

Package ‘difR’

January 13, 2010

Type Package

Title Collection of methods to detect dichotomous differential item functioning (DIF) in psychometrics

Version 2.1

Date 2010-01-13

Author Sebastien Beland (UQAM), David Magis (U Liege) and Gilles Raiche (UQAM)

Maintainer David Magis <david.magis@ulg.ac.be>

Depends R (>= 2.8.0), lme4, ltm

Description The difR package contains several traditional methods to detect DIF in dichotomously scored items. Both uniform and non-uniform DIF effects can be detected, with methods relying upon item response models or not. Some methods deal with more than one focal group.

License file LICENSE

LazyLoad yes

R topics documented:

breslowDay	2
contrastMatrix	4
dichoDif	5
difBD	9
difGenLogistic	11
difGenLord	15
difGMH	20
difLogistic	23
difLord	27
difLRT	32
difMH	35
difRaju	38
difStd	42
genDichoDif	45
genLogistik	48
genLordChi2	51
genMantelHaenszel	53

itemPar1PL	54
itemPar2PL	56
itemPar3PL	57
itemPar3PLconst	59
itemParEst	60
itemRescale	62
Logistik	64
LordChi2	67
LRT	68
mantelHaenszel	70
RajuZ	72
selectDif	74
selectGenDif	76
stdPDIF	79
verbal	81

Index	82
--------------	-----------

breslowDay	<i>Breslow-Day DIF statistic</i>
------------	----------------------------------

Description

Computes Breslow-Day statistics for DIF detection.

Usage

```
breslowDay(data, member, anchor=1:ncol(data))
```

Arguments

<code>data</code>	numeric: the data matrix (one row per subject, one column per item).
<code>member</code>	numeric: the vector of group membership with zero and one entries only. See Details .
<code>anchor</code>	a vector of integer values specifying which items (all by default) are currently considered as anchor (DIF free) items. See Details .

Details

`breslowDay` computes the Breslow-Day statistic (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 not allowed.

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.

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.

Value

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.

Note

The statistic currently implemented is the Breslow-Day statistic for the generalm test of odds ratio heterogeneity, and not the statistic for the test of trend in heterogeneity as displayed by Penfield (2003).

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

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.

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])

# 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	<i>Contrast matrix for computing generalized Lord's chi-square DIF statistic</i>
----------------	--

Description

This command sets the appropriate contrast matrix *C* for computing the generalized Lord's chi-square 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-square statistic. It is designed to perform accurate tests of equality of item parameters across 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 acronym (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 pre-defined 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)

Sebastien Beland
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<sebastien.beland.1@hotmail.com>
David Magis
Research Group of Quantitative Psychology and Individual Differences
Katholieke Universiteit Leuven
<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Université du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

See Also

[genLordChi2](#), [difGenLord](#)

Examples

```
# 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")
```

dichoDif

Comparison of DIF detection methods

Description

This function compares the specified DIF detection methods with respect to the detected items.

Usage

```
dichoDif(Data, group, focal.name, method, alpha=0.05,
correct=TRUE, thr=0.1, type="both", model="2PL", c=NULL,
engine="ltm", irtParam=NULL, same.scale=TRUE,
purify=FALSE, nrIter=10)
## 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 .

<code>focal.name</code>	numeric or character indicating the level of group which corresponds to the focal group.
<code>method</code>	character vector specifying the different methods to be compared. See Details .
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>correct</code>	logical: should the Mantel-Haenszel continuity correction be used? (default is TRUE).
<code>thr</code>	numeric: the threshold (cut-score) for standardized P-DIF statistic (default is 0.10).
<code>type</code>	a character string specifying which DIF effects must be tested (default is "both"). See Details .
<code>model</code>	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL"). Default is "2PL".
<code>c</code>	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See Details .
<code>engine</code>	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
<code>irtParam</code>	matrix with $2J$ rows (where J is the number of items) and at most 9 columns containing item parameters estimates. See Details .
<code>same.scale</code>	logical: are the item parameters of the <code>irtParam</code> matrix on the same scale? (default is "TRUE"). See Details .
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	result from a <code>dichoDif</code> class object.
<code>...</code>	other generic parameters for the <code>print</code> function.

Details

`dichoDif` is a generic function which calls one or several DIF detection methods and summarize their output. The possible methods are: "MH" for mantel-Haenszel (Holland and Thayer, 1988), "Std" for standardization (Dorans and Kullick, 1986), "Logistic" for logistic regression (Swaminathan and Rogers, 1990), "BD" for Breslow-Day method (Penfield, 2003), "Lord" for Lord's chi-square test (Lord, 1980), "Raju" for Raju's area method (Raju, 1990), and "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. Missing values are not allowed. 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`.

For Lord and Raju methods, one can specify either the IRT model to be fitted (by means of `model`, `c` and `engine` 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 (through the `thr` argument), while for the other methods it is depending on the significance level set by `alpha`.

Option `correct` specifies whether the continuity correction has to be applied to Mantel-Haenszel statistic (see `difMH`).

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"`). See `Logistik` 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.

Value

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

<code>DIF</code>	a character matrix with one row per item and whose columns refer to the different specified detection methods. See Details .
<code>correct</code>	the value of <code>correct</code> option.
<code>alpha</code>	the significance level <code>alpha</code> .
<code>thr</code>	the value of <code>thr</code> option.
<code>model</code>	the value of <code>model</code> option.
<code>c</code>	the value of <code>c</code> option.
<code>irtParam</code>	the value of <code>irtParam</code> option.
<code>same.scale</code>	the value of <code>same.scale</code> option.
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	an integer vector (of length equal to the number of methods) with the number of iterations in the purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	a logical vector (of length equal to the number of methods) indicating whether the iterative purification process converged. Returned only if <code>purify</code> is <code>TRUE</code> .

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

Dorans, N. J. and Kullick, 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.

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

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

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.

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.

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

[difMH](#), [difStd](#), [difBD](#), [difLogistic](#), [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 Mantel-Haenszel, standardization and logistic regression
# Standardization threshold of 0.08, no continuity correction,
# with item purification
dichoDif(verbal, group=25, focal.name=1, method=c("MH", "Std", "Logistic"),
correct=FALSE, thr=0.08, purify=TRUE)

# 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, alpha=0.05, purify=FALSE,
nrIter=10)
## S3 method for class 'BD':
print(x, ...)
## S3 method for class 'BD':
plot(x, pch=8, number=TRUE, col="red", ...)
```

