



VYBRANÁ TÉMATA Z PSYCHOMETRIE

NMST 570



Patrícia Martinková

Odd. statistického modelování, ÚI AV ČR
ÚVRV, Pedagogická fakulta UK

www.cs.cas.cz/martinkova

martinkova@cs.cas.cz

Úvod do psychometrie

1

- **Psychometrie**

Vědní oblast zabývající se teorií a technikami psychologického měření

- Měření schopností, znalostí
 - Přijímací testy, maturita, mezinárodní studie (PISA, PIRLS, TIMSS,...)
- Měření osobnostních rysů, postojů
 - BIG 5, testy emoční inteligence, testy psychických poruch,
- Měření únavy, bolesti, kvality života, atp.

Kombinuje poznatky z psychologie, pedagogiky, matematické statistiky

Seminář NMST570

2

- **Cíle semináře**
 - Představit vybraná témata z psychometrie
 - Demonstrovat metody na reálných datech
 - Propojit studenty matematické statistiky se studenty dalších oborů

- **Kritéria pro zápočet**
 - Aktivní účast (2 a méně absencí, vypracování úkolů zadaných v hodině)
 - Report / poster / prezentace (22. 5. 2018)

Předběžný harmonogram

3

- 27. února: Úvod
- 13. března: Chyby měření a reliabilita
- 27. března (nebo 3. dubna): IRT modely
- 10. dubna: **Návrhy projektů (á 5-10 min.)**, IRT modely
- 24. dubna: Modely přidané hodnoty škol
- **8. květen: Den osvobození (výuka se nekoná)**
- 22. května: **Postery a prezentace (á 10 min.)**

Projekt PRIMUS

4

“Centrum tvorby didaktických testů a psychometrického modelování”

<http://pages.pedf.cuni.cz/uvrv/cemp-cz/>

- 2018 – 2020
- Pokrývá metodologii psychometrických modelů
- Hlavní zaměření: rozdíly mezi skupinami
 - danými např. pohlavím, krajem, typem studované školy, etnickou příslušností, rodným jazykem aj.
 - v odpovědích na položky znalostních testů a v dalším hodnocení ve vzdělávání

Konference

5

- Psychometric Society
 - Konference IMPS, 10.-13. července 2018, New York
 - www.psychometricsociety.org
- FRontiers in Educational Measurement Oslo (FREMO)
 - 12.-13. září, 2018, Oslo
 - <http://www.uv.uio.no/cemo/english/conferences/fremo2018/>
- Psychologická diagnostika Brno
 - listopad 2017, Brno
 - www.psychodiagnostika.fss.muni.cz
- Psychoco
 - únor 2018, Tübingen, Německo
 - únor 2019, Praha, ČR
 - www.psychoco.org

Vybraná témata z psychometrie

Vybraná témata z psychometrie

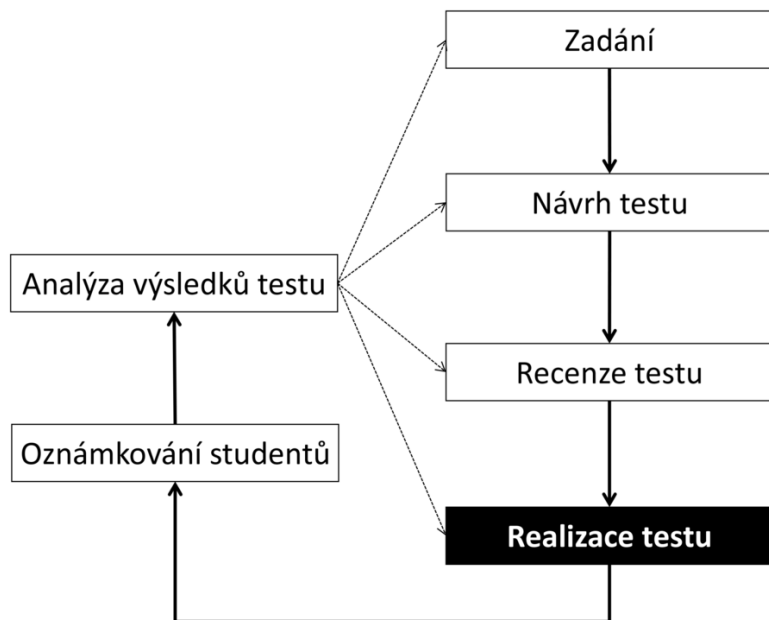
7

- Proces vývoje a validizace testů
- Spolehlivost (reliabilita) měření
- Validita
- Klasická analýza položek (obtížnost, diskriminační schopnost)
- Modely teorie odpovědi na položku (Item Response Theory, IRT)
- Detekce odlišného fungování položek pro skupiny (DIF, DDF)
- Počítačové adaptivní testování (Computerized Adaptive Testing, CAT)
- Propojení více forem testu (Equating and linking)
- Přidaná hodnota škol a učitelů (Value Added Models, VAM)
- ... a další

Vývoj a validizace testů

8

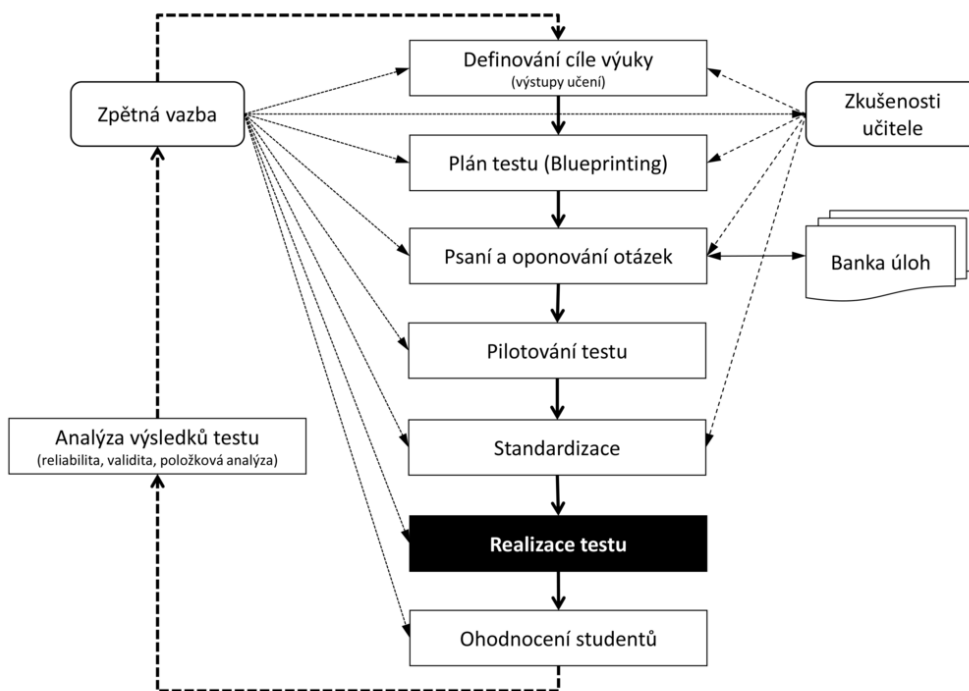
- Viz také [Seminář 2 v ZS 2017](#)
- Cyklus vývoje testu



