



ÚSTAV INFORMATIKY
Akademie věd České republiky



PEDAGOGICKÁ FAKULTA
Ústav výzkumu a rozvoje vzdělávání
Univerzita Karlova

Analýzy znalostních testů: Od klasické testové teorie po IRT modely a další témata



Workshop pro odbornou veřejnost
23. 10. 2023

T **A**
Č **R**

Projekt TL05000008 Výzvy pro hodnocení znalostí: Analytická podpora tvorby znalostních testů je spolufinancován se státní podporou Technologické agentury ČR v rámci Programu ÉTA 5.

Projekt EduTest – prezentující, autoři analýz

- doc. RNDr. Patrícia Martinková, Ph.D. Pedagogická fakulta UK & Ústav informatiky AV ČR
- Mgr. Eva Potužníková, Ph.D. Pedagogická fakulta UK
- Mgr. Jan Netík Ústav informatiky AV ČR, doktorand PedF UK
- Mgr. Adéla Hladká, Ph.D. Ústav informatiky AV ČR

- Mgr. Lucie Zelená Centrum pro zjišťování výsledků ve vzdělávání (aplikační garant)

Další autoři prezentovaných analýz:

- Mgr. Michaela Vařejková Ústav informatiky AV ČR, doktorandka MFF UK
- Mgr. Jana Dlouhá Ústav informatiky AV ČR, doktorandka FF UK
- MUDr. Ing. Lubomír Štěpánek Ústav informatiky AV ČR, doktorand 1. LF UK
- Mgr. Iván Pérez Ústav informatiky AV ČR, doktorand MFF UK

Harmonogram workshopu

I. Projekt EduTest: od výzkumu k praxi (9:00–9:40)

- Představení projektu, přehled výstupů projektu

Přestávka (9:40–10:00)

II. Od tradičních metod po modelování odpovědí na položky (10:00–11:20)

- Klasická testová teorie, tradiční položková analýza
- Modely teorie odpovědi na položku (IRT), odlišné fungování položek (DIF)

Přestávka (11:20–11:40)

III. Další témata (11:40–12:35)

- Vyrovnávání více verzí testů
- Textová analýza
- Adaptivní testování

Diskuse a závěr (12:35–13:00)

Elektronické materiály:

cs.cas.cz/comps/workshop

I. Projekt EduTest

Od výzkumu k praxi



Projekt EduTest

*Výzvy pro hodnocení znalostí:
Analytická podpora tvorby znalostních testů*

- Financováno Technologickou agenturou ČR
- Řešitelé:
 - Pedagogická fakulta Univerzity Karlovy
(Ústav výzkumu a rozvoje vzdělávání,
Centrum tvorby didaktických testů a psychometrického modelování, CEMP)
 - Ústav informatiky AV ČR
(Oddělení statistického modelování,
Skupina výpočetní psychometrie, COMPS)
- Aplikační garant:
 - Centrum pro zjišťování výsledků vzdělávání (CZVV)
- Doba řešení: 4/2021 – 12/2023



[EduTest.cz](https://edutest.cz)

Projekt EduTest – Cíle

- Vytvořit soubor postupů a nástrojů pro komplexní vyhodnocování výsledků znalostních testů
- Podpořit využívání navržených postupů a nástrojů při vyhodnocování stávajících přijímacích, maturitních a jiných znalostních testů a při vývoji nových testů
- Podpořit informované rozhodování na úrovni škol, krajů a státu

Projekt EduTest – Plánované výstupy

- Soubor postupů a nástrojů pro zkvalitnění tvorby znalostních testů pomocí psychometrických modelů
 - textová část, softwarová část
- Participativní workshopy
 - pro tvůrce testů (aplikační garant, duben 2022)
 - pro uživatele testů (školy apod., prosinec 2022)
 - **pro odbornou veřejnost (2023)**

Aplikace výzkumu psychometrických metod

- EduTest má za cíl posunout náš teoretický výzkum do praxe
- Náš výzkum v oblasti psychometrie
 - Teoretické články v oblasti reliability, detekce odlišného fungování položek, aj.
 - Vývoj algoritmů v oblasti psychometrie
 - Kniha
- Naše výuka v oblasti psychometrie
 - „Vybraná témata z psychometrie“, „Statistické metody v psychometrii“
 - Probíhá na UK
 - [Seminář z psychometrie](#)
 - Počítačové adaptivní testování (19.4.2022)
 - Test equating (semináře v roce 2021)
 - Detekce podvádění, aj.

Výzkum psychometrických metod, na který navazujeme

- Modelování fungování položek

- Martinková, Drabinová et al. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education*, 16(2), rm2. <https://doi.org/10.1187/cbe.16-10-0307>
- Drabinová & Martinková (2017). Detection of Differential Item Functioning with Nonlinear Regression: A Non-IRT Approach Accounting for Guessing. *Journal of Educational Measurement*, 54(4), 498-517. <https://doi.org/10.1111/jedm.12158>
- Martinková, Hladká, & Potužníková (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, 66, 101286. <https://doi.org/10.1016/j.learninstruc.2019.101286>
- Hladká, Martinková, & Magis (2023). Combining item purification and multiple comparison adjustment methods in detection of differential item functioning. *Multivariate Behavioral Research*. <https://doi.org/10.1080/00273171.2023.2205393>

- Reliabilita

- Martinková, Bartoš, & Brabec M (2023). Assessing inter-rater reliability with heterogeneous variance components models: Flexible approach accounting for contextual variables. *Journal of Educational and Behavioral Statistics*, 48(3), 349–383. <https://doi.org/10.3102/10769986221150517>
- Goldhaber, Grout, Wolff, & Martinková (2021). Evidence on the dimensionality and reliability of professional references' ratings of teacher applicants. *Economics of Education Review*, 83, 102130. <https://doi.org/10.1016/j.econedurev.2021.102130>
- Erosheva, Martinková, & Lee (2021). When zero may not be zero: A cautionary note on the use of inter-rater reliability in evaluating grant peer review. *Journal of the Royal Statistical Society – Series A*, 184(3), 904-919. <https://doi.org/10.1111/rssa.12681>

Výzkum psychometrických metod, na který navazujeme

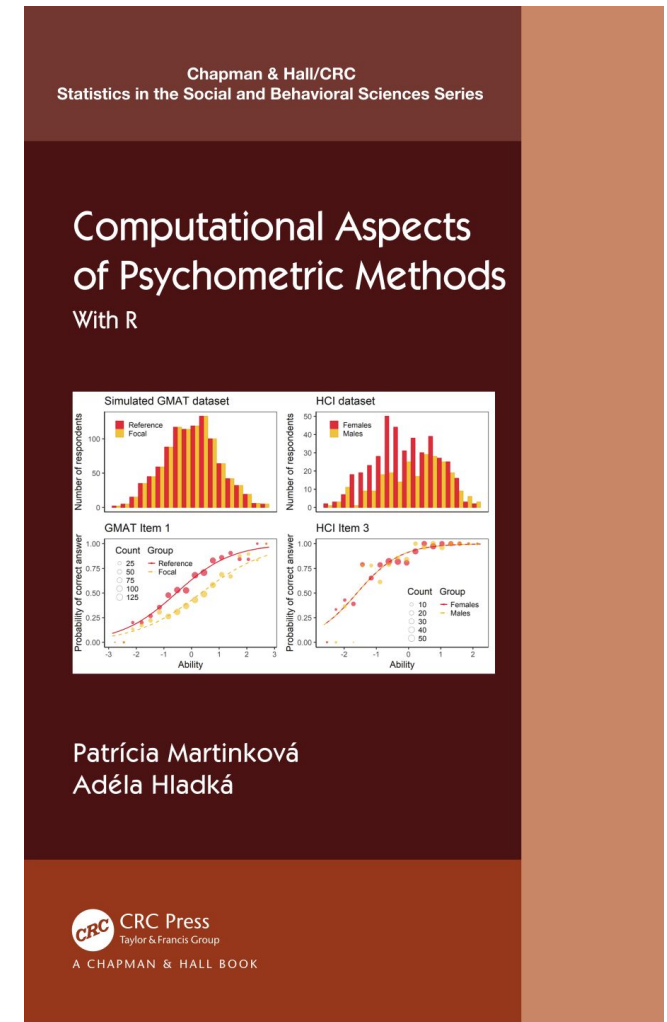
- Machine learning a textová analýza pro odhad obtížnosti položek
 - Štěpánek, Dlouhá, & Martinková (2023). Item difficulty prediction using item text features: Comparison of predictive performance across machine-learning algorithms. *Mathematics*, 11(19), 4104. <https://doi.org/10.3390/math11194104>
- Počítačové adaptivní testování
 - Štěpánek & Martinková (2020). Feasibility of computerized adaptive testing evaluated by Monte-Carlo and post-hoc simulations. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)* (pp. 359-367). IEEE. <https://doi.org/10.15439/2020F197>
- Implementace v rámci statistického software
 - Martinková & Hladká (2018). ShinyItemAnalysis for Teaching Psychometrics and to Enforce Routine Analysis of Educational Tests. *R Journal*, 10(2). <https://doi.org/10.32614/RJ-2018-074>
 - Hladká & Martinková (2020). difNLR: Generalized Logistic Regression Models for DIF and DDF Detection. *R Journal*, 12(1). <https://doi.org/10.32614/RJ-2020-014>
 - Martinková P, Drabinová A, Houdek J (2017). ShinyItemAnalysis: Analýza přijímacích a jiných znalostních či psychologických testů. *Testforum*, 9, 16-35. <https://doi.org/10.5817/TF2017-9-129>

Kniha

- Ilustrace důležitých výpočetních aspektů
- Vývoj IRT modelů krok za krokem
- Implementace metod v různých R balících a v ShinyItemAnalysis
- Příklady na reálných i simulovaných datech
- Reprodukovatelný R kód pro analýzy
- Příklady pro výuku

Obsah:

1. Introduction
2. Validity
3. Internal structure of the test and factor analysis
4. Reliability
5. Traditional item analysis
6. Item functioning with regression models
7. Item response theory models
8. More complex IRT models
9. Differential item functioning
10. Outlook on applications and more advanced psychometric topics



Martinková & Hladká (2023). Computational Aspects of Psychometric Methods: With R (1st ed.). Chapman and Hall/CRC.
<https://doi.org/10.1201/9781003054313>

IMPS 2024

Prague, Czech Republic

July 16-19, 2024 • Short Courses July 15



ShinyItemAnalysis

- Námi vyvíjená interaktivní aplikace
- Využití ve výuce i pro prezentaci výzkumu
- Možnost nahrání a analýzy vlastních dat
- Využívaná po celém světě
- Možnost přístupu offline i online



[Data](#)[Basic summary](#)[Data exploration](#)

Data

For demonstration purposes, the 20-item dataset **GMAT** is used. While on this page, you may select one of several other toy datasets or you may upload your own dataset (see below). To return to the demonstration dataset, click on the **Unload data** button.

Training datasets

Select dataset

GMAT ([Martinkova et al., 2017](#)) is a generated dataset based on the parameters of a real Graduate Management Admission Test (GMAT; Kingston et al., 1985) from the **difNLR** package. This dataset represents the responses of 2,000 subjects (1,000 males coded as '0', 1,000 females coded as '1') to a multiple-choice test of 20 items. It also contains generated continuous criterion intended to be predicted by the test. As discussed in [Martinkova et al. \(2017\)](#), the dataset was simulated in order to provide an example of DIF items present even in the case of identical total score distributions. To replicate the example provided in [Martinkova et al. \(2017\)](#), select the **GMAT** dataset and go to **DIF/Fairness** section.

Upload your own datasets

Here you can upload your own dataset. Select all necessary files and use the **Upload data** button on bottom of this page. For sample .csv data and details on input format, check the Supplementary material of the [Martinkova and Drabinova \(2018\)](#) article.

Choose data (CSV file)

The main **data** file should contain the responses of individual respondents (rows) to given items (columns). Data need to be either binary, nominal (e.g. in ABCD format), or ordinal (e.g. in Likert scale). The header may contain item names, however, no row names should be included. In all data sets, the **header** should be either included or excluded. If you want to rename items



SIA moduly

- Zásuvné moduly pro ShinyItemAnalysis (SIA) napsané v {shiny}
- Oboustranná komunikace s hlavní aplikací
- Možnost využít existující infrastrukturu SIA
- Např. možnost nahrání volně dostupných dat CZVV přímo v Excelu bez potřeby data jakkoliv editovat



Moduly a aplikace projektu EduTest

1. SIA modul „CERMAT tools“

- načtení volně dostupných položkových dat (maturita, JPZ)
- jsou detekovány odpovědi na položky, skupinové proměnné a kovariáty
- vyžaduje načtení metadat: soubor s popisem položek
- možnost poslat data do hlavní ShinyItemAnalysis a využít její funkcionality
shiny.cs.cas.cz/ShinyItemAnalysisCermat
- specifická položková analýza
- IRT model specifický pro různé typy položek

[Data](#)[CTT](#)[IRT](#)

Data

Item data

Browse...

mz_2022_spring_ma.x

Upload complete

Options

 Include only the cases with "termín = řádný", "prvomaturant", and "zkouska_p Display preview of the data

Item metadata

Browse...

mz_2022_spring_ma_l

Upload complete

Use data in the main app

Here you can pass your data you have just loaded in this module to the main SIA app. You can chose which column of the uploaded dataset will be considered a group or criterion variable.

Group

none ▾

Criterion

none ▾

[↔ Pass data to SIA](#)

Description

About the modules

Modules

CERMAT tools

CEFR level analysis

Computerized Adaptive Tests

Moduly a aplikace projektu EduTest

1. SIA modul „CERMAT tools“

- načtení volně dostupných položkových dat, specifická položková analýza
- možnost poslat data do hlavní ShinyItemAnalysis a využít její funkcionality
shiny.cs.cas.cz/ShinyItemAnalysisCermat

2. Aplikace pro střední školy – maturitní zkoušky

- využití dat školy (speciální export pro školy od CZVV)
- zpětná vazba pro SŠ, lze využít i v rámci přípravy na maturitu
edutest.shinyapps.io/maturita_data (pouze ukázková data)

Nahrání dat

Vložte soubor s výsledky pro Vaši školu. Soubor můžete do pole i přetáhnout.

Procházet soubory...

nevybrán žádný soubor

Volba zkoušky

Zvolte zkoušku, s níž chcete pracovat.

Český jazyk a literatura

Nabídka zkoušek vyskytujících se v nahraných datech

Aktuální data

Popis nahraných dat

Celek

zkouška	žáků	skupina mat. oborů	žáků
Matematika	54 (34 %)	Lyceum	64 (20 %)
Anglický jazyk	115 (72 %)	SOŠ technické	261 (80 %)
Český jazyk a literatura	156 (98 %)		

Vybraná zkouška

skupina mat. oborů	žáků	celkový výsledek	žáků	druh termínu	žáků
Lyceum	31 (20 %)	neuspěl(a)	2 (1 %)	řádný	156 (100 %)
SOŠ technické	125 (80 %)	-	6 (4 %)		
		uspěl(a)	148 (95 %)		
prvomaturant	žáků	povinná zkouška	žáků	konali zkoušku	žáků
ne	1 (1 %)	ano	156 (100 %)	ne	6 (4 %)
ano	155 (99 %)			ano	150 (96 %)

Výběr srovnávané a srovnávací skupiny

Srovnávané obory na
Vaší škole:

Lyceum
SOŠ technické

Dostupné obory na Vaší škole

Srovnávací skupina:

Lyceum
SOŠ technické
Gymnázium 8leté
Gymnázium 6leté
Gymnázium 4leté
SOŠ technologické
SOŠ ekonomické
SOŠ hotelové a podni

Lze vybrat i jiné obory,
než jsou obory Vaší školy

Specifikace srovnávací skupiny

Ve výchozím stavu se srovnávají všichni žáci všech oborů Vaší školy s relevantními žáky všech škol, které se daného testu zúčastnili. Porovnávají se tedy žáci stejných oborů. Složení srovnávané a srovnávací skupiny ale můžete změnit výběrem požadovaných oborů myší.

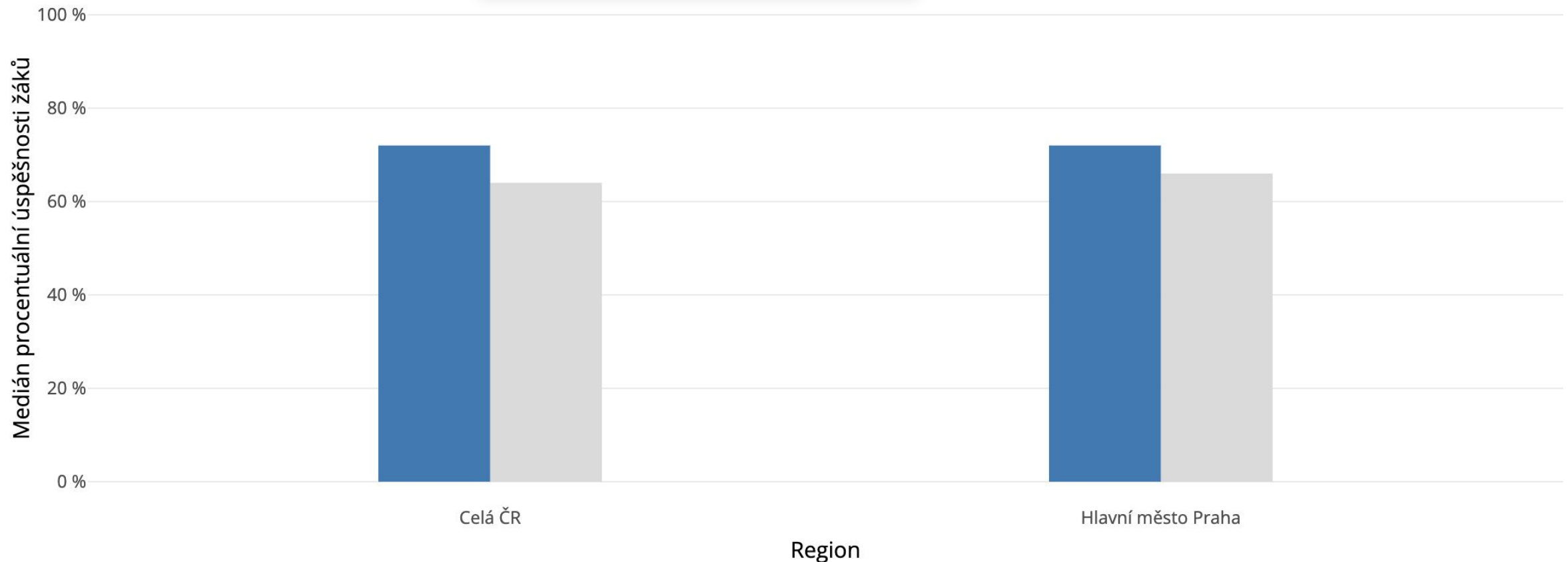
Pro výběr souvislé řady klikněte na její počátek a za držení klávesy **Shift** klikněte na konec, nebo klikněte na začátek a táhněte myší až k požadovanému oboru. Držením **Ctrl / ⌘** a klikáním myší vyberete více různě rozmístěných oborů. Pro zrušení výběru konkrétního již vybraného oboru podržte klávesu **Ctrl / ⌘** a klikněte na něj.

Celkový výsledek Vaší školy

Modrý sloupec ukazuje mediánovou procentuální úspěšnost žáků **Vaší školy** ve srovnání s mediánovou procentuální úspěšností všech žáků ze zvolené **srovnávací skupiny** (šedý sloupec), a to jak v rámci celé ČR, tak v rámci kraje, v němž se Vaše škola nachází. V rámci ČR byli žáci Vaší školy v didaktickém testu z českého jazyka a literatury o **8,0** procentních bodů **úspěšnější** než žáci ze srovnávací skupiny. V regionu jste **úspěšnější** o **6,0** procentních bodů.

Legenda přímo v textu

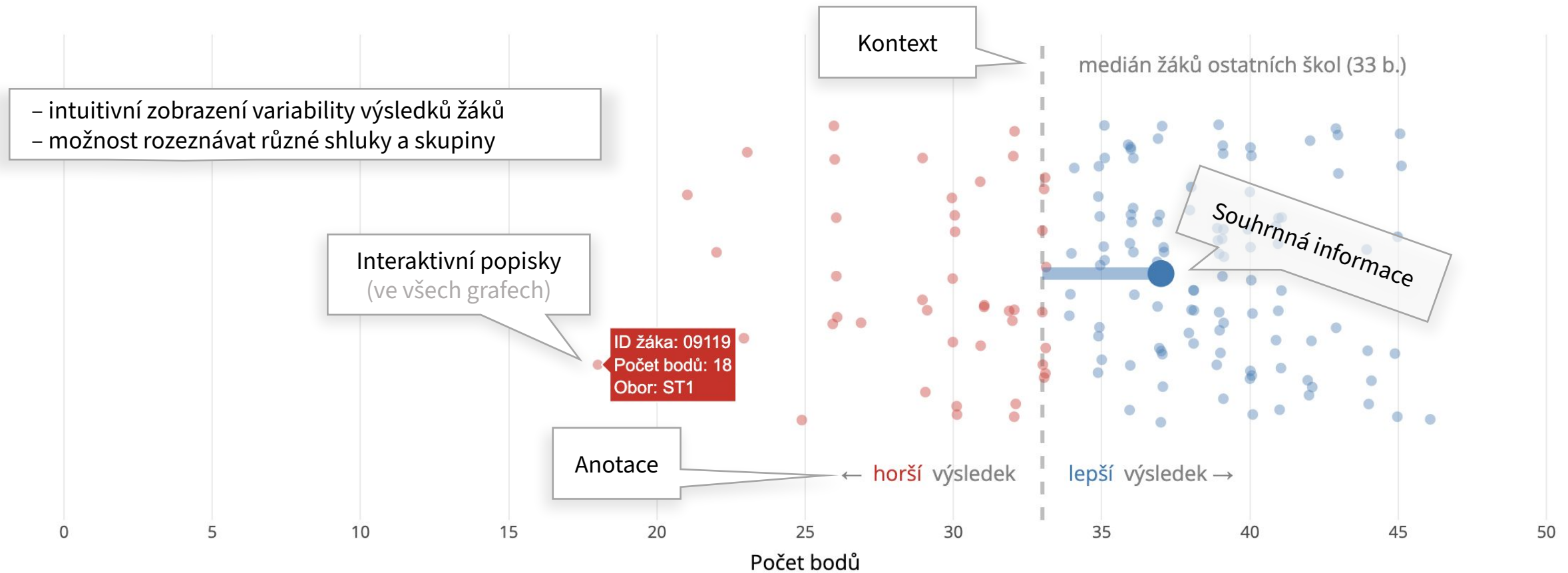
Automatizované interpretace, detaily



Výsledky Vašich žáků

V následující vizualizaci ve větším detailu ukazujeme výsledky jednotlivých žáků Vaší školy (malé průhledné body v grafu), které dáváme do kontextu s celkovými výsledky – mediánem – všech zapojených škol daného oboru (šedá svíslá čárkovaná čára). Pro snazší interpretaci prezentujeme žáky **s nižším než mediánovým výsledkem všech škol** červeně a žáky, **kteří dosáhli lepšího než mediánového výsledku** modře. Silná vodorovná čára zakončená větším bodem shrnuje pozici Vaší školy a ukazuje medián bodového zisku Vašich žáků. Lze říci, že žáci Vaší školy mají v testu celkově **o 4 body více** než ostatní.

„Kódování“ informace barvou



– intuitivní zobrazení variability výsledků žáků
– možnost rozeznávat různé shluky a skupiny

Interaktivní popisky
(ve všech grafech)

ID žáka: 09119
Počet bodů: 18
Obor: ST1

Anotace

Kontext

medián žáků ostatních škol (33 b.)

