ShinyItemAnalysis: Test and Item Analysis

Acknowledgements

Project was supported by Czech Science Foundation grant GJ15-15856Y 'Estimation of psychometric measures as part of admission test development' and by Charles University grant PRIMUS/17/HUM/11 'Center for Educational Measurement and Psychometrics (CEMP)'.

Preface

Educational and psychological testing is present in many areas of our everyday life, including assessing academic achievement, certifying qualifications and proficiency, or assessing one's fatigue, depression or pain. In many situations, testing is a key moment of people's life with long-reaching consequences, such as in university admissions or hiring of new employees. For these reasons, assessments that are used to measure ability, knowledge or other latent traits need to produce valid, reliable and fair scores.

While many methodological books exits that present the methodology for test validation, often practical examples of item and test analysis are missing, or are presented in commercial software which may be costly and thus unavailable. As an alternative, freely available statistical software `R` has been present for many years and many `R` packages have been developed to cover general psychometric concepts or specific psychometric topics. However for those who are new to `R`, it may be hard to overcome the initial burden of `R` code-based environment.

This book introduces selected topics in psychometrics and covers test and item validation with practical examples in `ShinyItemAnalysis` (Martinková, Drabinová, Leder, & Houdek, 2018). `ShinyItemAnalysis` is an `R` package and a freely available online application which provides interactive online environment to support teaching of psychometric concepts and test development with `R`, and to enforce routine validation of educational and psychological measurement worldwide.

The book is prepared as manual for `ShinyItemAnalysis` and explains methodology as well as practical features of each of its sections. It can thus serve those who use `ShinyItemAnalysis` - teachers who assess knowledge of their students, educators who develop new assessments, psychologists and researchers who use established or newly developed instruments in their projects, or even university stakeholders who want to introduce routine validation in their admission tests or classroom assessments. Inside the interactive environment of `ShinyItemAnalysis`, we introduce the reader also to examples in `R`programming language. Individual sections include Selected R code, as well as Exercises. The book can thus be very well used also in graduate courses of measurement and psychometrics and can serve as a gentle introduction to these topics in (or without) `R`.

Recent news can be found at www.ShinyItemAnalysis.org. The `ShinyItemAnalysis` application and `R` package were created with the aim to strengthen understanding of psychometric concepts and to support teaching of these concepts, to empower routine analysis of tests and also to present novel psychometric research. We hope the book you are reading will help fulfill these goals.

Contents

Preface **ii**

1 Introduction **1**
1.1 Psychometrics, Measurement, Test development . . . 1
1.2 Software . . . 1
1.3 Book overview . . . 1

2 Getting started **3**
2.1 Getting started with ShinyItemAnalysis . . . 3
2.2 Getting started with R . . . 3

3 Measurement data **6**
3.1 Toy data . . . 6
3.2 Data upload . . . 6
3.3 Data summary and exploration . . . 7
3.4 Total scores . . . 7
3.5 Selected `R` code . . . 9
3.6 Exercises . . . 10

4 Reliability **12**
4.1 Reliability estimates based on multiple administrations . . . 13
4.1.1 Test-retest reliability . . . 13
4.1.2 Reliability of parallel forms . . . 13
4.2 Reliability estimates making use of composite measurements . . . 13
4.2.1 Split-half coefficient . . . 13
4.2.2 Cronbach's alpha . . . 13
4.2.3 Revelle's beta . . . 14
4.2.4 McDonald's omega . . . 14
4.3 Further reliability estimates . . . 14
4.4 Selected `R` code . . . 14
4.5 Exercises . . . 14

5 Validity **15**
5.1 Construct validity . . . 15
5.2 Criterion validity . . . 16
5.3 Selected `R` code . . . 16
5.4 Exercises . . . 17

6 Traditional item analysis **18**
6.1 Selected `R` code . . . 18
6.2 Exercises . . . 20

7 Regression models for description of item properties **21**
7.1 Selected `R` code . . . 21
7.2 Exercises . . . 26

8 Item Response Theory models **27**
8.1 Selected `R` code . . . 28
8.2 Exercises . . . 29

9 Differential item functioning 30
9.1 Selected R code . . . 30
9.2 Exercises . . . 32

10 Further topics in psychometrics 33

11 Reports generation 34

A Appendices 37
A1 Installation of R and RStudio . . . 37
A1.1 Windows . . . 37
A1.2 Mac OS X . . . 38
A2 Installation of `ShinyItemAnalysis` . . . 38
A3 Installation of TEXdistribution . . . 39

References 40

Acronyms 42

Chapter 1

Introduction

1.1 Psychometrics, Measurement, Test development

Psychometrics is a field of study concerned with the theory and technique of psychological, educational and behavioral measurement. It is concerned with objective measurement (testing, assessment) of skills, knowledge, abilities, educational achievement, attitudes, personality traits and other.

As outlined in the Standards for educational and psychological testing AERA, APA and NCME 2014, assessments that are used to measure students' ability or knowledge need to produce valid, reliable and fair scores. To achieve these standards, many aspects of test development need to be taken care of (Haladyna & Downing, 2011). While core aspects of measurement are the same no matter the type of measurement, some specific topics may arise in different areas (see (Brennan, 2006) for overview on topic of Educational measurement).

Psychometric analysis and routine validation of tests is usually present in development of standardized tests, especially those used as admission tests to higher education (SAT, ACT, TOEFL, MCAT, BMAT, etc.), international large scale assessments (PISA, TIMSS, PIRLS), or annual testing of students performed in some countries and states (e.g. ...). Testing companies developing and administering these tests (e.g. College Board, ETS, etc.) often have departments or units taking care of item and test properties.

Test analysis is nowadays being more present also in regions and scientific areas where psychometrics does not have a long tradition. Complex test analyses can be found in development and validation of conceptual assessments (see e.g McFarland et al., 2017) - tests of students' conceptual understanding of key topics in certain fields.

1.2 Software

This book introduces selected topics in psychometrics with practical examples in `ShinyItemAnalysis` (Martinková et al., 2018), which provides gentle introduction to psychometric analyses. It uses powers of the many psychometric `R` packages in user-friendly interface. Many interactive features are present to support understanding of presented concepts.

1.3 Book overview

Several topics which we find most important parts of test analysis, "base stones", are covered in individual chapters. **Chapter 2** helps the reader to get started with the software.

Chapter 3 offers introduction to measurement data.

Chapter 4 provides introduction to measurement error and various ways to get proofs of reliability of the test.

Chapter 5 describes analyses which may help to provide proofs of test validity.

In **Chapter 6**, traditional item analysis is provided, including various item characteristics based on ratios or correlations, mainly describing item difficulty, discrimination power, guessing or response rate. Detailed distractor analysis is presented to provide better understanding to functioning of all offered options in multiple-choice tests for low as well as high performing students.

Chapter 7 introduces regression models for description of item properties.

Chapter 8 explains various item response theory models for binary, ordinal as well as nominal data describing tests including dichotomous, partial credit, Likert-scale, multiple-choice and other types of items.

Chapter 9 covers topic of differential item functioning and presents various methods for detection of DIF.

Further topics not yet covered by the `ShinyItemAnalysis`application are described in **Chapter 10**.

Chapter 11 describes how reports may be generated with `ShinyItemAnalysis`.

Appendices provide detailed guidance about how to install `R`, LaTeX, etc.

In summary, the `ShinyItemAnalysis` software as well as this book cover the basic topics in psychometrics and test analysis to provide a solid base stone. Moreover, by introducing these concepts in R, we aim to open the door to much wider and rich methodology for quantitative analysis in education, psychology and sociology to advance quantitative methodology in the behavioral sciences.

Chapter 2

Getting started

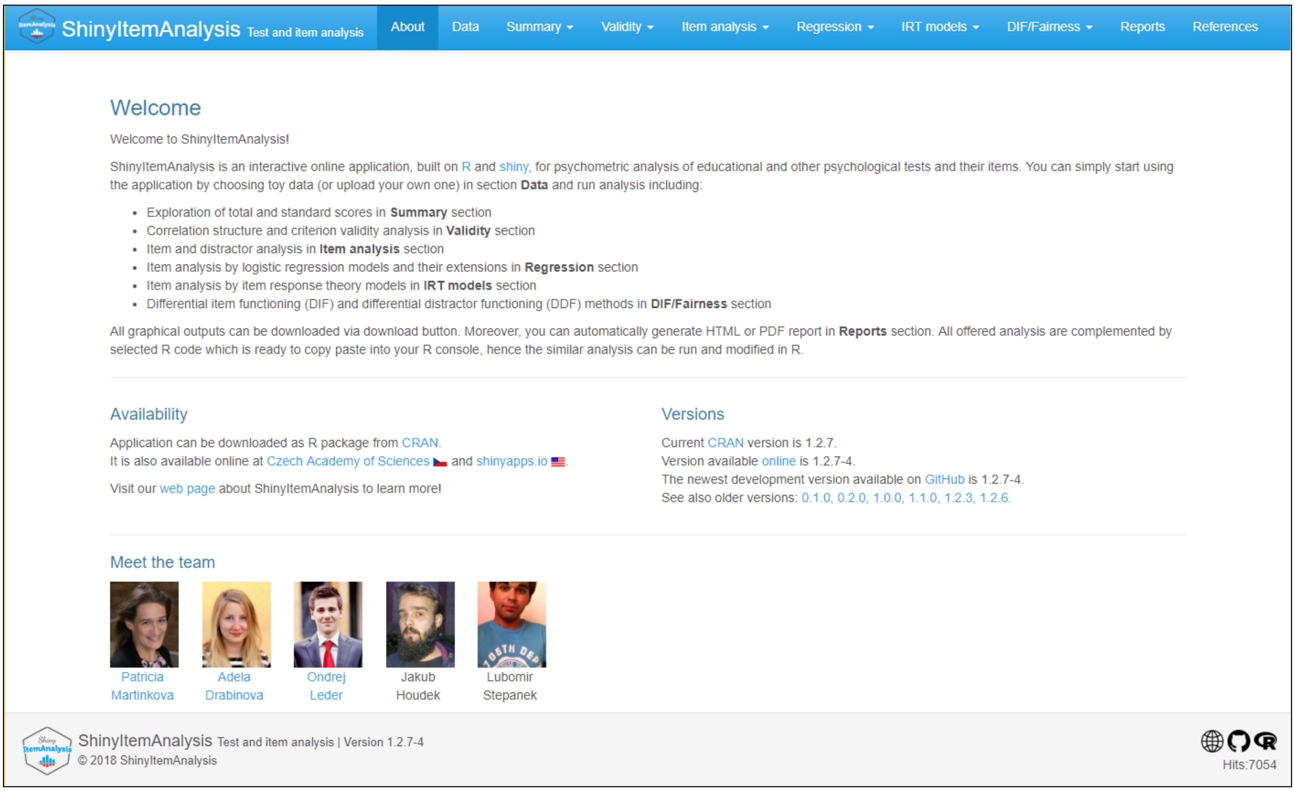
2.1 Getting started with ShinyItemAnalysis

ShinyItemAnalysis online application is available at

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

Other mirrors are specified at http://www.ShinyItemAnalysis.org. As we discuss below, it is also possible to run the application locally.