Vývoj a validizace testů

9

- Viz také [Seminář 2 v ZS 2017](#)
- Cyklus vývoje testu

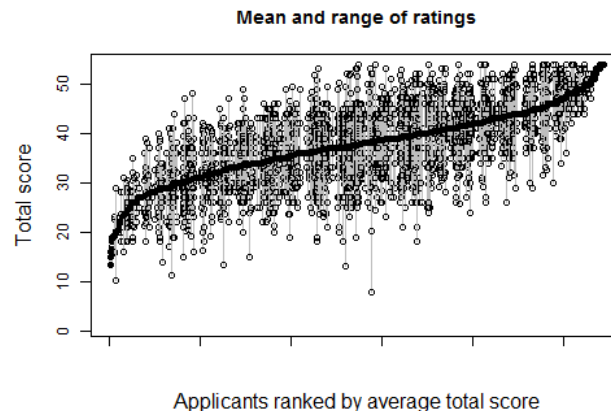
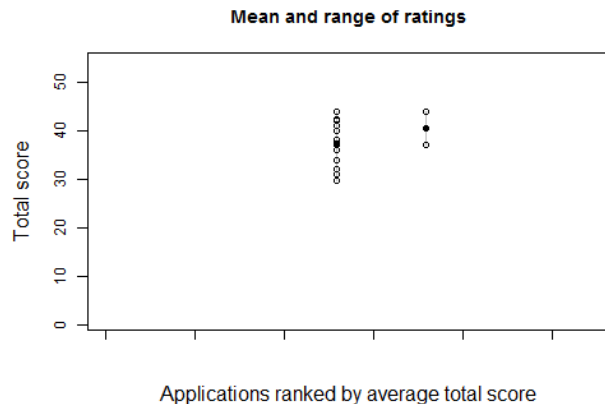


Reliabilita

10

Reliabilita - test „není zatížený chybami“

- **Test-retest reliabilita:** Výsledky se nezmění, pokud měření provedeme (nezávisle) znovu
- **Vnitrořídňní konzistence (Cronbachovo alfa):** Položky měří stejný konstrukt
- **Inter-rater reliabilita:** hodnocení jednotlivých hodnotitelů (nebo komisí) jsou konzistentní



Validita

11

Důkazy validity - obsahové:

- **Zjevná validita** (face validity)
- **Obsahová validita** (content validity)
- **Konstruktová validita** (construct validity)

Plynou z procesu vývoje testu:

- Definice testované oblasti (**Conceptual Framework**)
- Rozvržení položek v kategoriích (**blueprint**)

Příklad ověření obsahové validity:

[McFarland, Wenderoth, et. al. \(2017\)](#)

Důkazy validity - empirické:

- **Predikční validita:** Počet bodů v testu predikuje budoucí úspěšnost
- **Inkrementální:** Počet bodů v testu vylepšuje predikci budoucí úspěšnosti studenta (nad tu obsaženou v jiných dostupných údajích)
- **Diskriminační:** Test dobře rozlišuje mezi skupinami, mezi kterými by měl rozlišovat
- **Souběžná:** Výsledky v testu korelují s výsledky v jiných testech

Příklad využití empirických metod pro ověření validity:

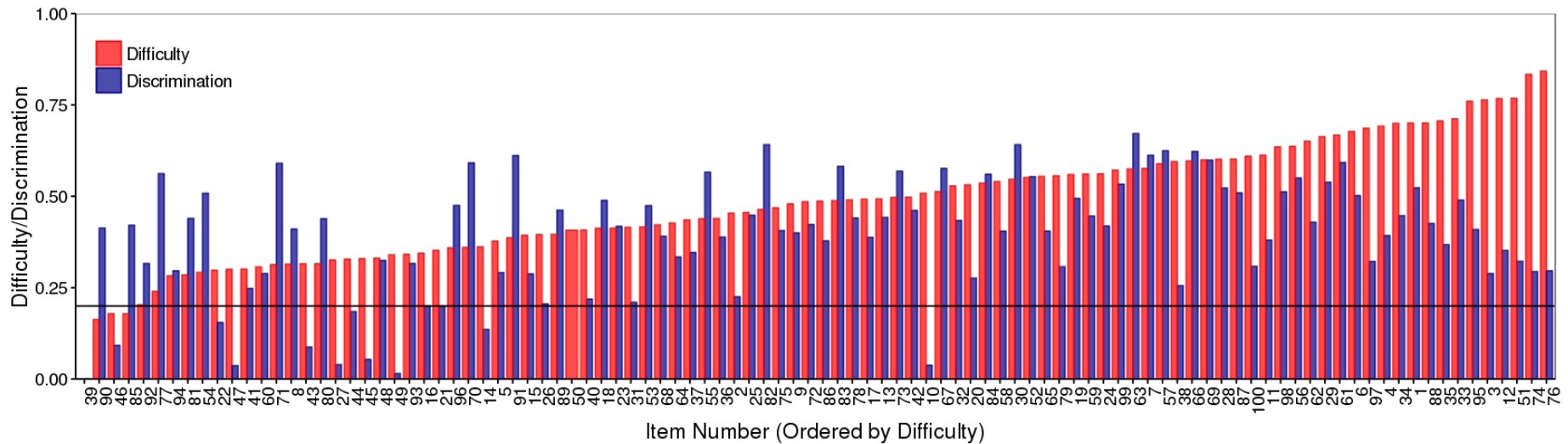
[McFarland, Price, et al. \(2017\)](#)

