

# ShinyItemAnalysis for Psychometric Training and to Enforce Routine Analysis of Educational Tests

Patrícia Martinková

Dept. of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences  
College of Education, Charles University in Prague

R meetup Warsaw, May 24, 2018

# Announcement 1: Save the date for Psychoco 2019!



International Workshop on Psychometric Computing

Psychoco 2019

February 21 - 22, 2019

Charles University & Czech Academy of Sciences, Prague

[www.psychoco.org](http://www.psychoco.org)

Since 2008, the international Psychoco workshops aim at bringing together researchers working on modern techniques for the analysis of data from psychology and the social sciences (especially in R).

## Announcement 2: Job offers

Job offers at Institute of Computer Science:

- CAS-ICS Postdoctoral position (deadline: August 30)
- ICS Doctoral position (deadline: June 30)
- ICS Fellowship for junior researchers (deadline: June 30)
- ... further possibilities to participate on grants

E-mail at [martinkova@cs.cas.cz](mailto:martinkova@cs.cas.cz) if interested in position in the area of

- Computational psychometrics
- Interdisciplinary statistics
- Other related disciplines

1. Introduction
2. ShinyItemAnalysis
3. Teaching psychometrics
4. Routine analysis of tests
5. Discussion

# Motivation

- To teach psychometric concepts and methods
  - Graduate courses "IRT models", "Selected topics in psychometrics"
  - Workshops for admission test developers
  - Active learning approach w/ hands-on examples
- To enforce routine analyses of educational tests
  - Admission tests to Czech Universities
  - Physiology concept inventories
  - ... tests of various purposes across the world
- Promotion of own psychometrics research
  - Detection of Differential Item Functioning (DIF)

**Need for user-friendly and freely available tool**

# ShinyItemAnalysis Application



# ShinyItemAnalysis

Interactive (and step by step) analysis of educational tests and their items

Available as:

- R package
  - Version 1.2.7 now on [CRAN](#)
  - Newest version on [GitHub](#)

```
startShinyItemAnalysis()
```

- Online shiny application
  - ICS server in Prague, CZ:

```
https://shiny.cs.cas.cz/ShinyItemAnalysis/
```

- shinyapps.io:

```
https://cemp.shinyapps.io/ShinyItemAnalysis/
```

# Authors and contributors



Patrícia Martinková<sup>1,2</sup>

Adéla Drabinová<sup>1,3</sup>

Jakub Houdek<sup>1,4</sup>

Ondřej Leder<sup>3</sup>

Lubomír Štěpánek<sup>4,5</sup>

<sup>1</sup> Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences

<sup>2</sup> College of Education, Charles University, Prague

<sup>3</sup> Department of Probability and Mathematical Statistics, Charles University, Prague

<sup>4</sup> Faculty of Informatics and Statistics, University of Economics, Prague

<sup>5</sup> First Faculty of Medicine, Charles University, Prague

# ShinyItemAnalysis application

ShinyItemAnalysis [Test and item analysis](#) [About](#) [Data](#) [Summary](#) [Validity](#) [Item analysis](#) [Regression](#) [IRT models](#) [DIF/Fairness](#) [Reports](#) [References](#)

## Description

**ShinyItemAnalysis** provides analysis of educational tests (such as admission tests) and their items including:

- Exploration of total and standard scores on **Summary** page.
- Correlation structure and predictive validity analysis on **Validity** page.
- Item and distractor analysis on **Item analysis** page.
- Item analysis by logistic models on **Regression** page.
- Item analysis by item response theory models on **IRT models** page.
- Differential item functioning (DIF) and differential distractor functioning (DDF) methods on **DIF/Fairness** page.

This application is based on the free statistical software R and its **shiny** package.

For all graphical outputs a download button is provided. Moreover, on **Reports** page HTML or PDF report can be created. Additionally, all application outputs are complemented by selected R code hence the similar analysis can be run and modified in R.

## Data

For demonstration purposes, by default, 20-item dataset **gwat** from R **difR** package is used. Other four datasets are available: **GWAT2** and **HSAT-8** from **difR** package and **Medical 100** and **HCI** from **ShinyItemAnalysis** package. You can change the dataset (and try your own one) on page **Data**.

## Availability

Application can be downloaded as R package from CRAN. It is also available online at Czech Academy of Sciences  and shinyapps.io .

## Version

Current version of **ShinyItemAnalysis** available on CRAN is 1.2.7. Version available online is 1.2.7. The newest development version available on GitHub is 1.2.7. See also older versions: 0.1.0, 0.2.0, 1.0.0, 1.1.0, 1.2.3, 1.2.6.

## Authors and contributors



Patricia  
Martinkova



Adela  
Drabikova



Ondrej  
Leder



Jakub  
Houdek



Lubomir  
Stepanek

## List of packages used

`library(corplot)`  
`library(CTT)`

`library(ggplot2)`  
`library(ggrid)`

`library(moments)`  
`library(msn)`

`library(rmarkdown)`  
`library(shiny)`



ShinyItemAnalysis Test and item analysis | Version 1.2.7  
© 2018 Patricia Martinkova, Adela Drabikova, Ondrej Leder and Jakub Houdek



# R package ShinyItemAnalysis downloads from CRAN

## Package CRAN downloads over time

Enter an R package to see the # of downloads over time from the RStudio CRAN Mirror. You can enter multiple packages to compare them