Souhrnná informace

← horší výsledek

lepší výsledek →

Počet bodů

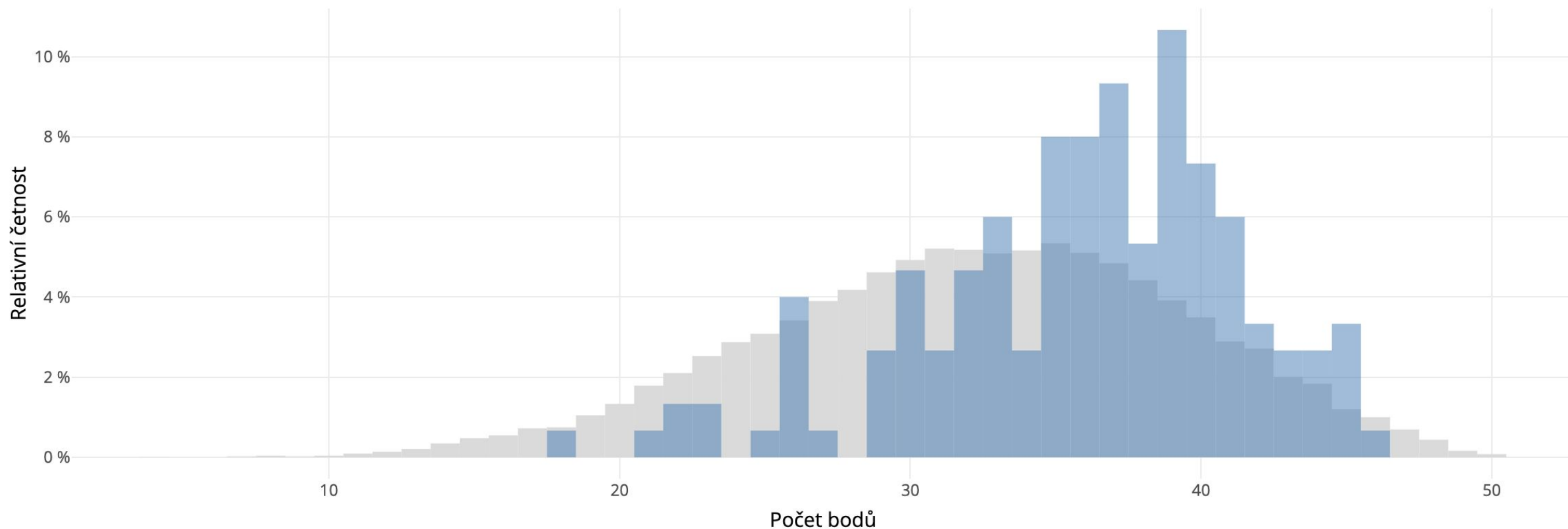
Rozdělení bodových výsledků

Jak často žáci ve zkoušce dosahovali určitého počtu bodů? Nabízíme vždy porovnání s výsledky žáků všech škol, kteří danou zkoušku konali. **Modré sloupce** znázorňují žáky Vaší školy, **šedé** všech škol dohromady (tedy i včetně Vás). Aby byla data porovnatelná, ukazujeme relativní četnosti, tj. procenta žáků, kteří dosáhli daného bodového výsledku v rámci Vaší školy, resp. všech škol. Nabízíme tři různé způsoby vizualizace stejných dat: histogramy pro obě skupiny přes sebe, histogramy s jednotlivými sloupci střídatě pro Vaši školu a celek a zrcadlové zobrazení, kde na levé straně je výsledek všech škol, na pravé ten Váš.

Přes sebe

Vedle sebe

Zrcadlově

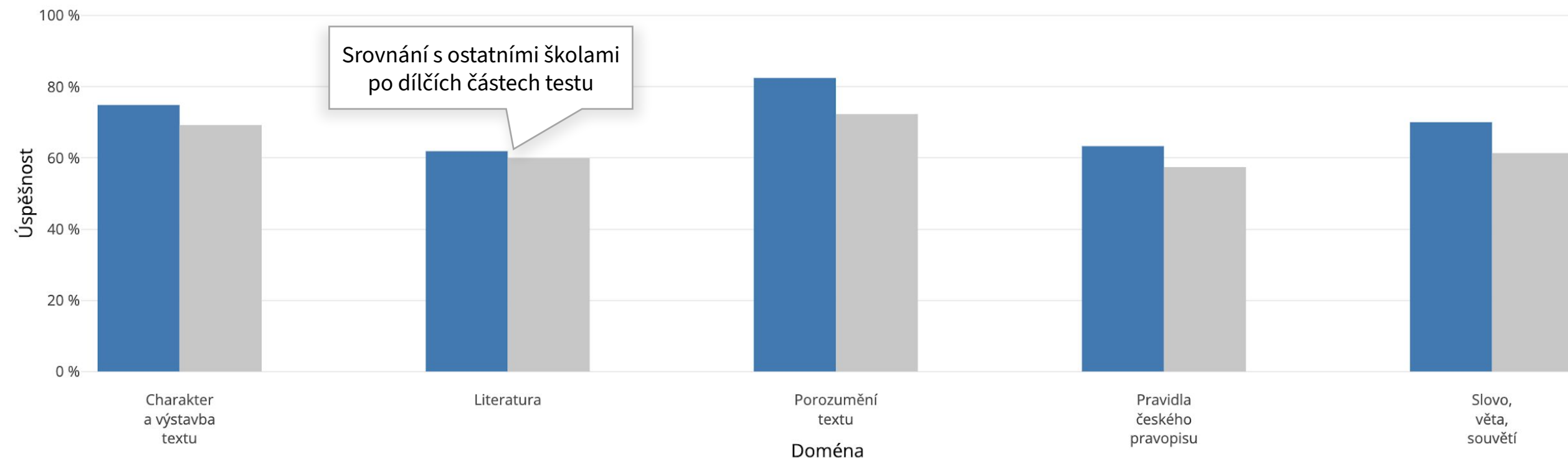


Úspěšnost v doménách

[Detailní analýza rozdílů](#)

Úspěšnost v doménách (oblastech) testu

Níže ukazujeme úspěšnost v jednotlivých částech testu. Úspěšnost je definovaná jako průměrné skóre v rámci dané skupiny přeškálované na interval (0–100 %), kde vyšší hodnota značí vyšší úspěšnost (lehčí doménu). Doporučujeme se podívat i na záložku „Detailní analýza rozdílů“, kde se srovnáváme žáky se stejným celkovým počtem bodů – přestože máte např. celkově nižší úspěšnost, v některých oblastech mohou mít Vaši žáci v porovnání se stejně skórujícími žáky ostatních škol paradoxně lepší výsledky.



Úspěšnost v úlohách testu

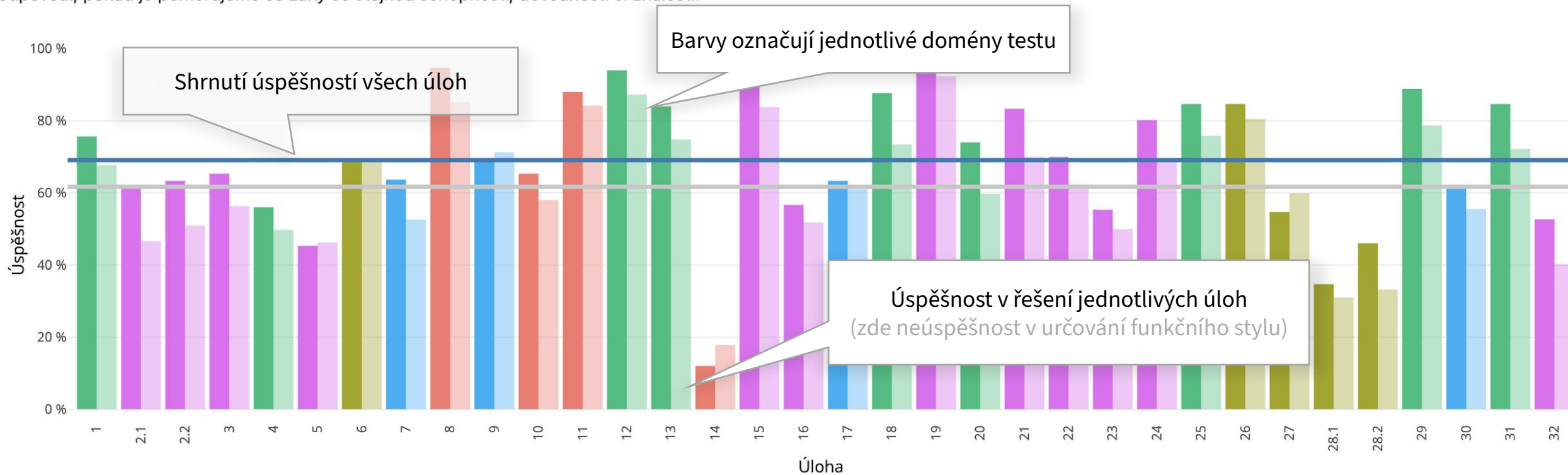
[Detailní analýza rozdílů](#)

Úspěšnost v úlohách testu

Graf

Graf zobrazuje úspěšnost žáků v jednotlivých testových úlohách, a to zvláště pro Vaši školu (syté sloupce) a žáky vybrané srovnávací skupiny. Úspěšnost je definována jako průměrný bodový skóre v poměru k maximálnímu počtu bodů, který bylo možné v dané úloze získat. V úlohách hodnocených maximálně 1 bodem tak úspěšnost odpovídá podílu žáků, kteří úlohu zodpověděli správně.

Pro detailnější pohled na rozdíly velmi doporučujeme navštívit záložku „Detailní analýza rozdílů“, která porovnává pravděpodobnost zisku určitého počtu bodů v jednotlivých úlohách u žáků se *stejným bodovým ziskem z celého testu*. Přestože máte např. nižší celkovou úspěšnost napříč úlohami, v některých úlohách mohou mít Vaši žáci paradoxně vyšší pravděpodobnost správné odpovědi, pokud je poměrujeme se žáky se stejnou schopností, dovedností či znalostí.



Tabulka

Tabulky v záložkách níže zobrazují parametry jednotlivých položek pro Vaši školu a celek. Kromě úspěšnosti je uvedeno také průměrné skóre v položce; to je totožné s úspěšností v případě dichotomních položek (položek hodnocených 0 – nesprávná nebo 1 – správná odpověď). Dále je uvedena směrodatná odchylka (standard deviation; SD), minimální a maximální dosažené skóre. Následuje podíl respondentů, kteří dosáhli maximálního možného skóre, tento podíl je opět totožný s odhadem úspěšnosti v případě dichotomních položek.

Vaše škola

[Všechny školy](#)

 Stejně informace podrobněji
v podobě tabulky

Možnost řazení, označování atp.

	Popis	Doména	Úspěšnost	Průměr	SD	Min. bodů	Max. bodů	Podíl max. bodů
14	charakter textu: určit funkční styl	Charakter a výstavba textu	0,12	0,12	0,33	0	1	0,12
28.1	literární historie: identifikovat literární směr, který je charakterizován ve výchozím textu (naturalismus)	Literatura	0,35	0,35	0,48	0	1	0,35
5	syntax: rozlišit větu jednoduchou x souvětí, identifikovat počet vět v souvětí	Slovo, věta, souvětí	0,45	0,45	0,50	0	1	0,45
28.2	literární historie: identifikovat spisovatele, o němž se píše ve výchozím textu (Zola)	Literatura	0,46	0,46	0,50	0	1	0,46
32	význam slov: posoudit, v kterých kontextech jsou určité dvojice slov antonymy (chuť x nechut; dostatek x nedostatek)	Slovo, věta, souvětí	0,53	0,53	0,50	0	1	0,53
27	literární teorie: identifikovat literární žánr (sonet)	Literatura	0,55	0,55	0,50	0	1	0,55
23	syntax: identifikovat větné členy (podmět, předmět)	Slovo, věta, souvětí	0,55	0,55	0,50	0	1	0,55
4	porozumění textu (práce s definicí – vokalizovaná x nevokalizovaná podoba předložek): najít v úsecích předložku, která může mít vokalizovanou podobu (pod)	Porozumění textu	0,56	0,56	0,50	0	1	0,56
16	syntax: identifikace příslovečných určení	Slovo, věta, souvětí	0,57	0,57	0,50	0	1	0,57
2.1	ovlivněné	Slovo, věta, souvětí	0,61	0,61	0,49	0	1	0,61

Report

Pro nahraná data a zvolenou zkoušku si můžete vygenerovat a následně vytisknout report, který nabízíme ve formátu PDF.

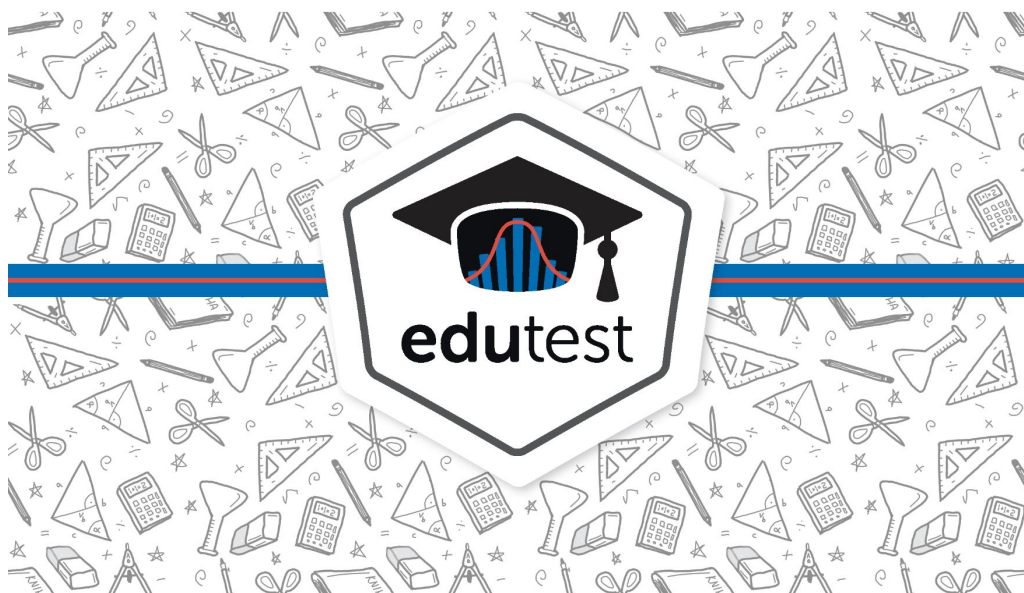
V současnosti používá report pro srovnání Vaší školy s ostatními hodnoty vybrané v předchozích záložkách. Pokud jste s výběrem skupin v aplikaci nemanipulovali, report použije výchozí hodnoty, tj. srovnávají se žáci všech oborů Vaší školy s žáky ostatních škol odpovídajících oborů.

Přizpůsobení

Sestrojit grafy jen pro různě fugující položky

Uložte si a vytiskněte to, co jste viděli v aplikaci

 Generovat a stáhnout



Český jazyk a literatura – jaro 2022

Výsledky z didaktického testu v rámci maturitní zkoušky

*Střední škola Ukázková, Testová 600/1, Testov 1,
příspěvková organizace (REDIZO: da3296532...)*

Vygenerováno 6. prosince 2022

© 2022

Skupina výpočetní psychometrie,
Ústav informatiky, Akademie věd ČR

Název smyšlené školy pro ukázkou

Obsah

Úvod	3
Celkové výsledky	4
Srovnání úspěšnosti Vaší školy a ostatních	4
Výsledky žáků Vaší školy v kontextu mediánu celku	5
Rozdělení dosažených bodů v testu	6
Výsledky v dílčích částech testu	7
Výsledky v testových úlohách	8
Detailní analýza rozdílů	10

Moduly a aplikace projektu EduTest

1. SIA modul „CERMAT tools“

- načtení volně dostupných položkových dat, specifická položková analýza
- možnost poslat data do hlavní ShinyItemAnalysis a využít její funkcionality shiny.cs.cas.cz/ShinyItemAnalysisCermat

2. Aplikace pro střední školy - maturitní zkoušky

- využití dat školy (speciální export CZVV)
- zpětná vazba pro SŠ, lze využít i v rámci přípravy na maturitu edutest.shinyapps.io/maturita_data (pouze ukázková data)

3. Aplikace pro základní školy – jednotné přijímací zkoušky

- přímé nahrání tzv. školní zprávy (Excel běžně zasílaný CZVV do škol)
- zpětná vazba pro ZŠ, lze využít i v rámci přípravy na JPZ edutest.shinyapps.io/prijimacky/

Náhled a možnost interaktivního procházení (řazení atp.) rozpoznaných položkových dat

Nahrání dat

Vložte soubor s výsledky pro Vaši školu, který jste obdrželi od CERMATu. Soubor můžete do pole i přetáhnout.

Procházet soubory... JPZ_skolni_zprava_ZS_2020

Upload complete

Nahrání „Excel“ souboru od CZVV

Volba zkoušky

Zvolte zkoušku, s níž chcete pracovat.

MA

MA

ČJL

Aut. rozpoznání dostupných položkových dat

Náhled dat

žák ↕	celkem bodů ↕	u1 ↕	u2.1 ↕	u2.2 ↕	u3.1 ↕	u3.2 ↕	u4.1 ↕	u4.2 ↕	u4.3 ↕	u5.1 ↕	u5.2 ↕
1	50	1	1	1	2	2	1	1	2	2	
2	50	1	1	1	2	2	1	1	2	2	
3	50	1	1	1	2	2	1	1	2	2	
4	49	1	1	1	2	2	1	1	1	2	
5	49	1	1	0	2	2	1	1	2	2	
6	48	1	1	1	2	2	1	1	2	2	
7	48	1	1	1	2	2	1	1	2	2	
8	48	1	1	1	2	2	1	1	2	2	
9	48	1	1	0	2	1	1	1	2	2	
10	48	1	1	1	2	2	1	1	2	2	

Zobrazují 1 až 10 z celkem 91 záznamů

Praktická část

Praktická část: výstupy projektu EduTest

? Která prezentovaná aplikace nebo modul jsou z vaší perspektivy nejvíce zajímavé, jak byste je využili?

Zatím představené interaktivní aplikace a moduly:

0. [ShinyItemAnalysis](#)
1. SIA modul [CERMAT tools](#) - pro tvůrce testů
2. [Aplikace pro střední školy](#) s cvičnými daty z maturity
3. [Aplikace pro základní školy](#) (JPZ)

Online materiály a odkazy: cs.cas.cz/comps/workshop

II. Od tradičních metod k modelování odpovědí na položky

CTT, IRT, DIF



Obsah

Od tradiční položkové analýzy po modelování odpovědí na položky

- Tradiční položková analýza
- Regresní modely pro popis položek
- Modely teorie odpovědi na položku (*Item Response Theory, IRT*)
- Odlišné fungování položek (*Differential Item Functioning, DIF*)
- Praktická část

POLOŽKOVÁ ANALÝZA UKÁZKA VÝSTUPŮ

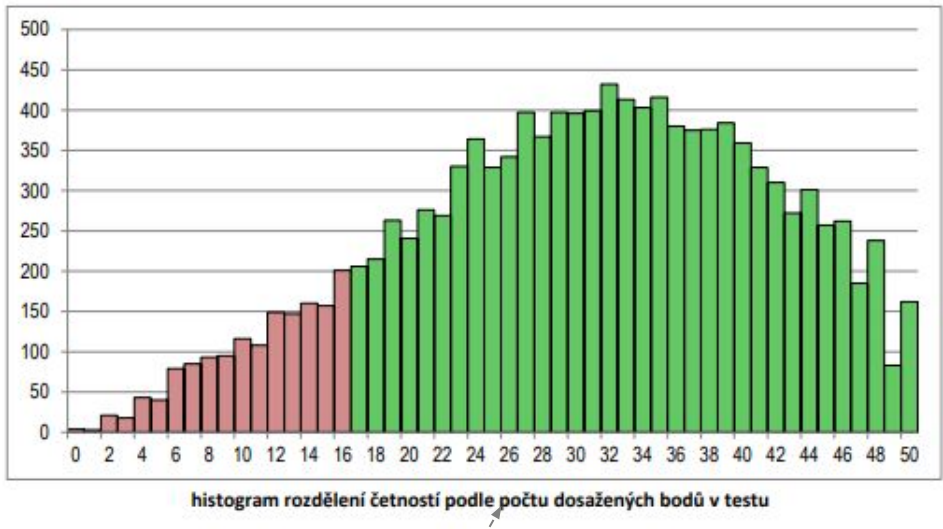
PROJEKT	MZ_2023j
PŘEDMĚT	matematika

DVANÁCTIMÍSTNÝ KÓD DIDAKTICKÉHO TESTU	MAMZD23C0T01
--	---------------------

ZÁKLADNÍ ÚDAJE O TESTU	počet položek	30
	počet elementů záznamového archu	28
	dosažitelný počet bodů	50
	cut-off skóre	17

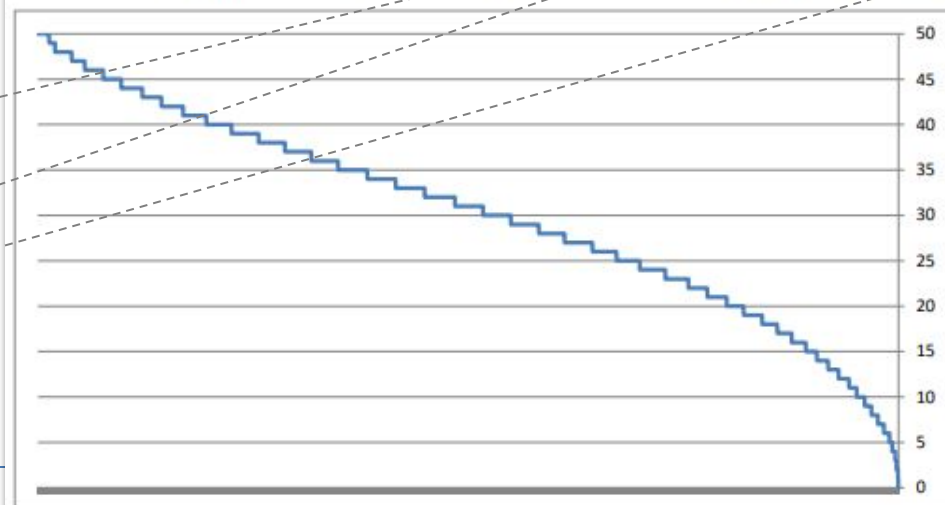
počet žáků	12247
------------	-------

program Propan ©JHu pro Cermat 2019-23



ZÁKLADNÍ INFORMACE O TESTU CELKEM

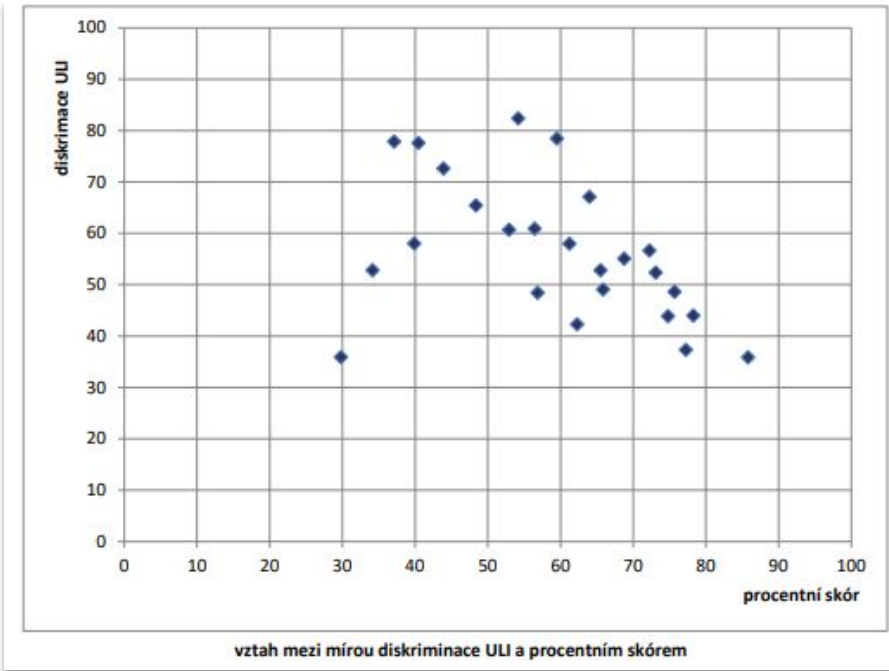
STATISTICKÉ CHARAKTERISTIKY	bodový skóre	procentní skóre	TESTOLOGICKÉ CHARAKTERISTIKY	
průměr	30,2	60,5	podíl neúspěšných (%)	12,4
směrodatná odchylka	10,9	21,8	variační koeficient (%)	36,0
decilové rozpětí	29	58,0	diskriminace ULI (%)	56,1
kvartilové rozpětí	16	32,0	Cronbachovo alfa	0,86



bodový skóre v závislosti na pořadí žáků od nejlepšího k nejhoršímu

- Četnost dosažených bodů
- Průměrný skóre
- Míry variability
- Podíl neúspěšných
- Reliabilita testu

PŘEHLED ÚLOH PRŮMĚRNÝ SKÓR VS. DISKRIMINACE



- Celkový pohled na jednotlivé úlohy v kombinaci % skór vs. diskriminace

	0:10	10:20	20:30	30:40	40:50	50:60	60:70	70:80	80:90	90:100
90:100										
80:90										
70:80				9 10	14	7				
60:70				8	21	3				
50:60				1	12	19 22	13 23			
40:50			24			15	6 17 20	18		
30:40			16			11	2	4 25	5	
20:30										
10:20										
0:10										

vztah mezi ULI a procentním skórem - uvedeno číslo položky

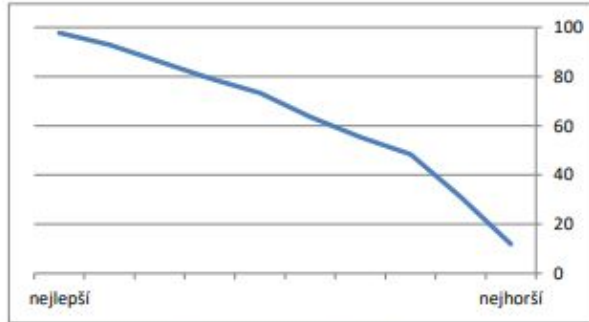
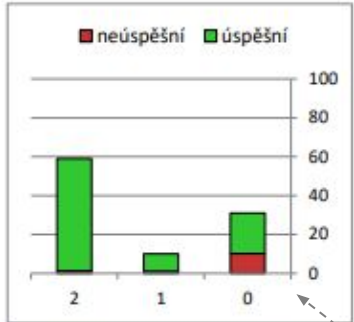
ÚDAJE O TESTOVÝCH POLOŽKÁCH

OTEVŘENÁ ÚLOHA

Položka 3

průměrný % skór: 64,0
 prům.% skór neúspěšných: 14,5
 diskriminace ULI (%): 67,1

úzce otevřená úloha
 počet vynechavších: 1134
 vynechanost (%): 9,3
 korelace RIR: 0,52



body	žáci celkem			neúspěšní		nejhorší %		nejlepší %		rozdíl
	počet	%	úsp.	počet	%	počet	%	počet	%	
0	3797	31,0	42,6	1236	81,4	2086	68,1	112	3,7	-64,5
1	1230	10,0	56,7	126	8,3	320	10,5	157	5,1	-5,3
2	7220	59,0	70,5	157	10,3	656	21,4	2793	91,2	69,8

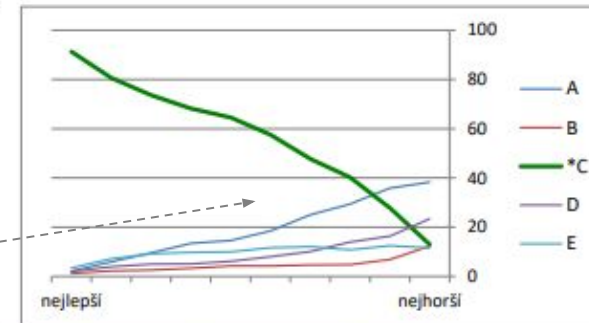
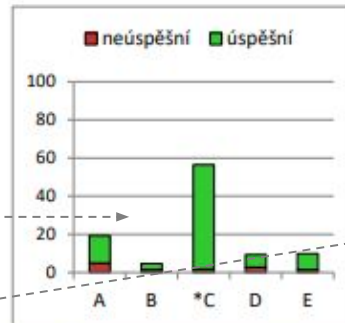
- Průměrný skór
- Diskriminace ULI
- Korelace RIR
- Vynechanost úlohy

UZAVŘENÁ ÚLOHA

Položka 19

průměrný % skór: 56,4
 prům.% skór neúspěšných: 14,3
 diskriminace ULI (%): 60,9

uzavřená úloha (multichoice)
 počet vynechavších: 50
 vynechanost (%): 0,4
 korelace RIR: 0,40



body	žáci celkem			neúspěšní		nejhorší %		nejlepší %		rozdíl
	počet	%	úsp.	počet	%	počet	%	počet	%	
A	2359	19,3	47,2	593	39,0	1121	36,6	143	4,7	-31,9
B	573	4,7	47,0	175	11,5	262	8,6	55	1,8	-6,8
*C	6912	56,4	69,5	217	14,3	714	23,3	2579	84,2	60,9
D	1151	9,4	46,2	339	22,3	579	18,9	99	3,2	-15,7
E	1202	9,8	55,6	176	11,6	357	11,7	184	6,0	-5,6

- Četnost dosažených bodů, resp. četnost volby alternativ
- Průměrný % skór, resp. míra volby alternativ podle výsledku v testu

[Data](#)[Scores ▾](#)[Validity ▾](#)[Reliability ▾](#)[Item analysis ▾](#)[Regression ▾](#)[IRT models ▾](#)[DIF/Fairness ▾](#)[Modules ▾](#)[Reports](#)[Data](#)[CTT](#)[IRT](#)

Data

Item data

Browse...

mz_2022_spring_ma.x

Upload complete

Options

 Include only the cases with "termín = řádný", "prvomaturant", and "zkouska_p Display preview of the data

Item metadata

Browse...

mz_2022_spring_ma_l

Upload complete

Use data in the main app

Here you can pass your data you have just loaded in this module to the main SIA app. You can chose which column of the uploaded dataset will be considered a group or criterion variable.