Arguments

<code>Data</code>	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See Details .
<code>group</code>	numeric or character: either the vector of group membership or the column indicator (within <code>Data</code>) of group membership. See Details .
<code>focal.name</code>	numeric or character indicating the level of <code>group</code> which corresponds to the focal group.
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is <code>FALSE</code>).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	the result from a <code>BD</code> class object.
<code>pch, col</code>	type of usual <code>pch</code> and <code>col</code> graphical options.
<code>number</code>	logical: should the item number identification be printed (default is <code>TRUE</code>).
<code>...</code>	other generic parameters for the <code>plot</code> or the <code>print</code> 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. Missing values are not allowed. 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 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 the degrees of freedom depend on the number of partial tables taken into account (Aguerri *et al.*, in press; 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.

Value

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

<code>BD</code>	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.
<code>alpha</code>	the significance level for DIF detection.
<code>DIFitems</code>	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>difPur</code>	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 <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIter</code> of allowed iterations. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>names</code>	the names of the items.

Note

The current test of Breslow-Day makes use of the statistic for testing odds ratio heterogeneity as given by Aguerri *et al.* (in press). This is not the Breslow-Day test of trend in odds ratio heterogeneity displayed by Penfield (2003).

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Université du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

- Aguerri, M.E., Galibert, M.S., Attorresi, H.F. and Maranon, P.P. (in press). Erroneous detection of nonuniform DIF using the Breslow-Day test in a short test. *Quality and Quantity*.
- 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.
- 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 item purification
difBD(verbal, group="Gender", focal.name=1, purify=TRUE)
difBD(verbal, group="Gender", focal.name=1, purify=TRUE, nrIter=5)

## 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, type="both", alpha=0.05,
purify=FALSE, nrIter=10)
## S3 method for class 'genLogistic':
print(x, ...)
## S3 method for class 'genLogistic':
plot(x, plot="lrStat", item=1, pch=8, number=TRUE,
col="red", colIC=rep("black",length(x$focal.names)+1),
ltyIC=1:(length(x$focal.names)+1), ...)
```

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.
type	a character string specifying which DIF effects must be tested (default is "both"). 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.
x	the result from a <code>Logistik</code> 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 <code>plot="itemCurve"</code> .
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 <code>plot="itemCurve"</code> .
...	other generic parameters for the plot or the print functions.

Details

The generalized logistic regression method (Magis, Raiche and Beland, 2009) 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 test score, 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"`). See [genLogistik](#) for further details.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. Missing values are not allowed. 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 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-square 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.

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 Jodoign and Gierl (2001). The output displays the ΔR^2 measures, together with the two classifications.

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. When one tests for nonuniform DIF (`type="nudif"`), $J + 1$ logistic curves are displayed, either parallel (if the item is not flagged as DIF) or with different slopes (if the item is flagged as DIF). When one tests for both DIF effects (`type="both"`) or only uniform DIF (`type="udif"`), only one logistic curve is drawn if the item is not flagged as DIF; otherwise, $J + 1$ curves are displayed with a corresponding legend (in the upper left corner). 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.

Value

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

<code>genLogistik</code>	the values of the generalized logistic regression statistics.
<code>logitPar</code>	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.
<code>deltaR2</code>	the differences in Nagelkerke's R^2 coefficients. See Details .
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the column indicators for the items which were detected as DIF items, or "No DIF item detected".

type	the value of type 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 of nrItem allowed iterations. Returned only if purify is TRUE.
names	the names of the items.
focal.names	the value of focal.names argument.

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.
- 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.
- Jodoin, M. G. & 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.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.
- Magis, D., Raiche, G. and Beland, S. (2009). A logistic regression procedure to detect differential item functioning among multiple groups. Unpublished manuscript.
- Zumbo, B. D. & 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](#)

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)

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

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

## End(Not run)
```

difGenLord

Generalized Lord's chi-square DIF method

Description

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

Usage

```
difGenLord(Data, group, focal.names, model, c=NULL, engine="ltm",
  irtParam=NULL, nrFocal=2, same.scale=TRUE, alpha=0.05,
  purify=FALSE, nrIter=10)
## 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), ...)
```

