Selected topics in psychometrics

Patrícia Martinková

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

Institute for Research and Development of Education
Faculty of Education, Charles University, Prague

NMST570, October 2, 2018

Table of contents

Psychometrics

Psychometrics is a field of study concerned with the theory and technique of psychological/educational/behavioral measurement.

Field is concerned with objective testing/measurement/assessment of
- skills, knowledge, abilities, educational achievement
- attitudes, personality traits

Psychometricians focus on
- construction and validation of assessment instruments
  - e.g. questionnaires, tests, raters' judgments, and personality tests
- research relating to measurement theory
  - e.g. item response models, intraclass correlation, etc.

Selected books

- Rao & Sinharay (Eds.): Handbook of statistics, vol. 26. Psychometrics. Elsevier, 2006.
- AERA, APA, NCME: Standards for educational and psychological testing. 2014.
- Downing & Haladyna (Eds.): Handbook of test development. LEA, 2006.
- Brennan (Ed.): Educational Measurement. ACE/Preaeger, 2006.

- Urbánek, Širůček, Denglerová: Psychometrika. Měření v psychologii. Portál, 2011.
- Štuka, Martinková et al.: Testování při výuce medicíny. Konstrukce a analýza znalostních testů. Karolinum, 2013.

Selected journals

- Psychometrika
- British Journal of Mathematical and Statistical Psychology (BJMSP)
- Journal of Educational and Behavioral Statistics (JEBS)
- Journal of Educational Measurement (JEM)
- Applied Psychological Measurement (APM)
- Educational and Psychological Measurement (EPM)
- Practical Assessment, Research and Evaluation (PARE)
- ...
- Testfórum (in Czech)

Announcement: 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/2019

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).

Psychometrics
OOOO

About this course
●OOO

Topics
OOOOOOOOOOO

Software
OOOOO

Course description

This course covers main topics in psychometrics, such as

- traditional item analysis
- reliability and validity of measurement
- use of logistic and nonlinear regression models
- item response theory models
- differential item functioning
- computerized adaptive testing, etc.

Methods are demonstrated using data from admission tests and other assessments.

Exercises are prepared in freely available statistical software R, other psychometric software is also introduced.

Course goals

After taking this course you should be able to:

1. Name the main topics studied in psychometrics
2. Find proofs about reliability and validity of assessment instrument
3. Apply traditional methods to describe item functioning (difficulty, discrimination, distractor analysis)
4. Explain how logistic regression may be used to describe item properties, apply regression models on real data and interpret results
5. Formulate IRT models for binary, polytomous and nominal items, apply them on real data and interpret results
6. Explain concept of differential item functioning, apply some DIF detection methods
7. Explain process of computerized adaptive testing, prepare your own adaptive test
8. Describe process of assessment development and validation and apply its steps in real situations

Course materials

- Lecture notes / ShinyItemAnalysis tutorial
- Rao & Sinharay (2006). Handbook of Statistics. Volume 26: Psychometrics. Amsterdam, NL: Elsevier.
- van der Linden (2016). Handbook of item response theory: Models, statistical tools, and applications (Vols.1-3). Boca Raton, FL: Chapman & Hall/CRC.

Grading policy

Course credit requirements:

- present at the exercise class sessions (two absences are tolerated)
- 10 HW assignments (at least 60% of points)

Exam/Grade:

- final project (own data or project assigned during the last lecture)
- oral exam (follow up questions on handed homework (50%) and on final project (50%))

You are welcome and encouraged to use your data for the project in lieu of the project assigned to the class. In such a case, you are expected to prepare written project proposal and submit it to the lecturer at least one week before the last lecture.

Psychometrics
oooo

About this course
oooo

Topics
●ooooooooooo

Software
ooooo

Course content / Tentative outline

- Introduction (2.10.2018)
- Reliability and measurement error (9.10.2018)
- Validity (16.10.2018)
- Traditional item analysis (23.10.2018)
- Regression models for item description (30.10.2018)
- IRT models for binary data (13.11.2018)
- IRT models for binary data (20.11.2018)
- IRT models for ordinal and nominal items (27.11.2018)
- Differential item functioning (1-2 lessons) (4.12.2018, 11.12.2018)
- Computerized adaptive testing (18.12.2018)
- Assessment development and validation. Final project (8.1.2019)