Group

none ▾

Criterion

none ▾

[↔ Pass data to SIA](#)

Description

About the modules

Modules

CERMAT tools

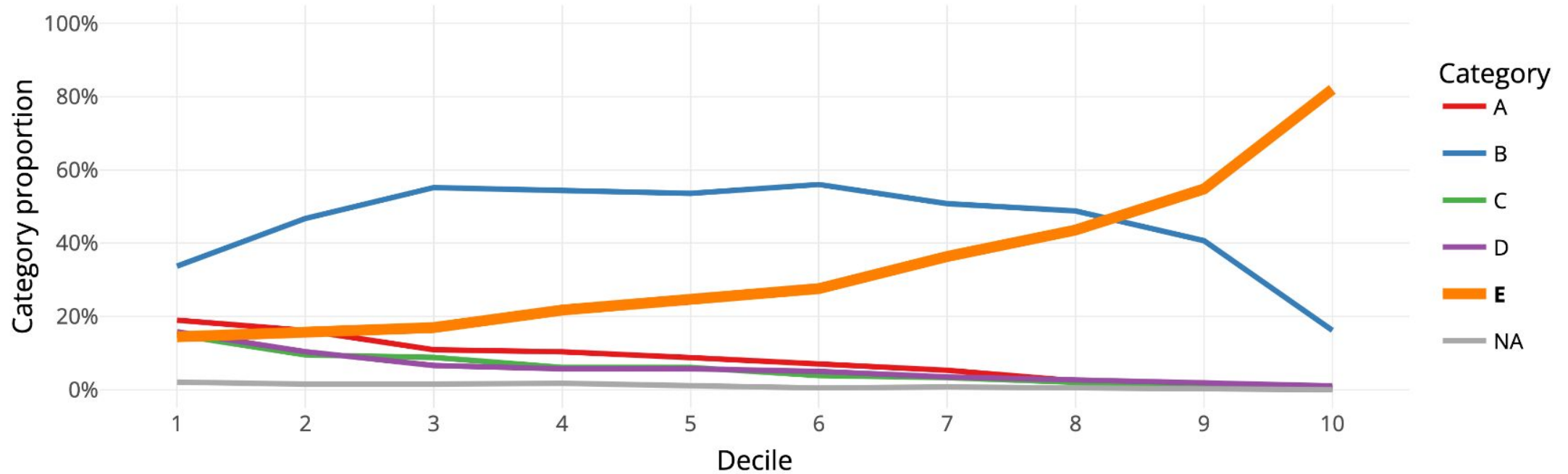
CEFR level analysis

Computerized Adaptive Tests

[Data](#)[Scores](#) ▾[Validity](#) ▾[Reliability](#) ▾[Item analysis](#) ▾[Regression](#) ▾[IRT models](#) ▾[DIF/Fairness](#) ▾[Modules](#) ▾[Reports](#)

Difficulty

Empirical item response curves

Item[Previous item](#)[Next item](#) Ignore NAs**Number of groups**

Item statistics

Difficulty	Mean	SD	RIR	ULI	gULI	Alpha.drop
0.35	0.71	0.96	0.32	0.44	0.49	0.87

Discrimination is by default estimated as the coRelation between Item and Total score (RIT index). Other options for the discrimination index include coRelation between Item and total score based on Rest of the items (RIR index). Discrimination can also be estimated as the difference in (scaled) item score in the upper and lower third of the respondents (Upper-Lower Index, ULI). ULI can be further customized by changing the number of groups and by changing which groups should be compared (see also Martinkova, Stepanek et al., 2017). By a rule of thumb, all items with a discrimination lower than 0.2 (threshold in the plot), should be checked for content. Lower discrimination is expectable in the case of very easy or very difficult items, or in ULI based on more homogeneous groups (such as 4th and last fifth). A threshold may be adjusted for these cases or may be set to 0.

Difficulty type:

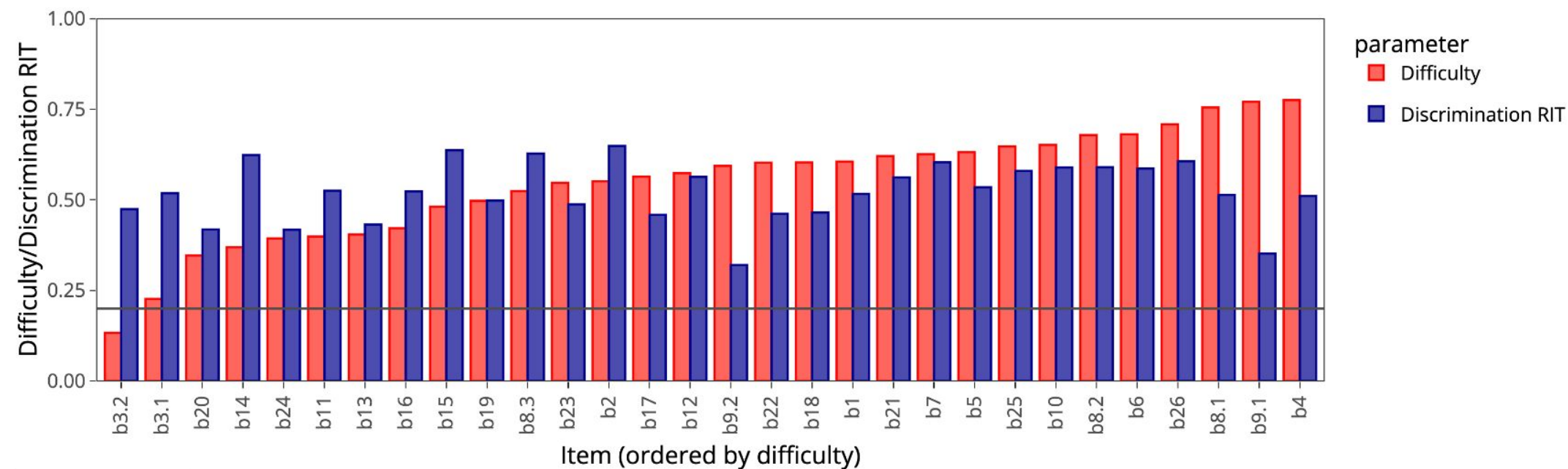
Average scaled score ▾

Discrimination type:

RIT ▾

Show threshold

Threshold:

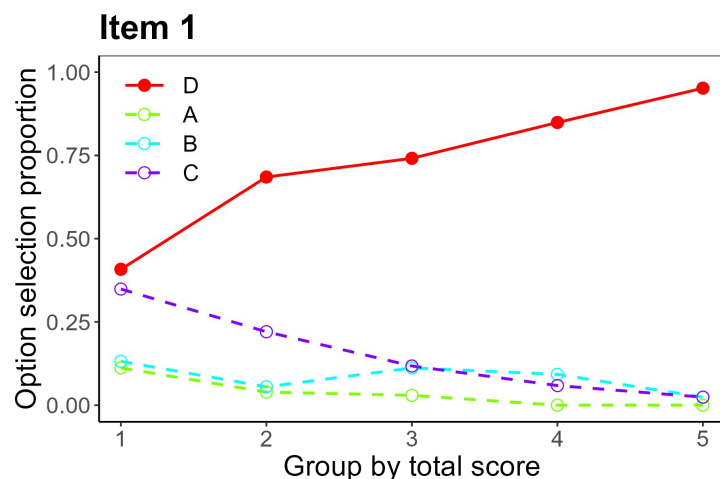


Download figure

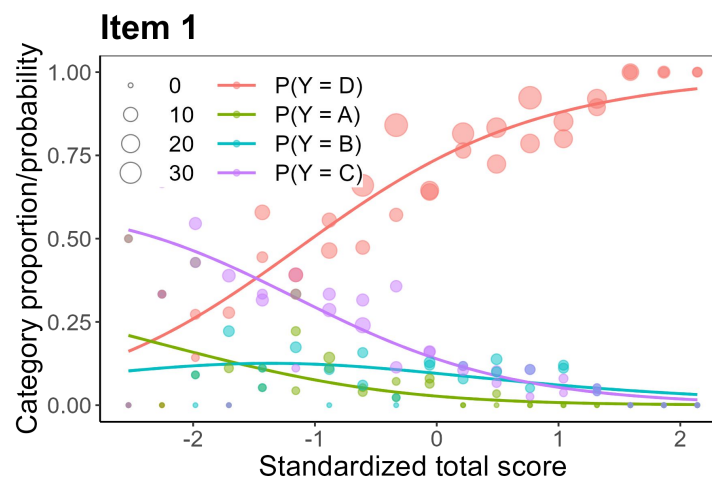
Od tradiční položkové analýzy k IRT

- Tradiční položková analýza využívá podílů, procent a korelací k popisu vlastností položky
- Regresní modely modelují pravděpodobnost dané (např. správné) odpověď' hladkou křivkou, odhadují se parametry křivky (obtížnost, diskriminace)
- Modely teorie odpovědi na položku (IRT) navíc předpokládají, že úroveň znalosti respondentů je neznámá a je třeba ji odhadnout společně s parametry položek

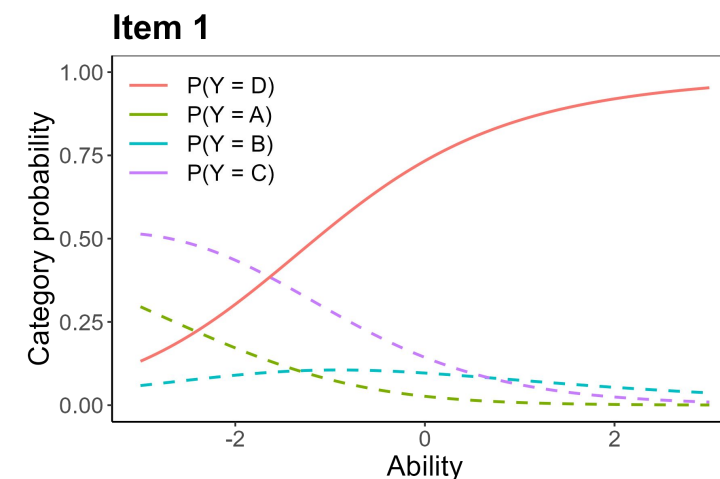
Tradiční položková analýza



Regresní modely



Modely IRT



VÝCHOZÍ TEXT K ÚLOZE 3

Vlak má tři vagony, všechny se stejným počtem míst. V každém vagonu je o 20 míst k stání více než k sezení.

Při odjezdu z Roztok byl vlak zaplněn přesně do poloviny své kapacity. V prvním a posledním vagonu byla všechna místa k sezení obsazená, ale ve druhém vagonu zůstalo 25 % míst k sezení volných.

(Kapacita vlaku je součet počtu všech míst k stání a sezení. Každý cestující obsadil buď jedno místo k stání, nebo jedno místo k sezení.)

(CZVV)

max. 2 body

3 Počet **míst k sezení** v jednom vagonu označme n .

Vyjádřete v závislosti na veličině n počet všech cestujících, kteří při odjezdu z Roztok

3.1 byli ve vlaku;

3.2 ve vlaku stáli.

Obtížné položky, dobře diskriminují

Položka b3.1

$a = 1.86\dots$ diskriminační schopnost

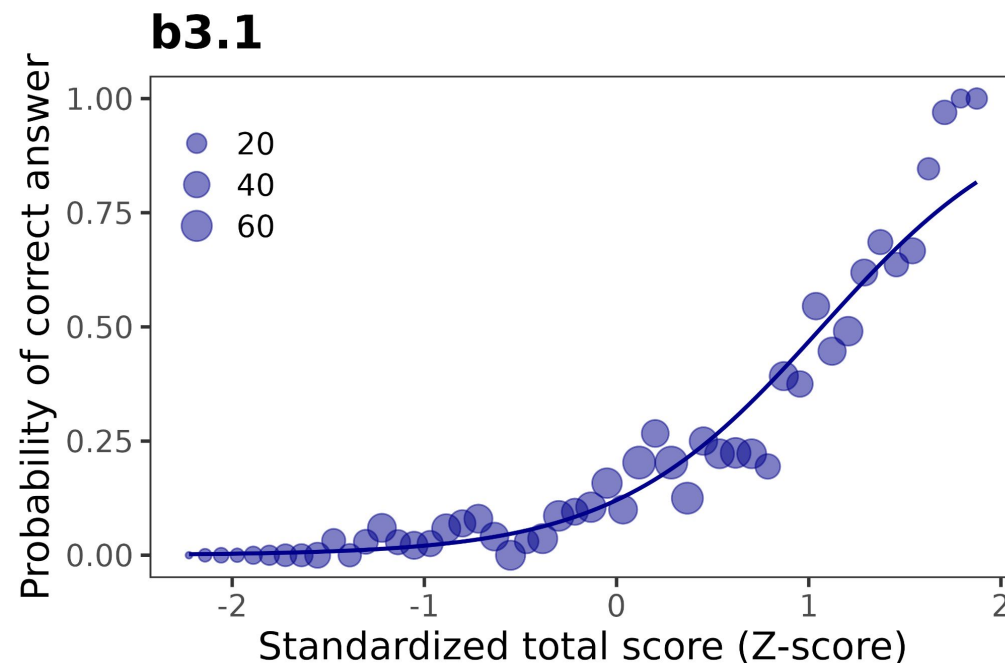
(sklon křivky v inflexním bodě)

$b = 1.07\dots$ obtížnost

(poloha inflexního bodu křivky)

Celkové skóre 1SD nad průměrem jsou potřeba k tomu aby položka byla

zodpovězena správně s pravděpodobností 0,5



max. 2 body

4 Pro $a \in \mathbb{R} \setminus \{-3; 0; 3\}$ zjednodušte:

$$\frac{1 + \frac{3}{a}}{\frac{a^2}{3} - 3} =$$

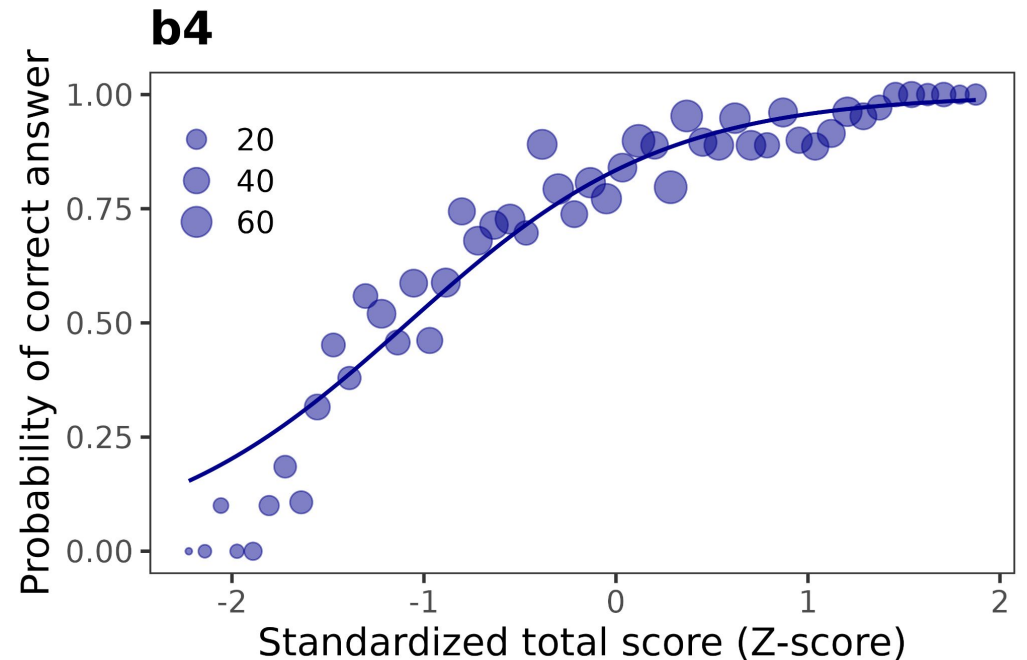
V záznamovém archu uveďte celý postup řešení.

Lehká položka, dobře diskriminuje

$a = 1.49 \dots$ diskriminační schopnost
(sklon křivky v inflexním bodě)

$b = -1.08 \dots$ obtížnost
(poloha inflexního bodu křivky)

Celkové skóre 1SD pod průměrem stačí k tomu aby položka byla zodpovězena správně s pravděpodobností 0,5

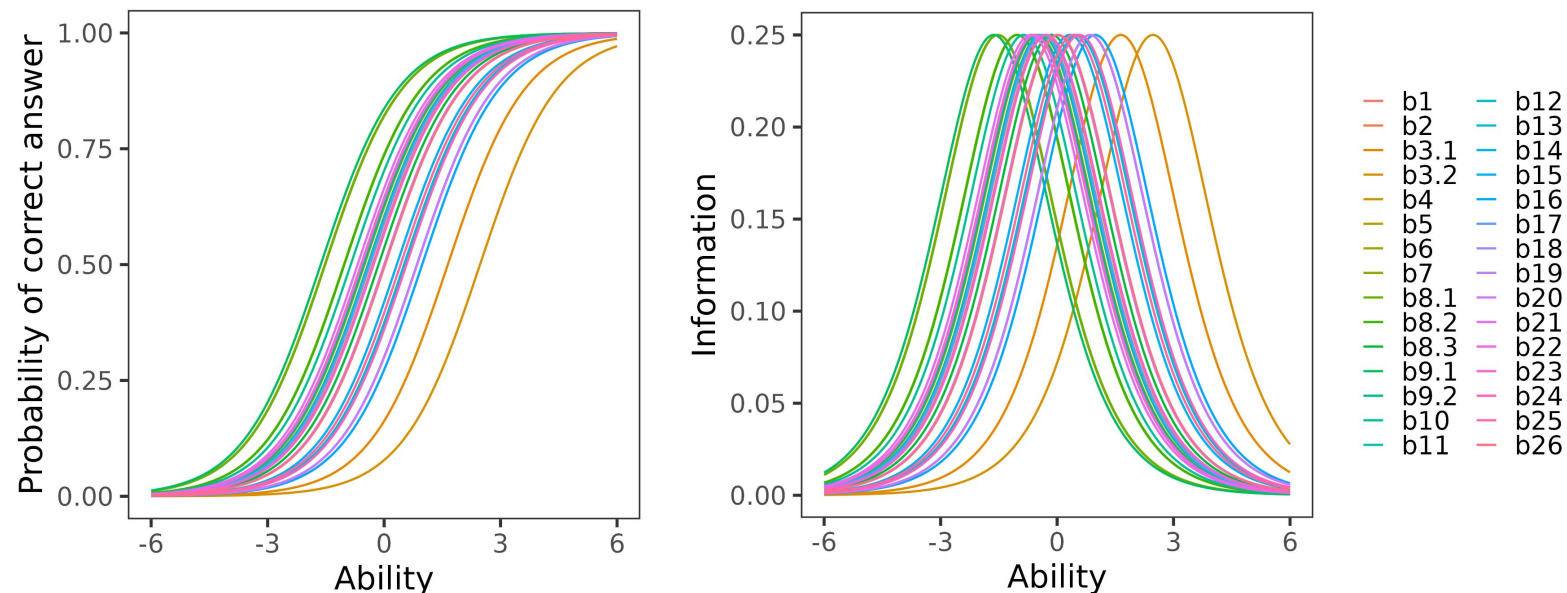


Modely teorie odpovědi na položku (IRT)

- Odhad znalosti respondentů společně s parametry položek
- Nutnost dostatečného množství pozorování (respondentů)
- 1–4PL modely pro binární data, modely pro ordinální i nominální položky

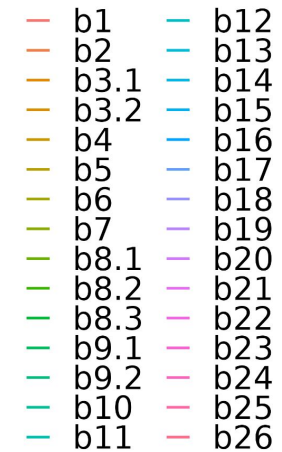
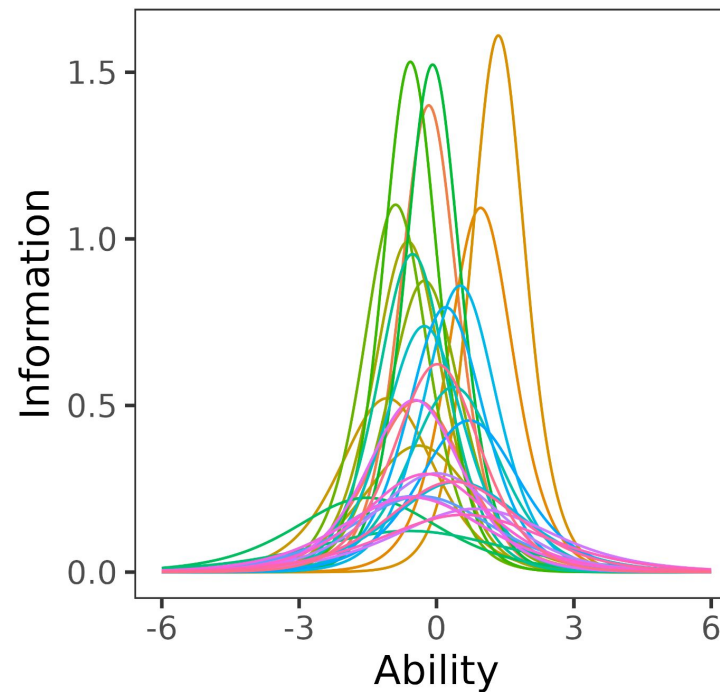
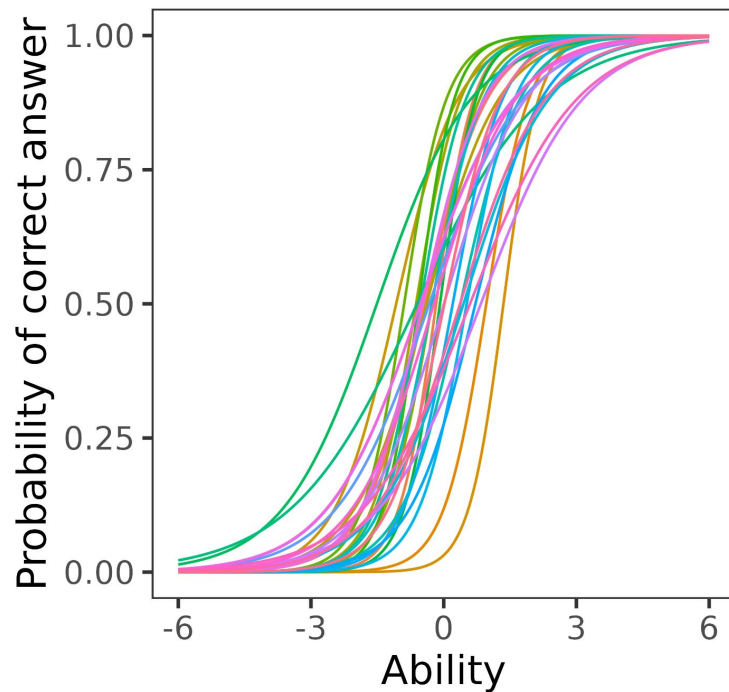
Raschův/1PL IRT model

