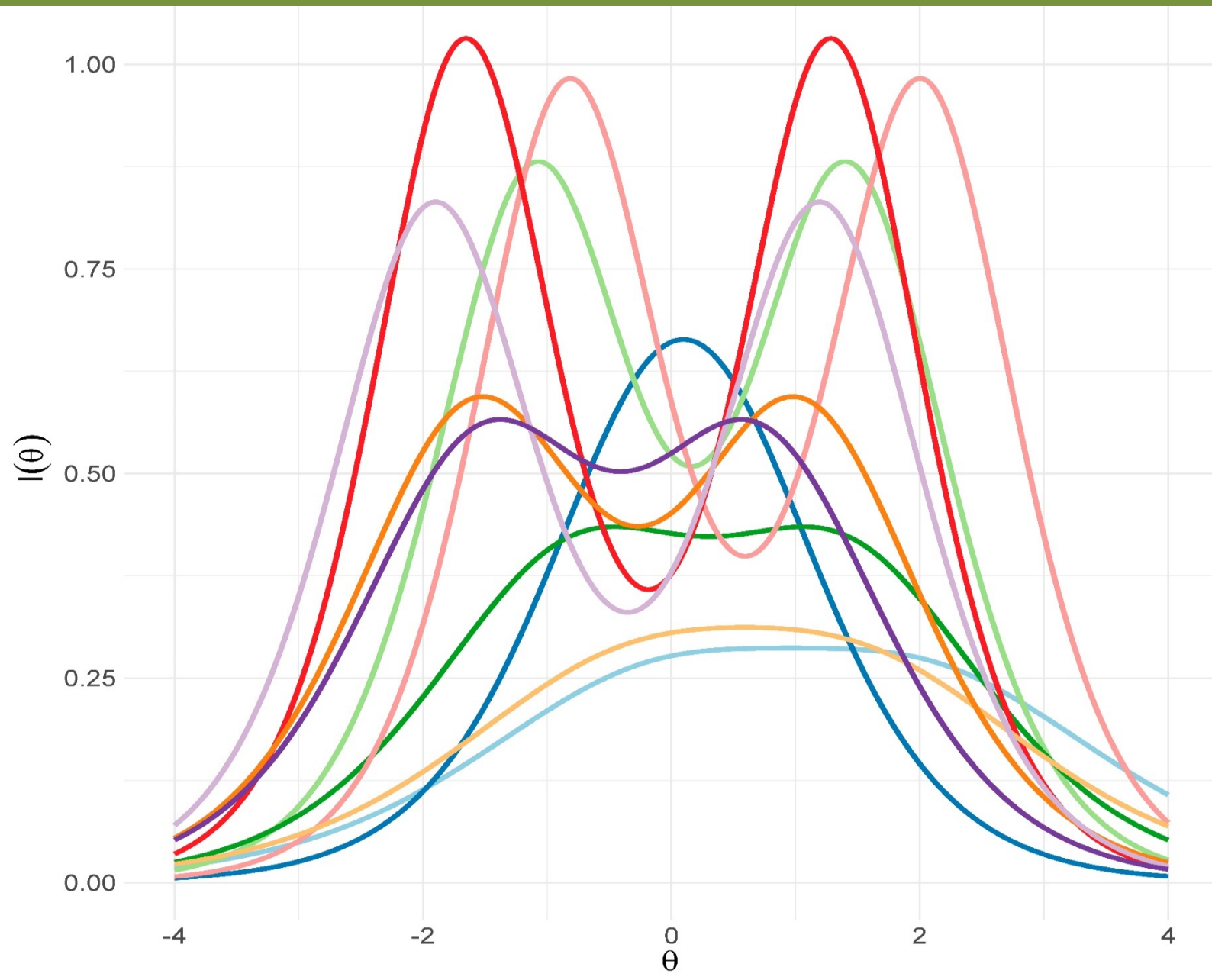
# >Jednodimenzionalni.IRT.modeli()

# >sa_primenom_u(R-u)




#Bojan Janičić

FILOZOFSKI FAKULTET · NOVI SAD ·

Bojan Janičić

# Jednodimenzionalni IRT modeli sa primenom u R-u

UNIVERZITET U NOVOM SADU
FILOZOFSKI FAKULTET NOVI SAD
21000 Novi Sad
Dr Zorana Đinđića 2
www.ff.uns.ac.rs

*Za izdavača*
Prof. dr Ivana Živančević Sekeruš

Bojan Janičić

**Jednodimenzionalni IRT modeli sa primenom u R-u**

*Recenzenti*
Prof. dr Oliver Tošković
Prof. dr. Petar Čolović

*Lektura*
Aleksandra Janičić

*Dizajn korica*
Bojan Janičić

*Tehnička priprema*
Bojan Janičić

ISBN 978-86-6065-861-8

Novi Sad, 2024.

# Sadržaj

# 1 Teorija odgovora na stavke

Iako je klasična teorija testa bila osnova za izradu najvećeg broja testova od njihovog nastanka do danas, Teorija odgovora na stavke ubrzano uzima primat u izradi sve većeg broja testova i postaje teorijska osnova merenja. Razlog za ovo su teorijski opravdaniji principi merenja i veći potencijal za rešavanje praktičnih problema merenja poput jednačenja merenja, konstrukcije adaptivnih testova[1] i ispitivanja diferencijalnog funkcionisanja stavki, odnosno provere invarijantnosti merenja (Embretson & Reise, 2000).

Za početak, osvrnućemo se na sam naziv ove teorije. U engleskom jeziku on glasi *Item Response Theory*, iako se češće upotrebljava skraćenica *IRT.* Na srpskom jeziku ovaj naziv prevodi se kao Teorija ajtemskog odgovora, Teorija stavskog odgovora ili Teorija odgovora na stavke. Ovaj poslednji prevod se čini najboljim i najvernije odslikava suštinu pa ćemo u ovoj knjizi koristiti njega. Međutim, kao skraćeni naziv koristićemo englesku skraćenicu *IRT* jer se ona najčešće koristi i na našem jeziku.

IRT nije jedan model već predstavlja (svakodnevno rastuću) porodicu matematičkih modela kojima se nastoji objasniti odnos između latentnih osobina i manifestnog ponašanja. Kada govorimo o manifestnom ponašanju najčešće mislimo na testovno ponašanje, odnosno, odgovaranje na stavke testova i upitnika, ali se može odnositi i na druge oblike manifestnog ponašanja.

Za razliku od klasične testne teorije (KTT) koja pripada tzv. psihometrijskoj tradiciji ***skorovanja*** testova, odnosno rezultata ispitanika na testovima, IRT pripada onome što nazivamo tradicijom ***skaliranja***. U KTT ukupni testni skor ispitanika računa se sabiranjem njegovih skorova na stavkama i on je pokazatelj pravog skora, odnosno ukazuje u kojoj meri ispitanik poseduje merenu osobinu. U IRT mera ispitanika računa se kao lokacija na kontinuumu latentne osobine koju test meri i određuje se statistički na osnovu *sklopa odgovora*[2] na stavke, uzimajući u obzir i parametre stavki u skladu sa odabranim modelom. Ono što je zanimljivo, parametar težine stavke takođe je definisan lokacijom na istoj skali, a to donosi izvesne prednosti prilikom konstrukcije testova o kojima će kasnije biti reči. Zbog ove osobine IRT se vrlo često naziva i Teorijom latentne crte (engl. *Latent Trait Theory*)(Fajgelj, 2020).

U centru pažnje IRT su stavke i odgovori ispitanika na njih. Ono što svi IRT modeli modeliraju i predviđaju je verovatnoća određenog odgovora ispitanika na stavku u zavisnosti od nivoa merene osobine

[1] Više o adaptivnom testiranju možete naći u odeljku 7.1.

[2] Sklop odgovora predstavlja niz odgovora ili oznaka za odgovore koje nam pokazuju kako je ispitanik odgovorio na stavke nekog testa. Npr. sklop odgovora ispitanika koji je odgovorio tačno ili potvrdno na prve tri od pet binarno skorovanih stavki (koliko sadrži test) mogao bi se predstaviti ovako: 11100. Sklop odgovora ispitanika koji je odgovorio tačno na prve dve i na četvrtu stavku izgledao bi ovako: 11010.

kod ispitanika i težine stavke/kategorije (Fajgelj, 2020). U IRT se i ne govori o „skoru" ispitanika već o „meri osobine"(Fajgelj & Janičić, 2008).

Imajući u vidu da je svim IRT modelima zajedničko da predviđaju verovatnoću određenog odgovora ispitanika na stavku u zavisnosti od prisustva merene latentne osobine, koreni IRT mogu se naći u radovima psihofizičara sa kraja XIX veka (Bock, 1997; Fajgelj, 2020; Van Der Linden & Hambleton, 1997) koji su se bavili odnosima fizičkih karakteristika stimulusa i njihove percepcije od strane ispitanika. Osim toga, koreni IRT mogu se videti i u radovima Terstona (Thurstone) sa početka XX veka i njegovim pokušajima da pozicionira zadatke iz Binnet-Simonove skale na skali kalendarskog uzrasta (Bock, 1997; Thissen & Steinberg, 2020) ili u njegovim istraživanjima percepcije stimulusa (Van Der Linden & Hambleton, 1997). Ipak, može se reći da se najbitniji doprinos Terstona skaliranju, a samim tim i razvoju IRT, ogleda u njegovom radu na *zakonu komparativnog suđenja*[3] (Thurstone, 1927). Takođe, u svojim radovima Terston je često koristio normalnu ogivu (kumulativnu krivu normalne raspodele)[4], u objašnjavanju povezanosti posmatranih pojava (Van Der Linden & Hambleton, 1997). Upotreba normalnih ili logističkih ogiva je takođe jedna od bitnih karakteristika IRT modela, odnosno to je osnovna funkcija kojom se objašnjava povezanost latentne osobine i verovatnoće izbora određenog odgovora na stavku.

Put ka formulisanju prvih pravih IRT modela dalje je popločan radovima Mariona Vebstera Ričardsona (Richardson, 1936), Džordža Endrua Fergusona (Ferguson, 1942), Daniela Najdžela Loulija (Lawley, 1943) i Ledijarda R. Takera (Tucker, 1946). Polovinom XX veka, Pol Lazarsfeld (Lazarsfeld, 1950) i Frederik Lord (Lord, 1952, 1953), reklo bi se nezavisno jedan od drugog, pominju modeliranje verovatnoće odgovora na stavku zavisnosti od prisustva *latentne crte* kod ispitanika, a to je upravo jedna od osnovnih karakteristika savremenih IRT modela. Frederic Lord je definisao ***dvoparametarski model*** za testove koji imaju binarno skorovane stavke (Lord, 1953). Njegov model je koristio normalnu ogivu, a nešto kasnije Alan Birnbaum (1968) definisao je logistički model koji je uključivao i parametar (pseudo)pogađanja. Džordž Raš (Rasch, 1960) je formulisao ***jednoparametarski logistički model*** (koji je i danas poznatiji pod nazivom

[3] Zakon komparativnog suđenja je merni model određivanja skalnih vrednosti intenziteta nekog stimulusa (objekta) na kontinuumu koji može, ali i ne mora biti fizički merljiv. Skaliranje se obavlja poređenjem objekata po parovima. Objekti se porede svaki sa svakim po prisustvu procenjivanog svojstva. Ako je procenjivano svojstvo objekata „zanimljivost" i kontinuum na koji će biti smešteni objekti biće „zanimljivost". Stimuluse procenjuje uzorak procenjivača, a na osnovu tih procena formira se kvadratna matrica. U ćelijama te matrice nalaze se proporcije procena u kojima je intenzitet procenjivanog svojstva nekog objekta X (predstavljeni u redovima matrice) procenjen kao jači od intenziteta objekta Y (u kolonama matrice). Na osnovu toga koliko često je intenzitet procenjivanog atributa nekog objekta procenjivan kao jači u odnosu na druge objekte, određuje se njegova lokacija na kontinuumu procenjivanog atributa i formira se intervalna skala. Više o zakonu komparativnog suđenja kao modelu merenja možete naći u knjizi *„Psihometrija – Metod i teorija psihološkog merenja"* (Fajgelj, 2020).

[4] Videti: Slika 6 – Uporedni prikaz standardne normalne i standardne logističke distribucije i njihovih ogiva.

***„Rašov model"***).

Iako je IRT potekla iz kognitivnog testiranja i prvi modeli bazirani na njoj su bili namenjeni dihotomno skorovanim kognitivnim stavkama, dalji razvoj IRT modela usledio je sedamdesetih i osamdesetih godina XX veka kada su formulisani modeli za politomne stavke. Prvi je bio dvoparametarski ***Model stepenovanog odgovaranja*** u normalnoj i logističkoj formi[5] (engl. *Graded response model* – GRM) Fumiko Samedžime (Samejima, 1969). Zatim je usledio Bokov (Bock, 1972) ***Nominalni model*** (engl. *Nominal response model*) namenjen ocenjivanju parametara stavki i latentne sposobnosti kada su ponuđeni odgovori dati u obliku liste od dve ili više međusobno isključivih kategorija.

Andersen (1977) i Andrič (Andrich, 1978a, 1978b) su predstavili slične modele namenjene politomnim stavkama sa uređenim kategorijama (poput Likertovih skala). Ovaj model sada se naziva ***Model skala procene*** (engl. *Rating Scale Model* – RSM) i predstavlja proširenje Rašovog modela. Radi se o jednoparametarskom modelu kod kojeg sve stavke imaju jednaku skalu odgovora (u smislu da su skalne vrednosti parametara težina kategorija jednako udaljene jedne od drugih – što ne znači nužno da su intervali između tih skalnih vrednosti unutar stavki jednaki). Andersen je kasnije (1983) predložio manje restriktivnu verziju ovog modela.

Još jedno proširenje Rašovog modela na politomne stavke sa uređenim kategorijama bio je ***Model stepenovanog ocenjivanja*** (engl. *Partial Credit Model* – PCM) Mastersa i Rajta (Masters, 1982; Wright & Masters, 1982). Za razliku od RSM ovaj model ne sadrži pretpostavku o ekvidistantnosti lokacija kategorija. Eidži Muraki je 1992. godine formulisao ***Generalizovani model stepenovanog ocenjivanja*** (engl. *Generalized Partial Credit Model* – GPCM) (Muraki, 1992). Ovaj model je modifikacija PCM modela Mastersa i Rajta, ali relaksira pretpostavku o jednakom parametru diskriminativnosti za sve stavke, pa u suštini predstavlja dvoparametarski model.

Dolazi i do razvoja multidimenzionalnih IRT modela (MIRT modeli). Ovi modeli su namenjeni testovima kod kojih u procesu odgovaranja na stavke učestvuje više latentnih osobina. MIRT modele možemo razvrstati u dve grupe. Prvu grupu čine linearni modeli. Ovi modeli prave linearnu kombinaciju procena osobina koje učestvuju u procesu odgovaranja (Reckase, 2009). Najčešće se nazivaju ***kompenzatornim*** jer se isti rezultat može dobiti različitim kombinacijama procenjenih latentnih osobina. Drugim rečima, dva ispitanika mogu ostvariti isti rezultat ako prvi ispitanik ima nizak nivo prve osobine, a visok nivo druge, a drugi ispitanik visok nivo prve osobine, a nizak nivo druge. Nizak nivo jedne osobine kompenzuje se višim nivoom druge.

Drugi tip MIRT modela su ***nekompenzatorni*** ili ***parcijalno kompenzatorni*** modeli. Kod ovih modela se kognitivni zadatak odgovaranja na test deli na delove (u zavisnosti od merenih dimenzija) i za svaki se koristi poseban jednodimenzionalni model (Reckase, 2009). Kod ovih modela verovatnoća predviđanog odgovora na određenu stavku ne može biti viša od najviše predviđene za bilo koju od merenih osobina. Dakle, verovatnoća odgovora definisana je najviše prisutnom osobinom ispitanika (od svih merenih).

[5] O razlikama između normalnih i logističkih modela više reči će biti u odeljku 2.3.2.

U praksi se češće upotrebljavaju kompenzatorni modeli (Reckase, 2009). S obzirom na to da MIRT modeli nisu predmet ove knjige, nećemo ih detaljno opisivati. Napomenućemo samo da su razvijeni modeli kako za binarno tako i za politomno skorovane stavke, u logističkoj i normalnoj formi, te u kompenzatornoj i parcijalno kompenzatornoj varijanti. Za razvoj ovih modela bitni su radovi Rejmonda Adamsa i saradnika, Suzan Vajtli (sada Embretson), Džejmsa B. Simpsona, Erika Marisa, Eidžija Murakija i Džejmsa Karlsona, Jirgena Rosta, Lihue Jao i Ričarda Švarca (Adams i ostali, 1997; Embretson, 1984, 1991; Maris, 1995; Muraki & Carlson, 1995; Rost, 1990; Sympson, 1978; Whitely, 1980; Yao & Schwarz, 2006).

Na početku je rečeno da se IRT modeli sve češće koriste, ali ako se uzme u obzir da su prvi modeli kreirani još pedesetih godina XX veka, može se reći da se koriste manje nego što bi se moglo očekivati s obzirom na prednosti koje donose. Razlog tome je što je njihovo izračunavanje matematički mnogo komplikovanije nego izračunavanje modela klasične testne teorije. Ovo posebno važi za IRT modele zasnovane na normalnoj ogivi. Zbog toga je bilo potrebno da računarska tehnologija postane dostupnija kako bi ovi modeli zaživeli u većoj meri. Pored računarskog hardvera, problem je bio i specifičan softver potreban za njihovo izračunavanje. Popularni komercijalni statistički paketi nisu podržavali IRT analize i u ovom trenutku ih još uvek ne podržavaju. Za IRT analize prvo se pojavio komercijalni softver poput programa „Winsteps“, „FACETS“, „PARSCALE“, „IRTPRO“, „Mplus“, „flexMIRT“, „EQSIRT“... Međutim, njegova cena je učinila da bude nedostupan širem krugu istraživača (Y.-J. Choi & Asilkalkan, 2019).

Pojava R programskog jezika i okruženja za statistička izračunavanja (R Core Team, 2023) koji je u potpunosti besplatan i pojava njegovih brojnih paketa namenjenih IRT modelovanju učinila je IRT analizu mnogo dostupnijom. Do 2019. godine bilo je dostupno 45 R paketa namenjenih IRT modelovanju (Y.-J. Choi & Asilkalkan, 2019).[6]

Tabela 1 – Lista odabranih R paketa za IRT analizu sa odabranim mogućnostima

| R paket | Raš | 2PL | 3PL | 4PL | PC | RS | GPC | GR | NR |
|---|---|---|---|---|---|---|---|---|---|
| cacIRT (Lathrop, 2015) | ✓ | ✓ | ✓ | | | | | | |
| catR (Magis & Barrada, 2017; Magis & Raîche, 2012) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| CDM (George i ostali, 2016) | ✓ | ✓ | | | ✓ | ✓ | | | |
| DFIT (Cervantes, 2017) | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | |
| difNLR (Drabinova & Martinkova, 2017) | ✓ | ✓ | ✓ | ✓ | | | | | |
| difR (Magis i ostali, 2010) | ✓ | ✓ | ✓ | | | | | | |
| equateIRT (Battauz, 2015) | ✓ | ✓ | ✓ | | | | | | |
| eRm (Mair & Hatzinger, 2007) | ✓ | | | | ✓ | ✓ | | | |
| EstCRM (Zopluoglu, 2022) | | | | | | | | | |

[6] Za detaljan pregled R paketa iz različitih oblasti koristan je i paket „ctv“ (Zeileis i ostali, 2023). Ovaj paket omogućava pregled i instalaciju paketa iz različitih oblasti. Npr. komandama:

```
install.packages("ctv")
ctv::install.views("Psychometrics")
```

instalirali biste sve R pakete relevantne za psihometriju. Više o ovom paketu možete saznati na linku https://cran.r-project.org/web/views/, a spisak i kratak opis psihometrijskih paketa možete naći na sledećoj adresi: https://cran.r-project.org/web/views/Psychometrics.html.

| R paket | Raš | 2PL | 3PL | 4PL | PC | RS | GPC | GR | NR |
|---|---|---|---|---|---|---|---|---|---|
| immer (Robitzsch & Steinfeld, 2022) | ✓ | ✓ | | | ✓ | ✓ | | | |
| irtoys (Partchev & Maris, 2022) | ✓ | ✓ | ✓ | | | | | | |
| kequate (Andersson i ostali, 2013) | | ✓ | ✓ | | | | ✓ | ✓ | |
| LNIRT (Fox i ostali, 2021) | | | | | | | ✓ | ✓ | |
| lordif (S. W. Choi i ostali, 2016) | ✓ | ✓ | | | | | | | |
| ltm (Rizopoulos, 2006) | ✓ | ✓ | ✓ | | | | ✓ | ✓ | |
| mirt (Chalmers, 2012) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| mirtCAT (Chalmers, 2016) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MLCIRTwithin (Bartolucci & Perugia, 2019) | ✓ | ✓ | | ✓ | | | ✓ | | |
| MultiLCIRT (Bartolucci i ostali, 2017) | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| pairwise (Heine, 2023) | ✓ | | | | ✓ | | | | |
| pcIRT (Hohensinn, 2018) | ✓ | | | | | ✓ | | | |
| PP (Reif & Steinfeld, 2021) | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | |
| psychomix (Frick i ostali, 2012) | ✓ | | | | | | | | |
| SNSequate (González, 2014) | ✓ | | ✓ | | | | | | |
| TAM (Robitzsch i ostali, 2022) | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| xxIRT (Luo, 2019) | | | ✓ | | | | ✓ | | |

(Preuzeto i prilagođeno iz: Y.-J. Choi & Asilkalkan, 2019)

*Legenda: Raš – Rašov jednoparametarski logistički model, 2PL – Dvoparametarski logistički, 3PL – Troparametarski logistički; 4PL – Četvoroparametarski logistički; PC – Model stepenovanog ocenjivanja; RS – Model skala procene; GPC – Generalizovani model stepenovanog ocenjivanja; GR – Model stepenovanog odgovaranja; NR – Nominalni model*

U tabeli (Tabela 1) prikazani su samo neki od R paketa i izlistane samo neke od njihovih mogućnosti. Neki od ovih paketa, poput *ltm* (Rizopoulos, 2006) i *mirt* (Chalmers, 2012), npr. imaju i mogućnost modelovanja multidimenzionalnih IRT modela, neki poput paketa *mirt*, *lordif* (S. W. Choi i ostali, 2016), *difNLR* (Hladka & Martinkova, 2023), *difR* (Magis i ostali, 2010) mogu se koristiti za ispitivanje diferencijalnog funkcionisanja stavki i testa, a neki poput *equateIRT* (Battauz, 2015), *SNSequate* (González, 2014), *kequate* (Andersson i ostali, 2013) mogu poslužiti za jednačenje testova. Takođe, moguće je proceniti i veliki broj modela koji u dosadašnjem delu teksta nisu navedeni jer izlaze iz nameravane oblasti koju ova knjiga obrađuje.

Iako R polako postaje standard statističke obrade podataka, još uvek postoji izvestan otpor korisnika iz nematematičkih oblasti. Verovatan razlog tome je nedostatak grafičkog korisničkog interfejsa i potreba da se naredbe „kuckaju“, te to što sprovođenje statističkih analiza češće liči na programiranje nego na ono na šta je većina istraživača navikla. Ono što ohrabruje je to što i proizvođači komercijalnog softvera sve više uviđaju potencijal R-a (a ne zaboravimo ni Python) za statistička izračunavanja i u poslednje vreme ga integrišu u svoje programske pakete. Tako na primer, IBM SPSS Statistics ima proširenje koje omogućava rad u R-u kroz grafički korisnički interfejs, pa tako i IRT analize. Kada pominjemo SPSS, vredelo bi pomenuti i besplatne makroe za IRT analizu u okviru ovog statističkog paketa. To su *KakaoBEJZ* (Fajgelj & Janičić,

2008) i *KakaoMIX* (Fajgelj & Janičić, 2009) namenjeni proceni jednoparametarskih i dvoparametarskih modela za binarno skorovane ili za mešavinu binarno i politomno skorovanih stavki[7]. Međutim, ni ovi makroi nemaju grafički korisnički interfejs, a sa druge strane mogućnosti su im slabije od gore nabrojanih R paketa. S druge strane, ako je neko navikao i vezan je za rad u SPSS-u, makroi mu mogu pružiti osnovne informacije potrebne za analizu i kreiranje testa.

## *1.1 Nova pravila merenja*

Zanimljivo poređenje IRT modela i modela baziranih na klasičnoj testnoj teoriji daju Van der Linden i Hamblton (Van Der Linden & Hambleton, 1997). Oni porede svako psihološko merenje sa eksperimentom. Psihološko merenje može se posmatrati kao serija eksperimenata u kojima istraživač beleži niz rezultata ispitanika na svakom od njih. Da bismo otklonili eksperimentalne greške potrebna je rigorozna kontrola uslova. Ona se sprovodi na tri načina: ujednačavanjem uslova u kojima se eksperiment sprovodi, randomizacijom ili statističkom kontrolom. Ujednačavanjem postižemo da svi ispitanici rade eksperiment pod istim uslovima i na taj način bilo kakve razlike koje nastanu ne mogu biti plod nekontrolisanih i neželjenih varijabli. Kada nije moguće kontrolisati neželjene varijable na način da ostvare isti uticaj u svim slučajevima, randomizacijom se postiže da njihov uticaj u proseku bude isti i njihov efekat ne bude sistematski. Statistička kontrola obavlja se naknadno i može se primenjivati kada prethodne dve nisu moguće. Za ovu metodu bitno je da sve varijable za koje smatramo da mogu uticati na rezultate budu izmerene. Takve varijable se zatim uvrštavaju u matematički model i njihov uticaj na rezultate se kontroliše, odnosno eliminiše. Ovo je moguće ako takav model odgovara podacima (saglasan je sa njima).

Nedostatak procedure ujednačavanja eksperimentalnih uslova je smanjena mogućnost generalizacije. Ovakvi rezultati mogu se generalizovati samo na situacije u kojima su kontrolisani isti uslovi kao i u samom eksperimentu (Van Der Linden & Hambleton, 1997).

Statistička kontrola nema ovakvo ograničenje. Van der Linden i Hamblton (Van Der Linden & Hambleton, 1997) dalje kažu da se klasična testna teorija oslanja na „ujednačavanje“ i „randomizaciju“, a pravi skor koji je centralna tačka ove teorije važi samo za određeni test. Sve sistematske varijacije u opaženim skorovima potiču od merene osobine, a ostali izvori varijacije su potpuno slučajni. Ujednačavanje se postiže strogom standardizacijom, a procedura randomizacije odnosi se na izbor ispitanika. Merne karakteristike testa određuju se na reprezentativnom uzorku i važe za populaciju koju on reprezentuje. Pravi skorovi ispitanika su oni koje oni ostvare na tom konkretnom testu, važe samo za taj test i razlikuju se od pravih skorova na nekom drugom testu čak i ako imaju isti predmet merenja. Različite stavke u dva različita testa imaju i različita svojstva koja sa aspekta eksperimentalne paradigme predstavljaju neželjene i nekontrolisane faktore i ometaju poređenje.

[7] Dostupni na https://www.stanef.in.rs/statistics/macros/75-cocoa ili zahtevom na mejl janicic@ff.uns.ac.rs

S druge strane, IRT se oslanja na statističku kontrolu. Mere ispitanika određuju se na osnovu modela koji uzima u obzir razlike među stavkama, a koje se odnose na njihovu težinu, diskriminativnost, podložnost pogađanju ili efekat nepažljivosti (omaške) (Hambleton i ostali, 1991; Van Der Linden & Hambleton, 1997). Za razliku od modela klasične testne teorije, IRT modeli nemaju ograničenu generalizabilnost i omogućavaju da se lakše reše problemi poput jednačenja testova, konstrukcije alternativnih formi i utvrđivanja jednakosti merenja.

IRT je promenila mnoga opštepoznata i prihvaćena pravila merenja bazirana na klasičnoj testnoj teoriji (KTT). Embretsonova (1996), a kasnije Embretsonova i Ris (Embretson & Reise, 2000) daju sažeti prikaz razlika starih (KTT) i novih (IRT) pravila merenja.

Reći ćemo nešto više o svakom od ovih pravila.

**Pravilo 1.** Ovo pravilo se odnosi na standardne greške merenja koje su prema IRT različite za različite skorove, ali su generalizabilne na različite populacije. Standardna greška merenja je vrednost suprotna pouzdanosti, odnosno preciznosti merenja. Skor ispitanika u KTT i mera osobine u IRT predstavljaju samo procene pravih vrednosti. Standardna greška merenja se u oba slučaja koristi za izgradnju intervala poverenja u kojem se sa određenim stepenom sigurnosti nalazi pravi skor ili mera.

KTT i IRT se razlikuju u pogledu na standardne greške merenja. U KTT smatra se da je standardna greška merenja jednaka za sve skorove unutar jedne populacije ili prema novijem shvatanju Lorda i Novika (Lord & Novick, 2008), standardna greška merenja najveća je kod srednjih skorova, a najmanja kod ekstremnih. Naime, kako je standardna greška merenja u stvari standardna devijacija dobijenih skorova oko pravog, kod ispitanika čiji je pravi skor blizak najnižem mogućem skoru na testu („pod“) ili najvišem mogućem skoru („plafon“) to variranje je nužno ograničeno sa jedne strane, pa je zato manje.

Osim toga, u KTT standardna greška merenja je karakteristična za određenu populaciju/uzorak. Za njeno izračunavanje (izraz [ 1 ]) potrebno je poznavati aritmetičku sredinu i standardnu devijaciju sirovih skorova u populaciji/uzorku.

$$sg_m = \sqrt{1 - r_{tt}}\sigma \qquad [\,1\,]$$

gde je $r_{tt}$ pouzdanost testa, a $\sigma$ standardna devijacija sumacionih skorova.

S druge strane, u IRT modelima izračunava se posebna standardna greška za svaki nivo osobine (skor ili sklop odgovora). Za razliku od KTT, u IRT standardne greške merenja su najmanje za srednje nivoe osobine, a najveće za ekstremne. Tačnije, standardne greške merenja su najmanje kada su stavke (ili test) prilagođene nivou osobine ispitanika i kada su diskriminativne (videti odeljak 7 o informativnosti). Pošto IRT ne računa jednu grešku merenja za čitavu populaciju već se one računaju za svaki nivo osobine posebno, one neće zavisiti od populacije. Greške merenja mogu se generalizovati na različite populacije. Poređenje mera ispitanika na osnovu ovakvih standardnih grešaka biće i preciznije (Embretson & Reise, 2000).

**Pravilo 2.** Pravilo se tiče odnosa dužine testa (broja stavki) i njegove pouzdanosti. Prema IRT kraći

testovi mogu biti pouzdaniji.

U jednačine pouzdanosti tipa interne konzistencije koje se zasnivaju na KTT (npr. koeficijenti α, S–B, $\lambda_6$) ugrađena je pretpostavka da pouzdanost testa zavisi od broja stavki. Ako se test dopuni homogenim stavkama (paralelnim indikatorima) pouzdanost će biti veća, a ako se test skrati pouzdanost će biti manja (Embretson & Reise, 2000; Fajgelj, 2020).

U IRT modelima to nije slučaj. Kao što je rečeno, standardna greška merenja (koja je veličina recipročna pouzdanosti) prema IRT biće manja što je test prilagođeniji nivou merene osobine kod ispitanika. Standardne greške merenja stavki mogu se sabirati[8] da bi se dobile standardne greške merenja za svaki nivo osobine.Ako test prilagodimo ispitaniku biranjem stavki koje su po težini bliske njegovom nivou osobine, taj nivo osobine možemo preciznije izmeriti kraćim testom. Zadavanje prelakih ili preteških stavki bi samo povećalo grešku merenja pa njihovim izbacivanjem (dakle, skraćivanjem testa) možemo ostvariti preciznije, odnosno pouzdanije merenje. Ovaj princip se primenjuje u adaptivnom testiranju (Embretson & Reise, 2000).

**Pravilo 3**. Ovo pravilo odnosi se na uporedivost skorova ispitanika dobijenih na različitim formama testova, a oslanja se na *pravilo 2*. Prema KTT, da bismo mogli porediti rezultate ispitanika na različitim formama testova, te forme moraju se jednačiti, odnosno moraju biti paralelne. Striktno paralelni indikatori moraju imati jednake prave skorove i jednake varijanse greške. Pošto su im jednaki pravi skorovi biće jednake i varijanse pravih skorova, a pošto su i varijanse greške jednake, biće jednake i ukupne varijanse (Fajgelj, 2020). Iz toga sledi da će imati i jednake aritmetičke sredine, odnosno *težine* (pošto je aritmetička sredina skorova uzorka ispitanika na indikatoru pokazatelj njegove težine). Ako su testovi paralelni u navedenom smislu onda su i skorovi ispitanika na njima jednaki, odnosno uporedivi.

Ako se težine razlikuju, indikatori neće biti paralelni, bar ne striktno. Pošto pouzdanost testa zavisi od njegove težine, testovi različitih težina neće imati jednaku pouzdanost na istom uzorku, niti će isti test imati jednaku pouzdanost na uzorcima iz populacija koje se razlikuju po nivou merene osobine. Sve to ugrožava poređenje skorova sa različitih formi testova (Fajgelj, 2020). Pošto je pouzdanost vrednost recipročna standardnoj grešci, nejednaka pouzdanost znači i različite standardne greške. Ako i nađemo dobru ujednačavajuću transformaciju za dva testa, ostaje i dalje problem nejednakih intervala poverenja oko skorova. Takođe, nijedna procedura jednačenja nije savršena i vezana je za *grešku jednačenja*. Greška jednačenja je najmanja kod umerenih skorova (kada je test prilagođen ispitanicima), a najveća kod ekstremnih.

S druge strane, primenom *pravila 2*, upotrebom IRT modela i adaptivnog testiranja moguće je smanjiti greške merenja. Logika adaptivnog testiranja je da ispitanik radi test koji je prilagođen nivou njegove osobine. Testovi različitih ispitanika ne moraju se sastojati od istih zadataka, niti moraju biti iste dužine. Kao procena merene latentne osobine dobija se lokacija na njenom kontinuumu. Za razliku od sumacionog skora, ona neće zavisiti od dužine testa. Birajući stavke prilagođene ispitanicima na ovaj način dobijaju se ***precizne*** mere na istoj skali. Imajući to u vidu, možemo reći da će mere ispitanika koji su radili testove

[8] Ako se prvo pretvore u informativnosti – videti odeljak 7.

različitih težina, a koji su tako namerno formirani da budu prilagođeni nivoima osobine kod ispitanika – biti uporedivije (Embretson & Reise, 2000).

**Pravilo 4.** Ovo pravilo odnosi se na zavisnost procenjenih mernih karakteristika stavki i testa od uzorka. Procena mernih karakteristika stavki u KTT zavisna je od uzorka. Ista stavka (ili test) u uzorku sa niskim nivoom merene osobine može biti ocenjena kao teška, a u uzorku sa visokim nivoom osobine kao laka. Takođe, zbog zakrivljenosti distribucije skorova i nelinearnog odnosa biće poremećena i procena biserijske korelacije skorova na stavkama i ukupnog skora kao mere diskriminativnosti stavki. Zato je prilikom utvrđivanja mernih karakteristika stavki i instrumenta u celini, kada se koristi pristup KTT, izuzetno važno da uzorak bude reprezentativan za ciljnu populaciju (Embretson & Reise, 2000).

U IRT pristupu se prilikom određivanja parametara stavki statistički kontrolišu parametri ispitanika tako da je manje bitno da li se parametri određuju na uzorku sa visokim ili niskim nivoom merene osobine (videti više u odeljku 1.3).

**Pravilo 5.** Ovo pravilo odnosi se na interpretaciju testovnih skorova. U KTT pristupu testovni skorovi se obično porede sa normama za ciljnu populaciju. Često se skorovi prevode u standardizovane vrednosti ili transformišu na neku od često primenjivanih skala (IQ, T, stenajn...) što omogućava pozicioniranje ispitanika u odnosu na referentnu grupu. Na osnovu toga možemo reći koji procenat ispitanika ima viši ili niži skor od posmatranog ispitanika. Ono što se zamera ovakvoj interpretaciji je to što ne govori šta taj ispitanik zaista može da uradi (Embretson, 1996).

U IRT modelima mere ispitanika možemo interpretirati kao verovatnoću da će on neki zadatak uraditi tačno (ili da će se složiti sa nekom tvrdnjom). Korisno svojstvo IRT modela je da su težine stavki i mere ispitanika izraženi u istim jedinicama, odnosno kao lokacije na kontinuumu latentne osobine. Poredeći meru ispitanika i parametar težine stavke možemo tačno reći kolika je verovatnoća da će je ispitanik tačno rešiti. Ako su ispitanik i stavka upareni, odnosno težina stavke jednaka je meri ispitanika, ispitanik ima verovatnoću od 0,5 (odnosno 50% izraženo u procentima) da taj zadatak uradi tačno. Ako je mera ispitanika viša od težine stavke ta verovatnoća je veća, a ako je niža, onda je verovatnoća manja od 50%. S druge strane, i IRT mere je takođe moguće normirati i koristiti takav pristup interpretaciji rezultata (Embretson & Reise, 2000).

**Pravilo 6.** Ovo pravilo odnosi se na uslove za intervalni nivo dobijene skale. Razlika u nivou merene osobine dva ispitanika koji imaju 5 i 6 bodova neće biti ista u slučaju da su oba ispitanika svoje skorove ostvarili tačnim odgovorima na lake stavke i u slučaju da su bodove ostvarili odgovarajući na teže stavke. Razlika će u oba slučaja iznositi 1 bod, ali taj jedan bod neće značiti istu razliku u prisustvu merene osobine.

Takođe, ako kao primer uzmemo dva testa od kojih je jedan lak, a drugi težak, za očekivati je da će oni imati različite distribucije ukupnih skorova. Lak test će imati negativno, a težak test pozitivno zakošenu distribuciju. Zbog toga, odnos skorova istog uzorka ispitanika na dva testa neće biti linearan, a intervali među skorovima neće biti jednaki. Na teškom testu intervali između niskih skorova biće širi nego na lakom testu, a u oblasti visokih skorova situacija će biti obrnuta. Iz toga sledi da sumacioni skorovi ne dostižu intervalni nivo merenja. Da bi merenje bilo na intervalnom nivou distribucija sumacionih skorova mora biti

normalna i pravi skorovi moraju imati osobine intervalne skale (Embretson, 1996). Intervalnost skale u KTT obezbeđuje se biranjem stavki testa tako da daju normalnu distribuciju dobijenih skorova ili normalizacijom skorova, a sve pod *pretpostavkom* da se pravi skorovi u populaciji distribuiraju normalno. Da bi bila ostvarena normalna distribucija test mora biti umerene težine. S druge strane, procena težine stavki i testa u KTT zavisi od uzorka, odnosno od populacije koju on reprezentuje. Na uzorku iz druge populacije isti test može biti prelak ili pretežak pa se gubi normalnost distribucije skorova i svojstvo intervalnosti. Drugi način da se obezbedi normalna distribucija je primena neke transformacije za normalizaciju. Ove transformacije su često nelinearne, odnosno menjaju udaljenosti (intervale) između skorova (Embretson & Reise, 2000).

U IRT modelima razlika u postignuću između dva ispitanika biće ista bez obzira da li su njihove mere dobijene na lakim ili teškim stavkama. Normalnost distribucije nije potrebna jer su mere izvorno intervalnog nivoa merenja. Ako model dovoljno dobro opisuje podatke onda neće biti bitno da li su ispitanici mereni lakim ili teškim stavkama ili su parametri stavki ocenjivani na reprezentativnom uzorku ili uzorcima sa niskim ili visokim nivoima osobine (Embretson, 1996). Treba ipak imati u vidu da IRT mere ne predstavljaju apsolutnu lokaciju na kontinuumu osobine. Za ovo bi bilo potrebno postojanje univerzalne skale koja je primenljiva na sve situacije merenja, a takva skala ne postoji. Intervalnost skale posebno važi za Rašov model i to će biti detaljnije objašnjeno u odeljku 4.1.1 (deo o specifičnoj objektivnosti).

**Pravilo 7.** Ovo pravilo se odnosi na primenu stavki različitog formata ocenjivanja u jednom testu. U klasičnoj testnoj teoriji, odnosno prilikom računanja ukupnog skora, različit format ocenjivanja stavki dovodi do toga da stavke sa većom varijansom (najčešće one sa najvećom skalom odgovora) dobijaju najveći značaj. Ta pojava se naziva autoponderisanje. Zbog toga se u KTT izbegava kombinovanje stavki različitog formata ocenjivanja. Jedan način da se to prevaziđe je standardizacija, međutim, ona je specifična za uzorak, a *z* skorovi zavise od prosečnog nivoa osobine u uzorku i varijabilnosti osobine, odnosno varijabilnosti skorova (Embretson & Reise, 2000).

IRT modeli uglavnom nemaju taj problem (Embretson & Reise, 2000). Ovo se odnosi na modele namenjene politomnim stavkama (jer su ostali namenjeni samo binarno skorovanim). Takvi su npr. Model stepenovanog ocenjivanja, Generalizovani model stepenovanog ocenjivanja i Model stepenovanog odgovaranja, ali ne i Model skala procene koji zahteva da stavke budu istog formata. IRT modeli ne računaju skor već statistički nalaze lokaciju ispitanika na kontinuumu latentne osobine.

**Pravilo 8.** Ovo pravilo se odnosi na poređenje skorova razlike (promene) i oslanja se na pravilo 6. Skorovi promene su razlike u skorovima u dva sukcesivna merenja. Da bi se razlike u skorovima mogle porediti potrebno je da mera bude na intervalnom nivou merenja. Kao što je pomenuto u pravilu 6, IRT mere to obezbeđuju dok sumacioni skorovi ne (Embretson & Reise, 2000).

**Pravilo 9**. Ovo pravilo odnosi se na faktorsku analizu binarno skorovanih stavki. Problem kod faktorske analize binarnih stavki je to što maksimalne vrednosti Pearsonovih koeficijenata korelacije (odnosno ϕ koeficijenata) među stavkama zavise od sličnosti u njihovim distribucijama. Distribucija binarnih stavki zavisi od njihove težine pa je rezultat toga da se umesto latentnih varijabli kao faktori izdvajaju artefakti –

faktori težine. Drugim rečima, zbog sličnosti u distribucijama najviše koreliraju stavke koje su slične po težini. Pokušaj da se to prevaziđe je upotreba tetrahoričkih korelacija[9]. Međutim, problem sa matricama tetrahoričkih korelacija je to što su vrlo često singularne, što onemogućava sprovođenje faktorske analize (Embretson & Reise, 2000)[10]. Čak i matrice tetrahoričkih korelacija umeju da daju faktore težine. Ovo je posebno slučaj kada ispitanici sa vrlo niskim nivoima osobine imaju određenu, nenultu verovatnoću pozitivnog odgovora na stavke (DeMars, 2010) (videti odeljak 2.3.3 u ovoj knjizi).

Navedeni problemi se prevazilaze upotrebom eksplorativne faktorske analize pune informacije (engl. *Full-Information Factor Analysis – FIFA*) (Bock i ostali, 1988; Bock & Schilling, 1997) koja je potekla iz IRT modela. FIFA predstavlja multidimenzionalno proširenje dvoparametarskog normalnog (2PN) modela Boka i saradnika (Embretson & Reise, 2000). Procena parametara u okviru ove analize bazira se na celokupnom sklopu odgovora (otud „puna informacija") i upotrebi procenitelja marginalne maksimalne verodostojnosti (engl. *Marginal Maximum Likelihood Estimator – MMLE*, videti odeljak 3.1.1.3.). U račun nisu uključene korelacije te ne postoji problem faktora težine.

Inače, pokazano je da su faktorska analiza kategorijalnih podataka (Muthén, 1983, 1984) i IRT dve formulacije istog modela (više u Takane & De Leeuw, 1987), te da se parametri iz ovakve faktorske analize bez dodatnih informacija mogu pretvoriti u IRT parametre (opterećenja u nagibe i pragovi u težine stavki

[9] Tetrahoričke korelacije se koriste prilikom računanja korelacija binarnih varijabli. Takav slučaj su stavke kod kojih su netačni odgovori skorovani sa 0, a tačni sa 1. Bitna pretpostavka na kojoj počiva računanje tetrahoričkih korelacija je da postoji kontinuirana latentna varijabla koja utiče na odgovaranje ispitanika na binarnoj skali odgovora. Odgovaranjem na stavku latentna varijabla biva binarizovana. Ako ispitanik ima nivo latentne osobine ispod određenog praga (tačke na latentnom kontinuumu), njegov skor biće 0, a ako njegov nivo osobine prelazi taj prag imaće skor 1. Cilj izračunavanja koeficijenta tetrahoričke korelacije je da se utvrdi kolika je produkt moment korelacija latentnih varijabli koje utiču na odgovoranje na stavke. U osnovi različitih manifestnih kategorijalnih varijabli mogu stajati različite latentne varijable, ali kada su u pitanju stavke jednodimenzionalnog testa latentna varijabla je obično ista za sve stavke. Tetrahorička korelacija može se smatrati posebnim slučajem polihoričke korelacije. Logika je ista, a razlika je samo u broju kategorija (više o tetrahoričkoj korelaciji videti u: Lord & Novick, 2008).

[10] Singularna matrica je matrica čija je determinanta jednaka ili bliska nuli i ima bar jedan karakteristični koren jednak nuli ili manji od nje. Singularnost matrice je osobina koja onemogućava njenu inverziju koja je potrebna za računanje različitih pokazatelja u okviru faktorske analize. Ako je matrica singularna, procena modela na osnovu nekih estimatora (npr. generalizovani najmanji kvadrati – GLS) neće biti moguća, dok će na osnovu drugih (npr. maksimalna verodostojnost – ML) rezultirati problemima u konvergenciji ili nestabilnim procenama parametara i nestabilnim pokazateljima saglasnosti, odnosno fita (Lorenzo-Seva & Ferrando, 2021). Matrica korelacija će biti singularna kada postoji izražena multikolinearnost, odnosno kada je neka od varijabli linearna transformacija druge ili više drugih varijabli u sistemu. Razlog može biti i prisustvo varijabli sa nultom varijansom. Osim ovoga, matrica korelacija će biti singularna kada je broj slučajeva manji od broja varijabli u matrici.

ili kategorija). Sama FIFA nije isključivo vezana za binarne stavke i primenljiva je i na politomnim stavkama.

**Pravilo 10.** Ovo pravilo se odnosi na osobine stavke kao stimulusa. U KTT jedna od najbitnijih pretpostavki je jednodimenzionalnost. Stavke se u test biraju na osnovu numeričkih mernih pokazatelja i specifične karakteristike u vezi sa njihovim sadržajem nisu toliko bitne. S druge strane, kod multidimenzionalnih IRT modela ove karakteristike mogu igrati bitnu ulogu. Ako u odgovaranju na stavke učestvuje više dimenzija (konstrukata) može doći do tzv. „pomeranja konstrukta" povezanog sa npr. težinom stavki, odnosno da u zavisnosti od težine stavki na odgovaranje počinje više da utiče nek drugi konstrukt, a to multidimenzionalni IRT modeli mogu uzeti u obzir (Embretson & Reise, 2000).

## *1.2 Modelski pristup*

IRT predstavlja „modelski" pristup merenju. Naučni modeli namenjeni su tumačenju stvarnosti. Da bi se mogli koristiti u tu svrhu oni moraju biti izomorfni stvarnosti, odnosno moraju predstavljati njenu dobru reprezentaciju (Fajgelj, 2004). U suprotnom, zaključci koji se o stvarnosti donose na osnovu modela (o onom njenom delu koji model modelira) ne moraju biti tačni. Izomorfnost modela, odnosno njegova saglasnost sa stvarnošću nikada nije potpuna. Ona se može proveriti empirijski, proverom saglasnosti predviđanja modela sa stvarnim podacima. Tek ako je model u dovoljnoj meri saglasan opaženim podacima, možemo reći da on dobro opisuje stvarnost i možemo ga koristiti za njeno tumačenje i analizu (Fajgelj, 2020).

### 1.2.1 Šta je merni model

Prema Embretson i Ris (Embretson & Reise, 2000) merni modeli su matematički modeli (jednačine, izrazi, formule) koji numerički kombinuju nezavisne varijable kako bi optimalno predvideli zavisne. IRT modeli spadaju u modele latentne varijable prema kojima latentni konstrukti uzrokuju manifestno ponašanje. Pošto nam latentne varijable nisu direktno dostupne o njima saznajemo preko manifestnog ponašanja, odnosno u slučaju psihološkog merenja, na osnovu odgovaranja na stavke. Dakle, u IRT modelima „nezavisna" varijabla je merena latentna osobina, a „zavisna" varijabla je manifestno ponašanje, odnosno odgovaranje na stavke.

Kao što je ranije već rečeno, IRT modeli predviđaju verovatnoću određenog odgovora ispitanika na stavku, uzimajući u obzir stepen prisustva latentne osobine i parametre stavke (kao što su težina, diskriminativnost, pogađanje...). IRT modeli merenja su kao i model baziran na KTT, opšti modeli merenja. Koriste se kako za konstrukciju i evaluaciju instrumenata, tako i za opis podataka dobijenih testiranjem. S obzirom na to da je IRT modelski pristup merenju, njegova predviđanja mogu se empirijski proveriti poređenjem sa realnim podacima. Ovaj postupak naziva se analiza saglasnosti (saobraznosti), ili kako je u govoru uobičajenije „analiza fitovanja".

## *1.3 Osobine IRT modela*

Pored činjenice da je IRT *modelski pristup* merenju, njegove važne osobine su *skaliranje* i *probabilistički pristup*.

Kada kažemo da je osobina IRT modela ***skaliranje***, time mislimo na cilj IRT modela da kvantitativno odredi položaj ispitanika i stavke na kontinuumu latentne crte (ili više njih). Radi boljeg razumevanja već ranije smo ovaj pristup kontrastirali sa pristupom *skorovanja* karakterističnim za KTT. Kao što je rečeno, u KTT se na osnovu odgovora na stavke računa ukupan skor ispitanika koji je indikator prisustva merene osobine (Fajgelj, 2020).[11].

IRT je ***probabilistički*** pristup, zbog činjenice da IRT modeli predviđaju *verovatnoću* određenog odgovora ispitanika na stavku. Uz predviđanja IRT modela uvek je vezan određeni stepen greške. Greška je efekat činilaca koji nisu deo modela (odnosno nezavisnih varijabli koje nisu uključene u model) (Lord & Novick, 2008).

Za razliku od probabilističkih modela, deterministički modeli pretpostavljaju da je zavisna varijabla određena nezavisnim varijablama u modelu. Varijansa greške je tako mala da se može zanemariti (Lord & Novick, 2008). U psihologiji i psihometriji su retki takvi modeli, s obzirom na to da su psihološke pojave suviše kompleksne i rezidualno variranje nikada nije toliko malo da se može zanemariti.

Reziduali u IRT modelima mogu se analizirati i analiziraju se u proceni saglasnosti (fita) modela i tu mogu pružiti značajne informacije o ispitanicima, stavkama i testu u celini (videti odeljak 6).

[11] Npr. ako bi ispitanik na binarno skorovanom testu od 5 stavki na prve 3 odgovorio tačno, a na preostale 2 netačno, dobio bi **skor** 3. Taj skor ukazuje na to da ispitanik ima viši nivo merenog svojstva od ispitanika koji je dobio skor 2 ili 1. U pristupu skaliranja, recimo IRT modelu, na osnovu modela bi bila određena lokacija tog ispitanika na kontinuumu merene latentne osobine i taj ispitanik bi dobio meru 1,25 logita (recimo). Ovaj ispitanik ima viši nivo merene osobine od ispitanika koji su dobili mere od (recimo) 0 ili –0,75 logita. Više o pojmu logita u odeljku 2.1.

# 2 Osnovni pojmovi jednodimenzionalnih IRT modela

U ovom odeljku pozabavićemo se osnovnim pojmovima IRT modela kao što su *logit, karakteristična kriva stavke* i *parametri.*

## *2.1 Logit*

Zanimljiva karakteristika IRT modela je da su parametri ispitanika i stavki izraženi u istim mernim jedinicama i na istoj skali, u *logitima* kada su u pitanju logistički modeli, ili u *z skorovima* ako su u pitanju modeli bazirani na normalnoj raspodeli.

Pojam bitan za objašnjavanje logita je ***dobitni odnos*** ili ***šanse*** (*engl. odds of success*). Dobitni odnos je odnos proporcija tačnih (potvrdnih) i netačnih (odričnih) odgovora.

Ovde ćemo napraviti malu digresiju da objasnimo pojmove *verovatnoće*, *šansi* i *verodostojnosti*. Termin **verovatnoća** (engl. *probability*) koristi se da označi koliko je predviđani događaj verovatan (u našem slučaju predviđamo tačan ili potvrdan odgovor). *Verovatnoća* se izražava u obliku proporcije i može imati vrednosti od 0 do 1. Množenjem sa 100 verovatnoća se može pretvoriti u procente i imati vrednosti od 0 do 100, pa kažemo: „Ispitanik ima verovatnoću od 49% da odgovori tačno“ ili „Ispitanik ima šansu od 49% da odgovori tačno“. Vidimo da se u našem jeziku termini *verovatnoća* i *šansa* mogu koristiti kao sinonimi. Međutim, ovde se termin ***šanse*** koristi kao prevod engleskog termina *odds* (ili *odds of success*). Termin ***šanse*** označava odnos verovatnoće tačnog odgovora i verovatnoće netačnog odgovora. Šanse nam govore *koliko puta* je verovatnije da ispitanik na stavku odgovori tačno nego netačno. Mogu imati vrednosti od 0 do +∞. Kada vrednost šansi iznosi 1, tačan i netačan odgovor su jednako verovatni. Kada je vrednost šansi veća od 1, tačan odgovor je verovatniji, a kada je manja od 1, onda je verovatniji netačan odgovor. Šanse se mogu pretvoriti u *verovatnoću* ako šanse ispitnika podelimo šansama stavke (videti izraz [ 6 ] kasnije u ovom odeljku).

Još jedan termin sa sličnim značenjem je ***verodostojnost*** (engl. *likelihood*). U srpskom jeziku se i umesto ovog termina često koristi termin *verovatnoća*. Verodostojnost [12] je proizvod verovatnoća odgovora ispitanika na stavke. Dok se *verovatnoća* (u značenju *probability*) obično odnosi na događaj koji se još nije desio, verodostojnost se odnosi na događaj koji se već odigrao (Embretson & Reise, 2000). Kada ispitanik

[12] Ni ovaj termin nije u potpunosti odgovarajući jer u srpskom jeziku on označava ono što je *verno*, *tačno* ili *autentično* (Vujanić i ostali, 2011). Terminu nedostaje značenje koje ukazuje na verovatnoću. Možda bi bolji termin bio „izvesnost“ ili eventualno „verovatnost“ (pri čemu se ovaj poslednji ne razlikuje dovoljno od termina „verovatnoća“). U svakom slučaju, termin „verodostojnost“ je prihvaćen i mi ćemo ga koristiti.

nije odgovarao na neki test, a poznati su nam parametri stavki i nivo osobine ispitanika, na osnovu jednačine modela moći ćemo da procenimo *verovatnoću* (*probability*) tačnog odgovora ispitanika na svaku od stavki. Kada je neki uzorak ispitanika odgovarao na test i poznati su nam odgovori na stavke, moći ćemo da procenimo za koje vrednosti parametara stavki i ispitanika su dobijeni sklopovi odgovora *najverodostojniji* (*likelihood*), odnosno, moći ćemo da procenimo vrednosti parametara na osnovu kojih je najverovatnije da je nastala takva (poznata) matrica odgovora.

Dakle, sva tri termina nam govore o nekoj vrsti verovatnoće, pa bismo mogli reći da je verovatnoća (probability) najopštiji termin koji obuhvata druga dva.

Vratimo se sada na *dobitni odnos*, odnosno *šanse*. Ako sa *p* označimo proporciju tačnih odgovora na stavke koje je dao neki ispitanik *i*, onda se dobitni odnos može izračunati na sledeći način:

$$DO_i = \frac{p}{1-p} \qquad [2]$$

Ako broj netačnih odgovora (1–*p*) označimo sa *q*, onda se prethodni izraz može jednostavno napisati kao $DO_i = p/q$. Dobitni odnos za ispitanika *i* koji je tačno rešio 3 od 4 stavke jednak je $DO_i = 0{,}75/0{,}25 = 3$. Dobitni odnos ispitanika nam govori o njegovom nivou osobine. Recimo, ako je drugi ispitanik *k* na istom testu od 4 stavke tačno rešio 2 (odnosno 50%), njegov dobitni odnos biće $DO_k$=1 i on najverovatnije ima niži nivo sposobnosti od ispitanika koji je tačno rešio 75% zadataka.

Dobitni odnos se lako može pretvoriti u proporciju pozitivnog odgovora na sledeći način:

$$P = \frac{DO}{1+DO} \qquad [3]$$

Odnosno *P*=3/(1+3)=0,75.

Za stavke, dobitni odnos može se izračunati na isti način, sa jednom razlikom. Naime, dobitni odnos stavke računa se kao:

$$DO_j = \frac{1-p}{p} \qquad [4]$$

gde je *p* proporcija ispitanika koji su tačno odgovorili na stavku *j*.

Kao što je već rečeno, dobitni odnos za ispitanike nam govori koliko puta su veće šanse da ispitanik odgovori tačno nego netačno. Što su veće šanse za tačan odgovor, ispitanik poseduje viši nivo osobine (sposobnosti, znanja) koju test meri. Kod stavki dobitni odnos bi trebalo da govori koji nivo osobine je potreban da bi ispitanik tačno odgovorio na stavku. Što manje ispitanika tačno reši stavku to je potrebniji viši nivo osobine. Dakle, što je veća proporcija onih koji su pogrešno odgovorili na stavku, dobitni odnos stavke je veći. Otud se u ovom izrazu u brojiocu nalazi *1–p*, što je proporcija ispitanika koji zadatak nisu rešili tačno.

Recimo to ovako: *to je proporcija ispitanika koje je zadatak „rešio"*. Dobitni odnos za stavke nam govori koliko puta su veće šanse da ispitanici pogreše na toj stavci, nego da odgovore tačno. To nam govori o težini stavke. Kada bi u brojiocu stajalo *p*, dobitni odnos bi govorio o njenoj lakoći.

Teža je stavka na koju je tačan odgovor dala manja proporcija ispitanika. Na primer, stavka *j* na koju je tačno odgovorilo samo 25% ispitanika imaće $DO_j$=3, a stavka *m* na koju je tačno odgovorilo duplo više (50%) ispitanika $DO_m$=1.

Logit je prirodni logaritam šansi:

$$logit = ln\left(\frac{p}{1-p}\right)$$

[ 5 ]

Na ovaj način šanse (dobitni odnosi) koje su varijabla racio nivoa prelaze na intervalnu skalu. Ako kažemo da je dobitni odnos za ispitanika mera osobine, a za stavke mera njene težine, onda bismo verovatnoću da će ispitanik *i* tačno odgovoriti na stavku *j* mogli dobiti kao količnik ta dva dobitna odnosa:

$$P(x_{ij} = 1) = \frac{DO_i}{DO_j}$$

[ 6 ]

Intervalizacijom dobitnih odnosa putem logaritmovanja – operacija deljenja postaje oduzimanje, a prethodni izraz možemo napisati kao:

$$\ln[P(x_{ij} = 1)] = ln(DO_i) - ln(DO_j)$$

[ 7 ]

Rečeno je da je $DO_i$ u stvari mera nivoa osobine ispitanika koja se u IRT modelima označava sa $\theta_i$, a $DO_j$ mera težine stavke koja se u IRT modelima označava sa $b_j$ (videti odeljak 2.3). U tom slučaju logit se može napisati i ovako (Fajgelj, 2020):

$$P(x_{ij} = 1|\theta, b) = \frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}} = \Psi(\theta_i - b_j) = \Psi Z$$

[ 8 ]

gde je Ψ kumulativna distribucija logističke raspodele, $\theta_i$ nivo osobine ispitanika, $b_j$ parametar težine stavke (videti odeljak 4.1.1 i izraze [ 24 ] i [ 25 ]), a *e* osnova prirodnog logaritma. To je matematička konstanta koja se naziva Ojlerov broj (Euler) ili Neperova konstanta (Napier) i približno iznosi 2,718. Ako *e* stepenujemo prirodnim logaritmom nekog broja *x* dobićemo taj isti broj x. To možemo napisati i ovako: $e^{ln(x)}=x$. Eksponencijacijom osnove *e* vrednošću $\theta_i - b_j$ izraz $\theta_i - b_j$ vraćamo u izvornu metriku dobitnog odnosa. Eksponencijacija je operacija inverzna logaritmovanju. Kako je izraz $e^{(\theta_i - b_j)}$ u suštini jednak $DO_i/DO_j$, izraz [ 8 ] možemo smatrati ekvivalentnim izrazu [ 3 ].

Naziv ***logit*** potiče od engleskog termina „log odds unit" (Fajgelj, 2020; Wu i ostali, 2016). U izrazu

[ 8 ] prikazan je logit jednoparametarskog logističkog modela. Logit se često označava i sa *Z*. U ovoj logaritamski transformisanoj varijanti parametri ispitanika ($\theta_i$) i parametri težine stavki ($b_j$) imaju simetričnu distribuciju. Mogu se kretati u opsegu od –∞ do ∞ (mada retko izlaze iz okvira od –3 do 3 logita).

## *2.2 Karakteristična kriva stavke*

Karakteristična kriva stavke – KKS (engl. *Item Characteristic Curve – ICC*) je matematička funkcija verovatnoće. Ova funkcija nam govori koja je verovatnoća određenog odgovora ispitanika *i* na stavku *j*, ako se u obzir uzmu određeni parametri modela. Za binarno skorovanu stavku *j* (1 - tačno, 0 - netačno) ona se može napisati kao (Fajgelj, 2020):

$$P_j(\theta) = Prob(X_{ij} = x \mid \theta) = f(\theta, \boldsymbol{\eta}, x) \qquad [9]$$

pri čemu je *x* vrednost modeliranog odgovora, a **η** je vektor parametara modela[13]. $X_{ij}$ označava skor ispitanika *i* na stavci *j*, a vrednost modeliranog odgovora je *x*. Funkcija daje vrednosti od 0 do 1. Iako smo rekli da se u ovoj knjizi nećemo baviti multidimenzionalnim IRT modelima, napomenućemo da se u tim modelima, usled većeg broja latentnih dimenzija, verovatnoća pozitivnih odgovora umesto krivom predstavlja površinom verovatnoće. Naravno, i takvi prikazi površina verovatnoće ograničeni su na trodimenzionalni prikaz koji može uključivati združeno delovanje samo dve dimenzije na jednom grafikonu.

Na slici (Slika 1) prikazana je karakteristična kriva binarno skorovane stavke *j* dobijena na osnovu Rašovog jednoparametarskog modela. Parametar težine ove stavke je $b_j$=0,41. U Rašovom jednoparametarskom modelu parametar diskriminativnosti jednak je za sve stavke, ne procenjuje se, ali se implicitno postavlja na $a_j$=1 (videti odeljak 4.1).

Na KKS možemo gledati i kao na nelinearnu regresiju ajtemskog odgovora na jednu ili više latentnih varijabli koje reprezentuju psihičku osobinu koja se meri (Fajgelj, 2020; Lord, 1975).

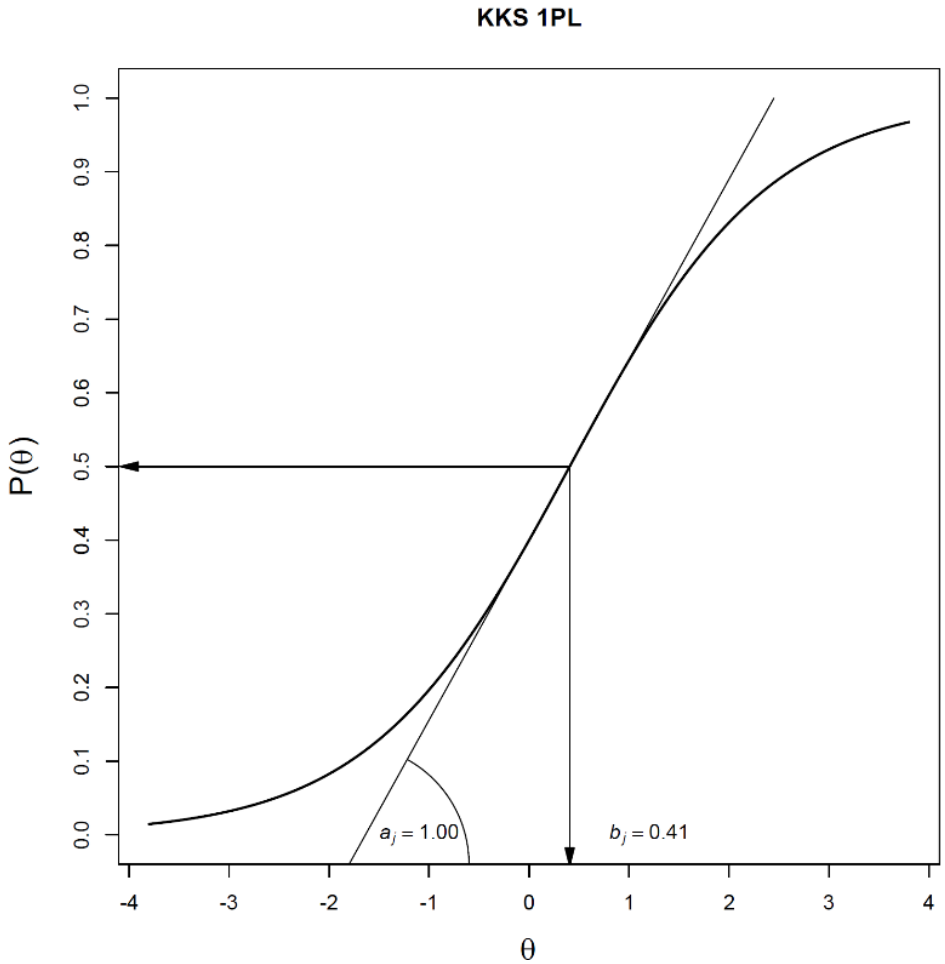
Šta možemo uočiti? Kao prvo, parametar nivoa osobine ispitanika $\theta_i$ i parametar težine stavke $b_j$, prikazani su na istoj skali (i na istoj osi - apscisa). Kriva ima oblik kumulativne krive logističke raspodele – ogive (slična je i kumulativna kriva normalne raspodele, samo što bi bila strmija).

Ako posmatramo KKS, vidimo da verovatnoća tačnog odgovora raste sa porastom nivoa merene osobine $\theta$. Taj rast je u početku relativno spor i potrebne su veće promene u nivou osobine da bi došlo do manjeg porasta u verovatnoći tačnog odgovora. Oko sredine krive, gde je nagib najveći, i mali porast u nivou osobine dovodi do velikog porasta u verovatnoći tačnog odgovora. Do sredine kriva ubrzano raste, a od te

[13] IRT modeli mogu se razlikovati po broju parametara koje procenjuju. „Vektor parametara" jednoparametarskog modela bi sadržao samo parametar težine (*b*), a dvoparametarskog parametre nagiba i težine (*a*, *b*). Za troparametarski model bi dodatno sadržao i parametar donje asimptote (*a*, *b*, *c*), a za četvoroparametarski i parametar gornje asimptote (a, b, c, d). Više o parametrima IRT modela u odeljku 2.3). $X_{ij}$ je odgovor ispitanika i na binarno skorovanoj stavci može imati vrednosti 0 ili 1. Modelski predviđena verovatnoća odgovora (*x*) može biti vrednost između 0 i 1.

tačke rast sve više usporava i opet doprinos u nivou osobine biva praćen sve manjim porastom u verovatnoći pozitivnog odgovora.

Središnja tačka krive naziva se *tačka preloma* i važna je iz više razloga.

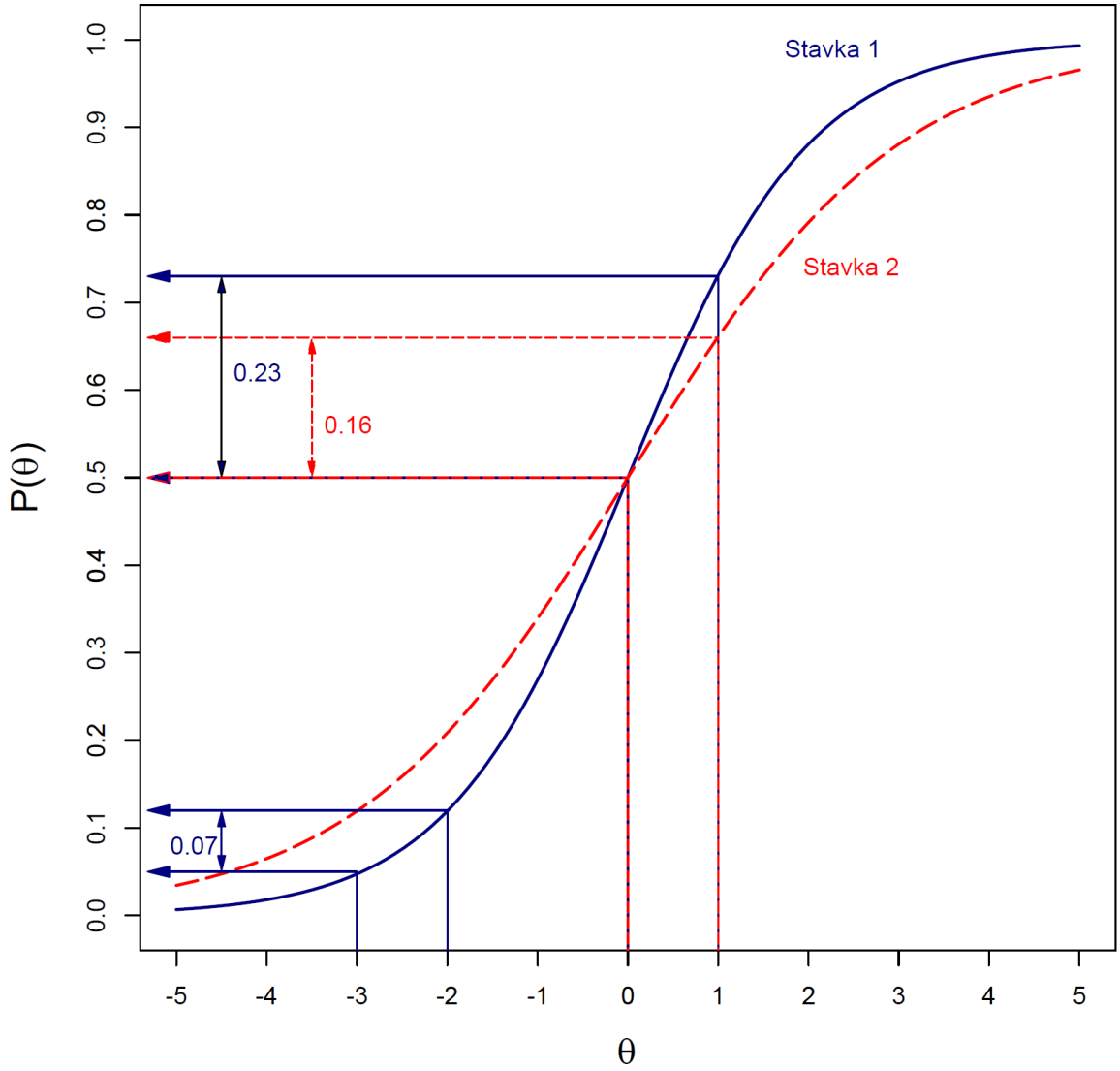



Slika 1 – Karakteristična kriva stavke u jednoparametarskom logističkom modelu

Prvo, projekcija tačke preloma na apscisu predstavlja parametar težine stavke $b_j$. Kao što je već rečeno, prikazana stavka ima parametar težine $b_j$=0,41 logita. Neka druga, teža stavka, bi se nalazila više desno, a lakša stavka više levo.

Drugo, parametar diskriminativnosti stavke $a_j$ definiše se kao nagib KKS u *tački preloma*. Treba napomenuti da ovo nedvosmisleno važi samo u dihotomnim modelima. U politomnim, ovaj parametar je vezan za kategorije odgovora (Fajgelj, 2020). Pošto se u prikazanom primeru radi o jednoparametarskom, Rašovom modelu, parametar diskriminativnosti ove stavke nije procenjivan, ali je implicitno postavljen na $a_j$=1. U jednoparametarskim modelima sve stavke imaju isti nagib. U višeparametarskim modelima gde se parametar diskriminativnosti procenjuje za svaku stavku, stavke sa višim $a_j$ imaće strmiji nagib. Savršeno diskriminativna stavka bi imala krivu u obliku stepenice, a savršeno nediskriminativna stavka bi imala krivu u obliku horizontalne ravne linije. Kao i kod obične linearne regresije nagib nam govori o tempu promene zavisne varijable sa promenom nezavisne. U ovom slučaju govori nam koliko se brzo menja verovatnoća tačnog odgovora za jediničnu promenu u nivou osobine.

Kao što smo rekli, nagib KKS nije svuda jednak. U oblasti ekstremnih nivoa osobine (visokih ili niskih), promena u nivou crte dovodi do malih promena u verovatnoći tačnog odgovora. Najveći efekat postoji u oblasti srednjeg nivoa merene osobine gde ista razlika u nivou osobine znači mnogo veću razliku u promeni verovatnoće tačnog odgovora nego na niskim ili visokim nivoima osobine (Slika 2).




Slika 2 – Karakteristična kriva dve stavke sa različitim parametrima diskriminativnosti (2PL model)

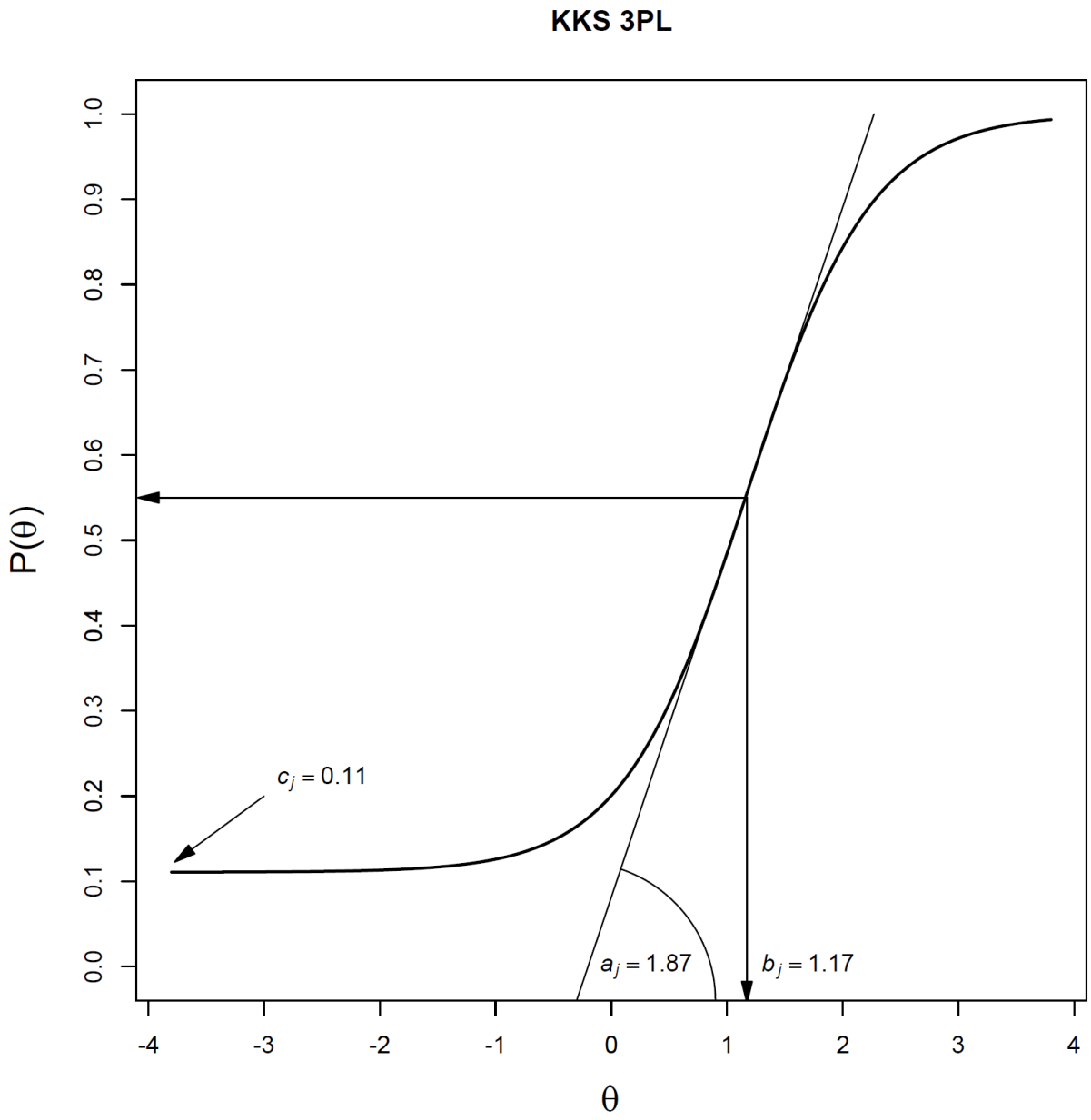
Isti odnos može se primetiti i u slučaju sa dve stavke koje imaju različite parametre diskriminativnosti (Slika 2). Na ovoj slici prikazane su KKS dve stavke procenjene na osnovu dvoparametarskog logističkog modela. Obe stavke imaju jednak parametar težine ($b$=0). *Stavka 1* ima veći parametar diskriminativnosti od *stavke 2* ($a_1$=1,5 i $a_2$=1).

Razlika u nivou osobine od 1 logita ne znači uvek isti doprinos u verovatnoći tačnog odgovora. Posmatrajmo stavku 1. Porast nivoa osobine od 1 logita na niskom nivou osobine, sa –3 na –2 logita, povećava verovatnoću tačnog odgovora na ovu stavku za 0,07. Isti toliki porast na umerenom nivou osobine, sa 0 na 1 logit, povećava verovatnoću tačnog odgovora za 0,23.

Takođe, razlika u verovatnoći tačnog odgovora između ispitanika koji imaju nivoe osobine 0 i 1 logit

veća je kod stavke koja ima viši parametar diskriminativnosti. Kod stavke 1 ova razlika iznosi 0,23 (0,73–0,50), dok kod stavke 2 razlika iznosi 0,16 (0,66–0,50). Dakle, za istu promenu u nivou crte veća razlika je kod stavke sa višim nivoom diskriminativnosti, što ukazuje da je verovatnoća tačnog odgovora više povezana sa nivoom latentne osobine. Samim tim i stavka je više povezana sa predmetom merenja testa. Logika je slična kao kod koeficijenta diskriminativnosti stavke u KTT. Podsetićemo se da se ovaj koeficijent u KTT računa kao korelacija skora na stavci i skora na testu (što je takođe procena nivoa merene osobine).

U troparametarskom modelu KKS izgleda malo drugačije (Slika 3).



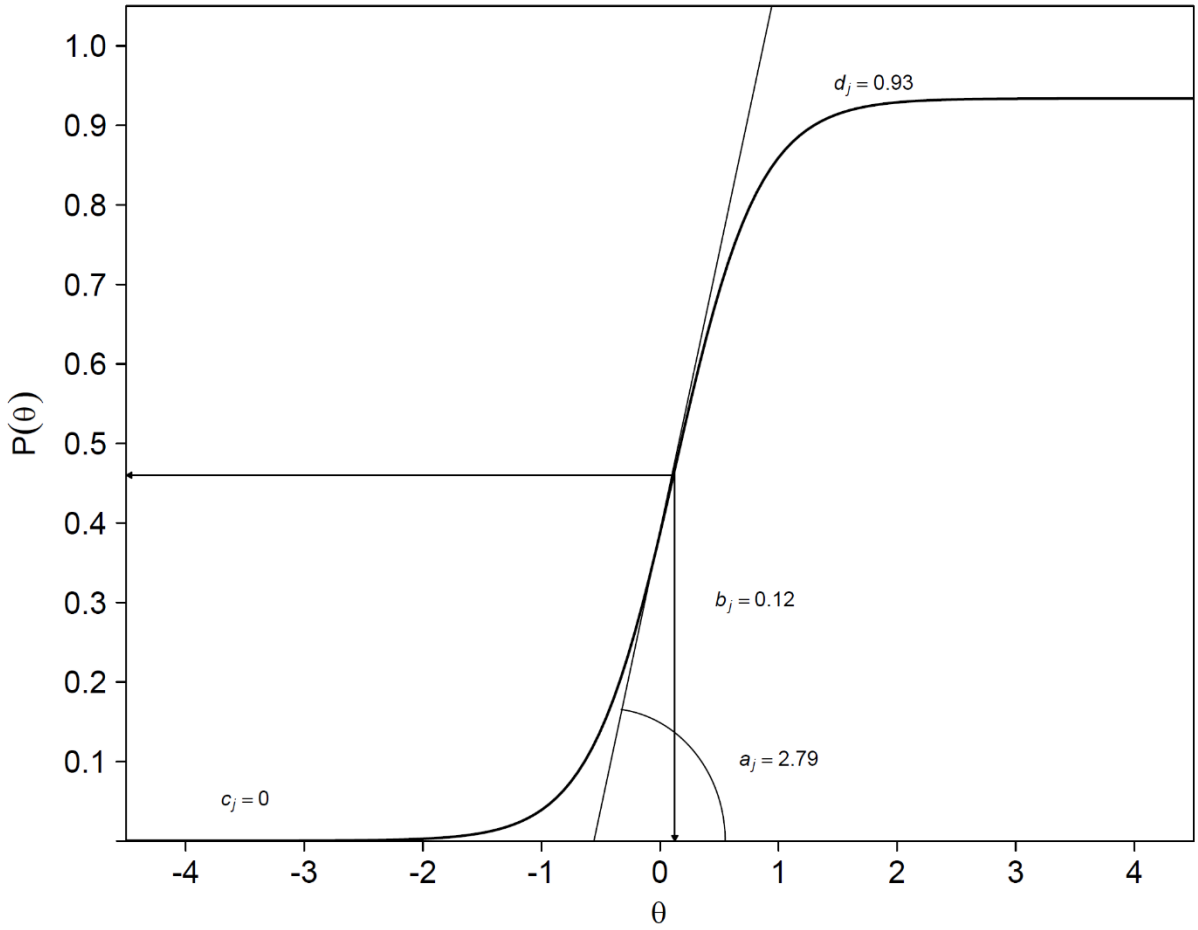

Slika 3 – KKS u troparametarskom logističkom modelu

Na slici (Slika 3) prikazana je karakteristična kriva binarno skorovane stavke dobijena primenom troparametarskog logističkog modela. Prikazana stavka ima parametar težine $b_j$=1,17 i parametar diskriminativnosti $a_j$=1,87. Novina u odnosu na prethodne modele je parametar $c_j$=0,11. Parametar $c$ naziva se parametrom pogađanja ili češće pseudopogađanja. Verovatno ga je najbolje nazvati donjom asimptotom

karakteristične krive stavke, odnosno njenim odsečkom na *y* osi (ordinata). On predstavlja najmanju verovatnoću tačnog odgovora kod ispitanika sa bilo kojim nivoom osobine. Smisao mu je sličan kao smisao odsečka u običnoj linearnoj regresiji. Podsetimo se, tamo je odsečak vrednost predviđane varijable *y* kada je vrednost prediktora *x*=0. Kontinuum latentne osobine proteže se od –∞ do +∞ i 0 ne znači odsustvo osobine već označava njen prosečan nivo (kao kod z skorova). Zato za odsečak kažemo da je najniža verovatnoća tačnog odgovora kod ispitanika sa bilo kojim nivoom osobine pa i onim najnižim, što bi bio ekvivalent 0 u linearnoj regresiji.

Ako ovaj odsečak tumačimo kao pogađanje, stavke kod kojih je mogućnost pogađanja veća imaju viši odsečak, a one kod kojih je mogućnost pogađanja manja imaju niži odsečak.

U četvoroparametarskim modelima karakteristična kriva stavke ima i gornji odsečak, odnosno gornju asimptotu (Slika 4). Ona se definiše kao najviša verovatnoća tačnog odgovora kod ispitanika sa bilo kojim nivoom latentne osobine. Gornji odsečak se naziva i parametrom nepažljivosti ili efektom omaške (Van Der Linden & Hambleton, 1997).




Slika 4 – KKS u četvoroparametarskom logističkom modelu

U jednodimenzionalnim IRT modelima za politomne stavke ne postoji karakteristična kriva stavke već *karakteristične krive kategorija* (KKK). Karakteristične krive kategorija su funkcije verovatnoće biranja određene kategorije odgovora u zavisnosti od nivoa latentne osobine ispitanika (Slika 11 u odeljku 4.2.1).

## *2.3 Parametri u IRT*

U zavisnosti od IRT modela može se procenjivati veći ili manji broj parametara.

Parametri koji se najčešće javljaju u IRT modelima su: nivo latentne osobine ispitanika (parametar ispitanika), težina, diskriminativnost, pogađanje i nepažljivost. Prikazani su parametri za dihotomne modele, dok će parametri koji se javljaju samo u politomnim modelima biti opisani naknadno, kada budu opisivani ti modeli.

Oznake ovih parametara su sledeće:

- Nivo latentne osobine ispitanika (engl. *ability, proficiency, trait*) – $\theta$ (a ponekad u starijim izvorima se označava i sa β)
- Parametar težine, lokacija stavke – $b$ (ali se nekada označava i sa δ ili opet $\beta$
- Parametar diskriminativnosti, odnosno nagib – $a$ (ili $\alpha$)
- Pseudopogađanje, „slučajni uspesi“, donja asimptota stavke – c (ili $\gamma$)
- Parametar nepažljivosti, nerazumevanje, loši uslovi testiranja, efekat omaške – $d$ (ili $z$)

Ovo su najčešće korišćeni parametri u IRT modelima, kako jednodimenzionalnim tako i u multidimenzionalnim. Ova lista parametara nije konačna. Npr. pominje se i model sa parametrom zakrivljenost stavke (Bazán i ostali, 2006), ali takvi modeli su retko u upotrebi i u ovoj knjizi ih nećemo obrađivati.

Naravno, u multidimenzionalnim modelima koji pretpostavljaju uticaj više latentnih osobina na proces odgovaranja postoje višestruki parametri nivoa latentne osobine ispitanika $\theta$ i parametri nagiba $a$ za svaku dimenziju posebno, a parametar težine se procenjuje odvojeno samo u parcijalno kompenzatornim modelima (DeMars, 2016).

### *2.3.1* Parametar težine stavke – *b*

Parametar težine[14] procenjuje se u svim modelima. U modelima za binarno skorovane stavke radi se o *parametru težine stavke*, dok se u modelima za politomne stavke sa uređenim kategorijama često procenjuju *parametri težine za kategorije*, ali ne i za stavku u celini. Kod binarno skorovanih stavki, parametar $b$ definiše se kao nivo $\theta$ u tački preloma KKS (tački infleksije) i odgovara lokaciji na kontinuumu latentne osobine na kojoj je verovatnoća tačnog odgovora od 50%. Podsetićemo, s obzirom na to da su parametar težine stavke $b_j$ i parametar ispitanika $\theta_i$ izraženi na istoj skali, može se reći i da je težina stavke definisana potrebnim nivoom osobine da bi ispitanik imao verovatnoću od 50% da na njoj da tačan odgovor (Fajgelj,

[14] Koristimo termin „težina“ pošto je to uobičajen naziv za ovaj parametar, mada nije i najbolji. Ima smisla u kognitivnim testovima i na binarno skorovanim stavkama, dok u nekognitivnim testovima i na politomnim stavkama može biti pomalo zbunjujući. Korisno ga je posmatrati na sledeći način: „koliko se ispitanik teško odlučuje da odabere neku kategoriju odgovora“. Alternativni termin koji se koristi je „lokacija“. Ona se može posmatrati kao lokacija na kontinuumu latentne osobine koja je potrebna za prelazak sa netačnog na tačan odgovor (sa odričnog na potvrdan, ili kod politomnih stavki, sa jedne na drugu kategoriju odgovora).

2020). Ovo što je rečeno se donekle razlikuje u tro- i četvoroparametarskim modelima. I dalje će se parametar težine nalaziti na projekciji tačke infleksije na latentnu osobinu, ali u ovim modelima to neće biti nivo osobine na kojem ispitanik ima verovatnoću od 50% da odgovori tačno. Lokacija tačke preloma u 3P modelima pomera se zbog donje (a u 4P modelima i zbog gornje) asimptote (Embretson & Reise, 2000). U troparametarskom modelu parametar težine stavke je lokacija gde ispitanik ima verovatnoću tačnog odgovora koja iznosi $(1+c_j)/2$ (Harris, 1989). Tako, ako je za neku stavku $j$ parametar $c_j$=0,11, onda se parametar težine $b_j$ nalazi na lokaciji gde je verovatnoća tačnog odgovora 1,11/2=.55 (Slika 3). Pošto četvoroparametarski model ima i parametar nepažljivosti koji ograničava maksimalnu verovatnoću tačnog odgovora, onda bi se parametar težine u ovom modelu nalazio na lokaciji gde je verovatnoća odgovora jednaka $(d_j-c_j)/2$.

Ono što važi u svim modelima je da je parametar težine stavke $b_j$ *lokacija* KKS na kontinuumu merene osobine, pa se zato u literaturi i softveru sreće i pod tim imenom. Vrednosti ovog parametra mogu se kretati od $-\infty$ do $\infty$ kada je $\theta$ skalirana da ima aritmetičku sredinu 0 i standardnu devijaciju 1 (uobičajeno). Iako je to mogući opseg, ovaj parametar retko izlazi iz opsega −3 do 3 u normalnim modelima i od −5 do 5 u logističkim. Isto važi i za parametar ispitanika (nivo osobine). Stavka sa parametrom težine $b_j$ =0 logita je stavka prosečne težine. Stavke sa negativnim parametrom težine su lakše, a sa pozitivnim teže.

Kao što je već rečeno, u politomnim IRT modelima često se ne računa parametar težine stavke već se računaju parametri težine kategorija – pragovi. Pragovi su lokacije na kontinuumu latentne osobine na kojima ispitanik ima jednaku verovatnoću da odabere neku od susedne dve kategorije (npr. „1 – uopšte se ne slažem“ i „2 – uglavnom se ne slažem“ na skali Likertovog tipa).

### *2.3.2* Parametar diskriminativnosti ili nagiba stavke – *a*

*Parametar diskriminativnosti* je drugi parametar u dvoparametarskim modelima. Uvođenje novih parametara obično poboljšava prediktivni kvalitet modela. Na stvarnim podacima neće sve stavke biti jednako diskriminativne, pa će ova nova dodatna informacija poboljšati prediktivnu moć modela. Međutim, Raš je bio mišljenja da model merenja ne bi trebalo da se prilagođava podacima, već je potrebno test konstruisati tako da zadovoljava osnovne principe merenja (Mead, 2008). To je glavni razlog zašto Rašov 1PL model ne uzima u obzir parametar diskriminativnosti. Ako stavke imaju različite parametre diskriminativnosti može doći do ukrštanja njihovih KKS. Takav slučaj prikazan je na donjem grafikonu (Slika 5) na kojem su prikazane karakteristične krive dve binarno skorovane stavke analizirane 2PL modelom.

Smisao parametra diskriminativnosti opisan je detaljnije u odeljku o KKS (2.2) i ovde to nećemo ponavljati. Podsetićemo samo da se on definiše kao nagib KKS u tački preloma i da nam govori o tempu promene verovatnoće biranja nekog odgovora u zavisnosti od nivoa latentne osobine. Što je nagib strmiji, tempo promene verovatnoće je brži, a stavka bolje razlikuje ispitanike sa različitim nivoima osobine.

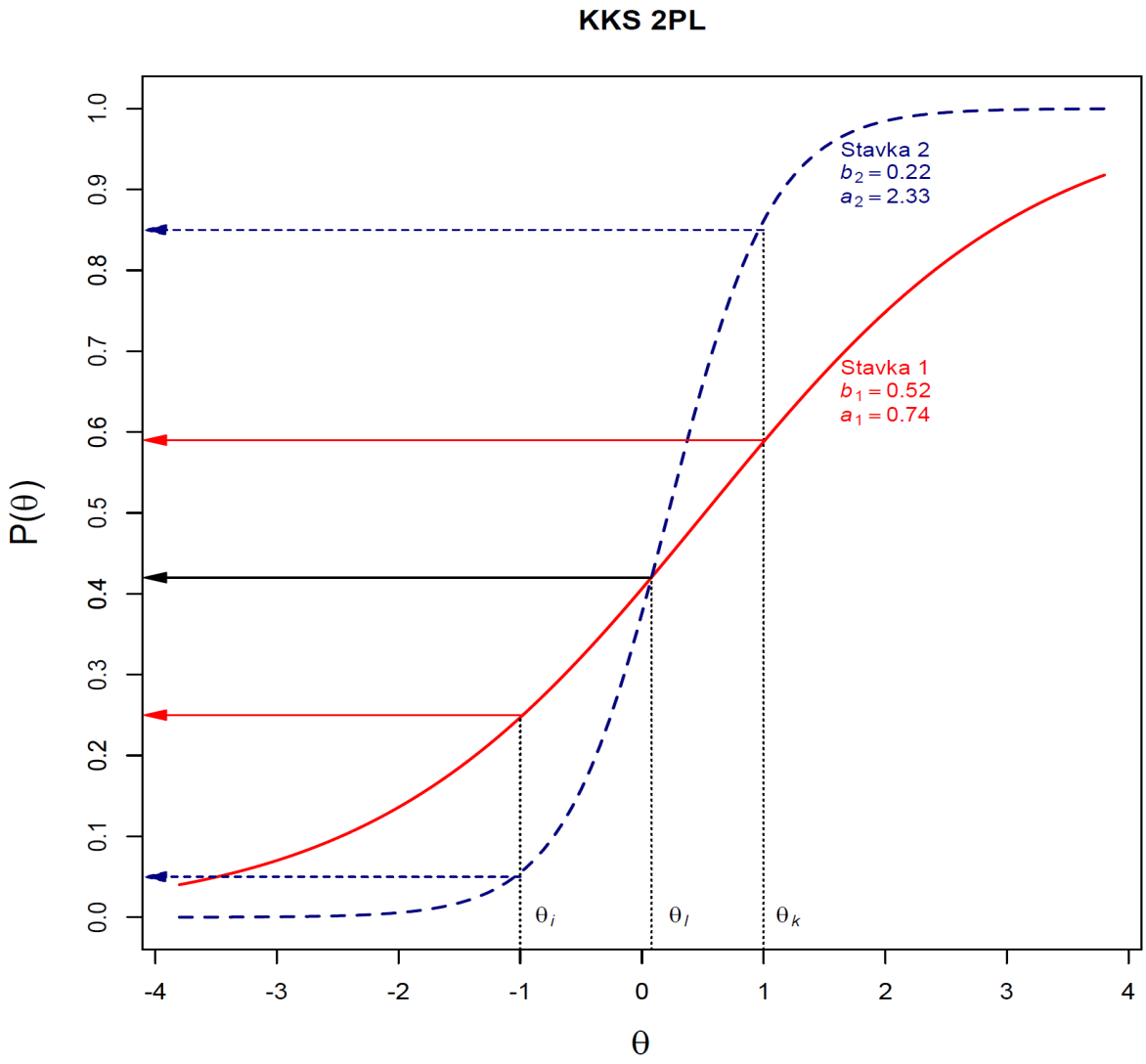
Zanimljiva je i veza parametra diskriminativnosti i koeficijenta biserijske korelacije koji se koristi kao koeficijent diskriminativnosti u KTT. Naime, Lord i Novik (Lord & Novick, 2008) pokazali su da se parametar diskriminativnosti stavke ekvivalentan onome iz dvoparametarskog modela baziranog na normalnoj ogivi može direktno izračunati iz biserijske korelacije stavke i ukupnog skora:

$$a_i \equiv \sqrt{\frac{r_{bis}}{1 - r_{bis}^2}}$$

[ 10 ]

Prema tome, mogli bismo reći da parametar diskriminativnosti u IRT modelima zavisi od korelacije stavke sa merenom osobinom.

Na sledećoj slici (Slika 5) prikazan je efekat parametra diskriminativnosti na verovatnoću tačnog odgovora u IRT modelima za binarne stavke (dvoparametarskim i troparametarskim), sa posebnim osvrtom na pojavu ukrštanja karakterističnih kriva stavki zbog različitih parametara diskriminativnosti.




Slika 5 – Ukrštanje dve KKS u 2PL modelu

*Stavka 1* ima težinu $b_1$=0,52 logita, a *stavka 2* ima težinu $b_2$=0,22 logita (Slika 5). Prema tome, obe stavke imaju težinu nešto iznad prosečne (0 logita), pri čemu je *stavka 2* nešto lakša. Parametar diskriminativnosti *stavke 1* iznosi $a_1$=0,74, a *stavke 2* iznosi $a_2$=2,33. To znači da je *stavka 2* diskriminativnija.

Na slici (Slika 5) vertikalnim isprekidanim linijama predstavljene su lokacije tri ispitanika, od tačaka preseka sa dve KKS strelice pokazuju verovatnoću tačnog odgovora ispitanika na svaku od stavki (pune crvene za stavku 1, isprekidane plave za stavku 2, a puna crna za tačku gde se karakteristične krive dve stavke seku).

Ispitanik *i* ima najniži nivo osobine $\theta_i$=−1 logit. Verovatnoća da će ovaj ispitanik odgovoriti tačno na stavku 1 je 0,25, dok je verovatnoća da će tačno odgovoriti na stavku 2 oko 0,05. Ako imamo u vidu da je stavka 1 teža, vidimo da ovaj ispitanik (koji ima niži nivo osobine) ima veću verovatnoću da tačno odgovori na težu stavku (što, priznaćete, deluje *paradoksalno*). S druge strane, *k* koji ima nivo osobine $\theta_k$=1 logit ima veću verovatnoću da odgovori tačno na lakšu stavku 2 (85%), nego na težu stavku 1 (59%), što je i očekivano. Takođe, ako zamislimo trećeg ispitanika *l*, koji ima nivo osobine $\theta_l$=0,08 logita, što je lokacija na kojoj se dve KKS ukrštaju, vidimo da on ima jednaku verovatnoću (42%) da tačno odgovori na obe stavke. Iz toga sledi da kada postoji ukrštanje KKS nije moguće jednoznačno reći koja je stavka teža. Ispitanicima sa nivoom osobine nižim od $\theta$=0,08 logita teža bi bila stavka 2, a onim sa višim nivoom osobine, teža bi bila stavka 1. Ukrštanje KKS može dovesti do toga da poredak ispitanika sa niskim i visokim nivoom crte nije isti, odnosno, do narušavanja *specifične objektivnosti* (videti odeljak 4.1.1 deo o specifičnoj objektivnosti).

Osim toga, ako poredak ispitanika nije isti na dve stavke, može se postaviti pitanje validnosti tih stavki, odnosno, da li one mere istu latentnu osobinu? Ako mere isti konstrukt stavke, ne bi trebalo da se značajno razlikuju po parametru diskriminativnosti s obzirom da je on u suštini korelacija stavke i merene latentne osobine. Poredak ispitanika na osnovu te stavke ne bi smeo da zavisi od dela kontinuuma latentne osobine.

Parametar diskriminativnosti teoretski može uzimati vrednosti od $-\infty$ do $\infty$, ali se retko sreće van raspona od −2 do 2 u logističkoj metrici (Hambleton i ostali, 1991), ili kako navodi Bejker (Baker, 2001) od −2,8 do 2,8. Iako parametar diskriminativnosti može biti negativan, to se kod binarno ocenjivanih stavki obično dešava kada postoji neki problem sa stavkom ili kada je ona pogrešno skorovana [15] (posebno ako se

[15] Zamislimo jednostavan matematički zadatak: 2+2=? sa ponuđenim odgovorima a) 4 i b) 5. Kada bi ovaj zadatak bio pogrešno skorovan (0 za odgovor pod a i 1 za b), verovatnoća biranja odgovora koji je skorovan kao tačan bi opadala sa rastom u nivou merenog znanja iz matematike. Nagib karakteristične krive ove stavke bio bi negativan, odnosno parametar diskriminativnosti bi bio negativan. Ispravnim skorovanjem ove stavke sve bi došlo na svoje mesto.

Takođe, možemo zamisliti upitnik koji se odnosi na podršku programima za očuvanje životne sredine i stavku koja bi glasila: „Podržavam rudarenje litijuma jer prelaskom na upotrebu električnih vozila smanjujemo emisiju ugljen-dioksida“ sa ponuđenim odgovorima DA i NE. Autor upitnika bi mogao očekivati da odgovor DA ukazuje na veću zainteresovanost za očuvanje životne sredine. Međutim, u nekom društvu bi ispitanici mogli više biti zabrinuti zbog narušavanja životne okoline do kojeg rudarenje litijuma može dovesti. U tom slučaju, oni koji su više zainteresovani za očuvanje životne okoline bi mogli češće birati odgovor NE, a ova stavka bi imala negativnu diskriminativnost. Ovakvu stavku bi trebalo izbaciti ili preformulisati.

radi o kognitivnom testu)(Baker, 2001; Hambleton i ostali, 1991). Greške u skorovanju se ispravljaju, a ako je stavka loše napisana koriguje se ili izbacuje. Zbog toga se u praksi retko javlja parametar diskriminativnosti niži od 0.

Interpretacije visine parametara diskriminativnosti u logističkim i normalnim modelima prikazane su u narednoj tabeli (Tabela 2).

Tabela 2 – Uporedne vrednosti parametra diskriminativnosti u logističkom i normalnom modelu

| | model | |
|---|---|---|
| | logistički | normalni |
| nema | 0 | 0 |
| vrlo nizak | 0,01–0,34 | 0,01–0,20 |
| nizak | 0,35–0,64 | 0,21–0,38 |
| umeren | 0,65–1,34 | 0,39–0,79 |
| visok | 1,35–1,69 | 0,80–0,99 |
| vrlo visok | >1,70 | >1 |
| savršen | +∞ | +∞ |

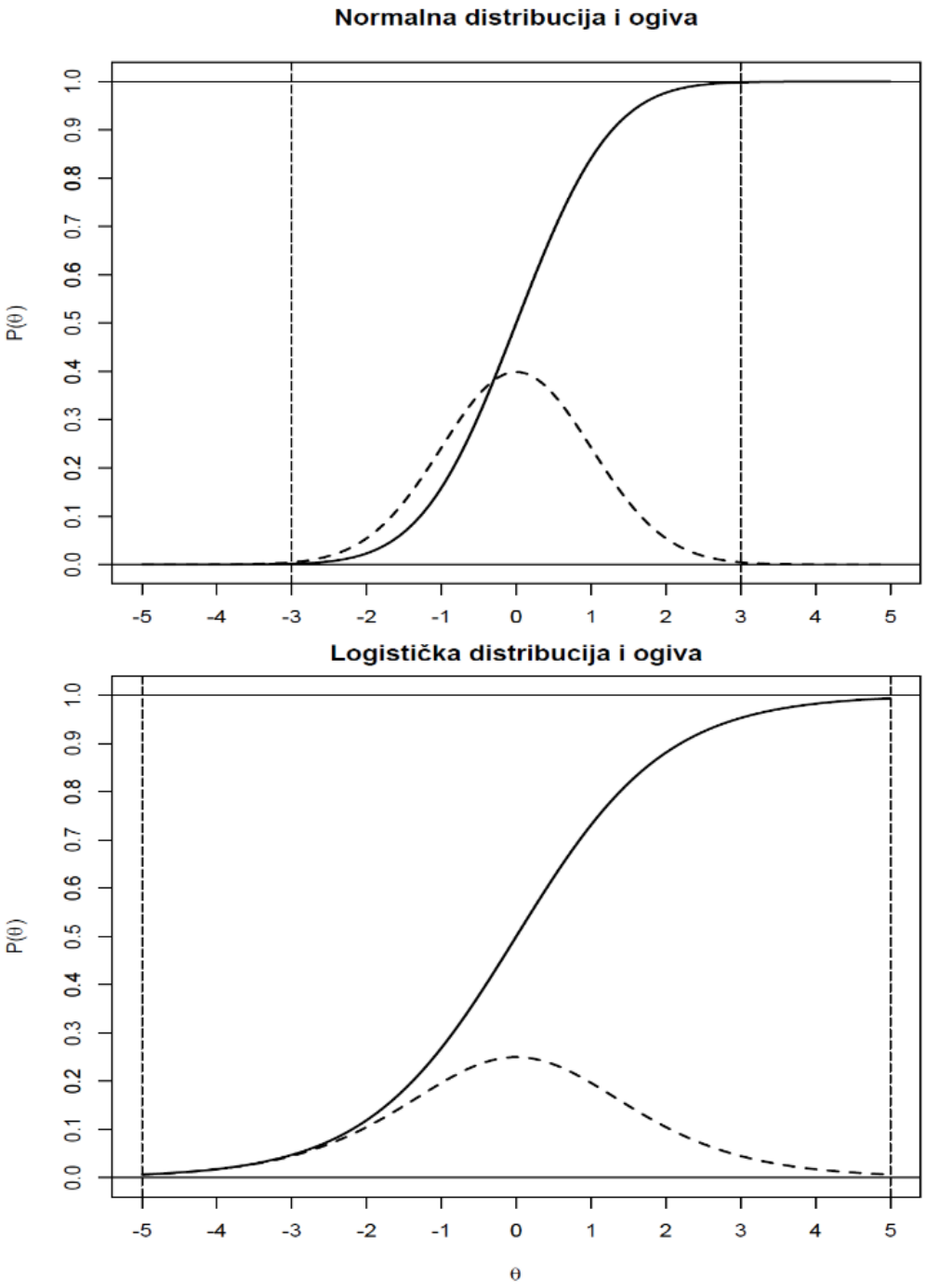
(prema: Baker, 2001)

Kao što se može videti iz tabele (Tabela 2), u poređenju sa logističkim modelima, niže vrednosti parametra *a* u normalnim modelima su dovoljne da bi se on smatrao umerenim, visokim ili vrlo visokim. Razlog tome je razlika u širini dve distribucije. U standardnoj normalnoj distribuciji u rasponu između –3*z* i 3*z* nalazi se 99,73% slučajeva, a u standardnoj logističkoj distribuciji se sličan broj slučajeva nalazi u rasponu od –5 do 5 logita (Slika 6, na slici su ogive prikazane punim linijama).

Kako se proporcija ispod krive u oba slučaja kreće u rasponu od 0 do 1, normalna ogiva (kumulativna kriva) mora biti strmija od logističke. Efektivna skala[16] latentne varijable u normalnim modelima ima raspon od 6 jedinica (–3 do 3*z*), a u logističkim od 10 (–5 do 5 logita). Možemo reći da je logit finija, manja jedinica od z skora. Razlika u nivou latentne crte od 1*z* veća je od razlike od 1 logita. Ranije smo rekli da je nagib tempo promene zavisne za jediničnu promenu nezavisne varijable. Logično je da će se verovatnoća tačnog odgovora manje promeniti kada je promena manja (1 logit) nego kada je ona veća (1 *z*). Ako posmatramo broj jedinica skale, on je u logističkom modelu 1,7 puta veći od broja jedinica u normalnom (10:6).

[16] Pod „efektivna skala“ misli se na raspon vrednosti koje se u praksi javljaju, mada je teorijski raspon obe skale od –∞ do +∞.

Što se tiče naziva ovog parametra, možda bi bilo bolje držati se njegovog tehničkog opisa kao *nagiba KKS u tački preloma*. Razlog tome je što taj nagib nije isti na svim delovima kontinuuma latentne osobine. Osim toga, u politomnim IRT modelima parametar nagiba računa se za stavku u celini, ali ne i za kategorije odgovora. Dalje, diskriminativnost kategorija ne zavisi samo od nagiba već i od razmaka između njih, odnosno širine razmaka između pragova. Što je ovaj interval uži kategorije će biti diskriminativnije (DeMars, 2010; Fajgelj, 2020). Dakle, ovde diskriminativnost zavisi i od nagiba i od razmaka između kategorija.




Slika 6 – Uporedni prikaz standardne normalne i standardne logističke distribucije i njihovih ogiva

### *2.3.3* Parametar pseudopogađanja – *c*

Parametar *pseudopogađanja* ili *slučajnog uspeha* $c_i$ predstavlja donju asimptotu KKS. On predstavlja najnižu verovatnoću tačnog odgovora ispitanika sa bilo kojim nivoom osobine na stavku. Pogađanje ili slučajni uspeh se povezuje sa kognitivnim testiranjem. To nije nešto što bi se moglo očekivati u testiranju konstrukata koji ne pripadaju kognitivnom domenu. Tamo se donja asimptota KKS može definisati kao verovatnoća irelevantnih ili socijalno poželjnih odgovora (Fajgelj, 2020). I opet, kao i u slučaju diskriminativnosti, uvođenje ovog parametra u model obično poboljšava njegove predikcije. Ako model ne sadrži ovaj parametar, varijacije koje potiču iz ovog izvora biće detektovane kao misfit (odnosno pogreške u predviđanju modela). Na taj način mogu se identifikovati ispitanici koji su pogađali.

Iako se parametar pseudopogađanja u jednačinama tro– i četvoroparametarskih modela tretira kao parametar stavke, postoje mišljenja da se radi o parametru uzorka ispitanika (Fajgelj, 2020). Pseudopogađanje definisano kao parametar stavke pretpostavlja da je pogađanje potpuno slučajno, iako ono to najčešće nije, a karakteristično je za ispitanike sa nižim nivoom osobine na težim stavkama.

Takođe, teško ga je razlikovati od parametra *lakoće* koji se javlja u nekim modelima (a definiše se isto kao donja asimptota KKS). Osim toga, procene ovog parametra su nestabilne i zahtevaju velike uzorke ispitanika i stavki.

Parametar *c* može imati vrednosti od 0 do 1, ali u praksi retko prelazi 0,3. Vrednosti parametra do 0,35 smatraju se prihvatljivim (Baker, 2001).

### *2.3.4* Parametar nepažljivosti – *d*

Parametar nepažljivosti definiše se kao gornja asimptota KKS, odnosno najveća verovatnoća tačnog odgovora na stavku kod ispitanika sa bilo kojim nivoom merene osobine. Zbog pojave nepažljivosti i uspavanosti, stavka ne mora imati verovatnoću od 100% da bude rešena, čak ni od strane ispitanika sa najvišim nivoom osobine. Zamerke koje su date na parametar pogađanja važe i za ovaj parametar. Iako se definiše kao parametar stavke, nekada se kao uzroci ove pojave, pored nepažljivosti i uspavanosti, navode i anksioznost, te ometanje prilikom procesa testiranja (što najčešće nema direktne veze sa samom stavkom).

Teorijski i ovaj parametar može imati raspon od 0 do 1, pri čemu viša vrednost označava manji efekat omaške (nepažljivost).

# 3 Procena parametara stavki i ispitanika

U svim IRT modelima potrebno je proceniti parametre stavki i parametre ispitanika jer su nam i jedni i drugi nepoznati. Embretson i Ris (Embretson & Reise, 2000) ovu situaciju porede sa sprovođenjem logističke regresije u kojoj su nam nepoznate vrednosti prediktora.

## *3.1 Maksimalna verodostojnost*

Iako su mogući različiti načini procene parametara, najčešće se koristi neki od metoda baziranih na proceni maksimalne verodostojnosti (engl. *maximum likelihood*). Verodostojnost i verovatnoća su slični pojmovi i ovde se odnose na proporciju tačnih/pozitivnih odgovora. Razlika između dva pojma je u tome da li se odnose na događaj koji se desio ili se nije desio. Verovatnoća se odnosi na događaje koji se još nisu desili, a verodostojnost na one koji jesu.[17] (Embretson & Reise, 2000).

Najpopularnija su tri metoda procene parametara: *procena združene maksimalne verodostojnosti* (engl. *Joint Maximal Likelihood Estimation* – JMLE), *procena uslovne maksimalne verodostojnosti* (engl. *Conditional Maximal Likelihood Estimation* – CMLE) i *procena marginalne maksimalne verodostojnosti* (engl. *Marginal Maximum Likelihood Estimation* – MMLE).

Prva dva metoda procene (JMLE i CMLE) ispitanike tretiraju kao fiksne faktore.[18] Verovatnoća odgovora na stavku dobijena na osnovu modela se može interpretirati kao verovatnoća da će ispitanik odgovoriti pozitivno na ponovljenu prezentaciju iste ili ekvivalentne stavke. U pristupima koji nivo osobine ispitanika tretiraju kao slučajan faktor (MMLE) mora se pretpostaviti neka distribucija te osobine. Verovatnoća

[17] O razlici između ovih pojmova bilo je više reči u odeljku 2.1)

[18] Kada govorimo o fiksnim i slučajnim faktorima na umu imamo značenje koje ovi pojmovi imaju u linearnom modelovanju. Ako smatramo da na zavisnu varijablu utiče konačan broj nivoa nezavisne varijable (faktora), zanima nas efekat samo tih nivoa (koji su baš zbog toga i prisutni u analiziranim podacima) i onda taj faktor nazivamo fiksnim. Ako faktor ima veliki ili beskonačan broj nivoa, a nivoe faktora koji su prisutni u podacima smatramo samo uzorkom svih mogućih nivoa faktora koji mogu uticati na zavisnu varijablu (a zanima nas efekat svih nivoa), onda taj faktor nazivamo slučajnim (McCulloch & Searle, 2001). Varijable same po sebi nisu fiksni ili slučajni faktori već to zavisi od toga kakve generalizacije na osnovu modela želimo da vršimo. Uzmimo za primer osobe/ispitanike. Ako rezultate nekog modela želimo da generalizujemo na populaciju kojoj ispitanici pripadaju, uzorak ispitanika na kojem vršimo analizu tretiraćemo kao slučajan faktor, a efekat ispitanika kao slučajan efekat. Međutim, ako nalaze ne želimo da generalizujemo na populaciju iz koje taj uzorak potiče, već nas zanima samo taj konkretan uzorak, onda ispitanike možemo tretirati kao fiksni faktor.

tačnog odgovora u uzorku, dobijena na osnovu takvog modela, tada zavisi od modelskih predikcija na osnovu parametara stavki, ali i od distribucije osobine u uzorku (Embretson & Reise, 2000).

Dakle, procena parametara modela u IRT obavlja se obično nekom od funkcija maksimalne verodostojnosti. Opšti oblik funkcije za sklop odgovora ispitanika $X_i$, na testu sa binarno skorovanim stavkama, izgleda ovako (Fajgelj, 2020):

$$L(\boldsymbol{\beta}|X_i,\theta_i) = \prod_{j=1}^{m} P_j(\theta_i)^{x_{ij}}\, Q_j(\theta_i)^{1-x_{ij}}$$

[ 11 ]

U suštini radi se o proizvodu verovatnoća odgovora ispitanika *i* na stavku *j* (*j*=1...m). $P_j(\theta_i)$ je verovatnoća tačnog odgovora ispitanika *i* na stavku *j* u zavisnosti od njegovog nivoa $\theta$, a $Q_j(\theta_i)$ je verovatnoća pogrešnog odgovora, odnosno 1– $P_j(\theta_i)$. Vektor $\boldsymbol{\beta}$ sadrži sve parametre svih stavki, a procena se dobija na osnovu podataka iz $X_i$ (da ne bude zabune $x_{ij}$ je opaženi odgovor ispitanika *i* na stavku *j*). *Verodostojnost podataka* predstavlja proizvod verovatnoća sklopova svih ispitanika. Moramo napomenuti da se ovako izračunava združena verovatnoća *nezavisnih* događaja i zato mora biti zadovoljena pretpostavka lokalne nezavisnosti (Embretson & Reise, 2000).

Rekli smo da je verodostojnost proizvod verovatnoća odgovora koje je ispitanik dao (bilo tačnih ili netačnih). U navedenom opštem izrazu za verodostojnost (izraz [ 11 ]) $P_j(\theta)$ se stepenuje sa $x_{ij,}$ što je opaženi odgovor ispitanika. Ako je taj odgovor tačan (1), onda se vrednost $P_j(\theta)$ stepenuje sa 1 i ostaje nepromenjena, ali ako je ispitanik odgovorio pogrešno (0), onda postaje jednaka 1 i ne utiče na proizvod. Isto važi i za $Q_j(\theta)$ koja se stepenuje sa 1–$x_{ij}$. Ako je ispitanik odgovorio tačno (1), onda je 1–$x_{ij}$=0 i $Q_j(\theta)$ dobija vrednost 1 pa ne utiče na proizvod (odnosno na verovatnoću sklopa). Ako je ispitanik odgovorio pogrešno (0), onda se $Q_j(\theta)$ stepenuje sa 1 i ne menja se, pa učestvuje u proizvodu.

Ovaj proizvod je jako mali broj pa se u računu umesto verodostojnosti obično koristi njen prirodni logaritam (engl. *LogLikelihood* – LL).[19]. Verovatnoće odgovora na stavke su brojevi između 0 i 1. Njihovim množenjem u cilju dobijanja združene verovatnoće za sklop odgovora ispitanika nastaju još manji brojevi (npr. 0,40*0,10=0,04). Ti brojevi su nekada toliko mali da ni računari ne mogu da ih izračunaju bez gubitka preciznosti (Embretson & Reise, 2000). Logaritamskom transformacijom verovatnoće sa racio nivoa merenja prelaze na intervalni nivo, a operacija množenja menja se u sabiranje. Tako LL (logaritam verodostojnosti) nekog sklopa odgovora nije proizvod verovatnoća tačnih odgovora na pojedinačne stavke već zbir prirodnih logaritama tih verovatnoća[20].

[19] Prirodni logaritam je logaritam sa osnovom *e* (Eulerov broj, matematička konstanta 2.71828). Podsećanje: prirodni logaritam od X je broj kojim bi trebalo stepenovati e da se dobije vrednost X.

[20] U izrazu [ 12 ] kao i u izrazu [ 11 ] na rezultat utiču samo verovatnoće opaženih odgovora ispitanika. Npr. ispitanik *i* je na pitanje *j* odgovorio tačno ($x_{ij}$=1), a modelska verovatnoća da tačno odgovori na to pitanje bila je 0,52. Kada to uvrstimo u izraz [ 12 ] to izgleda ovako (zanemarimo sumaciju po stavkama): 1*$ln$(0,52)+(1–1)$ln$(0,48) pošto je $Q$=1–$P$. Kada malo sredimo izraz dobijamo ln(0,52)+0*$ln$(0,48)= $ln$(0,52).

$$LL(\boldsymbol{\beta}|x_i, \theta_i) = \sum_{j=1}^{m} [x_{ij} ln\,(P_j(\theta_i)) + (1 - x_{ij}) ln(Q_j(\theta_i))]$$

[ 12 ]

*Prirodni logaritam verodostojnosti matrice podataka* predstavlja sumu prirodnih logaritama verovatnoća pojedinačnih sklopova odgovora ponderisanih njihovim brojem u matrici podataka (Fajgelj, 2020)..

Metod maksimalne verodostojnosti je procedura pomoću koje tražimo vrednosti parametara nekog modela za koje je najverovatnije da stoje iza procesa generacije podataka, odnosno da će biti ostvarena „maksimalna verodostojnost" (Embretson & Reise, 2000).

Na primer, ako tražimo lokaciju na kontinuumu $\theta$ (nivo osobine) na kojoj je neki sklop odgovora najverovatniji (uzevši u obzir parametre stavki), ono što bismo mogli da uradimo je da izračunamo navedenu funkciju za svaki nivo osobine i jednostavno vidimo gde je verodostojnost najveća. Ovakav metod daje najbolje procene, ali je računski vrlo zahtevan (Embretson & Reise, 2000). Osim toga, treba imati u vidu i veliki broj mogućih različitih sklopova odgovora i iterativnu prirodu procene IRT parametara, koja još više usložnjava ovu proceduru.

Zbog toga se za pronalaženje moda (maksimuma) funkcije verodostojnosti koriste procedure poput *Njutn-Rafsonove* (*Newton-Raphson*), koje znatno smanjuju broj potrebnih izračunavanja i olakšavaju nalaženje maksimuma funkcije. Potreba za ovakvim procedurama postaje sve manja sa sve jačim i dostupnijim računarima.

Svejedno, ukratko ćemo opisati Njutn-Rafsonovu proceduru za funkciju LL. Za određenu vrednost $\theta$ se izračunaju prvi (LL') i drugi izvod (LL'') funkcije logaritma verodostojnosti. Prvi izvod predstavlja vrednost nagiba tangente na funkciju LL u toj tački $\theta$.

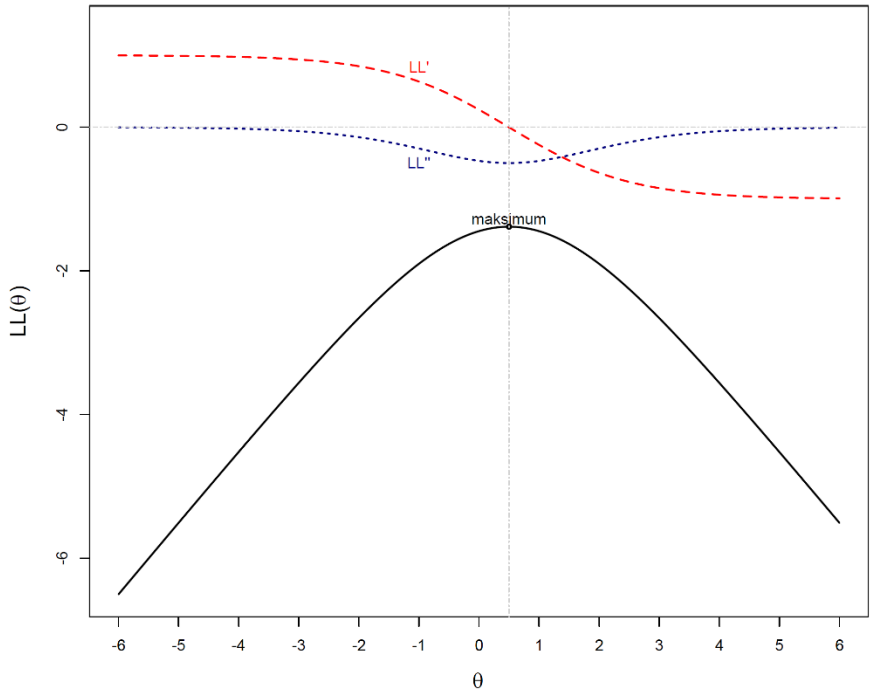
Ova funkcija je po pravilu konkavna (na gore) i izgleda kao na slici (Slika 7 – crna puna linija). Prikazana funkcija LL ima maksimum u tački $\theta$=0,5 logita i to je označeno vertikalnom sivom isprekidanom linijom i tačkom na liniji funkcije.

U suštini, ono što mi tražimo je tačka na kontinuumu $\theta$ u kojoj funkcija LL ima najviši nivo (njen mod). U toj tački, nagib funkcije, odnosno njen prvi izvod je 0. Naime, uobičajeno je da funkcija LL raste do određene vrednosti (maksimuma) i onda kreće da opada, odnosno – konkavna je. Dok funkcija raste njen prvi izvod je pozitivan, a kada opada negativan. Na slici (Slika 7 – crvena isprekidana linija) vidimo da su vrednosti prvih izvoda za vrednosti $\theta$ koje su niže od one u kojoj LL funkcija ima maksimum – pozitivne, a za one više – negativne. U maksimumu funkcije ta vrednost je 0 (na slici je to presek vertikalne i horizontalne

Drugim rečima, svodi se na prirodni logaritam verovatnoće opaženog odgovora $x_{ij}$. Da je isti ispitanik odgovorio pogrešno rezultat bi bio: $0*ln(0{,}52)+(1-0)ln(0{,}48)=0*ln(0{,}52)+1*ln(0{,}48)= ln(0{,}48)$.

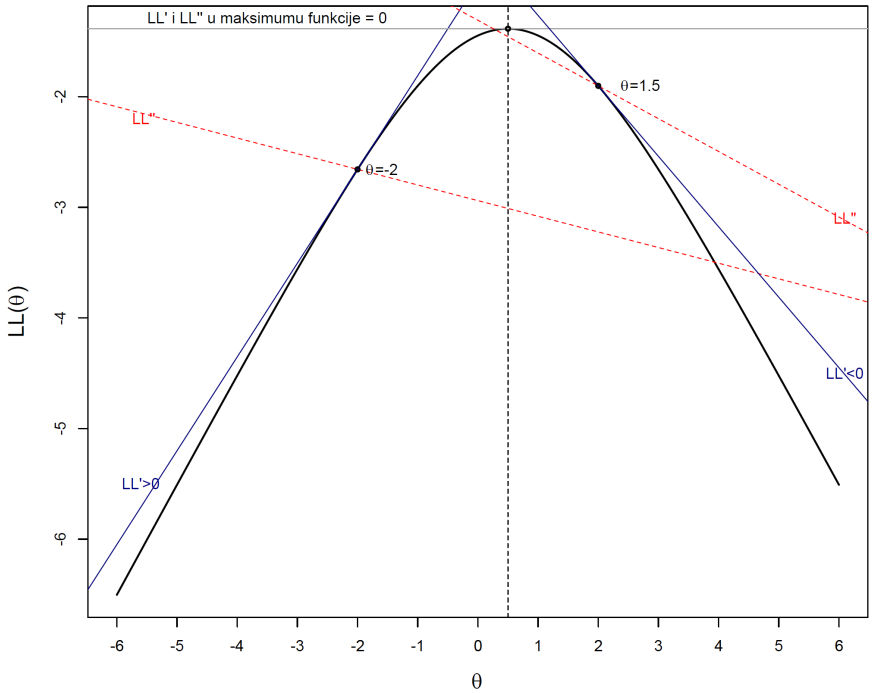
Inače, ovaj račun se ponavlja za svaku stavku i to se sumira da bi se dobila vrednost LL za određeni sklop odgovora.

sive linije).




Slika 7 – Funkcija logaritma verodostojnosti (LL) i funkcije njenih prvih (LL') i drugih (LL'') izvoda

Dakle, ako je vrednost $\theta$ potcenjena, prvi izvod (nagib) će biti pozitivan. Na slici (Slika 8) je to slučaj za $\theta$=–2. Ako je vrednost $\theta$ precenjena, prvi izvod biće negativan. Na slici (Slika 8) je to slučaj za $\theta$=1.5. Prvi izvodi označeni su plavim punim linijama i oznakama LL'.




Slika 8 – Funkcija LL sa prikazanim prvim i drugim izvodima za vrednosti $\theta$ od –2 i 1.5

Prema tome, ako smo dobili pozitivan prvi izvod za testiranu veličinu $\theta$, to znači da je vrednost koju tražimo veća i da bi i sledeća testirana vrednost trebalo da bude veća. Odnos prvog i drugog izvoda funkcije LL nam govori za koliko bi ona trebalo da bude veća. Drugi izvod nam govori o nagibu prvog izvoda. Na slici (Slika 8) drugi izvodi su predstavljeni crvenim isprekidanim linijama i oznakama LL''. Kod konkavnih funkcija kao što je LL, drugi izvodi su uvek negativni[21]. Što je LL funkcija bliža maksimumu, vrednost drugog izvoda je manja[22] (Slika 7), nagib funkcije LL brže se menja (usporava), što ukazuje da je funkcija LL bliža vrhu. Prema tome, što je drugi izvod funkcije LL niži, bliži smo traženoj vrednosti i ne bi trebalo da previše menjamo vrednost koju testiramo. Matematički se vrednost za koju bi trebalo promeniti testirani nivo $\theta$ izračunava kao količnik prvog i drugog izvoda (Embretson & Reise, 2000). Zbog činjenice da je drugi izvod uvek negativan, ako je prvi izvod pozitivan, onda će količnik biti negativan i obrnuto. Kao sledeća vrednost $\theta$ koju bi trebalo testirati uzima se razlika između tekuće vrednosti $\theta$ i količnika prvog i drugog izvoda. U primeru sa slike (Slika 8), ako je $\theta$=1.5 logita, prvi izvod je LL'=–0,46, a drugi LL''=–0,39. Količnik prvog i drugog izvoda biće 1,18. Prema tome, sledeća vrednost koju bi trebalo da testiramo je 1,5–1,18=0,32 logita, što je nešto niže od traženog maksimuma (0,50 logita).

Ovaj postupak se ponavlja sve dok vrednost količnika prvog i drugog izvoda ne postane manja od neke male vrednosti (recimo 0,001), odnosno dok korekcija ne postane beznačajna.

Prilikom sprovođenja ove procedure potrebno je postaviti ograničenje na visinu korekcije (na 1 ili neku sličnu vrednost), jer u pojedinim slučajevima korekcija može biti vrlo velika. Razlog tome je što nekada prvi izvod može biti relativno veliki broj, a drugi izvod jako mali. Količnik ova dva broja biće vrlo veliki broj i može proces pretrage odvesti daleko od tražene vrednosti (Thompson, 2009). Ovo se dešava na delovima funkcije LL gde je nagib visok i neznatno se menja.

Alternativna procedura za nalaženje maksimuma funkcije LL je procedura *polovljenja* (engl. *bisection*) (Thompson, 2009). Potrebno je odabrati dve tačke ($a$ i $b$) na funkciji LL koje odgovaraju širokom rasponu osobine (recimo –3 i 3 logita) i tačku ($m$) koja se nalazi tačno između njih. Zatim se porede vrednosti funkcije u dve tačke. Deo funkcije između srednje tačke i tačke koja ima nižu vrednost se izbacuje iz dalje pretrage, jer u tom opsegu funkcija ima niže vrednosti. Određuje se nova srednja tačka i postupak se ponavlja dok se ne dođe do maksimuma.

Druga varijanta ovog postupka je da se polovljenje radi na funkciji prvih izvoda funkcije LL (Thompson, 2009). Na isti način se odrede dve krajnje i centralna tačka. Zatim se porede prvi izvodi funkcije LL u tačkama $a$ i $b$. Ako je jedan izvod pozitivan, a drugi negativan, to znači da je maksimum funkcije LL između te dve tačke (jer je u njemu prvi izvod jednak 0). Nakon toga se izračuna i prvi izvod LL funkcije u središnjoj tački $m$. Ako je on pozitivan, iz pretrage možemo eliminisati deo funkcije na kojem se nalazi tačka

[21] Ako se procedura primenjuje na funkciji verodostojnosti (a ne njenog logaritma), potrebno je uvesti dodatna ograničenja jer funkcija ima drugačiji oblik i drugi izvodi nisu uvek negativni pa je potrebno kontrolisati predznak (Thompson, 2009).

[22] Njegova apsolutna vrednost je viša, ali zbog negativnog predznaka vrednost je niža. Drugim rečima, nagib je strmiji, ali negativan. To znači da nagib funkcije LL sve brže opada.

u kojoj je prvi izvod takođe pozitivan (jer se 0 ne nalazi između njih). Ako je taj prvi izvod tačke *m* negativan, onda se eliminiše deo funkcije prema tački u kojoj je izvod takođe negativan. Tačka *m* i preostala tačka sada čine novi raspon i između njih je potrebno odrediti novu središnju tačku. Procedura se ponavlja dok se ne pronađe tačka u kojoj je prvi izvod 0, a kao što smo rekli, to je maksimum funkcije LL.

Procedura polovljenja je manje efikasna od Njutn–Rafsonove procedure i osetljiva je na asimetrije funkcije LL (ali ne i kada se koristi varijanta procedure na prvim izvodima funkcije LL) (Thompson, 2009).

U nastavku ćemo opisati tri najčešće korišćene metode procene u IRT modelima: JMLE, CMLE i MMLE. Bitno je znati da ovi načini procene neće dati identične rezultate, ali će oni u velikoj meri biti slični. Svaki od ovih metoda ima svoje prednosti i mane na koje ćemo se osvrnuti posle opisa.

### 3.1.1 JMLE

JMLE se može sresti i pod nazivom *bezuslovna procena maksimalne verodostojnosti* i skraćenicom UCON (engl. *Unconditional Maximum Likelihood Estimation*). Ovaj postupak je vrlo čest u softveru za IRT, a jedan od razloga za to je taj što je relativno lak za programiranje (Embretson & Reise, 2000). Iako postoje varijante ovog postupka namenjene višeparametarskim modelima (Hambleton i ostali, 1991; Harris, 1989), ovde će zbog jednostavnosti biti prikazan postupak namenjen binarnim stavkama i jednoparametarskom modelu (Moulton, 2003).

Postupak ima dve faze. U prvoj fazi procenjuju se parametri ispitanika. Da bismo metodom maksimalne verodostojnosti procenili parametre ispitanika potrebni su nam i parametri stavki. Verovatnoća nekog sklopa odgovora zavisi i od parametara ispitanika i od parametara stavki. Problem je što su nam i jedni i drugi nepoznati. Problem sa nepoznatim parametrima stavki rešava se tako što se odrede neki provizorni, privremeni parametri, recimo mogu biti postavljeni na 0 (Embretson & Reise, 2000).

Nakon što se uz pomoć provizornih parametara stavki procene parametri ispitanika, oni se koriste za procenu parametara stavki. Postupak je iterativan i u svakoj narednoj iteraciji parametri stavki i ispitanika se unapređuju dok model ne konvergira (Embretson & Reise, 2000).

U Rašovom modelu procena parametara stavki i ispitanika se obavlja na relativno jednostavan način i može se izračunati (recimo) ručno (Fajgelj, 2020). Matrica podataka može biti u obliku klasične tabele u kojoj su u redovima ispitanici, u kolonama stavke, a u ćelijama odgovori ispitanika na stavke. Takođe, može biti u sažetom obliku gde su ispitanici grupisani po ukupnim skorovima (videti odeljak 3.1.1.2 koji se odnosi na CMLE za objašnjenje zašto je to moguće). U ovakvim sažetim matricama, u ćelijama su proporcije tačnih odgovora grupa ispitanika na stavke. Ako je to slučaj, ove proporcije moraju se pretvoriti u prirodne logaritme šansi $\ln(p/(1-p))$. *Šanse* (*dobitni odnosi*) su racio varijable i nisu aditivni. Zato se u računu koristi njihov prirodni logaritam koji ih intervalizuje i čini aditivnim. Na osnovu proseka ovih dobitnih odnosa za svaku stavku računaju se privremene Rašove mere težine, a prosek po grupama formiranim po skorovima daje privremene Rašove mere nivoa osobine (za svaki ukupan skor). U slučaju stavki, prikazani način računanja dobitnog odnosa ne daje *težinu* već *lakoću* i da bi se to korigovalo množi se sa –1 (ekvivalentno bi bilo da se težina stavki računa kao $\ln((1-p)/p)$). Tako je $\theta_i=\ln(p/(1-p))$, a $b_j= \ln((1-p)/p)$ ili $b_j=-\ln(p/(1-p))$.

Ako je matrica u izvornom obliku *stavke * ispitanici*, za svakog ispitanika provizorni početni nivo osobine računa se jednostavno na osnovu proporcije tačnih odgovora koje je dao na testu. On je takođe u formi *prirodnog logaritma dobitnog odnosa* $\ln(p/(1-p))$, gde je $p$ proporcija tačnih odgovora koje je ispitanik dao na testu. Na isti način računa se privremena težina stavke, s tim da je u ovom slučaju $p$ proporcija tačnih odgovora na tu stavku u uzorku. Naravno, i u ovom slučaju potrebno je takvu meru pomnožiti sa –1. Problem predstavljaju maksimalni ekstremni skorovi ispitanika i stavki. Naime, ako je proporcija tačnih odgovora ispitanika $p_i$=0, onda je dobitni odnos 0, a nije moguće izračunati prirodni logaritam za tu vrednost. Takođe, ako je proporcija tačnih odgovora ispitanika $p_i$=1 (odgovorio na sve stavke tačno) neće biti moguće izračunati dobitni odnos. U tom slučaju delilac $(1-p)$ će biti jednak 0, a deljenje sa nulom nije definisano, pa je dobitni odnos neizračunljiv. U tom slučaju se pribegava dodavanju ili oduzimanju neke male vrednosti (recimo 0,05) na vrednost $p_i$. Ako je proporcija tačnih odgovora ispitanika $p_i$=0 dodaje se 0,05, a ako je $p_i$=1 oduzima se 0,95. Isto se može primeniti i na stavke (i na grupisane podatke).

Dalje se postupak ne razlikuje bez obzira da li je matrica podataka sažeta ili ne.

Dobijene mere težine stavki se uprosečavaju da bi se dobila prosečna nekorigovana težina stavki.

Težine stavki se dalje koriguju tako što se pomenuti prosek oduzima od svake težine stavke. Na taj način prosečna težina stavki se postavlja na 0, odnosno, vrednosti se pretvaraju u devijacije od aritmetičke sredine.

Na osnovu tako procenjenih i korigovanih težina stavki i nivoa osobine, računaju se verovatnoće pozitivnih odgovora koristeći formulu modela [ 21 ]. Ove procene porede se sa stvarnim podacima i na osnovu njih se računaju odstupanja, odnosno reziduali.

Reziduali su odstupanja modelskih predviđanja od stvarnih odgovora ispitanika. Računaju se tako što se od odgovora ispitanika na stavku oduzme procenjena verovatnoća tačnog odgovora na osnovu modela.

$$\varepsilon_{ij} = x_{ij} - p_{ij}$$

[ 13 ]

U izrazu [ 13 ] $x_{ij}$ je opaženi odgovor ispitanika $i$ na stavku $j$, a $p_{ij}$ je predviđeni odgovor (odnosno verovatnoća predviđanog odgovora). U slučaju sa sažetim matricama podataka, umesto opaženog odgovora ispitanika uzima se prirodni logaritam opaženih šansi za neki *skor* (grupu formiranu po ukupnom skoru). Takođe, kao predviđeni odgovor uzima se predviđanje za grupu ispitanika sa određenim skorom. Tada, indeks $i$ označava grupu ispitanika sa istim skorom[23].

Reziduali se sumiraju po ispitanicima (skorovima) i po stavkama. Cilj cele procedure procene je smanjiti sume reziduala. Ovakvo rešenje se naziva maksimalnom verodostojnošću (MLE), a daje slično rešenje metodi najmanjih kvadrata (Embretson & Reise, 2000).

[23] I u nastavku opisa procedure, kada kažemo ispitanik ili se pojavi indeks *i* imajte u vidu da se u radu sa sažetim matricama podataka to odnosi na grupu ispitanika sa istim skorom.

Za stavke se ove sume reziduala množe sa –1. To se radi zato što nam procene na osnovu modela govore koliko je verovatno da će ispitanik tačno rešiti stavku. Što je verovatnoća viša, ispitanik ima viši nivo osobine. Kada isti podatak gledamo sa aspekta stavke, viša verovatnoća odgovora na neku stavku znači manju težinu stavke. Pošto je nama potrebna težina, sume reziduala se moraju pomnožiti sa –1.

U toku procene računa se i *ukupna suma reziduala* za celu matricu podataka. Ta suma se kvadrira i predstavlja podatak koji nam govori o kvalitetu modela u celini. Cilj procedure procene je da ova suma bude što bliža 0 (Embretson & Reise, 2000).

Na osnovu predviđenih verovatnoća računaju se *varijanse predviđanja* za svakog ispitanika na svakoj stavci i sumiraju po ispitanicima (i po stavkama). *Varijanse predviđanja* računaju se kao $p_{ij}(1-p_{ij})$, gde je $p_{ij}$ predviđena verovatnoća tačnog odgovora ispitanika *i* na stavku *j* dobijena na osnovu formule modela.

Dobijene sume se množe sa –1 i koriste za korekciju nivoa osobine i težina stavke. *Varijanse predviđanja* su kvadrati standardnih grešaka (za ispitanike i za stavke). Korekcija se vrši tako što se od prethodno procenjenih nivoa osobine nekog ispitanika (ili težine neke stavke) oduzme količnik sume reziduala i negativne sume varijansi predviđanja za tog ispitanika (ili stavku). Za stavke se pored ove korekcije vrši još jedna. Računa se prosek korigovanih težina. Taj prosek se oduzima od svake korigovane težine i na taj način prosečna težina stavki opet postaje 0 (radi se centriranje).

Potom se na osnovu ovako korigovanih nivoa osobine i težina stavki vrši nova procena verovatnoća tačnog odgovora na osnovu modela i postupak se ponavlja dok *ukupna kvadrirana suma reziduala* ne bude bliska 0 i tada se proces procene zaustavlja. Iz reziduala se može izračunati i RMSE (engl. *Root Mean Squared Error*). Kao što samo ime kaže to je kvadratni koren iz prosečne vrednosti kvadriranih reziduala, odnosno, prosečni rezidual izračunat preko RMS transformacije.

Nakon toga, kvadratni koren recipročne vrednosti sume varijansi predviđanja za nekog ispitanika predstavlja standardnu grešku za tog ispitanika (Embretson & Reise, 2000):

$$SE_i = \sqrt{\frac{1}{\sum_{j=1}^{j} p_{ij}(1-p_{ij})}}$$

[ 14 ]

Isto važi i za stavke, samo što se koristi suma varijansi po ispitanicima. Standardna greška biće manja što je suma varijansi predviđanja veća. Kako su u ovom primeru pitanju binarne stavke, ova suma biće veća što je verovatnoća tačnih odgovora ispitanika bliža 0,5 (50%), a to je slučaj kada su težine stavki bliske nivou osobine ispitanika. Za što preciznije određivanje mere ispitanika to znači da test ne sme biti ni suviše težak, ni suviše lak. Isto tako, za što preciznije određivanje težine *stavke* bitno je da uzorak ispitanika bude takav da mu stavka nije ni preteška, ni prelaka. Ranije je rečeno da mere stavki u IRT modelima ne zavise od uzorka, odnosno od distribucije osobine u njemu – i to je tačno, međutim, standardne greške procena tih mera zavise (DeMars, 2010; Embretson & Reise, 2000).

### 3.1.2 CMLE

*Procena uslovne maksimalne verodostojnosti* je drugi način procene parametara. U engleskom jeziku javlja se i pod nazivom *Fully Conditional Maximum Likelihood Estimation* (*FCON*). Ovaj metod se primenjuje samo u Rašovim, jednoparametarskim modelima. Postupak je sličan kao JMLE za jednoparametarske modele. Razlika je u tome što se verovatnoća tačnog odgovora računa na osnovu ukupnog skora ispitanika, a ne na osnovu parametara ispitanika koji su nepoznati (Embretson & Reise, 2000). Drugim rečima, bazira se na proceni maksimalne verodostojnosti dobijanja ukupnih skorova koji se koriste kao polazne procene nivoa osobine ispitanika (Mair & Hatzinger, 2007). Problem nepoznatih nivoa latentne osobine u modelu rešava se tako što se oni ni ne uključuju u proces procene parametara stavki i tretiraju se kao ometajući (engl. *nuisance*) parametri. Ovo je moguće zato što je u jednoparametarskim modelima ukupan skor dovoljan statistik za nivo osobine. Ispitanici sa istim ukupnim skorovima (ne nužno na istim stavkama) dobijaju istu meru osobine. Isto tako, za procenu parametra težine stavke dovoljan statistik je broj ispitanika koji je tačno ili potvrdno odgovorio na stavku. Sve stavke na koje je tačno ili potvrdno odgovorio jednak broj ispitanika imaće istu procenu težine, bez obzira na konkretne ispitanike koji su tako odgovorili (Embretson & Reise, 2000). Ono što će se razlikovati i kod ispitanika i kod stavki su standardne greške. Zbog toga što je ukupan skor dovoljan statistik za procenu nivoa osobine, on može da ga zameni u procesu procene parametara stavki. Za procenu parametara metodom CMLE nisu neophodni sirovi podaci niti matrica odgovora na stavke. Dovoljni su ukupni skorovi za ispitanike i broj tačnih odgovora po stavci za stavke.

Tek kada su procenjeni parametri stavki, oni se tretiraju kao poznati i na osnovu njih se procenjuju parametri ispitanika.

U ovom metodu, da bi model mogao biti identifikovan neki parametri moraju biti fiksirani na 0. U *Rašovom modelu za binarne stavke* (odeljak 4.1.1) to je težina jedne stavke, u *modelu skala procene* (odeljak 4.2.1.2) to su parametri kategorija 0 i 1, a u *modelu stepenovanog ocenjivanja* (odeljak 4.2.1.3) to je parametar kategorije 0 kod svih stavki i parametar još jedne kategorije (Mair & Hatzinger, 2007).

### 3.1.3 MMLE

U proceni metodom marginalne maksimalne verodostojnosti (*Marginal Maximum Likelihood estimation* – MMLE) se verovatnoće sklopova odgovora tretiraju kao očekivanja iz distribucije u populaciji, a opaženi podaci tretiraju se kao slučajan uzorak. Na taj način MMLE modelira verovatnoću javljanja sklopova odgovora u populaciji (Embretson & Reise, 2000), odnosno tretira ih kao slučajni, a ne fiksni efekat.

MMLE razdvaja procenu parametara stavki od procene parametara ispitanika. Prvo se procenjuju samo parametri stavki. Tek kada su oni poznati procenjuju se parametri ispitanika primenom metode maksimalne verodostojnosti ili Bejzijanskim pristupom (Embretson & Reise, 2000).

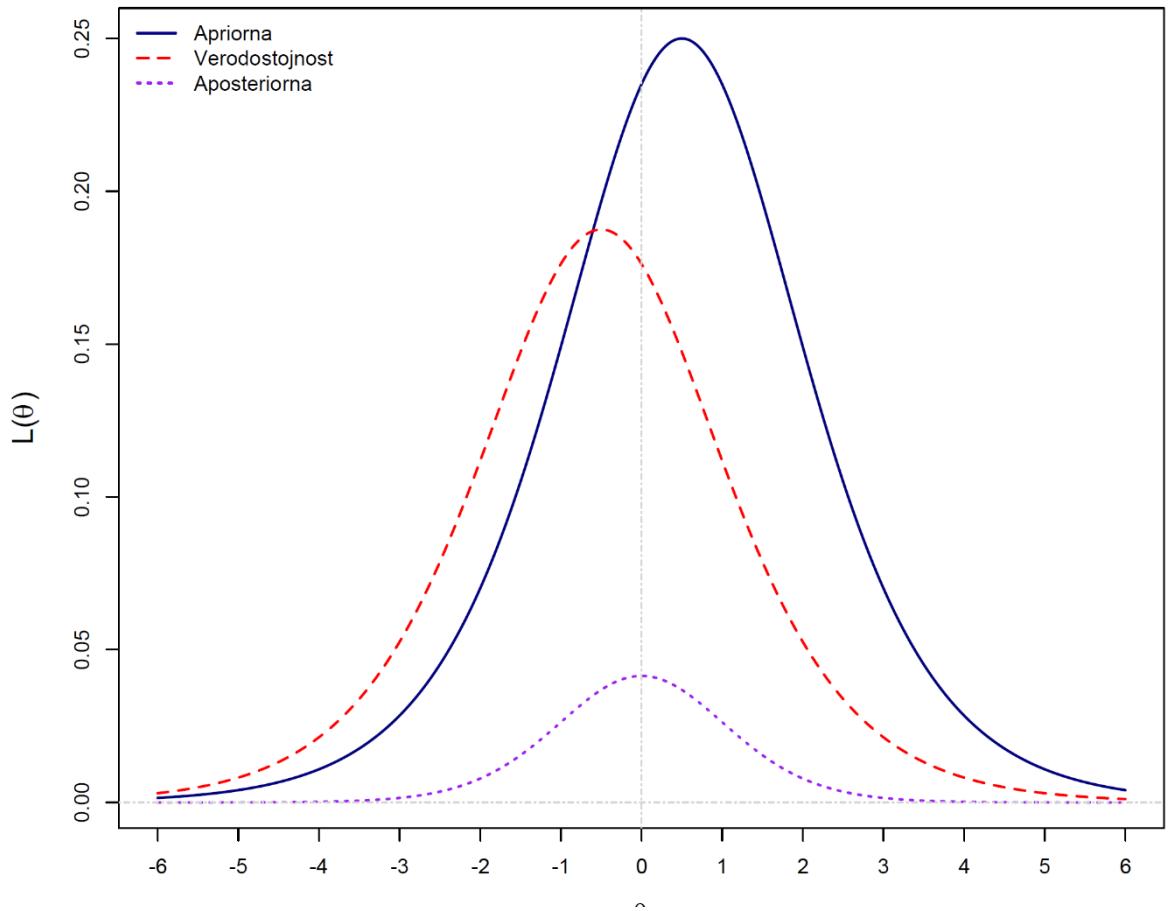
Ako se uzorak ispitanika tretira kao slučajan uzorak iz neke populacije, nije više neophodno procenjivati nivoe osobine ispitanika za potrebe procene parametara stavki. Možemo da pretpostavimo oblik distribucije nivoa osobine i na osnovu takve distribucije izračunamo verovatnoće javljanja svakog od nivoa. U tu svrhu najčešće se koristi standardizovana normalna distribucija (sa $M$=0 i $s^2$=1), ali moguće su i druge

vrste distribucija koje se čine primerenim (Embretson & Reise, 2000). Ako je uzorak dovoljno velik, verovatnoće određenih nivoa osobine možemo i empirijski da izračunamo na osnovu podataka koje imamo.

U terminima Bejzijanske statistike ova pretpostavka normalne raspodele predstavlja *apriornu* distribuciju (engl. *prior*). To je naše uverenje ili pretpostavka o verovatnoći javljanja nekog nivoa osobine u populaciji (Embretson & Reise, 2000).

U Bejzijanskom pristupu sledeći korak bi bio da pribavimo nove podatke o pojavi koja nas zanima. U ovom slučaju to je verovatnoća javljanja sklopa odgovora za neki nivo osobine, a na osnovu predviđanja modela.

Kada je to učinjeno, preostaje da izračunamo aposteriornu verodostojnost (engl. *posterior*) javljanja određenog sklopa odgovora u zavisnosti od nivoa osobine. Aposteriorna verodostojnost predstavlja proizvod verodostojnosti na osnovu modela i apriorne verovatnoće i može se predstaviti kao na narednoj slici (Slika 9). To je verodostojnost sklopa odgovora na različitim nivoima osobine u populaciji, ažurirana na osnovu predviđanja modela.




Slika 9 – Apriorna, aposteriorna i funkcija verodostojnosti

Verovatnoća nekog sklopa odgovora $X_s$ u zavisnosti od nivoa $\theta$ ispitanika i parametara modela $\boldsymbol{\beta}$ izračunava se kao:

$$P(X_s|\theta,\boldsymbol{\beta}) = \prod_{\boldsymbol{j=1}}^{\boldsymbol{m}} p_j^{x_{ij}} (1-p_j)^{(1-x_{ij})}$$

[ 15 ]

gde je $p_j$ verovatnoća tačnog odgovora na stavku $j$ na osnovu modela, $x_{ij}$ opaženi odgovor ispitanika $i$ na stavku $j$, a $m$ ukupan broj stavki.

Verovatnoća javljanja sklopa odgovora za dati nivo $\theta_q$ tada predstavlja proizvod verovatnoće javljanja tog nivoa $\theta$ u populaciji $P(\theta_q)$ i verovatnoće javljanja sklopa odgovora na tom nivou osobine $P(X_s|\theta_q,\boldsymbol{\beta})$ koja je izračunata na osnovu modela.

Da bi se dobila *ukupna verovatnoća sklopa* odgovora potrebno je izračunati sumu prethodno navedenih proizvoda za različite nivoe $\theta$ (izraz [ 16 ]).

$$P(X_s|\boldsymbol{\beta}) = \sum_{q}^{\theta} P(X_s|\theta_q,\boldsymbol{\beta})\, P(\theta_q)$$

[ 16 ]

U prethodnom izrazu $X_s$ – označava sklop odgovora, $\boldsymbol{\beta}$ – je vektor parametara modela, a $P(\theta_q)$ – je verovatnoća javljanja nekog $q$ nivoa $\theta$ u populaciji.

Kako je $\theta$ kontinuirana varijabla, broj potencijalnih nivoa osobine je beskonačan (npr. 1,2 logita, 1,25 logita, zatim 1,253 logita...). Stoga se ovi proizvodi (izraz [ 16 ]) ne izračunavaju za svaki mogući pojedinačni nivo osobine, već za određene raspone nivoa (na to se odnosi $q$ u izrazu). Verovatnoće se izračunavaju kao integrali tih raspona (intervala osobine). Distribucija se deli na više intervala koji se nazivaju *Gausove kvadrature*, pri čemu je svaka od kvadratura predstavljena jednom reprezentativnom vrednošću koja se naziva *kvadraturnom tačkom* ili *kvadraturnim čvorom* (engl. *quadrature node*). Izračunava se verovatnoća javljanja nivoa $\theta$ za svaki od intervala.

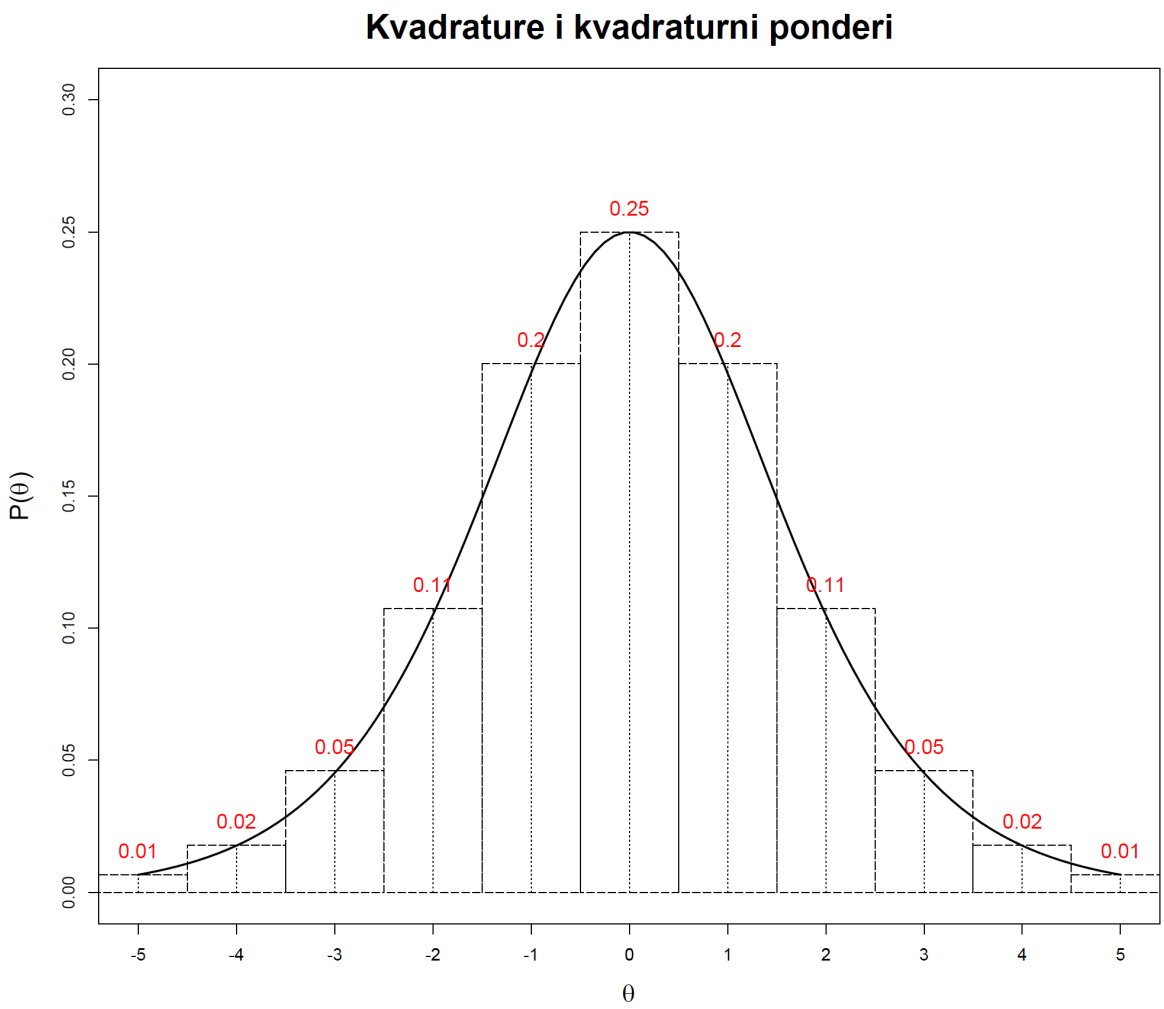
Umesto da se površina ispod krive računa kao integral u tom rasponu osobine, u praksi se koristi jednostavniji način. Zamislimo distribuciju osobine kao diskretnu distribuciju sastavljenu od četvorouglova (stubića), nalik histogramu sastavljenom od kvadratura (Slika 10). Za svaku kvadraturnu tačku ($X_r$) vezan je odgovarajući *kvadraturni ponder* (engl. *quadrature weight*). Kvadraturni ponder (označićemo ga sa $A(X_r)$) odgovara visini funkcije raspodele u kvadraturnoj tački. Na slici su lokacije kvadraturnih tačaka označene vertikalnim tačkastim linijama po sredini stubića koji označavaju kvadrature. Iznad su ispisani kvadraturni ponderi. Koristeći kvadraturne pondere, verovatnoće sklopova odgovora mogu se izračunati na sledeći način (de Ayala, 2022):

$$P(X_s|\boldsymbol{\beta})^* = \sum_{r}^{R} P(X_s|X_r,\boldsymbol{\beta})\, A(X_r)$$

[ 17 ]

gde $R$ označava broj kvadratura $r$, $X_s$ – sklop odgovora, $\boldsymbol{\beta}$ – vektor parametara modela, $A$ – kvadra-

turni ponder, $P(X_s|X_r, \boldsymbol{\beta})$ – predstavlja verovatnoću sklopa odgovora $X_s$ za kvadraturu $r$ na osnovu parametara modela.




Slika 10 – Kvadrature i kvadraturni ponderi standardne logističke raspodele

Što je veći broj kvadratura dobijene verovatnoće će bolje odslikavati distribuciju. Intervali kvadratura ne moraju biti jednaki. Primer su Hermit-Gausove (Hermit-Gauss) kvadrature koje su nejednake veličine, ali daju optimalne rezultate kod normalne distribucije (Embretson & Reise, 2000).

