## ITEM RESPONSE THEORY (IRT)

Item Response Theory (IRT) je sodobna psihometrična metoda, ki omogoča poglobljeno analizo odgovorov na testne postavke (Harvey & Hammer, 1999). V nasprotju s klasično testno teorijo (KTT), ki obravnava skupni rezultat testa, se IRT osredotoča na razumevanje interakcije med latentno sposobnostjo posameznika in značilnostmi posameznih postavk (De Champlain, 2010). Z uporabo IRT lahko natančneje ocenimo tako sposobnosti posameznikov kot tudi težavnost in diskriminativnost testnih postavk. V Jamoviju lahko IRT analize izvajamo z modulom snowIRT.

V Jamovi lahko IRT analizo izvedemo z modulom snowIRT (Slika 1 in Slika 2).

A screenshot of a computer

AI-generated content may be incorrect.

Slika 1: Uvoz dodatnega modula.

A screenshot of a computer

AI-generated content may be incorrect.

Slika 2: Uvoz modula IRT.

IRT modeli predpostavljajo, da je verjetnost pravilnega odgovora na postavko (ali določenega odgovora na postavko z več kategorijami) odvisna od dveh ključnih dejavnikov:

* **Sposobnosti posameznika (θ):** Latentna spremenljivka, ki predstavlja raven sposobnosti, znanja ali lastnosti, ki jo test meri.
* **Parametrov postavke:** Značilnosti posamezne postavke, ki vplivajo na verjetnost odgovora.

### Modeli

V teoriji odziva na postavke obstaja več različnih modelov, ki se razlikujejo glede na število parametrov, ki jih vključujejo, in vrsto podatkov, ki jih obravnavajo. Osnovna razlika med modeli je v tem, kako obravnavajo razmerje med sposobnostjo posameznika in značilnostmi postavk. Najenostavnejši modeli, kot je Raschov model, predpostavljajo, da je edini pomemben parameter težavnost postavke, medtem ko bolj kompleksni modeli vključujejo tudi parametre, kot sta diskriminativnost in verjetnost ugibanja. Poleg tega ločimo med dihotomskimi modeli, ki so primerni za postavke z dvema možnima odgovoroma (npr. pravilen/napačen), in politomskimi modeli, ki se uporabljajo pri ocenjevanju postavk z več kategorijami odgovorov (npr. Likertove lestvice). V nadaljevanju bomo predstavili nekaj najpogosteje uporabljenih modelov v IRT, njihove predpostavke ter primere uporabe.

**Dihotomski Rasch model** (Kreiner, 2012)je najpreprostejši model IRT, ki uporablja samo eno parameter – težavnost vprašanja. Model predpostavlja, da je verjetnost pravilnega odgovora na vprašanje odvisna samo od razlike med sposobnostjo udeleženca in težavnostjo vprašanja. Uporablja se povsod, kjer je pravilen le en odgovor, torej npr. tudi za vprašanja večkratne izbire (angl. *multiple choice)*.

**Politomski Rasch model** (Andersen, 1995) je razširitev dihotomskega modela, ki omogoča analizo vprašanj z več kot dvema možnima pravilnima odgovoroma (npr. Likertove lestvice). Model uporablja več parametrov za opis verjetnosti odgovora na vsako kategorijo.

### Zanesljivost modela in analiza postavk

V psihometriji zanesljivost pomeni stopnjo natančnosti oziroma doslednosti merjenja – torej, kako stabilno in ponovljivo test meri neko latentno lastnost. V okviru teorije odziva na postavke (IRT) se zanesljivost ne ocenjuje le na ravni celotnega testa, temveč tudi na ravni posameznih postavk in različnih ravni sposobnosti. Namesto enotne ocene, kot jo ponuja klasična testna teorija (npr. Cronbachov alfa), IRT omogoča podrobnejšo analizo informativnosti postavk in testa skozi različne metrike, ki jih predstavljamo v nadaljevanju.

Zanesljivost modela IRT se ocenjuje z različnimi metodami, kot so:

* **Informacijska funkcija vprašanja (Item Information Function - IIF)**: Meri, koliko informacij vprašanje prispeva k oceni sposobnosti udeleženca. Visoka vrednost IIF pomeni, da je postavka na tej ravni sposobnosti zelo informativna in dobro diskriminira med posamezniki.
* **Informacijska funkcija testa (Test Information Function – TIF):** Je vsota informacijskih funkcij vseh postavk in prikazuje skupno informacijo, ki jo celoten test zagotavlja za oceno sposobnosti na različnih ravneh. Višja TIF pomeni večjo zanesljivost testa. Tega izbran modul v JAMOVI ne omogoča.
* **Pearsonova zanesljivost** (Person Reliability; Anselmi idr., 2019): Predstavlja natančnost merjenja latentne sposobnosti posameznikov. Višje vrednosti kažejo na večjo stabilnost ocen, podobno kot Cronbachov alfa v klasični testni teoriji, vendar temelji na parametrih IRT. Vrednosti nad 0.8 običajno kažejo na visoko zanesljivost in so primerne za individualne diagnostične namene, medtem ko so vrednosti med 0.6 in 0.8 sprejemljive za skupinske raziskave ali primerjalne analize. Pri vrednostih pod 0.6 moramo biti previdni, saj so ocene manj stabilne in manj zanesljive.

Poleg ocene zanesljivosti na ravni celotnega testa je v teoriji odziva na postavke (IRT) ključnega pomena analiza posameznih postavk. Ta analiza omogoča vpogled v značilnosti vsakega vprašanja, kot sta njegova težavnost in sposobnost razlikovanja med posamezniki z različnimi ravnmi latentne sposobnosti. S tem zagotovimo, da test ne vsebuje prelahkih ali pretežkih vprašanj, in ocenimo, ali postavke pokrivajo celoten spekter sposobnosti, ki jih želimo meriti. V okviru Raschovega modela in povezanih IRT pristopov se za oceno lastnosti postavk pogosto uporabljata naslednja kazalnika:

* **Delež pravilnih odgovorov (Proportion correct):** Predstavlja enostaven in intuitiven pokazatelj težavnosti postavke. Višji delež pomeni lažjo postavko (več udeležencev odgovori pravilno), nižji delež pa težjo postavko. Čeprav ta indikator hitro osvetli razmere v podatkih, ne upošteva odnosa med sposobnostjo udeležencev in verjetnostjo pravilnega odgovora, kot ga definira IRT.
* **Težavnost postavke (*Measure;*** Embretson & Reise, 2013**):** V okviru Raschovega modela parameter *b* (tudi *item difficulty*) označuje raven latentne sposobnosti (θ), pri kateri ima oseba 50 % verjetnost pravilnega odgovora. V programski opremi, kot je Jamovi (modul snowIRT), je ta parameter prikazan v stolpcu Measure in izražen v logitih. Višja vrednost *Measure* pomeni težjo postavko, saj zahteva višjo sposobnost za dosego 50 % verjetnosti pravilnega odgovora, medtem ko nižje (negativne) vrednosti označujejo lažje postavke. Ocene v logitih omogočajo primerljiva in linearna merjenja, neodvisna od porazdelitve sposobnosti v vzorcu.

### Grafične predstavitve

Grafične predstavitve v teoriji odziva na postavke igrajo ključno vlogo pri interpretaciji rezultatov analize, saj omogočajo intuitivno in vizualno razumevanje odnosa med sposobnostmi preizkušancev ter značilnostmi testnih postavk. S pomočjo grafov lahko hitro prepoznamo, ali test ustrezno meri ciljno populacijo, kako informativne so posamezne postavke ter ali test pokriva celoten spekter latentne sposobnosti.

