

Analýza položek přijímacích testů na lékařskou fakultu v semi-reálném čase

—
4. ročník konference Psychologická diagnostika Brno 2017

Lubomír Štěpánek¹
Adéla Drabinová²
Martin Vejražka³

Patrícia Martinková²
Jakub Houdek²
Čestmír Štuka¹



¹Ústav biofyziky a informatiky

³Ústav lékařské biologie a laboratorní diagnostiky

1. lékařská fakulta, Univerzita Karlova



²Ústav informatiky
Akademie věd České republiky

(2017) Lubomír Štěpánek, Patrícia Martinková, Adéla Drabinová,
Jakub Houdek, Martin Vejražka, Čestmír Štuka, CC BY 3.0 (CZ)



Dílo lze dále svobodně šířit a upravovat, ovšem s uvedením původního autora. Jiná omezení nejsou kladena. Autoři vynaložili nezanedbatelné úsilí, aby byla uvedená fakta nejsprávnější možná, a práci sepsali podle svého nejlepšího vědomí a svých „nejlepších“ znalostí problematiky.

Obsah

- 1 Úvod
- 2 Aktuální stav věci
- 3 Analýza položek a testů
- 4 Aplikace ShinyItemAnalysis
- 5 Shrnutí
- 6 Zdroje

Rychlý úvod

- přijímací řízení je důležitým procesem pro uchazeče i instituce
- proto je vhodná rutinní evaluace
- rychlá a efektivní metodologie a online nástroj pro hodnocení v semi-reálném čase může být cestou

Náš tým



Patricie
Martinková¹



Adela
Drabinová¹



Jakub
Houdek¹



Lubomír
Štěpánek²



Martin
Vejražka³



Čestmír
Štuka²



¹Ústav informatiky
Akademie věd České republiky



²Ústav biofyziky a informatiky

³Ústav lékařské biochemie a laboratorní diagnostiky
1. lékařská fakulta, Univerzita Karlova

- podpořeno grantem GAČR GJ15-15856Y

Hodnocení přijímacích řízení obecně

- v některých zemích je standardizované testování rutinou
- v Česku zájem o standardní testování variuje napříč institucemi
 - robustní psychometrické testování spíše vzácně

Dostupné metody pro hodnocení testů a položek

- CTT (Classical Test Theory), klasická testová teorie
 - přehledové statistiky, celková skóre
 - histogramy celkových skóre
 - korelační heatmapy
 - Cronbachovo α
 - obtížnost a diskriminace položky

ULI (Upper-Lower Index)

ULI (Upper-Lower Index) $\stackrel{\text{def}}{=}$ rozdíl podílu správných odpovědí v nejúspěšnější a nejméně úspěšné třetině testovaných (dle celkového skóre)

Dostupné metody pro hodnocení testů a položek

- IRT (Item Response Theory), teorie odpovědi na položku
 - používá se j -PL (parametrový logistický) model pro $j \in \{1, 2, 3, 4\}$ tak, že parametry j -PL modelu odpovídají prvním j elementům z množiny

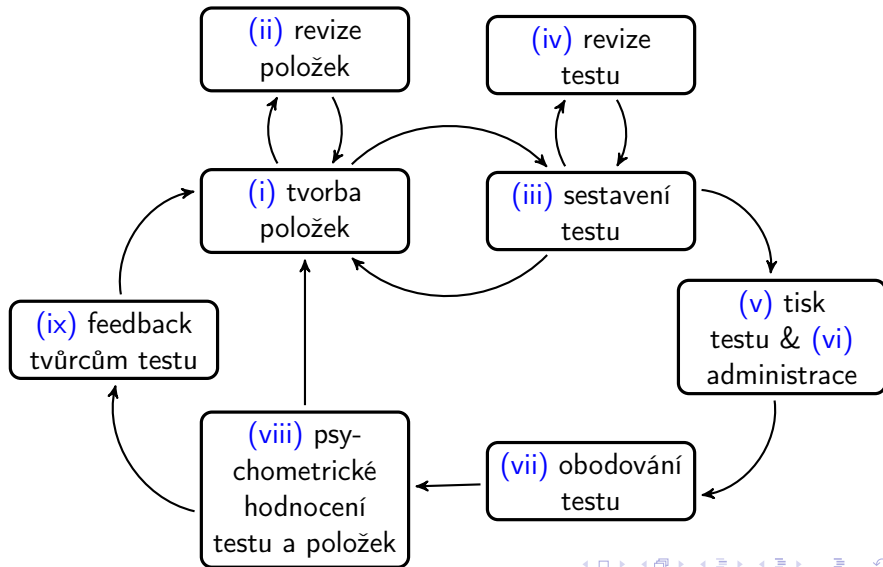
{obtížnost, diskriminace, „uhádnutí odpovědi“, nepozornost}

