

# JAMOVI ML

# GRUČENJE OZIROMA ZDRUŽEVANJE

## Uvod v združevanje

Združevanje (clustering) je metoda razvrščanja podatkov v skupine (grozde), kjer so podatki znotraj iste skupine bolj podobni drug drugemu kot podatkom v drugih skupinah. Cilj združevanja je ugotoviti latentne strukture in segmentirati podatke na podlagi podobnosti.

### K-means algoritem

K-means je eden najpogostejše uporabljenih algoritmov za združevanje. Deluje po naslednjem postopku:

Število grozdov ( $k$ ): Uporabnik določi število grozdov, ki jih algoritem išče.

Naključna inicializacija: Algoritem naključno izbere  $k$  začetnih središč (centroidov).

Dodelitev točk: Vsaka točka podatkovnega nabora je dodeljena tistemu središču, ki je najbližje (glede na evklidsko razdaljo).

Posodobitev središč: Za vsak grozd se izračuna novo središče kot povprečje vseh točk v tem grozdu.

Ponavljanje: Koraki dodelitve in posodobitve se ponavljajo, dokler se središča grozdov ne prenehajo spreminjati ali dokler ni doseženo največje število iteracij.

Izbira optimalnega števila grozdov

Optimalno število grozdov ( $k$ ) je pogosto neznano, zato uporabimo različne metode za njegovo določitev:

Gap statistika: Primerja variabilnost znotraj grozdov s pričakovano variabilnostjo pri naključnih podatkih. Optimalno število grozdov je tisto, pri katerem je razlika (gap) največja.

Elbow metoda: Izriše vsoto kvadratov razdalj med točkami in središči grozdov za različne vrednosti  $k$ . Optimalno število grozdov je tam, kjer se krivulja "zlomi" (točka komolca).

Silhouette koeficient: Meri, kako podobne so točke znotraj grozda v primerjavi s točkami v drugih grozdih. Večji koeficient pomeni boljšo ločljivost grozdov.

### Standardizacija podatkov

Pred uporabo algoritma K-means je pomembno, da podatke standardiziramo (npr. z metodo z-ocen), še posebej, če imajo spremenljivke različne enote ali skale. Standardizacija zagotavlja, da vse spremenljivke enakovredno prispevajo k izračunu razdalj.

### Grafična predstavitev grozdov

Cluster plot: Izris grozdov v 2D ali 3D prostoru. Točke so obarvane glede na pripadnost posameznim grozdom, središča grozdov pa so običajno prikazana kot posebni simboli.

Plot srednjih vrednosti: Po združitvi lahko izračunamo in prikažemo povprečne vrednosti posameznih značilk za vsak grozd, kar omogoča interpretacijo razlik med njimi.

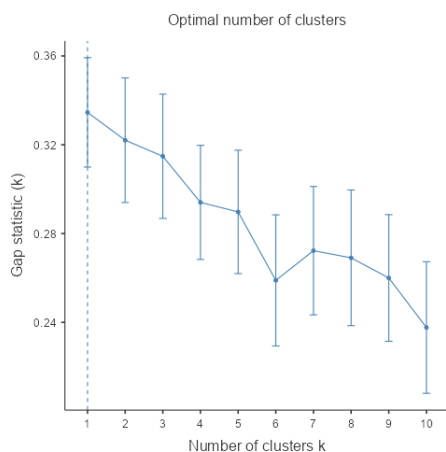
Gap statistika graf: Prikazuje razlike v variabilnosti za različne vrednosti  $k$ , kar pomaga pri izbiri optimalnega števila grozdov.

## Primer

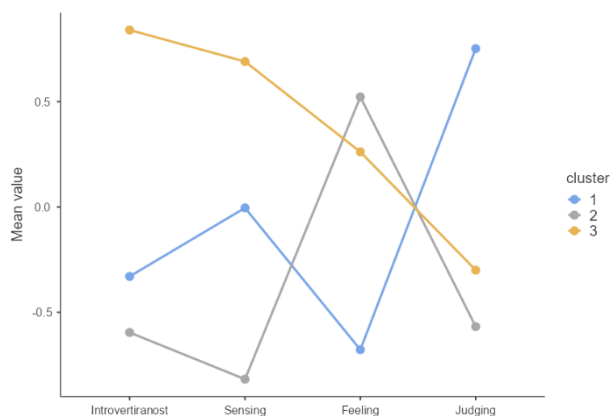
V enakem datasetu želimo združiti dijake s podobnimi rezultati MBTI testa. V modulu snowCluster tokrat izberemo K-means Clustering in izberemo zvezne spremenljivke, glede na katere želimo podatke združiti.

Pogosto vključimo standardizacijo. Program ponuja, da že vnaprej povemo, kolikšno število skupin bi radi imeli. Izberemo, da želimo grafe povprečij glede na skupine (to nam oriše situacijo, po kateri smo skupine združevali) in optimalno število grozdov (Gap statistika).

Najprej pregledamo optimalno število skupin. Iščemo višje točke oziroma točke, preden funkcija strmo pade. Na sliki (med rezultati pogledamo graf »Optimal number of clusters«) vidimo, da bi bile dobre izbire 3, 5, ali 7 (če nam ustreza večje število skupin). Za primer vzemimo  $k = 3$ .



V rezultatih nam najprej javi tabelo napak in natančnosti. V posameznih grupah vidimo varianco med samimi enotami v skupini in glede na vse. Velika vrednost »Between clusters« in majhna v posameznih skupinah nam pove, da je model ustrezen. V grafu povprečij glede na skupine lahko opazujemo trend, da...



## Vaje

V isti podatkovni zbirki primera za število skupin  $k = 2$  dobimo sledečo tabelo. Je rezultat ustrezen?

Sum of squares Table

	Value
Cluster 1	178.3
Cluster 2	102.9
Between clusters	70.8
Total	352.0

### Vaja 3:

V isti podatkovni zbirki gruči dijake glede na motivacijo in anksioznost in komentiraj. Lahko sklepaš kakšno zvezo med opazovanima spremenljivama in če ja, kako bi nadaljeval raziskavo?

## ITEM RESPONSE THEORY (IRT)

Item Response Theory (IRT) je sodobna psihometrična metoda, ki omogoča poglobljeno analizo odgovorov na testne postavke. 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. 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.

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 ( $\theta$ ):** Latentna spremenljivka, ki predstavlja raven sposobnosti, znanja ali lastnosti, ki jo test meri.
- **Parametrov postavke:** Značilnosti posamezne postavke, ki vplivajo na verjetnost odgovora.

## Modeli

**Dihotomski Rasch model** 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** 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

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.
- **Zanesljivost osebe (Person Reliability):** Ocena zanesljivosti merjenja sposobnosti za posameznika. Visoka zanesljivost pomeni, da je ocena sposobnosti posameznika stabilna in natančna. Jamovi pri Rasch modelu izračuna to zanesljivost, podobno kot Cronbachov alfa, vendar na osnovi IRT parametrov.

## Grafične predstavitve

1. **Wrightova mapa:** Grafični prikaz, ki združuje porazdelitev sposobnosti posameznikov in težavnosti postavk na isti lestvici (logiti). Omogoča vizualno preverjanje, ali je test primerno težak za populacijo, ki jo merimo, in ali so postavke dobro razporejene po celotnem spektru sposobnosti. Idealen test ima postavke, ki pokrivajo celotno paleto sposobnosti preizkušancev.
2. **Karakteristična krivulja postavke (Item Characteristic Curve; ICC):** Graf, ki prikazuje verjetnost pravilnega odgovora na item glede na sposobnost udeleženca. ICC pomaga razumeti, kako dobro vprašanje razlikuje med udeleženci z različnimi sposobnostmi.

## 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:

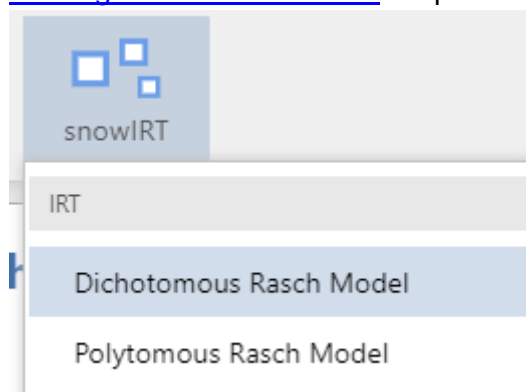
Velikost vzorca: IRT modeli so statistično bolj zahtevni in so manj zanesljivi pri majhnih vzorcih, saj ocene parametrov postanejo nestabilne. Priporoča se vzorec vsaj 200–500 preizkušancev, odvisno od kompleksnosti modela in števila postavk.

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

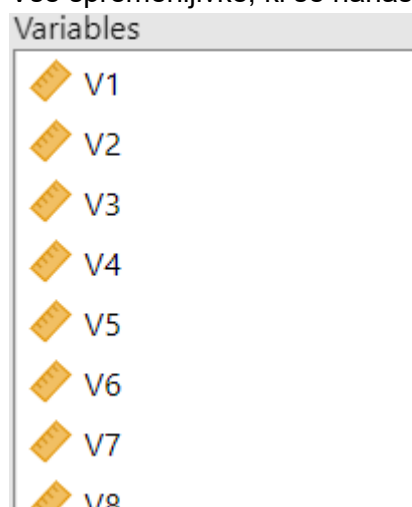
Predpostavke modela: IRT modeli temeljijo na določenih predpostavkah (npr. unidimenzionalnost, lokalna neodvisnost), ki jih je treba pred analizo preveriti.

### Primer (dihotomni Rasch model)

Najprej uvozimo modul snowIRT. Ker so naši podatki [JAMOVl-ucbenik/irt\\_data.csv at main · borbregant/JAMOVl-ucbenik](#) le tipa 0 ali 1, bomo izbrali dihotomski model.



