, W. H., and Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 2*, 95-106. doi:10.1111/j.17453984.1973.tb00787.x

Magis, D., and Facon, B. (2012). Angoff's Delta method revisited: improving the DIF detection under small samples. *British Journal of Mathematical and Statistical Psychology, 65*, 302-321. doi:10.1111/j.20448317.2011.02025.x

Magis, D., and Facon, B. (2013). Item purification does not always improve DIF detection: a counter-example with Angoff's Delta plot. *Educational and Psychological Measurement, 73*, 293-311. doi:10.1177/0013164412451903

Magis, D. and Facon, B. (2014). *deltaPlotR*: An R Package for Differential Item Functioning Analysis with Angoff's Delta Plot. *Journal of Statistical Software, Code Snippets, 59(1)*, 1-19. doi:10.18637/jss.v059.c01

See Also

`deltaPlot`, `diagPlot`, `dichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal) != "Anger"]

 # Three equivalent settings of the data matrix and the group membership
 r <- difTID(verbal, group = 25, focal.name = 1)
 difTID(verbal, group = "Gender", focal.name = 1)
 difTID(verbal[,1:24], group = verbal[,25], focal.name = 1)

 # With item purification and threshold 1
 r2 <- difTID(verbal, group = "Gender", focal.name = 1, purify = TRUE, thrTID = 1)

 # Saving the output into the "TIDresults.txt" file (and default path)
 difTID(verbal, group = 25, focal.name = 1, save.output = TRUE,
   output = c("TIDresults", "default"))

 # Graphical devices
 plot(r2)
 plot(r2, plot = "delta")

 # Plotting results and saving it in a PDF figure
 plot(r2, save.plot = TRUE, save.options = c("plot", "default", "pdf"))

 # Changing the path, JPEG figure
 path <- "c:/Program Files/"
 plot(r2, save.plot = TRUE, save.options = c("plot", path, "jpeg"))

## End(Not run)
```

`genDichoDif` *Comparison of DIF detection methods among multiple groups*

Description

This function compares the specified DIF detection methods among multiple groups, with respect to the detected items.

Usage

```
genDichoDif(Data, group, focal.names, method, anchor = NULL, match = "score",
  type = "both", criterion = "LRT", alpha = 0.05, model = "2PL", c = NULL,
  engine = "ltm", discr = 1, irtParam = NULL, nrFocal = 2, same.scale = TRUE,
  purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
```

```
  output = c("out", "default"))
## S3 method for class 'genDichoDif'
print(x, ...)
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
`focal.names`	numeric or character vector indicating the levels of `group` which correspond to the focal groups.
`method`	character: the name of the selected methods. See **Details**.
`anchor`	either `NULL` (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `Data`. See **Details**.
`type`	a character string specifying which DIF effects must be tested (default is `"both"`). See **Details**.
`criterion`	character: the type of test statistic used to detect DIF items with generalized logistic regression. Possible values are `"LRT"` (default) and `"Wald"`. See **Details**.
`alpha`	numeric: significance level (default is 0.05).
`model`	character: the IRT model to be fitted (either `"1PL"`, `"2PL"` or `"3PL"`). Default is `"2PL"`.
`c`	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
`engine`	character: the engine for estimating the 1PL model, either `"ltm"` (default) or `"lme4"`.
`discr`	either `NULL` or a real positive value for the common discrimination parameter (default is 1). Used onlky if `model` is `"1PL"` and `engine` is `"ltm"`. See **Details**.
`irtParam`	matrix with $2J$ rows (where J is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
`nrFocal`	numeric: the number of focal groups (default is 2).
`same.scale`	logical: are the item parameters of the `irtParam` matrix on the same scale? (default is "TRUE"). See **Details**.
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.

save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.
x	result from a genDichoDif class object.
...	other generic parameters for the print function.

Details

genDichoDif is a generic function which calls one or several DIF detection methods among multiple groups, and summarize their output. The possible methods are: "GMH" for Generalized Mantel-Haenszel (Penfield, 2001), "genLogistic" for generalized logistic regression (Magis, Raiche Beland and Gerard, 2011) and "genLord" for generalized Lord's chi-square test (Kim, Cohen and Park, 1995).

If method has a single component, the output of genDichoDif is exactly the one provided by the method itself. Otherwise, the main output is a matrix with one row per item and one column per method. For each specified method and related arguments, items detected as DIF and non-DIF are respectively encoded as "DIF" and "NoDIF". When printing the output an additional column is added, counting the number of times each item was detected as functioning differently (Note: this is just an informative summary, since the methods are obviously not independent for the detection of DIF items).

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from either the computation of the sum-scores, the fitting of the logistic models or the IRT models (according to the method).

The vector of group membership must hold at least three different values, either as numeric or character. The focal groups are defined by the values of the argument focal.names.

For generalized Mantel-Haenszel and generalized logistic methods, the matching criterion can be either the test score or any other continuous or discrete variable to be passed in the DIF function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

For the generalized logistic regression method, the argument type permits to test either both uniform and nonuniform effects simultaneously (with type="both"), only uniform DIF effect (with type="udif") or only nonuniform DIF effect (with type="nudif"). Furthermore, the argument criterion defines which test must be used, either the Wald test ("Wald") or the likelihood ratio test ("LRT"). See difGenLord for further details.

For generalized Lord method, one can specify either the IRT model to be fitted (by means of model, c, engine and discr arguments), or the item parameter estimates with arguments irtParam and same.scale. See difGenLord for further details.

The threshold for detecting DIF items depends on the method and is depending on the significance level set by alpha.

Item purification can be requested by specifying `purify` option to `TRUE`. Recall that item purification process is slightly different for IRT and for non-IRT based methods. See the corresponding methods for further information.

Adjustment for multiple comparisons is possible with the argument `p.adjust.method`. See the corresponding methods for further information.

A pre-specified set of anchor items can be provided through the `anchor` argument. For non-IRT methods, anchor items are used to compute the test score (as matching criterion). For IRT methods, anchor items are used to rescale the item parameters on a common metric. See the corresponding methods for further information.

The output of the `genDichoDif` function can be stored in a text file by fixing `save.output` and `output` appropriately. See the help file of `selectGenDif` function (or any other DIF method) for further information.

Value

Either the output of one of the DIF detection methods, or a list of class "genDichoDif" with the following arguments:

`DIF`	a character matrix with one row per item and whose columns refer to the different specified detection methods. See **Details**.
`alpha`	the significance level `alpha`.
`method`	the value of `method`argument.
`match`	the value of `match` argument.
`type`	the value of `type` argument.
`criterion`	the value of the `criterion` argument.
`model`	the value of `model` option.
`c`	the value of `c` option.
`engine`	The value of the `engine` argument.
`discr`	the value of the `discr` argument.
`irtParam`	the value of `irtParam` option.
`same.scale`	the value of `same.scale` option.
`p.adjust.method`	the value of the `p.adjust.method` argument.
`purification`	the value of `purify` option.
`nrPur`	an integer vector (of length equal to the number of methods) with the number of iterations in the purification process. Returned only if `purify` is TRUE.
`convergence`	a logical vector (of length equal to the number of methods) indicating whether the iterative purification process converged. Returned only if `purify` is TRUE.
`anchor.names`	the value of the `anchor` argument.
`save.output`	the value of the `save.output` argument.
`output`	the value of the `output` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education, 14*, 235-259. doi:10.1207/S15324818AME1403_3

See Also

difGMH, difGenLogistic, difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and trait
 # anger score ("Low" or "High")
 group <- rep("WomanLow", nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)
```

```
 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Comparing the three available methods
 # with item purification
 genDichoDif(Verbal, group = 25, focal.names = names, method = c("GMH", "genLogistic",
             "genLord"), purify = TRUE)

 # Same analysis, but saving the output into the 'genDicho' file
 genDichoDif(Verbal, group = 25, focal.names = names, method = c("GMH", "genLogistic",
             "genLord"), purify = TRUE, save.output = TRUE,
             output = c("genDicho", "default"))

## End(Not run)
```

---

genLogistik *Generalized logistic regression DIF statistic*

Description

Calculates the "generalized logistic regression" likelihood-ratio or Wald statistics for DIF detection among multiple groups.

Usage

```
genLogistik(data, member, match = "score", anchor = 1:ncol(data),
  type = "both", criterion = "LRT")
```

Arguments

data	numeric: the data matrix (one row per subject, one column per item).
member	numeric: the vector of group membership with zero and positive integer entries only. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of data. See **Details**.
anchor	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "both" (default), "udif" and "nudif". See **Details**.
criterion	character: the type of test statistic used to detect DIF items. Possible values are "LRT" (default) and "Wald". See **Details**.

Details

This command computes the generalized logistic regression statistic (Magis, Raiche, Beland and Gerard, 2011) in the specific framework of differential item functioning among $(J+1)$ groups and J is the number of focal groups. It forms the basic command of `difGenLogistic` and is specifically designed for this call.

The three possible models to be fitted are:

$$M_0 : logit(\pi_i) = \alpha + \beta X + \gamma_i + \delta_i X$$

$$M_1 : logit(\pi_i) = \alpha + \beta X + \gamma_i$$

$$M_2 : logit(\pi_i) = \alpha + \beta X$$

where π_i is the probability of answering correctly the item in group *i* ($i = 0, ..., J$) and X is the matching criterion. Parameters α and β are the common intercept and the slope of the logistic curves, while γ_i and δ_i are group-specific parameters. For identification reasons the parameters γ_0 and δ_0 of the reference group are set to zero. The set of parameters $\{\gamma_i : i = 1, ..., J\}$ of the focal groups ($g = i$) represents the uniform DIF effect across all groups, and the set of parameters $\{\delta_i : i = 1, ..., n\}$ is used to model nonuniform DIF effect across all groups. The models are fitted with the `glm` function.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `Logistik` function. This is specified by the `match` argument. By default, it takes the value `"score"` and the test score (i.e. raw score) is computed. The second option is to assign to `match` a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the `data` matrix.

Two tests are available: the Wald test and the likelihood ratio test. With the likelihood ratio test, two nested models are fitted and compared by means of Wilks' Lambda (or likelihood ratio) statistic (Wilks, 1938). With the Wald test, the model parameters are statistically tested using an appropriate contrast matrix. Each test is set with the `criterion` argument, with the values `"LRT"` and `"Wald"` respectively.

The argument `type` determines the type of DIF effect to be tested. The three possible values of `type` are: `type="both"` which tests the hypothesis $H_0 : \gamma_i = \delta_i = 0$ for all *i*; `type="nudif"` which tests the hypothesis $H_0 : \delta_i = 0$ for all *i*; and `type="udif"` which tests the hypothesis $H_0 : \gamma_i = 0 | \delta_i = 0$ for all *i*. In other words, `type="both"` tests for DIF (without distinction between uniform and nonuniform effects), while `type="udif"` and `type="nudif"` test for uniform and nonuniform DIF, respectively. Whatever the tested DIF effects, this is a simultaneous test of the equality of focal group parameters to zero.

The data are passed through the `data` argument, with one row per subject and one column per item. Missing values are allowed but must be coded as `NA` values. They are discarded from the fitting of the logistic models (see `glm` for further details).

The vector of group membership, specified with `member` argument, must hold only zeros and positive integers. The value zero corresponds to the reference group, and each positive integer value corresponds to one focal group. At least two different positive integers must be supplied.

Option `anchor` sets the items which are considered as anchor items for computing the logistic regression DIF statistics. Items other than the anchor items and the tested item are discarded. `anchor` must hold integer values specifying the column numbers of the corresponding anchor items. It is mainly designed to perform item purification.

In addition to the results of the fitted models (model parameters, covariance matrices, test statistics), Nagelkerke's R^2 coefficients (Nagelkerke, 1991) are computed for each model and the output returns the differences in these coefficients. Such differences are used as measures of effect size by the `difGenLogistic` command; see Gomez-Benito, Dolores Hidalgo and Padilla (2009), Jodoin and Gierl (2001) and Zumbo and Thomas (1997).

Value

A list with nine components:

- `stat` — the values of the generalized logistic regression DIF statistics (that is, the likelihood ratio test statistics).
- `R2M0` — the values of Nagelkerke's R^2 coefficients for the "full" model.
- `R2M1` — the values of Nagelkerke's R^2 coefficients for the "simpler" model.
- `deltaR2` — the differences between Nagelkerke's R^2 coefficients of the tested models. See **Details**.
- `parM0` — a matrix with one row per item and $2 + J * 2$ columns (where J is the number of focal groups), holding successively the fitted parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}_i$ and $\hat{\delta}_i$ ($i = 1, ..., J$) of the "full" model (M_0 if `type="both"` or `type="nudif"`, M_1 if `type="udif"`).
- `parM1` — the same matrix as `parM0` but with fitted parameters for the "simpler" model (M_1 if `type="nudif"`, M_2 if `type="both"` or `type="udif"`).
- `covMat` — a 3-dimensional matrix of size *p* x *p* x *K*, where *p* is the number of estimated parameters and *K* is the number of items, holding the *p* x *p* covariance matrices of the estimated parameters (one matrix for each tested item).
- `criterion` — the value of the `criterion` argument.
- `match` — a character string, either `"score"` or `"matching variable"` depending on the `match` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Gomez-Benito, J., Dolores Hidalgo, M. and Padilla, J.-L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology, 5*, 18-25. doi:10.1027/1614-2241.5.1.18

Jodoin, M. G. and Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349. doi:10.1207/S15324818AME1404_2

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692. doi:10.1093/biomet/78.3.691

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60-62. doi:10.1214/aoms/1177732360

Zumbo, B. D. and Thomas, D. R. (1997). A measure of effect size for a model-based approach for studying DIF. Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.

See Also

`difGenLogistic`, `genDichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender (0 or 1) and trait anger score
 # ("Low" or "High")
 # Reference group: women with low trait anger score (<=20)
 group <- rep(0,nrow(verbal))
 group[Anger>20 & Gender==0] <- 1
 group[Anger<=20 & Gender==1] <- 2
 group[Anger>20 & Gender==1] <- 3

 # Testing both types of DIF simultaneously
 # With all items
 genLogistik(verbal[,1:24], group)
 genLogistik(verbal[,1:24], group, criterion = "Wald")

 # Removing item 6 from the set of anchor items
 genLogistik(verbal[,1:24], group, anchor = c(1:5, 7:24))
 genLogistik(verbal[,1:24], group, anchor = c(1:5, 7:24), criterion = "Wald")

 # Testing nonuniform DIF effect
 genLogistik(verbal[,1:24], group, type = "nudif")
 genLogistik(verbal[,1:24], group, type = "nudif", criterion="Wald")
```

```
# Testing uniform DIF effect
genLogistik(verbal[,1:24], group, type = "udif")
genLogistik(verbal[,1:24], group, type = "udif", criterion="Wald")

# Using trait anger score as matching criterion
genLogistik(verbal[,1:24], group, match = verbal[,25])

## End(Not run)
```

---

genLordChi2 *Generalized Lord's chi-squared DIF statistic*

Description

Calculates the generalized Lord's chi-squared statistics for DIF detection among multiple groups.