Pošto se u MMLE za procenu parametara stavki ne koriste procene parametara osobine ispitanika iz uzorka, već marginalne verovatnoće javljanja određenih nivoa osobine na osnovu pretpostavljene distribucije $\theta$, parametri stavki ne zavise od procene parametara ispitanika, ali zavise od njihove marginalne distribucije (Harwell i ostali, 1988). MMLE zato daje stabilnije procene parametara stavki jer sa povećanjem uzorka ne dolazi do povećanja broja parametara ispitanika i modela u celini.

*Verodostojnost podataka* predstavlja proizvod verovatnoća svih sklopova. Kao što je već rečeno, ovaj proizvod je jako mali broj, pa se umesto njega obično koristi njegov prirodni logaritam. Prirodni logaritam verodostojnosti (engl. *LogLikelihood* – LL) matrice podataka predstavlja sumu prirodnih logaritama

verovatnoća pojedinačnih sklopova odgovora ponderisanih njihovim brojem u matrici podataka (Embretson & Reise, 2000).

Za procenu parametara koristi se *Expectation Maximization* (EM) algoritam. Ovaj algoritam je tehnika maksimalne verodostojnosti koja se koristi kada postoje slučajne varijable koje nisu dostupne, a u ovom slučaju nisu dostupni nivoi latentne osobine ispitanika (Fajgelj, 2020; Harwell i ostali, 1988).

EM je iterativna procedura koja ima dve faze. Prva je faza u kojoj određujemo *očekivanja* (engl. *expectations*) i označava se sa E. U njoj se računa verovatnoća da će ispitanici tačno rešiti stavku i očekivani broj (proporcija) ispitanika na svakom od nivoa osobine predstavljenim kvadraturama. Kada je to utvrđeno, prelazi se u fazu *maksimizacije* (engl. *maximization* – označava se sa M) u kojoj se vrednosti iz E faze uvrštavaju u jednačinu modela i traže se vrednosti parametara koje će maksimizirati verodostojnost. Zatim se procene parametara dobijene u M fazi koriste u novoj E fazi da se unaprede očekivanja. Nakon toga sledi nova M faza u kojoj se unapređuju procene parametara i tako dok rešenje ne konvergira (Fajgelj, 2020).

### 3.1.4 Razlike između metoda procene parametara

Kao što je već rečeno, različite metode procene parametara daće i donekle različite rezultate. Tačne vrednosti procenjenih parametara neće se poklapati, ali će poredak i odnosi biti uporedivi.

Svaki od prikazanih metoda procene ima svoje prednosti i mane.

Trenutno se najčešće koristi metod procene parametara MMLE. Razlog je to što je primenljiv na *svim* IRT modelima, jednako je efikasan na kratkim i dužim testovima, a procene standardnih grešaka parametara stavki su dobre aproksimacije očekivane uzoračke varijanse parametara. Može izračunati procene za savršene skorove (svi tačni ili svi netačni) pa nema gubitka informacija za ispitanike, kao što je slučaj sa JMLE i CMLE. Takođe, omogućava poređenje ugnježdenih modela [24] po prirodnom logaritmu verodostojnosti (LL), odnosno, može se koristiti hi–kvadrat test za količnik verodostojnosti (engl. *Likelihood Ratio Chi-Squared Test* – LR). Nedostatak MMLE je što se mora pretpostaviti neka distribucija osobine, pa ako pretpostavka nije dobra parametri mogu biti pristrasni i lošiji od procena dobijenih na osnovu JMLE i CMLE (Embretson & Reise, 2000).

JMLE je takođe primenljiv na različitim IRT modelima, jedno- i višeparametarskim. Pošto se u proceni parametara stavki koriste provizorne procene nivoa osobine ispitanika, procene parametara mogu biti pristrasne, ali za to postoje korekcije koje tu pristrasnost u velikoj meri mogu korigovati. Zbog pristrasnosti procena parametara i procene njihovih standardnih grešaka mogu biti manje pouzdane. Kod testova fiksne dužine ne daje konzistentne ocene parametara. Sa povećanjem uzorka kvalitet procena se ne popravlja, jer se povećava broj parametara koji se procenjuje (de Ayala, 2022; Embretson & Reise, 2000). S obzirom na to

[24] Dva modela su ugnježdena ako parametri jednog modela predstavljaju podskup parametara drugog modela. Npr. Raschov model je ugnježden u dvoparametarskom logističkom modelu. Jednak je dvoparametarskom modelu za binarne stavke u kojem su parametri diskriminativnosti fiksirani na 1.

da JMLE procenjuje združenu verodostojnost podataka, izbacivanje neke od stavki iz testa zato što se pokazala lošom, zahteva da model bude ponovo procenjen kako bi se uklonili efekti loše stavke na procene osobine (de Ayala, 2022). I ovaj metod procene ima problema sa identifikacijom[25] modela, pa se težina jedne ili više stavki mora arbitrarno fiksirati na 0 (Fajgelj, 2020). Takođe, ne funkcioniše dobro na troparametarskom modelu (Fajgelj, 2020).

Prednost CMLE je to što nije potrebno pretpostaviti distribuciju latentne osobine, pa se tu ne može ni pogrešiti. Parametri stavki su nezavisni od nepoznatih nivoa osobine ispitanika jer su oni izostavljeni iz računa. Parametri dobijeni CMLE procenom u najvećoj meri zadovoljavaju princip invarijantnosti (Embretson & Reise, 2000). Za standardne greške parametara stavki je opravdanije reći da predstavljaju varijansu uzorkovanja nego kod JMLE. Takođe, omogućava poređenje ugnježdenih modela LR testom. Međutim, nedostatak je to što je ovaj način procene primenljiv samo na Rašovim modelima (jer je jedino tu ukupan skor dovoljan statistik). Takođe, ovim pristupom nije moguća procena parametara za stavke sa savršenim skorovima (tačno odgovorili svi ispitanici ili nijedan nije tačno odgovorio).

Problem sa JMLE i CMLE je što nivo osobine tretiraju kao fiksni efekat. Kreće se od procena mera ispitanika na osnovu konkretnih podataka. To rezultira nestabilnim procenama parametara sa povećanjem uzorka (de Ayala, 2022; Fajgelj, 2020; Linden, 2010), a to nije slučaj u MMLE.

## *3.2 Procena parametara ispitanika u MMLE*

U MMLE proceduri, tek kada je završena procena parametara stavki i utvrđeno da je model saglasan sa podacima pristupa se proceni parametara ispitanika.

Obično se koristi metod maksimalne verodostojnosti (ML) ili neka od dve varijante procene zasnovane na Bejzijanskom pristupu: *maximum a posteriori* (MAP) ili *expected a posteriori* (EAP) (de Ayala, 2022; Embretson & Reise, 2000).

U ML proceni traži se nivo osobine na kojem je sklop odgovora ispitanika najverovatniji. Verovatnoća sklopa odgovora računa se kao proizvod verovatnoća odgovora koje je ispitanik dao, a na osnovu skupa parametara stavki i formule modela. Model obično predviđa verovatnoću tačnog ili pozitivnog odgovora, ali odatle nam je poznata i verovatnoća netačnog odgovora. Ako je verovatnoća tačnog odgovora 0,7, onda je verovatnoća pogrešnog 0,3 (kada su u pitanju binarno skorovane stavke).

Pošto ne znamo nivo osobine koji ispitanik ima, izraz [ 15 ] bi trebalo izračunati za svaki mogući nivo $\theta$ i naći onaj kod kojeg je dobijena vrednost najveća. Taj nivo $\theta$ uzima se kao mera ispitanika. Međutim, kao i kod procene parametara stavki, pošto je broj potrebnih izračunavanja ogroman, koriste se procedure

[25] Kažemo da je model identifikovan ako je teoretski moguće izračunati jedinstvene procene svih njegovih parametara (Kline, 2015). Na primer, model 10=X+Y nije identifikovan jer parametri X i Y mogu uzeti bezbroj različitih vrednosti (2 i 8, 9 i 1, 10 i 0, -50 i 60...). Ako jedan od parametara fiksiramo na određenu vrednost, model postaje izračunljiv. Npr. ako X fiksiramo na 0, Y je lako izračunati.

poput polovljenja ili Njutn–Rafsonove, koje optimizuju proces traženja maksimuma funkcije logaritma verodostojnosti (procedura je detaljnije opisana u odeljku 3.1.1.3).

U jednoparametarskom modelu svi ispitanici sa jednakim sirovim skorom dobiće istu meru i standardnu grešku, bez obzira na to da li su tačne odgovore dali na lake ili teške stavke (Embretson & Reise, 2000).

U dvoparametarskom modelu, dovoljan statistik za procenu nivoa osobine je suma proizvoda kvadrata parametra diskriminativnosti i skorova na stavci ($\sum_j^m a_j^2 x_{ij}$). To praktično znači da se pre sumiranja skorovi na stavkama ponderišu parametrom diskriminativnosti. Višu procenu nivoa osobine dobiće ispitanici koji su tačno odgovorili na diskriminativnije stavke. Moguće je da ispitanik koji ima manje tačnih odgovora dobije višu procenu osobine, ako je odgovorio na manji broj ali visoko diskriminativnih stavki (Embretson & Reise, 2000). Težina tačno odgovorenih stavki utiče na nivo osobine jer određuje lokaciju LL funkcije pa i njen maksimum.

Nusproizvod ML procedure je procena informativnosti testa. Naime, apsolutna vrednost drugog izvoda funkcije LL daje informativnost testa koja je u IRT-u osnovni pokazatelj preciznosti merenja. Preko informativnosti moguće je izračunati standardnu grešku (videti odeljak 7).

ML procedura na velikim uzorcima daje asimptotski nepristrasne procene, greške su normalno distribuirane i efikasna je. Ovo važi ako je model saglasan sa podacima. Metod bolje funkcioniše na većem broju stavki (Embretson & Reise, 2000).

Kriva LL je obično simetrična, ali nekada se mogu javiti anomalije u vidu lokalnih maksimuma („izbočina“ na krivoj) ili asimetrije. Lokalni maksimumi se često javljaju u troparametarskom modelu i mogu uticati da maksimum funkcije LL bude pogrešno ocenjen (Thompson, 2009).

Procena maksimalne verodostojnosti nije moguća ako ispitanik nije odgovorio tačno ili netačno bar na jedno pitanje, što predstavlja problem za primenu u *računarskom adaptivnom testiranju* (Embretson & Reise, 2000). Takođe, ako je ispitanik tačno odgovorio na sve stavke, kriva verodostojnosti biće monotono rastuća sa maksimumom u +∞, a ako nije odgovorio ni na jednu biće monotono opadajuća i maksimum će biti u –∞. U praksi, to znači da nećemo moći da nađemo odgovarajuću procenu nivoa $\theta$.

Način da se to prevaziđe je uvođenje prethodnih znanja u izračunavanje verodostojnosti. To se ostvaruje uvođenjem *apriori* distribucija. Jedan od takvih načina procene je *maximum a posteriori* (MAP). Kao što je rečeno, ovde se pretpostavlja distribucija verovatnoća osobine, a ispitanici se posmatraju kao slučajan uzorak iz te distribucije. Mogu se koristiti razne distribucije koje se čine primerenim, ali najčešće je to standardizovana normalna distribucija (sa M=0 i $s^2$=1). Maksimum logaritma funkcije verodostojnosti određuje se na osnovu *aposteriori* funkcije logaritma verodostojnosti koja nastaje množenjem modelski procenjenih verovatnoća apriornom distribucijom. Na taj način se rešava problem sa monotono rastućim ili opadajućim funkcijama LL. Proces traženja maksimuma funkcije isti je kao kod ML metoda, odnosno, moguće je izračunati sve vrednosti funkcije i naći maksimum ili koristiti procedure polovljenja ili Njutn-Rafsonovu.

*Expected a posteriori* – EAP je druga procedura zasnovana na Bejzijanskom pristupu (Wu i ostali, 2016). Razlika u odnosu na MAP je ta što se ne traži maksimum (mod) aposteriorne funkcije LL, već njena aritmetička sredina (statističko očekivanje) i što se koristi diskretna apriorna distribucija (Embretson & Reise, 2000). Ovaj pristup je pogodan kada sumnjamo da postoji lokalni maksimum funkcije LL. Međutim, EAP je mnogo zahtevnija procedura kada su izračunavanja u pitanju (Thompson, 2009) jer je potrebno izračunati sve vrednosti funkcije i naći prosek.

Uticaj apriorne distribucije veći je što je test kraći. U tom slučaju procena nivoa osobine biće više pomerena ka aritmetičkoj sredini apriorne distribucije. Ako je test vrlo dugačak, efekat apriorne distribucije može biti neznatan ili nikakav (Embretson & Reise, 2000). Vrlo je bitno odabrati odgovarajuću apriornu distribuciju jer u protivnom procene parametara mogu biti pristrasne.

# 4 IRT modeli

Pod IRT modelima podrazumevaćemo matematičke funkcije kojima se opisuju karakteristične krive, ili verovatnoće odgovora u zavisnosti od nivoa latentne osobine (Thissen & Steinberg, 1986). Možemo ih podeliti na više načina.

Prema pretpostavci o ***odnosu latentne varijable i manifestnog ponašanja*** (odgovora), IRT modele možemo podeliti na *normalne i logističke*. U suštini, razlika je u pitanju oblika KKS, odnosno vezne funkcije koja dovodi u vezu manifestno ponašanje i latentnu varijablu. Za razliku od logističkih modela u kojima su parametri izraženi u logitima, u normalnim modelima se vrednosti parametara izražavaju u *z* skorovima.

S obzirom na to da su normalni modeli matematički znatno kompleksniji, najčešće se izračunavanja obavljaju po logističkim modelima koji se jednostavnom transformacijom (množenjem logita modela sa faktorom $D$=1,702) reskaliraju u metriku karakterističnu za normalne modele (Fajgelj, 2020)[26]. Tako reskalirana logistička kriva je u oblasti srednjih vrednosti skoro podudarna sa normalnom, dok se nešto veće (ali ne velike) razlike javljaju u oblasti niskih i visokih vrednosti (Baker & Kim, 2004; Thissen & Wainer, 2001).

U odnosu na ***broj parametara*** razlikujemo jednoparametarske (1P), dvoparametarske (2P), troparametarske (3P), a u poslednje vreme i četvroparametarske (4P), pa i petoparametarske modele (5P), koji uvode parametar zakrivljenosti stavki (Bazán, Branco, & Bolfarine, 2006). Takođe, trebalo bi pomenuti i neparametarske modele (Mokken, 1971).

Prema ***formatu ocenjivanja*** stavki kojima su namenjeni, možemo razlikovati modele namenjene binarno skorovanim stavkama i one namenjene politomnim stavkama. Ova podela odnosi se na kategorijalno (u smislu diskretno) skorovane odgovore kakvi su najčešći u psihološkim mernim instrumentima. Pored ovakvih postoje i modeli za kontinuirane odgovore (npr. kada tražimo od ispitanika da procene jačinu nekog svog stava na skali od 0 do 100 ili merimo vreme reakcije), ali njih nećemo razmatrati u ovoj knjizi.

Kada je ***dimenzionalnost*** merenja u pitanju, razlikujemo jednodimenzionalne i višedimenzionalne modele. Ni višedimenzionalni modeli neće biti predmet ove knjige.

Ove podele mogu se kombinovati pa su moguće razne kombinacije.

---

[26] Pod logitom podrazumevamo izraz koji definiše kako verovatnoća odgovora zavisi od nivoa osobine ispitanika i parametara stavki. Npr. u 1PL modelu logit je $(\theta_i - b_j)$, a u dvoparametarskom $a_j(\theta_i - b_j)$. Prema tome, reskaliranje u normalnu metriku iz 1PL modela bi bilo $1{,}7(\theta_i - b_j)$, a iz 2PL modela $1{,}7a_j(\theta_i - b_j)$.

Više o ovome možete videti i u odeljku 2.3.2.

U predstavljanju najpoznatijih IRT modela krenućemo od jednostavnijih modela, namenjenih binarno skorovanim stavkama. S obzirom na to da su jednostavniji, lakše ih je razumeti, a logika na kojoj se baziraju primenjuje se i u složenijim modelima namenjenim politomno skorovanim stavkama.

## *4.1 IRT modeli za binarno skorovane stavke*

Pod binarno skorovanim stavkama podrazumevamo sve stavke koje se tako skoruju, bez obzira na format prezentacije koji nekada može biti politoman. Na primer, pitanja sa višestrukim izborom su po formatu prezentacije politomna (imaju više neuređenih kategorija / ponuđenih odgovora), ali su po formatu ocenjivanja najčešće binarno skorovana (0 za netačan odgovor, 1 za tačan).

Predstavljanje modela za binarne stavke počećemo od najsloženijeg takvog modela koji će biti prikazan u ovoj knjizi. Sledeći izraz [ 18 ] predstavlja formulu četvoroparametarskog logističkog modela za izračunavanje verovatnoće da će ispitanik $i$ sa nivoom latentne crte $\theta_i$ dati pozitivan (potvrdan, tačan, skorovan sa 1) odgovor na binarno skorovanoj stavci $j$ težine $b_j$, čija je diskriminativnost $a_j$, donja asimptota (pseudopogađanje) $c_j$, a gornja asimptota (nepažljivost) $d_j$ (Liao, Ho, Yen, & Cheng, 2012). Formula modela je:

$$P(x_{ij} = 1 | \theta, a, b, c, d) = c_j + (d_j - c_j) \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

[ 18 ]

Predviđena verovatnoća odgovora mora biti u rasponu od 0 do vrednosti gornje asimptote. Ako je donja asimptota $c_j$=0,10, a gornja $d_j$=0,95 onda deo koji predviđa razlomak u izrazu [ 18 ] ne sme biti veći od 0,95–0,10=0,85 i zato se on skalira tom vrednošću ($d_j$–$c_j$). Ako razlomački deo predviđa maksimalnu verovatnoću tačnog odgovora ispitanika (1) i to skaliramo sa $d_j$–$c_j$=0,85 (0,85) i na to dodamo i vrednost donje asimptote od 0,10, dobijamo ukupnu verovatnoću $p_{ij}$=0,95. Predviđena verovatnoća neće preći vrednost gornje asimptote KKS.

U predstavljanju modela krenuli smo od najsloženijeg jer se uz određene pretpostavke on postepeno može svesti (izjednačiti) sa jednostavnijim modelima. Jednostavniji modeli su ugnježdeni u složenijim. Ova ugnježdenost nam dalje omogućava njihovo statističko poređenje i na taj način informisanu odluku o tome koji od modela je primereniji podacima.

Na primer, ako pretpostavimo da nema nepažljivosti, odnosno da je gornja asimptota KKS jednaka 1, onda se gornji izraz [ 18 ] može svesti na jednačinu troparametarskog logističkog modela [ 19 ]. Drugim rečima, jednačina 3PL modela je poseban slučaj jednačine 4PL modela kada je parametar $d_j$=1.

Jednačina 3PL modela (Birnbaum, 1968) procenjuje verovatnoću da će ispitanik $i$ sa nivoom latentne crte $\theta_i$ dati pozitivan (potvrdan, tačan, skorovan sa 1) odgovor na binarno skorovanoj stavci $j$ težine $b_j$, čija je diskriminativnost $a_j$, a donja asimptota (pogađanje) $c_j$.

$$P(x_{ij} = 1|\theta, a, b, c) = c_j + (1 - c_j)\frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

[ 19 ]

Ako nema pogađanja (npr. radi se o nekognitivnoj stavci), parametar $c_j$ može se izjednačiti sa nulom, pa se gornja jednačina [ 19 ] može svesti na jednačinu 2PL modela za dihotomno skorovane stavke (Birnbaum, 1968):

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

[ 20 ]

Dakle, osnovna jednačina 2PL modela je poseban slučaj jednačine 3PL modela, kada je parametar $c_j$=0.

Ukoliko pretpostavimo da se diskriminativnosti stavki ne razlikuju i parametar diskriminativnosti za sve stavke postavimo na 1, kao što je to slučaj u Rašovom modelu, jednačina 2PL modela može se svesti na jednačinu 1PL modela, koja je matematički identična jednačini Rašovog modela za binarne stavke:

$$P(x_{ij} = 1|\theta, b) = \frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}}$$

[ 21 ]

U Rašovom modelu parametar $a_j$ se ne postavlja eksplicitno na 1, ali pošto ne figurira u modelu implicitno jeste jednak.

Kod 1PL modela jedini parametar koji se procenjuje je težina. Modelira se verovatnoća odgovora ispitanika na stavku u zavisnosti od nivoa osobine ispitanika i težine stavke. Nivo osobine i težina stavke se iskazuju u istim mernim jedinicama i na istoj skali. Parametri pogađanja i nepažljivosti se ne procenjuju. Parametar diskriminativnosti se ne procenjuje, ili se eventualno procenjuje, ali se podrazumeva da je jednak za sve stavke.

Varijanta jednoparametarskog logističkog modela sa procenjenim parametrom diskriminativnosti izgleda ovako:

$$P(x_{ij} = 1|\theta, a, b) = \frac{e^{a(\theta_i - b_j)}}{1 + e^{a(\theta_i - b_j)}}$$

[ 22 ]

Primetite da u izrazu ne postoji indeks stavke *j* za parametar *a* jer je on jednak za sve stavke. Od izraza za 2PL model [ 20 ] razlikuje se samo u nedostatku pomenutog indeksa.

### 4.1.1 Rašov model

Rašov model je jednoparametarski logistički model, a osnovna jednačina modela za binarne stavke

je data u izrazu [ 21 ]. Matematički je jednak 1PL IRT modelu. Model je razvijen zasebno i često se kaže da i nije IRT model već model objektivnog merenja jer dovodi u vezu latentnu osobinu i odgovor na stavku (Fajgelj, 2020).

Prvobitno, Rašov model bio je definisan u formi *dobitnog odnosa*:

$$P(x_{ij} = 1 | A_i B_j) = \frac{A_i B_j}{1 + A_i B_j}$$

[ 23 ]

U ovom slučaju $A_i$ je parametar ispitanika, a $B_i$ parametar stavke (Rasch, 1960; Reckase, 2009)

Da bi funkcija dala rezultat u intervalu od 0 do 1 karakterističnom za proporcije, proizvod parametra ispitanika i stavke nije smeo biti negativan, te su parametri definisani u rasponu od 0 do ∞. Pošto su parametri na racio nivou, ispitanik koji ima parametar $A$ jednak 0 ima nultu verovatnoću da odgovori na bilo koju stavku, odnosno, možemo govoriti o postojanju apsolutne nule. Isto tako, verovatnoća tačnog odgovora na stavku sa parametrom $B$ jednakim 0 je nulta, bez obzira na nivo osobine onog ko pokušava da odgovori.

Ispitanici koji imaju verovatnoću manju od 0,5 da tačno odgovore na stavku imaće procenu nivoa osobine manju od 1 (donja granica je 0), a oni čija je verovatnoća tačnog odgovora veća od 0,5 imaće procenu osobine >1 (do ∞). Kod testova prosečne težine distribucija mera ispitanika je izrazito pozitivno zakošena. Skala parametara modela ima apsolutnu 0 i osobine racio skale (Reckase, 2009).

Danas se model najčešće navodi u logistički transformisanom obliku datom u izrazima [ 8 ] i [ 21 ].

Procena parametra ispitanika dobija se kao prirodni logaritam dobitnog odnosa:

$$\theta_i = ln\left(\frac{p_j}{1 - p_j}\right)$$

[ 24 ]

gde je $p_j$ proporcija stavki na koje je ispitanik $i$ tačno odgovorio.

Procena težine stavke se računa na sličan način:

$$b_j = -ln\left(\frac{p_i}{1 - p_i}\right) = ln\left(\frac{1 - p_i}{p_i}\right)$$

[ 25 ]

gde je $p_i$ proporcija ispitanika koji su tačno odgovorili na stavku $j$.

Parametar težine je ujedno jedini parametar stavki u ovom modelu. Model pretpostavlja da stavke koje imaju isti predmet merenja moraju imati i (približno) jednake diskriminativnosti. Stavke koje se razlikuju po korelaciji sa predmetom merenja mogu meriti različite stvari i onda se javlja problem validnosti. Osim toga, razlike u parametrima diskriminativnosti mogu dovesti do ukrštanja KKS, što dalje stvara problem u jednoznačnom određivanju težina stavki.

Kao što je već rečeno, često se kaže da je parametar diskriminativnosti u Rašovom modelu jednak

1. To nije nigde eksplicitno rečeno, ali činjenica da logit ($\theta_i - b_j$) ne množimo ničim čini ga ekvivalentnim izrazu $a(\theta_i - b_j)$ ukoliko je $a$=1 za sve stavke.

Višeparametarski modeli uzimaju u obzir realnost da nagibi stavki često nisu identični, te da donje asimptote nisu jednake nuli i zbog toga neretko bolje opisuju podatke, odnosno, saglasniji su sa njima (de Ayala, 2022). Kao što je već rečeno, Rašov model nije napravljen da dobro opiše podatke već bi testovi trebalo da zadovolje model koji definiše uslove objektivnog merenja. Prvi uslov je da mere ispitanika ne smeju zavisiti od stavki kojima su mereni, niti parametri stavki smeju zavisiti od uzorka na kojem su stavke kalibrisane. Drugo, mere ispitanika i parametri stavki moraju biti na istoj skali, što omogućava statističku analizu. Na kraju, merenje mora zadovoljavati zahteve kombinovanog merenja, odnosno, mora biti aditivno (Shaw, 1991). Uslovi kombinovanog merenja su zadovoljeni kada je rezultirajuća varijabla aditivna funkcija druge dve varijable koje su merene bar na ordinalnom nivou merenja (Embretson & Reise, 2000; Luce & Tukey, 1964). Aditivnost se može odnositi i na monotone transformacije merenih varijabli (uključujući i logaritamsku). Rašov model zadovoljava ovaj uslov jer je odnos nivoa osobine i težine stavke aditivan. 2P i 3P modeli takođe ostvaruju aditivnost nivoa osobine i težine stavke, ali parametri diskriminativnosti i pogađanja komplikuju ovaj odnos (Embretson & Reise, 2000).

Jedna od bitnih osobina Rašovog modela je ***specifična objektivnost***. Specifična objektivnost je svojstvo IRT modela po kojem mere ispitanika ne zavise od onoga šta je mereno ni čime je mereno. Da bi model posedovao specifičnu objektivnost on mora biti u aditivnoj formi ($\theta_i - b_j$), a ne multiplikativnoj $a_j(\theta_i - b_j)$ (Ostini & Nering, 2006). To znači da samo jednoparametarski modeli, prevashodno Rašov 1PL i njegove ekstenzije, zadovoljavaju ovaj uslov. Kod svih IRT modela nivo osobine i težina stavki su dati na istoj skali pa je zadovoljen i taj uslov objektivnog merenja.

Specifična objektivnost podrazumeva ***invarijantnost poretka***, a to u suštini znači da razlika u očekivanom postignuću između dva ispitanika zavisi samo od njihovih nivoa osobine $\theta$, ne i od stavki kojima su mereni. Očekivano postignuće ispitanika $i$ na stavci $j$ zavisi od razlike njegovog nivoa osobine i težine stavke ($\theta_i - b_j$). Isto važi i za nekog ispitanika $k$ ($\theta_k - b_j$). Razliku između očekivanih postignuća ova dva ispitanika možemo izračunati na sledeći način:

$$(\theta_i - b_j) - (\theta_k - b_j) = \theta_i - b_j - \theta_k + b_j = \theta_i - \theta_k \quad [26]$$

Nakon oslobađanja od zagrada imamo dva parametra $b_j$ suprotnog predznaka. Oni se potiru i na kraju ostaje samo razlika između dva nivoa osobine. Ako bismo ovo primenili na neku drugu stavku, rezultat bi bio isti. Iz toga možemo zaključiti da razlika u postignuću dva ispitanika zavisi samo od njihovih nivoa osobine. Isto tako, jednaka razlika u nivou osobine znači jednaku razliku u postignuću bez obzira da li je to razlika između ispitanika koji ima nivo crte 1,5 logita i ispitanika koji ima nivo osobine 0,5 logita ili između druga dva ispitanika sa nivoima osobine 2 i 1 logit. Isti princip se može primeniti i na stavke.

U dvoparametarskom modelu razlike u očekivanom postignuću ne zavise samo od nivoa osobine već i od parametara diskriminativnosti stavki. Logit 2P modela je $a_j(\theta_i - b_j)$. Prema tome, razlika u očekivanom postignuću između dva ispitanika biće $a_j(\theta_i - \theta_k)$.

$$a_j(\theta_i - b_j) - a_j(\theta_k - b_j) = a_j\theta_i - a_jb_j - a_j\theta_k + a_jb_j = a_j\theta_i - a_j\theta_k = a_j(\theta_i - \theta_k)$$

[ 27 ]

Zbog toga se razlika između nivoa osobine dva ispitanika mora skalirati diskriminativnošću stavke da bismo dobili razliku u očekivanom postignuću.

Iz toga sledi i da razlika u očekivanom postignuću ova dva ispitanika neće biti ista na dve stavke koje imaju različite parametre diskriminativnosti, ali će biti proporcionalna odnosu tih parametara.

Zamislimo da dve stavke imaju isti parametar težine $b_j=b_m=1$ logit, ali različite parametre nagiba $a_j=0,7$ i $a_m=1,3$. Neka ispitanici $i$ i $k$ imaju nivoe osobine $\theta_i=1,5$ i $\theta_k=-0,5$. Razlika u očekivanom postignuću između njih će na stavci $j$ biti $0,7(1,5-0,5)=0,7*2=1,4$, a na stavci $m$ $1,3(1,5-0,5)=1,3*2=2,6$. Dakle, $1,4/2,6=0,7/1,3$.

S druge strane, jednaka razlika u nivou osobine dva ispitanika znači jednaku razliku u očekivanom postignuću na istoj stavci bez obzira da li se radi o razlici na višem ili nižem nivou osobine (Embretson & Reise, 2000). Na primer, ako je razlika između mera dva ispitanika 1 logit (npr. $\theta_i=-1$ i $\theta_k=0$), a razlika između druga dva ispitanika ista (npr. $\theta_l=1.5$ i $\theta_m=2.5$) samo na višem nivou osobine, razlika u očekivanom postignuću na nekoj stavci $j$ biće ista između oba para ($a_j*1$).

Ni poredak ispitanika u 2P modelima ne mora biti očuvan. KKS ne moraju biti i najčešće nisu paralelne, a mogu se i ukrštati. U slučaju ukrštanja može se desiti da u zavisnosti od dela kontinuuma latentne crte poredak ispitanika po postignuću na dve stavke bude čak i obrnut (Slika 5 i objašnjenje uz sliku). Naravno, ako ne dođe do ukrštanja poredak će ostati isti, ali zato intervali (razlike u postignuću) ne moraju biti jednaki.

Specifična objektivnost važi ako je model saglasan sa podacima (fituje) i može biti ostvarena u većoj ili manjoj meri. Ovo svojstvo bitno je za jednačenje testova, ispitivanje diferencijalnog funkcionisanja stavki (DIF), pravljenje banki stavki i adaptivno testiranje.

Na ovom svojstvu zasnovana je i nezavisnost procene parametara stavki od populacije. U KTT težina stavki zavisiće od nivoa osobine u populaciji. Ista stavka u populaciji ispitanika sa niskim nivoom osobine može biti teška (nizak prosečni skor na stavci), a u populaciji sa višim nivoom osobine laka (visok prosečni skor na stavci). U IRT modelima to nije tako. Parametri stavki dobijeni na populacijama sa različitim distribucijama osobine biće isti ili takvi da se mogu ujednačiti linearnom transformacijom (množenjem konstantom i/ili dodavanjem konstante – videti odeljak 8 deo o jednačenju). Ovo pre svega važi za parametre populacija. Na uzorcima nijedna transformacija neće parametre bez greške svesti na iste vrednosti. Razlog tome su standardne greške procene parametara. Jednačeni parametri uzoraka biće vrlo slični, ali ne nužno isti (DeMars, 2010). U KTT najčešće nije moguće jednačenje težina stavki obaviti linearnom transformacijom. Moguće je u slučajevima kada su stavke umerene težine, a populacije se ne razlikuju previše. Pod pretpostavkom normalne raspodele merene osobine i odsustva tačnog pogađanja, jednačenje je moguće nelinearnim transformacijama (DeMars, 2010).

U 1P modelu ukupan skor je dovoljan statistik za procenu nivoa osobine kod ispitanika. Ispitanici

sa istim skorom imaju istu procenu latentne osobine bez obzira na sklop odgovora. Rang korelacija između 1P mera osobine ispitanika i sumacionih skorova je savršena. U dvoparametarskom modelu to nije tako. Tamo će ispitanici sa *istim ponderisanim skorom* imati istu procenu latentne osobine (Fajgelj, 2020). Ponderisani skor dobija se kao suma skorova na stavkama koji su ponderisani kvadratom parametra diskriminativnosti ($a_j^2 x_{ij}$) (Embretson & Reise, 2000).

## *4.2 Modeli za politomno skorovane stavke*

S obzirom na to da je IRT nastala za potrebe pedagoškog, kognitivnog testiranja, prvi modeli bili su namenjeni binarno skorovanim kognitivnim stavkama. Uskoro su ovi modeli uz izvesna proširenja prilagođeni politomnim stavkama. Politomne modele možemo podeliti na: 1) modele namenjene stavkama sa uređenim kategorijama i one 2) namenjene stavkama sa nominalnim kategorijama.

Karakteristika ovih modela je da predviđaju verovatnoće izbora za svaku od ponuđenih kategorija. Ono što je u binarnim modelima bila KKS u ovim modelima je KKK (karakteristična kriva kategorije ili kriva kategorijskog odgovora). Kod nekih od ovih modela postoje parametri težine stavke (u celini), a kod nekih ne. Dalje, ovi modeli se mogu podeliti na direktne (verovatnoće se računaju iz jedne formule) i indirektne (računanje verovatnoća se obavlja u dve faze). Podela na direktne i indirektne koincidira sa podelom na *modele razlike* (engl. *difference models*) i *modele deljenja totalom* ili *modele deljenja ukupnim verovatnoćama* (engl. *divide-by-total*) (Thissen & Steinberg, 1986). Prvoj grupi pripada *model stepenovanog odgovaranja* (Samejima, 1969), a drugoj *model skala procene* (Andrich, 1978b, 1978a; Wright & Masters, 1982)*, model stepenovanog ocenjivanja* (Masters, 1982)*, generalizovani model stepenovanog ocenjivanja* (Muraki, 1990) i *nominalni model* (Bock, 1972).

### 4.2.1 Modeli za stavke sa uređenim kategorijama

U stavkama sa uređenim kategorijama, opcije odgovora su poređane duž kontinuuma pretpostavljene kontinuirane latentne osobine koju mere tako da koreliraju sa tim kontinuumom. U ovaj tip stavke spadaju stavke skala procene i skale Likertovog tipa. Na skale procene odgovaraju procenjivači koji procenjuju osobine neke druge osobe. Za skale procene karakteristično je da se podrazumeva ekvidistantnost ponuđenih kategorija. Likertove skale namenjene su proceni sopstvenih stavova, a kategorije na skali odgovora su obično date u formi petostepene semantičke ili numeričke skale. Ako se radi o semantičkoj skali, kategorije se obično kreću od „uopšte se ne slažem“ do „u potpunosti se slažem“. Uobičajeno je da se ove kategorije skoruju brojevima od 1 do 5 (ili od 0 do 4). Ako su kategorije odgovora date u numeričkoj formi, onda se one kreću od 1 do 5 ili od 0 do 4, pri čemu niži brojevi označavaj neslaganje, a viši slaganje sa tvrdnjom iz stabla stavke. Format stepenovanog ocenjivanja ne spada u format prezentacije stavki kao prethodna dva već u format ocenjivanja. Ovim formatom se obično ocenjuju kognitivne stavke u formi pitanja sa konstruisanim odgovorom („otvorena pitanja“), gde ispitanik može dobiti različit broj bodova u zavisnosti od kompletnosti i tačnosti odgovora.

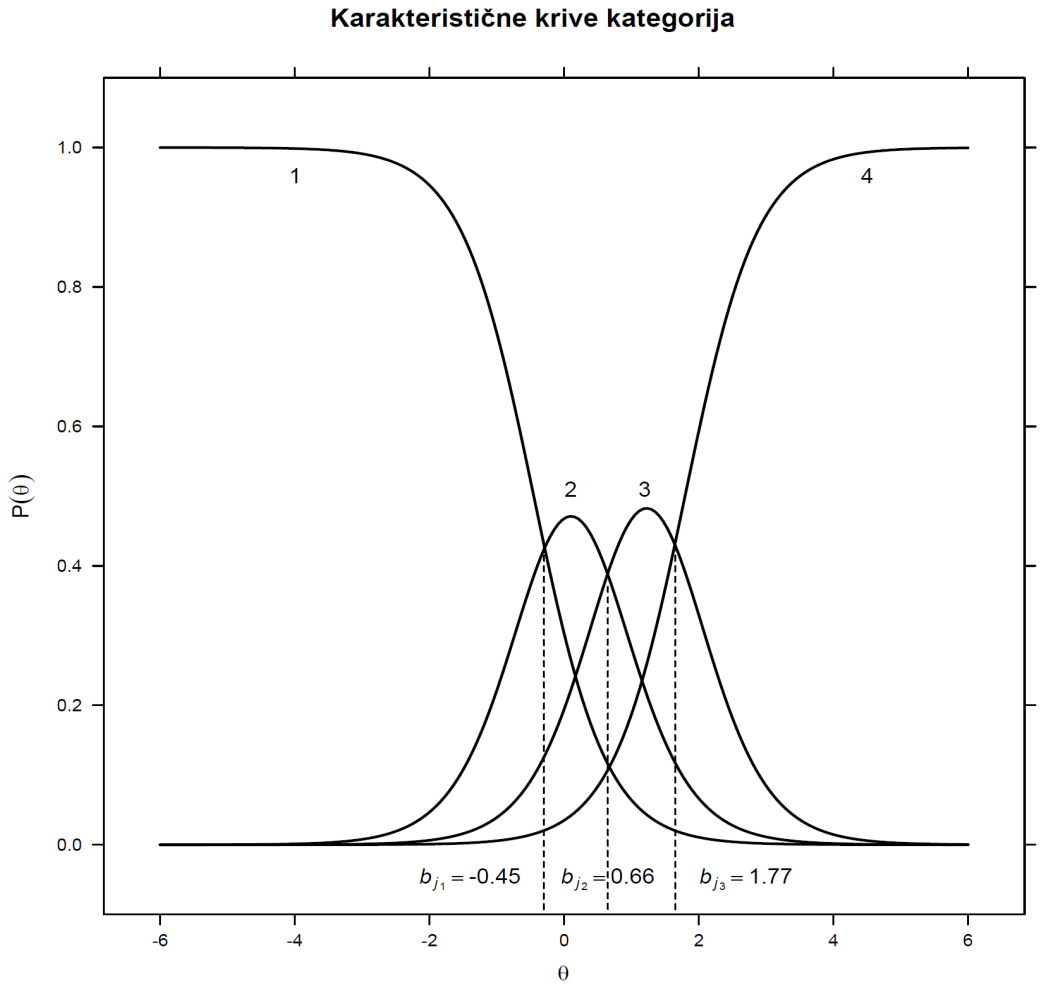
Biće prikazana tri najpoznatija modela.

#### *4.2.1.1 Model stepenovanog odgovaranja (Graded Response Model)*

Model stepenovanog odgovaranja Fumiko Samedžime (Samejima, 1969) namenjen je svim vrstama skala procene, pri čemu stavke ne moraju imati jednak broj kategorija.

Model pretpostavlja da stavka ima dve ili više kategorija odgovora. Ako stavka ima dve kategorije odgovora, GRM se svodi na dvoparametarski model za binarno skorovane stavke koji je predstavljen u izrazu [ 20 ].

Modelira se kumulativna verovatnoća da se ispitanikov odgovor nalazi u nekoj kategoriji *h*. Prema ovom modelu stavke se mogu posmatrati kao zadatak koji se sastoji iz niza koraka (koji su predstavljeni kategorijama odgovora). Da bi ispitanik ostvario neki od tih koraka, on mora da ostvari i sve prethodne (Reckase, 2009). Ovo nije slučaj u modelu stepenovanog ocenjivanja. Tamo se delovi zadatka mogu ostvariti nezavisno jedan od drugog, a svaki deo predstavlja kategoriju odgovora.




Slika 11 – Karakteristične krive kategorija politomne stavke sa 4 kategorije

U modelima za politomne stavke ne postoje karakteristične krive stavki, ali postoje karakteristične krive kategorija (Slika 11). Na grafikonu su kategorije označene brojevima. Karakteristične krive najniže i najviše kategorije su monotone. Karakteristična kriva najviše kategorije izgleda kao karakteristična kriva

binarno skorovane stavke (videti Slika 1). Verovatnoća biranja ove kategorije raste sa porastom nivoa osobine ispitanika.

Karakteristična kriva najniže kategorije je obrnuta (nalik KKS obrnuto skorovane binarne stavke). Kod nje verovatnoća biranja kategorije opada sa porastom nivoa osobine ispitanika. Krive srednjih kategorija nisu monotone. Verovatnoća biranja ovih kategorija raste sa porastom nivoa osobine, ali samo do jedne tačke kada kreće da opada. Istovremeno, raste verovatnoća biranja naredne kategorije. Lokacije tački preseka susednih kategorija na kontinuumu latentne osobine nazivaju se pragovi (engl. *thresholds*) ili granice, a pragova ima za 1 manje od broja kategorija. To su nivoi osobine na kojima ispitanik ima jednaku verovatnoću biranja jedne od susednih kategorija (na slici su označeni vertikalnim linijama)[27].

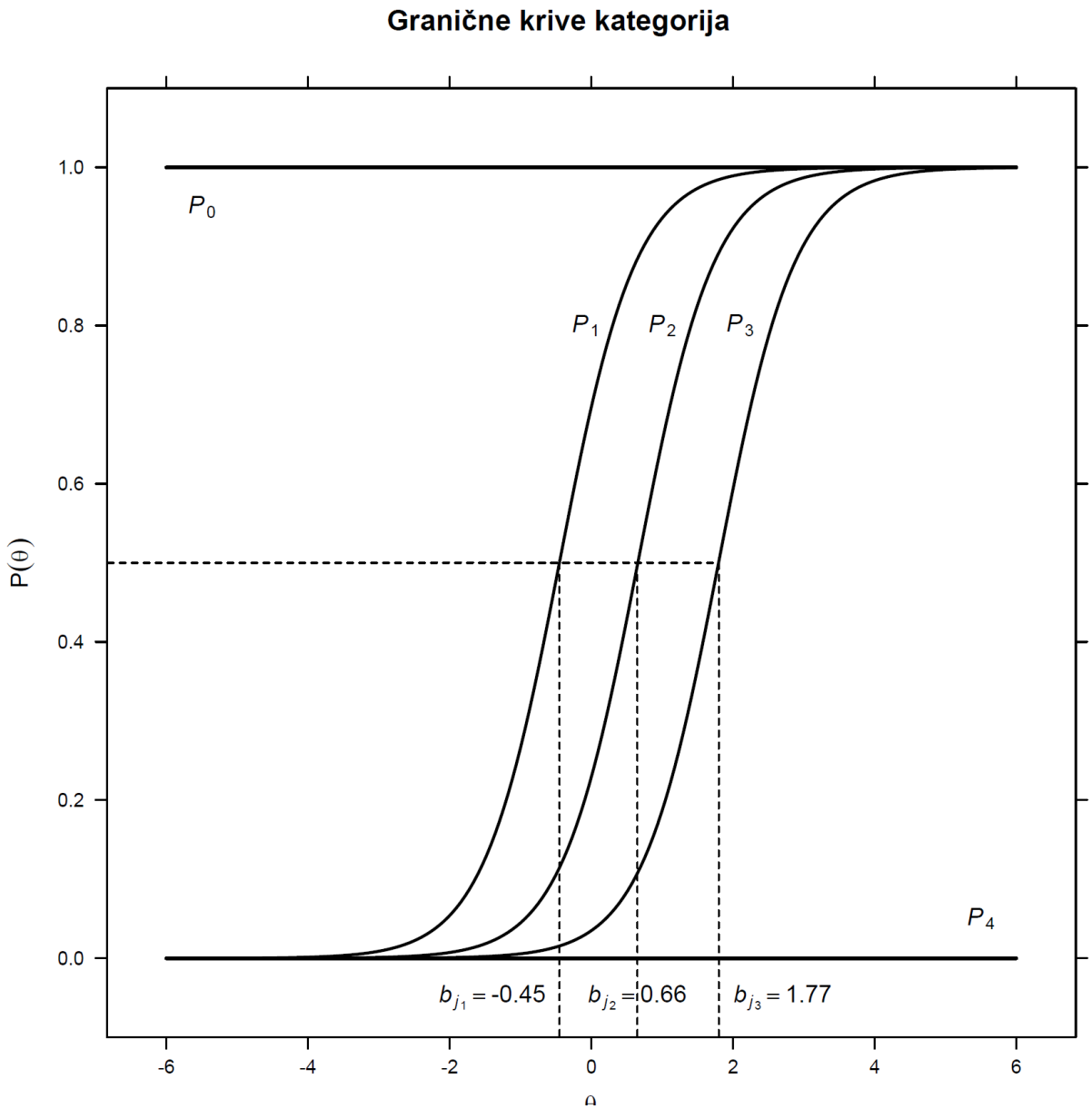
Na grafikonu sa *karakterističnim krivama kategorija* (Slika 11) pragovi su predstavljeni kao mesta gde se krive kategorija seku. Funkcije prikazane na grafikonu predstavljaju verovatnoću izbora neke od kategorija stavke u zavisnosti od nivoa osobine ispitanika.

Iako su na prikazanom grafikonu pragovi jednako udaljeni, oni to ne moraju biti. Model ne zahteva da razmak između pragova u okviru jedne stavke bude jednak, niti da su razmaci između pragova jednaki za sve stavke (kao što je to u modelu za skale procene – videti odeljak 4.2.1.2). Međutim, pragovi moraju biti sekvencijalni, odnosno, lokacija praga niže kategorije mora biti niže na kontinuumu osobine od lokacije praga više kategorije i obrnuto (ovo nije slučaj u modelu za stepenovano ocenjivanje – videti odeljak 4.2.1.3).

U *indirektnim* modelima, kao što je GRM, da bismo izračunali KKK potreban nam je jedan prethodni korak, a to su o*perativne karakteristične krive* (engl. *operating characteristic curve* – OCC) ili *granične krive kategorija* (engl. *category boundary response function* – CBRF). To su binarne funkcije kojima je opisana verovatnoća odgovaranja na granicama kategorija (DeMars, 2010). Ovih funkcija ima koliko i pragova, odnosno granica kategorija. *Granične krive kategorija* (GKK) modeliraju odgovaranje na prelazu sa jedne kategorije na drugu. To su funkcije binarnih odluka, odnosno, da li će ispitanik odabrati neku višu kategoriju (1) ili nižu (0).

Na narednoj slici (Slika 12) prikazane su granične krive kategorija za politomnu stavku sa 4 kategorije (1–4), istu onu za koju su prethodno prikazane KKK. Takva stavka ima 3 praga i isto toliko GKK. Krive kategorija su paralelne, što znači da imaju iste nagibe. GRM je dvoparametarski model i za svaku stavku procenjuje parametar diskriminativnosti. Ovaj parametar se odnosi na stavku u celini pa zato sve kategorije imaju isti nagib.

[27] Grafikon KKK bi se mogao napraviti i za binarno skorovane stavke. On bi sadržavao samo dve monotone krive: jednu opadajuću (za kategoriju 0) i jednu rastuću (za kategoriju 1). Tačka preseka te dve krive nalazila bi se na lokaciji parametra težine stavke.




Slika 12 – Granične krive kategorija politomne stavke sa 4 kategorije

Na grafikonu su označene još dve posebne „krive“: $P_0$ i $P_4$. Verovatnoća biranja odgovora višeg od 0 (0 ne postoji u skali odgovora) je 100%, jer ispitanik na ovoj stavci mora zaokružiti bar najnižu kategoriju 1. To je prikazano krivom $P_0$. Isto tako, kriva označena sa $P_4$ nam kaže da je verovatnoća biranja kategorije više od 4 nulta (jer ni kategorija 5 ne postoji). Ove dve krive su nam potrebne da bismo mogli izračunati verovatnoću biranja odgovora 1 i odgovora 4. Da bismo izračunali KKK kategorije *h* (odnosno verovatnoću biranja te kategorije) potrebno je da prvo izračunamo verovatnoću biranja te kategorije ili neke od viših. Za to se koristi sledeći izraz:

$$P_{jh}^{*}(\theta) = \frac{e^{a_j(\theta_i - b_{jh})}}{1 + e^{a_j(\theta_i - b_{jh})}}$$

[ 28 ]

gde indeks $h$ označava kategoriju odgovora na stavku $j$ i $h \in \{1, 2 \dots k\}$, $a_j$ je parametar diskriminativnosti stavke, $b_{jh}$ je parametar težine (odnosno lokacije) kategorije, a $\theta_i$ je parametar nivoa osobine ispitanika.

Možete primetiti da je izraz u suštini isti kao izraz dvoparametarskog modela za binarno skorovane stavke (izraz [ 20 ]), s tom razlikom što se umesto parametra težine stavke $b_j$, koristi parametar težine kategorije $b_{jh}$.

$P_{jh}^*$ označava verovatnoću biranja kategorije $h$ i neke od viših kategorija (dakle ne isključivo kategorije $h$). Da bismo dobili verovatnoću biranja samo kategorije $h$ potrebno je da od verovatnoće dobijene izrazom [ 28 ] oduzmemo istu tu verovatnoću za kategoriju $h$+1:

$$P(x_{ij} = h|\theta, \theta_i, a_j, b_{jh}) = \frac{e^{a_j(\theta_i - b_{jh})}}{1 + e^{a_j(\theta_i - b_{jh})}} - \frac{e^{a_j(\theta_i - b_{jh+1})}}{1 + e^{a_j(\theta_i - b_{jh+1})}} = P(\theta)^*_{jh} - P(\theta)^*_{jh+1}$$

[ 29 ]

Ovo je razlog zbog kojeg su Tisen i Stajnberg u svojoj taksonomiji IRT modela (Thissen & Steinberg, 1986) ovaj model svrstali u *modele razlike*.

Kriva $P_1$ (Slika 12) predstavlja verovatnoću biranja kategorije 2 ili neke od viših, u odnosu na biranje kategorije 1. Kriva označena sa $P_2$ predstavlja verovatnoću biranja kategorije 3 ili neke od viših u odnosu na verovatnoću biranja kategorije 2 ili neke od nižih. Ako želimo da znamo verovatnoću biranja neke pojedinačne kategorije, onda moramo znati verovatnoću *pozitivnog* odgovora na prethodnoj graničnoj funkciji i verovatnoću *pozitivnog odgovora* na narednoj graničnoj funkciji. Npr. za verovatnoću biranja odgovora 2 potrebno nam je da znamo verovatnoću pozitivnog odgovora na funkciji $P_1$ i verovatnoću pozitivnog odgovora na funkciji $P_2$. Zamislimo da ispitanik sa nivoom osobine –0,45 logita, na prvoj granici (pragu) ima verovatnoću 50% (0,50) da odabere odgovor 2 ili neki od viših, a nas zanima kolika je verovatnoća da odabere baš kategoriju 2. Da bismo to saznali potrebno je da znamo kolika je verovatnoća da na sledećoj granici odabere pozitivan odgovor, odnosno da odabere kategoriju 3 ili višu. Recimo da je to 10% (0,10). Na osnovu tih podataka možemo doći do toga da verovatnoća biranja kategorije 2 iznosi 50%–10%=40%. Dodatne krive (u primeru ove stavke to su krive $P_0$ i $P_4$) potrebne su nam za izračunavanje verovatnoća biranja krajnjih kategorija (1 i 4).

Model procenjuje parametre težine za kategorije, ali ne i za stavku u celini. O težini stavke možemo zaključivati na osnovu proseka parametara kategorija (Ali i ostali, 2015).

Podsetićemo da se parametar nagiba u ovom modelu ne može tumačiti kao diskriminativnost kategorija. Diskriminativnost kategorija će zavisiti od nagiba, ali i od udaljenosti kategorija. Što je nagib strmiji, i što su kategorije bliže, biće diskriminativnije (Embretson & Reise, 2000). Ako je parametar $a_j$ veći, GKK će biti strmije, a KKK zašiljenije. Ako je ovaj parametar niži, GKK su manje strme, a KKK platikurtičnije.

Model Eidžija Murakija (1990) predstavlja modifikaciju Samedžiminog modela za primenu na skalama procene, pa se model nekada tako i naziva (engl. *Rating Scale Graded Response Model* – RS GRM) (Ostini & Nering, 2006). U ovom modelu težina (prag) kategorije $b_{jh}$ podeljena je na težinu stavke $b_j$ i prag kategorije

$\tau_h$, pa će verovatnoća da će ispitanik odgovoriti kategorijom $h$ ili nekom višom biti:

$$P_{jh}^{*}(\theta) = \frac{e^{a_j(\theta_i-(b_j-\tau_h))}}{1+e^{a_j(\theta_i-(b_j-\tau_h))}} = \frac{e^{a_j(\theta_i-b_j+\tau_h)}}{1+e^{a_j(\theta_i-b_j+\tau_h)}}$$

[ 30 ]

Obratite pažnju da parametar $\tau_h$ nema indeks stavke, što znači da se ne procenjuje za svaku stavku već postoji set parametara kategorija (pragova) koji je zajednički za sve stavke, što je u duhu skala procene. Zbog toga Murakijeva verzija modela stepenovanog odgovaranja zahteva da sve stavke imaju jednak broj kategorija (Fajgelj, 2020). Osim toga, na ovaj način broj parametara koje model procenjuje se smanjuje. U-mesto procene seta parametara kategorija za svaku pojedinačnu stavku, ovde se procenjuje jedan set pragova koji važi za sve stavke i parametri težine i diskriminativnosti za svaku od stavki. Npr. na testu od 10 stavki tipa petostepene Likertove skale u Samedžiminom GRM modelu bilo bi potrebno proceniti 50 parametara (10 $a_j$ i 40 $b_{hj}$), a u Murakijevoj modifikaciji 24 parametra (10 $a_j$, 10 $b_j$ i 4 $\tau_h$).

#### *4.2.1.2 Model skala procene (Rating Scale Model)*

Model skala procene Andriča (Andrich, 1978b, 1978a) zasnovan je na Rašovom modelu i namenjen je politomnim stavkama tipa skale procene (sa ekvidistantnim kategorijama). Model je namenjen instrumentima koji služe za procenu stavova ili osobina ličnosti. Stavke u takvim instrumentima često se sastoje od stabla stavke u obliku tvrdnje ili pitanja i skale odgovora. Skala odgovora sačinjena je od ordinalnih kategorija koje su poređane duž kontinuuma stava. Primer ovakvih stavki su stavke Likertovog tipa. Uobičajene kategorije odgovora izražavaju stepen slaganja sa tvrdnjom od „uopšte se ne slažem“ do „u potpunosti se slažem“. Pored slaganja, kategorije odgovora mogu se odnositi na frekvenciju („nikad“–„uvek“), ocenjivanje („neprihvatljivo“–„odlično“) itd.

Prema ovom modelu, između kategorija na skali odgovora nalaze se pragovi (engl. *thresholds*). Pragovi se označavaju sa $\tau_h$, i ima ih za jedan manje od broja kategorija na skali odgovora (de Ayala, 2022). Primetite da prag ima samo indeks kategorije $h$, ali ne i indeks stavke. To znači da se u modelu procenjuje samo jedan set pragova koji je isti za sve stavke. To takođe znači da model skala procene zahteva da sve stavke imaju istu skalu odgovora kako u pogledu broja kategorija, tako i u pogledu jezičkih oznaka tih kategorija (Ostini & Nering, 2006). Za pragove se kaže da su ekvidistantni. To ne znači da je razmak između pragova prve i druge kategorije jednak razmaku pragova između druge i treće kategorije. Ekvidistantnost u ovom smislu znači da je razmak između prve i druge kategorije jednak za sve stavke, a isto važi i za razmake između ostalih kategorija. Drugim rečima, sve stavke imaju istu skalu odgovora, što je u duhu skala procene.

Ovi pragovi, iako su ekvidistantni, ne nalaze se na istim lokacijama za sve stavke. U modelu skala procene procenjuje se i parametar težine stavke $b_j$. Ovaj parametar predstavlja lokaciju *stavke* na kontinuumu osobine. Parametri pragova nisu apsolutne lokacije na kontinuumu latentne osobine već predstavljaju odstupanja pragova od lokacije stavke (de Ayala, 2022; Ostini & Nering, 2006). Lokacije pragova na kontinuumu osobine računaju se kao $b_j+\tau_h$.

Zamislimo da imamo 4 praga sa vrednostima: $\tau_1$=–1,3, $\tau_2$=–0,7, $\tau_3$=0,8, $\tau_4$=1,2 i parametar lokacije

(težine) stavke $b_j$=0,2. Za ovu stavku lokacije pragova bi bile 0,2+(−1,3)=−1,1, 0,2+(−0,7)=−0,5, 0,2+0,8=1 i 0,2+1,2=1,4 (respektivno). Za stavku koja ima parametar težine stavke $b_k$=−0,5, lokacije pragova bi bile −0,5+(−1,3)=−1,8, −0,5+(−0,7)=−1,2, −0,5+0,8=0,3 i −0,5+1,2=0,7. U oba slučaja razmaci između lokacija pragova su isti i iznose: 0,6 između 1. i 2. praga, 1,5 između 2. i 3, i 0,4 između 3. i 4. praga.

Prema modelu skala procene, verovatnoća da će ispitanik odabrati neku kategoriju zavisi od nivoa njegove osobine $\theta_i$ i lokacije konkretnog praga na kontinuumu osobine.

$$P\left(x_{ij}\middle|\theta_i, b_j, \boldsymbol{\tau}\right) = e^{\sum_{h=0}^{x_j}\left(\theta_i-(b_j+\tau_h)\right)} \Big/ \sum_{k=0}^{m} e^{\sum_{h=0}^{k}\left(\theta_i-(b_j+\tau_h)\right)} =$$

$$e^{-\sum_{h=0}^{x_j}\tau_h+x_j(\theta_i-b_j)} \Big/ \sum_{k=0}^{m} e^{-\sum_{h=0}^{k}\tau_h+k(\theta_i-b_j)}$$

[ 31 ]

Ovaj izraz određuje verovatnoću da će ispitanik $i$ sa osobinom $\theta_i$ preći $x_j$ od $h$ pragova h={0, 1...j...k} iz skupa pragova $\tau$, na stavci težine $b_j$. Ovaj model prema taksonomiji Tisena i Stajnberga (Thissen & Steinberg, 1986) spada u modele deljenja totalom, jer se verovatnoća odgovora u kategoriji $h$ izračunava tako što se šanse odgovora do kategorije $h$ (brojilac izraza [ 31 ]) dele šansama za odgovor u bilo kojoj od kategorija (imenilac).

Ako u izrazu [ 31 ] član $-\sum_{h=1}^{k}\tau_h$ označimo sa $\kappa_h$ onda ga možemo napisati kao (de Ayala, 2022):

$$P\left(x_{ij}\middle|\theta_i, b_j, h, \kappa_h\right) = \frac{e^{[x_j(\theta_i-b_j)+\kappa_{x_j}]}}{\sum_{k=0}^{m} 1 + e^{[k(\theta_i-b_j)+\kappa_h]}}$$

[ 32 ]

$\kappa_h$ se naziva *koeficijentom kategorije* i jednak je 0 kada je broj pređenih pragova 0, a u ostalim situacijama jednak je negativnoj sumi pređenih pragova:

$$\kappa_h = -\sum_{h=1}^{k}\tau_h$$

[ 33 ]

Suma svih pragova je u RSM-u ograničena da bude 0, odnosno $\kappa_k$=0 (de Ayala, 2022).

Pokušaćemo da pojasnimo koncept prelaženja pragova. Ako stavka ima 5 kategorija, imaće 4 praga. Prag 1 nalazi se između prve i druge kategorije, prag 2 između druge i treće, prag 3 između treće i četvrte, a prag 4 između četvrte i pete. Ako ispitanik ima nivo osobine niži od lokacije prvog praga, najverovatnije je da će izabrati kategoriju 1, a ako je taj nivo osobine viši od prvog praga, potrebno je proveriti kakav je taj nivo osobine u odnosu na ostale pragove. Ako je nivo osobine niži od drugog praga, najverovatnije je da će ispitanik odabrati kategoriju 2, a ako je viši trebalo bi proveriti da li je viši i od nekog narednog praga. Prema

tome, kako bi ispitanik odabrao neku kategoriju odgovora njegov nivo osobine mora preći određen broj pragova.

Da bi model bio izračunljiv mora se pretpostaviti da stavke imaju istu diskriminativnost, pa nema parametra $a_j$ (Fajgelj, 2020).

I u ovom modelu sumacioni skor je dovoljan statistik za procenu nivoa osobine ispitanika, a suma odgovora po ispitanicima je dovoljan statistik za procenu parametara težine stavki. Ispitanici koji imaju isti sumacioni skor imaće istu procenu osobine, a stavke koje imaju isti sumacioni skor po ispitanicima imaće istu procenu parametra težine. Za procenu koeficijenta kategorije $\kappa_h$ dovoljan statistik je ukupni broj ispitanika koji je na bilo kojoj stavci odgovorio u kategoriji $h$ (de Ayala, 2022).

#### *4.2.1.3 Model stepenovanog ocenjivanja (Partial Credit Model)*

Model stepenovanog ocenjivanja (Wright & Masters, 1982) namenjen je ocenjivanju kognitivnih stavki kod kojih ispitanik može dobiti bodove za rešavanje različitih delova zadatka. Primer takve stavke bi mogao biti zadatak iz matematike $(4+2)^2/9=?$ Ovakav zadatak bi mogao da se ocenjuje binarno, 0 ako ispitanik odgovori pogrešno i 1 ako ima tačno konačno rešenje. Međutim, ovaj zadatak se sastoji od više matematičkih operacija koje je potrebno ispravno obaviti da bi se došlo do tačnog konačnog rešenja. Binarno skorovanje dovodi do gubitka dela informacija o nivou merene osobine jer nam i podatak o tome da li ispitanik zna da obavi bar neku od traženih operacija govori nešto o njegovom nivou osobine (de Ayala, 2022). Zbog toga bi se mogao bodovati i tako da ispitanik dobije bod za svaku od ispravno obavljenih operacija. U tom slučaju govorimo o stepenovanom ocenjivanju.

Prva i najlakša operacija je sabiranje $(4+2=6)$. Druga je stepenovanje $(6^2=36)$. Na njoj je moguće dobiti ispravno rešenje samo ako je prva operacija ispravno obavljena. Treća operacija je deljenje $(36/9=4)$. U ovakvom načinu skorovanja pretpostavka je da nije moguće uraditi tačno složeniju operaciju, ako prethodne lakše operacije nisu urađene ispravno. Ispitanik koji uradi tačno sve tri operacije dobio bi 3 boda, a onaj koji ne uradi nijednu 0. Maksimalan broj bodova jednak je broju operacija koje bi trebalo obaviti. Broj kategorija odgovora je za jedan veći, zbog mogućnosti da ispitanik ne odradi ni jednu operaciju tačno i dobije 0. Na taj način dobijamo ordinalno politomno skorovanu stavku.

Mastersov IRT model stepenovanog ocenjivanja (Masters, 1982) modelira ovakve stavke tako što kategorije skora deli na parove koji čine ordinalne susedne kategorije ocena/skorova (de Ayala, 2022) i zatim svaki par analizira po dihotomnom modelu.

Zamislimo da su kategorije skorova paralelne sa kontinuumom latentne osobine. Ako se fokusiramo na jedan par susednih kategorija (recimo 1 i 2) na kontinuumu latentne osobine, možemo odrediti tačku takvu da će ispitanik sa nivoom osobine nižim od nje odgovoriti u kategoriji 1, a ako je ispitanikov nivo osobine iznad te tačke – u kategoriji 2. Ispravnije je reći da će ispitanik sa nivoom osobine iznad prelazne tačke verovatnije odgovoriti u kategoriji 2, a onaj sa nivoom osobine ispod prelazne tačke u kategoriji 1. Ovakve tačke nazivaju se *parametrima lokacije prelaza* (engl. *transition location parameter*) ili *koracima*, odnosno, *parametrima koraka* ili *težinama/lokacijama koraka* (engl. *step parameter*, *step difficulty*) (de

Ayala, 2022). To su ujedno tačke u kojima ispitanik ima jednaku verovatnoću da odgovori u bilo kojoj od dve susedne kategorije. Analogni su pragovima u prethodnim politomnim modelima. Ovih tačaka ima za jedan manje od broja kategorija skorova.

U našem primeru sa matematičkim zadatkom imamo 4 kategorije skora (0–4) i 3 koraka. *Lokacije koraka* procenjuju se za svaku stavku posebno pa svaki parametar lokacije koraka ima u indeksu oznaku stavke *j* i oznaku koraka *h*. U našem primeru biće označeni sa $b_{j1}$, $b_{j2}$ i $b_{j3}$.

Vratimo se na dihotomne modele za korake. Dihotomni model za tačku $b_{j1}$ modelira verovatnoću skora 1, koju možemo označiti sa $p_{j1}$, odnosno prelaz sa skora 0 na skor 1. Ujedno, komplementarna verovatnoća $1–p_{j1}$ bila bi verovatnoća odgovora 0 (označićemo je sa $p_{j0}$). Tačka $b_{j2}$ predstavlja prelaz sa skora 1 na skor 2, a tačka $b_{j3}$, sa skora 2 na skor 3. Za svaku od ovih tačaka imamo poseban dihotomni model koji modelira verovatnoću prelaza sa niže na višu (susednu) kategoriju skora. Prelazak nekog od koraka viših kategorija nije moguć ako ispitanik nije prešao sve prethodne korake. Ovaj dihotomni model može se predstaviti izrazom [ 34 ]:

$$P\left(x_{ij}=0\middle|x_{ij}=1\right)=\frac{e^{(\theta_i-b_{j1})}}{1-e^{(\theta_i-b_{j1})}}$$

[ 34 ]

Primetite da je jednačina ekvivalentna jednačini Rašovog modela za binarne stavke (izraz [ 21 ]), osim što umesto parametra lokacije stavke $b_j$ imamo parametar lokacije koraka $b_{j1}$. Rašov model za binarne stavke je poseban slučaj PCM kada stavke imaju dve kategorije skora (Wu i ostali, 2016).

Ono što ovakvi dihotomni modeli ne uzimaju u obzir je ordinalna priroda kategorija skorova. Naime, da bi neko prešao korak između skora 1 i 2, morao je prvo preći korak između skorova 0 i 1. Isto važi i za naredne kategorije skorova ukoliko ih ima. To uzima u obzir jednačina modela (de Ayala, 2022; Wright & Masters, 1982):

$$P\left(x_{ij}\middle|\theta_i,b_{jh}\right)=\frac{e^{\sum_{h=0}^{x_{ij}}(\theta_i-b_{jh})}}{e^0+\sum_{k=1}^{m_j}e^{\sum_{h=0}^{k}(\theta_i-b_{jh})}}=\frac{e^{\sum_{h=0}^{x_{ij}}(\theta_i-b_{jh})}}{\sum_{k=0}^{m_j}e^{\sum_{h=0}^{k}(\theta_i-b_{jh})}}$$

[ 35 ]

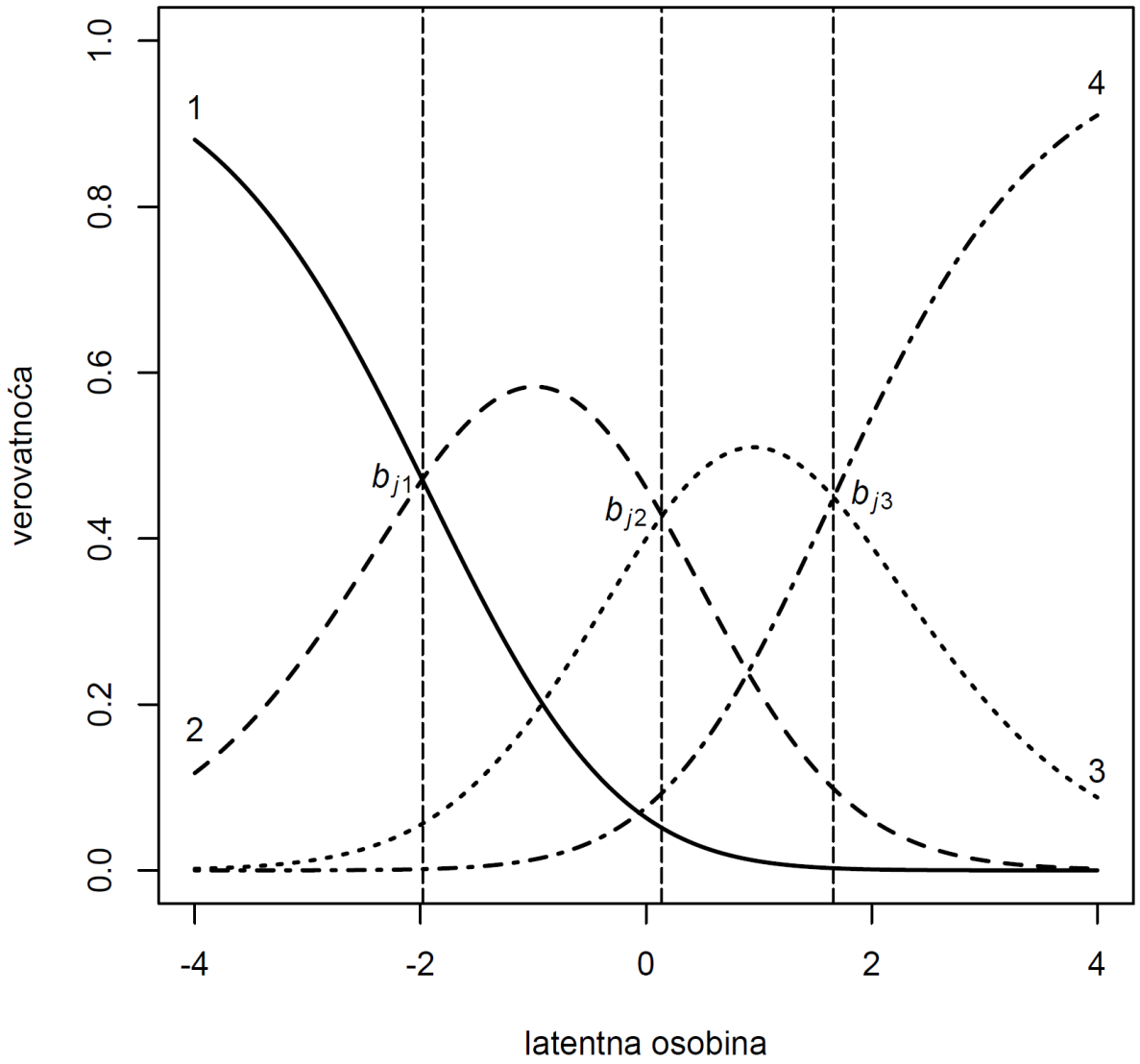
U ovom izrazu $x_{ij}$ je skor ispitanika *i* na stavci *j*, odnosno broj koraka koje je ispitanik prešao, *k* je indeks kategorije skora koji se kreće od 0 do $m_j$ (maksimalni skor na stavci *j*), *h* je indeks koraka između kategorije *k* i *k*–1, a $b_{jh}$ je lokacija *h*-tog koraka.

Ako je $x_{ij}$=2, deo izraza $e^{\sum_{h=0}^{x_{ij}}(\theta_i-b_{jh})}$ možemo napisati i kao $e^{0+(\theta_i-b_{j1})+(\theta_i-b_{j2})}$. Naime, da bi ispitanik postigao skor 2, on mora proći skor 0, prelaznu tačku (korak) za skor 1 i prelaznu tačku za skor 2. Sabiranje logita za korake ($\theta_i–b_{jh}$) moguće je jer se oni ne preklapaju, odnosno, uzajamno su isključivi (de Ayala, 2022). Ako ispitanik ne pređe prelaznu tačku za skor 1, verovatnoća da će ostvariti skor 0 je jednaka $e^0$ (odnosno 1). Da bi stigao do skora 2 ispitanik mora proći i skor 0 i otuda 0 u eksponentu.

U brojiocu izraza [ 35 ] nalaze se samo težine pređenih koraka, dok se u imeniocu nalaze svi koraci

(Wright & Masters, 1982). I opet, zbog ovoga Tisen i Stajnberg (Thissen & Steinberg, 1986) ovaj model svrstavaju u modele deljenja totalom.

Na slici (Slika 13) su prikazane karakteristične krive kategorija stavke *j* sa četiri kategorije odgovora/skora. Parametri lokacije koraka su tačke gde se seku karakteristične krive susednih kategorija. Na tim lokacijama verovatnoća odgovora u dve susedne kategorije skora biće jednaka, ali ne nužno i 50%.




Slika 13 – Karakteristične krive kategorija i koraci po PCM

Model ne zahteva da sve stavke imaju jednak broj koraka. Analizirani instrument može se sastojati od kombinacije dihotomnih i politomnih stavki sa različitim brojem kategorija skorova (de Ayala, 2022).

Tačke $b_{j1}$, $b_{j2}$ i $b_{j3}$ dele kontinuum latentne osobine na 4 dela (Slika 13). Za ispitanike sa nivoom osobine od $-\infty$ do $b_{j1}$, najverovatniji skor na ovoj stavci je 0, između $b_{j1}$ i $b_{j2}$ skor 1, između $b_{j2}$ i $b_{j3}$ skor 2, a od $b_{j3}$ do $+\infty$ skor 3. Međutim, maksimalna verovatnoća skora 2 u opsegu osobine između $b_{j2}$ i $b_{j3}$ je negde oko 0,5. To znači da su i združene verovatnoće ostalih skorova takođe oko 0,5. Drugim rečima, ako neki skor

ima verovatnoću višu od ostalih skorova on je najverovatniji *pojedinačni* skor, ali može se desiti da druge kategorije imaju veću združenu verovatnoću javljanja (Wu i ostali, 2016).

Osim toga, u modelu za stepenovano ocenjivanje lokacije koraka ne moraju biti sekvencijalne (mada sami skorovi moraju biti ordinalni). To znači da se npr. lokacija koraka za prelaz sa skora 2 na skor 3 može naći pre lokacije koraka za prelazak sa skora 1 na skor 2 (Slika 14). Razlog za poremećaj redosleda koraka može biti taj što je operacija za koju je prilikom ocenjivanja procenjeno da zahteva npr. veće znanje, u stvarnosti lakša ili jednako teška (a moguće je da postoji nepredviđeni alternativni i lakši način da se dođe do rešenja za nju). Ako se ovaj model primeni na nekognitivne stavke (recimo skalu stavova Likertovog tipa), do poremećaja redosleda može dovesti nesrazmerno biranje određenih kategorija odgovora. Na primer, kod nejasnih stavki srednju kategoriju odgovora mogu birati oni sa srednjim nivoom osobine, ali i oni koji je ne razumeju ili iz nekog razloga ne žele da odgovore (Fajgelj, 2020).

Poremećaj redosleda koraka ne znači nužno da je neka stavka loša. Kao što je rečeno, model ne zahteva sekvencijalnost (Wu i ostali, 2016). Kod ovakvih stavki vredelo bi razmisliti o promeni načina ocenjivanja i eventualnom sažimanju kategorija skale odgovora.

Podsetićemo, lokacije koraka su tačke na kontinuumu latentne osobine gde je verovatnoća da ispitanik odgovori u jednoj od dve susedne kategorije skora jednaka. Iz jednačine modela [ 35 ] vidimo da verovatnoća da ispitanik odgovori u nekoj kategoriji skora zavisi i od verovatnoća da odgovori u nekoj od preostalih kategorija, ne samo od dve susedne. Poremećaj redosleda lokacija koraka može biti posledica malog broja skorova u nekoj od kategorija (videti odeljak 10.5.4.4). Npr. u našem matematičkom zadatku može se desiti da su svi koji su znali da izvrše operaciju kvadriranja, znali i da izvrše operaciju deljenja (ili samo retki nisu znali). U tom slučaju kategorija skora 2 bi bila odsutna ili jako retka u podacima i to bi moglo dovesti do poremećaja redosleda koraka. Kada se nešto tako desi vredi razmotriti spajanje susednih kategorija (Wu i ostali, 2016). Opet, u primeru matematičkog zadatka to bi značilo da bi trebalo da sažmemo kategorije skorova 2 i 3. Tako bi spitanik koji ne uradi nijednu operaciju tačno dobio 0 bodova, onaj koji tačno izvrši sabiranje 1, a onaj koji tačno izvrši kvadriranje i deljenje 2 boda. Ova stavka bi imala 3 kategorije skora umesto 4 i 2 parametra lokacije koraka umesto 3.

Model stepenovanog ocenjivanja može imati i drugačiju parametrizaciju (de Ayala, 2022; Wu i ostali, 2016). Naime, umesto $k$ parametara težina kategorija $b_{jh}$ model može sadržati parametar težine stavke koji je jednak:

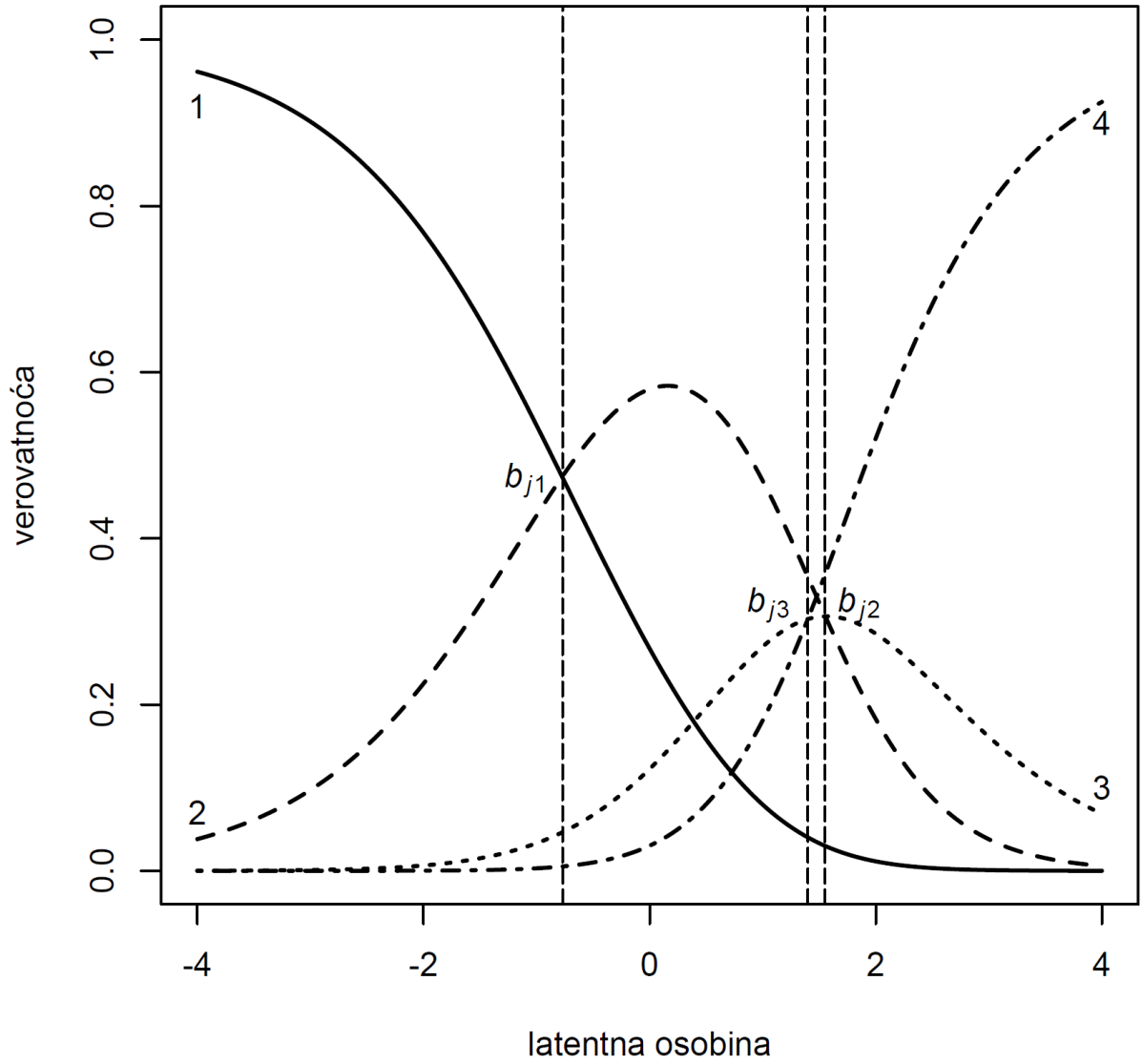
$$b_j = \frac{\sum_{h=1}^{k} b_{jh}}{k}$$

[ 36 ]

Ovaj parametar jednak je proseku parametara lokacije koraka. Pored njega procenjuje se i $k$ parametara koraka $\tau_{jk}$, koji su jednaki razlici između parametara lokacije koraka i parametra težine/lokacije stavke:

$$\tau_{jk} = b_j - b_{jh}$$

[ 37 ]




Slika 14 – Karakteristične krive kategorija sa poremećajem redosleda koraka u PCM

Pošto je $b_j$ prosek parametara $b_{jk}$, onda je suma parametara $\tau_{jk}$ jednaka 0. Tako $b_j$ predstavlja težinu stavke $j$, a parametri $\tau_{jk}$ predstavljaju odstupanja kategorija od te prosečne težine. Postojanje parametra težine stavke nam je zgodno kada opisujemo test i stavke, ali parametri koraka nam ne znače puno sami za sebe. Ako su nam potrebne lokacije koraka, njih možemo dobiti na sledeći način:

$$b_{jk} = b_j - \tau_{jk}$$

[ 38 ]

Model u ovakvoj parametrizaciji izgleda ovako:

$$P(x_{ij}|\theta_i, b_j, \tau_{jh}) = \frac{e^{\sum_{h=0}^{x_{ij}}(\theta_i - b_j - \tau_{jh})}}{\sum_{k=0}^{m_j} e^{\sum_{h=0}^{k}(\theta_i - b_j - \tau_{jh})}}$$

[ 39 ]

Pošto je suma parametara koraka jednaka 0, broj parametara modela se ovakvom parametrizacijom ne povećava, iako na prvi pogled izgleda tako. Umesto $k$ parametara lokacije koraka, ovde se procenjuje $k$–1 parametara koraka, jer poslednji mora biti takav da suma svih $\tau_{jk}$ parametara bude 0 (Wu i ostali, 2016).

I u modelu stepenovanog ocenjivanja sumacioni skor je dovoljan statistik za procenu nivoa osobine ispitanika. Ispitanici sa istim sirovim skorom imaju istu procenu $\theta$. Kada su u pitanju lokacije koraka, dovoljan statistik je broj ispitanika koji ima odgovor u određenoj kategoriji skora. Kategorije u kojima je odgovorio isti broj ispitanika imaju istu procenu parametra lokacije koraka (de Ayala, 2022). Ovo je moguće jer su nagibi svih stavki u PCM ograničeni da budu jednaki. Podsetićemo, PCM model pripada Rašovoj familiji modela i pretpostavlja se da stavke imaju jednake diskriminativnosti pa se ovaj parametar ne procenjuje.

Eidži Muraki (Muraki, 1992) je predložio model zasnovan na PCM koji uključuje parametar diskriminativnosti. Taj model se naziva *generalizovani model stepenovanog ocenjivanja* (engl. *Generalized Partial Credit Model* – GPCM) i namenjen je istom tipu stavki kao PCM. Jednačina ovog modela je:

$$P(x_{ij} = h|\theta_i, a_j, b_{jk}) = \frac{e^{\sum_{h=1}^{k_j} a_j(\theta_i - b_{jh})}}{\sum_{c=1}^{m} e^{\sum_{h=1}^{c} a_j(\theta_i - b_{jh})}}$$

[ 40 ]

Izračunava verovatnoću da će osoba $i$ sa nivoom osobine $\theta_i$ na pitanju $j$ dobiti $h$ bodova. $c$ je kategorija odgovora (skor na stavci) sa maksimalnim skorom $m$, $k$ je broj kategorija skorova.

Ovaj model je direktan i na osnovu njega može se direktno izračunati očekivani skor ispitanika $i$ na stavci $j$.

$$E_{x_{ij}} = \sum_{h=0}^{k-1} hP(x_{ij} = h|\theta_i, a_j, b_{jh})$$

[ 41 ]

gde je $P(x_{ij}=h|\theta_i,a_j,b_{jh})$ verovatnoća izbora kategorije $h$ (postizanja skora $h$) iz modela [ 40 ]. Ovaj izraz važi i za model skala procene (Fajgelj, 2020).

Ni GPCM ne postavlja zahtev da broj kategorija skorova bude jednak na svim stavkama. Isto kao u PCM, koraci ne moraju biti sekvencijalni.

S druge strane, uvođenjem parametra diskriminativnosti stavki, sirovi skor više nije dovoljan statistik za nivo osobine. Ispitanici sa istim sirovim skorom više ne moraju imati istu meru. Bitan je sklop odgovora. Ispitanici sa jednakom sumom ponderisanih skorova na stavkama imaće istu meru. Skorove na stavkama potrebno je ponderisati kvadratom parametra diskriminativnosti (Embretson & Reise, 2000). Isto tako, broj ispitanika koji je odgovorio u nekoj kategoriji skora više nije dovoljan statistik za korake.

### 4.2.2 Nominalni modeli

Nominalni modeli služe da se procene verovatnoće izbora kategorija kod politomnih stavki kod kojih kategorije nisu uređene, odnosno, ne može se napraviti njihov poredak u odnosu na kontinuum merene osobine. U ovakve stavke spadaju pitanja sa višestrukim izborom (PVI), bilo da im je format ocenjivanja nominalni (nekognitivna) ili binarni (kognitivna). Kod kognitivnih pitanja nominalni modeli se najčešće primenjuju u analizi kvaliteta distraktora. U kognitivnom testiranju PVI se najčešće skoruju binarno, pri čemu tačan odgovor nosi jedan bod a svi ostali 0, pa se za njih primenjuju binarni IRT modeli. U ovakvom načinu skorovanja gubi se deo informacija koje potencijalno mogu nositi distraktori. Naime, distraktori mogu biti kreirani tako da odslikavaju vrste zabluda, odnosno pogrešnih shvatanja, koje se često javljaju u vezi sa onim što stavka meri (de Ayala, 2022). Te zablude mogu biti povezane sa nivoom merene osobine, a nominalni model nam može omogućiti da identifikujemo kako. Uzmimo za primer pitanje iz testa znanja iz biologije: *Koji od navedenih organizama spada u klasu insekata? a) kišna glista b) vinogradarski puž c) krpelj d) pčela*. Primenom nominalnog modela možemo saznati da su neke zablude (npr. distraktori a i b) karakteristični za jako niske nivoe znanja, a na srednjem nivou osobine najčešća zabluda je ona izražena u trećem distraktoru (c) [28].

Iako je model prvobitno namenjen upravo analizi distraktora u pitanjima sa višestrukim izborom koja se skoruju dihotomno, može se primenjivati i u analizi nekognitivnih mernih instrumenata. Može se koristiti u svim situacijama kada odgovori nisu poređani paralelno sa kontinuumom latentne osobine (Embretson & Reise, 2000). To mogu biti i stavke čija je skala odgovora nominalna ili parcijalno nominalna (npr. „nikad ponekad, uvek, ne želim da odgovorim“). Kod nekognitivnih pitanja, nominalni model može nam pružiti i odgovor na pitanje da li nam je potrebna višestepena skala odgovora (recimo petostepena) ili je dovoljan manji broj kategorija odgovora (recimo 3) ili stavka možda funkcioniše kao dihotomna (Thissen & Steinberg, 1988). Ovo se odnosi i na skale sa ordinalnim i na skale sa nominalnim kategorijama. Kada se primenjuje na skale sa ordinalnim kategorijama model je sličan GPCM modelu. U stvari, i PCM i GPCM mogu se posmatrati kao posebni slučajevi nominalnog modela, a takođe i 2PL (de Ayala, 2022).

Kod nominalno skorovanih politomnih stavki formiranje ukupnog skora nema previše smisla. Nominalni model može proceniti nivo osobine ispitanika pa predstavlja mogućnost za procenu osobine na osnovu testa koji sadrži ovakva pitanja (de Ayala, 2022).

Najpoznatiji od nominalnih modela je Bokov model (Bock, 1972), poznat kao ***model za nominalne odgovore*** (engl. *Nominal Response Model* – NRM). Model je direktan i verovatnoća izbora kategorije *h* od *k* kategorija može se izračunati na sledeći način (Fajgelj, 2020).

[28] Kišna glista spada u klasu maločekinjastih crva, vinogradarski puž u klasu gastropoda, krpelj u klasu arahnida, a jedino pčela u klasu insekata. Kišna glista i vinogradarski puž po velikom broju osobina odudaraju od klase insekata i osobi sa iole većim poznavanjem biologije bi trebalo da bude jasno da ne spadaju u tu klasu. S druge strane, krpelj deli neke karakteristike sa insektima (segmentirano telo, spoljašnji skelet, zglavci) pa je potrebno veće znanje da bi se ovaj distraktor eliminisao.

$$P(x_{ij} = h|\theta_i, a_{jh}, c_{jh}) = \frac{e^{a_{jh}\theta_i + c_{jh}}}{\sum_{l=1}^{k} e^{a_{jl}\theta_i + c_{jl}}}$$

[ 42 ]

U prikazanom izrazu $h$ je indeks kategorije, $k$ je broj kategorija, $j$ indeks stavke, $a_{jh}$ je parametar nagiba kategorije, a $c_{jh}$ njen odsečak.

Prikazana jednačina modela predviđa verovatnoću izbora kategorije $h$. Brojilac sadrži verovatnoću da će ispitanik određenog nivoa osobine odgovoriti u kategoriji $h$, a imenilac sumu verovatnoća njegovih odgovora za svih $k$ kategorija. Verovatnoća da će ispitanik odgovoriti baš u kategoriji $h$ predstavlja odnos te dve veličine. I ovaj model se može svrstati u *modele deljenja totalom* (Thissen & Steinberg, 1986).

Model procenjuje nagib i donju asimptotu karakteristične krive, ali samo za kategorije. Nema parametara stavki, ali su parametri nagiba i donje asimptote kategorija specifični za svaku stavku $j$ i otuda taj indeks u oznakama $a_{jh}$ i $c_{jh}$.

Da bi model bio izračunljiv postavljeno je ograničenje da suma logita ($a_{jh}\theta_i+c_{jh}$) za jednu stavku mora biti 0 (Ostini & Nering, 2006). To automatski utiče i na parametre nagiba i donje asimptote kategorija[29] pa i suma parametara nagiba (diskriminativnosti) za sve kategorije unutar jedne stavke mora biti 0. Isto važi i za sumu donjih asimptota. Iz toga sledi da neke od kategorija mogu imati negativne diskriminativnosti. Ako to posmatramo u kontekstu distraktora, logično je da u PVI verovatnoća biranja slabijih distraktora pada sa porastom nivoa merene osobine (znanja, sposobnosti...).

Drugi način da model bude identifikovan je da se parametri nagiba i donje asimptote za prvu kategoriju svake stavke postave na 0 (de Ayala, 2022; Ostini & Nering, 2006).

Logit ovog modela $z_{jh}= a_{jh}\theta_i+c_{jh}$ se u linearnom obliku može izraziti i kao:

$$z_{jh} = \zeta_{jh} + \lambda_{jh}\theta$$

[ 43 ]

gde je $\zeta_{jh}$ odsečak a $\lambda_{jh}$ parametar nagiba.

U nominalnom modelu se parametar $\lambda$ može interpretirati kao $a$, a parametar težine $b$ se može dobiti na osnovu transformacije (Baker & Kim, 2004; Ostini & Nering, 2006):

$$b_{jh} = -\frac{\zeta_{jh}}{\lambda_{jh}}$$

[ 44 ]

odnosno, kao negativni količnik odsečka i nagiba. Ovde parametar $b$ nema isti smisao kao u diho-

[29] Parametar donje asimptote stavke smo u drugim modelima nazivali *pogađanje* ili *pseudopogađanje*, ali u ovom modelu on se naziva *lakoća* biranja kategorije (Fajgelj, 2020). U nominalnom modelu taj parametar je samo odsečak linearne jednačine (Ostini & Nering, 2006).

tomnim i politomnim modelima sa uređenim kategorijama. Podsetićemo, tamo ovaj parametar definiše graničnu funkciju između dve uzastopne kategorije, a u nominalnom modelu ne može se reći koje su kategorije susedne jer nije definisan poredak. Parametar lokacije $b$ određuje lokaciju karakteristične krive kategorije. Ove krive ne moraju biti monotone, a njihov oblik zavisi i od kombinacije parametara svih kategorija stavke (Ostini & Nering, 2006).

S druge strane, De Ajala (de Ayala, 2022) parametar $b$ u nominalnom modelu definiše na sledeći način:

$$b_{jh} = \frac{\zeta_{jh-1} - \zeta_{jh}}{\lambda_{jh} - \lambda_{jh-1}} \qquad [45]$$

odnosno kao tačku na kontinuumu $\theta$ gde se seku karakteristične krive susednih kategorija.

Oba ova načina računanja lokacije kategorije na osnovu nominalnog modela ukazuju da se na ovaj način može odrediti implicitni poredak iz na izgled neuređenih kategorija (Ostini & Nering, 2006).

Karakteristične krive kategorija u nominalnom modelu slične su krivama u modelima za politomne stavke sa uređenim kategorijama. Kriva kategorije sa najmanjom vrednošću $a$ je monotono opadajuća sa verovatnoćom 1 u $\theta$=–∞ i verovatnoćom 0 u $\theta$=∞. Kriva kategorije sa najvećim $a$ je monotono rastuća sa verovatnoćom 0 u $\theta$=–∞ i verovatnoćom 1 u $\theta$=∞, poput KKS za binarno skorovane stavke. Krive ostalih kategorija mogu biti unimodalne i asimptotski se približavati 0 (Ostini & Nering, 2006), a mogu po obliku biti slične kategorijama sa najnižim i najvišim parametrima $a$. Ako je moguće napraviti poredak kategorija po parametrima $a_{jh}$, moguće je i proceniti implicitni poredak kategorija, odnosno njihove lokacije $b_{jh}$.

## *4.3 Neparametarski modeli*

Ovde ćemo kratko predstaviti dva neparametarska IRT modela koji su osnova Mokenove analize skala (engl. *Mokken Scale Analysis* – MSA) (Mokken, 1971). Mokenova analiza skala je neparametarski probabilistički IRT model. Kao takav, ne pretpostavlja nikakav određeni oblik karakteristične krive stavke (Wind, 2017) i može se koristiti kada su pretpostavke parametarskih modela narušene i/ili uzorak nije dovoljne veličine za stabilnu procenu parametarskih modela. Takođe, može se upotrebljavati kada odluke na osnovu testiranja ne zahtevaju intervalnu skalu (dovoljna je ordinalna). Korisna je za formiranje jednodimenzionalnih skala i proveru ili obezbeđivanje invarijantnosti poretka ispitanika i stavki (više u odeljku 9.3). Posebno je korisna kada proces koji stoji iza odgovaranja nije u potpunosti jasan (afektivne varijable i postignuće u obrazovnom okruženju). Takođe, MSA je korisna kada je potrebno utvrditi u kojoj meri procene na osnovu testova zadovoljavaju fundamentalne karakteristike merenja, te koliko je opravdana upotreba ukupnih skorova (Wind, 2017).

Prvi model na kojem se zasniva MSA je *model monotone homogenosti* ili *proste homogenosti* (engl. *Monotone Homogeneity Model* – MHM*)*. Njegove pretpostavke su jednodimenzionalnost, monotonost i lokalna nezavisnost, kao kod ostalih jednodimenzionalnih modela (videti odeljak 5). Ovaj model ne dozvoljava procenu parametra ispitanika $\theta$, ali se na osnovu njega opravdava upotreba ukupnog skora za formiranje

poretka ispitanika po merenoj osobini. Naime, ako je KKS monotono neopadajuća, verovatnoća tačnog rešavanja stavke ili biranja više kategorije raste (ili bar ne opada) sa porastom nivoa osobine. U tom slučaju viši ukupan skor ukazuje na viši nivo osobine.

Da bismo ovo razumeli zamislimo obrnutu situaciju u kojoj nije zadovoljena pretpostavka monotonosti na trostepeno skorovanoj (skala 0–2) stavci *j* iz nekog testa sposobnosti. Uzorak ispitanika podelićemo na tri grupe prema ukupnom skoru izračunatom bez stavke *j* (obično se podela vrši na više grupa, ali zbog jednostavnosti mi ćemo se zadovoljiti ovim brojem grupa). Ukupan skor je indikator merene osobine. Viši ukupan skor ukazuje na višu sposobnost. Analizom smo utvrdili da ispitanici iz grupe sa najnižim ukupnim skorovima na stavci *j* u najvećem broju slučajeva ostvaruju skor 0. Ispitanici iz sledeće grupe, sa nešto višim ukupnim skorom, najčešće na stavci *j* postižu skor 1. Za sada je sve u redu. Ispitanici sa višim ukupnim skorom, postižu viši skor na stavci. Međutim, analiza je pokazala da ispitanici iz grupe sa najvišim ukupnim skorom, na stavci *j* najčešće postižu skor 0 (a očekivali bismo skor 2). U ovom slučaju, porast sposobnosti nije doveo do porasta skora na stavci, već do pada. Dakle, na stavci *j* je narušena pretpostavka monotonosti. Viši nivo osobine ne dovodi do višeg skora na stavci. Kada bismo sada skor na stavci *j* želeli da uvrstimo u ukupan skor (saberemo ga sa ukupnim skorom), da li bi tako dobijen ukupni skor bio dobar indikator merene sposobnosti? Ukupni skor bi trebalo da raste sa nivoom sposobnosti. Ovde to nije slučaj.

Drugi model je *model dvostruke monotonosti* (engl. *Double Monotonicity Model* – DMM). Ovaj model je posebni slučaj MHM i opravdava formiranje poretka stavki na osnovu aritmetičkih sredina stavki na uzorku (odnosno prosečnog skora na stavkama). Da bi to bilo moguće i imalo smisla DMM povrh pretpostavki koje ima MHM dodaje i zahtev da se karakteristične krive stavki ne smeju ukrštati. Kada se KKS ukrštaju nije moguće jednoznačno odrediti njihov poredak po težini (videti odeljak 2.3.2, Slika 5).

Mokenova analiza skala se zasniva na pretpostavkama dvostruke monotonosti. Prva pretpostavka je monotonost karakterističnih kriva stavki. KKS stavki su monotono neopadajuće. Druga pretpostavka ovog modela je da se karakteristične krive stavki ne smeju ukrštati, odnosno da im se nagibi ne smeju previše razlikovati. Ako su ove dve pretpostavke zadovoljene za neki test, onda na osnovu ukupnog skora na njemu možemo napraviti poredak ispitanika po merenoj osobini (Sijtsma & van der Ark, 2017). Ako se KKS ukrštaju, onda nisu jednake težine na različitim lokacijama kontinuuma latentne osobine.

I neparametarski i parametarski IRT modeli imaju pretpostavku monotonosti, odnosno, nagibi KKS ne smeju biti negativni, ali smeju varirati u rasponu od 0 naviše. To praktično znači da mogu imati i vrednost 0, pa se postavlja pitanje korisnosti takvih stavki. U Mokenovoj analizi koja se zasniva na dva prethodno navedena neparametarska modela izračunava se indeks skalabilnosti stavke. On ima funkciju utvrđivanja korisnosti stavke (Sijtsma & van der Ark, 2017).

Indeks skalabilnosti računa se za parove stavki:

$$H_{jk} = {\sigma_{jk}}\Big/{\sigma_{jk}^{max}}$$

[ 46 ]

gde je $\sigma_{jk}$ kovarijansa dve stavke, a $\sigma_{jk}^{max}$ maksimalna moguća varijansa dve stavke imajući u vidu

marginalne distribucije skorova dve stavke.

Indeks skalabilnosti može se izračunati za stavku:

$$H_j = {\sigma_{x_j R_j}}\Big/{\sigma_{x_j R_j}^{max}}$$

[ 47 ]

Indeks skalabilnosti stavke $H_j$ je maksimalna moguća kovarijansa stavke i *skora ostatka*. Skor ostatka je ukupni skor bez stavke *j*.

Oba indeksa mogu uzimati vrednosti iz intervala 0–1, odnosno ne mogu biti negativni.

U Mokenovoj analizi računa se i indeks skalabilnosti za test u celini:

$$H = \frac{\sum_{j=1}^{J} \sigma_{x_j R_j}}{\sum_{j=1}^{J} \sigma_{x_j R_j^{max}}}$$

[ 48 ]

Prihvatljive vrednosti *H* indeksa za skalu (ali i za stavke) su sledeće:

- $0{,}3 \leq H < 0{,}4$ slaba skala,
- $0{,}4 \leq H < 0{,}5$ srednja skala,
- $H \geq 0{,}5$ jaka skala,

a za skup stavki čiji je $H<0{,}3$ kažemo da nije skalabilan (Sijtsma & van der Ark, 2017).

## *4.4 Višedimenzionalni IRT modeli*

Pošto višedimenzionalni modeli nisu predmet ove knjige, o njima ćemo reći samo nekoliko reči.

U višedimenzionalnim IRT modelima ne važi zahtev da stavke moraju imati samo jedan predmet merenja. Postavlja se novi zahtev da svaka stavka mora meriti više dimenzija, i to onih koje su modelirane.

Ovi modeli sadrže više parametara osobine ispitanika $\theta_{id}$ i više parametara nagiba $a_{jd}$ (za svaku modeliranu dimenziju *d*).

Razlikujemo kompenzatorne i nekompenzatorne modele. U kompenzatornim modelima polazi se od pretpostavke da, ako ispitanik ima visok nivo na jednoj osobini, to mora biti kompenzovano nižim nivoom na drugoj. U modelima tog tipa to se ogleda u činjenici da je odsečak stavke ($\lambda_j$) jednak za sve modelirane dimenzije. U nekompenzatornim modelima nema ovog ograničenja, ali je verovatnoća pozitivnog odgovora (u modelima za binarno skorovane stavke) određena najnižom osobinom (Reckase, 2009).

Prikazaćemo višedimenzionalni troparametarski kompenzatorni model za binarno skorovane stavke (Fajgelj, 2020):

$$P\left(Xx_{ij}=1\middle|\boldsymbol{\theta}_{\boldsymbol{i}},\mathbf{a}_{\boldsymbol{j}},\lambda_j\right)=\mathrm{c_j}+(1-\mathrm{c_j})\frac{\mathrm{e}^{(\sum_{\mathrm{d}}\mathrm{a_{jd}}\theta_{\mathrm{id}}+\lambda_{\mathrm{j}})}}{1+\mathrm{e}^{(\sum_{\mathrm{d}}\mathrm{a_{jd}}\theta_{\mathrm{id}}+\lambda_{\mathrm{j}})}}$$

[ 49 ]

gde je $\boldsymbol{\theta}_{\mathrm{i}}$ vektor koji sadrži parametre ispitanika $\theta_{id}$ za svaku od dimenzija $d$, $\mathbf{a}_j$ je vektor koji sadrži parametre nagiba stavki $a_{jd}$ za svaku od dimenzija $d$, a $\lambda_j$ je odsečak koji je jednak za sve dimenzije u okviru stavke. Značaj stavke za ukupan skor zavisi od parametra nagiba. Što je parametar nagiba veći, stavka će više doprinositi ukupnom skoru

## *4.5 Izbor modela*

Kada se odlučujemo za IRT model koji ćemo koristiti, najčešće polazimo od ***instrumenta*** koji analiziramo. Pre svega, ako se radi o instrumentu sa binarno skorovanim stavkama, odabraćemo neki od modela koji su namenjeni takvim stavkama, a ako su stavke politomne sa uređenim kategorijama, odabraćemo neki od politomnih modela.

Neki modeli za politomne stavke ne stavljaju ograničenje kada je u pitanju broj kategorija. Npr. ako primenimo neki od tih modela na binarno skorovane stavke obično ćemo dobiti rezultate koji odgovaraju modelima za binarne stavke. Npr. model za stepenovano odgovaranje primenjen na binarno skorovane stavke ekvivalentan je dvoparametarskom modelu za dihotomne stavke. Takođe, model stepenovanog ocenjivanja primenjen na binarnim stavkama ekvivalentan je Rašovom modelu za binarne stavke.

Kada imamo kombinaciju stavki različitih formata ocenjivanja, možemo se opredeliti za neki od modela koji ne pretpostavljaju jednak broj kategorija odgovora poput GRM, PCM ili GPCM. Takođe, moguće rešenje je da instrument podelimo na grupe stavki koje imaju isti format, svaku grupu analiziramo zasebno, a na kraju dobijene rezultate spojimo (npr. izračunamo informativnost instrumenta koji kombinuje stavke različitih formata) (Embretson & Reise, 2000; Fajgelj, 2020).

Ako želimo ***merno-teorijski čist model***, logičan izbor biće Rašov model ili neka od njegovih ekstenzija (Fajgelj, 2020). Svojstva objektivnog merenja koje Rašov model zadovoljava date su u odeljku 4.1.1. Zbog toga bi modele iz Rašove familije modela posebno trebalo imati u vidu kada kreiramo novi test.

Međutim, ako konstrukt koji želimo da merimo ima više poddomena bez obzira što je suštinski jednodimenzionalan, ne možemo očekivati da će sve stavke imati jednake ***diskriminativnosti***. U tom slučaju može se desiti da primenom jednoparametarskog modela kao loše eliminišemo čitave poddomene čije stavke usled niže diskriminativnosti imaju lošije pokazatelje fita. Tada bi vredelo razmotriti upotrebu dvoparametarskog modela (Embretson & Reise, 2000; Linden, 2010).

Model možemo izabrati ***poređenjem pokazatelja fita*** više modela. Rečeno je da višeparametarski modeli obično bolje opisuju podatke, odnosno da imaju bolje pokazatelje fita (de Ayala, 2022). Međutim, to nije uvek tako. Kao mera odstupanja modela od podataka koristi se –2LL (–2 * logaritam verodostojnosti podataka). Poređenjem –2LL vrednosti dva modela možemo utvrditi da li postoji *statistički značajna razlika* u rezidualima dva modela i odabrati onaj koji bolje opisuje podatke. Razlika između –2LL vrednosti dva

modela ima hi–kvadrat distribuciju sa onoliko stepeni slobode koliko se razlikuje broj parametara dva modela. Bez obzira što višeparametarski model možda ima bolju vrednost –2LL, ukoliko razlika nije statistički značajna možemo se opredeliti za parsimoničniji model[30].

U zavisnosti od svrhe testiranja i analize može nam biti posebno bitno da procenimo ***određene parametre***. Npr. nekada to mogu biti parametri težine stavki, a nekada nam mogu bitniji biti parametri ispitanika. Ako se dvoumimo u vezi sa izborom modela, korisno je proceniti parametre koristeći više modela (npr. 1PL i 2PL – ako su to modeli između kojih se dvoumimo) i regresirati parametre od interesa dobijene jednim modelom, na iste parametre dobijene drugim. Raspršenje procena parametara oko regresione linije nam može reći kolike su i kakve razlike između dva modela, a u zavisnosti od svrhe testiranja možemo se odlučiti koji model nam više odgovara (Embretson & Reise, 2000).

Procene parametara ispitanika u Rašovom modelu su donekle otporne na narušavanje pretpostavke jednakih diskriminativnosti i odsustva pogađanja. Korelacije između mera ispitanika dobijenih Rašovim modelom na podacima koji u suštini odgovaraju troparametarskom modelu su vrlo visoke i bliske jediničnoj. Međutim, ono što se značajno razlikuje su procene standardnih grešaka, pa ako se rezultati testiranja koriste za klasifikaciju ispitanika[31] i u tom procesu intervali poverenja oko mera igraju neku ulogu, onda nije svejedno koji model odabrati (de Ayala, 2022), a procene će biti preciznije u višeparametarskim modelima.

U situaciji kada želimo da što bolje opišemo konkretne podatke, odnosno želimo da što preciznije procenimo parametre ispitanika kako bismo mogli da napravimo njihov poredak po merenoj osobini[32], izbor može pasti na neki od višeparametarskih modela. Složeniji modeli koji uzimaju u obzir činjenicu da nisu sve stavke jednako diskriminativne i da najniža verovatnoća tačnog odgovora ispitanika nije nulta na svim stavkama, obično imaju veću saglasnost sa konkretnim podacima, odnosno, bolje pokazatelje fita.

S druge strane, postoji mogućnost da prefitujemo model. Model koji savršeno opisuje konkretne

[30] Parsimoničnost se ovde odnosi na broj parametara modela. Jednoparametarski model je parsimoničniji od dvoparametarskog modela, dvoparametarski od troparametarskog....

[31] Nije retkost da u testovima koji mere određene psihološke poremećaje postoje granični (engl. *cut–off*) skorovi koji ispitanike dele na različite grupe izraženosti. Npr. za Skalu depresivnosti, anksioznosti i stresa (DASS, Lovibond & Lovibond, 1995) postoje granični skorovi za svaku od supskala koji služe za svrstavanje ispitanika u grupe sa blagim, umerenim, teškim i izrazito teškim poremećajima. Intervali poverenja oko pravih skorova daju nam raspon u kojem se nalazi pravi skor sa 95% sigurnosti. Za precizniju i sigurniju klasifikaciju bitno nam je da ti intervali poverenja budu što uži.

[32] U situaciji selekcije za posao (ili studije), gde nam je zadatak da izaberemo određeni broj najboljih kandidata (nema kritičnog skora), intervali poverenja ne igraju bitnu ulogu kao u situaciji klasifikacije. U svakom slučaju biće odabran zacrtan broj najboljih kandidata. Rekli smo da procene parametara ispitanika jednoparametarskog i višeparametarskih modela visoko koreliraju, ali se razlikuju u veličini standardne greške. S obzirom na to da nam intervali poverenja u ovoj situaciji nisu toliko bitni, mogli bismo koristiti i jednostavnije modele.

podatke teško će se generalizovati na neke druge podatke (Wright, 1998). Ako nemamo ambiciju da nalaze generalizujemo van uzorka na kojem radimo analizu i to nam može biti zadovoljavajuće.

Rašov model i njegove ekstenzije dobar su izbor ukoliko je bitna ***lakoća skorovanja***. U varijantama Rašovih modela ukupni skor je dovoljan statistik. Poredak ispitanika po Rašovim merama i po ukupnim skorovima biće jednak. Ako koristimo dvoparametarske modele, skorovi na stavkama moraju se ponderisati parametrima diskriminativnosti da bismo dobili poredak ispitanika koji odgovara njihovim IRT merama.

Kada se odlučujemo za IRT model koji ćemo primeniti možemo se rukovoditi i ***brojem parametara*** koji model procenjuje i veličinom uzorka koju smo ostvarili. Što je veći broj parametara, da bi oni bili dobro ocenjeni potreban nam je ***veći uzorak***. Ako uzorak nije dovoljne veličine, onda je možda bolja opcija odabrati dvoparametarski umesto troparametarskog modela ili jednoparametarski umesto dvoparametarskog.

Kada su u pitanju modeli za politomne stavke, treba imati u vidu da model stepenovanog ocenjivanja (PCM) procenjuje veći broj parametara od modela skala procene (RSM). Kod prvog se procenjuju pragovi kategorija za svaku od stavki, a kod drugog po jedan parametar težine za svaku stavku i jedan set parametara kategorija koji važi za sve stavke. Na primer, ukoliko imamo test sa 10 petostepenih stavki, u PCM bi bilo procenjivano 40 parametara kategorija, a u RSM 10 parametara težina stavki i samo 4 praga kategorija (14 ukupno). Razlika u broju procenjivanih parametara je veća što je broj kategorija u stavkama veći. Model stepenovanog odgovaranja (GRM) ima još više parametara od PCM. On za svaku stavku procenjuje i parametar diskriminativnosti. U prethodnom primeru na 40 parametara kategorija trebalo bi dodati još 10 parametara diskriminativnosti (o preporučenim veličinama uzorka pogledajte u odeljku 5.4).

Osim toga, politomni modeli su različito ***osetljivi na niske učestalosti*** odgovora po kategorijama. Naime, kod politomnih stavki broj ispitanika koji je na nekoj stavci odabrao određenu kategoriju nekad može biti vrlo mali. Tada će biti teško proceniti parametar te kategorije za konkretnu stavku, odnosno, procena će biti nestabilna. U tom slučaju, model skala procene koji pretpostavlja jednaku skalu odgovora i procenjuje jedan set parametara kategorija za sve stavke – ima prednost. Za procenu parametra kategorije verovatno će biti dovoljno odgovora u istoj kategoriji na drugim stavkama (Linacre, 2000).

Kada se porede model za stepenovano odgovaranje i generalizovani model za stepenovano ocenjivanje (koji su po mnogo čemu slični), GPCM se pokazuje otpornijim na veću količinu nedostajućih podataka, mali uzorak i nisku diskriminativnost stavki. Sva tri navedena činioca negativno utiču na stabilnost procene parametara, ali taj uticaj je manji kod GPCM (Dai i ostali, 2021). Tačnije, Dai i saradnici kažu da GRM u suštini daje tačnije procene parametara stavki i ispitanika kada nedostajanje podataka nije preveliko (do 20%), ali kada postane veće (30%) procene GPCM su stabilnije. S druge strane, GRM obično ima veću informativnost. Autori preporučuju da je bolje da izbor između ovih modela bude na osnovu informacionih kriterijuma (AIC, BIC) i vrednosti logLik, a ne na osnovu informativnosti testa na koju negativno utiču nedostajući podaci, niska diskriminativnost i mali uzorci.

Trebalo bi imati u vidu da biranje parsimoničnijeg modela zbog veličine uzorka može dovesti do

***pristrasne procene parametara***. Na primer, ako podatke koji su po osobinama troparametarski, pokušamo da opišemo koristeći dvoparametarski model, biće potcenjeni parametri težine i nagiba (DeMars, 2010; Yen, 1981).

# 5 Pretpostavke IRT modela

IRT modeli se nazivaju jakim modelima pošto postoje stroge pretpostavke koje moraju biti zadovoljene (Embretson & Reise, 2000). Osnovne pretpostavke IRT modela su: *monotonost KKS*, *dimenzionalnost* i *lokalna nezavisnost.*

## *5.1 Monotonost KKS*

Za karakterističnu krivu stavke (KKS) može se reći da predstavlja regresiju verovatnoće određenog odgovora na latentnu osobinu (u slučaju dihotomno skorovanih stavki) (Embretson, Reise, 2000). Monotonost podrazumeva da KKS mora imati specifičan oblik. To znači da *verovatnoća modeliranog odgovora na stavku mora rasti ili bar biti konstantna sa porastom nivoa osobine koju ona meri, ali ne sme opadati*[33]. Međutim, ova pretpostavka nije obavezna za sve IRT modele.

Monotonost KKS jeste zahtev većine IRT modela koji se koriste u praksi, ali ne svih. Postoji i velika familija takozvanih *razvijajućih* (engl. *unfolding*), odnosno, modela *idealne tačke* (engl. *ideal point*) (Liu & Chalmers, 2018; Liu & Wang, 2019; Maydeu-Olivares i ostali, 2006). Pretpostavka ovih modela je da je proces odgovaranja na stavku *razvijajućeg* tipa. Liu i Čalmers (Liu & Chalmers, 2018) prave razliku između stavki po *procesu* koji stoji iza odgovaranja na stavku i po *tipu stimulusa*. Po tipu stimulusa stavke mogu biti stavke sa direktnim odgovaranjem i stavke sa poređenjem. Kod stavki sa *direktnim odgovaranjem* ispitanik tvrdnju koja se iznosi u stablu stavke direktno ocenjuje na ponuđenoj skali odgovora. Kod stavki sa *poređenjem*, objekat koji je predmet tvrdnje poredi se sa drugim objektima i utvrđuje (recimo) preferencija. I kod jednog i kod drugog tipa stavki moguća su dva tipa *procesa* odgovaranja: kumulativni i razvijajući. Kod kumulativnog procesa odgovaranja verovatnoća biranja odgovora (tačnog, potvrdnog ili neke kategorije u politomnom ordinalnom pitanju) raste sa porastom nivoa osobine, a kada je stavka/kategorija po težini uparena (jednaka) sa nivoom osobine ispitanika ta verovatnoća iznosi 50%. Kod razvijajućeg procesa odgovaranja verovatnoća tačnog odgovora je *najveća* upravo u slučaju kada su stavka (odnosno kategorija) i nivo osobine ispitanika upareni. Na nižim i višim nivoima osobina verovatnoća biranja tog konkretnog odgovora biće niža, što znači da karakteristična kriva nije monotona. Ona sa porastom osobine raste do određene tačke, a zatim opada. Razvijajući IRT modeli namenjeni su pre svega stavkama sa direktnim odgovaranjem i razvijajućim procesom odgovaranja, odnosno, instrumentima koji mere stavove i instrumentima iz domena osobina ličnosti. Postoje modeli kako za binarno skorovane stavke, tako i za skale Likertovog tipa.

[33] Međutim, ako se modelira negativan ili netačan odgovor (što nije čest slučaj, ali je moguće), verovatnoća takvog odgovora mora padati (ili biti konstantna) sa rastom nivoa latentne osobine.

Podsetićemo, skale Likertovog tipa sastoje se od stabla kojim se izriče neka tvrdnja i ordinalne skale odgovora (recimo od „uopšte se ne slažem“ do „u potpunosti se slažem“). Ispitanik se sa izrečenom tvrdnjom može slagati u određenom stepenu, ali to može biti zato što je izrečeni stav „prejak“, ali i zato što je „preslab“. Npr. zamislimo stavku „Veronauka bi trebalo da bude izborni predmet u osnovnim školama“. Ispitanik koji je odabrao srednju kategoriju na ovom pitanju možda je to učinio zato što smatra da bi veronauka trebalo da bude obavezni predmet, ili zato što smatra da veronauka ne bi trebalo da bude uopšte zastupljena u školama. Razvijajući, odnosno modeli idealne tačke, namenjeni su rešavanju ovakvih problema i često imaju bolje pokazatelje saglasnosti (fita) od uobičajenih IRT modela (Chernyshenko i ostali, 2001). U ovom kontekstu pomenuti su samo kao primer modela koji ne moraju zadovoljavati pretpostavku monotonosti KKS.

## *5.2 (Jedno)dimenzionalnost*

I pored postojanja i razvoja ranije pomenutih multidimenzionalnih IRT modela, uobičajena praksa u psihometrijskom testiranju je da merni instrumenti imaju jedan predmet merenja. Ovo važi kako u modelima klasične testne teorije tako i u IRT modelima. Gde god imamo jedan skor ili meru kao rezultat testa, on mora meriti jednu stvar. U protivnom ispitanik do istog skora ili mere može doći na različite načine. Zamislimo test koji ima dva predmeta merenja, A i B. Recimo da 10 stavki mere konstrukt A, a drugih 10 konstrukt B. Dva ispitanika mogu dobiti isti ukupan skor tako što jedan ima visoke skorove na stavkama koje mere konstrukt A, a niske na onima koji mere konstrukt B, dok drugi ispitanik isti skor može dobiti visokim skorovima na stavkama koje mere konstrukt B, a niskim skorovima na stavkama koje mere konstrukt B. Ukoliko ova dva konstrukta nisu visoko korelirana ovakvi ukupni skorovi ne bi imali smisla i ne bismo mogli reći da su ova dva ispitanika ista o merenoj osobini (Hattie, 1984).

Prema tome i kod jednodimenzionalnih IRT modela koji su dominantno u upotrebi, jedna od jakih pretpostavki modela bila bi da stavke moraju meriti samo jednu latentnu osobinu. Međutim, potpuna jednodimenzionalnost nije moguća (de Ayala, 2022; Fajgelj, 2020). Na odgovore ispitanika utiču različiti kognitivni činioci, činioci ličnosti i same situacije testiranja. Latentna osobina i konceptualno može biti kombinacija više različitih sposobnosti i osobina ličnosti. Ako su stavke homogene kombinacije svih tih činilaca, uslov jednodimenzionalnosti i dalje može biti zadovoljen (Irwing i ostali, 2018). Danas se ova pretpostavka najčešće shvata tako da je dovoljno da postoji samo *jedna dominantna zajednička osobina* koju mere *sve stavke* jednog testa (Fajgelj, 2020; Hambleton i ostali, 1991). Na taj način, ako kao alat za utvrđivanje dimenzionalnosti testa koristimo faktorsku analizu, ovu pretpostavku će zadovoljavati testovi koji daju jednofaktorsko rešenje, ali i testovi koji imaju više koreliranih faktora, oni koji u faktorskoj analizi drugog reda takođe daju jednofaktorsko rešenje, te bifaktorska rešenja sa jakim generalnim faktorom (Irwing i ostali, 2018).

Sa razvojem višedimenzionalnih modela ova pretpostavka je modifikovana pa sad ona zahteva da sve stavke moraju imati tačno onoliko predmeta merenja koliko ima modeliranih latentnih osobina $\theta$ (de Ayala, 2022; Embretson & Reise, 2000). Ovo je sasvim logično. Ako na proces odgovaranja na stavke utiče više latentnih osobina, ukoliko neku od njih ne uvrstimo u model to će dovesti do pogrešnih predviđanja

modela jer je izostavljen činilac koji dovodi do sistematskih razlika u odgovaranju između ispitanika. Primena jednodimenzionalnih modela na podatke koji su u suštini višedimenzionalni rezultiraće pogrešnim predviđanjima. Pošto, višedimenzionalni modeli nisu predmet ove knjige, ovu temu ćemo ostaviti po strani.

## *5.3 Lokalna nezavisnost*

Lokalna nezavisnost i jednodimenzionalnost testa su slični koncepti, ali nisu sinonimi. Lokalna nezavisnost može postojati i kod višedimenzionalnih testova, ali i kod testova bez predmeta merenja (Embretson & Reise, 2000; Hattie, 1984). Pod *testom bez predmeta merenja* podrazumevamo skup stavki koje su međusobno statistički potpuno nepovezane (Hattie, 1984). Malo je verovatno da bi neko takav skup stavki mogao nazvati testom u uobičajenom značenju tog termina, ali ovde služi da ilustruje činjenicu da je lokalna nezavisnost drugačiji koncept od jednodimenzionalnosti.

Termin „lokalna“ u sintagmi lokalna nezavisnost označava nivo merene osobine. Pod tim se ne podrazumeva nužno IRT mera osobine, već se može odnositi i na sumacioni skor. Ova pretpostavka podrazumeva da korelacije skorova na stavkama za ispitanike na određenoj lokaciji merene osobine (jednaka mera ili jednaki skor) moraju biti nulte. Drugim rečima, ako kontrolišemo merenu osobinu[34], stavke ne smeju međusobno korelirati (de Ayala, 2022; Embretson & Reise, 2000; Fajgelj, 2020; Hattie, 1984).

Ovaj uslov može delovati paradoksalno jer je logično očekivati da stavke koje mere isti predmet merenja (visoko) koreliraju. Međutim, ovde je ključ u tome što se kontroliše nivo osobine. Ispitanici sa istim nivoom osobine bi trebalo da imaju istu verovatnoću odgovora na iste stavke, pa bi prema tome varijansa skorova na stavkama trebalo da bude nulta (ne računajući slučajne greške). Samim tim bi i kovarijanse i korelacije morale biti nulte jer slučajna greška ne korelira ni sa čim.

Embretson i Reise (2000) pod pretpostavkom lokalne nezavisnosti podrazumevaju dovoljnost IRT modela. Naime, sve korelacije između stavki moraju biti objašnjene parametrima (kako stavki tako i osoba) koji su definisani u primenjenom modelu. Drugim rečima, verovatnoća odgovora na stavku ne sme zavisiti od odgovora na neku drugu stavku, ako se kontrolišu parametri modela.

Kada je narušena pretpostavka lokalne nezavisnosti, procena pouzdanosti i informativnosti je veštački uvećana, a procene standardnih grešaka umanjene. Na osnovu toga možemo steći pogrešan utisak o kvalitetu instrumenta i merenja (i donositi pogrešne odluke na osnovu njega). Ako je lokalna zavisnost znatna, može postati dimenzija merenja, narušiti i pretpostavku dimenzionalnosti i onda je teško razlučiti šta zapravo test meri (Baghaei, 2008).

## *5.4 Veličina uzorka*

Veličina uzorka ne spada eksplicitno u pretpostavke IRT modela, ali smo ovu temu stavili u ovaj odeljak zato što predstavlja uslov za pravilnu procenu parametara modela. IRT modeli zahtevaju relativno velike uzorke tako da se može reći, što veći uzorak to bolje. Veći uzorak utiče na bolju procenu parametara stavki.

[34] Ili više njih u višedimenzionalnim modelima.

Veličina potrebnog uzorka zavisi od broja stavki. Veći broj stavki znači i veći broj parametara modela, što znači da duži testovi obično zahtevaju veće uzorke. Kao što je pomenuto ranije, povećanje broja stavki ne utiče jednako na povećanje broja parametara u svim modelima. Npr. u jednoparametarskim modelima za binarno skorovane stavke, ako broj stavki uvećamo za $m_d$ stavki, broj procenjivanih parametara će se uvećati za $m_d$, jer se za svaku procenjuje samo jedan parametar. Ako je u pitanju dvoparametarski model za binarno skorovane stavke, broj parametara će se uvećati za $2m_d$, a u troparametarskom, pogađate – za $3m_d$.

Stvar se usložnjava kada su u pitanju modeli za politomno skorovane stavke. U modelu stepenovanog odgovaranja (GRM), koji je dvoparametarski model za politomne stavke, ako sve stavke imaju isti broj kategorija $k$ (a ne moraju imati), broj parametara iznosi $m + m(k-1)$. U ovom modelu se procenjuje $m$ parametara nagiba i $k-1$ parametara težina kategorija (za svaku stavku). Samim tim, ako test uvećamo za $m_d$ stavki broj parametara će se uvećati za $m_d + m_d(k-1)$. Sve što je rečeno za GRM važi i za generalizovani model stepenovanog ocenjivanja (GPCM), pošto procenjuju isti broj parametara. Izvorni model stepenovanog ocenjivanja (PCM) je jednoparametarski i nema parametre nagiba za stavke. U PCM će se broj parametara uvećati za $m_d(k-1)$. U modelu skala procene (RSM) procenjuje se parametar težine za svaku stavku i jedan set od $k$ parametara kategorija koji je isti za sve stavke. U RSM će se broj parametara povećati samo za $m_d$.

Povećanje broja stavki može imati i pozitivan efekat na procenu parametara stavki. Ovaj uticaj je posredan preko poboljšanja procene $\theta$ ispitanika. Poboljšane procene parametara stavki dalje mogu dovesti do bolje procene $\theta$ (DeMars, 2010). Ovaj uticaj je veći kada se kao metode procene koriste JMLE i CMLE (videti odeljak 3). Dakle potrebna veličina uzorka zavisi i od metoda procene.

Neke opšte preporuke su sledeće. Rašov 1PL model može se relativno dobro proceniti na uzorcima 100–200 ispitanika.

Za parametar nagiba $a$, ukoliko se upotrebljava MMLE (odeljak 3.1.1.3) potrebno je bar 500 ispitanika. Ovaj uzorak je dovoljan ako su stavke umerenih težina, a merena osobina ima približno normalnu distribuciju[35] u uzorku (DeMars, 2010).

Da bi parametar pseudopogađanja ($c$) bio adekvatno procenjen, potrebni su mnogo veći uzorci. Demars (DeMars, 2010) preporučuje uzorak od bar 1000 ispitanika. Čak ni tada nije uvek moguće dobro proceniti ovaj parametar, a problemi se mogu javiti kod lakih a nediskriminativnih stavki. U tom slučaju se preporučuje da se ovaj parametar ne procenjuje za svaku stavku posebno već da se fiksira na neku plauzibilnu vrednost za sve stavke ili grupe stavki. Alternativno, može se prilikom procene zadati neki razuman opseg vrednosti parametra. Uobičajene vrednosti ovog parametra kreću se do 0,35 (Baker, 2001). Problem sa procenom parametra može se uočiti na osnovu visokih standardnih greški procene.

Kada su u pitanju politomni IRT modeli, pomenuli smo da je važan broj kategorija odgovora na stavkama (zbog broja procenjivanih parametara). Npr. za PCM sa 3 kategorije odgovora, ako se koristi

[35] Normalna distribucija $\theta$ ili parametara stavki nije pretpostavka IRT modela, ali može da olakša procenu.

MMLE može biti dovoljno i 250 ispitanika, a sa 6 kategorija taj broj može rasti i do 1000 (DeMars, 2010; He & Wheadon, 2013). Za GRM i GPCM sa 5 kategorija odgovora potrebno je oko 500 ispitanika (Reise & Yu, 1990).

Kod politomnih modela bitan je i broj ispitanika koji bira određene kategorije. Parametri kategorija biće bolje procenjeni ako je svaku kategoriju birao dovoljan broj ispitanika (dovoljno je 10-20 ispitanika prema: Wu i ostali, 2016), a najbolje je ako su kategorije odgovora ujednačeno birane. Ovaj zahtev odnosi se na svaku stavku. Izuzetak je RSM u kojem se procenjuje jedan set parametara kategorija za ceo test. U tom slučaju dobro je da kategorije budu dovoljno često i po mogućstvu ujednačeno birane na nivou testa.

Za IRT mere važi da su procene parametara nezavisne od uzorka i da za baždarenje (kalibrisanje) testa nije nužno da uzorak ispitanika bude reprezentativan. To je donekle tačno, a najviše važi za Rašov model. U njemu se obično metrika određuje tako što se prosečna težina stavki postavlja na 0, a parametar nagiba na 1. Metrika tada nije određena uzorkom, odnosno distribucijom osobine u uzorku, već zavisi od stavki. Razlika parametara dobijenih na različitim uzorcima može se tada opisati linearnom transformacijom i komponentom slučajne greške. Moramo napomenuti da mere u logitima nisu apsolutne već relativne (DeMars, 2010). Ako dve stavke iz dva različita testa koji imaju isti predmet merenja imaju i istu meru, to ne znači nužno da su jednako teške. Da bismo to mogli da tvrdimo trebalo bi da proverimo kako su utvrđene arbitrarne nule na skali, ali i jedinice merenja. Logiti u dve kalibracije ne moraju biti ni iste širine (DeMars, 2010). Arbitrarna nula na skali se najčešće postavlja kao prosečna težina stavki u testu ili prosečni nivo osobine u uzorku. U 1PL modelu jedinica merenja definisana je parametrom *a*, koji je za sve stavke jednak 1. U 2P modelu jedinica merenja definiše se tako što se varijansa parametara ispitanika fiksira na neku vrednost (1 ili neku drugu), ili se suma parametara *a* ograniči na neki broj (npr. broj stavki). Ova ograničenja su potrebna i zbog identifikacije modela (DeMars, 2010).

Kao što je rečeno, u IRT modelima vrednosti parametara stavki ne bi trebalo previše da zavise od distribucije osobine u uzorku, niti bi procene parametara ispitanika trebalo da zavise od distribucije težina stavki u testu. Međutim, standardne greške procene tih parametara će zavisiti (DeMars, 2010). Standardne greške parametara stavki biće manje (a informativnost veća) što je više ispitanika koji imaju nivo crte blizak parametru težine stavke. Isto tako, standardne greške procenjenih mera osobine biće manje što je veći broj stavki koje su po težini bliske procenjenom nivou osobine ispitanika. Drugim rečima, dobro je ako se raspon težina stavki poklapa sa rasponom nivoa osobine u uzorku. Ako je taj odnos dobar, i manji uzorci od gore preporučenih mogu biti dovoljni.

Za dobru procenu parametra nagiba takođe je potrebno da u uzorku postoji širi raspon nivoa osobine. Već smo rekli da na KKS treba gledati kao na nelinearnu regresiju. Kao i kod obične linearne regresije, ako postoji restrikcija opsega na prediktoru ili kriterijumu, teško je ispravno utvrditi nagib (DeMars, 2010).

Prilikom procene parametra pseudopogađanja potrebno je imati dovoljno ispitanika sa niskim nivoima osobine kod kojih KKS postaje paralelna sa apscisom. Ako u uzorku nedostaju ispitanici iz tog opsega osobine, stavka se može pokazati samo kao laka i da parametar *c* ne bude adekvatno procenjen. S druge strane, neadekvatna procena parametra pseudopogađanja može dovesti do pogrešne procene parametara

diskriminativnosti i težine (DeMars, 2010), što može dati i pogrešnu procenu informativnosti.

Prilikom odlučivanja o potrebnoj veličini uzorka treba se voditi i ciljem kalibracije. Ako nam je cilj da utvrdimo norme za određenu populaciju, i dalje je najbolje da imamo reprezentativan uzorak.

Za kraj, De Ajala (de Ayala, 2022) kaže da se navedenih preporuka za potrebne veličine uzorka ne moramo držati striktno. Za kalibraciju stavki obično je potrebno nekoliko stotina ispitanika, ali na to ne bi trebalo gledati kao na minimum već kao na poželjan cilj (de Ayala, 2022, str. 46).

# 6 Ocenjivanje saglasnosti modela i podataka (ocena fita)

Da bi se neki model mogao tumačiti za opisivanje stvarnosti, mora joj biti saobrazan. Provera *fitovanja* modela je upravo provera saobraznosti, odnosno, saglasnosti modela i opaženih podataka. Kao što je već rečeno, IRT je modelski pristup. Svaki od IRT modela ima predviđanja koja se mogu uporediti sa opaženim podacima i utvrditi koliko su tačna (Fajgelj, 2020). Odstupanja modelskih predviđanja od podataka nazivaju se *rezidual* ili *misfit*.

U IRT modelima možemo proveriti fit modela u celini, fit pojedinačnih stavki i fit ispitanika.

Pokazatelji fita nam ujedno govore da li su zadovoljene pretpostavke modela (Embretson & Reise, 2000). Jednoparametarski model je najstroži, jer osim pretpostavki monotonosti, jednodimenzionalnosti[36] i lokalne nezavisnosti, koje važe i za ostale modele, zahteva odsustvo pogađanja i jednake nagibe stavki (diskriminativnosti). Realni podaci koji nastaju kao rezultat primene testa na uzorku ispitanika najteže će zadovoljiti pretpostavke 1P modela. Nešto relaksiraniji je dvoparametarski model. On dopušta nejednake nagibe, ali i on zahteva odsustvo pogađanja.

Prema tome, pokazatelji misfita u jednoparametarskom modelu biće značajni ako je narušena dimenzionalnost, monotonost, lokalna nezavisnost, ali i ako postoji pogađanje ili se razlikuju nagibi stavki. Neke od tih pojava mogu imati isto poreklo. Npr. ako je kod neke od stavki narušena pretpostavka monotonosti, to može ukazivati na to da stavka meri nešto drugo, a ne ono što mere ostale stavke. Kada bi stavka merila istu stvar kao i ostale, sa rastom nivoa merene osobine, trebalo bi da raste i verovatnoća tačnog/pozitivnog odgovora na tu stavku. Ako to nije slučaj, moguće je da stavka meri (i) nešto drugo, a to je onda problem dimenzionalnosti. Ovakva stavka će imati nižu vrednost parametra diskriminativnosti, a može biti narušena i lokalna nezavisnost. Naravno, ostaje mogućnost i da je stavka jednostavno loša i ne meri ništa.

Kod dvoparametarskih modela, različiti nagibi neće biti detektovani kao misfit, ali donje asimptote različite od 0 hoće. U troparametarskom, ni donje asimptote različite od 0 neće biti identifikovane kao misfit, ali gornje asimptote različite od 1 verovatno hoće.

U slučaju postojanja značajnog misfita, a u zavisnosti od ciljeva analize, možemo se opredeliti za različite pravce delovanja. Prilikom konstrukcije mernog instrumenta možemo odlučiti da instrument prilagodimo željenom modelu. Na primer, ukoliko smo se opredelili za Rašov model a test koji smo kreirali pokazuje značajan misfit, možemo izbaciti ili korigovati stavke koje su problematične. S druge strane, ako smo koristili neki instrument sa određenim praktičnim ciljem kao što je selekcija kandidata, možemo se opredeliti da prilagodimo primenjeni model, odnosno da odaberemo onaj koji bolje opisuje podatke i ima

[36] Zato što ovde pričamo o jednodimenzionalnim modelima.

bolje pokazatelje fita i manje standardne greške (videti odeljak 4.5). Naravno, treba imati u vidu da modeli sa većim brojem parametara zahtevaju veće uzorke. Ako veličina uzorka koji imamo nije dovoljna za složeniji model, možda ćemo morati da se vratimo prilagođavanju testa.

Kada se radi o stavkama koje nisu saglasne sa modelom, njih možemo korigovati, isključiti iz testa i tumačiti ih zasebno. Između ostalog, nije nužno da isti model bude primenjen na sve stavke u testu (Embretson & Reise, 2000)

Kada se radi o ispitanicima čiji odgovori nisu saglasni sa modelom, njih takođe možemo isključiti iz analize i/ili njihove odgovore posebno tumačiti (Linacre, 2010).

## *6.1 Pokazatelji fita modela u celini*

Pokazatelji *saglasnosti (fita) modela* u celini nam govore koliko je odstupanje predviđanja zasnovanih na modelu od podataka. Takvi pokazatelji bi trebalo da imaju poznatu uzoračku distribuciju i obično se nazivaju *Goodness of Fit* (*GoF*) pokazateljima, *omnibus* ili *globalnim* pokazateljima fita. Ako je to tako, mogu se koristiti za testiranje hipoteze da li primenjeni model dobro opisuje proces generisanja podataka (Maydeu-Olivares, 2013).

Ovakvi pokazatelji postoje već duže vremena u modelima baziranim na Rašovom. Omnibus pokazatelji fita za višeparametarske modele su novijeg datuma i najčešće su zasnovani na hi–kvadrat testu. Dugo je akcenat analize saglasnosti u IRT modelima bio na proceni saglasnosti stavki, jer, kao što je već pomenuto, nije nužno da isti model bude primenjen na sve stavke (Embretson & Reise, 2000), pa se nije toliko obraćala pažnja na omnibus pokazatelje fita. S druge strane, pokazatelji fita koji su zasnovani na hi–kvadrat testu problematični su na velikim uzorcima. Naime, određeni rezidual uvek postoji, a čak i mala odstupanja na velikim uzorcima mogu biti statistički značajna i voditi odbacivanju modela (Linden, 2010). Drugim rečima, svaki (IRT) model se može odbaciti, samo je potreban dovoljno veliki uzorak (Embretson & Reise, 2000), a podsetićemo, za IRT modele potrebni su relativno veliki uzorci kako bi parametri bili procenjeni nepristrasno (videti odeljak 5.4).

I pored postojanja omnibus statistika, konačnu odluku da li je model saglasan sa podacima ili nije ne donosimo samo na osnovu jednog numeričkog pokazatelja. Konačna odluka zasniva se na proceduri koja podrazumeva ispitivanje različitih pokazatelja koji mogu biti numerički ili grafički. Ove pokazatelje integrišemo i donosimo odluku da li su odstupanja modelskih predviđanja prihvatljiva ili ne za svrhu kojoj će instrument služiti (Lamprianou, 2019).

Zato se, posebno u višeparametarskim modelima, provera fita modela često sastoji od provere zadovoljenosti pretpostavki modela, provere invarijantnosti parametara i provere modelskih predviđanja (Linden, 2010). Kako u ovoj knjizi govorimo o jednodimenzionalnim modelima, pretpostavke koje se na njih odnose opisane su u odeljku 5, a neki od načina provere biće opisani u odeljku 9. Ove provere uključuju proveru dimenzionalnosti, lokalne nezavisnosti i monotonosti.

Osim toga, kod jednoparametarskih modela mogu uključivati i proveru biserijske ajtem-total kore-

lacije, koja može pružiti uvid u jednakost parametara diskriminativnosti (videti izraz [ 10 ] za procenu parametra diskriminativnosti iz biserijske korelacije). Ako je varijabilnost ovih koeficijenata velika, to bi moglo da znači da je dvoparametarski model primereniji.

Uvid u skorove na težim stavkama kod grupe ispitanika sa najnižim ukupnim skorovima može da ukaže da li bi trebalo koristiti troparametarski model. Naime, ako ispitanici sa niskim ukupnim skorovima postižu relativno visoke skorove na teškim stavkama, to bi moglo da ukaže na prisustvo pogađanja.

Takođe, trebalo bi proveriti da li je test ubrzan, ako postoje indicije za to. Podsetićemo, kod ubrzanih testova vreme je izvor varijanse skorova[37], dok korelacije između stavki zavise i od položaja stavki u testu, a ne samo od predmeta merenja. To narušava pretpostavku lokalne nezavisnosti.

Zaključak koji bi trebalo da ponesemo odavde je da se procena fita modela ne završava proverom jednog konačnog pokazatelja fita (Douglas, 1982). Čak i kada takav pokazatelj ukazuje da model fituje, odnosno, da model ne moramo odbaciti, uvek je korisno proveriti zadovoljenost pretpostavki modela i lokalni fit pojedinačnih stavki (Maydeu-Olivares, 2013). Isto tako, ako omnibus pokazatelj fita pokaže da model nije saglasan sa podacima, uvid u fit pojedinačnih stavki reći će nam koje su razmere odstupanja i da li su one eventualno prihvatljive.

### 6.1.1 Pokazatelji zasnovani na $\chi^2$ testu

Klasični pokazatelj fita modela u celini je pokazatelj zasnovan na Pirsonovom hi–kvadrat testu (Maydeu-Olivares, 2013):

$$\chi^2 = n\sum_c (p_c - \hat{\pi}_c)^2/\hat{\pi}_c$$

[ 50 ]

gde je $n$ broj ispitanika, $p_c$ opažena proporcija sklopa $c$, a $\hat{\pi}_c$ modelski predviđena verovatnoća sklopa $c$ iz tabele kontingencije $C$. Tabela kontingencije $C$ je višedimenzionalna tabela i za test koji ima $m$ stavki sa po $k$ kategorija (kada sve stavke imaju jednak broj kategorija) sadrži $k^m$ ćelija. Svaka ćelija predstavlja potencijalni sklop odgovora[38].

Sličan statistik baziran na logaritmu verodostojnosti je $G^2$:

[37] Zadaci u ovakvim testovima su jako lagani, a ponekad i trivijalno laki. Kada bi ispitanici imali dovoljno vremena sve zadatke bi rešili tačno. Koliko će koji ispitanik zadataka tačno uraditi zavisi od toga koliko brzo radi, pa je na taj način vreme izvor varijanse skorova.

[38] Potencijalnih sklopova odgovora ima $k^m$ zato što su moguće kombinacije svih kategorija odgovora na svim stavkama. Recimo da imamo tri stavke sa mogućim odgovorima A, B i C. U tom slučaju imamo $3^3$ potencijalnih sklopova odgovora. Mogući su sledeći sklopovi: AAA, BAA, CAA, ABA, BBA, CBA, ACA, BCA, CCA, AAB, BAB, CAB, ABB, BBB, CBB, ACB, BCB, CCB, AAC, BAC, CAC, ABC, BBC, CBC, ACC, BCC i CCC.

$$G^2 = 2n \sum_c p_c ln(p_c/\hat{\pi}_c)$$

[ 51 ]

Oba navedena pokazatelja imaju hi–kvadrat distribuciju sa *df*=*C*–*q*–1, gde je *q* broj parametara modela, a C je broj ćelija gore pomenute matrice kontingencije (Maydeu-Olivares, 2013).

Problem kod ovakvih pokazatelja je to što greška tipa I ne sledi uvek teorijsku raspodelu. Poznato je da niske frekvencije u tabelama kontingencije utiču na povećanje greške tipa I, a u kontekstu pokazatelja fita to dovodi do neopravdanog odbacivanja modela. Višedimenzionalne tabele kontingencije odgovora ispitanika na stavke testa sadržaće veliki broj ćelija (sklopova) sa nultim ili niskim frekvencijama. Što ima više ovakvih sklopova, a uzorak je manji, ovaj problem biće veći. Problem manje pogađa $\chi^2$ statistik (Maydeu-Olivares & Joe, 2006). Korisno je uporediti *p* vrednosti $\chi^2$ i $G^2$ statistika. Ukoliko se ne razlikuju, onda problem ne postoji. Ukoliko se malo razlikuju, onda bi trebalo imati više poverenja u $\chi^2$, a ako je razlika velika, najverovatnije je da oba greše (Maydeu-Olivares, 2013).

Jedan od načina da se prevaziđe problem nepoznate distribucije grešaka tipa I je upotreba *samouzorkovanja* (engl. *bootstrapping*). Drugi način je sažimanje ćelija, a treći upotreba pokazatelja zasnovanih na ograničenim informacijama[39], odnosno na marginama nižeg reda u tabelama kontingencije.

Direktan način da se sažimanje ćelija obavi pre analize je upotreba margina nižeg reda. Majdu-Olivares i Džo (Maydeu-Olivares & Joe, 2006) su predstavili familiju takvih pokazatelja $M_r$ zasnovanih na rezidualima. U oznaci pokazatelja indeks *r* označava dimenzionalnost reziduala. Pokazatelj $M_1$ se zasniva na univarijatnim, $M_2$ na univarijatnim i bivarijatnim rezidualima, a $M_r$ se svodi na pokazatelj pune informacije (Maydeu-Olivares, 2013; Maydeu-Olivares & Joe, 2006). Najčešće korišćeni pokazatelj fita iz ove familije je $M_2$. U matričnoj notaciji izraz za izračunavanje $M_2$ statistika može se napisati ovako:

$$M_2 = n(\boldsymbol{p_2} - \hat{\boldsymbol{\pi}}_{\boldsymbol{2}})' \hat{\boldsymbol{C}}_2 (\boldsymbol{p_2} - \hat{\boldsymbol{\pi}}_{\boldsymbol{2}})$$

[ 52 ]

gde su $(\boldsymbol{p_2} - \hat{\boldsymbol{\pi}}_{\boldsymbol{2}})$ momenti distribucije za univarijatne i bivarijatne reziduale, a $\hat{\boldsymbol{C}}_2$ matrica modelskih verovatnoća svih opaženih sklopova odgovora. Ovaj pokazatelj sledi hi kvadrat raspodelu za:

$$df_2 = n(k-1) + \frac{m(m-1)}{2}(K-1)^2 - q$$

[ 53 ]

ukoliko je metod procene maksimalna verodostojnost[40], jer je distribucija univarijatnih i bivarijatnih reziduala tada standardizovana normalna distribucija (Maydeu-Olivares, 2013).

$M_2$ je pokazatelj bliske saglasnosti (engl. *close–fit*), a za očekivati je da u modelima sa velikim brojem

[39] Za razliku od $\chi^2$ i $G^2$ koji su zasnovani na punoj informaciji.

[40] Ne nužno ako se koriste druge metode procene parametara.

stepeni slobode on uvek bude značajan. U tom slučaju možemo koristiti i pokazatelje približnog (engl. *approximate*) fita koji se porede sa nekim graničnim vrednostima (engl. *cut-off*). Jedan do njih je $RMSEA_2$ (engl. *root mean square error of approximation*), koja se na osnovu $M_2$ statistika može izračunati na sledeći način:

$$RMSEA_2 = \sqrt{Max\frac{M_2 - df_2}{n * df_2}, 0}$$

[ 54 ]

gde je $n$ broj ispitanika. Ukoliko želimo da testiramo bliski fit, odnosno nultu hipotezu da je $RMSEA_2$ manja ili jednaka nekoj graničnoj vrednosti, to možemo učiniti na sledeći način:

$$p = 1 - Pr(\chi^2_{df_2}(n * df_2 * c_2^2) \leq M_2)$$

[ 55 ]

gde je $c_2$ granična vrednost za reziduale do bivarijatnog nivoa. Uobičajena granična vrednost je 0,05 (Maydeu-Olivares, 2013).

Cai i Hansen (2013) su predložili statistik $M_2^*$. Namenjen je politomno skorovanim stavkama u situacijama sa niskim frekvencijama odgovora po kategorijama i omogućava sažimanje univarijatnih i bivarijatnih kategorija odgovora, odnosno margina prvog i drugog reda (Chalmers, 2012). U slučaju binarno skorovanih stavki ekvivalentan je statistiku $M_2$.

Cai i Monro (2014) su takođe predložili i statistik $C_2$, namenjen politomnim ordinalnim stavkama. Naime, malopre pomenuti $M_2^*$ statistik suviše agresivno obavlja sažimanje kada su u pitanju testovi sa manjim brojem stavki, što može dovesti do pojave negativnih stepeni slobode i neizračunljivosti modela (model nije identifikovan). $C_2$ se i dalje zasniva na marginama prvog i drugog reda, ali ne sažima univarijatne kategorije, odnosno margine prvog reda koje su najčešće dobro popunjene odgovorima. Ovaj statistik predstavlja kompromis između statistika $M_2$ i $M_2^*$. Takođe, za binarno skorovane stavke i ovaj statistik je ekvivalentan statistiku $M_2$ (Cai & Monroe, 2014).

### 6.1.2 Andersenov Likelihood Ratio test

Andersenov LR (engl. *likelihood ratio*) test je omnibus pokazatelj fita namenjen Rašovoj familiji modela. Logika iza analize fitovanja Andersenovim LR testom je invarijantnost procene parametara na različitim poduzorcima. Poduzorci se najčešće formiraju po ukupnom skoru na analiziranom testu, ali je moguće koristiti i neki eksterni kriterijum. Kada se koristi eksterni kriterijum, Andersenov LR test može poslužiti i za ispitivanje diferencijalnog funkcionisanja testa – DTF (što je u suštini provera invarijantnosti merenja – odeljak 8). Ako model fituje, parametri procenjeni na poduzorku ispitanika sa nižim i na poduzorku sa višim nivoom osobine ne bi smeli da se razlikuju više od standardne greške procene parametara. U ovom pristupu uobičajeno je da se uzorak deli na ispitanike koji imaju sumacione skorove niže ili jednake i više od aritmetičke sredine ili medijane sumacionih skorova na uzorku[41]. Moguća je i podela na $J$–1 grupa u zavisnosti od

[41] Podsetićemo, u Rašovoj familiji modela ukupan skor je dovoljan statistik za mere ispitanika.

broja različitih skorova *r* (*r=1,2,...J–1*).

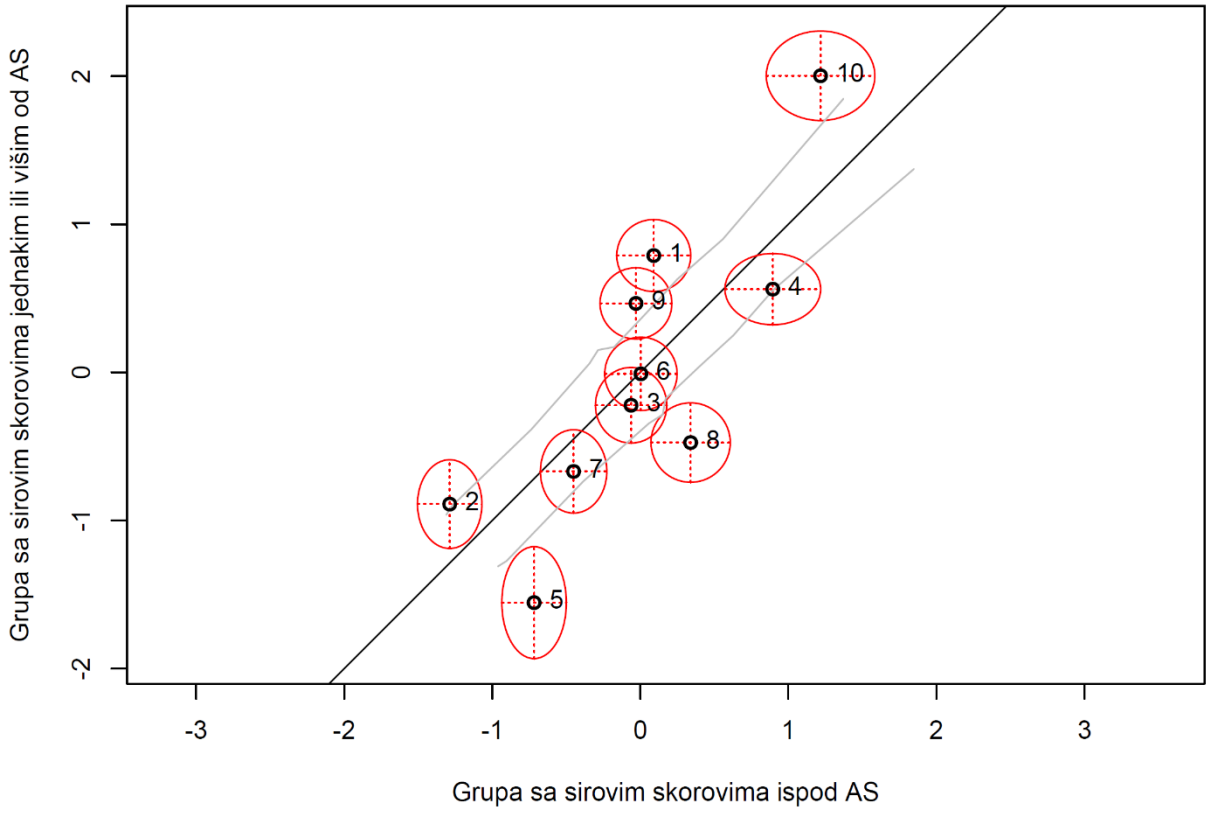
Izračunava se na sledeći način (Mair i ostali, 2008):

$$LR = Z = 2\left(\sum_{r=1}^{J} lnL_C^{(r)} - \ln L_C\right)$$

[ 56 ]

gde su $L_c^{(r)}$ funkcije verodostojnosti za poduzorke i $L_c$– ukupna funkcija verodostojnosti.

Kada model fituje, procene parametara stavki bi trebalo da budu približno iste bez obzira na visinu osobine u uzorku. Na grafičkom prikazu (Slika 15) tačkama su predstavljene procene parametara težine stavki ($b_{ij}$). Ako model fituje, idealno bi bilo da se tačke nalaze na dijagonalnoj liniji, a prihvatljivo je ako se nalaze u okviru 95% intervala poverenja oivičenog sivim linijama. Elipse oko oznaka stavki predstavljaju oblasti poverenja procena parametara. Ukupan fit modela ne mora biti narušen ako odstupanje postoji kod malog broja stavki i ako ono nije veliko. U prikazanom primeru Z statistik ukazuje da model ne fituje ($Z_{(9)}$=69,06, p=0,000).




Slika 15 – Grafička provera fita modela

Andersenov LR test ima $\chi^2$ distribuciju za (Linden, 2010):

$$df = [m(k-1)-1](r-1)$$

[ 57 ]

gde je *m* broj stavki, *k* broj kategorija, a *r* broj poduzoraka.

Problem sa ovim pokazateljem je to što različite podele uzorka mogu voditi različitim zaključcima. Takođe, kod malih uzoraka neki ukupni skorovi se ne moraju uopšte javiti u nekoj od podgrupa. Tada nije moguće proceniti neke parametre. Kada su u pitanju politomni modeli, u poduzorcima se može desiti da na pojedinim stavkama neke kategorije imaju nultu frekvenciju. Takve stavke se isključuju iz izračunavanja. Osim toga, ovaj pokazatelj može se koristiti samo kod Rašovih modela kada je metoda procene parametara CMLE (odeljak 3.1.1.2).

Anedersenov LR test je osetljiv na narušavanje dimenzionalnosti i na nejednake nagibe KKS (Mair, 2018).

## *6.2 Poređenje modela*

Pored omnibus procene fita modela koju koristimo da proverimo da li model odgovara podacima, moguće je i poređenje više ugnježdenih modela kako bismo proverili koji od njih je saglasniji sa podacima.

Statistik koji se za to koristi je količnik verodostojnosti (engl. *Likelihood Ratio* – LR) koji se nekada označava i sa $\Delta G^2$:

$$\Delta G^2 = (-2lnL_R) - (-2lnL_P) = G_R^2 - G_P^2$$

[ 58 ]

gde je $L_R$ maksimum funkcije verodostojnosti redukovanog modela, a $L_P$ maksimum funkcije verodostojnosti punog modela.

$\Delta G^2$ statistik ima $\chi 2$ distribuciju za broj stepeni slobode koji je jednak razlici između broja parametara dva modela (de Ayala, 2022).

$\Delta G^2$ se koristi za upoređivanje hijerarhijskih, ugnježdenih modela. Npr. 1PL i 2PL modeli se mogu smatrati ugnježdenim modelima jer je 1PL poseban slučaj 2PL modela kod kojeg su svi parametri nagiba ograničeni da budu jednaki (a nekada fiksirani na vrednost 1). Isto tako, 2PL model se može smatrati posebnim slučajem 3PL modela kod kojeg su svi parametri donje asimptote ograničeni na vrednost 0. I 1PL model je poseban slučaj 3PL modela kod kojeg su ograničeni i parametri nagiba i donje asimptote, na gore navedene načine.

Ako je $\Delta G^2$ značajan, dva modela se značajno razlikuju i onda bismo trebali da se opredelimo za onaj koji ima višu vrednost $-2lnL$. Ako razlika nije značajna, onda procena dodatnih parametara ne doprinosi dovoljno tačnosti modela i trebalo bi se opredeliti za jednostavniji. Tim pre što će za datu veličinu uzorka bolje biti procenjen model sa manje parametara.

Ako je $\Delta G^2$ značajan, može nas zanimati koliko je bolji model saglasniji sa podacima. Ovaj statistik se takođe izračunava na osnovu verodostojnosti punog i redukovanog modela, a označava se sa $R_\Delta^2$. Radi se o proporciji promene $R^2$ uvođenjem dodatnih parametara u model (de Ayala, 2022):

$$R_{\Delta}^2 = \frac{G_R^2 - G_P^2}{G_R^2}$$

[ 59 ]

Rezultat je u formi proporcije, a ako ga pomnožimo sa 100 dobićemo procentualno izraženu promenu. Na nama je da odlučimo da li procenat poboljšanja opravdava uvođenje dodatnih parametara.

Pokazatelji saglasnosti na nivou modela su i informacioni kriterijumi AIC, BIC, AICc i saBIC. Normalno je da modeli sa više parametara pokazuju bolje pokazatelje saglasnosti. To je ekvivalentno višestrukoj linearnoj regresiji gde modeli sa više prediktora obično bolje objašnjavaju kriterijum.