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



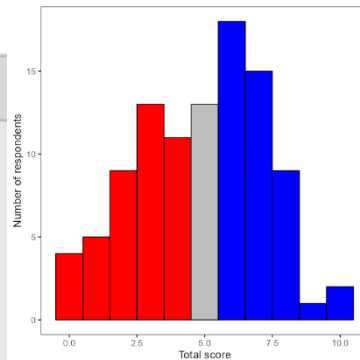
V oknu standard scores lahko odključamo opcije, ki se nanašajo na deskriptivno analizo

▼ | Standard scores

**Total score**

☒ Summary

☒ Histogram



V analizi nas najprej zanimajo podatki o samem modelu in podatki o težavnosti itemov, zato obkljukamo

▼ | Item Analysis

Item measures	Model fit
<input checked="" type="checkbox"/> Proportion	<input checked="" type="checkbox"/> Reliability

Dobimo

Model Fit

Person Reliability	
scale	0.649

Item Statistics

	Proportion
V1	0.830
V2	0.690
V3	0.700
V4	0.640
V5	0.560
V6	0.410
V7	0.370
V8	0.290
V9	0.180
V10	0.180

**Person Reliability (0.649):** Ta vrednost kaže na zmerno zanesljivost merjenja sposobnosti posameznikov. Je sprejemljiva za splošne raziskave ali primerjave skupin, vendar morda

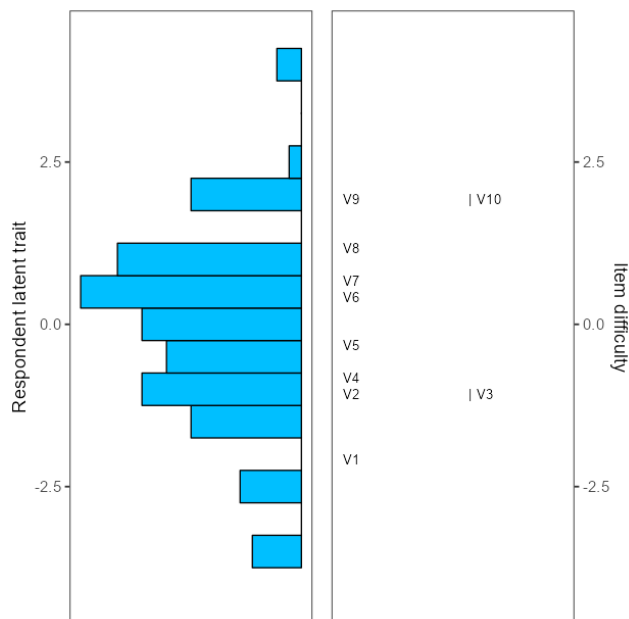
prenizka za visoko tvegane individualne odločitve (npr. diagnoze). Višja vrednost (bližje 1) pomeni večjo natančnost ocen.

Stolpec "Proportion" prikazuje delež pravih odgovorov na posamezno postavko. To je enostaven indikator težavnosti postavke.

- **V1 (0.830):** Relativno lahka postavka, saj je 83% udeležencev odgovorilo pravilno.
- **V2 (0.690):** 69% pravih odgovorov.
- ...
- **V10 (0.180):** Relativno težka postavka, saj je le 18% udeležencev odgovorilo pravilno.

V zavihku **Plots** 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.



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

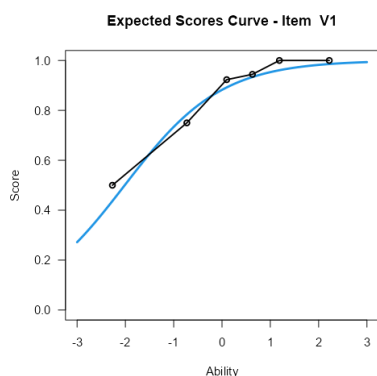
- **Razpon postavk:** Naše postavke pokrivajo dober razpon latentne spremenljivke, od zelo lahkih (V1) do zelo težkih (V9, V10).
- **Prekrivanje sposobnosti in težavnosti:** Obstaja dobro prekrivanje med razponom sposobnosti udeležencev in težavnosti postavk. To pomeni, da so za večino udeležencev na 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.

### Item Characteristic Curves (ICC):

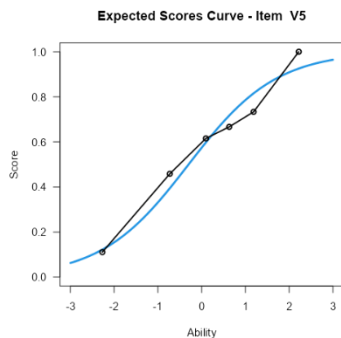
Za postavko V1 (lahka postavka):



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 pravih 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):



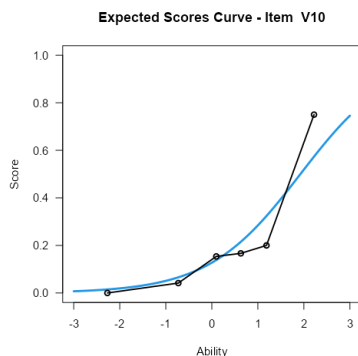
Interpretacija ICC za postavko V5 (srednje težka):

- **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 pravih 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):



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 pravih odgovorov (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 pravih 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/JAMOVU-ucbenik/blob/main/irt\\_data\\_vaja.csv](https://github.com/borbregant/JAMOVU-ucbenik/blob/main/irt_data_vaja.csv)

## Primer (politomski model)

Najprej komentirajmo navodila, ki nam jih JAMOVU 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 izvirno 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.

**Primer scenarija testa: ...**

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

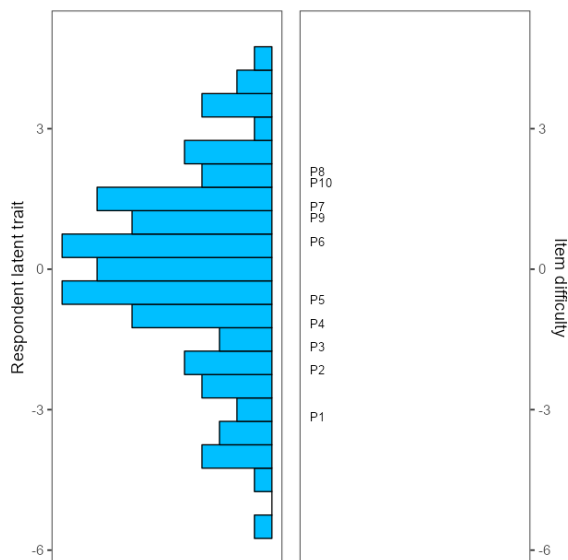
	tau parameters			
	1	2	3	4
P1	-28.11	6.622	9.208	12.28
P2	-4.11	-0.508	0.554	4.06
P3	-3.88	-1.010	0.564	4.32
P4	-29.72	7.366	9.728	12.63
P5	-4.77	-0.830	1.355	4.24
P6	-3.97	-0.916	0.902	3.98
P7	-4.26	-0.736	1.359	3.64
P8	-4.16	-0.615	1.461	3.32
P9	-3.16	0.839	2.324	NaN
P10	-2.95	0.531	2.419	NaN

**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**.

- **$\tau_1$  (tau 1):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 0 enaka verjetnosti izbire kategorije 1.
- **$\tau_2$  (tau 2):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 1 enaka verjetnosti izbire kategorije 2.
- **$\tau_3$  (tau 3):** Predstavlja točko na latentni lestvici, kjer je verjetnost izbire kategorije 2 enaka verjetnosti izbire kategorije 3.
- **$\tau_4$  (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.  $\tau_1 < \tau_2 < \tau_3 < \tau_4$ ). 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  $\tau_1$  izjemno nizek (-28.11), kar pomeni, da je prehod iz kategorije 0 v 1 izjemno lahek. Po drugi strani so višji pragovi (npr.  $\tau_4$  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  $\tau_4$ ):** "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.  $\tau_1 < \tau_2 < \tau_3 < \tau_4$ ). Če temu ni tako (npr.  $\tau_1$  je višji od  $\tau_2$ ), 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.

