OSNOVE PSIHOMETRIJE

ZA STUDENTE PSIHOLOGIJE

Dženana Husremović

Filozofski fakultet Univerziteta u Sarajevu

Osnove psihometrije

Za studente psihologije

Dženana Husremović

Sarajevo, 2016.

Dženana Husremović
OSNOVE PSIHOMETRIJE ZA STUDENTE PSIHOLOGIJE

Urednik:
Dr. Salih Fočo

Recenzenti:
Dr. Nermin Đapo
Dr. Đorđe Čekrlija

Lektor:
Meliha Kešmer

Izdanje:
Prvo

Izdavač:
Filozofski fakultet Univerziteta u Sarajevu

Sarajevo, 2016.

Elektronsko izdanje
http://www.ff-eizdavastvo.ba/Knjige.aspx

CIP - Katalogizacija u publikaciji
Nacionalna i univerzitetska biblioteka
Bosne i Hercegovine, Sarajevo
159.938.3(075.8)
HUSREMOVIĆ, Dženana
Osnove psihometrije [Elektronski izvori] : za studente psihologije / Dženana Husremović. –
Elektronski tekstualni podaci. - Sarajevo : Filozofski fakultet, 2016
Nisu navedeni sistemski zahtjevi .
Način dostupa (URL): http://www.ff-eizdavastvo.ba/Knjige.aspx. - Nasl. s naslovnog ekrana.
ISBN 978-9958-625-58-9
COBISS.BH-ID 22918662

TRIPUT MJERI, JEDNOM SIJECI.

Narodna poslovica

SAŽETAK

„Osnove psihometrije za studente psihologije“ je priručnik namijenjen lakšem razumijevanju i praćenju nastave iz predmeta Psihometrija. U priručniku su prikazani osnovni pojmovi, koncepti i odnosi u psihološkim mjerenjima. Prvi dio knjige odnosi se na opis osnovnih postulata mjerenja u društvenim znanostima. Definira se mjerenje, pravila mjerenja, te problemi s primjenom pravila u ispitivanju ljudskih karakteristika. Potom se definira psihološki test i njegove karakteristike koje ga razlikuju kao dijagnostičko sredstvo od drugih načina prikupljanja podataka. U poglavlju koje se odnosi na formiranje uratka na testu detaljno se obrazlažu načini kako različite kombinacije u formiranju uratka utječu na metrijske karakteristike testa. Najvažnija poglavlja u priručniku odnose se na ispitivanje metrijskih karakteristika: pouzdanosti, osjetljivosti, valjanosti i pristranosti. U poglavlju o pouzdanosti diskutira se o tome što je pouzdanost, kako se empirijski provjerava pouzdanost testa, determinante pouzdanosti, glavne kritike određivanja pouzdanosti prema Klasičnoj testnoj teoriji, te se prikazuju osnovni koncepti Teorije generalizabilnosti. U poglavlju o osjetljivosti testa, daje se pregled glavnih determinanti osjetljivosti zadataka i testa te se pravi poveznica između osjetljivosti, pouzdanosti i valjanosti. Poglavlje o valjanosti daje pregled suvremenih pogleda na valjanost i njene osnovne tipove – konstruktnu valjanost kao centralni pojam, te dokaze o valjanosti kroz sadržaj testa, unutarnju strukturu, povezanost sa drugim mjerama, psihološke procese i posljedice testiranja. Poseban značaj u priručniku dat je pristranosti u testiranju koja u današnjem vremenu ima veliki značaj. Poglavlje o pristranosti pruža informacije o važnosti pažljivog prijevoda instrumenata, metodama utvrđivanja pristranosti i načinima minimizacije i prevencije pristranosti kako se testovima ne bi poticala nejednakost u tretmanu grupa ili pojedinaca.

PREDGOVOR

Dragi studenti,

Na samom početku, da se razumijemo – vaš život i način kako danas doživljavate sebe sigurno je dijelom oblikovan psihološkim testiranjima. Vi ste kao učenici bili ispitanici u psihološkim testiranjima, ako ništa onda u vidu testova znanja. Vjerovatno niste razmišljali o njima na znanstveni način (a i ko bi kada je u tim trenucima samo bitno dobiti dobru ocjenu), ali ti testovi su svojim kvalitetom utjecali na to da vi o sebi razmišljate kao o nekome kome npr. dobro ide matematika, ali nije baš dobar u nekom drugom predmetu. Psihološka testiranja mogu imati i mnogo ozbiljnije posljedice od toga da li će neko imati 4 ili 5 iz školskih predmeta. Na primjer, psihološkim testiranjima utvrđuje se stepen uračunljivosti neke osobe koja tvrdi da je u trenutku kada je ubila nekoga bila neuračunljiva. Ukoliko se utvrdi postojanje neuračunljivosti, osoba će služiti blažu kaznu (ili je možda neće ni služiti). Pitanje koje sebi postavljamo je: kako možemo biti sigurni da rezultat testa zaista odražava neuračunljivost? U Sjedinjenim Američkim Državama psihološki testovi mogu biti pitanje života i smrti. U državama u kojima je dozvoljena smrtna kazna, osobi koja ima ograničene mentalne sposobnosti smrtna kazna ne može biti izrečena. Kako možemo biti sigurni da postignuti rezultat na testovima mentalnih sposobnosti doista odražava stvarno stanje neke karakteristike i, posljedično, kako možemo biti sigurni da smrtna kazna nije izrečena osobi s ograničenim sposobnostima? Činjenica da čitate ovu skriptu govori o tome da ćete u budućnosti, u svojoj karijeri biti ne samo korisnik testova nego i njihov kreator, te da ćete biti u situaciji u kojoj ćete morati donositi odluke na osnovu rezultata na psihološkim testovima. Stoga je poznavanje psihometrije jedna od ključnih kompetencija svakog budućeg psihologa. Ova skripta namijenjena je baš vama kako bi vam olakšala razumijevanje psihometrije kao psihološke discipline i kao osnove znanstvene psihologije, te kako bi vas pripremila za vašu buduću karijeru socijalnog znanstvenika.

Moji studenti naučili su me mnogo toga što me vodilo u pisanju ovog priručnika.

1. Naučila sam da nemali broj studenata upisuje psihologiju, misleći kako na tom studiju neće morati učiti ništa vezano za matematiku. Vjeruju kako će konačno zaboraviti i osnovne računske operacije, a onda se jako iznenade kada shvate da su prve godine studija psihologije posvećene uglavnom Psihološkoj metodologiji i predmetima kao što su Metodologija istraživanja, Statistika i Psihometrija. Ovaj raskorak u očekivanjima i stvarnim zahtjevima utječe na njihovu percepciju samoefikasnosti. Vide sebe kao nedovoljno efikasne u razumijevanju psihometrije i razvijaju ciljnu orijentaciju u učenju koja se najbolje može objasniti rečenicom – *Samo da se ovo položi*. Zato sam željela da ova knjiga bude lagana za korištenje, da objašnjenja budu pristupačna i razumljiva studentima koji nisu vjerovali da će se na studiju baviti matematikom i da im pošalje poruku kako svako ko želi može savladati psihometriju i uživati u njoj.
2. Također sam naučila da studenti ne vole „debele“ knjige. U pripremama za ispit, oni se teško snalaze u debelim knjigama pri traženju najvažnijih informacija. Za njih je bolja strategija pisanja priručnika fokusiranih na određeni sadržaj, praćenih sažecima za predavanja sa

sadržajima koje je nastavnik u međuvremenu pripremio, te upućivanje studenata na specifičnu literaturu koja im može biti od pomoći. U duhu „agilnog razvoja“, cilj ovog priručnika je da studentima definira glavna područja i pruži osnovna znanja u psihometriji, te da služi kao vodič u gradnji stručnih kompetencija. Ovaj priručnik nikako ne pretendira da pokrije sve što studenti trebaju naučiti iz područja mjerenja i konstrukcije testova.

3. Još jednu stvar koju sam naučila tokom suradnje sa studentima je da su oni, ukoliko im literatura za gradivo ne nudi objašnjenja u pripovjedačkom narativu, skloniji koristiti prezentacije i samostalno pripremljene skripte za pripremu ispita. To obično završi tako da njihovo znanje ostaje na nivou prepoznavanja i opisivanja, dok stvarno razumijevanje i primjena na drugim kolegijima izostaju. Prema konstruktivističkim teorijama učenja i podučavanja (Biggs, 1996), ciljevi podučavanja, metode i ispitivanje trebaju biti konstruktivno usklađeni kako bi polaznici zaista stekli kompetencije potrebne za budući rad. Namjera ovog priručnika jeste da bude dodatno sredstvo za podučavanje i da omogući studentima da na predavanjima aktivno slušaju i razumijevaju, a ne da provedu vrijeme prepisujući s table.
4. Jednom prilikom, razgovarala sam s prijateljicom koja je pokrenula svoju organizaciju za veselo podučavanje fizike. U razgovoru ona mi je rekla: „Znaš, mislim da smo mnogo toga upropastili uvođenjem dugmića u prirodoslovne muzeje. Naša djeca sada vjeruju da se fizikalna pojava desi zato što su oni pritisnuli neko dugme. Oni potpuno gube pojam o tome da se između pritiska na dugme i onoga što vide nalazi niz reakcija koje treba razumjeti da bi znali zašto se nešto dešava.“ Vrlo brzo nakon toga, jedan student mi je rekao: „Ali, meni nije jasno zašto mi sve ovo učimo kada danas imamo računare i dovoljno je samo da ubacimo podatke i pritisnemo dugme i sve se samo završi. Zašto nas tjerate da ovo znamo?“ I zaista, nova tehnologija, koliko god nam je olakšala život i omogućila razvoj psihologije kao znanosti, toliko nam je postavila izazove u smislu dokazivanja da računari ne mogu ništa uraditi sami ako mi ne znamo što radimo. Nije dovoljno samo stisnuti dugme jer je dugmadi mnogo i iza svakog se nalazi neka filozofija, neki koncept. Da bismo znali što se dešava kada stisnemo dugme na računaru i da bismo znali što smo dobili kao rezultat, moramo poznavati osnove psihološke metodologije. Uvijek je dobro podsjetiti se na onu čuvenu rečenicu „Garbage in – garbage out“. Ovaj priručnik se ne bavi obukom studenata za korištenje nekog programskog paketa za obrade podataka (iako sadrži tabele iz SPSS-a) nego se bavi konceptima i procedurama važnim za razumijevanje mjerenja u psihologiji.

Psihometrija je jako zabavna, prava avantura i ujedno područje u kojem najviše shvatamo naše mogućnosti i ograničenja kao psiholozi. Stoga je učenje o mjerenju i testovima, ustvari, promišljanje o mogućnostima psihologije.

SADRŽAJ

1 ČIME SE BAVI PSIHOMETRIJA?

Što je psihometrija i zašto je važna disciplina za sve psihologe?

Kao što samo ime ove znanstvene discipline kaže, psihometrija s bavi mjerenjem u psihologiji. Brown (2000) kaže da je psihometrija "...disciplina društvenih znanosti koja se bavi kvantifikacijom i analizom razlika kod ljudi". Analizirajući ovu definiciju vidimo da je glavni zadatak psihometrije provjera kvantifikatora, odnosno provjera kvalitete indikatora psiholoških karakteristika, te analiza razlika koje se pojavljuje u psihološkim mjerenjima. Iz ove definicije također vidimo da se psihometrija ne bavi ljudima niti njihovim osobinama, nego načinom kako su te osobine procijenjene i da li te procjene mogu biti uzete kao dovoljno kvalitetne za daljnje interpretacije i zaključivanja. Dakle, u psihometriji nas ne zanima primjena neke mjere, nego kvaliteta njene konstrukcije.

Široko gledajući, psihometrija se bavi teoretskim i tehničkim aspektima mjerenja (posebno konstrukcijom psiholoških mjernih instrumenata) te, u novije vrijeme i razvojem teoretskih pristupa mjerenju.

1.1 ZAŠTO JE VAŽNO ZNATI PSIHOMETRIJU?

Svaka disciplina, da bi bila znanost, mora imati objekt proučavanja i metode proučavanja. Metode proučavanja u znanosti temelje se na mjerenju. Znanstvena psihologija oslanja se na mjerenje kao način identifikacije prisutnosti i intenziteta neke psihološke karakteristike. Na primjer, u kliničkoj psihologiji, ako želimo ispitati koliko je klijent depresivan, to radimo tako da mu zadajemo određene upitnike, te opažamo ponašanje, bilježeći učestalost i intenzitet njihovih javljanja. Da bismo razvili adekvatne upitnike ili pripremili materijale za bilježenje ponašanja, potrebno je da znamo kako pripremiti pitanja, kako bodovati odgovore i koliko će krajnji rezultati zaista biti objektivni, precizni i mjeriti ono što smo zaista željeli mjeriti. Sve ovo uči se na Psihometriji.

1.2 KOJI SU CILJEVI PREDMETA PSIHOMETRIJA?

Prof. Stanislav Fajgelj u svojoj knjizi Psihometrija (2003) kaže da su osnovni ciljevi Psihometrije kao predmeta sljedeći:

1. Razumijevanje koncepta psihološkog mjerenja
    - Šta mjeriti (npr. šta je inteligencija, je li to obim glave ili uradak na zadacima)
    - Kako mjeriti (koje mjere koristiti)
    - Šta je fenomen mjere
2. Praćenje i razumijevanje gradiva drugih stručnih predmeta – bez razumijevanja metrijskih karakteristika teško možemo čitati stručnu literaturu i zaključivati o kvaliteti radova

3. Pravilna primjena testova i interpretacija testovnih rezultata – kada kao diplomirani psiholog budete radili u praksi ili istraživanjima, od velike važnosti će biti da znate odabrati testove, da ih znate pravilno primijeniti, da znate interpretirati dobivene rezultate, te da znate ograničenja vaših zaključaka

4. Stjecanje znanja potrebnih za konstrukciju novih testova ili za standardizaciju i adaptaciju postojećih – od psihologa se očekuje ne samo da znaju primijeniti testove, nego i da znaju konstruirati test za mjerenje neke karakteristike. Konstrukcija testa, kao i proizvodnja bilo kojeg drugog mjernog instrumenta, zahtijeva vještog konstruktora, a sutra to trebate biti upravo vi.

2 MJERENJE U PSIHOLOGIJI

Što je mjerenje u psihologiji? Po čemu se razlikuju mjerenje u fizici i mjerenje u psihologiji? Što je mjerenje preferencije čokolade, a što stava prema konzumaciji čokolade?

Znanost je definirana sa dvije stvari: ona ima objekt proučavanja i ima metode proučavanja.

Objekt proučavanja može biti fizičke prirode, ali može biti i nefizičke u užem smislu riječi.

Mjerenje je izuzetno važno u svakoj znanosti. Iako mjerenje u znanosti ima vrlo dugu tradiciju (Leahey, 2001), sama znanost o mjerenju posebno se razvila u posljednjih stotinu godina. Taj značajni razvoj znanosti o mjerenju možemo pripisati potrebi stvaranja preduvjeta za razvoj drugih znanosti koje se bave istraživanjem, primjenom znanstvenih spoznaja u svakodnevnom životu i povećanjem korpusa znanja o različitim fenomenima (posebno društvenim). Svi znanstvenici zainteresirani su za stjecanje spoznaja koje su objektivne, precizne i provjerljive. Cilj svakog znanstvenika je da prikupi podatke o nekom objektu, fenomenu ili sistemu na što objektivniji i precizniji način. Osim toga, objekti ili fenomeni vrlo rijetko egzistiraju izolirani jedni od drugih. To je razlog zašto znanstvenici teže otkriti odnose među objektima, objektivno procijeniti kvalitetu tih odnosa i donijeti zaključak o tome kako se taj odnos mijenja u međuzavisnosti svih elemenata. Na taj način znanstvenici mogu identificirati i objasniti fenomene i sisteme te predviđati promjene koje mogu nastati u budućnosti.

Kroz mjerenje i kvantifikaciju, uz pomoć matematičkih modela i statistike, znanstvenici mogu razlučiti objekte i fenomene jedne od drugih, razlikovati njihove osobine, razumijevati povezanosti i predviđati promjene u budućnosti, i to s visokim stepenom tačnosti i detaljnosti. Možemo zaključiti da je mjerenje jedan od osnovnih elemenata znanstvenog istraživanja i otkrivanja.

2.1 DEFINICIJA MJERENJA

Mjerenje je moguće definirati na različite načine, stoga i postoji više definicija mjerenja. Definicija koju je dao Stevens (1951) kaže da je mjerenje "*dodjeljivanje brojeva objektima ili događajima na osnovu nekog pravila*".

Druga definicija kaže da je mjerenje poređenje nekog predmeta mjerenja s mjernom jedinicom i ovo je dosta standardna definicija u prirodnim znanostima.

Definicija koja je nama najiskoristivija u smislu obuhvatanja svih bitnih elemenata u objašnjavanju mjerenja je: "Mjerenje je svaki proces primjene nekog skupa pravila za pridruživanje brojeva pojedinim atributima proučavanih objekata."

Ova definicija je u mnogome slična Stevensovoj definiciji. I jedna i druga kažu – nema mjerenja bez pravila.

2.1.1 Pravila mjerenja

Stevens je dao osnovna pravila mjerenja koja važe kako u prirodnim, tako i u humanističkim i društvenim znanostima. Pravila postoje da bi se opisao proces mjerenja i omogućilo njegovo provođenje. Osim toga, bez pravila nema standardizacije procesa mjerenja. Zamislite kako bi to bilo da nekome mjerimo visinu palcem, a nekome kažiprstom i onda kažemo: "Osoba A je visoka 50 palaca, a osoba B je visoka 38 kažiprsta". Pravila postoje kako bi mjerenje bilo jednako provedeno u svim situacijama i na svim ispitanicima, te kako bi dobiveni rezultati bili usporedivi.

Pravila koja ćemo ovdje navesti dao je Campbell 1938. (prema Reese, 1943), a kasnije ćemo pročitati nešto više o njegovoj teoriji mjerenja.

Pravilo 1. Identičnost brojeva – brojevi su ili isti ili različiti ($a = b$ ili $a \neq b$)

Pravilo 2. Simetričnost jednakosti (ako je $a = b$ onda je i $b = a$)

Pravilo 3. Jednakost brojeva (dvije stvari koje su jednake nekoj trećoj i same su među sobom jednake; ako je $a=b$ i $b=c$, onda je i $a=c$)

Dakle, ako je osoba A odgovorila sa 1, osoba B odgovorila sa 1 i osoba C odgovorila sa 1, mi vjerujemo da su sve tri osobe jednake prema tome koliko im se često dešava ponašanje iz date tvrdnje.

Pravilo 4. Asimetričnost relacije „veći“ (ako je $a>b$, onda $b \not> a$)

Pravilo 5. Odnos tranzitivnosti ako je $a>b$ i $b>c$, onda je i $a>c$

Pravilo 6. Mogućnost sabiranja brojeva (ako je $a=p$ i $b>0$, onda je $a+b>0$)

Pravilo 7. Redoslijed sabiranja brojeva ne utječe na rezultat ($a+b =b+a$)

Pravilo 8. Identične stvari mogu pri sabiranju zamjenjivati jedna drugu (ako je $a=p$ i $b=g$ onda je $a+b = p+g$)

Pravilo 9. Kombinacije u sabiranju brojeva ne utječu na krajnji rezultat ($(a+b) + c = a + (b + c)$)

2.1.1.1 Pravila mjerenja u psihologiji

Sigurno se pitate: "Zašto bi meni, kao studentu psihologije, bilo važno da poznajem ova pravila?" Zbog toga što je poštovanje ovih pravila najteži zadatak u konstrukciji testova. Da bismo ovo shvatili, krenimo od razlike između **direktnog i indirektnog** mjerenja.

Direktno mjerenje podrazumijeva da je ono što mjerimo vidljivo, odnosno da ga možemo direktno opažati (kao što je dužina stola). Osnovna karakteristika direktnog mjerenja je da postoji jednakost između predmeta mjerenja i jedinice mjerenja. Dužina stola mjeri se metrom, odnosno mjernim instrumentom kojim utvrđujemo dužinu. Druga bitna karakteristika direktnog mjerenja je da postoji jednakost među veličinskim razlikama. Na primjer, razlika između 1 i 2 cm je ista kao i razlika između 100 i 101 cm.

Indirektno mjerenje definira se kao standardizirani postupak pomoću kojeg se neki psihički proces izaziva, pa se efekti tog procesa (tj. ispitanikove reakcije) registriraju, a zatim se reakcije pojedinog ispitanika mjere i vrednuju uspoređujući ih s reakcijama dobivenim od drugih sličnih ispitanika u jednakoj situaciji. Za razliku od direktnog mjerenja, kod indirektnih mjerenja ne možemo direktno opažati mjerenu karakteristiku. Ova onemogućenost direktnog opažanja dovodi nas do niza problema. Prvi problem je kako mi uopće definiramo predmet mjerenja (npr. šta mislimo pod terminom "numeričke sposobnosti"). Kada se i dogovorimo oko definiranja predmeta mjerenja, imamo problem što ga ne možemo direktno opažati, nego o njemu možemo zaključivati samo na osnovu nekih vanjskih manifestacija za koje vjerujemo da su povezane s predmetom mjerenja. Na primjer, o numeričkim sposobnostima učenika zaključujemo na osnovu njegovog uratka na aritmetičkom testu. I tu dolazimo do dilema vezanih za navedena pravila mjerenja.

U prvom pravilu kaže se da su dvije vrijednosti ili iste ili različite. Zamislimo situaciju da ispitanici odgovaraju na pitanje:

Tabela 2-1 prikaz jedne tvrdnje iz Upitnika za mjerenje kvalitete timskog rada

Tvrdnja	Nikada	Vrlo rijetko	Ponekad	Često	Gotovo uvijek
	1	2	3	4	5
Članovi tima u kojem radim dobivaju povratne informacije o poslu koji obavljaju					

Ukoliko dva ispitanika (Marko i Damir) stave oznaku ispod broja 1 (Nikada) vjerovat ćemo da se ono što je opisano u tvrdnji ne dešava nikada, niti jednom niti drugom, dok ukoliko označe različite brojeve (npr. 1 i 2) vjerovat ćemo da se Marku opisano ponašanje ne dešava nikada, a Damiru vrlo rijetko. Međutim, vrlo je moguće da naš zaključak nije valjan. Neki od razloga zašto rezultati ne moraju biti valjani su:

1. Ispitanici nisu imali dovoljno podioka da se izraze. Na primjer, možda se to desilo jednom ili dvaput, ali ni Marko ni Damir nisu našli broj koji bi ukazao na tu učestalost i odlučili su se da zaokruže 1 "Nikada". Dakle, mi ćemo vjerovati da se to nikada nije desilo, ali to nije prava istina.

2. Možda se to Marku nije nikada desilo, a Damir nikada nije razmišljao o tome, pa je odlučio da zaokruži 1. Dakle, oni opet nisu odgovarali iz iste tačke, ali mi zaključujemo da su jednaki. Time smo prekršili pravilo o identičnosti brojeva jer ono što stoji iza brojeva nije identično.

3. Možda se to njima dešava, ali oni ne žele priznati i daju odgovor 1. Mi zaključujemo da su oni jednaki i da se njima ovo ponašanje nikada ne dešava, što opet nije tačno.

4. Možda im se podjednak broj puta desilo da ne dobivaju povratne informacije, ali je Marko to percipirao kao "Nikada" jer mu to nije bitno, a Damir kao "Vrlo rijetko" jer mu je to zasmetalo. Mi zaključujemo da između njih postoji razlika, ali u osnovi ne vezana za opisano ponašanje, nego za njihove percepcije. I opet nismo u pravu.

Samo na ovom jednom primjeru vidimo da je vrlo teško zaključiti da postoji odnos identičnosti brojeva, simetričnost jednakosti te jednakosti brojeva. Jer zapamtimo, indirektno mjerenje je takvo da mi različitim intenzitetima neke pojave pripisujemo brojeve, o onda vrlo brzo zaboravljamo da iza tih brojeva i ne moraju biti iste stvari. Ukoliko nisu zadovoljena prva pravila, posljedično će se pojaviti problemi i na drugim pravilima, a sve ovo značajno ugrožava mjerenje i zaključke koje donosimo o ispitivanim fenomenima.

Zadatak za vas: Zadatak dva ispitanika bio je da ocijene koliko doživljavaju sljedeće emocije na skali od 1 do 5. Koje biste probleme mogli pretpostaviti kod mjerenja emocija na ovoj skali?

Tabela 2-2: prikaz mjerenja iz Upitnika raspoloženja

	Uopće ne ili vrlo malo	**Malo**	**Umjereno**	**Dosta**	**Jako**
	1	2	3	4	5
Zainteresirano					
Budno					
Pozorno					

2.1.2 Pridruživanje brojeva

Drugi dio definicije kaže da je mjerenje "pridruživanje brojeva". Teorija mjerenja se prvenstveno zanima za konstrukciju „lenjira“ na osnovu kojeg znanstvenik može mjeriti svojstva objekata ili fenomena. To je proces koji uključuje povezivanje numeričkih simbola s ljudima, objektima, događajima ili fenomenima prema unaprijed određenim pravilima.

Ova procedura dodjele numeričkih simbola odvija se tako da svaki simbol označava jednu instancu ili jedan intenzitet karakteristike koju mjerimo. Na primjer, ocjene od 5 do 10 su dodijeljene karakteristici koju zovemo "akademsko postignuće" u nekom predmetu pri čemu je numerički simbol 5 dodijeljena stepenu akademskog postignuća ispod 55 % od ukupnog broja bodova, dok je numerički simbol 10 dodijeljena akademskom postignuću iznad 93 % od ukupnog broja bodova. U ovom slučaju mjerili smo neko kvantitativno svojstvo ili atribut, a to je količina usvojenog znanja. Možemo reći da je osoba koja je dobila 10 uradila više tokom studiranja određenog predmeta, nego osoba koja je dobila 6.

Isti princip važi za dodjeljivanje numeričkih simbola u svim psihološkim testovima. Na primjer, na pitanje "Da li uvijek perete ruke prije jela?" odgovoru „Ne“ dodjeljujemo numerički simbol 0, a odgovoru „Da“ dodjeljujemo 1.

U nekim slučajevima numerički simboli koji se dodjeljuju nivoima mjerene karakteristike nemaju nikakvo kvantitativno značenje. Na primjer na pitanje "Više volim čitati knjigu, nego izaći napolje s društvom" osoba može odgovoriti sa „Ne“ i pri tome će joj biti dodijeljena 0, ili sa „Da“ pri čemu će dobiti vrijednost 1. Ono što treba da zapazimo je da ovi brojevi ne znače da je osoba B bolja od osobe A ili obrnuto nego su one samo različite. Ovo je slučaj kod mjerenja svih kvalitativnih svojstava.

Simbole koje koristimo da označimo kvalitativne aspekte mjerene karakteristike nazivamo **numerali**. Numerali nemaju kvantitativno značenje. Simbole koje dodjeljujemo kvantitativnim svojstvima zovemo **brojevi**. I numerali (ukoliko su numerički simboli) i brojevi nam omogućavaju da provodimo statističke obrade kako bismo opisali, analizirali i predviđali. Prema tome, brojevi su nam neophodni da bismo mogli raditi statističke obrade i tako otkrivali nove informacije o objektima, fenomenima ili sistemima.

2.1.3 Atributi

Atributi su svojstva nekog objekta ili fenomena. U prirodnim znanostima obično se koristi termin "svojstva" kako bi se razlikovao neki kvalitet ili kvantitet određenog fizičkog entiteta, dok u socijalnim znanostima se češće koristi termin "atribut" kako bi se referiralo na neki kvantitet/kvalitet ljudskog ili socijalnog fenomena.

Zašto priča o atributima? Zbog toga što znanstvenici nikada ne ispituju čitav objekt. Kada kažemo „ispitivanje ljudi“, to nam ne znači ništa. Ali, kada kažemo "ispitivanje inteligencije" ukazujemo da ispitujemo inteligenciju kao atribut ili svojstvo koje posjeduju ljudi.

2.2 DVIJE TEORIJE MJERENJA VAŽNE ZA PSIHOLOGIJU

2.2.1 Campbellova teorija mjerenja

Norman Robert Campbell (1880-1949) bio je fizičar i filozof znanosti. U svojim radovima, koje je objavljivao od 1920. do 1938, postavio je osnove moderne teorije mjerenja. Njegovo objašnjenje što je mjerenje glasi: *"Mjerenje je proces dodjeljivanja brojeva koji reprezentiraju svojstva, a cilj mjerenja je da se omogući korištenje moćnog oruđa u vidu matematičkih analiza primijenjenih na neko područje znanosti"*

Ono što je najvažnije da znamo vezano za njegovu teoriju je:

1. Kada proučavamo ili mjerimo određene objekte, mi proučavamo ili mjerimo njihove karakteristike ili atribute. Ovo je vrlo bitno zbog toga što se, u svakodnevnom životu, često na osnovu procjene jednog atributa zaključuje o cijelom objektu. Na primjer, ako neko ima pozitivan stav prema fizičkom kažnjavanju djece, on se interpretira kao "strog, autoritaran, hladan, bešćutan...". A mi smo samo saznali da osoba smatra da je fizičko kažnjavanje djece prihvatljivo. Kada donosimo odluku o promociji na poslu, tada mjerimo tehničke vještine kandidata. Tehničke vještine su samo jedna od osobina koje osoba posjeduje, ne i cijela osoba. Osobina i osoba nikako ne smiju biti poistovjećivani. Međutim, kada poznajemo neke osobine osobe, možemo reći da donekle poznajemo tu osobu jer osoba jeste specifični, jedinstveni skup osobina.

2. Fizički objekti posjeduju dvije vrste atributa ili osobina – kvalitativne i kvantitativne osobine. Razlika, prema Campbellu, je u sljedećem: niti jedna osobina se ne bi trebala tretirati kao kvantitativna dok nije potpuno mjerljiva i dok se ne potvrde nalazi da proces sabiranja kod ove karakteristike u potpunosti poštuje zakonitosti mjerenja. Na primjer, kada težinu jednog tijela od 2 kilograma dodamo težini drugog tijela, koje teži isto 2 kilograma, dobijemo njihovu zajedničku

težinu od 4 kilograma. S druge strane, kada kombiniramo dva tijela koja imaju jednaku viskoznost ili gustoću uvijek dobijemo tu istu viskoznost, a ne dva puta veću viskoznost. Iz ovog opisa vidimo koliko je teško razlikovanje kvantitativnih i kvalitativnih varijabli primijeniti u psihologiji. Iako varijable kao što su sposobnosti ili znanje često tretiramo kao kvantitativne, prema ovoj definiciji one nisu kvantitativne (osoba čiji je DQI 120 nikako nije dva puta inteligentnija od osobe koja ima DQI 60).

Prema Cambellu, mjerenje se provodi i na jednoj i na drugoj vrsti varijabli. Ali, kvantitativne varijable možemo mjeriti direktno, dok kvalitativne možemo mjeriti isključivo indirektno.

Glavne teze teorije mjerenja prema Campbellu su:

1. Ne može se mjeriti objekt nego samo njegove osobine

2. Postoje dvije vrste osobina: kvantitativne i kvalitativne

3. Mjerenje se provodi i na kvantitativnim i na kvalitativnim osobinama, ali su različiti pristupi mjerenju

4. Mjerenje kvantitativnih osobina je preciznije i na višem nivou nego mjerenje kvalitativnih osobina

5. Kvantitativne osobine povinuju se zakonitostima ili pravilima koje važe sa sabiranje i mjere se direktno

6. Kvalitativne osobine se ne povinuju zakonitostima sabiranja i mjere se indirektno.

2.2.2 Stevensova teorija mjerenja

Nakon Campbella, veliki doprinos teoriji mjerenja dao je Stanley Smith Stevens, profesor psihologije na Harvard University. Njegova glavna područja interesa bila su psihofizika i psihoakustika. Dao je veliki doprinos mjerenju u psihologiji kroz Stevensov zakon jačine kojim je definirao odnos između intenziteta podražaja i percepcije intenziteta.

U teoriji mjerenja poznat je po definiranju skala mjerenja (Stevens S. , 1946). Kao i Campbell, smatrao je da je mjerenje pridruživanje nekog numeričkog sistema opaženim razlikama u osobinama kod ljudi, objekata ili događaja prema nekom pravilu. Stevens je istaknuo da pridruživanje brojeva prema različitim pravilima ili konvencijama vodi ka različitim skalama mjerenja: nominalnoj, ordinalnoj, intervalnoj i omjernoj skali. Kasnije je predložio i logaritamsku intervalnu skalu koja je izvedena skala. Stevens navodi (1951):

> *"Skale postoje prije svega jer postoji izomorfizam između atributa numeričkih serija i empirijskih operacija koje se mogu provoditi nad određenim aspektima objekata. Ovaj izomorfizam je parcijalan jer je nemoguće upariti sve osobine brojeva s osobinama objekata na potpuno sistematičan način. Ali neke osobine objekata mogu biti povezane semantičkim pravilima s nekim osobinama numeričkih serija. U radu s različitim aspektima objekata, možemo koristiti empirijske operacije za a) određivanje kvalitete (osnove za klasificiranje stvari), b)*

rangiranja po veličini i c) određivanje kada su veličine razlika i omjera između aspekta mjerenih objekata jednake."

2.3 MJERNE SKALE

Podjela mjernih skala koju koristimo u psihometriji upravo je ona koju je predložio Stevens u svojim radovima. Stevens (1946) kaže da skala mjerenja zavisi od vrste empirijskih operacija koje se mogu izvoditi s podacima. Ove operacije limitirane su prirodom karakteristike koja se mjeri / skalira i odabirom procedura.

U tabeli 2-1 dat je prikaz karakteristika 4 nivoa skala mjerenja.

Tabela 2-3 Osnovne karakteristike skala mjerenja

Skala	**Osnovna empirijska operacija**	**Matematička grupna struktura**	**Dozvoljena statistika**
Nominalna	Određivanje jednakosti	*Permutacijske grupe* f(x) = x' pri čemu je f(x) bilo koja zamjena jedan za jedan	Broj slučajeva Modus Kontingencijski koeficijent
Ordinalna	Određivanje veće - manje	Izotone grupe F(x) = x' pri čemu je f(x) bilo koja monotona rastuća funkcija	Centralna vrijednost Percentili
Intervalna	Određivanje jednakosti intervala ili razlika	Generalne linearne grupe F(x) = ax + b	Aritmetička sredina Standardna devijacija Pearsonova korelacija
Omjerna	Određivanje jednakosti omjera	Slične grupe F(x) = ax	Koeficijent varijabilnosti (koji je isti samo u slučaju kada se množi s konstantom)

2.3.1 Nominalna skala

Nominalna skala predstavlja najmanje restriktivan način dodjeljivanja brojeva. U slučaju nominalne skale, brojevi se dodjeljuju kao simboli ili opisi i mogu biti zamijenjeni bilo kojim drugim simbolima, recimo slovima ili riječima. Na primjer, u jednoj školi odjeljenja su numerirana brojevima (VIII1, VIII2), a u drugoj slovima (VIIIa, VIIIb). Suština korištenja broja je samo da se označi grupa. S obzirom na to da je svrha obilježavanja zadovoljena i u situaciji kada obilježja zamijene mjesta (na primjer, VIII1 postane VIII2, a VIII2 postane VIII1) ova skala ostaje invarijatna prilikom zamjene ili permutacije brojeva kao simbola. Unutar kategorija obilježenih brojevima nalazi se određeni broj ispitanika, a ono što je moguće od statističkih postupaka je određivanje najfrekventnije grupe (modusa) i korištenje kontingencijskih metoda za testiranje distribucije slučajeva između grupa (na primjer χ^2 test).

Nominalna skala je najjednostavnija forma mjerenja, a neki autori (prema Stevensu, 1946) smatraju da nije primjereno ovo zvati mjerenjem. Međutim, ako je mjerenje "dodjeljivanje brojeva prema nekom pravilu", onda poštivanje pravila "ne smije se dodijeliti isti broj kao oznaka za dvije različite grupe" kvalificira nominalnu skalu kao mjerenje.

2.3.2 Ordinalna skala

Ordinalna skala proizlazi iz operacije „redanje po veličini“. Kako nijedna transformacija koja čuva rangove neće utjecati na variranje, ova skala se može zvati i izotonična.

Većina mjerenja u psihologiji su, u osnovi, ordinalna mjerenja. Prema Stevensu (1946) mjerenja osobina ličnosti, inteligencije, školske ocjene... sve su to ordinalne skale. Na ordinalnoj skali sve statističke procedure koje uključuju aritmetičku sredinu i standardnu devijaciju ne mogu se koristiti. S druge strane, postoji i opravdanje za korištenje parametrijskih metoda jer nas, u velikom broju slučajeva, vode ka vrlo korisnim rezultatima.

Ordinalna skala daje informaciju o poretku objekata u nekoj grupi, ali ne postoji jednakost u razlikama među rangovima.

2.3.3 Intervalna skala

Intervalnom skalom dolazimo do pojma "kvantitativno". Intervalna skala nam daje informaciju o poretku, posjeduje jednakost razlika među intervalima, ali nema apsolutne nule. Nula na intervalnoj skali je arbitrarno određena, odnosno rezultat 0 ne znači da neki atribut uopće ne postoji, nego da njegovo prisustvo nije zahvaćeno tom metodom mjerenja.

Primjeri intervalne skale u mjerenju su skale za mjerenje temperature u celzijima i farenhajtima. Na primjer, temperatura od 30 celzija je 88 stepeni farenhajta. Da je ovo omjerna skala, onda bi temperatura od 15 stepeni celzija (upola manja od početne) bila 44 farenhajta, a to nije tako.

Formula za transformaaciju temperature iz jedne u drugu skalu je:

2-1

$$\frac{F-32}{9} = \frac{C}{5}$$

Temperatura od 30 celzija je prema formuli:

$$\frac{F-32}{9} = \frac{30}{5}$$

F - 32 = 6*9

F = 54 + 32

F = 86

Temperatura od 15 celzija je:

$$\frac{F - 32}{9} = \frac{15}{5}$$

F - 32 = 3*9

F = 27+32

F = 59

Iz ovog primjera vidimo da dva puta veći broj na skali celzija nije i dva puta veći broj u farenhajtima. Ovo važi i u obrnutom smjeru. Osim toga, unutar same skale kvalitativna razlika između temperatura nije proporcionalna njihovoj kvantitativnoj razlici. Temperatura od 30 celzija ne znači da je dva puta toplije od temperature od 15 celzija.
Ovaj problem se velikim dijelom javlja zbog nepostojanja apsolutne nule kao tačke u kojoj neko svojstvo ne postoji. Nedostatak apsolutne nule onemogućava korištenje čvrste uporišne tačke ili sidra na osnovu koje bi se komparirala ostala opažanja.

Psiholozi teže da većinu mjerenja obavljaju na intervalnom nivou i ponekada u tome uspijevaju. Mi mjerenja tretiramo kao intervalne skale prilikom određivanja interesa, stavova, ličnosti motivacije, vrijednosti i drugih psiholoških karakteristika. Skale dizajniramo tako da počnemo od arbitrarne nule kao tačke u kojoj neka osobina ne postoji ili neko ponašanje se nikada ne ispoljava i skaliramo da najviših intenziteta ili najvećih frekvencija.

Primjer: Zamislimo da određujemo učestalost odlaska na pozorišne predstave česticom datom ispod:

PITANJE: Koliko često odlazite na pozorišne predstave?

0. **Nikada**
1. **Rijetko (jednom do dva puta godišnje)**
2. **Ponekada (jednom do dva puta mjesečno)**
3. **Često (jednom sedmično)**

Ako je zaokružena 0, zaključujemo da ispitanik nikada nije bio u pozorištu. Ukoliko zaista nikada nije bio, onda je to stvarna 0. Ali mnogo je teži zadatak kreirati mjernu skalu na kojoj će pojedini podioci precizno reprezentirati jednake razlike u sukcesivnom povećanju kvantiteta ispitivane varijable. Ako pogledamo u ovom pitanju razliku između odgovora 1 i 2 i odgovora 2 i 3, lako uočavamo da to nisu iste razlike. Koliko god da je matematički razlika 1 i 2 jednaka razlici 2 i 3 u frekvenciji odlazaka u pozorište, ta razlika nije jednaka. Ukoliko nismo uspjeli postići jednakost intervala, bez obzira na to što se intervali čine jednakim, mjerenje ostaje na ordinalnom nivou.

2.3.4 Omjerna skala

Omjerna skala, koju uglavnom nalazimo u mjerenjima u prirodnim znanostima, je skala koja ima zadovoljene sve uvjete: veličinski odnos, jednakost intervala i jednakost omjera, te apsolutnu nulu. Ovo je skala s kojom se mogu raditi sve statističke operacije, rezultati se mogu transformirati tako da ih samo pomnožimo s nekom konstantom. Nula na omjernoj skali znači da neko svojstvo ne postoji (čak i onda kada se to nikada nije desilo, ali se može desiti).

2.4 POGREŠKE U MJERENJIMA

Mjerenje sa sobom uvijek nosi mogućnost pogreške, odnosno pogrešnog očitavanja. Što je, ustvari pogreška mjerenja? Pogreška mjerenja je razlika između stvarnog rezultata ispitanika na nekom testu i onoga koji smo mi zabilježili, odnosno opaženog rezultata na testu. Na primjer, stvarni rezultat jednog djeteta na testu inteligencije je 120, ali je u izradi testa njegov rezultat iznosio 117. Sljedeći put, kada smo mu dali isti test, njegov rezultat je bio 115. Ovo odstupanje se javlja zbog nepreciznih mjernih instrumenata. Kako mi nikada ne znamo, niti imamo mogućnost da izmjerimo stvarni rezultat ispitanika, svjesni smo da opaženi rezultat u sebi sadrži i pogrešku mjerenja.

Pogreške mjerenja dijelimo u dvije osnovne grupe:

1. Slučajne pogreške ili pogreške koje nastaju uslijed djelovanja nesistematskih faktora

2. Sistematske pogreške ili pogreške koje nastaju zbog djelovanja sistematskih faktora

2.4.1 Slučajne pogreške

Slučajne pogreške nastaju kao rezultat djelovanja nesistematskih faktora. Na primjer, za vrijeme mjerenja jednostavnog vremena reakcije na svjetlo ispitanik je imao 20 pokušaja. U prvom pokušaju njegovo vrijeme bilo je 141 ms, u drugom pokušaju se malo pomjerio na stolici, pa mu je rezultat bio 143 ms, u desetom pokušaju je osjetio malu neugodnost u predjelu vrata i rezultat mu je bio 145 ms, onda je prst držao malo bliže dugmetu i ostvario rezultat od 137 ms. Na kraju serije od 20 pokušaja vidjeli smo da je naš ispitanik u svakom pokušaju imao nešto drukčiji rezultat. U nekim pokušajima bio je niži od prosjeka serije, a u nekim pokušajima viši od prosjeka serije. Ovaj primjer opisuje dvije ključne karakteristike kako nesistematski faktori djeluju na rezultat:

1. djeluju po slučaju tako da se vjerovatnoća pogreške mijenja po slučaju

2. djeluju u oba smjera tako da nekada povećavaju, a nekada smanjuju rezultat.

2.4.2 Sistematske „pogreške“

Sistematske pogreške u mjerenju javljaju se pod utjecajem varijabli koje na rezultate djeluju tako da ih mijenjaju kod svih ispitanika u istom smjeru. Ove varijable mogu biti namjerno manipulirane/ uvođene kako bi se odredio utjecaj te varijable na rezultate ispitanika. Sistematski faktori su predvidljiviji i njima ili manipuliramo (u slučaju eksperimentalnih varijabli) ili ih nastojimo držati pod kontrolom za vrijeme ispitivanja. Na primjer, ukoliko ispitujemo inteligenciju kod učenika, nastojimo da u svim razredima to radimo u isto doba dana kako bismo izbjegli utjecaj vremena ispitivanja na rezultate učenika. Ali nikada nismo sigurni da smo sve sistematske utjecaje stavili pod kontrolu. Kako ćemo vidjeti u ovom priručniku, metoda mjerenja je jedna od najčešćih sistematskih pogrešaka koje utječu na rezultate.

3 PSIHOLOŠKI TEST

Kada neki instrument možemo nazvati psihološkim testom? Kako dijelimo psihološke testove? Koja je svrha korištenja psiholoških testova?

Psihološki test predstavlja glavno sredstvo za rad psihologa. Riječ TEST koristi se u različitim kontekstima tako da je važno znati što podrazumijevamo pod terminom psihološki test.

3.1 DEFINICIJA PSIHOLOŠKOG TESTA

Cronbach (1990, str. 32) je dao definiciju testa kao „sistematskog postupka za opažanje ponašanja i njegovo opisivanje uz pomoć numeričkih ljestvica ili uspostavljenih kategorija“.

Urbina (2004) detaljnije definira test kao sistematsku proceduru prikupljanja podataka o uzorcima ponašanja relevantnih za kognitivno ili afektivno funkcioniranje, njihovo mjerenje i ocjenjivanje u odnosu na standarde.