Intro page (Figure 2.1) includes general information about the application. Various tools are included in separate tabs with logical ordering into separate sections.



Welcome

Welcome to ShinyItemAnalysis!

ShinyItemAnalysis is an interactive online application, built on R and shiny, for psychometric analysis of educational and other psychological tests and their items. You can simply start using the application by choosing toy data (or upload your own one) in section **Data** and run analysis including:

- Exploration of total and standard scores in **Summary** section
- Correlation structure and criterion validity analysis in **Validity** section
- Item and distractor analysis in **Item analysis** section
- Item analysis by logistic regression models and their extensions in **Regression** section
- Item analysis by item response theory models in **IRT models** section
- Differential item functioning (DIF) and differential distractor functioning (DDF) methods in **DIF/Fairness** section

All graphical outputs can be downloaded via download button. Moreover, you can automatically generate HTML or PDF report in **Reports** section. All offered analysis are complemented by selected R code which is ready to copy paste into your R console, hence the similar analysis can be run and modified in R.

Availability

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

Visit our web page about ShinyItemAnalysis to learn more!

Versions

Current CRAN version is 1.2.7.
Version available online is 1.2.7-4.
The newest development version available on GitHub is 1.2.7-4.
See also older versions: 0.1.0, 0.2.0, 1.0.0, 1.1.0, 1.2.3, 1.2.6.

Meet the team

Patricia Martinkova
Adela Drabinova
Ondrej Leder
Jakub Houdek
Lubomir Stepanek

ShinyItemAnalysis Test and item analysis | Version 1.2.7-4
© 2018 ShinyItemAnalysis

Figure 2.1: Intro page.

2.2 Getting started with R

While you may read this book and try most of the exercises without even opening R, to get most of the book, we recommend to also download and use R. Two main benefits you get from using R are as follows: First, you may then use the application locally, which may become faster and more efficient than the online version. Second, besides running the online application locally in your `R` , you may instead use `R` console to try `R` examples provided inside `ShinyItemAnalysis` and modify them for your purposes.

`R` can be installed from

https://cran.r-project.org/

Detailed instructions can be found in appendix. [Consider mentioning RStudio.] Basic introductory of `R` can be given e.g. by Paradis (2002).

Once you have your `R` installed, it is easy to install the last stable version of `ShinyItemAnalysis` package from **CRAN** by writing the following code

```
install.packages("ShinyItemAnalysis")
```

The newest development version can be instead downloaded from **GitHub** (with `devtools` package) using the following lines

```
devtools::install_github("patriciamar/ShinyItemAnalysis")
```

Once the `ShinyItemAnalysis` package is installed it can be loaded and the online application can be run locally:

```
library(ShinyItemAnalysis)
startShinyItemAnalysis()
```

Main function `startShinyItemAnalysis()` launches an interactive application as described above.

`ShinyItemAnalysis` uses several `R` packages to provide wide palette of psychometric tools to analyze data (see Table 2.1). Overview of many other psychometric libraries is provided on Psychometric CRAN Task (Mair, 2018).

One can also use `R` console to try selected `R` code provided in `ShinyItemAnalysis`, or to modify the code as needed.

Table 2.1: R packages used for developing ShinyItemAnalysis.

R package	Citation	Title
corrplot	(Wei & Simko, 2017)	Visualization of a correlation matrix
CTT	(Willse, 2018)	Classical test theory functions
data.table	(Dowle & Srinivasan, 2017)	Extension of data.frame
deltaPlotR	(Magis & Facon, 2014)	Identification of dichotomous differential item functioning using Angoff's delta plot method
difNLR	(Drabinová, Martinková, & Zvára, 2018)	DIF and DDF detection by non-linear regression models
difR	(Magis, Beland, Tuerlinckx, & De Boeck, 2010)	Collection of methods to detect dichotomous differential item functioning
DT	(Xie, 2018)	A wrapper of the JavaScript library 'datatables'
ggplot2	(Wickham, 2016)	Create elegant data visualisations using the grammar of graphics
gridExtra	(Auguie, 2017)	Miscellaneous functions for "grid" graphics
knitr	(Xie, 2015)	A general-purpose package for dynamic report generation in R
lattice	(Sarkar, 2008)	Trellis graphics for R
latticeExtra	(Sarkar & Andrews, 2016)	Extra graphical utilities based on lattice
lme4	(Bates, Mächler, Bolker, & Walker, 2015)	Linear mixed-effects models using Eigen and S4
ltm	(Rizopoulos, 2006)	Latent trait models under IRT
MASS	(Venables & Ripley, 2002)	Support functions and datasets for Venables and Ripley's MASS
Matrix	(Bates & Maechler, 2017)	Sparse and dense matrix classes and methods
mirt	(Chalmers, 2012)	Multidimensional item response theory
moments	(Komsta & Novomestky, 2015)	Moments, cumulants, skewness, kurtosis and related tests
msm	(Jackson, 2011)	Multi-state Markov and hidden Markov models in continuous time
multilevel	(Bliese, 2016)	Multilevel functions
nlme	(Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017)	Linear and nonlinear mixed effects models
nnet	(Venables & Ripley, 2002)	Feed-forward neural networks and multinomial log-linear models
plotly	(Sievert et al., 2017)	Create interactive web graphics via 'plotly.js'
polycor	(Fox, 2016)	Polychoric and polyserial correlations
psych	(Revelle, 2018)	Procedures for psychological, psychometric, and personality research
psychometric	(Fletcher, 2010)	Applied psychometric theory
RColorBrewer	(Neuwirth, 2014)	Colorbrewer palettes
reshape2	(Wickham, 2007)	Flexibly reshape data: A reboot of the reshape package
rmarkdown	(Allaire et al., 2017)	Dynamic documents for R
shiny	(Chang, Cheng, Allaire, Xie, & McPherson, 2017)	Web application framework for R
shinyBS	(Bailey, 2015)	Twitter bootstrap components for shiny
shinydashboard	(Chang & Borges Ribeiro, 2018)	Create dashboards with shiny
shinyjs	(Attali, 2018)	Easily improve the user experience of your shiny apps in seconds
stringr	(Wickham, 2018)	Simple, consistent wrappers for common string operations
WrightMap	(Irribarra & Freund, 2014)	IRT item-person map with 'conquest' integration
xtable	(Dahl, 2016)	Export tables to LaTeXor HTML

Chapter 3

Measurement data

Throughout the book, we will be working with several measurement datasets. `ShinyItemAnalysis` also allows the users to upload their own datasets. These mostly contain responses of students (in rows) to test items (columns). Responses may be binary (i.e., true/false, or 1/0), ordinal (e.g., on Likert scale 1−2−3−4−5) or nominal (e.g., A−B−C−D for multiple-choice items). [consider mentioning mix of item formats]

Some further respondent covariates may be present, e.g. group membership: gender, ethnicity, etc. Besides, some criterion variable may be present, such as repondents' IQ, their future study success, study Grade Point Average (GPA), etc.

3.1 Toy data

In `ShinyItemAnalysis`, five training datasets may be uploaded using the **Select dataset** button in section **Data**:

GMAT is a simulated dataset from `ShinyItemAnalysis` `R` package. The dataset represents responses of 2,000 subjects (1,000 males, 1,000 females) to multiple-choice test of 20 items. The answers were generated using parameters of real Graduate Management Admission Test (GMAT) (Kingston, Leary, & Wightman, 1985). The distribution of total scores is the same for both groups. However, first two items were manipulated to function differently for the two groups. See Martinková et al. (2017) for further discussion. GMAT dataset also containts simulated continuous criterion variable.

Similarly, **GMAT2** is a simulated dataset based on parameters of real GMAT (Kingston et al., 1985) from `difNLR` `R` package (Drabinová et al., 2018). The dataset represents responses of 1,000 subjects (500 males, 500 females) to multiple-choice test of 20 items. Also in this dataset, the first two items were simulated to function differently in uniform and non-uniform way respectively.

Medical 100 is a real dataset of admission test to medical school from `ShinyItemAnalysis` `R` 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. Medical 100 contains criterion variable - indicator whether student remained in the study after one year or not.

MSAT-B is a subset of real Medical School Admission Test in Biology (MSAT-B) in Czech Republic from `difNLR` `R` package (Drabinová et al., 2018). The dataset represents responses of 1,407 subjects (484 males, 923 females) to selection of 20 multiple-choice items. First item was previously detected as functioning differently for the two genders. For more details on this dataset, see Drabinová and Martinková (2017).

HCI (McFarland et al., 2017) is a real dataset of Homeostasis Concept Inventory (HCI) offered by `R` package `ShinyItemAnalysis`. The dataset represents responses of 651 subjects (405 males, 246 females) to multiple-choice test of 20 items. HCI contains criterion variable - indicator whether student plans to major in the life sciences.