Usage

```
genLordChi2(irtParam, nrFocal)
```

Arguments

irtParam	numeric: the matrix of item parameter estimates. See **Details**.
nrFocal	numeric: the number of focal groups.

Details

This command computes the generalized Lord's chi-squared statistic (Kim, Cohen and Park, 1995), also called the *Qj* statistics, in the specific framework of differential item functioning with multiple groups. It forms the basic command of `difGenLord` and is specifically designed for this call.

The `irtParam` matrix has a number of rows equal to the number of groups (reference and focal ones) times the number of items *J*. The first *J* rows refer to the item parameter estimates in the reference group, while the next sets of *J* rows correspond to the same items in each of the focal groups. The number of columns depends on the selected IRT model: 2 for the 1PL model, 5 for the 2PL model, 6 for the constrained 3PL model and 9 for the unconstrained 3PL model. The columns of `irtParam` have to follow the same structure as the output of `itemParEst` command (the latter can actually be used to create the `irtParam` matrix).

In addition, the item parameters of the reference group and the focal groups must be placed on the same scale. This can be done by using `itemRescale` command, which performs equal means anchoring between two groups of item estimates (Cook and Eignor, 1991).

The number of focal groups has to be specified with argument `nrFocal`.

Value

A vector with the values of the generalized Lord's chi-squared DIF statistics.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

itemParEst, itemRescale, difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and
 # trait anger score ("Low" or "High")
 group <- rep("WomanLow",nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # Splitting the data into the four subsets according to "group"
 data0 <- data1 <- data2 <- data3 <- NULL
 for (i in 1:nrow(verbal)){
 if (group[i]=="WomanLow") data0 <- rbind(data0, as.numeric(verbal[i,1:24]))
 if (group[i]=="WomanHigh") data1 <- rbind(data1, as.numeric(verbal[i,1:24]))
 if (group[i]=="ManLow") data2 <- rbind(data2, as.numeric(verbal[i,1:24]))
 if (group[i]=="ManHigh") data3 <- rbind(data3, as.numeric(verbal[i,1:24]))
 }
```

```
# Estimation of the item parameters (1PL model)
m0.1PL <- itemParEst(data0, model = "1PL")
m1.1PL <- itemParEst(data1, model = "1PL")
m2.1PL <- itemParEst(data2, model = "1PL")
m3.1PL <- itemParEst(data3, model = "1PL")

# merging the item parameters with rescaling
irt.scale <- rbind(m0.1PL, itemRescale(m0.1PL, m1.1PL), itemRescale(m0.1PL, m2.1PL),
                   itemRescale(m0.1PL, m3.1PL))

# Generalized Lord's chi-squared statistics
genLordChi2(irt.scale, nrFocal = 3)

## End(Not run)
```

`genMantelHaenszel` *Generalized Mantel-Haenszel DIF statistic*

Description

Calculates the generalized Mantel-Haenszel statistics for DIF detection among multiple groups.

Usage

```
genMantelHaenszel(data, member, match = "score", anchor = 1:ncol(data))
```

Arguments

`data`	numeric: the data matrix (one row per subject, one column per item).
`member`	numeric: the vector of group membership with zero and positive integer entries only. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of `data`. See **Details**.
`anchor`	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.

Details

This command computes the generalized Mantel-Haenszel statistic (Somes, 1986) in the specific framework of differential item functioning. It forms the basic command of `difGMH` and is specifically designed for this call.

The data are passed through the `data` argument, with one row per subject and one column per item. Missing values are allowed but must be coded as `NA` values. They are discarded from sum-score computation.

The vector of group membership, specified with member argument, must hold only zeros and positive integers. The value zero corresponds to the reference group, and each positive integer value corresponds to one focal group. At least two different positive integers must be supplied.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the genMantelHaenszel function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the data matrix.

Option anchor sets the items which are considered as anchor items for computing generalized Mantel-Haenszel statistics. Items other than the anchor items and the tested item are discarded. anchor must hold integer values specifying the column numbers of the corresponding anchor items. It is primarily designed to perform item purification.

Value

A vector with the values of the generalized Mantel-Haenszel DIF statistics.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education, 14*, 235-259. doi:10.1207/S15324818AME1403_3

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician, 40*, 106-108. doi:10.2307/2684866

See Also

difGMH

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender (0 or 1) and trait anger
 # score ("Low" or "High")
 # Reference group: women with low trait anger score (<=20)
 group <- rep(0, nrow(verbal))
 group[Anger>20 & Gender==0] <- 1
 group[Anger<=20 & Gender==1] <- 2
 group[Anger>20 & Gender==1] <- 3

 # Without continuity correction
 genMantelHaenszel(verbal[,1:24], group)

 # Removing item 6 from the set of anchor items
 genMantelHaenszel(verbal[,1:24], group, anchor = c(1:5, 7:24))

## End(Not run)
```

---

itemPar1PL *Item parameter estimation for DIF detection using Rasch (1PL) model*

Description

Fits the Rasch (1PL) model and returns related item parameter estimates.

Usage

```
itemPar1PL(data, engine = "ltm", discr = 1)
```

Arguments

data	numeric: the data matrix.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
discr	either NULL or a real positive value for the common discrimination parameter (default is 1). Not used if engine is "lme4". See **Details**.

Details

`itemPar1PL` permits to get item parameter estimates from the Rasch or 1PL model. The output is ordered such that it can be directly used with the general `itemParEst` command, as well as the methods of Lord (`difLord`) and Raju (`difRaju`) and Generalized Lord's (`difGenLord`) to detect differential item functioning.

The `data` is a matrix whose rows correspond to the subjects and columns to the items.

Missing values are allowed but must be coded as `NA` values. They are discarded for item parameter estimation.

The estimation engine is set by the `engine` argument. By default (`engine="ltm"`), the Rasch model is fitted using marginal maximum likelihood, by means of the function `rasch` from the `ltm` package (Rizopoulos, 2006). The other option, `engine="lme4"`, permits to fit the Rasch model as a generalized linear mixed model, by means of the `glmer` function of the `lme4` package (Bates and Maechler, 2009).

With the `"ltm"` engine, the common discrimination parameter is set equal to 1 by default. It is possible to fix another value through the argument`discr`. Alternatively, this common discrimination parameter can be estimated (though not returned) by fixing `discr` to `NULL`. See the functionalities of `rasch` command for further details.

Value

A matrix with one row per item and two columns, the first one with item parameter estimates and the second one with the related standard errors.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

itemPar2PL, itemPar3PL, itemPar3PLconst, itemParEst, difLord, difRaju, difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Getting item parameter estimates ('ltm' engine)
 itemPar1PL(verbal[, 1:24])

 # Estimating the common discrimination parameter instead
 itemPar1PL(verbal[, 1:24], discr = NULL)

 # Getting item parameter estimates ('lme4' engine)
 itemPar1PL(verbal[, 1:24], engine = "lme4")

## End(Not run)
```

---

itemPar2PL — *Item parameter estimation for DIF detection using 2PL model*

Description

Fits the 2PL model and returns related item parameter estimates, standard errors and covariances between item parameters.

Usage

```
itemPar2PL(data)
```

Arguments

data	numeric: the data matrix.

Details

itemPar2PL permits to get item parameter estimates from the 2PL model. The output is ordered such that it can be directly used with the general itemParEst command, as well as the methods of Lord (difLord) and Raju (difRaju) and Generalized Lord's (difGenLord) to detect differential item functioning.

The data is a matrix whose rows correspond to the subjects and columns to the items.

Missing values are allowed but must be coded as NA values. They are discarded for item parameter estimation.

The 2PL model is fitted using marginal maximum likelihood by means of the functions from the ltm package (Rizopoulos, 2006).

Value

A matrix with one row per item and five columns: the estimates of item discrimination *a* and difficulty *b* parameters, the related standard errors *se(a)* and *se(b)*, and the covariances *cov(a,b)*, in this order.

Note

The 2PL model is fitted under the linear parametrization in ltm, the covariance matrix is extracted with the vcov() function, and final standard errors and covariances are derived by the Delta method. See Rizopoulos (2006) for further details, and the Note.pdf document in the difR package for mathematical details.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

itemPar1PL, itemPar3PL, itemPar3PLconst, itemParEst, difLord, difRaju,

difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Getting item parameter estimates
 itemPar2PL(verbal[,1:24])

## End(Not run)
```

`itemPar3PL` *Item parameter estimation for DIF detection using 3PL model*

Description

Fits the 3PL model and returns related item parameter estimates.

Usage

```
itemPar3PL(data)
```

Arguments

`data`	numeric: the data matrix.

Details

`itemPar3PL` permits to get item parameter estimates from the 3PL model. The output is ordered such that it can be directly used with the general `itemParEst` command, as well as the methods of Lord (`difLord`) and Raju (`difRaju`) and Generalized Lord's (`difGenLord`) to detect differential item functioning.

The output consists of nine columns which are displayed in the following order. The first three columns hold the estimates of item discrimination *a*, difficulty *b* and pseudo-guessing *c* parameters. In the next three columns one can find the related standard errors *se(a)*, *se(b)* and *se(c)*. Eventually, the last three columns contain the covariances between item parameters, respectively *cov(a,b)*, *cov(a,c)* and *cov(b,c)*.

The `data` is a matrix whose rows correspond to the subjects and columns to the items.

Missing values are allowed but must be coded as `NA` values. They are discarded for item parameter estimation.

The 3PL model is fitted using marginal maximum likelihood by means of the functions from the `ltm` package (Rizopoulos, 2006).

Value

A matrix with one row per item and nine columns. See **Details**.

Note

The 3PL model is fitted under the linear parametrization in `tpm`, the covariance matrix is extracted with the `vcov()` function, and final standard errors and covariances are derived by the Delta method. See Rizopoulos (2006) for further details, and the `Note.pdf` document in the `difR` package for mathematical details.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

`itemPar1PL`, `itemPar2PL`, `itemPar3PLconst`, `itemParEst`, `difLord`, `difRaju`, `difGenLord`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Getting item parameter estimates
 itemPar3PL(verbal[,1:24])

## End(Not run)
```

itemPar3PLconst — *Item parameter estimation for DIF detection using constrained 3PL model*

Description

Fits the 3PL model with constrained pseudo-guessing values and returns related item parameter estimates.

Usage

```
itemPar3PLconst(data, c=rep(0,ncol(data)))
```

Arguments

`data`	numeric: the data matrix.
`c`	numeric value or vector of constrained pseudo-guessing parameters. See **Details**.

Details

`itemPar3PLconst` permits to get item parameter estimates from the 3PL model for which the pseudo-guessing parameters are constrained to some fixed values. The output is ordered such that it can be directly used with the general `itemParEst` command, as well as the methods of Lord (`difLord`) and Raju (`difRaju`) and Generalized Lord's (`difGenLord`) to detect differential item functioning.

The output is similar to that of `itemPar2PL` method to fit the 2PL model; an additional column is included and holds the fixed pseudo-guessing parameter values.

The `data` is a matrix whose rows correspond to the subjects and columns to the items.

Missing values are allowed but must be coded as `NA` values. They are discarded for item parameter estimation.

The argument `c` can be either a single numeric value or a numeric vector of the same length of the number of items. In the former case, the pseudo-guessing parameters are considered to be all identical to the given `c` value; otherwise `c` is directly used to constraint these parameters.

The constrained 3PL model is fitted using marginal maximum likelihood by means of the functions from the `ltm` package (Rizopoulos, 2006).

Value

A matrix with one row per item and six columns: the item discrimination *a* and difficulty estimates *b*, the corresponding standard errors *se(a)* and *se(b)*, the covariances *cov(a,b)* and the constrained pseudo-guessing values *c*.

Note

The constrained 3PL model is fitted under the linear parametrization in `tpm`, the covariance matrix is extracted with the `vcov()` function, and final standard errors and covariances are derived by the Delta method. See Rizopoulos (2006) for further details, and the `Note.pdf` document in the `difR` package for mathematical details.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
`<sebastien.beland@umontreal.ca>`
Gilles Raiche
Universite du Quebec a Montreal
`<raiche.gilles@uqam.ca>`

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

`itemPar1PL`, `itemPar2PL`, `itemPar3PL`, `itemParEst`, `difLord`, `difRaju`,
`difGenLord`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Constraining all pseudo-guessing parameters to be equal to 0.05
 itemPar3PLconst(verbal[,1:24], c = 0.05)

 # Constraining pseudo-guessing values to  0.1 for the first 10 items,
 # and to 0.05 for the remaining ones
 itemPar3PLconst(verbal[,1:24], c = c(rep(0.1, 10), rep(0.05, 14)))

## End(Not run)
```

itemParEst *Item parameter estimation for DIF detection*

Description

Fits a specified logistic IRT model and returns related item parameter estimates.

Usage

```
itemParEst(data, model, c = NULL, engine = "ltm", discr = 1)
```

Arguments

data	numeric: the data matrix.
model	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL").
c	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
discr	either NULL or a real positive value for the common discrimination parameter (default is 1). Used only if model is "1PL" and engine is "ltm". See **Details**.

Details

itemParEst permits to get item parameter estimates of some pre-specified logistic IRT model, together with estimates of the standard errors and the covariances between item parameters, if any. The output is ordered such that it can be directly used with the methods of Lord (difLord) and Raju (difRaju) and Generalized Lord's (difGenLord) to detect differential item functioning.