Arguments

<code>Data</code>	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See Details .
<code>group</code>	numeric or character: either the vector of group membership or the column indicator (within <code>Data</code>) of group membership. See Details .
<code>focal.names</code>	numeric or character vector indicating the levels of <code>group</code> which correspond to the focal groups.
<code>model</code>	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL").
<code>c</code>	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See Details .
<code>engine</code>	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
<code>irtParam</code>	matrix with $2J$ rows (where J is the number of items) and at most 9 columns containing item parameters estimates. See Details .
<code>nrFocal</code>	numeric: the number of focal groups (default is 2).
<code>same.scale</code>	logical: are the item parameters of the <code>irtParam</code> matrix on the same scale? (default is TRUE). See Details .
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	the result from a <code>GenLord</code> class object.
<code>plot</code>	character: the type of plot, either "lordStat" or "itemCurve". See Details .
<code>item</code>	numeric or character: either the number or the name of the item for which ICC curves are plotted. Used only when <code>plot="itemCurve"</code> .
<code>pch, col</code>	type of usual <code>pch</code> and <code>col</code> graphical options.
<code>number</code>	logical: should the item number identification be printed (default is TRUE).
<code>colIC, ltyIC</code>	vectors of elements of the usual <code>col</code> and <code>lty</code> arguments for ICC curves. Used only when <code>plot="itemCurve"</code> .
<code>...</code>	other generic parameters for the plot or the print functions.

Details

The generalized Lord's chi-square method (Kim, Cohen and Park, 1995), also referred to as Q_j 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-square statistic.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. Missing values are not allowed. 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 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). 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-square 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.

Two types of plots are available. The first one is obtained by setting `plot="lordStat"` and it is the default option. The chi-square 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.

Value

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

<code>genLordChi</code>	the values of the generalized Lord's chi-square statistics.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>df</code>	the degrees of freedom of the asymptotic null distribution of the statistics.
<code>DIFitems</code>	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>difPur</code>	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 <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIterof</code> allowed iterations. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>model</code>	the value of <code>model</code> argument.
<code>c</code>	The value of the <code>c</code> argument.
<code>engine</code>	The value of the <code>engine</code> argument.
<code>itemParInit</code>	the matrix of initial parameter estimates, with the same format as <code>irtParam</code> either provided by the user (through <code>irtParam</code>) or estimated from the data (and displayed after rescaling).
<code>itemParFinal</code>	the matrix of final parameter estimates, with the same format as <code>irtParam</code> , obtained after item purification. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>estPar</code>	a logical value indicating whether the item parameters were estimated (<code>TRUE</code>) or provided by the user (<code>FALSE</code>).
<code>names</code>	the names of the items.
<code>focal.names</code>	the value of the <code>focal.names</code> argument.

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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, "lme4" engine
difGenLord(Verbal, group="group", focal.name=names, model="1PL",
engine="lme4")

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

# 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)

## End(Not run)

```

Description

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

Usage

```
difGMH(Data, group, focal.names, alpha=0.05, purify=FALSE,
        nrIter=10)
## S3 method for class 'GMH':
print(x, ...)
## S3 method for class 'GMH':
plot(x, pch=8, number=TRUE, col="red", ...)
```

Arguments

<code>Data</code>	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See Details .
<code>group</code>	numeric or character: either the vector of group membership or the column indicator (within <code>Data</code>) of group membership. See Details .
<code>focal.names</code>	numeric or character vector indicating the levels of <code>group</code> which correspond to the focal groups.
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	the result from a GMH class object.
<code>pch, col</code>	type of usual <code>pch</code> and <code>col</code> graphical options.
<code>number</code>	logical: should the item number identification be printed (default is TRUE).
<code>...</code>	other generic parameters for the <code>plot</code> or the <code>print</code> 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. Missing values are not allowed. 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 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-square distribution with lower-tail probability of one minus `alpha` and with as many degrees of freedom as the number of focal groups.

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.

Value

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

<code>GMH</code>	the values of the generalized Mantel-Haenszel statistics.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is TRUE.
<code>difPur</code>	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 <code>purify</code> is TRUE.
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIter</code> of allowed iterations. Returned only if <code>purify</code> is TRUE.
<code>names</code>	the names of the items.
<code>focal.names</code>	the value of <code>focal.names</code> argument.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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.

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

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)

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

## End(Not run)
```

difLogistic

Logistic regression DIF method

Description

Performs DIF detection using logistic regression method.

Usage

```
difLogistic(Data, group, focal.name, type="both", alpha=0.05,
purify=FALSE, nrIter=10)
## S3 method for class 'Logistic':
print(x, ...)
## S3 method for class 'Logistic':
plot(x, plot="lrStat", item=1, pch=8, number=TRUE, col="red",
colIC=rep("black",2), ltyIC=c(1,2), ...)
```

Arguments

<code>Data</code>	numeric: either the data matrix only, or the data matrix plus the vector of group membership. See Details .
<code>group</code>	numeric or character: either the vector of group membership or the column indicator (within <code>data</code>) of group membership. See Details .
<code>focal.name</code>	numeric or character indicating the level of <code>group</code> which corresponds to the focal group.
<code>type</code>	a character string specifying which DIF effects must be tested (default is "both"). See Details .
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	the result from a <code>Logistik</code> class object.
<code>plot</code>	character: the type of plot, either "lrStat" or "itemCurve". See Details .
<code>item</code>	numeric or character: either the number or the name of the item for which logistic curves are plotted. Used only when <code>plot="itemCurve"</code> .
<code>pch, col</code>	type of usual <code>pch</code> and <code>col</code> graphical options.
<code>number</code>	logical: should the item number identification be printed (default is TRUE).
<code>colIC, ltyIC</code>	vectors of two elements of the usual <code>col</code> and <code>lty</code> arguments for logistic curves. Used only when <code>plot="itemCurve"</code> .
<code>...</code>	other generic parameters for the <code>plot</code> or the <code>print</code> 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 test score, 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"`). See `Logistik` for further details.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. Missing values are not allowed. 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 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 (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.