Analýza položek

13

Tradiční analýza položek

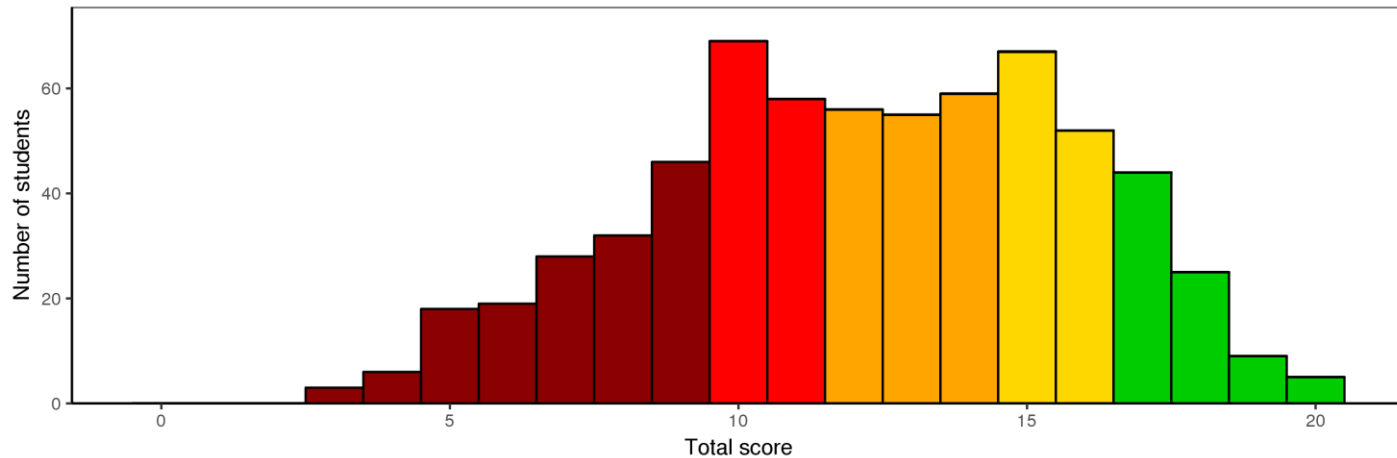
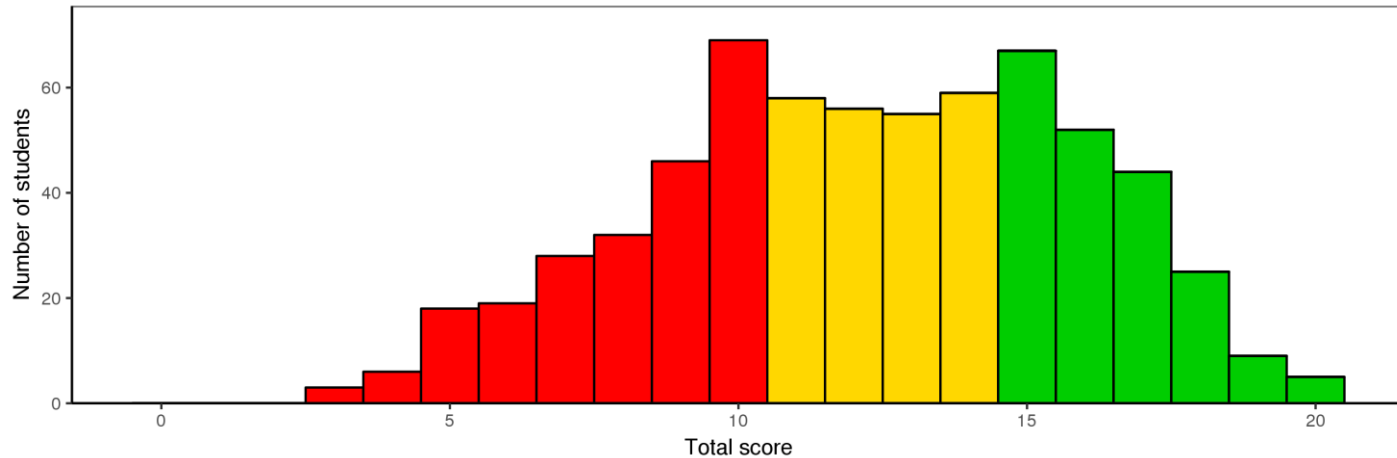
- Obtížnost (procento správných odpovědí)
- Diskriminační schopnost (Upper-Lower Index, ULI)