Reliability

Describes amount of error in measurement

- Test-retest
- Internal consistency (split-half, Cronbach's alpha)
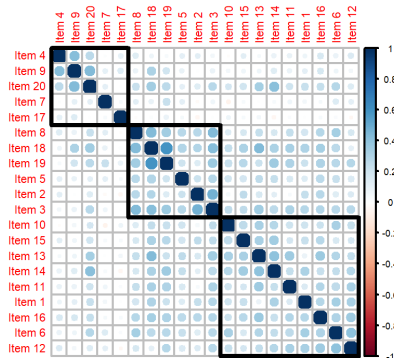- Inter-rater reliability
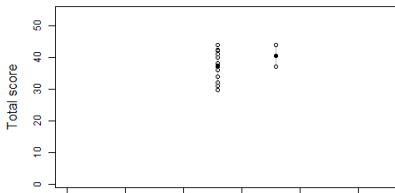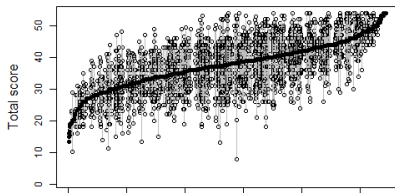
Validity

- Proofs based on content
- Empirical proofs
    - Criterion, predictive, incremental
   - Discriminant, etc.
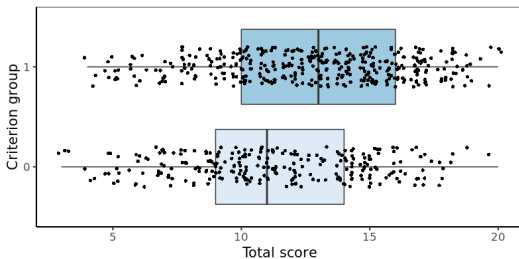
Validity

Correlation structure

McFarland, Price, Wenderoth, Martinková, et al. Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35, 2017.

Traditional item analysis

Uses ratios and correlations to describe item properties

- Difficulty
  - Ratio of correct answers p
- Discrimination, ULI, RIR, RIT, ...
  - Upper–Lower Index (ULI)
  - biserial correlation Item–Rest total (RIR), Item–Test total (RIT)
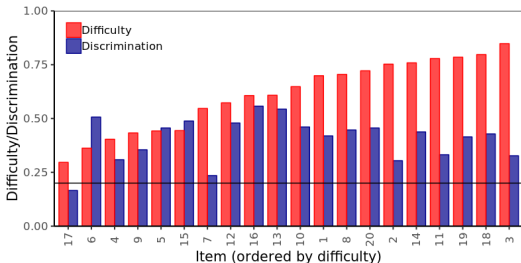
Traditional item analysis

Distractor analysis
- Respondents are divided into (3 or more) groups by total score
- Option selection is displayed with respect to group
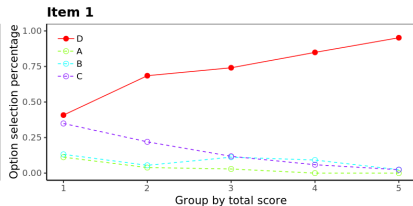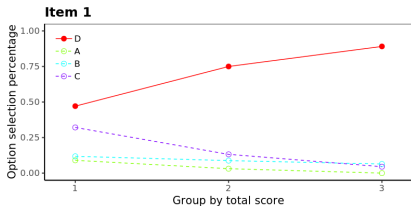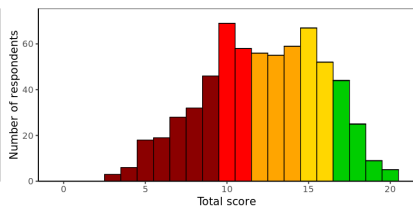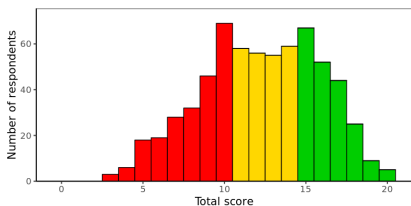
Regression models for item description

Models describing mean item score or probability of seclection of given answer with respect to total (standardized total) score
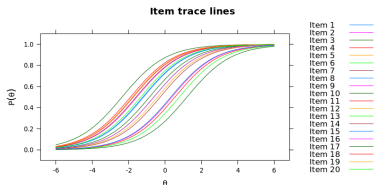
- Logistic regression
- Nonlinear regression
- Multiomial regression

Psychometrics
○○○○

About this course
○○○○

Topics
○○○○○○○○●○○○

Software
○○○○○
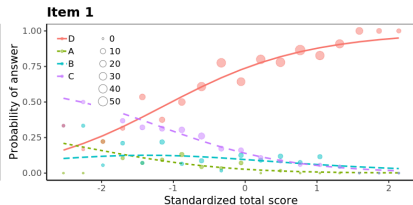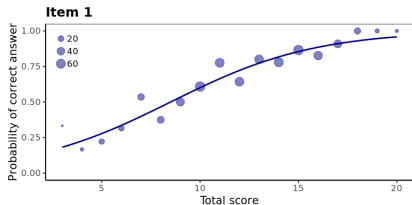
Item Response Theory (IRT) models