**Packages:**

**Data Transformation:**

Daily

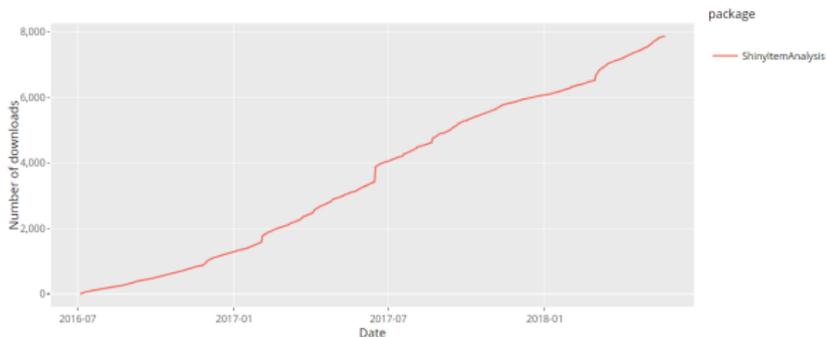
Weekly

Cumulative

**Date range: yyyy-mm-dd**

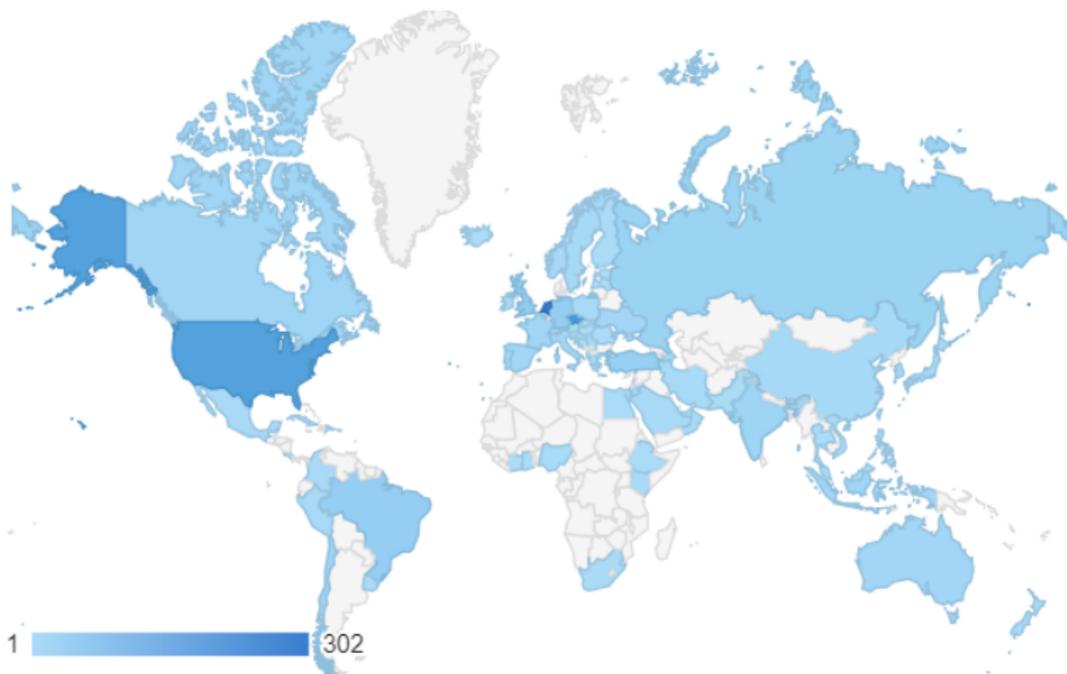
 to 

Created using the [cranlogs](#) package. This app is not affiliated with RStudio or CRAN. You can find the code for the app [here](#), or read more about it [here](#).



ShinyItemAnalysis  
Title: Test and Item Analysis via Shiny  
Latest version: 1.2.7

# ShinyItemAnalysis online app is used worldwide!



# ShinyItemAnalysis for teaching psychometrics

## Who do we teach:

- Graduate students of different fields (Psychometrics ▶ NMST570)
- Faculties, university stakeholders

## Some helpful features:

- Toy datasets, allows to upload own data
- Building models in a step-by-step way
- Models, estimates, interactive interpretation of results
- Interactive training and exercises
- Provides sample R code

# Datasets

- Five toy datasets are available
- Allows to upload and preview one's own dataset

ShinyItemAnalysis Test and Item analysis
About
Data
Summary
Validity
Item analysis
Regression
IRT models
DIF/Fairness
Reports
References

HCI (McFarland et al., 2017) is a real dataset of Homostatic Concept Inventory from [ShinyItemAnalysis](#) R package. The dataset represents responses of 601 subjects (400 males, 246 females) to multiple-choice test of 20 items. **HCI** contains criterion variable - indicator whether student plans to major in the life sciences.

**Select dataset**

HCI

**Upload your own datasets**

Main **data** file should contain responses of individual respondents (rows) to given items (columns). Header may contain item names, no row names should be included. If responses are in unscored ABCD format, the **key** provides correct response for each item. If responses are scored 0-1, key is vector of 1s.

**Group** is 0-1 vector, where 0 represents reference group and 1 represents focal group. Its length need to be the same as number of individual respondents in main dataset. If the group is not provided then it won't be possible to run DIF and DCF detection procedures on **DIF/Fairness** page.

**Criterion variable** is either discrete or continuous vector (e.g. future study success or future GPA in case of admission tests) which should be predicted by the measurement. Again, its length needs to be the same as number of individual respondents in the main dataset. If the criterion variable is not provided then it won't be possible to run validity analysis in **Predictive validity** section on **Validity** page.

In all data sets **header** should be either included or excluded. Columns of dataset are by default renamed to item and number of particular column. If you want to keep your own names, check box **Keep item names** below. Missing values in scored dataset are by default evaluated as 0. If you want to keep them as missing, check box **Keep missing values** below.

**Choose data (csv file)**

Browse... HCI\_ABCD.csv

Upload console

**Choose key (csv file)**

Browse... HCI\_key.csv

Upload console

**Choose groups for DIF (optional)**

Browse... HCI\_group.csv

Upload console

**Choose criterion variable (optional)**

Browse... HCI\_major.csv

Upload console

Submit Data

✓ Your data were successfully uploaded. Check them in **Data exploration** tab.

