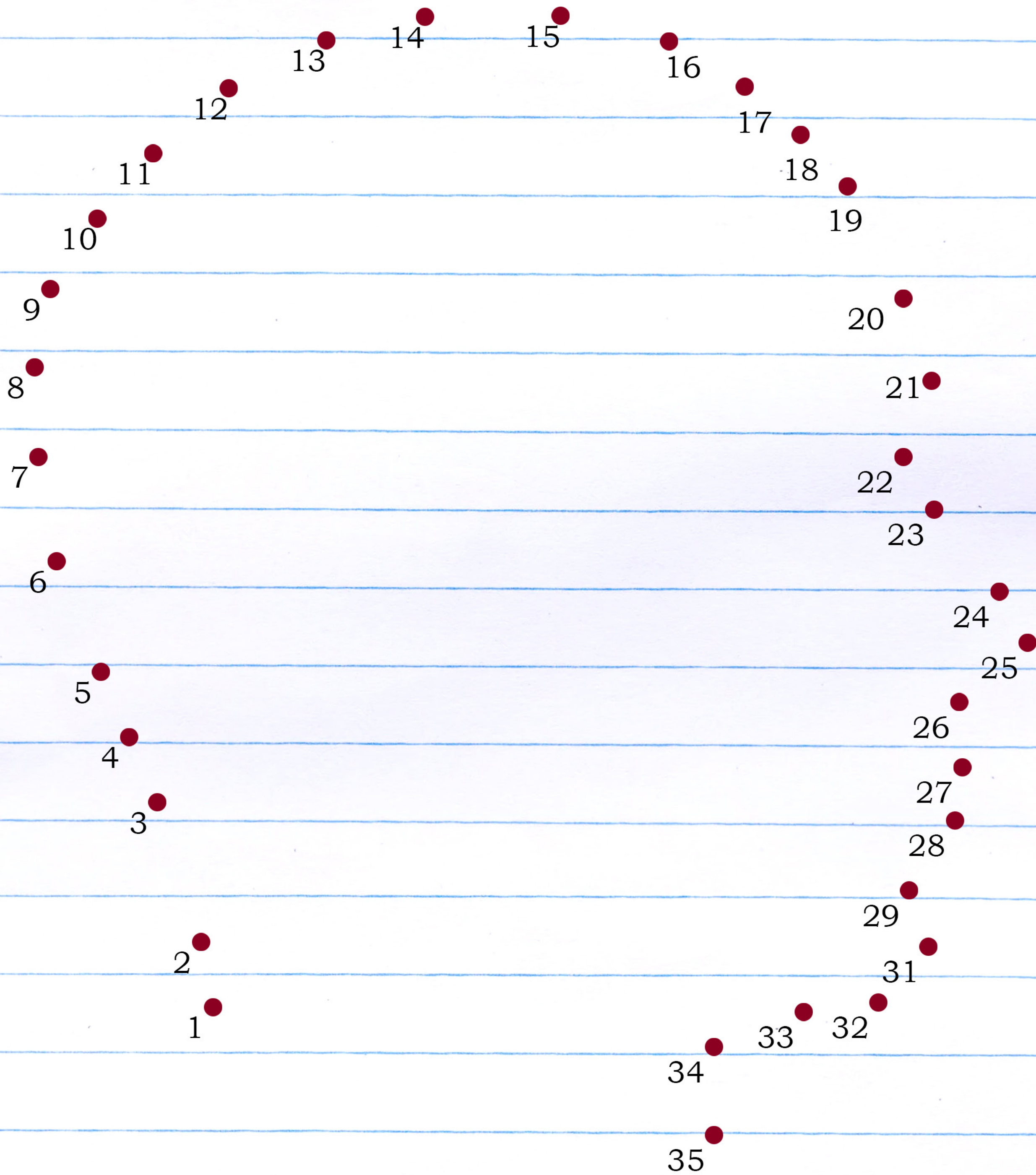
Danka Purić

Interne metrijske karakteristike
psiholoških instrumenata

Vodič kroz Psihometriju 1



1838
УНИВЕРЗИТЕТ У БЕОГРАДУ
ФИЛОЗОФСКИ ФАКУЛТЕТ

Institut za psihologiju
Filozofski fakultet, Beograd

Danka Purić

Interne metrijske karakteristike psiholoških instrumenata

Vodič kroz Psihometriju 1

Institut za psihologiju
Beograd, 2023.

Danka Purić
INTERNE METRIJSKE KARAKTERISTIKE PSIHOLOŠKIH INSTRUMENATA. VODIČ KROZ PSIHOMETRIJU 1

Izdavač:
Institut za psihologiju
Filozofski fakultet, Univerzitet u Beogradu
Čika Ljubina 18-20, Beograd, Srbija

Za izdavača:
dr Zora Krnjaić

Recenzenti:
dr Goran Opačič, vanredni profesor, Filozofski fakultet, Univerzitet u Beogradu
dr Predrag Teovanović, vanredni profesor, Fakultet za specijalnu edukaciju i rehabilitaciju, Univerzitet u Beogradu
dr Bojana Dinić, vanredna profesorka, Filozofski fakultet, Univerzitet u Novom Sadu

Lektura i korektura:
dr Vesna Đorđević, Institut za srpski jezik SANU

Ilustracija na koricama:
Jakša Lakićević

Štampa:
Institut za psihologiju
Filozofski fakultet, Univerzitet u Beogradu, Beograd
Tiraž: 10

ISBN: 978-86-6427-202-5

Beograd, 2023.

Preporučeno citiranje: Purić, D. (2023). *Interne metrijske karakteristike psiholoških instrumenata. Vodič kroz Psihometriju 1.* Beograd: Institut za psihologiju.

SADRŽAJ

REČ AUTORKE

Dragi čitaoci,

Knjiga pred vama nastala je sa idejom da studentima psihologije na Filozofskom fakultetu Univerziteta u Beogradu olakša polaganje ispita iz predmeta Psihometrija 1. Naravno, ako se pokaže da je ona interesantna i korisna i drugima – biću veoma radosna. Ona ne pretenduje da bude obuhvatni pregled svih relevantnih oblasti iz psihometrije, pa čak ni iz onog njenog dela usmerenog na interne metrijske karakteristike psiholoških instrumenata, ali nastoji da oblasti koje pokriva prikaže na jednostavan i razumljiv način. Želja mi je bila da ovu meni izuzetno zanimljivu oblast približim i onima koji o njoj i dalje ne znaju mnogo ili je se malo i plaše. Iz mog dosadašnjeg iskustva, ako joj se pristupi bez predrasuda, psihometrija ne samo što nije teška već, uz razumevanje, čak i lakše ulazi u korpus trajno stečenih znanja iz psihologije od mnogih drugih oblasti.

Ovaj udžbenik svojim najvećim delom pokriva gradivo kursa Psihometrija 1, odnosno oslanja se na kurikulum koji je moj dragi kolega i prijatelj, profesor Goran Opačić, godinama razvijao (još pre nego što sam počela da studiram, a kasnije i predajem). Gorane, hvala ti puno na nesebičnom deljenju svih resursa, konsultacijama i savetima od početka moje akademske karijere. Bez ogromnog posla koji si ti uradio a ja se na njega nadovezala, ovaj udžbenik bi sigurno izgledao potpuno drugačije.

Želim iskreno da se zahvalim i profesoru Goranu Kneževiću – tvoje mentorstvo, praktično od početka studija do danas, bilo mi je nezamenljivo, a tvoja strast prema istraživačkom radu me je inspirisala da se i sama posvetim ovoj profesiji.

Ipak, čini mi se da najveću zahvalnost za nastanak ove knjige dugujem svojim studentima. Vaša pitanja i komentari, ali i klimanje glavom i zbunjeni pogledi, omogućili su mi da bolje razumem šta je to u psihometriji lakše, a šta teže intuitivno razumljivo, kojim pitanjima treba pristupiti iz kojih uglova i kako određene teme predstaviti jednostavno, ali ipak ne ostati samo na površini. Takođe, vi ste glavni razlog zbog kog sam i odlučila da napišem ovaj udžbenik i ono što mi je održalo motivaciju tokom čitavog procesa njegovog nastajanja. Nadam se da će vam knjiga koristiti, ne samo dok spremate ispit već i nakon toga, a uverena sam da će me vaše povratne informacije uvek podsticati da se dodatno usavršavam.

Danka

PREDGOVOR

Knjiga *Interne metrijske karakteristike psiholoških instrumenata: Vodič kroz Psihometriju 1* kao svoje osnovne teme postavlja one psihometrijske karakteristike instrumenata koje se mogu utvrditi samo na osnovu rezultata na datom testu, bez njegovog dovođenja u vezu sa drugim konstruktima. U skladu sa tim, kao osnovne teme pojavljuju se diskriminativnost, objektivnost, pouzdanost, homogenost i reprezentativnost testa. Iako se i neki aspekti valjanosti testa (strukturna, odnosno faktorska valjanost) mogu utvrditi i bez uzimanja drugih konstrukata u obzir, dokazivanje valjanosti testa je višeslojan proces čiji različiti vidovi izlaze iz okvira unutrašnjih mernih svojstava instrumenata, zbog čega valjanost ovde nisam posmatrala kao internu metrijsku karakteristiku. Dodatno, ovaj vodič pokriva i različite modele merenja, kao što su Gutmanova teorija merenja i teorija odgovora na stavke, a koje se mogu koristiti komplementarno sa klasičnim pristupom prilikom utvrđivanja internih metrijskih karakteristika instrumenata. Na kraju, iako u procesu izrade instrumenta sledi nakon dokazivanja valjanosti, prikazan je i proces normiranja psiholoških instrumenata, jer je i ovo procedura za koju nisu potrebni dodatni podaci osim odgovora na sam test.

Vodič kroz Psihometriju 1 zamišljen je kao vodič i kroz teoriju i kroz praksu provere metrijskih karakteristika instrumenata. Svako poglavlje počinje kratkim uvodom u temu iz laičke perspektive, zatim se uvode i objašnjavaju osnovni teorijski pojmovi, da bi na kraju njihova primena bila ilustrovana primerima na konkretnim podacima. Iako formule nisu previše popularne među studentima psihologije, smatram da su one ipak veoma korisne za dublje razumevanje mnogih psihometrijskih koncepata, jer ekonomično prikazuju odnose između varijabli na koje se odnose, te nisam želela da bežim od njihove upotrebe. Ipak, kako tekst sadrži nemali broj formula, potrudila sam se da svaku formulu detaljno objasnim i prevedem na „svakodnevni” jezik i/ili istaknem ono što je u formuli ključno, kako bi njena suština mogla lako da se razume i zapamti. Potrudila sam se i da sve psihometrijske koncepte ilustrujem jednostavnim, razumljivim primerima, kako ne bi ostali previše apstraktni, već kako bi studenti uvideli njihovu primenu u redovnim situacijama psihološkog testiranja. Za razumevanje teksta potrebno je izvesno bazično statističko predznanje, ali većina studenata druge godine psihologije već ima u svom korpusu znanja sve statističke pojmove neophodne za praćenje izlaganja.

Svako poglavlje takođe sadrži i detaljne instrukcije kako se opisani postupci provere metrijskih karakteristika sprovode u IBM© SPSS© statističkom softveru, odnosno, u slučaju teorije odgovora na stavke, u programu Winsteps/Ministep. Takođe, postoji posebno, dodatno poglavlje u kome je detaljno objašnjen rad u IBM SPSS sintaksi, kao i poglavlje koje opisuje proveru metrijskih karakteristika instrumenta korišćenjem makroa RTT10G, koji se pokreće preko SPSS sintakse. Sve ove instrukcije sam uključila kako bih studentima olakšala samostalni rad na podacima koji se koriste na vežbama na kursu Psihometrija 1, ali pre svega rad na sopstvenim podacima, bilo u okviru kursa, bilo van njega.

Nadam se da sam uspela u svojoj nameri da gradivo Psihometrije 1 učinim pristupačnim i jednostavno razumljivim svima koji će se jednog dana baviti psihologijom, pa čak i onima čije se oblasti interesovanja ne preklapaju previše sa oblašću merenja. Takođe se nadam da će znanja proistekla iz čitanja ove knjige biti nešto što studenti mogu i kasnije (i nakon završetka studija) da primene u svojoj praksi – bilo na nivou kritičke evaluacije postojećih testova i publikacija koje govore o testovima i njihovoj upotrebi, bilo na nivou praktične upotrebe na podacima dobijenim testiranjem u različitim oblastima psihologije.

1) DISKRIMINATIVNOST

Iz svakodnevnog iskustva poznato nam je da se ljudi međusobno razlikuju po brojnim osobinama. Neko je duhovitiji, dok neko ima manje razvijen smisao za humor; neko ima veliku tremu pred javni nastup, neko drugi u njemu uživa, a neko treći oseća samo blago uzbuđenje. Diskriminativnost testa ili ajtema je njihovo svojstvo da „uhvate" ove individualne razlike, odnosno osobina testa/ajtema da razlikuje ispitanike prema stepenu, tj. nivou izraženosti merene osobine (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009).

1.1) Koncepti diskriminativnosti

Ako razmišljamo o diskriminativnosti kao o metrijskoj karakteristici testa (ili ajtema), obično želimo da stepen u kom je test (ili ajtem) diskriminativan na neki način brojčano izrazimo, odnosno kvantifikujemo. Postoji nekoliko načina za konceptualizaciju, a posledično i za numeričko izražavanje diskriminativnosti, a koji se donekle međusobno preklapaju. Pored diskriminativnosti, koristi se i termin *osetljivost* jer test koji dobro diskriminiše ispitanike ujedno je i osetljiv na razlike između njih.

Jedno od viđenja diskriminativnosti jeste da je test utoliko više diskriminativan ukoliko može da registruje više različitih skorova. Na prvi pogled, može delovati da će diskriminativnost testa onda zavisiti samo od broja stavki koje test ima, po principu – što više stavki, to više skorova, te i veća diskriminativnost. Međutim, to što teorijski postoji veliki broj skorova na testu ne znači nužno da će u uzorku biti ispitanika koji postižu sve skorove. Na primer, ako je test lak, može se desiti da niko ne postigne najniži skor ili ako je test veoma težak, da nekoliko najviših skorova ne dosegne niko od ispitanika. U takvim situacijama, diskriminativnost neće odgovarati broju skorova na testu.

Mera diskriminativnosti koja ukazuje na stepen u kom je „diskriminativni potencijal" instrumenta zaista i iskorišćen je Fergusonov delta koeficijent (Ferguson, 1949), koji se definiše kao količnik empirijski ostvarenog broja razlikovanja između skorova na datom instrumentu i teorijski maksimalnog mogućeg broja razlikovanja:

$$\delta = \frac{n^2 - \sum_{i=0}^{k} f_i^2}{N^2 - \frac{N^2}{k+1}}$$

gde je k broj ajtema, N broj ispitanika, a f_i frekvenca pojedinačnog skora. Kao što vidimo, na vrednost maksimalnog mogućeg broja razlikovanja (izraz ispod razlomačke crte) utiču samo broj stavki i broj ispitanika, dok na vrednost empirijski ostvarenog razlikovanja (izraz iznad razlomačke crte) utiču broj ispitanika i raspodela njihovih skorova u postojeće kategorije[1]. Izvorno, delta koeficijent je definisan za dihotomne stavke, a može se izračunati i korišćenjem pojednostavljene (izvedene) formule:

[1] Posredno, na broj ostvarenih razlikovanja utiče i broj kategorija tako što određuje maksimalan broj kategorija u koje se ispitanici mogu svrstati, iako na podacima dati broj može biti i manji.

$$\delta = \frac{(k+1)(N^2 - \sum_i f_i^2)}{kN^2}$$

gde je k broj ajtema, N broj ispitanika, a f_i frekvenca pojedinačnog skora. Prikazana formula može se generalizovati i na stavke sa više ponuđenih odgovora (Hankins, 2007), u kom slučaju primenjujemo sledeći, opštiji obrazac:

$$\delta = \frac{(1 + k(m-1))(N^2 - \sum_i f_i^2)}{k(m-1)N^2}$$

gde je m broj ponuđenih odgovora na stavkama. Delta koeficijent se kreće od vrednosti 0, koja odgovara situaciji gde ne postoje ostvarena razlikovanja skorova, do vrednosti 1, koja odgovara slučaju u kom su sva teorijski moguća razlikovanja između skorova i ostvarena. Drugim rečima, delta koeficijent ukazuje na proporciju (procenat) ostvarenih razlikovanja u odnosu na sva moguća razlikovanja koja bi test trebalo da omogući. Ferguson je pokazao da za uniformnu distribuciju delta koeficijent iznosi 1, za normalnu oko .90, a da je za asimetrične distribucije vrednost još niža (Hankins, 2007).

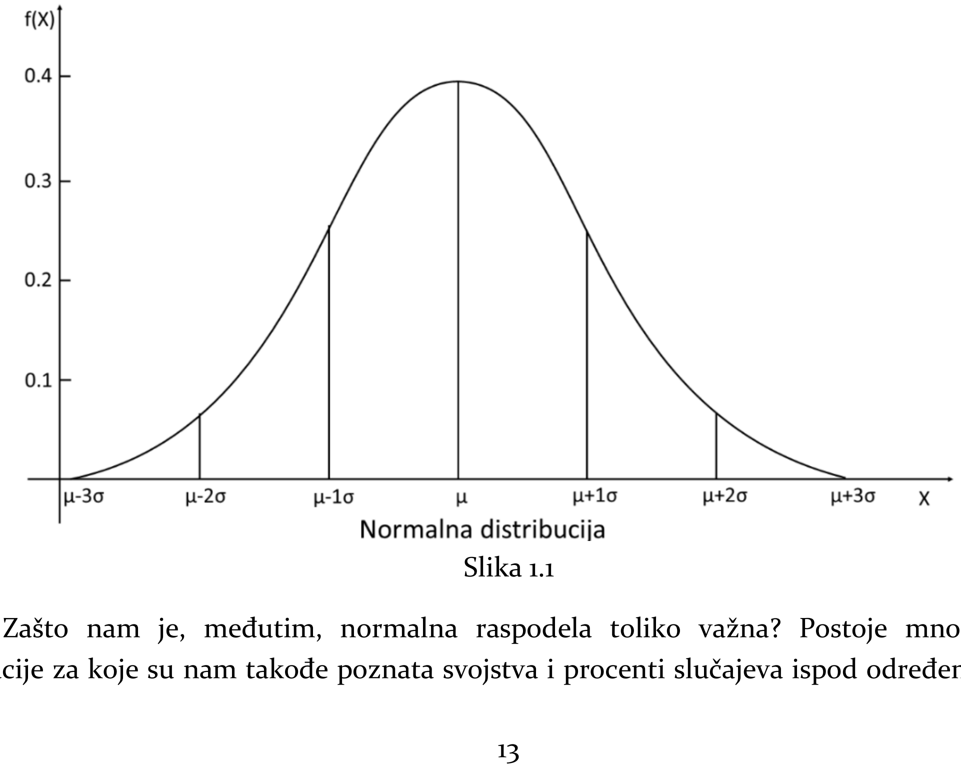
Fergusonovom delta koeficijentu se, međutim, mogu uputiti i izvesne zamerke (npr., u slučaju uniformne raspodele delta je uvek 1, bez obzira na to da li instrument ima 2 ili 20 stavki; kategorije odgovora se uvek tretiraju kao nominalne/kategoričke, te se ne koriste dodatni podaci koje pružaju ordinalne skale; zanemaruje se pouzdanost pa u ostvarene razlike ulaze i one nastale na osnovu greške, a ne samo pravog skora ispitanika), te njegova primena nije univerzalno preporučljiva (Terluin et al., 2009).

Drugi koncept diskriminativnosti odgovara viđenju diskriminativnosti kao maksimalne varijanse. Logika iza ovog koncepta je da maksimalna varijansa ukazuje na maksimalno razlikovanje skorova na testu, a minimalna varijansa na minimalno razlikovanje. Možemo onda postaviti pitanje kakva distribucija skorova nam daje minimalnu, a kakva maksimalnu varijansu. U slučaju minimalne varijanse, odgovor je jednostavan – u pitanju je distribucija kod koje svi ispitanici postižu isti skor, odnosno kada umesto varijable zapravo imamo konstantu, te uopšte ne možemo razlikovati ispitanike po merenoj osobini. A kakva distribucija proizvodi maksimalnu varijansu? Odgovor na ovo pitanje zavisi od graničnih uslova koje postavimo, ali načelno najveću varijansu testa dobijamo ukoliko je distribucija uniformna, tj. ukoliko su svi skorovi podjednako frekventni, odnosno podjednako verovatni (Dekking et al., 2005). Tabela 1.1 prikazuje razliku u varijansi skorova za normalnu i uniformnu distribuciju u situaciji kada je ukupan broj kategorija, kao i ispitanika, jednak za dve distribucije. Ipak, treba imati u vidu da uniformna distribucija daje maksimalnu entropiju („količinu neizvesnosti”) u slučaju konačnog, ograničenog broja skorova, dok u slučaju neograničenog broja vrednosti, odnosno skorova, normalna distribucija ima maksimalnu entropiju za poznatu aritmetičku sredinu i standardnu devijaciju (Park & Bera, 2009).

Time dolazimo i do trećeg, poslednjeg shvatanja diskriminativnosti koje ćemo ovde prikazati. U pitanju je ocenjivanje diskriminativnosti konkretnog testa poređenjem njegove raspodele skorova sa normalnom raspodelom, sa idejom da što je veće odstupanje dobijene raspodele od normalne, to je diskriminativnost lošija. Normalna raspodela (normalna distribucija, normalna kriva, zvonasta kriva ili Gausova kriva) jeste teorijska raspodela frekvenci skorova na nekom testu gde je za svaki skor tačno poznata frekvenca sa kojom se on javlja (R. J. Cohen & Swerdlik, 2018). Prosečni skor je najfrekventniji, a kako se udaljavamo od proseka tako skorovi postaju sve manje i manje frekventni. Za normalnu raspodelu poznate su i površine ispod krive, te tako znamo da aritmetička sredina preseca normalnu raspodelu tačno na polovini – 50% skorova nalazi se ispod, a 50% iznad aritmetičke sredine (drugim rečima, aritmetička sredina i medijana se za normalnu raspodelu poklapaju – imaju istu vrednost), kao i da se između -1 i +1 standardne devijacije nalazi 68% slučajeva, 95.45% slučajeva nalazi se između -2 i +2 SD, dok se unutar -3 i +3 SD nalazi 99.73% skorova ispitanika (Slika 1.1).

Tabela 1.1. Varijansa za normalnu i uniformnu raspodelu sa istim brojem kategorija (k = 14) i ispitanika (N = 1500)

	Normalna raspodela		**Uniformna raspodela**	
	$\mathbf{f_n}$	$\mathbf{f_n^2}$	$\mathbf{f_u}$	$\mathbf{f_u^2}$
0			100	10000
1	2	4	100	10000
2	7	49	100	10000
3	25	625	100	10000
4	66	4356	100	10000
5	138	19044	100	10000
6	225	50625	100	10000
7	287	82369	100	10000
8	287	82369	100	10000
9	225	50625	100	10000
10	138	19044	100	10000
11	66	4356	100	10000
12	25	625	100	10000
13	7	49	100	10000
14	2	4	100	10000
Σ	1500	314144	1500	150000
SD^2		0,92		1,00



Slika 1.1

Zašto nam je, međutim, normalna raspodela toliko važna? Postoje mnoge teorijske distribucije za koje su nam takođe poznata svojstva i procenti slučajeva ispod određenih odsečaka krive. Razlog se krije u prirodi konstrukata kojima se bavimo u psihologiji. U slučaju većine psiholoških mernih instrumenata možemo očekivati da je hipotetički konstrukt koji pokušavamo da izmerimo normalno raspodeljen u populaciji. Kao primeri mogu poslužiti konstrukti inteligencije (Thorndike, 1973) ili bazičnih crta ličnosti kao što su neuroticizam, ekstraverzija/introverzija, saradljivost (Jokić & Purić, 2019; Knežević et al., 2004).

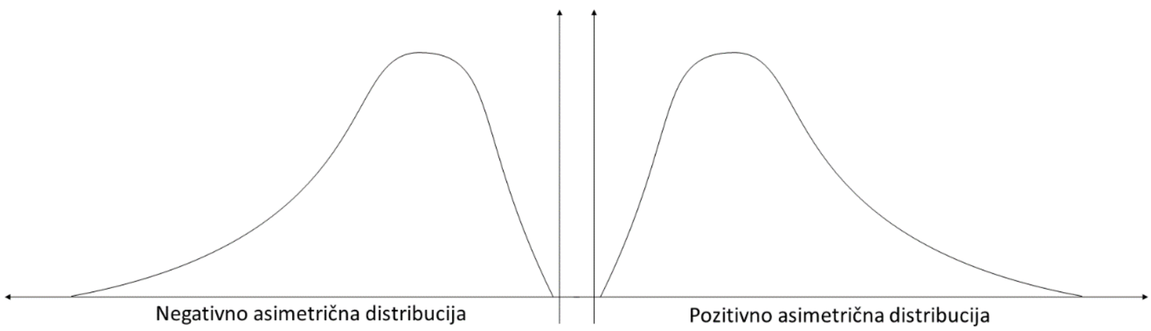
S obzirom na to da ovde navedeni modeli procene diskriminativnosti nisu u potpunosti komplementarni, postavlja se pitanje koji pristup je bolje primeniti. U većini slučajeva, diskriminativnost ćemo proveravati tako što ćemo empirijsku distribuciju testa porediti sa normalnom distribucijom. Ovo ne znači da nam varijansa instrumenta nije važna – u nekim slučajevima (npr., kada razmatramo diskriminativnost pojedinačnih stavki) varijansa može biti i informativnija od raspodele. Takođe, u slučaju određenih, specifičnih konstrukata moguće je da ćemo teorijski očekivati neki drugi oblik distribucije, a ne normalnu raspodelu. Ipak, iako savršeno normalna raspodela predstavlja idealni slučaj, empirijske raspodele često veoma nalikuju normalnoj distribuciji, posebno ukoliko je uzorak dovoljno veliki, te se ona koristi kao standard za poređenje prilikom procene diskriminativnosti (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018).

1.2) Odstupanja od normalne distribucije

Kada kažemo da ćemo proveravati diskriminativnost poređenjem empirijske distribucije sa normalnom tu pre svega mislimo na utvrđivanje da li je distribucija (približno) normalna ili postoje određena odstupanja od normalnosti koja su dovoljno velika da ih možemo smatrati značajnim. Raspodela može odstupati od normalne po dve ose – horizontalnoj, kada govorimo o asimetriji, i vertikalnoj, kada govorimo o spljoštenosti / izduženosti distribucije.

1.2.1) Odstupanja po horizontalnoj osi

Odstupanja po horizontalnoj osi odnose se na simetričnost, odnosno asimetričnost distribucije. Postoje dva vida asimetrije – pozitivno i negativno asimetrična distribucija (Slika 1.2).



Slika 1.2

Kod **pozitivno asimetrične distribucije** većina skorova se grupiše oko niskih vrednosti, dok su ekstremne vrednosti na pozitivnom polu. Može se reći i da je desni „rep" distribucije duži od levog. S obzirom na to da znamo da je aritmetička sredina osetljivija na ekstremne skorove od medijane kao mere centralne tendencije, kada su skorovi većine ispitanika niski, medijana će obično biti niža od aritmetičke sredine. Drugim rečima, kod pozitivno asimetrične distribucije aritmetička sredina će biti pomerena udesno, ka pozitivnom polu.

U kojoj situaciji možemo očekivati pozitivno asimetričnu distribuciju, odnosno kako možemo interpretirati ovo odstupanje od normalnosti? Ukoliko se radi o testu znanja ili sposobnosti, onda se može reći da je test isuviše težak za dati uzorak ispitanika (te su ga oni generalno loše uradili). Ukoliko se pak radi o testu ličnosti, onda je ispitivana osobina kod ispitanika prisutna u maloj meri. Neke od varijabli relevantnih u psihologiji, a koje tipično imaju pozitivno aismetričnu distribuciju su, na primer, vreme reakcije (Moscoso Del Prado Martin, 2009) ili mesečna primanja (Lebergott, 1959). Jedan od razloga za ovaj oblik distribucije je što za obe mere postoji apsolutna nula, koja je ujedno teorijski minimum (vreme reakcije mora biti pozitivna vrednost, kao i primanja). Samim tim, ekstremne vrednosti mogu se javiti samo na pozitivnom polu distribucije.

Kao što je iz primera verovatno jasno, kod pozitivno asimetrične distribucije imamo sniženu diskriminativnost na negativnom polu (teško razlikujemo ispitanike kada veliki broj njih ima niske skorove), dok je diskriminativnost bolja na pozitivnom polu (ukoliko većina ima niske skorove, visok skor je diskriminativniji, odnosno informativniji). Ovo odgovara i samom izgledu distribucije koja je na negativnom polu zbijenija nego što je to slučaj kod normalne raspodele, dok je na pozitivnom polu razvučenija.

Nasuprot ovome, **negativno asimetričnu distribuciju** karakteriše grupisanje skorova na pozitivnom polu distribucije, dok se ekstremne vrednosti nalaze na negativnom polu distribucije, odnosno levi „rep" distribucije je duži od desnog. Pošto su skorovi grupisani oko viših vrednosti, aritmetička sredina biće (obično) niža u poređenju sa medijanom, odnosno pomerena ulevo.

U slučaju testova znanja i sposobnosti, negativno asimetrična distribucija se interpretira kao pokazatelj da je test isuviše lak za dati uzorak ispitanika. Na primer, ukoliko bismo članovima Mense zadali test inteligencije namenjen opštoj populaciji, mogli bismo da očekujemo negativno asimetričnu distribuciju. U slučaju testova ličnosti, negativno asimetrična kriva ukazuje na osobinu koja je u većoj meri prisutna kod skoro svih članova uzorka, što se ponekad može primetiti u slučaju socijalno poželjnih osobina, poput altruizma, kada ispitanici (iz različitih razloga) teže da se prikažu u što boljem svetlu.

Diskriminativnost je kod negativno asimetrične distribucije snižena za visoke skorove, dok je dobra za niske skorove. I ovde sniženje diskriminativnosti prati oblik distribucije, te sniženu diskriminativnost imamo na desnom kraju distribucije, gde su skorovi zbijeniji nego što bi bili u slučaju savršeno normalne distribucije.

Mera asimetričnosti distribucije je skjunis (eng. *skewness*, Sk). Skjunis se tipično definiše kao treći standardizovani momentni koeficijent distribucije i računa se preko sledeće formule:

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$

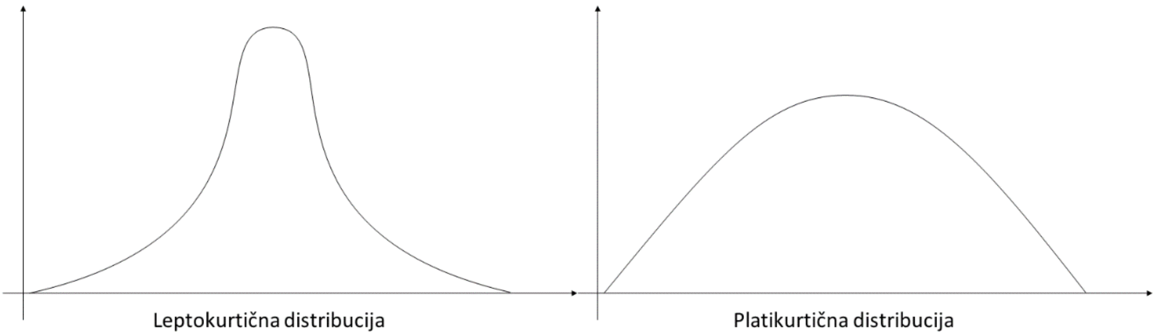
gde je E operator očekivane vrednosti (koji je za ove potrebe dovoljno razumeti kao prosečnu vrednost dobijenu na velikom broju slučajeva), X sirovi skor, μ aritmetička sredina, σ standardna devijacija, a μ_3 treći centralni moment distribucije[2] (Blanca et al., 2013; Groeneveld & Meeden, 1984; Joanes & Gill, 1998).

[2] Više o centralnim momentima može se pročitati u Tenjovićevom priručniku *Statistika u psihologiji* (Tenjović, 2000). Standardni momenti su centralizovani momenti podeljeni standardnom devijacijom (ili nekim njenim stepenom). Ovaj postupak za cilj ima standardizaciju vrednosti, odnosno čini da vrednost momenta ne zavisi od merne skale koja je upotrebljena. Slično kao što centralizovani skor ukazuje samo na odstupanje sirovog skora od aritmetičke sredine, a standardni skor ukazuje na odstupanje sirovog skora od aritmetičke sredine izraženo u jedinicama standardne devijacije, tako i

Za normalnu distribuciju vrednost Sk će biti nula, za pozitivno asimetričnu distribuciju će biti pozitivna, dok će za negativno asimetričnu distribuciju biti negativna. Ukoliko raspodela ne skreće ni u levo ni u desno, odnosno ukoliko nema (značajnih) odstupanja po horizontalnoj osi, odnosno ukoliko se aritmetička sredina i medijana poklapaju – možemo reći da je distribucija simetrična. Međutim, da bismo rekli da je ona i normalna, neophodno je da ne postoje (značajna) odstupanja od normalnosti ni po vertikalnoj osi.

1.2.2) Odstupanja po vertikalnoj osi

Odstupanja po vertikalnoj osi odnose se na situacije u kojima raspršenje skorova oko aritmetičke sredine ne odgovara normalnoj distribuciji. Možemo razlikovati spljoštenu, odnosno razvučenu distribuciju i izduženu, odnosno zbijenu distribuciju (Slika 1.3).



Slika 1.3

Izdužena distribucija naziva se **leptokurtična distribucija** i ona odgovara situaciji kada je veći broj slučajeva grupisan oko aritmetičke sredine nego što je to slučaj kod normalne distribucije, dok postoji manji broj ekstrema, slučajeva daleko od aritmetičke sredine. Suprotno ovome, kod razvučene ili **platikurtične distribucije** frekvenca slučajeva koji imaju prosečne ili skorove blizu prosečnih je manja, a veći broj slučajeva je udaljen od aritmetičke sredine. Kod leptokurtične distribucije varijabilitet je sužen u poređenju sa normalnom distribucijom, naročito u centralnom delu distribucije, odnosno ona je manje diskriminativna posebno za prosečne skorove ispitanika. Kod platikurtične distribucije varijabilitet je veći duž celog raspona, što ukazuje na dobru diskriminativnost za sve skorove, ali srazmerno nižu diskriminativnost skorova oko proseka u poređenju sa niskim i visokim skorovima. Povećana diskriminativnost platikurtične distribucije može delovati kao pozitivna osobina i, zaista, ne mora predstavljati problem u svim situacijama. Ipak, ako se setimo poređenja uniformne i normalne distribucije, jasno je da distribucija skorova na testu treba da odgovara pretpostavljenoj distribuciji konstrukta koji testom želimo da izmerimo, što je u najvećem broju slučajeva normalna distribucija.

Ukoliko je reč o testu sposobnosti, leptokurtična distribucija ukazuje na veći broj srednje teških zadataka, a manji broj lakih i teških, dok je kod platikurtične obrnuto – veći broj lakih i teških zadataka, a proporcionalno manji broj srednje teških. U slučaju nekognitivnih testova, platikurtična distribucija upućuje na veći broj ekstremnih tvrdnji koje ispitanici ili izrazito prihvataju ili izrazito odbacuju, dok kod leptokurtične distribucije imamo dosta tvrdnji sa kojima se većina ispitanika umereno slaže ili je neodlučna.

---

centralizovani moment pokazuje moment distribucije oko aritmetičke sredine, dok standardizovani moment ukazuje na standardizovano odstupanje distribucije od aritmetičke sredine.

Mera odstupanja distribucije od normalne po vertikalnoj osi naziva se kurtozis (eng. *kurtosis*, Ku) i tipično se izražava kao četvrti standardizovani momentni koeficijent distribucije, koji se može izračunati prema sledećoj formuli:

$$Kurt[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}$$

gde je E operator očekivane vrednosti, X sirovi skor, μ aritmetička sredina, σ standardna devijacija, a μ_4 četvrti centralni moment distribucije (Blanca et al., 2013; Groeneveld & Meeden, 1984; Joanes & Gill, 1998).

Kurtozis normalne raspodele iznosi 3, vrednosti niže od 3 ukazuju na platikurtičnu raspodelu, a vrednosti preko 3 ukazuju na leptokurtičnu raspodelu. Međutim, neki autori umesto kurtozisa kao meru izduženosti / spljoštenosti distribucije koriste tzv. korigovani kurtozis (eng. *excess kurtosis*), koji je jednak vrednosti kurtozisa umanjenog za tri. Ova vrednost se koristi i u programu SPSS, tako da normalnu ili mezokurtičnu raspodelu karakteriše kurtozis jednak nuli, pozitivan kurtozis ukazuje na leptokurtičnu, a negativan na platikurtičnu distribuciju. Ekstremno negativan kurtozis može ukazivati i na bimodalnu distribuciju.

Ukoliko distribucija ne odstupa značajno od normalne po vertikalnoj osi, možemo reći da je mezokurtična ili da ima raspon koji odgovara rasponu normalne raspodele. Ne zaboravite, da bismo rekli da je distribucija normalna, pored mezokurtičnosti, ona mora biti i simetrična.

1.2.3) Druga odstupanja od normalnosti

U određenim slučajevima mogu se primetiti i drugačija odstupanja od normalnosti, na primer, čest slučaj su višemodalne distribucije. Najjednostavniji primer višemodalne distribucije je bimodalna distribucija – distribucija sa dve najčešće vrednosti, odnosno dva „vrha" (moda). Ukoliko bi konstrukt koji procenjujemo trebalo da bude normalno distribuiran, višemodalnost predstavlja ozbiljnije odstupanje od ove pretpostavke, te narušava diskriminativnost merenja. Sa druge strane, ovakve distribucije mogu ukazivati na to da uzorak nije izvučen iz homogene, već iz heterogene populacije. Ukoliko ispitujemo stavove ljudi prema određenom društvenom fenomenu prema kome postoje polarisani stavovi, kao što je, na primer, usvajanje dece od strane homoseksualnih parova, možemo očekivati da će mali broj ispitanika imati neutralan stav, dok će većina biti ili jasno za ovu praksu ili jasno protiv nje. Ukoliko distribucija odstupa od normalnosti na ovaj način, važno je pre svega razumeti uzroke modalnosti, pa u zavisnosti od prirode podataka primeniti adekvatne statističke postupke za obradu podataka.

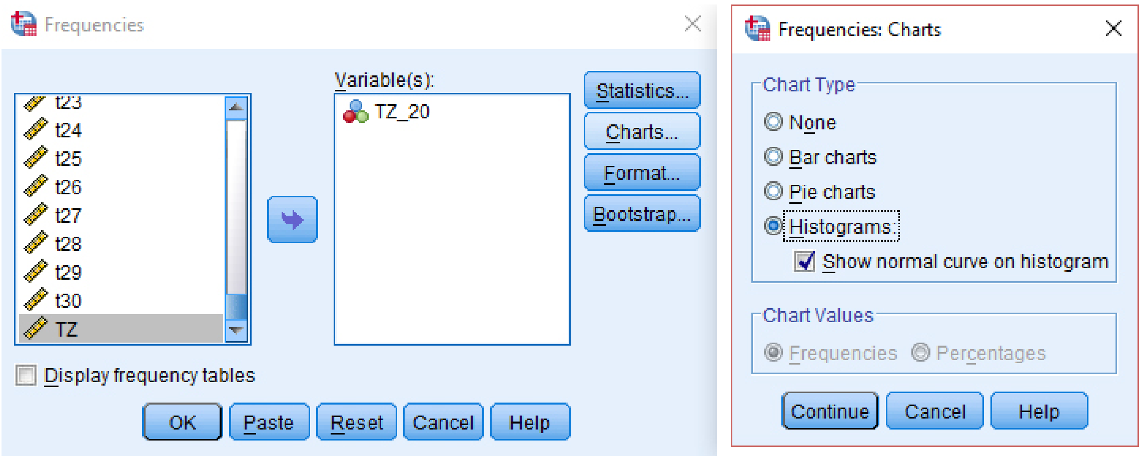
1.3) Provera diskriminativnosti u SPSS-u

Kako možemo proveriti da li je distribucija konkretnog testa normalna ili ne? Postoji više metoda utvrđivanja normalnosti distribucije (Pallant, 2005; Yazici & Yolacan, 2007). Mnoge od njih su implementirane u IMB® SPSS® statistički softver (IBM SPSS Statistics software, "SPSS"), a sada ćemo ih ilustrovati na primeru distribucije testa znanja iz psihologije koji se polaže na prijemnom ispitu za upis osnovnih studija psihologije. Za potrebe ove analize u obzir ćemo uzeti odgovore kandidata na prvih 20 pitanja. Ova pitanja imaju po 5 ponuđenih odgovora, od kojih je samo jedan tačan, a podaci su rekodirani tako da u novim varijablama imamo informaciju o tačnosti odgovora na datih 20 pitanja (videti Dodatak 1). Na osnovu rekodiranih varijabli izračunat je ukupan skor sa teorijskim rasponom 0–20.

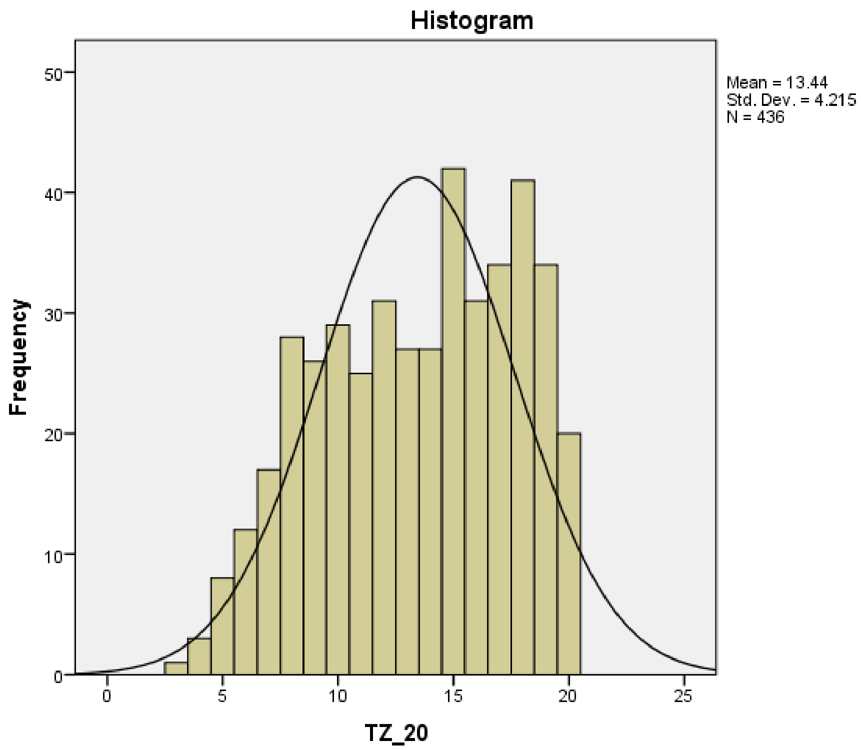
1.3.1) Provera diskriminativnosti na osnovu histograma

Najjednostavniji način da steknemo predstavu o normalnosti distribucije rezultata na testu (ili odstupanju od normalnosti) jeste uvidom u njen izgled. Distribuciju kontinuirane varijable prikazujemo histogramom, koji se u SPSS-u može pronaći u opciji **Analyze** → **Descriptive Statistics** → **Frequencies**, kao i u okviru opcije **Graphs** (**Chart Builder** ili **Legacy Dialogues** → **Histogram**).

Pošto će nam za kasnije analize biti potrebni određeni deskriptivni pokazatelji distribucije, sada ćemo koristiti opciju **Frequencies**. U prozoru koji se otvara u polje **Variables** kao varijablu ubacujemo sumarni skor na testu znanja, a zatim idemo na dugme **Charts** i biramo opciju **Histograms**. Zarad lakšeg uočavanja eventualnih odstupanja od normalnosti možemo uključiti i opciju prikazivanja normalne krive na histogramu (**Show normal curve on histogram**, Slika 1.4). Slika 1.5 prikazuje SPSS ispis.



Slika 1.4



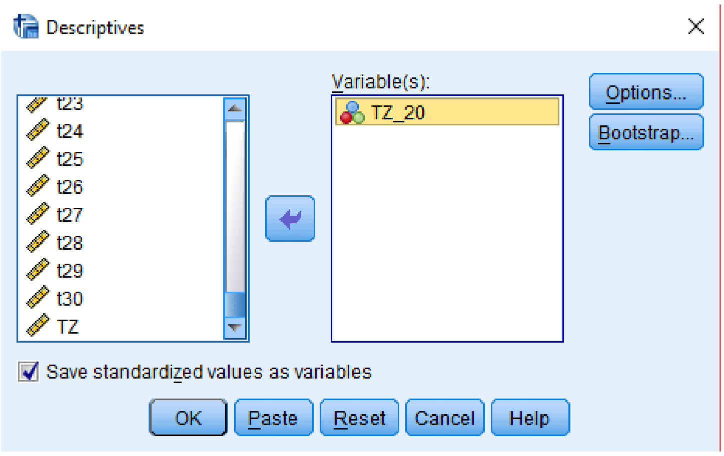
Slika 1.5

Šta možemo zaključiti o obliku raspodele na osnovu prikazanog histograma? Za početak, vidimo da postoji grupisanje skorova na desnom kraju distribucije, što ukazuje na negativno asimetričnu distribuciju. Drugim rečima, veliki broj kandidata dobro je uradio test znanja i postigao visok broj poena. Takođe, možemo primetiti da, sa izuzetkom nekoliko najnižih skorova – skoro svi skorovi imaju relativno visoke i ujednačene frekvence, što sugeriše da je distribucija platikurtična. Samo na osnovu inspekcije histograma obično nećemo doneti konačni zaključak o obliku distribucije skorova, već ćemo uzeti u obzir i dodatne pokazatelje. Ipak, inspekcija histograma je korak u proveravanju diskriminativnosti testa koji ne treba preskočiti jer može pomoći da se uoče autlajeri, kao i druge eventualne atipičnosti u distribuciji, koje treba dodatno proveriti (npr., „nepostojeći skorovi” koje treba rekodirati u nedostajuće vrednosti ili slično).

1.3.2) Provera diskriminativnosti na osnovu deskriptivnih pokazatelja

Kao što je prethodno rečeno, aritmetička sredina je osetljivija na ekstremne vrednosti od medijane, te poređenjem ove dve mere centralne tendencije takođe možemo saznati nešto o obliku distribucije. Vrednosti aritmetičke sredine i medijane u SPSS-u možemo dobiti korišćenjem opcije **Frequencies,** gde u okviru opcije **Statistics** treba označiti željene pokazatelje. Vrednosti aritmetičke sredine i medijane za ove podatke iznose redom, M = 13.44 i Mdn = 14. Vidimo da je medijana veća od aritmetičke sredine, odnosno da je aritmetička sredina malo pomerena ulevo u poređenju sa medijanom. Ovo se može interpretirati kao da je distribucija negativno asimetrična (jer su ekstremne vrednosti na levom, negativnom kraju distribucije), što je u skladu sa zaključkom koji smo doneli i na osnovu inspekcije histograma. Iako ovo nije pouzdan način utvrđivanja oblika distribucije (jer se često dešava da odnos M i Mdn ne bude očekivan), bitno je razumeti da samo na osnovu poznavanja vrednosti M i Mdn možemo imati određene pretpostavke o obliku distribucije, konkretno o njenoj simetričnosti. Sa druge strane, na osnovu ovih statistika ne možemo izvući nikakve zaključke o razvučenosti distribucije.

Poznato je da se oko 99.73% svih vrednosti normalne distribucije nalazi između -3 i +3 standardne devijacije. Stoga, može biti informativno proveriti raspon distribucije standardizovanih skorova na testu. Jedan od načina standardizovanja varijable je korišćenje predefinisane SPSS opcije za standardizaciju, koja se nalazi u okviru **Analyze** → **Descriptive Statistics** → **Descriptives**. U polje **Variables** treba ubaciti ukupni skor na testu, a zatim štiklirati opciju da se standardizovane vrednosti sačuvaju kao nova varijabla (**Save standardized values as variables**, Slika 1.6).

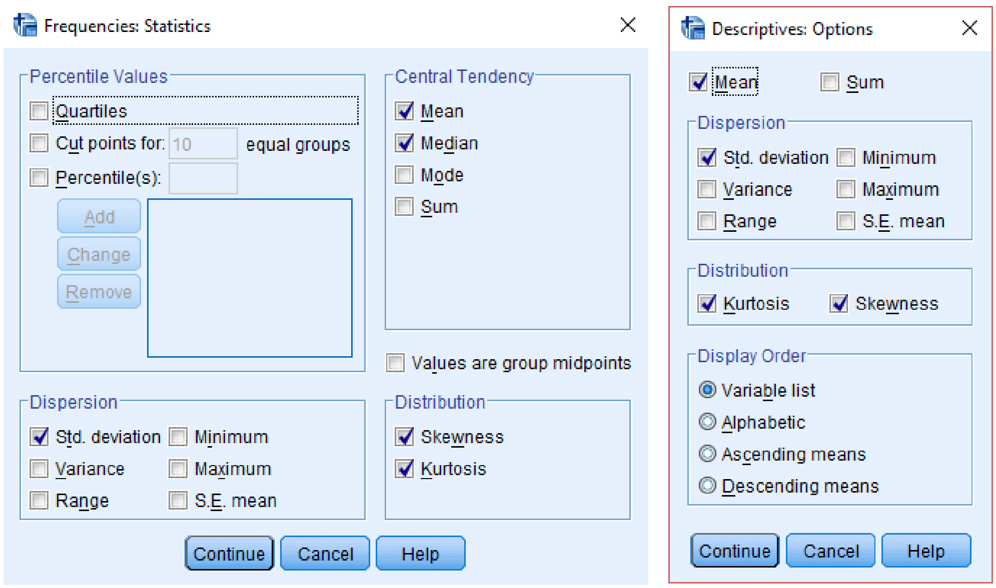


Slika 1.6

Sada u fajlu sa podacima imamo novu varijablu čije je ime ZTZ_20 (slovo Z dodaje se ispred imena originalne varijable u kojoj su sačuvani sirovi skorovi), i u čijem nazivu (eng. *label*) piše da je u pitanju Z skor za varijablu TZ_20. Da bismo videli raspon standardizovanih skorova, možemo koristiti bilo opciju **Frequencies**, bilo **Descriptives**, u okviru kojih je potrebno označiti da se prikažu minimum i maksimum za novu, standardizovanu varijablu. Ako koristite opciju **Descriptives**, ne zaboravite da isključite opciju standardizovanja varijabli (jer ćete u suprotnom dobiti novu varijablu koja u sebi sadrži standardizovane vrednosti standardizovanih vrednosti originalne varijable).

Minimalni z-skor distribucije poena kandidata na testu znanja je -2.48, dok je maksimalna vrednost 1.56. Primećujemo da je levi „rep" distribucije duži od desnog, što znamo da ukazuje na negativno asimetričnu distribuciju. Ovaj zaključak u skladu je i sa izgledom histograma i sa vrednostima aritmetičke sredine i medijane. Takođe, možemo videti da je raspon distribucije sužen u poređenju sa rasponom normalne distribucije koji treba da iznosi šest standardnih devijacija (tri ispod i tri iznad aritmetičke sredine). Sužen raspon skorova ne ukazuje jednoznačno ni na leptokurtičnu ni na platikurtičnu distribuciju, ali svakako govori o tome da test nije uspeo da uhvati kompletan raspon varijabilnosti znanja kandidata. S obzirom na to da test (odnosno, jedan njegov deo) ima samo 20 pitanja, ovo i nije previše neočekivano.

Prethodne dve metode provere diskriminativnosti daju nam veoma grubu procenu oblika distribucije, te je moguće da ćemo, primenjujući ih, doneti pogrešne zaključke o stvarnom obliku distribucije. Deskriptivni pokazatelji koji najdirektnije govore o obliku distribucije su njen skjunis i kurtozis. Ove vrednosti možemo dobiti i u okviru opcije **Frequencies** i u okviru opcije **Descriptives**, samo je potrebno štiklirati Sk i Ku u odgovarajućim prozorima (Slika 1.7).



Slika 1.7

Vrednost skjunisa za ovu distribuciju iznosi Sk = -0.268, dok je vrednost kurtozisa Ku = -0.981. Negativna vrednost skjunisa implicira da se radi o negativno asimetričnoj distribuciji, dok negativna vrednost kurtozisa sugeriše da je distribucija platikurtična. Ali da li je zaista tako?

Možda ste se, čitajući prethodne pasuse, zapitali koliko velika treba da bude neka vrednost ili razlika da bismo je smatrali značajnom. Konkretnije, ako bi vrednost aritmetičke sredine za ove podatke bila 13.99, a vrednost medijane 14 – da li bismo rekli da je distribucija asimetrična? Ako je levi rep distribucije na -2.48, a desni na +2.39, da li je levi rep zaista duži od desnog? Ako je skjunis jednak -0.008 a kurtozis -0.081, da li je distribucija zaista asimetrična, odnosno platikurtična?

Na stvarnim podacima, dobijenim na bilo kom konkretnom uzorku, uvek će postojati izvesna razlika u vrednostima medijane i aritmetičke sredine, a repovi distribucije neće biti sasvim jednaki niti će Sk i Ku biti tačno nula. Ali, kao što već znamo, nije svaka numerička razlika značajna. Prilikom poređenja aritmetičke sredine i medijane, kao i prilikom razmatranja raspona empirijske distribucije z-skorova, ne postoje definisane vrednosti koje nedvosmisleno ukazuju da je razlika značajna. Ovo je obično prepušteno proceni istraživača, što može biti problematično zbog subjektivnosti ovako definisanih kriterijuma. Objektivne mere na koje se možemo osloniti prilikom donošenja zaključaka su statistički testovi odstupanja distribucije od normalnosti (koji će biti prikazani u nastavku teksta), kao i vrednosti (standardizovanih) skjunisa i kurtozisa.

1.3.3) Provera diskriminativnosti na osnovu standardizovanih skjunisa i kurtozisa

Da bismo objasnili na koji način nam standardizovani skjunis i kurtozis ukazuju na odstupanja od normalnosti, napravićemo malu digresiju o statističkom ocenjivanju parametara.

Pretpostavimo za trenutak da je stvarna populacijska distribucija skorova na testu znanja iz psihologije normalna. Slučajnim uzorkovanjem iz date populacije, odnosno zadavanjem testa jednoj generaciji kandidata, dobijamo podatke na osnovu kojih možemo proceniti aritmetičku sredinu ove distribucije. Ako bismo skorove izrazili kao z-skorove, populacijska aritmetička sredina bila bi jednaka nuli, dok bi se aritmetička sredina dobijena na uzorku (tzv. statistik) od nje verovatno razlikovala, u manjoj ili većoj meri. Zamislimo dalje da smo u mogućnosti da veoma veliki broj puta uzorkujemo iz ove populacije, na primer, da isti test zadajemo više godina različitim grupama kandidata (ovde moramo pretpostaviti da će rezultat na testu svaki put biti samo rezultat znanja kandidata, a ne i poznatosti sadržaja već zadavanog testa). Svaka uzoračka procena aritmetičke sredine bila bi onda nezavisna procena aritmetičke sredine populacije. Ukoliko bismo histogramom prikazali frekvencu pojavljivanja uzoračkih vrednosti aritmetičke sredine, dobili bismo normalnu distribuciju. Drugim rečima, ako je populacijska aritmetička sredina jednaka nuli – većina uzoračkih aritmetičkih sredina imale bi vrednosti blizu nule, dok bi manji broj uzoračkih aritmetičkih sredina bio umereno pozitivan ili negativan, a izuzetno mali broj aritmetičkih sredina bi činile visoke pozitivne i visoke negativne vrednosti. Aritmetička sredina ove distribucije bila bi populacijska aritmetička sredina, odnosno njena vrednost bi bila nula. Standardna devijacija ove distribucije poznata je kao standardna greška aritmetičke sredine i ukazuje na stepen poverenja koji možemo imati u svoju procenu parametra (populacijske vrednosti). Ona nam takođe omogućava da testiramo da li je aritmetička sredina jednaka određenoj vrednosti ili se od nje značajno razlikuje[3].

Potpuno analogno ovome, ako pretpostavljamo da je raspodela skorova na testu znanja iz psihologije normalna, onda je populacijska vrednost skjunisa i kurtozisa jednaka nuli. Analogiju ćemo dalje nastaviti samo na primeru skjunisa, iako potpuno isto važi i za kurtozis. Dakle, zamislimo li da uzorkujemo veoma veliki broj puta iz populacije kandidata za upis na osnovne studije psihologije, vrednosti skjunisa koje budemo dobijali neće u svakom uzorku biti jednake nuli, iako su uzorci izvučeni iz populacije gde je znanje normalno raspodeljeno. Najveći broj uzoračkih skjunisa biće jednak nuli ili veoma blizu ove vrednosti, manji broj će biti umereno pozitivne i umereno negativne vrednosti, dok će u veoma malom procentu slučajeva skjunis za uzorak izvučen iz normalne distribucije biti visoka pozitivna ili negativna vrednost. Distribucija uzoračkih procena

[3] Za više informacija videti poglavlje VIII kod Tenjovića (Tenjović, 2000).

populacijskog skjunisa takođe je normalna distribucija. Njena aritmetička sredina je populacijska vrednost skjunisa, odnosno nula, a njena standardna devijacija naziva se standardnom greškom skjunisa (SE_{Sk}). Standardna greška skjunisa dobija se prema sledećoj formuli $SE_{Sk} = \sqrt{\frac{6}{N}}$, gde je N veličina uzorka (Blanca et al., 2013; Groeneveld & Meeden, 1984; Joanes & Gill, 1998). Kao što je već rečeno, potpuno isto važi i za kurtozis, tako da je aritmetička sredina uzoračke distribucije kurtozisa jednaka nuli, dok je standardna devijacija ove distribucije standardna greška kurtozisa (SE_{Ku}). Standardna greška kurtozisa računa se kao $SE_{Ku} = \sqrt{\frac{24}{N}}$.

Vratimo se sada standardizovanom skjunisu i kurtozisu. Već smo rekli da na osnovu ovih vrednosti možemo odrediti da li raspodela značajno odstupa od normalne. Na koji tačno način to možemo učiniti? Pre svega, potrebno je standardizovati vrednosti Sk i Ku. Kao i bilo koji drugi skor, skjunis i kurtozis se standardizuju tako što se od sirovog skora oduzme aritmetička sredina, a zatim ova razlika podeli standardnom devijacijom. U slučaju skjunisa i kurtozisa, kao što smo već pokazali, aritmetička sredina je jednaka nuli, dok je standardna devijacija poznata kao standardna greška skjunisa, odnosno kurtozisa. Na primeru koji smo dali, vrednost skjunisa je bila Sk = -0.268, dok je vrednost kurtozisa Ku = -0.981. Standardna greška skjunisa iznosi SE_{Sk} = 0.117, a standardna greška kurtozisa SE_{Ku} = 0.233. Ako ubacimo ove vrednosti u formulu, dobijamo da je $z_{Sk} = \frac{Sk - M_{Sk}}{SD_{Sk}} = \frac{Sk - 0}{SE_{Sk}} = \frac{-.268 - 0}{.117} = \frac{-.268}{.117} = -2.29$ i da je $z_{Ku} = \frac{Ku - M_{Ku}}{SD_{Ku}} = \frac{Ku - 0}{SE_{Ku}} = \frac{-.981 - 0}{.233} = \frac{-.981}{.233} = -4.21$. Šta nam ove vrednosti govore? Kako na osnovu njih određujemo da li je distribucija normalna ili nije, odnosno koje vrednosti su dovoljno visoke da bismo rekli da distribucija *značajno* odstupa od normalne?

Poznato je da se 95% slučajeva nalazi između -1.96 i +1.96 standardne devijacije, dok se 99% slučajeva nalazi između -2.58 i +2.58 standardne devijacije. Drugim rečima, ukoliko je neki standardizovani skor veći od 1.96 ili manji od -1.96, to znači da se on nalazi u 5% najviših, odnosno najnižih skorova distribucije. Slično tome, skor veći od 2.58 ili niži od -2.58 spada u 1% najređih skorova normalne distribucije. Zbog čega su nam važne baš ove vrednosti? Verovatno ste već pretpostavili da je to zbog konvencionalno određenih nivoa statističke značajnosti od .05 i .01, odnosno od 5% i 1%. Ako govorimo u terminima testiranja hipoteza, nulta hipoteza je da je dati z-skor izvučen iz normalne distribucije (sa aritmetičkom sredinom 0 i standardnom devijacijom 1). Ukoliko je apsolutna vrednost z-skora veća od ±1.96, verovatnoća da je taj skor izvučen iz normalne distribucije je manja od 5%, na osnovu čega je uobičajeno odbaciti nultu hipotezu. Isto važi i za apsolutnu vrednost z-skora od ±2.58, s tim što u ovom slučaju nultu hipotezu odbacujemo na višem nivou značajnosti od .01. Dakle, vrednosti ±1.96 i ±2.58 predstavljaju kritične vrednosti za određivanje značajnosti skjunisa i kurtozisa. Ukoliko je vrednost standardizovanog skjunisa ili kurtozisa izvan opsega ±1.96 (ili ±2.58), možemo reći da distribucija značajno odstupa od normalne.

Da se podsetimo, značajna pozitivna vrednost skjunisa ukazuje na pozitivnu asimetriju, značajna negativna vrednost na negativnu asimetriju, dok značajna pozitivna vrednost kurtozisa ukazuje na leptokurtičnost, a značajna negativna na platikurtičnost distribucije. Što su odstupanja od normalnosti veća, smatramo da test lošije diskriminiše ispitanike. Vrednost standardizovanog skjunisa koju smo dobili za test znanja iznosi z_{Sk} = -2.29. Kako je ova vrednost značajno različita od nule, na nivou .05 (ali ne i na nivou od .01), zaključujemo da je distribucija negativno asimetrična, što je u skladu i sa prethodno razmotrenim pokazateljima. Standardizovani kurtozis ove distribucije je z_{Ku} = -4.21 i ova vrednost značajno odstupa od nule na nivou .01 (dakle, samim tim i na nivou od

.05), pa možemo zaključiti da je distribucija platikurtična. Ponovo, ovaj zaključak je konzistentan sa zaključkom donetim na osnovu uvida u histogram.

Zapamtite – ukoliko su vrednosti standardizovanog skjunisa ili kurtozisa unutar raspona ±1.96, to znači da ne možemo odbaciti nultu hipotezu da je populacijska distribucija iz koje je ovaj uzorak izvučen normalna (simetrična ili mezokurtična). Drugim rečima, ako vrednost standardizovanog skjunisa ne izlazi iz opsega ±1.96, ne odbacujemo nultu hipotezu i zaključujemo da je distribucija simetrična. Analogno tome, ako je vrednost standardizovanog kurtozisa unutar opsega od ±1.96, zaključujemo da je distribucija mezokurtična. Tako bismo, na primer, za distribuciju čiji je standardizovani skjunis jednak $z_{Sk} = 1.53$ rekli da je simetrična, iako ova vrednost numerički odstupa od nule. Ključ je u tome što ovo odstupanje nije statistički značajno.

Jedan od nedostataka metode proveravanja diskriminativnosti na osnovu standardizovanih skjunisa i kurtozisa jeste što su ove mere veoma osetljive na veličinu uzorka. Kao što je iz formula za standardnu grešku skjunisa i kurtozisa jasno – sa povećanjem broja ispitanika u uzorku standardna greška se smanjuje. Što je standardna greška manja, to će za istu vrednost skjunisa ili kurtozisa standardizovana vrednost biti veća. Samim tim, na velikim uzorcima ($N > 200$) može se desiti da čak i mala vrednost Sk ili Ku bude statistički značajna (Field, 2009; Pallant, 2005). Zbog toga se u ovim situacijama preporučuje da se umesto standardizovanih, razmatraju vrednosti sirovog skjunisa i kurtozisa. Postoje različite okvirne preporuke za kritične vrednosti iznad kojih se može govoriti o odstupanju od normalnosti raspodele. Tako, prema jednom strožem kriterijumu, vrednosti u okviru ±0.5 ukazuju na normalnu raspodelu, vrednosti između ±0.5 i ±1 na umerena odstupanja od normalnosti, dok vrednosti veće od ±1 sugerišu ekstremna odstupanja od normalnosti (Bulmer, 1979). Po jednom blažem kriterijumu, međutim, normalnim se mogu smatrati distribucije čiji se Sk i Ku nalaze u okviru vrednosti ±2 (George & Mallery, 2003). Ipak, u slučaju velikih uzoraka, pre svega treba izvršiti vizuelnu inspekciju distribucije skorova (Field, 2009; Pallant, 2005).

1.3.4) Provera diskriminativnosti hi-kvadrat testom

Hi-kvadrat test jedan je od najčešće upotrebljavanih statističkih testova i ima brojne primene (Cochran, 1952). U kontekstu testiranja normalnosti, hi-kvadrat test se koristi da bi se utvrdilo da li empirijska raspodela značajno odstupa od pretpostavljene, odnosno u ovom slučaju normalne distribucije. Što je razlika između empirijske i normalne distribucije veća, njena diskriminativnost je lošija. Hi-kvadrat test računa se prema sledećoj opštoj formuli:

$$\chi^2 = \sum_{i=1}^{j} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

gde je j broj razreda, f_o opservirana ili empirijska frekvenca u datom razredu (slovo *o* je od engleskog *observed*), a f_e očekivana ili teorijska frekvenca (slovo *e* je od engleskog *expected*). Za datu vrednost χ^2 i broj stepeni slobode $df = j - 1 - 2$ možemo izračunati verovatnoću da se dobije toliko visok χ^2, na osnovu čega donosimo odluku o prihvatanju ili odbacivanju nulte hipoteze.

Na primeru 20 pitanja na testu znanja iz psihologije, ukoliko pretpostavimo da je svaki teorijski moguć skor na testu jedan razred, onda vrednosti f_o odgovaraju dobijenim frekvencama po razredima, odnosno frekvencama pojedinačnih skorova na testu. Očekivane frekvence su frekvence koje bi trebalo da dobijemo u svim razredima pod pretpostavkom da je distribucija normalna. U potpunosti automatizovano testiranje normalnosti distribucije ne postoji kao posebna opcija u SPSS-u, zbog čega se ova metoda nešto ređe i koristi. Dodatak 2 daje detaljni prikaz svih koraka koje treba proći kako bi se hi-kvadrat testom proverila pretpostavka o normalnosti instrumenta, kao i rezultate ove analize na primeru 20 pitanja sa testa znanja iz psihologije. Primena hi-kvadrat testa na ovim podacima ukazuje na značajna odstupanja od normalnosti, a na osnovu uvida u rezidualne frekvencije možemo naslutiti i da je test bio lakši nego što bi bilo poželjno (jer distribucija pokazuje niže frekvence za srednje skorove, a više frekvence za više skorove). Ipak, sam hi-kvadrat test nam ne pruža direktnu informaciju o obliku distribucije, odnosno o prirodi odstupanja od normalnosti.

1.3.5) Provera diskriminativnosti Kolmogorov–Smirnov testom

Još jedan od statističkih testova koji testiraju hipotezu o normalnosti distribucije varijable jeste Kolmogorov–Smirnov test (Lilliefors, 1969; Massey, 1951). Ovaj test počiva na poređenju empirijske (opservirane) i teorijske (očekivane) kumulativne distribucije rezultata, te utvrđivanju najveće, odnosno maksimalne razlike između ove dve distribucije. Drugim rečima, za svaku tačku (skor) distribucije se porede empirijska i teorijska vrednost i traži se najveća razlika. Kolmogorov–Smirnov D statistik (D od engleskog *difference*) računa se prema sledećoj formuli:

$$D = \max |F_o(x) - F_e(x)|$$

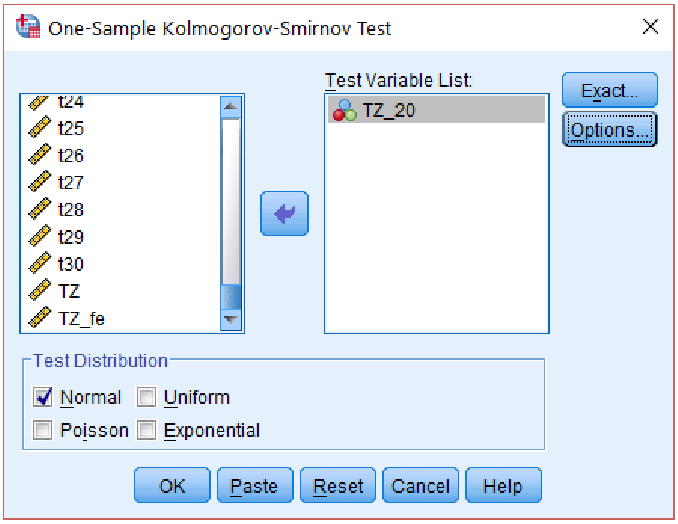
gde su F_o i F_e, redom, opservirana i očekivana kumulativna frekvenca.

U slučaju savršeno normalne distribucije D statistik bi iznosio nula, odnosno ne bi bilo razlike između dobijene i očekivane kumulativne distribucije frekvenci. Da bi se utvrdilo da li je D statistik statistički značajno veći od nule, odnosno da li raspodela značajno odstupa od normalne, on se poredi sa kritičnom vrednošću. Kritična vrednost D statistika, D_k, iznosi:

$$D_k = \frac{k}{\sqrt{N}}$$

gde je N veličina uzorka, a k jednako 1.36 za nivo značajnosti .05 i 1.63[4] za nivo značajnosti .01 (Lilliefors, 1969; Massey, 1951; Miller, 1956). Ukoliko je $D > D_k$, odbacujemo nultu hipotezu na datom nivou značajnosti, dok je ako je $D < D_k$, prihvatamo.

Kolmogorov–Smirnov test se u SPSS-u nalazi u okviru neparametrijskih testova i može se pokrenuti preko **Analyze** → **Nonparametric Test** → **Legacy Dialogues** → **1-Sample K-S** ili **Analyze** → **Nonparametric Test** → **One Sample**. Ukoliko koristimo **Legacy Dialogues** opciju, u polje *Variables* treba uneti željenu varijablu i ostaviti automatski štikliranu opciju *Normal* u okviru *Test Distribution* (Slika 1.8).



Slika 1.8

U ispisu iz ove analize dobijamo informacije o broju ispitanika, aritmetičkoj sredini i standardnoj devijaciji varijable, kao i o najvećoj pozitivnoj, negativnoj i apsolutnoj razlici (Slika 1.9). Najveća apsolutna razlika je D statistik i ona se može

[4] Navedene vrednosti se koriste kod dvosmernog testiranja značajnosti, dok se kod jednosmernog testiranja značajnosti koriste vrednosti 1.22 za nivo značajnosti .05 i 1.52 za nivo značajnosti .01 (Miller, 1956).

uporediti sa D_k vrednošću kako bi se doneo zaključak o značajnosti odstupanja distribucije od normalne. Pored D statistika, SPSS daje i Kolmogorov–Smirnov Z statistik i njegovu pripadajuću značajnost.

One-Sample Kolmogorov-Smirnov Test

		TZ_20
N		436
Normal Parameters[a,b]	Mean	13.44
	Std. Deviation	4.215
Most Extreme Differences	Absolute	.108
	Positive	.077
	Negative	-.108
Kolmogorov-Smirnov Z		2.256
Asymp. Sig. (2-tailed)		.000

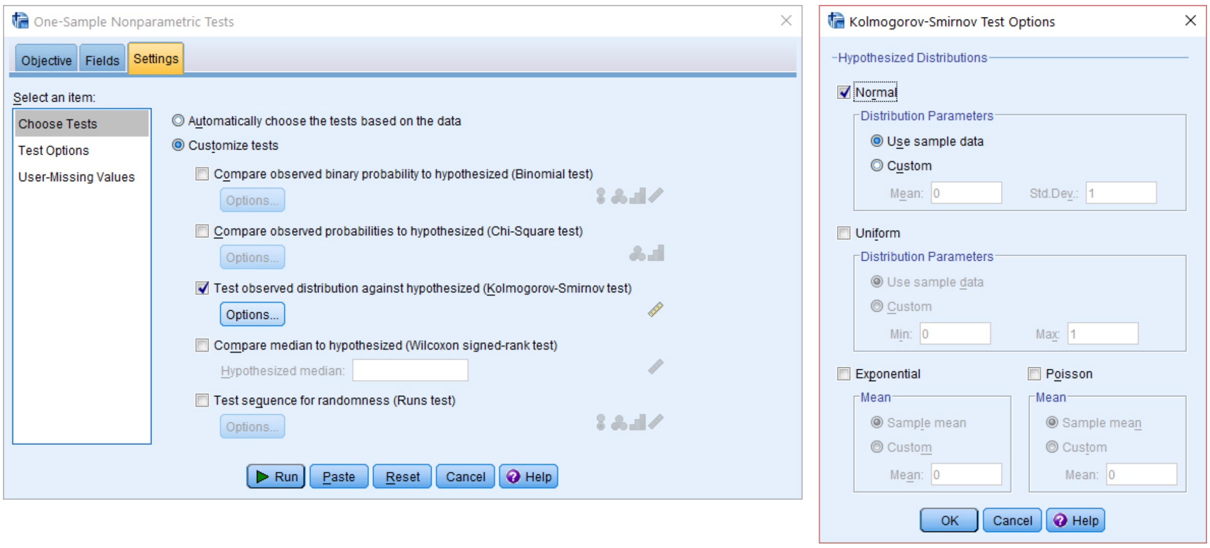
a. Test distribution is Normal.

b. Calculated from data.

Slika 1.9

U slučaju našeg testa, najveća apsolutna razlika između empirijske i teorijske kumulativne distribucije iznosi D = .108. Kako je D_k = 1.36 / √436 = 1.36 / 20.88 = .065 za nivo značajnosti .05 i D_k = 1.63 / √436 = 1.63 / 20.88 = .078 za nivo značajnosti .01, možemo zaključiti da distribucija testa znanja iz psihologije značajno odstupa od normalnosti, na oba nivoa značajnosti (jer je $D > D_k$). Do istog zaključka dolazimo i na osnovu značajnosti Kolmogorov–Smirnov stastistika iz SPSS tabele.

Ukoliko za računanje Kolmogorov–Smirnov statistika koristimo opciju **Analyze → Nonparametric Test → One Sample**, potrebno je u nekoliko koraka podesiti željenu analizu. Kao i prilikom računanja hi-kvadrata, u kartici *Objective* biramo *Customize analysis*, a u kartici *Fields* iz polja *Test Fields* izbacujemo sve varijable osim one čiju normalnost testiramo. U kartici *Settings* ponovo biramo *Customize Tests*, ali ovog puta biramo opciju *Test observed distribution against hypothesized (Kolmogorov-Smirnov test)*, a zatim u *Options* definišemo da je očekivana distribucija normalna – *Normal* (Slika 1.10).



Slika 1.10

Ispis ove analize drugačiji je od **Legacy Dialogues** ispisa i sadrži samo informaciju o nultoj hipotezi, primenjenom statističkom testu, njegovoj značajnosti i zaključku koji treba doneti. U ovom slučaju, vidimo da je Kolmogorov–Smirnov test značajan na nivou $p < .001$ (vrednost je ista onoj koju dobijamo u **Legacy Dialogues** ispisu), te da nultu hipotezu o normalnosti distribucije treba odbaciti (Slika 1.11).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of TZ_20 is normal with mean 13.44 and standard deviation 4.21.	One-Sample Kolmogorov-Smirnov Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Slika 1.11

Kao i kod hi-kvadrat testa, SPSS ispis za K-S test ne uključuje informacije o vrsti odstupanja od normalnosti, ukoliko ono postoji. Ispis nam daje samo informaciju da li je distribucija normalna ili nije, te u slučaju da se nulta hipoteza odbaci treba dalje utvrditi oblik distribucije uvidom u skjunis i kurtozis.

Videli smo da i Hi-kvadrat test i Kolmogorov–Smirnov test počivaju na poređenju empirijskih sa teorijskim frekvencama, međutim, prilikom računanja hi-kvadrat testa u obzir se uzimaju sve razlike, dok K-S test uzima u obzir samo najveću razliku. Takođe, hi-kvadrat test računa odstupanja na običnim, a K-S test na kumulativnim frekvencijama. Zbog toga je moguće da se u određenim situacijama desi da jedan test ukazuje na odstupanje distribucije od normalnosti, dok drugi ukazuje na odsustvo odstupanja. Na primer, ukoliko postoji veliki broj malih odstupanja empirijskih od teorijskih frekvenci, hi-kvadrat test će biti značajan, dok K-S ne mora biti. Sa druge strane, jedno veliko odstupanje može dovesti do toga da K-S bude značajan, dok hi-kvadrat neće biti. Zbog toga je veoma važno razumeti na koji način svi pomenuti testovi proveravaju hipotezu o normalnosti distribucije kada se donosi konačan zaključak o diskriminativosti testa.

1.3.6) Provera diskriminativnosti – zaključak

U prethodnim odeljcima prikazali smo nekoliko metoda za određivanje normalnosti distribucije, odnosno za proveru diskriminativnosti testa. Iako to ne mora uvek biti slučaj (zbog različitih principa koji stoje u osnovi računanja pomenutih pokazatelja i testova), na primeru testa znanja iz psihologije svi pokazatelji ukazivali su na isti obrazac rezultata. To je negativno asimetrična i platikurtična distribucija. Šta nam ovaj oblik distribucije govori o testu, odnosno o ispitanicima? Negativna asimetrija ukazuje da je test bio lak, barem za ovaj uzorak. Moguće je da bi isti test na nekom drugom uzorku, na primer, uzorku kandidata koji žele da upišu matematiku, bio srednje težak ili veoma težak. Metrijske karakteristike koje dobijamo na nekom uzorku su uvek metrijske karakteristike skorova, a ne samog testa (American Educational Research Association [AERA] et al., 2014). Međutim, u ovom slučaju ne radi se samo o drugačijem uzorku, već i o drugačijoj populaciji. Ako smatramo da je uzorak kandidata kojima je test zadat bio reprezentativan za populaciju iz koje je izvučen, možemo očekivati da bi se test slično ponašao i na drugim uzorcima iz iste populacije. Dakle, možemo zaključiti da je u testu postojalo više lakih pitanja nego što bi to bilo poželjno. Platikurtičnost distribucije, u kombinaciji sa asimetrijom, pokazuje da je diskriminativnost testa bila nešto veća za srednje i niske skorove, ali nešto niža za visoke skorove nego što bi bilo optimalno. Imajući u vidu da se radi o testu na prijemnom ispitu gde se prima oko četiri puta manji broj kandidata nego što se prijavljuje, poželjno bi bilo pojačati diskriminativnost upravo na desnom kraju distribucije, kako bi se maksimizirale razlike upravo između najboljih kandidata.

1.4) Normalizacija distribucije

Kao što je već rečeno, normalna distribucija je očekivana populacijska distribucija za većinu psiholoških fenomena kojima se bavimo, a normalna distribucija uzoračkih skorova daje najbolju diskriminativnost testa. Posmatrano iz te perspektive, odstupanja od normalnosti predstavljaju „nedostatak” testa, pa istraživači i konstruktori testova treba da teže dobijanju normalne raspodele prilikom dobijanja testa. Šta onda možemo uraditi ukoliko dobijemo raspodelu koja odstupa od normalne? Odgovor na ovo pitanje zavisi od toga da li se nalazimo u fazi konstrukcije testa ili koristimo već formiran instrument (Bukvić, 1996). Ukoliko smo mi konstruktori testa, onda imamo mogućnost da ga revizijom ajtema unapredimo tako da postignemo bolju diskriminativnost konačne verzije testa. Ukoliko pak ne postoji mogućnost promene samog testa, mogu se izvršiti određene transformacije samih podataka ili se mogu upotrebiti alternativne, neparametrijske metode obrade podataka (Pek et al., 2018).

1.4.1) Normalizacija testa

Zamislimo da smo test znanja iz psihologije iz prethodnog primera bili u mogućnosti da prvo zadamo jednom manjem uzorku srednjoškolaca zainteresovanih za psihologiju, te na osnovu diskriminativnosti na ovom „probnom” uzorku modifikujemo test tako da obezbedi idealnu diskriminativnost. Naravno, u realnim situacijama selekcije (naročito prijemnih ispita, državne mature i sl.) često nemamo mogućnost pilot testiranja, ali ćemo to sada zanemariti. Pretpostavimo dalje da smo na probnom uzorku dobili distribuciju koja je prikazana i detaljno analizirana u prethodnim odeljcima, odnosno negativno asimetričnu i platikurtičnu distribuciju. Objasnili smo kako interpretiramo ova odstupanja od normalnosti – test je ispitanicima bio prelak i proporcionalno je bilo nešto manje teških zadataka i zadataka srednje težine nego što bi trebalo.

Kako onda možemo izmeniti test da on u svojoj konačnoj formi proizvodi normalnu distribuciju skorova? Da bismo otežali test, možemo dodati teške zadatke ili izbaciti neke lake, ili oba, a kako bismo distribuciju učinili mezokurtičnom, potrebno je dodati srednje teške zadatke ili izbaciti neke lake i teške. U kombinaciji, ako želimo da zadržimo broj od 20 pitanja sa ponuđenim odgovorima, trebalo bi izbaciti neke lake stavke, a umesto njih dodati neke srednje teške i teške. Ovakve izmene trebalo bi da vode normalizaciji distribucije pri zadavanju nove verzije testa.

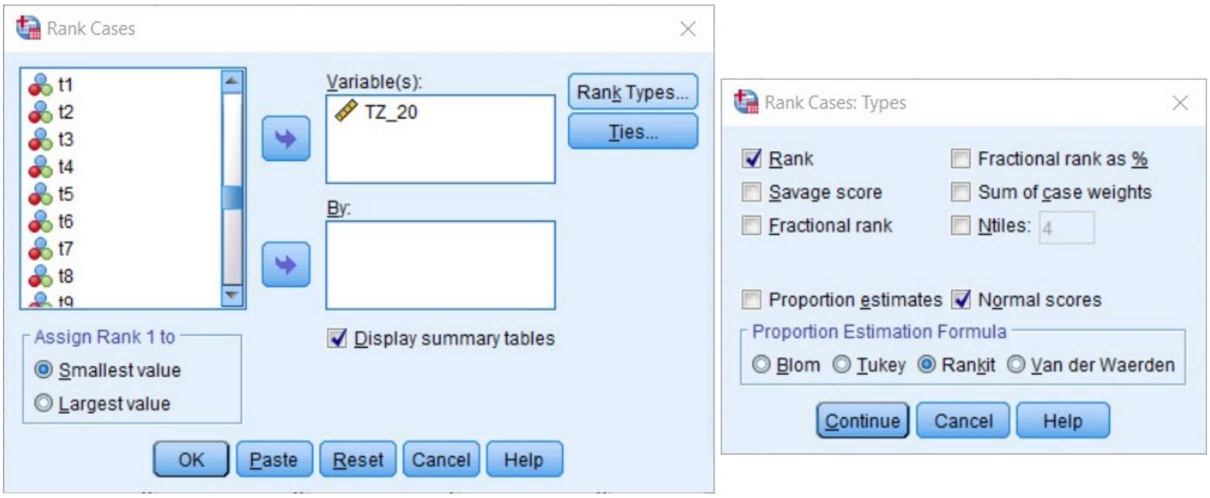
Načelno, negativno asimetričnu distribuciju korigujemo tako što dodajemo teške stavke ili izbacujemo lake. Još jedna mogućnost jeste skraćivanje vremena za rad. Suprotno tome, pozitivno asimetričnu distribuciju približavamo normalnoj tako što dodajemo lake zadatke, izbacujemo teške ili produžavamo vreme rada. Kako istraživači obično kreću u pilot testiranje sa nešto većim brojem stavki nego što će imati konačna verzija testa, normalizacija se češće postiže izbacivanjem nego dodavanjem stavki. Što se skraćivanja / produžavanja vremena rada tiče, ova procedura ima smisla samo kod testova znanja i sposobnosti, ali ne i kod nekognitivnih testova (npr., testova ličnosti), pa je primena ove metode u tom smislu ograničena.

Kako bi leptokurtična distribucija postala mezokurtična, potrebno je dodati lake i teške zadatke ili izbaciti neke srednje teške. Sa druge strane, platikurtična distribucija se približava normalnoj ukoliko se izbace neki laki i teški zadaci ili dodaju srednje teški. Naravno, u slučaju kada distribucija odstupa od normalne i po horizontalnoj i po vertikalnoj osi, potrebno je napraviti takve izmene koje kombinuju sve prethodno navedene preporuke za normalizaciju.

1.4.2) Normalizacija skorova

Koristeći isti primer, zamislimo sada drugačiju situaciju. Zamislimo (što je u ovom konkretnom primeru i verovatnije) da nemamo mogućnost da promenimo test koji je zadat uzorku, ali želimo da vidimo postoje li razlike u skorovima kandidata iz različitih tipova srednjih škola (gimnazija prirodni smer, gimnazija društveni smer, gimnazija opšti smer, srednja stručna škola itd.). Za utvrđivanje razlika između grupa na kontinuiranoj varijabli obično bismo upotrebili ANOVA-u, međutim, jedan od uslova za primenu ove analize jeste da kontinuirana varijabla bude normalno raspodeljena, što ovde nije slučaj. Jedno rešenje jeste da primenimo neparametrijsku tehniku obrade podataka, u ovom slučaju Kraskal–Volisov test (eng. *Kruskal–Wallis*), dok je drugo rešenje da transformacijom skorova normalizujemo distribuciju i zatim primenimo željenu parametrijsku tehniku.

Normalizacija varijable se u SPSS-u vrši preko komande **Transform → Rank Cases** tako što se u polje *Variable(s)* ubaci željena varijabla, a zatim u okviru *Rank Types* izaberu normalni skorovi – *Normal Scores (Proportion Estimation Formula – Rankit*[5]*)* (Slika 1.12). Automatski štikliranu opciju *Rank* možemo i odštiklirati, zato što ona daje rang datog skora, što nam u ovoj situaciji samo po sebi nije od interesa.



Slika 1.12

U ispisu nakon ove transformacije podataka (Slika 1.13) dobijamo informaciju da je SPSS kreirao novu varijablu, kako se ona zove i koji joj je naziv (eng. *label*), kao i na osnovu koje varijable je kreirana primenom koje konkretne funkcije. U našem primeru, izvorna varijabla TZ_20 je normalizovana primenom funkcije Normal Score. Novokreirana varijabla dobila je prefiks N i zove se NTZ_20, a njen naziv dodatno opisuje o kakvoj je varijabli reč.

Created Variables[a]

Source Variable	Function	New Variable	Label
TZ_20[b]	Normal Score	NTZ_20	Normal Score of TZ_20 using Rankit's Formula

a. Mean rank of tied values is used for ties.

b. Ranks are in ascending order.

Slika 1.13

[5] Iako su razlike između ponuđenih metoda normalizacije male, Rankit procedura se pokazuje najtačnijom, te se ona preporučuje kao podrazumevana (Solomon & Sawilowsky, 2009).

Pregledom vrednosti u varijabli *NTZ_20* (Slika 1.14) možemo primetiti da se one okvirno kreću u rasponu od -3 do +3, drugim rečima, normalizovani skor je dat u standardnim jedinicama, odnosno proces normalizacije u SPSS-u podrazumeva da se varijabla prvo standardizuje, a tek onda normalizuje. Zato je normalizovane skorove teško porediti direktno sa sirovim skorovima, ali se oni mogu porediti sa standardizovanim skorovima (koje od ranije imamo sačuvane u varijabli *ZTZ_20*). Upoređivanjem vrednosti varijabli *ZTZ_20* i *NTZ_20* vidimo da su neki skorovi pomereni ulevo, dok su neki pomereni udesno. Na primer, standardni skor -2.47604 nakon normalizacije iznosi -2.9818, odnosno pomeren je ulevo, ka negativnom kraju distribucije. Sa druge strane, standardizovanom skoru -1.52698 odgovara normalizovani skor -1.4411, što znači da je normalizacijom ovaj skor pomeren udesno. Za većinu skorova ova pomeranja neće biti velika, ali će na nivou celog uzorka dovesti do normalizacije distribucije. Pregledom deskriptivnih pokazatelja za normalizovani skor na testu znanja možemo videti da su aritmetička sredina i standardna devijacija približno jednake 0 i 1 (pošto je varijabla i standardizovana), dok su vrednosti Sk i Ku primetno niže nego što su bile za izvornu distribuciju (Slika 1.15). Ukoliko izračunamo standardizovani skjunis i kurtozis, dobićemo ZSk = -0.086 / 0.117 = -0.735 i Zku = -0.275 / 0.233 = -1.094. Obe vrednosti se nalaze u okviru kritičnih vrednosti ±1.96 (a samim tim i unutar vrednosti ±2.58), što znači da normalizovana distribucija ne odstupa značajno od normalne. Do istog zaključka mogli bismo doći i primenom Kolmogorov–Smirnov ili hi-kvadrat testa.