- Různé obtížnosti položek
- Model předpokládá stejný sklon pro všechny položky
- Pro 1PL IRT model je odhad obtížnosti ekvivalentní s celkovým skóre (pokud jsou položky binární)



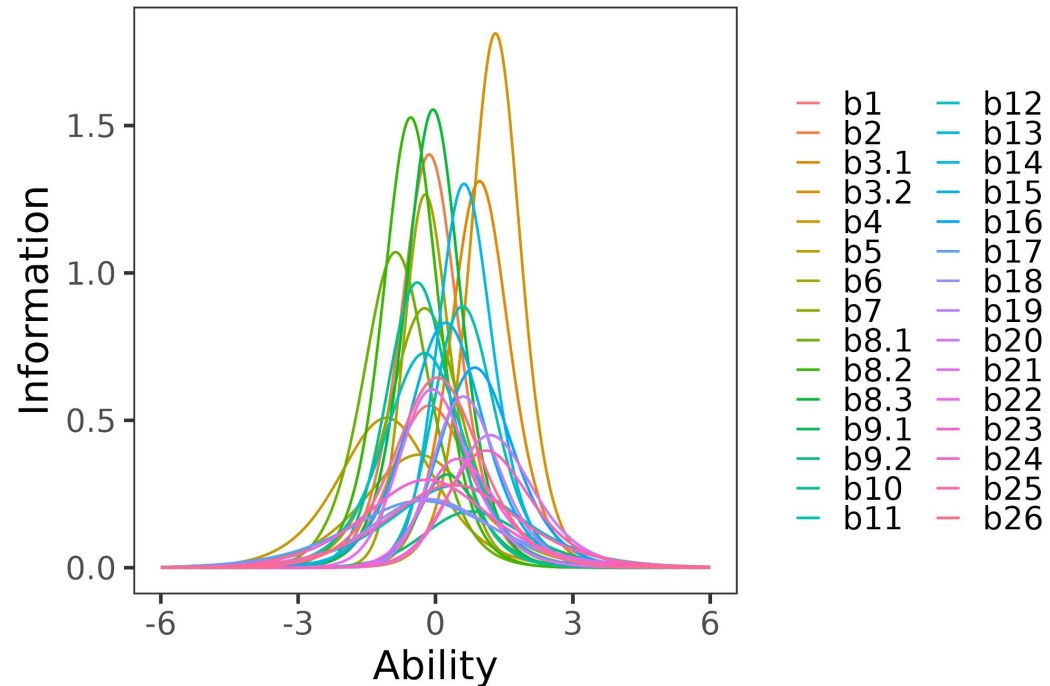
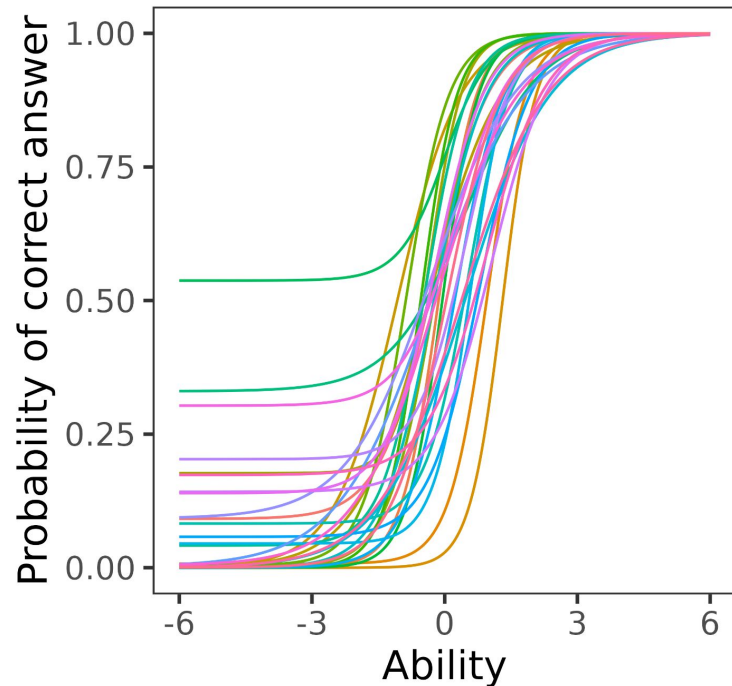
2PL IRT model

- Různé obtížnosti i sklony pro položky
- Každá položka je různě informativní



3PL IRT model

- Různé obtížnosti i sklony pro položky
- Modeluje možné hádání správné odpovědi (pseudo-guessing)



22 V geometrické posloupnosti platí:

$$a_2 = \sqrt[3]{3}$$

$$a_3 = -\sqrt[3]{9}$$

Jaká je hodnota součtu $a_1 + a_4$?

- A) 2
- B) 1
- C) 0
- D) -1
- E) jiná hodnota

Položka s možným hádáním správné odpovědi

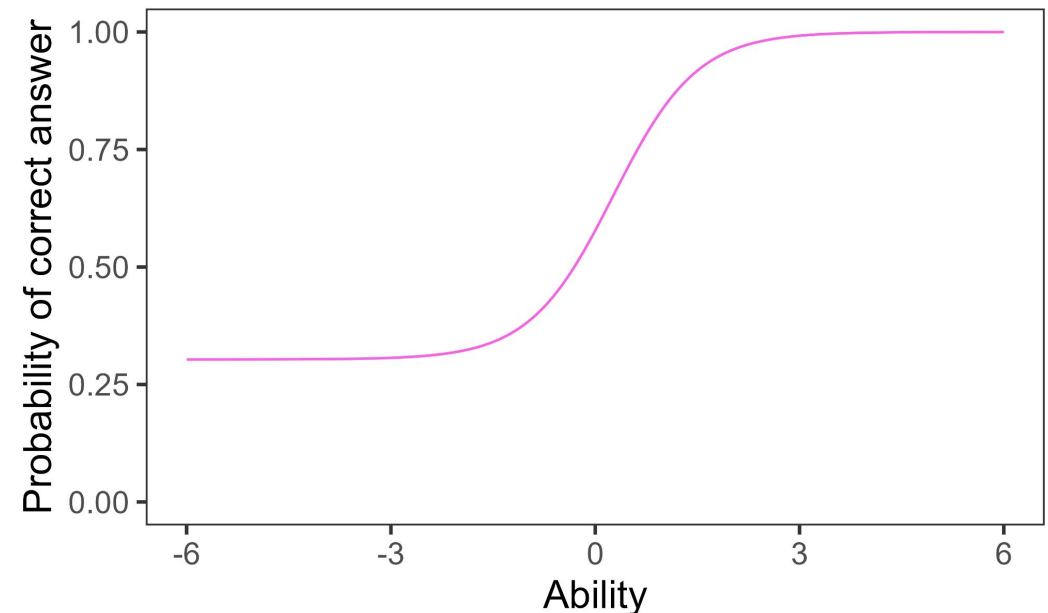
Odhadnuté parametry položky:

$a = 1.63$... diskriminační schopnost
(sklon křivky v inflexním bodě)

$b = 0.26$... obtížnost
(poloha inflexního bodu křivky)

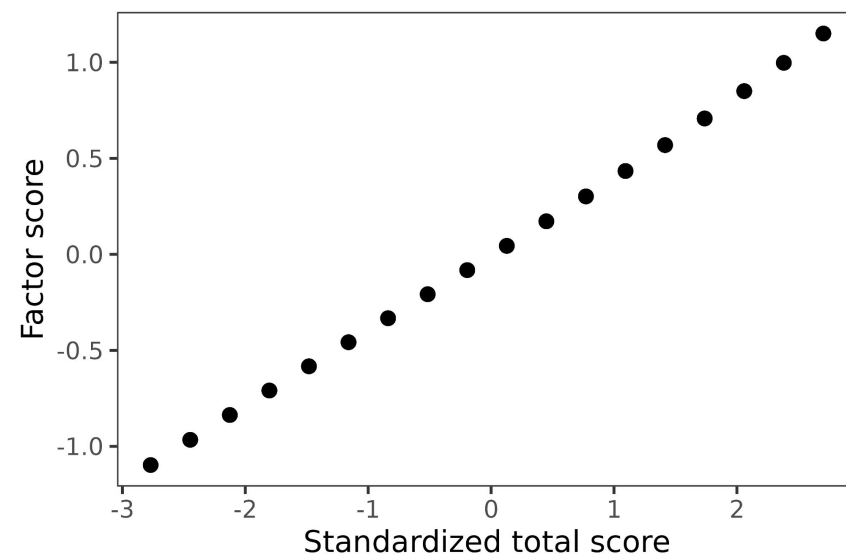
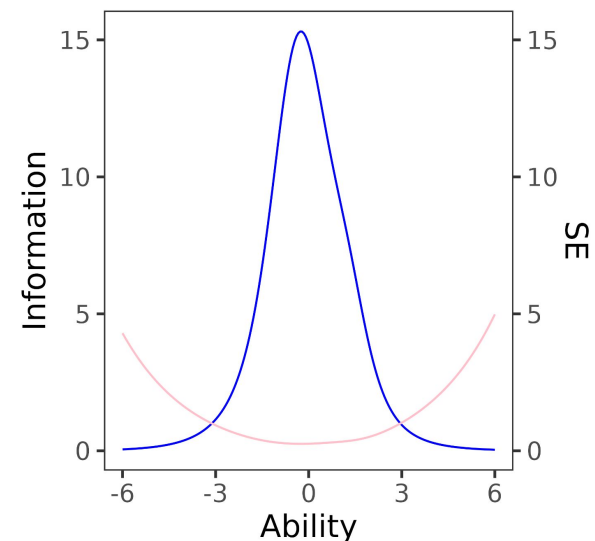
$c = 0.30$... uhádnutelnost
(dolní asymptota křivky)

b22



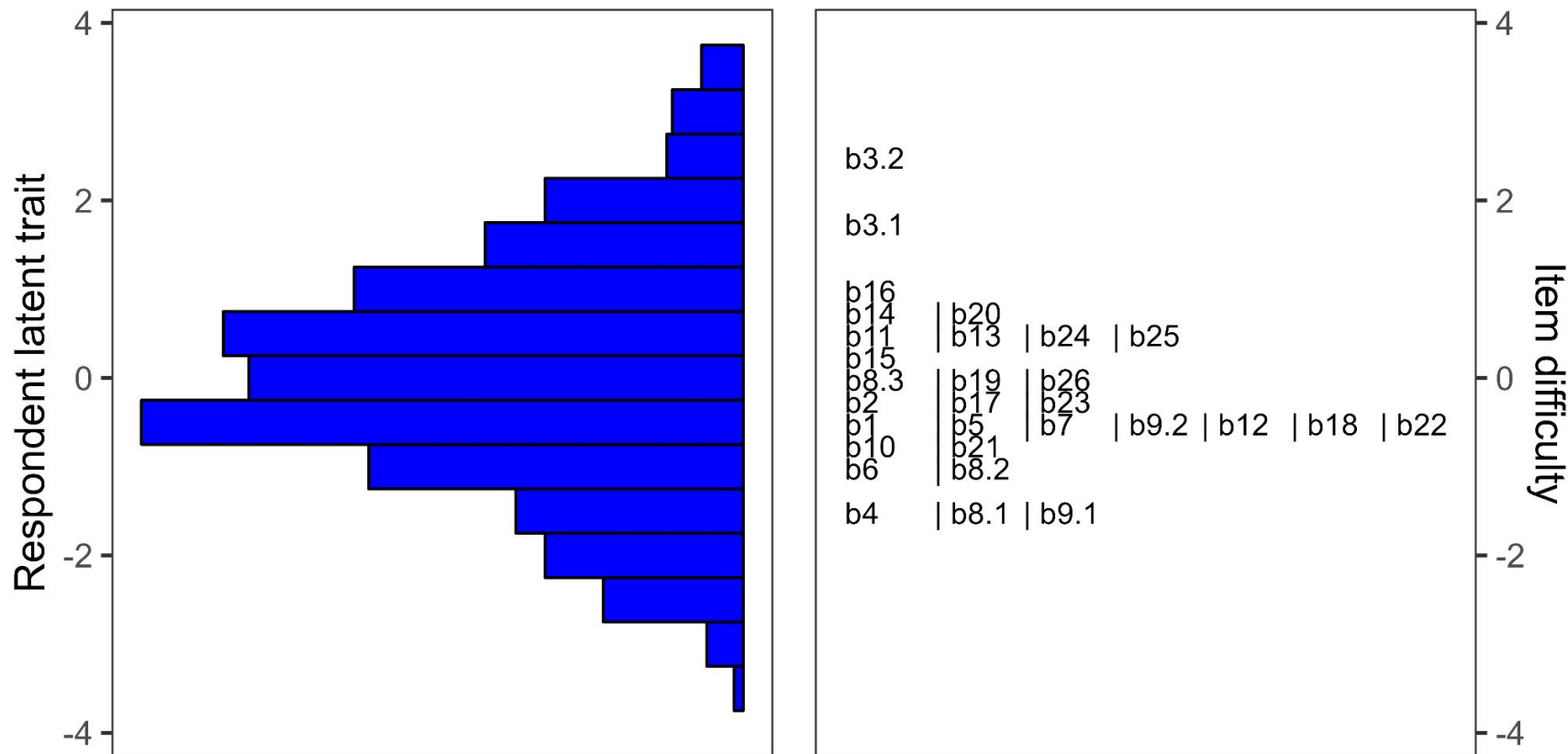
Využití modelů teorie odpovědi na položku

- IRT modely poskytnou konfidenční interval pro odhad znalosti žáka
 - Reliabilita testu není konstantní, ale je funkcí znalosti žáka
 - Pro 1PL model je odhadnutá znalost ekvivalentní s celkovým skóre
- Odhad znalosti v případě více verzí testu
- Informační funkce jednotlivých položek
 - Lze dále využít: počítačové adaptivní testování, vyvažování (equating), ...



Využití modelů teorie odpovědi na položku

- Interpretace obtížnosti položek vůči znalostem žáků (Wright Map)



[Data](#)[CTT](#)[IRT](#)

IRT

By pressing the button below, you can fit an IRT model based on raw, unscored responses for "k" items. Unlike the IRT section of the main app, this module allows you to define individual IRT "models" for each item, as specified in the metadata file (which you must upload before fitting the model). Below, you can review the model that is going to be fitted:

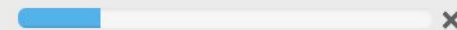
Model	Items
3PL	k15.1, k15.2, and k15.3
gpcm	k1, k2, k3, k4, k5.1, k5.2, k5.3, k6, k7, k8.1, k8.2, k9, k10, k11.1, k11.2, k12.1, k12.2, k13, and k14
nominal	k16, k17, k18, k19, k20, k21, k22, k23, k24, k25.1, and k25.2

[Fit the model](#)

Results

[Test](#)[Items](#)

Test information curve



Fitting the model This may take a while...