The data is a matrix whose rows correspond to the subjects and columns to the items.

Missing values are allowed but must be coded as NA values. They are discarded for item parameter estimation.

If the model is not the 1PL model, or if engine is equal to "ltm", the selected IRT model is fitted using marginal maximum likelihood by means of the functions from the ltm package (Rizopoulos, 2006). Otherwise, the 1PL model is fitted as a generalized linear mixed model, by means of the glmer function of the lme4 package (Bates and Maechler, 2009). With the "ltm" engine, the common discrimination parameter can be either fixed to a constant value using the discr argument, or it can be estimated (though not returned) by specifying discr to NULL. The default value of the common discrimination is 1.

The 3PL model can be fitted either unconstrained or by fixing the pseudo-guessing values. In the latter case the argument c holds either a numeric vector of same length of the number of items, with one value per item pseudo-guessing parameter, or a single value which is duplicated for all the items. If c is different from NULL then the 3PL model is always fitted (whatever the value of model).

Each row of the output matrix corresponds to one item of the data set; the number of columns depends on the fitted model. At most, nine columns are produced, with the unconstrained 3PL model. The order of the columns is the following: first, the estimates of item discrimination *a*, difficulty *b* and pseudo-guessing *c*; second, the corresponding standard errors *se(a)*, *se(b)* and *se(c)*; finally, the covariances between the item parameters, *cov(a,b)*, *cov(a,c)* and *cov(b,c)*.

If the 2PL model is fitted, only five columns are displayed: *a*, *b*, *se(a)*, *se(b)* and *cov(a,b)*. In case of the 1PL model, only *b* and *se(b)* are returned. If the constrained 3PL is considered, the output matrix holds six columns, the first five being identical to those from the 2PL model, and the last one holds the fixed pseudo-guessing parameters.

Value

A matrix with one row per item and at most nine columns, with item parameter estimates, standard errors and covariances, if any. See **Details**.

Note

Whenever making use of the `ltm` package to fit the IRT models, the linear parametrization is used, the covariance matrix is extracted with the `vcov()` function, and final standard errors and covariances are derived by the Delta method. See Rizopoulos (2006) for further details, and the `Note.pdf` document in the `difR` package for mathematical details.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
`<sebastien.beland@umontreal.ca>`
Gilles Raiche
Universite du Quebec a Montreal
`<raiche.gilles@uqam.ca>`

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1–25. doi:10.18637/jss.v017.i05

See Also

`itemPar1PL`, `itemPar2PL`, `itemPar3PL`, `itemPar3PLconst`, `difLord`, `difRaju`,
`difGenLord`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Estimation of the item parameters (1PL model, "ltm" engine)
 items.1PL <- itemParEst(verbal[,1:24], model = "1PL")

 # Estimation of the item parameters (1PL model, "ltm" engine,
 # estimated common discrimination parameter)
 items.1PL <- itemParEst(verbal[,1:24], model = "1PL", discr = NULL)

 # Estimation of the item parameters (1PL model, "lme4" engine)
 items.1PL <- itemParEst(verbal[,1:24], model = "1PL", engine = "lme4")

 # Estimation of the item parameters (2PL model)
 items.2PL <- itemParEst(verbal[,1:24], model = "2PL")

 # Estimation of the item parameters (3PL model)
 # items.3PL <- itemParEst(verbal[,1:24], model = "3PL")

 # Constraining all pseudo-guessing values to be equal to 0.05
 items.3PLc <- itemParEst(verbal[,1:24], model = "3PL", c = 0.05)

## End(Not run)
```

---

itemRescale *Rescaling item parameters by equal means anchoring*

Description

Rescale the item parameters from one data set to the scale of the parameters from another data set, using equal means anchoring.

Usage

```
itemRescale(mR, mF, items = 1:nrow(mR))
```

Arguments

mR	numeric: a matrix of item parameter estimates (one row per item) which constitutes the reference scale. See **Details**.
mF	numeric: a matrix of item parameter estimates (one row per item) which have to be rescaled. See **Details**.
items	a numeric vector of integer values specifying which items are used for equal means anchoring. See **Details**.

Details

The matrices mR and mF must have the same format as the output of the command itemParEst and one the possible models (1PL, 2PL, 3PL or constrained 3PL). The number of columns therefore equals two, five, nine or six, respectively.

Rescaling is performed by equal means anchoring (Cook and Eignor, 1991). The items involved in the anchoring process are specified by means of their row number in either mR or mF, and are passed through the items argument.

itemRescale primarily serves as a routine for item purification in Lord (difLord) and Raju (difRaju) Generalized Lord's (difGenLord) methods of DIF identification (Candell and Drasgow, 1988).

Value

A matrix of the same format as mF with the rescaled item parameters.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Candell, G.L. and Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253–260. doi:10.1177/014662168801200304

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

itemPar1PL, itemPar2PL, itemPar3PL, itemPar3PLconst, difLord, difRaju,

difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Splitting the data set into reference and focal groups
 nF <- sum(Gender)
 nR <- nrow(verbal)-nF
 data.ref <- verbal[,1:24][order(Gender),][1:nR,]
 data.focal <- verbal[,1:24][order(Gender),][(nR+1):(nR+nF),]

 # Estimating item parameters in each data set with 1PL model
 mR <- itemPar1PL(data.ref)
 mF <- itemPar1PL(data.focal)

 # Rescaling focal group item parameters, using all items for anchoring
 itemRescale(mR, mF)

 # Rescaling focal group item parameters, using the first 10 items for anchoring
 itemRescale(mR, mF, items = 1:10)

 # Estimating item parameters in each data set with 2PL model
 mR <- itemPar2PL(data.ref)
 mF <- itemPar2PL(data.focal)

 # Rescaling focal group item parameters, using all items for anchoring
 itemRescale(mR, mF)

## End(Not run)
```

LassoData — *Rearrange the data matrix for the Detection of DIF using the Lasso Approach (Magis et al. (2015)*

Description

A Function that rearrange the matrix to use the lasso DIF detection for dichotomous items.

Usage

```
LassoData(Data, group)
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership.

`group` numeric or character: either the vector of group membership or the column indicator (within Data) of group membership.

Details

This function rearranges the data matrix for use in lasso-based DIF detection with dichotomous items. It requires a matrix of dichotomous item responses and a vector indicating group membership.

Value

A matrix of five columns where, respectively,:

- `SCORE` is the total score.
- `GROUP` is the group membership.
- `PERS` is the number of the respondent.
- `Y` is the dichotomous answer to the item. Only "0" and "1" are allowed.
- `ITEM` is the item name (must be a character).

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
`<sebastien.beland@umontreal.ca>`
Carl F. Falk
Department of Psychology
McGill University (Canada)
`<carl.falk@mcgill.ca>`, https://www.mcgill.ca/psychology/carl-f-falk

References

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of Differential Item Functioning Using the Lasso Approach. Journal of Educational and Behavioral Statistics, 40(2), 111–135. https://doi.org/10.3102/1076998614559747

Examples

```
## Not run:

# Example with the verbal data

data(verbal)

LassoData(Data=verbal[,1:24], group=verbal[,26])

# Example with SimDichoDif to generate uniform DIF
```

```
It   <- 15 # number of items
ItDIFa <- NULL
ItDIFb <- c(1,3)
NR   <- 100 # number of responses for group 1 (reference)
NF   <- 100 # number of responses for group 2 (focal)
a    <- rep(1,It)
b    <- rnorm(It,1,.5)
Gb   <- rep(2,2)          # Group value for U-DIF
Ga   <- 0                 # Group value for NU-DIF: need to be fix to 0 for U-DIF
Out1 <- SimDichoDif(It,ItDIFa,ItDIFb,NR,NF,a,b,Ga,Gb)
Data<-Out1$data[,1:15]
Member<-Out1$data[,16]

LassoData(Data=Data, group=Member)


## End(Not run)
```

`lassoDIF.ABWIC` *Detection of Differential Item Functioning Using the Lasso Approach: Selection of Optimal λ Value*

Description

Performs DIF detection using a lasso-penalized logistic regression model for dichotomous items and selects the optimal value of the penalty parameter λ using an information criterion.

Usage

```
lassoDIF.ABWIC(Data, group, type = "AIC", N = NULL, lambda = NULL, ...)
```

Arguments

`...`	Additional arguments passed to internal methods.
`Data`	A numeric data frame or matrix: either only the item responses or the item responses with a group membership column.
`group`	A numeric or character vector: either a vector of group membership or a column index/name indicating group membership in `Data`.
`type`	Character string indicating the criterion used to select the optimal λ value. Must be one of `"AIC"`, `"BIC"`, or `"WIC"`.
`N`	Integer: total sample size. If `NULL`, it is inferred from the number of rows in `Data`.
`lambda`	Optional numeric vector of λ values to be used in the penalization path. If `NULL`, a default sequence is used.

Details

This function detects uniform DIF using a penalized logistic regression model based on the 2PL model. The model includes item-by-group interaction terms that are subject to lasso penalization. The optimal λ value is selected based on either the AIC, BIC, or WIC criterion.

For the selected λ^*, the function returns DIF parameters for all items, and flags items whose corresponding DIF parameters are non-zero.

Note: the function's behavior is sensitive to input parameters (e.g., criterion type, sample size, λ grid). It is strongly recommended to explore different settings and validate findings before interpreting DIF detection results.

Value

A list with the following components:

`DIFitems`	Indices of items flagged as exhibiting DIF.
`DIFpars`	Matrix of estimated DIF parameters for each item.
`crit.value`	Numeric vector of criterion values (e.g., AIC or BIC) across the λ path.
`crit.type`	The criterion used to select the optimal λ (either `"AIC"`, `"BIC"`, or `"WIC"`).
`lambda`	Vector of λ values considered.
`opt.lambda`	The optimal λ value selected.
`glmnet.fit`	Fitted `glmnet` model object.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Carl F. Falk
Department of Psychology
McGill University (Canada)
<carl.falk@mcgill.ca>, https://www.mcgill.ca/psychology/carl-f-falk
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of Differential Item Functioning Using the Lasso Approach. Journal of Educational and Behavioral Statistics, 40(2), 111–135. https://doi.org/10.3102/1076998614559747

Examples

```
## Not run:

# With the Verbal data set

data(verbal)

Dat    <-verbal[,1:20]
Member <-verbal[,26]

# Using AIC for selection
lassoDIF.ABWIC(Dat, Member, type="AIC")

# Using BIC for selection
lassoDIF.ABWIC(Dat, Member, type="BIC")

# With simulated data

It   <- 15 # number of items
ItDIFa <- NULL
ItDIFb <- c(1,3)
NR   <- 100 # number of responses for group 1 (reference)
NF   <- 100 # number of responses for group 2 (focal)
a    <- rep(1,It)          # for tests: runif(It,0.2,.5)
b    <- rnorm(It,1,.5)
Gb   <- rep(2,2)           # Group value for U-DIF
Ga   <- 0                  # Group value for NU-DIF: need to be fix to 0 for U-DIF
Out1 <- SimDichoDif(It,ItDIFa,ItDIFb,
NR,NF,a,b,Ga,Gb)
Dat<-Out1$data[,1:15]
Member<-Out1$data[,16]

# Using AIC for selection
lassoDIF.ABWIC(Dat, Member, type="AIC")

# Using BIC for selection
lassoDIF.ABWIC(Dat, Member, type="BIC")

# This plot shows how the estimated DIF effects for each item evolve
# as the lasso penalty (lambda) increases

aic.res <- lassoDIF.ABWIC(Dat, Member, type="AIC")
plot_lasso_paths(aic.res$glmnet.fit)
bic.res <- lassoDIF.ABWIC(Dat, Member, type="BIC")
plot_lasso_paths(bic.res$glmnet.fit)


## End(Not run)
```

lassoDIF.CV — *Detection of Differential Item Functioning Using the Lasso Approach: Selection of Optimal λ via Cross-Validation*

Description

Performs DIF detection using a lasso-penalized logistic regression model for dichotomous items and selects the optimal penalty parameter λ via cross-validation.

Usage

```
lassoDIF.CV(Data, group, nfold = 5, lambda = NULL, ...)
```

Arguments

`...`	Additional arguments passed to internal methods.
`Data`	A numeric data frame or matrix: either only the item responses or the item responses with a group membership column.
`group`	A numeric or character vector: either a vector of group membership or a column index/name indicating group membership in `Data`.
`nfold`	Integer: the number of folds used in cross-validation. Default is 5.
`lambda`	Optional numeric vector of λ values to be used in the penalization path. If `NULL`, a default sequence is used.

Details

This function detects uniform differential item functioning (DIF) using a lasso-penalized logistic regression model and selects the penalty parameter λ^* that minimizes cross-validation error. For this selected value, the function returns the estimated DIF parameters for all items and flags those with non-zero DIF effects.

Note: The performance of the method depends on choices such as the number of folds and the grid of λ values. We strongly recommend testing different configurations to assess the robustness of the results before interpretation.

Value

A list with the following components:

`DIFitems`	Indices of items flagged as exhibiting DIF.
`DIFpars`	Matrix of estimated DIF parameters for each item.
`crit.value`	Cross-validation criterion values (deviance) across the λ path.
`crit.type`	The type of criterion used, here `"cv"`.
`lambda`	Vector of λ values considered.
`opt.lambda`	The optimal λ value selected via cross-validation.
`glmnet.fit`	Fitted `glmnet` model object.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Carl F. Falk
Department of Psychology
McGill University (Canada)
<carl.falk@mcgill.ca>, https://www.mcgill.ca/psychology/carl-f-falk
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of Differential Item Functioning Using the Lasso Approach. Journal of Educational and Behavioral Statistics, 40(2), 111–135. https://doi.org/10.3102/1076998614559747

Examples

```
## Not run:

# With the Verbal data set

data(verbal)

Dat    <-verbal[,1:20]
Member <-verbal[,26]