---

**Data specification**

Header

Keep item names

Keep missing values

**Separator**

Comma

Semicolon

Tab

**Quote**

None

Double Quote

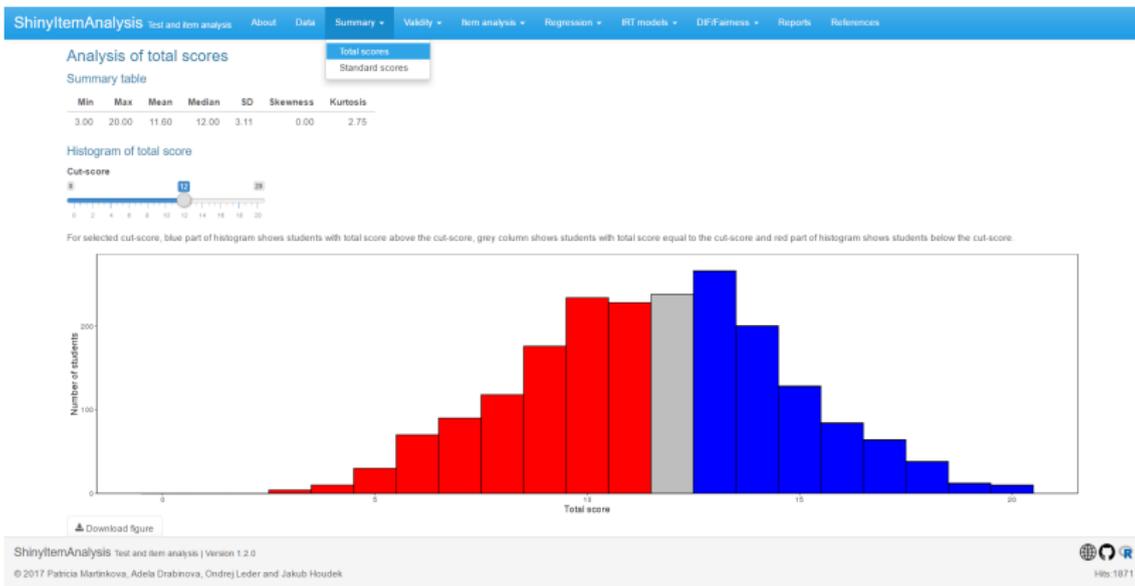
Single Quote

ShinyItemAnalysis Test and Item analysis | Version 1.2.7  
© 2018 Patricija Martinkova, Adela Drabinkova, Ondrej Leder and Jakub Houdek

HE5-5776

# Summary of Total Scores

- Summary statistics
- Interactive histogram



# Criterion validity

- Only when criterion variable is available (study success, GPA, etc.)
- Available for total score as well as for items



# Correlation structure

- Correlations between items
- Item clusters

ShinyItemAnalysis [Test and Item analysis](#) [About](#) [Data](#) [Summary](#) [Validity](#) [Item analysis](#) [Regression](#) [IRT models](#) [DIF/Fairness](#) [Reports](#) [References](#)

Correlation structure  
Predictive validity

### Correlation structure

**Polychoric correlation heat map**

Polychoric correlation heat map is a correlation plot which displays a polychoric correlations of items. The size and shade of circles indicate how much the items are correlated (larger and darker circle means larger correlation). The color of circles indicates in which way the items are correlated - blue color shows positive correlation and red color shows negative correlation.

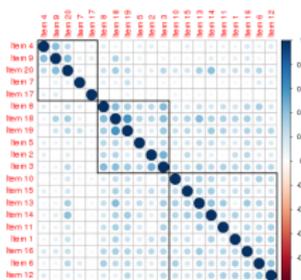
Polychoric correlation heat map can be reorder using hierarchical **clustering method** below. Ward's method aims at finding compact clusters based on minimizing the within-cluster sum of squares. Ward's n. 2 method used squared dissimilarities. Single method connects clusters with the nearest neighbours, i.e. the distance between two clusters is calculated as the minimum of distances of observations in one cluster and observations in the other clusters. Complete linkage with farthest neighbours, i.e. maximum of distances. Average linkage method used the distance based on weighted average of the individual distances. With McQuitty method used unweighted average. Median linkage calculates the distance as the median of distances between an observation in one cluster and observation in the other cluster. Centroid method used distance between centroids of clusters.

With **number of clusters** larger than 1, the rectangles representing clusters are drawn.

**Number of clusters**

**Clustering method**

Ward's
▼



Download figure

ShinyItemAnalysis | Test and Item analysis | Version 1.2.6  
© 2018 Patricija Martinková, Adela Drabíková, Ondřej Leder and Jakub Houdek

# Traditional Item Analysis

- Difficulty, discrimination
- Cronbach's alpha w/o item, index RIT, RIR, etc.

ShinyItemAnalysis [Test and item analysis](#) [About](#) [Data](#) [Summary](#) [Validity](#) [Item analysis](#) [Regression](#) [IRT models](#) [DEFairness](#) [Reports](#) [References](#)

Traditional item analysis **Traditional item analysis**  
Distractors

Traditional item analysis uses proportions of correct answers or correlations to estimate item properties.

Item difficulty/discrimination plot

Displayed is difficulty (red) and discrimination (blue) for all items. Items are ordered by difficulty.  
Difficulty of items is estimated as percent of students who answered correctly to that item.  
Discrimination is described by difference of percent correct in upper and lower third of students (Upper-Lower Index, ULI). By rule of thumb it should not be lower than 0.2 (borderline in the plot), except for very easy or very difficult items.