- differential item functioning (DIF)
- a další

Dostupné nástroje pro hodnocení testů a položek

- obecný statistický software či programovací jazyky
 - jazyk R
 - SPSS
 - STATA
 - SAS
 - a další
- specializovaný software pro položkovou analýzu
 - Iteman
 - Rogo
 - Winsteps
 - IRTPRO
 - ConQuest
 - a další

Vývoj položek a testů na 1. lékařské fakultě UK



Vývoj položek a testů na 1. lékařské fakultě UK

- body (i) – (vi), (ix) zaberou relativně dlouhý čas v pretestovém období („asynchronní úlohy“)
- body (vii) – (viii) by naopak měly být provedeny co nejdříve, jakmile uchazeči dopíší test („synchronní úlohy, úlohy v semi-reálném čase“)
 - detekovaná suspektní položka by neměla být započítána do celkového skóre
 - v bodě (ix) je taková položka poslána zpět autorovi k přeformulování či vyřazení z položkové banky

Analýza položek a testů na 1. lékařské fakultě UK

- přijímací testy na 1. lékařskou fakultu UK jsou založeny na *multiple true-false* položkách
- byla vyvinuta grafická prezentace vlastností položek
 - diagram distraktorů
- protože je přijímána jen nejúspěšnější pětina uchazečů, jsou v definici ULI uvažovány kvintily místo tercilů

Charakter dat sesbíraných z odpovědních archů uchazečů

- v bodě (vii) je test automatizovaně obodován
- výstupem je vektor celkových skóre uchazečů a dataset obsahující odpovědní schémata všech uchazečů na všechny položky
 - každý řádek tvořen odpověďmi uchazeče
 - každý sloupec obsahuje odpovědi všech uchazečů na jednu položku

Metodologie hodnocení položek a testů na 1. LF UK

- nad vektorem celkových skóre uchazečů daného testu spočítány q -kvantily, kde $q \in \{2, 3, 4, \dots\}$, dělíci uchazeče do q skupin
- pro $q = 5$ spočítány kvintily $\{Q_0, Q_1, \dots, Q_5\}$ a i -tý uchazeč s celkovým skóre x_i zařazen do k -té skupiny tak, že

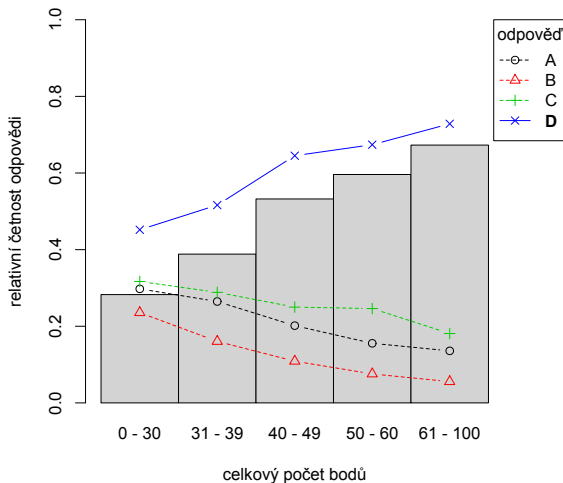
$$Q_{k-1} \leq x_i < Q_k,$$

kde $k \in \{1, 2, \dots, 5\}$

- pro k -tou skupinu, kde postupně $k \in \{1, 2, \dots, q\}$, je pro danou položku t spočítán podíl $u_{k,t}^{\{q\}}$ správných odpovědí ku všem odpovědím