Vidimo da ova definicija sadrži neke ključne pojmove koji razlikuju psihološke testove od drugih testiranja:

1. Sistematska procedura podrazumijeva da ispitivač prikuplja informacije, odnosno provodi testiranje jednako sa svim ispitanicima u istoj i primjerenoj situaciji.
2. Uzorci ponašanja – test sadrži jedan mali uzorak ponašanja koji je vezan za latentnu karakteristiku koju pokušavamo mjeriti. To isto tako znači da test sadrži reprezentativan uzorak ponašanja prema njihovoj empirijskoj ili praktičnoj važnosti, te omogućava efikasno ispitivanje s obzirom na ograničeno vrijeme trajanja testiranja.
3. Mjerenje i ocjenjivanje podrazumijeva da se u testu primjenjuje numerički ili kategorički sistem prema ranije utvrđenim pravilima.
4. Standardi – cilj mjerenja i testiranja jeste da se ustanovi položaj rezultata ispitanika u odnosu na rezultate skupine kojoj on pripada.

Prema Urbinoj (2004) riječ TEST treba koristiti kao termin koji označava one instrumente i procedure u kojima se odgovori ispitanika procjenjuju na osnovu njihove kvalitete ili tačnosti. To su uglavnom instrumenti kojima se utvrđuju neki aspekti kognitivnog funkcioniranja, znanja, vještina ili sposobnosti. S druge strane, u psihologiji je možda i više instrumenata u kojima odgovor ispitanika ne može biti tretiran kao tačno / pogrešno, već kao mišljenje o samoj tvrdnji. Na primjer, na pitanje „Ja više volim da ostanem čitati knjigu kod kuće, nego da izađem s društvom“ ne postoji tačan ili pogrešan odgovor. Svaki odgovor je tačan ukoliko on predstavlja stvarno slaganje ispitanika s tvrdnjom. U ovu grupu instrumenta spadaju oni koji mjere psihološke konstrukte kao što su motivacija, stavovi, vrijednosti, emocije, karakteristične reakcije ispitanika na ljude, situacije ili druge podražaje. Ove instrumente obično nazivamo inventarima, upitnicima, ček listama, projektivnim tehnikama i često su grupirani pod širokim nazivom „testovi ličnosti“.

U nastavku knjige koristit ćemo termin test kako bismo označili bilo koji instrument, ali prethodno napisano treba imati na umu kada razmišljamo o razlikama u vrednovanju i interpretiranju rezultata na testovima i drugima instrumentima.

3.2 DRUGI TERMINI KOJI SE KORISTE U VEZI S TESTOVIMA

Postoji niz drugih termina koje koristimo u vezi s testovima. Jedan od tih termina je *skala*. Termin skala može opisivati različite stvari:

1. Test napravljen od nekoliko dijelova, npr. Stanford – Binetova skala
2. Subtestove ili niz pitanja unutar testa koji mjere odvojene i specifične karakteristike, npr. Skala depresije unutar Minesota multifaznog upitnika ličnosti (MMPI);
3. Niz subtestova koji imaju zajedničku karakteristiku, npr. verbalne skale unutar Wechslerovog testa inteligencije
4. Zasebni instrument koji sadrži pitanja za mjerenje jedne karakteristike, npr. Skala samopoštovanja (Rosenberg, 1965)
5. Numerički sistem koji se koristi da bi se ocijenio ili validirao neki mjereni konstrukt, npr. skala od 1 do 5 pri čemu 1 znači „nimalo se ne slažem", a 5 znači „u potpunosti se slažem".

Iako se termin „skala" koristi da se označe različite stvari, u psihometrijskom smislu ovaj termin je jasan i označava grupu pitanja koja mjere jednu varijablu i koja su poredana prema njihovoj težini ili intenzitetu. Proces kojim se dolazi do poretka pitanja u testu naziva se skaliranje.

Drugi termin koji trebamo poznavati je *baterija testova.* Ovaj termin označava *grupu* testova ili subtestova koji se zadaju jednoj osobi u jednoj vremenskoj tački. Kada izdavač pakuje nekoliko testova zajedno zbog specifičnih razloga, onda se riječ baterija pojavljuje i u naslovu i čitava grupa testova se onda tretira kao jedan jedinstven instrument. Na primjer, baterija testova koja se koristi uglavnom za profesionalnu selekciju je *DAT-BOS – Diferencijalni testovi sposobnosti za selekciju – Baterija općih sposobnosti* (Bennet, Seashore i Wesman, 2006). Dakle, termin baterija testova koristi se za specifično odabranu grupu testova kako bi se odgovorilo na neko dijagnostičko pitanje.

I na kraju, još jedan termin važan za testove i testiranje je *testna serija.* Testna serija označava grupu testova koji su pripremljeni tako da mjere istu osobinu, ali na različitim razvojnim stepenima. Primjer testne serije su upravo Wechslerovi testovi za mjerenje inteligencije koji imaju tri nivoa – za djecu od 4 do 6 godina starosti, za djecu od 6 do 17 godina, te za odrasle. Osnovno obilježje testnih serija je da testovi mjere iste konstrukte, ali prilagođeno razvojnim grupama, te se rezultati ispitanika mogu uspoređivati i testovi se mogu koristiti za longitudinalno praćenje promjena.

3.3 VRSTE PSIHOLOŠKIH TESTOVA

Postoji mnogo podjela psiholoških testova, kao što postoji i mnogo područja u kojima se koriste psihološki testovi. Bucik (1997) u svojoj knjizi ponudio je vrlo iscrpnu listu kriterija na osnovu kojih se dijele testovi. U našem priručniku prikazat ćemo osnovne podjele bez kojih ne možemo pratiti druge sadržaje.

1) Podjela testova s obzirom na stepen upotrebe

Prema načinu upotrebe testove dijelimo na standardizirane i nestandardizirane. Prema Urbinoj (2004), psihološki testovi se tretiraju kao standardizirani kada zadovoljavaju dva kriterija koja se tiču objektivnosti procesa testiranja. Prvi kriterij je uniformnost procedure testiranja u svim važnim aspektima, počevši od zadavanja testa, bodovanja do interpretacije rezultata. Cilj standardizacije procedura je da se sve varijable vezane za ispitivača stave pod kontrolu tako da svaki ispitanik koji učestvuje u testiranju bude u istoj situaciji. Drugi kriterij je postojanje standarda ili normi za evaluaciju rezultata. Norme (ili standardi) predstavljaju rezultate velikih grupa ispitanika (koje zovemo i normativnim uzorkom) koji su prikupljeni u procesu razvijanja testa. Ovi zbirni rezultati, njihova distribucija i parametri distribucije (prosjeci i varijabilitet) se prikazuju u priručniku za test, a rezultat našeg ispitanika procjenjujemo na osnovu usporedbe njegovog individualnog rezultata s rezultatima normativnog uzorka. Testovi koji ne zadovoljavaju ove kriterije, odnosno nemaju izrađene norme niti standardiziranu proceduru testiranja, zovemo nestandardiziranim testovima ili kako bi profesor Bucik (1997) rekao – testovima za domaću upotrebu.

2) Podjela testova s obzirom na vrstu mjerene karakteristike

U psihologiji mjerimo različite karakteristike. Neke od njih su takve da na kraju testiranja možemo reći: „On je bolji, a on je slabiji“, dok za neke druge karakteristike to ne možemo reći, nego konstatirati: „Ispitanici se razlikuju“. Prvu grupu testova zovemo *testovima maksimalnog učinka ili testovima sposobnosti,* dok drugu grupu testova zovemo *testovi tipičnih ponašanja.*

U testove maksimalnog učinka spadaju:

- Testovi inteligencije
- Testovi postignuća ili uspjeha (engl. aptitude test), znanja ili spretnosti
- Testovi nadarenosti, kreativnosti ili talenta

U testove tipičnih ponašanja spadaju testovi koji mjere karakteristike za koje nema tačnih i pogrešnih odgovora nego se pretpostavlja da su odgovori ispitanika odraz neke njegove osobine, kao što je osobina ličnosti, stava, interesa i slično. U ovu grupu testova spadaju:

- Testovi strukture ličnosti (u užem smislu riječi ispituju strukturu ličnosti)
- Testovi interesa, stavova, vrijednosti...

Osim testova maksimalnog učinka i testova tipičnih ponašanja, posebnu grupu tvore *testovi za neuropsihološku procjenu* kojima se ustanovljavaju različite promjene u neuropsihološkom području na senzornoj, kognitivnoj, motornoj i ponašajnoj razini.

3) Podjela testova s obzirom na dimenzionalnost

Testovi mogu biti jednodimenzionalni i višedimenzionalni. Jednodimenzionalnim testovima mjeri se samo jedna karakteristika, dok se višedimenzionalnim testovima mjeri više karakteristika. Rezultati se u pravilu formiraju za svaku karakteristiku posebno, osim ako teorija dozvoljava formiranje jednog rezultata.

4) Podjela testova s obzirom na vrijeme rješavanja

Prema vremenu koje ispitanik ima na raspolaganju za rješavanje, testove dijelimo na *testove snage* i *testove brzine*. Testovi snage obično sadrže zadatke koji su različiti po težini – od vrlo laganih do vrlo teških. Ispitanik ima dovoljno vremena da pokuša riješiti sve zadatke, a koje zadatke će uspjeti da riješi zavisi od njegovih znanja ili sposobnosti. U testovima brzine zadaci su obično lagani do srednje teški, ali je vrijeme rada ograničeno tako da ispitanik mora dosta brzo procesirati informacije kako bi riješio više zadataka. Dakle, kod testova snage rezultat najviše zavisi od sposobnosti, znanja ili vještine ispitanika, dok kod testova brzine rezultat zavisi od brzine procesiranja informacija (što može biti bitno za razlikovanje ispitanika).

5) Podjela testova s obzirom na interpretacije rezultata

Prema interpretaciji rezultata testove dijelimo na direktne i indirektne. Direktne testove još zovemo i psihometrijskim testovima. U njima se ispitanike direktno pita da procijene stepen slaganja ili učestalost nekog ponašanja relevantnog za mjerenje psihološke karakteristike. Indirektni testovi baziraju se na projekcijskoj hipotezi prema kojoj lične interpretacije ispitanika nekog nestrukturiranog ili nejasnog materijala predstavljaju projekciju nesvjesnih ili potisnutih potreba, motiva ili konflikata. Kod indirektnih testova, ispitaniku se daje određeni nestrukturirani ili nejasni materijal za koji on daje svoju interpretaciju. Interpretacija se onda uspoređuje s katalozima odgovora i donosi se zaključak o prisustvu ili intenzitetu nekog motiva, potrebe ili konflikta koji ispitanik ne bi drukčije prijavio. U ovu grupu spadaju asocijacijske tehnike (npr. Rorschachov test mrlja), tehnike dopunjavanja (npr. Test nedovršenih rečenica), konstrukcijske tehnike (npr. Test tematske apercepcije), te ekspresivne tehnike (npr. Test crtanja čovjeka)

6) Podjela testova s obzirom na broj ispitanika

Prema broju ispitanika koji učestvuju u jednom navratu, testove dijelimo na *individualne i grupne*. Kod individualnih testova, u trenutku ispitivanja prisutan je samo jedan ispitanik, dok se grupni testovi mogu zadavati većem broju ispitanika. Grupno primjenjivi testovi mogu se koristiti i sa jednim ispitanikom, dok to vrlo često nije slučaj s individualnim testovima.

7) Podjela testova s obzirom na način ocjenjivanja rezultata

Ponekad položaj rezultata ispitanika ocjenjujemo u odnosu na grupne rezultate njemu sličnih ispitanika (normativne grupe), a ponekada u odnosu na neki postavljeni kriterij. Na osnovu toga testove dijelimo na *normativne i kriterijske*. Kod normativnih testova rezultat ispitanika se uspoređuje s rezultatima neke grupe i tako se ustanovi položaj njegovog rezultata. Na primjer, na testu iz Psihometrije rezultat jednog studenta možemo usporediti s distribucijom rezultata generacije i reći: „Ovaj student je među 5 % najslabijih na testu“. Kriterijski testovi su testovi u kojima je unaprijed određen neki kriterij s kojim se uspoređuje rezultat ispitanika i njegov položaj. Na primjer, ukoliko je kriterij za dobivanje ocjene 9 (B) bio 85 %, a naš ispitanik postigao 86 % on će dobiti ocjenu 9(B) bez obzira na to što se unutar grupe nalazi među 5 % najslabijih.

8) Podjela testova s obzirom na korištenje jezičkih sadržaja

Neki testovi sadrže zadatke koji zahtijevaju od ispitanika da je opismenjen (da zna čitati ili da može čuti verbalnu uputu, te da zna pisati ili usmeno odgovoriti). Ovakve testove zovemo *verbalnim testovima*. Drugi testovi su konstruirani tako da ispitanici ne moraju poznavati jezik da bi riješili test jer se koriste figure, crteži, simboli i slično. Ovakve testove zovemo *neverbalnim testovima*.

Iz ovih podjela vidimo da u psihologiji imamo mogućnost korištenja raznolikih mjernih instrumenata. Odluka o tome koju vrstu testa ćemo koristiti mora biti donesena uzimajući u obzir sve faktore, počevši od karakteristika ispitanika do situacije u kojoj će se provoditi testiranje.

3.4 CILJEVI TESTIRANJA

Testirati samo radi testiranja je gubitak vremena i energije. Testiranje se uvijek radi s nekim ciljem. Ciljeve testiranja možemo svrstati u nekoliko grupa:

1) Testiranje radi donošenja odluka

Testiranje nam služi kako bismo donosili valjane i informirane odluke. Te odluke se obično tiču selekcije kandidata, unapređivanja ili postavljanja na određena mjesta, klasificiranja, dijagnosticiranja ili bilo kojeg drugog odlučivanja o osobi, grupi, organizaciji, državi i programu. Donošenje odluka je često praćeno i negativnim reakcijama ukoliko odluka nije u skladu sa željama onih koji su učestvovali u testiranju. To je razlog zašto psiholozi vode računa da koriste metrijski provjerene testove jer samo tako možemo braniti naše odluke pred zainteresiranim stranama.

“Ukoliko mjerenje išta znači, onda je to zbog toga što ima neki vidljivi efekat na odluke i ponašanja. Ukoliko ne možemo identificirati odluku na koju će se odraziti mjerenje... onda mjerenje jednostavno nema vrijednosti“

(Hubbard, 2014)

2) Istraživanja

Testovi se koriste u svim područjima psihologije za istraživanja. Oni su naše najvažnije sredstvo za razumijevanje razvoja i odnosa među kognitivnim, afektivnim i ponašajnim karakteristikama.

3) Samospoznaja i lični razvoj

Osim što predstavljaju sredstvo koje koristi ispitivač kako bi bolje razumio stanje i karakteristike ispitanika, testovi imaju vrlo važnu ulogu u okviru ličnog razvoja pojedinca. U tom smislu, testovi se pripremaju tako da sam ispitanik u procesu testiranja stječe uvid u vlastite rezultate i time potiče vlastiti rast i razvoj što su Finn i Tonsager (1997) nazvali terapeutskim modelom ispitivanja. Ovu svrhu testova najviše vidimo u područjima kliničke i savjetodavne psihologije, ali je sigurno da će u budućnosti korištenje testova u svrhu razumijevanja sebe i vlastitog razvoja biti sve više u upotrebi.

4 OSNOVNA STATISTIKA – PONAVLJANJE JE MAJKA ZNANJA

Što je varijanca, a što kovarijanca? Kakva je razlika između kovarijance i korelacije? Kako organiziramo podatke u matrice i kako provodimo računske operacije s matricama?

Psihološka mjerenja postoje jer se ljudi razlikuju prema svojim karakteristikama. Psihologija se bavi istraživanjima i objašnjavanjem individualnih razlika među ljudima. Postoje najmanje dvije vrste razlika za koje smo zainteresirani: interindividualne i intraindividualne razlike. Interindividualne razlike odnose se na razlike među pojedincima, dok se intraindividualne razlike odnose na različite rezultate koje ista osoba postiže na nekom testiranju pod utjecajem bilo sistematskih, bilo nesistematskih faktora. Dakle, varijabilnost je u srcu psiholoških istraživanja i primjene psiholoških spoznaja u realnom životu, a kvaliteta istraživanja i intervencija bazira se na mogućnostima mjernih postupaka da detektiraju postojeće razlike.

Rezultate ispitanika prikazujemo u obliku neke funkcije. Ukoliko prezentiramo rezultate ispitanika na jednoj varijabli, onda se na apscisi nalaze rezultati, a na ordinati frekvencije ispitanika. U slučaju kad prikazujemo odnose između dvije varijable, tada na apscisi imamo vrijednosti jedne, a na ordinati vrijednosti druge varijable.

Da bismo mogli usvojiti nova znanja iz psihometrije, potrebno je da se podsjetimo na važne pojmove koje smo učili u Statistici. To su prije svega **aritmetička sredina i varijanca** kao parametri distribucije rezultata, te **korelacija i kovarijanca** kao parametri odnosa među dvije varijable.

Zamislimo da smo ispitali grupu od 12 zaposlenika Upitnikom zadovoljstva poslom. Njihovi rezultati na testu prikazani su u tabeli 4-1 u koloni 2:

Tabela 4-1 Određivanje varijance rezultata

Ispitanik	**Rezultat X**	**X – M** **Devijacijski rezultat**	**$(X-M)^2$** **Kvadrirani devijacijski rezultat**
1.	58	11	121
2.	53	6	36
3.	54	7	49
4.	56	9	81
5.	40	-7	49
6.	47	0	0
7.	42	-5	25
8.	41	-6	36
9.	38	-9	81
10.	37	-10	100
11.	46	-1	1
12.	51	4	16

4.1 ARITMETIČKA SREDINA

Aritmetička sredina kao mjera centralne tendencije je jedan od osnovnih parametara distribucije. To je prosječni rezultat koji ostvaruje grupa ispitanika ili jedan ispitanik na nizu ponovljenih mjerenja. Matematička formula za aritmetičku sredinu je

4-1

$$M = \frac{\sum X}{N}$$

U našem primjeru, aritmetička sredina bi iznosila:

$$M = \frac{58+53+54+56+40+47+42+41+38+37+46+51}{12} = 46{,}9 \approx 47$$

4.2 ARITMETIČKA SREDINA I VARIJANCA BINARNE VARIJABLE

U psihologiji često susrećemo binarne varijable. Binarna varijabla je ona varijabla koja može poprimiti samo dvije vrijednosti. Na primjer, u testovima postignuća rezultati na zadatku se kvalificiraju kao Tačno ili Pogrešno, pri čemu se tačnom odgovoru dodjeljuje 1 bod, a netačnom 0 bodova. I u upitnicima tipičnih ponašanja srećemo binarne varijable. Na primjer, u upitniku ličnosti pitanje može biti „Volim upoznavati nove ljude“ i od ispitanika se traži da odgovore sa Da ili Ne. Ukoliko odgovori Da – dobije 1 bod, ukoliko odgovori Ne – dobije 0 bodova. Važno je (opet) primijetiti da 1 na testu znanja i 1 na upitniku ličnosti ne znače isto iako je korišten isti numerički simbol. Na testu znanja rezultat 1 znači da je ispitanik znao zadatak i da je bolji od onog koji nije znao, dok je na upitniku ličnosti to kvalitativna razlika i ne možemo reći da je neko bolji zato što voli upoznavati nove ljude.

Ukoliko je binarna varijabla kodirana sa 0 i 1, onda aritmetička sredina te varijable predstavlja proporciju odgovora kodiranih sa 1 i označava se malim slovom p:

4-2

$$M = p = \frac{\sum X}{N}$$

Suprotno od p, kao proporcije odgovora koji su kodirani sa 1, je q – kao proporcija odgovora kodiranih sa 0. S obzirom na to da su p i q proporcije, važi:

$$p + q = 1$$

Da vidimo čemu je jednaka varijanca binarne varijable:

$$V_{bin} = \frac{\sum (X-p)_i^2}{N} = \frac{\sum (X^2 - 2pX + p^2)}{N} =$$

$$V_{bin} = \frac{\sum X^2}{N} + \frac{\sum p^2}{N} - \frac{2p\sum X}{N}$$

U slučaju kada imamo binarnu varijablu tada je $1^2 = 1$, što znači da će i $\frac{\sum X^2}{N} = \frac{\sum X}{N}$, a to znači da će naša gornja formula biti

$$V_{bin} = p + \frac{\sum p^2}{N} - 2p^2$$

S obzirom na to da je ukupan broj p jednak broju ispitanika, odnosno p je konstanta za sve ispitanike, možemo napisati da je $\frac{\sum p^2}{N} = \frac{Np^2}{N} = p^2$.

Naša formula će sada izgledati

$$V_{bin} = p + p^2 - 2p^2 = p - p^2 = p(1-p)$$

4-3

$$V_{bin} = p*q$$

Varijanca binarne varijable jednaka je umnošku proporcije odgovora kodiranih sa 1 i proporcije odgovora kodiranih sa 0. Ovo je vrlo važno jer vidimo da varijanca binarne varijable ovisi o visini proporcija p i q. Varijanca binarne varijable bit će najveća kada su ove proporcije jednake i iznose 0,5. To znači da u zadatku imamo najviše varijabilnosti kada imamo podjednak broj odgovora kodiranih sa 1 i 0.

4.3 VARIJANCA

Varijanca i s njom blisko povezana standardna devijacija pokazuju u kojoj mjeri pojedinačni rezultati odstupaju od aritmetičke sredine. Varijanca se računa prema formuli:

4-4

$$V = \frac{\sum (X-M)_i^2}{N} = \frac{\sum d_i^2}{N}$$

Pojedinačni rezultati odstupaju od aritmetičke sredine. Razliku između rezultata i aritmetičke sredine zovemo devijacijski ili otklonski rezultat i označavamo ga slovom ***d***. Varijanca je prosječna kvadrirana suma devijacijskih rezultata i ujedno najmanja prosječna kvadrirana devijacija koja se u nekom skupu brojeva može izračunati od neke konstante, pri čemu je ta konstanta M ili aritmetička sredina. Svaka druga konstanta dala bi veću varijancu.

U našem primjeru zaposlenika i mjerenja zadovoljstva poslom varijanca bi iznosila:

$$V = \frac{(58-47)^2 + (53^{-}47)^2 + \cdots + (51-47)^2}{12} = \frac{121 + 36 + .. + 16}{12} = 49{,}58$$

Brojnik u formuli varijance zovemo uobičajeno ***„suma kvadrata".***

Standardna devijacija je drugi korijen iz varijance i u našem slučaju bi iznosila:

$$st.dev = \sqrt{49{,}58} = 7{,}04$$

Kao parametar varijabilnosti standardna devijacija ima jednu prednost a to je da odražava varijabilnost u terminima sirovih rezultata (rezultata testa) pa je intuitivnija i lakše je razumijemo. Kada nam neko kaže da je aritmetička sredina rezultata 50 bodova, a standardna devijacija 7,69 bodova, nama je sve jasno. Varijanca odražava varijabilnost u terminima kvadriranih devijacijskih rezultata i stoga nam je njeno razumijevanje manje intuitivno. Veličina varijance je određena variranjem rezultata i skalom mjerenja. Varijanca i standardna devijacija vežu se za skalu sirovih rezultata. Prilikom određivanja koliko rezultati variraju uvijek trebamo znati koliki je hipotetski raspon rezultata na testu. Na primjer, ako se rezultati na testu mogu kretati od 0 do 600, a varijanca je 100, to je manje variranje nego da je skala bila od 0 do 60, a varijanca 100.

4.4 KOVARIJANCA

Na što vas podsjeća riječ kovarijanca? Osnova riječi je varijanca, a prefiks ko- ukazuje da se odnosi na nešto zajedničko. Ako varijanca predstavlja variranje rezultata u jednoj varijabli, onda kovarijanca predstavlja kovariranje rezultata za dvije varijable. Dakle, kovarijanca je stepen povezanosti varijabiliteta između dvije distribucije rezultata. Zamislimo da, pored mjere zadovoljstva poslom, za iste ispitanike imamo i rezultate na upitniku stresa na poslu:

Tabela 4-2 Određivanje kovarijance rezultata

Ispitanik	**Rezultat na Skali zadovoljstva poslom (X)**	**Rezultat na Skali stresa (Y)**	**Devijacijski rezultat X ($X - M_X$)**	**Devijacijski rezultat Y ($Y - M_Y$)**	**Umnožak devijacijskih rezultata ($X - M_X$) ($Y - M_Y$)**
1.	58	1,53	11	-0,55	-6,05
2.	53	1,74	6	-0,34	-2,04
3.	54	1,58	7	-0,5	-3,5
4.	56	1,53	9	-0,55	-4,95
5.	40	1,68	-7	-0,4	2,8
6.	47	2,42	0	0,34	0
7.	42	1,70	-5	-0,38	1,9
8.	41	2,79	-6	0,71	-4,26
9.	38	2,53	-9	0,45	-4,05
10.	37	2,74	-10	0,66	-6,6
11.	46	3,26	-1	1,18	-1,18
12.	51	1,53	4	-0,55	-2,2
Aritmetička sredina	46,9	2,1			

Kovarijanca se računa prema formuli:

4-5

$$c_{xy} = \frac{\sum(X - M_X)(Y - M_Y)}{N} = \frac{\sum d_X d_Y}{N}$$

U našem slučaju kovarijanca će iznositi:

$$c_{xy} = -2{,}51$$

Sada se sigurno pitate kako je kovarijanca negativna kada varijanca može biti samo pozitivnog predznaka? Ovaj primjer je namjerno izabran kako biste uvidjeli da su varijanca i kovarijanca dvije različite stvari. Kovarijanca pruža informaciju i o smjeru povezanosti dvije varijable. Kada je kovarijanca pozitivna, onda postoji pozitivna povezanost između varijabli, odnosno porast rezultata u jednoj varijabli praćen je tendencijom porasta rezultata u drugoj varijabli. Kada je kovarijanca negativna onda je povezanost takva da porast rezultata u jednoj varijabli prati tendencija snižavanja rezultata u drugoj varijabli. Da li vas ovo podsjeća na nešto što ste učili u Statistici? U pravu ste, kovarijanca je slična korelaciji. Da vidimo kakva je veza kovarijance i korelacije.

Kao što smo mogli vidjeti u prethodnom primjeru, kovarijanca se izračunava iz sirovih rezultata i na visinu kovarijance utječu skale na kojima su izraženi sirovi rezultati. S druge strane, korelacija se izračunava na osnovu rezultata transformiranih u z-vrijednosti i njena formula je:

4-6

$$r_{xy} = \frac{\sum\left(z_x z_y\right)_i}{N} = \frac{\sum\left(d_x d_y\right)_i}{N s_x s_y} = \frac{\sum\left[(X - M_x)\left(Y - M_y\right)\right]_i}{N s_x s_y}$$

Ako usporedite formulu za kovarijancu i korelaciju, primijetit ćete razliku u brojniku. U kovarijanci koristimo sirovi (bruto) devijacijski rezultat, a u korelaciji z-vrijednost. U nazivniku kod kovarijance imamo samo N, dok kod kovarijance imamo i standardne devijacije. Korelacija je, ustvari, standardizirana kovarijanca. Njihov odnos objašnjava se formulom:

4-7

$$c_{XY} = r_{XY} s_X s_Y \text{ , odnosno } r_{xy} = \frac{c_{XY}}{s_X s_Y}$$

Koeficijent korelacije je jednostavniji za interpretaciju linearnih veza. Kreće se u rasponu od -1 do +1 i ne ovisi o rasponu bruto rezultata. Kada nam neko kaže da je koeficijent korelacije r_{xy} =0,4 ili r_{xy} = -0,4, mi već mnogo znamo o visini i smjeru povezanosti jer korelacija nije vezana za metriku varijabli, dok za interpretaciju kovarijance trebamo informaciju i o varijabilnosti bruto rezultata.

Poznavanje osnovnih pojmova (varijance, kovarijance i korelacije) olakšat će nam razumijevanje psihometrije jer se određivanje metrijskih karakteristika testa umnogome naslanja na ove parametre.

4.5 VEKTORI I MATRICE U PSIHOMETRIJI

U psihologiji, podaci su najčešće prikazani u matricama. Stoga je važno da poznajemo osnovne matrice i njihove karakteristike.

4.5.1 Vektori

Vektor je niz brojeva poredanih po određenom redoslijedu u jednom stupcu (koloni) ili jednom redu. Označavaju se malim slovima. U geometrijskoj prezentaciji, elementi vektora su tačke na osama koordinatnog sistema tako da je, geometrijskim rječnikom rečeno, broj elemenata vektora određen brojem osi koordinatnog sistema. Svaki element vektora ukazuje na poziciju na odgovarajućoj osi. U koordinatnom sistemu sa dvije osi, prvi element vektora je tačka na X osi, a drugi element vektora tačka na Y osi. Pojedini elementi vektora x označavaju se imenom vektora i odgovarajućim indeksom, na primjer, vektor sa deset elemenata označava se sa x_i pri čemu je

i = 1,…,10.

Kada vektori imaju brojeve poredane u stupcima (kolonama), zovemo ih stupčani, a kada su brojevi poredani u redu, vektor zovemo retčani.

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \; stup\check{c}ani\ vektor$$

$$a = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \; ret\check{c}ani\ vektor$$

Dva vektora su jednaka ako i samo ako imaju sve elemente jednake i istog su oblika:

a=(8,5); b=(8,4); c=(8,5) d= $\begin{pmatrix} 8 \\ 5 \end{pmatrix}$

a≠b; a = c; a ≠ d; b ≠c; b ≠d; c ≠d

4.5.1.1 Operacije s vektorima

4.5.1.1.1 Sabiranje i oduzimanje vektora

Vektori koji imaju jednak broj elemenata i jednak oblik sabiraju se i oduzimaju tako da se saberu ili oduzmu korespondentni elementi u vektorima

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \\ a_3 - b_3 \end{pmatrix}$$

4.5.1.1.2 Množenje vektora

Kada množimo vektor s nekim brojem (skalarom), kao produkt dobivamo vektor čiji su elementi jednaki umnošku elemenata vektora i skalara.

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} * S = \begin{pmatrix} a_1 * S \\ a_2 * S \\ a_3 * S \end{pmatrix}$$

Da bismo mogli množiti dva vektora, mora biti zadovoljen isti kriterij kao i kod matrica (jer su vektori samo posebni oblik matrice sa jednim redom ili jednom kolonom) – broj kolona u prvom vektoru mora biti jednak broju redova u drugom vektoru. Rezultat množenja dva vektora dobijemo tako da pomnožimo korespondentne elemente, a njihove produkte saberemo. Krajnji produkt množenja dva vektora je broj ili skalar.

$$\begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} * \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3 = S$$

4.5.1.2 Jedinični i nul vektori

Jedinični i nul vektori predstavljaju važne vektore u određivanju valjanosti testova (o kojoj ćemo kasnije učiti) te ih trebamo upoznati.

Jedinični vektor je vektor koji kao jedan element ima jedinicu, a sve ostale elemente 0

$$j_1 = (1,0,0,0,...,0)$$

$$j_2 = (0,1,0,0,...,0)$$

Umnožak jediničnih vektora je 0 kada su $j_i \neq j_k$

Nul vektori su vektori čije su sve komponente jednake nuli.

Vektori predstavljaju samo specifični oblik matrica koje imaju ili jedan red ili jednu kolonu (stubac). Na primjer retčani vektor sa tri elementa je, ustvari, matrica 1x3, dok je stupčani vektor sa tri elementa matrica 3x1.

4.5.2 Matrice

Matrica je skup brojeva poredanih u *m* redova i *n* kolona (Fulgosi, 1988). Matricu opisujemo i kao mxn matricu (na primjer matrica 4x5 znači da matrica ima 4 reda i 5 kolona). Svaki uneseni rezultat zovemo elementom matrice, označavamo ga malim slovom sa dva broja u supskriptu pri čemu je prvi broj uvijek broj reda, a drugi broj uvijek broj kolone. Na primjer, c_{23} je element c u matrici C koji se nalazi u 2. redu i 3. koloni. Matrice se označavaju velikim slovima (X-matrica bruto rezultata u testu, C – matrica kovarijanci i sl.). Bruto rezultate prikazujemo u matrici bruto rezultata u kojoj se u redovima nalaze ispitanici, a u kolonama rezultati na zadacima ili dijelovima ispitivanja. Kovarijančna matrica ili matrica varijanci i kovarijanci je matrica koja kao elemente sadrži vrijednosti varijanci i kovarijanci između dva seta rezultata:

Tabela 4-3 Primjer kovarijančne matrice

	Varijabla 1	**Varijabla 2**	**Varijabla 3**	**Varijabla 4**
Varijabla 1	$\mathbf{V_1}$	c_{12}	c_{13}	c_{14}
Varijabla 2	c_{21}	$\mathbf{V_2}$	c_{23}	c_{24}
Varijabla 3	c_{31}	c_{32}	$\mathbf{V_3}$	c_{34}
Varijabla 4	c_{41}	c_{42}	c_{43}	$\mathbf{V_4}$

Na primjeru kovarijančne matrice naučit ćemo nekoliko pojmova:

1. Kvadratna matrica – matrica je kvadratna kada ima jednak broj elemenata u redu i koloni
2. Simetrična matrica – matrica je simetrična kada su elementi $e_{ij} = e_{ji}$
3. Glavna dijagonala je dijagonala koju nalazimo u kvadratnoj matrici i spaja elemente e_{ij} pri čemu je i = j (npr. $e_{11}, e_{22}, e_{33}, \ldots$)
4. Kvadratna simetrična matrica dobije se svaki put kada ukrstimo elemente sa samim sobom. U ovakvim matricama je $e_{ij} = e_{ji}$, odnosno elementi gornjeg i donjeg trokuta su jednaki.

U matrici u kojoj imamo ukrštene varijable sa samim sobom, ukupan broj različitih elemenata iznosi:

4-8

$$BROJ\ ELEMENATA\ SIMETRIČNE\ MATRICE = \frac{k(k-1)}{2}$$

pri čemu je k – broj varijabli.

U slučaju kada imamo matricu u kojoj su u redovima i kolonama druge varijable, broj različitih elemenata jednak je ukupnom broju elemenata matrice. U ovakvim matricama se u glavnoj dijagonali ne nalaze varijance nego kovarijance. Primjer takve matrice je:

Tabela 4-4 Primjer kovarijančne matrice varijabli 1,2 i 3 s varijablama X,Y I Z

	Varijabla X	**Varijabla Y**	**Varijabla Z**
Varijabla 1	C_{1x}	C_{1y}	C_{1z}
Varijabla 2	C_{2x}	C_{2y}	C_{2z}
Varijabla 3	C_{3x}	C_{3y}	C_{3z}

Još jedan važan pojam za matrice je **transpon matrice**. Transpon matrice je matrica izvedena iz druge matrice, pri čemu su elementi redova i stubaca zamijenili mjesta.

4.5.2.1 Operacije s matricama

4.5.2.1.1 Sabiranje i oduzimanje matrica

Dvije matrice možemo sabirati onda kada su istog oblika. Kao sumu dvije matrice dobijemo matricu čiji su elementi sume korespondentnih elemenata matrica sabiraka.

4-9

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{31} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \\ a_{31}+b_{31} & a_{32}+b_{32} \end{bmatrix}$$

Na primjer:

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 3 & 3 \end{bmatrix} + \begin{bmatrix} 4 & 3 \\ 3 & 4 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} 1+4 & 2+3 \\ 0+3 & 0+4 \\ 3+5 & 3+5 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 3 & 4 \\ 8 & 8 \end{bmatrix}$$

Oduzimanje matrica radimo po istom principu:

4-10

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{31} \end{bmatrix} = \begin{bmatrix} a_{11}-b_{11} & a_{12}-b_{12} \\ a_{21}-b_{21} & a_{22}-b_{22} \\ a_{31}-b_{31} & a_{32}-b_{32} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 3 & 3 \end{bmatrix} - \begin{bmatrix} 4 & 3 \\ 3 & 4 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} 1-4 & 2-3 \\ 0-3 & 0-4 \\ 3-5 & 3-5 \end{bmatrix} = \begin{bmatrix} -3 & -1 \\ -3 & -4 \\ -2 & -2 \end{bmatrix}$$

4.5.2.1.2 Množenje matrica

Kada množimo matricu nekim brojem, to radimo na isti način kao kod vektora. Rezultat množenja bit će matrica čiji će svaki element biti produkt elementa početne matrice s brojem (skalarom).

Dvije matrice možemo množiti samo kada je **broj kolona (stubaca) prve matrice jednak broju redova druge matrice.** Produktna matrica ima isti broj redova kao prva matrica i isti broj kolona (stubaca) kao druga matrica. Matrice se množe tako da se j-ti red prve matrice (shvaćen kao retčani vektor) pomnoži sa k-tim stupcem druge matrice (shvaćenim kao stupčani vektor). Na primjer, ako množimo matricu A reda 3x2 s matricom B reda 2x3, dobit ćemo produktnu matricu C reda 3x3, a ako množimo matricu B s matricom A, dobit ćemo matricu D reda 2x2. Dakle, za matrice ne važi pravilo komutacije u množenju.

Zamislimo da imamo dvije matrice A i B.

A = $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ i B = $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$

Njihov umnožak će biti jednak:

4-11

AxB = $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} x \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} =$

= $\begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{bmatrix}$

U slučaju da množimo BxA, dobijemo matricu:

4-12

BxA = $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$ x $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ = $\begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31} & b_{11}a_{12} + b_{12}a_{22} + b_{13}a_{32} \\ b_{21}a_{11} + b_{22}a_{21} + b_{32}a_{31} & b_{21}a_{12} + b_{22}a_{22} + b_{23}a_{32} \end{bmatrix}$

Primjer:

A = $\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$ B = $\begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$

A x B = $\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} x \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$ = $\begin{bmatrix} 1*7+4*10 & 1*8+4*11 & 1*9+4*12 \\ 2*7+5*10 & 2*8+5*11 & 2*9+5*12 \\ 3*7+6*10 & 3*8+6*11 & 3*9+6*12 \end{bmatrix} =$

$= \begin{bmatrix} 47 & 52 & 57 \\ 64 & 71 & 78 \\ 81 & 90 & 99 \end{bmatrix}$

4.5.2.2 Determinanta, minori i kofaktori

Svaka kvadratna matrica determinira jedan broj – determinantu matrice. Determinanta se piše između dvije ravne crte (kao što pišemo apsolutnu vrijednost).

Ako zadanoj matrici oduzmemo jedan ili više redova i stubaca, dobivamo minore matrice. Uvijek se oduzima jednak red redova i stubaca. Ako se iz zadane matrice oduzme red j i stubac k, koji se sijeku u elementu a_{jk}, preostala determinanta se zove minor elementa a_{jk}.

Elementima determinante se prilikom izvođenja operacija dodjeljuju predznaci tako da dva susjedna elementa nemaju isti predznak. Ako minoru elementa a_{jk} pridružimo pozicioni znak elementa, dobivamo **kofaktor elementa a_{jk}.**

Kako određujemo kofaktor elementa? Pretpostavimo da želimo odrediti kofaktor elementa a_{11}. To ćemo uraditi tako da iz matrice odstranimo elemente prvog reda i prvog stupca. Elementi koji ostaju tvore minor elementa a_{11} . Pošto je ovo prvi element on će kao kofaktor zadržati svoj predznak. Element a_{21} je njemu susjedni element, tako da prilikom određivanja znaka za njegov kofaktor – mijenjamo mu predznak .

Pogledajmo na primjeru. Uzmimo matricu A s elementima $\begin{bmatrix} 4 & 5 & 3 \\ 6 & 7 & 2 \\ 8 & 9 & 1 \end{bmatrix}$

Minor elementa a_{11} dobijemo tako da odstranimo elemente prvog reda i prvog stupca. Dobit ćemo minor $\begin{bmatrix} 7 & 2 \\ 9 & 1 \end{bmatrix}$. Ovo je prvi element i on će kao kofaktor zadržati svoj predznak. Minor elementa a_{12} dobijemo tako da iz matrice oduzmemo prvi red i drugi stubac te nam ostane $\begin{bmatrix} 6 & 2 \\ 8 & 1 \end{bmatrix}$. Ovo je drugi element i njegov kofaktor bi promijenio predznak (ako bi bio pozitivan, postat će negativan broj i obrnuto).

Determinantu matrice 2x2 određujemo tako da pomnožimo suprotne elemente i oduzmemo ih:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb$$

Ono što smo uradili je odredili minor elemenata u koloni 1 (elementi a i c). U matrici 2x2 minor elemenata a je element d, a minor elementa c je element b. Svaki element pomnožimo s njegovim minorom i oduzmemo ih zbog toga što kofaktor elementa c mijenja predznak.

Za matrice većeg reda determinantu određujemo preko kofaktora:

4-13

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}$$

U slučaju matrica višeg reda, ovaj račun postaje složeniji. Tada za određivanje determinanti matrice koristimo Excel ili neki od statističkih paketa koji imaju opcije za izračunavanje determinante.

Pogledajmo na primjeru kako bismo izračunali determinantu matrice A:

$$\begin{vmatrix} 4 & 5 & 8 \\ 6 & 7 & 2 \\ 8 & 9 & 1 \end{vmatrix} = 4 * \begin{vmatrix} 7 & 2 \\ 9 & 1 \end{vmatrix} \text{ - } 6 * \begin{vmatrix} 5 & 8 \\ 9 & 1 \end{vmatrix} + a_3 \begin{vmatrix} 5 & 8 \\ 7 & 2 \end{vmatrix} =$$

$$= 4 * (7 - 18) - 6 * (5 - 72) + 8 * (10 - 56) = -10$$

4.5.2.3 Inverz matrice

Inverz matrice je matrica koja, pomnožena sa zadanom matricom, daje matricu identiteta, odnosno matricu u kojoj imamo jedinice u glavnoj dijagonali, a sve ostale vrijednosti u trokutovima su 0. Inverz matrice određujemo tako da:

1. Od matrice A formiramo matricu M čiji se elementi vrijednosti minora elemenata zadane matrice A
2. Odredimo predznake tako da dobijemo matricu kofaktora
3. Transponiramo matricu kofaktora K i dobijemo K’
4. K’ se zove adjunkta matrice A
5. Svaki element podijelimo s vrijednošću determinante A. Dobivena matrica je inverzna matrica matrice A i označava se sa A^{-1}

Odredimo inverz matrice A iz prethodnog primjera. Prvo nam trebaju vrijednosti svih minora kako bismo dobili matricu minora M.

$$A = \begin{bmatrix} 4 & 5 & 8 \\ 6 & 7 & 2 \\ 8 & 9 & 1 \end{bmatrix} \quad \text{M} = \begin{bmatrix} 7*1-2*9 & 6*2-8*1 & 6*9-8*7 \\ 5*8-9*1 & 4*8-8*1 & 4*9-8*5 \\ 5*8-7*2 & 4*2-8*6 & 4*7-6*5 \end{bmatrix} = \begin{bmatrix} -11 & 4 & -2 \\ 31 & 24 & -4 \\ 26 & -40 & -2 \end{bmatrix}$$

Elementima matrice minora pridružimo predznake kako bismo dobili matricu kofaktora. Predznaci se pridružuju prema sljedećem rasporedu (pridružuju u smislu da ukoliko je znak pozitivan, onda to znači da broj zadržava svoj originalni predznak, a ako je negativan, to znači da će broj promijeniti predznak):

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$$

Naša matrica minora će sada postati matrica kofaktora K: $\begin{bmatrix} -11 & -4 & -2 \\ -31 & 24 & 4 \\ 26 & 40 & -2 \end{bmatrix}$

Matricu kofaktora transponiramo, odnosno zamijenimo mjesta redovima i stupcima kako bismo dobili transpon matrice K – K' ili adjunktu matrice K:

$$\text{K'} = \begin{bmatrix} -11 & -31 & 26 \\ -4 & 24 & 40 \\ -2 & 4 & -2 \end{bmatrix}$$

Svaki element adjunkte podijelimo s determinantom matrice A i dobijemo inverznu matricu matrice A koju označavamo sa A^{-1}:

$$A^{-1} = \begin{bmatrix} -11 & -31 & 26 \\ -4 & 24 & 40 \\ -2 & 4 & -2 \end{bmatrix} \text{/(-10)} = \begin{bmatrix} 1{,}1 & 3{,}1 & -2{,}6 \\ 0{,}4 & -2{,}4 & -4{,}0 \\ 0{,}2 & -0{,}4 & 0{,}2 \end{bmatrix}$$

5 URADAK NA TESTU

Zašto služe transformacije rezultata? Što se dešava s parametrima distribucije kada se rezultati na zadacima ili testovima transformiraju u neku novu skalu? Kako možemo formirati ukupni uradak na testu? Kakav je odnos parametara testa prema parametrima njegovih elemenata?

5.1 TRANSFORMACIJE REZULTATA

Transformacije rezultata vama nisu nepoznata stvar. Na Statistici ste, među prvim lekcijama, učili kako rezultate ispitanika nekoliko testova ne možemo uspoređivati prije nego ih sve transformiramo na zajedničku skalu. Dakle, da bismo mogli uspoređivati rezultate, moramo ih pripremiti tako što ćemo ih sve transformirati na jedinstvenu skalu. Sada ćemo malo proširiti priču o transformacijama.