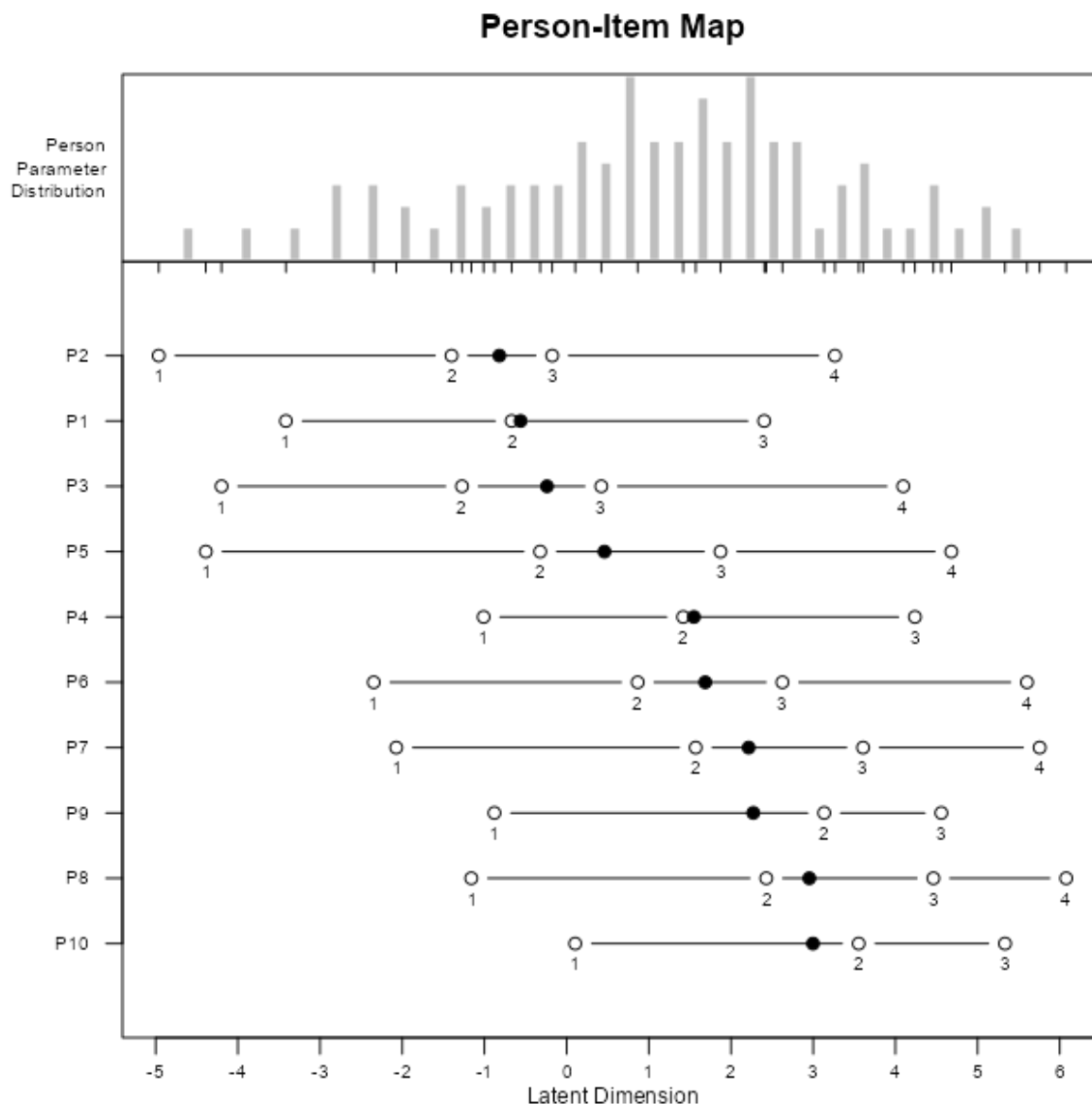
**Wright Map**



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.

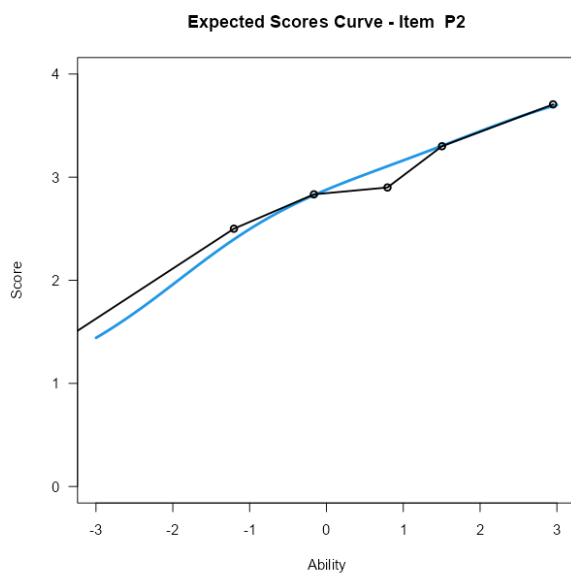
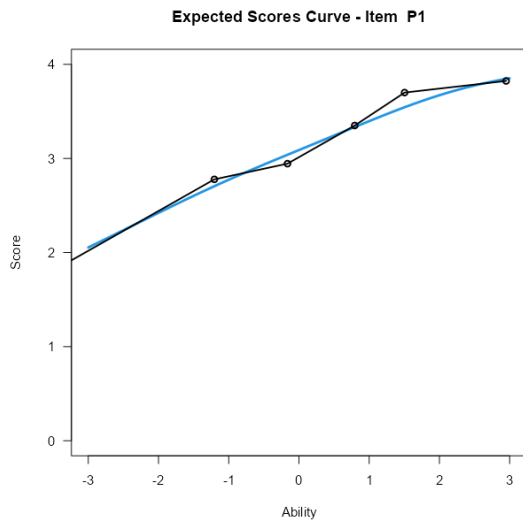
#### **Person-item map for PCM**



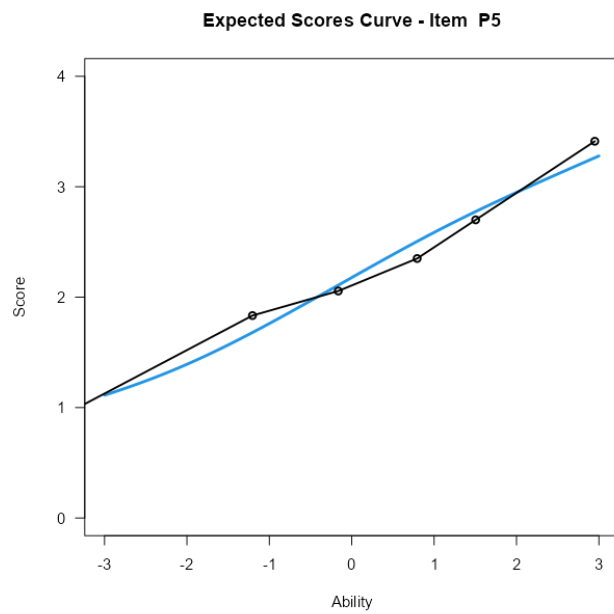
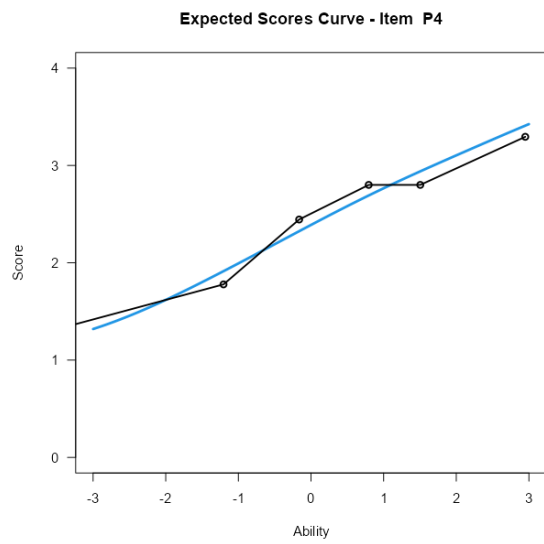
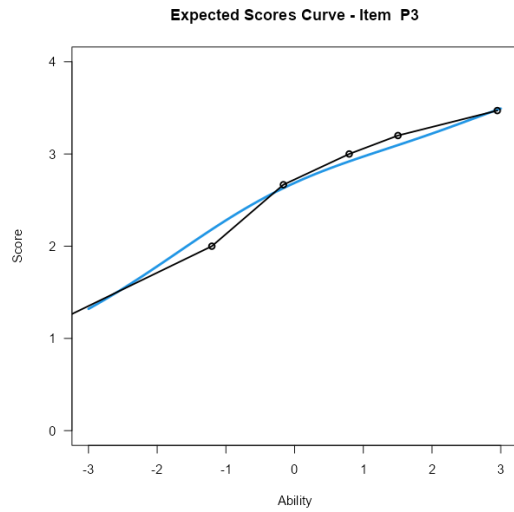
- **Komentar:** Wrightova mapa za Partial Credit Model prikazuje porazdelitev sposobnosti posameznikov in vse posamezne težavnostne pragove za vsako postavko.
- **Razlika od dihotomne:** Pri dihotomnih modelih vsaka postavka prispeva le eno točko (težavnost). Pri PCM pa vsaka postavka prispeva več točk – 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).

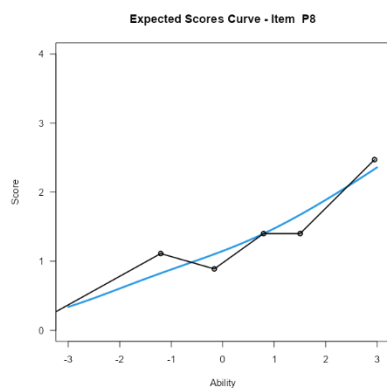
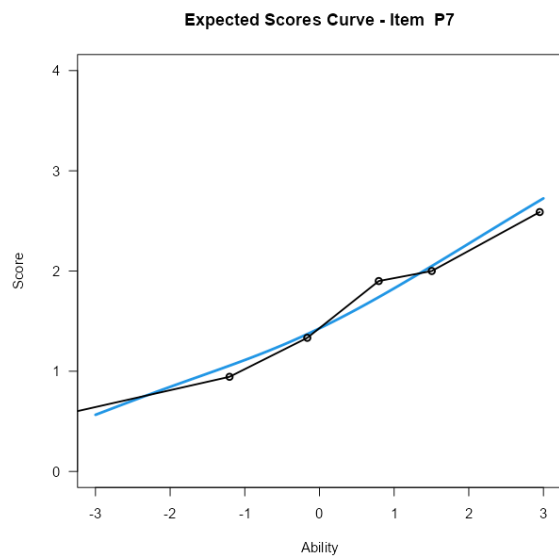
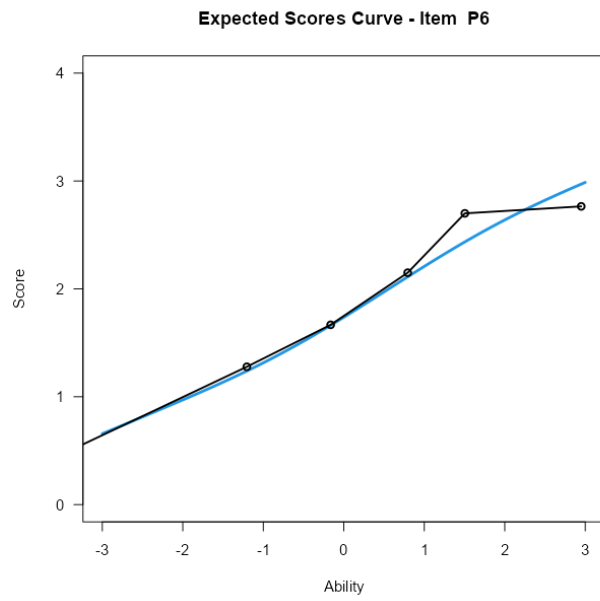
### Expected Score Curve