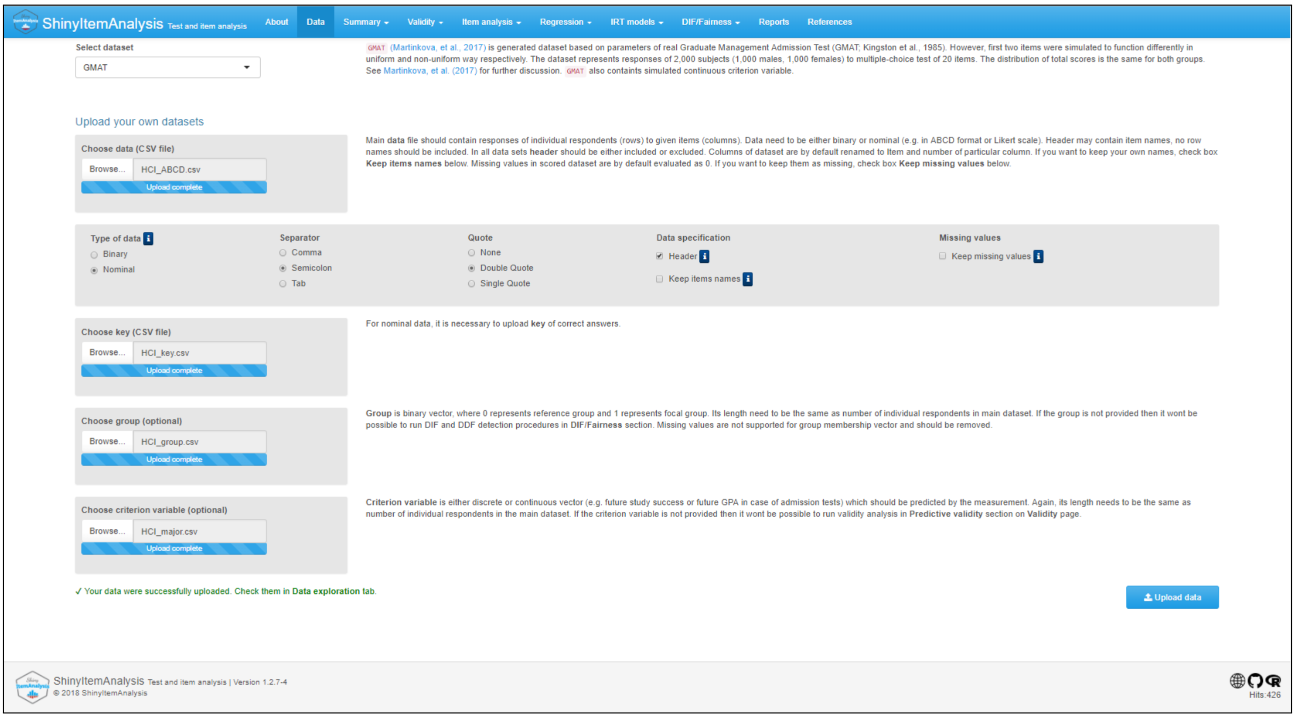
3.2 Data upload

Own data may be uploaded as csv files and previewed in the **Data** section. Main data file should contain responses of individual respondents (rows) to given items (columns). Data need to be either binary or nominal (e.g., in A-B-C-D format). Ordinal data (such as Likert scale) are currently treated as nominal.

Individual items need to be separated by comma, semicolon or tab, this is specified in check box "Separator". Around each reponse value, double-quote or quote may be present (this is often typical in nominal data and it may be specified using check box "Quote"). Header may contain item names, no row names should be included. In all data sets header should be either included or excluded, this is specified by check box "Header". Columns of dataset are by default renamed to Item and number of particular column, however, keeping original names may be forced using check box "Keep items names". Missing values in scored dataset are by default evaluated as 0, however treating them as missing may be forced using check box "Keep missing values".

For nominal data, it is necessary to upload key of correct answers. Group vector may also be included, this is a binary 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. Missing values are not supported for group membership vector and should be removed. Finally, criterion variable can be included, this is either discrete or continuous vector (e.g. future study success or future GPA in case of admission tests). Again, its length needs to be the same as number of individual respondents in the main dataset.

To replicate examples involving HCI dataset (McFarland et al., 2017), csv files for upload are provided on [include link]



Select dataset: GMAT

GMAT (Martinkova, et al., 2017) is generated dataset based on parameters of real Graduate Management Admission Test (GMAT; Kingston et al., 1985). However, first two items were simulated to function differently in uniform and non-uniform way respectively. The dataset 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. GMAT also containts simulated continuous criterion variable.

Upload your own datasets

Choose data (CSV file): HCI_ABCD.csv

Main **data** file should contain responses of individual respondents (rows) to given items (columns). Data need to be either binary or nominal (e.g. in ABCD format or Likert scale). Header may contain item names, no row names should be included. 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 items 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.

Type of data: ○ Binary ● Nominal

Separator: ○ Comma ● Semicolon ○ Tab

Quote: ○ None ● Double Quote ○ Single Quote

Data specification: ☑ Header ☐ Keep items names

Missing values: ☐ Keep missing values

Choose key (CSV file): HCI_key.csv

For nominal data, it is necessary to upload **key** of correct answers.

Choose group (optional): HCI_group.csv

Group is binary 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 wont be possible to run DIF and DDF detection procedures in **DIF/Fairness** section. Missing values are not supported for group membership vector and should be removed.

Choose criterion variable (optional): HCI_major.csv

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 wont be possible to run validity analysis in **Predictive validity** section on **Validity** page.

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

ShinyItemAnalysis Test and item analysis | Version 1.2.7-4
© 2018 ShinyItemAnalysis

Figure 3.1: Page to select or upload data.

3.3 Data summary and exploration

Data inspection is a crucial first step in any data analysis. Summary statistics should always be checked before proceeding to further analyses. Summary tab offers basic summaries of data including counts in nominal and binary categories in items, counts for groups and basic statistics for criterion variable (Figure 3.2).

Further checks of any suspicious data may be done in Data exploration tab (Figure 3.3).

3.4 Total scores

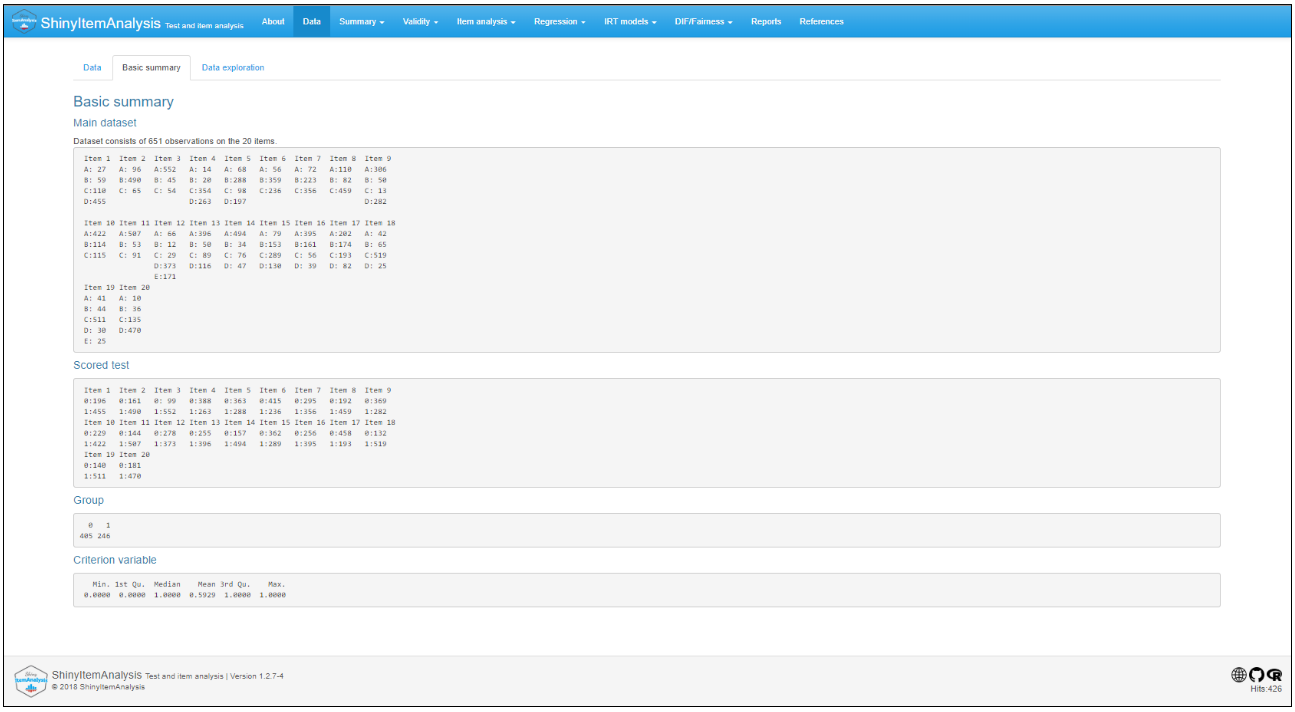
Total score also known as raw score or sum score is a total number of correct answers. In what follows we label total score of person p as X_p. Let $\bar{X} = \frac{1}{n}\sum_{p=1}^{n} X_p$ be a sample mean of total scores X_p and $s^2 = \frac{1}{n-1}\sum_{p=1}^{n}\left(X_p - \bar{X}\right)^2$ their sample variance. Z-score or also standardized score is a linear transformation of total score with a mean of 0 and with variance of 1, that is Z-score for person p is given by

$$Z_p = \frac{X_p - \bar{X}}{s}.$$

T-score is transformed Z-score with a mean of 50 and standard deviation of 10, that is

$$T_p = 10Z_p + 50.$$

Section **Summary** offers summary statistics and histogram of the total scores. The summary table (Table 3.1) offers their basic characteristics such as minimum and maximum, mean, median, standard deviation, skewness



Basic summary

Main dataset

Dataset consists of 651 observations on the 20 items.

```
Item 1  Item 2  Item 3  Item 4  Item 5  Item 6  Item 7  Item 8  Item 9
A: 27   A: 96   A:552   A: 14   A: 68   A: 56   A: 72   A:110   A:306
B: 59   B:490   B: 45   B: 20   B:288   B:359   B:223   B: 82   B: 50
C:110   C: 65   C: 54   C:354   C: 98   C:236   C:356   C:459   C: 13
D:455                   D:263   D:197                           D:282

Item 10 Item 11 Item 12 Item 13 Item 14 Item 15 Item 16 Item 17 Item 18
A:422   A:507   A: 66   A:396   A:494   A: 79   A:395   A:202   A: 42
B:114   B: 53   B: 12   B: 50   B: 34   B:153   B:161   B:174   B: 65
C:115   C: 91   C: 29   C: 89   C: 76   C:289   C: 56   C:193   C:519
                D:373   D:116   D: 47   D:130   D: 39   D: 82   D: 25
                E:171
Item 19 Item 20
A: 41   A: 10
B: 44   B: 36
C:511   C:135
D: 30   D:470
E: 25
```

