

The *difR* package

A toolbox for the identification of dichotomous differential item functioning

David Magis

University of Liège and K.U. Leuven, Belgium
david.magis@ulg.ac.be

Outline

- DIF (in 5 minutes)

Outline

- DIF (in 5 minutes)
- DIF methods (in 2 minutes)

Outline

- DIF (in 5 minutes)
- DIF methods (in 2 minutes)
- The *difR* package (in 5 minutes)

Outline

- DIF (in 5 minutes)
- DIF methods (in 2 minutes)
- The *difR* package (in 5 minutes)
- Application (until Florian's signal or lunch time 😊)

Outline

- DIF
- DIF methods
- The *difR* package
- Application

Outline

- DIF
- DIF methods
- The *difR* package
- Application

DIF

- Framework:
 - One test with **dichotomous** items
 - Two (or more) **groups**
 - One **reference** group, one (or more) **focal** group(s)
 - Question of interest: are the items functioning similarly in all groups?

DIF (2)

- Item is said to have differential functioning (to be **DIF**) if examinees from **different groups**, but with the **same ability level**, have **different probabilities** of answering the item correctly
- Goals of DIF research:
 - To develop methods to detect DIF
 - To identify and remove DIF items

DIF (3)

- Four main aspects:
 - IRT vs non-IRT
 - Uniform DIF vs nonuniform DIF
 - Two vs more than two groups
 - Item purification

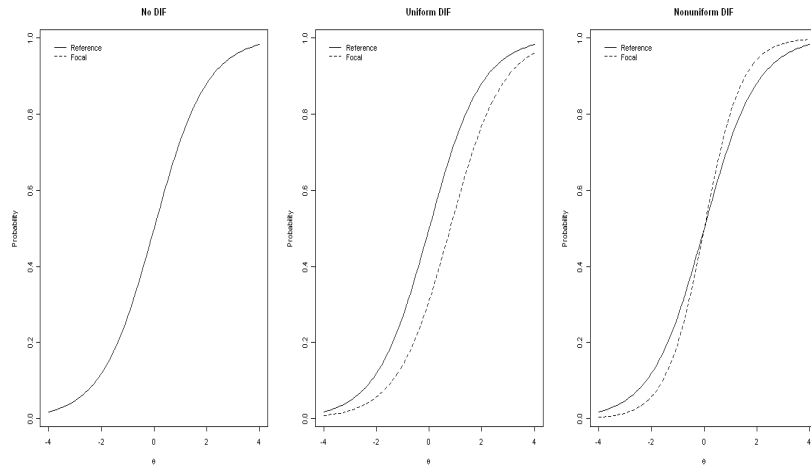
DIF (4)

- IRT vs non-IRT:
 - Early first methods rely on statistical aspects (Mantel-Haenszel, logistic regression, SIBTEST...) and don't require fitting IRT models
 - Other methods fit IRT models and compare model fits (LRT) or item parameters (Lord, Raju)

DIF (5)

- Uniform vs nonuniform:
 - DIF effect is **uniform** if the item-group interaction is independent of the ability level, and **nonuniform** otherwise
 - Non-IRT methods: conditional association between item response and group membership is independent of matching variable (i.e. sum score)
 - IRT methods:

DIF (6)



DIF (7)

- Two vs more than two groups:
 - Most methods deal with **two** groups (reference and focal)
 - Some are specifically designed for simultaneous comparison of **more than two** groups

DIF (8)

- Item purification:
 - DIF items can affect the validity of the measures of DIF
 - Some known effects:
 - **Type I error inflation:** non-DIF items are incorrectly flagged as DIF
 - **Masking effect:** Items with large DIF effect can mask the presence of other DIF items but with smaller DIF effects

DIF (9)

- Proposed solution: **item purification**
- **Iterative** process that successively **removes** items flagged as DIF from
 - the computation of sum scores (non-IRT)
 - the rescaling of item parameters (IRT)
- Process stops when
 - no DIF item is detected
 - two successive steps of the process yield the same classification of items

DIF (10)

- Item purification
 - controls for Type I error inflation
 - usually yields increased power
- but
 - can be time consuming
 - no guarantee that the iterative process stops