Stoga bi prilikom procene saglasnosti modela trebalo u obzir uzeti i broj parametara modela. Za to nam služe pokazatelji koji se nazivaju *informacionim kriterijumima*. Najpoznatiji su AIC (*Akaike Information Criterion*) (Akaike, 1974) i BIC (*Bayesian Information Criterion*) (Schwarz, 1978).

Ova dva pokazatelja predstavljaju modifikaciju pokazatelja $-2lnL$. AIC ga koriguje (penalizuje) za broj procenjivanih parametara modela (de Ayala, 2022):

$$AIC = -2lnL + 2q$$

[ 60 ]

gde $q$ označava broj parametara modela. Od dva modela bolji je onaj koji ima nižu vrednost AIC, pa prema tome, ako dva modela imaju istu vrednost $-2lnL$, onda će kao bolji biti procenjen onaj koji procenjuje manji broj parametara..

Korigovana verzija AICc (Sugiura, 1978) je pokušaj rešavanja problema koji postoji sa običnim AIC kriterijumom na malim uzorcima. Naime, kada je odnos broja parametara i broja slučajeva ($q/n$) preveliki, AIC favorizuje modele sa većim brojem parametara. U formulu AIC dodaje se faktor za korekciju pristrasnosti:

$$AICc = AIC + \frac{2q(q+1)}{n-q-1}$$

[ 61 ]

BIC takođe predstavlja modifikaciju AIC pokazatelja penalizujući broj ispitanika, ali i broj parametara (de Ayala, 2022):

$$BIC = -2lnL + \ln(n)2q$$

[ 62 ]

gde je $ln(n)$ prirodni logaritam broja ispitanika.

saBIC ili *SSBIC* (Sclove, 1987) je za BIC ono što je AICc za AIC. Naime i BIC pati od istog problema kao AIC kada je odnos $q/n$ preveliki. Formula saBIC kriterijuma namenjenog malim uzorcima je:

$$saBIC = -2lnL + q\left[\ln\left(\frac{n+2}{24}\right)\right]$$

[ 63 ]

Još jedan informacioni kriterijum koji se ređe pominje, ali se pojavljuje u paketu *mirt*, je Hanan-Kvin informacioni kriterijum (engl. *Hannan-Quinn information criterion* – HQ):

$$HQ = -2lnL + 2q\ln\big(\ln(n)\big)$$

[ 64 ]

HQ predstavlja pokušaj da informacioni kriterijum bude dosledniji. Ovde se deo koji se odnosi na penalizaciju (*2q ln(ln)n*)) menja sporije kao funkcija veličine uzorka (Miche & Lendasse, 2009).

Kod svih informacionih kriterijuma, preferirani model je onaj sa nižom vrednošću. Međutim, ne postoji definisana granična vrednost razlike koja je dovoljna da neki od modela proglasimo boljim. Uobičajeno je da se kao dovoljna, ali slaba, uzima i razlika manja od 2. Razlika u rasponu 2–6 je umeren dokaz, a ako je veća od 6, onda se smatra jakim dokazom u korist modela sa nižom vrednošću informacionog kriterijuma (Raftery, 1995).

## *6.3 Provera fita stavki*

Bez obzira da li omnibus pokazatelj saglasnosti modela govori u prilog ili protiv odbacivanja modela, potrebno je analizu nastaviti analizirajući fit pojedinačnih stavki. Ukoliko je model dobar (fituje) to ne znači da je svaka pojedinačna stavka dobra, odnosno da je odabrani model dobro opisuje. Ukoliko model nije saglasan sa podacima, analize saglasnosti pojedinačnih stavki nam mogu dati odgovor zašto model nije dobar (Maydeu-Olivares, 2013).

Postoje tri šire grupe pokazatelja fita za stavke. Pokazatelji fita bazirani na hi–kvadrat testu između modelski predviđenih i opaženih odgovora, testovi zasnovani na standardizovanim rezidualima (porede se opaženi i predviđeni odgovori ispitanika na stavke) i eksploratorni i neparametarski testovi narušavanja različitih pretpostavki (Wu i ostali, 2016).

### 6.3.1 Pokazatelji zasnovani na hi–kvadrat testu

Pokazatelji zasnovani na hi–kvadrat testu utvrđuju razliku između modelski predviđenih skorova na stavkama i onih koji su opaženi. Ne izračunavaju se na svakom pojedinačnom rezultatu već je uobičajeno da se raspon osobine podeli na određeni broj intervala. Potrebno je ispitanike poređati po nivou osobine i podeliti ih na određen broj grupa sa približno jednakim brojem ispitanika. Ove grupe se nazivaju $\theta$ intervali. Za svaki $\theta$ interval se zatim izračunaju opažene i očekivane proporcije tačnih/potvrdnih odgovora. Opažene proporcije dobijamo na osnovu podataka, a očekivane proporcije računaju se na osnovu formule modela. Vrednost $\theta$ koja se uvrštava u formulu modela može biti medijana ili aritmetička sredina $\theta$ intervala. Na osnovu tako dobijenih vrednosti računa se Pirsonov hi–kvadrat za svaku stavku (Irwing i ostali, 2018):

$$\chi_j^2 = \sum_{v=1}^{u}\sum_{x=0}^{1} n_v \frac{\big(O_{jvx} - E_{jvx}\big)^2}{E_{jvx}\big(1 - E_{jvx}\big)}$$

[ 65 ]

gde je *j* indeks stavke, *v* indeks $\theta$ intervala kojih ima *u*, $n_v$ broj ispitanika u grupi/inetrvalu, $O_{jvx}$ opažena proporcija odgovora *x*, $E_{jvx}$ modelski predviđena proporcija odgovora *x* na stavku.

Pokazatelj približno sledi hi–kvadrat raspodelu sa $u–q$ stepeni slobode, gde je $q$ broj parametara modela.

Ukoliko je broj $\theta$ intervala jednak 10, a za izračunavanje modelskih verovatnoća se koristi aritmetička sredina osobine u grupi, onda se radi o $Q_1$ statistiku Jenove (Yen, 1981), a ako se koristi medijana i broj intervala nije striktno određen u pitanju je Bokov statistik (Bock, 1972).

Mekinli i Mils (McKinley & Mills, 1985) su predložili sličan statistik – $G^2$ koji se zasniva količniku verodostojnosti (Orlando & Thissen, 2000):

$$G_j^2 = 2\sum_{v=1}^{10}\sum_{x=0}^{X_j} n_v \left[O_{jvx} ln\left(\frac{O_{jvx}}{E_{jv}}\right) + \left(1 - O_{jvx}\right) ln\left(\frac{1 - O_{jvx}}{1 - E_{jvx}}\right)\right]$$

[ 66 ]

Oznake su iste kao u izrazu [ 65 ], a statistik takođe približno sledi hi–kvadrat distribuciju sa $10–q$ stepeni slobode. Prikazan je izraz za politomne stavke sa $X_j$ kategorija odgovora.

Problem sa $G^2$ i $Q_1$ (i ostalim pokazateljima baziranim na hi–kvadratu) je taj što modelske predikcije ne mogu biti direktno poređene sa opaženim rezultatima, već zavise od samog modela. Naime, za razliku od Rašovog modela u kojem je sumacioni skor (opaženi rezultat) dovoljan statistik, u višeparametarskim modelima to nije slučaj. U Rašovim modelima intervale osobine ispitanika mogli bismo formirati po ukupnim skorovima i oni bi bili jednaki kao i oni formirani po procenama $\theta$. Počev od dvoparametarskog modela dovoljan statistik za nivo osobine $\theta$ je suma skorova na stavkama ponderisanih diskriminativnošću. U tom slučaju ispitanici sa jednakim sumacionim skorovima ne moraju imati iste nivoe $\theta$, a s druge strane, ispitanici sa različitim sumacionim skorovima mogu imati istu procenu osobine. Dakle, da bismo sortirali ispitanike po nivou osobine model mora biti prethodno primenjen i osobina mora biti procenjena na osnovu modela. Zbog toga se proporcije opaženih skorova računaju na osnovu modela, a trebalo bi da budu dostupne pre njegovog izračunavanja. Tako procenjene opažene proporcije tačnih odgovora su zavisne od modela pa procena distribucije pokazatelja fita nije pouzdana, prevashodno zbog nejasnih stepeni slobode (Orlando & Thissen, 2000).

Drugi problem je to što je podela u jednake $\theta$ intervale na gore opisani način zavisna od uzorka. Osim toga, rezultat prethodno navedenih statistika ($G^2$, $\chi^2$ i $Q_1$) zavisiće od broja intervala i njihovih granica.

Orlando i Tisen (Orlando & Thissen, 2000) su ponudili način da se ovi problemi prevaziđu. Predložili su postupak zasnovan na združenoj distribuciji verodostojnosti za svaki mogući ukupan skor. Opažene proporcije određenog odgovora (npr. tačno ili netačno) za svaki ukupan skor porede se sa očekivanjima zasnovanim na zajedničkoj distribuciji verovatnoće, a na osnovu verovatnoće svakog ukupnog skora za sve sklopove odgovora sa tim ukupnim skorom (Stone & Zhang, 2003).

Formula tog statistika je identična izrazu [ 65 ] osim što su indeksi grupa izmenjeni da naglase da se opažene i očekivane frekvencije računaju drugačije, za svaku grupu sa jednakim brojem tačnih odgovora posebno (Orlando & Thissen, 2000):

$$S-\chi_j^2=\sum_{k=1}^{r-1}\sum_{x=0}^{1}n_k\frac{\left(O_{jkx}-E_{jkx}\right)^2}{E_{jkx}\left(1-E_{jkx}\right)}$$

[ 67 ]

gde je *k* indeks grupe formirane po ukupnom skoru kojih ima *r*.

Postoji i varijanta analogna $G^2$ statistiku (Orlando & Thissen, 2000):

$$S-G_j^2=2\sum_{k=1}^{r-1}\sum_{x=0}^{X_j}n_k\left[O_{jk}ln\left(\frac{O_{jk}}{E_{jk}}\right)+\left(1-O_{jk}\right)ln\left(\frac{1-O_{jk}}{1-E_{jk}}\right)\right]$$

[ 68 ]

Opet, u odnosu na originalni $G^2$ razlika je samo u tome kako se računaju opažene i očekivane proporcije tačnih odgovora.

$S$–$\chi^2$ i $S$–$G^2$ približno slede hi–kvadrat raspodelu sa $r$–1–$q$ stepeni slobode.

Za $S$–$\chi^2$ postoji i generalizovani oblik namenjen politomnim stavkama (Kang & Chen, 2007).

Pored ovih statistika trebalo bi pomenuti i $\chi^{2*}$ statistik koji je definisao Stoun (Stone, 2000) u kojem se opažene frekvencije računaju na osnovu aposteriorne distribucije verovatnoća. S obzirom na to da u tom slučaju nije poznata distribucija statistika, koristi se procedura parametarskih simulacija (Monte Carlo). Postoji i varijanta ovog statistika $\chi^{2*}df$ , koja takođe koristi simulacije za izračunavanje skaliranog hi–kvadrat testa i stepeni slobode. Poslednja dva indikatora fita pokazali su dobra svojstva. Imaju veću snagu od $S$–$\chi^2$ istovremeno kontrolišući grešku tipa I i održavajući je u okviru zadatog nivoa značajnosti (de Ayala, 2022).

### 6.3.2 Pokazatelji fita stavki zasnovani na standardizovanim rezidualima

Rezidual je odstupanje opaženog odgovora od modelskih predikcija. Ako na primer, 1PL model za stavke sa binarnim skorovanjem predviđa da ispitanik *i* ima verovatnoću od 0,3 da odgovori tačno na stavku *j*, $P\left(X_{ij}=1\middle|\theta\right)=x_{ij}^E=0{,}3$, a on odgovori tačno ($x_{ij}=1$), rezidual bi iznosio $x_{ij}-x_{ij}^E=1-0{,}3=0{,}7$. Da je taj ispitanik odgovorio netačno, rezidual bi iznosio –0,3 (odnosno $0-0{,}3=-0{,}3$). Poredeći reziduale za dva odgovora, vidimo da je netačan odgovor više u skladu sa modelom (apsolutna vrednost reziduala je manja). Npr. neki drugi ispitanik *k* na istu stavku prema modelu ima verovatnoću tačnog odgovora od 0,9 ($P\left(X_{kj}=1\middle|\theta\right)=x_{kj}^E=0{,}9$) i na nju je odgovorio tačno ($x_{kj}=1$). Za ovoga ispitanika bi rezidual na posmatranoj stavci iznosio 0,1 ($1-0{,}9=0{,}1$). Ovaj odgovor ispitanika *k* na ovu stavku je više u skladu sa modelom od bilo kog odgovora ispitanika *i*.

Da bi se reziduali mogli sumirati potrebno ih je standardizovati. Verovatnoća odgovora ispitanika u Rašovim modelima (recimo PCM) računa se po izrazu (Linden, 2010):

$$E_{ij}\equiv_{x_{ij}}=\sum_{h=0}^{k_j}hP(x_{ijh}=h|\theta_i,b_{jh})$$

[ 69 ]

gde je $h$ kategorija skora, $k$ broj kategorija skorova na stavci $j$, a $P(x_{ijh}=h|\theta_i, b_{jh})$ verovatnoća da ispitanik ostvari skor u određenoj kategoriji uzevši u obzir parametre modela. Prikazani izraz je za politomni PCM model koji se, kada je odgovor skorovan binarno, svodi na izraz modela za binarno skorovane stavke.

Reziduale možemo standardizovati ako ih podelimo varijansom skora $x_{ij}$ koja se računa na sledeći način (Linden, 2010):

$$W_{ij} \equiv V(x_{ij}) = \sum_{h=0}^{k_j} (h - E_{x_{ij}})^2 P_{x_{ijh}}$$

[ 70 ]

Standardizovani rezidual tada će biti:

$$z_{ij} = \frac{x_{ij} - E_{ij}}{\sqrt{W_{ij}}}$$

[ 71 ]

Standardizovani reziduali mogu se sabirati za stavke i za ispitanike i na jednostavan način dobiti pokazatelji fita, kako za ispitanike – tako i za stavke. Ovaj način analize fitovanja dominantan je u softveru za jednoparametarske modele. Pokazatelji fita bazirani na standardizovanim rezidualima mogu se dalje transformisati u dva oblika koji se nazivaju ***autfit*** i ***infit***. I ova dva statistika predstavljaju formu hi–kvadrat testa (Linacre, 1993).

#### *6.3.2.1 Autfit*

Autfit (engl. *outfit*) je pokazatelj misfita (odstupanja od modelskih predviđanja) u oblasti ekstremnih nivoa osobine ili težina stavki. Najčešće se koristi u varijanti neponderisanih srednjih kvadrata (engl. *unweighted mean-square* – MSQ outfit) i označava sa MSQ autfit. To je pokazatelj fita osetljiv na autlajere, a zasnovan je na srednjem kvadratu reziduala (Lamprianou, 2019).

Ovaj pokazatelj računa se na standardizovanim rezidualima. Ranije smo opisali kako se računaju standardizovani reziduali (izrazi [ 69 ]–[ 71 ]). Njih uvrštavamo u formule za računanje infita i autfita.

Autfit za stavke se računa na sledeći način:

$$u_j = \frac{\sum_i^n z_{ij}^2}{n}$$

[ 72 ]

gde je $z$ standardizovani rezidual, a $n$ broj ispitanika.

#### *6.3.2.2 Infit*

Infit je misfit u oblasti srednjih nivoa osobine ispitanika ili srednjih težina stavki. Značajniji je pokazatelj misfita od autfita jer su izvesna odstupanja u zoni ekstremnih težina stavki donekle očekivana (pogađanje i nepažljivost). Infit ukazuje na ozbiljnija odstupanja od modela (npr. nejednake diskriminativnosti) (Fajgelj, 2020).

Formula za računanje misfita za stavke u obliku srednjih kvadrata ponderisanih informativnošću (engl. *information-weighted mean-square* – MSQ infit) je:

$$v_j = \frac{\sum_i^n W_{ij} z_{ij}^2}{\sum_i^n W_{ij}}$$

[ 73 ]

gde je $z$ standardizovani rezidual, a $W_{ij}$ ponder, odnosno, varijansa stavke (npr. p(1–p) ako je stavka binarno skorovana).

#### *6.3.2.3 Granične vrednosti infita i autfita*

MSQ infit i autfit mogu uzeti vrednosti od 0 do +∞, pri čemu je očekivana (normalna) vrednost 1.

Kada su u pitanju stavke, oba ukazuju na poremećaj u parametru nagiba stavke u jednoparametarskim modelima. To znači da stavka ima nagib koji značajno odstupa od modelskog. Podsetićemo, u jednoparametarskom modelu pretpostavka je da sve stavke imaju isti parametar nagiba, koji se u Rašovom modelu ne procenjuje, ali je implicitno postavljen na $a$=1. Niske vrednosti ukazuju na strmiji nagib, visoke vrednosti na blaži. Ni infit ni autfit ne govore koliko je opažena KKS udaljena od modelske. Ukoliko se nagibi opažene i modelske KKS ne razlikuju, ovi pokazatelji ne bi detektovali udaljenost (Wu i ostali, 2016).

Kao moguće granične vrednosti misfita u MSQ metrici navode se (Fajgelj, 2020; Linacre, 2002):

- <0,5 –*prigušeni šum* ili *overfit*. Ukazuje na odgovaranje koje je suviše predvidivo, odnosno, „suviše u skladu s modelom“ ili u skladu sa Gutmanovim modelom merenja. Ako je u formi *infita*, ukazuje na lokalnu zavisnost ili postojanje testleta[42].
- 0,5–1,5 –varijacije koje su produktivne za merenje
- 1,5–2 – varijacije koje nisu produktivne, ali ne degradiraju merenje
- >2 – šum, misfit koji degradira merenje
- misfit manji od 0,5 i veći od 1,5 smatra se „upozoravajućim“.

Međutim, ima i drugačijih predloga graničnih vrednosti. Kao uobičajeni prihvatljivi opseg MSQ misfita uzima se raspon 0,7–1,3 (Lamprianou, 2019; Mair, 2018; Wright & Linacre, 1994). Rajt i Linejker (Wright & Linacre, 1994) za testiranja sa „velikim ulozima“[43] (poput selekcionih) preporučuju stroži opseg

[42] Testleti su grupe stavki koje mere istu stvar, istu sadržinsku oblast, a konstruisani su da to čine na isti način i da imaju iste merne karakteristike. Mogu biti konstruisani kao celina (engl . *testlet* – mali test, testić) da bi bili zadavani zajedno, ali mogu služiti i kao alternativna pitanja koja mogu zamenjivati jedno drugo u alternativnim formama testova (Wainer i ostali, 2007). Primer testleta bio bi zadatak u kojem bi ispitaniku kao stimulus bila predočena Slika 5 – Ukrštanje dve KKS u 2PL modelu, koja bi bila praćena pitanjima: 1) Koja od dve prikazane stavke je lakša? 2) Koja od dve prikazane stavke je diskriminativnija? 3) Za koju od dve stavke je verovatnije da će ispitanik sa nivoom osobine –1 logit dati tačan odgovor?

[43] Možda bi bilo bolje reći sa „značajnim posledicama“.

0,8–1,2, a za ankete sa skalama procene 0,6–1,4. Kao što možemo videti, ovi preporučeni rasponi zavise od vrste instrumenta i namene testiranja.

Inače, ovi opsezi zasnovani su na distribuciji MSQ statistika. Očekivana vrednost statistika je 1. Na uzorcima od 200 ispitanika, 95% interval poverenja je 0,8–1,2. Na uzorcima od 2000 ispitanika taj interval je 0,94–1,06 (Wu i ostali, 2016). Prilikom interpretacije bi trebalo i o tome voditi računa.

Vrednosti niže od donje granice ovih raspona ukazuju na *prigušeni šum*, a više na *šum*. U svakom slučaju ovo su *predloženi* rasponi kojih se ne moramo držati po svaku cenu ako imamo dobar razlog da od njih odstupimo.

Pokazatelji infita i autfita imaju i standardizovanu formu (ZStd), koja ima normalnu raspodelu sa $M$=0 i $\sigma$=1. Ovaj pokazatelj se često označava sa $t$ i donekle rešava problem promenljivih opsega prihvatljivih vrednosti misfita u zavisnosti od veličine uzorka. Međutim, to stvara drugi problem. Ovaj statistik će uvek biti značajan na velikim uzorcima. Ako je uzorak dovoljno velik, svaka razlika će biti značajna bez obzira na njen praktični značaj koji može biti i zanemarljiv (Wu i ostali, 2016). Zbog toga su nam pokazatelji u MSQ metrici bitniji i tek ako oni ukazuju na značajan misfit tumače se standardizovane varijante, koje se inače mogu zanemariti. ZStd se smatra značajnim ako izlazi iz opsega ±1,96. Negativne vrednosti ukazuju na prigušeni šum, pozitivne na šum.

Ovi pokazatelji misfita nam ne pokazuju apsolutni kvalitet stavki već samo relativni. Oni nam govore kakva je neka stavka u odnosu na ostale stavke u testu. Pošto nam govore o nagibu stavki, veliki misfit može da znači da je stavka manje diskriminativna od ostalih stavki u testu. Međutim, kada bismo tu istu stavku prebacili u test koji je sastavljen od sličnih, manje diskriminativnih stavki, pokazatelji ne bi otkrili misfit. To dovodi do sledeće zanimljive posledice, a to je da možemo imati test koji fituje Rašov model, a da on bude sastavljen od jednako nediskriminativnih stavki koje ne mere ništa. Kada bismo na takvom testu izračunali koeficijent pouzdanosti iz klasične testne teorije (recimo Kronbahov α) dobili bismo nisku vrednost. Prilikom konstrukcije testa trebalo bi proveriti i koeficijente diskriminativnosti (korigovane ajtem-total korelacije). Ako koristimo pokazatelje fita za izbor stavki, preporuka je da se biraju oni sa vrednostima nešto nižim od 1 ili 0,9 (recimo 0,7–0,9 u MSQ metrici) jer će to biti diskriminativnije stavke (Wu i ostali, 2016).

Iako su nastali u okviru Rašovog modela, infit i autfit su primenljivi i u drugim IRT modelima (Meijer & Sijtsma, 2001).

### 6.3.3 Valdov (Wald) test

I ovaj pokazatelj fita stavki namenjen je Rašovim modelima. Logika koja stoji iza njega je invarijantnost procene parametara. Ako model fituje, parametri procenjeni na poduzorku ispitanika sa višim i na poduzorku sa nižim nivoom osobine ne bi smeli da se razlikuju (nije nužno da podela uzorka bude po visini osobine). Za bilo koju podelu uzorka na dva dela sa zasebnim procenama $\hat{b}_j^{(1)}$ i $\hat{b}_j^{(2)}$, te $\hat{\sigma}_b^{(1)}$ i $\hat{\sigma}_b^{(2)}$ važi:

$$s_j = \frac{\hat{b}_j^{(1)} - \hat{b}_j^{(2)}}{\sqrt{\hat{\sigma}_b^{(1)} - \hat{\sigma}_b^{(2)}}} \approx N(0,1)$$

[ 74 ]

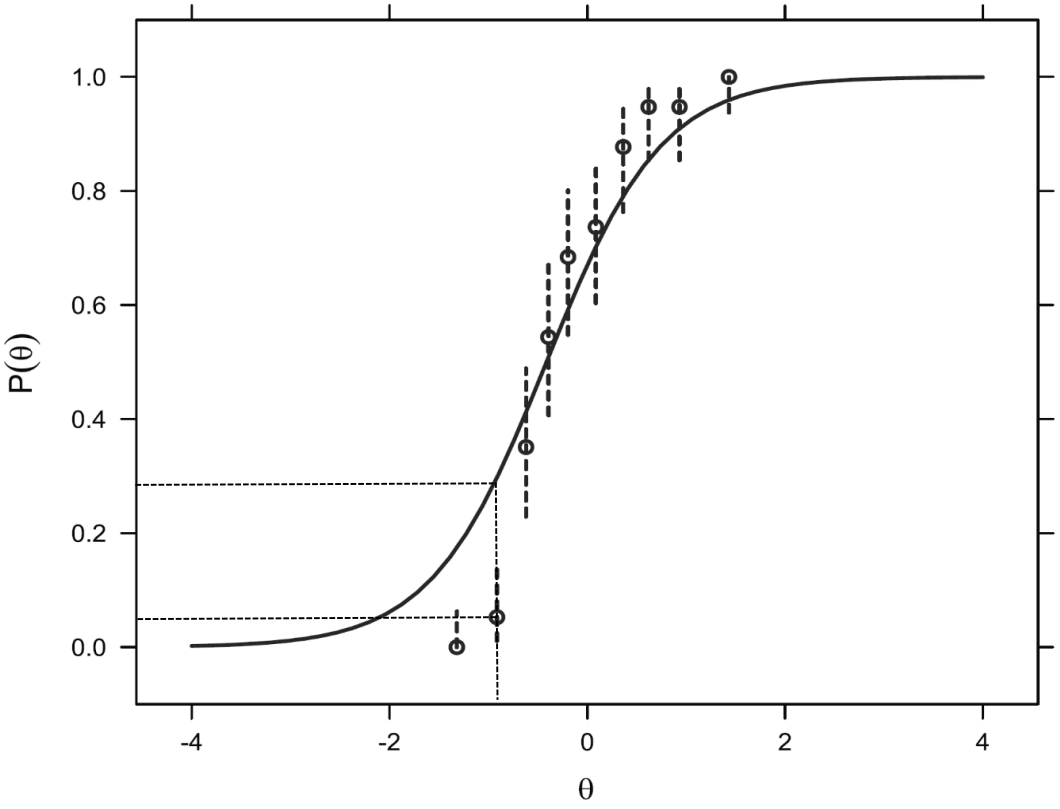
gde su $\hat{b}_j$ procene parametara, a $\hat{\sigma}_b$ standardne devijacije tih procena.

Ovaj test može poslužiti i za otkrivanje DIF (diferencijalnog funkcionisanja stavki – odeljak 8), a tada se grupe formiraju na osnovu neke eksterne varijable.

### 6.3.4 Grafička provera fita stavke

Jedan od prvih načina provere fita za stavke bio je grafički. Postupak je jednostavan i sastoji se u tome da na grafikonu istovremeno predstavimo modelsku karakterističnu krivu stavki i empirijsku krivu. Primenljiv je u svim modelima.

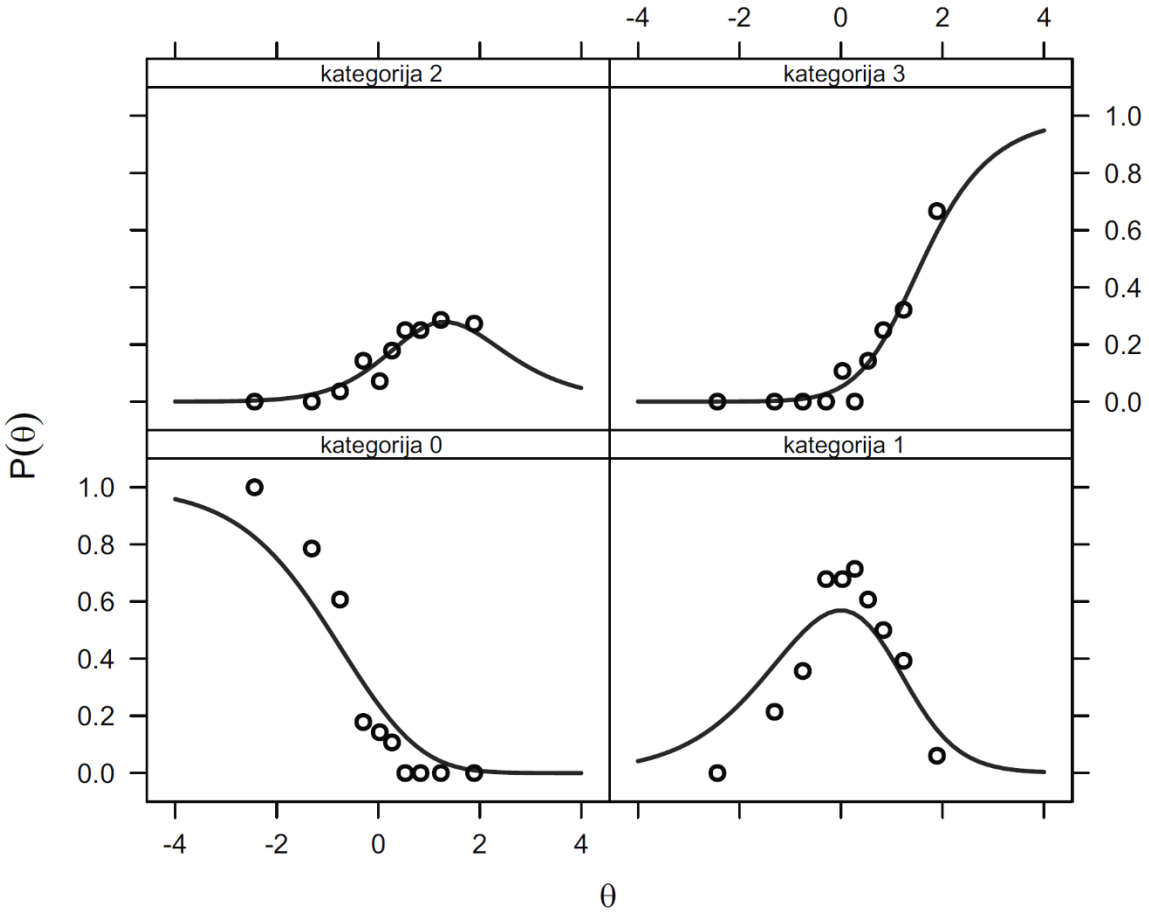
Na slici (Slika 16) punom linijom prikazana je modelska karakteristična kriva binarno skorovane stavke. Kružićima su prikazane empirijski opažene proporcije tačnih odgovora na različitim nivoima $\theta$. Kružići ne predstavljaju jedan nivo osobine već jedan interval $\theta$ koji je reprezentovan prosečnom vrednošću. Isprekidanim linijama dati su 95% intervali poverenja za te verovatnoće. Intervali poverenja delom zavise od veličine uzorka, pa kada su uzorci veliki nije ih nužno prikazivati (de Ayala, 2022).




Slika 16 – Modelska i empirijska karakteristična kriva binarno skorovane stavke

Stavka je saglasna sa podacima ako se kružići nalaze što bliže modelskoj KKS, odnosno, kada intervali poverenja obuhvataju modelsku krivu. Na slici vidimo da su predviđanja modela za ovu stavku uglavnom tačna, osim na nivou $\theta$ ispod –0,5 logita. Npr. na nivou osobine od –0,9 logita model predviđa verovatnoću od 0,28, a opažena je oko 0,04.

Da li će ovaj misfit stavke biti prihvatljiv zavisiće i od toga da li nam je opseg osobine u kojem su uočena odstupanja bitan za merenje koje planiramo. Ako planiramo da merimo (recimo) darovite ispitanike i očekujemo da će u uzorku nivo merene osobine bit natprosečan, onda nam ovaj misfit može biti prihvatljiv, ali ako planiramo da merenje vršimo na opštoj populaciji, onda verovatno neće (de Ayala, 2022).




Slika 17 – Modelske i empirijske karakteristične krive kategorija politomno skorovane stavke

Na slici (Slika 17) prikazane su modelske i empirijske karakteristične krive kategorija politomno skorovane stavke (model stepenovanog ocenjivanja – PCM, skorovi 0–3). Vidimo da model dobro predviđa odgovaranje u višim kategorijama skorova (2 i 3), dok u nižim kategorijama postoje izvesna odstupanja. Skorovi u kategoriji 0 imaju veću opaženu proporciju nego što model predviđa kod osoba sa nižim nivoima osobine (od –3 do –1 logita), a manju kod prosečnih i nadprosečnih (od 0 do 2 logita). Kada su u pitanju skorovi u kategoriji 1, model precenjuje verovatnoću odgovaranja u ovoj kategoriji za ispod prosečne ispitanike, a potcenjuje za natprosečne. Reklo bi se da bi parametri nagiba u ovim kategorijama trebalo da budu strmiji.

Iako su nam dostupni i numerički pokazatelji fita, kada se utvrdi postojanje misfita za neku stavku svakako je korisno pogledati grafički prikaz koji nam može ukazati na to koliki je taj misfit i u kojoj oblasti kontinuuma latentne osobine se javlja. Međutim, svi ovi numerički pokazatelji misfita nam govore samo da li on postoji ili ne. Uzroke moramo utvrditi kvalitativnom analizom.

## *6.4 Provera fita osoba*

Pokazatelji saglasnosti za ispitanike (engl. *appropriateness* ili *person-fit*) nam govore da li u odgovaranju ispitanika ima nešto neobično. Npr. ispitanik je odgovorio tačno na stavku koja po težini mnogo prevazilazi nivo njegove osobine, ili nije odgovorio na stavku koja je mnogo lakša. Većina pokazatelja saglasnosti, odnosno, fitovanja osoba bazira se na saglasnosti vektora (sklopa) odgovora ispitanika sa pretpostavljenim modelom (Meijer & Sijtsma, 2001). Razlikuju se po primenljivosti u različitim IRT modelima, a najveći broj njih napravljen je za binarno skorovane stavke (Embretson & Reise, 2000).

Ovde ćemo prikazati pokazatelje saglasnosti koji se najčešće javljaju u softveru za IRT, a mogu se koristiti i u binarnim i u politomnim modelima.

Dva pokazatelja misfita za ispitanike koja poreklo vode iz Rašovog modela su ***autfit*** i ***infit***. Varijante ovih statistika za proveru fita stavki prikazane su u odeljcima 6.3.2.1 – 6.3.2.3.

Autfit u MSQ metrici za ispitanike računa se na sledeći način:

$$u_i = \frac{\sum_j^m z_{ij}^2}{m}$$

[ 75 ]

gde je *z* standardizovani rezidual, a *m* broj stavki u testu.

Jedina razlika između izraza za autfit stavki [ 72 ] i [ 75 ] je u tome što se u izrazu za ispitanike sumacija reziduala vrši po stavkama i izraz deli brojem stavki *m,* a u izrazu za stavke sumacija se obavlja po ispitanicima i izraz se deli brojem ispitanika *n.*

Analogno izrazu za infit stavki, formula za računavanje infita za ispitanike u obliku srednjih kvadrata ponderisanih informativnošću (*information-weighted mean-square* – MSQ infit) je:

$$w_i = \frac{\sum_j^m W_{ij} z_{ij}^2}{\sum_j^m W_{ij}}$$

[ 76 ]

gde *z* standardizovani rezidual, a $W_{ij}$ ponder, odnosno, varijansa stavke.

Jedina razlika od izraza za stavke [ 73 ] je u tome što se tamo sumacija reziduala vrši po ispitanicima, a ovde po stavkama.

I za ispitanike postoje pokazatelji infita i autfita u standardizovanoj metrici. Interpretacija i granične vrednosti infita i autfita za ispitanike jednake su kao i za stavke (videti odeljak 6.3.2.3). Osim u Rašovom, primenljivi su i u drugim modelima (Meijer & Sijtsma, 2001).

Još jedan od često korišćenih pokazatelja fita za osobe je statistik Drasgova, Levina i Viliamsa (Drasgow i ostali, 1985) poznat kao $Z_3$. Zasniva se na logaritmu verodostojnosti sklopa odgovora ispitanika koji je dat u izrazu [ 12 ], koji je ovde napisan u formi za nivo osobine $\theta_i$ (Drasgow i ostali, 1985):

$$LL|\theta_i = x_{ij} ln(P_j(\theta_i)) + (1 - x_{ij}) ln(Q_j(\theta_i))$$

[ 77 ]

Očekivani logaritam verodostojnosti sklopa ispitanikovih odgovora na nekom skupu stavki zavisiće od procenjenog nivoa osobine ispitanika. Zato se sam *LL* ne može koristiti kao pokazatelj fita već ga je potrebno standardizovati na neki način.

*LL* se standardizuje tako što se umanji za očekivanu (prosečnu) vrednost distribucije uzorkovanja logaritma verodostojnosti koja se izračunava sumiranjem prethodnog izraza na svim ispitanicima:

$$E(LL|\theta_i) = \sum [P_j(\theta_i)\, ln(P_j(\theta_i)) + Q_j(\theta_i) ln(Q_j(\theta_i))]$$

[ 78 ]

a zatim se ta razlika podeli varijansom uzorkovanja *LL* za određeni nivo osobine, koja se računa na sledeći način:

$$V(LL|\theta_i) = \sum P_j(\theta_i)\, Q_j(\theta_i) [\ln(P_j(\theta_i))/\, Q_j(\theta_i)]^2$$

[ 79 ]

pa izraz za $Z_3$ do kraja izgleda ovako:

$$Z_3(\theta_i) = \frac{\sum LL|\theta_i - \sum E(LL|\theta_i)}{\sqrt{\sum V(LL|\theta_i)}}$$

[ 80 ]

$Z_3$ ima standardizovanu normalnu distribuciju sa aritmetičkom sredinom 0 i standardnom devijacijom 1 i interpretira se u skladu sa tim. Jako niske vrednosti (ispod –1,96) ukazuju na *misfit šumnog tipa*. Odgovori ispitanika su nepredvidivi i nisu u skladu sa modelom. Vrednosti pokazatelja preko 1,96 ukazuju na suviše predvidivo odgovaranje, odnosno na *prigušeni šum*. Neki autori (de Ayala, 2022; Embretson & Reise, 2000) ukazuju da $Z_3$ nema uvek standardizovanu normalnu distribuciju i da činioci koji mogu uticati na lošu procenu parametara (nedostajući podaci, broj stavki, fit modela u celini) mogu uticati i na ovaj pokazatelj.

Prikazan je pokazatelj $Z_3$ namenjen binarno skorovanim stavkama, a za uređene politomne modele na raspolaganju je pokazatelj $Z_h$ (de Ayala, 2022; Drasgow i ostali, 1985).

# 7 Informativnost

U IRT modelima pouzdanost nema centralno mesto kao u KTT. Pokazatelj koji zauzima mesto pouzdanosti u IRT-u je *informativnost*. Ona je pokazatelj kvaliteta, efikasnosti i grešaka merenja (Fajgelj, 2020).

Standardna greška merenja u IRT modelima za binarno skorovane stavke definiše se kao standardna greška proporcije:

$$\sigma_{Eij} = \sqrt{\frac{1}{p_{ij}(1-p_{ij})}}$$

[ 81 ]

gde je $p_{ij}$ verovatnoća odgovora dobijena iz formule modela. Treba primetiti da će standardna greška merenja u ovom slučaju biti najmanja kada je $p_{ij}$=0,5, odnosno kada je težina stavke uparena sa nivoom osobine ispitanika i on ima verovatnoću od 50% da je reši tačno (ili pogrešno).

Pošto se iskazuje u jedinicama parametra[44] $\theta$ i njegova je funkcija, standardna greška nije aditivna, pa se ne može sabirati za stavke kako bi se dobila standardna greška merenja za test niti za ispitanika (Fajgelj, 2020).

Britanski statističar Ronald Elmer Fišer (Ronald Aylmer Fisher) definisao je ***informativnost*** kao preciznost sa kojom se može proceniti neki parametar (prema: Baker, 2001) ili kao količinu informacija koju neka opaziva slučajna varijabla nosi o nekom nepoznatom parametru koji izaziva tu varijablu (Fajgelj, 2020). Problem neaditivnosti standardnih grešaka je rešio uvidevši aditivnost ? recipročnih kvadrata standardnih grešaka, a informativnost je operacionalizovao upravo tako:

$$I(\theta) = \frac{1}{\sigma_{Eij}^2}$$

[ 82 ]

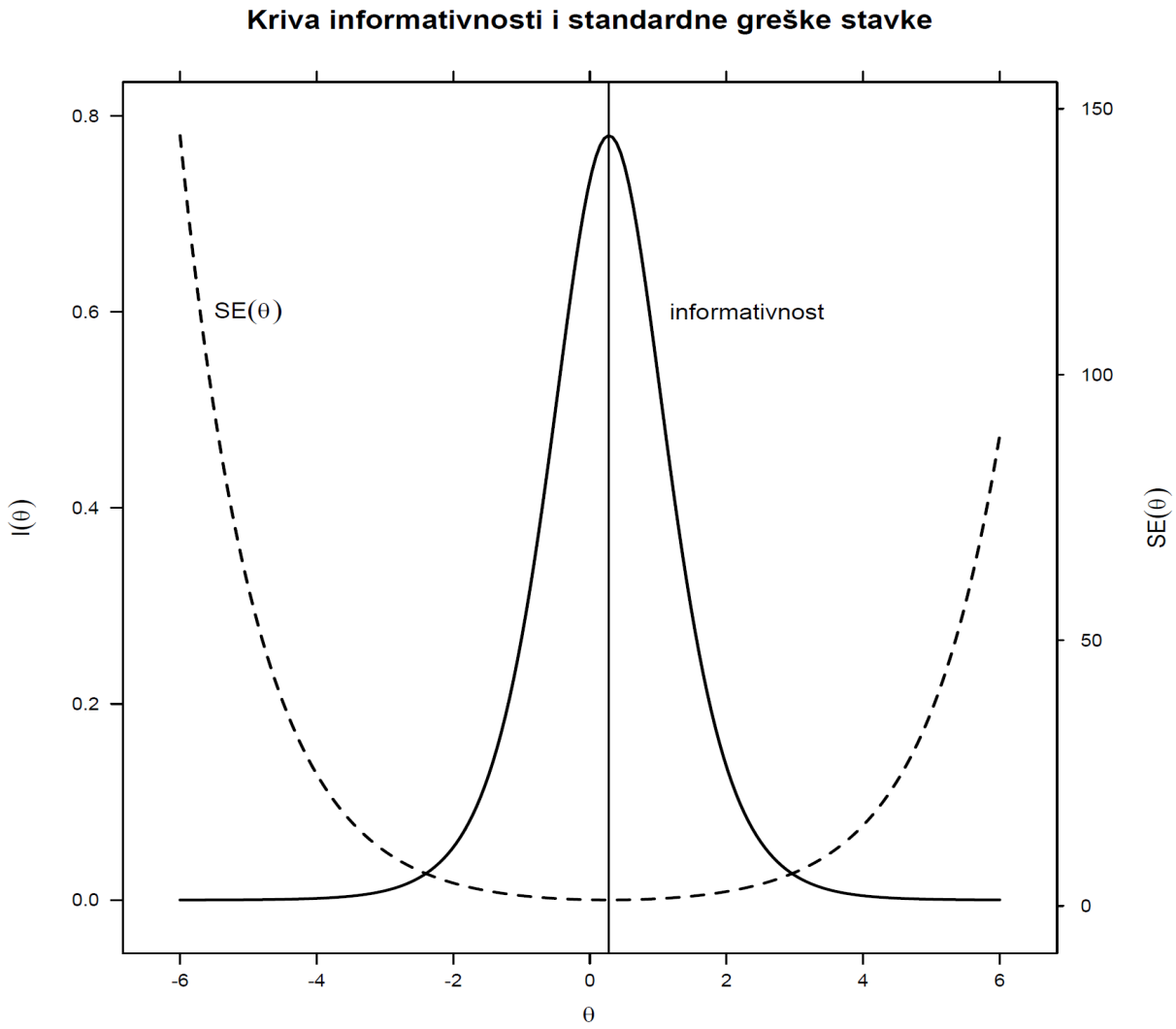
U 1PL modelu informativnost stavke *j* je (Fajgelj, 2020):

$$I_j(\theta) = \sum_{i=1}^{n} P_j(\theta_i)Q_j(\theta_i) = P_j(\theta)Q_j(\theta) = \sum_{i=1}^{n} \frac{1}{\sigma_{Eij}^2}$$

[ 83 ]

gde je $P_j(\theta)$ funkcija verovatnoće modela $Q_j = 1 - P_j(\theta)$.

[44] Što je zgodno za formiranje intervala poverenja.



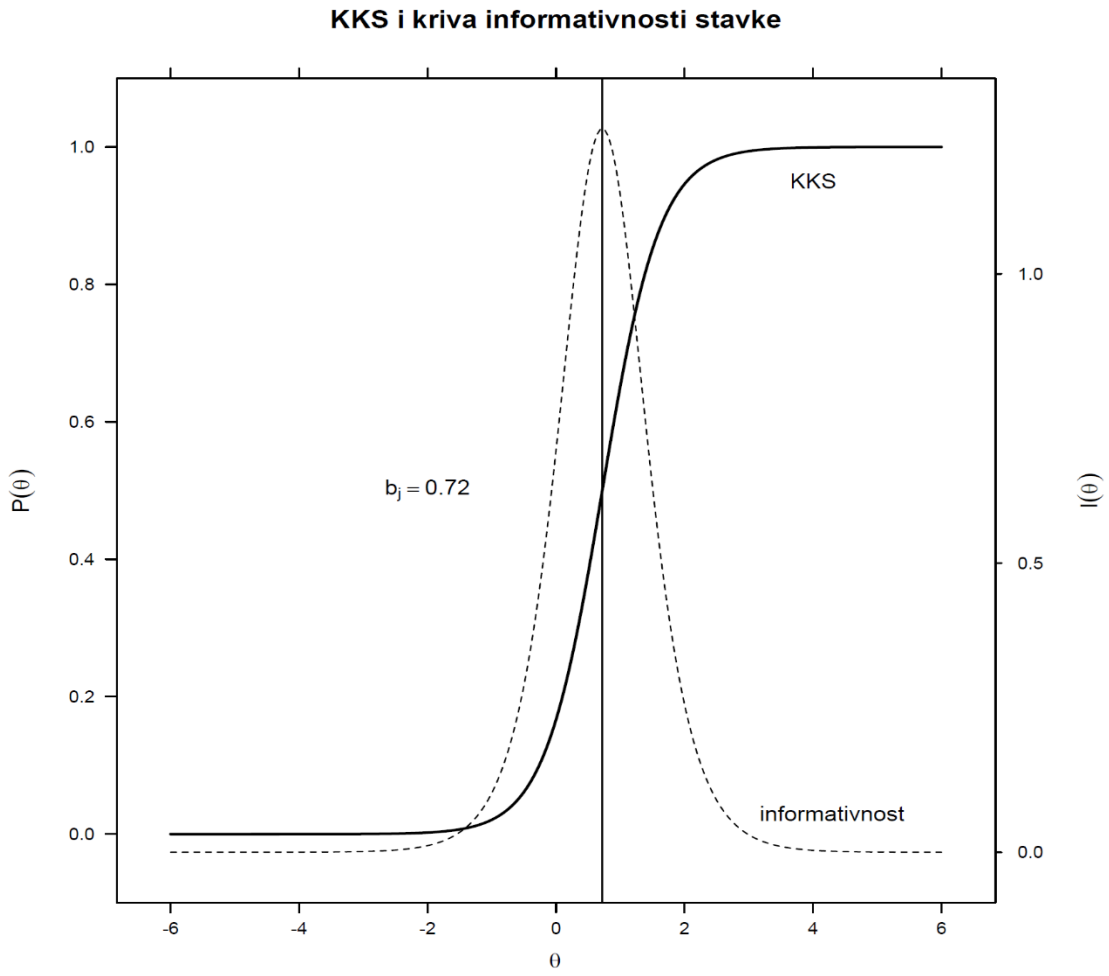

Slika 18 – Odnos informativnosti i standardne greške merenja na primeru jedne stavke

Gornja slika prikazuje odnos informativnosti i standardne greške merenja u IRT modelima. Vidimo da su vrednosti recipročne, ali da odnos nije linearan.

U prvom delu izraza [ 83 ] $\theta_i$ ima indeks ispitanika da bi bilo jasno da se informativnost stavke za ispitanike sa različitim nivoima osobine može sumirati i na taj način dobiti ukupna informativnost stavke na uzorku. U drugom delu izraza su indeks ispitanika i sumacija izbačeni zbog preglednosti, da bismo videli da se u suštini radi o varijansi odgovora (pošto se radi o binarno skorovanoj stavci, varijansa je proizvod proporcija dva moguća skora na stavci).

Dakle, informativnost je recipročna kvadratu standardne greške. Ako znamo da je standardna greška standardna devijacija uzoračkih parametara oko pravog, možemo reći da je to varijansa uzoračkih skorova oko pravog. Pošto se informativnost dobija sabiranjem tih recipročnih vrednosti, može se reći da je ona *varijansa skora*. S obzirom na to da je ovako definisana informativnost čisti broj, ona je aditivna. Može se izračunati za svaku stavku, za test u celini, pa čak i za ispitanika.

Informativnost stavke je funkcija nivoa osobine, pošto su ostali parametri za jednu stavku konstantni. Otuda se ona označava sa $I_j(\theta)$. U softveru za IRT analizu potrebno je obično zadati nivo osobine (ili raspon) za koji se računa informativnost stavke ili testa. Ovo je interesantno svojstvo jer na taj način možemo utvrditi za koji nivo osobine je test koji analiziramo najinformativniji, odnosno, najprecizniji. Softver za IRT nam obično omogućava prikaz krive informativnosti stavke i/ili testa, gde možemo videti na kom nivou osobine je stavka ili test najinformativnija.




Slika 19 – Uporedni prikaz KKS i informativnosti stavke

Na slici (Slika 19) su prikazane karakteristična kriva binarno skorovane stavke i njena kriva informativnosti nastale na osnovu 2PL modela. Primetite da se maksimum krive informativnosti nalazi na lokaciji stavke, odnosno nivou latentne osobine koji odgovara njenoj težini.

Kada smo izračunali informativnost stavke ili testa, moguće je na osnovu njenog odnosa sa standardnom greškom merenja izračunati grešku merenja za određeni nivo osobine:

$$\sigma_{E\theta} = \frac{1}{\sqrt{I(\theta)}}$$

[ 84 ]

Iz ovoga proizilazi i da je standardna greška merenja stavke (pa i testa) funkcija nivoa osobine. I

stvarno, kao što je ranije već rečeno u IRT modelima standardne greške merenja su različite za različite nivoe osobine. Standardne greške će biti najmanje kada je stavka ili test uparena sa nivoom osobine ispitanika (ili uzorka). One su najmanje kod srednjih nivoa osobine, a rastu ka ekstremnim nivoima. Intervali poverenja oko dobijenih mera se formiraju na osnovu standardnih grešaka merenja, ali nisu jednake širine duž celog kontinuuma latentne osobine.

Formula za izračunavanje informativnosti stavke u troparametarskom logističkom modelu za dihotomne stavke je (Baker, 2001):

$$I_j(\theta) = \left[a_j^2 \frac{1 - P_j(\theta)}{P_j(\theta)}\right]\left[\frac{(P_j(\theta) - c_j)^2}{(1 - c_j)^2}\right]$$

[ 85 ]

Ako nema pogađanja i u izrazu [ 85 ] izjednačimo parametar $c_j$ sa 0, dobijamo formulu informativnosti za dvoparametarski model za dihotomne stavke (Baker, 2001):

$$I_j(\theta) = \left[a_j^2 \frac{1 - P_j(\theta)}{P_j(\theta)}\right] * P_j(\theta)^2 = a_j^2 P_j(\theta) Q_j(\theta)$$

[ 86 ]

Ako je parametar nagiba $a_j$ jednak za sve stavke (recimo 1, kao u Rašovom modelu), onda se formula svodi na formulu informativnosti za jednoparametarski dihotomni model [ 83 ].

Parametar nagiba značajno utiče na informativnost. Već je rečeno da je maksimalna informativnost binarno skorovanih stavki u jednoparametarskom logističkom modelu jednaka 0,25. Stavka će imati najvišu informativnost na nivou crte gde je uparena sa ispitanikom, odnosno kada ispitanik ima jednaku verovatnoću da stavku reši tačno ili da pogreši. Zamislimo sada takvu maksimalnu informativnost iz 1PL modela i pogledajmo izraz za informativnost za binarne stavke [ 86 ] u 2PL modelu. U izrazu za dvoparametarski model vrednost dobijenu izrazom za jednoparametarski model potrebno je pomnožiti kvadratom parametra diskriminativnosti. Ako parametar diskriminativnosti iznosi 1 (kao što je to implicitno određeno u 1PL modelu), ništa se ne menja. Međutim, kada bi parametar diskriminativnosti bio $a_j$=0,70, onda bi informativnost bila jednaka $0{,}70^2$*0,25=0,12. S druge strane, ako bi parametar diskriminativnosti iznosio $a_j$=1,30, onda bi informativnost bila jednaka $1{,}30^2$*0,25=0,42. Što je parametar diskriminativnosti stavke viši, kriva informativnosti biće zašiljenija i viša, a što je on niži – niža i spljoštenija.

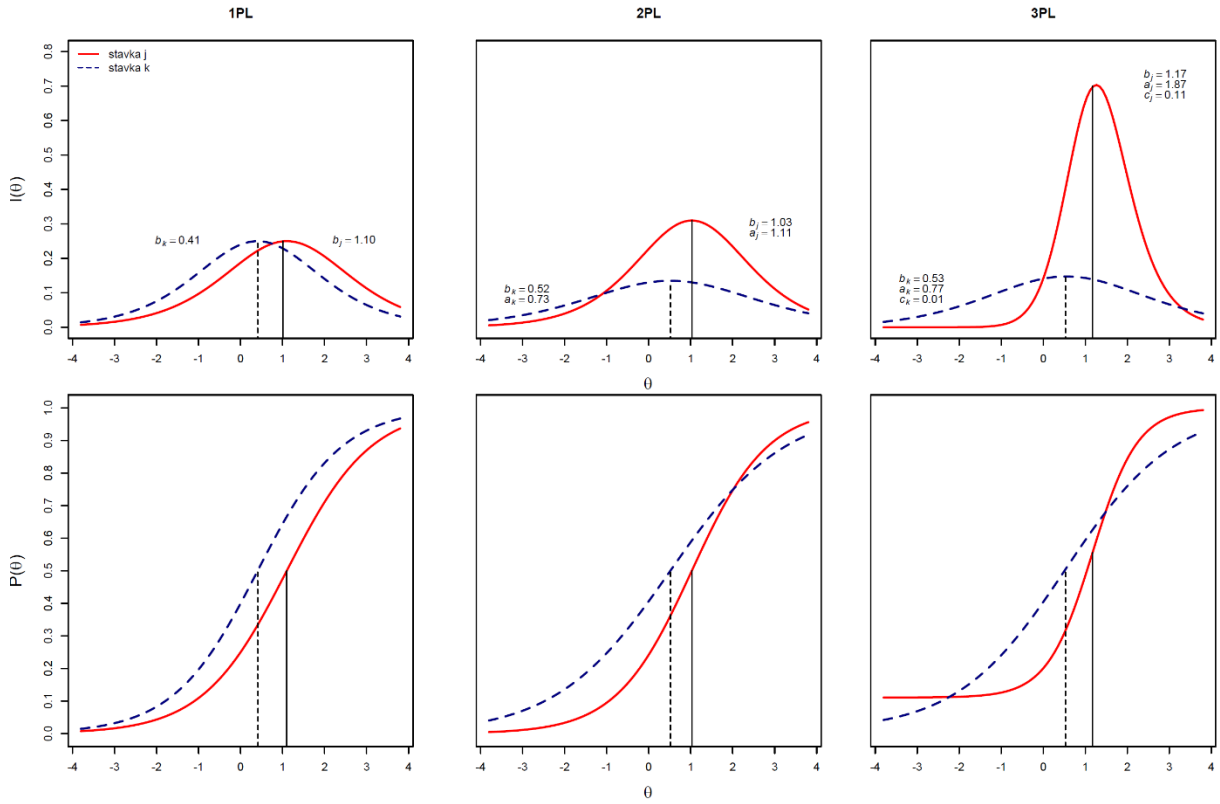
No, nije sve ni tako jednostavno.

Stavka sa višim parametrom diskriminativnosti imaće značajno višu informativnost na delu kontinuuma crte koji je blizak njenoj težini, od stavke slične težine, ali niže diskriminativnosti. Međutim, ista ta stavka (sa višim parametrom diskriminativnosti) može biti manje informativna (od one sa nižim) na delu kontinuuma latentne osobine udaljenijem od njene težine (Slika 20 – Paneli 2PL i 3PL gornji red: manje diskriminativna stavka je informativnija u nižim oblastima latentne osobine). Ovo bi trebalo imati u vidu jer svaka stavka doprinosi informativnosti testa za nivoe osobine na kojima je informativna (u bilo kom iznosu – što više to bolje).

Da bismo videli kako parametar pseudopogađanja utiče na informativnost zamislimo opet maksimalnu informativnost iz 1PL modela (0,25), da parametar diskriminativnosti iznosi $a_j$=1, i pogledajmo izraz [ 85 ]. Drugi deo izraza biće manji što je parametar $c_j$ viši, pa će samim tim i proizvod (informativnost) biti manji. Razlomak u prvom delu izraza daje proporciju Q($\theta$) u odnosu na P($\theta$). U našem slučaju, pošto je parametar diskriminativnosti $a_j$=1, a P($\theta$)=Q($\theta$)=0,5, prvi deo izraza biće 1.

Vrednost drugog dela izraza se smanjuje srazmerno vrednosti parametra $c_j$. Kada je parametar $c_j$=0, drugi deo izraza je jednak kvadratu P($\theta$), odnosno u ovom slučaju $0{,}5^2$=0,25 i informativnost je jednaka kao u 1PL modelu. Bilo koja vrednost $c_j$ osim 0 umanjuje drugi deo izraza i konačan proizvod. U našem primeru, ako sve ostane isto, a parametar $c_j$ uzme vrednost 0,20, informativnost pada na 0,14, a ako je $c_j$=0,23 informativnost se prepolovljuje u odnosu na prvobitnu i postaje 0,123. Dakle, viši parametar pseudopogađanja utiče na snižavanje informativnosti, a merenje postaje nepreciznije.

U jednoparametarskom i dvoparametarskom modelu za binarne stavke informativnost stavke je najviša na nivou crte koji odgovara parametru težine stavke. U troparametarskom modelu to nije slučaj. Informativnost stavke je tu najviša u tački koja je pomerena malo udesno od lokacije stavke (Hambleton i ostali, 1991) (videti: Slika 20, panel 3PL).




Slika 20 – Uporedni prikaz funkcija informativnosti i karakterističnih kriva dve stavke u 1PL, 2PL i 3PL modelu

Na slici (Slika 20) su prikazane krive informativnosti (gore) i karakteristične krive (dole) stavki *j* i

*k* dobijene na osnovu tri različita modela. Ako pogledamo panel 1PL, vidimo da je maksimalna informativnost obe stavke 0,25, a razlika u krivama je samo u njihovoj lokaciji. Uporedni grafikoni krivih informativnosti u 1PL modelima sami ne donose puno informacija. Sve krive su iste, jedino su im lokacije različite. Međutim, lokacija stavke nam je bitna informacija prilikom konstrukcije testa, kada nam je cilj da napravimo test prilagođen određenoj ciljnoj populaciji.

Na panelu 2PL (Slika 20) vidimo iste dve stavke modelovane dvoparametarskim modelom. Stavka *k* je lakša i ima parametar diskriminativnosti ispod 1 ($a_k$=0,73). Samim tim njena maksimalna informativnost je manja od 0,25 i iznosi 0,13, a nalazi se na lokaciji stavke ($b_k$=0,52). S druge strane, stavka *j* ima parametar diskriminativnosti preko 1 ($a_j$=1,11), a njena maksimalna informativnost je takođe na lokaciji stavke ($b_j$=1,03) i iznosi 0,31 (viša je od maksimalne moguće informativnosti binarno skorovane stavke u 1PL modelu).

Na kraju, pogledajmo panel 3PL (Slika 20). Vrednosti za stavku *k* se nisu puno promenile zato što ona ima vrlo nisku vrednost parametra pseudopogađanja $c_k$=0,01. Promena kod stavke *j* je znatno veća. Ona ima parametar $c_j$=0,11. Ako pogledamo krivu informativnosti ove stavke, vidimo da je informativnost ove stavke za nivoe crte ispod 0 logita praktično nulta. To je ujedno i lokacija na kontinuumu latentne varijable do koje je verovatnoća tačnog odgovora jednaka 0,11, bez obzira na nivo osobine (Slika 20, panel dole desno). Tek odatle se verovatnoća tačnog odgovora menja sa porastom nivoa osobine, a zbog suženog raspona osobine nagib postaje strmiji a stavka diskriminativnija (ali u užem rasponu merene osobine). Umanjenje maksimalne informativnosti koje je posledica povišenog $c_j$ kompenzuje se povišenim $a_j$, tako da ova stavka ima maksimalnu informativnost od 0,70 na nivou osobine malo višem od lokacije stavke, što se vidi po tome što je vrh krive informativnosti pomeren udesno u odnosu na težinu stavke $b_j$=1,17 logita.

*Informativnost testa* dobijamo sumiranjem informativnosti stavki po nivoima latentne osobine. Birnbaum (Birnbaum, 1968) kao jednačinu za izračunavanje informativnosti testa predlaže sledeći izraz:

$$I(\theta) = \sum_{j=1}^{m} I_j(\theta)$$

[ 87 ]

gde *j* označava stavku, a *m* broj stavki u testu.

Ono što je zanimljivo i što čini upotrebu IRT modela posebno privlačnom za konstrukciju testova je to što možemo izračunati informativnost testa sastavljenog od bilo koje kombinacije stavki (Baker, 2001; DeMars, 2010). Nije nužno da takav test bude prethodno primenjen na uzorku, ali je potrebno da informativnosti stavki budu poznate. To praktično znači da ukoliko imamo banku stavki sa poznatim mernim karakteristikama, iz nje po potrebi možemo sastaviti test sa unapred poznatim mernim osobinama.

Kao i za stavke, tako i za test, informativnost zavisi od nivoa osobine. Koliko će instrument biti informativan na nekom od nivoa latentne osobine zavisi od pokrivenosti tih nivoa osobine stavkama. Kriva informativnosti testa sastavljenog od binarno skorovanih stavki ne mora biti simetrična pa ni unimodalna,

a njen oblik takođe će zavisiti od pokrivenosti pojedinih nivoa osobine odgovarajućim stavkama. Ukoliko prilikom kreiranja testa koristimo banku stavki, jedna od mogućih strategija može biti da biramo stavke koje će biti prilagođene očekivanom nivou osobine u uzorku na kojem želimo da obavimo merenje. Ako nemamo neka posebna očekivanja vezana za distribuciju osobine u uzorku, onda bi dobra strategija bila birati stavke koje će pokriti širi opseg osobine, s tim da bi veći broj stavki trebalo da pokriva nivoe osobine bliže proseku nego ekstremne nivoe.

Zanimljivo je što se informativnost može izraziti i u metrici pouzdanosti iz KTT-a. Rekli smo da je informativnost vrednost recipročna kvadratu standardne greške merenja [ 82 ]. S druge strane, i standardna greška merenja je vrednost suprotna pouzdanosti i može se preko nje izračunati na sledeći način (Embretson & Reise, 2000):

$$\sigma_E = \sigma\sqrt{1 - r_{tt}}$$

[ 88 ]

gde je σ standardna devijacija instrumenta, a $r_{tt}$ njegova pouzdanost.

Ako pretpostavimo da su skorovi standardizovani, onda se izraz [ 88 ] može napisati samo kao:

$$\sigma_E = \sqrt{1 - r_{tt}}$$

[ 89 ]

Iz toga sledi da je pouzdanost jednaka:

$$r_{tt} = 1 - \sigma_E^2$$

[ 90 ]

A ako standardnu grešku zamenimo desnom stranom izraza [ 84 ] dobijamo formulu za računanje pouzdanosti preko informativnosti:

$$r_{tt\theta} = 1 - \frac{1}{I(\theta)}$$

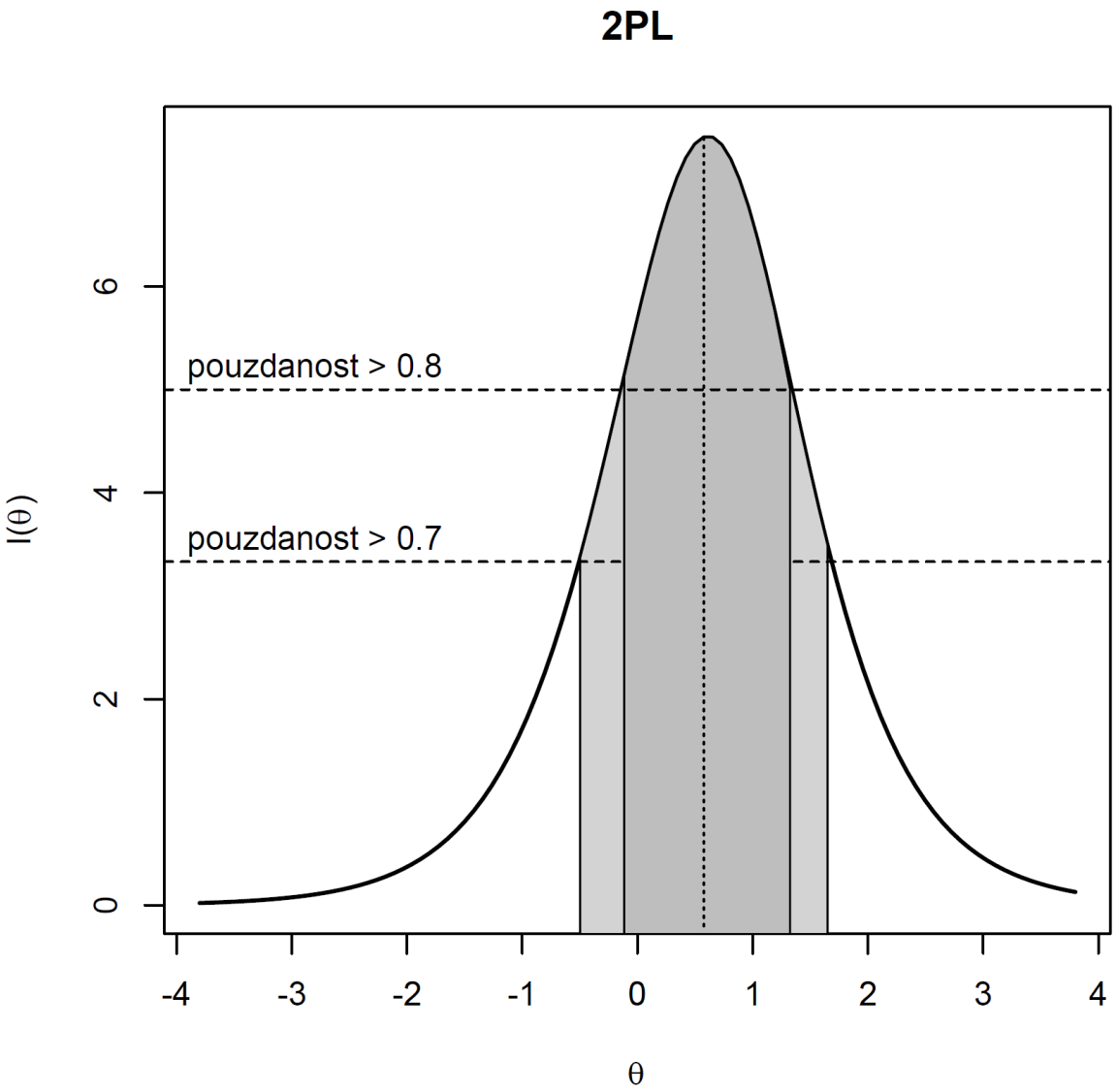
[ 91 ]

Treba imati u vidu da se ovde radi o *uslovnoj* (engl. *conditional reliability*) ili *marginalnoj pouzdanosti* (engl. *marginal reliability*) vezanoj za određeni nivo osobine. Na taj način postaje očigledna činjenica da psihološki merni instrumenti nisu jednako pouzdani u svim oblastima kontinuuma merene osobine.

Osim toga, na ovaj način imamo i neke smernice za interpretaciju vrednosti informativnosti. Npr. informativnost jednaka 5 odgovara pouzdanosti 1-1/5=1-0,2=0,80. Pouzdanosti od 0,7 odgovara informativnost od 3,33, a pouzdanosti od 0,9 odgovara informativnost koja je jednaka 10. U modelu EFPA (Evers i ostali, 2013) date su sledeće preporuke za tumačenje informativnosti testa:

- $I(\theta)$<3,33 neadekvatna
- 3,33≤ $I(\theta)$ <5,00 adekvatna
- 5,00 ≤ $I(\theta)$ <10,00 dobra
- $I(\theta)$ > 10 odlična.

Ovo je moguće i grafički predstaviti i na taj način steći utisak u kom delu kontinuuma merene latentne osobine će test biti najpouzdaniji ili bar dovoljno pouzdan. Isto nam govori i kriva informativnosti, ali svejedno, teško je zamisliti naučni rad ili priručnik za psihološki merni instrument koji ne pominje pouzdanost.



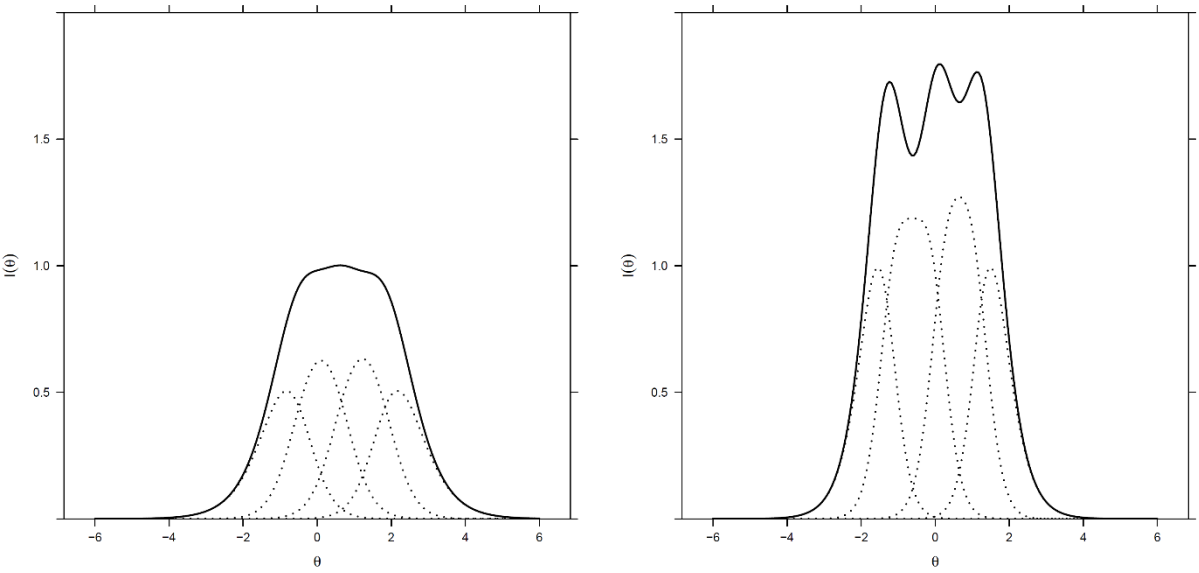

Slika 21 – Kriva informativnosti testa sa označenim nivoima pouzdanosti >0,7 i >0,8

Kod politomnih stavki svaka kategorija ima svoju informativnost i njome doprinosi informativnosti stavke i testa. Informativnosti kategorija se sumiraju i daju informativnost stavke, a suma informativnosti stavki daje informativnost testa. Informativnost stavke raste sa brojem kategorija odgovora, pa politomne stavke obezbeđuju veću informativnost od binarnih (Ostini & Nering, 2006). U jednoparametarskim politomnim modelima stavke sa jednakim brojem kategorija imaju jednaku *ukupnu informativnost*. Mogu se razlikovati po delu kontinuuma latentne osobine za koji su informativne (Dodd & De Ayala, 1994). U modelima koji sadrže parametar nagiba, on značajno određuje ukupnu informativnost kategorija i stavke.

Ako su lokacije kategorija simetrično raspoređene (npr. –1,5, –0,5, 0, 0,5, 1,5), maksimalna informativnost stavke biće na lokaciji težine stavke. Međutim, ako raspored lokacija nije simetričan (na primer

–2, –1,7, –1, 0,2, i 2) maksimum informativnosti biće pomeren u stranu gde su lokacije zbijenije.

U zavisnosti od razmaka kategorija, odnosno razmaka njihovih lokacija/težina, kriva informativnosti može imati različite oblike (de Ayala, 2022). Ako su kategorije zbijene, onda će kriva obično biti unimodalna sa vrhom na lokaciji bliskoj težini stavke (videti: Slika 22 – na slici su prikazane krive informativnosti kategorija i stavki sa po 4 kategorije dobijene na osnovu GRM). S druge strane, ako su kategorije razmaknute, kriva može biti zaravnjena i pokrivati širi raspon latentne osobine. Tada kategorije dodaju informativnost na različitim nivoima $\theta$ (DeMars, 2010). Ako su kategorije više razmaknute, onda kriva informativnosti stavke može biti i polimodalna (Ostini & Nering, 2006). Za merenje bi najbolje bilo kada bi kriva infor-




Slika 22 – Informativnost kategorija i stavke na primeru dve stavke sa po 4 kategorije (GRM)

mativnosti stavke ili testa bila ravna horizontalna linija na višem nivou (Baker, 2001). To bi značilo da su stavka ili test jednako i visoko precizni na svim nivoima merene osobine. To je u praksi retko ostvarljivo.

U nominalnom modelu informativnost stavke predstavljaju sumu proizvoda informativnosti svake od kategorija i verovatnoće biranja te kategorije (de Ayala, 2022). Kao i uvek, informativnost je funkcija nivoa $\theta$ i računa se na osnovu verovatnoće biranja određene kategorije za određeni nivo osobine.

$$I_j(\theta) = \sum_{h=1}^{k} I_{jh}(\theta) P_{jh}(\theta)$$

[ 92 ]

## *7.1 Primena informativnosti u konstrukciji testa*

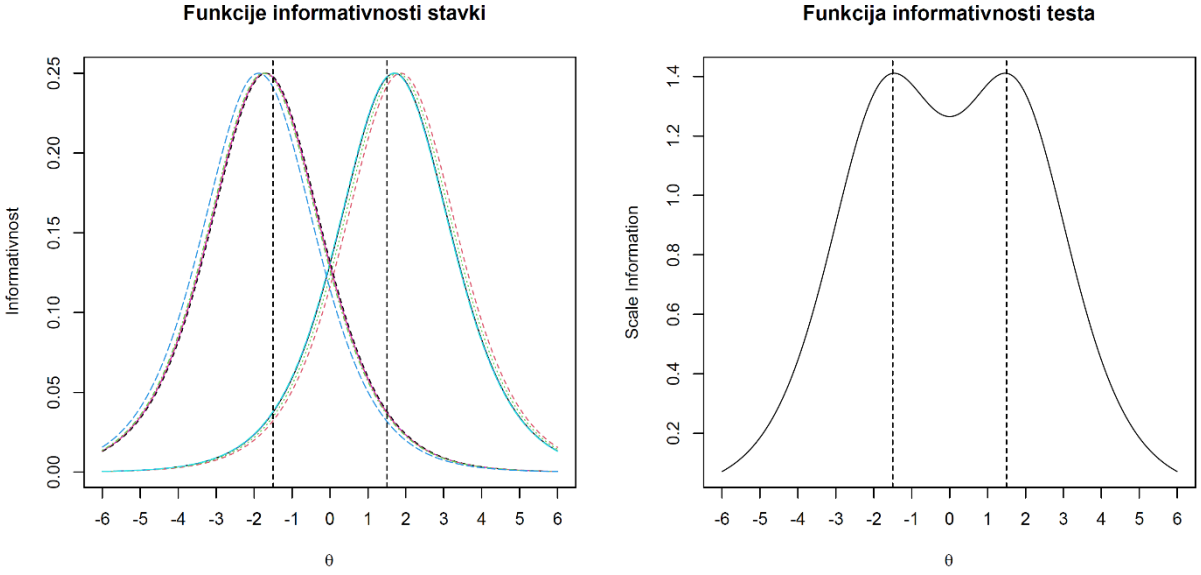
Informativnost stavki se koristi kao kriterijum za izbor stavki prilikom konstrukcije testa. Ono što bi trebalo da nam na početku bude poznato je ***ciljna kriva informativnosti testa*** (Hambleton & Swaminathan, 1985). Kako će ona izgledati zavisi od namene testa i cilja merenja.

Ako želimo da test bude primeren opštoj populaciji i precizan u širokom rasponu merene osobine, potrebno je da u njega uvrstimo informativne stavke različitih težina. Ciljna kriva informativnosti testa bi

bila platikurtična kriva koja pokriva širok raspon merene osobine. U ovom slučaju bi trebalo birati stavke u rasponu težina od –2 do +2 logita (eventualno –3 do +3) jer se očekuje da nivo osobine najvećeg broja ispitanika bude upravo u tom rasponu.

S druge strane, nekada nam cilj merenja ne obuhvata ceo kontinuum latentne osobine. Na primer, ukoliko želimo da konstruišemo test sposobnosti namenjen darovitim ispitanicima, zanimaće nas viši deo kontinuuma osobine. U tom slučaju želimo da informativnost bude najviša u delu kontinuuma iznad proseka, pa bi u test trebalo birati stavke koje su visoko informativne na nivoima od +1 do +3 logita.

Ako želimo da ispitanike podelimo u nekoliko grupa (recimo 3) po merenoj osobini, onda možemo želeti da nam test bude najinformativniji (najpreciziji) na nivoima latentne osobine koji predstavljaju kritične tačke za razdvajanje tih grupa. Recimo da su te kritične tačke –1,5 i 1,5 logit. Ciljna kriva informativnosti u ovom slučaju biće bimodalna sa vrhovima u kritičnim tačkama za razdvajanje tri grupe (Slika 23). U tom slučaju biraćemo stavke koje su visoko informativne u oblastima kritičnih tačaka. Na slici su kritične tačke označene vertikalnim isprekidanim linijama.




Slika 23 – Funkcije informativnosti stavki i testa namenjenog razdvajanju ispitanika na tri grupe (1PL model)

Ovakav način biranja stavki u test predstavlja vrstu adaptivnog testiranja u kojem se test prilagođava određenoj grupi ispitanika (Lord, 1977). Postupak je sledeći: 1) definisati ciljnu krivu informativnosti; 2) izabrati stavke koje će popuniti ciljnu funkciju informativnosti; 3) izračunati funkciju informativnosti testa i uporediti je sa ciljnom i 4) dodavati stavke dok funkcija informativnosti testa na bude odgovarala ciljnoj.

Stavke možemo birati po kriterijumu ***maksimalne informativnosti***. Na osnovu ovog kriterijuma u test biramo stavke koje imaju maksimalnu informativnost na lokacijama na kojima nam je to potrebno (Hambleton & Swaminathan, 1985). Npr. ako želimo test prosečne težine biramo stavke koje imaju najveću informativnost na nivou osobine od 0 logita. Ako postoji više tačaka na kontinuumu latentne osobine koje

želimo da pokrijemo, stavke možemo birati tako da u test uvrštavamo one čija je *prosečna informativnost za sve tačke od interesa* najviša. Npr. ako su nam bitne lokacije –1, +1, +2 logita, možemo za svaku stavku izračunati prosek njene informativnosti u ove tri tačke i izabrati onu koja ima najvišu vrednost. U ovakvom pristupu može se desiti da visoke prosečne vrednosti informativnosti većine stavki budu ostvarene na osnovu visoke informativnosti samo u nekoj od ovih tačaka. Recimo, većina izabranih stavki je najinformativnija u tački +1 logit. U tom slučaju druge dve lokacije mogu ostati nepokrivene.

Da bismo izbegli tu situaciju stavke možemo birati pristupom koji Hamblton i Svaminatan (Hambleton & Swaminathan, 1985) nazivaju pristupom ***gore–dole***. U ovom pristupu prvo se bira stavka najinformativnija u jednoj tački, zatim duga stavka koja je najinformativnija u drugoj tački, pa treća stavka u trećoj i tako u krug dok ne dođemo do željenog broja stavki ili ciljne krive informativnosti.

Prilikom adaptacije testa možemo se rukovoditi sličnim principom. Možemo izbacivati stavke informativne u delu kontinuuma osobine koji nam nije interesantan, dodavati nove tamo gde nam je potrebna veća preciznost i porediti krive informativnosti originalnog i modifikovanog testa. ***Kriva relativne efikasnosti*** je kriva odnosa informativnosti originalnog i modifikovanog testa po nivoima $\theta$ (videti odeljak 10.5.5.10.1). Na osnovu nje možemo videti da li su i koliko modifikacije testa promenile njegovu funkciju informativnosti, u kom smeru i u kom opsegu osobine.

Bejker (Baker, 2001) daje uputstva za kreiranje različitih vrsta testova na osnovu informativnosti stavki. Kada su u pitanju testovi koji se koriste za ***skrining*** potrebno planirati ih tako da budu najinformativniji na kritičnom nivou osobine koji predstavlja graničnu (engl. *cut-off*) vrednost. Naime testovi za skrining se kod nas nazivaju i *trijažnim testovima*. Služe za grubu klasifikaciju ispitanika u dve ili više grupa. Na primer, mogu se koristiti da ispitanike razvrstaju u grupu sa određenim psihičkim poremećajem i u grupu bez psihičkog poremećaja. Kod ovakvih testova kriva informativnosti bi trebalo da bude što strmija i da maksimum dostiže na graničnoj vrednosti, jer nam je tu preciznost najpotrebnija. Za to je potrebno birati stavke sa najvećom informativnošću, odnosno, najdiskriminativnije stavke čiji parametar težine se nalazi na lokaciji granične vrednosti nivoa osobine.

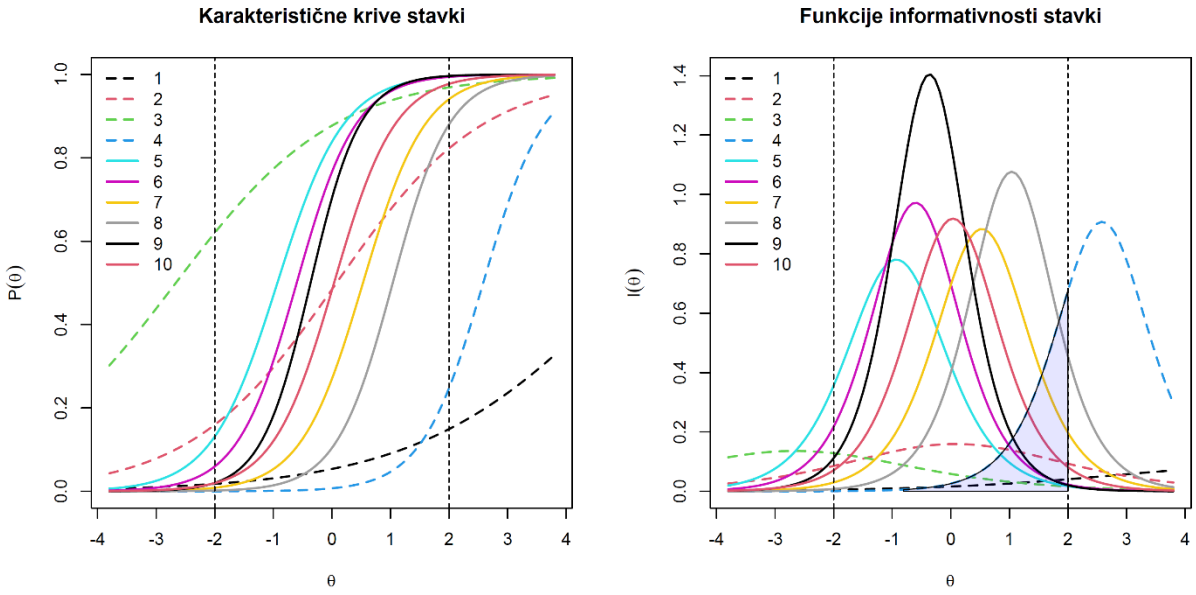
Kada su u pitanju testovi namenjeni merenju ***širokog opsega*** osobine, kriva informativnosti bi trebalo da dostiže maksimum u središnjoj tački željenog opsega (Baker, 2001). Takođe, ta kriva bi trebalo da bude platikurtična (idealno horizontalno linearna najvećim delom opsega). Da bismo ostvarili takvu krivu potrebno je da parametri odabranih stavki budu uniformno raspoređeni duž celog željenog opsega merenja. S druge strane, ako bismo želeli da ostvarimo horizontalno linearnu funkciju informativnosti, potrebno je birati stavke umerenih ili niskih diskriminativnosti, a distribucija parametara težina takvih stavki bi trebalo da ima oblik slova U, odnosno, da biramo stavke koje su po težinama bliske granicama željenog opsega merenja. Problem sa takvim načinom biranja je što su takve stavke generalno slabo informativne pa će merenje biti ujednačeno ali i nisko precizno u celom opsegu. Prema tome, ipak je praktičnije da kriva informativnosti bude platikurtična sa maksimumom na sredini željenog opsega merenja.

Treći tip testova su testovi za ***zašiljenom*** krivom informativnosti (Baker, 2001). I kod ovakvih

testova imamo željeni opseg merenja u kojem želimo da ono bude dovoljno precizno. Funkcija informativnosti je umereno strma, a maksimum se nalazi na središnjoj tački željenog opsega preciznog merenja. Za ovakav test bi trebalo birati najdiskriminativnije stavke čiji su parametri težine koncentrisani oko željene središnje tačke, ali ne blisko kao kod skrining testova.

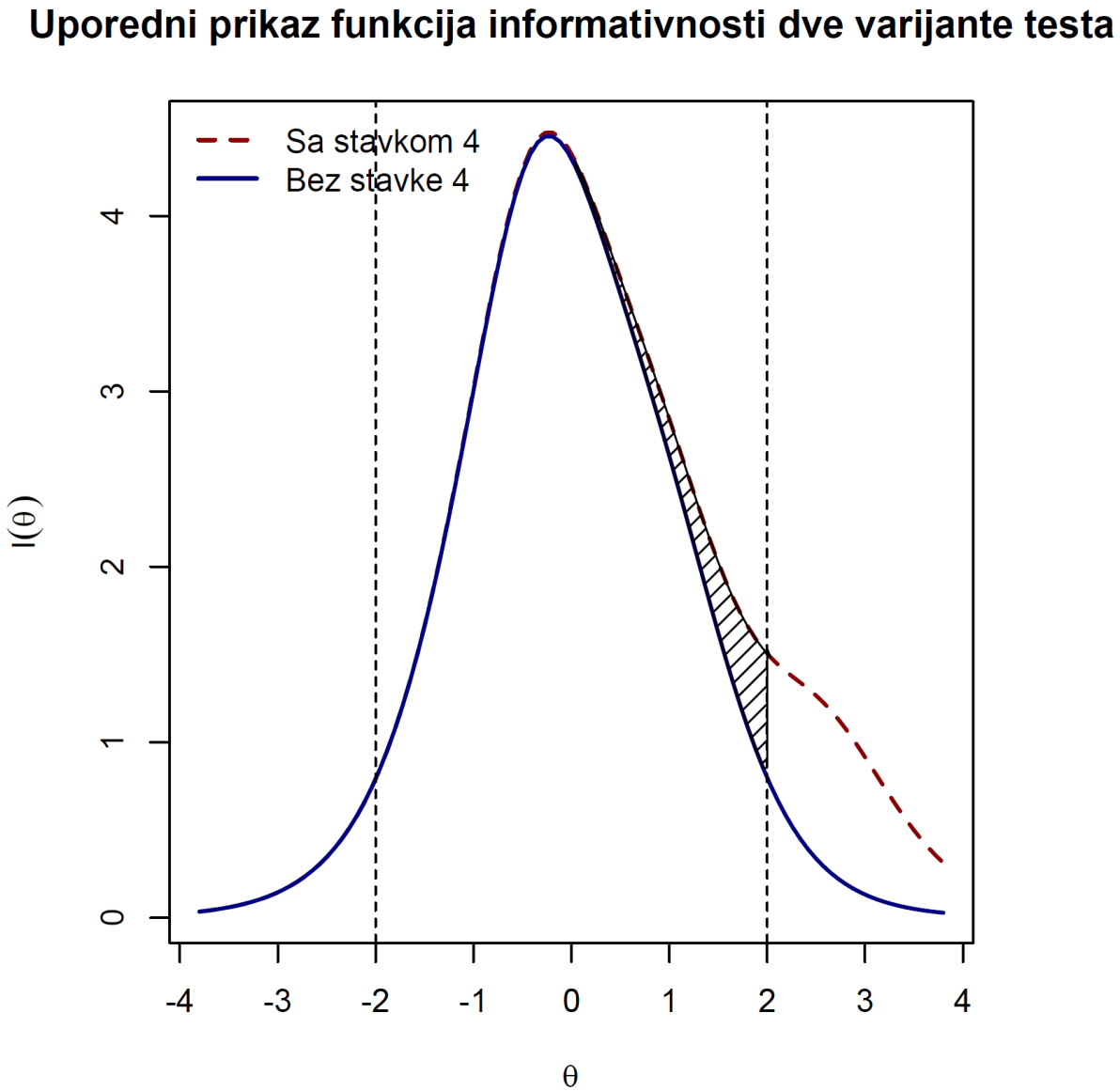
Prilikom planiranja krive informativnosti testa, trabalo bi voditi računa o ***primenjenom modelu*** i ***broju stavki***. U Rašovom ***modelu*** maksimalna informativnost stavke iznosi 0,25 i maksimalna informativnost testa može se dobiti kada se ta vrednost pomnoži brojem stavki (Baker, 2001). Nema parametra diskriminativnosti i on ne može uticati na informativnost testa. U dvo– i troparametarskim modelima parametar diskriminativnosti utiče na informativnost pri čemu, ako je veći od 1 maksimalna informativnost stavke biće veća od one u Rašovom modelu, a ako je manji – manja (pod pretpostavkom da je parametar pseudopogađanja 0). Viši parametar pseudopogađanja umanjuje informativnost pod uslovom da su parametri težine i diskriminativnosti nepromenjeni (Baker, 2001).

Kada je **broj stavki** u pitanju, može se reći da povećanje broja stavki utiče na nivo funkcije informativnosti. Veći broj stavki dovešće do više informativnosti, ali na kom delu kontinuuma zavisi od parametara težine stavki. To znači da na oblik krive informativnosti dodavanje novih stavki neće imati uticaja ako se distribucija parametara težine i diskriminativnosti ne razlikuje od distribucije istih parametara kod stavki koje su već prisutne u testu. Najbolje rešenje je da u testu imamo veći broj visokodiskriminativnih stavki čiji parametri težine pokrivaju opseg osobine u kojem želimo da vršimo merenje (Baker, 2001). Takva situacija je ilustrovana na narednoj slici (Slika 24). Ako zamislimo da je raspon latentne osobine u kojem nam je potrebno da merenje bude najpreciznije između –2 i 2 logita (označeno vertikalnim isprekidanim linijama), onda bi u test trebalo uvrstiti stavke označene punim linijama. One su visoko diskriminativne i težine su im u željenom opsegu osobine.




Slika 24 – Karakteristične krive i funkcije informativnosti stavke u 2PL modelu

Za razliku od njih stavke označene isprekidanim linijama su nisko diskriminativne, sa izuzetkom stavke 4. Ova stavka je visoko diskriminativna, ali njena težina izlazi iz željenog opsega. Ako pogledamo njenu funkciju informativnosti, vidimo da i ona doprinosi informativnosti testa u opsegu od –2 do 2 logita (površina označena plavom bojom), ali je najveći deo njene informativnosti izvan njega. Stavke označene isprekidanim linijama nam nisu potrebne u testu jer ne doprinose puno preciznosti merenja. Opet, kada je stavka 4 u pitanju, možemo razmotriti opciju da je zadržimo ili da je izbacimo. Informativnost testa u opsegu osobine između –2 i 2 logita nije puno viša ako je zadržimo (Slika 25).




Slika 25 – Funkcije informativnost testa sa i bez stavke 4

Na slici (Slika 25) isprekidanom linijom označena je funkcija informativnosti testa u kojem su zadržane stavke od 4 do 10, a punom linijom stavke od 5 do 10. Vidimo da je doprinos stavke 4 informativnosti testa u željenom opsegu merenja (oivičen vertikalnim isprekidanim linijama) relativno mali, tako da bismo je mogli izbaciti iz testa bez velikog gubitka u preciznosti merenja. Doprinos stavke 4 informativnosti testa u opsegu osobine od –2 do 2 logita prikazan je šrafiranom površinom.

Na principu biranja najinformativnijih stavki bazira se i ***računarsko adaptivno testiranje*** na osnovu IRT modela (engl. *Computerized Adaptive Testing* – CAT). U CAT ispitaniku se zadaje stavka koja je za njegov nivo osobine najinformativnija, odnosno najpreciznija. To je stavka koja je po parametru težine najbliža nivou osobine ispitanika[45]. Ako nam nije poznat nivo osobine ispitanika, obično se zadaje stavka prosečne težine u rasponu između –0,5 i +0,5 logita, a nekada se za početak biraju lake stavke da umanje testovnu anksioznost ispitanika (Embretson & Reise, 2000). Na osnovu odgovora ispitanika na stavku procenjuje se njegov nivo osobine i na osnovu tog, korigovanog nivoa $\theta_i$, bira se nova najinformativnija stavka. Ako je ispitanik odgovorio pogrešno, kao sledeća se bira lakša stavka, a ako je odgovorio tačno – teža. Nove stavke se zadaju dok se ne postigne dovoljna preciznost, odnosno dok se standardna greška merenja ne spusti ispod neke zadate vrednosti. Osim ovog, kao kriterijum za prekidanje testa može biti i broj stavki na koje je ispitanik odgovarao (Embretson & Reise, 2000). Ovaj pomoćni kriterijum je koristan kada standardnu grešku nije moguće spustiti ispod zadate vrednosti, ili kada je iz nekog razloga potrebno da svi ispitanici odgovaraju na jednak broj stavki.

Ukoliko posedujemo dovoljno veliki skup stavki poznatih mernih svojstava namenjen merenju neke latentne osobine, moguće je konstruisati test čija će nam merna svojstva biti poznata i pre nego što ga primenimo. Naime, rekli smo da informativnost testa predstavlja sumu informativnosti stavke po nivoima latentne osobine, te nešto ranije, da su procene parametara stavki u IRT modelima nezavisne od uzorka. Biranjem stavki na osnovu njihovih informativnosti mi projektujemo informativnost novog ili modifikovanog testa po nivoima osobine. Jezikom klasične testne teorije, projektujemo na kom nivou osobine će novi test biti najpouzdaniji (i koliko će biti).

Za konstrukciju testa prilagođenog ispitaniku ili grupi ispitanika bitan je skup stavki od kog krećemo. Za potrebe CAT-a formiraju se ***banke ajtema.*** To su skupovi stavki sa poznatim mernim karakteristikama (Embretson & Reise, 2000; Fajgelj, 2020; Millman & Arter, 1984). Da bi bilo moguće kreiranje testova željenih mernih karakteristika banka ajtema mora biti dovoljno velika i sadržavati dovoljan broj kvalitetnih stavki različitih nivoa težine.

## *7.2 Separacija i separaciona pouzdanost*

Iako je rečeno da u IRT modelima informativnost igra ulogu koju u KTT-u ima pouzdanost, pouzdanost je i ovde prisutna. U odeljku 7 već je prikazano izračunavanje *uslovne pouzdanosti* preko informativnosti. S druge strane, u Rašovom modelu koriste se dva pokazatelja pouzdanosti. To su *separacija (separacioni*

[45] U zavisnosti od primenjenog modela, na informativnost stavke neće uticati samo njena lokacija već i parametar diskriminativnosti (2P i 3P) i parametar donje asimptote (3P).

*odnos)* i *separaciona pouzdanost*. *Separacija* je broj statistički različitih stratuma koji se mogu identifikovati na osnovu testa i može se kretati u rasponu $0<G<\infty$. Može se definisati i kao reproducibilnost relativnih lokacija mera (de Ayala, 2022).

Separacija se označava sa $G$ i predstavlja odnos prave i pogrešne standardne devijacije (Fajgelj, 2020):

$$G = \frac{\sigma_a}{\sigma_{E\theta}}$$

[ 93 ]

gde je $\sigma_a$ prava standardna devijacija koja se računa kao:

$$\sigma_a = \sqrt{\sigma^2 - \sigma_{E\theta}^2}.$$

[ 94 ]

U prethodnom izrazu, $\sigma^2$ je ukupna varijansa, a $\sigma_{E\theta}^2$ kvadrat standardne greške merenja.

Rajt i Masters (Wright & Masters, 1982) stratume definišu kao statistički različite nivoe osobine koji se međusobno razlikuju za 3 greške merenja i kažu da se broj razdvojivih stratuma na osnovu separacije može izračunati na sledeći način (Wright & Masters, 1982):

$$H = (4G + 1)/3$$

[ 95 ]

Separacioni odnos i separaciona pouzdanost stoje u sledećem odnosu:

$$G = \frac{\sigma_a}{\sigma_{E\theta}} = \sqrt{\frac{r_{tt}}{1 - r_{tt}}}$$

[ 96 ]

pa se pouzdanost na osnovu separacije može izračunati kao:

$$r_{tt} = \frac{G^2}{1 + G^2} = \frac{\sigma_a^2}{\sigma^2}$$

[ 97 ]

Separaciona pouzdanost kao i pouzdanost u KTT-u predstavlja odnos prave i ukupne varijanse. Predstavlja formu pouzdanosti tipa interne konzistencije, analogna je i ima vrednost sličnu Kronbahovoj alfi (ili K–$R_{20}$) (Fajgelj, 2020). Pokazatelji iz klasične testne teorije po pravilu će biti nešto viši od separacione pouzdanosti pošto su u njihovo računanje uključeni svi skorovi, dok se prilikom računanja separacione pouzdanosti obično izbacuju ekstremni skorovi kod kojih standardna greška merenja ne može biti procenjena (Linacre, 1993).

Separaciona pouzdanost biće veća što je veća varijansa osobine u uzorku, što je test duži, stavke imaju više kategorija, a težina stavki usklađena sa nivoom osobine u uzorku. Ne zavisi od veličine uzorka i

od fita modela (Linacre, 1993).

Separacionu pouzdanost moguće je izračunati i za stavke, ali i za ispitanike (Fajgelj, 2020).

# 8 Diferencijalno funkcionisanje stavki i testa

Parafraziraćemo Angofa (Angoff, 1993) i ***diferencijalno funkcionisanje stavke*** (engl. *Differential Item Functioning* – DIF) definisati kao ispoljavanje drugačijih statističkih svojstava stavke kod različitih manifestnih grupa koje su ujednačene po meri osobine. Pod manifestnim grupama obično se podrazumevaju grupe formirane po određenim demografskim karakteristikama, a najčešće su to pol, maternji jezik, nacionalnost. U svakom slučaju, radi se o trećim varijablama koje nisu deo merenog konstrukta.

DIF se može definisati i kao razlika karakterističnih kriva (iste) stavke u različitim grupama (populacijama) (Embretson & Reise, 2000). To znači da ispitanici sa jednakim nivoom osobine, a koji pripadaju različitim grupama, postižu različite skorove na stavci (odnosno, imaju različite modelske verovatnoće da tačno odgovore na stavku ili da odgovore u istoj kategoriji skora na politomnim stavkama).

U IRT pristupu, DIF predstavlja narušavanje pretpostavke invarijantnosti parametara stavki u različitim uzorcima, odnosno narušavanje specifične objektivnosti (de Ayala, 2022).

Diferencijalno funkcionisanje stavki može, ali ne mora, dovesti i do toga da ispitanici sa istim nivoom osobine (iz različitih demografskih grupa) imaju različite skorove *na testu*. Drugim rečima, može ali ne mora dovesti do ***diferencijalnog funkcionisanja testa*** (engl. *Differential Test Functioning* – DTF).

Pošto se merenje vrlo često vrši sa nekim praktičnim ciljem (npr. selekcija), ako na sve ispitanike, bez obzira kojoj grupi pripadali, primenjujemo iste procene parametara modela može doći do sistematskog potcenjivanja ili precenjivanja jedne grupe, a samim tim i do pogrešnih odluka koje zasnivamo na rezultatima takvog merenja (Embretson & Reise, 2000).

Grupe koje poredimo prilikom postupaka za ispitivanje DIF-a najčešće se nazivaju *referentna* i *fokalna* (de Ayala, 2022; Embretson & Reise, 2000; Fajgelj, 2020). Referentna grupa je ona na kojoj su stavke kalibrisane, odnosno, na kojoj su određeni parametri stavki.

DIF je u suštini pitanje jednakosti i validnosti merenja. Naime, ako stavke različito funkcionišu u različitim grupama ispitanika formiranim po trećim varijablama, to znači da na meru osobine utiče treća varijabla koja nije deo merenog konstrukta (Chen & Revicki, 2014; Fajgelj, 2020). Kada proveravamo postojanje DIF-a, treće varijable koje ćemo uzeti u obzir su obično one za koje je poznato da su povezane sa sprovođenjem merenja uopšte ili sa konkretnim predmetom merenja.

DIF se u IRT modelima može ispoljiti kao razlika u parametrima težine stavke ili kao razlika u parametrima nagiba u dve grupe.

Kada postoji razlika u parametrima težine, govorimo o ***uniformnom DIF-u***. Naziva se uniformnim zato što jedna od dve grupe ima veću verovatnoću tačnog odgovora duž celog kontinuuma $\theta$ i ta razlika u

verovatnoćama je manje-više jednaka duž celog kontinuuma latentne osobine.

Kada postoji razlika u parametrima diskriminativnosti (nagiba) stavki, govorimo o ***neuniformnom DIF-u***. U slučaju postojanja neuniformnog DIF-a *razlika* u verovatnoćama tačnog odgovora na stavku za dve grupe menja se u različitim delovima kontinuuma $\theta$. U tom slučaju može se desiti i da se KKS ukrštaju pa će jedna grupa imati veću verovatnoću tačnog odgovora u jednom delu kontinuuma $\theta$, a druga u drugom.

Da bismo utvrdili da li postoji DIF koristeći IRT pristup, potrebno je da uporedimo karakteristične krive stavki u dve grupe. Za to je potrebno proceniti model u svakoj od grupa. Tako procenjene parametre mogli bismo direktno porediti pod pretpostavkom da je distribucija osobine u dve grupe ista, ali to ne možemo uvek znati. Procene parametara u IRT modelima su nezavisne od uzorka, međutim, parametri koje dobijemo ne moraju biti nužno na istoj skali. Skala latentne osobine prilikom svake procene parametara modela arbitrarno se postavlja tako da ima aritmetičku sredinu 0 i standardnu devijaciju 1. Svrha ovakvog definisanja skale je identifikacija modela (Embretson & Reise, 2000). To i dalje ne znači da su dve skale iste.

Da bi se mogle porediti KKS iz dve grupe skale se moraju ujednačiti, što se čini procedurama vezivanja (engl. *linking*) (Embretson & Reise, 2000). Vezivanje se može obaviti na više različitih načina, a ovde će biti opisana logika jedne od jednostavnijih procedura.

Pre svega, da bi se parametri stavki u dve grupe mogli vezati potrebno je da dve grupe odgovaraju na potpuno isti skup stavki ili da postoji bar nekoliko stavki na koje su obe grupe odgovarale. Ove zajedničke stavke nazivaju se ***sidra*** i služe da „usidre“ skalu. Za sidrenje, prilikom ispitivanja DIF-a potrebno je da sidra budu *proverene stavke* za koje znamo da jednako funkcionišu u obe grupe.

Ako su obe grupe odgovarale na isti skup stavki, potrebno je proceniti parametre u obe grupe i izračunati njihove aritmetičke sredine i standardne devijacije. Recimo da su $\bar{b}_R$ i $\bar{b}_F$ prosečni parametri težine, a $\sigma_R$ i $\sigma_F$ standardne devijacije tih parametara u referentnoj i fokalnoj grupi. Na osnovu procena parametara u dve grupe računaju se konstante za vezivanje ($x$ i $y$), odnosno, vrednosti na osnovu kojih ćemo transformisati parametre dobijene na jednom uzorku na skalu parametara dobijenu na drugom uzorku (Embretson & Reise, 2000).

Da bismo metriku jedne grupe (recimo fokalne – F) prebacili u metriku druge grupe (referentne – R) potrebno je da parametre fokalne grupe pomnožimo i dodamo im konstante za vezivanje, kao što je prikazano u narednom izrazu (Embretson & Reise, 2000):

$$b_F^* = xb_F + y \qquad [98]$$

gde je:

$$x = \frac{\sigma_R}{\sigma_F} \qquad [99]$$

i:

$$y = \bar{b}_R - x(\bar{b}_F)$$

[ 100 ]

U situaciji kada dve grupe nisu radile isti test potrebno je da postoji set stavki za sidrenje na koji su odgovarale obe grupe. Recimo da imamo dve grupe ispitanika R i F. Set stavki koji su radile obe grupe označićemo sa $S_S$, set stavki na koji su odgovarali samo ispitanici iz referentne grupe R sa $S_R$, a set na koji su odgovarali samo ispitanici iz fokalne grupe F sa $S_F$. U toj situaciji potrebno je proceniti parametre zasebno za sva tri seta, a zatim parametre setova $S_R$ i $S_F$ prebaciti na metriku seta $S_S$ na gore opisani način.

Svaka procedura za proveru DIF-a mora proći i ovu fazu vezivanja, pa analiza DIF-a u suštini predstavlja dvostepeni proces (Embretson & Reise, 2000).

Postoji više načina za proveru diferencijalnog funkcionisanja stavki, ali tri najčešća su upotreba Mantel-Hanselovog (Mantel-Haenszel) statistika, logistička regresija i metod zasnovan na količniku verodostojnosti (*likelihood ratio*) IRT modela. Od tri navedena metoda samo poslednji je zasnovan na IRT pristupu, mada postoje hibridni modeli zasnovani na logističkoj regresiji i IRT modelima (S. W. Choi i ostali, 2016).

## *8.1 Mantel-Hanselov postupak*

Mantel-Hanselov hi-kvadrat (Holland & Thayer, 1986) je neparametarska procedura za ispitivanje DIF-a. Zasniva se na hi-kvadrat testu po grupama ispitanika koje su ujednačene po skoru na testu, a namenjen je binarno skorovanim stavkama.

Procedura predviđa da se prvo formiraju grupe ispitanika koje imaju iste ukupne skorove na testu. Umesto skora na testu može se koristiti i neka njegova modifikacija.

Za svaki nivo ukupnog skora $u$ (osim savršenih) formira se tabela kontingencije 2x2 (Tabela 3). Za savršene skorove 0 i $m$ ($m$ je broj stavki) nema smisla praviti tabele kontingencije jer u grupama savršenih skorova postoji samo jedna vrsta odgovora (0 kod ukupnog skora 0, i 1 kod ukupnog skora $m$). Takve tabele bi imale samo jednu kolonu. Ako ukupnih skorova ima $U$, onda se za svaku stavku formira $U$–2 tabela kontingencije.

Tabela 3 – Tabela kontingencije za jedan nivo ukupnog skora i jednu stavku

| | Odgovor na stavku | | |
|---|---|---|---|
| Grupa | 0 | 1 | Ukupno |
| Referentna | $b_u$ | $a_u$ | $n_{Ru}$ |
| Fokalna | $d_u$ | $c_u$ | $n_{Fu}$ |
| Ukupno | $n_{0u}$ | $n_{1u}$ | |

Napomena: $u$ je indeks grupe po ukupnom skoru kojih ima $U$, a $n$ je broj ispitanika u grupi $u$, indeksi R i F označavaju referentnu i fokalnu grupu.

Na osnovu ovakvih tabela računa se Mantel-Hanselov hi-kvadrat (prema: de Ayala, 2022):

$$MH\chi^2 = \frac{\left(\left|\sum_{u=1}^{U-1}\left(a_u - E(a_u)\right)\right| - 0{,}5\right)^2}{\sum_{u=1}^{U-1} var(a_u)}$$

[ 101 ]

gde je –0,5 Jejtsova (Yates) korekcija, a $E(a_u)$:

$$E(a_t) = \frac{n_{Ru}n_{1u}}{n_u}$$

[ 102 ]

a $var(a_u)$:

$$var(a_u) = \frac{n_{Ru}n_{Fu}n_{1u}n_{0u}}{n_u^2(n_u - 1)}$$

[ 103 ]

MH $\chi^2$ je test uslovne nezavisnosti i testira nultu hipotezu da nema razlike u frekvencijama različitih skorova u referentnoj i fokalnoj grupi kada se kontroliše ukupan skor na testu. Ima hi-kvadrat distribuciju za 1 stepen slobode (de Ayala, 2022).

Značajan MH $\chi^2$ nam ne govori kolika je razlika između dve grupe. U tu svrhu računa se odnos šansi za svih $U$–1 tabela kontingencije:

$$\alpha_{MH} = \frac{\sum_{u=1}^{U-1}\left(\frac{a_u d_u}{n_u}\right)}{\sum_{u=1}^{U-1}\left(\frac{b_u c_u}{n_u}\right)}$$

[ 104 ]

$\alpha_{MH}$ je odnos (racio nivo) i može uzimati vrednosti od 0 do ∞. Ako je jednak 1, onda su odnosi šansi za dve grupe jednaki i nema DIF-a za posmatranu stavku. Ako je veći od 1, onda pripadnici referentne grupe imaju više uspeha na posmatranoj stavci, a ako je manji od 1, onda su pripadnici fokalne grupe bolji. Rekli smo da je $\alpha_{MH}$ na racio nivou, pa tako vrednost $\alpha_{MH}$=3 znači da ispitanici iz referentne grupe imaju 3 puta veće šanse da odgovore tačno na posmatranu stavku od ispitanika iz fokalne grupe koji imaju isti nivo osobine (ukupni skor). Ako je vrednost statistika $\alpha_{MH}$=0,5, to bi značilo da su šanse ispitanika iz referentne grupe da odgovore tačno na posmatranu stavku upola manje nego šanse ispitanika iz fokalne grupe (sa istim nivoom osobine). Iz perspektive članova fokalne grupe to bi se moglo reći ovako: šanse ispitanika iz fokalne grupe da tačno odgovore na stavku su 2 puta (1/0,5) veće nego šanse ispitanika iz referentne grupe.

Osim $\alpha_{MH}$ pokazatelja za opisivanje veličine DIF-a koristi se njegov prirodni logaritam koji se označava sa $\beta_{MH}$. Ovaj pokazatelj može imati vrednosti od –∞ do +∞. U tom slučaju 0 znači da DIF ne postoji, negativne vrednosti da stavka favorizuje fokalnu, a pozitivne da favorizuje referentnu grupu.

Holand i Tajer (Holland & Thayer, 1986) su predložili $\Delta_{MH}$ pokazatelj koji dalje transformiše $\alpha_{MH}$:

$$\Delta_{MH} = -2{,}35\ln(\alpha_{MH}) = -2{,}35\beta_{MH}$$

[ 105 ]

Ovako ocenjen DIF razvrstava se u tri kategorije po veličini efekta (Magis i ostali, 2010):

A) $|\Delta_{MH}| < 1$: zanemarljiv

B) $1 \leq |\Delta_{MH}| \leq 1{,}5$: umeren

C) $1{,}5 < |\Delta_{MH}|$: veliki

Negativne vrednosti ukazuju da pripadnici fokalne grupe imaju *manju* verovatnoću da odgovore tačno na posmatranu stavku od pripadnika referentne grupe sa istim nivoom osobine i obrnuto (Wu i ostali, 2016).

Iako jednostavan Mantel-Hanselov postupak se dobro pokazao, ali je osetljiv na veličinu uzorka. Kada je $n$<500 propušta da otkrije DIF u više od 50% slučajeva, a nije efikasan kod neuniformnog DIF-a (Fajgelj, 2020). Iz izraza [ 101 ] možemo zaključiti da će MH statistik biti pozitivan ukoliko referentna grupa postiže više tačnih odgovora na stavku nego što je to očekivano, a negativan ako postiže manji broj. Kod neuniformnog DIF-a kada se KKS za dve grupe ukrštaju, jedna grupa će imati veću verovatnoću tačnog odgovora u nižem delu kontinuuma $\theta$ (niži ukupni skorovi), a manju u višem delu (viši ukupni skorovi). U tom slučaju MH statistici bi mogli na nižim skorovima biti negativni, a na višim pozitivni. Prilikom sumiranja takve vrednosti bi se potirale pa bi ukupna vrednost MH $\chi^2$ bila bliska 0 i ne bi bila statistički značajna, a neuniformni DIF bi prošao neopaženo. Zbog toga se Mantel-Hanselov postupak ne preporučuje za ispitivanje neuniformnog DIF-a. S druge strane, značajan MH $\chi^2$ ne znači nužno da je DIF uniforman (de Ayala, 2022).

Pored opisanog postupka za binarno skorovane stavke razvijen je i postupak namenjen politomnim stavkama i za više od jedne fokalne grupe (French i ostali, 2019).

## *8.2 Postupak zasnovan na IRT i LR testu*

Postupak zasnovan na IRT modelima i količniku verodostojnosti (LR testu – $G^2$) predložili su Tisen, Stajnberg i Vainer (Thissen i ostali, 1993).

U ovoj proceduri prvo je potrebno proceniti iste IRT modele za dve grupe sa ograničenjem da parametri stavki moraju biti jednaki za obe. Zatim se procenjuju modeli koji za jednu po jednu stavku dozvoljavaju da parametri budu različiti po grupama. Porede se logaritmi verodostojnosti ograničenog i modela sa slobodnim parametrima i ako je verodostojnost značajno bolja u modelu sa slobodnim parametrima, to ukazuje da stavka za koju su u tom krugu analize oslobođeni parametri različito funkcioniše u dve grupe[46]. Nulta hipoteza koju testira $G^2$ u ovom slučaju je da nema razlike u procenama parametara u dve grupe.

Ako postoji DIF, na osnovu parametara stavki možemo videti koja grupa je favorizovana.

Kao što smo ranije rekli, da bismo mogli porediti parametre stavki procenjene u dve grupe potrebno je obaviti vezivanje, odnosno, postaviti parametre na istu skalu. Da bismo to mogli da učinimo potrebne

[46] Procedura može biti i obrnuta. Može se krenuti od potpuno slobodnih modela za obe grupe, a zatim nastaviti sa ograničavanjem parametara po stavkama.

su nam stavke za sidrenje. Pošto su obe grupe (referentna i fokalna) odgovarale na iste stavke, možemo kao sidra odabrati neke od stavki. To možemo učiniti na osnovu prethodnih saznanja ili ekspertskog mišljenja (Embretson & Reise, 2000), a možemo proceniti i na osnovu podataka.

Ako koristimo procenu na osnovu podataka, postupak analize DIF se obavlja iterativno. U prvom krugu koristi se gore opisani postupak i procenjuje se DIF svake stavke koristeći sve ostale kao sidra. Za svaku stavku računa se $G^2$ i ako je statistički značajan, smatramo da stavka pokazuje DIF. Ovaj postupak se često označava sa AOAA (engl. *All Others as Anchors*). Nedostatak ovog postupka je što se i stavke koje pokazuju znake DIF-a koriste kao sidra prilikom analize drugih stavki (Meade & Wright, 2012).

Za rešavanje ovog problema predloženo je više načina na koje se utvrđuje set sidrenih stavki.

Prvi smo već naveli, a zasniva se na teoriji ili prethodnim istraživanjima. Kao sidrene stavke koristićemo one za koje na osnovu teorije smatramo da su invarijantne po posmatranim grupama.

Najčešće korišćeni metod je prečišćavanje (engl. *purification*). Radi se o iterativnom postupku gde se u prvom koraku koristi AOAA pristup, pa ako se na taj način detektuje DIF kod nekih stavki, u drugom koraku se sve stavke kod kojih nije detektovan DIF koriste kao sidra i DIF ponovo procenjuje. Drugi korak se ponavlja sve dok se otkrivaju nove stavke koje diferencijalno funkcionišu. Ovaj način daje bolje rezultate od AOAA pristupa, ali nekada ne rezultira stabilnim skupom stavki koje pokazuju invarijantnosti ili DIF (Meade & Wright, 2012). Inače, ovakvi postupci često su deo softverskih rešenja koja DIF otkrivaju po drugim principima (MH ili logistička regresija).

Sledeći način je procedura koja se odvija u dva koraka. U prvom koraku se koristi AOAA pristup kako bi se utvrdilo koje stavke su sumnjive na DIF, a koje ne. Za ovu proceduru mogu biti korišćeni svi ovde navedeni postupci za ispitivanje DIF-a ($G^2$, MH, logistička regresija), a kao sidra se mogu koristiti sve ili samo neke stavke koje u toj prvoj fazi ne pokazuju znake različitog funkcionisanja po grupama.

Lopez Rivas i saradnici (Lopez Rivas i ostali, 2009) predlažu da se kao sidra biraju stavke koje nemaju značajan DIF u prvom krugu analize, a imaju najviše parametre diskriminativnosti. Prema ovim autorima, korišćenje jednog visoko diskriminativnog sidra rezultiralo je povećanjem snage $G^2$ testa za ispitivanje DIF-a kod binarnih i politomnih stavki. U slučaju da je veličina efekta DIF-a mala, bolji rezultati su se postizali sa više sidrenih stavki. Preporuka je da se odabere maksimum 5 sidrenih stavki (Meade & Wright, 2012). Kao interesantan nalaz Mid i Rajt (Meade & Wright, 2012) navode da sidra po lokaciji bliska prosečnom nivou osobine u fokalnoj grupi unapređuju otkrivanje DIF-a kod binarno skorovanih stavki. Takođe, nalaze da uključivanje stavki koje pokazuju DIF u sidra smanjuje snagu $G^2$ i povećava grešku tipa I.

Vudsova (Woods, 2009) predlaže da se kao kriterijum za izbor sidrenih stavki koristi vrednost $G^2$ iz AOAA koraka. Pod pretpostavkom jednakog broja kategorija stavki (odnosno parametara lokacije), one sa neznačajnim i nižim vrednostima $G^2$ bi trebalo da pokazuju manju količinu DIF-a. Međutim Mid (Meade, 2010) je pokazao da to ne mora biti slučaj i sugeriše da se stavke koje će se koristiti kao sidra biraju na osnovu pokazatelja veličine efekta koje je sam predložio. Radi se o pokazateljima zasnovanim na očekivanim skorovima na stavkama i testu. Kao pokazatelj veličine efekta koji najviše obećava, Mid i Rajt (Meade & Wright, 2012) izdvajaju *apsolutnu razliku skora na stavci u uzorku* (engl. *Unsigned Item Difference in the*

*Sample* – UIDS). Ovaj pokazatelj se računa na ispitanicima iz fokalne grupe. Potrebno je izračunati *očekivane skorove na stavci* na osnovu $\theta$ ispitanika i parametara stavki dobijenih na modelu za fokalnu grupu. Takođe, potrebno je izračunati iste takve očekivane skorove koristeći parametre stavki iz modela za referentnu grupu. UIDS predstavlja prosečnu apsolutnu razliku skorova ispitanika dobijenih na ta dva načina.

Osim ovog pokazatelja Mid (Meade, 2010) navodi i SIDS (engl. *Signed Item Difference in the Sample*). Radi se o *prosečnoj razlici skora na stavci na uzorku* koja vodi računa o predznaku. Računa se na isti način kao UIDS, s tim da se vodi računa o predznacima razlika. Razlike suprotnog predznaka mogu se potirati i u slučaju neuniformnog DIF-a prosečna razlika može biti bliska 0. Zato SIDS koji je značajno manji od UIDS može biti znak neuniformnog DIF-a. Prethodni pokazatelji izraženi su u testnim bodovima (Meade, 2010).

*Standardizovana razlika očekivanog skora* (engl. *Expected Score Standardized Difference* – ESSD) dobija se tako što se SIDS podeli zajedničkom standardnom devijacijom parametara stavki fokalne i parametara stavki referentne grupe. ESSD je izražen u jedinicama standardnih devijacija i interpretira se kao Koenovo (Cohen) d:

- ESSD < 0,2 – zanemarljiva veličina efekta
- 0,2 ≤ ESSD < 0,5 – mali efekat
- 0,5 ≤ ESSD < 0,8 – srednji efekat
- 0,8 ≤ ESSD – veliki efekat (Meade, 2010)

Slični pokazatelji postoje i za test u celini i biće opisani kasnije (odeljak 10.6.3).

## *8.3 Postupak zasnovan na logističkoj regresiji*

Kao ni Mantel-Hanselova procedura, ni ispitivanje DIF-a upotrebom logističke regresije nije zasnovano na IRT modelima, mada se u poslednje vreme kombinuju logistička regresija i IRT u tzv. hibridnom pristupu (S. W. Choi i ostali, 2016).

Logistička regresija je postupak koji se koristi za predviđanje (verovatnoća) binarnih ili politomnih ishoda na osnovu skupa prediktora koji mogu biti bilo kog nivoa merenja. Za ispitivanje DIF-a kod binarno skorovanih stavki koristi se binarna logistička, a kod politomnih, ordinalna logistička regresija.

Predviđani ishod je odgovor na stavku, a nominalni prediktor je pripadnost referentnoj ili fokalnoj grupi. Pored grupne pripadnosti, kao prediktor se može uvrstiti i nivo osobine ispitanika, te interakcija nivoa osobine i grupne pripadnosti.

Logiku ovog postupka objasnićemo na binarnom logističkom modelu. Prilikom provere DIF-a procenjuju se tri modela. Najjednostavniji model je (prema: de Ayala, 2022):

$$\hat{z} = \beta_0 + \beta_1 \Lambda \qquad [106]$$

gde je $\hat{z}$ odgovor ispitanika na stavku, $\beta_0$ odsečak, a $\beta_1$ regresioni koeficijent za nivo osobine $\Lambda$ (može biti IRT mera ili ukupan skor).

Ovaj model predviđa odgovor ispitanika u zavisnosti od nivoa osobine koji može biti izražen kao

sumacioni skor ili kao parametar ispitanika iz nekog IRT modela. Ovaj model ćemo zvati *Model 1.*

Sledeći model je nešto složeniji i kao kategorijalni prediktor uključuje grupnu pripadnost (prema: de Ayala, 2022):

$$\hat{z} = \beta_0 + \beta_1 \Lambda + \beta_2 G$$

[ 107 ]

gde je $G$ grupna pripadnost, a $\beta_2$ regresioni koeficijent za grupnu pripadnost. *Model 2* predviđa odgovor ispitanika u zavisnosti od nivoa osobine $\Lambda$ i grupne pripadnosti $G$.

I na kraju procenjuje se i puni model logističke regresije – *Model 3* (prema: de Ayala, 2022):

$$\hat{z} = \beta_0 + \beta_1 \Lambda + \beta_2 G + \beta_3 G\Lambda$$

[ 108 ]

gde je $\beta_3$ regresioni koeficijent za efekat interakcije pripadnosti grupi i nivoa osobine. Model osim dva glavna efekta (nivoa osobine i grupne pripadnosti) procenjuje i efekat interakcije.

Rekli smo da je *Model 3* puni model. *Model 2* predstavlja redukciju *Modela 3* i njegov poseban slučaj kada je $\beta_3$ ograničen na 0. Ovo ograničenje praktično znači da je odnos nivoa osobine i odgovora ispitanika isti u dve grupe. Kada je $\beta_3$=0 to znači da je nagib isti u dve grupe, odnosno, da ne postoji neuniformni DIF.

*Model 1* je redukovani *Model 2* kada je $\beta_2$ ograničen na 0. Ovo ograničenje znači da skorovi ispitanika na stavci ne zavise od grupne pripadnosti, odnosno, da ne postoji uniformni DIF.

U sva tri modela je kao prediktor uključen nivo osobine, što znači i da je statistički kontrolisan.

Referentna grupa u odnosu na koju se interpretiraju koeficijenti prediktora je obično i referentna grupa analize DIF-a.

Postoji više načina utvrđivanja DIF-a na osnovu pomenuta tri modela. **Prvi način** zasniva se na značajnosti razlika u devijaciji (-2logLik) tri navedena modela. Modeli se porede LR testom ($\Delta G^2$).

Ako $\Delta G^2$ između *Modela 3* i *Modela 2* nije statistički značajan, to znači da uvođenje efekta interakcije grupne pripadnosti i nivoa osobine ne popravlja model. U tom slučaju taj efekat nam je nepotreban, odnosno, možemo reći da ne postoji neuniformni DIF. Suprotno, ako je ovaj $\Delta G^2$ značajan, odbacujemo nultu hipotezu da ne postoji neuniformni DIF (možemo reći da postoji).

Slično, ako $\Delta G^2$ između *Modela 2* i *Modela 1* nije značajan, onda to znači da uvođenje efekta grupne pripadnosti ne popravlja *Model 1*. U tom slučaju nam efekat grupne pripadnosti nije potreban, odnosno, možemo reći da ne postoji uniformni DIF. Suprotno, ako je $\Delta G^2$ značajan, odbacujemo nultu hipotezu da ne postoji uniformni DIF.

**Drugi način** je da se nakon utvrđivanja značajnosti razlika između devijacija tri modela, uporede i pseudo $R^2$ za modele. Kao mera veličine efekta posmatra se promena u pseudo $R^2$ pokazateljima (Zumbo, 1999):

$$\Delta R^2 = R_P^2 - R_R^2$$

[ 109 ]

gde indeks *P* označava puni, a *R* redukovani model. Najčešće se koriste Nagelkirkov pseudo $R^2$ (Nagelkerke, 1991) ili $R^2$ na bazi ponderisanih najmanjih kvadrata (de Ayala, 2022).

Veličina efekta u formi $\Delta R^2$ može se klasifikovati na sledeći način (Jodoin & Gierl, 2001):

- $\Delta R^2 < 0{,}035$ – zanemarljiv efekat
- $0{,}035 \leq \Delta R^2 < 0{,}070$ – srednji efekat
- $\Delta R^2 > 0{,}070$ – veliki efekat

Ove veličine efekta se označavaju i oznakama A, B i C (respektivno) kao kod Mantel-Hanselovog postupka. U opticaju je i podela Zumba i Tomasa (Magis i ostali, 2010; Zumbo, 1999): $\Delta R^2$ manja od 0,13 – zanemarljiv efekat, između 0,13 i 0,26 – srednji, preko 0,26 – veliki. Oznake su iste. Da bi stavka bila klasifikovana kao stavka koja pokazuje DIF, potrebno je da *p*–vrednost hi–kvadrata za model koji uključuje grupnu pripadnost i interakciju ne bude veća od 0,01 i da $\Delta R^2$ bude veća od 0,035 (Gelin & Zumbo, 2003).

**Treći način** (Crane i ostali, 2004) je da se neuniformni DIF detektuje na osnovu LR testa (kao u prvom pristupu), dok se za uniformni DIF posmatra uticaj glavnog efekta grupe na razlike u $\beta_1$ koeficijentu u modelima 1 i 3. Razlika veća od 5% ukazuje na značajan uniformni DIF.

Vratimo se na postupak *prečišćavanja* koji smo pominjali kod Mantel-Hanselove procedure, a koristi se i u postupcima zasnovanim na logističkoj regresiji. Prečišćavanje je često bilo predmet kritike. Neki autori su se zalagali da stavke kod kojih postoji sumnja na DIF budu isključene iz izračunavanja skora po kojem se grupe uparuju (Zumbo, 1999), a neki su opet bili protiv toga jer takav postupak dovodi do gubitka informacija (Reise i ostali, 1993). Za prevazilaženje ovog problema kod metoda analize DIF-a koje za procenu nivoa osobine koriste IRT mere i logističku regresiju, Krejn i saradnici (Crane i ostali, 2006) predložili su proceduru koja u proceni nivoa osobine koji služe za uparivanje grupa koriste i stavke kod kojih je uočen DIF. Tačnije, prilikom procene nivoa osobine ispitanika za stavke kod kojih se sumnja na DIF uzimaju se parametri koji su specifični za grupe, dok se za ostale koriste zajednički parametri za obe grupe[47]. Tako korigovane mere ispitanika koriste se za novi krug analize DIF-a upotrebom logističke regresije. Postupak se ponavlja sve dok u dve uzastopne iteracije ne budu iste stavke označene kao one koje imaju DIF (S. W. Choi i ostali, 2011).

## *8.4 Šta ako DIF postoji?*

Ako otkrijemo postojanje DIF-a, postoji više puteva kojima možemo poći.

Kao prvo, ovde će biti opisani statistički postupci koji će nam ukazati da li DIF postoji ili ne. Dobićemo numeričke pokazatelje koji će nam to reći. Ono što nam neće reći je zašto DIF postoji, odnosno, zašto

[47] Iako ovde stalno govorimo o dve grupe, analiza DIF-a može uključivati i više fokalnih grupa.

jedno pitanje „favorizuje“ jednu grupu u odnosu na drugu. Za to nam je potrebna kvalitativna analiza, mišljenje eksperata iz oblasti predmeta merenja analiziranog testa, kognitivni intervjui sa ispitanicima...

Stavke koje pokazuju DIF možemo izbaciti iz testa, ali ne moramo. Ako DIF neke stavke ne dovodi i do diferencijalnog funkcionisanja testa, onda ga možemo ignorisati (Crane i ostali, 2007). U slučaju da DIF menja prosečne skorove (mere) grupa, onda bi u izračunavanju mera trebalo koristiti posebne parametre za različite grupe ili razmisliti o korekciji ili eliminaciji problematičnih stavki.

# 9 Provera pretpostavki modela

Provera pretpostavki modela je sastavni deo IRT modelovanja. Rečeno je da ona spada u proveru saglasnosti modela sa podacima. Takođe, značajan misfit nam najčešće ukazuje da su narušene pretpostavke modela. Ako pretpostavke modela ne proverimo pre procene određenog IRT modela, najčešće ćemo morati na to da se vratimo kada otkrijemo misfit kako bismo otkrili njegovo poreklo.

## *9.1 Provera dimenzionalnosti testa*

Proveru dimenzionalnosti testa možemo obaviti na više načina.

Prva mogućnost koja nam obično pada na pamet je svakako *faktorska analiza*. U ovu svrhu može se koristiti bilo eksplorativna, bilo konfirmativna faktorska analiza. Ako se upuštate u IRT analizu, pretpostavka je da imate dovoljno znanja u primeni faktorske analize pa nećemo detaljnije ulaziti u njeno objašnjavanje.

Ako koristimo *eksplorativnu faktorsku analizu* (EFA), jednodimenzionalnost će biti potvrđena ako dobijemo jednofaktorsko rešenje. Problem u EFA je to što je potrebno je doneti odluku koji broj faktora je opravdano zadržati, a za to možemo koristiti različite kriterijume. Uobičajeni su upotreba scree-dijagrama, Gutman-Kajzerovog kriterijuma ili metod *paralelne analize* (Horn, 1965), koja u poslednje vreme predstavlja metod izbora za ovu svrhu.

Podsetićemo, *paralelna analiza* poredi karakteristične korenove dobijene na podacima koje analiziramo sa karakterističnim korenovima dobijenim na velikom broju slučajno generisanih matrica istih dimenzija kao matrica originalnih podataka. Zadržavamo samo one faktore koji imaju karakteristični koren viši od 95. percentila vrednosti karakterističnih korenova dobijenih na slučajnim podacima (za isti faktor). Paralelnu analizu moguće je sprovesti u R-u, u paketima „psych“ (Revelle, 2018) i „EFA.MRFA“ (Navarro-Gonzalez & Lorenzo-Seva, 2020), a dostupna je i u besplatnim softverima „JASP“ (JASP Team, 2024) i „Jamovi“ (The jamovi project, 2024).

U okviru paketa „psych“ dostupne su procedure za još neke kriterijume za određivanje broja faktora koje je opravdano zadržati. Neke od njih su *Very Simple Structure* – VSS (Starkweather, 2014), *Minimum Average Partial* – MAP (Velicer, 1976) i Bayesian Information Criterion – BIC (sa i bez korekcije za veličinu uzorka). Prilikom izbora kriterijuma koji ćemo koristiti trebalo bi voditi računa da su neki od njih pre svega namenjeni za upotrebu u analizi glavnih komponenti (MAP, Gutman-Kajzerov kriterijum), da neki imaju različite varijante za faktorsku i komponentnu analizu (paralelna analiza), a da neki softveri (npr. SPSS) scree-dijagram crtaju na osnovu karakterističnih korenova iz inicijalnog rešenja (što je u stvari komponentna analiza). Takođe, bez obzira koji kriterijum koristimo uvek postoji izvesna doza subjektivnosti koja

može uticati na to da ne bude izabrano optimalno rešenje.

Dimenzionalnost se može proveriti i primenom konfirmativne faktorske analize (KFA). Prednost KFA je što imamo pokazatelje fita koji nam govore koliko je testirani model saglasan sa podacima i što postoje preporuke za interpretaciju tih pokazatelja (Hu & Bentler, 1999; MacCallum i ostali, 1996). S druge strane, pri interpretaciji ovih pokazatelja neki autori pokazuju kreativnost što i u ovaj metod uvodi izvesnu subjektivnost. I ovde test možemo smatrati dovoljno jednodimenzionalnim ako jednofaktorski model ima prihvatljive indekse fita (saglasnosti).

Bez obzira koja metoda je korišćena, idealna situacija je kada dobijemo jednofaktorsko rešenje. Tada možemo reći da je instrument jednodimenzionalan i primeniti neki jednodimenzionalni IRT model. U kontekstu provere pretpostavki jednodimenzionalnih IRT modela, nas i ne zanima koji faktorski model najbolje opisuje podatke. Zanima nas da li su podaci dovoljno jednodimenzionalni da bi određeni IRT model mogao biti primenjen a da procenjeni parametri ne budu pristrasni u velikoj meri.

Čak i ako se faktorskom analizom izdvoji više faktora, to još uvek ne znači da instrument nije pogodan za IRT modelovanje. Ako smo faktorskom analizom dobili više faktora, a ti faktori su relativno visoko korelirani, te prvi faktor objašnjava najveći deo varijanse i pri tom postoji teorijska osnova da izolovane faktore tumačite kao poddomene šireg konstrukta, instrument može biti tretiran kao suštinski jednodimenzionalan. Jednodimenzionalnost se može potvrditi i faktorskom analizom drugog reda ukoliko izolovani faktori prvog reda daju jednofaktorsko rešenje u hijerarhijskoj faktorskoj analizi (Fajgelj, 2020).

S druge strane, kada dobijemo jednodimenzionalno rešenje koristeći EFA ili KFA, to ne mora uvek značiti da stavke mere samo jednu stvar. Teorijski, ako sve stavke mere (recimo) dva konstrukta u istim proporcijama, jednofaktorsko rešenje može se pokazati kao najbolje. Takođe, u slučaju da stavke mere dva konstrukta (ne nužno u istim proporcijama), ali *u uzorku* jedan od njih nema varijansu[48], to će dovesti do jednofaktorskog rešenja. Nedostatak varijanse može biti osobina uzorka (ispitanici su ujednačeni po drugom merenom konstruktu) ili situacija merenja (sama situacija merenja nije zahtevala upotrebu druge merene osobine – npr. motivacije)(DeMars, 2010). Na drugom uzorku ili u drugoj situaciji merenja, isti test se može pokazati kao višedimenzionalan.

Drugi mogući način testiranja jednodimenzionalnosti je *Stautov test suštinske jednodimenzionalnosti* (engl. *Stout's Test of Essential Unidimensionality*), koji se nekada naziva i *Stautovim indeksom suštinske jednodimenzionalnosti*, a obično se označava sa *T* (DeMars, 2016; Hattie i ostali, 1996). Nulta hipoteza ovog testa je da su prosečne kovarijanse parova stavki kod ispitanika sa istim nivoom latentne osobine jednake 0.

Logika testa je sledeća, ako ispitanici imaju isti nivo osobine, trebalo bi da na istim stavkama postižu iste skorove, pa bi varijansa, a samim tim i kovarijansa tih skorova trebalo da bude 0 ili vrlo bliska toj vrednosti (ostaje samo slučajna greška koja ne korelira ni sa čim). Ovako definisan test liči na test koji bi testirao postojanje *lokalne zavisnosti*, ali sama *procedura sprovođenja testa* omogućava proveru dimenzionalnosti.

[48] Ili bar ne dovoljno veliku varijansu.

Iz testa čiju dimenzionalnost testiramo potrebno je izdvojiti dva (odnosno 3) suptesta. Prvi suptest naziva se *Test za procenu 1* (engl. *Assessment Subtest 1* – AT1). U ovaj suptest izdvajaju se stavke za koje imamo opravdani razlog da sumnjamo da mere drugu dimenziju. Stavke možemo izabrati na osnovu teorijskih razloga, na osnovu mišljenja eksperata ili empirijski. Empirijski možemo to učiniti na osnovu analize glavnih komponenti (AGK) na matrici polihoričkih korelacija ili na osnovu klaster analize. Na osnovu AGK, u AT1 izabraćemo stavke koje su najviše zasićene drugom nerotiranom komponentom. Imajte u vidu da ako koristimo empirijski pristup, potrebno je uzorak podeliti na dva dela i izdvajanje stavki u suptestove obaviti na jednom poduzorku, a proveru dimenzionalnosti na drugom (DeMars, 2010).

U drugi suptest (AT2) od preostalih stavki biraju se stavke koje su po težini i dimenzionalnosti slične stavkama u AT1. Ovaj suptest služi za korekciju pristrasnosti procene indeksa *T* koja može nastati usled pomeranja aritmetičke sredine zbog biranja stavki homogenih po težini u suptestove (Hattie i ostali, 1996).

Preostale stavke se koriste za rangiranje ispitanika po nivou osobine. Za to se obično koriste sumacioni skorovi na ovom skupu stavki. Ovaj suptest naziva se *Test za podelu* (engl. *Partitioning Subtest* – PT*).*

Za svaku od $g$ grupa formiranih po skorovima na PT-u računa se uobičajena procena varijanse na stavkama koje čine AT1. Što je ova varijansa viša, to više ukazuje na odstupanje od jednodimenzionalnosti. Osim varijanse, računa se i njena standardna greška.

Takođe, računa se i procena ukupne jednodimenzionalne varijanse kao suma varijanse u G grupa.

Na kraju, izračunava se indeks *T*.

$$T_L = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \frac{\left|\hat{\sigma}_g^2 - \hat{\sigma}_{Tg}^2\right|}{S_g}$$

[ 110 ]

gde je $G$ ukupan broj grupa $g$ formiranih po skorovima, $\hat{\sigma}_g^2$ – procena varijanse skorova u grupi $g$, $\hat{\sigma}_{Tg}^2$– suma procena varijansi po grupama, a $S_g$ – standardna greška procene varijanse u grupi $g$.

*T* u izrazu [ 110 ] ima indeks *L* koji označava da se radi o indeksu za skup stavki AT1. Ovaj indeks je osetljiv na odstupanje od jednodimenzionalnosti i na izvore pristrasnosti. Isti takav indeks računa se i na skupu stavki AT2 i nosi indeks *B* ($T_B$). On nije osetljiv na dimenzionalnost, ali jeste na izvore pristrasnosti i zato se koristi za korekciju indeksa. Konačni indeks *T* se zato računa na sledeći način (Hattie i ostali, 1996):

$$T = \frac{T_L - T_B}{\sqrt{2}}$$

[ 111 ]

Ovaj indeks ima normalnu raspodelu, pa nultu hipotezu odbacujemo ako je *T* veći od vrednosti *z skora* vezanog za određeni α nivo (npr. >1,96 za 95%).

Na žalost, R ne poseduje funkciju za izračunavanje ovog indeksa pa on neće biti prikazan u praktičnom delu ove knjige, ali će biti prikazani neki dodatni načini koji su sastavni deo paketa za IRT modelovanje.

## 9.2 Provera pretpostavke lokalne nezavisnosti

Provera narušavanja pretpostavke lokalne nezavisnosti se može obaviti na više načina. Najčešće u okviru paketa za IRT analize postoje procedure za proveru, a u zavisnosti od paketa mogu se zasnivati na različitim principima.

U paketu *eRm* (Mair & Hatzinger, 2007) namenjenom Rašovim modelima, za proveru lokalne nezavisnosti *binarno skorovanih stavki* koriste se testovi koji predstavljaju varijantu Fišerovog egzaktnog testa. Oslanjaju se na činjenicu da je u Rašovom modelu sumacioni skor dovoljan statistik za nivo osobine, a broj ispitanika koji je tačno odgovorio na stavku dovoljan statistik za parametar težine stavke. To omogućava da se na osnovu matrice **A** (matrica podataka sa ispitanicima u *n* redova i stavkama u *p* kolona) kreiraju sve moguće matrice sa istim parametrima (matrice koje imaju iste marginalne vrednosti). Obično se ne kreiraju baš sve moguće matrice već zadati broj *B*. Izračunavaju se pokazatelji $T_b$ (odgovarajući pokazatelji za svaki od *B* simuliranih uzoraka) i kao *p* vrednost uzima proporcija javljanja vrednosti izračunate na opaženoj matrici $\mathbf{A_0}$ u vrednostima dobijenim na simuliranim/kreiranim matricama od $\mathbf{A_1}$ do $\mathbf{A_B}$. Za simulaciju se koriste tehnike na bazi Monte Karlo (Monte Carlo) pristupa (Koller & Hatzinger, 2013).

Jedan od testova koji se koristi je $T_{11}$. Ovo je globalni test neodgovarajućih interajtemskih korelacija i homogenosti:

$$T_{11}(\boldsymbol{A}) = \sum_{jk} \left| r_{jk} - \tilde{r}_{jk} \right|$$

[ 112 ]

gde je $r_{jk}$ korelacija između stavki *j* i *k*, a $\tilde{r}_{jk}$ predstavlja procenjenu prosečnu korelaciju stavki *j* i *k* dobijenu na osnovu simuliranih matrica. Sumacija se vrši po svim mogućim parovima stavki.

Kao vrednost $T_0$ računa se suma apsolutnih odstupanja $r_{jk}$ od $\tilde{r}_{jk}$ za svaki par stavki u opaženoj matrici. U narednom koraku računa se isti statistik za svaku od simuliranih matrica $T_1 ... T_B$. Statistička značajnost se zatim proveravana na sledeći način:

$$p = \frac{1}{B} \sum_{b}^{B} t_b$$

[ 113 ]

gde je $t_b$=1 ako je $T_b(\boldsymbol{A}_b) \geq T_0(\boldsymbol{A}_0)$, a $t_b$=0 u ostalim slučajevima.

Značajan test $T_{11}$ može ukazivati na lokalnu zavisnost i/ili na niske korelacije među stavkama (Koller & Hatzinger, 2013).

Još jedan test koji se primenjuje u paketu *eRm* je test $T_{1l}$. Ovaj test je namenjen parovima stavki. Ispituje da li je učestalost ekstremnih sklopova odgovora na dve stavke (npr. 00 ili 11) veća nego što to model predviđa, odnosno da li je dobijena učestalost takvih sklopova veća nego što se dobija u 95% slučajeva na simuliranim matricama. Ukoliko je ova učestalost prevelika, korelacije između stavki biće procenjene kao prevelike (Koller & Hatzinger, 2013). Test se računa na sledeći način:

$$T_{1l}(\boldsymbol{A}) = \sum_{i}^{n} \delta_{ijk}$$

[ 114 ]

gde je $\delta_{ijk}$=1 ako su odgovori ispitanika *i* na stavke *j* i *k* jednaki, a $\delta_{ijk}$=0 ako su različiti.

Statistička značajnost testa se procenjuje na sledeći način:

$$p = \frac{1}{B}\sum_{b}^{B} t_b$$

[ 115 ]

gde je $t_b$=1 ako je $T_b \geq T_0$, a $t_b$=0 u ostalim slučajevima.

Drugi načini provere lokalne zavisnosti baziraju se na korelacijama reziduala.

Za otkrivanje lokalne zavisnosti se između ostalog koriste Pirsonov hi–kvadrat test, ili hi–kvadrat test maksimalne verodostojnosti $G^2$, sa Kramerovim (Cramer) V ili ϕ koeficijentom kao pokazateljem korelacije. $\chi^2$ i $G^2$ su asimptotski ekvivalentni, ali na uzorcima mogu davati drugačije rezultate (Chen & Thissen, 1997). Kako bi bio uhvaćen smer eventualne lokalne zavisnosti, koriste se varijante hi–kvadrata sa predznakom (engl. *signed chi-squared test*).

Predložena granična vrednost Pirsonovog hi–kvadrata iznad koje se smatra da postoji lokalna zavisnost je 3,84, a za $G^2$ 6,63 (Chen & Thissen, 1997). De Ajala (de Ayala, 2022) predlaže upotrebu Koenove podele (Cohen, 1988) veličine efekta za Kramerovo V kao osnovu za interpretaciju na ovaj način izračunate lokalne zavisnosti (Tabela 4).

Tabela 4 – Veličine efekta za Kramerovo V u zavisnosti od broja kategorija odgovora stavke

| | Broj kategorija odgovora | | | | |
|---|---|---|---|---|---|
| efekat | 2 | 3 | 4 | 5 | 6 |
| mali | 0,100 | 0,071 | 0,058 | 0,050 | 0,045 |
| srednji | 0,300 | 0,212 | 0,173 | 0,150 | 0,134 |
| veliki | 0,500 | 0,354 | 0,289 | 0,250 | 0,224 |

Jedan od najčešće korišćenih pokazatelja je $Q_3$ pokazatelj Jenove (Yen, 1984) zasnovan na Pirsonovim produkt-moment korelacijama reziduala. Interpretacija ovog pokazatelja zasnovana je na veličini efekta a ne na statističkoj značajnosti, što ima smisla s obzirom da je veličina uzoraka koji se koriste u IRT analizi dovoljna da korelacija reziduala koja bi mogla biti razlog za brigu bude i statistički značajna (DeMars, 2010). Čen i Tisen (Chen & Thissen, 1997) predlažu graničnu vrednost od $r$>0,20 kao znak da bi mogla postojati nezanemarljiva lokalna zavisnost. U opticaju su i druge vrednosti poput 0,36 (Stone & Zhang, 2003) kao granice srednje veličine efekta po Koenovoj klasifikaciji (Cohen, 1988).

Drugi način interpretacije $Q_3$ statistika je procena varijanse reziduala koju dve stavke dele. Kako je

ovaj pokazatelj koeficijent korelacije, njegov kvadrat je koeficijent determinacije i predstavlja deljenu varijansu reziduala dve stavke. Možemo smatrati značajnim narušavanjem lokalne nezavisnosti ako dve varijable dele više od 5% ove varijanse (de Ayala, 2022). Taj procenat odgovara korelaciji od nešto većoj od 0,22. Međutim, ova granična vrednost dovodi do nešto niže snage ovog pokazatelja u poređenju sa drugim pokazateljima (Chen & Thissen, 1997).

Mare (Marais, 2013) je utvrdila da korelacije reziduala zavise od broja stavki u instrumentu te da su, kada je broj stavki manji, i korelacije reziduala niže (pod pretpostavkom jednake lokalne zavisnosti). Iz tog razloga bi korelaciju reziduala dve stavke trebalo posmatrati u odnosu na korelacije ostalih parova stavki. Ona predlaže da se ne upotrebljava jedinstvena granična vrednost za sve testove, već da se pojedinačne korelacije reziduala parova stavki porede sa prosečnom korelacijom reziduala u matrici podataka (Marais, 2013). Prosečna korelacija reziduala, odnosno $Q_{3,}$ izračunava se na sledeći način:

$$\bar{Q}_3 = \frac{\sum_{j>k} Q_{3,jk}}{m(m-1)/2}$$

[ 116 ]

gde je $m$ broj stavki.

Na osnovu prosečne vrednosti $Q_3$ i maksimalne vrednosti $Q_3$ u matrici ($Q_{3,max}$) definisala je pokazatelj:

$$Q_{3,*} = Q_{3,max} - \bar{Q}_3$$

[ 117 ]

$Q_{3,*}$ i $Q_{3,max}$ su pokazatelji koji govore da li postoji lokalna zavisnost u skupu stavki. Na osnovu simulacija i empirijskih podataka Kristensen i saradnici (Christensen i ostali, 2017) su zaključili da ne postoji univerzalna granična vrednost za ove pokazatelj, već je poželjno za svaki poseban slučaj utvrditi je upotrebom simulacija. Od ova dva pokazatelja preporučuju $Q_{3,*}$ jer manje zavisi od broja stavki, a granična vrednost mu je zadovoljavajuće stabilna oko vrednosti 0,2. Što se tiče $Q_3$ pokazatelja za stavke kažu da je svaka vrednost koja je viša za 0,2 od prosečne vrednosti u matrici indikator lokalne zavisnosti, a vrednost koja je za 0,3 viša je malo moguća kod nezavisnih stavki (Christensen i ostali, 2017).

$Q_3$ pokazatelj je pokazao jednaku ili veću snagu od $\chi^2$ pokazatelja uz manju stopu lažno pozitivnih nalaza (Mislevy i ostali, 2012). Treba imati na umu da je hi–kvadrat osetljiv na očekivane frekvencije bliske 0, a da tu pojavu češće možemo očekivati kod politomnih stavki sa većim brojem kategorija. U tom slučaju se povećava verovatnoća greške tipa I (lažno pozitivni).

Ovde smo razmatrali samo numeričke pokazatelje koji nam govore da među stavkama možda postoje lokalno zavisne. Kada dobijemo takav pokazatelj, trebalo bi obratiti pažnju na sadržaj stavki i proveriti da li postoji razlog zbog kojeg bi stavke bile lokalno zavisne. Lokalna zavisnost stavki može poticati od dodatne latentne varijable koju dve stavke mere ali to ne dele sa ostalima. S druge strane, lokalna zavisnost može biti i „površinska“, odnosno da zavisi od položaja stavki u testu i pojavne sličnosti, a ne od dubinskih karakteristika (Chen & Thissen, 1997).

Kada se utvrdi postojanje lokalne zavisnosti, možemo izbaciti jednu od stavki u paru ili primeniti neki drugi model. Na primer, možemo sumirati skorove na dve stavke i na tako sumirane skorove primeniti model stepenovanog ocenjivanja (tretirajući ih u analizi kao jednu stavku). Takođe, moguće je primeniti IRT modele namenjene testletima ili bifaktorski multidimenzionalni model. Kristensen i saradnici kažu da je kombinovanje stavki i primena PCM-a dalo najbolje pokazatelje fita i informativnosti testa (Christensen i ostali, 2017)

## *9.3 Provera pretpostavke monotonosti*

Pretpostavku monotonosti KKS (o kojoj smo više rekli u odeljku 5.1) možemo testirati primenom metoda karakterističnih za Mokenovu analizu skala. Mokenova analiza skala zasniva se na dva neparametarska IRT modela. Prvi model je *model monotone homogenosti,* a drugi *model dvostruke monotonosti* (videti odeljak 4.3).

Kao pokazatelj narušavanja monotonosti u Mokenovoj analizi koristi se *Crit*. *Crit* je veličina efekta narušavanja monotonosti i predstavlja integraciju više drugih pokazatelja poput indeksa skalabilnosti (videti izraz [ 46 ]), broj narušavanja monotonosti, značajnost narušavanja monotonosti, sumu svih narušavanja monotonosti i još nekih (detaljnije prilikom prikaza procedure provere u odeljcima 10.3.2 i 10.5.2.2).

Osim ovih numeričkih pokazatelja za procenu monotonosti, mogu se koristiti i grafički prikazi KKS na osnovu neparametarskih modela.

Procedure koje se koriste u okviru Mokenove analize skala nam omogućavaju i testiranje pretpostavke paralelnosti KKS (invarijantnosti poretka stavki), ali to nije pretpostavka svih IRT modela.

# 10 Procena IRT modela u R–u

Za prolazak kroz primere analiza koji slede, pretpostavlja se da imate neko znanje i iskustvo u radu sa R-om. Ako nemate, dobro mesto za početak je sledeća veb adresa https://www.bigbookofr.com/.

Čak i bez toga prikazane primere možete reprodukovati kopiranjem prikazanih komandi u R (R Core Team, 2023) ili Rstudio (Posit team, 2024). Naravno, podrazumeva se da imate instalirane obe aplikacije.

## *10.1 Instaliranje i pozivanje potrebnih paketa (biblioteka)*

Za početak, dobro je instalirati paket (biblioteku) pod nazivom *pacman* (Rinker & Kurkiewicz, 2018) i učitati je. Ova biblioteka ima funkciju *p_load()* koja proverava da li imate instaliranu biblioteku koju navedete u komandi (npr. *p_load(mirt)*) i ako nemate, instalira je i učita. Praktično izvrši komande *install.packages("naziv_biblioteke")* i *require("naziv_biblioteke")*. U narednom delu koda komanda *install.packages("pacman")* je komentarisana sa znakom # ispred. Kada ispred komande stoji #, ona neće biti izvršena. Možete je izvršiti ako obrišete znak #.

```
#install.packages("pacman")
#Instalirati ukoliko je potrebno
library(pacman)
```

## *10.2 Simulacija podataka*

Kako bi primeri iz ove knjige bili reproducibilni koristićemo simulirane podatke. Za to ćemo koristiti paket *mirt* (Chalmers, 2012). Na ovom mestu simuliraćemo podatke za binarno skorovane stavke, a podaci će biti smešteni u objekat pod nazivom *b.pod*.

Za simulaciju podataka potrebno je da definišemo željene parametre stavki. Objekat *nagibi* sadržaće parametre diskriminativnosti, objekat *lokacije* parametre težine, a objekat *pogadjanje* donje asimptote, odnosno, parametre pseudopogađanja. Sva tri objekta su numerički vektori dužine 10 – pošto simuliramo 10 stavki. Svakom elementu vektora možemo pristupiti preko indeksa koji je dat u uglastim zagradama (npr. ako izvršite komandu *nagibi[8]* biće vam ispisan parametar diskriminativnosti za stavku 8). Da bi podaci bili reproducibilni bitna je vrednost *set.seed* koja se zadaje prilikom simulacije. Ona je u primeru definisana kao promenljiva *ss* na početku koda.

```
ss <- 3
set.seed(ss)
nagibi <- runif(10, 1, 2.2)
# Nagibi/diskriminativnosti ajtema simulirani su iz uniformne distribucije, sa
minimalnom vrednošću 1, a maksimalnom 2.2
lokacije <- runif(10, -3, 3)
# Lokacije/težine ajtema simulirane iz uniformne distribucije, sa minimalnom
```

```
vrednošću -3, a maksimalnom 3
pogadjanje <- runif(10, 0, 0)
# Parametri pseudopogađanja ajtema simulirani iz uniformne distribucije, sa
minimalnom vrednošću 0, a maksimalnom takođe 0 (nema pogađanja)
# Zbog reproducibilnosti potrebno je pre simulacije parametara uvek primeniti
komandu set.seed(ss). Vrednost ss je postavljena na početku.
set.seed(ss)
thete <- rnorm(1000, mean=0)
# Nivoi osobine ispitanika simulirani iz normalne distribucije sa AS=0 i SD=1

set.seed(ss)
b.pod <- data.frame(mirt::simdata(a=nagibi, d=lokacije, u=1, N=1000,
 itemtype = "dich"))
# Argument itemtype="dich" definiše da želimo da stavke budu dihotomne (0,1)
names(b.pod) <- paste0("it",1:10)
# Imenovanje varijabli (stavki)

# Prikaz prva tri reda matrice podataka

head(b.pod,3)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## 1   0   0   0   0   1   1   0   1   1    0
## 2   0   1   0   0   0   1   0   1   1    0
## 3   0   1   1   0   1   1   0   1   1    0

# Deskriptivni pokazatelji

psych::describe(b.pod)

##      Vars    n  mean   sd  median trimmed   mad  min  max  range   skew  kurtosis   se
## it1     1 1000  0.52 0.50    1     0.53     0    0    1     1    -0.09    -1.99 0.02
## it2     2 1000  0.49 0.50    0     0.49     0    0    1     1    -0.09    -1.99 0.02
## it3     3 1000  0.52 0.50    1     0.52     0    0    1     1    -0.08    -2.00 0.02
## it4     4 1000  0.55 0.50    1     0.57     0    0    1     1    -0.21    -1.96 0.02
## it5     5 1000  0.80 0.40    1     0.88     0    0    1     1    -1.51     0.29 0.01
## it6     6 1000  0.79 0.41    1     0.86     0    0    1     1    -1.39    -0.06 0.01
## it7     7 1000  0.13 0.34    0     0.04     0    0    1     1     2.21     2.89 0.01
## it8     8 1000  0.71 0.45    1     0.76     0    0    1     1    -0.92    -1.16 0.01
## it9     9 1000  0.81 0.39    1     0.89     0    0    1     1    -1.59     0.52 0.01
## it10   10 1000  0.28 0.45    0     0.22     0    0    1     1     0.98    -1.04 0.01
```

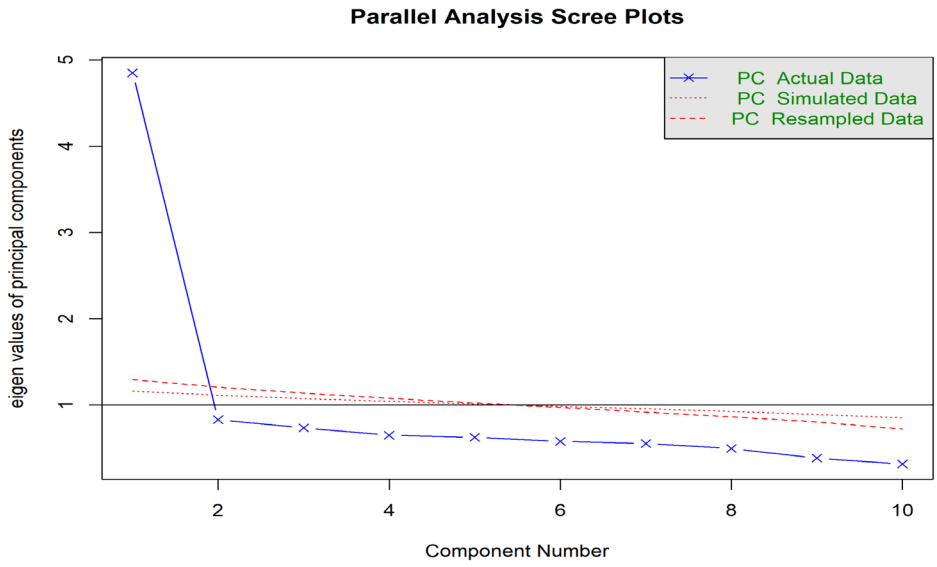
## *10.3 Provera pretpostavki modela*

Sada kada imamo podatke prvo je potrebno da proverimo da li su zadovoljene pretpostavke modela.