# Using cross-validation
set.seed(1234)

cv.res <- lassoDIF.CV(Dat, Member, nfold=5)
cv.res

# With simulated data

It    <- 15 # number of items
ItDIFa <- NULL
ItDIFb <- c(1,3)
NR    <- 100 # number of responses for group 1 (reference)
NF    <- 100 # number of responses for group 2 (focal)
a     <- rep(1,It)          # for tests: runif(It,0.2,.5)
b     <- rnorm(It,1,.5)
Gb    <- rep(2,2)           # Group value for U-DIF
Ga    <- 0                  # Group value for NU-DIF: need to be fix to 0 for U-DIF
Out1 <- SimDichoDif(It,ItDIFa,ItDIFb,NR,NF,a,b,Ga,Gb)
Dat<-Out1$data[,1:15]
Member<-Out1$data[,16]
```

```
set.seed(1234) # appears to be sensitive to random number seed

cv.res <- lassoDIF.CV(Dat, Member, nfold=5)
cv.res


## End(Not run)
```

`liu_agresti_ccor` *Liu-Agresti Common Cumulative Odds Ratio*

Description

Computes the Liu-Agresti estimate of the common cumulative odds ratio (Ψ) and its reciprocal (α) for ordinal data from two independent groups. This statistic quantifies the direction and strength of ordinal association between groups.

Usage

```
liu_agresti_ccor(responses, group)
```

Arguments

`responses`	A numeric vector of ordinal item responses. Categories must be coded as integers (e.g., 1 to 5 for a Likert-type scale).
`group`	A grouping vector indicating the group to which each observation belongs. It must contain exactly two unique values (e.g., "ref" and "foc").

Details

This function creates a 2 x J contingency table, where J is the number of distinct ordinal response categories. It computes cumulative marginal frequencies and estimates the odds ratio using Liu and Agresti's formulation (1996, Eq. 2). The variance of the log-transformed estimate is computed according to their Eq. 3.

The estimate $\hat{\Psi}$ is based on cumulative frequencies and is designed for ordinal response categories. It quantifies the association between group membership and the likelihood of higher category responses.

The function does not support missing values; observations with `NA` should be removed prior to use.

If one of the response categories is completely absent from one group, then the cumulative margins used in the computation may contain zero values. In such cases, either the numerator or the denominator of the Liu-Agresti formula will be zero, making the estimate undefined. When this occurs, the function returns `NA` and issues a warning.

About the notation: In the original article by Liu and Agresti (1996), the cumulative logistic model uses the parameters β and θ. To avoid any confusion with a logistic model or the IRT framework, the symbol ψ is used here to denote the group effect.

Value

A matrix with one row and three columns containing:

Psi_hat	The Liu-Agresti estimate of the common cumulative odds ratio ($\hat{\Psi}$).
Alpha_hat	The reciprocal of $\hat{\Psi}$.
SE_log_Psi	The standard error of $\log(\hat{\Psi})$, which can be used to construct confidence intervals or conduct hypothesis testing.

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Liu, I., & Agresti, A. (1996). Mantel-Haenszel-Type Inference for Cumulative Odds Ratios with a Stratified Ordinal Response. *Biometrics*, 52(4), 1223–1234.

Examples

```
# Simulated balanced example
set.seed(123)

group <- rep(c("ref", "foc"), each = 100)
stopifnot(length(group) == 200)

responses <- sample(1:4, size = length(group), replace = TRUE)
stopifnot(length(responses) == length(group))

liu_agresti_ccor(as.integer(responses), factor(group))
```

Logistik *Logistic regression DIF statistic*

Description

Calculates the "logistic regression" likelihood-ratio statistics and effect sizes for DIF detection.

Usage

```
Logistik(data, member, member.type = "group", match = "score",
 anchor = 1:ncol(data), type = "both", criterion = "LRT", all.cov = FALSE)
```

Arguments

`data`	numeric: the data matrix (one row per subject, one column per item).
`member`	numeric or factor: the vector of group membership. Can either take two distinct values (zero for the reference group and one for the focal group) or be a continuous vector. See **Details**.
`member.type`	character: either `"group"` (default) to specify that group membership is made of two groups, or `"cont"` to indicate that group membership is based on a continuous criterion. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the total test score based on the anchor items, or `"restscore"` to compute the matching score while excluding the item currently being tested. This prevents contamination of the matching variable by the item itself. Alternatively, any numeric vector with the same length as the number of rows in `data` can be supplied as an external matching variable.
`anchor`	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. Ignored if `match` is not `"score"`. See **Details**.
`type`	a character string specifying which DIF effects must be tested. Possible values are `"both"` (default), `"udif"` and `"nudif"`. See **Details**.
`criterion`	a character string specifying which DIF statistic is computed. Possible values are `"LRT"` (default) or `"Wald"`. See **Details**.
`all.cov`	logical: should *all* covariance matrices of model parameter estimates be returned (as lists) for both nested models and all items? (default is `FALSE`.

Details

This command computes the logistic regression statistic (Swaminathan and Rogers, 1990) in the specific framework of differential item functioning. It forms the basic command of `difLogistic` and is specifically designed for this call.

If the `member.type` argument is set to `"group"`, the `member` argument must be a vector with two distinct (numeric or factor) values, say 0 and 1 (for the reference and focal groups respectively). Those values are internally transformed onto factors to denote group membership. The three possible models to be fitted are then:

$$M_0 : logit(\pi_g) = \alpha + \beta X + \gamma_g + \delta_g X$$
$$M_1 : logit(\pi_g) = \alpha + \beta X + \gamma_g$$
$$M_2 : logit(\pi_g) = \alpha + \beta X$$

where π_g is the probability of answering correctly the item in group g and X is the matching variable. Parameters α and β are the intercept and the slope of the logistic curves (common to all groups), while γ_g and δ_g are group-specific parameters. For identification reasons the parameters γ_0 and δ_0 for reference group ($g = 0$) are set to zero. The parameter γ_1 of the focal group ($g = 1$) represents the uniform DIF effect, and the parameter δ_1 is used to model nonuniform DIF effect. The models are fitted with the `glm` function.

If `member.type` is set to `"cont"`, then "group membership" is replaced by a continuous or discrete variable, given by the `member` argument, and the models above are written as

$$M_0 : logit(\pi_g) = \alpha + \beta X + \gamma Y + \delta XY$$
$$M_1 : logit(\pi_g) = \alpha + \beta X + \gamma Y$$
$$M_2 : logit(\pi_g) = \alpha + \beta X$$

where `Y` is the group variable. Parameters γ and δ act now as the γ_1 and δ_1 DIF parameters.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `Logistik` function. This is specified by the `match` argument. By default, it takes the value `"score"` and the test score (i.e. raw score) is computed. The second option is to assign to `match` a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the `data` matrix.

Two types of DIF statistics can be computed: the likelihood ratio test statistics, obtained by comparing the fit of two nested models, and the Wald statistics, obtained with an appropriate contrast matrix for testing the model parameters (Johnson and Wichern, 1998). These are specified by the argument `criterion`, with respective values `"LRT"` and `"Wald"`. By default, the LRT statistics are computed.

If `criterion` is `"LRT"`, the argument `type` determines the models to be compared by means of the LRT statistics. The three possible values of `type` are: `type="both"` (default) which tests the hypothesis $H_0 : \gamma_1 = \delta_1 = 0$ (or $H_0 : \gamma = \delta = 0$) by comparing models M_0 and M_2; `type="nudif"` which tests the hypothesis $H_0 : \delta_1 = 0$ (or $H_0 : \delta = 0$) by comparing models M_0 and M_1; and `type="udif"` which tests the hypothesis $H_0 : \gamma_1 = 0$ (or $H_0 : \gamma = 0$) by comparing models M_1 and M_2 (assuming that $\delta_1 = 0$ or $\delta = 0$). In other words, `type="both"` tests for DIF (without distinction between uniform and nonuniform effects), while `type="udif"` and `type="nudif"` test for uniform and nonuniform DIF, respectively.

If `criterion` is `"Wald"`, the argument `type` determines the logistic model to be considered and the appropriate contrast matrix. If `type=="both"`, the considered model is model M_0 and the contrast matrix has two rows, (0,0,1,0) and (0,0,0,1). If `type=="nudif"`, the considered model is also model M_0 but the contrast matrix has only one row, (0,0,0,1). Eventually, if `type=="udif"`, the considered model is model M_1 and the contrast matrix has one row, (0,0,1).

The data are passed through the `data` argument, with one row per subject and one column per item. Missing values are allowed but must be coded as `NA` values. They are discarded from the fitting of the logistic models (see `glm` for further details).

The vector of group membership, specified with `member` argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

Option `anchor` sets the items which are considered as anchor items for computing the test scores and related logistic regression DIF statistics. Items other than the anchor items and the tested item are discarded. `anchor` must hold integer values specifying the column numbers of the corresponding anchor items. It is mainly designed to perform item purification. Note that this option is discarded when `match` is not `"score"`.

The output contains: the selected DIF statistics (either the LRT or the Wald statistic) computed for each item, two matrices with the parameter estimates of both models (for each item) and two matrices of related standard error values. In addition, Nagelkerke's R^2 coefficients (Nagelkerke, 1991) are computed for each model and the output returns both, the vectors of R^2 coefficients for each model and the differences in these coefficients. Such differences are used as measures of effect size by the difLogistic command; see Gomez-Benito, Dolores Hidalgo and Padilla (2009), Jodoin and Gierl (2001) and Zumbo and Thomas (1997). The criterion and member.type arguments are also returned, as well as a character argument named match that specifies the type of matching criterion that was used.

Value

A list with nine components:

- `stat` — the values of the logistic regression DIF statistics.
- `R2M0` — the values of Nagelkerke's R^2 coefficients for the "full" model.
- `R2M1` — the values of Nagelkerke's R^2 coefficients for the "simpler" model.
- `deltaR2` — the differences between Nagelkerke's R^2 coefficients of the tested models. See **Details**.
- `parM0` — a matrix with one row per item and four columns, holding successively the fitted parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}_1$ and $\hat{\delta}_1$ of the "full" model (M_0 if type="both" or type="nudif", M_1 if type="udif").
- `parM1` — the same matrix as parM0 but with fitted parameters for the "simpler" model (M_1 if type="nudif", M_2 if type="both" or type="udif").
- `seM0` — a matrix with the standard error values of the parameter estimates in matrix parM0.
- `seM1` — a matrix with the standard error values of the parameter estimates in matrix parM1.
- `cov.M0` — either NULL (if all.cov argument is FALSE) or a list of covariance matrices of parameter estimates of the "full" model (M_0) for each item (if all.cov argument is TRUE).
- `cov.M1` — either NULL (if all.cov argument is FALSE) or a list of covariance matrices of parameter estimates of the "reduced" model (M_1) for each item (if all.cov argument is TRUE).
- `criterion` — the value of the criterion argument.
- `member.type` — the value of the member.type argument.
- `match` — a character string, either "score" or "matching variable" depending on the match argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>

References

Gomez-Benito, J., Dolores Hidalgo, M. and Padilla, J.-L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology, 5*, 18-25. doi:10.1027/1614-2241.5.1.18

Jodoin, M. G. and Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349. doi:10.1207/S15324818AME1404_2

Johnson, R. A. and Wichern, D. W. (1998). *Applied multivariate statistical analysis (fourth edition)*. Upper Saddle River, NJ: Prentice-Hall.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692. doi:10.1093/biomet/78.3.691

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

Zumbo, B. D. and Thomas, D. R. (1997). A measure of effect size for a model-based approach for studying DIF. Prince George, Canada: University of Northern British Columbia, Edgeworth Laboratory for Quantitative Behavioral Science.

See Also

`difLogistic`, `dichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Testing both types of DIF simultaneously
 # With all items, test score as matching criterion
 Logistik(verbal[,1:24], verbal[,26])

 # Returning all covariance matrices of model parameters
 Logistik(verbal[,1:24], verbal[,26], all.cov = TRUE)

 # Testing both types of DIF simultaneously
 # With all items and Wald test
 Logistik(verbal[,1:24], verbal[,26], criterion = "Wald")

 # Removing item 6 from the set of anchor items
 Logistik(verbal[,1:24], verbal[,26], anchor = c(1:5, 7:24))
```

```
  # Testing for nonuniform DIF
  Logistik(verbal[,1:24], verbal[,26], type = "nudif")

  # Testing for uniform DIF
  Logistik(verbal[,1:24], verbal[,26], type = "udif")

  # Using the "anger" trait variable as matching criterion
  Logistik(verbal[,1:24],verbal[,26], match = verbal[,25])

  # Using the "anger" trait variable as group membership
  Logistik(verbal[,1:24],verbal[,25], member.type = "cont")

## End(Not run)
```

LogistikPoly — *Detection of DIF in polytomous (ordinal) items using cumulative logistic regression*

Description

This function implements a method for detecting Differential Item Functioning (DIF) in ordinal response items using cumulative logistic regression (`vglm` with the `propodds` family).

Usage

```
LogistikPoly(data, member, member.type = "group", match = "score",
             anchor = 1:ncol(data), type = "both", criterion = "LRT",
             all.cov = FALSE)
```

Arguments

`data`	A data.frame or matrix of item responses (ordinal scale), with one row per subject, one column per item.
`member`	A vector indicating group membership (e.g., reference vs. focal group).
`member.type`	Type of the group variable. Use `"group"` (default) for a categorical variable; a continuous covariate may also be provided.
`match`	matching variable: `"score"`, `"restscore"`, or an external numeric vector.
`anchor`	Indices of items used to compute the matching score (default is all items).
`type`	Type of DIF tested: `"both"` (uniform and non-uniform), `"udif"` (only uniform), or `"nudif"` (only non-uniform).
`criterion`	Model comparison criterion. Use `"LRT"` (likelihood-ratio test) or `"Wald"` (Wald test).
`all.cov`	Logical; if `TRUE`, returns the variance-covariance matrices of the model parameters for each item.

Details

This function compares nested cumulative logistic regression models to detect DIF in polytomous (ordinal) items. The full model includes group membership and its interaction with the matching variable (depending on the selected `type`).

If `match = "score"`, the total test score (based on anchor items) is used as the matching variable. This is the classical approach and allows for the application of iterative purification, whereby items identified as DIF are progressively excluded from the anchor set and the matching score is updated. If `match = "restscore"`, the matching score is computed by excluding the item currently being tested from the total score. However, since the matching score varies across items, purification cannot be applied under this setting.