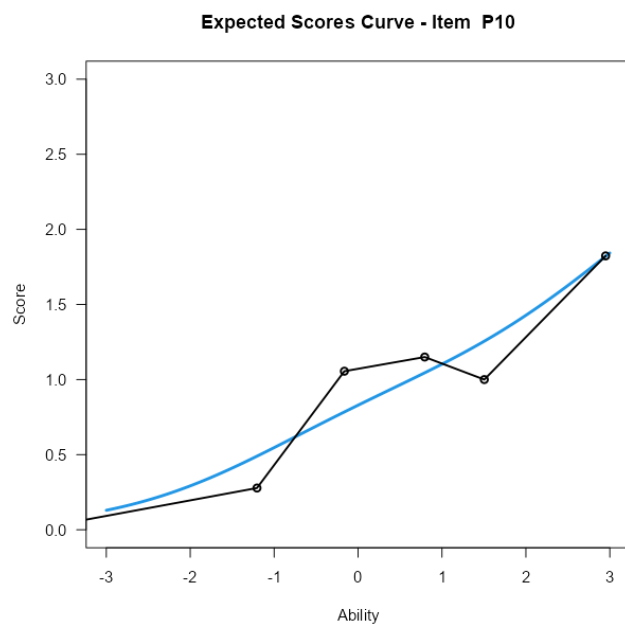
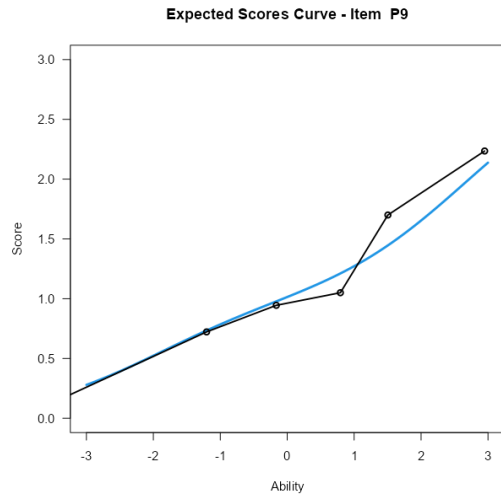
**Komentar:** Namesto verjetnosti pravilnega odgovora (kot pri dihotomnem modelu), ICC pri politomnih modelih prikazuje pričakovani rezultat na postavki kot funkcijo sposobnosti posameznika. Zato se krivulja dviga od najnižje do najvišje kategorije (npr. od 0 do 4).



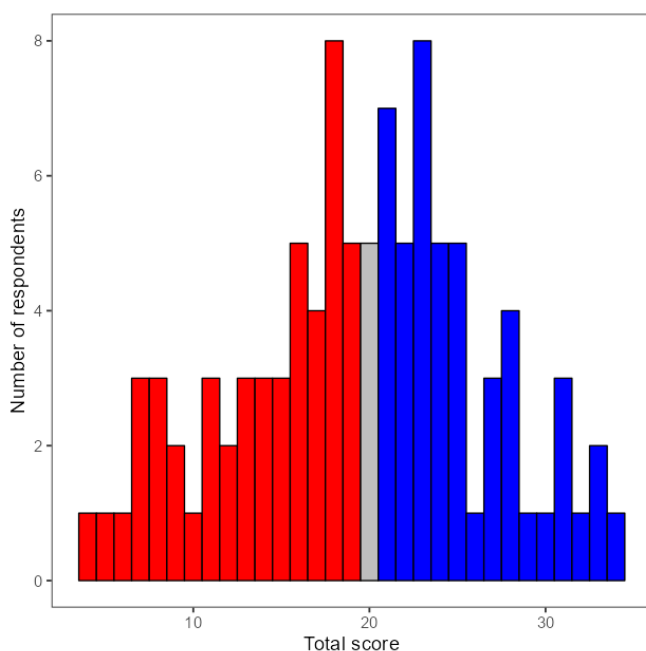








**Rating Scale Deltas/thresholds**  
**Histogram of total score**



## Nadaljnje možnosti v modulu snowIRT

### 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.

### 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.

## Meta-Analiza

### Uvod v metanalizo

Metanaliza je statistična metoda, ki združuje rezultate več študij, da bi pridobili bolj zanesljive in splošne zaključke. Gre za sistematičen pregled in analizo podatkov iz različnih raziskav, ki preučujejo podobno vprašanje. Metanaliza omogoča povečanje statistične moči, zmanjšanje negotovosti in identificiranje vzorcev, ki jih posamezne študije same ne morejo razkriti.

### Postopek metanalize

Metanaliza poteka po naslednjih korakih:

Določitev cilja in vprašanja: Jasno opredelite raziskovalno vprašanje in kriterije za vključitev študij.

1. Iskanje študij: Sistematično iščemo študije v znanstvenih bazah podatkov, kot so PubMed, Web of Science ali drugi relevantni viri.
2. Izbira študij: Študije se izberejo na podlagi vnaprej določenih kriterijev (npr. metodološka kakovost, velikost vzorca, relevantnost).
3. Ekstrakcija podatkov: Iz vsake študije se izluščijo ključni podatki, kot so velikost učinka, standardne napake, intervali zaupanja in druge statistične mere.
4. Analiza podatkov: Podatki se združijo in analizirajo s statističnimi metodami, kot so modeli s fiksnimi ali naključnimi učinki.
5. Interpretacija rezultatov: Rezultati se interpretirajo v kontekstu raziskovalnega vprašanja, pri čemer se upoštevajo morebitne omejitve in pristranskosti.

### Modeli metanalize

Metanaliza lahko uporablja dva glavna modela:

Model s fiksnimi učinki: Predpostavlja, da so vsi učinki v študijah enaki in da razlike med študijami nastajajo le zaradi naključne variabilnosti.

Model z naključnimi učinki: Upošteva, da lahko med študijami obstajajo resnične razlike v učinkih, ki niso posledica le naključja.

### Ocena heterogenosti

Heterogenost meri, kako različni so rezultati med študijami. Pomembni statistični kazalci za oceno heterogenosti vključujejo:

1. Q-test: Preverja, ali so razlike med študijami statistično pomembne.

2.  $I^2$ -statistika: Meri delež variabilnosti, ki je posledica heterogenosti namesto naključja. Višje vrednosti  $I^2$  kažejo na večjo heterogenost.

## Grafična predstavitev rezultatov

1. Forest plot: Prikazuje učinke posameznih študij in skupni učinek metanalize. Vsaka študija je predstavljena s črto, ki prikazuje interval zaupanja, skupni učinek pa je prikazan z diamantom.
2. Funnel plot: Uporablja se za preverjanje pristranskosti. Simetričen funnel plot kaže na odsotnost pristranskosti, medtem ko asimetrični plot lahko nakazuje prisotnost manjkajočih študij.

## Omejitve in izzivi

Metanaliza ima tudi svoje omejitve, kot so:

Pistranskost publikacij: Študije z negativnimi rezultati so pogosto manj verjetno objavljene.

Heterogenost študij: Razlike v metodologiji ali populacijah lahko otežujejo primerjavo.

Kakovost podatkov: Rezultati metanalize so odvisni od kakovosti vključenih študij.

## Primer

Za meta-analize v JAMOVU, bomo najprej uvozili modul MAJOR, ki omogoča analizo korelacijskih koeficientov, dihotomske modele, analizo velikosti učinka, analizo srednjih razlik, analizo deležev dogodkov. In analizo notranje zanesljivosti vprašalnikov. Na primeru si bomo natančneje pogledali analizo srednjih razlik, analizo velikosti učinka in analizo notranje konsistentnosti, ostale paprepuščamo bralcu.



### MAJOR - Meta-Analysis for JAMOVU 1.2.4

W. Kyle Hamilton

An interface for Jamovi and the R package Metafor

INSTALLED

#### Meta Analysis

Correlation Coefficients (r, N)

Dichotomous Models

Effect Sizes and (Sampling Variances or Standard Errors)

Mean Differences (n, M, SD)

Proportions

Reliability Generalization

Za nabor podatkov bomo uporabili datoteko XXX, ki je računalniško generirana in vsebuje umetne podatke o 100 študijah, kjer so zbrane srednje vrednosti, mere razpršenosti, mere učinka (te je včasih v pedagoškem raziskovanju potrebno dodati, ne bodo vsi nabori v študijah a priori lepi za obravnavo), in mere za notranje konsistentnosti.

Najprej si pogledjmo analizo srednjih razlik. Izberemo ustrezne stolpce, kot prikazuje slika X.

Random-Effects Model (k = 100)

	Estimate	se	Z	p	CI Lower Bound	CI Upper Bound
Intercept	-0.475	0.0197	-24.1	< .001	-0.514	-0.436
	.	.	.	.	.	.