Nakon što se završi testiranje i ispitanici odlože svoje testove, mi dobivamo sirove ili bruto rezultate. Međutim, to ne moraju biti rezultati koje ćemo koristiti direktno za formiranje ukupnog uratka ili za daljnje analize. Na primjer, u testu je bio neki jako težak zadatak postavljen kao binarna varijabla. Mi za taj zadatak unaprijed možemo odlučiti kako ćemo umjesto jednog boda za tačan uradak dati 3 boda, odnosno sirove rezultate ispitanika na tom zadatku pomnožiti sa 3. Tako ćemo umjesto 0 i 1, dobiti 0 i 3 kao moguće rezultate.

Transformacije rezultata predstavljaju aplikaciju matematičke modifikacije na vrijednosti varijable. Postoji niz mogućih transformacija (baš koliko ima i funkcija u matematici) od jednostavnog dodavanja konstante, preko korjenovanja ili kvadriranja rezultata, pa do logaritamskih transformacija, čak i trigonometrijskih transformacija. Međutim, ono što nas zanima u ovom trenutku su transformacije važne za početnike u psihometriji. Koje transformacije imamo na raspolaganju i koje su korisne svakom psihologu? Transformacije, prije svega, dijelimo prema tome da li one kao posljedicu imaju promjenu distribucije rezultata ili distribucija rezultata ostaje istog oblika. U tom smislu imamo linearne i nelinearne transformacije. Osim toga, linearne transformacije mogu podrazumijevati samo jednu jednostavnu aritmetičku operaciju, a mogu podrazumijevati i više aritmetičkih operacija u formuli, te ih onda dijelimo na jednostavne linearne transformacije i složene linearne transformacije.

Zamislimo da je osam ispitanika radilo jedan zadatak i da je zadatak bio binarnog tipa. U tabeli 5-1 vidimo što se desi s rezultatima ukoliko nad njima provedemo osnovne računske operacije.

Jednostavne linearne transformacije uključuju sabiranje s konstantom, oduzimanje konstante, množenje s konstantom i dijeljenje s konstantom.

Tabela 5-1 Jednostavne linearne transformacije rezultata

Ispitanik	Sirovi rezultat	Sabiranje s konstantom 1 Y = X+1	Oduzimanje konstante 1 Y = X-1	Množenje s konstantom 3 Y = 3X	Dijeljenje s konstantom 3 Y=X/3
1	**1**	**2**	**0**	**3**	**1/3**
2	**0**	**1**	**-1**	**0**	**0**
3	**1**	**2**	**0**	**3**	**1/3**
4	**1**	**2**	**0**	**3**	**1/3**
5	**0**	**1**	**-1**	**0**	**0**
6	**0**	**1**	**-1**	**0**	**0**
7	**1**	**2**	**0**	**3**	**1/3**
8	**0**	**1**	**-1**	**0**	**0**

5.1.1 Jednostavne linearne transformacije

5.1.1.1 Sabiranje s konstantom

Sabiranje s konstantom definirano je funkcijom X' = X + A, pri čemu je A – konstanta i različita je od 0. Kada u praksi koristimo ovu transformaciju? Uglavnom onda kada želimo da rezultate pomjerimo prema većim vrijednostima (na primjer, kada ne želimo da bilo ko dobije 0 jer je „ljepše“ dobiti 2 iako ovo pomjeranje ne znači da je osoba odjednom postala „bolja“).

Ukoliko dodamo konstantu na rezultate, aritmetička sredina će iznositi:

5-1

$$M' = \frac{\sum X_i'}{N} = \frac{\sum (X+A)_i}{N} = \frac{\sum X_i}{N} + \frac{NA}{N} = M + A$$

Vidimo da će se, u slučaju dodavanja konstante, aritmetička sredina povećati za vrijednosti konstante.

Dodavanje konstante neće utjecati na vrijednost varijance:

5-2

$$V' = \frac{\sum (X' - M')_i^2}{N} = \frac{\sum [(X + A) - (M + A)]_i^2}{N} = \frac{\sum [X + A - M - A]_i^2}{N} = \frac{\sum [X - M]_i^2}{N} = V$$

Možemo zaključiti: ukoliko rezultatima dodajemo konstantu A, aritmetička sredina rezultata se poveća za visinu konstante, dok varijanca ostaje ista.

5.1.1.2 Oduzimanje konstante

Oduzimanje konstante definirano je funkcijom X' = X – A. Utjecaj oduzimanja konstante na parametre distribucije je isti kao i kod sabiranja. Ukoliko od rezultata oduzmemo konstantu, aritmetička sredina će se smanjiti za vrijednosti konstante, dok će varijanca ostati ista.

5-3

$$M' = \frac{\sum X_i'}{N} = \frac{\sum (X-A)_i}{N} = \frac{\sum X_i}{N} - \frac{NA}{N} = M - A$$

5-4

$$V' = \frac{\sum (X' - M')_i^2}{N} = \frac{\sum [(X-A) - (M-A)]_i^2}{N} = \frac{\sum [X - A - M + A]_i^2}{N} = \frac{\sum [X - M]_i^2}{N} = V$$

5.1.1.3 Množenje s konstantom

Množenje s konstantom je dosta često u psihologiji. Već smo spomenuli da množenjem pojedinim zadacima dajemo veću važnost u formiranju ukupnog rezultata. Množenje s konstantom predstavljeno je funkcijom X' =X*A. U ovom slučaju doći će do promjene i aritmetičke sredine i varijance (poenta množenja i jeste povećati varijancu).

5-5

$$M' = \frac{\sum X_i'}{N} = \frac{\sum (XA)_i}{N} = \frac{A \sum X_i}{N} = M * A$$

5-6

$$V' = \frac{\sum (X' - M')_i^2}{N} = \frac{\sum [XA - MA]_i^2}{N} = \frac{\sum [A(X-M)]_i^2}{N} = \frac{A^2 \sum [X - M]_i^2}{N} = A^{2*}V$$

5-7

$$ST.DEV = \sqrt{V} = \sqrt{A^{2*}V} = A * ST.DEV$$

Vidimo da će se, ukoliko rezultate pomnožimo s konstantom, aritmetička sredina povećati za umnožak originalne aritmetičke sredine s konstantom, dok će se varijanca povećati za umnožak kvadrata konstante.

5.1.1.4 Dijeljenje s konstantom

Dijeljenje s konstantom izraženo je funkcijom X' = X / A. U slučaju kada rezultate podijelimo s konstantom, aritmetička sredina novih rezultata bit će količnik između originalne aritmetičke sredine i konstante, dok će varijanca biti količnik originalne varijance i kvadrata konstante.

5-8

$$M' = \frac{\sum X_i'}{N} = \frac{\sum\left(\frac{X}{A}\right)_i}{N} = \frac{\frac{\sum X_i}{A}}{N} = \frac{\sum X_i}{NA} = \frac{M}{A}$$

5-9

$$V' = \frac{\sum(X' - M')_i^2}{N} = \frac{\sum\left[\frac{X}{A} - \frac{M}{A}\right]_i^2}{N} = \frac{\sum\left[\frac{(X - M)}{A}\right]_i^2}{N} = \frac{\sum(X - M)_i^2}{NA^2} = \frac{V}{A^2}$$

5-10

$$ST.DEV = \sqrt{V} = \sqrt{\frac{V}{A^2}} = \frac{ST.DEV}{A}$$

5.1.2 Složene (višestruke) linearne transformacije

5.1.2.1 Transformacija u z-vrijednosti

U praksi, kada ispitanik radi psihološki test, on ili ona dobiju sirove rezultate. Ti rezultati mogu biti izraženi na različite načine. To mogu biti bodovi (na primjer, Marko je osvojio 50 bodova na testu) ili procenti (Marko je osvojio 80 % na testu). Međutim, sam sirovi rezultat nam ništa ne znači. Na primjer, da bismo znali da li je Marko uradio test dobro ili loše, mi moramo poznavati distribuciju rezultata grupe ispitanika koja je slična Marku i koja je također radila ovaj test. Dakle, potrebno nam je da znamo aritmetičku sredinu i standardnu devijaciju testa. Kada znamo ova dva parametra, onda rezultate možemo prevesti u z-vrijednosti. Transformacija sirovih rezultata u z-vrijednosti je osnovna složena linearna transformacija u psihologiji. Z-vrijednosti nazivamo standardiziranim rezultatima pri čemu oni predstavljaju udaljenost rezultata koji je ispitanik postigao od aritmetičke sredine izražene u terminima standardne devijacije (Furr, 2013, str. 58). Prisjetit ćemo se formule za određivanje z-vrijednosti:

5-11

$$z_X = \frac{X - M}{s}$$

z-skala ima aritmetičku sredinu 0 i standardnu devijaciju 1 što možemo dokazati na sljedeći način:

$$M = \frac{\sum z_i}{N} = \frac{\sum\left(\frac{X - M}{s}\right)_i}{N} = \frac{s^{-1}(\sum X - NM)}{N} = s^{-1}\left[\frac{\sum X}{N} - \frac{NM}{N}\right] = s^{-1}(M - M) = s^{-1} * 0 = 0$$

$$V = \frac{\sum(X - M)_i^2}{N} = \frac{\sum(z - 0)_i^2}{N} = \frac{\sum(z)_i^2}{N} = \frac{\sum\left(\frac{X - M}{s}\right)_i^2}{N} = \frac{\sum(\boldsymbol{X} - \boldsymbol{M})_{\boldsymbol{i}}^{\boldsymbol{2}}}{s^2\boldsymbol{N}} = \frac{V}{V} = 1$$

5.1.2.2 Višestruke linearne transformacije na skale sa zadanim parametrima

Kada rezultate transformiramo u z-skalu, onda smo u mogućnosti da lako ustanovimo položaj ispitanika i poredimo rezultate testova. Na primjer, kada psihologu kažemo da je rezultat ispitanika z=1, on odmah zna da je taj ispitanik udaljen jednu standardnu devijaciju od aritmetičke sredine, te da je bolji od oko 84 % drugih ispitanika. Međutim, z-vrijednosti nisu razumljive za nepsihologe tako da rezultate transformiramo u druge standardne skale koje imaju zadane aritmetičke sredine i standardne devijacije. Primjer takve standardne skale je skala devijacijskog kvocijenta inteligencije (DQI) na kojoj izražavamo rezultate ispitanika na testovima inteligencije. Njena aritmetička sredina je 100 i standardna devijacija 15.

Sve višestruke linearne transformacije rade se prema formuli:

5-12

$$X' = az_x + b$$

pri čemu su a i b konstante, gdje **a** predstavlja standardnu devijaciju nove skale, a **b** predstavlja aritmetičku sredinu nove skale.

Pretpostavimo da je ispitanik na testu inteligencije, čija je aritmetička sredina bila 55, a standardna devijacija 8, postigao rezultat 65.

Njegov z-rezultat iznosio bi:

$$z_X = \frac{X - M}{s} = \frac{65 - 55}{8} = 1{,}25$$

Kada bismo željeli izraziti njegov rezultat na skali **DQI** (devijacijskog kvocijenta inteligencije) koja ima aritmetičku sredinu 100 i standardnu devijaciju 15, njegov rezultat bi iznosio:

$$X' = az_x + b = 15 * 1{,}25 + 100 = 18{,}25 + 100 = 118{,}25 \ \approx 118$$

Ukoliko želimo "preskočiti" računanje z-vrijednosti i sirovi rezultat direktno transformirati u novu skalu sa zadanim parametrima, dovoljno je samo da u formulu 5-12 ubacimo formulu za z-vrijednost i dobit ćemo opću formulu za linearnu transformaciju rezultata s jedne skale na drugu:

5-13

$$X' = \frac{s'}{s}\,(X - M) + M'$$

Pri čemu je:

s' – standardna devijacija nove skale

s – standardna devijacije stare skale

M' – aritmetička sredina nove skale

M – aritmetička sredina stare skale

Osim DQI skale postoje i druge standardne skale koje se koriste. Jedna od njih je **T-skala** čija je aritmetička sredina 50, a standardna devijacija 10.

Rezultati transformirani prema višestrukoj linearnoj transformaciji zadržat će oblik distribucije, što je jedna od prednosti ovih transformacija. Međutim, linearne transformacije vezane su za aritmetičku sredinu i standardnu devijaciju te se stoga mogu provoditi samo u slučaju kada rezultati ne odstupaju značajno od normalne distribucije. Ukoliko na testiranju dobijemo rezultate koji značajno odstupaju od normalne distribucije, onda ne možemo jednostavno izvršiti linearnu transformaciju nego rezultate prvo treba "normalizirati". Koju funkciju ćemo koristiti kako bi distribuciju rezultata normalizirali, zavisi od oblika distribucije. Prema Tabachnik i Fidell (2007,str. 87), transformacijske funkcije koje trebamo poznavati za različite vrste asimetrija su:

Tabela 5-2 Vrste transformacija radi normalizacije rezultata

Vrsta asimetrije	**Transformacijska funkcija**
Umjerena pozitivna asimetrija	$X' = \sqrt{X}$
Umjerena negativna asimetrija	$X' = \sqrt{k - X}$*
Značajna pozitivna asimetrija	$X' = logX$
Značajna negativna asimetrija	$X' = \log(\mathrm{k} - \mathrm{X})$
Krivulja u obliku obrnutog slova J	$X' = \frac{1}{X}$
Krivulje u obliku slova J	$X' = \frac{1}{k - X}$

***Konstanta k određuje se tako da se pronađe najviši rezultat u distribuciji i onda odredi visina k tako da k bude veće od najvećeg rezultata (obično za 1). Onda se kreira nova varijabla tako da se od konstante k oduzmu rezultati i negativna asimetrija pretvori u pozitivnu asimetriju. Vodite računa o smjeru povezanosti kada interpretirate ovako reflektiranu varijablu. Ako ste dobili pozitivnu korelaciju, onda to znači da u stvarnosti imate negativnu korelaciju jer su vam u transformiranoj skali najviši rezultati postali najniži i obrnuto.*

Ukoliko se pojavi potreba za transformacijom radi normalizacije funkcije, potrebno je da probate nekoliko transformacija, ali ni to ne znači da ćete uspjeti normalizirati distribuciju. Ako transformacijama ne uspijete normalizirati distribuciju, probajte kategorizirati varijablu. U svakom slučaju, trebate pročitati više od ovoga što je dato u našem priručniku da biste bolje razumjeli transformacije radi normalizacije.

5.1.3 Nelinearne transformacije

Dosta čest način prezentiranja i interpretacije rezultata je u percentilnoj skali, odnosno u percentilnim rangovima koji daju informaciju o tome u kojem percentilu se nalazi ispitanikov rezultat. Percentilna skala je vrlo popularna jer većina ljudi razumije kada kažemo: „On je među prvih 10 % učenika“. Ukoliko imamo sve sirove rezultate ispitanika, onda se percentilna transformacija radi tako da odredimo ukupan broj ispitanika koji su ostvarili manji rezultat od našeg ispitanika, tome dodamo polovinu broja ispitanika koji su ostvarili isti rezultat kao naš ispitanik, taj broj podijelimo s ukupnim brojem ispitanika u testiranju i količnik pomnožimo sa 100.

5-14

$$PR = \frac{f_{-} + 0{,}5 * f_{=}}{N} * 100$$

Na primjer, ukoliko je u istraživanju učestvovalo 80 ispitanika, Damir je ostvario rezultat koji je bolji od njih 60, njegov rezultat ostvarila su još 3 ispitanika, njegov percentilni rang bit će:

$$PR = \frac{f_{-} + 0{,}5 * f_{=}}{N} * 100 = \frac{60 + 0{,}5 * 4}{80} * 100 = \frac{62}{80} * 100 = 77{,}5 \approx 76$$

Dakle, Damirov percentilni rang je 76-i.

Drugi način da izračunamo percentilni rang je da koristimo z-vrijednost. Mi znamo za svaku z-vrijednost iz statističkih tablica koliko ispitanika je slabije, odnosno bolje od ispitanika za kojeg računamo percentilni rang. Međutim, ovo podrazumijeva da su rezultati normalno distribuirani te ovaj način određivanja percentilnog ranga možemo koristiti samo ako distribucija ne odstupa značajno od normalne.

Najveća prednost percentilnih transformacija je što ne zahtijevaju normalnu distribuciju. Ukoliko rezultate tranformiramo u percentilne rangove, tada ćemo i normalnu distribuciju pretvoriti više u pravougaonu. Međutim, moramo znati da percentilnom transformacijom gubimo podatak o veličini razlika među pojedinim ispitanicima. Na primjer, Damir je postigao rezultat 60, Sanja je postigla sljedeći najbolji rezultat 61, a Dijana je postigla sljedeći najbolji rezultat 70. Njihovi percentilni rangovi će biti: Damir – 76, Sanja – 77, Dijana -78. Vidimo da imamo njihov poredak na ordinalnoj skali, ali više ne možemo govoriti o veličini razlika među njihovim rezultatima.

5.2 FORMIRANJE UKUPNOG URATKA NA TESTU – LINEARNE KOMBINACIJE

U psihologiji, određivanje prisustva ili intenziteta neke karakteristike obično se radi tako da se ispitanicima postavlja više pitanja ili zadataka. Na kraju se odgovori na pitanja kombiniraju i formira se ukupni uradak. Taj ukupni uradak nazivamo i kompozit, a formira se najčešće kao linearna kombinacija.

Prije nego samo saberemo bodove, trebamo odlučiti da li bi jednostavno sabiranje bilo i najbolji način za formiranje ukupnog uratka. Možemo reći da postoje dvije osnovne dimenzije na osnovu kojih dijelimo linearne kombinacije:

1. Da li koristimo sabiranje ili oduzimanje za formiranje uratka? Prema ovom kriteriju linearne kombinacije dijelimo na aditivne i suptraktivne.
2. Da li samo sabiremo rezultate ili nekim rezultatima dajemo više važnosti tako što ih množimo s nekom konstantom? Prema ovom kriteriju linearne kombinacije dijelimo na jednostavne i diferencijalno ponderirane.

Kada ove dvije dimenzije ukrstimo, dobijemo četiri mogućnosti za formiranje linearnih kombinacija:

1. Jednostavna aditivna linearna kombinacija

2. Jednostavna suptraktivna linearna kombinacija
3. Diferencijalno ponderirana aditivna linearna kombinacija
4. Diferencijalno ponderirana suptraktivna linearna kombinacija

5.2.1 Aritmetičke sredine i varijance linearnih kombinacija

5.2.1.1 Aritmetička sredina i varijanca jednostavne aditivne linearne kombinacije

Jednostavna aditivna linearna kombinacija opisana je formulom:

5-15

$$Y = X_1 + X_2 + X_3 + ... + X_K$$

Njena aritmetička sredina je jednaka:

$$M_Y = \frac{\sum(X_1 + X_2 + X_3 + \cdots + X_K)}{N}$$

$$M_Y = \frac{\sum X_1}{N} + \frac{\sum X_2}{N} + \frac{\sum X_3}{N} + \cdots + \frac{\sum X_K}{N}$$

5-16

$$M_Y = M_1 + M_2 + M_3 + \cdots + M_K$$

Aritmetička sredina jednostavne aditivne linearne kombinacije jednaka je zbiru aritmetičkih sredina pojedinih elemenata. Ukoliko su varijable binarne i kodirane sa 0 i 1, onda će aritmetička sredina linearne kombinacije biti jednaka zbiru proporcija odgovora kodiranih sa 1. Ukoliko su varijable standardizirane na z-vrijednosti, onda će njihove aritmetičke sredine iznositi 0, pa će i aritmetička sredina aditivne linearne kombinacije biti jednaka 0.

Kako bismo vidjeli čemu je jednaka varijanca linearne kombinacije, zamislimo jednostavni primjer u kojem je jednostavna aditivna kombinacije formirana od samo dva zadataka.

$$Y = X_1 + X_2$$

Varijanca jednostavne aditivne linearne kombinacije iznosi:

$$V_Y = \frac{\sum(Y - M_Y)_i^2}{N}$$

$$V_Y = \frac{\sum[(X_1 + X_2) - (M_1 + M_2)]_i^2}{N}$$

$$V_Y = \frac{\sum[X_1 + X_2 + -M_1 - M_2]_i^2}{N}$$

$$V_Y = \frac{\sum[(X_1 - M_1) + (X_2 - M_2)]_i^2}{N} = \frac{\sum[d_1 + d_2]_i^2}{N} = \frac{\sum(d_1^2 + 2d_1d_2 + d_2^2)_i}{N}$$

$$V_Y = \frac{\sum(d_1^2)_i}{N} + \frac{2\sum(d_1d_2)_i}{N} + \frac{\sum(d_2^2)_i}{N}$$

Ranije smo vidjeli da su $\frac{\sum(d_1^2)_i}{N}$ i $\frac{\sum(d_2^2)_i}{N}$ ništa drugo do varijance prvog i drugog dijela ove linarne kombinacije, a da je $\frac{2\sum(d_1d_2)_i}{N}$ kovarijanca prvog i drugog dijela. To znači da je varijanca jednostavne linearne kombinacije od dva člana jednaka:

5-17

$$V_Y = V_1 + V_2 + 2COV_{12}$$

Opća formula za varijancu jednostavne aditivne linearne kombinacije bi onda bila:

5-18

$$V_Y = \sum V_i + 2\sum cov_{ij}$$

Dakle, varijanca jednostavne linearne kombinacije jednaka je sumi svih varijanci dijelova linearne kombinacije i dvostrukoj sumi kovarijanci koju imaju svi parovi varijabli. Ukoliko pogledamo kovarijančnu matricu (tabela 4-3), vidjet ćemo da je varijanca linearne kombinacije jednaka sumi svih elemenata kovarijančne matrice elemenata linearne kombinacije.

5.2.1.2 Aritmetička sredina i varijanca jednostavne suptraktivne linearne kombinacije

Ponekad ukupni uradak formiramo tako da od rezultata jednog dijela testa oduzmemo rezultat na drugom dijelu testa. Na primjer, ukoliko mjerimo vrijeme potrebno za savladavanja labirinta može nas zanimati kolika je promjena u sekundama između prvog i drugog prolaska. U takvim slučajevima linearna kombinacija definirana je kao:

5-19

$$Y = X_1 - X_2$$

Aritmetička sredina jednostavne suptraktivne kombinacije određena je kao:

5-20

$$M_Y = M_1 - M_2$$

Varijanca je određena kao:

$$V_Y = \frac{\sum(Y - M_Y)_i^2}{N}$$

$$V_Y = \frac{\sum[(X_1 - X_2) - (M_1 - M_2)]_i^2}{N}$$

$$V_Y = \frac{\sum[X_1 - X_2 + -M_1 + M_2]_i^2}{N}$$

$$V_Y = \frac{\sum[(X_1 - M_1) - (X_2 + M_2)]_i^2}{N} = \frac{\sum[d_1 - d_2]_i^2}{N} = \frac{\sum(d_1^2 - 2d_1d_2 + d_2^2)_i}{N}$$

$$V_Y = \frac{\sum(d_1^2)_i}{N} - \frac{2\sum(d_1d_2)_i}{N} + \frac{\sum(d_2^2)_i}{N}$$

5-21

$$V_Y = V_1 + V_2 - 2COV_{12}$$

5.2.1.3 Aritmetička sredina i varijanca diferencijalno ponderirane aditivne kombinacije

Diferencijalno ponderirane kombinacije su dosta češte u formiranju ukupnog uratka na testovima, posebno na testovima znanja jer zadaci na tim testovima nisu jednako teški. Ponderiranje, odnosno omogućavanje različitog bodovnog učešća zadataka u formiranju ukupnog uratka, ima najmanje dvije funkcije. Prva i važnija funkcija u psihometrijskom smislu je omogućavanje boljeg razlikovanja ispitanika prema mjerenoj karakteristici. Druga funkcija, iako u psihometrijskom smislu ne mora biti toliko važna, je pravednost u ocjenjivanju. Zaista ne bi bilo pravedno da dva ispitanika dobiju istu ocjenu kada je prvi uradio više laganih zadataka, a drugi uradio isto toliko, ali jako teških zadataka. Prilikom ponderiranja moramo voditi računa o visini pondera tako da ih dodjeljujemo prema nekom objektivnom kriteriju. To je obično težina zadatka ili njegova važnost za mjerenje karakteristike.

Ukupan uradak kod diferencijalno ponderirane linearne kombinacije je defniran formulom:

5-22 $$Y_{dplk} = W_1X_1 + W_2X_2 + W_3X_3 + ... + W_KX_K$$

Kada su rezultati svake varijable ponderirani, njihove pojedinačne aritmetičke sredine bit će uvećane za umnožak konstante kao što smo vidjeli u formuli 5-6. To podrazumijeva da će i aritmetička sredina linearne kombinacije biti određena kao suma ponderiranih aritmetičkih sredina:

5-23 $$M_{dplk} = W_1M_1 + W_2M_2 + W_3M_3 + ... + W_KM_K$$

Smisao ponderiranja rezultata je povećanje varijance i, samim tim, razlikovanja ispitanika. U slučaju diferencijalno ponderiranih linearnih kombinacija varijanca značajno raste. U sekciji 5.1.1.3 vidjeli smo da je varijanca ponderirane varijable uvećana za umnožak s kvadratom konstante, odnosno standardna devijacija je uvećana za umnožak s konstantom. Kada ovo uvrstimo u formulu za varijancu linearne kombinacije, dobijemo:

$$V_Y = \sum V_i + 2\sum cov_{ij}$$

Znamo da je V'=V*w^2 ; također znamo da je $c_{ij} = r_{ij}s_i s_j$, a kako je s'= s*w; onda dobijemo da je varijanca diferencijalno ponderirane linearne kombinacije jednaka:

5-24

$$V_Y = \sum V_i\, w_i^2 + 2\sum r_{ij}\, s_i s_j w_i w_j = \sum V_i\, w_i^2 + 2\sum cov_{ij}\, w_i w_j$$

Ma kraju priče o formiranju ukupnog rezultata treba napomenuti kako se ukupan uradak može formirati i kao aritmetička sredina. McDonald (1999) je istaknuo da je razuman izbor formiranja ukupnog uratka kod pitanja Likertovog tipa (pitanja sa skalama) prosječni rezultat, odnosno aritmetička sredina svih pitanja u skali:

5-25

$$Y = \frac{\sum X}{m}$$

pri čemu je $\mathrm{m} - \mathrm{broj\ pitanja\ u\ skali}$. Ovaj način određivanja uratka ništa ne mijenja u smislu razlikovanja ispitanika, ali nama olakšava razumijevanje ukupnih uradaka. Kada ispitanik osvoji 21 bod na skali Likertovog tipa od 1 do 5, mi trebamo informaciju o broju pitanja u skali da bismo se orijentirali da li je ovo visok, srednji ili nizak rezultat. Ali, ako odredimo uradak kao aritmetičku sredinu (21 podijelimo s brojem pitanja i dobijemo 3,8), mnogo se lakše orijentiramo o položaju ispitanika.

5.2.2 Korelacije linearnih kombinacija i drugih varijabli

U psihologiji često imamo potrebu da odredimo korelaciju između neke linearne kombinacije i druge varijable. Na primjer, u profesionalnoj selekciji koristimo niz instrumenata kao prediktore da bismo predvidjeli buduće radno ponašanje kao kriterij. Isto tako, imamo potrebu da izračunavamo korelaciju između grupe pitanja kao prediktora i nekog drugog pitanja u testu (na primjer, kako bismo mogli upotpuniti nedostajuće vrijednosti jer je ispitanik preskočio odgovor na to pitanje). Stoga nam je potrebno da znamo kako odrediti korelaciju između:

1. Grupe varijabli kao prediktora i neke vanjske varijable kao kriterija (korelacija linearne kombinacije i jednostavne vanjske varijable)
2. Grupe varijabli kao prediktora i neke članice te grupe varijabli kao kriterija (spuriozna korelacija)
3. Dvije linearne kombinacije

5.2.2.1 *Korelacija linearne kombinacije i neke jednostavne vanjske varijable*

Neka je linearna kombinacija predstavljena formulom:

$$U_i = X_1 + X_2 + X_3$$

Kako bismo sve varijable imali na istoj skali mjerenja, odnosno sveli ih na istu metriku, transformiramo rezultate prediktora i kriterija u z-vrijednosti. Tada će ova linearna kombinacija biti definirana kao:

$$U_{zi} = z_1 + z_2 + z_3$$

Kriterijsku varijablu Y također transformiramo u z metriku i ona će biti z_Y.

Znamo da je korelacija između dvije varijable definirana formulom:

$$r_{xy} = \frac{\sum (d_x d_y)_i}{N s_x s_y}$$

U slučaju linearne kombinacije i kriterijske varijable možemo pisati:

$$r_{Uy} = \frac{\sum (d_U d_y)_i}{N s_U s_y} = \frac{\sum [(U - M_U)(z_y - M_y)]_i}{N s_U s_y}$$

Pošto znamo da su aritmetičke sredine z-vrijednosti jednake 0, te da je standardna devijacija jednaka 1 (a u našem slučaju kriterijska varijabla je standardizirana, dok je ukupan uradak na prediktorima zbir standardiziranih vrijednosti prediktora) ova formula poprima sljedeći oblik:

$$r_{Uy} = \frac{\sum [U * z_y]_i}{N s_U}$$

Vidjeli smo ranije da je standardna devijacija jednostavne linearne kombinacije jednaka zbiru varijanci svakog dijela kombinacije, te dvostrukoj sumi kovarijanci njenih elemenata. Kako su nam ovdje dijelovi linearne kombinacije standardizirani, onda su im i standardne devijacije jednake 1. Osim toga, vidjeli smo da je korelacija standardizirana kovarijanca, odnosno da, kada imamo standardizirane varijable njihova kovarijanca jednaka je njihovoj korelaciji.

U formuli to izgleda ovako:

$$V_Y = \sum V_i + 2 \sum cov_{ij}$$

$$s_Y = \sqrt{\sum V_i + 2 \sum cov_{ij}}$$

V_i svake standardizirane varijable jednaka je 1, te je $\sum s_i = k$, odnosno broju članica (u našem slučaju je to 3). cov_{ij} kod standardiziranih varijabli jedanaka je r_{ij} te ova formula sada glasi:

$$s_Y = \sqrt{k + 2 \sum r_{ij}}$$

Naša formula za korelaciju izgledat će sada ovako:

$$r_{Uy} = \frac{\sum\left[(z_1 + z_2 + z_3) * z_y\right]_i}{N\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

$$r_{Uy} = \frac{\sum\left(z_1 z_y\right)_i + \sum\left(z_2 z_y\right)_i + \sum\left(z_3 z_y\right)_i}{N\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

Prisjetimo se sada što predstavlja količnik sume umnožaka standardiziranih rezultata i broja ispitanika? To nije ništa drugo do korelacija te dvije varijable.

$$\frac{\sum\left(z_1 z_y\right)_i}{N} = r_{1y}$$

$$\frac{\sum\left(z_2 z_y\right)_i}{N} = r_{2y}$$

$$\frac{\sum\left(z_3 z_y\right)_i}{N} = r_{3y}$$

Sada možemo pisati čemu je jednaka korelacija između tri članice linearne kombinacije i neke vanjske jednostavne varijable:

$$r_{Uy} = \frac{r_{1y} + r_{2y} + r_{3y}}{\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

Opći oblik formule za korelaciju između linearne kombinacije i kriterija je:

5-26

$$r_{Uy} = \frac{\sum r_{iy}}{\sqrt{k + 2\sum r_{ij}}}$$

Pri čemu je r_{iy} korelacija svake članice linearne kombinacije s kriterijem, a r_{ij} korelacije među članicama linearne kombinacije.

Da vidimo gdje se nalaze ovi elementi formule u korelacijskoj matrici:

Tabela 5-3 Korelacijska matrica tri prediktora i kriterija

	Zadatak 1	**Zadatak 2**	**Zadatak 3**	**Kriterij**
Zadatak 1	1	r_{12}	r_{13}	r_{1Y}
Zadatak 2	r_{21}	1	r_{23}	r_{2Y}
Zadatak 3	r_{31}	r_{12}	1	r_{3Y}

Vidimo da je ova matrica sastavljena od korelacija među elementima linearne kombinacije i korelacija elemenata linearne kombinacije s kriterijem. Korelacije iz zadnje kolone su one koje idu u brojnik, dok je u nazivniku zbir svih elemenata matrice korelacija i varijanci elemenata linearne kombinacije. Jednostavno, zar ne?

Kada će korelacija između linearne kombinacije i neke kriterijske varijable biti veća? Iz matematike znamo da će količnik biti veći što je brojnik veći, a nazivnik manji. To znači da će korelacije između linearne kombinacije i neke kriterijske varijable biti veće kada su korelacije pojedinačnih članica linearne korelacije s kriterijem veće, a međusobne korelacije članica linearne kombinacije manje.

5.2.2.2 Korelacija između linearne kombinacije i neke članice linearne kombinacije

Ova korelacija zove se spuriozna korelacija. Koristimo je u raznim situacijama – kod određivanja povezanosti pitanja s drugim pitanjima, kod item analize ili kod zamjene nedostajućih vrijednosti, te je stoga potrebno da je upoznamo.

Pretpostavimo da je linearna kombinacija sastavljena od tri elementa:

$$U_i = X_1 + X_2 + X_3$$

te da su rezultati standardizirani u z-vrijednosti:

$$U_{zi} = z_1 + z_2 + z_3$$

Pretpostavimo da želimo izračunati korelaciju elementa z_1 s ostalim elementima linearne kombinacije kao prediktorima. Sada će nam z_1 biti kriterij. Krenut ćemo od formule za korelaciju:

$$r_{Uy} = \frac{\sum(d_U d_1)_i}{N s_U s_1} = \frac{\sum[(U - M_U)(z_1 - M_1)]_i}{N s_U s_1}$$

I ovdje znamo znamo da su aritmetičke sredine z-vrijednosti jednake 0, te da je standardna devijacija standardiziranih varijabli jednaka 1, što znači da će naša s_1 biti jednaka 1. Kada uvrstimo poznate elemente u formulu dobit ćemo:

$$r_{U1} = \frac{\sum[U * z_1]_i}{N s_U}$$

U prethodnoj sekciji vidjeli smo da je standardna devijacija linearne kombinacije jednaka

$$s_Y = \sqrt{k + 2\sum r_{ij}}$$

te da sada ovu formulu možemo pisati na sljedeći način:

$$r_{U1} = \frac{\sum[(z_1 + z_2 + z_3) * z_1]_i}{N\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

$$r_{U1} = \frac{\sum(z_1 z_1)_i + \sum(z_2 z_1)_i + \sum(z_3 z_1)_i}{N\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

Da vidimo čemu su jednaki pojedini dijelovi ove formule:

$$\frac{\sum(z_1 z_1)_i}{N} = V_1 = 1$$

$$\frac{\sum(z_2 z_1)_i}{N} = r_{21}$$

$$\frac{\sum(z_3 z_1)_i}{N} = r_{31}$$

Kada opet poznate elemente uvrstimo u formulu dobijemo:

$$r_{U1} = \frac{1 + r_{21} + r_{31}}{\sqrt{k + 2\sum(r_{12} + r_{13} + r_{23})}}$$

U svom općem obliku, formula za spurioznu korelaciju glasi:

5-27

$$r_{U1} = \frac{1 + \sum r_{i1}}{\sqrt{k + 2\sum r_{ij}}}$$

pri čemu u brojniku imamo sumu korelacija koju svaka članica linearne kombinacije ima s onom članicom za koju računamo spurioznu korelaciju.

Ako pogledamo u matrici korelacija 5-4 gdje se nalaze elementi brojnika i nazivnika, vidimo da su elementi brojnika ustvari elementi reda varijable za koju računamo spurioznu korelaciju, dok su elementi nazivnika suma svih elemenata korelacijske matrice.

Tabela 5-4 Korelacijska matrica tri zadatka

	Zadatak 1	**Zadatak 2**	**Zadatak 3**
Zadatak 1	1	$\boldsymbol{r_{12}}$	$\boldsymbol{r_{13}}$
Zadatak 2	r_{21}	1	r_{23}
Zadatak 3	r_{31}	r_{12}	1

5.2.2.3 Korelacija dvije linearne kombinacije

Korelacija dvije linearne kombinacije određuje se kao standardizirana kovarijanca te dvije kombinacije.

Pretpostavimo da je data kovarijančna matrica dvije linearne kombinacije, od kojih je prva definirana kao:

$$U = A + B + C$$

dok je druga definirana kao:

$$Y = G + H$$

Njihova kovarijančna matrica je:

Tabela 5-5 Kovarijančna matrica dvije linearne kombinacije

	A	*B*	*C*	*G*	*H*
A	V_A	cov_{AB}	cov_{AC}	cov_{AG}	cov_{AH}
B	cov_{BA}	V_B	cov_{BC}	cov_{BG}	cov_{BH}
C	cov_{CA}	cov_{CB}	V_C	cov_{CG}	cov_{CH}
G	cov_{GA}	cov_{GB}	cov_{GC}	V_G	cov_{GH}
H	cov_{HA}	cov_{HB}	cov_{HC}	cov_{HG}	V_H

Ovu složenu matricu možemo podijeliti u tri različita dijela:

1. Kovarijančna matrica testa U koja sadrži varijance i kovarijance dijelova A, B i C (obojena roza)
2. Kovarijančna matrica testa Y koja sadrži varijance i kovarijance testa Y (obojena plavo)
3. Kovarijančna matrica dijelova oba testa (obojena zeleno)

Suma elemenata dijela matrice U*Y u kojem se nalaze sve pojedinačne kovarijance dijelova testa je kovarijanca između ove dvije linearne kombinacije.

Korelacija između dvije linearne kombinacije određuje se tako da:

1. Sumiramo sve elemente kovarijančne matrice U*Y (zeleni dio)
2. Odredimo standardne devijacije dvije linearne kombinacije tako da sumiramo sve elemente njihovih kovarijančnih matrica i odredimo drugi korijen (kao što smo naučili u sekciji 5.2.1.1) i dobivene elemente uvrstimo u formulu koja prikazuje odnos kovarijance i korelacije:

5-28

$$r_{UY} = \frac{c_{UY}}{s_U s_Y}$$

5.2.3 Što možemo koristiti za ponderiranje rezultata?

Ponderiranje rezultata, odnosno određivanje visine konstante kojom množimo rezultate na pojedinim zadacima je ozbiljan zadatak. Kako odrediti koliko je neki zadatak "važniji" za razlikovanje ispitanika u odnosu na neki drugi zadatak? Za diferencijalno ponderiranje, Nunnally i Bernstein (1994) u svom klasiku „Psychometric Theory“ kažu da postoje različita pravila kako ponderirati pitanja. Visina pondera može biti određena visinom regresijskog koeficijenta pitanja, odnosno korelacijom pitanja s ukupnim uratkom određenim kao jednostavna linearna kombinacija. Ipak, oni ne preporučuju korištenje korelacije jer 1) generalno nije mudro da se pitanja konstruiraju prema korelaciji s kriterijem te 2) korelacija pitanja s ukupnim uratkom u sebi sadrži dosta greške. Također ističu da ponderiranje ima smisla i pravi razliku između jednostavne i diferencijalno ponderirane kombinacije onda kada ponderirana pitanja imaju veću pouzdanost (o pouzdanosti ćemo govoriti uskoro). Osim toga, dokazi govore o tome da su korelacije između ponderiranih i neponderiranih ukupnih skorova obično jako visoke ukoliko test ima najmanje 20 pitanja. Dakle, ponderiranje daje nešto više koristi ako test ima manje od 20 pitanja, ako korelacije pitanja i ukupnog rezultata jako variraju te je uzorak ispitanika veliki (najmanje nekoliko stotina).

Neki osnovni načini određivanja pondera su:

1. Ponderiranje prema indeksu težine zadatka – na primjer ukoliko je to jako težak zadatak koji rješava samo 10 % ispitanika pri čemu je q=0,9, kao broj bodova koji se dobiva za tačno urađen zadatak možemo odrediti 10*q, odnosno 9 bodova.
2. Ponderiranje prema proporciji učešća u rezultatu – ukoliko na primjer želimo da neki segment testiranja učestvuje u konačnoj ocjeni sa 20 %, tada odredimo rezultat na tom segmentu kao

$$\mathrm{w} = \frac{rezultat\ ispitanika}{maksimalni\ rezultat} * 100 * proporcija\ učešća$$

3. Regresijsko ponderiranje
   Prema Fajgelju (2003) ovo je najpoznatije optimalno ponderiranje. U ovom ponderiranju koriste se regresijski koeficijenti (koje ste spominjali u regresijskoj analizi na Statistici, a o kojima će biti riječi u poglavlju o valjanosti).

6 POUZDANOST

Što je pouzdanost testa? Kako određujemo koliko je neki test pouzdan? Kako uzorak i broj zadataka u testu utječu na pouzdanost? Što je generalizabilnost?

Psihološka mjerenja su neophodna za istraživanja i aplikaciju psiholoških spoznaja u realnom životu. Jednostavno je nemoguće zamisliti psihološko istraživanje i provođenje nekog psihosocijalnog programa bez mjerenja. Da bi mjerenje bilo kvalitetno, testovi moraju biti pouzdani.

Zapitajmo se: ko je pouzdan čovjek? Vaš odgovor vjerovatno uključuje opise kao što su – onaj čovjek kome mogu vjerovati, onaj ko je tačan, onaj ko neku aktivnost uvijek radi na isti način, neko na koga se mogu osloniti.

Što je, onda, pouzdan test? Slično kao i čovjek, to je test kojem možemo vjerovati da će svaki put kada je upotrebljen s istim ispitanikom dati iste ili vrlo slične rezultate i na koje se možemo osloniti prilikom zaključivanja. Zamislite da odete kod liječnika i on vam dva puta mjeri krvni tlak. Prvi put rezultat pokaže da je vaš krvni tlak u fiziološkim granicama, a drugi put vam tim istim tlakomjerom izmjeri vrijednosti koje ukazuju da vi hitno morate na terapiju zbog visokog tlaka?! Sada zamislite kako bi bilo da koristimo nepouzdan psihološki test u situacijama kada odlučujemo ko će se upisati u školu, koga ćemo zaposliti na neko radno mjesto ili koga ćemo hospitalizirati. Nepouzdan test bi nam mogao jednom pokazati da neki učenik ne može biti upisan u školu, a drugi put da može ili u prvom mjerenju zaključimo da osobu treba hospitalizirati, a u drugom mjerenju zaključimo da je možemo pustiti na kućni tretman. Korištenje nepouzdanih testova predstavljalo bi opasnu prijetnju po ljude.

Pouzdanost testa predstavlja jednu od osnovnih metrijskih karakteristika koja govori o konzistenciji i preciznosti rezultata mjernog procesa. Kako bismo mogli vjerovati testnim rezultatima, mi moramo biti uvjereni da ako će ispitanici, ako im damo test nekoliko puta, postići slične ili iste rezultate. Nedostatak pouzdanosti znači nepreciznost i nekonzistenciju, odnosno pogrešku mjerenja. Pogreška mjerenja definira se kao promjena u rezultatu koja je nastala pod utjecajem faktora povezanih s procesom mjerenja, ali nerelevantnih za ono što je mjereno. Prema tome, pouzdanost kao metrijska karakteristika govori o tome koliko su rezultati konzistentni i oslobođeni od pogreške mjerenja (Urbina, 2004, str. 117).

6.1 KLASIČNA TESTNA TEORIJA

Da bismo razumjeli pouzdanost, moramo krenuti od teorija koje objašnjavaju ovu metrijsku karakteristiku. Prva teorija od koje krećemo je *Klasična testna teorija* koja se još naziva i Teorija pravog rezultata. Začeci ove teorije bili su radovi Spearmanaa, Yulea, Kellyja, Guttmana i čitave plejade psihologa. Njeno uobličavanje kako je danas znamo dao je Novick (1966), a objašnjena je i u klasicima kao što su Lord i Novick (1968) i Allen i Yen (1979).

Klasična testna teorija nastala je nakon što su se konceptualizirale tri osnovne ideje ili pretpostavke:

1. U mjerenju postoji pogreška, rezultat koji se dobije mjerenjem u sebi uvijek sadrži pogrešku
2. Pogreška nastaje zbog djelovanja nesistematskih faktora i ona je slučajna
3. Korelacija može biti korištena za izračunavanje pouzdanosti

Visina rezultata ispitanika na testu ili pitanju određena je:

1. Karakteristikom koju mjerimo – ako testom mjerimo depresiju, onda će izraženost depresivnosti kod ispitanika učestvovati u formiranju rezultata
2. Sistematskim faktorima –varijablama koje učestvuju u formiranju rezultata tako da kod svih ispitanika mijenjaju rezultat u istom smjeru. Na primjer, ako mjerimo raspoloženje, a ispitanici su prije toga gledali neku komediju i od srca se smijali, to će uticati tako da će kod većine dovesti do pomjeranja rezultata na upitniku prema višim vrijednostima
3. Nesistematskim faktorima – varijablama koje djeluju po slučaju i rezultate mijenjaju u oba smjera i sa različitim intenzitetom.

Teorija pouzdanosti se bavi djelovanjem **nesistematskih** faktora.