ZTZ_20	NTZ_20
-2.47604	-2.9818
-2.23878	-2.5111
-2.23878	-2.5111
-2.23878	-2.5111
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-2.00151	-2.0830
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.76425	-1.7330
-1.52698	-1.4411
-1.52698	-1.4411
-1.52698	-1.4411
-1.52698	-1.4411
-1.52698	-1.4411
-1.52698	-1.4411
-1.52698	-1.4411

Slika 1.14

Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Normal Score of TZ_20 using Rankit's Formula	436	-.005061	.9799776	-.086	.117	-.255	.233
Valid N (listwise)	436						

Slika 1.15

Ipak, treba biti oprezan u korišćenju normalizacije. Ona ima smisla samo ukoliko je odstupanje empirijske raspodele od normalne umereno (u skladu sa prethodno iznetim kriterijumima) i ukoliko imamo jasna očekivanja o normalnosti populacijske distribucije za dati konstrukt (Bukvić, 1996). Ukoliko empirijska distribucija veoma odstupa od normalne ili postoji mogućnost da je i populacijska distribucija zapravo asimetrična ili lepto/platikurtična (kao, na primer, u slučaju vremena reakcije, prosečnih primanja, psihopatoloških crta itd.), onda normalizacija nije preporučljiva. Takođe, ukoliko primetimo da normalizovani skorovi i dalje odstupaju od normalne raspodele – varijablu ne možemo smatrati uspešno normalizovanom. Umesto toga, treba koristiti neparametrijske tehnike obrade podataka ili (ako je moguće) na osnovu dostupnih podataka izračunati neki indeks koji je pogodan za primenu parametrijskih tehnika.

1.5) Faktori koji utiču na diskriminativnost testa

Više faktora utiče na diskriminativnost testa. Neke od njih (pre svega težinu stavki) već smo pomenuli kada smo razmatrali potencijalne uzroke odstupanja od normalnosti, kao i načine normalizovanja distribucije testa. Sada ćemo ukratko razmotriti na koji način broj i kvalitet stavki, način ocenjivanja i pouzdanost testa utiču na diskriminativnost testa.

1.5.1) Broj stavki i diskriminativnost testa

Broj stavki utiče na diskriminativnost testa na prilično očigledan način. Ukoliko na osnovu rezultata testa ispitanike možemo da svrstamo u samo dve kategorije – na primer, one koji su savladali gradivo za prijemni ispit iz psihologije i one koji ga nisu savladali, ostvarili smo izvestan vid diskriminativnosti. Ipak, jasno je da je ovo manja diskriminativnost nego da ispitanike razvrstavamo u deset kategorija, što je opet manja diskriminativnost nego da imamo dvadeset kategorija itd. Drugim rečima, broj stavki direktno je proporcionalan diskriminativnosti testa – veći broj stavki vodi i boljoj diskriminativnosti (Bukvić, 1996; Fajgelj, 2009).

Postoje, međutim, izvesni ograničavajući uslovi za ovu tvrdnju. Kvalitet stavki bitno utiče na diskriminativnost testa, tako da puko povećavanje broja stavki neće nužno poboljšati diskriminativnost testa, ukoliko stavke koje dodajemo nisu i same diskriminativne (više o ovome ćemo reći u narednom odeljku). Takođe, povećavanje broja stavki u instrumentu produžava vreme rada, što se može negativno odraziti i na motivaciju ispitanika, ali i proizvesti zamor kao mogući negativni uticaj na postignuće. Zato je važno pronaći balans između dovoljno velikog broja stavki, koji će voditi dobrim metrijskim karakteristikama (ne samo diskriminativnosti) i dovoljno malog, koji će instrument učiniti lakim i praktičnim za zadavanje.

1.5.2) Kvalitet stavki i diskriminativnost testa

Da bismo ilustrovali na koji način kvalitet stavki utiče na diskriminativnost testa, zamislimo sledeći primer. Ukoliko imamo 15 stavki u testu, ali je pet stavki toliko lako da ih *svi* ispitanici tačno rešavaju, onda će minimalan skor koji će bilo ko od ispitanika postići biti 5. Raspon skorova na testu neće biti 0–15, kao što bismo očekivali na osnovu broja stavki, već 5–15, pošto će samo odgovori na preostalih 10 stavki zapravo razlikovati ispitanike. Stavke koje rešavaju svi ispitanici ili one koje ne rešava niko nisu diskriminativne, pa ne doprinose ni diskriminativnosti testa. Distribucija skorova nakon isključenja ovih stavki će zapravo biti identična distribuciji sa njima (samo će se kretati od 0 do 10, umesto od 5 do 15), pa se ne može reći da one „kvare", odnosno negativno utiču na diskriminativnost. Ipak, postavlja se pitanje njihove svrhe u instrumentu. Ukoliko nam ne omogućavaju da razlikujemo ispitanike, nediskriminativne stavke samo predstavljaju dodatno opterećenje i za ispitanike i za ocenjivače, te ih zato uvek treba isključiti.

U manje ekstremnoj varijanti, kao što smo već videli kada smo razmatrali diskriminativnost testa, veoma lake stavke doprinosiće samo razlikovanju ispitanika sa najlošijim znanjem, dok će veoma teške stavke diskriminisati samo one čije je znanje najbolje. Time dolazimo do jednog aspekta kvaliteta stavki koji utiče na diskriminativnost, a to je težina stavke. Dobar test treba da ima takvu kombinaciju težine stavki da omogući najbolju diskriminativnost, odnosno normalnu raspodelu skorova. To obično podrazumeva izvestan broj lakih i teških zadataka i veći broj zadataka srednje težine.

Kako na podacima procenjujemo težinu stavki? Kod testova sa binarnim odgovorima (u koje spada najveći broj testova znanja i sposobnosti i neki testovi ličnosti), to je proporcija tačnih odgovora (R. J. Cohen & Swerdlik, 2018; DeVellis, 2006; Fajgelj, 2009). Što je ova proporcija niža, to je stavka teža, odnosno što je veći procenat ispitanika tačno uspeo da reši stavku, to je ona lakša. Težinu stavki u SPSS-u možemo najlakše dobiti preko **Analyze → Descriptive Statistics → Descriptives**, tako što ubacimo sve stavke u polje *Variables*, i u *Options* u delu *Display Order* izaberemo opciju *Descending means*, ukoliko želimo da stavke budu poređane od najlakše do najteže, ili *Ascending means*, ukoliko želimo da budu poređane od najteže do najlakše. Slika 1.16 pokazuje težinu stavki u našem testu znanja iz psihologije. Možemo videti da je najlakši ajtem, ajtem **t4**, rešilo 96% ispitanika (prosečna vrednost je .96). Ovaj ajtem je očigledno bio previše lak za dati uzorak ispitanika i doprineo je negativno asimetričnom obliku distribucije skorova. I sledeći najlakši ajtem, **t20**, ima prilično visok procenat tačnog rešavanja, kao i narednih nekoliko ajtema. Sa druge strane, najteži ajtem (**t11**) rešila je oko trećina ispitanika (težina je .34), pa se još jednom potvrđuje da težina ajtema nije bila idealno prilagođena nivou znanja kandidata.

Descriptive Statistics

	N	Mean	Std. Deviation
t4	436	.96	.194
t20	436	.86	.352
t1	436	.84	.365
t16	436	.83	.372
t7	436	.82	.384
t9	436	.81	.391
t14	436	.76	.425
t10	436	.71	.455
t5	436	.69	.461
t3	436	.64	.481
t19	436	.63	.483
t15	436	.63	.484
t12	436	.62	.487
t18	436	.60	.491
t2	436	.59	.492
t13	436	.57	.496
t6	436	.56	.497
t17	436	.51	.500
t8	436	.46	.499
t11	436	.34	.475
Valid N (listwise)	436		

Slika 1.16

Drugo važno svojstvo stavki koje utiče na diskriminativnost testa, pored težine, jeste njihova diskriminativnost. Diskriminativnost stavki može se procenjivati na sličan način kao i diskriminativnost celog testa. Sa jedne strane, možemo posmatrati distribuciju skorova za svaki pojedinačni skor, ali kod većine zadataka koji imaju tačne odgovore, ajtemi su binarnog tipa, te će distribucija teško nalikovati normalnoj. Zato za procenu diskriminativnosti binarnih stavki obično koristimo varijansu stavke, pri čemu se stavka sa većom varijansom skorova smatra diskriminativnijom. U slučaju testova ličnosti ili drugih testova gde se odgovori daju na višestepenoj Likertovoj skali, distribucije će više nalikovati normalnoj, ali broj stepeni skale je i dalje isuviše mali da bi omogućio pravu normalnu raspodelu. Kod stavki ovog tipa, za procenu diskriminativnosti možemo kombinovati dva pristupa – stavku smatramo utoliko diskriminativnijom ukoliko distribucija odgovora više nalikuje normalnoj distribuciji i ukoliko joj je varijansa veća.

Slika 1.16, pored aritmetičkih sredina stavki, uključuje i njihove standardne devijacije. Kako je standardna devijacija koren varijanse, ove vrednosti možemo koristiti za procenu diskriminativnosti stavki u testu. Ovde ne postoje univerzalni kriterijumi o tome šta su dobre ili loše vrednosti, već varijabilnost pojedinačne stavke treba da posmatramo u kontekstu dobijenih vrednosti za sve stavke u instrumentu. Možemo primetiti da je diskriminativnost većine stavki dobra jer postoji varijabilitet skorova, s tim što je varijansa nešto niža za najlakše ajteme (najniža varijansa je dobijena upravo za najlakši ajtem, **t4**). Ako znamo da se kod binarnih ajtema varijansa dobija kao V = p * q, gde je p proporcija tačnih odgovora, a q proporcija negativnih, jasno je da će maksimalna varijansa postojati za ajteme kod kojih je proporcija tačnih odgovora 0.5, te da će se smanjivati kako idemo ka lakšim i težim ajtemima, što prikazuje i Slika 1.16. Na primer, za stavku **t7**, proporcija tačnih odgovora je p = .82; samim tim proporcija netačnih odgovora je q = 1 − p = 1 - .82 = .18; varijansa stavke je onda V = p * q = .82 * .18 = .148, a standardna devijacija je koren ove vrednosti, odnosno SD = √V = √.148 = .384. Sa druge strane, za stavku **t17**, koja ima p = .51 i q = .49, varijansa je praktično maksimalna V = 0.2499, odnosno SD = 0.5.

Još jedan način da proverimo diskriminativnost stavki jeste preko njihove korelacije sa ukupnim skorom (R. J. Cohen & Swerdlik, 2018; DeVellis, 2006; Fajgelj, 2009). Ovaj postupak podrazumeva da će kvalitetna stavka dobro razlikovati, diskriminisati ispitanike s osobinom koju test meri od onih što ne poseduju datu osobinu. Iako se u ovoj knjizi nećemo detaljno baviti pojmom valjanosti (ili validnosti), važno je istaći da se ona odnosi na to u kojoj meri test adekvatno procenjuje željeni konstrukt. Korelacije stavke sa ukupnim skorom mogu se zato, u izvesnom smislu, smatrati i pokazateljem „interne valjanosti" stavki, ako pretpostavimo da test meri ono što bi trebalo da meri. Naravno, ovo možemo znati samo u slučaju da postoji i neki dokaz „eksterne valjanosti" samog testa.

U slučaju binarnih ajtema, korelacija stavke (binarne varijable) sa ukupnim skorom (kontinuiranom varijablom) jeste point-biserijska korelacija i može se izračunati prema sledećoj formuli:

$$r_{pb} = \frac{M_p - M_q}{\sigma_x} * \sqrt{p * q}$$

gde su M_p i M_q prosečni skorovi ispitanika koji su tačno i netačno odgovorili na datu stavku, σ_x standardna devijacija skora na testu, a p i q proporcija tačnih i netačnih odgovora.

Kako je point-biserijska korelacija numerički jednaka Pirsonovoj korelaciji binarne i kontinuirane varijable, za procenu korelacija ajtema sa ukupnim skorom u SPSS-u možemo iskoristiti opciju **Analyze → Correlate → Bivariate**. Korišćenjem sintakse i komande **WITH** (videti

Dodatak 1) možemo dobiti pregledan ispis (Slika 1.17)[6]. Primećujemo da sve stavke značajno pozitivno koreliraju sa ukupnim skorom, a raspon korelacija ide od niskih do srednje visokih vrednosti. Najnižu korelaciju sa testom ima stavka **t4**, r_{pb} = .164, što nije neočekivano imajući u vidu da je ovo najlakša stavka, koja je pokazala i najnižu diskriminativnost iskazanu preko varijanse stavke. Najviša korelacija dobijena je sa ajtemom **t15,** koji je imao umerenu težinu. Generalno ajtemi su imali umerene do umereno visoke korelacije sa ukupnim skorom, što ukazuje na dobru diskriminativnost / „unutrašnju valjanost" ajtema.

Jedan potencijalni problem sa point-biserijskom korelacijom skora na stavki i ukupnog skora jeste što je i data stavka ušla u računanje konačnog skora. Drugim rečima, korelirali smo varijablu sa linearnim kompozitom čiji je ona deo. Zbog toga se, umesto obične, često koristi korigovana korelacija stavke sa ukupnim skorom (Fajgelj, 2009). Korigovana korelacija dobija se prema sledećoj formuli:

$$r_{s(x-s)} = \frac{r_{sx}\sigma_x - \sigma_s}{\sqrt{\sigma_x^2 + \sigma_s^2 - 2r_{sx}\sigma_x\sigma_s}}$$

gde je r_{sx} korelacija stavke sa ukupnim skorom, a σ_x i σ_s standardna devijacija testa i stavke, redom. Ukoliko pogledamo samo korenovani izraz ispod razlomačke crte, vidimo da ga čini razlika varijansi testa i stavke i njihove dvostruke kovarijanse (pošto znamo da je kovarijansa jednaka proizvodu korelacije i standardnih devijacija varijabli koje koreliramo). Drugim rečima, korenovani izraz je prosto varijansa testa bez date stavke, tako da je ceo izraz ispod razlomačke crte standardna devijacija testa bez stavke sa kojom koreliramo ovaj skor. Iznad razlomačke crte nalazi se proizvod standardne devijacije testa i korelacije stavke i testa, odnosno onaj deo standardne devijacije (uslovno, varijanse) testa što korelira sa datom stavkom, umanjen za standardnu devijaciju stavke (koja je izbačena iz računanja ukupnog skora). Korigovana korelacija ajtema i testa bi u idealnom slučaju trebalo da bude preko .30, mada se u slučaju instrumenata sa široko definisanim predmetom merenja mogu dobiti i niže vrednosti (Fajgelj, 2009).

Correlations

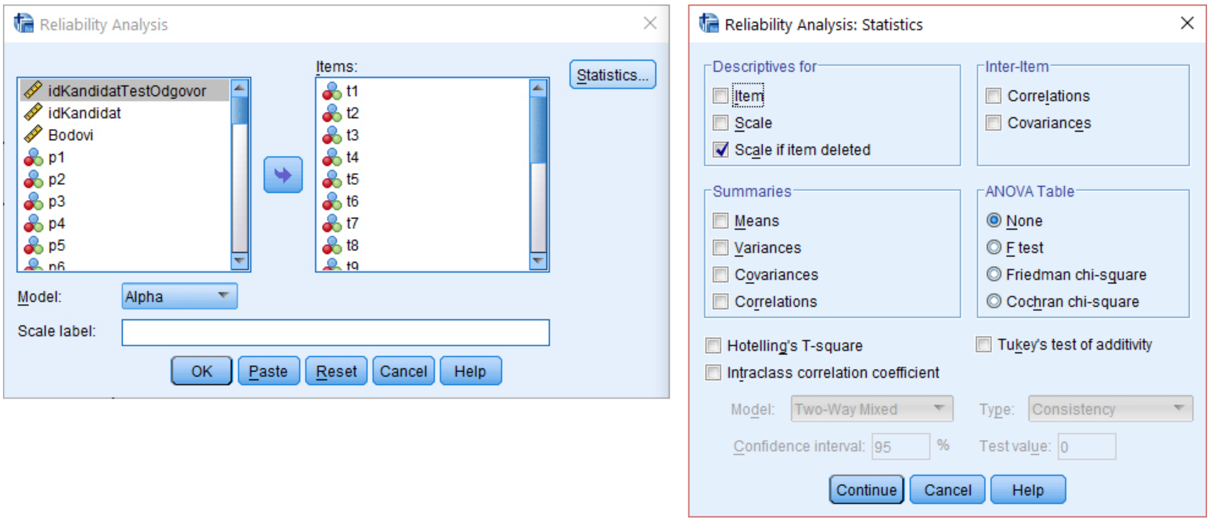
		TZ_20
t1	Pearson Correlation	.430**
t2	Pearson Correlation	.599**
t3	Pearson Correlation	.602**
t4	Pearson Correlation	.164**
t5	Pearson Correlation	.434**
t6	Pearson Correlation	.535**
t7	Pearson Correlation	.490**
t8	Pearson Correlation	.496**
t9	Pearson Correlation	.345**
t10	Pearson Correlation	.532**
t11	Pearson Correlation	.416**
t12	Pearson Correlation	.466**
t13	Pearson Correlation	.345**
t14	Pearson Correlation	.315**
t15	Pearson Correlation	.647**
t16	Pearson Correlation	.463**
t17	Pearson Correlation	.489**
t18	Pearson Correlation	.577**
t19	Pearson Correlation	.574**
t20	Pearson Correlation	.407**

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 1.17

Korigovana korelacija ajtema sa ukunim skorom nalazi se u okviru procedure **Analyze → Scale → Reliability Analysis** u SPSS-u. O pouzdanosti će biti više reči u posebnom poglavlju ove knjige, a za sada ćemo samo pokazati koje opcije treba izabrati kako bismo dobili korigovanu ajtem-total korelaciju (eng. *corrected item-total correlation*). U polje *Items* ubacujemo sve stavke na istrumentu, a zatim u okviru *Statistics* biramo *Descriptives for Scale if item deleted* opciju (Slika 1.18).

[6] Iz ovog ispisa su, radi preglednosti, dodatno obrisani redovi sa brojem ispitanika i značajnošću korelacija, pošto su vrednosti bile iste za sve korelacije: N = 436, p < .001.



Slika 1.18

U ispisu dobijamo nekoliko tabela, ali nas trenutno zanima samo četvrta kolona tabele *Item-Total Statistics*, u kojoj se nalaze korigovane korelacije svih ajtema sa ukupnim skorom (*Corrected Item-Total Correlation*). Kao što Slika 1.19 upućuje, ove korelacije razlikuju se od nekorigovanih, ali ne u prevelikoj meri. Pošto je stavka sada izbačena iz ukupnog skora, korigovane korelacije su obično niže od nekorigovanih, ali je njihov poredak po veličini veoma sličan. Ponovo, stavka koja najniže korelira sa ukupnim skorom je stavka **t4**, dok stavka **t15** ima najvišu korelaciju. Što je broj stavki u instrumentu veći, to će korigovane i nekorigovane korelacije biti međusobno sličnije, jer će „uticaj" pojedinačne stavke biti srazmerno manji, no korekciju u svakom slučaju treba primeniti (Fajgelj, 2009).

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
t1	12.59	16.573	.355	.811
t2	12.84	15.520	.516	.802
t3	12.80	15.555	.521	.802
t4	12.47	17.533	.119	.819
t5	12.74	16.289	.339	.812
t6	12.88	15.769	.443	.806
t7	12.61	16.325	.417	.808
t8	12.98	15.928	.399	.809
t9	12.62	16.778	.260	.815
t10	12.73	15.932	.447	.806
t11	13.09	16.324	.316	.813
t12	12.82	16.089	.368	.811
t13	12.87	16.566	.236	.818
t14	12.67	16.814	.220	.818
t15	12.81	15.360	.572	.799
t16	12.60	16.452	.389	.810
t17	12.92	15.951	.391	.809
t18	12.84	15.616	.491	.804
t19	12.80	15.662	.489	.804
t20	12.58	16.681	.333	.812

Slika 1.19

Osim korelacija ajtema sa ukupnim skorom, za diskriminativnost testa bitne su i korelacije između ajtema. Ovde se ponovo vraćamo na ideju o diskriminativnosti kao o varijansi skorova. Podsetimo se formule varijanse linearnog kompozita:

$$V(\sum_{i=1}^{m} X_i) = \sum_{i,j=1}^{m} Cov(X_i, X_j) = \sum_{i=1}^{m} V(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

gde je *m* broj ajtema u testu, a *X* varijabla koja predstavlja skor na ajtemu. Drugim rečima, varijansa linearnog kompozita (u ovom slučaju ukupnog skora na testu), predstavljena prvim izrazom, jednaka je kovarijansi svih varijabli (pojedinačnih ajtema) koje čine kompozit (drugi izraz). Ova kovarijansa uključuje i kovarijansu varijable sa samom sobom i kovarijanse različitih varijabli (pri čemu se svaka kovarijansa različitih varijabli računa dva puta, jednom kao kovarijansa *ij*, a drugi put kao kovarijansa *ji*). Ako ovo dalje razložimo (treći izraz u jednačini), videćemo da varijansu linearnog kompozita možemo razložiti na zbir varijansi pojedinačnih varijabli koje ga čine (pojedinačnih ajtema) i dvostruku kovarijansu svih parova različitih varijabli (ajtema). U matričnom obliku, varijansu linearnog kompozita možemo prikazati kao sumu elemenata matrice kovarijanse (Tabela 1.2), odnosno matrice korelacija kada su varijable date u standardizovanom obliku (Tabela 1.3).

Tabela 1.2. Matrica kovarijansi nestandardizovanih varijabli

σ_1	$r_{21}\,\sigma_2\,\sigma_1$	$r_{31}\,\sigma_3\,\sigma_1$	$r_{41}\,\sigma_4\,\sigma_1$	…	$r_{i1}\,\sigma_i\,\sigma_1$
$r_{12}\,\sigma_1\,\sigma_2$	σ_2	$r_{32}\,\sigma_3\,\sigma_2$	$r_{42}\,\sigma_4\,\sigma_2$	…	$r_{i2}\,\sigma_i\,\sigma_2$
$r_{13}\,\sigma_1\,\sigma_3$	$r_{23}\,\sigma_2\,\sigma_3$	σ_3	$r_{43}\,\sigma_4\,\sigma_3$	…	$r_{i3}\,\sigma_i\,\sigma_3$
$r_{14}\,\sigma_1\,\sigma_4$	$r_{24}\,\sigma_2\,\sigma_4$	$r_{34}\,\sigma_3\,\sigma_4$	σ_4	…	$r_{i4}\,\sigma_i\,\sigma_4$
…	…	…	…	…	…
$r_{1i}\,\sigma_1\,\sigma_i$	$r_{2i}\,\sigma_2\,\sigma_i$	$r_{3i}\,\sigma_3\,\sigma_i$	$r_{4i}\,\sigma_4\,\sigma_i$	…	σ_i

Tabela 1.3. Matrica kovarijansi standardizovanih varijabli

1	r_{21}	r_{31}	r_{41}	…	r_{i1}
r_{12}	1	r_{32}	r_{42}	…	r_{i2}
r_{13}	r_{23}	1	r_{43}	…	r_{i3}
r_{14}	r_{24}	r_{34}	1	…	r_{i4}
…	…	…	…	…	…
r_{1i}	r_{2i}	r_{3i}	r_{4i}	…	1

Ako znamo da je ukupan skor na testu linearni kompozit skorova na pojedinačnim stavkama, jasno je da će varijansa testa biti utoliko veća ukoliko su kovarijanse, odnosno korelacije ajtema veće.

Ovde pak dolazimo do izvesnog paradoksa (Fajgelj, 2009). Rekli smo da sa povećavanjem kovarijansi između ajtema raste i diskriminativnost testa. Šta će se, međutim, desiti, ukoliko korelacije između ajtema dostignu maksimalnu teorijsku vrednost od jedan? Ilustrovaćemo ovo primerom testa koji ima samo dva ajtema. Mogući skorovi na testu koji ima dva ajtema su 0, 1 i 2. Međutim, ukoliko su ajtemi savršeno korelirani, onda će svako ko je tačno odgovorio na prvi ajtem ujedno tačno odgovoriti i na drugi ajtem. To znači da neće biti ispitanika koji su tačno odgovorili na jedan, ali ne i na drugi ajtem, odnosno skor 1 empirijski neće postojati, a jedini skorovi na testu će biti 0 i 2. Ukoliko zamislimo da umesto 2 imamo 10 ajtema, situacija je veoma slična. Test koji ima 10 ajtema omogućava dobijanje 11 različitih skorova (0–10). Ali ako su ajtemi savršeno korelirani, onda će ispitanik koji je tačno odgovorio na prvi ajtem imati tačne odgovore i na preostalih 9 ajtema, čime će ostvariti maksimalni skor na testu. Sa druge strane, ispitanik koji je netačno odgovorio na prvi ajtem će netačno odgovoriti i na sve druge i postići minimalni skor. Dakle, raspodela skorova na testu na kom su ajtemi savršeno korelirani biće bimodalna, a ne normalna, čime se značajno narušava diskriminativnost takvog testa. U ovoj hipotetičkoj situaciji, jedan ajtem je dovoljan da obezbedi razlikovanje ispitanika, a svi drugi ajtemi postaju redundantni zbog toga što ne dodaju nove informacije u odnosu na one koje nam je već pružio prvi ajtem. Zato je u praksi poželjno da korelacije između ajtema budu što veće, ali ne i maksimalne. Zapravo, smatra se i da nije poželjno da korelacije ajtema budu veće od .80, kako stavke ne bi bile previše međusobno slične, odnosno redundantne.

1.5.3) Način ocenjivanja i diskriminativnost testa

Pod načinom ocenjivanja ovde podrazumevamo način izračunavanja konačnog rezultata na testu. U većini slučajeva konačni skor se računa kao suma skorova na pojedinačnim pitanjima ili zadacima. Moguće je, međutim, ukupan skor izračunati i na druge načine (Fajgelj, 2009). Ovo se, pre svega, odnosi na negativne poene i korekcije za pogađanje.

Negativni poeni se dobijaju za netačne odgovore na nekim ili svim zadacima u testu. Na primer, na testu znanja iz psihologije postoji 10 DA/NE pitanja, gde je verovatnoća slučajnog pogađanja 50%. Kada ne bismo primenjivali korekciju za pogađanje, ispitanici koji ne poseduju nikakvo znanje bi, u proseku, na osnovu slučajnog pogađanja ostvarili 5 poena na ovom delu testa. Uz korekciju za pogađanje, ispitanici koji ne poseduju znanje bi trebalo da osvoje 0 poena (jer će 5 netačnih odgovora poništiti 5 slučajno tačnih odgovora). Ukoliko konstruktor testa proceni da je neko drugo rešenje optimalno, negativni poeni mogu imati i manji ponder od jedinice, te tako jedan netačan odgovor može oduzimati polovinu ili četvrtinu poena koje nosi jedan tačan odgovor (Bukvić, 1996).

Korekciju za pogađanje možemo primeniti i u slučaju zadataka koji imaju više od dva ponuđena odgovora, kao i više od jednog tačnog, primenjujući sledeću formulu:

$$X = T - \frac{P}{k - a}$$

gde je X ukupan skor na testu, T broj tačnih, P broj netačnih odgovora, k broj ponuđenih odgovora za svako pitanje, a a broj odgovora koje treba obeležiti (Diamond & Evans, 1971). Vidimo da je korekcija u slučaju binarnih stavki samo poseban slučaj prethodne formule, gde je izraz $k – a = 2 – 1 = 1$, pa se ukupan skor računa kao X = T – P.

Još jedan aspekt načina ocenjivanja s uticajem na diskriminativnost testa jeste da li se priznaju delimično tačni odgovori ili je potrebno da ispitanik tačno odgovori na sve delove / aspekte pitanja da bi dobio poen. Priznavanje delimično tačnih odgovora se zapravo svodi na to da imamo više ajtema unutar jednog, pri čemu ti ajtemi nose neku proporciju poena celog ajtema.

Iz svega navedenog očigledno je da način ocenjivanja može bitno promeniti oblik distribucije skorova na testu. Kako ne postoje univerzalne preporuke za upotrebu najboljeg načina ocenjivanja, istraživač / konstruktor testa mora pažljivo da razmotri sve prednosti i nedostatke metode koju planira da primeni (R. J. Cohen & Swerdlik, 2018).

1.5.4) Pouzdanost i diskriminativnost testa

O pouzdanosti će biti više reči u jednom od narednih poglavlja, ali je za sada možemo definisati kao stepen preciznosti merenja. Što je pouzdanost veća, to je preciznost veća, a greška merenja manja, odnosno pravimo manju grešku prilikom procene pravog skora ispitanika na osnovu njegovog dobijenog skora. Greška merenja određuje interval poverenja u kom verujemo da se nalazi pravi skor ispitanika, sa određenim stepenom verovatnoće (obično 95% ili 99%). Što je greška merenja manja, to je ovaj interval uži, a stvarna diskriminativnost testa veća (R. J. Cohen & Swerdlik, 2018).

Ilustrujmo ovo na primeru testa koji meri inteligenciju. Recimo da je jedan ispitanik postigao skor od 120 IQ jedinica, dok je drugi ispitanik postigao skor od 123 IQ jedinice. Da li se ova dva ispitanika zaista razlikuju po stepenu inteligencije? Odgovor na ovo pitanje zavisi od pouzdanosti testa kojim je izmerena inteligencija, odnosno od greške merenja. Ako pretpostavimo da je greška merenja veoma mala, interval poverenja može obuhvatati, recimo, dva IQ skora. To bi značilo da se pravi skor prvog ispitanika nalazi u rasponu od 119 do 121, dok se skor drugog ispitanika nalazi u rasponu od 122 do 124 jedinice. Kako se intervali ne preklapaju, zaključujemo da se ova dva ispitanika, odnosno ova dva skora značajno razlikuju. U slučaju veće greške merenja i šireg intervala poverenja, od, na primer, četiri IQ jedinice, skor prvog ispitanika bio bi u rasponu od 118 do 122, a skor drugog ispitanika u rasponu 121 do 125 IQ jedinica. Kako se ovi intervali preklapaju, postoji mogućnost da je pravi skor dva ispitanika isti. Drugim rečima, sa većom greškom merenja razlika između ova dva skora nije značajna.

Zaključujemo da veća pouzdanost vodi boljoj stvarnoj diskriminativnosti testa, ali odnos diskriminativnosti i pouzdanosti nije jednosmeran. Da bi test pouzdano merio neki konstrukt, pre svega mora registrovati postojeće razlike između ispitanika, odnosno mora biti diskriminativan, pa se može reći i da je diskriminativnost preduslov pouzdanosti (Ebel, 1967).

1.6) Rezime

Diskriminativnost testa kao metrijsku karakteristiku možemo posmatrati iz nekoliko uglova – kao diskriminativnu moć instrumenta, njegovu varijansu ili oblik distribucije. Najčešće se rukovodimo time da dobra diskriminativnost odgovara normalnoj distribuciji skorova na instrumentu, jer očekujemo da se skorovi na najvećem broju psiholoških varijabli normalno distribuiraju. Postoje brojni načini provere odstupanja diskriminativnosti od normalne, koji uključuju kako vizuelne – histogram, tako i statističke – hi-kvadrat test, Kolmogorov–Smirnov test, standardizovani skjunis i kurtozis. Različiti pokazatelji normalnosti distribucije neće se uvek savršeno poklapati, u kom slučaju je potrebno doneti konačni zaključak na osnovu svih dostupnih informacija, uzimajući u obzir način na koji se svaki od pokazatelja dobija.

Diskriminativnost testa zavisi od više faktora, uključujući broj, težinu i diskriminativnost stavki, te korelacije između njih, kao i način ocenjivanja testa i njegovu pouzdanost. Ukoliko se pokaže da distribucija testa odstupa od normalne, u fazi konstrukcije instrumenta je moguće izvršiti korekcije u samom instrumentu za koje očekujemo da će voditi boljoj diskriminativnosti. Ukoliko ne postoji mogućnost izmene samih stavki, a odstupanja od normalnosti su umerena, moguće je izvršiti i statističku normalizaciju skorova.

1.7) Preporučena literatura

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology*, *9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Bukvić, A. (1996). *Načela izrade psiholoških testova*. Zavod za udžbenike i nastavna sredstva, Beograd.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment: An Introduction to Tests and Measurement. Ninth Edition*. New York, NY: McGraw-Hill Education.

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Ferguson, G. A. (1949). On the Theory of Test Discrimination. *Psychometrica*, *14*(1), 61–68.

Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of the American Statistical Association*, *64*(325), 387–389.
Tenjović, L. (2000). *Statistika u psihologiji: priručnik*. Beograd: Centar za primenjenu psihologiju Društva psihologa Srbije.

2) OBJEKTIVNOST

Psihometrijsko značenje objektivnosti odgovara „svakodnevnoj” upotrebi ovog pojma. Test je objektivan u onoj meri u kojoj rezultat na testu ne zavisi od onoga koji je (pr)ocenjivao osobinu ispitanika (Bukvić, 1996). Ako se ponovo poslužimo primerom testa znanja iz psihologije koji se polaže kao deo prijemnog ispita za upis psihologije, jasno je da je test tako sačinjen da bi svako ko ga ocenjuje prema unapred definisanom ključu došao do istog broja poena za svakog kandidata. Prijemni ispiti na nekim drugim fakultetima, međutim, po svojoj prirodi ne mogu biti savršeno objektivni. Ako želimo da procenimo talenat kandidata za glumu, ples ili vizuelne umetnosti, teško ćemo osmisliti i implementirati sasvim objektivan test. Zato se, u ovakvim situacijama, obično pribegava komisijskom ocenjivanju postignuća kandidata, te više procenjivača nezavisno ocenjuje svakog kandidata, a zatim se te ocene integrišu u konačnu ocenu – broj poena kandidata.

2.1) Izvori varijabilnosti testovnih skorova

Da bismo bolje razumeli psihometrijski koncept objektivnosti (a kasnije i pouzdanosti), treba, pre svega, da razmotrimo izvore varijabilnosti skorova na testu. Kada zadajemo test i interpretiramo rezultate dobijene na osnovu testa, obično razmišljamo samo o izvoru varijabiliteta koji potiče od ispitanika. Drugim rečima, skor koji je neko postigao na testu tumačimo kao (isključivi) rezultat znanja, sposobnosti ili osobina ličnosti tog ispitanika. Zapravo, svaki skor odražava nekoliko različitih faktora (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009):

A) Merenu osobinu (predmet merenja) i

B) Grešku merenja, koja se može razložiti na dve komponente

   a. Sistematsku grešku merenja, odnosno merni postupak (način zadavanja i ocenjivanja)

   b. Slučajnu grešku merenja

U sistematsku grešku merenja ulaze različiti faktori koji sistematski utiču na rezultate ispitanika. To, na primer, mogu biti nejednaki uslovi zadavanja testa različitim grupama ispitanika. Ako jednoj grupi ispitanika damo više vremena za rešavanje testa, njihovi skorovi će biti viši od skorova grupe koja je imala kraće vreme za rad. Takođe, ukoliko je test tako konstruisan da je pristrasan prema nekim grupama, skorovi će odražavati ovaj sistematski faktor. Test čiji je sadržaj bio unapred poznat delu ispitanika takođe će dovoditi do razlike u skorovima onih ispitanika upoznatih sa testom pre rešavanja i onih bez takvog „predznanja”. O ovim faktorima je važno voditi računa prilikom konstrukcije testa kako bi se predupredile sistematske greške. Podrazumeva se i da se moraju obezbediti jednaki uslovi rada za sve ispitanike. O slučajnoj grešci će biti više reči u odeljku o pouzdanosti, a za sada je možemo posmatrati kao sve one faktore koji deluju na ispitanike, a istraživaču po pravilu nisu poznati, niti deluju na identičan način na sve ispitanike.

Ako ovako definišemo sistematsku i slučajnu grešku merenja, postavlja se pitanje da li se objektivnost merenja odnosi na sistematsku ili pak slučajnu grešku merenja. Najtačnije bi bilo reći da ovo zavisi od konkretne situacije testiranja, a prevashodno od broja procenjivača. Zamislimo, na primer, da su dve grupe ispitanika ocenjivala dva procenjivača s različitim kriterijumima ocenjivanja – ovaj scenario se dobro uklapa u definiciju sistematske greške, jer će jedna grupa imati konzistentno niže/više skorove od druge grupe. Ukoliko je, nasuprot tome, svakog ispitanika ocenjivao drugi procenjivač – uticaj „strogosti" procenjivača na skor ispitanika nije poznat istraživaču i neće delovati na isti način na sve ispitanike, što bi odgovaralo definiciji slučajne greške. Važno je napomenuti i da se slaganje procena različitih procenjivača često razmatra kao jedan aspekt pouzdanosti, koji se nekad naziva i „pouzdanošću procenjivača" (eng. *inter-rater reliability* ili *inter-scorer reliability*) (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Zapravo, većina metoda procene objektivnosti podrazumeva da je nekoliko procenjivača ocenjivalo iste ispitanike – te grešku u proceni tretiraju kao slučajnu. U situacijama kada samo jedan procenjivač procenjuje svakog od ispitanika, o objektivnosti je bolje razmišljati u kontekstu smanjivanja uticaja sistematske greške.

Za razliku od nekih od prethodno navedenih izvora sistematske greške, način na koji će ocenjivači ocenjivati odgovore ispitanika je nešto teže kontrolisati. Ono što možemo i što je svakako preporučljivo učiniti jeste jasno definisati ključ za ocenjivanje odgovora i/ili obezbediti obuku za procenjivače i dati im priliku da razlike u svojim ocenama prodiskutuju (R. J. Cohen & Swerdlik, 2018), no ni ovo neće nužno dovesti do savršene objektivnosti ocena. Navedeno nikako ne znači da testovi (i drugi vidovi psihološke procene) koji nisu savršeno objektivni ne mogu biti dobri, već samo ističe potrebu za procenom objektivnosti i tumačenjem rezultata spram nivoa objektivnosti procene.

2.2) Vidovi procene objektivnosti

Na šta tačno mislimo kada kažemo da ocene različitih procenjivača treba da budu objektivne? Kako možemo kvantifikovati objektivnost? Može delovati kao da je neka vrsta korelacije ocena procenjivača najbolja mera objektivnosti, ali lako možemo zamisliti situaciju u kojoj jedan procenjivač daje ocene 1, 1, 2, 2, 3, 4, 4, a drugi procenjivač daje ocene 2, 2, 3, 3, 4, 5, 5. Korelacija između ova dva niza ocena je savršena, ali ocenivanje nije objektivno jer prvi procenjivač konzistentno daje ocenu manje svakom ispitaniku od drugog procenjivača. Dakle, u ovom primeru postoji korelacija ocena procenjivača, ali postoje i razlike između njih, te merenje nije objektivno. Sa druge strane, očigledno je da ni puko nepostojanje značajnih razlika između procenjivača samo po sebi ne garantuje objektivnost. Na primer, u situaciji kada jedan procenjivač daje ocene 1, 1, 2, 2, 3, 4, 4, dok drugi daje ocene 4, 1, 3, 2, 2, 4, 1, prosečna ocena koju su dala oba procenjivača je jednaka, ali između njihovih ocena ne postoji (praktično) nikakvo slaganje. Drugim rečima, ako ne postoje razlike u prosečnim procenama, ali nema ni korelacije između njihovih procena – merenje takođe neće biti objektivno (Bukvić, 1996).

U literaturi se mogu sresti dva termina kojima se označava objektivnost procene. Jedan je već pomenuta „pouzdanost" procenjivača (eng. *inter-rater reliability, IRR*), a drugi je slaganje procenjivača (eng. *inter-rater agreement, IRA*). Iako se ovi termini ponekad koriste kao sinonimi, između njih postoji važna distinkcija – pouzdanost procenjivača odnosi se samo na korelacije između njihovih procena, dok u slučaju slaganja procenjivača govorimo istovremeno i o korelaciji i o nepostojanju razlika između procenjivača (Graham et al., 2012; Shweta et al., 2015). U određenim (najčešće istraživačkim) kontekstima, može se zamisliti da je dovoljno pokazati da postoji pouzdanost procenjivača; recimo, važno je samo da se utvrdi relativni rang ispitanika koje procenjujemo (ko je bolji, a ko lošiji na nekom testu), a konkretne vrednosti procene su od sekundarne važnosti. Ipak, u najvećem broju situacija postoji potreba i da se procena „veže" za neke konkretne ocene (ko je dobro, a ko loše uradio test), te su mere slaganja procenjivača od veće važnosti.

Procena objektivnosti za konkretan merni instrument zavisiće od specifičnosti podataka za koje utvrđujemo objektivnost. Na raspolaganju su nam brojne metode zavisne ne samo od vida objektivnosti koji procenjujemo, već i od broja procenjivača i nivoa merenja. Neke od metoda odnose se samo na procenu „pouzdanosti" procenjivača, te ih je obično potrebno dopuniti metodama utvrđivanja (ne)postojanja razlika između procenjivača, a neke obuhvataju obe komponente objektivnosti, te se odnose na slaganje procenjivača. Tabela 2.1 daje pregled najčešćih metoda pogodnih za procenu objektivnosti testa.

Tabela 2.1. Pregled najčešćih metoda za utvrđivanje objektivnosti testa

Nivo merenja	Vid procene objektivnosti	
	„Pouzdanost" procenjivača	Slaganje procenjivača
Intervalni	Pirsonova korelacija, Intraklasni koeficijent korelacije	Intraklasni koeficijent korelacije
Ordinalni	Spirmanova korelacija, Kendalovo W, Intraklasni koeficijent korelacije	Intraklasni koeficijent korelacije, Ponderisani kapa koeficijent, Fleisov kapa koeficijent
Nominalni	Procenat slaganja, Kapa koeficijent	Procenat slaganja, Kapa koeficijent

2.3) Procena objektivnosti za intervalni nivo merenja

Zamislimo da smo ispitanicima dali tzv. „zadatak upotrebe", u kome treba da navedu što više što kreativnijih načina upotrebe nekog svakodnevnog predmeta, na primer, kašike. Iako je u teoriji moguće osmisliti ključ za ocenjivanje ovakvog testa, nemoguće je predvideti sve pojedinačne odgovore koje ispitanici mogu produkovati. Zato se prilikom bodovanja obično pribegava ocenjivanju svakog odgovora od strane nekoliko procenjivača. Pretpostavimo da je pet procenjivača ocenjivalo kreativnost 50 odgovora ispitanika na skali od 1 (nimalo kreativno) do 5 (izuzetno kreativno). Kako u ovom primeru možemo proceniti objektivnost ocena koje smo dobili?

Objektivnost merenja bismo u ovom slučaju mogli utvrditi kombinacijom Pirsonove korelacije, kojom utvrđujemo „pouzdanost" procenjivača, i ANOVA-e za ponovljena merenja, za utvrđivanje postojanja razlika u njihovim procenama. Da smo imali samo dva procenjivača, pored ANOVA-e, prikladno bi bilo računati i t-test za zavisne uzorke. Ipak, najprimereniji način procene objektivnosti za intervalne podatke jeste primena intraklasnog koeficijenta korelacije (eng. *Intraclass Correlation Coefficient*, ICC), koji će nam dati informaciju o stepenu slaganja procenjivača.

2.3.1) Pirsonova korelacija

Najjednostavniji način da procenimo „pouzdanost" procenjivača jeste da koreliramo njihove procene i izračunamo prosečnu vrednost. U našem primeru postoji pet procenjivača, što znači da ćemo u matrici korelacija dobiti 10 jedinstvenih korelacija koje zatim treba da uprosečimo.

Koristeći opciju **Analyze → Correlate → Bivariate**, uz ubacivanje procena svih pet procenjivača u polje Variables, dobijamo tabelu koju ilustruje Slika 2.1. Prosečni koeficijent korelacije dobijamo uprosečavanjem svih 10 jedinstvenih korelacija[7] $\bar{r} = \frac{.318+.578+.510+.608+.554+.645+.455+.516+.377+.501}{10} = .506$.

Correlations

		R1	R2	R3	R4	R5
R1	Pearson Correlation	1	.318*	.578**	.608**	.455**
	Sig. (2-tailed)		.024	.000	.000	.001
	N	50	50	50	50	50
R2	Pearson Correlation	.318*	1	.510**	.554**	.516**
	Sig. (2-tailed)	.024		.000	.000	.000
	N	50	50	50	50	50
R3	Pearson Correlation	.578**	.510**	1	.645**	.377**
	Sig. (2-tailed)	.000	.000		.000	.007
	N	50	50	50	50	50
R4	Pearson Correlation	.608**	.554**	.645**	1	.501**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	50	50	50	50	50
R5	Pearson Correlation	.455**	.516**	.377**	.501**	1
	Sig. (2-tailed)	.001	.000	.007	.000	
	N	50	50	50	50	50

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 2.1

Kako možemo tumačiti dobijenu vrednost? Pre svega, poznato nam je da koeficijent korelacije može uzimati vrednosti između -1 i 1. U slučaju potpuno objektivne procene, korelacije svih procenjivača trebalo bi da budu savršene i pozitivne, odnosno da iznose 1. Samim tim i prosečna korelacija, koja se u ovom kontekstu može nazivati koeficijentom ili indeksom objektivnosti (Bukvić, 1996), takođe bi trebalo da iznosi 1. Ne postoje jasne preporuke o graničnim vrednostima dobre objektivnosti merenja kada je iskazana kao prosečna korelacija, ali svakako treba težiti što većoj vrednosti, odnosno što većoj sličnosti procena različitih procenjivača. Kao što vidimo, u konkretnom primeru koji smo prikazali objektivnost nije na zadovoljavajućem nivou, s obzirom na to da je koeficijent od .50 prilično udaljen od idealne vrednosti.

Kako bismo imali potpune podatke o objektivnosti testa, bilo bi neophodno da proverimo i da li između procenjivača ima značajnih razlika u procenama. U slučaju više procenjivača, iskoristili bismo ANOVA-u za ponovljena merenja (videti Dodatak 2), dok bismo u slučaju dva procenjivača koristili t-test za zavisne uzorke.

[7] Zapravo, ispravna procedura bi podrazumevala Fišerovu Z-transformaciju, gde se korelacije prvo pretvore u standardne skorove, zatim se tako dobijeni standardni skorovi uproseče, a onda se rezultujući Z skor pretvori „nazad" u korelaciju. Kako su razlike u vrednostima male, za potrebe ilustracije ovde smo direktno uprosečili koeficijente korelacije.

2.3.2) Intraklasni koeficijent korelacije (ICC)

Intraklasni koeficijent korelacije (ICC predstavlja jednu od najčešće korišćenih tehnika procene „pouzdanosti" i slaganja procenjivača. Iako je inicijalno nastao kao modifikacija Pirsonovog koeficijenta korelacije, ICC je zapravo zasnovan na analizi varijanse, te daje i procenu stepena koreliranosti procena procenjivača i procenu stepena njihovog slaganja (nepostojanja razlika) (Koo & Li, 2016; Shweta et al., 2015). ICC teorijski uzima vrednosti između 0 i 1, a način njegovog računanja je takav da se vrednost ICC može tumačiti kao proporcija ukupne varijanse koja se može pripisati objektu procene (Liljequist et al., 2019). Neke starije preporuke navode da su vrednosti ICC ispod .40 loše, između .40 i .59 umerene, između .60 i .74 dobre i preko .75 odlične (Cicchetti, 1994). Ipak, teško da bismo situaciju u kojoj više od polovine varijanse „odlazi" na procenjivače mogli okarakterisati kao „umerenu" objektivnost merenja. Ovo je reflektovano i u nekim novijim konvencijama, prema kojima treba obratiti pažnju na 95% interval poverenja za ICC, a za koji se vrednosti ispod .50 smatraju niskom objektivnošću, vrednosti između .50 i .75 umerenom, između .75 i .90 dobrom, a preko .90 odličnom objektivnošću, odnosno slaganjem procenjivača (Koo & Li, 2016).

Postoji čak 10 različitih načina računanja intraklasnog koeficijenta korelacije za iste podatke u zavisnosti od **modela**, **tipa** i **definicije** koje upotrebljavamo (Denham, 2017; Koo & Li, 2016). Kako se vrednosti mogu bitno razlikovati u zavisnosti od odabira koeficijenta, važno je da razumemo razlike između njih i na osnovu toga izaberemo onaj najprimereniji ne samo podacima koje imamo već i budućoj nameni testa.

Postoje tri osnovne klase ICC **modela** (Koo & Li, 2016; Shrout & Fleiss, 1979), a koju ćemo izabrati u konkretnom slučaju zavisi od odgovora na dva pitanja: 1) da li su sve objekte ocenjivali isti procenjivači ili ne? i 2) da li su procenjivači slučajno izvučeni iz šire populacije potencijalnih procenjivača ili oni čine celu populaciju? Tri klase ICC modela su:

1. **Model jednosmernih slučajnih efekata** (eng. *one-way random effects model*). Ukoliko su različite objekte ocenjivali različiti procenjivači, onda je jednosmerni slučajni model najprimereniji. U našem primeru ovo bi odgovaralo situaciji u kojoj postoji ukupno 10 procenjivača, ali je svaki odgovor procenjivalo nasumice izabranih pet. Tako se može desiti da odgovor „za jedenje supe" procenjuje istih pet procenjivača kao i odgovor „za mešanje kafe", da postoji delimično preklapanje u procenjivačima, ali i da svi procenjivači koji procenjuju kreativnost ova dva odgovora budu različiti. Prilikom korišćenja ove procedure dobro je kontrabalansirati procenjivače i objekte procene tako da svaki procenjivač ima jednak broj procena (npr., ako imamo 50 odgovora koje treba da proceni po 5 procenjivača od ukupno 10, svaki procenjivač treba da napravi 50*5/10 = 25 procena). Pošto su procenjivači i objekti procene na ovaj način ukršteni, a ponekad čak i nemamo informaciju koji je tačno procenjivač procenjivao određeni objekat, efekti ova dva izvora variranja se ne mogu precizno razdvojiti. Zbog toga se model naziva jednosmernim, jer se procenjivači i objekti procene tretiraju kao jedan, zajednički efekat. Ovaj efekat je slučajan jer se pretpostavlja da su i objekti procene i procenjivači slučajno uzorkovani iz šire populacije. U poređenju sa druga dva modela, jednosmerni slučajni ICC će uvek imati najnižu vrednost[8].

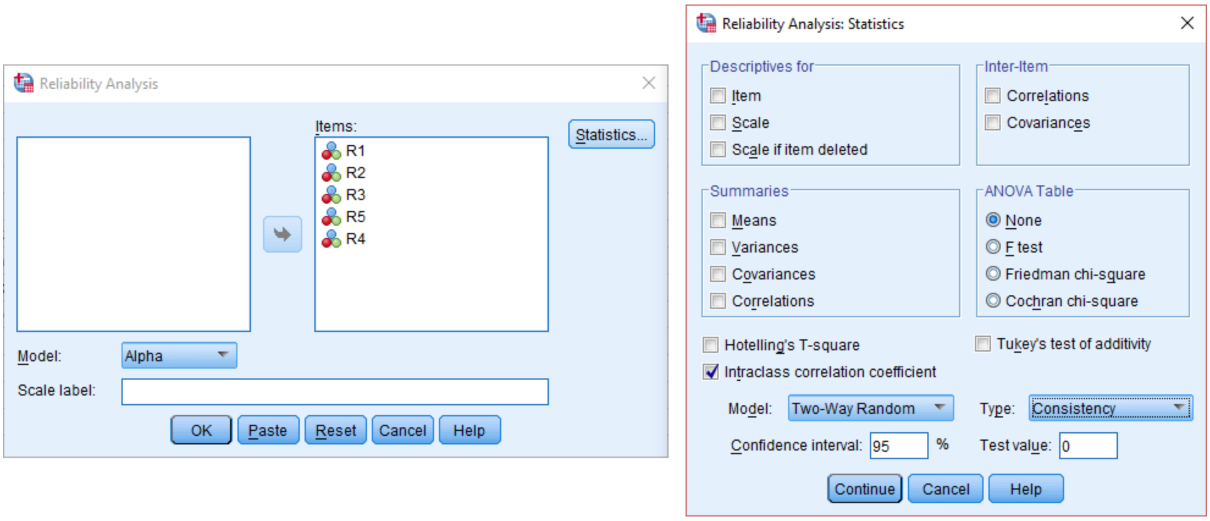
[8] Jednosmerni slučajni model uvek daje najniže vrednosti ICC upravo zato što ne može da razdvoji efekat procenjivača i objekta procene. To znači da se u jednosmernom modelu deo varijanse koja se odnosi na skorove ispitanika „meša" sa varijansom procenjivača i ne može se od nje statistički razdvojiti. Nasuprot tome, kod dvosmernih efekata se efekti procenjivača i objekata procene mogu razdvojiti, što čini procenu greške preciznijom.

2. **Model dvosmernih slučajnih efekata** (eng. *two-way random effects model*). Ukoliko su sve objekte procenjivali isti procenjivači, pri čemu su ti procenjivači slučajno izvučeni iz šire populacije potencijalnih procenjivača – treba upotrebiti dvosmerni slučajni model. U našem primeru, svaki odgovor je ocenjivalo istih pet procenjivača, pri čemu ovi procenjivači nisu jedini mogući koji bi mogli vršiti procene – lako je moguće zamisliti i da smo imali druge procenjivače izvučene iz iste populacije. Zato je u ovom slučaju model dvosmernih slučajnih efekata najprimereniji. Naziv modela potiče od toga što se procenjivači i objekti procene u modelu tretiraju kao odvojeni izvori varijanse, odnosno kao odvojeni efekti – te se procenjuju ukupno dva slučajna efekta. Kao i za prethodni model, svaki od ovih efekata posmatramo kao slučajan, pošto pretpostavljamo da su i objekti procene i procenjivači slučajno izvučeni iz populacije.
3. **Model dvosmernih mešovitih efekata** (eng. *two-way mixed effects model*). Ukoliko su sve objekte procenjivali isti procenjivači i ovo su jedini procenjivači od interesa za istraživača, odnosno predstavljaju celokupnu populaciju procenjivača, onda se koristi model dvosmernih mešovitih efekata. U primeru koji smo dali, ovo bi odgovaralo (malo verovatnoj hipotetičkoj) situaciji gde je pet procenjivača koje smo zamolili da procene odgovore ispitanika zapravo pet vrhunskih stručnjaka za kreativnost, stalnih članova odbora za donošenje odluka o nagradama za kreativna postignuća. U takvoj situaciji, procene ovih pet procenjivača su jedine relevantne i nijedan procenjivač ne bi mogao biti zamenjen nekim drugim, jer suštinski predstavljaju jedinu populaciju značajnu za nas. Model se naziva dvosmernim mešovitim zato što se posmatraju efekti objekta procene koji su slučajni i efekti procenjivača koji su fiksirani (jer procenjivači nisu slučajno uzorkovani iz populacije). ICC dobijen u dvosmernom mešovitom modelu će uvek biti numerički identičan onom dobijenom u dvosmernom slučajnom modelu, ali se njihova interpretacija, pre svega generalizacija, razlikuje. Dok za dvosmerni slučajni ICC možemo očekivati da će biti repliciran i na drugom uzorku od pet procenjivača koji procenjuju druge kreativne odgovore ispitanika, za dvosmerni mešoviti model ovo ne važi.

Pored tri osnovne klase ICC, važno je napraviti i distinkciju između dva **tipa** ICC koeficijenta. Koji tip koeficijenta ćemo upotrebiti zavisi od odgovora na pitanje „Da li će u narednim primenama testa odgovore ocenjivati jedan procenjivač ili nekoliko?" (Koo & Li, 2016). U našem primeru, ukoliko smo pet procenjivača angažovali samo kako bismo proverili stepen objektivnosti testa, a ubuduće planiramo da se oslanjamo na procenu jednog procenjivača – koristićemo tip ICC koeficijenta za jednog procenjivača (eng. *single rater*). Ukoliko pak planiramo da pri svakoj upotrebi testa do konačnog rezultata ispitanika dođemo uz pomoć procena 5 procenjivača – koristićemo ICC za prosek procenjivača (eng. *mean of raters*). Jasno je, naravno, da će objektivnost uvek biti utoliko veća ukoliko imamo više nezavisnih procenjivača (jer će se subjektivne razlike tako u najvećoj meri „nivelisati" i poništiti), ali je u određenim situacijama u praktičnom smislu neizvodljivo da veći broj ljudi vrši procenu. U takvim okolnostima, potrebno je posebnim treningom procenjivača obezbediti visoku objektivnost svakog procenjivača pojedinačno.

Konačno, potrebno je da odlučimo koju od dve **definicije** ICC koeficijenata želimo da koristimo – konzistentnost ocena procenjivača (eng. *consistency*) ili apsolutno slaganje među procenjivačima (eng. *absolute agreement*) (Koo & Li, 2016). Apsolutno slaganje se odnosi na to da li su procenjivači dali identične ocene različitim objektima procene, dok se konzistentnost odnosi samo na to da li su ocene bile usaglašene (ako jedan procenjivač daje višu ocenu nekom objektu, onda to čini i drugi procenjivač). ICC kao konzistentnost zapravo se odnosi na koncept objektivnosti kao „pouzdanosti” procenjivača (eng. *inter-rater reliability*), dok se ICC kao apsolutno slaganje odnosi na objektivnost kao slaganje procenjivača (eng. *inter-rater agreement*). Ukoliko će konačna procena biti zasnovana na uprosečavanju ili sabiranju ocena različitih procenjivača, može se računati ICC kao konzistentnost ocena procenjivača. Naime, čak i ako jedan procenjivač daje sistematski više ocene od drugog, u konačnom zbiru ovo neće uticati na poredak kandidata jer će svi imati podjednako „uvećane” ocene (varijabla je uvećana za konstantu). U našem primeru, ako jedan od procenjivača sve odgovore procenjuje kao kreativnije nego što to čine preostali procenjivači – to će dovesti do više prosečne ocene za sve odgovore, ali neće narušiti redosled odgovora po kreativnosti. Sa druge strane, ako nam je važno da znamo da li su procene procenjivača bile međusobno jednake, odnosno ako želimo da okarakterišemo određene odgovore kao, na primer, „veoma kreativne”, onda treba da računamo ICC kao apsolutno slaganje.

U IBM SPSS statističkom softveru se intraklasni koeficijent korelacije može pronaći u okviru analize pouzdanosti instrumenta **Analyze → Scale → Reliability Analysis**. Ovo ne treba da iznenađuje, s obzirom na to da smo već pomenuli kako se objektivnost tretira i kao „pouzdanost” procenjivača (eng. *inter-rater reliability)*. U polje *Items* treba ubaciti varijable koje se odnose na različite procenjivače (u našem slučaju to je pet varijabli), a u okviru menija *Statistics* treba štiklirati opciju da se izračuna ICC – *Intraclass Correlation Coeffiecient*, kao i izabrati odgovarajući model i definiciju[9] koeficijenta (Slika 2.2). Kao što je prethodno objašnjeno, za potrebe naše analize izabraćemo dvosmerni slučajni model i slaganje procenjivača izraženo kao konzistentnost jer najviše odgovaraju prirodi podataka i zaključcima koje želimo da donesemo.



Slika 2.2

U ispisu ove analize dobijamo nekoliko pokazatelja. Pre svega, kako se ICC računa u okviru analize pouzdanosti instrumenta, dobićemo vrednost Kronbahovog alfa koeficijenta pouzdanosti. Za sada ovaj koeficijent nećemo interpretirati, već ćemo pogledati vrednosti u tabeli *Intraclass Correlation Coeffiecient* (Slika 2.3).

[9] Ono što je u SPSS-u označeno kao ICC „tip” (type) u tekstu smo označavali kao definiciju ICC, prema podeli datoj u Koo & Li (2016).

Reliability Statistics

Cronbach's Alpha	N of Items
.831	5

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.495[a]	.367	.630	5.910	49	196	.000
Average Measures	.831	.743	.895	5.910	49	196	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

Slika 2.3

Primećujemo da u koloni *Intraclass Correlation* imamo dve vrednosti – *Single Measures*, odnosno pojedinačna vrednost slaganja i *Average Measures*, odnosno prosečna vrednost slaganja. Pojedinačna vrednost se odnosi na vrednost ICC-a za jednog procenjivača, dok se prosečna vrednost odnosi na ICC za sve procenjivače uzete zajedno. Kao što smo već rekli, prosečna vrednost će uvek biti viša od pojedinačne, što je vidljivo i na našim podacima, pri čemu je pojedinačna vrednost veoma niska i nezadovoljavajuća, dok je prosečna vrednost dobra. Treba primetiti i da je vrednost pojedinačne mere slaganja slična prosečnoj Pirsonovoj korelaciji (r = .506), ali je nešto niža od nje. Koja od ove dve mere onda odražava pravu objektivnost, odnosno slaganje procenjivača – pojedinačna ili prosečna? Odgovor pre svega zavisi od toga na koji način će se primenjivati protokol ocenjivanja. Ukoliko i ubuduće planiramo da procene kreativnosti odgovora ispitanika zasnivamo na proceni pet različitih procenjivača, onda je prosečna vrednost ICC adekvatna procena objektivnosti. Ukoliko, međutim, planiramo da u budućnosti svaki odgovor ocenjuje samo jedan procenjivač, onda nas zanima pojedinačna vrednost ICC.

Na ovom mestu kratko ćemo prokomentarisati odnos između koeficijenta korelacije i intraklasnog koeficijenta. Jedna od pretpostavki računanja ICC-a je da se razlike u varijansi procenjivača tumače kao neslaganje, dok ovakve razlike u varijansi neće uticati na visinu korelacije (McGraw & Wong, 1996). Samim tim, koeficijent korelacije predstavlja gornju granicu ICC-a, odnosno ICC će uvek imati jednaku ili nižu vrednost nego što je vrednost Pirsonove korelacije za iste podatke.

Još jedna pojedinost koju ste možda primetili u ispisu ove analize jeste da je prosečna vrednost ICC jednaka vrednosti Kronbahovog alfa koeficijenta. I za dvosmerni slučajni i za dvosmerni mešoviti model ICC će (ukoliko koristimo definiciju konzistentnosti, a ne slaganja) imati numerički istu vrednost kao i Kronbahov koeficijent pouzdanosti. Ovo odgovara definiciji objektivnosti kao pouzdanosti (procena različitih) procenjivača. Više o pouzdanosti biće rečeno u narednom poglavlju.

2.4) Procena objektivnosti za ordinalni nivo merenja

Procenu objektivnosti za ordinalni nivo merenja ilustrovaćemo sledećim hipotetičkim primerom – tročlani žiri na plesnom takmičenju posmatra nastupe 35 plesača i plesačica. Svaki član žirija rangira sve plesače dodeljujući im jedinstvene rangove od 1 (najbolji) do 35 (najlošiji). Kako možemo utvrditi stepen slaganja procenjivača? Kako Tabela 2.1 pokazuje, za procenu „pouzdanosti" procenjivača za ordinalne podatke može se iskoristiti Spirmanov koeficijent korelacije ili Kendalov W koeficijent. I jedan i drugi koeficijent mogu se, u zavisnosti od konteksta, tumačiti i kao da se odnose na stepen slaganja procenjivača. Kako bismo utvrdili da li između članova žirija ima značajnih razlika u prosečnom rangu, možemo koristiti Fridmanov test (pošto se radi o više od dva procenjivača).

2.4.1) Spirmanova korelacija

Analogno situaciji koju smo imali za intervalni nivo procene, najjednostavniji način da utvrdimo stepen u kom su procene članova žirija slične jeste da izračunamo međusobne korelacije i uprosečimo ih. U ovom primeru, imamo tri procenjivača, te i tri korelacije koje je potrebno uprosečiti.

Koristeći opciju **Analyze → Correlate → Bivariate**, uz označavanje opcije **Spearman** kao željenog koeficijenta korelacije, za tri člana žirija iz primera dobijamo određene vrednosti (Slika 2.4). Vidimo da je slaganje svaka dva člana žirija bilo visoko, preko .90. Ako koeficijent objektivnosti izrazimo kao prosečnu korelaciju, dobijamo vrednost $\bar{r} = \frac{.926+.934+.917}{3} = .926$. Na osnovu ove vrednosti objektivnost procenjivača bismo ocenili kao veoma visoku.

Correlations

			p1	p2	p3
Spearman's rho	p1	Correlation Coefficient	1.000	.926**	.934**
		Sig. (2-tailed)	.	.000	.000
		N	35	35	35
	p2	Correlation Coefficient	.926**	1.000	.917**
		Sig. (2-tailed)	.000	.	.000
		N	35	35	35
	p3	Correlation Coefficient	.934**	.917**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	35	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 2.4

2.4.2) Kendalovo W

Kendalovo W ili Kendalov koeficijent konkordance je neparametrijski statistik koji ukazuje na stepen slaganja procenjivača. Može se koristiti za ordinalne podatke ili intervalne podatke čija distribucija odstupa od normalne, a vrednosti mu se kreću od 0 – potpuno odsustvo slaganja do 1 – potpuno slaganje procenjivača (Kendall & Babington Smith, 1939).

Računanje Kendalovog W počiva na rangiranju objekata procene. Prvo se svakom objektu dodeljuje rang u okviru svakog procenjivača (utvrđujemo da li je dati plesač od nekog člana žirija dobio viši ili niži rang u poređenju sa drugim plesačima), a zatim se računa suma rangova za svaki objekat po formuli $R_i = \sum_{j=1}^{m} r_{ij}$, gde je m broj procenjivača (članova žirija), i oznaka za objekat procene (plesača), a j za procenjivača (člana žirija). Pod pretpostavkom da među članovima žirija postoji apsolutno slaganje, najbolji plesač bi imao sumarni rang 3 (bio bi rangiran kao prvi od strane sva tri člana žirija), dok bi najlošiji plesač imao sumarni rang 105 (tri puta bi bio rangiran kao 35.). Suprotno tome, ukoliko među procenjivačima nema slaganja, sumarni rangovi plesača se neće bitno razlikovati jedni od drugih. U sledećem koraku računa se opšta suma rangova (za sve objekte procene zajedno) $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$, gde je n broj objekata procene (plesača). Ukupna suma rangova će biti ista bez obzira na stepen slaganja procenjivača, jer zavisi samo od broja procenjivača i broja objekata procene, i može se dobiti prema formuli $\bar{R} = \frac{mn(n+1)}{2}$. Zatim, računamo sumu kvadriranih odstupanja prosečnog ranga pojedinačnog objekta od opšteg prosečnog ranga $S = \sum_{i=1}^{n}(R_i - \bar{R})^2$. Kendalovo W se onda računa po sledećoj formuli:

$$W = \frac{12S}{m^2(n^3 - n)}$$

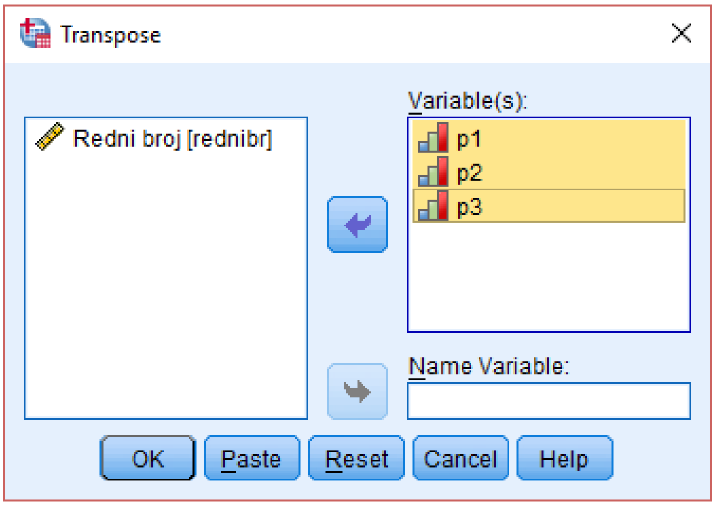
gde su m i n, redom, broj procenjivača i objekata procene.

Kendalovo W se takođe može izračunati i na osnovu prosečnog koeficijenta Spirmanove korelacije između procenjivača, prema formuli:

$$W = \frac{(m-1)\bar{r} + 1}{m}$$

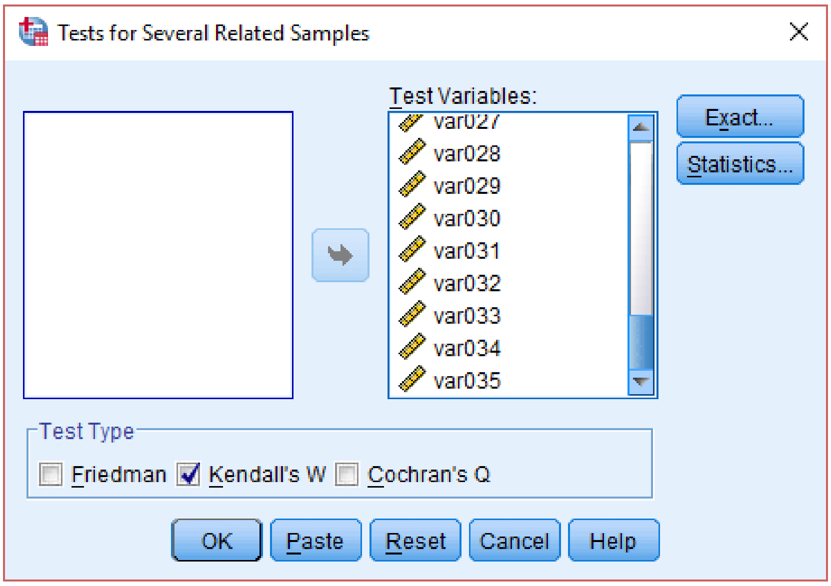
gde je m broj procenjivača, a $\bar{r}$ prosečna Spirmanova korelacija između parova procenjivača.

U SPSS-u se Kendalovo W nalazi u okviru neparametrijskih tehnika namenjenih analizi k povezanih (zavisnih) uzoraka. Bitno je, međutim, napomenuti da SPSS algoritam Kendalovo W računa tako što uprosečava rangove **po kolonama**. To znači da je, za dobijanje tačne vrednosti ovog statistika, potrebno da su podaci tako organizovani da se procenjivači nalaze u redovima, a objekti procene u kolonama. Kako je ovo suprotno uobičajenom načinu organizacije podataka u bazama, pre sprovođenja analize neophodno je transponovati podatke komandom **Data** → **Transpose**. Kao varijable koje želimo da transponujemo ubacujemo isključivo varijable koje sadrže rangove procenjivača (Slika 2.5). Ukoliko bismo transponovali još neku varijablu, ona bi bila uključena u računanje Kendalovog W, što bi dovelo do dobijanja pogrešne vrednosti ovog statistika.



Slika 2.5

Pokretanje ove komande kreira novi SPSS fajl gde su podaci transponovani. Tri varijable koje su bile u kolonama (članovi žirija) sada su tri reda u novom fajlu sa podacima, dok je 35 redova inicijalnog fajla (takmičari) sada pretvoreno u varijable u zasebnim kolonama. Na ovako transponovanom fajlu sa podacima računamo Kendalovo W preko procedure **Analyze** → **Nonparametric Test** → **Legacy Dialogues** → **K related samples**. U polje *Test Variables* treba ubaciti sve objekte procene, odnosno sve varijable koje postoje u podacima (izuzev nominalne varijable CASE_LBL, u službi identifikovanja inicijalne varijable koja je transponovana) (Slika 2.6). Kao tip testa treba izabrati Kendalovo W (*Test Type – Kendall's W*).



Slika 2.6

U ispisu analize dobijamo dve tabele. U prvoj tabeli **Ranks**, nalazi se prosečan rang za svaku kolonu, odnosno prosečan rang svakog plesača. U drugoj tabeli ispisa (Slika 2.7) nalazi se samo Kendalovo W, kao i Hi-kvadrat test, koji testira nultu hipotezu da je Kendalovo W jednako nuli. U našem primeru vidimo da Kendalovo W iznosi W = .950 i značajno je, što ukazuje na zavidno slaganje članova žirija po pitanju kvaliteta plesača na takmičenju.

Test Statistics

N	3
Kendall's W[a]	.950
Chi-Square	96.943
df	34
Asymp. Sig.	.000

a. Kendall's Coefficient of Concordance

Slika 2.7

Rekli smo da se Kendalovo W takođe može izračunati i na osnovu prosečne Spirmanove korelacije rangova procenjivača. Kao što smo već pokazali, prosečna Spirmanova korelacija iznosi $\bar{r} = \frac{.926+.934+.917}{3} = .925667$. Ukoliko ovu vrednost zamenimo u formuli za Kendalovo W, dobijamo: $W = \frac{(m-1)\bar{r}+1}{m} = \frac{(3-1)*.926+1}{3} = .95044$, odnosno istu vrednost koju smo dobili i u odgovarajućoj SPSS analizi. Dakle, podaci nedvosmisleno ukazuju na dobro slaganje procenjivača, odnosno na visoku objektivnost merenja.

Da bismo stekli potpunu sliku o objektivnosti procene, (prosečnu) Spirmanovu korelaciju ili Kendalovo W mogli bismo dopuniti analizama postojanja razlika za ordinalne podatke, kao što su test predznaka ili Vilkoksonov test za dva procenjivača, odnosno Fridmanov test za tri ili više procenjivača (videti Dodatak 2). Takođe, kao meru slaganja procenjivača i na ordinalnim podacima možemo računati intraklasni koeficijent korelacije (ICC).

2.5) Procena objektivnosti za nominalni nivo merenja

Konačno, zamislimo situaciju gde dva klinička psihologa nezavisno jedan od drugog treba da postave dijagnoze nizu pacijenata. Kako ćemo proceniti u kojoj meri se njihove procene poklapaju, odnosno u kojoj meri su dijagnoze objektivno date? U slučaju nominalnih podataka, procena objektivnosti je nešto jednostavnija nego za ordinalne ili intervalne podatke. To je zbog toga što kod nominalnih podataka postojanje savršene korelacije između procenjivača nužno implicira odsustvo razlika, te je za procenu objektivnosti dovoljno izračunati jedan koeficijent slaganja koji se istovremeno odnosi na oba aspekta objektivnosti. Najčešće korišćene metode za utvrđivanje objektivnosti na nominalnim podacima su procenat apsolutnog slaganja i kapa koeficijent.

2.5.1) Procenat apsolutnog slaganja

Najjednostavniji način da procenimo stepen slaganja procenjivača jeste da utvrdimo procenat slučajeva gde postoji apsolutno slaganje (Graham et al., 2012; Shweta et al., 2015). Numerički, ovo postižemo tako što saberemo broj situacija u kojima su procenjivači dali identične procene i podelimo to ukupnim brojem situacija.

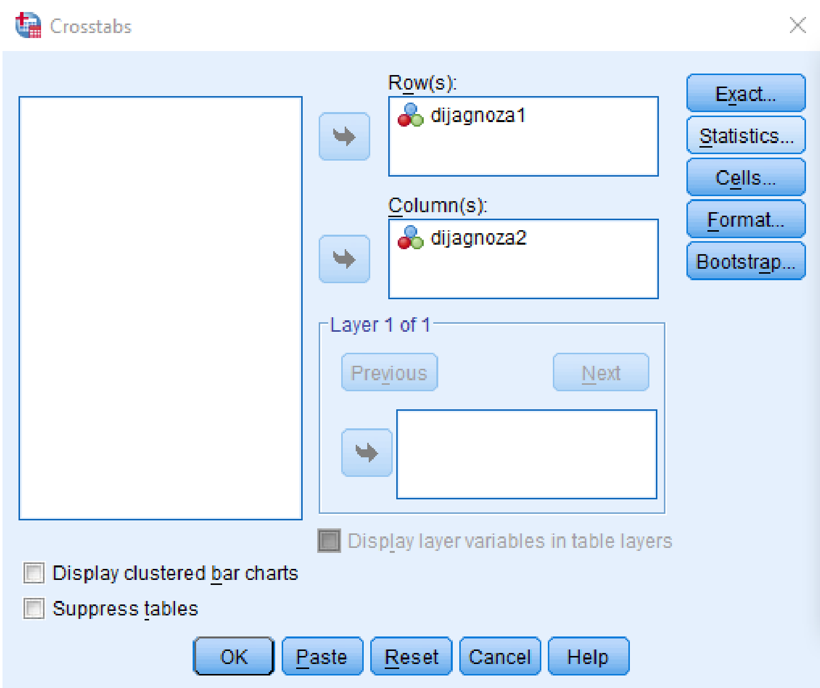
U našem primeru, pretpostavimo da su dva kliničara davala samo sledeće četiri dijagnoze: anksiozni poremećaj, depresija, bipolarni poremećaj i šizofrenija. Ukoliko ukrštanje dijagnoza prikažemo u tabeli s četiri reda i četiri kolone, ćelije na dijagonali odnose se na proporciju slaganja, dok se vandijagonalne ćelije odnose na neslaganje u dijagnozi (Tabela 2.2).

Tabela 2.2. Slaganje procenjivača u slučaju nominalnih podataka

		Kliničar 1				
		1	2	3	4	Total
Kliničar 2	1. anksiozni poremećaj	**slaganje**	neslaganje	neslaganje	neslaganje	n_{anks2}
	2. depresija	neslaganje	**slaganje**	neslaganje	neslaganje	n_{depr2}
	3. bipolarni poremećaj	neslaganje	neslaganje	**slaganje**	neslaganje	n_{bipol2}
	4. šizofrenija	neslaganje	neslaganje	neslaganje	**slaganje**	$n_{šizo2}$
	Total	n_{anks1}	n_{depr1}	n_{bipol1}	$n_{šizo1}$	N

Procenat slaganja dobijamo tako što saberemo dijagonalne elemente ove matrice, podelimo ih ukupnim brojem procena (što u našem primeru odgovara ukupnom broju pacijenata) i to pomnožimo sa 100 (kako bismo dobili procente, a ne proporcije).

U SPSS-u do procenta apsolutnog slaganja procenjivača dolazimo preko **Analyze → Descriptive Statistics→ Crosstabs**. U polje *Row(s)* ubacujemo jednu od nominalnih varijabli, a drugu ubacujemo u polje *Column(s)* (Slika 2.8).



Slika 2.8

U ispisu (Slika 2.9) dobijamo informaciju o broju pacijenata kojima su kliničari postavili iste, odnosno različite dijagnoze. Već na prvi pogled možemo primetiti da je generalno postojalo slaganje između kliničara u postavljanju dijagnoze, jer je broj slučajeva van glavne dijagonale primetno manji od broja slučajeva na dijagonali. Sabiranjem vrednosti na dijagonali, deljenjem ukupnim brojem pacijenata i množenjem sa 100, dobijamo vrednost $p = \frac{22+20+19+25}{100} * 100 = 86$, odnosno slaganje kliničara je postojalo u 86% slučajeva. Uzimajući u obzir preporuke da se procenat apsolutnog slaganja od 75% smatra prihvatljivim, a od 90% visokim (Graham et al., 2012), dobijenu vrednost bismo okarakterisali kao dobro slaganje. Ono što je, međutim, nedostatak procenta apsolutnog slaganja kao mere objektivnosti jeste što ne uzima u obzir da je do određenog procenta slaganja moglo da dođe i na osnovu slučaja. Ovaj nedostatak prevazilati kapa koeficijent, koji ćemo prikazati u narednom delu.

dijagnoza1 * dijagnoza2 Crosstabulation

Count

		dijagnoza2				Total
		anksiozni poremećaj	depresija	bipolarni poremećaj	šizofrenija	
dijagnoza1	anksiozni poremećaj	22	2	1	1	26
	depresija	2	20	2	1	25
	bipolarni poremećaj	2	0	19	0	21
	šizofrenija	0	1	2	25	28
Total		26	23	24	27	100

Slika 2.9

Radi jednostavnosti, u našem primeru postojala su samo dva procenjivača, no procenat apsolutnog slaganja se može računati i za veći broj procenjivača, prateći istu logiku – utvrđuje se procenat slučajeva u kojima su svi procenjivači dali istu (pr)ocenu. Ponekad se u praksi koristi i procenat susednog slaganja (eng. *percentage of adjacent agreement*), gde se pod slaganjem podrazumeva ne samo davanje identične, već i davanje susedne ocene (Graham et al., 2012; Shweta et al., 2015). Naravno, procenat susednog slaganja će uvek biti jednak ili veći od procenta apsolutnog slaganja. Da bi se računao procenat susednog slaganja, neophodno je da između kategorija postoje barem odnosi ordinalne uređenosti (u našem primeru, davanje susednih dijagnoza ne bi moglo ukazivati ni na koji oblik slaganja procenjivača).

2.5.2) Kapa koeficijent

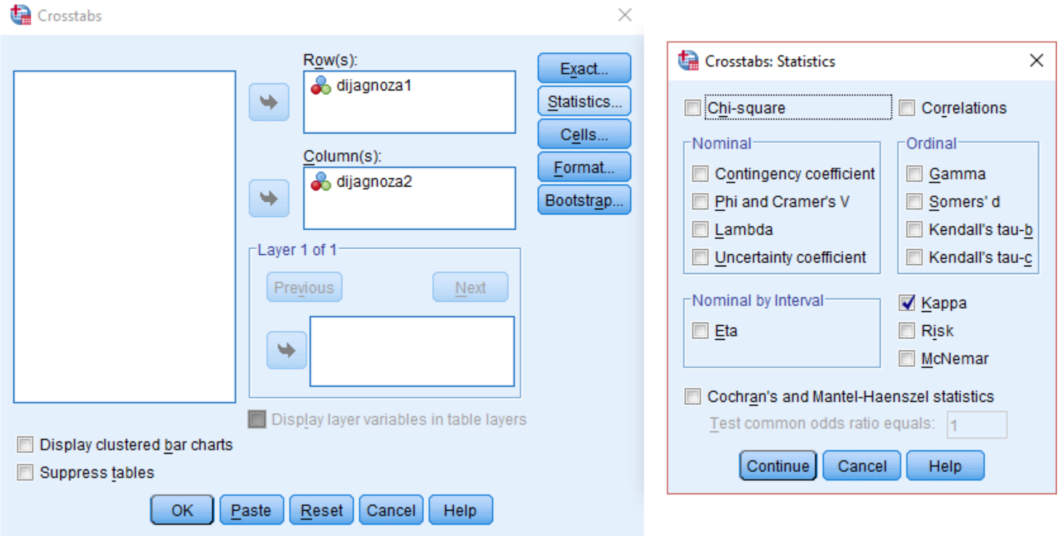
Koenov kapa koeficijent (eng. *Cohen's kappa*) predstavlja meru objektivnosti namenjenu nominalnim podacima (J. Cohen, 1960). Važna karakteristika kapa koeficijenta jeste što uzima u obzir (u vidu korekcije) mogućnost slučajnog slaganja procenjivača. Kapa koeficijent se računa prema sledećoj formuli:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

gde je p_o opservirano, odnosno dobijeno slaganje procenjivača, a p_e očekivano slaganje, odnosno slaganje koje se može očekivati na osnovu slučaja. I dobijeno i očekivano slaganje se iskazuju kao proporcija objekata/subjekata za koje postoji slaganje procenjivača, odnosno koji su svrstani u istu kategoriju. Maksimalna vrednost kapa koeficijenta je 1, i to samo u slučaju kada postoji potpuno slaganje između procenjivača, odnosno kada procenjivači identično svrstavaju sve objekte/subjekte procene. Kada je slaganje procenjivača na nivou slučajnog, kapa koeficijent jednak je nuli. Ukoliko je slaganje procenjivača niže od onog očekivanog na osnovu slučaja, može se desiti i da vrednost kapa koeficijenta bude negativna. Postoje različite preporuke o prihvatljivim vrednostima kapa koeficijenta, pri čemu se najčešće koeficijent viši od .60 smatra dobrim, a preko .80 odličnim (Denham, 2017; Graham et al., 2012; Shweta et al., 2015).

U našem primeru (Tabela 2.2), proporciju slaganja (p_o) dobijamo tako što saberemo proporcije svake od četiri dijagonalne ćelije, odnosno broj slučajeva na glavnoj dijagonali podelimo brojem ispitanika. Proporciju slaganja koja se može očekivati na osnovu slučaja (p_e) dobijamo kao proizvod marginalnih proporcija, odnosno proizvod marginalnih frekvenci podeljen kvadratom broja ispitanika $p_e = \frac{1}{N^2} \sum_k n_{k1} n_{k2}$, gde su n_{k1} i n_{k2} broj ispitanika svrstanih u datu kategoriju od strane prvog i drugog procenjivača, redom.

Kapa koeficijent se u SPSS-u može dobiti u okviru **Analyze** → **Descriptive Statistics**→ **Crosstabs**. U polje *Row(s)* ubacujemo jednu od nominalnih varijabli, dok drugu ubacujemo u polje *Column(s)*, a u okviru prozora *Statistics* biramo opciju da se izračuna *Kappa* koeficijent (Slika 2.10).



Slika 2.10

U ispisu ove analize relevantne su nam dve tabele (Slika 2.11). Prvu tabelu (*Crosstabulation*) već smo razmatrali prilikom izračunavanja procenta apsolutnog slaganja i ona nam pokazuje kako su kliničari klasifikovali pacijente. Kao što je već objašnjeno, vrednosti na glavnoj dijagonali predstavljaju pacijente oko kojih je postojalo slaganje u dijagnozi, dok su vrednosti van glavne dijagonale pacijenti za koje su dva kliničara dala različite dijagnoze. Visok kapa koeficijent (kappa = .813), dat u drugoj tabeli (*Symmetric Measures*), ukazuje na dobro slaganje između kliničara.

dijagnoza1 * dijagnoza2 Crosstabulation

Count

		dijagnoza2				
		anksiozni poremećaj	depresija	bipolarni poremećaj	šizofrenija	Total
dijagnoza1	anksiozni poremećaj	22	2	1	1	26
	depresija	2	20	2	1	25
	bipolarni poremećaj	2	0	19	0	21
	šizofrenija	0	1	2	25	28
Total		26	23	24	27	100

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.813	.046	14.075	.000
N of Valid Cases		100			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Slika 2.11

Poređenjem kapa koeficijenta sa procentom apsolutnog slaganja vidimo da je kapa koeficijent niži, što je i očekivano s obzirom na korekciju za slučajno slaganje. U ispisu još vidimo i da je kapa koeficijent značajan, što samo ukazuje da je njegova vrednost značajno veća od nule, ali nema dodatnu interpretativnu vrednost, jer značajnost kapa koeficijenta ne implicira nužno objektivnost merenja (i nizak koeficijent može biti značajan na većem broju ispitanika).

Originalni, Koenov kapa koeficijent se može računati za bilo koji broj kategorija, ali samo za dva procenjivača. Kako procenjujemo slaganje većeg broja procenjivača ukoliko su podaci dati na nominalnoj skali? Uglavnom se koriste neke od modifikacija kapa koeficijenta, najčešće Fleisova kapa (eng. *Fleiss' kappa*) (J. Fleiss, 1971). I ova mera počiva na istoj ideji kao i Koenova kapa – proporcija slaganja procenjivača upoređuje se sa proporcijom slaganja koja se može očekivati na osnovu slučaja, ali je njeno računanje kompleksnije. Kako nije ugrađena u SPSS programski paket, razvijene su sintakse preko kojih se ona može izračunati, praćene uputstvima za njihovo pokretanje (King, 2004). Fleisova kapa se, osim za procenu objektivnosti nominalnih podataka, može koristiti i za procenu objektivnosti ordinalnih podataka (za dva ili više procenjivača), a na ordinalnim podacima (za dva procenjivača) možemo upotrebiti i Koenovu ponderisanu kapu (eng. *weighted kappa*) (Ranganathan et al., 2017). Može se takođe pokazati i da su različite verzije kapa koeficijenta numerički ekvivalentne određenim formama intraklasnog koeficijenta korelacije (J. L. Fleiss & Cohen, 1973; Rae, 1988).

2.6) Rezime

U veoma velikom broju situacija psihološko merenje može biti savršeno objektivno, ali u onim situacijama kada to nije, izuzetno je važno proceniti koji deo varijanse skorova se može pripisati procenjivanim osobama, a koji deo procenjivačima. Objektivnost se može posmatrati na dva osnovna načina – kao stepen u kom su procene procenjivača međusobno korelirane, što nazivamo pouzdanošću procenjivača, i kao stepen u kom su procene procenjivača međusobno iste, što nazivamo slaganjem procenjivača. Slaganje procenjivača u sebi podrazumeva pouzdanost procenjivača, ali uzima u obzir i (ne)postojanje razlika između procena različitih procenjivača.

U zavisnosti od nivoa merenja, postoje različite tehnike provere objektivnosti. Tako, za intervalne podatke možemo koristiti Pirsonove korelacije ili intraklasni koeficijent korelacije (ICC), za ordinalne podatke Spirmanove korelacije ili Kendalovo W, dok za nominalni nivo merenja najčešće koristimo procenat apsolutnog slaganja ili kapa koeficijent.

2.7) Preporučena literatura

Bukvić, A. (1996). *Načela izrade psiholoških testova*. Zavod za udžbenike i nastavna sredstva, Beograd.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment: An Introduction to Tests and Measurement. Ninth Edition*. New York, NY: McGraw-Hill Education.

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings*. Center fir Educator Compensation Reform.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Shweta, Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, *41*(Special Issue 3), 20–27.

3) POUZDANOST

Ukoliko vodimo računa o svojoj težini, redovno merenje na kućnoj vagi će nam sa dovoljnom preciznošću pokazati da li smo se ugojili, smršali ili imamo relativno stabilnu telesnu težinu. Ista ova vaga, međutim, ne bi nam bila od velike koristi ukoliko bismo želeli da napravimo kolač čiji sadržaj podrazumeva 135g brašna, 250g putera i 400g čokolade. Razlog je jasan – za pravljenje kolača potrebno je finije merenje nego za kontrolisanje telesne težine odrasle osobe. Lako je zamisliti i situacije (npr., u okviru farmakološke industrije) kada je merenje potrebno izvršiti sa još većim nivoom preciznosti, a težinu iskazati u mikrogramima ili nanogramima.