### 10.3.1 Jednodimenzionalnost

Prikazaćemo proveru dimenzionalnosti upotrebom paralelne analize (Horn, 1965), kriterijuma vrlo jednostavne strukture (VSS)(Revelle & Rocklin, 1979), Veliserovog (Velicer, 1976) kriterijuma minimalne prosečne parcijalne korelacije (Minimum Average Partial - MAP) i BIC-a, svi dostupni u paketu *psych* (Revelle, 2018).

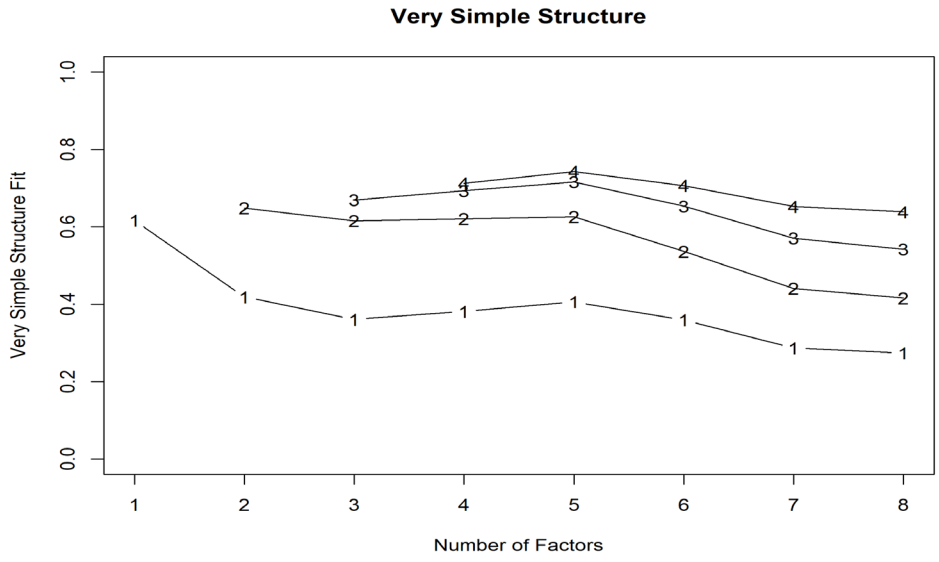
```
p_load(psych)
PA <- fa.parallel(b.pod, n.iter = 1000,cor="tet", quant=.95, fa="pc")
```




Slika 26 – Skri-dijagram paralelne analize

```
## Parallel analysis suggests that the number of factors =  NA  and the number of
components =  1

# Kao broj slučajeva se postavlja broj redova matrice podataka (ako nema
nedostajućih podataka)
VSS <- vss(b.pod, n.obs = nrow(b.pod))
```

```
summary(VSS)

##
## Very Simple Structure
## VSS complexity 1 achieves a maximimum of 0.62  with  1  factors
```




Slika 27 – Dijagram kriterijuma vrlo jednostavne strukture (VSS)

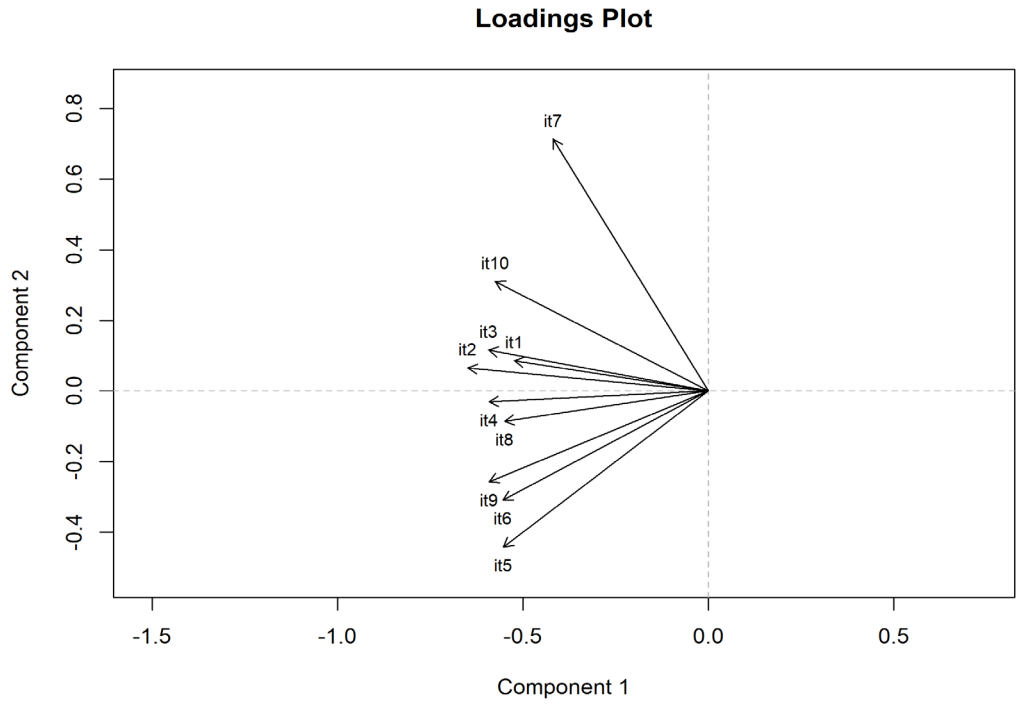
```
## VSS complexity 2 achieves a maximimum of 0.65  with  2  factors
##
## The Velicer MAP criterion achieves a minimum of 0.46  with  1  factors
##
```

Revel (Revelle, 2017) kaže da je paralelna analiza osetljiva na veličinu uzorka. Kod velikih uzoraka karakteristični korenovi simuliranih podataka biće niski, što dovodi do toga da paralelna analiza precenjuje broj faktora. Iz tog razloga smo koristili paralelnu analizu po modelu glavnih komponenti koja daje realniju procenu broja faktora.

Paralelna analiza sugeriše da je broj faktora koje bi trebalo zadržati 1, a isto sugerišu i VSS i MAP (Slika 27). Skri dijagram sa tačkom preloma na drugom faktoru, takođe sugeriše jednodimenzionalnost (Slika 26).

Još jedan način da proverimo dimenzionalnost je upotreba kategorijalne analize glavnih komponenti. Ona je pogodna zbog grafičkog prikaza. Koristićemo paket *Gifi* (Mair & Leeuw, 2022) i njegovu funkciju *princals().*

```
p_load(Gifi)
PCd <- princals(b.pod)
plot(PCd)
```




Slika 28 – Grafikon opterećenja kategorijalne analize glavnih komponenti

Kada je skup stavki jednodimenzionalan, strelice na grafikonu (Slika 28) bi trebalo da budu okrenute manje-više u istom smeru (Mair, 2018). Na osnovu grafikona možemo zaključiti da je u pitanju jednodimenzionalan test, s tim da stavka 7 donekle odstupa od onoga što mere ostale stavke.

Još jedan način provere može biti procena jednodimenzionalnih i višedimenzionalnih modela upotrebom faktorske analize stavki (engl. *item factor analysis* – IFA). IFA je dostupna u paketima *mirt* i *ltm.* Postupak se sastoji u proceni jednodimenzionalnog i dvodimenzionalnog modela i njihovim kasnijim poređenjem upotrebom LR testa (test količnika verodostojnosti).

```
# Učitaćemo biblioteku mirt
p_load(mirt)
mt1 <- mirt(b.pod, 1, verbose = FALSE)
mt2 <- mirt(b.pod, 2, technical = list(NCYCLES=10000), verbose = FALSE)
# U mt2 argument technical = list(NCYCLES=10000) bio je potreban da se poveća broj
EM iteracija kako bi model konvergirao.
anova(mt1, mt2)

##          AIC    SABIC      HQ      BIC    logLik     X2 df     p
## mt1 10402.10 10436.73 10439.4 10500.25 -5181.049
## mt2 10406.41 10456.62 10460.5 10548.73 -5174.203 13.693  9 0.134

# Ako je test značajan bolji je onaj sa višim logLik (pošto je negativan onaj sa
nižom apsolutnom vrednošću), odnosno model sa nižim AIC ili BIC kriterijumom.
detach("package:mirt", unload=TRUE)
# U ovom trenutku nam više ne treba paket mirt

# Ista procedura može se izvesti i u paketu "ltm"
p_load(ltm)
# Jednodimenzionalni model
lt1 <- ltm::ltm(b.pod ~ z1)
# Dvodimenzionalni model
lt2 <- ltm::ltm(b.pod ~ z1+z2)
# Likelihood Ratio test
anova(lt1, lt2)

##
##  Likelihood Ratio Table
##          AIC      BIC  log.Lik   LRT df p.value
## lt1 10402.08 10500.23 -5181.04
## lt2 10406.66 10553.89 -5173.33 15.42 10   0.118

detach("package:ltm", unload=TRUE)
```

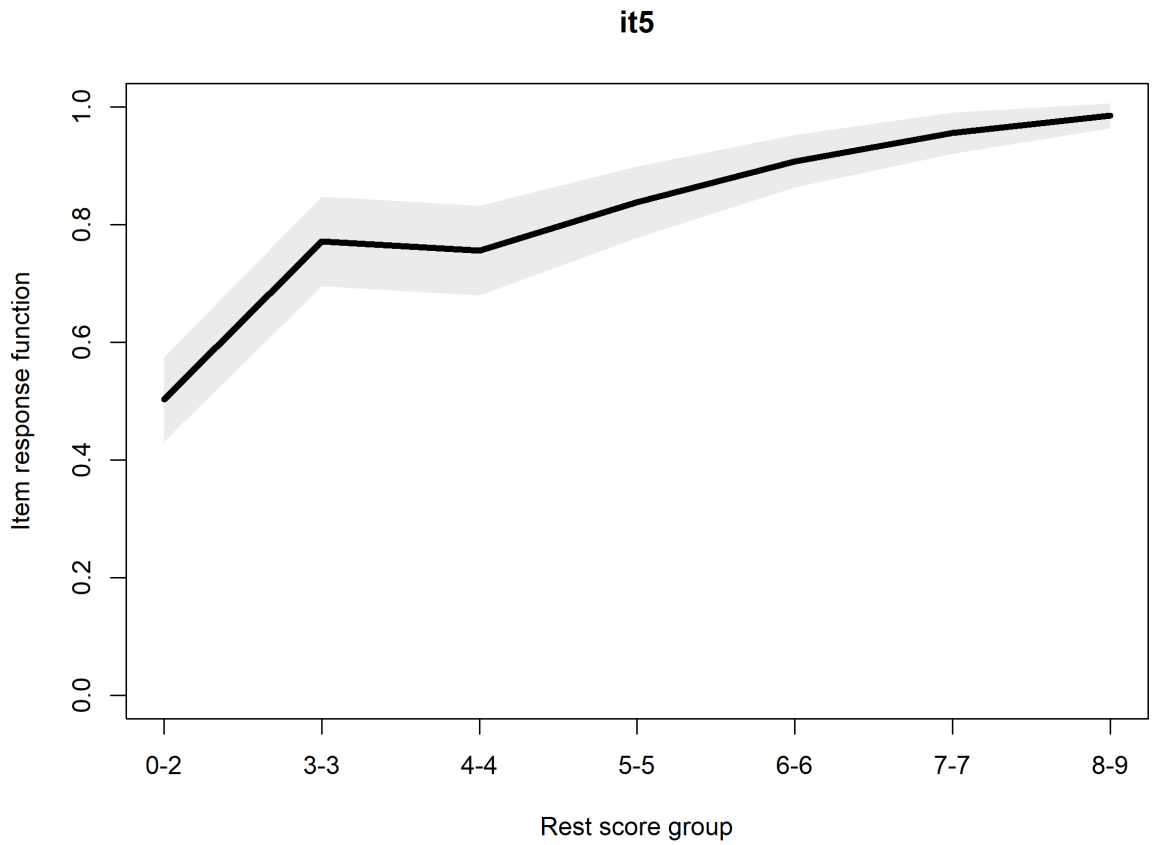
U oba slučaja razlika između modela nije značajna. Jednodimenzionalni model imao je niže vrednosti AIC i BIC što mu daje prednost, ali i nižu vrednost logaritma verodostojnosti što daje prednost dvodimenzionalnom modelu. Kako LR test nije značajan opredelićemo se za jednostavniji model, odnosno, možemo reći da je test dovoljno jednodimenzionalan.

### 10.3.2 Monotonost

Proveru monotonosti obavićemo primenom paketa *mokken* (Van der Ark, 2007). Ovaj paket namenjen je Mokenovoj analizi skala (odeljak 4.3).

```
p_load(mokken)
monotonost <- check.monotonicity(b.pod, minvi=.03)
summary(monotonost)

##      ItemH #ac #vi #vi/#ac maxvi sum sum/#ac zmax #zsig crit
## it1   0.32  21   0       0     0   0       0    0     0    0
## it2   0.42  21   0       0     0   0       0    0     0    0
## it3   0.37  15   0       0     0   0       0    0     0    0
## it4   0.37  21   0       0     0   0       0    0     0    0
## it5   0.40  21   0       0     0   0       0    0     0    0
## it6   0.39  21   0       0     0   0       0    0     0    0
## it7   0.53  15   0       0     0   0       0    0     0    0
## it8   0.37  21   0       0     0   0       0    0     0    0
## it9   0.44  21   0       0     0   0       0    0     0    0
## it10  0.51  15   0       0     0   0       0    0     0    0

skalabilnost <- coefH(b.pod)

## $Hij
##      it1     se      it2     se      it3     se      it4     se      it5
## it1                  0.245  (0.034)  0.249  (0.032)  0.273  (0.034)  0.392
## it2   0.245  (0.034)                  0.341  (0.034)  0.337  (0.036)  0.566
## it3   0.249  (0.032)  0.341  (0.034)                  0.267  (0.034)  0.484
## it4   0.273  (0.034)  0.337  (0.036)  0.267  (0.034)                  0.479
## it5   0.392  (0.058)  0.566  (0.056)  0.484  (0.056)  0.479  (0.054)
## it6   0.357  (0.056)  0.608  (0.052)  0.433  (0.055)  0.442  (0.052)  0.293
## it7   0.415  (0.079)  0.651  (0.064)  0.533  (0.073)  0.549  (0.076)  0.452
## it8   0.310  (0.047)  0.459  (0.047)  0.384  (0.046)  0.335  (0.044)  0.259
## it9   0.444  (0.059)  0.610  (0.055)  0.470  (0.058)  0.493  (0.055)  0.314
## it10  0.446  (0.049)  0.427  (0.047)  0.480  (0.048)  0.457  (0.051)  0.693
##      se      it6     se      it7     se      it8     se      it9     se
## it1  (0.058)  0.357  (0.056)  0.415  (0.079)  0.310  (0.047)  0.444  (0.059)
## it2  (0.056)  0.608  (0.052)  0.651  (0.064)  0.459  (0.047)  0.610  (0.055)
## it3  (0.056)  0.433  (0.055)  0.533  (0.073)  0.384  (0.046)  0.470  (0.058)
## it4  (0.054)  0.442  (0.052)  0.549  (0.076)  0.335  (0.044)  0.493  (0.055)
## it5           0.293  (0.040)  0.452  (0.133)  0.259  (0.045)  0.314  (0.040)
## it6  (0.040)                  0.602  (0.111)  0.268  (0.043)  0.280  (0.041)
## it7  (0.133)  0.602  (0.111)                  0.600  (0.094)  0.836  (0.080)
## it8  (0.045)  0.268  (0.043)  0.600  (0.094)                  0.396  (0.046)
## it9  (0.040)  0.280  (0.041)  0.836  (0.080)  0.396  (0.046)
## it10 (0.068)  0.666  (0.068)  0.397  (0.057)  0.558  (0.063)  0.792  (0.059)
##      it10    se
## it1   0.446  (0.049)
## it2   0.427  (0.047)
## it3   0.480  (0.048)
## it4   0.457  (0.051)
## it5   0.693  (0.068)
## it6   0.666  (0.068)
## it7   0.397  (0.057)
## it8   0.558  (0.063)
## it9   0.792  (0.059)
## it10
##
## $Hi
##      Item H  se
## it1    0.319 (0.023)
## it2    0.417 (0.022)
```

```
## it3    0.366 (0.022)
## it4    0.366 (0.023)
## it5    0.401 (0.027)
## it6    0.393 (0.027)
## it7    0.530 (0.040)
## it8    0.367 (0.025)
## it9    0.444 (0.026)
## it10   0.506 (0.027)
##
## $H
## Scale H      se
##   0.397 (0.016)

skalabilnost$H

## Scale H      se
##   0.397 (0.016)

plot(monotonost, items=5, ask=FALSE, curves="IRF")
```




Slika 29 – Funkcija odgovora na stavku 5 u zavisnosti od skora ostatka

Rezultat funkcije *check.monotonicity()* je tabela sa većim brojem pokazatelja.

Bitan pojam u ovoj analizi je *skor ostatka* (engl. – *restscore*). Radi se o ukupnom skoru bez stavke za koju se proverava monotonost.

Raspon skorova ostatka deli se na grupe, odnosno na intervale. Broj grupa zavisiće od minimalnog

broja ispitanika u njima. Minimalna veličina grupe definisana je argumentom *minsize=.* Ako se ovaj argument izostavi primenjuju se podrazumevane vrednosti. Podrazumevana vrednost je n/10 ako je n≥500, n/5 ako je 250≤n<500, i *max*(n/3, 50) ako je n<250.

Broj grupa zavisi od definisane minimalne veličine, a ako u nekoj ima manje ispitanika susedne grupe se spajaju. Oznaka #ac u ispisu funkcije *check.monotonicity()* označava ovaj broj grupa skorova ostatka.

*ItemH* je koeficijent skalabilnosti stavke. On je mera kvaliteta i korisnosti stavke. Predstavlja odnos kovarijanse skora stavke i skora ostatka i *maksimalne moguće* kovarijanse skora stavke i skora ostatka .

*#vi* označava broj slučajeva narušavanja monotonosti, *#zsig* broj statistički značajnih narušavanja monotonosti, *maxvi* je maksimalna vrednost narušavanja monotonosti, *sum* je suma svih narušavanja.

zmax i #zsig odnose se na standardizovana odstupanja svakog od narušavanja monotonosti..

*Crit* je veličina efekta narušavanja monotonosti i predstavlja rekapitulaciju ostalih pokazatelja i izračunava se na sledeći način :

$$\mathrm{Crit}_j = 50(0.3 - H_j) + \sqrt{\#\mathrm{vi}} + 100\frac{\#\mathrm{vi}}{\#\mathrm{ac}} + 100\mathrm{maxvi} + 10\sqrt{\mathrm{sum}} + 1000\frac{\mathrm{sum}}{\#\mathrm{ac}} + 5\mathrm{zmax} + 10\sqrt{\mathrm{zsig}} + 100\frac{\mathrm{zsig}}{\#\mathrm{ac}} \quad [\,118\,]$$

Smatra se da je monotonost ozbiljno narušena ako je Crit>80. Ako je 40<Crit<80 postoje naznake narušavanja, ali nisu jednoznačne, a ako je Crit<40 nema dokaza da postoji narušavanje monotonosti. Međutim, postoje autori koji svaku vrednost Crit pokazatelja veću od 40 interpretiraju kao ozbiljno narušavanje monotonosti (Crişan i ostali, 2019).

U našem primeru ne vidimo značajno narušavanje pretpostavke monotonosti. Kao primer, prikazana je funkcija odgovora na stavku 5 u zavisnosti od skorova ostatka (Slika 29). Na ovakvom grafikonu funkcija bi trebala da bude monotono neopadajuća (ne mora u svakoj tački rasti, ali ne sme opadati). U interpretaciji grafikona potrebno je uzeti u obzir osenčenu oblast poverenja (95%).

Kada je u pitanju indeks skalabilnosti za test, pravila njegove interpretacije su sledeća:

- $0{,}3 \le H_j < 0{,}4$ – slaba skala (ili skalabilnost stavke)
- $0{,}4 \le H_j < 0{,}5$ – srednja skala (ili skalabilnost stavke)
- $H_j \ge 0{,}5$ – jaka skala (ili skalabilnost stavke)
- a za skup stavki čiji je $H<0{,}3$ kažemo da nije skalabilan.

Indeksi skalabilnosti stavki u gornjem primeru su u opsegu prihvatljivih vrednosti, ali uglavnom slabi ili osrednji. Takođe, indeks skalabilnosti testa je prihvatljiv, ali slab.

Pored monotonosti KKS, možemo ispitati i da li se KKS ukrštaju, odnosno, proveriti invarijantnost poretka stavki. Ovo nije pretpostavka svih IRT modela već samo onih iz Rašove familije.

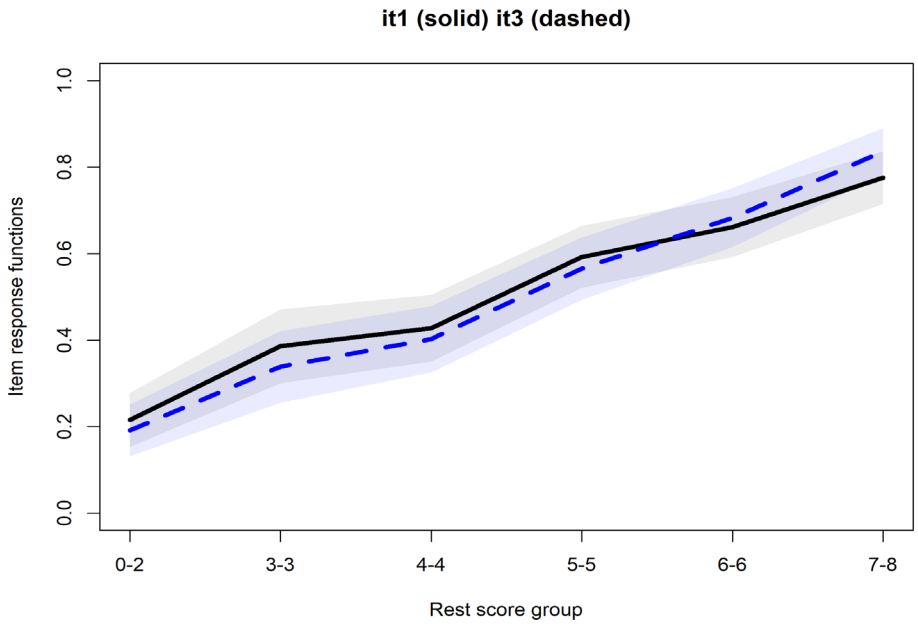
```
inv.poretka <- check.iio(b.pod)
summary(inv.poretka)$item.summary
```

```
##       ItemH #ac #vi #vi/#ac maxvi  sum sum/#ac zmax #zsig crit
## it9    0.44  50   1    0.02  0.07 0.07  0.0015 1.30     0   13
## it5    0.40  52   2    0.04  0.07 0.11  0.0021 1.30     0   19
## it6    0.39  51   1    0.02  0.03 0.03  0.0007 0.64     0    7
## it8    0.37  49   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it4    0.37  49   1    0.02  0.04 0.04  0.0008 0.81     0   10
## it1    0.32  48   3    0.06  0.07 0.17  0.0036 1.65     1   42
## it3    0.37  46   1    0.02  0.06 0.06  0.0013 1.35     0   16
## it2    0.42  48   1    0.02  0.07 0.07  0.0015 1.65     1   29
## it10   0.51  44   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it7    0.53  49   0    0.00  0.00 0.00  0.0000 0.00     0    0
```

Stavka koja ima upozoravajuću Crit vrednost je stavka 1.

Pošto stavka 1 ima upozoravajuću *Crit* vrednost, proverićemo grafikone sa parovima funkcija odgovora. Odabrali smo dva grafikona (od 45 mogućih) koji bi mogli biti zanimljivi. Na predstavljenim grafikonima vidimo ukrštanja funkcija odgovora kod stavki 1 i 3, ali se tu 95% intervali poverenja (osenčene površine) preklapaju (Slika 30).

```
plot(inv.poretka, ask = FALSE, item.pairs = 36:37)
```

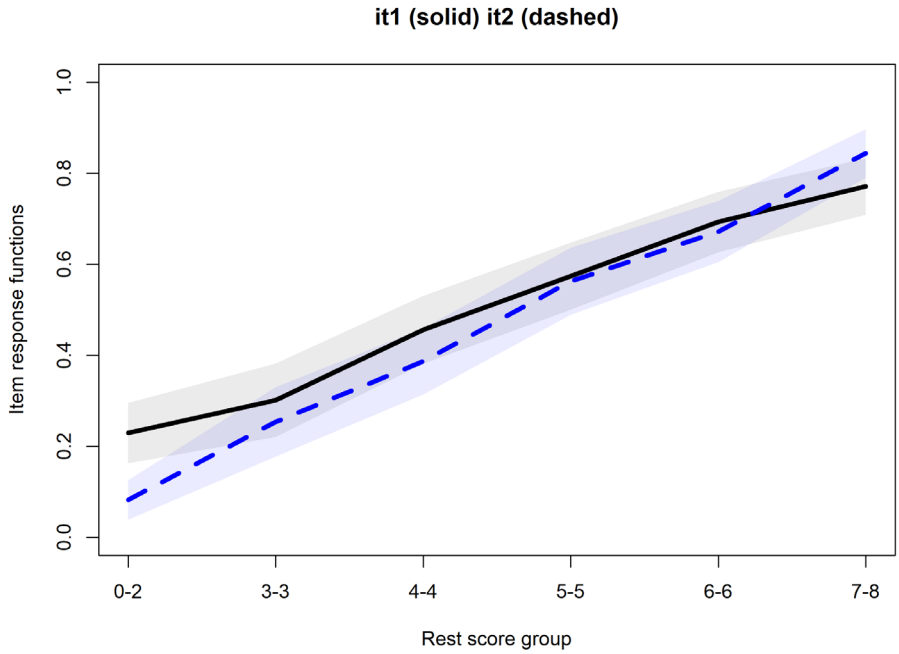



Slika 30 – Funkcije odgovora za stavke 1 (puna linija) i 3 (isprekidana linija)

```
# Ako ne navedemo argument item.pairs biće nacrtani svi parovi stavki. Argument
ask=FALSE određuje da ne morate da potvrdite crtanje svakog grafikona. Mi smo ovde
izdvojili samo one parove stavki koji su nam se učinili zanimljivim
```

Takođe, postoji ukrštanje funkcija odgovora stavki 1 i 2 (Slika 31), ali se intervali poverenja funkcija odgovora ove dve stavke ne preklapaju samo u grupi sa skorovima ostatka 0–2. Stavke su jednako teške u svim ostalim grupama, a u grupi skora ostatka 0–2 lakša je stavka 2. Iako je to jedino narušavanje koje je detektovano kao značajno (što možemo zaključiti iz prethodne tabele sa *Crit* vrednostima), ova razlika nije

nešto što bi trebalo da nas zabrine. Na ove stavke bi trebalo obratiti pažnju u jednoparametarskom modelu pošto do ukrštanja KKS dolazi zbog različitih nagiba.




Slika 31 – Funkcije odgovora za stavke 1 (puna linija) i 2 (isprekidana linija)

### 10.3.3 Lokalna nezavisnost

Proveru lokalne nezavisnosti na ovom mestu nećemo detaljno prikazivati. Da bismo zaista procenili lokalnu nezavisnost potrebno je prethodno proceniti željeni model. Načini provere lokalne nezavisnosti biće prikazani nakon procene pojedinih modela.

Ovde ćemo proceniti jednofaktorski model u paketu *psych()* koristeći funkciju *fa()* i nakon toga proveriti korelacije reziduala koristeći funkciju *residuals()*.

```
m1f <- fa(b.pod, nfactors=1, cor="tet")

# Jednofaktorski model na matrici tetrahoričkih korelacija
LDres <- residuals(m1f, diag = FALSE)
# Kreiranje matrice korelacija reziduala na modelu (bez vrednosti u dijagonali)
LDres[abs(LDres)<.2] <- NA
# Zbog preglednosti zamenićemo sve korelacije koje imaju apsolutnu vrednost <0,2
nedostajućim vrednostima
LDres

##       it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA
## it2   NA  NA
## it3   NA  NA  NA
## it4   NA  NA  NA  NA
## it5   NA  NA  NA  NA  NA
## it6   NA  NA  NA  NA  NA  NA
## it7   NA  NA  NA  NA  NA  NA  NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA
```

```
## it9  NA  NA  NA  NA  NA  NA  NA  NA  NA
## it10 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA

# Prikaz matrice korelacije reziduala
```

Kao što možemo videti nema stavki čije korelacije reziduala prelaze kritičnu apsolutnu vrednost od 0,2, pa možemo smatrati da je ispunjen uslov lokalne nezavisnosti. Ukoliko postoje takve korelacije trebalo bi izbaciti iz testa po jednu stavku iz parova koje pokazuju preveliku korelaciju reziduala.

Vrednost korelacije reziduala veća od 0,2 je upozoravajuća, ali ako ne prelazi puno tu vrednosti i broj parova koji koreliraju je mali to ne mora nužno narušiti saglasnost modela sa podacima. Postoje mišljenja da bi trebalo reagovati na korelacije koje su bar srednje veličine (>0,36) prema Koenovoj klasifikaciji (Cohen, 1988).

Ako gledamo pokazatelje fita za stavke u Rašovim modelima, stavke sa narušenom lokalnom zavisnošću će imati značajan prigušeni šum (overfit). Prigušeni šum se smatra manjim problemom i takve stavke se ne eliminišu uvek iz testa (Kean i ostali, 2018). Takođe, biranje stavki kod kojih pokazatelji fita idu u smeru prigušenog šuma rezultira pouzdanijim testom i nekada se preporučuje (Wu i ostali, 2016). Drugim rečima, ako koristimo pokazatelje fita zasnovane na standardizovanim rezidualima kao kriterijum za izbor stavki, bolje je birati stavke koji imaju MSQ u opsegu 0,7–0,9, nego one koje imaju MSQ=1.

## *10.4 IRT modeli za binarne stavke*

Ovo su modeli za stavke koje poseduju samo dve kategorije odgovora. Potrebno je da odgovori budu kodirani sa 0 i 1. Takav način kodiranja prirodan je za kognitivne testove gde se tačni odgovori skoruju sa 1, a netačni sa 0. Kod nekognitivnih testova gde kategorije mogu biti *tačno* i *netačno, ne slažem se* i *slažem se*, *ne* i *da*, ili neke druge dve kategorije, potrebno je odgovore kodirati na isti način. Takođe, stavke je potrebno kodirati da oznaka 1 uvek označava više prisustvo osobine, a 0 niže. Može i obrnuto, ali je potrebno da sve stavke budu kodirane na isti način.

### 10.4.1 Procena Rašovog modela u paketu eRm

Paket *eRm* (Mair & Hatzinger, 2007) namenjen je proceni modela iz Rašove familije modela. Kao metod procene parametara koristi metod uslovne maksimalne verodostojnosti (CMLE).

```
# Učitavanje biblioteke
p_load(eRm)
# Za procenu Rašovog modela za binarno skorovane stavke koristi se funkcija RM()
mr.e <- RM(b.pod)
# Procenjeni model smešten je u objekat mr.e

mr.e

##
## Results of RM estimation:
##
## Call:  RM(X = b.pod)
##
## Conditional log-likelihood: -2912.337
## Number of iterations: 11
```

```
## Number of parameters: 9
##
## Item (Category) Difficulty Parameters (eta):
##                  it2        it3        it4         it5        it6       it7
## Estimate 0.42166177 0.25293099 0.06048448 -1.57136771 -1.4411636 2.9151082
## Std.Err  0.07283497 0.07279473 0.07302969  0.08817746  0.0859907 0.1074704
##                   it8         it9       it10
## Estimate -0.88439814 -1.64752484 1.66388673
## Std.Err   0.07869732  0.08954728 0.08092088

# Pozivanjem imena objekta u koji smo upisali model dobijamo njegov ispis. Na ovaj
način ne dobijamo parametar za prvu stavku pošto je on fiksiran za potrebe
identifikacije modela
# Dobija se i vrednost LL (log-likelihood) koja može poslužiti za poređenje modela
summary(mr.e)

##
## Results of RM estimation:
##
## Call:  RM(X = b.pod)
##
## Conditional log-likelihood: -2912.337
## Number of iterations: 11
## Number of parameters: 9
##
## Item (Category) Difficulty Parameters (eta): with 0.95 CI:
##      Estimate Std. Error lower CI upper CI
## it2     0.422      0.073    0.279    0.564
## it3     0.253      0.073    0.110    0.396
## it4     0.060      0.073   -0.083    0.204
## it5    -1.571      0.088   -1.744   -1.399
## it6    -1.441      0.086   -1.610   -1.273
## it7     2.915      0.107    2.704    3.126
## it8    -0.884      0.079   -1.039   -0.730
## it9    -1.648      0.090   -1.823   -1.472
## it10    1.664      0.081    1.505    1.822
##
## Item Easiness Parameters (beta) with 0.95 CI:
##           Estimate Std. Error lower CI upper CI
## beta it1    -0.230      0.073   -0.373   -0.088
## beta it2    -0.422      0.073   -0.564   -0.279
## beta it3    -0.253      0.073   -0.396   -0.110
## beta it4    -0.060      0.073   -0.204    0.083
## beta it5     1.571      0.088    1.399    1.744
## beta it6     1.441      0.086    1.273    1.610
## beta it7    -2.915      0.107   -3.126   -2.704
## beta it8     0.884      0.079    0.730    1.039
## beta it9     1.648      0.090    1.472    1.823
## beta it10   -1.664      0.081   -1.822   -1.505

# Komandom summary() dobijamo rezime modela.
```

```
# Ispis težina za sve stavke možete dobiti na sledeći način:
cbind(mr.e$betapar*(-1))

##                  [,1]
## beta it1   0.23038215
## beta it2   0.42166177
## beta it3   0.25293099
## beta it4   0.06048448
## beta it5  -1.57136771
## beta it6  -1.44116363
## beta it7   2.91510819
## beta it8  -0.88439814
## beta it9  -1.64752484
## beta it10  1.66388673

# Pozvali smo beta parametre iz objekta koji sadrži model. Pošto se radi o
parametrima lakoće, pomnožili smo ih sa -1.
# Komanda cbind() služi samo da ih predstavimo u vertikalnom rasporedu
```

Pošto je Rašov model jednoparametarski jedini parametar koji ćemo dobiti je lokacija, odnosno težina stavke $b_j$. Niže vrednosti ukazuju na lakše, a više vrednosti na teže stavke. U našem primeru najlakša je stavka 9 ($b_9$=–1,65), a najteža stavka 7 ($b_7$=2,91). Stavke će biti najinformativnije upravo na ovim lokacijama.

Parametri težine nam dosta govore i o testu u celini. Na osnovu njihove distribucije znaćemo da li je test lak ili težak, a znaćemo i gde će biti najpreciznji, odnosno, najpouzdaniji i najinformativniji. Podsetićemo, informativnost testa predstavlja sumu informativnosti stavki i funkcija je nivoa osobine. Kako u ovom testu imamo veći broj laganih stavki, možemo pretpostaviti da će test biti informativniji za ispitanike sa nivoom osobine nešto ispod proseka.

#### *10.4.1.1 Procena parametara ispitanika*

*eRm* koristi CMLE metod procene parametara. Procena parametara stavki nezavisna je od procene parametara (nivoa osobine) ispitanika. Zato je njih potrebno posebno proceniti. Sačuvaćemo ih u objekat *mr.ePI* pošto će nam biti potrebni za neka kasnija izračunavanja.

```
mr.ePI <- person.parameter(mr.e)
mr.ePI

##
## Person Parameters:
##
##  Raw Score    Estimate Std.Error
##          0 -3.78203657        NA
##          1 -2.79705110 1.1020455
##          2 -1.87577564 0.8614377
##          3 -1.21335219 0.7782270
##          4 -0.63659070 0.7462584
##          5 -0.08534653 0.7428523
##          6  0.48010094 0.7658275
##          7  1.10634645 0.8235886
##          8  1.87214684 0.9380939
##          9  2.96908543 1.1939947
##         10  4.14580154        NA
```

U ovoj tabeli, krajnja leva kolona je sumacioni skor, u drugoj koloni su Rašove mere, a u poslednjoj njihove standardne greške merenja. Vidimo da svi ispitanici sa istim sumacionim skorom imaju jednake mere i standardne greške merenja. Takođe, možemo videti da standardne greške nisu iste za različite nivoe osobine, te da su manje što je mera ispitanika bliža prosečnoj (0 logita). Za savršene ukupne skorove (ispitanik odgovorio sve tačno ili sve netačno) nije moguće izračunavanje standardnih greški.

#### *10.4.1.2 Pokazatelj globalnog fita modela*

eRm za procenu globalnog fita modela računa Andersenov LR statistik.

```
mr.eLR <- LRtest(mr.e, splitcr = "mean")
# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
mr.eLR

##
## Andersen LR-test:
## LR-value: 13.732
## Chi-square df: 9
## p-value:  0.132
```

Pošto Andersenov LR test nije značajan, nećemo odbaciti model, odnosno, možemo reći da model u celini fituje.

Logika iza analize fita Andersenovim LR testom je invarijantnost procene parametara na različitim poduzorcima. Uzorak se deli na 2 ili više poduzoraka i proverava invarijantnost. Ukoliko neke kategorije odgovora nisu prisutne na nekoj stavci u poduzorcima, dobićete poruku ovakve sadržine: *Warning: The following items were excluded due to inappropriate response patterns within subgroups:* spisak stavki koje su isključene

10.4.1.2.1 Grafička provera modela

##### *10.4.1.2.1.1 Invarijantnost procene parametara stavki*

Invarijantnost procenjenih parametara stavki možemo grafički proceniti na osnovu sledeća dva grafikona. Na prvom su na apscisi predstavljene procene parametara u uzorku sa niskim nivoom osobine, a na ordinati sa visokim. Idealan slučaj bi bio kada bi se tačke koje predstavljaju procene parametara nalazile na crnoj liniji, ali je u redu ako se nalazi u okviru intervala poverenja oivičenog sivim linijama. Takođe, u redu je ako elipse koje predstavljaju 95% oblasti poverenja procena parametara obuhvataju crnu liniju. U oba slučaja ne možemo reći da se procene parametara razlikuju u dva poduzorka.

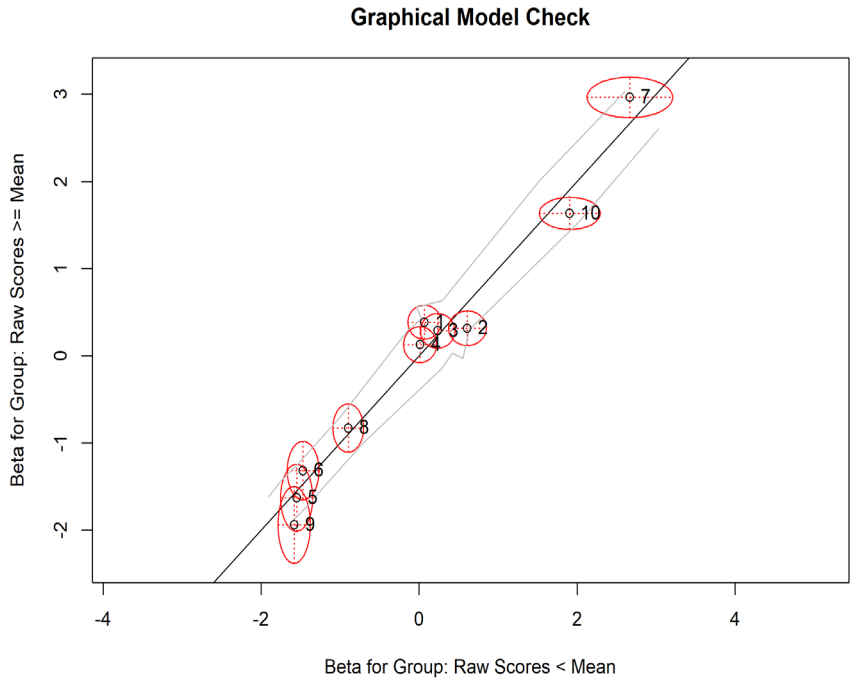
Grafikon (Slika 32) se zasniva na LR testu[49] i podela uzorka se radi na osnovu podele koja je u njemu korišćena.

Grafikon potvrđuje ono što nam je rekao LR test. Procene parametara na poduzorcima sa visokim i niskim skorovima se ne razlikuju značajno. Sve stavke se nalaze u okviru 95% intervala poverenja. Elipse

[49] Objekat funkcije *plotGOF()* je objekat sa LR testom – *mr.eLR*

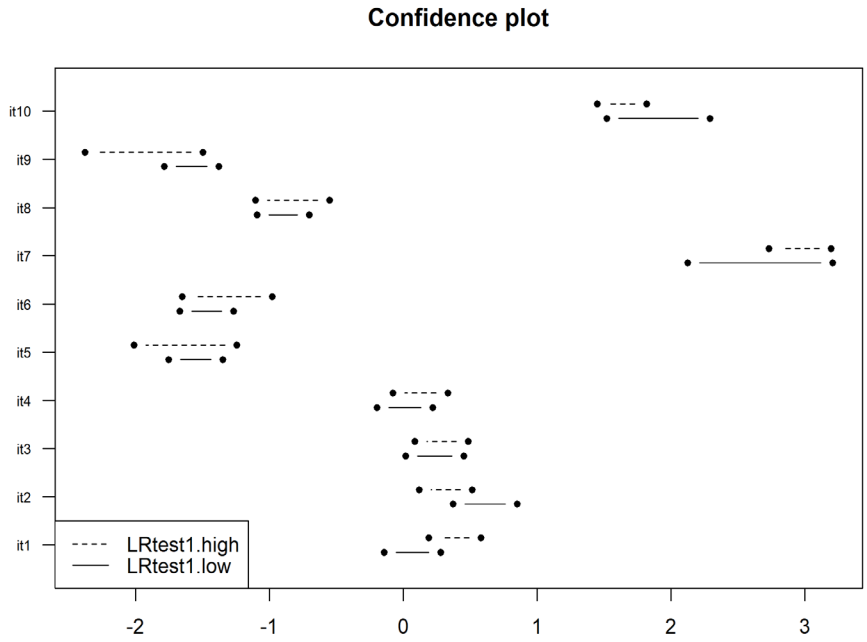
95% oblasti poverenja procena parametara za sve stavke obuhvataju centralnu liniju.

```
plotGOF(mr.eLR, ctrline = list(col = "gray"), conf = list(), tlab = "number")
```




Slika 32 – Grafička provera modela u paketu *eRm*

```
plotDIF(mr.eLR, leg=TRUE)
```




Slika 33 – Intervali poverenja procene parametara težine stavki na dva poduzorka

Sličan grafikon je onaj koji uporedo predstavlja intervale poverenja procene parametara u dva poduzorka (Slika 33). Ako se ovi intervali poverenja za dva poduzorka preklapaju, ne možemo reći da je invarijantnost procene parametara narušena. Kao i u prethodnom slučaju, funkcija *plotDIF()* koristi objekat u kojem smo sačuvali rezultate Andersenovog LR testa.

Vidimo da se 95% intervali poverenja procene parametara za sve stavke na dva poduzorka preklapaju (Slika 33), što znači da se procene parametara ne razlikuju značajno u uzorcima sa niskim i visokim skorovima. Stavka koja deluje najproblematičnije je stavka 1 gde se dva intervala poverenja sasvim malo preklapaju.

Širine intervala poverenja se razlikuju za lake i teške stavke. Lake stavke imaju široke intervale poverenja (veće standardne greške procene parametara) u poduzorku sa višim skorovima, a teške u poduzorku sa nižim skorovima. To je u skladu sa činjenicom da je merenje najpreciznije kada su stavke i ispitanici upareni.

#### *10.4.1.3 Pokazatelji fita za stavke*

U paketu eRm moguće je dobiti pokazatelje fita za stavke bazirane na hi–kvadratu i na standardizovanim rezidualima. Takođe, moguće je dobiti i Valdov test.

```
# Pošto i eRm i mirt imaju funkciju itemfit(), moguće je da funkcija neće raditi
ako ste posle eRm paketa učitali paket mirt. U tom slučaju morate ukloniti mirt
sa:
detach("package:mirt", unload=TRUE)
# ili umesto donje naredbe napisati eRm::itemfit(mr.ePI) i tako funkciju pozvati
direktno iz tog paketa
ItF <- itemfit(mr.ePI)
ItF

##
## Itemfit Statistics:
##         Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it1  1053.994 943   0.007      1.117     1.067    1.808   1.854   0.385
## it2   783.854 943   1.000      0.830     0.885   -2.823  -3.378   0.551
## it3   857.312 943   0.978      0.908     0.967   -1.489  -0.924   0.473
## it4   900.425 943   0.836      0.954     0.977   -0.714  -0.644   0.471
## it5   781.453 943   1.000      0.828     0.915   -1.346  -1.671   0.415
## it6   905.074 943   0.808      0.959     0.919   -0.299  -1.656   0.418
## it7   727.362 943   1.000      0.771     0.854   -1.165  -2.359   0.217
## it8   985.849 943   0.162      1.044     1.003    0.507   0.075   0.417
## it9   690.525 943   1.000      0.731     0.843   -2.118  -3.109   0.470
## it10  716.346 943   1.000      0.759     0.852   -2.489  -3.653   0.437

# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
WT=Waldtest(mr.e,splitcr = "mean")
# Argument splitcr ima isto značenje i opcije kao u funkciji LRtest()
WT

##
## Wald test on item level (z-values):
##
##           z-statistic p-value
```

```
## beta it1        2.162   0.031
## beta it2       -1.861   0.063
## beta it3        0.337   0.736
## beta it4        0.782   0.434
## beta it5       -0.347   0.728
## beta it6        0.763   0.446
## beta it7        0.985   0.324
## beta it8        0.399   0.690
## beta it9       -1.442   0.149
## beta it10      -1.252   0.210
```

U tabeli *Itemfit Statistics* prve tri kolone odnose se na hi–kvadrat test. p vrednosti ispod 0,05 ukazuju na misfit. Takođe, date su vrednosti infita i autfita. Vrednosti koje izlaze iz okvira 0,7–1,3 ukazuju na misfit. Vrednosti manje od 0,7 na prigušeni šum (overfit), a preko 1,3 na šum. Pored toga, prikazane su i standardizovane vrednosti ovih pokazatelja. Niže od –1,96 ukazuju na prigušeni šum, a preko 1,96 na šum. Veći značaj bi trebalo pridavati pokazateljima u MSQ metrici (videti odeljak 6.3.2.3). U poslednjoj koloni date su diskriminativnosti izračunate kao u klasičnoj testnoj teoriji. Ukoliko postoji misfit kod neke stavke, znatno drugačija diskriminativnost može biti izvor. Podsetićemo, jednoparametarski model predviđa da su nagibi stavki ujednačeni.

Pokazatelji misfita se mogu koristiti kao kriterijum za izbor stavki u test. Zadržaćemo stavke koje imaju pokazatelje misfita u preporučenom opsegu, a izbacićemo one čiji su pokazatelji izvan njega. Setite se da preporučene granice prihvatljivog opsega zavise od vrste testa i njegove namene (videti odeljak 6.3.2.3).

Takođe, setite se da je za pouzdanost testa dobro da pokazatelji misfita budu bliže donjoj granici prihvatljivog opsega (prema overfitu)(Wu i ostali, 2016).

Na ovom mestu, napomenućemo da su pokazatelji infita i autfita relativni. Ako neka stavka ima veći nagib od drugih stavki u testu, pokazatelji misfita će naginjati overfitu, a ako stavka ima manji nagib od ostalih, pokazatelji će ukazivati na šum. Ako stavku koja je u okviru jednog skupa stavki pokazivala overfit, stavimo u drugi test gde stavke imaju jednake nagibe kao ona, stavka će imati savršene pokazatelje fita (Wu i ostali, 2016).

Takođe, trebalo bi napomenuti da pokazatelji infita i autfita zavise od veličine uzorka. Ako stavke isključujemo na osnovu misfita i koristimo MSQ pokazatelje, na velikim uzorcima pokazatelji za većinu stavki će biti uglavnom dobri. S druge strane, ako koristimo standardizovane pokazatelje, odbacili bismo veliki broj stavki (Wu i ostali, 2016).

U našem primeru stavka 1 ima značajan hi–kvadrat test što ukazuje na značajan misfit. Međutim, vrednosti infita i autfita su u preporučenom opsegu (0,7–1,3), a čak ni standardizovani infit i autfit ne izlaze iz preporučenog opsega. Uz stavku 8 ovo je jedina stavka čiji pokazatelji misfita idu u pravcu šuma, a ne overfita. Primetno je da ova stavka ima nižu diskriminativnost od većine ostalih stavki (najnižu ima stavka 7), pa je moguće da ne meri isti konstrukt kao ostale stavke (bar ne u celini).

Stavke 2, 7, 9 i 10 imaju standardizovani infit i autfit koji izlazi iz preporučenog opsega (izuzetak je autfit za stavku 7). Pošto su sve vrednosti niže od –1,96, možemo reći da se radi o suviše predvidivom odgovaranju, odnosno, prigušenom šumu ili overfitu. Podsetićemo da postoji preporuka da se standardizovana

varijanta tumači tek ako i MSQ izlazi izvan prihvatljivog opsega. Takođe, veći značaj bi trebalo pridati infitu (videti odeljak 6.3.2.3).

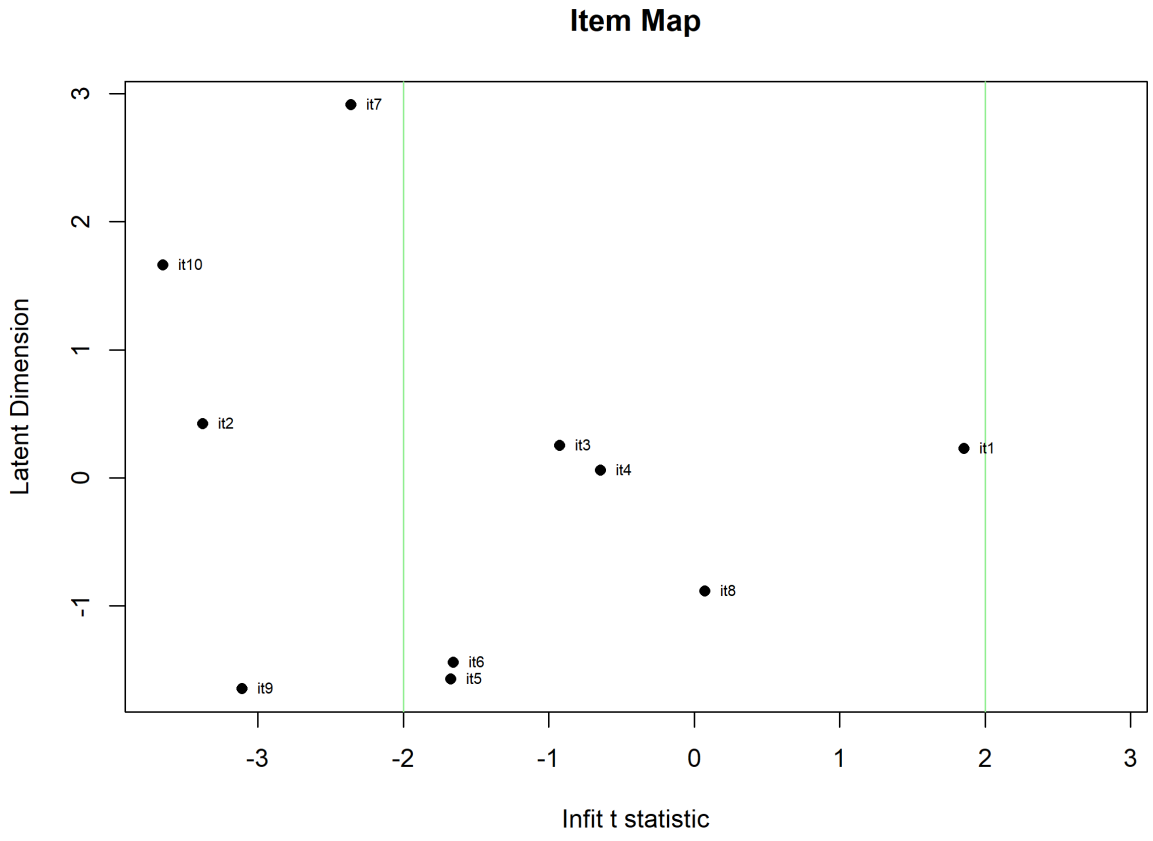
Kod Valdovog testa p<0,05 signalizira narušavanje invarijantnosti procene parametara, pa samim tim i na misfit. I Valdov test je značajan za stavku 1.

#### 10.4.1.3.1 Grafički prikaz standardizovanog infita za stavke

Sledeći grafikon prikazuje standardizovani infit za stavke. Prihvatljivi opseg infita se nalazi između dve vertikalne (zelene) linije. Levo se nalazi prigušeni šum ili overfit, a desno šum, odnosno nepredvidivo odgovaranje.

```
plotPWmap(mr.e, imap=TRUE, pmap=FALSE)
```

```
# Argument imap=TRUE određuje da će biti prikazan infit za stavke, a pmap=FALSE da
neće biti prikazan za ispitanike. Moguće je napraviti isti grafikon i za
ispitanike (imap=FALSE, pmap=TRUE), ili grafikon koji će objediniti i stavke i
ispitanike (imap=TRUE, pmap=TRUE)
# Ukoliko želite grafikon i za ispitanike potrebno je dodati argument pp=mr.ePI
# mr.ePI je objekat u kojem smo sačuvali parametre ispitanika
```




Slika 34 – Standardizovani infit za stavke

Na grafikonu (Slika 34) vidimo da problematična stavka 1 ne izlazi iz okvira prihvatljivog infita, mada naginje šumnom.

S druge strane, vidimo da četiri stavke imaju neprihvatljiv standardizovani infit tipa prigušenog šuma, odnosno suviše predvidljivo odgovaranje. To su stavke 2, 7, 9 i 10. Prigušeni šum može ukazivati na lokalnu zavisnost ili postojanje testleta. Pošto su ovo simulirani podaci, mi se ne možemo baviti otkrivanjem razloga overfita. Kod simuliranih podataka overfit je čest slučaj.

#### *10.4.1.4 Procena lokalne zavisnosti – T11 i T1l*

Neparametarski test za procenu lokalne zavisnosti na nivou testa $T_{11}$ bazira se na sumi odstupanja između opaženih i očekivanih ajtemskih korelacija. Koristi se metoda samouzorkovanja (bootstrapping) kojom se iz matrice podataka izvlači željeni broj uzoraka iste veličine kao originalni (uzorkovanje sa povratom).

```
set.seed(ss)
T11 <- NPtest(as.matrix(b.pod), n = 1000, method = "T11")
# Objekat funkcije mora biti u obliku matrice pa je objekat tipa data.frame
pretvoren u matricu pomoću funkcije as.matrix().
# Argument n = 1000 označava željeni broj bootstrapp uzoraka.
T11

## Nonparametric RM model test: T11 (global test - local dependence)
## (sum of deviations between observed and expected inter-item
##     correlations)
## Number of sampled matrices: 1000
## one-sided p-value: 0.004
```

Ako je prethodni test značajan, onda bi ugrožavanje lokalne nezavisnosti moglo ugrožavati fit modela. U tom slučaju bi trebalo proveriti koje stavke su izvor lokalne zavisnosti.

U našem primeru $T_{11}$ je statistički značajan što ukazuje na postojanje lokalne zavisnosti, a to je u skladu i sa prethodnim nalazom o overfitu 4 stavke.

U tu svrhu možemo iskoristiti neparametarski test lokalne zavisnosti za stavke $T_{1l.}$ Bazira se na povećanim korelacijama među stavkama. Broji slučajeve (ispitanike) koji imaju iste odgovore na parovima stavki i prijavljuje one kod kojih je ta povezanost značajna.

```
set.seed(ss)
T1 <- NPtest(as.matrix(b.pod), n = 1000, method = "T1l")
T1

## Nonparametric RM model test: T1 (learning - based on item pairs)
##     (counting cases with reponsepattern (1,1) for item pair)
## Number of sampled matrices: 1000
## Number of sampled matrices: 1000
## Number of Item-Pairs tested: 45
## Item-Pairs with one-sided p < 0.05
##  (2,6)  (2,7)  (2,9)  (5,9) (9,10)
##  0.008  0.016  0.017  0.045  0.010
```

Rezultati ovog testa potvrđuju ono što smo videli kroz pokazatelje misfita za stavke. Iz testa bi trebalo ukloniti po jednu stavku iz svakog para, a ako se neka stavka javlja u više parova, najekonomičnije je ukloniti tu stavku. U ovom slučaju to bi bile stavke 2 i 9.

Nakon uklanjanja stavki koje narušavaju lokalnu nezavisnost, dobro je proveriti ponovo globalni

pokazatelj testom T11.

Još jedan pokazatelj lokalne (ne) zavisnosti je pokazatelj $Q_3$ koji predstavlja produkt–moment korelacije reziduala.

```
TQ3 <- NPtest(as.matrix(b.pod),n = 1000, method = "Q3h")
# TQ3 je objekat tipa list koji sadrži tri elementa. U elementu prop nalaze se
p vrednosti korelacija reziduala, a koje se nalaze u elementu Q3hmat
# Pošto je element prop vektor, a Q3hmat matrica, da bismo ih uporedili
pretvorićemo prop u matricu # Prvo ćemo kreirati matricu u kojoj su sve vrednosti
nedostajuće (NA), a ima
redova i kolona koliko imamo stavki
mat <- matrix(NA, nrow = 10, ncol = 10)
# Zatim ćemo vrednosti iznad dijagonale popuniti vrednostima vektora prop
mat[upper.tri(mat, diag = FALSE)] <- TQ3$prop
# Pošto matrica Q3h ima vrednosti samo ispod dijagonale, matricu p vrednosti ćemo
transponovati
mat <- t(mat)
# Iako postoji i funkcija lower.tri koja vektor smešta direktno ispod dijagonale,
ona ne daje ispravan raspored elemenata pa smo ovo odradili zaobilaznim putem

# Zbog preglednosti ćemo sve p vrednosti koje nisu manje od 0.05 pretvoriti u
nedostajuće vrednosti
mat[!mat<0.05] <- NA
# Ako su nam u matrici ostale p vrednosti manje od 0,05, očitamo broj reda i broj
kolone i to će nam reći kod kojih parova stavki postoji lokalna zavisnost
# Kolika je možemo videti iz matrice Q3hmat
LD <- TQ3$Q3hmat
LD
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
##  [1,]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##  [2,] 0.979    NA    NA    NA    NA    NA    NA    NA    NA    NA
##  [3,] 0.336 0.283    NA    NA    NA    NA    NA    NA    NA    NA
##  [4,] 0.157 0.658 0.880    NA    NA    NA    NA    NA    NA    NA
##  [5,] 0.740 0.287 0.298 0.173    NA    NA    NA    NA    NA    NA
##  [6,] 0.822 0.008 0.648 0.331 0.210    NA    NA    NA    NA    NA
##  [7,] 0.855 0.009 0.440 0.412 0.998 0.840    NA    NA    NA    NA
##  [8,] 0.617 0.315 0.369 0.682 0.925 0.789 0.621    NA    NA    NA
##  [9,] 0.720 0.354 0.911 0.592 0.162 0.897 0.142 0.058    NA    NA
## [10,] 0.139 0.959 0.327 0.720 0.322 0.261 0.329 0.619 0.055    NA
```

```
# ili
LD[is.na(mat)] <- NA
LD
```

```
##       [,1] [,2]  [,3]  [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [2,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [3,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [4,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [5,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [6,]   NA   NA 0.648 0.331   NA   NA   NA   NA   NA    NA
##  [7,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [8,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
##  [9,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
## [10,]   NA   NA    NA    NA   NA   NA   NA   NA   NA    NA
```

Vrednosti $Q_3$ pokazatelja veće od 0,2 smatraju se upozoravajućim, dok se neki autori (Zhang i ostali,

2019) pozivaju na Koenovu (Cohen, 1988) podelu veličine efekta kod koje na vrednosti veće od 0,36 (srednji efekat) treba obratiti pažnju. Za više informacija o interpretaciji $Q_3$ pogledajte odeljak 9.2.

$Q_3$ nam pokazuje da su značajne korelacije reziduala između stavki 4 i 6, te 3 i 6. Najviša korelacija reziduala je između stavki 3 i 6. Kada bismo želeli da uklonimo neku od stavki kako bi se rešili problema lokalne zavisnosti, najekonomičnije bi bilo izbaciti stavku 6 koja se nalazi u oba para koja nezanemarljivo koreliraju rezidualima.

Iako se $T_{1l}$ i $Q_3$ pokazatelji slažu da je narušena lokalna nezavisnost, ne slažu se o kojim stavkama se radi.

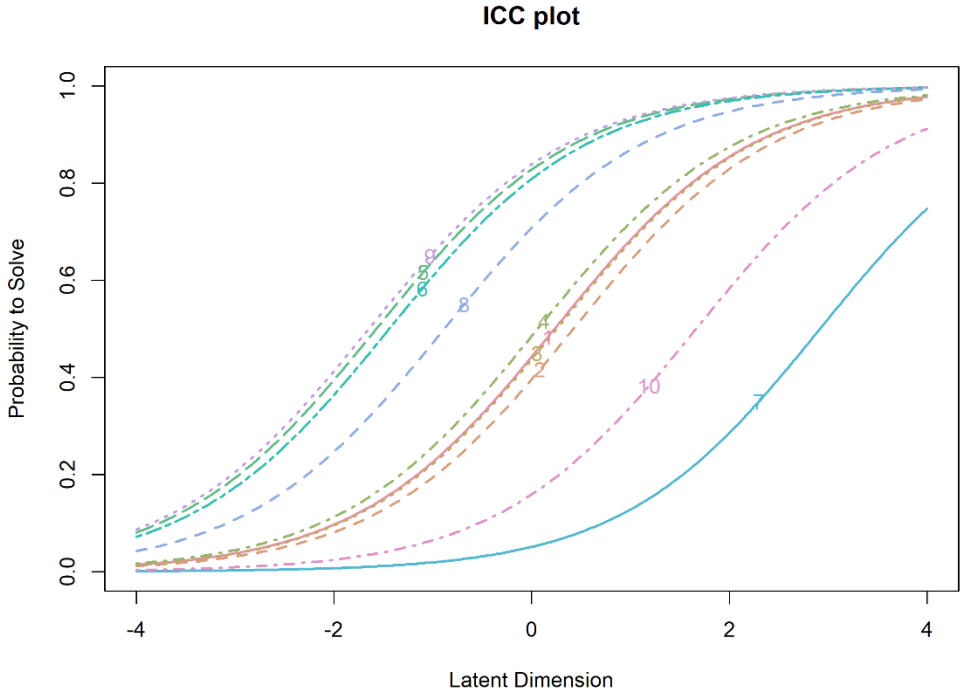
Podsetićemo da provera lokalne zavisnosti na jednofaktorskom modelu nije ukazala da postoji izražena lokalna zavisnost.

#### *10.4.1.5 Grafički prikaz karakterističnih kriva stavki*

Grafički prikaz karakterističnih kriva stavki u Rašovom modelu nije naročito informativan (Slika 35). Sve krive izgledaju isto (isti nagib, isti odsečak) samo se razlikuje njihova lokacija. U poslednje vreme se češće prikazuju krive (funkcije) informativnosti stavki.

U paketu eRm možete prikazati krive za sve stavke.

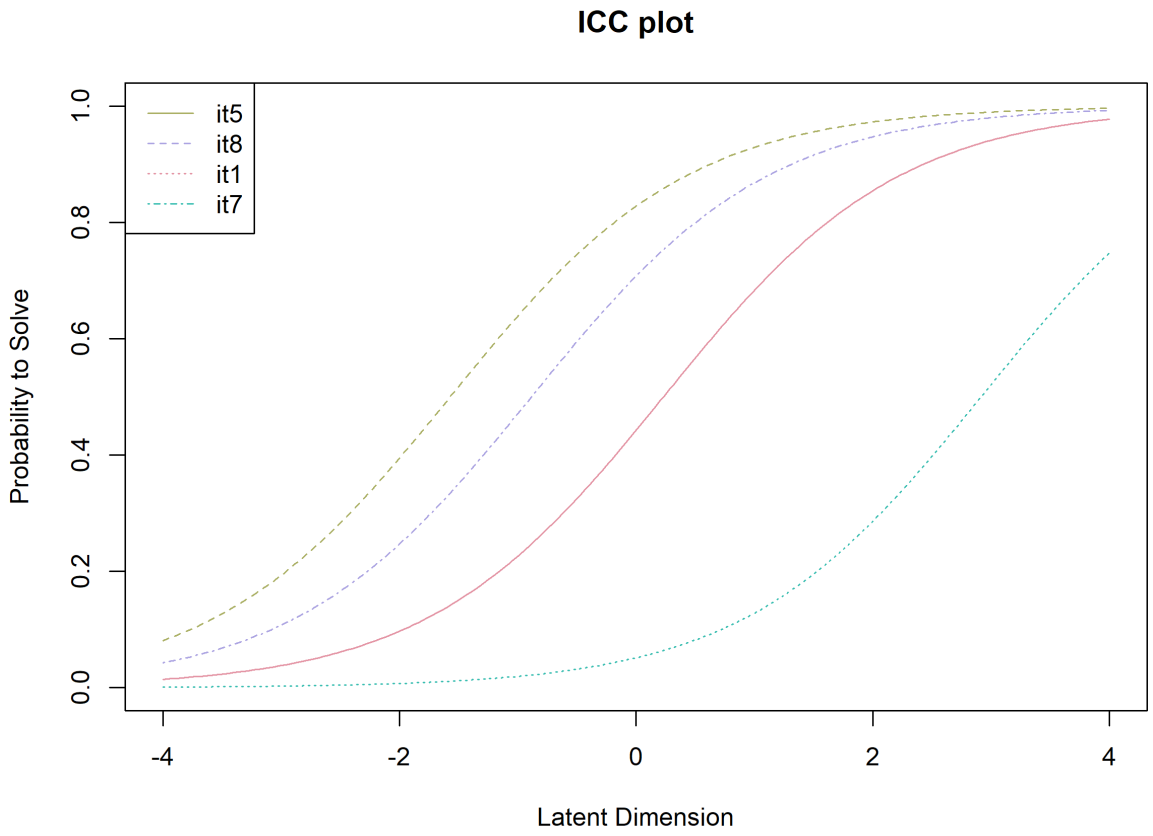
```
plotjointICC(mr.e, item.subset = "all", legend = F, lty=c(1:10), lwd=2)

# Argument lwd=2 određuje debljinu linija (bez njega debljina je 1)
# Argument lty=c(1:10) daje instrukciju da se koriste različiti tipovi linija
# c(1:10) je vektor brojeva od 1 do 10 c(1:10)
```




Slika 35 – Karakteristične krive stavki u Rašovom modelu (sve stavke)

Možete prikazati krive samo za određene stavke.

```
plotjointICC(mr.e, item.subset = c(1, 5, 7, 8), legend = T, lty=c(1:4), lwd=1)

# Argument item.subset = c(1, 5, 7, 8) određuje da budu prikazane 1, 5, 7. i 8. stavka
# Argument legend = T zahteva ispis legende
```




Slika 36 – Karakteristične krive odabranih stavki

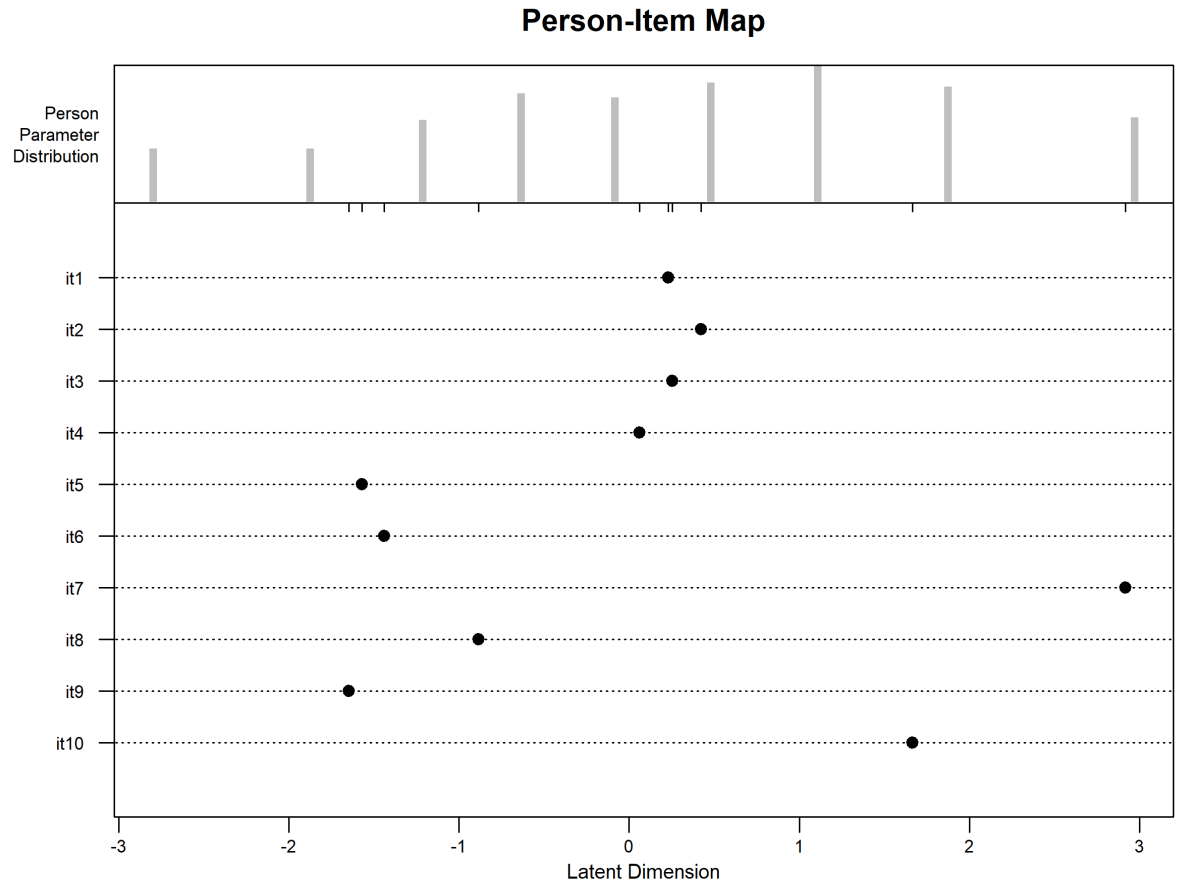
#### *10.4.1.6 Uporedni prikaz distribucije parametara ispitanika i stavki*

Iako, ako Rašov model fituje, procene parametara stavki ne bi smele da zavise od distribucije osobine u uzorku ispitanika, standardne greške će zavisiti. Bolje će biti procenjeni parametri u delovima kontinuuma osobine koji su pokriveni ispitanicima sa uparenim nivoima osobine.

```
plotPImap(mr.e, sorted = FALSE)
```

```
cbind("b"=mr.e$betapar*(-1),"se b"=mr.e$se.beta )

##                       b       se b
## beta it1   0.23038215 0.07280670
## beta it2   0.42166177 0.07283497
## beta it3   0.25293099 0.07279473
```

```
## beta it4   0.06048448 0.07302969
## beta it5  -1.57136771 0.08817746
## beta it6  -1.44116363 0.08599070
## beta it7   2.91510819 0.10747040
## beta it8  -0.88439814 0.07869732
## beta it9  -1.64752484 0.08954728
## beta it10  1.66388673 0.08092088

# Gornjom komandom smo iz objekta mr.e sa Rašovim modelom izvukli parametre lakoće
(betapar) i množenjem sa -1 pretvorili u težine, i spojili u tabelu zajedno sa
pripadajućim standardnim greškama (se.beta). Kolonama smo dodelili nazive "b" i
"se b"
```




Slika 37 – Uporedni prikaz distribucije parametara stavki i ispitanika

Sa grafikona (Slika 37) vidimo da su ispitanicima slabije pokriveni viši nivoi osobine, pa su i standardne greške procene parametara veće za teže stavke (one sa višim *b* – it7).

Osim toga, ovaj grafikon nam može poslužiti da vidimo koji nivoi osobine u populaciji koju želimo da merimo nisu dovoljno pokriveni stavkama testa. Ovo je povezano sa informativnošću, a u IRT modelima informativnost je mera preciznosti merenja i recipročna je kvadratu standardne greške (odeljak 7).

Naš test je prilagođeniji ispitanicima sa prosečnim i ispodprosečnim nivoima osobine. Veći broj stavki ima težinu oko 0 logita i nižu (Slika 37). Ako želimo da prilagodimo test opštoj populaciji, vredelo bi razmisliti o dodavanju stavki koje bi bile teške između 0,5 i 2,5 logita.

U Rašovom modelu stavke su najinformativnije na lokaciji koja odgovara parametru težine stavke (odeljci 4.1.1 i 7). Birajući stavke za test u skladu sa parametrom težine istovremeno biramo deo kontinuuma latentne osobine u kojem će merenje biti najpreciznije.

Populacija koju želimo da merimo ne mora biti opšta. To mogu biti i osobe sa visokim nivoom osobine ili one sa ispodprosečnim. U tom slučaju u test možemo birati stavke koje su njima prilagođene. Nema smisla ispitanicima sa visokim nivoom osobine davati jako lagane stavke jer znamo da će na njih odgovoriti tačno (i obrnuto).

Kada pravimo test moguće su različite strategije:

1. da bude precizan u nekom uskom kritičnom rasponu osobine
2. da bude precizan u širokom rasponu osobine (uniformno) i
3. da ima normalnu raspodelu parametara težina stavki

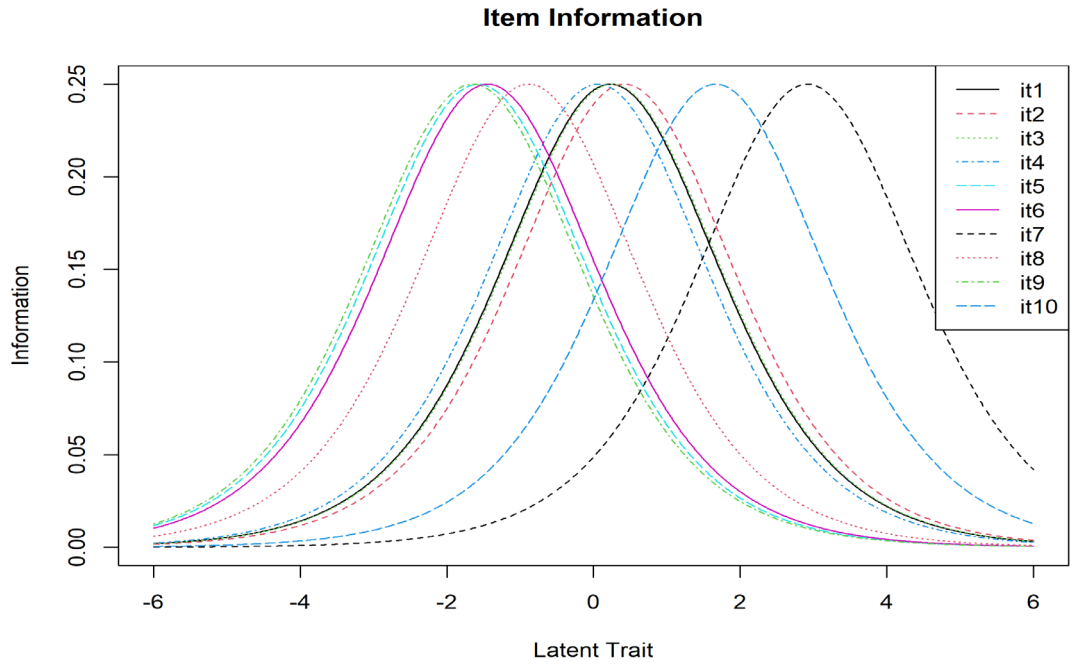
Poslednja dva načina se malo razlikuju u preciznosti, a lakše je praviti uniforman test (Wright and Stone 1979).

Jedno od jednostavnijih pravila je da interval osobine koji stavke testa pokrivaju bi trebalo da bude 4 puta veći od standardne devijacije merene osobine u ciljnoj populaciji (Wright and Stone 1979).

#### *10.4.1.7 Informativnost stavki i testa*

Ponovićemo još jednom, u IRT modelima informativnost je najvažniji pokazatelj preciznosti i kvaliteta merenja. Stavke će biti najinformativnije u merenju nivoa osobine koji odgovara njihovoj lokaciji, odnosno parametru težine stavke. Test će biti najinformativniji kada je prilagođen ispitaniku, ni prelak – ni pretežak.

```
plotINFO(mr.e, type="item")
```



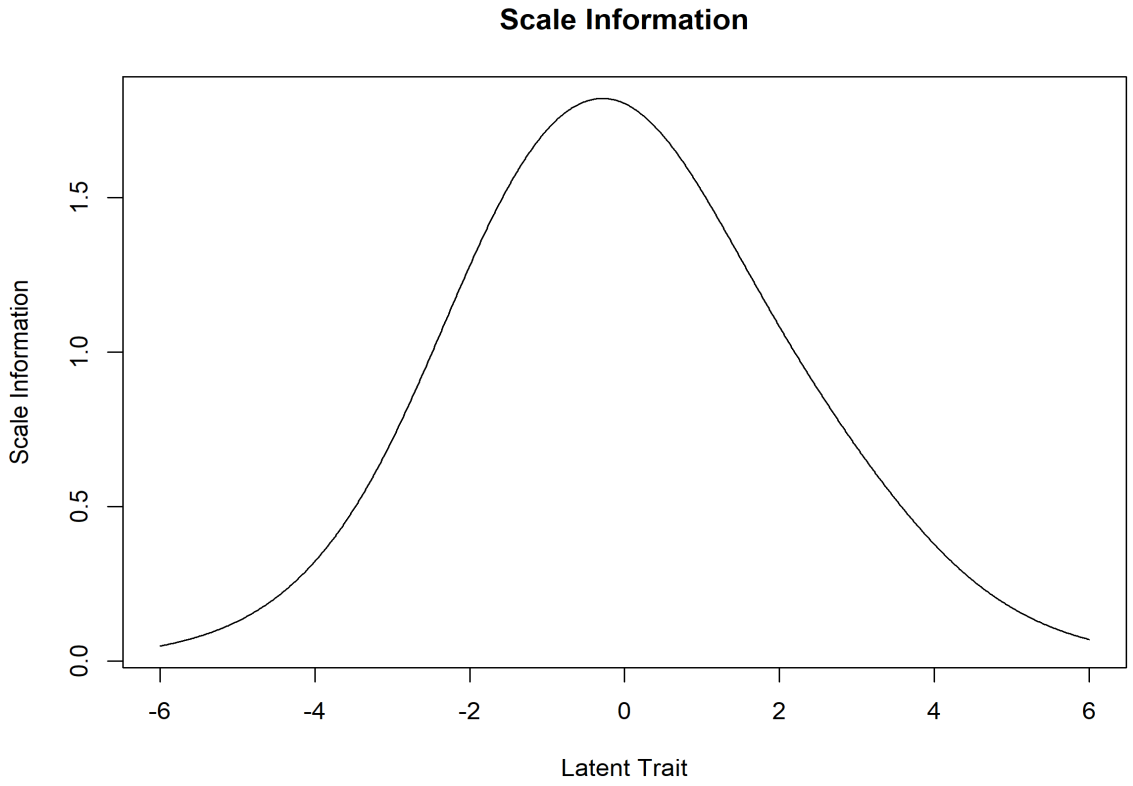

Slika 38 – Krive informativnosti stavki u Rašovom modelu

Uz pomoć grafikona informativnosti stavki (Slika 38) i testa (Slika 39) postaje očiglednije ono što je ranije rečeno. Na njima možemo videti u kojem delu kontinuuma latentne osobine je merenje najpreciznije.

Na grafikonima za stavke vidimo gde je koja stavka najpreciznija, a na grafikonu za test vidimo istu informaciju za test.

U Rašovom modelu za binarno skorovane stavke (Slika 38), sve stavke imaju jednaku maksimalnu informativnost (0,25) samo je pitanje lokacije na kojoj će biti najinformativnije, a to je upravo lokacija stavke, odnosno $b_j$.

```
plotINFO(mr.e, type="test")
```




Slika 39 – Kriva informativnosti testa

Informativnost testa odgovara sumi informativnosti stavki i visina funkcije informativnosti na pojedinim lokacijama zavisi od pokrivenosti tog dela kontinuuma latentne osobine stavkama. Deo kontinuuma u kojem želimo da merenje bude informativnije trebalo bi pokriti većim brojem stavki odgovarajuće težine.

Iz gornjeg grafikona (Slika 39) vidimo da je naš test najinformativniji na nivou osobine nešto nižem od prosečnog (tačnije –0,3 logita, a kako smo do toga došli vidite u sledećem odeljku).

#### 10.4.1.7.1 Numerički pokazatelji informativnosti

Informativnost se može prikazati i numerički. Pokazaćemo na primeru informativnosti testa, ali je moguće isto učiniti za svaku stavku.

Pošto je informativnost funkcija nivoa osobine potrebno je zadati raspon $\theta$.

```
TETE <- seq(-3, 3, 0.1)
# Kreiramo vektor sa željenim rasponom i nivoima tete
# Sekvenca od -3 do 3 u koracima po 0.1
TINFO <- test_info(mr.e, theta=TETE)
# Vektor TETE prosleđujemo argumentu theta funkcije test_info()
# Rezultat je vektor informativnosti testa za pojedine nivoe latentne osobine
# Ispis tabele informativnosti po nivoima osobine
TI <- data.frame(cbind(TETE, TINFO))
# Sa funkcijama cbind() i data.frame() povezali smo vektore nivoa osobine i
informativnost u matricu podataka
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n", TI[TI[,2]==max(TI[,2]),1], "logita")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  -0.3 logita

cat("\nInformativnost u rasponu od -3 do 3 logita: \n")

##
## Informativnost u rasponu od -3 do 3 logita:

inf33 <- sum(TINFO)
inf33

## [1] 82.47289

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 3 logita:  \n")

##
## Informativnost u rasponu od 0 do 3 logita:

inf03 <- sum(test_info(mr.e, theta=seq(0, 3, 0.1)))
inf03

## [1] 39.91717

cat(" \nProcenat informativnosti u rasponu od 0 do 3 logita, u odnosu na raspon od
-3 do 3 logita:  \n")

##
## Procenat informativnosti u rasponu od 0 do 3 logita, u odnosu na raspon od
-3 do 3 logita:

inf03/inf33*100

## [1] 48.40036
```

Vidimo da je informativnost nešto manja (48,40%) u opsegu natprosečnih nivoa osobine nego kod onih ispod proseka (51,60%).

Koristeći funkciju *item_info* možemo proceniti informativnost stavki za različite nivoe latentne osobine. Ove podatke možemo iskoristiti za izbor stavki u test. Biće prikazan primer računanja prosečne informativnosti stavki za tri tačke pomenute u odeljku 7.1. To su tačke –1, 1 i 2 logita (mogu biti i neke druge, ali su ove uzete za primer). Potrebno je definisati vektor nivoa latentne osobine koji nas zanimaju. U primeru je to objekat *tete*. Rezultat će biti objekat tipa lista (nazvali smo ga *it_inf*) koji sadrži informativnosti

svih stavki i njihovih kategorija na zadatim nivoima latentne osobine. Pomoću indeksa *[[i]]$i.info* gde *i* predstavlja broj stavke, pozvali smo vektore sa vrednostima informativnosti za pojedinačne stavke, a funkcijom *mean()* izračunali njihove prosečne vrednosti.

Koristili smo petlju *for* da izračunamo prosečne informativnosti za sve stavke.

```
tete <- c(-1, 1, 2)
it_inf <- item_info(mr.e, theta=tete)
for(i in 1:ncol(b.pod)){
ii <- mean(it_inf[[i]]$i.info)
cat(paste0("Prosečna informativnost it", i, " za lokacije ", paste(tete,
 collapse=", "), " logita iznosi: ", round(ii,2), "\n"))
}
```

```
## Prosečna informativnost it1 za lokacije -1, 1, 2 logita iznosi: 0.17
## Prosečna informativnost it2 za lokacije -1, 1, 2 logita iznosi: 0.18
## Prosečna informativnost it3 za lokacije -1, 1, 2 logita iznosi: 0.17
## Prosečna informativnost it4 za lokacije -1, 1, 2 logita iznosi: 0.17
## Prosečna informativnost it5 za lokacije -1, 1, 2 logita iznosi: 0.11
## Prosečna informativnost it6 za lokacije -1, 1, 2 logita iznosi: 0.11
## Prosečna informativnost it7 za lokacije -1, 1, 2 logita iznosi: 0.11
## Prosečna informativnost it8 za lokacije -1, 1, 2 logita iznosi: 0.14
## Prosečna informativnost it9 za lokacije -1, 1, 2 logita iznosi: 0.1
## Prosečna informativnost it10 za lokacije -1, 1, 2 logita iznosi: 0.18
```

Najvišu prosečnu informativnost za ova tri nivoa latentne osobine imaju stavke 2 i 10, a najnižu stavka 9.

#### *10.4.1.8 Eliminacija stavki*

Paket eRm poseduje funkciju za automatsku eliminaciju stavki korak po korak, po nekom od kriterijuma fitovanja (itemfit – hi–kvadrat test, LR test ili na osnovu Valdovog statistika). Stavke se uklanjaju jedna po jedna na osnovu značajnog misfita.

```
itemfit(mr.ePI)
```

```
##
## Itemfit Statistics:
##         Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it1  1053.994 943   0.007      1.117     1.067    1.808   1.854   0.385
## it2   783.854 943   1.000      0.830     0.885   -2.823  -3.378   0.551
## it3   857.312 943   0.978      0.908     0.967   -1.489  -0.924   0.473
## it4   900.425 943   0.836      0.954     0.977   -0.714  -0.644   0.471
## it5   781.453 943   1.000      0.828     0.915   -1.346  -1.671   0.415
## it6   905.074 943   0.808      0.959     0.919   -0.299  -1.656   0.418
## it7   727.362 943   1.000      0.771     0.854   -1.165  -2.359   0.217
## it8   985.849 943   0.162      1.044     1.003    0.507   0.075   0.417
## it9   690.525 943   1.000      0.731     0.843   -2.118  -3.109   0.470
## it10  716.346 943   1.000      0.759     0.852   -2.489  -3.653   0.437
```

```
ELIMINACIJA=stepwiseIt(mr.e, criterion = list("itemfit"))
```

```
## Eliminated item - Step 1: it1
```

```
ELIMINACIJA

##
## Results for stepwise item elimination:
## Number of steps: 1
## Criterion: Itemfit
##
##                Chisq  df p-value
## Step 1: it1 1053.994 944   0.007

# Mogući argumenti criterion=list("Wald") ili list("LRtest")
ELIMINACIJA=stepwiseIt(mr.e, criterion = list("LRtest"))

## Eliminated item - Step 1: it1

ELIMINACIJA

##
## Results for stepwise item elimination:
## Number of steps: 1
## Criterion: Andersen's LR-test
##
##              LR-value Chisq df p-value
## Step 1: it1    17.678        9   0.039
## Step 2: None    9.084        8   0.335
```

Procedura za automatsku eliminaciju predlaže uklanjanje stavke 1 na osnovu značajnog $\chi^2$ i LR testa.

Ako sve stavke imaju zadovoljavajuće pokazatelje fita nijedna neće biti eliminisana. Dobićete poruku: *Warning: No items eliminated! Each of them fits the rasch model!*

#### *10.4.1.9 Izbacivanje stavki*

Kada koristimo paket *eRm* trebalo bi imati u vidu da on koristi CMLE procenu parametara stavki. Podsetićemo, u svim IRT modelima nepoznati su nam i parametri stavki i parametri ispitanika. *eRm* koristeći CMLE procenu prvo procenjuje parametre stavki. Kada su nam parametri stavki poznati, na osnovu njih, koristeći JMLE procenu *eRm* procenjuje parametre ispitanika.

Kada se odlučimo da izbacimo neku stavku, promeniće se i ukupni skorovi i nova procena parametara stavki daće druge rezultate. Zato bi stavke trebalo uklanjati po jednu u svakom koraku.

Zamislimo sada da smo rešili da izbacimo stavku 1 kako nam predlaže funkcija *stepwiseIt()*. Izbacićemo je iz matrice podataka i ponovo proceniti model.

Kada uklanjamo stavke nije nužno da ih brišemo iz podataka. Dovoljno je da u indeksu kolone matrice podataka navedemo redne brojeve stavki koje zadržavamo, kao u donjem primeru. Drugi način je da navedemo redne brojeve stavki koje uklanjamo sa znakom minus ispred, recimo p.pod[,–c(1)]. Ako uklanjate više stavki njihove redne brojeve, odvojene zarezima, dodajte u zagradu, npr. p.pod[,–c(1, 3)].

```
# Prvo ćemo izbaciti stavku 1 iz matrice podataka. Novu matricu podataka nazvaćemo
b.podR. Redukovanu matricu podataka čuvamo pod drugim imenom kako bismo sačuvali i
originalne podatke. Nije nužno tako, ali je korisno
b.podR <- b.pod[,2:10]
```

```
mr.eR <- RM(b.podR)

# Uporedni prikaz parametara na originalnom i na redukovanom skupu stavki
cbind("originalni"=mr.e$betapar*(-1), "redukovani"=c(NA, mr.eR$betapar*(-1)))

##            originalni redukovani
## beta it1   0.23038215         NA
## beta it2   0.42166177  0.4517879
## beta it3   0.25293099  0.2795252
## beta it4   0.06048448  0.0833948
## beta it5  -1.57136771 -1.5746559
## beta it6  -1.44116363 -1.4424688
## beta it7   2.91510819  3.0039745
## beta it8  -0.88439814 -0.8767581
## beta it9  -1.64752484 -1.6519092
## beta it10  1.66388673  1.7271096
```

Promene u ovom slučaju nisu velike, ali svakako postoje.

#### *10.4.1.10 Separaciona pouzdanost*

Separaciona pouzdanost je pouzdanost tipa interne konzistencije koja se izračunava u Rašovim modelima (Fajgelj, 2020). Bliska je, ali ne identična Kronbahovoj alfi.

```
SEP.POUZD <- SepRel(mr.ePI)
# Funkcija SepRel() kao objekat uzima objekat sa parametrima ispitanika
SEP.POUZD

## Separation Reliability: 0.672

# Za izračunavanje Kronbahove alfe koristićemo funkciju alpha() iz paketa psych
# Ispisaćemo samo sirovu alfu (bez $total$raw_alpha biće ispisane tabele sa raznim
pokazateljima)
cat("Kronbahova alfa: \n")

## Kronbahova alfa:

psych::alpha(b.pod, check.keys=TRUE)$total$raw_alpha

## [1] 0.7572566
```

Separaciona pouzdanost ovog testa je nešto ispod donje granice prihvatljive pouzdanosti (0,7).

Pouzdanost IRT modela koji ima dobre pokazatelje fita ne mora nužno biti visoka. Pokazatelji misfita kod jednoparametarskih modela će ukazati na kršenje pretpostavke jednakosti nagiba. Ako ta pretpostavka nije narušena, model može imati dobre pokazatelje fita. Oni ništa ne govore o samom nagibu, a on može biti jednako nizak za sve stavke. Drugim rečima, sve stavke mogu biti jednako slabo diskriminativne, a da model još uvek fituje (Wu i ostali, 2016).

##### 10.4.1.10.1 Izračunavanje separacionog odnosa i broja razdvojivih stratuma

Separacioni odnos nam govori na koliko statistički različitih stratuma se mogu razvrstati ispitanici.

Separacioni odnos se označava sa $G$ (videti odeljak 7.2).

```
G <- sqrt(SEP.POUZD$sep.rel/(1-SEP.POUZD$sep.rel))
cat("G =", G)
```

```
## G = 1.431517

BRS=(4*G+1)/3
cat("Broj razdvojivih stratuma =", BRS)

## Broj razdvojivih stratuma = 2.242023
```

Na osnovu ovog testa po nivou osobine možemo izdvojiti dva, eventualno tri, značajno različita stratuma ispitanika. Kažemo „eventualno tri“ zato što vrednost prelazi 2 za 0,24.

#### *10.4.1.11 Parametri (mere) ispitanika (θ)*

Iz ranije konstruisanog objekta mr.ePI možemo izvući parametre ispitanika. Ovo ima smisla raditi samo ako je model saglasan sa podacima. Nećemo ih prikazivati zbog uštede prostora.

```
MERE <- mr.ePI$theta.table[,1]
# Iz objekta mr.ePI iz tabele theta.table izdvajamo prvu kolonu {,1}
# Ovu varijablu možemo kasnije koristiti u analizama, npr:

mean(MERE)

## [1] 0.3778818

# Prosečna mera u logitima
```

#### *10.4.1.12 Pokazatelji fita za ispitanike*

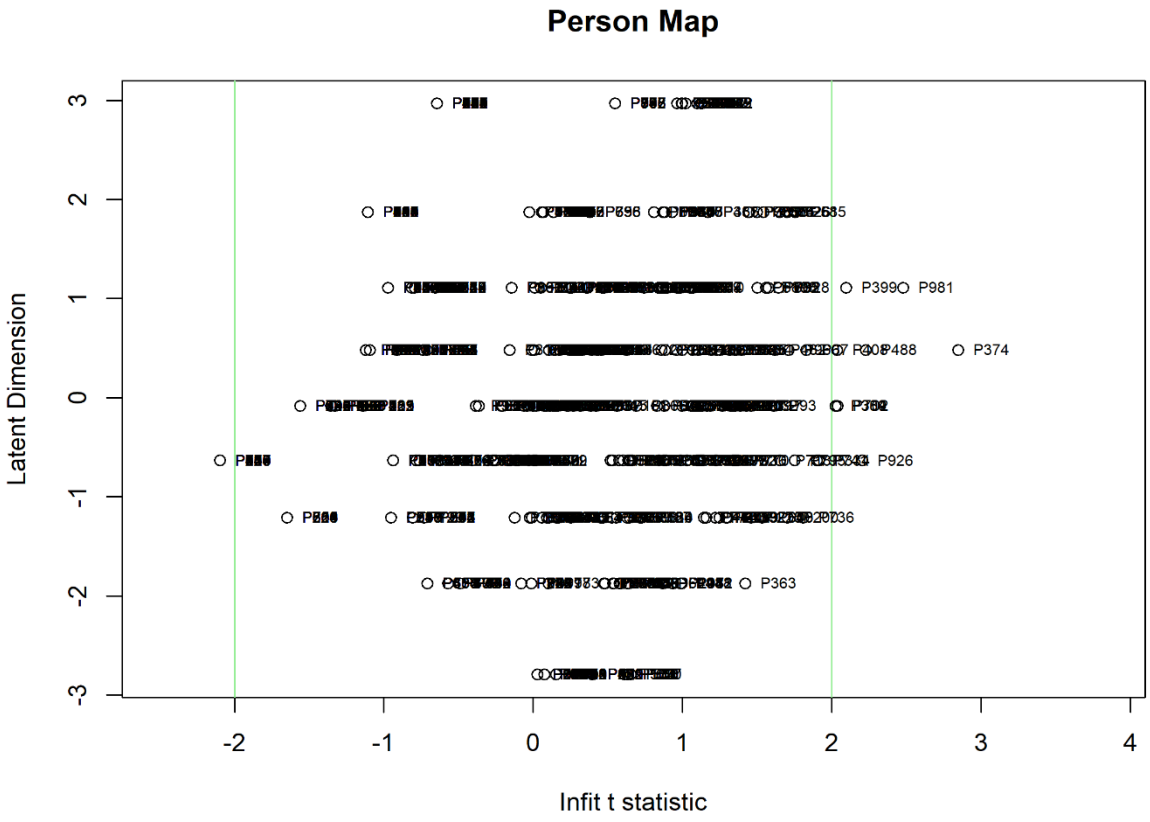
Za ispitanike možemo dobiti iste pokazatelje kao i za stavke. Interpretacija pokazatelja je identična. Prikazaćemo samo prvih 10 ispitanika.

```
fit.isp <- personfit(mr.ePI)

fit.isp

##
## Personfit Statistics:
##         Chisq df p-value Outfit MSQ Infit MSQ Outfit t Infit t
## P1      3.789  9   0.925      0.379     0.444    -0.98   -2.10
## P2      8.478  9   0.487      0.848     0.956    -0.01   -0.03
## P3      5.588  9   0.780      0.559     0.736    -0.77   -0.71
## P4      8.773  9   0.458      0.877     1.015     0.17    0.16
## P5      9.818  9   0.365      0.982     1.008     0.15    0.13
## P6      7.419  9   0.594      0.742     1.051    -0.16    0.26
## P7      5.334  9   0.804      0.533     0.652    -0.83   -1.14
## P9      6.017  9   0.738      0.602     0.937     0.10   -0.01
## P10     4.953  9   0.838      0.495     1.243     0.32    0.55
...
```

10.4.1.12.1 Grafikon misfita za ispitanike

Kao i za stavke, možemo dobiti i grafikon misfita za ispitanike. Prikazan je infit u standardizovanoj metrici (Slika 40). Vertikalne zelene linije ograničavaju prihvatljivi opseg vrednosti. Levo od prihvatljivog opsega su ispitanici čije je odgovaranje suviše predvidljivo (prigušeni šum), a desno oni ispitanici čije odgovaranje model ne predviđa dobro (šum).

```
plotPWmap(mr.e, pmap=T, imap=F)
```




Slika 40 – Grafikon standardizovanog infita za ispitanike

Jedan ispitanik ima misfit (infit) u formi overfita, a veći broj u obliku šuma.

#### *10.4.1.13 Formiranje sirovog skora i upis u data.frame*

Mere koje smo kreirali u jednom od prethodnih koraka možemo koristiti u daljim analizama, a možemo ih dodati i u matricu podataka (*data.frame*). Mi to na ovom mestu nećemo raditi, ali komanda za to bi bila *b.pod$MERE <- MERE*. Umesto toga, kreiraćemo ukupan sumacioni skor na testu kao novu varijablu, a zatim ćemo spojiti SKOR i MERE u novi *data.frame* pod nazivom *df* i izračunati korelacije sirovih skorova i Rašovih mera.

```
SKOR <- rowSums(b.pod[1:ncol(b.pod)])
df <- data.frame(cbind(SKOR, MERE))
head(df, 5)

##   SKOR        MERE
## 1    4 -0.63659070
## 2    4 -0.63659070
## 3    6  0.48010094
## 4    3 -1.21335219
## 5    5 -0.08534653

#View(RASCH.POD)
cat("Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera: \n")
```

```
## Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera:

cor(SKOR, MERE, method="pearson")

## [1] 0.9916267

cat(" \nSpirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:
\n")

##
## Spirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:

cor(SKOR, MERE, method="spearman")

## [1] 1
```

Pirsonov koeficijent korelacije sumacionog skora i Rašovih mera je vrlo visok, a Spirmanov koeficijent rang korelacije je jednak 1. Poredak ispitanika po sumacionim skorovima i Rašovim merama je isti.

### 10.4.2 Procena Rašovog modela u paketu ltm

Paket *ltm* (Rizopoulos 2006) je paket koji može da procenjuje jednodimenzionalne i višedimenzionalne IRT modele. Kao metod procene parametara koristi marginalnu maksimalnu verodostojnost (MMLE). Može da procenjuje jedno–, dvo– i troparametarske modele, za binarno i politomno skorovane stavke.

Učitaćemo ga.

```
p_load(ltm)
```

Paket *ltm* može da proceni Rašov model na više različitih načina.

Prvi je upotrebom funkcije *rasch().*

```
m1p.l <- rasch(b.pod)
m1p.l

##
## Call:
## rasch(data = b.pod)
##
## Coefficients:
##  Dffclt.it1   Dffclt.it2   Dffclt.it3   Dffclt.it4   Dffclt.it5   Dffclt.it6
##      -0.086        0.044       -0.070       -0.200       -1.299       -1.211
##  Dffclt.it7   Dffclt.it8   Dffclt.it9  Dffclt.it10       Dscrmn
##       1.741       -0.837       -1.350        0.888        1.480
##
## Log.Lik: -5193.881
```

Ovo je jednoparametarski model, ali nije pravi Rašov model. Nagib/diskriminativnost je jednak za sve stavke, ali nije jednak 1 već je njihova vrednost procenjena (*Dscrmn=1,480*).

Za pravi Rašov model potrebno je dodati argument *constraint = cbind(ncol(b.pod) + 1, 1)*.

Ovim ograničavamo 11. parametar na vrednost 1. Naime, u modelu su prvih 10 parametara parametri težine stavki (zato što imamo 10 stavki), a prvi sledeći je nagib koji je jednak za sve. Ograničenje se daje u vidu matrice u kojoj je u prvoj koloni broj parametra koji se ograničava, a u drugoj vrednost na koju

se ograničava. Prva komanda u narednom kodu je upravo ta matrica koju prosleđujemo argumentu *constraint*.

```
cbind(ncol(b.pod) + 1, 1)
```

```
##      [,1] [,2]
## [1,]   11    1
```

```
# Gornja komanda je samo ilustracija kako izgleda matrica koju prosleđujemo
argumentu constraint
# Umesto ncol(b.pod)+1 mogli smo napisati i 11, ali na ovaj način, ako obrišemo
ili dodamo stavku u matricu podataka b.pod nema potrebe da menjamo komandu
mr.1 <- rasch(b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
mr.1
```

```
##
## Call:
## rasch(data = b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
##
## Coefficients:
##  Dffclt.it1   Dffclt.it2   Dffclt.it3   Dffclt.it4   Dffclt.it5   Dffclt.it6
##      -0.111        0.061       -0.091       -0.264       -1.731       -1.613
##  Dffclt.it7   Dffclt.it8   Dffclt.it9  Dffclt.it10       Dscrmn
##       2.326       -1.112       -1.800        1.184        1.000
##
## Log.Lik: -5249.971
```

```
cbind(mr.1$coefficients)
```

```
##           beta.i beta
## it1   0.11123755    1
## it2  -0.06063400    1
## it3   0.09098949    1
## it4   0.26371818    1
## it5   1.73103257    1
## it6   1.61341573    1
## it7  -2.32609594    1
## it8   1.11215465    1
## it9   1.79999042    1
## it10 -1.18353587    1
```

U prvoj koloni poslednje tabele su parametri težine, a u drugoj diskriminativnosti.

Drugi način procene Rašovog modela u paketu *ltm* je upotrebom funkcije *tpm()*, koja je namenjena proceni troparametarskog modela. Kao i kod funkcije *rasch()* moguće je postaviti ograničenja na određene parametre. Ako ograničimo parametar diskriminativnosti za sve stavke na 1, a parametar pseudopogađanja na 0, dobićemo Rašov model. Ako ograničimo samo parametar pseudopogađanja, dobićemo dvoparametarski model. Kako su ovi modeli ugnježđeni moći ćemo da obavimo njihovo poređenje na osnovu logaritma verodostojnosti i vidimo koji bolje fituje.

Ograničenja se prosleđuju argumentu *constraint* u obliku matrice sa tri kolone i onoliko redova koliko ima stavki kojima želimo da ograničimo neki parametar. U prvoj koloni matrice su redni brojevi stavki čije parametre fiksiramo. U drugoj koloni su numeričke oznake parametra koji želimo da fiksiramo (1 – pseudopogađanje, 2 – lakoća i 3 – nagib). U trećoj koloni su vrednosti na koje želimo da fiksiramo parametre.

```
cm <- matrix(c(1:ncol(b.pod), rep(3, ncol(b.pod)), rep(1, ncol(b.pod))),
 nrow = ncol(b.pod), ncol=3, byrow = FALSE)
cm

##       [,1] [,2] [,3]
##  [1,]    1    3    1
##  [2,]    2    3    1
##  [3,]    3    3    1
##  [4,]    4    3    1
##  [5,]    5    3    1
##  [6,]    6    3    1
##  [7,]    7    3    1
##  [8,]    8    3    1
##  [9,]    9    3    1
## [10,]   10    3    1

mr.lt <- tpm(b.pod, constraint=cm, max.guessing = 0)
mr.lt

## Call:
## tpm(data = b.pod, constraint = cm, max.guessing = 0)
##
## Coefficients:
##       Gussng  Dffclt  Dscrmn
## it1        0  -0.111       1
## it2        0   0.061       1
## it3        0  -0.091       1
## it4        0  -0.264       1
## it5        0  -1.731       1
## it6        0  -1.613       1
## it7        0   2.326       1
## it8        0  -1.112       1
## it9        0  -1.800       1
## it10       0   1.184       1
##
## Log.Lik: -5249.971

# Uporedite sa modelom mr.l koji smo ranije procenili funkcijom rasch()
cbind(mr.l$coefficients, mr.lt$coefficients[,2:3])

##           beta.i beta     beta.1i beta.2i
## it1   0.11123755    1  0.11123755       1
## it2  -0.06063400    1 -0.06063400       1
## it3   0.09098949    1  0.09098949       1
## it4   0.26371818    1  0.26371818       1
## it5   1.73103257    1  1.73103257       1
## it6   1.61341573    1  1.61341573       1
## it7  -2.32609594    1 -2.32609594       1
## it8   1.11215465    1  1.11215465       1
## it9   1.79999042    1  1.79999042       1
## it10 -1.18353587    1 -1.18353587       1

# Ovo je pravi Rašov model sa parametrima nagiba jednakim 1
```

Možemo proceniti i jednoparametarski logistički model sa procenjenim parametrom diskriminativnosti jednakim za sve stavke (ovde ćemo ga označiti sa 1PL). Umesto argumenta *constraint* koristićemo *type="rasch"*. On samo ograničava diskriminativnosti da budu jednake za sve stavke.

```
m1p.lt <- tpm(b.pod, type="rasch", max.guessing = 0)
m1p.lt

## Call:
## tpm(data = b.pod, type = "rasch", max.guessing = 0)
##
## Coefficients:
##       Gussng  Dffclt  Dscrmn
## it1        0  -0.086    1.48
## it2        0   0.044    1.48
## it3        0  -0.070    1.48
## it4        0  -0.200    1.48
## it5        0  -1.299    1.48
## it6        0  -1.211    1.48
## it7        0   1.741    1.48
## it8        0  -0.837    1.48
## it9        0  -1.350    1.48
## it10       0   0.888    1.48
##
## Log.Lik: -5193.881
```

Ako ne koristimo argumente *constraint* i *type="rasch"*, a koristimo argument *max.guessing=0*, dobićemo dvoparametarski model za binarno skorovane stavke.

```
m2p.lt <- tpm(b.pod, max.guessing = 0)
m2p.lt

## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Coefficients:
##       Gussng  Dffclt  Dscrmn
## it1        0  -0.104   1.095
## it2        0   0.040   1.782
## it3        0  -0.073   1.417
## it4        0  -0.208   1.384
## it5        0  -1.257   1.575
## it6        0  -1.197   1.514
## it7        0   1.756   1.459
## it8        0  -0.900   1.304
## it9        0  -1.195   1.901
## it10       0   0.822   1.730
##
## Log.Lik: -5181.038
```

Na kraju, ako isključimo i argument *max.guessing=0*, dobićemo 3PL model.

Procene parametara pseudopogađanja mogu biti nestabilne i to će nekada rezultirati upozorenjem. Nekada se problem može rešiti postavljanjem početnih vrednosti za ove parametre. Ako nemate neke očekivane vrednost možete ih postaviti na slučajne vrednosti argumentom *start.val="random".*

```
m3p.lt <- tpm(b.pod, start.val =  "random" )
m3p.lt

##
## Call:
## tpm(data = b.pod, start.val = "random")
##
## Coefficients:
```

```
##       Gussng  Dffclt  Dscrmn
## it1    0.000  -0.102   1.098
## it2    0.000   0.042   1.789
## it3    0.000  -0.070   1.423
## it4    0.000  -0.207   1.380
## it5    0.000  -1.248   1.591
## it6    0.000  -1.188   1.531
## it7    0.000   1.754   1.458
## it8    0.235  -0.422   1.744
## it9    0.144  -0.991   2.132
## it10   0.000   0.821   1.731
##
## Log.Lik: -5179.289
```

#### *10.4.2.1 Fit modela u celini*

U *ltm* paketu dostupan je pokazatelj fita modela u celini za Rašove model. Pokazatelj je zasnovan na Pirsonovom hi–kvadrat testu. Zbog ranije pominjanih problema sa nepoznatom uzoračkom distribucijom pokazatelja (odeljak 6.3.1), procedura *GoF.rasch()* nudi opciju Monte Karlo simulacija (parametarsko samo-uzorkovanje) za izračunavanje *p* vrednosti. Za tu opciju potrebno je uključiti argumente *simulate.p.value=TRUE* i *B=1000* sa željenim brojem uzoraka (u ovom slučaju definisano je 1000 uzoraka). Uključivanje ovog argumenta bitnije je ako je *p* vrednost blizu graničnoj *p*=0,05. Ovo napominjem zato što u zavisnosti od snage računara izračunavanje pokazatelja može uzeti dosta vremena.

Ako želite reproducibilne rezultate, potrebno je dodati i argument *seed=123* (umesto 123 možete staviti bilo koji broj, ali u analizi sa drugim brojem možete dobiti rezultate koji će malo odstupati od prethodnih). Funkcija *GoF.rasch()* radi samo sa modelima procenjenim funkcijom *rasch()*.

```
GoF.rasch(mr.1)

##
## Bootstrap Goodness-of-Fit using Pearson chi-squared
##
## Call:
## rasch(data = b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
##
## Tobs: 808.39
## # data-sets: 50
## p-value: 0.94

GoF.rasch(mr.1, simulate.p.value=TRUE, B=1000, seed=123)

##
## Bootstrap Goodness-of-Fit using Pearson chi-squared
##
## Call:
## rasch(data = b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
##
## Tobs: 808.39
## # data-sets: 1001
## p-value: 0.893
```

Na osnovu ovog omnibus pokazatelja fita ne možemo odbaciti model, odnosno možemo reći da *model fituje*. Razlika u *p*–vrednosti pokazatelja fita dobijenog sa i bez Monte Karlo simulacija nije velika, ali bi

u situaciji kada je *p* blizu kritične vrednosti mogla napraviti razliku o zaključku o fitu modela.

Fitovanje Rašovog modela funkcijom *tpm()* možda je delovalo kao nepotrebno. Međutim, mogućnost da procenjujemo ugnježđene modele sa različitim brojem parametara koristeći istu funkciju, osim što je edukativno[50], daje nam mogućnost da modele poredimo upotrebom LR testa (odeljak 4.5).

```
#Poređenje Rašovog i 2PL modela
anova(mr.lt, m2p.lt)

##
##  Likelihood Ratio Table
##             AIC      BIC  log.Lik    LRT df p.value
## mr.lt  10539.94 10638.10 -5249.97
## m2p.lt 10422.08 10569.31 -5181.04 137.87 10  <0.001
```

Ranije smo utvrdili da je Rašov model dovoljno saglasan sa podacima. Međutim, to ne znači da je on i najbolji model za opisivanje ovih podataka. Poređenje modela LR testom će nam pokazati da li ima boljih modela.

Značajne *p* vrednosti ukazuju na značajne razlike u fitu modela. Ako one postoje, bolje fituju modeli sa nižim vrednostima *AIC* i *BIC*, a višim vrednostima *log.Lik* (logaritam verodostojnosti). Kod interpretacije *log.Lik* obratite pažnju da su vrednosti negativne, pa manja apsolutna vrednost označava višu vrednost.

U poređenju Rašovog i 2PL modela značajan LRT ukazuje da bolji fit ima 2PL model jer ima višu vrednost *log.Lik* (i niže vrednosti AIC i BIC).

Ako su podaci saglasni sa strožim modelom (onim koji postavlja više zahteva), biće saglasni i sa manje strogim modelima u kojima je on ugnježđen. Ako Rašov model ima zadovoljavajući fit, imaće i 2PL i 3PL.

Kada imamo ovakvu situaciju (oba modela fituju, ali jedan bolje), odluka o tome za koji ćemo se model opredeliti zavisi od toga šta želimo da uradimo. Ako želimo da kreiramo merno–teorijski čist test ili nam je bitna lakoća skorovanja, opredelićemo se za Rašov model. Ako želimo da iskoristimo sve informacije koje nam nudi 2PL model kako bismo napravili poredak ispitanika za neku praktičnu svrhu, onda možemo birati taj model. Imajte na umu da još nismo detaljnije pogledali 2PL model.

```
anova(mr.lt, m3p.lt)

##
##  Likelihood Ratio Table
##             AIC      BIC  log.Lik    LRT df p.value
## mr.lt  10539.94 10638.10 -5249.97
## m3p.lt 10418.58 10565.81 -5179.29 141.36 10  <0.001
```

Poređenjem Rašovog i 3PL modela vidimo da LRT ukazuje da bolji fit ima 3PL.

```
anova(m1p.lt, m2p.lt)

##
##  Likelihood Ratio Table
##             AIC      BIC  log.Lik   LRT df p.value
```

[50] Pokazuje nam da su model ugnježđeni.

```
## m1p.lt 10429.76 10532.82 -5193.88
## m2p.lt 10422.08 10569.31 -5181.04 25.69  9   0.002
```

U poređenju 1PL (jednoparametarski sa procenjenim parametrom diskriminativnosti) i 2PL modela značajan LRT ukazuje da bolji fit ima 2PL model.

```
anova(m1p.lt, m3p.lt)

##
##  Likelihood Ratio Table
##             AIC      BIC  log.Lik   LRT df p.value
## m1p.lt 10429.76 10532.82 -5193.88
## m3p.lt 10418.58 10565.81 -5179.29 29.18  9   0.001
```

LRT za 1PL i 3PL modele ukazuje da bolji fit ima 3PL model.

```
anova(mr.l, m1p.l)

##
##  Likelihood Ratio Table
##            AIC      BIC  log.Lik    LRT df p.value
## mr.l  10519.94 10569.02 -5249.97
## m1p.l 10409.76 10463.75 -5193.88 112.18  1  <0.001
```

Poređenje Rašovog i 1PL modela ukazuje da je bolji 1PL model. Ovde smo koristili modele koje smo procenili funkcijom *rasch()*, a ne *tpm()* kao u ostalim primerima.

#### 10.4.2.1.1 Pokazatelji fita za stavke

Pokazatelje saglasnosti za stavke u paketu *ltm* dobijamo funkcijom *item.fit()* koja može da radi sa modelima procenjenim funkcijama *rasch()*, *tpm()* i *ltm()*. I ovi pokazatelji bazirani su na hi–kvadrat testu kao i omnibus pokazatelji.

Dodatni argumenti funkcije *item.fit()* su *G=10* (koji određuje u koliko intervala formiranih po nivou osobine da deli uzorak) i *FUN=median* (kojom vrednošću $\theta$ će biti predstavljeni intervali). Kombinacija *G=10* i *FUN=median* definiše $Q_1$ pokazatelj fita Jenove. U Bokovom hi–kvadrat testu broj intervala nije tačno definisan, a grupe su predstavljene aritmetičkom sredinom.

I ova funkcija nudi opciju Monte Karlo simulacija, a to definišemo argumentima *simulate.p.value=TRUE* i *B=1000*.

```
item.fit(mr.l, G=10, FUN= median, simulate.p.value=TRUE, B=1000)

##
## Item-Fit Statistics and P-values
##
## Call:
## rasch(data = b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
##
## Alternative: Items do not fit the model
## Ability Categories: 10
## Monte Carlo samples: 1000
##
##           X^2 Pr(>X^2)
## it1   35.9669     0.98
## it2   97.0727    0.003
```

```
## it3   66.1906   0.3157
## it4   58.5186   0.5125
## it5   52.3958   0.3077
## it6   49.5453   0.4106
## it7   75.0583    0.017
## it8   42.7984   0.7732
## it9   70.9004     0.03
## it10 115.9762    0.001
```

Ovde pokazatelji fita ukazuju na druge stavke kao problematične. Naime značajne hi–kvadrat testove imaju stavke 2, 7, 9 i 10. Iako značajan misfit postoji na nivou stavki, omnibus pokazatelj ne ukazuje na misfit. Setite se da su pokazatelji u paketu *eRm* kod većine stavki ukazivali na overfit. Ovde nema podele na šum i prigušeni šum. Značajni hi–kvadrati ukazuju na oba tipa misfita.

#### 10.4.2.1.2 Pokazatelji fita za ispitanike

Dostupni su i pokazatelji fita za ispitanike. Pokazatelji koji se ovde računaju i prikazuju su $L_0$ i $L_z$ i zasnivaju se na logaritmu verodostojnosti (odeljak 6.4).

$L_z$ je standardizovani pokazatelj i označava se još i sa $Z_h$ (kada je test sastavljen od politomnih stavki) ili sa $Z_3$ (kada su u pitanju dihotomne). Vrednosti standardizovanog pokazatelja preko 1,96 ukazuju na šum, a one ispod –1,96 na prigušeni šum.

Definisanje funkcije *person.fit()* analogno je funkciji *item.fit()* za stavke, osim što postoji dodatni argument *alternative="two.sided"*. Ovaj argument definiše alternativnu hipotezu. Navedena opcija *"two.sided"* namenjena je otkrivanju bilo kakvog misfita. Ako je p<0,05, postoji sumnja na misfit.

Osim navedene opcije, u argumentu *alternative* moguće su i opcija *"less"* koji je namenjena otkrivanju šuma i *"greater"* koja je namenjena otkrivanju prigušenog šuma (overfit).

I ova funkcija radi sa modelima procenjenim funkcijama *rasch()*, *tpm()* i *ltm()*.

Rezultat funkcije je objekat koji sadrži pokazatelje fita za *sklopove odgovora*, ne za ispitanike.

Da biste utvrdili o kojim ispitanicima se radi morate spojiti pokazatelje misfita sa matricom podataka.

```
mr.lPF <- person.fit(mr.l, simulate.p.value=T, B=1000, alternative="two.sided")
# Pozivanjem naziva objekta u koji smo smestili pokazatelja fita dobijemo ispis
sklopova pokazatelja fita i njihove statističke značajnosti
#m2p.ltPF
# Nećemo ih sada ispisati zbog prostora, ali ćemo podatke o sklopovima odgovora,
pokazateljima i njihovoj statističkoj značajnosti iz objekta izdvojiti u
data.frame
mr.lMF <- data.frame(mr.lPF$resp.patterns, mr.lPF$Tobs, mr.lPF$p.values)
# Zatim ćemo ih spojiti sa podacima, kao što smo to radili za mere u Rašovom
modelu
# Prvo pravimo kopiju podataka kako bismo sačuvali originalne
b.podK <- b.pod
# Zatim ih spajamo sa pokazateljima misfita
b.podK <- merge(b.podK, mr.lMF, all.x = TRUE, by=names(b.podK))
```

```
# Donjom komandom možemo iz matrice podataka odabrati samo ispitanike koji imaju
značajan misfit ili one koji nemaju (Lz.1 je naziv kolone u kojoj je smeštena p
vrednost za Lz pokazatelj)
b.podK[b.podK$Lz.1<.05,]
```

```
##     it1 it2 it3 it4 it5 it6 it7 it8 it9 it10         L0        Lz        L0.1
## 103   0   0   0   0   1   0   1   1   0    0  -8.139238 -1.966343 0.041958042
## 173   0   0   0   1   0   1   1   0   1    0  -8.341564 -2.043853 0.035964036
## 342   0   1   0   1   0   1   1   0   1    0  -8.973226 -2.462292 0.015984016
## 343   0   1   0   1   0   1   1   0   1    0  -8.973226 -2.462292 0.015984016
## 411   0   1   1   1   0   1   1   1   1    1  -7.763552 -1.846521 0.047952048
## 415   0   1   1   1   1   0   1   0   1    1  -9.667418 -2.914267 0.015984016
## 485   1   0   0   0   0   0   1   0   1    0  -9.071197 -2.660667 0.016983017
## 498   1   0   0   0   0   1   1   1   1    0  -7.952918 -1.726735 0.043956044
## 546   1   0   0   1   0   1   1   1   1    1  -8.591998 -2.230318 0.022977023
## 586   1   0   1   0   0   0   0   1   1    1  -8.332784 -2.000587 0.039960040
## 636   1   0   1   1   1   0   1   0   1    0  -8.695938 -2.254334 0.025974026
## 713   1   1   0   1   0   0   0   1   0    0  -8.266116 -1.987658 0.039960040
## 785   1   1   1   0   1   0   1   0   1    0  -9.020290 -2.475167 0.013986014
## 821   1   1   1   1   0   1   1   0   0    0 -10.674179 -3.601208 0.004995005
##            Lz.1
## 103 0.044955045
## 173 0.035964036
## 342 0.016983017
## 343 0.016983017
## 411 0.049950050
## 415 0.016983017
## 485 0.014985015
## 498 0.044955045
## 546 0.025974026
## 586 0.043956044
## 636 0.031968032
## 713 0.044955045
## 785 0.014985015
## 821 0.004995005
```

```
# Ovo je varijanta za prikaz onih koji nemaju misfit
#b.podK[!b.podK$Lz.1<.05,]

# Ovo su varijante za otkrivanje samo šuma...
#person.fit(mr.l, simulate.p.value=T, B=1000, alternative="less")
# ili samo prigušenog šuma (overfita)
#person.fit(mr.l, simulate.p.value=T, B=1000, alternative="greater")
```

Od 1000 slučajeva samo 12 ima značajan misfit, što nije puno. Ako je merenje obavljeno sa nekom praktičnom svrhom poput selekcije, odgovore misfitujućih ispitanika bi trebalo posebno tumačiti. Takođe, trebalo bi pregledati njihove testove i pogledati šta je tačno neuobičajeno u njihovom odgovaranju.

Sa argumentom *alternative* i opcijama *"less"* ili "*greater"* mogli bismo proveriti radi li se o neočekivano predvidljivom ili nepredvidljivom odgovaranju.

#### *10.4.2.2 Mere ispitanika*

Procena mera ispitanika ima smisla ako smo pokazali da je model saglasan sa podacima.

U paketu *ltm* mere ispitanika možete dobiti funkcijom *factor.scores(naziv_modela).* Na sledećem

primeru biće prikazano kako da ih spojite sa vašim podacima koji sadrže odgovore na stavke. Takve mere možete dalje koristiti za razne analize.

```
# Pošto će nam izvorni podaci i dalje trebati, napravićemo kopiju pod drugim
imenom
b.podK <- b.pod
# Mere ispitanika ćemo upisati u novi data.frame
MereL <- factor.scores(mr.l)$score.dat
# Objekat funkcije je model koji smo ranije procenili pomoću paketa ltm
# Objekat MereL ne sadrži mere za sve ispitanike već samo za sklopove koji
postoje, a ovde ih ima
nrow(MereL)

## [1] 263

# Za spajanje ćemo koristiti funkciju merge() kojoj je potrebno navesti imena
matrica podataka koje želimo da spojimo
b.podM <- merge(b.podK, MereL, all.x = TRUE, by=names(b.podK))
# Argument all.x=TRUE znači da želimo da zadržimo sve redove iz prve matrice
(b.podK), a argument by=names(b.podK) znači da bi dve matrice trebalo da budu
uparene po varijablama sadržanim u vektoru names(b.podK), što su nazivi varijabli
u matrici b.podK. Ove varijable moraju postojati u obe matrice da bi mogle da se
koriste za uparivanje. Na ovaj način će mere za određeni sklop odgovora na stavke
iz matrice MereL biti pripisane svim ispitanicima koji imaju isti takav sklop u
matrici b.podK i sve to će biti upisano u novi objekat b.podM
head(b.podM)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10 Obs     Exp        z1     se.z1
## 1   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361
## 2   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361
## 3   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361
## 4   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361
## 5   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361
## 6   0   0   0   0   0   0   0   0   0    0  19 6.28197 -2.064065 0.6443361

# Pored mera u matricu će biti upisane opažene i očekivane frekvencije svakog
sklopa, mera ispitanika kao varijabla z1, i standardna greška mere kao varijabla
se.z1
```

#### *10.4.2.3 Test jednodimenzionalnosti*

Paket *ltm* ima i funkciju za testiranje jednodimenzionalnosti testa. Ova funkcija zasnovana je na varijanti paralelne analize (Drasgow & Lissak, 1983). Testira nultu hipotezu da drugi karakteristični koren u matrici podataka nije značajno veći od drugog karakterističnog korena na podacima dobijenim Monte Karlo simulacijama po određenom IRT modelu. Ako je $p$<0,05, onda test nije jednodimenzionalan.

Procedura je namenjena binarno skorovanim stavkama i ne funkcioniše na politomnim.

```
unidimTest(mr.l, b.pod, B=1000)

##
## Unidimensionality Check using Modified Parallel Analysis
##
## Call:
## rasch(data = b.pod, constraint = cbind(ncol(b.pod) + 1, 1))
##
## Matrix of tertachoric correlations
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
```

```
## it1  1.0000 0.3528 0.3781 0.3936 0.3604 0.3388 0.3095 0.3390 0.4005 0.4383
## it2  0.3528 1.0000 0.4838 0.4506 0.4926 0.5426 0.5167 0.4670 0.5232 0.4431
## it3  0.3781 0.4838 1.0000 0.3831 0.4408 0.4065 0.4028 0.4136 0.4212 0.4735
## it4  0.3936 0.4506 0.3831 1.0000 0.4596 0.4379 0.3969 0.3849 0.4644 0.4287
## it5  0.3604 0.4926 0.4408 0.4596 1.0000 0.4758 0.2211 0.3495 0.5173 0.4685
## it6  0.3388 0.5426 0.4065 0.4379 0.4758 1.0000 0.3184 0.3727 0.4486 0.4575
## it7  0.3095 0.5167 0.4028 0.3969 0.2211 0.3184 1.0000 0.3554 0.4846 0.4513
## it8  0.3390 0.4670 0.4136 0.3849 0.3495 0.3727 0.3554 1.0000 0.4995 0.4199
## it9  0.4005 0.5232 0.4212 0.4644 0.5173 0.4486 0.4846 0.4995 1.0000 0.5472
## it10 0.4383 0.4431 0.4735 0.4287 0.4685 0.4575 0.4513 0.4199 0.5472 1.0000
##
## Alternative hypothesis: the second eigenvalue of the observed data is
substantially larger
##          than the second eigenvalue of data under the assumed IRT model
##
## Second eigenvalue in the observed data: 0.3745
## Average of second eigenvalues in Monte Carlo samples: 0.3601
## Monte Carlo samples: 1000
## p-value: 0.3746
```

Test nije značajan, što znači da test možemo smatrati jednodimenzionalnim.

### 10.4.3 Procena Rašovog modela u paketu *mirt*

Paket *mirt* (Chalmers, 2012) moćan je i fleksibilan paket koji može da procenjuje jednodimenzionalne i višedimenzionalne IRT modele. Poput paketa *ltm* kao metod procene parametara koristi marginalnu maksimalnu verodostojnost (MMLE). Može da procenjuje jednoparametarske i višeparametarske modele (do petoparametarskog), a njegove mogućnosti daleko prevazilaze ono što će biti prikazano u ovoj knjizi.

Za početak učitaćemo biblioteku *mirt*.

```
p_load(mirt)
```

#### *10.4.3.1 Rašov model u paketu mirt*

Kao što smo uradili sa paketom *ltm* i ovde ćemo krenuti sa prikazivanjem procene Rašovog modela.

```
mr.m <- mirt(b.pod, 1, itemtype="Rasch", SE=T, verbose = FALSE)
# Argumenti u komandi: broj 1 predstavlja broj dimenzija (skraćeno pisanje
argumenta model=1), itemtype="Rasch" definiše da su parametri nagiba jednaki 1, a
SE=T ili TRUE traži izračunavanje standardnih grešaka
mr.m

##
## Call:
## mirt(data = b.pod, model = 1, itemtype = "Rasch", SE = T, verbose = FALSE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 22 EM iterations.
## mirt version: 1.41
## M-step optimizer: nlminb
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix =  5.441375
```

```
##
## Log-likelihood = -5193.887
## Estimated parameters: 11
## AIC = 10409.77
## BIC = 10463.76; SABIC = 10428.82
## G2 (1012) = 564.23, p = 1
## RMSEA = 0, CFI = NaN, TLI = NaN
```

Ako pozovemo naziv objekta u koji smo zapisali model (mr.m), dobićemo pokazatelj fita $G^2$ zasnovan na logaritmu verodostojnosti i punoj informaciji (odeljak 6.1.1).

```
# Parametre stavki i standardne greške procene parametara dobićemo funkcijom
coef()
coef(mr.m)
```

```
## $it1
##         a1      d  g  u
## par      1  0.126  0  1
## CI_2.5  NA -0.048 NA NA
## CI_97.5 NA  0.300 NA NA
##
## $it2
##         a1      d  g  u
## par      1 -0.065  0  1
## CI_2.5  NA -0.239 NA NA
## CI_97.5 NA  0.109 NA NA
##
## $it3
##         a1      d  g  u
## par      1  0.104  0  1
## CI_2.5  NA -0.070 NA NA
## CI_97.5 NA  0.278 NA NA
##
## $it4
##         a1     d  g  u
## par      1 0.296  0  1
## CI_2.5  NA 0.121 NA NA
## CI_97.5 NA 0.471 NA NA
##
## $it5
##         a1     d  g  u
## par      1 1.921  0  1
## CI_2.5  NA 1.715 NA NA
## CI_97.5 NA 2.127 NA NA
##
## $it6
##         a1     d  g  u
## par      1 1.792  0  1
## CI_2.5  NA 1.590 NA NA
## CI_97.5 NA 1.993 NA NA
##
## $it7
##         a1      d  g  u
## par      1 -2.576  0  1
## CI_2.5  NA -2.811 NA NA
## CI_97.5 NA -2.342 NA NA
##
```

```
## $it8
##         a1     d  g  u
## par      1 1.238  0  1
## CI_2.5  NA 1.051 NA NA
## CI_97.5 NA 1.425 NA NA
##
## $it9
##         a1     d  g  u
## par      1 1.997  0  1
## CI_2.5  NA 1.788 NA NA
## CI_97.5 NA 2.206 NA NA
##
## $it10
##         a1      d  g  u
## par      1 -1.314  0  1
## CI_2.5  NA -1.502 NA NA
## CI_97.5 NA -1.126 NA NA
##
## $GroupPars
##         MEAN_1 COV_11
## par          0  2.187
## CI_2.5      NA  1.879
## CI_97.5     NA  2.495
```

```
# Ako želimo samo parametre težine i diskriminativnosti možemo koristiti funkcije
MDIFF() i MDISC()
cbind(MDIFF(mr.m), MDISC(mr.m))
```

```
##          MDIFF_1
## it1  -0.12625506 1
## it2   0.06504211 1
## it3  -0.10372187 1
## it4  -0.29590631 1
## it5  -1.92123844 1
## it6  -1.79168151 1
## it7   2.57643212 1
## it8  -1.23783078 1
## it9  -1.99711544 1
## it10  1.31442113 1
```

Primetite da u ispisu koji dobijemo funkcijom *coef()* kao oznake parametara stoje a1, d, g, u.

- *a1* je oznaka za parametar nagiba/diskriminativnosti. Broj 1 je tu zato što *mirt* može da procenjuje i višedimenzionalne modele gde svaka dimenzija ima svoj parametar nagiba.
- *d* je oznaka za parametar težine (engl. *difficulty*). I on može imati brojčane dodatke kada su u pitanju politomne stavke jer tu označava parametre kategorija.
- *g* je oznaka za parametar pseudopogađanja (engl. *guessing*).
- *u* je oznaka za parametar gornje asimptote (engl. *upper*).

Dakle, moguće je proceniti jednodimenzionalne ili višedimenzionalne modele za binarne ili politomne stavke, od jednoparametarskih do petoparametarskih.

Uporedićemo procene parametara sa onim dobijenim paketima *eRm* i *ltm.*

```
tezine <- round(data.frame(cbind(mr.e$betapar*(-1), coef(mr.1)[,1],
 MDIFF(mr.m))),3)
names(tezine) <- c("eRm", "ltm", "mirt")

##              eRm    ltm   mirt
## beta it1   0.230 -0.111 -0.126
## beta it2   0.422  0.061  0.065
## beta it3   0.253 -0.091 -0.104
## beta it4   0.060 -0.264 -0.296
## beta it5  -1.571 -1.731 -1.921
## beta it6  -1.441 -1.613 -1.792
## beta it7   2.915  2.326  2.576
## beta it8  -0.884 -1.112 -1.238
## beta it9  -1.648 -1.800 -1.997
## beta it10  1.664  1.184  1.314

cat("Pirsonove korelacije procena parametara iz tri paketa:\n")

## Pirsonove korelacije procena parametara iz tri paketa:

cor(tezine, method = "pearson")

##            eRm       ltm      mirt
## eRm  1.0000000 0.9999941 0.9999941
## ltm  0.9999941 1.0000000 0.9999994
## mirt 0.9999941 0.9999994 1.0000000

##
## Spirmanove korelacije procena parametara iz tri paketa:

##      eRm ltm mirt
## eRm    1   1    1
## ltm    1   1    1
## mirt   1   1    1
```

Iako procene parametara nisu identične, one skoro savršeno koreliraju kada koristimo Pirsonov koeficijent korelacije. Rang korelacija je savršena.

Sličnije su procene parametara iz paketa koji koriste isti metod procene parametara, a to su *ltm* i *mirt.*

Upotrebom funkcije *summary()* na modelima procenjenim u paketu *mirt* nećemo dobiti IRT parametre već faktorska opterećenja i komunalitete.

```
summary(mr.m)

##         F1   h2
## it1  0.656 0.43
## it2  0.656 0.43
## it3  0.656 0.43
## it4  0.656 0.43
## it5  0.656 0.43
## it6  0.656 0.43
## it7  0.656 0.43
## it8  0.656 0.43
## it9  0.656 0.43
## it10 0.656 0.43
##
## SS loadings:  4.302
```

```
## Proportion Var:  0.43
##
## Factor correlations:
##
##    F1
## F1  1
```

Identično rešenje dobili bismo primenom konfirmativne faktorske analize u paketu *lavaan* (Rosseel, 2012) kada bismo varijable tretirali kao ordinalne i koristili *DWLS* procenitelj. Pošto Rašov model pretpostavlja jednake parametre nagiba za sve stavke i u *lavaanu* bi trebalo ograničiti opterećenja svih stavki da budu jednaka.

#### *10.4.3.2 Pokazatelji fita u paketu mirt*

U paketu *mirt* moguće je dobiti omnibus pokazatelje fita za model i pokazatelje fita za stavke.

10.4.3.2.1 Pokazatelji fita modela u celini

Jedan od pokazatelja fita već je prikazan ($G^2$). Drugi pokazatelj fita bazira se na ograničenim informacijama i radi se o pokazatelju $M_2$ (videti odeljak 6.1.1). Osim pokazatelja $M_2$, dostupni su i $M_2^*$ i $C_2$. Poslednja dva primenjuju sažimanje margina prvog i drugog reda ili samo drugog reda (respektivno). $C_2$ je koristan kod modela sa malim brojem stepeni slobode[51]. U paketu *mirt* i funkciji $M_2()$ podrazumevana je varijanta $M_2^*$, dok je $M_2$ dostupan samo za binarno skorovane stavke.

$M_2$ je pokazatelj bliskog fita i u modelima sa velikim brojem stepeni slobode možemo očekivati da će biti značajan što ukazuje na misfit (Maydeu-Olivares, 2013). S obzirom na to, prilikom interpretacije moraćemo se osloniti i na pokazatelje približnog fita, a ovde su to pokazatelji inače karakteristični za strukturalno modelovanje: *CFI*, *TLI*, $RMSEA_2$ i *SRMR*[52]. Za ove poslednje možemo koristiti uobičajene kritične (cut–off) vrednosti koje se koriste i u strukturalnom modelovanju (Hu & Bentler, 1999). Vrednosti *CFI* i *TLI* ≥0,90 smatraju se prihvatljivim fitom, a ≥0,95 dobrom saglasnošću modela i podataka. Vrednost *RMSEA* bi trebalo da bude <0,08 ako je fit prihvatljiv, a <0.05 za dobar fit. Poželjno je da granice intervala poverenja $RMSEA_2$ ne obuhvataju vrednost 0,08 pošto se preko te vrednosti fit ne smatra prihvatljivim. Vrednost *SRMR* (standardizovani prosečni rezidual dobijen *root mean square* transformacijom) bi trebalo da bude <0,08 ako je fit dobar.

Ako je uzorak veliki, možemo zanemariti značajan $M_2$ statistik i fit modela tumačiti na osnovu preostalih pokazatelja (de Ayala, 2022).

[51] Ako dobijete poruku:

```
'M2() statistic cannot be calculated due to too few degrees of freedom' probajte s
a C2.
```

[52] Ovde se $RMSEA_2$ i ostali pokazatelji računaju na osnovu pokazatelja $M_2$ (videti odeljak 6.1.1)(Chalmers, 2012).

```
M2(mr.m, type="M2*")

##             M2 df          p      RMSEA     RMSEA_5   RMSEA_95      SRMSR
## stats 61.68182 44 0.04024028 0.02005646 0.004462893 0.03108157 0.03492453
##             TLI       CFI
## stats 0.9930448 0.9931993

# Ako ne navedemo argument type="M2", podrazumevan je pokazatelj "M2*". Moguće je
izabrati i "C2".
```

Pokazatelj bliskog fita $M_2$ je statistički značajan (mada je blizu granične vrednosti), ali ostali pokazatelji ukazuju na dobar fit ovog modela. $RMSEA_2$=0,02 (90% IP 0,004–0,031) što je manje od granične vrednosti za dobar fit od 0,05, i *TLI* i *CFI* su viši od 0,99 a granična vrednost za dobar fit za oba pokazatelja je >0,95. *SRMSR*=0,035, a granična vrednost ovog pokazatelja za dobar fit je <0,08. S obzirom na veličinu uzorka možemo reći da je model dovoljno saglasan sa podacima.

#### *10.4.3.2.1.1 Pokazatelji fita za stavke*

Paket *mirt* nudi pokazatelje fita za stavke bazirane na hi–kvadrat testu, logaritmu verodostojnosti i na standardizovanim rezidualima. Mogu se naći svi pokazatelji prikazani u odeljku 6.3 i neki dodatni.

Funkcija kojom dobijamo pokazatelje fita za stavke je *itemfit()*. Ranije je pomenuto da funkcija sa istim nazivom postoji i u biblioteci *eRm* i koja će od njih biti pozvana zavisi od toga koja biblioteka je kasnije učitana. Ako želite da pozovete funkciju baš iz biblioteke *mirt* možete to eksplicitno tražiti na sledeći način *mirt::itemfit(naziv_modela).*

Argument kojim zahtevate određene pokazatelje fita je *fit_stats*=c("S_X2", "Zh", "X2", "G2", "PV_Q1", "PV_Q1*", "X2*", "X2*_df", "infit"). U primeru je spisak u formi slovnog vektora koji može sadržati više različitih pokazatelja.

"SX_2" označava hi–kvadrat po proceduri Orlanda i Tisena, "Zh" – pokazatelj zasnovan na logaritmu verodostojnosti Drazgova, Levina i Viliamsa, "X2" – Bokov hi–kvadrat test (mada je podrazumevan $Q_1$ statistik Jenove sa 10 grupa i medijanom kao procenom osobine u grupama), "G2" – hi–kvadrat Mekinlija i Milsa zasnovan na količniku verodostojnosti, "PV_Q1" i "PV_Q1*" – $Q_1$ statistik zasnovan na plauzibilnim vrednostima sa i bez Monte Karlo simulacija (Chalmers & Ng, 2017), "X2*" i "X2*_df" – hi–kvadrat statistici zasnovani na Monte Karlo simulacijama (Stone, 2000) i "infit" – pokazatelji infita i autfita zasnovani na standardizovanim rezidualima (MSQ i t).

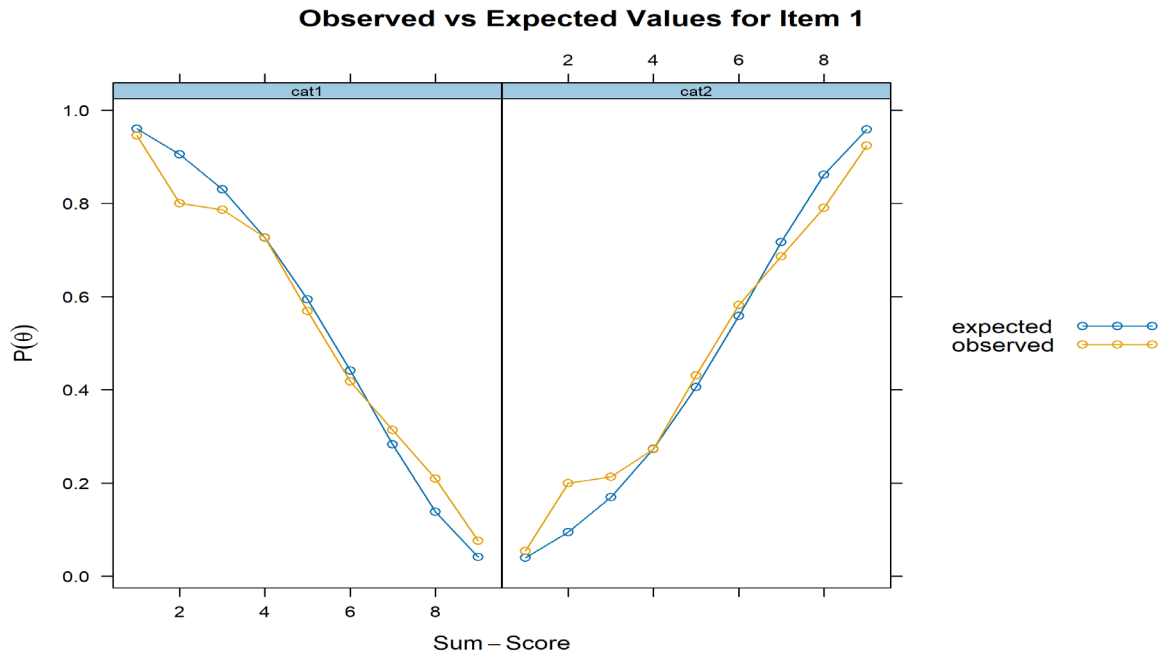
"SX_2" i "Zh" nije moguće dobiti ako postoje nedostajući podaci. U tom slučaju mogu se računati samo na kompletnim podacima ili je potrebno izvršiti imputaciju nedostajućih podataka. Ispitanici sa nedostajućim podacima mogu se ukloniti dodavanjem argumenta *na.rm=TRUE*. Imputacija se može obaviti upotrebom paketa *mice* (Buuren & Groothuis-Oudshoorn, 2011), ali ovde je nećemo objašnjavati.

```
# Ovom naredbom dobijamo podrazumevani hi-kvadrat po proceduri Orlanda i Tisena
itemfit(mr.m)

##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1 18.289       7      0.040  0.011
## 2   it2  9.658       7      0.019  0.209
```

```
## 3   it3 11.676       7      0.026  0.112
## 4   it4  2.562       7      0.000  0.922
## 5   it5  5.893       6      0.000  0.435
## 6   it6  2.723       6      0.000  0.843
## 7   it7  5.524       5      0.010  0.355
## 8   it8 14.391       7      0.033  0.045
## 9   it9  7.569       6      0.016  0.271
## 10 it10  7.000       6      0.013  0.321

# Ovom naredbom smo zahtevali hi-kvadrat po proceduri Orlanda i Tisena i
pokazatelje zasnovane na standardizovanim rezidualima
itemfit(mr.m,  fit_stats=c("S_X2", "infit"))

##    item outfit z.outfit infit z.infit   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1  0.879   -2.598 0.934  -2.138 18.289       7      0.040  0.011
## 2   it2  0.715   -6.556 0.797  -7.001  9.658       7      0.019  0.209
## 3   it3  0.779   -4.962 0.861  -4.658 11.676       7      0.026  0.112
## 4   it4  0.791   -4.525 0.870  -4.257  2.562       7      0.000  0.922
## 5   it5  0.740   -2.519 0.893  -2.138  5.893       6      0.000  0.435
## 6   it6  0.802   -2.003 0.888  -2.360  2.723       6      0.000  0.843
## 7   it7  0.755   -1.559 0.918  -1.207  5.524       5      0.010  0.355
## 8   it8  0.834   -2.273 0.919  -2.069 14.391       7      0.033  0.045
## 9   it9  0.664   -3.239 0.848  -3.013  7.569       6      0.016  0.271
## 10 it10  0.696   -4.081 0.806  -5.364  7.000       6      0.013  0.321
```

Stavke kod kojih postoji značajan misfit na osnovu hi–kvadrat testa su stavke 1 i 7, s tim da je kod stavke 7 on blizu granice značajnosti. Ako posmatramo pokazatelje infita i autfita u MSQ metrici, nijedna stavka nije upadljiva. Obe navedene stavke naginju overfitu (prigušenom šumu). Standardizovani infit i autfit su izvan preporučenog opsega, ali s obzirom da u MSQ metrici stavke ne izlaze iz preporučenog okvira, nećemo tome pridavati veću pažnju.

```
itemfit(mr.m,  fit_stats=c("S_X2", "infit"), S_X2.plot = 1)
```




Slika 41 – Opažene i modelske karakteristične krive kategorija stavke 1

```
# Argument S_X2.plot = 1 traži prikaz modelske i empirijske KKK za stavku koja je
označena brojem (u ovom slučaju to je stavka 1)
```

Što se tiče grafikona (Slika 41), iako je stavka 1 binarno skorovana ovde su prikazane kategorije odgovora 0 i 1. Krive su komplementarne i iste stvari vidimo na obe samo iz različitih uglova. Npr. na prikazu za kategoriju 1 (što je skor 0) vidimo da model malo precenjuje verovatnoću odgovora 0 na stavku 1 kod ispitanika sa ukupnim skorovima 2 i 3.

Ako gledamo prikaz za kategoriju 2 (što je ajtemski skor 1), možemo reći da model potcenjuje verovatnoću skora 1 na stavku 1 kod ispitanika sa ukupnim skorom 2 i 3.

Kod realnih podataka ovakvi grafikoni mogu nam pomoći da shvatimo razlog zbog kojeg neka stavka ima misfit.

#### *10.4.3.2.1.2 Pokazatelji fita za ispitanike u paketu mirt*

Pokazatelje fita za ispitanike u paketu *mirt* dobijamo funkcijom *personfit().* Podrazumevani su pokazatelji fita zasnovani na standardizovanim rezidualima (infit i autfit u MSQ i standardizovanoj metrici), te pokazatelj $Z_h$ Drazgova i saradnika, baziran na logaritmu verodostojnosti (ovde je prikazan u standardizovanoj metrici).

```
PF.m <- personfit(mr.m, method = "ML")
head(PF.m)
```

```
##      outfit     z.outfit     infit     z.infit           Zh
## 1 0.3797929 -0.968456135 0.4448185 -2.09880634  1.589058910
## 2 0.8465402 -0.004622173 0.9557541 -0.03230262  0.164273771
## 3 0.5590177 -0.763093274 0.7368196 -0.70969903  0.779779500
## 4 0.8758526  0.173163891 1.0144080  0.15561853 -0.002152688
## 5 0.9795414  0.149579361 1.0075185  0.13000269 -0.026363013
## 6 0.7436884 -0.159594649 1.0534475  0.26598428  0.063179647
```

Zbog prostora prikazani su pokazatelji samo za prvih 6 ispitanika. Interpretacija ovih pokazatelja opisana je ranije (odeljak 6.4). Kao kratki podsetnik, prihvatljivi opseg infita i autfita je 0,7–1,3 u MSQ metrici, –1,96–1,96 u standardizovanoj metrici. Za $Z_h$ važi isti opseg kao za standardizovani infit i autfit.

Način da napravite rezime fita ispitanika biće prikazan kasnije (npr. odeljak 10.5.5.3).

Niske vrednosti *infita* i *autfita* ukazuju na prigušeni šum, a visoke na šum. Kod $Z_h$ pokazatelja je obrnuto.

Inače objekat u koji smo smestili rezultate je klase *data.frame* isto kao matrica podataka i lako se mogu spojiti. Pokazatelji misfita se kasnije mogu koristiti za selekciju ispitanika.

```
# Spojene matrice ćemo zapisati u novi data.frame
b.podMF <- data.frame(cbind(b.pod, PF.m))
```

#### 10.4.3.2.2 Lokalna zavisnost

Proveru lokalne zavisnosti u paketu *mirt* možemo obaviti na više načina.

Prvi se zasniva na hi kvadratu za parove stavki. Dobijamo ga komandom *residuals* ako koristimo

argument *type="LD"*.

```
residuals(mr.m, type="LD") # Ako dodamo argument suppres=.20 neće biti prikazane
korelacije manje od .20 (korelacije veće od te vrednosti ukazuju na lokalnu
zavisnost)

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.078  -0.032   0.006   0.003   0.031   0.077
##
##        it1    it2    it3    it4    it5    it6    it7    it8    it9  it10
## it1     NA -0.061 -0.043 -0.032 -0.039 -0.055 -0.055 -0.063 -0.013 0.006
## it2  3.747     NA  0.035  0.010  0.045  0.077  0.059  0.025  0.064 0.009
## it3  1.869  1.206     NA -0.039  0.012 -0.012 -0.003 -0.012  0.001 0.031
## it4  1.011  0.102  1.550     NA  0.023  0.008 -0.005 -0.032  0.027 0.001
## it5  1.542  2.042  0.140  0.540     NA  0.033 -0.078 -0.046  0.059 0.034
## it6  3.035  5.977  0.134  0.069  1.090     NA -0.030 -0.033  0.017 0.028
## it7  3.052  3.475  0.010  0.026  6.062  0.881     NA -0.019  0.039 0.021
## it8  3.991  0.648  0.151  1.049  2.098  1.110  0.343     NA  0.048 0.002
## it9  0.171  4.137  0.002  0.714  3.513  0.273  1.509  2.322     NA 0.072
## it10 0.041  0.081  0.931  0.002  1.144  0.760  0.422  0.004  5.140    NA

residuals(mr.m, type=c("LD"), df.p=T, suppres=.20)

## Degrees of freedom (lower triangle) and p-values:
##
##      it1   it2   it3   it4   it5   it6   it7   it8   it9  it10
## it1   NA 0.053 0.172 0.315 0.214 0.081 0.081 0.046 0.679 0.840
## it2    1    NA 0.272 0.750 0.153 0.014 0.062 0.421 0.042 0.777
## it3    1 1.000    NA 0.213 0.708 0.714 0.919 0.698 0.968 0.335
## it4    1 1.000 1.000    NA 0.463 0.793 0.871 0.306 0.398 0.969
## it5    1 1.000 1.000 1.000    NA 0.297 0.014 0.147 0.061 0.285
## it6    1 1.000 1.000 1.000 1.000    NA 0.348 0.292 0.602 0.383
## it7    1 1.000 1.000 1.000 1.000 1.000    NA 0.558 0.219 0.516
## it8    1 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.128 0.952
## it9    1 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.023
## it10   1 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA
##
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.078  -0.032   0.006   0.003   0.031   0.077
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Ako ne zatražimo ispis stepeni slobode (*df.p=T*), u gornjem trouglu matrice biće date korelacije

(Kramerovo V), a u donjem vrednosti hi–kvadrata. U protivnom, u donjem trouglu matrice biće prikazani stepeni slobode.

Vidimo da na našim podacima ne postoji lokalna zavisnost (u tabeli *upper triangle summary* nema vrednosti većih od 0,20, pa su sve zamenjene sa NA zbog argumenta *suppres=.20*).

Drugi način provere lokalne zavisnosti je hi–kvadrat zasnovan na logaritmu verodostojnosti ($G^2$), takođe za parove stavki. Njega dobijamo argumentom *type="LDG2"*.

```
residuals(mr.m, type="LDG2")

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.071  -0.032   0.006   0.003   0.031   0.079
##
##        it1    it2    it3    it4    it5    it6    it7    it8    it9  it10
## it1     NA -0.061 -0.043 -0.032 -0.039 -0.054 -0.054 -0.063 -0.013 0.006
## it2  3.704     NA  0.035  0.010  0.046  0.079  0.061  0.026  0.066 0.009
## it3  1.854  1.214     NA -0.039  0.012 -0.012 -0.003 -0.012  0.001 0.031
## it4  1.005  0.102  1.539     NA  0.023  0.008 -0.005 -0.032  0.027 0.001
## it5  1.509  2.106  0.141  0.545     NA  0.033 -0.071 -0.046  0.059 0.035
## it6  2.956  6.298  0.133  0.069  1.089     NA -0.028 -0.033  0.017 0.028
## it7  2.898  3.711  0.010  0.026  4.971  0.812     NA -0.018  0.042 0.021
## it8  3.914  0.655  0.150  1.040  2.090  1.107  0.330     NA  0.048 0.002
## it9  0.169  4.346  0.002  0.723  3.502  0.272  1.766  2.335     NA 0.077
## it10 0.041  0.081  0.944  0.002  1.211  0.793  0.424  0.004  5.948    NA

residuals(mr.m, type="LDG2", df.p=T, suppres=.20)

## Degrees of freedom (lower triangle) and p-values:
##
##      it1   it2   it3   it4   it5   it6   it7   it8   it9  it10
## it1   NA 0.054 0.173 0.316 0.219 0.086 0.089 0.048 0.681 0.840
## it2    1    NA 0.270 0.749 0.147 0.012 0.054 0.418 0.037 0.776
## it3    1 1.000    NA 0.215 0.707 0.715 0.919 0.698 0.968 0.331
## it4    1 1.000 1.000    NA 0.460 0.793 0.871 0.308 0.395 0.969
## it5    1 1.000 1.000 1.000    NA 0.297 0.026 0.148 0.061 0.271
## it6    1 1.000 1.000 1.000 1.000    NA 0.367 0.293 0.602 0.373
## it7    1 1.000 1.000 1.000 1.000 1.000    NA 0.565 0.184 0.515
## it8    1 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.126 0.952
## it9    1 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.015
## it10   1 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA
##
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.071  -0.032   0.006   0.003   0.031   0.079
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

```
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Ispis za ovaj način provere lokalne zavisnosti isti je kao i za prethodni, a i nalazi u pogledu postojanja lokalne zavisnosti su isti.

I na kraju, ranije opisani pokazatelj $Q_3$, odnosno, korelacija reziduala (odeljak 9.2). Dobijamo ga korišćenjem argumenta *type="Q3"*.

```
residuals(mr.m, type="Q3")

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.193  -0.117  -0.087  -0.087  -0.055  -0.003
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.193 -0.126 -0.106 -0.110 -0.129 -0.105 -0.130 -0.104 -0.078
## it2  -0.193  1.000 -0.098 -0.130 -0.055 -0.007 -0.006 -0.080 -0.055 -0.152
## it3  -0.126 -0.098  1.000 -0.163 -0.072 -0.102 -0.070 -0.097 -0.117 -0.087
## it4  -0.106 -0.130 -0.163  1.000 -0.061 -0.080 -0.062 -0.128 -0.086 -0.117
## it5  -0.110 -0.055 -0.072 -0.061  1.000 -0.065 -0.099 -0.157 -0.048 -0.036
## it6  -0.129 -0.007 -0.102 -0.080 -0.065  1.000 -0.054 -0.138 -0.124 -0.038
## it7  -0.105 -0.006 -0.070 -0.062 -0.099 -0.054  1.000 -0.051 -0.003 -0.087
## it8  -0.130 -0.080 -0.097 -0.128 -0.157 -0.138 -0.051  1.000 -0.042 -0.078
## it9  -0.104 -0.055 -0.117 -0.086 -0.048 -0.124 -0.003 -0.042  1.000 -0.007
## it10 -0.078 -0.152 -0.087 -0.117 -0.036 -0.038 -0.087 -0.078 -0.007  1.000

residuals(mr.m, type="Q3", suppres=.2)

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.193  -0.117  -0.087  -0.087  -0.055  -0.003
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1    1
## it2        1
## it3            1
## it4                1
## it5                    1
## it6                        1
## it7                            1
## it8                                1
## it9                                    1
## it10                                       1
```

Kada se koristi granična vrednost korelacije reziduala od 0,36 koja kao značajan rezidual uzima korelaciju srednje veličine (Cohen, 1988; Zhang i ostali, 2019) u našim podacima ne postoji lokalna zavisnost. Ni kada granicu spustimo na 0,20 nema indicija da postoji lokalna zavisnost.

Alternativni način je da poredimo vrednost $Q_3$ za stavke sa prosečnom vrednošću u matrici.

```
Q3m <- as.matrix(residuals(mr.m, type="Q3"))

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.193  -0.117  -0.087  -0.087  -0.055  -0.003
```

```
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.193 -0.126 -0.106 -0.110 -0.129 -0.105 -0.130 -0.104 -0.078
## it2  -0.193  1.000 -0.098 -0.130 -0.055 -0.007 -0.006 -0.080 -0.055 -0.152
## it3  -0.126 -0.098  1.000 -0.163 -0.072 -0.102 -0.070 -0.097 -0.117 -0.087
## it4  -0.106 -0.130 -0.163  1.000 -0.061 -0.080 -0.062 -0.128 -0.086 -0.117
## it5  -0.110 -0.055 -0.072 -0.061  1.000 -0.065 -0.099 -0.157 -0.048 -0.036
## it6  -0.129 -0.007 -0.102 -0.080 -0.065  1.000 -0.054 -0.138 -0.124 -0.038
## it7  -0.105 -0.006 -0.070 -0.062 -0.099 -0.054  1.000 -0.051 -0.003 -0.087
## it8  -0.130 -0.080 -0.097 -0.128 -0.157 -0.138 -0.051  1.000 -0.042 -0.078
## it9  -0.104 -0.055 -0.117 -0.086 -0.048 -0.124 -0.003 -0.042  1.000 -0.007
## it10 -0.078 -0.152 -0.087 -0.117 -0.036 -0.038 -0.087 -0.078 -0.007  1.000
```

```
diag(Q3m)<- NA
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\n\nProsečna vrednost Q3:", Q3mean, "\n\n")
```

```
##
## Prosečna vrednost Q3: -0.08733607
```

```
Q3c <- Q3m-Q3mean
Q3c
```

```
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.106 -0.039 -0.018 -0.023 -0.041 -0.018 -0.043 -0.017  0.009
## it2  -0.106     NA -0.010 -0.043  0.033  0.081  0.081  0.008  0.033 -0.065
## it3  -0.039 -0.010     NA -0.075  0.016 -0.014  0.018 -0.010 -0.030  0.001
## it4  -0.018 -0.043 -0.075     NA  0.027  0.007  0.026 -0.041  0.002 -0.030
## it5  -0.023  0.033  0.016  0.027     NA  0.023 -0.012 -0.069  0.039  0.051
## it6  -0.041  0.081 -0.014  0.007  0.023     NA  0.033 -0.051 -0.036  0.049
## it7  -0.018  0.081  0.018  0.026 -0.012  0.033     NA  0.037  0.085  0.001
## it8  -0.043  0.008 -0.010 -0.041 -0.069 -0.051  0.037     NA  0.046  0.009
## it9  -0.017  0.033 -0.030  0.002  0.039 -0.036  0.085  0.046     NA  0.080
## it10  0.009 -0.065  0.001 -0.030  0.051  0.049  0.001  0.009  0.080     NA
```

U poslednjoj prikazanoj matrici su vrednosti $Q_3$ umanjene za prosečnu vrednost u matrici. Pošto je prosečna vrednost negativna (–0,87, a obično i jeste negativna), pozitivne vrednosti $Q_3$ su praktično uvećane za apsolutnu vrednost proseka $Q_3$. Ako ovakve vrednosti prelaze 0,20, to ukazuje na postojanje lokalne zavisnosti. Takvih vrednosti ovde nema.

#### *10.4.3.3 Informativnost*

Informativnost u određenom opsegu osobine može se izračunati funkcijom *areainfo()*.