Larger test statistics values may indicate potential DIF.

McKelvey-Zavoina pseudo R^2 is used to compute model fit for both the full and reduced models, and their difference (`deltaR2`) is also provided.

For each item, the DIF analysis is performed using only complete cases. Respondents with missing data on the item being tested, the matching variable, or the group variable are excluded from the estimation for that item.

Value

A list with the following elements:

`stat`	DIF test statistic (LRT or Wald) for each item.
`R2M0`	McKelvey-Zavoina pseudo R^2 for the full model (with group).
`R2M1`	McKelvey-Zavoina pseudo R^2 for the reduced model (without group).
`deltaR2`	Difference in R^2 between full and reduced models.
`parM0`	Matrix of parameter estimates for the full model.
`parM1`	Matrix of parameter estimates for the reduced model.
`seM0`	Standard errors for the parameters in the full model.
`seM1`	Standard errors for the parameters in the reduced model.
`cov.M0`	List of variance-covariance matrices for the full model (if `all.cov = TRUE`).
`cov.M1`	List of variance-covariance matrices for the reduced model (if `all.cov = TRUE`).
`criterion`	Criterion used for DIF detection ("LRT" or "Wald").
`member.type`	Type of group membership variable.
`match`	Indicates the type of matching method used ("score" or custom variable).

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential ItemFunctioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Examples

```
## Not run:

# With real data

attach(SCS)

# With Wald procedure
LogistikPoly(data=SCS[,1:10], member=SCS[,11],criterion = "Wald")

# Testing for non-uniform DIF
LogistikPoly(data=SCS[,1:10], member=SCS[,11],type = "nudif")

# Testing for uniform DIF
LogistikPoly(data=SCS[,1:10], member=SCS[,11],type = "udif")

# Use of the rest scores
LogistikPoly(data=SCS[,1:10], member=SCS[,11], match = "restscore")

# With simulated data

set.seed(1234)

# original item parameters
a <- rlnorm(10,-.5) # slopes
b <- runif(10, -2, 2) # difficulty
d <- list() # step parameters
d[[1]] <- c(0, 2, .5, -.15, -1.1)
d[[2]] <- c(0, 2, .25, -.45, -.75)
d[[3]] <- c(0, 1, .5, -.65, -1)
d[[4]] <- c(0, 2, .5, -.85, -2)
d[[5]] <- c(0, 1, .25, -.05, -1)
d[[6]] <- c(0, 2, .5, -.95, -1)
d[[7]] <- c(0, 1, .25, -.35, -2)
d[[8]] <- c(0, 2, .5, -.15, -1)
d[[9]] <- c(0, 1, .25, -.25, -2)
d[[10]] <- c(0, 2, .5, -.35, -1)

# Change only a few item parameters
# Uniform DIF
It <- 10
```

```
NR <- 1000
NF <- 1000
ItDIFa <- NULL
Ga <- NULL
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2) # 2 items w/ difficulty parameter that is higher in group 2

Out.Unif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d, ncat=5, Ga, Gb)
#Out.Unif
Out.Unif$ipars
Data <- Out.Unif$data

# With Wald procedure
LogistikPoly(data=Out.Unif$data[,1:10], member=Out.Unif$data[,11], criterion = "Wald")

# Testing for non-uniform DIF
LogistikPoly(data=Out.Unif$data[,1:10], member=Out.Unif$data[,11], type = "nudif")

# Testing for uniform DIF
LogistikPoly(data=Out.Unif$data[,1:10], member=Out.Unif$data[,11], type = "udif")

# Use of the rest scores
LogistikPoly(data=Out.Unif$data[,1:10], member=Out.Unif$data[,11], match = "restscore")


## End(Not run)
```

LordChi2 *Lord's chi-square DIF statistic*

Description

Calculates the Lord's chi-square statistics for DIF detection.

Usage

```
LordChi2(mR, mF)
```

Arguments

mR	numeric: the matrix of item parameter estimates (one row per item) for the reference group. See **Details**.
mF	numeric: the matrix of item parameter estimates (one row per item) for the focal group. See **Details**.

Details

This command computes the Lord's chi-square statistic (Lord, 1980) in the specific framework of differential item functioning. It forms the basic command of difLord and is specifically designed for this call.

The matrices mR and mF must have the same format as the output of the command itemParEst with one the possible models (1PL, 2PL, 3PL or constrained 3PL). The number of columns therefore equals two, five, nine or six, respectively. Moreover, item parameters of the focal must be on the same scale of that of the reference group. If not, make use of e.g. equal means anchoring (Cook and Eignor, 1991) and itemRescale to transform them adequately.

Value

A vector with the values of the Lord's chi-square DIF statistics.

Note

WARNING: the previous versions of LordChi2 were holding an error: under the 3PL model, the covariance matrices Sig_1 and Sig_2 were wrongly computed as the variance of the pseudo-guessing parameters were replaced by the parameter estimates. This has been fixed from version 4.0 of difR. Many thanks to J. Patrick Meyer (Curry School of Education, University of Virginia) for having discovered this mistake.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

itemParEst, itemRescale, difLord, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Splitting the data into reference and focal groups
 nF <- sum(Gender)
 nR <- nrow(verbal)-nF
 data.ref <- verbal[, 1:24][order(Gender),][1:nR,]
 data.focal <- verbal[, 1:24][order(Gender),][(nR+1):(nR+nF),]

 # Pre-estimation of the item parameters (1PL model)
 mR <- itemParEst(data.ref, model = "1PL")
 mF <- itemParEst(data.focal, model = "1PL")
 mF <- itemRescale(mR, mF)
 LordChi2(mR, mF)

 # Pre-estimation of the item parameters (2PL model)
 mR <- itemParEst(data.ref, model = "2PL")
 mF <- itemParEst(data.focal, model = "2PL")
 mF <- itemRescale(mR, mF)
 LordChi2(mR, mF)

 # Pre-estimation of the item parameters (constrained 3PL model)
 mR <- itemParEst(data.ref, model = "3PL", c = 0.05)
 mF <- itemParEst(data.focal, model = "3PL", c = 0.05)
 mF <- itemRescale(mR, mF)
 LordChi2(mR, mF)

## End(Not run)
```

LRT *Likelihood-Ratio Test DIF statistic*

Description

Calulates Likelihoo-Ratio Test (LRT) statistics for DIF detection.

Usage

```
LRT(data, member)
```

Arguments

data	numeric: the data matrix (one row per subject, one column per item).
member	numeric: the vector of group membership with zero and one entries only. See **Details**.

Details

This command computes the likelihood-ratio test statistic (Thissen, Steinberg and Wainer, 1988) in the specific framework of differential item functioning. It forms the basic command of `difLRT` and is specifically designed for this call.

The data are passed through the `data` argument, with one row per subject and one column per item. Missing values are allowed but must be coded as `NA` values.

The vector of group membership, specified with `member` argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

The LRT DIF statistic is computed for each item separately, using all other items as anchor items.

Value

A vector with the values of the LRT DIF statistics.

Note

Because of the fitting of the modified Rasch model with `glmer` the process can be very time consuming (see the **Details** section of `difLRT`).

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Bates, D. and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-31. http://CRAN.R-project.org/package=lme4

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See Also

`difLRT`, `dichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal)!="Anger"]

 # Keeping the first 5 items and the first 50 subjects
 # (this is an artificial simplification to reduce the computational time)
 # Sixth column holds the group membership
 verbal <- verbal[1:50, c(1:5, 25)]

 # Likelihood-ratio statistics
 LRT(verbal[,1:5], verbal[,6])

## End(Not run)
```

`mantelHaenszel` *Mantel-Haenszel DIF statistic*

Description

Calculates Mantel-Haenszel statistics for DIF detection.

Usage

```
mantelHaenszel(data, member, match = "score", correct = TRUE, exact = FALSE,
  anchor = 1:ncol(data))
```

Arguments

`data`	numeric: the data matrix (one row per subject, one column per item).
`member`	numeric: the vector of group membership with zero and one entries only. See **Details**.
`match`	specifies the type of matching criterion. Can be either `"score"` (default) to compute the total test score based on the anchor items, or `"restscore"` to compute the matching score while excluding the item currently being tested. This prevents contamination of the matching variable by the item itself. Alternatively, any numeric vector with the same length as the number of rows in `data` can be supplied as an external matching variable.
`correct`	logical: should the continuity correction be used? (default is `TRUE`).
`exact`	logical: should an exact test be computed? (default is `FALSE`).
`anchor`	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.

Details

This command basically computes the Mantel-Haenszel (1959) statistic in the specific framework of differential item functioning. It forms the basic command of `difMH` and is specifically designed for this call.

The data are passed through the `data` argument, with one row per subject and one column per item.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded from sum-score computation.

The vector of group membership, specified with `member` argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the `mantelHaenszel` function. This is specified by the `match` argument. By default, it takes the value `"score"` and the test score (i.e. raw score) is computed. The second option is to assign to `match` a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the `data` matrix.

By default, the continuity correction factor -0.5 is used (Holland and Thayer, 1988). One can nevertheless remove it by specifying `correct=FALSE`.

By default, the asymptotic Mantel-Haenszel statistic is computed. However, the exact statistics and related P-values can be obtained by specifying the logical argument `exact` to `TRUE`. See Agresti (1990, 1992) for further details about exact inference.

Option `anchor` sets the items which are considered as anchor items for computing Mantel-Haenszel statistics. Items other than the anchor items and the tested item are discarded. `anchor` must hold integer values specifying the column numbers of the corresponding anchor items. It is primarily designed to perform item purification.

In addition to the Mantel-Haenszel statistics to identify DIF items, `mantelHaenszel` computes the estimates of the common odds ratio α_{MH} which are used for measuring the effect size of the items (Holland and Thayer, 1985, 1988). They are returned in the `resAlpha` argument of the output list. Moreover, the logarithm of α_{MH}, say λ_{MH}, is asymptotically distributed and its variance is computed and returned into the `varLambda` argument. Note that this variance is the one proposed by Philips and Holland (1987), since it seems the most accurate expression for the variance of λ_{MH} (Penfield and Camilli, 2007).

Value

A list with several arguments:

`resMH`	the vector of the Mantel-Haenszel DIF statistics (either asymptotic or exact).
`resAlpha`	the vector of the (asymptotic) Mantel-Haenszel estimates of the common odds ratios. Returned only if `exact` is `FALSE`.
`varLambda`	the (asymptotic) variance of the λ_{MH} statistic. Returned only if `exact` is `FALSE`.
`Pval`	the exact P-values of the MH test. Returned only if `exact` is `TRUE`.
`match`	a character string, either `"score"` or `"matching variable"` depending on the `match` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7*, 131-177. doi:10.1214/ss/1177011454

Holland, P. W. and Thayer, D. T. (1985). An alternative definition of the ETS delta scale of item difficulty. *Research Report RR-85-43*. Princeton, NJ: Educational Testing Service.

Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Ed.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Penfield, R. D., and Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao and S. Sinharray (Eds.), *Handbook of Statistics 26: Psychometrics* (pp. 125-167). Amsterdam, The Netherlands: Elsevier.

Philips, A., and Holland, P. W. (1987). Estimators of the Mantel-Haenszel log odds-ratio estimate. *Biometrics, 43*, 425-431. doi:10.2307/2531824

See Also

difMH, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # With and without continuity correction
 mantelHaenszel(verbal[,1:24], verbal[,26])
```

```
 mantelHaenszel(verbal[,1:24], verbal[,26], correct = FALSE)

 # Exact test
 mantelHaenszel(verbal[,1:24], verbal[,26], exact = TRUE)

 # Removing item 6 from the set of anchor items
 mantelHaenszel(verbal[,1:24], verbal[,26], anchor = c(1:5,7:24))

## End(Not run)
```

`plot_lasso_paths` *Plot coefficient paths from LASSO DIF*

Description

This function displays coefficient trajectories from LASSO-regularized DIF detection.

Usage

```
plot_lasso_paths(
  out,
  nr.lambda = 100,
  highlight = NULL,
  title = "Regularization Paths of DIF Effects",
  ...
)
```

Arguments

`out`	A fitted object returned by `lassoDIF()`.
`nr.lambda`	Number of lambda values to evaluate and display (default is 100).
`highlight`	Optional: indices of items to highlight in color.
`title`	Main title of the plot.
`...`	Additional graphical parameters passed to `plot()`.

Value

A base R plot of coefficient paths.

RajuZ *Raju's area DIF statistic*

Description

Calculates the Raju's statistics for DIF detection.

Usage