Pouzdanost testa odnosi se na nivo konzistentnosti merenja koji datim testom možemo postići (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018). U psihometrijskom smislu, ona se najčešće upotrebljava u ovom značenju, ali se takođe odnosi i na vremensku stabilnost merenja. Vremenska stabilnost podrazumeva da ukoliko istu stvar merimo dva puta, dobijemo i isti rezultat. Na primer, ako danas izmerim svoj mobilni telefon i utvrdim da je težak 145g, s obzirom na to da se težina telefona ne menja iz dana u dan, mogu očekivati da će i svako naredno (pouzdano) merenje dati težinu od 145g.

3.1) Klasična teorija testa

Kada smo govorili o objektivnosti merenja, rekli smo da se varijabilitet skorova na nekom testu može razložiti na dve osnovne komponente. Prva komponenta odnosi se na varijabilitet skorova koji potiče od ispitanika, dok se drugi deo odnosi na varijabilitet potekao od različitih faktora greške, pri čemu ove faktore možemo podeliti na slučajne i sistematske. Ovakvo razumevanje varijabiliteta skorova naziva se teorijom pravog skora, odnosno klasičnom teorijom testa (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; DeVellis, 2006; Fajgelj, 2009; Kline, 2000; Novick, 1966). Koristeći matričnu notaciju, klasičnu teoriju testa možemo napisati u sledećem obliku:

$$X = T + E$$

gde je X vektor ukupnih skorova na nekom testu, T vektor pravih skorova ispitanika (deo varijabiliteta koji zavisi od ispitanika), a E vektor grešaka (deo varijabiliteta koji odlazi na grešku).

Ukoliko bismo navedeno ilustrovali primerom testa znanja iz psihologije – ukupan skor ispitanika je broj poena koje je on osvojio na testu znanja, na primer, 25. Moguće je, međutim, da ovaj ispitanik poseduje veće znanje psihologije i da je umeo da reši 26 pitanja, ali se kod jednog zbunio (npr., nije zaokružio odgovarajući broj u listu za odgovore). U tom slučaju, pravi skor ispitanika bio bi 26, a skor greške bi bio -1. Naravno, moguće je isto tako i da je ispitanik zapravo znao rešenje na 24 pitanja, a da je na preostalih 6 nasumice odgovarao, od čega je na jedno pitanje „pogodio" odgovor. U ovom slučaju, pravi skor ispitanika je 24, a skor greške je 1.

Ono što nam je, kao istraživačima, jedino dostupno jeste ukupan, dobijeni skor ispitanika na testu. Ni za jednog pojedinačnog ispitanika ne možemo sa sigurnošću odrediti koji deo skora je pravi skor, a koji je greška merenja. Drugim rečima, nismo sigurni da li je pravi skor ispitanika koji je osvojio 25 poena na testu znanja zapravo 24 ili 26 ili možda baš 25. Postoje, međutim, metode na osnovu kojih se može proceniti udeo slučajne greške u ukupnom varijabilitetu, a samim tim i udeo prave, pouzdane varijanse u ukupnoj varijansi. Ove metode nazivaju se metodama procene pouzdanosti (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Informacija o pouzdanosti instrumenta važna nam je kako bismo znali u kojoj meri možemo imati poverenje u rezultate dobijene testom. Na primer, ako se na testu znanja iz psihologije pravi skor ispitanika nalazi u okviru ±2 poena na testu, možemo zaključiti da se najveći deo varijanse (razlika između ispitanika) može objasniti razlikama u njihovom znanju. Ako se pak pravi skor ispitanika nalazi u okviru ±10 poena na testu, imamo razloga za ozbiljne sumnje u pouzdanost skorova dobijenih takvim testom.

Treba napomenuti i da se prilikom procenjivanja pouzdanosti uzima u obzir samo slučajna greška. Drugim rečima, podrazumeva se da je merenje izvršeno na objektivan način, kao i da su eliminisani svi faktori koji bi doveli do eventualne sistematske promene skorova ispitanika (eliminisanje ovih faktora može se postići pre testiranja obezbeđivanjem dobrih uslova testiranja, ali i nakon testiranja određenim statističkim tehnikama).

3.1.1) Osobine slučajne greške

O sistematskoj grešci je već bilo reči, ali kako možemo razumeti slučajnu grešku? Slučajnu grešku čini delovanje svih onih faktora koji su istraživaču nepoznati i koji se na nepredvidljiv način menjaju od situacije do situacije testiranja (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Kline, 2000). Na primer, možemo pretpostaviti da će jaka buka tokom testiranja otežati rad ispitanicima, ali ne bismo očekivali da se slične okolnosti ponove u svakoj situaciji testiranja. U slučajne faktore mogu spadati i trenutno stanje ispitanika (umor, trema, raspoloženje, specifična situacija u kojoj se ispitanik nalazi ili koje se seti tokom rada na testu i dr.), kao i „sreća” (neko je pročitao samo deo literature i dobije pitanje baš iz tog dela gradiva, dok je neko drugi čitao sve sem onog dela koji je bio na testu). Zajednički element svih navedenih faktora jeste, dakle, da oni deluju na slučajan način, koji je teško ili nemoguće predvideti i kontrolisati.

Ukoliko smo na ovaj način definisali slučajnu grešku, šta možemo izdvojiti kao njene bitne osobine?

- Pre svega, ukoliko je greška zaista slučajna, na nivou celog uzorka ispitanika ona neće dovesti ni do povećanja ni do smanjenja skorova. Iako za pojedinačne ispitanike slučajna greška može povećati ili smanjiti pravi skor, na nivou grupe njena prosečna vrednost mora biti nula (u suprotnom, radi se o sistematskom ometajućem ili olakšavajućem faktoru). Ovu osobinu slučajne greške možemo iskazati izrazom: $M_E = 0$.
- Takođe, slučajna greška bi morala sa jednakom verovatnoćom delovati na sve ispitanike, bez obzira na njihov nivo osobine. Drugim rečima, ukoliko je greška veća za ispitanike koji poseduju neku osobinu u većoj meri nego za ispitanike koji je poseduju u manjoj meri, onda se ponovo radi o nekom sistematskom, a ne slučajnom uticaju. Samim tim, kao jednu od bitnih osobina slučajne greške možemo izdvojiti to da je njena korelacija sa pravim skorom jednaka nuli, odnosno $r_{TE} = 0$.
- Konačno, korelacija dve slučajne greške (na različitim testovima, formama testova ili prilikom različitih trenutaka zadavanja istog testa) mora biti nula, s obzirom na to da se ne može očekivati da dva slučajna događaja budu na bilo koji način, pozitivno ili negativno, povezana. Drugim rečima, $r_{EjEk} = 0$.

Možemo zaključiti da slučajna greška predstavlja šum koji doprinosi nepreciznosti merenja, ali ne utiče na prosečan skor uzorka ni na kakav konzistentan način (Bukvić, 1996; R. J. Cohen &

Swerdlik, 2018; Kline, 2000). S obzirom na navedene karakteristike slučajne greške (posebno na poslednju), jasno je da ni „količina” greške neće uvek biti jednaka, odnosno isti test neće na svim uzorcima nužno pokazivati istu pouzdanost. Upravo zbog toga se, kao što je već pomenuto, preporučuje da se, umesto o pouzdanosti testa, govori o pouzdanosti podataka (AERA et al., 2014; DeVellis, 2006; Fajgelj, 2009).

3.1.2) Opšti obrazac za procenu pouzdanosti

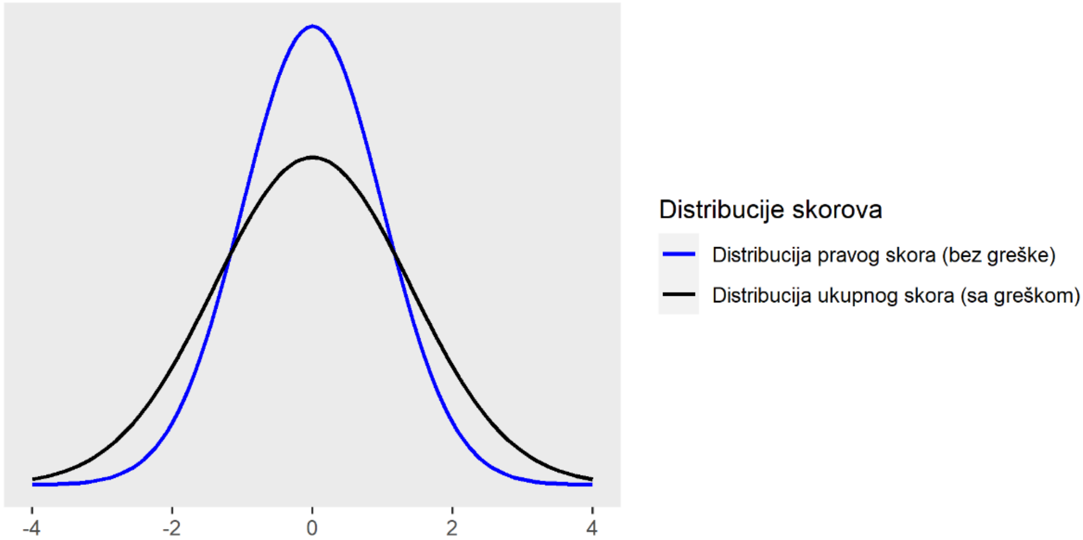
Već smo rekli da se, prema teoriji pravog skora, svaki pojedinačni skor na instrumentu može razložiti na pravi skor i skor greške, odnosno da se vektor koji sadrži ukupne skorove ispitanika može razložiti na vektor pravih skorova i vektor skorova greške. Čemu će onda biti jednaka varijansa ukupnog skora? Ako znamo da je ukupan skor na testu linearni kompozit pravog skora i skora greške, kao i da je varijansa svakog linearnog kompozita jednaka zbiru varijansi komponenti i njihove dvostruke kovarijanse, varijansu ukupnog skora možemo prikazati na sledeći način:

$$\sigma_X^2 = \sigma_{T+E}^2 = \sigma_T^2 + \sigma_E^2 + 2r_{TE}\sigma_T\sigma_E$$

gde je σ_X^2 varijansa ukupnog skora, σ_T^2 varijansa pravog skora, σ_E^2 varijansa greške, a r_{TE} korelacija pravog skora i skora greške. Kako je jedna od osobina slučajne greške to da ne korelira sa pravim skorom, izraz $2r_{TE}\sigma_T\sigma_E$ biće jednak nuli, te se formula može napisati i kao:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Dakle, ukupna varijansa na testu može se razložiti na varijansu pravog skora i varijansu greške (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Novick, 1966). Ukoliko bismo ovo prikazali grafički (Slika 3.1), videli bismo da je distribucija bez greške uža od distribucije skorova koji sadrže grešku, s obzirom na to da varijansa greške doprinosi ukupnom varijabilitetu (i to podjednako za sve skorove[10]). Što je greška veća, to će distribucija izmerenih skorova manje nalikovati distribuciji pravih skorova i obrnuto – manja greška znači da će distribucija ukupnog skora biti utoliko sličnija distribuciji pravog skora.



Slika 3.1

[10] Ovo ne znači da slučajna greška podjednako deluje na sve *ispitanike*, jer se može desiti da slučajna greška za jednog ispitanika bude veća nego za nekog drugog, već da nema razloga da verujemo da će greška biti drugačija za različite *skorove* na instrumentu. Odnosno, slučajna greška bi podjednako trebalo da deluje i na niske skorove (npr., 5 poena na testu znanja iz psihologije) i na srednje skorove (npr., 15 poena) i na visoke skorove (npr., 25 poena).

Pouzdanost se obično definiše kao udeo prave varijanse u ukupnoj varijansi, odnosno kao $r_{tt} = \frac{\sigma_T^2}{\sigma_X^2}$, ali, kao što smo već rekli, varijansa pravog skora, kao ni skora greške nam nije poznata. Matematički gledano, procena pouzdanosti odgovara situaciji rešavanja jedne jednačine sa dve nepoznate (Fajgelj, 2009). Kako bi se došlo do rešenja, potrebno je napraviti dodatne pretpostavke koje jednačinu čine matematički rešivom. U praksi, pouzdanost se najčešće procenjuje preko procene udela varijanse greške u ukupnoj varijansi, a metode procene se razlikuju po načinu na koji vrše ovu procenu. Ukoliko izraz $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ podelimo ukupnom varijansom, dobićemo:

$$\frac{\sigma_X^2}{\sigma_X^2} = \frac{(\sigma_T^2 + \sigma_E^2)}{\sigma_X^2}$$

odnosno:

$$1 = \frac{\sigma_T^2}{\sigma_X^2} + \frac{\sigma_E^2}{\sigma_X^2}$$

Kada količnik varijanse greške i varijanse ukupnog skora prebacimo na drugu stranu znaka jednakosti, dobijamo da je pouzdanost jednaka:

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

što je opšti oblik većine obrazaca pouzdanosti koji će biti prikazani u nastavku teksta[11] (Bukvić, 1996; Fajgelj, 2009).

3.2) Metode procene pouzdanosti

Postoji više metoda procene pouzdanosti različitih i po načinu na koji konceptualizuju i procenjuju grešku merenja, ali i po tome na koji vid pouzdanosti se odnose. Na primer, test-retest metoda se odnosi na vremensku konzistentnost (da li test daje iste rezultate u dve različite vremenske tačke), metoda deljenja testa i analiza stavki spadaju u metode interne konzistencije i, pre svega, govore o preciznosti merenja koje datim instrumentom možemo postići, dok se metoda alternativnih formi može upotrebljavati na oba načina (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Ove dve vrste pouzdanosti su komplementarne i ne treba ih izjednačavati. Dodatno, kao poseban vid pouzdanosti može se izdvojiti i pouzdanost procenjivača (nazivali smo je i objektivnošću), a koja se odnosi na konzistentnost skorova u situaciji kada o osobini sudimo na osnovu procena nekoliko procenjivača (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009).

3.2.1) Test-retest metoda

Verovatno najjednostavnija i intuitivno najrazumljivija metoda procene pouzdanosti jeste test-retest metoda. U ovoj metodi se instrument (npr., test znanja iz psihologije ili test anksioznosti) zadaje grupi ispitanika u jednoj vremenskoj tački, a zatim se istoj grupi ispitanika isti instrument zadaje ponovo u narednoj vremenskoj tački, te se za procenu pouzdanosti skorovi na testu i retestu koreliraju (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009).

Od ranije nam je poznato da je procenat varijanse koji dve varijable dele jednak kvadratu njihove korelacije. Zašto se onda pouzdanost, definisana kao procenat varijanse koji odlazi na pravi skor, računa kao obična (a ne kvadrirana) korelacija testa i retesta? Odgovor leži u osobinama slučajne greške, a sada ćemo ga demonstrirati izvođenjem formule. Korelaciju dve varijable možemo prikazati kao sumu proizvoda njihovih standardizovanih skorova podeljenu brojem ispitanika:

[11] Za označavanje pouzdanosti u literaturi se najčešće koriste skraćenice r_{tt} (*t* je od oznake za pravi skor), r_{xx} (*x* je oznaka za skor na testu) ili r_{nn} (gde *n* predstavlja broj delova testa).

$$r_{xy} = \frac{\sum_{i=1}^{n}(z_x z_y)}{N}$$

gde su z_x i z_x standardizovani skorovi na varijablama, a N broj ispitanika. S obzirom na to da znamo da se skor na testu i retestu, odnosno na alternativnim formama testa može razložiti na pravi skor i skor greške, formulu možemo razložiti na sledeći način:

$$r_{x1x2} = \frac{\sum_{i=1}^{n}(z_1 z_2)}{N} = \frac{\sum_{i=1}^{n}(t_1 + g_1)(t_2 + g_2)}{N} = \frac{\sum_{i=1}^{n}((t_1 * t_2) + (t_1 * g_2) + (t_2 * g_1) + (g_1 * g_2))}{N}$$

Pošto znamo da je korelacija pravog skora i skora greške jednaka nuli, kao i da je korelacija dve slučajne greške takođe jednaka nuli, poslednja tri elementa iznad razlomačke crte se mogu obrisati. Takođe, znamo i da je, po definiciji, pravi skor na testu i retestu jednak, pa umesto t_1 i t_2 možemo pisati samo t (jer se radi o istom pravom skoru u oba slučaja). Ono što ostaje jeste:

$$r_{x1x2} = \frac{\sum_{i=1}^{n}(tt)}{N}$$

odnosno kovarijansa (standardizovanih) pravih skorova na testu. Dakle, udeo prave varijanse u ukupnoj varijansi jednak je koeficijentu korelacije testa i retesta. Na primer, ukoliko je test-retest korelacija testa znanja iz psihologije $r = .87$, onda je pouzdanost ovog testa $r_{tt} = .87$, odnosno 87% ukupne varijanse testa (dobijenih poena) čini varijansa pravih skorova ispitanika (stvarnog znanja ispitanika), dok 13% varijanse odlazi na slučajnu grešku (pogađanje, propuste itd.).

Implicitna pretpostavka kod računanja koeficijenta pouzdanosti test-retest metodom jeste da se sama osobina (poznavanje gradiva psihologije, odnosno anksioznost) nije vremenom promenila. Za određeni broj psiholoških konstrukata, kao što su, recimo, inteligencija (Hunt, 2010) ili crte ličnosti (Bazana & Stelmack, 2004), pokazano je da su relativno stabilni kroz vreme, odnosno da ne postoje velike razlike u skorovima ispitanika od jednog trenutka do sledećeg. Čak i za određene psihološke konstrukte gde bismo teorijski očekivali da su podložniji promenama, empirijski se pokazuje nezanemarljiv nivo vremenske stabilnosti, na primer, za raspoloženja (McConville & Cooper, 1997) i stavove (Gawronski et al., 2017).

Međutim, čak i ako pretpostavimo da se sama osobina koju procenjujemo instrumentom nije vremenom promenila, situacija inicijalnog testiranja može uticati na postignuće ispitanika kada drugi put bude radio isti test (ispitanik se može, na primer, setiti pitanja sa testa znanja, ali i toga kako je odgovorio na neku stavku na testu anksioznosti) (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Kline, 2000). Ovo je posebno uočljivo kod testova sposobnosti gde se može govoriti o efektima uvežbavanja. Zbog ovog nedostatka test-retest metode, predložena je metoda alternativnih formi.

3.2.2) Metoda alternativnih formi

Metoda alternativnih formi nadovezuje se na osnovnu ideju test-retest procene pouzdanosti, ali umesto da se isti test zadaje u dve vremenske tačke, zadaju se *alternativne forme* testa. Alternativne forme testova su zasebni testovi, koji sadrže različite ajteme, ali su namenjeni merenju iste osobine i tako napravljeni da imaju iste ili približno iste metrijske karakteristike (istu ukupnu težinu, kao i jednake težine pojedinačnih ajtema, jednaku korelaciju ajtema sa ukupnim skorom itd.) (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). U slučaju kada alternativne forme imaju i iste aritmetičke sredine i (što je za procenu pouzdanosti važnije) varijanse dobijenih skorova može se govoriti o *paralelnim formama* testa (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). O paralelnosti merenja biće više reči u nastavku teksta, u delu o modelima merenja.

Na primer, u slučaju testa anksioznosti umesto ajtema „Optimističan/na sam u pogledu budućnosti" iz prve forme, u drugoj alternativnoj formi mogli bismo ispitanicima prikazati ajtem

„Pesimističan/na sam u pogledu budućnosti" (negativno kodiran ajtem koji se odnosi na isti sadržaj) ili pak „Ne opterećujem se previše time šta će biti" (drugačije formulisan ajtem koji se odnosi na isti sadržaj). Ili, u testu znanja iz psihologije umesto pitanja koje se odnosi na uslovljavanje na vreme, možemo postaviti pitanje koje se tiče uslovljavanja na trag. Zadavanje alternativnih formi eliminiše (u najvećoj meri) efekat uvežbavanja, jer ispitanici u jednoj alternativnoj formi (uslovno rečeno „testu") odgovaraju na drugačije ajteme nego u drugoj alternativnoj formi (uslovno rečeno „retestu") (Bukvić, 1996).

Prateći istu logiku prikazanu u delu o test-retest pouzdanosti, i kod metode alternativnih formi koeficijent pouzdanosti računa se kao korelacija alternativnih formi. Dakle, proporcija varijanse koja odlazi na pravi skor jednaka je korelaciji dve alternativne forme testa. Moguće je napraviti i nekoliko alternativnih formi testa, u kom slučaju se pouzdanost računa kao prosečna korelacija alternativnih formi (Bukvić, 1996). Isto važi i za test-retest metodu ukoliko se test zadaje više od dva puta.

3.2.3) Metoda deljenja testa

Nadovezujući se dalje na ideju testiranja pouzdanosti metodom alternativnih formi, možemo se zapitati: ako već imamo alternativne forme testa – zašto moramo da čekamo da prođe neko vreme između dva zadavanja? Zašto ne bismo zadali dve forme istovremeno i onda ih korelirali? Dve alternativne forme testa koje se istovremeno zadaju mogu se posmatrati i kao dva dela jednog testa, što nas dovodi do toga da pouzdanost testa možemo utvrditi i deljenjem na dva ili više delova.

Treba imati u vidu da se značenje pouzdanosti sada menja. S obzirom na to da su dva dela testa zadata u istom vremenskom trenutku, više ne možemo govoriti o vremenskoj, već isključivo o unutrašnjoj konzistentnosti ili doslednosti testa.

Takođe, ako delove testa posmatramo kao paralelne forme – koreliranjem delova testa dolazimo do pouzdanosti delova, ali ne i celog testa. Jedan od najčešće korišćenih obrazaca za računanje pouzdanosti metodom deljenja testa jeste Spirman–Braunov obrazac ili S-B obrazac (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009), nazvan po imenima dvojice psihologa, koji su nezavisno jedan od drugog došli do iste formule za izračunavanje pouzdanosti celog testa na osnovu pouzdanosti delova(Brown, 1910; Spearman, 1910). Originalna namena Spirman–Braunove formule bila je da se proceni pouzdanost testa u slučaju skraćivanja ili produživanja, zbog čega se za nju nekad koristi i naziv „proročka" formula. Pouzdanost testa S-B obrascem procenjujemo kao:

$$r_{nn} = \frac{nr_{xx}}{1 + (n-1)r_{xx}}$$

gde je r_{nn} pouzdanost celog testa, n broj delova testa, a r_{xx} korelacija dva dela testa ili prosečna korelacija u slučaju više delova testa. Ovaj obrazac računanja pouzdanosti pretpostavlja da su delovi testa paralelni, odnosno da su im varijanse jednake, kao i da su korelacije delova testa (u slučaju kada se test deli na više od dva dela) takođe međusobno jednake.

U najjednostavnijem slučaju deljenja testa na dva dela (eng. *split-half* postupak, koji je izvorno i predložen), pouzdanost testa se računa kao:

$$r_{nn} = \frac{2r_{xx}}{1 + r_{xx}}$$

i jasno je da će pouzdanost celog testa biti jednaka 1 ukoliko je korelacija dva dela testa savršena, kao i da će pouzdanost biti 0 ukoliko između dva dela testa ne postoji korelacija. Takođe, pouzdanost testa veća je od pouzdanosti delova, odnosno od njihove korelacije. Isto važi i u slučaju deljenja testa na više delova, imajući u vidu da je *split-half* obrazac samo specijalni slučaj opštijeg obrasca za n = 2.

S obzirom na to da je test na više načina moguće podeliti na dva dela, postavlja se pitanje koji način je najbolji, odnosno koji način će obezbediti najveću paralelnost delova testa. Kako je kod testova sposobnosti uobičajeno da se stavke ređaju od najlakše do najteže, podela na prvu i drugu polovinu testa dovela bi do neujednačenosti težine ova dva dela. Zato se obično pribegava deljenju testa na parne i neparne stavke, jer se na taj način postiže da delovi budu što ujednačeniji mogući po težini (Bukvić, 1996).

Alternativna ideja procene pouzdanosti u slučaju deljenja testa na dva dela oslanja se na opšti obrazac pouzdanosti koji smo prethodno prikazali, $r_{tt} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$, odnosno zasniva se na proceni udela varijanse greške u ukupnoj varijansi. Na ovom principu počiva računanje dva obrasca pouzdanosti – Rulonovog i Gutmanovog – koji se mogu primenjivati u slučaju kada se test deli na dva dela (Bukvić, 1996).

Rulon je pošao od pretpostavke da se razlike u skorovima između dve polovine testa ili dve paralelne forme testa mogu smatrati indikatorom greške (Williams & Zimmerman, 1966). Logika koja leži u osnovi ove pretpostavke je sledeća. Ukoliko test podelimo na dva dela i znamo da su pravi skorovi ispitanika na ta dva dela isti, onda se razlike u skorovima ispitanika na jednom i drugom delu testa mogu pripisati samo slučajnoj grešci. Dakle, ako je neko na prvih 10 pitanja na testu znanja iz psihologije dobio skor 8, a na drugom delu istog testa skor 7 – očigledno postoji razlika u dobijenim skorovima. Iako nam pravi skor ispitanika nije poznat (on može biti i 7 i 8 i 7.5 ili neki drugi broj), znamo da je on isti na oba dela testa, tako da je razlika koju registrujemo zapravo posledica delovanja slučajne greške.

U skladu sa navedenim, Rulonov obrazac udeo varijanse greške u ukupnoj varijansi procenjuje kao količnik varijanse skorova razlike na dva dela testa i ukupne varijanse, a obrazac za izračunavanje pouzdanosti je:

$$r_{tt} = 1 - \frac{\sigma_D^2}{\sigma_X^2}$$

gde je r_{tt} pouzdanost testa, σ_D^2 varijansa skorova razlike na dva dela testa, a σ_X^2 varijansa celog testa. Ovaj obrazac može se koristiti samo kada test delimo na dva dela.

Na sličan način pouzdanost u slučaju deljenja testa na dva dela definiše i Gutman:

$$r_{tt} = 2(1 - \frac{\sigma_{X1}^2 + \sigma_{X2}^2}{\sigma_X^2})$$

gde je r_{tt} pouzdanost testa, σ_{X1}^2 i σ_{X2}^2 varijansa skorova na dva dela testa, a σ_X^2 varijansa celog testa (Bukvić, 1996; Callender & Osburn, 1979; Guttman, 1945).

Dakle, umesto varijanse razlike skorova koju imamo u Rulonovom obrascu, u Gutmanovom obrascu se za procenu greške koristi zbir varijansi delova testa. Da bismo razumeli logiku ovog obrasca, treba da se podsetimo da je izraz ispod razlomačke crte, varijansa celog testa, zapravo varijansa linearnog kompozita sačinjenog od dva dela testa. Drugim rečima, varijansa testa jednaka je zbiru varijansi delova testa (što je ujedno i izraz iznad razlomačke crte) i dvostrukoj kovarijansi delova testa, što možemo napisati i ovako:

$$r_{tt} = 2(1 - \frac{\sigma_{X1}^2 + \sigma_{X2}^2}{\sigma_{X1}^2 + \sigma_{X2}^2 + 2r_{X1X2}\sigma_{X1}\sigma_{X2}})$$

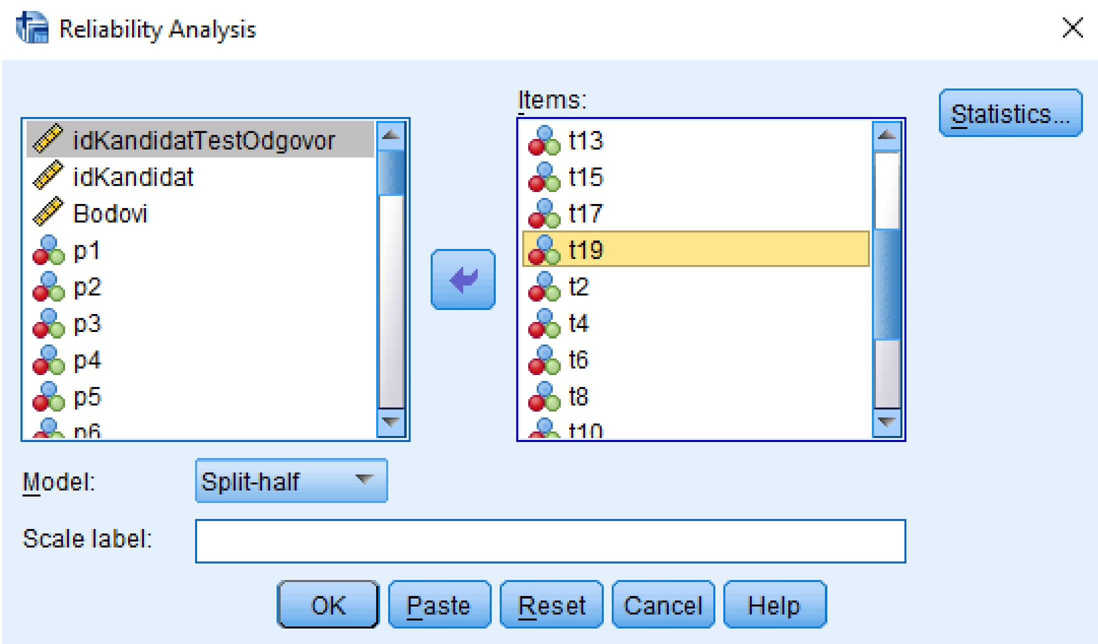
Što je korelacija između delova testa veća, to će σ_X^2 biti veća od zbira σ_{X1}^2 i σ_{X2}^2, te će i vrednost razlomka biti manja, odnosno pouzdanost testa veća. U ekstremnom slučaju kada između delova ne postoji korelacija, izraz ispod razlomačke crte biće jednak izrazu iznad razlomačke crte, te će pouzdanost biti 0.

Opštiji oblik Gutmanovog obrasca, čiji je prethodno prikazani obrazac samo specijalni slučaj (kada je broj delova testa dva), glasi:

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} \sigma_i^2}{\sigma_X^2}\right)$$

gde je n broj delova testa, σ_i^2 varijansa skorova na jednom od delova testa, a σ_X^2 varijansa celog testa. Opet, što je korelacija između stavki veća, to će i varijansa celog testa biti veća od zbira varijansi delova, te će i koeficijent pouzdanosti imati veću vrednost. Gutman je ovaj obrazac imenovao lambda 3, odnosno λ_3, dok je specijalni slučaj lambde 3, kada je broj delova testa n = 2 nazvao lambda 4 (λ_4) (Callender & Osburn, 1979).

Spirman–Braunov i Gutmanov koeficijent pouzdanosti deljenja testa na dva dela se u IBM SPSS-u mogu dobiti preko opcije **Analyze** → **Scale** → **Reliability Analysis**, biranjem opcije *Split-half* u okviru *Model* (Slika 3.2). Koristićemo kao primer prvih 20 pitanja na testu znanja iz psihologije. Iako je ovo test znanja, a ne sposobnosti, te pitanja nisu poređana po težini, radi demonstracije smo pitanja uneli redosledom koji odgovara par-nepar podeli testa na dva dela. Naime, treba naglasiti da SPSS kao prvu polovinu stavki prepoznaje one koje su prve unete u polje *Items*. Dakle, ako unesemo pitanja onim redom kojim se ona pojavljuju u testu, t1, t2, t3... do t20 – SPSS će pitanja t1 do t10 tretirati kao prvu, a pitanja t11 do t20 kao drugu polovinu testa. Zato, ako želimo da stavke podelimo na parne i neparne, moramo prvo uneti sve parne, a zatim sve neparne stavke ili obrnuto (prvo sve neparne, pa sve parne, kao što Slika 3.2 prikazuje).



Slika 3.2

U ispisu ove analize dobijamo dve tabele, od kojih ćemo se sada fokusirati na drugu, tabelu *Reliability Statistics* (Slika 3.3). U slučaju računanja pouzdanosti deljenjem testa na dva dela, ova tabela sadrži nekoliko pokazatelja pouzdanosti. Pre svega možemo videti pouzdanost svakog dela testa pojedinačno (*Part 1, Part 2*), pri čemu je pouzdanost delova izračunata kao Kronbahova alfa (*Cronbach's Alpha*, o kojoj će biti više reči u narednom odeljku). Pored pouzdanosti delova, u ispisu imamo informaciju i o broju stavki koje su ušle u svaki deo testa i njihovim imenima (komentar ispod tabele). Ovde vidimo da su stavke zaista podeljene na onih 10 koje su prvo unete i 10 koje su naredne unete. Takođe, imamo i informaciju o korelaciji između dva dela testa (*Correlation Between Forms*). Konačno, imamo i informaciju o Spirman–Braunovom koeficijentu pouzdanosti (*Spearman-Brown Coefficient*), kao i o Gutmanovom koeficijentu pouzdanosti deljenja testa na dva dela (*Guttman Split-Half Coefficient*).

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.691
		N of Items	10[a]
	Part 2	Value	.680
		N of Items	10[b]
	Total N of Items		20
Correlation Between Forms			.715
Spearman-Brown Coefficient	Equal Length		.834
	Unequal Length		.834
Guttman Split-Half Coefficient			.833

a. The items are: t1, t3, t5, t7, t9, t11, t13, t15, t17, t19.

b. The items are: t2, t4, t6, t8, t10, t12, t14, t16, t18, t20.

Slika 3.3

Spirman–Braunov koeficijent ima dve vrednosti, date u dva reda – za slučaj kada su polovine testa jednake dužine (*Equal Length*) i za slučaj kada nisu (*Unqual Length*). U našem slučaju, polovine su imale po 10 stavki, odnosno bile su jednake dužine, pa gledamo gornji red tabele. Da smo imali nejednak broj stavki, gledali bismo donji red[12], mada su u ovom konkretnom slučaju obe vrednosti numerički iste. Pouzdanost ovih 20 pitanja na testu znanja iz psihologije, iskazana Spirman–Braunovim koeficijentom je r_{tt} = .834, što znači da 83.4% varijanse ukupnih skorova potiče od prave varijanse, odnosno varijanse pravih skorova. Da li ovu vrednost možemo smatrati zadovoljavajućom? Postoje različite konvencije o tome koje vrednosti koeficijenata pouzdanosti[13] se smatraju dobrim (Clark & Watson, 1995; Peterson, 1994), a mi ćemo se rukovoditi time da je pouzdanost ispod .70 nezadovoljavajuća, preko .70 prihvatljiva (posebno u slučaju preliminarne verzije testa), preko .80 dobra, odnosno visoka, a preko .90 odlična, odnosno veoma visoka (Bukvić, 1996; Clark & Watson, 1995; R. J. Cohen & Swerdlik, 2018). Dakle, možemo zaključiti da je pouzdanost dobijena na osnovu ovih 20 pitanja bila dobra. Do istog zaključka dolazimo i ukoliko gledamo Gutmanov *split-half* koeficijent, koji iznosi je r_{tt} = .833. O konkretnim implikacijama (ovakve) pouzdanosti biće više reči u delu koji se bavi standardnom greškom i intervalima poverenja.

Na šta nam sve Slika 3.3 ukazuje, odnosno šta možemo primetiti gledajući ispis na njoj? Pre svega, možemo videti da je, kao što je i očekivano, pouzdanost celog testa veća od pouzdanosti njegovih delova. Takođe, možemo primetiti i da je pouzdanost veća od prosečne korelacije delova testova. Ako se setimo formule za Spirman–Braunov koeficijent pouzdanosti, te korelaciju delova testa zamenimo vrednošću dobijenom za test znanja iz psihologije (iz ispisa), dobijamo:

[12] U slučaju nejednakog broja stavki poželjno je prilikom definisanja analize prvo uneti neparne, a zatim parne stavke, jer SPSS po automatizmu deli stavke tako da prva polovina bude veća od druge (a neparnih stavki će uvek biti isto ili više nego parnih).

[13] Date preporuke formulisane su za Kronbahov alfa koeficijent interne konzistencije, ali ih možemo koristiti i za druge koeficijente istog tipa, odnosno druge koeficijente interne konzistencije.

$$r_{nn} = \frac{2r_{xx}}{1 + r_{xx}} = \frac{2 * .715}{1 + .715} = .834$$

odnosno vrednost Spirman–Braunovog koeficijenta pouzdanosti. Na kraju, primećujemo da su razlike u vrednosti pouzdanosti koju dobijamo na osnovu Spirman–Braunovog i Gutmanovog postupka deljenja testa na dva dela minimalne i zanemarljive. Ovo će biti slučaj za sve instrumente, tako da su oba obrasca podjednako informativna u pogledu pouzdanosti instrumenta (Bukvić, 1996).

3.2.4) Analiza stavki

Rekli smo da je istorijski najstarija metoda procene pouzdanosti bila test-retest metoda. Ovu metodu sledilo je korišćenje alternativnih/paralelnih formi testa, a zatim i deljenje testa na dva ili više delova. Nastavljajući ovaj način razmišljanja, dolazimo do toga da je test moguće podeliti na onoliko delova koliko on sadrži stavki. U ovom slučaju svaku pojedinačnu stavku posmatramo kao „mini" paralelnu formu testa, što zapravo znači da skor na svakom ajtemu posmatramo kao zbir pravog skora ispitanika (njegove osobine) i skora slučajne greške. Može se zato reći da metoda analize stavki predstavlja samo specijalan slučaj metode deljenja testa i ona se takođe odnosi na pouzdanost kao unutrašnju konzistentnost merenja (a ne na vremensku stabilnost skorova) (Bukvić, 1996).

Iako ovo može delovati pomalo apstraktno, možemo o tome razmišljati na sledeći način. Ako je pravi skor ispitanika na testu znanja iz psihologije koji ima 20 pitanja 15, to znači da taj ispitanik ima onoliko znanja koliko je potrebno za tačno rešavanje 75% pitanja na testu. Da je test imao 32 pitanja, pravi skor ispitanika bi bio 24 tačna odgovora, dok bi u slučaju 4 pitanja pravi skor bio 3 tačna odgovora. A šta ako test ima samo jedno pitanje? Onda bi pravi skor ispitanika bio 0.75. Ovakav skor, naravno, ne možemo dobiti ni na jednom pitanju, ali pretpostavljamo da će delovanje slučajne greške uticati na to da ispitanik nekad dobije skor 1, a nekad skor 0 (pri čemu bi verovatnoća tačnog odgovora bila tri puta veća od verovatnoće pogrešnog odgovora). Kao što smo već rekli, na osnovu dobijenog skora za pojedinačnog ispitanika ne možemo zaključiti o pravom skoru, ali možemo na većem uzorku ispitanika razložiti varijansu ukupnih skorova na varijansu pravih skorova i varijansu greške, koristeći neku od metoda procene pouzdanosti.

Najpoznatiji i najčešće korišćeni obrazac izračunavanja pouzdanosti (ne samo u okviru metode analize stavki, već i generalno) jeste Kronbahov alfa koeficijent pouzdanosti ili, kraće, Kronbahova alfa (eng. *Cronbach's alpha*) ili čak samo alfa (α) (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Peterson, 1994). Alfa se računa prema obrascu:

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} \sigma_s^2}{\sigma_X^2}\right)$$

gde je n broj stavki, σ_s^2 varijansa skorova na pojedinačnoj stavki, a σ_X^2 varijansa celog testa. Ukoliko uporedimo ovaj obrazac sa Gutmanovim opštim obrascem pouzdanosti za deljenje testa, vidimo da su identični, odnosno da je Kronbahova alfa isto što i Gutmanova lambda 3. Kronbahova alfa se može izvesti i iz Spirman–Braunovog obrasca pouzdanosti, a za zainteresovane čitaoce Dodatak 2 daje prikaz celog dokaza, odnosno izvođenja jedne formule iz druge.

Kronbahova alfa se u literaturi može pronaći i u nekim drugim oblicima (notacijama) (Bukvić, 1996), kao što je na primer:

$$r_{tt} = \frac{n}{n-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{n} \sigma_s^2}{\sigma_X^2}\right)$$

ali prostim izvođenjem jednačine možemo videti da su ova dva obrasca ekvivalentna:

$$r_{tt} = \frac{n}{n-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{n}\sigma_s^2}{\sigma_X^2}\right) = \frac{n}{n-1}\left(\frac{\sigma_X^2}{\sigma_X^2} - \frac{\sum_{i=1}^{n}\sigma_s^2}{\sigma_X^2}\right) = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n}\sigma_s^2}{\sigma_X^2}\right)$$

Još jedan od obrazaca za računanje Kronbahove alfe jeste i obrazac za standardizovane varijable:

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{n}{\sum\sum r_{ij}}\right)$$

gde je n broj stavki, a $\sum\sum r_{ij}$ suma elemenata matrice korelacija standardizovanih varijabli (stavki). Ukoliko su stavke standardizovane, suma varijansi pojedinačnih stavki jednaka je broju stavki – zato što je varijansa svake stavke jednaka 1. Takođe, od ranije nam je poznato da je varijansa linearnog kompozita jednaka zbiru varijansi pojedinačnih stavki i njihovih dvostrukih kovarijansi (odnosno zbiru elemenata matrice kovarijansi), a u slučaju kada su varijable standardizovane zbiru broja stavki (jer svaka stavka ima varijansu 1) i njihovih dvostrukih korelacija (odnosno zbiru elemenata matrice korelacija). Znajući ovo, uviđamo da je Kronbahov alfa obrazac za standardizovane varijable suštinski ekvivalentan obrascu za nestandardizovane varijable:

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n}\sigma_s^2}{\sigma_X^2}\right) = \frac{n}{n-1}\left(1 - \frac{n}{\sum\sum r_{ij}}\right)$$

pošto je suma varijansi stavki u slučaju nestandardizovanih varijabli $\sum_{i=1}^{n}\sigma_s^2$, a u slučaju standardizovanih n, dok je varijansa testa u nestandardizovanom obliku jednaka σ_X^2, a u standardizovanom $\sum\sum r_{ij}$.

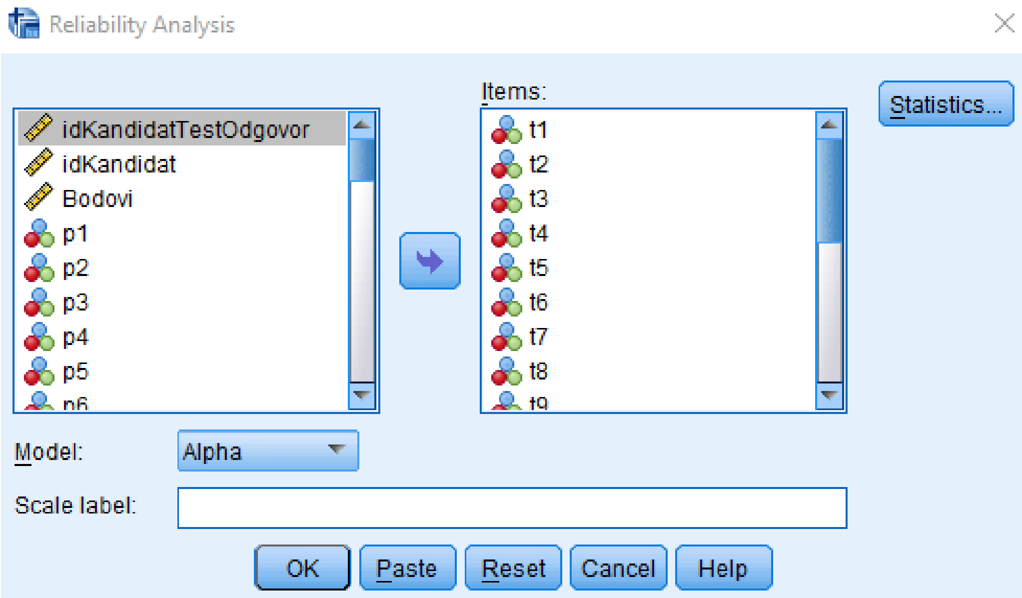
Konačno, u slučaju kada su stavke u testu binarne, odnosno kada su jedini mogući odgovori na pojedinačnu stavku 0 i 1, tj. tačno i netačno, Kronbahova alfa se može izračunati i preko sledećeg obrasca:

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum pq}{\sigma_X^2}\right)$$

gde su p i q, redom, proporcija tačnih i netačnih odgovora na pojedinačnu stavku. Proizvod proporcije tačnih i netačnih odgovora je zapravo varijansa stavke, te je jasno da je ova formula samo specijalni slučaj opštijeg obrasca Kronbahove alfe, kada su stavke binarne (Bukvić, 1996; Fajgelj, 2009). Ovaj obrazac se inače naziva Kuder–Ričardson obrazac ili KR-20 (eng. *Kuder–Richardson Formula 20*), prema autorima koji su ga prvi definisali.

Pregledom i upoređivanjem svih navedenih obrazaca pouzdanosti možemo primetiti dve pojedinosti. Prvo je da svi obrasci prate format prethodno navedenog opšteg obrasca za izračuvanje pouzdanosti, $r_{tt} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$. Sličnost je najuočljivija kod Rulonovog obrasca, ali se može primetiti i kod svih ostalih obrazaca za računanje pouzdanosti. Druga stvar koju primećujemo jeste da svi obrasci osim Rulonovog sadrže jedan dodatni element, najčešće $\frac{n}{n-1}$. Ovaj izraz predstavlja korekciju koja omogućava da vrednost koeficijenta pouzdanosti dostigne maksimalnu vrednost od 1. Kako Rulonov obrazac ne sadrži korekciju, koeficijenti pouzdanosti računati na osnovu Rulonovog obrasca uvek će biti niži od drugih koeficijenata pouzdanosti računatih na istim podacima.

Kronbahova alfa, odnosno Kuder–Ričardsonov obrazac 20 (u slučaju kada su stavke binarne), u SPSS-u se može dobiti preko **Analyze → Scale → Reliability Analysis** i biranjem opcije *Aplha* (što je inače i podrazumevana opcija) u okviru *Model* (Slika 3.4). U slučaju računanja alfa koeficijenta (ili bilo kog drugog koeficijenta koji počiva na metodi analize stavki), redosled kojim ćemo uneti ajteme u analizu nije bitan, zato što se svaka stavka posmatra kao zasebni „mini-test".



Slika 3.4

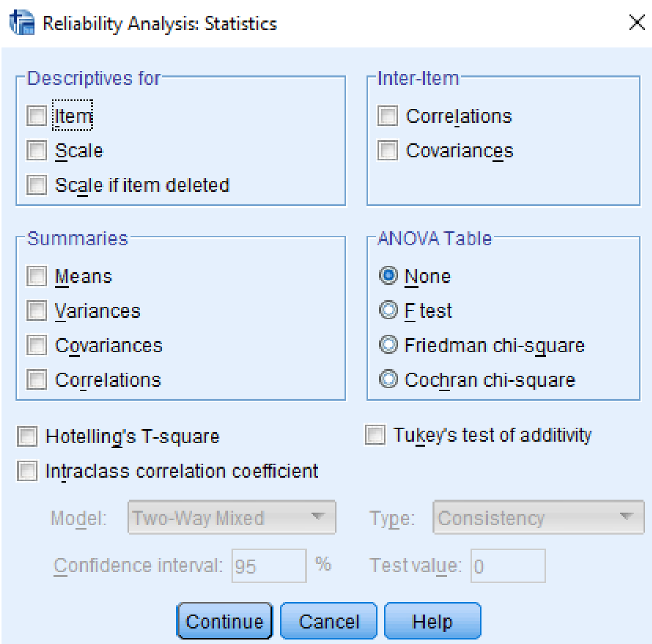
Osnovni ispis analize pouzdanosti je u ovom slučaju veoma mala tabela koja daje informaciju o vrednosti alfa koeficijenta i broju ajtema na osnovu kojih je on izračunat (Slika 3.5). U primeru testa znanja iz psihologije pouzdanost testa iznosi α = .817, što je slično vrednosti koju smo dobili primenom split-half metode i ukazuje na dobru pouzdanost instrumenta.

Reliability Statistics

Cronbach's Alpha	N of Items
.817	20

Slika 3.5

Pored mere pouzdanosti celog instrumenta, međutim, u okviru analize pouzdanosti u SPSS-u možemo dobiti brojne dodatne informacije o testu i stavkama. Klikom na dugme *Statistics* u okviru prozora *Reliability Analysis* otvara se novi prozor u kom možemo uključiti željene dodatne opcije (Slika 3.6). Da se podsetimo, u okviru ovog prozora nalazi se, između ostalog, opcija za računanje intraklasnog koeficijenta korelacije, koji se koristi za procenu pouzdanosti procenjivača (ovu opciju smo posmatrali kao vid procene objektivnosti merenja).



Slika 3.6

U okviru grupe opcija *Descriptives for* možemo izabrati da nam se prikažu deskriptivni pokazatelji za pojedinačne ajteme (*Item*), za ceo instrument (*Scale*) ili za instrument u slučaju da se dati ajtem izbaci iz instrumenta (*Scale if item deleted*). Označavanjem sve tri opcije u ispisu dobijamo tri dodatne tabele. Tabela *Item statistics* (Slika 3.7) prikazuje aritmetičke sredine i standardne devijacije svih stavki, što može biti korisno za procenjivanje težine i diskriminativnosti stavki. Tabela *Scale statistics* (Slika 3.8) daje informacije o celom instrumentu – aritmetičku sredinu, varijansu, standardnu devijaciju i broj ajtema. Standardna devijacija instrumenta važna je za računanje greške merenja, o čemu će biti više reči u narednom odeljku.

Item Statistics

	Mean	Std. Deviation	N
t1	.84	.365	436
t2	.59	.492	436
t3	.64	.481	436
t4	.96	.194	436
t5	.69	.461	436
t6	.56	.497	436
t7	.82	.384	436
t8	.46	.499	436
t9	.81	.391	436
t10	.71	.455	436
t11	.34	.475	436
t12	.62	.487	436
t13	.57	.496	436
t14	.76	.425	436
t15	.63	.484	436
t16	.83	.372	436
t17	.51	.500	436
t18	.60	.491	436
t19	.63	.483	436
t20	.86	.352	436

Slika 3.7

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
13.44	17.764	4.215	20

Slika 3.8

Tabela *Item-Total Statistics* (Slika 3.9) pruža informacije o tome šta bi se desilo sa instrumentom ukoliko bismo neku od stavki izbacili. Pre svega, možemo videti na koji način bi se promenile aritmetička sredina i varijansa testa. Poređenjem sa vrednostima u tabeli *Scale Statistics* primećujemo da izbacivanje svih stavki dovodi do smanjenja aritmetičke sredine instrumenta (što je sasvim logično, s obzirom na to da bi test onda imao stavku manje), kao i njegove varijanse (pri čemu stavke sa većom varijansom više doprinose smanjenju). Takođe, imamo informaciju o korigovanoj korelaciji stavke sa ukupnim skorom (*Corrected Item-Total Correlation*), već pominjanu u okviru lekcije o diskriminativnosti, a koja ukazuje na diskriminativnost, odnosno unutrašnju valjanost stavke. Konačno, u poslednjoj koloni tabele, *Cronbach's Alpha if Item Deleted*, vidimo šta bi se desilo sa pouzdanošću celog instrumenta ukoliko bismo izbacili neku stavku. Za većinu stavki važi da bi se njihovim izbacivanjem pouzdanost smanjila (pošto su vrednosti u koloni niže od dobijene pouzdanosti α = .817), ali postoje i tri ajtema čije bi izbacivanje povećalo pouzdanost (u pitanju su ajtemi t4, čije izbacivanje podiže pouzdanost na α = .819, i t13 i t14, čije izbacivanje menja vrednost pouzdanosti na α = .818).

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
t1	12.59	16.573	.355	.811
t2	12.84	15.520	.516	.802
t3	12.80	15.555	.521	.802
t4	12.47	17.533	.119	.819
t5	12.74	16.289	.339	.812
t6	12.88	15.769	.443	.806
t7	12.61	16.325	.417	.808
t8	12.98	15.928	.399	.809
t9	12.62	16.778	.260	.815
t10	12.73	15.932	.447	.806
t11	13.09	16.324	.316	.813
t12	12.82	16.089	.368	.811
t13	12.87	16.566	.236	.818
t14	12.67	16.814	.220	.818
t15	12.81	15.360	.572	.799
t16	12.60	16.452	.389	.810
t17	12.92	15.951	.391	.809
t18	12.84	15.616	.491	.804
t19	12.80	15.662	.489	.804
t20	12.58	16.681	.333	.812

Slika 3.9

Ukoliko obratimo pažnju na korigovane korelacije ova tri ajtema sa ukupnim skorom, možemo videti da su one najniže u tabeli (redom, .119, .236 i .220), i primetno niže od korelacija za druge ajteme. Sve ovo ukazuje da date stavke možda ne mere sasvim isti konstrukt kao i preostale stavke u instrumentu (pretpostavljamo da je to znanje iz psihologije) ili ga barem ne mere na najadekvatniji način.

Šta bi se desilo ukoliko bismo izbacili ove tri stavke iz testa i izračunali pouzdanost na preostalim ajtemima? Slika 3.10 prikazuje rezultat analize pouzdanosti za preostalih 17 ajtema. Možemo videti da je, uprkos manjem broju ajtema, test sada nešto pouzdaniji. Ipak, ova razlika je veoma mala (ukoliko bismo pouzdanost zaokruživali na dve decimale, u oba slučaja bismo je zaokružili na vrednost .82), što je zapravo u skladu sa očekivanim poboljšanjem pouzdanosti koje smo mogli videti u tabeli *Item-Total Statistics*. Naravno, ukoliko nam je cilj da skratimo instrument,

Reliability Statistics

Cronbach's Alpha	N of Items
.821	17

Slika 3.10

a zadržimo dobre metrijske karakteristike, trudićemo se da izbacujemo one stavke s najlošijim metrijskim karakteristikama (po svim relevantnim kriterijumima).

Još neke od opcija dostupnih u okviru *Reliability Analysis: Statistics* jesu *Inter-Item Correlations* i *Covariances* (Slika 3.6), koje u ispisu daju matricu korelacija, odnosno kovarijansi svih stavki u instrumentu. Ove matrice zbog veličine nećemo prikazivati u tekstu, ali mogu biti važne za uočavanje eventualnih neočekivanih korelacija ajtema. Na primer, postojanje negativnih korelacija između ajtema može ukazivati na to da ajtemi nisu (adekvatno) rekodirani, izostanak korelacije jedne varijable sa ostalima može ukazivati na to da je varijansa date varijable nulta (te je treba isključiti iz analize), maksimalna korelacija dve varijable takođe ukazuje na problem u podacima (npr., ista varijabla je dva puta uneta), što sve treba rešiti pre nego što pristupimo analizi pouzdanosti.

U okviru *Reliability Analysis: Statistics* možemo još da dobijemo i deskriptivne pokazatelje za aritmetičke sredine, varijanse, kovarijanse i korelacije ajtema (*Summaries: Means, Variances, Covariances, Correlations*; Slika 3.6). Biranjem opcija *Summaries: Means, Variances* i *Correlations* u ispisu dobijamo tabelu sa zgodnim pregledom aritmetičkih sredina i varijansi svih ajtema, kao i njihovih međusobnih korelacija (Slika 3.11).

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	.672	.342	.961	.619	2.812	.024	20
Item Variances	.198	.038	.250	.213	6.668	.003	20
Inter-Item Correlations	.178	-.056	.375	.432	-6.672	.007	20

Slika 3.11

Vrednosti iz prva dva reda su nam većinski već poznate iz tabele *Item Statistics*, dok se vrednosti iz reda *Inter-Item Correlations* mogu izračunati na osnovu matrice korelacija, a posebno su relevantne informacije o prosečnoj korelaciji između stavki, kao i o minimalnoj i maksimalnoj korelaciji. Primećujemo da je prosečna korelacija ajtema .178, što se može smatrati niskom korelacijom (ali ne nužno preniskom, s obzirom na to da smo dobili dobru pouzdanost celog instrumenta). Zanimljivo je da postoje dva ajtema čija je korelacija (minimalnih) -.056. Iako je predznak ove korelacije negativan, ona je veoma malog intenziteta, što implicira da nije značajna, odnosno da je korelacija nula. Ipak, pregledom pune matrice korelacija trebalo bi identifikovati ova dva pitanja, razmotriti njihov sadržaj i po potrebi izbaciti jedno od njih (ono koje ima nižu korigovanu korelaciju sa ukupnim skorom i manje doprinosi pouzdanosti instrumenta) ili oba. Maksimalna korelacija je, sa druge strane, .375, što takođe nije previše visoka vrednost, ali treba imati u vidu da svako pitanje na testu zahvata drugačije aspekte znanja psihologije, pa i ne možemo očekivati veoma visoke korelacije između njih. Prema nekim autorima, u slučaju široko definisanih konstrukata prosečna korelacija između ajtema od .15 - .20 smatra se poželjnom, dok se za instrumente namenjene merenju usko definisanih konstrukata očekuje viša prosečna korelacija od .40 - .50 (Clark & Watson, 1995).

3.3) Modeli merenja u klasičnoj teoriji testa

Klasična teorija testa, iskazana izrazom X = T + E, zapravo predstavlja jednu jednačinu sa dve nepoznate. Već smo rekli da nam je u bilo kojoj situaciji realnog merenja jedino poznat ukupni skor ispitanika, ali da su njegov pravi skor i skor greške nepoznati (25 = 24 + 1, ali 25 = 26 + (-1) ili 25 = 23 + 2). Rekli smo isto tako i da se računanje koeficijenata pouzdanosti zasniva na određenim procenama varijanse pravog skora, odnosno (češće) varijanse greške. Da bismo mogli da dođemo do procena pouzdanosti, potrebno je da napravimo određene pretpostavke o prirodi podataka koje omogućavaju da rešimo ovu jednačinu sa dve nepoznate (Fajgelj, 2009). Neke od tih pretpostavki već su bile pomenute, ali se njima nismo eksplicitnije bavili.

Većina obrazaca pouzdanosti počiva na nekoj varijanti tzv. *modela paralelnih indikatora*. Šta su paralelni indikatori? U zavisnosti od metode procene pouzdanosti indikator može biti ceo test (u jednom trenutku zadavanja), paralelna forma testa, deo testa ili pojedinačna stavka (Fajgelj, 2009). A na šta se odnosi paralelnost? U osnovnom modelu paralelnih indikatora paralelnost se definiše preko jednakosti pravih skorova i jednakosti varijanse greške (Bukvić, 1996; DeVellis, 2006; Fajgelj, 2009; Novick, 1966). Ako je pravi skor ispitanika na dva dela testa isti, a iste su i varijanse grešaka – onda su ta dva dela testa paralelna. Ako su pravi skor ispitanika i varijanse grešaka na testu i retestu jednaki – takođe se radi o paralelnim indikatorima. Paralelne stavke su stavke na kojima je pravi skor ispitanika jednak i koji imaju jednake varijanse grešaka. Iz ove pretpostavke slede i neke druge pretpostavke kao što su jednakost varijansi i kovarijansi ukupnih skorova, a u nekim strožim varijantama modela paralelnih indikatora i jednakost aritmetičkih sredina i varijansi pravih skorova (Bukvić, 1996; Fajgelj, 2009).

Postoje, sa druge strane, i modeli koji postavljaju manje stroge uslove paralelnosti ajtema. Na primer, Kronbahov alfa koeficijent počiva na nešto blažim pretpostavkama modela tzv. *tau-ekvivalentnih indikatora*. Za ove indikatore se pretpostavlja da su im samo pravi skorovi (obeleženi grčkim slovom *tau*, τ) jednaki, dok varijanse ne moraju biti, kao ni varijanse grešaka (Dunn et al., 2014; Fajgelj, 2009; Morera & Stokes, 2016; Zinbarg et al., 2005). Iako možda tako ne deluje na prvi pogled, i tau-ekvivalentnost je teško ostvariva u praksi, jer podrazumeva da je svaki ajtem nekog testa jednako dobar indikator konstrukta na koji se odnosi (Dunn et al., 2014; Fajgelj, 2009; Morera & Stokes, 2016).

Još manje restriktivni su tzv. *kongenerički modeli*, koji dozvoljavaju da i pravi skorovi indikatora budu različiti, a pretpostavljaju samo da se indikatori odnose na isti konstrukt. U koeficijente pouzdanosti zasnovane na kongeneričkom modelu spada MekDonaldova (McDonald) omega (ω), čije računanje počiva na hijerarhijskom faktorskom modelu. Omega nije uključena u SPSS statistički softver i ovde je nećemo detaljnije prikazivati, ali je važno naglasiti da se pokazalo da je u situacijama kada su pretpostavke tau-ekvivalentnog modela zadovoljene, omega koeficijent jednako dobar pokazatelj pouzdanosti testa kao i alfa, dok je u situacijama kada ove pretpostavke nisu ispunjene, on superiorniji, odnosno tačnije procenjuje stvarnu pouzdanost testa (Dunn et al., 2014; Fajgelj, 2009; Morera & Stokes, 2016; Zinbarg et al., 2005).

Kao što smo videli, svi modeli merenja predstavljaju aproksimacije stvarnosti, ali se razlikuju po tome koliko stroge uslove postavljaju za računanje koeficijenata pouzdanosti. Što su modeli restriktivniji, to je veća verovatnoća da dođe do narušavanja njihovih pretpostavki. Kakve ovo posledice ima po merenje? Odstupanja od pretpostavki modela impliciraju da je procena koju dobijamo u izvesnoj meri pogrešna. Greška koju pravimo tipično ide u smeru potcenjivanja stvarne pouzdanosti testa[14], ali je u konkretnoj situaciji teško proceniti o kakvoj i kolikoj grešci se radi (Novick & Lewis, 1967; Sheng & Sheng, 2012). Zbog toga se sve više preporučuje upotreba modela za koje je manje verovatno da će im uslovi biti narušeni, a jedan od njih je Gutmanov model merenja.

[14] Pouzdanost ćemo češće potceniti nego preceniti zbog toga što narušavanje pretpostavki modela najčešće vodi povećanoj proceni slučajne greške pri izračunavanju pouzdanosti (odstupanje od modela se smatra greškom).

3.4) Procena pouzdanosti u Gutmanovom modelu merenja

Gutmanov model merenja polazi od drugačijih pretpostavki i na drugačiji način procenjuje varijansu pravih skorova i skorova greške u poređenju sa prethodno prikazanim koeficijentima pouzdanosti. Gutman smatra da je prava varijansa varijable onaj deo njene varijanse predvidljiv na osnovu svih drugih varijabli koje mere isti konstrukt (Fajgelj, 2009; Guttman, 1953). Ovde se misli na sve moguće varijable u univerzumu, odnosno na sve potencijalne stavke koje bismo mogli sastaviti da merimo željeni konstrukt (npr., znanje psihologije). Ova varijansa se u Gutmanovom modelu naziva varijansom imaža (eng. i franc. *image*). Imaž, odnosno prava varijansa konkretne stavke, može se proceniti kao predviđena varijansa na osnovu svih drugih stavki u testu, odnosno kao koeficijent multiple determinacije (R^2) u regresionom modelu. Sa druge strane, varijansa greške ili varijansa antiimaža je u Gutmanovom modelu definisana kao deo varijanse pojedinačne varijable (indikatora, stavke) koji se ne može predvideti drugim varijablama (indikatorima, stavkama) u univerzumu. Antiimaž se procenjuje kao koeficijent alijenacije ($1 - R^2$) u regresionom modelu gde pojedinačnu stavku predviđamo svim drugim stavkama u instrumentu.

Gutman je predložio šest različitih obrazaca za računanje pouzdanosti, koje je označio grčkim slovom lambda (λ) i numerisao od λ_1 do λ_6 (Callender & Osburn, 1979; Guttman, 1945). On je smatrao da svi ovi koeficijenti predstavljaju donje granice pouzdanosti (prikladne za upotrebu u različitim situacijama), s obzirom na to da stvarna pouzdanost testa mora biti jednaka ili veća vrednosti dobijenoj korišćenjem datih obrazaca (Guttman, 1945). Koeficijent lambda 6 (λ_6) zapravo predstavlja pouzdanost dobijenu na osnovu podele varijanse testa na varijansu imaža i antiimaža i dobija se prema sledećoj formuli:

$$\lambda_6 = 1 - \frac{\sum e_j^2}{\sigma_x^2}$$

gde su e_j^2 varijanse grešaka pojedinačnih ajtema, odnosno varijanse pojedinačnih antiimaža dobijene na osnovu regresije, a σ_x^2 varijansa testa. Vidimo da je u ovom koeficijentu model imaža i antiimaža najdirektnije moguće primenjen na računanje pouzdanosti. Koeficijenti lambda 2 (λ_2) i lambda 5 (λ_5) takođe počivaju na proceni varijanse imaža, ali ih ovde nećemo prikazivati.

Gutmanov koeficijent lambda 1 (λ_1) računa se prema formuli $\lambda_1 = 1 - \frac{\sum \sigma_j^2}{\sigma_x^2}$, gde je σ_j^2 varijansa pojedinačnog ajtema. Kako ovaj obrazac ne sadrži korekciju $\frac{n}{n-1}$, on će od svih Gutmanovih koeficijenata uvek imati najnižu vrednost (Callender & Osburn, 1979; Guttman, 1945), zbog čega se ponekad naziva apsolutnom donjom granicom pouzdanosti (Knežević & Momirović, 1996). Koeficijent pouzdanosti λ_3 se dobija primenom date korekcije na koeficijent λ_1 i, kao što je već pomenuto, ekvivalentan je Kronbahovom alfa koeficijentu: $\lambda_3 = \frac{n}{n-1}\lambda_1 = \frac{n}{n-1}(1 - \frac{\sum \sigma_j^2}{\sigma_x^2})$. Takođe, Gutmanova λ_4 zapravo je već pomenuti Gutmanov split-half koeficijent pouzdanosti[15], koji se koristi u slučaju deljenja testa na dva dela $\lambda_4 = 2(1 - \frac{\sigma_{x1}^2 + \sigma_{x2}^2}{\sigma_x^2})$.

Sve Gutmanove koeficijente pouzdanosti u SPSS-u možemo dobiti ukoliko u okviru **Analyze → Scale → Reliability Analysis** izaberemo opciju *Guttman* u okviru *Model*. U ispisu ove analize (Slika 3.12) dobijamo šest koeficijenata pouzdanosti. Poređenjem dobijenih vrednosti svih lambda koeficijenata uočavamo da je vrednost lambda 1 najniža, kao što bi po definiciji i trebalo da bude.

---

[15] Pokazuje se da u određenim okolnostima (ukoliko se izabere najadekvatniji mogući način podele stavki na dva dela) koeficijent λ_4 može imati najveću vrednost među Gutmanovim koeficijentima, te se može smatrati „najvećom donjom granicom” pouzdanosti (Callender & Osburn, 1979).

Takođe, ako uporedimo vrednost lambde 4 sa vrednošću Gutmaovog split-half koeficijenta iz odeljka o deljenju testa, primećujemo da je to ista vrednost od .833 (treba napomenuti da smo i sada prvo uneli neparne, a zatim parne stavke), kao što je vrednost lambde 3 jednaka Kronbahovom alfa koeficijentu i iznosi .817.

Reliability Statistics

Lambda	1	.777
	2	.824
	3	.817
	4	.833
	5	.806
	6	.
N of Items		20

Slika 3.12

Primećujemo, međutim, da u ispisu nismo dobili vrednost lambde 6 (ovo je praćeno i porukom upozorenja koju SPSS izbacuje u ispisu). Takođe, u tabeli *Item-Total Statistics* dobijamo dodatnu kolonu *Squarred Multiple Correlation*, u kojoj se nalaze vrednosti imaža za pojedinačne stavke, ali je ona prazna (Slika 3.13). Do ovoga najčešće dolazi u slučaju kada matrica kovarijansi iz nekog razloga[16] nije pogodna za analizu (SPSS će izbaciti poruku upozorenja na početku ispisa koja nas o tome obaveštava). Jedno od rešenja jeste da se umesto sirovih varijabli koriste standardizovane. Varijanse sirovih varijabli su slične (s obzirom na to da imamo istu skalu odgovora – tačno ili netačno), ali ne i iste – dok su varijanse svih standardizovanih stavki međusobno jednake i iznose 1. Stoga standardizacija varijabli može učiniti da matrica kovarijansi (određena korelacijama između varijabli, koje se standardizacijom ne menjaju, i varijansama, koje se ujednačavaju) postane pogodna za analizu.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
t1	12.59	16.573	.355	.	.811
t3	12.80	15.555	.521	.	.802
t5	12.74	16.289	.339	.	.812
t7	12.61	16.325	.417	.	.808
t9	12.62	16.778	.260	.	.815
t11	13.09	16.324	.316	.	.813
t13	12.87	16.566	.236	.	.818
t15	12.81	15.360	.572	.	.799
t17	12.92	15.951	.391	.	.809
t19	12.80	15.662	.489	.	.804
t2	12.84	15.520	.516	.	.802
t4	12.47	17.533	.119	.	.819
t6	12.88	15.769	.443	.	.806
t8	12.98	15.928	.399	.	.809
t10	12.73	15.932	.447	.	.806
t12	12.82	16.089	.368	.	.811
t14	12.67	16.814	.220	.	.818
t16	12.60	16.452	.389	.	.810
t18	12.84	15.616	.491	.	.804
t20	12.58	16.681	.333	.	.812

Slika 3.13

[16] Ovo je najčešće posledica nulte ili veoma niske varijanse ajtema ili pak maksimalnih ili veoma visokih korelacija između parova ajtema.

Kako bismo prikazali kompletan ispis u slučaju primene Gutmanovog modela, ponovićemo analizu ali na standardizovanim stavkama testa znanja iz psihologije. Tabele koeficijenata pouzdanosti i *Item-Total Statistics* su sada kompletne (Slika 3.14 i Slika 3.15).

Reliability Statistics

Lambda	1	.772
	2	.817
	3	.812
	4	.852
	5	.799
	6	.819
N of Items		20

Slika 3.14

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Zt1 Zscore(t1)	0E-7	80.175	.357	.173	.805
Zt2 Zscore(t2)	0E-7	77.500	.516	.316	.797
Zt3 Zscore(t3)	0E-7	77.415	.521	.294	.796
Zt4 Zscore(t4)	0E-7	84.372	.120	.075	.818
Zt5 Zscore(t5)	0E-7	80.595	.333	.179	.807
Zt6 Zscore(t6)	0E-7	78.922	.431	.243	.801
Zt7 Zscore(t7)	0E-7	79.085	.421	.205	.802
Zt8 Zscore(t8)	0E-7	79.661	.387	.206	.804
Zt9 Zscore(t9)	0E-7	81.783	.265	.111	.810
Zt10 Zscore(t10)	0E-7	78.734	.442	.245	.801
Zt11 Zscore(t11)	0E-7	81.030	.308	.153	.808
Zt12 Zscore(t12)	0E-7	79.927	.372	.181	.805
Zt13 Zscore(t13)	0E-7	82.490	.225	.088	.812
Zt14 Zscore(t14)	0E-7	82.514	.224	.112	.812
Zt15 Zscore(t15)	0E-7	76.567	.572	.372	.794
Zt16 Zscore(t16)	0E-7	79.588	.392	.227	.804
Zt17 Zscore(t17)	0E-7	79.687	.386	.181	.804
Zt18 Zscore(t18)	0E-7	77.889	.492	.309	.798
Zt19 Zscore(t19)	0E-7	77.955	.488	.272	.798
Zt20 Zscore(t20)	0E-7	80.497	.339	.167	.806

Slika 3.15

Koeficijenti pouzdanosti koje sada dobijamo malo su drugačiji od onih dobijanih na nestandardizovanim stavkama (s obzirom na to da matrica kovarijansi na osnovu koje su računati nije identična[17]), ali njihovi međusobni odnosi su isti. Lambda 1 ima najnižu vrednost, dok su

[17] Alfa na standardizovanim stavkama se nekad naziva i standardizovanom alfom, a računa se na osnovu matrice korelacija umesto matrice kovarijansi. Standardizovana alfa podrazumeva da svi ajtemi imaju jednake varijanse, što u praksi često nije slučaj. Ukoliko planiramo da rezultat na testu izrazimo

vrednosti Lambde 3 i 4 jednake Kronbahovoj alfi i Gutmanovom split-half koeficijentu kada bismo ih računali na standardizovanim stavkama. Vrednost lambde 6, odnosno pouzdanost imaža slična je vrednosti Kronbahove alfe, ali je od nje nešto viša.

Ovde treba prokomentarisati da bi razlike u koeficijentima pouzdanosti koje se dobijaju primenom različitih obrazaca trebalo da budu male i da generalno vode istom zaključku o pouzdanosti instrumenta. Kada su razlike između lambde 6 i drugih koeficijenata pouzdanosti male (kao u našem primeru), to ukazuje da su pretpostavke o paralelnosti indikatora u najvećoj meri bile ispunjene. U slučaju većih razlika između koeficijenata pouzdanosti možemo zaključiti da indikatori nisu (u dovoljnoj meri) paralelni, te je za procenu pouzdanosti smislenije koristiti koeficijente zasnovane na Gutmanovom modelu imaža i antiimaža.

Konačno, videli smo da u slučaju kada analizu pouzdanosti radimo na sirovih 20 ajtema ne dobijamo vrednosti kolone *Squarred Multiple Correlation* u okviru *Item-Total Statistics* tabele, odnosno nemamo informaciju o imažu pojedinačnih stavki. Na koji još način možemo doći do imaža, odnosno pouzdanosti pojedinačne stavke? Pošto znamo da se imaž stavke procenjuje kao multiplo R^2 iz regresije u kojoj je data stavka kriterijum, a sve ostale stavke na testu prediktori, primenom regresione analize lako možemo doći do pouzdanosti svih pojedinačnih stavki.

Ilustrovaćemo ovo na primeru prvih 20 stavki testa znanja iz psihologije. Recimo da nas zanima koliki je imaž stavke 1. U regresionu analizu kao kriterijum unosimo skor na ovoj stavki (t1), dok kao prediktore unosimo stavke t2 do t20. U ispisu analize (Slika 3.16) vidimo da koeficijent multiple determinacije iznosi $R^2 = .173$, što je upravo ona vrednost koju smo mogli videti kao *Squarred Multiple Correlation* za ovu stavku u tabeli *Item-Total Statistics*[18]. Drugim rečima, oko 17% varijanse ove pojedinačne stavke može se objasniti drugim stavkama u instrumentu. Po pravilu, imaž stavke je osetljiv na broj stavki instrumenta – što instrument ima više stavki, to je moguće predvideti i veći procenat varijanse bilo koje od njih na osnovu svih drugih. S obzirom na to da smo u ovom primeru predviđali stavku na osnovu samo 19 drugih stavki, relativno niska vrednost imaža nije iznenađujuća.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.416[a]	.173	.136	.340

a. Predictors: (Constant), t20, t14, t13, t4, t11, t9, t16, t8, t5, t17, t6, t12, t7, t10, t19, t18, t3, t2, t15

Slika 3.16

3.5) Pouzdanost, greška merenja i interval poverenja

Na početku ovog poglavlja rekli smo da se pouzdanost odnosi na stepen preciznosti merenja koji možemo postići određenim instrumentom. Svakako, viša vrednost koeficijenta pouzdanosti (u slučaju da se radi o internoj konzistenciji testa) znači da je instrument precizniji, ali kako su tačno ove dve stvari povezane?

---

kao sumu ili prosek sirovih skorova (nasuprot standardizovanim skorovima) na ajtemima, prikladnije je upotrebiti nestandardizovanu alfu (Falk & Savalei, 2011).

[18] Ista vrednost bi se dobila i da smo standardizovani skor z1 predviđali standardizovanim skorovima z2 do z20. Ovo je zbog toga što je imaž isti kada se računa na osnovu standardizovanih i nestandardizovanih varijabli, s obzirom na to da su za računanje multiplog R^2 važne korelacije, a ne kovarijanse između varijabli.

Podsetimo se ponovo klasične testovne teorije koja predviđa da se varijansa skorova na testu može razložiti na varijansu pravih skorova i varijansu greške, $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$. Ovu formulu smo dalje razložili do opšteg obrasca računanja pouzdanosti $r_{tt} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$. Sada ćemo, koristeći istu notaciju, izvesti formulu za standardnu grešku merenja, koja je kvadratni koren varijanse greške merenja:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$
$$\sigma_E^2 = \sigma_X^2 - \sigma_T^2$$
$$\frac{\sigma_E^2}{\sigma_X^2} = \frac{\sigma_X^2}{\sigma_X^2} - \frac{\sigma_T^2}{\sigma_X^2}$$
$$\frac{\sigma_E^2}{\sigma_X^2} = 1 - r_{tt}$$
$$\sigma_E^2 = \sigma_X^2(1 - r_{tt})$$
$$\sigma_E = \sqrt{\sigma_X^2(1 - r_{tt})}$$
$$\sigma_E = \sigma_X\sqrt{(1 - r_{tt})}$$

gde je σ_E standardna greška merenja, σ_X standardna devijacija skorova na testu i r_{tt} pouzdanost testa (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Kline, 2000).

Ako se vratimo na primer testa znanja iz psihologije, standardna devijacija testa iznosi SD = 4.215, dok je pouzdanost jednaka α = .817. Standardnu grešku merenja onda dobijamo po formuli $SEm = SD\sqrt{(1-\alpha)} = 4.215 * \sqrt{1 - .817} = 4.215 * \sqrt{.183} = 1.803$. Šta ovaj broj predstavlja? Standardna greška merenja sama po sebi ukazuje na standardnu devijaciju skorova greške na instrumentu i bitno je zapamtiti da je izražena u jedinicama skorova testa (pošto u formulu za njeno računanje ulazi standardna devijacija skorova na testu). U našem primeru, skorovi na testu imaju teorijski raspon od 0 do 20 (koliko je bilo pitanja), tako da se vrednost 1.803 odnosi na ove skorove. Da smo skor na testu iskazali kao z skor, standardna greška merenja bi takođe bila z skor. Standardna greška merenja, međutim, retko se koristi i interpretira sama za sebe. Ona uglavnom služi kao osnova za izračunavanje intervala poverenja. Interval poverenja ili interval pouzdanosti predstavlja interval u kom verujemo da se, sa određenim stepenom sigurnosti, nalazi pravi skor ispitanika (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018).

Ovde ćemo se ponovo podsetiti procene aritmetičke sredine populacije na osnovu vrednosti dobijene na uzorku. Rekli smo da je distribucija uzorkovanja normalna distribucija, te da će najveći broj uzoraka davati procene parametra koje su veoma blizu njegovoj stvarnoj vrednosti, dok će manji broj njih precenjivati ili potcenjivati parametar, a tek veoma mali broj uzoraka će puno precenjivati ili potcenjivati populacijsku vrednost. Istu logiku primenjivali smo kada smo određivali značajnost skjunisa i kurtozisa, a istu logiku primenjujemo i sada, kada želimo da utvrdimo u kom intervalu se najverovatnije nalazi pravi skor ispitanika.

Ukoliko više puta merimo osobinu ispitanika (npr., znanje iz psihologije) istim testom (pretpostavljajući za potrebe primera da ne postoje efekti prethodnih testiranja), usled delovanja slučajne greške dobijeni skorovi ispitanika neće uvek savršeno odražavati njegov pravi skor. Kako je bitna osobina slučajne greške upravo to da ona slučajno deluje – dobijeni skor ispitanika nekada će biti veći, a nekada manji od njegovog pravog skora. Ukoliko je, recimo, pravi skor ispitanika 25 tačnih odgovora od 30, dobijeni skor ispitanika nekada će biti 24, nekada 26, ponekad će biti upravo 25, ali se ponekad može desiti da ispitanik dobije i skor 20 ili 30. Verovatnoća pojavljivanja ovih skorova takođe prati normalnu distribuciju, te će skor 25 biti najčešći, takođe česti će biti i skorovi 24 i 26, dok će najređi biti skorovi 20 i 30. Šta su aritmetička sredina i standardna devijacija ove distribucije?

Aritmetička sredina je, naravno, pravi skor ispitanika, jer je to vrednost koju ćemo najčešće dobijati. A šta određuje stepen raspršenja dobijenih skorova oko pravog skora? U pitanju je standardna greška merenja, odnosno standardna devijacija skorova greške na instrumentu.

Znajući ovo, lako možemo utvrditi u kom rasponu će se nalaziti pravi skor ispitanika sa verovatnoćom 95%, odnosno 99%. Formula za računanje intervala poverenja (eng. *confidence interval*, CI) je:

$$CI_{95} = X \pm 1.96 * SEm$$
$$CI_{99} = X \pm 2.58 * SEm$$

gde je X dobijeni skor ispitanika[19], a SEm standardna greška merenja (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018). Prikazane formule se odnose na klasičan pristup računanju intervala poverenja (oko dobijenog skora ispitanika), a postoje i alternativni pristupi koji koriste procenjeni pravi skor ispitanika, regresioni pristup i/ili linearne transformacije (Charter & Feldt, 2000), ali se njima nećemo detaljnije baviti.

Primenimo ovo na primeru 20 pitanja testa znanja iz psihologije i zamišljenog kandidata koji je na ovom delu testa osvojio 10 poena. U kom rasponu se nalazi njegov pravi skor? Sa 95% sigurnosti možemo tvrditi da se pravi skor ovog ispitanika nalazi između X – 1.96*SEm = 10 – 1.96*1.803 = 6.466 i X + 1.96*SEm = 10 + 1.96*1.803 = 13.534 poena na ovom delu testa. Drugim rečima, interval poverenja obuhvata oko 3.5 skora ispod i iznad dobijenog skora ispitanika.

Kako se interval poverenja menja sa promenom pouzdanosti? Zamislimo, recimo, da je pouzdanost umesto .817 bila .917. Standardna greška merenja bi onda bila $SEm = 4.215 * \sqrt{1 - .917} = 1.214$, a interval poverenja za kandidata koji je osvojio 10 poena bio bi između 10 – 1.96*1.214 = 7.620 i 10 + 1.96*1.214 = 12.380. Za pouzdanost 0.977 dobijamo standardnu grešku merenja 0.639 i interval poverenja 8.747 do 11.253, dok je za pouzdanost .997 standardna greška merenja 0.231, a interval poverenja 9.548 do 10.452. Dakle, što je pouzdanost veća, to je standardna greška merenja manja, a interval poverenja uži. Zbog toga je veoma važno da, sa jedne strane, težimo što većoj pouzdanosti instrumenta i, sa druge strane, razumemo implikacije pouzdanosti po preciznost merenja i stepen poverenja u zaključke koje na osnovu merenja donosimo.

3.6) Faktori koji utiču na pouzdanost

Od čega sve zavisi pouzdanost instrumenta? Poznavanje faktora koji utiču na pouzdanost, kao i odnosa pouzdanosti i drugih metrijskih karakteristika važno je iz nekoliko razloga. Pre svega, ovo znanje nam pomaže da razumemo zbog čega jedan instrument ima veću pouzdanost od drugog, zatim šta možemo učiniti da bismo povećali pouzdanost nekog instrumenta, ali i na koji način treba interpretirati rezultate testa koji ima nezadovoljavajuću pouzdanost.

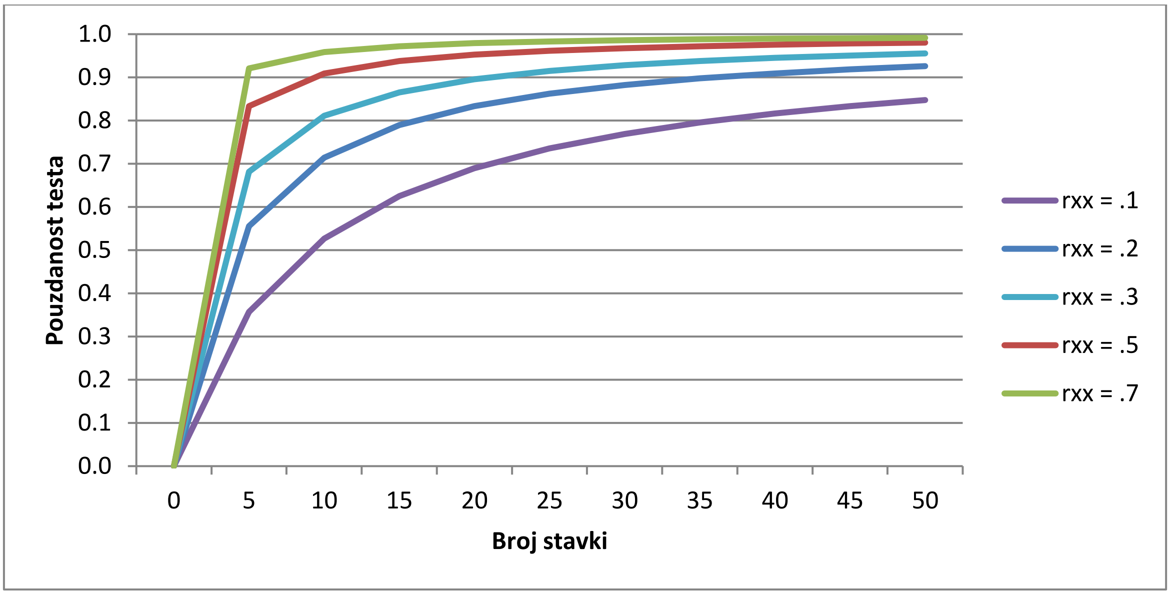
3.6.1) Broj i kvalitet stavki i pouzdanost

Slično kao i kod diskriminativnosti, broj stavki direktno je proporcionalan pouzdanosti instrumenta. Drugim rečima, što više stavki instrument ima, to će biti pouzdaniji (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Ukoliko se setimo Spirman–Braunovog obrasca pouzdanosti, $r_{nn} = \frac{nr_{xx}}{1+(n-1)r_{xx}}$, vidimo da povećavanje broja delova testa (stavki) nužno dovodi i do povećanja pouzdanosti.

Ovaj odnos, međutim, nije linearan, već odgovara negativno ubrzanoj krivoj (Slika 3.17). Ovo znači da je (pod pretpostavkom jednakih korelacija između ajtema) povećanje testa za, na primer,

[19] Zamena za pravi skor koja nam je direktno dostupna, mada ne nužno i najbolja (Fajgelj, 2009).

pet stavki kod niskih pouzdanosti dovoljno da primetno poboljša pouzdanost, dok za više vrednosti isti skok u broju ajtema dovodi do daleko manjih promena u pouzdanosti.



Slika 3.17

Ovaj grafik možemo posmatrati i iz drugog ugla – koliko puta je potrebno povećati broj stavki instrumenta da bismo dostigli određenu željenu pouzdanost. Iz Spirman–Braunovog obrasca pouzdanosti može se, jednostavnim rešavanjem po n (videti Dodatak 2), izvesti tzv. proročka formula koja nam daje upravo ovu informaciju:

$$n = \frac{r_{nn}(1 - r_{xx})}{r_{xx}(1 - r_{nn})}$$

gde je r_{xx} trenutna pouzdanost testa, r_{nn} željena pouzdanost testa, a n broj puta koliko treba uvećati test (odnosno broj kojim treba pomnožiti trenutni broj stavki) da bi se dobila željena pouzdanost. Naravno, ovde se pretpostavlja da stavke koje dodajemo ne menjaju prosečnu korelaciju stavki, odnosno da je prosečna korelacija stavki koje dodajemo jednaka prosečnoj korelaciji stavki već prisutnih u testu.

Ilustrujmo ovo primerom. Ako imamo test od 15 stavki čija je pouzdanost .65, a želimo da povećamo pouzdanost na .85, koliko će nam stavki biti potrebno? Zamenjivanjem vrednosti u izrazu $n = \frac{r_{nn}(1-r_{xx})}{r_{xx}(1-r_{nn})} = \frac{.85(1-.65)}{.65(1-.85)} = \frac{.85*.35}{.65*.15} = 3.05$, dobijamo da test treba produžiti oko 3 puta, tako da ima ukupno 46 stavki. Sa druge strane, test znanja iz psihologije imao je 20 pitanja i pouzdanost od .817. Koliko bi nam stavki trebalo da bismo podigli pouzdanost ovog testa na, npr., .90? S obzirom na to da test treba produžiti $\frac{.9(1-.817)}{.817(1-.9)} = \frac{.9*.183}{.817*.1} = 2.02$ puta, trebalo bi nam 40 stavki. Isto se može izračunati za bilo koji broj stavki, stvarnu i željenu pouzdanost testa.

Slika 3.17 nam ukazuje i na važnost kvaliteta stavki za pouzdanost instrumenta. Ukoliko poredimo različite linije na grafiku, uočavamo da je za isti broj stavki (npr., 10) pouzdanost najniža za test kod koga je prosečna korelacija stavki .1, nešto viša za test sa prosečnom korelacijom .2 i sve veća i veća kako se prosečna korelacija stavki povećava. Ono što smo takođe već videli jeste da ponekad izbacivanje (loših) stavki može povećati pouzdanost instrumenta. Drugim rečima, puko povećavanje broja stavki neće uvek dovesti do (bitnog) povećanja pouzdanosti. Nužno je da stavke koje dodajemo budu kvalitetne i da se odnose na isti predmet merenja kao i postojeće stavke.

Naravno, mogućnost da dodajemo ili izbacujemo stavke iz testa ćemo imati samo ako smo mi konstruktori testa.

3.6.2) Pouzdanost i diskriminativnost

O odnosu diskriminativnosti i pouzdanosti testa već je bilo reči u lekciji o diskriminativnosti, pa ćemo se sada osvrnuti i na već pomenute odnose ove dve metrijske karakteristike, ali i razmotriti neka dodatna pitanja.