Scored test

```
Item 1  Item 2  Item 3  Item 4  Item 5  Item 6  Item 7  Item 8  Item 9
0:196   0:161   0: 99   0:388   0:363   0:415   0:295   0:192   0:369
1:455   1:490   1:552   1:263   1:288   1:236   1:356   1:459   1:282
Item 10 Item 11 Item 12 Item 13 Item 14 Item 15 Item 16 Item 17 Item 18
0:229   0:144   0:278   0:255   0:157   0:362   0:256   0:458   0:132
1:422   1:507   1:373   1:396   1:494   1:289   1:395   1:193   1:519
Item 19 Item 20
0:140   0:181
1:511   1:470
```

Group

```
  0   1
405 246
```

Criterion variable

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.0000  1.0000  0.5929  1.0000  1.0000
```

ShinyItemAnalysis Test and item analysis | Version 1.2.7-4
© 2018 ShinyItemAnalysis

Figure 3.2: Page of basic data summary.

Table 3.1: Summary table of total scores for HCI dataset

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3.00	20.00	12.21	12.00	3.64	−0.20	2.35

and kurtosis. The kurtosis here is estimated by sample kurtosis

$$g_2 = \frac{m_4}{s^4} = \frac{\frac{1}{n}\sum\limits_{p=1}^{n}\left(X_p - \bar{X}\right)^4}{\left[\frac{1}{n}\sum\limits_{p=1}^{n}\left(X_p - \bar{X}\right)^2\right]^2}.$$

The skewness is estimated by sample skewness

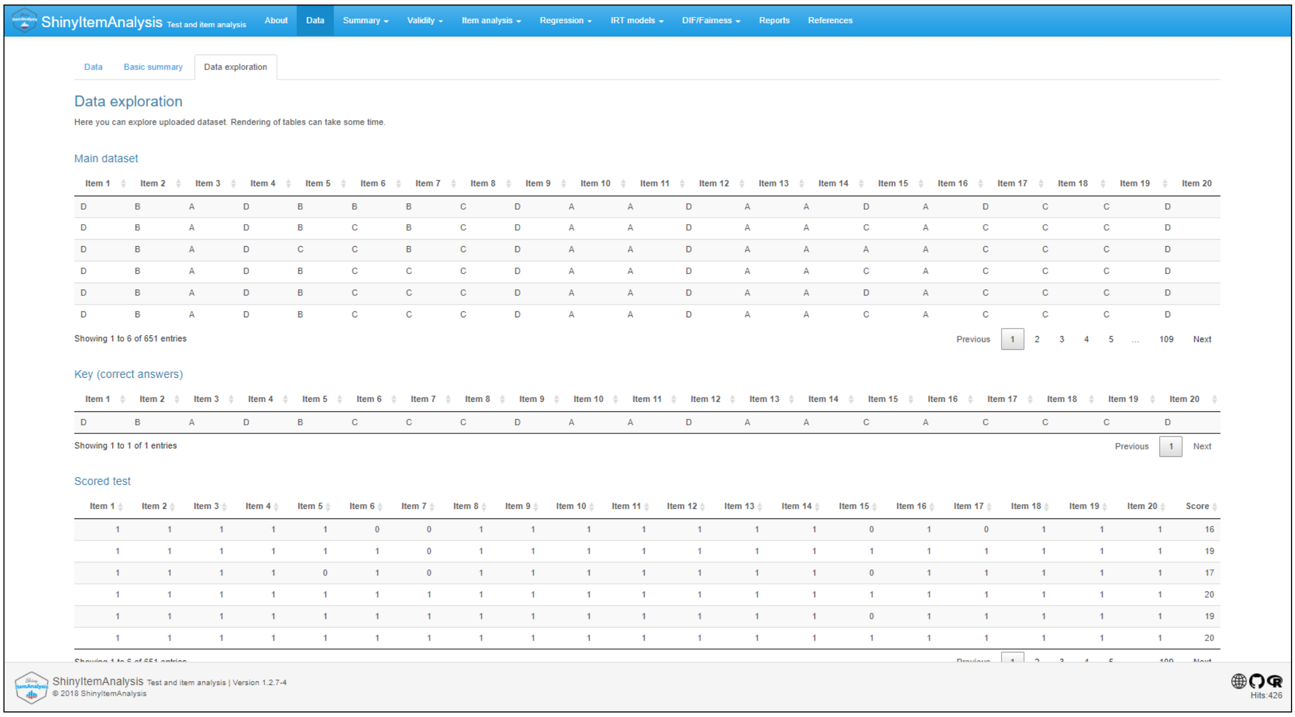
$$b_1 = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum\limits_{p=1}^{n}\left(X_p - \bar{X}\right)^3}{\left[\frac{1}{n-1}\sum\limits_{p=1}^{n}\left(X_p - \bar{X}\right)^2\right]^{3/2}}.$$

The kurtosis for normally distributed scores is near the value of 3 and the skewness is near the value of 0.

Besides the summary statistics the histogram of total scores is provided to describe the distribution of total scores (Figure 3.4). The cut-score may be specified to better visualize the distribution of total scores in two groups (e.g. of those who passed a test and those who did not). For selected cut-score, blue part of histogram shows respondents with total score above the cut-score, grey column shows respondents with total score equal to the cut-score and red part of histogram shows respondents below the cut-score. Bell-shaped histograms are typical for normally distributed data. On the other hand, two-peaked histograms may signalize that the data is actually composed out of two different subgroups.

Total scores with various standard scores (Z-scores, T-scores) are summarized, together with percentile and success rate for each level of total score in one table (Table 3.2).

From numbers in Table 3.2, we can for example read that students with 16 points were in 87th percentil with 80% success rate.



Data exploration

Here you can explore uploaded dataset. Rendering of tables can take some time.

Main dataset

Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
D	B	A	D	B	B	B	C	D	A	A	D	A	A	D	A	D	C	C	D
D	B	A	D	B	C	B	C	D	A	A	D	A	A	C	A	C	C	C	D
D	B	A	D	C	C	B	C	D	A	A	D	A	A	A	A	C	C	C	D
D	B	A	D	B	C	C	C	D	A	A	D	A	A	C	A	C	C	C	D
D	B	A	D	B	C	C	C	D	A	A	D	A	A	D	A	C	C	C	D
D	B	A	D	B	C	C	C	D	A	A	D	A	A	C	A	C	C	C	D

Showing 1 to 6 of 651 entries

Key (correct answers)

Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
D	B	A	D	B	C	C	C	D	A	A	D	A	A	C	A	C	C	C	D

Showing 1 to 1 of 1 entries

Scored test

Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Score
1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	16
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	17
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	19
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20

Showing 1 to 6 of 651 entries

Figure 3.3: Data exploration tab.

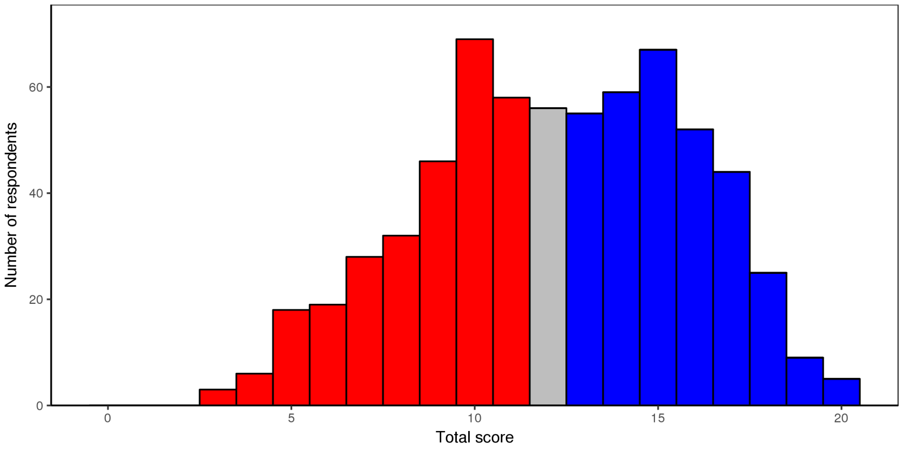


Figure 3.4: Histogram of total scores for HCI data.

3.5 Selected R code

```
library(difNLR)
library(ggplot2)
library(moments)

# loading data
data(GMAT)
data <- GMAT[, 1:20]

# total score calculation
score <- apply(data, 1, sum)

# summary of total score
c(min(score), max(score), mean(score), median(score), sd(score),
  skewness(score), kurtosis(score))

# colors by cut-score
```

Table 3.2: Standard scores for HCI dataset.

Total score	Percentile	Success rate	Z-score	T-score
3.00	0.00	15.00	−2.53	24.69
4.00	0.01	20.00	−2.26	27.44
5.00	0.04	25.00	−1.98	30.19
6.00	0.07	30.00	−1.71	32.93
7.00	0.11	35.00	−1.43	35.68
8.00	0.16	40.00	−1.16	38.43
9.00	0.23	45.00	−0.88	41.18
10.00	0.34	50.00	−0.61	43.92
11.00	0.43	55.00	−0.33	46.67
12.00	0.51	60.00	−0.06	49.42
13.00	0.60	65.00	0.22	52.17
14.00	0.69	70.00	0.49	54.91
15.00	0.79	75.00	0.77	57.66
16.00	0.87	80.00	1.04	60.41
17.00	0.94	85.00	1.32	63.16
18.00	0.98	90.00	1.59	65.90
19.00	0.99	95.00	1.87	68.65
20.00	1.00	100.00	2.14	71.40

```
cut <- median(score) # cut-score
color <- c(rep("red", cut - min(score)),
           "gray",
           rep("blue", max(score) - cut))
df <- data.frame(score)

# histogram
ggplot(df, aes(score)) +
  geom_histogram(binwidth = 1, fill = color, col = "black") +
  xlab("Total score") +
  ylab("Number of respondents") +
  theme_app()