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.

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. When one tests for nonuniform DIF (`type="nudif"`), two logistic curves are displayed, either parallel (if the item is not flagged as DIF) or with different slopes (if the item is flagged as DIF). When one tests for both DIF effects (`type="both"`) or only uniform DIF (`type="udif"`), only one logistic curve is drawn if the item is not flagged as DIF; otherwise, two curves are displayed with a corresponding legend (in the upper left corner). 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.

Value

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

<code>Logistik</code>	the values of the logistic regression statistics.
<code>logitPar</code>	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.
<code>deltaR2</code>	the differences in Nagelkerke's R^2 coefficients. See Details .
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the column indicators for the items which were detected as DIF items, or "No DIF item detected".
<code>type</code>	the value of <code>type</code> argument.
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is TRUE.
<code>difPur</code>	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 <code>purify</code> is TRUE.
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number of <code>nrItem</code> allowed iterations. Returned only if <code>purify</code> is TRUE.
<code>names</code>	the names of the items.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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.

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.

Jodoin, M. G. & 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.

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

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.

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. & 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, dichuDif](#)

Examples

```
## Not run:
# Loading of the verbal data
data(verbal)
```

```

# Excluding the "Anger" variable
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)

# 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")

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

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

## End(Not run)

```

difLord

Lord's chi-square DIF method

Description

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

Usage

```

difLord(Data, group, focal.name, model, c=NULL, engine="ltm",
irtParam=NULL, same.scale=TRUE, alpha=0.05, purify=FALSE,
nrIter=10)
## 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), ...)

```

Arguments

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

<code>group</code>	numeric or character: either the vector of group membership or the column indicator (within <code>data</code>) of group membership. See Details .
<code>focal.name</code>	numeric or character indicating the level of <code>group</code> which corresponds to the focal group.
<code>model</code>	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL").
<code>c</code>	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See Details .
<code>engine</code>	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".
<code>irtParam</code>	matrix with $2J$ rows (where J is the number of items) and at most 9 columns containing item parameters estimates. See Details .
<code>same.scale</code>	logical: are the item parameters of the <code>irtParam</code> matrix on the same scale? (default is "TRUE"). See Details .
<code>alpha</code>	numeric: significance level (default is 0.05).
<code>purify</code>	logical: should the method be used iteratively to purify the set of anchor items? (default is FALSE).
<code>nrIter</code>	numeric: the maximal number of iterations in the item purification process. Default is 10.
<code>x</code>	the result from a <code>Lord</code> class object.
<code>plot</code>	character: the type of plot, either "lordStat" or "itemCurve". See Details .
<code>item</code>	numeric or character: either the number or the name of the item for which ICC curves are plotted. Used only when <code>plot="itemCurve"</code> .
<code>pch, col</code>	type of usual <code>pch</code> and <code>col</code> graphical options.
<code>number</code>	logical: should the item number identification be printed (default is TRUE).
<code>colIC, ltyIC</code>	vectors of two elements of the usual <code>col</code> and <code>lty</code> arguments for ICC curves. Used only when <code>plot="itemCurve"</code> .
<code>...</code>	other generic parameters for the <code>plot</code> or the <code>print</code> functions.

Details

Lord's chi-square 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-square statistic.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. Missing values are not allowed. 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`.

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). 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-square 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.

Two types of plots are available. The first one is obtained by setting `plot="lordStat"` and it is the default option. The chi-square 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.

Value

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

<code>LordChi</code>	the values of the Lord's chi-square statistics.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.

nrPur	the number of iterations in the item purification process. Returned only if <code>purify</code> 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 <code>purify</code> is TRUE.
convergence	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIterof</code> allowed iterations. Returned only if <code>purify</code> is TRUE.
model	the value of <code>model</code> argument.
c	The value of the <code>c</code> argument.
engine	The value of the <code>engine</code> argument.
itemParInit	the matrix of initial parameter estimates, with the same format as <code>irtParam</code> either provided by the user (through <code>irtParam</code>) or estimated from the data (and displayed without rescaling).
itemParFinal	the matrix of final parameter estimates, with the same format as <code>irtParam</code> , obtained after item purification. Returned only if <code>purify</code> 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.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/i05/>

See Also

[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)
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")

# 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)

# 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)

## 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)
## S3 method for class 'LRT':
print(x, ...)
## S3 method for class 'LRT':
plot(x, pch=8, number=TRUE, col="red", ...)

```

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.
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).
...	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 not allowed. 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

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

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-square 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.

Value

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

<code>LRT</code>	the values of the likelihood-ratio statistics.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number of allowed iterations (10 by default). Returned only if <code>purify</code> is <code>TRUE</code> .
<code>names</code>	the names of the items.

Note

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

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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>

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

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

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

## End(Not run)
```

difMH

Mantel-Haenszel DIF method

Description

Performs DIF detection using Mantel-Haenszel method.

