Package 'difNLR'

June 21, 2016

Type Package

Title Detection of Dichotomous Differential Item Functioning (DIF) by Non-Linear Regression Function

Version 0.1.0

Date 2016-05-23

Author Adela Drabinova (CUNI, ICS CAS), Patricia Martinkova (ICS CAS) and Karel Zvara (CUNI)

Maintainer Adela Drabinova <adela.drabinova@gmail.com>

Depends R (>= 3.2.2), ggplot2, gridExtra, methods, stats

Description Detection of differential item functioning among dichotomously scored items with non-linear regression procedure.

License GPL-3

LazyData no

NeedsCompilation no

Repository CRAN

Date/Publication 2016-06-21 08:36:14

R topics documented:

difNLR-package *Detection of Dichotomous Differential Item Functioning (DIF) by Non-Linear Regression Function*

Description

The difNLR package containts DIF detection method based on Non-Linear Regression. Both uniform and non-uniform DIF effects can be detected when considering one focal group.

Details

Package: difNLR
Type: Package
Version: 0.1.0
Date: 2016-05-23
Depends: R (>= 3.2.2), ggplot2, grid
License: GPL-3

Note

This package was supported by grant funded by Czech Science foundation under number GJ15-15856Y.

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Drabinova, A. and Martinkova P. (2016). Detection of Differenctial Item Functioning Based on Non-Linear Regression, Technical Report, V-1229, http://hdl.handle.net/11104/0259498

Kingston, N., Leary, L., and Wightman, L. (1985). An Exploratory Study of the Applicability of Item Response Theory Methods to the Graduate Management Admission Test. ETS Research Report Series, 1985(2) : 1-64.

Swaminathan, H. and Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. Journal of Educational Measurement, 27, 361-370.

`difNLR` *Non-Linear Regression DIF Method*

Description

Performs DIF detection using Non-Linear Regression method.

Usage

```
difNLR(data, group,  type = "both", p.adjust.method = "BH", start,...)

## S3 method for class 'difNLR'
print(x, ...)
## S3 method for class 'difNLR'
plot(x, plot.type = "cc", item = "all", col = c("tomato", "turquoise"),
alpha = .5, shape = 21, size = .8, linetype = c(1, 2), title, ...)
## S3 method for class 'difNLR'
fitted(object, item = "all", ...)
## S3 method for class 'difNLR'
predict(object, item = "all", score, group,...)
```

Arguments

`data`	numeric: binary data matrix. See **Details**.
`group`	numeric: binary vector of group membership. "0" for reference group, "1" for focal group.
`type`	character: type of DIF to be tested (either "both" (default), "udif" or "nudif"). See **Details**.
`p.adjust.method`	character: method for multiple comparison correction. See **Details**.
`start`	numeric: matrix with n rows (where n is the number of items) and at most 5 columns containing initial item parameters estimates. See **Details**.
`x, object`	an object of `"difNLR"` class
`plot.type`	character: type of plot to be plotted (either `"cc"` for characteristic curve, or "stat" for F-test statistics). See **Details**.
`item`	either character (`"all"`), or numeric vector, or single number corresponding to column indicators. See **Details**.
`col`	character: single value, or vector of two values representing colors for plot.
`alpha`	numeric: transparency parameter for plot.
`shape`	integer: shape parameter for plot.
`size`	numeric: line width in plot.

linetype	line type for reference and focal group in plot.
title	string corresponding to title of plot.
score	numeric: standardized total score of subject.
...	not used.

Details

DIF detection procedure based on Non-Linear Regression is the extension of Logistic Regression procedure (Swaminathan and Rogers, 1990).

The data is a matrix whose rows represents examinee answers ("1" correct, "0" incorrect) and columns correspond to the items. The group must be a vector of the same length as nrow(data).

The type corresponds to type of DIF to be tested. Possible values are "both" to detect any DIF (uniform and/or non-uniform), "udif" to detect only uniform DIF or "nudif" to detect only non-uniform DIF.

The start is a matrix with a number of rows equal to number of items. The number of columns correspond to number of parameters in model in alternative hypothesis (5 for values "both" and "nudif" in type, 4 for "udif" in type). If start missing, initial values are calculated by startNLR() function.

The p.adjust.method is a character for p.adjust function from the stats package. Possible values are "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".

The output of the difNLR is displayed by the print.difNLR function.

Two types of plots are available. The first one is obtained by setting plot.type = "cc"(default). The characteristic curve for item specified in item option is plotted. For default option "all" of item, characteristic curves of all converged items are plotted. The drawn curves represent best model.

The second plot is obtained by setting plot.type = "stat". The F-statistics are displayed on the Y axis, for each coverged item. The detection threshold is displayed by a horizontal line and items detected as DIF are printed with the red color. Only parameters size and title are used.

Fitted values are extracted by the fitted.difNLR function for item(s) specified in item argument.

Predicted values are produced by the predict.difNLR function for item(s) specified in item argument. score represents standardized total score of new subject and group argument represents group membership of new subject.

Missing values are not allowed either for responses or for group membership.

Value

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

DIF	either the column indicators of the items which were detected as DIF, or "NONE".
Fval	the values of F-test statistics.
Pval	the p-values by F-test.
df	the degrees of freedom of F-test.
coef	the matrix of estimated item parameters.
group	the vector of group membership.