# scores calculations
score <- apply(data, 1, sum) # Total score
tosc <- sort(unique(score)) # Levels of total score
perc <- cumsum(prop.table(table(score))) # Percentiles
sura <- 100 * (tosc / max(score)) # Success rate
zsco <- sort(unique(scale(score))) # Z-score
tsco <- 50 + 10 * zsco # T-score
```

3.6 Exercises

Ex. 3.1 Run ShinyItemAnalysis and try basic data exploration. Using default dataset, answer following questions.

- What is its name?
- How many observations does dataset consist of?
- How many observations do come from focal and reference group?
- What are the maximum and minimum values of criterion variable?

Ex. 3.2 Upload data into ShinyItemAnalysis and explore them.

- What is mean and standard deviation of total scores?
- Calculate Z-score for student with total score 10. Provide whole calculation.

- Calculate T-score for student with total score 10. Provide whole calculation.
- How many points did student with 90th percentile receive?

Ex. 3.3 Create short `R` script including following tasks.

- Upload data from previous section.
- Calculate total scores for uploaded dataset, their mean, median and their standard deviation.
- Draw histogram of total scores. Values smaller than median should be red, values larger than median should be blue, median should be gray.
- Calculate Z-score for uploaded dataset.
- Calculate T-score for uploaded dataset.

Chapter 11

Reports generation

To support routine usage of psychometric methods in test development, `ShinyItemAnalysis` offers possibility to upload data for analysis as csv files, and to generate PDF or HTML reports. Sample PDF report and csv files used for its generation are provided in Supplemental Materials.

Report generation uses `rmarkdown` templates and `knitr` for compiling (see Figure 11.1). LaTeXis used for creating PDF reports. Latest version of LaTeXwith properly set paths is needed to generate PDF reports locally.

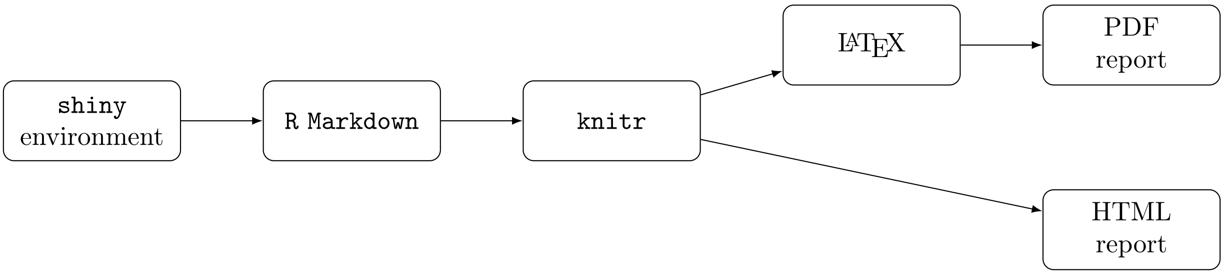
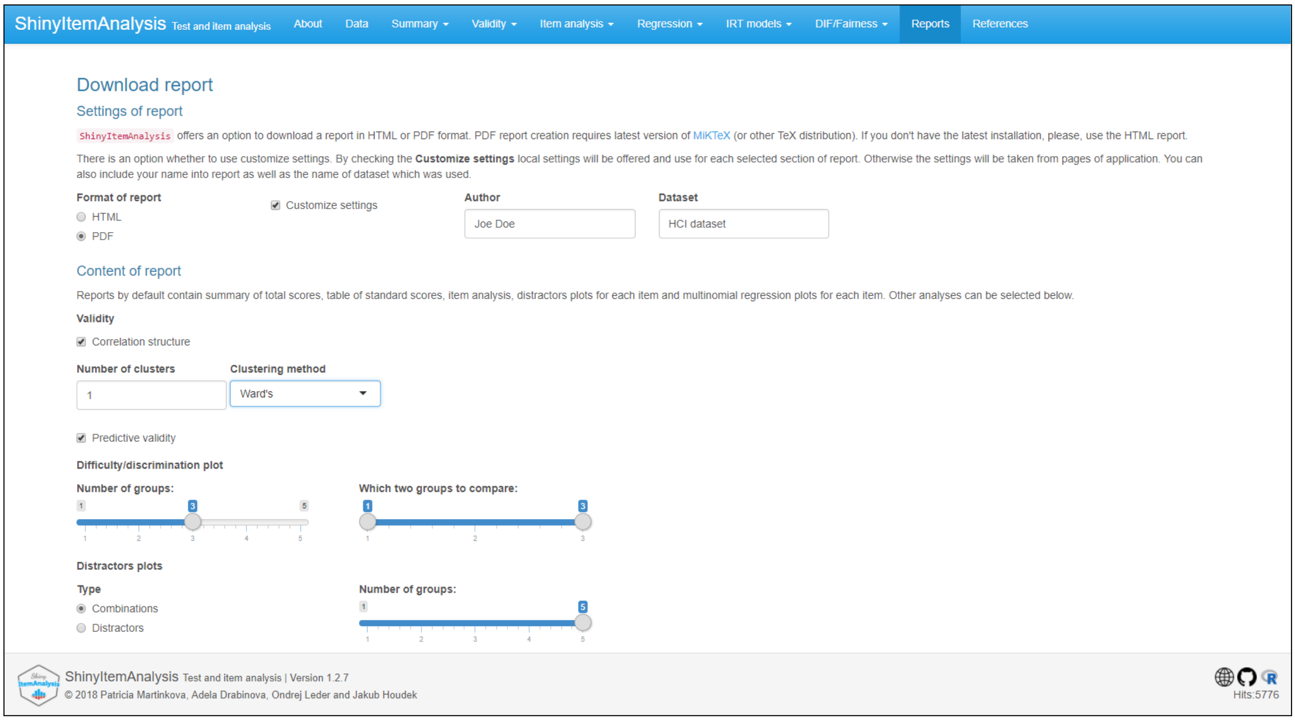


Figure 11.1: Report generation workflow.

Report page setting allows to specify dataset name, to include name of person who generated the report, to select from available methods and to customize settings (see Figures 11.2 and 11.3). **Generate report** button starts analyses needed for report generation. Subsequently, **Download report** button initializes compiling the text, tables and figures into PDF or HTML file.



Download report

Settings of report

`ShinyItemAnalysis` offers an option to download a report in HTML or PDF format. PDF report creation requires latest version of MiKTeX (or other TeX distribution). If you don't have the latest installation, please, use the HTML report.

There is an option whether to use customize settings. By checking the **Customize settings** local settings will be offered and use for each selected section of report. Otherwise the settings will be taken from pages of application. You can also include your name into report as well as the name of dataset which was used.

Format of report
○ HTML
● PDF

☑ Customize settings

Author: Joe Doe

Dataset: HCI dataset

Content of report

Reports by default contain summary of total scores, table of standard scores, item analysis, distractors plots for each item and multinomial regression plots for each item. Other analyses can be selected below.

Validity

☑ Correlation structure

Number of clusters: 1

Clustering method: Ward's

☑ Predictive validity

Difficulty/discrimination plot

Number of groups: 3

Which two groups to compare: 1 – 3

Distractors plots

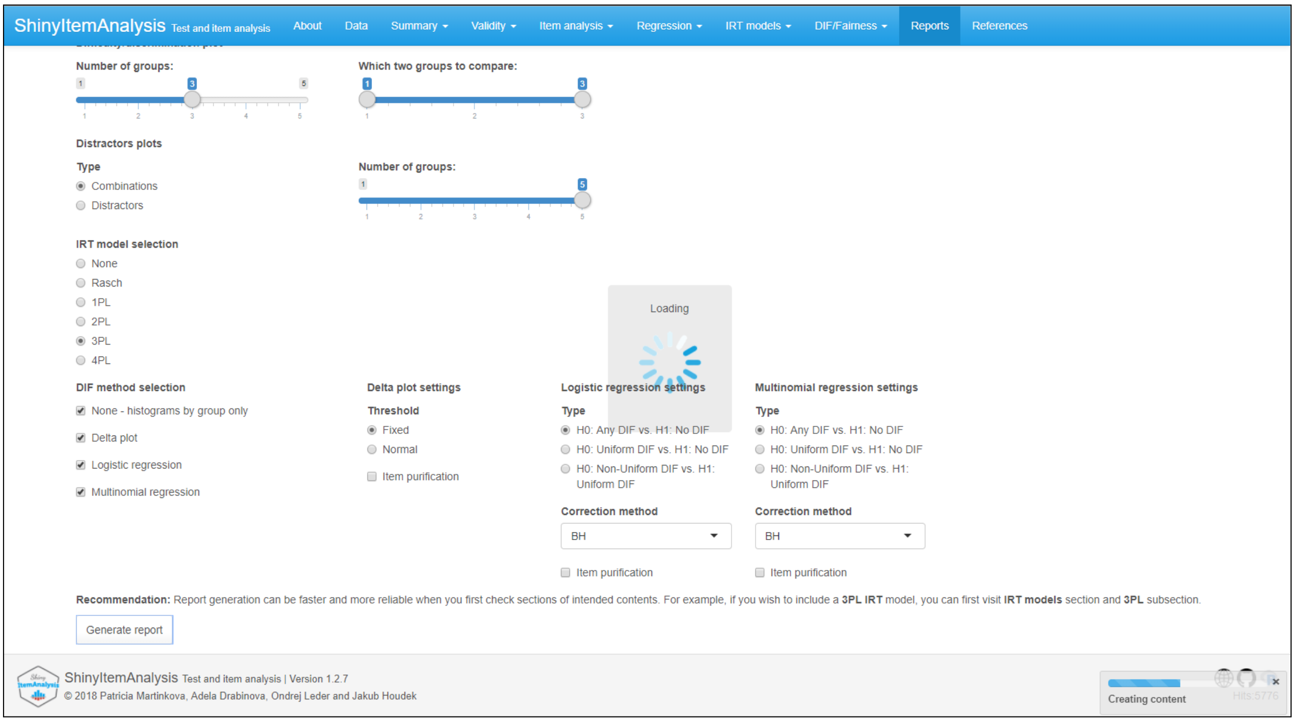
Type
● Combinations
○ Distractors

Number of groups: 5

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

Hits:5776

Figure 11.2: Report setting of HCI data analysis.



Number of groups: 3

Which two groups to compare: 1 – 3

Distractors plots

Type

● Combinations
○ Distractors

Number of groups: 5

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

BH

□ 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

BH

□ Item purification

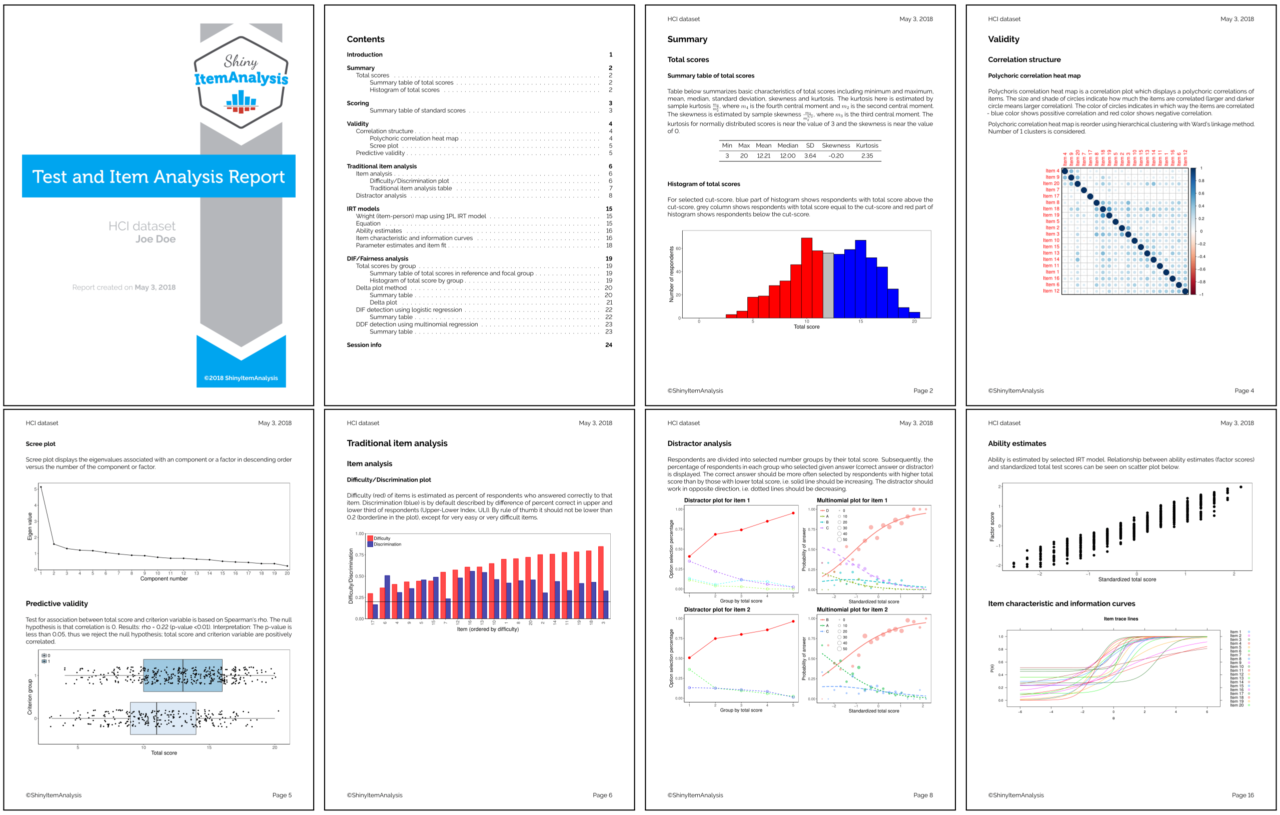
Recommendation: Report generation can be faster and more reliable when you first check sections of intended contents. For example, if you wish to include a **3PL IRT** model, you can first visit **IRT models** section and **3PL** subsection.

Generate report

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

Figure 11.3: Report setting of HCI data analysis.

Sample pages of PDF report on HCI dataset are displayed in Figure 11.4. Reports provide quick overview of test characteristics and may be a helpful material for test developers, item writers and institutional stakeholders.



Shiny ItemAnalysis

Test and Item Analysis Report

HCI dataset

Joe Doe

Report created on **May 3, 2018**

©2018 ShinyItemAnalysis

Contents

Introduction 1

Summary 2
Total scores 2
Summary table of total scores 2
Histogram of total scores 2

Scoring 3
Summary table of standard scores 3

Validity 4
Correlation structure 4
Polychoric correlation heat map 4
Scree plot 5
Predictive validity 5

Traditional item analysis 6
Item analysis 6
Difficulty/Discrimination plot 6
Traditional item analysis table 7
Distractor analysis 8

IRT models 15
Wright (item-person) map using 1PL IRT model 15
Equation 15
Ability estimates 16
Item characteristic and information curves 16
Parameter estimates and item fit 18

DIF/Fairness analysis 19
Total scores by group 19
Summary table of total scores in reference and focal group 19
Histogram of total score by group 19
Delta plot method 20
Summary table 20
Delta plot 21
DIF detection using logistic regression 22
Summary table 22
DDF detection using multinomial regression 23
Summary table 23

Session info 24

HCI dataset May 3, 2018

Summary

Total scores

Summary table of total scores

Table below summarizes basic characteristics of total scores including minimum and maximum, mean, median, standard deviation, skewness and kurtosis. The kurtosis here is estimated by sample kurtosis $\frac{m_4}{m_2^2}$, where m_4 is the fourth central moment and m_2 is the second central moment. The skewness is estimated by sample skewness $\frac{m_3}{m_2^{3/2}}$, where m_3 is the third central moment. The kurtosis for normally distributed scores is near the value of 3 and the skewness is near the value of 0.

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3	20	12.21	12.00	3.64	-0.20	2.35

Histogram of total scores

For selected cut-score, blue part of histogram shows respondents with total score above the cut-score, grey column shows respondents with total score equal to the cut-score and red part of histogram shows respondents below the cut-score.

©ShinyItemAnalysis Page 2

HCI dataset May 3, 2018

Validity

Correlation structure

Polychoric correlation heat map

Polychoris 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 possitive correlation and red color shows negative correlation.

Polychoric correlation heat map is reorder using hierarchical clustering with Ward's linkage method. Number of 1 clusters is considered.

©ShinyItemAnalysis Page 4

HCI dataset May 3, 2018

Scree plot

Scree plot displays the eigenvalues associated with an component or a factor in descending order versus the number of the component or factor.

Predictive validity

Test for association between total score and criterion variable is based on Spearman's rho. The null hypothesis is that correlation is 0. Results: rho = 0.22 (p-value <0.01). Interpretation: The p-value is less than 0.05, thus we reject the null hypothesis; total score and criterion variable are positively correlated.

©ShinyItemAnalysis Page 5

HCI dataset May 3, 2018

Traditional item analysis

Item analysis

Difficulty/Discrimination plot

Difficulty (red) of items is estimated as percent of respondents who answered correctly to that item. Discrimination (blue) is by default described by difference of percent correct in upper and lower third of respondents (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.