1. **Wrightova mapa** (Hilaliyah idr., 2019)**:** Gre za grafični prikaz, ki na eni lestvici (izraženi v logitih) združuje porazdelitev latentnih sposobnosti preizkušancev in težavnosti postavk. Wrightova mapa omogoča vizualno preverjanje, ali je test primerno težak glede na populacijo, ki jo merimo, in kako dobro so postavke razporejene po celotnem spektru sposobnosti. Idealen test vsebuje postavke, ki pokrivajo celoten razpon latentnih sposobnosti ciljnih udeležencev, brez prekomernih praznin ali koncentracij na posameznih ravneh. Interpretacija Wrightove mape omogoča ugotavljanje morebitnih efektov stropa in tal, ko postavke ne pokrivajo skrajnih ravni sposobnosti, kar pomeni, da test lahko podcenjuje ali precenjuje sposobnosti najbolj izstopajočih posameznikov. Prav tako je mogoče prepoznati, ali je test pretežak ali prelahk za merjeno populacijo.
2. **Karakteristična krivulja postavke (Item Characteristic Curve; ICC,** Rosenbaum, (

1987**)**: Prikazuje verjetnost pravilnega odgovora na posamezno postavko glede na raven latentne sposobnosti posameznika, izraženo na logit lestvici. Vsaka postavka ima svojo ICC, ki omogoča vpogled v več ključnih psihometričnih značilnosti: **Težavnost postavke** je odvisna od položaja krivulje na horizontalni osi. Če je krivulja pomaknjena proti levi, pomeni, da že posamezniki z nižjo sposobnostjo (negativni logiti) dosegajo visoko verjetnost pravilnega odgovora – takšne postavke so **lahke**. Če je krivulja pomaknjena proti desni, to pomeni, da je postavka **težka**, saj jo pravilno rešijo le posamezniki z visoko sposobnostjo. **Diskriminativnost postavke** je razvidna iz naklona krivulje. Strma krivulja pomeni, da postavka učinkovito ločuje med posamezniki z različnimi ravnmi sposobnosti okoli določene točke (običajno pri verjetnosti 50 %). Plitvejši naklon kaže na manjšo občutljivost postavke pri razlikovanju med preizkušanci. **Ustreznost postavke modelu** se oceni s primerjavo med pričakovano krivuljo (običajno modra, modelirana na podlagi parametrov) in opazovano krivuljo (npr. črna črta, empirično izračunana iz dejanskih podatkov). Če se opazovana krivulja dobro ujema s teoretično, to pomeni, da se vedenje postavke sklada z napovedmi IRT-modela. Večja odstopanja pa lahko kažejo na težave pri delovanju postavke, kot so slaba prilagoditev modelu, sistematični vzorci napačnih odgovorov ali posebnosti v populaciji. Skupaj to omogoča poglobljeno presojo kakovosti postavke: kako težka je, kako dobro razlikuje med posamezniki in ali se obnaša skladno z modelskimi pričakovanji. ICC je zato ključno orodje pri preverjanju delovanja posameznih postavk ter nujno pri konstrukciji, validaciji in optimizaciji merilnih instrumentov.

### Analiza skupnega rezultata

Skupni rezultat udeležencev je pomemben kazalec uspešnosti testa. Analiza skupnega rezultata vključuje:

1. **Povzetek skupnega rezultata**:
   * Povprečje, standardni odklon, minimum in maksimum skupnega rezultata.
   * Opisna statistika, ki omogoča hitro oceno uspešnosti udeležencev.
2. **Histogram skupnega rezultata**:
   * Histogram prikazuje porazdelitev skupnih rezultatov udeležencev.
   * Omogoča vizualno oceno normalnosti porazdelitve in prepoznavanje morebitnih izjemnih vrednosti.

### Praktični izzivi in omejitve IRT

Kljub številnim prednostim ima IRT tudi nekatere omejitve:

Velikost vzorca (Svetina Valdivia & Dai, 2024): IRT modeli so statistično bolj zahtevni in so manj zanesljivi pri majhnih vzorcih, saj ocene parametrov postanejo nestabilne. Priporoča se vzorec vsaj 200–500 preizkušancev, odvisno od kompleksnosti modela in števila postavk.

Pristranskost v podatkih (Michaelides, 2010): Nekontrolirana pristranskost pri izboru postavk ali udeležencev lahko vpliva na rezultate IRT analize in vodi do napačnih zaključkov.

Predpostavke modela (Reise & Rodriguez, 2016): IRT modeli temeljijo na določenih predpostavkah (npr. unidimenzionalnost, lokalna neodvisnost), ki jih je treba pred analizo preveriti. *Unidimenzionalnost* pomeni, da vse postavke v testu merijo eno samo latentno značilnost (npr. matematično sposobnost), kar je osnova za smiselno interpretacijo parametrov. *Lokalna neodvisnost* pa pomeni, da so odgovori na posamezne postavke med seboj statistično neodvisni, če je sposobnost posameznika (θ) znana – torej, da med postavkami ni dodatnih povezav, ki jih latentna spremenljivka ne pojasnjuje.

### Primer (dihotomni Rasch model)

Najprej uvozimo modul snowIRT. Ker so naši podartki [JAMOVI-ucbenik/irt\_data.csv at main · borbregant/JAMOVI-ucbenik](https://github.com/borbregant/JAMOVI-ucbenik/blob/main/irt_data.csv) le tipa 0 ali 1, bomo izbrali dihotomski model kot prikazuje Slika 3.

A screenshot of a computer

AI-generated content may be incorrect.

Slika 3: Izbor ustreznega (dihotomskega Rasch) IRT modela.

Vse spremenljivke, ki se nanašajo na vprašanja, torej V1 do V10 nesemo v okno Variables (Slika 4).

A screenshot of a computer

AI-generated content may be incorrect.

Slika 4: Izbor ustreznih opazovanih spremenljivk (postavk).

V oknu “standard scores” lahko odkljukamo (Slika 5) opcije, ki se nanašajo na deskriptivno analizo. Izrišeta se nam histogram (Slika 6) in tabela deskriptivnih informacij (Preglednica 1), ki ju ne bomo komentirali.

A screenshot of a computer screen

AI-generated content may be incorrect.

Slika 5: Izbor opcij za deskriptivno analizo.

A graph of different colored bars

AI-generated content may be incorrect.

Slika 6: Histogram rezultatov

Preglednica 1: Deskriptivne informacije.

| Summary of total score | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | | **Minimum** | | **Maximum** | | **Mean** | | **Median** | | **SD** | | **SE** | | **Skewness** | | **Kurtosis** | |
| 100 |  | 0.00 |  | 10.0 |  | 4.85 |  | 5.00 |  | 2.35 |  | 0.235 |  | -0.182 |  | 2.30 |  |
|  | | | | | | | | | | | | | | | | | |

V analizi nas najprej zanimajo podatki o samem modelu in podatki o težavnosti postavk, zato izberemo opcije, ki jih prikazuje Slika 7.

A screenshot of a computer

AI-generated content may be incorrect.

Slika 7: Zanesljivost modela in analiza postavk.

Pojavi se nam sledeči izpis (Preglednica 2 in Preglednica 3).