6.1.1 Struktura bruto rezultata

Nakon što ispitanik popuni test, rezultat koji vidimo zove se bruto, opaženi ili sirovi rezultat. Samo ime „bruto" govori da se u ovom rezultatu nalaze određene komponente. Bruto rezultat je prema Klasičnoj testnoj teoriji sastavljen od dva dijela – pravog rezultata i komponente pogreške.

$$X_b = X_t + X_e$$

Pravi rezultat (engl. true score) je, jednostavno rečeno, stvarna izraženost neke karakteristike koju mi mjerimo. Da nema pogreške mjerenja i da smo potpuno precizni u mjerenju, mi bismo registrirali pravi rezultat. Međutim, zbog djelovanja pogreške mjerenja mi registriramo bruto rezultat koji je „opterećen" pogreškom mjerenja. Kada bismo mjerenje ponovili veliki broj puta sa istim ispitanicima, onda bi prosjek mjerenja za svakog ispitanika bio njegov pravi rezultat jer bi se pogreške poništile (pošto se pojavljuju po slučaju i po slučaju mijenjaju pojedinačne rezultate). Ovo nas navodi na još jednu važnu pretpostavku Klasične testne teorije, a to je da je predmet mjerenja stabilan kroz vrijeme.

Da sumiramo, osnovne pretpostavke od kojih polazi Klasična testna teorija su:

1. Predmet mjerenja je stabilan kroz vrijeme (inače pravi rezultat ne bi mogao biti isti)
2. Pogreške mjerenja su slučajne (i mijenjaju rezultat po slučaju)
3. Bruto rezultat je zbir pravog rezultata i komponente pogreške

Zamislimo sada jednu matricu bruto rezultata grupe od N ispitanika na K mjerenja pri čemu je svako mjerenje isti test / item / procedura :

Tabela 6-1 Matrica bruto rezultata prema Klasičnoj testnoj teoriji

	TESTOVI				PARAMETRI PO ISPITANIKU	
ISPITANICI	1	2	..	K	M	V
1	${}_bX_{11} = {}_tX_{11} + {}_eX_{11}$	${}_bX_{12} = {}_tX_{12} + {}_eX_{12}$	..	X_{1k}	${}_bM_{1.}, {}_tM_{1.}, {}_eM_{1.}$	${}_bV_{1.}, {}_tV_{1.}, {}_eV_{1.}$
2	${}_bX_{21} = {}_tX_{21} + {}_eX_{21}$	${}_bX_{22} = {}_tX_{22} + {}_eX_{22}$	..	X_{2k}	${}_bM_{2.}, {}_tM_{2.}, {}_eM_{2.}$	${}_bV_{2.}, {}_tV_{2.}, {}_eV_{2.}$
3	${}_bX_{31} = {}_tX_{31} + {}_eX_{31}$	${}_bX_{32} = {}_tX_{32} + {}_eX_{32}$	..	X_{3k}	${}_bM_{3.}, {}_tM_{3.}, {}_eM_{3.}$	${}_bV_{3.}, {}_tV_{3.}, {}_eV_{3.}$
…	…	…		…	…	…
N	${}_bX_{N1} = {}_tX_{N1} + {}_eX_{N1}$	${}_bX_{N2} = {}_tX_{N2} + {}_eX_{N2}$	..	X_{Nk}	${}_bM_{N.}, {}_tM_{N.}, {}_eM_{N.}$	${}_bV_{N.}, {}_tV_{N.}, {}_eV_{N.}$
PARAMETRI PO TESTU						
M	${}_bM_{.1}, {}_tM_{.1}, {}_eM_{.1}$	${}_bM_{.2}, {}_tM_{.2}, {}_eM_{.2}$		${}_bM_{.N}, {}_tM_{.N}, {}_eM_{.N}$		
V	${}_bV_{.1}, {}_tV_{.1}, {}_eV_{.1}$	${}_bV_{.2}, {}_tV_{.2}, {}_eV_{.2}$		${}_bV_{.N}, {}_tV_{.N}, {}_eV_{.N}$		

Svaki bruto rezultat možemo posmatrati kao linearnu kombinaciju pravog rezultata i komponente pogreške. U slučaju ovako definirane linearne kombinacije (jednostavne i aditivne) znamo da je:

1. Aritmetička sredina linearne kombinacije jednaka zbiru aritmetičkih sredina dijelova, što znači da je aritmetička sredina bruto rezultata jednaka zbiru aritmetičke sredine pravih rezultata i aritmetičke sredine pogreške
2. Varijanca linearne kombinacije jednaka je zbiru varijanci elemenata i dvostrukom zbiru kovarijanci elemenata. To znači da je varijanca bruto rezultata hipotetski jednaka zbiru varijanci pravih rezultata, varijanci pogreške te dvostrukoj kovarijanci između pravih rezultata i komponente pogreške.

Uzimajući u obzir osnovne pretpostavke Klasične testne teorije, te osnovnu matematiku vezanu za aritmetičku sredinu i varijancu linearnih kombinacija, možemo reći ponešto o aritmetičkoj sredini i varijanci za ispitanika i za mjerenje.

Parametri po ispitaniku:

Ako važe pretpostavke KTT-a koje kažu da su pogreške slučajne i da po slučaju mijenjaju rezultat, onda možemo zaključiti da je:

1. Aritmetička sredina pravih rezultata u osnovi sam pravi rezultat
$$ {}_tM_{i.} = {}_tX_{i.} $$
2. Aritmetička sredina komponenti pogreške nula jer se pogreške ponište
$$ {}_eM_{.1} = {}_eM_{.2} = \ldots = {}_eM_{.j} = 0 $$

3. Aritmetička sredina bruto rezultata (Mb = Mt + Me) je sam pravi rezultat
$$_{b}M_{i.} = {}_{t}X_{i.}$$
4. Varijanca pravih rezultata je nula jer je pravi rezultat isti u ponovljenim mjerenjima
$$_{t}V_{i.} = {}_{t}V_{i.} = 0$$
5. Varijanca bruto rezultata (Vb = Vt + Ve) je jednaka varijanci komponente pogreške
$$_{b}V_{i.} = {}_{e}V_{i.}$$

Parametri po testu:

Ako važe pretpostavke Klasične testne teorije onda će za parametre testova važiti sljedeće:

1. Aritmetička sredina pravih rezultata na svakom testu biće ista
$$_{t}M_{.i} = {}_{t}M_{.j}$$
2. Aritmetičke sredine komponenti pogreške će biti nula (pod pretpostavkom da imamo veliki broj ispitanika)
$$_{e}M_{.i} = {}_{e}M_{.j} = 0$$
3. Aritmetička sredina bruto rezultata biće jednaka aritmetičkoj sredini pravih rezultata
$$_{b}M_{.i} = {}_{t}M_{.i}$$
4. Varijanca pravih rezultata u svakom testu biće ista (jer svaki ispitanik ima svoj pravi rezultat, a on se ponavlja kroz sva mjerenja)
$$_{t}V_{.i} = {}_{t}V_{.j}$$
5. Varijance komponente pogreške biće jednake u svakom mjerenju
$$_{e}V_{.i} = {}_{e}V_{.j}$$
6. Varijanca bruto rezultata biće jednaka zbiru varijance pravih rezultata i varijance komponente pogreške
$$_{b}V_{.i} = {}_{t}V_{.i} + {}_{e}V_{.i}$$

Ako ste pažljivo pročitali dio u kojem je rečeno da je varijanca linearne kombinacije jednaka zbiru varijanci elemenata i dvostrukom zbiru kovarijanci elemenata sada se pitate: „Gdje nam je kovarijanca u jednačini?“ Odgovor na ovo pitanje potražit ćemo u osnovnoj pretpostavci Klasične testne teorije – pogreške su slučajne. Kada se nešto pojavljuje po slučaju ne može korelirati ni sa jednom drugom varijablom. Prema tome, kovarijanca (ili korelacija kao standardizirana kovarijanca) između komponente pogreške i bilo koje druge varijable, pa i pravog rezultata, je uvijek 0.

6.1.2 Što je pouzdanost?

Furr i Bacharach (2013) su predložili jedan koristan način da mislimo o pouzdanosti kao odnosu između signala i šuma. Signal bi bio naš pravi rezultat, a šum je pogreška mjerenja koja ometa percepciju signala i otežava nam da ga detektiramo. Pouzdanost bi bila odnos između količine signala i ukupne količine signala i šuma, odnosno:

$$POUZDANOST = \frac{\text{SIGNAL}}{\text{SIGNAL} + \text{ŠUM}}$$

Iz ove perspektive, pouzdanost je veća kada je signal jak, a šum mali.

U Klasičnoj testnoj teoriji postoje najmanje četiri načina da promišljamo o pouzdanosti, a svi se tiču odnosa između bruto rezultata, pravog rezultata i komponente pogreške. Ova četiri pristupa okarakterizirana su sa dvije dimenzije:

- da li se pouzdanost tumači kao odnos bruto rezultata prema pravom rezultatu ili prema komponenti pogreške.
- da li se pouzdanost tretira kao „proporcija varijance“ ili kao „korelacija“

6.1.2.1 Pouzdanost kao proporcija varijance pravih rezultata i bruto rezultata / komponente pogreške i bruto rezultata

Tumačenje pouzdanosti kao proporcije varijance pravih rezultata u varijanci bruto rezultata, je najčešći koncept razumijevanja pouzdanosti. Pravi rezultati grupe ispitanika variraju i mi želimo da variraju. Da nije tako, mi bismo mjerili konstante. Ali, osim variranja pravih rezultata, varira i pogreška. U dijelu 6.1. vidjeli smo da je varijanca bruto rezultata zbir varijance pravih rezultata i varijance pogreške. Koeficijent pouzdanosti je prema formuli:

6-1

$$r_{xx} = \frac{V_t}{V_B}$$

Primijetite da oznaka za pouzdanost u supskriptu ima dva ista slova. To znači da se ovaj koncept odnosi samo na jedan test i razlikuje se od oznake za korelaciju gdje u supskriptu imamo dva različita slova za dva različita testa.

Pretpostavimo da smo ispitivali grupu ispitanika i da smo mogli zabilježiti i njihov pravi rezultat i njihov bruto rezultat. Pretpostavimo da je varijanca pravih rezultata bila 100, a varijanca bruto rezultata bila 125. Prema gore navedenoj formuli koeficijent pouzdanost bi iznosio:

$$r_{xx} = \frac{100}{125} = 0{,}8$$

odnosno, 80% variranja bruto rezultata objašnjeno je variranjem pravih rezultata, dok je 20% variranja u bruto rezultatima posljedica pogreške. Koeficijent pouzdanost je proporcija varijance pravih rezultata unutar varijance bruto rezultata. Kao i svaka proporcija, koeficijent pouzdanosti se kreće u rasponu od 0 do 1 (u procentima – od 0 do 100 %). Veća vrijednost koeficijenta pouzdanosti ukazuje da imamo precizniji test jer je veća proporcija variranja objašnjena pravim rezultatima. Osim toga, primijetite da će pouzdanost testa biti nula ukoliko je varijanca pravih rezultata jednaka nula što još jednom naglašava činjenicu da je pouzdanost vezana za variranje karakteristike među ispitanicima. Koeficijent pouzdanosti će biti 1 onda kada u mjerenju nije bilo nimalo pogreške, što se u realnom životu nikada ne događa, posebno ne u društvenim znanostima. Iako ne postoji recept koliko visok treba da bude koeficijent pouzdanosti, može se preporučiti „Što viša pouzdanost, to bolje“. U istraživanjima se često zadovoljavamo koeficijentima pouzdanosti koji su veći od 0,7 ili 0,8, dok u situacijama kada moramo donijeti važne odluke na osnovu testiranja težimo još višim koeficijentima pouzdanosti. U svakom slučaju, ukoliko nam je koeficijent pouzdanosti niži od 0,5 trebamo se zabrinuti i rezultate ne koristiti za donošenje zaključaka.

Kada posmatramo pouzdanost iz perspektive odnosa varijance pogreške i bruto rezultata konceptualno ćemo zaključiti isto kao i u prethodnom slučaju. Radi matematike, da vidimo čemu je jednaka pouzdanost kada usporedimo varijancu pogreške i varijancu bruto rezultata.

Znamo da je $V_B = V_t + V_e$ odnosno da je $V_t = V_B - V_e$. Kada ovo uvrstimo u formulu za pouzdanost, dobijemo:

6-2

$$r_{xx} = \frac{V_t}{V_B} = \frac{V_B - V_e}{V_B} = 1 - \frac{V_e}{V_B}$$

U našem ranijem primjeru, ako je varijanca pravih rezultata bila 100, varijanca bruto rezultata bila 125, to znači da je varijanca komponente pogreške 25. Stoga bi pouzdanost prema ovoj formuli bila:

$$r_{xx} = 1 - \frac{25}{100} = 1 - 0{,}25 = 0{,}75$$

Kao rezultat imamo isti koeficijent pouzdanosti.

6.1.2.2 Pouzdanost kao kvadrirana korelacija između pravih i bruto rezultata

Iako mi nikada ne možemo „vidjeti" prave rezultate, sva priča o pouzdanosti govori o tome da je pouzdanost stepen u kojem su bruto rezultati konzistentni sa pravim rezultatima, odnosno da između bruto rezultata i pravih rezultata postoji visoka povezanost. Da vidimo čemu je jednaka korelacija između bruto i pravih rezultata:

$$r_{bt} = \frac{\sum(t+e)t}{Ns_b s_t} = \frac{\sum(t^2+et)}{Ns_b s_t} = \frac{\sum t^2}{Ns_b s_t} + \frac{\sum e \sum t}{Ns_b s_t}$$

Poznato nam je iz sekcije 6.1.1.da je suma otklonskih rezultata pogreške 0 jer se pogreške ponište, što znači da će drugi dio ovog izraza biti 0. Ostane nam:

$$r_{bt} = \frac{\sum t^2}{Ns_b s_t}$$

Sada znamo da je $\frac{\sum t^2}{N}$ ništa drugo do varijanca pravih rezultata tako da imamo:

$$r_{bt} = \frac{V_t}{s_b s_t} = \frac{s_t}{s_b} = \sqrt{\frac{V_t}{V_b}}$$

6-3

$$r_{bt} = \sqrt{r_{xx}}$$

Kao što vidimo, korelacija bruto i pravih rezultata je drugi korijen koeficijenta pouzdanost i zove se INDEKS POUZDANOSTI (Ghiselli, Campbell i Zedeck, 1981). Možemo reći da je koeficijent pouzdanost u stvari koeficijent determinacije za koeficijent korelacije pravih i bruto rezultata i određuje proporciju zajedničke varijance pravih i bruto rezultata.

6.1.3 Pouzdanost i standardna pogreška mjerenja

Iako nam je koeficijent pouzdanost izuzetno važna informacija o kvaliteti mjernog instrumenta, iz njega ne saznajemo puno o samoj veličini pogreške. Mi dobijemo informaciju kolika je proporcija varijance objašnjena pravim rezultatima i pogreškom, ali ništa ne znamo o tome kolika je pogreška. Zamislimo sada ispitanika Admira kojem smo mjerili vrijeme reakcije na svjetlo 20 puta. Pretpostavimo da je Admirov pravi rezultat na ispitivanju vremena reakcije 130 ms. Ipak, on nije svaki put postigao svoj pravi rezultat zbog djelovanja pogreške. Njegovi rezultati su bili (u milisekundama) na prvom mjerenju 132ms, na drugom 129, na trećem 127 i tako dalje. Rezultati su varirali. Standardna devijacija rezultata koje postiže ispitanik na nizu paralelnih mjerenja naziva se **standardna pogreška mjerenja**. Standardna pogreška mjerenja je jedan od najvažnijih koncepata u mjerenju i označava se sa $\mathbf{s_e}$. Što je standardna pogreška mjerenja veća, to je veća razlika između bruto rezultata i pravih rezultata te je samim tim test manje pouzdan.

Kako računamo standardnu pogrešku mjerenja? Vidjeli smo da je

$$r_{xx} = 1 - \frac{V_e}{V_B}$$

Iz toga slijedi:

$$\frac{V_e}{V_B} = 1 - r_{xx}$$

$$V_e = V_B(1 - r_{xx})$$

$$s_e = \sqrt{V_B(1 - r_{xx})}$$

6-4

$$s_e = s_B\sqrt{(1 - r_{xx})}$$

Ova formula pokazuje da je standardna pogreška mjerenja povezana s koeficijentom pouzdanosti. Kada je pouzdanost $r_{xx} = 1$ tada je standardna pogreška mjerenja jednaka 0, odnosno ako nema greške u mjerenju tada je pouzdanost savršena. Također vidimo da standardna pogreška mjerenja ne može biti veća od standardne devijacije bruto rezultata.

6.1.4 Paralelne forme

Sigurna sam da ste se do sada zapitali: „U ovoj knjizi se stalno spominju pravi rezultati, a mi nikada ne vidimo prave rezultate. Kako da znam pravi rezultat?“. I zaista, do sada smo pričali o tome kako je pouzdanost odnos varijance pravih i bruto rezultata, odnosno kvadrat korelacije pravih i bruto rezultata. Pažljivom čitaocu ova priča može izgledati kao nemoguća za praktično određivanje pouzdanosti. Međutim, teoretičari Klasične testne teorije riješili su ovaj problem *pretpostavkom o paralelnosti mjerenja*. Oni kažu da dva mjerenja možemo konstruirati tako da su paralelna. Dva mjerenja su paralelna onda kada su:

1. pravi rezultati na prvom i drugom mjerenju potpuno jednaki, odnosno $X_{t1} = X_{t2}$, te je njihove korelacija 1. Kada testovi imaju jednake prave rezultate, onda će i aritmetičke sredine pravih rezultata, aritmetičke sredine bruto rezultata kao i varijance pravih rezultata biti jednake.
2. varijance pogreške jednake $V_{e1} = V_{e2}$. Ukoliko su jednake varijance pogreške to znači da će i varijance bruto rezultata biti jednake.
3. Korelacije ovih testova sa vanjskim kriterijskim varijablama jednake.

Kako bismo provjerili da li su dva mjerenja paralelna, potrebno je da testiramo statističku značajnost razlika aritmetičkih sredina, statističku značajnost razlika varijanci, te provjerimo korelaciju među testovima. Ustanovljavanje paralelnosti testova podrazumijeva i faktorsku analizu kojom provjeravamo da li ova mjerenja mjere istu latentnu varijablu, ali o faktorskoj analizi govorit ćemo u poglavlju o valjanosti. Ukoliko rezultati pokažu da imamo jednake aritmetičke sredine i varijance, te da se rezultati ova dva mjerenja mogu objasniti istom latentnom varijablom, onda kažemo da su to dva paralelna mjerenja. Pretpostavka o paralelnosti mjerenja je jako važna jer nas ona uvodi u priču o praktičnom određivanju koeficijenata pouzdanosti.

Pretpostavimo da imamo rezultate na dva mjerenja koja zadovoljavaju pretpostavku o paralelnosti (na primjer dva pitanja na testu profesionalnih interesa – socijalni tip interesa):

Koliko vam se dopadaju sljedeće aktivnosti:

1. Pomagati drugima
2. Pomagati djeci sa poteškoćama u učenju

Ova dva pitanja mjere stepen sviđanja unutar istog tipa interesa. Odgovor na prvo pitanje trebao bi da bude sličan odgovoru na drugo pitanje. Prema Klasičnoj testnoj teoriji, korelacija između ova dva pitanja je pouzdanost svakog od pitanja. Dakle, kada možemo pretpostaviti da imamo dva paralelna mjerenja, onda je koeficijent korelacije između mjerenja ujedno i koeficijent pouzdanosti tih mjerenja. Da bismo dokazali ovu tvrdnju poći ćemo od osnovne pretpostavke Klasične testne teorije da je bruto rezultat zbir pravog rezultata i komponente pogreške:

$$\mathbf{X_b = X_t + X_e}$$

Osim toga znamo da je korelacija između dva mjerenja X i Y jednaka:

$$r_{xy} = \frac{c_{XY}}{s_X s_Y}$$

Ukoliko je zadovoljena pretpostavka o paralelnosti mjerenja onda će standardne devijacije bruto rezultata ova dva testa biti jednake.

$$s_X = s_Y$$

Kako je bruto rezultat jednostavna linearna kombinacija pravog rezultata i komponente pogreške, to će kovarijanca između bruto rezultata na dva mjerenja biti jednaka sumi kovarijanci svih članica ove dvije linearne kombinacije:

$$r_{X_bY_b} = \frac{cov_{X_tY_t} + cov_{X_tY_e} + cov_{X_eY_t} + cov_{X_eY_e}}{s_{X_b}s_{Y_b}}$$

Kako znamo da je pogreška slučajna i ne korelira ni sa čim, možemo reći da su kovarijance bilo kojeg rezultata sa komponentom pogreške jednake 0.

$$cov_{X_tY_e} = cov_{X_eY_t} = cov_{X_eY_e} = 0$$

Slijedi da je:

$$r_{X_bY_b} = \frac{cov_{X_tY_t}}{s_{X_b}s_{Y_b}}$$

Prisjetimo se da je kovarijanca jednaka:

$$c_{XY} = r_{XY}s_Xs_Y$$

Kod određivanja korelacije između dvije paralelne mjere, korelacija između pravih rezultata je jednaka 1, varijanca pravih rezultata je jednaka i varijance pogreške su jednake, pa samim tim i varijance bruto rezultata:

$$cov_{X_tY_t} = r_{X_tY_t}s_{X_t}s_{Y_t}$$

$$r_{X_tY_t} = 1$$

$$s_{X_t} = s_{Y_t}; s_{X_e} = s_{Y_e} \Rightarrow s_{X_b} = s_{Y_b}$$

Dobijemo da je korelacija između bruto rezultata dvije paralelne forme jednaka:

6-5

$$r_{X_bY_b} = \frac{{s_{X_t}}^2}{{s_{X_b}}^2} = \frac{V_t}{V_B}$$

Vidimo da je korelacija dvije paralelne forme jednaka odnosu između varijance pravih rezultata i varijance bruto rezultata, odnosno koeficijentu pouzdanosti.

6.2 EMPIRIJSKO ODREĐIVANJE POUZDANOSTI

Metode za određivanje pouzdanosti mogli bismo podijeliti prema tome kako je pouzdanost tretirana, iako se sve metode određivanja pouzdanosti zasnivaju na pretpostavkama o paralelnosti mjerenja (Furr i Bacharach, 2013). Postoje dva osnovna koncepta razumijevanja pouzdanosti i skladu s tim i metode kojima se pouzdanost određuje.

6.2.1 Pouzdanost kao stabilnost karakteristike u vremenu

Prema ovom konceptu pretpostavljamo da je karakteristika koju mjerimo stabilna u vremenu, odnosno da nije pod utjecajem različitih faktora kao što su zrenje, učenje, socijalna situacija i slično. Pretpostavimo da mjerimo karakteristiku koja je dostigla razvojnu stabilnost, koja nije pod utjecajem učenja u školi, niti se tako drastično mijenja pod utjecajem svakodnevnih životnih situacija ili politike. U tom slučaju pretpostavljamo da će test, primijenjen više puta u nekoliko vremenskih tačaka na istim ispitanicima mjeriti tu karakteristiku na isti način, odnosno da će ispitanici na nizu ponovljenih mjerenja imati isti pravi rezultat.

Na raspolaganju imamo najmanje dvije metode za određivanje koeficijenta pouzdanosti kao mjere stabilnosti karakteristike u vremenu.

6.2.1.1 Test – retest koeficijent pouzdanosti

Test-retest pouzdanost određuje se tako da se ispitanici testiraju u dvije vremenske tačke istim testom. Nakon toga se odredi visina korelacije između rezultata i ta korelacija predstavlja test-retest koeficijent pouzdanosti testa.

6-6

$$r_{X_1X_2} = r_{XX}$$

Iako djeluje vrlo jednostavno, postoji nekoliko pretpostavki koje leže u osnovi „vjerovanja“ ovom koeficijentu. Prva pretpostavka je da se pravi rezultati ispitanika ne mijenjaju tokom vremena. Mi trebamo biti uvjereni da je karakteristika koju mjerimo toliko stabilna da na nju neće ništa uticati tokom vremena koje protekne između prvog i drugog mjerenja. Druga pretpostavka je da će varijanca pogreške u prvom mjerenju biti jednaka varijanci pogreške u drugom mjerenju**.** Međutim, koliko je ova pretpostavka opravdana? Mi znamo da se varijanca pogreške javlja zbog slučajnih faktora koji djeluju u svakom trenutku i koji mijenjaju pravi rezultat u oba smjera i sa različitim intenzitetom. Da bi bili sigurni u ovu pretpostavku za drugo mjerenje moramo kreirati situaciju koja će biti gotovo identična situaciji na prvom mjerenju. Ispitanike bi trebali smjestiti u istu prostoriju; testirati ih u isto doba dana; imati istog ispitivača jer, kako navodi Bucik (1997,str 106), neki ispitivači mogu dati jasnu i potpunu uputu, a drugi ne; formirati iste grupe kako bi socijalni kontekst bio isti i tako dalje. Ukoliko i pretpostavimo da se možemo nositi sa legitimnošću ove pretpostavke, opet nam ostaje pitanje stabilnosti karakteristike u vremenu. Zašto? Prvo, zato jer su psihološki konstrukti rijetko tako stabilni u vremenu. Pretpostavimo da mjerimo ekstraverziju. Ekstraverzija se smatra stabilnom karakteristikom. Pretpostavimo da su neka od pitanja na testu bila:

- Zanimaju me drugi ljudi.
- Posvećujem vrijeme drugim ljudima.
- Šutljiv sam pored nepoznatih ljudi

Pitanja iz IPIP 100

Međutim, u periodu između dva testiranja ispitanik je mogao proći kroz neke životne događaje koji su ga naveli da promjeni svoja ponašanja.Mogao je imati jako puno obaveza na fakultetu i ne posvećivati pažnju drugim ljudima. Osoba, zahvaljujući toj situaciji u kojoj se zatekla, stekne dojam o sebi da je ljudi pretjerano ne zanimaju, te da im i ne posvećuje previše pažnje. Sljedeći put kada dođe na testiranje, naš ispitanik može odgovarati drugačije. Njegov latentni pravi rezultat se možda i nije promijenio, ali se njegov dojam o sebi promijenio. Odgovarajući iz tog novog dojma, ispitanik više neće imati isti manifestni pravi rezultat koji mi jedino možemo mjeriti. Ipak, treba reći da za neke karakteristike, kao što su sposobnosti, test-retest metoda daje kvalitetne podatke o pouzdanosti. Drugo, problem koji utiče na obje pretpostavke od koje polazimo je problem vremena koje je potrebno da prođe između dva testiranja. Pretpostavimo da smo mjerili inteligenciju kod grupe ispitanika, te da smo isti test ponovo primijenili za 7 dana. Velika je šansa da su 1) ispitanici zapamtili neke od svojih odgovora ili postupaka kako su došli do rješenja, te da su 2) rezimirali mjesta gdje su griješili i da su u ovih sedam dana došli do pravog rješenja. Ovdje bismo imali problem dosjećanja, što bi dovelo da toga da ispitanici nemaju jednake prave rezultate na ponovljenim mjerenjima. S druge strane, predugo čekanje nas dovodi u opasnost da različiti faktori zaista utiču na mjerenu karakteristiku i da se ona promijeni u tom periodu. Na primjer, ako koristimo neki test rječnika i čekamo godinu dana, velika je vjerovatnoća da će osoba u procesu obrazovanja, ali i u svakodnevnim situacijama obogatiti svoj rječnik i bolje razumjeti riječi koje možda nije znala prvi put. Ovaj primjer pokazuje još jedan problem o kojem moramo razmišljati, a to je životni ili razvojni period ispitanika u kojem se radi testiranje. Pod utjecajem određenih specifičnih faktora u pojedinim periodima života (školovanje kod djece, promjene u socijalnom statusu ili radnom statusu kod odraslih, rođenje, smrti,...) dolazi do promjena psiholoških karakteristika, pa čak i onih za koje vjerujemo da su stabilne. Stoga u određivanju test-retest koeficijenta pouzdanosti moramo dobro promisliti koliko je karakteristika zaista stabilna.

6.2.1.2 Alternativne forme

Jedan od načina kako pokušamo prevladati problem vremenskog intervala kod test-retest metode je korištenje alternativnih formi.

Alternativne forme su dvije forme testa koje mjere istu stvar i mjere je na isti način. Ispitanike onda možemo testirati doslovno u jednom danu (na primjer, ujutro i poslijepodne), a korelacija između rezultata na alternativnim formulama je koeficijent pouzdanosti svake od tih formi. Na taj način rješavamo problem vremenskih intervala i svih drugih problema koji proističu iz njega.

Međutim, upotreba alternativnih formi također nije bez zadrške. Bucik (1997) daje primjer kako i u ovom slučaju vrijeme između zadavanja dvije forme utiče na visinu pouzdanosti. Zamislimo da imamo dvije alternativne forme testa koji mjeri numerički faktor. Zamislimo da ova dva testa zadajemo učenicima prvih razreda srednje škole u različitim vremenskim razmacima:

1. Situacija 1 – test A smo zadali u jutarnjim satima, a test B u poslijepodnevnim satima. Dobili smo korelaciju od 0,95, što nam govori da je koeficijent pouzdanosti 0,95, odnosno da je samo 5% varijance pogreške u rezultatima.
2. Situacija 2- test A i B zadajemo u vremenskom razmaku od mjesec dana. Korelacija između rezultata pada na 0,90. što znači da je sada udio varijance pogreške 10%.
3. Situacija 3 – test A i B zadajemo u razmaku od godinu dana i korelacija sada iznosi 0,80. Sada je udio pogreške procijenjen na 20% što je već jako puno za test sposobnosti.

Što bi mogli biti potencijalni razlozi za ovu promjenu?

Prvi razlog može biti da same alternativne forme nisu paralelne, odnosno da ne zadovoljavaju sve kriterije paralelnosti mjerenja. Alternativne forme imaju donekle različit sadržaj zadataka, što može voditi ka tome da zadaci nisu jednako teški ili razumljivi ispitanicima. Ovo možda i nije toliki problem kod zadataka u testu numeričkih sposobnosti, ali na testiranjima ličnosti paralelnim formama, gdje pitanja mogu mjeriti druge varijable, bez čvrstog dokaza o paralelnosti mjerenja ne možemo vjerovati koeficijentu pouzdanosti dobivenom na ovaj način. Drugi razlog može biti utjecaj prvog testiranja na drugo testiranje efektom prelijevanja (Furr, 2013,str.127). Na primjer, nakon što su uradili prvu formu, zbog efekta uvježbavanja, usmjeravanja pažnje, raspoloženja, osjećanja samoefikasnosti, manje anksioznosti ili nečega drugog, ispitanici u drugom testiranju postižu sistematski drugačije rezultate. Ovi faktori neće podjednako uticati na sve njih, ali greška više nije u potpunosti slučajna nego ima sistematskog faktora u njenom formiranju.

Vidimo da određivanje pouzdanosti kao stabilnosti karakteristike u vremenu jeste matematički jednostavno, ali konceptualno zahtijeva mnogo promišljanja o svim faktorima koji utječu na mjerenje i njihovo kontroliranje kako bismo bili sigurni korelaciju možemo sa visokom sigurnošću tretirati kao koeficijent pouzdanosti.

6.2.2 Pouzdanost kao unutarnja konzistencija

Ovaj način ocjenjivanja pouzdanosti se zasniva na analizi svojstava i odnosa pitanja unutar testa, te zahtijeva samo jednu primjenu testa. Koeficijent pouzdanosti prema unutarnjoj konzistenciji govori o tome koliko pitanja u testu slično mjere određenu karakteristiku. Prije nego krenemo sa određivanjem koeficijenta pouzdanosti kao unutarnje konzistencije moramo razjasniti na što se odnosi ovaj termin „unutarnja konzistencija" i koja je razlika između unutarnje konzistencije i homogenosti testa. Unutarnja konzistencija se odnosi sa međusobnu povezanost seta pitanja u testu, dok se homogenost pitanja odnosi na jednodimenzionalnost seta pitanja (pitanja mjere isti faktor i to samo jedan faktor). Unutarnja konzistencija jeste potreban, ali ne i dovoljan uvjet za homogenost (Schmitt, Uses and Abuses of Coefficient Alpha, 1996). Razliku između ova dva pojma najlakše je opisati matricom korelacija između dva seta pitanja kako je to uradio i Schmitt. Pretpostavimo da imamo dva seta od po šest pitanja na testu inteligencije pri čemu su njihove korelacije date u dvije korelacijske matrice:

Tabela 6-2 Korelacije dva seta varijabli jednake pouzdanosti različite dimenzionalnosti

Varijabla	1	2	3	4	5	6	Varijabla	1	2	3	4	5	6
1	-						1	-					
2	0,8	-					2	0,5	-				
3	0,8	0,8	-				3	0,5	0,5	-			
4	0,3	0,3	0,3	-			4	0,5	0,5	0,5	-		
5	0,3	0,3	0,3	0,8	-		5	0,5	0,5	0,5	0,5	-	
6	0,3	0,3	0,3	0,8	0,8	-	6	0,5	0,5	0,5	0,5	0,5	-
	α=0,86							**α=0,86**					

*Primjer iz Schmitt (1996)

Pogledajmo ove dvije korelacijske matrice. Za oba seta podataka dobili smo isti koeficijent pouzdanosti od 0,86. Međutim, u prvom setu vidimo da između prva tri pitanja i druga tri pitanja imamo visoke korelacije, te da obrazac korelacija ukazuje na postojanje zajedničkog generalnog faktora. Kada bismo eliminirali generalni faktor, elementi matrice u kojima se nalaze korelacije prva tri pitanja sa druga tri pitanja ostali bi prazni, odnosno njihove korelacije bez generalnog faktora bile bi 0. S druge strane, oni međusobno unutar dvije grupe imaju visoke korelacije koje bi se zadržale (ali ne tako visoko) i kada bismo „izbacili" generalni faktor. To nam govori je ovo test koji mjeri više dimenzija. Test ima visok koeficijent unutarnje konzistencije, ali nije homogen. Odstranjivanje generalnog faktora iz rezultata u drugoj matrici doveo bi do nultih korelacija što govori o tome da nema nekih grupno specifičnih i item specifičnih faktora koji utiču na odgovore ispitanika te je ovaj test jednodimenzionalan, odnosno homogen. Stoga je jako važno da razlikujemo homogenost kao jednodimenzionalnost i unutarnju konzistenciju (kao pomalo neprecizan termin) kao povezanost između pitanja.

Iz perspektive unutarnje konzistencije (za koju sada znamo da nije homogenost), dva su ključna faktora koja utiču na pouzdanost testa. Prvi faktor je (očigledno) konzistencija pitanja. Ako su pitanja povezana među sobom tada će i test biti pouzdan, odnosno ukoliko su opažene razlike među rezultatima na pitanjima konzistentne, onda možemo zaključiti da imamo bruto rezultate konzistentne sa pravim rezultatima. Drugi faktor je dužina testa. Što je test duži, to je njegova pouzdanost veća, s tim što dužina testa podrazumijeva da imamo kvalitetna pitanja. O ovome će biti više govora u dijelovima koji slijede.

Naučit ćemo dvije metode kako se određuje koeficijent pouzdanosti po unutarnjoj konzistenciji: split- half koeficijent ili prepolovljeni koeficijent pouzdanosti i Cronbachov α (alfa) koeficijent pouzdanosti.

6.2.2.1 Split – half koeficijent pouzdanosti

Ova metoda podrazumijeva da test podijelimo na dva dijela koji su dvije „paralelne" forme. Ovo namjerno stavljamo pod navodnike jer je teško podijeliti test u dvije paralelne forme. Možemo ga podijeliti po principu par-nepar ili prva polovina – druga polovina. Odabir načina kako ćemo podijeliti test zavisi od toga kako je test dizajniran (Urbina, 2004, str. 127). Dvije stvari o kojima treba voditi računa su 1) da li se pitanja sistematski razlikuju kroz test i 2) da li je brzina igrala ulogu u testiranju, odnosno da li je bio test brzine:

1. Testovi u kojima se pitanja razlikuju po težini (kao što je to slučaj u testovima postignuća gdje počinjemo sa lakšim pitanjima, a završavamo sa teškim pitanjima) ne mogu biti podijeljeni tako da u jednoj polovini budu laka pitanja, a u drugoj polovini teška pitanja. Neki multidimenzionalni testovi, kao što su testovi ličnosti, imaju pitanja poredana tako da ide pitanje iz jedne dimenzije, pa druge dimenzije, pa treće dimenzije, itd. U ovakvim situacijama, test ne možemo dijeliti par-nepar niti možemo samo prepoloviti test. Moramo u svakoj polovini imati podjednak broj pitanja prema svakoj dimenziji.
2. Ukoliko imamo testove brzine, ispitanici imaju ograničeno vrijeme rješavanja i pretpostavka je da neće stići do zadnjih pitanja. Kada bismo ove testove samo prepolovili dobili bismo u prvoj polovini gotovo sve riješene zadatke, a u drugoj više neriješenih zadataka.

Prilikom podjele trebamo voditi računa da u svakom dijelu testa imamo zadatke koji su različite težine i da svaki zadatak u prvom dijelu ima svoj par u drugom dijelu.

Nakon što smo test podijelili, izračunamo ukupne uratke na dvije polovine i odredimo koeficijent korelacije između njih. Ovaj koeficijent korelacije između dvije polovine testa je koeficijent pouzdanosti za svaku od polovina testa.

Sada prelazimo na korak 2 , a to je određivanje pouzdanosti za čitav test. Za ovu svrhu koristimo Spearman – Brownovu formulu ili Spearman – Brownov popravak koju su ova dva autora nezavisno izveli 1910. godine (prema Fajgelj, 2003, str.230). Ova formula je za slučaj kada se testu kojem je poznat koeficijent pouzdanost poveća dužina za dva puta:

6-7

$$S - B = \frac{2r_{pp\prime}}{1 + r_{pp\prime}}$$

Ovu formulu ćemo, u njenom općem obliku spomenuti i kada budemo govorili o determinantama pouzdanosti.

Dakle, sada nakon što smo dobiveni koeficijent pouzdanosti polovine testa uvrstili u S-B formulu dobili smo koeficijent pouzdanosti za čitav test. Važno je da primjetimo razliku između split-half koeficijenta i koeficijenta dobivenog na alternativnim formama ili test-retest koeficijenta. Posljednja dva su dobivena kao korelacije između čitavih testova, dok je korelacija u određivanju split – half koeficijenta dobivena na dvije polovine istog testa, te je stoga mjera unutarnje konzistencije, a ne stabilnosti karakteristike u vremenu.

Da pogledamo jedan primjer. Zamislimo da je grupa od 12 učenika radila test iz numeričkog faktora gdje su mogli na zadatku odvojiti 0; 0,5 i 1 bod.

Tabela 6-3 Split – half koeficijent pouzdanosti

	Zadaci							Split half		Split half	
Isp	Z1	Z2	Z3	Z4	Z5	Z6	**U**	U neparni zadaci	U parni zadaci	U prva 3	U druga 3
1	1	1	1	1	0,5	1	**5,5**	2,5	3,0	3,0	2,5
2	1	1	1	1	1	1	**6,0**	3,0	3,0	3,0	3,0
3	1	0,5	0	1	0,5	0,5	**3,5**	1,5	2,0	1,5	2,0
4	1	1	0	1	0,5	1	**4,5**	1,5	3,0	2,0	2,5
5	1	1	1	1	0	1	**5,0**	2,0	3,0	3,0	2,0
6	1	1	0	1	0,5	1	**4,5**	1,5	3,0	2,0	2,5
7	1	0,5	1	1	1	1	**5,5**	3,0	2,5	2,5	3,0
8	1	0	0	1	0,5	0,5	**3,0**	1,5	1,5	1,0	2,0
9	1	1	1	1	0,5	1	**5,5**	2,5	3,0	3,0	2,5
10	1	1	1	1	0,5	0,5	**5,0**	2,5	2,5	3,0	2,0
11	1	1	1	1	0,5	1	**5,5**	2,5	3,0	3,0	2,5
12	1	1	1	1	1	1	**6,0**	3,0	3,0	3,0	3,0
M	1	0,8	0,7	1,0	0,6	0,9	**5,0**	2,3	2,7	2,5	2,5
St.dev	0	0,33	0,49	0,00	0,29	0,23	**0,94**	0,62	0,50	0,71	0,40

Koraci:

1. Podijelimo test na dva dijela i formiramo ukupne uratke na ta dva dijela. U ovom primjeru test je podijeljen dva puta: jednom u parne i neparne zadatke, a drugi put na prva tri i druga tri zadatka. U kolonama od 9 do 12 dati su rezultati na ovim polovinama.
2. Izračunamo korelaciju između dvije polovine. U našem slučaju
    a. Korelacija između parnih i neparnih zadataka: 0,40
    b. Korelacija između prve i druge polovine testa: 0,41

   Ove korelacije su ujedno i koeficijenti pouzdanosti za polovine testa
3. Ovu korelaciju uvrstimo u S-B formulu i dobijemo split-half koeficijent pouzdanosti. U našem slučaju oba koeficijenta iznose r_{xx}=0,58.

6.2.2.2 Cronbachov α (alfa) koeficijent pouzdanosti

Cronbach (1951) je definirao jedan od najvažnijih koeficijenata pouzdanosti u Klasičnoj testnoj teoriji. Split-half koeficijent pouzdanosti tretira dva dijela testa kao dvije paralelne forme, dok Cronbach ide dalje i koristi pristup na nivou pitanja, odnosno logiku konzistencije dva dijela testa spušta na nivo konzistencije između samih pitanja. Sada posmatramo svako pitanje kao jednu paralelnu mjeru i u jednom testu imamo onoliko mjera koliko ima pitanja. Povezanost između pitanja će biti mjera unutarnje konzistencije. Ovaj pristup se donekle razlikuje u svojoj primjeni ukoliko imamo 1) različite forme pitanja (binarne u odnosu na nebinarne zadatke), 2) različite pretpostavke o paralelnosti mjerenja (da li zaista imamo pitanja koja su paralelne mjere ili su to malo fleksibilnije shvaćene pretpostavke) i 3) različite ulazne podatke (da li su nam ulazni podaci za računanje varijance pitanja, kovarijančna matrica pitanja ili korelacijska matrica pitanja).

Prvi korak – određivanje koeficijenta α započinje određivanjem parametara zadataka i testa. Prvo odredimo varijance zadataka i varijancu testa. Pretpostavimo da smo imali set podataka ispitivanja studenata o kvaliteti 5 ispitnih pitanja na testu iz Psihometrije. Na osnovu njihovih rezultata odredili smo kovarijančnu matricu:

Tabela 6-4 Kovarijančna matrica pet pitanja

	Pitanje 1	**Pitanje 2**	**Pitanje 3**	**Pitanje 4**	**Pitanje 5**
Pitanje 1	**1,2**	0,4	0,5	0,6	0,3
Pitanje 2	0,4	**0,4**	0,2	0,2	0,2
Pitanje 3	0,5	0,2	**1,2**	0,6	0,3
Pitanje 4	0,6	0,2	0,6	**1,2**	0,4
Pitanje 5	0,3	0,2	0,3	0,4	**0,6**

Kao i u svakoj kovarijančnoj matrici, u glavnoj dijagonali nalaze se varijance pitanja, a u trokutovima kovarijance između pitanja. Na osnovu kovarijančne matrice možemo odrediti ukupnu varijancu testa tako što saberemo sve elemente matrice.

Dobit ćemo da je varijanca testa Vu=11,8.

Drugi korak – osnovna logika je da ukoliko su pitanja konzistentna to će njihove kovarijance biti veće, a samim tim i veća varijanca testa. Ukoliko pitanja nisu konzistentna onda su njihove kovarijance jako male i varijanca testa bit će gotovo pa ista kao ukupna varijanca zadataka. Na ovoj logici počiva Cronbachova formula:

6-8

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum V_i}{V_Y}\right]$$

Pri čemu je k – broj zadataka u testu, Vi varijance zadataka, a Vy varijanca testa.

U našem primjeru koeficijent α iznosit će:

$$\alpha = \frac{5}{5-1}\left[1 - \frac{1{,}2 + 0{,}4 + 1{,}2 + 1{,}2 + 0{,}6}{11{,}8}\right]$$

$$\alpha = 1{,}25 * [1 - 0{,}39]$$

$$\alpha = 1{,}25 * 0{,}61$$

$$\alpha = 0{,}49$$

Vidimo da naša mjera nije bila pouzdana jer je samo 49 % varijance objašnjeno pravim rezultatima, dok je 51 % varijance koju možemo pripisati pogreški.

Da još jednom ponovimo logiku koeficijenta α. Svako pitanje se tretira kao jedna mjera. Što su pitanja konzistentnija, to je njihova povezanost veća, odnosno njihove kovarijance se veće. Ako su njihove kovarijance veće, to će varijanca testa biti veća i više će se razlikovati od jednostavne sume varijanci pojedinih zadataka. Pri tome će količnik sume varijanci zadataka i varijance testa biti manji, a kada ga oduzmemo od 1 razlika će biti veća. Dio izraza na početku k/k-1 utječe na visinu koeficijenta pouzdanosti jer što imamo više zadataka to će i pouzdanost biti veća.

