

Teaching Psychometrics and Analysing Educational Tests with ShinyItemAnalysis

Patricia Martinkova et al.
useR!2017, Brussels, July 7

Department of Statistical Modelling, Institute of Computer Science
Czech Academy of Sciences

Joint work...



Patricia Martinkova¹



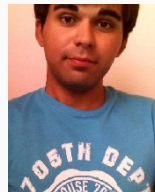
Adela Drabinova^{1,2}



Jakub Houdek^{1,3}



Ondrej Leder²



Lubomir Stepanek^{3,4}

¹Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences

²Department of Probability and Mathematical Statistics, Charles University, Prague

³Faculty of Informatics and Statistics, University of Economics, Prague

⁴First Faculty of Medicine, Charles University, Prague

Acknowledgement: Czech Science Foundation grant GJ15-15856Y

Table of contents

1. Introduction
2. ShinyItemAnalysis for TEACHING
3. ShinyItemAnalysis for RESEARCH
4. ShinyItemAnalysis for PRODUCTION
5. Conclusion

Introduction

Motivation



- Admission tests development
 - to select the best applicants
 - need for valid, reliable and fair tests, well functioning items
- Development of tests for classroom use
 - to keep students interested in the subject
 - to avoid discrimination of minorities
- To teach psychometric models and concepts
- To promote our own psychometric research

Need for user-friendly and freely available tool

ShinyItemAnalysis

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

Available as:

- R package
 - Version 1.2.0 now on 
 - Newest version on 
- Online shiny application
 - ICS server in Prague, CZ:

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

- shinyapps.io:

<https://patriciamar.shinyapps.io/ShinyItemAnalysis/>

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 [GNAT](#) from R [diRKE](#) package is used. Other three datasets are available: [GNAT2](#) and [Medical 20 DIF](#) from [diRKE](#) package and [Medical 100](#) 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](#). In case of busy server you can try other mirrors: [Charles University](#) or [shinyapps.io](#).

Version


Current version of [ShinyItemAnalysis](#) available on [CRAN](#) is 1.2.0. Version available [online](#) is 1.2.0. The newest development version available on [GitHub](#) is 1.2.0.

See also older versions: [0.1.0](#), [0.2.0](#), [1.0.0](#), [1.1.0](#).


Authors and contributors




Patricia Martinkova




Adela Drabinova



Ondrej Leder




Jakub Houdek



Lubomir Stepanek

ShinyItemAnalysis | test and item analysis | Version 1.2.0



HIS.1871

ShinyItemAnalysis **for TEACHING**

ShinyItemAnalysis for Teaching

Who do we teach:

- students of educational measurement
- faculties, university stakeholders

Some helpful features:

- Interactive plots
- Example datasets, allows to upload own data
- Shows model equations, provides interpretation of results
- Allows to download plots, generate extensive reports
- Provides sample R code

Datasets

- Four toy datasets are available
- Allows to upload one's own dataset

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

Data

For demonstration purposes, 20-item dataset **GMAT** and dataset **GMATkey** from R **difxlr** package are used. On this page, you may select one of four datasets offered from **difxlr** and **ShinyItemAnalysis** packages or you may upload your own dataset (see below). To return to demonstration dataset, refresh this page in your browser (**F5**).

Used dataset **GMAT** (Martinkova, et al, 2017) is generated based on parameters of real Graduate Management Admission Test (GMAT) data set (Kingston et al., 1985). However, first two items were generated to function differently in uniform and non-uniform way respectively. The data set represents responses of 2,000 subjects (1,000 males, 1,000 females) to multiple-choice test of 20 items. The distribution of total scores is the same for both groups. See [Martinkova, et al \(2017\)](#) for further discussion.

Dataset **GMAT2** (Drabinoва & Martinkova, 2016) is also generated based on parameters of GMAT (Kingston et al., 1985) from R **difxlr** package. Again, first two items were generated to function differently in uniform and non-uniform way respectively. The data set represents responses of 1,000 subjects (500 males, 500 females) to multiple-choice test of 20 items.

Dataset **Medical 20 DIF** (Drabinoва & Martinkova, 2016) is a subset of real admission test to medical school from R **difxlr** package. First item was previously detected as functioning differently. The data set represents responses of 1,407 subjects (484 males, 923 females) to multiple-choice test of 20 items. For more details of item selection see Drabinoва & Martinkova (2016).