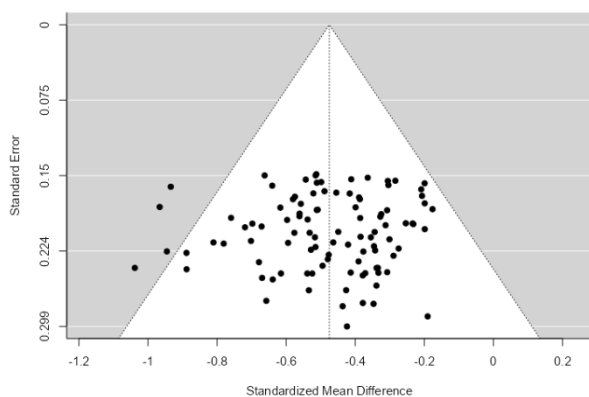
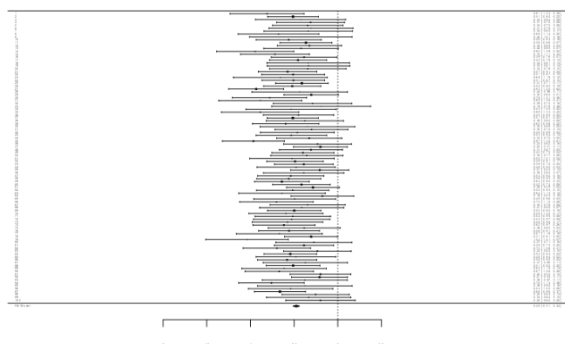
Note. Tau<sup>2</sup> Estimator: Restricted Maximum-Likelihood

V prvi izhodni tabeli iz negativnega “Estimate” vidimo, da je učinek v negativno smer, torej je v eksperimentalni skupini vrednost opazovane spremenljivke manjša za povprečno 0.475. P vrenost nam pove, da je velikost učinka statistično pomembna. Druga izhodna tabela (prepuščamo bralcu, da jo generira) nam z raznimi metrikami pove, da heterogenost ni prisotna, torej statističnih razlik med študijami ni.

Vse te podatke nam modul v JAMOVl tudi pove v vezanem besedilu na sliki X z interpretacijo, kar olajša interpretacijo. Uporabi še nekaj metrik, ki jih nismo izbrali; Ključna opazka je npr., da imata dve vključeni študiji zelo velik učinek, torej moramo biti do njih posebno pozorni.

A total of k=100 studies were included in the analysis. The observed standardized mean differences ranged from -1.0382 to -0.1763, with the majority of estimates being negative (100%). The estimated average standardized mean difference based on the random-effects model was  $\hat{\mu} = -0.4751$  (95% CI: -0.5138 to -0.4364). Therefore, the average outcome differed significantly from zero ( $z = -24.0767$ ,  $p < 0.0001$ ). According to the Q-test, there was no significant amount of heterogeneity in the true outcomes ( $Q(99) = 87.6656$ ,  $p = 0.7854$ ,  $\tau^2 = 0.0000$ ,  $I^2 = 0.0000\%$ ). An examination of the studentized residuals revealed that none of the studies had a value larger than  $\pm 3.4808$  and hence there was no indication of outliers in the context of this model. According to the Cook's distances, two studies (27; 45) could be considered to be overly influential. Neither the rank correlation nor the regression test indicated any funnel plot asymmetry ( $p = 0.4450$  and  $p = 0.5389$ , respectively).

V študijo pogosto vključimo še Forest plot na sliki X, ki smo ga že komentirali in Funnel plot na sliki X, ki prikazuje, da so naše raziskave “simetrične”, torej ni bilo pristranskosti pri izbiri raziskav.



Podobno lahko naredimo za effect size:

Postopek je zelo podoben, a pazimo, da v Model options izberemo Standard errors, saj imamo le take podatke (če bi imeli oz. Naredili novo spremenljivko variance, bi bilo seveda drugače). Za effect size smo uporabili Cohenov d, analogno bi lahko delali tudi z drugimi velikostmi učinka.

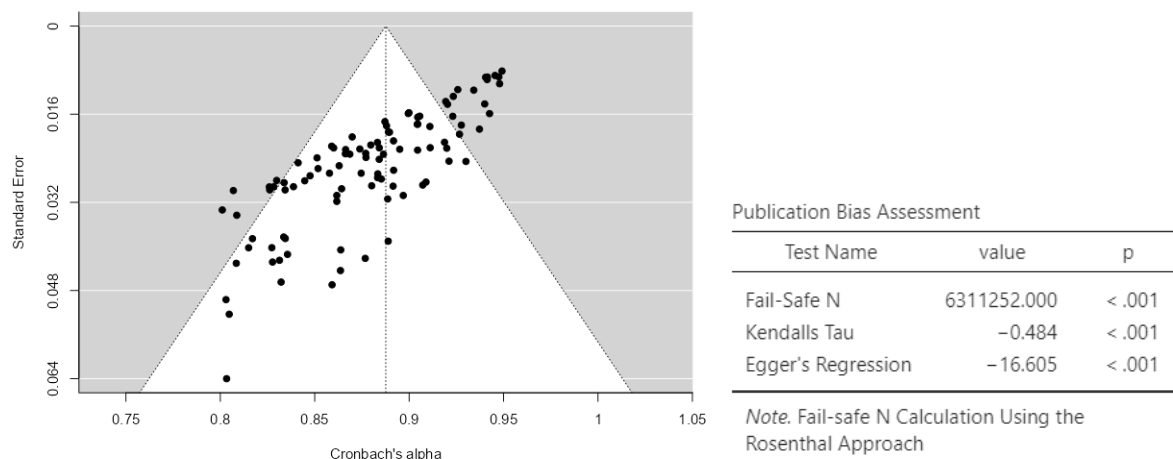
Podobno lahko delamo z analizo zanesljivosti. Spomnimo se, da Cronbachov alpha koeficient meri, ali vprašanja v vprašalniku, ki določajo eno spremenljivko merijo isti konstrukt. Seveda je nujno, da študije zajete v meta-analizo uporabljajo enak vprašalnik oziroma so razlike lahko zgolj vpliv prilagoditev na kulturni konteksts ipd.

Cronbach's Alpha	
<input type="checkbox"/> Cronbach Alpha	<input type="checkbox"/>
Number of Items	
<input type="checkbox"/> Items	<input type="checkbox"/>
Sample Size	
<input type="checkbox"/> N1	<input type="checkbox"/>
Study Label	
<input type="checkbox"/> Study	<input type="checkbox"/>



V rezultatih namrežultat 0.89 pove, da je povprečna zanesljivost študij zelo dobra in statistično razvidna (majhna p vrednost). Visoke vrednosti statistik heterogenosti (Tau,  $I^2$ ,  $H^2$ , Q) nam kažejo, da je heterogenost prisotna, torej razlike v študijah niso zgolj posledica napak.

### Funnel Plot



Funnel plot nam tokrat sugerira asimetrijo, kar lahko pomeni pristranskost pri izbiri publikacij. To nam potrdi majhna p vrednost Eggerjevega testa. Manj točk spodaj desno sugerira, da imamo manj raziskav z majhno napako in velikim vplivom (zanesljivostjo). To se v statistiki pogosto dogaja, saj so »boljši« rezultati večkrat objavljeni.

## Dodatne možnosti v JAMOV

### Analiza korelacijskih koeficientov

Analiza korelacijskih koeficientov v metaanalizi omogoča združevanje rezultatov več študij, ki so preučevale povezanost med dvema spremenljivkama. Ključni podatki, ki jih uporabimo v analizi, so ocenjena korelacija ( $r$ ) in velikost vzorca ( $n$ ) za vsako posamezno študijo.

### Postopek analize

1. **Transformacija korelacij:** Ker korelacijski koeficienti niso normalno porazdeljeni, jih običajno transformiramo z uporabo Fisherjeve z-transformacije, da omogočimo pravilno statistično analizo. To je v MAJOR JAMOV modulu že privzeto, kot prikazuje slika

▼ | Model Options

Model estimator: Restricted Maximum-Likelihood ▼

Model measures: Fisher's r-to-z transformed correlation coefficient ▼

Moderator type:
 

- Raw correlation coefficient
- Raw correlation coefficient (bias corrected)
- Fisher's r-to-z transformed correlation coefficient

Tests and Confidence intervals: Z test ▼

Confidence interval level: 95 %

2. **Izračun uteženih povprečij:** Vsak korelacijski koeficient prispeva k skupnemu rezultatu glede na svojo velikost vzorca, pri čemer večji vzorci dobijo večjo težo.
3. **Ocena heterogenosti:** Uporabimo statistike, kot so Q-test in  $I^2$ , da preverimo, ali obstajajo pomembne razlike med študijami.
4. **Interpretacija rezultatov:** Povprečna korelacija in njeni intervali zaupanja nam omogočajo vpogled v moč in smer povezanosti med spremenljivkama.
5. **Grafična predstavitev**
  - **Forest plot:** Prikazuje ocene korelacij posameznih študij in skupno oceno.
  - **Funnel plot:** Preverja morebitno pristranskost publikacij pri objavljenih rezultatih.

## Dihotomski modeli

Dihotomski modeli se uporabljajo, kadar imamo rezultate v obliki diskretnih dogodkov, torej kadar vsaka študija poroča o številu dogodkov v eksperimentalni in kontrolni skupini ter skupni velikosti vzorca. To je pogosto pri kliničnih raziskavah, kjer merimo učinek intervencije.

### Ključni koraki analize

1. **Pridobivanje podatkov:** Zberemo podatke o številu dogodkov v eksperimentalni (E) in kontrolni (K) skupini ter velikost vzorca v obeh skupinah.
2. **Izračun velikosti učinka:** Najpogosteje uporabljene mere vključujejo:
  - **Razmerje obetov (Odds Ratio, OR):** Izračunano kot  $(E1 / (N1 - E1)) / (K1 / (N2 - K1))$ .
  - **Relativno tveganje (RR):** Določa razmerje med tveganjem v eksperimentalni in kontrolni skupini.
  - V modulu MAJOR to izbiramo v podpoglavju Model options, model measures, kjer je privzeta možnost Razmerje obetov kot prikazuje slika. V modulu je privzeto logaritmiranje podatkov, kar lahko izklopimo z uporabo inverzne transformacije *Back-Transform* na sliki.

Model Options

Model estimator: Restricted Maximum-Likelihood

Model measures: Log odds ratio

Moderator type:

☐ Back-Transform

Confidence interval level: 95 %

3. **Izračun heterogenosti:** Podobno kot pri drugih metaanalizah uporabimo Q-test in  $I^2$  za oceno variabilnosti med študijami. Če heterogenost ni prisotna, je morda bolj primeren model Fixed Effects, kar prav tako izberemo v podpoglavju model options, tokrat model estimator.
4. **Vizualizacija rezultatov**
  - **Forest plot:** Prikazuje razmerja obetov za vsako študijo ter skupno oceno.
  - **Funnel plot:** Pomaga odkriti pristranskost objavljenih raziskav.