Analýza distraktorů

14

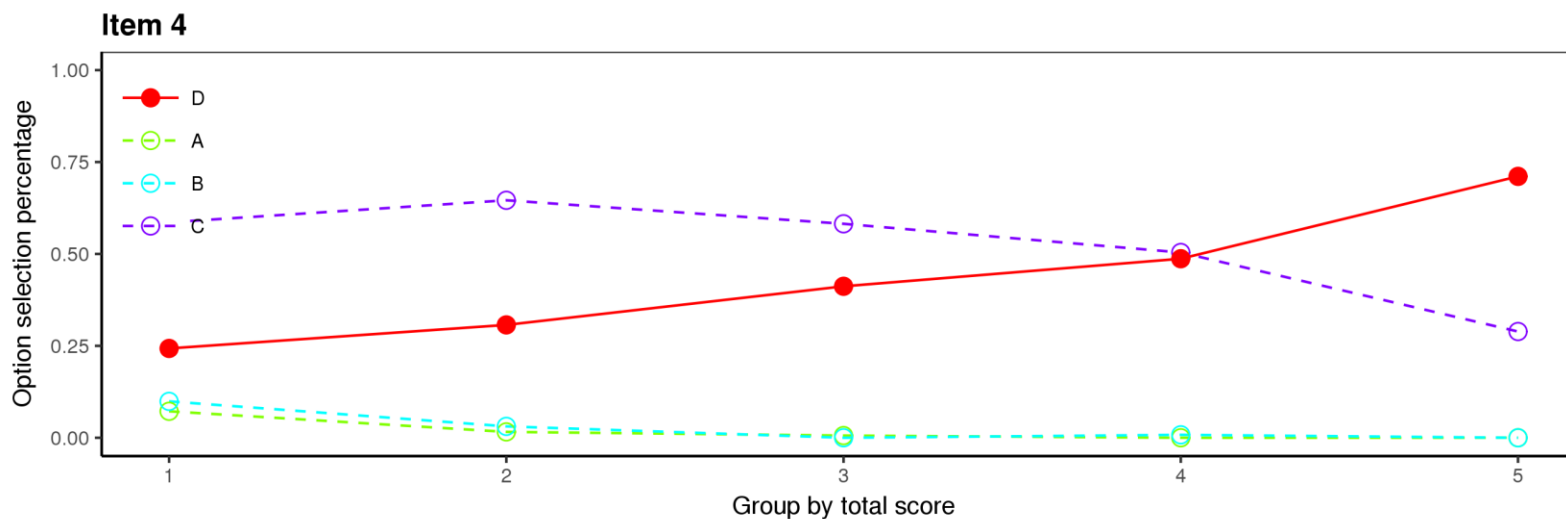
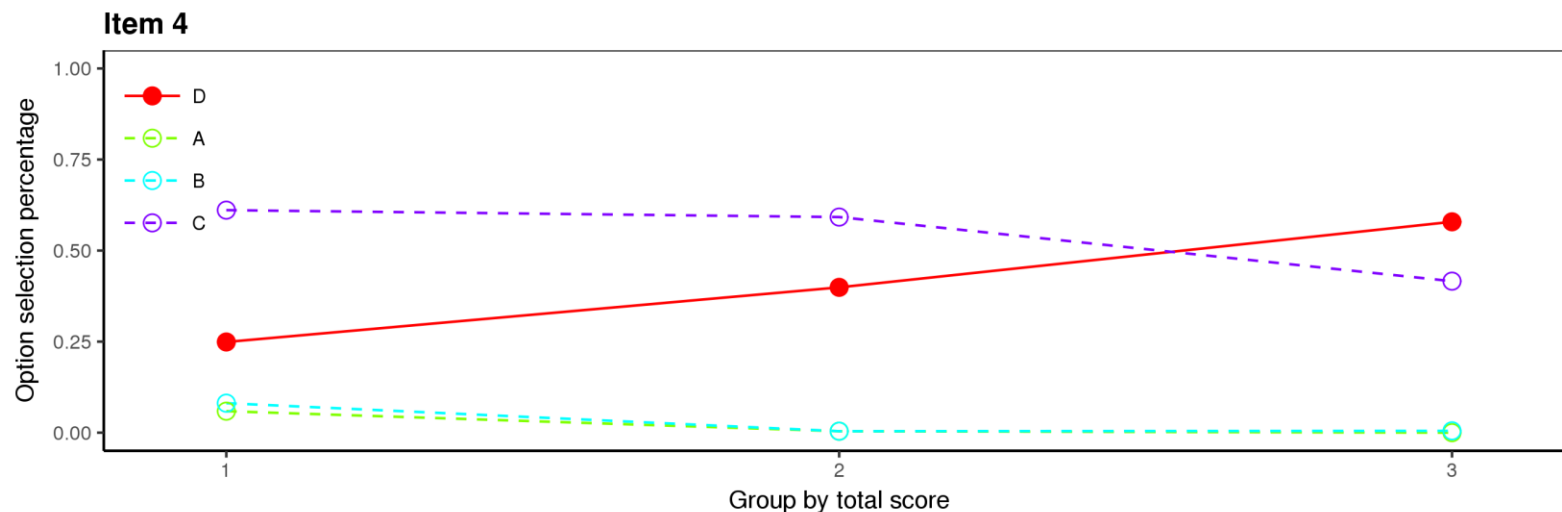
Rozdělení studentů do skupin dle celkového skóre