Preglednica 2: Pearsonova analiza zanesljivosti modela.

| Model Fit | | | |
| --- | --- | --- | --- |
|  | | **Person Reliability** | |
| scale |  | 0.649 |  |
|  | | | |

Preglednica 3: Analiza postavk modela.

| Item Statistics | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | **Proportion** | | **Measure** | |
| V1 |  | 0.830 |  | -2.014 |  |
| V2 |  | 0.690 |  | -1.023 |  |
| V3 |  | 0.700 |  | -1.084 |  |
| V4 |  | 0.640 |  | -0.733 |  |
| V5 |  | 0.560 |  | -0.298 |  |
| V6 |  | 0.410 |  | 0.487 |  |
| V7 |  | 0.370 |  | 0.703 |  |
| V8 |  | 0.290 |  | 1.162 |  |
| V9 |  | 0.180 |  | 1.924 |  |
| V10 |  | 0.180 |  | 1.924 |  |
|  | | | | | |

**Person Reliability (0.649):** Vrednost nakazuje zmerno zanesljivost merjenja sposobnosti posameznikov.

Stolpec **"Proportion"** prikazuje delež udeležencev, ki so na posamezno postavko odgovorili pravilno, kar predstavlja enostaven in intuitiven kazalnik težavnosti. Višje vrednosti (npr. 0.830) pomenijo, da je postavka relativno lahka, saj je pravilno odgovorilo večje število oseb. Nasprotno pa nižje vrednosti (npr. 0.180) nakazujejo, da gre za težjo postavko, ki jo je pravilno rešilo le malo udeležencev.

Stolpec **"Measure"** poroča o težavnosti postavke v smislu Raschovega parametra. Nižje logit vrednosti (npr. −2.014) pomenijo, da je postavka lahko dosegljiva tudi osebam z nižjo sposobnostjo, torej gre za lažjo postavko. Višje vrednosti (npr. 1.924) kažejo, da je potrebna visoka raven sposobnosti, zato gre za zahtevno postavko.

V zavihku **Plots** (Slika 8) lahko izberemo grafične predstavitve. Za dihotomni Rasch model sta ključni:

* **Wright Map:** Za prikaz porazdelitve sposobnosti in težavnosti.
* **Item Characteristic Curves (ICC):** Za vsako postavko posebej.

A screenshot of a computer

AI-generated content may be incorrect.

Slika 8: Izbor grafičnih predstavitev.

Najprej se nam izriše Wrigtova mapa (Slika 9):

A graph of a bar chart

AI-generated content may be incorrect.

Slika 9: Wrightova mapa.

**Interpretacija Wrightove mape:** Wrightova mapa grafično združuje porazdelitev sposobnosti *Respondents* (levo) in težavnosti *Items* (desno) na isti logit lestvici.

* **Razpon postavk:** Naše postavke pokrivajo dober razpon latentne spremenljivke, od zelo lahkih (V1) do zelo težkih (V9, V10).
* **Prekrivanje sposobnosti in težavnosti:** Obstaja dobro prekrivanje med razponom sposobnosti udeležencev in težavnosti postavk. To pomeni, da so za večino udeležencev na voljo postavke, ki so zanje primerno zahtevne.
* **Potencialni efekti stropa in tal:**
  + Opazimo določeno število udeležencev na zelo visokem koncu lestvice sposobnosti (nad 2.5 logita), za katere ni dovolj zahtevnih postavk. Postavki V9 in V10 sta najtežji, vendar so nekateri udeleženci sposobnejši od teh postavk. To nakazuje na **učinek stropa**, kar pomeni, da test ne omogoča polne diferenciacije med visoko sposobnimi posamezniki.
  + Podobno obstajajo udeleženci na zelo nizkem koncu lestvice sposobnosti (pod -2.5 logita), za katere ni dovolj lahkih postavk, da bi natančno ocenili njihovo minimalno sposobnost. To nakazuje na **učinek tal**.
* **Sklep:** Wrightova mapa kaže, da ima test dober razpon postavk in je v splošnem dobro usklajen s sposobnostmi udeležencev. Vendar pa izpostavlja potencialne učinke stropa in tal, kar pomeni, da test morda ni optimalno zahteven za najsposobnejše in preveč zahteven za najmanj sposobne udeležence. To je pomembna informacija za morebitno revizijo testa.

Za vsako postavko se nam izriše tudi ICC

Za postavko V1 (lahka postavka, Slika 10):

A blue line in the sky

AI-generated content may be incorrect.

Slika 10: ICC postavke V1.

Interpretacija ICC za postavko V1 (lahka):

* Težavnost: Modra krivulja je postavljena daleč na levo. Tudi udeleženci s sposobnostmi okoli -2.0 logitov imajo več kot 50% verjetnosti pravilnega odgovora. To potrjuje, da je V1 lahka postavka, kar se sklada z visokim deležem pravilnih odgovorov (0.830) in njeno pozicijo na Wrightovi mapi.
* Diskriminativnost: Krivulja kaže tipičen nagib Rasch modela, kar kaže na dobro diskriminativnost za to težavnostno raven.
* Ustreznost modela: Črna opazovana črta relativno dobro sledi modri pričakovani krivulji. To kaže, da V1 dobro ustreza pričakovanjem modela za lahko postavko; ljudje z zelo visoko sposobnostjo dosledno odgovarjajo pravilno.

Za postavko V5 (srednje težka postavka Slika 11):

A blue line on a black background

AI-generated content may be incorrect.

Slika 11: ICC postavke V5.

Interpretacija ICC za postavko V5 (srednje težka):

* Težavnost: Modra krivulja je bolj centrirana proti sredini lestvice sposobnosti (okoli 0.0 do 0.5 logita). Udeleženci s sposobnostmi okoli 0.0 logita imajo približno 60% verjetnosti pravilnega odgovora, verjetnost 50% pa doseže okoli -0.5 do 0.0. To potrjuje, da je V5 srednje težka postavka, kar se ujema z njenim deležem pravilnih odgovorov (0.560) in centralno pozicijo na Wrightovi mapi.
* Diskriminativnost: Krivulja kaže dober, strm nagib, kar pomeni, da je učinkovita pri razlikovanju med udeleženci okoli povprečne ravni sposobnosti.
* Ustreznost modela: Črna opazovana črta v splošnem dobro sledi modri pričakovani krivulji, še posebej v osrednjem razponu. To kaže na dobro ustreznost postavke V5.

Za postavko V10 (težka postavka, Slika 12):

A blue line in a black background

AI-generated content may be incorrect.

Slika 12: ICC postavke V10.

Interpretacija ICC za postavko V10 (težka):

* Težavnost: Modra krivulja je znatno premaknjena v desno. Tudi udeleženci s sposobnostmi okoli 1.0 ali 2.0 logitov imajo nizko verjetnost pravilnega odgovora (med 0.2 in 0.6). Le tisti z zelo visokimi sposobnostmi imajo dobro verjetnost. To potrjuje, da je V10 težka postavka, kar se ujema z njenim nizkim deležem pravilnih odgovorov (0.180) in visoko pozicijo na Wrightovi mapi.
* Diskriminativnost: Začetni nagib krivulje kaže na diskriminativnost za zelo visoke sposobnosti.
* Ustreznost modela: Črna opazovana črta nekoliko odstopa od modre pričakovane krivulje, še posebej pri višjih ravneh sposobnosti. Opazovani rezultati se ne dvignejo tako strmo, kot napoveduje model za višje sposobnosti. To kaže, da lahko za V10 obstaja določena neustreznost na višjem koncu spektra sposobnosti, kar pomeni, da se postavka morda ne obnaša povsem, kot je pričakovano za zelo sposobne posameznike (morda je celo težja, kot model v celoti zajame, ali pa obstaja nenavaden vzorec odgovorov za visoko sposobne ljudi). To je vredno zabeležiti kot potencialno področje za nadaljnje raziskave.

#### Vaja

Analiziraj dihotomski model <https://github.com/borbregant/JAMOVI-ucbenik/blob/main/irt_data_vaja.csv>

### Primer (politomski model)

Najprej komentirajmo navodila, ki nam jih JAMOVI pripne.