Diagram distraktorů

**Psychometrické charakteristiky, položka 2014
všeobecné lékařství**



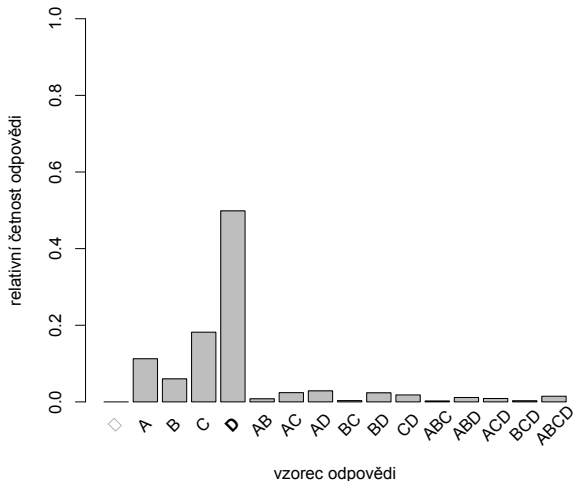
Odpovědní schéma uchazečů na danou položku

- všechny položky přijímacích testů jsou typu *multiple true-false* s možnými odpověďmi A, B, C, D
- existuje právě $2^{|\{A,B,C,D\}|} = 2^4 = 16$ možností (odpovědních schémat), jak na danou položku odpovědět, tedy pro odpovědní schéma \mathcal{O} platí

$$\begin{aligned}\mathcal{O} \in \{ & \emptyset, \\ & A, B, C, D, \\ & AB, AC, AD, BC, BD, CD, \\ & ABC, ABD, ACD, BCD, \\ & ABCD \}\end{aligned}$$

Diagram odpovědních schémat uchazečů na danou položku

Relativní četnost jednotlivých kombinací odpovědí, položka 2014
všeobecné lékařství



Obtížnost a diskriminace položky

- obtížnost diffc_t položky t je podíl všech uchazečů, kteří na položku odpověděli nesprávně, ku všem uchazečům, kteří na ni odpověděli, tedy

$$\text{diffc}_t = 1 - \sum_{k=1}^q u_{k,t}^{\{q\}}$$

- diskriminace $\text{discr}_t^{\{q\}}(l_1, l_2)$ položky t je rozdíl mezi dvěma podíly správných odpovědí, a to v l_1 -té a l_2 -té skupině, tedy

$$\text{discr}_t^{\{q\}}(l_1, l_2) = u_{l_2,t}^{\{q\}} - u_{l_1,t}^{\{q\}},$$

kde $l_1, l_2 \in \{1, 2, \dots, q\}$

Diskriminace položky a upper-lower index (ULI)

- zřejmě je $ULI_t = \text{discr}_t^{\{3\}}(1, 3)$
- pro naše účely je lepší volit $q = 5$ namísto tradičních $q = 3$, protože ke studiu je přijímána nejlepší cca pětina uchazečů
- speciálně nás u každé položky t zajímá kromě diskriminace mezi první a pátou pětinou, $\text{discr}_t^{\{5\}}(1, 5)$, i diskriminace mezi čtvrtou a pátou pětinou, $\text{discr}_t^{\{5\}}(4, 5)$
- čtvrtému kvintilu Q_4 celkového skóre zhruba odpovídá cut-off přijetí ke studiu

Diagram obtížnost-diskriminace (test chemie 2016)