Dataset **Medical 100** is a real data set of admission test to medical school from R **ShinyItemAnalysis** package. The data set represents responses of 2,392 subjects (750 males, 1,633 females and 9 subjects without gender specification) to multiple-choice test of 100 items.

Select dataset

GMAT

Upload your own datasets

Main **data** file should contain responses of individual students (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 students in main dataset. If the group is not provided then it won't be possible to run DIF and DDF 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 students 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 below.

Missing values in scored dataset are by default evaluated as 0. If you want to keep them as missing, check box below.

Choose data (csv file)

Browse...

No file selected

Choose key (csv file)

Browse...

No file selected

Choose groups for DIF (optional)

Browse...

No file selected

Choose criterion variable (optional)

Browse...

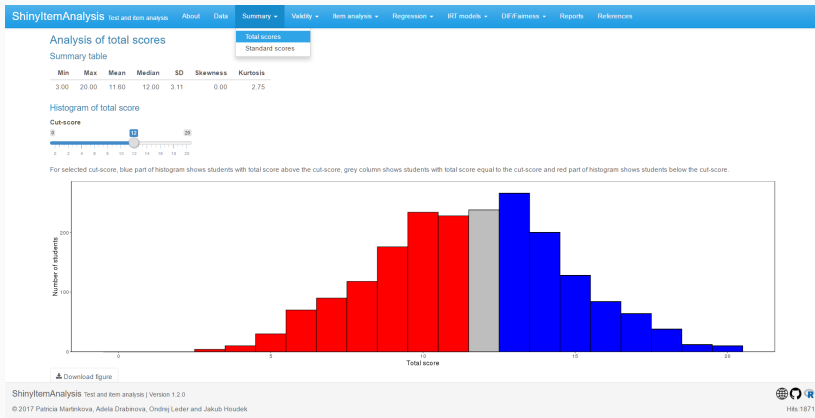
No file selected

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

Hits: 1871

Summary of Total Scores

- Summary statistics
- Interactive histogram



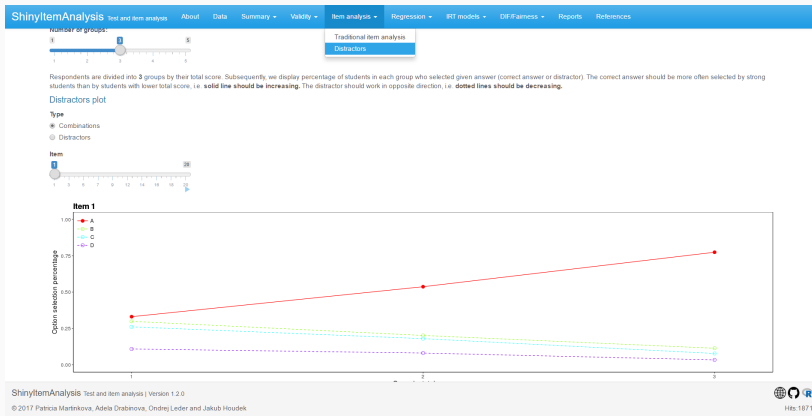
Traditional Item Analysis

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



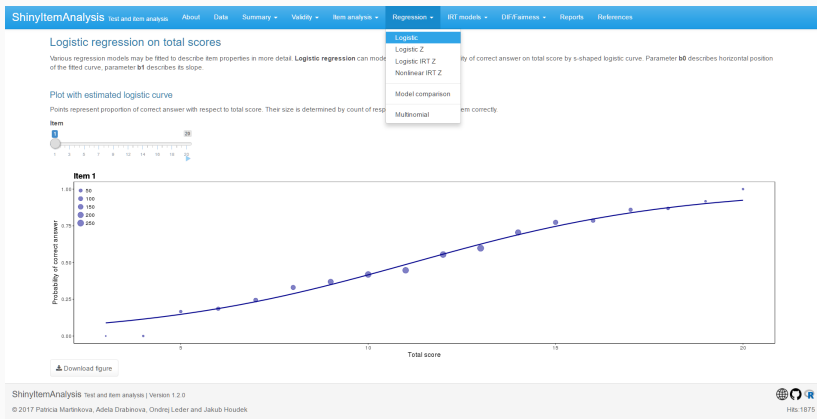
Distractor Analysis

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