Item	Difficulty (red)	Discrimination (blue)
1	0.42	0.48
2	0.44	0.38
3	0.46	0.42
4	0.48	0.45
5	0.48	0.28
6	0.52	0.42
7	0.54	0.32
8	0.54	0.42
9	0.56	0.38
10	0.58	0.32
11	0.58	0.32
12	0.60	0.38
13	0.60	0.32
14	0.62	0.35
15	0.64	0.38
16	0.64	0.30
17	0.68	0.38
18	0.70	0.28
19	0.72	0.25
20	0.78	0.28

[Download figure](#)

**Cronbach's alpha**  
Cronbach's alpha is an estimate of the reliability of a psychometric test. It is a function of the number of items in a test, the average covariance between item pairs, and the variance of the total score (Cronbach, 1951).

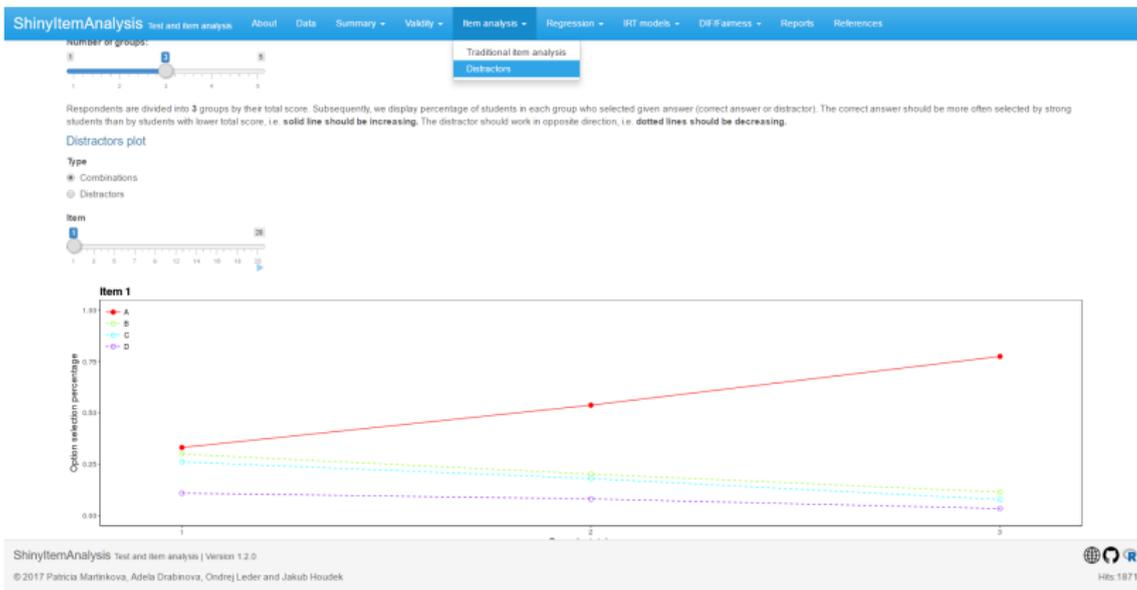
Estimate	SD
0.55	0.16

ShinyItemAnalysis Test and item analysis | Version 1.2.0  
© 2017 Patricie Martinkova, Adela Drabesova, Ondrej Leder and Jakub Houdek

Hb:1871

# Distractor Analysis

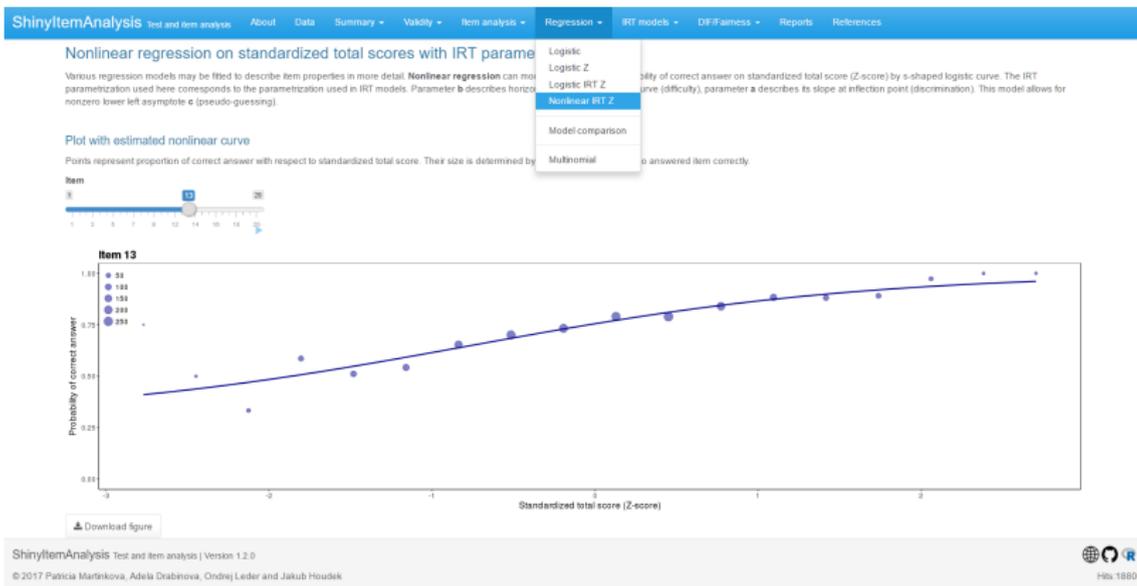
- Displays option selection percentage by total score group
- Number of groups can be changed