Analýza distraktorů

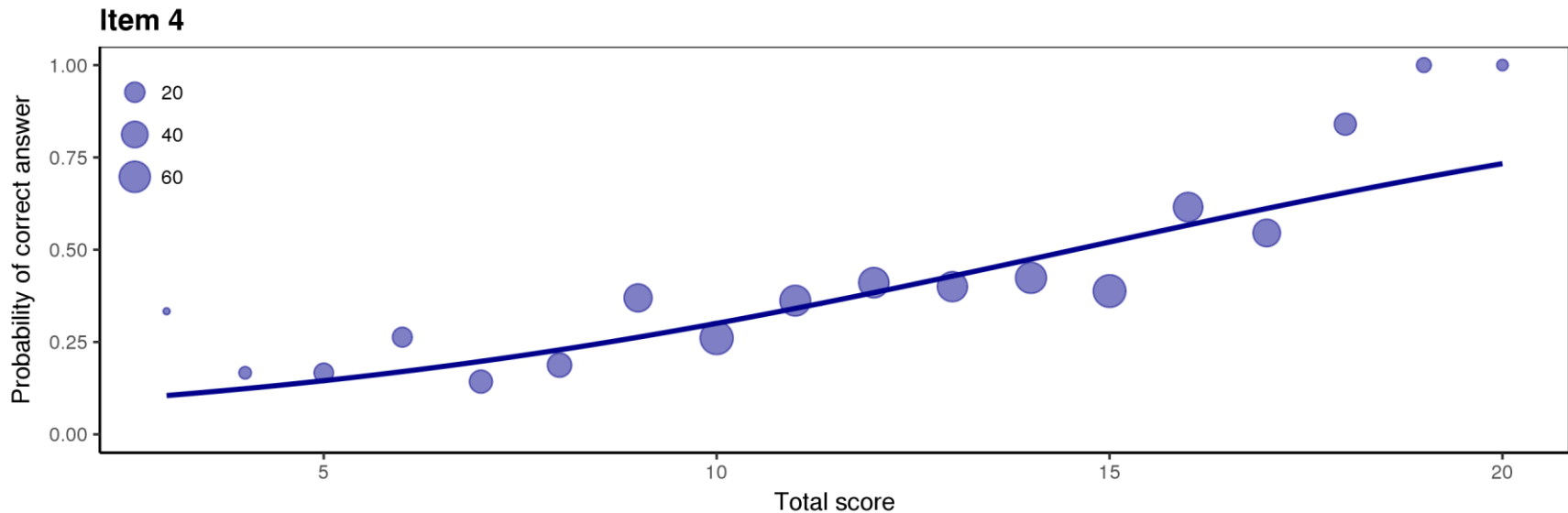
15

Procento volby odpovědi dle celkového skóre



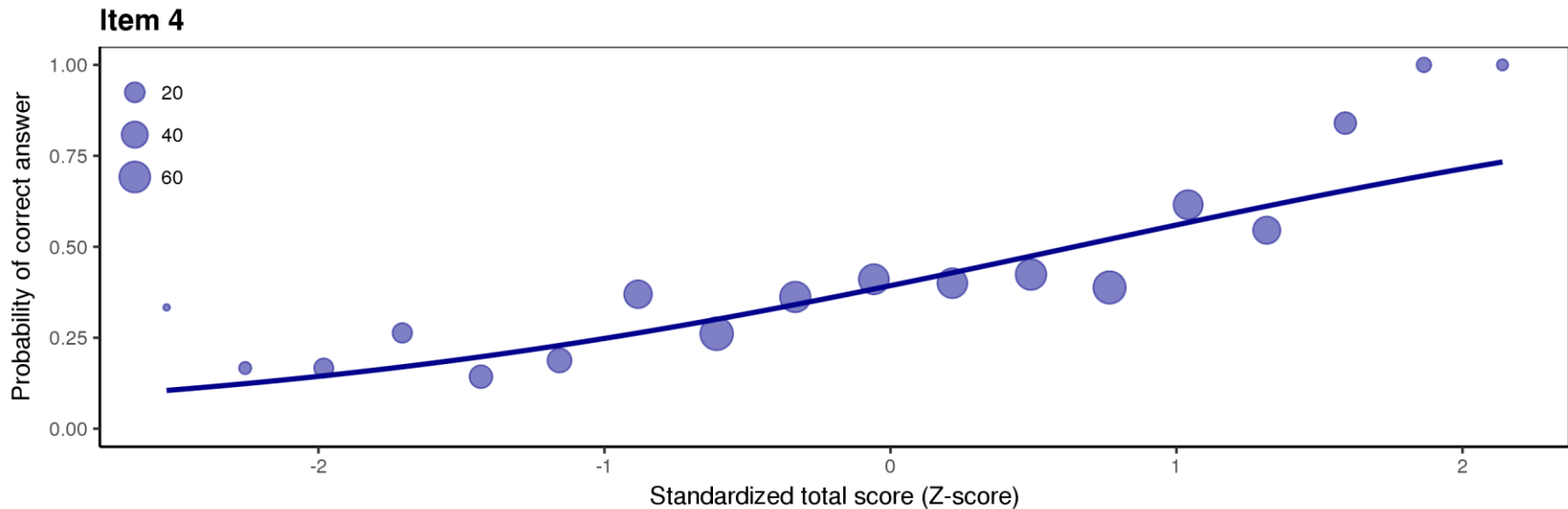
Logistická regrese

- Modeluje pravděpodobnost správné odpovědi v závislosti na celkovém skóre



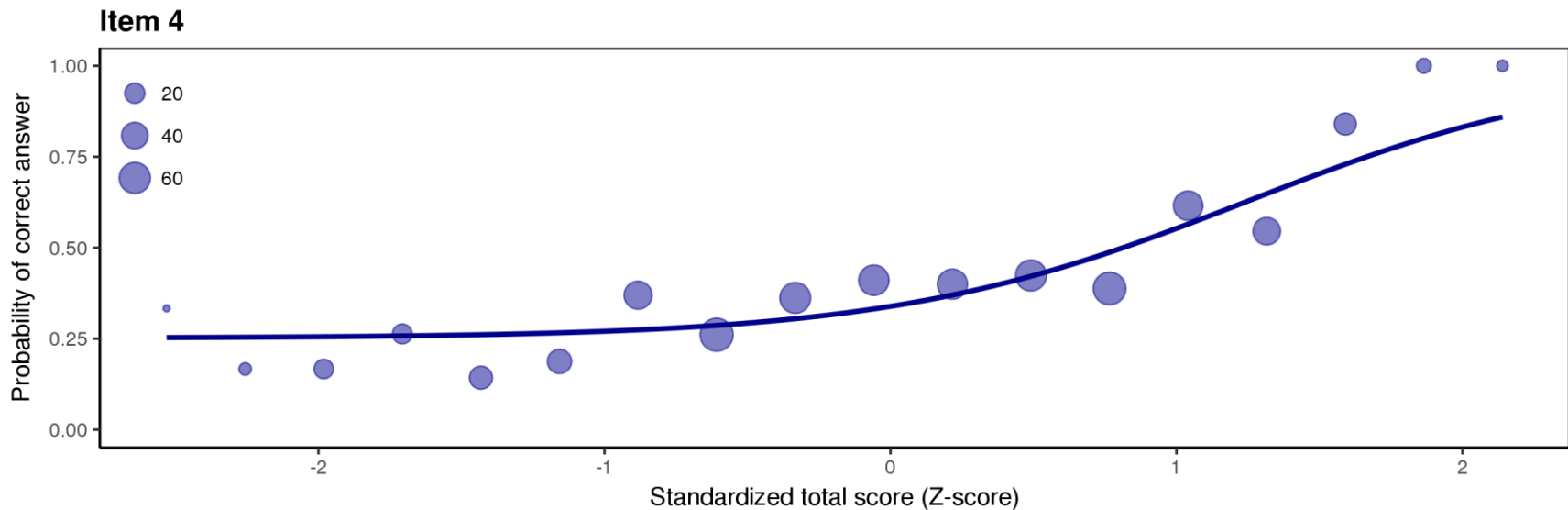
Logistická regrese

- Lze také modelovat pravděpodobnost správné odpovědi v závislosti na *standardizovaném skóre*
Lze měnit parametrizaci
- Stejná křivka, jiné odhady parametrů a interpretace



Nelineární regrese

- Modeluje pravděpodobnost správné odpovědi v závislosti na celkovém/standardizovaném skóre
- Dovoluje nenulovou dolní asymptotu (uhádnutelnost) resp. horní asymptotu různou od jedné (nepozornost)