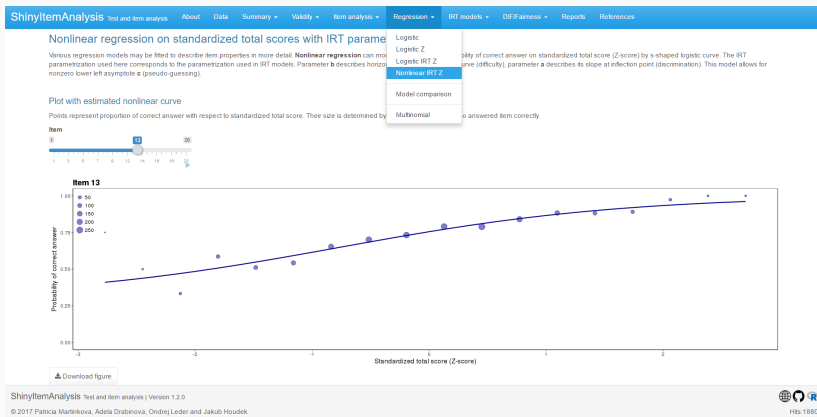
Logistic Regression

- Displays probability of correct answer by total score
- Parameterization can be changed (Z scores, IRT parameterization)



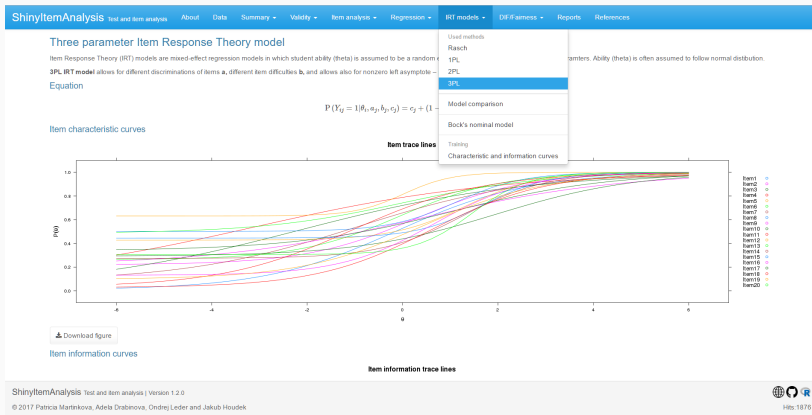
Nonlinear Regression

- Allows for guessing (and inattention)



IRT Models

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



Selected R Code

- Sample R code may be run in separate R session

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

Selected R code

```

library(difR)
library(mirt)
data(GHWT)
data <- GHWT[, 1:20]


# Model
fit <- mirt(data, model = 1, itemtype = "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 = "infoSI")
# 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(fiscorrel(fit))
sts <- as.vector(scale(apply(data, 1, sum)))
plot(fs ~ sts)

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

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

```

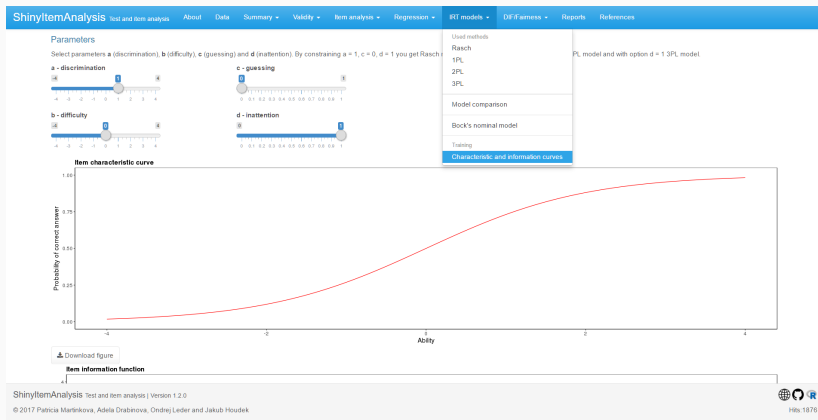
ShinyItemAnalysis Test and item analysis | Version 1.2.0
© 2017 Patricia Martinkova, Adele Drabnova, Ondrej Leder and Jakub Houdek



HRS 1880

IRT Models - construct your own item

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



ShinyItemAnalysis **for RESEARCH**

ShinyItemAnalysis for Research

To widespread novel methodology:

- write scientific paper
- ... and provide R code
- ... and provide dataset
- ... and write accompanying R package
- ... and **prepare shiny application**