```
RajuZ(mR, mF, signed = FALSE)
```

Arguments

mR	numeric: the matrix of item parameter estimates (one row per item) for the reference group. See **Details**.
mF	numeric: the matrix of item parameter estimates (one row per item) for the focal group. See **Details**.
signed	logical: should the *signed* area be computed, or the *unsigned* (i.e. in absolute value) ara? Default is `FALSE`, i.e. the unsigned area. See **Details**.

Details

This command computes the Raju's area statistic (Raju, 1988, 1990) in the specific framework of differential item functioning. It forms the basic command of `difRaju` and is specifically designed for this call.

The matrices `mR` and `mF` must have the same format as the output of the command `itemParEst` and one the possible models (1PL, 2PL or constrained 3PL). The number of columns therefore equals two, five or six, respectively. Note that the unconstrained 3PL model cannot be used in this method: all pseudo-guessing parameters must be equal in both groups of subjects. Moreover, item parameters of the focal must be on the same scale of that of the reference group. If not, make use of e.g. equal means anchoring (Cook and Eignor, 1991) and `itemRescale` to transform them adequately.

By default, the *unsigned* area, given by Equation (57) in Raju (1990), is computed. It makes use of Equations (14), (15), (23) and (46) for the numerator, and Equations (17), (33) to (39), and (52) for the denominator of the Z statistic. However, the *signed* area, given by Equation (56) in Raju (1990), can be used instead. In this case, Equations (14), (21) and (44) are used for the numerator, and Equations (17), (25) and (48) for the denominator. The choice of the type of area is fixed by the logical *signed* argument, with default value `FALSE`.

Value

A list with two components:

res	a matrix with one row per item and three columns, holding respectively Raju's area between the two item characteristic curves, its standard error and the Raju DIF statistic (the latter being the ratio of the first two columns).
signed	the value of the `signed` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Cook, L. L. and Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice, 10*, 37-45.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495-502. doi:10.1007/BF02294403

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. doi:10.1177/014662169001400208

See Also

itemParEst, itemRescale, difRaju, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Splitting the data into reference and focal groups
 nF <- sum(Gender)
 nR <- nrow(verbal)-nF
 data.ref <- verbal[,1:24][order(Gender),][1:nR,]
 data.focal <- verbal[,1:24][order(Gender),][(nR+1):(nR+nF),]

 # Pre-estimation of the item parameters (1PL model)
 mR <- itemParEst(data.ref,model = "1PL")
 mF <- itemParEst(data.focal,model = "1PL")
 mF <- itemRescale(mR, mF)
```

```
# Signed and unsigned Raju statistics
RajuZ(mR, mF)
RajuZ(mR, mF, signed = TRUE)

# Pre-estimation of the item parameters (2PL model)
mR <- itemParEst(data.ref, model = "2PL")
mF <- itemParEst(data.focal, model = "2PL")
mF <- itemRescale(mR, mF)

# Signed and unsigned Raju statistics
RajuZ(mR, mF)
RajuZ(mR, mF, signed = TRUE)

# Pre-estimation of the item parameters (constrained 3PL model)
mR <- itemParEst(data.ref, model = "3PL", c = 0.05)
mF <- itemParEst(data.focal, model = "3PL", c =0 .05)
mF <- itemRescale(mR, mF)

# Signed and unsigned Raju statistics
RajuZ(mR, mF)
RajuZ(mR, mF, signed = TRUE)

## End(Not run)
```

SCS *Sexual Compulsivity Scale Data Set*

Description

The items were rated on a likert scale (1=Not at all like me, 2=Slightly like me, 3=Mainly like me, 4=Very much like me):

- $Q1$: My sexual appetite has gotten in the way of my relationships.
- $Q2$: My sexual thoughts and behaviors are causing problems in my life.
- $Q3$: My desires to have sex have disrupted my daily life.
- $Q4$: I sometimes fail to meet my commitments and responsibilities because of my sexual behaviors.
- $Q5$: I sometimes get so horny I could lose control.
- $Q6$: I find myself thinking about sex while at work.
- $Q7$: I feel that sexual thoughts and feelings are stronger than I am.
- $Q8$: I have to struggle to control my sexual thoughts and behavior.
- $Q9$: I think about sex more than I would like to.
- $Q10$. It has been difficult for me to find sex partners who desire having sex as much as I want to.
- $Gender$: chosen from drop down list (1=male, 2=female).

Format

The SCS matrix consists of 3215 rows (one per subject) and 11 columns (one per item).

Source

The full dataset is available at the following URL: *https://openpsychometrics.org/_rawdata/*

References

Kalichman, S. C., & Rompa, D. (1995). Sexual sensation seeking and sexual compulsivity scales: Reliability, validity, and predicting HIV risk behavior. *Journal of Personality Assessment*, 65(3), 586–601. https://doi.org/10.1207/s15327752jpa6503_16

---

selectDif — *Selection of one of the DIF detection methods*

Description

This function performs DIF detection for one pre-specified method and is applicable only to methods designed for dichotomous items.

Usage

```
selectDif(Data, group, focal.name, method, anchor = NULL, props = NULL,
  thrTID = 1.5, alpha = 0.05, MHstat = "MHChisq", correct = TRUE,
  exact = FALSE, stdWeight = "focal", thrSTD = 0.1, BDstat = "BD",
  member.type = "group", match = "score", type = "both", criterion = "LRT",
  model = "2PL", c = NULL, engine = "ltm", discr = 1, irtParam = NULL,
  same.scale = TRUE, signed = FALSE, purify = FALSE, purType = "IPP1",
  nrIter = 10, extreme = "constraint", const.range = c(0.001, 0.999),
  nrAdd = 1, p.adjust.method = NULL, save.output = FALSE,
  output = c("out", "default"))
```

Arguments

`Data`	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
`group`	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
`focal.name`	numeric or character indicating the level of `group` which corresponds to the focal group.
`method`	character: the name of the selected method. Possible values are `"TID"`, `"MH"`, `"Std"`, `"Logistic"`, `"BD"`, `"SIBTEST"`, `"Lord"`, `"Raju"` and `"LRT"`. See **Details**.

anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
props	either NULL (default) or a two-column matrix with proportions of success in the reference group and the focal group. See **Details** .
thrTID	numeric: the threshold for detecting DIF items with TID method (default is 1.5).
alpha	numeric: significance level (default is 0.05).
MHstat	character: specifies the DIF statistic to be used for DIF identification. Possible values are "MHChisq" (default) and "logOR". See **Details** .
correct	logical: should the continuity correction be used? (default is TRUE).
exact	logical: should an exact test be computed? (default is FALSE).
stdWeight	character: the type of weights used for the standardized P-DIF statistic. Possible values are "focal" (default), "reference" and "total". See **Details**.
thrSTD	numeric: the threshold (cut-score) for standardized P-DIF statistic (default is 0.10).
BDstat	character specifying the DIF statistic to be used. Possible values are "BD" (default) and "trend". See **Details**.
member.type	character: either "group" (default) to specify that group membership is made of two groups, or "cont" to indicate that group membership is based on a continuous criterion. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "both" (default), "udif" and "nudif". See **Details**.
criterion	a character string specifying which DIF statistic is computed. Possible values are "LRT" (default) or "Wald". See **Details**.
model	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL"). Default is "2PL".
c	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
discr	either NULL or a real positive value for the common discrimination parameter (default is 1). Used onlky if model is "1PL" and engine is "ltm". See **Details**.
irtParam	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
same.scale	logical: are the item parameters of the irtParam matrix on the same scale? (default is "TRUE"). See **Details**.
signed	logical: should the Raju's statistics be computed using the signed (TRUE) or unsigned (FALSE, default) area? See **Details**.
purify	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).

purType	character: the type of purification process to be run. Possible values are "IPP1" (default), "IPP2" and "IPP3". Ignored if purify is FALSE or if method is not "TID".
nrIter	numeric: the maximal number of iterations in the item purification process (default is 10).
extreme	character: the method used to modify the extreme proportions. Possible values are "constraint" (default) or "add". Ignored if method is not "TID".
const.range	numeric: a vector of two constraining proportions. Default values are 0.001 and 0.999. Ignored if method is not "TID" or if extreme is "add".
nrAdd	integer: the number of successes and the number of failures to add to the data in order to adjust the proportions. Default value is 1. Ignored if method is not "TID" or if extreme is "constraint".
p.adjust.method	either NULL (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
save.output	logical: should the output be saved into a text file? (Default is FALSE).
output	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or "default" (default value). See **Details**.

Details

This is a generic function which calls one of the DIF detection methods and displays its output. It is mainly used as a routine for dichoDif command.

The possible methods are:

1. "TID" for Transformed Item Difficulties (TID) method (Angoff and Ford, 1973),
2. "MH" for mantel-Haenszel (Holland and Thayer, 1988),
3. "Std" for standardization (Dorans and Kulick, 1986),
4. "BD" for Breslow-Day method (Penfield, 2003),
5. "Logistic" for logistic regression (Swaminathan and Rogers, 1990),
6. "SIBTEST" for SIBTEST (Shealy and Stout) and Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996) methods,
7. "Lord" for Lord's chi-square test (Lord, 1980),
8. "Raju" for Raju's area method (Raju, 1990), and
9. "LRT" for likelihood-ratio test method (Thissen, Steinberg and Wainer, 1988).

The Data is a matrix whose rows correspond to the subjects and columns to the items. In addition, Data can hold the vector of group membership. If so, group indicates the column of Data which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, group must be a vector of same length as nrow(Data).

Missing values are allowed for item responses (not for group membership) but must be coded as NA values. They are discarded from either the computation of the sum-scores, the fitting of the logistic models or the IRT models (according to the method).

The vector of group membership must hold only two different values, either as numeric or character. The focal group is defined by the argument focal.name.

For "MH", "Std", "Logistic" and "BD" methods, the matching criterion can be either the test score or any other continuous or discrete variable to be passed in the selected DIF function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

For Lord and Raju methods, one can specify either the IRT model to be fitted (by means of model, c, engine and discr arguments), or the item parameter estimates with arguments irtParam and same.scale. See difLord and difRaju for further details.

The threshold for detecting DIF items depends on the method. For standardization it has to be fully specified (with the thr argument), as well as for the TID method (through the thrTID argument). For the other methods it is depending on the significance level set by alpha.

For Mantel-Haenszel method, the DIF statistic can be either the Mantel-Haenszel chi-square statistic or the log odds-ratio statistic. The method is specified by the argument MHstat, and the default value is "MHChisq" for the chi-square statistic. Moreover, the option correct specifies whether the continuity correction has to be applied to Mantel-Haenszel statistic. See difMH for further details.

By default, the asymptotic Mantel-Haenszel statistic is computed. However, the exact statistics and related P-values can be obtained by specifying the logical argument exact to TRUE. See Agresti (1990, 1992) for further details about exact inference.

The weights for computing the standardized P-DIF statistics are defined through the argument stdWeight, with possible values "focal" (default value), "reference" and "total". See stdPDIF for further details.

For Breslow-Day method, two test statistics are available: the usual Breslow-Day statistic for testing homogeneous association (Aguerri, Galibert, Attorresi and Maranon, 2009) and the trend test statistic for assessing some monotonic trend in the odds ratios (Penfield, 2003). The DIF statistic is supplied by the BDstat argument, with values "BD" (default) for the usual statistic and "trend" for the trend test statistic.

The SIBTEST method (Shealy and Stout, 1993) and its modified version, the Crossing-SIBTEST (Chalmers, 2018; Li and Stout, 1996) are returned by the difSIBTEST function. SIBTEST method is returned when type argument is set to "udif", while Crossing-SIBTEST is set with "nudif" value for the type argument. Note that type takes the by-default value "both" which is not allowed within the difSIBTEST function; however, within this fucntion, keeping the by-default value yields selection of Crossing-SIBTEST.

The difSIBTEST function is a wrapper to the SIBTEST function from the **mirt** package (Chalmers, 2012) to fit within the difR framework (Magis et al., 2010). Therefore, if you are using this function for publication purposes please cite Chalmers (2018; 2012) and Magis et al. (2010).

For logistic regression, the argument type permits to test either both uniform and nonuniform effects simultaneously (type="both"), only uniform DIF effect (type="udif") or only nonuniform DIF effect (type="nudif"). The criterion argument specifies the DIF statistic to be computed, either the likelihood ratio test statistic (with criterion="LRT") or the Wald test (with criterion="Wald"). Moreover, the group membership can be either a vector of two distinct values, one for the reference group and one for the focal group, or a continuous or discrete variable that acts as the "group" membership variable. In the former case, the member.type argument is set to "group" and the focal.name defines which value in the group variable stands for the focal group. In the latter case, member.type is set to "cont", focal.name is ignored and each value of the group represents one "group" of data (that is, the DIF effects are investigated among participants relying on different values of some discrete or continuous trait). See Logistik for further details.

For Raju's method, the type of area (signed or unsigned) is fixed by the logical signed argument, with default value FALSE (i.e. unsigned areas). See RajuZ for further details.

Item purification can be requested by specifying purify option to TRUE. Recall that item purification is slightly different for IRT and for non-IRT based methods. See the corresponding methods for further information.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. See the corresponding methods for further information.

A pre-specified set of anchor items can be provided through the anchor argument. For non-IRT methods, anchor items are used to compute the test score (as matching criterion). For IRT methods, anchor items are used to rescale the item parameters on a common metric. See the corresponding methods for further information. Note that anchor argument is not working with "LRT" method.

The output of the selected method can be stored in a text file by fixing save.output and output appropriately. See the help file of the corresponding method for further information.

Value

The output of the selected DIF detection method.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7*, 131-177. doi:10.1214/ss/1177011454

Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (2009). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity, 43*, 35-44. doi:10.1007/s1113500791302

Angoff, W. H., and Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 2*, 95-106. doi:10.1111/j.17453984.1973.tb00787.x

Chalmers, R. P. (2012). mirt: A Multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48(6)*, 1-29. doi:10.18637/jss.v048.i06

Chalmers, R. P. (2018). Improving the Crossing-SIBTEST statistic for detecting non-uniform DIF. *Psychometrika, 83*(2), 376–386. doi:10.1007/s1133601795838

Dorans, N. J. and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368. doi:10.1111/j.17453984.1986.tb00255.x

Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Dirs.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Li, H.-H., and Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*, 647–677. doi:10.1007/BF02294041

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Penfield, R.D. (2003). Application of the Breslow-Day test of trend in odds ratio heterogeneity to the detection of nonuniform DIF. *Alberta Journal of Educational Research, 49*, 231-243.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. doi:10.1177/014662169001400208

Shealy, R. and Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194. doi:10.1007/BF02294572

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370. doi:10.1111/j.1745-3984.1990.tb00754.x

Thissen, D., Steinberg, L. and Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer and H. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

See Also

`difTID`, `difMH`, `difStd`, `difBD`, `difLogistic`, `difSIBTEST`, `difLord`, `difRaju`, `difLRT`, `dichoDif`

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Excluding the "Anger" variable
 verbal <- verbal[colnames(verbal)!="Anger"]
```

```
# Calling Mantel-Haenszel
selectDif(verbal, group = 25, focal.name = 1, method = "MH")

# Calling Mantel-Haenszel and saving output in 'MH.txt' file
selectDif(verbal, group = 25, focal.name = 1, method = "MH",
   save.output = TRUE, output = c("MH", "default"))

# Calling Lord method
# 2PL model, with item purification
selectDif(verbal, group = 25, focal.name = 1, method = "Lord", model = "2PL",
          purify = TRUE)

## End(Not run)
```

---

selectGenDif *Selection of one of the DIF detection methods among multiple groups*

Description

This function performs DIF detection among multiple groups for one pre-specified method. This function can only be used with dichotomous items.

Usage