6.2.2.3 Standardizirani Cronbachov α koeficijent

U gornjem primjeru koeficijent α je dobiven na sirovim podacima. U slučaju da koristimo standardizirane rezultate na zadacima ili z-vrijednosti, tada dobivamo tzv. standardizirani koeficijent α. Postupak standardizacije rezultata na pitanjima, prije nego se krene u određivanje pouzdanosti, ima posebno smisla kada se varijance pitanja jako razlikuju. Ukoliko se varijance pitanja razlikuju, onda će neka pitanja imati više utjecaja na pouzdanost, stoga se preporučuje da se rezultati prvo standardiziraju, a potom da se njima operira. Nunnally i Bernstein (1994, str.232) ističu kako je *„teško precijeniti važnost ovog pristupa teoriji pogreške mjerenja"*. I sami kažu kako je određivanje standardizirane α jednostavnije i, za veliki broj ljudi, razumljivije. Kod „sirove" α kao ulazne podatke koristimo kovarijančnu matricu, dok u slučaju standardizirane α koristimo korelacije među pitanjima.

U prvom koraku potrebno je odrediti korelacije među svim parovima pitanja, potom odredimo prosječnu korelaciju $\overline{r_{ij}}$ među pitanjima.

U drugom koraku računamo koeficijent pouzdanost tako što prosječnu korelaciju uvrstimo u opći oblik Spearman – Brownove formule:

6-9

$$r_{xx} = \frac{k * \overline{r_{ij}}}{1 + (k-1)\overline{r_{ij}}}$$

što je u stvari formula za standardizirani α koeficijent.

Zamislimo da smo dobili korelacijsku matricu među pet pitanja na testu numeričkog faktora:

Tabela 6-5 Korelacijska matrica šest pitanja testa numeričkog faktora

	1	**2**	**3**	**4**	**5**
1	1				
2	0,56	1			
3	0,40	0,60	1		
4	0,47	0,30	0,46	1	
5	0,31	0,30	0,36	0,51	1

Prosječna korelacija za naših pet pitanja i deset pojedinačnih korelacija iznosi $\overline{r_{ij}} = 0{,}43$. Kada ovo uvrstimo u formulu dobit ćemo da je naš standardizirani koeficijent α:

$$r_{xx} = \frac{5 * 0{,}43}{1 + (5-1) * 0{,}43}$$

$$r_{xx} = \frac{5 * 0{,}43}{1 + (5 - 1) * 0{,}43}$$

$$r_{xx} = 0{,}79$$

6.2.2.3.1 Izračunavanje prosječne korelacije

Koeficijenti korelacije se ne mogu tek tako sabrati i podijeliti sa ukupnim brojem korelacija da bi se dobila prosječna vrijednost. Oni nisu aditivni jer su to indeksi i nisu izraženi na intervalnoj skali.

Tabela 6-6 Odnos koeficijenta korelacije i koeficijenta determinacije

Koeficijenti korelacije	Koeficijenti determinacije
-0,5	0,25
-0,3	0,09
-0,1	0,01
0,1	0,01
0,3	0,09
0,5	0,25

Iako se nama čini da je razlika između koeficijenta korelacije 0,1 i 0,3 u odnosu sa 0,3 i 0,5 jednaka, kada ove korelacije kvadriramo i pretvorimo u proporcije (koeficijente determinacije) vidimo da odnos između 0,01 i 0,09 nije isti kao i odnos 0,09 i 0,25. U prvom slučaju razlika je 0,08, a u drugom 0,14. Osim toga, što uraditi kada imamo i pozitivne i negativne korelacije (kao što je prikazano u tabeli), kolika će biti aritmetička sredina korelacija? U našem slučaju ona bi bila 0. Da li trebamo reći kako u prosjeku ove varijable ne koreliraju kada jasno vidimo da između njih ima povezanosti? Da bismo izračunali prosječnu korelaciju, koeficijente korelacije trebamo transformirati u intervalnu skalu. Za transformaciju korelacije koristimo Fischer-ovu z transformaciju, odnosno transformaciju korelacija u Fisherove z koeficijente (Fisher, 1915).

Napomena: ove z-koeficijente ne treba pomiješati sa z-vrijednostima, odnosno sa standardiziranim rezultatima.

Formula za određivanje Fisherovih z-koeficijenata je

6-10

$$z = \frac{\ln(1 + r) - \ln(1 - r)}{2}$$

Pri čemu je ln – prirodni logaritam. Ove transformacije možemo izvršiti pomoću kalkulatora ili uz pomoć formule koja se u Excel-u zove FISHER.

Nakon što smo odredili z-koeficijent svih korelacija pitanja, saberemo ih i odredimo prosječni z-koeficijent, koji onda transformiramo u korelaciju preko formule:

6-11

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Vidimo da sada imamo eksponencijalnu funkciju u kojoj je e – eksponencijalna konstanta i njena vrijednost je oko 2.718. U Excelu za povratak z-koeficijenta u koeficijent korelacije koristimo funkciju FISHERINV.

6.2.2.4 Kuder – Richardsonova formula KR20

Kuder i Richardson (1937) su pripremili varijantu obrasca za izračunavanje koeficijenta pouzdanosti kada u testu imamo binarne varijable pri čemu se tačan odgovor boduje sa 1 bod, a netačan sa 0. Ova formula glasi:

6-12

$$KR20 = \frac{k}{k-1}\left[1 - \frac{\sum(pq)_i}{V_Y}\right]$$

6.2.2.5 Kuder-Richardsonova formula KR21

.
Druga formula, koju su dali isti autori, odnosi se na određivanje koeficijenta pouzdanosti u situacijama kada su nam poznate samo aritmetička sredina testa i varijanca testa.

6-13

$$KR20 = \frac{k}{k-1}\left[1 - \frac{M_x\left(1 - \frac{M_x}{k}\right)}{V_X}\right]$$

6.2.3 Pouzdanost linearnih kombinacija

Kao što smo ranije vidjeli, linearne kombinacije ne moraju samo da budu pitanja u testu. Linearne kombinacije mogu da budu grupe subtestova namijenjenih mjerenju nekog agregiranog konstrukta. Na primjer, profesionalno sagorijevanje je kao konstrukt, prema Maslach i Jackson (1981), sastavljeno od tri komponente (emocionalne iscrpljenosti, depersonalizacije i osjećaja smanjenog postignuća). U upitniku su date tri subskale za koje možemo izračunati pouzdanost, ali nas zanima i kolika je pouzdanost čitave skale jer je to agregirani konstrukt.

Pouzdanost linearnih kombinacija zavisi od 1) pouzdanosti testova kao njenih dijelova i 2) korelacija među njima. Ovdje biste sebi mogli postaviti razumno pitanje: „Zašto jednostavno ne upotrijebimo Cronbachovu formulu?“ Neko bi možda došao u iskušenje da to uradi, ali bi nas rezultat naveo na pogrešne zaključke ukoliko se ne radi o testovima koji mjere istu latentnu varijablu. Pouzdanost uzoraka zadataka iz istog domena ovisi o prosječnoj korelaciji između njih (kao što smo vidjeli u računanju standardiziranog koeficijenta pouzdanosti), što nije istina za uzorke zadataka iz različitog domena. Pretpostavimo da želimo odrediti koeficijent pouzdanosti za 3 subskale nekog upitnika. Također pretpostavimo da svaka od subskala ima visok koeficijent pouzdanosti, ali su međusobno u nultim korelacijama. U ovom slučaju koeficijent α bi bio nula. Međutim, pogrešno bi bilo zaključiti da je pouzdanost takve linearne kombinacije nula. Najispravnije bi bilo da pouzdanost određujemo koristeći alternativne forme. U tom slučaju bi koeficijent korelacije između dvije alternativne forme bio koeficijent pouzdanosti za svaku od tih formi. Ali, ukoliko nije moguće administrirati dvije alternativne forme, tada možemo procijeniti pouzdanost linearne kombinacije na sljedeći način:

1. Znamo da je koeficijent pouzdanosti odnos između varijance pravih rezultata linearne kombinacije i varijance bruto rezultata te kombinacije
2. Varijanca bruto rezultata linearne kombinacije jednaka je zbiru varijanci elemenata i dvostrukoj sumi kovarijanci. Dakle, varijancu ove velike linearne kombinacije određujemo kao sumu složene matrice linearne kombinacije (pogledati matricu 5-5)
3. Varijanca pravih rezultata linearne kombinacije biće jednaka sumi varijanci pravih rezultata dijelova linearne kombinacije i dvostrukoj sumi kovarijanci elemenata. Znamo da se kovarijanca između rezultata javlja isključivo zbog pravih rezultata jer pogreška ne korelira ni sa čim, te je stoga lako doći do podatka o kovarijanci linearne kombinacije. Ostaje nam da se prisjetimo da je koeficijent pouzdanosti jednak omjeru varijance pravih rezultata i varijance bruto rezultata, te da varijancu pravih rezultata za elemente linearne kombinacije možemo izračunati tako da pomnožimo varijancu bruto rezultata sa koeficijentom pouzdanosti.
4. Nakon toga izračunamo varijancu pravih rezultata velike linearne kombinacije, koju podijelimo sa varijancom bruto rezultata i dobijemo koeficijent pouzdanosti.

Zamislimo da imamo jednu kovarijančnu matricu dva testa koja mjere dvije vrste profesionalnih interesa:

Tabela 6-7 Kovarijančna matrica subtestova profesionalnih interesa

	Sa21	*Sa22*	*Sa23*	*Sa24*	*Ea25*	*Ea27*	*Ea28*	*Ea30*
Sa21	**0,34**	0,09	0,11	0,08	0,13	0,1	0,13	0,1
Sa22	0,09	**0,3**	0,11	0,07	0,17	0,11	0,09	0,06
Sa23	0,11	0,11	**0,46**	0,21	0,19	0,13	0,13	0,1
Sa24	0,08	0,07	0,46	**0,65**	0,12	0,09	0,15	0,1
Ea25	0,13	0,17	0,21	0,12	**0,52**	0,2	0,11	0,16
Ea27	0,1	0,11	0,19	0,09	0,2	**0,47**	0,1	0,22
Ea28	0,13	0,09	0,13	0,15	0,11	0,1	**0,42**	0,1
Ea30	0,1	0,06	0,13	0,1	0,16	0,22	0,1	**0,48**

1. Bruto rezultati:
    - Varijanca bruto rezultata testa Sa (prve linearne kombinacije – zbir svih elemenata njene kovarijančne matrice) = 3,34
    - Varijanca bruto rezultata testa Ea (druge linearne kombinacije – zbir svih elemenata njene kovarijančne matrice) = 3,67

- Kovarijanca dvije linearne kombinacije (zbir elemenata kovarijančne matrice između dva testa) =1,9
- Ukupna varijanca ove dvije linearne kombinacije (zbir svih elemenata složene kovarijančne matrice) =10,92

2. Koeficijenti pouzdanosti testova Sa i Ea

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum V_i}{V_Y}\right]$$

$$\alpha_{Sa} = \frac{4}{3}\left[1 - \frac{0,34 + 0,3 + 0,46 + 0,65}{3,34}\right] = 0,63$$
$$\alpha_{Ea} = \frac{4}{3}\left[1 - \frac{0,52 + 0,47 + 0,42 + 0,48}{3,67}\right] = 0,65$$

3. Varijance pravih rezultata testova Sa i Ea

$$V_t = r_{xx}V_b$$

$$V_{Sa} = 0,63 * 3,34 = 2,12$$

$$V_{Ea} = 0,65 * 3,67 = 2,37$$

4. Varijanca pravih rezultata složene linearne kombinacije

$$V_{tslk} = V_{tSa} + V_{tEa} + 2COV_{SaEa}$$

$$V_{tslk} = 2,12 + 2,37 + 2 * 1,9 = 8,29$$

5. Pouzdanost složene linearne kombinacije

$$r_{xx} = \frac{V_t}{V_b} = \frac{8,29}{10,92} = 0,76$$

Drugi način bi bio preko formule (Nunnally i Bernstein, 1994, str. 268):

6-14

$$r_{YY} = 1 - \frac{\sum \sigma_i^2 - \sum r_{ii}\sigma_i^2}{\sigma_Y^2}$$

Pri čemu je

σ_Y^2 – varijanca linearne kombinacije testova

σ_i^2 – varijanca svakog testa

6.2.4 Određivanje pouzdanosti analizom varijance

Vidjeli smo da je pouzdanost određena omjerom varijanci pravih i bruto rezultata. Dakle, svaki metod pomoću kojeg možemo procijeniti varijancu pravih rezultata može biti korišten za ocjenu pouzdanosti (Nunnally i Bernstein, 1994, str. 274). Hoyt (1941) je u svom članku dao pregled kako se može koristiti analiza varijance za određivanje koeficijenta pouzdanosti. Ovaj pristup ćemo kasnije spomenuti i u poglavlju o teoriji generalizabilnosti.

Analiza varijance se podučava kao metoda određivanja razlika među grupama. Međutim, ANOVA nije ništa drugo nego specijalni slučaj multiple regresijske analize u kojoj kao prediktore koristimo kategoričke varijable pri čemu svaki ispitanik spada u neku određenu grupu. Rezultat ispitanika X_{ij} definiran je kao linearna kombinacija tri elementa: 1) univerzalnog efekta koji se javlja kod svih ispitanika, 2) efekta tretmana specifičnog za grupu kojem ispitanik pripada i 3)slučajne pogreške jedinstvene za svako mjerenje.

1. Univerzalni efekat ili konstanta predstavlja prosječnu vrijednost ili aritmetičku sredinu svih mjera i označava se sa **X..** Njena vrijednost **μ** zavisi od toga kako su mjere skalirane i ne mora biti od posebnog značaja.
2. Efekat tretmana predstavlja odstupanje rezultata grupe od ukupne prosječne vrijednosti čitave grupe. Prosječna vrijednost čitave grupe označava se sa **X.j** a njena visina sa μ_j. Ova razlika između vrijednosti **X.j - X..** nam je najznačajnija jer upravo analizom varijance testiramo njenu značajnost.
3. Slučajna greška ili e_{ij} predstavlja variranje rezultata među ispitanicima koji pripadaju istoj grupi. Prema Klasičnoj testnoj teoriji, ona je slučajna i njena visina se procjenjuje kao razlika između individualnog rezultata Xij i grupne aritmetičke sredine X.j

Kada oduzmemo X.. od pojedinačnih rezultata Xij dobijemo devijacijske rezultate. Kvadriranjem devijacijskih rezultata, te njihovim sumiranjem dobivamo Ukupnu sumu kvadrata, koju, kada podijelimo sa ukupnim brojem mjera minus 1 i dobijemo vrijednost totalnog prosječnog kvadrata (MStot). Ovaj totalni prosječni kvadrat se može pokazati kao jednak prosječnoj sumi kvadrata između grupa ili prosječnoj sumi kvadrata unutar grupa. Utvrđivanje razlike se odnosi na utvrđivanje da li je varijanca između grupa jednaka varijanci unutar grupa ili je ona, zbog tretmana ili intervencije značajno povećana.

U određivanju pouzdanosti, imamo ispitanika koji odgovara na više pitanja (pitanja sada odgovaraju onome što smo u ANOVI zvali „tretman“). Odgovori ispitanika dati su u matrici bruto rezultata. Ovi podaci predstavljaju ponovljena mjerenja i traže od nas da primijenimo dizajn analize varijance za zavisne uzorke. I dok jednostavna analiza varijance ima samo tri varijabiliteta (ukupno variranje, između grupa i unutar grupa), u slučaju našeg dizajna imamo četiri: 1) ukupni varijabilitet, 2) varijabilitet među ispitanicima, 3) varijabilitet među pitanjima i 4) rezidualni varijabilitet koji nastaje kao interakcija između pitanja i ispitanika. Ovaj rezidual odražava činjenicu da odgovor ispitanika na neko pitanje (X_{ij}) ne može biti objašnjeno niti karakteristikama ispitanika niti karakteristikama pitanja. U skladu sa osnovnim principom pouzdanosti, a to je da pouzdanost zavisi od varijabiliteta među ispitanicima, mi hoćemo da nam varijabilitet među ispitanicima bude što veći, dok rezidualni varijabilitet treba da bude što manji (jer je neobjašnjiv, odnosno on je pogreška).

U tabeli 6-8 su dati svi podaci koje računamo u analizi varijance:

Tabela 6-8 Izvori varijabiliteta u analizi varijance

Izvor varijance	**Suma kvadrata**	**Stepeni slobode**	**Što sadržava prosječni kvadrat kod ponovljenih mjerenja**
Ispitanici (i)	$j\Sigma\,(X_{ij} – X_{i.})^2$	i-1	$jV_{isp} + V_{rezidual}$
Suci, tretman, zadaci (j)	$i\,\Sigma(X_{ij} – X_{.j})^2$	j-1	$jV_{zad} + V_{rezidual}$
Rezidual	$\Sigma\,(X_{ij} – X_{i.}\text{-}X_{.j}\text{-}X_{..})^2$	(i-1) (j-1)	$V_{rezidual}$
Total	$\Sigma(Xi_{ij}– X_{..})^2$	ij-1	

i – ispitanici, j – zadaci, pitanja ili procjenjivači

Koeficijent pouzdanosti se računa kao odnos između razlike prosječnog kvadrata individua i reziduala s prosječnim kvadratom individua:

6-15

$$r_{11} = \frac{MS_{isp} - MS_{rez}}{MS_{isp}}$$

Primjer: zamislimo da je 12 ispitanika odgovaralo 4 četiri pitanja:

Tabela 6-9 Matrica bruto rezultata ispitanika za analizu varijance

Ispitanik	**Zadatak 1**	**Zadatak 2**	**Zadatak 3**	**Zadatak 4**	$X_{i.}$	$X_{i.}^2$
1	1	1	1	1	4	16
2	1	1	1	1	4	16
3	1	0	0	0	1	1
4	1	1	0	1	3	9
5	0	1	1	1	3	9
6	0	1	0	0	1	1
7	0	0	1	1	2	4
8	0	0	0	1	1	1
9	1	1	1	1	4	16
10	0	1	1	1	3	9
11	0	1	1	1	3	9
12	1	1	1	1	4	16
$X_{.j}$	6	9	8	10		
$X_{.j}^2$	36	81	64	100		

Dobivena je kovarijančna matrica za ove rezultate:

Tabela 6-10 Kovarijančna matrica 4 zadatka za analizu varijance

	Zadatak 1	**Zadatak 2**	**Zadatak 3**	**Zadatak 4**
Zadatak 1	0,273	0,045	0	0
Zadatak 2	0,045	0,205	0,091	0,045
Zadatak 3	0,000	0,091	0,242	0,121
Zadatak 4	0,000	0,045	0,121	0,152

Prema formuli za određivanje koeficijenta α, pouzdanost ovog testa iznosila bi

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum V_i}{V_Y}\right]$$

$$\alpha = \frac{4}{3}\left[1 - \frac{0{,}871}{1{,}477}\right]$$

$$\alpha = \frac{4}{3}\left[1 - \frac{0{,}871}{1{,}477}\right] = 0{,}56$$

Računanje koeficijenta pouzdanosti analizom varijance uključuje sljedeće korake (s operativnim formulama):

1. **Određivanje totalne sume kvadrata**:

6-16

$$SK_{tot} = \sum X_{ij}^2 - \frac{\left(\sum X_{ij}\right)^2}{N}$$

Pri čemu je

Xij – rezultat ispitanika na svakom pojedinačnom mjerenju. U našem primjeru imamo ukupno 48 rezultata.

N – ukupan broj mjerenja, u našem slučaju to je 48 mjerenja

Malo podsjećanje na osnove matematike: $\sum X_{ij}^2$ znači a da svaki rezultat kvadriramo, pa onda saberemo, dok $\left(\sum X_{ij}\right)^2$ znači da prvo sve rezultate saberemo, a onda taj zbir kvadriramo.

$$SK_{tot} = 33 - \frac{1089}{48} = 10$$

2. **Određivanje sume kvadrata ispitanika**

6-17

$$SK_{isp} = \frac{\sum X_{i.}^2}{j} - \frac{\left(\sum X_{ij}\right)^2}{N}$$

Pri čemu je:

$\sum X_{i.}^2$ – kvadrirana suma rezultata za svakog ispitanika (zadnja kolona tabele rezultata ispitanika)

j – broj zadataka

$$SK_{isp} = \frac{107}{4} - \frac{1089}{48} = 4{,}06$$

3. **Određivanje sume kvadrata po zadacima**

6-18

$$SK_{zad} = \frac{\sum X_{.j}^2}{i} - \frac{\left(\sum X_{ij}\right)^2}{N}$$

Pri čemu je:

$\sum X_{.j}^2$ – suma kvadriranih ukupnih rezultata po zadacima

i – broj ispitanika

$$SK_{zad} = \frac{281}{12} - \frac{1089}{48} = 0{,}72$$

4. Određivanje sume kvadrata reziduala

6-19

$$SK_{rez} = SK_{tot} - SK_{isp} - SK_{zad}$$
$$SK_{rez} = 10 - 4{,}06 - 0{,}72 = 5{,}22$$

Tabela 6-11 Izvori varijance u određivanju koeficijenta pouzdanosti na primjeru

Izvor varijance	**Suma kvadrata**	**Stepeni slobode**	**MS (prosječni kvadrat)**
Ispitanici (i)	4,06	i-1	0,37
Suci, tretman, zadaci (j)	0,72	j-1	0,24
Rezidual	5,22	(i-1) (j-1)	0,15
Total	10	ij-1	

$$r_{11} = \frac{MS_{isp} - MS_{rez}}{MS_{isp}} = \frac{0,37-0,15}{0,37} = 0,57$$

Vidimo da imamo malu razliku u ova dva koeficijenta (α i koeficijenta pouzdanosti po Hoytovoj metodi) što je rezultat zaokruživanja brojeva.

Na ovom primjeru naučili smo kako određujemo pouzdanost preko analize varijance i na taj način stvorili podlogu za priču o Teoriji generalizabilnosti.

6.3 O ČEMU OVISI POUZDANOST TESTOVA?

Vidjeli smo da pouzdanost testova ovisi o mnogo faktora. U ovoj sekciji ćemo sistematizirati sve faktore i vidjeti kako možemo popravljati pouzdanost.

6.3.1 Homogenost uzorka i pouzdanost

Koeficijent pouzdanosti ovisi o heterogenosti uzorka prema mjerenoj varijabli. Ako se prisjetimo da je pouzdanost odnos između varijance pravih rezultata i varijance bruto rezultata, onda heterogenost uzorka, koja povećava varijancu pravih rezultata, povećava i pouzdanost testa. U određivanju metrijskih karakteristika testa želimo imati što veću heterogenost uzorka prema mjerenoj varijabli. Ovo nas upućuje na jednu važnu implikaciju, a to je da je pouzdanost metrijska karakteristika rezultata testa, a ne samog testa. To znači da u jednom vrlo heterogenom uzorku možemo dobiti visoku pouzdanost, a u drugom uzorku mnogo nižu pouzdanost. Situacija da imamo homogeniziran uzorak u testiranju nije rijetka u psihologiji. Na primjer, neki test ličnosti čije su norme pripremljene na velikom uzorku, u populaciji kandidata za upis na studij može imati značajno smanjenu pouzdanost. Situaciju kada test primjenjujemo na homogeniziranom uzorku ili ograničenom obimu uzorka moramo uzeti u obzir prilikom interpretacije koeficijenta pouzdanosti.

Evo jednog primjera. Koristili smo test inteligencije na uzorku studenata Elektrotehničkog fakulteta. Dobili smo sljedeće rezultate:

Tabela 6-12 Varijabilitet u homogeniziranom uzorku

	Originalni normativni uzorak	**Selekcionirani uzorak**
Varijanca bruto rezultata	225	135
Varijanca pogreške	22,5	29,7
Varijanca pravih rezultata	202,5	105,3
Koeficijent pouzdanosti	0,90	0,78

Istraživači su se zapitali kako je moguće da su dobili ovako nisku pouzdanost, te im je onda psihometričar objasnio da je to zbog ograničenosti obima uzorka.

Kako bismo provjerili koliki bi bio koeficijent pouzdanosti testa primijenjenog na homogenom uzorku, odnosno na uzorku ograničenog obima, koristimo formulu koju zovemo popravak za omeđenost obima:

6-20

$$r_c = \frac{r_{oo} * \frac{s_n}{s_o}}{\sqrt{1 - r_{oo}^2 + r_{oo}^2 * \frac{V_n}{V_o}}}$$

Pri čemu je

r_c – korigirani koeficijent pouzdanosti

r_{oo} – koeficijent pouzdanosti dobiven na omeđenom uzorku

s_n – standardna devijacija rezultata na neomeđenom uzorku

s_o – standardna devijacija rezultata dobivenih na omeđenom uzorku

Trebamo primijetiti da je ova korekcija moguća samo onda kada imamo sve parametre na raspolaganju.

6.3.2 Broj pitanja i pouzdanost

Kada nekoga pitamo jedno pitanje, on nam da jedan odgovor. Ali, ako ga tri puta pitamo i tri puta dobijemo isti odgovor, onda smo sigurniji da smo nešto precizno izmjerili. Što imamo više pitanja to će test biti pouzdaniji. Dodavanjem pitanja mi povećavamo varijancu pravih rezultata mnogo više nego varijancu pogreške. Kada imamo samo jedno mjerenje i dodamo još jedno mjerenje, mi smo povećali broj mjerenja za dva puta. Što će se desiti s varijancom pravih rezultata?

$$Y = X1 + X2$$

Varijanca pravog rezultata sada, u ovoj linearnoj kombinaciji bit će jednaka:

$$V_{t-duplo} = V_{t1} + V_{t2} + 2r_{t1t2}s_{t1}s_{t2}$$

Znamo da je korelacija između pravih rezultata paralelnih mjera jednaka 1 i varijance su jednake, te dobivamo da je:

$$V_{t-duplo} = 2 * V_t + 2 * V_t$$

$$V_{t-duplo} = 4 * V_{t\ jednog\ mjerenja}$$

Varijanca pravih rezultata, kada jednom mjerenju dodamo još samo jedno mjerenje, bit će povećana 4 puta. Što će biti s pogreškom? Istom ovom logikom, ako jednom mjerenju dodamo još samo jedno mjerenje varijanca pogreške će biti:

$$V_{e-duplo} = V_{e1} + V_{e2} + 2r_{e1e2}s_{e1}s_{e2}$$

Pošto znamo da je pogreška slučajna i ne korelira s bilo čim, to će dio izraza $2r_{e1e2}s_{e1}s_{e2}$ biti jednak 0, a varijanca pogreške će biti:

$$V_{e-duplo} = 2 * V_{e\ jednog\ mjerenja}$$

Vidimo da varijanca pravih rezultata raste mnogo brže nego varijanca pogreške, a posljedično raste i koeficijent pouzdanosti. Upravo su na ovoj logici Spearman i Brown pripremili formulu kojom provjeravamo odnos između dužine testa i koeficijenta pouzdanosti i koja glasi:

6-21

$$\alpha_{k\prime} = \frac{n * \alpha_k}{1 + (n-1)\alpha_k}$$

Pri čemu je:

$\alpha_{k\prime}$ – procjena koeficijenta pouzdanosti u slučaju promjene dužine testa

α_k – koeficijent pouzdanosti dobiven u istraživanju

n – omjer između broja zadataka u izmijenjenoj verziji testa i broja zadataka u originalnoj verziji testa

$$n = \frac{broj\ zadataka\ u\ novoj\ verziji\ testa}{broj\ zadataka\ u\ orginalnoj\ verziji\ testa}$$

Na primjer, dobili smo da test koji sadrži 20 zadataka ima koeficijent pouzdanosti od 0,7. Zanima nas koliko bi se promijenila pouzdanost kada bismo testu dodali još 10 zadataka?

$$n = \frac{20 + 10}{20} = \frac{30}{20} = 1{,}5$$

$$\alpha_{k\prime} = \frac{1{,}5 * 0{,}7}{1 + (1{,}5 - 1) * 0{,}7}$$

$$\alpha_{k\prime} = \frac{1{,}05}{1{,}35}$$

$$\alpha_{k\prime} = 0{,}78$$

Vidimo da bi se pouzdanost povećala na 0,78.

Druga verzija ove formule nam omogućava da računamo koliko treba da produžimo test – da promijenimo dužinu testa da bismo dobili željeni koeficijent pouzdanosti. Ta formula je:

6-22

$$k' = k * \frac{\alpha_{k\prime}(1 - \alpha_k)}{\alpha_k(1 - \alpha_{k\prime})}$$

Na primjer, imamo test od 20 pitanja koji ima pouzdanost od 0,7. Mi želimo da nam pouzdanost bude 0,78. Koliko zadataka trebamo dodati?

$$k' = 20 * \frac{0{,}78 * (1 - 0{,}7)}{0{,}7 * (1 - 0{,}78)}$$

$$k' = 20 * \frac{0{,}78 * (1 - 0{,}7)}{0{,}7 * (1 - 0{,}78)}$$

$$k' = 20 * \frac{0{,}234}{0{,}154} = 30$$

Dakle, nova forma treba da ima 30 zadataka, odnosno potrebno je dodati još 10 zadataka.

Možemo zaključiti da, ukoliko sve drugo ostaje isto, produžavanjem testa, odnosno dodavanjem zadataka, povećavamo pouzdanost testa. Međutim, ovdje moramo biti oprezni. Prvo, dodani zadaci moraju biti konzistentni s ostalim pitanjima. Drugo, ne možemo test produžavati do unedogled. Moramo voditi računa da ispitanici gube koncentraciju i da nakon što vide kako odgovaraju na prva pitanja dolazi do udešavanja u odgovaranju. Oni prestanu misliti o pitanjima i samo zaokružuju odgovore koje vide da preferiraju. Stoga je važno razmotriti odnos kvalitete testiranja kao procesa i kvalitete prikupljenih podataka naspram povećavanja dužine testa. Trebamo znati da je odnos između dužine testa i koeficijenta pouzdanosti objašnjen krivuljom negativne akceleracije. To znači da na početku, kada imamo kratke testove, dodavanje pitanja značajno povećava koeficijent pouzdanosti, ali kada je test već prilično dug, porast koeficijenta je mnogo manji i nikada ne dosegne jedan. O svemu ovome treba voditi računa jer kriterij dobrog testa nije samo pouzdanost, nego i njegova ekonomičnost i praktičnost upotrebe.

6.3.3 Pogreška mjerenja – korelacija – pouzdanost

Kada bismo mogli mjeriti prave rezultate, mi bismo bili mnogo sretni. Međutim, u praksi to nikada nije slučaj. Rezultati u sebi uvijek sadrže komponentu pogreške. Ova pogreška utječe na visinu rezultata, ali i na korelacije među rezultatima u dvije mjere.

Pretpostavimo da imamo 12 ispitanika koji su radili dva paralelna testa. Njihovi pravi rezultati na ova dva testa su identični i prikazani u prvim kolonama. Ono što smo mi zabilježili kao njihove bruto rezultate dato je u drugim kolonama:

Tabela 6-13 Pravi i bruto rezultati grupe ispitanika

Isp	Pravi rezultat prvo mjerenje	Pravi rezultat drugo mjerenje	Bruto rezultat prvo mjerenje	Bruto rezultat drugo mjerenje
1	1	1	1	1
2	2	2	1	2
3	3	3	3	1
4	4	4	2	4
5	5	5	5	3
6	6	6	6	6
7	7	7	5	6
8	8	8	8	8
9	9	9	7	8
10	10	10	10	8
11	11	11	12	11
12	12	12	12	10

Korelacija između pravih rezultata je 1. Međutim, kada izračunamo korelaciju između bruto rezultata zaključimo da je ona niža – 0,93. Pogreška mjerenja utječe na korelaciju tako da je umanjuje ili atenuira. Na visinu korelacije između dvije varijable utječu:

1. Prirodna povezanost između te dvije varijable
2. Pouzdanost prvog testa
3. Pouzdanost drugog testa

Nas, u svakom slučaju, može zanimati kolika je korelacija između pravh rezultata ili kolika bi bila korelacija među mjerama kada bismo povećali pouzdanost? Ove korigirane korelacije zovemo disatenuirane, a formulu za određivanje korelacije idealnih mjera zovemo korekcija za atenuaciju.

Osnovni obrazac za disatenuaciju glasi:

6-23

$$r_{X_tY_t} = \frac{r_{XY}}{\sqrt{r_{xx}r_{yy}}}$$

Pri čemu je:

$r_{X_tY_t}$ – disatenuirana korelacija između mjere X i Y

r_{XY} – opažena korelacija

r_{xx} – koeficijent pouzdanosti mjere X

r_{yy} – koeficijent pouzdanosti mjere Y

U svom općem obliku, korekcija za atenuaciju nam omogućava da provjeravamo koliko bi se promijenio koeficijent korelacije ukoliko mijenjamo koeficijente pouzdanosti testova:

6-24

$$r_{X_tY_t} = \frac{r_{XY}\sqrt{r_{xx}'r_{yy}'}}{\sqrt{r_{xx}r_{yy}}}$$

Pri čemu su dodana još dva elementa:

r_{xx}' – promijenjeni koeficijent pouzdanosti za mjeru X

r_{yy}' – promijenjeni koeficijent pouzdanosti za mjeru Y

6.3.4 Konzistencija pitanja i pouzdanost

Pitanja u testu trebaju biti konzistentna, a njihovu konzistenciju procjenjujemo preko korelacija između pitanja ili inter-item korelacija. Očekujemo da sva pitanja, ako su konzistentna, pozitivno koreliraju. Pogledajmo jedan primjer izračunavanja koeficijenta pouzdanosti u SPSS-u na Upitniku depresije. Tabele koje slijede kopirane su iz SPSS-ovog izvještaja (outputa) i njihov format je zadržan u originalnom obliku kako biste ih mogli prepoznati kada budete radili obrade podataka u SPSS-u na računaru.

Podaci koje smo dobili su sljedeći:

Tabela 6-14 SPSS Broj ispitanika u obradi

Case Processing Summary

		N	%
Cases	Valid	1921	93.1
	Excluded[a]	143	6.9
	Total	2064	100.0

U ovoj tabeli vidimo da je bio ukupno 1921 ispravno unesen podatak, te da su 143 ispitanika isključena iz analize. U narednoj tabeli dobili smo podatak o visini pouzdanosti α:

Tabela 6-15 SPSS tabela s koeficijentom pouzdanosti

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.503	.525	5

Vidimo da je α na sirovim rezultatima 0,503, standardizirana α 0,525 i broj pitanja uključenih u analizu 5. Dobiveni koeficijent pouzdanosti je nizak. Dolazimo do dijela izvještaja koji nam daje više detalja za interpretaciju koeficijenta. Tabela 6-15 je inter-item korelacijska matrica:

Tabela 6-16 SPSS tabela inter-item korelacija

Inter-Item Correlation Matrix

	Pitanje 1	Pitanje 2	Pitanje 3	Pitanje 4	Pitanje 5
Pitanje 1	1.000	.045	.534	.375	.045
Pitanje 2	.045	1.000	.059	.059	.278
Pitanje 3	.534	.059	1.000	.405	.012
Pitanje 4	.375	.059	.405	1.000	-.006
Pitanje 5	.045	.278	.012	-.006	1.000

U matrici vidimo da je pitanje 2 u veoma niskim korelacijama s ostalim pitanjima, a pitanja 3 i 5, te 4 i 5 su u gotovo nultim korelacijama. Ovaj nalaz nam ukazuje da pitanja u skali nisu konzistentna. Najbolje korelacije su između pitanja 1, 3 i 4. Da pogledamo sadržaj pitanja:

1. osjećam potištenost i neraspoloženje
2. najbolje se osjećam ujutro rec
3. imam razdoblja plača ili osjećaja kao da ću zaplakati
4. imam poteškoća sa spavanjem tokom cijele noći
5. jedem uobičajeno kao i ranije rec

Vidimo da su pitanja 2 i 5 obrnuto bodovana pitanja. Ova pitanja su formulirana u suprotnom smjeru od ostalih pitanja. U pitanjima 1, 3 i 4 viši rezultat podrazumijeva veću izraženost simptoma, dok kod pitanja 2 i 5 viši rezultat znači manje izraženu depresiju. Iako su ova pitanja rekodirana (ukoliko je ispitanik odgovorio sa 1 dobije 5 bodova) i dalje nisu konzistentna s ostatkom upitnika. Ovo je jedan specifičan problem koji traži opreznost u konstrukciji instrumenata i korištenju obrnuto bodovanih pitanja. Ona često nisu konzistentna s ostatkom upitnika. Dodatnu informaciju o tome što uraditi s nekonzistentnim pitanjima dobivamo iz sljedeće tabele:

Tabela 6-17 SPSS Korelacije pitanja sa skalom

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
Pitanje 1	9.51	5.377	.395	.317	.376
Pitanje 2	8.60	5.735	.185	.082	.511
Pitanje 3	9.66	5.315	.395	.335	.374
Pitanje 4	9.61	5.546	.317	.202	.422
Pitanje 5	8.22	5.894	.138	.080	.544

U ovoj tabeli nam je najzanimljivija zadnja kolona koja govori o tome što će se desiti s koeficijentom pouzdanosti ukoliko neko pitanje izbacimo. Vidimo da, ukoliko izbacimo pitanja 2 i 5, koeficijent pouzdanosti će porasti, a u slučaju izbacivanja pitanja 1, 3 ili 4, on će opasti. Koristeći se ovom statistikom, možemo donositi odluke koja pitanja ima smisla ostavljati u upitnicima, a koja ne. Ipak, kod odlučivanja o izbacivanju pitanja, moramo voditi računa o sadržaju pitanja jer neka pitanja ne moraju biti visoko konzistentna s drugim, mogu biti takva da ih rijetki rješavaju ili zaokružuju, ali da im je i dalje mjesto u upitniku. Jedan primjer je pitanje u upitnicima depresije koje se odnosi na suicidalne misli, osjećanja ili ponašanja. Ova pitanja tipično nisu visoko povezana s drugim pitanjima niti ih ispitanici često zaokružuju, ali za dijagnostiku i intervencije ovo pitanje ima ključnu važnost. Moramo balansirati između konzistencije pitanja i sadržajne heterogenosti pitanja. Ranije smo govorili o homogenosti i heterogenosti pitanja kada smo započeli priču o pouzdanosti kao unutarnjoj konzistenciji. Ali, sada, kada smo naučili što je unutarnja konzistencija i koji faktori utječu na pouzdanost, vrijeme je da unaprijedimo priču o homogenosti i heterogenosti pitanja. Prilikom konstrukcije testa moramo sadržajno pokriti različite domene (situacije ili ponašanja) u kojima se neka karakteristika ispoljava. To znači da moramo dopustiti određeni nivo heterogenosti pitanja kako bismo zahvatili sam konstrukt. Ovaj zadatak je utoliko zahtjevniji što je varijabla varijabilnija i što se ispoljava u više konteksta i vremenskih okvira. Sadržaj pitanja se nalazi pod kontrolom psihologa koji razvija test i njegov zadatak je da ocijeni koliko heterogen treba biti sadržaj pitanja, pri čemu odluku treba bazirati na literaturi i poznavanju karakteristike za koju priprema test, te namjene testa i populacije za koju se test priprema. Kako Urbina naglašava: *„Stepen u kojem je test namjerno dizajniran da sadrži heterogena pitanja ne može biti tretiran kao izvor pogreške. Heterogenost testa ili kognitivnih funkcija koje mjeri test i koje su potrebne za uradak na testu postaju izvor pogreške samo onda kada je namjera bila da test bude homogen kroz sva ili većinu pitanja"* (2004, str.130).

Urbina u istom poglavlju daje sjajan pregled relativnosti homogenosti i heterogenosti u testovima. Pitanje homogenosti i heterogenosti odnosi se na dva elementa: uzorke ponašanja korištene u pitanjima i grupe ispitanika. Oba elementa utječu na statističke parametre kojima se provjeravaju metrijske karakteristike testa (npr. ograničenost obima, inter-item korelacije, atenuiranje korelacije...) te je važno imati na umu:

1. Heterogenost i homogenost su uvijek u psihologiji relativni pojmovi. Dva su pitanja, samom činjenicom da su drukčije formulirana, heterogena. Dva čovjeka su uvijek heterogena jer nisu identični.
2. Definiranje pitanja ili ispitanika kao homogene ili heterogene grupe uvijek mora biti urađeno na osnovu varijable ili varijabli koje služe za ustanovljavanje sličnosti i razlika. Na primjer, dva pitanja mogu biti homogena jer mjere isti profesionalni interes, ali su heterogena u odnosu na sadržaj (na primjer pitanja: Volim čitati znanstvene članke i Volim izvoditi znanstvene eksperimente). Isto tako, grupa ljudi može biti heterogena prema spolu ili prema mjestu stanovanja, ali homogena prema nekoj drugoj varijabli (mogu biti sve medicinski tehničari na klinici).

6.3.5 Kako pouzdanost primjenjujemo u praksi – regresijski model pouzdanosti

U praksi, psihološka testiranja koristimo da donesemo važne odluke za živote ljudi. Psiholozi procjenjuju sposobnosti djeteta da prati standardni nastavni program u školama. U zavisnosti od rezultata procjene može se donijeti odluka da dijete zbog umanjenih sposobnosti pohađa školu prema prilagođenom nastavnom programu. Danas se, u velikom broju zemalja, koriste testovi znanja na kraju srednje škole. Na osnovu rezultata formiraju se liste upisa na fakultete (eksterna matura). Klinički psiholozi na osnovu testovnih rezultata pišu nalaz i daju preporuke za pacijente. Sve su ovo vrlo važne odluke. S druge strane, znamo da su rezultati testova uvijek opterećeni pogreškom, tako da nas interesiraju odgovori na pitanja: „Ako je ispitanik osvojio toliko poena, u kojem rasponu se nalazi njegov pravi rezultat?“ ili „Ako je ovo rezultat ispitanika na jednom mjerenju, u kojem rasponu bi bili njegovi rezultati na ponovljenim mjerenjima?“ Mi smo svjesni činjenice da na raspolaganju nemamo veliki broj mjerenja (štaviše, obično imamo samo jednu priliku za mjerenje), a ipak moramo biti sigurni u svoje odluke. Zato su nam predviđanja raspona rezultata i donošenje odluka na osnovu intervala pouzdanosti jako važna.

Podsjetimo se još jednom, mi ne možemo registrirati ispitanikov pravi rezultat, ali koristimo testove da bismo najbolje procijenili stvarni nivo izraženosti karakteristike. Pri tome određujemo pouzdanost psiholoških testova kako bismo bili sigurniji u kvalitetu dobivenih rezultata. Dvije osnovne informacije potrebne za donošenje odluke su:

1. Koja je najbolja procjena ispitanikovog pravog rezultata?
2. Koji je interval pouzdanosti u kojem se nalazi ispitanikov pravi ili budući bruto rezultat?

Iz ova dva pitanja možete zaključiti da nam je potrebno razumijevanje regresijskih modela predviđanja rezultata.

6.3.5.1 Procjena pravog rezultata ispitanika

Iz linearne regresije znamo da možemo procijeniti devijacijski rezultat jednog testa na osnovu poznavanja korelacije između tog testa i drugog testa kao prediktora. Ovu procjenu radimo prema formuli:

$$y = \frac{s_y}{s_x} * x * r_{xy}$$

Kada ovo prevedemo u termine pravih i bruto rezultata gdje mi na osnovu bruto rezultata pokušavamo predvidjeti pravi rezultat, ova formula poprima sljedeći oblik:

6-25

$$t = \frac{s_t}{s_b} * x * r_{bt}$$

Mi znamo da je $s_t = s_b * \sqrt{r_{xx}}$ (pogledati u dijelu 6.1.2). Znamo i da je $\sqrt{r_{xx}} = r_{bt}$, odnosno da je drugi korijen koeficijenta pouzdanosti korelacija između bruto i pravih rezultata i da tu korelaciju zovemo indeks pouzdanosti. Kada ove poznate elemente uvrstimo u formulu, dobijemo:

$$t = \frac{s_b * \sqrt{r_{xx}}}{s_b} * x * r_{bt}$$

$$t = \sqrt{r_{xx}} * x * \sqrt{r_{xx}}$$

6-26

$$t = r_{xx} * x$$

Dakle, pravi devijacijski ili otklonski rezultat jednak je umnošku koeficijenta pouzdanosti i bruto otklonskog rezultata.

Pravi rezultat procjenjujemo na osnovu bruto rezultata koristeći formulu:

$$\bar{X}_t = r_{xx}(X_b - M) + M$$

pri čemu je $X_b - M = x$, odnosno devijacijski bruto rezultat ispitanika od aritmetičke sredine.

$$\bar{X}_t = r_{xx} * x + M$$

Na primjer, ukoliko je ispitanik na testu, čija je aritmetička sredina 50, a koeficijent pouzdanosti 0,80, postigao rezultat 58, njegov procijenjeni pravi rezultat će biti:

$$\bar{X}_t = 0{,}80 * (58 - 50) + 50 = 56{,}4 \approx 56$$

U procjeni pravog rezultata na osnovu bruto rezultata primjećujemo efekat koji zovemo regresija prema sredini. To je efekat koji se odnosi na vjerovatnoću da će drugi rezultat ispitanika biti bliži aritmetičkoj sredini grupe nego prvi rezultat. Zbog ovog efekta, kada je ispitanikov bruto rezultat veći od aritmetičke sredine, njegov predviđeni pravi rezultat je uvijek manji od bruto rezultata, a kada je ispitanikov bruto rezultat manji od aritmetičke sredine, njegov predviđeni pravi rezultat je uvijek veći od bruto rezultata ili bliži aritmetičkoj sredini.