Već je rečeno da pouzdanost utiče na diskriminativnost tako što određuje stepen „stvarnog" razlikovanja ispitanika. Koristeći terminologiju klasične teorije testa, možemo reći i da pouzdanost određuje stepen u kom test pravi diskriminaciju, odnosno omogućava razlikovanje između različitih pravih nasuprot dobijenih skorova ispitanika. Naime, dva susedna skora na testu predstavljaju različite dobijene skorove, ali se postavlja pitanje da li su pravi skorovi ispitanika koji su ih postigli takođe različiti.

U odeljku o intervalima poverenja izračunali smo da se pravi skor ispitanika koji je osvojio 10 poena na testu znanja iz psihologije sa 95% sigurnosti nalazi između 6.466 i 13.534 poena. Koristeći formulu za interval poverenja za ispitanika koji je osvojio 12 poena na testu dobjamo vrednosti 8.466 i 15.534, odnosno za ovog ispitanika bismo sa 95% sigurnosti verovali da se njegov pravi skor nalazi u datom opsegu. Da li se onda pravi skorovi ispitanika koji su postigli 10 i 12 poena razlikuju? S obzirom na to da se intervali poverenja preklapaju, postoji mogućnost da su pravi skorovi ova dva ispitanika zapravo isti, što znači da između ova dva skora ne postoji statistički značajna razlika (na nivou .05, pošto smo koristili 95% interval poverenja). Sa druge strane, pravi skor ispitanika koji je na testu znanja osvojio 18 poena nalazi se u rasponu od 14.466 do 21.534, odnosno razlikuje se statistički značajno od skora 10, jer se intervali poverenja ova dva skora ne preklapaju.

Na koji način pouzdanost zavisi od diskriminativnosti? Već je rečeno da je za pouzdano merenje neophodno pre svega da postoji razlikovanje, odnosno diskriminacija skorova. Ako diskriminativnost tretiramo kao varijansu skorova, posmatrajući opšti obrazac računanja pouzdanosti $r_{tt} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$, uviđamo da se pouzdanost povećava kako varijansa skorova na testu (σ_X^2) raste (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Pretpostavka je da varijansa skorova na testu prevashodno zavisi od varijabiliteta pojave, odnosno od varijanse pravog skora, a manje (ili, po Bukviću, uopšte ne) od varijanse greške.

Ponekad se u praksi dešava da pouzdanost računamo na uzorcima sa smanjenim varijabilitetom. Na primer, u situaciji selekcije možemo imati informacije o skorovima samo onih kandidata koji su prošli selekciju, odnosno kandidata sa najvišim postignućem. Jasno je da će se ovakvo sužavanje raspona i smanjenje varijanse skorova negativno odraziti na dobijeni koeficijent pouzdanosti instrumenta, iako se može pretpostaviti da bi njegova pouzdanost na uzorku većeg varijabiliteta skorova bila bolja. Da bismo procenili pouzdanost testa na neselekcionisanom uzorku, možemo primeniti sledeću formulu:

$$r_{x1x1} = 1 - \frac{\sigma_x^2(1 - r_{xx})}{\sigma_{x1}^2}$$

gde su σ_x^2 i r_{xx} varijansa i pouzdanost testa na selekcionisanom uzorku, a σ_{x1}^2 očekivana (ili poznata na osnovu drugog istraživanja) varijansa testa na neselekcionisanom uzorku (Bukvić, 1996). Naravno, formula se može primeniti i u obrnutoj situaciji – kada nas zanima kolika bi bila pouzdanost na selekcionisanom uzorku, ukoliko znamo pouzdanost na neselekcionisanom.

Ako uzemo da su pouzdanost (α = .817) i varijansa (SD^2 = 17.764) koje smo računali za test znanja iz psihologije dobijeni na neselekcionisanom uzorku, odnosno na uzorku svih prijavljenih kandidata, možemo izračunati kolika bi bila pouzdanost instrumenta kada bismo je računali samo na uzorku primljenih kandidata. Recimo da je varijansa skorova ovih kandidata niža i da iznosi SD^2_1 = 12.346. Pouzdanost testa bi onda bila $\alpha_1 = 1 - \frac{SD^2(1-\alpha)}{SD_1^2} = 1 - \frac{17.764(1-.817)}{12.346} = 1 - .263 = .737$. Drugim rečima, kada se primeni na uzorku suženog varijabiliteta, pouzdanost testa je manja. Da smo isti test zadali uzorku s još većom varijansom od uzorka kandidata koji su se pripremali za upis (recimo, zadamo test i kandidatima koji su se spremali za upis matematike), na primer, SD^2_2 = 22.607, pouzdanost bi bila $\alpha_2 = 1 - \frac{17.764(1-.817)}{22.607} = 1 - .144 = .856$, odnosno bila bi veća od pouzdanosti na inicijalnom uzorku (čija je varijansa sužena u poređenju sa drugim uzorkom).

Treba još pomenuti i da gotovo svi navedeni obrasci računanja pouzdanosti uzimaju u obzir korelacije između stavki/delova testa. Kao što nam je poznato, veća korelacija stavki vodi većoj diskriminativnosti, a i većoj pouzdanosti, te je ovo još jedan način na koji su ova dva svojstva testa povezana.

3.6.3) Pouzdanost i valjanost

Psihometrijski pojam valjanosti (ili validnosti) ima više značenja i valjanost instrumenta se može procenjivati na različite načine (obično je potrebno demonstrirati da instrument poseduje nekoliko vidova valjanosti da bi se praktično primenjivao)[20]. Mnogi od načina procene valjanosti podrazumevaju koreliranje instrumenta čija se valjanost procenjuje sa nekim drugim instrumentom, bilo da se radi o instrumentu namenjenom proceni istog ili drugačijih konstrukata (Bukvić, 1996; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Na primer, test anksioznosti koji smo sami konstruisali bismo mogli korelirati sa nekim drugim, već postojećim i „proverenim" testom anksioznosti (dva testa mere isti konstrukt), kako bismo pokazali da meri upravo onu osobinu koju bi i trebalo da meri. Takođe, test znanja iz psihologije bismo mogli korelirati i sa kasnijim uspehom na studijama (korelacija različitih konstrukata) da bismo videli da li i u kojoj meri uspeh na prijemnom ispitu može predvideti uspeh u studiranju. U oba slučaja valjanost se procenjuje preko koeficijenta korelacije, ali treba podvući da valjanost testa ne treba izjednačavati sa korelacijom, jer je to samo jedan od mnogobrojnih načina njene provere.

Koja je najveća korelacija koju jedan instrument može imati sa nekim drugim instrumentom, odnosno koju mogu imati dve varijable? Teorijski, maksimalna korelacija je 1 ili -1 (maksimalna pozitivna, odnosno negativna korelacija). Da li ćemo u praksi, međutim, ikad dobiti ovako visoku vrednost? Nećemo, zato što je, kao što smo već pomenuli, svako merenje nesavršeno, odnosno svako merenje u sebi sadrži neku grešku. Ako se podsetimo definicije slučajne greške – ona je nekorelirana kako sa pravim skorom, tako i sa bilo kojim drugim skorom greške. Drugim rečima, varijansa skorova na testu sačinjena je od dela koji može korelirati sa nečim drugim – varijanse pravog skora i dela koji po definiciji neće korelirati ni sa čim – varijanse slučajne greške. Imajući ovo u vidu, stvarna korelacija dva testa nikad neće moći da dostigne maksimalnu vrednost, čak i kada bi konstrukti bili savršeno korelirani (Bukvić, 1996). Standardno tumačenje visine koeficijenta korelacije je da su korelacije oko .10 niske, oko .30 srednje i oko .50 visoke (J. Cohen, 1992), mada neki noviji pregledi pokazuju da, okvirno, najniža trećina koeficijenata korelacije o kojima se izveštava u psihološkim

[20] Načelno, ukoliko valjanost instrumenta nije pokazana, sve tvrdnje koje se odnose na njegov predmet merenja treba shvatiti uslovno.

istraživanjima ima vrednost ispod .20, srednja trećina između .20 i .30, dok najviša trećina korelacija ima vrednost preko .30 (Hemphill, 2003).

Na koji tačno način pouzdanost testa čiju valjanost procenjujemo utiče na koeficijent korelacije sa drugim testom? Često se kaže da pouzdanost predstavlja gornju granicu valjanosti, odnosno da test ne može korelirati ni sa čim više nego sa samim sobom. Kao što je već prikazano, pouzdanost se može izračunati i kao korelacija skorova istog testa zadatog dva ili više puta (uslovno, „korelacija testa sa samim sobom"). Iz toga bi sledilo da je vrednost koeficijenta pouzdanosti maksimalna korelacija koju test može imati sa nečim drugim, ali ovo nije sasvim tačno. Setimo se da je korelacija testa sa samim sobom zapravo udeo prave u ukupnoj varijansi, odnosno proporcija varijanse skorova koju test deli sam sa sobom. To onda znači da je koeficijent pouzdanosti jednak maksimalnoj proporciji varijanse koju test može deliti sa nekom drugom varijablom (a ne maksimalnoj korelaciji). Kako nam je poznato da se proporcija varijanse koju dve varijable dele dobija kvadriranjem koeficijenta korelacije između njih, sledi da je maksimalna korelacija koju test može imati sa nekom drugom varijablom zapravo koren iz koeficijenta pouzdanosti:

$$r_{max} = \sqrt{r_{tt}}$$

gde je r_{max} maksimalna korelacija testa sa drugom varijablom, a r_{tt} pouzdanost testa. Koren iz koeficijenta pouzdanosti naziva se i indeksom pouzdanosti (Fajgelj, 2009). Ako se ponovo vratimo na primer testa znanja iz psihologije, čija je pouzdanost iznosila .817, možemo izračunati da je maksimalna korelacija koju on može ostvariti sa drugim testom psihologije ili uspehom u studiranju (ili bilo kojom trećom varijablom) jednaka $\sqrt{.817} = .904$, što je veoma visoka vrednost ali ne i maksimalna.

Ovde smo, međutim, u obzir uzeli samo nesavršenost merenja znanja iz psihologije našim testom, ali smo (pošto nam to nije bilo u fokusu) zanemarili preciznost merenja drugog konstrukta. Ako uzmemo i grešku merenja drugog instrumenta u obzir, maksimalni koeficijent korelacije biće još niži. Kao što nas prilikom merenja zanimaju pravi, a ne izmereni skorovi ispitanika, tako nas zapravo zanimaju i korelacije pravih, a ne ukupnih skorova na dva instrumenta. Da bismo dobili korelaciju pravih skorova dva testa, odnosno koeficijent korelacije za zamišljenu idealnu situaciju kada su pouzdanosti oba instrumenta savršene (imaju maksimalnu vrednost 1) možemo primeniti tzv. korekciju za atenuaciju (eng. *correction for attenuation*) (Bukvić, 1996; Fajgelj, 2009). Atenuacija znači smanjenje, a odnosi se na smanjenje koeficijenta korelacije usled nesavršene pouzdanosti instrumenata. Koeficijent atenuacije računa se po sledećoj formuli:

$$r_{t1t2} = \frac{r_{x1x2}}{\sqrt{r_{tt1} r_{tt2}}}$$

gde je r_{t1t2} korelacija pravih skorova dva testa, r_{x1x2} korelacija njihovih dobijenih skorova, a r_{tt1} i r_{tt2} pouzdanosti oba instrumenta.

Pretpostavimo da smo test znanja iz psihologije korelirali sa testom opšte informisanosti čija je pouzdanost .902 i dobili korelaciju .45. Na osnovu ovih informacija možemo izračunati da je korelacija pravih skorova ova dva testa, odnosno korelacija znanja iz psihologije i opšte informisanosti zapravo jednaka $\frac{.45}{\sqrt{.817 * .902}} = \frac{.45}{.858} = .524$. U ovom slučaju, pouzdanosti oba instrumenta su bile dobre, pa se korelacija pravih skorova ne razlikuje drastično od korelacije izmerenih skorova, ali ovo naravno neće biti slučaj za niže pouzdanosti bilo jednog, bilo oba instrumenta.

Korekciju za atenuaciju možemo primeniti i u slučaju kada nam je poznata pouzdanost samo jednog instrumenta. U tom slučaju govorimo o delimičnoj ili jednostranoj korekciji za atenuaciju

(Bukvić, 1996; Fajgelj, 2009), a formula za njeno izračunavanje je identična, samo se kao vrednost pouzdanosti drugog testa uzima vrednost jedan.

3.7) Rezime

Kada govorimo o pouzdanosti, pre svega je važno da napravimo razliku između vremenske i interne konzistencije testa – prva se odnosi na sličnost skorova „kroz vreme" i procenjuje se test-retest metodom ili metodom alternativnih formi, dok se druga odnosi na sličnost skorova „kroz ajteme" testa i procenjuje se metodom deljenja testa ili analizom stavki. Većina obrazaca pouzdanosti zasniva se na primeni klasične teorije testa, konkretno na modelu paralelnih indikatora. Klasična teorija testa pretpostavlja da se varijansa skorova na testu može razložiti na varijansu pravih skorova i varijansu greške. Računanje pouzdanosti počiva na proceni varijanse greške merenja.

Rulonov obrazac pouzdanost procenjuje na osnovu razlika u skorovima na dva dela testa, te se smatra donjom granicom pouzdanosti. Spirman–Braunov obrazac se računa na osnovu korelacija između delova testa, dok se Kronbahova alfa oslanja na odnos varijanse celog testa i njegovih delova, ali se može pokazati da se zapravo i u ovom obrascu koriste korelacije između ajtema. U okviru Gutmanovog modela varijansa skorova se razlaže na imaž koji odgovara pravom skoru i antiimaž koji odgovara skoru greške.

Na osnovu pouzdanosti testa možemo odrediti i standardnu grešku merenja testa koja se koristi prilikom računanja intervala pouzdanosti za pojedinačne skorove na testu. Pouzdanost testa raste sa porastom stavki, ali ne linearno, jer je i kvalitet stavki, izražen preko njihove međusobne korelacije, takođe važan. Načelno, što je diskriminativnost testa (izražena kao varijansa) veća, to je i pouzdanost veća, a što je pouzdanost veća – veća je i njegova valjanost (izražena kao korelacija sa drugim testom).

3.8) Preporučena literatura

Bukvić, A. (1996). *Načela izrade psiholoških testova*. Zavod za udžbenike i nastavna sredstva, Beograd.

Callender, J. C., & Osburn, H. G. (1979). An Empirical Comparison of Coefficient Alpha, Guttman's Lambda-2, and MSPLIT Maximized Split-Half Reliability Estimates. *Journal of Educational Measurement*, *16*(2).

Clark, L. A., & Watson, D. (1995). *Constructing Validity : Basic Issues in Objective Scale Development The Centrality of Psychological Measurement*. *7*(3), 309–319.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment: An Introduction to Tests and Measurement. Ninth Edition*. New York, NY: McGraw-Hill Education.

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Kline, P. (2000). *The Handbook of Psychological Testing, Second Edition*. London and New York: Routledge.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

4) GUTMANOVA TEORIJA

U poglavlju o pouzdanosti rekli smo da Gutmanov model, zasnovan na njegovoj teoriji imaža i antiimaža, predstavlja alternativu klasičnom modelu paralelnih indikatora i pruža drugačiji metod procene prave varijanse i varijanse greške. Pored procene pouzdanosti, Gutmanova teorija koristi se i u formulama za izračunavanje reprezentativnosti testa i stavki, u nekim merama homogenosti, a svoju primenu pronalazi i u faktorskoj analizi (gde se matrica imaž kovarijansi može koristiti kao ulazna matrica za analizu). Imajući u vidu važnost Gutmanove teorije, u ovom poglavlju ćemo detaljnije objasniti osnovne principe na kojima ona počiva i ilustrovati ih na primeru konkretnih podataka.

4.1) Imaž i antiimaž skorovi

Gutman je pravi skor ispitanika na testu imenovao kao imaž, a definisao ga kao onaj deo skora koji se može predvideti na osnovu odgovora na druge stavke na testu. Analogno ovome, grešku pojedinačnog skora Gutman naziva antiimažom, a operacionalno je definiše kao onaj deo skora koji se ne može predvideti na osnovu drugih varijabli u modelu, odnosno kao rezidual (Guttman, 1953).

Ukoliko bismo ovo želeli da predstavimo u matričnom obliku mogli bismo napisati Z = T + E, gde je *Z* matrica standardizovanih skorova ispitanika, *T* matrica pravih skorova, odnosno skorova imaža, a *E* matrica reziduala, odnosno antiimaža (Tabela 4.1.). U skladu sa uobičajenim načinom prikazivanja, ispitanici se nalaze u redovima, a varijable u kolonama tabele. Dakle, skor svakog ispitanika na svakoj varijabli može se razložiti na skor imaža (predviđeni skor na osnovu svih drugih varijabli/stavki) i skor antiimaža (rezidual). Tako se, na primer, standardizovani skor četvrtog ispitanika na trećoj varijabli Z_{43} može razložiti na pravi skor, odnosno skor imaža T_{43} i skor greške, odnosno skor antiimaža E_{43} četvrtog ispitanika na trećoj varijabli. Skor šestog ispitanika na drugoj varijabli Z_{62} dobija se sabiranjem pravog skora, tj. skora imaža tog ispitanika T_{62} i skora greške, odnosno antiimaža E_{62}.

Tabela 4.1. Razlaganje standardizovanih skorova na skorove imaža i skorove antiimaža

	Varijable – standardizovani skor					Varijable – skor imaža					Varijable – skor antiimaža			
Ispitanici	Z_{11}	Z_{12}	Z_{13}	Z_{14}		T_{11}	T_{12}	T_{13}	T_{14}		E_{11}	E_{12}	E_{13}	E_{14}
	Z_{21}	Z_{22}	Z_{23}	Z_{24}		T_{21}	T_{22}	T_{23}	T_{24}		E_{21}	E_{22}	E_{23}	E_{24}
	Z_{31}	Z_{32}	Z_{33}	Z_{34}	=	T_{31}	T_{32}	T_{33}	T_{34}	+	E_{31}	E_{32}	E_{33}	E_{34}
	Z_{41}	Z_{42}	Z_{43}	Z_{44}		T_{41}	T_{42}	T_{43}	T_{44}		E_{41}	E_{42}	E_{43}	E_{44}
	Z_{51}	Z_{52}	Z_{53}	Z_{54}		T_{51}	T_{52}	T_{53}	T_{54}		E_{51}	E_{52}	E_{53}	E_{54}
	Z_{61}	Z_{62}	Z_{63}	Z_{64}		T_{61}	T_{62}	T_{63}	T_{64}		E_{61}	E_{62}	E_{63}	E_{64}

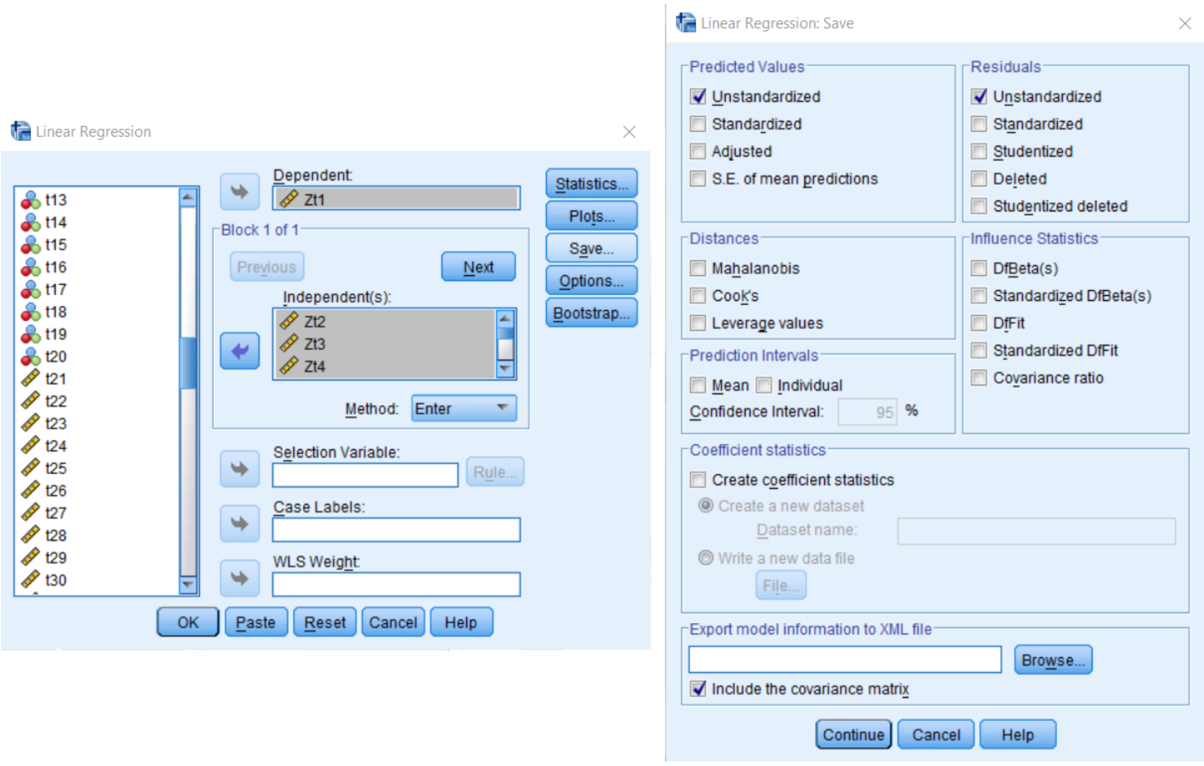
Sada ćemo ilustrovati računanje imaža i antiimaža stavki na primeru 20 pitanja na testu znanja iz psihologije. Za početak, varijable ćemo standardizovati, kako bi svaka varijabla imala jednaku, jediničnu varijansu. Korišćenjem opcije *Save standardized values as variables* u okviru **Descriptives** dobijamo 20 novih varijabli čije su vrednosti Z-skorovi (Slika 4.1 prikazuje samo Z-skorove prvih 20 ispitanika na prvih 10 varijabli). Možemo primetiti da u okviru svake varijable postoje samo dve moguće vrednosti Z-skora, što odgovara binarnoj prirodi izvornih skorova (tačni odgovori su samo 0 i 1). Moguće vrednosti se razlikuju od varijable do varijable, a zavise od težine stavke, odnosno proporcije tačnih i netačnih odgovora (što je više tačnih odgovora, to će pozitivni Z-skor biti bliži nuli, i obrnuto, jer prosek na celom testu mora biti 0).

	Zt1	Zt2	Zt3	Zt4	Zt5	Zt6	Zt7	Zt8	Zt9	Zt10
1	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	.46624	-.91528	-2.07537	-1.55804
2	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
3	.43310	-1.20254	-1.32494	.20120	-1.50764	.88605	-2.13991	-.91528	-2.07537	-1.55804
4	-2.30361	-1.20254	-1.32494	.20120	-1.50764	.88605	-2.13991	-.91528	.48074	-1.55804
5	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	.46624	-.91528	-2.07537	-1.55804
6	.43310	-1.20254	-1.32494	-4.95888	.66177	-1.12602	.46624	-.91528	-2.07537	-1.55804
7	.43310	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
8	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
9	-2.30361	-1.20254	-1.32494	.20120	.66177	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
10	-2.30361	-1.20254	-1.32494	.20120	.66177	-1.12602	-2.13991	-.91528	-2.07537	.64036
11	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	.48074	-1.55804
12	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	.46624	-.91528	.48074	-1.55804
13	.43310	.82966	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
14	.43310	-1.20254	-1.32494	.20120	.66177	-1.12602	-2.13991	-.91528	-2.07537	.64036
15	-2.30361	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	.48074	.64036
16	-2.30361	-1.20254	-1.32494	.20120	.66177	-1.12602	-2.13991	-.91528	.48074	-1.55804
17	.43310	-1.20254	-1.32494	.20120	-1.50764	-1.12602	-2.13991	-.91528	-2.07537	-1.55804
18	.43310	-1.20254	-1.32494	.20120	-1.50764	-1.12602	.46624	1.09006	.48074	-1.55804
19	.43310	-1.20254	-1.32494	.20120	-1.50764	-1.12602	.46624	-.91528	-2.07537	.64036
20	-2.30361	-1.20254	-1.32494	.20120	-1.50764	.88605	.46624	-.91528	-2.07537	-1.55804

Slika 4.1

Sada možemo izračunati skorove imaža i antiimaža kao predviđene vrednosti, odnosno reziduale u multiploj regresiji gde je standardizovani skor na jednoj varijabli kriterijum, a standardizovani skorovi na svim drugim varijablama prediktori. U okviru **Analyze → Regression → Linear**, ubacićemo prvo varijablu Zt1 kao kriterijum, varijable Zt2 do Zt20 kao prediktore, a u okviru opcije *Save* štikliraćemo da želimo da sačuvamo nestandardizovane[21] predviđene vrednosti i nestandardizovane reziduale (Slika 4.2). Proceduru ćemo ponoviti još 19 puta tako da svaka od 20 varijabli bude kriterijumska u po jednoj analizi. Na taj način smo svaku standardizovanu varijablu razložili na imaž i antiimaž.

[21] Biramo da čuvamo nestandardizovane, umesto standardizovanih vrednosti, kako bi se sabirale do izvornog (u našem slučaju Z) skora. Kada bismo sačuvali standardizovane vrednosti, one zbog standardizacije više ne bi bile aditivne.



Slika 4.2

U podacima dobijamo dva seta od po 20 novih varijabli, od toga 20 varijabli se odnosi na skorove imaža, dok se preostalih 20 odnosi na skorove antiimaža. Ukoliko grupišemo sve varijable koje se odnose na predviđene vrednosti (a čiji nazivi počinju prefiksom PRE_ u SPSS-u) dobijamo matricu imaž-skorova (Slika 4.3 prikazuje deo ove matrice). Isto tako, ako grupišemo sve varijable koje se odnose na reziduale (prefiks RES_ u SPSS-u), dobijamo matricu antiimaž skorova (Slika 4.4 prikazuje deo ove matrice).

	PRE_1	PRE_2	PRE_3	PRE_4	PRE_5	PRE_6	PRE_7	PRE_8	PRE_9	PRE_10
1	-1.01097	-1.00852	-1.09149	-.16641	-1.23277	-.94772	-1.22538	-.57823	-.52172	-1.12743
2	-.84817	-.88893	-1.11274	.16998	-.81127	-.47101	-1.08050	-.37243	-.48942	-.69933
3	-1.10434	-.62446	-.76755	-.29769	-.57239	-1.08864	-1.19760	-.54572	-1.00633	-1.11191
4	-1.13875	-1.34785	-1.16859	-.24395	-.71928	-1.01052	-.73749	-1.10019	-.76516	-.97168
5	-.48007	-.53321	-.71963	-.12673	-.79670	-.65986	-.96667	-.67743	-.25700	-.84894
6	-.81391	-1.07460	-1.23899	-.53786	-.29990	-.61223	-1.19329	-.51325	-.42171	-.60761
7	-.95613	-.40074	-.74027	-.24986	-.59238	-.65439	-1.23046	-.32151	-.87552	-1.03073
8	-.50658	-.99713	-1.05580	.16268	-.53891	-.93280	-.67651	-1.01226	-.27710	-.61623
9	-1.12900	-.92510	-1.13614	-.62083	-1.02601	-.75601	-1.25514	-.61968	-.85366	-.86773
10	-.60177	-.75869	-1.02679	-.59386	-.79743	-.28072	-1.32708	-.46624	-.86372	-.86356
11	-.97423	-1.25928	-1.13648	-.28303	-.92438	-.77887	-.79741	-.84181	-.67764	-.95066
12	-.98183	-1.12143	-1.04191	-.18207	-.89617	-.77624	-.81321	-.75067	-.40996	-1.08640
13	-.59788	-.67305	-.92941	.06037	-.38291	-.69273	-.78007	-.60451	-.88509	-1.03465
14	-.74089	-1.06966	-1.17799	-.56662	-.87385	-.99563	-.92650	-.80195	-.59907	-.61999
15	-.73549	-1.30072	-1.06246	-.16470	-.83276	-.64154	-.60596	-.69070	-.53424	-.63887
16	-.49764	-1.17752	-1.08655	-.34778	-.48761	-.93262	-.50224	-1.24915	-.41700	-.74117
17	-1.04805	-.95679	-1.11182	.01515	-.76220	-.98422	-.79151	-.61382	-.42256	-.48361
18	-.91425	-.86559	-.83682	-.06446	-.98889	-.90448	-.72819	-.37787	-.43260	-.98198
19	-.62546	-.80977	-.86328	-.02035	-.91651	-.85035	-.95476	-.41422	-.24234	-.82802
20	-1.01003	-.79756	-.73988	-.16489	-.87995	-1.01844	-.81801	-.71902	-.34442	-.44356

Slika 4.3

	RES_1	RES_2	RES_3	RES_4	RES_5	RES_6	RES_7	RES_8	RES_9	RES_10
1	-1.29264	-.19402	-.23344	.36761	-.27486	-.17830	1.69162	-.33705	-1.55365	-.43061
2	-1.45544	-.31361	-.21220	.03121	-.69637	-.65500	-1.05941	-.54285	-1.58596	-.85871
3	1.53744	-.57808	-.55739	.49889	-.93525	1.97469	-.94231	-.36956	-1.06904	-.44613
4	-1.16487	.14530	-.15635	.44514	-.78836	1.89657	-1.40242	.18491	1.24589	-.58636
5	-1.82354	-.66934	-.60531	.32792	-.71094	-.46616	1.43291	-.23785	-1.81837	-.70910
6	1.24702	-.12794	-.08595	-4.42102	.96167	-.51379	1.65952	-.40203	-1.65366	-.95043
7	1.38924	-.80181	-.58467	.45106	-.91526	-.47162	-.90945	-.59377	-1.19985	-.52731
8	-1.79703	-.20542	-.26913	.03852	-.96873	-.19322	-1.46340	.09698	-1.79827	-.94181
9	-1.17462	-.27744	-.18880	.82202	1.68777	-.37001	-.88477	-.29560	-1.22172	-.69031
10	-1.70185	-.44386	-.29815	.79505	1.45920	-.84530	-.81283	-.44904	-1.21166	1.50392
11	-1.32938	.05674	-.18846	.48423	-.58326	-.34715	-1.34250	-.07347	1.15838	-.60738
12	-1.32178	-.08111	-.28303	.38326	-.61147	-.34978	1.27945	-.16461	.89070	-.47164
13	1.03099	1.50271	-.39553	.14083	-1.12473	-.43329	-1.35984	-.31077	-1.19028	-.52339
14	1.17400	-.13289	-.14694	.76781	1.53562	-.13039	-1.21341	-.11333	-1.47630	1.26035
15	-1.56812	.09818	-.26248	.36590	-.67488	-.48448	-1.53395	-.22458	1.01498	1.27923
16	-1.80597	-.02502	-.23839	.54897	1.14938	-.19340	-1.63767	.33387	.89774	-.81687
17	1.48115	-.24575	-.21312	.18605	-.74544	-.14180	-1.34840	-.30146	-1.65281	-1.07443
18	1.34736	-.33696	-.48812	.26566	-.51875	-.22154	1.19443	1.46792	.91334	-.57606
19	1.05857	-.39277	-.46166	.22155	-.59113	-.27567	1.42099	-.50106	-1.83303	1.46838
20	-1.29358	-.40499	-.58506	.36608	-.62769	1.90449	1.28425	-.19626	-1.73095	-1.11448

Slika 4.4

Dakle, kako pokazuje Tabela 4.1, Z-skorove smo razložili na predviđene vrednosti i reziduale. Ukoliko za bilo kog ispitanika saberemo skor imaža i skor antiimaža za istu varijablu – dobićemo vrednost koja odgovara Z-skoru. Na primer, imaž skor ispitanika 4 na varijabli 3 (Slika 4.3) iznosi I_{43} = -1.16859, dok je njegov antiimaž skor (Slika 4.4) jednak A_{43} = -.15635. Sabiranjem ovih dveju vrednosti dobijamo standardizovani skor ispitanika $Z_{43} = T_{43} + E_{43} = -1.16859 + (-.15635) = -1.32494$. Ovo je vrednost koju i Slika 4.1 odražava. Dakle, Z-skor koji je ispitanik dobio na jednom pitanju na testu znanja iz psihologije smo, uzimajući u obzir njegove odgovore na sve druge stavke u testu, uspeli da razložimo na procenjeni pravi skor (imaž) i procenjeni skor greške (antiimaž). Na isti način, naravno, možemo dobiti i bilo koji drugi Z-skor sabiranjem imaža i antiimaža za datu stavku.

Možemo primetiti da, iako je svaka od standardizovanih varijabli sadržala samo dva moguća skora (s obzirom na binarnu prirodu odgovora na testu znanja), predviđene vrednosti i reziduali sadrže mnogo veći broj različitih skorova. Ovo je sasvim logično, s obzirom na to da se predviđene vrednosti u regresiji računaju na osnovu skorova na prediktorima, a različiti ispitanici su imali različite kombinacije tačnih i netačnih odgovora na pitanjima na testu, što vodi i njihovim različitim predviđenim vrednostima (procenjenim pravim skorovima). Kako se reziduali računaju kao razlika između izvornih i predviđenih skorova, i oni pokazuju isti varijabilitet kao i predviđeni skorovi.

4.2) Varijansa i kovarijansa imaža i antiimaža

U skladu sa definicijama imaža i antiimaža, varijansu imaža možemo definisati kao procenat varijanse određene stavke koji može biti objašnjen na osnovu preostalih varijabli u modelu (instrumentu). Drugim rečima, imaž pojedinačne stavke jednak je multiplom R^2, odnosno koeficijentu multiple determinacije u regresiji gde su prediktori sve ostale stavke. Po Gutmanu, varijansa imaža se uzima za donju granicu pouzdanosti[22] (Guttman, 1945). Varijansa greške pojedinačne stavke se pak u Gutmanovom modelu procenjuje preko varijanse antiimaža i nju čine reziduali, odnosno neobjašnjena varijansa u modelu ($1 - R^2$). Ovo je varijansa koju data stavka ne deli ni sa jednom preostalom stavkom u modelu. S obzirom na to da je varijansa antiimaža definisana

[22] Varijansa imaža je donja granica pouzdane varijanse zato što je zamislivo da varijabla ima deo varijanse nedeljen sa drugim varijablama u modelu, ali koji i dalje ne predstavlja grešku, već prosto specifičnost date varijable. Sa druge strane, varijansa deljena sa drugim stavkama koje se odnose na isti predmet merenja bi morala biti pouzdana varijansa.

kao rezidual, jasno je da se varijanse imaža i antiimaža sabiraju do 1, koliko iznosi varijansa jedne standardizovane varijable.

Dakle, razlaganje koje prikazuje Tabela 4.1 važi ne samo za pojedinačne skorove, već i za varijanse i kovarijanse varijabli. Varijansa standardizovanih skorova može se razložiti na varijansu imaža i varijansu antiimaža, dok se kovarijansa standardizovanih skorova (odnosno korelacija) može razložiti na kovarijansu imaža i kovarijansu antiimaža (Tabela 4.2).

Tabela 4.2. Razlaganje kovarijanse niza varijabli na kovarijansu imaža i kovarijansu antiimaža

	Varijable – standardizovani skor					Varijable – skor imaža					Varijable – skor antiimaža			
Varijable	1	r_{12}	r_{13}	r_{14}		R^2_1	r'_{12}	r'_{13}	r'_{14}		$1 - R^2_1$	a_{12}	a_{13}	a_{14}
	r_{21}	1	r_{23}	r_{24}		r'_{21}	R^2_2	r'_{23}	r'_{24}		a_{21}	$1 - R^2_2$	a_{23}	a_{24}
	r_{31}	r_{32}	1	r_{34}	=	r'_{31}	r'_{32}	R^2_3	r'_{34}	+	a_{31}	a_{32}	$1 - R^2_3$	a_{34}
	r_{41}	r_{42}	r_{43}	1		r'_{41}	r'_{42}	r'_{43}	R^2_4		a_{41}	a_{42}	a_{43}	$1 - R^2_4$

Ovde treba napomenuti da su, kao i matrica korelacija/kovarijansi standardizovanih skorova, i matrica kovarijansi imaža i matrica kovarijansi antiimaža simetrične matrice. Tako se, na primer, vrednost r_{12} odnosi na korelaciju varijable 1 sa varijablom 2, i ova vrednost jednaka je vrednosti r_{21}, odnosno korelaciji varijable 2 sa varijablom 1. Zatim, kovarijansa imaža treće varijable sa imažom četvrte varijable r'_{34} jednaka je kovarijansi imaža četvrte varijable sa imažom treće r'_{43}. Isto tako, kovarijansa antiimaža varijabli 2 i 4 (a_{24}) jednaka je kovarijansi antiimaža varijabli 4 i 2 (a_{42}).

Pogledajmo prvo dijagonalne elemente ove tri matrice – matrice kovarijansi standardizovanih skorova, matrice kovarijansi skorova imaža i matrice kovarijansi skorova antiimaža. Kao što nam je poznato, kovarijansa varijable sa samom sobom jednaka je njenoj varijansi. Ukoliko su varijable standardizovane, ova varijansa jednaka je jednici. Ujedno, korelacija svake varijable same sa sobom je jedan, te su dijagonalni elementi matrice kovarijansi standardizovanih skorova jedinice.

Prateći istu logiku, znamo da je kovarijansa skorova imaža sa samim sobom jednaka varijansi imaža, odnosno multiplom R^2 – onom delu varijanse koji varijabla deli sa drugim varijablama u modelu. Ove vrednosti su zapravo proporcije, jer multiplo R^2 može imati vrednosti između 0 i 1. Isto tako, kovarijansa skorova antiimaža sa samim sobom jednaka je varijansi antiimaža, odnosno $1 – R^2$, što predstavlja deo varijanse jedinstven za određenu varijablu i koji ona ne deli sa drugim varijablama u modelu. I ove vrednosti su decimalni brojevi u rasponu od 0 do 1. Ponovo, vidimo da se varijanse imaža i antiimaža sabiraju do 1.

Kako bismo ovo ilustrovali, iskoristićemo regresionu analizu u kojoj je Z-skor za stavku 3 kriterijumska varijabla, a Z-skorovi svih ostalih varijabli su prediktori. U ispisu analize (Slika 4.5) vidimo da je multiplo $R^2 = .294$, te zaključujemo da je varijansa imaža ove stavke jednaka upravo $R^2_3 = .294$. To znači da na osnovu odgovora na sve druge stavke na testu, možemo predvideti 29.4% varijanse odgovora na treće pitanje u testu znanja iz psihologije, te da je udeo pravog skora u ukupnom, odnosno pouzdanost ovog pitanja jednaka .294. Varijansa antiimaža, odnosno proporcija neobjašnjene varijanse bi samim tim bila $1 - R^2_3 = .706$, što znači da 70.6% varijanse odgovora na trećem pitanju predstavlja slučajnu grešku. Kao što smo već napomenuli, pouzdanost pojedinačnih stavki po pravilu je relativno niska (inače nam ne bi bio potreban veći broj stavki u instrumentu da bismo precizno merili neku osobinu), tako da ove vrednosti možemo smatrati zadovoljavajućim.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.542[a]	.294	.262	.85916691

a. Predictors: (Constant), Zscore(t4), Zscore(t10), Zscore(t13), Zscore(t9), Zscore(t5), Zscore(t14), Zscore(t11), Zscore(t1), Zscore(t20), Zscore(t17), Zscore(t8), Zscore(t6), Zscore(t7), Zscore(t12), Zscore(t16), Zscore(t19), Zscore(t18), Zscore(t2), Zscore(t15)

Slika 4.5

S obzirom na to da smo prethodno već sačuvali skorove imaža i antiimaža za svaku varijablu, razlaganje varijanse dodatno možemo proveriti računanjem varijansi standardizovanog, predviđenog i rezidualnog skora za varijablu 3. Kao što vidimo (Slika 4.6), varijansa standardizovane varijable je 1, varijansa imaža je .294, dok je varijansa antiimaža .706. Potpuno isto važilo bi i za bilo koju drugu varijablu u testu.

Descriptive Statistics

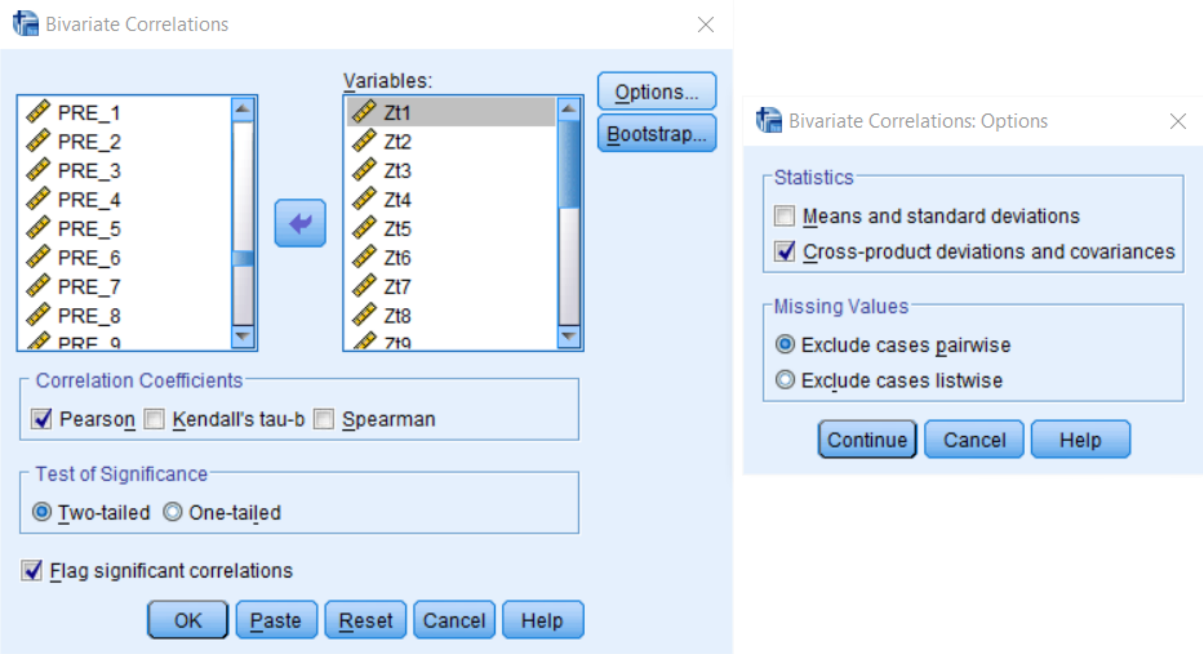
	N	Mean	Variance
Zscore(t3)	436	0E-7	1.000
Unstandardized Predicted Value	436	0E-7	.294
Unstandardized Residual	436	0E-7	.706
Valid N (listwise)	436		

Slika 4.6

Ako se podsetimo poglavlja o pouzdanosti, konkretno dela koji se tiče procene pouzdanosti u Gutmanovom modelu, videćemo da u okviru ispisa analize pouzdanosti, u tabeli *Item-Total Statistics* (Slika 3.15) postoji kolona *Squared Multiple Correlation.* Vrednosti u ovoj koloni predstavljaju kvadriranu multiplu korelaciju, odnosno imaž svakog ajtema. Vrednost kvadrirane multiple korelacije/varijansa imaža za ajtem 3 je, kao što smo i sada izračunali, .294, na osnovu čega lako možemo izračunati i vrednost antiimaža od .706.

Razumevanje vandijagonalnih elemenata matrica koje prikazuje Tabela 4.2 je nešto teže, ali ne i nelogično. Ono što nam je od ranije poznato jeste šta su vandijagonalni elementi matrice kovarijansi standardizovanih skorova, odnosno vandijagonalni elementi matrice korelacija. To su korelacije različitih varijabli. Vrednosti korelacija su decimalni brojevi koji mogu imati vrednosti između -1 (savršena negativna korelacija) i +1 (savršena pozitivna korelacija).

Izračunaćemo sada matricu kovarijansi standardizovanih skorova za 20 pitanja na testu znanja iz psihologije. Kovarijanse varijabli dobijamo tako što u **Analyze → Correlate → Bivariate** ubacimo sve standardizovane skorove u polje *Variables*, a zatim u okviru *Options* štikliramo opciju *Cross-product deviations and covariances* (Slika 4.7).



Slika 4.7

U ispisu, pored standardnih vrednosti korelacije, njene pripadajuće značajnosti i broja ispitanika, dobijamo još dve vrednosti. od kojih nam je sada relevantna vrednost kovarijanse (radi preglednosti Slika 4.8 daje samo jedan isečak matrice korelacija i kovarijansi). Ukoliko gledamo red *Covariance* primećujemo, dakle, da su dijagonalni elementi matrice kovarijansi jednaki jedinicama (što odgovara varijansi standardizovanih varijabli), dok su vandijagonalni elementi jednaki korelacijama između varijabli (budući da su one standardizovane). Možemo primetiti da su kovarijanse pojedinačnih varijabli najčešće niske vrednosti, što, iako se sve stavke odnose na isti predmet merenja, nije iznenađujuće imajući u vidu niske pouzdanosti pojedinačnih stavki, kao i različite sadržaje na koje se odnose pojedinačne stavke.

		Zscore(t1)	Zscore(t2)	Zscore(t3)	Zscore(t4)	Zscore(t5)	Zscore(t6)
Zscore(t1)	Pearson Correlation	1	.241**	.222**	.043	.177**	.109*
	Sig. (2-tailed)		.000	.000	.376	.000	.023
	Sum of Squares and Cross-products	435.000	104.726	96.649	18.494	76.896	47.436
	Covariance	1.000	.241	.222	.043	.177	.109
	N	436	436	436	436	436	436
Zscore(t2)	Pearson Correlation	.241**	1	.306**	.146**	.240**	.297**
	Sig. (2-tailed)	.000		.000	.002	.000	.000
	Sum of Squares and Cross-products	104.726	435.000	133.000	63.543	104.494	129.270
	Covariance	.241	1.000	.306	.146	.240	.297
	N	436	436	436	436	436	436
Zscore(t3)	Pearson Correlation	.222**	.306**	1	.119*	.195**	.302**
	Sig. (2-tailed)	.000	.000		.013	.000	.000
	Sum of Squares and Cross-products	96.649	133.000	435.000	51.891	84.762	131.375
	Covariance	.222	.306	1.000	.119	.195	.302
	N	436	436	436	436	436	436
Zscore(t4)	Pearson Correlation	.043	.146**	.119*	1	-.056	.012
	Sig. (2-tailed)	.376	.002	.013		.241	.798
	Sum of Squares and Cross-products	18.494	63.543	51.891	435.000	-24.468	5.334
	Covariance	.043	.146	.119	1.000	-.056	.012
	N	436	436	436	436	436	436

Slika 4.8

Šta se, međutim, nalazi u vandijagonalnim elementima matrica kovarijansi imaža i kovarijansi antiimaža? Vandijagonalni elementi matrice kovarijansi imaža su kovarijanse imaža različitih varijabli, na primer, kovarijansa imaža varijable 1 sa imažom varijable 2, r'_{12}. Dakle, to je kovarijansa onog dela varijanse varijable 1 koji objašnjavaju ostale varijable u modelu (u našem primeru, varijable 2, 3 i 4) sa onim delom varijanse varijable 2 koji objašnjavaju preostale varijable (u našem primeru, varijable 1, 3 i 4). Može se takođe reći i da vrednost r'_{12} predstavlja onaj deo kovarijanse varijabli 1 i 2 objašnjiv na osnovu varijanse koju te dve varijable dele sa ostalim varijablama. Ukoliko se setimo da, po Gutmanovom modelu, imaž stavke prestavlja pravi skor, vidimo da je kovarijansa imaža zapravo kovarijansa pravih skorova varijabli.

Kao što smo izračunali kovarijansu standardizovanih skorova, možemo izračunati i kovarijansu imaža. Ovog puta ćemo kao varijable u analizu ubaciti 20 predviđenih vrednosti sačuvanih u multiplim regresijama. Slika 4.9 daje isečak pune matrice kovarijansi imaža.

		Unstandardized Predicted Value	Unstandardized Predicted Value	Unstandardized Predicted Value	Unstandardized Predicted Value	Unstandardized Predicted Value	Unstandardized Predicted Value
Unstandardized Predicted Value	Pearson Correlation	1	.843**	.855**	.480**	.825**	.823**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	Sum of Squares and Cross-products	75.406	85.814	84.021	23.869	63.232	73.427
	Covariance	.173	.197	.193	.055	.145	.169
	N	436	436	436	436	436	436
Unstandardized Predicted Value	Pearson Correlation	.843**	1	.962**	.429**	.797**	.872**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	Sum of Squares and Cross-products	85.814	137.420	127.536	28.843	82.448	105.043
	Covariance	.197	.316	.293	.066	.190	.241
	N	436	436	436	436	436	436
Unstandardized Predicted Value	Pearson Correlation	.855**	.962**	1	.491**	.806**	.859**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	Sum of Squares and Cross-products	84.021	127.536	127.922	31.803	80.504	99.867
	Covariance	.193	.293	.294	.073	.185	.230
	N	436	436	436	436	436	436
Unstandardized Predicted Value	Pearson Correlation	.480**	.429**	.491**	1	.413**	.410**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	Sum of Squares and Cross-products	23.869	28.843	31.803	32.826	20.907	24.123
	Covariance	.055	.066	.073	.075	.048	.055
	N	436	436	436	436	436	436

Slika 4.9

Šta sve možemo primetiti posmatrajući matricu kovarijansi imaža (dakle, gledajući redove *Covariance*)? Pre svega, vidimo da dijagonalni elementi nisu jednaki jedinicama, što bi bio slučaj kod korelacija, već su to vrednosti između 0 i 1. Kao što već znamo, ovo su vrednosti multiplog R^2 kada se skor na varijabli predviđa na osnovu svih drugih skorova na testu. Tako, vidimo da se na dijagonali za varijablu 3 nalazi vrednost .294, što smo prethodno demostrirali da je varijansa imaža ove stavke.

Vandijagonalni elementi predstavljaju kovarijanse imaža, odnosno kovarijanse pravih skorova. Ukoliko uporedimo kovarijanse sa korelacijama imaža, primećujemo da kovarijanse imaju niže vrednosti. Ako se podsetimo da je kovarijansa dve varijable jednaka proizvodu njihove korelacije i njihovih varijansi, jasno je da će kovarijansa imaža nužno biti manja od korelacije, s obzirom na to da su varijanse imaža decimalni brojevi manji od 1. Kovarijanse imaža su takođe, po pravilu, niže od kovarijansi standardizovanih skorova (opet, zbog niže varijanse pravih skorova u poređenju sa ukupnom, jediničnom varijansom).

Analogno ovome, u skladu s tabelarnim podacima (Tabela 4.2), vandijagonalni elementi matrice kovarijansi antiimaža su kovarijanse antiimaža različitih varijabli, na primer, kovarijansa antiimaža varijable 1 sa antiimažom varijable 2, a_{12}. To je, dakle, kovarijansa jedinstvene varijanse varijable 1 (neobjašnjene na osnovu varijabli 2, 3 i 4) sa jedinstvenom varijansom varijable 2 (neobjašnjene na osnovu varijabli 1, 3 i 4). Ponovo, kako nam je poznato da varijansa antiimaža zapravo odgovara varijansi greške, vidimo da se na vandijagonalnim elementima matrice antiimaža nalaze kovarijanse grešaka varijabli.

Izračunaćemo sada i kovarijanse antiimaža za 20 pitanja na testu znanja iz psihologije, tako što ćemo kao varijable u analizu uključiti rezidualne vrednosti sačuvane u multiplim regresijama. Slika 4.10 daje isečak matrice kovarijansi antiimaža.

		Unstandardized Residual	Unstandardized Residual	Unstandardized Residual	Unstandardized Residual	Unstandardized Residual	Unstandardized Residual
Unstandardized Residual	Pearson Correlation	1	-.058	-.038	.014	-.038	.076
	Sig. (2-tailed)		.228	.429	.769	.427	.115
	Sum of Squares and Cross-products	359.594	-18.912	-12.628	5.375	-13.664	25.991
	Covariance	.827	-.043	-.029	.012	-.031	.060
	N	436	436	436	436	436	436
Unstandardized Residual	Pearson Correlation	-.058	1	-.018	-.100*	-.068	-.077
	Sig. (2-tailed)	.228		.707	.036	.159	.107
	Sum of Squares and Cross-products	-18.912	297.580	-5.464	-34.700	-22.046	-24.227
	Covariance	-.043	.684	-.013	-.080	-.051	-.056
	N	436	436	436	436	436	436
Unstandardized Residual	Pearson Correlation	-.038	-.018	1	-.057	-.013	-.099*
	Sig. (2-tailed)	.429	.707		.234	.789	.039
	Sum of Squares and Cross-products	-12.628	-5.464	307.078	-20.088	-4.258	-31.508
	Covariance	-.029	-.013	.706	-.046	-.010	-.072
	N	436	436	436	436	436	436
Unstandardized Residual	Pearson Correlation	.014	-.100*	-.057	1	.120*	.052
	Sig. (2-tailed)	.769	.036	.234		.012	.282
	Sum of Squares and Cross-products	5.375	-34.700	-20.088	402.174	45.375	18.789
	Covariance	.012	-.080	-.046	.925	.104	.043
	N	436	436	436	436	436	436

Slika 4.10

Gledajući dijagonalne elemente ove matrice (red *Covariance*), primećujemo da su to varijanse grešaka, odnosno rezidualne varijanse, $1 - R^2$. Kao što smo prethodno demonstrirali, varijansa antiimaža za varijablu 3 iznosi .706, što je upravo i vrednost koju pronalazimo u matrici antiimaž kovarijansi i varijansi.

Kako greške merenja dve varijable ne bi trebalo da koreliraju, vandijagonalni elementi matrice kovarijansi antiimaža su vrednosti veoma bliske nuli. Ovde, međutim, uočavamo jednu od osnovnih razlika između modela paralelnih indikatora i Gutmanovog modela. Dok model paralelnih indikatora zahteva da kovarijansa grešaka dve varijable bude jednaka nuli, Gutmanov model dozvoljava postojanje ovih kovarijansi (odnosno dozvoljava da njihove vrednosti numerički budu različite od nule) (Guttman, 1953). Kao i kod kovarijansi imaža, kovarijanse antiimaža niže su od korelacija antiimaža. Ovo je zbog toga što se kovarijanse dobijaju kao proizvod korelacija i varijansi, a varijanse antiimaža su decimalni brojevi manji od 1.

Numerički, kovarijanse/korelacije antiimaža predstavljaju inverze parcijalnih kovarijansi/korelacija izvornih varijabli[23]. Drugim rečima, ukoliko bismo korelirali varijablu Zt1 sa varijablom Zt2, tako da držimo sve druge varijable konstantnim (što odgovara definiciji parcijalne korelacije), dobili bismo vrednost koja odgovara antiimaž korelaciji, ali bi predznak bio suprotan. Tako, Slika 4.11 prikazuje parcijalnu korelaciju standardizovanih skorova za prvo i drugo pitanje na testu, uz držanje svih drugih pitanja (3–20) konstantnim. Ukoliko uporedimo ovu vrednost sa analognom vrednošću (korelacijom) koju Slika 4.10 odražava, primećujemo da se zaista radi o numerički istoj vrednosti, samo suprotnog predznaka.

Correlations

Control Variables			Zscore(t1)	Zscore(t2)
Zscore(t3) & Zscore(t4) & Zscore(t5) & Zscore(t6) & Zscore(t7) & Zscore(t8) & Zscore(t9) & Zscore(t10) & Zscore(t11) & Zscore(t12) & Zscore(t13) & Zscore(t14) & Zscore(t15) & Zscore(t16) & Zscore(t17) & Zscore(t18) & Zscore(t19) & Zscore(t20)	Zscore(t1)	Correlation	1.000	.058
		Significance (2-tailed)	.	.238
		df	0	416
	Zscore(t2)	Correlation	.058	1.000
		Significance (2-tailed)	.238	.
		df	416	0

Slika 4.11

Ukoliko bismo sada uporedili vrednosti kovarijansi standardizovanih imaž i antiimaž skorova i razmotrili njihov međusobni odnos, možemo videti da se kovarijansa između dve izvorne, standardizovane varijable može razložiti na kovarijansu predviđenih skorova i inverz kovarijanse grešaka. Drugim rečima, kovarijansa se može razložiti na onaj deo potekao od varijanse prisutne u ostatku testa (koja se može predvideti na osnovu drugih varijabli u modelu – prava/imaž varijansa) i na onaj deo koji potiče od varijanse neprisutne u ostatku testa (kada se sve druge varijable drže konstantnim – parcijalna kovarijansa, odnosno inverz kovarijanse grešaka/antiimaža). Tako, kovarijansa standardizovanih varijabli 2 i 3 iznosi r_{23} = .306, kovarijansa pravih skorova ove dve varijable je r'_{23} = .293, dok je kovarijansa njihovih grešaka a_{23} = -.013 (odnosno parcijalna kovarijansa je .013). Ukoliko saberemo kovarijansu pravih skorova sa inverzom kovarijanse grešaka (parcijalnom kovarijansom), dobijamo kovarijansu izvornih standardizovanih varijabli (Guttman, 1953), u našem primeru $r_{23} = r'_{23} - a_{23} = .293 - (-.013) = .293 + .013 = .306$.

4.3) Ekstenzija Gutmanovog modela

Primer koji smo do sada razmatrali (Tabela 4.1 i Tabela 4.2) podrazumevao je da naš test ima samo četiri varijable, što nije veoma verovatno da se desi u praksi, dok smo za ilustraciju u SPSS-u koristili 20 stavki testa znanja iz psihologije, što se takođe može smatrati relativno malim brojem stavki. Razmislimo sada šta se dešava sa vrednostima u matricama kovarijansi standardizovanih skorova, imaža i antiimaža, kada broj varijabli teži beskonačnosti. Ni ovaj slučaj se (na sreću ispitanika) u praksi neće desiti, ali je korisna mentalna vežba za razumevanje postavki Gutmanovog modela.

[23] Suprotan predznak je matematička posledica načina računanja antiimaž kovarijansi (kovarijansi grešaka) i parcijalnih kovarijansi (kovarijansi ukupnih skorova, kada se sve druge varijable drže konstantnim) (Guttman, 1953).

Kada broj varijabli teži beskonačnosti, vrednosti u matrici kovarijansi standardizovanih skorova ostaju iste. Korelacija varijable sa samom sobom uvek će biti 1, dok korelacija jedne stavke sa drugom ne zavisi od ukupnog broja stavki u nekom instrumentu (pojedinačne korelacije su u ovom smislu nezavisne). Vrednosti u matricama imaž i antiimaž kovarijansi će se, međutim, promeniti (Guttman, 1953).

Kao što nam je iz multiple regresije već poznato, dodavanjem novog prediktora u model procenat objašnjene varijanse se nikako ne može smanjiti. On može ostati isti ili se povećati, u zavisnosti od toga koliko novih informacija donosi novi prediktor. Ukoliko novi prediktor deli svoju kompletnu varijansu sa postojećim prediktorima u modelu, ukupan procenat objašnjene varijanse će ostati isti[24]. Ukoliko, međutim, novi prediktor nosi i neke nove informacije – procenat ukupne objašnjene varijanse će se nužno povećati. Da li će ovo povećanje biti značajno ili ne zasebno je pitanje, dok čisto numerički gledano objašnjena varijansa raste sa povećanjem broja prediktora.

Primenjujući ovo znanje na model imaža i antiimaža zaključujemo da će se sa povećanjem broja stavki u modelu i imaž stavke povećati. Naime, veći broj stavki u instrumentu vodiće boljoj predikciji, odnosno većem multiplom R^2. Dakle, dijagonalni elementi matrice kovarijansi imaža će se povećati i težiće pravoj pouzdanosti stavke, odnosno varijansi pravog skora. Suprotno tome, dijagonalni elementi matrice antiimaža, odnosno $1 - R^2$ će se sa povećanjem broja stavki smanjivati, tj. varijansa pojedinačnih stavki koja ne može biti objašnjena modelom će postajati manja i težiti pravoj varijansi greške (Guttman, 1953).

Šta se dešava sa vandijagonalnim elementima matrice kovarijansi imaža i antiimaža kada broj stavki teži beskonačnosti? Rekli smo da se sa povećavanjem broja stavki imaž sve više i više približava pravoj varijansi stavke (pravom skoru). Samim tim, kovarijanse imaža će sve više i više ličiti na kovarijanse, odnosno korelacije pravih skorova, koje bismo mogli nazvati i „pravim" kovarijansama skorova. S obzirom na to da imaž teži pravoj varijansi stavke, kovarijanse imaža teže pravim, izvornim kovarijansama/korelacijama stavki. Ovo ne znači nužno da će se kovarijanse imaža numerički povećavati (one se mogu i smanjivati, ukoliko kovarijansa imaža precenjuje stvarnu korelaciju, mada je ovo ređi slučaj), već samo da će se približavati pravim kovarijansama varijabli, odnosno da će postajati sve „tačnije".

Sa druge strane, vandijagonalni elementi matrice antiimaž kovarijansi težiće nuli, zato što će deo korelacija stavki koji ne može biti objašnjen modelom postajati sve manji i manji kako se broj varijabli u modelu povećava. Ili, posmatrano iz drugog ugla, sa povećavanjem broja varijabli varijansa antiimaža se smanjuje, a što je varijansa nekih varijabli manja, to je i njihova kovarijansa manja. Matrica kod koje su vandijagonalni elementi jednaki nuli naziva se dijagonalnom matricom, pa se može reći i da matrica antiimaž kovarijansi teži da postane dijagonalna matrica kada broj stavki teži beskonačnosti (Fajgelj, 2009). Ova matrica se još naziva i matricom unikviteta, jer su njeni dijagonalni elementi jedinstvena, unikvitetna varijansa svake stavke (Guttman, 1953). Naziv „dijagonalna matrica" odnosi se na matematička svojstva matrice antiimaž kovarijansi (koje numeričke vrednosti se nalaze na dijagonali, a koje van nje), dok se naziv „matrica unikviteta" odnosi na njena psihometrijska svojstva (šta je značenje vrednosti na dijagonali i van nje) i koristi terminologiju koja se najčešće vezuje za faktorsku analizu, analizu koja počiva na razlaganju varijanse niza varijabli na zajedničku i jedinstvenu – unikvitetnu.

[24] Iako teorijski zamisliv, ovakav scenario se u praksi gotovo i ne pojavljuje, osim kada je novi prediktor samo linearni kompozit jednog ili više postojećih prediktora.

Neke od razlika između Gutmanovog modela i modela paralelnih indikatora sada su jasnije. Kao što smo iz matrice antiimaž kovarijansi mogli videti, Gutmanov model dozvoljava korelaciju grešaka različitih varijabli. Takođe, greška jedne varijable može korelirati sa pravim skorom druge varijable. Jedino pravi skor i skor greške, odnosno imaž i antiimaž iste varijable ne mogu biti korelirani. Ove vrednosti su aditivne, odnosno sabiraju se do vrednosti 1, što podrazumeva nultu korelaciju između njih. Ipak, u situaciji kada broj varijabli teži beskonačnosti, a korelacije grešaka teže nuli – Gutmanov model se približava modelu paralelnih indikatora. Zato se može reći i da je model paralelnih indikatora zapravo samo specijalni slučaj Gutmanovog modela, kada broj stavki teži beskonačnosti (Fajgelj, 2009; Guttman, 1953).

Gutmanova teorija ima široku primenu u psihometriji. Već smo prikazali jednu od Gutmanovih formula za računanje pouzdanosti testa koja počiva na razlaganju varijanse na imaž i antiimaž, lambdu 6, a koeficijenti pouzdanosti lambda 2 (λ_2) i lambda 5 (λ_5) takođe se računaju po sličnom principu. Praktično sve mere reprezentativnosti se takođe zasnivaju na Gutmanovoj teoriji, kao i neke od mera homogenosti, što ćemo prikazati u narednim odeljcima. U okviru faktorske analize, posebna metoda – faktorizacija imaža, kao ulaznu matricu za izdvajanje faktora koristi upravo matricu imaž kovarijansi. Imajući sve ovo u vidu, razumevanje osnovnih principa Gutmanove teorije je višestruko korisno za razumevanje psihometrije i njenih brojnih pokazatelja.

4.4) Rezime

Osnovna postavka Gutmanove teorije jeste razlaganje izvornih (tipično standardizovanih) skorova ispitanika na nekom testu na skorove imaža i skorove antiimaža, kao i razlaganje varijansi i kovarijansi skorova na kovarijansu/varijansu imaža i kovarijansu/varijansu antiimaža. Imaž odgovara pouzdanoj varijansi, a operacionalizovan je kao onaj deo varijanse koji se može objasniti na osnovu drugih varijabli u modelu, odnosno kao multiplo R^2. Antiimaž odgovara varijansi greške, a operacionalizovan je kao onaj deo varijanse varijable koji se ne može objasniti na osnovu drugih varijabli u modelu, odnosno kao 1 – multiplo R^2. Kao što se skorovi imaža i antiimaža sabiraju do 1, tako su i njihove varijanse i kovarijanse aditivne (do 1), što istovremeno znači da ne mogu korelirati. U Gutmanovom modelu, međutim, a za razliku od modela paralelnih indikatora, dozvoljene su međusobne korelacije grešaka što ovaj model čini opštijim, a model paralelnih indikatora samo specijalnim slučajem Gutmanovog modela. Primena Gutmanovog modela je široka, od pouzdanosti, preko reprezentativnosti i homogenosti, do faktorske analize.

4.5) Preporučena literatura

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Guttman, L. (1953). Image Theory for the Structure of Quantitative Variates. *Psychometrika*, *18*(4), 277–296.

5) REPREZENTATIVNOST

Zamislimo test opšte informisanosti koji sadrži samo tri pitanja. Da li bismo ovakav test smatrali dobrim testom? Svakako da ne bismo, s obzirom na to da je izvesno da mnogi relevantni aspekti opšte informisanosti takvim testom nisu pokriveni. U ovom slučaju, može se zaključiti da dati test nije reprezentativan za domen koji želi da izmeri – domen opšte informisanosti. Šta bismo, međutim, rekli ukoliko bi test sadržao 30 pitanja, a šta ukoliko bi sadržao 90? Jasno je da odluka o reprezentativnosti ne može biti binarna (jeste / nije reprezentativno), već da se radi o stepenu u kom stavke testa dobro „pokrivaju" željeni domen merenja. Možemo, zato, izračunati numeričke pokazatelje reprezentativnosti radi indikovanja mere reprezentativnosti datog testa za konstrukt koji teži da meri.

5.1) Odnos uzorka i populacije

Kada govorimo o reprezentativnosti[25], zapravo govorimo o stepenu u kom dati uzorak dobro predstavlja populaciju iz koje je izvučen. Ukoliko uzorak dobro predstavlja populaciju kojoj pripada, onda ga smatramo reprezentativnim za tu populaciju, i obrnuto – nereprezentativan je ukoliko loše predstavlja datu populaciju. Ali na koju populaciju se tačno misli? Najčešće je to populacija ispitanika u istraživanju, pa bi, recimo, reprezentativni uzorak studenata u Srbiji trebalo da obuhvati studente različitih fakulteta i različitih univerziteta, oba pola, različitih starosti, nivoa studija (osnovne, master, doktorske) itd. Celokupna populacija studenata u Srbiji nam iz različitih razloga neće biti dostupna, ali je, bez obzira na to, relativno jednostavno možemo definisati – kao sve osobe upisane na bilo koji fakultet u Srbiji u datom trenutku.

Definicija reprezentativnosti analogna je bilo da se radi o uzorku ispitanika ili o uzorku stavki. Kako možemo definisati populaciju stavki u odnosu na koju procenjujemo reprezentativnost stavki u nekom instrumentu? Ovu populaciju, koja se ponekad naziva i univerzumom stavki, čine sve moguće stavke koje se odnose na neki predmet merenja (Momirović et al., 1999; Momirović & Hošek, 1995). Na primer, ukoliko želimo da merimo znanje iz psihologije, univerzum stavki bi činila sva teorijski zamisliva pitanja koja bi se mogla sastaviti, a ispituju znanje iz psihologije. Isto tako, ako merimo anksioznost, zanima nas koliko su stavke našeg instrumenta reprezentativne za sve stavke za merenje anksioznosti koje bi iko ikada mogao smisliti.

Kada bismo imali beskonačan broj stavki koje mere neku dimenziju, na primer, znanje iz psihologije, bili bismo sigurni da je takav test savršeno reprezentativan, odnosno da savršeno opisuje populaciju – celokupno znanje iz psihologije. Međutim, ovakav test nije moguće napraviti ni teorijski, niti u praktičnom smislu. Srećom, kao i u slučaju uzorka ispitanika – uopšte nam nije neophodno da na raspolaganju imamo celokupnu populaciju. Rezultati koje dobijamo biće

[25] Pored termina *reprezentativnost*, koji ćemo koristiti u ovom tekstu, može se naići i na termin *adekvatnost uzorka* (eng. *sampling adequacy*), koji se specifično odnosi na reprezentativnost stavki (a ne i ispitanika).

relevantni i generalizabilni i ako su dobijeni na uzorku, pod uslovom da je uzorak, u ovom slučaju uzorak stavki za merenje određene dimenzije, kvalitetan i reprezentativan.

Reprezentativnost se, naravno, pre svega obezbeđuje dobrom i jasnom definicijom konstrukta, odnosno predmeta merenja, tako da nijedan važan aspekt konstrukta ne bude propušten. U tom smislu, logička i teorijska analiza su veoma važne za obezbeđivanje reprezentativnosti bilo kog instrumenta. Pored ovoga, postoje i numerički pokazatelji repezentativnosti, definisani kao mere reprezentativnosti uzorka stavki nekog instrumenta za univerzum stavki koje se odnose na isti predmet merenja (Knežević & Momirović, 1996; Momirović et al., 1999). Najčešće korišćeni pokazatelji reprezentativnosti su Kajzer–Majer–Olkinova (Kaiser–Meyer–Olkin) mera reprezentativnosti KMO, nazvana tako po autorima koji su je predložili, njena normalizovana varijanta KMO_n i Momirovićeva mera reprezentativnosti PSI-AIK. Svi navedeni pokazatelji počivaju na Gutmanovoj teoriji imaža i antiimaža (Fajgelj, 2009).

5.2) Mere reprezentativnosti

Prema postavkama Gutmanove teorije, kako broj stavki teži beskonačnosti tako imaž varijanse teže varijansama pravih skorova, a imaž kovarijanse „pravim" korelacijama, dok antiimaž kovarijanse teže nuli (Guttman, 1953). Ujedno, kako se broj stavki približava beskonačnosti tako možemo reći da uzorak stavki sve bolje predstavlja populaciju, odnosno ceo univerzum stavki koje mere dati konstrukt (Momirović et al., 1999; Momirović & Hošek, 1995). Zato se mere reprezentativnosti izvode iz odnosa kovarijansi ili korelacija imaža i kovarijansi/korelacija standardizovanih skorova stavki nekog psihološkog mernog instrumenta (Fajgelj, 2009).

Kajzer, Majer i Olkin su predložili meru MSA (eng. *measure of sampling adequacy*), koja se danas po početnim slovima tvoraca naziva KMO i koja se računa prema sledećoj formuli:

$$KMO = 1 - \frac{\sum\sum_{j\neq k} a_{jk}^2}{\sum\sum_{j\neq k} r_{jk}^2}$$

gde su a_{jk}^2 vandijagonalni elementi matrice antiimaž korelacija, dok su r_{jk}^2 vandijagonalni elementi matrice korelacija standardizovanih varijabli, odnosno matrice izvornih korelacija (Kaiser, 1970). U slučaju kada broj varijabli teži beskonačnosti, vandijagonalni elementi matrice antiimaž korelacija teže nuli, te i ceo količnik u izrazu teži nuli, odnosno reprezentativnost teži jedinici, svojoj teorijski maksimalnoj vrednosti. Što su pak korelacije antiimaža (odnosno grešaka u modelu) veće – to će reprezentativnost biti niža.

Iako se u praksi retko dobijaju vrednosti KMO ispod .50, njegov teorijski minimum je negativna beskonačnost, u slučaju kada je suma vandijagonalnih elemenata matrice korelacija jednaka nuli (odnosno kada varijable međusobno ne koreliraju) (Kaiser, 1970). Takođe, kako se nekad pokazalo da je KMO nestabilna mera (pre svega u slučaju niskih reprezentativnosti), Olkin je predložio izmenu KMO mere koja bi je svela na teorijski raspon vrednosti od 0 do 1 (Kaiser & Rice, 1974). Ova mera naziva se normalizovani KMO (KMO_n) i računa se prema formuli:

$$KMO_n = \frac{\sum\sum_{j\neq k} r_{jk}^2}{\sum\sum_{j\neq k} r_{jk}^2 + \sum\sum_{j\neq k} a_{jk}^2}$$

gde su r_{jk}^2 vandijagonalni elementi matrice korelacija standardizovanih varijabli, dok su a_{jk}^2 vandijagonalni elementi matrice antiimaž korelacija. Jasno je da i u slučaju KMO_n kako broj varijabli teži beskonačnosti, a korelacija antiimaža teži nuli – količnik teži maksimalnoj vrednosti 1. Sa druge strane, u slučaju nultih korelacija između varijabli reprezentativnost će biti nulta.

Kajzer je, pored formule za izračunavanje numeričkog indikatora reprezentativnosti, ponudio i okvirne vrednosti za njegovo tumačenje. Imajući u vidu da se empirijski gotovo nikad ne dobijaju vrednosti ispod .40 ili .50, pokazateljima dobre reprezentativnosti se smatraju vrednosti preko .80, a odlične preko .90 (Kaiser, 1970). Ponekad se kao vodič za tumačenje reprezentativnosti navode sledeće slikovite odrednice koje je Kajzer ponudio: vrednosti iznad .90 su divne (eng. *marvelous*), preko .80 pohvalne (eng. *meritorious*), preko .70 prilične (eng. *middling*), preko .60 osrednje (eng. *mediocre*), preko .50 bedne (eng. *miserable*), a ispod .50 neprihvatljive (eng. *unacceptable*) (Dziuban & Shirkey, 1974; Kaiser & Rice, 1974)[26]. Postoje i smernice da su vrednosti ispod .50 neprihvatljive, između .50 i .70 osrednje, između .70 i .80 dobre, između .80 i .90 odlične, a preko .90 izvanredne (Field, 2009).

Momirović je ponudio još jednu meru reprezentativnosti – PSI-AIK, koja predstavlja generalizaciju izvornog KMO. Ova mera reprezentativnosti počiva na istom principu izračunavanja kao i KMO, samo što se umesto vandijagonalnih elemenata matrice antiimaž korelacija koriste vandijagonalni elementi matrice antiimaž kovarijansi. Ona se računa prema formuli:

$$PSI - AIK = 1 - \frac{\sum\sum_{j\neq k} a'^2_{jk}}{\sum\sum_{j\neq k} r^2_{jk}}$$

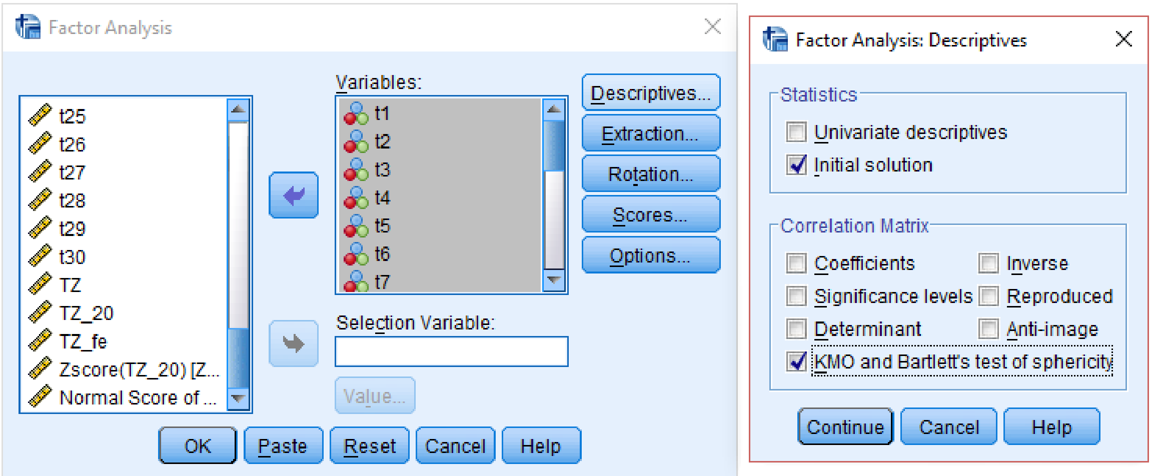
gde su a'^2_{jk} vandijagonalni elementi matrice antiimaž kovarijansi, dok su r^2_{jk} vandijagonalni elementi matrice kovarijansi standardizovanih varijabli, odnosno matrice izvornih korelacija (Knežević & Momirović, 1996; Momirović et al., 1999; Momirović & Hošek, 1995). U kakvom su međusobnom odnosu antiimaž kovarijanse i antiimaž korelacije? Od ranije znamo da je korelacija samo standardizovana kovarijansa, odnosno kovarijansa podeljena standardnim devijacijama dveju varijabli. Drugim rečima, kovarijansa je jednaka korelaciji pomnoženoj standardnim devijacijama. Takođe znamo i da su antiimaž varijanse (pa, samim tim, i njihove standardne deviijacije) decimalni brojevi manji od jedan. Kada ovo povežemo, sledi da će antiimaž kovarijanse biti manje od antiimaž korelacija (jer množimo brojevima manjim od jedan), odnosno da će ceo izraz biti veći (jer se količnik oduzima od jedan). Ovo, dalje, znači da će Momirovićeva mera reprezentativnosti PSI-AIK uvek biti veća od KMO, za isti set varijabli i ispitanika.

Pored mera reprezentativnosti celog testa, odnosno niza stavki, Kajzer i saradnici definisali su i mere reprezentativnosti pojedinačnih stavki. Ove mere računaju se na osnovu (gotovo) istih formula kao i mere reprezentativnosti celog testa, odnosno počivaju na istoj logici. Jedina razlika je što se, umesto sume svih vandijagonalnih elemenata matrice korelacija antiimaža i matrice korelacija izvornih skorova, uzimaju samo one korelacije koje se odnose na pojedinačnu varijablu (Kaiser, 1970; Kaiser & Rice, 1974). Na primer, ako bismo želeli da izračunamo reprezentativnost testa od četiri stavke (Tabela 4.2), bila bi nam potrebna suma antiimaž korelacija a_{12}, a_{13}, a_{14}, a_{23}, a_{24}, a_{34}, kao i suma izvornih korelacija r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}. Za računanje reprezentativnosti stavke 1, međutim, u računicu bismo uključili samo antiimaž korelacije a_{12}, a_{13} i a_{14} i izvorne korelacije r_{12}, r_{13} i r_{14}. U zavisnosti od pokazatelja koji želimo da izračunamo, date vrednosti bismo uključili u odgovarajuću formulu (KMO, KMO_n ili PSI-AIK), uz napomenu da bismo za PSI-AIK koristili antiimaž kovarijanse umesto korelacija.

[26] Ove vrednosti su izvorno definicane za jedan drugi indeks, indeks jednostavnosti faktorskog rešenja (eng. *index of factorial simplicity*) (Kaiser, 1974), međutim, Kaiser ih smatra primerenim i u slučaju mera reprezentativnosti (Kaiser & Rice, 1974).

5.3) Provera reprezentativnosti u SPSS-u

Od pomenuta tri pokazatelja reprezentativnosti, samo jedan je ugrađen u IBM SPSS statistički softver. U pitanju je normalizovani KMO koeficijent (označen kao *Kaiser-Meyer-Olkin Measure of Sampling Adequacy*) i on se u SPSS-u može naći u okviru faktorske analize **Analyze** → **Dimension Reduction** → **Factor**. Zadovoljavajuća vrednost KMO se smatra preduslovom faktorske analize, te je stoga ovaj pokazatelj reprezentativnosti ugrađen upravo u nju. U polje *Variables* treba ubaciti sve stavke za koje želimo da izračunamo reprezentativnost, a zatim u okviru *Desriptives* treba štiklirati odgovarajuću opciju – *KMO and Bartlett's test of sphericity* (Slika 5.1).



Slika 5.1

I prilikom računanja reprezentativnosti ćemo kao primer korististi prvih 20 pitanja na testu znanja iz psihologije, za koje se prethodno pokazalo da imaju zadovoljavajuću diskriminativnost i pouzdanost. Faktorskom analizom se sada nećemo baviti, tako da nećemo uključivati nikakve dodatne opcije u analizi, dok ćemo se u ispisu analize fokusirati isključivo na vrednost KMO, koja se može naći u prvoj tabeli u ispisu, tabeli *KMO and Bartlett's Test* (Slika 5.2).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.887
Bartlett's Test of Sphericity	Approx. Chi-Square	1354.430
	df	190
	Sig.	.000

Slika 5.2

Kao što možemo videti, vrednost reprezentativnosti je preko .80, a približava se .90, što ukazuje na dobru reprezentativnost ovog uzorka stavki. Dakle, uprkos relativno malom broju pitanja (samo 20 od potencijalno ogromnog broja koji bi se mogao postaviti), naš test znanja je uspešno pokrio mereni domen – znanje iz psihologije.

Ukoliko nas dodatno zanima kakva je bila reprezentativnost pojedinačnih stavki, na primer, kako bismo identifikovali najviše, odnosno najmanje reprezentativne stavke, u SPSS-u možemo dobiti i vrednosti normalizovanog KMO za stavke. Potrebno je da u dijalogu *Descriptives* (Slika 5.1) označimo opciju *Anti-image*, što će nam u ispisu dati matricu antiimaž kovarijansi i antiimaž korelacija. Naziv ove tabele u ispisu je *Anti-image matrices* i na dijagonali njene donje polovine, matrice antiimaž korelacija, nalaze se mere reprezentativosti pojedinačnih stavki – MSA (*Measures of Sampling Adequacy*). Radi preglednosti, ovde dajemo samo isečak ove tabele (Slika 5.3).

	t20	-.039	.012	.001	-.036	-.051	.016	-.082	.069	-.070	-.094	.007	-.031	-.001
Anti-image Correlation	t1	.903[a]	-.058	-.038	.014	-.038	.076	-.043	-.013	.006	-.073	-.013	-.028	.060
	t2	-.058	.903[a]	-.018	-.100	-.068	-.077	-.109	-.045	.096	.017	-.054	-.053	-.063
	t3	-.038	-.018	.923[a]	-.057	-.013	-.099	-.112	-.017	-.027	-.033	-.084	-.062	-.025
	t4	.014	-.100	-.057	.651[a]	.120	.052	-.021	.049	-.015	.060	.002	-.125	.056
	t5	-.038	-.068	-.013	.120	.869[a]	-.061	.038	.004	-.037	.008	-.025	-.019	-.028
	t6	.076	-.077	-.099	.052	-.061	.899[a]	-.030	-.019	-.003	-.069	.008	-.068	-.086
	t7	-.043	-.109	-.112	-.021	.038	-.030	.909[a]	-.091	-.128	.006	-.032	-.026	.007
	t8	-.013	-.045	-.017	.049	.004	-.019	-.091	.883[a]	-.010	-.099	-.027	-.125	-.080
	t9	.006	.096	-.027	-.015	-.037	-.003	-.128	-.010	.848[a]	.012	.032	-.104	-.004
	t10	-.073	.017	-.033	.060	.008	-.069	.006	-.099	.012	.895[a]	-.113	-.040	-.021
	t11	-.013	-.054	-.084	.002	-.025	.008	-.032	-.027	.032	-.113	.862[a]	-.128	-.031
	t12	-.028	-.053	-.062	-.125	-.019	-.068	-.026	-.125	-.104	-.040	-.128	.867[a]	-.022
	t13	.060	-.063	-.025	.056	-.028	-.086	.007	-.080	-.004	-.021	-.031	-.022	.868[a]
	t14	.012	.015	-.025	-.026	.009	.028	-.074	-.050	-.058	-.134	.120	-.081	.003
	t15	-.092	-.131	-.062	-.049	-.119	-.184	-.039	-.083	-.066	-.148	-.089	.051	.044
	t16	-.153	-.085	-.076	.002	-.137	-.004	-.079	.078	-.047	-.036	.079	.026	-.034
	t17	-.060	-.069	-.098	.019	.065	-.039	-.008	-.075	-.041	-.063	-.060	-.071	-.065
	t18	-.068	-.161	-.177	-.063	-.138	-.034	-.007	-.141	-.007	.018	-.005	-.050	-.004
	t19	-.047	-.130	-.086	-.020	-.025	-.131	-.020	-.052	-.056	-.086	-.068	.047	-.085
	t20	-.047	.015	.001	-.041	-.061	.020	-.101	.085	-.082	-.118	.009	-.038	-.002

a. Measures of Sampling Adequacy(MSA)

Slika 5.3

Posmatrajući dijagonalne elemente ove matrice primećujemo da sve stavke imaju generalno dobre vrednosti reprezentativnosti, pri čemu je ova vrednost najviša za stavku 3, a najniža (i ne tako dobra) za stavku 4. Ukoliko se podsetimo analize diskriminativnosti i pouzdanosti za ove stavke, videćemo da je stavka 4 ispitanicima bila prelaka, kao i da je pokazivala lošu pouzdanost. Često, mada ne uvek, desiće se da različiti pokazatelji ukazuju na iste stavke kao najbolje, odnosno najlošije. Ipak, možemo zaključiti da stavke u globalu dobro predstavljaju domen znanja iz psihologije, koji bi i trebalo da mere.

5.4) Rezime

Reprezentativnost predstavlja svojstvo testa da stavke koje ga čine u dobroj meri predstavljaju univerzum svih mogućih stavki za merenje datog konstrukta. Načelno, sa povećavanjem broja stavki povećava se i njegova reprezentativnost, a mere reprezentativnosti nam daju brojčani pokazatelj ovog stepena. Najčešće korišćene mere reprezentativnosti su izvorni i normalizovani KMO, odnosno Kajzer–Majer–Olkinova mera reprezentativnosti zasnovani na matrici antiimaž korelacija, kao i Momirovićeva mera PSI-AIK, koja se temelji na matrici antiimaž kovarijansi, te je nužno veća od izvornog KMO. Pored mera reprezentativnosti testa, mogu se računati i mere reprezentativnosti pojedinačnih stavki. Osim što je važna kao preduslov sprovođenja faktorske analize, reprezentativnost nam pruža važne informacije o testu i stavkama sama po sebi.

5.5) Preporučena literatura

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/h0036316

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Kaiser, H. F. (1970). A Second Generation Little Jiffy. *Psychometrika*, *35*(4), 401–415. Retrieved from http://www.springerlink.com/index/417580617713668.pdf

Kaiser, H. F., & Rice, J. (1974). *Little jiffy, Mark IV*. *34*, 111–117.

Knežević, G., & Momirović, K. (1996). RTT9G i RTT10G: Programi za analizu metrijskih karakteristika. In P. Kostić (Ed.), *Merenje u psihologiji* (pp. 37–56). Institut za kriminološka i sociološka istraživanja, Beograd.

Momirović, K., Wolf, B., & Popović, D. A. (1999). *Uvod u teoriju merenja i interne metrijske karakteristike kompozitnih mernih instrumenata*. Univerzitet u Prištini, Fakultet za fizičku kulturu.

6) HOMOGENOST

Pretpostavimo da želimo da napravimo test koji meri anksioznost. Već smo navodili primere stavki ovakvog testa, poput: „Često se osećam zabrinuto” ili „Pesimističan/na sam u pogledu budućnosti”. Samim uvidom u sadržaj stavki, jasno je da se one odnose na različita konkretna ponašanja/osećanja. Moguće je, zato, da će se neki ispitanik složiti sa prvom tvrdnjom, ali ne i sa drugom (često je zabrinut, ali nije i pesimističan) ili da se će se složiti sa drugom, ali ne i prvom tvrdnjom (ne oseća se zabrinuto, ali je pesimista). Ipak, mi obe stavke smatramo pokazateljima anksioznosti, jer verujemo da dele zajednički izvor, osnovu, a to je stepen anksioznosti osobe koja popunjava test. Imajući to u vidu, očekujemo da će se većina ispitanika koji se slože sa prvom izjavom složiti i sa drugom (i obrnuto). Naravno, uvek govorimo o stepenu u kom se stavke odnose na isti predmet merenja, odnosno stepen u kom se njihovi predmeti merenja razlikuju. Svojstvo testa da se njegove stavke odnose na isti predmet merenja naziva se homogenošću ili jednodimenzionalnošću (Fajgelj, 2009).

Pored ovakvog načina definisanja homogenosti, može se naići i na tvrdnju da je homogenost stepen u kom se stavke kompozitnog mernog instrumenta odnose na isti sadržaj ili na istu kombinaciju različitih sadržaja (Knežević & Momirović, 1996; Momirović et al., 1999). Šta pod tim podrazumevamo? Ovde se misli na to da je test homogen ako sve stavke mere konstrukt na identičan način. U jednom pojednostavljenom primeru, test mentalne rotacije bi trebalo da obuhvata i sposobnost kreiranja mentalne reprezentacije trodimenzionalnog objekta i sposobnost njene rotacije. Da bismo ovaj test mogli smatrati homogenim, odnosno jednodimenzionalnim, važno je da svaka stavka zahvata oba aspekta, kao i da to čini u podjednakoj meri, odnosno da udeo kreiranja i rotiranja mentalne reprezentacije bude jednak za sve stavke.