IRT models can be conceptualized as mixed effect models
- Abilities are considered **latent** (unobserved) and are estimated
 - Modelled as random effects (often with normal distribution)

- Allow to put items and repondents on one scale (Item–Person map)

Psychometrics
oooo

About this course
oooo

Topics
ooooooooo●oo

Software
ooooo
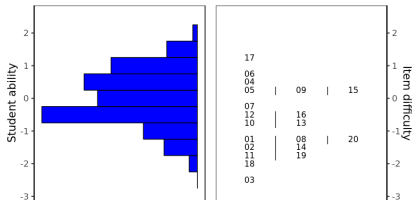
Item Response Theory (IRT) models

IRT models have many advantages

- Provide more precise estimation of ability than total score
  - Account for item difficulty and discrimination
- Provide detailed description of item and test functioning
  - Item characteristic curve
  - Item information curve
  - Test information curve
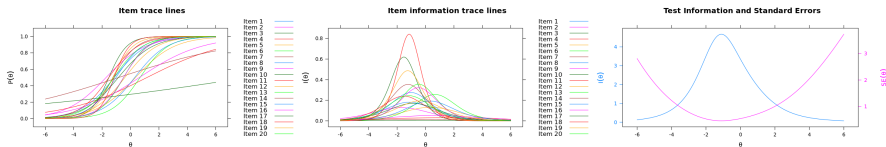
Differential item functioning (DIF)

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

- Potentially unfair items
- Uniform vs. nonuniform DIF
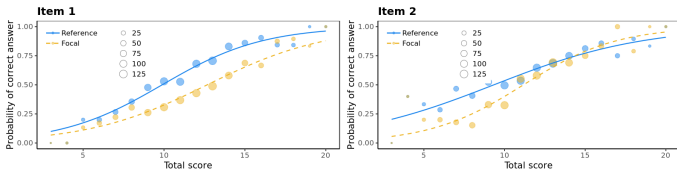
- DIF can be unrelated to total score differences, you may get:
  - No DIF item, but significant total score difference btw groups
  - DIF items present but total score distributions are the same for groups

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

Computerized adaptive testing (CAT)

Tailored testing
- The next administered item(s) depend on previous responses
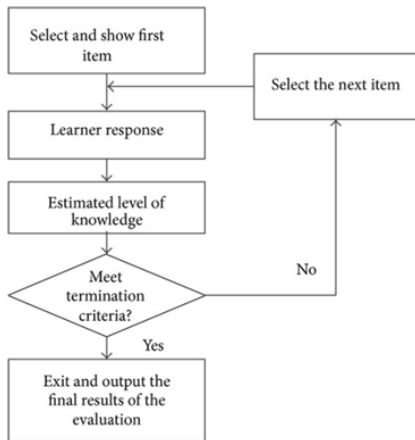
Software for psychometric analyses

General statistical software

- R https://cran.r-project.org/
  - Psychometric libraries see
    https://CRAN.R-project.org/view=Psychometrics
- SPSS
- Stata
- SAS, etc.

IRT software

- IRTPRO
- flexMIRT
- Winsteps
- ConQuest, etc.

ShinyItemAnalysis

ShinyItemAnalysis is an R package and an online application for
- teaching/learning psychometric methods (CTT, IRT, DIF,...)
- routine analysis of educational (e.g. admission) tests

With the aim of widespreading the metodology into distant fields and geographic regions.

www.ShinyItemAnalysis.org

Thank you for your attention!

www.cs.cas.cz/martinkova

martinkova@cs.cas.cz

References

- Martinková, Drabinová & Houdek (2017): ShinyItemAnalysis: Analysis of admission and other educational and psychological tests. Testfórum, 9, str.16-35. http://dx.doi.org/10.5817/TF2017-9-129
- Martinková, Drabinová, Leder & Houdek (2018): ShinyItemAnalysis: Test and Item Analysis with Shiny. Version 1.2.8. www.ShinyItemAnalysis.org https://shiny.cs.cas.cz/ShinyItemAnalysis https://CRAN.R-project.org/package=ShinyItemAnalysis
- McFarland, Price, Wenderoth, Martinková, et al. Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35, 2017.
- 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

Vocabulary

- Latent variable
- Reliability, measurement error
- Validity
- Traditional item analysis
- Item response theory models (IRT)
- Differential item functioning (DIF)
- Computerized adaptive testing (CAT)