```
selectGenDif(Data, group, focal.names, method, anchor = NULL, match = "score",
  type = "both", criterion = "LRT", alpha = 0.05, model = "2PL", c = NULL,
  engine = "ltm", discr = 1, irtParam = NULL, nrFocal = 2, same.scale = TRUE,
  purify = FALSE, nrIter = 10, p.adjust.method = NULL, save.output = FALSE,
  output = c("out", "default"))
```

Arguments

Data	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See **Details**.
group	numeric or character: either the vector of group membership or the column indicator (within data) of group membership. See **Details**.
focal.names	numeric or character vector indicating the levels of group which correspond to the focal groups.
method	character: the name of the selected method. See **Details**.
anchor	either NULL (default) or a vector of item names (or identifiers) to specify the anchor items. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of Data. See **Details**.

`type`	a character string specifying which DIF effects must be tested. Possible values are `"both"` (default), `"udif"` and `"nudif"`. See **Details**.
`criterion`	character: the type of test statistic used to detect DIF items with generalized logistic regression. Possible values are `"LRT"` (default) and `"Wald"`. See **Details**.
`alpha`	numeric: significance level (default is 0.05).
`model`	character: the IRT model to be fitted (either `"1PL"`, `"2PL"` or `"3PL"`). Default is `"2PL"`.
`c`	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See **Details**.
`engine`	character: the engine for estimating the 1PL model, either `"ltm"` (default) or `"lme4"`.
`discr`	either `NULL` or a real positive value for the common discrimination parameter (default is 1). Used onlky if `model` is `"1PL"` and `engine` is `"ltm"`. See **Details**.
`irtParam`	matrix with *2J* rows (where *J* is the number of items) and at most 9 columns containing item parameters estimates. See **Details**.
`nrFocal`	numeric: the number of focal groups (default is 2).
`same.scale`	logical: are the item parameters of the `irtParam` matrix on the same scale? (default is "TRUE"). See **Details**.
`purify`	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
`nrIter`	numeric: the maximal number of iterations in the item purification process (default is 10).
`p.adjust.method`	either `NULL` (default) or the acronym of the method for p-value adjustment for multiple comparisons. See **Details**.
`save.output`	logical: should the output be saved into a text file? (Default is `FALSE`).
`output`	character: a vector of two components. The first component is the name of the output file, the second component is either the file path or `"default"` (default value). See **Details**.

Details

This is a generic function which calls one of the DIF detection methods for multiple groups, and displays its output. It is mainly used as a routine for `genDichoDif` command.

There are three possible methods currently implemented: `"GMH"` for Generalized Mantel-Haenszel (Penfield, 2001), `"genLogistic"` for generalized logistic regression (Magis, Raiche, Beland and Gerard, 2010) and `"genLord"` for generalized Lord's chi-square test (Kim, Cohen and Park, 1995).

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. In addition, `Data` can hold the vector of group membership. If so, `group` indicates the column of `Data` which corresponds to the group membership, either by specifying its name or by giving the column number. Otherwise, `group` must be a vector of same length as `nrow(Data)`.

Missing values are allowed for item responses (not for group membership) but must be coded as `NA` values. They are discarded from either the computation of the sum-scores, the fitting of the logistic models or the IRT models (according to the method).

The vector of group membership must hold at least three different values, either as numeric or character. The focal groups are defined by the values of the argument focal.names.

For "GMH" and "genLogistic" methods, the matching criterion can be either the test score or any other continuous or discrete variable to be passed in the selected DIF function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix.

For the generalized logistic regression method, the argument type permits to test either both uniform and nonuniform effects simultaneously (with type="both"), only uniform DIF effect (with type="udif") or only nonuniform DIF effect (with type="nudif"). Furthermore, the argument criterion defines which test must be used, either the Wald test ("Wald") or the likelihood ratio test ("LRT").

For generalized Lord method, one can specify either the IRT model to be fitted (by means of model, c, engine and discr arguments), or the item parameter estimates with arguments irtParam, nrFocal and same.scale. Moreover, the matching criterion can be either the test score or any other continuous or discrete variable to be passed in the Logistik function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the Data matrix. See difGenLord for further details.

The threshold for detecting DIF items depends on the method and is depending on the significance level set by alpha.

Item purification can be requested by specifying purify option to TRUE. Recall that item purification is slightly different for IRT and for non-IRT based methods. See the corresponding methods for further information.

Adjustment for multiple comparisons is possible with the argument p.adjust.method. See the corresponding methods for further information.

A pre-specified set of anchor items can be provided through the anchor argument. For non-IRT methods, anchor items are used to compute the test score (as matching criterion). For IRT methods, anchor items are used to rescale the item parameters on a common metric. See the corresponding methods for further information.

The output of the selected method can be stored in a text file by fixing save.output and output appropriately. See the help file of the corresponding method for further information.

Value

The output of the selected DIF detection method.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education

Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Kim, S.-H., Cohen, A.S. and Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276. doi:10.1111/j.17453984.1995.tb00466.x

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education, 14*, 235-259. doi:10.1207/S15324818AME1403_3

See Also

difGMH, difGenLogistic, difGenLord

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender ("Man" or "Woman") and trait
 # anger score ("Low" or "High")
 group <- rep("WomanLow", nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)

 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Calling generalized Mantel-Haenszel
 selectGenDif(Verbal, group = 25, focal.names = names, method = "GMH")

 # Calling generalized Mantel-Haenszel and saving output in 'GMH.txt' file
```

```
selectGenDif(Verbal, group = 25, focal.name = names, method = "GMH",
             save.output = TRUE, output = c("GMH", "default"))

# Calling generalized logistic regression
selectGenDif(Verbal, group = 25, focal.names = names, method = "genLogistic")

# Calling generalized Lord method (2PL model)
selectGenDif(Verbal, group = 25, focal.names = names, method = "genLord",
             model = "2PL")

## End(Not run)
```

---

sibTest *SIBTEST DIF statistic*

Description

Calculates the SIBTEST statistics for DIF detection.

Usage

```
sibTest(data, member, anchor = 1:ncol(data), type = "udif")
```

Arguments

data	numeric: the data matrix (one row per subject, one column per item).
member	numeric or factor: the vector of group membership. Can either take two distinct values (zero for the reference group and one for the focal group) or be a continuous vector. See **Details**.
anchor	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.
type	a character string specifying which DIF effects must be tested. Possible values are "udif" (default) and "nudif". See **Details**.

Details

This command computes the SIBTEST Beta coefficients and relatif DIF statistics, both for uniform (Shealy and Stout, 1993) and nonuniform (or crossing-SIBTEST; Chalmers, 2018) DIF effects. It forms the basic command of difSIBTEST function and is specifically designed for this call. This function provides a wrapper to the SIBTEST function from the **mirt** package (Chalmers, 2012) to fit within the difR framework (Magis et al., 2010). Therefore, if you are using this function for publication purposes please cite Chalmers (2018; 2012).

The data are passed through the data argument, with one row per subject and one column per item.

The vector of group membership, specified with member argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

Option anchor sets the items which are considered as anchor items for computing the test scores and related SIBTEST DIF statistics. anchor must hold integer values specifying the column numbers of the corresponding anchor items. If all columns of data are specified as anchor items, then all items are tested for DIF with the all-other-items-as-anchor strategy. If a smaller set of items is defined as the anchor set, then only items outside the anchor set will be tested for DIF; items belonging to this anchor set are not tested and corresponding NA values are returned instead. It is mainly designed to perform item purification.

The output contains: the SIBTEST Beta statistics and related standard errors; the X2 statistics that follow an asymptotic chi-square distribution; the degrees of freedom and the corresponding p-values. The default type value is also returned.

Value

A list with six components:

Beta	the values of the Beta SIBTEST statistics.
SE	the standard errors of Beta values.
X2	the values of X^2 statistics for SIBTEST method.
df	the degrees of freedom for each X2 statistic.
p.value	the p-values of the SIBTEST statistics.
type	the value of the type argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium

References

Chalmers, R. P. (2012). mirt: A Multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48(6)*, 1-29. doi:10.18637/jss.v048.i06

Chalmers, R. P. (2018). Improving the Crossing-SIBTEST statistic for detecting non-uniform DIF. *Psychometrika, 83*(2), 376–386. doi:10.1007/s1133601795838

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Shealy, R. and Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detect test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194. doi:10.1007/BF02294572

See Also

difSIBTEST, dichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)

 # Testing uniform DIF with all items
 sibTest(verbal[,1:24], verbal[,26])

 # Testing nonuniform DIF with all items
 sibTest(verbal[,1:24], verbal[,26], type = "nudif")

 # Removing item 6 from the set of anchor items
 sibTest(verbal[,1:24], verbal[,26], anchor = c(1:5, 7:24))

 # Considering items 3 to 9 as the set of anchor items
 sibTest(verbal[,1:24], verbal[,26], anchor = 3:9)


## End(Not run)
```

`SimDichoDif` *Generation of DIF for dichotomous items*

Description

Function to generate DIF for dichotomous items using the 2PL model.

Usage

```
SimDichoDif(It, ItDIFa, ItDIFb, NR, NF,
            a = rep(1, It), b,
            Ga = rep(0, length(ItDIFa)), Gb = rep(0, length(ItDIFb)),
            D = 1, thR = NULL, thF = NULL,
            muR = 0, muF = 0, sigR = 1, sigF = 1)
```

Arguments

It	It: Number of items
ItDIFa	Vector of integers specifying which items have DIF for a parameters.
ItDIFb	Vector of integers specifying which items have DIF for b parameters.
NR	Number of respondents for reference group.
NF	Number of respondents for focal group (generalize to multiple focal groups).
a	Item slope for reference group.
b	Item difficulty for reference group.

Gb	Vector of difference in b's for focal group(s).
Ga	Vector of difference in a's for focal group(s).
D	Scaling parameter for 2PL. Defaults to 1.
thR	Optional vector of latent variable values for reference group.
thF	Optional vector of latent variable values for focal group.
muR	Mean of latent variable for reference group. Used if latent scores not supplied.
muF	Mean of latent variable for reference group. Used if latent scores not supplied.
sigR	Standard deviation of latent variable for reference group. Used if latent scores not supplied.
sigF	Standard deviation of latent variable for reference group. Used if latent scores not supplied.

Details

This function is based on the 2PL model to test uniform, non-uniform of both DIF. To use the Rasch model, please restrict *a* parameter to 1.

Value

A list with several arguments:

data	the matrix with DIF items.
ipars	the item parameters.
thetas	the person parameters.

Author(s)

Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Carl F. Falk
Department of Psychology
McGill University (Canada)
<carl.falk@mcgill.ca>, https://www.mcgill.ca/psychology/carl-f-falk

References

Berger, M., & Tutz, G. (2016). Detection of Uniform and Nonuniform Differential Item Functioning by Item-Focused Trees. *Journal of Educational and Behavioral Statistics, 41*(6), 559–592. https://doi.org/10.3102/1076998616659371

Examples

```
## Not run:

# test to generate UDIF

It   <- 15 # number of items
ItDIFa <- NULL
ItDIFb <- c(1,3)
NR   <- 100 # number of responses for group 1 (reference)
NF   <- 100 # number of responses for group 2 (focal)
a    <- rep(1,It)          # for tests: runif(It,0.2,.5)
b    <- rnorm(It,1,.5)
Gb   <- rep(2,2)           # Group value for U-DIF
Ga   <- 0                  # Group value for NU-DIF: need to be fix to 0 for U-DIF
#Type <- "UDIF"
#seed <- 1

Out1 <- SimDichoDif(It,ItDIFa,ItDIFb,NR,NF,a,b,Ga,Gb)
Out1
Out1$ipars

# Test to generate NUDIF

It   <- 15                # Nb of items with DIF
ItDIFa <- c(1,3)
ItDIFb <- c(1,3)
NR   <- 100              # N for Ref.
NF   <- 100              # N for Focal
a    <- rep(1,It)        # For Rasch or any value for 1PL
b    <- rnorm(It,1,.5)   # Item difficulties from random normal
Gb   <- rep(.8,2)        # Group value for U-DIF
Ga   <- rep(1.2,2)       # Group value for NU-DIF
#Type <- "NUDIF"
#seed <- 1

Out2 <- SimDichoDif(It,ItDIFa,ItDIFb,NR,NF,a,b,Ga,Gb)
Out2
Out2$ipars

# Generates a mix of UDIF and NUDIF

It   <- 15                # Nb of items with DIF
ItDIFa <- c(1)
ItDIFb <- c(1,3)
NR   <- 100              # N for Ref.
NF   <- 100              # N for Focal
a    <- rep(1,It)        # For Rasch or any value for 1PL
b    <- rnorm(It,1,.5)   # Item difficulties from random normal
Gb   <- rep(.8,2)        # Group value for U-DIF
Ga   <- 1.2              # Group value for NU-DIF
#Type <- "NUDIF"
#seed <- 1
```

```
Out3 <- SimDichoDif(It,ItDIFa,ItDIFb,NR,NF,a,b,Ga,Gb)
Out3
Out3$ipars


## End(Not run)
```

SimPolyDif *Generation of DIF for polytomous items*

Description

Function to generate DIF for polytomous items using the GPCM.

Usage

```
SimPolyDif(It, ItDIFa, ItDIFb,
                      NR, NF, a, b, d, ncat=3,
                      Ga=rep(0,ItDIFa), Gb=rep(0,ItDIFb),
                      D=1,
                      thR=NULL,thF=NULL,muR=0,muF=0,sigR=1,sigF=1,
                     ItDIFd=NULL, Gd = lapply(1:It, function(x){rep(0,ncat)}))
```

Arguments

It	It: Number of items
ItDIFa	Vector of integers specifying which items have DIF for a parameters.
ItDIFb	Vector of integers specifying which items have DIF for b parameters.
NR	Number of respondents for reference group.
NF	Number of respondents for focal group (generalize to multiple focal groups).
a	Item slope for reference group.
b	Item difficulty for reference group.
d	Step parameters, as a list whose length is the same as the number of items, for the reference group.
ncat	Number of categories per item. Currently the same number for all items.
Gb	Vector of difference in b's for focal group(s).
Ga	Vector of difference in a's for focal group(s).
D	Scaling parameter for GPCM. Defaults to 1.
thR	Optional vector of latent variable values for reference group.
thF	Optional vector of latent variable values for focal group.

muR	Mean of latent variable for reference group. Used if latent scores not supplied.
muF	Mean of latent variable for reference group. Used if latent scores not supplied.
sigR	Standard deviation of latent variable for reference group. Used if latent scores not supplied.
sigF	Standard deviation of latent variable for reference group. Used if latent scores not supplied.
ItDIFd	Vector of integers specifying which items have DIF for step parameters.
Gd	List of differences in d's for focal group(s).

Details

This function is based on traditional parameterizations of the GPCM that have an overall difficulty parameter and step parameters.

Value

A list with several arguments:

data	the matrix with DIF items.
ipars	the item parameters.
thetas	the person parameters.

Author(s)

Carl F. Falk
Department of Psychology
McGill University (Canada)
<carl.falk@mcgill.ca>, https://www.mcgill.ca/psychology/carl-f-falk
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>

References

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Examples