Nasuprot tome, ukoliko konstrukt zahvata različite aspekte ponašanja (koja se odnose na dati konstrukt), odnosno obuhvata različite indikatore iste osobine, ali neke stavke pretežno mere jedan, a neke druge stavke pretežno drugi indikator – onda ne možemo govoriti o potpunoj jednodimenzionalnosti predmeta merenja datog testa. U takvom slučaju bismo rekli da je test multidimenzionalan.

6.1) Mere homogenosti

Kako bismo mogli numerički proceniti stepen u kom se stavke nekog testa odnose na isti predmet merenja? Istorijski prva ideja počivala je na korelacijama stavki, dok se savremene konceptualizacije homogenosti oslanjaju na modele zajedničke varijanse svih stavki u instrumentu (Fajgelj, 2009).

Intuitivno je logično da je korelacija između stavki dobar pokazatelj u kojoj meri se one odnose na isti predmet merenja. Što je korelacija stavki veća, to je predmet njihovog merenja sličniji, odnosno to je test homogeniji. Prateći ovu ideju, meru homogenosti koja se obično označava kao H_1 (ili h_1 ili MIC od eng. *mean inter-item correlation*) možemo dobiti prema sledećoj formuli:

$$\overline{r_{xx}} = \frac{\sum\sum r_{ij} - n}{n(n-1)}$$

gde su r_{ij} korelacije pojedinačnih stavki, a n broj stavki nekog instrumenta (Clark & Watson, 1995; Knežević & Momirović, 1996). Dakle, mera homogenosti H1 nije ništa drugo do prosečna korelacija stavki. I u slučaju kada je tumačimo kao meru homogenosti, odnosno jednodimenzionalnosti testa, važi preporuka da prosečna korelacija treba da bude oko .15-.20 za šire definisane konstrukte, odnosno oko .40-.50 za uže definisane konstrukte (Clark & Watson, 1995).

Iako naizgled jednostavna i elegantna, ova mera homogenosti ima nekoliko problema (Fajgelj, 2009). Pre svega, ona ne uzima u obzir to da izmereni skorovi obuhvataju kako pravi skor, tako i skor greške. To, sa jedne strane, znači da će korelacije stavki zavisiti od njihove pouzdanosti (setimo se odnosa pouzdanosti i valjanosti), te se može desiti da prosečna korelacija stavki bude niska ne zato što se one ne odnose na isti predmet merenja, već zato što ga neprecizno mere. Sa druge strane, može se desiti i da neki par stavki veštački visoko korelira zato što su im greške korelirane. Takođe, iako je teorijski maksimalna prosečna korelacija stavki jednaka 1, kada bi ona bila empirijski dobijena za neki test, to ne bi ukazivalo na njegove dobre metrijske karakteristike, naprotiv! Prosečna korelacija stavki od 1 podrazumeva maksimalne korelacije između svaka dva ajtema, što dalje znači da je ispitanik koji je (ne)tačno rešio jedan ajtem ujedno (ne)tačno rešio i sve ostale, čime se obesmišljava test kao takav, jer bi ga mogao zameniti samo jedan ajtem.

Alternativni pristup proceni homogenosti počiva na faktorskoj analizi ili analizi glavnih komponenti. Iako to nije predmet ove knjige, na ovom mestu ćemo, u najkraćim crtama i na pojednostavljen način, objasniti one principe analize glavnih komponenti i faktorske analize neophodne za razumevanje mera homogenosti.

Ukoliko pogledamo korelacije između stavki na nekom (bilo kom) testu, po pravilu ćemo uočiti da su ove korelacije pozitivne i niskog do umerenog intenziteta. Kao što smo već rekli, uzrok ovim pozitivnim korelacijama jeste to što stavke imaju isti predmet merenja. Sa druge strane, razlog njihovom „niskom” intenzitetu jeste postojanje greške merenja, koja nas sprečava da bilo kojom pojedinačnom stavkom potpuno precizno izmerimo dati konstrukt. Dakle, bilo koje dve stavke u instrumentu dele jedan deo svoje varijanse, potekao od predmeta merenja, dok svaka ima i neku svoju jedinstvenu varijansu, koju ne deli sa preostalim stavkama.

Analiza glavnih komponenti kao statistički postupak omogućava nam da izdvojimo komponentu koja će u sebi objediniti što više varijanse u osnovi svih stavki nekog instrumenta (Abdi & Williams, 2010; Cudeck & MacCallum, 2007; Tenjović, 2000). Rezultat analize glavnih komponenti, međutim, nije samo jedna, već veći broj glavnih komponenti[27], pri čemu prva glavna komponenta uvek ima najveću varijansu, odnosno „pokriva“ najveći deo ukupne varijanse stavki. Varijansa glavne komponente naziva se još i svojstvenom vrednošću ili karakterističnim korenom te glavne komponente, a označava se grčkim slovom lambda (λ). Što su korelacije između ajtema veće, to će varijansa prve glavne komponente biti veća.

Faktorska analiza pak počiva na razlaganju ukupne varijanse na testu na varijansu zajedničku svim ajtemima, za koju pretpostavljamo da se odnosi na glavni predmet merenja i jedinstvenu varijansu specifičnu za svaki od ajtema. Faktori se u faktorskoj analizi izdvajaju na isti način kao kod analize glavnih komponenti, samo što se izdvajaju na osnovu zajedničke varijanse (dakle, samo na onom delu varijanse koji stavke međusobno dele), za razliku od analize glavnih

[27] Broj glavnih komponenti je zapravo uvek jednak izvornom broju varijabli, odnosno broju stavki testa (n). Ovim brojem glavnih komponenti kompletna varijansa testa je iscrpljena, pa nije moguće izdvojiti više glavnih komponenti.

komponenti gde se izdvajaju glavne komponente na ukupnoj varijansi stavki (uključujući i zajedničku i jedinstvenu). I za faktorsku analizu važi da će prvi faktor uvek imati najveću svojstvenu vrednost, odnosno objašnjavaće najveću proporciju zajedničke varijanse datog skupa ajtema.

S obzirom na to kako se računaju, jasno je da će se prva glavna komponenta ili prvi faktor odnositi na ono što je stavkama u instrumentu zajedničko, odnosno na onu varijansu koju sve stavke dele. Što je udeo varijanse prve glavne komponente/prvog faktora u ukupnoj varijansi veći – to se varijacije u izvornim ajtemima bolje mogu objasniti jednim zajedničkim izvorom variranja, što implicira homogenost, odnosno jednodimenzionalnost.

Momirović je, sa saradnicima, predložio nekoliko mera homogenosti koje počivaju na udelu varijanse prve glavne komponente u ukupnoj varijansi datog testa (Knežević & Momirović, 1996; Momirović et al., 1999). Drugim rečima, Momirovićeve mere homogenosti koncipirane su tako da ukazuju na proporciju varijanse testa koja odlazi na glavni predmet merenja. Tako se, recimo, mera homogenosti H2 računa prema sledećoj formuli:

$$H2 = \frac{\lambda_{max}}{\sum \lambda_i}$$

gde je λ_{max} svojstvena vrednost prve glavne komponente (koja je, kao što smo pomenuli, uvek najveća svojstvena vrednost), dok su λ_i svojstvene vrednosti svih glavnih komponenti koje se mogu identifikovati. Što je udeo varijanse prve glavne komponente u ukupnoj varijansi veći, to je test homogeniji, jer stavke dele veći procenat zajedničke varijanse. Kao i pokazatelji reprezentativnosti, i mera H2 počiva na Gutmanovoj teoriji imaža, zato što se glavne komponente identifikuju na osnovu matrice imaž kovarijansi stavki. Preostale mere homogenosti se dobijaju na osnovu obične matrice korelacija.

Za razliku od mere H2, prilikom računanja mere H5 ne uzimaju se u obzir svojstvene vrednosti svih glavnih komponenti već samo onih čija je svojstvena vrednost veća od jedan, odnosno čija je pouzdanost veća od 0[28]:

$$H5 = 1 - \frac{\sum \lambda_p - \lambda_{max}}{n - \lambda_{max}}$$

gde su λ_p svojstvene vrednosti pouzdanih glavnih komponenti, a *n* broj stavki instrumenta, odnosno maksimalan mogući broj glavnih komponenti. Dakle, sada varijansu prve glavne komponente ne poredimo sa ukupnom, već samo sa pouzdanom varijansom. Što je udeo prve glavne komponente u pouzdanoj varijansi veći – to će mera homogenosti H5 biti veća. U slučaju kada postoji samo jedna glavna komponenta čija je svojstvena vrednost veća od jedan (što mora biti prva glavna komponenta, jer je njena varijansa uvek najveća), izraz iznad razlomačke crte postaje 0, a homogenost dostiže teorijski maksimalnu vrednost od 1. Ukoliko su pak sve komponente pouzdane, onda će suma njihovih varijansi biti jednaka n, čime vrednost razlomka postaje 1, a homogenost 0.

Mera homogenosti H7 počiva na sličnom principu kao i H5, zato što se homogenost procenjuje preko odnosa varijanse prve glavne komponente sa varijansom svih pouzdanih komponenti:

$$H7 = \frac{\lambda_{max} - 1}{\sum \lambda_p - 1}$$

Ponovo, ukoliko samo jedna glavna komponenta ima svojstvenu vrednost veću od 1, to mora biti prva glavna komponenta i onda će vrednost iznad razlomačke crte biti jednaka vrednosti ispod

[28] Smatra se da komponenta čija je pouzdanost manja od nule meri samo varijansu greške, te samim tim nije relevantna za konstrukt koji želimo da merimo.

razlomačke crte, čime će vrednost homogenosti dostići maksimalnih 1. Sa povećanjem broja glavnih komponenti čije su svojstvene vrednosti veće od 1 homogenost se smanjuje.

Pored navedenih, Momirović je predložio još nekoliko mera homogenosti. One su sve zasnovane na istoj osnovnoj ideji, a to je da se proceni koji deo varijanse testa potiče od merenog konstrukta. Što je ova proporcija veća, to je test homogeniji, odnosno njegove stavke se u većoj meri odnose na zajednički predmet merenja.

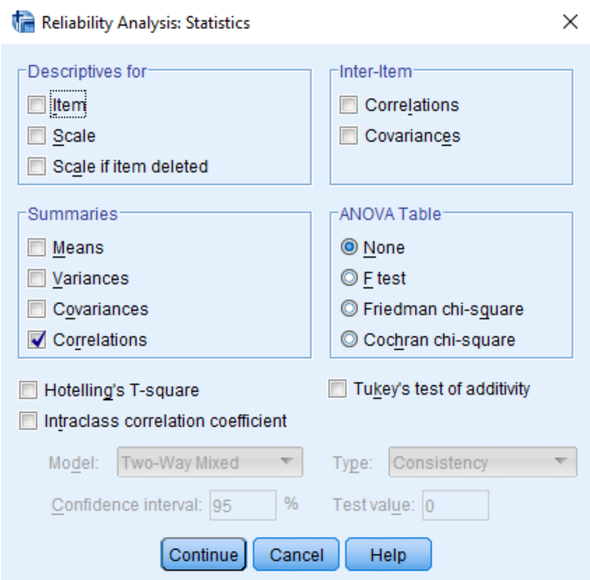
Bitno je naglasiti da u slučaju mera H2, H5 i H7 (i svih drugih mera koje počivaju na analizi glavnih komponenti) vrednost homogenosti od 1 ne predstavlja paradoks u kom jedna stavka može zameniti ceo test. Takođe, ovo nije teorijski idealna vrednost koja je u praksi nedostižna, već se lako može desiti da homogenost nekog instrumenta zaista bude 1. Ovo samo znači da je prva glavna komponenta bila jedina pouzdana komponenta, odnosno da je objašnjavala svu zajedničku varijansu stavki.

Ne postoje jasne preporuke o poželjnim vrednostima homogenosti. Svakako, za usko definisan konstrukt homogenost može biti visoka (npr., oko .70 ili viša), pa čak i maksimalna, dok je za šire definisan konstrukt teško očekivati jednako visoke vrednosti, pa bi se i umerene vrednosti homogenosti (od, na primer, .40 ili .50) mogle smatrati adekvatnim. Tako, na primer, ako imamo test koji meri anksioznost – možemo očekivati visoke vrednosti homogenosti, ali ako računamo homogenost za test neuroticizma (koji obuhvata više faceta, od kojih je anksioznost samo jedan), očekivali bismo srazmerno nižu homogenost. Takođe, na homogenost posredno utiče i broj stavki instrumenta. U slučaju manjeg broja stavki, veća je verovatnoća da će njihov predmet merenja biti sličan, te da će se dobiti visoke vrednosti homogenosti, dok se kod instrumenata sa većim brojem stavki lakše može desiti da obuhvate konstrukt na širi način, vodeći nižoj homogenosti.

Homogenost pojedinačne stavke može se izračunati prema formuli koja se oslanja na H2 meru homogenosti. Homogenost stavke dobija se kao kovarijansa imaža date stavke i prve glavne komponente (kreirane na osnovu matrice imaž kovarijansi) (Knežević & Momirović, 1996; Momirović et al., 1999). Što imaž stavke deli veću proporciju varijanse sa zajedničkim predmetom merenja, to je homogenost date stavke bolja.

6.2) Provera homogenosti u SPSS-u

Preko korisničkog interfejsa u SPSS-u možemo izračunati samo jednu od pomenutih mera homogenosti – meru H1, odnosno prosečnu interkorelaciju stavki. Sa ovom merom smo se, zapravo, već susreli kada smo govorili o pouzdanosti testa, jer se ona i dobija u okviru analize pouzdanosti. Imajući u vidu da, pored njegove homogenosti, prosečna korelacija stavki odražava i grešku merenja, odnosno pouzdanost instrumenta, ovo nije iznenađujuće. Da se podsetimo, do prosečne korelacije stavki instrumenta možemo doći preko opcije **Analyze → Scale → Reliability Analysis**, označavanjem opcije *Summaries: Correlations* u okviru *Statistics* (Slika 6.1). Kao



Slika 6.1

varjable ćemo u analizu ubaciti prvih 20 stavki na testu znanja iz psihologije.

U ispisu nas sada zanima samo tabela *Summary Item Statistics*, koja, budući da ništa drugo nismo označili, ima samo jedan red *Inter-Item Correlations* (Slika 6.2).

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.178	-.056	.375	.432	-6.672	.007	20

Slika 6.2

Kao što smo i prilikom analize pouzdanosti konstatovali, i sada primećujemo da, u proseku, stavke našeg testa koreliraju relativno nisko jedna sa drugom. Kakav zaključak o homogenosti na osnovu toga možemo doneti? Imajući u vidu da se radi o testu koji bi trebalo da zahvati širok raspon znanja iz psihologije, niske vrednosti su zapravo i očekivane, te bi se moglo zaključiti da je homogenost instrumenta adekvatna. Dalje, s obzirom na to da su korelacije većinom pozitivne (minimalna vrednost je jako blizu nuli, dok je maksimalna u umerenom opsegu) možemo zaključiti i da postoji neki zajednički predmet merenja na koji se odnose sve, ili barem većina stavki. No, koju proporciju ovaj zajednički predmet merenja objašnjava, na osnovu mere H1 ne možemo utvrditi. O preostalim pomenutim merama homogenosti biće više reči u narednom poglavlju.

6.3) Rezime

Homogenost predstavlja svojstvo testa da se njegove stavke odnose na isti, zajednički predmet merenja, a mere homogenosti ukazuju na stepen u kom stavke testa procenjuju isti konstrukt. Starije konceptualizacije homogenosti su je izjednačavale sa prosečnom korelacijom stavki (mera H1), dok se savremene operacionalizacije oslanjaju na metode faktorske analize i analize glavnih komponenti i na različite načine procenjuju relativni udeo varijanse glavnog zajedničkog predmeta merenja u ukupnoj (mera H2) ili u pouzdanoj varijansi instrumenta (mere H5 i H7). Homogenost za usko definisane konstrukte treba da bude visoka (čak može biti i maksimalna), dok je za šire definisane konstrukte prihvatljivo i ako je umerena.

6.4) Preporučena literatura

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Knežević, G., & Momirović, K. (1996). RTT9G i RTT10G: Programi za analizu metrijskih karakteristika. In P. Kostić (Ed.), *Merenje u psihologiji* (pp. 37–56). Institut za kriminološka i sociološka istraživanja, Beograd.

Momirović, K., Wolf, B., & Popović, D. A. (1999). *Uvod u teoriju merenja i interne metrijske karakteristike kompozitnih mernih instrumenata*. Univerzitet u Prištini, Fakultet za fizičku kulturu.

7) PROVERA METRIJSKIH KARAKTERISTIKA INSTRUMENTA MAKROOM RTT10G

Kao što se iz dosadašnjeg teksta moglo videti, različiti pokazatelji metrijskih karakteristika instrumenta se u SPSS-u nalaze u okviru različitih analiza. Ako želimo da dobijemo informacije o pouzdanosti, reprezentativnosti i homogenosti instrumenta, potrebno je da pokrenemo nekoliko različitih analiza i uključimo veći broj dodatnih opcija koje nisu uvek podrazumevane za date analize (npr., mera reprezentativnosti KMO, mera homogenosti H1). Određeni broj pokazatelja i nije dostupan u okviru redovnih SPSS analiza, kao što su izvorni, nenormalizovani KMO ili PSI-AIK koeficijent reprezentativnosti ili mere homogenosti H2, H5 i H7. Slična situacija je i sa pokazateljima kvaliteta pojedinačnih stavki – neki se mogu dobiti u SPSS-u, uz manje ili više truda (npr., mere pouzdanosti stavki mogu podrazumevati sprovođenje većeg broja multiplih regresija), dok neki nisu dostupni.

U cilju obezbeđivanja obuhvatne analize metrijskih karakteristika psiholoških mernih instrumenata, Knežević i Momirović su napisali makro RTT10G, koji računa veći broj ovih pokazatelja u okviru iste analize (Knežević & Momirović, 1996). Makro je napisan u programskom jeziku Matrix, koji koristi SPSS sintaksa, što ga čini lako primenljivim za sve korisnike SPSS-a.

Pre svega ćemo objasniti kako se ovaj makro pokreće, odnosno aktivira, a zatim ćemo proći kroz delove ispisa koji se odnose na reprezentativnost instrumenta, njegovu pouzdanost, zatim homogenost, i na kraju na metrijske karakteristike svake stavke ponaosob. Da bismo rezultate mogli da poredimo sa onima dobijenim u SPSS-u, ponovo ćemo kao primer koristiti podatke sa testa znanja iz psihologije (preciznije, odgovore na prvih 20 pitanja u testu).

7.1) Pokretanje makroa RTT10G

Makro RTT10G se aktivira pokretanjem veoma jednostavne sintakse:

```
INCLUDE 'D:\Danka\fax\rtt10g.sps'.
RTT10G VARS = t1 to t20.
```

Dakle, za sprovođenje analize dovoljno je napisati samo dve linije komandi, koje ćemo sada detaljnije analizirati. U prvoj liniji komandi mi zapravo pozivamo makro, koji mora biti sačuvan na kompjuteru kao zaseban fajl. Prva ključna reč – INCLUDE govori SPSS-u da treba da „pozove" nešto, dok drugi deo komande ukazuje na to šta treba pozvati. Da bi se makro pravilno pozvao, potrebno je navesti celokupnu putanju do makroa na kompjuteru. U našem primeru, ja sam makro sačuvala sa svom kompjuteru, na disku *D*, u okviru foldera *Danka* i podfoldera *fax*. Stoga, putanja do fajla na mom kompjuteru je *D:\Danka\fax\rtt10g.sps*. Neophodno je, pored lokacije samog fajla napisati i njegov naziv i ekstenziju, odnosno dodati i *rtt10g.sps* jer će u suprotnom SPSS imati samo putanju do foldera, ali neće „znati" koji fajl u datom folderu da pozove. Fajl koji pozivamo je sam makro, koji je sintaksni fajl sa ekstenzijom .sps, a ne fajl sa podacima na kojima se analiza vrši (ovaj fajl mora biti već otvoren u SPSS-u da bismo ga analizirali, te nema potrebe da ga dodatno pozivamo). Primećujemo da se putanja do fajla nalazi u jednostrukim navodnicima, kao i da se na kraju reda nalazi tačka. Kao što smo ranije pominjali, SPSS nije osetljiv na mala i velika slova, ali je veoma osetljiv na znakove interpunkcije. Zato je neophodno upotrebiti jednostruke, a ne dvostruke znake navoda, kao i završiti komandu tačkom.

Ukoliko prvi red komande nije dobro napisan, SPSS će prijaviti grešku (Slika 7.1) kojom nas obaveštava ili da nije pronašao željeni fajl ili da ne prepoznaje prvu reč druge linije komandi kao SPSS komandu (što zapravo takođe znači da nije pronađen fajl sa makroom jer data ključna reč postoji samo u okviru makroa). U primeru na slici možda primećujete da ime makroa nije bilo dobro napisano, umesto rtt10g piše samo rtt10 (fali slovo *g*), što je dovelo do greške. Neuspešno pokretanje makroa je najčešće posledica grešaka upravo u prvoj liniji komandi, tako da je dobro uvek pažljivo proveriti da li je putanja do fajla dobro napisana, da li su upotrebljeni ispravni navodnici i da li se na kraju cele linije komandi nalazi tačka.

```
>Error # 31 in column 9.  Text: D:\Danka\fax\psmel\rtt10.sps
>File not found.
>Execution of this command stops.

>Error # 1.  Command name: rtt10g
>The first word in the line is not recognized as an SPSS Statistics comm
and.
>Execution of this command stops.
```

Slika 7.1

Druga linija komandi takođe ima dva dela. Prvi deo je ključna reč RTT10G (može biti napisana i malim slovima), koja zapravo aktivira makro jednom kada je on pozvan (prvom linijom komandi). Kao što smo rekli, ovo je ključna reč definisana samim makroom, tako da neće funkcionisati ukoliko on nije aktiviran. Nakon ove ključne reči treba definisati varijable na kojima će se analiza vršiti. To činimo upisivanjem imena varijabli nakon VARS (skraćeno od *variables*) = . Varijable možemo pobrojati sve ili iskoristiti komandu *to* i navesti samo prvu i poslednju varijablu (ukoliko su adekvatno poređane u fajlu sa podacima) koju želimo da uključimo u analizu. I ovu liniju komandi neophodno je završiti tačkom. Važno je naglasiti i da se moraju uneti imena, a ne nazivi (*label*) varijabli da bi sintaksa radila. Ukoliko se pogrešno unesu imena varijabli, SPSS će prijaviti grešku „*The variable name XX not found in the dictionary*“.

Prvi deo ispisa iz analize je kompletna matrica interkorelacija varijabli (koju zbog veličine nećemo prikazivati). U našem primeru u pitanju je matrica dimenzija 20x20, pošto smo u analizu uključili 20 stavki sa testa znanja iz psihologije. Iz prethodnih poglavlja, kao i ovog, mogli smo videti da računanje svih mera pouzdanosti, reprezentativnosti i homogenosti počiva na interkorelacijama stavki, te nije iznenađujuće što je upravo matrica korelacija, kao osnov izračunavanja svih prikazanih mera, prva tabela koju ćemo videti u ispisu.

Iako je matricu korelacija često teško sagledati zbog njene veličine, važno je obratiti pažnju na dve stvari koje mogu sprečiti da se izračunaju željeni pokazatelji. Jedno je postojanje varijabli sa nultom varijansom, a drugo postojanje savršene korelacije između dve varijable. Kakva je to varijabla koja ima nultu varijansu? U pitanju je varijabla na koju su svi ispitanici isto odgovorili. Najčešće se radi o ekstremno lakom ili teškom ajtemu koji su svi ispitanici tačno, odnosno netačno rešili (npr., Ko je začetnik psihoanalize?). Zbog čega je ovaj ajtem problematičan za analizu? Zbog toga što nulta varijansa proizvodi i nultu kovarijansu, odnosno varijabla koja ne varira ne može ni kovarirati/korelirati sa drugim varijablama. Ovo će dovesti do toga da se na osnovu matrice korelacija ne mogu izračunati željeni pokazatelji, zato što sama matrica korelacija ne zadovoljava preduslove da bude analizirana. Slične probleme proizvode i savršeno korelirane stavke, odnosno one na koje su ispitanici istovetno odgovarali. Iako može delovati da je malo verovatno da se ovo desi u praksi, ponekad na veoma lake ili veoma teške ajteme (ne)tačne odgovore daje mali broj istih ljudi, što može voditi maksimalnoj korelaciji između tih stavki.

U slučaju kada matrica korelacija nije pogodna za analizu program će takođe prijaviti grešku. Ovo će obično biti duži tekst u ispisu, a može se (više puta) pojaviti poruka kao što je „*An attempt has been made to use previously undefined matrix (or scalar). Execution of this command stops. Matrix - 'XX' is undefined*". Ponekad će se u ispisu pojaviti i matrica korelacija, ali ne uvek. Zato, ukoliko je sintaksa za pozivanje makroa ispravno napisana, ali makro ne daje željene pokazatelje metrijskih karakteristika, obavezno treba detaljno pogledati varijanse svih stavki i njihove korelacije kako bi se iz analize odstranile eventualne problematične stavke. Osim što dovode do nemogućnosti pokretanja analize, ove stavke su svakako loše – nulta varijansa ne doprinosi razlikovanju ispitanika, dok savršena korelacija dve stavke jednu od njih čini redundantnom. Dakle, pre pokretanja analize treba iz instrumenta izbaciti sve stavke sa nultom varijansom, kao i po (barem) jednu stavku iz svakog para savršeno koreliranih stavki.

Nakon matrice korelacija u ispisu možemo pronaći pokazatelje reprezentativnosti, pouzdanosti i homogenosti celog instrumenta, kao i pokazatelje metrijskih karakteristika stavki. Svi ovi pokazatelji nalaze se ispod odgovarajućih podnaslova koji ukazuju na to o kojim se metrijskim karakteristikama radi.

7.2) Provera reprezentativnosti preko makroa RTT10G

Od mera reprezentativnosti, preko programa RTT10G, možemo dobiti sve tri prethodno pomenute mere. Nenormirani KMO je označen grčim slovom Ψ_1 (PSI 1), normirani KMO (KMO_n) označen je grčkim slovom Ψ_2 (PSI 2), a naziva se i Kajzer Rajs merom reprezentativnosti (prema članku u kom je prvi put objavljena), dok je generalizacija KMO na antiimaž kovarijanse označena kao Ψ_{AIK} (PSI-AIK) (Slika 7.2).

```
MERE REPREZENTATIVNOSTI TESTA

Kaiser-Mayer-Olkin, mera reprezentativnosti, PSI 1 (Kaiser, 1970)
    .8728

Kaiser, Rice, mera reprezentativnosti, PSI 2 (KMO mera koju generisu komercijalni statisticki paketi, Kaiser & Rice, 1974)
    .8871

Kaiser-Mayer-Olkin, mera reprezentativnosti, PSI-AIK (varijanta PSI 1 sa anti-image kovarijansama)
    .9240
```

Slika 7.2

Vidimo da je reprezentativnost izražena kao PSI-AIK odlična, dok je vrednost nenormiranog i normiranog KMO dobra. Primećujemo da PSI-AIK mera repezentativnosti ima višu vrednost od KMO mere, što je u skladu sa načinom njihovog izračunavanja o kom je ranije bilo reči (PSI-AIK počiva na kovarijansama, a KMO na korelacijama antiimaža). Ukoliko vrednost normiranog KMO uporedimo sa normiranim KMO pokazateljem reprezentativnosti iz SPSS-a (Slika 5.2), vidimo da se radi o istoj vrednosti, s tim što je ispis preko makroa RTT10G prikazuje sa decimalom više. Važno je primetiti da, iako se računaju na drugačiji način, sve tri mere reprezentativnosti vode istom zaključku, a to je da je ovih 20 stavki testa znanja iz psihologije odabrano tako da dobro predstavljaju ceo univerzum potencijalnih pitanja na testu.

7.3) Provera pouzdanosti preko makroa RTT10G

Makro RTT10G daje veći broj mera pouzdanosti, organizovanih u nekoliko grupa u zavisnosti od toga na koji način se računa ukupan skor na testu (Slika 7.3).

```
MERE POUZDANOSTI POD KLASICNIM SUMACIONIM MODELOM

Guttman, LAMBDA 1
    .7716

Spearman-Brown-Kuder-Richardson-Guttman-Cronbach, ALFA
    .8122

Guttman, LAMBDA 6
    .8185


MERE POUZDANOSTI PRVE GLAVNE KOMPONENTE

Momirovic-Knezevic-Radovic, MI1
    .7843

Lord-Kaiser-Caffrey, BETA
    .8256

Momirovic, donja granica pouzdanosti, BETA 6
    .8364


MERE POUZDANOSTI POD GUTTMANOVIM MODELOM MERENJA

Momirovic, Knezevic, donja granica pouzdanosti, GAMA
    .8370

Guttman-Nicewander, RHO
    .8382
```

Slika 7.3

Tako imamo mere pouzdanosti pod klasičnim sumacionim modelom, što jednostavno znači da su to mere pouzdanosti za instrumente kod kojih se ukupni skor dobija sabiranjem (sumacijom) skorova na svim stavkama na instrumentu. Ovo su mere sa kojima smo se već susretali u odeljku o pouzdanosti i mogu se izračunati preko podrazumevanih SPSS analiza. Prvi koeficijent je Gutmanova lambda 1, za koju je već rečeno da predstavlja donju granicu pouzdanosti instrumenta (jer obrazac za njeno izračunavanje ne sadrži korekciju $\frac{n}{n-1}$). Ovaj koeficijent će zato uvek imati najnižu vrednost. Sledeći koeficijent je Gutmanova lambda 3, odnosno Kronbahova alfa. Kako je ovaj koeficijent numerički identičan i Spirman–Braunovom koeficijentu pouzdanosti, odnosno Kuder–Ričardson obrascu 20 u slučaju binarnih stavki – njegovo ime u ispisu uključuje sve navedene autore koji su, na ovaj ili onaj način, definisali dati koeficijent. Ipak, najčešća oznaka ovog koeficijenta, koja je i u makrou RTT10G odabrana kao dominantna, jeste upravo alfa koeficijent. Konačno, u ispisu dobijamo i Gutmanovu lambdu 6, koja se računa na osnovu udela varijanse imaža u ukupnoj varijansi i po pravilu je malo viša od lambde 3.

Gledajući samo vrednosti mera pouzdanosti za klasični sumacioni model, primećujemo da je pouzdanost najniža ako se izrazi kao λ_1, dok su vrednosti α i λ_6 više i međusobno slične. Ovde ćemo se podsetiti da mala razlika između λ_6 i α ukazuje na to da su pretpostavke o paralelnosti indikatora većinski ispunjene, te da se mogu računati i „klasični" pokazatelji pouzdanosti. λ_1 ukazuje na nešto nižu, ali zadovoljavajuću pouzdanost testa znanja iz psihologije, dok α i λ_6 ukazuju na dobru pouzdanost.

Ukoliko vrednosti dobijene makroom RTT10G uporedimo sa ispisom iz SPSS analize pouzdanosti, možemo primetiti da se ove vrednosti malo razlikuju od vrednosti dobijenih na nestandardizovanim varijabalama (Slika 3.12), ali da su identične vrednostima dobijenim na standardizovanim varijablama (Slika 3.14), uzimajući zaokruživanje na različit broj decimala u obzir. Ovo je posledica toga što RTT10G sve pokazatelje računa na osnovu matrica standardizovanih podataka. Kao što smo već konstatovali u poglavlju o pouzdanosti, razlike između pokazatelja dobijenih na standardizovanim i nestandardizovanim varijablama su minimalne i ne utiču na zaključke o pouzdanosti instrumenta koje donosimo.

Preostale dve grupe koeficijenata pouzdanosti su mere pouzdanosti prve glavne komponente i mere pouzdanosti pod Gutmanovim modelom merenja. Mere pouzdanosti prve glavne komponente se odnose na pouzdanost skora na instrumentu kada se skor računa kao prva glavna komponenta. Već smo rekli da je prva glavna komponenta varijabla koja objašnjava najveći procenat zajedničke varijanse svih pojedinačnih stavki, zbog čega pretpostavljamo da se odnosi na pravi predmet merenja. Kako prva glavna komponenta ne uključuje u sebe kompletnu varijansu, još jedno očekivanje je da se varijansa preostalih komponenti odnosi na varijansu greške. Drugim rečima, izdvajanjem prve glavne komponente eliminišemo deo varijanse greške, a zadržavamo pravu varijansu testa. Samim tim mere pouzdanosti bi trebalo da budu veće ukoliko se skor računa kao prva glavna komponenta, nego ako se računa sabiranjem svih stavki (kada pored zajedničke varijanse, skor uključuje i varijansu greške).

Iako tačne formule za njihovo izračunavanje ovde nećemo navoditi, mere μ_1, β i β_6 su analogne, redom, λ_1, α i λ_6. Vidimo da su sve mere pouzdanosti u modelu prve glavne komponente veće od analognih mera pouzdanosti sumacionog modela. Drugim rečima, kao što je očekivano, skor izračunat kao prva glavna komponenta u sebi sadrži manje varijanse greške od običnog sumacionog skora, te je udeo prave varijanse (pouzdanost) za ovakav skor veći. Ponovo, prva mera daje nešto nižu, ali i dalje zadovoljavajuću vrednost pouzdanosti za test znanja iz psihologije, dok preostale dve mere ukazuju na dobru pouzdanost testa.

Poslednje dve mere pouzdanosti su mere koje se dobijaju u slučaju kada se ukupan skor na instrumentu računa kao prva glavna komponenta, ali dobijena na osnovu matrice imaž kovarijansi. Ni njihove formule nećemo navoditi, samo ćemo kratko objasniti kako ih treba tumačiti. Pošto je imaž deo varijanse koji data varijabla deli sa drugim varijablama u modelu, smatra se da je imaž donja granica pouzdanosti – te se obe ove mere takođe mogu smatrati donjim granicama prave pouzdanosti instrumenta. U primeru testa znanja iz psihologije obe mere ukazuju na dobru pouzdanost.

7.4) Provera homogenosti preko makroa RTT10G

Za procenu homogenosti na osnovu makroa RTT10G na raspolaganju su nam mere H1, H2 i H5 (Slika 7.4). Primećujemo da je mera H1 jednaka prosečnoj interkorelaciji ajtema koju smo dobili i u SPSS ispisu analize pouzdanosti (Slika 6.2). Ova mera homogenosti značajno je niža i od H2 i od H5, što je očekivano imajući u vidu način njenog računanja i činjenicu da obično potcenjuje stvarnu homogenost. Mere H2 i H5 pak daju sličnu, povoljnu sliku o homogenosti testa. Deluje kao da je veliki deo varijanse testa objašnjen prvom glavnom komponentom, koja u ovom slučaju pretpostavljamo da predstavlja znanje iz psihologije. Drugim rečima, stavke testa se zaista u velikoj meri odnose na jedan isti konstrukt, a ne zahvataju druge konstrukte ili to čine u maloj meri.

```
MERE HOMOGENOSTI TESTOVA

Prosecna korelacija varijabli, H 1
    .1778

Momirovic, mera homogenosti, H 2
    .7916

Knezevic-Momirovic, mera homogenosti, H 5
    .7696
```

Slika 7.4

7.5) Analiza stavki preko makroa RTT10G

Još jedna od prednosti programa RTT10G u poređenju sa regularnim analizama u SPSS-u jeste što daje nekoliko pokazatelja kvaliteta stavki. Pritom, svi ovi pokazatelji nalaze se u jednoj tabeli (Slika 7.5), što bitno olakšava identifikovanje najboljih i najlošijih stavki. U redovima tabele nalaze se stavke koje su ušle u analizu dok se u kolonama nalaze, redom, mere reprezentativnosti (REP), pouzdanosti (REL) i dve mere valjanosti stavki (H i B).

REPREZENTATIVNOST, POUZDANOST I VALJANOST VARIJABLI

	REP	REL	H	B
t1	.9025	.1733	.4507	.4489
t2	.9032	.3159	.6162	.5918
t3	.9231	.2941	.6176	.5964
t4	.6507	.0755	.1594	.2247
t5	.8690	.1792	.4299	.4265
t6	.8995	.2428	.5363	.5159
t7	.9093	.2054	.5070	.5071
t8	.8825	.2061	.4792	.4764
t9	.8478	.1107	.3256	.3630
t10	.8947	.2449	.5357	.5259
t11	.8618	.1527	.3958	.4032
t12	.8675	.1813	.4397	.4622
t13	.8682	.0877	.2939	.3252
t14	.7844	.1123	.2844	.3239
t15	.8952	.3716	.6676	.6417
t16	.8703	.2267	.4848	.4803
t17	.9173	.1806	.4779	.4750
t18	.8815	.3085	.5906	.5711
t19	.9134	.2716	.5854	.5675
t20	.8634	.1674	.4200	.4317

Slika 7.5

Kao mera reprezentativnosti koristi se vrednost normiranog KMO za stavke, te će vrednosti iz ove tabele biti iste kao i vrednosti iz matrice antiimaž korelacija iz SPSS-a (Slika 5.2). I sada vidimo da je reprezentativnost stavke t4 upadljivo lošija od reprezentativnosti drugih stavki. Većina preostalih stavki ima ili odličnu reprezentativnost, preko .90, ili nešto nižu, ali i dalje dobru reprezentativnost, u opsegu od .80 do .90.

Kao mera pouzdanosti u RTT10G makrou prikazana je varijansa imaža datih stavki, odnosno proporcija varijanse date stavke koju objašnjavaju sve preostale stavke u instrumentu. Vrednosti u ovoj koloni identične su onima koje dobijamo u SPSS-u u okviru analize pouzdanosti prema Gutmanovom modelu, u koloni *Squarred Multiple Correlation* u okviru *Item-Total Statistics* tabele (Slika 3.15). Podsetimo se još jednom, ova vrednost odgovara multiplom R^2, koje se dobija kada se data stavka ubaci kao kriterijum u regresionu analizu, a sve ostale stavke koriste kao prediktori (Slika 3.16).

Razmatrajući vrednosti u drugoj koloni tabele metrijskih karakteristika stavki primećujemo ponovo da je stavka t4 najproblematičnija – ona ima najlošiju pouzdanost. Pored nje, nisku pouzdanost pokazuje i stavka t13, odnosno skor na ove dve stavke se slabo može predvideti na osnovu drugih stavki testa, što ukazuje da one imaju veći udeo jedinstvene varijanse. Generalno, pouzdanosti pojedinačnih stavki su niske, ali imajući u vidu da se imaž svake stavke procenjuje na osnovu 19 (a ne 29 ili 49) drugih stavki, ove vrednosti su zadovoljavajuće. U kontekstu testa znanja iz psihologije adekvatnom pouzdanošću bismo mogli smatrati vrednosti iznad .15 ili .20, pri čemu se kao posebno pouzdane izdvajaju stavke t2, t15 i t18. Da smo imali veći broj stavki u instrumentu, pouzdanosti svih pojedinačnih stavki bi izvesno bile više, te bi i granicu zadovoljavajućih vrednosti trebalo podići na višu vrednost.

Konačno, ispis daje i dve mere interne valjanosti stavki. Valjanost se odnosi na stepen u kom test ili stavka mere ono što bi trebalo da mere. Valjanost se može proveravati i dokazivati na više načina (čiji opis prevazilazi obim ove knjige), ali, kao što smo već pomenuli, po pravilu podrazumeva i dovođenje testa u vezu sa nekim spoljnim varijablama koje potvrđuju da on meri željeni konstrukt. Bitno je zato naglasiti da se ovde ne radi o valjanosti u „punom" smislu, već samo o internoj valjanosti stavki. Ali na šta se onda odnosi interna valjanost, kako je računamo? Ako nemamo nikakve spoljašnje kriterijume valjanosti stavke, kako najbolje procenjujemo u kojoj meri je stavka povezana sa konstruktom? U nedostatku drugih pokazatelja, najboljom merom konstrukta smatraćemo ukupni skor na instrumentu, te se mere interne valjanosti odnose na povezanost stavki sa glavnim predmetom merenja testa.

Već su nam poznata dva moguća načina računanja ukupnog skora na instrumentu – sumacija skorova na stavkama i računanje prve glavne komponente – i mere interne valjanosti stavki se odnose upravo na ova dva shvatanja ukupnog skora. Tako valjanost u Hotelingovom (Hotelling) prostoru (H) predstavlja korelaciju stavke sa prvom glavnom komponentom, odnosno varijablom koja objašnjava najveći deo zajedničke varijanse, dok se valjanost u Bartovom (Burt) prostoru (B) odnosi na korelaciju stavke sa sumarnim skorom na instrumentu. Pošto RTT10G radi sa standardizovanim varijablama, ovo se odnosi na korelaciju standardizovanih skorova na stavkama sa sumarnim skorom (ili prvom glavnom komponentom, u slučaju mere H) izračunatim na osnovu standardizovanih varijabli. Korišćenjem komande *WITH* u SPSS-u (videti Dodatak 1) lako možemo proveriti da su vrednosti u koloni B zaista obične korelacije stavki sa ukupnim skorom (Slika 7.6).

Correlations

		ZTZ_20_Z
Zt1 Zscore(t1)	Pearson Correlation	.449**
Zt2 Zscore(t2)	Pearson Correlation	.592**
Zt3 Zscore(t3)	Pearson Correlation	.596**
Zt4 Zscore(t4)	Pearson Correlation	.225**
Zt5 Zscore(t5)	Pearson Correlation	.426**
Zt6 Zscore(t6)	Pearson Correlation	.516**
Zt7 Zscore(t7)	Pearson Correlation	.507**
Zt8 Zscore(t8)	Pearson Correlation	.476**
Zt9 Zscore(t9)	Pearson Correlation	.363**
Zt10 Zscore(t10)	Pearson Correlation	.526**
Zt11 Zscore(t11)	Pearson Correlation	.403**
Zt12 Zscore(t12)	Pearson Correlation	.462**
Zt13 Zscore(t13)	Pearson Correlation	.325**
Zt14 Zscore(t14)	Pearson Correlation	.324**
Zt15 Zscore(t15)	Pearson Correlation	.642**
Zt16 Zscore(t16)	Pearson Correlation	.480**
Zt17 Zscore(t17)	Pearson Correlation	.475**
Zt18 Zscore(t18)	Pearson Correlation	.571**
Zt19 Zscore(t19)	Pearson Correlation	.568**
Zt20 Zscore(t20)	Pearson Correlation	.432**

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 7.6

I ovog puta se stavka t4 pokazuje kao stavka koja najlošije korelira sa ukupnim skorom na instrumentu, bilo da gledamo Hotelingov ili Bartov prostor. Većina stavki korelira i sa sumarnim skorom i sa prvom glavnom komponentom preko .40, što ukazuje na dobru internu valjanost. Neke korelacije čak prelaze vrednost .60 (stavke t2, t3, t15), na osnovu čega možemo zaključiti da one posebno dobro odražavaju glavni predmet merenja testa znanja iz psihologije, odnosno znanje iz ove oblasti. Generalno će vrednosti u kolonama H i B biti slične pod uslovom da prva glavna komponenta objašnjava nezanemarljiv deo ukupne varijanse, te su sumarni skor i prva glavna komponenta međusobno slične (korelirane). Na primeru testa znanja iz psihologije vidimo da je interna valjanost gotovo svih stavki bila veoma dobra.

Informacije koje dobijamo na osnovu makroa RTT10G mogu nam biti korisne i za davanje globalne ocene o metrijskim karakteristikama instrumenta, ali i za analizu pojedinačnih stavki, posebno u situacijama kada želimo da skraćivanjem poboljšamo metrijska svojstva testa. Lako možemo kopirati ispis ajtem analize u neki od programa za rad sa tabelama (npr., *Excel*), te sortirati stavke po određenom kriterijumu ili izračunati prosečnu vrednost svakog od pokazatelja, što, u kombinaciji sa drugim informacijama o instrumentu i stavkama, može olakšati donošenje odluka o zadržavanju/izbacivanju stavki iz instrumenta.

7.6) Rezime

Većina pokazatelja globalnih metrijskih karakteristika mernih instrumenata može se dobiti u okviru predefinisanih opcija IBM SPSS statističkog softvera, a isto važi i za jedan deo pokazatelja metrijskih karakteristika stavki. Ipak, korišćenjem makroa RTT10G na jednom mestu dobijamo daleko veći broj informacija o instrumentu i njegovim stavkama – od nekoliko mera reprezentativnosti, preko mera pouzdanosti pod klasičnim sumacionim modelom, modelom prve glavne komponente i Gutmanovim modelom imaž varijanse, do tradicionalnih i savremenih mera homogenosti testa. Od pokazatelja kvaliteta stavki na raspolaganju su nam mere reprezentativnosti, pouzdanosti i dve mere interne valjanosti stavki. Makro se veoma jednostavno pokreće, što olakšava njegovo rutinsko korišćenje.

7.7) Preporučena literatura

Knežević, G., & Momirović, K. (1996). RTT9G i RTT10G: Programi za analizu metrijskih karakteristika. In P. Kostić (Ed.), *Merenje u psihologiji* (pp. 37–56). Institut za kriminološka i sociološka istraživanja, Beograd.

Momirović, K., Wolf, B., & Popović, D. A. (1999). *Uvod u teoriju merenja i interne metrijske karakteristike kompozitnih mernih instrumenata*. Univerzitet u Prištini, Fakultet za fizičku kulturu.

8) TEORIJA ODGOVORA NA STAVKE

Ukoliko razmišljamo o bilo kojoj osobini koja bi mogla biti predmet psihološkog merenja, na primer, o znanju psihologije, anksioznosti ili političkim uverenjima – jedna od osnovnih ideja jeste da se ispitanici razlikuju po stepenu prisustva ove osobine. Upravo ova ideja je i preduslov svakog merenja. Sa druge strane, i stavke se razlikuju po stepenu u kom se odnose na osobinu. Kako ispitanik može posedovati više ili manje znanja iz psihologije, tako i stavka u testu znanja može biti lakša (zahtevati manje znanja da bi se tačno rešila) ili teža (takva da zahteva viši nivo znanja za tačan odgovor). Slično ovome, ispitanici mogu biti vrlo anksiozni, umereno ili veoma malo anksiozni, a stavke kojima merimo anksioznost mogu takođe biti „lake", odnosno odnositi se na nizak nivo anksioznosti (npr., „Osećam tremu samo pred veoma važne ispite") ili „teške", koje se odnose na visok nivo anksioznosti (npr., „Izbegavam određene situacije zbog straha od paničnog napada"). Teorija odgovora na stavke odnosi se upravo na model odnosa nivoa osobine kod ispitanika i stavki (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Lord & Novick, 2008).

Teorija odgovora na stavke u literaturi se može naći i pod nazivima teorija odgovora na ajteme/teorija ajtemskog odgovora (TAO) (Fajgelj, 2009; Fajgelj & Kosanović, 2001), dok je izvorni engleski termin *Item Response Theory* (IRT) (Lord, 1980). U daljem tekstu služićemo se skraćenicom IRT.

U prethodnim poglavljima bavili smo se metrijskim karakteristikama testa u skladu sa klasičnom teorijom testa i razmotrili pokazatelje diskriminativnosti, objektivnosti, pouzdanosti, homogenosti i reprezentativnosti. Sada ćemo prikazati osnovne pretpostavke teorije odgovora na stavke, kao i dodatne mere koje se u ovom modelu mogu dobiti.

8.1) IRT kao alternativa klasičnoj teoriji testa

Do sada smo pored klasične teorije testa i njene specifikacije u vidu modela paralelnih indikatora, pominjali još i Gutmanov imaž model merenja. IRT predstavlja još jedan pogled na teoriju merenja i počiva na proceni verovatnoće da će dati ispitanik tačno rešiti određenu stavku. Greška merenja se u modelu može izračunati kao razlika između dobijenog skora i izračunate verovatnoće (koja počiva na teorijskom modelu ponašanja ispitanika i stavki). Ovakvo računanje greške omogućava nam da direktno testiramo da li se (i u kojoj meri) podaci ponašaju u skladu sa našim mernim modelom (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Lord, 1980).

Kao jedna od osnovnih razlika između IRT-a i modela koji počivaju na klasičnoj testovnoj teoriji ističe se i da se klasični modeli bave *skorovanjem* ispitanika (dodeljivanjem skorova na testu, najčešće sabiranjem skorova na pojedinačnim ajtemima), dok se IRT bavi *skaliranjem* ispitanika, odnosno njihovim pozicioniranjem na kontinuum merene latentne dimenzije (Fajgelj, 2009; Fajgelj & Kosanović, 2001; Harvey & Hammer, 1999). Na isti ovaj kontinuum se pozicioniraju i stavke, jer se i one odnose na latentnu dimenziju.

Ideja skaliranja tesno je povezana sa idejom o jednodimenzionalnosti merenja, s obzirom na to da ispitanike i stavke možemo pozicionirati na kontinuumu latentne dimenzije samo ukoliko se zaista i radi o jednoj latentnoj dimenziji[29] (An & Yung, 2014; de Ayala, 2009; Fajgelj, 2009; Hambleton & Jones, 1993). *Pretpostavka jednodimenzionalnosti* je u okviru IRT-a veoma eksplicitna, odnosno direktno se testira da li su podaci jednodimenzionalni ili nisu (najčešće se sprovodi analiza glavnih komponenti na rezidualima koji nisu objašnjeni modelom) (Fajgelj, 2009; Linacre, 2022; Lord & Novick, 2008). Iako nisu nužno međusobno povezane, iz pretpostavke o jednodimenzionalnosti sledi i pretpostavka o *uslovnoj*, odnosno *lokalnoj nezavisnosti* stavki (de Ayala, 2009; Fajgelj, 2009; Lord & Novick, 2008). Uslovna nezavisnost stavki podrazumeva da su odgovori na stavke međusobno nezavisni, odnosno nekorelirani, kada uzmemo u obzir poziciju ispitanika na latentnoj varijabli. Drugim rečima, za ispitanike koji se međusobno ne razlikuju po nivou latentne dimenzije (npr., znanja iz psihologije) verovatnoća tačnog odgovora na jednu stavku ne bi trebalo da bude povezana sa verovatnoćom tačnog odgovaranja na drugu stavku. Naime, korelacije između ajtema bi trebalo da budu posledica samo merene latentne dimenzije i ničeg drugog, te kada statistički držimo pod kontrolom latentnu dimenziju – očekujemo i da će korelacije pojedinačnih ajtema postati nulte. Takođe, u IRT-u postoji i pretpostavka *monotonosti*, u skladu sa kojom očekujemo da će sa porastom latentne varijable rasti i verovatnoća tačnog odgovora na stavke, odnosno da između osobine i skorova postoji monoton odnos (Linacre, 2022).

8.2) Očekivanja u IRT

Zamislimo da imamo dva ispitanika, ispitanika A i ispitanika B, koji su radili test znanja iz psihologije koji ima 30 pitanja. Recimo da je ispitanik A odgovorio tačno na samo 5 pitanja, dok je ispitanik B tačno odgovorio na 25 pitanja (odnosno samo na 5 pitanja je dao pogrešan odgovor). Pretpostavili bismo da su 5 stavki koje je rešio ispitanik A lakše stavke, te da ih je rešio i ispitanik B (koji poseduje veći nivo znanja). Isto tako, očekivali bismo da su stavke koje ispitanik B nije rešio teške, te da ih samim tim nije rešio ni ispitanik A (koji ima manji nivo znanja).

Uzmimo sada drugi primer. Test znanja od pet ajtema uradilo je petoro ispitanika (Tabela 8.1). Dakle, neki ispitanici su tačno rešili veći broj stavki, dok su neki rešili manje. Neke stavke je rešio veći broj ispitanika, a neke manji. Za svakog ispitanika možemo izračunati njegov skor na testu kao proporciju tačnih odgovora. Isto tako možemo dobiti težinu svake stavke kao proporciju ispitanika koji su je tačno rešili. Obratite pažnju na to da niža vrednost/manja proporcija tačnih odgovora označava veću težinu stavke (jer ju je manji procenat ispitanika tačno rešio).

Tabela 8.1. Uobičajen izgled podataka sa testa

	Stavka 1	Stavka 2	Stavka 3	Stavka 4	Stavka 5	Skor ispitanika
Ispitanik 1	0	1	1	0	0	0.4
Ispitanik 2	0	1	1	1	1	0.8
Ispitanik 3	0	1	1	0	1	0.6
Ispitanik 4	1	1	1	1	1	1
Ispitanik 5	0	1	0	0	0	0.2
Težina stavke	0.2	1	0.8	0.4	0.6	

[29] Postoje zapravo i multidimenzionalni IRT modeli (de Ayala, 2009; Drasgow & Hulin, 1990), ali se njima nećemo detaljnije baviti jer su jednodimenzionalni modeli daleko češći u praktičnoj upotrebi.

Možemo, međutim, tabelu organizovati i na drugačiji način, tako da ispitanike poređamo po ukupnom skoru od onih koji su tačno rešili najmanji, do onih koji su rešili najveći broj zadataka, a stavke po težini od onih koje su najčešće do onih koje su najređe bile tačno rešene. Ako to uradimo, dobijamo pregledniji prikaz istih podataka (Tabela 8.2.).

Tabela 8.2. Podaci sortirani po znanju ispitanika i težini stavki

	Stavka 2	Stavka 3	Stavka 5	Stavka 4	Stavka 1	Skor ispitanika
Ispitanik 5	1	0	0	0	0	0.2
Ispitanik 1	1	1	0	0	0	0.4
Ispitanik 3	1	1	1	0	0	0.6
Ispitanik 2	1	1	1	1	0	0.8
Ispitanik 4	1	1	1	1	1	1
Težina stavke	1	0.8	0.6	0.4	0.2	

Sada se mnogo jasnije primećuje pravilnost u podacima – da su najlakše stavke rešavali svi ili skoro svi ispitanici, dok su teže stavke rešavali samo neki, i to oni koji su posedovali više znanja. Takođe, ispitanici sa malo znanja su uspeli da reše samo najlakše stavke, dok su ispitanici koji su više znali tačno rešili i najteže stavke. Tako dolazimo i do osnovnih očekivanja u IRT-u – što su ispitanici sposobniji (imaju više znanja, izraženiju osobinu), to je veća verovatnoća da će tačno odgovarati na stavke (odnosno da će se slagati sa tvrdnjama) (Lord, 1980). Ujedno, što je stavka teža, manja je verovatnoća da će ispitanici na nju tačno odgovoriti.

Kroz sve do sada opisano nazire se suština IRT-a. Postoji kontinuum merne osobine na kome su pozicionirani kako ispitanici tako i stavke. Ispitanici poseduju osobinu u određenom stepenu, dok stavke mere datu osobinu na nekom nivou, odnosno „zahtevaju" određeni stepen izraženosti osobine da bi bile tačno rešene (de Ayala, 2009; Fajgelj, 2009; Fajgelj & Kosanović, 2001). Ono što je ključno jeste odnos ova dva nivoa – ukoliko je nivo osobine ispitanika izraženiji od nivoa koji je potreban za tačno rešavanje stavke – ispitanik bi trebalo da je tačno reši. Ukoliko je nivo osobine koji zahteva stavka veći od nivoa osobine koju poseduje ispitanik – on ne bi trebalo tačno da je reši. Što je razlika u nivou osobine između ispitanika i stavke veća – to smo uvereniji da će je on tačno, odnosno netačno rešiti (ukoliko je više „iznad", odnosno „ispod" nivoa stavke). I ne samo to – u IRT-u možemo izračunati tačnu verovatnoću sa kojom bi ispitanik datog nivoa osobine trebalo da tačno reši stavku date težine! O ovome će biti više reči kasnije.

8.2.1) Mapa ispitanika i stavki

S obzirom na to da je kontinuum osobine koju ispitanici poseduju, a stavke mere, jedan jedinstven kontinuum – moguće je prikazati i ispitanike i stavke na istom grafiku. Ova vrsta grafika naziva se mapom ispitanika i stavki (eng. *person-item map*) i ovde ćemo se služiti grafičkim prikazom koji se dobija u programima Winsteps/Ministep (Linacre, 2022). Slika 8.1 pokazuje kako bi izgledalo ukoliko bismo podatke iz prethodnog primera (Tabela 8.1, odnosno Tabela 8.2) prikazali na jednom grafiku[30].

[30] Zapravo, sa samo 5 ispitanika i samo 5 stavki ne bismo mogli da napravimo pravu mapu ispitanika i stavki, ali je ovde navodimo za potrebe ilustracije.

Vertikalna linija na sredini označava merenu dimenziju, u ovom slučaju znanje ispitanika. Kao što u vrhu mape piše, sa leve strane se nalaze ispitanici, dok su stavke sa desne strane. Sasvim levo napisane su vrednosti koje se odnose na skorove ispitanika, u ovom slučaju one se kreću od minimalnih 0 do maksimalnih 1. Ispitanici su poređani tako da se na vrhu mape nalaze oni sa većim skorovima, odnosno oni sa većim znanjem, a nivo znanja se smanuje kako idemo ka donjem delu mape. Stavke su pak poređane od onih najređe tačno rešavanih, odnosno od najtežih, do najlakših – onih u najvećem broju slučajeva tačno rešenih. Možemo tako videti da je najlakša stavka 2, koja odgovara minimalnom mogućem skoru koji bi se mogao postići na ovom testu, dok nijedan ispitanik nije postigao minimalan skor. Suprotno tome, na najvišem nivou znanja nalazi se ispitanik 4, koji je tačno rešio sve stavke i nijedna stavka nije na ovom nivou, odnosno ne zahteva „maksimalno znanje" ispitanika da bi bila tačno rešena. Ispitanik 3 i stavka 4 su pak „izjednačeni" po nivou znanja, odnosno stavka 4 je svojom težinom primerena upravo nivou znanja ispitanika 3.

```
             Ispitanici   Stavke
                  <više>|<ređe>
 1          Ispitanik 4 +
                        |
                        |
0.8         Ispitanik 2 |  Stavka 1
                        |
                        |
0.6         Ispitanik 3 |  Stavka 4
                        |
                        |
0.4         Ispitanik 1 |  Stavka 5
                        |
                        |
0.2         Ispitabik 5 |  Stavka 3
                        |
                        |
 0                      +  Stavka 2
                 <manje>|<češće>
```

Slika 8.1

Kao što je već napomenuto, ovo je zamišljeni primer, a prava mapa ispitanika i stavki mogla bi izgledati kao što ilustruje Slika 8.2. Sada imamo nešto veći broj kako stavki, tako i ispitanika. Ovde smo kao primer iskoristili već pominjani test anksioznosti od 16 ajtema. Ispitanici su na mapi označeni tačkom i znakom #, dok su ajtemi označeni svojim rednim brojem (npr., I0015 označava Item 0015, odnosno 15. stavku). Pošto u slučaju velikih uzoraka ne bi bilo moguće prikazati svakog pojedinačnog ispitanika na mapi – znak # obično označava veći broj ispitanika (npr., 2 kao u našem primeru, ili neki veći broj, u zavisnosti od veličine uzorka i postizanja najpreglednijeg prikaza), dok je pojedinačni ispitanik prikazan tačkom.

```
      Persons MAP OF Items
             <more>|<rare>
  2                +
                   |
                   |
                   |
                   |T
                   |
                   |
                   |
                   |
                   |  I0001  I0009
  1                +
                   |S I0008  I0016
                .  |  I0002  I0012
                .  |
                   |
                . T|  I0007
                #  |
          .######  |  I0013
               ## S|
           ######  |M I0015
  0   .########### M+
         .#######  |  I0005
          ###### S|  I0006
               .#  |
               .#  |
               ## T|
                   |S I0011
                .  |  I0004
                   |  I0003  I0014
                   |  I0010
 -1                +
             <less>|<frequ>
EACH '#' IS 2.
```

Slika 8.2

Ukoliko bismo ovu mapu ispitanika i stavki zarotirali za 90º udesno (tako da ispitanici budu iznad, a stavke ispod linije koja definiše mernu dimenziju) primetili bismo da gornji deo grafika – distribucija skorova ispitanika – nalikuje histogramu. U ovom slučaju distribucija deluje približno normalno. U idealnom slučaju i distribucija skorova ispitanika i distribucija težine varijabli će biti normalne. Distribucija stavki, međutim, deluje relativno uniformno, s tim što postoji određeni broj stavki koje su veoma „lake", odnosno „teške" za ispitanike. Drugim rečima, neke stavke su tako formulisane da se velika većina ispitanika sa njima slagala (stavke na dnu mape), te nam ne pomažu da razlikujemo ispitanike po nivou anksioznosti, dok su druge stavke tako formulisane da se sa njima skoro niko nije slagao (stavke na vrhu mape), te takođe nisu informativne. Generalno, distribucija ajtema treba da odgovara željenoj diskriminativnosti ispitanika prema merenoj osobini. Drugim rečima, poželjno je da imamo najveći broj stavki oko onog dela distribucije gde nam je potrebno najpreciznije razlikovanje (An & Yung, 2014; Lord, 1980). Najčešće, to će biti u središnjem delu distribucije (gde očekujemo da će se grupisati najveći broj ispitanika), ali u nekim situacijama može nam biti posebno važno da razlikujemo ispitanike sa najniže/najviše izraženom latentnom osobinom.

Takođe, na mapi možemo videti i kako su generalno pozicionirane stavke u odnosu na ispitanike. Slovom M sa leve strane vertikalne linije označena je aritmetička sredina za ispitanike, dok je aritmetička sredina za stavke takođe označena slovom M, ali sa desne strane linije. U idealnom slučaju, kada je težina testa dobro podešena znanju / sposobnosti / osobini ispitanika – aritmetičke sredine će se poklapati ili biti veoma blizu jedna druge. Ukoliko je aritmetička sredina za ispitanike iznad aritmetičke sredine za stavke, radi se o lakom testu – jer je većina ispitanika imala više znanja nego što stavke zahtevaju. Ukoliko je aritmetička sredina ispitanika ispod aritmetičke sredine stavki, radi se o teškom testu – s obzirom na to da su ispitanici većinom znali manje od onog što su stavke zahtevale za tačno rešavanje. Naš test anksioznosti pokazao se kao dobro podešen za ovaj uzorak ispitanika.

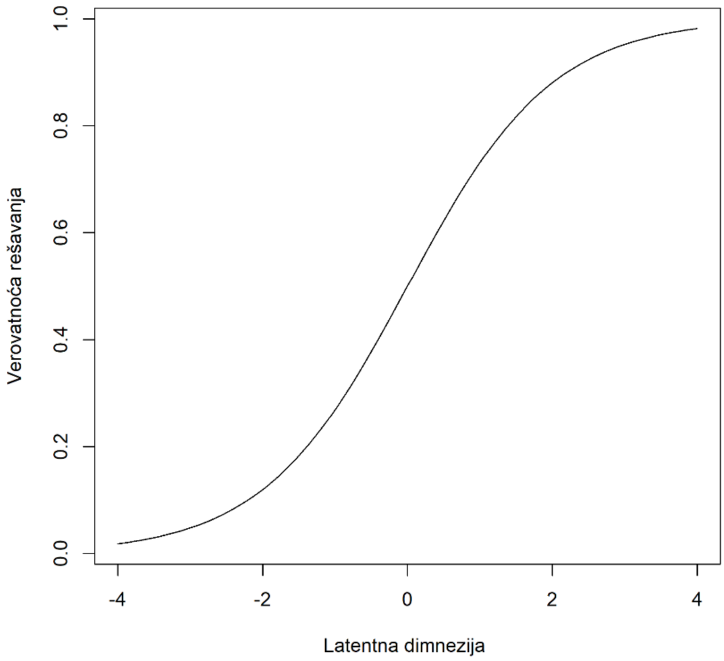
Kako bismo interpretirali „težak”, odnosno „lak” test kada se radi o osobini ličnosti? Ako je test „lak”, možemo reći da su ispitanici bili skloni da se većinski slažu sa stavkama u instrumentu, tako da su stavke bile diskriminativne pretežno za ispitanike sa nižim skorovima anksioznosti, dok visoko anksiozne ispitanike nisu mogle dobro da razlikuju. Drugim rečima, stavke su bile „blage” i formulisane na takav način da je sa njima bilo lako složiti se. Obrnuto, ako je test „težak”, to znači da su stavke bile diskriminativne samo za veoma anksiozne ispitanike, jer su tako formulisane da je teško složiti se sa njima, te će samo neko sa visokim skorom na anksioznosti označiti veći stepen slaganja sa tvrdnjom.

8.2.2) Karakteristična kriva stavke

Prethodno pominjana očekivanja u IRT-u mogu se grafički predstaviti i preko karakteristične krive stavke – KKS (engleski *Item Characteristic Curve, ICC*). Ova kriva prikazuje teorijski model porasta verovatnoće tačnog rešavanja pojedinačne stavke sa porastom nivoa osobine ispitanika (Baker, 2001; Fajgelj, 2009; Hambleton & Jones, 1993; Lord, 1952; Lord & Novick, 2008). Slika 8.3 prikazuje tipičnu karakterističnu krivu binarno skorovane stavke. Na njoj možemo videti kako verovatnoća tačnog rešavanja raste od nepostojeće kada je nivo osobine ispitanika veoma nizak u poređenju sa težinom stavke, do izvesnosti tačnog rešavanja kada je nivo osobine ispitanika veoma visok u poređenju sa težinom stavke. Kriva ima oblik normalne ili logističke ogive[31], odnosno S oblik, a najveći nagib je u središnjem delu, gde su nivo osobine ispitanika i težina stavke izjednačeni. Važno je primetiti da odnos između nivoa latentne osobine i verovatnoće tačnog odgovora nije linearan. To znači da mali porast u latentnoj dimenziji nekad može voditi velikom porastu verovatnoće tačnog rešavanja stavke (ako su sposobnost ispitanika i težina stavke blizu) i obrnuto, veliki porast u nivou

[31] Zbog jednostavnijeg računanja parametara, logistička ogiva i model koji na njoj počiva (logit model) češće se koriste od normalne ogive i modela koji na njoj počiva (probit model) (Baker, 2001; Drasgow & Hulin, 1990; Linacre, 2022).

izraženosti latentne dimenzije može voditi malom porastu verovatnoće tačnog rešavanja (kada su sposobnost ispitanika i težina ajtema međusobno udaljene).



Slika 8.3. Karakteristična kriva binarno skorovane stavke

Šta se dešava kada su nivo osobine ispitanika i težina stavke izjednačeni? U tom slučaju postoji verovatnoća od 50% da će ispitanik tačno rešiti stavku. Pošto i ispitanik i stavka „imaju" isti nivo osobine, podjednako je verovatno da će stavka „pobediti", odnosno da je ispitanik neće tačno rešiti i da će „pobediti" ispitanik tako što će tačno rešiti stavku.

8.3) IRT modeli po broju parametara

Kao što smo rekli, karakteristična kriva stavke na grafički način predstavlja teorijski, odnosno matematički model odnosa sposobnosti (ili druge latentne dimenzije) ispitanika i verovatnoće tačnog rešavanja stavke. Drugim rečima, verovatnoća tačnog rešavanja stavke je *u funkciji* sposobnosti ispitanika. Međutim, poznavanje samo sposobnosti ispitanika nije dovoljno da bismo odredili sa kojom verovatnoćom će on rešiti neku stavku. Jednačina kojom dolazimo do karakteristične krive stavke uključuje još neke numeričke vrednosti, odnosno parametre. Spram toga koliko parametara uključuju, kao osnovne, možemo razlikovati jednoparametarske, dvoparametarske i troparametarske IRT modele (An & Yung, 2014; Baker, 2001; de Ayala, 2009; Drasgow & Hulin, 1990; Fajgelj, 2009; Hambleton & Jones, 1993; Lord & Novick, 2008). U nastavku teksta prikazaćemo modele koji su formulisani za binarne, odnosno dihotomne ajteme, ali se analogni modeli mogu primenjivati i za politomne ajteme, gde se odgovori daju na ordinalnoj ili intervalnoj skali sa većim brojem ponuđenih odgovora. U ovom slučaju, parametri se procenjuju za svaku od kategorija odgovora, na osnovu čega se dobija informacija o celom ajtemu.

8.3.1) Jednoparametarski modeli

U jednoparametarskim logističkim modelima (1PL) pretpostavljamo da se stavke međusobno razlikuju samo po jednom svojstvu (parametru), a to je njihova **težina**. Ovaj model u potpunosti odgovara do sada iznetim očekivanjma u okviru IRT-a. Dakle, verovatnoća da će ispitanik tačno rešiti stavku zavisi od njegove sposobnosti i težine stavke. To se može prikazati i sledećom formulom:

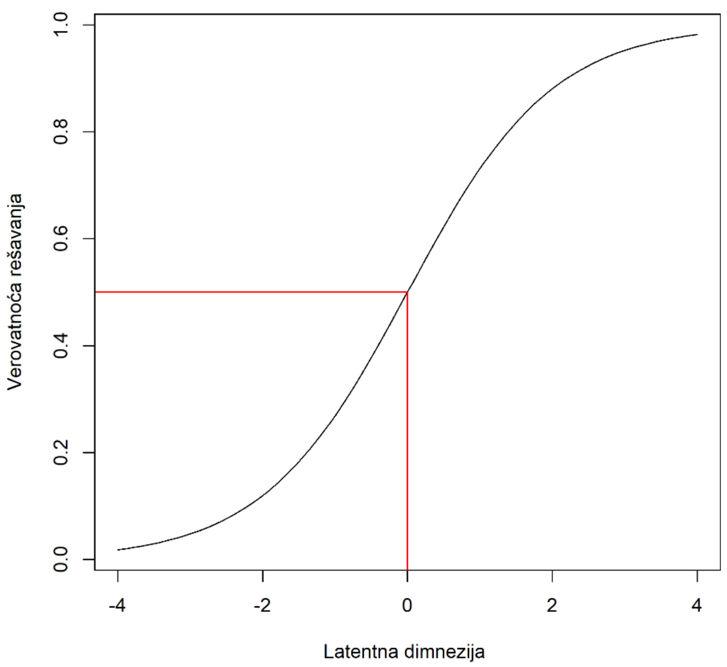
$$P_i(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}}$$

gde θ označava kontinuum sposobnosti ispitanika, b_i težinu stavke *i*, a *e* je osnova prirodnog logaritma. $P_i(\theta)$ se odnosi na verovatnoću da će ispitanik određenog nivoa sposobnosti tačno rešiti datu stavku, budući da je verovatnoća u funkciji sposobnosti ispitanika, te nije ista za sve nivoe sposobnosti. Vrednosti parametra težine zavise od merne skale koju koristimo, a u slučaju standardizovanih skorova kreću se obično između vrednosti -2 i +2 (u ovom opsegu se nalazi 95.45% slučajeva ispod normalne krive).

Šta se dešava ako su sposobnost ispitanika i težina stavke jednakog nivoa? Zamenjivanjem u jednačini dobijamo:

$$P_i(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}} = \frac{e^0}{1 + e^0} = \frac{1}{1 + 1} = .5$$

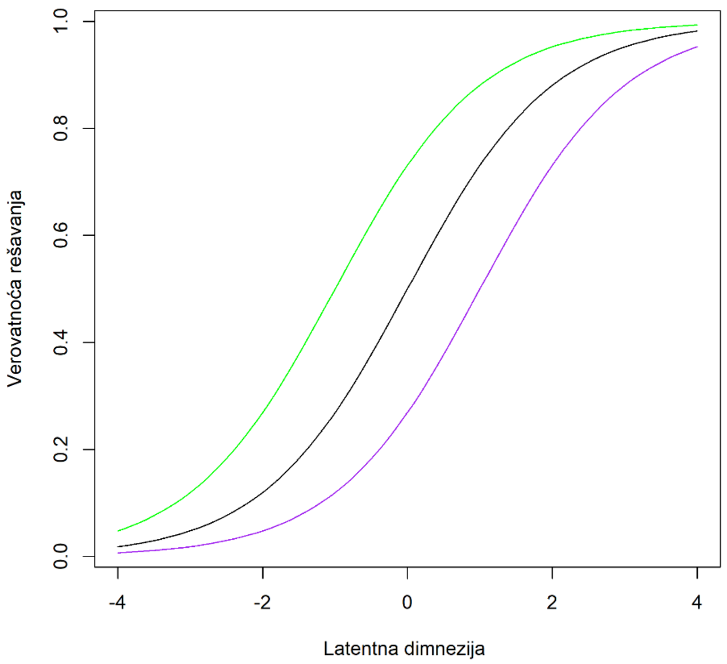
odnosno verovatnoću od 50% da ispitanik tačno reši stavku, kao što smo prethodno i rekli. Ova verovatnoća nam takođe može poslužiti i da na KKS odredimo težinu stavke. To činimo tako što povlačimo liniju sa *y* ose iz tačke gde je verovatnoća rešavanja .5 dok ne presečemo KKS, a zatim pod pravim uglom „spustimo" liniju na *x* osu. Vrednost *x* ose koju tako očitamo odgovara težini stavke. Drugim rečima, težinu stavke možemo razumeti kao njenu **lokaciju**, položaj na kontinuumu latentne osobine (Baker, 2001; Drasgow & Hulin, 1990; Fajgelj, 2009). Slika 8.4. prikazuje primer gde težina stavke odgovara skoru 0, odnosno prosečnom nivou sposobnosti ispitanika (ukoliko pretpostavimo da su skorovi centrirani).



Slika 8.4

Na sličan način možemo uporediti nekoliko stavki po težini. Slika 8.5 prikazuje tri stavke koje se razlikuju po težini. Kako možemo odrediti koja od njih je najlakša, a koja najteža? Možemo primeniti isti metod, odnosno povući liniju iz vrednosti .5 na *y* osi i videti gde ona preseca ove tri stavke. Jasno je da je stavka prikazana zelenom bojom, čiju KKS ova linija prvo preseče, najlakša jer kada bismo spustili liniju na *x* osu, dobijamo najniži nivo osobine. Sledeća je crna KKS, srednje težine, dok je ljubičasta KKS, poslednja koju zamišljena linija preseče, najteža (pošto je da bi se rešila sa 50% verovatnoće potreban najveći nivo sposobnosti / znanja / osobine).

Alternativno, možemo krenuti od *x* ose i povući liniju iz, na primer, vrednosti 0, odnosno prosečne sposobnosti ispitanika. Krećući se na gore, prva linija koju presečemo je ljubičasta i ona pokazuje najmanju verovatnoću da ispitanik prosečnog nivoa sposobnosti tačno reši datu stavku – odnosno ukazuje na najtežu stavku. Sledeća je crna linija, za koju je verovatnoća srednja, te ukazuje na srednju težinu stavke, a poslednja presećena linija je zelena, kod koje postoji najveća verovatnoća tačnog rešavanja, odnosno ova stavka ima najmanju težinu. I jedan i drugi metod vode nas istom zaključku – da je skroz leva stavka najlakša (čija je KKS zelena), dok je ona skroz desno (ljubičasta KKS) najteža.



Slika 8.5

Treba još reći i da se jednoparametarski modeli često nazivaju i Rašovi (Rasch) modeli, prema svom tvorcu Georgu Rašu, danskom matematičaru, statističaru i psihometričaru. Rašov model je jednoparametarski logistički model, kod kog je diskriminativnost (pojam koji će biti objašnjen u narednoj sekciji) svih ajtema fiksirana na vrednost 1. S obzirom na to da postoje i drugačiji jednoparametarski modeli, Rašov model predstavlja uži pojam od jednoparametarskog modela. Takođe, Rašov model polazi od ideje da ne treba prilagođavati model tako da odgovara podacima, već kreirati takav instrument koji će dobro odgovarati teorijskom modelu (de Ayala, 2009; Fajgelj, 2009; Hambleton & Jones, 1993). Jedan od najčešće korišćenih softvera za IRT analizu, Winsteps, počiva upravo na jednoparametarskom modelu (Linacre, 2022).

8.3.2) Dvoparametarski modeli

Dvoparametarski logistički modeli (2PL) pretpostavljaju da se stavke međusobno razlikuju po dve osobine – po **težini** (kao u jednoparametarskom modelu) i po **diskriminativnosti**. Diskriminativnost se ovde razume na suštinski isti način kako smo je i do sada posmatrali – kao svojstvo stavke da razlikuje ispitanike koji imaju, odnosno nemaju neku osobinu ili, preciznije rečeno, koji je imaju u manjem i većem stepenu. Jednačina koja definiše dvoparametarske modele je sledeća:

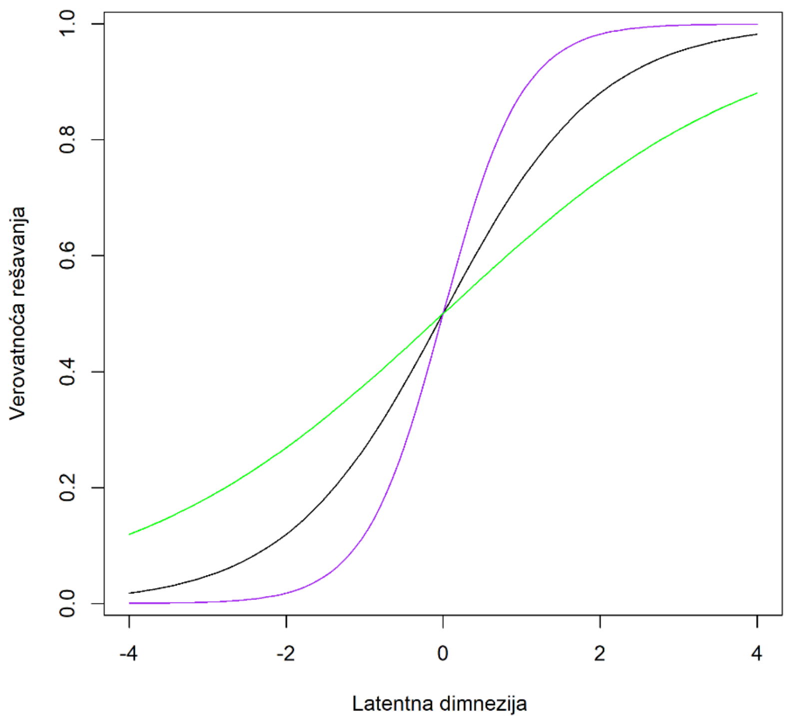
$$P_i(\theta) = \frac{e^{a_i*(\theta-b_i)}}{1+e^{a_i*(\theta-b_i)}}$$

gde je a_i diskriminativnost stavke i, a sve druge oznake iste kao kod jednoparametarskih modela. Vrednosti dikriminativnosti se mogu kretati od nule – što odgovara potpuno nediskriminativnoj stavci, koja uopšte ne razlikuje ispitanike – do pozitivne beskonačnosti u slučaju savršeno diskriminativne stavke. Veoma niskom diskriminativnošću (u slučaju logističkog modela) smatraju se vrednosti do .34, niskom od .35 do .64, umerenom od .65 do 1.34, visokom od 1.35 do 1.69, a veoma visokom više od 1.70 (Baker, 2001). Vraćajući se na formulu za dvoparametarski model, možemo primetiti da fiksiranjem diskriminativnosti svih ajtema na a = 1 (što odgovara Rašovom modelu) dobijamo jednoparametarski model, te se on može posmatrati kao poseban slučaj dvoparametarskog modela.

Kakve posledice ima uvođenje drugog parametra, diskriminativnosti, na izgled KKS? Diskriminativnost definiše **nagib** krive u tački preloma tako što veći nagib odgovara većoj diskriminativnosti (An & Yung, 2014; Baker, 2001; Drasgow & Hulin, 1990; Fajgelj, 2009; Linacre, 2022). Ukoliko zamislimo idealno diskriminativnu krivu, koja savršeno razlikuje ispitanike – mogli bismo je prikazati kao potpuno vertikalnu liniju. Svi ispitanici čiji je nivo osobine ispod nivoa stavke imaju 0% verovatnoće da je tačno reše, dok svi ispitanici čiji je nivo sposobnosti iznad nivoa stavke imaju 100% verovatnoću da je reše. Sa druge strane, drugi ekstremni slučaj je savršeno nediskriminativna stavka, koja uopšte ne razlikuje ispitanike. Ovakva stavka mogla bi se predstaviti potpuno horizontalnom linijom gde svi ispitanici, bez obzira na nivo osobine, imaju jednaku verovatnoću da reše stavku. Na primer, ukoliko u testu znanja postavimo besmisleno pitanje sa 5 ponuđenih odgovora – svi ispitanici će imati jednaku verovatnoću da ga tačno reše (20%), bilo da znaju puno ili malo iz date oblasti.

Slika 8.6. prikazuje tri stavke koje se razlikuju po diskriminativnosti, odnosno nagibu. Najdiskriminativnija je stavka čija je KSS prikazana ljubičastom bojom i koja ima najveći nagib, dok je najmanje diskriminativna stavka sa najmanjim nagibom KKS, koja je na slici prikazana zelenom bojom.

Na ovoj slici možemo uočiti još jednu zanimljivu stvar. Dok se diskriminativnosti ove tri stavke međusobno razlikuju – njihove težine su potpuno iste. Sve tri stavke imaju verovatnoću rešavanja od 50% na nivou osobine 0, odnosno sve tri stavke svojom težinom odgovaraju prosečnom nivou sposobnosti ispitanika. Slika 8.5, ukoliko je ponovo pogledamo, pokazuje da je ovde situacija obrnuta. Stavke se razlikuju po težini, ali sve imaju isti nagib, odnosno diskriminativnost im je ista. Na osnovu ovoga možemo zaključiti da su parametri težine i diskriminativnosti stavke međusobno nezavisni (Baker, 2001).



Slika 8.6

8.3.3) Troparametarski modeli

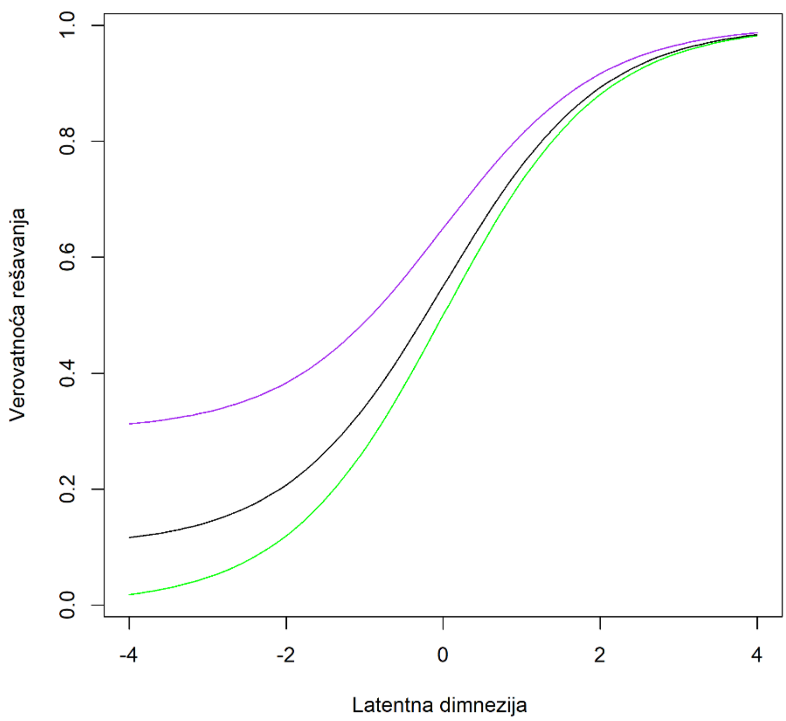
U troparametarskim logističkim modelima (3PL), pored parametara **težine** i **diskriminativnosti**, dodaje se i treći parametar – **pogađanja**. Naime, stavke se razlikuju prema tome sa kojom verovatnoćom ih rešava ispitanik s minimalnim nivoom sposobnosti/znanja/osobine. Matematički, ovaj model se prikazuje sledećom formulom:

$$P_i(\theta) = c_i + (1 - c_i) * \frac{e^{a_i * (\theta - b_i)}}{1 + e^{a_i * (\theta - b_i)}}$$

gde je c_i parametar pogađanja, odnosno verovatnoća tačnog rešavanja stavke i sa minimalnim nivoom osobine, a sve druge oznake iste kao kod dvoparametarskih modela. Dakle, bez obzira na nivo osobine, svi ispitanici imaju istu „početnu" verovatnoću c da tačno odgovore na stavku, dok je preostala verovatnoća do 100% distribuirana u skladu sa dvoparametarskim modelom. Vrednosti parametra pogađanja, logično, mogu se kretati od nule, kada nema slučajnog pogađanja, do 1, kada je verovatnoća pogađanja 100%. Ipak, u praksi se kao gornja prihvatljiva granica parametra pogađanja najčešće uzima vrednost od .35 (Baker, 2001). I jednoparametarski i dvoparametarski modeli su posebni slučajevi troparametarskog modela – fiksiranjem verovatnoće pogađanja za sve ajteme na c = 0 dobijamo dvoparametarski model, a dodatnim fiksiranjem i diskriminativnosti na a = 1 dobijamo jednoparametarski.

Svakako, parametar pogađanja najveću relevantnost ima kod stavki sa ponuđenim odgovorima (jer je u slučaju stavki otvorenog tipa ova verovatnoća po pravilu minimalna do nepostojeća). Ukoliko imamo 5 ponuđenih odgovora, obično je verovatnoća da ispitanik koji ne zna tačan odgovor tačno reši stavku 20%. Naravno, ova verovatnoća može biti i veća ukoliko su neki ponuđeni odgovori loše formulisani, te ih ispitanici mogu odbaciti i bez poznavanja tačnog odgovora, ili pak manja ukoliko među ponuđenim odgovorima postoji dobar atraktor koji „odvlači" od tačnog odgovora one ispitanike bez traženog znanja.

Kako grafički prikazujemo parametar pogađanja? Već smo rekli da je to verovatnoća da ispitanici tačno odgovore na stavku čak i kada ništa ne znaju, te je parametar pogađanja prikazan kao tačka u kojoj KKS preseca *y* osu, odnosno **intercept** (osečak) na y osi. Slika 8.7 prikazuje tri stavke međusobno različite po parametru pogađanja – najmanju verovatnoću pogađanja ima stavka čija KKS preseca *y* osu u vrednosti 0 (zelena linija), zatim stavka čija KKS preseca *y* osu u vrednosti (okvirno) .10 (crna linija), a najveću stavka čija KKS počinje na oko .30 (ljubičasta linija).



Slika 8.7

Možemo primetiti i da se ove tri stavke, osim po vrednosti parametra pogađanja, razlikuju i po težini i po diskriminativnosti. Kako se parametar pogađanja povećava, tako se (po pravilu) i težina i diskriminativnost stavke smanjuju (Baker, 2001; Harvey & Hammer, 1999). Ovo je sasvim logično, jer bi dobra stavka trebalo da bude takva da odgovor na nju zavisi samo od znanja (sposobnosti, osobine) ispitanika, a ne od spoljašnjih faktora kao što je slučajno pogađanje. Zato je, naravno, u interesu istraživača da verovatnoću slučajnog pogađanja svede na minimum.

8.4) Pokazatelji fita modela

Karakteristične krive stavki koje smo prethodno prikazali predstavljaju teorijske modele, odnosno naša očekivanja o tome kako podaci treba da izgledaju. One pretpostavljaju da za svakog ispitanika tačno možemo da izračunamo verovatnoću sa kojom će rešiti određenu stavku. Ovo, naravno, odgovara idealnoj situaciji gde se svi ispitanici i sve stavke ponašaju u skladu sa očekivanjima modela. To konkretno znači da, ukoliko stavke poređamo po težini, za svakog ispitanika imamo neprekinut niz tačnih odgovora (jedinica), nakon čega sledi neprekinut niz netačnih odgovora (nula). Drugim rečima, ispitanik je uspešno rešavao sve stavke koje su za njega bile dovoljno lake, a onda je došao do stavke preteške za njega, nakon čega više nije mogao da daje tačne odgovore na testu.

Realni podaci koje ćemo dobijati nikada neće u potpunosti odgovarati teorijskom modelu. Postoje dva osnovna vida odstupanja od modela – neočekivani neuspeh i neočekivani uspeh (Tabela 8.3) (Linacre, 2022). Neočekivani neuspeh se odnosi na situaciju kada je niz jedinica prekinut nulom, odnosno ispitanik nije odgovorio na stavku na koju bi bilo očekivano da je odgovorio. Kao najverovatniji razlog odstupanja u ovom slučaju uzima se nepažnja ispitanika. Sa druge strane, neočekivani uspeh je situacija kada je niz nula prekinut jedinicom, odnosno kada je ispitanik tačno rešio stavku za koju na osnovu njegovih drugih odgovora ne bismo očekivali da ima znanja/sposobnosti da je reši. U ovom slučaju, prepisivanje ili sreća se smatraju najverovatnijim razlozima odstupanja podataka od modela.