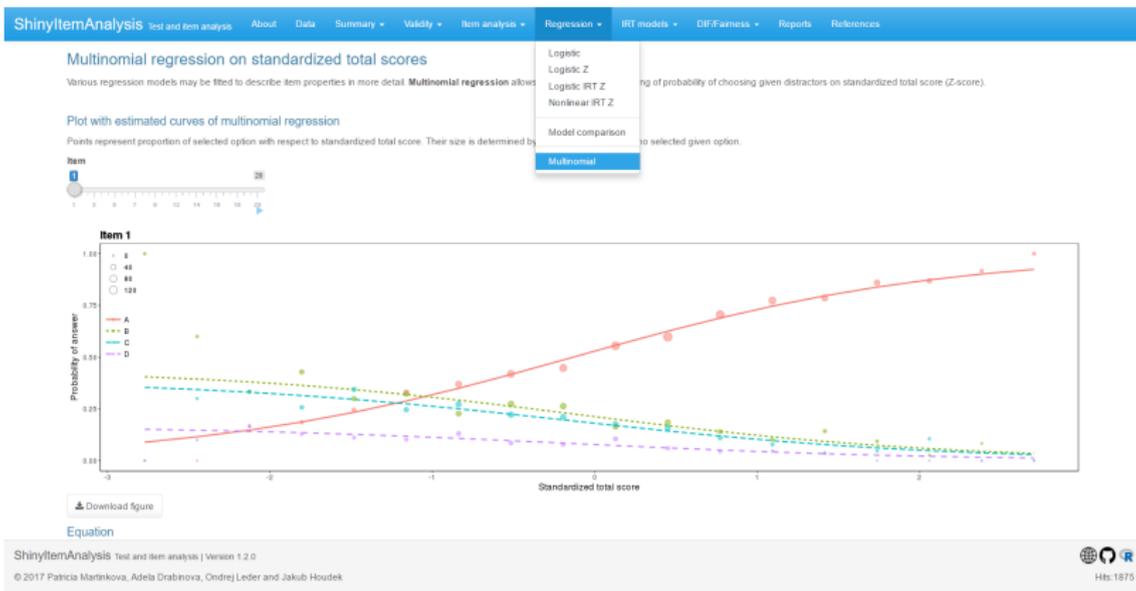
# Nonlinear Regression

- Allows for guessing (and inattention)



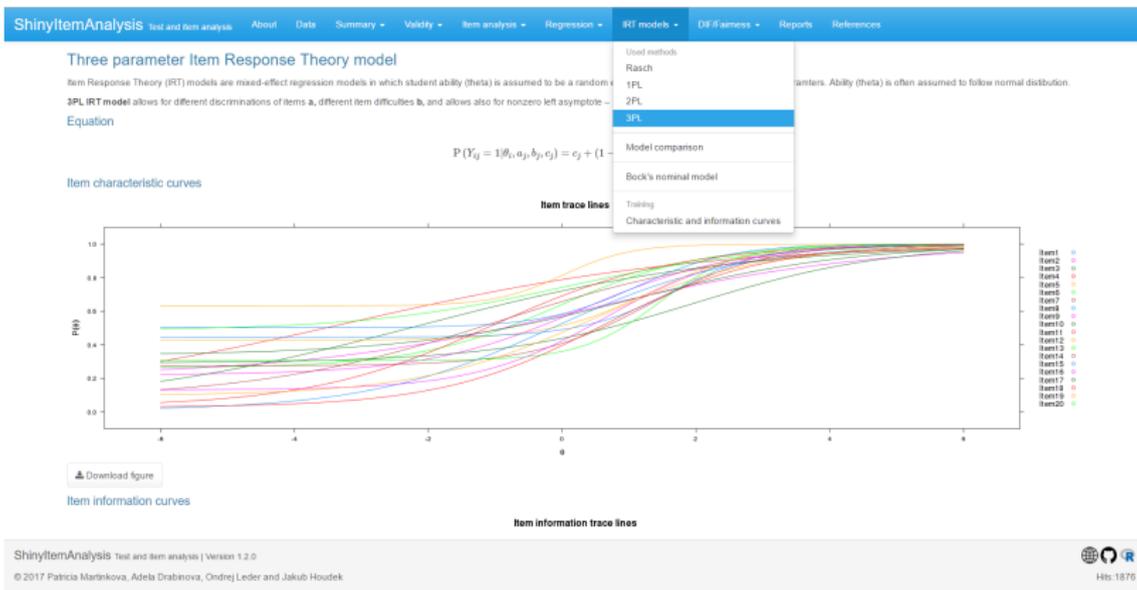
# Multinomial Regression

- Allows for joint modeling of distractors



# IRT Models

- Conceptualized as nonlinear mixed effect models
- More precise ability estimation



# Dichotomous IRT Models - interactive training

- Plots Item Characteristic and Information Curves (ICC and IIC) based on selected parameters

ShinyItemAnalysis [test and item analysis](#) [About](#) [Data](#) [Summary](#) [Validity](#) [Item analysis](#) [Regression](#) [IRT models](#) [DIF analysis](#) [Reports](#) [References](#)

Parameters

Select parameters  $a$  (discrimination),  $b$  (difficulty),  $c$  (guessing) and  $d$  (inflation). By constraining  $a = 1$ ,  $c = 0$ ,  $d = 1$  you get Rasch model. With option  $c = 0$  and  $d = 1$  you get 2PL model and with option  $d = 1$  3PL model. When different curve parameters describe properties of the same item but for different groups of respondents, this phenomenon is called Differential Item Functioning (DIF). See further section for more information.

**a - discrimination**  **b - difficulty**  **c - guessing**  **d - inflation**

**a - discrimination**  **b - difficulty**  **c - guessing**  **d - inflation**

Select also the value of latent ability  $\theta$  to see the interpretation of the item characteristic curves.

$\theta$  - latent ability

Equations

$$P(\theta|a, b, c, d) = c + (d - c) \cdot \frac{e^{a(\theta - b)}}{1 + e^{a(\theta - b)}}$$

$$I(\theta|a, b, c, d) = a \cdot (d - c) \cdot \frac{e^{a(\theta - b)}}{[1 + e^{a(\theta - b)}]^2}$$

Interpretation: The probability of correct answer with latent ability  $\theta = 0$  in red item with parameters  $a = 1$ ,  $b = 0$ ,  $c = 0$ ,  $d = 1$  is equal to 0.50. The probability of correct answer with latent ability  $\theta = 0$  in blue item with parameters  $a = 2$ ,  $b = 0.5$ ,  $c = 0$ ,  $d = 1$  is equal to 0.27.