`data`	the binary data matrix.
`type`	character: type of DIF that was tested.
`conv_fail`	numeric: number of convergence issues.
`conv_fail_which`	the column indicators of the items which did not converge.
`p.adjust.method`	character: method for multiple comparison correction which was applied.

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Drabinova, A. and Martinkova P. (2016). Detection of Differenctial Item Functioning Based on Non-Linear Regression, Technical Report, V-1229, http://hdl.handle.net/11104/0259498

Swaminathan, H. and Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. Journal of Educational Measurement, 27, 361-370.

See Also

`p.adjust`

Examples

```
## Not run:
# loading data based on GMAT
data(GMAT)

data  <- GMAT[, colnames(GMAT) != "group"]
group <- GMAT[, "group"]

# Testing both DIF effects simultaneously
x <- difNLR(data, group)

# Testing both DIF effects with none multiple comparison correction
difNLR(data, group, type = "both", p.adjust.method = "none")
```

```
# Testing uniform DIF effects
difNLR(data, group, type = "udif")

# Testing non-uniform DIF effects
difNLR(data, group, type = "nudif")

# Graphical devices
plot(x)
plot(x, item = x$DIF)
plot(x, plot.type = "stat")

# Fitted values
fitted(x)
fitted(x, item = 1)

# Predicted values
predict(x)
predict(x, item = 1)

# Predicted values for new subjects
predict(x, item = 1, score = 0, group = 1)
predict(x, item = 1, score = 0, group = 0)

## End(Not run)
```

GMAT *Dichotomous Data Set Based on Graduate Management Admission Test*

Description

The GMAT data set is generated data set based on parameters from Graduate Management Admission Test (GMAT) data set (Kingston et al., 1985). First two items were considered to function differently in uniform and non-uniform way respectively. The data set represents responses of 1,000 subjects to multiple-choice test of 20 items. A correct answer is coded as 1 and incorrect answer as 0. The column group represents group membership, where 0 represents reference group and 1 represents focal group.

Usage

```
data(GMAT)
```

Format

A GMAT data frame consists of 1,000 observations on the following 21 variables.

The first 20 columns represent dichotomously scored items of the test. The 21st column is vector of group membership; values 0 and 1 refer to reference and focal group.

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Kingston, N., Leary, L., and Wightman, L. (1985). An Exploratory Study of the Applicability of Item Response Theory Methods to the Graduate Management Admission Test. ETS Research Report Series, 1985(2) : 1-64.

Examples

```
data(GMAT)
```

GMATkey *Key of Correct Answers for* GMATtest *Data Set*

Description

The GMATkey is a vector of factors representing correct answers of generated GMATtest data set based on Graduate Management Admission Test (GMAT) data set (Kingston et al., 1985).

Usage

```
data(GMATkey)
```

Format

A data frame with 20 values representing correct answers to items of GMATtest data set. For more details see GMATtest.

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Kingston, N., Leary, L., and Wightman, L. (1985). An Exploratory Study of the Applicability of Item Response Theory Methods to the Graduate Management Admission Test. ETS Research Report Series, 1985(2) : 1-64.

See Also

GMATtest, GMAT

Examples

```
data(GMATkey)
```

GMATtest *Data Set Based on Graduate Management Admission Test*

Description

The GMATtest data set is generated data set based on parameters from Graduate Management Admission Test (GMAT) data set (Kingston et al., 1985). First two items were considered to function differently in uniform and non-uniform way respectively. The data set represents responses of 1,000 subjects to multiple-choice test of 20 items. Aditionally, 4 possible answers on all items were generated, coded A, B, C and D. The column group represents group membership, where 0 represents reference group and 1 represent focal group.

Usage

```
data(GMATtest)
```

Format

A GMAT data frame consists of 1,000 observations on the following 21 variables.

The first 20 columns represents answers of subject to an items of the test. The 21st column is vector of group membership; values 0 and 1 refer to reference and focal group.

Correct answers were D, C, C, C, B, B, B, A, A, D, A, A, A, C, C, B, C, B, D, A respectively. For more details see data GMATkey

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Kingston, N., Leary, L., and Wightman, L. (1985). An Exploratory Study of the Applicability of Item Response Theory Methods to the Graduate Management Admission Test. ETS Research Report Series, 1985(2) : 1-64.

See Also

GMATkey, GMAT

Examples

```
data(GMATtest)
```

---

startNLR *Calculates Starting Values*

Description

Calculates starting values for function `difNLR` function based on linear approximation.

Usage

```
startNLR(data, group, parameterization = "IRT")
```

Arguments

`data`	a matrix or data frame of binary data.
`group`	a binary vector of group membership. "0" represents reference group, "1" represents focal group.
`parameterization`	character: parameterization of regression coefficients. Possible options are "IRT" (Item response theory, default option) and "LR" (Logistic regression)

Author(s)

Adela Drabinova
Institute of Computer Science, The Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University in Prague
adela.drabinova@gmail.com

Patricia Martinkova
Institute of Computer Science, The Czech Academy of Sciences
martinkova@cs.cas.cz

Karel Zvara
Faculty of Mathematics and Physics, Charles University in Prague

References

Drabinova, A. and Martinkova P. (2016). Detection of Differenctial Item Functioning Based on Non-Linear Regression, Technical Report, V-1229, http://hdl.handle.net/11104/0259498

Examples

```
# loading data based on GMAT
data(GMAT)

data  <- GMAT[, colnames(GMAT) != "group"]
group <- GMAT[, "group"]

# starting values
startNLR(data, group)
```

Index