Tabela 8.3. Primer odstupanja podataka od modela za pojedinačnog ispitanika

Tip odgovora ispitanika	Primer podataka
Idealna situacija	111111110000000000
Neočekivani neuspeh	111011110000000000
Neočekivani uspeh	111111110000001000

Slika 8.8 prikazuje ove dve vrste odstupanja od modela za 20 ispitanika koji su radili test znanja iz psihologije. Neočekivani neuspesi, odnosno pogrešni odgovori tamo gde bismo očekivali da su ispitanici znali tačan odgovor obeleženi su simbolom @, dok su neočekivani uspesi označeni slovom A. Naravno, tačni odgovori predstavljeni su jedinicama, a netačni nulama.

```
150 +111111@1@111@A00A000
151 +1111111@11@@10000AA0
152 +1@11@11@1@1@1AA0AA00
153 +1111@111111@@000AA00
154 +11111@111@@1@AA00A00
155 +1@@11111@11@@0A0AAA0
156 +1@@11@1@11@11AA00A0A
157 +11111111@@1@@A0A0A00
158 +11111@111@1@@0AA0A00
159 +11111111@11@@0AA0000
160 +11111111@1@@1A000A00
161 +11111111@@111000A000
162 +11@11111@@@1100A0AA0
163 +11111@1@1@11@AA000A0
164 +111@1111@1@@10000AAA
165 +111111@@1@@1@AA0AA00
166 +11@11@111@@1@00AAA0A
167 +1111111@11@1@000A0A0
168 +11@11111@@11@A0A000A
169 +111111@@@1@@@AAA0A0A
```

Slika 8.8

Zajedničko ime za sva odstupanja podataka od modela je misfit (Linacre, 2022). U misfit ulaze i neočekivani uspesi, kao i neočekivani neuspesi ispitanika. Međutim, kada govorimo o vrstama misfita, češće govorimo o podeli na infit i autfit (eng. *outfit*). Infit se odnosi na odstupanja blizu ispitanikovog procenjenog nivoa sposobnosti, dok se autfit odnosi na odstupanja koja su daleko od ispitanikovog procenjenog nivoa sposobnosti. Na primer, ukoliko je ispitanik tačno rešio 15 stavki, ali je pogrešno odgovorio na 14. stavku po težini, ovo odstupanje je blizu njegovog nivoa procenjene sposobnosti, pa bismo ovakvo odstupanje nazivali infitom. Sa druge strane, ako je ovaj isti ispitanik pogrešio u odgovoru na 2. najlakšu stavku, greška se desila „daleko" od njegovog procenjenog nivoa sposobnosti, pa bismo je nazvali autfitom.

Ali šta ako je ispitanik pogrešno odgovorio na 8. po težini stavku u instrumentu? Da li ovu grešku treba smatrati infitom ili autfitom? Odnosno, šta je dovoljno „blizu", odnosno dovoljno „daleko" od ispitanikovog nivoa sposobnosti da bismo datu grešku označili kao infit ili autfit? Zapravo, svi neočekivani odgovori ispitanika ulaze u procenu kako infita, tako i autfita, ali su različito ponderisani. Ovo znači da što je greška bliža procenjenom nivou sposobnosti ispitanika, imaće veći udeo u računanju infita, a manji udeo u računanju autfita i obrnuto (Linacre, 2022). Ne treba zaboraviti da u IRT-u po pravilu možemo izračunati sve pokazatelje kako za ispitanike, tako i za stavke. To važi i za pokazatelje fita, odnosno misfita podataka, te i u slučaju računanja infita i autfita za stavke sve greške ulaze u računicu. Neočekivani odgovori onih ispitanika koji su po svojoj sposobnosti blizu težine stavke će biti više ponderisani pri računanju infita, a neočekivani odgovori ispitanika čiji je nivo sposobnosti udaljen od težine stavke će se više uzimati u obzir prilikom računanja autfita.

Infit i autfit se obično računaju kao prosečni kvadrat (eng. *mean-square, MNSQ*), odnosno χ^2 podeljen svojim stepenima slobode, tako da bi u slučaju dobrog poklapanja podataka sa modelom trebalo da iznose oko 1 (Linacre, 2022). Ako je vrednost fita veća od 1 – to ukazuje na nedovoljan fit modela (eng. *underfit*), odnosno na postojanje „šuma" u podacima, pa se ovaj fit ponekad naziva i „šumnim" (eng. *noisy*). Ako je vrednost fita manja od 1 – ona ukazuje na preterani fit modela (eng. *overfit*), što znači da se podaci ponašaju previše pravilno i predvidljivo. Ovakav fit se ponekad naziva i „prigušenim" (eng. *muted*) (Linacre, 2022).

Na nivou celog instrumenta, fit modela će obično biti veoma blizu idealne vrednosti od 1, ali određeni ajtemi mogu pokazivati veći ili manji misfit. Autori Winsteps programa navode da, u većini praktičnih situacija, misfit koji se kreće u opsegu od 0.5 do 1.5 ukazuje na ajteme produktivne za merenje (Linacre, 2022; Wright & Linacre, 1994). Misfit niži od 0.5 ukazuje na preveliku regularnost podataka, te može proizvesti nerealistično visoke vrednosti pouzdanosti i separacije (o kojima će biti više reči u nastavku teksta). On ukazuje na neproduktivne ajteme, odnosno ajteme koji ne donose previše novih informacija, ali oni ne degradiraju merni model kao takav. Slično tome, misfit između 1.5 i 2 se smatra neproduktivnim, ali ne i degradirajućim po kvalitet merenja, dok misfit veći od 2 narušava merni model jer odgovori na ajtemima u većoj meri odstupaju od očekivanja modela (Linacre, 2022; Wright & Linacre, 1994). Autori, međutim, navode i da je moguće postaviti strože kriterijume za fit modela, na primer, da se u slučaju kada je potrebna visoka preciznost dobrim fitom smatraju vrednosti misfita između 0.8 i 1.2 (Linacre, 2022; Wright & Linacre, 1994), a da u „redovnim" situacijama vrednosti od 0.7 do 1.3 ukazuju na dobar fit (Wright & Linacre, 1994), što prihvataju i drugi autori (de Ayala, 2009). U daljem tekstu i mi ćemo se služiti navedenim referentnim vrednostima od 0.7 do 1.3, ali je važno imati u vidu da one nisu apsolutne, te ih po potrebi treba prilagoditi konkretnom testu.

Pokazalo se da je visok misfit veća pretnja validnosti testa od niskog misfita, te ovo treba uzeti u obzir prilikom interpretiranja dobijenih vrednosti. Takođe, visoke vrednosti infita i autfita se ne interpretiraju na isti način i nemaju jednake posledice po kvalitet merenja. Naime, visok autfit pokazuje da se ajtem loše ponaša za one ispitanike čiji je nivo osobine daleko iznad ili ispod težine ajtema, odnosno za ispitanike kojima i nije primarno namenjen. Na primer, u slučaju teških ajtema visok autfit često može biti posledica slučajnih tačnih odgovora (pogađanja ili prepisivanja) malog broja ispitanika sa niskim skorovima, te ovo svakako treba proveriti pre donošenja konačnog suda o kvalitetu ajtema. Sa druge strane, loš infit ukazuje na to da se ajtem loše ponaša baš za one ispitanike čijem nivou sposobnosti bi trebalo da bude najbolje podešen, što je veći problem za validnost merenja (ali je ujedno i teže odrediti tačan uzrok ovakvog misfita) (Linacre, 2022).

Posmatrajući ajteme testa, dakle, vrednosti fita veće od 1.3 bismo smatrali da indikuju loše ajteme, koji potencijalno narušavaju kvalitet merenja, posebno ukoliko je u pitanju previsok infit. Vrednosti fita manje od 0.7 bi pak ukazivale na ajteme koji se previše regularno ponašaju, pa je moguće da su redundantni, ali ne narušavaju kvalitet merenja u jednakom stepenu. Samim tim, ajteme s prevelikim misfitom je preporučljivo izbaciti iz instrumenta, dok ajteme koji imaju nizak misfit možemo i zadržati.

I pojedinačni ispitanici mogu imati previsoke ili preniske vrednosti fita. U slučaju visokog misfita, obično se radi o osobi koja je bila nepažljiva, koja je prepisivala ili je sklona davanju ekstremnih odgovora, dok nizak misfit ukazuje na osobu koja je bila oprezna u davanju odgovora ili sklona biranju srednjih kategorija. Iako (tipično) mali broj ispitanika sa lošim fitom neće imati osetan uticaj na fit celog modela, treba da budemo oprezni prilikom tumačenja procenjenog nivoa osobine datih ispitanika, jer oni nisu adekvatno procenjeni u okviru modela.

8.5) Dodatni pokazatelji karakteristični za IRT

Da bismo procenili kvalitet merenja nekog instrumenta, pored pokazatelja fita potrebno je uzeti u obzir još neke parametre karakterstične za IRT. Ovde ćemo opisati pojmove informativnosti, pouzdanosti i separacije, kao i način na koji se oni obično operacionalizuju i tumače.

8.5.1) Informativnost

Jedan od ključnih pojmova IRT-a jeste pojam informativnosti. Često se kaže da je informativnost u IRT-u analogna pouzdanosti u klasičnom modelu merenja, zato što nam ona govori o stepenu preciznosti merenja (An & Yung, 2014; Baker, 2001; Harvey & Hammer, 1999). Ono što, međutim, razlikuje informativnost od mera koje tipično dobijamo u okviru klasične testovne teorije jeste to što je informativnost *u funkciji* nivoa sposobnosti ispitanika (An & Yung, 2014; Baker, 2001; Fraley et al., 2000; Harvey & Hammer, 1999). Drugim rečima, jedna ista stavka neće biti jednako informativna za svakog ispitanika. Šta to znači?

Zamislimo da imamo jednu umereno laku stavku, na primer, čija je težina 6 jedinica sirovog skora. Ovu stavku je rešavalo dvoje ispitanika, ispitanik A i ispitanik B. Ispitanik A je na celom testu postigao ukupan skor od 5 poena, dok je ispitanik B postigao ukupan skor od 25 poena. Pretpostavimo da su oba ispitanika tačno odgovorila na stavku. Koliko je ova informacija za nas značajna u jednom, a koliko u drugom slučaju? Posmatrajući odnos težine stavke i sposobnosti ispitanika, nimalo nas ne iznenađuje što je ispitanik B, koji je postigao skor 25, uspešno rešio stavku. Njegov nivo sposobnosti daleko je iznad nivoa težine stavke, tako da nam činjenica da ju je tačno rešio malo pomaže da ga tačno pozicioniramo na dimenziji latentne osobine koju procenjujemo. U slučaju ispitanika B, dakle, ova stavka nosi malo informacija. Ispitanik A je, međutim, po nivou svoje sposobnosti blizu stavke koju je rešavao. U tom smislu, možemo zamisliti i scenario gde je on uspešno, ali i gde je neuspešno rešava. Zbog toga, za ispitanika A ista ova stavka nosi više informacija, jer nam je relevantnija za precizno određivanje njegovog položaja na dimenziji latentne osobine.

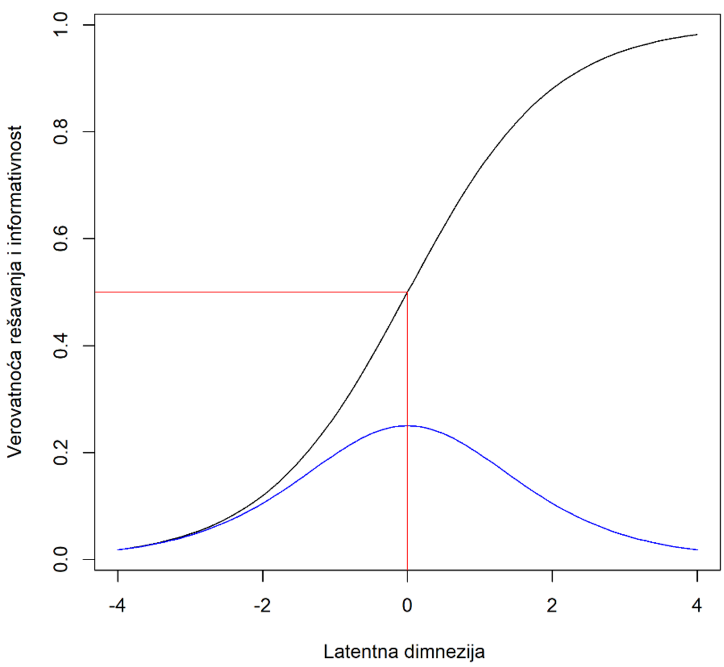
Možemo zaključiti da je informativnost stavke različita za različite ispitanike. Kao što smo već rekli – informativnost stavke je u funkciji nivoa sposobnosti ispitanika. Za koje ispitanike će onda stavka biti najinformativnija? Kakav treba da bude odnos nivoa sposobnosti ispitanika i težine stavke da bi nam činjenica da je ispitanik tačno ili netačno rešio stavku pružila najviše informacija o njegovom nivou sposobnosti? Najveću informativnost imamo kada je sposobnost ispitanika jednaka nivou težine stavke. Dakle, ne kada je ispitanik po svojoj sposobnosti daleko iznad ili ispod nivoa težine stavke, već onda kada su stavka i ispitanik „izjednačeni”. Kao što već znamo, ovo odgovara situaciji kada je verovatnoća tačnog rešavanja stavke 50%.

Upravo ovo možemo videti i u formuli za izračunavanje informativnosti u jednoparametarskom logističkom modelu:

$$I_i(\theta) = P_i(\theta)Q_i(\theta)$$

gde je I_i informativnost pojedinačne stavke i, θ kontinuum sposobnosti ispitanika, a P_i i Q_i verovatnoće tačnog, odnosno netačnog odgovora na stavku. Možemo primetiti da su verovatnoća tačnog i netačnog odgovora, a samim tim i informativnost stavke u funkciji sposobnosti ispitanika.

Slika 8.9 daje grafički prikaz odnosa verovatnoće tačnog rešavanja stavke (prikazane crnom bojom) i njene informativnosti (prikazane plavom bojom). Plava linija koja se odnosi na informativnost ajtema naziva se kriva informativnosti ajtema, odnosno stavke (eng. *item information curve, IIC*) (Baker, 2001; Fraley et al., 2000; Linacre, 2022). Možemo videti da je informativnost maksimalna za onaj nivo sposobnosti koji stavka procenjuje, a da zatim opada kako idemo ka višim, odnosno nižim nivoima latentne osobine. Kada se veoma udaljimo od nivoa težine stavke, njena informativnost je praktično nulta.



Slika 8.9

U prethodnoj formuli kao jedini parametar uzeli smo težinu stavke (izraženu preko P i Q). Na informativnost može uticati još i diskriminativnost, i to tako što su diskriminativnije stavke ujedno i informativnije. Ovo je logično, jer diskriminativnost stavke ukazuje na njenu sposobnost da dobro razlikuje ispitanike sa niskim, odnosno visokim nivoom osobine. Samim tim, ukoliko je stavka diskriminativnija, to ona nosi više informacija koje su relevantne za pozicioniranje ispitanika na kontinuumu merene osobine (Baker, 2001; Fraley et al., 2000). Ukoliko u formulu uključimo i parametar diskriminativnosti, ona će izgledati ovako:

$$I_i(\theta) = a_i^2 P_i(\theta) Q_i(\theta)$$

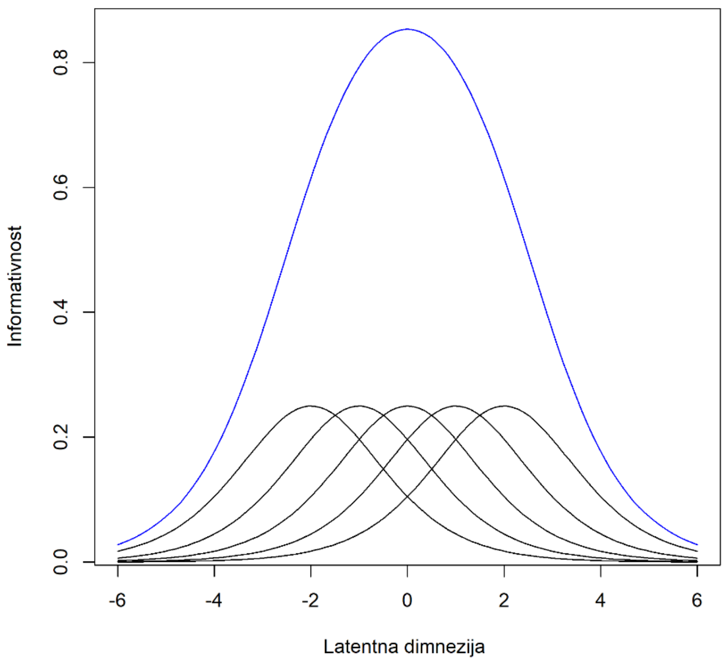
odnosno krive informativnosti ajtema sa većom diskriminativnošću će imati više „vrhove”, za razliku od kriva informativnosti ajtema niže diskriminativnosti, koji će biti „ravniji”.

Informativnost pojedinačnih stavki je aditivna, što znači da se informativnosti pojedinačnih stavki mogu sabirati (Baker, 2001; Fraley et al., 2000; Linacre, 2022). Tako je informativnost celog testa definisana kao suma informativnosti njegovih stavki, odnosno:

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta)$$

što se grafički može prikazati krivom informativnosti testa (eng. *test information curve, TIC*) (Baker, 2001; Fraley et al., 2000; Linacre, 2022).

Primećujemo da je i informativnost celog testa u funkciji nivoa sposobnosti ispitanika. To znači da, kao ni pojedinačne stavke, ni test nije jednako informativan za sve ispitanike. On će tipično biti najinformativniji za one ispitanike čiji su skorovi oko prosečnih vrednosti, a zatim će informativnost opadati kako idemo ka ekstremnijim skorovima, kako veoma niskim tako i veoma visokim. Ukoliko je distribucija skorova na testu normalna, informativnost će ravnomerno opadati i ka niskim i ka visokim skorovima, dok će u slučaju asimetričnih distribucija opadanje biti neravnomerno. Ukoliko test sadrži veći broj lakih ajtema, informativnost će biti veća za niske skorove, a ukoliko sadrži veći broj teških ajtema za visoke skorove. Slika 8.10 ilustruje odnos informativnosti pojedinačnih stavki (prikazanih crnom bojom) i informativnosti testa (prikazane plavom bojom), gde možemo primetiti da se sa povećanjem broja stavki načelno povećava i informativnost testa.



Slika 8.10

Iako bi bilo idealno postići visoku informativnost na svim nivoima merene osobine, ovo nije realno očekivanje od bilo kog testa. Svaki instrument namenjen je primeni na određenoj ciljnoj populaciji i potrebno je da bude informativan upravo za najveći broj pripadnika te populacije. Kako se većina fenomena kojima se bavimo normalno distribuira u populaciji, najveća informativnost nam je upravo i potrebna za skorove koji su blizu aritmetičke sredine, jer tek mali broj ispitanika postiže veoma niske ili veoma visoke skorove. Stoga, da bi test bio upotrebljiv, najvažnije je da kriva informativnosti testa pokazuje visoku informativnost na srednjim nivoima merenja (ili, u slučaju kriterijumskog ocenjivanja, na nivoima gde postoje prelomni skorovi), kao i da pokriva željeni opseg skorova (Linacre, 2022).

Slično kao i kod računanja pouzdanosti, informativnost stavke i testa računaju se na osnovu standardne greške merenja (kojoj su obrnuto proporcionalni), što nam ukazuje na to da se i standardna greška merenja u okviru IRT-a razlikuje na različitim nivoima merenja (Baker, 2001; Fraley et al., 2000; Linacre, 2022). Ukoliko uzmemo da su informativnost i standardna greška obrnuto srazmerni, odnosno da je $I(\theta) = \frac{1}{\sigma^2} = \frac{1}{SEm(\theta)^2}$, standardnu grešku merenja možemo izraziti preko sledeće formule (za dvoparametarski logistički model):

$$SEm(\theta) = \frac{1}{\sqrt{I(\theta)}} = \frac{1}{\sqrt{\sum_{i=1}^{N} a_i^2 P(\theta) Q(\theta)}}$$

Vidimo da će za one ispitanike blizu nivoa osobine koji stavka procenjuje (za koji postoji maksimalna informativnost), greška merenja biti veoma mala. Odnosno, kada je sposobnost ispitnika na nivou težine stavke, greška merenja dostiže svoju minimalnu vrednost. Kako se udaljavamo od ove vrednosti, greška merenja će postajati sve veća i veća, te će za ispitanike sa ekstremnim skorovima najčešće biti velika, odnosno nivo sposobnosti ovih ispitanika nećemo moći precizno da izmerimo našim instrumentom. Takođe, greška merenja se smanjuje sa porastom diskriminativnosti stavke.

8.5.2) Pouzdanost

Iako se u IRT-u informativnost smatra merom analognom pouzdanosti u klasičnom modelu, IRT takođe daje i mere pouzdanosti. Kao za mnoge druge pokazatelje, dobijamo mere pouzdanosti i za ispitanike i za stavke. Pouzdnost u IRT-u se pre svega odnosi na to koliko je verovatno da bismo u narednim testiranjima dobili iste procene težine stavki, odnosno sposobnosti ispitanika. Program Winsteps daje donju i gornju granicu pouzdanosti – donja se naziva stvarnom pouzdanošću (eng. *real reliability*), dok se gornja nazva pouzdanošću modela (eng. *model reliability*), a očekivanje je da se prava pouzdanost (eng. *true reliability*) nalazi negde između ove dve vrednosti (Linacre, 2022). IRT pouzdanost se ponekad naziva i separacionom pouzdanošću, zbog njene bliske povezanosti sa merom separacije, odnosno zbog načina na koji se u okviru IRT-a praktično koristi informacija o pouzdanosti, o čemu će biti više reči u narednom odeljku.

Interpretacija pouzdanosti za ispitanike je ekvivalentna standardnoj interpretaciji pouzdanosti testa (Linacre, 2022). Zapravo, kao što smo već pominjali, ni u klasičnoj testovnoj teoriji pouzdanost se ne smatra „trajnim" svojstvom testa, već ona zavisi od uzorka kom je test zadat (DeVellis, 2006; Fajgelj, 2009; Linacre, 2022), te je naziv pouzdanost za ispitanike lako razumljiv. Dakle, IRT pouzdanost za ispitanike nam govori koliko smo precizno uspeli da procenimo skorove ispitanika na osnovu uzorka ajtema koje smo im zadali. Iako je način računanja nešto drugačiji, u slučaju kada nema ekstremnih ispitanika (sa minimalnim ili maksimalnim mogućim skorovima), vrednost pouzdanosti za ispitanike najčešće će se malo razlikovati od vrednosti Kronbahove alfe, i to obično tako što će potcenjivati pouzdanost dobijenu u okviru klasične testovne teorije. U skladu sa uobičajenim preporukama, vrednosti preko .80 se smatraju dobrim, a preko .90 odličnim.

Pouzdanost za stavke pak predstavlja meru karakterističnu samo za IRT (Linacre, 2022). Njeno tumačenje je analogno tumačenju koeficijenta pouzdanosti za ispitanike. Pouzdanost za stavke nam govori koliko smo precizno uspeli da procenimo težinu stavki na osnovu skorova uzorka ispitanika koji su radili dati test. Kako bismo precizno locirali položaj ajtema na latentnoj dimenziji koju bi trebalo da procenjujemo, potreban nam je dovoljno veliki i varijabilan uzorak. Stoga, niske vrednosti pouzdanosti obično ukazuju na to da je uzorak bio mali ili selekcionisan – te nije obuhvatio kompletan raspon mogućih skorova na latentnoj varijabli (Linacre, 2022). Po pravilu, uzorak ispitanika biće veći (nekada i više puta) od uzorka stavki, te možemo očekivati da ćemo težinu stavki proceniti veoma precizno. U skladu sa tim, preporučene vrednosti pouzdanosti za stavke su preko .90.

8.5.3) Separacija

U poglavlju o diskriminativnosti testa govorili smo o odnosu pouzdanosti i diskriminativnosti. Tada smo rekli da stvarna diskriminativnost testa zavisi od stepena preciznosti merenja. Ukoliko je preciznost visoka, onda ćemo moći da razlikujemo i skorove koji su veoma međusobno bliski, a ukoliko je preciznost niska, moći ćemo da razlikujemo samo međusobno prilično udaljene skorove. Ova osnovna ideja prisutna je u meri koju u okviru IRT-a nazivamo separacijom (Linacre, 2022).

Separaciju možemo posmatrati i kao odnos „prave" varijanse i varijanse greške (Linacre, 2022). Što je udeo prave (pouzdane) varijanse u ukupnim skorovima veći, to je udeo greške manji, a samim tim će i separacija biti veća. U IRT-u, separacija se definiše kao funkcija pouzdanosti, prema formuli:

$$SEP = \sqrt{\frac{r_{nn}}{1 - r_{nn}}}$$

gde je r_{nn} pouzdanost testa za ispitanike ili za stavke.

U slučaju kada je pouzdanost instrumenta .80, separacija će iznositi $SEP = \sqrt{\frac{r_{nn}}{1-r_{nn}}} = \sqrt{\frac{.80}{1-.80}} = \sqrt{\frac{.80}{.20}} = \sqrt{4} = 2$. Za pouzdanost .90, možemo na isti način izračunati da će separacija biti jednaka 3. Drugim rečima, preporuka da pouzdanost za ispitanike treba da bude veća od .80, a za stavke od .90 istovremeno znači da je poželjno da separacija za ispitanike bude veća od 2, a separacija za stavke preko 3.

Koje je praktično značenje separacije? Separacija nam govori o broju statistički različitih nivoa osobine, odnosno stratuma, koje možemo razlikovati na osnovu datog testa. Ovo važi pod pretpostavkom da populacija ima istu standardnu grešku merenja kao i uzorak i da su ekstremne vrednosti posledica slučajne greške (a ne pravih skorova). Ukoliko pak smatramo da ekstremni skorovi odražavaju prave skorove ispitanika, broj stratuma dobijamo primenom korekcije $Strata = \frac{(4*SEP+1)}{3}$ na dobijenu vrednost separacije (Linacre, 2022).

Separacija/broj stratuma za ispitanike ukazuje na broj distinktivnih kategorija ispitanika prema njihovom nivou osobine koje možemo razlikovati u našem uzorku. Separacija/broj stratuma za stavke daje informaciju o broju kategorija stavki različitih po težini koje možemo identifikovati. Tako, na primer, separacija od četiri za ispitanike ukazuje na to da možemo razlikovati četiri grupe ispitanika koji postižu 1) veoma niske, 2) umereno niske, 3) umereno visoke i 4) veoma visoke skorove na testu. Isto tako, separacija od deset za stavke govori da se stavke koje smo zadali ispitanicima mogu grupisati u deset nivoa težine, od najlakših (u prvom) do najtežih (u poslednjem stratumu).

8.6) IRT u programu Winsteps/Ministep

Za razliku od klasičnog ili Gutmanovog modela, IRT nije implementiran u IBM SPSS statistički softver, što znači da pokazatelje karakteristične za IRT ne možemo dobiti u okviru standardnih SPSS analiza i ispisa. Za ilustraciju IRT analiza koristićemo program Winsteps, odnosno njegovu besplatnu verziju Ministep (Linacre, 2022). Winsteps/Ministep ima nešto skromniji korisnički interfejs u poređenju sa SPSS-om, ali je relativno jednostavan za korišćenje. Da bismo pokrenuli analizu, potrebno je da pravilno definišemo tzv. kontrolni fajl (eng. *control file*). Nakon toga, pokrećemo analizu i biramo neki od ispisa koji će nam dati željene pokazatelje.

Važno je još jednom napomenuti da Winsteps počiva na jednoparametarskom IRT modelu, što znači da polazi od početne pretpostavke da se stavke međusobno razlikuju samo po težini. Empirijski, stavke će se, osim po težini, razlikovati i po diskriminativnosti, odnosno nagibu. Winsteps pruža mogućnost post-hoc procene diskriminativnosti stavki, pri čemu pretpostavljena vrednost iznosi 1 za sve stavke (niže vrednosti ukazuju na lošiju, a više na bolju diskriminativnost od očekivane) (Linacre, 2022), ali se time nećemo detaljnije baviti u ovom tekstu.

Kontrolni fajl u Winstepsu je tekstualni, najčešće .txt fajl, koji se može podeliti na tri celine, redom: kontrolni fajl u užem smislu, nazive varijabli (*labels*) i podatke.

8.6.1) Podešavanje kontrolnog fajla – podaci

Iako se podaci nalaze na samom kraju kontrolnog fajla, počećemo objašnjenje od njih, kako bismo lakše razumeli na šta se odnose druga podešavanja kontrolnog fajla. Podaci na kojima obično radimo sačuvani su u nekoj vrsti tabele. Najčešće, to je SPSS fajl sa podacima, čiji *Data View* u redovima sadrži ispitanike, a u kolonama varijable. Slično tome, ukoliko podatke sačuvamo (ili izvezemo iz nekog onlajn upitnika) u Excel-u – ponovo ćemo imati tabelu čiji redovi predstavljaju ispitanike, a kolone varijable. Međutim, podatke je moguće sačuvati i u tekstualnom, .txt fajlu. I u ovom slučaju ispitanici će biti u redovima, a varijable u kolonama, ali nam ni redovi ni kolone neće biti „označeni" u samom programu.

Srećom, konvertovanje fajlova iz jednog formata u drugi je jednostavno i brzo, tako da bez obzira na to u kom formatu su nam podaci inicijalno sačuvani imamo mogućnost da ih pretvorimo u format potreban za bilo koju željenu analizu/program. Dodatak 2 prikazuje kako fajl sačuvan u SPSS .sav formatu možemo sačuvati kao .txt fajl potreban za IRT analizu u Winstepsu.

Najjednostavniji način organizovanja podataka u tekstualnom fajlu jeste da svaka varijabla ima fiksiranu širinu, odnosno da uvek zauzima jednak broj kolona. Na primer, ukoliko želimo da unesemo redni broj ispitanika, treba unapred da odredimo koliko kolona će nam biti potrebno za ovo. Pretpostavimo da imamo manje od 1000 ispitanika, te da će redni brojevi imati vrednost od 1 do 999. Kako je najveći broj trocifren – potrebne su nam tri kolone da bismo mogli da napišemo redne brojeve svih ispitanika u okviru ove varijable. Kada unosimo redni broj koji ima manje od tri cifre, obično ćemo početne kolone ostaviti praznim, kako bi podaci bili adekvatno „poravnati". Na primer, ako unosimo broj 1, ostavićemo dve prazne kolone ispred kolone u koju upisujemo broj 1, kako bi varijabla *redni broj* ukupno zauzimala 3 kolone. Ili, ukoliko unosimo redni broj 78, čije cifre zauzimaju dve kolone, ostavićemo jednu praznu kolonu ispred njega, ponovo kako bi ukupan broj kolona koje ovaj redni broj zauzima bio jednak 3. Konačno, ukoliko unosimo redni broj 225, nema potrebe da ostavljamo bilo koju kolonu praznu, jer smo već popunili sve tri kolone, koliko je predviđno da redni broj zauzima. Tekstualni fajl sa podacima u kome se nalazi samo varijabla *redni broj*, mogao bi izgledati ovako:

1
3
78
6
124
55
225
17

Vidimo da su svi brojevi poravnati po desnoj ivici i da zauzimaju ukupno tri kolone, ukoliko brojimo od početka reda u fajlu. Ukoliko povučemo zamišljene vertikalne linije koje odvajaju kolone, možemo jasnije videti da svaka cifra zauzima jednu kolonu, da varijabla *redni broj* zauzima ukupno tri kolone, kao i da su za jednocifrene i dvocifrene brojeve neke kolone prazne:

		1
		3
	7	8
		6
1	2	4
	5	5
2	2	5
	1	7

Ukoliko bismo sada hteli da dopunimo fajl sa podacima dodatnim varijablama, za informaciju o polu, na primer, bila bi nam dovoljna jedna kolona, jer ćemo pol označiti jednom cifrom (npr., 0 i 1 ili 1 i 2). Ukoliko želimo da unesemo godine ispitanika, za to ćemo najverovatnije predvideti dve kolone. Za odgovore na pojedinačne stavke, u najvećem broju slučajeva, dovoljna je samo jedna kolona (bilo da su odgovori 0 i 1 ili 1, 2, 3, 4 ili 5 ili nešto treće). Dopunjen tekstualni fajl sa podacima bi izgledao ovako (uneli smo odgovore na 5 binarnih ajtema):

		1	1	2	2	0	1	1	1	1
		3	2	2	5	1	1	0	1	1
	7	8	1	1	8	0	0	1	0	1
		6	1	3	2	1	0	0	1	1
1	2	4	2	2	7	1	1	1	1	1
	5	5	2	2	1	0	1	0	1	1
2	2	5	2	1	9	1	0	1	0	0
	1	7	1	2	4	0	0	1	1	0

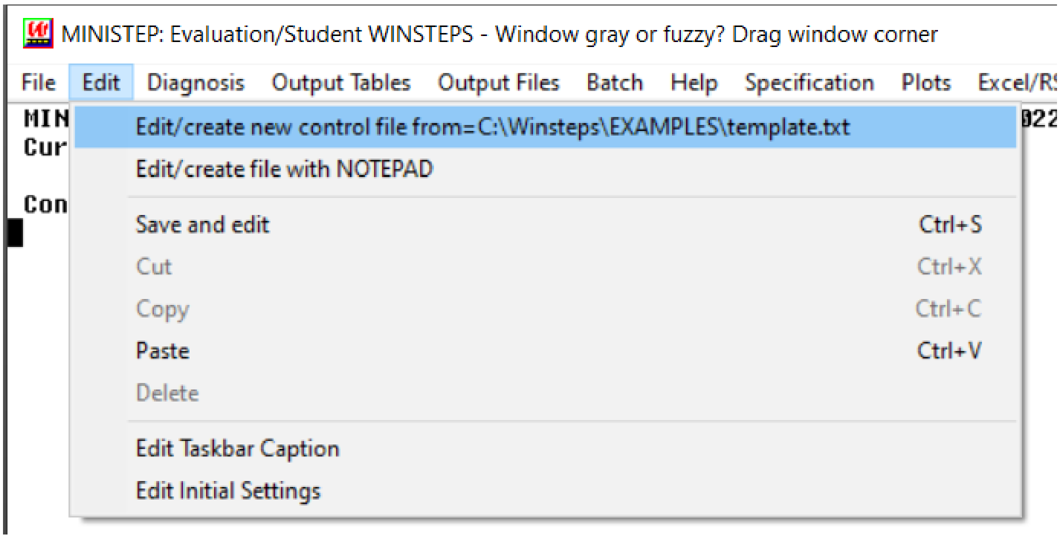
Naši podaci zauzimaju ukupno 11 kolona – 3 kolone odlaze na redni broj, 1 na pol ispitanika, 2 na njegov uzrast i 5 na odgovore ispitanika na stavke u zamišljenom testu. Ukoliko bismo ih prikazali bez zamišljenih vertikalnih linija, podaci bi izgledali ovako:

112201111
322511011
7811800101
613210011
12422711111
5522101011
22521910100
1712400110

Primećujemo da neke varijable zauzimaju veći broj kolona, dok neke zauzimaju samo po jednu. Ono što je važno jeste da je broj kolona koji varijabla zauzima **isti** za sve ispitanike, odnosno da je dužina varijabli „fiksirana” na određeni broj kolona. Samim tim, znamo da je informacija o polu za sve ispitanike u 4. koloni, kao i da je odgovor na prvu stavku u 7. koloni, a na petu stavku u 11. koloni.

8.6.2) Podešavanje kontrolnog fajla – kontrolni fajl u užem smislu

Kontrolni fajl u užem smislu definiše relevantne ulazne parametre za analizu, kao što su broj stavki, njihov položaj u delu sa podacima, položaj informacija o ispitaniku u podacima, mogući odgovori ispitanika itd. Prilikom konstruisanja kontrolnog fajla ne moramo krenuti „od nule”, već možemo iskoristiti šablon (eng. *template*) koji Winsteps nudi (Slika 8.11).



Slika 8.11

Na početku ovog Notepad fajla nalaze se sledeće tri linije teksta koje služe kao uputstvo korisnicima:

; this is a WINSTEPS specification control file template.

; Save it with your own name, e.g., control.txt

; a semi-colon means a comment: remove semi-colons as needed.

Dakle, Winsteps nas informiše da je ovo primer kontrolnog fajla koji možemo modifikovati i sačuvati pod novim imenom koje sami izaberemo. Takođe, veoma važna je i informacija o tome da znak ; odnosno tačka sa zapetom na početku reda znači da se taj red odnosi na komentar, odnosno na deo teksta koji neće biti pokrenut kao komanda. U SPSS-u se za komentare koristi znak *, dok je u Winstepsu to ;. U šablonu koji Winsteps nudi, velikom broju komandi prethodi tačka sa zapetom. Ovo je zbog toga što šablon u sebi sadrži različite moguće opcije analize koje korisnik može, ali i ne mora iskoristiti. Ove komande su samo neke od najčešće korišćenih, ali ne predstavljaju iscrpnu listu svih Winsteps komandi, te je moguće upisivati i dodatne komande u kontrolni fajl. Sve redove koji počinju ; program preskače, te ukoliko želimo da aktiviramo neku komandu koja je u šablon fajlu „skrivena” kao komentar, potrebno je da obrišemo ovaj znak. Naravno, ukoliko želimo da ostavljamo svoje komentare u fajlu – i to možemo učiniti uz stavljanje odgovarajućeg znaka na početak reda.

Na samom početku kontrolnog fajla nalazi se nekoliko linija kojima ne prethodi tačka sa zapetom, jer su ovo neophodna podešavanja da bi se analiza pokrenula. U pitanju su sledeći redovi:

NAME1	=	??	; column of start of person information
NAMLEN	=	??	; maximum length of person information
ITEM1	=	??	; column of first item-level response
NI	=	??	; number of items = test length
XWIDE	=	1	; number of columns per response
PERSON	=	Person	; Persons are called ...
ITEM	=	Item	; Items are called ...

Iza svake komande nalazi se komentar koji pojašnjava njeno značenje. Tako, NAME1 treba da sadrži informaciju o koloni u kojoj počinju informacije o ispitaniku, dok NAMLEN definiše dužinu (izraženu u kolonama) informacije o ispitaniku. Iza ovih komandi nalaze se znakovi pitanja, jer su ovo vrednosti koje korisnik programa treba da unese. Iako se, naravno, sve informacije u jednom redu odnose na ispitanika, pod „informacijama o ispitaniku" podrazumevaju se informacije koje **nisu** odgovori na stavke u instrumentu. To su najčešće neke identifikacione informacije kao što su redni broj, pol, godine i slično.

Ukoliko bismo kao podatke uzeli zamišljeni primer sa početka ovog odeljka, koju vrednost bismo upisali za NAME1, a koju za NAMLEN? Prva kolona u kojoj se nalaze informacije o ispitanku je kolona 1, jer u njoj počinje varijabla *redni broj* (koja zauzima prve tri kolone, odnosno kolone 1, 2 i 3). Iako je kod nekih ispitanika ova kolona prazna, to nije slučaj za sve ispitanike, tako da ne smemo izostaviti ovu kolonu. Stoga, kao vrednost NAME1 upisali bismo 1. Maksimalna dužina informacija o ispitaniku u slučaju našeg fajla sa podacima jeste 6. Kao što smo već rekli, redni broj zauzima tri kolone (kolone 1, 2 i 3), pol jednu kolonu (kolonu 4), a uzrast dve kolone (kolone 5 i 6), što u zbiru iznosi 3+1+2=6 kolona koje se odnose na informacije o ispitaniku (a koje nisu odgovori na stavke). Važno je naglasiti da za pokretanje analize Winsteps zahteva barem jednu kolonu informacija o ispitaniku, tako da je pored stavki neohodno imati barem neku dodatnu varijablu koja se odnosi na ispitanike u fajlu sa podacima (najčešće je to redni broj).

Sledeće dve komande tiču se položaja stavki u podacima. ITEM1 definiše prvu kolonu sa informacijama o stavkama, dok NI definiše broj stavki koje ulaze u analizu, odnosno dužinu testa. Opet, gledajući primer sa početka odeljka – koje vrednosti bismo upisali u ITEM1 i NI? Već smo videli da se odgovor na prvi ajtem za sve ispitanike nalazi u sedmoj koloni, te bi vrednost ITEM1 bila 7. Primećujemo da je ovo prva sledeća kolona nakon informacija o ispitaniku. Ovo će najčešće biti slučaj i sa drugim podacima, jer identifikacione informacije o ispitaniku obično prethode odgovorima na stavke u fajlu sa podacima. Kao vrednost NI upisaćemo 5, jer naš zamišljeni test ima ukupno pet stavki. Dakle, kao vrednost NI ne treba upisivati poslednju kolonu u kojoj se nalazi neki od ajtema (poslednja kolona u našem primeru je 11), već ukupan broj ajtema koje želimo da analiziramo. Program će sam „stati" kada dođe do poslednje kolone na osnovu informacije o početnoj koloni i ukupnom broju kolona koje treba analizirati.

Komanda XWIDE se odnosi na broj kolona koji zauzima svaka od varijabli, odnosno stavki. Podrazumevana vrednost za ovu komandu je 1, zato što je najčešće jedna kolona sasvim dovoljna za potencijalne odgovore ispitanika na pojedinačne stavke. Naravno, moguće je podesiti i da broj kolona bude veći od 1, ali za tim retko ima potrebe. Treba primetiti i da je za pravilno čitanje podataka neophodno da sve varijable koje ulaze u analizu zauzimaju isti broj kolona (bilo da je to jedna kolona ili neki veći broj).

Na kraju, komande PERSON i ITEM nam omogućavaju da definišemo sopstvene nazive za testirane osobe i ajteme, te ovde možemo upisati „osoba" ili „ispitanik" za PERSON, a „stavka",

„ajtem”, „pitanje” ili „zadatak” za ITEM, ili pak bilo koji drugi naziv koji odgovara prirodi naših podataka.

Pored ovih komandi sa početka fajla, postoji još nekoliko redova komandi kojima ne prethodi tačka sa zapetom, zbog toga što predstavljaju važna podešavanja analize koja je obično potrebno barem pregledati, ako ne i modifikovati, pre nego što se pristupi analizi.

Pri sredini kontrolnog fajla možemo pronaći komandu

CODES = "01" ; valid response codes

Ova komanda definiše validne kodove za odgovore ispitanika, odnosno moguće vrednosti u varijablama koje se odnose na stavke. Podrazumevana vrednost ovog polja je "01". Ovo znači da su mogući odgovori na stavke 0 i 1, što odgovara većini testova znanja i sposobnosti (odgovor je ili netačan ili tačan). Primećujemo da su mogući odgovori ispitanika navedeni u okviru znakova navoda, kao i da su napisani jedan za drugim bez razmaka. Takođe primećujemo da svaki odgovor zauzima samo jednu kolonu, s obzirom na to da smo prethodno definisali da je širina svake varijable jednaka jednoj koloni.

Ukoliko naš test ima drugačije ponuđene odgovore, potrebno je da promenimo vrednost CODES. Na primer, ukoliko je u pitanju test znanja koji dozvoljava i polovično tačne odgovore, onda bi mogući odgovori ispitanika bili 0 – netačno, 1 – delimično tačno i 2 – potpuno tačno, te bismo ih u Winstepsu definisali kao CODES = "012". Ukoliko smo pak zadali test ličnosti ili stava gde se odgovori daju na, na primer, petostepenoj Likertovoj skali, onda bi mogući odgovori ispitanika bili definisani kao CODES = "12345". Jako je važno da pravilno definišemo sve moguće odgovore na stavku, jer će samo ti odgovori biti uzeti u obzir u okviru analize, te u suprotnom nećemo dobiti tačne rezultate.

Pred kraj kontrolnog fajla (u užem smislu) nalazi se sledeći set komandi:

```
 UMEAN  =  0   ; item mean – default is 0.00
 USCALE =  1   ; measure units – default is 1.00
 UDECIM =  2   ; reported decimal places – default is 2
 MRANGE =  0   ; half-range on maps – default is 0 (auto-scaled)
```

UMEAN i USCALE definišu aritmetičku sredinu i standardnu devijaciju za stavke. Podrazumevane vrednosti (kao što piše u komentaru) su 0 i 1, što odgovara aritmetičkoj sredini i standardnoj devijaciji z-skorova. Najčešće ćemo i koristiti upravo ove vrednosti, jer su one lakše za interpretaciju od jedinica sirovih skorova na testu.

UDECIM definiše podrazumevani broj decimalnih mesta u ispisu analize. Najčešće će nam podrazumevane dve decimale pružati dovoljan stepen preciznosti, ali ukoliko želimo da dobijemo još preciznije vrednosti IRT pokazatelja, možemo podesiti broj decimala i na neki veći broj (npr., 3). MRANGE definiše središnju vrednost na mapama koje dobijamo u ispisu IRT analize. Ukoliko smo aritmetičku sredinu stavki podesili na 0, odnosno ukoliko koristimo z-skorove u analizi – najlogičnije je da upravo ova vrednost bude uzeta i kao središnja tačka grafičkih prikaza analize.

8.6.3) Podešavanje kontrolnog fajla – nazivi varijabli

Na samom kraju kontrolnog fajla u užem smislu nalazi se red sa komandom &END koja označava da se tu završavaju komande koje definišu analizu. Nakon ovog reda, slede još dva reda:

;Put item labels here for NI= lines
END LABELS

Prvi red počinje znakom ;, što znači da je u pitanju komentar, odnosno uputstvo za korisnika. Naime, ukoliko želimo da definišemo nazive varijabli u testu, to možemo učiniti tako što ćemo svaki naziv upisati u poseban red. Na primer, ukoliko imamo 5 stavki, treba da upišemo:

Stavka1
Stavka2
Stavka3
Stavka4
Stavka5

Broj redova sa nazivima varijabli mora odgovarati broju varijabli da bi ih program prepoznao. Upisivanje naziva varijabli nije obavezno, ali može olakšati čitanje ispisa. Na primer, pretpostavimo da je stavka 2 pokazala loše metrijske karakteristike i da smo odlučili da je izbacimo iz analize. Winsteps automatski imenuje varijable onim redom kojim se one nalaze u fajlu sa podacima, tako da prva stavka dobija ime I0001, sledeća I0002, pa I0003 itd. Ukoliko smo izbacili stavku 2 iz analize (i podataka), ime I0002 biće dodeljeno stavci 3, dok će I0003 biti naša „izvorna" stavka 4, što može biti zbunjujuće. Ukoliko, međutim, unesemo nazive stavki:

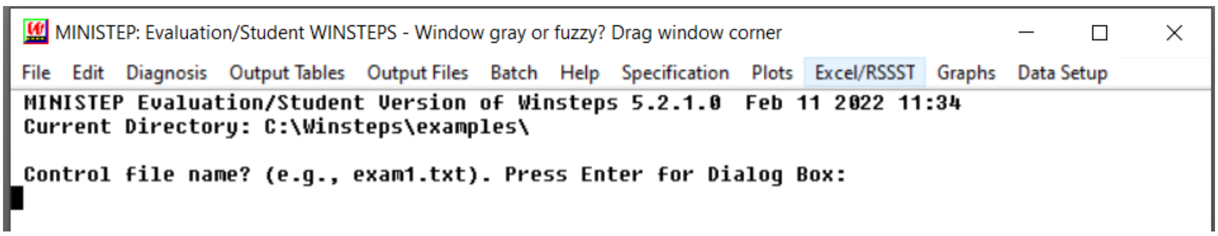
Stavka1
Stavka3
Stavka4
Stavka5

biće nam mnogo jednostavnije da pratimo kako se preostale stavke ponašaju u modelu, jer će nam njihov naziv ukazivati na početnu poziciju u testu, pre izbacivanja stavke 2.

Komanda END LABELS ukazuje na to da se deo sa nazivima varijabli završava. Nakon komande END LABELS ubacujemo podatke koje treba analizirati. Kao što je već rečeno, podaci treba da budu u tekstualnom formatu i sa fiksiranim brojem kolona za svaku varijablu.

8.6.4) Pokretanje analize

Nakon što smo definisali kontrolni fajl tako da odgovara prirodi podataka koje ćemo analizirati i sačuvali ga pod željenim imenom, možemo pokrenuti analizu u Winstepsu. Kada otvorimo Winsteps, program će nam sam odmah zatražiti kontrolni fajl (Slika 8.12).

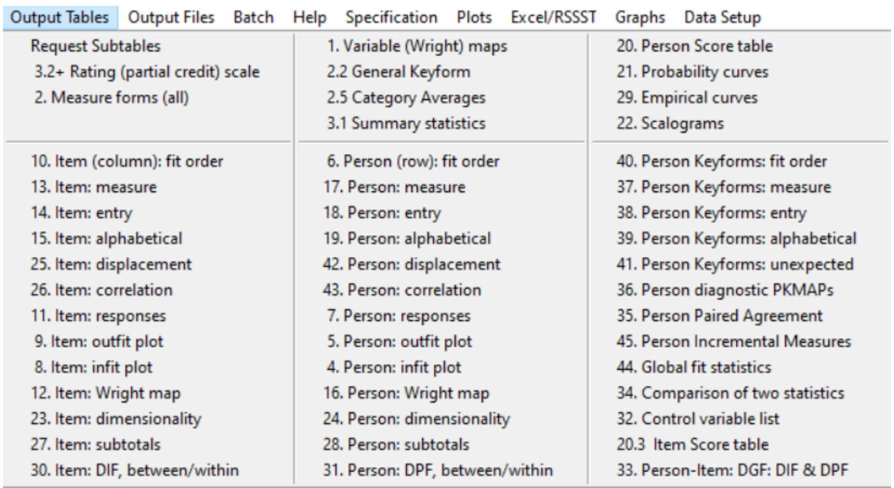


Slika 8.12

Pritiskom na *Enter* na tastaturi otvara nam se prozor preko koga možemo izabrati fajl za analizu. Zatim, pritiskom na *Enter* još dva puta, pokrećemo analizu i u glavnom prozoru programa dobijamo „preliminarni" ispis iz analize (Slika 8.13).

```
>=====================================<
|       7          -.20       -.0569      4     13*                                     |
>=====================================<
|       8          -.17       -.0481      4     13*                                     |
>=====================================<
|       9          -.14       -.0406      4     13*                                     |
>=====================================<
|      10          -.12       -.0342      4     13*                                     |
>=====================================<
|      11          -.10       -.0289      4     13*                                     |
-----------------------------------------------------------------------------------------
 Calculating JMLE Fit Statistics
>=====================================<
Time for estimation: 0:0:2.6
Output to C:\Winsteps\examples\ZOU905WS.TXT
KNOX CUBE TEST
-----------------------------------------------------------------------------------------
| KID       35 INPUT       35 MEASURED                    INFIT          OUTFIT    |
|           TOTAL      COUNT      MEASURE   REALSE     IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN        9.7       18.0         -.37     1.16       .99    -.2    .68    -.1|
| P.SD        2.4         .0         2.22      .34       .94    1.2   1.29     .7|
| REAL RMSE    1.21 TRUE SD     1.86  SEPARATION  1.55  KID    RELIABILITY  .70|
|-----------------------------------------------------------------------------------|
| TAP       18 INPUT       18 MEASURED                    INFIT          OUTFIT    |
|           TOTAL      COUNT      MEASURE   REALSE     IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN       18.9       35.0         -.76      .99       .96     .0    .68    -.1|
| P.SD       14.0         .0         4.26      .49       .28     .7    .58     .5|
| REAL RMSE    1.10 TRUE SD     4.12  SEPARATION  3.73  TAP    RELIABILITY  .93|
-----------------------------------------------------------------------------------------
Output written to C:\Winsteps\examples\ZOU905WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus
```

Slika 8.13



Slika 8.14

Za detaljniji uvid u rezultate možemo otvoriti neke od tabela koje se nude u okviru *Output Tables* padajućeg menija (Slika 8.14). Prva tabela koju obično želimo da pogledamo jeste tabela 3.1 *Summary statistics*, odnosno tabela sumarnih statistika koja daje opšti pregled modela – njegovog fita, pouzdanosti, separacije itd., kako za ispitanike, tako i za stavke. Ukoliko želimo da proverimo fit, odnosno misfit pojedinačnih stavki, otvorićemo i tabelu 10. *Item (column): fit order*, koja sortira stavke po stepenu misfita. Takođe, korisno je pogledati i mapu ispitanika i stavki, odnosno tabelu 12. *Item: map*. Naravno, u zavisnosti od potrebnih informacija, možemo koristiti i neke od drugih tabela koje su nam na raspolaganju.

Kao primer preko koga ćemo ilustrovati tumačenje rezultata IRT analize ponovo ćemo koristiti prvih 20 pitanja na testu znanja iz psihologije. Kako ćemo analize raditi u MINISTEP verziji Winsteps programa, koja ima ograničenje od najviše 25 ajtema i najviše 75 ispitanika, uključićemo nasumično izabranih 75 ispitanika koji su polagali test znanja iz psihologije.

8.6.5) Tumačenje rezultata

Za početak, otvorićemo tabelu 3.1. *Summary statistics*. Ovaj ispis, zapravo sadrži dve (ili tri) tabele, koje imaju iste redove i kolone, ali se prva (i druga) odnosi na ispitanike (Slika 8.15), dok se poslednja odnosi na stavke (Slika 8.16). Kao što smo već rekli, u IRT-u se gotovo svi pokazatelji računaju i za ispitanike i za stavke, zbog čega je i ispis ovako organizovan.

```
TABLE 3.1 Put your page heading here            ZOU855WS.TXT  Feb 11 2022 14:33
INPUT: 75 Person  20 Item  REPORTED: 75 Person  20 Item  2 CATS MINISTEP 5.2.1.0
--------------------------------------------------------------------------------

     SUMMARY OF 69 MEASURED (NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      13.9      20.0         .00    .63       1.00    .09    .93    .05 |
|  SEM        .5        .0         .15    .02        .03    .11    .05    .10 |
| P.SD       3.9        .0        1.27    .18        .21    .92    .44    .85 |
| S.SD       4.0        .0        1.28    .18        .22    .93    .44    .86 |
| MAX.      19.0      20.0        2.13   1.05       1.64   3.03   2.57   2.69 |
| MIN.       5.0      20.0       -2.51    .49        .64  -2.04    .23  -1.53 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .68 TRUE SD    1.08  SEPARATION  1.60  Person RELIABILITY  .72 |
|MODEL RMSE    .65 TRUE SD    1.09  SEPARATION  1.67  Person RELIABILITY  .74 |
| S.E. OF Person MEAN = .15                                                   |
-------------------------------------------------------------------------------
  MAXIMUM EXTREME SCORE:      6 Person 8.0%

     SUMMARY OF 75 MEASURED (EXTREME AND NON-EXTREME) Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      14.3      20.0         .27    .73                                 |
|  SEM        .5        .0         .18    .04                                 |
| P.SD       4.1        .0        1.53    .37                                 |
| S.SD       4.2        .0        1.54    .37                                 |
| MAX.      20.0      20.0        3.41   1.85                                 |
| MIN.       5.0      20.0       -2.51    .49                                 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .83 TRUE SD    1.29  SEPARATION  1.55  Person RELIABILITY  .70 |
|MODEL RMSE    .82 TRUE SD    1.30  SEPARATION  1.59  Person RELIABILITY  .72 |
| S.E. OF Person MEAN = .18                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .96 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82  SEM = 1.73 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .86
```

Slika 8.15

```
     SUMMARY OF 20 MEASURED (NON-EXTREME) Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      53.8      75.0       -1.21    .32       1.00    .06    .93   -.04 |
|  SEM       2.3        .0         .23    .02        .03    .24    .07    .24 |
| P.SD      10.0        .0        1.01    .07        .14   1.04    .31   1.05 |
| S.SD      10.3        .0        1.03    .07        .15   1.07    .32   1.08 |
| MAX.      71.0      75.0         .83    .53       1.26   1.93   1.60   1.58 |
| MIN.      30.0      75.0       -3.41    .28        .77  -1.86    .42  -1.71 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .34 TRUE SD     .95  SEPARATION  2.79  Item   RELIABILITY  .89 |
|MODEL RMSE    .33 TRUE SD     .95  SEPARATION  2.86  Item   RELIABILITY  .89 |
| S.E. OF Item MEAN = .23                                                     |
-------------------------------------------------------------------------------
Item RAW SCORE-TO-MEASURE CORRELATION = -.98 (approximate due to missing data)
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Slika 8.16

Pri samom vrhu stranice imamo informaciju o broju ispitanika i stavki koji su analizirani – *INPUT: 75 Person, 20 Item*. Važno je uvek obratiti pažnju na to da li se ovi brojevi poklapaju sa brojem ispitanika i stavki koje smo želeli da uključimo u analizu, jer u suprotnom to može biti indikator lošeg podešavanja kontrolnog fajla. U naslovu prve tabele možemo videti da je u analizu ušlo samo 69, umesto 75 ispitanika – „SUMMARY OF 69 MEASURED (NON-EXTREME) Person", dok u narednoj tabeli možemo videti pokazatelje ukoliko su uključeni ovi „ekstremni" ispitanici – „SUMMARY OF 75 MEASURED (EXTREME AND NON-EXTREME) Person". O čemu se radi, kakvi su to ekstremni ispitanici? U pitanju su ispitanici koji postižu neki od dva ekstremna skora – maksimalan ili minimalan. Drugim rečima, od 75 ispitanika u našem uzorku njih 6 je postiglo ili skor 0 ili skor 20. Ispod prve tabele imamo i precizniju informaciju da je svih 6 ispitanika (odnosno 8% celog uzorka) zapravo postiglo maksimalni skor od 20 tačnih odgovora – „MAXIMUM EXTREME SCORE: 6 Person". U okviru Winsteps verzije IRT-a smatra se da ovi ispitanici imaju beskonačno nisku, odnosno visoku poziciju na kontinuumu latentne dimenzije, zbog čega je njihovu poziciju, odnosno pravi skor nemoguće proceniti. Stoga, oni se po pravilu isključuju iz analize (mada postoji opcija u kontrolnom fajlu da se uključe i ekstremni ispitanici). U većini slučajeva, procenat ovih ispitanika u uzorku je veoma mali, te njihovo isključivanje neće značajno uticati na rezultate. U slučaju kada u uzorku nema ekstremnih ispitanika, umesto dve tabele u ispisu ćemo dobiti samo jednu (prvu prikazanu).

U gornjem delu sve tri tabele, nalaze se (u redovima) aritmetička sredina (MEAN), standardna greška merenja (SEM), procenjena populacijska, kao i uzoračka standardna devijacija (P.SD. i S.SD., redom), minumum (MIN) i maksimum (MAX) za nekoliko pokazatelja (u kolonama). Sada ćemo se posebno osvrnuti na sirovi skor (RAW SCORE), izražen u jedinicama samog instrumenta, položaj na latentnoj varijabli koju želimo da merimo (MEASURE), infit i autfit modela.

U tabeli za ispitanike (bez ekstremnih vrednosti) vidimo da je prosečan sirovi skor ispitanika bio 13.9 jedinica, odnosno ispitanici su u proseku tačno rešilli 13.9 zadataka od ukupno 20 (broj stavki vidimo u koloni COUNT). Standardna devijacija na uzorku je 4.0, dok su minimum i maksimum, redom, 5 i 19 tačno rešenih zadataka. Primećujemo da je maksimalni skor 19, a ne 20, jer smo isključili iz analize sve ispitanike koji su postizali ovaj skor. U tabeli za stavke vidimo da je u proseku svaku stavku tačno rešilo 53.8 ispitanika (od 75 koji su ušli u analizu, što je opet vidljivo i u koloni COUNT). Uzoračka standardna devijacija je 10.3, najlakšu stavku je tačno rešio 71 ispitanik, a najtežu 30.

Gledajući kolonu MEASURE, koja se odnosi na položaj ispitanika na latentnoj dimenziji, vidimo da je prosečan skor ispitanika 0. Ovo je u skladu sa tim kako smo podesili analizu – da aritmetička sredina latentne dimenzije bude 0, što piše i u samom dnu ispisa UMEAN=.000 USCALE=1.000. Standardna devijacija je 1.05, dok su minimum i maksimum -2.51 i 2.13. Sa druge strane, prosečna vrednost MEASURE je za stavke -1.21, što znači da je prosečni položaj stavki na latentnoj dimenziji -1.21 jedinica standardizovanog skora. Standardna devijacija za stavke je 1.03, dok su minimum i maksimum -3.41 i 0.28. Upoređivanjem vrednosti za ispitanike i stavke možemo zaključiti da je test za naše ispitanike bio lak. Pre svega, ovo vidimo na osnovu toga što su stavke u proseku za oko jednu standardnu devijaciju „ispod" nivoa znanja ispitanika, a uvid u raspone skorova ispitanika i težine stavki dodatno potkrepljuje ovaj zaključak (stavke su dosledno „ispod" ispitanika). U težinu testa dodatno ćemo se uveriti inspekcijom mape ispitanika i stavki (o čemu će biti više reči kasnije).

Ispod tabela za ispitanike i stavke, imamo i informaciju o korelaciji između sirovog skora i pozicije na latentnoj dimenziji (do koje smo došli skaliranjem odgovora), RAW SCORE-TO-MEASURE CORRELATION. Ova korelacija je data kako za ispitanike, tako i za stavke. U slučaju kada podaci dobro odgovaraju modelu, korelacije će biti blizu 1 za ispitanike, a blizu -1 za ajteme (zato što veći skor za ajteme označava manju težinu).

U desnom delu tabele nalaze se informacije o infitu i autfitu podataka. Svaka od mera pokazatelja fita ima dve kolone – MNSQ, odnosno prosečni kvadrat (eng. *mean square*) i ZSTD, odnosno Z standardizovani[32] skor (eng. *Z-standardized*). U delu o računanju fita modela rekli smo da je prosečni kvadrat zapravo χ^2 podeljen svojim stepenima slobode i da bi, u slučaju dobrog fita podataka modelu, vrednosti MNSQ trebalo da budu blizu 1. Vrednost Z skora ukazuje na značajnost prosečnog kvadrata, odnosno na verovatnoću da se data vrednost χ^2 dobije slučajno. Kao što je već poznato, apsolutne vrednosti veće od 1.96 ukazuju na značajnost na nivou .05, a veće od 2.58 na značajnost na nivou .01. Kako obe vrednosti obično vode istim zaključcima, za interpretaciju rezultata je najčešće dovoljno da protumačimo samo kolonu MNSQ.

Vidimo da su prosečni infit i autfit za ispitanike, redom, 1.00 i 0.93. Obe vrednosti su veoma blizu teorijski idealne vrednosti od 1, na osnovu čega zaključujemo da podaci dobro odgovaraju modelu. Isto tako, prosečna vrednost infita i autfita za stavke je 1.00 i 0.93, što ponovo ukazuje na dobar fit podataka. Uvid u minimalne i maksimalne vrednosti infita i autfita, međutim, pokazuje da postoje određeni ispitanici i određene stavke čiji fit izlazi iz okvira preporučenih vrednosti od 0.7 do 1.3. Iako loš fit za pojedinačne ispitanike znači da postoji mogućnost da njihovo znanje nismo adekvatno procenili, ono što nas, kada se bavimo konstrukcijom instrumenta i proverom njegovih metrijskih karakteristika, više zanima jeste fit stavki. Tabela fita stavki (koju ćemo prikazati kasnije) daje dodatne informacije o fitu pojedinačnih stavki u instrumentu.

U donjem desnom delu obe tabele nalaze se vrednosti pouzdanosti (RELIABILITY) i separacije (SEPARATION) za ispitanike i stavke. U tabeli su prikazana dva seta vrednosti, jedan koji je označen kao stvarni (REAL) i drugi koji je označen kao vrednost za model (MODEL). Kao što smo već rekli, „stvarna" vrednost predstavlja donju, a vrednost za model gornju granicu populacijske vrednosti. Po pravilu, razlika između ove dve vrednosti će biti veoma mala i neće bitno uticati na interpretaciju rezultata.

Pouzdanost za ispitanike je .72, odnosno .74. Ovo je niža vrednost od one koju smo dobili klasičnom analizom, što je posledica različitog načina računanja pouzdanosti u IRT-u (mada su rezultati dobijeni i na manjem uzorku, pa nisu direktno uporedivi). U skladu sa nešto nižom pouzdanošću, separacija je manja od 2 i iznosi 1.60, odnosno 1.67. Ove vrednosti možemo smatrati prihvatljivim, jer ukazuju da možemo razlikovati oko dva statistički različita nivoa postignuća ispitanika (ako bismo zaokruživali na ceo broj). Sa druge strane, pouzdanost za stavke je visoka i iznosi .89. Ova vrednost je pretežno posledica toga što smo imali dovoljno veliki uzorak ispitanika koji su postizali različite skorove na testu znanja. Separacija za stavke je takođe dobra, 2.79, odnosno 2.86, što nam govori da možemo razlikovati oko 3 statistički različita nivoa težine stavki u testu.

Konačno, u samom dnu tabele za ispitanike vidimo kakve bi se vrednosti dobile ukoliko bismo u analizu uključili i 6 ekstremnih ispitanika. Aritmetička sredina i uzoračka standardna devijacija bi bile 0.27 i 1.54, što se ne razlikuje previše od vrednosti dobijenih bez ekstremnih ispitanika. Činjenica da bi prosečan skor bio 0.27, dakle, veći od dobijenog, sama po sebi ukazuje na to da je većina ekstremnih ispitanika (ako ne i svi) postigla maksimalni skor od 20 poena. Naravno, ovo je u skladu sa informacijom iz ispisa da su svi ekstremni ispitanici postigli maksimalni skor.

[32] Termin *standardizovani* se ovde odnosi na to da je vrednost trasformisana u vrednost koja prati jediničnu normalnu distribuciju, mada se u literaturi ovi koeficijenti nazivaju i t-statisticima.

Vrednosti pouzdanosti i separacije su takođe veoma slične u slučaju da se i ekstremni ispitanici uključe u analizu i vode istim zaključcima o pouzdanosti instrumenta.

Imajući u vidu da su minimalne i maksimalne vrednosti infita i autfita za pojedinačne stavke izlazile izvan okvira preporučenih vrednosti, sada ćemo detaljnije pogledati fit pojedinačnih stavki. Ove informacije nalaze se u tabeli 10. *Item (column): fit order*. U ispisu dobijamo nekoliko tabela, ali ćemo se sada fokusirati na prvu (Slika 8.17). U ovoj tabeli u redovima se nalaze stavke, a u kolonama različiti pokazatelji, od kojih ćemo se sada usredsrediti na redni broj stavke (ENTRY NUMBER), njen sirovi skor (RAW SCORE), poziciju na latentnoj dimenziji (MEASURE), vrednosti infita i autfita, kao i korelaciju ajtema sa ukupnim skorom – dobijenu na podacima i očekivanu na osnovu modela (PTMEASUR-AL CORR. i EXP., redom). Pošto nismo upisali nazive stavki, Winsteps im je sam dodelio imena koja odgovaraju njihovom rednom broju, što se može videti u poslednjoj koloni tabele. Stavke su u ovoj tabeli, kao što je iz njenog imena očigledno, sortirane po stepenu (mis)fita, od stavki sa najvećim do stavki sa najmanjim vrednostima infita i autfita. Stavke u sredini tabele su one koje imaju najbolji fit (najbliži idealnoj vrednosti od 1).

Item STATISTICS: MISFIT ORDER

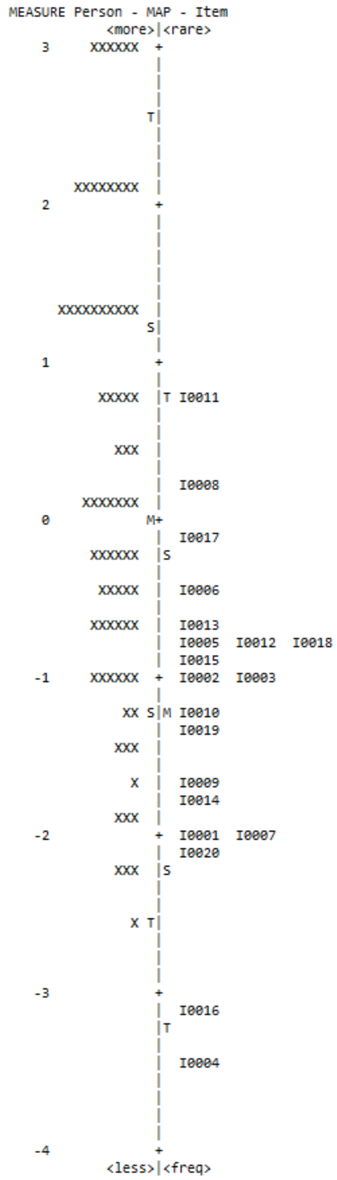
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
14	61	75	-1.78	.33	1.08	.53	1.60	1.41	A .32	.40	84.1	81.4	I0014
18	50	75	-.77	.29	1.18	1.38	1.37	1.57	B .39	.50	65.2	73.1	I0018
8	38	75	.18	.28	1.26	1.93	1.30	1.58	C .43	.56	65.2	72.6	I0008
13	49	75	-.68	.29	1.18	1.44	1.28	1.27	D .41	.51	66.7	72.7	I0013
12	50	75	-.77	.29	1.21	1.61	1.20	.90	E .39	.50	68.1	73.1	I0012
17	42	75	-.13	.28	1.08	.69	1.20	1.13	F .49	.55	69.6	71.9	I0017
11	30	75	.83	.29	1.07	.53	1.16	.75	G .55	.58	73.9	75.4	I0011
9	60	75	-1.68	.32	1.09	.58	1.08	.33	H .36	.42	79.7	80.5	I0009
1	63	75	-2.01	.35	1.06	.35	.72	-.54	I .38	.38	79.7	83.3	I0001
19	56	75	-1.29	.30	1.02	.22	.92	-.16	J .45	.45	73.9	77.0	I0019
10	55	75	-1.19	.30	.98	-.10	.81	-.61	j .49	.46	75.4	76.2	I0010
4	71	75	-3.41	.53	.96	.03	.42	-.47	i .29	.23	94.2	94.2	I0004
6	46	75	-.44	.28	.89	-.88	.80	-1.03	h .59	.53	75.4	72.0	I0006
20	64	75	-2.14	.36	.89	-.49	.61	-.78	g .44	.37	87.0	84.4	I0020
7	63	75	-2.01	.35	.87	-.64	.75	-.46	f .45	.38	85.5	83.3	I0007
16	70	75	-3.15	.49	.85	-.32	.63	-.25	e .34	.26	92.8	92.7	I0016
5	50	75	-.77	.29	.84	-1.27	.70	-1.40	d .59	.50	76.8	73.1	I0005
2	53	75	-1.02	.29	.83	-1.29	.76	-.95	c .57	.48	81.2	74.8	I0002
3	53	75	-1.02	.29	.83	-1.33	.66	-1.44	b .58	.48	78.3	74.8	I0003
15	52	75	-.93	.29	.77	-1.86	.62	-1.71	a .62	.49	82.6	74.1	I0015
MEAN	53.8	75.0	-1.21	.32	1.00	.06	.93	-.04			77.8	78.0	
P.SD	10.0	.0	1.01	.07	.14	1.04	.31	1.05			8.3	6.5	

Slika 8.17

Vidimo da su stavke 14 i 18 imale autfit veći od preporučenog, dok nijedna stavka nije imala preveliki infit (što govori pozitivno o validnosti merenja). Takođe, stavke 3 i 15 su imale nizak autfit, mada smo već rekli da takve stavke ne narušavaju kvalitet merenja. Previsok misfit obično ukazuje na ajteme koji narušavaju kvalitet merenja, te bi, uzimajući i pokazatelje iz klasične testovne teorije u obzir, trebalo razmotriti njihovo izbacivanje iz instrumenta (ukoliko smo u fazi konstrukcije i revizije instrumenta). Šta na osnovu vrednosti u ovoj tabeli možemo zaključiti o ovim ajtemima? Gledajući korelaciju sa ukupnim skorom, deluje da je ona, u slučaju ajtema sa nešto višim misfitom, niža od korelacija većine drugih ajtema sa ukupnim skorom. Težinu ovih ajtema možemo videti u koloni MEASURE i primećujemo da je, na primer, ajtem 14 bio lak ispitanicima, odnosno dovoljno je da ispitanik bude 1.78 standardnih devijacija ispod uzoračkog proseka da bi tačno rešio ovaj ajtem.

U dnu tabele imamo i prosečne vrednosti i standardne devijacije za sve pokazatelje. Ukoliko ih uporedimo sa vrednostima iz tabele koja se odnosi na stavke u okviru ispisa 3.1. *Summary statistics*, jasno je da se radi o identičnim vrednostima.

Konačno, da bismo što bolje sagledali odnos težine testa i znanja ispitanika, pogledaćemo mapu ispitanika i stavki koja se u Winstepsu nalazi u okviru ispisa 12. *Item: map*. I ovaj ispis daje dve mape, ali ćemo se baviti samo prvom jer ona ispitanike i stavke organizuje na intuitivno razumljiviji način (Slika 8.18).



Slika 8.18

Kao što već znamo, centralna linija predstavlja kontinuum znanja ispitanika i težine stavki. Sa leve strane se nalaze ispitanici pri čemu jedan znak X označava jednog ispitanika, na vrhu mape su ispitanici koji su bolje uradili test, a na dnu oni koji su ga lošije uradili. Sa desne strane su stavke, imenovane na osnovu svog rednog broja u podacima (jer im mi nismo uneli nazive). Na vrhu mape su stavke koje su ispitanici ređe tačno rešavali, odnosno teške stavke, a na dnu stavke koje su ispitanici češće tačno rešavali, odnosno lake stavke.

Pre svega, ukoliko uporedimo položaj aritmetičkih sredina za ispitanike i stavke (slovo M sa leve i desne strane centralne linije), potvrđujemo ranije donet zaključak o tome da je test bio lak za ispitanike (ili barem ovih 20 stavki testa). Dodatno, vidimo da stavke dobro pokrivaju niži deo raspona skorova ispitanika, ali da nema stavki na vrhu mape. Drugim rečima, nedostaju teške stavke koje bi dobro razlikovale ispitanike sa najvišim nivoom znanja. Stavka 4 (I0004) je problematična u pogledu težine. Naime, ova stavka je prelaka i nalazi se na delu kontinuuma merene dimenzije na kome nema ispitanika. To znači da ona ne pomaže u diskriminaciji ispitanika nijednog nivoa znanja koji je bio prisutan u uzorku. Najtežim se pokazao ajtem 11, a zatim ajtemi 8 i 17 i ovi ajtemi jesu diskriminativni jer postoje ispitanici sa datim nivoom znanja.

Primećujemo i da je vrednost na donjem kraju kontinuuma -4, što okvirno odgovara težini najlakšeg ajtema (-3.41 za ajtem 4, gde je on i pozicioniran na mapi), dok je vrednost na gornjem kraju kontinuma +3, što odgovara skorovima ekstremnih ispitanika. Sledeća grupa (neekstremnih) ispitanika nalazi se na nivou +2.13. Minimalne i maksimalne vrednosti merene dimenzije određuju se na osnovu minimalnih i maksimalnih skorova bilo za ispitanike, bilo za stavke, tako da se dobije potreban raspon za pozicioniranje svih ispitanika i svih stavki na mapi. Ove vrednosti smo takođe mogli videti u tabelama ispisa 3.1. *Summary statistics*. Položaj svake pojedinačne stavke na mapi odgovara njenoj težini koju, između ostalog, možemo pronaći i u tabeli 10. *Item (column): fit order*.

U idealnom slučaju, distribucije ispitanika i stavki trebalo bi da prate normalnu distribuciju i da budu „ogledalo" jedna drugoj – tako da najveći broj ajtema bude na onom nivou merenja gde se nalazi i najviše ispitanika. U slučaju ovog testa, takvo poklapanje ne postoji. Ukoliko bismo želeli da poboljšamo merenje, za početak bismo mogli izbaciti stavku 4 iz instrumenta, uzimajući u obzir njenu neadekvatnu težinu. Nakon izbacivanja ove stavke, imajući njen ekstremno nizak skor težine u vidu, izvesno je da bi se distribucije ispitanika i stavki približile jedna drugoj. Nakon ovoga, dalje bismo mogli modifikovati instrument kako bismo postigli optimalno merenje.

8.7) Rezime

IRT predstavlja teoriju merenja koja počiva na ideji da se verovatnoća da ispitanici određenog nivoa sposobnosti tačno rešavaju stavke instrumenta može iskazati određenim matematičkim funkcijama, koje se obično grafički prikazuju preko karakterističnih kriva stavki. U zavisnosti od broja parametara koji ove funkcije uključuju možemo razlikovati tri osnovne klase IRT modela: u jednoparametarskim modelima verovatnoća tačnog rešavanja stavke zavisi od sposobnosti ispitanika i težine stavke, u dvoparametarskim modelima se dodaje i parametar diskriminativnosti stavke, dok troparametarski modeli dodatno uključuju i parametar verovatnoće pogađanja za datu stavku. S obzirom na to da ovi modeli definišu očekivanja, neophodno je utvrditi u kojoj meri podaci odgovaraju očekivanjima, za šta se koriste pokazatelji misfita – infit i autfit.

IRT je takođe teorija zasnovana na skaliranju, odnosno na pozicioniranju i ispitanika i stavki na zajednički kontinuum latentne dimenzije. Grafiički prikaz ovog pozicioniranja čini mapu ispitanika i stavki koja na jednostavan način govori o sposobnosti ispitanika i težini stavki.

Jedan od osnovnih pokazatelja uspešnosti merenja u IRT-u je informativnost, koja predstavlja stepen informacija koji određena stavka ili test u celini nose za ispitanike različitih nivoa sposobnosti. Kao i informativnost, i standardna greška je u funkciji stepena izraženosti sposobnosti ispitanika, a kako su one obrnuto proporcionalne, najmanju grešku pravimo za ispitanike za koje je data stavka/test najinformativnija. Za razliku od klasične testovne teorije, pouzdanost u IRT-u pre svega služi računanju separacije, odnosno statistički različitih nivoa merenja koje smo uspeli da ostvarimo.

8.8) Preporučena literatura

An, X., & Yung, Y. (2014). Item Response Theory : What It Is and How You Can Use the IRT Procedure to Apply It. *SAS Institute Inc.*, 1–14. Retrieved from https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation.

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Fajgelj, S., & Kosanović, B. (2001). Nova i stara ajtem analiza – poređenje. *Psihologija*, *1–2*, 83–110.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, *12*(3), 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Linacre, J. M. (2022). *Winsteps® Rasch measurement computer program User's Guide. Version 5.2.1.* Portland, Oregon: Winsteps.com.

9) NORMIRANJE

Bez obzira na užu oblast psihologije kojom se bave, psiholozi će često kao deo svojih profesionalnih zadataka obavljati psihološko testiranje. Ovo testiranje, međutim, retko će biti na velikim (još ređe reprezentativnim) uzorcima ispitanika, a često će biti na malom broju pojedinaca ili čak samo na jednoj osobi. Pritom, od psihologa će se očekivati da protumače rezultate testiranja, a na osnovu datog tumačenja neretko se donose odluke koje imaju značajne posledice po testirane osobe – na primer, da li će osoba biti primljena na posao, upućena na lečenje, poslata na dodatnu obuku, usmerena na upis određene grupacije fakulteta itd. Kako bi psiholozi mogli da donose utemeljene zaključke o psihološkim karakteristikama testiranih osoba, potrebno je da se oslanjaju na norme.

Pretpostavimo da je psihološkinja kandidatkinji zadala test inteligencije i da je ona tačno uradila 24 zadatka. Šta na osnovu ove informacije možemo zaključiti o intelektualnim sposobnostima kandidatkinje? Zapravo, ništa, s obzirom na to da ne znamo ni koliko je zadataka imao test, ni (još važnije) kakvo je tipično postignuće ispitanika koji pripadaju istoj populaciji kao i kandidatkinja. Ukoliko bismo imali informaciju o tome da je prosečno postignuće na ovom testu 18 poena, zaključili bismo da je kandidatkinja natprosečno inteligentna, ali ne bismo znali tačno u kojoj meri je njeno postignuće natprosečno. Ukoliko bismo dodatno imali i informaciju o tome da je standardna devijacija na ovom testu 3 poena, onda bismo postignuće kandidatkinje mogli da ocenimo kao veoma visoko, jer se ona nalazi dve standardne devijacije iznad tipičnog postignuća za populaciju kojoj pripada. Ukoliko bismo njen skor izrazili kao IQ skor, rekli bismo da kandidatkinja ima IQ = 130 jedinica, odnosno da se ona nalazi u najboljih 2% populacije po intelektualnim sposobnostima. Ove informacije nam ne bi bile dostupne samo na osnovu testiranja date kandidatkinje, već bismo ih tipično pronašli u tabeli normi. S obzirom na izuzetan rezultat koji je postigla, psihološkinja bi (ukoliko ne postoje faktori koji ukazuju drugačije) u ovoj situaciji najverovatnije preporučila da kadidatkinja bude primljena na posao.

9.1) Pojam normiranja

Pored termina *normiranje*, u upotrebi su i pojmovi *standardizacija* ili *baždarenje testa*. I norma i standard imaju dva uobičajena značenja u svakodnevnom govoru – deskriptivno i preskriptivno. **Preskriptivno značenje** se odnosi na ono što se propisuje, što se obično smatra poželjnim, pa je u tom kontekstu norma ili standard ponašanje koje očekujemo da osoba ispoljava u nekoj situaciji (npr., rukovanje sa osobom koju smo tek upoznali). Kada se pojam norme upotrebljava u ovom značenju, on sa sobom po pravilu nosi i evaluativnu komponentu, te se odstupanje od norme smatra nepoželjnim. U psihometriji se, međutim, pojmovi norme i standarda shvataju **deskriptivno**, te oni predstavljaju uobičajeno, odnosno tipično ponašanje ispitanika (norma je, npr., da se tačno odgovori na 15 od 30 pitanja na testu opšte informisanosti) (Bukvić, 1996; Chadha, 2009; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Samim tim, odstupanje od norme ne nosi sa sobom nikakve evaluativne konotacije, budući da norma samo opisuje ono što je prosečno u određenoj populaciji.

Treba napomenuti da je pojam *standardizacije* nešto širi od pojma normiranja. Pod standardizacijom se tako, u psihometrijskom kontekstu, može misliti na standardizaciju uslova testiranja, načina zadavanja testa, ocenjivanja odgovora i sl., kao i tipičnog postignuća ispitanika (Fajgelj, 2009). Standardizacija se, pritom, ponekad koristi i da označi transformaciju sirovih skorova na standardnu, Z-skalu. Pojam *normiranje*, sa druge strane, ima veoma precizno značenje koje se odnosi na izradu normi, odnosno standarda sa kojima se mogu porediti rezultati pojedinačnih ispitanika.

Termin *baždarenje*, mimo upotrebe u psihometrijskom kontekstu, ima značenje kalibracije, proveravanja preciznosti mernih instrumenata i njihovo podešavanje kako bi bili ispravni (Bukvić, 1996). Na veoma sličan način možemo razumeti i baždarenje psihološkog mernog instrumenta – to je proces kojim proveravamo preciznost merenja, ali i vršimo izvesne korekcije kada je to potrebno. U daljem tekstu ćemo termine *baždarenje* i *normiranje* koristiti kao sinonime, dok termin *standardizacija* nećemo koristiti u ovom značenju.

9.2) Transformacije skorova pri normiranju

U statističkom smislu, proces normiranja predstavlja transformaciju sirovog skora u neki drugi skor koji je izražen u poznatoj skali, sa ciljem olakšavanja njegove interpretacije (Bukvić, 1996; Chadha, 2009; Fajgelj, 2009). Kao što smo na početku poglavlja naveli, sirovi skor kandidatkinje na testu inteligencije pretvorili smo u IQ skor, a takođe smo ga izrazili i kao percentilni rang. Ovim smo ujedno ilustrovali i dve osnovne vrste transformacija sirovih skorova – linearne i nelinearne (Bukvić, 1996).

Kod linearnih transformacija odnosi između sirovih skorova ostaju nepromenjeni nakon transformacije. Drugim rečima, znamo da je sirovi skor 21 za onoliko veći od sirovog skora 18 za koliko je 24 veće od 21. Samim tim, potpuno isto mora važiti i za odgovarajuće IQ skorove – skor od 115 IQ jedinica je onoliko veći od IQ skora 100 koliko je skor od 130 veći od skora 115. Ovakvi odnosi između skala su, naravno, posledica načina njihovog izračunavanja, gde između sirovih i standardizovanih skorova postoji linearan odnos, zbog čega ove transformacije i nazivamo linearnim.

Kod nelinearnih transformacija, kao što im ime govori, odnos između sirovih i transformisanih skorova neće biti linearan. To znači da razlika između transformisanih sirovih skorova 18 i 21 može biti manja ili veća od razlike između transformisanih skorova 21 i 24. Upravo ovo je slučaj kod, na primer, percentilnih rangova, budući da se oni ne dobijaju primenom bilo koje „formule", već zavise od proporcije ispitanika koja je ostvarila dati skor. Ukoliko se skorovi raspodeljuju (približno) normalno, možemo očekivati da će veći procenat ispitanika imati skorove između 18 i 21 nego što će ih biti između skorova 21 i 24. Ukoliko skorove na testu izrazimo kao percentilne rangove – primenili smo nelinearnu transformaciju.

9.2.1) Linearne transformacije

Teorijski, postoji beskonačno mnogo linearnih transformacija koje bismo mogli primeniti na nekim podacima. U praksi se, međutim, najčešće koristi ograničeni broj standardnih skala na koje se prevode sirovi skorovi ispitanika. Ove skale su izabrane zbog toga što poseduju određena poželjna svojstva koja ih čine lakim i intuitivnim za razumevanje i interpretaciju. Sada ćemo dati kratki pregled najčešće korišćenih standardnih skala (Bukvić, 1996; Chadha, 2009; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Kline, 2000).

Z-skala ima aritmetičku sredinu M = 0 i standardnu devijaciju σ = 1. Ovo je psiholozima možda i najpoznatija skala, jer se standardizovane varijable koriste u velikom broju statističkih i psihometrijskih analiza. Ona je i „najjednostavnija" skala, te vrlo dobro znamo kako da interpretiramo skorove ukoliko su dati na Z-skali. Sa druge strane, nepsiholozima je Z-skala uglavnom nepoznata, a razumevanje otežavaju i činjenice da su skorovi po pravilu iskazani kao decimalni brojevi i da je polovina skorova ispod nule, odnosno čine je negativni brojevi. Uprkos tome što je većini psihologa ovo lako razumljivo, laicima Z-skala nije „intuitivna", te se ona ne koristi u situacijama kada rezultate testiranja treba saopštiti nepsiholozima (što je najveći broj situacija testiranja).

IQ-skala ima aritmetičku sredinu M = 100 i standardnu devijaciju σ = 15. Ova skala veoma je dobro poznata kako psiholozima, tako i nepsiholozima. Sam naziv IQ predstavlja skraćenicu engleskog termina *intelligence quotient*, odnosno *količnik inteligencije* i potiče iz ranih istraživanja dečije inteligencije. U skladu sa tim, IQ skala se tipično koristi da izrazi postignuće ispitanika na testovima sposobnosti, a ponekad je i rezervisana za ukupni skor ispitanika dobijen na osnovu nekoliko (sub)testova intelektualnih sposobnosti. Ono što IQ skalu čini jednostavnom za razumevanje jeste upotreba celih, pozitivnih brojeva za izražavanje postignuća i njena opšta poznatost.

T-skala ima aritmetičku sredinu M = 50 i standardnu devijaciju σ = 10. Ona se obično vezuje za primenu u kliničkom kontekstu, jer se T-skala koristi u jednom od najčešće koriščenih kliničkih instrumenata – MMPI (eng. *Minnesota Multiphasic Personality Inventory*), ali nije ograničena samo na njega. Kao i kod IQ skale, pozitivna svojstva T-skale predstavlja to što su skorovi izraženi kao celi, pozitivni brojevi, te ih je lako intuitivno razumeti.

PISA skala ima aritmetičku sredinu M = 500 i standardnu devijaciju σ = 100. Osmišljena je i najčešće se koristi u kontekstu procene učeničkog postignuća. Njen raspon je tako osmišljen da razlika u jednom poenu na PISA skali odgovara veličini efekta izraženoj kao Koenovo d (eng. *Cohen's d*) od 0.01, dok razlika od 10 poena na PISA skali odgovara efektu veličine 0.10 (OECD, 2019).

Nešto ređe je u upotrebi **stenajn** skala (eng. *stanine*, od *standard nine*), čije su aritmetička sredina M = 5 i standardna devijacija σ = 2 (negde se navodi i $\sigma \approx 2$ ili σ = 1.96). Skorovi na ovoj skali nalaze se u opsegu od 1 do 9, a izražavaju se kao celi brojevi. Zbog relativno uskog raspona skorova i zbog toga što nema specifične prednosti u poređenju sa već prikazanim standardnim skorovima, nije široko rasprostranjena.

Opšta formula za prevođenje sirovih skorova na neku od standardnih skala je

$$X_n = \frac{\sigma_n}{\sigma_s}(X_s - M_s) + M_n$$

gde X označava skor, M aritmetičku sredinu, a σ standardnu devijaciju, pri čemu indeks s označava sirove („stare"), a n normirane („nove") vrednosti. Ovako izračunati standardni skorovi se nekad nazivaju i izvedenim standardnim skorovima. Možemo primetiti da su u ovoj opštoj formuli zapravo sažeta dva koraka – standardizacija, odnosno prevođenje sirovih skorova na Z-skalu, a zatim prevođenje Z skorova na drugu, izabranu skalu. Ako bismo ovo razložili, prevođenje skorova na Z-skalu vršimo sledećom transformacijom:

$$Z = \frac{(X_s - M_s)}{\sigma_s}$$

dok standardne skorove prevodimo na drugu skalu primenom formule:

$$X_n = Z * \sigma_n + M_n$$

9.2.2) Nelinearne transformacije

Ponovo, iako bismo sirove podatke mogli nelinearno transformisati na beskonačno mnogo načina, prilikom izrade normi po pravilu se koriste fraktilne[33], odnosno percentilne norme. **Percentilne norme** počivaju na frekvenci skorova u distribuciji. Imajući u vidu da se većina fenomena koje psiholozi izučavaju raspodeljuje (približno) normalno, jasno je da između sirovih skorova i percentila ne može postojati linearan odnos. Kao što smo prethodno ilustrovali, „pomeranje" za konstantan broj poena na skali sirovih skorova neće uvek biti praćeno jednakom promenom u frekvenci skorova ispitanika, što predstavlja i osnovni nedostatak ovakvih normi (Bukvić, 1996; Fajgelj, 2009; Kline, 2000).

Kada govorimo o percentilnim normama, važno je razlikovati percentile od percentilnih rangova (Bukvić, 1996; Chadha, 2009; Fajgelj, 2009). **Percentil** je sirovi skor ispod koga se nalazi određeni procenat slučajeva. Na primer, ukoliko je medijana distribucije skor 16, onda je taj skor percentil 50, što možemo napisati i kao $P_{50} = 16$. Dakle, percentil ispod koga se nalazi 50% ispitanika je sirovi skor 16. Sa druge strane, **percentilni rang** je procenat slučajeva koji se nalazi ispod određenog skora. Tako bi percentilni rang za skor 16 bio 50%, budući da se ispod datog skora nalazi 50% ispitanika. Ovo bismo mogli napisati $PR_{16} = 50$, odnosno percentilni rang skora 16 je 50 (procenata).

Kada navodimo percentilne norme, po pravilu smo zainteresovani za percentilne rangove. Zbog čega je to tako? Zbog toga što nas zanima da postignuće ispitanika uporedimo sa postignućem referentne grupe. Stoga, zanima nas koji procenat slučajeva se nalazi ispod skora koji je postigao ispitanik koga smo testirali, a ovo odgovara percentilnom rangu. Percentili su nam obično važniji kada opisujemo distribuciju (koji su to skorovi koji dele distribuciju na pola, na četvrtine i sl.), dok su percentilni rangovi važniji za interpretiranje skorova pojedinačnih ispitanika (od koliko ispitanika je ispitanik sa ovim skorom bolji).

Ovde nećemo navoditi formule za računanje percentila ni percentilnih rangova, ali je važno napomenuti da se one oslanjaju isključivo na empirijski dobijenu distribuciju skorova. Dakle, za računanje percentila i percentilnih rangova nije neophodno da distribucija ispunjava bilo kakve uslove, odnosno percentili i percentilni rangovi se mogu računati i za distribucije koje odstupaju od normalne (Chadha, 2009; Fajgelj, 2009).

Na početku poglavlja govorili smo o tome da termin *baždarenje* između ostalog podrazumeva i kalibrisanje, odnosno podešavanje mernih instrumenata tako da budu na zadovoljavajućem nivou preciznosti. U slučaju izrade normi, ovo podešavanje odnosi se upravo na odstupanja od modela normalne distribucije. Naime, iako je naše očekivanje za veliku većinu psiholoških konstrukata da će se normalno distribuirati u populaciji, uzorci koje izvlačimo iz populacije će po pravilu pokazivati manja ili veća odstupanja od normalnosti. Ukoliko su ta odstupanja veća, obično se postavlja pitanje reprezentativnosti uzorka ili pak pravog oblika populacijske distribucije, dok u slučaju manjih odstupanja i dalje verujemo da je uzorak izvučen iz normalno distribuirane populacije, a manja odstupanja objašnjavamo delovanjem slučajne greške.

---

[33] Termin *fraktilne norme* je opštiji od termina *percentilne norme*, jer pored percentilnih (koje dele distribuciju na 100 jednakih grupa) obuhvata i sve druge vrste fraktilnih normi, kao što su, npr., kvartilne (koje dele distribuciju na 4 dela) ili decilne (na 10 delova) (Fajgelj, 2009).

Ukoliko bismo, međutim, norme napravili na osnovu sirovih skorova dobijenih na ovakvom uzorku, izmerena greška bi bila „ugrađena” i u norme. Kako bismo norme učinili što univerzalnijim, odnosno kako bismo umanjili (neutralisali) delovanje greške, najčešće pribegavamo procesu normalizacije skorova.

Normalizacija se takođe može smatrati nelinearnom transformacijom sirovih skorova, koja se vrši sa ciljem približavanja izvorne distribucije skorova normalnoj (teorijski očekivanoj) distribuciji. U odeljku o diskriminativnosti objasnili smo kako se skorovi mogu normalizovati i primetili da su normalizovani skorovi uvek izraženi kao standardni, Z-skorovi (s obzirom na to da su to jedinice standardne normalne distribucije). Drugim rečima, prilikom normalizacije skorova mi smo, pored nelinearne, primenili i linearnu transformaciju sirovih skorova – preveli smo ih u Z-skorove. Naravno, normalizovane skorove možemo naknadno prevesti i na bilo koju drugu željenu skalu, a kombinovanje linearnih i nelinearnih transformacija je uobičajen postupak prilikom izrade normi, pri čemu nelinearne transformacije obično prethode linearnim (Bukvić, 1996; Fajgelj, 2009; Kline, 2000).

9.3) Konstrukcija i upotreba normi

Vratimo se sada na primer sa početka poglavlja, odnosno na psihologe u praksi koji treba da primene određeni merni instrument i protumače rezultate koje ispitanici postižu. Iako bi, naravno, svi psiholozi (pod uslovom da imaju sve potrebne informacije) umeli da primene odgovarajuće linearne i nelinearne transformacije kako bi dobili standardni skor ispitanika, ovo bi nesumnjivo oduzimalo jako puno vremena ukoliko se testiranje vrši relativno često (što je obično slučaj). Zbog toga, kako bi se korišćenje testa što više olakšalo konačnim korisnicima (psiholozima u praksi), kao sastavni deo priručnika o korišćenju testa navodi se i tabela normi (Bukvić, 1996; Chadha, 2009; Fajgelj, 2009).

Tabela normi je tabela koja po pravilu sadrži sirove i standardne skorove/percentilne rangove, a može sadržati i standardnu grešku merenja i interval poverenja za standardne skorove, kao i kategorije ispitanika koje se mogu razlikovati na osnovu postignuća na testu. Ona ima onoliko redova koliko postoji teorijski mogućih skorova na nekom instrumentu, dok broj kolona zavisi od količine informacija koje su date budućem korisniku i obično pokriva navedene vrste skorova i dodatnih informacija. Ukoliko test sadrži nekoliko subskala – potrebno je napraviti tabelu normi za svaku od njih, kao i za ukupni skor na instrumentu (ukoliko ga je moguće izračunati na osnovu subskala). Uz tabelu normi obično se navodi i kratka informacija o tome kako su skorovi normirani (koje transformacje su izvršene), kao i o tome kako tabelu treba koristiti. Ideja je da jednostavnim pregledom tabele normi korisnik ili korisnica testa može na osnovu dobijenog sirovog skora utvrditi standardni skor, percentilni rang, interval u kome se nalazi pravi skor ispitanika itd.

Kako bismo došli do normi za određeni instrument, potrebno je da ga primenimo na tzv. normativnom uzorku. **Normativni uzorak** je dovoljno veliki i dovoljno reprezentativni uzorak na kome očekujemo da će se empirijski dobiti svi teorijski mogući skorovi u populaciji kojoj je test namenjen i to tako da empirijska distribucija što više odgovara teorijskoj (Bukvić, 1996; Chadha, 2009; R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009; Kline, 2000). Ukoliko se pokaže da u normativnom uzorku postoje razlike[34] između određenih širokih kategorija, kao što su polne, uzrasne razlike,

[34] S obzirom na to da su normativni uzorci po pravilu veliki, i male razlike između grupa mogu biti statistički značajne. Zato je bolje, umesto razmatranja značajnosti, obratiti pažnju na veličinu efekta,

razlike po obrazovanju, mestu stanovanja ispitanika (selo–grad) i sl., preporučuje se izrada zasebnih normi za različite subpopulacije. U ovom slučaju, važno je da normativni uzorak bude dovoljno veliki i reprezentativan i za svaku od subpopulacija za koju se izrađuju norme. Po potrebi, mogu se izraditi i lokalne norme za specifične populacije kojima će test biti zadavan. Kada interpretiramo rezultate koje je ispitanik postigao na testu, jako je važno da ga poredimo sa odgovarajućom referentnom grupom, odnosno da primenimo adekvatne norme. U suprotnom, nećemo dobiti pravu sliku o postignuću ispitanika.

Sa prolaskom vremena, dešava se da norme „zastare", te je s vremena na vreme potrebno ponovo primeniti test na novom normativnom uzorku i na osnovu njega ponuditi nove norme (R. J. Cohen & Swerdlik, 2018; Fajgelj, 2009). Moguće je i da se, vremenom, razlike između određenih subpopulacija smanjuju/gube ili pak pojačavaju, što je takođe važno uzeti u obzir kako prilikom ponovnog normiranja, tako i prilikom upotrebe „starih" normi u testiranju.

9.4) Normiranje u SPSS-u

Linearna transfromacija izvornih skorova na bilo koju željenu skalu veoma je jednostavna u SPSS-u, a dostupna je i opcija normalizacije skorova. Moguće je, kao korak ka izračunavanju percentilnih rangova, dobiti i kumulativnu distribuciju frekvencija.

Kao primer normiranja koristićemo već pominjanu skalu anksioznosti od 16 ajtema. Kako bismo izvorne skorove mogli da pretvorimo u bilo koju drugu skalu, prateći prethodno date formule, neophodno je da znamo koliko iznose aritmetička sredina i standardna devijacija sumarnog skora, kada je on izražen u jedinicama izvorne skale. Za ovo ćemo iskoristiti opciju **Descriptives** u SPSS-u (a možemo koristiti i **Frequencies**). U ispisu (Slika 9.1) dobijamo da je aritmetička sredina M = 61.7097, dok je standardna devijacija izvornih skorova SD = 8.40236.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
anx	93	34.00	78.00	61.7097	8.40236
Valid N (listwise)	93				

Slika 9.1

Imajući jednostavnost komande u vidu, umesto korišćenja dijaloga **Transform → Compute Variable**, za linearno transformisanje izvornih skorova na drugu skalu koristićemo SPSS sintaksu i komandu **compute**. Naravno, prilikom prikazivanja ispisa, SPSS je zaokružio vrednosti M i SD na određeni broj decimala, ali dvostrukim klikom na bilo koju od njih možemo dobiti informaciju i o dodatnim decimalama. Radi ilustracije, napravićemo dve nove varijable koje sadrže standardizovane skorove sa izvorne skale, samo ćemo koristiti drugačije zaokružene vrednosti M i SD. Komande koje ćemo upisati u SPSS sintaksu su:

```
compute Z1 = (anx - 61.7097) / 8.40236.
compute Z2 = (anx - 61.709677) / 8.402356.
execute.
```

Standardizovani skor takođe možemo dobiti i korišćenjem opcije *Save standardized values as variables* u okviru **Descriptives**. Na ovaj način postižemo najveći stepen preciznosti, budući da SPSS automatski uključuje sve decimale aritmetičke sredine i standardne devijacije izvorne skale u računanje Z skora.

---

npr., računanjem Koenovog d (eng. *Cohen's d*), prilikom donošenja odluke o pravljenju zasebnih ili jedinstvenih normi.

Ukoliko sada proverimo aritmetičku sredinu i standardnu devijaciju novih varijabli, vidimo da smo uspešno preveli varijablu na Z-skalu (Slika 9.2), odnosno da su aritmetička sredina i standardna devijacija sve tri varijable jednake M = 0 i SD = 1, kao što bi i trebalo da budu, pri čemu najveći stepen preciznosti dobijamo ukoliko koristimo opciju standardizacije ugrađenu u SPSS.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Z1	93	-3.30	1.94	.0000	1.00000
Z2	93	-3.30	1.94	.0000	1.00000
Zscore(anx)	93	-3.29785	1.93878	0E-7	1.00000000
Valid N (listwise)	93				

Slika 9.2

Ukoliko bismo dalje želeli da skorove sa ovog upitnika anksioznosti prevedemo na neku drugu (izvedenu) standardnu skalu, na primer, T-skalu, koja odgovara prirodi konstrukta, mogli bismo ponovo napisati jednostavnu sintaksu (prateći odgovarajuću formulu i koristeći SPSS standardizovani skor):

```
compute Tanx = Zanx * 10 + 50.
execute.
```

Deskriptivni pokazatelji nove varijable potvrđuju da smo standardne skorove preveli na skalu čije su aritmetička sredina i standardna devijacija M = 50 i SD = 10 (Slika 9.3).

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Tanx	93	17.02	69.39	50.0000	10.00000
Valid N (listwise)	93				

Slika 9.3

Budući da se radi o linearnim transformacijama, korelacija izvornih skorova i skorova prevedenih na bilo koju od standardnih skala biće jednaka jedinici, koliko će iznositi i međusobne korelacije standardnih skorova, što možemo videti i u tabeli korelacija (Slika 9.4). Dakle, bez obzira na tip linearne transformacije koji primenimo, odnosi između skorova ostaju nepromenjeni.

Correlations

		anx	Z1	Z2	Zscore(anx)	Tanx
anx	Pearson Correlation	1	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	93	93	93	93	93
Z1	Pearson Correlation	1.000**	1	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	93	93	93	93	93
Z2	Pearson Correlation	1.000**	1.000**	1	1.000**	1.000**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	93	93	93	93	93
Zscore(anx)	Pearson Correlation	1.000**	1.000**	1.000**	1	1.000**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	93	93	93	93	93
Tanx	Pearson Correlation	1.000**	1.000**	1.000**	1.000**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	93	93	93	93	93

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 9.4

Da je distribucija skorova odstupala od normalnosti (pod uslovom da je odstupanje malo), verovatno bismo pre standardizacije primenili normalizaciju skorova. Kao što je u poglavlju o diskriminativnosti objašnjeno, za ovo koristimo opciju **Transform → Rank Cases**, a zatim kao *Rank Types* izaberemo *Normal Scores (Proportion Estimation Formula – Rankit)*. Normalizovani skorovi će biti izraženi u Z skali, te ukoliko želimo da ih prebacimo na neku drugu skalu koristimo istu formulu kao i za obične standardne skorove, samo koristimo drugu varijablu:

```
compute TNanx = Nanx * 10 + 50.
execute.
```

S obzirom na to da normalizacija predstavlja nelinearnu transformaciju podataka, vidimo da normalizovani skorovi (bilo da su izraženi u Z ili T-skali) ne koreliraju savršeno sa izvornom varijablom (Slika 9.5). Korelacija je, međutim, u našem primeru visoka, što ukazuje na to da je izvorna varijabla već bila (barem približno) normalno distribuirana.

Correlations

		anx	Normal Score of anx using Rankit's Formula	TNanx
anx	Pearson Correlation	1	.991**	.991**
	Sig. (2-tailed)		.000	.000
	N	93	93	93
Normal Score of anx using Rankit's Formula	Pearson Correlation	.991**	1	1.000**
	Sig. (2-tailed)	.000		.000
	N	93	93	93
TNanx	Pearson Correlation	.991**	1.000**	1
	Sig. (2-tailed)	.000	.000	
	N	93	93	93

**. Correlation is significant at the 0.01 level (2-tailed).

Slika 9.5

Kako bismo u tabeli normi imali sve važne informacije, potrebno je da za svaki standardizovani skor damo donju i gornju granicu intervala poverenja, određenog na osnovu standardne greške merenja. Standardnu grešku merenja dobijamo po već poznatoj formuli $\sigma_{Em} = \sigma_X\sqrt{(1 - r_{tt})}$, odnosno $SEm = SD\sqrt{(1-\alpha)}$, gde su SD standardna devijacija skale i α pouzdanost instrumenta. Interval poverenja zatim dobijamo kao $CI_{95} = X \pm 1.96 * SEm$, odnosno $CI_{99} = X \pm 2.58 * SEm$.

Korišćenjem opcije **Analyze → Scale → Reliability Analysis** dobijamo da je pouzdanost skale anksioznosti nešto niža od optimalne, ali zadovoljavajuća – α = .769. Kako bismo izračunali standardnu grešku, potrebna nam je još i informacija o standardnoj devijaciji skale. Ovde je važno povesti računa o tome da se uzme odgovarajuća vrednost, odnosno da se koristi standardna devijacija one skale za koju računamo standardnu grešku. Ukoliko želimo da izračunamo standardnu grešku sirovih skorova, iskoristićemo standardnu devijaciju sirovih skorova (u našem primeru ona iznosi SD = 8.40236). Ukoliko računamo standardnu grešku standardnih skorova, koristićemo standardnu devijaciju Z-skorova (SD = 1), a ako računamo standardnu grešku T-skorova koristićemo standardnu devijaciju za T-skalu (SD = 10).

S obzirom na to da smo odlučili da sirove skorove prevedemo na T-skalu, najlogičnije je da interval poverenja izračunamo upravo za T-skorove. Primenom formule dobijamo da je $SEm = SD\sqrt{(1-\alpha)} = 10 * \sqrt{(1-.769)} = 4.806$, a da je 95% interval poverenja $CI_{95} = X \pm 1.96 * SEm =$

$X \pm 9.42$. Drugim rečima, pravi skor ispitanika se sa 95% sigurnosti nalazi u okviru granica određenih intervalom poverenja, pri čemu primećujemo da je standardna greška merenja jednaka za sve skorove. Tako bi se pravi skor ispitanika koji je ostvario T-skor od 20 nalazio između T-skorova 11 i 29, dok bi se pravi skor ispitanika koji je ostvario T-skor od 42 nalazio između skorova 33 i 51.

Za računanje percentilnih rangova potrebno nam je da utvrdimo proporciju ispitanika koji su ostvarili svaki teorijski mogući sirovi skor na instrumentu, kao i kumulativni procenat ispitanika koji ostvaruju dati skor ili niži. Kako sirovi skorovi savršeno koreliraju sa svojim linearnim transformacijama, ovo će ujedno biti i proporcija ispitanika koji postižu korespondentne standardne skorove. Procente i kumulativne procente možemo dobiti preko *Frequency Table* u okviru **Frequencies** analize u SPSS-u (Slika 9.6), a izuzev prve kolone, ispis bi bio potpuno isti i da smo u analizu ubacili neku od varijabli standardnih skorova.

anx

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	34.00	1	1.1	1.1	1.1
	40.00	1	1.1	1.1	2.2
	45.00	2	2.2	2.2	4.3
	46.00	1	1.1	1.1	5.4
	49.00	2	2.2	2.2	7.5
	50.00	2	2.2	2.2	9.7
	51.00	1	1.1	1.1	10.8
	53.00	4	4.3	4.3	15.1
	54.00	2	2.2	2.2	17.2
	55.00	4	4.3	4.3	21.5
	56.00	4	4.3	4.3	25.8
	57.00	3	3.2	3.2	29.0
	58.00	3	3.2	3.2	32.3
	59.00	2	2.2	2.2	34.4
	60.00	7	7.5	7.5	41.9
	61.00	5	5.4	5.4	47.3
	62.00	4	4.3	4.3	51.6
	63.00	8	8.6	8.6	60.2
	64.00	2	2.2	2.2	62.4
	65.00	3	3.2	3.2	65.6
	66.00	3	3.2	3.2	68.8
	67.00	7	7.5	7.5	76.3
	68.00	1	1.1	1.1	77.4
	69.00	3	3.2	3.2	80.6
	70.00	4	4.3	4.3	84.9
	71.00	4	4.3	4.3	89.2
	72.00	3	3.2	3.2	92.5
	74.00	2	2.2	2.2	94.6
	75.00	1	1.1	1.1	95.7
	76.00	1	1.1	1.1	96.8
	77.00	2	2.2	2.2	98.9
	78.00	1	1.1	1.1	100.0
	Total	93	100.0	100.0	

Slika 9.6

Prva kolona u tabeli odnosi se na sirove skorove ispitanika (ili standardne ukoliko smo analizirali varijablu standardnih skorova), druga na njihove frekvence. U trećoj koloni imamo procenat ispitanika koji ostvaruju dati skor, a u četvrtoj „validni" procenat, odnosno procenat kada u analizu ulaze samo ispitanici koji nemaju nedostajuće podatke. Kako u našem primeru nije bilo nedostajućih podataka, ove dve vrednosti su jednake. Konačno, poslednja kolona daje kumulativni procenat slučajeva, odnosno procenat ispitanika koji su ostvarili dati skor ili manji. Primenom odgovarajućih formula, na osnovu frekvenci i kumulativnih frekvenci skorova možemo izračunati percentilne rangove, ali se time ovde nećemo detaljnije baviti. Od podataka koji su nam direktno dostupni kao deo SPSS ispisa kumulativni procenat najbolje aproksimira percentilni rang datog skora, mada mu nije jednak.

Možemo primetiti da, iako su teorijski minimalni i maksimalni skor na instrumentu 16 i 80, empirijski maksimum i minimum ne odgovaraju ovim vrednostima (što nije previše iznenađujuće imajući relativno mali broj ispitanika u vidu). Posebno u slučaju minimalnih vrednosti empirijska vrednost odstupa od teorijske. Kako je u tabeli normi potrebno dati percentilne rangove za sve teorijski moguće skorove, morali bismo „ručno" da dopunimo vrednosti za sve skorove koji nedostaju. To nisu samo najniži i najviši skorovi na instrumentu, već i drugi skorovi koji nisu prisutni u distribuciji (npr., skorovi između 34 i 40 ili 40 i 45, kao i skor 73, na primer). S obzirom na to da niko od ispitanika nije postigao dati skor – njegova frekvenca/procenat će biti nulti, dok će kumulativna frekvenca/procenat biti isti kao i za niži razred. Na primer, kako niko od ispitanika nije postigao skor 73, njegova frekvenca je 0, procenat je takođe 0, a kumulativni procenat je isti kao i kumulativni procenat za skor 72, odnosno 92.5. U slučaju minimalnih skorova koje niko od ispitanika nije ostvario (skorovi 33 i niži), njihova kumulativna frekvenca/procenat će biti nulti.

9.5) Normiranje u programu Winsteps/Ministep

U okviru IRT-a normiranje se vrši prevođenjem procenjenih pravih skorova na odgovarajuću standardnu skalu. Pretpostavka prilikom normiranja jeste da procenjene pozicije i ispitanika i stavki na kontinuumu latentne dimenzije predstavljaju njihove prave skorove.

Normiranje u IRT-u ilustrovaćemo u programu Winsteps/Ministeps (Linacre, 2022). Nakon što smo izvezli podatke iz SPSS-a (videti Dodatak 2), podesićemo kontrolni fajl na odgovarajući način. Pored uobičajenih podešavanja, potrebno je da posebnu pažnju obratimo na komande UMEAN i USCALE. Kao što smo već rekli, podrazumevane vrednosti su UMEAN = 0 i USCALE = 1, što odgovara aritmetičkoj sredini i standardnoj devijaciji Z-skale. Ukoliko želimo da prave skorove ispitanika izrazimo na nekoj drugoj skali, potrebno je da unesemo odgovarajuće vrednosti aritmetičke sredine i standardne devijacije. Na primer, kada bismo želeli da prebacimo skorove na IQ skalu, upisali bismo UMEAN = 100 i USCALE = 15. U našem primeru, kao i kada smo normiranje vršili u SPSS-u, izabraćemo T-skalu, te ćemo upisati UMEAN = 50 i USCALE = 10.

U ispisu ćemo sada otvoriti Tabelu 20. *Score table*, koja daje skorove ispitanika prevedene na željenu standardnu skalu. I ovaj ispis sadrži nekoliko elemenata. Prva tabela *Table of measures on complete test*, odnosno tabela mera (pozicija na latentnoj dimenziji) za ceo test, daje pregled osnovnih informacija (Slika 9.7). Drugim rečima, u ovoj tabeli za svaki sirovi skor možemo pronaći odgovarajuću poziciju na latentnoj dimenziji, kada je ona iskazana u jedinicama standardne skale, kao i pripadajuću standardnu grešku merenja.

TABLE OF MEASURES ON COMPLETE TEST

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
16	-6.02E	18.24	38	36.71	2.31	60	47.89	2.42
17	5.97	9.96	39	37.24	2.28	61	48.48	2.46
18	12.83	7.02	40	37.75	2.26	62	49.10	2.50
19	16.82	5.71	41	38.26	2.24	63	49.74	2.56
20	19.62	4.93	42	38.76	2.23	64	50.40	2.61
21	21.79	4.40	43	39.25	2.22	65	51.10	2.68
22	23.56	4.01	44	39.74	2.21	66	51.84	2.75
23	25.05	3.72	45	40.23	2.20	67	52.62	2.84
24	26.34	3.48	46	40.71	2.20	68	53.45	2.94
25	27.48	3.29	47	41.19	2.20	69	54.35	3.05
26	28.51	3.12	48	41.68	2.20	70	55.32	3.18
27	29.44	2.99	49	42.16	2.20	71	56.38	3.34
28	30.30	2.87	50	42.65	2.21	72	57.56	3.53
29	31.10	2.78	51	43.14	2.22	73	58.89	3.77
30	31.84	2.69	52	43.63	2.23	74	60.42	4.06
31	32.55	2.62	53	44.13	2.24	75	62.23	4.45
32	33.22	2.55	54	44.64	2.26	76	64.44	4.98
33	33.85	2.50	55	45.15	2.28	77	67.29	5.75
34	34.46	2.45	56	45.67	2.30	78	71.32	7.06
35	35.05	2.40	57	46.21	2.32	79	78.24	9.99
36	35.62	2.37	58	46.75	2.35	80	90.27E	18.26
37	36.17	2.33	59	47.31	2.38			

TO SET MEASURE RANGE AS 0-100, UMEAN=58.177 USCALE=10.385
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=53.233 USCALE=6.647

Slika 9.7

U dnu tabele imamo date vrednosti UMEAN i USCALE, koje bi trebalo uneti kako bi se procenjeni pravi skorovi kretali u okviru vrednosti 0–100 (u prvom redu), odnosno kako bi odgovarali rasponu sirovih skorova (u drugom redu). Ipak, uglavnom ćemo koristiti uobičajene standardne skale za normiranje instrumenata.

U tabeli *Table of sample norms (500/100) and frequencies corresponding to complete test*, odnosno tabeli uzoračkih normi i frekvencija za ceo test nalaze se iste ove informacije, ali i neke dodatne (Slika 9.8). Kao i u prethodnoj tabeli, kolona SCORE odnosi se na sirovi skor ispitanika, odnosno na obični sumarni skor na instrumentu. Kolona MEASURE predstavlja procenjeni položaj ispitanika na latentnoj dimenziji, izražen kao T-skor, jer smo tako podesili kontrolni fajl. Uz T-skor navedena je i pripadajuća standardna greška merenja – S.E., koja je takođe data u jedinicama T-skale. Za ispitanike koji su postigli najniži, odnosno najviši skor na testu procena pravog skora se vrši na nešto drugačiji način nego za ostale skorove, zbog čega su oni obeleženi slovom E pored standardnog skora. E se odnosi na ekstrapolirano, ali i na ekstremno, jer su ovo uvek ekstremni skorovi (Linacre, 2022). Možemo primetiti da je minimalni standardni skor u ovom primeru negativan broj (-6.02). Ovo je posledica toga što se standardni skorovi automatski računaju primenom odgovarajuće formule, ali kako znamo da T-skala ne obuhvata negativne brojeve, u tabeli normi bismo ovaj skor zamenili skorom 0.

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
16	-6.02E	18.24	-392	290	0	.0	0	.0	0
17	5.97	9.96	-201	159	0	.0	0	.0	0
18	12.83	7.02	-92	112	0	.0	0	.0	0
19	16.82	5.71	-28	91	0	.0	0	.0	0
20	19.62	4.93	16	79	0	.0	0	.0	0
21	21.79	4.40	51	70	0	.0	0	.0	0
22	23.56	4.01	79	64	0	.0	0	.0	0
23	25.05	3.72	103	59	0	.0	0	.0	0
24	26.34	3.48	123	55	0	.0	0	.0	0
25	27.48	3.29	141	52	0	.0	0	.0	0
26	28.51	3.12	158	50	0	.0	0	.0	0
27	29.44	2.99	173	48	0	.0	0	.0	0
28	30.30	2.87	186	46	0	.0	0	.0	0
29	31.10	2.78	199	44	0	.0	0	.0	0
30	31.84	2.69	211	43	0	.0	0	.0	0
31	32.55	2.62	222	42	0	.0	0	.0	0
32	33.22	2.55	233	41	0	.0	0	.0	0
33	33.85	2.50	243	40	0	.0	0	.0	0
34	34.46	2.45	253	39	1	1.1	1	1.1	1
35	35.05	2.40	262	38	0	.0	1	1.1	1
36	35.62	2.37	271	38	0	.0	1	1.1	1
37	36.17	2.33	280	37	0	.0	1	1.1	1
38	36.71	2.31	288	37	0	.0	1	1.1	1
39	37.24	2.28	297	36	0	.0	1	1.1	1
40	37.75	2.26	305	36	1	1.1	2	2.2	2
41	38.26	2.24	313	36	0	.0	2	2.2	2
42	38.76	2.23	321	35	0	.0	2	2.2	2
43	39.25	2.22	329	35	0	.0	2	2.2	2
44	39.74	2.21	337	35	0	.0	2	2.2	2
45	40.23	2.20	344	35	2	2.2	4	4.3	3
46	40.71	2.20	352	35	1	1.1	5	5.4	5
47	41.19	2.20	360	35	0	.0	5	5.4	5
48	41.68	2.20	367	35	0	.0	5	5.4	5
49	42.16	2.20	375	35	2	2.2	7	7.5	6
50	42.65	2.21	383	35	2	2.2	9	9.7	9
51	43.14	2.22	391	35	1	1.1	10	10.8	10
52	43.63	2.23	399	35	0	.0	10	10.8	11
53	44.13	2.24	407	36	4	4.3	14	15.1	13
54	44.64	2.26	415	36	2	2.2	16	17.2	16
55	45.15	2.28	423	36	4	4.3	20	21.5	19
56	45.67	2.30	431	37	4	4.3	24	25.8	24
57	46.21	2.32	440	37	3	3.2	27	29.0	27
58	46.75	2.35	448	37	3	3.2	30	32.3	31
59	47.31	2.38	457	38	2	2.2	32	34.4	33
60	47.89	2.42	466	39	7	7.5	39	41.9	38
61	48.48	2.46	476	39	5	5.4	44	47.3	45
62	49.10	2.50	486	40	4	4.3	48	51.6	49
63	49.74	2.56	496	41	8	8.6	56	60.2	56
64	50.40	2.61	506	42	2	2.2	58	62.4	61
65	51.10	2.68	518	43	3	3.2	61	65.6	64
66	51.84	2.75	529	44	3	3.2	64	68.8	67
67	52.62	2.84	542	45	7	7.5	71	76.3	73
68	53.45	2.94	555	47	1	1.1	72	77.4	77
69	54.35	3.05	569	49	3	3.2	75	80.6	79
70	55.32	3.18	585	51	4	4.3	79	84.9	83
71	56.38	3.34	602	53	4	4.3	83	89.2	87
72	57.56	3.53	620	56	3	3.2	86	92.5	91
73	58.89	3.77	642	60	0	.0	86	92.5	92
74	60.42	4.06	666	65	2	2.2	88	94.6	94
75	62.23	4.45	695	71	1	1.1	89	95.7	95
76	64.44	4.98	730	79	1	1.1	90	96.8	96
77	67.29	5.75	775	92	2	2.2	92	98.9	98
78	71.32	7.06	840	112	1	1.1	93	100.0	99
79	78.24	9.99	950	159	0	.0	93	100.0	100
80	90.27E	18.26	1141	291	0	.0	93	100.0	100

Slika 9.8

Važno je primetiti da, za razliku od klasičnog modela gde je standardna greška merenja ista za sve ispitanike, ovo nije slučaj u IRT-u. Kao što smo objasnili u poglavlju o IRT-u, standardna greška merenja je u funkciji informativnosti testa, odnosno nije ista na svim nivoima osobine. Jasno se vidi da je greška merenja najmanja za skorove blizu aritmetičke sredine stavki (u ovom primeru, između

45 i 49 poena), te da raste kako idemo ka višim, odnosno nižim skorovima. Najveću grešku u proceni osobine pravimo za ispitanike sa minimalnim, odnosno maksimalnim skorom, dok smo najpreciznije procenili one koji su postizali skorove na onom nivou gde je informativnost testa najveća (odnosno blizu aritmetičke sredine stavki).

Kolone NORMED i njena pripadajuća standardna greška merenja (S.E.) predstavljaju prave skorove izražene u jedinicama PISA skale, koja, kao što smo rekli, ima aritmetičku sredinu M = 500, a standardnu devijaciju SD = 100 (što je napisano i u nazivu tabele), ali ovi skorovi su nam od manjeg značaja, pa ih nećemo posebno tumačiti.

Narednih pet kolona odnosi se na frekvencu (FREQUENCY), odnosno procenat (%) ispitanika koji postižu dati sirovi skor, zatim kumulativnu frekvencu (CUM. FREQ.) i kumulativni procenat (%), te percentilni rang ispitanika (PERCENTILE). Iako je (pomalo zbunjujuće) poslednja kolona imenovana PERCENTILE, što bismo možda preveli kao percentil, ona zapravo daje informaciju o procentu ispitanika koji se nalaze ispod datog skora, što odgovara definiciji percentilnog ranga.

Ukoliko ovu tabelu uporedimo sa ispisom analize u SPSS-u, primećujemo da Winsteps/Ministep automatski daje standardne skorove, frekvence/procente i kumulativne frekvence/procente za SVE teorijski moguće skorove na instrumentu, dok smo u SPSS-u dobijali informaciju samo za one vrednosti koje su empirijski bile prisutne u uzorku. Ovo je zbog toga što IRT počiva na skaliranju u skladu sa teorijskim modelom, te je za svaki skor moguće odrediti njegov položaj na dimenziji latentne osobine, čak i ako ga niko od ispitanika u uzorku nije postigao. Dakle, iako su frekvence određenih sirovih skorova jednake 0, oni su svejedno zastupljeni u tabeli normi i pridružen im je odgovarajući standardni skor, kao i percentilni rang.

Tabela normi koja predstavlja sastavni deo priručnika o korišćenju testa obično će veoma ličiti na tabelu uzoračkih normi datu u ispisu IRT analize, uz napomenu da se standardni skorovi (osim ukoliko se radi o Z-skali) po pravilu zaokružuju na cele (pozitivne) brojeve, kao i da se često navode intervali poverenja za skorove (opet dati kao celi brojevi). Intervale poverenja lako možemo izračunati na osnovu standardnog skora i pripadajuće standardne greške merenja, a tom prilikom ćemo još jednom moći da primetimo da je preciznost merenja nejednaka na različitim nivoima latentne osobine (konkretno, da opada kako idemo ka ekstremnim skorovima). Na primer, 95% interval poverenja za T-skor 20 (u tabeli dat na dve decimale kao 19.62) iznosi $CI_{95} = 19.62 \pm 1.96 * 4.93 = X \pm 9.66$, odnosno pravi skor ispitanika čiji je T-skor 20 nalazi se u intervalu 10 – 29. Sa druge strane, interval poverenja za T-skor 42 (42.16 u tabeli) je $CI_{95} = 42.16 \pm 1.96 * 2.20 = X \pm 4.31$, te se pravi skor ispitanika koji postižu ovaj standardni skor kreće u mnogo užem intervalu između 38 i 46.

9.6) Rezime

Normiranje predstavlja veoma važan korak u konstrukciji psihološkog instrumenta jer omogućava korisnicima testa – psiholozima u praksi – da jednostavno i ispravno tumače rezultate dobijene testom. Norme testa se dobijaju primenom testa na dovoljno velikom i reprezentativnom uzorku koji se naziva normativnim uzorkom, a rezultati normiranja prikazuju se u tabelama normi.

U matematičkom smislu, normiranje je primena određenih linerarnih i/ili nelinearnih transformacija na sirovim skorovima. Pomoću linearnih transformacija sirovi skorovi se prevode na neku od standardnih skala kao što su: Z, IQ, T ili PISA skala, dok se nelinearnim transformacijama najčešće utvrđuje relativni rang postignuća osobe na testu u poređenju sa postignućem drugih osoba iz iste populacije i vrši normalizacija sirovih skorova kako bi norme što bolje odražavale pretpostavljene populacijske vrednosti.

9.7) Preporučena literatura

Bukvić, A. (1996). *Načela izrade psiholoških testova*. Zavod za udžbenike i nastavna sredstva, Beograd.

Chadha, N. K. (2009). Applied Psychometry. In *Applied Psychometry*. New Delhi, India: SAGE Publications India Pvt Ltd.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment: An Introduction to Tests and Measurement. Ninth Edition*. New York, NY: McGraw-Hill Education.

Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.

Kline, P. (2000). *The Handbook of Psychological Testing, Second Edition*. London and New York: Routledge.

10) DODATAK 1 – TRANSFORMACIJE PODATAKA KORIŠĆENJEM SPSS SINTAKSE

Zamislimo da smo zadali ispitanicima neki psihološki test i pred sobom imamo bazu podataka koja sadrži sirove odgovore ispitanika, a želimo da utvrdimo psihometrijske osobine datog testa. Šta bi se desilo ako bismo u analizu pouzdanosti istovremeno ubacili i stavke na kojima se slaganje sa stavkom odnosi na prisustvo osobine (npr., „Volim knjige" i „Volim da čitam") i one kod kojih se slaganje sa stavkom odnosi na odsustvo osobine (npr., „Ne čitam ako ne moram" i „Knjige su dosadne")? Ili ako bismo skor ispitanika na testu opšte informisanosti izračunali tako što bismo sabrali njihove sirove odgovore (drugim rečima, redne brojeve odgovora koje su izabrali), zanemarujući tačnost tih odgovora? Jasno je da bismo dobili besmislene vrednosti koje ne odražavaju metrijska svojstva testa kojim se bavimo, te da ne bismo ništa saznali o onome što nas interesuje.

Pre nego što pređemo na bilo koju analizu podataka, što uključuje i psihometrijsku analizu, odnosno utvrđivanje metrijskih karakteristika psihološkog instrumenta, neophodno je da izvršimo sve potrebne transformacije podataka i uredimo bazu tako da bude pogodna za analizu. Osnovne transformacije podataka podrazumevaju rekodiranje varijabli i računanje ukupnih skorova, a često je potrebno i promeniti redosled varijabli kako bi se baza lakše koristila.

10.1) Prednosti korišćenja SPSS sintakse za transformacije i obradu podataka

SPSS sintaksa je programski jezik IMB® SPSS® statističkog softvera (IBM SPSS Statistics software, "SPSS"), koji omogućava korisniku da putem pisanih komandi, umesto preko korisničkog interfejsa, izvede različite operacije – od najjednostavnijih transformacija podataka do složenih analiza (IBM Corporation, 2021). Iako korišćenje sintakse ponekom može delovati zastrašujuće ili možda samo nepotrebno, poznavanje osnova rada u SPSS sintaksi je višetstruko korisno, jer njome možemo:

- automatizovati određene procese i značajno uštedeti vreme u njihovom inicijalnom izvođenju,
- ponoviti analizu na novom ili proširenom uzorku ispitanika,
- omogućiti drugim istraživačima da pokrenu iste trasnsformacije/analize,
- dodatno specifikovati analize koje već postoje u okviru SPSS opcija,
- pokrenuti analize koje nisu ugrađene u SPSS.

Uz sve ovo, zahvaljujući opciji *Paste*, koja postoji u praktično svakom SPSS dijalogu, sintaksu nije neophodno zasebno učiti već se, barem na početku, možemo osloniti na opcije koje su nam dostupne preko korisničkog interfejsa, a zatim ih doraditi i prilagoditi svojim potrebama.

10.2) Rekodiranje varijabli

Gotovo da ne postoji psihološki merni instrument za koji nije potrebno izvršiti neki tip transformacije sirovih skorova pre nego što izračunamo ukupni skor na instrumentu. U slučaju testova znanja ili sposobnosti sirove odgovore je potrebno rekodirati u tačne odgovore, dok je kod testova ličnosti ili drugih nekognitivnih testova uobičajeno da postoje negativno reflektovane stavke, koje je potrebno rekodirati tako da veće vrednosti odgovaraju manjoj izraženosti osobine i obrnuto.

10.2.1) Rekodiranje u iste varijable

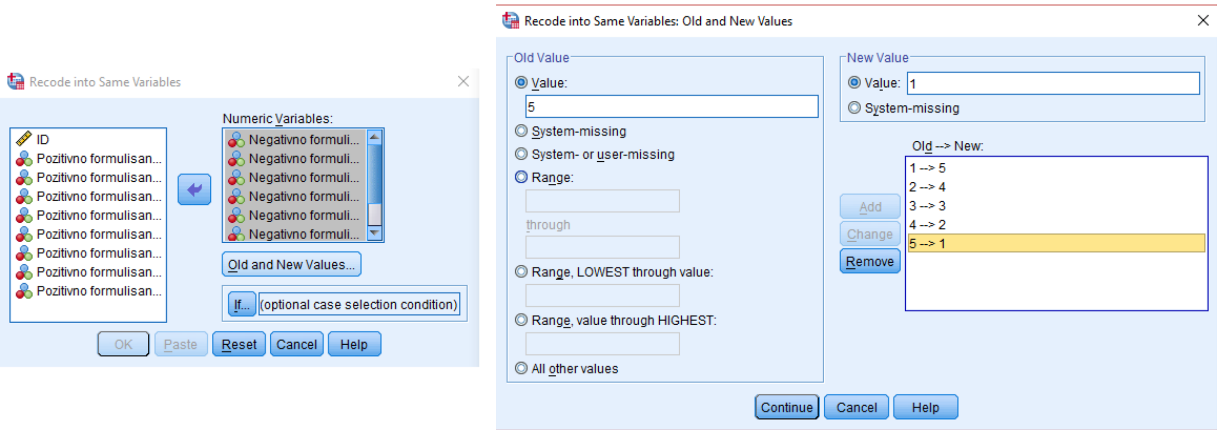
Počećemo sa primerom zamišljenog testa anksioznosti koji sadrži 16 stavki na koje se odgovori daju na skali od 1 do 5, pri čemu veći broj označava veće slaganje sa sadržajem tvrdnje. Osam tvrdnji ovog instrumenta formulisano je tako da se veće slaganje odnosi na veće prisustvo anksioznosti, na primer: „Često se osećam zabrinuto”, dok je preostalih osam stavki tako formulisano da se veće slaganje sa njima interpretira kao niži stepen anksioznosti, na primer: „Optimističan/na sam u pogledu budućnosti”. Očito je da bismo sabiranjem sirovih skorova ispitanika na ovih 16 stavki dobili skor koji se ne može interpretirati, jer se sve tvrdnje ne odnose na isti pol dimenzije anksioznost (što se može videti i na osnovu negativnih korelacija između pozitivno i negativno formulisanih stavki, Slika 10.1), te bismo sabiranjem praktično „poništili” individualne razlike između ispitanika – što nam je upravo cilj da izmerimo. Zato ćemo negativno reflektovane stavke (one kod kojih se veće slaganje odnosi na manju anksioznost) rekodirati tako što ćemo vrednost 1 pretvoriti u 5, vrednost 2 u 4, 3 će ostati 3, vrednost 2 će postati 4, a vrednost 1 će postati 5.

		Negativno formulisana stavka	Negativno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Negativno formulisana stavka	Negativno formulisana stavka	Negativno formulisana stavka
Negativno formulisana stavka	Pearson Correlation	1	.311**	-.400**	-.081	-.345**	-.250*	-.023	.310**	.307**
	Sig. (2-tailed)		.002	.000	.438	.001	.016	.824	.003	.003
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.311**	1	-.343**	-.135	-.084	.030	.059	.234*	.151
	Sig. (2-tailed)	.002		.001	.196	.423	.772	.572	.024	.149
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	-.400**	-.343**	1	.056	.184	.097	.007	-.174	-.059
	Sig. (2-tailed)	.000	.001		.595	.077	.353	.949	.096	.574
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	-.081	-.135	.056	1	.056	.206*	-.282**	-.248*	-.002
	Sig. (2-tailed)	.438	.196	.595		.593	.048	.006	.017	.983
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	-.345**	-.084	.184	.056	1	.378**	-.173	-.486**	-.059
	Sig. (2-tailed)	.001	.423	.077	.593		.000	.096	.000	.574
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	-.250*	.030	.097	.206*	.378**	1	-.129	-.228*	.018
	Sig. (2-tailed)	.016	.772	.353	.048	.000		.217	.028	.865
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	-.023	.059	.007	-.282**	-.173	-.129	1	.358**	.157
	Sig. (2-tailed)	.824	.572	.949	.006	.096	.217		.000	.133
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.310**	.234*	-.174	-.248*	-.486**	-.228*	.358**	1	.080
	Sig. (2-tailed)	.003	.024	.096	.017	.000	.028	.000		.447
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.307**	.151	-.059	-.002	-.059	.018	.157	.080	1
	Sig. (2-tailed)	.003	.149	.574	.983	.574	.865	.133	.447	
	N	93	93	93	93	93	93	93	93	93

Slika 10.1

Svaki put kada rekodiramo stavke imamo izbor da li želimo da nove vrednosti upišemo u već postojeće, izvorne varijable (SPSS opcija *Recode into Same Variables*) ili želimo da kreiramo nove varijable u kojima će se nalaziti rekodirane vrednosti (SPSS opcija *Recode into Different Variables*). Prednost rekodiranja u novu varijablu jeste što zadržavamo i izvornu varijablu tako da smo uvek sigurni šta su sirovi odgovori, a šta rekodirani. U slučaju testa anksioznosti, međutim, izvorne varijable ne nose nikakve dodatne informacije u poređenju sa rekodiranim – zadržavamo isti raspon odgovora i za svakog ispitanika možemo (istom komandom *recode*) u bilo kom trenutku „vratiti” njegov sirov skor. Drugim rečima, rekodiranje u okviru iste varijable ne dovodi ni do kakvog gubitka informacija, tako da ćemo se u ovom slučaju odlučiti za opciju *Recode into Same Variables*.

Rekodiranje u okviru iste varijable se u SPSS-u izvodi preko opcija **Transform → Recode into Same Variables**. U polje *Variables* ubacujemo sve varijable koje treba rekodirati, a zatim u okviru opcije Old and New values definišemo pravila za rekodiranje (Slika 10.2). U slučaju kada rekodiramo unutar istih varijabli, nije nužno da definišemo posebno pravilo za vrednosti koje ostaju iste (u našem primeru, sirovi skor 3 ostaje 3 i nakon rekodiranja, pa to pravilo nismo morali da pišemo).



Slika 10.2

Kada bismo u glavnom dijalogu opcije Recode kliknuli *OK*, SPSS bi rekodirao sve ove varijable. Ukoliko, međutim, kliknemo na dugme *Paste*, koje se nalazi odmah pored *OK* – otvoriće nam se novi, sintaksni prozor, u kome će se nalaziti sledeći tekst:

DATASET ACTIVATE DataSet1.

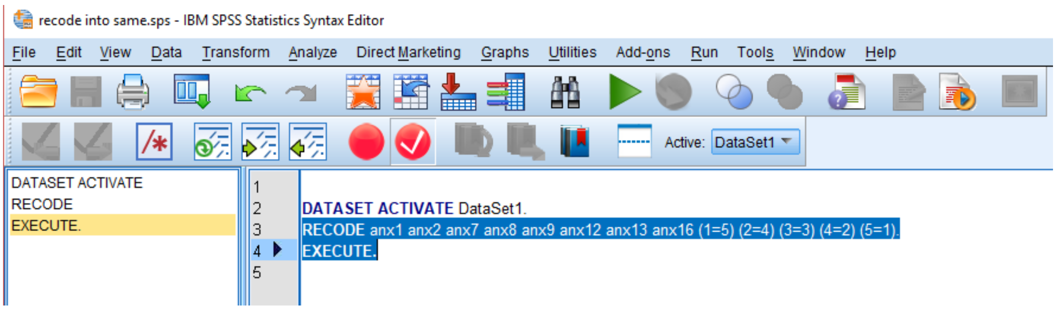
RECODE anx1 anx2 anx7 anx8 anx9 anx12 anx13 anx16 (1=5) (2=4) (3=3) (4=2) (5=1).

EXECUTE.

Na primeru ove kratke sintakse od samo tri linije komandi, možemo uočiti nekoliko stvari. Prvo, vidimo da se svaka linija komandi završava tačkom. U SPSS sintaksi tačka označava kraj komande i obavezna je. Ukoliko zaboravimo da stavimo tačku na kraj komande – ona se neće pokrenuti. Drugo, primećujemo da su određene reči napisane plavom bojom. Plava boja koda označava da se radi o nazivu procedure koju treba primeniti. Treće, vidimo da su sve procedure napisane velikim slovima. Ovo, zapravo, nije važno, jer SPSS sintaksa nije osetljiva na mala i velika slova, tako da sve komande (kao i imena varijabli) možemo pisati i malim i velikim slovima. Na kraju, s obzirom na to da znamo tačno šta ovaj niz komandi treba da predstavlja (jer smo ga prethodno definisali preko korisničkog interfejsa), lako možemo „dešifrovati" značenje same sintakse.

Prva linija komandi – DATASET ACTIVATE – odnosi se na aktiviranje baze podataka i obično bude prva linija koju SPSS generiše prilikom pravljenja nove sintakse na osnovu komande *Paste*. Ova komanda nije neophodna da bi sintaksa radila (možete je slobodno i obrisati). Druga linija komandi je RECODE linija i ona daje komandu da se varijable koje su pobrojane u nastavku komande rekodiraju. Struktura komande je takva da se prvo daje ključna reč (RECODE), zatim se, sa razmacima, navode varijable koje treba transfromisati i na kraju pravila za transformaciju. Svako pravilo dato je u zasebnoj zagradi, a brojevi unutar zagrade odgovaraju starim vrednostima (sa leve strane znaka jednakosti) i novim vrednostima (sa desne strane znaka jednakosti). Poslednja linija koda je samo jedna reč – EXECUTE. Komanda EXECUTE je neophodna za sve transformacije na podacima i ona SPSS-u govori da izvrši transformacije koje su predviđene prethodnim linijama komandi.

Ukoliko želimo da pokrenemo sintaksu i izvršimo željenu transformaciju imamo nekoliko opcija. U okviru menija *Run* imamo opcije *Run All*, *Run Selection* i *Run To End*. *Run All* pokreće sve komande koje su ispisane u datoj sintaksi. U našem slučaju, pošto imamo samo jednu transformaciju podataka, opcija *Run All* će pokrenuti samo nju, ali generalno ovu opciju treba koristiti oprezno jer se može desiti da pokrenemo i komande koje nismo želeli (i koje mogu transformisati podatke na način na koji u datom trenutku ne želimo). Bolje je zato koristiti opciju *Run Selection*, koja pokreće samo deo sintakse koji smo označili mišem. Umesto da kliknemo na *Run Selection* u sintaksnom meniju, nakon što označimo željeni deo sintakse, možemo kliknuti i na dugme sa zelenom strelicom ili na tastaturi pritisnuti **CTRL + R** (Slika 10.3). Ovo pokreće komandu koja rekodira vrednosti varijabli, ali ne proizvodi nikakav ispis u SPSS-u. Zato je važno da vodimo računa da komandu pokrenemo samo jedanput, jer ukoliko je pokrenemo dva puta, zapravo ćemo u varijable „vratiti” sirove skorove umesto rekodiranih.



Slika 10.3

Da bismo proverili da li su se varijable rekodirale na način na koji smo želeli, možemo ponovo izračunati korelacije između njih. Korelacije bi sada trebalo da su sve pozitivne, jer se veći skorovi na svim varijablama odnose na veću izraženost anksioznosti. Kao što pokazuje Slika 10.4, sve korelacije su istog intenziteta kao i pre rekodiranja (Slika 10.1), ali su sve pozitivne, pa zaključujemo da je komanda uspešno izvršena.

		Negativno formulisana stavka	Negativno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Pozitivno formulisana stavka	Negativno formulisana stavka	Negativno formulisana stavka	Negativno formulisana stavka
Negativno formulisana stavka	Pearson Correlation	1	.311**	.400**	.081	.345**	.250*	-.023	.310**	.307**
	Sig. (2-tailed)		.002	.000	.438	.001	.016	.824	.003	.003
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.311**	1	.343**	.135	.084	-.030	.059	.234*	.151
	Sig. (2-tailed)	.002		.001	.196	.423	.772	.572	.024	.149
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	.400**	.343**	1	.056	.184	.097	-.007	.174	.059
	Sig. (2-tailed)	.000	.001		.595	.077	.353	.949	.096	.574
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	.081	.135	.056	1	.056	.206*	.282**	.248*	.002
	Sig. (2-tailed)	.438	.196	.595		.593	.048	.006	.017	.983
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	.345**	.084	.184	.056	1	.378**	.173	.486**	.059
	Sig. (2-tailed)	.001	.423	.077	.593		.000	.096	.000	.574
	N	93	93	93	93	93	93	93	93	93
Pozitivno formulisana stavka	Pearson Correlation	.250*	-.030	.097	.206*	.378**	1	.129	.228*	-.018
	Sig. (2-tailed)	.016	.772	.353	.048	.000		.217	.028	.865
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	-.023	.059	-.007	.282**	.173	.129	1	.358**	.157
	Sig. (2-tailed)	.824	.572	.949	.006	.096	.217		.000	.133
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.310**	.234*	.174	.248*	.486**	.228*	.358**	1	.080
	Sig. (2-tailed)	.003	.024	.096	.017	.000	.028	.000		.447
	N	93	93	93	93	93	93	93	93	93
Negativno formulisana stavka	Pearson Correlation	.307**	.151	.059	.002	.059	-.018	.157	.080	1
	Sig. (2-tailed)	.003	.149	.574	.983	.574	.865	.133	.447	
	N	93	93	93	93	93	93	93	93	93

Slika 10.4

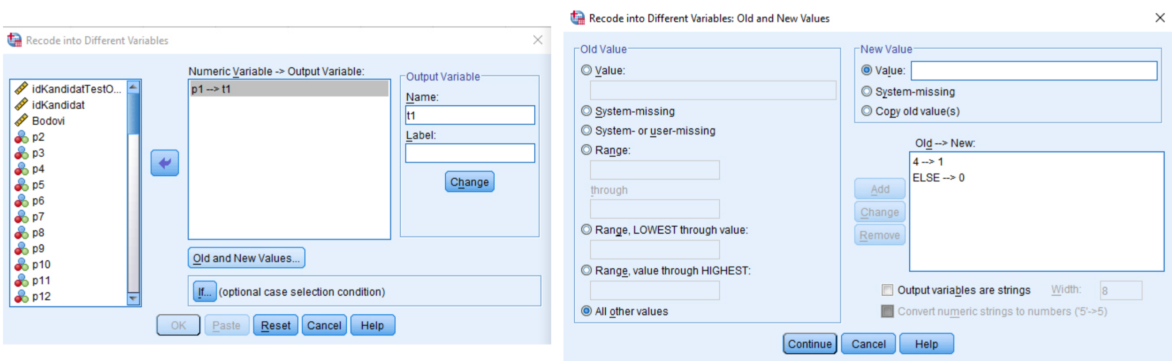
Šta smo u ovom slučaju dobili korišćenjem sintakse? Komande koje smo zadali preko korisničkog interfejsa nisu bile ni dugačke ni komplikovane. Iako, pod uslovom da smo već ovladali ključnim rečima i načinom pisanja sintakse, pa ne moramo da koristimo opciju *Paste*, pisanje dve linije komandi jeste vremenski kraće, ušteda vremena možda nije dovoljna da bi nas motivisala da koristimo sintaksu za ovakve situacije. Ono što dobijamo, međutim, jeste mogućnost da sačuvamo sintaksu, odnosno trag o transformacijama koje smo primenili na podacima. To nam omogućava i da se vratimo na sintaksu ukoliko nam se učini da smo napravili grešku (npr., nije trebalo da rekodiramo varijablu anx7 već anx6), da grešku lako i brzo ispravimo (promenom sintakse), kao i da istu sintaksu pokrenemo ponovo na novom ili dopunjenom fajlu sa podacima. Sintaksu možemo podeliti i sa drugim istraživačima, tako da svi koji primenjuju instrument mogu, pokretanjem samo jedne komande, na ispravan način rekodirati sirove podatke na datom testu.

10.2.2) Rekodiranje u nove varijable

Zamislimo sada drugi primer – test znanja iz psihologije koji se polaže na prijemnom ispitu za upis osnovnih studija psihologije. Ovaj test ima ukupno 30 pitanja od čega prvih 20 pitanja imaju po pet ponuđenih odgovora, od kojih je samo jedan tačan, dok poslednjih 10 pitanja imaju po dva ponuđena odgovora od kojih je samo jedan tačan, dok netačan odgovor donosi negativne poene. Ponovo, očigledno je da ukupan rezultat kandidata ne možemo dobiti na osnovu sirovih skorova, već je prethodno potrebno rekodirati podatke tako da za svako pitanje imamo podatak o tačnosti odgovora, odnosno o broju poena osvojenom za dato pitanje. Ovaj broj poena može biti 1 ili 0 za prvih 20 pitanja, a -1, 0 ili 1 za poslednjih 10.

Koji tip rekodiranja ćemo iskoristiti u ovom slučaju? S obzirom na to da nam informacija o sirovim odgovorima ispitanika može biti značajna u naknadnim analizama (npr., koji netačan odgovor – distraktor su ispitanici najčešće birali, da li su neki distraktori bili previše „očigledni", pa ih niko od ispitanika nije birao, da li je na nekom pitanju većina ispitanika izabrala neki od distraktora umesto tačnog odgovora itd.), korisno bi bilo da ih sačuvamo, a da informacije o tačnosti predstavimo novim varijablama. Drugim rečima, u ovom i sličnim slučajevima izabrali bismo opciju *Recode into Different Variables*.

Rekodiranje u nove varijable se u SPSS-u izvodi preko opcija **Transform → Recode into Different Variables**. Kako pravilo za rekodiranje nije isto za sve varijable (jer tačan odgovor nije uvek na istoj poziciji), rekodiranje počinjemo tako što u polje *Variables* ubacujemo prvu varijablu koje treba rekodirati, a zatim u okviru opcije Old and New values definišemo pravilo za rekodiranje (Slika 10.5). Kako je za prvo pitanje tačan odgovor bio pod 4, u pravilu definišemo da ova vrednost nosi 1 poen, dok sve ostale vrednosti nose 0 poena.



Slika 10.5

Ukoliko umesto opcije *OK* izaberemo *Paste*, u SPSS sintaksi ćemo dobiti ispisane sledeće komande.

RECODE p1 (4=1) (ELSE=0) INTO t1.
EXECUTE.

Vidimo da je struktura sintakse gotovo identična kao kada smo rekodirali u okviru istih varijabli. Prva komanda RECODE definiše koje varijable treba rekodirati i prema kom pravilu. Vrednost sa leve strane znaka jednakosti je vrednost u izvornoj varijabli, dok je vrednost sa desne strane vrednost nove varijable. Dodatna ključna reč u poređenju sa rekodiranjem u okviru istih varijabli jeste komanda *INTO*, koja specifikuje da nove vrednosti treba sačuvati u varijabli čije ime sledi, u našem slučaju kao vrednosti nove varijable t1. Komanda EXECUTE iz drugog reda ponovo govori SPSS-u da izvrši tražene transformacije podataka.

Ukoliko pretpostavimo da je odgovor 4 tačan odgovor ne samo za prvo pitanje, već i za neka druga pitanja (npr., 6, 9, 16 i 19), sintaksu bismo mogli veoma lako dopuniti dodatnim varijablama tako da se rekodiranje izvrši i na njima. Sintaksa bi u tom slučaju izgledala ovako:

RECODE p1 p6 p9 p16 p19 (4=1) (ELSE=0) INTO t1 t6 t9 t16 t19.
EXECUTE.

Naravno, ovom komandom smo kreirali nove varijable samo za ona pitanja na kojima je tačan odgovor bio 4. Potrebno je proširiti sintaksu tako da rekodiramo i ona pitanja gde su tačni odgovori bili označeni brojevima 1, 2, 3 ili 5. U ovom slučaju, kako je pozicija tačnog odgovora drugačija i pravilo za rekodiranje će biti promenjeno. Promena pravila rekodiranja podrazumeva i dodavanje novog reda komandi u SPSS sintaksi. Ovde treba napomenuti da svaka RECODE komanda mora biti praćena komandom EXECUTE, ali ukoliko imamo više RECODE komandi jednu ispod druge – dovoljno je da damo komandu EXECUTE samo jednom. Da bi SPSS ovo prepoznavao kao niz povezanih komandi, međutim, neophodno je da se one nalaze direktno jedne ispod drugih, bez praznih redova, kao u sledećem primeru:

RECODE p11 p13 p17 (1=1) (ELSE=0) INTO t11 t13 t17.
RECODE p3 p7 p8 (2=1) (ELSE=0) INTO t3 t7 t8.
RECODE p2 p10 p12 p15 p18 p20 (3=1) (ELSE=0) INTO t2 t10 t12 t15 t18 t20.
RECODE p1 p6 p9 p16 p19 (4=1) (ELSE=0) INTO t1 t6 t9 t16 t19.
RECODE p4 p5 p14 (5=1) (ELSE=0) INTO t4 t5 t14.
EXECUTE.

Ovim komandama smo obuhvatili prvih 20 stavki testa znanja, na kojima se odgovori daju biranjem jednog od pet ponuđenih odgovora. Pokretanje sintakse će kreirati 20 novih varijabli u kojima se nalaze informacije o tačnosti odgovora ispitanika na ova pitanja u testu.

Na sličan način možemo rekodirati i preostalih 10 pitanja u testu, naravno, uz drugačije, adekvatno pravilo rekodiranja:

RECODE p21 p23 p26 p27 p28 p29 (1=1) (2=-1) (ELSE = 0) INTO t21 t23 t26 t27 t28 t29.
RECODE p22 p24 p25 p30 (1=-1) (2=1) (ELSE = 0) INTO t22 t24 t25 t30.
EXECUTE.

10.3) Sortiranje varijabli

Važno je znati da SPSS kreira nove varijable redosledom kojim su one pomenute u sintaksi, tako da će redosled varijabli u fajlu sa podacima, na osnovu prethodno prikazane sintakse, biti t11, t13, t17, t3, t7... do t14 (za prvih 20 varijabli, odnosno do t30 za sve varijable). Iako može delovati da je redosled varijabli u fajlu sa podacima irelevantan, za pokretanje određenih analiza i interpretaciju dobijenih rezultata može biti praktičnije da su varijable poređane redom koji odgovara redosledu pitanja u testu ili po relevantnim indikatorima ako postoje.

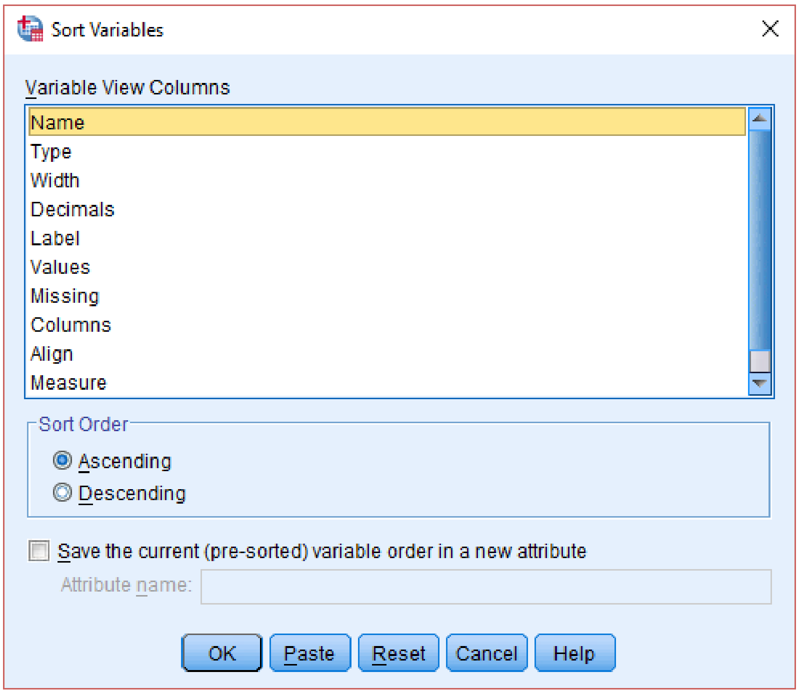
Željeni redosled varijabli možemo postići na nekoliko načina. Najočigledniji način da ovo uradimo jeste da u samom fajlu sa podacima „ručno" promenimo pozicije varijabli tako da one idu od prve do poslednje stavke u instrumentu. Ovo rešenje je, međutim, jednokratno – jer će važiti samo u datom fajlu sa podacima, pa ako budemo u situaciji da na drugom fajlu sa podacima primenjujemo istu proceduru rekodiranja u nove varijable, moraćemo ponovo da promenimo redosled novokreiranih varijabli. Uređivanje redosleda varijabli preko sintakse je zato elegantnije i opštije rešenje.

Jedna opcija je da, umesto da grupišemo sva pitanja koja imaju isti tačan odgovor u jedan red sintakse, imamo po jedan red sintakse za svako pitanje, tako da komande idu redosledom kojim želimo da pitanja budu poređana u fajlu sa podacima. Takva sintaksa bi mogla izgledati ovako:

```
recode p1 (4=1)(ELSE=0) into t1.
recode p2 (3=1)(ELSE=0) into t2.
recode p3 (2=1)(ELSE=0) into t3.
recode p4 (5=1)(ELSE=0) into t4.
recode p5 (5=1)(ELSE=0) into t5.
recode p6 (4=1)(ELSE=0) into t6.
recode p7 (2=1)(ELSE=0) into t7.
recode p8 (2=1)(ELSE=0) into t8.
recode p9 (4=1)(ELSE=0) into t9.
recode p10 (3=1)(ELSE=0) into t10.
execute.
```

Sintaksa se može nastaviti dodatnim varijablama koje želimo da rekodiramo. Primetite da su komande RECODE i EXECUTE napisane malim slovima – kao što smo već rekli, SPSS sintaksni jezik nije osetljiv na veličinu slova tako da će ova sintaksa biti izvršena bez problema.

Druga opcija za postizanje željenog redosleda varijabli jeste da dodatnom komandom promenimo redosled varijabli u fajlu. Ova opcija se u korisničkom interfejsu može naći u okviru **Data → Sort Variables**. U prozoru koji se otvara možemo izabrati po kom kriterijumu želimo da sortiramo varijable (po imenu, tipu, širini, broju decimala itd.), kao i da li želimo da varijable budu sortirane uzlaznim (eng. *ascending*) ili silaznim (eng. *descending*) redosledom (Slika 10.6).



Slika 10.6

Ukoliko umesto *OK* kliknemo na opciju *Paste*, dobijamo sledeću, veoma jednostavnu sintaksu:

SORT VARIABLES BY NAME (A).

Prve dve reči komande – SORT VARIABLES – jesu ključne reči koje definišu samu proceduru sortiranja, dok preostale ključne reči – BY NAME (A) – preciziraju na koji način ova procedura treba da bude izvršena, konkretno prema kom kriterijumu treba sortirati varijable (u našem primeru prema imenu varijabli) i na koji način (u našem primeru uzlaznim redosledom, pošto je A skraćeno od *ascending*). Tamno crvena i narandžasta boja se u SPSS sintaksi odnose na dodatne opcije osnovnih komandi.

Pokretanje sintakse sortiraće varijable željenim redosledom, međutim, opcija sortiranja varijabli po imenu u SPPS-u se odnosi na sortiranje *svih* varijabli u fajlu sa podacima, uključujući i varijable u kojima se nalaze sirovi odgovori ispitanika, kao i druge informacije o njima, na primer, pol, godine itd. Zato ova komanda u nekim situacijama neće u potpunosti odgovarati potrebama istraživača, ali se korišćenjem određenih (naprednijih) komandi može postići željeni rezultat.

Da bismo definisali redosled varijabli u fajlu sa podacima koji u potpunosti odgovara željenom redosledu (a možda se ne poklapa sa sortiranjem po imenu ili tipu varijabli), komanda koju treba primeniti je sledeća:

ADD FILES FILE = * /KEEP= ID to p30 t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 t21 t22 t23 t24 t25 t26 t27 t28 t29 t30 ALL.

Generalno, komanda ADD FILES služi spajanju aktivnog fajla sa podacima sa dodatnim fajlom, u slučaju kada postoje iste varijable a različiti ispitanici u ova dva fajla sa podacima. U našem primeru, međutim, komandu ADD FILES ne koristimo na ovaj način, već dodajemo aktivni fajl sa podacima samom sebi, tako što definišemo da je fajl koji se dodaje aktivni fajl (deo komande koji ovo specifikuje je FILE = *). Ovo neće, naravno, uticati na broj ispitanika (i samo po sebi nema nikakvog efekta na podatke), ali nam daje mogućnost da isključimo neke varijable ili, što nam je sada važnije, definišemo u kom redosledu želimo da zadržimo i organizujemo postojeće varijable. Ovo radimo korišćenjem ključne reči KEEP, a nakon nje navodimo sve varijable koje želimo da zadržimo u fajlu sa podacima, onim redom kojim želimo da ih zadržimo. Zelene reči u SPSS sintaksi odnose se na potkomande za određenu proceduru.

Deo sintakse nakon ključne reči KEEP može se podeliti na tri dela. Prvi deo je definisanje varijabli koje želimo da zadržimo a prethode varijablama koje želimo da sortiramo. U našem primeru to je deo sintakse „ID to p30" – kojim smo definisali da želimo da zadržimo sve varijable od prve varijable u fajlu sa podacima (ID ispitanika) do poslednje varijable koja prethodi informaciji o tačnosti odgovora (p30 – sirovi odgovor na pitanje 30). Drugi deo sintakse definiše varijable koje zadržavamo u fajlu, a čiji redosled želimo da promenimo. Sve ove varijable je potrebno poimence navesti u sintaksi, željenim redosledom. U našem slučaju varijable koje želimo da sortiramo su t1, t2, t3... do t30 i one su upravo tim redosledom i definisane. Na kraju, ključna reč ALL daje komandu da se zadrže (u svom izvornom redosledu) i sve varijable koje slede nakon poslednje varijable čiji redosled smo promenili. U našem primeru, to bi bile sve varijable nakon t30 (ukoliko ih ima).

10.4) Računanje nove varijable

Još jedan nezaobilazni postupak u radu sa bilo kojim psihološkim mernim instrumentom jeste računanje ukupnog ili totalnog skora na tom instrumentu. Ovaj skor se obično iskazuje kao prosek ili suma odgovora ispitanika, ali u nekim slučajevima ukupni skor se može dobiti i na složenije načine. U zavisnosti od tipa instrumenta, nekad ćemo imati samo jedan ukupni skor, dok ćemo u drugim slučajevima pored ukupnog skora za ceo instrument imati i skorove za facete / domene / aspetke instrumenta ili se čak ukupni skor i neće računati već samo skorovi za facete.

Na ovom mestu treba napraviti i jedan kratki osvrt na upotrebu proseka, odnosno sume kao ukupnog skora na instrumentu. Statistički posmatrano, nema nikakve suštinske razlike između korišćenja sume i korišćenja proseka kao mere izraženosti osobine koju psihološki test procenjuje. Kako je prosek samo linearna transformacija sume (delimo svaki skor istom, konstantnom vrednošću – brojem varijabli), korelacija ove dve varijable biće uvek 1. To znači da će poredak ispitanika (individualne razlike između njih) biti identičan u oba slučaja, distribucija skorova će biti ista, kao i korelacije sa drugim varijablama. Da li je onda potpuno svejedno koju od ove dve mere izračunamo? Zapravo, ne sasvim. Iako nije *pogrešno* upotrebiti ni sumu ni prosek u bilo kojoj situaciji, nekad je *prikladnije* upotrebiti jednu umesto druge mere.

Vratimo se na primer testa anksioznosti koji sadrži 16 stavki na koje se odgovori daju na skali od 1 do 5. Ukoliko ukupni skor anksioznosti iskažemo kao sumu odgovora na pojedinačnim stavkama, teorijski minimum skorova će biti 16, dok će teorijski maksimum biti 80. Ukoliko pak skor iskažemo kao prosek, teorijski minimum će biti 1, dok će teorijski maksimum biti 5. Iako između sumarnog skora ispitanika od 56 i prosečnog skora od 3.5 postoji statistička ekvivalencija – oni nisu jednako intuitivni za interpretaciju. Prosečan skor je lakše uporediti sa teorijskim minimumom, maksimumom i prosekom od sumarnog, pa bi u ovoj situaciji upotreba proseka bila opravdanija od upotrebe sume. Takođe, ukoliko u bazi imamo nedostajuće podatke, korišćenjem proseka (uzimajući u obzir za svakog ispitanika samo one stavke na koje jeste odgovorio) možemo tačnije odrediti skor ispitanika nego da smo računali sumarni skor.

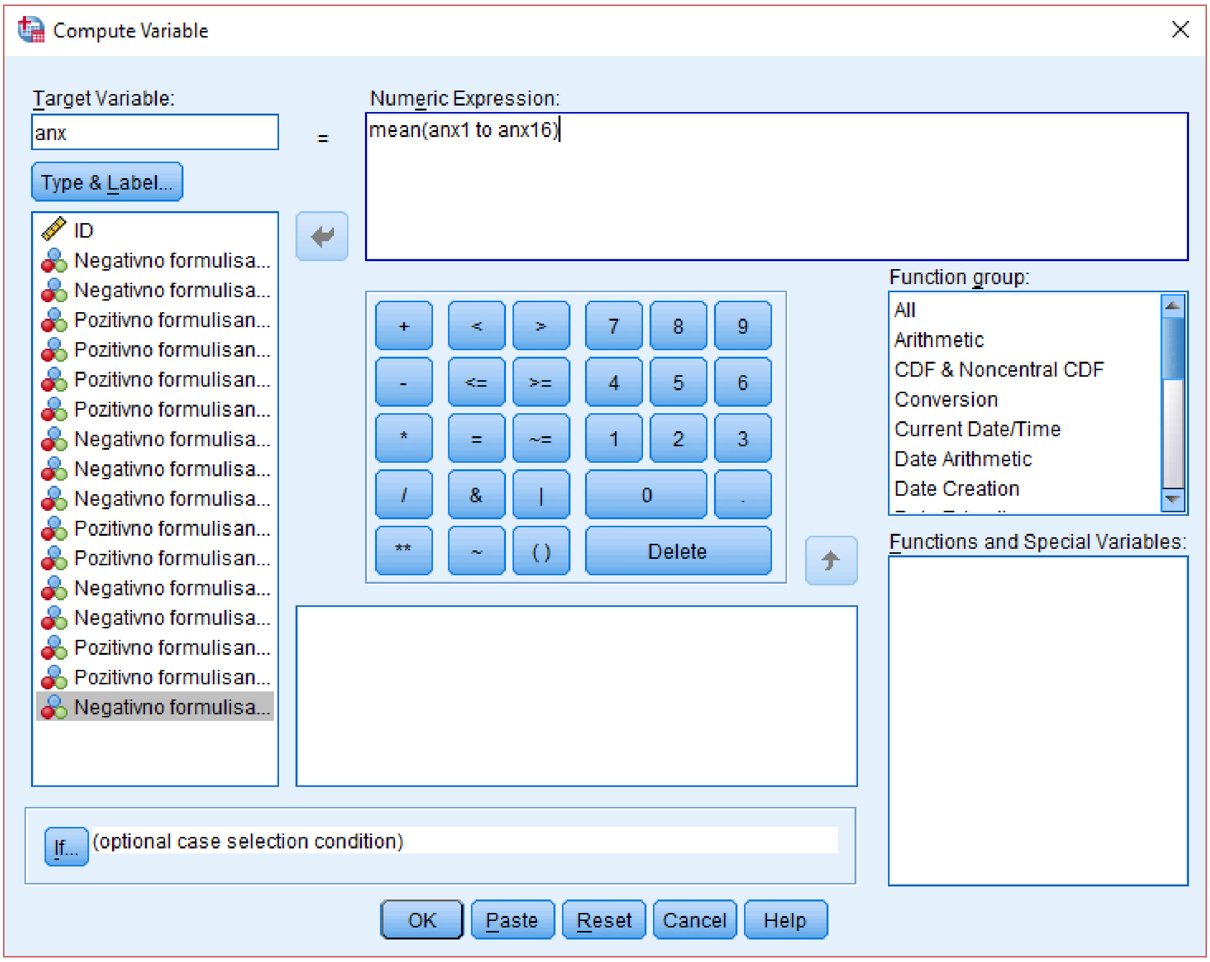
U slučaju testa znanja iz psihologije, prosečan skor na instrumentu govori o proporciji tačnih odgovora koje je svaki ispitanik imao, dok sumarni skor govori o broju pitanja na koja je ispitanik tačno odgovorio. I jedna i druga mera mogu biti pogodne za različite situacije, ali ako (kao što je slučaj u situaciji prijema kandidata na fakultet) želimo da saberemo skor na testu znanja sa nekim drugim skorovima kandidata (prethodni uspeh u školi, uspeh na testu opšte informisanosti), onda će sumarni skor biti smisleniji.

Novu varijablu u SPSS-u pravimo preko komandi **Transform → Compute Variable** (Slika 10.7). U polje *Target Variable* upisujemo ime nove varijable, dok u *Numeric Expression* upisujemo formulu za računanje nove varijable, najčešće na osnovu postojećih varijabli. Na raspolaganju nam je veliki broj predefinisanih funkcija, od koji su za računanje ukupnog skora na instrumentu najvažnije MEAN (prosečna vrednost) i SUM (sumarni skor).

Ponovo, ukoliko umesto *OK*, kliknemo na dugme *Paste*, u sintaksnom prozoru ćemo dobiti sledeću komandu:

```
COMPUTE anx=mean(anx1 to anx16).
EXECUTE.
```

I ova komanda je jednostavna za razumevanje i pamćenje. Prvi red komande odnosi se na računanje nove varijable – ključna reč COMPUTE, pri čemu se sa leve strane znaka jednakosti nalazi ime varijable koju želimo da kreiramo (izračunamo), dok se sa desne strane znaka jednakosti nalazi izraz čiju vrednost treba izračunati. Ukoliko su varijable u fajlu sa podacima poređane redom koji odgovara imenima stavki (od stavke 1 do stavke 16), umesto da navodimo pojedinačne varijable koje želimo da uprosečimo (razdvojene zarezima), možemo iskoristiti ključnu reč *TO* da bismo obuhvatili sve varijable od anx1 do anx16. Kao i kod rekodiranja varijabli, i u slučaju računanja nove varijable potrebno je SPSS-u dati komandu EXECUTE da bi se primenjene transformacije podataka i izvršile.



Slika 10.7

Na primeru testa znanja iz psihologije, komande za računanje sumarnog skora na svim varijablama, odnosno na prvih 20 varijabli izgledale bi ovako:

compute TZ = sum (t1 to t30).

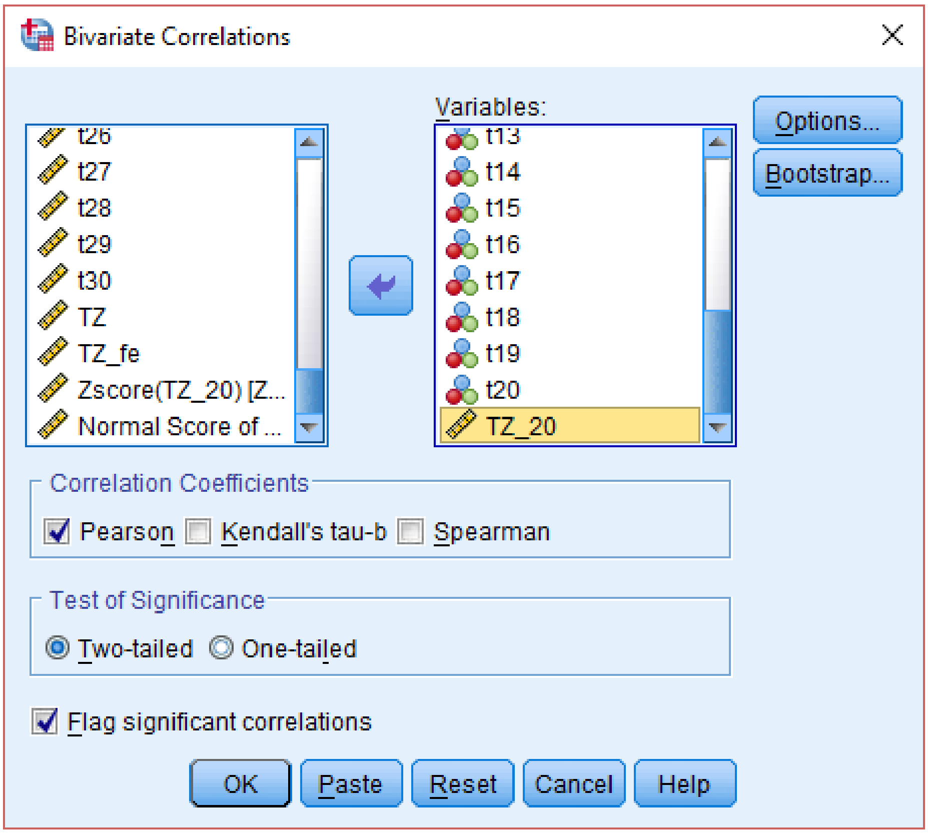
compute TZ_20 = sum (t1 to t20).

execute.

Ponovo primećujemo kako je dovoljno napisati jednu EXECUTE komandu za nekoliko komandi koje želimo da pokrenemo, sve dok se komande nalaze u redovima jedna ispod druge bez praznih redova. Iako je računanje nove varijable jednostavna i brza procedura i kada se koristi korisnički interfejs, u slučaju kada računamo veći broj novih varijabli (ukupnih skorova na facetima instrumenta) upotreba sintakse može biti i brža jer se delovi sintakse mogu kopirati i/ili spojiti.

10.5) Modifikovanje matrice korelacija

Iako ne spada u transformacije podataka, provera korelacija između varijabli predstavlja standardni postupak u većini psiholoških istraživanja. U kontekstu ispitivanja metrijskih karakteristika instrumenta često je važno imati uvid u korelacije pojedinačnih ajtema, kao i njihove korelacije sa ukupnim skorom. Ukoliko želimo da proverimo u kojoj meri odgovori ispitanika na 20 pitanja u testu znanja koreliraju sa ukupnim skorom (samo za prvih 20 varijabli), koristićemo komande **Analyze** → **Correlate** → **Bivariate**, tako što ćemo u polje *Variables* ubaciti prvih 20 pitanja na testu znanja i ukupni skor na prvih 20 pitanja (Slika 10.8).



Slika 10.8

Klik na dugme *Paste* daje sledeću sintaksu:

```
CORRELATIONS
  /VARIABLES=t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 TZ_20
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Iako nešto kompleksnija od prethodnih sintaksnih komandi koje smo prikazali, ni ova komanda nije teška za razumevanje. Prva ključna reč CORRELATIONS ukazuje na analizu koju treba izvršiti, dok potkomande VARIABLES, PRINT i MISSING preciznije definišu analizu. Potkomanda VARIABLES definiše varijable koje treba korelirati, a potkomanda PRINT određuje šta treba da se nađe u ispisu analize. Ključna reč TWOTAIL ukazuje na to da treba prikazati vrednost „dvosmerne” značajnosti, što je podrazumevana opcija[35], dok NOSIG ukazuje na to da u ispisu treba (zvezdicama) obeležiti značajne korelacije. Konačno, potkomanda MISSING definiše na koji način treba tretirati eventualne nedostajuće vrednosti u podacima, pri čemu je PAIRWISE takođe podrazumevana opcija[36]. Možemo primetiti da ova komanda nije praćena ključnom reči EXECUTE. To je slučaj sa

[35] „Dvosmerna” značajnost se odnosi na situaciju kada nemamo unapred određene pretpostavke o smeru korelacije i u tom smislu se smatra eksplorativnom analizom. Ukoliko imamo jasne pretpostavke o smeru povezanosti varijabli, možemo koristiti opciju „jednosmerne” značajnosti. Ona se u SPSS-u dobija tako što se umesto TWOTAIL upiše ključna reč ONETAIL.

[36] U slučaju da ispitanik ima nedostajuću vrednost samo na jednoj varijabli (recimo varijabli 5) u PAIRWISE opciji ovaj ispitanik će biti isključen iz onih korelacija koje uključuju varijablu na kojoj postoji nedostajuća vrednost (npr., korelacija varijabli 1 i 5 ili 10 i 5), dok neće biti isključen iz analiza u kojima se koreliraju varijable na kojima ima podatke (npr., iz korelacije varijabli 1 i 2 ili 10 i 20). Sa druge strane, u LISTWISE opciji ispitanik će biti isključen iz svih analiza.

svim komandama koje se odnose na statističke analize, odnosno koje ne podrazumevaju direktne transformacije podataka.

Tabela koju dobijamo pokretanjem ove analize je puna matrica korelacija, koja ima onoliko redova i kolona koliko varijabli smo ubacili u analizu. U našem primeru puna matrica korelacija ima 21 red i 21 kolonu (20 skorova na pojedinačnim pitanjima i ukupan skor), što je može učiniti nedovoljno preglednom (u slučaju većeg broja varijabli tabela postaje još nepreglednija). Ukoliko nas zanima samo korelacija pitanja sa ukupnim skorom, a ne i korelacija pitanja jednih sa drugim (što je čest slučaj), informativan nam je zapravo samo jedan red, odnosno jedna kolona ove tabele (poslednji red / kolona). SPSS sintaksa pruža mogućnost da u ispisu dobijemo samo ovaj red / kolonu. To se postiže korišćenjem ključne reči WITH.

```
CORRELATIONS
  /VARIABLES=t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 WITH TZ_20
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Ukoliko pokrenemo ovu sintaksu, u ispisu dobijamo matricu korelacija koja ima 20 redova i jednu kolonu (radi preglednosti, Slika 10.9 prikazuje samo isečak tabele u ispisu), kao što je sintaksom bilo i definisano jer je 20 varijabli prethodilo ključnoj reči WITH, dok je samo jedna varijabla sledila nakon nje.

Ukoliko bismo obrnuli redosled varijabli pre i nakon ključne reči WITH, umesto tabele koja ima 20 redova i jednu kolonu dobili bismo tabelu sa 20 kolona i jednim redom (Slika 10.10 prikazuje isečak tabele). U ovom slučaju smo, dodatno, umesto pisanja pojedinačnih varijabli, korišćenjem komande TO skratili sintaksu, obuhvatajući i dalje sve pojedinačne stavke od t1 do t20.

```
CORRELATIONS
  /VARIABLES= TZ_20 WITH t1 to t20
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

		TZ_20
t1	Pearson Correlation	.430**
	Sig. (2-tailed)	.000
	N	436
t2	Pearson Correlation	.599**
	Sig. (2-tailed)	.000
	N	436
t3	Pearson Correlation	.602**
	Sig. (2-tailed)	.000
	N	436
t4	Pearson Correlation	.164**
	Sig. (2-tailed)	.001
	N	436
t5	Pearson Correlation	.434**
	Sig. (2-tailed)	.000
	N	436
t6	Pearson Correlation	.535**
	Sig. (2-tailed)	.000
	N	436
t7	Pearson Correlation	.490**
	Sig. (2-tailed)	.000
	N	436
t8	Pearson Correlation	.496**
	Sig. (2-tailed)	.000
	N	436
t9	Pearson Correlation	.345**
	Sig. (2-tailed)	.000
	N	436
t10	Pearson Correlation	.532**
	Sig. (2-tailed)	.000
	N	436
t11	Pearson Correlation	.416**
	Sig. (2-tailed)	.000
	N	436
t12	Pearson Correlation	.466**
	Sig. (2-tailed)	.000
	N	436

Slika 10.9

Correlations

		t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15	t16	t17
TZ_20	Pearson Correlation	.430**	.599**	.602**	.164**	.434**	.535**	.490**	.496**	.345**	.532**	.416**	.466**	.345**	.315**	.647**	.463**	.489**
	Sig. (2-tailed)	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	436	436	436	436	436	436	436	436	436	436	436	436	436	436	436	436	436

**. Correlation is significant at the 0.01 level (2-tailed).

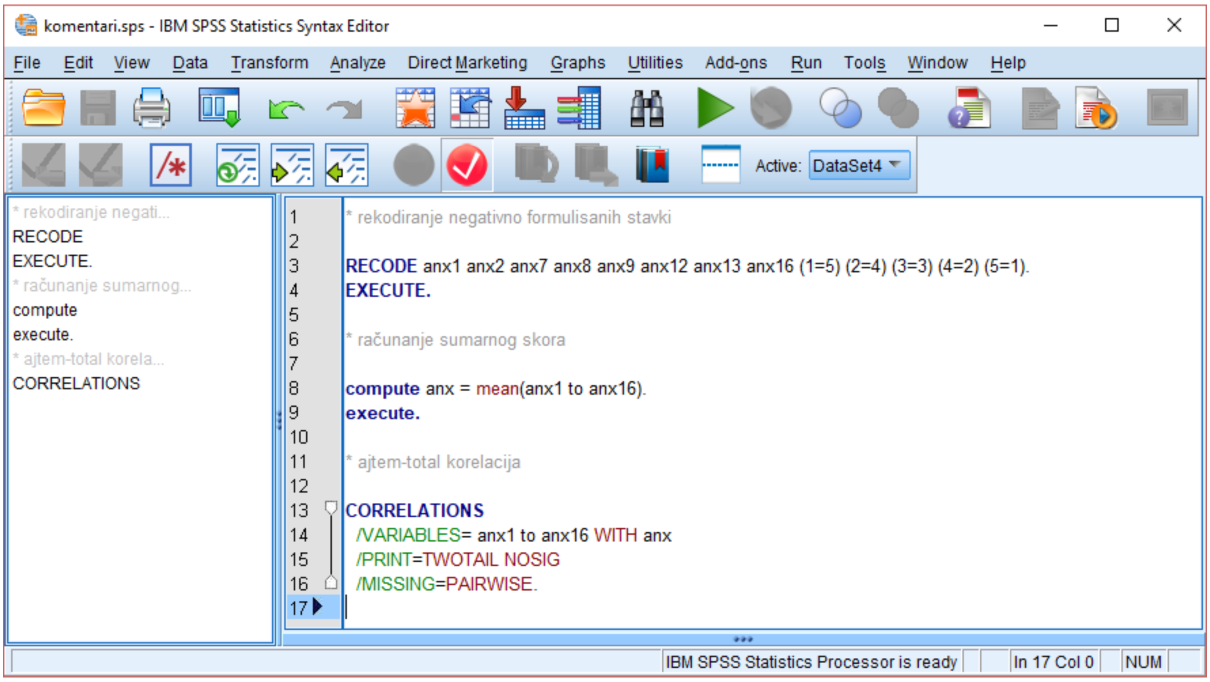
Slika 10.10

Korišćenjem komande WITH, dakle, dobili smo pregledniju i informativniju tabelu korelacija. Jednom malom izmenom u sintaksi bitno smo olakšali čitanje i interpretiranje navedene tabele korelacija. Ova opcija, pritom, nije dostupna iz korisničkog interfejsa, već se može definisati samo sintaksno. Komanda WITH može biti korisna i u situacijama kada želimo da koreliramo dva skupa varijabli (npr., prediktorski i kriterijumski skup ili dve grupe varijabli koje procenjuju iste dimenzije, ali su dobijene drugačijim načinom merenja). U tom slučaju jedan od skupova navodimo pre ključne reči WITH, dok drugi skup navodimo nakon nje.