Usage

```
difMH(Data, group, focal.name , correct=TRUE, alpha=0.05,
       purify=FALSE, nrIter=10)
## S3 method for class 'MH':
print(x, ...)
## S3 method for class 'MH':
plot(x, pch=8, number=TRUE, col="red", ...)
```

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.
correct	logical: should the continuity correction be used? (default is TRUE)
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.
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).
...	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. Missing values are not allowed. 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 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.

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.

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.

Value

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

<code>MH</code>	the values of the Mantel-Haenszel DIF statistics.
<code>alphaMH</code>	the values of the mantel-Haenszel estimates of common odds ratios.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
<code>correct</code>	the value of <code>correct</code> option.
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>difPur</code>	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 <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIter</code> of allowed iterations. Returned only if <code>purify</code> is <code>TRUE</code> .

names the names of the items.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

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.

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, New Jersey: Lawrence Erlbaum Associates.

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.

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.

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.

See Also

[mantelHaenszel](#), [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
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 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)

## 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",
  irtParam=NULL, same.scale=TRUE, alpha=0.05, purify=FALSE,
  nrIter=10)
## S3 method for class 'Raj':
print(x, ...)
## S3 method for class 'Raj':
plot(x, pch=8, number=TRUE, col="red", ...)
```

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".
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 .
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.
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`).
`...` 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.

The `Data` is a matrix whose rows correspond to the subjects and columns to the items. Missing values are not allowed. 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`.

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). 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 normal distribution with lower-tail probability of one minus $\alpha/2$.

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.

Value

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

`RajuZ` the values of 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".
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.
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.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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

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

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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")

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

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

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

# 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)

## 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)

## End(Not run)
```

difStd *Standardization DIF method*

Description

Performs DIF detection using standardization method.

Usage

```
difStd(Data, group, focal.name,thr=0.1, purify=FALSE, nrIter=10)
## S3 method for class 'PDIF':
print(x, ...)
## S3 method for class 'PDIF':
plot(x, pch=8, number=TRUE, col="red", ...)
```

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.
thr	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.
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).
...	other generic parameters for the plot or the print functions.

Details

The method of standardization (Dorans and Kullick, 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. Missing values are not allowed. 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 threshold (or cut-score) for classifying items as DIF has to be set by the user by the argument `thr`. 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.

In addition, the standardized P-DIF statistics are used to measure the effect sizes of the items. 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 $|St - P - DIF| \geq 0.10$. The values of the effect sizes, together with the Dorans, Schmitt and Bleistein (DSB) 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.

Value

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

<code>PDIF</code>	the values of the Standardization statistics.
<code>alpha</code>	the value of <code>alpha</code> argument.
<code>thr</code>	the threshold (cut-score) for DIF detection.
<code>DIFitems</code>	either the column indicators of the items which were detected as DIF items, or "No DIF item detected".
<code>purification</code>	the value of <code>purify</code> option.
<code>nrPur</code>	the number of iterations in the item purification process. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>difPur</code>	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 <code>purify</code> is <code>TRUE</code> .
<code>convergence</code>	logical indicating whether the iterative item purification process stopped before the maximal number <code>nrIter</code> of allowed iterations. Returned only if <code>purify</code> is <code>TRUE</code> .
<code>names</code>	the names of the items.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

Dorans, N. J. and Kullick, 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.

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.

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

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

## 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, type="both",
alpha=0.05, model="2PL", c=NULL, engine = "ltm", irtParam=NULL,
nrFocal=2, same.scale=TRUE, purify=FALSE, nrIter=10)
## 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 .
type	a character string specifying which DIF effects must be tested (default is "both"). 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".
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.
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 and Beland, 2009) 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. Missing values are not allowed. 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 at least three different values, either as numeric or character. The focal groups are defined by the values of the argument focal.names.

For the generalized logistic regression method, 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").

For generalized Lord method, one can specify either the IRT model to be fitted (by means of model, c and engine 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.

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 methodargument.
type	the value of type argument.
model	the value of model option.
c	the value of c option.
irtParam	the value of irtParam option.
same.scale	the value of same.scale option.
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.

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>

David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

Magis, D., Raiche, G. and Beland, S. (2009). A logistic regression procedure to detect differential item functioning among multiple groups. Unpublished manuscript.

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.

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)

## End(Not run)
```

genLogistik *Generalized logistic regression DIF statistic*

Description

Calculates the "generalized logistic regression" likelihood-ratio statistics for DIF detection among multiple groups.

Usage