```
# Za ceo test
MRI <- areainfo(mr.m, c(-3,3))
MRI # Objekat u kojem smo sačuvali podatke o informativnosti
```

```
##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 8.111699        10  0.8111699     10
```

```
# Samo za određene stavke
areainfo(mr.m, c(-3,3), which.items = c(2,7:10))
```

```
##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 3.900145         5  0.7800291      5
```

```
# stavke 2 i od 7 do 10
```

Tabela (mada ne liči na tabelu) prikazuje donju i gornju granicu intervala za koji smo tražili informativnost, ukupnu informativnost u tom opsegu, ukupnu informativnost u opsegu od –10 do 10 logita (*TotalInfo*), proporciju informativnosti u traženom opsegu u odnosu na *TotalInfo* i broj stavki.

Informativnost po nivoima merene osobine možemo dobiti na sledeći način:

```
# Prvo definišemo vrednosti latentne osobine za koje želimo da izračunamo
informativnost (sekvenca u rasponu od -3 do 3 sa koracima od 0,1)
thetaM <- seq(-3,3, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(mr.m, Theta = thetaM)
# Prikaz prvih 6 redova (uklonite head() za ispis cele tabele)
head(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
```

```
##      nivo osobine informativnost
## [1,]         -3.0      0.9062269
## [2,]         -2.9      0.9631966
## [3,]         -2.8      1.0209857
## [4,]         -2.7      1.0792363
## [5,]         -2.6      1.1375661
## [6,]         -2.5      1.1955757
```

```
# Maksimalna informativnost
max(infoM)
```

```
## [1] 1.820823
```

```
# Ispis nivoa osobine na kom je informativnost maksimalna
thetaM[infoM==max(infoM)]
```

```
## [1] -0.6
```

```
# Ispis raspona u kojem je informativnost veća od 3,33 što odgovara pouzdanosti od
0,7
thetaM[infoM>3.33]
```

```
## numeric(0)
```

Vidimo da je maksimalna informativnost testa 1,82 na nivou osobine od –0,6 logita. Najčešće nam nije toliko bitna maksimalna informativnost, koliko nam je bitan raspon u kojem je informativnost prihvatljiva. Prihvatljivu informativnost možemo definisati preko odnosa sa pouzdanošću iz klasične testne teorije, a taj odnos je $I=1/(1-r_{tt})$. Ako kao prihvatljivu pouzdanost uzmemo vrednost od 0,7, to odgovara informativnosti od 3,33. U poslednjoj liniji koda tražili smo da nam prikaže vrednosti nivoa osobine kod kojih je informativnost veća od 3,33. Pošto smo u ispisu dobili *numeric(0)*, to znači da ne postoji takav nivo osobine, odnosno da test ne dostiže željeni nivo informativnosti ni na jednom nivou latentne osobine.

```
# Možemo izračunati pouzdanost po nivoima osobine na osnovu poznatog odnosa sa
informativnošću rtt=1-(1/infoM)
rtt <- ifelse(!(1-(1/infoM))<0, 1-(1/infoM), 0)
# Možemo prikazati opseg nivoa osobine u kojem je pouzdanost iznad nekog zadatog
nivoa
cat("Opseg nivoa osobine u kojem je pouzdanost iznad 0,7: \n")
```

```
## Opseg nivoa osobine u kojem je pouzdanost iznad 0,7:
```

```
thetaM[rtt>.7]
```

```
## numeric(0)

# Ako kao rezultat dobijete ispis numeric(0), to znači da pouzdanost nikad nije
viša od zadate vrednosti
# Možemo izračunati standardnu grešku merenja po nivoima osobine na osnovu
poznatog odnosa sa informativnošću se=1/sqrt(infoM)
se <- 1/sqrt(infoM)
# Možemo prikazati opseg nivoa osobine u kojem je standardna greška merenja niža
od nekog zadatog nivoa
thetaM[se<.548]

## numeric(0)

head(cbind("nivo osobine"=thetaM, "informativnost"=round(infoM, 3),
 "pouzdanost"=round(rtt,3), "sgm"=round(se, 3)))

##      nivo osobine informativnost pouzdanost   sgm
## [1,]         -3.0          0.906      0.000 1.050
## [2,]         -2.9          0.963      0.000 1.019
## [3,]         -2.8          1.021      0.021 0.990
## [4,]         -2.7          1.079      0.073 0.963
## [5,]         -2.6          1.138      0.121 0.938
## [6,]         -2.5          1.196      0.164 0.915
```

Inače, paket *mirt* ima i funkciju za računanje empirijske i marginalne pouzdanosti. Marginalna pouzdanost bazira se na prosečnoj varijansi greške. Prosečna varijansa greške računa se ponderisanjem varijanse greške na svakom od nivoa osobine proporcijom tog nivoa osobine u populaciji (pretpostavlja se standardizovana normalna distribucija)(Thissen & Wainer, 2001). Izraz za računanje marginalne pouzdanosti za standardizovanu $\theta$ je:

$$\bar{\rho} = 1 - \overline{SE}^2(\theta) \qquad [119]$$

Empirijska pouzdanost se zasniva na istom principu, ali koristi vrednosti iz uzorka.

Za računanje empirijske pouzdanosti potrebne su nam procene nivoa osobine ispitanika i standardne greške vezane za njih, a za marginalnu objekat sa modelom koji je procenjen.

```
# Parametre ispitanika i standardne greške dobijamo funkcijom fscores() i smeštamo
u objekat THETA_SE koji prosleđujemo funkciji za računanje empirijske pouzdanosti
THETA_SE <- fscores(mr.m, full.scores.SE = TRUE)
head(THETA_SE)

##              F1     SE_F1
## [1,] -0.7963690 0.6767522
## [2,] -0.7963690 0.6767522
## [3,]  0.1296586 0.6926259
## [4,] -1.2608076 0.6886454
## [5,] -0.3387503 0.6783190
## [6,]  0.6273521 0.7206811

cat("\nEmpirijska pouzdanost:", empirical_rxx(THETA_SE))

##
## Empirijska pouzdanost: 0.753239

cat("\nMarginalna pouzdanost:", marginal_rxx(mr.m))
```

```
##
## Marginalna pouzdanost: 0.6102668

cat("\nKronbahova alfa:", psych::alpha(b.pod)$total$raw_alpha)

##
## Kronbahova alfa: 0.7572566
```

Empirijska pouzdanost bliska je pouzdanosti iz klasične testne teorije, konkretno Kronbahovoj alfi. Međutim, pokazatelj marginalne pouzdanosti (ukoliko je tačna pretpostavka o normalnoj distribuciji osobine) ukazuje da ovaj model ne bi dao dovoljno pouzdana merenja.

Kao što je rečeno, iako je lepo dobiti procenu pouzdanosti u obliku jednog broja, korisnije nam je znati u kom rasponu osobine je merenje prihvatljivo precizno, odnosno, informativno.

##### 10.4.3.3.1 Mere ispitanika u paketu *mirt*

Mere ispitanika na osnovu procenjenog modela u paketu *mirt* dobijamo funkcijom *fscores()*, kao što je već prikazano u prethodnom odeljku. Prikazaćemo kako da ih povežete sa sirovim podacima.

```
# fscores(mr.m)
# Za razliku od paketa ltm gde smo mere dobijali za sklopove odgovora, a ne za
ispitanike, ovde dobijamo meru za svakog ispitanika i možemo je dodati direktno u
matricu podataka
# Opet ćemo napraviti kopiju podataka kako bismo sačuvali originalnu matricu
b.podK2 <- b.pod
# Nakon toga jednostavno definišemo novu varijablu u matrici koja će biti jednaka
rezultatu funkcije fscores()
b.podK2$MERE <- fscores(mr.m)
head(b.podK2)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10         F1
## 1   0   0   0   0   1   1   0   1   1    0 -0.7963690
## 2   0   1   0   0   0   1   0   1   1    0 -0.7963690
## 3   0   1   1   0   1   1   0   1   1    0  0.1296586
## 4   1   0   0   0   1   1   0   0   0    0 -1.2608076
## 5   0   0   1   1   1   1   0   1   0    0 -0.3387503
## 6   0   0   1   1   1   1   0   1   1    1  0.6273521
```

### 10.4.4 Dvoparametarski logistički model (2PL) za binarne stavke

Paket *eRm* je namenjen Rašovoj familiji modela pa se ne može koristiti za procenu višeparametarskih modela. Za ove modele koristićemo pakete *ltm* i *mirt.*

#### *10.4.4.1 2PL model za binarno skorovane stavke u paketu ltm*

Učitaćemo paket *ltm*.

```
p_load(ltm)
```

Ranije smo već pokazali način procene dvoparametarskog modela u paketu *ltm* koristeći funkciju *tpm()*.

```
m2p.lt <- tpm(b.pod, max.guessing = 0)
m2p.lt

##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Coefficients:
##       Gussng  Dffclt  Dscrmn
## it1        0  -0.104   1.095
## it2        0   0.040   1.782
## it3        0  -0.073   1.417
## it4        0  -0.208   1.384
## it5        0  -1.257   1.575
## it6        0  -1.197   1.514
## it7        0   1.756   1.459
## it8        0  -0.900   1.304
## it9        0  -1.195   1.901
## it10       0   0.822   1.730
##
## Log.Lik: -5181.038
```

Isti model moguće je dobiti funkcijama *grm()* i *ltm()* na sledeći način.

```
m2p.lg <- grm(b.pod)
m2p.lg

##
## Call:
## grm(data = b.pod)
##
## Coefficients:
##       Extrmt1  Dscrmn
## it1    -0.104   1.095
## it2     0.040   1.782
## it3    -0.073   1.417
## it4    -0.208   1.384
## it5    -1.257   1.575
## it6    -1.197   1.514
## it7     1.756   1.459
## it8    -0.900   1.304
## it9    -1.195   1.901
## it10    0.822   1.730
##
## Log.Lik: -5181.038

m2p.ll <- ltm(b.pod~z1)
m2p.ll

##
## Call:
## ltm(formula = b.pod ~ z1)
##
## Coefficients:
##       Dffclt  Dscrmn
## it1   -0.104   1.095
## it2    0.040   1.782
## it3   -0.073   1.417
## it4   -0.208   1.384
## it5   -1.257   1.575
```

```
## it6   -1.197   1.514
## it7    1.756   1.459
## it8   -0.900   1.304
## it9   -1.195   1.901
## it10   0.822   1.730
##
## Log.Lik: -5181.038
```

Funkcija *grm()* namenjena je modelu za stepenovano odgovarnje, koji je opet namenjen stavkama sa bilo kojim brojem uređenih, ordinalnih kategorija. Dakle, ova funkcija može procenjivati modele i za politomne stavke, a broj kategorija po stavkama ne mora biti isti (ali sve moraju počinjati od iste najniže kategorije)[53]. Kada su sve stavke binarno skorovane, ovaj model se svodi na 2PL model za binarno skorovane stavke.

Funkcija *ltm()* može da procenjuje i dvodimenzionalne modele. Model se definiše formulom čiji su elementi podaci na kojima se vrši analiza i broj dimenzija koji se određuje argumentima ~z1+z2 (`b.pod~z1+z2` je primer za dvodimenzionalni, a u gornjoj naredbi je dat primer za jednodimenzionalni model).

Ako uporedimo rezultate procene različitim funkcijama videćemo da su identični.

```
cbind(round(coef(m2p.lt)[,2:3], 3), coef(m2p.lg), round(coef(m2p.ll),3))
```

```
##      Dffclt Dscrmn Extrmt1 Dscrmn Dffclt Dscrmn
## it1  -0.104  1.095  -0.104  1.095 -0.104  1.095
## it2   0.040  1.782   0.040  1.782  0.040  1.782
## it3  -0.073  1.417  -0.073  1.417 -0.073  1.417
## it4  -0.208  1.384  -0.208  1.384 -0.208  1.384
## it5  -1.257  1.575  -1.257  1.575 -1.257  1.575
## it6  -1.197  1.514  -1.197  1.514 -1.197  1.514
## it7   1.756  1.459   1.756  1.459  1.756  1.459
## it8  -0.900  1.304  -0.900  1.304 -0.900  1.304
## it9  -1.195  1.901  -1.195  1.901 -1.195  1.901
## it10  0.822  1.730   0.822  1.730  0.822  1.730
```

Prve dve kolone su nastale funkcijom *tpm()*, druge dve funkcijom *grm()*, a poslednje dve funkcijom *ltm()*.

Kao što ste do sada već videli, ispis parametara modela vrši se pozivanjem imena objekta u kojem je model smešten ili funkcijom *coef()*.

```
m2p.lt
```

```
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Coefficients:
##       Gussng  Dffclt  Dscrmn
## it1        0  -0.104   1.095
## it2        0   0.040   1.782
```

[53] Npr. mogu biti 3 stavke sa kategorijama: 1–5, 1–5, 1–3, ali ne mogu imati kategorije 1–5, 1–5, 2–4. U poslednjem slučaju potrebno je poslednju stavku rekodirati u kategorije 1–3.

```
## it3         0  -0.073   1.417
## it4         0  -0.208   1.384
## it5         0  -1.257   1.575
## it6         0  -1.197   1.514
## it7         0   1.756   1.459
## it8         0  -0.900   1.304
## it9         0  -1.195   1.901
## it10        0   0.822   1.730
##
## Log.Lik: -5181.038
```

U ovom modelu imamo dva skupa parametara: težine i diskriminativnosti. Parametri težine su u modelima dobijenim funkcijama *tpm()* i *ltm()* označeni sa *Dffclt*, a u modelu dobijenom funkcijom *grm()* sa *Extrmt.1*. Razlog tome je što je osnovna namena funkcije *grm()* procena modela stepenovanog odgovaranja za politomne stavke koji ne procenjuje težinu stavki već težine kategorija kojih može biti više (otud brojčana oznaka). Diskriminativnosti su u svim funkcijama označene sa *Dscrmn*.

Parametri težine stavki nam i dalje govore kom nivou osobine su stavke najprimerenije i gde najpreciznije mere. Informativnost stavki će i dalje biti najveća u tački preloma KKS (na $b_j$), ali ovoga puta će zavisiti i od parametra diskriminativnosti. Što je parametar diskriminativnosti viši, biće i viša informativnost stavke.

Najnižu diskriminativnost ima stavka 1 $a_1$=1,095, a najvišu stavka 9 $a_9$=1,901.

Prihvatljivi parametri diskriminativnosti kreću se u rasponu od 0,8 do 2,5 (de Ayala, 2022). Stavke sa nižim parametrima diskriminativnosti ne razlikuju ispitanike sa različitim nivoima osobine i ne doprinose merenju. Stavke sa nagibima većim od 2,5 su obično diskriminativne u vrlo uskom opsegu merene osobine. Karakteristična kriva ovakve stavke je strma, skoro vertikalna i od verovatnoće 0 do 1 stigne u vrlo uskom rasponu osobine.

Inače, neobično visoki parametri (ne samo diskriminativnosti) koji višestruko odudaraju od parametara ostalih stavki mogu biti rezultat narušavanja nekih pretpostavki modela (odeljak 2.2 Slika 3) ili izostanka konvergencije modela (Mair, 2018). Visoki parametri diskriminativnosti uticaće i na visoku procenu informativnosti stavke u višeparametarskim modelima. Takve stavke mogu izgledati kao najbolje, ali treba biti obazriv jer ako su visoki parametri diskriminativnosti rezultat navedenih anomalija, naši zaključci biće pogrešni.

```
# Spajamo parametre Rašovog i 2PL modela u jednu tabelu
por.par <- data.frame(cbind(coef(mr.l), coef(m2p.lt)[,2:3]))
por.par
```

```
##           Dffclt Dscrmn    Dffclt.1 Dscrmn.1
## it1  -0.11123755      1 -0.10382191 1.095200
## it2   0.06063400      1  0.04013580 1.781958
## it3  -0.09098949      1 -0.07259405 1.417127
## it4  -0.26371818      1 -0.20837164 1.384487
## it5  -1.73103257      1 -1.25695673 1.574540
## it6  -1.61341573      1 -1.19669836 1.514250
## it7   2.32609594      1  1.75602355 1.458674
## it8  -1.11215465      1 -0.89962189 1.304368
```

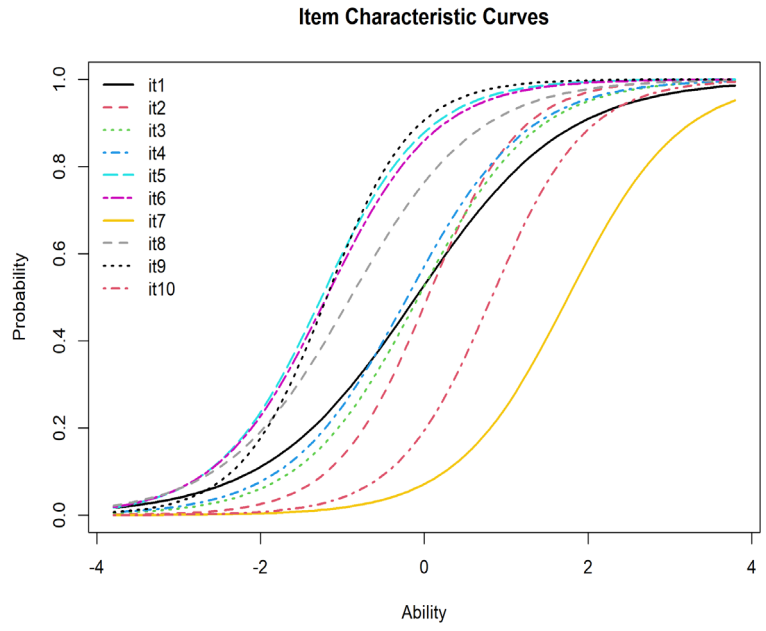
```
## it9  -1.79999042      1 -1.19463714 1.900745
## it10  1.18353587      1  0.82151587 1.729871

cat("Pirsonove korelacije parametara težine iz Rašovog i 2PL modela\n")

## Pirsonove korelacije parametara težine iz Rašovog i 2PL modela

cor(por.par[,c(1,3)], method = "pearson")

##           Dffclt Dffclt.1
## Dffclt   1.00000  0.99834
## Dffclt.1 0.99834  1.00000

cat("Spirmanove korelacije parametara težine iz Rašovog i 2PL modela\n")

## Spirmanove korelacije parametara težine iz Rašovog i 2PL modela

cor(por.par[,c(1,3)], method = "spearman")

##             Dffclt  Dffclt.1
## Dffclt   1.0000000 0.9636364
## Dffclt.1 0.9636364 1.0000000
```

Korelacije parametara težine dobijene pomoću dva modela su vrlo visoke, ali nisu jedinične.

#### 10.4.4.1.1 Karakteristične krive stavki

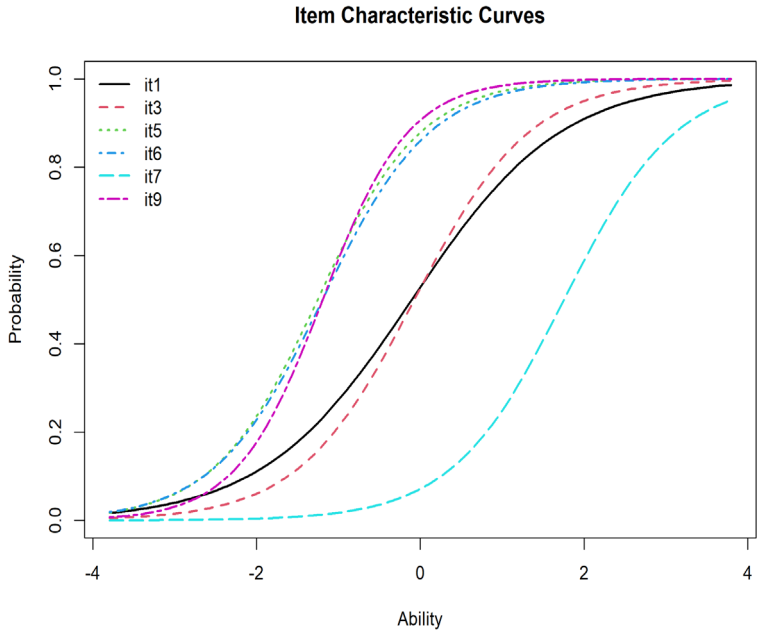
U dvoparametarskom logističkom modelu grafikoni karakterističnih kriva stavki nisu nezanimljivi kao u jednoparametarskom modelu. Ovde se krive pored lokacije, mogu razlikovati i po nagibu.

```
# Za funkciju plot podrazumevani tip su KKS, ali ako je model grm() onda KKK
# Argument type="ICC" određuje da bude predstavljena KKS, ali se može izostaviti
(osim u grm())
# Ako nema parametra items= biće prikazane krive za sve stavke
plot(m2p.lt, type="ICC", lty=c(1:10), lwd=2, legend=T)
```




Slika 42 – KKS u 2PL modelu

```
# Ako parametru items= prosledimo vektor sa rednim brojevima stavki biće prikazane
samo one koje smo naveli
plot(m2p.lt, items=c(1, 3,5:7,9), type="ICC", lty=c(1:10), lwd=2, legend=T)
```




Slika 43 – KKS odabranih stavki u 2PL modelu

Na prvom grafikonu (Slika 42) vidimo sa se kriva stavke 1, koja ima najniži parametar diskriminativnosti, ukršta sa krivama nekih drugih stavki. Da li će doći do ukrštanja zavisi od razlike u nagibima i udaljenosti lokacija stavki. Ukrštanje će se pre dogoditi što su razlike u nagibima veće i što su lokacije stavki bliže, odnosno, što su stavke sličnije težine.

#### 10.4.4.1.2 Pokazatelji fita

Paket *ltm* nema omnibus pokazatelj fita za višeparametarske modele. Međutim, pošto koristi isti metod procene (MMLE) kao paket *mirt*, pokazatelj fita možemo izračunati u tom paketu. Možemo i celu analizu obaviti u paketu *mirt*. Međutim, paket *mirt* nema pregledne grafikone (odeljak 10.4.4.1.3.1, Slika 46 i Slika 45), a to paket *ltm* ima. Ova dva paketa se na taj način dopunjuju.

```
p_load(mirt)
# Procenićemo model u paketu mirt
m2p.m <- mirt(b.pod, 1, SE=TRUE, verbose = FALSE)

# Uporedni prikaz parametara iz paketa mirt (levo) i ltm (desno). Na trećoj
decimali skoro da i nema razlike
round(cbind(MDIFF(m2p.m), MDISC(m2p.m), coef(m2p.lt)[,2:3]), 3)

##      MDIFF_1       Dffclt Dscrmn
## it1   -0.104 1.095 -0.104  1.095
## it2    0.040 1.782  0.040  1.782
## it3   -0.073 1.417 -0.073  1.417
## it4   -0.208 1.384 -0.208  1.384
## it5   -1.257 1.574 -1.257  1.575
## it6   -1.197 1.514 -1.197  1.514
```

```
## it7    1.756 1.458  1.756  1.459
## it8   -0.900 1.304 -0.900  1.304
## it9   -1.195 1.900 -1.195  1.901
## it10   0.822 1.730  0.822  1.730

# Izračunaćemo omnibus pokazatelje fita u paketu mirt
M2(m2p.m)

##             M2 df         p       RMSEA RMSEA_5   RMSEA_95      SRMSR       TLI
## stats 38.18272 35 0.3268395 0.009540746       0 0.02530255 0.02073589 0.9984261
##             CFI
## stats 0.9987759

# Uklonićemo paket mirt iz memorije pošto nam trenutno više ne treba
detach("package:mirt", unload = TRUE)
```

$M_2$ statistik (bliskog) fita nije značajan, pa možemo reći da je model u celini saglasan sa podacima. I ostali pokazatelji (približnog) fita su odlični.

To što je model u celini saglasan, ne znači da nije potrebno pogledati pokazatelje fita za stavke.

#### *10.4.4.1.2.1 Pokazatelji fita za stavke*

Pokazatelji saglasnosti za stavke su isti kao za Rašov model (u paketu *ltm*), pa ih nećemo detaljnije opisivati. Trebalo bi napomenuti da funkcija *item.fit()* ne može da se koristi sa modelom procenjenim funkcijom *grm()*, već samo *rasch(), tpm()* i *ltm().*

```
item.fit(m2p.lt, simulate.p.value=T, B=1000)

##
## Item-Fit Statistics and P-values
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Alternative: Items do not fit the model
## Ability Categories: 10
## Monte Carlo samples: 1000
##
##           X^2 Pr(>X^2)
## it1  107.0995     0.02
## it2  102.2674   0.1479
## it3   52.6213   0.8452
## it4   50.6463   0.8492
## it5   36.3126   0.8561
## it6   34.0779   0.9421
## it7  112.2195    0.048
## it8   38.7962   0.7912
## it9   44.5941   0.8881
## it10  95.1305   0.1978
```

Prema prikazanim pokazateljima fita, dve stavke imaju značajan misfit, a to su stavke 1 i 7. (setite se da je tako bilo i u Rašovom modelu) pri čemu je pokazatelj fita za stavku 7 blizu granice značajnosti.

Stavka 1 ima najnižu diskriminativnost i mogli bismo da posumnjamo da ne meri isto što i ostale stavke, bar ne u potpunosti. Ova stavka ima i najniži indeks skalabilnosti $H_j$ (videti odeljak 10.3.2) i svakako

je kandidat za izbacivanje.

S druge strane, stavka 7 ima umerenu diskriminativnost (gledano u kontekstu ostalih stavki), ali je ubedljivo najteža.

Osim ovih, u paketu *ltm* postoje pokazatelji fita stavki zasnovani na hi–kvadratima reziduala margina drugog i trećeg reda (Rizopoulos, 2006).

```
margins(m2p.lt, type="two-way")
```

```
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Fit on the Two-Way Margins
##
## Response: (0,0)
##   Item i Item j Obs    Exp (O-E)^2/E
## 1      5      8  94 100.87      0.47
## 2      2      6 173 164.28      0.46
## 3      8      9 108 101.79      0.38
##
## Response: (1,0)
##   Item i Item j Obs   Exp (O-E)^2/E
## 1      2      6  41 49.56      1.48
## 2      3      9  52 46.48      0.66
## 3      7      9   4  5.60      0.46
##
## Response: (0,1)
##   Item i Item j Obs   Exp (O-E)^2/E
## 1      5      7  14  7.22      6.36 ***
## 2      2     10  82 71.98      1.39
## 3      2      7  23 28.90      1.20
##
## Response: (1,1)
##   Item i Item j Obs    Exp (O-E)^2/E
## 1      2     10 198 207.64      0.45
## 2      5      7 115 121.86      0.39
## 3      2      7 106 100.19      0.34
##
## '***' denotes a chi-squared residual greater than 3.5
```

Ova procedura umesto da računa hi–kvadrat na kompletnim sklopovima odgovora to čini na parovima ili tripletima stavki. Kao značajne vrednosti hi–kvadrata zvezdicom su označene one preko 3,5, ali se to može promeniti dodavanjem argumenta *rule=3* (upisuje se broj nove željene granične vrednosti hi–kvadrata). U prikazanim tabelama zvezdicom su označeni parovi ili tripleti stavki kod kojih postoje neočekivane kombinacije odgovora (koje nisu u skladu sa modelom). U našem primeru za parove stavki značajan je samo hi–kvadrat za par 5–7. Sklop odgovora 0–1 (0 na lakšu stavku 5 i 1 na težu stavku 7) češći je nego što model predviđa.

```
margins(m2p.lt, type="three-way")
```

```
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
```

```
## Fit on the Three-Way Margins
##
## Response: (0,0,0)
##   Item i Item j Item k Obs   Exp (O-E)^2/E
## 1      3      5      6  81 69.67      1.84
## 2      1      6      8  70 80.11      1.28
## 3      5      7      8  89 98.87      0.98
##
## Response: (1,0,0)
##   Item i Item j Item k Obs   Exp (O-E)^2/E
## 1      2      3      6  14 26.17      5.66 ***
## 2      3      5      6   7 14.49      3.87 ***
## 3      2      6      7  34 45.36      2.84
##
## Response: (0,1,0)
##   Item i Item j Item k Obs   Exp (O-E)^2/E
## 1      5      7     10  11  5.34      6.00 ***
## 2      5      7      8   5  2.01      4.47 ***
## 3      1      2      6  18 25.55      2.23
##
## Response: (1,1,0)
##   Item i Item j Item k Obs   Exp (O-E)^2/E
## 1      1      3      6  36 27.46      2.65
## 2      1      4      9  28 20.92      2.39
## 3      3      6      9  41 32.59      2.17
##
## Response: (0,0,1)
##   Item i Item j Item k Obs  Exp (O-E)^2/E
## 1      3      5      7   8 3.47      5.92 ***
## 2      1      5      7   7 3.44      3.69 ***
## 3      6      9     10   0 3.08      3.08
##
## Response: (1,0,1)
##   Item i Item j Item k Obs   Exp (O-E)^2/E
## 1      2      5      7   9  3.52      8.55 ***
## 2      1      2     10  55 41.45      4.43 ***
## 3      4      5      7   8  4.02      3.94 ***
##
## Response: (0,1,1)
##   Item i Item j Item k Obs  Exp (O-E)^2/E
## 1      5      6      7  13 5.81      8.88 ***
## 2      5      7      9  13 6.03      8.04 ***
## 3      5      7      8   9 5.22      2.74
##
## Response: (1,1,1)
##   Item i Item j Item k Obs    Exp (O-E)^2/E
## 1      5      6      7 105 114.92      0.86
## 2      2      4     10 159 169.68      0.67
## 3      1      6      7  85  90.95      0.39
##
## '***' denotes a chi-squared residual greater than 3.5
```

Kada su u pitanju tripleti stavki, hi–kvadrat je značajan za sledeće: 2–3–6, 3–5–6 (sklop 1–0–0), 5–7–10, 5–7–8 (sklop 0–1–0) i 3–5–7 (sklop 0–0–1). U najvećem broju misfitujućih tripleta učestvuje jedini par sa značajnim misfitom 5–7. Obratićemo više pažnje na odnos te dve stavke, ali iz sklopova je prilično jasno da je u pitanju neočekivano veća proporcija tačnih odgovora na najtežu stavku 7, u odnosu na neke

lakše. U preostala dva misfitujuća tripleta javljaju se stavke 3 i 6. Stavka 6 je jedna od najlakših i u ovim misfitujućim tripletima sklop 1–0–0 je ređe zastupljen nego što model predviđa. Stavka 5 se javlja u najvećem broju tripleta i radi se o najlakšoj stavci.

Iako učestalosti pojedinih kombinacija odgovora na različite stavke odstupaju od modelski predviđenih, ta odstupanja nisu prevelika i ne narušavaju fit većine stavki, niti modela u celini.

#### *10.4.4.1.2.2 Pokazatelji fita za ispitanike*

Pokazatelji saglasnosti za ispitanike su isti kao za Rašov model (u paketu *ltm* – videti odeljak 10.4.2.1.2*).* Dobijaju se funkcijom *person.fit()*. Funkcija *person.fit()* ne može da se koristi sa modelima procenjenim funkcijom *grm()*, već samo sa onima procenjenim funkcijama *rasch(), tpm()* i *ltm().* Rezultat funkcije je objekat koji sadrži pokazatelje fita za *sklopove odgovora*, ne za ispitanike. Da biste utvrdili o kojim ispitanicima se radi morate spojiti pokazatelje misfita sa matricom podataka.

```
m2p.ltPF <- person.fit(m2p.lt, simulate.p.value=T, B=1000,
 alternative="two.sided")
# Pozivanjem naziva objekta u koji smo smestili pokazatelja fita dobijemo ispis
sklopova pokazatelja fita i njihove statističke značajnosti.
#m2p.ltPF
# Na ovom mestu nećemo to uraditi, ali ćemo podatke o sklopovima odgovora,
pokazateljima i njihovoj statističkoj značajnosti iz objekta izdvojiti u
data.frame
m2p.ltMF <- data.frame(m2p.ltPF$resp.patterns, m2p.ltPF$Tobs, m2p.ltPF$p.values)
# Zatim ćemo ih spojiti sa podacima, kao što smo to radili za mere u Rašovom
modelu
# Prvo pravimo kopiju podataka kako bismo sačuvali originalne
b.podK <- b.pod
# Zatim ih spajamo sa pokazateljima misfita
b.podK <- merge(b.podK, m2p.ltMF, all.x = TRUE, by=names(b.podK))
# Na ovaj način možemo iz matrice podataka odabrati samo ispitanike koji imaju
značajan misfit ili one koji nemaju (ako tako želimo)
b.podK[b.podK$Lz.1<.05,]
```

```
##     it1 it2 it3 it4 it5 it6 it7 it8 it9 it10         L0        Lz        L0.1
## 342   0   1   0   1   0   1   1   0   1    0  -8.386218 -1.990733 0.030969031
## 343   0   1   0   1   0   1   1   0   1    0  -8.386218 -1.990733 0.030969031
## 415   0   1   1   1   1   0   1   0   1    1  -9.106329 -2.800911 0.017982018
## 546   1   0   0   1   0   1   1   1   1    1  -8.312508 -2.185258 0.030969031
## 586   1   0   1   0   0   0   0   1   1    1  -8.273383 -1.869302 0.044955045
## 636   1   0   1   1   1   0   1   0   1    0  -8.084065 -1.847826 0.045954046
## 722   1   1   0   1   1   1   0   0   0    1  -8.187593 -1.945159 0.041958042
## 785   1   1   1   0   1   0   1   0   1    0  -8.440201 -2.132222 0.037962038
## 821   1   1   1   1   0   1   1   0   0    0 -10.309346 -3.380828 0.003996004
##            Lz.1
## 342 0.031968032
## 343 0.031968032
## 415 0.015984016
## 546 0.030969031
## 586 0.047952048
## 636 0.045954046
## 722 0.040959041
## 785 0.038961039
## 821 0.003996004
```

```
# Ovo je varijanta za prikaz onih koji nemaju misfit
#b.podK[!b.podK$Lz.1<.05,]

# Moguće su varijante pokazatelja misfita sa alternative="less" koja služi za
detekciju šuma...
person.fit(m2p.lt, simulate.p.value=T, B=1000, alternative="less")

##
## Person-Fit Statistics and P-values
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Alternative: Inconsistent response pattern under the estimated model
## Monte Carlo samples: 1000
##
##     it1 it2 it3 it4 it5 it6 it7 it8 it9 it10       L0 Pr(<L0)      Lz Pr(<Lz)
## 1     0   0   0   0   0   0   0   0   0    0  -2.4715  0.9141  1.5763   0.985
## 2     0   0   0   0   0   0   0   0   1    0  -4.0212  0.8841  1.2936  0.9241
## 3     0   0   0   0   0   0   0   1   0    0  -3.8846  0.8671  1.1944  0.9041
## 4     0   0   0   0   0   0   0   1   1    0  -4.6992  0.7642  1.0597  0.8282
## 5     0   0   0   0   0   1   0   0   0    0  -3.7258  0.8771  1.3795  0.9341
## 6     0   0   0   0   0   1   0   0   0    1  -6.1975  0.3846 -0.1578  0.4116
## 7     0   0   0   0   0   1   0   0   1    0  -4.4039  0.8521  1.3285  0.9221
## 8     0   0   0   0   0   1   0   0   1    1  -6.4925  0.2877 -0.3627  0.3097
## 9     0   0   0   0   0   1   0   1   0    0  -4.5476  0.8182  1.1189  0.8671
## 10    0   0   0   0   0   1   0   1   1    0  -4.5997  0.8282  1.2118  0.9071
...

# i sa alternative="greater" koja služi za detekciju prigušenog šuma (overfita)
person.fit(m2p.lt, simulate.p.value=T, B=1000, alternative="greater")

##
## Person-Fit Statistics and P-values
##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Alternative: More consistent response pattern than the model predicts
## Monte Carlo samples: 1000
##
##     it1 it2 it3 it4 it5 it6 it7 it8 it9 it10       L0 Pr(>L0)      Lz Pr(>Lz)
## 1     0   0   0   0   0   0   0   0   0    0  -2.4715  0.0729  1.5763   0.017
## 2     0   0   0   0   0   0   0   0   1    0  -4.0212  0.1389  1.2936  0.0769
## 3     0   0   0   0   0   0   0   1   0    0  -3.8846  0.1608  1.1944  0.1439
## 4     0   0   0   0   0   0   0   1   1    0  -4.6992  0.2258  1.0597  0.1588
## 5     0   0   0   0   0   1   0   0   0    0  -3.7258  0.1309  1.3795  0.0659
## 6     0   0   0   0   0   1   0   0   0    1  -6.1975  0.6054 -0.1578  0.5734
## 7     0   0   0   0   0   1   0   0   1    0  -4.4039  0.1588  1.3285  0.0739
## 8     0   0   0   0   0   1   0   0   1    1  -6.4925  0.7093 -0.3627  0.6773
## 9     0   0   0   0   0   1   0   1   0    0  -4.5476  0.2068  1.1189  0.1538
## 10    0   0   0   0   0   1   0   1   1    0  -4.5997  0.1798  1.2118  0.1119
...
```

Nema puno misfitujućih ispitanika (ima ih 8). U slučaju da ste merenje obavljali sa nekom konkretnom svrhom (npr. selekcija) odgovore ovih ispitanika bi trebalo posebno tumačiti i pogledati, kako biste videli šta je tačno neuobičajeno u njihovom odgovaranju.

Sa argumentom *alternative* i opcijama *"less"* ili "*greater"* mogli biste proveriti radi li se o neočekivano predvidljivom ili nepredvidljivom odgovaranju.

#### 10.4.4.1.3 Informativnost testa

Prilikom procene informativnosti testa i stavki mora biti definisan opseg osobine koji nas zanima jer od njega i zavisi informativnost. S druge strane, opseg koji nas zanima zavisi od ciljne populacije ili od lokacije na kontinuumu osobine na kojem su nam bitne precizne odluke.

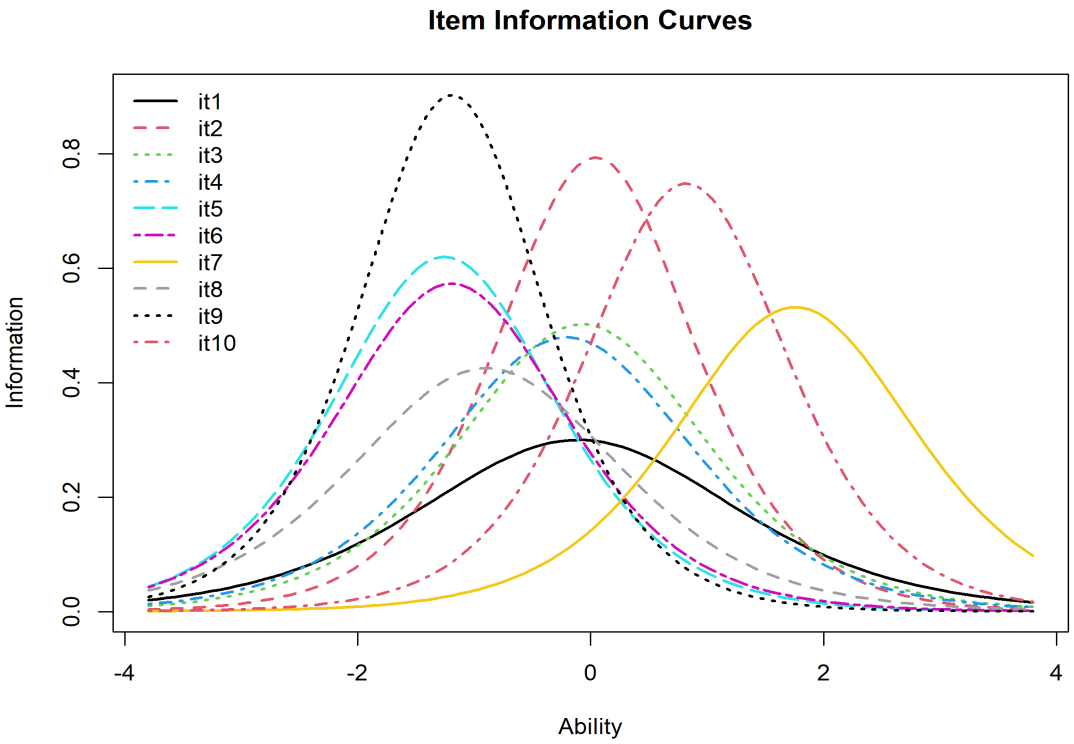
```
# Informativnost testa

information(m2p.lt, range=c(-5, 5))

##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Total Information = 15.16
## Information in (-5, 5) = 15.11 (99.69%)
## Based on all the items

# Informativnost stavke (items=)
information(m2p.lt, range=c(-3,3), items=1)

##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Total Information = 1.1
## Information in (-3, 3) = 1.02 (92.75%)
## Based on items 1

# Informativnost stavki (može biti definisano više stavki)
information(m2p.lt, c(-3,3), items=c(1, 3:5, 7))

##
## Call:
## tpm(data = b.pod, max.guessing = 0)
##
## Total Information = 6.93
## Information in (-3, 3) = 6.46 (93.26%)
## Based on items 1, 3, 4, 5, 7
```

Ovde stavke ne moraju imati istu maksimalnu informativnost. Informativnost u 2PL modelu jednaka je informativnosti iz 1PL modela ponderisanoj kvadratom parametra diskriminativnosti.

##### *10.4.4.1.3.1 Grafički prikaz informativnosti stavki*

```
# Za funkciju plot podrazumevani tip su KKS, ali ako je model grm() onda KKK
# Argument type="IIC" određuje da bude predstavljena kriva informativnosti

# Ako nema argumenta items= biće prikazane krive za sve stavke
plot(m2p.lt, type="IIC", lty=c(1:10), lwd=2, legend=T)
```

```
# Ako parametru items= prosledimo vektor sa rednim brojevima stavki biće prikazane
samo one koje smo naveli (da biste izvršili komandu prethodno uklonite #)
#plot(m2p.lt, items=c(2,7:10), type="IIC", lty=c(1:10), lwd=2, legend=T)
```



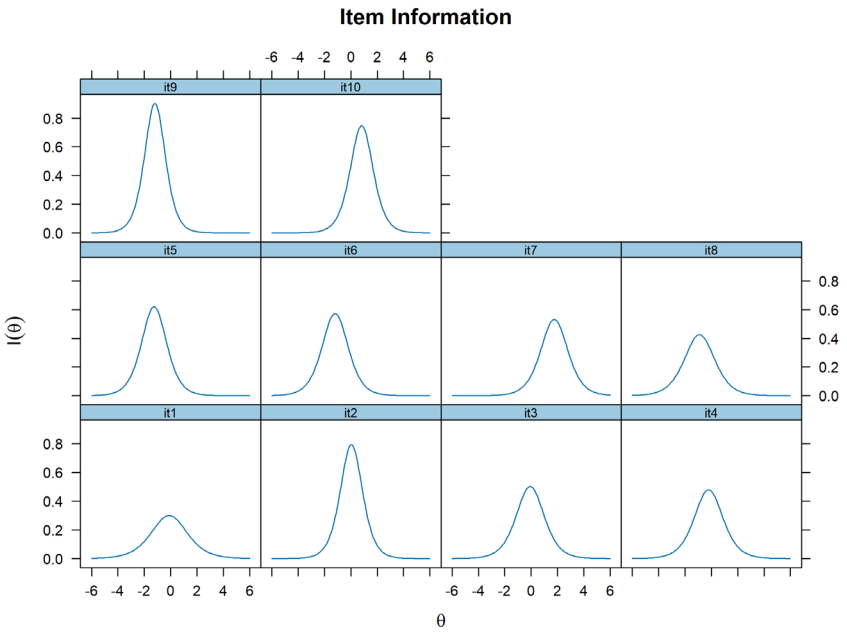

Slika 44 – Funkcije informativnosti stavki u 2PL modelu (paket *ltm*)

```
cat("Maksimalna informativnost najinformativnije stavke 9 je:",
 round(coef(m2p.lt)[9,3]^2*.25,2), "na lokaciji:",
 round(coef(m2p.lt)[9,2],2), "logita")

## Maksimalna informativnost najinformativnije stavke 9 je: 0.9 na lokaciji: -1.19
logita

cat("Maksimalna informativnost najmanje informativne stavke 1 je:",
 round(coef(m2p.lt)[1,3]^2*.25,2), "na lokaciji:", round(coef(m2p.lt)[1,2],2),
 "logita")

## Maksimalna informativnost najmanje informativne stavke 1 je: 0.3 na lokaciji:
 -0.1 logita
```

Najinformativnija stavka je stavka 9 (Slika 44). Maksimalna informativnost stavke 9 je 0,89 na njenoj lokaciji, a to je $b_9$=–1,2 logita. Najlošija, odnosno, najmanje informativna stavka je stavka 1. Podsetićemo, to je ujedno i jedina stavka koja je imala značajan misfit. Njena maksimalna informativnost je 0,30 na lokaciji $b_1$=–0,1 logita. Ovo smo izračunali tako što smo maksimalnu informativnost u Rašovom modelu za binarne stavke (0,25) pomnožili kvadratom parametra diskriminativnosti stavke (u ovom slučaju stavke 9, a posle i stavke 1).

Radi poređenja prikazaćemo isti ovaj grafik iz paketa *mirt* (Slika 45).

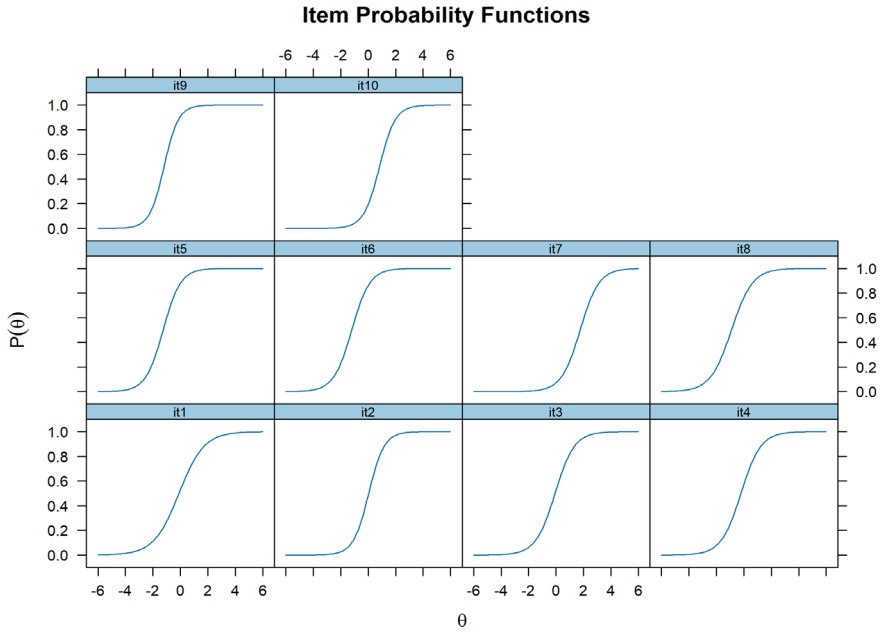
```
mirt::plot(m2p.m, type = 'infotrace', auto.key = FALSE)
```




Slika 45 – Funkcije informativnosti stavki u 2PL modelu (paket *mirt*)

Ono što se autoru ove knjige učinilo nepreglednim je što su sve krive prikazane zasebno. Stavke donekle možemo vizuelno uporediti po maksimalnoj informativnosti, ali ne i po lokaciji na kojoj je postižu. Čini se preglednijim grafikon iz paketa *ltm* gde su sve stavke predstavljene na istoj osi. Slično je i sa grafikonom KKS (Slika 46).

```
mirt::plot(m2p.m, type = 'trace')
```



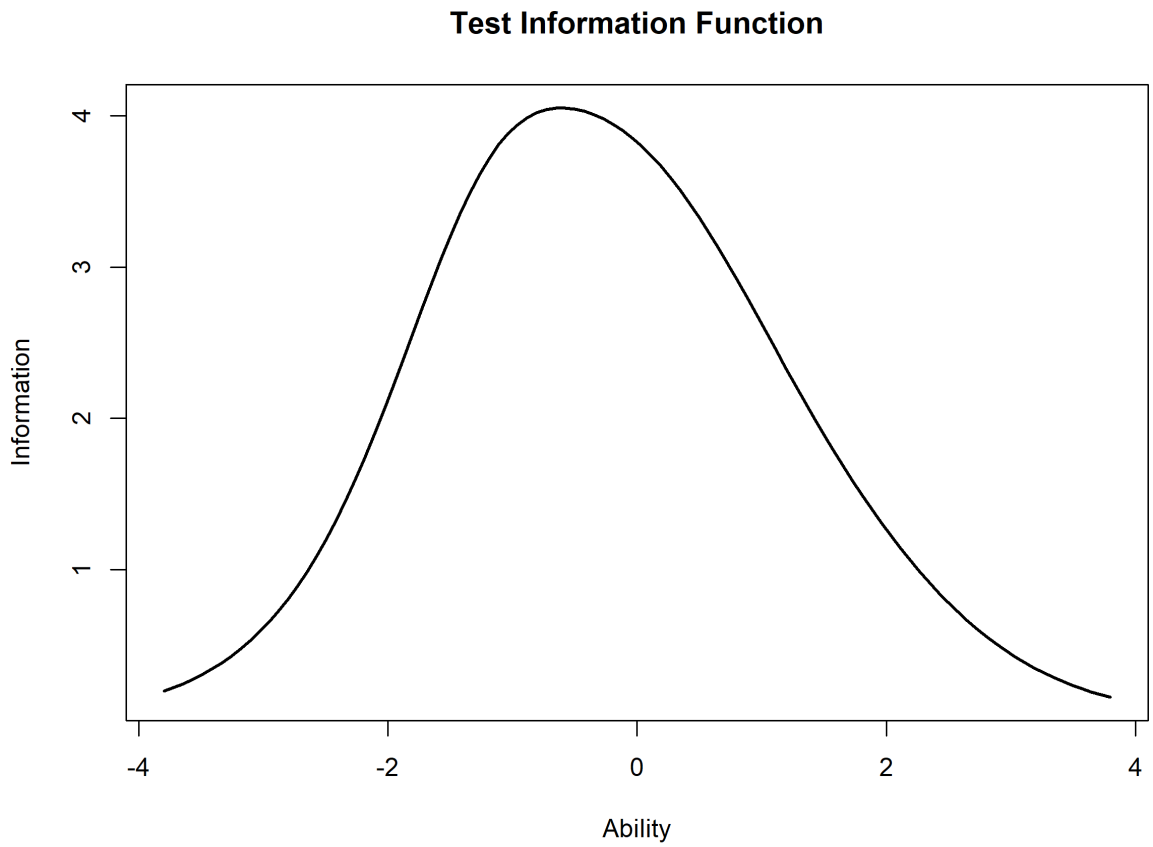

Slika 46 – KKS stavki u 2PL modelu (paket *mirt*)

Manje razlike u nagibima nisu uočljive, a pošto su stavke prikazane na zasebnim grafikonima nisu uočljive ni razlike u lokacijama.

#### *10.4.4.1.3.2 Grafički prikaz informativnosti testa*

Informativnost testa je i ovde jednaka sumi informativnosti stavki. Test će biti informativniji što ima više visoko informativnih stavki. Biće najinformativniji na delu kontinuuma latentne crte gde ima najviše (visoko informativnih) stavki. Kada biramo stavke za test, pored pokazatelja fita vodimo računa i o informativnosti stavki i lokaciji gde je neka stavka najinformativnija. Drugim rečima, potrebne su nam visoko diskriminativne stavke, odgovarajućih težina. Ako želimo da merenje bude preciznije u oblasti visokog nivoa osobine, biraćemo teže i visoko diskriminativne stavke. Ako želimo da merenje obavljamo u oblasti nižeg nivoa osobine, biraćemo lakše i visoko diskriminativne stavke. Za opštu populaciju gledaćemo da ceo raspon osobine u ciljnoj populaciji pokrijemo diskriminativnim stavkama.

```
# Argument items=0 određuje da bude prikazana informativnost testa
plot(m2p.lt, items=0, type="IIC", lwd=2)
```




Slika 47 – Funkcija informativnosti testa u 2PL modelu

Dok su krive informativnosti binarih stavki u 1PL i 2PL modelima simetrične, krive informativnosti testa ne moraju biti. Ove krive ne moraju čak biti ni unimodalne. Bimodalna kriva bi se mogla javiti kada bismo imali dve grupe visoko informativnih stavki – grupu lakih i grupu teških. Gore prikazana kriva je asimetrična, a informativnost je veća u delu kontinuuma u kojem postoji veći broj informativnijih stavki (nešto ispod 0 logita).

#### 10.4.4.1.4 Standardne greške i pouzdanost

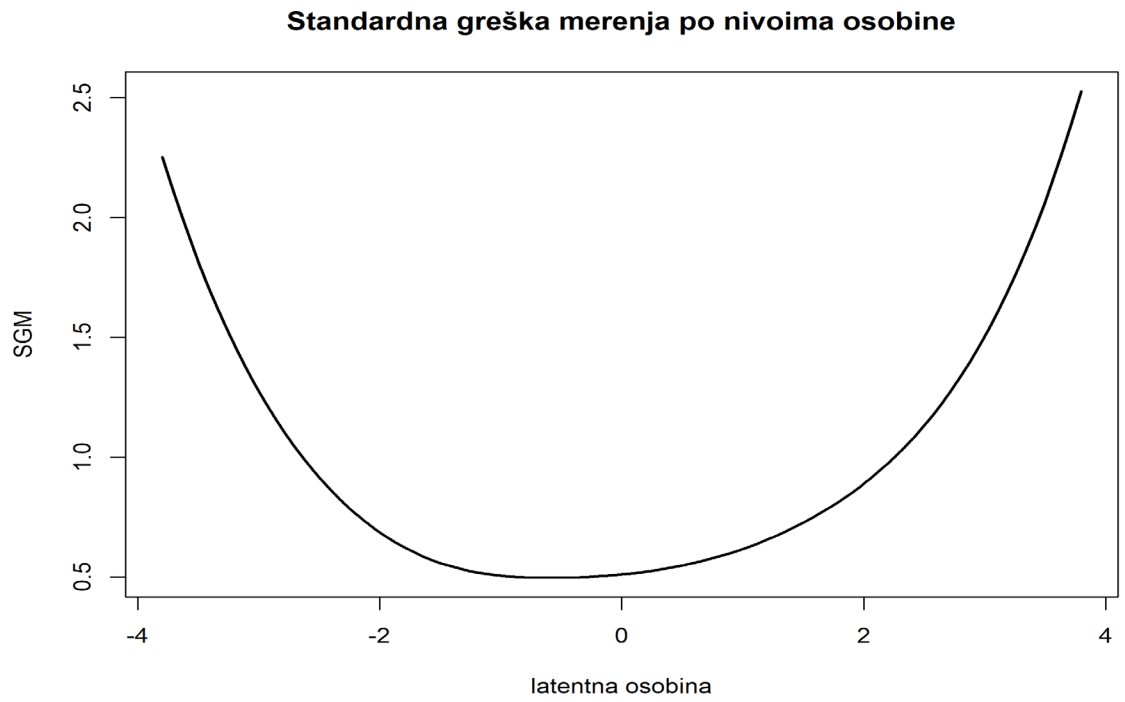
Podatke koji se koriste za crtanje grafikona informativnosti testa u paketu *ltm* možemo iskoristiti za procenu standardnih grešaka i pouzdanosti merenja u zavisnosti od nivoa osobine. Sačuvaćemo podatke u objekat *Info*. U pitanju je matrica sa kolonom *z* koja sadrži nivoe osobine i kolonom *info* koja sadrži informativnost testa na odgovarajućem nivou osobine

```
Info <- data.frame(plot(m2p.lt, type="IIC", items=0, plot=FALSE))

# Možemo naći lokaciju gde je informativnost testa maksimalna
MI <- max(Info$info)
Info[Info$info==MI,]

##             z     info
## 43 -0.5757576 4.052333

cat("Maksimalna informativnost testa iznosi:", round(MI,2), "na nivou osobine od",
 round(Info[Info$info==MI,1], 2), "logita")

## Maksimalna informativnost testa iznosi: 4.05 na nivou osobine od -0.58 logita
```

Znajući da je standardna greška recipročna korenu informativnosti ($SE(\theta) = 1/\sqrt{I(\theta)}$) možemo je izračunati (`1/sqrt(info)`) i prikazati na grafikonu (Slika 48).

```
with(Info, plot(z, 1/sqrt(info), type = "l", lwd = 2, xlab = "latentna osobina",
 ylab = "SGM", main = "Standardna greška merenja po nivoima osobine"))
```

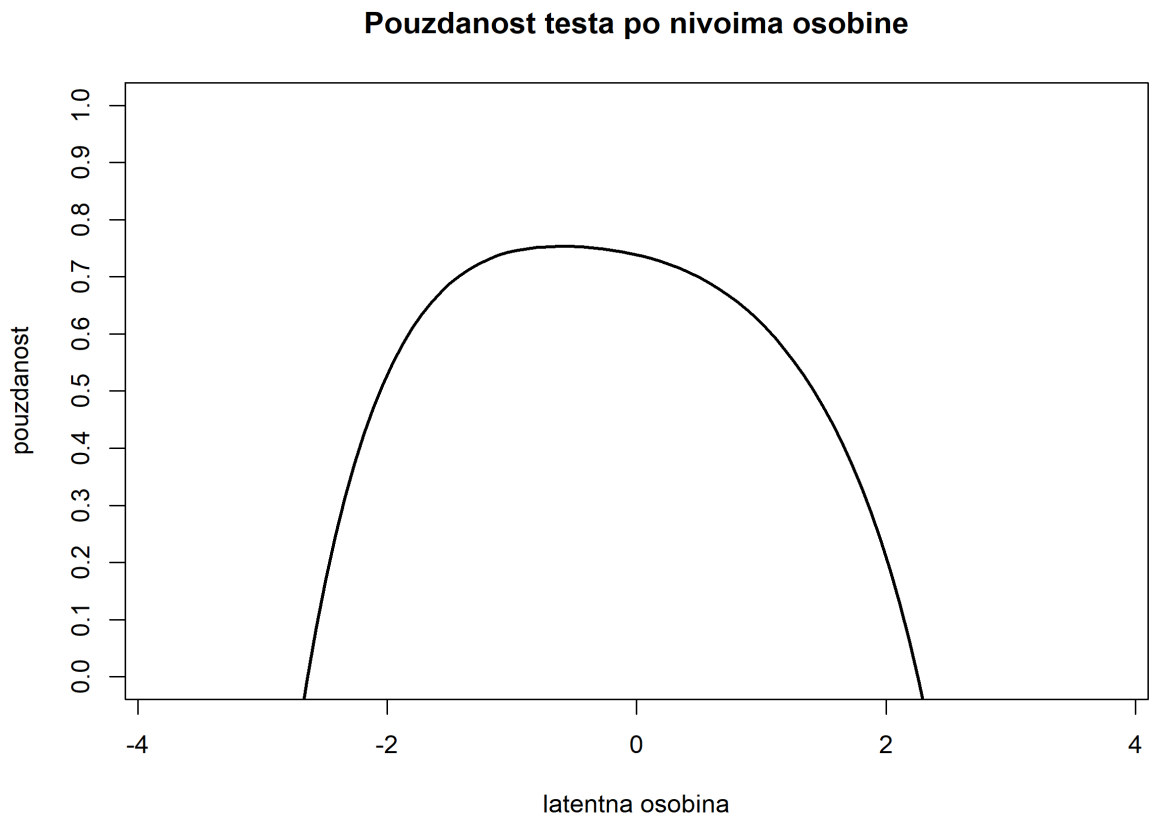
Slično, pošto je pouzdanost jednaka $1–SE^2$, onda se preko informativnosti ona može izračunati kao




Slika 48 – Standardna greška merenja po nivoima latentne osobine

$1–1/I(\theta)$ i to predstaviti grafički (Slika 49).

```
with(Info, plot(z, 1-(1/info), type = "l", lwd = 2, xlab = "latentna osobina",
 ylab = "pouzdanost", main = "Pouzdanost testa po nivoima osobine",
 yaxp=c(0, 1, 10), ylim=c(0,1)))

cronbach.alpha(b.pod)
```




Slika 49 – Pouzdanost testa za različite nivoe osobine (izračunata preko informativnosti)

```
##
## Cronbach's alpha for the 'b.pod' data-set
##
## Items: 10
## Sample units: 1000
## alpha: 0.757
```

10.4.4.1.5 Mere ispitanika

Kao i kod Rašovog modela, upisivanje mera ispitanika u matricu podataka u paketu *ltm* zahteva malo truda.

```
# Pošto će nam izvorni podaci i dalje biti potrebni, napravićemo kopiju pod drugim
imenom
b.podK <- b.pod
# Mere ispitanika ćemo upisati u novi data.frame
MereL <- factor.scores(m2p.lt)$score.dat
# Objekat funkcije je model koji smo ranije procenili pomoću paketa ltm
# Objekat MereL ne sadrži mere za sve ispitanike već samo za sklopove koji
postoje, a ovde ih ima
nrow(MereL)
```

```
## [1] 263

# Za spajanje ćemo koristiti funkciju merge() kojoj je potrebno navesti imena
matrica podataka koje želimo da spojimo
b.podM <- merge(b.podK, MereL, all.x = TRUE, by=names(b.podK))
# Argument all.x=TRUE znači da želimo da zadržimo sve redove iz prve matrice
(b.podK), a argument by=names(b.podK) znači da bi dve matrice trebalo da budu
uparene po varijablama sadržanim u vektoru names(b.podK), što su nazivi varijabli
u matrici b.podK. Ove varijable moraju postojati u obe matrice da bi mogle da se
koriste za uparivanje. Na ovaj način mere za određeni sklop odgovora na stavke iz
matrice MereL biće pripisane svim ispitanicima koji imaju isti takav sklop u
matrici b.podK i sve to će biti upisano u novi objekat b.podM
head(b.podM)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10 Obs    Exp        z1     se.z1
## 1   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293
## 2   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293
## 3   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293
## 4   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293
## 5   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293
## 6   0   0   0   0   0   0   0   0   0    0  19 22.232 -1.903569 0.5474293

# Pored mera u matricu će biti upisane opažene i očekivane frekvencije svakog
sklopa, mera ispitanika kao varijabla z1 i standardna greška mere kao varijabla
se.z1
```

Podsećamo, mere imaju smisla ako je model saglasan sa podacima.

#### 10.4.4.1.6 Deskriptivni pokazatelji

Paket *ltm* ima korisnu funkciju koja prikazuje različite deskriptivne pokazatelje koji nam mogu pomoći prilikom dijagnostike eventualnih problema. Ova funkcija kao objekat uzima matricu podataka i nije vezana za konkretan model.

```
ltm::descript(b.pod)

##
## Descriptive statistics for the 'b.pod' data-set
##
## Sample:
##  10 items and 1000 sample units; 0 missing values
##
## Proportions for each level of response:
##          0     1   logit
## it1  0.477 0.523  0.0921
## it2  0.511 0.489 -0.0440
## it3  0.481 0.519  0.0760
## it4  0.447 0.553  0.2128
## it5  0.198 0.802  1.3988
## it6  0.214 0.786  1.3010
## it7  0.871 0.129 -1.9098
## it8  0.291 0.709  0.8905
## it9  0.189 0.811  1.4565
## it10 0.720 0.280 -0.9445
##
## Frequencies of total scores:
##       0  1  2  3   4   5   6   7   8  9 10
## Freq 19 55 55 89 121 116 134 153 129 92 37
```

```
##
##
## Point Biserial correlation with Total Score:
##      Included Excluded
## it1    0.5507   0.3878
## it2    0.6432   0.5013
## it3    0.6019   0.4500
## it4    0.5989   0.4472
## it5    0.5372   0.4090
## it6    0.5420   0.4104
## it7    0.4211   0.3019
## it8    0.5529   0.4073
## it9    0.5667   0.4459
## it10   0.5752   0.4357
##
##
## Cronbach's alpha:
##                  value
## All Items       0.7573
## Excluding it1   0.7427
## Excluding it2   0.7247
## Excluding it3   0.7329
## Excluding it4   0.7334
## Excluding it5   0.7390
## Excluding it6   0.7387
## Excluding it7   0.7515
## Excluding it8   0.7390
## Excluding it9   0.7346
## Excluding it10 0.7350
##
##
## Pairwise Associations:
##    Item i Item j p.value
## 1       5      7   0.009
## 2       6      7   2e-04
## 3       7      8   5e-06
## 4       1      7   2e-06
## 5       7      9   2e-06
## 6       4      7   3e-09
## 7       1      6   1e-09
## 8       3      7   8e-10
## 9       5      8   4e-10
## 10      1      5   2e-10
```

Ovom komandom dobijamo Kronbahovu alfu za ceo test ali i kako bi se kretala vrednost alfe kada bismo uklonili pojedine stavke (pretposlednja tabela u gornjem ispisu).

Takođe, dobijamo i tabelu sa najnižim korelacijama između parova stavki (poslednja tabela). Ako stavke mere istu latentnu osobinu, trebalo bi da su povezane. Stavka koja ne korelira, ili jako nisko korelira sa ostalim, može biti problematična. U našem primeru nema takvih stavki. Sve koreliraju značajno, ali vidimo da se stavka 7 javlja u prvih 8 parova najnižih korelacija. To bi moglo biti rezultat toga što ta stavka ima donekle drugačiji predmet merenja (na šta bismo mogli pomisliti gledajući ajtem–total korelacije u tabeli sa point–biserijskim korelacijama), ali i rezultat smanjene varijanse jer se radi o najtežoj stavci na koju je tačno odgovorilo samo 13% ispitanika (ovo vidimo iz tabele sa proporcijama odgovora).

Tabela sa proporcijama odgovora u svakoj od kategorija za svaku od stavki je prva tabela u gornjem ispisu. Ta tabela može biti vrlo korisna u slučaju da dobijemo čudne procene parametara. One su vrlo često rezultat niske ili nulte frekvencije odgovora u nekoj od kategorija.

#### *10.4.4.2 2PL model za binarno skorovane stavke u paketu mirt*

Ranije smo već prikazali funkciju za procenu ovog modela, ali ćemo je ovde ponoviti. Argument *SE=TRUE* traži procenu standardnih grešaka i potreban je ako želimo da bude procenjena i informativnost. Broj *1* je skraćeno pisanje argumenta *model=1* koji definiše dimenzionalnost modela. Argument *verbose=FALSE* sprečava ispis podataka o iteracijama.

```
p_load(mirt)
m2p.m <- mirt(b.pod, 1, SE=TRUE, verbose = FALSE)
# Prikaz parametara stavki
round(cbind(MDIFF(m2p.m), MDISC(m2p.m)), 3)
```

```
##      MDIFF_1
## it1   -0.104 1.095
## it2    0.040 1.782
## it3   -0.073 1.417
## it4   -0.208 1.384
## it5   -1.257 1.574
## it6   -1.197 1.514
## it7    1.756 1.458
## it8   -0.900 1.304
## it9   -1.195 1.900
## it10   0.822 1.730
```

```
# Izračunaćemo omnibus pokazatelje fita
M2(m2p.m)
```

```
##             M2 df         p       RMSEA RMSEA_5   RMSEA_95      SRMSR       TLI
## stats 38.18272 35 0.3268395 0.009540746       0 0.02530255 0.02073589 0.9984261
##             CFI
## stats 0.9987759
```

Kao što smo ranije već rekli, ovaj model ima zadovoljavajuće pokazatelje fita. $M_2$ pokazatelj bliskog fita nije značajan, a i pokazatelji približnog fita su u preporučenim granicama.

##### 10.4.4.2.1 Pokazatelji fita za stavke

Ovo su pokazatelji koji su već opisani kod primera za Rašov model (odeljak 10.4.3.2.1.2). Nećemo ih ponovo opisivati.

```
itemfit(m2p.m)
```

```
##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1  3.811       7      0.000  0.801
## 2   it2  6.727       7      0.000  0.458
## 3   it3 10.583       7      0.023  0.158
## 4   it4  2.274       7      0.000  0.943
## 5   it5  6.300       6      0.007  0.390
## 6   it6  3.193       6      0.000  0.784
## 7   it7  5.545       5      0.010  0.353
## 8   it8  8.616       7      0.015  0.281
```

```
## 9   it9  2.369       6      0.000  0.883
## 10 it10  3.722       6      0.000  0.714

# Ovom naredbom dobijamo podrazumevani S-X2 po proceduri Orlanda i Tisena
```

Na osnovu *p* vrednosti za hi–kvadrat testove i na osnovu pripadajućih *RMSEA* vidimo da nema misfitujućih stavki. Prikazaćemo još neke alternativne pokazatelje. Sledećom komandom dobijamo $Q_1$ pokazatelj Jenove i $G^2$ pokazatelj Mekinlija i Milsa.

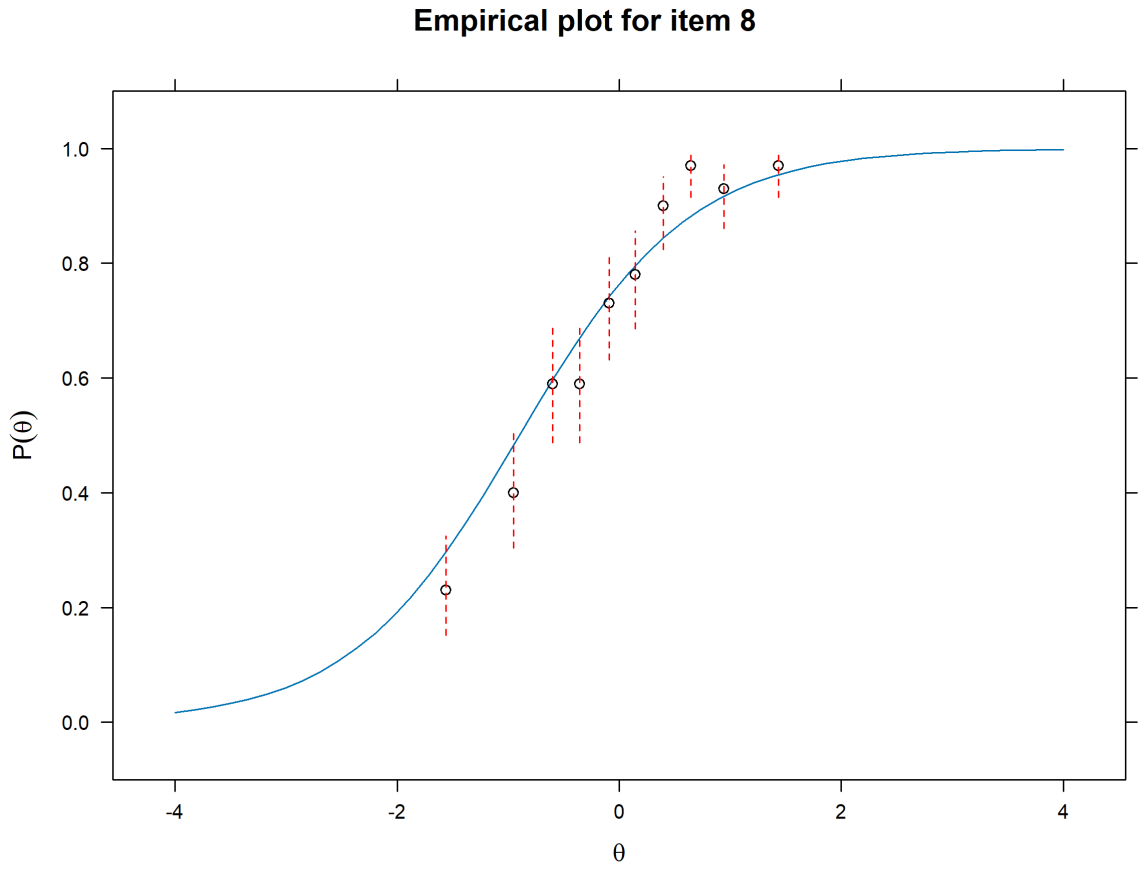
```
itemfit(m2p.m,  fit_stats=c("X2", "G2"))

##    item     X2 df.X2 RMSEA.X2  p.X2     G2 df.G2 RMSEA.G2  p.G2
## 1   it1 41.979     8    0.065 0.000 47.603     8    0.070 0.000
## 2   it2 55.383     8    0.077 0.000 70.748     7    0.095 0.000
## 3   it3 26.757     8    0.048 0.001 32.116     8    0.055 0.000
## 4   it4 19.883     8    0.039 0.011 22.182     8    0.042 0.005
## 5   it5 25.901     8    0.047 0.001 33.269     6    0.067 0.000
## 6   it6 27.754     8    0.050 0.001 31.985     7    0.060 0.000
## 7   it7 34.819     8    0.058 0.000 35.145     7    0.063 0.000
## 8   it8 18.649     8    0.037 0.017 21.834     8    0.042 0.005
## 9   it9 28.718     8    0.051 0.000 36.962     6    0.072 0.000
## 10 it10 57.693     8    0.079 0.000 87.036     6    0.116 0.000
```

Ovde je slika potpuno drugačija. Prema $Q_1$ 5 stavki ima značajan misfit, a prema $G^2$ sve. Ovi nalazi su kontradiktorni omnibus pokazatelju fita koji kaže da je model saglasan sa podacima. Naredni pokazatelj je Stounov hi–kvadrat koji se zasniva na samouzorkovanju. Za njega se pokazalo da ima veću snagu od S–$\chi^2$ i da istovremeno dobro kontroliše grešku tipa I (de Ayala, 2022). Zbog procedure Monte Karlo simulacija izračunavanje može da potraje neko vreme (minuta) u zavisnosti od brzine računara. Iz istog razloga pri ponovljenim računanjima *p* vrednosti neće bisti sasvim iste. Kada postoje *p* vrednosti koje su bliske 0,05 dobro je ponoviti izračunavanje nekoliko puta da bismo bili sigurni u zaključke koje donosimo.

```
itemfit(m2p.m,  fit_stats=c("X2*"), return.tables = FALSE)

##    item X2_star p.X2_star
## 1   it1   0.405     0.738
## 2   it2   0.484     0.412
## 3   it3   0.391     0.630
## 4   it4   0.474     0.566
## 5   it5   0.651     0.552
## 6   it6   1.027     0.227
## 7   it7   0.551     0.940
## 8   it8   1.038     0.243
## 9   it9   0.362     0.922
## 10 it10   0.231     0.958
```

Prema ovom pokazatelju nema misfitujućih stavki, a *p* vrednosti su dovoljno daleko od kritične da ne moramo da ponavljamo izračunavanje. Inače funkcija *itemfit()* ima opciju prikazivanja empirijske KKS uz modelsku, što nam može pružiti uvid u misfit stavke. Kako ovde nemamo misfitujućih stavki, za ilustraciju prikazaćemo onu sa najvećom vrednošću hi–kvadrata (stavka 8).

```
itemfit(m2p.m, empirical.plot=8)
```




Slika 50 – Modelska i empirijska karakteristična kriva stavke 8

Ispitanici su grupisani u 10 grupa po nivou osobine (Slika 50). Kružići predstavljaju empirijske proporcije predviđanog odgovora, a crvene isprekidane linije 95% intervale poverenja. Ako intervali poverenja ne zahvataju modelsku KKS, onda možemo reći da model ne predviđa dobro odgovore ispitanika. Na osnovu grafikona možemo videti da za većinu grupa (9/10) model dobro predviđa odgovaranje. Međutim, u grupi sa nivoom osobine oko 0,5 logita to nije tako. Ispitanici sa tim nivoom osobine češće odgovaraju tačno/potvrdno nego što to model predviđa.

#### 10.4.4.2.2 Pokazatelji fita za ispitanike

Ako ne navedemo pokazatelje fita koje želimo, *mirt* će prikazati infit, autfit (MSQ i z) i $Z_h$ pokazatelj Drazgova i saradnika (odeljak 6.4). Infit i autfit su pokazatelji koji se koriste u Rašovom modelu i nisu uobičajeni za modele koji uključuju diskriminativnost pa ćemo se osloniti na $Z_h$. Da ne bude zabune, infit i autfit su primenljivi i izvan Rašove familije modela (Meijer & Sijtsma, 2001).

```
PF.m <- personfit(m2p.m, method="EAP", fit_stats="Zh")

# Zbog prostora smo prikazali samo ispitanike kod kojih je apsolutna vrednost Zh
statistika veća od 1,96, što znači da postoji misfit
PF.m[abs(PF.m$Zh)>1.96,]
```

```
##       outfit  z.outfit    infit  z.infit        Zh
## 67  2.235144 1.7732771 1.787348 1.944245 -2.217973
## 117 2.727943 2.3158141 1.606388 1.658576 -2.179717
## 200 5.547438 3.0399073 1.600961 1.748875 -2.845275
## 297 2.801725 2.3633423 1.504008 1.440651 -1.994919
## 309 2.916249 2.4855933 1.849861 2.163669 -2.790830
## 334 1.558781 0.9538047 1.745998 2.022638 -2.009948
## 374 3.694699 3.0757594 2.335181 2.984004 -4.206270
## 384 2.916249 2.4855933 1.849861 2.163669 -2.790830
## 399 2.766765 1.9737259 2.075255 2.220184 -2.759063
## 408 2.614016 2.1667169 1.826722 2.049368 -2.546985
## 462 2.447096 1.9389054 1.664342 1.671710 -2.078253
## 488 2.649497 2.1407053 1.961331 2.249071 -2.782348
## 519 2.975690 2.1997548 1.657928 1.561248 -2.136599
## 585 3.735897 1.8878674 1.970559 1.644240 -2.193513
## 602 2.533791 1.7791663 1.710184 1.603471 -2.019869
## 606 2.467550 1.6370035 1.612112 1.761457 -2.216100
## 640 2.467550 1.6370035 1.612112 1.761457 -2.216100
## 702 2.008502 1.5691463 1.791494 2.075055 -2.347704
## 718 2.039040 1.4896318 1.647794 1.799270 -2.110888
## 735 2.533791 1.7791663 1.710184 1.603471 -2.019869
## 770 3.610907 2.7334231 1.523559 1.516876 -2.300328
## 774 5.510840 2.9767690 1.358783 1.146458 -2.206856
## 926 1.604717 0.9657306 1.823118 2.208490 -2.230511
## 981 3.172074 1.9912023 2.373290 2.428966 -3.223167
```

```
# Sledeća komanda je zbog prostora komentarisana i nije prikazan rezultat. Njome
se biraju svi ispitanici koji na infitu ili autfitu imaju vrednosti van opsega
0,7-1,3, a da je Zh u opsegu -1,96-1,96. To znači da infit i autfit ukazuju na
misfit, a Zh ne. Takvih ispitanika ima 691
#PF.m[(abs(1-PF.m$outfit)>0.3 | abs(1-PF.m$infit)>0.3) & !abs(PF.m$Zh)>1.96,]
```

Mere ispitanika izračunate su korišćenjem *EAP* metoda (*Expected A Posteriori*). Moguća je primena još nekih metoda, od kojih su *ML* (maksimalna verodostojost) i *MAP* (*Maximum A Posteriori*) opisani u odeljku 3.2.

Zbog prostora smo prikazali samo ispitanike kod kojih je apsolutna vrednost $Z_h$ statistika veća od 1,96, što znači da postoji misfit. U ovom slučaju kod svih ispitanika $Z_h$ je negativan što ukazuje na šum, odnosno, na nepredvidivo odgovaranje. Ako uporedimo vrednosti $Z_h$ sa pokazateljima autfita i infita, vidimo da se u ovim slučajevima slažu u pogledu postojanja misfita, a i u pogledu njegovog tipa (šum ili prigušeni šum). Međutim, postoji 691 ispitanik za koga bi se po kriterijumima za infit ili autfit moglo smatrati da postoji misfit, a da ga $Z_h$ ne detektuje. Takođe, ne deluje verovatno da 70% ispitanika ima misfit, a da je model saglasan sa podacima.

#### 10.4.4.2.3 Lokalna zavisnost

Proveru lokalne zavisnosti u paketu *mirt* možemo obaviti na više načina. Prvi se zasniva na hi–kvadratu za parove stavki. Dobijamo ga komandom *residuals* ako koristimo argument *type="LD".*

```
residuals(m2p.m, type="LD")

# Ako dodamo argument suppres=.20 neće biti prikazane korelacije manje od .20
(korelacije veće od te vrednosti ukazuju na lokalnu zavisnost)
```

```
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.084  -0.013   0.003   0.001   0.015   0.049
##
##        it1    it2   it3    it4    it5    it6    it7    it8    it9   it10
## it1     NA -0.034 0.015  0.029 -0.004 -0.013 -0.016  0.003 -0.001  0.033
## it2  1.143     NA 0.012 -0.009  0.010  0.049  0.041  0.020 -0.002 -0.050
## it3  0.226  0.133    NA -0.021  0.009 -0.009  0.004  0.015 -0.031  0.013
## it4  0.860  0.076 0.431     NA  0.023  0.014  0.005 -0.002 -0.001 -0.013
## it5  0.013  0.102 0.072  0.545     NA  0.020 -0.084 -0.036  0.014  0.008
## it6  0.171  2.383 0.080  0.205  0.418     NA -0.031 -0.018 -0.023  0.006
## it7  0.253  1.676 0.016  0.020  7.018  0.934     NA -0.004  0.022  0.003
## it8  0.007  0.400 0.221  0.003  1.307  0.320  0.013     NA  0.032 -0.001
## it9  0.002  0.003 0.972  0.001  0.197  0.524  0.502  1.048     NA  0.027
## it10 1.087  2.466 0.166  0.174  0.061  0.033  0.012  0.001  0.728     NA
```

```
residuals(m2p.m, type=c("LD"), df.p=T, suppres=.20)
```

```
## Degrees of freedom (lower triangle) and p-values:
##
##      it1   it2   it3   it4   it5   it6   it7   it8   it9  it10
## it1   NA 0.285 0.634 0.354 0.910 0.679 0.615 0.933 0.966 0.297
## it2    1    NA 0.715 0.782 0.750 0.123 0.195 0.527 0.957 0.116
## it3    1 1.000    NA 0.512 0.788 0.777 0.900 0.638 0.324 0.683
## it4    1 1.000 1.000    NA 0.460 0.650 0.887 0.958 0.970 0.676
## it5    1 1.000 1.000 1.000    NA 0.518 0.008 0.253 0.657 0.804
## it6    1 1.000 1.000 1.000 1.000    NA 0.334 0.571 0.469 0.856
## it7    1 1.000 1.000 1.000 1.000 1.000    NA 0.910 0.479 0.913
## it8    1 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.306 0.976
## it9    1 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.394
## it10   1 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA
##
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.084  -0.013   0.003   0.001   0.015   0.049
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Ako ne zatražimo ispis stepeni slobode (*df.p=T*), u gornjem trouglu matrice date su korelacije (Kramerovo V), a u donjem vrednosti hi–kvadrata. U protivnom, u donjem trouglu matrice biće prikazani stepeni slobode. Vidimo da na našim podacima ne postoji lokalna zavisnost (u tabeli *upper triangle summary* nema vrednosti većih od 0,20 pa su sve zamenjene sa NA zbog argumenta *suppress*).

Drugi način provere lokalne zavisnosti je hi–kvadrat ($G^2$) zasnovan na logaritmu verodostojnosti, takođe za parove stavki. Njega dobijamo argumentom *type="LDG2"*.

```
residuals(m2p.m, type="LDG2")

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.075  -0.013   0.003   0.001   0.015   0.050
##
##        it1    it2   it3    it4    it5    it6    it7    it8    it9   it10
## it1     NA -0.034 0.015  0.029 -0.004 -0.013 -0.016  0.003 -0.001  0.033
## it2  1.136     NA 0.012 -0.009  0.010  0.050  0.042  0.020 -0.002 -0.049
## it3  0.227  0.134    NA -0.021  0.009 -0.009  0.004  0.015 -0.031  0.013
## it4  0.864  0.076 0.429     NA  0.023  0.014  0.005 -0.002 -0.001 -0.013
## it5  0.013  0.102 0.073  0.551     NA  0.020 -0.075 -0.036  0.014  0.008
## it6  0.170  2.469 0.080  0.206  0.417     NA -0.029 -0.018 -0.023  0.006
## it7  0.250  1.762 0.016  0.020  5.651  0.859     NA -0.004  0.024  0.003
## it8  0.007  0.403 0.222  0.003  1.303  0.320  0.013     NA  0.032 -0.001
## it9  0.002  0.003 0.950  0.001  0.197  0.523  0.553  1.052     NA  0.028
## it10 1.101  2.411 0.167  0.173  0.062  0.033  0.012  0.001  0.774     NA

residuals(m2p.m, type="LDG2", df.p=T, suppres=.20)

## Degrees of freedom (lower triangle) and p-values:
##
##      it1   it2   it3   it4   it5   it6   it7   it8   it9  it10
## it1   NA 0.286 0.634 0.353 0.910 0.680 0.617 0.933 0.966 0.294
## it2    1    NA 0.715 0.782 0.749 0.116 0.184 0.525 0.957 0.121
## it3    1 1.000    NA 0.512 0.788 0.777 0.900 0.637 0.330 0.683
## it4    1 1.000 1.000    NA 0.458 0.650 0.886 0.959 0.970 0.678
## it5    1 1.000 1.000 1.000    NA 0.518 0.017 0.254 0.657 0.803
## it6    1 1.000 1.000 1.000 1.000    NA 0.354 0.572 0.470 0.856
## it7    1 1.000 1.000 1.000 1.000 1.000    NA 0.910 0.457 0.913
## it8    1 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.305 0.976
## it9    1 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.379
## it10   1 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA
##
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.075  -0.013   0.003   0.001   0.015   0.050
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Ispis za ovaj način provere lokalne zavisnosti isti je kao i za prethodni, a i nalazi u pogledu postojanja lokalne zavisnosti su isti.

I na kraju $Q_3$, odnosno korelacija reziduala. Dobijamo ga korišćenjem argumenta *type="Q3"*.

```
residuals(m2p.m, type="Q3")

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.230  -0.115  -0.080  -0.088  -0.055  -0.009
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.160 -0.066 -0.044 -0.076 -0.089 -0.071 -0.070 -0.085 -0.044
## it2  -0.160  1.000 -0.123 -0.149 -0.089 -0.032 -0.025 -0.088 -0.119 -0.230
## it3  -0.066 -0.123  1.000 -0.136 -0.071 -0.096 -0.064 -0.070 -0.140 -0.103
## it4  -0.044 -0.149 -0.136  1.000 -0.058 -0.071 -0.054 -0.095 -0.107 -0.126
## it5  -0.076 -0.089 -0.071 -0.058  1.000 -0.079 -0.102 -0.143 -0.110 -0.052
## it6  -0.089 -0.032 -0.096 -0.071 -0.079  1.000 -0.055 -0.118 -0.181 -0.052
## it7  -0.071 -0.025 -0.064 -0.054 -0.102 -0.055  1.000 -0.044 -0.009 -0.115
## it8  -0.070 -0.088 -0.070 -0.095 -0.143 -0.118 -0.044  1.000 -0.051 -0.080
## it9  -0.085 -0.119 -0.140 -0.107 -0.110 -0.181 -0.009 -0.051  1.000 -0.032
## it10 -0.044 -0.230 -0.103 -0.126 -0.052 -0.052 -0.115 -0.080 -0.032  1.000

residuals(m2p.m, type="Q3", suppres=.2)

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.230  -0.115  -0.080  -0.088  -0.055  -0.009
##
##      it1   it2 it3 it4 it5 it6 it7 it8 it9  it10
## it1    1
## it2       1.00                              -0.23
## it3              1
## it4                  1
## it5                      1
## it6                          1
## it7                              1
## it8                                  1
## it9                                      1
## it10     -0.23                                1.00
```

Kada se koristi granična vrednost korelacije reziduala od 0,36 koja kao značajan rezidual uzima korelaciju srednje veličine (Cohen, 1988) u našim podacima ne postoji lokalna zavisnost. Kada kao graničnu vrednost koristimo 0,20 postoji korelacija između stavki 2 i 10 koja je viša od nje. Kada $Q_3$ pokazatelje poredimo sa prosečnom vrednošću u matrici ne nalazimo dokaze lokalne zavisnosti.

```
Q3m <- as.matrix(residuals(m2p.m, type="Q3"))

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.230  -0.115  -0.080  -0.088  -0.055  -0.009
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.160 -0.066 -0.044 -0.076 -0.089 -0.071 -0.070 -0.085 -0.044
## it2  -0.160  1.000 -0.123 -0.149 -0.089 -0.032 -0.025 -0.088 -0.119 -0.230
## it3  -0.066 -0.123  1.000 -0.136 -0.071 -0.096 -0.064 -0.070 -0.140 -0.103
## it4  -0.044 -0.149 -0.136  1.000 -0.058 -0.071 -0.054 -0.095 -0.107 -0.126
## it5  -0.076 -0.089 -0.071 -0.058  1.000 -0.079 -0.102 -0.143 -0.110 -0.052
```

```
## it6  -0.089 -0.032 -0.096 -0.071 -0.079  1.000 -0.055 -0.118 -0.181 -0.052
## it7  -0.071 -0.025 -0.064 -0.054 -0.102 -0.055  1.000 -0.044 -0.009 -0.115
## it8  -0.070 -0.088 -0.070 -0.095 -0.143 -0.118 -0.044  1.000 -0.051 -0.080
## it9  -0.085 -0.119 -0.140 -0.107 -0.110 -0.181 -0.009 -0.051  1.000 -0.032
## it10 -0.044 -0.230 -0.103 -0.126 -0.052 -0.052 -0.115 -0.080 -0.032  1.000
```

```
Q3m[Q3m==1] <- NA
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", Q3mean, "\n\n")
```

```
##
## Prosečna vrednost Q3: -0.0883636
```

```
Q3c <- Q3m-Q3mean
Q3c
```

```
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.072  0.023  0.045  0.012 -0.001  0.017  0.018  0.004  0.044
## it2  -0.072     NA -0.035 -0.060 -0.001  0.056  0.063  0.001 -0.030 -0.141
## it3   0.023 -0.035     NA -0.048  0.017 -0.007  0.024  0.019 -0.051 -0.015
## it4   0.045 -0.060 -0.048     NA  0.031  0.017  0.035 -0.007 -0.019 -0.038
## it5   0.012 -0.001  0.017  0.031     NA  0.009 -0.013 -0.055 -0.022  0.036
## it6  -0.001  0.056 -0.007  0.017  0.009     NA  0.033 -0.030 -0.093  0.037
## it7   0.017  0.063  0.024  0.035 -0.013  0.033     NA  0.045  0.079 -0.027
## it8   0.018  0.001  0.019 -0.007 -0.055 -0.030  0.045     NA  0.037  0.008
## it9   0.004 -0.030 -0.051 -0.019 -0.022 -0.093  0.079  0.037     NA  0.056
## it10  0.044 -0.141 -0.015 -0.038  0.036  0.037 -0.027  0.008  0.056     NA
```

```
Q3c[Q3c<.20] <- NA
Q3c
```

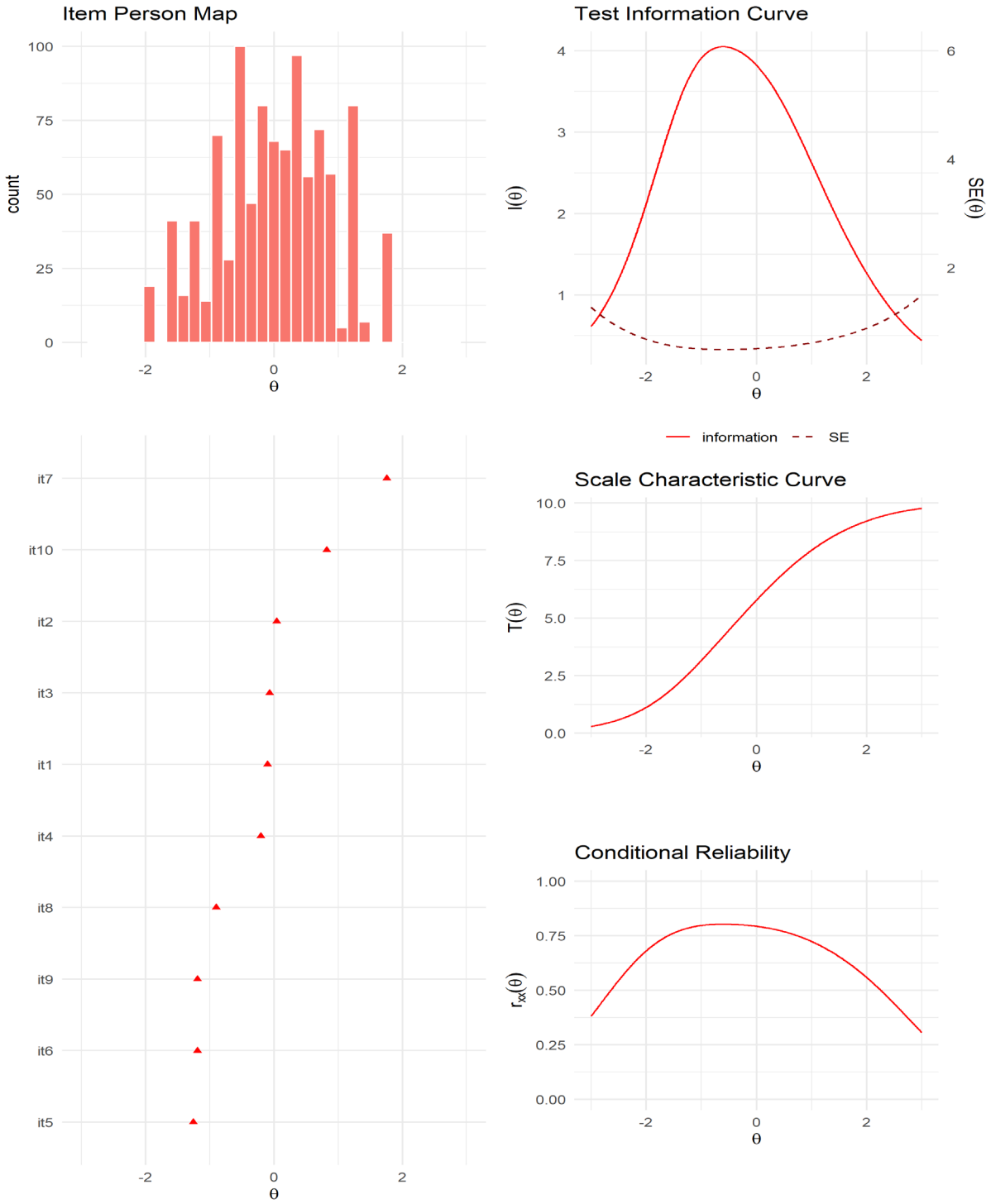
```
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

#### 10.4.4.2.4 Grafički prikaz

Što se tiče grafikona, u paketu *mirt* najproblematičniji su oni za karakteristične krive i krive informativnosti. Postoji dodatni paket za grafičke prikaze u *mirtu* pod nazivom *ggmirt* (Masur, 2022). U trenutku pisanja ove knjige paket je još u eksperimentalnoj fazi i nema ga na CRAN–ovim serverima, ali može se instalirati sa platforme *github.* Opisaćemo postupak.

```
# Potrebno je da imate instaliran paket devtools. Ako nemate, uklonite komentar sa
sledeće naredbe i izvršite je
#install.packages("devtools")

#Zatim, na sledeći način instalirajte ggmirt
#devtools::install_github("masurp/ggmirt")
library(ggmirt)
# Biće nam potrebna i biblioteka ggplot2
```

```
p_load(ggplot2)
# i RColorBrewer (zbog dodatnih boja za krive stavki)
p_load(RColorBrewer)
# Prikazaćemo grafikon koji sadrži više informacija na jednom mestu
summaryPlot(m2p.m, theta_range = c(-3, 3), adj_factor = 1.5)

## Warning: Removed 2 rows containing missing values or values outside the scale r
ange
## (`geom_bar()`).
```




Slika 51 – Dijagnostički grafikoni za test (paket *ggmirt*)

Ovaj grafikon (Slika 51) na jednom mestu daje niz pokazatelja koji su bitni za prikaz modela. Prikazana je mapa ispitanika i stavki na osnovu koje možemo videti koliko se nivo osobine u uzorku i distribucija parametara težina stavki preklapaju. Biće bolje procenjene mere ispitanika čiji nivoi osobine su dobro pokriveni stavkama testa odgovarajuće težine.

Pored ovog grafikona prikazana je i kriva informativnosti testa (zajedno sa standardnom greškom). Kriva nam može poslužiti da procenimo da li smo bitne delove kontinuuma latentne osobine dobro pokrili stavkama, odnosno, dovoljno precizno izmerili (Slika 51).

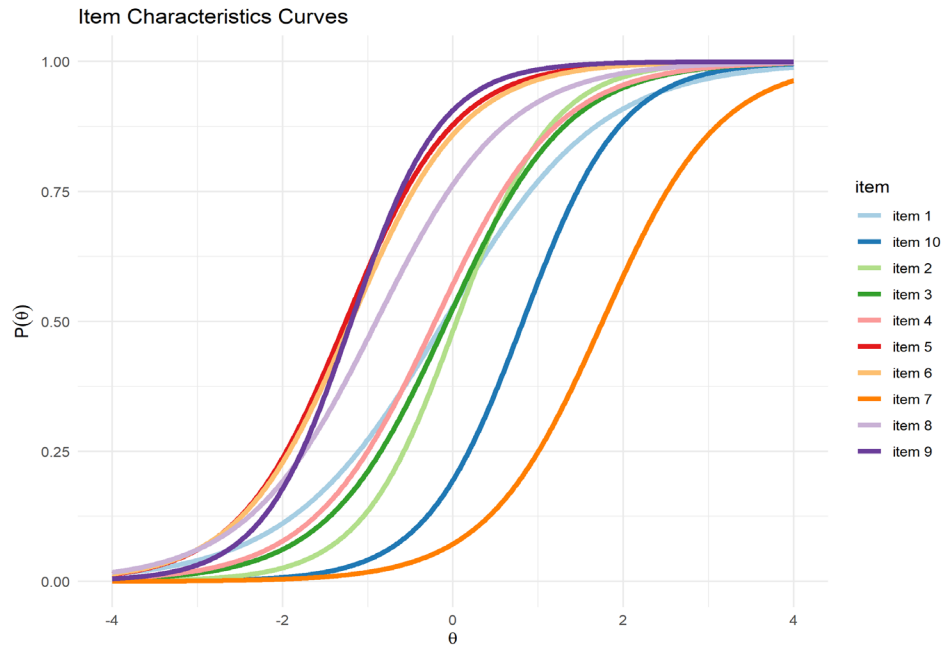
Takođe, tu je i grafikon uslovne pouzdanosti (Slika 51). Ovaj tip pouzdanosti opisan je ranije (odeljak 10.4.3.3).

Osim toga imamo i karakterističnu krivu testa (Slika 51). Na ovoj krivi prikazan je očekivani skor u zavisnosti od nivoa osobine. U Rašovom modelu to je prilično jednostavno jer osobe sa istim ukupnim skorom imaju istu meru. U modelima koji uključuju parametar diskriminativnosti istu meru imaju ispitanici koji imaju isti ponderisani skor, a kao što je rečeno to je suma ajtemskih skorova ponderisanih diskriminativnošću (odeljak 4.1.1).

Grafikone KKS i funkcija informativnosti prikazali smo ranije kao primer nepreglednih grafikona. Ovde ćemo sada prikazati grafikone dobijene pomoću paketa *ggmirt*.

I karakteristične krive stavki (Slika 52) i krive informativnosti (Slika 53) moguće je prikazati sve na jednom grafikonu ili odvojeno. Argument koji to kontroliše je *facet=TRUE* ili *facet=FALSE.*

```
p_load(RColorBrewer)
ggmirt::tracePlot(m2p.m, facet = FALSE, legend = TRUE)+
 ggplot2::scale_color_brewer(palette = "Paired") + ggplot2::geom_line(size = 1.5)
```

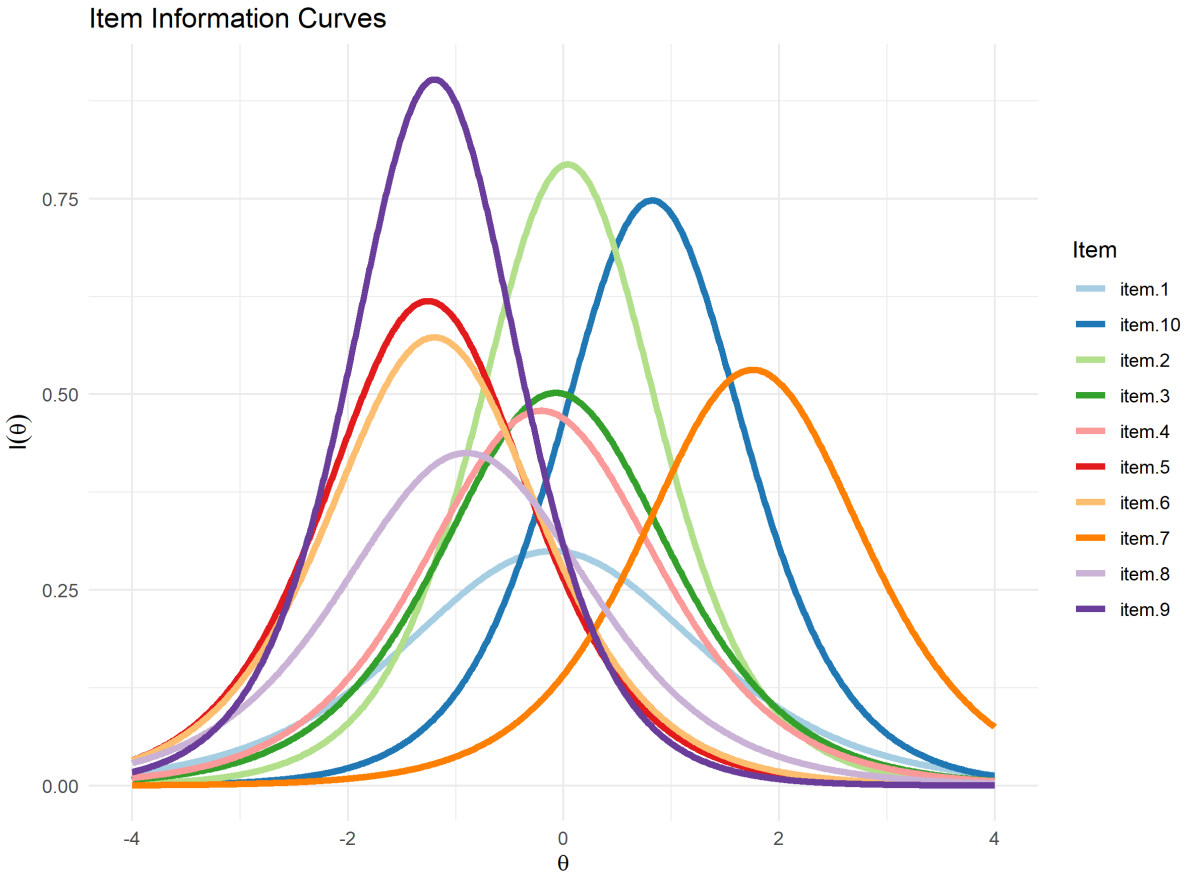



Slika 52 – KKS u 2PL modelu (paketi *mirt* i *ggmirt*)

Na grafikonu KKS (Slika 52) vidimo da najmanji nagib ima kriva stavke 1, a najveći kriva stavke 9. Kriva stavke 2 (druga po diskriminativnosti) ukršta se sa većim brojem drugih kriva.

Krive informativnosti stavki dobijamo funkcijom *itemInfoPlot().*

```
ggmirt::itemInfoPlot(m2p.m, legend = TRUE, facet = FALSE) +
 ggplot2::scale_color_brewer(palette = "Paired") +
  ggplot2::geom_line(size = 1.5)
```

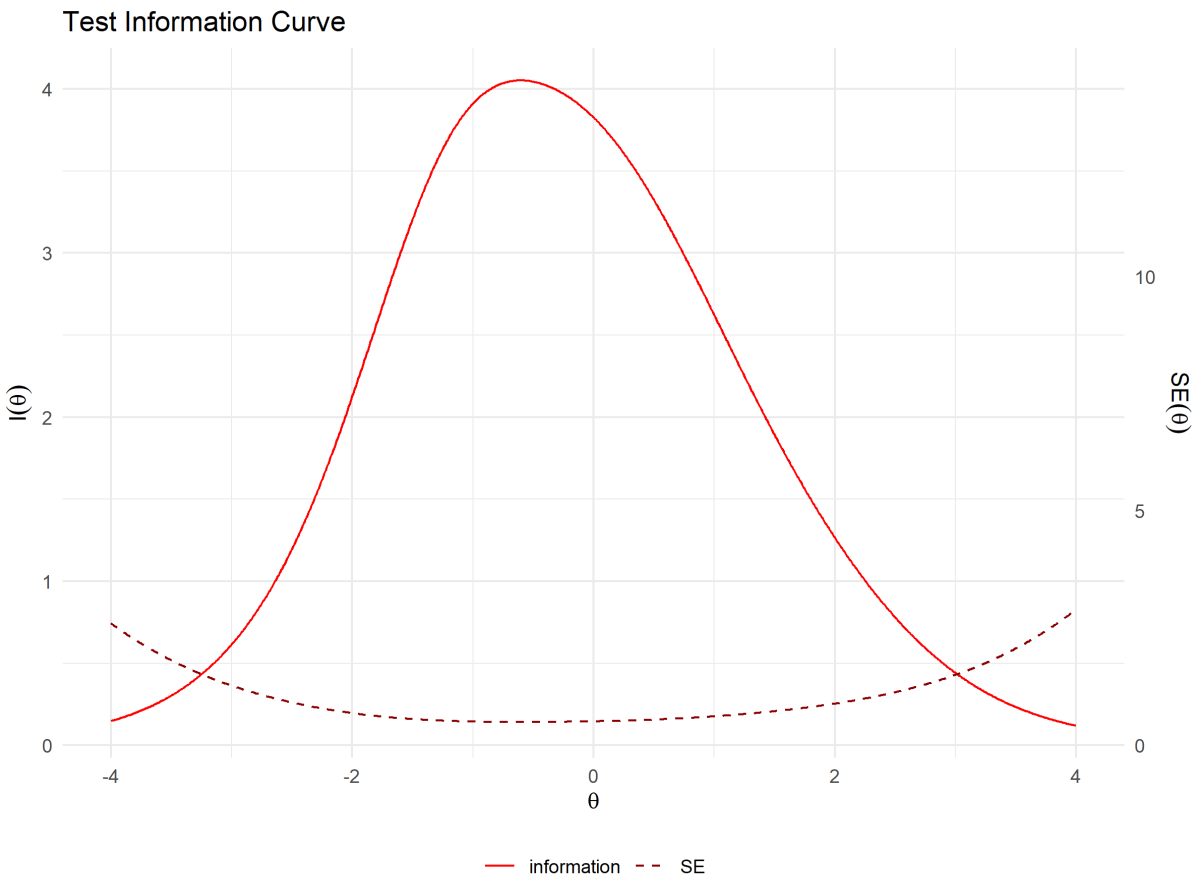



Slika 53 – Funkcije informativnosti stavki u 2PL modelu

Najinformativnije stavke su stavke 9, 2 i 10 koje pokrivaju različite nivoe latentne osobine (Slika 53). To je dobro ako želimo da test primenjuemo na opštoj populaciji. Veći broj stavki je prilagođen nivou osobine ispod proseka, pa bi valjalo dopuniti test stavkama koje su nešto teže.

Krivu informativnosti i standardne greške testa možemo dobiti funkcijom *testInfoPlot().*

```
testInfoPlot(m2p.m)
```




Slika 54 – Krive informativnosti i standardne greške testa u 2PL modelu

Informativnost testa najveća je na –0,60 (–0,58) logita (Slika 54). To se doduše ne vidi iz grafikona, ali znamo na osnovu izračunavanja koja smo obavili u paketu *ltm* (odeljak 10.4.4.1.3), a možemo i ovde izračunati.

#### 10.4.4.2.5 Informativnost

Informativnost u određenom opsegu osobine može se izračunati funkcijom *areainfo()*.

```
# Za ceo test
M2PI <- areainfo(m2p.m, c(-3,3))
M2PI

##  LowerBound UpperBound  Info TotalInfo Proportion nitems
##          -3          3 14.39  15.15825  0.9493178     10

# Samo za određene stavke
areainfo(m2p.m, c(-3,3), which.items = c(2,7:10)) # stavke 2 i od 7 do 10

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 7.762161  8.173876  0.9496304      5
```

Informativnost po nivoima merene osobine možemo dobiti na sledeći način:

```
# Prvo definišemo vrednosti latentne osobine za koje želimo da izračunamo
informativnost (sekvenca u rasponu od -3 do 3 sa koracima od 0,1)
thetaM <- seq(-3,3, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(m2p.m, Theta = thetaM)
# Prikaz prvih 6 redova (uklonite head() za ispis cele tabele)
head(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
```

```
##      nivo osobine informativnost
## [1,]         -3.0      0.6149720
## [2,]         -2.9      0.7056687
## [3,]         -2.8      0.8082465
## [4,]         -2.7      0.9237018
## [5,]         -2.6      1.0529128
## [6,]         -2.5      1.1965542
```

```
cat("Maksimalna informativnost:\n")
```

```
## Maksimalna informativnost:
```

```
max(infoM)
```

```
## [1] 4.051365
```

```
cat("Nivo latentne osobine na kojem je informativnost maksimalna:\n")
```

```
## Nivo latentne osobine na kojem je informativnost maksimalna:
```

```
thetaM[infoM==max(infoM)]
```

```
## [1] -0.6
```

```
cat("Raspon latentne osobine u kojem je informativnost >3,33:\n")
```

```
## Raspon latentne osobine u kojem je informativnost >3,33:
```

```
thetaM[infoM>3.33]
```

```
##  [1] -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1  0.0
## [16]  0.1  0.2  0.3  0.4
```

Vidimo da je maksimalna informativnost testa 4,087 na nivou osobine od –0,6 logita. Raspon u kojem je informativnost prihvatljiva (>3,33) je od –1,4 do 0,4 logita.

```
# Možemo izračunati pouzdanost po nivoima osobine na osnovu poznatog odnosa sa
informativnošću rtt=1-(1/infoM)
rtt <- ifelse(!(1-(1/infoM))<0, 1-(1/infoM), 0)
cat("Raspon latentne osobine u kojem je pouzdanost >0,70:\n")
```

```
## Raspon latentne osobine u kojem je pouzdanost >0,70:
```

```
thetaM[rtt>.7]
```

```
##  [1] -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1  0.0
## [16]  0.1  0.2  0.3  0.4
```

```
# Možemo izračunati standardnu grešku merenja po nivoima osobine na osnovu
poznatog odnosa sa informativnošću se=1/sqrt(infoM)
```

```
se <- 1/sqrt(infoM)
cat("Raspon latentne osobine u kojem je SE <0,548:\n")

## Raspon latentne osobine u kojem je SE <0,548:

thetaM[se<.548]

##  [1] -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1  0.0
## [16]  0.1  0.2  0.3  0.4

head(cbind("nivo osobine"=thetaM, "informativnost"=round(infoM, 3),
"pouzdanost"=round(rtt,3), "sgm"=round(se, 3)))

##      nivo osobine informativnost pouzdanost   sgm
## [1,]         -3.0          0.615      0.000 1.275
## [2,]         -2.9          0.706      0.000 1.190
## [3,]         -2.8          0.808      0.000 1.112
## [4,]         -2.7          0.924      0.000 1.040
## [5,]         -2.6          1.053      0.050 0.975
## [6,]         -2.5          1.197      0.164 0.914
```

Ako prihvatljivu informativnost definišemo kao informativnost preko 3,33 (što odgovara pouzdanosti preko 0,7), onda je raspon u kojem je naš test dovoljno precizan između –1,4 i 0,5 logita. Sledeće pitanje je da li nam to odgovara? Merenje ovim testom će biti dovoljno precizno na ispodprosečnom nivou osobine i u uskom opsegu oko proseka. Ako želimo da precizno merimo i natprosečne nivoe osobine, trebalo bi dodati još stavki, ali ne bilo kojih. Potrebne su nam informativne stavke na delu kontinuuma osobine gde nam je informativnost nedovoljna. Recimo, stavke težine iznad 0,5 logita, ravnomerno raspoređene duž kontinuuma osobine.

Za trenutak ćemo se prebaciti na informativnost stavki. Koristeći funkciju *iteminfo()* možemo izračunati informativnost stavki za različite nivoe latentne osobine. Ove podatke možemo iskoristiti za izbor stavki u test. Biće prikazan primer računanja prosečne informativnosti stavki za tri tačke pomenute u odeljku 7.1. To su tačke –1, 1 i 2 logita (mogu biti i neke druge, ali su ove uzete za primer). Potrebno je definisati numerički vektor koji sadrži nivoe latentne osobine koji nas zanimaju. U primeru je to objekat *tete*. Takođe, iz modela (*m2p.m*) potrebno je izvući podatke koji se odnose na željenu stavku. To smo uradili koristeći funkciju *extract.item(m2p.m, item=i)* koju smo stavili unutar funkcije *item.info()*. Rezultat će biti vektor koji sadrži informativnost stavke na zadatim nivoima latentne osobine. Ovaj vektor smo prosledili funkciji *mean()* i dobili prosečnu informativnost stavke na zadatim nivoima latentne osobine. Sve to smo upisali u objekat *ii*. Koristili smo petlju *for* da izračunamo prosečne informativnosti za sve stavke.

```
tete <- c(-1, 1, 2)
for(i in 1:ncol(b.pod))
  {ii <- mean(iteminfo(extract.item(m2p.m, item=i), Theta = tete,
   total.info = FALSE))
  cat(paste0("Prosečna informativnost it", i, " za lokacije ", paste(tete,
  collapse=", "), " iznosi: ", round(ii, 2), "\n"))
}

## Prosečna informativnost it1 za lokacije -1, 1, 2 iznosi: 0.09
## Prosečna informativnost it2 za lokacije -1, 1, 2 iznosi: 0.15
## Prosečna informativnost it3 za lokacije -1, 1, 2 iznosi: 0.12
## Prosečna informativnost it4 za lokacije -1, 1, 2 iznosi: 0.12
```

```
## Prosečna informativnost it5 za lokacije -1, 1, 2 iznosi: 0.11
## Prosečna informativnost it6 za lokacije -1, 1, 2 iznosi: 0.11
## Prosečna informativnost it7 za lokacije -1, 1, 2 iznosi: 0.16
## Prosečna informativnost it8 za lokacije -1, 1, 2 iznosi: 0.1
## Prosečna informativnost it9 za lokacije -1, 1, 2 iznosi: 0.16
## Prosečna informativnost it10 za lokacije -1, 1, 2 iznosi: 0.19
```

Najvišu prosečnu informativnost za zadate lokacije imaju stavke 10 i 7, a najnižu stavka 1.

Izračunaćemo empirijsku i marginalnu (uslovna) pouzdanost.

```
# Parametre ispitanika i standardne greške dobijamo funkcijom fscores i smeštamo u objekat THETA_SE koji prosleđujemo funkciji za računanje empirijske pouzdanosti
THETA_SE <- fscores(m2p.m, full.scores.SE = TRUE)
head(THETA_SE)
```

```
##              F1     SE_F1
## [1,] -0.5078283 0.4522012
## [2,] -0.4652698 0.4522777
## [3,]  0.1621199 0.4683784
## [4,] -0.9442547 0.4606318
## [5,] -0.3231324 0.4535247
## [6,]  0.4645001 0.4857366
```

```
cat("\nEmpirijska pouzdanost:", empirical_rxx(THETA_SE))
```

```
##
## Empirijska pouzdanost: 0.7582367
```

```
cat("\nMarginalna pouzdanost:", marginal_rxx(m2p.m))
```

```
##
## Marginalna pouzdanost: 0.7540355
```

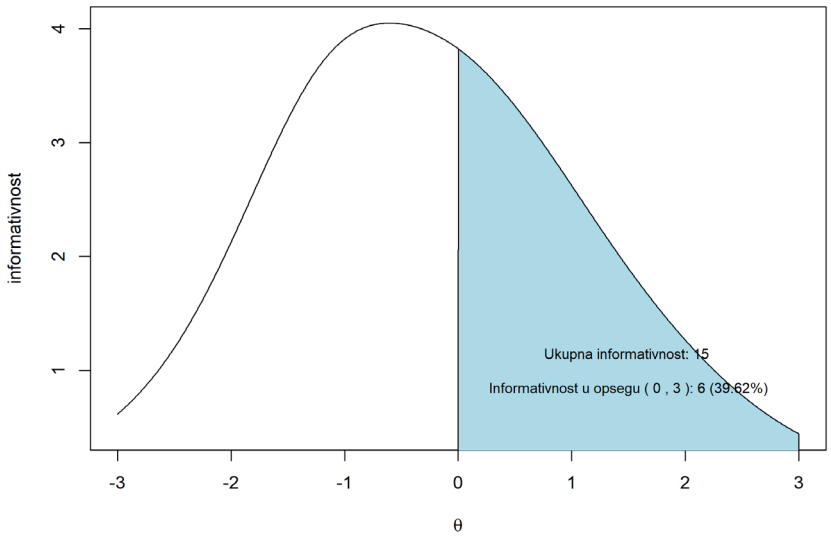
```
cat("\nKronbahova alfa:", psych::alpha(b.pod)$total$raw_alpha)
```

```
##
## Kronbahova alfa: 0.7572566
```

Kao što je rečeno, od jednog pokazatelja pouzdanosti korisnije nam je da znamo u kom rasponu osobine je merenje prihvatljivo precizno, odnosno informativno. Narednim komandama možemo grafički predstaviti informativnost u željenom rasponu (Slika 55).

```
# Za crtanje površine informativnosti u željenom rasponu potrebno je označiti sledeće komande i pokrenuti ih
# Ako želite da prikažete drugi raspon, potrebno je zameniti brojeve u zagradi kod raspona
# Kopirali smo objekat u kojem je smešten model kako bismo lakše mogli da koristim o donji skript. U slučaju da želite da prikažete ovaj grafikon za drugi model, sam o kopirajte taj drugi model u objekat pod imenom mod (prva linija)
mod <- m2p.m
par(mfrow=c(1,1))
raspon <- c(0,3) # ovde
povrsina <- areainfo(mod, raspon)
Teta <- matrix(seq(-3,3, length.out=1000))
informativnost <- testinfo(mod, Teta)
plot(informativnost ~ Teta, type = 'l', xlab=expression(theta))
izbor <- Teta >= raspon[1] & Teta <=raspon[2]
polygon(c(raspon[1], Teta[izbor], raspon[2]), c(0.29, informativnost[izbor],
 0.29), col='lightblue')
```

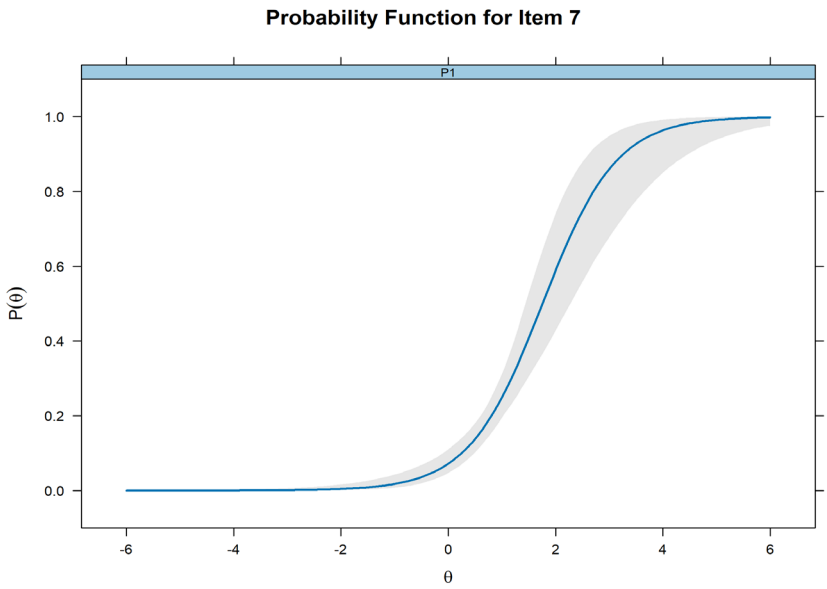
```
text(x = mean(raspon), y = 1, labels = paste("Ukupna informativnost:",
 round(povrsina$TotalInfo, 0),"\n\nInformativnost u opsegu (", raspon[1],",",
 raspon[2], "):", round(povrsina$Info, 1),
 paste("(", round(100 * povrsina$Proportion, 2), "%)", sep = "")), cex = .8)
```




Slika 55 – Površina informativnosti u opsegu osobine od 0 do 3 logita

U nastavku su date komande za različite grafikone. Prvi je grafikon KKS sa regionom poverenja (engl. *confidence envelope*) koji dobijamo argumentom *CE=T(RUE)* (Slika 56). Argument *CEalpha* određuje širinu regiona poverenja (0.05 daje 95% region), a *CEdraws* određuje broj bootstrapp uzoraka na kojima se region određuje.

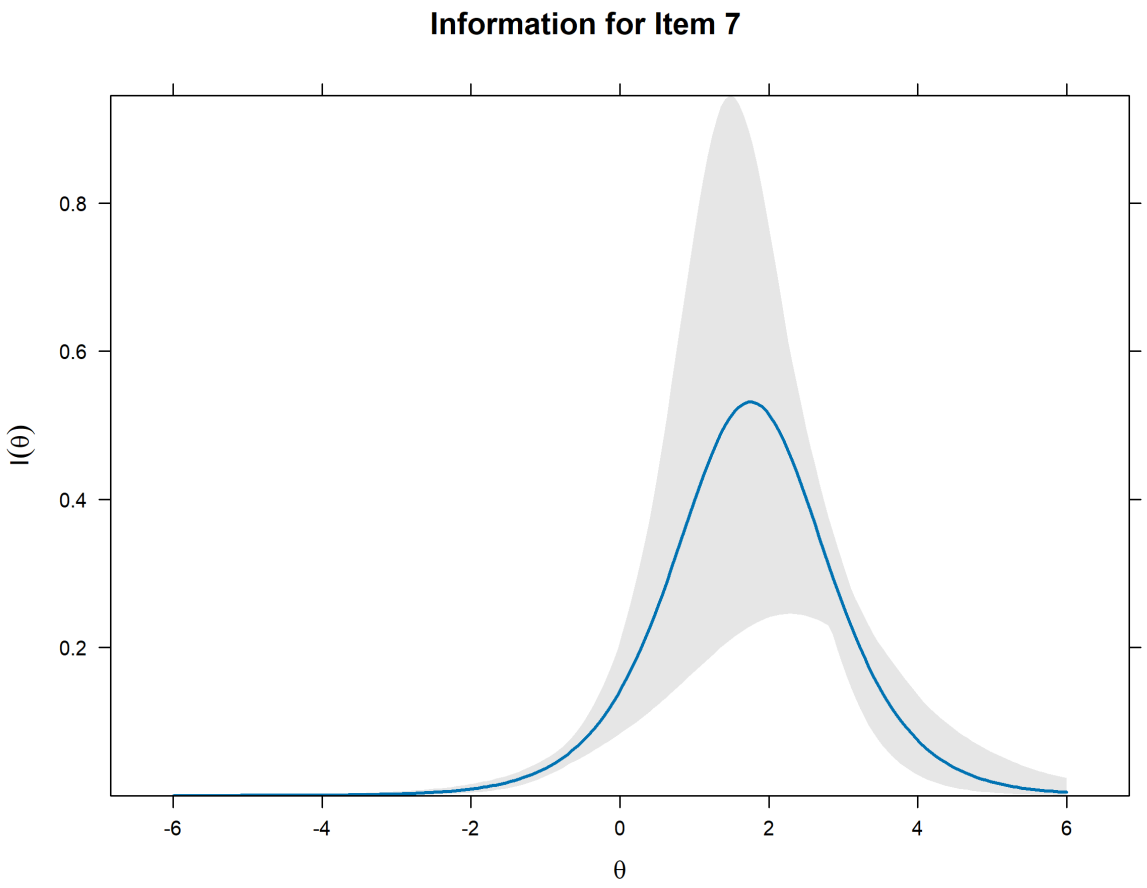
```
itemplot(m2p.m, item=7, lwd=2, type="trace", CE=T, CEalpha=0.05, CEdraws=1000)
```




Slika 56 – Karakteristična kriva stavke 7 sa regionom poverenja

I informativnost stavke možemo prikazati sa regionom poverenja (Slika 57). Argumenti koji to određuju su isti (osim argumenta *type="info"*). Sa ovim regionom ima smisla prikazivati informativnost pojedinačne stavke.

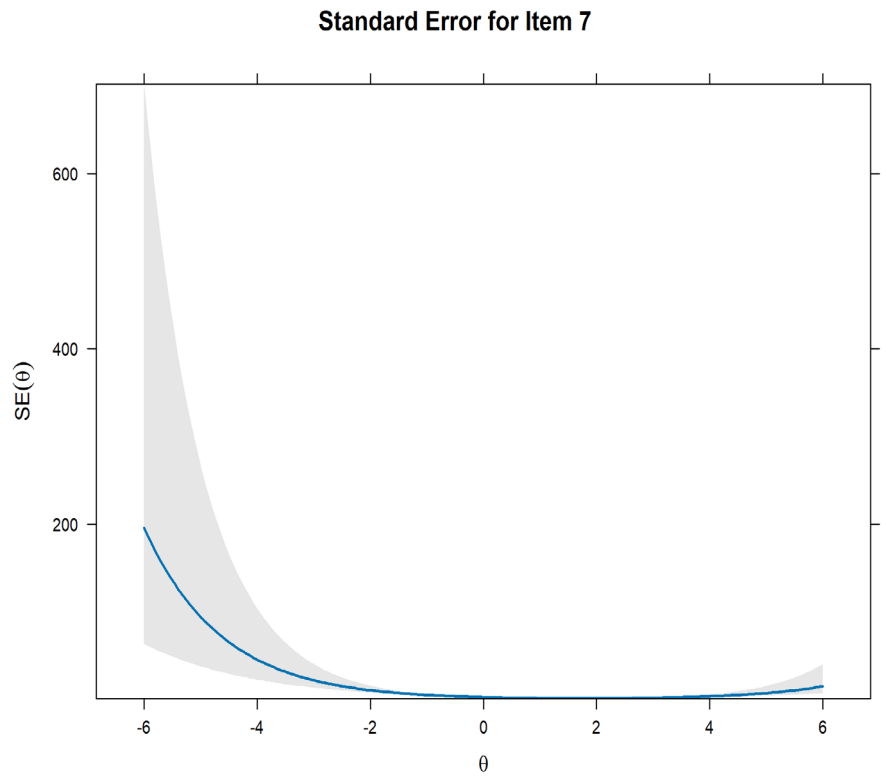
```
# Informativnost stavke (prvi broj je broj stavke)
itemplot(m2p.m, item=7, lwd=2, type="info", CE=T, CEalpha=0.05, CEdraws=1000)
```




Slika 57 – Funkcija informativnosti stavke 7 sa regionom poverenja

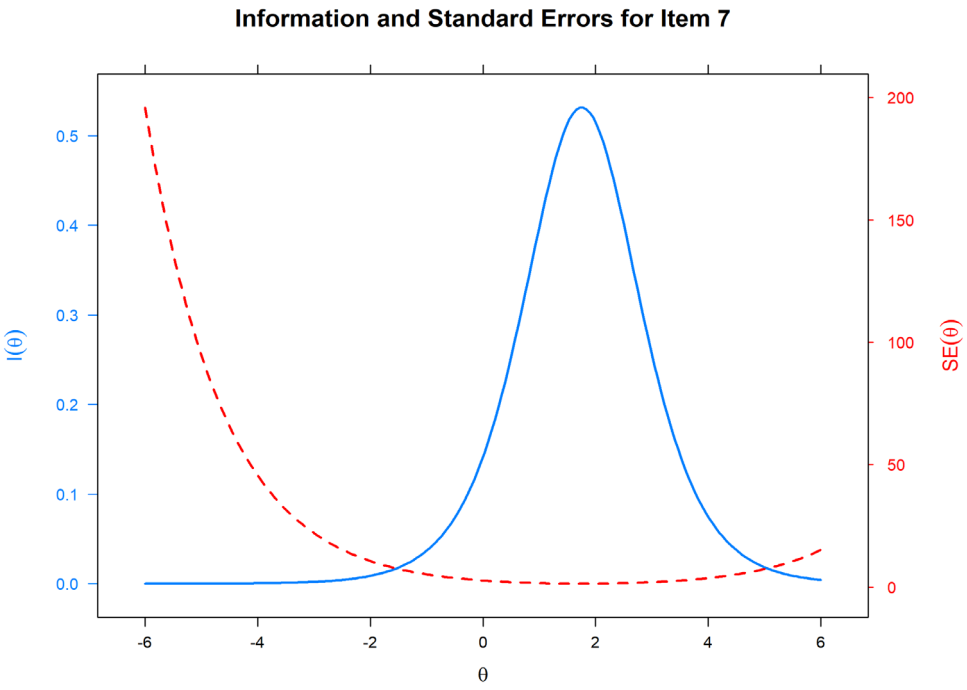
Dodatni grafikoni za stavke:

```
# Standardna greška stavke (argument item određuje broj stavke)
itemplot(m2p.m, item=7, lwd=2, type="SE", CE=T, CEalpha=0.05, CEdraws=1000)
```



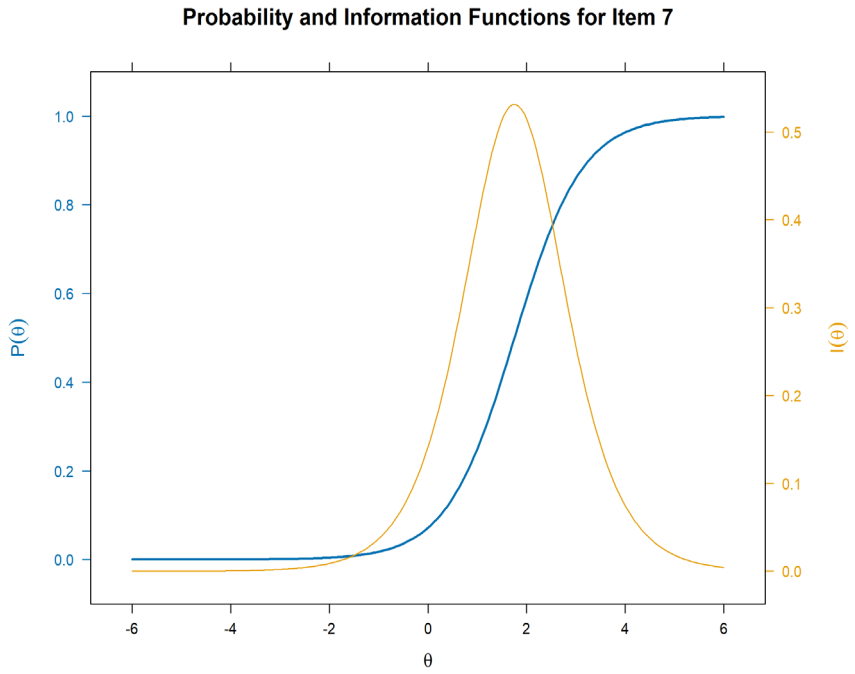

Slika 58 – Kriva standardne greške stavke 7 sa regionom poverenja

```
# Informativnost i standardna greška stavke
itemplot(m2p.m, item=7, lwd=2, type="infoSE")
```




Slika 59 – Kriva informativnosti i standardne greške stavke 7

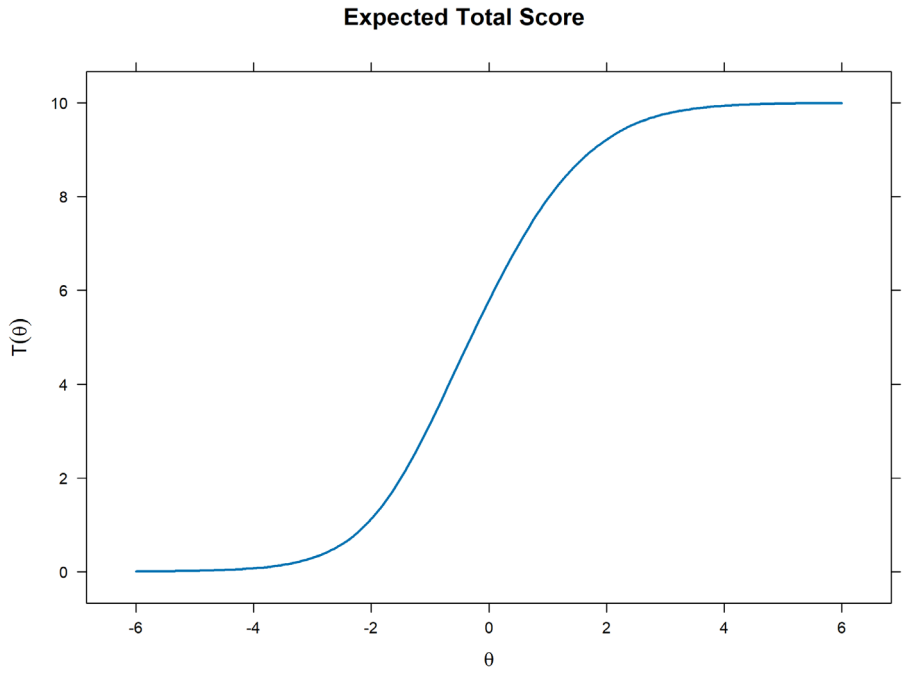
```
# KKS i informativnost stavke
itemplot(m2p.m, item=7, lwd=2, type="infotrace")
```




Slika 60 – Karakteristična kriva i funkcija informativnosti stavke 7

Za test u celini možemo dobiti i druge korisne grafikone. Prvi je očekivani ukupni skor na testu u zavisnosti od nivoa osobine (Slika 61).

```
plot(m2p.m, lwd=2)
```




Slika 61 – Kriva očekivanog ukupnog skora na testu u zavisnosti od nivoa osobine

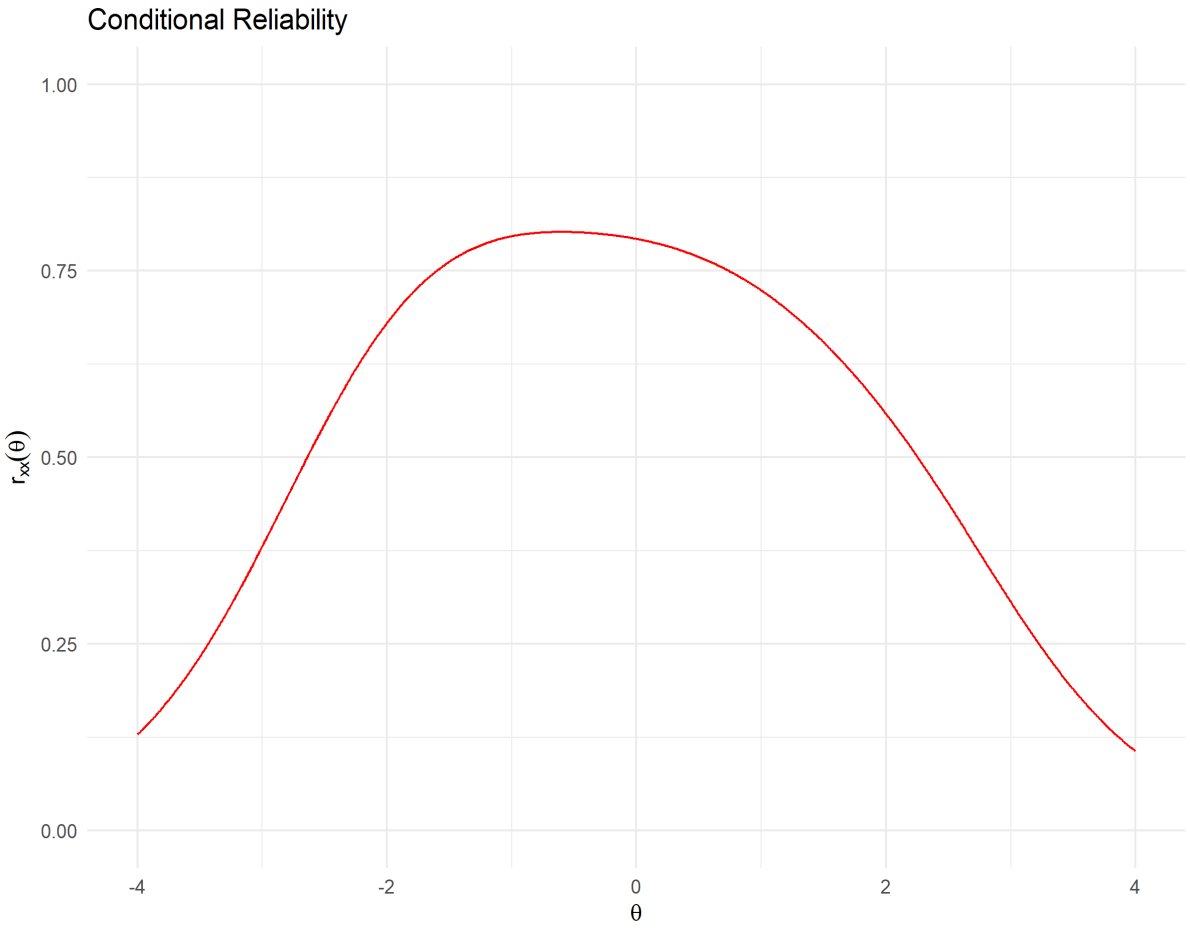
Te podatke možemo dobiti i u obliku tabele:

```
# Koristićemo ranije kreirani objekat (vektor) thetaM, ali ga moramo dati kao
matricu
OS <- expected.test(m2p.m, Theta = as.matrix(thetaM))
# Prikazano je 10 najviših nivoa osobine
tail(cbind("nivo osobine"=thetaM, "očekivani skor"=round(OS, 2),
 "skor"=round(OS)), 10)
```

```
##       nivo osobine očekivani skor skor
## [52,]          2.1           9.30    9
## [53,]          2.2           9.38    9
## [54,]          2.3           9.44    9
## [55,]          2.4           9.50   10
## [56,]          2.5           9.56   10
## [57,]          2.6           9.61   10
## [58,]          2.7           9.66   10
## [59,]          2.8           9.70   10
## [60,]          2.9           9.73   10
## [61,]          3.0           9.77   10
```

Zanimljiv može biti i grafikon uslovne pouzdanosti (Slika 62).

```
conRelPlot(m2p.m) # Iz paketa ggmirt
```




Slika 62 – Kriva uslovne pouzdanosti

```
# Ovo su izvorni grafikoni iz paketa mirt (oni iz paketa ggmirt su ranije
prikazani)
#plot(m2p.m, type="rxx", lwd=2)
#plot(m2p.m, type="SE", lwd=2)
#plot(m2p.m, type="trace", lwd=2)
#plot(m2p.m, type="infotrace", lwd=2)
```

### 10.4.5 Procena 3PL modela za binarno skorovane stavke u paketu *mirt*

Procena 3PL modela u paketu *mirt* obavlja se na sledeći način.

```
p_load(mirt)
m3p.m <- mirt(b.pod, 1, itemtype="3PL", SE=T, verbose = FALSE)
```

Argument *itemtype="3PL"* traži troparametarski model, a argument *SE=T* zahteva izračunavanje standardnih grešaka i informativnosti. "T" u ovom argumentu je skraćena oznaka za "TRUE". Vodite računa da ne kreirate objekat koji se zove "T" jer onda nećete moći da koristite skraćenu oznaku. Pošto se parametri pogađanja relativno teško procenjuju nekada je potrebno povećati broj iteracija *technical=list(NCYCLES=1500)* kako bi model konvergirao (broj može biti i veći od 1500).

```
m3p.m <- mirt(b.pod, 1, itemtype="3PL", SE=TRUE, technical=list(NCYCLES=1500),
 verbose=FALSE)

coef(m3p.m, simplify=TRUE, IRTpars=TRUE)$items[,1:3]

##             a           b            g
## it1  1.097714 -0.10182323 6.690748e-05
## it2  1.788729  0.04259161 1.977625e-05
## it3  1.438802 -0.05756428 6.497094e-03
## it4  1.380400 -0.20632029 2.399592e-05
## it5  1.589644 -1.24845734 2.723581e-04
## it6  1.529424 -1.18817379 3.004672e-05
## it7  1.458056  1.75451447 6.793823e-06
## it8  1.742871 -0.42270858 2.343267e-01
## it9  2.119085 -1.00032906 1.370202e-01
## it10 1.730854  0.82127446 1.216603e-05
```

Model možemo proceniti i u paketu *ltm* funkcijom *tpm()*. Nećemo učitavati paket *ltm* već ćemo funkciju pozvati direktno.

```
m3p.lt <- ltm::tpm(b.pod)
# Ako na ovaj način prijavi grešku, probati sa argumentom, start.val="random"
m3p.lt <- ltm::tpm(b.pod, start.val="random")
# Ako ni tada ne uspe, pokušajte ponovo ili zadajte neke plauzibilne vrednosti
parametara kao start.val
BS <- ncol(b.pod) #Varijabli BS dali smo vrednost broja stavki u matrici podataka
# Kreiramo matricu u kojoj su početne vrednosti parametra pogađanja za sve stavke
(prva kolona) jednake .05, sve početne vrednosti težina (druga kolona) jednake 0,
a početne vrednosti parametra diskriminativnosti za sve stavke (treća kolona)
jednake 1
SV <- matrix(c(rep(.05, BS), rep(0, BS), rep(1, BS)), nrow=BS, ncol=3,
 byrow=FALSE)
# Zatim tu matricu prosleđujemo argumentu start.val
m3p.lt <- ltm::tpm(b.pod, start.val=SV)
```

Uporedićemo procene na osnovu dva paketa. Ovo nije nužno da radite, ali je ovde dato kao ilustracija u sklopu prikaza paketa.

```
round(cbind(coef(m3p.m, simplify=TRUE, IRTpars=TRUE)$items[,1:3],
 coef(m3p.lt)[, 3:1]), 3)
```

```
##          a      b     g Dscrmn Dffclt Gussng
## it1  1.098 -0.102 0.000  1.098 -0.102  0.000
## it2  1.789  0.043 0.000  1.789  0.042  0.000
## it3  1.439 -0.058 0.006  1.423 -0.070  0.000
## it4  1.380 -0.206 0.000  1.380 -0.207  0.000
## it5  1.590 -1.248 0.000  1.591 -1.248  0.000
## it6  1.529 -1.188 0.000  1.531 -1.187  0.000
## it7  1.458  1.755 0.000  1.458  1.754  0.000
## it8  1.743 -0.423 0.234  1.744 -0.422  0.235
## it9  2.119 -1.000 0.137  2.132 -0.991  0.143
## it10 1.731  0.821 0.000  1.731  0.821  0.000
```

Pošto koriste isti način procene (MMLE) razlike postoje tek na trećoj decimali (na stavci 3 i 9 i na drugoj). Parametar pogađanja različit od 0 imamo na stavkama 8 i 9. Pogađanje nije nešto što želimo da imamo u testu. Viši parametar pogađanja smanjuje varijansu pravog skora. Npr. kada je pogađanje $c_8$=0,24 to znači da variranje verovatnoće tačnog odgovora koje zavisi od nivoa osobine može biti samo 1–0,24=0,76 iako, teorijski, može varirati od 0 do 1. Vrednosti parametra $c_j$ manje od 0,35 smatraju se prihvatljivim (Baker, 2001). Vrednosti parametara za stavke 8 i 9 ne prelaze tu vrednost. Uporedićemo parametre 2PL i 3PL modela da vidimo kako procena parametra pogađanja utiče na procene ostalih parametara stavki.

```
round(cbind(coef(m2p.m, simplify=TRUE, IRTpars=TRUE)$items[,1:2],
 coef(m3p.m, simplify=TRUE, IRTpars=TRUE)$items[,1:3]),3)
```

```
##          a      b     a      b     g
## it1  1.095 -0.104 1.098 -0.102 0.000
## it2  1.782  0.040 1.789  0.043 0.000
## it3  1.417 -0.073 1.439 -0.058 0.006
## it4  1.384 -0.208 1.380 -0.206 0.000
## it5  1.574 -1.257 1.590 -1.248 0.000
## it6  1.514 -1.197 1.529 -1.188 0.000
## it7  1.458  1.756 1.458  1.755 0.000
## it8  1.304 -0.900 1.743 -0.423 0.234
## it9  1.900 -1.195 2.119 -1.000 0.137
## it10 1.730  0.822 1.731  0.821 0.000
```

Ono što možemo primetiti je da su promene parametara težine i diskriminativnosti manje što je parametar pseudopogađanja bliži 0. Najveće su kod stavke 8. Povećana je težina i diskriminativnost stavke. U troparametarskim modelima parametar težine stavke ne nalazi se na nivou osobine na kom je verovatnoća da ispitanik odgovori tačno 0,5 već 0,5+$c_j$/2. Osim toga, pošto do određenog nivoa osobine svi ispitanici imaju jednaku verovatnoću ($c_j$) da odgovore tačno, raspon nivoa osobine u kojem ona utiče na verovatnoću tačnog odgovora često je sužen pa nagib KKS postaje strmiji, odnosno stavka diskriminativnija.

#### *10.4.5.1 Globalni fit modela*

Globalni fit modela procenićemo pokazateljem $M_2$ i pokazateljima približnog fita.

```
M2(m3p.m) # model ima dobre pokazatelje fita
```

```
##             M2 df         p RMSEA RMSEA_5   RMSEA_95      SRMSR      TLI CFI
## stats 22.41761 25 0.6115303     0       0 0.02212776 0.02056219 1.001788   1
```

Svi pokazatelji ukazuju na to da je model saglasan sa podacima, što nije neočekivano jer su bili saglasni i restriktivniji modeli (Rašov i 2PL).

Pošto smo te modele ranije procenili, iskoristićemo mogućnost da ih uporedimo LR testom i vidimo da li nam je uopšte neophodan kompleksniji model. Koristićemo činjenicu da je Rašov model ugnježden u 2PL modelu, a oba u 3PL modelu, pa ćemo funkciji *anova()* kao objekte proslediti sva tri procenjena modela (od najjednostavnijeg ka najkompleksnijem).

```
anova(mr.m, m2p.m, m3p.m)
```

```
##            AIC    SABIC       HQ      BIC    logLik     X2 df     p
## mr.m  10409.77 10428.82 10430.29 10463.76 -5193.887
## m2p.m 10402.10 10436.73 10439.40 10500.25 -5181.049 25.675  9 0.002
## m3p.m 10418.62 10470.57 10474.58 10565.85 -5179.308  3.482 10 0.968
```

Kada posmatramo značajnost LR testa ($p=0,004$) vidimo da 2PL model ima značajno bolju saglasnost od Rašovog (mada i on ima zadovoljavajuće pokazatelje fita). AIC se slaže sa tom procenom (niži je kod 2PL modela), ali BIC prednost daje Rašovom modelu. Između 2PL i 3PL modela nema značajne razlike ($p$=0,951) kada posmatramo LR test, a AIC i BIC prednost daju 2PL modelu.

#### *10.4.5.2 Pokazatelji fita za stavke*

Pogledaćemo pokazatelje koje smo koristili i kod prethodnih modela.

```
itemfit(m3p.m) # Ovom naredbom dobijamo podrazumevani S-X2 po proceduri Orlanda i
Tisena
```

```
##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1  3.973       6      0.000  0.680
## 2   it2  6.739       6      0.011  0.346
## 3   it3 10.486       6      0.027  0.106
## 4   it4  2.022       6      0.000  0.918
## 5   it5  6.018       5      0.014  0.304
## 6   it6  2.765       5      0.000  0.736
## 7   it7  5.499       4      0.019  0.240
## 8   it8  6.540       6      0.009  0.365
## 9   it9  2.247       5      0.000  0.814
## 10 it10  3.772       5      0.000  0.583
```

Ni ovde nema stavki koje pokazuju značajan misfit (gledamo $p$ vrednosti i *RMSEA* za S–$\chi^2$ test). Prikazaćemo još samo Stounov hi–kvadrat ($\chi^{2*}$). Već smo rekli da se za njega se pokazalo da ima veću snagu od S–$\chi^2$ i da istovremeno dobro kontroliše grešku tipa I (de Ayala, 2022). Podsećamo da zbog procedure samouzorkovanja izračunavanje može da potraje malo duže i da iz istog razloga pri ponovljenim računanjima $p$ vrednosti neće biti sasvim iste. Takođe, kada postoje $p$ vrednosti koje su bliske graničnoj, dobro je ponoviti izračunavanje nekoliko puta da bismo bili sigurniji u zaključke o misfitu.

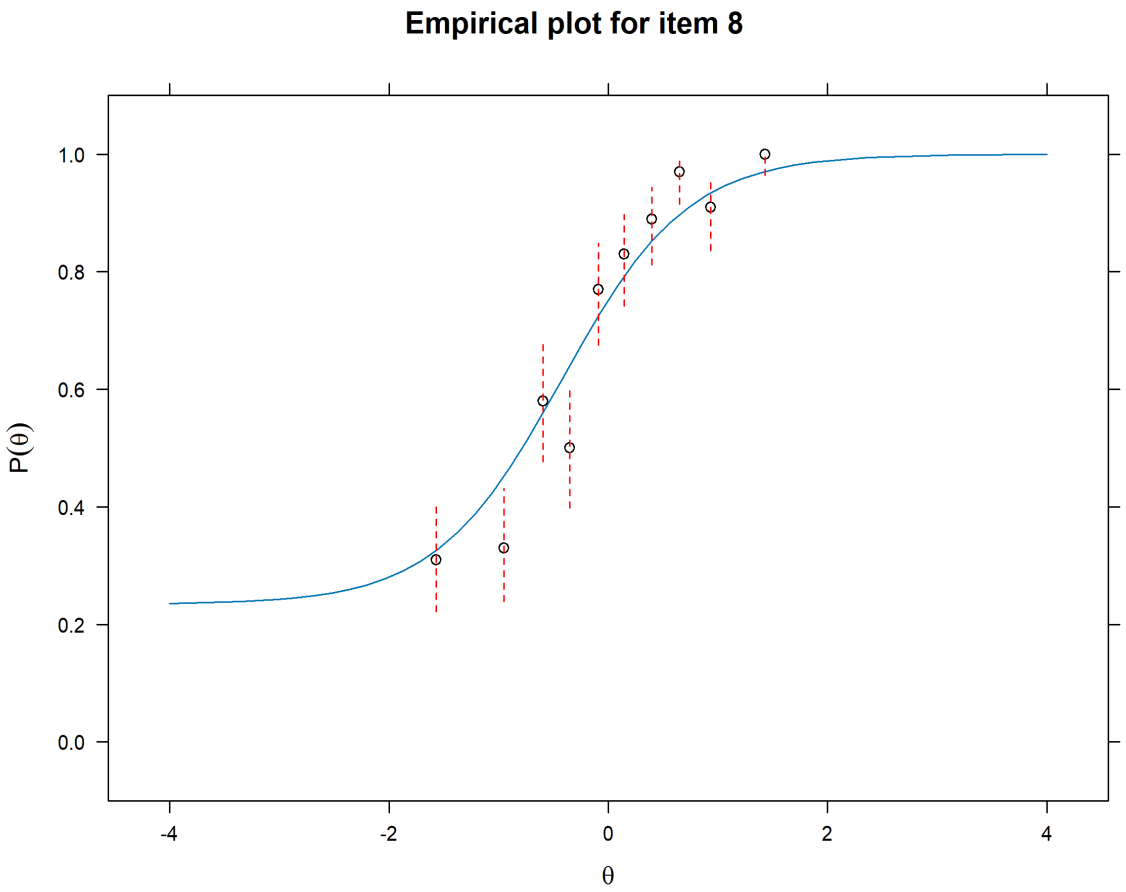
```
itemfit(m2p.m,  fit_stats=c("X2*"), return.tables = FALSE)
```

```
##    item X2_star p.X2_star
## 1   it1   0.405     0.732
## 2   it2   0.484     0.408
## 3   it3   0.391     0.633
## 4   it4   0.474     0.539
## 5   it5   0.651     0.561
```

```
## 6   it6   1.027     0.267
## 7   it7   0.551     0.945
## 8   it8   1.038     0.241
## 9   it9   0.362     0.916
## 10 it10   0.231     0.962
```

Nema misfitujućih stavki, a nema potrebe ni za ponavljanjem izračunavanja jer su *p* vrednosti dovoljno daleko od kritične. Prikazaćemo empirijsku i modelsku karakterističnu krivu stavke 8. Ona nam je interesantna zbog relativno visokog parametra pogađanja i najviše vrednosti pokazatelja misfita (mada ne značajne).

```
itemfit(m3p.m, empirical.plot=8)
```




Slika 63 – Modelska i empirijska karakteristična kriva stavke 8

Modelska kriva u najvećem delu kontinuuma osobine sledi empirijsku (Slika 63). Značajnija odstupanja vidimo kod grupa sa nivoom osobine oko –1 i –0,4 logita gde modelska kriva precenjuje verovatnoću tačnog odgovora.

#### *10.4.5.3 Pokazatelji fita za ispitanike*

Prikazaćemo pokazatelje fita za ispitanike. Ograničićemo se na pokazatelj $Z_h$.

```
PF.m <- personfit(m3p.m, method="EAP", fit_stats="Zh")
# Zbog prostora smo prikazali samo ispitanike kod kojih je apsolutna vrednost Zh s
```

```
tatistika veća od 1,96, što znači da postoji misfit
PF.m[abs(PF.m$Zh)>1.96,]

##       outfit  z.outfit    infit  z.infit        Zh
## 67  2.124929 1.6916687 1.767967 1.953793 -2.187496
## 117 2.708205 2.3085381 1.610630 1.682613 -2.201549
## 200 5.478055 3.0299875 1.586361 1.728486 -2.817415
## 297 2.778443 2.3599759 1.510238 1.471336 -2.024305
## 309 2.917025 2.4958563 1.814274 2.122991 -2.744393
## 322 9.231293 2.9044653 1.622458 1.122790 -2.094202
## 374 3.637783 3.0711391 2.309311 3.006619 -4.193201
## 384 2.917025 2.4958563 1.814274 2.122991 -2.744393
## 399 2.782834 1.9465339 2.088458 2.233018 -2.773462
## 408 2.616110 2.1867991 1.798627 2.024614 -2.514767
## 462 2.450679 1.9233507 1.670653 1.683313 -2.090595
## 488 2.645747 2.1603832 1.937599 2.241884 -2.762949
## 519 3.022985 2.2440850 1.656725 1.577797 -2.174745
## 585 3.789320 1.8414256 1.999260 1.674664 -2.236972
## 602 2.557513 1.7557700 1.723461 1.617363 -2.036555
## 606 2.439555 1.6247662 1.588497 1.728572 -2.174801
## 640 2.439555 1.6247662 1.588497 1.728572 -2.174801
## 702 2.018355 1.5910628 1.794107 2.107633 -2.386832
## 718 2.050510 1.5075134 1.654853 1.846711 -2.167722
## 735 2.557513 1.7557700 1.723461 1.617363 -2.036555
## 770 3.576204 2.7231146 1.489763 1.461013 -2.243118
## 774 5.643558 2.9832876 1.354432 1.140951 -2.210765
## 926 1.594333 0.9514037 1.817961 2.237396 -2.243357
## 981 3.115185 2.0040632 2.366787 2.498136 -3.293247

# Ako želite da vidite ispitanike koji imaju misfit u vidu prigušenog šuma
PF.m[PF.m$Zh>1.96,]

## [1] outfit   z.outfit infit    z.infit  Zh
## <0 rows> (or 0-length row.names)

# Ako želite da vidite ispitanike koji imaju misfit u vidu šuma
PF.m[PF.m$Zh<(-1.96),]

##       outfit  z.outfit    infit  z.infit        Zh
## 67  2.124929 1.6916687 1.767967 1.953793 -2.187496
## 117 2.708205 2.3085381 1.610630 1.682613 -2.201549
## 200 5.478055 3.0299875 1.586361 1.728486 -2.817415
## 297 2.778443 2.3599759 1.510238 1.471336 -2.024305
## 309 2.917025 2.4958563 1.814274 2.122991 -2.744393
## 322 9.231293 2.9044653 1.622458 1.122790 -2.094202
## 374 3.637783 3.0711391 2.309311 3.006619 -4.193201
## 384 2.917025 2.4958563 1.814274 2.122991 -2.744393
## 399 2.782834 1.9465339 2.088458 2.233018 -2.773462
## 408 2.616110 2.1867991 1.798627 2.024614 -2.514767
## 462 2.450679 1.9233507 1.670653 1.683313 -2.090595
## 488 2.645747 2.1603832 1.937599 2.241884 -2.762949
## 519 3.022985 2.2440850 1.656725 1.577797 -2.174745
## 585 3.789320 1.8414256 1.999260 1.674664 -2.236972
## 602 2.557513 1.7557700 1.723461 1.617363 -2.036555
## 606 2.439555 1.6247662 1.588497 1.728572 -2.174801
## 640 2.439555 1.6247662 1.588497 1.728572 -2.174801
## 702 2.018355 1.5910628 1.794107 2.107633 -2.386832
## 718 2.050510 1.5075134 1.654853 1.846711 -2.167722
## 735 2.557513 1.7557700 1.723461 1.617363 -2.036555
```

```
## 770 3.576204 2.7231146 1.489763 1.461013 -2.243118
## 774 5.643558 2.9832876 1.354432 1.140951 -2.210765
## 926 1.594333 0.9514037 1.817961 2.237396 -2.243357
## 981 3.115185 2.0040632 2.366787 2.498136 -3.293247
```

Zbog prostora smo prikazali samo ispitanike kod kojih je apsolutna vrednost $Z_h$ statistika veća od 1,96, što znači da postoji misfit.