Najnižja kategorija odgovorov na vaši Likertovi lestvici (ali podobni ordinalni lestvici) kodirana z vrednostjo 0. Če bi imeli npr. lestvico od 1 do 5, bi jo morali pred analizo v Jamoviju pretvoriti v 0 do 4. Naša generirana Python koda to zahtevo že upošteva.

Ko izvedete IRT analizo v Jamoviju, bo modul **snowIRT** rezultate analize posameznikov (npr. ocenjene sposobnosti theta, standardne napake ocene, ipd.) samodejno dodal kot nove stolpce v vašo izvorno podatkovno tabelo (datasheet) na levi strani Jamovijevega vmesnika.

Uporabili bomo nabor podatkov GITHUB s stolpci P1 do P10 in vrednostmi od 0 do 4) bi lahko izvirali iz vprašalnika, ki meri določeno latentno lastnost, kjer udeleženci odgovarjajo na postavke na Likertovi lestvici.

| Summary of total score | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N** | | **Minimum** | | **Maximum** | | **Mean** | | **Median** | | **SD** | | **SE** | | **Skewness** | | **Kurtosis** | |
| 100 |  | 4.00 |  | 34.0 |  | 19.5 |  | 20.0 |  | 6.93 |  | 0.693 |  | -0.143 |  | 2.49 |  |
|  | | | | | | | | | | | | | | | | | |

Ta tabela prikazuje opisne statistike skupnega seštevka odgovorov (seštevek vseh kategorij od 0 do 4 za vsakega udeleženca) za celoten test.

* **Minimum (4.00) in Maximum (34.0):** Kažeta razpon doseženih skupnih točk. Z 10 postavkami in 5 kategorijami (0-4), je teoretični minimum 0 (10x0) in maksimum 40 (10x4). Naši podatki se gibljejo znotraj tega razpona.
* **Mean (19.52) in Median (20.0):** Skupno povprečje točk je blizu sredine teoretičnega razpona (20), kar kaže, da test ni ekstremno lahek ali težak na splošno. Mediana je zelo blizu povprečja, kar nakazuje na relativno simetrično porazdelitev.
* **SD (6.93):** Standardni odklon kaže na variabilnost skupnih točk med udeleženci.
* **Skewness (-0.143):** Rahla negativna asimetrija, kar pomeni, da je porazdelitev rahlo nagnjena proti višjim rezultatom, vendar je vrednost blizu 0, kar kaže na precej simetrično porazdelitev.
* **Kurtosis (2.49):** Vrednost pod 3 (za normalno porazdelitev) kaže na rahlo plosko porazdelitev (platikurtično), kar pomeni, da so repi porazdelitve tanjši in je manj ekstremnih vrednosti kot pri normalni porazdelitvi.

| Model Fit | | | |
| --- | --- | --- | --- |
|  | | **Person Reliability** | |
| Scale |  | 0.927 |  |
|  | | | |

Vrednost **Person Reliability (0.927)** je visoka in kaže na **odlično zanesljivost** ocenjenih sposobnosti posameznikov. To pomeni, da so ocene latentne lastnosti za vsakega udeleženca zelo natančne in stabilne, kar je ključnega pomena za individualne odločitve ali interpretacije. Ta zanesljivost je bistveno višja kot pri dihotomnem primeru (0.649), kar je pogosto pričakovano pri politomskih lestvicah z več informacijami na postavko.

| Item statistics of the rating scale model | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | **Measure** | | **S.E.Measure** | |
| P1 |  | -3.031 |  | 0.182 |  |
| P2 |  | -2.122 |  | 0.173 |  |
| P3 |  | -1.566 |  | 0.169 |  |
| P4 |  | -1.088 |  | 0.166 |  |
| P5 |  | -0.544 |  | 0.164 |  |
| P6 |  | 0.528 |  | 0.164 |  |
| P7 |  | 1.244 |  | 0.168 |  |
| P8 |  | 2.102 |  | 0.176 |  |
| P9 |  | 1.168 |  | 0.183 |  |
| P10 |  | 1.869 |  | 0.192 |  |
|  | | | | | |

* Ta tabela prikazuje ocenjene težavnosti (Measure) posameznih postavk v logit lestvici za Rating Scale Model.
  + **Interpretacija težavnosti (Measure):** Nižje (bolj negativne) vrednosti pomenijo, da je postavka lažja, višje (bolj pozitivne) vrednosti pa, da je postavka težja. Lahko jih interpretiramo kot točko na latentni lestvici, kjer je verjetnost, da posameznik odgovori v najvišjo kategorijo, ali, bolj natančno, povprečje pragov (vzporednih težavnosti) za to postavko.
  + **Vrsta postavk po težavnosti:**
    - **Najlažje postavke:** P1 (-3.031), P2 (-2.122), P3 (-1.566) – To so postavke, s katerimi se bodo udeleženci najverjetneje popolnoma strinjali ali izbrali visoke kategorije. V našem primeru bi to lahko bile trditve, ki so splošno sprejete glede inovacij.
    - **Srednje težke postavke:** P4 (-1.088), P5 (-0.544), P6 (0.528) – Te postavke bodo razločevale med udeleženci s povprečnimi sposobnostmi.
    - **Najtežje postavke:** P7 (1.244), P8 (2.102), P9 (1.168), P10 (1.869) – To so postavke, s katerimi se bodo popolnoma strinjali le udeleženci z zelo visoko latentno lastnostjo. Opazimo, da so P8 in P10 najtežji.
  + **S.E. (Standard Error):** Standardne napake ocene mer (težavnosti). Nižje vrednosti pomenijo natančnejšo oceno. Na splošno so vrednosti nizke, kar kaže na stabilne ocene.

| Delta-tau paramaterization of the partial credit model | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **tau parameters** | | | | | | | |
|  | | **1** | | **2** | | **3** | | **4** | |
| P1 |  | -28.11 |  | 6.622 |  | 9.208 |  | 12.28 |  |
| P2 |  | -4.11 |  | -0.508 |  | 0.554 |  | 4.06 |  |
| P3 |  | -3.88 |  | -1.010 |  | 0.564 |  | 4.32 |  |
| P4 |  | -29.72 |  | 7.366 |  | 9.728 |  | 12.63 |  |
| P5 |  | -4.77 |  | -0.830 |  | 1.355 |  | 4.24 |  |
| P6 |  | -3.97 |  | -0.916 |  | 0.902 |  | 3.98 |  |
| P7 |  | -4.26 |  | -0.736 |  | 1.359 |  | 3.64 |  |
| P8 |  | -4.16 |  | -0.615 |  | 1.461 |  | 3.32 |  |
| P9 |  | -3.16 |  | 0.839 |  | 2.324 |  | NaN |  |
| P10 |  | -2.95 |  | 0.531 |  | 2.419 |  | NaN |  |
|  | | | | | | | | | |

**Komentar:** Ta tabela je specifična in zelo pomembna za razumevanje delovanja politomskih postavk. Prikazuje ocenjene **težavnostne pragove (tau parametri)** med posameznimi kategorijami odgovorov **za vsako postavko posebej**.