```
genLogistik(data, member, anchor=1:ncol(data), type="both")
```

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 .
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 (default is "both"). See Details .

Details

This command computes the generalized logistic regression statistic (Magis, Raiche and Beland, 2009) 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:

$$M_0 : \text{logit}(\pi_i) = \alpha + \beta X + \gamma_i + \delta_i X$$

$$M_1 : \text{logit}(\pi_i) = \alpha + \beta X + \gamma_i$$

$$M_2 : \text{logit}(\pi_i) = \alpha + \beta X$$

where π_i is the probability of answering correctly the item in group i ($i = 0, \dots, J$) and X is the sum score. 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 $\{\gamma_i : i = 1, \dots, J\}$ of the focal groups ($g = i$) represents the uniform DIF effect across all groups, and the set of parameters $\{\delta_i : i = 1, \dots, n\}$ is used to model nonuniform DIF effect across all groups. The models are fitted with the `glm` function.

The argument `type` determines the models to be compared by means of the likelihood ratio statistic. The three possible values of `type` are: `type="both"` which tests the hypothesis $H_0 : \gamma_i = \delta_i = 0$ for all i , by comparing models M_0 and M_2 ; `type="nudif"` which tests the hypothesis $H_0 : \delta_i = 0$ for all i , by comparing models M_0 and M_1 ; and `type="udif"` which tests the hypothesis $H_0 : \gamma_i = 0$ for all i , by comparing models M_1 and M_2 (and assuming that $\delta_1 = 0$). 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. 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 not allowed.

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 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.

The output contains: the likelihood ratio statistics computed for each item, and two matrices with the parameter estimates of both models, for each item. In addition, 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 `difLogistic` command; see Gomez-Benito, Dolores Hidalgo and Padilla (2009), Jodoign and Gierl (2001) and Zumbo and Thomas (1997).

Value

A list with four components:

<code>deviance</code>	the values of the generalized logistic regression DIF statistics.
<code>deltaR2</code>	the differences between Nagelkerke's R^2 coefficients of the tested models. See Details .
<code>parM0</code>	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 $\hat{\delta}_i$ ($i = 1, \dots, J$) of the "full" model (M_0 if <code>type="both"</code> or <code>type="nudif"</code> , M_1 if <code>type="udif"</code>).
<code>parM1</code>	the same matrix as <code>parM0</code> but with fitted parameters for the "simpler" model (M_1 if <code>type="nudif"</code> , M_2 if <code>type="both"</code> or <code>type="udif"</code>).

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

Jodoin, M. G. & 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.

Magis, D., Raiche, G. and Beland, S. (2009). A logistic regression procedure to detect differential item functioning among multiple groups. Unpublished manuscript.

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

Zumbo, B. D. & 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)

# Removing item 6 from the set of anchor items
genLogistik(verbal[,1:24], group, anchor=c(1:5,7:24))

# Testing nonuniform DIF effect
genLogistik(verbal[,1:24], group, type="nudif")

# Testing uniform DIF effect
genLogistik(verbal[,1:24], group, type="udif")

## End(Not run)
```

`genLordChi2`*Generalized Lord's chi-square DIF statistic*

Description

Calculates the generalized Lord's chi-square 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-square statistic (Kim, Cohen and Park, 1995), also called the Q_j 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-square DIF statistics.

Author(s)

Sebastien Beland
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<sebastien.beland.1@hotmail.com>
David Magis
Research Group of Quantitative Psychology and Individual Differences
Katholieke Universiteit Leuven
<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
Gilles Raiche
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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-square 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, 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 .
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 not allowed.

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 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)

Sebastien Beland
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<sebastien.beland.1@hotmail.com>
David Magis
Research Group of Quantitative Psychology and Individual Differences
Katholieke Universiteit Leuven
<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
Gilles Raiche
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

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.

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

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 (IPL) model

Description

Fits the Rasch (IPL) model and returns related item parameter estimates.

Usage

```
itemPar1PL(data, engine="ltm")
```

Arguments

data	numeric: the data matrix.
engine	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".

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 not allowed.

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).

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)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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>

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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])

# 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.

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 not allowed.

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.

Author(s)

Sebastien Beland
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<sebastien.beland.1@hotmail.com>
David Magis
Research Group of Quantitative Psychology and Individual Differences
Katholieke Universiteit Leuven
<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
Gilles Raiche
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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 not allowed.

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**.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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	<i>Item parameter estimation for DIF detection using constrained 3PL model</i>
-----------------	--

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 not allowed.

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 .

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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")
```

Arguments

<code>data</code>	numeric: the data matrix.
<code>model</code>	character: the IRT model to be fitted (either "1PL", "2PL" or "3PL").
<code>c</code>	optional numeric value or vector giving the values of the constrained pseudo-guessing parameters. See Details .
<code>engine</code>	character: the engine for estimating the 1PL model, either "ltm" (default) or "lme4".

Details

`itemParEst` permits to get item parameter estimates of some prespecified 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 not allowed.

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). 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**.

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Université du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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>

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17, 1-25. URL: <http://www.jstatsoft.org/v17/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, "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 of 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)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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

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)
```

 Logistik

Logistic regression DIF statistic

Description

Calculates the "logistic regression" likelihood-ratio statistics and effect sizes for DIF detection.

Usage

```
Logistik(data, member, anchor=1:ncol(data), type="both")
```

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 .
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 (default is "both"). See Details .

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.

The three possible models to be fitted are:

$$M_0 : \text{logit}(\pi_g) = \alpha + \beta X + \gamma_g + \delta_g X$$

$$M_1 : \text{logit}(\pi_g) = \alpha + \beta X + \gamma_g$$

$$M_2 : \text{logit}(\pi_g) = \alpha + \beta X$$

where π_g is the probability of answering correctly the item in group g and X is the sum score. 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.