Data

Scores ▾

Validity ▾

Reliability ▾

Item analysis ▾

Regression ▾

IRT models ▾

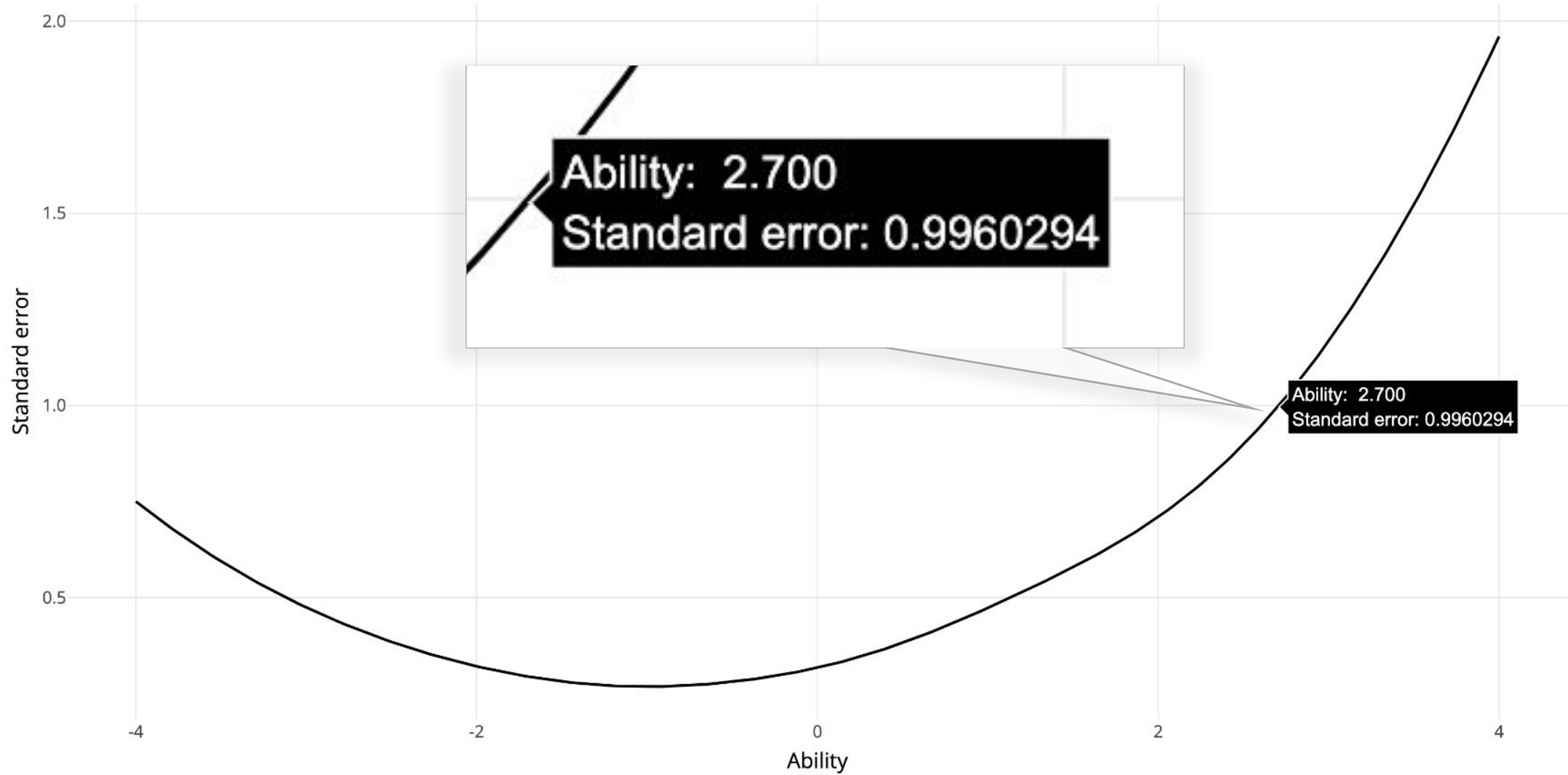
DIF/Fairness ▾

Modules ▾

Reports



Test standard errors



Test reliability



Data

Scores ▾

Validity ▾

Reliability ▾

Item analysis ▾

Regression ▾

IRT models ▾

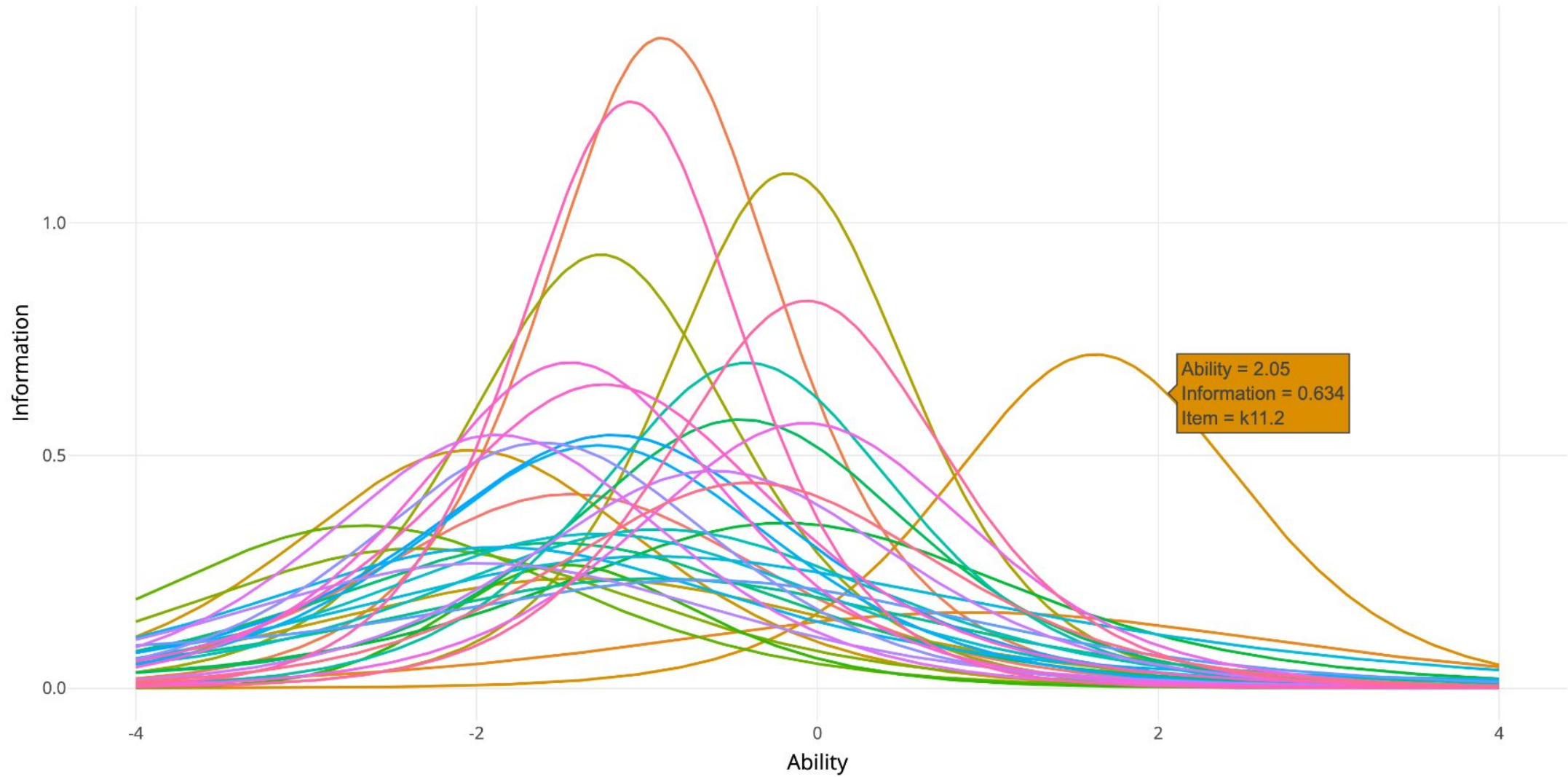
DIF/Fairness ▾

Modules ▾

Reports



Item information curves



Item characteristic curve



Data

Scores ▾

Validity ▾

Reliability ▾

Item analysis ▾

Regression ▾

IRT models ▾

DIF/Fairness ▾

Modules ▾

Reports



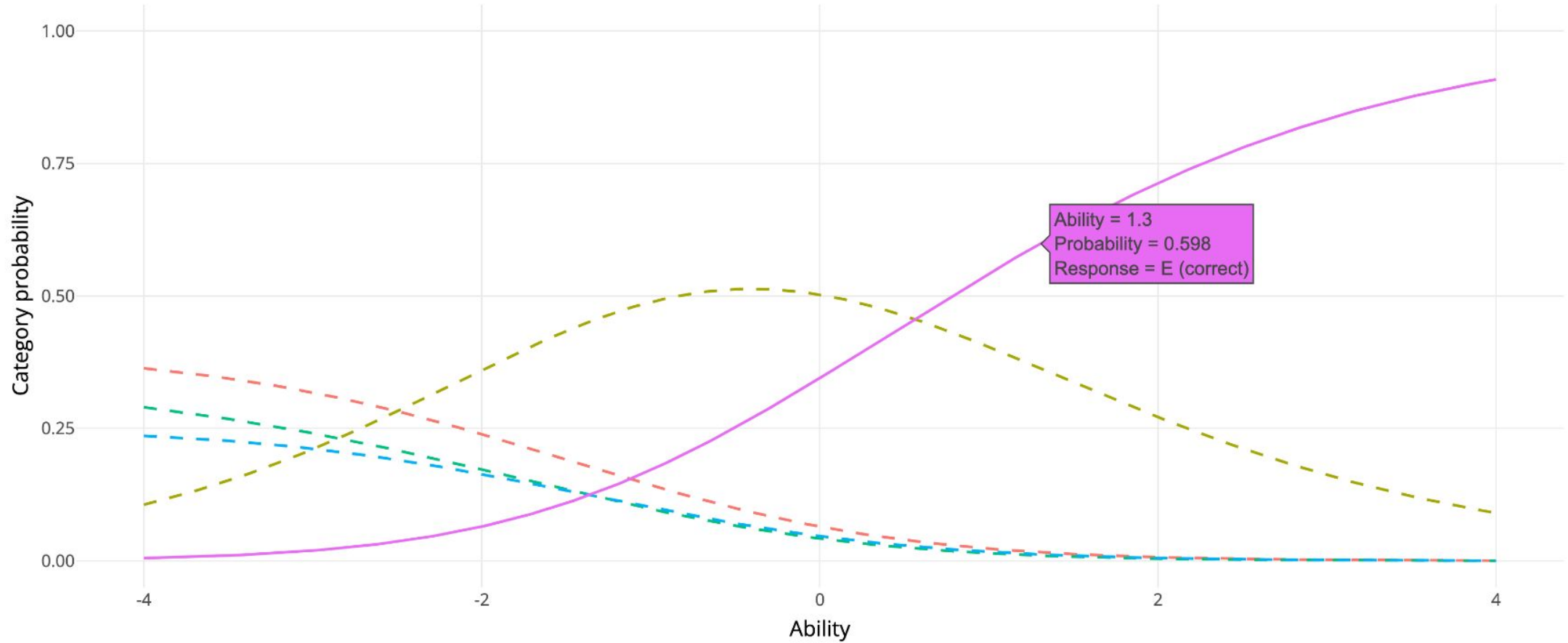
Item characteristic curve

Item

k22 ▾

Previous item

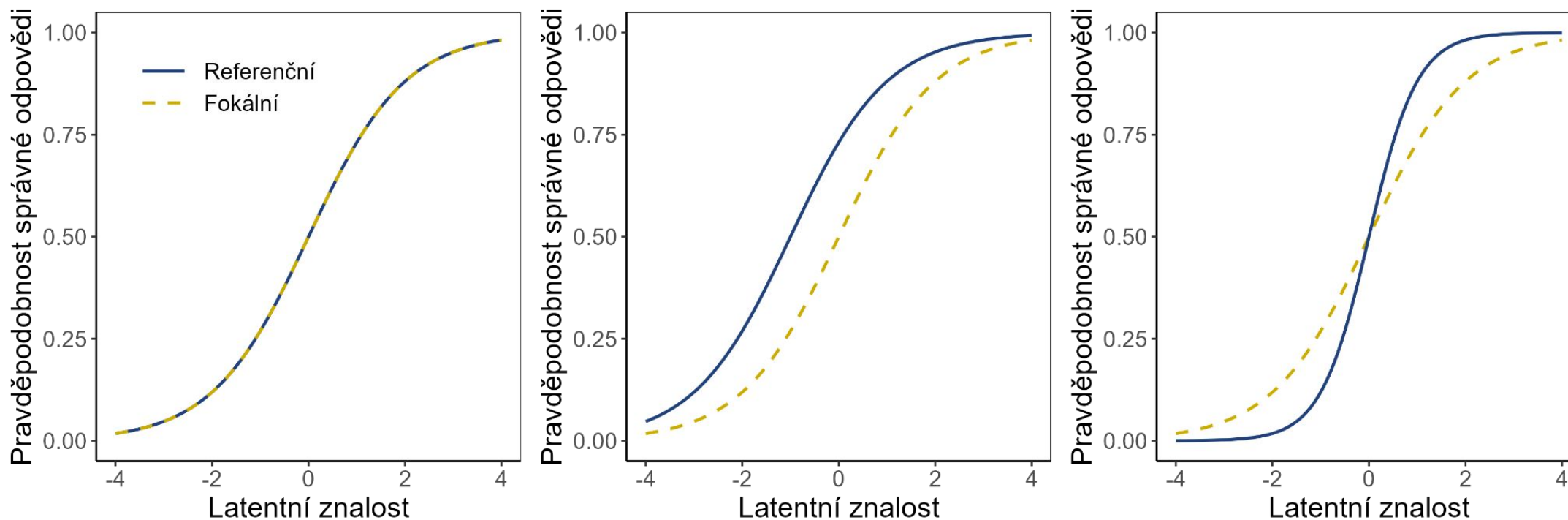
Next item



Odlišné fungování položek

DIF, z angl. Differential Item Functioning

= respondenti z různých skupin se stejnou znalostí mají odlišné pravděpodobnosti správné odpovědi na položku



Odlišné fungování položek

- Může poukazovat na neférovost položek (položka měří něco, co nemá)
- Ne všechny položky, které fungují odlišně, jsou neférové (položka měří, co má; rozdíl není způsoben neférovostí)
- DIF analýza
 - přináší hlubší poznání o fungování položek, miskonceptcích, senzitivitě na instruktáž
 - pomáhá informovat tvůrce testu o charakteristikách položek
 - pomáhá informovat o silných a slabých stránkách skupin

Odlišné fungování položek – příklady

DIF položka v testu SAT z 60. let – **označena jako neférová**

„Běžec je k maratonu jako“

- A. vyslanec k velvyslanectví
- B. mučedník k masakru
- C. veslař k regatě**
- D. rozhodčí k turnaji
- E. koně ke stájím

častěji správně odpověděli bílí než afroameričtí studenti se stejnými celkovými znalostmi

- ? **Co způsobilo rozdíly mezi skupinami v odpovídání na tuto položku?**
- ? **Byl rozdíl způsobený znalostí něčeho, co test neměl testovat? – ANO**

Odlišné fungování položek – příklady

DIF položka z přijímacího testu na lékařské fakulty – **NE**označena jako neférová

„Nedostatek vitamínu D v dětství může způsobit“

- A. křivici
- B. kurděje
- C. nanismus
- D. mentální retardaci

častěji správně odpověděly studentky než studenti se stejnými celkovými znalostmi

- ? Co způsobilo rozdíly mezi skupinami v odpovídání na tuto položku?
- ? Byl rozdíl způsobený znalostí něčeho, co test neměl testovat? – **NE**

Příklady využití DIF analýzy

- Pro tvůrce testů
 - ověření férovosti, vybalancování položek napříč skupinami
- Pro tvůrce státních politik
 - porozumění vzdělávacím výsledkům různých skupin
- Pro školy a vyučující
 - odhalení silných a slabých stránek

DIF analýza: Příklad 1

- Využití DIF analýzy pro tvůrce testů, pro vyučující
 - JPZ 2023, matematika
 - UA vs. CZ žáci
-
- Položky fungují odlišně, ale nejsou neférové (testují znalost matematiky)

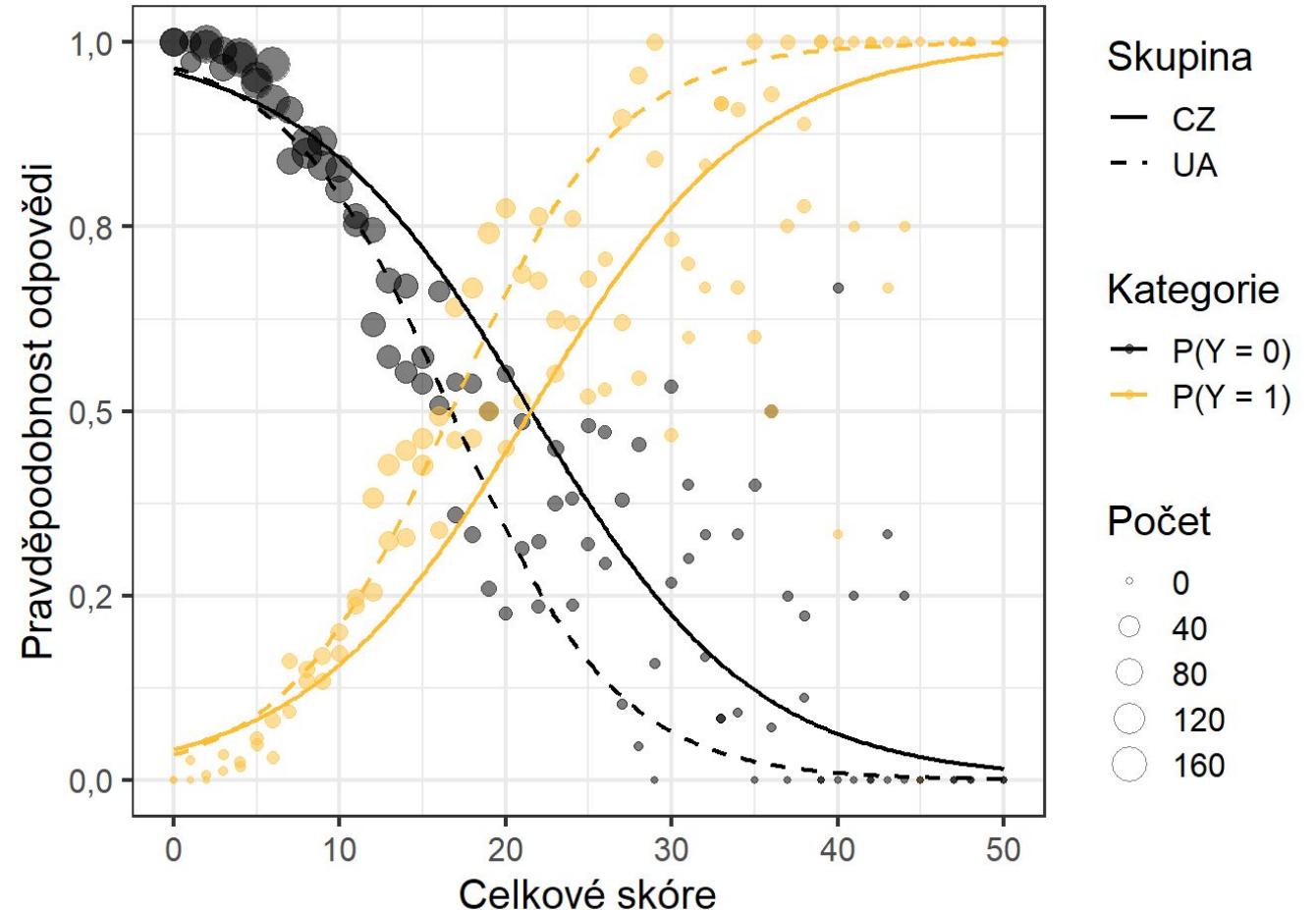
4.1 Upravte a rozložte na součin vytknutím:

$$2 \cdot (x^2 - x) + x =$$

Ukrajínští žáci mají **větší** pravděpodobnost správné odpovědi při stejném celkovém skóre

Možné důvody:

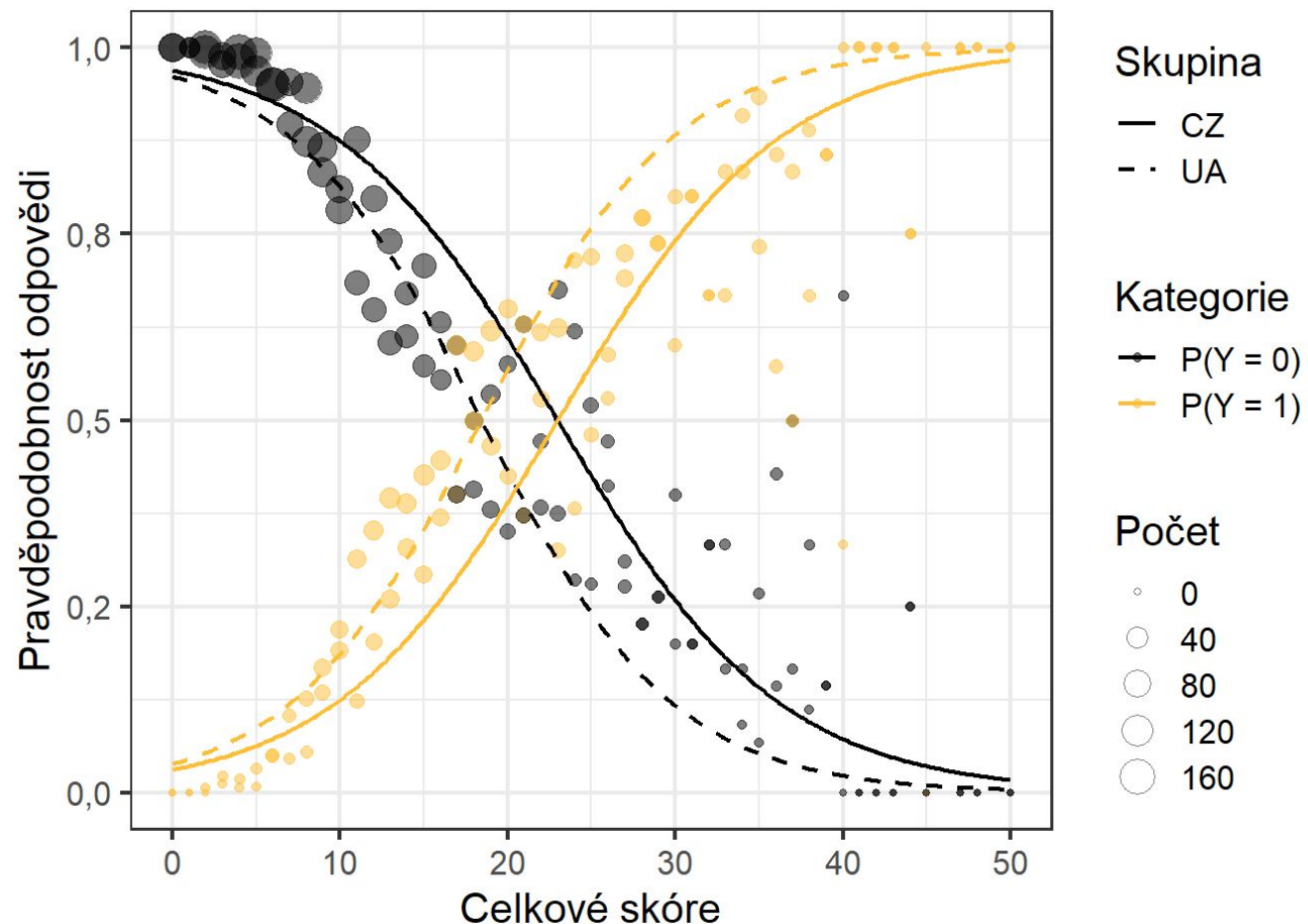
- jiné kurikulum
- málo textu



4.2 Umocněte a zjednodušte (výsledný výraz nesmí obsahovat závorky):

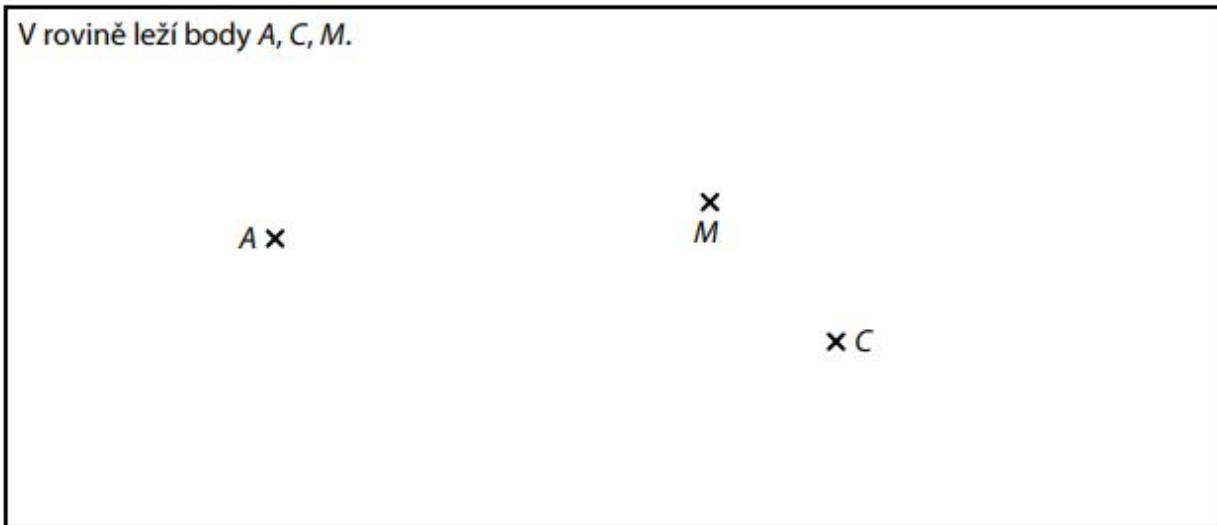
$$\left(\frac{2}{3}a - 3\right)^2 =$$

Ukrajínští žáci mají **větší** pravděpodobnost správné odpovědi při stejném celkovém skóre



VÝCHOZÍ TEXT A OBRÁZEK K ÚLOZE 9

V rovině leží body A, C, M .



(CZVV)

max. 2 body

9 Body A, C jsou vrcholy obdélníku $ABCD$.
Bod M leží na úhlopříčce BD tohoto obdélníku.

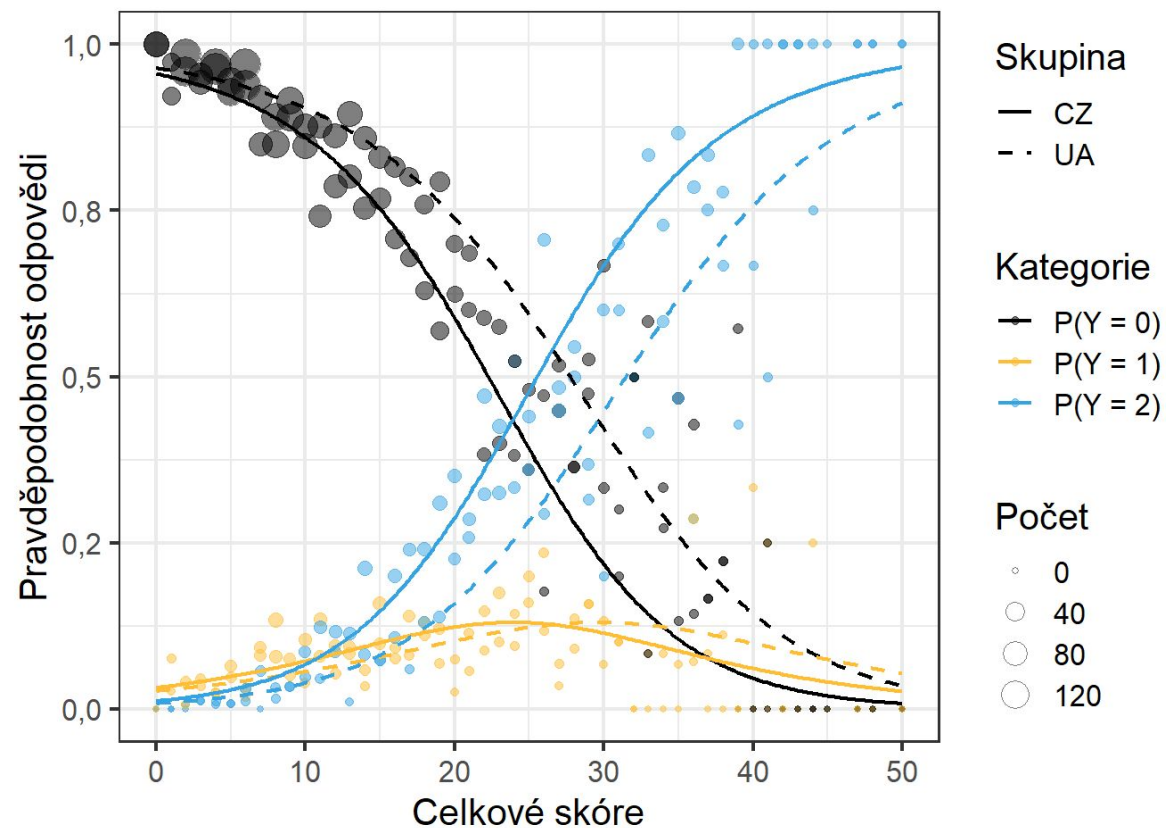
Sestrojte vrcholy B, D obdélníku $ABCD$, **označte** je písmeny a obdélník **narýsujte**.

V záznamovém archu obtáhněte celou konstrukci **propisovací tužkou** (čáry i písmena).

Ukrajínští žáci mají **menší** pravděpodobnost zisku plného počtu bodů

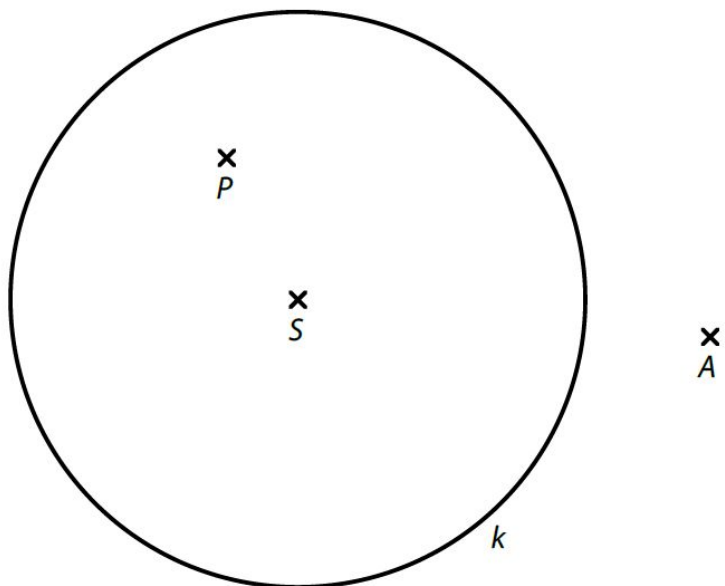
Možné důvody:

- jiné kurikulum
- více textu



VÝCHOZÍ TEXT A OBRÁZEK K ÚLOZE 10

V rovině leží body A, P a kružnice k se středem S .



(CZV)

max. 3 body

10 Bod A je vrchol **rovnoramenného** trojúhelníku ABC , jehož **základna** leží na přímce AP .

Vrcholy B, C tohoto trojúhelníku leží na kružnici k .

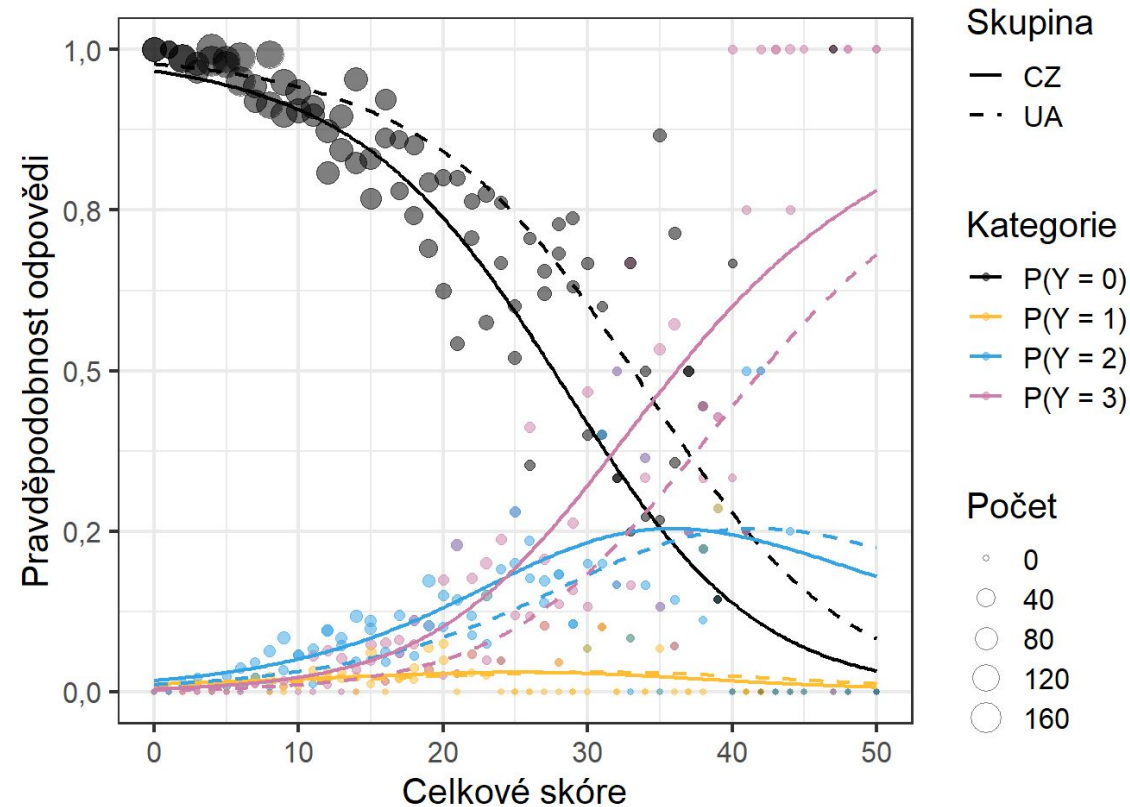
Sestrojte vrcholy B, C trojúhelníku ABC , **označte** je písmeny a trojúhelník **narýsujte**. Najděte všechna řešení.

V záznamovém archu obtáhněte celou konstrukci **propisovací tužkou** (čáry i písmena).