Outline

- DIF
- **DIF methods**
- The *difR* package
- Application

DIF methods

Method	DIF effect	Groups	
		Two	More than two
NON-IRT	Uniform	TID, MH, Std, logReg, SIBTEST	GMH, genLogReg, genTID
NON-IRT	Nonuniform	MH*, BD, logReg, SIBTEST*	genLogReg
IRT	Uniform	Lord, Raju, LRT	genLord
IRT	Nonuniform	Lord, Raju, LRT	genLord

DIF methods (2)

Method	DIF effect	Groups	
		Two	More than two
NON-IRT	Uniform	TID, MH, Std, logReg, SIBTEST	GMH, genLogReg, genTID
NON-IRT	Nonuniform	MH*, BD, logReg, SIBTEST*	genLogReg
IRT	Uniform	Lord, Raju, LRT	genLord
IRT	Nonuniform	Lord, Raju, LRT	genLord

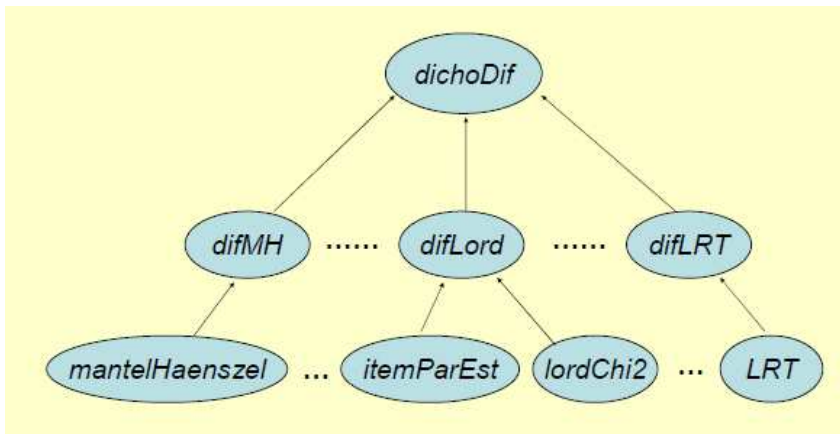
Outline

- DIF
- DIF methods
- **The *difR* package**
- Application

The *difR* package

- Jointly developed by
 - **Sébastien Béland** (UQAM, Canada),
 - **Francis Tuerlinckx** (K.U. Leuven, Belgium)
 - **Paul De Boeck** (University of Amsterdam, The Netherlands and K. U. Leuven, Belgium)

The *difR* package (2)



The *difR* package (3)

- Three levels of R functions:
 - **Low level**: Working functions, do the computational job
 - **Middle level**: DIF functions, of the form “dif...” to call a specific method (e.g. `difMH` for Mantel-Haenszel)
 - **High level**: `dichoDif` function, calls several middle level functions and merge their output

The *difR* package (4)

- Generic **input** parameters:
 - `Data`: the data matrix
 - `group`: the vector of group membership
 - `focal.name(s)`: the name(s) of focal group(s)
 - `purify`: should item purification be performed? (default is FALSE)
 - `save.output`: should the output be saved into a text file? (default is FALSE)
 - `output`: specifies the name and the place to save the output

The *difR* package (5)

- Specific input parameters:
 - Depend on the method
 - Can specify:
 - the **DIF statistic** (e.g. Mantel-Haenszel)
 - the **type** of logistic **model** (e.g. logistic regression) or IRT model (e.g. Lord, Raju)
 - The DIF classification **thresholds** (e.g. standardization)
 - The matrix of item parameters (e.g. Lord, Raju)
 - Etc.