The argument `type` determines the models to be compared by means of the likelihood ratio statistic. The three possible values of `type` are: `type="both"` (default) which tests the hypothesis $H_0 : \gamma_1 = \delta_1 = 0$ by comparing models M_0 and M_2 ; `type="nudif"` which tests the hypothesis $H_0 : \delta_1 = 0$ by comparing models M_0 and M_1 ; and `type="udif"` which tests the hypothesis $H_0 : \gamma_1 = 0$ by comparing models M_1 and M_2 (assuming that $\delta_1 = 0$). 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.

The data are passed through the `data` argument, with one row per subject and one column per item. Missing values are not allowed.

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 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.

The output contains: the likelihood ratio statistics computed for each item, and two matrices with the parameter estimates of both models, for each item. In addition, 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 `difLogistic` command; see Gomez-Benito, Dolores Hidalgo and Padilla (2009), Jodoign and Gierl (2001) and Zumbo and Thomas (1997).

Value

A list with four components:

<code>deviance</code>	the values of the logistic regression DIF statistics.
<code>deltaR2</code>	the differences between Nagelkerke's R^2 coefficients of the tested models. See Details .
<code>parM0</code>	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 <code>type="both"</code> or <code>type="nudif"</code> , M_1 if <code>type="udif"</code>).
<code>parM1</code>	the same matrix as <code>parM0</code> but with fitted parameters for the "simpler" model (M_1 if <code>type="nudif"</code> , M_2 if <code>type="both"</code> or <code>type="udif"</code>).

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.
- Jodoin, M. G. & 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.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.
- Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- Zumbo, B. D. & 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
Logistik(verbal[,1:24], verbal[,26])

# 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")
```

```
## 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.

Author(s)

Sebastien Beland
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<sebastien.beland.1@hotmail.com>
David Magis
Research Group of Quantitative Psychology and Individual Differences
Katholieke Universiteit Leuven
<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
Gilles Raiche
Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
Universite du Quebec a Montreal
<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

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

Calculates Likelihood-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 not allowed.

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)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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>
 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	<i>Mantel-Haenszel DIF statistic</i>
----------------	--------------------------------------

Description

Calculates the Mantel-Haenszel statistics for DIF detection.

Usage

```
mantelHaenszel(data, member, correct=TRUE, 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 .
correct	logical: should the continuity correction be used? (default is TRUE).
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 Mantel-Haenszel (1959) statistic in the specific framework of differential item functioning. It forms the basic command of `difFMH` 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 not allowed.

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.

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

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).

Value

A list with two components:

<code>resMH</code>	the vector of the Mantel-Haenszel DIF statistics.
<code>resAlpha</code>	the vector of the Mantel-Haenszel estimates of the common odds ratios.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

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.

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, New Jersey: Lawrence Erlbaum Associates.

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.

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)

# Removing item 6 from the set of anchor items
mantelHaenszel(verbal[,1:24], verbal[,26], anchor=c(1:5,7:24))

## End(Not run)

```

RajuZ

Raju's area DIF statistic

Description

Calculates the Raju's statistics for DIF detection.

Usage

```
RajuZ(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 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.

Value

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).

Author(s)

Sebastien Beland
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <sebastien.beland.1@hotmail.com>
 David Magis
 Research Group of Quantitative Psychology and Individual Differences
 Katholieke Universiteit Leuven
 <David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>
 Gilles Raiche
 Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)
 Universite du Quebec a Montreal
 <raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 495-502.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197-207.

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)
RajuZ(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)
RajuZ(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)
RajuZ(mR, mF)

## End(Not run)

```

selectDif

Selection of one of the DIF detection methods

Description

This function performs DIF detection for one pre-specified method.

Usage

```

selectDif(Data, group, focal.name, method, alpha=0.05,
correct=TRUE, thr=0.1, type="both", model="2PL", c=NULL,
engine="ltm", irtParam=NULL, same.scale=TRUE,
purify=FALSE, nrIter=10)

```

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. See Details .
alpha	numeric: significance level (default is 0.05).
correct	logical: should the continuity correction be used? (default is TRUE).
thr	numeric: the threshold (cut-score) for standardized P-DIF statistic (default is 0.10).
type	a character string specifying which DIF effects must be tested (default is "both"). 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".
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 .
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.

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: "MH" for mantel-Haenszel (Holland and Thayer, 1988), "Std" for standardization (Dorans and Kullick, 1986), "Logistic" for logistic regression (Swaminathan and Rogers, 1990), "BD" for Breslow-Day method (Penfield, 2003), "Lord" for Lord's chi-square test (Lord, 1980), "Raju" for Raju's area method (Raju, 1990), and "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. Missing values are not allowed. 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 of the argument `focal.name`.

For Lord and Raju methods, one can specify either the IRT model to be fitted (by means of `model`, `c` and `engine` 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), while for the other methods it is depending on the significance level set by `alpha`.

Option `correct` specifies whether the continuity correction has to be applied to Mantel-Haenszel statistic (see `difMH`).

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"`). See `Logistik` 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.

Value

The output of the selected DIF detection method.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

- Dorans, N. J. and Kullick, 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.
- Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. I. Braun (Eds.), *Test validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- 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

[difMH](#), [difStd](#), [difBD](#), [difLogistic](#), [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 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.

Usage

```
selectGenDif(Data, group, focal.names, method, type="both",
alpha=0.05, model="2PL", c=NULL, engine = "ltm", irtParam=NULL,
nrFocal=2, same.scale=TRUE, purify=FALSE, nrIter=10)
```

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 .
type	a character string specifying which DIF effects must be tested (default is "both"). 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".
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.