* **τ1​ (tau 1):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 0 enaka verjetnosti izbire kategorije 1.
* **τ2​ (tau 2):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 1 enaka verjetnosti izbire kategorije 2.
* **τ3​ (tau 3):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 2 enaka verjetnosti izbire kategorije 3.
* **τ4​ (tau 4):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 3 enaka verjetnosti izbire kategorije 4.
* **Interpretacija in pomen za PCM:**
  + V Partial Credit Modelu (PCM) se pričakuje, da so ti pragovi **različni za vsako postavko**. To omogoča, da se vsaka postavka obnaša edinstveno glede prehodov med kategorijami, kar je zelo fleksibilno in realistično.
  + **Trend pragov:** Pri smiselni postavki bi morali pragovi naraščati (npr. τ1​<τ2​<τ3​<τ4​). To pomeni, da je za prehod v višjo kategorijo potrebna vedno večja sposobnost.
  + **Problematični pragovi (primeri iz rezultatov):**
    - **Ekstremne vrednosti (npr. P1, P4):** Vrednosti kot so -28.11 ali 12.28 kažejo na to, da je prehod med določenimi kategorijami za te postavke izjemno težak ali izjemno lahek, skoraj nemogoč ali vedno prisoten, kar nakazuje na težave z ustreznostjo postavke ali kategorije. Na primer, pri P1 je τ1​ izjemno nizek (-28.11), kar pomeni, da je prehod iz kategorije 0 v 1 izjemno lahek. Po drugi strani so višji pragovi (npr. τ4​ za P1 je 12.28) izjemno visoki, kar pomeni, da je za doseganje kategorije 4 potrebna ekstremno visoka sposobnost. To je posledica našega simuliranega "idealnega" obnašanja postavke P1 kot "zelo lahke".
    - **"NaN" vrednosti (P9, P10 pri τ4​):** "NaN" (Not a Number) pomeni, da se ta prag ni mogel oceniti. To se običajno zgodi, kadar v vzorcu ni dovolj podatkov za izračun prehoda med kategorijami. Najpogosteje to pomeni, da nihče (ali zelo malo ljudi) ni izbral najvišje kategorije (v tem primeru kategorije 4), zato model ne more določiti točke prehoda iz kategorije 3 v 4. To je pomemben znak, da je ta postavka pretežka ali pa najvišja kategorija ni dosegljiva za večino populacije.
    - **Neurejeni pragovi (npr. P1: 6.622, 9.208, 12.28):** Tudi če so vrednosti veljavne, je pomembno preveriti, ali so **monotono naraščajoče** (tj. τ1​<τ2​<τ3​<τ4​). Če temu ni tako (npr. τ1​ je višji od τ2​), to lahko pomeni, da kategorije niso dobro diferencirane ali da je postavka problematična. V vaših rezultatih, razen ekstremov pri P1 in P4, se večinoma zdijo urejeni, vendar je to ključna točka za preverjanje.

**Wright Map**

A blue and white graph

AI-generated content may be incorrect.

Na osi vidimo posameznike in nato oznake postavk, ki so ločene tudi s simboli za vsak prehod med kategorijami (npr. P1:0|1, P1:1|2, P1:2|3, P1:3|4). Idealno bi si želeli, da so vsi pragovi lepo razporejeni po celotnem latentnem kontinuumu, in da se prekrivajo s porazdelitvijo sposobnosti udeležencev.

Kot pri dihotomnem modelu, če so "Respondents" skoncentrirani na določenem delu lestvice, kjer ni ustreznih pragov ali postavk, lahko to nakazuje na učinke stropa/tal ali pomanjkanje informacij v tem delu lestvice. Iz vaše slike je videti, da so naši simulirani podatki precej dobro porazdeljeni, vendar so P9 in P10 spet na skrajnem desnem koncu, kar kaže na njihovo visoko težavnost. Ekstremne vrednosti tau parametrov se prav tako odražajo na mapi.

**Person-item map for PCM**

A black screen with white lines

AI-generated content may be incorrect.

 **Komentar:** Wrightova mapa za Partial Credit Model prikazuje porazdelitev sposobnosti posameznikov in vse posamezne težavnostne pragove za vsako postavko.

 **Razlika od dihotomne:** Pri dihotomnih modelih vsaka postavka prispeva le eno točko (težavnost). Pri PCM pa vsaka postavka prispeva več točk – po eno za vsak prag med kategorijami (npr. 4 točke za lestvico 0-4). To omogoča podrobnejši vpogled v to, kako postavka deluje na različnih ravneh sposobnosti.

 **Interpretacija:** Preverjamo, ali so pragovi lepo razporejeni po lestvici in ali se ujemajo z razponom sposobnosti udeležencev. Idealno bi si želeli, da so pragovi dobro porazdeljeni, da zajamejo celoten spekter sposobnosti in da ni velikih "lukenj" brez informacij. Ponovno so jasno vidni ekstremni pragovi in nekatere težave (npr. P9 in P10 so na skrajni desni, kar potrjuje, da so težke in da najvišje kategorije morda niso dosegljive).

**Expected Score Curve**

**Komentar: Namesto verjetnosti pravilnega odgovora (kot pri dihotomnem modelu), ICC pri politomnih modelih prikazuje pričakovani rezultat na postavki kot funkcijo sposobnosti posameznika. Zato se krivulja dviga od najnižje do najvišje kategorije (npr. od 0 do 4).**

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

A blue line in the sky

AI-generated content may be incorrect.

**Rating Scale Deltas/thresholds**

**Histogram of total score**

A graph of red and blue bars

AI-generated content may be incorrect.

### Diferencična funkcija itemov

Diferencična funkcija itemov (DIF) se nanaša na situacijo, ko imajo itemi različne lastnosti za različne skupine udeležencev, kljub temu da imajo te skupine enako raven latentnega konstrukta (npr. znanja ali sposobnosti). DIF lahko nastane zaradi različnih dejavnikov, kot so kulturni vplivi, spolne razlike ali druge sistemske pristranskosti.

**Metode za prepoznavanje DIF**

Rajuova metoda (Raju's Area Method): Primerja površine med informacijskimi funkcijami odgovora na postavko (IRF) za dve skupini. Če se IRF bistveno razlikujeta, to kaže na prisotnost DIF.

Logistična regresija: Uporablja se za preverjanje, ali skupinska pripadnost pomembno napoveduje pravilni odgovor, ko je kontrolirana sposobnost posameznika.

**Primer testa za Raju DIF analizo** [**JAMOVI-ucbenik/irt\_dif\_data.csv at main · borbregant/JAMOVI-ucbenik**](https://github.com/borbregant/JAMOVI-ucbenik/blob/main/irt_dif_data.csv)

**Generirani podatki (irt\_dif\_data.csv) simulirajo scenarij, kjer test ocenjuje določeno znanje (npr. iz naravoslovja), udeleženci pa so razdeljeni v dve skupini, recimo glede na to, ali so referenčna skupina (0) domači govorci in fokalna skupina (1) učenci, za katere slovenščina ni materni jezik.**

**Scenarij:**

**Predstavljajte si, da so ti podatki zbrani na testu znanja iz fizike za srednješolce. Test vsebuje 10 dihotomnih postavk (pravilno/napačno). Učenci so razdeljeni v dve skupini:**

* **Skupina 0 (Referenčna): Domači govorci slovenskega jezika.**
* **Skupina 1 (Fokalna): Učenci, ki se slovenščine učijo kot drugi jezik.**