Diagram obtížnosti vs. diskriminace
všeobecné lékařství

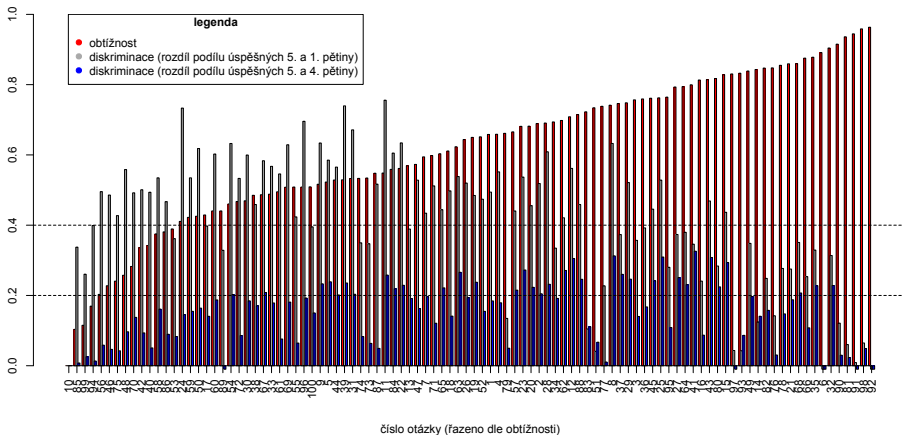
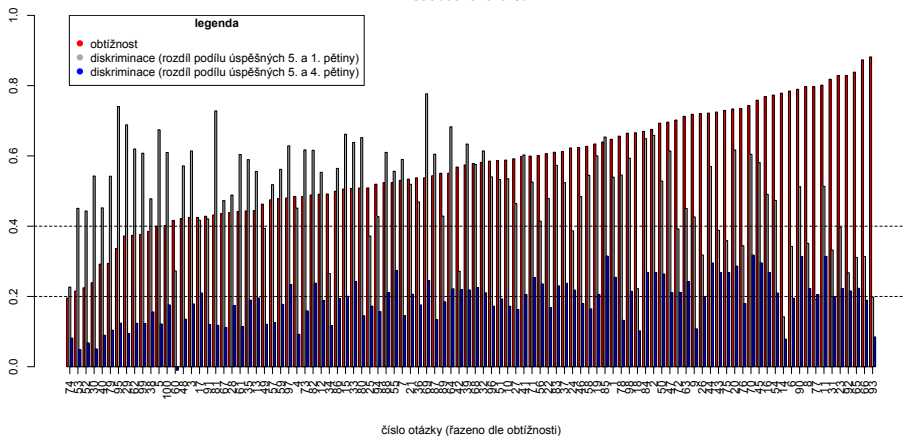


Diagram obtížnost-diskriminace (test chemie 2017)

Diagram obtížnosti vs. diskriminace
všeobecné lékařství



Stručně o ShinyItemAnalysis aplikaci

- metodologie implementována v online volně dostupné aplikaci a R-kovém balíčku ShinyItemAnalysis
- aplikace dostupná na

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

- balíček dostupný na [CRAN](#)
- nejnovější verze na [GitHub](#)
- zdrojový kód napsán v jazyce R

Co je jazyk R

- jde o *free-as-in-beer* a *free-as-in-speech* programovací jazyk a prostředí pro statistické výpočty a grafické výstupy
- široce užíván statistiky, ekonometry či biologi

```
1 | # příklad R-kového kódu =====
2 |
3 | isPrime <- function(n){
4 |
5 |     # '''
6 |     # Vrací TRUE, pokud je n prvočíslo, jinak vrací FALSE
7 |     # '''
8 |
9 |     sum(unlist(lapply(1:n, function(x) n %% x == 0))) == 2
10 |
11 | }
12 |
13 | isPrime(6)    ## FALSE
14 | isPrime(17)   ## TRUE
```

Rozhraní ShinyItemAnalysis aplikace

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 **QWAT** from R **difR** package is used. Other three datasets are available: **QWAT2** and **Medical 20 DIF** from **difR** 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 1.2.3. Version available [online](#) is 1.2.3. The newest development version available on [GitHub](#) is 1.2.3.