### Analiza proporcev

V analizi proporcev preučujemo pogostost določenega dogodka v različnih študijah. Podatki vključujejo število primerov dogodka (X) in skupno velikost vzorca (N) za vsako študijo.

#### Postopek analize

1. **Pridobivanje podatkov:** Za vsako študijo zabeležimo število primerov dogodka (X) in velikost vzorca (N).
2. **Izračun proporcev:** Vsaka študija prispeva vrednost  $X/N$ , ki predstavlja relativno pogostost dogodka. Za razliko od dihotomskih modelov, modul JAMOVI privzame surove (ang. *raw*) podatke in jih ne logaritmirajo. To lahko spremenimo v razdelku Model Options, kjer izberemo Effect size model measures. Prav tako lahko spreminjamo model. Kot drugje je npr. Če je prisotna heterogenost je privzet model največjega verjetja ustrezen, sicer lahko izberemo model Fixed-Effect.
3. **Združevanje rezultatov:** Uporabimo uteženo povprečje proporcev, pri čemer večji vzorci dobijo večjo težo.
4. **Preverjanje heterogenosti:** Statistični testi (Q-test,  $I^2$ ) ocenijo variabilnost rezultatov med študijami.
5. **Grafična predstavitev**
  - **Forest plot:** Vizualizira rezultate posameznih študij in skupno oceno proporcev.
  - **Funnel plot:** Omogoča preverjanje morebitne pristranskosti publikacij.

Vsaka od teh analiz omogoča vpogled v vzorce podatkov iz različnih študij in pomaga pri oblikovanju zanesljivih zaključkov na podlagi združenih rezultatov.

## STRUKTURNA ENAČBA MODELA (SEM)

Modeliranje strukturnih enačb (SEM) je napredna statistična tehnika, ki omogoča preučevanje kompleksnih odnosov med latentnimi konstrukti in njihovimi merjenimi kazalci.

SEM združuje elemente konfirmatorne faktorske analize (CFA) in potnih modelov, kar omogoča hkratno analizo merilnih in strukturnih odnosov.

## Osnovne značilnosti SEM

SEM je linearni model, ki vključuje:

- **Povezave med opazovanimi spremenljivkami** (npr. regresijske koeficiente),
- **Povezave med latentnimi in opazovanimi spremenljivkami** (npr. faktorske uteži v CFA),
- **Povezave med latentnimi spremenljivkami** (npr. strukturni modeli).

Ta pristop omogoča preverjanje hipotez o vzročnih povezavah, posrednih učinkih in kovariancah, hkrati pa upošteva tudi merilne napake.

Ključni koncepti

1. **Merilni model (CFA)**
  - Preverja, ali izbrani kazalci ustrezno merijo osnovne latentne konstrukte.
  - Vključuje ocenjevanje faktorskih uteži, zanesljivosti konstruktov (npr. Cronbach's alpha, McDonald's omega) in diskriminantne veljavnosti.
2. **Strukturni model**
  - Analizira odnose med latentnimi spremenljivkami.
  - Vključuje:
    - **Neposredne vplive** (npr. vpliv enega konstrukta na drugega),
    - **Posredne vplive** (npr. preko mediatorjev),
    - **Kovariance** (povezanost med konstrukti).
3. **Ocena ustreznosti modela**
  - **$\chi^2$ -test**: Preverja, ali model ustreza podatkom (ničelna hipoteza predpostavlja dobro prileganje).
  - **Primerjalni indeksi (CFI, TLI)**: Vrednosti nad 0.95 kažejo dobro ujemanje.
  - **RMSEA in SRMR**: Vrednosti pod 0.08 so zaželeni.

## Grafična predstavitev in interpretacija

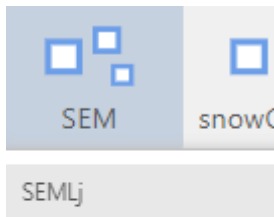
- **Potni diagram**: Vizualizacija povezav med konstrukti in kazalci.
- **Standardizirane rešitve**: Prikazujejo moč in smer povezav.
- **Modifikacijski indeksi**: Namigujejo na možne izboljšave modela.

## Omejitve in izzivi

- Za zanesljive rezultate je potreben **dovolj velik vzorec** ( $n > 200$ ).
- **Normalnost porazdelitve** je pomembna pri uporabi metod največje verjetnosti (ML).
- **Prekomerna kompleksnost modela** lahko zmanjša zanesljivost ocen.

Primer uporabe: (NI OK KER JE UBISTVU SAMO MULTIVAR ANALIZA!) – brezvezen primer za SEM, ampak se da tudi s tem to narediti...

**Naložite podatke** v jamovi in odprite SEMLj modul. Modul ponuja sintaksno rešitev, ki je primerna, če nam ustreza bolj ustreza programerski pristop in interaktivno, ki jo bomo uporabili mi.



SEM (syntax)

SEM (interactive)

Uporabili bomo podatke [here at UCLA](#). Vzorec sestavlja 500 študentov. Opazovanih je bilo 9 spremenljivk: Motivation (motiv), Harmony (harm), Stability (stabi), Negative Parental Psychology (ppsy), SES, Verbal IQ (verbal), Reading (read), Arithmetic (arith) and Spelling (spell). Vse spremenljivke so bile merjene na zvezni skali.

Suppose you are a researcher studying the effects of student background on academic achievement. The lab recently finished collecting and uploading the dataset of N=500 students each with 9 observed variables: Motivation, Harmony, Stability, Negative Parental Psychology, SES, Verbal IQ, Reading, Arithmetic and Spelling. The principal investigator hypothesizes three latent constructs Adjustment, Risk, Achievement measured with its corresponding to the following codebook mapping:

#### Adjustment

- **motiv** Motivation
- **harm** Harmony
- **stabi** Stability

#### Risk

- **ppsy** (Negative) Parental Psychology
- **ses** SES
- **verbal** Verbal IQ

#### Achievement

- **read** Reading
- **arith** Arithmetic
- **spell** Spelling

Postavimo model:

- **Odvisne (endogene) spremenljivke** – Spremenljivke, ki so pojasnjene z drugimi spremenljivkami v modelu (npr. *read*, *arith*).
- **Neodvisne (eksogene) spremenljivke** – Spremenljivke, ki vplivajo na druge, same pa niso pojasnjene z modelom (npr. *motiv*, *ppsy*). Motiv vpliva tako na *read*, kot na *arith*, *ppsy* pa le na *read*.

Najprej izberemo latentne spremenljivke, ki smo jih določili. V zavihku endogenous models izberemo katere neodvisne spremenljivke vplivajo na katere odvisne spremenljivke.

Izberemo še opcijo izrisa diagrama poti.

### Structural Equation Modeling

**Manifest Variables**

- motiv
- harm
- stabi
- ppsyh
- ses
- verbal
- read
- arith**
- spell

**Latent Endogenous Variables**

readd  
✕  
read

---

arithh  
✕  
arith

Add New Latent

**Latent Exogenous Variables**

psy  
✕  
ppsyh

---

mot  
✕  
motiv

Add New Latent

Endogenous models

Defined Factors	→	Endogenous Variables Models
readd	→ → ▼	<b>Endogenous = readd</b>
arithh		psy
psy		mot
mot		<b>Endogenous = arithh</b>
		<span>mot</span>

Path diagram

☒ Path diagram    **Path diagram**

☐ show residuals    ☐ show intercepts

Paths	Layout	Nodes
<input checked="" type="radio"/> Coefficients <input type="radio"/> Betas <input type="radio"/> Labels <input type="radio"/> None	Type: Tree-like Rotate: Exog. left	Node Size: Medium Manifest shapes: Rectangles Latent shapes: Circles Abbreviate: 5

Najprej lahko interpretiramo statistiko, ki jo že poznamo v Parameters estimates tabeli. Iz korelacijskega koeficienta beta in p-vrednosti lahko preberemo, da so postavljeni dejavniki smiselni. Koeficienti beta so prikazani tudi v diagramu poti.

### Parameters estimates

				95% Confidence Intervals		$\beta$	z	p
Dep	Pred	Estimate	SE	Lower	Upper			
readd	psy	-0.216	0.0296	-0.274	-0.158	-0.218	-7.29	< .001

## Parameters estimates

Dep	Pred	Estimate	SE	95% Confidence Intervals		$\beta$	z	p
				Lower	Upper			
readd	mot	0.476	0.0369	0.404	0.548	0.482	12.92	< .001
arithh	mot	0.600	0.0358	0.530	0.670	0.600	16.77	< .001

Če želimo preveriti ali motiv in psych vplivata različno na read, moramo najprej prebrati, kako je jamovi shranil parameter, kar preverimo s klikom show parameter options v menuju Output options. V parameter estimates nato vidimo, da v našem primeru ustreza p5 in p6. V constraint test vidimo, da je hipoteza, da bi bila vpliva enaka zavrnjena.

Output options		Label	Dep	Pred
Information	<input checked="" type="checkbox"/> Show parameters labels	p5	readd	psy
		p6	readd	mot

Zdaj lahko izberemo v custom model settings p5==p6.