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 and Beland, 2009) 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. Missing values are not allowed. 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 at least three different values, either as numeric or character. The focal groups are defined by the values of the argument `focal.names`.

For the generalized logistic regression method, 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"`).

For generalized Lord method, one can specify either the IRT model to be fitted (by means of `model`, `c` and `engine` arguments), or the item parameter estimates with arguments `irtParam`, `nrFocal` 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 is slightly different for IRT and for non-IRT based methods. See the corresponding methods for further information.

Value

The output of the selected DIF detection method.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

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.

Magis, D., Raiche, G. and Beland, S. (2009). A logistic regression procedure to detect differential item functioning among multiple groups. Unpublished manuscript.

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.

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 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)

```

stdPDIF

Standardization DIF statistic

Description

Calculates standardized P-difference statistics for DIF detection.

Usage

```
stdPDIF(data, member, 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 .
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 standardization statistic in the specific framework of differential item functioning (Dorans and Kullick, 1986). It forms the basic command of `difStd` 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 not allowed.

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 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 vector with the values of the standardization DIF statistics.

Author(s)

Sebastien Beland

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<sebastien.beland.1@hotmail.com>

David Magis

Research Group of Quantitative Psychology and Individual Differences

Katholieke Universiteit Leuven

<David.Magis@psy.kuleuven.be>, <http://ppw.kuleuven.be/okp/home/>

Gilles Raiche

Centre sur les Applications des Modeles de Reponses aux Items (CAMRI)

Universite du Quebec a Montreal

<raiche.gilles@uqam.ca>, <http://www.er.uqam.ca/nobel/r17165/>

References

Dorans, N. J. and Kullick, 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.

See Also

[difStd](#), [dichoDif](#)

Examples

```
## Not run:
# Loading of the verbal data
data(verbal)

# All items as anchor items
stdPDIF(verbal[,1:24], verbal[,26])

# Removing item 6 from the set of anchor items
stdPDIF(verbal[,1:24], verbal[,26], anchor=c(1:5,7:24))

## End(Not run)
```

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.berkeley.edu/EIRM/> permits to get access to the full data set.

References

- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73, 533-559.
- De Boeck, P. and Wilson, M. (2004). *Explanatory item response models: a generalized linear and nonlinear approach*. New-York: Springer.
- Smits, D., De Boeck, P. and Vansteelandt, K. (2004). The inhibition of verbal aggressive behavior. *European Journal of Personality*, 18, 537-555.
- 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

*Topic datasets

- verbal, 80
- breslowDay, 1, 10
- contrastMatrix, 3
- dichoDif, 3, 5, 10, 26, 30, 34, 37, 40, 44, 66, 67, 69, 71, 73, 74, 76, 80
- difBD, 2, 3, 8, 8, 76
- difGenLogistic, 11, 47, 48, 50, 78
- difGenLord, 4, 15, 46, 47, 51, 54–63, 77, 78
- difGMH, 20, 22, 47, 53, 78
- difLogistic, 8, 12, 23, 49, 64–66, 76
- difLord, 6, 8, 27, 54–63, 66, 67, 75, 76
- difLRT, 8, 31, 68, 69, 76
- difMH, 6, 8, 21, 22, 34, 70, 71, 75, 76
- difRaju, 6, 8, 37, 54–63, 72, 73, 75, 76
- difStd, 8, 41, 76, 79, 80
- genDichoDif, 14, 44, 50, 77
- genLogistik, 12, 14, 47
- genLordChi2, 4, 50
- genMantelHaenszel, 52
- glm, 48, 64
- itemPar1PL, 54, 56, 58, 59, 61, 63
- itemPar2PL, 55, 55, 58, 59, 61, 63
- itemPar3PL, 55, 56, 57, 59, 61, 63
- itemPar3PLconst, 55, 56, 58, 58, 61, 63
- itemParEst, 18, 30, 40, 51, 54–59, 60, 62, 66, 67, 72, 73
- itemRescale, 17, 29, 51, 62, 67, 72, 73
- Logistik, 6, 24, 26, 63, 75
- LordChi2, 66
- LRT, 34, 68
- mantelHaenszel, 37, 70
- plot.BD (*difBD*), 8
- plot.genLogistic (*difGenLogistic*), 11
- plot.GenLord (*difGenLord*), 15
- plot.GMH (*difGMH*), 20
- plot.Logistic (*difLogistic*), 23
- plot.Lord (*difLord*), 27
- plot.LRT (*difLRT*), 31
- plot.MH (*difMH*), 34
- plot.PDIF (*difStd*), 41
- plot.Raj (*difRaju*), 37
- print.BD (*difBD*), 8
- print.dichoDif (*dichoDif*), 5
- print.genDichoDif (*genDichoDif*), 44
- print.genLogistic (*difGenLogistic*), 11
- print.GenLord (*difGenLord*), 15
- print.GMH (*difGMH*), 20
- print.Logistic (*difLogistic*), 23
- print.Lord (*difLord*), 27
- print.LRT (*difLRT*), 31
- print.MH (*difMH*), 34
- print.PDIF (*difStd*), 41
- print.Raj (*difRaju*), 37
- RajuZ, 40, 71
- selectDif, 73
- selectGenDif, 76
- stdPDIF, 44, 79
- verbal, 80