©ShinyItemAnalysis Page 6

HCI dataset May 3, 2018

Distractor analysis

Respondents are divided into selected number groups by their total score. Subsequently, the percentage of respondents in each group who selected given answer (correct answer or distractor) is displayed. The correct answer should be more often selected by respondents with higher total score than by those with lower total score, i.e. solid line should be increasing. The distractor should work in opposite direction, i.e. dotted lines should be decreasing.

Distractor plot for item 1
Multinomial plot for item 1
Distractor plot for item 2
Multinomial plot for item 2
Option selection percentage
Group by total score
Probability of answer
Standardized total score

©ShinyItemAnalysis Page 8

HCI dataset May 3, 2018

Ability estimates

Ability is estimated by selected IRT model. Relationship between ability estimates (factor scores) and standardized total test scores can be seen on scatter plot below.

Item characteristic and information curves

©ShinyItemAnalysis Page 16

Figure 11.4: Selected pages of report on HCI data.

Appendices

A1 Installation of R and RStudio

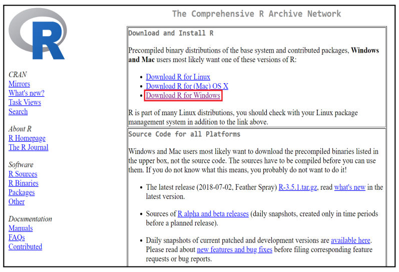
Here we provide detailed instruction for installation of R in Windows and Mac OS X. We also recommend you to install RStudio.

A1.1 Windows

1. Go to

   https://cran.r-project.org/

   and click on **Download R for Windows**, then **base** and finally **Download R 3.x.x for Windows**.



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

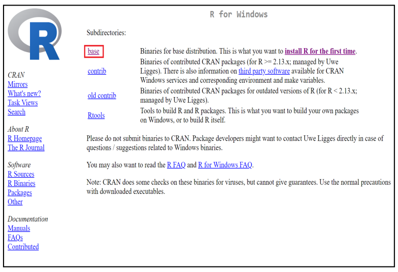
- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-07-02, Feather Spray) R-3.5.1.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.



R for Windows

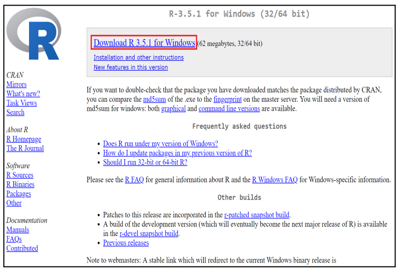
Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.



R-3.5.1 for Windows (32/64 bit)

Download R 3.5.1 for Windows (62 megabytes, 32/64 bit)

Installation and other instructions

New features in this version

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server. You will need a version of md5sum for windows: both graphical and command line versions are available.

Frequently asked questions

- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

Other builds

- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

Note to webmasters: A stable link which will redirect to the current Windows binary release is

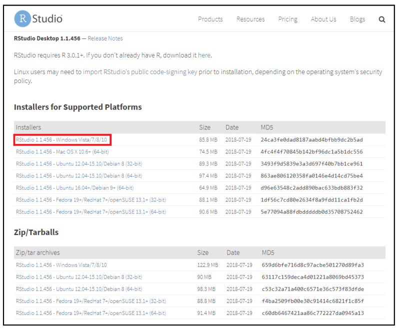
   This starts downloading of installer R-3.X.X-win.exe.

2. Open downloaded installer and follow instructions to install R. Leave all default settings in the installation options.

3. Go to

   https://www.rstudio.com/products/rstudio/download/

   and click on **RStudio 1.1.XXX - Windows Vista/7/8/10** on the bottom of the page.



RStudio Desktop 1.1.456 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it here.

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	85.8 MB	2018-07-19	24ca3fe0dad8187aabd4bfbb9dc2b5ad
RStudio 1.1.456 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-07-19	4fc4f4f70845b142bf96dc1a5b1dc556
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-07-19	3493f9d5839e3a3d697f40b7bb1ce961
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-07-19	863ae806120358fa0146e4d14cd75be4
RStudio 1.1.456 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.9 MB	2018-07-19	d96e63548c2add890bac633bdb883f32
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-07-19	1df56c7cd80e2634f8a9fdd11ca1fb2d
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-07-19	5e77094a88fdbddddb0d35708752462

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	122.9 MB	2018-07-19	659d6bfe716d8c97acbe501270d89fa3
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-07-19	63117c159deca4d01221a8069bd45373
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-07-19	c53c32a71a400c6571e36c573f83dfde
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-07-19	f4ba2509fb00e30c91414c6821f1c85f
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-07-19	c60db6467421aa86c772227da0945a13

   This starts downloading of installer RStudio-1.1.XXX.exe.

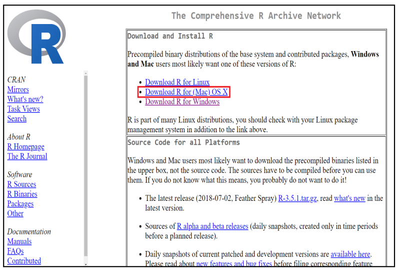
4. Open downloaded installer and follow instructions to install RStudio. Leave all default settings in the installation options.

A1.2 Mac OS X

1. Go to

   https://cran.r-project.org/

   and click on **Download R for Mac OS X** and then **Download R-3.x.x.pkg**.



The Comprehensive R Archive Network

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

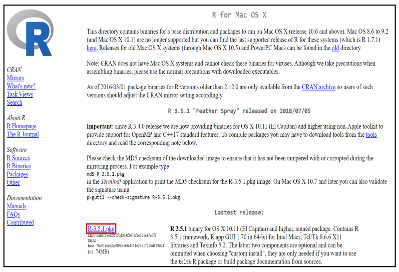
- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-07-02, Feather Spray) R-3.5.1.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature



R for Mac OS X

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) here. Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the old directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

As of 2016/03/01 package binaries for R versions older than 2.12.0 are only available from the CRAN archive so users of such versions should adjust the CRAN mirror setting accordingly.

R 3.5.1 "Feather Spray" released on 2018/07/05

Important: since R 3.4.0 release we are now providing binaries for OS X 10.11 (El Capitan) and higher using non-Apple toolkit to provide support for OpenMP and C++17 standard features. To compile packages you may have to download tools from the tools directory and read the corresponding note below.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type
`md5 R-3.5.1.pkg`
in the *Terminal* application to print the MD5 checksum for the R-3.5.1.pkg image. On Mac OS X 10.7 and later you can also validate the signature using
`pkgutil --check-signature R-3.5.1.pkg`

Lastest release:

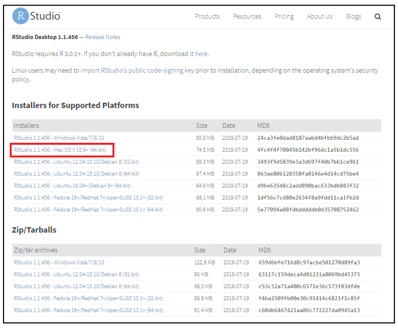
R-3.5.1.pkg
MD5-hash: 58eaff65fbd024f267ef1e521e17e7f8
SHA1-hash: 76c01bfa62a6896d5f4a4511e25d17276d149621
(ca. 74MB)

R 3.5.1 binary for OS X 10.11 (El Capitan) and higher, signed package. Contains R 3.5.1 framework, R.app GUI 1.70 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 5.2. The latter two components are optional and can be ommitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

2. Install R. Leave all default settings in the installation options.
3. Go to

   https://www.rstudio.com/products/rstudio/download/

   and click on **RStudio 1.1.XXX - Mac OS X 10.6+ (64-bit)** on the bottom of the page.



RStudio Desktop 1.1.456 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it here.

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	85.8 MB	2018-07-19	24ca3fe0dad8187aabd4bfbb9dc2b5ad
RStudio 1.1.456 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-07-19	4fc4f4f70845b142bf96dc1a5b1dc556
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-07-19	3493f9d5839e3a3d697f40b7bb1ce961
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-07-19	863ae806120358fa0146e4d14cd75be4
RStudio 1.1.456 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.9 MB	2018-07-19	d96e63548c2add890bac633bdb883f32
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-07-19	1df56c7cd80e2634f8a9fdd11ca1fb2d
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-07-19	5e77094a88fdbddddb0d35708752462

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.456 - Windows Vista/7/8/10	122.9 MB	2018-07-19	659d6bfe716d8c97acbe501270d89fa3
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-07-19	63117c159deca4d01221a8069bd45373
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-07-19	c53c32a71a400c6571e36c573f83dfde
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-07-19	f4ba2509fb00e30c91414c6821f1c85f
RStudio 1.1.456 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-07-19	c60db6467421aa86c772227da0945a13

4. Install RStudio by dragging the application icon to your Applications folder.

A2 Installation of ShinyItemAnalysis

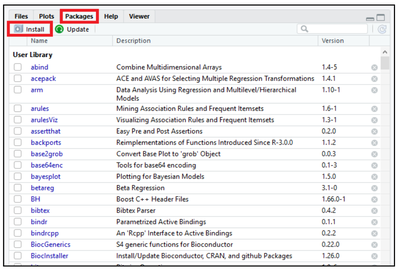
1. Open RStudio (or R) and install and load ShinyItemAnalysis with typing following commands into console:

```
install.packages("ShinyItemAnalysis")
library(ShinyItemAnalysis)
```

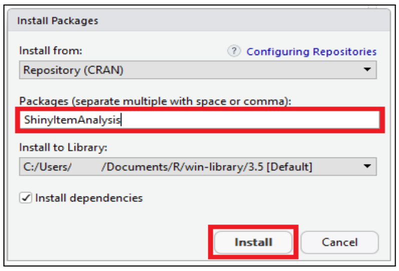
2. In case that some dependency packages have not been installed automatically, you can use command

```
install.packages("MISSING-PACKAGE")
```

   where MISSING-PACKAGE is replaced by name of not installed package. To install packages, you can use also clickable environment of RStudio:



	Name	Description	Version
	User Library		
□	abind	Combine Multidimensional Arrays	1.4-5
□	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
□	arm	Data Analysis Using Regression and Multilevel/Hierarchical Models	1.10-1
□	arules	Mining Association Rules and Frequent Itemsets	1.6-1
□	arulesViz	Visualizing Association Rules and Frequent Itemsets	1.3-1
□	assertthat	Easy Pre and Post Assertions	0.2.0
□	backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.2
□	base2grob	Convert Base Plot to 'grob' Object	0.0.3
□	base64enc	Tools for base64 encoding	0.1-3
□	bayesplot	Plotting for Bayesian Models	1.5.0
□	betareg	Beta Regression	3.1-0
□	BH	Boost C++ Header Files	1.66.0-1
□	bibtex	Bibtex Parser	0.4.2
□	bindr	Parametrized Active Bindings	0.1.1
□	bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2
□	BiocGenerics	S4 generic functions for Bioconductor	0.22.0
□	BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.26.0



Install Packages

Install from: Configuring Repositories
Repository (CRAN)

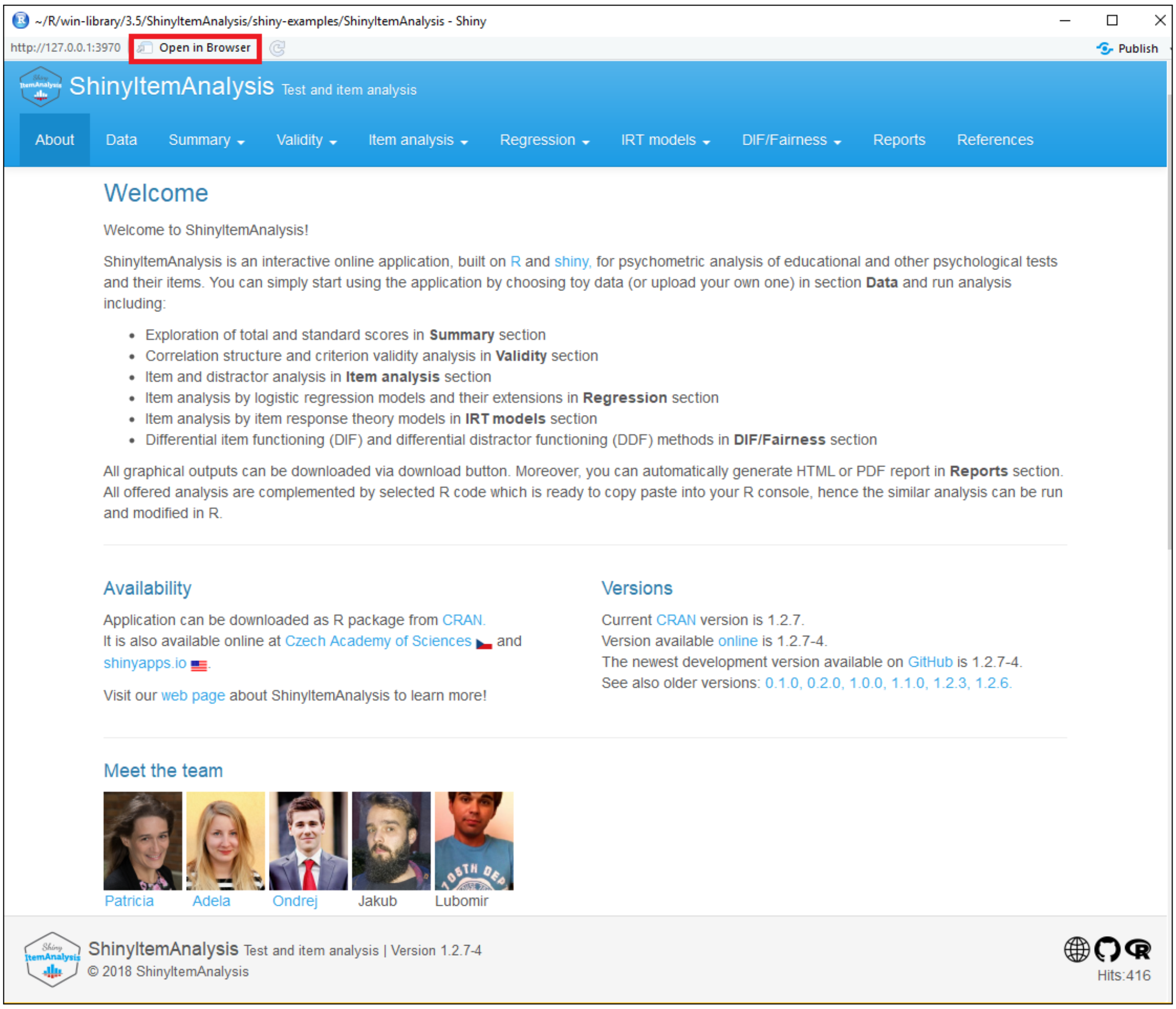
Packages (separate multiple with space or comma):
ShinyItemAnalysis

Install to Library:
C:/Users/ /Documents/R/win-library/3.5 [Default]

☑ Install dependencies

3. Now, `ShinyItemAnalysis` is ready to run. To launch the application, type into console