**Item characteristic curve**

**Item information function**

ShinyItemAnalysis Test and item analysis | Version 1.2.7  
© 2018 Patricja Martinková, Adèle Duboisová, Ondřej Lešter and Jakub Houdek

HRM 1776

# Polytomous IRT Models - interactive training

- Plots Category Response Curves and Expected Item Score based on selected parameters

ShinyItemAnalysis [Test and item analysis](#) [About](#) [Data](#) [Summary](#) [Validity](#) [Item analysis](#) [Regression](#) [IRT models](#) [Diff. IRT models](#) [Reports](#) [References](#)

Parameters

Select number of responses and difficulty for cumulative probabilities b and common discrimination parameter a. Cumulative probability  $P(Y \geq k)$  is always equal to 1 and it is not displayed, corresponding category probability  $P(Y = k)$  is displayed with black color.

Highest score  
4

a - discrimination  
-4 -2 0 2 4

b1 - difficulty  
-4 -2 0 2 4

b2 - difficulty  
-4 -2 0 2 4

b3 - difficulty  
-4 -2 0 2 4

b4 - difficulty  
-4 -2 0 2 4

Equations

$$\pi_k = P(Y \geq k | \theta, a, b_k) = \frac{e^{a(\theta - b_k)}}{1 + e^{a(\theta - b_k)}}$$

$$\pi_k = P(Y = k | \theta, a, b_1, b_2) = \pi_k^* - \pi_{k+1}^*$$

$$E(Y | \theta, a, b_1, \dots, b_k) = \sum_{k=1}^K k \pi_k^*$$

Plots

**Cummulative probabilities**

**Category probabilities**

**Expected item score**

Ability: 0.28  
Category probability: 0.179207113  
 $\pi(Y = 4)$

ShinyItemAnalysis | Test and item analysis | Version 1.2.7  
© 2018 Patrycja Martinková, Adèle Diabouine, Ondřej Leifer and Jakub Myska

HRM 5776

# Dichotomous IRT Models - check your understanding

ShinyItemAnalysis Test and Item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

### Exercise 1

Consider the following 2PL items with parameters  
 Item 1:  $a = 2.5, b = -0.5$   
 Item 2:  $a = 1.5, b = 0$

For these items fill the following exercises with an accuracy of up to 0.05. Then click on **Submit answers** button. If you need a hint, click on blue button with question mark.

- Sketch item characteristic and information curves **?** **X**
- Calculate probability of correct answer for latent abilities  $\theta = -2, -1, 0, 1, 2$  **?**

Item 1:	$\theta = -2$ ✓ <input type="text" value="0"/>	$\theta = -1$ <b>X</b> <input type="text" value="0.3"/>	$\theta = 0$ <b>X</b> <input type="text" value="0.4"/>	$\theta = 1$ <b>X</b> <input type="text" value="0.9"/>	$\theta = 2$ <b>X</b> <input type="text" value="0.9"/>
Item 2:	$\theta = -2$ ✓ <input type="text" value="0"/>	$\theta = -1$ ✓ <input type="text" value="0.2"/>	$\theta = 0$ ✓ <input type="text" value="0.5"/>	$\theta = 1$ ✓ <input type="text" value="0.6"/>	$\theta = 2$ <b>X</b> <input type="text" value="0.9"/>
- For what level of ability  $\theta$  are the probabilities equal? **?** **X**
- Which item provides more information for weak ( $\theta = -2$ ), average ( $\theta = 0$ ) and strong ( $\theta = 2$ ) students? **?**

$\theta = -2$	<input type="radio"/> Item 1	<input checked="" type="radio"/> Item 2	<b>X</b>
$\theta = 0$	<input checked="" type="radio"/> Item 1	<input type="radio"/> Item 2	✓
$\theta = 2$	<input checked="" type="radio"/> Item 1	<input type="radio"/> Item 2	<b>X</b>

Help  
 Look at the figure on the right side.  
 Which curve does have larger value for desired level of ability  $\theta$ ?

40% correct. Try again.

### Exercise 2

Consider now 2 items with following parameters  
 Item 1:  $a = 1.5, b = 0, c = 0, d = 0$   
 Item 2:  $a = 1.5, b = 0, c = 0.2, d = 0$

For these items fill the following exercises with an accuracy of up to 0.05. Then click on **Submit answers** button.

- What is the lower asymptote for items? **?**  
 Item 1:



ShinyItemAnalysis Test and Item analysis | Version 1.2.6  
 © 2016 Patricija Martinkova, Adela Drabizova, Ondrej Leder and Jakub Houdek



HIS 6062

# Selected R Code

- Sample R code may be run and modified in separate R session

```
ShinyItemAnalysis Test and Item analysis About Data Summary Validity Item analysis Regression IRT models DF-Fairness Reports References

Selected R code

library(diffr)
library(mirt)
data(GMAT)
data <- GMAT[, 1:20]

# Model
fit <- mirt(data, model = 1, IRTtype = "2PL", SE = T)
# Item Characteristic Curves
plot(fit, type = "trace", facet_items = F)
# Item Information Curves
plot(fit, type = "infotrace", facet_items = F)
# Test Information Function
plot(fit, type = "infofst")
# Coefficients
coef(fit, simplify = TRUE)
coef(fit, IRTpars = TRUE, simplify = TRUE)
# Item fit statistics
itemfit(fit)
# Factor scores vs Standardized total scores
fs <- as.vector(scores(fit))
sta <- as.vector(scale(apply(data, 1, sum)))
plot(fs = sta)

# You can also use IRT library for IRT models
library(diffr)
library(ltm)
data(GMAT)
data <- GMAT[, 1:20]

# Model
fit <- ltm(data ~ xi, IRT.param = TRUE)
# Item Characteristic Curves
plot(fit)
# Item Information Curves
plot(fit, type = "IIC")
# Test Information Function
plot(fit, type = "TIF")

ShinyItemAnalysis Test and Item analysis | Version 1.2.0
© 2017 Patricija Martinkova, Adela Drabíková, Ondřej Leder and Jakub Houdek
```