Ukrajínští žáci mají **menší** pravděpodobnost zisku plného počtu bodů

Možné důvody:

- jiné kurikulum
- více textu, specifické pojmy



DIF analýza: Příklad 2

- Využití DIF analýzy pro tvůrce státních politik
- Maturitní test, matematika, jaro 2019
- Gymnázia vs. technické střední školy (ST1, ST2)

- Položky fungují odlišně, ale nejsou neférové (testují znalost matematiky)
- DIF analýza poukazuje na silné stránky různých typů škol

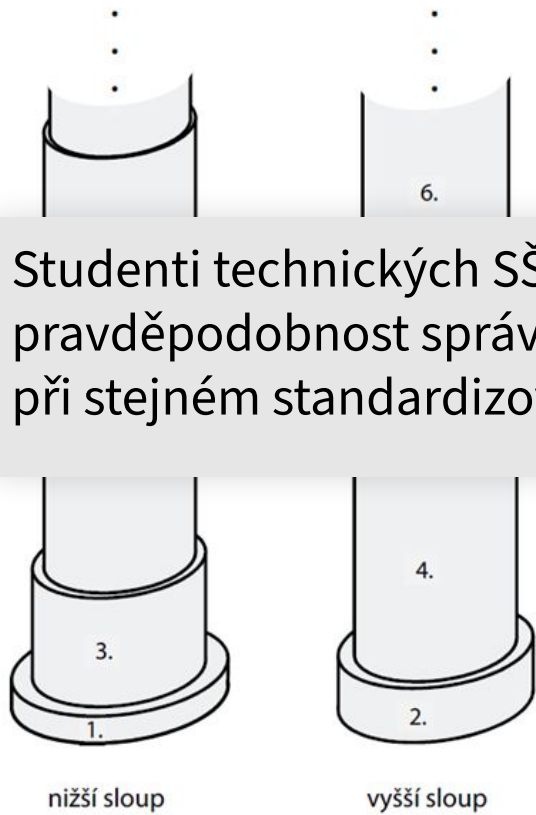
VÝCHOZÍ TEXT A OBRÁZEK K ÚLOZE 9

V Kocourkově navrhli nereálný plán stavby dvou sloupů sahajících do nebe.

Na stavbu se má použít celkem 20 válců. Jednotlivé válce jsou podle výšky označeny pořadovými čísly od 1 do 20.

Nejnižší je 1. válec s výškou 1 m, 2. válec má výšku 2 m a rovněž každý další válec je dvakrát vyšší než válec s pořadovým číslem o 1 nižším. (Tedy 3. válec má výšku 4 m, 4. válec 8 m atd.)

Nižší sloup bude postaven ze všech válců označených lichými pořadovými čísly od 1 do 19, vyšší sloup ze všech válců označených sudými pořadovými čísly od 2 do 20.



Studenti technických SŠ mají **větší** pravděpodobnost správné odpovědi při stejném standardizovaném skóre

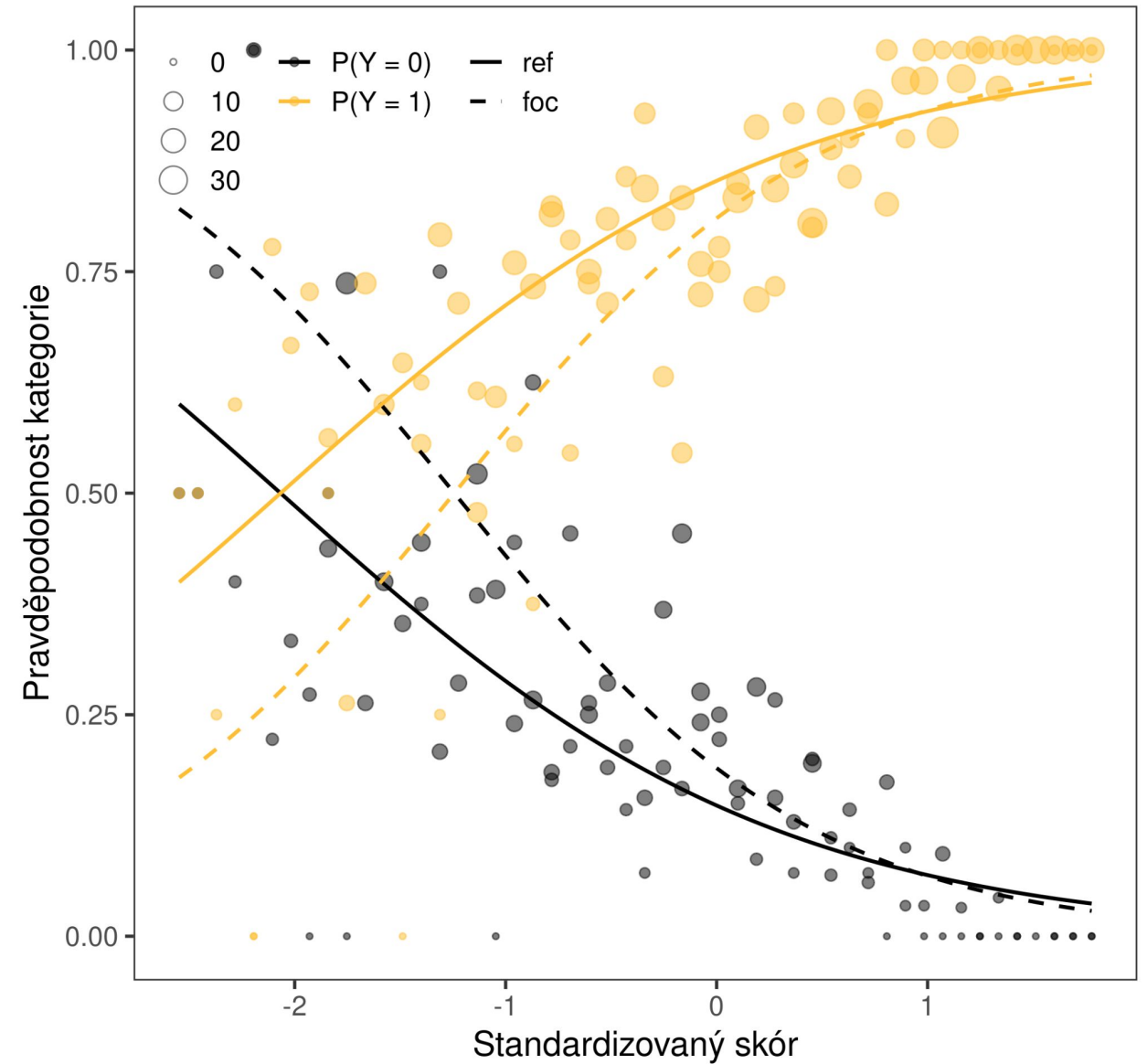
(CZVV)

max. 2 body

9 Určete v metrech

9.1 výšku 20. válce;

b9.1



DIF analýza: Příklad 3

- Využití DIF analýzy pro školy
 - Maturitní test, český jazyk a literatura, jaro 2022
 - Vybraná škola vs. jiné školy stejného typu
-
- Položky fungují odlišně, ale nejsou neférové (testují znalost českého jazyka a literatury)
 - DIF analýza poukazuje na silné stránky dané školy

Úspěšnost v úlohách testu

Detailní analýza rozdílů

Analýza rozdílného fungování úloh testu pro různé skupiny

Úvod do složitějšího tématu

Analýza meziskupinových rozdílů

Tato analýza porovnává úspěšnost žáků Vaší školy s úspěšností žáků zvolené srovnávací skupiny vždy pro žáky se stejným celkovým bodovým skórem. Díky tomu umožňuje zjistit silnější a slabší stránky Vašich žáků v případě, že celková úspěšnost Vaší školy se výrazně liší od průměru srovnávací skupiny a tento výrazný celkový rozdíl překrývá možné odchylky napříč jednotlivými testovými úlohami.

I když mají jednotliví žáci stejný celkový bodový výsledek, jednotlivé úlohy testu pro ně mohou být odlišně obtížné. **I žáci s naprosto identickým celkovým výsledkem tedy mohou mít jinou pravděpodobnost, že na danou úlohu odpoví správně.** Pokud se podobné rozdíly v odpovědích objevují systematicky u různých zdatných žáků z jedné skupiny, může to naznačovat silné nebo naopak slabé stránky této skupiny.

Vyznačení úloh, na které se zaměřit

U jakých úloh jsme detekovali meziskupinové rozdíly?

Jedná se o 4 úlohy: „6“ (7), „14“ (15), „18“ (19) a „27“ (28). V závorkách uvádíme pořadová čísla pro výběr na posuvníku níže. Popis úloh a jejich domény nabízíme napravo od posuvníku.

Jednotlivé úlohy

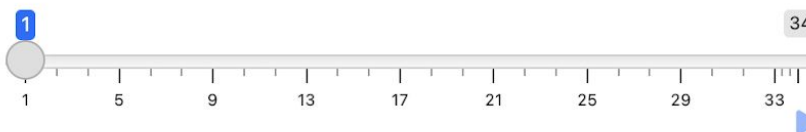
Shrnutí

Krátký popis úlohy, její zařazení

Vyberte úlohu

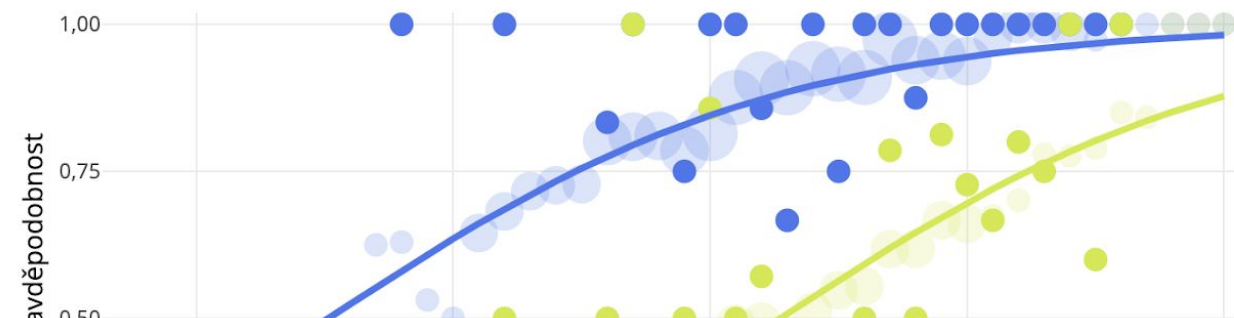
Na posuvníku vpravo vyberte úlohu, která Vás zajímá.

Tlačítkem ▶ můžete automaticky projít všechny.



Úloha „1“

Popis: porozumění textu: rozhodnout, zda tvrzení vyplývá z výchozího textu, či nevyplývá
Doména: Porozumění textu



U jakých úloh jsme detekovali meziskupinové rozdíly?

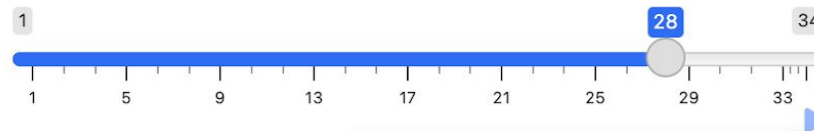
Jedná se o 5 úloh: „6“ (7), „9“ (10), „14“ (15), „18“ (19) a „27“ (28). V závorkách uvádíme pořadová čísla pro výběr na posuvníku níže. Popis úloh a jejich domény nabízíme napravo od posuvníku.

Jednotlivé úlohy [Shrnutí](#)

Vyberte úlohu

Na posuvníku vpravo vyberte úlohu, která Vás zajímá.

Tlačítkem ▶ můžete automaticky projít všechny.

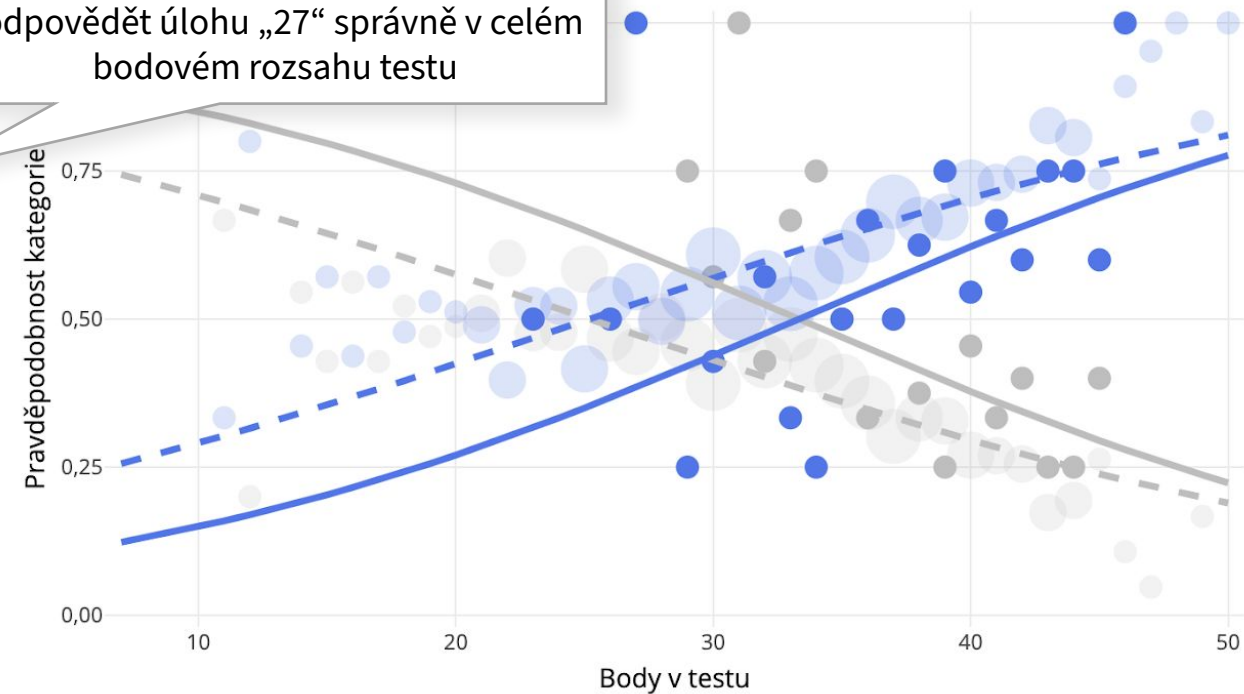
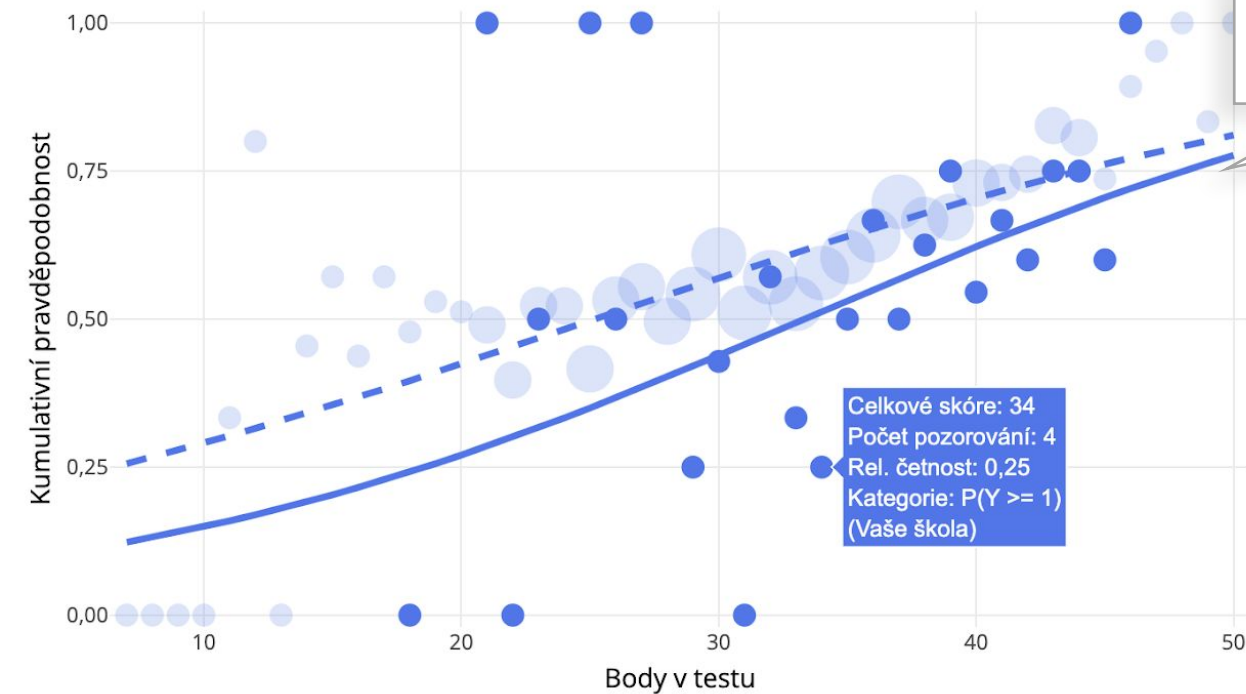


Úloha „27“

Popis: literární teorie: identifikovat literární žánr (sonet)

Doména: Literatura

Vaše škola má **menší** pravděpodobnost zodpovědět úlohu „27“ správně v celém bodovém rozsahu testu



U jakých úloh jsme detekovali meziskupinové rozdíly?

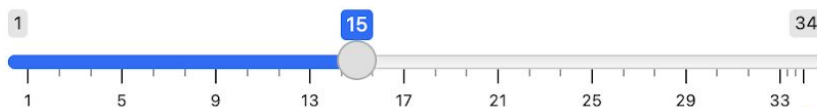
Jedná se o 5 úloh: „6“ (7), „9“ (10), „14“ (15), „18“ (19) a „27“ (28). V závorkách uvádíme pořadová čísla pro výběr na posuvníku níže. Popis úloh a jejich domény nabízíme napravo od posuvníku.

[Jednotlivé úlohy](#)
[Shrnutí](#)

Vyberte úlohu

Na posuvníku vpravo vyberte úlohu, která Vás zajímá.

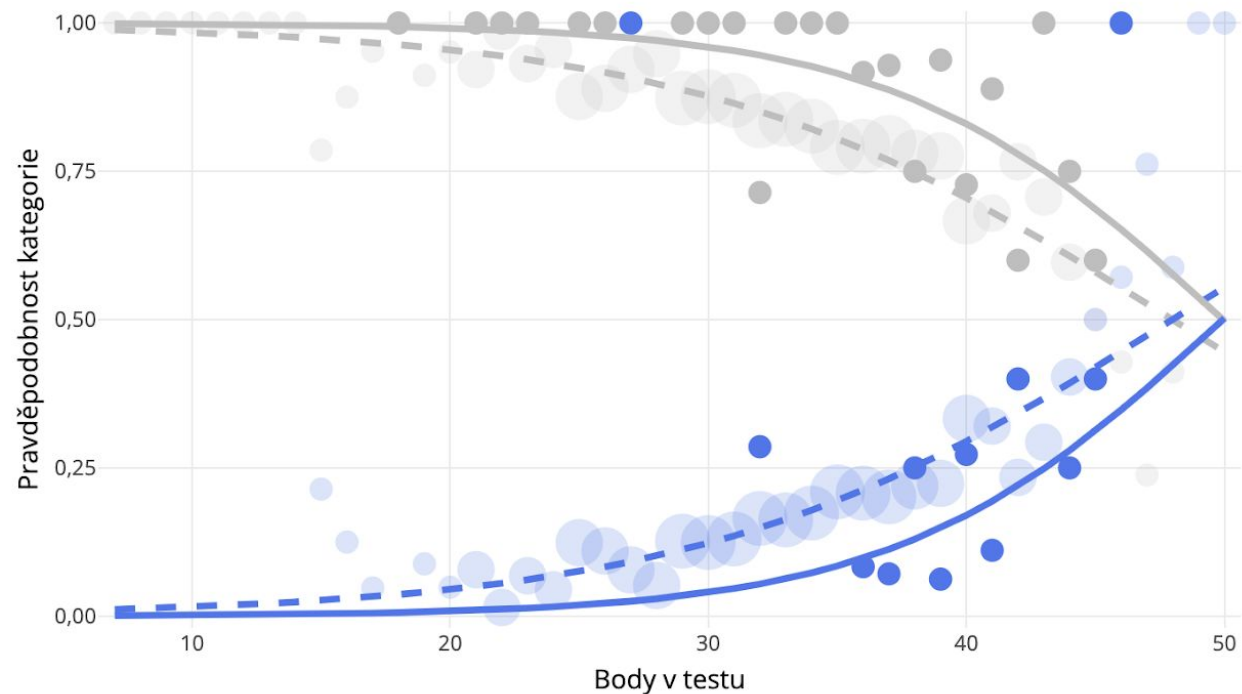
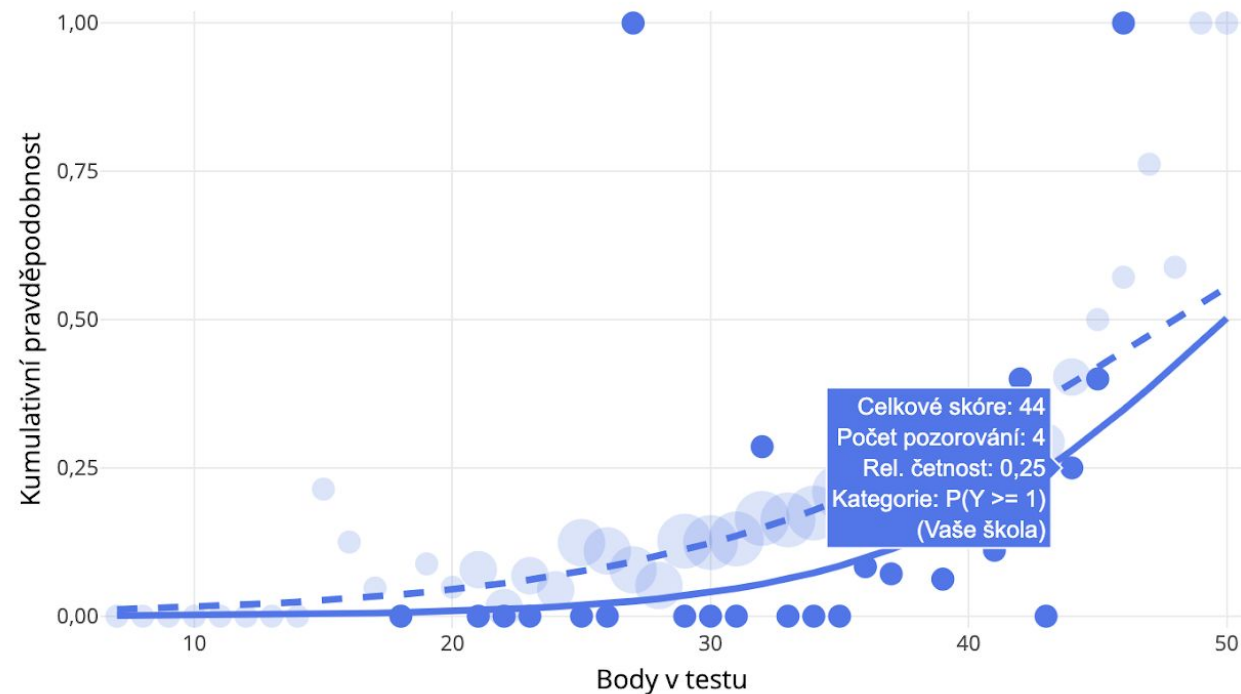
Tlačítkem ▶ můžete automaticky projít všechny.



Úloha „14“

Popis: charakter textu: určit funkční styl

Doména: Charakter a výstavba textu



U jakých úloh jsme detekovali meziskupinové rozdíly?

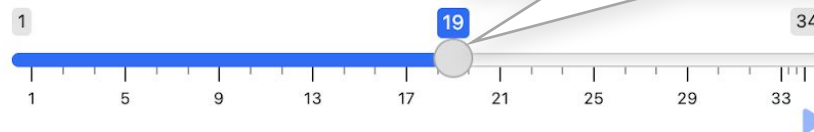
Jedná se o 5 úloh: „6“ (7), „9“ (10), „14“ (15), „18“ (19) a „27“ (28). V závorkách uvádíme pořadová čísla pro výběr na posuvníku níže. Popis úloh a jejich domény nabízíme napravo od posuvníku.

[Jednotlivé úlohy](#)
[Shrnutí](#)

Vyberte úlohu

Na posuvníku vpravo vyberte úlohu, která Vás zajímá.

Tlačítkem ▶ můžete automaticky projít všechny.