```
startShinyItemAnalysis()
```



Welcome

Welcome to ShinyItemAnalysis!

ShinyItemAnalysis is an interactive online application, built on R and shiny, for psychometric analysis of educational and other psychological tests and their items. You can simply start using the application by choosing toy data (or upload your own one) in section **Data** and run analysis including:

- Exploration of total and standard scores in **Summary** section
- Correlation structure and criterion validity analysis in **Validity** section
- Item and distractor analysis in **Item analysis** section
- Item analysis by logistic regression models and their extensions in **Regression** section
- Item analysis by item response theory models in **IRT models** section
- Differential item functioning (DIF) and differential distractor functioning (DDF) methods in **DIF/Fairness** section

All graphical outputs can be downloaded via download button. Moreover, you can automatically generate HTML or PDF report in **Reports** section. All offered analysis are complemented by selected R code which is ready to copy paste into your R console, hence the similar analysis can be run and modified in R.

Availability

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

Visit our web page about ShinyItemAnalysis to learn more!

Versions

Current CRAN version is 1.2.7.
Version available online is 1.2.7-4.
The newest development version available on GitHub is 1.2.7-4.
See also older versions: 0.1.0, 0.2.0, 1.0.0, 1.1.0, 1.2.3, 1.2.6.

Meet the team

Patricia Adela Ondrej Jakub Lubomir

ShinyItemAnalysis Test and item analysis | Version 1.2.7-4
© 2018 ShinyItemAnalysis

Hits:416

Click on **Open in browser** button to open application in your favourite browser.

A3 Installation of TeXdistribution

Here we provide links to detailed tutorials for MiKTex installation. MiKTeX is TeXdistribution and can be freely downloaded:

https://miktex.org/download

Please, follow tutorials for your choice of operation system at this webpage.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Chang, W. (2017). rmarkdown: Dynamic documents for r [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rmarkdown (R package version 1.8)

American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Ames, A. J., & Penfield, R. D. (2015). An ncme instructional module on item-fit statistics for item response theory models. *Educational Measurement: Issues and Practice*, *34*(3), 39–48.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573.

Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, *10*(2), 95–105.

Attali, D. (2018). shinyjs: Easily improve the user experience of your shiny apps in seconds [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=shinyjs (R package version 1.0)

Auguie, B. (2017). gridextra: Miscellaneous functions for "grid" graphics [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=gridExtra (R package version 2.3)

Bailey, E. (2015). shinybs: Twitter bootstrap components for shiny [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=shinyBS (R package version 0.61)

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bates, D., & Maechler, M. (2017). Matrix: Sparse and dense matrix classes and methods [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=Matrix (R package version 1.2-12)

Bliese, P. (2016). multilevel: Multilevel functions [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=multilevel (R package version 2.6)

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*(1), 29–51.

Brennan, R. L. (2006). *Educational measurement.* Praeger.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 1904-1920*, *3*(3), 296–322.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. doi: 10.18637/jss.v048.i06

Chang, W., & Borges Ribeiro, B. (2018). shinydashboard: Create dashboards with 'shiny' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=shinydashboard (R package version 0.7.0)

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). shiny: Web application framework for r [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=shiny (R package version 1.0.5)

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Dahl, D. B. (2016). xtable: Export tables to latex or html [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=xtable (R package version 1.8-2)

Dowle, M., & Srinivasan, A. (2017). data.table: Extension of ‘data.frame‘ [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=data.table (R package version 1.10.4-3)

Drabinová, A., & Martinková, P. (2017). Detection of differential item functioning with nonlinear regression: A non-IRT approach accounting for guessing. *Journal of Educational Measurement*, *54*(4), 498–517.

Drabinová, A., & Martinková, P. (2018). difnlr: Generalized logistic regression models for dif and ddf detection. *R Journal*. (Submitted)

Drabinová, A., Martinková, P., & Zvára, K. (2018). difnlr: Dif and ddf detection by non-linear regression models. [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=difNLR (R package version 1.2.2)

Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, *14*(2), 352–364.

Fletcher, T. D. (2010). psychometric: Applied psychometric theory [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=psychometric (R package version 2.2)

Fox, J. (2016). polycor: Polychoric and polyserial correlations [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=polycor (R package version 0.7-9)

Haladyna, T. M., & Downing, S. M. (2011). *Handbook of test development.* Routledge.

Irribarra, D. T., & Freund, R. (2014). Wright map: Irt item-person map with conquest integration [Computer software manual]. Retrieved from http://github.com/david-ti/wrightmap

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, *38*(8), 1–29. Retrieved from http://www.jstatsoft.org/v38/i08/

Kingston, N., Leary, L., & 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), 1–56.

Komsta, L., & Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and related tests [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=moments (R package version 0.14)

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Routledge.

Magis, D., Beland, 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.

Magis, D., & Facon, B. (2014). deltaPlotR: An R package for differential item functioning analysis with angoff's delta plot. *Journal of Statistical Software, Code Snippets*, *59*(1), 1–19. Retrieved from http://www.jstatsoft.org/v59/c01/

Mair, P. (2018). *CRAN task view: Psychometric models and methods.* Retrieved 2018-08-16, from https://CRAN.R-project.org/view=Psychometrics

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, *22*(4), 719–748.

Martinková, P., Drabinová, A., Leder, O., & Houdek, J. (2018). ShinyItemAnalysis: Test and item analysis via shiny [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=ShinyItemAnalysis (R package version 1.2.6)

Martinková, P., Drabinová, A., Liaw, Y.-L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE-Life Sciences Education*, *16*(2), rm2.

Martinková, P., Štěpánek, L., Drabinová, A., Houdek, J., Vejražka, M., & Štuka, v. (2017). Semi-real-time analyses of item characteristics for medical school admission tests. In *Computer science and information systems (fedcsis), 2017 federated conference on* (pp. 189–194).

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

McFarland, J. L., Price, R. M., Wenderoth, M. P., Martinková, P., Cliff, W., Michael, J., ... Wright, A. (2017). Development and validation of the homeostasis concept inventory. *CBE-Life Sciences Education*, *16*(2), ar35.

Muraki, E. (1992). A generalized partial credit model: Application of an em algorithm. *ETS Research Report Series*, *1992*(1).

Neuwirth, E. (2014). Rcolorbrewer: Colorbrewer palettes [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=RColorBrewer (R package version 1.1-2)

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (2nd ed.). McGraw-Hill New York.

Paradis, E. (2002). *R for beginners.* Montpellier (F): University of Montpellier. Retrieved from https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). nlme: Linear and nonlinear mixed effects models [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=nlme (R package version 3.1-131)

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, *53*(4), 495–502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, *14*(2), 197–207.

Rasch, G. (1960). *Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.* Nielsen & Lydiche.

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74.

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych (R package version 1.8.3)

Rizopoulos, D. (2006). ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25. Retrieved from http://www.jstatsoft.org/v17/i05/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rust, J., & Golombok, S. (2014). *Modern psycometrics* (3rd ed.). Routledge.

Samejima, F. (1970). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *35*(1), 139–139.

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org (ISBN 978-0-387-75968-5)

Sarkar, D., & Andrews, F. (2016). latticeextra: Extra graphical utilities based on lattice [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=latticeExtra (R package version 0.6-28)

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2017). plotly: Create interactive web graphics via 'plotly.js' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=plotly (R package version 4.7.1)

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 1904-1920*, *3*(3), 271–295.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement*, *27*(4), 361–370.

van der Linden, W. J. (2017). *Handbook of item response theory*. CRC Press.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4 (ISBN 0-387-95457-0)

Wei, T., & Simko, V. (2017). R package "corrplot": Visualization of a correlation matrix [Computer software manual]. Retrieved from https://github.com/taiyun/corrplot ((Version 0.84))

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. Retrieved from http://www.jstatsoft.org/v21/i12/

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2018). stringr: Simple, consistent wrappers for common string operations [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=stringr (R package version 1.3.0)

Willse, J. T. (2018). Ctt: Classical test theory functions [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=CTT (R package version 2.3.2)

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://yihui.name/knitr/ (ISBN 978-1498716963)

Xie, Y. (2018). Dt: A wrapper of the javascript library 'datatables' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=DT (R package version 0.4)

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, revelle's β, and mcdonald's ω h: Their relations with each other and two alternative conceptualizations of reliability. *psychometrika*, *70*(1), 123–133.

Acronyms

DDF Differential Distractor Functioning.

GMAT Graduate Management Admission Test.
GPA Grade Point Average.
GPCM Generalized Partial Credit Model.
GRM Graded Response Model.

HCI Homeostasis Concept Inventory.

ICC Item Characteristic Curve.
IIC Item Information Curve.

MSAT-B Medical School Admission Test in Biology.

NRM Nominal Response Model.

PCM Partial Credit Model.

RSM Rating Scale Model.

ULI Upper-Lower Index.

Index

correlation, 15, 16
  Pearson, 15, 18
  polychoric, 15
criterion variable, 16
Cronbach's alpha, 13

differential distractor functioning, 30
differential item functioning, 30
difficulty, 18, 21
discrimination, 18, 21, 27
distractor, 18, 21, 30

guessing, 21

hierarchical clustering, 15

inattention, 21
item response theory, 13, 27, 30

kurtosis, 8

logistic regression, 21

McDonald's omega, 14

reliability, 12
  test-retest, 13
Revelle's beta, 14

score
  latent, 27
  observed, 12
  percentile, 8
  standardized, 7, 21, 27
  success rate, 8
  T-score, 7
  total, 21
  total score, 7
  true, 12
  Z-score, *see* standardized
skewness, 8

validity, 15
  construct, 15
  content, 15
  criterion, 15