Kod svih ispitanika $Z_h$ je negativan što ukazuje na šum, odnosno, na nepredvidivo odgovaranje. Kada smo zatražili tabelu sa prigušenim šumom dobili smo ispis `<0 rows> (or 0-length row.names)`, što znači da takvi ispitanici ne postoje.

#### *10.4.5.4 Lokalna zavisnost*

Proveru lokalne zavisnosti obavićemo na ranije prikazane načine.

Hi–kvadrat za parove stavki (argument *type="LD"*).

```
residuals(m3p.m, type="LD")

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.085  -0.014   0.004   0.000   0.013   0.047
##
##        it1    it2   it3    it4    it5    it6    it7    it8    it9   it10
## it1     NA -0.035 0.014  0.029 -0.005 -0.015 -0.016 -0.002 -0.002  0.033
## it2  1.213     NA 0.010 -0.009  0.008  0.047  0.041  0.012 -0.004 -0.050
## it3  0.201  0.105    NA -0.021  0.007 -0.011  0.004  0.010 -0.032  0.012
## it4  0.865  0.083 0.444     NA  0.022  0.013  0.005 -0.005  0.002 -0.013
## it5  0.027  0.068 0.051  0.503     NA  0.018 -0.085 -0.029  0.017  0.006
## it6  0.216  2.179 0.116  0.179  0.307     NA -0.031 -0.012 -0.020  0.004
## it7  0.262  1.653 0.015  0.028  7.178  0.992     NA -0.014  0.020  0.004
## it8  0.005  0.136 0.100  0.025  0.859  0.154  0.199     NA  0.043 -0.014
## it9  0.003  0.016 1.009  0.004  0.303  0.403  0.412  1.828     NA  0.024
## it10 1.058  2.542 0.138  0.165  0.041  0.018  0.013  0.197  0.555     NA

residuals(m2p.m, type=c("LD"), df.p=T, suppres=.20)

## Degrees of freedom (lower triangle) and p-values:
##
##      it1   it2   it3   it4   it5   it6   it7   it8   it9  it10
## it1   NA 0.285 0.634 0.354 0.910 0.679 0.615 0.933 0.966 0.297
## it2    1    NA 0.715 0.782 0.750 0.123 0.195 0.527 0.957 0.116
## it3    1 1.000    NA 0.512 0.788 0.777 0.900 0.638 0.324 0.683
## it4    1 1.000 1.000    NA 0.460 0.650 0.887 0.958 0.970 0.676
## it5    1 1.000 1.000 1.000    NA 0.518 0.008 0.253 0.657 0.804
## it6    1 1.000 1.000 1.000 1.000    NA 0.334 0.571 0.469 0.856
## it7    1 1.000 1.000 1.000 1.000 1.000    NA 0.910 0.479 0.913
## it8    1 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.306 0.976
## it9    1 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA 0.394
## it10   1 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000    NA
##
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.084  -0.013   0.003   0.001   0.015   0.049
##
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Prema ovom pokazatelju na našim podacima ne postoji lokalna zavisnost (u tabeli *upper triangle summary* nema vrednosti većih od 0,20 pa su sve zamenjene sa NA zbog argumenta *suppres*). Preskočićemo $G^2$ pokazatelj i prikazati $Q_3$. Dobijamo ga korišćenjem argumenta *type="Q3"*.

```
residuals(m3p.m, type="Q3", suppres=.2)
```

```
## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.225  -0.110  -0.092  -0.089  -0.055  -0.010
##
##      it1    it2 it3 it4 it5 it6 it7 it8 it9   it10
## it1    1
## it2       1.000                              -0.225
## it3               1
## it4                   1
## it5                       1
## it6                           1
## it7                               1
## it8                                   1
## it9                                       1
## it10     -0.225                                1.000
```

Kada kao graničnu vrednost koristimo 0,20 izdvaja se par stavki 2 i 10. Međutim, ako $Q_3$ pokazatelje poredimo sa prosečnom vrednošću u matrici, nema indicija za postojanje lokalne zavisnosti (videti odeljke 9.2 i 10.4.4.2.3).

```
Q3m <- as.matrix(residuals(m3p.m, type="Q3"))
```

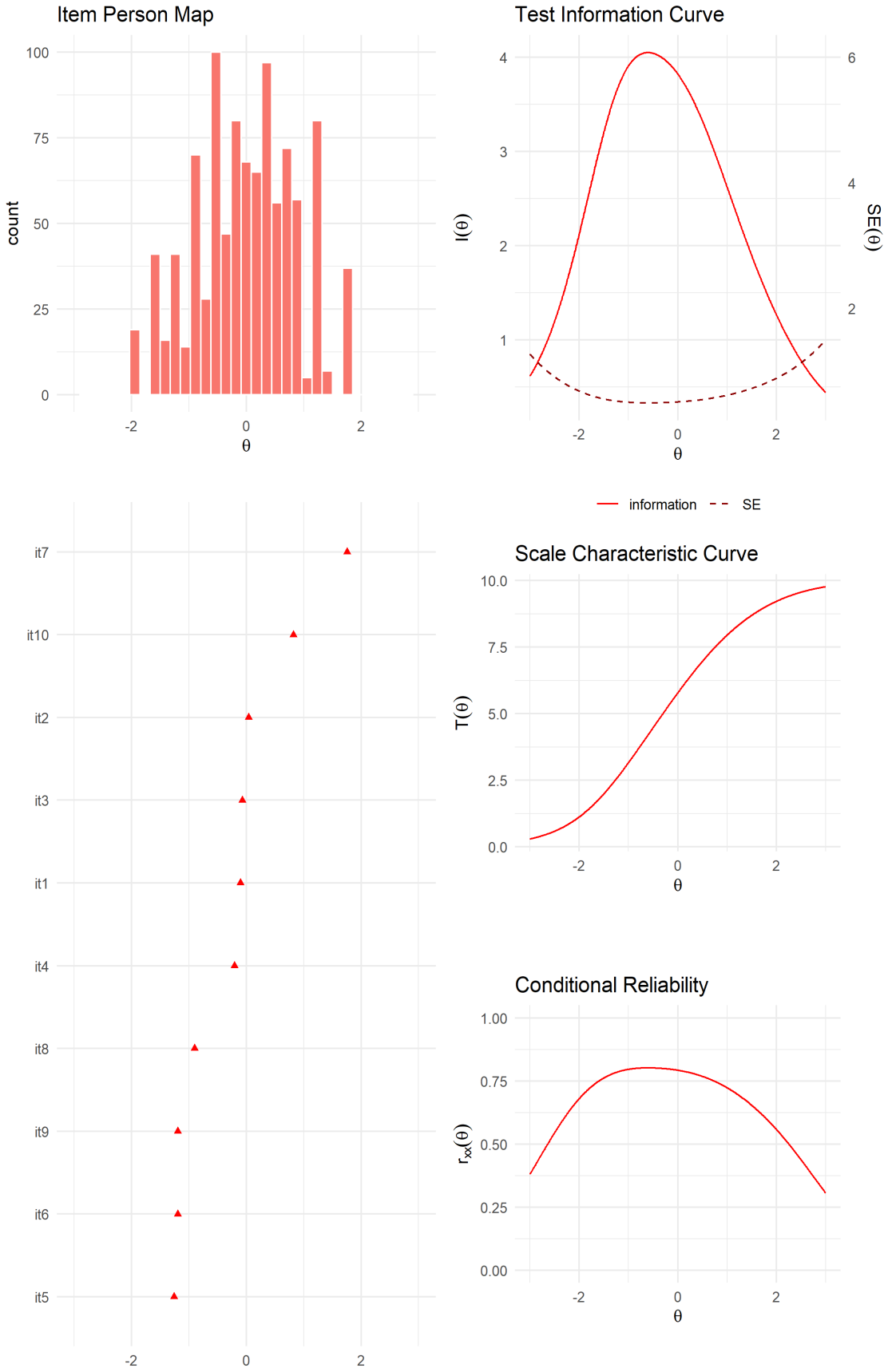
```
## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.225  -0.110  -0.092  -0.089  -0.055  -0.010
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.158 -0.065 -0.042 -0.081 -0.094 -0.070 -0.074 -0.086 -0.041
## it2  -0.158  1.000 -0.122 -0.144 -0.092 -0.035 -0.023 -0.100 -0.122 -0.225
## it3  -0.065 -0.122  1.000 -0.134 -0.075 -0.100 -0.063 -0.077 -0.140 -0.101
## it4  -0.042 -0.144 -0.134  1.000 -0.061 -0.074 -0.051 -0.098 -0.107 -0.121
## it5  -0.081 -0.092 -0.075 -0.061  1.000 -0.095 -0.102 -0.134 -0.111 -0.053
## it6  -0.094 -0.035 -0.100 -0.074 -0.095  1.000 -0.055 -0.109 -0.183 -0.052
## it7  -0.070 -0.023 -0.063 -0.051 -0.102 -0.055  1.000 -0.051 -0.010 -0.110
## it8  -0.074 -0.100 -0.077 -0.098 -0.134 -0.109 -0.051  1.000 -0.036 -0.094
## it9  -0.086 -0.122 -0.140 -0.107 -0.111 -0.183 -0.010 -0.036  1.000 -0.035
## it10 -0.041 -0.225 -0.101 -0.121 -0.053 -0.052 -0.110 -0.094 -0.035  1.000
```

```
Q3m[Q3m==1] <- NA
#Q3m <- Q3m[upper.tri(Q3m)]
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", Q3mean, "\n\n")

##
## Prosečna vrednost Q3: -0.08904057

Q3c <- Q3m-Q3mean
Q3c

##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.069  0.024  0.047  0.008 -0.005  0.019  0.015  0.003  0.048
## it2  -0.069     NA -0.033 -0.055 -0.003  0.054  0.066 -0.011 -0.033 -0.136
## it3   0.024 -0.033     NA -0.045  0.014 -0.011  0.027  0.012 -0.051 -0.012
## it4   0.047 -0.055 -0.045     NA  0.028  0.015  0.038 -0.009 -0.018 -0.032
## it5   0.008 -0.003  0.014  0.028     NA -0.006 -0.013 -0.045 -0.022  0.036
## it6  -0.005  0.054 -0.011  0.015 -0.006     NA  0.034 -0.020 -0.094  0.037
## it7   0.019  0.066  0.027  0.038 -0.013  0.034     NA  0.038  0.079 -0.021
## it8   0.015 -0.011  0.012 -0.009 -0.045 -0.020  0.038     NA  0.053 -0.005
## it9   0.003 -0.033 -0.051 -0.018 -0.022 -0.094  0.079  0.053     NA  0.054
## it10  0.048 -0.136 -0.012 -0.032  0.036  0.037 -0.021 -0.005  0.054     NA

Q3c[Q3c<.20] <- NA
Q3c

##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

#### *10.4.5.5 Grafički prikaz*

Na početku prikazaćemo grafički rezime modela.

```
library(ggmirt)
p_load(ggplot2)
p_load(RColorBrewer)
summaryPlot(m2p.m, theta_range = c(-3, 3), adj_factor = 1.5)
```

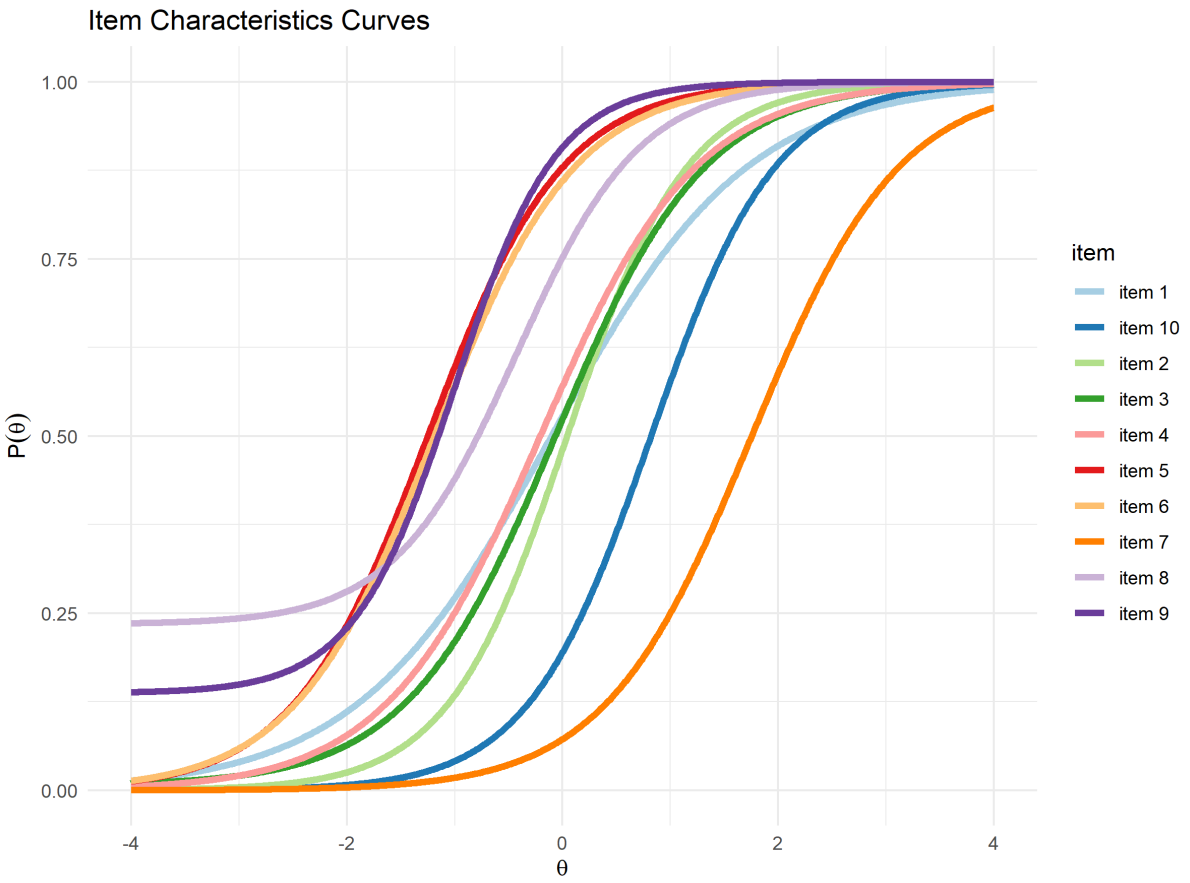



Slika 64 – Dijagnostički grafikon testa (3PL model)

Na prethodnom grafikonu (Slika 64) prikazana je mapa ispitanika i stavki. Na osnovu nje možemo videti da su stavkama bolje pokriveni ispitanici čiji su nivoi osobine prosečni ili ispod proseka. U natprosečnom delu kontinuuma osobine manje je stavki pa će i mere tih ispitanika biti manje precizno procenjene. To nam pokazuju i krive informativnosti i uslovne pouzdanosti (Slika 64). Informativnost i pouzdanost su najveće na nivou osobine ispod proseka. Na kom tačno nivou videćemo nešto kasnije kada izračunamo numeričke pokazatelje.

Pogledaćemo karakteristične krive stavki. One se sada razlikuju i po nagibu i po donjoj asimptoti (Slika 65).

```
p_load(RColorBrewer) # Nije potrebno ako je paket ranije učitan
ggmirt::tracePlot(m3p.m, facet = FALSE, legend = TRUE)+
 ggplot2::scale_color_brewer(palette = "Paired") + ggplot2::geom_line(size = 1.5)

## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.
```



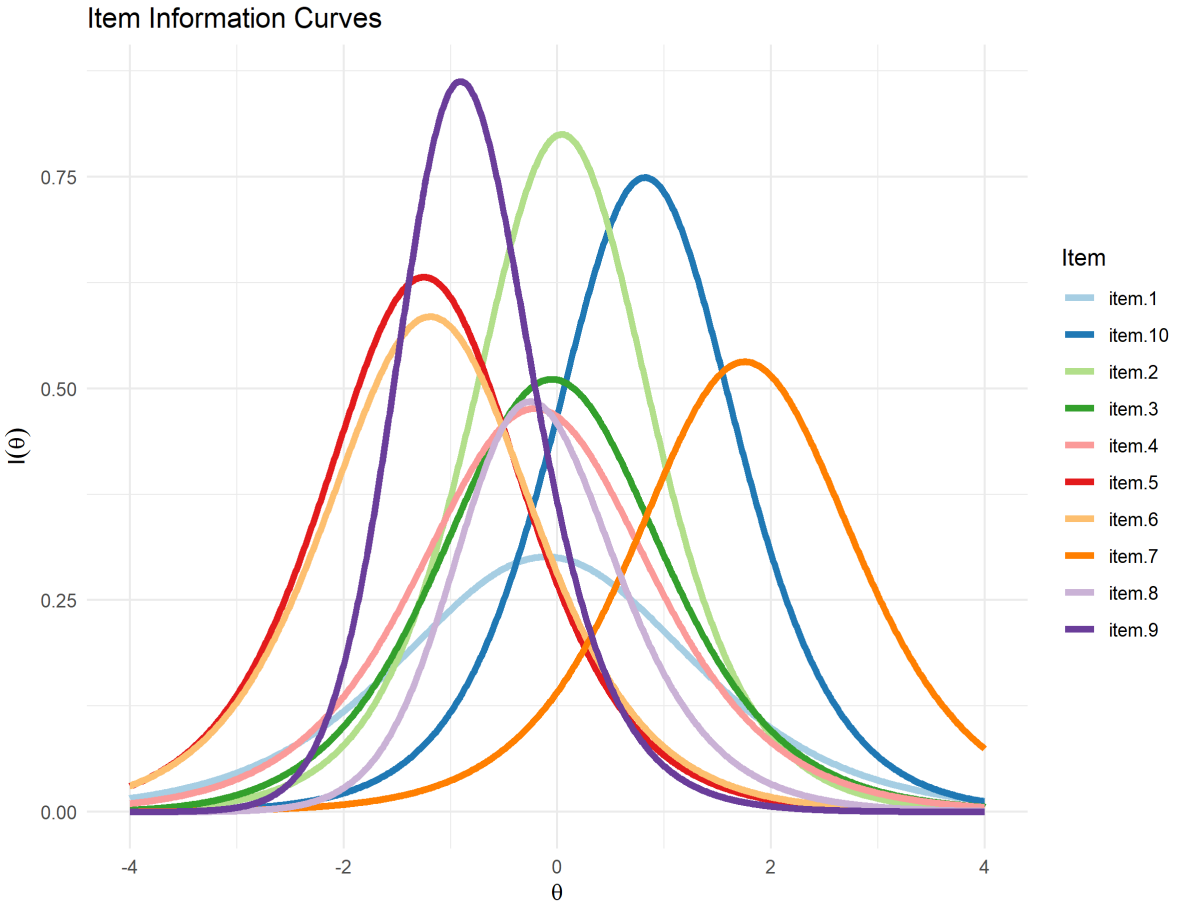

Slika 65 – Karakteristične krive stavki u 3PL modelu (paketi *mirt* i *ggmirt*)

I dalje najmanji nagib ima kriva stavke 1, a najveći kriva stavke 9 (Slika 65). Ovaj nagib (stavka 9) je sada viši nego u 2PL modelu (sada je 2,13, a bio je 1,88). Verovatnoća tačnog odgovora na ovu stavku nikada nije niža od 15% bez obzira na nivo osobine. Do nivoa osobine oko –3 logita verovatnoća je tolika a onda se menja u zavisnosti od nivoa osobine. Slična situacija je sa krivom stavke 8. Najniža verovatnoća

tačnog odgovora na ovu stavku je 24%.

#### *10.4.5.6 Informativnost*

```
ggmirt::itemInfoPlot(m3p.m, legend = TRUE, facet = FALSE) +
 ggplot2::scale_color_brewer(palette = "Paired") + ggplot2::geom_line(size = 1.5)
```

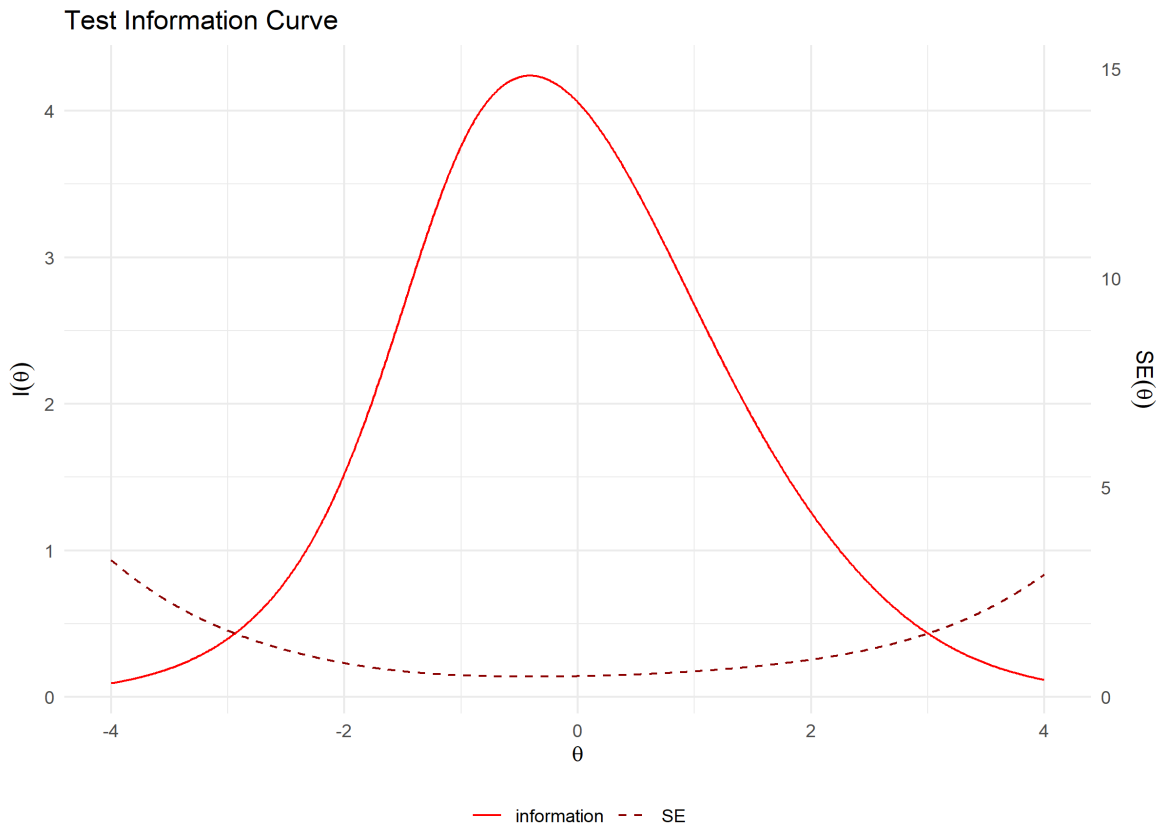



Slika 66 – Funkcije informativnosti stavki u 3PL modelu

Najinformativnije stavke su i dalje stavke 9, 2 i 10 koje pokrivaju različite nivoe latentne osobine (Slika 66). Ipak, viši nivoi osobine pokriveni su manjim brojem stavki koje uz to imaju i nešto nižu informativnost.

Ukoliko želimo da primenjujemo test na opštoj populaciji, bilo bi dobro povećati broj nešto težih stavki. Podsećamo da se informativnosti stavki sumiraju po nivoima osobine da bi dale informativnost testa (Slika 67). Ovakav kakav jeste, test preciznije meri ispitanike sa ispodprosečnim nivoom osobine.

```
testInfoPlot(m3p.m)
```




Slika 67 – Krive informativnosti i standardne greške testa

Informativnost testa (Slika 67) najveća je na nivou osobine od –0,4 logita. To vidimo iz narednih izračunavanja.

Koristimo funkciju *areainfo()*.

```
# Za ceo test
M3PI <- areainfo(m3p.m, c(-3,3))
M3PI

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 13.77663  14.37964   0.958065     10

# Samo za određene stavke
M3PI_1 <- areainfo(m3p.m, c(-3,3), which.items = c(2:10)) # Bez stavke 1
M3PI_1

##  LowerBound UpperBound    Info TotalInfo Proportion nitems
##          -3          3 12.7582  13.28263  0.9605175      9

cat("Procenat informativnosti koji ostaje u opsegu od -3 do 3 logita kada izbacimo stavku 1:")

## Procenat informativnosti koji ostaje u opsegu od -3 do 3 logita kada izbacimo s
tavku 1:

round(M3PI_1[,3]/M3PI[,3]*100,2)

## [1] 92.61
```

Informativnost po nivoima merene osobine:

```
# Prvo definišemo vrednosti latentne osobine za koje želimo da izračunamo
informativnost (sekvenca u rasponu od -3 do 3 sa koracima od 0,1)
thetaM <- seq(-3,3, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(m3p.m, Theta = thetaM)
# Prikaz prvih 6 redova (uklonite head() za ispis cele tabele)
head(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
```

```
##      nivo osobine informativnost
## [1,]         -3.0      0.3981289
## [2,]         -2.9      0.4583505
## [3,]         -2.8      0.5270840
## [4,]         -2.7      0.6053891
## [5,]         -2.6      0.6944233
## [6,]         -2.5      0.7954330
```

```
# Maksimalna informativnost
cat("Maksimalna informativnost:", max(infoM), "\n")
```

```
## Maksimalna informativnost: 4.239889
```

```
# Ispis nivoa osobine na kom je informativnost maksimalna
cat("Nivo osobine na kojem je informativnost maksimalna:",
 thetaM[infoM==max(infoM)], "\n")
```

```
## Nivo osobine na kojem je informativnost maksimalna: -0.4
```

```
# Ispis raspona u kojem je informativnost veća od 3,33 što odgovara pouzdanosti od
0,7
cat("Nivoi osobine na kojima je informativnost preko 3,33: od",
 min(thetaM[infoM>3.33]), "do", max(thetaM[infoM>3.33]), "logita")
```

```
## Nivoi osobine na kojima je informativnost preko 3,33: od -1.2 do 0.5 logita
```

Vidimo da je maksimalna informativnost testa 4,29 na nivou osobine od –0,4 logita. Informativnost je nešto viša nego u 2PL modelu (neznatno) i na nešto višem nivou osobine (odeljak 10.4.4.3).

```
# Možemo izračunati pouzdanost po nivoima osobine na osnovu poznatog odnosa sa
informativnošću rtt=1-(1/infoM)
rtt <- ifelse(!(1-(1/infoM))<0, 1-(1/infoM), 0)
cat("Raspon nivoa osobine u kojem je pouzdanost ≥0,7:\n")
```

```
## Raspon nivoa osobine u kojem je pouzdanost ≥0,7:
```

```
thetaM[rtt>.7]
```

```
##  [1] -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2
## [16]  0.3  0.4  0.5
```

```
# Možemo izračunati standardnu grešku merenja po nivoima osobine na osnovu
poznatog odnosa sa informativnošću se=1/sqrt(infoM)
se <- 1/sqrt(infoM)
cat("Raspon nivoa osobine u kojem je SE <0,548:\n")
```

```
## Raspon nivoa osobine u kojem je SE <0,548:

thetaM[se<.548]

##  [1] -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2
## [16]  0.3  0.4  0.5

head(cbind("nivo osobine"=thetaM, "informativnost"=round(infoM, 3),
"pouzdanost"=round(rtt,3), "sgm"=round(se, 3)))

##      nivo osobine informativnost pouzdanost   sgm
## [1,]         -3.0          0.398          0 1.585
## [2,]         -2.9          0.458          0 1.477
## [3,]         -2.8          0.527          0 1.377
## [4,]         -2.7          0.605          0 1.285
## [5,]         -2.6          0.694          0 1.200
## [6,]         -2.5          0.795          0 1.121
```

Prihvatljivu informativnost definisali smo kao informativnost preko 3,33 (što odgovara pouzdanosti preko 0,70). Raspon u kojem je naš test dovoljno precizan je između –1,2 i 0,5 logita. Ako je to raspon osobine u kojem želimo da vršimo precizno merenje, možemo se zadovoljiti time. Ako to nije slučaj, trebalo bi dodavati stavke koje su informativne izvan ovog raspona. Kao što smo videli, merenje nije jednako precizno na svim nivoima osobine.

Izračunaćemo empirijsku i marginalnu pouzdanost. Podsećamo, za računanje pouzdanosti potrebne su nam procene nivoa osobine ispitanika i standardne greške vezane za njih.

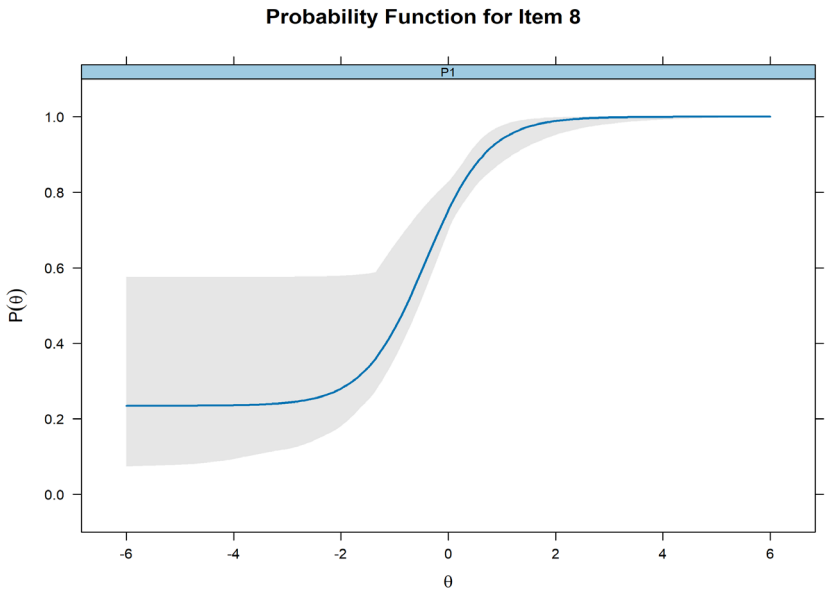
```
# Parametre ispitanika i standardne greške dobijamo funkcijom fscores i smeštamo u
objekat THETA_SE koji prosleđujemo funkciji za računanje empirijske pouzdanosti
THETA_SE <- fscores(m3p.m, full.scores.SE = TRUE)
head(THETA_SE)

##              F1     SE_F1
## [1,] -0.5003921 0.4655923
## [2,] -0.4574367 0.4633830
## [3,]  0.1793770 0.4648576
## [4,] -0.9130737 0.4406089
## [5,] -0.3326846 0.4500345
## [6,]  0.4742148 0.4810067

cat("\nEmpirijska pouzdanost:", empirical_rxx(THETA_SE))

##
## Empirijska pouzdanost: 0.7576106

cat("\nMarginalna pouzdanost:", marginal_rxx(m3p.m))

##
## Marginalna pouzdanost: 0.7526228

cat("\nKronbahova alfa:", psych::alpha(b.pod)$total$raw_alpha)

##
## Kronbahova alfa: 0.7572566
```

Empirijska pouzdanost se ne razlikuje bitno od one iz 2PL modela (tek na trećoj decimali).

Pogledaćemo grafikon karakteristične krive stavke 8 (Slika 68), koja ima najveću vrednost $\chi^{2*}$ (iako

ne značajnu). To je ujedno stavka sa najvećim parametrom pogađanja i druga po diskriminativnosti.

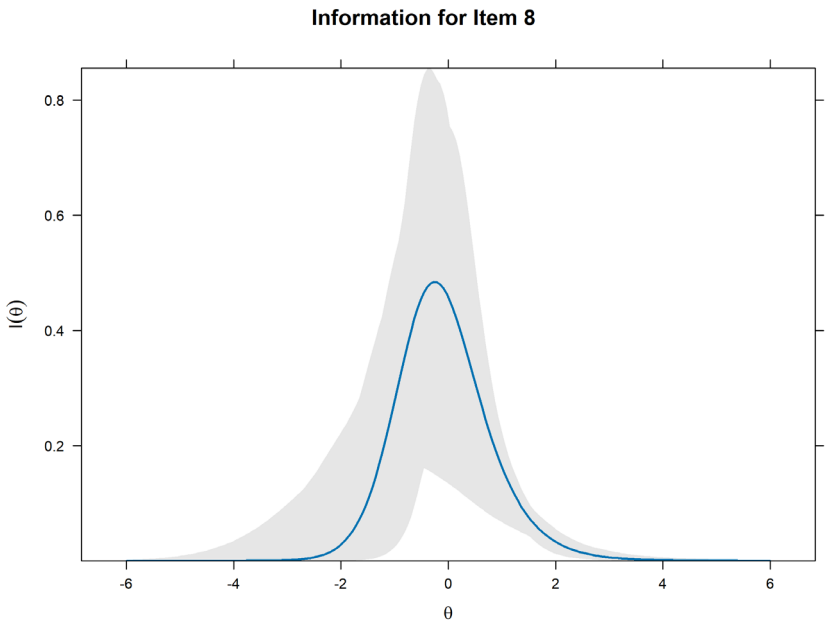
```
itemplot(m3p.m, item=8, lwd=2, type="trace", CE=T, CEalpha=0.05, CEdraws=1000)
```




Slika 68 – Karekteristična kriva stavke 8 sa regionom poverenja

Upadljivo je da je region poverenja najširi u oblasti u kojoj je verovatnoća tačnog odgovora procenjena na nivou donje asimptote (Slika 68). Isto tako, primećujemo da je i region poverenja funkcije informativnosti značajno širi u delu kontinuuma latentne osobine gde verovatnoća tačnog odgovora ne zavisi od nivoa osobine (Slika 69).

```
# Informativnost stavke
itemplot(m3p.m, item=8, lwd=2, type="info", CE=T, CEalpha=0.05, CEdraws=1000)
```

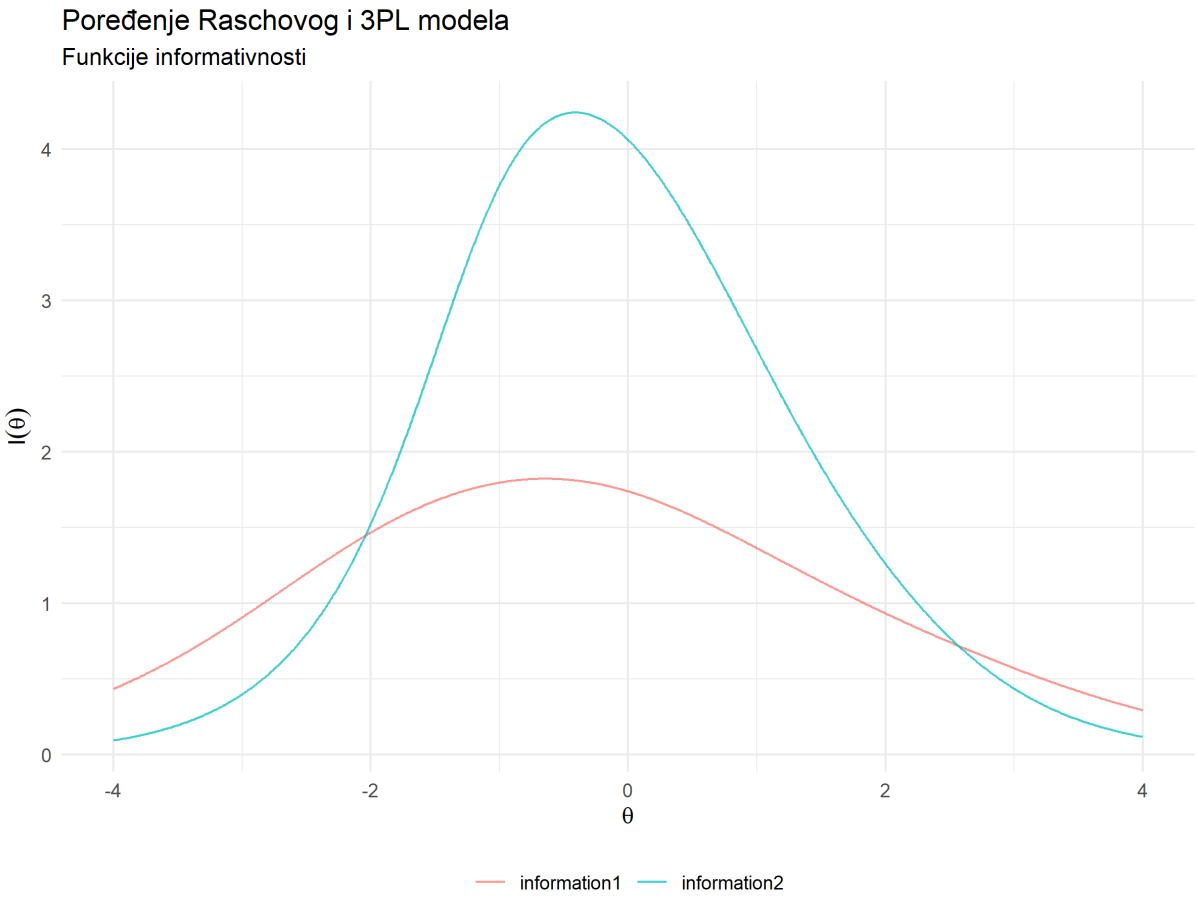



Slika 69 – Funkcija informativnosti stavke 8 sa regionom poverenja

#### *10.4.5.7 Poređenje informativnosti 3 modela*

Za kraj ćemo uporediti funkcije informativnosti tri modela (Rašovog, 2PL i 3PL). Koristićemo grafičko poređenje i funkciju *testInfoCompare()* iz paketa *ggmirt*.

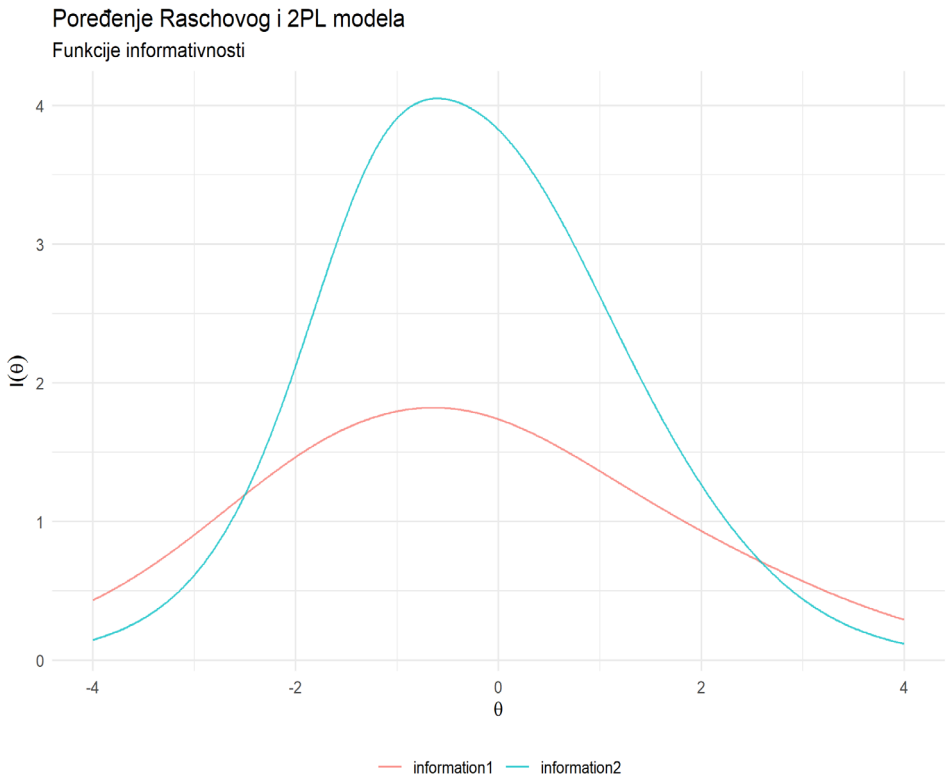
```
ggmirt::testInfoCompare(mr.m, m3p.m, title = "Poređenje Raschovog i 3PL modela",
 subtitle="Funkcije informativnosti")
```




Slika 70 – Poređenje funkcija informativnosti Raschovog i 3PL modela

Kada uporedimo 3PL i Rašov model vidimo da je generalno viša informativnost 3PL modela, osim u ekstremnim oblastima nivoa osobine gde je Rašov model informativniji (Slika 70).

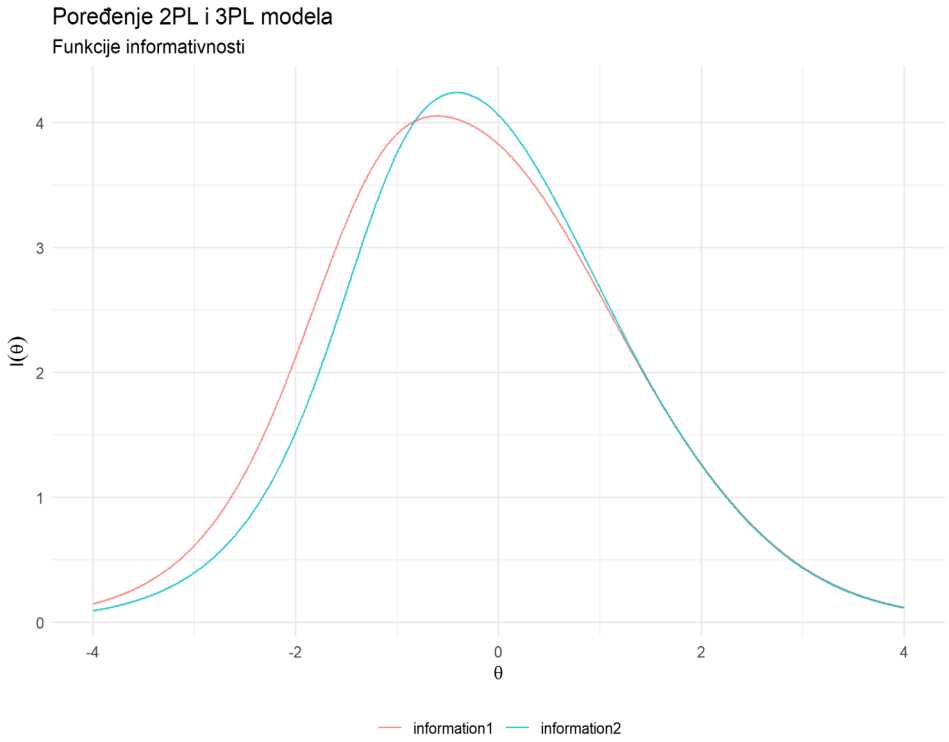
```
ggmirt::testInfoCompare(mr.m, m2p.m, title = "Poređenje Raschovog i 2PL modela",
 subtitle="Funkcije informativnosti" )
```




Slika 71 – Poređenje funkcija informativnosti Raschovog i 2PL modela

Isto primećujemo i kada poredimo 2PL i Rašov model (Slika 71). Rašov model je nešto informativniji u oblastima ekstremnih skorova, ali je 2PL značajno informativniji u oblasti od –2,5 do 2,5 logita.

```
ggmirt::testInfoCompare(m2p.m, m3p.m, title = "Poređenje 2PL i 3PL modela",
 subtitle="Funkcije informativnosti")
```




Slika 72 – Poređenje funkcija informativnosti 2PL i 3PL modela

Između funkcija informativnosti 2PL i 3PL modela razlika je vrlo mala (Slika 72), a kako ni LR test ne ukazuje na značajnu razliku između ovih modela, nema potrebe da koristimo kompleksniji 3PL model. Uporedićemo i numeričke pokazatelje informativnosti, ukupne i u rasponu od –3 do 3 logita (to je raspon u kojem možemo očekivati najveći broj ispitanika).

Pogledaćemo prvo korelacije mera ispitanika dobijenih na osnovu tri modela.

```
# Korelacije mera ispitanika iz tri modela
UMM <- data.frame(cbind(fscores(mr.m), fscores(m2p.m), fscores(m3p.m)))
names(UMM) <- c("mRasch", "m2PL", "m3PL")
cat("Pirsonove korelacije mera dobijenih na osnovu tri modela:\n")

## Pirsonove korelacije mera dobijenih na osnovu tri modela:

round(cor(UMM),3)

##        mRasch  m2PL  m3PL
## mRasch  1.000 0.997 0.997
## m2PL    0.997 1.000 0.999
## m3PL    0.997 0.999 1.000

cat("Spirmanove korelacije mera dobijenih na osnovu tri modela:\n")

## Spirmanove korelacije mera dobijenih na osnovu tri modela:

round(cor(UMM, method = "spearman"),3)

##        mRasch  m2PL  m3PL
## mRasch  1.000 0.986 0.987
## m2PL    0.986 1.000 0.999
## m3PL    0.987 0.999 1.000
```

Vidimo da su Pirsonove korelacije vrlo visoke – veće između višeparametarskih modela. Spirmanove korelacije višeparametarskih modela i Rašovog modela su takođe visoke, ali neznatno niže nego Pirsonove.

```
UIM <- data.frame(rbind(MRI[,3:4], M2PI[,3:4], M3PI[,3:4]))
rownames(UIM) <- c("Rasch", "2PL", "3PL")
UIM

##            Info TotalInfo
## Rasch  8.111699  10.00000
## 2PL   14.389995  15.15825
## 3PL   13.776629  14.37964
```

Kada je u pitanju informativnost u opsegu od –3 do 3 logita, vidimo da je ona najveća kod 2PL modela, a isto važi i za informativnost u opsegu od –10 do 10 logita. Pošto LR test ne nalazi statistički značajnu razliku između 2PL i 3PL modela, birajući između ova dva modela mogli bismo se opredeliti za jednostavniji 2PL model, posebno ako nam je cilj da procenimo mere ispitanika i na osnovu ovog test obavimo neki kon-

kretan posao (selekciju ispitanika). Ako sa druge strane želimo merno–teorijski čist model i lakoću skorovanja[54], onda bismo mogli da se opredelimo i za Rašov model s obzirom na to da i on ima prihvatljive pokazatelje fita. U tom slučaju test bi trebalo dopuniti diskriminativnijim stavkama u opsegu nivoa osobine koji nije dovoljno pokriven. Ova poslednja primedba važi i u slučaju da se opredelimo za neki od višeparametarskih modela. Problem sa Rašovim modelom koji ima dobre pokazatelje fita je taj što on može sadržavati jednako nediskriminativne stavke i još uvek imati zadovoljavajuće pokazatelje fita.

```
detach("package:ggmirt", unload = TRUE)
detach("package:mirt", unload = TRUE)
```

## *10.5 IRT modeli za politomne stavke sa uređenim kategorijama*

Za prikaz IRT modela za politomne stavke sa uređenim kategorijama prvo ćemo simulirati podatke.

### 10.5.1 Simulacija podataka u paketu *catIrt*

Simulaciju podataka ćemo ovoga puta obaviti upotrebom paketa *catIrt* (Nydick, 2022). Ovaj paket sadrži funkcije za simulaciju podataka po jednodimenzionalnim IRT modelima i za simulaciju računarskog adaptivnog testiranja. Simulirali smo podatke po modelu stepenovanog odgovaranja (GRM). Funkciji *simIrt* potrebno je proslediti matricu sa *m* redova (*m* je broj stavki), koja će u prvoj koloni sadržati *nagibe.* Odabrali smo slučajan uzorak iz uniformne distribucije u rasponu vrednosti od 1 do 2. U ostalim kolonama bi trebalo da se nalaze parametri težine kategorija (pragovi). Broj kolona sa parametrima težine određuje koliko će simulirane stavke imati kategorija odgovora (imaće za jednu više). Mi smo simulirali dva niza parametara kategorija, pa će stavke imati tri kategorije odgovora. Najniža vrednost kategorije odgovora biće 1, a najviša 3. Za niže parametre kategorija podatke smo simulirali iz uniformne distribucije sa vrednostima između –2 i 0 logita, a za više sa vrednostima između 0 i +2 logita. Da bi podaci bili reproducibilni bitna je vrednost *set.seed* koja se zadaje prilikom simulacije. Ona je u primeru definisana kao promenljiva *ss* na početku koda.

```
#install.packages("catIrt")
# Ako nemate instaliran ovaj paket instalirajte ga
ss <- 123 # Dodeljivanje vrednosti varijabli ss koja služi kao vrednost funkcije
set.seed() radi reproducibilnosti podataka
set.seed(ss) # Ne preskakati izvršenje ove komande
thete <- rnorm(1000, mean=0) # Simuliranje nivoa osobine iz standardne normalne
distribucije za 1000 ispitanika
set.seed(ss)
parametri <- cbind(runif(10, 1, 2), b1 = runif(10, -2, 0), b2 = runif(10, 0, 2))
# Parametri nagiba su simulirani iz uniformne distribucije u rasponu od 1 do 2
# Parametri lokacije su simulirani iz uniformne distribucije u rasponu od -2 do 0
za prvi i od 0 do 2 za drugi parametar
# Sve je smešteno u objekat pod nazivom parametri klase matrix, array
# Pošto smo definisali samo parametre b1 i b2 imaćemo 3 kategorije odgovora (za 1
više od broja parametara lokacije). Kada bismo želeli 4 kategorije odgovora bilo
bi potrebno definisati i parametar b3 itd.
parametri # Ispis matrice sa parametrima za simulaciju
```

[54] Prisetimo se da svi ispitanici sa jednakim sumacionim skorom imaju iste Rašove mere.

```
##                          b1        b2
##  [1,] 1.287578 -0.08633331 1.7790786
##  [2,] 1.788305 -1.09333169 1.3856068
##  [3,] 1.408977 -0.64485873 1.2810136
##  [4,] 1.883017 -0.85473320 1.9885396
##  [5,] 1.940467 -1.79415063 1.3114116
##  [6,] 1.045556 -0.20035006 1.4170609
##  [7,] 1.528105 -1.50782453 1.0881320
##  [8,] 1.892419 -1.91588093 1.1882840
##  [9,] 1.551435 -1.34415856 0.5783195
## [10,] 1.456615 -0.09099270 0.2942273

# Ukoliko želite namerno da postavite neki parametar određene stavke na neku tačno
definisanu vrednost (npr. za vežbu) možete to uraditi na sledeći način
# r <- 2 # Varijabli r dodelite vrednost koja odgovara rednom broju stavke čije
parametre želite da modifikujete
# k <- 1 # Varijabli k dodelite vrednost kolone (parametra) koju želite da
izmenite (1 je nagib, 2 i dalje su parametri kategorija)
# parametri[r, k] <- 0 # Ovako bismo postavili nagib stavke 2 na vrednost 0
# Gore navedene izmene morate odraditi pre sledeće komande
mod <- catIrt::simIrt(params = parametri, mod = "grm", theta = thete)

##
## Graded response model simulation:
##     1000  simulees,  10  items,  3  levels per item

# Prethodnom komandom simuliramo podatke po modelu stepenovanog odgovaranja (grm),
koji su smešteni u objekat mod klase list i u matricu resp u okviru njega
p.pod <- data.frame(mod$resp)
# Izvlačimo podatke u matricu podataka pod nazivom p.pod
names(p.pod) <- paste0("it",1:10)
# Dodeljujemo imena varijablama
head(p.pod,3)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## 1   3   2   2   2   1   1   2   2   2    1
## 2   1   2   2   2   2   1   2   2   2    3
## 3   1   3   3   2   3   2   2   3   3    3

# Prikazujemo prva tri reda matrice podataka

# Deskriptivni pokazatelji

psych::describe(p.pod)

##      vars    n mean   sd median trimmed  mad min max range  skew kurtosis   se
## it1     1 1000 1.67 0.72      2    1.59 1.48   1   3     2  0.59    -0.91 0.02
## it2     2 1000 1.94 0.60      2    1.92 0.00   1   3     2  0.02    -0.22 0.02
## it3     3 1000 1.88 0.72      2    1.85 1.48   1   3     2  0.19    -1.07 0.02
## it4     4 1000 1.80 0.55      2    1.79 0.00   1   3     2 -0.09    -0.13 0.02
## it5     5 1000 2.06 0.51      2    2.08 0.00   1   3     2  0.09     0.70 0.02
## it6     6 1000 1.78 0.78      2    1.72 1.48   1   3     2  0.41    -1.25 0.02
## it7     7 1000 2.10 0.62      2    2.12 0.00   1   3     2 -0.07    -0.45 0.02
## it8     8 1000 2.11 0.51      2    2.11 0.00   1   3     2  0.15     0.55 0.02
## it9     9 1000 2.16 0.68      2    2.20 0.00   1   3     2 -0.21    -0.86 0.02
## it10   10 1000 1.94 0.95      2    1.92 1.48   1   3     2  0.12    -1.88 0.03
```

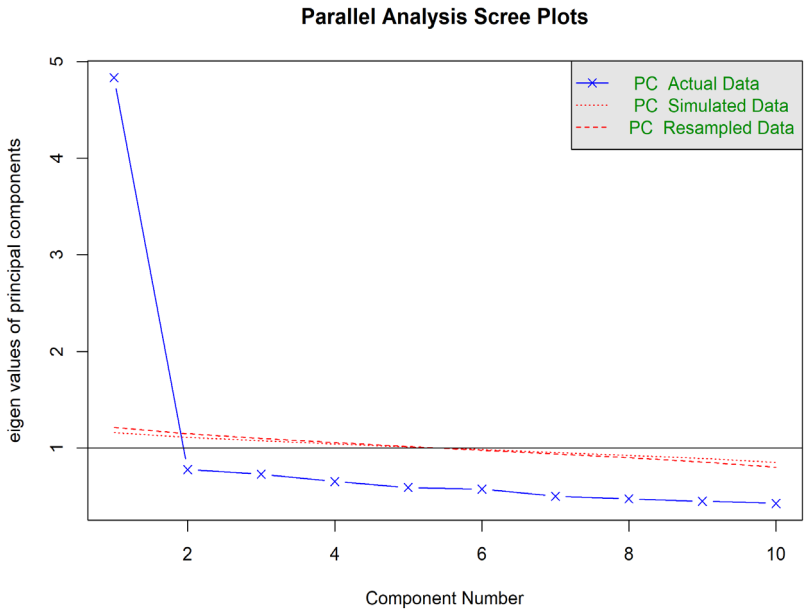
### 10.5.2 Provera pretpostavki modela

Opet ćemo prvo proveriti da li su zadovoljene pretpostavke modela, koje su iste kao kod modela za binarno skorovane stavke.

#### *10.5.2.1 Jednodimenzionalnost*

Primenićemo paralelnu analizu (Horn, 1965), kriterijum vrlo jednostavne strukture (VSS) (Revelle & Rocklin, 1979), Veliserov (Velicer, 1976) kriterijum minimalne prosečne parcijalne korelacije (Minimum Average Partial – MAP) i BIC, svi dostupni u paketu *psych* (Revelle, 2017).

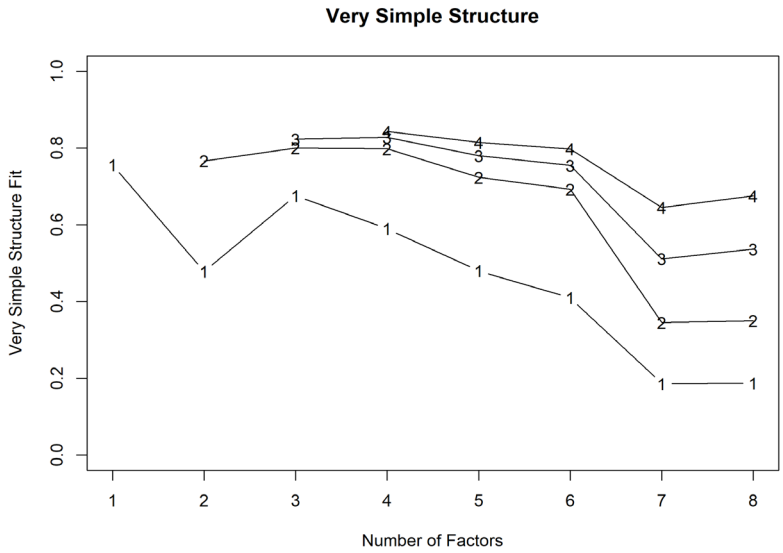
```
p_load(psych)
PA <- fa.parallel(p.pod, n.iter = 1000,cor="poly", quant=.95, fa="pc")
```




Slika 73 – Skri–dijagram paralelne analize

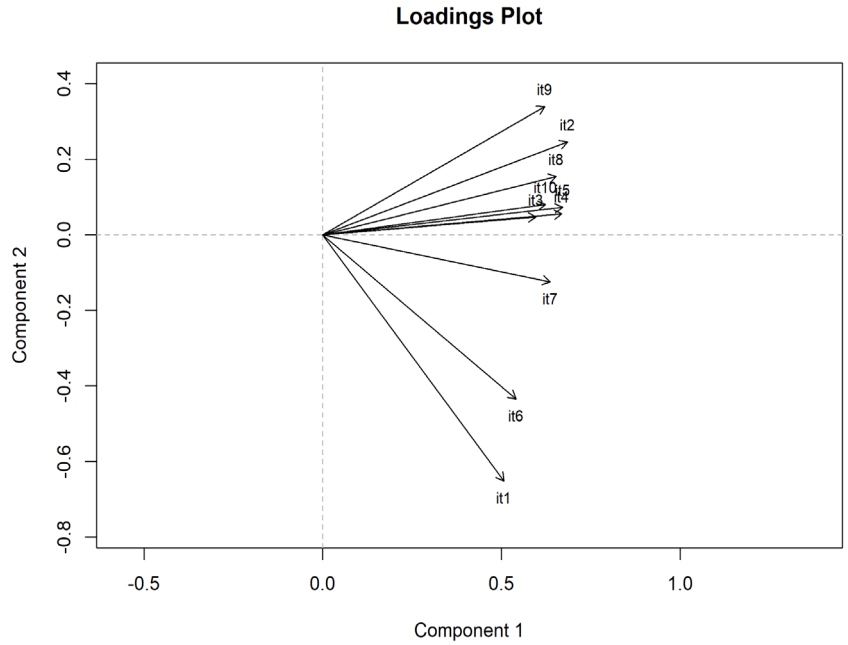
```
## Parallel analysis suggests that the number of factors =  NA  and the number of
components =  1

# Kao broj slučajeva se postavlja broj redova matrice podataka (ako nema
nedostajućih podataka)
VSS <- vss(p.pod, n.obs = nrow(p.pod))

summary(VSS)

##
## Very Simple Structure
## VSS complexity 1 achieves a maximimum of 0.76  with  1  factors
## VSS complexity 2 achieves a maximimum of 0.8  with  3  factors
##
## The Velicer MAP criterion achieves a minimum of 0.5  with  1  factors
##
```

Zbog činjenice da imamo veliki uzorak a da je paralelna analiza osetljiva na veličinu uzorka i da kod velikih uzoraka precenjuje broj faktora (Revelle, 2017), koristili smo paralelnu analizu po modelu glavnih komponenti. Paralelna analiza sugeriše da je broj faktora koje bi trebalo zadržati 1, a isto sugerišu i VSS i MAP (Slika 74). Skri-dijagram sa tačkom preloma na drugom faktoru takođe sugeriše jednodimenzionalnost (Slika 73). Grafički ćemo proveriti dimenzionalnost upotrebom kategorijalne analize glavnih komponenti.




Slika 74 – Dijagram kriterijuma vrlo jednostavne strukture (VSS)

```
# Upotrebom kategorijalne analize glavnih komponenti
p_load(Gifi)
PCd <- princals(p.pod)
plot(PCd)
```




Slika 75 – Dijagram opterećenja kategorijalne analize glavnih komponenti

Ranije smo već rekli da bi kod jednodimenzionalnog skupa stavki strelice na grafikonu opterećenja kategorijalne analize glavnih komponenti trebalo da budu usmerene uglavnom u istom smeru. Na osnovu grafikona (Slika 75) možemo zaključiti da je skup stavki uglavnom jednodimenzionalan, s tim da stavke 1 i 6 donekle odstupaju od onoga što mere ostale stavke.

Uporedićemo jednodimenzionalni i dvodimenzionalni model upotrebom faktorske analize stavki (IFA). IFA je dostupna u paketima *mirt* i *ltm*, ali paket *ltm* nije namenjen politomnim stavkama. Postupak je isti kao kod binarno skorovanih stavki. Potrebno je proceniti jednodimenzionalni i dvodimenzionalni model (može i višedimenzionalni) i uporediti LR testom. Ukoliko je višedimenzionalni model značajno bolji, onda je pretpostavka jednodimenzionalnosti narušena.

```
# Učitaćemo biblioteku mirt
p_load(mirt)
mt1 <- mirt(p.pod, 1, verbose = FALSE)
mt2 <- mirt(p.pod, 2, verbose = FALSE)
# Ukoliko je potrebno može se dodati argument technical = list(NCYCLES=10000) i
povećati broj EM iteracija ako je to potrebno kako bi model konvergirao
anova(mt1, mt2)

##          AIC    SABIC       HQ      BIC    logLik     X2 df    p
## mt1 16340.59 16392.54 16396.54 16487.82 -8140.293
## mt2 16347.19 16414.73 16419.94 16538.59 -8134.595 11.395  9 0.25

# Ako je LR test značajan, bolji je model sa višom vrednošću log.Lik, odnosno onaj
sa nižim AIC ili BIC kriterijumom
detach("package:mirt", unload=TRUE, force = TRUE)
# U ovom trenutku nam više ne treba paket mirt
```

Ako se oslonimo na LR test, razlika između modela nije statistički značajna. Jednodimenzionalni model ima niže vrednosti AIC i BIC, što mu daje prednost. Možemo zaključiti da se jednodimenzionalni i dvodimenzionalni model ne razlikuju značajno, ali da informacioni kriterijumi daju prednost jednodimenzionalnom modelu. Zbog toga nećemo reći da je narušena pretpostavka jednodimenzionalnosti. Na osnovu procenjenih modela možemo dobiti i faktorska opterećenja ova dva modela:

```
m1df <- data.frame(summary(mt1)$rotF)
m2df <- data.frame(summary(mt2)$rotF)
m12df <- cbind(m1df, m2df)
m12df
summary(mt2)$fcor

             F1           F1            F2
it1   0.5084252    0.0973351     0.61876878
it2   0.7393491    0.7860367    -0.05862282
it3   0.5986873    0.5533263     0.06673221
it4   0.7581906    0.7059351     0.07616974
it5   0.7661069    0.6714757     0.13705907
it6   0.5195897    0.4256130     0.13491307
it7   0.6677311    0.5460455     0.17598763
it8   0.7305874    0.8199689    -0.11849163
it9   0.6547507    0.6933021    -0.04919170
it10  0.6717105    0.6158075     0.08124929
          F1        F2
F1 1.0000000 0.6348309
F2 0.6348309 1.0000000
```

Prva kolona ove tabele su faktorska opterećenja u jednofaktorskom modelu, a druga i treća su opterećenja stavki na dva faktora dvofaktorskog modela. Na dvofaktorskom modelu vidimo indiciju da stavka 1 odudara od ostalih po predmetu merenja. Iako u jednodimenzionalnom modelu ima zadovoljavajuće faktorsko opterećenje, u dvofaktorskom modelu njeno opterećenje je najveće na drugom "faktoru" (koji se i sastoji samo od te stavke). Korelacija dva faktora je visoka (0,635)

#### *10.5.2.2 Monotonost*

Proveru monotonosti obavićemo primenom paketa *mokken* (Van der Ark 2007). Ovaj paket namenjen je Mokenovoj analizi skala (videti odeljke 4.3 i 9.3).

```
p_load(mokken)
monotonost <- check.monotonicity(p.pod, minvi=.03)
summary(monotonost)

##      ItemH #ac #vi #vi/#ac maxvi  sum sum/#ac zmax #zsig crit
## it1   0.32  42   1    0.02  0.08 0.08  0.0018 0.99     0   19
## it2   0.45  42   1    0.02  0.03 0.03  0.0008 0.43     0    4
## it3   0.37  42   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it4   0.49  42   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it5   0.47  30   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it6   0.33  56   1    0.02  0.03 0.03  0.0006 0.38     0    9
## it7   0.41  42   1    0.02  0.07 0.07  0.0017 1.37     0   16
## it8   0.44  30   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it9   0.40  42   0    0.00  0.00 0.00  0.0000 0.00     0    0
## it10  0.43  42   0    0.00  0.00 0.00  0.0000 0.00     0    0

skalabilnost <- coefH(p.pod)

## $Hij
##      it1     se      it2     se      it3     se      it4     se      it5
## it1                  0.332  (0.037)  0.303  (0.036)  0.409  (0.042)  0.402
## it2   0.332  (0.037)                 0.401  (0.035)  0.481  (0.035)  0.482
## it3   0.303  (0.036)  0.401  (0.035)                 0.416  (0.036)  0.415
## it4   0.409  (0.042)  0.481  (0.035)  0.416  (0.036)                  0.589
## it5   0.402  (0.039)  0.482  (0.035)  0.415  (0.040)  0.589  (0.041)
## it6   0.256  (0.034)  0.370  (0.039)  0.273  (0.034)  0.445  (0.040)  0.382
## it7   0.389  (0.039)  0.439  (0.034)  0.361  (0.034)  0.463  (0.040)  0.451
## it8   0.312  (0.041)  0.493  (0.039)  0.419  (0.038)  0.566  (0.044)  0.398
## it9   0.310  (0.040)  0.475  (0.033)  0.391  (0.037)  0.527  (0.038)  0.481
## it10  0.282  (0.035)  0.576  (0.040)  0.384  (0.038)  0.589  (0.041)  0.628
##      se      it6     se      it7     se      it8     se      it9     se
## it1  (0.039)  0.256  (0.034)  0.389  (0.039)  0.312  (0.041)  0.310  (0.040)
## it2  (0.035)  0.370  (0.039)  0.439  (0.034)  0.493  (0.039)  0.475  (0.033)
## it3  (0.040)  0.273  (0.034)  0.361  (0.034)  0.419  (0.038)  0.391  (0.037)
## it4  (0.041)  0.445  (0.040)  0.463  (0.040)  0.566  (0.044)  0.527  (0.038)
## it5           0.382  (0.040)  0.451  (0.036)  0.398  (0.035)  0.481  (0.038)
## it6  (0.040)                  0.394  (0.035)  0.406  (0.038)  0.309  (0.036)
## it7  (0.036)  0.394  (0.035)                  0.413  (0.034)  0.368  (0.034)
## it8  (0.035)  0.406  (0.038)  0.413  (0.034)                  0.463  (0.039)
## it9  (0.038)  0.309  (0.036)  0.368  (0.034)  0.463  (0.039)
## it10 (0.044)  0.292  (0.035)  0.470  (0.041)  0.565  (0.045)  0.393  (0.038)
##      it10    se
## it1   0.282  (0.035)
## it2   0.576  (0.040)
## it3   0.384  (0.038)
```

```
## it4   0.589  (0.041)
## it5   0.628  (0.044)
## it6   0.292  (0.035)
## it7   0.470  (0.041)
## it8   0.565  (0.045)
## it9   0.393  (0.038)
## it10
##
## $Hi
##      Item H  se
## it1    0.321 (0.023)
## it2    0.447 (0.020)
## it3    0.366 (0.021)
## it4    0.493 (0.021)
## it5    0.466 (0.021)
## it6    0.332 (0.021)
## it7    0.413 (0.020)
## it8    0.443 (0.021)
## it9    0.402 (0.021)
## it10   0.427 (0.022)
##
## $H
## Scale H      se
##   0.404 (0.015)

skalabilnost$H

## Scale H      se
##   0.404 (0.015)
```

Pokazatelje prikazane u tabeli koja je rezultat funkcije *check.monotonicity()* opisali smo u delu koji se odnosi na binarno skorovane stavke, pa to nećemo ponavljati (videti 10.3.2).

Podsetićemo se interpretacije vrednosti *Crit* (Crişan i ostali, 2019):

- crit<40 – nema dokaza da postoji narušavanje monotonosti
- 40<crit<80 – postoje naznake narušavanja monotonosti, ali nisu jednoznačne
- crit>80 – monotonost je ozbiljno narušena

Na našim podacima nema ozbiljnijeg narušavanja pretpostavke monotonosti. Najveće je kod stavke 6 koja ima *Crit* vrednost 19, ali iz vrednosti u koloni *zsig* vidimo da narušavanje nije statistički značajno. Najveće standardizovano pojedinačno narušavanje monotonosti ima stavka 7 (*zmax*=1,37), ali ni ono nije statistički značajno. Inače, ova stavka je druga najlošija po vrednosti *Crit* (16).

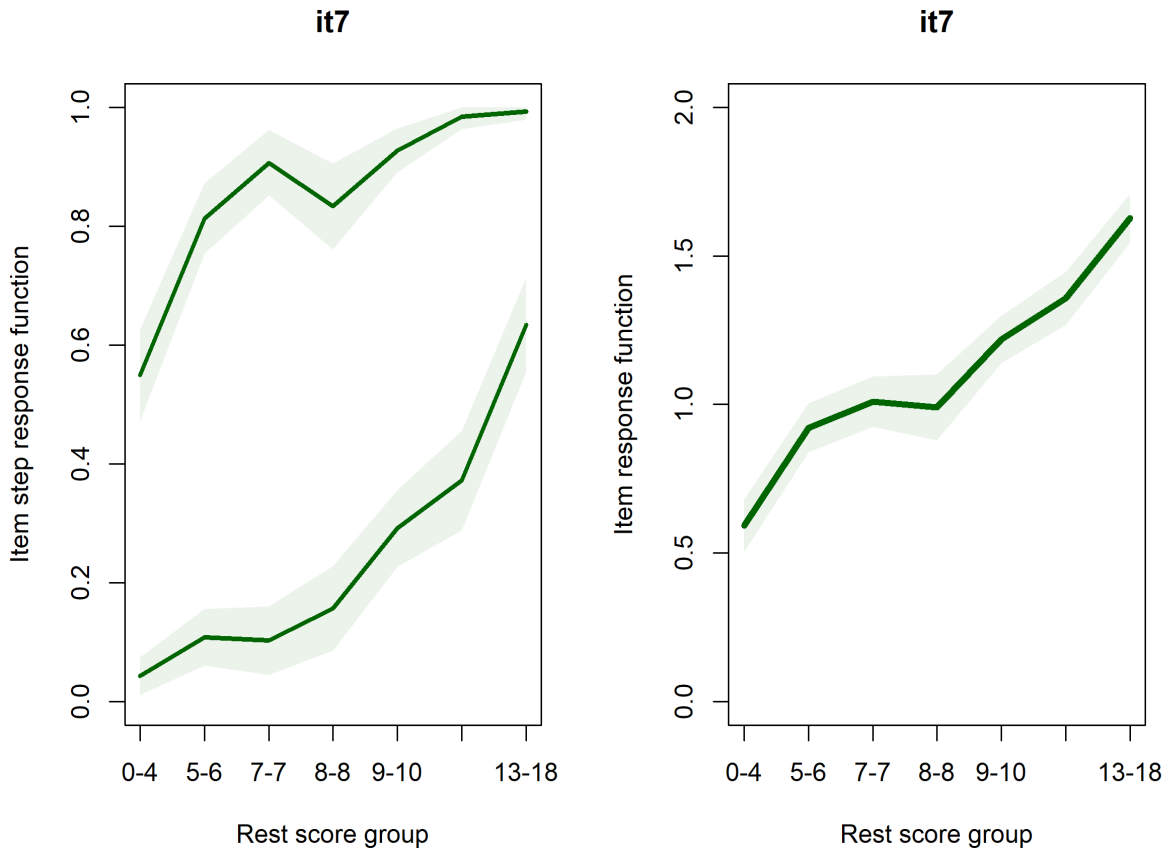
Pravila interpretacije indeksa skalabilnosti su sledeća (Sijtsma & van der Ark, 2017):

- $0,3 \leq H_j < 0,4$ – slaba skala (ili skalabilnost stavke)
- $0,4 \leq H_j < 0,5$ – srednja skala (ili skalabilnost stavke)
- $H_j \geq 0,5$ – jaka skala (ili skalabilnost stavke)
- a za skup stavki čiji je $H<0,3$ kažemo da nije skalabilan.

Indeksi skalabilnosti stavki su u opsegu prihvatljivih vrednosti, uglavnom slabi ili osrednji. Najnižu skalabilnost imaju stavke 1 i 6. Stavke 1 i 6 su identifikovane kao potencijalno problematične prilikom provere dimenzionalnosti, pa nije čudno što su takve i kada posmatramo pretpostavku monotonosti. Ako bar delom mere (i) nešto drugo osim latentne osobine koju mere ostale stavke, skor na njima neće nužno rasti

sa porastom nivoa zajedničkog predmeta merenja. 95% interval poverenja indeksa skalabilnosti $H_j \pm 1.96SE_{Hj}$ za navedene dve stavke obuhvata vrednost 0,3 što znači da bi u populaciji mogao biti i niži od granice prihvatljivih vrednosti. Indeks skalabilnosti testa je osrednji (0,404), na granici ka slabom.

```
plot(monotonost, items=7, ask=FALSE, color="dark green")
```




Slika 76 – Krive odgovora između kategorija (levo) i kriva odgovora na stavku (desno)

Za grafički prikaz smo odabrali stavku 7 jer ima najvišu vrednost *#zsig*. Na grafikonima (Slika 76) su predstavljene krive odgovora između kategorija (levo) i kriva odgovora na stavku (desno). Za proveru narušavanja monotonosti bitnija nam je kriva za stavku u celini. Na krivi se uočava narušavanje monotonosti u grupi *skorova ostatka* 7–8. Na tom mestu kriva blago opada, ali s obzirom na interval poverenja ne može se reći da je to narušavanje monotonosti značajno. U tabeli sa pokazateljima monotonosti nema nijednog značajnog narušavanja (kolona *zsig*). Na grafikonu sa krivama odgovora između kategorija takođe uočavamo naznake narušavanja monotonosti u grupi skorova ostatka 7–8 na delu krive za odgovore između prve i druge kategorije (gornja kriva). Ni ovde se ne može reći da je narušavanje statistički značajno. Čak i kada bi bilo, narušavanje monotonosti kategorija nije problem ako ne utiče na narušavanje monotonosti stavke u celini (Sijtsma & van der Ark, 2017).

#### *10.5.2.3 Lokalna nezavisnost*

Kako reziduali zavise od primenjenog modela, proveru lokalne nazavisnost detaljnije ćemo obaviti nakon procene modela u paketima u kojima budemo vršili analize. Ovde ćemo proceniti jednofaktorski model u paketu *psych* koristeći funkciju *fa()* i nakon toga proveriti korelacije reziduala koristeći funkciju *residuals()*.

```
m1f <- fa(p.pod, nfactors=1, cor="poly")
# Jednofaktorski model na matrici polihoričkih korelacija
LDres <- residuals(m1f, diag = FALSE)
# Kreiranje matrice korelacija reziduala na modelu (bez vrednosti u dijagonali)
LDres[abs(LDres)<.2] <- NA
# Zbog preglednosti zamenićemo sve korelacije koje imaju apsolutnu vrednost <0,2
sa nedostajućim vrednostima
LDres
```

```
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA
## it2   NA  NA
## it3   NA  NA  NA
## it4   NA  NA  NA  NA
## it5   NA  NA  NA  NA  NA
## it6   NA  NA  NA  NA  NA  NA
## it7   NA  NA  NA  NA  NA  NA  NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
```

```
# Prikaz matrice korelacije reziduala
```

Nema stavki sa povišenim korelacijama reziduala (>0,2) pa stoga (preliminarno) možemo smatrati da je ispunjen uslov lokalne nezavisnosti.

### 10.5.3 Model skale procena (RSM) u paketu *eRm*

Počećemo sa Rašovom familijom modela (RSM i PCM) u paketima *eRm* i *mirt*, a zatim preći na generalizovani model stepenovanog ocenjivanja (GPCM) u paketu *mirt* i model stepenovanog odgovaranja (GRM) u paketima *mirt* i *ltm*.

*Model skala procene* pripada familiji Rašovih modela i procenićemo ga prvo pomoću paketa *eRm* (Mair & Hatzinger, 2007) koji je tome i namenjen. Metod procene parametara u ovom paketu je metod uslovne maksimalne verodostojnosti (CMLE). Ovaj model je jednoparametarski i računa samo parametre težine stavki i parametre kategorija (pragove). Nagibi svih stavki su isti ($a$=1), a parametri kategorija ekvidistantni. *Ekvidistantnost* ne znači da su razmaci između sekvencijalnih pragova u okviru iste stavke jednaki, već da su razmaci između istih pragova u svim stavkama jednaki. Zato se procenjuje samo jedan set pragova koji je jednak za sve stavke. Jednaki setovi pragova za sve stavke impliciraju da stavke moraju biti istog formata. Za ovaj model *eRm* zahteva da najniža kategorija ajtemskog skora bude 0. Ako u matrici podataka to nije slučaj, *eRm* će interno prilagoditi podatke tom uslovu i obavestiće porukom (neće menjati matricu podataka).

```
# Učitavanje biblioteke
p_load(eRm)
# Za procenu modela skala procene koristi se funkcija RSM()
mRSM.e <- RSM(p.pod, se=T)

## Warning:
## The following items have no 0-responses:
## it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## Responses are shifted such that lowest category is 0.

# Funkcija RSM zahteva da najniža vrednost bude 0 pa su sve vrednosti interno
umanjene za 1 (ne menja se matrica podataka)
# Procenjeni model smešten je u objekat mRSM.e
parRSM.e <- thresholds(mRSM.e)
parRSM.e

##
## Design Matrix Block 1:
##      Location Threshold 1 Threshold 2
## it1   2.11928     0.85400     3.38455
## it2   1.27696     0.01169     2.54224
## it3   1.46433     0.19905     2.72960
## it4   1.69758     0.43230     2.96285
## it5   0.90413    -0.36114     2.16941
## it6   1.77638     0.51110     3.04165
## it7   0.78756    -0.47772     2.05284
## it8   0.76295    -0.50233     2.02822
## it9   0.58970    -0.67558     1.85498
## it10  1.27390     0.00862     2.53918

# Funkcijom thresholds() dobijamo ispis parametara modela: težinu stavki lokacije
pragova
```

Pošto RSM pripada Rašovoj familiji modela, ne procenjuje se parametar diskriminativnosti. Prikazani su parametri težine stavki i pragovi. U ovom modelu pragovi su lokacije na kontinuumu latentne osobine gde ispitanici imaju jednaku verovatnoću da odaberu neku od dve susedne kategorije. Ranije smo rekli da su pragovi ekvidistantni za sve stavke. Ilustrovaćemo to na primeru našeg modela. Izračunaćemo razliku između lokacija pragova na svim stavkama. Od oba praga oduzećemo parametar težine stavke da bismo dobili parametre pragova.

```
parRSM.e <- data.frame(parRSM.e$threshtable)
names(parRSM.e) <- c("bj", "t1", "t2")
parRSM.e$interval <- parRSM.e$t2-parRSM.e$t1 # Od drugog praga oduzećemo prvi da
vidimo interval između njih (biće jednak za sve stavke)
parRSM.e$b1 <- parRSM.e$t1-parRSM.e$bj # Od vrednosti prvog praga oduzećemo
parametar lokacije stavke da dobijemo lokaciju prvog praga
parRSM.e$b2 <- parRSM.e$t2-parRSM.e$bj # Ponovićemo to isto za drugi prag
round(parRSM.e,3)

##         bj     t1    t2 interval     b1    b2
## it1  2.119  0.854 3.385    2.531 -1.265 1.265
## it2  1.277  0.012 2.542    2.531 -1.265 1.265
## it3  1.464  0.199 2.730    2.531 -1.265 1.265
## it4  1.698  0.432 2.963    2.531 -1.265 1.265
## it5  0.904 -0.361 2.169    2.531 -1.265 1.265
## it6  1.776  0.511 3.042    2.531 -1.265 1.265
## it7  0.788 -0.478 2.053    2.531 -1.265 1.265
```

```
## it8  0.763 -0.502 2.028    2.531 -1.265 1.265
## it9  0.590 -0.676 1.855    2.531 -1.265 1.265
## it10 1.274  0.009 2.539    2.531 -1.265 1.265
```

Vidimo da je interval između lokacija dva praga jednak za sve stavke, a isto tako i parametri pragova su identični za sve stavke. Ovaj model procenjuje *m* parametara težina stavki i samo jedan set parametara pragova kojih ima *k–1* (gde je *k* broj kategorija odgovora). Parametri težine nam dosta govore o testu u celini. Na osnovu njihove distribucije znaćemo da li je test lak ili težak, a znaćemo i gde će biti najprecizniji, odnosno najpouzdaniji i najinformativniji. Podsetićemo, informativnost testa predstavlja sumu informativnosti stavki i funkcija je nivoa osobine. Analogno tome, informativnost stavki je suma informativnosti kategorija. Kako u ovom testu imamo veći broj stavki koje su nešto teže od prosečne težine od 0 logita (videti gornju tabelu), možemo pretpostaviti da će test biti informativniji za ispitanike sa nivoom osobine u tom delu kontinuuma. Isti model možemo proceniti i pomoću paketa *mirt*.

```
mRSM.m <- mirt::mirt(p.pod, 1, itemtype = "rsm", SE=T, verbose = FALSE)

parRSM.m <- data.frame(coef(mRSM.m, simplify=TRUE, IRTpars=TRUE)$items[,-1])
parRSM.m # Ovo su parametri pragova

##              b1      b2         c
## it1  -0.2423744 2.30049 0.0000000
## it2  -0.2423744 2.30049 0.8434691
## it3  -0.2423744 2.30049 0.6558982
## it4  -0.2423744 2.30049 0.4223224
## it5  -0.2423744 2.30049 1.2165671
## it6  -0.2423744 2.30049 0.3433972
## it7  -0.2423744 2.30049 1.3332046
## it8  -0.2423744 2.30049 1.3578313
## it9  -0.2423744 2.30049 1.5311948
## it10 -0.2423744 2.30049 0.8465328

# Izračunaćemo njihove lokacije
parRSM.m$t1 <- parRSM.m$b1 - parRSM.m$c
parRSM.m$t2 <- parRSM.m$b2 - parRSM.m$c
parRSM.m

##              b1      b2         c         t1        t2
## it1  -0.2423744 2.30049 0.0000000 -0.2423744 2.3004900
## it2  -0.2423744 2.30049 0.8434691 -1.0858435 1.4570210
## it3  -0.2423744 2.30049 0.6558982 -0.8982726 1.6445918
## it4  -0.2423744 2.30049 0.4223224 -0.6646968 1.8781676
## it5  -0.2423744 2.30049 1.2165671 -1.4589415 1.0839229
## it6  -0.2423744 2.30049 0.3433972 -0.5857716 1.9570928
## it7  -0.2423744 2.30049 1.3332046 -1.5755790 0.9672854
## it8  -0.2423744 2.30049 1.3578313 -1.6002057 0.9426588
## it9  -0.2423744 2.30049 1.5311948 -1.7735692 0.7692953
## it10 -0.2423744 2.30049 0.8465328 -1.0889071 1.4539573

# Uporedićemo rezultate eRm i mirt paketa
PP <- data.frame(cbind(parRSM.e[, c("bj","t1","t2")],
 parRSM.m[, c("c","t1","t2")]))
names(PP) <- c("bj.eRm", "t1.eRm","t2.eRm", "bj.mirt", "t1.mirt","t2.mirt")
round(PP,3)

##      bj.eRm t1.eRm t2.eRm bj.mirt t1.mirt t2.mirt
## it1   2.119  0.854  3.385   0.000  -0.242   2.300
```

```
## it2   1.277  0.012  2.542   0.843  -1.086   1.457
## it3   1.464  0.199  2.730   0.656  -0.898   1.645
## it4   1.698  0.432  2.963   0.422  -0.665   1.878
## it5   0.904 -0.361  2.169   1.217  -1.459   1.084
## it6   1.776  0.511  3.042   0.343  -0.586   1.957
## it7   0.788 -0.478  2.053   1.333  -1.576   0.967
## it8   0.763 -0.502  2.028   1.358  -1.600   0.943
## it9   0.590 -0.676  1.855   1.531  -1.774   0.769
## it10  1.274  0.009  2.539   0.847  -1.089   1.454

# Korelacije lokacija pragova iz paketa eRm i mirt
cor(PP$t1.eRm, PP$t1.mirt)

## [1] 1

cor(PP$t2.eRm, PP$t2.mirt)

## [1] 1

# Razlike između lokacija pragova u paketu eRm i mirt
round(cbind("t1dif"=PP$t1.eRm-PP$t1.mirt,
 "t2dif"=PP$t2.eRm-PP$t2.mirt), 1)

##       t1dif t2dif
##  [1,]   1.1   1.1
##  [2,]   1.1   1.1
##  [3,]   1.1   1.1
##  [4,]   1.1   1.1
##  [5,]   1.1   1.1
##  [6,]   1.1   1.1
##  [7,]   1.1   1.1
##  [8,]   1.1   1.1
##  [9,]   1.1   1.1
## [10,]   1.1   1.1
```

Vidimo da se procene lokacija pragova u dva paketa razlikuju, ali su savršeno korelirane i predstavljaju linearne transformacije (ako zaokružimo na 1 decimalu, parametri pragova su za 1.1 viši u paketu *eRm*).

#### *10.5.3.1 Procena parametara ispitanika*

Pošto još nisu procenjeni, procenićemo parametre ispitanika. Sačuvaćemo ih u objekat *mRSM.ePI* pošto će nam biti potrebni za kasnija izračunavanja.

```
mRSM.ePI <- person.parameter(mRSM.e)
mRSM.ePI

##
## Person Parameters:
##
##  Raw Score   Estimate Std.Error
##          0 -3.1782637        NA
##          1 -2.3069286 1.0519377
##          2 -1.5038914 0.7830074
##          3 -0.9818200 0.6734035
##          4 -0.5702552 0.6145489
##          5 -0.2154697 0.5793635
##          6  0.1066867 0.5574201
```

```
##          7  0.4092169 0.5436901
##          8  0.6999345 0.5354038
##          9  0.9839249 0.5309531
##         10  1.2648125 0.5294817
##         11  1.5455881 0.5307506
##         12  1.8292867 0.5350869
##         13  2.1196560 0.5434051
##         14  2.4219531 0.5573342
##         15  2.7441713 0.5796271
##         16  3.0994854 0.6152604
##         17  3.5121826 0.6745882
##         18  4.0361409 0.7845888
##         19  4.8418537 1.0536677
##         20  5.7160312        NA
```

Kao i kod modela za binarno skorovane stavke, krajnja leva kolona prikazuje sumacioni skor, druga Rašove mere, a poslednja njihove standardne greške merenja. I ovde svi ispitanici sa istim sumacionim skorom imaju jednake mere i standardne greške merenja. Standardne greške su najmanje na nivou osobine oko 0 logita, a najveće kod ekstremnih skorova. Za savršene ukupne skorove ovde nije moguće izračunavanje standardnih greški.

#### *10.5.3.2 Pokazatelj fita modela u celini*

Izračunaćemo Andersenov LR statistik kao pokazatelj fita modela u celini.

```
mRSM.eLR <- LRtest(mRSM.e, splitcr = "mean")
# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
mRSM.eLR
```

```
##
## Andersen LR-test:
## LR-value: 292.654
## Chi-square df: 10
## p-value:  0
```

Andersenov LR test je značajan pa ne možemo reći da je model saglasan sa podacima. To nije iznenađenje jer smo simulirali podatke po modelu stepenovanog ocenjivanja koji ne predviđa jednake nagibe niti ekvidistantne pragove.

Ukoliko neke kategorije odgovora nisu prisutne na nekoj stavci u poduzorcima dobićete poruku ovakve sadržine: *Warning: The following items were excluded due to inappropriate response patterns within subgroups:* (spisak stavki koje su isključene).

Proverićemo i globalni fit modela koji smo procenili u paketu *mirt*.

```
mirt::M2(mRSM.m)
```

```
##             M2 df          p      RMSEA    RMSEA_5   RMSEA_95      SRMSR
## stats 106.6783 43 2.5007e-07 0.03850159 0.02936771 0.04773025 0.04840372
##             TLI       CFI
## stats 0.9840781 0.9804388
```

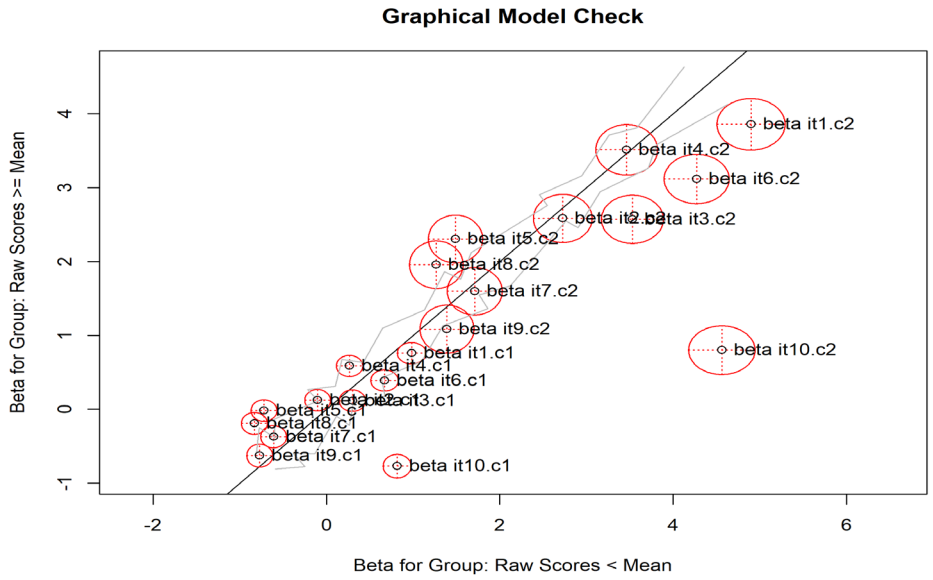
Pokazatelj bliskog fita $M_2$ je statistički značajan, što ukazuje da model nije saglasan sa podacima.

Međutim, pokazatelji približnog fita RMSEA$_2$, SRMSR, TLI i CFI imaju zadovoljavajuće vrednosti. Podsetićemo da *eRm* i *mirt* na različite načine procenjuju parametre stavki (CMLE i MMLE respektivno).

#### 10.5.3.2.1 Grafička provera modela

Pogledaćemo grafikone procene parametara stavki na dva poduzorka kako bismo videli da li je zadovoljena invarijantnost procene parametara koja je karakteristika Rašovih modela.

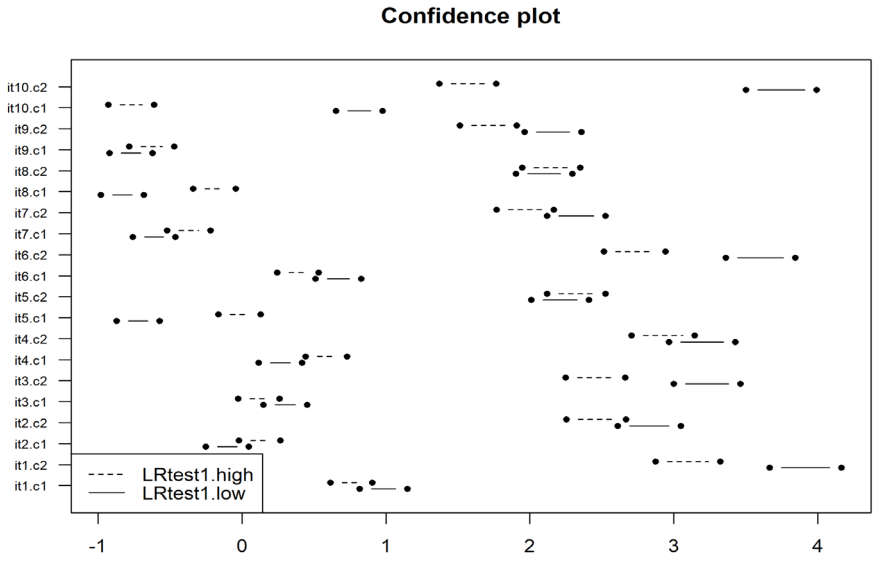
```
plotGOF(mRSM.eLR, ctrline = list(col = "gray"), conf = list())
```




Slika 77 – Grafička provera fita modela (invarijantnost procena parametara)

Problematična je stavka 10, ali nije jedina (Slika 77). U slučaju politomnih modela pregledniji je grafikon sa uporednim prikazom intervala poverenja procena parametara na dva poduzorka.

```
plotDIF(mRSM.eLR, leg=TRUE)
```




Slika 78 – Intervali poverenja procene parametara u grupama sa niskim i visokim

Najproblematičnija je stavka 10, kod koje se 95% intervali poverenja za obe kategorije razlikuju na poduzorcima (ne preklapaju se, što znači da se procene parametara razlikuju na dva poduzorka – Slika 78). Nije to jedina stavka kod koje uočavamo problem. Razlike postoje kod kategorije 2 stavke 9, kategorije 1 stavke 8, kategorije 2 stavke 6, kategorije 1 stavke 5, kategorije 2 stavke 3, kategorije 2 stavke 1 i možda kategorije 1 stavke 4. Izbacivanjem stavke 10 promenili bi se i ukupni skorovi na osnovu kojih se radi procena parametara stavki u CMLE pristupu, pa bi bilo dobro videti šta se menja i da li će se nešto promeniti u pogledu procene parametara preostalih stavki.

#### 10.5.3.2.2 Pokazatelji fita za stavke

Prikazaćemo pokazatelje fita za stavke.

```
# Pošto i eRm i mirt imaju funkciju pod istim nazivom - itemfit(), moguće je da
funkcija iz paketa eRm neće raditi ako ste posle njega učitali paket mirt (i
obrnuto). U tom slučaju morate ukloniti mirt sa:
#detach("package:mirt", unload=TRUE)
# ili umesto donje naredbe napisati eRm::itemfit(mr.ePI) i tako funkciju pozvati
direktno iz tog paketa
ItF <- itemfit(mRSM.ePI)
ItF

##
## Itemfit Statistics:
##         Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it1  1210.796 985   0.000      1.228     1.248    4.594   5.629   0.424
## it2   656.179 985   1.000      0.665     0.647   -9.060 -10.048   0.603
## it3  1016.502 985   0.237      1.031     1.054    0.739   1.331   0.505
## it4   604.563 985   1.000      0.613     0.582  -10.449 -12.250   0.592
## it5   568.099 985   1.000      0.576     0.525  -11.905 -14.333   0.586
## it6  1266.258 985   0.000      1.284     1.328    6.003   7.378   0.453
## it7   746.126 985   1.000      0.757     0.740   -6.237  -7.067   0.556
## it8   598.103 985   1.000      0.607     0.562  -10.778 -12.926   0.556
## it9   897.524 985   0.978      0.910     0.901   -2.118  -2.495   0.545
## it10 1621.405 985   0.000      1.644     1.744   12.906  15.160   0.547

# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
WT=Waldtest(mRSM.e,splitcr = "mean")
# Argument splitcr ima isto značenje i opcije kao u funkciji LRtest()
WT

## Wald test on item level (z-values):
##
##              z-statistic p-value
## beta it1.c1       -1.999   0.046
## beta it1.c2       -3.884   0.000
## beta it2.c1        2.125   0.034
## beta it2.c2       -0.594   0.552
## beta it3.c1       -1.709   0.087
## beta it3.c2       -3.884   0.000
## beta it4.c1        3.000   0.003
## beta it4.c2        0.190   0.849
## beta it5.c1        6.622   0.000
## beta it5.c2        3.555   0.000
## beta it6.c1       -2.577   0.010
```

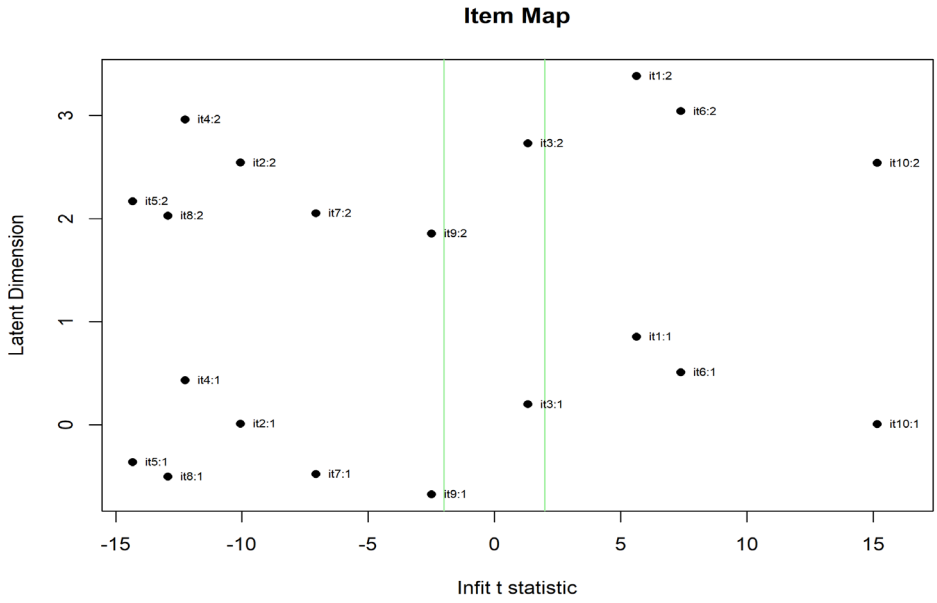
```
## beta it6.c2       -4.502   0.000
## beta it7.c1        2.220   0.026
## beta it7.c2       -0.490   0.624
## beta it8.c1        5.970   0.000
## beta it8.c2        3.033   0.002
## beta it9.c1        1.305   0.192
## beta it9.c2       -1.327   0.185
## beta it10.c1     -13.665   0.000
## beta it10.c2     -14.601   0.000
```

Značajan misfit na osnovu hi–kvadrat testa imaju stavke: 10, 6 i 1. Na osnovu pokazatelja infita i autfita kao misfitujuće stavke možemo identifikovati stavku 10 i 6 (ako kao prihvatljivi opseg MSQ misfita koristimo 0,7–1,3). Obe stavke imaju misfit u obliku šuma. Ako gledamo pokazatelje misfita u standardizovanoj metrici, tu ima više stavki koje izlaze iz opsega prihvatljivih vrednosti (–1,96–+1,96). Ipak, samo tri ranije navedene imaju vrednosti koje ukazuju na šum, dok je kod ostalih u pitanju prigušeni šum, a hi–kvadrati i pokazatelji u MSQ metrici nisu upadljivi. Prema Valdovom statistiku misfit postoji kod stavke 10, 5, 8, 6, 5 i 1 (za oba parametra praga), i kod stavki 7, 3, 4 i 2 za po jednu od kategorija. Ajtem–total korelacije stavki 1 i 6 su nešto niže od ostalih. Podsetićemo, ove dve stavke su identifikovane kao potencijalno problematične prilikom analize jednodimenzionalnosti i monotonosti.

#### *10.5.3.2.2.1 Grafički prikaz standardizovanog infita za stavke*

Pošto veliki broj stavki ima loše pokazatelje standardizovanog infita, a istovremeno nema značajne pokazatelje infita u MSQ metrici niti značajan hi–kvadrat, ovaj grafikon (Slika 79) nam nije informativan i nećemo mu pridavati veliku pažnju. Ipak, uočite da stavke sa značajnim misfitom imaju značajan standardizovani infit šumnog tipa (odgovaranje koje model ne predviđa dobro – desno od desne zelene linije). Ostale stavke imaju standardizovani infit u obliku prigušenog šuma – suviše predvidljivo odgovaranje).

```
plotPWmap(mRSM.e, imap=TRUE, pmap=FALSE)
```




Slika 79 – Standardizovani infit za stavke

```
# Argument imap=TRUE određuje da će biti prikazan infit za stavke, a pmap=FALSE da
neće biti prikazan za ispitanike. Moguće je napraviti isti grafikon i za
ispitanike (imap=FALSE, pmap=TRUE), ili grafikon koji će objediniti i stavke i
ispitanike (imap=TRUE, pmap=TRUE).
# Ukoliko želite grafikon i za ispitanike potrebno je dodati argument pp=mr.ePI
# mr.ePI je objekat u kojem smo sačuvali parametre ispitanika
```

#### *10.5.3.3 Procena lokalne zavisnosti*

Pokazatelji dostupni za procenu lokalne zavisnosti u paketu *eRm* koje smo koristili kod modela namenjenih binarno skorovanim stavkama ($T_{11}$, $T_{1l}$, $Q_{3h}$) ovde nisu primenljivi jer su namenjeni samo binarnim stavkama. Lokalnu zavisnost ćemo pokušati da ocenimo na osnovu modela koji smo procenili u paketu *mirt*.

```
mirt::residuals(mRSM.m, type="LD", suppress=.212)

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.726  -0.154   0.296   0.175   0.366   0.750
##
##          it1     it2     it3      it4      it5      it6     it7      it8
## it1       NA  -0.245      NA    0.296    0.344   -0.314      NA   -0.354
## it2  119.797      NA      NA    0.312    0.366   -0.301   0.216    0.366
## it3       NA      NA      NA    0.259    0.309   -0.276      NA    0.312
## it4  175.045 194.559 134.003       NA    0.406    0.354   0.268    0.390
## it5  237.253 268.205 190.494  329.243       NA    0.407   0.340    0.447
## it6  197.395 181.429 152.735  250.218  330.997       NA   0.273    0.399
## it7       NA  92.920      NA  144.092  231.216  149.170      NA    0.323
## it8  250.491 267.183 195.068  304.415  398.810  318.597 209.060       NA
## it9       NA      NA      NA  125.180  183.719  134.672      NA  178.813
## it10 940.348 907.888 820.620 1034.132 1125.045 1054.483 846.775 1123.806
##          it9   it10
## it1       NA -0.686
## it2       NA  0.674
## it3       NA -0.641
## it4    0.250  0.719
## it5    0.303  0.750
## it6   -0.259 -0.726
## it7       NA  0.651
## it8    0.299  0.750
## it9       NA  0.627
## it10 786.851     NA

mirt::residuals(mRSM.m, type="LD", suppress=.354)

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.726  -0.154   0.296   0.175   0.366   0.750
##
##          it1     it2    it3      it4      it5      it6     it7      it8     it9
## it1       NA      NA     NA       NA       NA       NA      NA       NA      NA
## it2       NA      NA     NA       NA    0.366       NA      NA    0.366      NA
## it3       NA      NA     NA       NA       NA       NA      NA       NA      NA
## it4       NA      NA     NA       NA    0.406       NA      NA    0.390      NA
```

```
## it5       NA 268.205     NA  329.243       NA    0.407      NA    0.447      NA
## it6       NA      NA     NA       NA  330.997       NA      NA    0.399      NA
## it7       NA      NA     NA       NA       NA       NA      NA       NA      NA
## it8       NA 267.183     NA  304.415  398.810  318.597      NA       NA      NA
## it9       NA      NA     NA       NA       NA       NA      NA       NA      NA
## it10 940.348 907.888 820.62 1034.132 1125.045 1054.483 846.775 1123.806 786.851
##        it10
## it1  -0.686
## it2   0.674
## it3  -0.641
## it4   0.719
## it5   0.750
## it6  -0.726
## it7   0.651
## it8   0.750
## it9   0.627
## it10     NA
```

Ispod dijagonale nalaze se vrednosti hi–kvadrata a iznad standardizovane vrednosti, odnosno Kramerovo V. Za interpretaciju Kramerovog V, De Ajala (de Ayala, 2022) predlaže Koenovu podelu veličina efekta (Cohen, 1988) (videti odeljak 9.2). Ako bismo srednji efekat posmatrali kao značajnu lokalnu zavisnost, onda bismo kao graničnu vrednost morali uzeti 0,212, a ako se odlučimo za veliki efekat onda je to vrednost 0,354. Kada primenimo prvu od ove dve vrednosti, lokalne zavisnosti ima jako puno, a sa drugom nešto manje. Stavka 8 se često javlja u lokalno zavisnim parovima, ali parovi u koje je uključena stavka 5 imaju najviše vrednosti. Proverićemo i lokalnu zavisnost na osnovu korelacija reziduala, odnosno pokazatelja $Q_3$. Kao nezanemarljive vrednosti $Q_3$ uzećemo one koje su za 0,2 veće od prosečne vrednosti u matrici.

```
Q3m <- as.matrix(mirt::residuals(mRSM.m, type="Q3"))
```

```
## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.171  -0.123  -0.091  -0.092  -0.070  -0.023
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.142 -0.093 -0.105 -0.081 -0.097 -0.069 -0.171 -0.134 -0.119
## it2  -0.142  1.000 -0.093 -0.042 -0.023 -0.141 -0.075 -0.030 -0.030 -0.080
## it3  -0.093 -0.093  1.000 -0.091 -0.126 -0.123 -0.116 -0.079 -0.074 -0.102
## it4  -0.105 -0.042 -0.091  1.000 -0.041 -0.094 -0.135 -0.081 -0.037 -0.083
## it5  -0.081 -0.023 -0.126 -0.041  1.000 -0.156 -0.050 -0.032 -0.071 -0.130
## it6  -0.097 -0.141 -0.123 -0.094 -0.156  1.000 -0.051 -0.085 -0.141 -0.126
## it7  -0.069 -0.075 -0.116 -0.135 -0.050 -0.051  1.000 -0.070 -0.085 -0.118
## it8  -0.171 -0.030 -0.079 -0.081 -0.032 -0.085 -0.070  1.000 -0.070 -0.149
## it9  -0.134 -0.030 -0.074 -0.037 -0.071 -0.141 -0.085 -0.070  1.000 -0.111
## it10 -0.119 -0.080 -0.102 -0.083 -0.130 -0.126 -0.118 -0.149 -0.111  1.000
```

```
# Uklonićemo jedinice iz dijagonale
diag(Q3m) <- NA
# Izračunaćemo prosečnu vrednost
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", round(Q3mean, 3), "\n\n")
```

```
##
## Prosečna vrednost Q3: -0.092
```

```
# Od vrednosti u matrici oduzećemo prosečnu vrednost
Q3c <- Q3m-Q3mean
Q3c
```

```
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.049 -0.001 -0.013  0.012 -0.005  0.024 -0.079 -0.041 -0.027
## it2  -0.049     NA -0.001  0.050  0.069 -0.049  0.018  0.062  0.062  0.012
## it3  -0.001 -0.001     NA  0.001 -0.034 -0.031 -0.024  0.013  0.018 -0.010
## it4  -0.013  0.050  0.001     NA  0.052 -0.002 -0.043  0.011  0.055  0.009
## it5   0.012  0.069 -0.034  0.052     NA -0.063  0.042  0.060  0.021 -0.038
## it6  -0.005 -0.049 -0.031 -0.002 -0.063     NA  0.041  0.007 -0.049 -0.034
## it7   0.024  0.018 -0.024 -0.043  0.042  0.041     NA  0.022  0.007 -0.025
## it8  -0.079  0.062  0.013  0.011  0.060  0.007  0.022     NA  0.022 -0.057
## it9  -0.041  0.062  0.018  0.055  0.021 -0.049  0.007  0.022     NA -0.019
## it10 -0.027  0.012 -0.010  0.009 -0.038 -0.034 -0.025 -0.057 -0.019     NA
```

```
# Prikazaćemo samo one veće od 0,20
Q3c[Q3c<.20] <- NA
Q3c
```

```
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

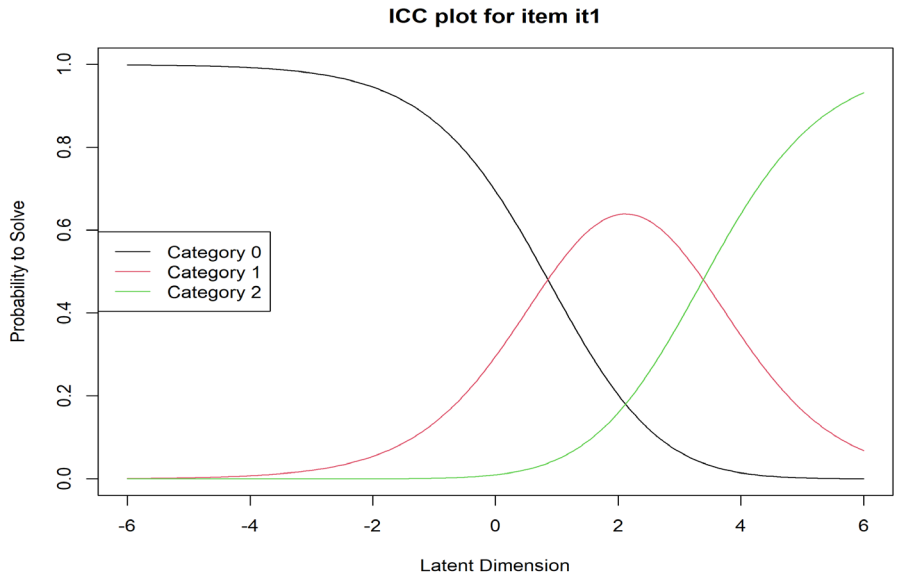
```
# Nema većih od 0,20
```

Obratite pažnju da smo od vrednosti $Q_3$ pokazatelja smeštenih u matricu Q3m oduzeli prosečnu vrednost pokazatelja u matrici. Kako je ta vrednost obično negativna, to znači da smo ih uvećali za apsolutnu vrednost prosečnog $Q_3$. Tako uvećane vrednosti smo zapisali u matricu Q3c čije smo vrednosti poredili sa 0,20 (sakrili smo sve niže – a to su sve). Prema modelu koji smo procenili u paketu *mirt* i na osnovu $Q_3$ vrednosti ne nalazimo povišenu lokalnu zavisnost. Za više informacija o interpretaciji $Q_3$ pogledajte odeljak 9.2.

#### *10.5.3.4 Grafički prikaz karakterističnih kriva kategorija*

U politomnim modelima dobijamo prikaz karakterističnih kriva kategorija, a ne stavki (Slika 80). Za razliku od modela za binarno skorovane stavke, nećemo dobiti prikaz svih stavki na jednom grafikonu, niti bi to ovde bilo pregledno i praktično. Ni ovde krive nisu posebno informativne. Zbog istog nagiba i istih intervala između pragova sve izgledaju isto, samo se razlikuju po lokaciji.

```
plotICC(mRSM.e, item.subset = 1, xlim=c(-6, 6))
```



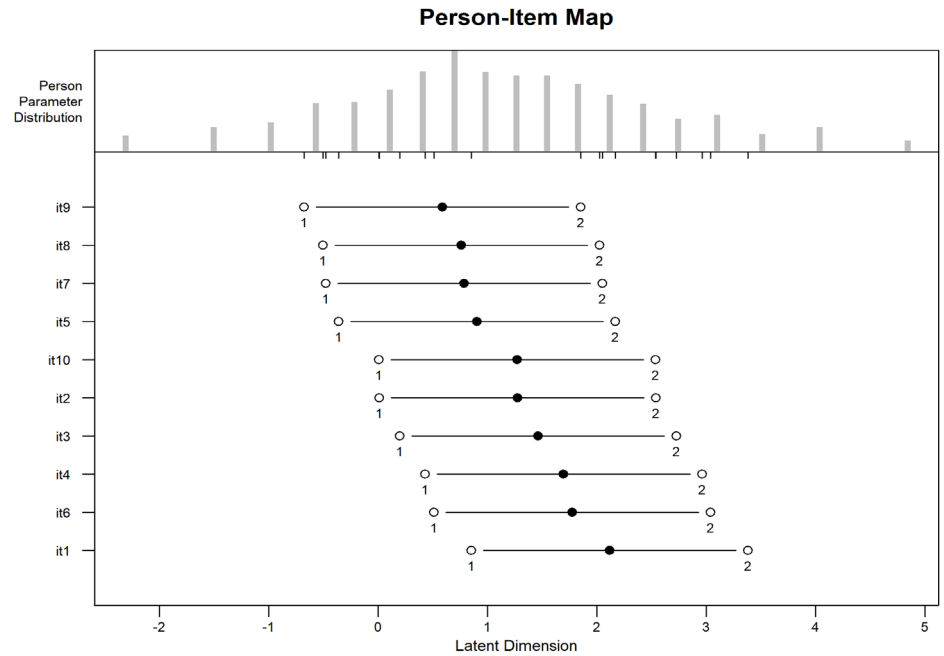

Slika 80 – Karakteristične krive kategorija stavke 1

Pošto je većina stavki nešto teža, prikazana je latentna osobina u rasponu od –6 do 6 logita (argument *xlim=c(–6, 6)*). Ako želite prikaz svih stavki (ali ne na istom grafikonu), onda vam je potreban argument *item.subset="all"*.

#### *10.5.3.5 Uporedni prikaz distribucije parametara ispitanika i stavki*

Na uporednom prikazu stavke smo sortirali po težini (argument *sorted=TRUE*).

```
plotPImap(mRSM.e, sorted = TRUE)
```



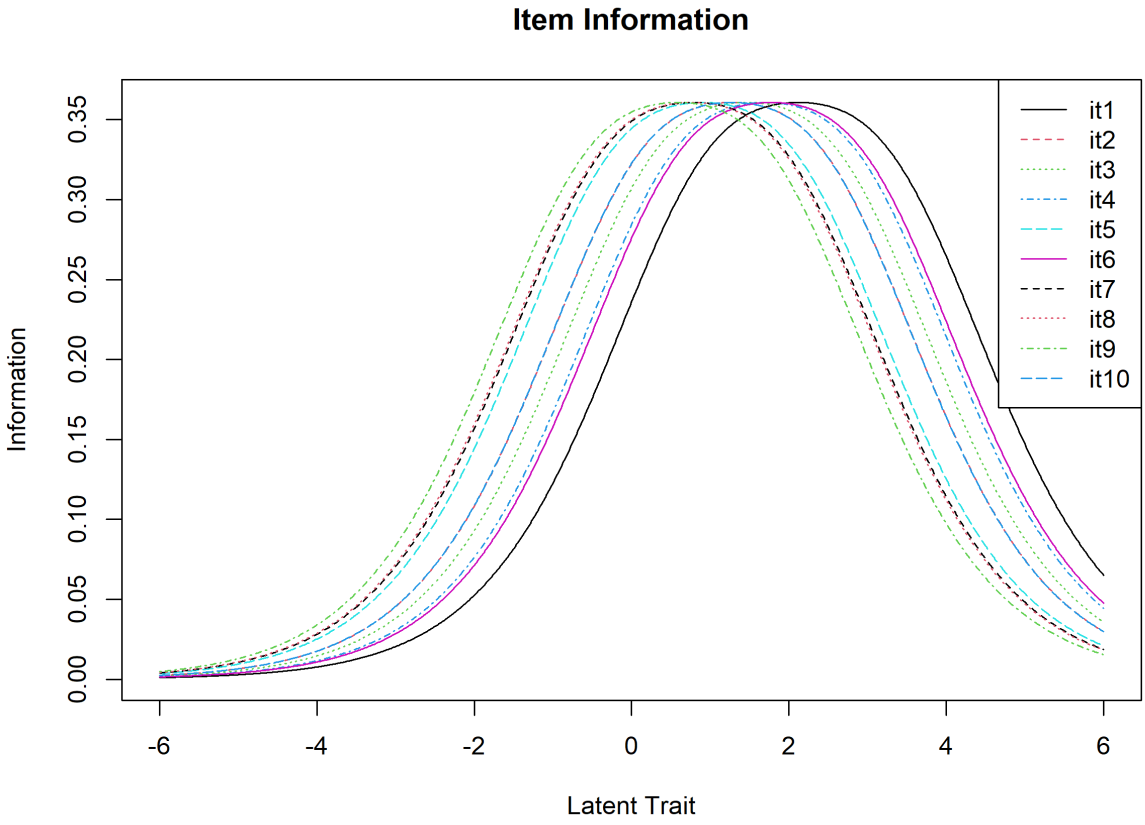

Slika 81 – Uporedni prikaz distribucije parametara ispitanika i stavki

Na grafikonu (Slika 81) su belim kružićima i brojevima označene lokacije pragova, a crnim kružićem lokacija stavke. Vidimo da je stavkama najbolje pokriven natprosečni deo kontinuuma latentne osobine između 0 i 3 logita. Slabije su pokriveni ispitanici sa nivoima osobine ispod proseka. Kada govorimo o pokrivenosti kontinuuma latentne osobine odgovarajućim stavkama u obzir bi trebalo uzeti i parametre kategorija, ne samo težinu stavke.

#### *10.5.3.6 Informativnost stavki i testa*

Ono što smo videli na prethodnom grafikonu potvrđuje i grafikon informativnosti stavki (Slika 82). Većina stavki maksimum funkcije informativnosti postiže u opsegu između –2 i 4 logita.

```
plotINFO(mRSM.e, type="item")
```

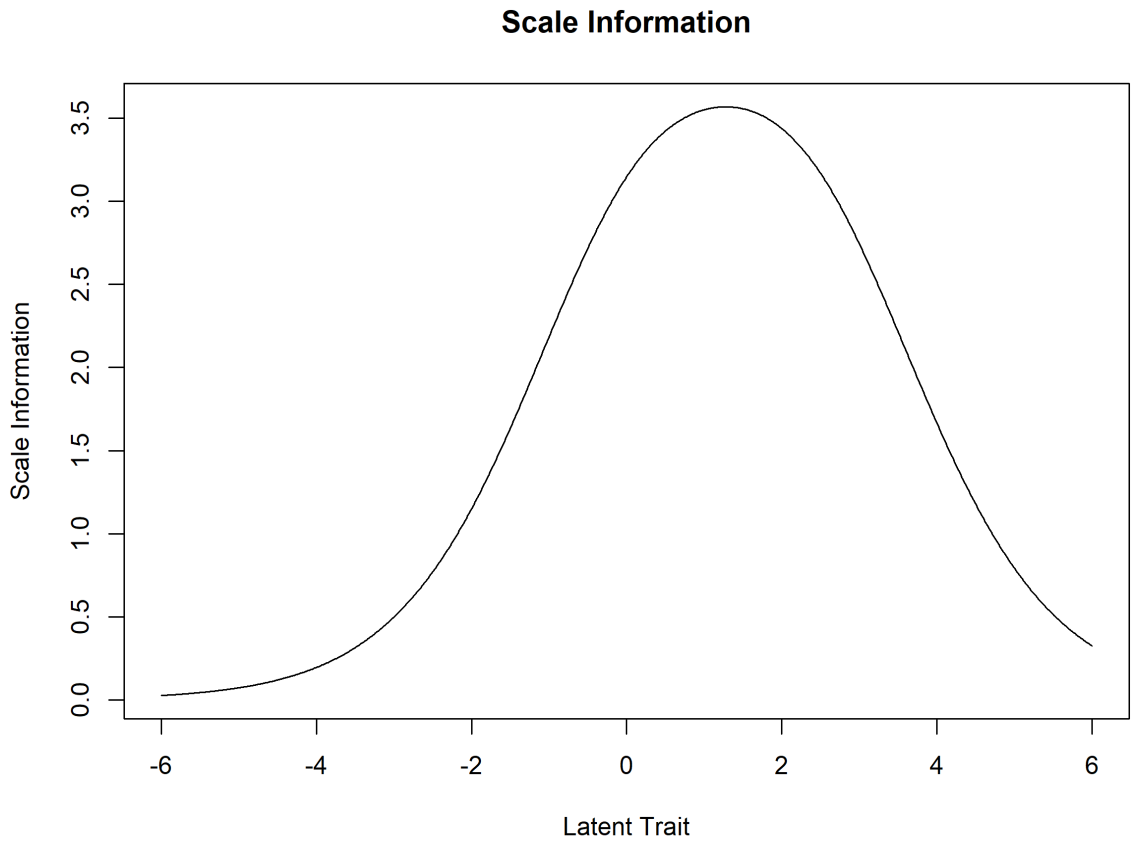



Slika 82 – Funkcije informativnosti stavki u modelu skala procene

Kod politomnih stavki (bez obzira na model) informativnost stavke je suma informativnosti kategorija po nivoima latentne osobine. Kolika će biti maksimalna informativnost zavisi od blizine kategorija. Što su bliže maksimalna informativnost biće veća, ali u užem opsegu latentne osobine. Pošto i ovaj model pripada Rašovoj familiji modela i sve stavke imaju jednaku diskriminativnost i setove pragova, maksimalna informativnost svih stavki biće ista samo je pitanje na kom delu kontinuuma latentne osobine. Najinformativnije će biti na lokaciji stavke.

Informativnost testa je jednaka sumi informativnosti stavki po nivoima osobine pa zavisi od njihovih lokacija (Slika 83).

```
plotINFO(mRSM.e, type="test")
```




Slika 83 – Funkcija informativnosti testa u modelu skala procene

Očekivano, test je najinformativniji u rasponu od –1 do 3 logita (Slika 83). Da bi test precizno merio u širem rasponu osobine u opštoj populaciji, trebalo bi dodati lakše stavke koje maksimalnu informativnost postižu u opsegu od –3 do –1 logita.

#### 10.5.3.6.1 Numerički pokazatelji informativnosti

Ono što smo gore prikazali grafički, izrazićemo numerički. Potrebno je navesti raspon $\theta$ koji nas zanima.

```
TETE <- seq(-6, 6, 0.1)
# Kreiramo vektor sa željenim rasponom i nivoima thete
# Sekvenca od -6 do 6 u koracima po 0.1
TINFO <- test_info(mRSM.e, theta=TETE)
# Vektor TETE prosleđujemo argumentu theta funkcije test_info()
# Rezultat je vektor informativnosti testa za pojedine nivoe latentne osobine
# Ispis tabele informativnosti po nivoima latentne osobine
TI <- data.frame(cbind(TETE, TINFO))
# Sa funkcijama cbind() i data.frame() povezali smo vektore nivoa osobine i
```

```
informativnost u matricu podataka
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n",  TI[TI[,2]==max(TI[,2]),1], "logita")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  1.3 logita

cat("\nInformativnost u rasponu od -6 do 6 logita: \n")

##
## Informativnost u rasponu od -6 do 6 logita:

inf66 <- sum(TINFO)
inf66

## [1] 196.5248

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 6 logita:  \n")

##
## Informativnost u rasponu od 0 do 6 logita:

inf06 <- sum(test_info(mRSM.e, theta=seq(0, 6, 0.1)))
inf06

## [1] 141.7496

cat(" \nProcenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od
-6 do 6 logita:  \n")

##
## Procenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od -6
do 6 logita:

inf06/inf66*100

## [1] 72.1281
```

Vidimo da je informativnost mnogo veća (72,13%) u opsegu natprosečnih nivoa osobine.

#### *10.5.3.7 Eliminacija stavki*

Paket eRm poseduje funkciju za automatsku eliminaciju stavki korak po korak, ali ova funkcija radi isključivo sa binarno skorovanim stavkama.

##### 10.5.3.7.1 Izbacivanje stavki

Kada se odlučimo da izbacimo neku stavku i izbacimo je, promeniće se i ukupni skorovi i nova procena parametara stavki daće drugačije rezultate. Zato bi stavke trebalo uklanjati po jednu u svakom koraku.

```
# Prvo ćemo izbaciti stavku 10 iz matrice podataka. Novu matricu podataka
nazvaćemo p.podR. Redukovanu matricu podataka čuvamo pod drugim imenom kako bismo
sačuvali i originalne podatke. Nije nužno tako, ali je korisno
p.podR <- p.pod[,1:9]
mRSM.eR <- RSM(p.podR, se=TRUE)

## Warning:
## The following items have no 0-responses:
```

```
## it1 it2 it3 it4 it5 it6 it7 it8 it9
## Responses are shifted such that lowest category is 0.

# Uporedni prikaz parametara na originalnom i na redukovanom skupu stavki
round(cbind("originalni"=data.frame(thresholds(mRSM.e)$threshtable),
 "redukovani"=data.frame(rbind(thresholds(mRSM.eR)$threshtable[[1]],
 rep(NA,3)))),3)

##      originalni.X1.Location originalni.X1.Threshold.1 originalni.X1.Threshold.2
## it1                   2.119                     0.854                     3.385
## it2                   1.277                     0.012                     2.542
## it3                   1.464                     0.199                     2.730
## it4                   1.698                     0.432                     2.963
## it5                   0.904                    -0.361                     2.169
## it6                   1.776                     0.511                     3.042
## it7                   0.788                    -0.478                     2.053
## it8                   0.763                    -0.502                     2.028
## it9                   0.590                    -0.676                     1.855
## it10                  1.274                     0.009                     2.539
##      redukovani.Location redukovani.Threshold.1 redukovani.Threshold.2
## it1                2.499                  0.960                  4.038
## it2                1.555                  0.016                  3.094
## it3                1.767                  0.228                  3.306
## it4                2.029                  0.491                  3.568
## it5                1.132                 -0.406                  2.671
## it6                2.118                  0.579                  3.656
## it7                1.000                 -0.539                  2.539
## it8                0.972                 -0.567                  2.511
## it9                0.776                 -0.763                  2.315
## it10                  NA                     NA                     NA

mRSM.eRPI <- person.parameter(mRSM.eR)
mRSM.eRLR <- LRtest(mRSM.eR, splitcr = "mean")
# Izračunali smo Andersenov LR test na redukovanom modelu
mRSM.eRLR

##
## Andersen LR-test:
## LR-value: 120.461
## Chi-square df: 9
## p-value:  0

# I dalje je značajan, mada je vrednost niža nego u originalnom modelu
mRSM.eLR

##
## Andersen LR-test:
## LR-value: 292.654
## Chi-square df: 10
## p-value:  0

itemfit(mRSM.eRPI)

##
## Itemfit Statistics:
##        Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it1 1293.958 984   0.000      1.314     1.342    6.219   7.571   0.420
## it2  705.832 984   1.000      0.717     0.715   -7.357  -7.577   0.597
## it3 1107.271 984   0.004      1.124     1.154    2.797   3.551   0.502
```

```
## it4  648.706 984   1.000      0.659     0.635   -8.973 -10.194   0.582
## it5  592.424 984   1.000      0.601     0.571  -10.816 -12.151   0.581
## it6 1375.802 984   0.000      1.397     1.436    8.106   9.363   0.453
## it7  798.672 984   1.000      0.811     0.808   -4.647  -4.889   0.555
## it8  619.272 984   1.000      0.629     0.603   -9.868 -11.053   0.553
## it9  969.988 984   0.619      0.985     0.986   -0.333  -0.329   0.547
```

Izbacili smo stavku 10. Promenili su se parametri ostalih stavki, ali model i dalje nije saglasan sa podacima. Andersenov LR test je i dalje značajan, a takođe postoji još stavki sa misfitom (stavka 6, 1 i 4). Sledeća stavka koju bi trebalo izbaciti je stavka 6 (značajan hi–kvadrat, šumni autfit i infit).

#### *10.5.3.8 Separaciona pouzdanost*

Separaciona pouzdanost je pouzdanost tipa interne konzistencije koja se izračunava u Rašovim modelima (videti odeljak 7.2). Bliska je, ali ne identična Kronbahovoj alfi. Takođe, u paketu *mirt* izračunaćemo marginalnu i empirijsku pouzdanost (videti odeljak 10.4.3.3), a u paketu *ltm* i Kronbahovu alfu.

```
SEP.POUZD <- SepRel(mRSM.ePI)
# Funkcija SepRel() kao objekat uzima objekat sa parametrima ispitanika
SEP.POUZD

## Separation Reliability: 0.7874

# Za izračunavanje Kronbahove alfe koristićemo funkciju cronbach.alpha() iz paketa
ltm
# Ispisaćemo samo sirovu alfu (zato smo dodali $alpha)
cat("Kronbahova alfa: \n")

## Kronbahova alfa:

ltm::cronbach.alpha(p.pod)$alpha

## [1] 0.8085455

cat("Marginalna pouzdanost: \n")

## Marginalna pouzdanost:

mirt::marginal_rxx(mRSM.m)

## [1] 0.7645136

cat("Empirijska pouzdanost: \n")

## Empirijska pouzdanost:

mirt::empirical_rxx(mirt::fscores(mRSM.m, full.scores.SE = TRUE))

##       F1
## 0.807321
```

Svi vidovi pouzdanosti za ovaj test su zadovoljavajući (>0,7).

Izračunaćemo i broj razdvojivih stratuma (videti odeljak 7.2).

```
G <- sqrt(SEP.POUZD$sep.rel/(1-SEP.POUZD$sep.rel))
cat("G =", G)

## G = 1.924508
```

```
BRS=(4*G+1)/3
cat("\nBroj razdvojivih stratuma =", BRS)

##
## Broj razdvojivih stratuma = 2.899343
```

Na osnovu ovog testa po nivou osobine možemo izdvojiti 3 značajno različita stratuma ispitanika.

#### *10.5.3.9 Parametri (mere) ispitanika*

Iz ranije konstruisanog objekta mRSM.ePI možemo izvući parametre ispitanika. Ovo ima smisla raditi samo ako je model saglasan sa podacima.

```
MERE <- mRSM.ePI$theta.table[,1]
# Iz objekta mr.ePI iz tabele theta.table izdvajamo prvu kolonu {,1}
# Ovu varijablu možemo kasnije koristiti u analizama, npr:
mean(MERE)

## [1] 1.089812

# Prosečna mera u logitima
```

#### *10.5.3.10 Pokazatelji fita za ispitanike*

Izračunaćemo pokazatelje fita za ispitanike.

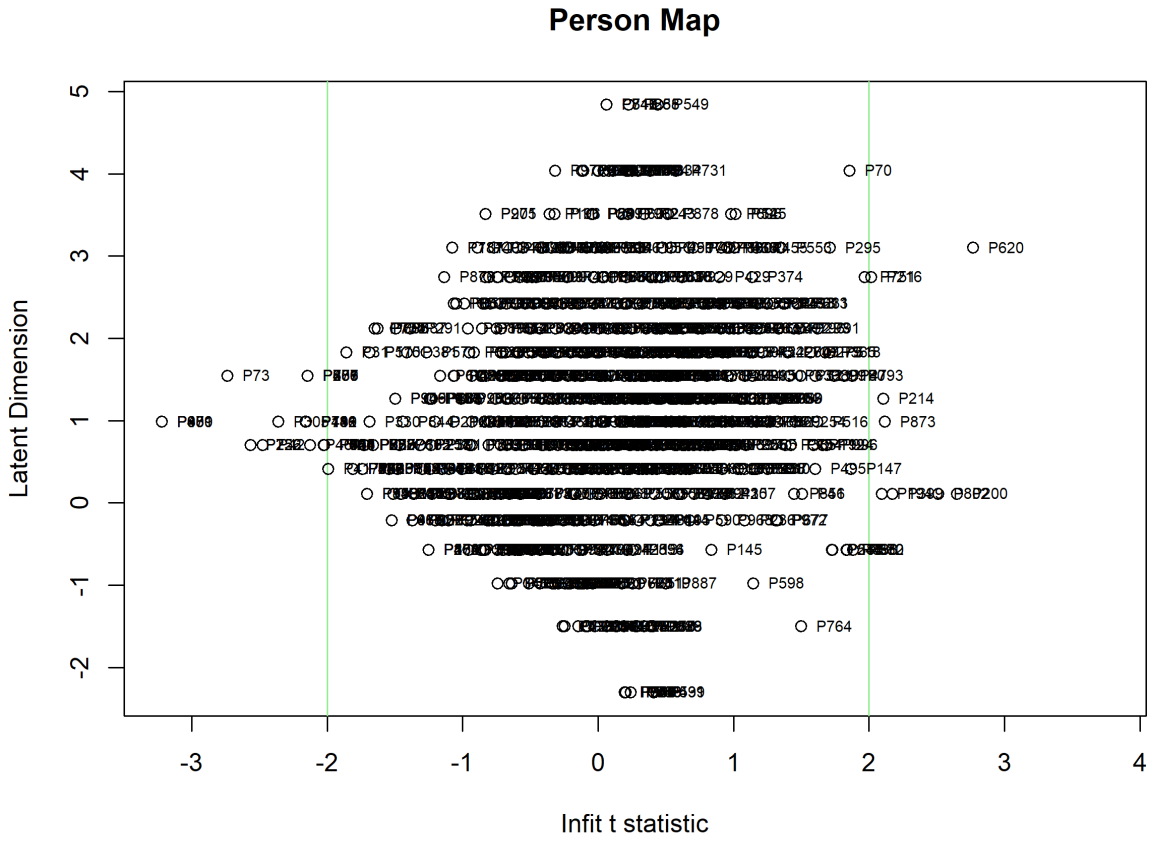
```
fit.isp <- personfit(mRSM.ePI)

Personfit Statistics:
       Chisq df p-value Outfit MSQ Infit MSQ Outfit t Infit t
P1    13.166  9   0.155      1.317     1.245     0.87    0.71
P2     6.342  9   0.705      0.634     0.629    -0.92   -0.93
P3    10.593  9   0.305      1.059     1.118     0.28    0.43
P4     7.526  9   0.582      0.753     0.757    -0.54   -0.53
P5     5.265  9   0.811      0.527     0.501    -1.39   -1.48
P6     9.484  9   0.394      0.948     0.898     0.13    0.01
P7     8.364  9   0.498      0.836     0.835    -0.29   -0.30
P8     8.432  9   0.491      0.843     0.829    -0.32   -0.36
P9     7.024  9   0.635      0.702     0.657    -0.77   -0.92
P10   15.379  9   0.081      1.538     1.497     1.34    1.25
```

Zbog prostora prikazano je samo prvih deset redova.

##### 10.5.3.10.1 Grafikon misfita za ispitanike

Kao i za stavke, može se dobiti grafikon misfita za ispitanike (Slika 84). Prikazan je infit u standardizovanoj metrici. Vertikalne zelene linije ograničavaju prihvatljivi opseg vrednosti. Levo od prihvatljivog opsega su ispitanici čije je odgovaranje suviše predvidljivo (prigušeni šum), a desno oni ispitanici čije odgovaranje model ne predviđa dobro (šum).

```
plotPWmap(mRSM.e, pmap=T, imap=F)
```




Slika 84 – Standardizovani infit za ispitanike

Manji broj ispitanika ima misfit u formi prigušenog šuma, a nešto veći broj u obliku šuma (Slika 84).

### *10.5.3.11 Formiranje sirovog skora i korelacije sa Rašovim merama*

Kreiraćemo ukupan skor i korelirati sa Rašovim merama ispitanika procenjenim na osnovu modela.

```
SKOR <- rowSums(p.pod[1:ncol(p.pod)])
df <- data.frame(cbind(SKOR, MERE))
head(df, 5)

##   SKOR      MERE
## 1   18 0.6999345
## 2   19 0.9839249
## 3   25 2.7441713
## 4   19 0.9839249
## 5   17 0.4092169

#View(RASCH.POD)
cat("Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera: \n")

## Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera:

cor(SKOR, MERE, method="pearson")

## [1] 0.9896955
```

```
cat(" \nSpirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:
\n")
```

```
##
## Spirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:
```

```
cor(SKOR, MERE, method="spearman")
```

```
## [1] 1
```

Pirsonov koeficijent korelacije sumacionog skora i Rašovih mera je vrlo visok, a Spirmanov koeficijent rang korelacije je jednak 1. Poredak ispitanika po sumacionim skorovima i Rašovim merama je isti i u modelu skala procene.

### 10.5.4 Model stepenovanog ocenjivanja (PCM) u paketu *eRM*

Model stepenovanog ocenjivanja je još jedan model iz Rašove familije modela. Za razliku od modela skala procene, ovde koraci (ekvivalent pragova) nisu ekvidistantni za sve stavke i izračunavaju se za svaku posebno. Zbog toga u ovom modelu nije nužno da sve stavke imaju jednak broj kategorija. I ovde najniža kategorija odgovora u podacima mora biti 0. Ako to nije tako, *eRm* će automatski pomeriti sve kategorije da ih prilagodi. Npr. ako su kategorije odgovora u podacima 1, 2 i 3, pretvoriće ih u 0, 1 i 2. Ako na nekoj stavci niko nije ostvario skorove 0 i 1, već postoje samo ajtemski skorovi 2 i 3, to će biti pretvoreno u 0 i 1 (respektivno). Demonstriraćemo to na našim podacima:

```
p.pod2 <- p.pod # Kopiramo podatke u novi objekat da sačuvamo originalne podatke
# Rekodiraćemo stavku 1 tako da vrednosti 1 i 2 postanu 2, a vrednosti 3 nećemo
menjati
p.pod2$it1 <- ifelse(p.pod$it1<3, 2,3)
mPCM.e2 <- PCM(p.pod2, se=T)
```

```
## Warning:
## The following items have no 0-responses:
## it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## Responses are shifted such that lowest category is 0.
```

```
parPCM.e2 <-data.frame(thresholds(mPCM.e2)$threshtable)
parPCM.e2
```

```
##      X1.Location X1.Threshold.1 X1.Threshold.2
## it1   2.50346241      2.5034624             NA
## it2   0.55730338     -1.2897748      2.4043815
## it3   0.66807497     -0.4035493      1.7396992
## it4   1.25048947     -0.9511039      3.4520828
## it5  -0.00465535     -2.3480394      2.3387287
## it6   0.87528940      0.2990416      1.4515372
## it7  -0.06168185     -1.7676159      1.6442522
## it8  -0.24686465     -2.6082308      2.1145015
## it9  -0.21214883     -1.5285415      1.1042438
## it10  0.42347480      1.5957888     -0.7488392
```

U prvoj koloni nalaze se težine (lokacije) stavki koje predstavljaju prosek lokacija koraka. Kod stavke 1, kod koje smo imali samo dve kategorije, imamo samo jedan parametar lokacije koraka (između 0 i 1) i to je ekvivalentno težini u Rašovom modelu za binarno skorovane stavke.

Zbog uporedivosti, analizu ćemo dalje nastaviti na originalnim podacima.

```
mPCM.e <- PCM(p.pod, se=T)

## Warning:
## The following items have no 0-responses:
## it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## Responses are shifted such that lowest category is 0.

matrix(coef(mPCM.e), nrow=10, ncol=2, byrow = TRUE,
 dimnames = list(paste0("it", 1:10), c("c1", "c2")))

##              c1          c2
## it1  -0.4091949 -2.35631888
## it2   1.2786345 -1.06731749
## it3   0.3991091 -1.28585809
## it4   0.9472235 -2.43001714
## it5   2.3266958  0.04061836
## it6  -0.2959992 -1.69010731
## it7   1.7482184  0.14771697
## it8   2.5840026  0.51783729
## it9   1.5075818  0.43994606
## it10 -1.5958427 -0.80692873

# Koeficijenti koji su ovde prikazani nisu lokacije koraka i nisu pogodni za
interpretaciju
# Lokacije koraka dobijamo na sledeći način
parPCM.e <-data.frame(thresholds(mPCM.e)$threshtable)
parPCM.e

##      X1.Location X1.Threshold.1 X1.Threshold.2
## it1   1.17815944      0.4091949       1.947124
## it2   0.53365875     -1.2786345       2.345952
## it3   0.64292905     -0.3991091       1.684967
## it4   1.21500857     -0.9472235       3.377241
## it5  -0.02030918     -2.3266958       2.286077
## it6   0.84505365      0.2959992       1.394108
## it7  -0.07385848     -1.7482184       1.600501
## it8  -0.25891864     -2.5840026       2.066165
## it9  -0.21997303     -1.5075818       1.067636
## it10  0.40346436      1.5958427      -0.788914

names(parPCM.e) <- c("bj", "b1", "b2")
# Od lokacije drugog koraka oduzećemo lokaciju prvog da vidimo interval između
njih (ovde neće biti jednak za sve stavke kao u RSM)
parPCM.e$interval <- parPCM.e$b2-parPCM.e$b1
# Od lokacije prvog koraka oduzećemo lokaciju stavke da dobijemo vrednost
parametra koraka
parPCM.e$t1 <- parPCM.e$b1-parPCM.e$bj
parPCM.e$t2 <- parPCM.e$b2-parPCM.e$bj # Ponovićemo to isto za drugi korak
# Izračunaćemo sume koraka
cat("Sume koraka su: \n")

## Sume koraka su:

parPCM.e$sumaT <- parPCM.e$t1+parPCM.e$t2
round(parPCM.e,3)

##          bj     b1     b2 interval     t1     t2 sumaT
## it1   1.178  0.409  1.947    1.538 -0.769  0.769     0
```

```
## it2   0.534 -1.279  2.346    3.625 -1.812  1.812     0
## it3   0.643 -0.399  1.685    2.084 -1.042  1.042     0
## it4   1.215 -0.947  3.377    4.324 -2.162  2.162     0
## it5  -0.020 -2.327  2.286    4.613 -2.306  2.306     0
## it6   0.845  0.296  1.394    1.098 -0.549  0.549     0
## it7  -0.074 -1.748  1.601    3.349 -1.674  1.674     0
## it8  -0.259 -2.584  2.066    4.650 -2.325  2.325     0
## it9  -0.220 -1.508  1.068    2.575 -1.288  1.288     0
## it10  0.403  1.596 -0.789   -2.385  1.192 -1.192     0
```

Intervali između lokacija koraka nisu jednaki. Sume koraka jednake su 0.

Postupak procene istog modela u paketu *mirt*:

```
mPCM.m <- mirt::mirt(p.pod, 1, itemtype = "Rasch", verbose = FALSE)
koefM <- coef(mPCM.m, simplify=TRUE, IRTpars=TRUE)$items[,]
# Poređenje lokacija koraka iz paketa eRm i mirt
razl <- round(cbind(parPCM.e[,2:3], koefM[,-1],
 "b1dif"=round(parPCM.e[,2]-koefM[,2],1),
 "b2dif"=round(parPCM.e[,3]-koefM[,3],1)),2)
names(razl) <- c("b1.eRm", "b2.eRm", "b1.mirt", "b2.mirt", names(razl)[5:6] )
razl

##      b1.eRm b2.eRm b1.mirt b2.mirt b1dif b2dif
## it1    0.41   1.95    0.12    1.68   0.3   0.3
## it2   -1.28   2.35   -1.58    2.06   0.3   0.3
## it3   -0.40   1.68   -0.71    1.42   0.3   0.3
## it4   -0.95   3.38   -1.25    3.06   0.3   0.3
## it5   -2.33   2.29   -2.60    2.00   0.3   0.3
## it6    0.30   1.39    0.00    1.14   0.3   0.3
## it7   -1.75   1.60   -2.04    1.32   0.3   0.3
## it8   -2.58   2.07   -2.85    1.78   0.3   0.3
## it9   -1.51   1.07   -1.81    0.79   0.3   0.3
## it10   1.60  -0.79    1.28   -1.04   0.3   0.3
```

Lokacije koraka procenjene u paketima *eRm* i *mirt* razlikuju se za 0,3 (zaokruženo na prvu decimalu).

#### *10.5.4.1 Pokazatelj fita modela u celini*

Izračunaćemo Andersenov LR statistik kao pokazatelj fita modela u celini. Zbog poređenja prikazaćemo i $M_2$ statistik za model izračunat u paketu *mirt*.

```
mPCM.eLR <- LRtest(mPCM.e, splitcr = "mean")
# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
mPCM.eLR

##
## Andersen LR-test:
## LR-value: 101.497
## Chi-square df: 19
## p-value:  0

mirt::M2(mPCM.m)

##             M2 df         p      RMSEA RMSEA_5 RMSEA_95      SRMSR       TLI
## stats 40.36354 34 0.2095523 0.01368759       0 0.027902 0.08002099 0.9979877
```

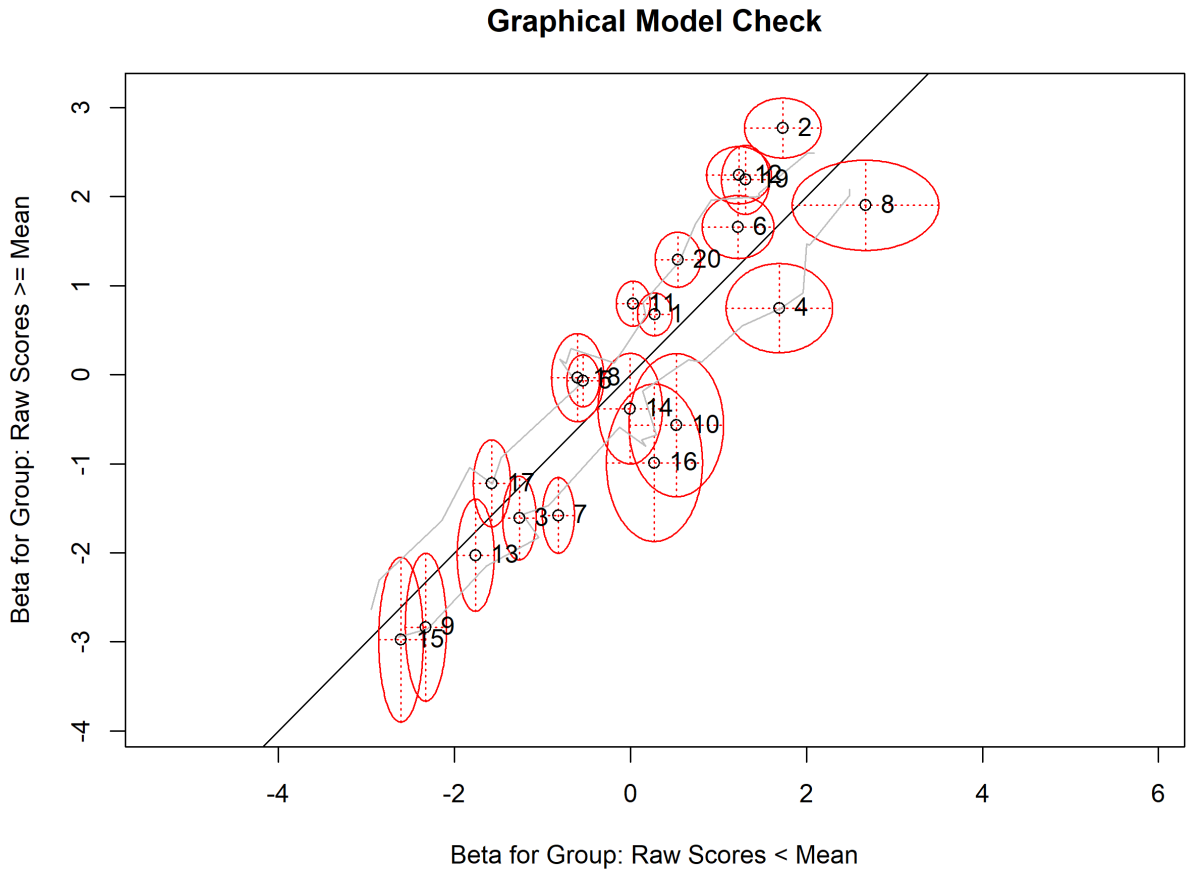
```
##             CFI
## stats 0.9980452
```

Prema Andersenovom LR testu model nije saglasan sa podacima. Kod modela procenjenog u paketu *mirt* pokazatelj bliskog fita $M_2$ kaže drugačije. I pokazatelji približnog fita ukazuju da je model saglasan sa podacima. Jedino *SRMSR* ima graničnu vrednost. Podsetićemo da *eRm* i *mirt* na različite načine procenjuju parametre stavki (CMLE i MMLE respektivno). CMLE računa uslovne verovatnoće odgovora za iste sumacione skorove, dok MMLE uzima u obzir marginalnu distribuciju merene osobine, za koju se pretpostavlja da je standardizovana normalna distribucija.

#### 10.5.4.1.1 Grafička provera modela

I za model stepenovanog ocenjivanja važi pretpostavka invarijantnosti procene parametara. Pogledaćemo grafikone zasnovane na Andersenovom LR testu koji testira tu pretpostavku.

```
plotGOF(mPCM.eLR, ctrline = list(col = "gray"), conf = list(), tlab ="number" )
```

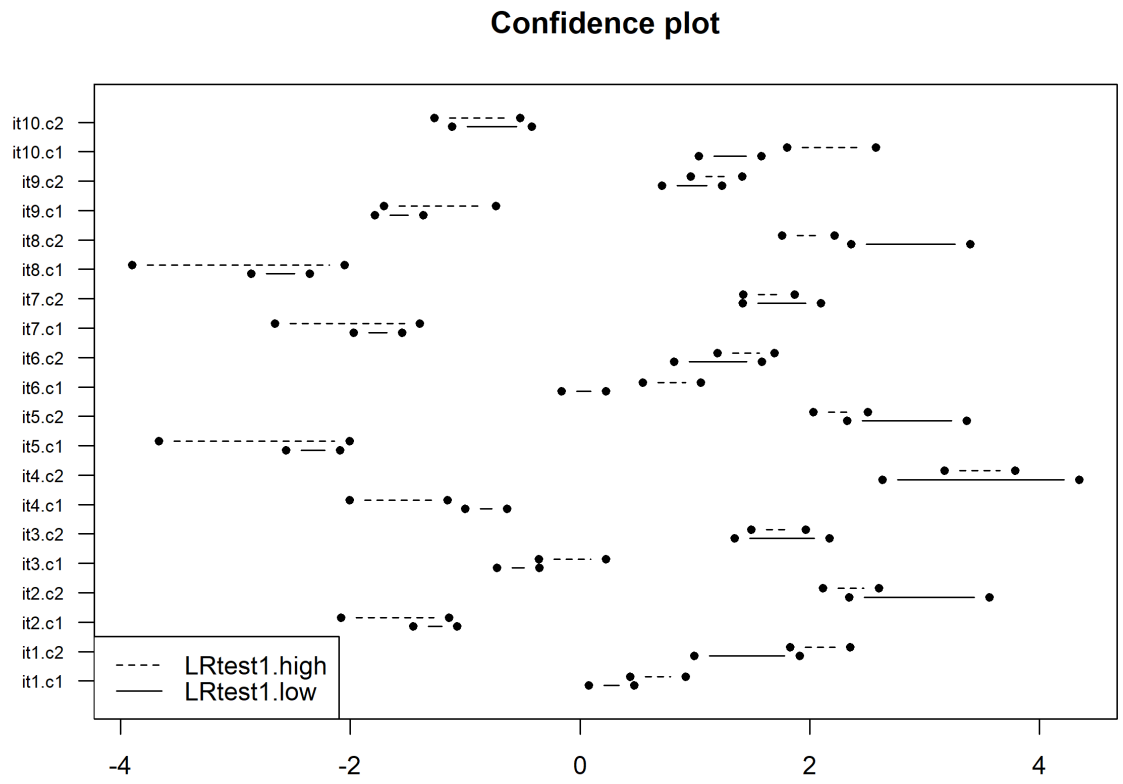



Slika 85 – Grafička provera fita modela (invarijantnost procene parametara)

Zbog preglednosti smo postavili argument *tlab="number"*, pa su umesto naziva ispisani redni brojevi parametara (Slika 85). Kako u našem slučaju imamo dva parametra koraka po stavci, brojevi 1 i 2 označavaju prvi i drugi parametar lokacije koraka prve stavke, brojevi 3 i 4 prvi i drugi parametar koraka druge stavke i tako dalje. Uočavamo da od pretpostavke invarijantnosti procene parametara odstupa parametar

lokacije koraka 1 četvrte stavke, parametar koraka 1 šeste stavke, parametar koraka 2 osme stavke i parametar koraka 1 desete stavke. To se bolje vidi na grafikonu sa uporednim prikazom intervala poverenja procene parametara na poduzorcima ispitanika sa skorom nižim od prosečnog i kod onih sa višim skorom (Slika 86 – kriterijum podele je definisan prilikom zadavanja LR testa, a ova funkcija koristi objekat u kojem smo sačuvali LR test).

```
plotDIF(mPCM.eLR, leg=TRUE)
```




Slika 86 – Intervali poverenja procene parametara u grupama sa niskim i visokim skorom

Obratite pažnju na činjenicu da su u poduzorku sa skorom ispod proseka, 95% intervali poverenja procene parametara širi za parametre koji su više na skali latentne osobine i obrnuto (Slika 86). Podsećamo, parametri stavki biće preciznije procenjeni ako su upareni sa uzorkom ispitanika i obrnuto.

### *10.5.4.2 Pokazatelji fita za stavke*

Pogledaćemo numeričke pokazatelje fita za stavke. Za to nam je potrebno da prethodno procenimo parametre ispitanika. Sačuvaćemo ih u objekat *mPCM.ePI*.

```
mPCM.ePI <- person.parameter(mPCM.e)
mPCM.ePI
```

```
## Person Parameters:
##
##  Raw Score    Estimate Std.Error
##          0 -4.53364270        NA
##          1 -3.57048958 1.0940711
##          2 -2.67502377 0.8409219
##          3 -2.05845231 0.7400931
##          4 -1.55467766 0.6831703
##          5 -1.11656808 0.6417872
##          6 -0.72776173 0.6056518
##          7 -0.38068425 0.5732278
##          8 -0.06756479 0.5470989
##          9  0.22152882 0.5297790
##         10  0.49729685 0.5220508
##         11  0.76974888 0.5232849
##         12  1.04763786 0.5321694
##         13  1.33850994 0.5474484
##         14  1.64942837 0.5687657
##         15  1.98870943 0.5976199
##         16  2.36921926 0.6385237
##         17  2.81515590 0.7017399
##         18  3.38100798 0.8137670
##         19  4.23869135 1.0806082
##         20  5.16704016        NA
```

U gornjoj tabeli imamo procene Rašovih mera za svaki sumacioni skor i standardne greške merenja vezane za taj skor (i nivo osobine). *eRm* ne može da izračuna standardne greške za savršene skorove. U nekim softverima se takvim skorovima dodaje verovatnoća 0,05 (ako je skor 0) ili se ista verovatnoća oduzima, ako je ispitanik imao maksimalni skor. Kada imamo procenjene parametre ispitanika možemo pristupiti izračunavanju pokazatelja fita za stavke.

```
# Pošto i eRm i mirt imaju funkciju itemfit() moguće je da funkcija neće raditi
ako ste posle eRm paketa učitali paket mirt. U tom slučaju morate ukloniti mirt
sa:
#detach("package:mirt", unload=TRUE)
# ili umesto donje naredbe napisati eRm::itemfit(mr.ePI) i tako funkciju pozvati
direktno iz tog paketa
ItF <- itemfit(mPCM.ePI)
ItF

##
## Itemfit Statistics:
##         Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it1  1192.828 985   0.000      1.210     1.113    3.318   2.610   0.424
## it2   794.655 985   1.000      0.806     0.816   -4.528  -4.405   0.603
## it3   984.820 985   0.496      0.999     0.995   -0.010  -0.112   0.505
## it4   759.084 985   1.000      0.770     0.791   -5.089  -5.011   0.592
## it5   747.180 985   1.000      0.758     0.780   -4.830  -4.726   0.586
## it6  1171.316 985   0.000      1.188     1.103    3.005   2.456   0.453
## it7   862.800 985   0.998      0.875     0.876   -2.931  -2.999   0.556
## it8   794.082 985   1.000      0.805     0.825   -3.848  -3.773   0.556
## it9   923.029 985   0.921      0.936     0.920   -1.446  -1.985   0.545
## it10 1126.550 985   0.001      1.143     0.984    1.172  -0.380   0.547

# Argument splitcr = "mean" određuje kako će biti izvršena podela testa na
poduzorke. U ovom slučaju je kriterijum aritmetička sredina, ali mogući su
medijana ("median") ili svi opaženi ukupni skorovi ("all.r")
```

Značajan misfit na osnovu hi–kvadrat testa imaju stavke: 1, 6 i 10. Na osnovu pokazatelja infita i autfita (ako kao opseg prihvatljivih vrednosti MSQ misfita koristimo 0,7–1,3) nijedna stavka ne izlazi iz dozvoljenih granica. Najbliže tome su stavke koje imaju značajan hi–kvadrat. Među njima, stavka 1 ima najviše vrednosti standardizovanog misfita u formi šuma, a posle nje stavka 6. Kod većine ostalih stavki misfit u obe metrike teži ka overfitu. Kao i kod RSM ajtem–total korelacije stavki 1 i 6 su nešto niže od ostalih. Setimo se da su u analizi dimenzionalnosti te dve stavke odudarale od drugih.

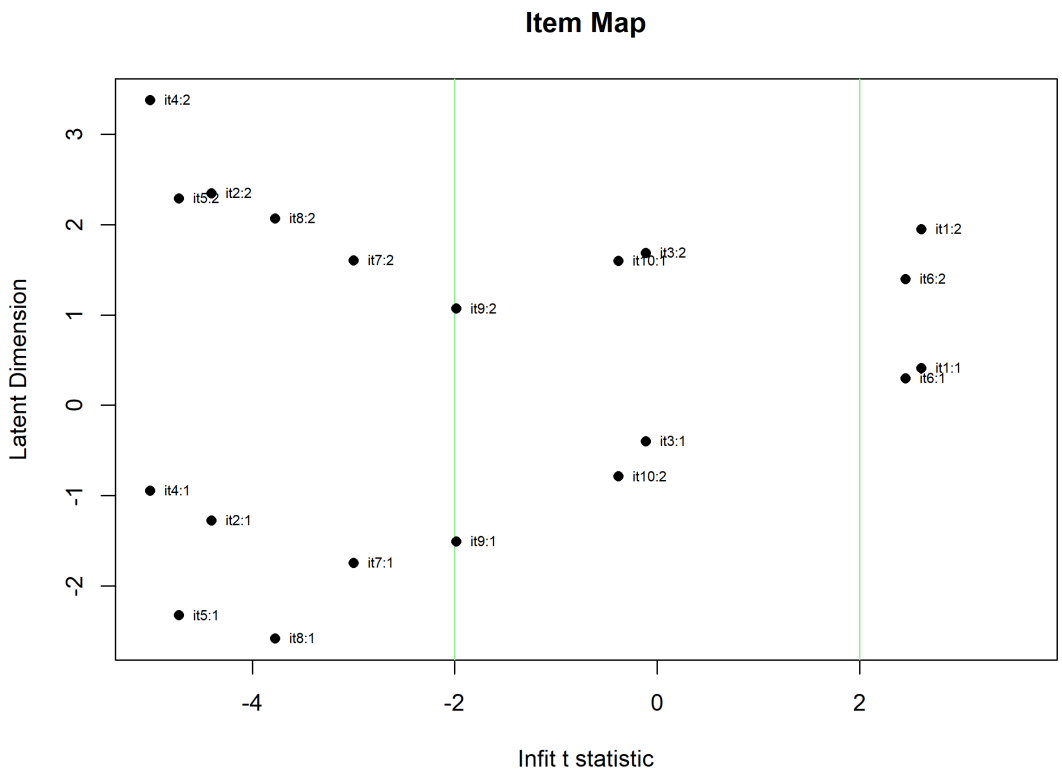
```
WT <- Waldtest(mPCM.e,splitcr = "mean")
# Argument splitcr ima isto značenje i opcije kao u funkciji LRtest()
WT

##
## Wald test on item level (z-values):
##
##              z-statistic p-value
## beta it1.c1        2.537   0.011
## beta it1.c2        3.703   0.000
## beta it2.c1       -1.341   0.180
## beta it2.c2       -2.353   0.019
## beta it3.c1        2.665   0.008
## beta it3.c2        1.588   0.112
## beta it4.c1       -3.214   0.001
## beta it4.c2       -1.542   0.123
## beta it5.c1       -1.158   0.247
## beta it5.c2       -2.201   0.028
## beta it6.c1        4.722   0.000
## beta it6.c2        4.040   0.000
## beta it7.c1       -0.787   0.431
## beta it7.c2       -1.026   0.305
## beta it8.c1       -0.749   0.454
## beta it8.c2       -2.376   0.017
## beta it9.c1        1.306   0.192
## beta it9.c2        1.922   0.055
## beta it10.c1       3.646   0.000
## beta it10.c2       3.667   0.000
```

Prema Valdovom statistiku značajan je misfit za oba parametra lokacije koraka stavki 6, 10 i 1 (redosled po visini misfita). Kod stavki 2–5 i 8, postoji značajan misfit za po jedan parametar lokacije koraka.

#### 10.5.4.2.1 Grafički prikaz standardizovanog infita za stavke

Ono što je rečeno kod opisa numeričkih pokazatelja fita vidi se na narednom grafikonu standardizovanog infita (Slika 87).