**Postavke testa:**

* **DIF\_Q1 - DIF\_Q4, DIF\_Q6 - DIF\_Q10 (Postavke brez DIF-a): To so splošne postavke iz fizike, ki merijo fizikalno znanje in se predvideva, da delujejo enako za obe skupini, ne glede na jezikovno ozadje. Npr. "Katera sila deluje na telo v prostem padu?"**
* **DIF\_Q5 (Postavka s simuliranim DIF-om): Ta postavka je bila umetno "otežena" za fokalno skupino (učence, za katere slovenščina ni materni jezik). To bi lahko bila postavka, ki vsebuje kompleksno besedilo, idiome, ali kulturno specifične reference, ki so domačim govorcem lažje razumljive, ne glede na njihovo dejansko fizikalno znanje. Na primer:**
  + **"V naslednjem odstavku je opisana situacija. Previdno preberi in določi, katera fizikalna veličina je ključna za razumevanje pojava: 'V delavnici je mojster koval in brusil kovino, pri čemer so nastajali oblaki dima in vročina je hitro napolnila prostor, kar je povzročalo neudobje prisotnim.'" (Vprašanje se v bistvu nanaša na prenos toplote, a je ovito v kompleksno, morda za tuje govorce nerazumljivo, besedilo, kar otežuje razumevanje naloge, ne fizike same).**

**V jamovi dobimo**

| **Raju’s area method** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Statistic** | | **p** | | **Adj.p** | | **deltaRaju** | | **Effect size** | |
| **DIF\_Q1** |  | **0.656** |  | **0.512** |  | **0.670** |  | **-0.4022** |  | **A** |  |
| **DIF\_Q2** |  | **0.937** |  | **0.349** |  | **0.614** |  | **-0.5324** |  | **A** |  |
| **DIF\_Q3** |  | **0.178** |  | **0.859** |  | **0.859** |  | **-0.0887** |  | **A** |  |
| **DIF\_Q4** |  | **0.619** |  | **0.536** |  | **0.670** |  | **-0.2970** |  | **A** |  |
| **DIF\_Q5** |  | **4.223** |  | **< .001** |  | **< .001** |  | **-2.0130** |  | **C** |  |
| **DIF\_Q6** |  | **-2.011** |  | **1.956** |  | **0.148** |  | **0.9360** |  | **A** |  |
| **DIF\_Q7** |  | **-1.025** |  | **1.695** |  | **0.614** |  | **0.4894** |  | **A** |  |
| **DIF\_Q8** |  | **-0.473** |  | **1.364** |  | **0.707** |  | **0.2421** |  | **A** |  |
| **DIF\_Q9** |  | **-2.048** |  | **1.959** |  | **0.148** |  | **1.1017** |  | **B** |  |
| **DIF\_Q10** |  | **-0.900** |  | **1.632** |  | **0.614** |  | **0.5642** |  | **A** |  |
| **Note. 1. Effect size(ETS Delta scale) for absolute values of 'deltaRaju' = A: negligible effect(<1.0), B: moderate effect(>1.0),C: large effect(>1.5).** | | | | | | | | | | | |
|  | | | | | | | | | | | |

**Razlaga stolpcev:**

* **Statistic: To je vrednost Raju Z statistike, ki meri, kako se krivulja značilnosti postavke razlikuje med referenčno in fokalno skupino. Večja kot je absolutna vrednost, večja je razlika.**
* **p: To je neprilagojena p-vrednost, ki kaže statistično značilnost DIF-a za posamezno postavko. Če je p < 0.05 (ali izbrano raven značilnosti), postavka izkazuje statistično značilen DIF.**
* **Adj.p: To je prilagojena p-vrednost (običajno korigirana za množične primerjave, npr. z Bonferronijevo korekcijo), ki zmanjšuje verjetnost napake tipa I (lažno pozitivnega DIF-a). To je pomembnejša p-vrednost za odločanje o prisotnosti DIF-a.**
* **deltaRaju: To je mera velikosti DIF-a na lestvici Delta, ki je standardizirana mera razlike med skupinskima krivuljama. Pozitivna vrednost pomeni, da je postavka lažja za fokalno skupino (kodirano kot 1) v primerjavi z referenčno skupino (kodirano kot 0). Negativna vrednost pomeni, da je postavka težja za fokalno skupino.**
* **Effect size: To je kvalitativna kategorizacija velikosti DIF-a (A, B, C) na podlagi absolutne vrednosti deltaRaju, kot je navedeno v opombi pod tabelo.**
  + **A: zanemarljiv učinek (< 1.0)**
  + **B: zmeren učinek (med 1.0 in 1.5)**
  + **C: velik učinek (> 1.5)**

**Analiza rezultatov:**

1. **Postavka DIF\_Q5: Izražen DIF**
   * **Statistic = 4.223: Visoka vrednost Z-statistike kaže na znatno razliko med skupinama za to postavko.**
   * **p < .001 in Adj.p < .001: Obe p-vrednosti sta izjemno nizki, kar pomeni, da je prisotnost DIF-a za to postavko statistično izjemno značilna. Zelo verjetno je, da ta postavka ne deluje enako za obe skupini.**
   * **deltaRaju = -2.0130: Negativna vrednost kaže, da je postavka DIF\_Q5 težja za fokalno skupino (skupina 1) v primerjavi z referenčno skupino (skupina 0), ob upoštevanju enake sposobnosti.**
   * **Effect size = C: Kategorija "C" pomeni, da gre za velik učinek DIF-a. To potrjuje naše predhodno simuliranje DIF-a za to postavko. V kontekstu našega primera (test fizike, kjer je fokalna skupina "ne-materni govorci"), to pomeni, da je ta postavka izrazito težja za učence, za katere slovenščina ni materni jezik, kar je lahko posledica jezikovne kompleksnosti ali kulturne pristranskosti postavke.**
2. **Postavka DIF\_Q9: Zmeren DIF, vendar ni statistično značilen po prilagojeni p-vrednosti**
   * **Statistic = -2.048: Relativno visoka absolutna vrednost Z-statistike.**
   * **p = 1.959: To je verjetno napaka v izpisu, saj p-vrednost ne more biti večja od 1. Predvidevam, da bi morala biti p-vrednost okoli 0.04 (če bi bila absolutna Z vrednost 2.048). Vendar je Adj.p = 0.148.**
   * **Adj.p = 0.148: Prilagojena p-vrednost (0.148) je višja od običajne ravni značilnosti (npr. 0.05), kar pomeni, da DIF za to postavko ni statistično značilen po Bonferronijevi ali podobni korekciji. Kljub temu, da bi bil morda mejno značilen brez korekcije, je pri večkratnih testih pomembna prilagojena p-vrednost.**
   * **deltaRaju = 1.1017: Pozitivna vrednost kaže, da je postavka DIF\_Q9 lažja za fokalno skupino (skupina 1) v primerjavi z referenčno skupino.**
   * **Effect size = B: To kaže na zmeren učinek DIF-a.**
   * **Sklep: Čeprav je velikost učinka zmerna, statistična značilnost ni potrjena po prilagoditvi za večkratne primerjave, kar bi nas moralo previdno pri interpretaciji. Možno je, da je prisoten nek trend, vendar ni dovolj močan za dokončno potrditev DIF-a pri tej velikosti vzorca.**
3. **Vse ostale postavke (DIF\_Q1, DIF\_Q2, DIF\_Q3, DIF\_Q4, DIF\_Q6, DIF\_Q7, DIF\_Q8, DIF\_Q10): Ni značilnega DIF-a**
   * **Vse te postavke imajo p in Adj.p vrednosti, ki so višje od 0.05, kar pomeni, da ni statistično značilnega DIF-a.**
   * **Njihove deltaRaju vrednosti so blizu 0 in Effect size je "A" (zanemarljiv učinek), kar potrjuje, da se te postavke med skupinama obnašajo zelo podobno. To se ujema z našo simulacijo, kjer smo le za eno postavko (DIF\_Q5) umetno ustvarili DIF.**