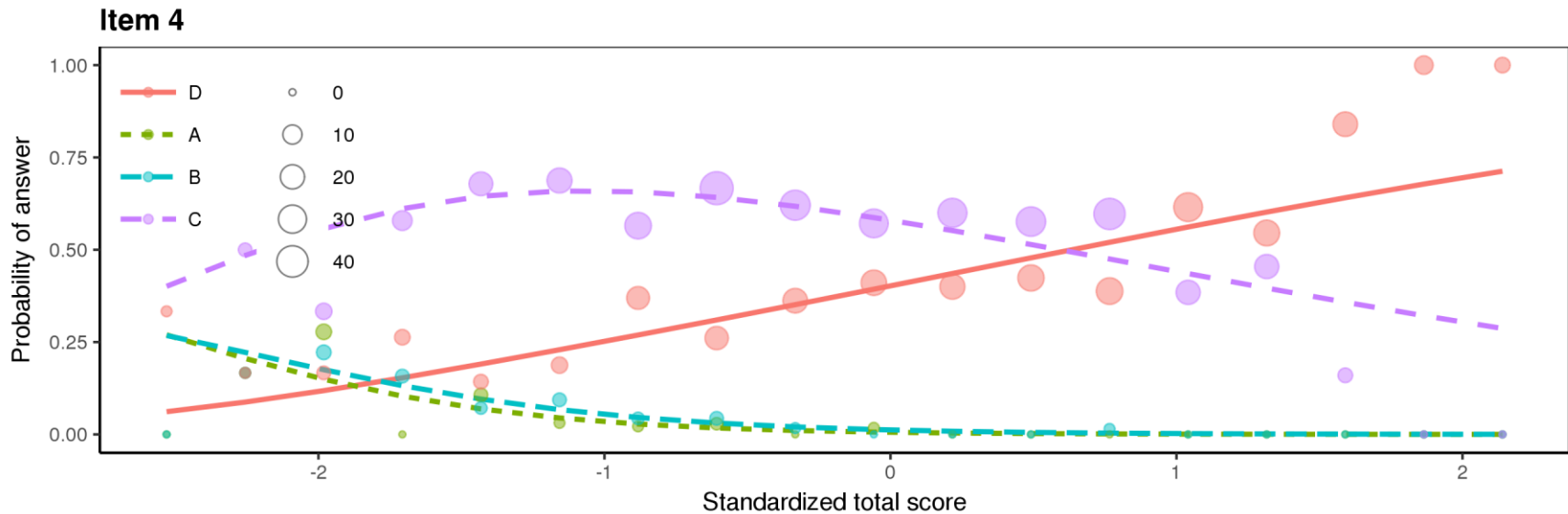


Využití regresních metod

19

Multinomická regrese

- Modeluje pravděpodobnost volby jednotlivých distraktorů v závislosti na celkovém/standardizovaném skóre



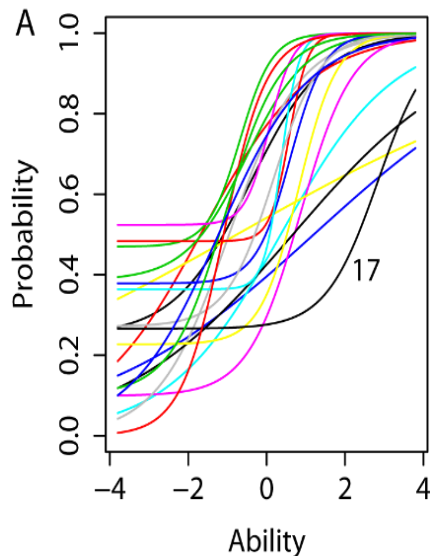
Item Response Theory (IRT) modely

20

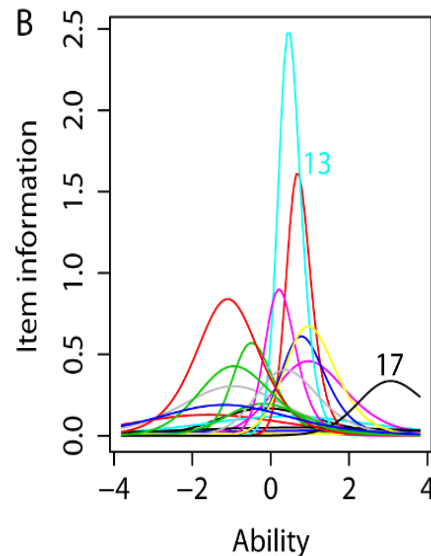
Modely Teorie odpovědi na položku

Viz také [Seminář 3 a 4 v ZS 2017](#)

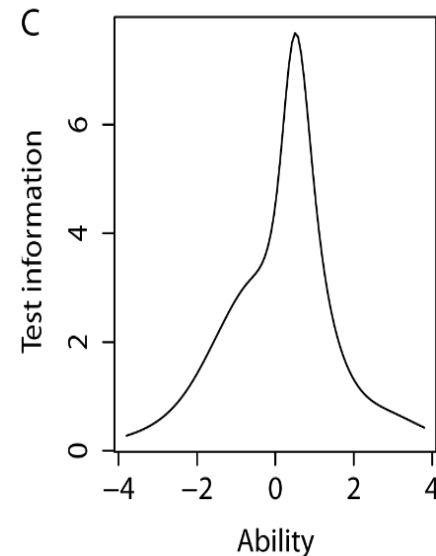
- Smíšené zobecněné (ne)lineární modely
- Znalost modelována jako náhodný efekt



Charakteristické funkce položek (ICC)



Informační funkce položek (IIC)



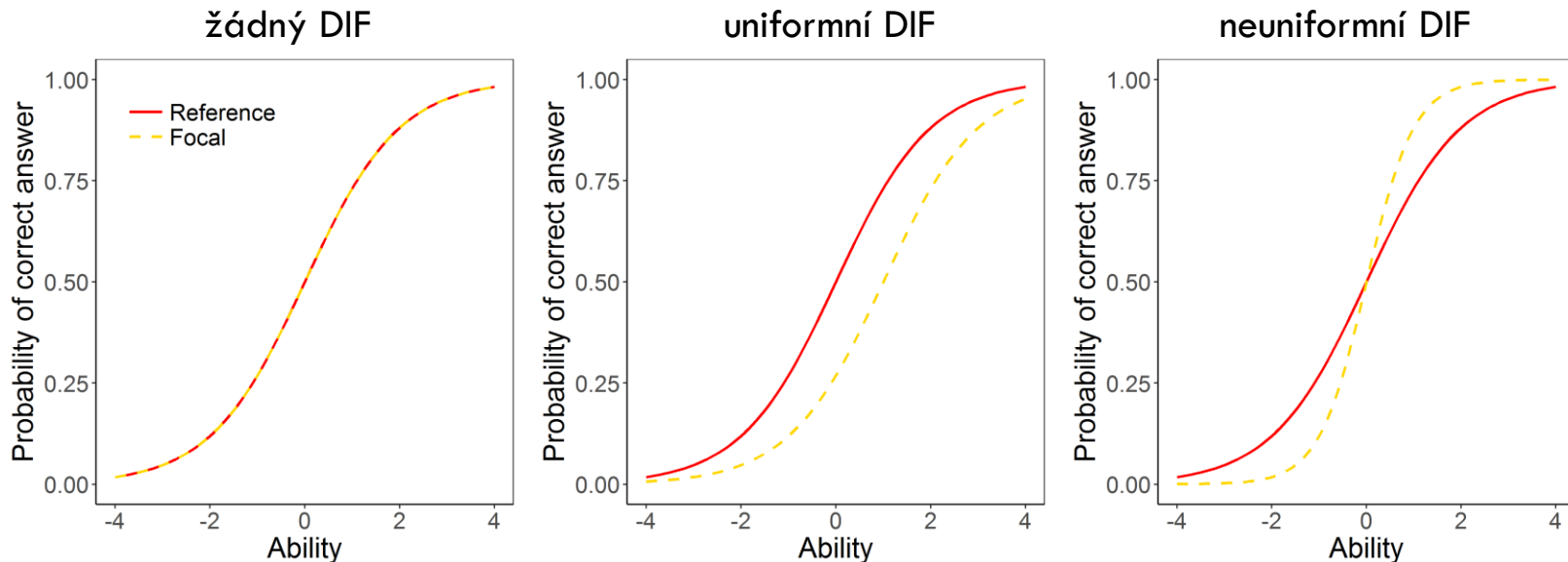
Informační funkce testu (TIC)