```
plotPWmap(mPCM.e, imap=TRUE, pmap=FALSE)
```




Slika 87 – Standardizovani infit za stavke

```
# Argument imap=TRUE određuje da će biti prikazan infit za stavke, a pmap=FALSE da
neće biti prikazan za ispitanike. Moguće je napraviti isti grafikon i za
ispitanike (imap=FALSE, pmap=TRUE), ili grafikon koji će objediniti i stavke i
ispitanike (imap=TRUE, pmap=TRUE)
# Ukoliko želite grafikon i za ispitanike potrebno je dodati argument pp=mr.ePI
# mr.ePI je objekat u kojem smo sačuvali parametre ispitanika
```

#### *10.5.4.3 Procena lokalne zavisnosti*

Prikazaćemo pokazatelje lokalne zavisnosti iz paketa *mirt*:

```
mirt::residuals(mPCM.m, type="LD")

## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.134  -0.054   0.042   0.010   0.071   0.121
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.047 -0.084 -0.041  0.036 -0.115 -0.031 -0.071 -0.091 -0.126
## it2   4.462     NA  0.059  0.105  0.121 -0.037  0.065  0.121  0.081  0.071
## it3  14.156  7.033     NA  0.055  0.053 -0.101 -0.054  0.075 -0.060 -0.100
## it4   3.344 21.986  6.004     NA  0.118 -0.041  0.042  0.096  0.075  0.069
## it5   2.631 29.396  5.544 28.020     NA -0.044  0.103  0.106  0.079  0.039
## it6  26.348  2.790 20.437  3.325  3.843     NA -0.022  0.051 -0.101 -0.134
## it7   1.932  8.457  5.935  3.604 21.323  0.942     NA  0.066  0.057 -0.048
```

```
## it8   9.953 29.049 11.400 18.340 22.459  5.243  8.724     NA  0.068  0.047
## it9  16.528 13.152  7.297 11.242 12.471 20.554  6.586  9.230     NA -0.073
## it10 31.925 10.034 19.844  9.404  2.966 35.833  4.604  4.346 10.558     NA

#mirt::residuals(mPCM.m, type="LD", suppress=.212)
#mirt::residuals(mPCM.m, type="LD", suppress=.354)
```

Pokazatelji zasnovani na hi–kvadratu i Kramerovom V koeficijentu ne nalaze narušavanje pretpostavke lokalne nezavisnosti, bilo da kao graničnu vrednost uzmemo vrednost Kramerovog V za veliki ili srednji efekat.

```
Q3m <- as.matrix(mirt::residuals(mPCM.m, type="Q3"))

## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.219  -0.139  -0.089  -0.086  -0.031   0.026
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.151 -0.114 -0.099 -0.073 -0.121 -0.079 -0.167 -0.151 -0.184
## it2  -0.151  1.000 -0.090  0.005  0.026 -0.153 -0.049  0.019 -0.020 -0.095
## it3  -0.114 -0.090  1.000 -0.076 -0.097 -0.153 -0.118 -0.052 -0.090 -0.174
## it4  -0.099  0.005 -0.076  1.000  0.023 -0.089 -0.095 -0.011 -0.020 -0.076
## it5  -0.073  0.026 -0.097  0.023  1.000 -0.139 -0.009  0.000 -0.031 -0.072
## it6  -0.121 -0.153 -0.153 -0.089 -0.139  1.000 -0.064 -0.072 -0.167 -0.219
## it7  -0.079 -0.049 -0.118 -0.095 -0.009 -0.064  1.000 -0.031 -0.079 -0.146
## it8  -0.167  0.019 -0.052 -0.011  0.000 -0.072 -0.031  1.000 -0.028 -0.102
## it9  -0.151 -0.020 -0.090 -0.020 -0.031 -0.167 -0.079 -0.028  1.000 -0.173
## it10 -0.184 -0.095 -0.174 -0.076 -0.072 -0.219 -0.146 -0.102 -0.173  1.000

diag(Q3m) <- NA
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", round(Q3mean, 3), "\n\n")

##
## Prosečna vrednost Q3: -0.086

Q3c <- Q3m-Q3mean
Q3c

##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.066 -0.028 -0.013  0.012 -0.035  0.007 -0.081 -0.065 -0.098
## it2  -0.066     NA -0.004  0.091  0.112 -0.068  0.037  0.105  0.066 -0.009
## it3  -0.028 -0.004     NA  0.010 -0.012 -0.068 -0.032  0.034 -0.004 -0.089
## it4  -0.013  0.091  0.010     NA  0.109 -0.003 -0.010  0.075  0.065  0.009
## it5   0.012  0.112 -0.012  0.109     NA -0.053  0.076  0.086  0.054  0.014
## it6  -0.035 -0.068 -0.068 -0.003 -0.053     NA  0.022  0.014 -0.081 -0.133
## it7   0.007  0.037 -0.032 -0.010  0.076  0.022     NA  0.055  0.006 -0.060
## it8  -0.081  0.105  0.034  0.075  0.086  0.014  0.055     NA  0.058 -0.016
## it9  -0.065  0.066 -0.004  0.065  0.054 -0.081  0.006  0.058     NA -0.087
## it10 -0.098 -0.009 -0.089  0.009  0.014 -0.133 -0.060 -0.016 -0.087     NA

Q3c[Q3c<.20] <- NA
Q3c

##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

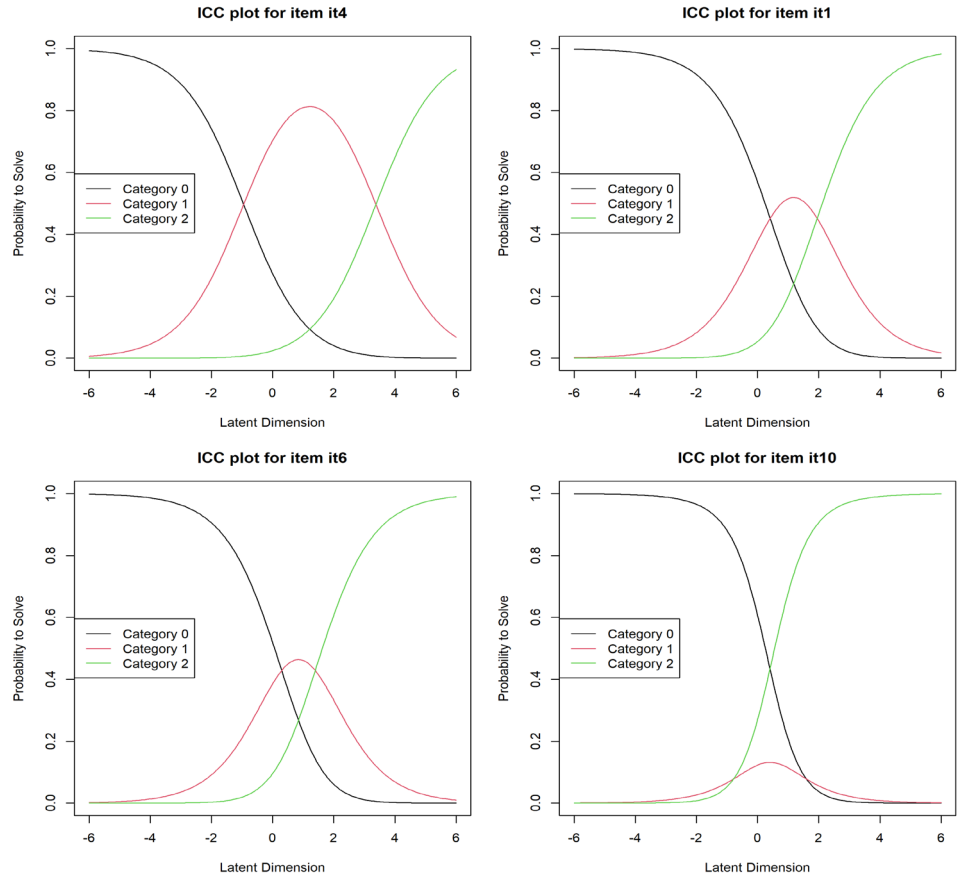
```
## it5   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## it6   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## it7   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## it8   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## it9   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## it10  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
```

Kada vrednosti $Q_3$ pokazatelja korigujemo za prosečnu vrednost u matrici ne postoje vrednosti veće od 0,2, tako da ni na ovaj način ne nalazimo znakove narušavanja lokalne nezavisnosti.

#### *10.5.4.4 Grafički prikaz karakterističnih kriva kategorija*

U modelu stepenovanog odgovaranja kao i u modelu skala procene dobićemo prikaz karakterističnih kriva kategorija, a ne stavki (Slika 88). Za razliku od modela skala procene, u ovom model KKK se razlikuju i po razmacima pragova, a ne samo po lokaciji. Rekli smo (odeljak 4.2.1.3) da koraci u modelu za stepenovano ocenjivanje ne moraju biti ni ekvidistantni ni sekvencijalni. Npr. lokacija koraka treće kategorije može na kontinuumu latentne crte biti pre koraka druge kategorije. Ako su koraci međusobno bliski, onda se može desiti da kategorija koja se nalazi između njih nikada ne bude najverovatniji odgovor. To sugeriše da bi stavku trebalo drugačije skorovati.

```
par(mfrow=c(2,2))
plotICC(mPCM.e, item.subset = 4, xlim=c(-6, 6))
plotICC(mPCM.e, item.subset = 1, xlim=c(-6, 6))
plotICC(mPCM.e, item.subset = 6, xlim=c(-6, 6))
plotICC(mPCM.e, item.subset = 10, xlim=c(-6, 6))
```




Slika 88 – Karakteristične krive kategorija stavki 4, 1, 6 i 10

```
par(mfrow=c(1,1))
# Ako želite prikaz svih stavki onda vam je potreban argument item.subset="all"
# Svaka stavka će biti prikazana na posebnom grafikonu
#plotICC(mPCM.e, item.subset = "all")
```

Pošto je većina stavki nešto teža, prikazana je latentna osobina u rasponu od –6 do 6 logita (argument *xlim=c(–6, 6)*). Prikazani su grafikoni za stavke 4, 1, 6 i 10. Stavka 4 je odabrana jer ima najveći interval između koraka (Slika 88 gore levo). Vidimo da je kod te stavke odgovor u kategoriji 1 (podsetimo odgovori od 1–3 su pomereni u kategorije 0–2) u relativno širokom rasponu latentne osobine najverovatniji. Uporedimo to sa stavkom 1 (Slika 88 gore desno) ili 6 (Slika 88 dole levo) gde je odgovor u toj kategoriji u vrlo malom rasponu latentne osobine najverovatniji. Poseban slučaj je stavka 10 (Slika 88 dole desno) u kojoj odgovor u kategoriji skora 1 nikada nije najverovatniji. Kod ove stavke došlo je do poremećaja redosleda koraka (koji je u ovom modelu dozvoljen i moguć). Parmetar lokacije drugog koraka ($b_{10\_2}$=–0,79) na kontinuumu latentne osobine dolazi pre prvog ($b_{10\_1}$=1,60). Zašto do toga dolazi? Setimo se načina ocenjivanja u modelu stepenovanog ocenjivanja. Zadaci koji se ocenjuju na ovaj način sastavljeni su od operacija koje bi trebalo da su poređane po težini, odnosno po nivou latentne osobine koji je potreban da bi bile pravilno izvršene. Moguće je da procena težine operacija nije bila dobra pa da se teža operacija ispostavi lakšom ili bar jednako teškom kao neka koja je procenjena kao teža. Npr. procenili smo da je u zadatku koji sadrži stepenovanje (1) i logaritmovanje (2) redosled operacija od lakše ka težoj onaj kojim su ovde navedene, a ispostavi se da su stepenovanje i logaritmovanje jednako teške operacije. U tom slučaju (skoro) svi koji su uradili stepenovanje, uradili su i logaritmovanje. Kategorija skora 1 neće biti zastupljena ili će biti zastupljena u vrlo malom procentu (proporciji). Kada se takva kategorija nalazi između koraka dve kategorije, može doći do poremećaja redosleda. Iskoristićemo funkciju *descript()* iz paketa *ltm* da vidimo opažene proporcije odgovora po kategorijama u našim podacima.

```
ltm::descript(p.pod)$perc
```

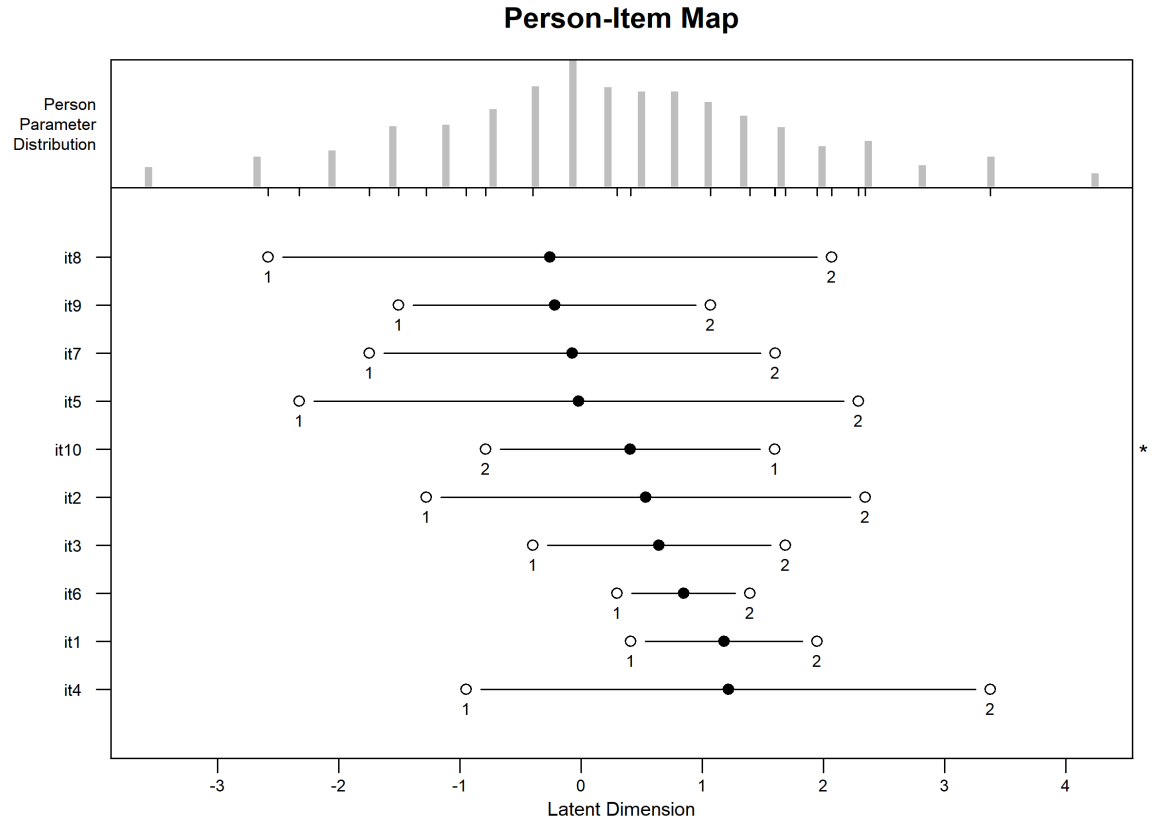
```
##          1     2     3
## it1  0.481 0.367 0.152
## it2  0.211 0.640 0.149
## it3  0.329 0.465 0.206
## it4  0.267 0.664 0.069
## it5  0.104 0.732 0.164
## it6  0.439 0.345 0.216
## it7  0.148 0.606 0.246
## it8  0.085 0.724 0.191
## it9  0.163 0.512 0.325
## it10 0.483 0.095 0.422
```

Vidimo da se kod stavke 10 srednja kategorija (2) javlja samo u 9,5% slučajeva.

#### *10.5.4.5 Uporedni prikaz distribucije parametara ispitanika i stavki*

Na uporednom prikazu distribucije parametara ispitanika i stavki, stavke smo sortirali po težini (argument *sorted=TRUE*) (Slika 89).

```
plotPImap(mPCM.e, sorted = TRUE)
```



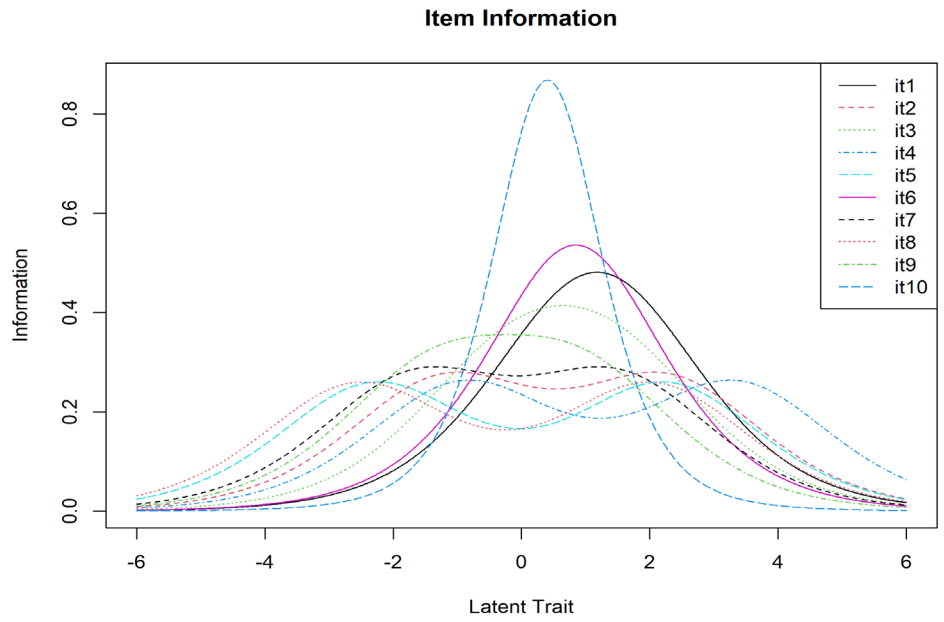

Slika 89 – Uporedni prikaz parametara ispitanika i stavki

Na grafikonu (Slika 89) su belim kružićima i brojevima označene lokacije koraka, a crnim kružićem lokacija stavke. Vidimo da je stavkama najbolje pokriven natprosečni deo kontinuuma latentne osobine između 0 i 3 logita. Slabije su pokriveni ispitanici sa nivoima osobine ispod proseka. Neke stavke obuhvataju širi opseg osobine (8 i 5), a neke uzak (1, 3 i 6). Primetite poremećaj redosleda pragova kod stavke 10. I ovde pokrivenost osobine zavisi i od raspona između lokacija koraka.

#### *10.5.4.6 Informativnost stavki i testa*

Ono što smo videli na prethodnom grafikonu (Slika 89) potvrđuje i grafikon informativnosti stavki (Slika 90). Krive informativnosti stavki u PCM modelu su simetrične kao i u RSM. Pošto se informativnosti kategorija po nivoima osobine sabiraju da bi dale informativnost stavke, ovde ne mora nužno postojati jedan nivo osobine na kojoj stavka ima maksimalnu informativnost. Od razmaka lokacija koraka zavisiće izgled krive informativnosti. Ako su koraci razmaknutiji, odnosno kategorije udaljenije na kontinuumu latentne osobine, onda će kriva informativnosti biti niža, ali šira. Stavka će doprinositi informativnosti testa u širem intervalu merene osobine. Kriva informativnosti takve stavke može biti polimodalna (kao krive stavki 7, 2, 4, 5 i 8).

```
plotINFO(mPCM.e, type="item")
```

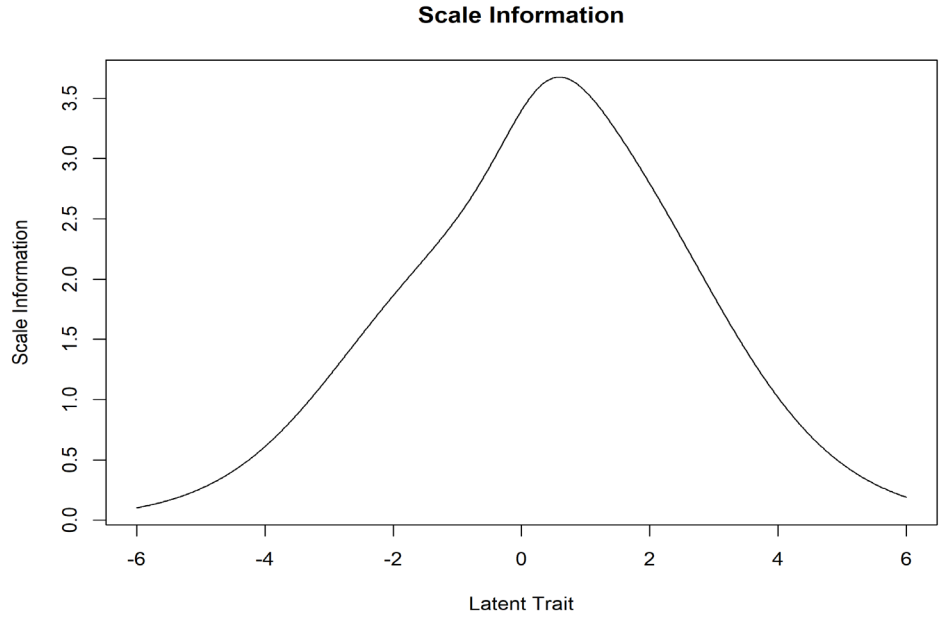



Slika 90 – Funkcije informativnosti stavki u modelu za stepenovano ocenjivanje

Krive stavki čiji su pragovi bliži biće više i obično unimodalne (jer se sabiraju informativnosti kategorija koje su bliske) (Slika 90). Vrh takve krive biće na proseku lokacija koraka (videti stavke 1, 6 i 10). Zanimljivo je da su baš ove tri navedene stavke one koje pokazuju misfit. Stavka 10 funkcioniše kao dihotomna stavka. Kategorija skora 2 (kategorija 1 kako je označava *eRm*) se vrlo retko javlja i nikada nije najverovatnija (videti Slika 88 dole desno i tabelu sa proporcijama odgovora po kategorijama u odeljku 10.5.4.4). Preostale dve stavke najmanje koreliraju sa predmetom merenja. PCM je jednoparametarski model i ne procenjuje parametar diskriminativnosti, ali to možemo videti na osnovu opterećenja na faktoru u jednofaktorskom rešenju (videti tabelu faktorskih opterećenja u odeljku 10.5.2.1 ili parametre diskriminativnosti stavki u GPCM modelu u odeljku 10.5.5). U tabeli sa proporcijama odgovora po kategorijama (pomenuta tabela proporcija u odeljku 10.5.4.4) vidimo da ove tri stavke imaju drugačije obrasce proporcija po kategorijama od ostatka stavki. Kod većine stavki najbrojniji su odgovori u srednjoj kategoriji. Rekli smo da je kod stavke 10 srednja kategorija odgovora najređa, dok kod stavki 1 i 6 proporcija odgovora po kategorijama opada od kategorije 1 ka kategoriji 3. Takođe, iz objekta *parPCM.e* i tabele sa procenjenim parametrima modela u odeljku 10.5.4 vidimo da je razmak između koraka kod stavki 1 i 6 najmanji, a stavke su relativno teške. Možemo pretpostaviti da se u odgovaranje na ove stavke meša još neka varijabla koja uzrokuje drugačiji proces odgovaranja. Imajući sve navedeno u vidu, iako ove tri stavke imaju više maksimalne informativnosti od ostalih stavki, trebalo bi razmisliti o njihovom izbacivanju jer potencijalno narušavaju validnost instrumenta.

Informativnost testa je jednaka sumi informativnosti stavki pa zavisi od njihovih lokacija (Slika 91).

```
plotINFO(mPCM.e, type="test")
```




Slika 91 – Funkcija informativnosti testa

Kriva informativnosti testa ne mora biti simetrična ni unimodalna. Ovaj test je najinformativniji nešto iznad prosečnog nivoa osobine (Slika 91). Ako kao dovoljan nivo informativnosti uzmemo 3,33, što je ekvivalentno pouzdanosti od 0,7, vidimo da je to u vrlo uskom opsegu osobine. Trebalo bi dodati još stavki, posebno onih čija težina je ispod prosečne vrednosti (ako želimo da merenje obavljamo u opštoj populaciji)

#### 10.5.4.6.1 Numerički pokazatelji informativnosti

Ono što smo gore prikazali grafički, izrazićemo numerički. Potrebno je navesti raspon latentne osobine koji nas zanima.

```
TETE <- seq(-6, 6, 0.1)
# Kreiramo vektor sa željenim rasponom i nivoima thete
# Sekvenca od -6 do 6 u koracima po 0.1
TINFO <- test_info(mPCM.e, theta=TETE)
# Vektor TETE prosleđujemo argumentu theta funkcije test_info()
# Rezultat je vektor informativnosti testa za pojedine nivoe latentne osobine
# Ispis tabele informativnosti po nivoima latentne osobine
TI <- data.frame(cbind(TETE, TINFO))
# Sa funkcijama cbind() i data.frame() povezali smo vektore nivoa osobine i
informativnost u matricu podataka
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n", TI[TI[,2]==max(TI[,2]),1], "logita")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  0.6 logita

cat("\nInformativnost u rasponu od -6 do 6 logita: \n")

##
## Informativnost u rasponu od -6 do 6 logita:
```

```
inf66 <- sum(TINFO)
inf66

## [1] 197.1287

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 6 logita:  \n")

##
## Informativnost u rasponu od 0 do 6 logita:

inf06 <- sum(test_info(mPCM.e, theta=seq(0, 6, 0.1)))
inf06

## [1] 117.5398

cat(" \nProcenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od
-6 do 6 logita:  \n")

##
## Procenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od -6 d
o 6 logita:

inf06/inf66*100

## [1] 59.62588
```

Vidimo da je informativnost dosta veća (59,63%) u opsegu natprosečnih nivoa osobine.

#### *10.5.4.7 Izbacivanje stavki*

Kada izbacimo neku stavku promeniće se i ukupni skorovi i nova procena parametara stavki daće drugačije rezultate. Zato bi stavke trebalo uklanjati po jednu u svakom koraku.

```
# Prvo ćemo izbaciti stavku 10 iz matrice podataka (a zatim 6 i 1). Novu matricu
podataka nazvaćemo p.podR. Redukovanu matricu podataka čuvamo pod drugim imenom
kako bismo sačuvali i originalne podatke. Nije nužno tako, ali je korisno
p.podR <- p.pod[,c(2:5,7:9)]
mPCM.eR <- PCM(p.podR, se=TRUE)

## Warning:
## The following items have no 0-responses:
## it2 it3 it4 it5 it7 it8 it9
## Responses are shifted such that lowest category is 0.

# Uporedni prikaz parametara na originalnom i na redukovanom skupu stavki
round(cbind("originalni"=data.frame(thresholds(mPCM.e)$threshtable),
 "redukovani"=data.frame(rbind(rep(NA,3),
 thresholds(mPCM.eR)$threshtable[[1]][1:4,], rep(NA,3),
 thresholds(mPCM.eR)$threshtable[[1]][5:7,], rep(NA,3)))),3)

##      originalni.X1.Location originalni.X1.Threshold.1 originalni.X1.Threshold.2
## it1                   1.178                     0.409                     1.947
## it2                   0.534                    -1.279                     2.346
## it3                   0.643                    -0.399                     1.685
## it4                   1.215                    -0.947                     3.377
## it5                  -0.020                    -2.327                     2.286
## it6                   0.845                     0.296                     1.394
## it7                  -0.074                    -1.748                     1.601
## it8                  -0.259                    -2.584                     2.066
```

```
## it9                  -0.220                    -1.508                     1.068
## it10                  0.403                     1.596                    -0.789
##      redukovani.Location redukovani.Threshold.1 redukovani.Threshold.2
## it1                   NA                     NA                     NA
## it2                0.959                 -1.039                  2.957
## it3                1.067                 -0.135                  2.269
## it4                1.707                 -0.674                  4.088
## it5                0.368                 -2.130                  2.866
## it6                   NA                     NA                     NA
## it7                0.289                 -1.545                  2.123
## it8                0.110                 -2.400                  2.621
## it9                0.116                 -1.311                  1.543
## it10                  NA                     NA                     NA

mPCM.eRPI <- person.parameter(mPCM.eR)
mPCM.eRLR <- LRtest(mPCM.eR, splitcr = "mean")
# Izračunali smo Andersenov LR test na redukovanom modelu
mPCM.eRLR

##
## Andersen LR-test:
## LR-value: 38.163
## Chi-square df: 13
## p-value:  0

itemfit(mPCM.eRPI)

##
## Itemfit Statistics:
##        Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## it2  776.711 978   1.000      0.793     0.808   -4.717  -4.630   0.598
## it3 1012.873 978   0.213      1.035     1.002    0.708   0.069   0.487
## it4  771.264 978   1.000      0.788     0.819   -4.441  -4.353   0.562
## it5  750.024 978   1.000      0.766     0.796   -4.438  -4.355   0.567
## it7  902.142 978   0.960      0.921     0.917   -1.719  -1.937   0.521
## it8  782.587 978   1.000      0.799     0.826   -3.760  -3.713   0.540
## it9  892.557 978   0.976      0.912     0.888   -1.859  -2.752   0.553
```

Izbacili smo stavku 10 jer je imala značajan misfit (i najveću vrednost hi–kvadrata)[55]. Promenili su se parametri ostalih stavki, ali model i dalje nije bio saglasan sa podacima. Andersenov LR test bio je i dalje značajan, a takođe bilo je još stavki sa misfitom (stavke 6 i 1). Sledeća stavka koju smo izbacili bila je stavka 6. Imala je značajan hi–kvadrat (najveći od preostalih stavki), šumni autfit i infit. I nakon njenog izbacivanja LR test ukazivao je da model nije saglasan sa podacima. Ostala je još jedna stavka sa značajnim misfitom – stavka 1. Nakon njenog izbacivanja nije bilo misfitujućih stavki, ali LR test je bio i dalje značajan. Preostale stavke težile su overfitu.

#### *10.5.4.8 Separaciona pouzdanost*

Izračunali smo separacionu, marginalnu i empirijsku pouzdanost za test (videti odeljak 7.2), a takođe i Kronbahovu alfu.

[55] Postupak izbacivanja stavki bio je iterativan, ali je zbog prostora prikazan samo završni korak.

```
SEP.POUZD <- SepRel(mPCM.ePI)
# Funkcija SepRel() kao objekat uzima objekat sa parametrima ispitanika
SEP.POUZD

## Separation Reliability: 0.8004

# Za izračunavanje Kronbahove alfe koristićemo funkciju cronbach.alpha() iz paketa
ltm
# Ispisaćemo samo sirovu alfu ($alpha)
cat("Kronbahova alfa: \n")

## Kronbahova alfa:

ltm::cronbach.alpha(p.pod)$alpha

## [1] 0.8085455

cat("Marginalna pouzdanost iz paketa mirt: \n")

## Marginalna pouzdanost iz paketa mirt:

mirt::marginal_rxx(mPCM.m)

## [1] 0.7559942

cat("Empirijska pouzdanost iz paketa mirt: \n")

## Empirijska pouzdanost iz paketa mirt:

mirt::empirical_rxx(mirt::fscores(mPCM.m, full.scores.SE = TRUE))

##        F1
## 0.8179874
```

Sve vrste pouzdanosti za ovaj test su zadovoljavajuće (>0,70), s tim da je marginalna nešto niža. Dok je izračunavanje preostalih vrsta pouzdanosti vezano za konkretni uzorak, marginalna se računa u odnosu na pretpostavljenu normalnu distribuciju merene osobine.

Izračunaćemo i broj razdvojivih stratuma.

```
G <- sqrt(SEP.POUZD$sep.rel/(1-SEP.POUZD$sep.rel))
cat("G =", G)

## G = 2.002526

BRS=(4*G+1)/3
cat("\nBroj razdvojivih stratuma =", BRS)

##
## Broj razdvojivih stratuma = 3.003369
```

Na osnovu ovog testa po nivou osobine možemo izdvojiti 3 značajno različita stratuma ispitanika.

#### *10.5.4.9 Parametri (mere) ispitanika*

Iz ranije konstruisanog objekta *mPCM.ePI* možemo izvući parametre ispitanika. Ovo ima smisla raditi samo ako je model saglasan sa podacima. Mere nećemo prikazivati zbog prostora.

```
MERE <- mPCM.ePI$theta.table[,1]
# Iz objekta mr.ePI iz tabele theta.table izdvajamo prvu kolonu {,1}
```

```
# Ovu varijablu možemo kasnije koristiti u analizama, npr:
mean(MERE)

## [1] 0.2755733

# Prosečna mera u logitima
```

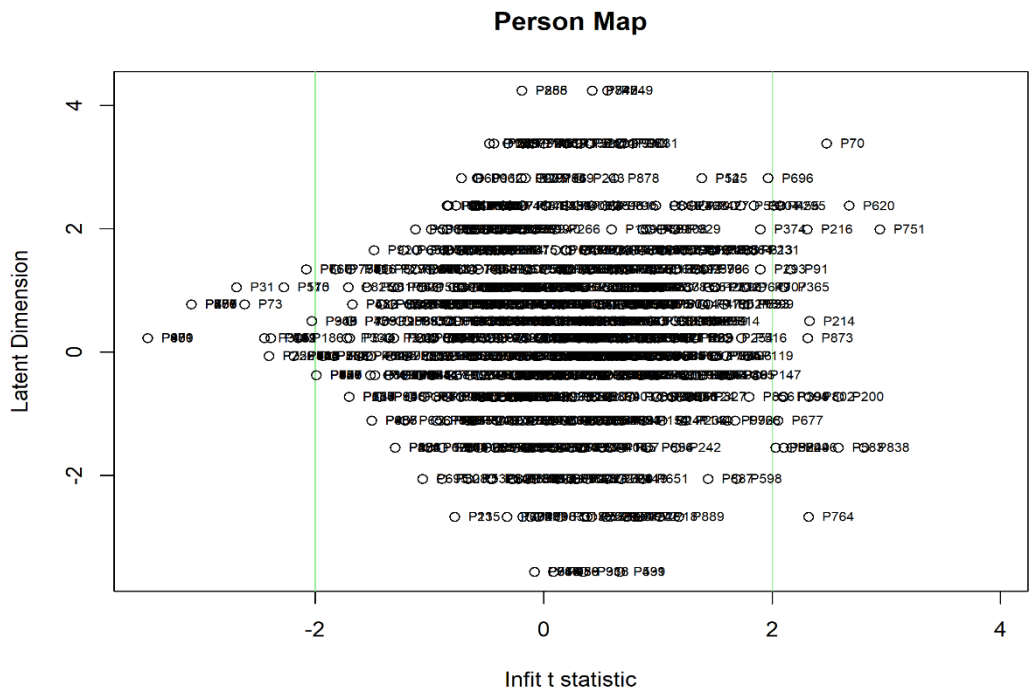
#### *10.5.4.10 Pokazatelji fita za ispitanike*

Izračunaćemo pokazatelje fita za ispitanike. Ni njih nećemo prikazivati.

```
fit.isp <- personfit(mPCM.ePI)
```

10.5.4.10.1 Grafikon misfita za ispitanike

```
plotPWmap(mPCM.e, pmap=T, imap=F)
```




Slika 92 – Standardizovani infit za ispitanike

Kao i kod modela skala procene, manji broj ispitanika ima misfit u formi prigušenog šuma, a nešto veći broj u obliku šuma (Slika 92).

#### *10.5.4.11 Formiranje sirovog skora i korelacije sa Rašovim merama*

Izračunaćemo ukupni skor i korelirati sa Rašovim merama ispitanika na osnovu modela stepenovanog ocenjivanja.

```
SKOR <- rowSums(p.pod[1:ncol(p.pod)])
df <- data.frame(cbind(SKOR, MERE))
head(df, 5)

##   SKOR        MERE
## 1   18 -0.06756479
```

```
## 2   19  0.22152882
## 3   25  1.98870943
## 4   19  0.22152882
## 5   17 -0.38068425

#View(RASCH.POD)
cat("Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera: \n")

## Pirsonov koeficijent korelacije između sumacionog skora i Rašovih mera:

cor(SKOR, MERE, method="pearson")

## [1] 0.985381

cat(" \nSpirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:
\n")

##
## Spirmanov koeficijent korelacije između sumacionog skora i Rašovih mera:

cor(SKOR, MERE, method="spearman")

## [1] 1
```

Pirsonov koeficijent korelacije sumacionog skora i Rašovih mera je vrlo visok (0,985), a Spirmanov koeficijent rang korelacije je jednak 1. Poredak ispitanika po sumacionim skorovima i Rašovim merama je isti i u modelu stepenovanog ocenjivanja.

### 10.5.5 Generalizovani model stepenovanog ocenjivanja (GPCM) u paketu *mirt*

Generalizovani model stepenovanog ocenjivanja Murakija (1992) zasnovan je na modelu stepenovanog ocenjivanja, ali uključuje parametar nagiba stavki. Ne zahteva jednak broj kategorija na svim stavkama, a koraci/pragovi ne moraju biti sekvencijalni. Pošto model nije jednoparametarski, ne može biti procenjen u paketu *eRm*, pa ćemo koristiti paket *mirt*.

```
p_load(mirt)
mGPCM.m <- mirt(p.pod, 1, itemtype = "gpcmIRT", SE=T, verbose = FALSE)
# Parametre koraka ćemo sačuvati u novoj matrici podataka
koefGPCM <- data.frame(coef(mGPCM.m, simplify=TRUE, IRTpars=TRUE)$items[,1:3])
# Izračunaćemo interval između lokacija koraka
koefGPCM$interval <- koefGPCM[,"b2"]-koefGPCM[,"b1"]
# Izračunaćemo lokaciju stavke kao prosek lokacija koraka
koefGPCM$bj <- rowMeans(koefGPCM[, c("b1", "b2")])
# Izračunaćemo tau za prvi korak...
koefGPCM$T1 <- koefGPCM$b1-koefGPCM$bj
# i za drugi
koefGPCM$T2 <- koefGPCM$b2-koefGPCM$bj
round(koefGPCM,3)

##         a1     b1     b2 interval     bj     T1     T2
## it1  0.770  0.246  1.652    1.407  0.949 -0.703  0.703
## it2  1.770 -1.087  1.416    2.503  0.165 -1.251  1.251
## it3  1.016 -0.615  1.251    1.866  0.318 -0.933  0.933
## it4  1.917 -0.824  2.003    2.827  0.589 -1.414  1.414
## it5  1.974 -1.664  1.290    2.954 -0.187 -1.477  1.477
## it6  0.783  0.140  1.033    0.893  0.586 -0.446  0.446
## it7  1.398 -1.558  1.011    2.569 -0.274 -1.285  1.285
```

```
## it8  1.758 -1.911  1.207    3.118 -0.352 -1.559  1.559
## it9  1.275 -1.451  0.631    2.082 -0.410 -1.041  1.041
## it10 0.892  1.564 -1.348   -2.912  0.108  1.456 -1.456
```

Novina u ovom modelu u odnosu na model stepenovanog ocenjivanja je to što se procenjuje parametar nagiba za stavku (koji je jednak za sve kategorije u okviru iste stavke). Najviše parametre nagiba imaju stavke 5 i 4. Podsetićemo da kod politomnih stavki diskriminativnost zavisi od nagiba i blizine kategorija odgovora. Što su kategorije bliže, diskriminativnost će biti veća. U slučaju stavki 4 i 5 i intervali između koraka su slični (uski), ali je nešto uži onaj kod stavke 4 koja ima manji nagib. Kao što ste verovatno primetili, kao i kod modela stepenovanog ocenjivanja intervali nisu jednaki, a kod stavke 10 postoji inverzija koraka.

#### *10.5.5.1 Pokazatelj fita modela u celini*

Kao omnibus pokazatelj fita izračunaćemo $M_2$ i uz njega pokazatelje približnog fita.

```
M2(mGPCM.m)
```

```
##              M2 df         p RMSEA RMSEA_5   RMSEA_95      SRMSR     TLI CFI
## stats 22.83802 25 0.5870214     0       0 0.02267783 0.01749283 1.00093   1
```

Svi pokazatelji ukazuju da model dobro opisuje podatke.

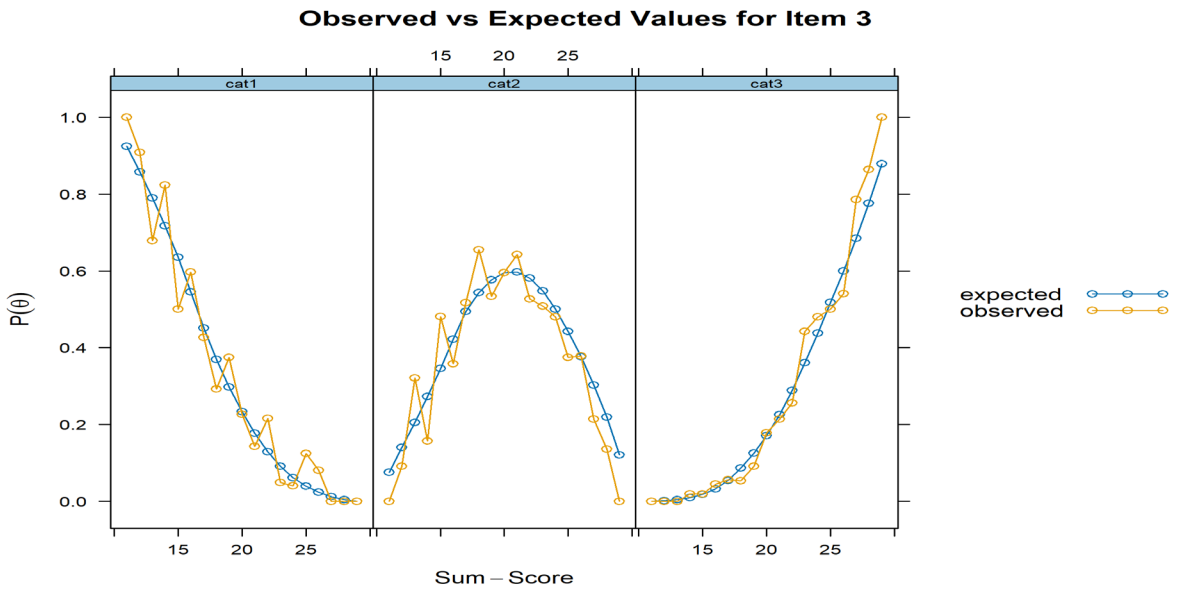
#### *10.5.5.2 Pokazatelji fita za stavke*

Zatražićemo $S–\chi^2$ pokazatelj po proceduri Orlanda i Tisena i $\chi^{2*}$ Stouna, te grafikon modelskih i empirijski karakterističnih kriva stavke 3.

```
# Ovom naredbom dobijamo podrazumevani hi-kvadrat po proceduri Orlanda i Tisena
mirt::itemfit(mGPCM.m)
```

```
##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1 25.287      25      0.003  0.446
## 2   it2 20.071      21      0.000  0.517
## 3   it3 38.517      24      0.025  0.031
## 4   it4 15.855      20      0.000  0.726
## 5   it5 24.723      19      0.017  0.170
## 6   it6 28.184      26      0.009  0.349
## 7   it7 18.358      22      0.000  0.685
## 8   it8 21.566      20      0.009  0.365
## 9   it9 31.729      23      0.019  0.106
## 10 it10 21.942      24      0.000  0.583
```

```
# Argument S_X2.plot = 3 traži prikaz modelske i empirijske KKK za stavku koja je označena brojem (u ovom slučaju to je stavka 3), ali onda se ne prikazuje tabela sa numeričkim pokazateljima
mirt::itemfit(mGPCM.m,  fit_stats=c("S_X2"), S_X2.plot = 3)
```




Slika 93 – Modelske i empirijske karakteristične krive kategorija stavke 3

```
# Ovom naredbom smo zahtevali hi-kvadrat pokazatelj po proceduri koju je definisao Stoun
mirt::itemfit(mGPCM.m,  fit_stats=c("X2*"))
```

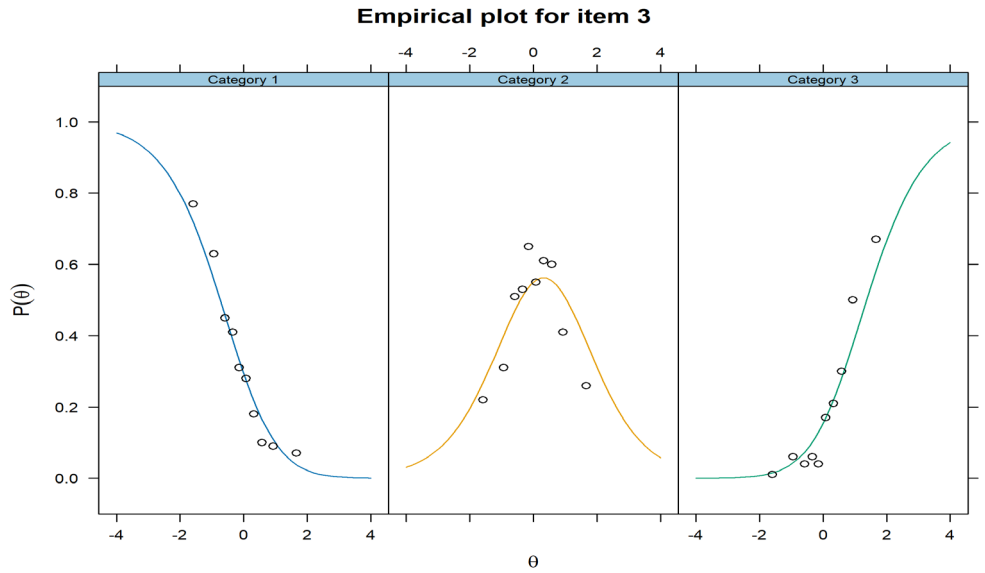
```
##    item X2_star p.X2_star
## 1   it1   3.136     0.488
## 2   it2   6.045     0.053
## 3   it3   8.257     0.015
## 4   it4   4.483     0.178
## 5   it5   3.246     0.379
## 6   it6   3.289     0.444
## 7   it7   2.513     0.571
## 8   it8   5.154     0.154
## 9   it9   5.540     0.079
## 10 it10   3.810     0.237
```

Oba pokazatelja se slažu da misfit postoji kod stavke 3. To je ujedno jedina stavka kod koje je detektovan misfit. Ako gledamo Stounov pokazatelj ($\chi^{2*}$), stavke 2 i 9 imaju *p*–vrednosti blizu granične. S obzirom na to da se ovaj statistik zasniva na Monte Karlo simulacijama trebalo bi ponoviti izračunavanje nekoliko puta kako bismo bili sigurni. U ponovljenim izračunavanjima koja nisu ovde prikazana i stavka 2 je imala značajan misfit u nekoliko navrata. Tretiraćemo je kao sumnjivu. Prikazane su i empirijske i modelske karakteristične krive kategorija za stavku 3 (Slika 93 i Slika 94). Kod ove stavke su problematične dve najniže kategorije odgovora. Vidimo odstupanje modelskih i empirijskih KKK, ali ono ne deluje sistematsko.

Skrenućemo pažnju da stavke 1 i 6, koje su u jednoparametarskom PCM modelu imale značajan misfit (odeljak 10.5.4.2 i 10.5.4.6), u ovom modelu imaju zadovoljavajuće pokazatelje fita. To ukazuje na to da je problem bio u parametrima nagiba koji se kod ovih stavki razlikuju od nagiba ostalih stavki u testu.

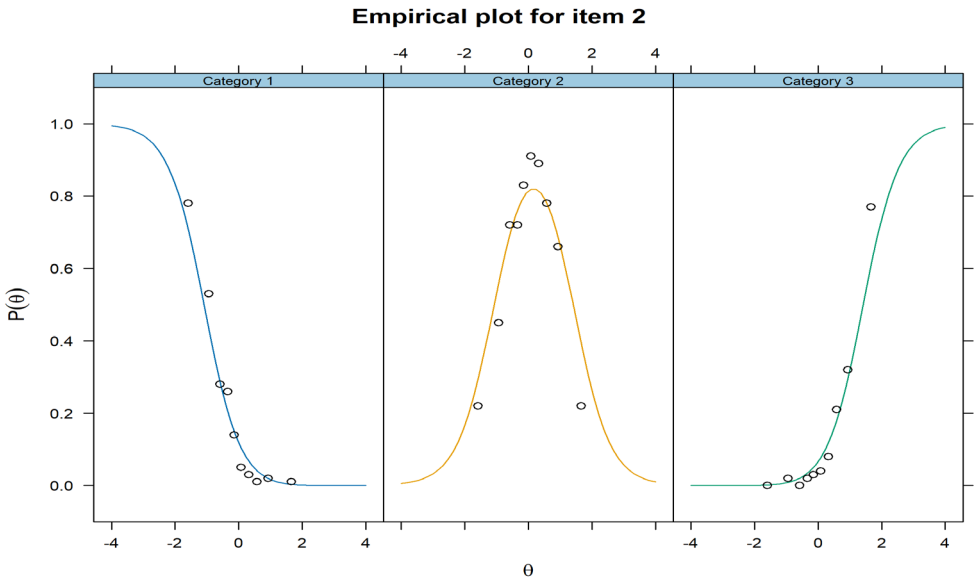
Drugi način da prikažemo empirijske i modelske krive kategorija je upotrebom funkcije *itemfit()*.

```
mirt::itemfit(mGPCM.m, empirical.plot=3)
```




Slika 94 – Modelske i empirijske karakteristične krive kategorija stavke 3 (drugi prikaz)

```
mirt::itemfit(mGPCM.m, empirical.plot=2)
```




Slika 95 – Modelske i empirijske karakteristične krive kategorija stavke 2

Kod stavke 2 (Slika 95) ne vidimo bitnija odstupanja osim u kategoriji skora 2 gde model malo potcenjuje verovatnoću ovog odgovora (nije zabrinjavajuće). Kod stavke 3 deluje kao da je nagib stavke blago potcenjen. Da su krive za nijansu strmije ne bi bilo velikih odstupanja (koja ni ovako ne deluju kao prevelika).

#### *10.5.5.3 Pokazatelji fita za ispitanike*

Podsećamo da pokazatelje fita za ispitanike u paketu *mirt* dobijamo funkcijom *personfit()*. Procedura pokazuje *infit* i *autfit* i pokazatelj $Z_h$ Drazgova i saradnika, baziran na logaritmu verodostojnosti u standardizovanoj metrici.

```
PF.m <- mirt::personfit(mGPCM.m, method = "EAP")
head(PF.m)
```

```
##      outfit   z.outfit     infit     z.infit         Zh
## 1 1.6617716  1.2012860 1.2637111  0.70449412 -0.5258398
## 2 0.3609491 -1.3761351 0.6416248 -0.96043078  1.4589217
## 3 0.8704501 -0.2330055 0.9617855  0.04335264  0.3042059
## 4 0.8192990 -0.1586265 0.7635944 -0.51977012  0.1101673
## 5 0.4284061 -1.2369086 0.4684968 -1.33035437  1.0742726
## 6 0.7150132 -0.4149865 0.6603825 -0.62005134  0.6187894
```

```
# Prikazaćemo proporcije misfitujućih ispitanika
# U donjem redu su proporcije misfitujućih ispitanika (dobro je kada su manje od
0,05)
p_load(dplyr) # Paket potreban za funkciju reframe()
reframe(PF.m, infit = prop.table(table(z.infit > 1.96 | z.infit < -1.96)),
 outfit = prop.table(table(z.outfit > 1.96 | z.outfit < -1.96)),
 Zh = prop.table(table(Zh > 1.96 | Zh < -1.96)))
```

```
##   infit outfit     Zh
## 1 0.948  0.949  0.968
## 2 0.052  0.051  0.032
```

Zbog prostora prikazani su pokazatelji samo za prvih 6 ispitanika. Interpretacija ovih pokazatelja opisana je ranije (odeljak 10.4.1.12). Kao kratki podsetnik, prihvatljivi opseg infita i autfita je 0,7–1,3 u MSQ metrici, –1,96–+1,96 u standardizovanoj metrici. Za $Z_h$ važi isti opseg kao za standardizovani infit i autfit. Niske vrednosti infita i autfita ukazuju na prigušeni šum, a visoke na šum. Kod $Z_h$ pokazatelja je obrnuto. Objekat u kojem smo sačuvali rezultate je klase *data.frame* isto kao matrica podataka i lako se mogu spojiti. Pokazatelji misfita se kasnije mogu koristiti za selekciju ispitanika. Prikazana je i tabela sa proporcijama ispitanika koji pokazuju infit i autfit. U donjem redu tabele su proporcije ispitanika sa misfitom. Dobro je kada su ove vrednosti niske, po mogućstvu ispod 0,05. Infit i autfit su na toj granici, a $Z_h$ je nešto ispod nje.

#### *10.5.5.4 Procena lokalne zavisnosti*

Postojanje lokalne zavisnosti utvrdićemo upotrebom $Q_3$ koeficijenta korelacije reziduala.

```
Q3m <- as.matrix(residuals(mGPCM.m, type="Q3"))
```

```
## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.174  -0.112  -0.095  -0.091  -0.073  -0.007
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.113 -0.034 -0.066 -0.041 -0.022 -0.020 -0.126 -0.095 -0.073
## it2  -0.113  1.000 -0.105 -0.134 -0.117 -0.120 -0.121 -0.094 -0.086 -0.095
## it3  -0.034 -0.105  1.000 -0.091 -0.120 -0.070 -0.097 -0.061 -0.070 -0.095
## it4  -0.066 -0.134 -0.091  1.000 -0.112 -0.061 -0.174 -0.119 -0.089 -0.076
## it5  -0.041 -0.117 -0.120 -0.112  1.000 -0.112 -0.088 -0.138 -0.103 -0.078
## it6  -0.022 -0.120 -0.070 -0.061 -0.112  1.000 -0.007 -0.033 -0.111 -0.100
```

```
## it7  -0.020 -0.121 -0.097 -0.174 -0.088 -0.007  1.000 -0.093 -0.100 -0.107
## it8  -0.126 -0.094 -0.061 -0.119 -0.138 -0.033 -0.093  1.000 -0.084 -0.094
## it9  -0.095 -0.086 -0.070 -0.089 -0.103 -0.111 -0.100 -0.084  1.000 -0.134
## it10 -0.073 -0.095 -0.095 -0.076 -0.078 -0.100 -0.107 -0.094 -0.134  1.000
```

```
diag(Q3m) <- NA
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", round(Q3mean, 3), "\n\n")
```

```
##
## Prosečna vrednost Q3: -0.091
```

```
Q3c <- Q3m-Q3mean
Q3c
```

```
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1      NA -0.022  0.057  0.024  0.050  0.068  0.071 -0.035 -0.004  0.018
## it2  -0.022     NA -0.014 -0.043 -0.026 -0.030 -0.030 -0.004  0.005 -0.004
## it3   0.057 -0.014     NA -0.001 -0.030  0.021 -0.006  0.029  0.021 -0.004
## it4   0.024 -0.043 -0.001     NA -0.022  0.030 -0.084 -0.029  0.001  0.015
## it5   0.050 -0.026 -0.030 -0.022     NA -0.022  0.003 -0.047 -0.012  0.012
## it6   0.068 -0.030  0.021  0.030 -0.022     NA  0.084  0.057 -0.020 -0.009
## it7   0.071 -0.030 -0.006 -0.084  0.003  0.084     NA -0.003 -0.010 -0.016
## it8  -0.035 -0.004  0.029 -0.029 -0.047  0.057 -0.003     NA  0.007 -0.004
## it9  -0.004  0.005  0.021  0.001 -0.012 -0.020 -0.010  0.007     NA -0.043
## it10  0.018 -0.004 -0.004  0.015  0.012 -0.009 -0.016 -0.004 -0.043     NA
```

```
Q3c[Q3c<.20] <- NA
Q3c
```

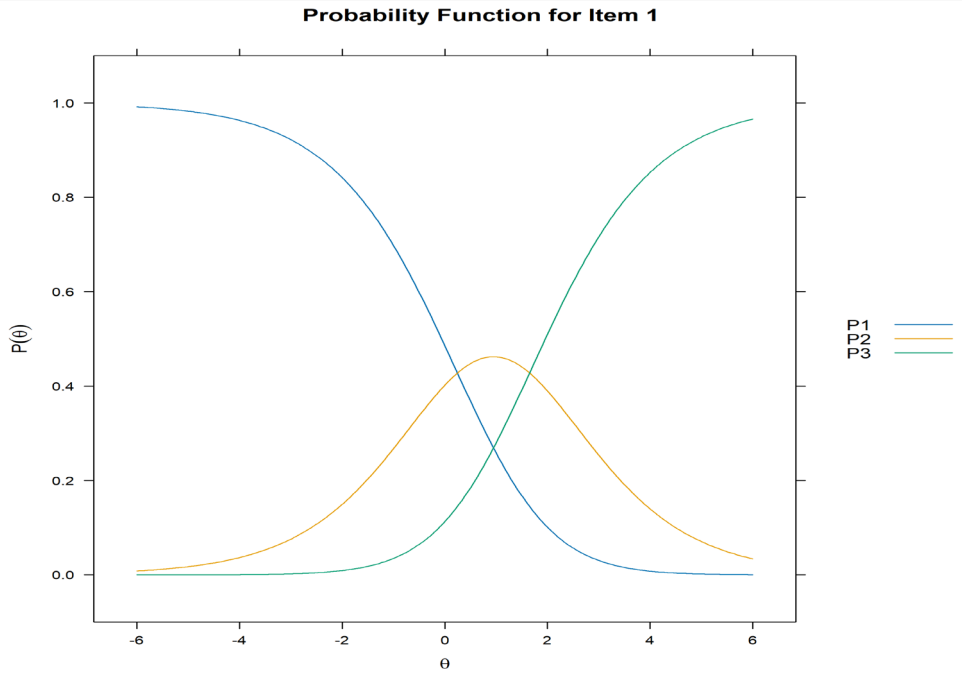
```
##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## it1   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it2   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it3   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it4   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it5   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it6   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it7   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it8   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it9   NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
## it10  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA
```

Vrednosti $Q_3$ pokazatelja veće od 0,2 smatraju se upozoravajućim, dok se neki autori pozivaju na Koenovu (1988) podelu veličine efekta i vrednosti veće od 0,36 (srednji efekat) smatraju kao one na koje bi trebalo obratiti pažnju. Za više informacija o interpretaciji $Q_3$ pogledajte odeljak 9.2. Mi smo ovde koristili pristup u kojem se vrednost $Q_3$ pokazatelja poredi sa prosečnom vrednošću u matrici i kao indikativne uzimaju se one koje su za 0,2 veće od proseka. Kako je prosečna vrednost u matrici obično negativna, to praktično snižava graničnu vrednost $Q_3$ pokazatelja (nešto ranije, ovo je opisano malo drugačije ali suštinski isto). $Q_3$ ne ukazuje da je u ovom testu narušena lokalna zavisnost.

#### *10.5.5.5 Grafički prikaz karakterističnih kriva kategorija*

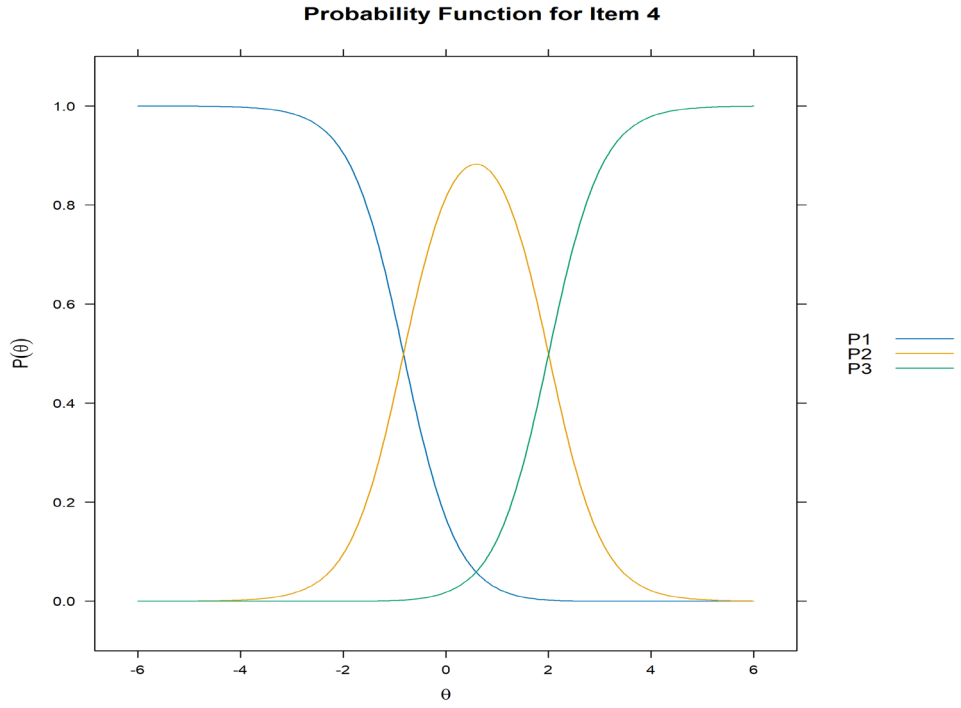
U ovom modelu takođe ćemo prikazati karakteristične krive kategorija, a ne stavki. Krive su slične kao u modelu stepenovanog ocenjivanja sa dodatkom različitog nagiba stavki.

```
itemplot(mGPCM.m, item=1, type = "trace")
```



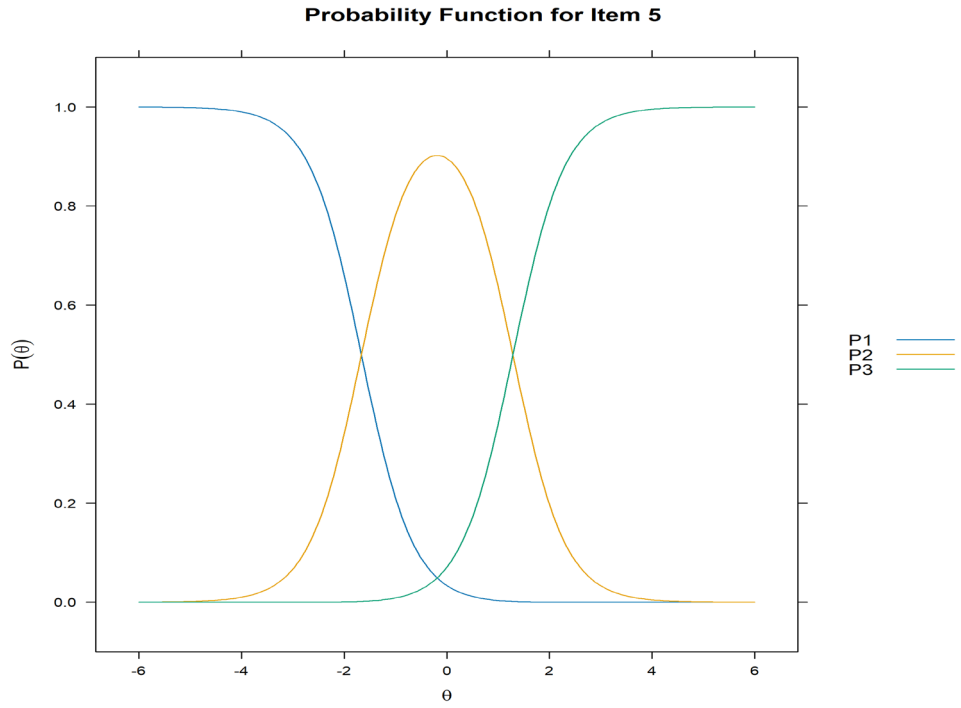

Slika 96 – Karakterične krive kategorija stavke 1

```
itemplot(mGPCM.m, item=4, type = "trace")
```



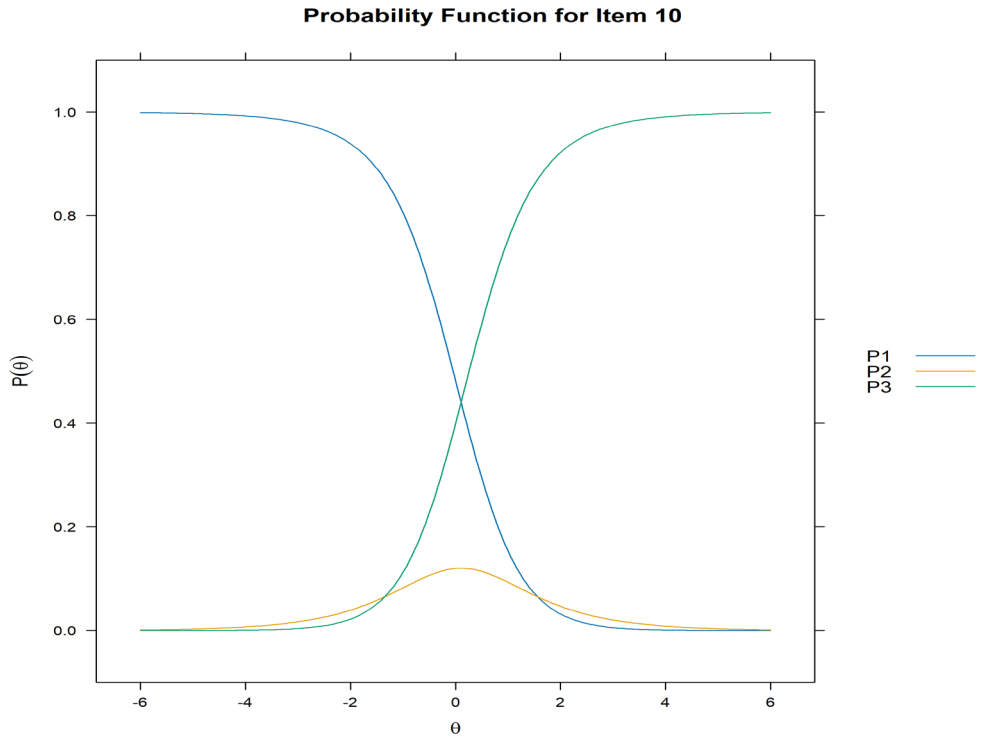

Slika 97 – Karakterične krive kategorija stavke 4

```
itemplot(mGPCM.m, item=5, type = "trace")
```




Slika 98 – Karakteristične krive kategorija stavke 5

```
itemplot(mGPCM.m, item=10, type = "trace")
```



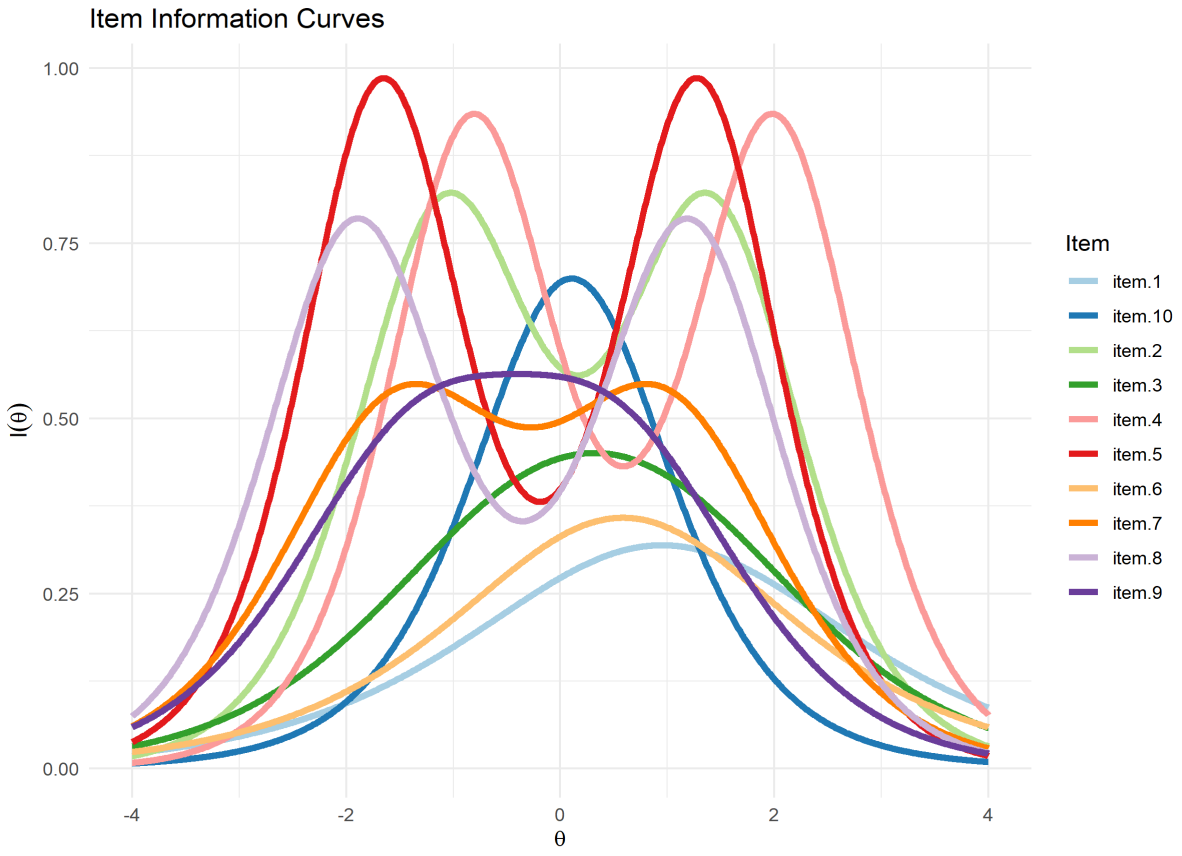

Slika 99 – Karakteristične krive kategorija stavke 10

Prikazane su 4 stavke: stavka 1 ima najniži nagib (Slika 96), a stavke 4 (Slika 97) i 5 (Slika 98) dva najviša. Stavka 5 ima viši nagib i veći interval između koraka, a stavka 4 manji nagib i manji raspon između koraka. Stavka 10 (Slika 99) je zanimljiva zbog poremećaja redosleda lokacija koraka, a razlozi za ovu pojavu opisani su kod modela stepenovanog ocenjivanja (odeljci 4.2.1.3 i 10.5.4.4).

#### *10.5.5.6 Informativnost stavki i testa*

Kao i kod modela stepenovanog ocenjivanja informativnost stavke u GPCM zavisi od informativnosti kategorija. Informativnosti kategorija se sabiraju po nivoima osobine i tako daju informativnost stavke. Za razliku od PCM, ovde stavke imaju različite nagibe, a informativnost kategorije zavisi od kvadrata nagiba. Što je nagib veći i informativnost stavke će biti veća. Oblik krive informativnosti stavke zavisi i od intervala između kategorija. Kada su ovi intervali veći, kriva će biti polimodalna. Ovo će više doći do izražaja što su nagibi veći. Za prikaz ćemo koristiti pakete *ggmirt*, *ggplot2* i *RColorBrewer*.

```
library(ggmirt)
p_load(ggplot2)
# i RColorBrewer (zbog dodatnih boja za krive stavki)
p_load(RColorBrewer)
itemInfoPlot(mGPCM.m, legend = T) + scale_color_brewer(palette = "Paired") +
 ggplot2::geom_line(size = 1.5)
```

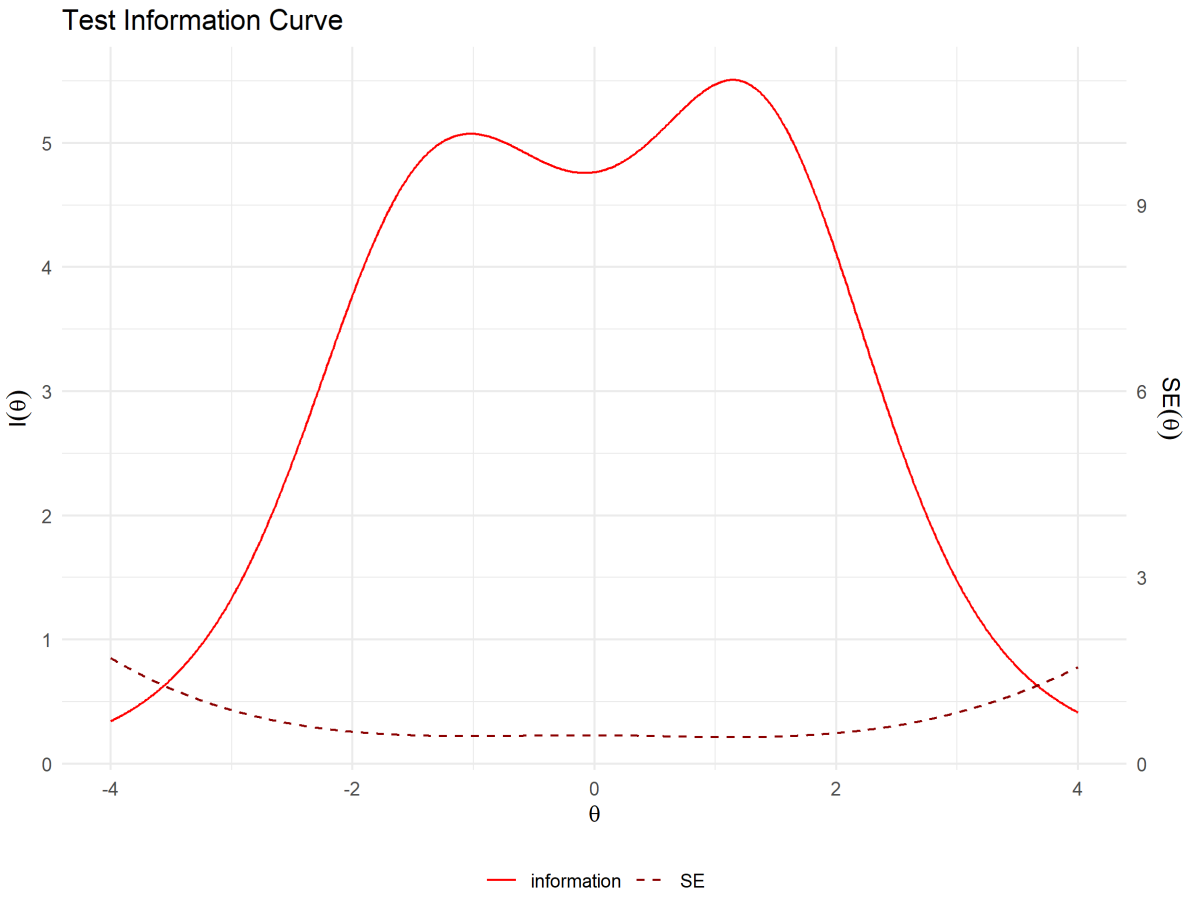



Slika 100 – Funkcije informativnosti stavki u generalizovanom modelu stepenovanog ocenjivanja

Krive stavki (Slika 100) kod kojih su lokacije koraka najbliže biće unimodalne i zašiljene (stavke 3, 6 i 10). Stavka 10 ima krivu koja najviše liči na krive informativnosti binarno skorovanih stavki zato što kod nje postoji poremećaj redosleda koraka. Usled toga, srednja kategorija nikad nije najverovatniji odgovor, pa stavka i funkcioniše kao binarna. Kako se lokacije koraka udaljavaju, krive postaju šire, a vrhovi su im manje šiljati (9 i 1), dok su kod stavki sa udaljenim lokacijama koraka krive bimodalne ili polimodalne (5, 4, 2, 8). Ako je sve ostalo isto, krive stavki sa većim parametrom nagiba biće više.

Kriva informativnosti testa predstavlja sumu informativnosti stavki po nivoima osobine, pa zavisi od visine i lokacije na kojoj su stavke informativne.

```
testInfoPlot(mGPCM.m, adj_factor = 2)
```




Slika 101 – Funkcije informativnosti i standardne greške testa

Kriva informativnosti ovog testa je bimodalna (Slika 101). Test je najinformativniji nešto iznad prosečnog nivoa osobine, ali postoji i drugi vrh krive (niži) na nivou osobine od –1 logita. Ako kao dovoljan nivo informativnosti uzmemo 3,33, što je ekvivalentno pouzdanosti od 0,7, vidimo da informativnost prelazi tu vrednost u opsegu osobine negde između –2 i +2 logita (u kom tačno videćemo kasnije). To nije loše, ali ako želimo da proširimo raspon osobine u kojem želimo dovoljno pouzdano merenje trebalo bi dodati još stavki koje dodaju informativnost izvan ovog opsega.

#### 10.5.5.6.1 Numerički pokazatelji informativnosti

Izračunaćemo numeričke pokazatelje informativnosti kako bismo potvrdili ono što smo videli na prethodnom grafikonu. Potrebno je navesti željeni raspon latentne osobine.

```
# Za ceo test
mGPCMpi <- areainfo(mGPCM.m, c(-3,3))
mGPCMpi

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 24.88573  27.10552  0.9181058     10

# Samo za određene stavke
areainfo(mGPCM.m, c(-3,3), which.items = c(1,2,4:10)) # Bez stavke 3

##  LowerBound UpperBound    Info TotalInfo Proportion nitems
##          -3          3 23.0925  25.07356  0.9209904      9
```

Informativnost po nivoima merene osobine možemo dobiti na sledeći način:

```
# Prvo definišemo vrednosti latentne osobine za koje želimo da izračunamo
informativnost (sekvenca u rasponu od -6 do 6 sa koracima od 0,1)
thetaM <- seq(-6,6, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(mGPCM.m, Theta = thetaM)
# Prikaz prvih 6 redova (ukolonite head() za ispis cele tabele)
TI <- data.frame(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
head(TI)

##   nivo.osobine informativnost
## 1         -6.0     0.02770761
## 2         -5.9     0.03104996
## 3         -5.8     0.03483887
## 4         -5.7     0.03913966
## 5         -5.6     0.04402788
## 6         -5.5     0.04959098

# Maksimalna informativnost
cat("Maksimalna informativnost:", max(infoM), "\n")

## Maksimalna informativnost: 5.503845

# Ispis nivoa osobine na kom je informativnost maksimalna
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n", thetaM[infoM==max(infoM)], "\n")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  1.1

# Ispis raspona u kojem je informativnost veća od 3,33 što odgovara pouzdanosti od
0,7
cat("Raspon latentne osobine u kojem je test zadovoljavajuće informativan: \n",
 min(thetaM[infoM>3.33]), "do", max(thetaM[infoM>3.33]), "logita")

## Raspon latentne osobine u kojem je test zadovoljavajuće informativan:
##  -2.1 do 2.2 logita

cat("\nInformativnost u rasponu od -6 do 6 logita: \n")

##
## Informativnost u rasponu od -6 do 6 logita:
```

```
inf66 <- sum(TINFO)
inf66

## [1] 197.1287

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 6 logita:  \n")

##
## Informativnost u rasponu od 0 do 6 logita:

inf06 <- sum(TINFO[!thetaM < 0], na.rm = TRUE)
inf06

## [1] 117.5398

cat(" \nProcenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od
-6 do 6 logita:  \n")

##
## Procenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od -6
do 6 logita:

round(inf06/inf66*100,2)

## [1] 59.63
```

Maksimalna informativnost testa je 5,50 na nivou osobine od 1,1 logita (drugi nešto niži vrh je na vrednosti od –1 logita, gde je informativnost 5,07). Vidimo da je informativnost veća (59,63%) u opsegu natprosečnih nivoa osobine. Informativnost je zadovoljavajuća u rasponu od –2,1 do +2,2 logita.

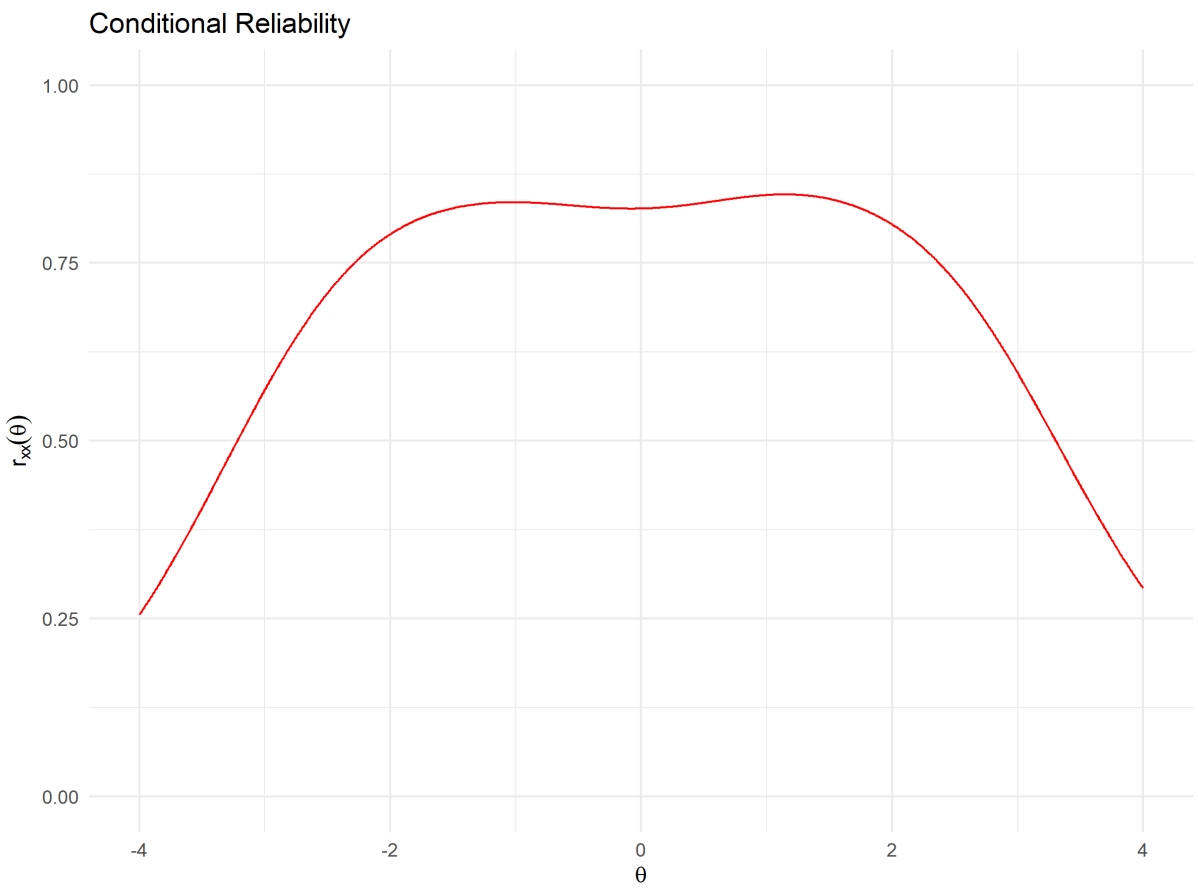
#### *10.5.5.7 Marginalna i empirijska pouzdanost*

Izračunaćemo empirijsku i marginalnu pouzdanost (videti odeljak 7.2).

```
# Parametre ispitanika i standardne greške dobijamo funkcijom fscores() i smeštamo
u objekat THETA_SE koji prosleđujemo funkciji za računanje empirijske pouzdanosti
THETA_SE <- fscores(mGPCM.m, full.scores.SE = TRUE)
cat("\nEmpirijska pouzdanost:", round(empirical_rxx(THETA_SE),3))

##
## Empirijska pouzdanost: 0.828

cat("\nMarginalna pouzdanost:", round(marginal_rxx(mGPCM.m),3))

##
## Marginalna pouzdanost: 0.828

cat("\nKronbahova alfa:", round(ltm::cronbach.alpha(p.pod)$alpha,3))

##
## Kronbahova alfa: 0.809
```

Tri procene pouzdanosti su vrlo slične i dobre (>0,8). Prikazaćemo i grafikon uslovne pouzdanosti (Slika 102). Uslovna pouzdanost nije nijedan od gore navedenih pokazatelja. Ona je funkcija nivoa osobine.

```
ggmirt::conRelPlot(mGPCM.m)
```




Slika 102 – Uslovna pouzdanost testa

#### *10.5.5.8 Mere ispitanika*

U paketu *mirt* mere ispitanika dobijaju se jednostavno. Sačuvaćemo ih u objekat pod nazivom *MERE* i iskoristiti nešto kasnije. Podsetićemo, mere ima smisla računati samo ako je model saglasan sa podacima. Na početku izračunavanja ovog modela $M_2$ statistik nije bio značajan pa smo zaključili da je model saglasan sa podacima.

```
MERE <- fscores(mGPCM.m)
```

#### *10.5.5.9 Izbacivanje stavki*

Iako je ovaj model sa svim stavkama saglasan sa podacima odlučili smo se da izbacimo stavke koje pokazuju misfit. Ako to bude konačno rešenje za koje ćemo se opredeliti na kraju, onda bi gore izračunate mere trebalo ponovo da izračunamo na konačnom modelu. Stavke bi trebalo izbacivati jednu po jednu i u svakom koraku proceniti učinjeno. Uklonićemo prvo stavku 3 jer je jedino ona pokazala značajan misfit, a zatim i stavku 2 koja je nakon izbacivanja stavke 3 pokazala misfit (zbog prostora nećemo prikazivati pojedinačne korake).

```
mGPCM.mR <- mirt(p.pod[,-c(2, 3)], 1, itemtype = "gpcmIRT", SE=T, verbose = FALSE)
```

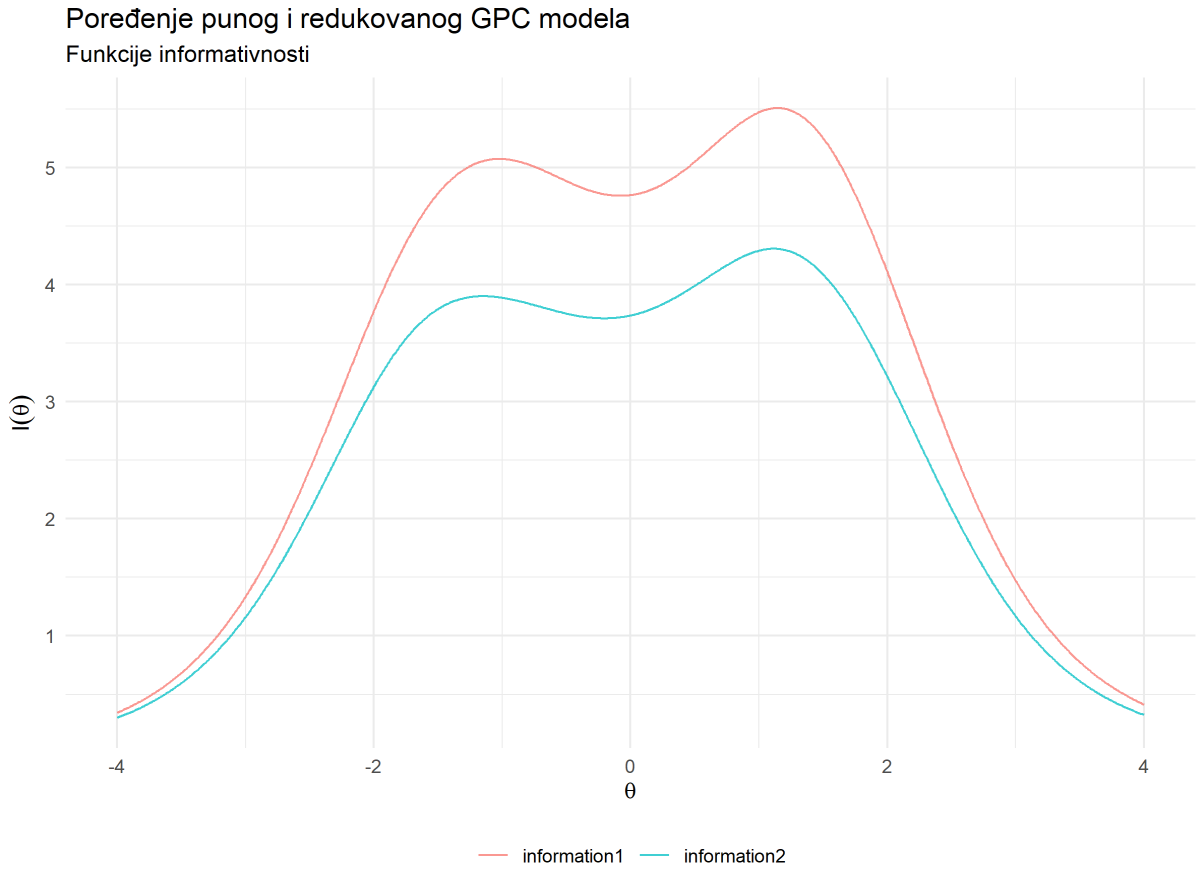
```
M2(mGPCM.mR)

##             M2 df         p      RMSEA RMSEA_5   RMSEA_95      SRMSR       TLI
## stats 17.21508 12 0.1416844 0.02085724       0 0.04122182 0.01935463 0.9949456
##             CFI
## stats 0.9969673

mirt::itemfit(mGPCM.mR,  fit_stats=c("X2*"))

##   item X2_star p.X2_star
## 1  it1   2.669     0.488
## 2  it4   3.228     0.288
## 3  it5   4.354     0.088
## 4  it6   4.308     0.144
## 5  it7   2.384     0.470
## 6  it8   3.898     0.179
## 7  it9   4.703     0.075
## 8 it10   3.066     0.240

PF.mR <- mirt::personfit(mGPCM.mR, method = "EAP")
p_load(dplyr) # Paket potreban za funkciju reframe()
reframe(PF.mR, infit = prop.table(table(z.infit > 1.96 | z.infit < -1.96)),
 outfit = prop.table(table(z.outfit > 1.96 | z.outfit < -1.96)),
 Zh = prop.table(table(Zh > 1.96 | Zh < -1.96)))

##   infit outfit     Zh
## 1 0.963  0.979  0.969
## 2 0.037  0.021  0.031

# U donjem redu je proporcija misfitujućih ispitanika (dobro je kada je manja od
0,05)


THETA_SE <- fscores(mGPCM.mR, full.scores.SE = TRUE)
cat("\nEmpirijska pouzdanost:", round(empirical_rxx(THETA_SE),3))

##
## Empirijska pouzdanost: 0.791

cat("\nMarginalna pouzdanost:", round(marginal_rxx(mGPCM.mR),3))

##
## Marginalna pouzdanost: 0.79

cat("\nKronbahova alfa:", round(psych::alpha(p.pod[,-c(2,3)])$total$raw_alpha,3))

##
## Kronbahova alfa: 0.765
```

Nakon izbacivanja stavke 3, u novom modelu je stavka 2 imala značajan misfit. Setite se da je ta stavka imala graničnu *p* vrednost za Stounov pokazatelj misfita i da je u ponovljenim primenama "šetala" ispod i iznad granične *p* vrednosti. Nakon izbacivanja ove dve stavke više nije bilo misfitujućih stavki. Proporcija misfitujućih ispitanika je niža (ispod 0,05). Pouzdanosti su nešto niže, ali ne drastično (za oko 0,04).

#### *10.5.5.10 Poređenje informativnosti dva modela*

Uporedićemo funkcije informativnosti punog i redukovanog modela. Koristićemo funkciju *testInfoCompare()* iz paketa *ggmirt*. Kao objekte funkcije navešćemo modele koje poredimo.

```
ggmirt::testInfoCompare(mGPCM.m, mGPCM.mR,
 title = "Poređenje punog i redukovanog GPC modela",
 subtitle="Funkcije informativnosti")
```




Slika 103 – Uporedni prikaz funkcija informativnosti punog i redukovanog generalizovanog modela stepenovanog ocenjivanja

Oblik krive informativnosti (Slika 103) nije se promenio, ali je kriva redukovanog modela niža (što je i očekivano jer je broj stavki manji).

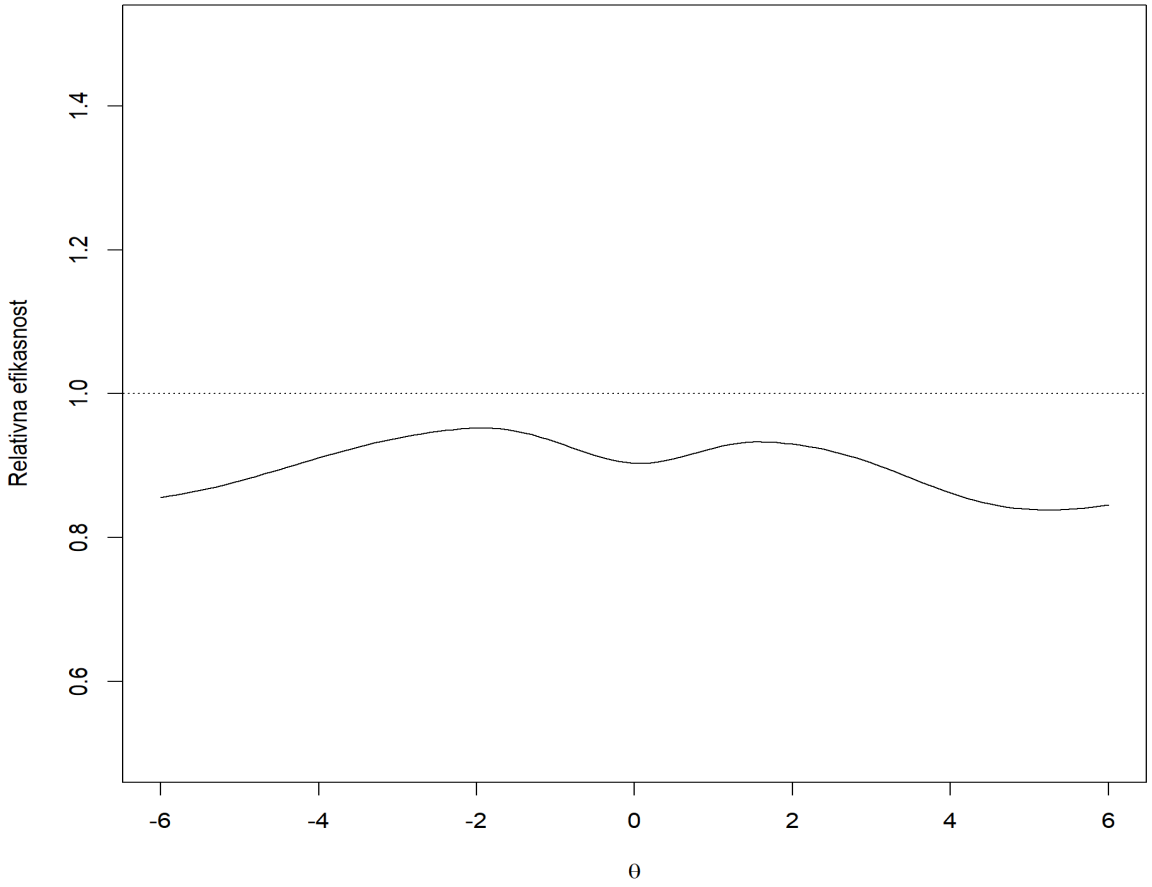
#### 10.5.5.10.1 Relativna efikasnost

Izračunaćemo i prikazati relativnu efikasnost redukovanog modela. Relativna efikasnost se dobija kada informativnost jednog instrumenta podelimo informativnošću drugog instrumenta po nivoima osobine (de Ayala, 2022):

$$RE = \frac{I(\theta, X)}{I(\theta, Y)} \qquad [120]$$

Instrumenti čije informativnosti poredimo mogu biti dva instrumenta koji imaju isti predmet merenja, dve forme istog instrumenta, isti instrument skorovan na različite načine (npr. različitim IRT modelima).

```
# Izračunaćemo informativnosti dva modela u rasponu osobine od -6 do 6 logita
thetaRI = seq(-6, 6, .1)
# informativnost potpunog testa
IP <- testinfo(mGPCM.m, Theta = thetaRI)
# informativnost redukovanog testa
IR <- testinfo(mGPCM.mR, Theta = thetaRI)
# Relativna efikasnost redukovanog testa
RE <- IR/IP
plot(thetaRI, RE, type="l", xlab=expression(theta), ylim=c(0.5,1.5),
 ylab="Relativna efikasnost")
abline(h=1, lty=3)
```




Slika 104 – Kriva relativne efikasnosti redukovanog generalizovanog modela stepenovanog ocenjivanja

Horizontalna isprekidana linija predstavlja jednaku efikasnost dve forme (Slika 104). Vidimo da redukovana forma ne dostiže efikasnost pune forme ni na jednom delu kontinuuma latentne osobine. Relativna efikasnost redukovane verzije testa kreće se u rasponu 0,85–0,95. Ako uzmemo u obzir da smo izbacili 20% stavki, onda to i nije tako loše.

#### *10.5.5.11 Formiranje sumacionog skora i korelacije sa merama*

Izračunaćemo sumacioni skor i korelirati sa IRT merama ispitanika na osnovu generalizovanog modela stepenovanog ocenjivanja.

```
SKOR <- rowSums(p.pod)
df <- data.frame(cbind(SKOR, MERE))
cat("Pirsonov koeficijent korelacije između sumacionog skora i IRT mera: \n")

## Pirsonov koeficijent korelacije između sumacionog skora i IRT mera:

cor(SKOR, MERE, method="pearson")

##             F1
## [1,] 0.9881862

cat(" \nSpirmanov koeficijent korelacije između sumacionog skora i IRT mera:  \n")

##
## Spirmanov koeficijent korelacije između sumacionog skora i IRT mera:

cor(SKOR, MERE, method="spearman")

##             F1
## [1,] 0.9885732
```

I Pirsonov i Spirmanov koeficijent korelacije sumacionog skora i IRT mera dobijenih generalizovanim modelom stepenovanog ocenjivanja su vrlo visoki (oba 0,988). Poredak ispitanika na osnovu ukupnog skora i mera ipak nije isti kao u Rašovim modelima. U generalizovanom modelu stepenovanog ocenjivanja ukupan skor nije dovoljan statistik za procenu mera osobine. S obzirom na to da stavke imaju različite diskriminativnosti, dovoljan statistik je skor zasnovan na sumi ajtemskih skorova ponderisanih diskriminativnošću.

### 10.5.6 Model stepenovanog odgovaranja (*Graded Response Model* – GRM)

Model stepenovanog odgovaranja je dvoparametarski model namenjen politomnim stavkama, odnosno stavkama sa bilo kojim brojem uređenih, ordinalnih kategorija odgovora. Ako stavke imaju samo dve kategorije, svodi se na 2PL model za binarno skorovane stavke. Model procenjuje parametar nagiba za stavke i parametre težine kategorija kojih ima za 1 manje nego samih kategorija. Ne zahteva da sve stavke budu istog formata, parametri kategorija ne moraju biti ekvidistantni, ali za razliku od modela stepenovanog ocenjivanja i generalizovanog modela stepenovanog ocenjivanja, moraju biti sekvencijalni. Od paketa u kojima smo do sada radili može se proceniti u paketima *ltm* i *mirt.* Kada stavke nemaju isti broj kategorija, paket *ltm* zahteva da ih rekodirate tako da svi imaju istu najnižu kategoriju (npr. moguće je da 3 stavke imaju kategorije 1–5, 1–3, 1–4, ali ne i 1–5, 2–4, 1–4). Paket *mirt* interno rekodira stavke bez promene u matrici podataka.

```
# U paketu ltm
p_load(ltm)
mGRM.l <- grm(p.pod, Hessian = TRUE)
```

```
coef(mGRM.l)

##      Extrmt1 Extrmt2 Dscrmn
## it1   -0.098   2.021  1.005
## it2   -1.093   1.418  1.869
## it3   -0.745   1.351  1.272
## it4   -0.823   2.009  1.977
## it5   -1.660   1.286  2.029
## it6   -0.308   1.487  1.035
## it7   -1.572   1.028  1.527
## it8   -1.906   1.198  1.822
## it9   -1.504   0.685  1.474
## it10  -0.081   0.273  1.543

# U paketu mirt
mGRM.m <- mirt::mirt(p.pod, 1, "graded", SE=T, verbose = FALSE)
coef(mGRM.m, simplify=TRUE)

## $items
##         a1    d1     d2
## it1  1.005 0.098 -2.031
## it2  1.869 2.040 -2.651
## it3  1.272 0.947 -1.719
## it4  1.979 1.626 -3.976
## it5  2.029 3.365 -2.611
## it6  1.035 0.318 -1.540
## it7  1.527 2.399 -1.571
## it8  1.821 3.470 -2.185
## it9  1.474 2.215 -1.011
## it10 1.543 0.124 -0.423
##
## $means
## F1
##  0
##
## $cov
##    F1
## F1  1
```

Oba paketa koriste isti metod procene parametara (MMLE) i procene su gotovo identične.

```
round(cbind(coef(mGRM.l),
coef(mGRM.m, simplify=TRUE, IRTpars=TRUE)$items),3)

##      Extrmt1 Extrmt2 Dscrmn     a     b1    b2
## it1   -0.098   2.021  1.005 1.005 -0.097 2.021
## it2   -1.093   1.418  1.869 1.869 -1.091 1.419
## it3   -0.745   1.351  1.272 1.272 -0.744 1.352
## it4   -0.823   2.009  1.977 1.979 -0.822 2.009
## it5   -1.660   1.286  2.029 2.029 -1.658 1.287
## it6   -0.308   1.487  1.035 1.035 -0.307 1.488
## it7   -1.572   1.028  1.527 1.527 -1.571 1.029
## it8   -1.906   1.198  1.822 1.821 -1.906 1.200
## it9   -1.504   0.685  1.474 1.474 -1.503 0.685
## it10  -0.081   0.273  1.543 1.543 -0.081 0.274
```

To nam ostavlja mogućnost da naizmenično koristimo dva paketa po potrebi. Među parametrima nema neobično visokih i onih koji značajno odskaču od ostalih što bi ukazivalo na misfit. Obratite pažnju na parametre kategorija stavke 10 koji su u dva modela stepenovanog ocenjivanja imali poremećaj redosleda.

U ovom modelu redosled kategorija je normalan. Ovaj model ne procenjuje parametre težine stavki, ali za potrebe ocenjivanja pokrivenosti kontinuuma latentne osobine odgovarajućim stavkama oni se mogu izraziti kao prosek parametara kategorija (Ali i ostali, 2015).

```
parm <- data.frame(coef(mGRM.l))
round(cbind(parm, "bj"=rowMeans(parm[,1:2]),
 "interval"=parm$Extrmt2-parm$Extrmt1),3)
```

```
##      Extrmt1 Extrmt2 Dscrmn     bj interval
## it1   -0.098   2.021  1.005  0.961    2.119
## it2   -1.093   1.418  1.869  0.162    2.511
## it3   -0.745   1.351  1.272  0.303    2.096
## it4   -0.823   2.009  1.977  0.593    2.832
## it5   -1.660   1.286  2.029 -0.187    2.946
## it6   -0.308   1.487  1.035  0.590    1.795
## it7   -1.572   1.028  1.527 -0.272    2.600
## it8   -1.906   1.198  1.822 -0.354    3.104
## it9   -1.504   0.685  1.474 -0.410    2.189
## it10  -0.081   0.273  1.543  0.096    0.354
```

Vidimo da nema prelakih ni preteških stavki. Najlakša stavka je stavka 9 sa težinom od –0,41 logita, a najteža stavka 1 sa težinom od 0,96 logita. Ovi podaci samo delimično nam govore o nivoima osobine koji su pokriveni stavkama jer ulogu igraju i kategorije. Kao i kod ostalih politomnih modela, što je širi raspon između parametara kategorija stavka pokriva širi opseg merene latentne osobine. Stavka 8 pokriva najširi interval, a stavka 10 najuži. Najlakša kategorija u bilo kojoj stavci je kategorija 1 stavke 8, a najteža je kategorija 2 stavke 1. Najviši parametar nagiba ima stavka 5 ($a_5$=2,029), a najniži stavka 1 ($a_1$=1,005).

#### *10.5.6.1 Pokazatelji fita modela u celini*

Paket *ltm* nema pokazatelj fita za model stepenovanog odgovaranja, pa ćemo koristiti pokazatelj iz paketa *mirt*.

```
mirt::M2(mGRM.m)
```

```
##             M2 df        p RMSEA RMSEA_5   RMSEA_95      SRMSR      TLI CFI
## stats 23.43049 25 0.552441     0       0 0.02342672 0.01766954 1.000675   1
```

$M_2$ nije značajan što ukazuje da je model saglasan sa podacima. I pokazatelji približnog fita navode na isti zaključak.

#### *10.5.6.2 Pokazatelji fita za stavke*

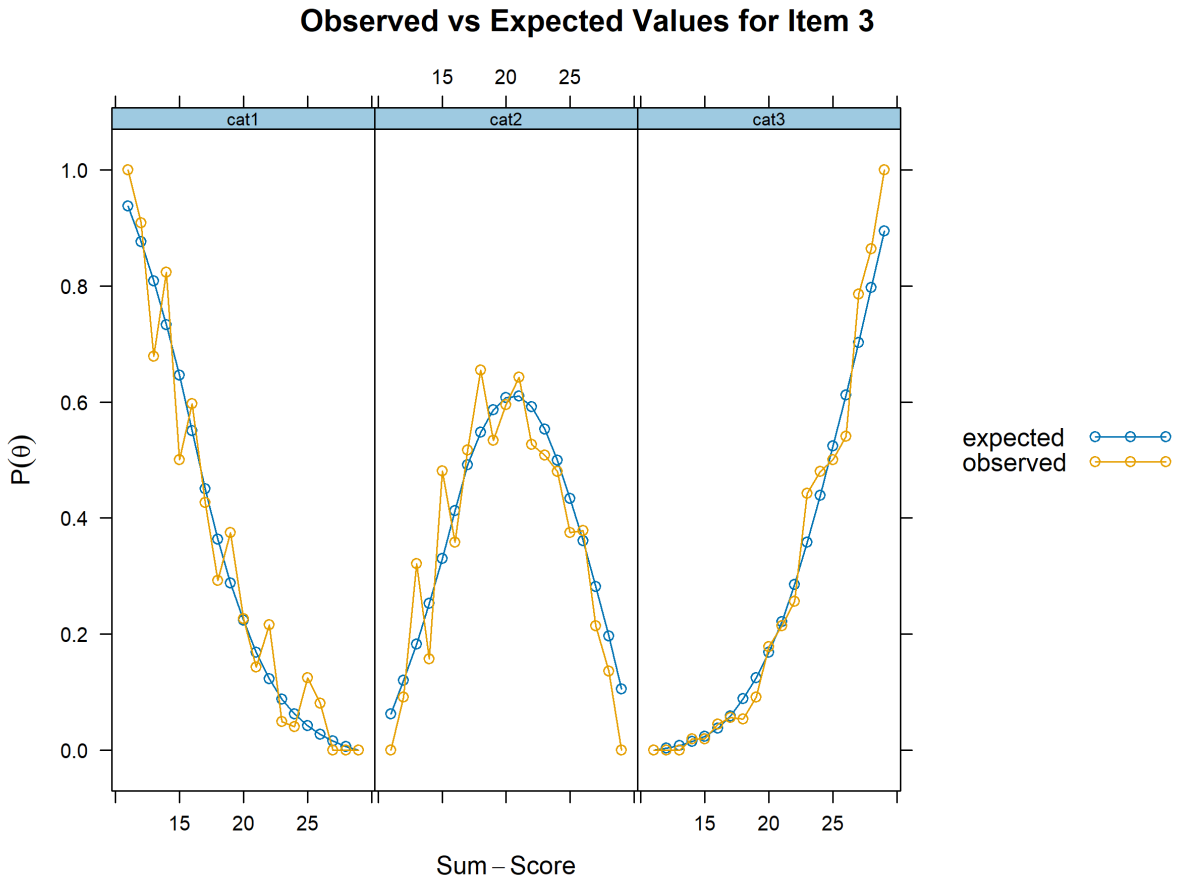
Pošto paket *ltm* nema uobičajene pokazatelje fita za stavke, i ovde ćemo koristiti paket *mirt*.

```
mirt::itemfit(mGRM.m, fit_stats = c("S_X2", "X2*"))
```

```
##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2 X2_star p.X2_star
## 1   it1 26.274      26      0.003  0.448   2.352     0.757
## 2   it2 19.827      21      0.000  0.532   5.171     0.125
## 3   it3 42.236      26      0.025  0.023   6.614     0.063
## 4   it4 15.927      21      0.000  0.774   3.575     0.388
## 5   it5 24.671      19      0.017  0.172   2.662     0.600
## 6   it6 31.272      27      0.013  0.260   6.113     0.089
## 7   it7 18.312      23      0.000  0.740   2.656     0.649
## 8   it8 20.056      20      0.002  0.454   4.785     0.204
```

```
## 9   it9 33.449      24      0.020  0.095   4.478     0.195
## 10 it10 24.167      25      0.000  0.510   4.020     0.183

# Ovom naredbom dobijamo podrazumevani hi-kvadrat po proceduri Orlanda i Tisena i
hi-kvadrat pokazatelj po proceduri koju je definisao Stoun
mirt::itemfit(mGRM.m,  fit_stats=c("S_X2"), S_X2.plot = 3)
```

```
# Argument S_X2.plot = 3 traži prikaz modelske i empirijske KKK za stavku koja je
označena brojem (u ovom slučaju to je stavka 3), ali onda se ne prikazuje tabela
sa numeričkim pokazateljima
```




Slika 105 – Modelske i empirijske karakteristične krive kategorija stavke 3

Prema izračunatim pokazateljima fita za stavke, jedino stavka 3 ima značajan misfit po $S–\chi^2$ statistiku, a po $\chi^{2*}$ statistiku na prikazanom nema, ali na ponovljenim računanjima ima. Podsetićemo $\chi^{2*}$ je zasnovan na parametarskim simulacijama i u ponovljenim izračunavanjima daje donekle različite rezultate. Ako stavka ima *p* vrednost blisku graničnoj, poželjno je ponoviti izračunavanje više puta (konkretno, vrednost prikazana u gornjoj tabeli nije značajna, ali se u ponovljenim računanjima dobijaju značajne vrednosti).

U paketu *ltm* možemo dobiti hi–kvadrate po rezidualima margina drugog i trećeg reda, odnosno, parova ili tripleta stavki.

```
margins(mGRM.l, type="two-way")

##
## Call:
## grm(data = p.pod, Hessian = TRUE)
##
## Fit on the Two-Way Margins
##
##      it1 it2   it3   it4   it5   it6   it7   it8   it9   it10
## it1  -    4.28  2.78  1.18  1.81  5.56  2.04  8.44  6.12  3.33
## it2       -     7.48  1.58  1.84  1.20  2.63  7.64  5.97 12.81
## it3             -     3.67  4.47  2.56  6.88  9.32  4.29  8.25
## it4                   -     0.89  5.99  2.02  6.25  0.96  8.39
## it5                         -     5.52  5.68  2.79  2.39  0.35
## it6                               -     4.17  8.37 10.41  6.21
## it7                                     -     1.43  3.57  3.94
## it8                                           -     1.45  4.34
## it9                                                 -     4.63
## it10                                                      -

margins(mGRM.l, type="three-way")

##
## Call:
## grm(data = p.pod, Hessian = TRUE)
##
## Fit on the Three-Way Margins
##
##     Item i Item j Item k (O-E)^2/E
## 1        1      2      3     18.24
## 2        1      2      4     13.85
## 3        1      2      5     14.56
## 4        1      2      6     18.76
## 5        1      2      7     16.17
## 6        1      2      8     26.12
## 7        1      2      9     27.87
## 8        1      2     10     28.25
## 9        1      3      4     14.64
## 10       1      3      5     18.13
## ...
## ...
## 119      7      9     10     15.87
## 120      8      9     10     14.99
```

U tabeli sa misfitom po parovima za politomne stavke postoji $m(m–1)/2$ parova stavki (gde je $m$ broj stavki). Značajan misfit je označen zvezdicama ispod dijagonale. U našoj tabeli takvih parova nema.

U izlazu za triplete stavki postoji $m(m–1)(m–2)/6$ tripleta. Značajan misfit je takođe označen zvezdicama desno od reda tripleta koji pokazuje misfit. U našem primeru takvih tripleta nema. Tabela je skraćena zbog uštede prostora.

Prema ovim pokazateljima nema značajnog misfita ni kada su u pitanju margine drugog ni kada su u pitanju margine trećeg reda.

#### *10.5.6.3 Procena lokalne zavisnosti*

I lokalnu zavisnost ćemo proceniti u paketu *mirt*, pošto *ltm* nema proceduru za to. Prvo ćemo prikazati procenu lokalne zavisnosti zasnovanu na hi–kvadratu i Kramerovim V koeficijentima.

```
residuals(mGRM.m, "LD")
```

```
## LD matrix (lower triangle) and standardized values.
##
## Upper triangle summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.072  -0.047  -0.024  -0.005   0.037   0.080
##
##        it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1     NA -0.046  0.037  0.024  0.030  0.053  0.032 -0.065 -0.055 -0.041
## it2  4.275     NA -0.061  0.028  0.030 -0.024 -0.036  0.062  0.055  0.080
## it3  2.775  7.479     NA -0.043 -0.047 -0.036 -0.059  0.068  0.046 -0.064
## it4  1.176  1.585  3.673     NA -0.021  0.055 -0.032 -0.056  0.022  0.065
## it5  1.808  1.845  4.467  0.891     NA -0.053  0.053 -0.037 -0.035  0.013
## it6  5.549  1.192  2.548  5.984  5.517     NA  0.046  0.065 -0.072 -0.056
## it7  2.038  2.631  6.875  2.028  5.687  4.173     NA  0.027 -0.042 -0.044
## it8  8.422  7.648  9.324  6.254  2.791  8.360  1.431     NA  0.027 -0.047
## it9  6.125  5.972  4.288  0.955  2.389 10.410  3.571  1.453     NA -0.048
## it10 3.329 12.797  8.248  8.385  0.350  6.200  3.939  4.335  4.623     NA
```

Ako kao kritičnu vrednost Kramerovog V uzmemo graničnu vrednost srednje veličine efekta po Koenu (Cohen, 1988) koja iznosi 0,212, vidimo da iznad dijagonale nema vrednosti koje su više, pa bismo mogli zaključiti da nema bitnije lokalne zavisnosti.

```
Q3m <- as.matrix(residuals(mGRM.m, type="Q3"))
```

```
## Q3 summary statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -0.170  -0.109  -0.094  -0.089  -0.074  -0.002
##
##         it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1   1.000 -0.108 -0.035 -0.067 -0.039 -0.019 -0.018 -0.124 -0.094 -0.071
## it2  -0.108  1.000 -0.105 -0.129 -0.114 -0.113 -0.116 -0.092 -0.082 -0.088
## it3  -0.035 -0.105  1.000 -0.097 -0.125 -0.069 -0.099 -0.065 -0.074 -0.096
## it4  -0.067 -0.129 -0.097  1.000 -0.110 -0.059 -0.170 -0.118 -0.086 -0.074
## it5  -0.039 -0.114 -0.125 -0.110  1.000 -0.109 -0.088 -0.140 -0.103 -0.077
## it6  -0.019 -0.113 -0.069 -0.059 -0.109  1.000 -0.002 -0.030 -0.107 -0.094
## it7  -0.018 -0.116 -0.099 -0.170 -0.088 -0.002  1.000 -0.095 -0.097 -0.102
## it8  -0.124 -0.092 -0.065 -0.118 -0.140 -0.030 -0.095  1.000 -0.085 -0.093
## it9  -0.094 -0.082 -0.074 -0.086 -0.103 -0.107 -0.097 -0.085  1.000 -0.131
## it10 -0.071 -0.088 -0.096 -0.074 -0.077 -0.094 -0.102 -0.093 -0.131  1.000
```

```
diag(Q3m) <- NA
Q3mean <-sum(Q3m, na.rm = TRUE)/(ncol(Q3m)*(ncol(Q3m)-1))
cat("\nProsečna vrednost Q3:", round(Q3mean, 3), "\n\n")
```

```
##
## Prosečna vrednost Q3: -0.089
```

```
# Matrica Q3 korigovana za prosečnu vrednost
Q3c <- Q3m-Q3mean
```

```
# Vrednosti veće od .20 ukazuju na lokalnu zavisnost
Q3c
```

```
##          it1    it2    it3    it4    it5    it6    it7    it8    it9   it10
## it1       NA -0.019  0.054  0.022  0.050  0.070  0.071 -0.035 -0.005  0.019
## it2   -0.019     NA -0.016 -0.040 -0.025 -0.024 -0.027 -0.003  0.007  0.001
## it3    0.054 -0.016     NA -0.008 -0.036  0.021 -0.010  0.024  0.015 -0.007
## it4    0.022 -0.040 -0.008     NA -0.021  0.030 -0.081 -0.029  0.003  0.015
## it5    0.050 -0.025 -0.036 -0.021     NA -0.020  0.001 -0.051 -0.014  0.012
## it6    0.070 -0.024  0.021  0.030 -0.020     NA  0.087  0.059 -0.018 -0.005
## it7    0.071 -0.027 -0.010 -0.081  0.001  0.087     NA -0.006 -0.008 -0.013
## it8   -0.035 -0.003  0.024 -0.029 -0.051  0.059 -0.006     NA  0.004 -0.004
## it9   -0.005  0.007  0.015  0.003 -0.014 -0.018 -0.008  0.004     NA -0.042
## it10   0.019  0.001 -0.007  0.015  0.012 -0.005 -0.013 -0.004 -0.042     NA
```

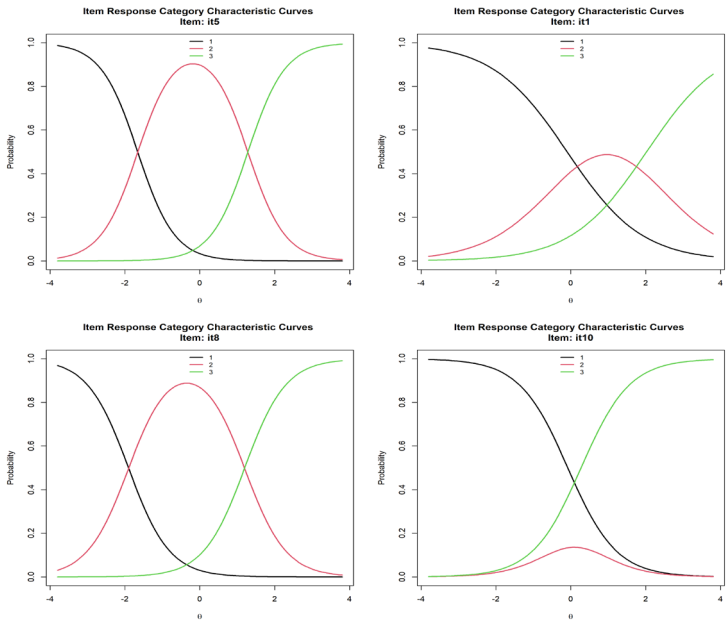
```
# Ako želite da u matrici vrednosti niže od 0,2 zamenite sa NA izvršite i sledeće
dve naredbe
#Q3c[Q3c<.20] <- NA
#Q3c
```

Ni na osnovu $Q_3$ pokazatelja ne bi se moglo reći da je narušena pretpostavka lokalne nezavisnosti. Matricu Q3c formirali smo tako što smo od vrednosti u matrici Q3m oduzeli prosečnu vrednost $Q_3$ pokazatelja u toj matrici. Kao pokazatelj postojanja lokalne zavisnosti smatrali smo vrednosti u matrici Q3c veće od 0,2 (videti odeljak 9.2).

#### *10.5.6.4 Grafički prikaz karakterističnih kriva kategorija*

I u paketu *ltm* ako zatražimo prikaz karakterističnih kriva stavke za politomne modele dobićemo ustvari prikaz karakterističnih kriva kategorija. Prikazaćemo samo neke, a koristeći paket *ggmirt* prikazaćemo sve.

```
par(mfrow=c(2,2))
plot(mGRM.l, items=c(5,1,8,10), type="ICC", lwd=2, legend=T,
 xlab=expression(theta), ask=F)
```



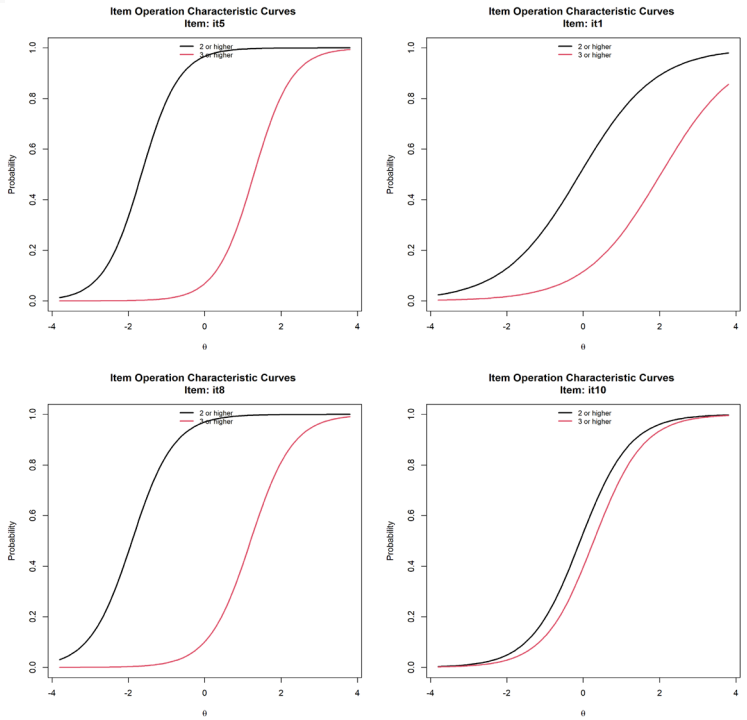

Slika 106 – Karakteristične krive kategorija stavki 5, 1, 8 i 10

```
par(mfrow=c(1,1))
```

Prikazane su 4 stavke (Slika 106). Stavka 5 je najdiskriminativnija i to se vidi po nagibu KKK. Vidimo da i kategorija 2, na nivou osobine nešto malo ispod 0 logita ima verovatnoću da bude birana skoro 0,9. Ova kategorija je u relativno širokom rasponu latentne osobine najverovatnija. Pored ove stavke vidimo stavku 1, koja ima najmanji parametar nagiba i mali razmak između kategorija. Srednja kategorija u malom opsegu latentne osobine je najverovatnija, a verovatnoća da bude izabrana ne prelazi 0,5. Stavka 8 je prikazana jer ima širok raspon između kategorija, a njena prva kategorija je najlakša u celom testu. I na kraju, stavka 10 koja je u modelu stepenovanog ocenjivanja imala poremećaj redosleda kategorija, ovde to nema, ali funkcioniše kao dihotomna stavka. Srednja kategorija nikada nije najverovatnija, a umesto KKK smo mogli prikazati samo krivu kategorije 3 (zelena) jer se proces odgovaranja praktično svodi na izbor između kategorije 1 ili 3 (nije sasvim tačno).

U paketu *ltm* možemo tražiti i prikaz operativnih kriva kategorija. To su funkcije binarnih odluka između niže kategorije ili neke od viših. One izgledaju isto kao KKS binarno skorovanih stavki i za svaku stavku ih ima za jednu manje od broja kategorija odgovora.

```
par(mfrow=c(2,2))
plot(mGRM.l, items=c(5,1,8,10), type="OCCu", lwd=2, legend=T,
 xlab=expression(theta), ask=F)
```




Slika 107 – Operativne (granične) krive kategorija stavki 5, 1, 8 i 10

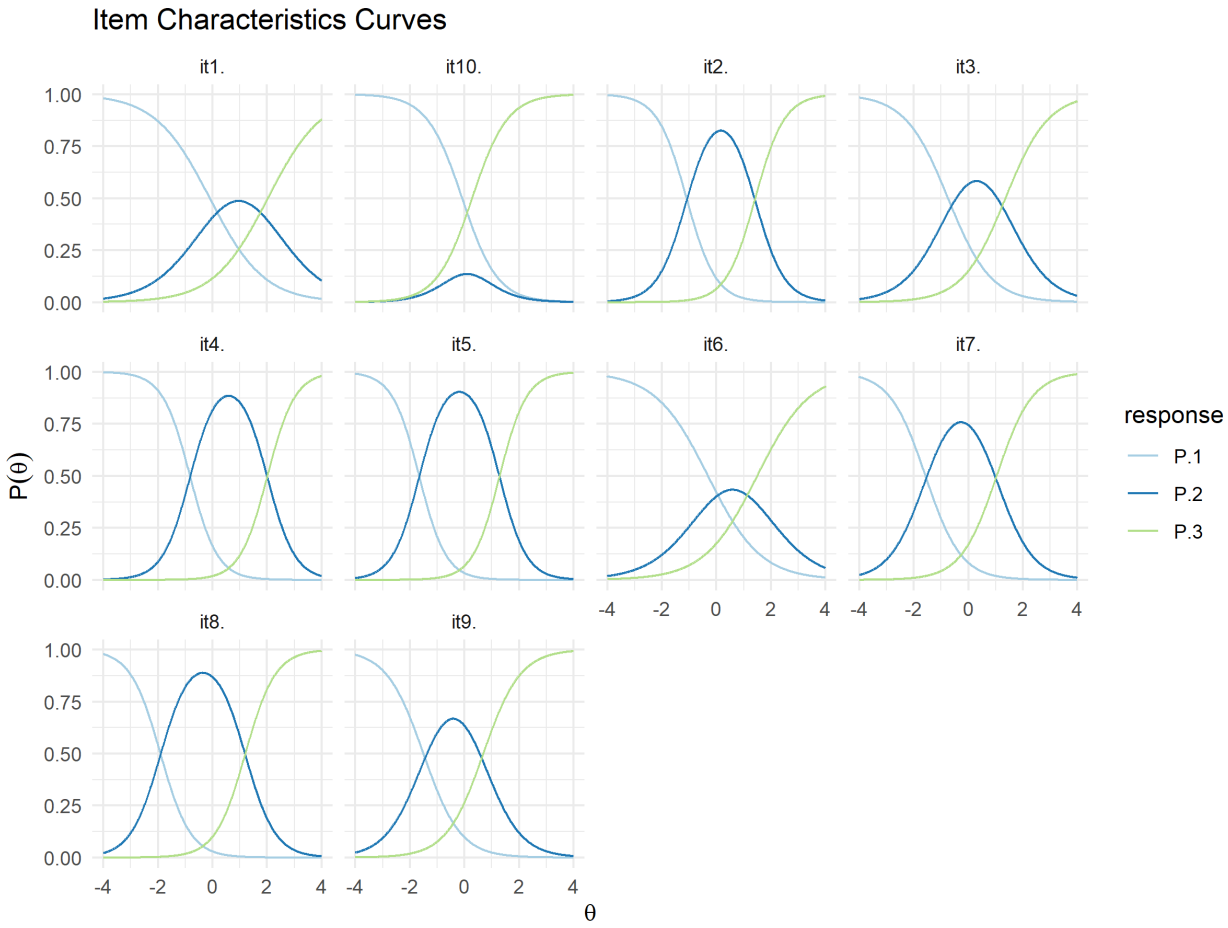
```
par(mfrow=c(1,1))
```

Na gornjem grafikonu (Slika 107) crnim linijama su označene operativne krive kategorije 2 za svaku od prikazanih stavki. One prikazuju verovatnoću da će ispitanik, u zavisnosti od nivoa osobine, izabrati kategoriju 2 ili neku od viših kategorija, umesto kategorije 1. Operativne krive kategorija za stavku 10 su vrlo

bliske, ali se ne preklapaju, tako da ono što je rečeno u vezi sa time da bi se proces odgovaranja na stavku 10 mogao svesti na izbor kategorija 1 ili 3 ne stoji u potpunosti.

Ovo je način da dobijete karakteristične krive kategorija koristeći paket *ggmirt*, a na osnovu modela procenjenog u paketu *mirt* (Slika 108).

```
library(ggmirt)
p_load(ggplot2)
p_load(RColorBrewer)
tracePlot(mGRM.m, facet = FALSE, legend = TRUE)+
 scale_color_brewer(palette = "Paired")
```

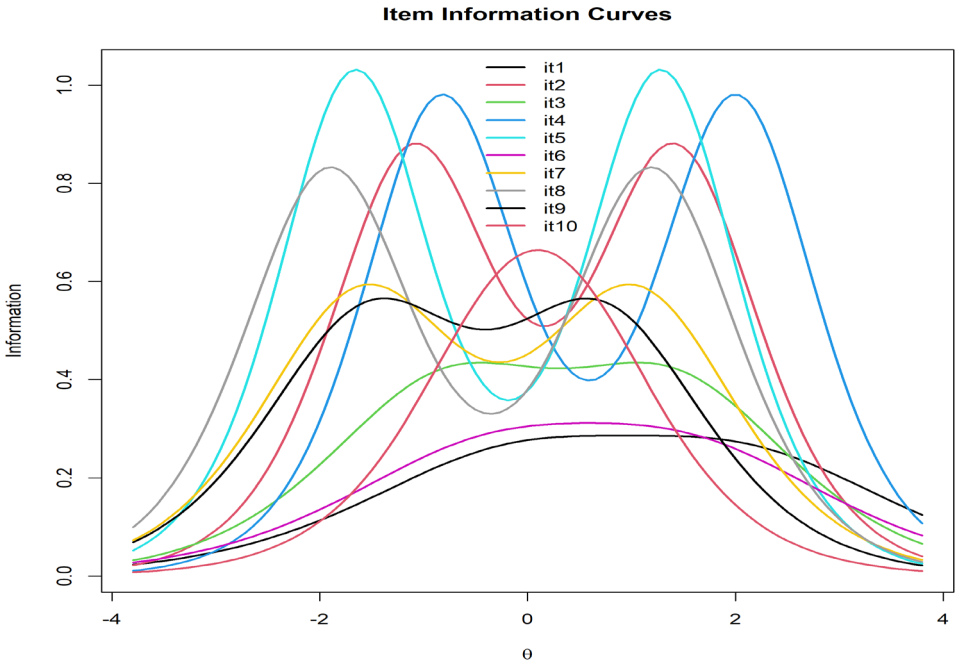



Slika 108 – Karakteristične krive kategorija u modelu stepenovanog odgovaranja

### *10.5.6.5 Informativnost stavki i testa*

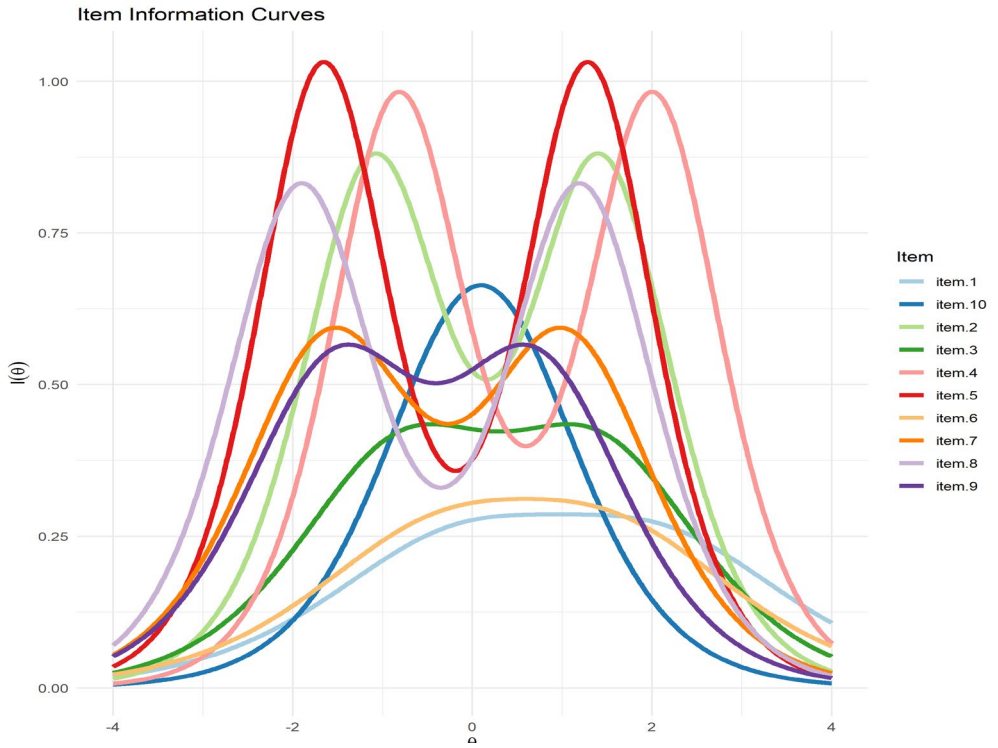
Ono što je rečeno za funkcije informativnosti stavki u generalizovanom modelu stepenovanog ocenjivanja, stoji i u modelu stepenovanog odgovaranja. Oblik i visina funkcije zavisiće od nagiba i razmaka kategorija. Prikazaćemo krive pomoću paketa *ltm* (Slika 109) i paketa *ggmirt*, *ggplot2* i *RColorBrewer* (Slika 110).

```
plot(mGRM.l, type="IIC", lwd=2, legend=T, xlab=expression(theta))
```




Slika 109 – Funkcije informativnosti stavki u modelu stepenovanog odgovaranja (*ltm*)

```
# type="IIC" određuje da budu prikazane krive informativnosti
library(ggmirt)
p_load(ggplot2)
p_load(RColorBrewer)
```



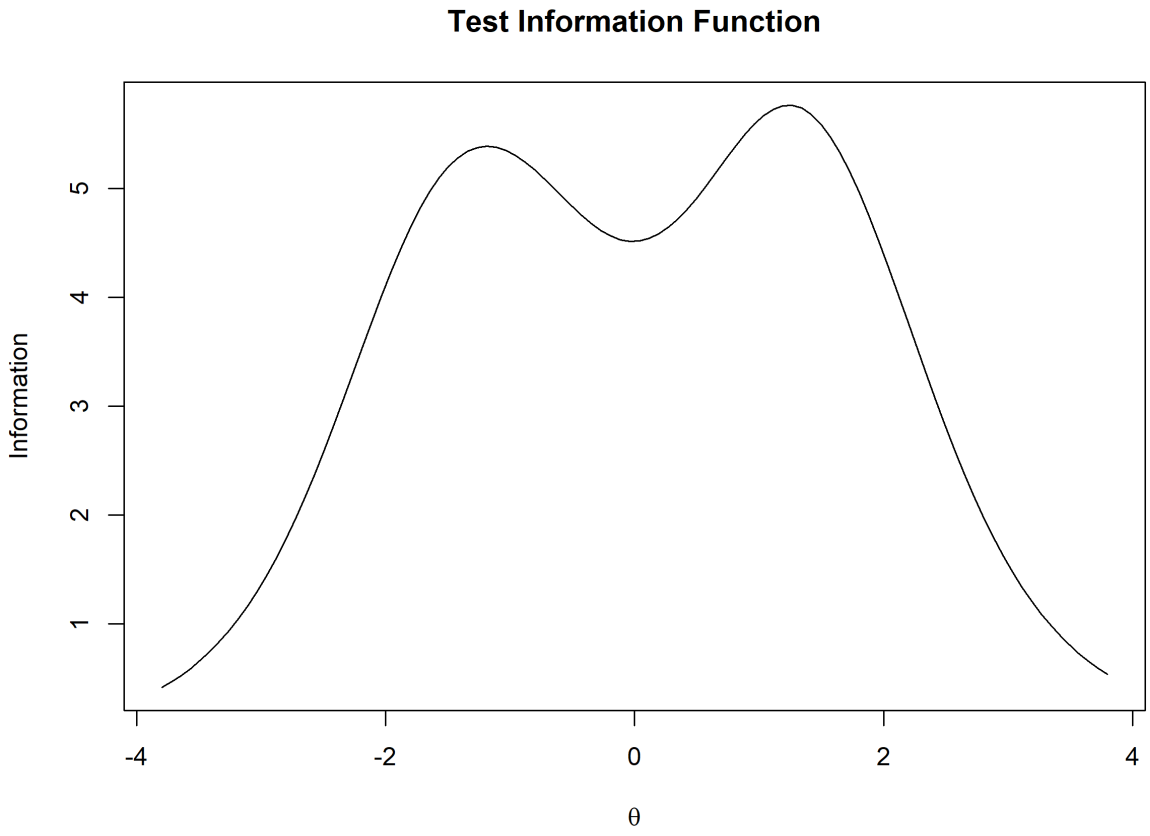

Slika 110 – Funkcije informativnosti stavki u modelu stepenovanog odgovaranja (*ggmirt*)

```
itemInfoPlot(mGRM.m, legend = T) + scale_color_brewer(palette = "Paired") +
 ggplot2::geom_line(size = 1.5)
```

Mada grafikoni (Slika 109 i Slika 110) izgledaju slično kao kod generalizovanog modela stepenovanog ocenjivanja, bimodalnost kriva je izraženija. Najinformativnije stavke su stavke 5 i 4, ali je kod njih i bimodalnost najizraženija. Slede stavke 2 i 8, a zatim 7, 9 i 3. Unimodalne i najniže funkcije informativnosti imaju stavke 6 i 1. Stavka 10 ima unimodalnu ali višu krivu, nalik onima za binarno skorovane stavke.

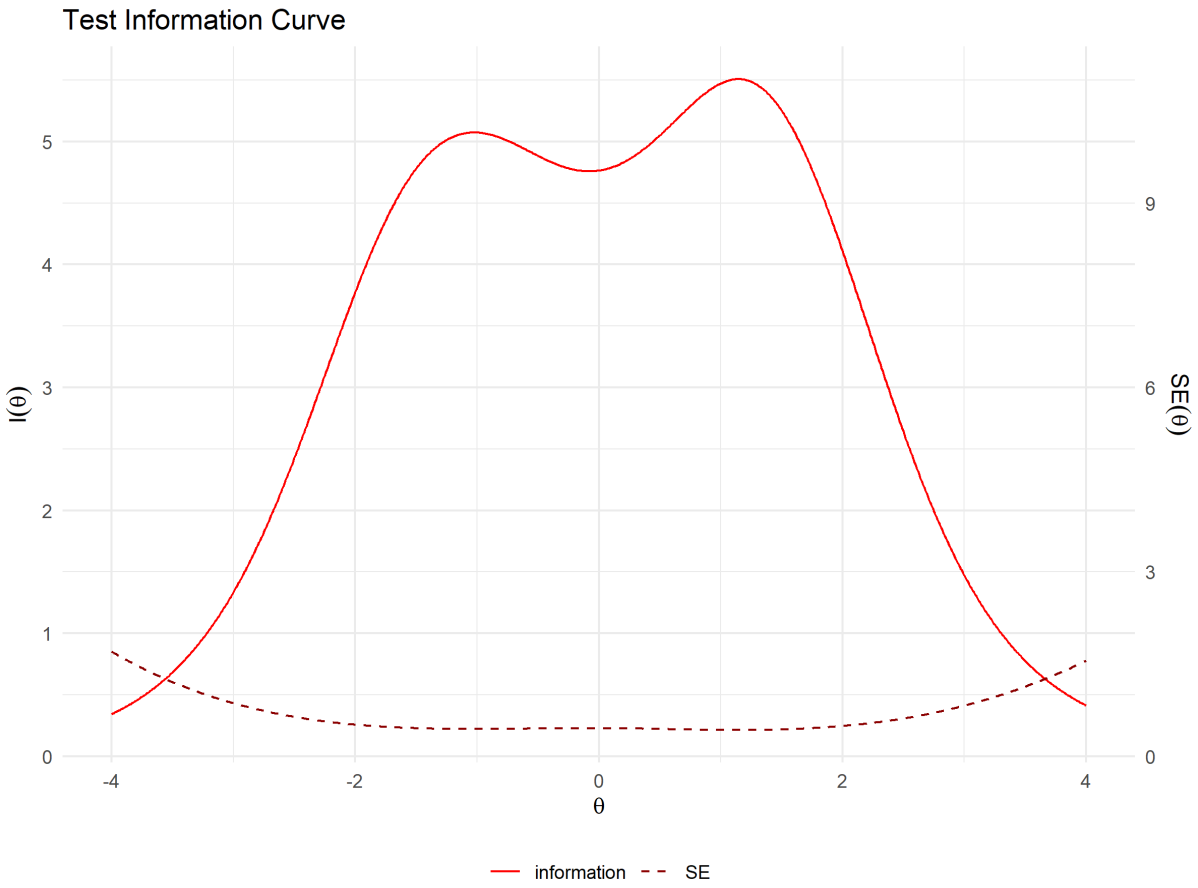
Funkcija informativnosti testa predstavlja sumu funkcija informativnosti stavki po nivoima latentne osobine. Zadovoljavajući nivo informativnosti (preko 3,33) ovaj model ima u rasponu između –2 i 2 logita (otprilike). Kriva je bimodalna (Slika 111 i Slika 112), a viši vrh se nalazi negde na nivou osobine oko 1 logita. Standardna greška (Slika 112) je prilično ujednačena između –2 i 2 logita, a van tog opsega je nešto viša.

```
# U paketu ltm
plot(mGRM.l, type="IIC", items=0, xlab=expression(theta))
```




Slika 111 – Funkcija informativnosti testa  u modelu za stepenovano odgovaranje

```
# U paketu ggmirt
testInfoPlot(mGPCM.m, adj_factor = 2)
```




Slika 112 – Funkcije informativnoti i standardne greške testa

#### 10.5.6.5.1 Numerički pokazatelji informativnosti

Na osnovu numeričkih pokazatelja preciziraćemo ono što je izrečeno u opisu grafikona. Izračunaćemo prvo informativnost u opsegu između –3 i 3 logita, gde očekujemo najveći broj ispitanika.

```
# Za ceo test
mGRMpi <- areainfo(mGRM.m, c(-3,3))
mGRMpi

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 25.75303  27.87002  0.9240406     10

# Samo za određene stavke
areainfo(mGPCM.m, c(-3,3), which.items = c(1,2,4:10)) # Bez stavke 3

##  LowerBound UpperBound    Info TotalInfo Proportion nitems
##          -3          3 23.0925  25.07356  0.9209904      9
```

Informativnost po nivoima merene osobine u paketu *ltm* možemo dobiti na sledeći način:

```
# Informativnost u paketu ltm možemo dobiti iz grafikona ako podatke za njegovo
crtanje sačuvamo u objekat, a argumentu plot dodelimo vrednost FALSE
Info <- data.frame(plot(mGRM.l, type="IIC", items=0, range=c(-6,6), plot=FALSE))

# U paketu mirt informativnost možemo dobiti na već više puta opisan način
thetaM <- seq(-6,6, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(mGRM.m, Theta = thetaM)
# Prikaz prvih 6 redova (ukolonite head() za ispis cele tabele)
TI <- data.frame(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
head(TI)

##   nivo.osobine informativnost
## 1         -6.0     0.01645987
## 2         -5.9     0.01889474
## 3         -5.8     0.02171019
## 4         -5.7     0.02496866
## 5         -5.6     0.02874320
## 6         -5.5     0.03311939

# Maksimalna informativnost
cat("Maksimalna informativnost:", max(infoM), "\n")

## Maksimalna informativnost: 5.760339

# Ispis nivoa osobine na kom je informativnost maksimalna
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n", thetaM[infoM==max(infoM)], "\n")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  1.2

# Ispis raspona u kojem je informativnost veća od 3,33 što odgovara pouzdanosti od
0,7
cat("Raspon latentne osobine u kojem je test zadovoljavajuće informativan: \n",
 min(thetaM[infoM>3.33]), "do", max(thetaM[infoM>3.33]), "logita")

## Raspon latentne osobine u kojem je test zadovoljavajuće informativan:
##  -2.2 do 2.3 logita

cat("\nInformativnost u rasponu od -6 do 6 logita: \n")

##
## Informativnost u rasponu od -6 do 6 logita:

inf66 <- sum(TINFO)
inf66

## [1] 197.1287

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 6 logita:  \n")

##
## Informativnost u rasponu od 0 do 6 logita:
```

```
inf06 <- sum(TINFO[!thetaM < 0], na.rm = TRUE)
inf06

## [1] 117.5398

cat(" \nProcenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od
-6 do 6 logita:  \n")

##
## Procenat informativnosti u rasponu od 0 do 6 logita, u odnosu na raspon od -6
do 6 logita:

round(inf06/inf66*100,2)

## [1] 59.63
```

Najviša informativnost testa je 5,76 na nivou osobine od 1,2 logita (drugi, niži vrh je na –1,2 gde je informativnost 5,39). Informativnost je veća (59,63%) u opsegu natprosečnih nivoa osobine. Zadovoljavajuća je u rasponu od –2,2 do +2,3 logita (za nijansu širem nego u generalizovanom modelu stepenovanog odgovaranja.

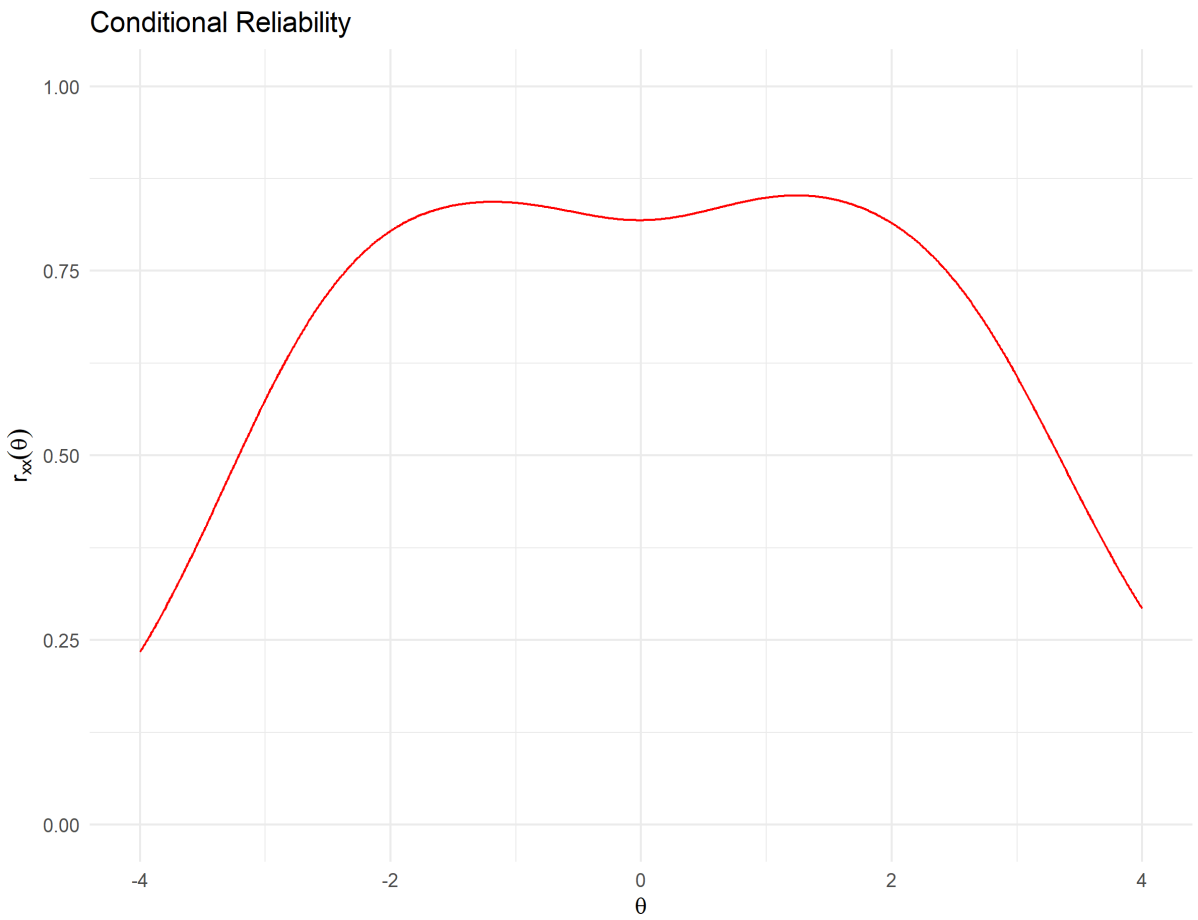
#### *10.5.6.6 Marginalna i empirijska pouzdanost*

Izračunaćemo empirijsku i marginalnu pouzdanost (u paketu *mirt*).

```
# Parametre ispitanika i standardne greške dobijamo funkcijom fscores() i smeštamo
u objekat THETA_SE koji prosleđujemo funkciji za računanje empirijske pouzdanosti
THETA_SE <- fscores(mGRM.m, full.scores.SE = TRUE)
cat("\nEmpirijska pouzdanost:", round(empirical_rxx(THETA_SE),3))

##
## Empirijska pouzdanost: 0.829

cat("\nMarginalna pouzdanost:", round(marginal_rxx(mGRM.m),3))

##
## Marginalna pouzdanost: 0.829

cat("\nKronbahova alfa:", round(ltm::cronbach.alpha(p.pod)$alpha,3))

##
## Kronbahova alfa: 0.809
```

Tri procene pouzdanosti su vrlo slične i dobre (>0,80) i ne razlikuju se puno od onih dobijenih na generalizovanom modelu stepenovanog ocenjivanja. Prikazaćemo i grafikon uslovne pouzdanosti (Slika 113).

```
ggmirt::conRelPlot(mGRM.m)
```




Slika 113 – Uslovna pouzdanost testa

#### *10.5.6.7 Mere ispitanika*

Mere ispitanika je moguće dobiti i u paketu *ltm,* ali nisu date za svakog ispitanika već po sklopovima odgovora. Podsećamo, svi ispitanici sa istim sklopovima odgovora imaju iste mere. U paketu *mirt* mere ispitanika dobijaju se jednostavnije i direktno se mogu spojiti sa izvornim podacima, pa ćemo ih izračunati na taj način.

```
MERE <- fscores(mGRM.m)
```

#### *10.5.6.8 Izbacivanje stavki*

Model sa svim stavkama saglasan je sa podacima, ali stavka 3 je imala značajan misfit. Zbog toga ćemo je izbaciti i ponovo proveriti model. Nakon njenog izbacivanja i stavka 6 je pokazala misfit pa smo i nju izbacili (prikazan je samo korak kada su izbačene obe stavke).

```
mGRM.mR <- mirt(p.pod[,-c(3,6)], 1, itemtype = "graded", SE=T, verbose = FALSE)

M2(mGRM.mR)

##             M2 df         p      RMSEA RMSEA_5   RMSEA_95      SRMSR        TLI
## stats 11.48524 12 0.4878534          0       0 0.03110692 0.01537241  1.000427
##             CFI
## stats         1
```

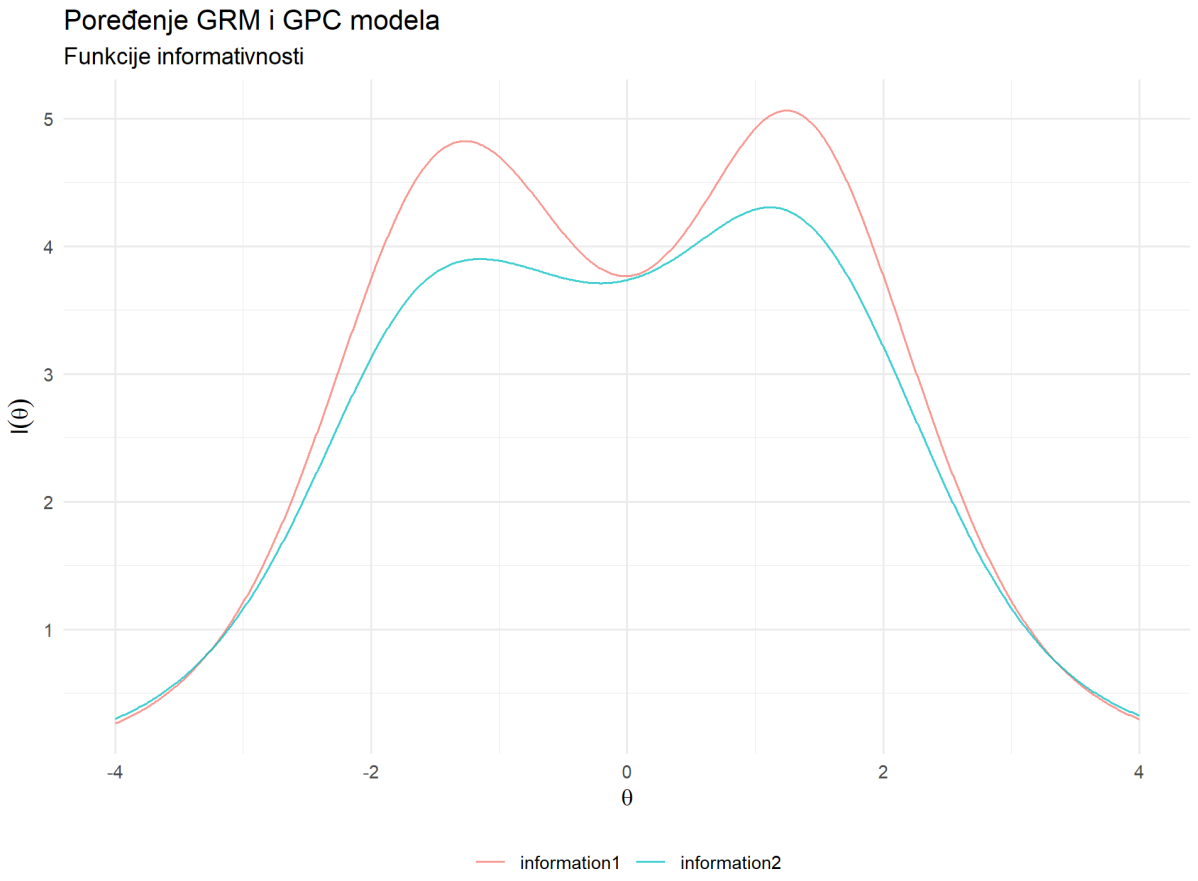
```
mirt::itemfit(mGPCM.mR,  fit_stats=c("X2*"))

##   item X2_star p.X2_star
## 1  it1   2.136     0.770
## 2  it2   4.346     0.137
## 3  it4   3.032     0.449
## 4  it5   3.154     0.317
## 5  it7   2.347     0.637
## 6  it8   3.113     0.475
## 7  it9   3.267     0.351
## 8 it10   4.162     0.114

PF.mR <- mirt::personfit(mGRM.mR, method = "EAP")
p_load(dplyr) # Paket potreban za funkciju reframe()
reframe(PF.mR, infit = prop.table(table(z.infit > 1.96 | z.infit < -1.96)),
 outfit = prop.table(table(z.outfit > 1.96 | z.outfit < -1.96)),
 Zh = prop.table(table(Zh > 1.96 | Zh < -1.96)))

##   infit outfit     Zh
## 1 0.973  0.982  0.970
## 2 0.027  0.018  0.030

# U donjem redu je proporcija misfitujućih ispitanika (dobro je kada je manja od 0
,05)
THETA_SE <- fscores(mGRM.mR, full.scores.SE = TRUE)
cat("\nEmpirijska pouzdanost:", round(empirical_rxx(THETA_SE),3))

##
## Empirijska pouzdanost: 0.807

cat("\nMarginalna pouzdanost:", round(marginal_rxx(mGRM.mR),3))

##
## Marginalna pouzdanost: 0.806

cat("\nKronbahova alfa:", round(psych::alpha(p.pod[,-c(2,3)])$total$raw_alpha,3))

##
## Kronbahova alfa: 0.784
```

Nakon izbacivanja stavke 3 i stavka 6 je pokazala značajan misfit pa je i ona izbačena. Prikazani su rezultati nakon izbacivanja obe stavke. Proporcija misfitujućih ispitanika je niža (ispod 5%). Pouzdanosti su nešto niže (nego pre izbacivanja stavki), ali to nije drastično (za oko 0,02). Marginalna i empirijska pouzdanost su dobre (>0,8).