Iz formule za procjenu pravog rezultata i poznajući efekat regresije prema sredini, možemo reći da visina procijenjenog pravog rezultata ovisi o:

1. pouzdanosti testa
2. udaljenosti bruto rezultata od aritmetičke sredine testa
3. smjeru razlike bruto rezultata od aritmetičke sredine testa

6.3.5.2 Predviđanje bruto rezultata na osnovu poznavanja pravog rezultata

Kada smo naučili kako procijenimo vrijednost pravog rezultata na osnovu bruto rezultata, da vidimo kako možemo postaviti interval pouzdanosti u kojem će se nalaziti bruto rezultati ispitanika. Znamo iz regresijske analize da uvijek predviđamo rezultat koji se nalazi na crti regresije, odnosno prosječni rezultat koji ispitanici s nekim rezultatom u varijabli X postižu u varijabli Y. U našem slučaju, kada predviđamo s pravih na bruto rezultate, varijabla X nam je pravi rezultat, a varijabla Y je bruto rezultat.

Jednačina za predviđanje bruto rezultata na osnovu pravog rezultata je:

6-27

$$\widehat{X_b} = r_{bt}\frac{s_b}{s_t}(T - \overline{T}) + \overline{X_b}$$

Znamo da je aritmetička sredina bruto rezultata ujedno i aritmetička sredina pravih rezultata, a da je korelacija pravih i bruto rezultata jednaka odnosu standardne devijacije pravih i standardne devijacije bruto rezultata, što znači da je:

$$r_{bt}\frac{s_b}{s_t} = \frac{s_t}{s_b} * \frac{s_b}{s_t} = 1$$

Kada jedinicu uvrstimo u formulu 6-27, dobijemo da je predviđeni bruto rezultat jednak pravom rezultatu:

$$\widehat{X_b} = T - \bar{T} + \bar{T}$$

6-28

$$\widehat{X_b} = T$$

Procijenjeni pravi rezultat je prosječni rezultat na crti regresije. Da bismo odredili interval pouzdanosti u kojem se s određenom sigurnošću nalazi pravi rezultat ispitanika, koji je na testiranju ostvario neki bruto rezultat, potreban nam je i podatak o raspršenju mogućih pravih rezultata oko crte regresije. Ovo raspršenje nije ništa drugo do standardna pogreška mjerenja. Standardna pogreška mjerenja je raspršenje bruto rezultata jednog ispitanika na nizu ponovljenih mjerenja ili standardna pogreška prognoze intervala pouzdanosti u kojem očekujemo bruto rezultate ispitanika koji ima neki pravi rezultat.

Formula za standardnu pogrešku mjerenja je:

6-29

$$s_e = s_B\sqrt{(1 - r_{xx})}$$

Interval pouzdanosti određuje se kao interval u kojem se sa 95 % ili 99 % sigurnosti očekuje da će biti bruto rezultat ispitanika i određuje se kao:

$T\underline{+}$ 1,96*s_e ili $\underline{+}$ 2,58*s_e

Na primjer, na testu čija je aritmetička sredina bila 50, standardna devijacija 8, a pouzdanost α= 0,80, ispitanik je postigao rezultat 60. U kojem intervalu sa 95 % sigurnosti se nalaze bruto rezultati ispitanika?

x= 56 – 50 = 6

$$T = 0{,}80 * 6 + 50$$

$$\boldsymbol{T = 54{,}8}$$

$$s_e = 8 * \sqrt{(1 - 0{,}80)}$$

$$s_e = 3{,}57$$

Gornja granica intervala pouzdanosti: 54,8 + 1,96*3,57= 61,8

Donja granica intervala pouzdanosti: 54,8 - 1,96*3,57= 47,8

Bruto rezultati ispitanika sa 95 % sigurnosti se nalaze u intervalu 48 – 62.

6.3.5.3 Predviđanje pravih rezultata na osnovu bruto rezultata

Vidjeli samo da na osnovu bruto rezultata možemo procijeniti pravi rezultat, ali isto tako se pitamo: „U kojem intervalu možemo očekivati da se nalazi ispitanikov pravi rezultat?" U ovom slučaju predviđamo s bruto rezultata na prave rezultate, ali kao tačku oko koje gradimo interval koristimo procijenjeni pravi rezultat jer je on oslobođen komponente pogreške. Kao mjeru raspršenja koristimo standardnu pogrešku pravog rezultata (Lord i Novick, 2008):

$$s_{t.b} = s_t\sqrt{(1 - r_{xx})}$$

Kako je:

$$s_{t.b} = s_t\sqrt{(1 - r_{xx})}$$

$$s_t = \sqrt{r_{xx}} * s_b$$

formula za standardnu pogrešku pravog rezultata glasi:

6-30

$$s_{t.b} = s_b\sqrt{r_{xx}(1 - r_{xx})}$$

Interval pouzdanosti određuje se kao interval u kojem se sa 95 % ili 99 % sigurnosti očekuje pravi rezultat ispitanika:

$$T \pm 1{,}96 * s_{t.x} \text{ ili } \pm 2{,}58 * s_{t.x}$$

6.3.5.4 Predviđanje rezultata na paralelnom mjerenju

Na isti način kako smo predviđali prave i bruto rezultate, možemo predviđati i rezultat koji će ispitanik imati na paralelnim mjerenjima, koristeći se formulom za standardnu pogrešku prognoze u paralelnim mjerenjima:

6-31

$$s_{y.x} = s_y\sqrt{1 - r_{xx}^2}$$

pri čemu u formuli imamo koeficijent pouzdanosti koji je na paralelnim mjerenjima korelacija između rezultata na dvije primjene testa.

6.4 KRITIKE KLASIČNE TESTNE TEORIJE

Klasična testna teorija počiva na pretpostavci da je pogreška slučajna, da komponente pogreške ne koreliraju međusobno niti da koreliraju s pravim rezultatom. Međutim, veliko je pitanje da li je to zaista tako. Pretpostavimo da dvije osobe trebaju ocijeniti depresivnost kod grupe zaposlenika. Jedna osoba može biti više okrenuta ka simptomima promjena u navikama hranjenja i spavanja, dok druga osoba može više posmatrati generalni dnevni aktivitet osobe. Zahvaljujući činjenici da oni različito doživljavaju konstrukt koji mjere / opažaju, njihovi zaključci mogu biti jako različiti. U svjetlu Klasične testne teorije, te razlike vidimo kao posljedicu pogreške. Ali ovo nije slučajna pogreška i ovo nije nepouzdanost kako je definira KTT. Druga važna stvar u razumijevanju kritika Klasične testne teorije odnosi se na paralelnost mjera. Važna pretpostavka KTT je da imamo paralelne mjere, odnosno mjere sa svim jednakim parametrima. U praksi, ovo je rijetko slučaj. Obično imamo različito teške zadatke u testovima postignuća ili različita ponašanja na upitnicima ličnosti i teško da ih možemo tretirati kao paralelne mjere. Pa ipak, mi im pripisujemo iste bodove i dajemo im jednaku važnost u formiranju ukupnog uratka. Treće, same karakteristike mjernog postupka utječu na rezultate ispitanika. Nije svejedno da li su ispitanici radili test ujutro ili u podne, s jednim ili drugim ispitivačem, u lijevoj ili desnoj prostoriji... I na kraju, KTT se u formiranju intervala pouzdanosti oslanja na to da je standardna pogreška mjerenja jednaka za sve rezultate. Međutim, šta je s ispitanikom koji je postigao maksimalni rezultat na testu? Da li to znači da pri postavljanju intervala pouzdanosti oko njegovog rezultata trebamo pretpostaviti da on može postići viši rezultat od maksimalno mogućeg rezultata na testu?

Može se reći da je mnogo toga vodilo psihologe da se zapitaju o kvaliteti određivanja pouzdanosti, koristeći pretpostavke Klasične testne teorije. Tri osnovne kritike koje sumiraju sve rečeno odnose se na paralelnost mjerenja, tau-ekvivalentnost i kongeneričnost.

1. Paralelnost mjerenja – Klasična testna teorija izgrađena je na pretpostavci da imamo paralelne mjere, odnosno da su korelacije među pravim rezultatima na ponovljenim mjerenjima jednake i da je varijanca pogreške jednaka. Da imamo test u kojem su pitanja potpuno paralelna, to bi značilo da ispitanici uvijek jednako odgovaraju na pitanja. Na primjer, u testu znanja na svakom zadatku bi dobili ili 0 ili 1. U tom slučaju, zašto trebamo više od jednog pitanja? Upravo o ovome je Tucker (1946) govorio u svom članku i pokazao korištenjem Teorije odgovora na zadatak da povećanje pouzdanosti testa u duhu Klasične testne teorije vodi ka opadanju valjanosti. Ako bismo željeli imati potpuno konzistentna pitanja, onda bismo se morali ograničiti na jednu domenu iz koje bismo uzorkovali ponašanja, a to bi jako umanjilo valjanost testa. Na primjer, ako hoćemo da mjerimo profesionalne interese realističkog tipa (interese za rad s mašinama, alatima, konkretne aktivnosti u prirodi ili rad sa životinjama), mi bismo se morali fokusirati na samo jednu domenu i reći: „Moja pitanja će biti vezana za interese za rad s mašinama i alatima. Sve drugo će mi smanjiti pouzdanost“. S druge strane, postoji šansa da je neko realistični tip, ali da njegovi interesi nisu usmjereni prema mašinama, nego prema radu sa životinjama. Ako mu damo test koji je visoko pouzdan, ali nije obuhvatio domenu koja je našem ispitaniku zanimljiva, mi ćemo zaključiti kako on nema realističnih interesa. Uvrštavanje pitanja iz ostalih domena smanjit će pouzdanost testa, ali će povećati valjanost testa,

odnosno tada će test mjeriti širi raspon ponašanja i moći ćemo reći da on zaista mjeri realistične interese. Korelacije među pitanjima su atenuirane, ali ova atenuacija nije klasična posljedica nepouzdanosti i ne znači nepouzdanost. Rješenje ovog problema treba tražiti u umjerenosti. Nemudro je očekivati da u testu imamo pitanja u jediničnim korelacijama. Treba težiti ka tome da pitanja koreliraju s ukupnim uratkom u testu u intervalu od 0,3 do 0,8, a da njihove međusobne korelacije budu od 0,1 do 0,6 (Guilford, 1965).

2. Kongeneričnost – jedna od osnovnih pretpostavki Klasične testne teorije je da pitanja mjere samo jednu karakteristiku, odnosno da su jednodimenzionalna. Kongeneričnost je u osnovi paralelnosti mjerenja. Ne možemo očekivati visoke korelacije između pitanja ukoliko ne mjere jednu karakteristiku. Međutim, u analizama valjanosti (o kojima ćemo pričati kasnije) gotovo uvijek se pokaže da pitanja mjere više dimenzija. Kongenerični model pretpostavlja i da nema korelacije među greškama, a jednodimenzionalnost nema jasno postavljenu pretpostavku o nekoreliranim pogreškama. Pogreške zaista mogu korelirati. Nakon što se analizom valjanosti identificira povezanost pitanja s nekom latentnom varijablom, ostaci varijanci pitanja (koju u Klasičnoj testnoj teoriji pripišemo pogreški) među sobom mogu korelirati. Korelacije pogrešaka javljaju se zato što je na uradak u zadatku djelovalo više faktora ili zbog uticaja metode ispitivanja (Podsakoff, MacKenzie i Podsakoff, 2013). Dakle, pogreške nisu uvijek slučajne, one mogu korelirati iz raznih razloga jer pitanja gotovo pa nikad ne mjere samo jedan faktor.
3. Tau-ekvivalentnost – Klasična testna teorija pretpostavlja da pitanja, ne samo da mjere jedan faktor nego taj faktor mjere u jednakoj količini. To bi značilo da svako pitanje na testu znanja treba imati jednak broj bodova za tačno urađen zadatak ili da svako pitanje u upitnicima ličnosti u jednakoj mjeri određuje neku osobinu ličnosti. U praksi, to nije tako. U testovima znanja pitanja skaliramo tako da imamo lakša i teža, na upitnicima ličnosti koristimo različite domene. U analizi valjanosti vidimo da ne doprinosi svako pitanje podjednako objašnjenju latentne varijable. Možemo konstatirati da pitanja u testovima nisu tau-ekvivalentna. Štaviše, njihova visoka tau-ekvivalentnost smanjila bi mogućnost da varijablu kvalitetno mjerimo i da razlikujemo ispitanike.

Na osnovu kritika Klasične testne teorije, teoretičari su postavili nove teorije kao što je Teorija generalizabilnosti.

6.5 TEORIJA GENERALIZABILNOSTI

Teoriju generalizabilnosti (G teorija) postavio je Lee J. Cronbach sa suradnicima (1972). Na nju možemo gledati kao na proširenje Klasične testne teorije, ali prema kojoj predmet mjerenja varira na ponovljenim mjerenjima. G teorija je jako dobar psihometrijski okvir za strategije kompleksnih mjerenja u kojima veliki broj elemenata (u terminima G teorije – faceta) može utjecati na kvalitetu mjerenja. Ova je najveća i najvažnija razlika između KTT i G teorije.

Varijabilitet u psihološkim mjerenjima potječe od različitih faceta i te facete utječu na kvalitetu mjerenja. Osnovna filozofija G teorije je da pojedinačno mjerenje predstavlja samo uzorak iz univerzuma prihvatljivih mjerenja za koja pretpostavljamo da su međusobno zamjenjiva (Cronbach, 1972).

Zamislimo da psiholog želi ispitati koliko su studenti u socijalnim situacijama agresivni. Da bi proveo istraživanje, mora donijeti nekoliko odluka:

1. Kako će definirati agresivnost? Psiholog uzorkuje neku definiciju agresivnosti i neke indikatore agresivnosti iz neograničenog skupa mogućih indikatora i definicija pri čemu ove skupove zovemo univerzumima.
2. Ko će mu biti ispitanici? Psiholog na raspolaganju ima čitavu populaciju studenata i on treba uzorkovati jednu grupu.
3. Koje socijalne situacije će izabrati? Psiholog odabire neke socijalne situacije u kojima se studenti nalaze iz čitavog univerzuma mogućih situacija.
4. Koju metodu će koristiti? Psiholog na raspolaganju ima čitav skup različitih metoda i mora se odlučiti za jednu ili nekoliko njih.
5. Koje mjere će koristiti, odnosno koja pitanja će odabrati ukoliko se odlučio za neki upitnik? Naš psiholog uzorkuje moguća pitanja iz univerzuma pitanja.
6. Kako će skalirati pitanja, kako će ih poredati u upitniku i koju skalu procjene će koristiti?
7. U kojem vremenskom periodu će prikupljati podatke i u koje doba dana će prikupljati podatke?
8. ...

Na kraju, kada postavimo istraživanje, imamo osjećaj da je jedino tako i moglo biti postavljeno. Ali, kada pogledamo kako smo došli do svih elemenata istraživanja, shvatimo da je sve uzorkovano iz neograničenih skupova. Pojam uzorkovanja se, iz navike, veže za uzorkovanje ljudi. G teorija nam otvara novu perspektivu na uzorkovanje jer kaže: „Sve je uzorkovano!“.
Dva osnovna pojma koja smo naučili do sada su:

1. Univerzum mjerenja – sva moguća mjerenja koja su jednako prihvatljiva onome ko ih primjenjuje i koja vidi kao zamjenjiva
2. Facete – izvori varijabiliteta (na primjer, format pitanja, skala procjene na pitanjima, vremenski period u kojem se obavlja mjerenje, dužina testova...)

Glavno pitanje je, kada znamo da smo sve uzorkovali, koliko se možemo osloniti da su rezultati naših mjerenja slični rezultatima koje bismo dobili da smo imali neki drugi uzorak iz facete? Na primjer, ako smo istraživanje radili u januaru, da li možemo reći da bismo iste rezultate agresivnosti studenata dobili u martu? Shavelson i Webb (1991) koriste termin „dependability“ umjesto „reliability“ kako bi ukazali da pouzdanost ne treba biti shvaćena samo kao trenutno dobiveni broj. Pouzdanost je pitanje mogućnosti oslanjanja na preciznost mjerenja i sada i u drugim domenama i vremenima. Zorne primjere uzorkovanja nalazimo i u sportu. U sportovima, kao što je umjetničko klizanje, gimnastika i skijaški skokovi, grupa sudaca ocjenjuje izvedbu sportaša. Svaki od njih gleda istog sportaša, a opet – da li vide i registriraju isto? Oni jesu kvalificirani da rade svoj posao suca, ali su samo mali uzorak iz čitave populacije sudaca koji su mogli biti tu u tom trenutku i ocjenjivati tog sportaša. Da li bi sportaš prošao bolje da su na mjestu sudaca sjedili neki drugi, isto tako kvalificirani suci? Za ocjene sudaca možemo izračunati klasični koeficijent pouzdanosti, ali on nam neće reći ništa o tome kako se ocjene jednog suca odnose prema ocjenama drugih sudaca. Sudac može biti visoko pouzdan, ali da se sistematski razlikuje od drugog suca jer više ocjenjuje odjeću klizača nego klizačke figure.

6.5.1 Tipovi univerzuma i tipovi faceta

Mjerenja, odnosno nacrti, razlikuju se prema broju faceta koje su uzorkovane. Nekada imamo relativno jednostavnu situaciju kada uzorkujemo samo jednu facetu, ali većina nacrta je mnogo složenija.

6.5.1.1 Univerzumi sa jednom facetom

U G teoriji mjerenje se posmatra kao uzorak iz čitavog skupa mogućih mjerenja. Univerzumi sa jednom facetom su oni univerzumi u kojima se generalizira mjerenje na samo jedan skup. Na primjer, odabrali smo 20 pitanja za mjerenje depresije i odlučili da ćemo generalizirati na sva moguća pitanja kojima smo mogli mjeriti depresiju. Na kraju ćemo donijeti dijagnostičku odluku. U ovom nacrtu imamo samo jednu facetu. Univerzumi sa jednom facetom imaju 4 izvora varijabiliteta (Shavelson i Webb, 1991):

1. Individualne razlike među ispitanicima ili u terminima G teorije „prava varijanca“
2. Različite težine pitanja (ako su pitanja facete) ili subjektivnost procjenjivača / suca
3. Interakcija između ispitanika i pitanja / suca
4. Greška mjerenja

6.5.1.2 Univerzumi sa dvije facete

Kada dođemo do uzorkovanja dvije facete i vidimo mogućnost G teorije da procjenjuje više faceta mjerenja u isto vrijeme, shvatimo koliko je ona korisna alternativa za KTT. Neki primjeri ovakvog nacrta bili bi: u psihologiji rada kada u procesu ocjene uspješnosti rada grupa odabranih ispitivača (prva faceta – odabrani posmatrači) u nekoliko vremenskih tačaka posmatra kvalitetu rada radnika u tvornici (druga faceta – vrijeme). Ovdje se nacrt usložnjava jer imamo više interakcija:

1. Individualne razlike
2. Subjektivnost posmatrača
3. Vremenska tačka
4. Interakcija radnici X vrijeme
5. Interakcija posmatrači X vrijeme
6. Interakcija radnici X posmatrači X vrijeme
7. Greška mjerenja

6.5.1.3 Univerzumi sa tri ili više faceta

U ovim slučajevima uzorkujemo najmanje tri facete, a nacrti postaju još složeniji. Primjer univerzuma sa tri facete bio bi mjerenje ponašanja nekim upitnikom u nekoliko vremenskih tačaka, ali s različitim ocjenjivačima. Konkretan primjer bio bi u školi kada učiteljica mjeri ponašanje učenika, a onda u starijem razredu to radi učenikova razrednica.

Što je veći broj faceta, to je veći broj efekata u analizi. Što je dizajn mjerenja kompleksniji, to je veći broj komponenti varijance.

6.5.2 G i D studije

Teorija generalizabilnosti koristi se za različite vrste analiza, ali osnova ove teorije je da se analize događaju u dva koraka. Prvi korak zovemo G studija ili studija generalizabilnosti (Marcoulides, 1996) u kojoj se procjenjuju komponente varijance iz prikupljenih podataka. U ovoj fazi se identificiraju faktori koji utječu na varijancu bruto rezultata i procjenjuje se njihov efekat preko izračunavanja koeficijenta generalizabilnosti. U drugom koraku, rezultati prvog koraka koriste se da bi se procijenila generalizabilnost kombinacija faceta i da bi se donijela odluka o budućem mjerenju. Kako je ovo faza u kojoj se donosi odluka o nacrtu budućeg mjerenja, ona se naziva i D studija (decision study). Na primjer, provjerava se koliko pitanja treba ostati u upitniku, a da on i dalje zadrži svoj nivo generalizabilnosti ili koja su to pitanja koja imaju najbolje koeficijente generalizabilnosti kako bi bila zadržana u budućem instrumentu.

6.5.3 Određivanje koeficijenta generalizabilnosti

Generički koeficijent generalizabilnosti smo već imali priliku upoznati jer se on određuje Hoytovom metodom (pogledati dio 6.2.4).

On se označava grčkim malim slovom ρ (rho) i određuje se kao:

6-32

$$\rho^2 = \frac{MS_{isp} - MS_{rez}}{MS_{isp}}$$

6.5.3.1 Određivanje koeficijenta generalizabilnosti – G studija

Zamislimo da smo imali pet ispitanika koji su odgovarali na tri pitanja iz upitnika ekstraverzije. Uputa je glasila: „Molimo vas da na skali od 1 do 7 procijenite u kojoj mjeri ste vi 1) veseli 2) druželjubivi i 3) otvoreni“.

Odgovori ispitanika su bili sljedeći:

Tabela 6-18 Rezultati ispitanika na tri pitanja – određivanje koeficijenta generalizabilnosti

Ispitanici	veseli	druželjubivi	otvoreni	Prosječni rezultat ispitanika
1	4	5	4	4,3
2	2	2	2	2,0
3	4	3	4	3,7
4	4	5	6	5,0
5	4	7	7	6,0
Prosječni rezultat na pitanju	3,6	4,4	4,6	

U prvoj fazi, odnosno G studiji, trebamo uraditi analizu varijance:

Tabela 6-19 Izvori varijabiliteta u studiji

Izvor varijance	**Suma kvadrata**	**Stepeni slobode**	**MS (prosječni kvadrat)**	**Komponente varijance**	**Proporcija varijance**
Ispitanici (i)	27,1	i-1	6,8	1,2	0,54
Pitanja (j)	2,8	j-1	1,4	0,2	0,09
Rezidual	6,5	(i-1)(j-1)	0,8	0,8	0,37
Total	36,4	ij-1		2,2	1,00

Dobiveni koeficijent generalizabilnosti iznosi

$$\rho^2 = \frac{6{,}8 - 0{,}8}{6{,}8} = 0{,}88$$

Za svaku facetu možemo izračunati komponentu varijance, odnosno proporciju varijance koja ukazuje na stepen u kojem je određena faceta uticala na rezultate ekstraverzije.

Komponente varijance računaju se prema formuli:

Tabela 6-20 Komponente varijance rezultata

Komponenta varijance	
Ispitanici	$V_{isp} = \frac{MS_{isp} - MS_{rez}}{n_i}$
Pitanja	$V_{pit} = \frac{MS_{pit} - MS_{rez}}{n_j}$
Rezidual	MS_{rez}

Pri čemu je n_i broj ispitanika , a n_j – broj pitanja.

Kako je ove komponente varijance teško interpretirati, iz njih izračunavamo proporciju varijance kao odnos komponente varijance i ukupne varijance.

Glavni efekat koji nas interesira je efekat ispitanika. Ovaj efekat nam govori o tome u kojem stepenu ispitanici utiču na razlike u prosječnim ocjenama na nekom testu. Vidimo u našem primjeru da je njihov efekat najveći, odnosno da objašnjavaju 54% variranja u rezultatima. Ovo je upravo u terminima signala i šuma, naš signal. Međutim, rezultati su opterećeni i šumom, a to je rezidual mijenja rezultate i u našem slučaju on utječe na 37 % varijance u rezultatima, što nije malo.

Da bismo stekli punu perspektivu o logici G studija, trebamo obratiti pažnju i na efekat pitanja i objasniti zašto nam sada pitanja više nisu izvor pogreške kada identificiramo razlike među ispitanicima. Iz zadnjeg reda u tabeli koja prikazuje rezultate naših ispitanika na tri pitanja ekstraverzije, vidimo da postoje razlike u aritmetičkim sredinama, odnosno da su naši ispitanici pokazali niži prosječni rezultat na pitanju koliko su veseli nego koliko su otvoreni. Međutim, mi i želimo da ispitujemo razlike i ne želimo da imamo iste ocjene na pitanjima, a naša glavna briga je da li u našim pitanjima imamo sličnost u rangiranju ispitanika, odnosno da li su nam pitanja konzistentna u smislu rangova koje ispitanici zauzimaju na svakom pitanju. Ako je Markov rezultat bio drugi po rangu na prvom pitanju, da li je on drugi ili blizu drugog ranga na ostalim pitanjima? Upravo zato što želimo različite prosječne rezultate na pitanjima, nama efekat pitanja nije bitan.

Kada smo vidjeli prvi korak u kojem smo odredili komponente varijance, da vidimo što radimo u drugom koraku studije, odnosno u D studiji.

6.5.3.2 Donošenje odluke za buduća mjerenja – D studije

U drugom koraku psiholog procjenjuje kvalitetu različitih mjernih strategija i donosi odluku kako će mjerenje u buduće izgledati. Zašto nam je ovo važno? Zato jer želimo da maksimiziramo kvalitetu mjerenja i efikasnost. Ovo nije tako jednostavno. Na primjer, mi bismo željeli da uključimo više pitanja kako bismo bili sigurni da rezultate mjerenja možemo generalizirati i da smo karakteristiku valjano izmjerili. S druge strane, važno nam je da imamo testove koji su dovoljno kratki kako bi ih ispitanici željeli ispunjavati i ispunjavali ih od početka do kraja pažljivo i koncentrirano. Ukoliko je mjerenje dizajnirano tako da postoji grupa procjenjivača, onda želimo procjenjivače koji se slažu u svojim ocjenama. Kao što Shavelson i Webb (1991, str. 8,9) kažu, problem sa nekonzistencijom ocjena procjenjivača /sudija dovodi do problema u generaliziranju podataka. Ukoliko su procjenjivači nekonzistentni onda, u ocjenjivanju kandidata, zaključci o kvaliteti uratka zavise od čiste sreće, a uspjeh kandidata zavisi od toga da li je dobio liberalnog ili strogog procjenjivača. U didaktičkom primjeru mjerenja ekstraverzije naveli smo tri pitanja, ali upitnici za mjerenje osobina ličnosti obično su mnogo duži. Isto tako, željeli bismo da uključimo što veće uzorke ispitanika, ali znamo da povećanje broja ispitanika ne mora nužno povećati kvalitetu ispitivanja ako nam uzorak nije dovoljno kvalitetan.

4 moguća dizajna mjerenja koja možemo pretpostaviti bi bili (prema Nunnaly i Bernstein, 1994):

1. Da samo jednim pitanjem ispitujemo svakog ispitanika (da imamo samo jednog procjenjivača za cijelu grupu ispitanika)
2. Da sa nekoliko pitanja (gdje njihov broj ne mora biti isti kao u G studiji) ispitujemo sve ispitanike (da odaberemo nekoliko procjenjivača za sve ispitanike)
3. Da svakom ispitaniku dodijelimo jedno pitanje(ili jednog procjenjivača) i da se pitanja ne ponavljaju
4. Da ispitanicima damo veći broj pitanja, ali da se pitanja ne ponavljaju kod ispitanika

1. U slučaju kada odaberemo strategiju da samo jednim pitanjem ocjenjujemo ispitanike koeficijent generalizabilnosti računamo preko formule

6-33

$$\rho^2 = \frac{MS_{isp} - MS_{rez}}{MS_{isp} + (j-1)MS_{rez}}$$

U našem slučaju koeficijent generalizabilnost za samo jedno pitanje bi iznosio

$$\rho^2 = \frac{6,8 - 0,8}{6,8 + (3-1) * 0,8} = \frac{6}{8,4} = 0,71$$

Vidimo da smanjivanjem broja pitanja smanjujemo koeficijent generalizabilnosti. Ova formula omogućava da provjerimo kolika bi promjena koeficijenta generalizabilnosti bila kada bismo ispitivanje radili sa jednim pitanjem. Nameće se drugo logično pitanje – koja pitanja su najbolja za korištenje u budućem testu? Mogli bismo koristiti koeficijente korelacije između pitanja i ukupnog uratka na testu i onda izabrati ona koja imaju najviše korelacije. Ali, Lord i Novick (Lord i Novick, 2008, str. 210) su nam dali formulu na osnovu koje možemo procijeniti koeficijent generalizabilnosti svakog pitanja koja glasi

6-34

$$\rho^2 = \frac{(j-1) * (\sum cov_{1i})^2}{2jV_1 \sum cov_{ij}}$$

pri čemu je:

j – broj pitanja

cov_{1i} – kovarijanca između pitanja za koje određujemo koeficijent generalizabilnosti i svih ostalih pitanja

V_1 – varijanca pitanja za koje određujemo koeficijent generalizabilnosti

cov_{ij} – kovarijanca između svih pitanja osim izuzimajući kovarijance sa pitanjem za koje određujemo koeficijent generalizabilnosti

2. U slučaju da želimo promijeniti broj pitanja u testu (bilo da smanjimo broj ili povećamo broj pitanja) koeficijent generalizabilnosti procjenjujemo preko formule:

6-35

$$\rho^2 = \frac{\frac{MS_{isp} - MS_{rez}}{j}}{MS_{isp} + \frac{(j-k)MS_{rez}}{k}}$$

pri čemu je j – broj pitanja u prvoj formi testa, a k- broj pitanja u drugoj formi testa.

3. U trećoj mogućoj strategiji imamo samo jedno pitanje za svakog ispitanika i pri tome se pitanja ne ponavljaju. Ovaj dizajn mjerenja možemo zamisliti u situaciji ispitivanja bračnih parova gdje bračni partneri procjenjuju jedno drugo (teško bi bilo zamisliti zašto bi neki psiholog želio da svakom ispitaniku da jedno pitanje i da se ta pitanja ne ponavljaju). Vidimo iz ovog primjera kako dizajn mjerenja u psihologiji može biti kreativan. U slučaju da imamo ovakvu situaciju, naša matematika postaje jednostavna analiza varijance jer više nemamo interakcije.

Formula za određivanje koeficijenta generalizabilnosti u ovom slučaju je

6-36

$$\rho^2 = \frac{MS_{isp} - MS_{rez}}{MS_{isp} + \frac{j * MS_{zad}}{i} + \frac{(ij - i - j) * MS_{rez}}{i}}$$

4. Na kraju, mjerna strategija može biti da imamo nekoliko pitanja ili ocjenjivača za svaku osobu pri čemu ocjenjivači ocjenjuju samo tu osobu. Takav primjer imamo kada radnici ocjenjuju kvalitetu rada svog rukovodioca. Rukovodilac je samo njihov i radnici sa drugih odjela ga i ne mogu ocjenjivati.

U tom slučaju formula bi bila

6-37

$$\rho^2 = \frac{MS_{isp} - MS_{rez}}{MS_{isp} + \frac{j * MS_{zad}}{i * k} + \frac{(ij - ik - j) * MS_{rez}}{ik}}$$

pri čemu je k – broj osoba u grupi koja ocjenjuje jednog ispitanika.

U ovom dijelu obradili smo samo G i D studije u slučaju jednofacetnog dizajna. Ovo je osnova koju trebamo razumjeti i na osnovu koje ćete graditi svoje znanje i vještine u određivanju metrijskih karakteristika testova u okviru teorije generalizabilnosti. G teorija širi tradicionalnu perspektivu psihometrijskih karakteristika jer 1) konceptualizira pouzdanost tako da uključuje vjerovatnoću da facete sistematski utiču na kvalitetu i strategije mjerenja; 2) daje statističke alate pomoću kojih možemo ispitivati efekte pojedinih faceta i planirati dizajn mjerenja tako da on bude i kvalitetan i efikasan.

Preporuke za daljnje čitanje

Kroz poglavlje o pouzdanosti spomenuli smo nekoliko puta ime velikog psihologa i psihometričara Lee J. Cronbach-a. Njegovi tekstovi i udžbenici su danas klasici, te stoga toplo preporučujem da se dodatno upoznate sa pouzdanošću čitajući izvorne tekstove. Posebno interesantan je tekst objavljen pod nazivom „My current thoughts on coefficient alpha i successor procedures“ (2004). Sjajnu kolekciju tekstova o pouzdanosti možete naći u knjizi pod nazivom „Score Reliability: Contemporary Thinking on Reliability Issues“ koju je napisao Bruce Thompson (2002)

7 OSJETLJIVOST TESTA

Koji su osnovni koraci u analizi zadataka? Kako prepoznamo i kako korigiramo pogađanje odgovora na testovima? Kako možemo odrediti osjetljivost testa?

Osjetljivost testa direktno je povezana sa pouzdanošću kroz priču o homogenosti, jednodimenzionalnosti i konzistentnosti. Iz didaktičkih razloga, mi ćemo se osjetljivošću testa baviti u izdvojenom poglavlju. Osjetljivost ili diskriminativnost testa odnosi se na potencijal testa da razlikuje ispitanike po određenoj karakteristici. Kada kažemo diskriminativnost, u psihologiji se referirano na razlikovanje. Ovaj termin je u osnovi neutralan, te je važno da ga ne miješamo sa negativnom konotacijom koju je poprimio u društvenom kontekstu „diskriminacija" kao oblik nepravednog postupanja sa određenim grupama ljudi. Za psihologe on ima svoje izvorno značenje – razlikovanje.

Možemo reći da je test osjetljiviji kada je „sposoban" da registrira i male razlike koje postoje u karakteristici među ispitanicima. Kao kad kuhamo kolač, da bi kolač bio dobar moramo imati dobre sastojke. Tako je i kod testa – da bi test bio osjetljiv njegova pitanja moraju biti osjetljiva. Zato ćemo prvo govoriti o osjetljivosti zadatka, a onda i o osjetljivosti testa.

Prije nego krenemo u priču o osjetljivosti zadataka, moramo reći nešto o pogađanju odgovora.

7.1 POGAĐANJE I KOREKCIJE ZA POGAĐANJE

Zamislimo sljedeću situaciju. Dva ispitanika, Mišo i Azra (tabela 7-1) su radili test od 4 pitanja sa dvije alternative i trebali su da zaokruže da li je neka tvrdnja tačna ili pogrešna. Mišo je je uradio što je znao i nije pogađao rezultate. Azra je odlučila da pokuša pogoditi odgovore koje nije znala. U koloni 2 i 3 su njihovi hipotetski rezultati bez pogađanja. Mišo je znao odgovor na prva tri pitanja, a Azra nije znala niti jedan odgovor. Ali, Azra je uspjela pogoditi odgovor na dva pitanja. Kada imamo dvije alternative vjerovatnoća da pogodimo tačan odgovor je 0,5. To znači da je na 4 pitanja, Azra slučajno mogla pogoditi odgovor na dva pitanja. Umjesto da razlikujemo Mišu i Azru sa tri boda, ova razlika se smanjila na samo jedan bod, odnosno ovaj test nije uspio da identificira veličinu razlike između Miše i Azre.

Spremnost ispitanika da pogađaju odgovore nalazimo na testovima postignuća u kojima su zadaci objektivnog tipa (višestrukog izbora, tačno – netačno, povezivanje,..). To su tipovi pitanja gdje ispitanici objektivno mogu koristiti pogađanje. Kod pitanja gdje trebaju dati kratki odgovor situacija je drugačija, mada istraživanja (Plumlee, 1952) pokazuju da pouzdanost kratkih odgovora nije jako različita od pitanja višestrukih izbora. Ispitanici imaju različite stilove odgovaranja. Neki ispitanici prvo rješavaju ono sto znaju, neki idu od lakšeg ka težem, neki se fokusiraju na jedne zadatke, ali pred kraj testa veliki broj ispitanika pokuša pogoditi odgovore koje ne znaju. Kod testova tipičnih ponašanja nemamo ovu vrstu pogađanja jer ne govorimo o tome da ispitanik ne zna odgovor, ali možemo govoriti o nekim drugim karakteristikama kao sto je socijalno poželjno odgovaranje.

Tabela 7-1 Odgovori dva ispitanika bez pogađanja i sa pogađanjem

	KOLONA 2	KOLONA 3	KOLONA 4	KOLONA 5
Zadatak	Mišo bez pogađanja	Azra sa pogađanjem	Mišo bez pogađanja	Azra sa pogađanjem
1	1	0	1	0
2	1	0	1	1
3	1	0	1	1
4	0	0	0	0
Ukupan rezultat	3	0	3	2

7.1.1 Korekcije za pogađanje na nivou zadatka

Pogađanje utiče na procjenu težine zadatka. Zamislimo da je zadatak višestrukog izbora sa 4 alternative radilo 100 ispitanika od kojih polovina zna riješiti zadatak, a druga polovina ne zna riješiti zadatak. To je zadatak čiji bi indeks *p* bio 0,5. Zamislimo da je ovih 50 ispitanika, koji nisu znali rješenje, pokušali da pogode tačan odgovor. Po slučaju bismo oko 12 pogodaka. To znači da bismo u konačnici imali 62 ispitanika koji su tačno riješili zadatak i indeks *p* bi se povećao na 0,62. Posljedice pogađanja, posebno kod neadekvatnih pitanja, čine zadatak lakšim ili težim nego što stvarno jeste, a to se odražava na ukupan rezultat testa.

Korekcije indeksa *p* radimo prema formulama koje su date u tabeli (Bucik, 1997):

Tabela 7-2 Korekcije za pogađanje na nivou zadatka

	Bez korekcije za pogađanje	Sa korekcijom za pogađanje
Test snage	$p = \frac{N_p}{N}$	$p = \frac{N_p - \frac{N_n}{A-1}}{N}$
Test brzine	$p = \frac{N_p}{N - N_z}$	$p = \frac{N_p - \frac{N_n}{A-1}}{N - N_z}$

Pri čemu je

Np – broj ispitanika koji je tačno riješio zadatak

Nn – broj ispitanika koji je netačno riješio zadatak

Nz – broj ispitanika koji nisu stigli do zadatka zbog ograničenog vremena (nepokušani zadaci)

7.1.2 Korekcije za pogađanje na testu

Kada pogledamo popunjene testove ispitanika, onda razlikujemo 4 vrste zadataka:

1. R – tačno riješeni zadaci – zadaci u kojima imamo tačan odgovor (R – right)
2. W – netačno riješeni zadaci – zadaci koji imaju pogrešan odgovor (W – wrong)
3. S – preskočeni zadaci – zadaci koji nisu riješeni, ali nakon njih vidimo makar još jedan pokušaj rješavanja (S – skiped)

4. U – nepokušani zadaci – zadaci iza kojih više nema pokušaja rješavanja (U – untried)

U testovima snage, odnosno testovima u kojima vrijeme nije ograničeno, očekujemo da vidimo tačne i netačne zadatke jer je ispitanik imao dovoljno vremena da proba riješiti sve zadatke od kojih je neke uspješno uradio, a neke nije uspješno uradio. U testovima brzine, koji su koncipirani tako da ispitanik može riješiti sve zadatke, očekujemo da vidimo tačne zadatke i nepokušane zadatke. I u jednom i u drugom slučaju ne očekujemo da vidimo preskočene zadatke. Uzimajući zdravo za gotovo činjenicu da ispitanici pogađaju odgovore na pitanja objektivnog tipa, mi možemo procijeniti koliko ispitanik zaista zna odgovoriti koristeći formulu za korigirani rezultat (Nunnally i Bernstein, 1994, str. 341):

7-1

$$R_{proc} = R - \frac{W}{A-1}$$

Pri čemu je A – broj alternativa, R – rezultat koji je ispitanik ostvario, a W – broj pogrešnih odgovora. Vidimo iz formule da je korekcija manja što imamo više alternativa. Dakle, kada imamo samo dvije alternative (na primjer da li je tvrdnja tačna ili pogrešna) tada je korekcija najveća.

U slučaju da smo u testu naišli na zadatke koje ispitanici nisu pokušavali riješiti (preskočene i nepokušane zadatke) korekciju radimo prema formuli:

7-2

$$R_{proc} = R + \frac{S+U}{A}$$

Kod testova brzine, u obzir moramo uzeti činjenicu da je ispitanik bio ograničen vremenom i da nije stigao probati da uradi sve zadatke. Za ovu korekciju koristimo formulu

7-3

$$R_{proc} = R - \frac{W}{C} - \frac{S}{D}$$

pri čemu su C i D konstante, gdje se preporučuje da C bude nešto manje od A-1, a D nešto veće od A, odnosno broja alternativa.

Pogađanje je veliki problem koji Fajgelj (2003, str. 400 – 405, 470-472) obrazlaže sa aspekta jednakih vjerovatnoća pogađanja i sa aspekta vrste pitanja u testovima, te se preporučuje dodatno proučavanje literature radi boljeg razumijevanja ovog fenomena.

7.2 ANALIZA ZADATAKA

Analiza zadataka (engl. item analysis) predstavlja skup statističkih postupaka koji nam daju informacije o karakteristikama pojedinih zadataka u testu.

U okviru analize zadataka provjeravamo nekoliko važnih stvari o kojima ovisi kvaliteta zadataka i njihova sposobnost da razlikuju ispitanike. Trebamo se još jednom podsjetiti na razlike u testovima postignuća I testovima tipičnih ponašanja. U testovima postignuća imamo kvantitativno mjerilo gdje možemo reci da je neko bolji ili lošiji. Kod testova tipičnih ponašanja, kao što su različiti upitnici vrijednosti, skale stavova itd., govorimo isključivo o kvalitativnim razlikama. One znače da je neko manje ili više ekstrovertan ili da je neko manje ili više cijeni neku vrijednost, ali nikako da je neko bolji ili gori. Psihometrijske analize se i za jedne i za druge testove provode se na isti način, ali je interpretacija različita. Tako na primjer, kada govorimo o težini zadatka, to za testove postignuća znači koliko ispitanika je tačno uradilo zadatak, a za testove tipičnih ponašanja koliko ispitanika se složilo sa tvrdnjom. Analiza zadataka uključuje kvalitativnu i kvantitativnu analizu.

7.2.1 Kvalitativna analiza zadataka

Kvalitativna analiza zadataka odnosi se na sam sadržaj zadatka. Obično se radi tako da eksperti iz oblasti ocijene sadržaj pitanja i stilističke karakteristike, kao i njihovu ispravnost i etičnost (Urbina, 2004). Kao glavni kriteriji u kvalitativnoj analizi koriste:

1. Prikladnost sadržaja pitanja namjeni testiranja i populaciji za koju će se koristiti – sadržaj pitanja treba da omogući da se ostvari svrha istraživanja. Osim toga, pitanje mora biti pripremljeno tako da ga ispitanici iz populacije, za koju je namijenjeno, mogu svi razumjeti.
2. Prikladnost formata pitanja – desi se da ispitivač koristi pogrešnu skalu (na primjer pita „Koliko se slažete sa tvrdnjom..“ a onda ponudi odgovore „nikada, ponekada, često“). U kvalitativnoj analizi, pitanja se provjeravaju i prema tome da li ima dovoljno prostora za odgovor, da li je taj prostor jasno povezan sa samim pitanjem, da li je čitavo pitanje na jednoj stranici i slično.
3. Jasnoća kojom je pitanje postavljeno – onaj ko čita pitanje ne smije imati nedoumica što je ispitivač mislio. Na primjer, ako pitamo: „Koliko ste zadovoljni životom u vašem gradu?“ svaki ispitanik može imati svoju interpretaciju što je za njega život u gradu.
4. Jezička ispravnost – pitanje mora biti jezički (gramatički i pravopisno) ispravno napisano.
5. Etičnost pitanja – pitanja moraju biti postavljena tako da nikoga ne vrijeđaju. Ovdje trebamo razdvojiti etičnost postavljanja pitanja i pojavu „škakljivih“ termina u upitnicima kada je namjena testiranja upravo ispitivanje odnosa ispitanika prema nekim socijalno osjetljivim temama.

7.2.2 Kvantitativna analiza zadataka

Kvantitativna analiza uključuje određivanje ili izračunavanje nekoliko parametara osjetljivosti:

1. Težine zadatka
2. Diskriminativne valjanosti zadatka ili „diskriminabilnosti“ zadatka
3. Analizu distraktora
4. Korelacija među zadacima
5. Karakteristične krivulje zadatka

7.2.2.1 *Težina zadatka*

Za binarne varijable, težina zadatka određuje se kao proporcija ispitanika koji su tačno riješili zadatak u odnosu na ukupan broj ispitanika i označava se sa slovom p. Kada je p bliže 1 ,tada je zadatak lakši, a kada je p blizu 0 zadatak je teži. Optimalna težina zadataka je 0,5. Iz perspektive osjetljivosti, pitanje je to osjetljivije što bolje razlikuje ispitanike, a to znači da ima veliku varijancu. Kako smo naučili u dijelu 4.5., varijanca binarne varijable je jednaka V=p*q i možemo zaključiti da će najveću varijancu imati upravo pitanje čiji je p=q=0,5.