Custom model settings

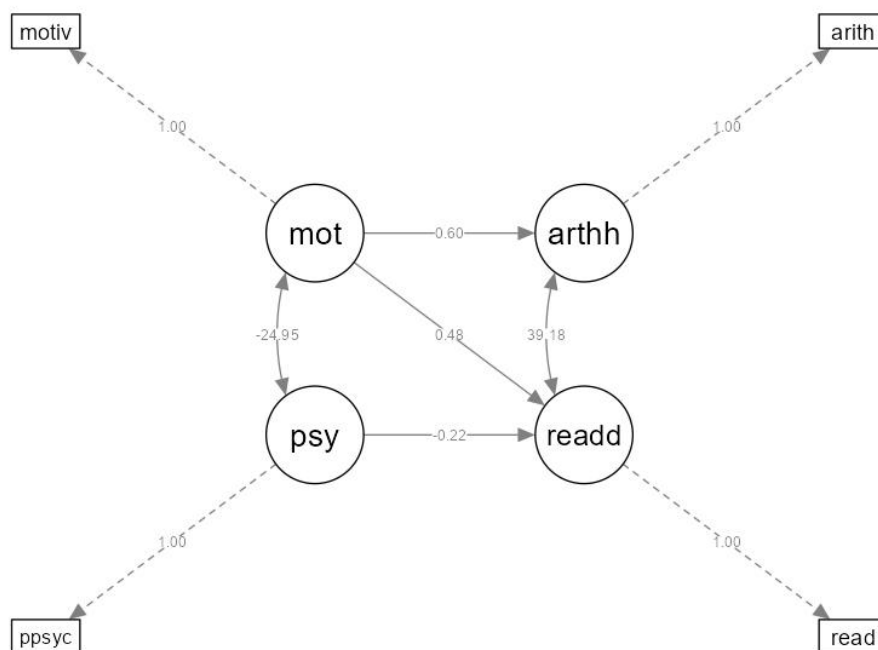
Insert constraints or defined parameters

p5==p6

Add directive

### Constraints score tests

Type	Par 1		Par 2	X <sup>2</sup>	df	p
Univariate	p5	==	p6	174	1	< .001
Total				174	1	< .001



## Vaja

### [SEM: Political Democracy](#)

Spremenljivka	Pomen v slovenščini
y1	Svoboda tiska, 1960
y2	Svoboda politične opozicije, 1960
y3	Poštenost volitev, 1960
y4	Učinkovitost izvoljenega zakonodajnega telesa, 1960
y5	Svoboda tiska, 1965
y6	Svoboda politične opozicije, 1965
y7	Poštenost volitev, 1965



Spremenljivka	Pomen v slovenščini
y8	Učinkovitost izvoljenega zakonodajnega telesa, 1965
x1	BDP na prebivalca, 1960
x2	Poraba energije na prebivalca, 1960
x3	Delež delovne sile v industriji, 1960

Podatkovni niz vsebuje **75 držav** z meritvami politične demokracije in gospodarskega razvoja v letih 1960 in 1965. Vsebuje štiri ključne indikatorje demokratičnosti, merjene v dveh časovnih obdobjih, ter tri gospodarske indikatorje iz leta 1960.

Model bo preučeval, kako gospodarski razvoj držav (merjen z BDP-jem, porabo energije in stopnjo industrializacije) vpliva na stopnjo demokratičnosti, pri čemer bo upošteval tudi trajnost demokracije skozi čas. Analiziral bo povezavo med gospodarskimi dejavniki in demokracijo v letu 1960, nato pa še vpliv teh dejavnikov in zgodnejšega stanja demokracije na demokracijo leta 1965, pri tem pa upošteval tudi povezave med posameznimi indikatorji demokratičnosti v obeh obdobjih.

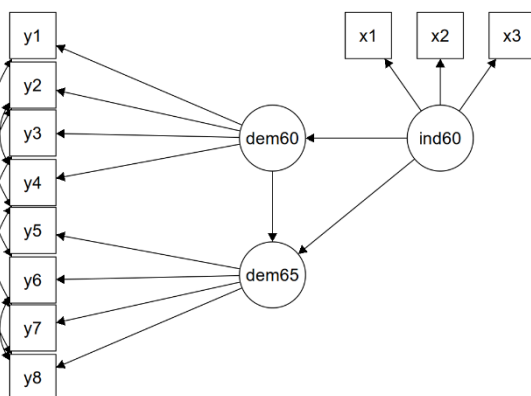
Model, ki ga želimo oceniti, je v osnovi posredniški model, kjer (latentna) stopnja industrializacije leta 1960 vpliva na (latentno) raven demokracije leta 1960, ta pa nato vpliva na (latentno) raven demokracije leta 1965.

Pomembno je omeniti, da ta model zahteva zgoraj omenjene latentne povezave in nekorelacije med opazovanimi spremenljivkami. Natančneje, želimo korelacije med naslednjimi pari opazovanih spremenljivk:

- y1 in y5
- y2 in y4
- y2 in y6
- y3 in y7
- y4 in y8
- y6 in y8

To so korelacije med ostanki opazovanih spremenljivk po odstranitvi njihovih obremenitev na latentne spremenljivke. V terminologiji faktorske analize gre za korelacije med unikančnostmi.

V okviru diagrama torej želimo



Zdaj lahko začnemo z interaktivnim modulom SEM. Najprej določimo, da ssta latentni spremenljivki dem60 in dem65 endogeni. Prvi pripadajo y1 do y4, drugi pa y5 do y8. Eksogena spremenljivka bo ind60, ki ji pripadajo x1,x2,x3.

Latent Endogenous Variables

dem60

- y1
- y2
- y3
- y4

dem65

- y5
- y6
- y7
- y8

Add New Latent

Latent Exogenous Variables

ind60

- x1
- x2
- x3

Add New Latent

Zdaj določimo, da dem60 napoveduje ind60, dem65 pa napovedujeta ind60 in dem60. To določimo v zavihku endogenous models.

Endogenous models

Defined Factors

dem60

dem65

ind60

→

→

Endogenous Variables Models

Endogenous = dem60

ind60

Endogenous = dem65

ind60

dem60

Zdaj določimo le še povezave med opazovanimi spremenljivkami y1 do y8. To določimo v zavihku variances and covariances.

Variances and covariances

y2

y3

y4

y5

y6

y7

y8

x1

→

y1

y2

y2

y3

y4

y6

y5

y4

y6

y7

y8

y8

Zadnje obkljukamo še opcijo path diagram:

35

▼ | Path diagram

☒ Path diagram

**Path diagram**

☐ show residuals

☐ show intercepts

**Paths**

☒ Coefficients

☐ Betas

☐ Labels

☐ None

Type

Tree-like ▼

Rotate

Exog. left ▼

**Nodes**

Node Size

Medium ▼

Manifest shapes

Rectangles ▼

Latent shapes

Circles ▼

Abbreviate

5 ▼

# STROJNO UČENJE V KLASIFIKACIJI

Strojno učenje je veja umetne inteligence, ki se osredotoča na razvoj algoritmov in modelov, ki omogočajo računalnikom, da se učijo iz podatkov in sprejemajo odločitve brez eksplicitnega programiranja. Glavni cilj strojnega učenja je prepoznati vzorce in odnose v podatkih ter na podlagi teh informacij napovedati ali klasificirati nove primere.

## DELITEV PODATKOV

Prvi korak v strojnem učenju je razdelitev podatkov na učni nabor (za treniranje modela), validacijski nabor (za izbiro hiperparametrov), in testni nabor (za končno oceno modela). Običajno se podatke deli v razmerju 70/30 (to je privzeto v Jamovi modulu, ki ga bomo uporabili) ali 80/20, pri čemer je pomembno, da je delitev naključna in da ohranimo reprezentativnost podatkov.

## MATRIKA ZMEDENOSTI IN KRIVULJA ROC

Za oceno kakovosti klasifikacijskega modela pogosto uporabimo matriko zmedenosti, ki prikazuje število pravilno in napačno napovedanih primerov v posameznih kategorijah. Poleg tega je ROC-krivulja (Receiver Operating Characteristic) uporabno orodje za vrednotenje delovanja modela pri različnih mejnih vrednostih napovedi. Merilo pod krivuljo ROC (AUC) nam pove, kako dobro model loči med razredi.

# MODELI STROJNEGA UČENJA

Obstaja veliko različnih modelov, ki se uporabljajo v strojnem učenju, od preprostih do naprednih. Med najpogostejšimi so:

Regressijski modeli: Uporabljajo se za napovedovanje zveznih vrednosti.

Naivni Bayes: Model, ki temelji na verjetnostih in predpostavki neodvisnosti med spremenljivkami.

Podporni vektorski stroji (SVM): Metoda, ki maksimizira mejo med razredi.

Odločitvena drevesa in naključni gozdovi (Random Forest): Naključni gozd je ansambelska metoda, ki združuje več odločitvenih dreves za robustnejše in natančnejše napovedi.

Nevronske mreže: Modeli, ki posnemajo delovanje možganov in so še posebej primerni za kompleksne nelinearne vzorce.

## PRIMERJAVA MODELOV IN KRIŽNA VALIDACIJA

Za primerjavo različnih modelov se pogosto uporablja metoda križne validacije (cross-validation). Pri tem podatke razdelimo na K skupin (npr.  $K = 5$  ali  $K = 10$ ), model pa treniramo in testiramo na različnih delih podatkov. Za bolj robustne rezultate lahko postopek križne validacije večkrat ponovimo (npr. 3 ali 5 ponovitev).

Križna validacija omogoča oceno splošne napake modela in je ključnega pomena pri izbiri najboljšega modela za določeno nalogo.

## DODATNE POMEMBNE TOČKE

Obdelava podatkov: Pravilna priprava podatkov (normalizacija, kodiranje kategorij, odstranjevanje manjkajočih vrednosti) je ključnega pomena za uspešnost modelov.

Hiperparametri: Mnogi modeli zahtevajo nastavitve hiperparametrov (npr. število dreves pri Random Forest), kar pomembno vpliva na njihovo delovanje.

Interpretacija rezultatov: Poleg metrike točnosti (accuracy) je pomembno upoštevati tudi druge kazalnike, kot so preciznost, priklic in F1-mera, še posebej pri neuravnoteženih podatkih.

## Primer

V podatkovni zbirki ... želimo napovedati spol učenca glede na raven Matematične motivacije in matematične anksioznosti. Ciljna spremenljivka (*Target*) je nominalna, napovedni (*Covariates*) pa obravnavamo zvezno.