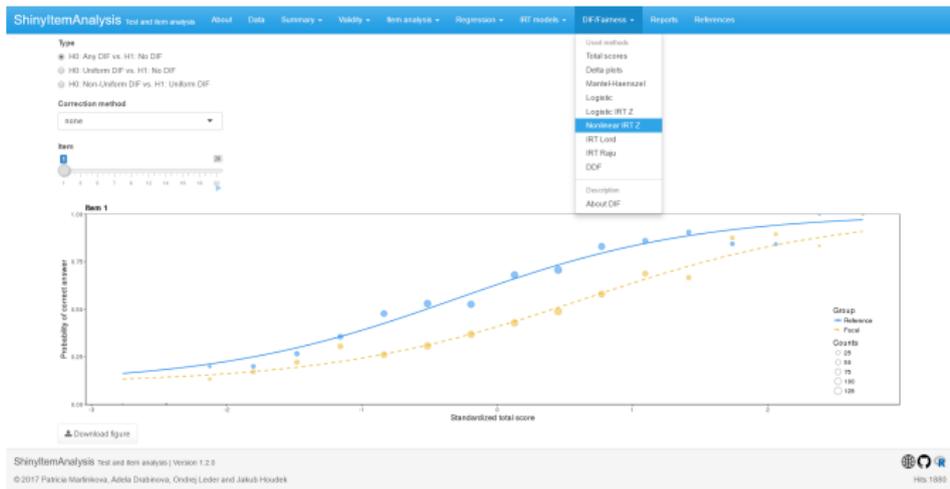
# ShinyItemAnalysis to promote our research

App promotes methods and research of our team:

- Detection of Differential Item Functioning (DIF)
- Detection of Differential Distractor Functioning (DDF)
- Why DIF/DDF analysis should be routine part of test development
- etc.

# Differential Item Functioning (DIF)

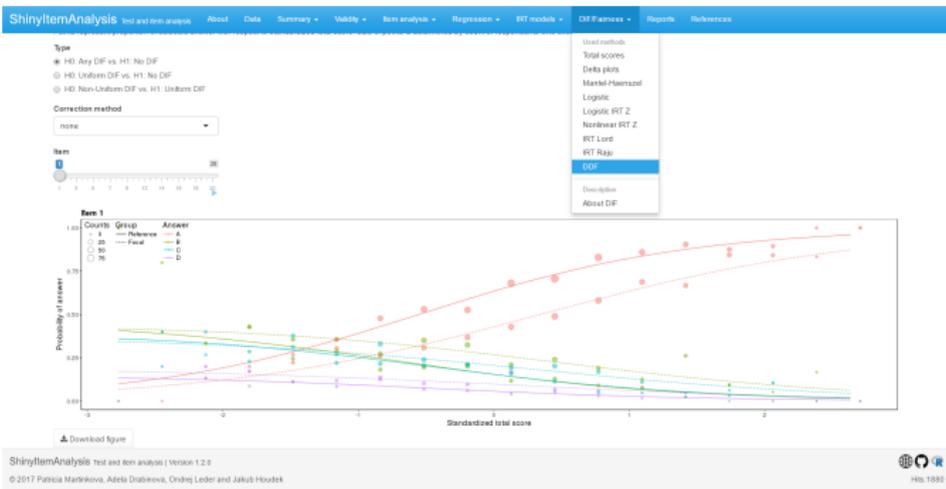
**DIF:** Students from two groups and *with the same underlying latent ability* have different probability of answering the item correctly.



Drabinová & Martinková (2017): Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. *Journal of Educational Measurement*, 54(4), pp. 498-517. doi [10.1111/jedm.12158](https://doi.org/10.1111/jedm.12158)

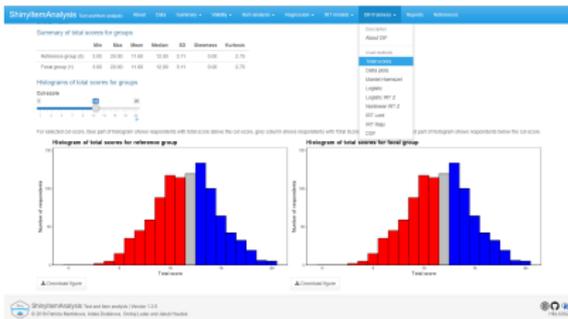
# Differential Distractor Functioning (DDF)

**DDF:** Students from two groups and *with the same underlying latent ability* have different probability of selecting given options.



# Why DIF Analysis Should Be Analyzed Routinely?

- Dataset HCI: significant difference in total score between males and females, yet no DIF item!
- Simulated GMAT data: total scores may have exactly the same distribution, yet there may be DIF present in some items!



Martinková, Drabinová, Liaw, Sanders, McFarland & Price (2017): Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. [doi:10.1187/cbe.16-10-0307](https://doi.org/10.1187/cbe.16-10-0307)

# Routine validation of educational tests

Supporting tool for routine validation of educational tests:

- Upload your own data
- Generate PDF/HTML report
- Local or online version

# Report generation - settings

- Chose methods, customize settings
- Chose report format (PDF/HTML)

The screenshot shows the 'Download report' settings page in the ShinyItemAnalysis application. The page has a blue header with navigation links: 'ShinyItemAnalysis', 'Test and item analysis', 'About', 'Data', 'Summary', 'Validity', 'Item analysis', 'Regression', 'IRT models', 'DIF/fairness', 'Reports', and 'References'. The main content area is titled 'Download report' and includes a 'Settings of report' section. Under 'Format of report', the 'PDF' option is selected. The 'Author' field contains 'Joe Doe' and the 'Dataset' field contains 'HCI dataset'. The 'Content of report' section includes a 'Validity' subsection with 'Correlation structure' checked. The 'Number of clusters' is set to 1, and the 'Clustering method' is 'Ward's'. 'Predictive validity' is also checked. There are two 'Difficulty/discrimination plot' sliders, both set to 0. The 'Distractors plots' section has 'Type' set to 'Combinations' and 'Number of groups' set to 0. The footer contains the ShinyItemAnalysis logo, version information (1.2.7), copyright notice (© 2018 Patricia Martinková, Adela Drabínová, Ondřej Leder and Jakub Houdek), and social media icons for GitHub, R, and a user ID 'HRs 5776'.