```
## Not run:
set.seed(1234)

# original item parameters
a <- rlnorm(10, -0.5)  # slopes
b <- runif(10, -2, 2)  # difficulty
d <- list()
```

```
d[[1]] <- c(0, 2, .5, -.15, -1.1)
d[[2]] <- c(0, 2, .25, -.45, -.75)
d[[3]] <- c(0, 1, .5, -.65, -1)
d[[4]] <- c(0, 2, .5, -.85, -2)
d[[5]] <- c(0, 1, .25, -.05, -1)
d[[6]] <- c(0, 2, .5, -.95, -1)
d[[7]] <- c(0, 1, .25, -.35, -2)
d[[8]] <- c(0, 2, .5, -.15, -1)
d[[9]] <- c(0, 1, .25, -.25, -2)
d[[10]] <- c(0, 2, .5, -.35, -1)

# Uniform DIF
It <- 10; NR <- 1000; NF <- 1000
ItDIFa <- NULL; Ga <- NULL
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2)

Out.Unif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                       ncat = 5, Ga = Ga, Gb = Gb)
Out.Unif$ipars
Data <- Out.Unif$data
difPolyLogistic(as.data.frame(Data[, 1:It]),
                group = Data[, It + 1], focal.name = "G2")

# Nonuniform DIF
ItDIFa <- c(1, 2)
Ga <- rep(.25, 2)
ItDIFb <- c(1, 3)
Gb <- rep(.5, 2)

Out.NUnif <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                        ncat = 5, Ga = Ga, Gb = Gb)
Out.NUnif$ipars
Data <- Out.NUnif$data
difPolyLogistic(as.data.frame(Data[, 1:It]),
                group = Data[, It + 1], focal.name = "G2")

# Also changing step parameters
ItDIFd <- c(2)
Gd <- list(c(0, .25, -.25, .25, -.25))

Out.NUnif2 <- SimPolyDif(It, ItDIFa, ItDIFb, NR, NF, a, b, d,
                         ncat = 5, Ga = Ga, Gb = Gb,
                         ItDIFd = ItDIFd, Gd = Gd)
Out.NUnif2$ipars
Data <- Out.NUnif2$data
difPolyLogistic(as.data.frame(Data[, 1:It]),
                group = Data[, It + 1], focal.name = "G2")


## End(Not run)
```

stdPDIF *Standardization DIF statistic*

Description

Calculates standardized P-difference statistics for DIF detection.

Usage

```
stdPDIF(data, member, match = "score", anchor = 1:ncol(data), stdWeight = "focal")
```

Arguments

data	numeric: the data matrix (one row per subject, one column per item).
member	numeric: the vector of group membership with zero and one entries only. See **Details**.
match	specifies the type of matching criterion. Can be either "score" (default) to compute the test score, or any continuous or discrete variable with the same length as the number of rows of data. See **Details**.
anchor	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See **Details**.
stdWeight	character: the type of weights used for the standardized P-DIF statistic. Possible values are "focal" (default), "reference" and "total". See **Details**.

Details

This command computes the standardized P-DIF statistic in the specific framework of differential item functioning (Dorans and Kulick, 1986). It forms the basic command of difStd and is specifically designed for this call. In addition, the standardized alpha values (Dorans, 1989) are also computed as a basis for effect size calculation.

The standardized P-DIF statistic is a weighted average of the difference in proportions of successes in the reference group and in the focal group. The average is computed across the test score strata. The weights can be of three kinds (Dorans, 1989; Dorans and Kulick, 1986) and are specified through the stdWeight argument: the proportion of focal groups examinees within each stratum (stdWeight="focal"), the proportion of reference group examinees within each stratum (stdWeight="reference"), and the proportion of examinees (from both groups) within each stratum (stdWeight="total"). By default, the weights are built from the focal group.

Similarly to the 'alpha' estimates of the common odds ratio for the Mantel-Haenszel method (see mantelHaenszel), the *standardized alpha values* can be computed as rough measures of effect sizes, after a transformation to the Delta Scale (Holland, 1985). See Dorans (1989, p.228, Eqn.15) for further details.

The data are passed through the data argument, with one row per subject and one column per item. Missing values are allowed but must be coded as NA values. They are discarded from sum-score computation.

The vector of group membership, specified with member argument, must hold only zeros and ones, a value of zero corresponding to the reference group and a value of one to the focal group.

The matching criterion can be either the test score or any other continuous or discrete variable to be passed in the stdPDIF function. This is specified by the match argument. By default, it takes the value "score" and the test score (i.e. raw score) is computed. The second option is to assign to match a vector of continuous or discrete numeric values, which acts as the matching criterion. Note that for consistency this vector should not belong to the data matrix.

Option anchor sets the items which are considered as anchor items for computing standardized P-DIF statistics. Items other than the anchor items and the tested item are discarded. anchor must hold integer values specifying the column numbers of the corresponding anchor items. It is mainly designed to perform item purification.

Value

A list with three arguments:

resStd	the vector of the standardized P-DIF statistics.
resAlpha	the vector of standardized alpha values.
match	a character string, either "score" or "matching variable" depending on the match argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning. Standardization and the Mantel-Haenszel method. *Applied Measurement in Education, 2*, 217-233. doi:10.1207/s15324818ame0203_3

Dorans, N. J. and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368. doi:10.1111/j.17453984.1986.tb00255.x

Holland, P. W. (1985, October). *On the study of differential item performance without IRT*. Paper presented at the meeting of Military Testing Association, San Diego (CA).

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

See Also

difStd, dichoDif, mantelHaenszel

Examples

```
 ## Not run:
 # Loading of the verbal data
 data(verbal)

 # All items as anchor items
 stdPDIF(verbal[,1:24], verbal[,26])

 # All items as anchor items, reference group weights
 stdPDIF(verbal[,1:24], verbal[,26], stdWeight = "reference")

 # All items as anchor items, both groups' weights
 stdPDIF(verbal[,1:24], verbal[,26], stdWeight = "total")

 # Removing item 6 from the set of anchor items
 stdPDIF(verbal[,1:24], verbal[,26], anchor = c(1:5,7:24))

## End(Not run)
```

---

subtestLogistic *Testing for DIF among subgroups with generalized logistic regression*

Description

Performs the Wald test to identify DIF items among a subset of groups of examinees, using the results of generalized logistic regression for all groups.

Usage

```
subtestLogistic(x, items, groups, alpha = 0.05)
## S3 method for class 'subLogistic'
print(x, ...)
```

Arguments

x	an object of class "genLogistic", typically the output of the difGenLogistic command.
items	numeric or character: a vector of items to be tested. See **Details**.
groups	numeric or character: a vector of groups of examinees to be compared. See **Details**.
alpha	numeric: the significance level (default is 0.05).
...	other generic parameters for the print function.

Details

This command makes use of the results from the generalized logistic regression to perform subtests between two or more groups of examinees (Magis, Raiche, Beland and Gerard, 2010). The Wald test is used with an appropriate contrast matrix.

The `subtestLogistic` command requires a preliminary output of the generalized logistic regression with all groups of examinees, preferable with the `difGenLogistic` command. The object `x` is an object of class "genLogistic" from which subtests can be performed. The same DIF effect (either uniform, nonuniform, or both types) is tested among the subset of groups of examinees as the one tested with all groups. It is provided b y the argument `type` argument of `x`.

The argument `items` is a vector of the names of the items to be tested, or their number in the data set. A single item can be specified.

The argument `groups` specifies which groups of examinees are considered in this subtest routine. It is a vector of either group names or integer values. In the latter case, the reference group is specified with the 0 (zero) value, while the focal groups are set up by their rank in the `x$focal.names` argument. At least two groups must be specified, and all groups can be included (which leads back to the generalized logistic regression with the Wald test).

The output provides, among others, the Wald statistics, the degrees of freedom and related asymptotic *p*-values for each tested item, as well as the contrast matrix.

Value

A list of class "subLogistic" with the following components:

`stats`	a table with as many rows as tested items, and four columns: the item number, the Wald statistic, the degrees of freedom and the asymptotic *p*-value.
`contrastMatrix`	the contrast matrix used for testing DIF among the groups set up by `groups`.
`items`	the value of the `items` argument.
`groups`	the value of the `groups` argument.
`type`	the value of the `x$type` argument.
`purification`	the value of the `x$purification` argument.
`alpha`	the value of the `alpha` argument.

Author(s)

David Magis
Data science consultant at IQVIA Belux
Brussels, Belgium
Sebastien Beland
Faculte des sciences de l'education
Universite de Montreal (Canada)
<sebastien.beland@umontreal.ca>
Gilles Raiche
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>

References

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Magis, D., Raiche, G., Beland, S. and Gerard, P. (2011). A logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing, 11*, 365–386. doi:10.1080/15305058.2011.602810

See Also

difGenLogistic, genDichoDif

Examples

```
## Not run:

 # Loading of the verbal data
 data(verbal)
 attach(verbal)

 # Creating four groups according to gender (0 or 1) and trait anger score
 # ("Low" or "High")
 # Reference group: women with low trait anger score (<=20)
 group <- rep("WomanLow",nrow(verbal))
 group[Anger>20 & Gender==0] <- "WomanHigh"
 group[Anger<=20 & Gender==1] <- "ManLow"
 group[Anger>20 & Gender==1] <- "ManHigh"

 # New data set
 Verbal <- cbind(verbal[,1:24], group)

 # Reference group: "WomanLow"
 names <- c("WomanHigh", "ManLow", "ManHigh")

 # Testing all types of DIF with all items
 rDIF <- difGenLogistic(Verbal, group = 25, focal.names = names)
 rUDIF <- difGenLogistic(Verbal, group = 25, focal.names = names, type = "udif")
 rNUDIF <- difGenLogistic(Verbal, group = 25, focal.names = names, type = "nudif")

 # Subtests between the reference group and the first two focal groups
 # for item "S2WantShout" (item 6) and the three types of DIF
 subGroups <- c("WomanLow", "WomanHigh", "ManLow")
 subtestLogistic(rDIF, items = 6, groups = subGroups)
 subtestLogistic(rUDIF, items = 6, groups = subGroups)
 subtestLogistic(rNUDIF, items = 6, groups = subGroups)

 # Subtests between the reference group and the first focal group
 # for items "S2WantShout" (item 6) and "S3WantCurse" (item 7)
 # (only both DIF effects)
 subGroups <- c("WomanLow", "WomanHigh")
 items1 <- c("S2WantShout", "S3WantCurse")
```

```
items2 <- 6:7
subtestLogistic(rDIF, items = items1, groups = subGroups)
subtestLogistic(rDIF, items = items2, groups = subGroups)

## End(Not run)
```

---

verbal *Verbal Aggression Data Set*

Description

The Verbal Aggression data set comes from Vansteelandt (2000) and is made of the responses of 316 subjects (243 women and 73 men) to a questionnaire of 24 items, about verbal aggression. All items describe a frustrating situation together with a verbal aggression response. A correct answer responses is coded as 0 and 1, a value of one meaning that the subject would (want to) respond to the frustrating situation in an aggressive way. In addition, the *Trait Anger* score (Spielberger, 1988) was computed for each subject.

Format

The `verbal` matrix consists of 316 rows (one per subject) and 26 columns.

The first 24 columns hold the responses to the dichotomously scored items. The 25th column holds the trait anger score for each subject. The 26th column is vector of the group membership; values 0 and 1 refer to women and men, respectively.

Each item name starts with `S` followed by a value between 1 and 4, referring to one of the situations below:

S1: A bus fails to stop for me.

S2: I miss a train because a clerk gave me faulty information.

S3: The grocery store closes just as I am about to enter.

S4: The operator disconnects me when I had used up my last 10 cents for a call.

The second part of the name is either *Want* or *Do*, and indicates whether the subject wanted to respond to the situation or actually did respond.

The third part of the name is one of the possible aggressive responses, either *Curse*, *Scold* or *Shout*.

For example, item `S1WantShout` refers to the sentence: "a bus fails to stop for me. I want to shout". The corresponding item response is 1 if the subject agrees with that sentence, and 0 if not.

Source

The Verbal aggression data set is taken originally from Vansteelandt (2000) and has been used as an illustrative example in De Boeck (2008), De Boeck and Wilson (2004) and Smits, De Boeck and Vansteelandt (2004), among others. The following URL *http://bear.soe.berkely.edu/EIRM/* permits to get access to the full data set.

References

De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73*, 533-559. doi:10.1007/s11336-0089092x

De Boeck, P. and Wilson, M. (2004). *Explanatory item response models: a generalized linear and nonlinear approach*. New York: Springer. doi:10.1007/9781475739909

Magis, D., Beland, S., Tuerlinckx, F. and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi:10.3758/BRM.42.3.847

Smits, D., De Boeck, P. and Vansteelandt, K. (2004). The inhibition of verbal aggressive behavior. *European Journal of Personality, 18*, 537-555. doi:10.1002/per.529

Spielberger, C.D. (1988). *State-trait anger expression inventory research edition. Professional manual*. Odessa, FL: Psychological Assessment Resources.

Vansteelandt, K. (2000). *Formal models for contextualized personality psychology*. Unpublished doctoral dissertation, K.U. Leuven, Belgium.

Index