The *difR* package (6)

- Output:
 - **List** with all useful information (input and output)
 - Displayed in a visually attractive way through `print(.)`
 - Can be **saved** into a text file
 - Can be **plotted** for visual representation of DIF statistics, through `plot(.)`

The *difR* package (7)

- `dichoDif` function:
 - Calls **one** or **several** DIF methods
 - Either for two groups, or for more than two groups
 - All **specific options** can be passed to `dichoDif`
 - Returns a **summary** of all requested methods
 - For direct **comparison** of method output

Outline

- DIF
- DIF methods
- The *difR* package
- **Application**

Application

- Data set: **verbal aggression** example
- **316 students** (243 females, 73 males), first year psychology (K.U. Leuven)
- **24 items** built by mixing
 - 4 frustrating situations
 - 3 possible aggressive responses
 - 2 possible actions related to aggressive responses

Application (2)

- Frustrating situations:
 - **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”

Application (3)

- Possible **actions**:
 - I want to...
 - I do...
- Possible **aggressive responses**:
 - To shout
 - To curse
 - To scold

Application (4)

- Examples:
 - **S1DoShout**: “A bus fails to stop for me. I shout”.
 - **S3WantCurse**: “The grocery store closes just as I am about to enter. I want to curse.”
 - Etc.

Application (5)

- “**Correct** response” if student responds in an aggressive way, that is, if he/she answers “yes”.
- Research question: do the items “function” similarly for males and females?
- Data collected by Vansteelandt (2000)
- Available in *difR*

Application (6)

- **Reference** group: female students
- **Focal** group: male students
- Columns 1-24: items
- Column 25: *Anger* (not used here)
- Column 26: *Gender* (group membership)

Application (7)

- Three DIF analyzes:
 - Using **Mantel-Haenszel**
 - Using **Lord's test** (and 1PL model)
 - Using `dichoDif` function and several DIF methods

Application (8)

- And now...



Application (9)

- **Reading** and **preparing** the data:

```
require(difR)
data(verbal)
verbal <-verbal[colnames(verbal)!="Anger"]
```

Application (10)

- Mantel-Haenszel analysis:
 - Focal group: 1 (males)
 - MH chi-square statistic (default)
 - Significance level: 5% (default)
 - No item purification (default)

```
difMH(verbal, group="Gender", focal.name=1)
```

Application (11)

- Output:

```
Detection of Differential Item Functioning using Mantel-Haenszel method
with continuity correction and without item purification
```

Application (12)

Mantel-Haenszel Chi-square statistic:

```
Stat. P-value
S1WantCurse 1.7076 0.1913
S1WantScold 2.1486 0.1427
S1WantShout 0.9926 0.3191
S2WantCurse 1.9302 0.1647
S2WantScold 2.9540 0.0857 .
S2WantShout 9.6032 0.0019 **
S3WantCurse 0.0013 0.9711
S3WantScold 0.6752 0.4112
S3WantShout 0.8185 0.3656
S4WantCurse 1.6292 0.2018
S4WantScold 0.0152 0.9020
S4WantShout 4.1188 0.0424 *
S1DoCurse 0.1324 0.7160
S1DoScold 2.7501 0.0972 .
S1DoShout 0.0683 0.7938
S2DoCurse 6.3029 0.0121 *
S2DoScold 6.8395 0.0089 **
S2DoShout 0.2170 0.6414
S3DoCurse 5.7817 0.0162 *
S3DoScold 3.8880 0.0486 *
S3DoShout 0.2989 0.5846
S4DoCurse 1.1220 0.2895
S4DoScold 1.4491 0.2287
S4DoShout 0.8390 0.3597
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Detection threshold: 3.8415 (significance level: 0.05)

Application (13)

Items detected as DIF items:

```
S2WantShout
S4WantShout
S2DoCurse
S2DoScold
S3DoCurse
S3DoScold
```

Application (14)

Effect size (ETS Delta scale):

Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect

```
alphaMH deltaMH
S1WantCurse 1.7005 -1.2476 B
S1WantScold 1.7702 -1.3420 B
S1WantShout 1.4481 -0.6701 A
S2WantCurse 1.9395 -1.5567 C
S2WantScold 1.9799 -1.6052 C
S2WantShout 2.8804 -2.4861 C
S3WantCurse 0.9439 0.1358 A
S3WantScold 0.7194 0.7741 A
S3WantShout 1.5281 -0.9965 A
S4WantCurse 1.6849 -1.2260 B
S4WantScold 1.0901 -0.2028 A
S4WantShout 2.3458 -2.0036 C
S1DoCurse 0.7967 0.5340 A
S1DoScold 0.4995 1.6313 C
S1DoShout 1.1765 -0.3821 A
S2DoCurse 0.3209 2.6709 C
S2DoScold 0.3746 2.3072 C
S2DoShout 0.7931 0.5447 A
S3DoCurse 0.4616 1.8165 C
S3DoScold 0.4727 1.7606 C
S3DoShout 0.6373 1.0585 B
S4DoCurse 0.6444 1.0327 B
S4DoScold 0.6385 1.0541 B
S4DoShout 1.6053 -1.1123 B
```

Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
(for absolute values of 'deltaMH')

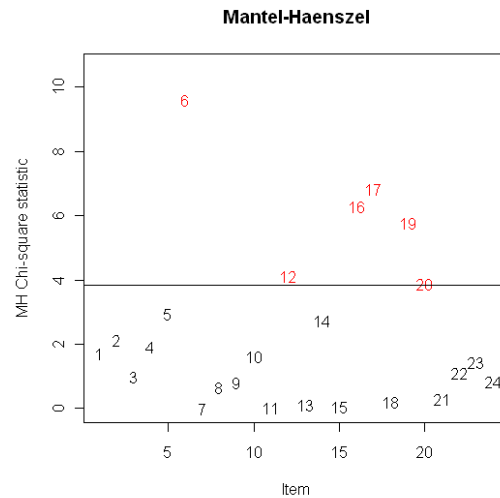
Output was not captured!

Application (15)

- Plotting the output:

```
plot(
  difMH(verbal, group="Gender", focal.name=1)
)
```

Application (16)



Application (17)

- Other possible options:
 - Significance level: `alpha = ...`
 - No continuity correction: `correct = FALSE`
 - Log OR DIF statistic: `MHstat = "logOR"`
 - Item purification: `purify = TRUE`
 - Number of iterations: `nrIter=...`
 - ...

Application (18)

- Lord's test:
 - Focal group: 1 (males)
 - 1PL model to be estimated from 'ltm' package
 - Significance level: 5% (default)
 - No item purification (default)

```
r <- difLord(verbal,  
group="Gender", focal.name=1,  
model="1PL", engine="ltm")
```

Application (19)

- Structure of the output (using `str(r)`):

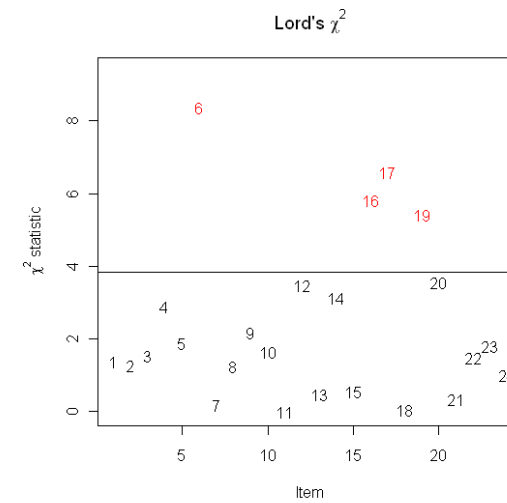
```
List of 13  
 $ LordChi      : num [1:24] 1.37 1.27 1.54 2.88 1.89 ...  
 $ alpha       : num 0.05  
 $ thr         : num 3.84  
 $ DIFitems    : int [1:4] 6 16 17 19  
 $ purification: logi FALSE  
 $ model       : chr "1PL"  
 $ c           : NULL  
 $ engine      : chr "ltm"  
 $ itemParInit : num [1:48, 1:2] -1.19 -0.557 -0.103 -1.747 -0.713 ...  
 .. attr(*, "dimnames")=List of 2  
 .. ..$ : chr [1:48] "Item1" "Item2" "Item3" "Item4" ...  
 .. ..$ : chr [1:2] "b" "se(b)"  
 $ estPar      : logi TRUE  
 $ names       : chr [1:24] "S1wantCurse" "S1wantScold" "S1wantShout" "S2wantCurse" ...  
 $ save.output : logi FALSE  
 $ output      : chr [1:2] "out" "default"  
 - attr(*, "class")= chr "Lord"
```

Application (20)

- Visualizing the results:

```
plot(r)
```

Application (21)



Application (22)

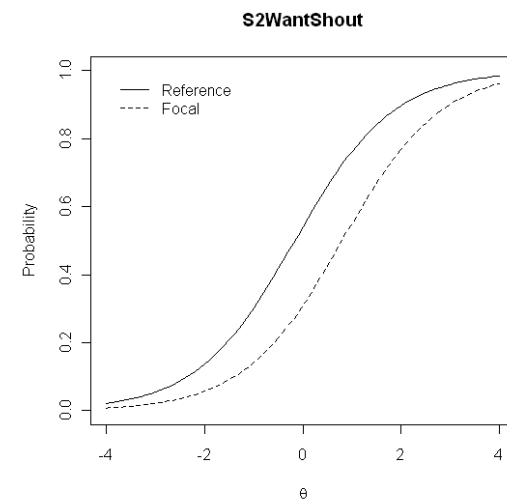
- Visualizing one item in particular:

```
plot(r, plot="itemCurve", item=6)
```

or

```
plot(r, plot="itemCurve",  
     item="S2WantShout")
```

Application (23)



Application (24)

- Other possible options:
 - Significance level: `alpha = ...`
 - Item purification: `purify = TRUE`
 - Number of iterations: `nrIter=...`
 - Provide the item parameters by yourself:
`irtParam = ...`
 - ...

Application (25)

- `dichoDif` use:
 - Focal group: 1 (males)
 - Methods: Mantel-Haenszel, Standardization, logistic regression, Lord's test (1PL), Raju's method (1PL)
 - Significance level: 5% (default)
 - No item purification (default)

```
dichoDif(verbal,group="Gender",
focal.name=1,method=c(
"MH","Std","Logistic","Lord","Raju"),
model="1PL)
```

Application (26)

Comparison of DIF detection results using 5 methods

Methods used: Mantel-Haenszel, Standardization, Logistic regression,
Lord's chi-square test, Raju's area

Parameters:

Significance level: 0.05

Standardization threshold: 0.1

Mantel-Haenszel DIF statistic: Chi-square statistic

Mantel-Haenszel continuity correction: Yes

Weights for standardized P-DIF statistic: based on the focal group

Logistic regression DIF statistic: LRT statistic

DIF effect(s) tested by logistic regression: uniform DIF effect

Item response model: 1PL

Item purification: No

Application (27)

Comparison of DIF detection results:

	M-H	Stand.	Logistic	Lord	Raju	#DIF
S1wantCurse	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S1wantScold	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S1wantShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S2wantCurse	NoDIF	NoDIF	DIF	NoDIF	NoDIF	1/5
S2wantScold	NoDIF	DIF	NoDIF	NoDIF	NoDIF	1/5
S2wantShout	DIF	DIF	DIF	DIF	DIF	5/5
S3wantCurse	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S3wantScold	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S3wantShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4wantCurse	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4wantScold	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4wantShout	DIF	DIF	NoDIF	NoDIF	NoDIF	2/5
S1doCurse	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S1doScold	NoDIF	NoDIF	DIF	NoDIF	NoDIF	1/5
S1doShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S2doCurse	DIF	DIF	DIF	DIF	DIF	5/5
S2doScold	DIF	DIF	DIF	DIF	DIF	5/5
S2doShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S3doCurse	DIF	DIF	DIF	DIF	DIF	5/5
S3doScold	DIF	DIF	DIF	NoDIF	NoDIF	3/5
S3doShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4doCurse	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4doScold	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5
S4doShout	NoDIF	NoDIF	NoDIF	NoDIF	NoDIF	0/5

Output was not captured!

Application (28)

- For an interpretation of DIF effects with this data set:

De Boeck, P. (2008). Random items IRT models. *Psychometrika*, 73, 533-559.

Further work

- Improve some methods (e.g. more flexibility in IRT model fitting, incorporating new packages, etc.)
- Include **SIBTEST** and **transformed item difficulties (TID)** methods
- Extend to **polytomous** items
- Allow for **missing data**

Final slide

- Further information:
 - Package:
<http://cran.r-project.org/web/packages/difR/>
 - Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847-862.
 - Slides and R script:
<http://hdl.handle.net/2268/65169>

THANK YOU!

... and if doesn't work ...

THANK YOU!