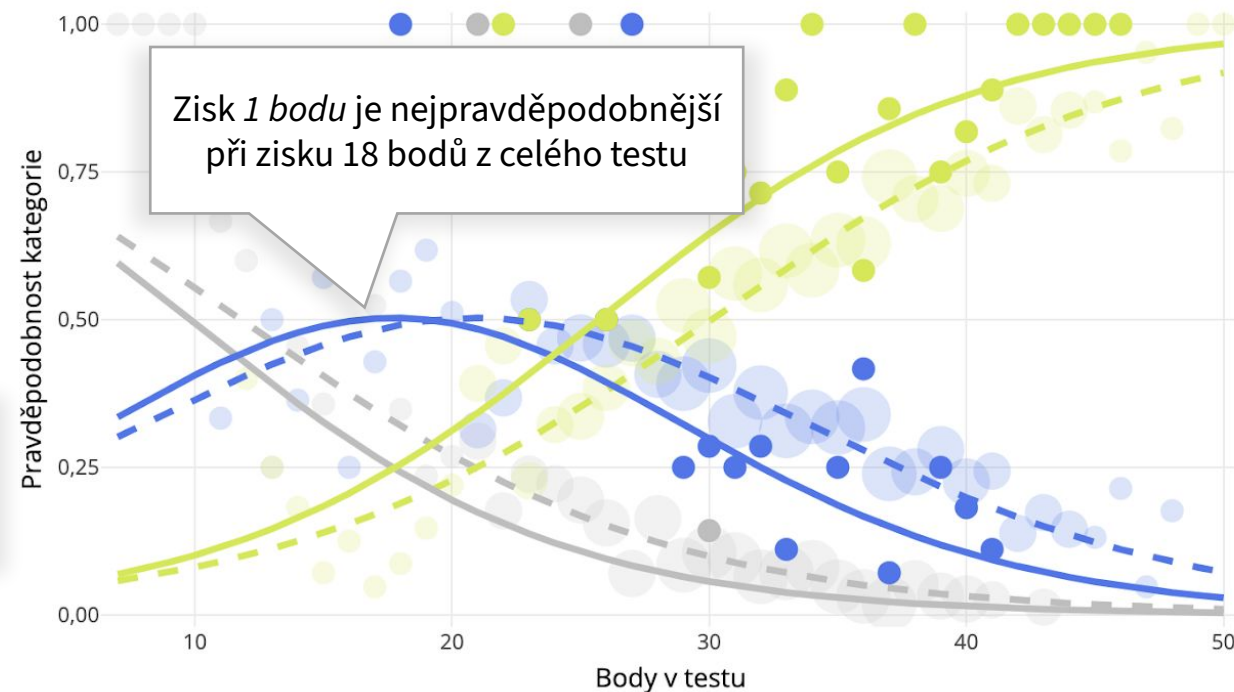
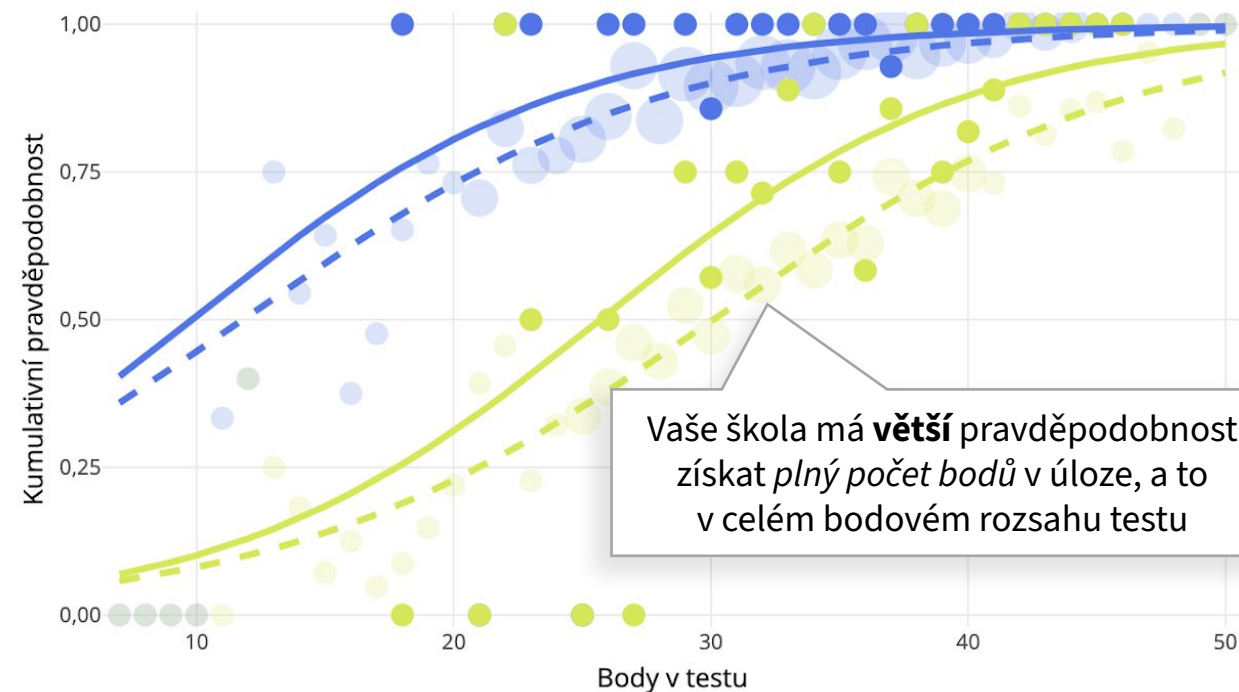


Komplikovanější úloha
s možným ziskem 0–2 body

Úloha „18“

Popis: porozumění textu: rozhodnout, zda tvrzení vyplývá z výchozího textu, či nevyplývá

Doména: Porozumění textu



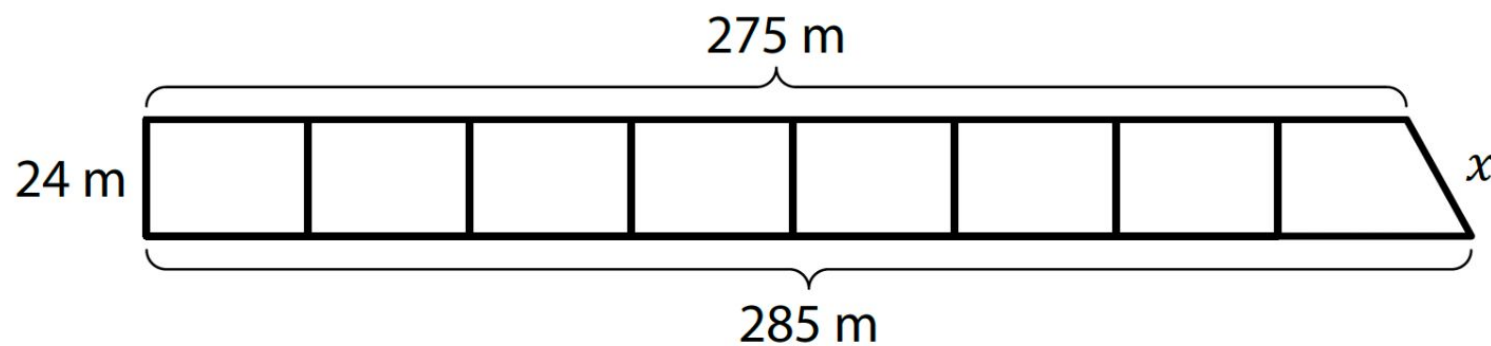
DIF analýza: Příklad 4

- didaktický test z matematiky, maturitní zkouška jaro 2022
- chlapci vs. dívky

- položky fungují odlišně, ale nejsou neférové
(testují znalost matematiky)

VÝCHOZÍ TEXT A OBRÁZEK K ÚLOZE 5

Pozemek má tvar pravoúhlého lichoběžníku s výškou 24 m a základnami délek 285 m a 275 m. Pozemek je rozdělen na **8 parcel** o stejné výměře. Prvních sedm parcel tvoří shodné obdélníky, poslední parcela má tvar pravoúhlého lichoběžníku.



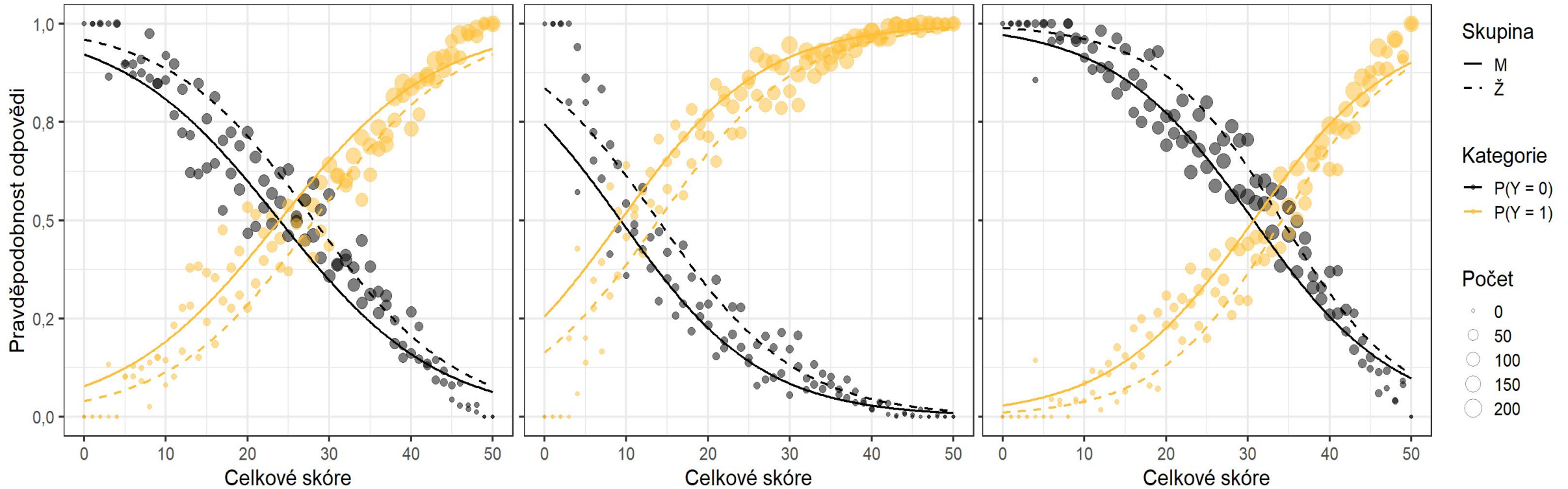
(CZVV)

max. 3 body

5 Vypočtete

- 5.1 v m^2 výměru jedné parcely,
- 5.2 v m chybějící délku x strany pozemku,
- 5.3 v m obvod **poslední** parcely.

„Podpoložky“ b5.1, b5.2 a b5.3



Dívky odpovídají správně s **menší** pravděpodobností než chlapci.
Možné důvody: nároky na prostorovou orientaci?

7 V oboru R řešte:

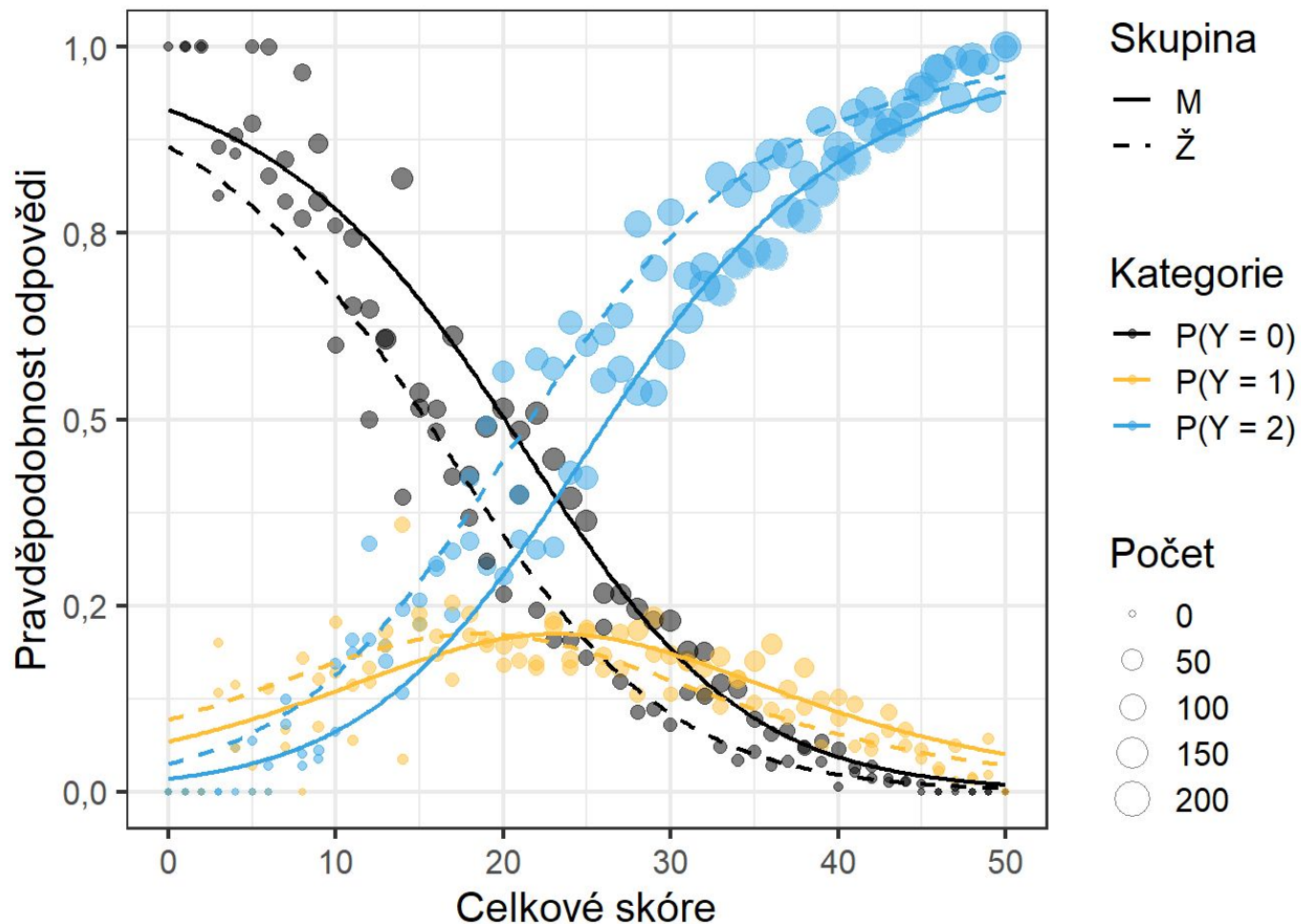
$$\frac{x-2}{x+2} \cdot \frac{3}{x} + \frac{16}{x^2+2x} = \frac{x}{x+2}$$

V záznamovém archu uveďte celý postup řešení.

max. 2 body

Dívky mají větší pravděpodobnost zisku plného počtu bodů.

Možné důvody: nároky na pečlivost/preciznost?



Praktická část

Praktická část a diskuse

- V aplikaci [ShinyItemAnalysis](#) (záložka Data) vyberte dataset CZmaturaS
 - (MZ z matematiky, jaro 2019; skupina: gymnázia vs. ostatní; vzorek 2 000 žáků)
 - Vyzkoušejte si dostupné analýzy v záložkách:
Item Analysis, Regression, IRT models, DIF/Fairness
- Prohlédněte si modul CERMAT tools (záložka Modules)
 - Nahrajte položková data M 2022 a příslušná metadata
 - Vyzkoušejte si dostupné analýzy
- ? Které analýzy využíváte při tvorbě a vyhodnocování testů?
- ? Které analýzy považujete za klíčové?

Online materiály a odkazy: cs.cas.cz/comps/workshop

III. Další témata

Vyvažování verzí testu,
odhadování obtížnosti z analýzy textu,
počítačové adaptivní testování (CAT)



Vyvažování verzí testu

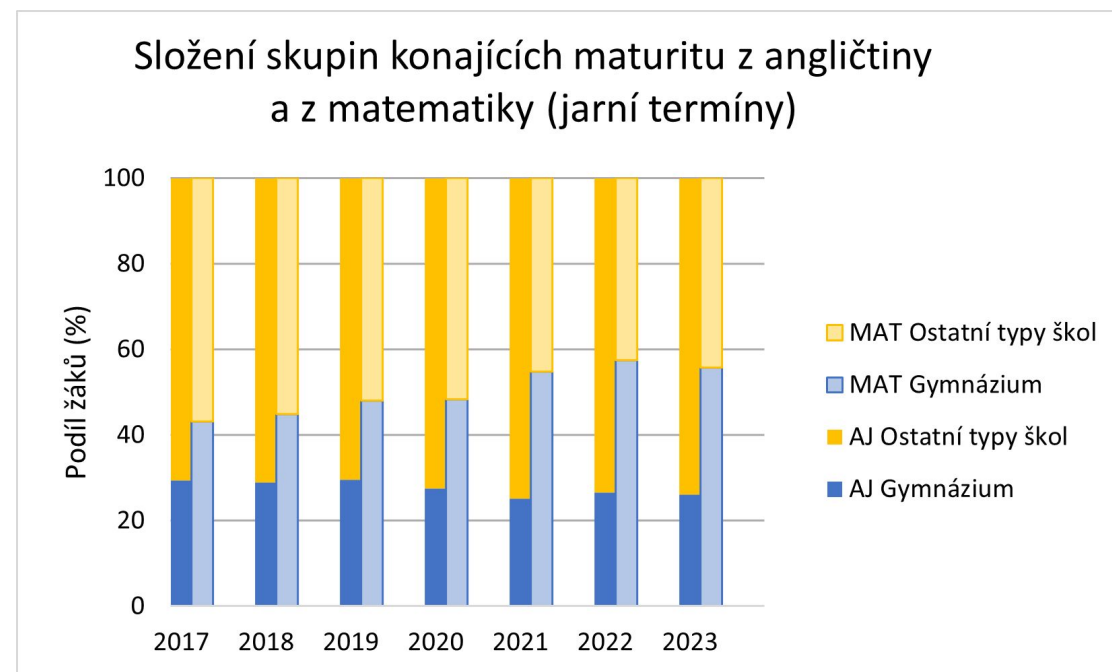
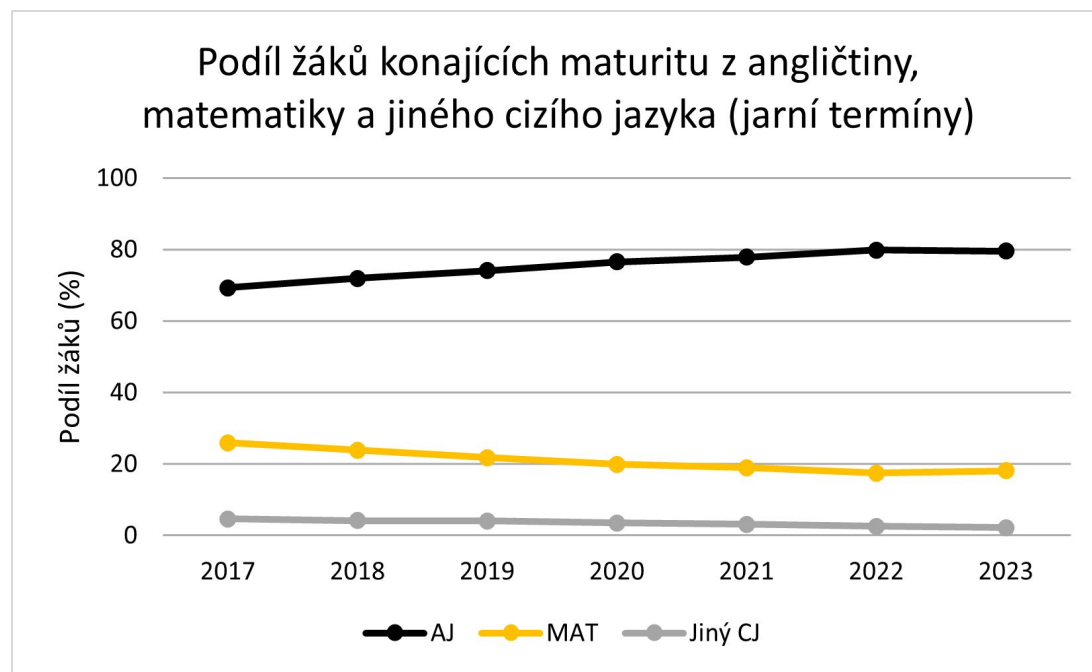
- Dle Standardů pro pedagogické a psychologické testování (AERA, APA, NCME, 2014) je třeba zajistit srovnatelnost škál všech paralelních forem testu, aby testové skóry byly tzv. „zaměnitelné“
- Paralelní formy testu – určené k měření stejného konstruktů, vytvořené podle stejné specifikace, klíčová je zde expertní role tvůrců testu
- Test equating – skupina statistických metod používaných k vyvažování (vyrovnávání) nominálně paralelních forem testu, které ve skutečnosti nemusí být perfektně paralelní
- Cílem je odlišit rozdíly v obtížnosti forem testu od rozdílů ve znalostech respondentů

Vyvažování verzí testu

- 2 hlavní přístupy k equatingu:
 - **stejně (ekvivalentní) skupiny** respondentů pro všechny formy testu – pozorované rozdíly jsou způsobeny rozdílnou obtížností testových forem
 - **stejně položky** do **všech** forem testu (tzv. *anchor test*) – pozorované rozdíly ve výsledcích *anchor testu* jsou způsobeny rozdílnými znalostmi skupin respondentů
 - zlatý standard v situacích, kdy nelze předpokládat ekvivalenci skupin

Vyvažování verzí testu – případ maturity

- možnost volby cizího jazyka nebo matematiky
 - složení žáků se může meziročně lišit



Vyvažování verzí testu – případ maturity

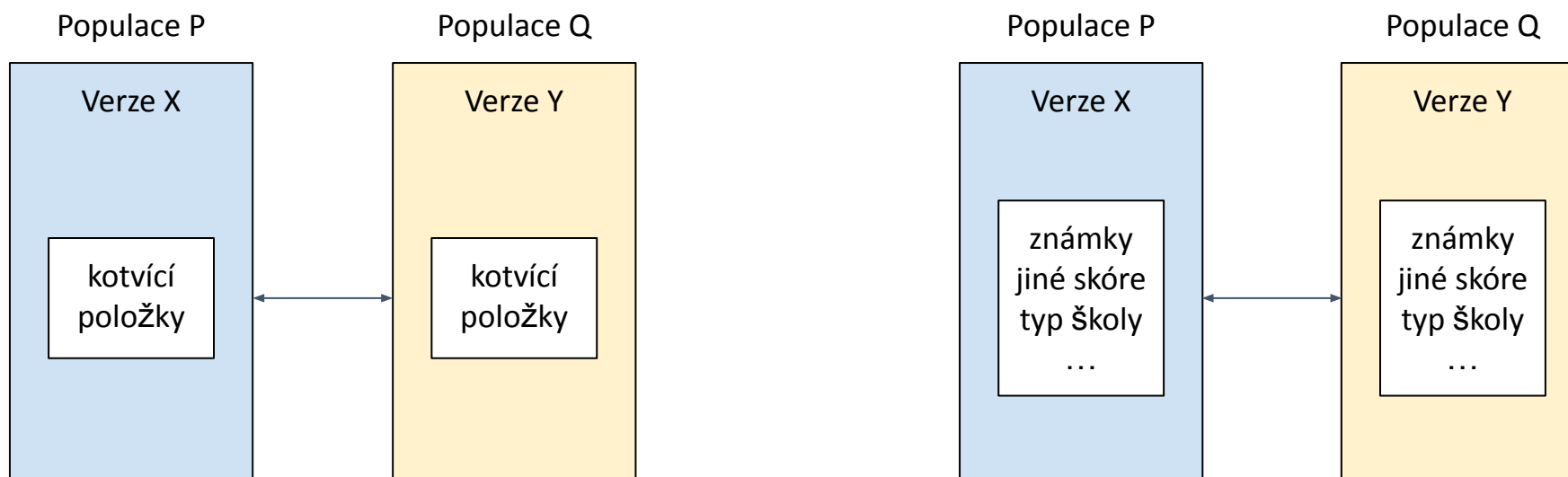
- jarní a podzimní termín
 - složení žáků se značně liší pro všechny předměty

Termín	ČJL	AJ	MAT
jaro 2017	32,14 (8,47)	71,40 (17,58)	21,89 (13,74)
podzim 2017	24,06 (6,56)	49,14 (18,25)	13,33 (8,77)
jaro 2018	31,81 (8,37)	70,22 (17,51)	21,15 (13,00)
podzim 2018	24,80 (5,70)	47,29 (18,59)	13,79 (8,22)
jaro 2019	31,64 (8,37)	77,42 (17,49)	24,15 (13,21)
podzim 2019	22,91 (6,20)	52,46 (20,06)	11,94 (7,84)

Pozn. Maximální počet bodů: ČJL = 50, AJ = 95, MAT = 50;
hranice úspěšnosti: ČJL = 22, AJ = 42, MAT = 17