Detekce odlišného fungování položek

21

Differential Item Functioning (DIF)

- Viz také [Seminář 5 v ZS 2017](#)
- Pro jedince se stejnou latentní znalostí ale z různých skupin se liší pravděpodobnost správného zodpovězení položky

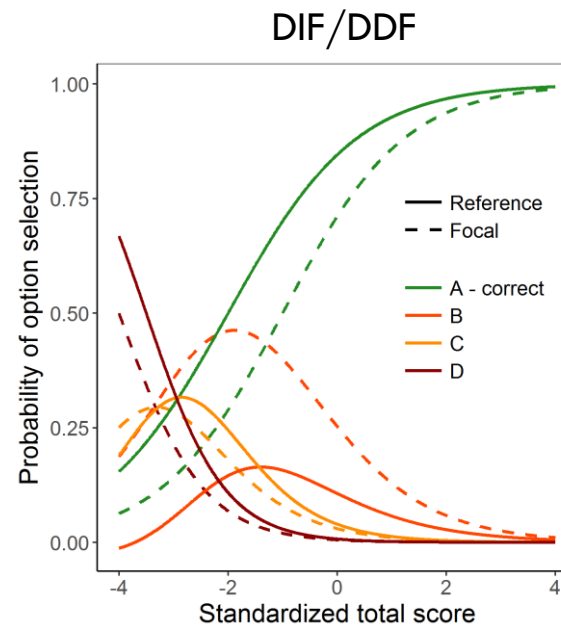
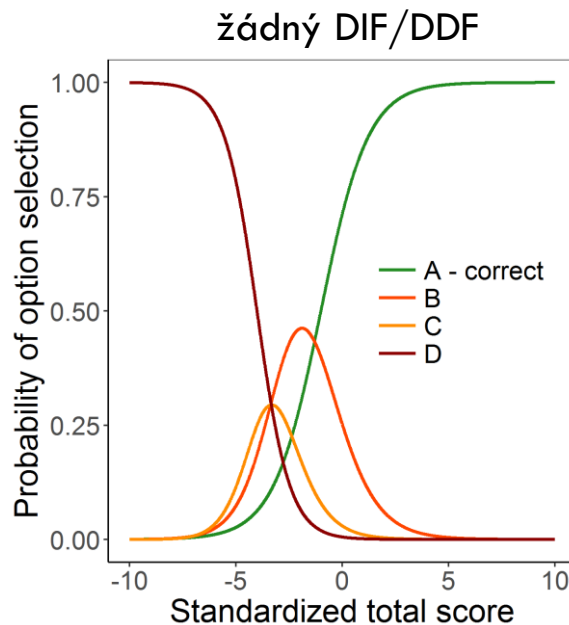


Detekce odlišného fungování položek

22

Differential Distractor Functioning (DDF)

- Viz také [Seminář 5 v ZS 2017](#)
- Pro jedince se stejnou latentní znalostí ale z různých skupin se liší pravděpodobnost volby jednotlivých distraktorů

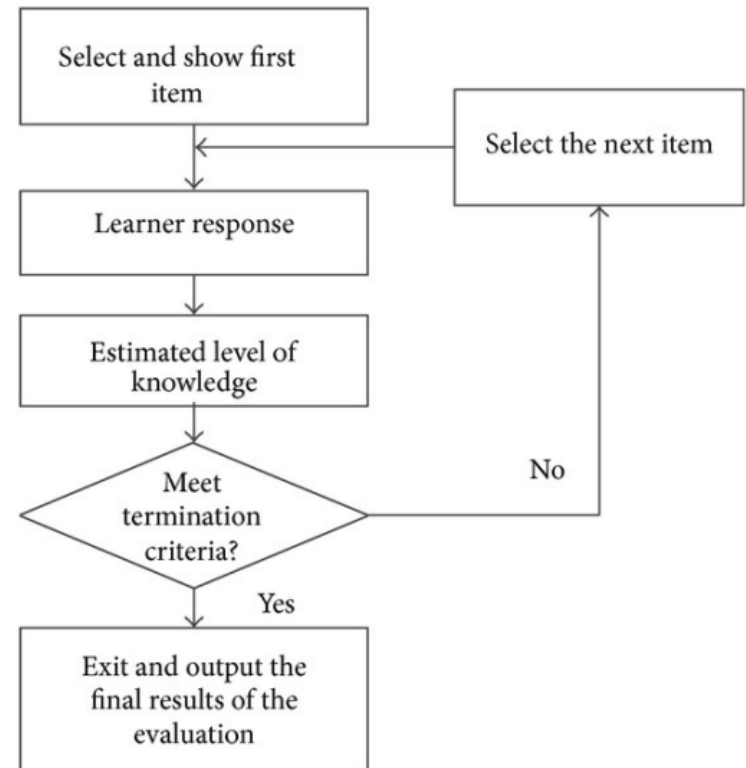
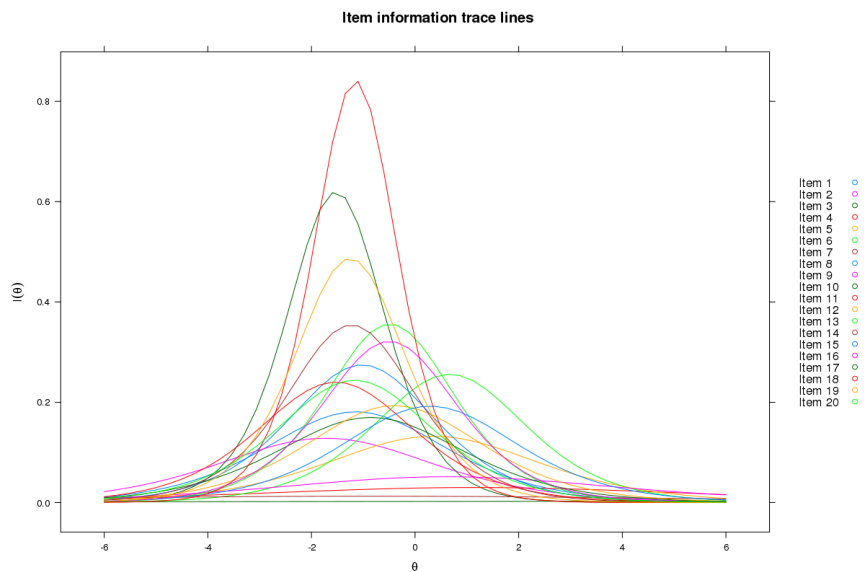


Adaptivní testy

23

Computerized Adaptive Tests (CAT)

- Viz také [Seminář 6 v ZS 2017](#)
- Položky „šité na míru“



Software

Software

25

Obecný statistický software:

- R <https://cran.r-project.org/>
 - Psychometrické knihovny v R: <https://CRAN.R-project.org/view=Psychometrics>
 - ShinyItemAnalysis
 - webové rozhraní <https://shiny.cs.cas.cz/ShinyItemAnalysis/>
 - R knihovna <https://CRAN.R-project.org/package=ShinyItemAnalysis>
- SAS, Stata, SPSS, ...