#### Nadaljnje možnosti DIF analize v Jamoviju

**Modul snowIRT v Jamoviju ponuja poleg Raju metode še več naprednih metod za odkrivanje DIF-a, ki omogočajo podrobnejšo in specifičnejšo analizo diferenciranega delovanja postavk. Izbira metode je odvisna od tipa podatkov (dihotomni, ordinalni) in specifičnih raziskovalnih vprašanj.**

**4.Z.1 Delta Metoda (Delta Method)**

* **Opis: Delta metoda je robustna in relativno enostavna neparametrična metoda za odkrivanje uniformnega DIF-a. Osredotoča se na razliko v deležih pravilnih odgovorov med skupinama za posamezno postavko, ki so transformirani v Z-vrednosti. Metoda je primerna za dihotomne postavke. V Jamoviju se pogosto uporablja kot preliminarna metoda za hitro prepoznavanje potencialnih DIF postavk, saj ne zahteva striktnih predpostavk o modelu IRT.**
* **Kdaj jo uporabiti: Kadar imamo dihotomne podatke in želimo hitro in pregledno preveriti prisotnost uniformnega DIF-a brez potrebe po predhodni specifikaciji in ocenjevanju kompleksnega IRT modela. Je odlična izbira za začetno presejalno analizo.**
* **V Jamoviju: V modulu snowIRT pri izbiri DIF analize izberete Delta Method. Rezultati so v enaki obliki kot Raju metodi**

**4.Z.2 DIF za ordinalne podatke (DIF for Ordinal Data)**

* **Opis: Medtem ko so Raju metoda in Delta metoda primarno namenjeni dihotomnim podatkom, modul snowIRT omogoča tudi analizo DIF-a za ordinalne podatke, kot so Likertove lestvice. Pri ordinalnih podatkih je DIF lahko bolj kompleksen, saj lahko vpliva na več prehodov med kategorijami. Metode za ordinalne podatke preverjajo, ali se pragovi med kategorijami obnašajo drugače med skupinami.**
* **Kdaj jo uporabiti: Ko imate postavke z več kot dvema kategorijama odgovorov (npr. 3-stopenjska, 5-stopenjska lestvica ipd.) in sumite na diferencirano delovanje postavk med skupinami.**
* **V Jamoviju: ...**

**4.Z.3 Conditional Likelihood Ratio Test (CLRT)**

* **Opis: Conditional Likelihood Ratio Test je parametriziran pristop k odkrivanju DIF-a, ki temelji na primerjavi fit modelov. Preizkuša, ali se osnovni model (brez DIF-a) statistično pomembno razlikuje od modela, ki vključuje parametre za DIF. Ta metoda se lahko uporablja za dihotomne in politomne podatke, ter omogoča prepoznavanje uniformnega in neuniformnega DIF-a (slednje, če se testirajo razlike v diskriminaciji poleg težavnosti). CLRT velja za eno bolj robustnih in močnih metod za odkrivanje DIF-a, saj ne temelji na združevanju podatkov, temveč na pogojni verjetnosti.**
* **Kdaj jo uporabiti: Kadar želite statistično močno in formalno testiranje DIF-a, in ko ste že ocenili ustrezen IRT model (npr. Rasch, 2PL, Rating Scale ali Partial Credit). Je posebej primeren za prepoznavanje tako uniformnega kot neuniformnega DIF-a.**
* **V Jamoviju: Po izbiri ustreznega IRT modela (dihotomnega ali politomnega) lahko v zavihku DIF analize izberete Conditional Likelihood Ratio Test. Rezultati bodo vključevali χ2 statistiko, stopinje prostosti in p-vrednost za vsako postavko, kar vam bo omogočilo odločitev o prisotnosti DIF-a.**

### Analiza distraktorjev pri večkratnih izbirah

Pri analizi večkratnih izbir je pomembno preučiti ne le pravilne odgovore, temveč tudi napačne odgovore (distraktorje). Distraktorji so ključni za oceno kakovosti itemov, saj pomagajo ugotoviti, ali so napačni odgovori privlačni za udeležence z nižjimi sposobnostmi.

**Postopek analize distraktorjev**

1. **Frekvenca distraktorjev**:
   * Izračunamo delež udeležencev, ki so izbrali posamezen distraktor.
   * Dobri distraktorji bi morali privlačiti udeležence z nižjimi sposobnostmi.
2. **Učinkovitost distraktorjev**:
   * Preverimo, ali so distraktorji učinkoviti tako, da analiziramo odnose med izbiro distraktorja in sposobnostjo udeleženca.
   * Učinkovit distraktor bi moral biti pogosteje izbran pri udeležencih z nižjimi sposobnostmi.
3. **Grafična predstavitev**:
   * **Distraktor Characteristic Curve (DCC)**: Graf, ki prikazuje verjetnost izbire distraktorja glede na sposobnost udeleženca.
   * DCC pomaga vizualno oceniti učinkovitost distraktorjev.

Uvozimo [JAMOVI-ucbenik/irt\_distractor\_data.csv at main · borbregant/JAMOVI-ucbenik](https://github.com/borbregant/JAMOVI-ucbenik/blob/main/irt_distractor_data.csv) obklujukamo za ketera vprašanja želimo podatke v distractor plot in dobimo  
A screenshot of a computer

AI-generated content may be incorrect.

**Counts of respondents**

| Item Q1 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 11 |  | 2 |  | 1 |  |
| B |  | 13 |  | 5 |  | 1 |  |
| C |  | 17 |  | 7 |  | 1 |  |
| \*D |  | 71 |  | 90 |  | 81 |  |
|  | | | | | | | |

| Item Q2 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 9 |  | 4 |  | 1 |  |
| B |  | 12 |  | 2 |  | 3 |  |
| \*C |  | 63 |  | 96 |  | 78 |  |
| D |  | 28 |  | 2 |  | 2 |  |
|  | | | | | | | |

| Item Q3 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 57 |  | 75 |  | 81 |  |
| B |  | 32 |  | 15 |  | 2 |  |
| C |  | 11 |  | 9 |  | 1 |  |
| D |  | 12 |  | 5 |  | 0 |  |
|  | | | | | | | |

| Item Q4 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 16 |  | 3 |  | 3 |  |
| B |  | 12 |  | 7 |  | 3 |  |
| C |  | 33 |  | 17 |  | 3 |  |
| \*D |  | 51 |  | 77 |  | 75 |  |
|  | | | | | | | |

| Item Q5 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 11 |  | 7 |  | 1 |  |
| B |  | 19 |  | 4 |  | 1 |  |
| C |  | 26 |  | 16 |  | 5 |  |
| \*D |  | 56 |  | 77 |  | 77 |  |
|  | | | | | | | |

| Item Q6 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 31 |  | 79 |  | 77 |  |
| B |  | 39 |  | 12 |  | 2 |  |
| C |  | 20 |  | 12 |  | 4 |  |
| D |  | 22 |  | 1 |  | 1 |  |
|  | | | | | | | |

| Item Q7 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 15 |  | 8 |  | 2 |  |
| B |  | 14 |  | 11 |  | 0 |  |
| C |  | 37 |  | 10 |  | 4 |  |
| \*D |  | 46 |  | 75 |  | 78 |  |
|  | | | | | | | |

| Item Q8 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 30 |  | 24 |  | 4 |  |
| \*B |  | 34 |  | 57 |  | 70 |  |
| C |  | 30 |  | 10 |  | 6 |  |
| D |  | 18 |  | 13 |  | 4 |  |
|  | | | | | | | |

| Item Q9 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 16 |  | 10 |  | 6 |  |
| B |  | 15 |  | 12 |  | 3 |  |
| C |  | 43 |  | 17 |  | 6 |  |
| \*D |  | 38 |  | 65 |  | 69 |  |
|  | | | | | | | |