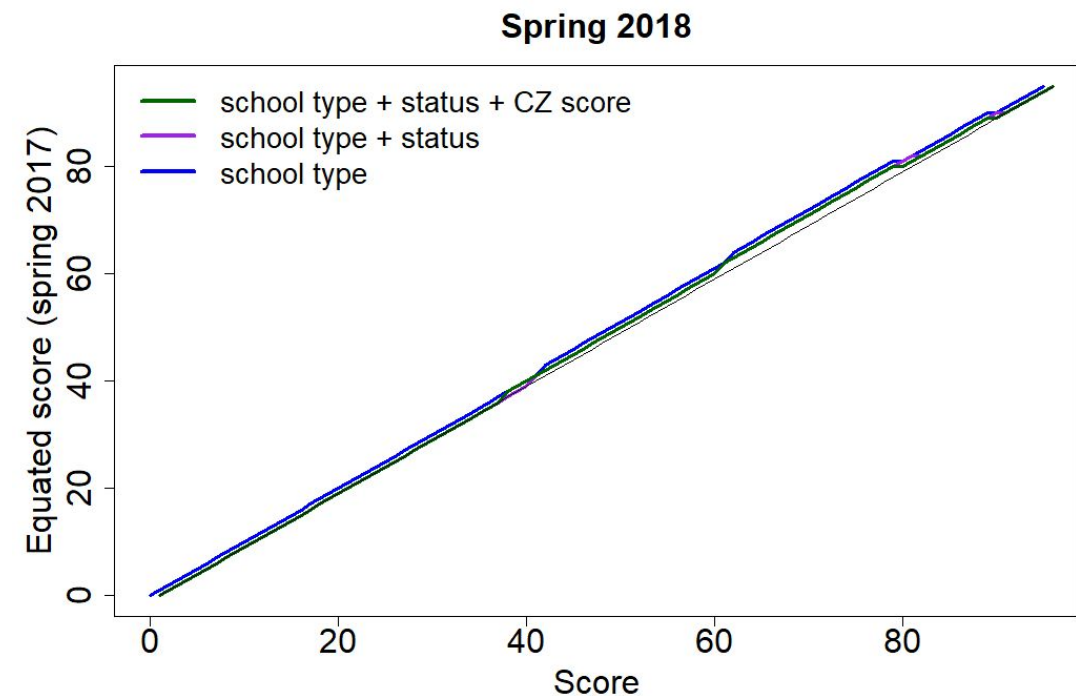
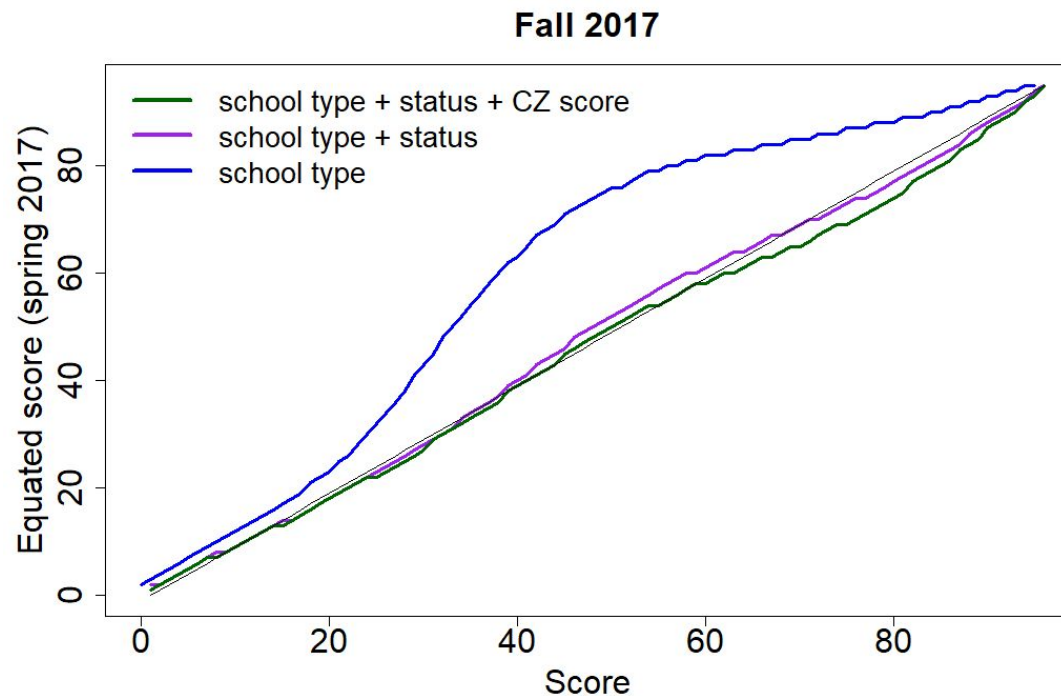
Vyvažování verzí testu – případ maturity

- Skupiny nejsou ekvivalentní a testy neobsahují společné položky
- Společné položky lze využít v pilotáži, ale znění se může změnit, navíc může být složité zajistit reprezentativní vzorek (malé cizí jazyky)
- Místo společných položek lze využít kovariáty - informace o respondentech, které mohou vysvětlit rozdíl ve znalostech mezi skupinami (př. SweSAT)



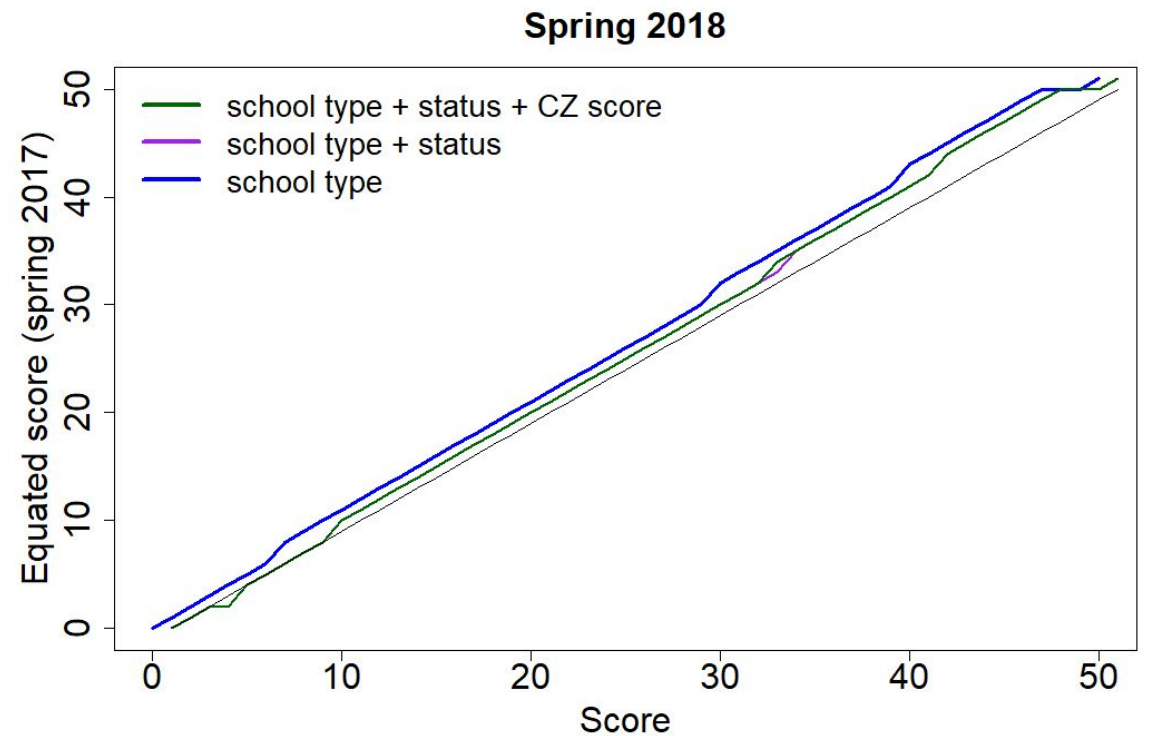
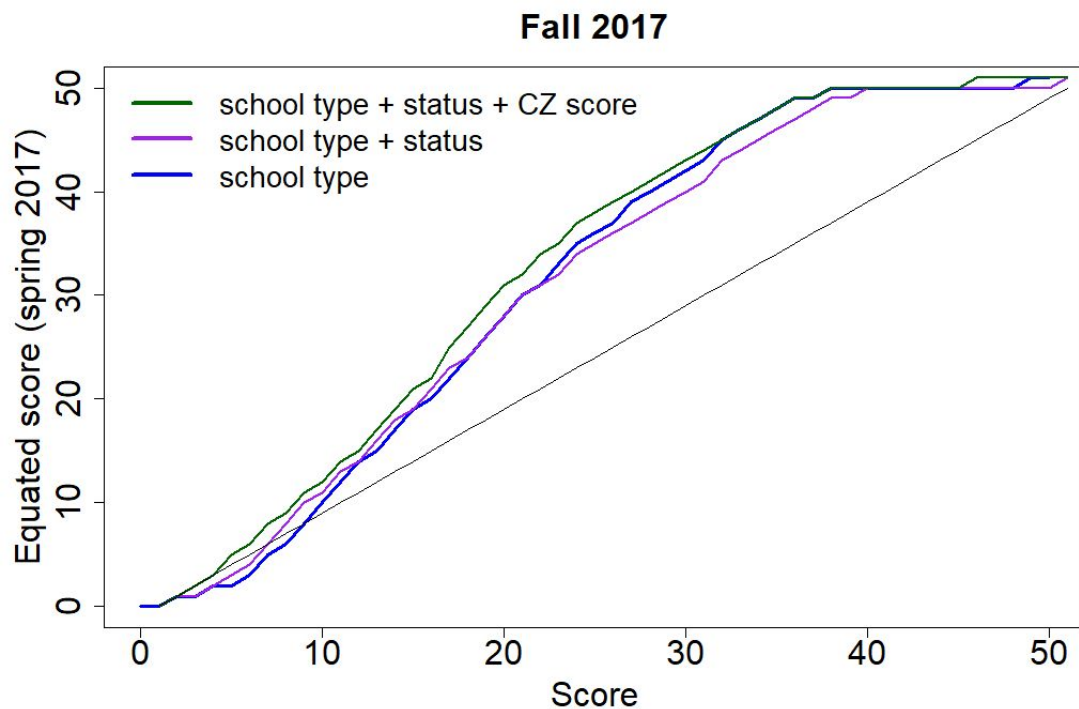
Vyvažování maturitních testů – anglický jazyk

- jaro 2017 vs. podzim 2017 a jaro 2018, různé kovariáty



Vyvažování maturitních testů – matematika

- jaro 2017 vs. podzim 2017 a jaro 2018, různé kovariáty



Vyvažování maturitních testů – závěry

- Při vyvažování testů pomocí kovariát je důležité zahrnout všechny relevantní kovariáty
- Při slabé korelaci kovariát s testovým skórem nejsme schopni s jejich pomocí dostatečně dobře vysvětlit rozdíly mezi skupinami – testové verze se mohou jevit různě obtížné, i když jsou fakticky srovnatelné
- Covid-19???
- Pokud chceme zaručit srovnatelnost testů, bez kotvících položek se do budoucna neobejdeme

Textová analýza

- Navazujeme na teoretický výzkum v oblasti využití machine learning a textové analýzy pro odhad obtížnosti položek (Štěpánek et al., 2023)
- Vycházíme ze srovnání textových charakteristik již administrovaných testových položek a jejich obtížnosti odhadnuté z žákovských odpovědí
- Pomocí machine learning algoritmů se snažíme
 - určit, které textové charakteristiky nejlépe predikují obtížnost položky
 - predikovat obtížnost nových položek
- Využití: doplnění a zpřesnění expertních odhadů tvůrců na nových položkách
- Položky na porozumění textu v cizím jazyce (části 5 a 7 maturitních testů)

Textová analýza

- 4 typy textových charakteristik
 - **počty** (vět, slov, slabik, znaků) a odvozené ukazatele (např. průměrná délka slov)
 - **běžnost/vzácnost slov** v textu s oporou o frekvenční slovníky a CEFR úrovně
 - **indexy čitelnosti textu** – většinou vycházejí z údajů o počtech a frekvencích
 - **lexikální podobnost** různých částí testové položky (např. výchozí text vs. správná odpověď, výchozí text vs. distraktory)
- Nejsilnější prediktory obtížnosti: celkový počet znaků, SD délky slova, čitelnost textu, podobnost mezi výchozím textem a distraktory
- V rámci SIA modulů ukázka určení CEFR úrovní jednotlivých slov a výpočtu základních textových charakteristik

[Data](#)[Scores ▾](#)[Validity ▾](#)[Reliability ▾](#)[Item analysis ▾](#)[Regression ▾](#)[IRT models ▾](#)[DIF/Fairness ▾](#)[Modules ▾](#)[Reports](#)[Description](#)[About the modules](#)[Modules](#)[CERMAT tools](#)[Text analysis](#)[Computerized Adaptive Tests](#)

Text analysis

Item passage:

In 2003, a 106-year-old woman from Os in Norway received an offer for free bus rides to school. Town officials sent it to her to attend school in autumn 2003. Ingeborg Thuen, born in 1897, is the oldest citizen of Os and actually started school in 1903. When the town hall of Os read the 97 of her birth year as 1997 and added her to the list of children starting the first grade in autumn 2003, she laughed and said, "Free rides are a very good idea, although now I live near the school. When I started school back then,

Question:

Why was the woman offered free bus rides? She was offered free bus rides:

Correct option:

Because of a computer mistake.

Incorrect option 1:

Because she was the oldest person in the town.

Incorrect option 2:

Because she wanted to start school.

Incorrect option 3:

Because of her hour-long walk to school every morning.

[▶ Analyze](#)

[Data](#)[Scores ▾](#)[Validity ▾](#)[Reliability ▾](#)[Item analysis ▾](#)[Regression ▾](#)[IRT models ▾](#)[DIF/Fairness ▾](#)[Modules ▾](#)[Reports](#)

CEFR level analysis

Item passage:

In 2003, a 106-year-old woman from Os in Norway received an offer for free bus rides to school. Town officials sent it to her because they thought she would first attend school in autumn 2003. Ingeborg Thuen, born in 1897, is the oldest citizen of Os and actually started school in 1903 when she was six years old. Computers in the town hall of Os read the 97 of her birth year as 1997 and added her to the list of children starting the first grade in autumn 2003. When Ingeborg received the offer, she laughed and said, Free rides are a very good idea, although now I live near the school. When I started school back then, I had to walk for an hour every morning, which was really hard. However, as I can already read, write and count, I will skip school this time, she joked.

Question:

Why was the woman offered free bus rides? She was offered free bus rides:

Correct option:

Because of a computer mistake.

Incorrect option 1:

Because she was the oldest person in the town.

Incorrect option 2:

Because she wanted to start school.

Incorrect option 3:

Because of her hour-long walk to school every morning.

CEFR legend: **A1** **A2** **B1** **B2** **C1** **C2**

Item features

Item feature	Combined text	Item passage	Question	Correct option	Incorrect option 1	Incorrect option 2	I
Number of characters	667.0	502.0	49.0	22.0	26.0	27.0	

[Data](#)[Scores ▾](#)[Validity ▾](#)[Reliability ▾](#)[Item analysis ▾](#)[Regression ▾](#)[IRT models ▾](#)[DIF/Fairness ▾](#)[Modules ▾](#)[Reports](#)

Item features

Item feature	Combined text	Item passage	Question	Correct option	Incorrect option 1	Incorrect option 2	I
Number of characters	667.0	502.0	49.0	22.0	26.0	27.0	
SD of word length (chars)	1.8	1.8	1.5	0.6	1.6	1.5	
Longest word length (chars)	12.0	12.0	7.0	8.0	7.0	7.0	
Text readability - FOG index	8.9	11.2	2.8	10.0	3.6	2.4	
Text readability - Dale-Chall index	6.1	6.7	0.3	7.0	0.4	0.3	
Text readability - Traenkle-Bailer index	-323.9	-383.3	-191.7	-287.4	-228.0	-256.2	
Text readability - SMOG index	8.8	10.0	3.1	8.8	3.1	3.1	
Number of tokens (words, numbers, interpunction)	133.0	102.0	11.0	3.0	5.0	5.0	
Number of multi-syllable words (>= 4)	0.8	1.0	0.0	0.0	0.0	0.0	
Number of monosyllabic words	59.4	61.8	72.7	0.0	40.0	60.0	

Please see the article below for more details and explanations of each item feature. Excluded words (i.e. "stopwords"): I, a, about, an, are, as, at, be, by, com, for, from, how, in, is, it, of, on, or, that, the, this, to, was, what, when, where, who, will, with, www.

References

CEFR levels based on [The Oxford 3000™ by CEFR level](#) and [The Oxford 5000™ by CEFR level](#)

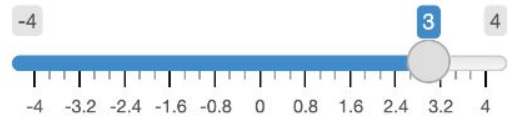
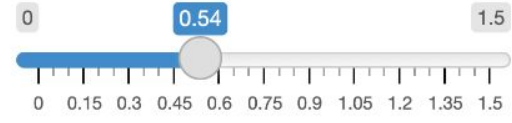
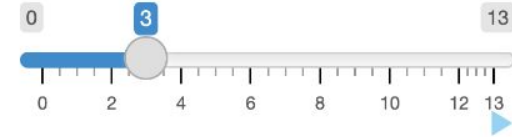
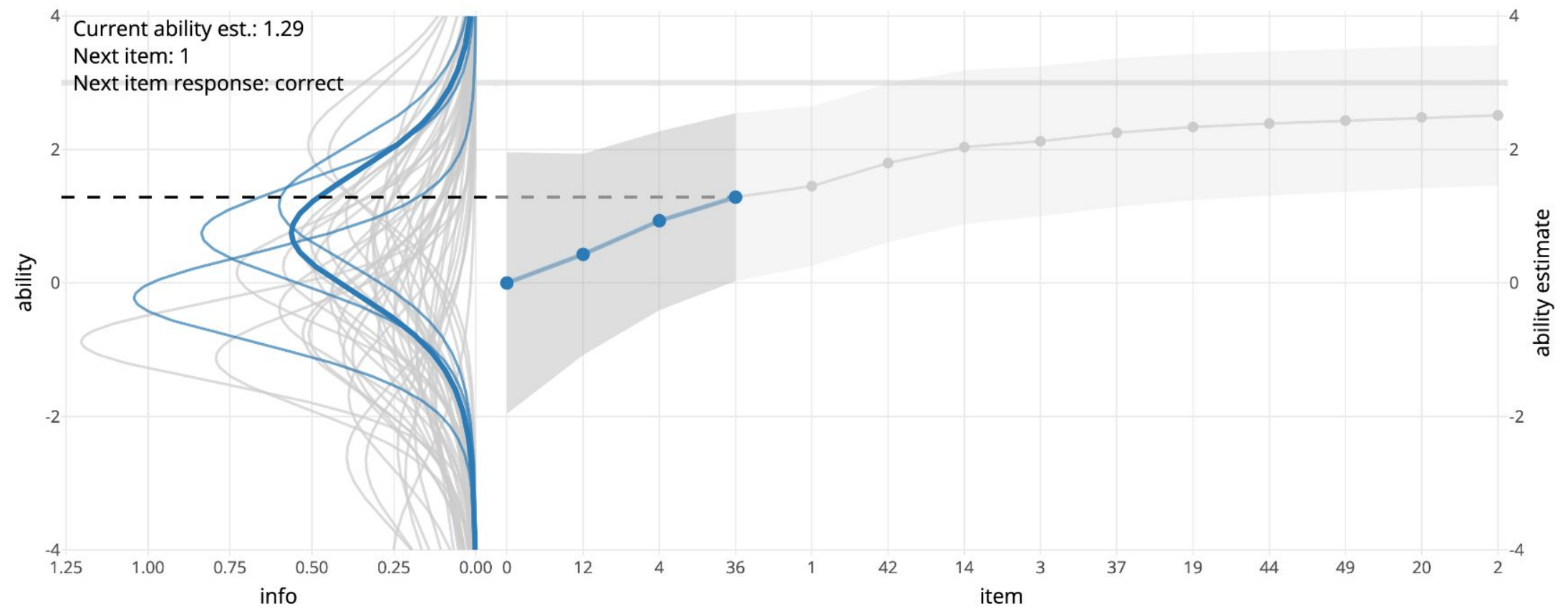
Štěpánek, L., Dlouhá, J., & Martinková, P. (2023). Item Difficulty Prediction Using Item Text Features: Comparison of Predictive Performance across Machine-Learning Algorithms. *Mathematics*, 11(19), 4104. <https://doi.org/10.3390/math11194104>

Počítačové adaptivní testování

- Využití informačních funkcí položek k adaptivní administraci
- Položka je žákovi zadána dle aktuálního odhadu znalosti (a znalosti parametrů položky)
- SIA modul „*Computerized Adaptive Tests*“
 - Tzv. post-hoc analýza, vhodná pro nastavení parametrů CATu
 - Simulované odpovědi na základě latentní znalosti a IRT modelu
- Adaptivní verze maturitního testu z matematiky
 - Využití např. pro cvičné testy (viz procvicprijimacky.cermat.cz)

[Data](#)[Scores](#) ▾[Validity](#) ▾[Reliability](#) ▾[Item analysis](#) ▾[Regression](#) ▾[IRT models](#) ▾[DIF/Fairness](#) ▾**Modules** ▾[Reports](#)

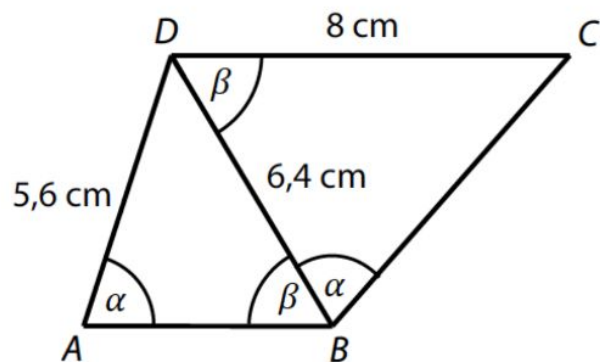
Module's example 2PL ▾

Respondent's true ability (θ)**Min. SE****CAT step** Show IICs for only administered items

9

Lichoběžník $ABCD$ je rozdělen úhlopříčkou na dva podobné trojúhelníky ABD a BDC . V trojúhelnících jsou vyznačeny dvě dvojice shodných úhlů α, β .

Platí: $|AD| = 5,6\text{ cm}$, $|BD| = 6,4\text{ cm}$, $|CD| = 8\text{ cm}$.



Rozhodněte, zda je pravdivé (A), či nikoli (N).

$$|AB| : |BD| = |BD| : |CD|$$

ANO

Obvod trojúhelníku BCD je 1, 25krát větší než obvod trojúhelníku ABD .

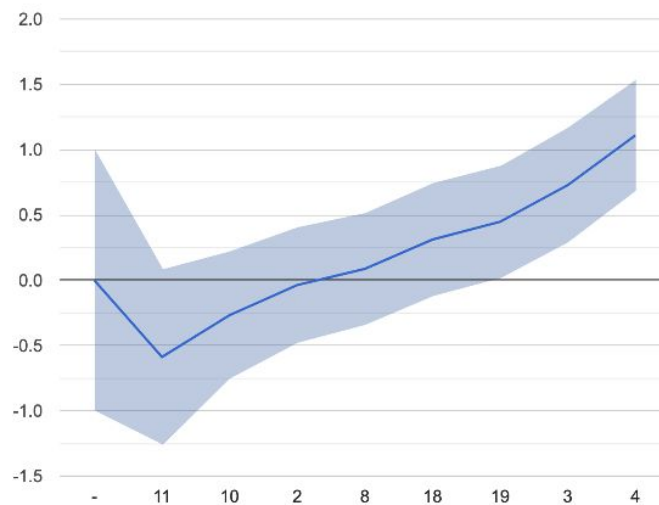
NE

$$|AB| = 5,12\text{ cm}$$

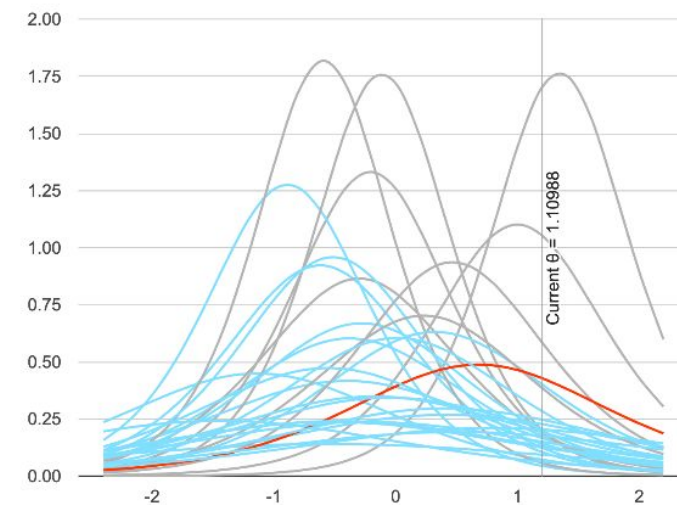
$$|BC| = 7\text{ cm}$$

ANO

Odeslat

Aktuální θ : 1.10988, Aktuální SE: 0.424789

Tato položka: 20



Praktická část

Praktická část a diskuse

- Vyzkoušejte si modul k textové analýze
 - Využijte zadání testů z angličtiny (úlohy z části 5) v online materiálech
- Vyzkoušejte si adaptivní test z matematiky
 - Modul adaptivní testování
 - Ukázková aplikace s položkami z maturitních testů
 - Vyzkoušejte různé kombinace správných a nesprávných odpovědí a sledujte, jak se mění odhad latentní znalosti

Online materiály a odkazy: cs.cas.cz/comps/workshop

- ? Řešíte v praxi vyvažování testů? Jak?
- ? Jak stanovujete obtížnost položek? Využíváte analýzu textu?
- ? Využíváte v praxi adaptivní testování?

Diskuse a závěr

Od klasické testové teorie (CTT)
po modely teorie odpovědi na položku (IRT)



Diskuse

- ? Co vše by mělo být součástí standardních výstupů z analýzy znalostních testů (např. pro odbornou komisi MŠMT)?
- ? Jaké výstupy by bylo vhodné zveřejňovat a proč?
- ? Jaké výstupy z dat CZVV byste využili a jak?
- ? Která metoda nebo nástroj Vám připadá nejdůležitější, resp. bylo by dobré ji dále rozvinout či rozšířit její využití?
- ? ...

Děkujeme za pozornost!

www.EduTest.cz