10.6) Pisanje komentara u sintaksi

Kao što je već pomenuto, čuvanje sintakse nam omogućava da transformacije podataka i analize koje smo uradili sačuvamo, vratimo im se u nekom drugom trenutku ili ih podelimo sa drugim istraživačima. Da bi se druge osobe lakše snašle u sintaksi, ali i da bismo se mi sami podsetili toga šta smo i zbog čega uradili – pored komandi koje su ispisane u SPSS-u možemo upisati i komentare, odnosno dodatna pojašnjenja za sebe i druge.

Komentari se pišu tako što se na početak reda stavi simbol *, a zatim se u nastavku reda pišu komentari. SPSS će komentare obojiti sivo, što označava da se radi o „neaktivnom” delu sintakse, dok su „aktivni” delovi crni, plavi, zeleni, bordo ili narandžasti. Aktivni deo sintakse treba da bude odvojen jednim praznim redom od komentara (jer se prvi red ispod komentara takođe tretira kao neaktivan). Primer sintakse sa komentarima prikazan je na Slika 10.11.



Slika 10.11

10.7) Rezime

Pre nego što pristupimo analizi metrijskih karaketristika instrumenta ili bilo kojoj drugoj statističkoj analizi, neophodno je da pripremimo bazu podataka tako što ćemo izvršiti odgovarajuće transformacije. Najveći broj psiholoških mernih instrumenata podrazumeva rekodiranje određenog broja varijabli – bilo kako bi sve stavke instrumenta ukazivale na prisustvo (a ne odsustvo) merenog konstrukta, bilo kako bi se uvažile informacije o tačnosti odgovora. U SPSS-u za rekodiranje varijabli koristimo opcije *Recode into Same Variables*, odnosno *Recode into Different Variables*, u zavisnosti od toga da li želimo da nove vrednosti sačuvamo u istim ili u novim varijablama. Računanje ukupnog skora, bilo da ga izražavamo kao sumu ili kao prosek skorova na pojedinačnim stavkama, spada u najčešće primenjivane transformacije podataka, a u SPSS-u se postiže korišćenjem opcije *Compute*.

Sve pomenute tranformacije varijabli, kao i mnoge druge, možemo izvršiti korišćenjem korisničkog interfejsa, ali i korišćenjem SPSS sintakse. Korišćenje sintakse ima puno važnih prednosti nad korišćenjem interfejsa: brzinu, ekonomičnost, mogućnost čuvanja i deljenja sintakse, kao i podešavanje dodatnih opcija koje inače ne bi bile dostupne. Kako su komande za osnovne transformacije podataka veoma jednostavne, mogu se lako i brzo usvojiti i koristiti kao rutinski deo istraživačkog rada.

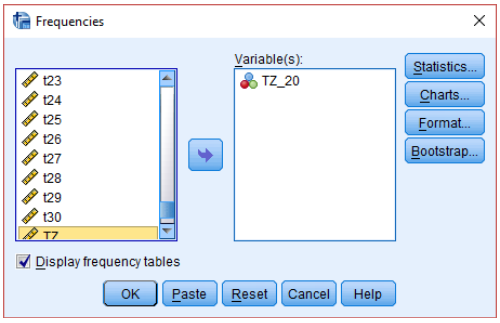
10.8) Preporučena literatura

IBM Corporation. (2021). *IBM SPSS Statistics 21 Command Syntax Reference*.

11) DODATAK 2 – DOPUNSKE ANALIZE, POSTUPCI I OBRASCI

11.1) Provera diskriminativnosti Hi-kvadrat testom u SPSS-u

Kao što je već rečeno, automatska provera odstupanja distribucije od normalne na osnovu Hi-kvadrat testa nije implementirana u SPSS, ali se može izvršiti u nekoliko jednostavnih koraka. Što se distribucije opserviranih frekvencija tiče, ona nam je već poznata, a možemo je prikazati u okviru **Analyze → Descriptive Statistics → Frequencies** tako što ćemo štiklirati (odnosno ostaviti štikliranu) opciju *Display Frequency Tables* (Slika 11.1 i Slika 11.2).



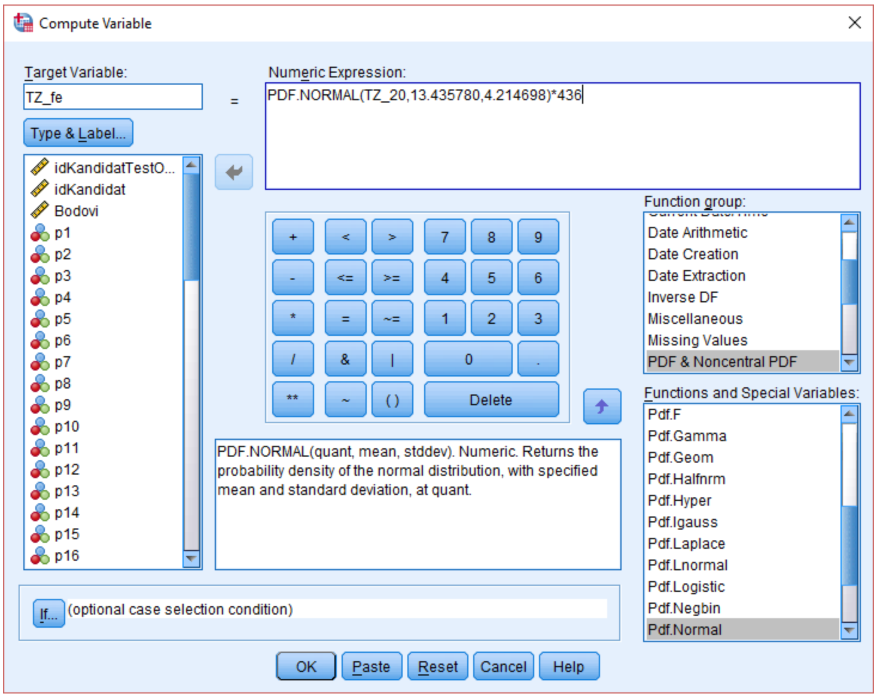
Slika 11.1

TZ_20

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	1	.2	.2	.2
	4	3	.7	.7	.9
	5	8	1.8	1.8	2.8
	6	12	2.8	2.8	5.5
	7	17	3.9	3.9	9.4
	8	28	6.4	6.4	15.8
	9	26	6.0	6.0	21.8
	10	29	6.7	6.7	28.4
	11	25	5.7	5.7	34.2
	12	31	7.1	7.1	41.3
	13	27	6.2	6.2	47.5
	14	27	6.2	6.2	53.7
	15	42	9.6	9.6	63.3
	16	31	7.1	7.1	70.4
	17	34	7.8	7.8	78.2
	18	41	9.4	9.4	87.6
	19	34	7.8	7.8	95.4
	20	20	4.6	4.6	100.0
	Total	436	100.0	100.0	

Slika 11.2

Teorijsku distribuciju skorova na testu možemo izračunati ili ručno, na osnovu tablice površine slučajeva pod normalnom krivom za dati broj razreda (skorova), ili korišćenjem funkcije *PDF.normal* u okviru **Transform → Compute Variable** dijaloga. PDF.normal funkcija (skraćeno od *probability density function*) daje verovatnoću pojavljivanja datog skora za normalnu distribuciju poznate aritmetičke sredine i standardne devijacije. Zato, pre nego što kreiramo novu varijablu sa teorijskim frekvencama, treba da korišćenjem opcija **Analyze → Descriptive Statistics → Frequencies** ili **Analyze → Descriptive Statistics → Descriptives** utvrdimo koliko iznose aritmetička sredina i standardna devijacija naše empirijske distribucije. Već znamo da je aritmetička sredina skorova na testu znanja 13.44, dok standardna devijacija iznosi 4.22. Bitno je i da znamo koliko ispitanika imamo u bazi, što je u ovom slučaju 436 kandidata. PDF.normal funkciju koristimo, što Slika 11.3 prikazuje, tako što prvo unesemo ime varijable za čije skorove želimo da dobijemo verovatnoću pojavljivanja u slučaju da je distribucija normalna (u našem slučaju varijablu TZ_20), zatim aritmetičku sredinu date varijable (13.44, ovde datu sa većim brojem decimala radi veće preciznosti) i njenu standardnu devijaciju (4.22, opet navedenu sa većim brojem decimala). Kao što joj ime kaže – PDF.normal daje *verovatnoću*, odnosno *procenat* slučajeva, dok su nama potrebne frekvence. Zato ćemo ceo izraz pomnožiti brojem slučajeva, odnosno veličinom uzorka (u slučaju naših podataka to je 436). Novu varijablu nazvali smo TZ_fe (*expected frequency*).



Slika 11.3

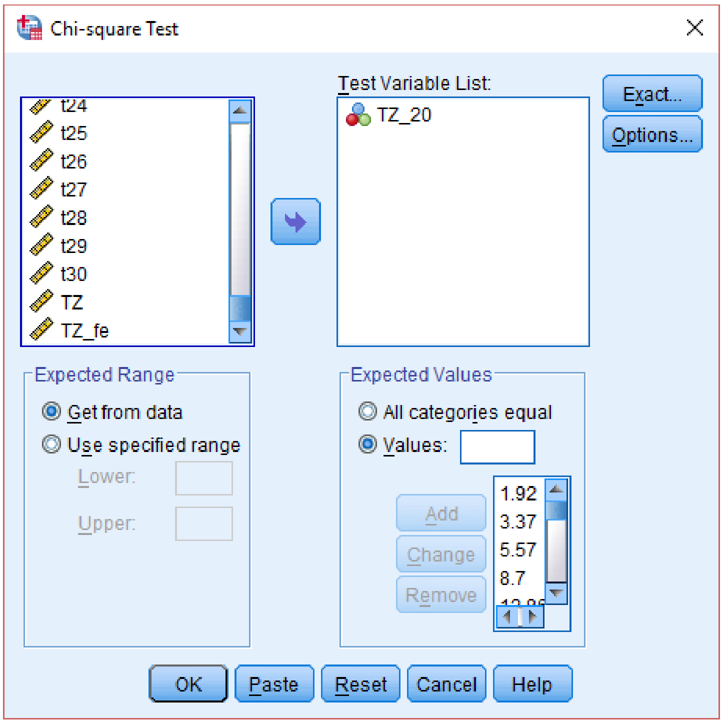
Teorijske frekvence koje smo ovako izračunali možemo videti u samom fajlu sa podacima (Slika 11.4). Vrednosti varijable TZ_fe su očekivane frekvence za varijablu TZ_20. Drugim rečima, skor 3, koji se u podacima pojavljuje samo jednom (što vidimo i iz samih podataka), trebalo bi da se pojavi 1.92 puta u slučaju normalne raspodele, dok bi skor 4 koji se pojavljuje tri puta, trebalo da se pojavi 3.37 puta itd. Ono što nam je potrebno za računanje hi-kvadrata su jedinstvene vrednosti očekivanih frekvenci, dakle, vrednosti 1.92, 3.37, 5.57, 8.70, 12.86 itd. do poslednje jedinstvene vrednosti. Drugačije rečeno, zanemarujemo koliko puta se određena teorijska frekvenca pojavljuje u podacima (jer je to njena empirijska frekvenca), već samo tražimo vrednosti očekivanih frekvenci za sve skorove koji se pojavljuju u empirijskoj distribuciji. Pošto niko od kandidata nije imao manje od 3 poena, to je ukupno 18 različitih vrednosti. U većini slučajeva, čak i kada postoje određena odstupanja od normalnosti, moći ćemo da primetimo kako očekivane frekvence rastu do neke tačke (empirijske aritmetičke sredine), a zatim počinju da opadaju. Ravnomernost ovog rasta i opadanja određuje stepen odstupanja distribucije od normalne.

TZ_20	TZ_fe
3	1.92
4	3.37
4	3.37
4	3.37
5	5.57
5	5.57
5	5.57
5	5.57
5	5.57
5	5.57
5	5.57
5	5.57
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
6	8.70
7	12.86
7	12.86
7	12.86
7	12.86

Slika 11.4

Hi-kvadrat test nalazi se u okviru **Analyze → Nonparametric Test → Legacy Dialogues → Chi-square** ili **Analyze → Nonparametric Test → One Sample**. U slučaju da koristimo **Legacy Dialogues** u polje *Test Variable List* ubacujemo varijablu čiju distribuciju želimo da testiramo (u ovom primeru, to je varijabla TZ_20), a u delu gde se definišu očekivane vrednosti

– *Expected Values* unosimo redom teorijske frekvence koje smo prethodno izračunali, od najnižeg do najvišeg skora (Slika 11.5). Veoma je važno uneti sve teorijske frekvence, i to odgovarajućim redosledom, kako bi računanje Hi-kvadrat statistika bilo adekvatno.



Slika 11.5

U ispisu ove analize dobijamo dve tabele (Slika 11.6). Prva tabela daje pregled empirijskih i očekivanih frekvenci za sve empirijski postojeće skorove. Možemo primetiti da su empirijske frekvence iste one koje dobijamo u ispisu **Analyze → Descriptive Statistics → Frequencies**, dok su očekivane frekvence one koje smo izračunali preko PDF.normal funkcije i uneli u *Expected Values* u SPSS-u.

TZ_20

	Observed N	Expected N	Residual
3	1	2.0	-1.0
4	3	3.6	-.6
5	8	5.9	2.1
6	12	9.2	2.8
7	17	13.6	3.4
8	28	18.9	9.1
9	26	25.0	1.0
10	29	31.2	-2.2
11	25	36.8	-11.8
12	31	41.0	-10.0
13	27	43.3	-16.3
14	27	43.1	-16.1
15	42	40.6	1.4
16	31	36.1	-5.1
17	34	30.4	3.6
18	41	24.2	16.8
19	34	18.2	15.8
20	20	12.9	7.1
Total	436		

Test Statistics

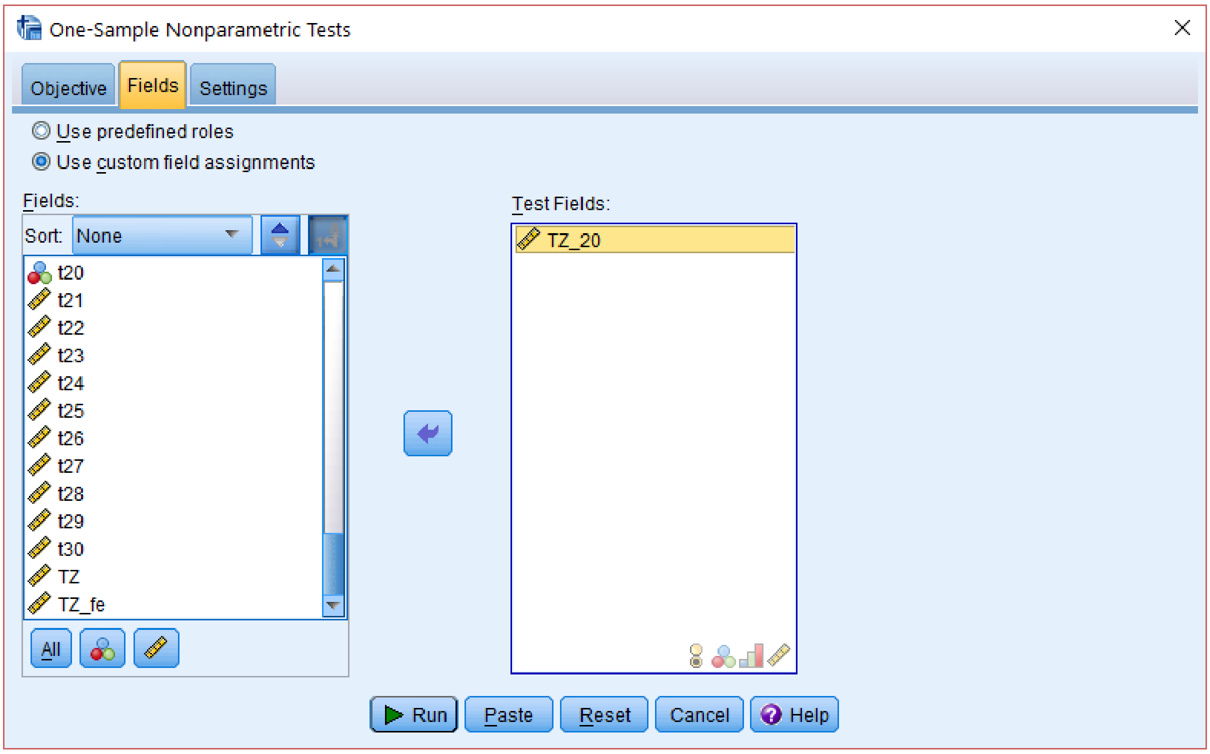
	TZ_20
Chi-Square	56.502[a]
df	17
Asymp. Sig.	.000

Slika 11.6

Pored ovoga, prva tabela u ispisu nam daje i rezidualne frekvence, odnosno razliku između teorijske i empirijske frekvence za svaki pojedinačni skor. Ukoliko bismo ove vrednosti kvadrirali i podelili očekivanim frekvencama, a zatim sumirali količnike – dobili bismo vrednost hi-kvadrata. Ova vrednost, zajedno sa brojem stepeni slobode i značajnošću nalazi se u drugoj tabeli ispisa. S obzirom na to da je vrednost hi-kvadrata za našu raspodelu 56.502, što je značajno na nivou .001 za 17 stepeni slobode, možemo zaključiti da raspodela značajno odstupa od normalne.

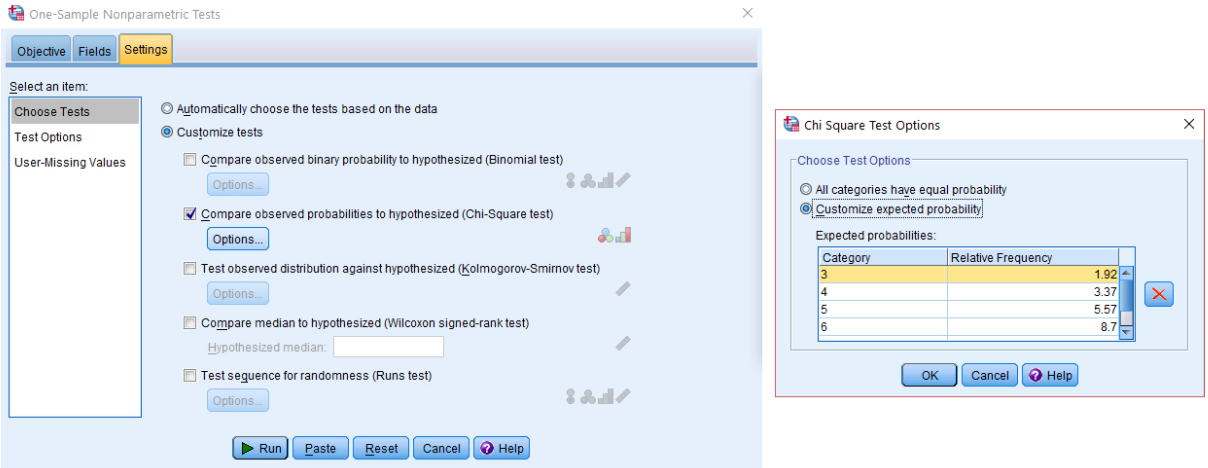
Ono što nam ovaj ispis za hi-kvadrat test ne daje kao informaciju jeste na koji način distribucija odstupa od normalne. Na osnovu pregleda empirijskih, teorijskih i rezidualnih frekvenci možemo naslutiti da bi bilo očekivano da postoji nešto veći broj rezultata u središnjem delu raspodele, oko aritmetičke sredine, i nešto manji na repovima, posebno na desnom repu distribucije. Verovatno je već jasno da su ovo iste informacije koje smo dobili na osnovu uvida u histogram (grafički prikaz empirijske distribucije frekvenci koji omogućava lakšu vizuelizaciju). Samim tim, one ukazuju na platikurtičnu i negativno asimetričnu distribuciju, što smo već i zaključili na osnovu drugih pokazatelja. Ipak, da bismo bili sigurni o kojoj vrsti odstupanja od normalnosti se radi, potrebno je da pogledamo vrednosti skjunisa i kurtozisa.

Drugi način da dođemo do vrednosti hi-kvadrat test statistika u SPSS-u jeste preko opcije **Analyze → Nonparametric Test → One Sample**. U prvoj kartici, *Objective*, treba izabrati opciju da sami podesimo analizu koju želimo da izaberemo, odnosno *Customize analysis*. U sledećoj kartici, *Fields*, u okviru polja *Test Fields* već se nalaze sve varijable koje postoje u bazi sa podacima. Kako ne želimo da testiramo normalnost svih ovih varijabli, treba da izbacimo varijable koje nisu fokus analize, a ostavimo samo one čiju normalnost distribucije želimo da testiramo (u našem slučaju, to je varijabla TZ_20). Kada ovo uradimo, opcija *Use predefined roles* će se automatski promeniti u *Use custom field assignments* (Slika 11.7).



Slika 11.7

U trećoj kartici, Settings, ponovo želimo da sami izaberemo testove koje ćemo pokrenuti, te biramo opciju *Customize tests*. Biramo hi-kvadrat test i u *Options* definišemo teorijske frekvence za sve empirijske skorove u podacima preko opcije *Customize expected probability* (Slika 11.8). Ovo radimo na sličan način kao i kod prethodno opisane metode, s tim što frekvence ne moraju nužno biti unete redom, već je za svaku teorijsku frekvencu potrebno reći na koju se kategoriju, razred, odnosno skor odnosi.



Slika 11.8

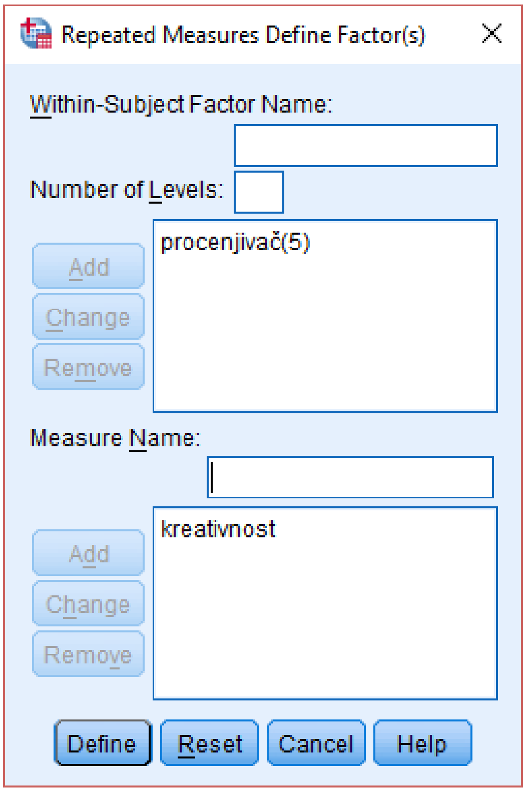
Ispis ove analize dosta je kraći od **Legacy Dialogues** ispisa i daje samo kratki pregled rezultata, odnosno

- nultu hipotezu koja je testirana – u ovom slučaju da distribucija frekvenci odgovara zadatoj (koju smo definisali kao normalnu distribuciju),
- informaciju o tome koji test je primenjen – hi-kvadrat,
- značajnost datog testa – sig = .000 što je identična vrednost kao i u prethodno prikaznom ispisu i
- zaključak koji treba doneti, koji je u ovom slučaju isti prethodno donetom zaključku – da nultu hipotezu treba odbaciti, odnosno da raspodela značajno odstupa od normalne.

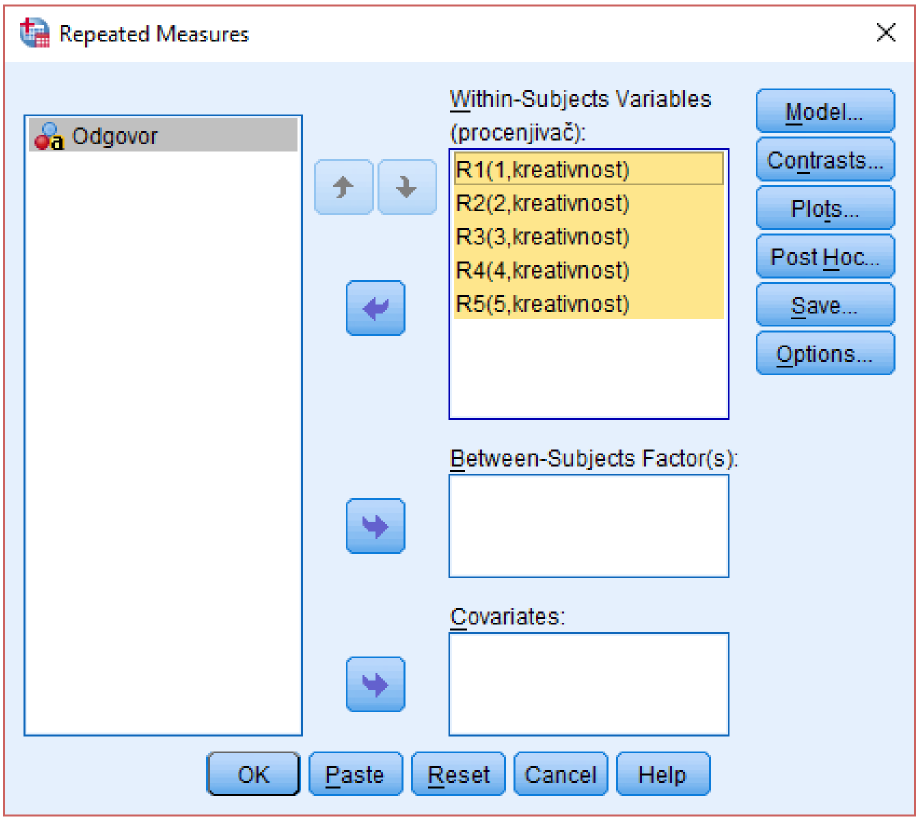
11.2) ANOVA za ponovljena merenja

Ukoliko želimo da procenimo da li su se procenjivači značajno razlikovali u svojim procenama, pod uslovom da su procene date na intervalnoj skali – upotrebićemo analizu varijanse za ponovljena merenja. Nastavićemo da koristimo isti primer sa pet procenjivača koji su ocenjivali odgovore ispitanika na zadatku upotrebe na skali od 1 do 5, korišćen i u glavnom tekstu.

ANOVA za ponovljena merenja se u SPSS-u pokreće preko koraka **Analyze → General Linear Model → Repeated Measures**. Pre svega, potrebno je definisati ponovljeni faktor (*Within-Subject Factor*) i zavisnu varijablu (*Measure*). U ovom slučaju, ponovljeni faktor je **procenjivač** i on ima pet nivoa, pošto imamo pet procenjivača. Zavisnu varijablu nazvaćemo **kreativnost** jer se procene procenjivača odnose na kreativnost odgovora ispitanika (Slika 11.9). U sledećem koraku (Slika 11.10) ubacujemo varijable sa procenama naših pet procenjivača kao nivoe nezavisne varijable u polju *Within-Subject Variables*, nakon čega možemo pokrenuti analizu.



Slika 11.9



Slika 11.10

U ispisu dobijamo nekoliko tabela, od kojih ćemo se trenutno zadržati samo na tabeli **Tests of Within-Subjects Effects**, jer se u njoj nalaze vrednosti F-količnika, koji testira hipotezu da su prosečne ocene svih pet procenjivača međusobno jednake. Posmatrajući red Greenhouse-Geisser[37] za faktor procenjivač u ovoj tabeli (Slika 11.11) vidimo da je F-količnik značajan, što znači da su se procenjivači značajno razlikovali u svojim procenama kreativnosti odgovora ispitanika.

Tests of Within-Subjects Effects

Measure: kreativnost

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
procenjivač	Sphericity Assumed	97.024	4	24.256	24.999	.000
	Greenhouse-Geisser	97.024	3.351	28.952	24.999	.000
	Huynh-Feldt	97.024	3.627	26.752	24.999	.000
	Lower-bound	97.024	1.000	97.024	24.999	.000
Error(procenjivač)	Sphericity Assumed	190.176	196	.970		
	Greenhouse-Geisser	190.176	164.210	1.158		
	Huynh-Feldt	190.176	177.716	1.070		
	Lower-bound	190.176	49.000	3.881		

Slika 11.11

Ako dalje želimo da utvrdimo između kojih tačno parova procenjivača postoje razlike, to bismo mogli učiniti uz pomoć unapred definisanih kontrasta (opcija **Contrasts**), dok bismo razlike grafički mogli prikazati dijagramom (opcija **Plots**). Kao što je već rečeno, međutim, razlike između procenjivača ne moraju predstavljati problem ukoliko konačna ocena služi samo tome da se uspostavi poredak objekata procene po stepenu izraženosti nekog svojstva, što je pristup koji bismo u našem primeru i primenili. U slučaju kada nije relevantan samo relativni poredak objekata procene, već i apsolutna vrednost ovih ocena (npr., kada postoje minimalne i maksimalne konačne ocene koje se mogu dodeliti), veoma je važno utvrditi da li neki procenjivač značajno odstupa od ostalih po svojim ocenama. Ukoliko se pokaže da određeni procenjivač bitno smanjuje objektivnost procene, treba razmotriti isključivanje njegovih procena iz dalje analize.

[37] Pošto je pretpostavka o sferičnosti narušena (što ovde nismo prikazivali), koristimo Greenhouse-Geisser korekciju.

11.3) Fridmanov test

Da bismo proverili da li je između tri ili više procenjivača bilo značajne razlike u prosečnim rangovima koje su dodelili plesačima, upotrebićemo Fridmanov test. Analizu ćemo prikazati na primeru članova žirija koji ocenjuju plesače na takmičenju, koji smo koristili da ilustrujemo procenu objektivnosti kada su podaci na ordinalnom nivou. Naravno, ukoliko su svi članovi žirija koristili isti raspon rangova 1–35 (koliko je plesača učestvovalo na takmičenju), i ne treba očekivati razlike u prosečnom rangu. Treba, međutim, imati u vidu da ova pretpostavka ne mora u svakom slučaju biti tačna, kao i da se Fridmanov test koristi i u slučaju intervalnih podataka kada distribucije skorova odstupaju od normalne ili nisu zadovoljeni neki drugi uslovi za primenu parametrijskih tehnika.

Fridmanov test je neparametrijski ekvivalent ANOVA-i za ponovljena merenja, te počiva na razlaganju ukupne varijanse na onu koja potiče od ponovljenog faktora (u ovom slučaju člana žirija) i onu koja potiče od unutargrupnih razlika (u ovom slučaju plesača) (Friedman, 1937).

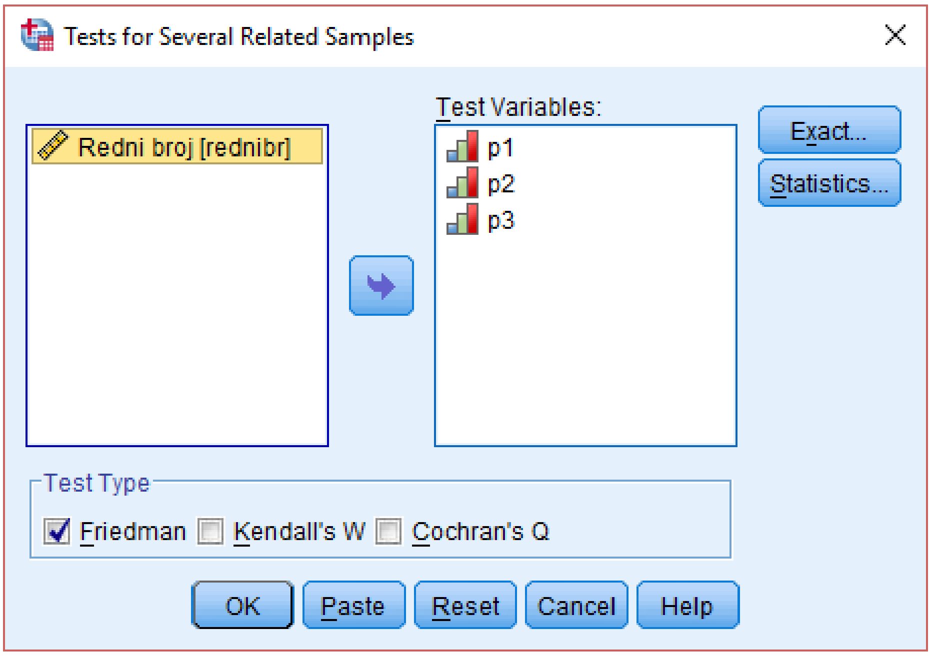
- Prvi korak u računanju Fridmanovog test statistika jeste rangovanje rezultata unutar svakog reda (plesača). Na ovaj način vidimo u kojoj od situacija je vrednost bila najveća, odnosno najmanja (koji od članova žirija je datog plesača ocenio najbolje, a koji najlošije).
- Sledeći korak je računanje prosečnog ranga po kolonama (odnosno prosečnog ranga za svakog člana žirija), prema formuli $\overline{r_j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$, gde je *r* oznaka za rang, *i* oznaka za red (plesača), *j* za kolonu (člana žirija), a *n* broj ispitanika (plesača). Ukoliko pretpostavimo da nema bitnih razlika između članova žirija, sve razlike u rangovima biće proizvod slučajnosti. Drugim rečima, očekivali bismo da prosečni rangovi budu međusobno jednaki.
- Ukoliko nema razlika između procenjivača, i ako je broj redova dovoljno veliki, distribucija rangova u svakoj koloni biće normalna, sa poznatom aritmetičkom sredinom i standardnom devijacijom. Aritmetička sredina ove distribucije je zapravo ukupni prosečni rang (koji se može izračunati kao $\bar{r} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} r_{ij}$, gde je *k* oznaka za broj procenjivača – članova žirija). Ukupan prosečni rang uvek mora biti jednak $\bar{r} = \frac{(k+1)}{2}$, odnosno u našem slučaju sa tri člana žirija, ukupni prosečni rang mora biti 2.
- Pretpostavka o nepostojanju razlika između procenjivača testira se računanjem Fridmanovog statistika koji prati χ^2 distribuciju sa k-1 stepeni slobode.

Najjednostavnija[38] formula za računanje Fridmanovog statistika jeste:

$$Q = \frac{12}{nk(k+1)}\sum_{j=1}^{k}\left(\sum_{i=1}^{n} r_{ij}\right)^2 - 3n(k+1)$$

Fridmanov test se u SPSS-u nalazi u okviru neparametrijskih tehnika namenjenih analizi k povezanih (zavisnih) uzoraka, odnosno u okviru opcija **Analyze → Nonparametric Test → Legacy Dialogues → K related samples**. U polje *Test variables* ubacujemo tri varijable koje se odnose na procene članova žirija, i kao test biramo Fridmanov test (*Test Type – Friedman*) (Slika 11.12).

[38] Ova formula daje iste rezultate kao i kompleksniji oblici, ali ne podrazumeva računanje prosečnih rangova i uključuje samo cele brojeve (Friedman, 1937).



Slika 11.12

U ispisu za Fridmanov test dobijamo dve tabele – tabelu prosečnih rangova po procenjivačima i tabelu χ^2 test statistika koji ukazuje na značajnost razlika između procenjivača. Kao što iz prve tabele možemo videti (Slika 11.13), ukupan prosečni rang za sva tri člana žirija iznosi $\frac{2+2.04+1.96}{2} = 2$, kao što je prethodno i rečeno (u skladu sa formulom $\bar{r} = \frac{(k+1)}{2}$). Ono što takođe možemo primetiti jeste da su prosečni rangovi sva tri procenjivača gotovo identični, što je i očekivano imajući u vidu da su procenjivači sve plesače rangirali dodeljujući im neki od rangova u rasponu 1–35, odnosno svi članovi žirija su koristili isti raspon skorova. U skladu sa ovim je i niska vrednost χ^2 test statistika, koji nije značajan, na osnovu čega možemo zaključiti da se procenjivači nisu značajno razlikovali u prosečnoj oceni koju su dodeljivali plesačima. Naravno, podsećamo da se Fridmanov test može koristiti i kada procenjivači daju odgovore na skali procene intervalnog tipa, u kom slučaju se njihove procene mogu značajno razlikovati.

Friedman Test

Ranks

	Mean Rank
p1	2.00
p2	2.04
p3	1.96

Test Statistics[a]

N	35
Chi-Square	.151
df	2
Asymp. Sig.	.927

a. Friedman Test

Slika 11.13

11.4) Izvođenje Kronbahovog alfa obrasca pouzdanosti iz Spirman–Braunovog obrasca

Spirman–Braunov opšti obrazac pouzdanosti u slučaju deljenja testa na više delova je:

$$r_{nn} = \frac{nr_{xx}}{1+(n-1)r_{xx}}$$

pri čemu je r_{xx} prosečna korelacija između delova testa (u slučaju kada se test deli na više od dva dela). Radi lakšeg razumevanja dokaza prosečnu korelaciju delova testa ćemo notirati kao $\overline{r_{xx}}$, te obrazac izgleda ovako:

$$r_{nn} = \frac{n\overline{r_{xx}}}{1+(n-1)\overline{r_{xx}}}$$

Poznato je da se prosečna korelacija stavki izračunava prema sledećoj formuli:

$$\overline{r_{xx}} = \frac{\sum\sum r_{ij} - n}{n(n-1)}$$

gde je n broj stavki, a $\sum\sum r_{ij}$ zbir korelacija svih mogućih parova stavke (uključujući i korelaciju stavke same sa sobom, kao i odvojeno računanje korelacije stavke a sa stavkom b od korelacije stavke b sa stavkom a). Zamenom ovog izraza u formuli i daljim izvođenjem dobijamo:

$$r_{nn} = \frac{n\overline{r_{xx}}}{1+(n-1)\overline{r_{xx}}} = \frac{n\frac{\sum\sum r_{ij}-n}{n(n-1)}}{1+(n-1)\frac{\sum\sum r_{ij}-n}{n(n-1)}} = \frac{\frac{n(\sum\sum r_{ij}-n)}{n(n-1)}}{\frac{n(n-1)+(n-1)(\sum\sum r_{ij}-n)}{n(n-1)}}$$

$$= \frac{n(\sum\sum r_{ij}-n)}{n(n-1)+(n-1)(\sum\sum r_{ij}-n)} = \frac{n}{n-1}\frac{\sum\sum r_{ij}-n}{n+\sum\sum r_{ij}-n} = \frac{n}{n-1}\frac{\sum\sum r_{ij}-n}{\sum\sum r_{ij}}$$

$$= \frac{n}{n-1}\left(\frac{\sum\sum r_{ij}}{\sum\sum r_{ij}} - \frac{n}{\sum\sum r_{ij}}\right) = \frac{n}{n-1}\left(1-\frac{n}{\sum\sum r_{ij}}\right)$$

što je Kronbahov alfa obrazac u slučaju standardizovanih varijabli.

11.5) Izvođenje formule za produženje testa iz Spirman–Braunovog obrasca

Na osnovu Spirman–Braunovog obrasca pouzdanosti $r_{nn} = \frac{nr_{xx}}{1+(n-1)r_{xx}}$ treba doći do broja puta koliko treba produžiti test da bi se dobila željena pouzdanost $n = \frac{r_{nn}(1-r_{xx})}{r_{xx}(1-r_{nn})}$. U ovom slučaju, značenje izraza u obrascu nešto je drugačije – r_{nn} je željena pouzdanost testa, r_{xx} je trenutna pouzdanost testa, a n je broj puta koliko treba produžiti test. Ovakvo čitanje notacije nije suštinski drugačije od onog koje se standardno koristi, s obzirom na to da se kod računanja pouzdanosti deljenjem testa prosečna korelacija stavki uzima kao procena pouzdanosti datog dela, a dodavanje novog dela testa doprinosi ukupnoj pouzdanosti.

Kako su svi elementi oba obrasca jednaki, vršimo prosto rešavanje gornje jednačine po n:

$$r_{nn} = \frac{nr_{xx}}{1+(n-1)r_{xx}}$$

$$r_{nn}(1+(n-1)r_{xx}) = nr_{xx}$$

$$r_{nn} + r_{nn}(n-1)r_{xx} = nr_{xx}$$

$$r_{nn} + nr_{nn}r_{xx} - r_{nn}r_{xx} = nr_{xx}$$

$$nr_{nn}r_{xx} - nr_{xx} = r_{nn}r_{xx} - r_{nn}$$

$$nr_{xx}(r_{nn} - 1) = r_{nn}(r_{xx} - 1)$$

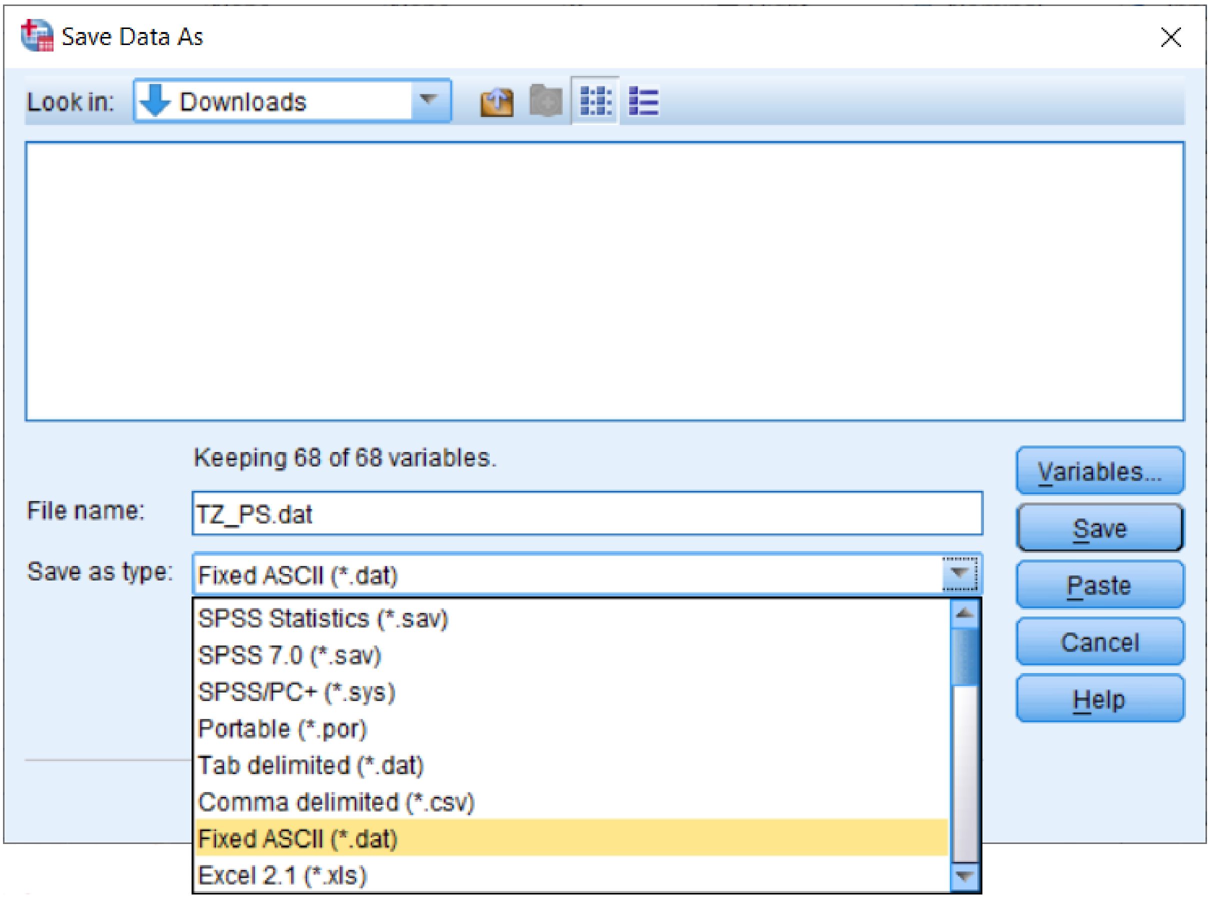
$$n = \frac{r_{nn}(1-r_{xx})}{r_{xx}(1-r_{nn})}$$

čime smo dobili traženi izraz.

11.6) Čuvanje SPSS fajla sa podacima u tekstualnom formatu

SPSS fajlovi sa podacima se mogu lako i jednostavno sačuvati i u različitim drugim formatima. Sada ćemo, koristeći podatke sa prvih 20 pitanja testa znanja iz psihologije demonstrirati kako se .sav fajl može sačuvati kao tekstualni fajl.

Nakon što smo otvorili fajl sa podacima, treba da odemo na opciju **File** → **Save As**. Otvoriće se novi prozor u kome možemo izabrati lokaciju čuvanja fajla, ali i njegov format u okviru opcije *Save as type* (Slika 11.14). Vidimo da nam je na raspolaganju veoma veliki broj formata koji mogu biti korisni u različitim situacijama. Kako nam je za korišćenje u programu Winsteps važno da svaka varijabla uvek zauzima jednak broj kolona, odnosno da broj kolona bude fiksiran, izabraćemo Fixed ASCII (*.dat) tekstualni format. Fajl kasnije možemo otvarati u editoru teksta kao što je Notepad i u .dat formatu, ali ga možemo i preimenovati (i time konvertovati) u .txt fajl bez gubitka informacija.



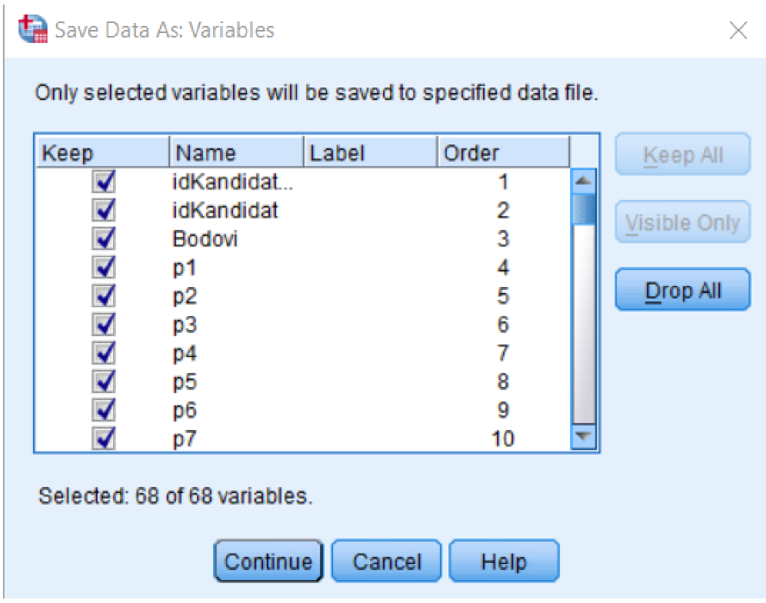
Slika 11.14

Novi fajl koji želimo da sačuvamo trebalo bi da sadrži neke informacije o ispitaniku (poput rednog broja, npr.), kao i odgovore na sve stavke koje želimo da analiziramo. Izvorni SPSS fajl, međutim, sadrži veći broj varijabli od onih koje su nam potrebne – npr., sirove odgovore ispitanika na svim stavkama, odgovore na deset DA/NE pitanja, kao i sumarni skor na testu. Iako ove varijable neće „smetati" prilikom analize, mogu učiniti kontrolni fajl nepreglednim i otežati određivanje položaja/kolone prvog ajtema koji želimo da analiziramo.

Klikom na dugme *Variables*, otvara se novi prozor (Slika 11.15) u kome možemo izabrati koje varijable želimo da sačuvamo u novom fajlu. Podrazumevana opcija je da su sve varijable označene, ali možemo deselektovati (odštiklirati) sve varijable koje ne želimo da izvezemo. U slučaju da je broj varijabli koje nam nisu potrebne u novom fajlu veći od broja varijabli koje želimo da zadržimo –

može biti brže da izaberemo opciju *Drop All* (koja deselektuje sve varijable istovremeno), a zatim označimo samo one varijable koje želimo da sačuvamo.

U našem primeru, izabraćemo drugu varijablu u fajlu sa podacima idKandidat, koja sadrži podatke o rednom broju kandidata, kao i varijable t1 do t20, u kojima se nalazi informacija o tačnosti odgovora na prvih 20 pitanja na testu, odnosno ukupno ćemo selektovati 21 varijablu. Nakon ovoga kliknućemo na *Continue*, a zatim u glavnom prozoru na *Save*.



Slika 11.15

SPSS će novi fajl sačuvati na izabranoj lokaciji, a takođe će kreirati i ispis, u kome se nalaze detaljne informacije o novom fajlu (Slika 11.16). U prvom redu ispisa, dobijamo informaciju o lokaciji gde je novi fajl sačuvan, dok se u trećem redu nalazi informacija o varijabalama koje će novi fajl sadržati.

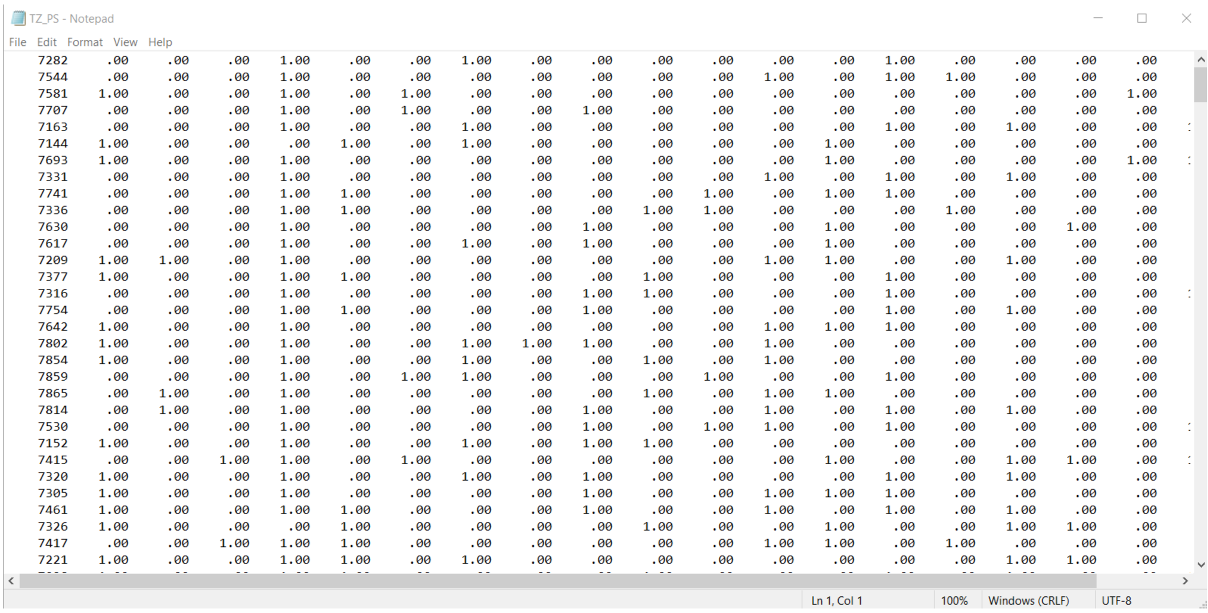
```
WRITE OUTFILE='C:\Users\Danka\Downloads\TZ_PS.dat'
  TABLE
  /idKandidat t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 .

Write will generate the following

Variable          Rec   Start    End  Format
idKandidat          1       1      8  F8.0
t1                  1       9     16  F8.2
t2                  1      17     24  F8.2
t3                  1      25     32  F8.2
t4                  1      33     40  F8.2
t5                  1      41     48  F8.2
t6                  1      49     56  F8.2
t7                  1      57     64  F8.2
t8                  1      65     72  F8.2
t9                  1      73     80  F8.2
t10                 1      81     88  F8.2
t11                 1      89     96  F8.2
t12                 1      97    104  F8.2
t13                 1     105    112  F8.2
t14                 1     113    120  F8.2
t15                 1     121    128  F8.2
t16                 1     129    136  F8.2
t17                 1     137    144  F8.2
t18                 1     145    152  F8.2
t19                 1     153    160  F8.2
t20                 1     161    168  F8.2
EXECUTE.
```

Slika 11.16

Nakon ovoga, ispis sadrži tabelu čiji su redovi varijable koje smo izabrali da sačuvamo u novom fajlu, dok nam kolone daju informaciju o tome kako je novi fajl organizovan, odnosno formatiran. Kolona *Variable* sadrži ime varijable koja je sačuvana, dok kolona *Rec* označava da je varijabla snimljena (eng. *recorded*) u novi fajl i ima vrednost 1 za sve varijable (pošto su sve uspešno sačuvane). Kolone *Start* i *End* daju informaciju o prvoj, odnosno poslednjoj koloni koja se odnosi na datu varijablu, dok kolona *Format* govori o formatu varijable.

Vidimo da format svih varijabli počinje slovom F, što znači da se radi o numeričkim varijablama. Za alfanumeričke varijable (*string*) koristi se početno slovo A, a i svi drugi formati imaju svoje skraćenice (npr., *date* za datumske varijable). Naredne dve cifre odnose se na ukupan broj kolona koje varijabla zauzima i broj decimalnih mesta. U slučaju alfanumeričkih varijabli, naravno, ne postoje decimalna mesta, pa bi format imao samo jedan broj koji govori o broju kolona (npr., A50 ili A15 za string varijable koje imaju 50, odnosno 15 karaktera). Osim prve varijable, svi ajtemi imaju format F8.2, što znači da se radi o numeričkim varijablama koje imaju po 8 kolona, od čega 2 kolone odlaze na decimalna mesta. Ovo je zapravo podrazumevano formatiranje numeričkih varijabli u SPSS-u, međutim, ono nije najpraktičnije za analizu u Winstepsu.

Ukoliko otvorimo fajl sa podacima u Notepad-u (Slika 11.17), vidimo da on ne izgleda „najurednije”. Varijable jesu poravnate, ali svaka varijabla zauzima više kolona, te iako smo sačuvali samo 20 stavki testa, ne možemo istovremeno sve da ih vidimo u prozoru. Takođe, s obzirom na to da su mogući odgovori na testu samo 0 i 1, odnosno jednocifreni brojevi bez decimalnih mesta – ovakav prikaz je nepotrebno nepregledan.


TZ_PS - Notepad

File Edit Format View Help

```
    7282     .00     .00     .00    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00    1.00     .00     .00     .00     .00
    7544     .00     .00     .00    1.00     .00     .00     .00     .00     .00     .00     .00    1.00     .00    1.00    1.00     .00     .00     .00
    7581    1.00     .00     .00    1.00     .00    1.00     .00     .00     .00     .00     .00     .00     .00     .00     .00     .00     .00    1.00
    7707     .00     .00     .00    1.00     .00    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00     .00     .00     .00
    7163     .00     .00     .00    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00    1.00     .00    1.00     .00     .00
    7144    1.00     .00     .00     .00    1.00     .00    1.00     .00     .00     .00     .00     .00    1.00     .00     .00     .00     .00     .00
    7693    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00     .00     .00    1.00     .00     .00     .00     .00    1.00
    7331     .00     .00     .00    1.00     .00     .00     .00     .00     .00     .00     .00    1.00     .00    1.00     .00    1.00     .00     .00
    7741     .00     .00     .00    1.00    1.00     .00     .00     .00     .00     .00    1.00     .00    1.00    1.00     .00     .00     .00     .00
    7336     .00     .00     .00    1.00    1.00     .00     .00     .00     .00    1.00    1.00     .00     .00     .00    1.00     .00     .00     .00
    7630     .00     .00     .00    1.00     .00     .00     .00     .00    1.00     .00     .00     .00    1.00     .00     .00     .00    1.00     .00
    7617     .00     .00     .00    1.00     .00     .00    1.00     .00    1.00     .00     .00     .00    1.00     .00     .00     .00     .00     .00
    7209    1.00    1.00     .00    1.00     .00     .00     .00     .00     .00     .00     .00    1.00    1.00     .00     .00    1.00     .00     .00
    7377    1.00     .00     .00    1.00    1.00     .00     .00     .00     .00    1.00     .00     .00     .00    1.00     .00     .00     .00     .00
    7316     .00     .00     .00    1.00     .00     .00     .00     .00    1.00    1.00     .00     .00     .00    1.00     .00     .00     .00     .00
    7754     .00     .00     .00    1.00    1.00     .00     .00     .00    1.00     .00     .00     .00     .00    1.00     .00    1.00     .00     .00
    7642    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00     .00    1.00    1.00    1.00     .00     .00     .00     .00
    7802    1.00     .00     .00    1.00     .00     .00    1.00    1.00    1.00     .00     .00    1.00     .00     .00     .00     .00     .00     .00
    7854    1.00     .00     .00    1.00     .00     .00    1.00     .00     .00    1.00     .00    1.00     .00     .00     .00     .00     .00     .00
    7859     .00     .00     .00    1.00     .00    1.00    1.00     .00     .00     .00    1.00     .00     .00    1.00     .00     .00     .00     .00
    7865     .00    1.00     .00    1.00     .00     .00     .00     .00     .00    1.00     .00    1.00    1.00     .00     .00     .00     .00     .00
    7814     .00    1.00     .00    1.00     .00     .00     .00     .00    1.00     .00     .00    1.00     .00    1.00     .00    1.00     .00     .00
    7530     .00     .00     .00    1.00     .00     .00     .00     .00    1.00     .00    1.00    1.00     .00    1.00     .00     .00     .00     .00
    7152    1.00     .00     .00    1.00     .00     .00    1.00     .00    1.00    1.00     .00     .00     .00     .00     .00     .00     .00     .00
    7415     .00     .00    1.00    1.00     .00    1.00     .00     .00     .00     .00     .00     .00    1.00     .00     .00    1.00    1.00     .00
    7320    1.00     .00     .00    1.00     .00     .00    1.00     .00    1.00     .00     .00     .00     .00    1.00     .00    1.00     .00     .00
    7305    1.00     .00     .00    1.00     .00     .00     .00     .00    1.00     .00     .00    1.00    1.00    1.00     .00     .00     .00     .00
    7461    1.00     .00     .00    1.00    1.00     .00     .00     .00    1.00     .00     .00    1.00     .00    1.00     .00    1.00     .00     .00
    7326    1.00     .00     .00     .00    1.00     .00     .00     .00     .00    1.00     .00     .00    1.00     .00     .00    1.00    1.00     .00
    7417     .00     .00    1.00    1.00    1.00     .00     .00     .00     .00     .00     .00    1.00    1.00     .00    1.00     .00     .00     .00
    7221    1.00     .00     .00    1.00    1.00     .00    1.00     .00     .00     .00     .00     .00     .00     .00     .00    1.00    1.00     .00
```

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

Slika 11.17

Iako je moguće analizirati ovako sačuvane podatke u okviru Winstepsa, to se može pokazati komplikovanim, jer je osim definisanja širine varijabli potrebno navesti kodove, odnosno validne odgovore u odgovarajućem formatu (8 kolona i 2 decimale), a često izmeniti i dodatne komande. Zbog toga, praktičnije rešenje je promeniti način na koji su varijable formatirane u SPSS fajlu sa podacima, a tek nakon toga ih sačuvati kao tekstualni fajl.

Ako se vratimo na SPSS fajl sa podacima, u *Data View* (Slika 11.18) prepoznajemo da su podaci sačuvani u F8.2 formatu (tip varijable je *Numeric*, svaka varijabla je široka 8 kolona i ima 2 decimale). Format lako možemo promeniti u željeni tako što ćemo prvo broj decimala smanjiti na 0, a zatim i ukupan broj kolona na 1 (suprotan redosled ne bi bio moguć jer broj decimala mora biti manji od broja kolona). Kada promenimo vrednost za prvu varijablu, možemo je kopirati i preneti na druge varijable koristeći opcije *Copy* i *Paste*.

	Name	Type	Width	Decimals
35	t1	Numeric	8	2
36	t2	Numeric	8	2
37	t3	Numeric	8	2
38	t4	Numeric	8	2
39	t5	Numeric	8	2
40	t6	Numeric	8	2
41	t7	Numeric	8	2
42	t8	Numeric	8	2
43	t9	Numeric	8	2
44	t10	Numeric	8	2
45	t11	Numeric	8	2
46	t12	Numeric	8	2
47	t13	Numeric	8	2
48	t14	Numeric	8	2
49	t15	Numeric	8	2
50	t16	Numeric	8	2
51	t17	Numeric	8	2
52	t18	Numeric	8	2
53	t19	Numeric	8	2
54	t20	Numeric	8	2

Slika 11.18

Nakon što smo promenili formatiranje varijabli, možemo ih ponovo sačuvati kao tekstualni fajl, prateći iste korake kao i ranije. Podaci koje sada dobijamo u tekstualnom fajlu (Slika 11.19) izgledaju mnogo preglednije i lakši su za korišćenje u Winsteps programu, budući da svaki ajtem sada zauzima samo jednu kolonu. Ovako sačuvane podatke možemo staviti na sam kraj kontrolnog fajla (nakon komande END LABELS) i nakon ispravnog podešavanja svih potrebnih komandi pokrenuti analizu.

TZ_PS - Notepad

File Edit Format View Help

```
72820001001000000001000000
75440001000000001011000000
75811001010000000000000100
77070001010010000000000001
71630001001000000001010010
71441000101000001000000001
76931001000000000100000110
73310001000000001010100001
77410001100000101100000000
73360001100001100010000000
76300001000010001000100001
76170001001010001000000001
72091101000000001100100000
73771001100001000100000001
73160001000011000100000011
77540001100010000101000001
76421001000000001110000001
78021001001110010000000000
78541001001001010000000001
78590001011000100100000001
78650101000001011000000001
78140101000010010101000000
75300001000010110100001000
71521001001011000000000001
74150011010000000100110100
73201001001010000101000001
73051001000010011100000001
74611001100010010101000000
73261000100001001001100100
74170011100000001101000001
```

Ln 1, Col 1 100% Windows (CRLF) UTF-8

Slika 11.19

Novi ispis iz SPSS-a sada kao format za sve ajteme na testu prikazuje F1.0, odnosno radi se o numeričkim varijablama koje zauzimaju po jednu kolonu i nemaju decimalna mesta (Slika 11.20). Da svaka varijabla zauzima samo po jednu kolonu, možemo videti i po tome što su prva (*Start*) i poslednja (*End*) kolona za svaku varijablu jednake. Drugim rečima, varijabla se „završava" u istoj koloni u kojoj i „počinje". Kako format varijable koja se odnosi na ID ispitanika nismo menjali, on je i dalje F8.0, odnosno u pitanju je numerička varijabla koja ima 8 kolona i nijedno decimalno mesto. U tekstualnom fajlu vidimo da su neke od ovih kolona prazne, budući da su redni brojevi zapravo četvorocifreni, ali to nije od značaja za analizu.

```
WRITE OUTFILE='C:\Users\Danka\Downloads\TZ_PS.dat'
  TABLE
  /idKandidat t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 .

Write will generate the following

Variable         Rec   Start     End  Format
idKandidat         1       1       8  F8.0
t1                 1       9       9  F1.0
t2                 1      10      10  F1.0
t3                 1      11      11  F1.0
t4                 1      12      12  F1.0
t5                 1      13      13  F1.0
t6                 1      14      14  F1.0
t7                 1      15      15  F1.0
t8                 1      16      16  F1.0
t9                 1      17      17  F1.0
t10                1      18      18  F1.0
t11                1      19      19  F1.0
t12                1      20      20  F1.0
t13                1      21      21  F1.0
t14                1      22      22  F1.0
t15                1      23      23  F1.0
t16                1      24      24  F1.0
t17                1      25      25  F1.0
t18                1      26      26  F1.0
t19                1      27      27  F1.0
t20                1      28      28  F1.0
EXECUTE.
```

Slika 11.20

Informacije koje dobijamo u ispisu veoma su važne za pravilno podešavanje kontrolnog fajla, konkretno za određivanje vrednosti NAME1, NAMLEN, ITEM1 i XWIDE. NAME1 je kolona u kojoj počinju informacije o ispitaniku i iz ispisa možemo videti da je početak varijable o ispitanicima (*Start*) u koloni 1. Takođe, možemo videti da se ove informacije završavaju u koloni 8 (*End*), što znači da je dužina informacija o ispitaniku, odnosno NAMLEN, jednaka 8 kolona. U slučaju da smo imali nekoliko varijabli koje se odnose na informacije o ispitaniku gledali bismo vrednost *Start* prve takve varijable za određivanje vrednosti NAME1, a vrednost *End* poslednje varijable koja se odnosi na informacije o ispitaniku kako bismo odredili vrednost NAMLEN.

Za određivanje vrednosti koju treba uneti u polje ITEM1 gledamo *Start* prvog ajtema u podacima. U našem primeru, to je ajtem t1 i on počinje (*Start*) u koloni 9. To znači da ajtemi na testu počinju u 9. koloni. XWIDE, odnosno broj kolona koje zauzima svaki ajtem možemo videti gledajući *Start* i *End* za ajteme, kao i na osnovu njihovog formata. Kako bismo odredili broj ajtema, odnosno NI, ne moramo da gledamo ispis iz SPSS-a, jer već znamo koliko varijabli ima naš instrument (ili pak koliko smo ih sačuvali u tekstualnom fajlu kako bismo ih analizirali u Winsteps-u, ukoliko analizu radimo samo na podskupu varijabli).

LITERATURA

Abdi, H., & Williams, L. J. (2010). Principal component analysis. wiley interdisciplinary reviews: computational statistics. *Wiley Interdisplinary Reviews: Computational Statistics*, *2*(4), 433–459.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

An, X., & Yung, Y. (2014). Item Response Theory : What It Is and How You Can Use the IRT Procedure to Apply It. *SAS Institute Inc.*, 1–14. Retrieved from https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf

Baker, F. B. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation.

Bazana, P. G., & Stelmack, R. M. (2004). Stability of Personality Across the Life Span: A Meta-Analysis. In *On the Psychobiology of Personality: Essays in Honor of Marvin Zuckerman* (pp. 113–144). https://doi.org/10.1016/B978-008044209-9/50009-9

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology*, *9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Brown, W. (1910). Some Experimental Results in the Correlation of Mental Abilities. *British Journal of Psychology, 1904-1920*, *3*(3), 296–322. https://doi.org/10.1111/j.2044-8295.1910.tb00207.x

Bukvić, A. (1996). *Načela izrade psiholoških testova*. Zavod za udžbenike i nastavna sredstva, Beograd.

Bulmer, M. G. (1979). *Principles of statistics*. Dover Publications, New York.

Callender, J. C., & Osburn, H. G. (1979). An Empirical Comparison of Coefficient Alpha, Guttman's Lambda-2, and MSPLIT Maximized Split-Half Reliability Estimates. *Journal of Educational Measurement*, *16*(2).

Chadha, N. K. (2009). Applied Psychometry. In *Applied Psychometry*. New Delhi, India: SAGE Publications India Pvt Ltd.

Charter, R. A., & Feldt, L. S. (2000). The relationship between two methods of evaluating an examinee's difference scores. *Journal of Psychoeducational Assessment*, *18*, 125–142.

Cicchetti, D. V. (1994). Interreliability standards in psychological evaluations. *Psychological Assessment*, *6*(4), 284–290.

Clark, L. A., & Watson, D. (1995). *Constructing Validity : Basic Issues in Objective Scale Development The Centrality of Psychological Measurement*. 7(3), 309–319.

Cochran, W. G. (1952). The χ^2 Test of Goodness of Fit. *The Annals of Mathematical Statistics*, *23*(3), 315–345. https://doi.org/10.1214/aoms/1177729380

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. Retrieved from http://epm.sagepub.com

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological Testing and Assessment: An Introduction to Tests and Measurement. Ninth Edition*. New York, NY: McGraw-Hill Education.

Cudeck, R., & MacCallum, R. C. (2007). Factor analysis at 100: Historical developments and future directions. In *Factor Analysis at 100: Historical Developments and Future Directions*. https://doi.org/10.4324/9780203936764

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York, USA: The Guilford Press, New York.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Science & Business Media,

London, UK.
Denham, B. E. (2017). *Categorical Statistics for Communication Research*. Oxford, UK: Wiley-Blackwell.
DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), S50–S59.
Diamond, J., & Evans, W. (1971). The Correction for Guessing. *Review of Educational Research*, *43*(2), 181–191.
Drasgow, F., & Hulin, C. L. (1990). Item Response Theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 577–636). Consulting Psychologists Press.
Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046
Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/h0036316
Ebel, R. L. (1967). The Relation of Item Discrimination to Test Reliability. *Journal of Educational Measurement*, *4*(3), 125–128.
Fajgelj, S. (2009). *Psihometrija. Metod i teorija psihološkog merenja, III dopunjeno izdanje*. Centar za primenjenu psihologiju, Beograd.
Fajgelj, S., & Kosanović, B. (2001). Nova i stara ajtem analiza – poređenje. *Psihologija*, *1–2*, 83–110.
Falk, C. F., & Savalei, V. (2011). The relationship between unstandardized and standardized alpha, true reliability, and the underlying measurement model. *Journal of Personality Assessment*, *93*(5), 445–453. https://doi.org/10.1080/00223891.2011.594129
Ferguson, G. A. (1949). On the Theory of Test Discrimination. *Psychometrica*, *14*(1), 61–68.
Field, A. (2009). *Discovering Statiscs Using SPSS, Third Edition*. Sage Publications, London, UK.
Fleiss, J. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, *76*(5), 378–382.
Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619. https://doi.org/10.1177/001316447303300309
Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). *An Item Response Theory Analysis of Self-Report Measures of Adult Attachment*. *78*(2), 350–365. https://doi.org/10.1037//0022-3514.78.2.350
Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, *32*(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522
Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal Stability of Implicit and Explicit Measures: A Longitudinal Analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312. https://doi.org/10.1177/0146167216684131
George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update*. Allyn and Bacon, Michigan.
Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings*. Center fir Educator Compensation Reform.
Groeneveld, R. A. ., & Meeden, G. (1984). Measuring Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *33*(4), 391–399.
Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/BF02288892
Guttman, L. (1953). Image Theory for the Structure of Quantitative Variates. *Psychometrika*, *18*(4), 277–296.
Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, *12*(3), 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x
Hankins, M. (2007). *Questionnaire discrimination : (re)-introducing coefficient δ*. *5*, 3–7. https://doi.org/10.1186/1471-2288-7-19
Harvey, R. J., & Hammer, A. L. (1999). Item Response Theory. *The Counseling Psychologist*, *27*(3), 353–383. https://doi.org/https://doi.org/10.1177/0011000099273004

Hemphill, J. F. (2003). Interpreting the Magnitudes of Correlation Coefficients. *American Psychologist*, *58*(1), 78–79. https://doi.org/10.1037/0003-066X.58.1.78
Hunt, E. (2010). *Human intelligence*. Cambridge University Press, New York, NY.
IBM Corporation. (2021). *IBM SPSS Statistics 21 Command Syntax Reference*.
Joanes, D. N., & Gill, C. A. (1998). Comparing Measures of Sample Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *47*(1), 183–189.
Jokić, B., & Purić, D. (2019). Relating rational and experiential thinking styles with trait emotional intelligence in broader personality space. *Europe's Journal of Psychology*, *15*(1), 140–158. https://doi.org/10.5964/ejop.v15i1.1692
Kaiser, H. F. (1970). A Second Generation Little Jiffy. *Psychometrika*, *35*(4), 401–415. Retrieved from http://www.springerlink.com/index/417580617713668.pdf
Kaiser, H. F. (1974). Analysis of factorial simplicity. *Psychometrika*, *39*(1), 31–36.
Kaiser, H. F., & Rice, J. (1974). *Little jiffy, Mark IV*. *34*, 111–117.
Kendall, M. G., & Babington Smith, B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*, *10*(3), 275–289. https://doi.org/10.12693/APhysPolA.130.87
King, J. E. (2004). Software Solutions for Obtaining a Kappa-Type Statistic for Use with Multiple Raters. *Annual Meeting of the Southwest Educational Research Association, Dallas, Texas*, 1–11.
Kline, P. (2000). *The Handbook of Psychological Testing, Second Edition*. London and New York: Routledge.
Knežević, G., Džamonja-Ignjatović, T., & Đurić-Jočić, D. (2004). *Petofaktorski model ličnosti*. Centar za primenjenu psihologiju, Beograd.
Knežević, G., & Momirović, K. (1996). RTT9G i RTT10G: Programi za analizu metrijskih karakteristika. In P. Kostić (Ed.), *Merenje u psihologiji* (pp. 37–56). Institut za kriminološka i sociološka istraživanja, Beograd.
Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
Lebergott, S. (1959). The Shape of the Income. *The American Economic Review*, *49*(3), 328–347.
Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation – A discussion and demonstration of basic features. In *PLoS ONE* (Vol. 14). https://doi.org/10.1371/journal.pone.0219854
Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of the American Statistical Association*, *64*(325), 387–389.
Linacre, J. M. (2022). *Winsteps® Rasch measurement computer program User's Guide. Version 5.2.1.* Portland, Oregon: Winsteps.com.
Lord, F. M. (1952). *A Theory of Test Scores (Psychometric Monograph No. 7)*. Retrieved from https://www.psychometricsociety.org/sites/default/files/pdf/MN07.pdf%0Ahttp://psycnet.apa.org/psycinfo/1954-01886-001
Lord, F. M. (1980). *Applications of Item Response Theory To Practical Testing Problems*. New York, USA and London, UK: Routlege.
Lord, F. M., & Novick, M. R. (2008). *Statistical Theories of Mental Test Scores*. Information Age Publiching, USA.
Massey, F. J. . (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78. Retrieved from https://www.jstor.org/stable/2280095
McConville, C., & Cooper, C. (1997). The temporal stability of mood variability. *Personality and Individual Differences*, *23*(1), 161–164. https://doi.org/10.1016/S0191-8869(97)00013-5
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlations coefficients. *Psychological Methods*, *1*(4), 390–390. https://doi.org/10.1037//1082-989x.1.4.390
Miller, L. H. (1956). Table of Percentage Points of Kolmogorov Statistics. *Journal of the American Statistical Association*, *51*(273), 111–121. https://doi.org/10.1080/01621459.1956.10501314
Momirović, K., & Hošek, A. (1995). Predlog jedne nove mere reprezentativnosti nekog uzorka

varijabli. *Psihologija*, *28*(1–2), 71–88.
Momirović, K., Wolf, B., & Popović, D. A. (1999). *Uvod u teoriju merenja i interne metrijske karakteristike kompozitnih mernih instrumenata*. Univerzitet u Prištini, Fakultet za fizičku kulturu.
Morera, O. F., & Stokes, S. M. (2016). Coefficient α as a measure of test score reliability: Review of 3 popular misconceptions. *American Journal of Public Health*, *106*(3), 458–461. https://doi.org/10.2105/AJPH.2015.302993
Moscoso Del Prado Martin, F. (2009). A Theory of Reaction Time Distributions. *Behavior and Brain Sciences*. Retrieved from http://cogprints.org/6310/
Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2
Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*(1), 1–13. https://doi.org/10.1007/BF02289400
OECD. (2019). How PISA results are reported: What is a PISA score? In *PISA 2018 Results (Volume I): What Students Know and Can Do* (pp. 41–47). OECD Publishing.
Pallant, J. (2005). *SPSS SURVIVAL MANUAL: A step by step guide to data analysis using SPSS for Windows (Version 12)*. Allen & Unwin, Australia.
Park, S. Y., & Bera, A. K. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*, *150*(2), 219–230. https://doi.org/10.1016/j.jeconom.2008.12.014
Pek, J., Wong, O., & Wong, A. C. M. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in Psychology*, *9*(NOV), 1–17. https://doi.org/10.3389/fpsyg.2018.02104
Peterson, R. A. (1994). A Meta-analysis of Cronbach' s Coefficient Alpha. *Journal of Consumer Research*, *21*(2), 381–391.
Rae, G. (1988). The equivalence o f multiple rater kappa statistics and intraclass correlation coefficients. *Educational and Psychological Measurement*, *48*, 367–374.
Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in Clinical Research*, *8*(4), 187–191. https://doi.org/10.4103/picr.PICR_123_17
Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, *3*, 1–13. https://doi.org/10.3389/fpsyg.2012.00034
Shrout, P. E., & Fleiss, J. L. (1979). *Intraclass Correlations : Uses in Assessing Rater Reliability*. *86*(2), 420–428.
Shweta, Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, *41*(Special Issue 3), 20–27.
Solomon, S. R., & Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *Journal of Modern Applied Statistical Methods*, *8*(2), 448–462. https://doi.org/10.22237/jmasm/1257034080
Spearman, C. (1910). Correlation Calculated From Faulty Data. *British Journal of Psychology, 1904-1920*, *3*(3), 271–295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x
Tenjović, L. (2000). *Statistika u psihologiji: priručnik*. Beograd: Centar za primenjenu psihologiju Društva psihologa Srbije.
Terluin, B., Knol, D. L., Terwee, C. B., & de Vet, H. C. W. (2009). Understanding Ferguson's δ: Time to say good-bye? *Health and Quality of Life Outcomes*, *7*(1), 1–7. https://doi.org/10.1186/1477-7525-7-38
Thorndike, E. L. (1973). *The Measurement of Intelligence*. Retrieved from https://babel.hathitrust.org/cgi/pt?id=mdp.39015000265028&view=1up&seq=9
Williams, R. H., & Zimmerman, D. W. (1966). An extension of the Rulon formula for test reliability: The case of correlated true and error components of scores. *ETS Research Bulletin Series*, *2*(i–8).
Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. In *Rasch Measurement Transactions* (pp. 370–371).

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, *77*(2), 175–183. https://doi.org/10.1080/10629360600678310

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7

POJMOVNIK

antiimaž – u Gutmanovom modelu merenja, deo skora ili varijanse skorova nedeljen sa drugim varijablama u modelu, odnosno koji se na osnovu njih ne može predvideti; izjednačen sa skorom greške

autfit – u teoriji odgovora na stavke, pokazatelj odstupanja podataka od modela na koji pretežno utiču greške daleko od procenjenog nivoa osobine ispitanika, odnosno stavke

bimodalna distribucija – distribucija skorova na testu koja ima dva moda, odnosno dve najčešće vrednosti

diskriminativnost – svojstvo testa da razlikuje ispitanike spram stepena merene osobine, odnosno da registruje individualne razlike između ispitanika u merenoj osobini

hi-kvadrat test – jedan od najčešće korišćenih neparametrijskih testova koji testira nultu hipotezu da je empirijska distribucija frekvenci jednaka teorijskoj

histogram – grafički prikaz frekvence odgovora na varijabli

homogenost – svojstvo testa da se njegove stavke odnose na jedan isti predmet merenja

imaž – u Gutmanovom modelu merenja, deo skora ili varijanse skorova koji je deljen sa drugim varijablama u modelu, odnosno koji se na osnovu njih može predvideti; izjednačen sa pravim skorom

infit – u teoriji odgovora na stavke, pokazatelj odstupanja podataka od modela na koji pretežno utiču greške blizu procenjenog nivoa osobine ispitanika, odnosno stavke

informativnost – u teoriji odgovora na stavke, pokazatelj stepena u kom test razlikuje ispitanike različitog nivoa merene osobine; u funkciji nivoa merene osobine

intraklasni koeficijent korelacije – statistički pokazatelj stepena u kom postoji slaganje procenjivača, odnosno u kome se ocene različitih procenjivača poklapaju

karakteristična kriva stavke – u teoriji odgovora na stavke, kriva koja grafički prikazuje odnos nivoa izraženosti latentne osobine i verovatnoće (tačnog) odgovora na datu stavku

klasična teorija testa – model merenja koji pretpostavlja da se opaženi, dobijeni skor ispitanika na nekom testu može razložiti na pravi skor i skor greške

kriva informativnosti stavke/testa – u teoriji odgovora na stavke, kriva koja grafički prikazuje odnos nivoa izraženosti latentne osobine i informativnosti stavke, odnosno testa

Kronbahova alfa – jedan od najčešće korišćenih koeficijenata pouzdanosti tipa interne konzistencije; zasnovan na analizi stavki

kurtozis – numerički pokazatelj izduženosti/spljoštenosti distribucije skorova na testu

leptokurtična distribucija – distribucija skorova koja pokazuje grupisanje skorova oko srednjih vrednosti

linearni kompozit – varijabla koja se dobija primenom linearnih (aditivnih) transformacija na drugim varijablama

mapa ispitanika i stavki – u teoriji odgovora na stavke, grafički prikaz u kome su ispitanici i stavke, na osnovu svojih procenjenih vrednosti osobine, pozicionirani na latentnom kontinuumu osobine

misfit – u teoriji odgovora na stavke, pokazatelj odstupanja podataka od modela

negativno asimetrična distribucija – distribucija skorova koja pokazuje grupisanje skorova oko viših vrednosti

normalizacija skorova – statistički postupak kojim se empirijska distribucija skorova na testu približava teorijskom modelu normalne distribucije

normalna distribucija – teorijski model distribucije skorova na testu koji odgovara većini psiholoških konstrukata

normiranje – postupak prevođenja sirovih skorova na testu na neku standardnu skalu, primenom linearnih ili nelinearnih transformacija

objektivnost – svojstvo testa da skorovi (pretežno) zavise od stepena izraženosti osobine ispitanika, a ne od karakteristika procenjivača

platikurtična distribucija – distribucija skorova koja pokazuje raspršenje skorova preko svih vrednosti

pouzdanost – svojstvo testa da na konzistentan način procenjuje merenu osobinu

pozitivno asimetrična distribucija – distribucija skorova koja pokazuje grupisanje skorova oko nižih vrednosti

pravi skor – u klasičnoj teoriji testa, deo dobijenog skora koji odražava nivo osobine ispitanika

Rašovi modeli – u teoriji odgovora na stavke, modeli koji pretpostavljaju da se stavke razlikuju isključivo spram parametra težine i kod kojih je parametar diskriminativnosti fiksiran na jedan

reprezentativnost – svojstvo testa da uzorak stavki od kog je sačinjen na adekvatan način predstavlja univerzum stavki datog konstrukta

separacija – u teoriji odgovora na stavke, pokazatelj broja statistički različitih nivoa osobine koji se testom može registrovati; odnos pouzdane varijanse i varijanse greške

sistematska greška merenja – u klasičnoj teoriji testa, deo skora greške koji se može pripisati sistematskim faktorima

skjunis– numerički pokazatelj simetrije, odnosno asimetrije distribucije skorova na testu

skor greške – u klasičnoj teoriji testa, deo dobijenog skora koji se ne odnosi na nivo osobine ispitanika, već je posledica greške merenja

slučajna greška merenja – u klasičnoj teoriji testa, deo skora greške koji se ne može pripisati nikakvim sistematskim ili predvidljivim faktorima

standardna greška merenja – standardna devijacija distribucije dobijenih skorova oko pravog skora; inverzno proporcionalna pouzdanosti

standardna skala – skala koja ima poznatu aritmetičku sredinu i standardnu devijaciju

teorija odgovora na stavke – model merenja koji pretpostavlja da je verovatnoća (tačnog) odgovora na stavku u funkciji stepena izraženosti latentne osobine i da se ovaj odnos matematički može izraziti preko funkcije koja može uključivati različit broj parametara

uniformna distribucija – distribucija kod koje su svi odgovori jednako frekventni ili verovatni

INDEKS POJMOVA

INDEKS AUTORA

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

159.9.072(075.8)(0.034.2)

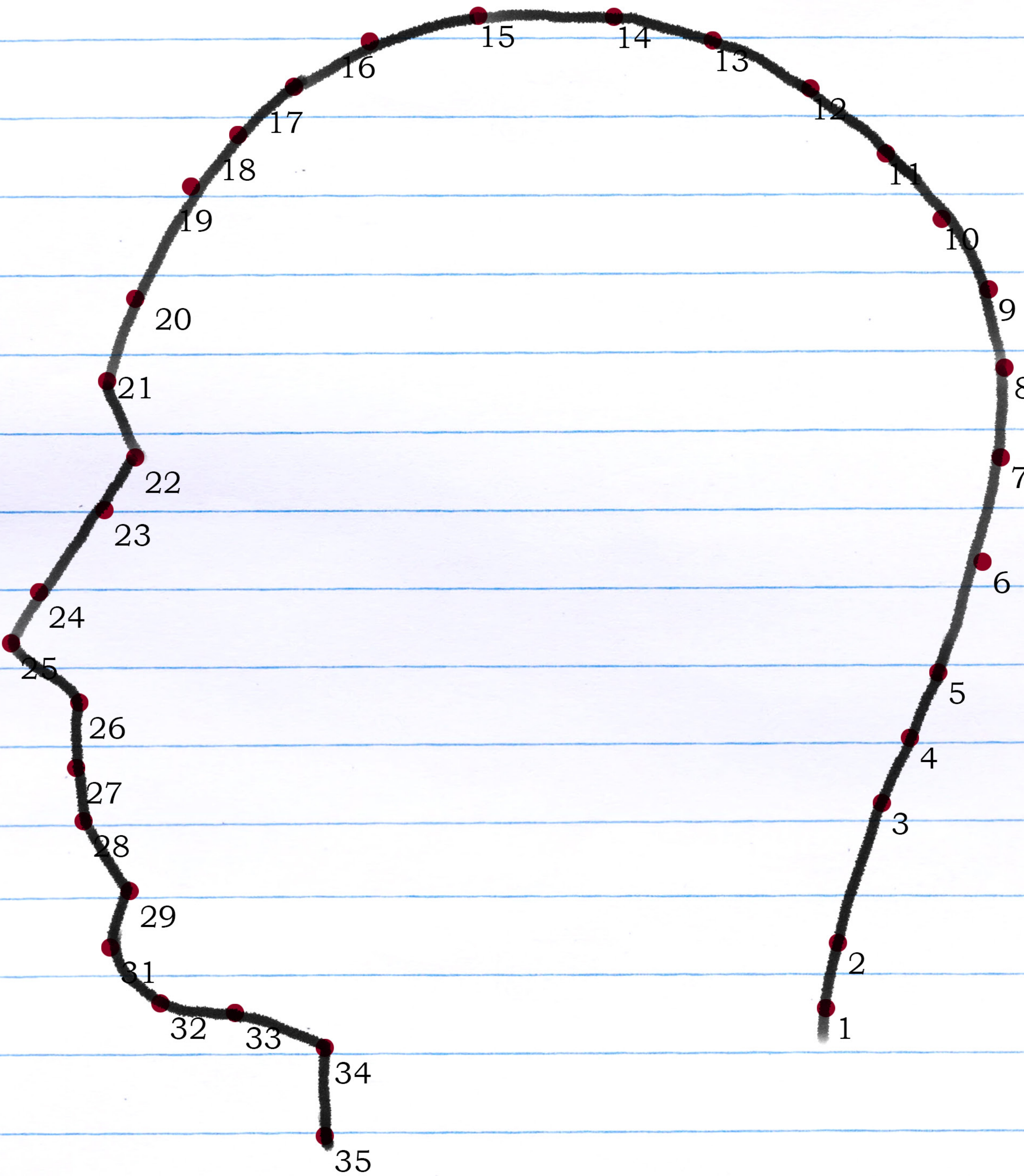
ПУРИЋ, Данка, 1985-
Interne metrijske karakteristike psiholoških instrumenata [Elektronski izvor]. Vodič kroz Psihometriju 1 / Danka Purić. - Beograd : Filozofski fakultet, Institut za psihologiju, 2023 (Beograd : Filozofski fakultet, Institut za psihologiju). - 1 elektronski optički disk (CD- ROM) ; 12 cm

Sistemski zahtevi: Nisu navedeni. - Nasl. sa naslovne strane dokumenta. - Tiraž 10. - Pojmovnik. - Preporučena literatura uz svako poglavlje. - Bibliografija. - Registri.

ISBN 978-86-6427-202-5

а) Психометрија

COBISS.SR-ID 116757001


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
31
32
33
34
35