IRT software:

- IRTPRO
- flexMIRT
- Winsteps
- ConQuest...

ShinyItemAnalysis

26

- Dostupná jako knihovna v R na CRAN
- Dostupná online:

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

- Martinková P, Drabinová A, Houdek, J. (2017) ShinyItemAnalysis: [Analýza přijímacích a jiných znalostních testů](#). *Testforum*, no. 9, pp.16-35.

Aplikace pro analýzy: ShinyItemAnalysis

27

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 **GMAT** from R **difNLR** package is used. Other three datasets are available: **GMAT2** and **Medical 20 DIF** from **difNLR** 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

© 2017 Patricia Martinkova, Adela Drabinova, Ondrej Leder and Jakub Houdek



Hits:1871

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

Praktické příklady

28

- ShinyItemAnalysis
 - <https://shiny.cs.cas.cz/ShinyItemAnalysis/> (aktuální verze z CRANu online)
 - <https://shiny.cs.cas.cz/ShinyItemAnalysisTest/> (nová verze)
 - Nebo lokálně v R (aktuální verze z CRANu je 1.2.3):

```
library(ShinyItemAnalysis)  
startShinyItemAnalysis()
```
- Pracujte dle [tutoriálu k Lekci 1](#) ze zimního semestru
 - Probádejte cvičný datový soubor GMAT
- Pracujte dle [tutoriálu k Lekci 2](#) ze zimního semestru
 - [Stáhněte si data HCI](#), projděte si analýzy a vygenerujte report
- Pracujte s novými daty
 - [Stáhněte si maturitní data](#) a proved'te analýzu.

Náměty pro projekty

Náměty na projekty

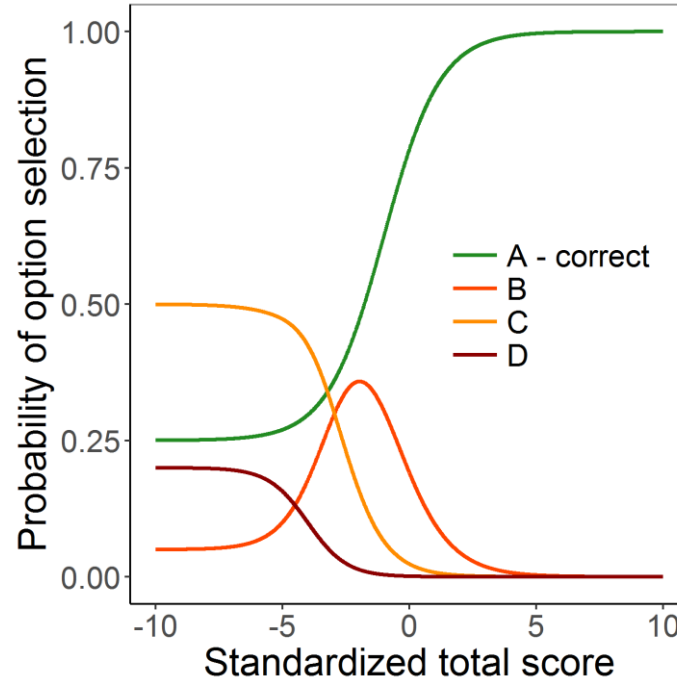
30

- Validizace testu
 - Zadejte nový nebo známý znalostní/psychologický test, nebo překlad známého testu vzorku respondentů
 - Proveďte validizaci, porovnejte funkčnost testu s literaturou
- Analýza odlišného fungování položek pro skupiny (DIF, DDF)
 - Využijte publikovaná data OECD (např. PISA, TIMSS)
 - Analyzujte DIF vzhledem ke státní příslušnosti, pohlaví, apod.
- Počítačové adaptivní testování (CAT)
 - Na základě známých parametrů fixního testu vytvořte adaptivní test
 - Proveďte simulační studii, jak by se test choval jako adaptivní

Náměty na projekty

31

- Multinomický model pro popis položek
 - Upravte multinomický model tak, aby bylo možné modelovat různou atraktivitu distraktorů
 - Porovnejte na reálných datech a simulacemi



Náměty na projekty

32

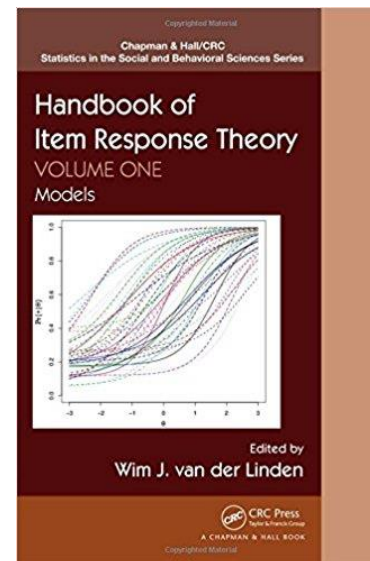
- IRT modely
 - Vyberte si jeden z komplexnějších IRT modelů, popište jej
 - Implementujte v R na vhodných datech
 - Příp. porovnejte s jiným modelem simulacemi

Literatura:

Wim J. van der Linden (Ed.):

Handbook of Item Response Theory.

Vol.1-3. CRC Press, 2017.



Děkuji za pozornost!

martinkova@cs.cas.cz

www.cs.cas.cz/martinkova

Reference

- Martinková P, Drabinová A, Houdek, J. (2017). ShinyItemAnalysis: [Analýza přijímacích a jiných znalostních testů](#). Test Fórum, no. 9, pp.16-35.
- Martinková P, Štěpánek L, Drabinová A, Houdek J, Vejražka M, Štuka Č. (2017). [Semi-real-time analyses of item characteristics for medical school admission tests](#). *Annals of Computer Science and Information Systems*, 11, pp. 189-194.
- McFarland JL, Price RM, Wenderoth MP, Martinková P, et al. [Development and Validation of the Homeostasis Concept Inventory](#). *CBE Life Sciences Education*, vol. 16 no. 2 ar35, 2017.