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

List of packages used

ShinyItemAnalysis [Test and item analysis](#) | Version 1.2.3

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



Hits 2466



Rozhraní ShinyItemAnalysis aplikace

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

Two parameter Item Response Theory model

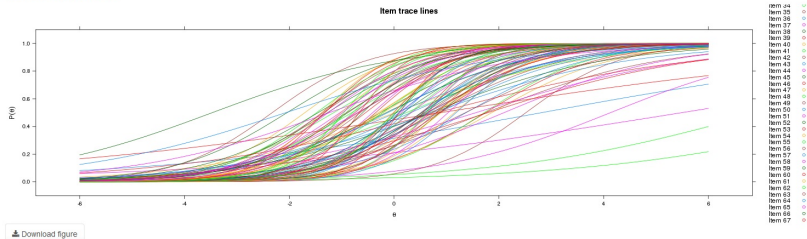
Item Response Theory (IRT) models are mixed-effect regression models in which student ability (θ) is assumed to be a random effect and is estimated together with item parameters. Ability (θ) is often assumed to follow normal distribution.

2PL IRT model allows for different slopes in inflection point – different discriminations **a**. Items can also differ in location of their inflection point – in item difficulty **b**.

Equation

$$P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

Item characteristic curves



Item information curves

ShinyItemAnalysis Test and item analysis | Version 1.2.3

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

Vlastnosti ShinyItemAnalysis aplikace

- mnohozáložkový layout
- každá záložka zobrazuje diagramy a další charakteristiky
- aplikace nabízí cvičné datasety, příklady R-kového kódu, rovnice a interpretaci
- aplikace umožňuje online sazbu $\text{T}_{\text{E}}\text{X}$ dokumentů a stáhnutelných .pdf a HTML reportů

Shrnutí

- Upper-Lower index byl generalizován pro individuální potřeby
- byla vytvořena ShinyItemAnalysis aplikace pro psychometrickou analýzu přijímacích testů a položek
- ShinyItemAnalysis nabízí grafický interface a webový framework, který otevírá funkcionality jazyka R široké audienci
- aplikace pokrývá řadu metod a nabízí příklady, cvičné datasety, rovnice modelů, odhady parametrů, interpretaci výsledků, R-kové kódy

Shrnutí

- aplikace je vhodná pro výukové účely i pro rutinní analýzy, včetně online generování online reportů
- využití ShinyItemAnalysis aplikace na lékařské fakultě potvrdilo užitečnost a dostupnost aplikace pro rutinní analýzy

Zdroje



MARTINKOVÁ, Patrícia; ŠTĚPÁNEK, Lubomír; DRABINOVÁ, Adéla; HOUDEK, Jakub; VEJRAŽKA, Martin; ŠTUKA, Čestmír. Semi-real-time analyses of item characteristics for medical school admission tests. *Annals of Computer Science and Information Systems*, M. Ganzha, L. Maciaszek, M. Paprzycki (ed). ACSIS. 2017, roč. 11, s. 189–194. Dostupné z DOI: 10.15439/2017F380.



MARTINKOVÁ, Patrícia; DRABINOVÁ, Adéla; LEDER, Ondřej; HOUDEK, Jakub. *ShinyItemAnalysis: Test and Item Analysis via Shiny*. 2017. Dostupné také z: <https://cran.r-project.org/package=ShinyItemAnalysis>.



MARTINKOVÁ, Patrícia; DRABINOVÁ, Adéla; HOUDEK, Jakub. ShinyItemAnalysis: Analyzing admission and other educational and psychological tests. *Testforum*. 2017. Dostupné z DOI: 10.5817/TF2017-9-129.

Děkuji za pozornost!

lubomir.stepanek@lf1.cuni.cz

lubomir.stepanek@fbmi.cvut.cz

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