#### *10.5.6.9 Poređenje funkcija informativnosti dva modela*

Uporedićemo funkcije informativnosti redukovanih modela stepenovanog odgovaranja i generalizovanog modela stepenovanog ocenjivanja.

```
ggmirt::testInfoCompare(mGRM.mR, mGPCM.mR, title = "Poređenje punog i redukovanog
GPC modela", subtitle="Funkcije informativnosti")
```



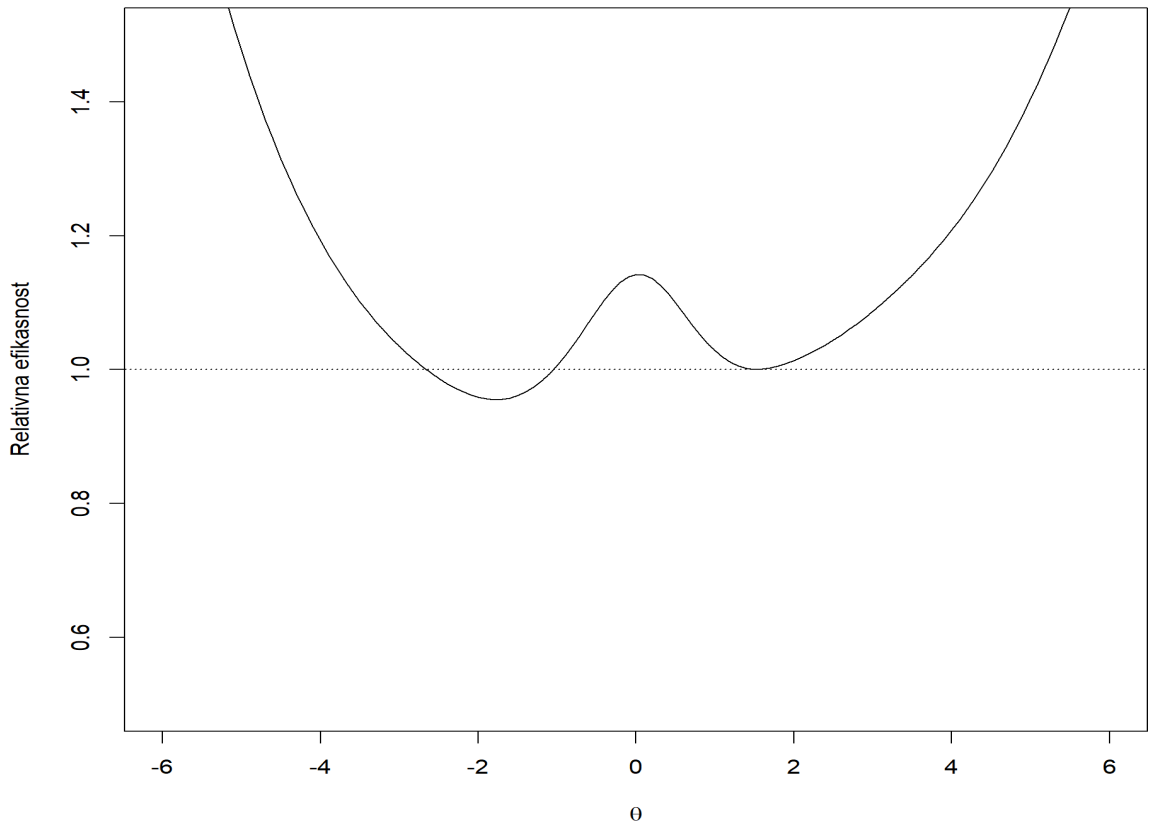

Slika 114 – Uporedni prikaz funkcija informativnosti redukovanog modela za stepenovano odgovaranje i redukovanog generalizovanog modela za stepenovano ocenjivanje

Oblik dve krive informativnosti (Slika 114) je sličan, ali funkcija informativnosti modela stepenovanog odgovaranja (GRM) je izraženije bimodalna i viša u rasponu osobine između –3 i +3 logita (osim u uskom delu oko 0 logita gde se krive dodiruju). Prikazaćemo to kroz relativnu efikasnost.

#### *10.5.6.10 Relativna efikasnost*

Izračunaćemo i prikazati relativnu efikasnost redukovanog modela generalizovanog modela stepenovanog ocenjivanja u odnosu na redukovanu verziju modela stepenovanog odgovaranja.

```
# Izračunaćemo informativnosti dva modela u rasponu osobine od -6 do 6 logita
thetaRI = seq(-6, 6, .1)
GR <- testinfo(mGRM.mR, Theta = thetaRI)
GPC <- testinfo(mGPCM.mR, Theta = thetaRI)
RE <- GPC/GR
plot(thetaRI, RE, type="l", xlab=expression(theta), ylim=c(0.5,1.5),
 ylab="Relativna efikasnost")
#axis(2, at=seq(.5, 1.5, .1), labels=seq(.5, 1.5, .1))
abline(h=1, lty=3)
```




Slika 115 – Relativna efikasnost generalizovanog modela stepenovanog ocenjivanja

Grafikon (Slika 115) predstavlja odnos funkcija informativnosti GPC i GGRM modela. Horizontalna isprekidana linija označava jednaku efikasnost dve forme. Generalizovani model stepenovanog ocenjivanja postiže veću efikasnost od modela stepenovanog odgovaranja na nivoima osobine nižim od –2,5 i višim od 2 logita. S obzirom na to da se u tim delovima kontinuuma latentne osobine ne očekuje puno ispitanika, to i nije toliko bitno. Bitnija je veća efikasnost u rasponu od –1 do +1 logita. Model stepenovanog odgovaranja efikasniji je samo na nivou osobine oko –2 logita.

### *10.5.6.11 Formiranje sirovog skora i korelacije sa merama*

Izračunaćemo ukupni skor i korelirati sa IRT merama ispitanika na osnovu modela stepenovanog odgovaranja.

```
SKOR <- rowSums(p.pod)
df <- data.frame(cbind(SKOR, MERE))
cat("Pirsonov koeficijent korelacije između sumacionog skora i IRT mera: \n")

## Pirsonov koeficijent korelacije između sumacionog skora i IRT mera:

cor(SKOR, MERE, method="pearson")

##             F1
## [1,] 0.9876094
```

```
cat(" \nSpirmanov koeficijent korelacije između sumacionog skora i IRT mera:  \n")
```

```
##
## Spirmanov koeficijent korelacije između sumacionog skora i IRT mera:
```

```
cor(SKOR, MERE, method="spearman")
```

```
##             F1
## [1,] 0.9879577
```

I Pirsonov i Spirmanov koeficijent korelacije sumacionog skora i IRT mera dobijenih modelom stepenovanog odgovaranja su vrlo visoki (oba 0,988), ali poredak ispitanika po sumacionim skorovima i IRT merama nije isti. I u ovom modelu ukupan skor nije dovoljan statistik za procenu mera osobine. Dovoljan statistik je skor zasnovan na sumi ajtemskih skorova ponderisanih diskriminativnošću.

#### *10.5.6.12 Poređenje svih modela*

Uporedićemo sve do sada prikazane politomne modele LR testom, a uporedićemo i vrednosti informacionih kriterijuma za svaki od njih.

```
anova(mRSM.m, mPCM.m, mGPCM.m)
```

```
##              AIC    SABIC       HQ      BIC    logLik       X2 df p
## mRSM.m  18162.06 18182.84 18184.45 18220.96 -9069.031
## mPCM.m  16510.15 16546.51 16549.32 16613.21 -8234.074 1669.915  9 0
## mGPCM.m 16346.54 16398.49 16402.50 16493.77 -8143.268  181.611  9 0
```

```
anova(mRSM.m, mPCM.m, mGRM.m)
```

```
##             AIC    SABIC       HQ      BIC    logLik       X2 df p
## mRSM.m 18162.06 18182.84 18184.45 18220.96 -9069.031
## mPCM.m 16510.15 16546.51 16549.32 16613.21 -8234.074 1669.915  9 0
## mGRM.m 16340.59 16392.54 16396.54 16487.82 -8140.293  187.562  9 0
```

```
anova(mGPCM.m, mGRM.m)
```

```
##              AIC    SABIC       HQ      BIC    logLik    X2 df   p
## mGPCM.m 16346.54 16398.49 16402.50 16493.77 -8143.268
## mGRM.m  16340.59 16392.54 16396.54 16487.82 -8140.293 5.951  0 NaN
```

Kada uporedimo model skala procene, model stepenovanog ocenjivanja i generalizovani model stepenovanog ocenjivanja vidimo da je po vrednostima AIC, BIC, HQ i logLik GPCM model najbolji od tri poređena modela. Značajno je bolji od PCM modela, koji je opet značajno bolji od RSM modela.

Sličnu situaciju imamo kada GRM uporedimo sa RSM i PCM. GRM ima najbolje pokazatelje AIC, BIC i logLik. Značajno je bolji od PCM.

GPCM i GRM imaju jednak broj parametara pa je broj stepeni slobode hi–kvadrata za poređenje razlike logLik jednak 0 i ne dobijamo *p* vrednost. Zato ćemo model uporediti na osnovu AIC i BIC. GRM ima za nijansu bolje pokazatelje, pa bismo mogli reći da on najbolje opisuje podatke (ovde nismo poredili redukovane modele).

### 10.5.7 Procena nominalnog modela u paketu mirt

Bokov (Bock, 1972) nominalni model namenjen je politomnim stavkama sa nominalnim skorovanjem, odnosno, neuređenim kategorijama odgovora. Može se koristiti za analizu distraktora u kognitivnim stavkama, ali nam može poslužiti i za analizu skala u nekognitivnim stavkama. Simuliraćemo podatke na kojima ćemo prikazati analizu.

#### *10.5.7.1 Simulacija podataka za nominalni model*

Simulacija nominalnih podataka u paketu *mirt.*

```
p_load(mirt)
ss <- 3
set.seed(ss)
nagibi <- runif(5, .9, 1.7)
# Nagibi/diskriminativnosti ajtema simulirani su iz uniformne distribucije, sa
minimalnom vrednošću 0.9, a maksimalnom 1.7
# Matrica odsečaka mora imati redova onoliko koliko želite stavki, i kolona
koliko želite kategorija odgovora
d <- matrix(c(
  0, .1, .7, 0,
  0, 1, .2, .1,
  0, 1, .5, 1.2,
  0, .2, .5, .7,
  3, 2, .5, 0),ncol=4,byrow=TRUE)
# Matrica ak parametara za kategorije
noma <- matrix(c(
  0, .2, 1, 2,
  0, .5, -.3, 2,
  0, .4, -.08, 2,
  0, .3, -.1, 1,
  1.5, .7, -.7, -.1
), ncol=4, byrow=TRUE)

# Zbog reproducibilnosti potrebno je pre simulacije parametara uvek primeniti
komandu set.seed(ss). Vrednost ss je postavljena na početku
set.seed(ss)
thete <- rnorm(1000, mean=0, sd=1)
# Nivoi osobine ispitanika simulirani iz normalne distribucije sa AS=0 i SD=1

set.seed(ss)
n.pod <- data.frame(mirt::simdata(a=nagibi, d=d, u=1, N=1000,
 itemtype = "nominal", nominal=noma))
# Imenovanje varijabli (stavki)
names(n.pod) <- paste0("it",1:5)
# Prikaz prva tri reda generisanih podataka
head(n.pod,3)

##   it1 it2 it3 it4 it5
## 1   1   2   0   2   3
## 2   1   2   1   2   0
## 3   0   2   3   0   0
```

#### *10.5.7.2 Procena modela*

Za procenu ovog modela koristićemo paket *mirt*. Postupak je isti kao i za ostale modele u ovom paketu, osim što je argument *itemtype* potrebno postaviti na *nominal*. Parametre za kategorije smo sačuvali

u objektu pod imenom *parametri*. Poslednje dve komande koje su komentarisane služe da prikažu da su sume parametara kategorija (kako nagiba, tako i odsečaka) ograničene na 0. Ukoliko želite da ih izvršite morate prethodno ukloniti #.

```
nom.m <- mirt(n.pod, 1, itemtype = "nominal", SE=T, verbose = FALSE)
nom.m

##
## Call:
## mirt(data = n.pod, model = 1, itemtype = "nominal", SE = T, verbose = FALSE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 21 EM iterations.
## mirt version: 1.41
## M-step optimizer: BFGS
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix =  84645269
##
## Log-likelihood = -5986.733
## Estimated parameters: 30
## AIC = 12033.47
## BIC = 12180.7; SABIC = 12085.42
## G2 (993) = 918.23, p = 0.9561
## RMSEA = 0, CFI = NaN, TLI = NaN

parametri <- coef(nom.m, simplify=TRUE, IRTpars=TRUE)$items
round(parametri, 3)

##         a1     a2     a3    a4     c1     c2     c3     c4
## it1 -0.766 -0.673  0.055 1.385 -0.176 -0.125  0.532 -0.231
## it2 -0.551 -0.071 -1.362 1.985 -0.326  0.693 -0.267 -0.101
## it3 -0.598 -0.235 -0.798 1.630 -0.617  0.363 -0.243  0.497
## it4 -0.425 -0.020 -0.396 0.841 -0.322 -0.120  0.131  0.311
## it5  0.872 -0.096 -1.774 0.998  1.540  0.466 -0.911 -1.095

# round(rowSums(parametri[,1:4]),3)
# round(rowSums(parametri[,5:8]),3)
```

Pozivanjem imena objekta u koji smo zapisali model (*nom.m*) ispisujemo podatke o toku procene modela, odnosno, o broju iteracija koji je bio potreban za procenu modela. Takođe, dobijamo vrednosti AIC, BIC, SABIC i vrednost LR testa, odnosno, $G^2$ omnibus pokazatelja fita modela. S obzirom na to da $G^2_{(993)}$=918,23, $p$=0,956 nije značajan, možemo reći da je model saglasan sa podacima.

Pošto nominalni model može biti numerički nestabilan, korisno je dati neke startne vrednosti parametrima kako bi vrednosti parametara bile procenjene u opsegu od 0 do $k$–1 (gde je $k$ broj kategorija). Koristićemo nemodelski pristup definisanju startnih vrednosti koji predlaže Čalmers (Chalmers, 2012; de Ayala, 2022). U ovom pristupu kao početne vrednosti parametara postavljaju se vrednosti ranga koji se na osnovu učestalosti biranja dodeljuju svakoj od kategorija. Na svakoj stavci najučestalija kategorija dobija

najviši rang, a najređa najniži, rangovi imaju vrednosti od 0 do *k*–1, ili u našem slučaju od 0 do 3. Najučestalija kategorija na svakoj stavci dobiće rang 3, a najređa 0. Ostale kategorije će dobiti rangove u skladu sa učestalošću biranja. Ti rangovi biće ujedno i startne vrednosti za *ak* parametre kategorija.

Prvo ćemo napraviti tabele proporcija biranja kategorija i odrediti rangove.

```
prop <- data.frame(ltm::descript(n.pod)$perc)
prop # Tabela proporcija

##        X0    X1    X2    X3
## it1 0.204 0.204 0.326 0.266
## it2 0.131 0.321 0.229 0.319
## it3 0.118 0.265 0.195 0.422
## it4 0.175 0.192 0.272 0.361
## it5 0.595 0.194 0.166 0.045

rang <- t(apply(prop, MARGIN = 1, FUN = rank))
rang # Tabela rangova

##      X0  X1 X2 X3
## it1 1.5 1.5  4  3
## it2 1.0 4.0  2  3
## it3 1.0 3.0  2  4
## it4 1.0 2.0  3  4
## it5 4.0 3.0  2  1
```

Zatim je potrebno definisati model koji će biti procenjen u paketu *mirt*. Pošto želimo da procenimo jednodimenzionalni model za 5 stavki, model ćemo definisati na sledeći način: md <- "F1=1-5". Zatim je potrebno dodati startne vrednosti za svaki parametar u formi *START = (1, ak0, 0.5)*. Definicija mora biti u novom redu i počinjati ključnom rečju *START* i znakom "=", nakon čega u zagradi stoji redni broj stavke, oznaka parametra i startna vrednost parametra. Kako je to prilično pipav i dosadan posao napravili smo sledeći algoritam koji to radi umesto nas (koristimo tabelu *rang* koju smo ranije formirali):

```
md <- "F1=1-5 \n"
for(i in 1:nrow(rang)){
  for(j in 1:ncol(rang)){
    md <- paste0(md, "START = (", i, ", ak", j-1,", ", rang[i,j]-1,") \n",
     collapse = "\n")
  }
}
md

## [1] "F1=1-5 \nSTART = (1, ak0, 0.5) \nSTART = (1, ak1, 0.5) \nSTART = (1, ak2,
3) \nSTART = (1, ak3, 2) \nSTART = (2, ak0, 0) \nSTART = (2, ak1, 3) \nSTART = (2,
ak2, 1) \nSTART = (2, ak3, 2) \nSTART = (3, ak0, 0) \nSTART = (3, ak1, 2) \nSTART
= (3, ak2, 1) \nSTART = (3, ak3, 3) \nSTART = (4, ak0, 0) \nSTART = (4, ak1, 1) \n
START = (4, ak2, 2) \nSTART = (4, ak3, 3) \nSTART = (5, ak0, 3) \nSTART = (5, ak1,
2) \nSTART = (5, ak2, 1) \nSTART = (5, ak3, 0) \n"

md <- mirt.model(md) # Funkcijom mirt.model() string md pretvaramo u objekat tipa
mirt.model
```

```
md

## $x
##       Type    Parameters
##  [1,] "F1"    "1-5"
##  [2,] "START" "(1,ak0,0.5)"
##  [3,] "START" "(1,ak1,0.5)"
##  [4,] "START" "(1,ak2,3)"
##  [5,] "START" "(1,ak3,2)"
##  [6,] "START" "(2,ak0,0)"
##  [7,] "START" "(2,ak1,3)"
##  [8,] "START" "(2,ak2,1)"
##  [9,] "START" "(2,ak3,2)"
## [10,] "START" "(3,ak0,0)"
## [11,] "START" "(3,ak1,2)"
## [12,] "START" "(3,ak2,1)"
## [13,] "START" "(3,ak3,3)"
## [14,] "START" "(4,ak0,0)"
## [15,] "START" "(4,ak1,1)"
## [16,] "START" "(4,ak2,2)"
## [17,] "START" "(4,ak3,3)"
## [18,] "START" "(5,ak0,3)"
## [19,] "START" "(5,ak1,2)"
## [20,] "START" "(5,ak2,1)"
## [21,] "START" "(5,ak3,0)"
##
## attr(,"class")
## [1] "mirt.model"
```

Kada imamo kreiran model, možemo ga proceniti.

```
nom.m <- mirt::mirt(n.pod, model=md, itemtype = "nominal", SE=T, verbose = FALSE)
nom.m

##
## Call:
## mirt::mirt(data = n.pod, model = md, itemtype = "nominal", SE = T,
##     verbose = FALSE)
##
## Full-information item factor analysis with 1 factor(s).
## Converged within 1e-04 tolerance after 27 EM iterations.
## mirt version: 1.41
## M-step optimizer: BFGS
## EM acceleration: Ramsay
## Number of rectangular quadrature: 61
## Latent density type: Gaussian
##
## Information matrix estimated with method: Oakes
## Second-order test: model is a possible local maximum
## Condition number of information matrix =  506.0909
##
## Log-likelihood = -5959.072
## Estimated parameters: 30
## AIC = 11978.14
## BIC = 12125.38; SABIC = 12030.09
## G2 (993) = 862.91, p = 0.9988
## RMSEA = 0, CFI = NaN, TLI = NaN
```

```
parametri <- coef(nom.m, simplify=TRUE, IRTpars=TRUE)$items
round(parametri, 3)

##         a1     a2     a3     a4     c1     c2     c3     c4
## it1 -0.770 -0.667  0.056  1.381 -0.178 -0.122  0.531 -0.232
## it2 -0.572 -0.048 -1.370  1.990 -0.330  0.698 -0.269 -0.099
## it3 -0.607 -0.231 -0.798  1.636 -0.619  0.366 -0.240  0.493
## it4 -0.431 -0.012 -0.398  0.841 -0.323 -0.119  0.132  0.310
## it5  1.569  0.375 -1.487 -0.457  1.592  0.584 -0.913 -1.263

round(prop, 3)

##        X0    X1    X2    X3
## it1 0.204 0.204 0.326 0.266
## it2 0.131 0.321 0.229 0.319
## it3 0.118 0.265 0.195 0.422
## it4 0.175 0.192 0.272 0.361
## it5 0.595 0.194 0.166 0.045

# round(rowSums(parametri[,1:4]),3)
# round(rowSums(parametri[,5:8]),3)
```

Model je konvergirao nakon 27 iteracija. $G^2_{(993)}$=862.91, $p$=0,999 nije značajan, pa možemo reći da je model saglasan sa podacima. Informacioni kriterijumi AIC, BIC i SABIC daju prednost ovom modelu u odnosu na onaj bez startnih vrednosti (uporediti vrednosti informacionih kriterijuma).

$M_2$ omnibus pokazatelj fita ovde nije moguće izračunati zbog nedovoljnog broja stepeni slobode.

Parametri nagiba kategorija označeni su sa *a* i brojem kategorije. Kategorije sa najvišim pozitivnim parametrom nagiba su one koje najčešće biraju ispitanici sa višim nivoom osobine. U kognitivnim stavkama to su tačni odgovori, a u nekognitivnim to su kategorije koje najčešće biraju ispitanici sa najvišim nivoom osobine. Parametri odsečka odgovaraju učestalosti biranja kategorija u uzorku.

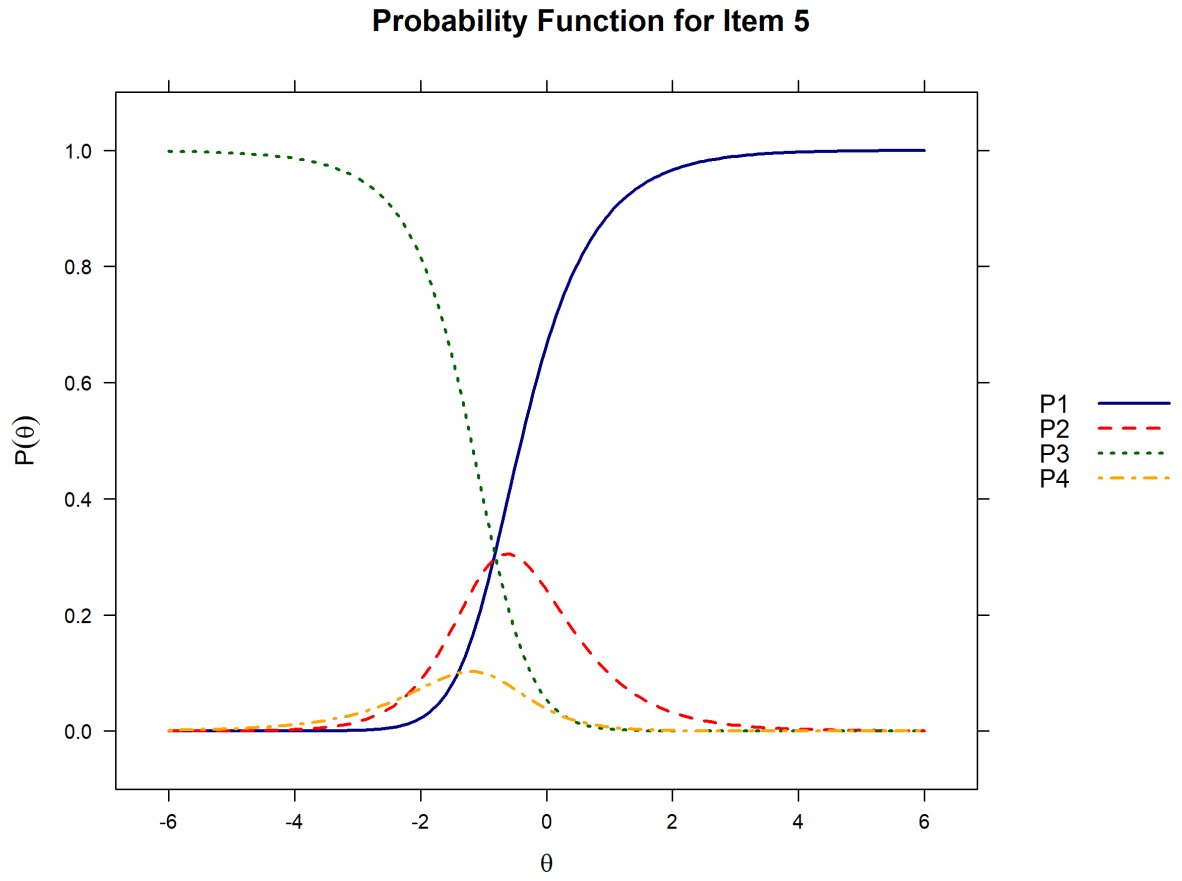
Pogledajmo na primeru stavke 5.

```
itemplot(nom.m, item=5,type = "trace",par.settings = simpleTheme(lty=1:4, lwd=2,
 col=c("navy", "red", "darkgreen", "orange")))
```

Uz prikaz karakterističnih kriva kategorija izračunate su i težine kategorija 1 i 2 na dva načina koji predlažu De Ajala (de Ayala, 2022) i Bejker i Kim (Baker & Kim, 2004) (videti odeljak 4.2.2).

```
cat("Tačka na kontinuumu latentne osobine u kojoj je jednaka verovatnoća biranja
kategorija 1 i 2: ",
 (parametri[5,"c1"]-parametri[5,"c2"])/(parametri[5,"a2"]-parametri[5,"a1"]),
 "(prema: de Ayala, 2022)\n")

## Tačka na kontinuumu latentne osobine u kojoj je jednaka verovatnoća biranja
kategorija 1 i 2:  -0.8442889 (prema: de Ayala, 2022)

cat("Procena parametra b za kategoriju 1: " ,
 -(parametri[5,"c1"]/parametri[5,"a1"]), "(prema: Baker i Kim, 2024)\n")

## Procena parametra b za kategoriju 1:  -1.014516 (prema: Baker i Kim, 2024)

cat("Tačka na kontinuumu latentne osobine u kojoj je jednaka verovatnoća biranja
kategorija 2 i 3: ",
 (parametri[5,"c2"]-parametri[5,"c3"])/(parametri[5,"a3"]-parametri[5,"a2"]),
 "(prema: de Ayala, 2022)\n")
```

```
## Tačka na kontinuumu latentne osobine u kojoj je jednaka verovatnoća biranja
kategorija 2 i 3:  -0.8038322 (prema: de Ayala, 2022)

cat("Procena parametra b za kategoriju 2: " ,
 -(parametri[5,"c2"]/parametri[5,"a2"]), "(prema: Baker i Kim, 2024)\n")

## Procena parametra b za kategoriju 2:  -1.556394 (prema: Baker i Kim, 2024)
```




Slika 116 – Karakteristične krive kategorija stavke 5

Na grafikonu (Slika 116) prikazane su karakteristične krive kategorija. Grafikon podseća na grafikon politomne stavke sa uređenim kategorijama iz GPCM modela. Karakteristična kriva kategorije 1 (*P1*, $a_{51}$=1,569, $c_{51}$=1,592) je kategorija odgovora čija verovatnoća konstantno raste sa porastom nivoa osobine. Kada bi ovo bila kognitivna stavka (a to ne znamo jer je simulirana), to bi bila kriva tačnog odgovora. Ovo je najčešće birana kategorija u uzorku. Karakteristična kriva kategorije 3 (*P3*, $a_{53}$=–1,487, $c_{53}$=0,132) opada monotono sa porastom nivoa osobine. Ovo je kriva sa najnižim parametrom nagiba (najvišim negativnim). Kategorija 4 (*P4* $a_{54}$=–0,457, $c_{54}$=–1,263) je privlačna ispitanicima na nivou osobine između –2 i –1 logita. Najslabije je povezana sa merenom osobinom i najređe birana. Kategorija 2 (*P2*, $a_{52}$=0,375, $c_{52}$=0,583) je privlačna ispitanicima sa nivoom osobine ispod proseka, između –1 i 0 logita. Kategorije 2 i 4 nikada nisu najverovatnije pojedinačne kategorije. Ova stavka bi dobro funkcionisala kao dihotomna.

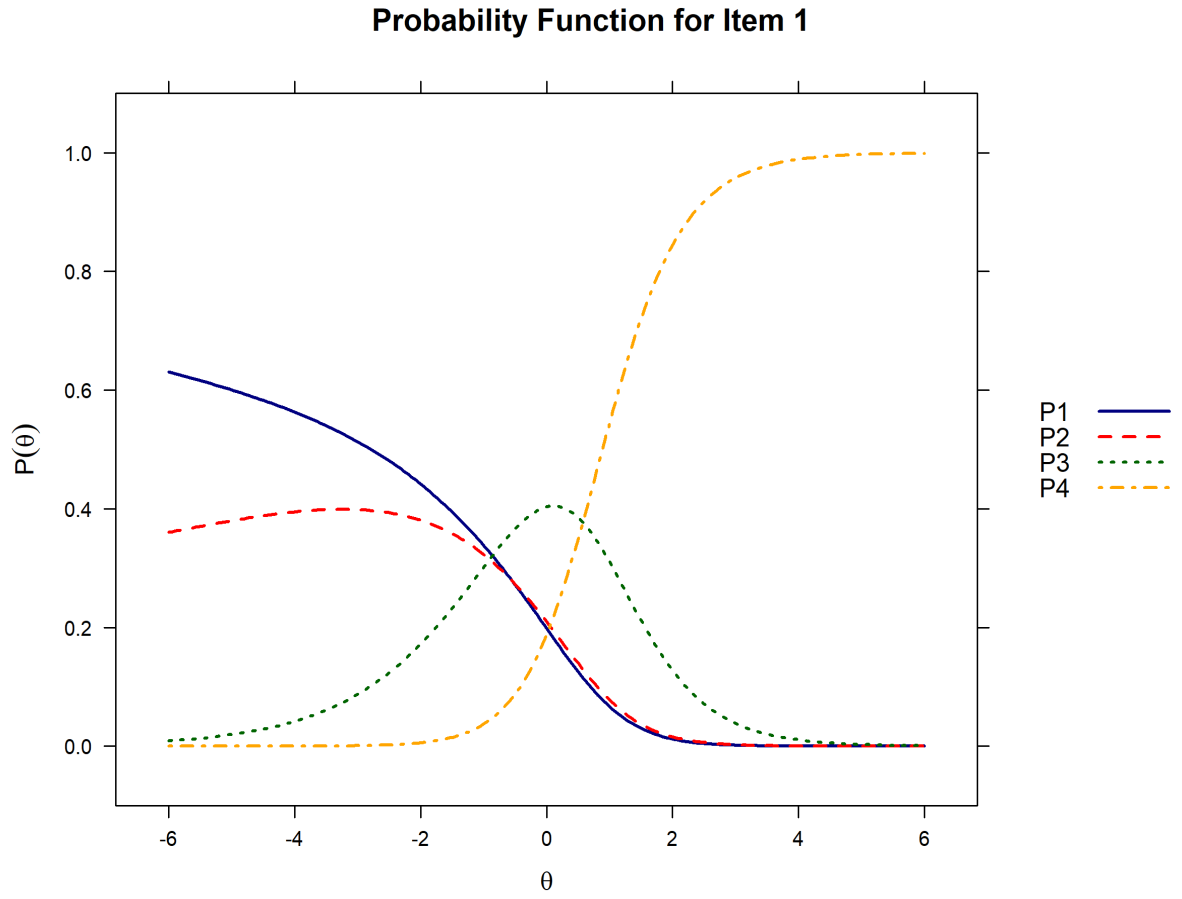
10.5.7.2.1 Pokazatelji fita za stavke

Kao pokazatelj fita za stavke koristićemo Stounov hi–kvadrat statistik.

```
itemfit(nom.m, fit_stats = "X2*")

##   item X2_star p.X2_star
## 1  it1   1.766     0.252
## 2  it2   0.958     0.484
## 3  it3   0.561     0.870
## 4  it4   0.723     0.929
## 5  it5   0.685     0.846
```

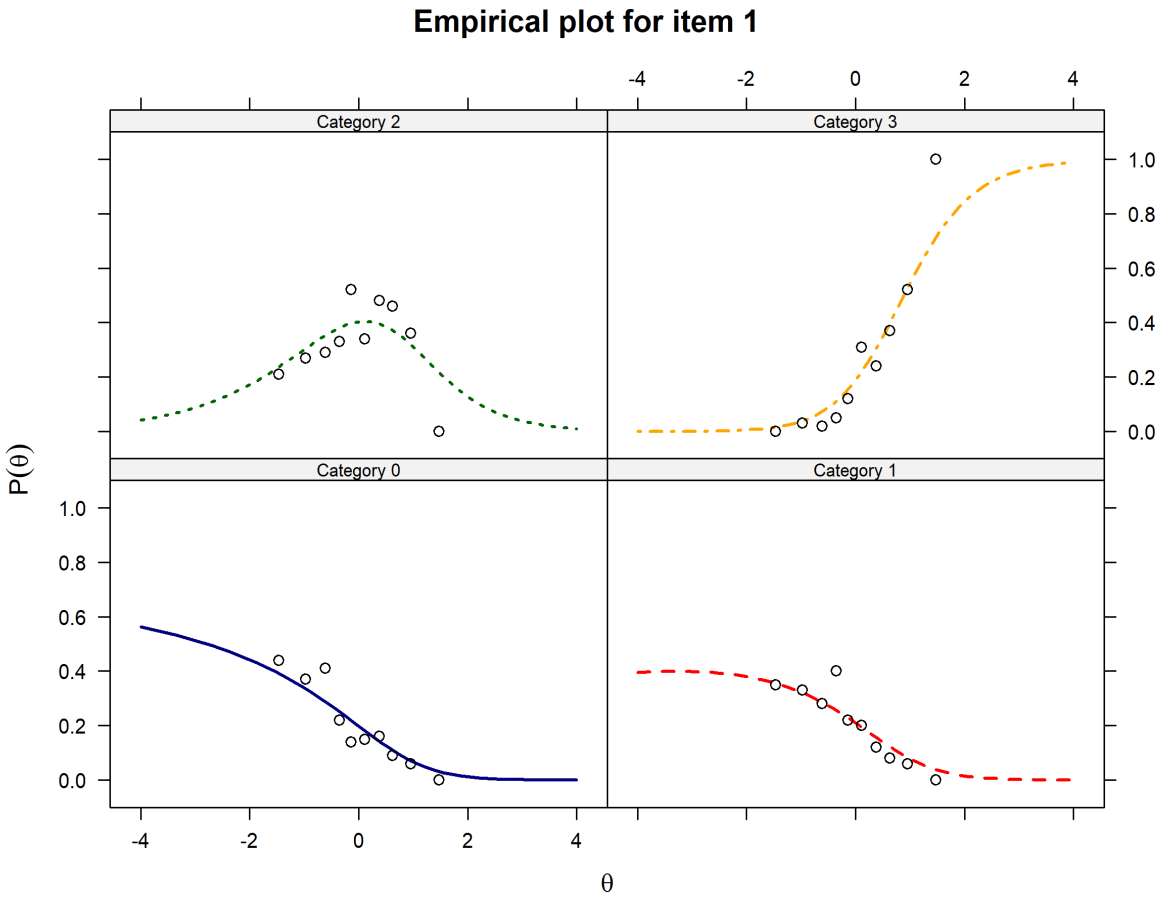
Nema misfitujućih stavki, a stavka 1 ima pokazatelj fita najbliži značajnosti pa ćemo pogledati grafikon sa empirijskim i modelskim krivama kategorija (Slika 117 i Slika 118). Uz prikaz karakterističnih kriva kategorija izračunate su i težine kategorija 3 i 4 na dva načina koji predlažu De Ajala (de Ayala, 2022) i Bejker i Kim (Baker & Kim, 2004) (videti odeljak 4.2.2).

```
itemplot(nom.m, item=1,type = "trace", par.settings = simpleTheme(lty=1:4, lwd=2,
 col=c("navy", "red", "darkgreen", "orange")))
```




Slika 117 – Karakteristične krive kategorija stavke 1

```
itemfit(nom.m, empirical.plot = 1, par.settings = simpleTheme(lty=1:4, lwd=2,
 col=c("navy", "red", "darkgreen", "orange")))
```




Slika 118 – Modelske i empirijske karakteristične krive kategorija stavke 1

Kategorija 4 (*P4*) ima najviši pozitivni parametar nagiba ($a_4$=1,38) i njena kriva liči na karakterističnu krivu binarno skorovane stavke. Verovatnoća biranja ove kategorije raste sa porastom nivoa latentne osobine. U uzorku je to druga najčešće birana kategorija ($c_4$=0,266).

Kategorije 1 i 2 imaju slične, negativne parametre nagiba. Kategorija 1 (*P1*) ima najniži parametar nagiba ($a_{11}$=–0,770) i verovatnoća njenog biranja monotono opada sa rastom nivoa latentne osobine. Ova kategorija je najređe birana ($c_{11}$=–0,177). Kategorija 2 (*P2*) ima nešto viši parametar nagiba ($a_{12}$=–0,667) i verovatnoća njenog biranja raste do –3 logita a zatim opada sa rastom nivoa latentne osobine. Ova kategorija je druga najređe birana ($c_{12}$=–0,122). Verovatnoća biranja kategorije 3 (*P3*, $a_{13}$=0,056) raste sa nivoom latentne osobine negde do prosečnog nivoa (0 logita), a zatim kreće da opada. U rasponu od –1 do 0,5 logita ova kategorija ima najveću pojedinačnu verovatnoću biranja. Povezanost ove kategorije sa merenom osobinom je bliska 0. Ovo je ujedno najčešće birana kategorija ($c_{13}$=0,531). U kontekstu kognitivnog testiranja ova kategorija bi mogla biti atraktor (najbolji distraktor). Pošto su podaci simulirani, možemo zamisliti da je stavka kognitivna pa bismo u tom slučaju mogli reći da distraktori 1 i 2 privlače gotovo istu grupu ispitanika (sudeći po nivou latentne osobine), tako da je jedan verovatno suvišan. Da je u pitanju nekognitivna

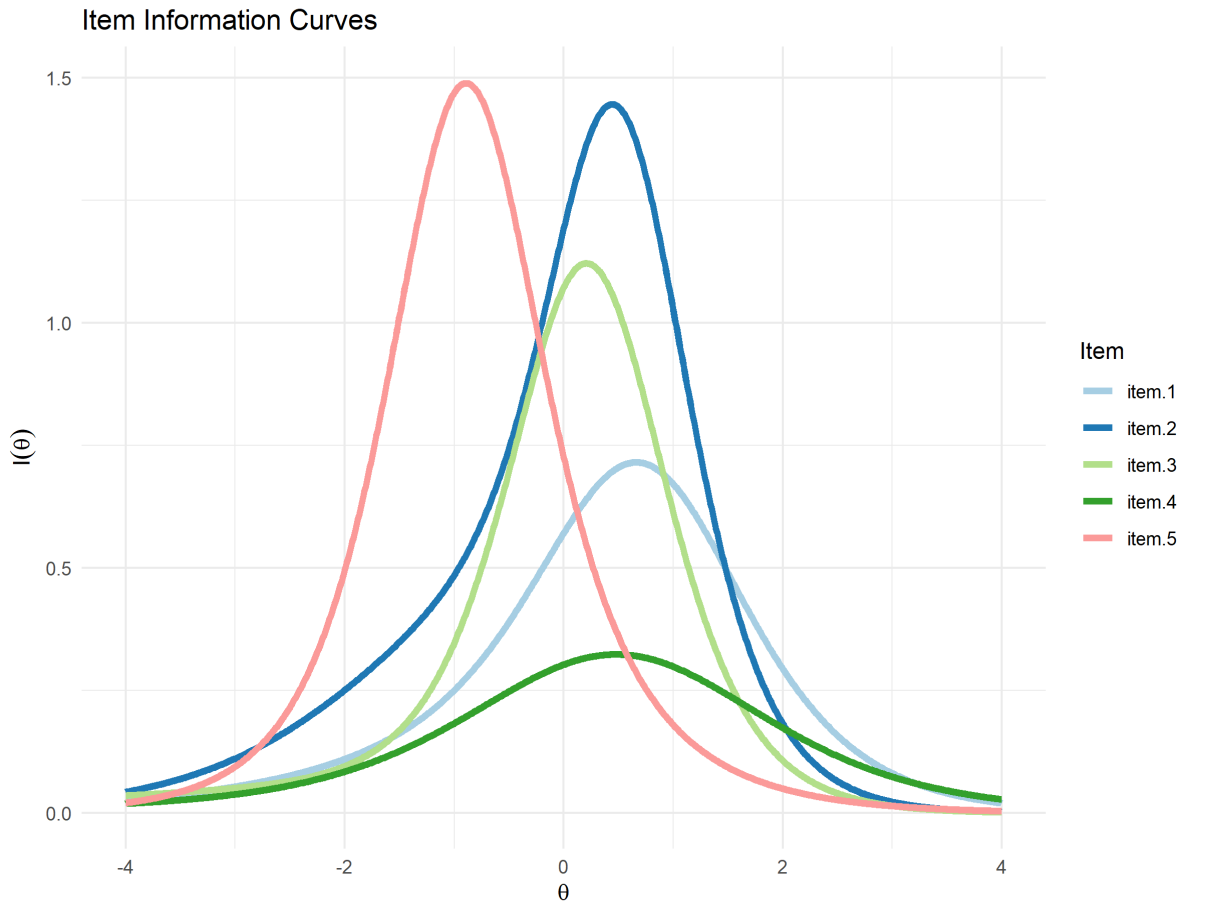
stavka, vredelo bi razmisliti o sažimanju kategorija odgovora 1 i 2. Ukoliko se opredelimo za sažimanje kategorija, ove kategorije bi trebalo rekodirati u jednu. Kategorije 1 i 2 bi trebalo sažeti u kategoriju 1, a kategorije 3 i 4 pretvoriti u 2 i 3 (respektivno). Podsetićemo, nominalni model ne zahteva jednak broj kategorija odgovora na svim stavkama. Nakon toga bi trebalo proceniti model sa tako rekodiranom stavkom i uporediti pokazatelje fita (AIC i BIC) sa prethodnim modelom. Ukoliko bi oni ukazivali na poboljšanje modela (snižavanje vrednosti AIC i BIC), takav model bismo mogli zadržati. Važno je napomenuti da bi sažimanje kategorija dovelo do smanjivanja informativnosti stavke i posledično smanjivanja informativnosti testa (De Ayala, 2022).

Kada pogledamo grafikon sa empirijskim karakterističnim krivama kategorija (Slika 118), vidimo da model uglavnom dobro predviđa kategorije 0, 1 i 3 (*P1*, *P2* i *P4*). Verovatnoću odgovora u kategoriji 2 (*P3*) neznatno precenjuje na nivoima osobine ispod prosečnog, a uglavnom potcenjuje na nivoima osobine iznad proseka.

#### 10.5.7.2.2 Informativnost stavki i testa

Prikazaćemo grafikon informativnosti stavki koristeći funkciju *itemInfoPlot()* iz paketa *ggmirt* koju ćemo direktno pozvati.

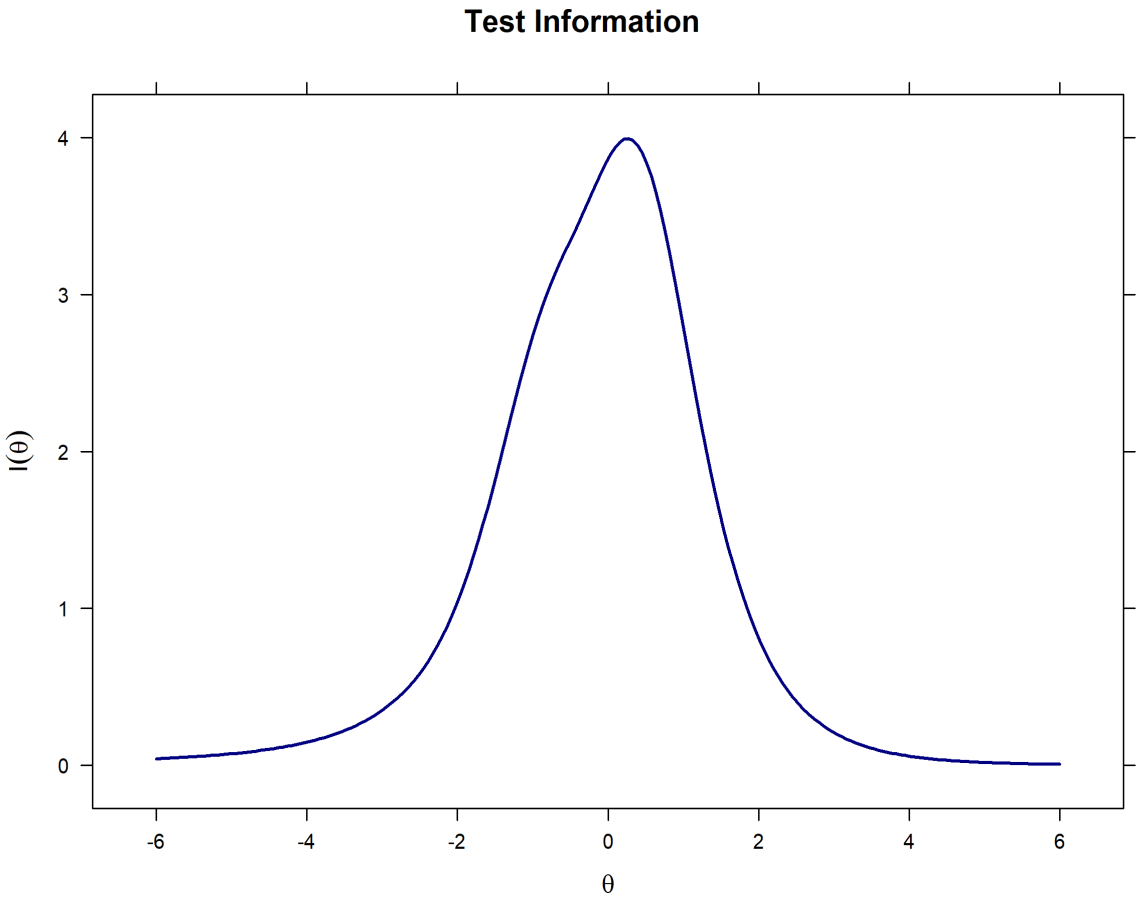
```
ggmirt::itemInfoPlot(nom.m, legend = T) + scale_color_brewer(palette = "Paired") +
  ggplot2::geom_line(size = 1.5)
```




Slika 119 – Funkcije informativnosti stavki u nominalnom modelu

Vidimo (Slika 119) da sve stavke osim stavke 5 maksimum informativnosti dostižu na nivou latentne osobine nešto iznad 0 logita. Najvišu informativnost imaju stavke 5 i 2, zatim 3 i 1, dok je najmanje informativna stavka 4. Prema lokacijama funkcija informativnosti možemo pretpostaviti da će test biti najinformativniji na nivou osobine nešto višem od prosečnog. Prikazaćemo krivu informativnosti testa (Slika 120).

```
plot(nom.m, type="info",par.settings = simpleTheme(lwd=2, col="navy"))
```




Slika 120 – Funkcija informativnosti testa

Grafikon (Slika 120) nam to potvrđuje, ali ćemo izračunati i numeričke pokazatelje.

Izračunaćemo prvo informativnost u opsegu između –3 i 3 logita.

```
# Za ceo test
mNOMpi <- areainfo(nom.m, c(-3,3))
mNOMpi

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 11.56707  12.27441  0.9423722      5

# Samo za određene stavke
areainfo(nom.m, c(-3,3), which.items = c(1,2,4:5)) # Bez stavke 3

##  LowerBound UpperBound     Info TotalInfo Proportion nitems
##          -3          3 9.295441  9.840104  0.9446487      4
```

95% informativnosti ovog testa koncentrisano je u zadatom opsegu osobine. Ako je to opseg osobine u kojem želimo da obavljamo merenje, možemo biti zadovoljni.

```
thetaM <- seq(-3,3, .1)
# Izračunamo informativnost za svaki od traženih nivoa
infoM <- testinfo(nom.m, Theta = thetaM)
# Prikaz prvih 6 redova (ukolonite head() za ispis cele tabele)
TI <- data.frame(cbind("nivo osobine"=thetaM, "informativnost"=infoM))
head(TI)

##   nivo.osobine informativnost
## 1         -3.0      0.3501953
## 2         -2.9      0.3863011
## 3         -2.8      0.4271699
## 4         -2.7      0.4735363
## 5         -2.6      0.5262475
## 6         -2.5      0.5862701

# Maksimalna informativnost
cat("Maksimalna informativnost:", max(infoM), "\n")

## Maksimalna informativnost: 3.991351

# Ispis nivoa osobine na kom je informativnost maksimalna
cat("Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
\n", thetaM[infoM==max(infoM)], "\n")

## Lokacija na kontinuumu latentne osobine na kojoj je test najinformativniji:
##  0.3

# Ispis raspona u kojem je informativnost veća od 3,33 što odgovara pouzdanosti od
0,7
cat("Raspon latentne osobine u kojem je test zadovoljavajuće informativan: \n",
 min(thetaM[infoM>3.33]), "do", max(thetaM[infoM>3.33]), "logita")

## Raspon latentne osobine u kojem je test zadovoljavajuće informativan:
##  -0.5 do 0.7 logita

cat("\nInformativnost u rasponu od -3 do 3 logita: \n")

##
## Informativnost u rasponu od -3 do 3 logita:

inf33 <- sum(TINFO)
inf33

## [1] 197.1287

# Ta vrednost sama po sebi nije puno korisna ali može služiti za poređenje
cat("\nInformativnost u rasponu od 0 do 3 logita:  \n")

##
## Informativnost u rasponu od 0 do 3 logita:

inf03 <- sum(TINFO[!thetaM < 0], na.rm = TRUE)
inf03

## [1] 92.42318

cat(" \nProcenat informativnosti u rasponu od 0 do 3 logita, u odnosu na raspon od
-3 do 3 logita:  \n")
```

```
##
## Procenat informativnosti u rasponu od 0 do 3 logita, u odnosu na raspon od -3
do 3 logita:

round(inf03/inf33*100,2)

## [1] 46.88
```

Test je najinformativniji na nivou osobine od 0,3 logita. Zadovoljavajuća informativnost (>3,33 što odgovara pouzdanosti >0,7) je u rasponu od –0,5 do 0,7 logita. To je vrlo uzan opseg osobine i trebalo bi dodati informativne stavke koje najviše informacije nose izvan tog opsega.

#### *10.5.7.2.2.1 Empirijska i marginalna pouzdanost*

I za nominalni model može se izračunati empirijska i marginalna pouzdanost.

```
empirical_rxx(fscores(nom.m, full.scores.SE = TRUE))

##        F1
## 0.7349187

marginal_rxx(nom.m)

## [1] 0.7255489
```

Empirijska i marginalna pouzdanost testa su zadovoljavajuće, odnosno prelaze vrednost 0,7. Međutim, ako se prisetimo da je test zadovoljavajuće pouzdan samo u rasponu od –0,5 do 0,7 logita, to baca drugo svetlo na ovaj podatak.

## *10.6 Ispitivanje diferencijanog funkcionisanja stavki i testa*

Prvo ćemo simulirati podatke za prikaz postupka ispitivanja DIF-a. Biće simulirani binarno skorovani podaci na isti način kao što smo simulirali podatke za IRT modele za binarno skorovane stavke. Jedina razlika je što će biti kreirani podaci za dve grupe, s tim što će za jednu od njih biti izmenjeni parametri nagiba i lokacije za jednu stavku. Ovde će biti prikazane procedure za ispitivanje DIF-a kod binarno skorovanih stavki, ali procedura je slična i za politomne stavke.

```
ss <- 3
set.seed(ss)
nagibi <- runif(10, 1, 2.2)
# Nagibi/diskriminativnosti
nagibi # Prikaz nagiba za prvu grupu

##  [1] 1.201650 1.969020 1.461931 1.393281 1.722521 1.725273 1.149560 1.353521
##  [9] 1.693132 1.757175

# Kopiraćemo nagibe
nagibi2 <- nagibi
# Postavljamo drugačiji parametar nagiba za ajtem 2 kako bismo dobili DIF
nagibi2[2] <- 1.5

# Lokacije/težine stavki
lokacije <- runif(10, -3, 3)
```

```
# Ispis lokacija stavki za prvu grupu
lokacije

##  [1]  0.07209539  0.03014348  0.20421212  0.34349661  2.20751693  1.97825216
##  [7] -2.33130508  1.22213015  2.38492959 -1.32160468

# Kopiranje lokacija za drugu grupu
lokacije2 <- lokacije
# Izmena parametra težine za drugu stavku
lokacije2[2] <- .7

pogadjanje <- runif(10, 0, 0)


set.seed(ss)
# Simulacija nivoa osobine u uzorku (ovde su isti za obe grupe, ali nije nužno da
budu isti)
thete <- rnorm(500, mean=0)
# Nivoi osobine ispitanika simulirani iz normalne distribucije sa AS=0 i SD=1

set.seed(ss)
# Prvi deo uzorka
b.pod1 <- data.frame(mirt::simdata(a=nagibi, d=lokacije, u=1, N=500,
 itemtype = "dich"))
# Drugi deo uzorka sa izmenjenim parametrima za stavku 2
b.pod2 <- data.frame(mirt::simdata(a=nagibi2, d=lokacije2, u=1, N=500,
 itemtype = "dich"))
# Argument itemtype="dich" definiše da želimo da stavke budu dihotomne (0,1)
# Spajanje dva poduzorka
b.podD <- rbind(b.pod1, b.pod2)
# Imenovanje varijabli (stavki)
names(b.podD) <- paste0("it",1:10)
head(b.podD,3)

##   it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## 1   0   0   0   0   1   1   0   1   1    0
## 2   0   1   0   1   1   1   0   1   1    1
## 3   0   0   0   1   1   1   0   1   0    0

tail(b.podD,3)

##      it1 it2 it3 it4 it5 it6 it7 it8 it9 it10
## 998    0   0   0   0   0   0   0   0   0    0
## 999    1   1   1   1   1   1   0   1   1    1
## 1000   1   1   1   1   1   1   0   1   1    1

# Kreiranje grupišuće varijable (prvih 500 referentna grupa, drugih 500 fokalna)
grupa <- factor(c(rep(0, 500), rep(1, 500)) , levels = 0:1, labels = c("R", "F"))
```

### 10.6.1 Mantel–Hanselov postupak

Prvo ćemo prikazati Mantel–Hanselov postupak. Biće korišćen paket *difR* (Magis i ostali, 2010) i funkcija *difMH()* namenjena ispitivanju DIF-a kod *binarno skorovanih stavki*.

Komandi *difMH()* prosledićemo argumente *group=grupa* koji definiše grupišuću varijablu, *focal.name = "F"* koji definiše oznaku fokalne grupe, *match = "score"* koji definiše meru nivoa osobine po kojoj bi trebalo da budu ujednačene grupe, *purify = TRUE* da bi stavke koje pokažu DIF trebalo da budu isključene

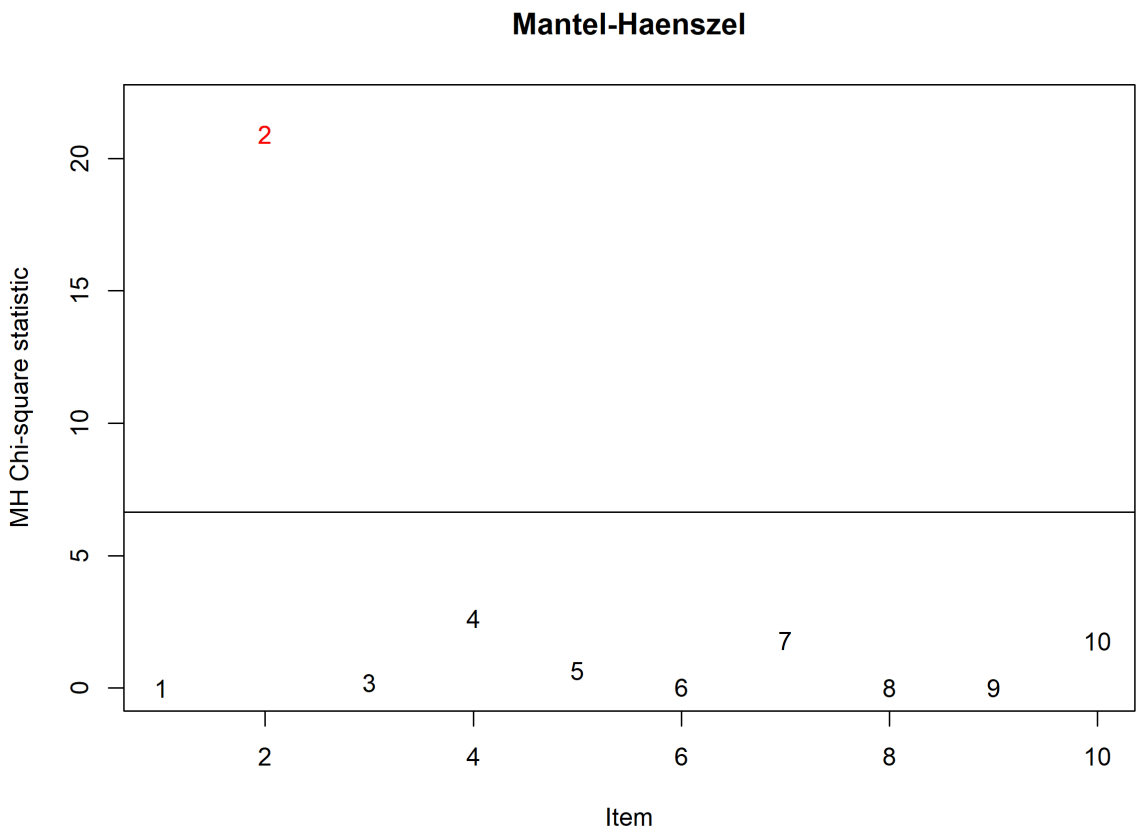
iz računanja skora po kojem se grupe ujednačavaju, *alpha = .01* koji definiše nivo značajnosti i *p.adjust.method = "BH"* kojim smo definisali da će se koristiti Benjamini–Hohbergova (Benjamini–Hochberg) korekcija *p* vrednosti za višestruka poređenja.

```
library("pacman")
p_load(difR)
MHD <- difMH(b.podD, group=grupa, focal.name = "F", match = "score",
 purify = TRUE, alpha = .01, p.adjust.method = "BH")
MHD

##
## Detection of Differential Item Functioning using Mantel-Haenszel method
## with continuity correction and with item purification
##
## Results based on asymptotic inference
##
## Convergence reached after 1 iteration
##
## Matching variable: test score
##
## No set of anchor items was provided
##
## Multiple comparisons made with Benjamini-Hochberg adjustement of p-values
##
## Mantel-Haenszel Chi-square statistic:
##
##      Stat.   P-value Adj. P
## it1   0.0001  0.9938  0.9938
## it2  20.9122  0.0000  0.0000 ***
## it3   0.2035  0.6519  0.9938
## it4   2.6237  0.1053  0.4548
## it5   0.6593  0.4168  0.8336
## it6   0.0310  0.8602  0.9938
## it7   1.7950  0.1803  0.4548
## it8   0.0062  0.9375  0.9938
## it9   0.0080  0.9287  0.9938
## it10  1.7819  0.1819  0.4548
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 6.6349 (significance level: 0.01)
##
## Items detected as DIF items:
##
##  it2
##
##
## Effect size (ETS Delta scale):
##
## Effect size code:
##  'A': negligible effect
##  'B': moderate effect
##  'C': large effect
##
##      alphaMH deltaMH
## it1   1.0110 -0.0258 A
## it2   0.4566  1.8422 C
## it3   0.9195  0.1972 A
```

```
## it4   1.3135 -0.6409 A
## it5   1.2185 -0.4644 A
## it6   1.0564 -0.1289 A
## it7   0.7213  0.7677 A
## it8   1.0011 -0.0026 A
## it9   1.0042 -0.0098 A
## it10  0.7457  0.6897 A
##
## Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
##  (for absolute values of 'deltaMH')
##
## Output was not captured!
```

Otkrivena je jedna stavka kod koje postoji DIF, i to je stavka 2. DIF je značajan na nivou $p$<0,01. Prema klasifikaciji veličine efekta DIF je veliki (ima oznaku C u poslednjoj tabeli). S obzirom na to da je vrednost $\alpha_{MH}$ manja od 1 a vrednost $\Delta_{MH}$ pozitivna, možemo zaključiti da na stavci 2 bolje rezultate postižu članovi fokalne grupe (videti odeljak 8.1). Na osnovu MH statistika ne znamo da li je DIF uniformni ili neuniformni, a ovom pokazatelju bi mogao promaći neuniformni DIF (isti odeljak).

```
plot(MHD)
```




Slika 121 – Grafikon Mantel–Hanselovog statistika

Grafikon (Slika 121) prikazuje stavke sa DIF-a (odnosno Mantel–Hanselov statistik). Horizontalna linija predstavlja graničnu vrednost MH statistika. Vrednosti iznad linije su redni brojevi stavki koje imaju značajan DIF.

Primetite da se u računu nigde ne pojavljuje IRT model niti parametri stavki.

### 10.6.2 Ispitivanje DIF-a upotrebom logističke regresione analize

Sledeći postupak koji ćemo prikazati je upotreba logističke regresione analize. I dalje ćemo koristiti paket *difR* i funkciju *diflogistic()*. Pošto se radi o istom paketu u kojem smo računali Mantel–Hanselov statistik, argumenti su uglavnom isti. Novina je argument *type="both"* kojim smo rekli da želimo da proverimo postojanje i uniformnog i neuniformnog DIF-a. Ako želimo proveru ili jednog ili drugog, argumente možemo dodeliti vrednost *"udif"* za uniformni ili *"nudif"* za neuniformni.

Zahtevali smo da *sidra* (videti odeljak 8.3) budu određena *prečišćavanjem* (*purify=TRUE*), ali ona mogu biti unapred zadata argumentom *anchors* kojem se mora dodeliti vektor imena sidrenih stavki (ili vektor rednih brojeva).

```
L_D <- difLogistic(b.podD, group=grupa, focal.name="F", p.adjust.method = "BH",
 type="both", purify = TRUE)
print.Logistic(L_D)

##
## Detection of both types of Differential Item Functioning
## using Logistic regression method, with item purification
## and with LRT DIF statistic
##
## Convergence reached after 1 iteration
##
## Matching variable: test score
##
## No set of anchor items was provided
##
## Multiple comparisons made with Benjamini-Hochberg adjustement of p-values
##
## Logistic regression DIF statistic:
##
##      Stat.   P-value Adj. P
## it1   1.1121  0.5735  0.9284
## it2  29.6695  0.0000  0.0000 ***
## it3   0.3325  0.8468  0.9284
## it4   3.6215  0.1635  0.4088
## it5   0.7126  0.7003  0.9284
## it6   0.1485  0.9284  0.9284
## it7   3.8510  0.1458  0.4088
## it8   0.7817  0.6765  0.9284
## it9   0.1870  0.9107  0.9284
## it10  4.5436  0.1031  0.4088
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 5.9915 (significance level: 0.05)
##
## Items detected as DIF items:
##
##  it2
##
##
## Effect size (Nagelkerke's R^2):
```

```
##
## Effect size code:
##  'A': negligible effect
##  'B': moderate effect
##  'C': large effect
##
##      R^2    ZT JG
## it1  0.0010 A  A
## it2  0.0252 A  A
## it3  0.0003 A  A
## it4  0.0032 A  A
## it5  0.0008 A  A
## it6  0.0002 A  A
## it7  0.0056 A  A
## it8  0.0008 A  A
## it9  0.0002 A  A
## it10 0.0036 A  A
##
## Effect size codes:
##  Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
##  Jodoin & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
##
##  Output was not captured!
```

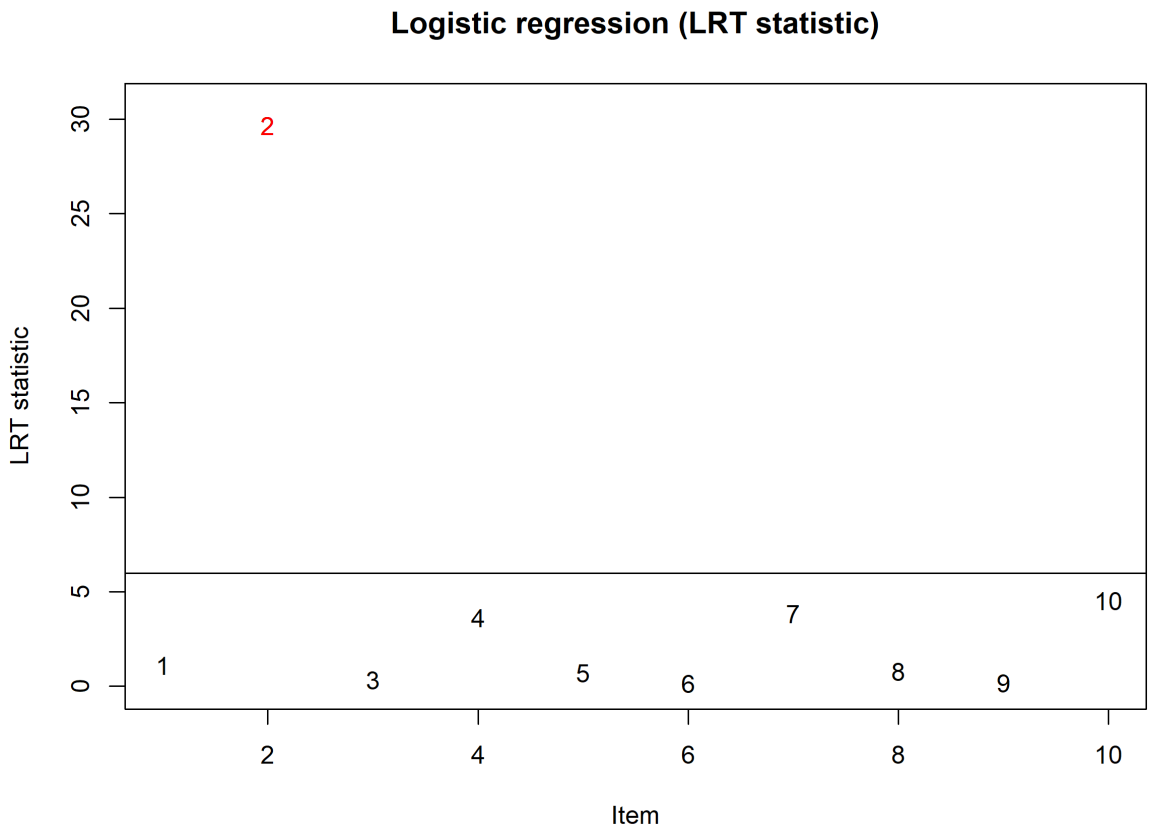
Procedura namenjena ispitivanju uniformnog i neunifromnog DIF-a detektuje stavku 2 kao jedinu koja pokazuje DIF. Veličina efekta procenjena na osnovu promene u Nagelkirkovom pseudo $R^2$ je zanemarljiva (oznaka A) bez obzira koju klasifikaciju koristimo, onu Zumba i Tomasa (Thomas) ili onu Žodona i Girla (Jodoin & Gierl, 2001). Podsetićemo se kriterijuma da hi–kvadrat za model koji uključuje grupnu pripadnost i interakciju grupne pripadnosti i nivoa osobine (gornji) mora biti značajan i da veličina efekta mora biti srednja da bismo rekli da stavka diferencijalno funkcioniše (Gelin & Zumbo, 2003). Stavka 2 zadovoljava samo prvi od ta dva kriterijuma, pa po tom kriterijumu kod nje ne postoji značajan DIF.

Sa *L_D$logitPar* zatražili smo koeficijente logističke regresije. Zanimaju nas samo kolone *GROUP* i *SCORE:GROUP* (interakcija nivoa osobine i grupe). Koeficijenti se interpretiraju u odnosu na referentnu grupu. Kako je koeficijent za grupu na stavci 2 pozitivan, to znači da pripadnici fokalne grupe postižu više skorove na ovoj stavci od ispitanika iz referentne grupe sa jednakim nivoom osobine. Značajan koeficijent za efekat interakcije ukazuje na neuniformni DIF.

```
L_D$logitPar
```

```
##       (Intercept)     SCORE    GROUP SCORE:GROUP
##  [1,]   -3.735266 0.7320156 0.000000   0.0000000
##  [2,]   -5.062990 0.8844363 2.199933  -0.2518068
##  [3,]   -3.731294 0.7500723 0.000000   0.0000000
##  [4,]   -3.703715 0.7742417 0.000000   0.0000000
##  [5,]   -2.076777 0.8999829 0.000000   0.0000000
##  [6,]   -2.405910 0.8626559 0.000000   0.0000000
##  [7,]   -7.214203 0.8530248 0.000000   0.0000000
##  [8,]   -2.753027 0.7607380 0.000000   0.0000000
##  [9,]   -2.181693 0.9312940 0.000000   0.0000000
## [10,]   -9.572695 1.4182664 0.000000   0.0000000
```

```
plot(L_D, plot="lrStat", itemFit = "best",
 pch = 8, number = T, col = "red", colIC = rep("black", 2), ltyIC = c(1, 2))


# Na ordinati su prikazane vrednosti LR statistika
# Granična vrednost za značajan DIF prikazana je horizontalnom linijom
```




Slika 122 – Grafikon LR testa za logističku regresiju

Grafikon (Slika 122), koji je u suštini isti kao i onaj za Mantel–Hanselov statistik, pokazuje da je jedina stavka koja pokazuje DIF stavka 2. Na ordinati su predstavljene vrednosti LR testa ($G^2$), a horizontalna linija označava graničnu vrednost. Brojevi (a i apscisa) označavaju redne brojeve stavki.

Ponovićemo proceduru posebno prvo za neuniformni pa za uniformni DIF.

```
# Neuniformni DIF
difLogistic(b.podD, group=grupa, focal.name="F", p.adjust.method = "BH",
 type="nudif", purify = TRUE)

##
## Detection of nonuniform Differential Item Functioning
## using Logistic regression method, with item purification
## and with LRT DIF statistic
##
## Convergence reached after 1 iteration
##
## Matching variable: test score
##
## No set of anchor items was provided
```

```
##
## Multiple comparisons made with Benjamini-Hochberg adjustement of p-values
##
## Logistic regression DIF statistic:
##
##      Stat.  P-value Adj. P
## it1  1.1015 0.2939  0.6517
## it2  7.1112 0.0077  0.0766 .
## it3  0.0463 0.8297  0.8747
## it4  0.7357 0.3910  0.6517
## it5  0.0249 0.8747  0.8747
## it6  0.0910 0.7629  0.8747
## it7  3.4427 0.0635  0.3177
## it8  0.7577 0.3840  0.6517
## it9  0.1852 0.6670  0.8747
## it10 2.6262 0.1051  0.3504
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 3.8415 (significance level: 0.05)
##
## Items detected as DIF items: No DIF item detected
##
## Effect size (Nagelkerke's R^2):
##
## Effect size code:
##  'A': negligible effect
##  'B': moderate effect
##  'C': large effect
##
##      R^2    ZT JG
## it1  0.0010 A  A
## it2  0.0060 A  A
## it3  0.0000 A  A
## it4  0.0006 A  A
## it5  0.0000 A  A
## it6  0.0001 A  A
## it7  0.0050 A  A
## it8  0.0007 A  A
## it9  0.0002 A  A
## it10 0.0021 A  A
##
## Effect size codes:
##  Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
##  Jodoin & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
##
##  Output was not captured!
```

Neuniformni DIF nije značajan mada je *p* vrednost za stavku 2 bliska kritičnoj (0,77).

```
# Uniformni DIF
difLogistic(b.podD, group=grupa, focal.name="F", p.adjust.method = "BH",
 type="udif", purify = TRUE) # "nudif" or "udif"

##
## Detection of uniform Differential Item Functioning
## using Logistic regression method, with item purification
## and with LRT DIF statistic
##
```

```
## Convergence reached after 2 iterations
##
## Matching variable: test score
##
## No set of anchor items was provided
##
## Multiple comparisons made with Benjamini-Hochberg adjustement of p-values
##
## Logistic regression DIF statistic:
##
##      Stat.   P-value Adj. P
## it1   0.0106  0.9180  0.9659
## it2  22.5583  0.0000  0.0000 ***
## it3   0.2862  0.5926  0.9659
## it4   2.8858  0.0894  0.4468
## it5   0.6877  0.4070  0.9659
## it6   0.0575  0.8105  0.9659
## it7   0.4084  0.5228  0.9659
## it8   0.0240  0.8769  0.9659
## it9   0.0018  0.9659  0.9659
## it10  1.9174  0.1661  0.5538
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 3.8415 (significance level: 0.05)
##
## Items detected as uniform DIF items:
##
##  it2
##
##
## Effect size (Nagelkerke's R^2):
##
## Effect size code:
##  'A': negligible effect
##  'B': moderate effect
##  'C': large effect
##
##      R^2    ZT JG
## it1  0.0000 A  A
## it2  0.0192 A  A
## it3  0.0003 A  A
## it4  0.0025 A  A
## it5  0.0008 A  A
## it6  0.0001 A  A
## it7  0.0006 A  A
## it8  0.0000 A  A
## it9  0.0000 A  A
## it10 0.0015 A  A
##
## Effect size codes:
##  Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
##  Jodoin & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
##
##  Output was not captured!
```

Uniformni DIF za stavku 2 je značajan, ali je veličina efekta izražena u $\Delta R^2$ (Nagelkirk) zanemarljiva, bilo da koristimo klasifikaciju Zumba i Tomasa ili onu Žodona i Girla (Jodoin & Gierl, 2001).

Prikazaćemo i proceduru iz paketa *lordif* (S. W. Choi i ostali, 2011) koji primenjuje hibrid ordinalne logističke regresione analize i IRT modelovanja. Namenjen je ispitivanju DIF-a kod binarno skorovanih i ordinalnih politomnih stavki. Kao meru osobine koristi IRT parametre ispitanika, a za njihovo izračunavanje oslanja se na paket *ltm* (Rizopoulos, 2006).

Funkciji *lordif()* prosledili smo sledeće argumente: *criterion = c("R2")* što znači da će se značajnost DIF-a utvrđivati i na osnovu $\Delta R^2$ između punog i redukovanih modela (videti odeljak 8.3), *pseudo.R2 = "Nagelkerke"* – za poređenje će se koristiti Negelkerkeov pseudo $R^2$, *R2.change = 0.035* – kao značajna promena $R^2$ smatraće se promena od 0,035 (videti isti odeljak), *model = "GRM"* – za uparivanje grupa koristiće se IRT mere dobijene modelom stepenovanog odgovaranja, *MonteCarlo = TRUE* – značajnost DIF-a će se određivati na osnovu Monte Karlo simulacija, *nr = 1000* – biće primenjeno 1000 replikacija. Opredelili smo se da sidra i ovde odredimo prečišćavanjem, mada funkcija dozvoljava da sidra definišemo unapred. Naime, ako argumentu *anchor* ne dodelimo stavke koje želimo da koristimo kao sidra, automatski se primenjuje prečišćavanje. Sidra dodeljujemo kao vektor imena stavki ili njihovih rednih brojeva.

```
p_load(lordif)
L_DIF <- lordif(b.podD, group=grupa, criterion = c("R2"),
 pseudo.R2 = "Nagelkerke", model = "GRM", MonteCarlo = TRUE, nr = 1000)

##  (mirt) | Iteration: 2, 1 items flagged for DIF (2)
## Monte Carlo simulation
## Start time: Mon May 20 16:28:08 2024
##
##
## 1 items flagged for DIF by Monte Carlo thresholds (2)
```

I ovde je kao jedina stavka koja ispoljava diferencijalno funkcionisanje u referentnoj i fokalnoj grupi izdvojena stavka 2. Granična vrednost $\Delta R^2$ određena je na osnovu Monte Karlo simulacija. Analiza je sačuvana u objektu *L_DIF,* a sa *L_DIF$ipar.sparse* ispisali smo kalibracije stavki iz GRM modela.

```
L_DIF$ipar.sparse

##             a         cb1
## I1   1.213135 -0.09857028
## I3   1.263900 -0.16285445
## I4   1.345676 -0.25397066
## I5   2.037193 -1.24498023
## I6   1.674160 -1.12621029
## I7   1.083010  2.01122475
## I8   1.419194 -0.78512000
## I9   1.882819 -1.28723338
## I10  2.374718  0.66430328
## I2.1 1.997670 -0.05614979
## I2.2 1.270051 -0.60485341
```

Vidimo da se poslednja dva reda tabele (I2.1 i I2.2) odnose na stavku 2. U pretposlednjem redu su parametri ove stavke u referentnoj grupi, a u poslednjem u fokalnoj. Vidimo da je stavka nešto lakša u fokalnoj grupi ($b_{2F}$ = –0,60, $b_{2R}$ = –0,06), dok je nagib veći u referentnoj nego u fokalnoj ($a_{2R}$ = 2, $a_{2F}$ = 1,27).

Ispisaćemo odabrane rezultate koji se nalaze u objektu *L_DIF$stats*. Ovaj objekat sadrži LR testove

između tri modela, promene u pseudo $R^2$ pokazateljima (Nagelkerke, Cox & Snell i McFaden), proporcija promene $\beta_1$ koeficijenta između modela 1 i 2, te stepene slobode za LR testove. Prvo ćemo zatražiti značajnosti LR testova između tri modela i proporciju promene $\beta_1$ koeficijenta.

```
L_DIF$stats[, c("item", "chi12", "chi13", "chi23", "beta12")]
```

```
##    item  chi12  chi13  chi23 beta12
## 1     1 0.8622 0.8487 0.5852 0.0003
## 2     2 0.0000 0.0000 0.0000 0.0437
## 3     3 0.6586 0.6350 0.3984 0.0009
## 4     4 0.0845 0.1584 0.4000 0.0008
## 5     5 0.3027 0.5812 0.8797 0.0004
## 6     6 0.6810 0.7371 0.5066 0.0004
## 7     7 0.6027 0.3302 0.1631 0.0026
## 8     8 0.9814 0.8853 0.6219 0.0000
## 9     9 0.8851 0.7054 0.4106 0.0002
## 10   10 0.0796 0.1039 0.2276 0.0079
```

Vidimo da samo kod stavke 2 postoje značajni LR testovi. Značajna su sva 3 što znači da postoji i uniformni i neuniformni DIF. Značajan LR test u prvoj koloni (*chi12*) ukazuje na uniformni DIF, u trećoj koloni (*chi23*) na neuniformni, a u drugoj koloni (*chi13*) je omnibus pokazatelj DIF-a (i uniformnog i neuniformnog). Kada je u pitanju promena $\beta_1$ koeficijenta, promena veća od 5% (0,05) se smatra pokazateljem postojanja uniformnog DIF-a. Vidimo da se stavka 2 približava toj granici (0,0437).

Zatražićemo vrednosti $\Delta R^2$ (Nagelkirk):

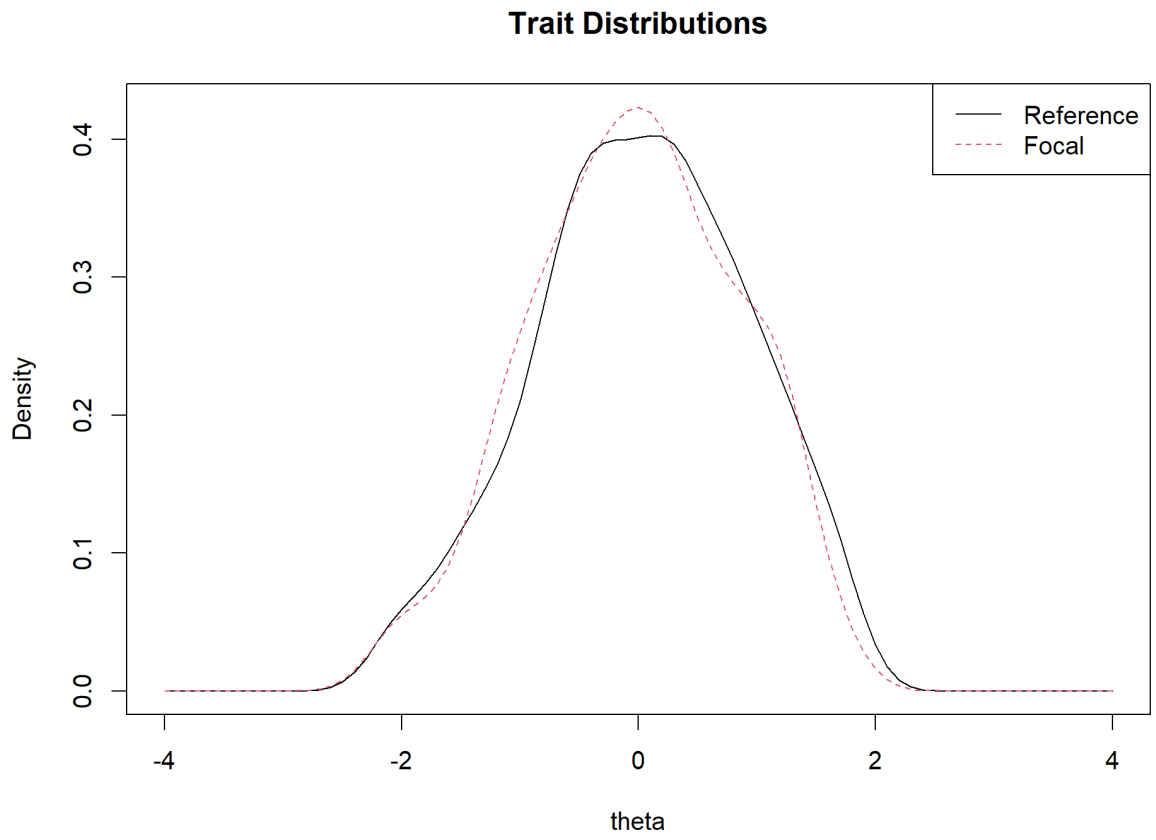
```
L_DIF$stats[, c("item", "pseudo12.Nagelkerke", "pseudo13.Nagelkerke",
 "pseudo23.Nagelkerke")]
```

```
##    item pseudo12.Nagelkerke pseudo13.Nagelkerke pseudo23.Nagelkerke
## 1     1              0.0000              0.0003              0.0003
## 2     2              0.0240              0.0384              0.0145
## 3     3              0.0002              0.0009              0.0007
## 4     4              0.0027              0.0034              0.0006
## 5     5              0.0012              0.0012              0.0000
## 6     6              0.0002              0.0006              0.0005
## 7     7              0.0004              0.0034              0.0030
## 8     8              0.0000              0.0002              0.0002
## 9     9              0.0000              0.0008              0.0008
## 10   10              0.0021              0.0032              0.0010
```

Vidimo da samo omnibus test za stavku 2 ima srednju veličinu efekta (>0,035) po podeli Žodona i Girla (Jodoin & Gierl, 2001), ali ova stavka ispunjava oba kriterijuma da bi se klasifikovala kao stavka koja ispoljava DIF.

```
plot(L_DIF)
```

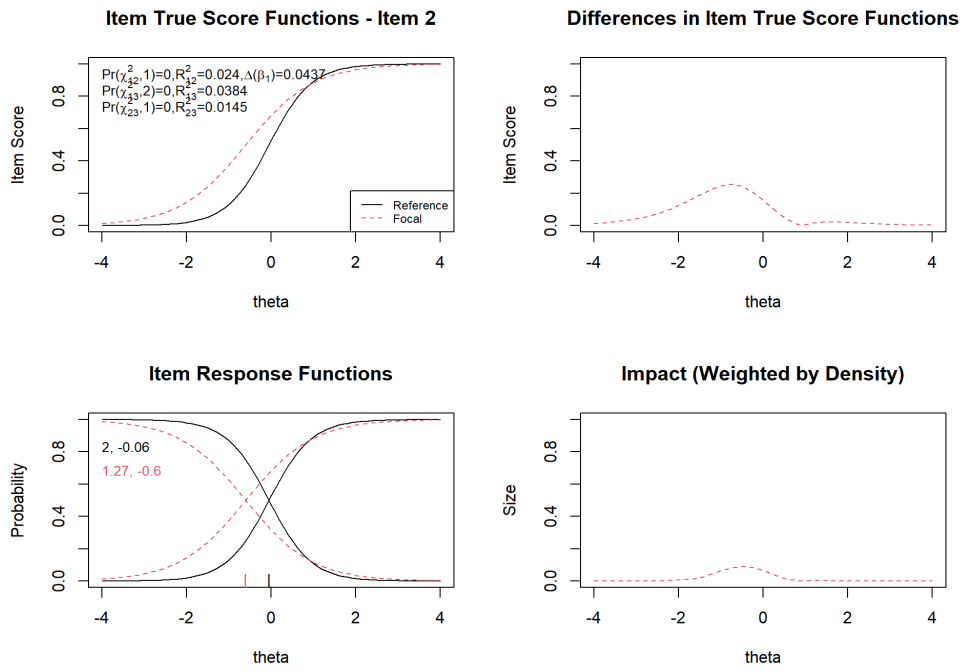
Komandom *plot(L_DIF)* dobili smo dijagnostičke grafikone. Na prvom (Slika 123) vidimo distribucije procenjenih nivoa osobine koje su u obe grupe vrlo slične, sa nešto više ispitanika sa nivoom osobine oko prosečnog nivoa u fokalnoj grupi.




Slika 123 – Distribucija latentne osobine u referentnoj i fokalnoj grupi

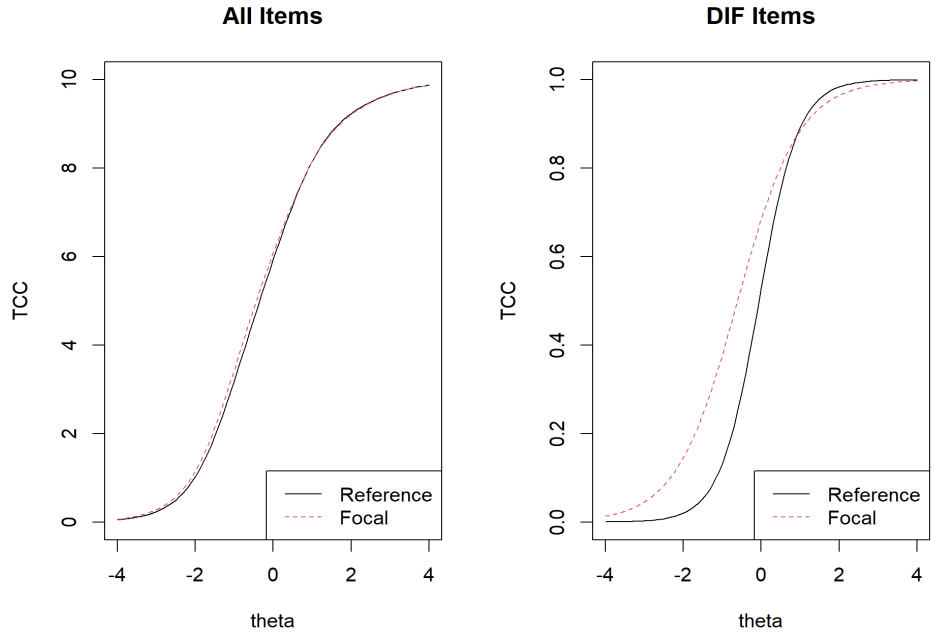
Sledeći grafikon sastoji se od četiri panela (Slika 124). Na panelima sa leve strane vidimo karakteristične krive (gore) i karakteristične krive kategorija (dole) stavke 2 koja pokazuje DIF. Crnom linijom označena je referentna, a crvenom fokalna grupa. Oba grafikona nam pokazuju isto – manji nagib u fokalnoj grupi i nešto niži parametar lokacije. Na grafikonu sa karakterističnim krivama kategorija (Slika 124 dole levo) navedeni su parametri nagiba i lokacije stavke 2 za dve grupe. Na grafikonu sa karakterističnom krivom stavke (Slika 124 gore levo) ispisani su rezultati LR testova za uniformni, neuniformni i sveukupni DIF. Test ukupnog DIF-a (indeks 13) je statistički značajan. Isto važi i za test uniformnog DIF-a (indeks 12) i test neuniformnog DIF-a (indeks 23). Dakle, za stavku 2 svi LR testovi su statistički značajni. Ako posmatramo vrednosti $\Delta R^2$, one su uglavnom male osim za sveukupni DIF koji ima vrednost $\Delta R^2_{13} = 0{,}0384$, što ga čini efektom srednje veličine. $\Delta(\beta_1)_{13} = 0{,}0437$ sugeriše da uniformni DIF za stavku 2 nije značajan.

Na panelima sa desne strane (Slika 124) prikazan je uticaj DIF-a na skorove u fokalnoj grupi. Na gornjem panelu vidimo da je uticaj DIF-a najveći na lokaciji nešto nižoj od prosečnog nivoa osobine i da maksimalno iznosi nešto preko 0,20 (20%). Na tom nivou ispitanici iz fokalne grupe imaju za 20% veću verovatnoću tačnog odgovora na stavku. Na donjem panelu (Slika 124) je prikazano isto, ali ponderisano zastupljenošću tog nivoa osobine u fokalnoj grupi. Vidimo da je uticaj mali.




Slika 124 – Funkcije pravog skora misfitujuće stavke; Razlike u pravom skoru stavke sa DIF-om; Karakteristične krive stavke sa DIF-om u dve grupe; Uticaj DIF-a

Na sledećem grafikonu (Slika 125) na levom panelu prikazana je karakteristična kriva testa sa svim stavkama uključujući i one sa DIF (na ordinati je prikazan očekivani skor). Vidimo da je uticaj stavke 2, koja pokazuje DIF, vrlo mali. Dve krive se skoro sasvim preklapaju. Na desnom panelu prikazana je ista ta kriva, ali na podskupu stavki koje pokazuju DIF. U našem slučaju to je samo stavka 2, pa je ovo opet karakteristična kriva stavke 2.




Slika 125 – Karakteristična kriva testa i karakteristična kriva stavki sa DIF-om

U nastavku je prikazana ista procedura kada se kao kriterijum za otkrivanje DIF-a koristi hi–kvadrat test, a ne koriste Monte Karlo simulacije.

```
L_DIF.chi <- lordif(b.podD, group=grupa, criterion = c("Chisqr"), model = "GRM")
```

```
##  (mirt) | Iteration: 2, 1 items flagged for DIF (2)
```

```
L_DIF.chi
```

```
## Call:
## lordif(resp.data = b.podD, group = grupa, criterion = c("Chisqr"),
##     model = "GRM")
##
##   Number of DIF groups: 2
##
##   Number of items flagged for DIF: 1 of 10
##
##   Items flagged: 2
##
##   Number of iterations for purification: 2 of 10
##
##   Detection criterion: Chisqr
##
##   Threshold: alpha = 0.01
##
##    item ncat  chi12  chi13  chi23
## 1     1    2 0.8622 0.8487 0.5852
## 2     2    2 0.0000 0.0000 0.0000
## 3     3    2 0.6586 0.6350 0.3984
## 4     4    2 0.0845 0.1584 0.4000
## 5     5    2 0.3027 0.5812 0.8797
## 6     6    2 0.6810 0.7371 0.5066
## 7     7    2 0.6027 0.3302 0.1631
## 8     8    2 0.9814 0.8853 0.6219
## 9     9    2 0.8851 0.7054 0.4106
## 10   10    2 0.0796 0.1039 0.2276
```

Izbor različitog kriterijuma (LR test ili $R^2$) ne utiče na sama izračunavanja već na to na osnovu kojeg pokazatelja će stavka biti označena da ima DIF. Ako se opredelimo za $\Delta R^2$) onda će to biti kriterijum, a u slučaju LR testa – $G^2$. Ako se uz $\Delta R^2$ opredelimo i za Monte Karlo simulacije onda će kritična vrednost biti određena na osnovu njih, a neće se koristiti kritične vrednosti koje predlažu Dželin i Zumbo (Gelin & Zumbo, 2003) ili Žodon i Girl (Jodoin & Gierl, 2001).

I na ovaj način stavka 2 je identifikovana kao stavka koja pokazuje diferencijalno funkcionisanje.

### 10.6.3 Ispitivanje DIF-a pomoću $G^2$ testa IRT modela

Ovde će biti prikazana procedura u paketu *mirt*. Prvo će biti procenjen ograničeni model za više grupa (engl. *multiple group*). Ograničenje se sastoji u tome da parametri za stavke moraju biti jednaki u obe grupe. Biće procenjen dvoparametarski logistički model za binarno skorovane stavke. Pre analize DIF-a pretpostavlja se da model u obe grupe ima dobre pokazatelje fita. Proverili smo.

```
modR <- mirt(b.podD[grupa=="R", ], model=1, itemtype="2PL", SE=T, verbose = FALSE)
M2(modR)

##             M2 df         p      RMSEA RMSEA_5   RMSEA_95      SRMSR       TLI
## stats 40.68253 35 0.2344305 0.01803793       0 0.03835166 0.03077471 0.9953584
##             CFI
## stats 0.9963899

itemfit(modR, fit_stats = "S_X2")

##    item  S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1 3.937       7      0.000  0.787
## 2   it2 5.699       6      0.000  0.458
## 3   it3 5.898       7      0.000  0.552
## 4   it4 6.730       7      0.000  0.458
## 5   it5 4.784       5      0.000  0.443
## 6   it6 2.897       6      0.000  0.822
## 7   it7 4.288       5      0.000  0.509
## 8   it8 6.930       6      0.018  0.327
## 9   it9 1.271       5      0.000  0.938
## 10 it10 5.919       4      0.031  0.205

modF <- mirt(b.podD[grupa=="F", ], model=1, itemtype="2PL", SE=T, verbose = FALSE)
M2(modF)

##             M2 df         p RMSEA RMSEA_5   RMSEA_95      SRMSR      TLI CFI
## stats 29.77769 35 0.7181747     0       0 0.02487768 0.02779944 1.005688   1

itemfit(modF, fit_stats = "S_X2")

##    item   S_X2 df.S_X2 RMSEA.S_X2 p.S_X2
## 1   it1  3.703       6      0.000  0.717
## 2   it2  4.349       7      0.000  0.739
## 3   it3  7.316       6      0.021  0.293
## 4   it4 10.731       6      0.040  0.097
## 5   it5  4.757       6      0.000  0.575
## 6   it6  4.608       6      0.000  0.595
## 7   it7  9.521       5      0.043  0.090
## 8   it8  6.643       7      0.000  0.467
## 9   it9  6.083       6      0.005  0.414
## 10 it10  4.746       5      0.000  0.448
```

2PL model ima dobre omnibus pokazatelje fita u obe grupe. Nema misfitujućih stavki.

Nakon toga pristupamo analizi DIF-a. Procenjujemo ograničeni multigrupni model sa jednakim parametrima u obe grupe.

```
mod.constr <- multipleGroup(b.podD, 1, itemtype="2PL", group=grupa,
 invariance=c(colnames(b.podD), 'free_means', 'free_var'), method="EM",
 verbose=FALSE, SE=TRUE)
# Argument invariance određuje koji su parametri ograničeni da budu jednaki u
različitim grupama. Potrebno je navesti vektor koji sadrži nazive stavki na koje
se ograničenje odnosi
# U osnovnom modelu to su sve stavke i AS i varijanse
invariance = c(colnames(b.podD), 'free_means', 'free_var')
# Postoje jos parametri "slopes" - nagibi i "intercepts" - težine
# Od ostalih argumenata bitno je navesti SE=TRUE, kako bi bila izračunata
informativnost
```

```
constr.par <- coef(mod.constr,simplify = TRUE,IRTpars=TRUE)[[1]][[1]]

# Ispis parametara osnovnog modela (parametri jednaki za obe grupe)
constr.par
```

```
##             a          b g u
## it1  1.260801 -0.1104828 0 1
## it2  1.582165 -0.3102045 0 1
## it3  1.320517 -0.1718701 0 1
## it4  1.393811 -0.2605322 0 1
## it5  2.096019 -1.2175544 0 1
## it6  1.737800 -1.0993676 0 1
## it7  1.132374  1.9126512 0 1
## it8  1.482677 -0.7686883 0 1
## it9  1.943852 -1.2569854 0 1
## it10 2.509321  0.6197300 0 1
```

Zatim testiramo modele sa oslobođenim parametrima za stavke. Oslobađaju se parametri za jednu po jednu stavku, dok su parametri ostalih ograničeni (odnosno koriste se kao sidra – AOAA pristup) (videti odeljak 8.2).

Poredi se logLik (logaritam verodostojnosti) takvih modela i osnovnog, ograničenog modela. Ako model sa oslobođenim parametrima za stavku ima bolji fit od ograničenog, stavka čiji su parametri oslobođeni pokazuje znake diferencijalnog funkcionisanja.

```
# Oslobađaju se parametri za svaku stavku pojedinačno, a svi ostali su sidra
# Paket parallel omogućava korišćenje više jezgara uz paket mirt
p_load(parallel)
ncores <- parallel::detectCores()
mirtCluster(ncores-1) # Ubrzanje obrade - sva jezgra osim jednog obrađuju
istovremeno
# Možete staviti onoliko jezgara (threadova procesora) koliko imate. Ne više od
toga

# Funkcija DIF(), uneti naziv osnovnog modela (mod.constr), which.par - koje
parametre osloboditi (a1 - nagib, d1, d2... težine kojih moze biti za jednu manje
od broja kategorija odgovora)
dif.drop <- DIF(mod.constr, c('a1','d'), scheme = 'drop', seq_stat = .05,
 items2test=1:10)
# Argument scheme - definiše kako se radi oslobađanje parametara:
"drop" - oslobađa jedan po jedan, "add" - svi su slobodni pa ograničava jedan po
jedan
# items2test - numerički vektor (npr. c(1,2,3,4,5)) koji sadrži brojeve stavki
koje treba testirati na DIF. Ako je neki izostavljen on se tretira kao sidro
# seq_stat = .05 definiše koji se kriterijum koristi. Može biti AIC, BIC, SABIC
ili HQ, a ako je numerički kao što je to ovde slučaj onda se koristi LR test

# Značajnosti razlika modela sa slobodnim parametrima i ograničenim
dif.drop
```

```
##      groups converged     AIC   SABIC      HQ     BIC     X2 df     p
## it1     R,F      TRUE   3.446   6.910   7.177  13.262  0.554  2 0.758
## it2     R,F      TRUE -23.077 -19.613 -19.346 -13.261 27.077  2     0
## it3     R,F      TRUE   3.161   6.624   6.891  12.976  0.839  2 0.657
## it4     R,F      TRUE  -1.499   1.964   2.232   8.317  5.499  2 0.064
## it5     R,F      TRUE   0.685   4.149   4.416  10.501  3.315  2 0.191
## it6     R,F      TRUE   2.432   5.896   6.163  12.248  1.568  2 0.457
```

```
## it7     R,F      TRUE   2.515   5.978   6.245  12.330  1.485  2 0.476
## it8     R,F      TRUE   3.526   6.990   7.257  13.342  0.474  2 0.789
## it9     R,F      TRUE   2.449   5.913   6.180  12.265  1.551  2 0.461
## it10    R,F      TRUE   1.087   4.550   4.817  10.902  2.913  2 0.233
```

U gornjoj tabeli *p* vrednost u poslednjoj koloni odnosi se na značajnost $G^2$ testa, odnosno govori nam da li je razlika između fita ograničenog modela i modela sa oslobođenim parametrima za ajtem koji posmatramo statistički značajna. Ako jeste, onda se parametri za stavku razlikuju u dve grupe, odnosno postoji DIF.

Sledećom komandom ćemo iz objekta *dif.drop* u koji smo upisali rezultate prethodne analize izdvojiti samo stavke koje pokazuju DIF.

```
dif.drop[dif.drop$p<.05,]
```

```
##     groups converged     AIC   SABIC      HQ     BIC     X2 df p
## it2    R,F      TRUE -23.077 -19.613 -19.346 -13.261 27.077  2 0
```

Prikazaćemo parametre stavki koje nemaju DIF. Pošto tabele *constr.par* i *dif.drop* imaju isti broj redova i poredak (po rednom broju stavki), koristimo indeks formiran na osnovu tabele *dif.drop* *`[!dif.drop$p<.05,]`* – samo one stavke koje nemaju značajan DIF (*p* nije manje od .05), da odaberemo stavke koje su invarijantne.

Sortiraćemo ih po opadajućem parametru nagiba pošto kao sidra želimo da izaberemo do 5 invarijantnih stavki sa najvišim nagibom (videti odeljak 8.2).

```
sidra <- as.data.frame(constr.par[!dif.drop$p<.05,])
sidra[order(-sidra$a),]
```

```
##             a          b g u
## it10 2.509321  0.6197300 0 1
## it5  2.096019 -1.2175544 0 1
## it9  1.943852 -1.2569854 0 1
## it6  1.737800 -1.0993676 0 1
## it8  1.482677 -0.7686883 0 1
## it4  1.393811 -0.2605322 0 1
## it3  1.320517 -0.1718701 0 1
## it1  1.260801 -0.1104828 0 1
## it7  1.132374  1.9126512 0 1
```

Kada smo se opredelili za sidrene stavke, procenićemo model u kojem će parametri samo ovih stavki biti ograničeni da budu jednaki u dve grupe (uz aritmetičke sredine i varijanse). Argumentu *invariance* prosleđujemo vektor sa imenima sidrenih stavki. Odabrali smo 3 stavke sa najvišim parametrom diskriminativnosti: it10, it5 i it9.

```
# Kreiramo vektor sa imenima sidrenih stavki
imena.sidra <- names(b.podD)[c(10, 5,9)]
# Procenjujemo model za dve grupe
model.sidra <- multipleGroup(b.podD, 1, itemtype="2PL", group=grupa,
                             invariance = c(imena.sidra, 'free_means', 'free_var'),
                             method="EM", verbose=FALSE, SE=TRUE)
# Prikazaćemo parametre modela za dve grupe
R.par <- coef(model.sidra, simplify=TRUE,IRTpars=TRUE)$R$items # Referentna
F.par <- coef(model.sidra, simplify=TRUE,IRTpars=TRUE)$F$items # Fokalna
```

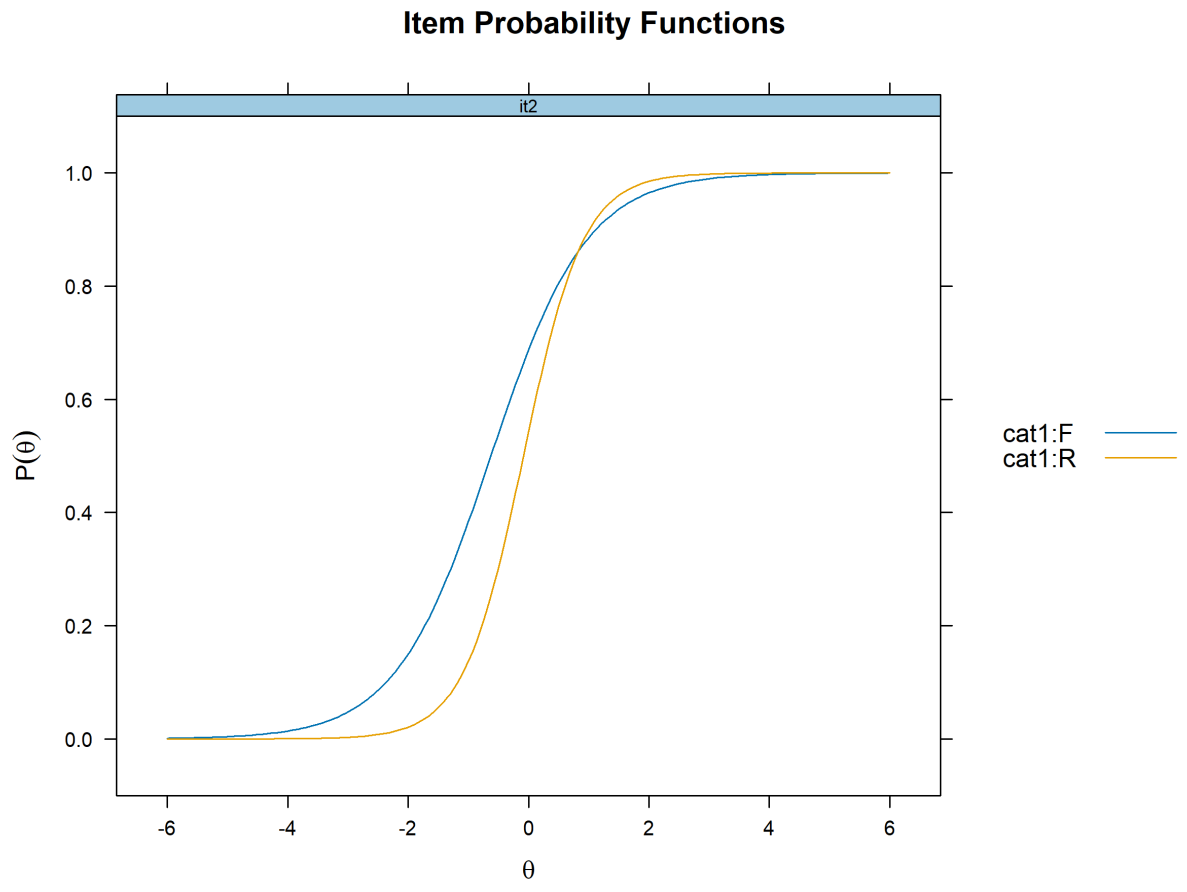
```
cbind(R.par, F.par)
```

```
##             a           b g u         a           b g u
## it1  1.321595 -0.13653103 0 1 1.1246519 -0.09803735 0 1
## it2  2.009936 -0.08420684 0 1 1.2615309 -0.62270951 0 1
## it3  1.202232 -0.17914857 0 1 1.3422966 -0.18934171 0 1
## it4  1.495228 -0.35694464 0 1 1.2133439 -0.18101058 0 1
## it5  2.039774 -1.26427656 0 1 2.0397735 -1.26427656 0 1
## it6  1.635616 -1.18731890 0 1 1.7350512 -1.10253351 0 1
## it7  1.308713  1.76901405 0 1 0.8950873  2.27841889 0 1
## it8  1.527032 -0.78189615 0 1 1.3223776 -0.82894383 0 1
## it9  1.871957 -1.31052865 0 1 1.8719574 -1.31052865 0 1
## it10 2.377522  0.64219271 0 1 2.3775217  0.64219271 0 1
```

```
# Levo su prikazani parametri referentne, a desno fokalne grupe
```

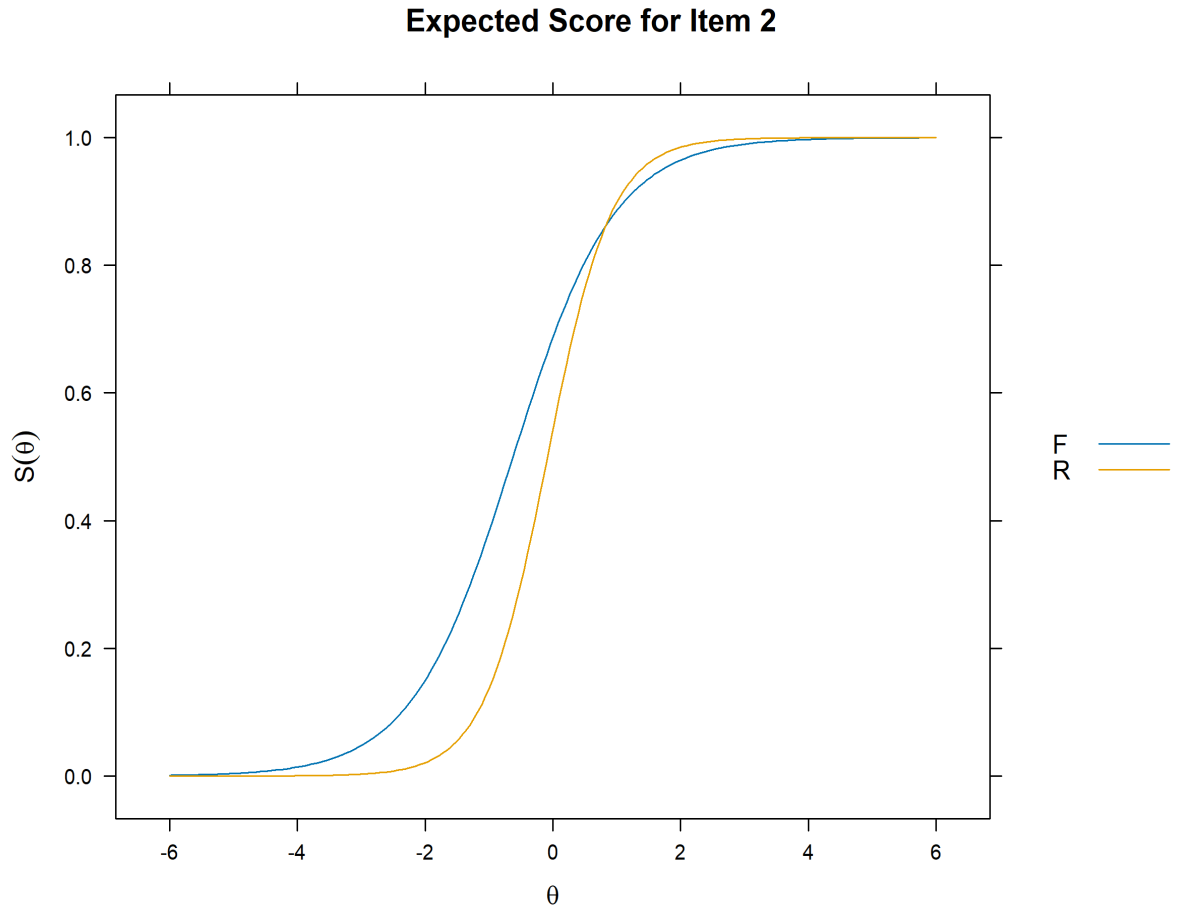
Uočite da su samo za sidrene stavke (10, 9 i 5) parametri jednaki za obe grupe. Na ovom modelu ponovićemo analizu DIF-a. Argument *plotdif = TRUE* bi zahtevao grafikone stavki koje imaju DIF (mi smo stavili *FALSE*).

```
# Potrebno je da napravimo vektor sa brojevima stavki koje ćemo testirati na DIF
test.stavke <- c(1:4,6:8) # To su sve stavke osim sidra
dif.stavke <- DIF(model.sidra, c('a1','d'), items2test = test.stavke,
 plotdif = TRUE, seq_stat = .05)
```




Slika 126 – Uporedni prikaz KKS za stavku 2 u referentnoj i fokalnoj grupi

```
dif.stavke

##      groups converged      AIC    SABIC       HQ     BIC      X2 df     p
## it1     R,F      TRUE    3.236    6.700    6.967 13.052   0.764  2 0.683
## it2     R,F      TRUE -18.605 -15.141 -14.874 -8.789 22.605  2     0
## it3     R,F      TRUE    3.662    7.125    7.392 13.477   0.338  2 0.844
## it4     R,F      TRUE    0.132    3.595    3.862  9.947   3.868  2 0.145
## it6     R,F      TRUE    3.758    7.221    7.488 13.573   0.242  2 0.886
## it7     R,F      TRUE    2.010    5.474    5.741 11.826    1.99  2  0.37
## it8     R,F      TRUE    3.470    6.934    7.201 13.286    0.53  2 0.767

# Narednom komandom možemo dobiti grafikon stavke sa DIF
itemplot(model.sidra, item=2, type="score")
```




Slika 127 – Uporedni prikaz funkcija očekivanog skora na stavci 2 u zavisnosti od nivoa osobine za referentnu i fokalnu grupu

Vidimo da samo stavka 2 ima značajni $G^2$ test, što ukazuje da se samo ova stavka ponaša drugačije u dve grupe. Na grafikonu sa karakterističnim krivama stavke 2 u dve grupe, vidimo da u fokalnoj grupi ova stavka ima manji nagib i težinu. Testiraćemo DIF posebno za nagib, a posebno za težinu. Prvo testiramo DIF za parametar težine, odnosno uniformni DIF. Argumentu *which.par=c('d')* smo prosledili samo parametar *d* koji označava težinu. U modelima za politomne stavke u obzir dolaze i oznake *d1*, *d2* i tako dalje, sve do broja koji je za jedan manji od broja kategorija odgovora.

```
dif.stavke.t <- DIF(model.sidra, which.par=c('d'), items2test = test.stavke,
 plotdif = FALSE)
```

```
dif.stavke.t
```

```
##      groups converged    AIC  SABIC     HQ    BIC     X2 df     p
## it1     R,F      TRUE  1.804  3.535  3.669  6.711  0.196  1 0.658
## it2     R,F      TRUE -9.561 -7.830 -7.696 -4.654 11.561  1 0.001
## it3     R,F      TRUE  1.944  3.676  3.810  6.852  0.056  1 0.814
## it4     R,F      TRUE -1.502  0.230  0.364  3.406  3.502  1 0.061
## it6     R,F      TRUE  1.990  3.722  3.855  6.898   0.01  1  0.92
## it7     R,F      TRUE  0.915  2.647  2.780  5.823  1.085  1 0.298
## it8     R,F      TRUE  1.767  3.498  3.632  6.674  0.233  1 0.629
```

Jedino stavka 2 ima značajan uniformni DIF, dok je za stavku 4 on blizak značajnosti.

Proverićemo i DIF za nagib, odnosno neuniformni DIF. Argumentu *which.par=c('a1')* smo prosledili samo parametar *a1* koji označava nagib za prvu dimenziju. Pošto paket *mirt* može da procenjuje i višedimenzionalne modele, u obzir dolaze i oznake *a2*, *a3*... po jedan za svaku dimenziju koju test meri.

```
dif.stavke.n <- DIF(model.sidra, which.par=c('a1'), items2test = test.stavke,
 plotdif = FALSE)
dif.stavke.n
```

```
##      groups converged    AIC  SABIC     HQ   BIC    X2 df     p
## it1     R,F      TRUE  1.356  3.088  3.221 6.264 0.644  1 0.422
## it2     R,F      TRUE -3.630 -1.898 -1.765 1.278  5.63  1 0.018
## it3     R,F      TRUE  1.685  3.417  3.550 6.593 0.315  1 0.575
## it4     R,F      TRUE  0.887  2.618  2.752 5.794 1.113  1 0.291
## it6     R,F      TRUE  1.919  3.651  3.784 6.827 0.081  1 0.776
## it7     R,F      TRUE  0.086  1.818  1.951 4.994 1.914  1 0.167
## it8     R,F      TRUE  1.493  3.225  3.359 6.401 0.507  1 0.477
```

I ovde je značajan DIF samo za stavku 2, što znači da se nagibi značajno razlikuju u referentnoj i fokalnoj grupi, odnosno da postoji neuniformni DIF.

Postojanje značajnog DIF-a za stavku 2 još uvek ne znači da je on toliki da narušava merenje. Procenićemo veličinu efekta kako bismo to utvrdili.

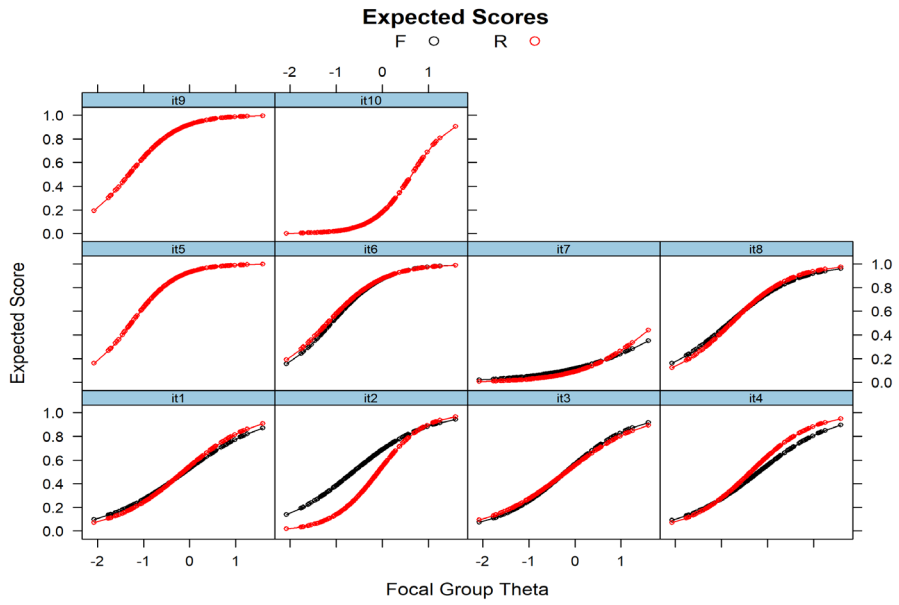
```
empirical_ES(model.sidra)
```

```
##        SIDS  UIDS   SIDN  UIDN   ESSD theta.of.max.D  max.D mean.ES.foc
## it1  -0.012 0.025 -0.011 0.025 -0.053          0.979 -0.043       0.514
## it2   0.127 0.133  0.124 0.130  0.486         -0.796  0.253       0.643
## it3   0.005 0.017  0.005 0.017  0.024          0.979  0.027       0.540
## it4  -0.053 0.056 -0.050 0.054 -0.227          0.478 -0.087       0.535
## it5   0.000 0.000  0.000 0.000  0.000         -1.577  0.000       0.839
## it6  -0.010 0.011 -0.011 0.012 -0.051         -1.539 -0.041       0.788
## it7   0.008 0.026  0.004 0.029  0.084          1.590 -0.091       0.133
## it8  -0.005 0.019 -0.003 0.020 -0.021         -1.709  0.043       0.695
## it9   0.000 0.000  0.000 0.000  0.000         -1.577  0.000       0.839
## it10  0.000 0.000  0.000 0.000  0.000         -1.577  0.000       0.281
##      mean.ES.ref
## it1        0.526
## it2        0.517
## it3        0.535
## it4        0.589
## it5        0.839
## it6        0.798
## it7        0.125
```

```
## it8         0.700
## it9         0.839
## it10        0.281
```

U gornjoj tabeli su navedene različite procene veličine efekta. SIDS je prosečna razlika u očekivanom skoru na stavci u fokalnoj grupi (opisano u odeljku 8.2). Pošto vodi računa o predznaku, negativne i pozitivne razlike mogu se potirati. To se može dogoditi kada postoji neuniformni DIF. UIDS je sličan pokazatelj koji ne vodi računa o predznaku (prosečna apsolutna razlika u očekivanom skoru na uzorku). Vidimo da se u slučaju stavke 2 za koju je otkriven DIF, ove dve vrednosti ne razlikuju puno, što znači da iako postoji neuniformni DIF kod ove stavke je značajniji uniformni. Sledeće dve kolone (SIDN i UIDN) su forme ovih pokazatelja koje nisu računate na uzorku već u odnosu na standardizovanu normalnu distribuciju. U našem slučaju ove vrednosti se ne razlikuju puno od onih dobijenih na uzorku jer smo nivo osobine u podacima simulirali upravo iz standardizovane normalne distribucije. ESSD je standardizovana razlika u očekivanom skoru. Računa se tako što se SIDS podeli zajedničkom standardnom devijacijom parametara ispitanika izračunatih na osnovu parametara stavki referentne i na osnovu parametara stavki fokalne grupe. U suštini, radi se o Koenovom d za očekivane skorove, pa se tako i interpretira (odeljak 8.2). ESSD za stavku 2 iznosi 0,486, što spada u malu veličinu efekta (na granici sa srednjom). Lokacija najveće opažene razlike je na nivou osobine od –0,796 logita. Ovaj podatak nam je bitan jer na osnovu njega možemo videti da li DIF ima uticaja na nivou osobine na kojem očekujemo najveći broj ispitanika ili da li se javlja na delu kontinuuma na kojem su nam odluke koje donosimo na osnovu merenja bitne ili ne. Maksimalna razlika je 0,253, odnosno ispitanici iz fokalne grupe imaju približno za 25% veću verovatnoću da tačno odgovore na ovu stavku od ispitanika iz referentne grupe koji imaju jednak nivo merene osobine. U poslednje dve kolone dati su prosečni očekivani skorovi u dve grupe. Vidimo da je prosečan očekivani skor viši u fokalnoj grupi, što znači da je ova stavka lakša za ispitanike iz te grupe.

```
# Grafikon očekivanih skorova na stavkama u dve grupe
empirical_ES(model.sidra, plot=TRUE)
```




Slika 128 – Uporedni prikaz funkcija očekivanog skora na stavkama u zavisnosti od nivoa osobine za referentnu i fokalnu grupu

Isto nam pokazuje i grafikon (Slika 128). Na grafikonu vidimo da se KKS u dve grupe razlikuju i za stavku 4, ali videli smo da te razlike ne dosežu nivo statističke značajnosti. Za sidrene stavke prikazana je po jedna KKS.

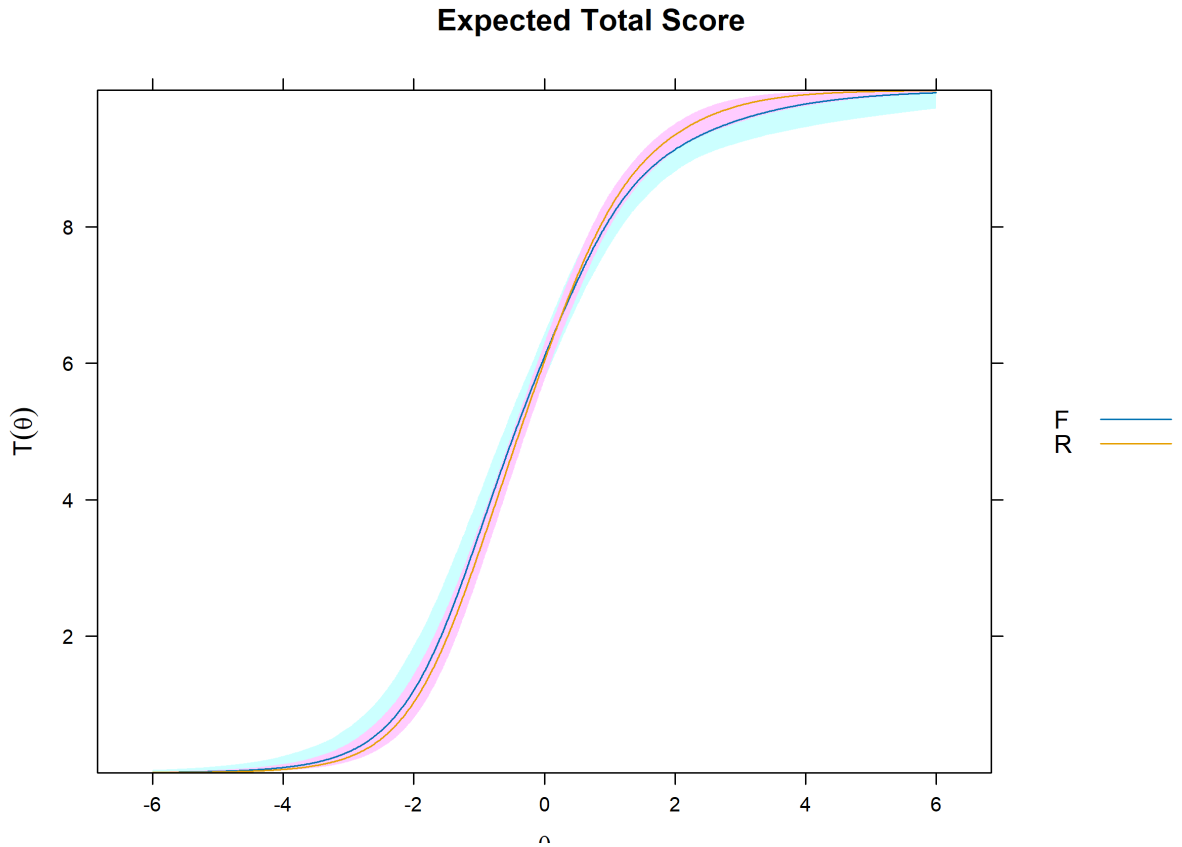
Osim pokazatelja za stavke, moguće je dobiti i pokazatelje za test u celini (argument *DIF=FALSE)*. Pokazatelji su ekvivalentni onima za stavke samo što se odnose na test u celini.

```
empirical_ES(model.sidra, DIF=FALSE)
```

```
##           Effect Size       Value
## 1                STDS  0.06031063
## 2                UTDS  0.28781852
## 3              UETSDS  0.14923051
## 4               ETSSD  0.02952498
## 5         Starks.DTFR  0.05725529
## 6               UDTFR  0.28711310
## 7              UETSDN  0.15267727
## 8 theta.of.max.test.D -1.03150906
## 9           Test.Dmax  0.26350438
```

Prvo ćemo pogledati standardizovanu razliku u skorovima na testu (ETSSD). S obzirom na to da ona iznosi samo 0,029 možemo reći da je veličina efekta trivijalna, te da DIF na stavci 2 ne utiče na očekivane ukupne skorove na testu. Imajući u vidu prethodno rečeno, možemo se osvrnuti na ostale pokazatelje.

```
# Grafikon očekivanih ukupnih skorova za dve grupe
DTF(model.sidra, draws=1000, plot="func")
```




Slika 129 – Uporedni prikaz funkcija očekivanih ukupnih skorova u zavisnosti od nivoa osobine za referentnu i fokalnu grupu

STDS i UTDS su prosečne razlike u ukupnim skorovima na testu. STDS vodi računa o predznaku, a UTDS koristi apsolutne vrednosti. Na osnovu razlike od 0,06 bodova (STDS) vidimo da je uticaj DIF-a mali na ukupne skorove, a s obzirom na to da je prosečna apsolutna razlika nešto veća, vidimo da uticaj nije u potpunosti uniforman. UETSDS je apsolutna razlika u očekivanim skorovima na testu izračunata na uzorku. Ovo je hipotetička razlika kada bi DIF bio isključivo uniforman. UETSDN je isti pokazatelj izračunat na standardizovanoj normalnoj distribuciji. UDTFR je apsolutna očekivana razlika u ukupnim skorovima na uzorku. Ovo je razlika u opaženim sumacionim skorovima koju bismo mogli očekivati u fokalnoj grupi sa standardizovanom normalnom distribucijom parametara ispitanika, kada bi DIF bio isključivo uniformne prirode. Lokacija najveće razlike je –1,03 logita, a maksimalna razlika u očekivanim skorovima je 0,26 bodova (podsetimo se, mogući raspon bodova na ovom testu je 0–10 ). Na grafikonu (Slika 129) vidimo da, iako postoje male razlike u opaženim skorovima za dve grupe, one ne dostižu nivo statističke značajnosti jer se 95% oblasti poverenja (obojene površine) preklapaju.

# 11 Reference


1. Adams, R. J., Wilson, M., & Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, *21*(1), 1–23. https://doi.org/10.1177/0146621697211001

2. Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

3. Ali, U. S., Chang, H.-H., & Anderson, C. J. (2015). Location Indices for Ordinal Polytomous Items Based on Item Response Theory: IRT Location Indices for Ordinal Polytomous Items. *ETS Research Report Series*, *2015*(2), 1–13. https://doi.org/10.1002/ets2.12065

4. Andersen, E. B. (1977). Sufficient Statistics and Latent Trait Models. *Psychometrika*, *42*(1), 69–81. https://doi.org/10.1007/BF02293746

5. Andersen, E. B. (1983). A General Latent Structure Model for Contingency Table Data. U H. Wainer & S. Messick (Ur.), *Principals of Modern Psychological Measurement* (str. 117–139). Lawrence Erlbaum Associates.

6. Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the Kernel Method of Test Equating with the Package kequate. *Journal of Statistical Software*, *55*(6), 1–25.

7. Andrich, D. (1978a). A Rating Formulation for Ordered Response Categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

8. Andrich, D. (1978b). Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. *Applied Psychological Measurement*, *2*(4), 581–594. https://doi.org/10.1177/014662167800200413

9. Angoff, W. H. (1993). Perspectives on Differential Item Functioning Methodology. U P. W. Holland & H. Wainer (Ur.), *Differential item functioning* (str. 3–23). Erlbaum.

10. Baghaei, P. (2008). Local Dependency and Rasch Measures. *Rasch Measurement Transactions*, *21*(3), 1105–1106.

11. Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd ed). ERIC Clearinghouse on Assessment and Evaluation.

12. Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques* (Second Edition). CRC Press.

13. Bartolucci, F., Bacci, S., & Perugia, M. G.-U. of. (2017). *MultiLCIRT: Multidimensional Latent Class Item Response Theory Models*. https://CRAN.R-project.org/package=MultiLCIRT

14. Bartolucci, F., & Perugia, S. B.-U. of. (2019). *MLCIRTwithin: Latent Class Item Response Theory (LC-IRT) Models under Within-Item Multidimensionality*. https://CRAN.R-project.org/package=MLCIRTwithin

15. Battauz, M. (2015). equateIRT: An R Package for IRT Test Equating. *Journal of Statistical Software*, *68*(7), 1–22. https://doi.org/10.18637/jss.v068.i07

16. Bazán, J. L., Bolfarine, H., & Branco, M. D. (2006). A Skew Item Response Model. *Bayesian Analysis*, *1*(4). https://doi.org/10.1214/06-BA128

17. Birnbaum, A. (1968). Some Latent Trait Models. U Lord, Frederic M. & M. R. Novick (Ur.), *Statistical theories of mental test scores (with contributions of Allan Birnbaum)* (str. 397–424). Addison-Wesley.

18. Bock, R. D. (1972). Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories. *Psychometrika*, *37*, 29–51. https://doi.org/10.1007/BF02291411

19. Bock, R. D. (1997). A Brief History of Item Response Theory. *Educational Measurement: Issues and Practice*, *16*(4), 21–33. https://doi.org/10.1111/j.1745-3992.1997.tb00605.x

20. Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-Information Item Factor Analysis. *Applied Psychological Measurement*, *12*(3), 261–280. https://doi.org/10.1177/014662168801200305

21. Bock, R. D., & Schilling, S. (1997). High-dimensional Full-information Item Factor Analysis. U M. Berkane (Ur.), *Latent Variable Modeling and Applications to Causality* (Sv. 120, str. 163–176).

Springer New York. https://doi.org/10.1007/978-1-4612-1842-5_8

22. Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

23. Cai, L., & Hansen, M. (2013). Limited-Information Goodness-of-Fit Testing of Hierarchical Item Factor Models: <i>Testing Hierarchical Item Factor Models</I>. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 245–276. https://doi.org/10.1111/j.2044-8317.2012.02050.x

24. Cai, L., & Monroe, S. (2014). *A New Statistic for Evaluating Item Response Theory Models for Ordinal Data* (CRESST Report 839). National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

25. Cervantes, V. H. (2017). DFIT: An R Package for Raju's Differential Functioning of Items and Tests Framework. *Journal of Statistical Software*, *76*(5), 1–24. https://doi.org/10.18637/jss.v076.i05

26. Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

27. Chalmers, R. P. (2016). Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications. *Journal of Statistical Software*, *71*(5), 1–39. https://doi.org/10.18637/jss.v071.i05

28. Chalmers, R. P., & Ng, V. (2017). Plausible-Value Imputation Statistics for Detecting Item Misfit. *Applied Psychological Measurement*, *41*(5), 372–387. https://doi.org/10.1177/0146621617692079

29. Chen, W.-H., & Revicki, D. (2014). Differential Item Functioning (DIF). U A. C. Michalos (Ur.), *Encyclopedia of Quality of Life and Well-Being Research* (str. 1611–1614). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_728

30. Chen, W.-H., & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265. https://doi.org/10.2307/1165285

31. Chernyshenko, O. S., Stark, S., Chan, K.-Y., Drasgow, F., & Williams, B. (2001). Fitting Item Re-

sponse Theory Models to Two Personality Inventories: Issues and Insights. *Multivariate Behavioral Research*, *36*(4), 523–562. https://doi.org/10.1207/S15327906MBR3604_03

32. Choi, S. W., Gibbons, with contributions from L. E., & Crane, P. K. (2016). *lordif: Logistic Ordinal Regression Differential Item Functioning using IRT*. https://CRAN.R-project.org/package=lordif

33. Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). **lordif**: An *R* Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software*, *39*(8). https://doi.org/10.18637/jss.v039.i08

34. Choi, Y.-J., & Asilkalkan, A. (2019). R Packages for Item Response Theory Analysis: Descriptions and Features. *Measurement: Interdisciplinary Research and Perspectives*, *17*(3), 168–175. https://doi.org/10.1080/15366367.2019.1586404

35. Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's $Q_3$: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, *41*(3), 178–194. https://doi.org/10.1177/0146621616677520

36. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). L. Erlbaum Associates.

37. Crane, P. K., Belle, G. V., & Larson, E. B. (2004). Test Bias in a Cognitive Test: Differential Item Functioning in the CASI. *Statistics in Medicine*, *23*(2), 241–256. https://doi.org/10.1002/sim.1713

38. Crane, P. K., Gibbons, L. E., Jolley, L., & Van Belle, G. (2006). Differential Item Functioning Analysis With Ordinal Logistic Regression Techniques: DIFdetect and difwithpar. *Medical Care*, *44*(Suppl 3), S115–S123. https://doi.org/10.1097/01.mlr.0000245183.28384.ed

39. Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., Hays, R. D., & Teresi, J. A. (2007). A Comparison of Three Sets of Criteria for Determining the Presence of Differential Item Functioning Using Ordinal Logistic Regression. *Quality of Life Research*, *16*(S1), 69–84. https://doi.org/10.1007/s11136-007-9185-5

40. Crişan, D. R., Tendeiro, J., & Meijer, R. (2019). *The Crit Value as an Effect Size Measure for Violations of Model Assumptions in Mokken Scale Analysis for Binary Data* [Preprint]. PsyArXiv.

https://doi.org/10.31234/osf.io/8ydmr

41. Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of Polytomous IRT Models With Rating Scale Data: An Investigation Over Sample Size, Instrument Length, and Missing Data. *Frontiers in Education*, *6*, 721963. https://doi.org/10.3389/feduc.2021.721963

42. de Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory* (2nd izd.). The Guilford Press.

43. DeMars, C. E. (2010). *Item Response Theory*. Oxford University Press.

44. DeMars, C. E. (2016). Partially Compensatory Multidimensional Item Response Theory Models: Two Alternate Model Forms. *Educational and Psychological Measurement*, *76*(2), 231–257. https://doi.org/10.1177/0013164415589595

45. Dodd, B., & De Ayala, R. (1994). Item Information as a Function of Threshold Values in the Rating Scale Model. U M. Wilson (Ur.), *Objective measurement: Theory into practice* (Sv. 2, str. 301–317). Ablex Norwood, NJ.

46. Douglas, G. (1982). Issues in the Fit of Data to Psychometric Models. *Education Research and Perspectives*, *9*(June 1982), 32–43.

47. Drabinova, A., & Martinkova, P. (2017). Detection of Differential Item Functioning with Nonlinear Regression: A Non-IRT Approach Accounting for Guessing. *Journal of Educational Measurement*, *54*(4), 498–517. https://doi.org/10.1111/jedm.12158

48. Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness Measurement with Polychotomous Item Response Models and Standardized Indices. *British Journal of Mathematical and Statistical Psychology*, *38*(1), 67–86. https://doi.org/10.1111/j.2044-8317.1985.tb00817.x

49. Drasgow, F., & Lissak, R. I. (1983). Modified Parallel Analysis: A Procedure for Examining the Latent Dimensionality of Dichotomously Scored Item Responses. *Journal of Applied Psychology*, *68*(3), 363–373. https://doi.org/10.1037/0021-9010.68.3.363

50. Embretson, S. E. (1984). A General Latent Trait Model for Response Processes. *Psychometrika*, *49*(2), 175–186.

51. Embretson, S. E. (1991). A Multidimensional Latent Trait Model for Measuring Learning and Change. *Psychometrika*, *56*, 495–515.

52. Embretson, S. E. (1996). The New Rules of Measurement. *Psychological Assessment*, *8*(4), 341–349. https://doi.org/10.1037/1040-3590.8.4.341

53. Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. L. Erlbaum Associates.

54. Evers, A., Hagemeister, C., Høstm, A., Lindley, P., Muñiz, J., & Sjöberg, A. (2013, srpanj 13). *EFPA Review Model For The Description And Evaluation Of Psychological And Educational Tests—Test Review Form And Notes For Reviewers Version*.

55. Fajgelj, S. (2004). *Metode istraživanja ponašanja*. Centar za primenjenu psihologiju.

56. Fajgelj, S. (2020). *Psihometrija—Metod i teorija psihološkog merenja*. Centar za primenjenu psihologiju.

57. Fajgelj, S., & Janičić, B. (2008). KakaoBEJZ: Makro za ajtem analizu dihotomno skorovanih ajtema - teorija ajtemskog odgovora. *Primenjena psihologija*, *1*(3–4), 223–241. https://doi.org/10.19090/pp.2008.3-4.223-241

58. Fajgelj, S., & Janičić, B. (2009). *KakaoBEJZ i KakaoMIKS: SPSS makroi za ajtem analizu – Teorija ajtemskog odgovora*. Unpublished paper, University of Novi Sad.

59. Ferguson, G. A. (1942). Item Selection by the Constant Process. *Psychometrika*, *7*(1), 19–29. https://doi.org/10.1007/BF02288601

60. Fox, J.-P., Klotzke, K., & Entink, R. K. (2021). *LNIRT: LogNormal Response Time Item Response Theory Models*. https://CRAN.R-project.org/package=LNIRT

61. French, B. F., Finch, W. H., & Immekus, J. C. (2019). Multilevel Generalized Mantel-Haenszel for Differential Item Functioning Detection. *Frontiers in Education*, *4*, 47. https://doi.org/10.3389/feduc.2019.00047

62. Frick, H., Strobl, C., Leisch, F., & Zeileis, A. (2012). Flexible Rasch Mixture Models with Package psychomix. *Journal of Statistical Software*, *48*(7), 1–25.

63. Gelin, M. N., & Zumbo, B. D. (2003). Differential Item Functioning Results May Change Depending On How An Item Is Scored: An Illustration With The Center For Epidemiologic Studies Depression Scale. *Educational and Psychological Measurement*, *63*(1), 65–74. https://doi.org/10.1177/0013164402239317

64. George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software*, *74*(2), 1–24. https://doi.org/10.18637/jss.v074.i02

65. González, J. (2014). SNSequate: Standard and Nonstandard Statistical Models and Methods for Test Equating. *Journal of Statistical Software*, *59*(7), 1–30. https://doi.org/10.18637/jss.v059.i07

66. Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory*. Springer Netherlands. https://doi.org/10.1007/978-94-017-1988-9

67. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications.

68. Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Educational Measurement: Issues and Practice*, *8*(1), 35–41. https://doi.org/10.1111/j.1745-3992.1989.tb00313.x

69. Harwell, M. R., Baker, F. B., & Zwarts, M. (1988). Item Parameter Estimation Via Marginal Maximum Likelihood and an EM Algorithm: A Didactic. *Journal of Educational Statistics*, *13*(3), 243–271. https://doi.org/10.3102/10769986013003243

70. Hattie, J. (1984). An Empirical Study of Various Indices for Determining Unidimensionality. *Multivariate Behavioral Research*, *19*(1), 49–78. https://doi.org/10.1207/s15327906mbr1901_3

71. Hattie, J., Krakowski, K., Jane Rogers, H., & Swaminathan, H. (1996). An Assessment of Stout's Index of Essential Unidimensionality. *Applied Psychological Measurement*, *20*(1), 1–14. https://doi.org/10.1177/014662169602000101

72. He, Q., & Wheadon, C. (2013). The Effect of Sample Size on Item Parameter Estimation for the Partial Credit Model. *International Journal of Quantitative Research in Education*, *1*(3), 297. https://doi.org/10.1504/IJQRE.2013.057692

73. Heine, J.-H. (2023). *pairwise: Rasch Model Parameters by Pairwise Algorithm*. https://CRAN.R-project.org/package=pairwise

74. Hladka, A., & Martinkova, P. (2023). *difNLR: DIF and DDF Detection by Non-Linear Regression Models*. https://CRAN.R-project.org/package=difNLR

75. Hohensinn, C. (2018). pcIRT: An R Package for Polytomous and Continuous Rasch Models. *Journal of Statistical Software, Code Snippets*, *84*(2), 1–14. https://doi.org/10.18637/jss.v084.c02

76. Holland, P. W., & Thayer, D. T. (1986). Differential Item Functioning and the Mantel-Haenszel Procedure. *ETS Research Report Series*, *1986*(2), i–24.

77. Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179–185.

78. Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

79. Irwing, P., Booth, T., & Hughes, D. J. (Ur.). (2018). *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development: Sv. I*. Wiley-Blackwell.

80. JASP Team. (2024). *JASP (Version 0.18.3)[Computer software]* [Software]. https://jasp-stats.org/

81. Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, *14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

82. Kang, T., & Chen, T. T. (2007). *An Investigation of the Performance of the Generalized SX 2 Item-Fit Index for Polytomous IRT Models. 2007-1. ACT, Inc.* (1; ACT Research Report Series, str. 25).

83. Kean, J., Bisson, E. F., Brodke, D. S., Biber, J., & Gross, P. H. (2018). An Introduction to Item Response Theory and Rasch Analysis: Application Using the Eating Assessment Tool (EAT-10). *Brain Impairment*, *19*(1), 91–102. https://doi.org/10.1017/BrImp.2017.31

84. Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th izd.). Guilford Publications.

85. Koller, I., & Hatzinger, R. (2013). Nonparametric Tests for the Rasch Model: Explanation, Development, and Application of Quasi-Exact Tests for Small Samples. *InterStat*, *11*, 1–16.

86. Lamprianou, I. (2019). *Applying the Rasch Model in Social Sciences Using R and BlueSky Statistics*. Routledge, Taylor and Francis Group.

87. Lathrop, Q. N. (2015). *cacIRT: Classification Accuracy and Consistency under Item Response Theory*. https://CRAN.R-project.org/package=cacIRT

88. Lawley, D. N. (1943). On Problems Connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edinburgh*, *62-A, Part I*(3), 74–81. https://doi.org/10.1017/S0080454100006282

89. Lazarsfeld, P. F. (1950). The Logical and Mathematical Foundation of Latent Structure Analysis. *Studies in social psychology in world war II Vol. IV: Measurement and prediction*, 362–412.

90. Linacre, J. M. (2000). Comparing and Choosing Between „Partial Credit Models“ (PCM) and „Rating Scale Models“ (RSM). *Rasch Measurement Transactions*, *14*(3), 768.

91. Linacre, J. M. (2002). What Do Infit and Outfit, Mean-Square and Standardized Mean. *Rasch measurement transactions*, *16*(2), 878.

92. Linacre, J. M. (2010). When to Stop Removing Items and Persons in Rasch Misfit Analysis. *Rasch Measurement Transactions*, *23*(4), 1241.

93. Linacre, J. M. (with Wright, B. D.). (1993). *A User's Guide to WINSTEPS® Ministep: Rasch-Model Computer Programs*. Mesa Press.

94. Linden, W. J. van der (Ur.). (2010). *Handbook of Modern Item Response Theory*. Springer.

95. Liu, C.-W., & Chalmers, R. P. (2018). Fitting Item Response Unfolding Models to Likert-Scale Data Using Mirt in R. *PLOS ONE*, *13*(5), e0196292. https://doi.org/10.1371/journal.pone.0196292

96. Liu, C.-W., & Wang, W.-C. (2019). A General Unfolding IRT Model for Multiple Response Styles. *Applied Psychological Measurement*, *43*(3), 195–210. https://doi.org/10.1177/0146621618762743

97. Lopez Rivas, G. E., Stark, S., & Chernyshenko, O. S. (2009). The Effects of Referent Item Parameters on Differential Item Functioning Detection Using the Free Baseline Likelihood Ratio Test. *Applied*

*Psychological Measurement*, *33*(4), 251–265. https://doi.org/10.1177/0146621608321760

98. Lord, F. M. (1952). A Theory of Test Scores. *Psychometric monographs*, *7*(x), 8.

99. Lord, F. M. (1953). The Relation of Test Score to the Trait Underlying the Test. *ETS Research Bulletin Series*, *13*, 517–548. https://doi.org/10.1002/j.2333-8504.1952.tb00926.x

100. Lord, F. M. (1975). The 'Ability' Scale in Item Characteristic Curve Theory. *Psychometrika*, *40*(2), 205–217. https://doi.org/10.1007/BF02291567

101. Lord, F. M. (1977). Practical Applications of Item Characteristic Curve Theory. *Journal of Educational Measurement*, *14*(2), 117–138.

102. Lord, F. M., & Novick, M. R. (2008). *Statistical Theories of Mental Test Scores* (Nachdr. der Ausg. Reading, Mass. [u.a.], 1968). Addison-Wesley.

103. Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not Positive Definite Correlation Matrices in Exploratory Item Factor Analysis: Causes, Consequences and a Proposed Solution. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 138–147. https://doi.org/10.1080/10705511.2020.1735393

104. Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scale, Sydney: The Psychological Foundation of Australia.* The Psychological Foundation of Australia.

105. Luce, R. D., & Tukey, J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, *1*(1), 1–27. https://doi.org/10.1016/0022-2496(64)90015-X

106. Luo, X. (2019). *xxIRT: Item Response Theory and Computer-Based Testing in R*. https://CRAN.R-project.org/package=xxIRT

107. MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, *1*(2), 130–149.

108. Magis, D., & Barrada, J. R. (2017). Computerized Adaptive Testing with R: Recent Updates of the Package catR. *Journal of Statistical Software, Code Snippets*, *76*(1), 1–19. https://doi.org/10.18637/jss.v076.c01

109. Magis, D., Beland, S., Tuerlinckx, F., & Boeck, P. D. (2010). A General Framework and an R Package for the Detection of Dichotomous Differential Item Functioning. *Behavior Research Methods*, *42*, 847–862.

110. Magis, D., & Raîche, G. (2012). Random Generation of Response Patterns under Computerized Adaptive Testing with the R Package catR. *Journal of Statistical Software*, *48*(8), 1–31. https://doi.org/10.18637/jss.v048.i08

111. Mair, P. (2018). *Modern Psychometrics with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-93177-7

112. Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, *20*(9). https://doi.org/10.18637/jss.v020.i09

113. Mair, P., & Leeuw, J. D. (2022). *Gifi: Multivariate Analysis with Optimal Scaling* [Software]. https://CRAN.R-project.org/package=Gifi

114. Mair, P., Reise, S. P., & Bentler, P. M. (2008). IRT Goodness-of-Fit Using Approaches from Logistic Regression. *Department of Statistics Papers, Department of Statistics, UCLA, UC Los Angeles*. http://escholarship.org/uc/item/1m46j62q

115. Marais, I. (2013). Local Dependance. U K. B. Christensen, S. Kreiner, & M. Mesbah (Ur.), *Rasch Models in Health* (str. 111–130). John Wiley & Sons.

116. Maris, E. (1995). Psychometric Latent Response Models. *Psychometrika*, *60*(4), 523–547. https://doi.org/10.1007/BF02294327

117. Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

118. Masur, P. K. (2022). *ggmirt: Plotting Functions to Extend „mirt" for IRT Analyses*.

119. Maydeu-Olivares, A. (2013). Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 71–101. https://doi.org/10.1080/15366367.2013.831680

120. Maydeu-Olivares, A., Hernández, A., & McDonald, R. P. (2006). A Multidimensional Ideal Point Item Response Theory Model for Binary Data. *Multivariate Behavioral Research*, *41*(4), 445–472.

https://doi.org/10.1207/s15327906mbr4104_2

121. Maydeu-Olivares, A., & Joe, H. (2006). Limited Information Goodness-of-fit Testing in Multidimensional Contingency Tables. *Psychometrika*, *71*(4), 713–732. https://doi.org/10.1007/s11336-005-1295-9

122. McCulloch, C. E., & Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons.

123. McKinley, R. L., & Mills, C. N. (1985). A Comparison of Several Goodness-of-Fit Statistics. *Applied Psychological Measurement*, *9*(1), 49–57. https://doi.org/10.1177/014662168500900105

124. Mead, R. (2008). A Rasch Primer: The Measurement Theory of Georg Rasch. U *Psychometrics services research memorandum 2008–001*. Data Recognition Corporation.

125. Meade, A. W. (2010). A Taxonomy of Effect Size Measures for the Differential Functioning of Items and Scales. *Journal of Applied Psychology*, *95*(4), 728–743. https://doi.org/10.1037/a0018966

126. Meade, A. W., & Wright, N. A. (2012). Solving the Measurement Invariance Anchor Item Problem in Item Response Theory. *Journal of Applied Psychology*, *97*(5), 1016–1031. https://doi.org/10.1037/a0027934

127. Meijer, R. R., & Sijtsma, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, *25*(2), 107–135. https://doi.org/10.1177/01466210122031957

128. Miche, Y., & Lendasse, A. (2009). A Faster Model Selection Criterion for OP-ELM and OP-KNN: Hannan-Quinn Criterion. U M. Verleysen (Ur.), *Proceedings of the 17th European Symposium on Artificial Neural Networks: Advances in Computational Intelligence and Learning (ESANN 2009). Bruges, Belgium* (str. 177–182). d-side publications.

129. Millman, J., & Arter, J. A. (1984). Issues in Item Banking. *Journal of Educational Measurement*, *21*(4), 315–330.

130. Mislevy, J. L., Rupp, A. A., & Harring, J. R. (2012). Detecting Local Item Dependence in Polytomous Adaptive Data. *Journal of Educational Measurement*, *49*(2), 127–147. https://doi.org/10.1111/j.1745-3984.2012.00165.x

131. Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis: With Applications in Political Research*. Mouton.

132. Moulton, M. H. (2003). *Rasch Estimation Demonstration Spreadsheet*. Rasch Estimation Demonstration Spreadsheet. https://www.rasch.org/moulton.htm

133. Muraki, E. (1990). Fitting a Polytomous Item Response Model to Likert-Type Data. *Applied Psychological Measurement*, *14*(1), 59–71. https://doi.org/10.1177/014662169001400106

134. Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. https://doi.org/10.1177/014662169201600206

135. Muraki, E., & Carlson, J. E. (1995). Full-Information Factor Analysis for Polytomous Item Responses. *Applied Psychological Measurement*, *19*(1), 73–90. https://doi.org/10.1177/014662169501900109

136. Muthén, B. (1983). Latent Variable Structural Equation Modeling with Categorical Data. *Journal of Econometrics*, *22*(1–2), 43–65. https://doi.org/10.1016/0304-4076(83)90093-3

137. Muthén, B. (1984). A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika*, *49*(1), 115–132.

138. Nagelkerke, N. J. D. (1991). Note on a General Definition of the Coefficient of Determination. *Biometrika*, *78*(3), 691–692.

139. Navarro-Gonzalez, D., & Lorenzo-Seva, U. (2020). *EFA.MRFA: Dimensionality Assessment Using Minimum Rank Factor Analysis*. https://CRAN.R-project.org/package=EFA.MRFA

140. Nydick, S. (2022). *catIrt: Simulate IRT-Based Computerized Adaptive Tests*. https://CRAN.R-project.org/package=catIrt

141. Orlando, M., & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, *24*(1), 50–64. https://doi.org/10.1177/01466216000241003

142. Ostini, R., & Nering, M. L. (2006). *Polytomous Item Response Theory Models*. Sage Publications.

143. Partchev, I., & Maris, G. (2022). *irtoys: A Collection of Functions Related to Item Response Theory (IRT)*. https://CRAN.R-project.org/package=irtoys

144. Posit team. (2024). *RStudio: Integrated Development Environment for R*. Posit Software, PBC.

http://www.posit.co/

145. R Core Team. (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* [Software]. https://www.R-project.org/

146. Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111. https://doi.org/10.2307/271063

147. Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Danish Institute for Educational Research.

148. Reckase, M. D. (2009). *Multidimensional Item Response Theory*. Springer New York. https://doi.org/10.1007/978-0-387-89976-3

149. Reif, M., & Steinfeld, J. (2021). *PP: Estimation of Person Parameters for the 1,2,3,4-PL Model and the GPCM*. https://github.com/jansteinfeld/PP

150. Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*, *114*(3), 552–566.

151. Reise, S. P., & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, *27*(2), 133–144. https://doi.org/10.1111/j.1745-3984.1990.tb00738.x

152. Revelle, W. (2017). *psych: Procedures for Personality and Psychological Research*. https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research

153. Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research* [Software]. Northwestern University, Evanston, Illinois. https://CRAN.R-project.org/package=psych

154. Revelle, W., & Rocklin, T. (1979). Very Simple Structure: An Alternative Procedure For Estimating The Optimal Number Of Interpretable Factors. *Multivariate Behavioral Research*, *14*(4), 403–414. https://doi.org/10.1207/s15327906mbr1404_2

155. Richardson, M. W. (1936). The Relation Between the Difficulty and the Differential Validity of a

Test. *Psychometrika*, *1*(2), 33–49.

156. Rinker, T. W., & Kurkiewicz, D. (2018). *pacman: Package Management for R*. http://github.com/trinker/pacman

157. Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*(5). https://doi.org/10.18637/jss.v017.i05

158. Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test Analysis Modules*. https://CRAN.R-project.org/package=TAM

159. Robitzsch, A., & Steinfeld, J. (2022). *immer: Item Response Models for Multiple Ratings*. https://CRAN.R-project.org/package=immer

160. Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

161. Rost, J. (1990). Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, *14*(3), 271–282.

162. Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika Monograph Supplement*, *34*(4, Pt. 2), 100–100.

163. Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 461–464.

164. Sclove, S. L. (1987). Application of Model-Selection Criteria to Some Problems in Multivariate Analysis. *Psychometrika*, *52*, 333–343.

165. Shaw, F. (1991). Descriptive IRT vs. Prescriptive Rasch. *Rasch Measurement Transactions*, *5:1*, 131.

166. Sijtsma, K., & van der Ark, L. A. (2017). A Tutorial on How to Do a Mokken Scale Analysis on Your Test and Questionnaire Data. *British Journal of Mathematical and Statistical Psychology*, *70*(1), 137–158. https://doi.org/10.1111/bmsp.12078

167. Starkweather, J. (2014). Identifying or Verifying the Number of Factors to Extract using Very Simple Structure. *Benchmarks RSS Matters*, *December*.

168. Stone, C. A. (2000). Monte Carlo Based Null Distribution for an Alternative Goodness-of-Fit Test Statistic in Irt Models. *Journal of educational measurement*, *37*(1), 58–75.

169. Stone, C. A., & Zhang, B. (2003). Assessing Goodness of Fit of Item Response Theory Models: A Comparison of Traditional and Alternative Procedures. *Journal of Educational Measurement*, *40*(4), 331–352. https://doi.org/10.1111/j.1745-3984.2003.tb01150.x

170. Sugiura, N. (1978). Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections: Further Analysis of the Data by Akaike's. *Communications in Statistics-theory and Methods*, *7*(1), 13–26.

171. Sympson, J. B. (1978). A Model for Testing with Multidimensional Items. *Proceedings of the 1977 computerized adaptive testing conference*, *14*.

172. Takane, Y., & De Leeuw, J. (1987). On the Relationship Between Item Response Theory and Factor Analysis of Discretized Variables. *Psychometrika*, *52*(3), 393–408. https://doi.org/10.1007/BF02294363

173. The jamovi project. (2024). *Jamovi* (Verzija 2.5) [Software]. https://www.jamovi.org

174. Thissen, D., & Steinberg, L. (1986). A Taxonomy of Item Response Models. *Psychometrika*, *51*(4), 567–577. https://doi.org/10.1007/BF02295596

175. Thissen, D., & Steinberg, L. (1988). Data Analysis Using Item Response Theory. *Psychological Bulletin*, *104*(3), 385–395. https://doi.org/10.1037/0033-2909.104.3.385

176. Thissen, D., & Steinberg, L. (2020). An Intellectual History of Parametric Item Response Theory Models in the Twentieth Century. *Chinese/English Journal of Educational Measurement and Evaluation*, *1*(1). https://doi.org/10.59863/GPML7603

177. Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. U P. W. Holland & H. Wainer (Ur.), *Differential item functioning* (str. 67–113). Erlbaum.

178. Thissen, D., & Wainer, H. (2001). *Test Scoring*. Routledge.

179. Thompson, N. A. (2009). *Ability Estimation with Item Response Theory: White Paper*. Assessment Systems Corporation.

180. Thurstone, L. L. (1927). A Law of Comparative Judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

181. Tucker, L. R. (1946). Maximum Validity of a Test with Equivalent Items. *Psychometrika*, *11*(1), 1–13. https://doi.org/10.1007/BF02288894

182. Van der Ark, L. A. (2007). Mokken Scale Analysis in R. *Journal of Statistical Software*, *20*(11), 1–19.

183. Van Der Linden, W. J., & Hambleton, R. K. (1997). Item Response Theory: Brief History, Common Models, and Extensions. U *Handbook of modern item response theory* (str. 1–28). Springer.

184. Velicer, W. F. (1976). Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika*, *41*, 321–327.

185. Vujanić, M., Nikolić, M., & Matica srpska (Novi Sad, Serbia) (Ur.). (2011). *Rečnik srpskoga jezika* (Izmenjeno i popravljeno izdanje). Matica srpska.

186. Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet Response Theory and Its Applications*. Cambridge University Press. http://public.eblib.com/choice/publicfullrecord.aspx?p=293364

187. Whitely, S. E. (1980). Multicomponent Latent Trait Models for Ability Tests. *Psychometrika*, *45*(4), 479–494. https://doi.org/10.1007/BF02293610

188. Wind, S. A. (2017). An Instructional Module on Mokken Scale Analysis. *Educational Measurement: Issues and Practice*, *36*(2), 50–66. https://doi.org/10.1111/emip.12153

189. Woods, C. M. (2009). Empirical Selection of Anchors for Tests of Differential Item Functioning. *Applied Psychological Measurement*, *33*(1), 42–57.

190. Wright, B. D. (1998). Model Selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? *Rasch Measurement Transactions*, *12*(3), 641–642.

191. Wright, B. D., & Linacre, J. M. (1994). Reasonable Mean-Square Fit Values. *Rasch Measurement Transactions*, *8*(3), 370–371.

192. Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Mesa Press.

193. Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational Measurement for Applied Researchers*. Springer Singapore. https://doi.org/10.1007/978-981-10-3302-5

194. Yao, L., & Schwarz, R. D. (2006). A Multidimensional Partial Credit Model With Associated Item and Test Statistics: An Application to Mixed-Format Tests. *Applied Psychological Measurement*,

*30*(6), 469–492. https://doi.org/10.1177/0146621605284537

195. Yen, W. M. (1981). Using Simulation Results to Choose a Latent Trait Model. *Applied Psychological Measurement*, *5*(2), 245–262. https://doi.org/10.1177/014662168100500212

196. Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

197. Zeileis, A., Bivand, R., Eddelbuettel, D., Hornik, K., & Vialaneix, N. (2023). CRAN Task Views: The Next Generation. *arXiv preprint arXiv:2305.17573*.

198. Zhang, Y., Wang, D., Gao, X., Cai, Y., & Tu, D. (2019). Development of a Computerized Adaptive Testing for Internet Addiction. *Frontiers in Psychology*, *10*, 1010. https://doi.org/10.3389/fpsyg.2019.01010

199. Zopluoglu, C. (2022). *EstCRM: Calibrating Parameters for the Samejima's Continuous IRT Model*. https://CRAN.R-project.org/package=EstCRM

200. Zumbo, B. D. (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF). *Ottawa: National Defense Headquarters*.

# 12 Indeks pojmova

## J



## K



## L



## M



## N



## O



## P

## R



## S



## T



## U



## V

# 13 Indeks grafikona

UNIVERZITET U NOVOM SADU
FILOZOFSKI FAKULTET NOVI SAD
21000 Novi Sad
Dr Zorana Đinđića 2
www.ff.uns.ac.rs

Elektronsko izdanje
https://digitalna.ff.uns.ac.rs/sadrzaj/2024/978-86-6065-861-8

CIP - Каталогизација у публикацији
Библиотеке Матице српске, Нови Сад

159.9.072:004.432.2R

**ЈАНИЧИЋ, Бојан, 1968-**
Jednodimenzionalni IRT modeli sa primenom u R-u [Elektronski izvor] /
Bojan Janičić. - Novi Sad : Filozofski fakultet, 2024

Način pristupa (URL): https://digitalna.ff.uns.ac.rs/sadrzaj/2024/978-86-6065-861-8. - Opis zasnovan na stanju na dan 11.9.2024. - Nasl. sa naslovnog ekrana. - Bibliografija. - Registri.

ISBN 978-86-6065-861-8

a) Статистички модели -- ИРТ анализа -- Програмски језик "R"

COBISS.SR-ID 151839753