Uporabili bomo modul snowCluster, kjer izberemo Machine learning, ki spada pod kategorijo Classification. Izberemo napovedne in ciljno spremenljivko, v zavihku »Analysis« pa določimo parametre, ki bodo diktirali strojno naučen model. »Split set« nam razdeli podatke na učne in testne in ga pustimo privzetega na 0.7. Training model pustimo privzel »pls«, lahko pa bi se tudi igrali z drugimi (zelo popularen v pedagoški znanosti je npr. »rf; random forest«). Zanima nas rezultat modela na »Test set«, zato tam vse odkljukamo, »Training set« pa pustimo prazen. Kategorija Plots nam omogoča osnovno statistiko že v sami knjižnici, da pred tem ne rabimo posebej izvajati deskriptivne statistike. V zavihku »Compare models« lahko naštejemo modele s katerimi bi radi našega izbranega primerjali in s kakšnimi kriteriji bi ta primerjava potekala (ponavadi izberemo »Accuracy«).

Target

Spol

Covariates

Motivacija

Anksioznost

Analysis

Data

Split set 0.7

Training Model

Model pls

Tuning parameters

Methods cv

Number 10

Repeats 5

Tune Length 10

Plots

☒ Box plot
 ☒ Density plot
 ☒ Model selection
 ☒ Variable importance

Training set

☐ Overall statistics
 ☐ Confusion matrix
 ☐ Statistics by class
 ☐ ROC curve
 

Model ctree

☐ Calibration curve

Test set

☒ Overall statistics
 ☒ Confusion matrix
 ☒ Statistics by class
 ☒ ROC curve

Save

☐ Prediction

V rezultatih nam Jamovi javi najprej osnovne informacije o modelu: Koliko je bila velikost vzorca, koliko napovednih spremenljivk, in koliko napovedanih razredov. V našem primeru imamo vzorec velik 63, napovedujemo pa razreda »Moski« in »Zenska«. Najbolj nas zanima natančnost modela, kar izraža »Accuracy«, ki smo ga dobili 0.7, kar pomeni, da lahko z natančnostjo 70% napovemo spol na podlagi anksioznosti in motivacije. Kappa 0.2 nam pa pove, da model vseeno ni najboljši. V praksi so modeli s Kappo več od 0.4 ustrezni. V tem primeru imamo še poseben podatek »ncomp was held constant as a value 1«, kar pomeni, da ena od spremenljivk ni bila uporabljena (morda zaradi nerelevantnosti ali redundantnosti). V »Variable importance plot« vidimo, da je bila za napoved ključna le Anksioznost.

Overall statistics with test set

Accuracy	Accuracy 95% CI		Kappa
	Lower	Upper	
0.692	0.482	0.857	0.224

Confusion matrix with test set

	Predicted	
	Moski	Zenska
Moski	3	3
Zenska	5	15

## Vaje

### VAJA 1

V isti podatkovni zbirki napovej Spol, tokrat na podlagi MBTI spremenljivk (Introvertiranost, Sensing, Feeling, Judging). Je model pls bolj ali manj natančen? Koliko komponent je model izbral za optimalno? Če je katero izpustil, katero in zakaj? V razdelku compare models preglej tudi, če bi bil kakšen drug model bolj optimalen.

## STROJNO UČENJE NAPOVEDOVANJE ZVEZNE

V Jamovi ne najdem ustreznega modula, a se bi mi zdelo super, da povemo alternativo regresije (napovedi zvezne) z naprednejšo metodo, npr. RF sprejme za vhod tudi nominalne spremenljivke.

## ČASOVNE VRSTE

Podobno kot zgoraj. Je sicer v Snow modulu, a je nerodna knjižnica brez programiranja.

## TEMPLATE

V nadaljevanju želimo preveriti, ali se povprečni spremenljivk »Q21c« — »Oblikovanje idej« in »Q21d« — »Preizkušanje idej« statistično značilno razlikujeta. V kolikor gre za dve spremenljivki znotraj istega vzorca učiteljic in učiteljev, je primerno uporabljati t-preizkus za odvisne vzorce. V meniju »Analyze« izberemo podmeni »tT-tTesti« in možnost »tT-tTest odvisnih vzorcev«. V okno »Parjene spremenljivke« vstavimo dve spremenljivki, ki ju želimo primerjati (vrstni red pri tem ni pomemben). Ti dve se bosta izpisali ena zraven druge (Slika 48). Jamovi avtomatično izbere ukaz »tT-test« ( ).

Slika 48: Pogled na t-preizkus odvisnih vzorcev.

### 4.3.4.1.1 PREVERJANJE POGOJEV

Preden pregledamo izpisane rezultate, preverimo pogoj normalnosti porazdelitve razlik med vrednostmi spremenljivk. Izberemo torej ukaz »Test normalnosti« ( ). Izpišejo se sledeči rezultati:

Ponazoritev 88: Tests of Normality

	statistic	p		
Q21c	Q21d	Shapiro-Wilk	0.639	< .001
		Kolmogorov-Smirnov	0.426	< .001
		Anderson-Darling	155	< .001

Opomba. Additional results provided by moretests

Iz preglednice je razvidno, da so vsi preizkusi normalnosti statistično značilni ( $p < 0,001$ ), kar kaže na to, da razlike niso normalno porazdeljene in je uporaba t-preizkusa za odvisne vzorce neupravičena. Z namenom prikaza preizkusa pa bomo vseeno nadaljevali z izvedbo t-preizkusa za odvisne vzorce.

Jamovi ne predvideva možnosti preizkušanja enakosti varianc za t-preizkuse za odvisne vzorce.

#### 4.3.4.1.2 REZULTATI PREIZKUSA

Potem, ko se prepričamo, da je uporaba t-preizkusa za odvisne vzorce upravičena, si oglejmo izpis rezultatov:

Ponazoritev 89: T-tTest odvisnih vzorcev

		statistika	df	p		
Q21c	Q21d	tT-tTest				
				6.27	810	< .001

Izpis simbolov:

Statistika ali t označuje vrednost t-preizkusa odvisnih vzorcev.

df označuje prostostne stopnje t-preizkusa.

p označuje stopnjo statistične pomembnosti t-preizkusa odvisnih vzorcev.

Iz preglednice razumemo, da so razlike med povprečjema spremenljivke »Q21c« in »Q21d« statistično značilne ( $p < 0,001$ ). Da preverimo, katera spremenljivka ima višje povprečje, si oglejmo opisno statistiko. Izberemo ukaz »Opisne statistike« ( ), in, če želimo tudi grafično predstavitev podatkov, izberemo tudi ukaz »Opisni izrisi« ( ). Dobimo sledeči izpis rezultatov in na Sliki 49 :

Ponazoritev 90: Opisne statistike

	N	Povprečna vrednost	Mediana	SD	SN	
Q21c	811	4.14	4	0.717	0.0252	
Q21d	811	4.03	4	0.749	0.0263	

Slika 49: Grafična ponazoritev razlik v povprečju med spremenljivkama »Q21c« in »Q21d«.

Iz preglednice lahko razumemo, da je povprečna vrednost spremenljivke »Q21c« ( $M = 4,14$ ;  $SD = 0,717$ ) statistično višja od povprečja spremenljivke »Q21d« ( $M = 4,03$ ;  $SD = 0,749$ ).

#### 4.3.4.1.3 VELIKOST UČINKA

Iz zgornje preglednice je vidno, da so učiteljice in učitelji višje ocenili svoje sposobnosti oblikovanja idej ocenili višje od preizkušanja idej. Čeprav je t-preizkus pokazal statistično značilne razlike v povprečjih ( $t(810) = 6,27$ ;  $p < 0,001$ ), želimo ugotoviti, ali je razlika med njima, ki jo lahko izračunamo z izbiro ukaza »Povprečna razlika« ( ), velika, zato si oglejmo mero velikosti učinka. To izberemo z ukazom »Velikost učinka« ( ). Izpis je torej sledeči:

Ponazoritev 91: T-tTest odvisnih vzorcev

		statistika	df	p	Povprečna razlika	Razlika SN
Velikost učinka						
Q21c	Q21d	tT-tTest				
				6.27	810	< .001
0.110	0.0175	Cohenov d			0.220	



#### Primer razlage

Rezultat t-preizkusa za odvisne vzorce ( $t = 6,27$ ;  $g = 810$ ;  $2P < 0,001$ ) kaže, da obstajajo statistično pomembne razlike v odgovorih na vprašanje trditvipostavki o oblikovanju idej in preizkušanju idej obstajajo statistično pomembne razlike. Analiza rezultatov kaže, da učiteljice in učitelji v povprečju večji pomen pripisujejo vplivu oblikovanja idej ( $M = 4,14$ ;  $SD = 0,717$ ) kot preizkušanju idej ( $M = 4,03$ ;  $SD = 0,749$ ). Mera velikosti vpliva (Cohenov  $d = 0,220$ ) kaže na majhen vpliv med spremenljivkama.

#### 4.3.4.2 VAJE

##### VAJA 1

- T-preizkus za odvisne vzorce.
- Preizkusili smo, ali obstajajo statistično značilne razlike v povprečju odgovorov na vprašanji postavki »Q23c« - »Samostojno delovanje za doseganje ciljev« in »Q23d« — »Načrtovanje kratko-, srednje- in dolgoročnih ciljev« obstajajo statistično značilne razlike.
- Predstavimo dobljene rezultate.
- Napišite obrazložitev.

Ponazoritev 92: T-tTest odvisnih vzorcev

		statistika	df	p	Velikost učinka	
Q23c	Q23d	tT-tTest		15.8	802	< .001
	Cohenov d	0.557				

Ponazoritev 93: Opisne statistike

	N	Povprečna vrednost	Mediana	SD	SN	
Q23c	803	4.19	4		0.756	0.0267
Q23d	803	3.76	4		0.823	0.0290

Obrazložitev:

---

---

---

---

---

---

---

---

---

—