Evo i primjera:

Tabela 7-3 Odnos težine zadataka i varijance

Pitanje	P	q	Varijanca
1	0,1	0,9	0,09
2	0,3	0,7	0,21
3	0,5	0,5	0,25
4	0,7	0,3	0,21
5	0,9	0,1	0,09

Vidimo da bilo koja druga kombinacija p i q neće dati tako visoku varijancu. Naravno, pitanja koja imaju veću varijancu povećat će varijancu testa i samim tim činiti da test bude osjetljiviji.

Napomena: u slučaju da pitanje ne bodujemo sa 0 i 1, nego sa 1 i 2 (dodavanje konstante kako bi se izbjegla nula) varijanca će ostati ista.

U testu želimo da imamo pitanja koja imaju visoke varijance. Ali, da li je to baš uvijek tako? Dva su razloga za oprez prilikom odgovaranja na ovo pitanje: 1) odnos između varijance pitanja i njegova važnost za test i 2) ko je odgovorio tačno ili zaokružio 1 na ovom pitanju.

Kada bismo test pravili samo od umjereno teških stavki, mi bismo 1) onemogućili ispitanike koji imaju neku karakteristiku malo izraženu da je uopće pokažu jer nemamo laganih zadataka; 2) onemogućili ispitanike koji imaju karakteristiku jako izraženu da je pokažu u punom kapacitetu. Na primjer, kada bismo test iz aritmetike konstruirali samo od umjereno teških zadataka, učenici koji nisu dosegli taj nivo znanja dobili bi (hipotetski) 0 i mi bismo zaključili kako oni „ništa ne znaju“. S druge strane, učenici koji matematiku jako dobro poznaju i mogu uraditi mnogo teže zadatke osvojili bi maksimum na našem testu, ali njihov pravi rezultat prevazilazi skalu našeg testa. I tada ne bismo dobro dijagnosticirali visinu izraženosti karakteristike. U kliničkim testovima, uvijek ima pitanja na koja samo rijetki daju pozitivan odgovor jer je to, na primjer, neki jako težak i rijedak simptom. Kao laici mogli bismo reći da to pitanje treba izbaciti iz testa, ali, kao psiholozi, mi smo svjesni važnosti sadržaja tog pitanja i činjenice da na njega pozitivno odgovara samo mali procenat populacije. Zato takva pitanja ne eliminiramo iz testova.

Indeks težine nam ništa ne govori o tome ko su ispitanici koji tačno rješavaju taj zadatak. Da bi zadatak bio osjetljiv, on ispitanike treba da razlikuje tako da ga bolje rade ili ga više zaokružuju oni koji karakteristiku imaju više izraženu. Može se desiti da je pitanje optimalne težine, ali da ga podjednako dobro rade i slabiji i bolji. Dakle, na osnovu odgovora na pitanje mi ne možemo predvidjeti ko je bolji, a ko lošiji. Stoga nam indeks težine ne može biti jedini parametar u odlučivanju u osjetljivost zadatka, iako u sastavljanju finalne forme testa o njoj treba voditi računa. Ukoliko imamo jako lagana pitanja za našu populaciju, distribucija rezultata više neće biti normalna, nego negativno asimetrična. Kada imamo previše teška pitanja za našu populaciju, distribucija rezultata na testu biće pozitivno asimetrična.

Optimalna težina zadataka zavisi i od broja alternativa. Jedan način da procijenimo kolika bi trebala biti optimalna težina pitanja u zavisnosti od broja alternativa dao je Gregory (1992)

7-4

$$optimalna\ težina = \frac{1+g}{2}$$

Pri čemu je g- vjerovatnoća pogađanja odgovora. Na primjer, kada imamo četiri alternative vjerovatnoća pogađanja tačnog odgovora je 0,25. Optimalna težina zadatka je fleksibilan pojam i nema recepta, ali se treba držati pravila da optimalna težina treba omogućiti najbolje razlikovanje ispitanika. Optimalna težina bi bila između 0,5 i 0,8, s tim da u testu trebamo imati i zadatke sa težinom od 0,15 pa do 0,90.

Da rezimiramo priču oko težine zadatka, pogađanja i diskriminativnosti. Korekcije za pogađanje radimo da bismo došli do procjene prave težine koja uzima u obzir činjenicu da su ispitanici spremni pogađati. Težina zadataka je bitan parametar diskriminativnosti, ali nije jedina mjera, te ne može objasniti kvalitetu razlikovanja u nekom testu.

7.2.2.2 Osjetljivost zadatka

Osjetljivost ili diskriminativna valjanost zadatka je parametar koji nam govori o tome kako zadatak razlikuje ispitanike u odnosu na ono što je mjereno testom. Dobar zadatak treba da mjeri karakteristiku na isti način kako je mjeri test. Drugim riječima, dobar zadatak je onaj koji uspješnije rješavaju bolji ispitanici na testu, a manje uspješno slabiji ispitanici na testu. Na primjer, ako imamo dihotomni zadatak, želimo da u grupi slabijih ispitanika na testu bude više onih koji ne urade taj zadatak, a što su ispitanici bolji, da to više njih tačno radi zadatak.

Da bismo ispitali osjetljivost zadatka trebamo odgovore tri pitanja:

- Šta je kriterij ili standard posjedovanja varijable koja je predmet mjerenja?
- Kako pripremiti podatke za utvrđivanje diskriminativnosti?
- Na osnovu kojih statističkih mjera izraziti stepen diskriminativnosti stavki?

Kada govorimo o kriteriju, govorimo o mjeri sa kojom ćemo usporediti zadatak. Osim testa, koji je unutarnji kriterij, možemo koristiti i neko vantestno ponašanje za koje smo ranije potvrdili da mjeri onu karakteristiku koju mjerimo zadatkom. To može biti rezultat na nekom drugom mjerenju. U tom slučaju govorimo o vanjskom kriteriju. Na primjer, ako imamo upitnik ekstraverzije u kojem se nalazi pitanje: „Imam široki krug prijatelja “, oni koji su ekstravertniji bi trebali odgovarati na pitanje sa 1, a oni koji su introvertniji bi trebali odgovarati sa 0. Ako je to pitanje koje zaista mjeri ekstraverziju, onda bi odgovori ispitanika trebali korespondirati sa rezultatom na nekom drugom testu koji potvrđeno mjeri ekstraverziju. Vanjski kriterij može biti i rezultat objektivnih opažanja. Ipak, ukupan uradak na testu je uobičajeni kriterij.

Diskriminativnu valjanost zadatka određujemo:

1. Metodom ekstremnih grupa
2. Korelacijama između zadatka i kriterija

7.2.2.2.1 Metoda ekstremnih grupa

Metoda ekstremnih grupa radi se tako da ispitanike podijelimo u grupe prema rezultatima koje su ostvarili na testu. Veličina grupa zavisi od veličine uzorka. Ukoliko imamo veliki uzorak, grupe možemo formirati od 27% najboljih i 27% najlošijih, dok male uzorke možemo podijeliti u trećine i u izračunavanje uzeti donju i gornju trećinu. Jedna od podjela je i prema rezultatu koji dijeli ispitanike u kvartile. Tako bi u slabijoj grupi bili ispitanici čiji je rezultat u prvom kvartilu, a u boljoj grupi bi bili ispitanici iznad rezultata trećeg kvartila.

Za dvije ektremne grupe odredimo indeks težine *p* zadatka za koji određujemo diskriminativnu valjanost. Kada oduzmemo indeks težine gornje grupe od indeksa težine donje grupe dobijemo vrijednost koju zovemo **indeks diskriminativnosti**.

7-5

$$D= p^+ - p^-$$

Što je indeks diskriminativnosti viši, to je diskriminativnost zadatka veća. Na primjer, imamo 1156 ispitanika koji su radili test matematike. Odredili smo rezultate na testu koji dijele ispitanike u prvom i trećem kvartilu. Rezultat Q1 = 9 i manje predstavlja najslabije, a Q3 = 16 i više su oni najbolji. U slabijoj grupi se nalazi 351 ispitanik, odnosno oni koji su postigli 9 ili manje bodova na testu, dok se u grupi boljih nalaze 373 ispitanika koji su postigli rezultat 16 ili više. Od 351 ispitanika, njih 88 je tačno uradilo prvi zadatak, a njih 263 netačno. U boljoj grupi, njih 267 je uradilo tačno prvi zadatak, a 106 netačno.

Tabela 7-4 Kontingencijska tablica frekvencija uspješnosti uratka na zadatku i testu

	Bolja grupa	**Slabija Grupa**
Riješili zadatak	267	88
Nisu riješili zadatak	106	263

$$p^- = 88/351 = 0,25$$

$$p^+= 267/373= 0,72$$

$$D = p^+ - p^- = 0,72 – 0,25 = 0,47$$

Vidimo da oni, koji su generalno bolji na testu, u većoj proporciji rješavaju zadatak, a oni slabiji u manjoj proporciji rješavaju i zadatak. To je osnovna logika indeksa diskriminativnosti, kao i diskriminativne valjanosti zadatka.

7.2.2.2.2 Korelacije između zadatka i testa (item-total korelacije)

Već na samom početku priče o diskriminativnoj valjanosti smo rekli da je dobar zadatak je onaj koji uspješnije rješavaju bolji ispitanici, a manje uspješno slabiji ispitanici. Iz ovoga smo mogli naslutiti da, kada određujemo osjetljivost zadatka, mi u stvari koristimo mjere povezanosti sa kriterijem, odnosno koeficijente korelacije.

Kada odlučujemo koji koeficijent korelacije koristiti u obzir trebamo uzeti mjernu skalu na kojem je izražena varijabla.

7.2.2.2.2.1 Point – biserijalni koeficijent korelacije

Ovo je slučaj kada imamo jednu prirodno dihotomnu varijablu i jednu numeričku varijablu na intervalnoj skali. U ovom slučaju koristimo point-biserijalni koeficijent korelacije čija formula glasi:

7-6

$$r_{pbis} = \frac{M_p - M_t}{SD_t}\sqrt{\frac{p}{q}}$$

pri čemu je:

Mp – prosječni rezultat na testu za one koji su zadatak tačno riješili; Mt – prosječni rezultat na testu za sve ispitanike; SDt – standardna devijacija za sve ispitanike

p- proporcija ispitanika koji su zadatak tačno riješili, q- proporcija ispitanika koji su zadatak pogrešno riješili

Na primjer, imamo grupu od 26 ispitanika koji su radili zadatak iz testa znanja. U tabeli su prikazani njihovi rezultati na zadatku i na testu:

Tabela 7-5 Rezultati ispitanika na dihotomnom zadatku i testu

Ispitanik	Zadatak	Test	Ispitanik	Zadatak	Test
1	0	8	**14**	0	3
2	1	9	**15**	1	9
3	1	7	**16**	0	7
4	1	6	**17**	0	4
5	1	5	**18**	1	8
6	0	8	**19**	1	9
7	0	4	**20**	0	6
8	1	5	**21**	1	5
9	0	5	**22**	0	8
10	1	10	**23**	0	5
11	1	10	**24**	1	8
12	0	9	**25**	0	8

13	1	9	**26**	1	8

- Odredimo aritmetičku sredinu testa za ispitanike koji su zadatak tačno riješili (ili koji su označili 1 na pitanju, ukoliko je test postignuća).ž

Mp= 7,7

- Odredimo aritmetičku sredinu i standardnu devijaciju testa za sve ispitanike

Mt = 7; SDt =2,01

- Odredimo p i q za zadatak

p = 0,54; q =0,46

- Sve parametre uvrstimo u formulu

$$r_{pbis} = \frac{7{,}7 - 7}{2{,}01}\sqrt{\frac{0{,}54}{0{,}46}}$$

$$r_{pbis} = 0{,}37$$

Napomena: kada računamo point-biserijalni koeficijent korelacije za zadatke u testovima brzine uzimamo u obzir samo one ispitanike koji su došli do tog zadatka. U tom slučaju Mt**,** koja je bila aritmetička sredina svih ispitanika, postaje Mt', odnosno aritmetička sredina na testu onih ispitanika koji su došli do zadatka. p i q se obavezno računaju samo sa rezultatima ispitanika koji su došli do određenog zadatka.

7.2.2.2.2.2 Biserijalni koeficijent korelacije

Biserijalni koeficijent korelacije računa se kada nemamo zadatak kao prirodno dihotomnu varijablu, nego smo varijablu naknadno dihotomizirali, dok s druge strane imamo rezultate testa izražene na intervalnoj skali. Na primjer, mogli smo pitati ispitanike da ocijene skali od 1 do 5 koliko često posjećuju kulturne događaje. 1 je nikako, 2 – rijetko, 3 – ponekad, 4 – često i 5 – stalno. Za potrebe istraživanja kažemo da su ispitanici sa odgovorima 1 i 2 jedna grupa koja ima nisku izloženost kulturnim sadržajima, a oni koji su odgovorili sa 3, 4 i 5 druga grupa sa većom izloženošću kulturnim sadržajima.

Biserijalni koeficijent se računa prema formuli

7-7

$$r_{pbis} = \frac{M_p - M_n}{SD_t} * \frac{pq}{y}$$

pri čemu je

Mp – prosječni rezultat na testu za one koji su zadatak tačno riješili

Mn – prosječni rezultat na testu za one koji su zadatak pogrešno riješili

SDt – standardna devijacija rezultata svih ispitanika

p i q – indeksi lakoće i težine zadatka

y – odsječak na y osi normalne distribucije za proporciju p koji se očitava u statističkim tablicama

7.2.2.2.2.3 Tetrahorični koeficijent korelacije

Ovaj koeficijent korelacije koristimo kada imamo dvije dihotomne varijable od kojih je makar jedna naknadno dihotomizirana. Mjerenje povezanosti između dihotomnih varijabli je područje koje se proučava još od ranih dana razvoja moderne statistike (Ekstrom, 2009). Karl Pearson je 1900. godine (prema Ekstrom,2009) u svom članku prvi put predložio koeficijent koji će kasnije postati poznati kao tetrahorični koeficijent korelacije. Osnovna ideja tetrahoričnog koeficijenta je da se, u određivanju povezanosti između dvije dihotomne varijable, koristi kontingencijska tablica 2x2 pri čemu je makar jedna varijabla naknadno dihotomizirana, odnosno u svojoj osnovi ima bivarijatnu standardnu normalnu distribuciju, te da se odredi parametar veličina dihotomiziranih varijabli pri čemu bivarijatna standardna normalna distribucija odgovara zajedničkim vjerovatnoćama kontingencijskih tablica. Camp (1933, prema Ekstrom, 2009) je istakao kako je i sam Pearson smatrao da je tetrahorični koeficijent jedan od njegovih najvažnijih doprinosa statističkoj teoriji, odmah nakon sistema krivulja, χ^2 testa i statistike za male uzorke. Dakle, nema nikakvog razloga da ne upoznamo ovako važan koeficijent iako je on prilično zanemaren od strane praktičara.

Imamo dvije varijable u po dvije kategorije koje formiraju kontingencijsku tablicu (kontingencija – međuovisnost):

Tabela 7-6 Kontingencijska tablica za tetrahorični koeficijent korelacije

Testni rezultat		**Zadatak**		
		pravilno	**pogrešno**	Suma
	Uspješni	A	b	a+b
	Neuspješni	C	d	c+d
	Suma	a+c	b+d	N

U kontingencijskoj tablici se nalaze 4 kategorije ispitanika. Kategorija „a" su oni koji su tačno uradili zadatak i uspješni su na testu, kategorija „b" su oni koji su pogrešno uradili zadatak, ali su bili uspješni na ispitu, kategorija „c" su oni koji su zadatak tačno uradili,ali neuspješni na testu i kategorija „d" oni koji su bili neuspješni i na prvoj i na drugoj mjeri. Proporcije a i d zovemo proporcije slaganja, a proporcije b i c proporcije neslaganja. Što je veća proporcija ispitanika u proporcijama slaganja (a manja u proporcijama neslaganja) to je i korelacija veća.

Formula za određivanje tetrahoričnog koeficijenta je:

7-8

$$r_{tet} = \cos\left[\pi \frac{bc}{\sqrt{bc} + \sqrt{ad}}\right]$$

Pri čemu je π=3,14. Sada vidimo zašto je ovaj koeficijent nepopularan među psiholozima – jer koristi kosinus funkciju za određivanje koeficijenta korelacije. Međutim, da vidimo što u geometrijskoj prezentaciji znači kosinus.

Kosinus ugla α određen je kao odnos između nalegle katete b i hipotenuze c. Kosinus ugla je veći ukoliko je ugao oštriji, a dužina ove dvije duži sličnija, odnosno što su dva vektora sličnija dužinom to je njihov kosinus ugla veći:

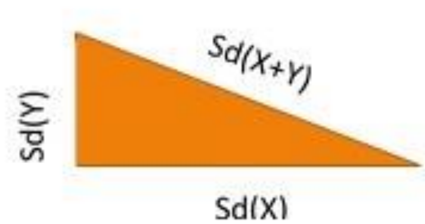
7-9

$$sličnost = \cos\alpha = \frac{A * B}{\|A\| * \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\ \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Pri čemu su A_i i B_i komponente vektora a i b. Ukoliko su vektori normalizirani tako da je svaka vrijednost vektora oduzeta od aritmetičke sredine vektora (A – Ma) tada se ova mjera sličnosti zove centrirana sličnost i jednaka je Pearsonovom koeficijentu korelacije (Wikipedia, 2016).

Kako bismo odnos kosinusa i korelacije još bolje razumjeli, sve ćemo staviti u termine varijance i standardne devijacije. Standardna devijacija dvije varijable u linearnoj kombinaciji čija je korelacija 0 jednaka je zbiru standardnih devijacija svake od varijabli. U geometriji, standardne devijacije možemo predstaviti kao dvije katete trougla koja (ako su nekorelirana) koje zaklapaju ugao od 90°:

Slika 7-1 Odnos kateta i hipotenuza u pravouglom trouglu izražen u mjerama standardnih devijacija



Ovaj odnos je tačan jer znamo da je u Pitagorinoj teoremi zbir kvadrata kateta jednak kvadratu hipotenuza. U terminima varijance i standardne devijacije, zbir kvadrata kateta su u stvari varijance dvije varijable, a kvadrat nad hipotenuzom je varijanca kombinacije dvije varijable. U ovoj situaciji same varijable imaju nultu korelaciju.

Sada pretpostavimo da su X i Y povezane. To znači da će njihova varijanca biti jednaka:

$$V_{X+Y} = V_X + V_Y + 2COV_{XY}$$

U terminima određivanja kosinusa ugla koji nije pravi analogna formula za određivanje kosinusa nepravog ugla je

$$c^2 = a^2 + b^2 - 2a * b * cos\ \alpha$$

Ako sada u ovu formulu uvrstimo da su a,b i c varijance dobije sljedeće:

$$V_{X+Y} = V_X + V_Y - 2V_X V_Y\ cos\alpha$$

$$2V_X V_Y\ cos\alpha = V_{X+Y} - V_X - V_Y$$

$$cos\alpha = \frac{V_{X+Y} - V_X - V_Y}{2V_X V_Y}$$

$$cos\alpha = \frac{V_X + V_Y + 2COV_{XY} - V_X - V_Y}{2V_X V_Y}$$

$$cos\alpha = \frac{2COV_{XY}}{2V_X V_Y}$$

Kako je $\frac{COV_{XY}}{V_X V_Y}$ = r_{XY}, to je onda

7-10

$$cos\alpha = r_{XY}$$

Što je varijabilitet u jednoj varijabli manji, a u drugoj veći, to će i korelacija među njima biti manja, a što varijable sličnije variraju to će i korelacija biti veća. Mi varijable možemo zamisliti kao dva vektora. Ukoliko je ugao koji oni zaklapaju oštriji to znači da su njihovi rezultati više sinhronizirani, odnosno povećanje rezultata na jednom vektoru, praćeno je povećanjem rezultata na drugom vektoru. Kosinus ugla koji ova dva vektora zaklapaju je u stvari njihova korelacija. Kada imamo pravi ugao, kosinus ugla je 0, odnosno korelacija između ove dvije varijable je 0. Ovakve vektore zovemo ortogonalne. Kada dva vektora zaklapaju tupi ugao, njihov kosinus je negativan, odnosno korelacija je negativna, a kod ugla od 180° kosinus je -1, odnosno korelacija je -1.

Pogledajmo primjer određivanja tetrahoričnog koeficijenta korelacije. Zamislimo da imamo kontingencijsku tablicu s rezultatima 100 ispitanika na zadatku i na testu:

Tabela 7-7 Kontingencijska tablica rezultata 100 ispitanika– tetrahorični koeficijent korelacije

Testni rezultat		**Zadatak**		
		pravilno	**pogrešno**	Suma
	Uspješni	60	10	70
	Neuspješni	10	20	30
	Suma	70	30	100

Tetrahorični koeficijent računamo kao:

$$r_{tet} = \cos\left[\pi \frac{10 * 10}{\sqrt{10 * 10} + \sqrt{60 * 20}}\right]$$

$$r_{tet} = \cos\left[\pi \frac{100}{44{,}6}\right]$$

$$r_{tet} = \cos\left[\pi \frac{100}{44{,}6}\right]$$

$$r_{tet} = \cos[3{,}14 * 3{,}14 * 2{,}24] = \cos 7{,}04 = 0{,}73$$

7.2.2.2.2.4 φ (fi) koeficijent korelacije

Tetrahorični koeficijent korelacije je povezanost između varijabli koje imaju u osnovi bivarijatnu normalnu distribuciju, a fi-koeficijent je povezanost između varijabli koje u osnovi imaju bivarijatnu diskretnu distribuciju. Ovaj koeficijent su neovisno predložili Boas, Pearson i Yule (prema Ekstrom, 2009).

ϕ koeficijent računamo koristeći istu kontingencijsku tablicu pri čemu je formula za računanje:

7-11

$$r_{\emptyset} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Zamislimo da smo imali grupu ispitanika kojima smo ispitivali sljepoću na boje. Ispitanik radi zadatak uspješno ili neuspješno, a na kraju je rezultat na testu ili da vidi boje ili da ne vidi boje.

Tabela 7-8 Kontingencijska matrica rezultata testa sljepoće na boje

Testni rezultat		**Zadatak**		
		pravilno	**pogrešno**	Suma
	Uspješni	60	10	70
	Neuspješni	10	20	30
	Suma	70	30	100

$$r_{\emptyset} = \frac{60 * 20 - 10 * 10}{\sqrt{70 * 30 * 70 * 30}}$$

$$r_{\emptyset} = 0{,}53$$

Napomena:

U literaturi (npr. Nunnaly i Bernstein, 1994) se navodi da u slučaju kada imamo dihotomne varijable koristimo ϕ koeficijent. Međutim, ono što napominje Fajgelj (2003) i što je važno da znamo je da, kada su podaci binarni, formule za ϕ i za Pearsonov r su identične, odnosno daju identične vrijednosti jer je ϕ samo poseban slučaj Pearsonovog koeficijenta korelacije. Isto tako, kada imamo jednu prirodno dihotomnu varijablu, a drugu kontinuiranu, tada je vrijednost point-biserijalnog koeficijenta identična vrijednosti koju bismo dobili da računamo obični r jer je point-biserijalni koeficijent specifični oblik formule Pearsonovog koeficijenta korelacije. Za razliku od ϕ i R_{pbis}, tetrahorični i biserijalni koeficijenti korelacije predstavljaju aproksimacije povezanosti jer su varijable vještački dihotomizirane.

Preporučuje se studentima da se „poigraju" podacima. Napravite dva niza podataka na intervalnoj skali, izračunajte njihov r, potom ih kategorizirajte i vidite što ćete dobiti. Kategorizacija dovodi do smanjenja korelacije. Međutim, Nunnaly je jednom prilikom uspoređivao r i r_{bis}. Prije dihotomizacije dobio je r=0,52m, dok je nakon dihotomizacije dobio r_{bis}= 0,71! Slično kao u našim primjerima – i mi smo dobili viši tetrahorični koeficijent nego ɸ. Ove razlike se najvjerovatnije javljaju kada se presjek podataka radi udaljeno od centralne vrijednosti ili (kod tetrahoričnog) apsolutne vrijednosti koje prelaze 1.

7.2.2.2.3 Kako odrediti kvalitetu zadatka na osnovu koeficijenta korelacije?

Prilikom interpretacije diskriminativne valjanosti zadatka, moramo imati na umu odnos između težine pitanja i koeficijenta korelacije. Ukoliko pitanje ima visok ili nizak p, onda to pretpostavlja da ga u velikoj proporciji rade ili ne rade ispitanici u gotovo svim grupama. Kada pitanje slabo razlikuje ispitanike, onda je prirodno da ima i nižu korelaciju s testom. Većina pitanja s jako visokim p ima korelacije s testom od 0 do 0,3. Zadaci u nultim i negativnim korelacijama s testom po pravilu nisu poželjni za test. Za ovakve zadatke imamo „poseban" tretman. Ukoliko se desi da neki zadatak ima negativnu korelaciju s testom, trebamo pogledati zašto se to desilo. Na primjer, ukoliko je test mjerio depresiju, a pitanje je bilo „Ja imam mnogo energije", ovo pitanje će biti negativno povezano s uratkom u testu. Ako smo zaboravili da imamo ovakvo inverzno pitanje, mi ćemo njegov rezultat pribrojiti ukupnom uratku i umjesto da imamo niži rezultat, dobit ćemo viši rezultat. Zato moramo biti oprezni i provjeriti da li smo rekodirali sva inverzna pitanja.

Procjenu kvalitete pitanja na osnovu korelacije dao je Ebel (Ebel, 1965). Ova podjela je i danas aktuelna (Bucik, 1997, str. 165):

Tabela 7-9 Procjena kvalitete pitanja prema visini koeficijenta korelacije s ukupnim uratkom

Visina korelacije	**Kvaliteta pitanja**
≥ 0,40	Vrlo dobro pitanje
0,3 – 0,39	Dobro pitanje
0,2 – 0, 29	Slabo pitanje
≤ 0,2	Loše pitanje

7.2.2.2.4 p-r dijagram

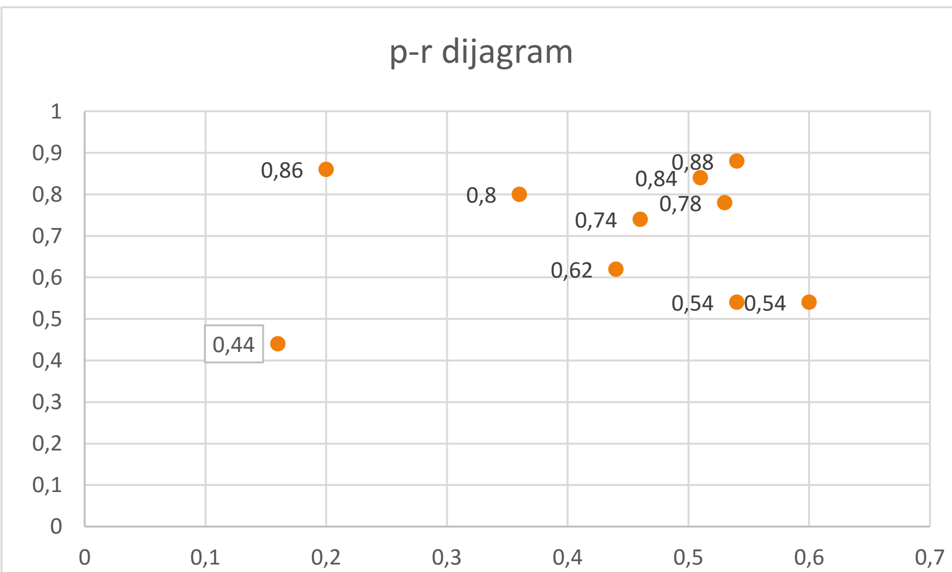
Vidimo da težina pitanja i diskriminativnost pitanja predstavljaju međuovisne parametre. Zato je koristan pristup u analizi ove parametre posmatrati zajedno. Za to koristimo p-r dijagram. To je dijagram u kojem se na apscisi nalaze vrijednosti koeficijenata korelacije između zadatka i testa (item-total korelacije), a na ordinati vrijednosti koeficijenata p. U dijagram smjestimo sva pitanja kao tačke koje povezuju visinu ova dva parametra.

Pogledajmo tabelu sa 10 pitanja višestrukog izbora sa po pet alternativa koje su rješavali učenici i njihove parametre (p, r_{pbis} i D – indeks diskriminativnosti). S obzirom na broj alternativa optimalni indeks težine bio bi oko 0,6.

Tabela 7-10 Indeksi težine, diskriminativne valjanosti i indeks diskriminativnosti 10 pitanja

pitanje	broj ispitanika	p	r_{pbis}	D
P 1	50	0,80	0,36	0,23
P 2	50	0,78	0,53	0,52
P 3	50	0,62	0,44	0,41
P 4	50	0,84	0,51	0,30
P 5	50	0,88	0,54	0,29
P 6	50	0,74	0,46	0,37
P 7	50	0,54	0,54	0,44
P 8	50	0,44	0,16	0,10
P 9	50	0,54	0,60	0,44
P 10	50	0,86	0,20	0,17

Slika 7-2 p-r dijagram



Sada vidimo tačno gdje su naša pitanja. Prvo, možemo zaključiti da većina pitanja ima korelacije s testom više od 0,4, a da tri pitanja imaju korelaciju nižu od 0,4. Pogledajmo pitanje koje ima korelaciju 0,2 i indeks p – 0,87. Ovo pitanje najbolje oslikava ranije opisani odnos – visok p povezan je s nižom korelacijom. Pitanje koje je ovdje najproblematičnije je P8 čiji je p=0,44, što je u ovom testu bilo najteže pitanje, ali i dalje osjetljivo (ima visoku varijancu). U kvalitativnoj analizi pitanja vidjelo se da pitanje nije bilo u potpunosti jasno studentima što je dovelo do toga da je p niži od ostalih pitanja. Uradak na zadatku je ovisio o pravilnom razumijevanju teksta, a ne poznavanja materije što je dovelo do toga da ima nisku korelaciju s uratkom na testu.

7.2.2.3 Homogenost zadataka – inter-item korelacije

O povezanosti među zadacima govorili smo u poglavlju o pouzdanosti i rekli da visoka povezanost među zadacima može biti problem jer tada pouzdanost raste, ali valjanost opada jer nismo sigurni da li smo pitanjima uspjeli zahvatiti varijabilnost karakteristike. Ipak, važno je da pitanja budu povezana i da im korelacije budu pozitivne. Međutim, desi se situacija da imamo zadatke koji su u nultim ili negativnim korelacijama sa drugim zadacima. U ovim situacijama se trebamo zapitati da li se radi o testu koji nije jednodimenzionalan. Da bismo odlučili želimo li više korelacije među pitanjima ili su nam i niže korelacije prihvatljive, moramo uzeti u obzir namjenu testa. Ukoliko se radi o testu postignuća sa zadacima različite težine i iz različitih domena, tu prirodno ne možemo očekivati visoke korelacije, mada možemo očekivati pozitivne korelacije. S druge strane, ako zaista mjerimo karakteristiku koja je homogenija (nego što je to znanje), onda očekujemo da između zadataka imamo više korelacije. Dakle, nema recepta za proglašavanje pitanja dobrim ili lošim. Sve ovisi o cilju testiranja, facetarnosti varijable i njenoj varijabilnosti. Pogledajmo dvije matrice korelacija na testu znanja i upitniku anksioznosti:

Tabela 7-71Korelacije pitanja na testu znanja i testu anksioznosti

	Test znanja					**Upitnik anksioznosti**					
	k1	k2	k3	k4	k5		A1	A2	A3	A4	A5
k1	1,0					A1	1,0				
k2	0,1	1,0				A2	0,4	1,0			
k3	-0,1	0,4	1,0			A3	0,4	0,5	1,0		
k4	0,2	0,2	0,0	1,0		A4	0,3	0,6	0,6	1,0	
k5	0,3	0,0	0,0	0,3	1,0	A5	0,4	0,5	0,6	0,6	1,0

Usporedbom korelacija u ove dvije matrice vidimo da u upitniku anksioznosti imamo visoke pozitivne korelacije, dok u testu znanja imamo niske korelacije. Razlog tome je što je test znanja multidimenzionalan, odnosno mjeri različita područja znanja, dok je anksioznost homogenija (makar kako je definirana za pripremu ovog upitnika). U poglavlju o valjanosti opet ćemo govoriti o sadržaju upitnika.

7.2.2.4 Analiza distraktora

U testovima postignuća, gdje imamo pitanja višestrukog izbora s nekoliko alternativa, osim analize tačnog odgovora provjeravamo i kvalitetu pogrešnih alternativa, odnosno distraktora. Distraktori nam služe da bi ispitanici koji ne znaju odgovor na pitanje imali malu šansu da pogode odgovor. Kažemo da je dobro pitanje ono u kojem su sve alternative ispitanicima koji ne znaju odgovor bile podjednako atraktivne. Ukoliko to nije slučaj, onda povećavamo mogućnost pogađanja tačnog odgovora. Na primjer, imamo pitanje na testu znanja iz psihologije za studente psihologije:

Čime se bavio Charles Spearman?

a) Psihologijom
b) Fizikom
c) Biologijom
d) Sociologijom

Možemo pretpostaviti da studenti psihologije znaju da je Charles Spearman jedan od najznačajnijih psihologa. Ukoliko neko ne zna odgovor, a polaže psihologiju, lako će pogoditi odgovor. Iako se nama čini da je vjerovatnoća pogađanja odgovora 0,25 (1/A) činjenica je da će studenti lako odgovoriti na pitanje sve da nisu „otvorili knjigu“. Distraktori mogu i zbuniti ispitanike koji znaju tačan odgovor. Kada su distraktori nejasni tada će i bolji ispitanici birati neku od alternativa koja im se čini tačna. Zamislimo da promijenimo alternative u gornjem pitanju, te da smo alternativu A definirali kao tačnu prilikom konstrukcije pitanja:

Čime se bavio Charles Spearman?

a) Inteligencijom
b) Statistikom
c) Biologijom
d) Sociologijom

Student koji je učio može biti zbunjen jer je Spearman, osim što se bavio inteligencijom, dao veliki doprinos razvoju statističkih metoda, te možemo reći da se bavio statistikom. To će dovesti do toga da ispitanici koji znaju, daju „pogrešan“ odgovor, pa će pitanje ispasti teže nego što stvarno jeste. Analiza distraktora se radi tako da ispitanike podijelimo u bolju i slabiju skupinu prema uratku na testu, podatke o zaokruženim alternativama prikažemo u tabeli i pregledamo relativne odnose između broja odgovora na distraktorima u boljoj i lošijoj skupini. Sjetite se da lošijoj skupini distraktori trebaju biti podjednako atraktivni, a bolju skupinu da ne zbune. Moramo voditi računa i o težini zadatka – kada je zadatak lakši, prirodno da ga bolje rade i ispitanici u slabijoj skupini.

Tabela 7-82 Analiza distraktora

	Pitanje 1				**Pitanje 2**				**Pitanje 3**				**Pitanje 4**			
	A	B*	C	D	A	B	C	D*	A*	B	C	D	A	B*	C	D
Bolja grupa	7	8	2	1	10	2	2	4	17	1	1	1	1	7	3	7
Slabija grupa	6	6	4	3	3	2	2	12	10	2	3	3	0	19	4	5

Ako pogledamo pitanje 1, u obje grupe imamo podjednak broj biranja alternative A i B. Možemo konstatirati da ovo pitanje podjednako uspješno rade i jedna i druga grupa, a da je alternativa A bila previše slična alternativi B. U drugom pitanju vidimo da je pitanje diskriminativno jer ga slabije rade ispitanici u slabijoj grupi, ali im je alternativa D bila previše privlačna i moguće da su neki od njih bili zbunjeni pitanjem. U pitanju 3 vidimo dobru raspodjelu po distraktorima i to je dobro udešeno pitanje, dok u pitanju 4 imamo inverziju – njega bolje radi grupa slabijih nego boljih. Ovo pitanje treba provjeriti – zašto je ono lakše slabijim ispitanicima?

7.3 OSJETLJIVOST TESTA

Osjetljivost testa je mogućnost testa da razlikuje ispitanike po mjerenom svojstvu na osnovu ukupnih bruto rezultata. Mjerni instrument smatramo osjetljivim onda kada može izmjeriti najmanje razlike u mjerenoj pojavi. Ovo je i glavna poenta testiranja jer mogućnost razlikovanja ispitanika čini da je test iskoristiv. Ako postoji neka razlika u nekoj karakteristici između dva ispitanika, a mi ne možemo da je identificiramo testom, onda taj test nema svoju funkciju jer će ta dva ispitanika biti tretirana kao jednaki, baš kao i neka druga dva ispitanika koja zaista jesu jednaki. Zašto je to važno? Zato jer smo kao psiholozi svjesni da razlike u psihološkim karakteristikama nisu samo kvantitativne prirode. Razlike su, prije svega, kvalitativne prirode. Na primjer, ako govorimo o inteligenciji, razlika od 10 DQI jedinica nije isto što i razlika u 10 cm u dužini, zato jer se iza ove razlike od 10 DQI jedinica nalaze kvalitativne razlike u mentalnim procesima. Osjetljivost testa važna je zbog kvalitativnih razlika.

7.3.1 Kako identificiramo osjetljivost testa

Ako je osjetljivost testa sposobnost razlikovanja, onda je prvi indikator za osjetljivost testa njegova varijanca. Do sada smo vidjeli da je varijanca testa određena varijancom zadataka i njihovom kovarijancom. Međutim, varijanca nam govori o varijabilitetu, ali ne i o tome koliko test razlikuje ispitanike, odnosno koliko ispitanici postižu različite rezultate. Kao mjeru razlikovanja ispitanika po ostvarenim rezultatima možemo koristiti BOR ili broj ostvarenih razlikovanja. To je ukupan broj parova različitih rezultata koje nalazimo u testu. Na primjer, imamo grupu ispitanika koji su radili 2 zadatka i gdje su mogući rezultati na testu bili 0, 1 i 2. Zamislimo da je 5 ispitanika dobilo rezultat 0, njih 25 dobilo rezultat 1 i njih 30 rezultat 2. Svaki ispitanik grupe 0 (njih 5) razlikuje se od svakog ispitanika iz grupe 1 (njih 25), a ukupan broj parova razlikovanja je 5 x 25, odnosno 125. Ispitanici iz grupe 0 razlikuju se i od ispitanika iz grupe 2 (njih 30), što je 150 različitih parova. Svaki ispitanik iz grupe 1 razlikuje se od ispitanika iz grupe 2 i ukupan broj parova različitih rezultata grupe 1 i 2 je 750 (30 x 25).

Tabela 7-13 Broj ostvarenih razlikovanja

Mogući rezultati na testu	Frekvencije (N = 60)	Broj ostvarenih razlikovanja (BOR)	Maksimalni broj ostvarenih razlikovanja (BOR_{max})
0	5	5x 25 = 125	20 x 20 = 400
1	25	5 x 30 = 150	20 x 20 = 400
2	30	25 x 30 = 750	20 x 20 = 400
		Ukupno 1025	Ukupno 1200

Vidimo da imamo ukupno 1025 parova različitih rezultata. Maksimalni broj parova različitih rezultata bi bio kada bi distribucija bila kvadratična, odnosno kada bi jednak broj ispitanika postizao svaki od mogućih rezultata na testu.

Da bismo odredili osjetljivost testa, koristimo Fergusonov δ (delta) koeficijent. Iako je Thurlow (1950) bio među prvima koji je prepoznao da postoji razlika između osjetljivosti i pouzdanosti testa, te pisao o osjetljivosti i ponudio koeficijent, ovaj koeficijent ostao je poznat kao Fergusonova delta. Thurlow i Ferguson (Ferguson, 1949) rekli su da za svaku veličinu uzorka postoji maksimalni mogući broj razlikovanja koji se može opaziti (BORmax). Maksimalni broj razlikovanja može se usporediti s brojem stvarno opaženih razlikovanja i taj odnos se može izraziti proporcijom. Fergusonov δ je odnos između opaženog broja razlikovanja i maksimalnog mogućeg broja razlikovanja:

7-12

$$\delta = \frac{BOR}{BOR_{max}}$$

Ukoliko su svi postigli isti rezultat, onda je δ=0. Ovaj koeficijent baziran je isključivo na ordinalnim karakteristikama podataka. Iako neki autori smatraju da je ovaj koeficijent nepotreban (Norman, 2008) i da je generalno „osjetljivost jednako pouzdanost“, drugi autori ističu da je ovaj koeficijent koristan jer predstavlja dodatni indeks, osim koeficijenata pouzdanosti i valjanosti. On se i računa kada postoje zadovoljene „glavne“ metrijske karakteristike – pouzdanost i valjanost.

Formula za određivanje Fergusonovog δ koeficijenta u slučaju kada imamo dihotomne varijable je:

7-13

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

Pri čemu je:

K – broj pitanja

N – broj ispitanika

F – frekvencija svakog rezultata

U slučaju kada imamo politomne varijable (varijable s više vrijednosti) Fergusonov δ koeficijent se računa kao:

7-14

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

Pri čemu je m – broj vrijednosti u odgovorima na pitanjima.

7.3.2 O čemu ovisi osjetljivost testa

7.3.2.1 Težinska primjerenost zadataka

Prva determinanta težine testa je težina samih zadataka. Ukoliko su zadaci u testu lagani za populaciju na kojoj je rađeno mjerenje, test će imati negativnu asimetriju i slabije će razlikovati ispitanike u području nižih vrijednosti. Isto tako, ako su zadaci bili teški, onda će ispitanici slabije raditi test, on će imati pozitivno asimetričnu distribuciju i slabije će razlikovati ispitanike u području viših vrijednosti. Neko ko nije dovoljno upućen u psihologiju rekao bi da onda trebamo imati zadatke srednje težinske primjerenosti. Ali, kao što smo ranije rekli, trebamo u testovima i zadatke koji su teži i lakši kako bismo svima omogućili da pokažu raspon u kojem posjeduju karakteristiku. To znači da moramo „balansirati“ u pripremi zadataka, a kriterij težinske primjerenosti zadatka ne smije biti jedini kriterij.

Kao grubu procjenu težinske primjerenost testa računamo prosječni indeks težine zadataka:

7-15

$$\bar{p} = \frac{\sum p_i}{k}$$

Pri čemu je p_i – indeksi težine zadataka, a k – broj zadataka.

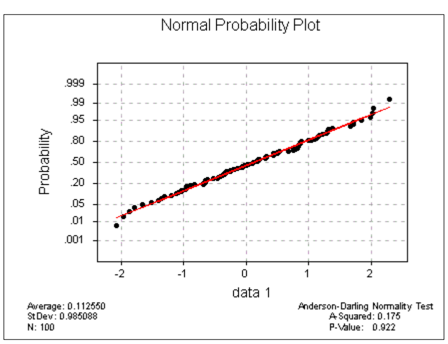
Ako nam nisu poznati indeksi lakoće zadataka, već samo aritmetička sredina bruto rezultata, dijeljenje M sa brojem zadataka $\bar{p} = \frac{M}{k}$ nam daje uvid u visinu prosječnog indeksa lakoće.

Ipak, kvaliteta i težina zadataka u testu zavisit će prije svega od namjene testa i populacije za koju je test namijenjen. Ako je namjena testa da ga koristimo na široj populaciji i da imamo norme, tada moramo voditi računa o sastavu samog testa tako da u njemu imamo zadatke koji omogućavaju postizanje normalne distribucije. Testovi mogu biti konstruirani tako da za opću populaciju budu jako lagani ili jako teški, ali oni imaju posebnu namjenu. Na primjer, test može biti namijenjen djeci s određenim deficitima u mentalnom funkcioniranju, ali to ne znači da se djeca ne razlikuju po svojim kapacitetima. Ovaj test bi imao izrazito negativnu asimetriju i jako slabo razlikovao djecu u široj populaciji, ali na populaciji djece s određenim mentalnim deficitom rezultati bi mogli imati normalnu raspodjelu. Ista situacija je i za izrazito nadarenu grupu. Nadarena grupa će uraditi test za opću populaciju bez ikakvih problema. Za njih trebamo test sa zadacima koji su jako teški za djecu u široj populaciji, ali će omogućiti da unutar specifične grupe detektiramo razlike. Dakle, namjena testa određuje i težine zadataka.

Kako provjeravamo da li je došlo do poremećaja diskriminativnosti?

1. Vizuelnim pregledom distribucije rezultata – već pogled na distribuciju u slučaju izrazitih odstupanja od normaliteta može nam reći da je to nedovoljno osjetljiv test. Ipak, ova „odokativna“ metoda ne može biti osnova za dokazivanje u psihologiji.

Slika 7-3 Normalni grafikon vjerovatnoće



2. Normalnim grafikonom vjerovatnoće u kojem se na apscisi nalaze opažene frekvencije, a na ordinati teoretske frekvencije. U slučaju normalne raspodjele, križanja se tada nalaze na pravcu **x=y**. Ako nije normalna raspodjela, onda ćemo imati devijaciju s pravca.
3. Indeks asimetrije i kurtoze (indeks varijabilnosti i visine distribucije) o kojim ste učili na Statistici ukazuju na to da li postoji problem. Kada imamo problem, onda će indeks asimetrije biti veći od standardne devijacije, a javit će se i problem s visinom distribucije.