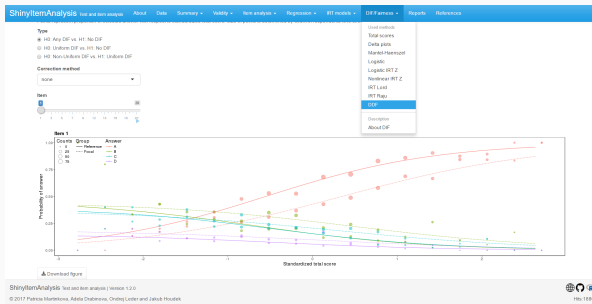
Differential Item Functioning (DIF)

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



Differential Distractor Functioning (DDF)

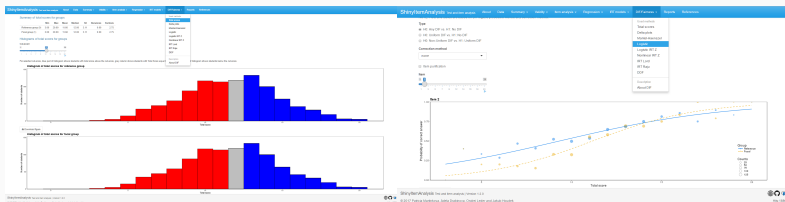
DDF: Students from two groups and *with the same underlying latent ability* have different probability of selecting given options.



To learn more, go to
 Adela Drabinova: `difNLR`: Detection of potential gender/minority bias with
 extensions of logistic regression (Thursday, July 6) [► UseR!2017](#)

Why DIF Analysis Should Be Analyzed Routinely?

- Simulated GMAT data: total scores may have exactly the same distribution, yet there may be DIF present in some items!



Martinkova et al. (2017): Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. [Online](#)

ShinyItemAnalysis for PRODUCTION

Report Generation

- Upload your own data
- Generate PDF/HTML report

ShinyItemAnalysis
test and item analysis
About
Data
Summary
Validity
Item analysis
Regression
IRT models
DIF fairness
Reports
References

Content of report

Reports by default contain summary of total scores, item analysis, distractors plots for each item and multinomial regression plots for each item.

Correlation structure

- ☒ None
- ☐ Corplot + Screenshot

Distractors plot

Type

- ☒ Combinations
- ☐ Distractors

IRT model selection

- ☐ None
- ☐ Rasch
- ☒ 1PL
- ☐ 2PL
- ☐ 3PL

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

none

- ☐ 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

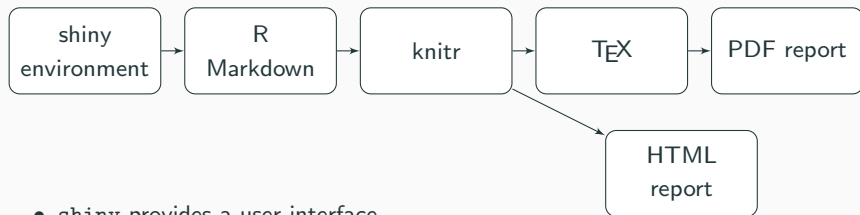
none

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.

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




Hits: 1876

Report Generation



- shiny provides a user interface.
- rmarkdown for creating templates for PDF/HTML report generation.
- knitr for compiling R markdown syntax into HTML/PDF
- T_EX for creating PDF reports (latest distribution of T_EX is needed).

To learn more see UseR!2017 poster by Jakub Houdek et al. [► Online](#)

Conclusion

Conclusion

ShinyItemAnalysis is a shiny 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
- PRODUCTION
 - 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`



References

- Martinkova, Drabinova, Leder & Houdek (2017): ShinyItemAnalysis: Test and Item Analysis with Shiny.
<https://shiny.cs.cas.cz/ShinyItemAnalysis/>
<https://CRAN.R-project.org/package=ShinyItemAnalysis>
- Drabinova, Martinkova & Zvara (2017): difNLR: Detection of Dichotomous DIF by Non-linear Regression.
<https://CRAN.R-project.org/package=difNLR>
- Drabinova & Martinkova (under review): Detection of DIF Based with Non-Linear Regression: Non-IRT Approach Accounting for Guessing.
- Martinkova, Drabinova, 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.
www.lifescied.org/content/16/2/rm2
- Martinkova, Drabinova & Houdek (2017): ShinyItemAnalysis: Analysis of admission and other educational and psychological tests. Testforum, Accepted.