| Item Q10 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 30 |  | 53 |  | 64 |  |
| B |  | 38 |  | 22 |  | 8 |  |
| C |  | 22 |  | 16 |  | 8 |  |
| D |  | 22 |  | 13 |  | 4 |  |
|  | | | | | | | |

| Item Q11 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 25 |  | 53 |  | 68 |  |
| B |  | 41 |  | 22 |  | 11 |  |
| C |  | 24 |  | 17 |  | 3 |  |
| D |  | 22 |  | 12 |  | 2 |  |
|  | | | | | | | |

| Item Q12 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 25 |  | 16 |  | 5 |  |
| B |  | 28 |  | 14 |  | 6 |  |
| C |  | 38 |  | 25 |  | 10 |  |
| \*D |  | 21 |  | 49 |  | 63 |  |
|  | | | | | | | |

| Item Q13 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 20 |  | 16 |  | 3 |  |
| B |  | 20 |  | 11 |  | 1 |  |
| \*C |  | 22 |  | 51 |  | 71 |  |
| D |  | 50 |  | 26 |  | 9 |  |
|  | | | | | | | |

| Item Q14 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 27 |  | 19 |  | 6 |  |
| B |  | 27 |  | 18 |  | 8 |  |
| \*C |  | 15 |  | 36 |  | 55 |  |
| D |  | 43 |  | 31 |  | 15 |  |
|  | | | | | | | |

| Item Q15 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 49 |  | 30 |  | 20 |  |
| \*B |  | 12 |  | 43 |  | 56 |  |
| C |  | 23 |  | 16 |  | 5 |  |
| D |  | 28 |  | 15 |  | 3 |  |
|  | | | | | | | |

| Item Q16 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 30 |  | 19 |  | 9 |  |
| B |  | 17 |  | 21 |  | 15 |  |
| \*C |  | 19 |  | 32 |  | 34 |  |
| D |  | 46 |  | 32 |  | 26 |  |
|  | | | | | | | |

| Item Q17 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 20 |  | 21 |  | 10 |  |
| B |  | 27 |  | 18 |  | 4 |  |
| C |  | 50 |  | 36 |  | 22 |  |
| \*D |  | 15 |  | 29 |  | 48 |  |
|  | | | | | | | |

| Item Q18 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 15 |  | 22 |  | 39 |  |
| B |  | 58 |  | 46 |  | 25 |  |
| C |  | 21 |  | 14 |  | 11 |  |
| D |  | 18 |  | 22 |  | 9 |  |
|  | | | | | | | |

| Item Q19 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| \*A |  | 9 |  | 26 |  | 40 |  |
| B |  | 53 |  | 43 |  | 24 |  |
| C |  | 24 |  | 18 |  | 8 |  |
| D |  | 26 |  | 17 |  | 12 |  |
|  | | | | | | | |

| Item Q20 | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **lower** | | **middle** | | **upper** | |
| A |  | 45 |  | 45 |  | 30 |  |
| \*B |  | 10 |  | 14 |  | 34 |  |
| C |  | 32 |  | 25 |  | 8 |  |
| D |  | 25 |  | 20 |  | 12 |  |
|  | | | | | | | |

| Item difficulty and discrimination index | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | | **Difficulty** | | **ULI** | | **RIT** | | **RIR** | |
| Q1 |  | 0.807 |  | 0.320 |  | 0.372 |  | 0.284 |  |
| Q2 |  | 0.790 |  | 0.400 |  | 0.457 |  | 0.372 |  |
| Q3 |  | 0.710 |  | 0.420 |  | 0.420 |  | 0.322 |  |
| Q4 |  | 0.677 |  | 0.430 |  | 0.443 |  | 0.343 |  |
| Q5 |  | 0.700 |  | 0.450 |  | 0.439 |  | 0.342 |  |
| Q6 |  | 0.623 |  | 0.620 |  | 0.532 |  | 0.438 |  |
| Q7 |  | 0.663 |  | 0.490 |  | 0.449 |  | 0.349 |  |
| Q8 |  | 0.537 |  | 0.560 |  | 0.480 |  | 0.378 |  |
| Q9 |  | 0.573 |  | 0.510 |  | 0.461 |  | 0.358 |  |
| Q10 |  | 0.490 |  | 0.510 |  | 0.441 |  | 0.334 |  |
| Q11 |  | 0.487 |  | 0.580 |  | 0.517 |  | 0.418 |  |
| Q12 |  | 0.443 |  | 0.530 |  | 0.480 |  | 0.378 |  |
| Q13 |  | 0.480 |  | 0.640 |  | 0.541 |  | 0.446 |  |
| Q14 |  | 0.353 |  | 0.470 |  | 0.421 |  | 0.318 |  |
| Q15 |  | 0.370 |  | 0.540 |  | 0.509 |  | 0.413 |  |
| Q16 |  | 0.283 |  | 0.230 |  | 0.243 |  | 0.135 |  |
| Q17 |  | 0.307 |  | 0.360 |  | 0.403 |  | 0.302 |  |
| Q18 |  | 0.253 |  | 0.310 |  | 0.344 |  | 0.245 |  |
| Q19 |  | 0.250 |  | 0.410 |  | 0.421 |  | 0.328 |  |
| Q20 |  | 0.193 |  | 0.280 |  | 0.360 |  | 0.271 |  |
| Note. ULI:Upper-Lower Index, RIT:Item-Total correlation, RIR: Item-Rest correlation. | | | | | | | | | |
|  | | | | | | | | | |

**Distractor plot**

A graph with red line and green line

AI-generated content may be incorrect.

Andersen, E. B. (1995). Polytomous Rasch Models and their Estimation. V G. H. Fischer & I. W. Molenaar (Ur.), *Rasch Models* (str. 271–291). Springer New York. https://doi.org/10.1007/978-1-4612-4230-7\_15

Anselmi, P., Colledani, D., & Robusto, E. (2019). A Comparison of Classical and Modern Measures of Internal Consistency. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02714

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education: Classical test theory and item response theory. *Medical Education*, *44*(1), 109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory* (0 izd.). Psychology Press. https://doi.org/10.4324/9781410605269

Harvey, R. J., & Hammer, A. L. (1999). Item Response Theory. *The Counseling Psychologist*, *27*(3), 353–383. https://doi.org/10.1177/0011000099273004

Hilaliyah, H., Agustin, Y., Setiawati, S., Hapsari, S. N., Rangka, I. B., & Ratodi, M. (2019). Wright-Map to investigate the actual abilities on math test of elementary students. *Journal of Physics: Conference Series*, *1318*(1), 012067. https://doi.org/10.1088/1742-6596/1318/1/012067

Kreiner, S. (2012). The Rasch Model for Dichotomous Items. V K. B. Christensen, S. Kreiner, & M. Mesbah (Ur.), *Rasch Models in Health* (1. izd., str. 5–26). Wiley. https://doi.org/10.1002/9781118574454.ch1

Michaelides, M. (2010). A review of the effects on IRT item parameter estimates with a focus on misbehaving common items in test equating. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00167

Reise, S. P., & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: Some empirical and conceptual issues and challenges. *Psychological Medicine*, *46*(10), 2025–2039. https://doi.org/10.1017/s0033291716000520

Rosenbaum, P. R. (1987). Comparing Item Characteristic Curves. *Psychometrika*, *52*(2), 217–233. https://doi.org/10.1007/bf02294236

Svetina Valdivia, D., & Dai, S. (2024). Number of Response Categories and Sample Size Requirements in Polytomous IRT Models. *The Journal of Experimental Education*, *92*(1), 154–185. https://doi.org/10.1080/00220973.2022.2153783