# Report generation

- Generate report (run analyses)
- Download report (compile text into HTML/PDF)

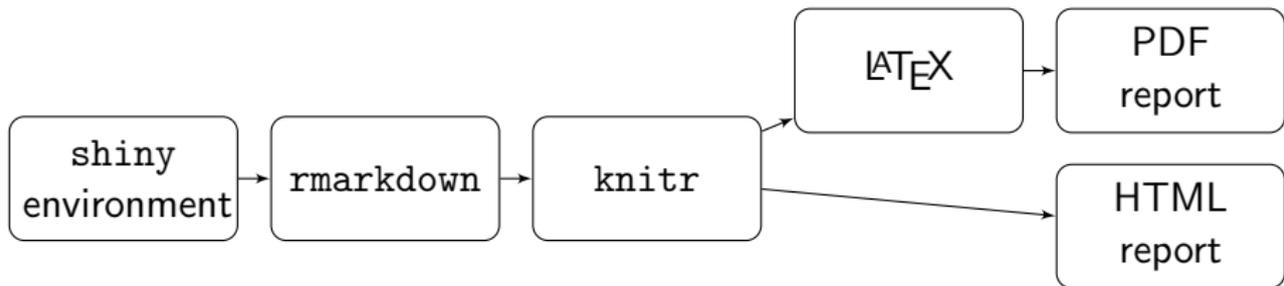
The screenshot shows the 'Reports' section of the ShinyItemAnalysis application. The navigation bar includes: ShinyItemAnalysis, Test and Item analysis, About, Data, Summary, Validity, Item analysis, Regression, IRT models, DIF/fairness, Reports, and References. The 'Reports' section contains several interactive elements:

- Number of groups:** A slider set to 3.
- Which two groups to compare:** A slider set to 3.
- Number of groups (bottom):** A slider set to 3.
- Distractors plots:**
  - Type:  Combinations,  Distractors
- IRT model selection:**
  - None
  - Rasch
  - 1PL
  - 2PL
  - 3PL
  - 4PL
- DIF method selection:**
  - None - histograms by group only
  - Delta plot
  - Logistic regression
  - Multinomial regression
- Delta plot settings:**
  - Threshold:  Fixed,  Normal
  - Item purification
- Logistic regression settings:**
  - Type:  H0: Any DIF vs. H1: No DIF,  H0: Uniform DIF vs. H1: No DIF,  H0: Non-Uniform DIF vs. H1: Uniform DIF
  - Correction method:
  - Item purification
- Multinomial regression settings:**
  - Type:  H0: Any DIF vs. H1: No DIF,  H0: Uniform DIF vs. H1: No DIF,  H0: Non-Uniform DIF vs. H1: Uniform DIF
  - Correction method:
  - Item purification

A 'Loading' spinner is visible in the center. At the bottom, a recommendation states: "Recommendation: Report generation can be faster and more reliable when you first check sections of intended contents. For example, if you wish to include a 3PL IRT model, you can first visit IRT models section and 3PL subsection." Below this is a 'Generate report' button. The footer includes the ShinyItemAnalysis logo, version 1.2.7, copyright information, and a 'Creating content' progress indicator.



# Report generation workflow



- shiny provides a user interface
- rmarkdown for creating templates for PDF/HTML report generation
- knitr for compiling R markdown syntax into HTML/PDF
- T<sub>E</sub>X for creating PDF reports (latest distribution of L<sub>A</sub>T<sub>E</sub>X is needed)

## Conclusion and Discussion

ShinyItemAnalysis is an R package and online application for interactive and step-by-step analysis of educational tests. It is useful for:

- TEACHING of psychometrics and educational measurement
  - offers example datasets, upload of new datasets
  - visualization, interpretation of results
  - sample R Code
- ROUTINE VALIDATION OF EDUCATIONAL TESTS
  - generates extensive reports for supplied data

ShinyItemAnalysis also promotes our RESEARCH in DIF/DDF detection

<https://shiny.cs.cas.cz/ShinyItemAnalysis/>

`martinkova@cs.cas.cz`



Thank you for your attention!

`www.cs.cas.cz/martinkova`

`martinkova@cs.cas.cz`



This research was supported by Czech Science Foundation under grant GJ15-15856Y

- Martinková, Drabinová, Leder & Houdek (2017). ShinyItemAnalysis: Test and Item Analysis with Shiny.  
<https://shiny.cs.cas.cz/ShinyItemAnalysis/>  
<https://CRAN.R-project.org/package=ShinyItemAnalysis>
- Martinková, Drabinová & Houdek (2017). ShinyItemAnalysis: Analýza přijímacích a jiných znalostních či psychologických testů. TESTFÓRUM, č.9, str. 16-35. doi 10.5817/TF2017-9-129
- McFarland, Price, Wenderoth, Martinková, et al. (2017). Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35. doi 10.1187/cbe.16-10-0305
- Martinková, Drabinová, Liaw, Sanders, McFarland & Price (2017). Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. doi 10.1187/cbe.16-10-0307
- Drabinová & Martinková (2017). Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement, 54(4), pp. 498-517. doi 10.1111/jedm.12158