4. Koeficijent varijabilnosti koji predstavlja odnos standardne devijacije i aritmetičke sredine i računa se prema formuli:

7-16

$$\mathrm{KV} = \frac{SD}{M} * 100$$

Što je koeficijent varijabilnosti manji, to je diskriminativnost manja.

7.3.2.2 Broj zadataka u testu

Što je veći broj zadataka, to je osjetljivost testa veća. Ukoliko test ima jedno pitanje, onda će se ispitanici podijeliti u samo dvije grupe – koji su riješili zadatak i koji nisu. Iako kratki testovi mogu biti korisni, njihov glavni problem je nemogućnost razlikovanja ispitanika. Što testovi imaju više zadataka, to imaju veću šansu da detektiraju postojanje razlika. Ne utječu svi zadaci podjednako na osjetljivost. Da bi zadatak poboljšao osjetljivost testa, on sam mora biti kvalitetan, odnosno moći da razlikuje ispitanike. S više zadataka, teoretski, može se postići veći broj različitih rezultata (BRR), a time i veći broj ostvarenih rezultata (BOR). BRR je indeks koji govori koliko teoretski možemo postići različitih zadataka u testu.

7-17

BRR = k+1

Na primjer, ako na testu imamo 5 zadataka na testu tada je:

BRR = 5+1=6

BOR se računa na osnovu tablice distribucije frekvencija. Njegova formula je:

7-18

$$BOR = \frac{N^2 - \sum f_i^2}{2}$$

7.3.2.3 Korelacije među zadacima

Što su korelacije među zadacima veće, to će i varijanca testa biti veća jer znamo da je varijanca linearne kombinacije $V_u = \Sigma V_1 + \Sigma r_{ij}\sigma_i\sigma_j$. Prema formuli za računanje varijance, vidimo da će ukupna varijanca biti najveća kada su zadaci u potpunoj korelaciji i kada su prosječne težine. Međutim, tu dolazimo do paradoksalne situacije o kojoj smo pričali u pouzdanosti. Ako imamo korelaciju među zadacima r= 1, to znači da će ispitanik koji uradi prvi zadatak uraditi i sve ostale zadatke. Ako nije uradio prvi zadatak, neće uraditi nijedan drugi. Ako takav test ima 10 zadataka, onda ćemo imati ispitanike koji su ostvarili rezultate 0 ili 10. Usprkos maksimalnoj varijanci, osjetljivost ovog testa je gotovo pa nikakva, jer smo grupirali ispitanike u dvije grupe. U slučaju kada je korelacija r=0, tada je varijanca minimalna, ali bi broj različitih individualnih ukupnih rezultata mogao biti čak k +1**.** Tada imamo problem da su zadaci potpuno heterogeni. Ovdje se osjetljivost kosi sa dvije stvari: s pouzdanošću (pouzdanost je bazirana na korelacijama) i činjenicom da heterogeni zadaci daju neupotrebljiv ukupni rezultat jer se ne zna koja je od karakteristika više utjecala na uradak. Kako rješavamo ova dva problema? U praksi, vodimo računa da korelacije među zadacima imamo u prihvatljivoj zoni i da nisu previsoke (jer to znači da imamo redundantne zadatke). Rješenje je da test sadrži zadatke nejednake težine, ali da oni budu u „pristojnim" korelacijama.

7.3.2.4 Način formiranja uratka na testu

Znamo da se ukupni rezultat može formirati kao jednostavna linearna kombinacija:

$$X_U = \Sigma x_i$$

Ukoliko želimo povećati osjetljivost testa, onda možemo formirati ukupni rezultat prema modelu diferencijalno ponderirane linearne kombinacije:

$$\mathbf{X_U = \Sigma W_i X_i}$$

Pogledajmo još jednom dvije ispitanice koje su radile 5 zadataka:

Tabela 7-194 Ukupan uradak dvije ispitanice definiran kao JLK i kao DPLK

Indeks težine	**0,2**	**0,4**	**0,6**	**0,8**	**0,9**	**JLK**	**DPLK**
	Z1	Z2	Z3	U4	Z5		
Mia	1	1	1	0	0	3	1,2
Ema	0	1	0	1	1	3	2,1

Mia je znala riješiti prva tri zadatka, a Ema drugi, četvrti i peti. Ako pogledamo indekse težine ovih zadataka, vidjet ćemo da je Ema znala riješiti teže zadatke, a Mia samo lakše. Kada rezultat formiramo kao JLK one imaju jednak rezutat i kažemo: „One su jednake". U školi bi dobile istu ocjenu. Ali, Ena i Mia nisu jednake. Ena je bolja. Ako ponderiramo njihove uratke s indeksima težine zadataka, onda vidimo tu razliku. Ukoliko dopustimo da težina zadataka utječe na bodovanje ispitanika, onda ćemo ih moći bolje razlikovati. Stoga, iako se u literaturi mogu naći argumenti protiv ponderiranja, ponderiranje vodi ka osjetljivijim testovima.

8 VALJANOST

Što je valjanost? Da li je to karakteristika testa ili testnih rezultata? Zbog čega je važna? Koji su dokazi valjanosti?

Zamislimo sljedeću situaciju. Miro želi postati pilot. Prijavio se na testiranje za pilota. Tamo ga je dočekalo mnogo ispitivanja – od senzornih sposobnosti, zdravstvenih provjera pa sve do psiholoških testova – inteligencije i ličnosti. Miro se potrudio, ali je za dvije sedmice dobio odbijenicu. Bio je jako ljut i uložio je žalbu. Dobio je odgovor da, prema psihološkom profilu, ne odgovara za mjesto pilota. Miro se zapitao – kako to oni znaju? Što se mjeri psihološkim testovima pa da bi mogli predvidjeti moj budući uspjeh kao pilota? Ovo pitanje, koje je sebi Miro postavio, jedno je od glavnih pitanja u psihometriji – što mjerimo našim testovima? Ovim pitanjem ćemo se baviti u poglavlju o valjanosti.

8.1 ŠTO JE VALJANOST?

Dva pitanja koja sebi stalno postavljamo su:

- Kako se može interpretirati psihička osobina, stanje ili proces nekim testom kao mjernim postupkom?
- U kojem stepenu ili *koliko dobro* ispitujemo predmet mjerenja nekim mjernim postupkom?

Valjanost je najvažnija karakteristika mjernih instrumenata. Koncept valjanosti ispituje se od samog početka razvoja psihometrije i u tom vremenu date su različite definicije valjanosti. Jedna od osnovnih definicija je da je valjanost „stepen u kojem test mjeri ono za što je namijenjen da mjeri“. Ova definicija, iako dosta direktna, previše pojednostavljuje pitanje valjanosti. Cronbach (1971) je istaknuo da je „validacija proces prikupljanja dokaza koji podržavaju interpretacije i zaključke do kojih se došlo na osnovu rezultata“. Možda najbolja (u svakom slučaju najnovija) definicija koju su dali Američko udruženje za edukacijska istraživanja, Američko društvo psihologa i Nacionalno vijeće za mjerenja u obrazovanju u reviziji Standarda za pedagoško i psihološko testiranje (2006), je da je valjanost *„stepen u kojem dokazi i teorija podržavaju interpretacije testnih rezultata prikupljenih u okviru predviđenih područja za korištenje*“. Ova definicija sumira tri važne pretpostavke o valjanosti.

Valjanost testnih rezultata dolazi od akumuliranih dokaza koji podržavaju interpretacije i korištenje. To znači da se valjanost uvijek određuje u nekom stepenu i da ona ne funkcionira po principu sve ili ništa. Proces validacije nekog mjernog instrumenta započinje samom definicijom psihologa o tome što je konceptualni okvir i pretpostavke na osnovu kojih će pripremati test. Razvijanje novih spoznaja unutar teorije i akumuliranje podataka koji podržavaju interpretacije testnih rezultata dovodi do toga da valjanost hipoteza, baziranih na rezultatima u različitim područjima mjerenja i mjerenja u različite svrhe, može da jača ili da se smanjuje. Kako se ističe u Standardima za pedagoško i psihološko testiranje (AERA, APA, i NCME, 2006) valjanost je zajednička odgovornost onoga ko razvija test (ko daje dokaze i racionalu za namjeravanu svrhu testa) i onoga ko koristi test u praksi (ko procjenjuje dostupne dokaze u kontekstu u kojem se test koristi).

Na primjer, Test općih sposobnosti (Smith i Whetton, 1999) sadrži 4 subtesta kojima se mjere Numerički, Verbalni, Spacijalni i Neverbalni faktor. Verbalni faktor sposobnosti mjeri se testom od 36 zadataka koji zahtijevaju traženje analogija među ponuđenim riječima. Autori su, pri konstrukciji testa, krenuli od definicije da je verbalni faktor „sposobnost razumijevanja riječi kroz korištenje analogija". U terminima valjanosti, ovih 36 zadataka sami od sebe nisu ni valjani ni nevaljani. Ali, autorove interpretacije rezultata mogu biti valjane ili nevaljane. Da li su autori ispravno interpretirali rezultate na svih 36 pitanja kao razumijevanje riječi kroz korištenje analogija? Valjanost je pitanje tačnosti i legitimiteta interpretacija testnih rezultata. U definiciji valjanosti se kaže „predviđena područja za korištenje". Testovi se mogu koristiti u različitim područjima. Test verbalnih sposobnosti možemo koristiti u istraživanju, ali i u profesionalnoj selekciji kako bismo utvrdili da li je kandidat dovoljno verbalno vješt da bi se nosio sa zahtjevima posla. Međutim, rezultati na našem testu možda jesu valjano interpretirani kao verbalni faktor, ali da li je ovaj rezultat zaista koristan u smislu predviđanja ponašanja budućeg zaposlenika? Svaki test ima više namjena, odnosno područja u kojima se testni rezultati mogu koristiti, a dokazi za interpretacije rezultata mogu biti izvedeni na različite načine. Valjanost je kao, na primjer, kašika. Možemo reći da je kašika koristan predmet. Ali za što je korisna? Kašika je korisna da jedemo supu ili varivo ili da izvadimo sladoled u činiju, manje je korisna da njome jedemo komad mesa, ali je gotovo beskorisna za rezanje kruha (iako u nedostatku noža i kašika može poslužiti). Tako je i s valjanošću mjernih instrumenata – za neka područja test je jako koristan, ali za neka potpuno beskoristan. Valjanost nije karakteristika testa, nego karakteristika interpretacije rezultata testa. Brojni psiholozi ipak i dalje govore o valjanosti testa, vjerovatno zbog toga što je tako jednostavnije, što ne vide nijanse valjanosti ili što im je lijeno da kažu valjanost rezultata, pa kažu „verbalni test je valjan". Valjanost testnih rezultata bazira se na dokazima i teoriji. To znači da korištenje testova mora biti posljedica poznavanja dokaza koji podržavaju korištenje tog testa u određenoj situaciji. Suvremeni pogledi na valjanost upravo naglašavaju važnost interpretacije testova u okviru stabilne psihološke teorije.

Valjanost testnih interpretacija pregleda se kroz osnovne tipove valjanosti. Tradicionalna podjela tipova valjanosti uključivala je sadržajnu, kriterijsku i konstruktnu valjanost. Međutim, u posljednjih 20 godina pogled na valjanost se promijenio. Ta promjena se, prije svega, ogleda u pozicioniranju konstruktne valjanosti kao centralnog koncepta koji povezuje druge koncepte, odnosno dokaze za identificiranje valjanosti rezultata. Osim toga, kao kriterij za ocjenu valjanosti promoviran je i kriterij socijalnih posljedica testiranja. U Standardima za pedagoško i psihološko testiranje (AERA, APA, i NCME, 2006) potcrtano je 5 osnovnih grupa dokaza za identificiranje valjanosti testnih interpretacija. Tih pet grupa prikazano je na slici 8-1:

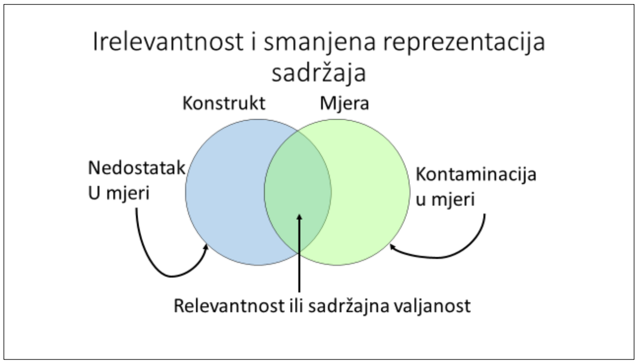
Slika 8-1 Tipovi valjanosti

Konstruktna valjanost interpretacija rezultata zavisi od sadržaja pitanja, unutarnje strukture testa, psiholoških procesa uključenih u odgovaranje, povezanosti s drugim varijablama te posljedicama korištenja testova. Od sada ćemo riječ „konstrukt“ često koristiti, pa je trebamo objasniti. Konstrukt predstavlja sve što se nalazi u ljudskom umu, a što nije moguće direktno opažati. U psihologiji, konstrukt označava koncepte kao što su osobine, ali označava i teoretske povezanosti između koncepata koje izvodimo na osnovu empirijskih dokaza iz ponašanja ljudi. Psihološki konstrukti razlikuju se po svojoj širini i kompleksnosti, potencijalu za korištenje i stepenu apstrakcije potrebnom da bi se o njima zaključivalo na osnovu prikupljenih podataka. Po pravilu, uže definirani konstrukti su manje apstraktni i imaju uže područje primjene (na primjer, lokus kontrole). Ali, kako je za njih jednostavnije doći do dogovora između znanstvenika što je taj konstrukt, oni se mnogo lakše mjere nego kompleksni multifacetni konstrukti koji mogu imati drukčija značenja između kultura, konteksta ili vremena.

8.2 DOKAZI O VALJANOSTI: SADRŽAJ PITANJA

Ovaj tip dokaza zovemo **sadržajna valjanost.** Sadržajna valjanost odnosi se na usklađenost sadržaja testa i sadržaja koji bi trebao biti u testu, odnosno koliko je sadržaj testa relevantan i reprezentativan za ono za što je namijenjen da mjeri. Da bi zaista mjerio psihološku prirodu konstrukta, test mora imati adekvatan sadržaj. Koliko je ovo važno znaju svi studenti kada polažu ispit. Ukoliko profesor ne vodi računa o sadržajnoj valjanosti testa i reprezentativnosti pitanja, studenti mogu biti ili jako obradovani („Tačno sam dobio ono što sam naučio“) ili jako tužni („Samo to nisam pročitao, a baš sva pitanja su bila iz tog poglavlja“). Ovo nam je dobra uvodna priča o prijetnjama sadržajnoj valjanosti.

Slika 8-2 Sadržajna valjanost, nedostatak i kontaminacija u mjeri



Prva prijetnja sadržajnoj valjanosti je ubacivanje sadržaja koji nisu relevantni za ispitivanje konstrukta. Pretpostavimo da je profesor trebao pripremiti test za parcijalni ispit iz Psihometrije u kojem je trebao ispitati znanje studenata o pouzdanosti. Tokom nastave radili su tri velike teme: što je pouzdanost, kako određujemo pouzdanost i predviđanje rezultata. Dobar test bi sadržavao pitanja iz ova tri područja, raspoređena tako da mjere sve važne koncepte i odnose. Prijetnja sadržajnoj valjanosti bi bila kada bi nastavnik zatražio od studenata da objasne glavne kritike Klasične testne teorije. Druga prijetnja sadržajnoj valjanosti zove se smanjena reprezentativnost konstrukta. Test ne smije imati sadržaje koji su nerelevantni za konstrukt, ali treba uključiti puni obim sadržaja (koliko je to moguće, s obzirom na vrijeme i dužinu testa) relevantnih za konstrukt. Na primjer, smanjena reprezentativnost konstrukta u našem slučaju bi bila kada bi profesor uključio samo jedno pitanje o određivanju pouzdanosti (pita samo za alternativne metode), a ne dotakne se unutarnje konzistencije. Zbog vremena testiranja treba voditi računa o količini sadržaja, ali i uključivanju sadržaja koji su reprezentativni i pokrivaju puni obim konstrukta. Sadržajna valjanost obično se ispituje kvalitativnom i kvantitativnom analizom (o kojima smo već govorili kada smo govorili o analizi zadataka). Kvalitativna analiza uključuje mišljenje eksperata o relevantnosti sadržaja (da li je sadržaj testa u cijelosti i pojedinostima relevantan za ono područje koje se želi ispitati); formalno-logičkom aspektu (da li su zadaci jasni i razumljivi, da li su rješivi, koliko je vremena potrebno za rješavanje, da li su predviđena rješenja zaista tačna), te procjenu opće reprezentativnosti čitavog testa. Kvantitativna analiza podrazumijeva korake koji se provode u analizi zadataka, tako da nam ona služi i za provjeru sadržajne valjanosti.

8.2.1 Facijalna valjanost

Koncept povezan sa sadržajnom valjanošću je facijalna valjanost. Facijalna valjanost je stepen u kojem se mjera "čini" kao povezana sa specifičnim konstruktom kada je koriste neeksperti (ispitanici). Iako facijalna valjanost nije ključna karakteristika iz perspektive psihometrije, ona ima važne implikacije za korištenje testa. Na primjer, autorica ovog priručnika je jednom prilikom, u okviru profesionalne selekcije, koristila upitnik ličnosti 16 PF (Cattell, Cattell i Cattell, 2000) kako bi procijenila konstrukte kao što su rasuđivanje, emocionalna stabilnost i socijalna odvažnost, a koji su bili važni za posao. Međutim, kandidati su, jedan za drugim, postizali ekstremno visoke rezultate na skali "Toplina" jer su, gledajući pitanja i intuitivno prosuđujući, zaključili kako je to poželjna karakteristika i željeli su se predstaviti u pozitivnom svjetlu. Sadržajna valjanost i facijalna valjanost nisu isto. Sadržajna valjanost se odnosi na reprezentativnost i relevantnost sadržaja vezanih za konstrukt koji mjerimo i mogu je ocjenjivati samo stručne osobe upoznate s konstruktom. Facijalna valjanost je stepen u kojem laici doživljavaju relevantnost testa i namjenu testiranja. Zbog toga je nama, u psihometrijskom smislu, mnogo važnija sadržajna valjanost.

8.3 DOKAZI O VALJANOSTI: POVEZANOST SA DRUGIM VARIJABLAMA

Povezanost sa drugim varijablama jedan je od osnovnih dokaza valjanosti testnih interpretacija. U suvremenom shvaćanju valjanosti naglašava se teoretsko razumijevanje konstrukta za koji se interpretiraju testni rezultati (konstruktna valjanost kao centralni pojam). Kao dokaz valjanosti testa koristi se povezanost rezultata s rezultatima drugih psiholoških konstrukata za koje se pretpostavlja da naš ciljni konstrukt ima teoretsku povezanost.

U tradicionalnom pogledu na valjanost, povezanost sa drugim varijablama nazivamo **kriterijska valjanost**. Međutim, u tradicionalnom shvatanju kriterijske valjanosti nema naglaska na konceptualno značenje interpretacije rezultata (Furr i Bacharbach,2013). Na primjer, psiholozi u profesionalnoj selekciji koriste testove čiji su testni rezultati povezani s radnim ponašanjima na mjestu za koje se bira osoba, te na osnovu testnih rezultata donose odluku ko je najbolji kandidat za posao. Pri tome, psihologe ne mora zanimati što zaista znače rezultati na testu (na primjer, što znači visok ili nizak rezultat na 16 PF skalama) sve dok se, na osnovu testnih rezultata, može s određenim uspjehom predviđati radno ponašanje. Iz tradicionalne perspektive kriterijske valjanosti kažemo da su rezultati testa valjani ukoliko su povezani s rezultatima na specifičnim kriterijskim varijablama. Kriterijska valjanost dijeli se na konkurentnu (dijagnostičku) i prognostičku. Konkurentna (dijagnostička) valjanost definira se kao stepen povezanosti testnih rezultata s rezultatima na kriterijskim varijablama kada su podaci prikupljani u isto vrijeme. Prognostička valjanost je povezanost testnih rezultata s kriterijskim varijablama pri čemu su rezultati na kriterijskim varijablama prikupljani u drugoj vremenskoj tački (obično nakon prikupljanja podataka na prediktorskim varijablama). U kriterijskoj valjanosti, prema tradicionalnom shvaćanju, ne bavimo se toliko psihološkim značenjem rezultata. Messick (1989) je dao svoju definiciju valjanosti prema kojoj valjanost predstavlja *„..integrativnu evaluacijsku procjenu stepena u kojem empirijski dokazi i teoretski okvir podržavaju adekvatnost i ispravnost zaključaka i mjera baziranih na testnim rezultatima ili drugim načinima mjerenja*“. On ističe da kriterijska valjanost potpada pod širi koncept konstruktne valjanosti. Ovo znači da nam ne može biti dovoljno da odredimo samo kriterijsku valjanost instrumenata, nego moramo poznavati konstruktnu valjanost testnih rezultata. Mi se u potpunosti slažemo, jer nepoznavanje konstruktne valjanosti testnih rezultata, a korištenje testova u profesionalnoj selekciji, bilo bi krajnje neozbiljno i neodbranjivo pred autoritetima. Međutim, to ne znači da ne trebamo poznavati metode određivanja povezanosti rezultata testova s kriterijskim varijablama. Stoga ćemo u nastavku ovog dijela prvo govoriti o tome kako određujemo povezanost rezultata na testovima s kriterijskim varijablama, a potom ćemo preći na određivanje povezanosti mjera tretiranu u suvremenom konceptu pouzdanosti. Tu ćemo govoriti o konvergentnoj i diskriminativnoj valjanosti.

8.3.1 Koeficijenti korelacije – povezanost između prediktora i kriterija

Već smo u priči o kriterijskoj valjanosti nekoliko puta spomenuli povezanost što nam je dalo naslutiti da su koeficijenti korelacije osnova određivanja kriterijske valjanosti. Još jednom napominjemo – ne postoji univerzalna kriterijska valjanost testnih rezultata. Svaki test ima onoliko koeficijenata kriterijske valjanosti koliko je kriterija za koji se test koristi kao prediktor.

8.3.1.1 Pearsonov koeficijent korelacije

Kada govorimo o koeficijentima korelacije, prvi koji nam pada na pamet je klasični Pearsonov koeficijent. Da se prisjetimo i formule za r:

8-1

$$r_{xy} = \frac{\sum(z_x z_y)_i}{N} = \frac{\sum(d_x d_y)_i}{N s_x s_y} = \frac{\sum[(X - M_x)(Y - M_y)]_i}{N s_x s_y}$$

Kada odredimo koeficijent korelacije, potrebno je da odredimo i njegovu statističku značajnost. Ona se računa preko t-distribucije prema formuli:

8-2

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Pri čemu su stepeni slobode df = N-2. U slučaju kada računamo ϕ koeficijent korelacije, statistička značajnost se računa preko χ^2, a formula je:

8-3

$$\chi^2 = N\Phi^2$$

pri čemu su stepeni slobode df =1

Dobivena korelacija je samo jedna od mogućih korelacija između prediktora i kriterija jer je dobivena na uzorku ispitanika. Stoga nas zanima u kojem intervalu možemo očekivati korelaciju između našeg testa kao prediktora i neke varijable kao kriterija. Da bismo odredili interval pouzdanosti, ne možemo koristiti „sirovu“ korelaciju (o tome smo govorili u dijelu 6.2.2) nego je moramo transformirati u Fisherov z-koeficijent.

$$z = \frac{\ln(1+r) - \ln(1-r)}{2}$$

Osim z-koeficijenta potrebna nam je i standardna pogreška kako bismo mogli postaviti interval pouzdanosti. Ona se računa prema formuli:

8-4

$$s_{zr} = \frac{1}{\sqrt{N-3}}$$

Nakon što smo postavili interval kao z $\underline{+}\, s_{zr} * 1{,}96(2{,}58)$ ponovo transformiramo z_r u r prema formuli:

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

8.3.1.1.1 Bonferronijeva korekcija

Statističko zaključivanje o značajnosti korelacije bazirano je na testiranju nul-hipoteze da varijable nisu u značajnoj vezi. Kada utvrdimo da je korelacija značajna (p < 0,05), mi odbacujemo nul-hipotezu o nepovezanosti varijabli. Pri tome ostavljamo prostora da smo pogriješili u odbacivanju nul-hipoteze (tip 1 pogreške ili α pogreška). U situacijama kada testiramo više hipoteza (provjeravamo korelacije jedne grupe varijabli sa drugom grupom varijabli kao u ispitivanju korelacija među pitanjima), povećavamo vjerovatnoću pojave rijetkog događaja – da smo neke korelacije proglasili značajnim, a one nisu značajne. Na primjer, ako smo računali korelacije 10 pitanja i dobili 100 značajnih korelacija, postoji vjerovatnoća da smo 5 od njih proglasili značajnim, a one to nisu.

Kako bismo umanjili α pogrešku, koristimo Bonferronijevu korekciju. Ova korekcija bazirana je na ideji da ispitivač koji testira m hipoteza, treba da testira značajnost na nivou od p_{ukupno}/m kako bi smanjio pogrešku tipa 1. Na primjer, ako smo testirali 8 korelacija pri čemu ukupni nivo vjerovatnoće za svih osam hipoteza treba biti 0,05, onda značajnost svake korelacije treba testirati na nivou od 0,05/8. Sada p više neće biti 0,05, nego 1/8, odnosno 0,00625. Dakle, Bonferronijeva nejednačina kaže da ukupna vjerovatnoća α pogreške treba biti jednaka ili manja od zbira vjerovatnoća α pogreške pojedinačnih korelacija ili:

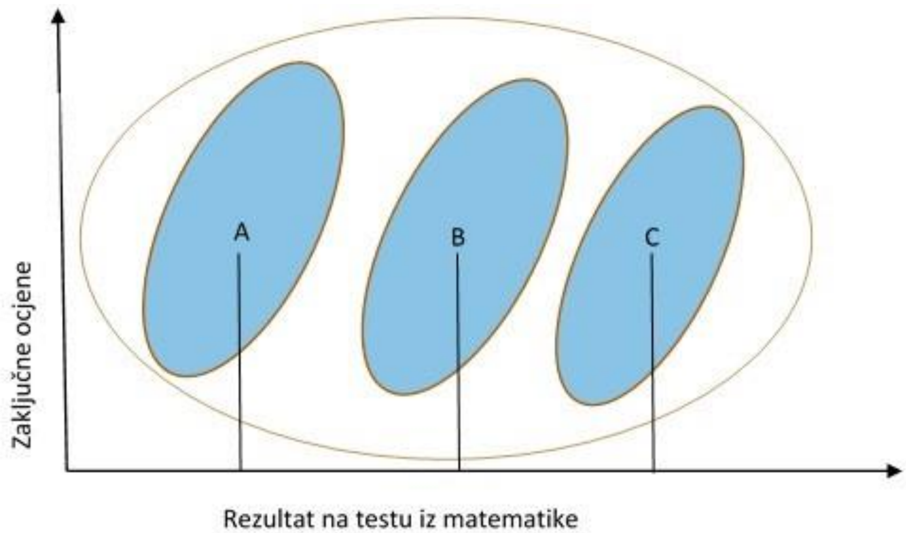
Ukupna α pogreška ≤ suma svih α pogreški pojedinačnih testova

Kritičari kažu da je Boneferronijeva korekcija dosta stroga, posebno kada imamo veliki broj testova i kada su oni u pozitivnim korelacijama, jer povećava vjerovatnoću pogreške tipa 2 ili β pogreške (da prihvatimo nul-hipotezu i proglasimo korelaciju neznačajnom, a korelacija je značajna). Druga kritika ovoj nejednačini upućena je na ideju o grupi hipoteza. Ne postoji neki definitivni koncenzus oko toga što je grupa hipoteza jer mi svaku korelaciju testiramo posebno.

8.3.1.1.2 Prijetnje koeficijentu korelacije

Postoji nekoliko faktora koji utječu na visinu korelacije između varijabli: pouzdanost instrumenta, omeđenost obima i heterogenost uzorka. O odnosu pouzdanosti i valjanosti govorili smo u dijelu 6.3.3, a o utjecaju omeđenosti obima u dijelu 6.3.3. Ovdje ćemo vidjeti kako heterogenost uzorka utječe na visinu korelacije. Pretpostavimo da smo ispitivali odnos između rezultata na standardnom testu matematike i zaključnih ocjena iz matematike u tri škole. U svakoj školi imamo djecu koja imaju zaključne ocjene od 1 do 5. Na testu znanja iz matematike pokazalo se da su prosječni rezultati jako različiti i da je škola A postigla niže rezultate od škole B, a škola C je bila najbolja. Njihovi rezultati prikazani su na slici 8-3:

Slika 8-3 Povezanost uratka na testu i zaključnih ocjena u tri škole



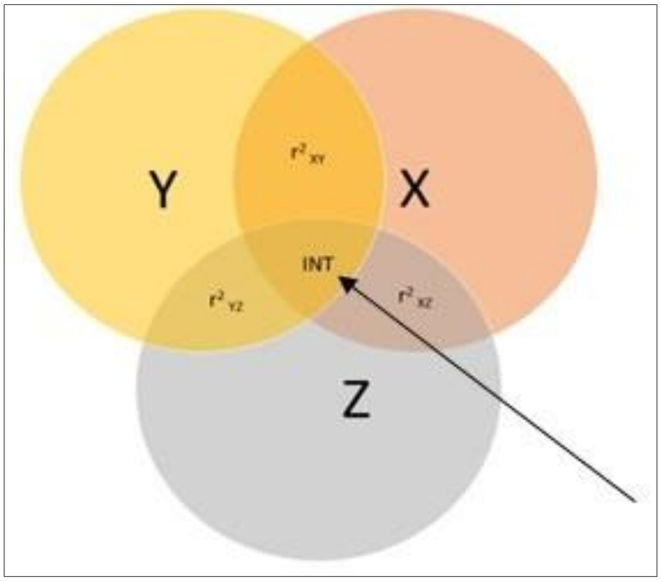
Kada pogledamo škole pojedinačno, vidimo da postoji korelacija između uratka na testu i zaključnih ocjena. Ali prosječni rezultati na testu matematike među školama se sistematski razlikuju i ovi učenici nisu jedinstvena populacija, odnosno imamo heterogen uzorak. Kada ovo ne bismo uzeli u obzir i sve rezultate ispitanika korelirali na ukupnom uzorku, dobili bismo da je ta korelacija blizu 0. Ovo bi bio pogrešan zaključak jer korelacija očigledno postoji. Kada imamo heterogene uzorke, potrebno je izračunati pojedinačne korelacije za svaki poduzorak i odrediti prosječnu korelaciju, i to tako da prvo korelacije transformiramo u Fisherove z-koeficijente, a potom dobiveni prosječni z-koeficijent transformiramo u koeficijent korelacije. Dobiveni prosječni koeficijent korelacije predstavlja reprezentantu odnosa testnih rezultata i zaključnih ocjena.

8.3.1.2 Parcijalna korelacija

Psihologija kao znanost vodi se principom parsimonije. To znači da se vodimo principom da s relativno malim setom latentnih varijabli (konstrukata ili dimenzija) pokušamo objasniti veći broj manifestnih varijabli, odnosno ponašanja. Da bi neka varijabla bila dijelom objašnjenje konstrukta, ona mora samostalno doprinositi njegovom objašnjenju. Stoga i koeficijent korelacije između dvije varijable treba ukazivati da između dvije varijable postoji povezanost, odnosno da one imaju zajedničku varijancu.

Pretpostavimo, iz didaktičkih razloga, da je povezanost između uspjeha u školi i broja sati koje dijete dnevno provede gledajući TV r= -0,40. Osim ove dvije varijable, mjerili smo i stepen agresivnosti djeteta i dobili smo da je korelacija između postignuća u školi i rezultata na testu agresivnosti r= -0,5, a korelacija između rezultata na testu agresivnosti i broja sati gledanja TV-a bila je r= 0,60.

Slika 8-4 Povezanost tri varijable – parcijalna korelacija



Vidimo da sve varijable međusobno koreliraju i njihov odnos prikazan je na slici 8-4. Imamo tri skupa koji predstavljaju varijance svake od varijable, a podskupovi su njihove zajedničke varijance i to su njihove ekskluzivne zajedničke varijance. INT (interakcija) je mjesto gdje se one sve tri preklapaju i koje nas dovodi u nedoumicu. Kada bismo posmatrali samo običnu bivarijatnu korelaciju između postignuća i agresivnosti, mi bismo joj pripisali i dio koji je pod utjecajem treće varijable. Pitamo se – ako je treća varijabla povezana i s prvom i s drugom i povećava njihovu korelaciju, te ako parcijaliziramo varijancu treće varijable, da li će ove druge dvije varijable i dalje biti u korelaciji? U našem konkretnom slučaju, možemo se pitati kolika bi bila korelacija između agresivnosti i postignuća u školi kada bismo kontrolirati utjecaj broja sati koje dijete provodi gledajući TV? Možemo se pitati i kolika bi bila korelacija između broja sati provedenih pred televizorom i postignuća u školi ako parcijaliziramo utjecaj agresivnosti?

Dakle, mi želimo parcijalizirati rezultate ispitanika u nekoj varijabli tako da „izbacimo“ onaj dio rezultata koji je u korelaciji s drugom varijablom:

$$z_{1-3} = z_1 - r_{13}z_3 \text{ ili } z_{2-3} = z_2 - r_{23}z_3.$$

U našem prvom pitanju (kolika bi bila korelacija između agresivnosti i postignuća u školi kada bismo kontrolirati utjecaj broja sati koje dijete provodi gledajući TV) prva varijabla je agresivnost, druga varijabla postignuće u školi, a treća, čiji utjecaj parcijaliziramo – broj sati gledanja TV-a. Formula za određivanje parcijalne korelacije je:

8-5

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - {r_{13}}^2}\sqrt{1 - {r_{23}}^2}}$$

U našem primjeru agresivnost je označena sa 1, postignuće u školi sa 2, a broj sati pred TV-om 3: $r_{12} = -0{,}5$; $r_{13} = -0{,}4$ i $r_{23} = 0{,}6$.

$$r_{12.3} = \frac{-0{,}5 - ((-0{,}4) * 0{,}6)}{\sqrt{1 - 0{,}16}\sqrt{1 - 0{,}36}}$$

$$r_{12.3} = -0{,}35$$

Dakle, parcijalna korelacija između agresivnosti i postignuća u školi iznosi -0,35 što je manje od njihove neparcijalizirane korelacije.

U drugom pitanju – kolika bi bila korelacija između broja sati provedenih pred televizorom i postignuća u školi ako parcijaliziramo utjecaj agresivnosti – dobili bismo sljedeći rezultat:

8-6

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1 - {r_{12}}^2}\sqrt{1 - {r_{13}}^2}}$$

$$r_{23.1} = \frac{0{,}6 - (-0{,}5 * -0{,}4)}{\sqrt{1 - 0{,}25}\sqrt{1 - 0{,}16}}$$

$$r_{23.1} = 0{,}5$$

Kao i svaki put, potrebno je da provjerimo statističku značajnost korelacije koja se u slučaju parcijalne korelacije računa preko formule:

8-7

$$t = \frac{r_{12.3}}{\sqrt{\frac{1 - {r_{12.3}}^2}{N - 1}}}$$

pri čemu su stepeni slobode df=N-3

Dakle, parcijalna korelacija je korelacija između dvije varijable kada treću varijablu držimo konstantnom. U našem prvom primjeru, parcijalna korelacija predstavlja povezanost između agresivnosti i postignuća u školi kada bi sva djeca imala podjednak broj sati gledanja TV-a, dok u drugom slučaju parcijalna korelacija predstavlja povezanost između broja sati pred TV-om i postignuća u školi kad bi sva djeca imala jednak rezultat na testu agresivnosti. Vodite računa da parcijalna korelacija ne ukazuje na kauzalnost, kao ni obična bivarijatna korelacija.

Visina parcijalne korelacije zavisi od predznaka korelacija. Ukoliko je r_{12} pozitivnog predznaka, $r_{12.3}$ će biti manja od je r_{12} ukoliko r_{13} i r_{23} imaju isti predznak. Ukoliko r_{13} i r_{23} imaju drugačije predznake, onda je obično r_{12} veća od $r_{12.3}$ (naravno ovo nije pravilo i ima primjera kada to nije tako, ali obično važi dato objašnjenje). U slučaju kada je r_{12} veća od $r_{12.3}$, pri čemu r_{13} i r_{23} imaju suprotne predznake, onda varijablu r_{23} nazivamo **supresor varijablom**. Pogledajmo primjer supresor varijable. Pretpostavimo da je neki strogi profesor dao ispit iz Psihometrije. Za svakog studenta imamo tri mjere:

- X = količina napora koji je student uložio pripremajući se za ispit
- Y = studentov rezultat na ispitu
- Z = rezultat na testu ispitanik anksioznosti koju je profesor isprovocirao kod studenta

Dobivene su sljedeće korelacije:

- r_{12} = +0,20
- r_{13} = +0,80

- r_{23} = —0,40

Zar nije čudno da je korelacija između uloženog napora i rezultata na testu tako niska? Ove korelacije nam govore da što je veći strah, to će student uložiti i veći napor (**r**$_{XZ}$= 0,80). S druge strane, što je veći strah, to će student biti slabiji na testu (**r**$_{YZ}$=—0,40). Ako parcijaliziramo utjecaj straha, dobijemo da je korelacija između napora i uratka jednaka:

$$r_{12.3} = \frac{0{,}2 + 0{,}32}{\sqrt{1 - 0{,}64}\sqrt{1 - 0{,}16}}$$

$$r_{12.3} = 0{,}94$$

Vidimo da je stvarna korelacija između napora i uratka mnogo veća od one koju smo imali na početku (što našim studentima iz primjera nije i ne mora biti neka utjeha), ali je nismo vidjeli zbog supresorskog efekta varijable straha.

8.3.1.3 Semiparcijalna korelacija

Semiparcijalna korelacija je korelacija između dvije varijable kada parcijaliziramo utjecaj treće varijable iz samo jedne od varijabli koje su nam u fokusu. U određivanju bivarijatnih korelacija više srećemo računanje parcijalnih korelacija. Međutim, u multiploj regresijskoj analizi i drugim multivarijatnim postupcima semiparcijalna korelacija postaje jako važna. Kvadrat semiparcijalne korelacije pokazuje nam koliko je poboljšanje u predviđanju rezultata kriterija kada uvedemo novu varijablu. Na primjer, želimo da predvidimo uspjeh na studiju i pri tome koristimo dvije varijable – test znanja iz psihologije i test opće informiranosti. Nakon što smo povezali test znanja iz psihologije i uspjeh na studiju, pitamo se: “Ukoliko uvedemo i test opće informiranosti, hoće li on značajno povećati predviđanje?“ Da bi odgovor bio potvrdan, test opće informiranosti mora imati značajnu samostalnu korelaciju s kriterijem bez njegove interakcije s testom znanja iz psihologije. O ovome ćemo ubrzo više pričati u poglavlju o multiploj regresiji.

Formula za određivanje semiparcijalne korelacije je:

8-8

$$r_{1(2.3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - {r_{23}}^2}}$$

Primijetite razliku u supskriptu između oznake za parcijalnu i semiparcijalnu korelaciju ($r_{12.3}\ i\ r_{1(2.3)}$). U supskriptu za parcijalnu korelaciju kažemo „odnos između jedan i dva kada smo iz obje parcijalizirali 3“, a kod semiparcijalne korelacije kažemo „odnos 1 i 2, kada smo iz dva parcijalizirali 3“.

Statistička značajnost semiparcijalne korelacije određuje se prema istoj formuli kao za parcijalnu korelaciju.

Pretpostavimo da želimo predvidjeti ocjene iz matematike na studiju. Imamo tri mjere – ocjene iz matematike kao kriterij, te rezultate na dva testa:

GPA – ocjene iz matematike

CLEP – rezultati na naprednom testu iz matematike na prijemnom ispitu

SAT – rezultati standardiziranog testa iz matematike

Želimo provjeriti da li test SAT doprinosi objašnjenju ocjena iz matematike kada iz njegovih rezultata parcijaliziramo povezanost s rezultatima CLEP testa. Dobili smo sljedeće bivarijatne korelacije na uzorku od 300 studenata:

$r_{GPA,SAT} = 0{,}72$;

$r_{GPA,CLEP} = 0{,}87$;

$r_{SAT,CLEP} = 0{,}88$

$$r_{1(2.3)} = \frac{0{,}72 - 0{,}87 * 0{,}88}{\sqrt{1 - 0{,}88^2}}$$

$$r_{1(2.3)} = -0{,}096$$

$$t = \frac{r_{12.3}}{\sqrt{\frac{1 - {r_{12.3}}^2}{N-1}}}$$

$$t = \frac{-0{,}096}{\sqrt{\frac{1 - 0{,}096^2}{300-1}}}$$

$t = -1{,}67$; $df = 297$; $p > 0{,}05$

Uvrštavanje rezultata sa SAT testa pored CLEP testa ne bi doprinijelo predviđanju ocjena iz matematike.

8.3.1.4 Multipla korelacija

Multipla korelacija predstavlja korelaciju između grupe varijabli kao prediktora i jednog kriterija. U određivanju metrijskih karakteristika testa multipla korelacija ima veoma značajnu ulogu. Spominjat ćemo je u kontekstu određivanja kriterijske valjanosti, ali i u kontekstu određivanja konstrukte valjanosti.

8.3.1.4.1 Multipla regresijska jednačina

U psihologiji rijetko možemo predviđati neko ponašanje samo na osnovu rezultata u jednom testu, odnosno na osnovu mjerenja jednog prediktora. Obično nam je potrebno više varijabli da bismo bili uspješni u predviđanju pri čemu je rezultat na kriteriju definiran kao:

$$\bar{Y} = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k$$

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e$$

Prva jednačina odnosi se na predviđanje prosječnog rezultata u kriteriju s datim rezultatima u prediktorima, dok je druga jednačina predviđanje stvarnog bruto rezultata koji uključuje e, odnosno otklon bruto rezultata od prosječnog rezultata $\bar{Y}$. Za standardizirane rezultate vrijedi jednačina:

$$\bar{z}_Y = \beta_1 z_1 + \beta_2 z_2 + . \ . \ . + \beta_k z_k$$

Koeficijente *b* u prvim jednačinama zovemo regresijskim koeficijentima ili nagibima pravca (engl. slope), dok je *a* konstanta veličina odsječka na Y osi (engl. intercept). Konstanta a govori nam o tome koliki bi bio rezultat na Y kada bi svi rezultati u prediktorima bili 0. U jednačini za standardizirane rezultate nalaze se standardizirani ponderi koje nazivamo beta ponderi (β) i time ih razlikujemo od koeficijenata učešća ili regresijskih koeficijenata bruto rezultata.

8.3.1.4.2 Slučaj sa dva prediktora

Sa dva prediktora, stvari su jednostavne i iskoristit ćemo tu jednostavnost da naučimo kako računamo beta pondere. Multipla korelacija označava se velikim slovom $R_{y.12}$. U općem obliku ova oznaka je $R_{y.1\ldots k}$, pri čemu je k – broj prediktora.

Pondere za dva prediktora računamo preko formula:

8-9

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}$$

8-10

$$\beta_1 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}$$

8-11

$$R^2 = \frac{{r_{y1}}^2 + {r_{y2}}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}$$

Primijetite da ovom formulom dobijemo kvadriranu multiplu korelaciju i da je rezultat potrebno korjenovati da bismo dobili visinu koeficijenta multiple korelacije.

U slučaju kada imamo više prediktora koristimo matrični račun (Rencher i Christensen, 2012). Vektor beta pondera određuje se kao:

8-12

$$\beta = R^{-1} * r\text{, odnosno}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad R= \begin{bmatrix} 1 & \cdots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{k1} & \cdots & 1 \end{bmatrix} \qquad r= \begin{pmatrix} r_{1y} \\ \vdots \\ r_{ky} \end{pmatrix}$$

pri čemu je R – matrica korelacija među prediktorima, a r – stupčani vektor korelacija prediktora s kriterijem. Primijetite da se u formuli koristi inverz matrice R, a ne originalna matrica. U dijelu 4.5.2.3 naučili smo što je inverz matrice. Vektor beta pondera određuje se kao umnožak inverza matrice korelacija među prediktorima i vektora korelacija prediktora s kriterijem. Ukoliko su prediktori u nultim korelacijama, tada će β ponderi biti jednaki korelacijama koje prediktor ima s kriterijem, odnosno:

$$r_{jk} = 0 \Rightarrow \beta_k = r_{ky}$$

Kako su β ponderi standardizirani regresijski koeficijenti b, regresijske koeficijente određujemo koristeći kovarijance:

b = C^{-1} *c, odnosno

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} \quad C= \begin{bmatrix} V_1 & \cdots & cov \\ \vdots & \ddots & \vdots \\ cov_{k1} & \cdots & V_k \end{bmatrix} \quad r= \begin{pmatrix} cov_{1y} \\ \vdots \\ cov_{ky} \end{pmatrix}$$

pri čemu je C – matrica kovarijanci među prediktorima a r – stupčani vektor kovarijanci prediktora s kriterijem.

Konstanta a ili odsječak na y -osi (intercept) računa se prema formuli:

8-13

$$a = M_y - \sum b_j M_{xj}$$

Pri čemu je M_y – aritmetička sredina kriterija, b – regresijski koeficijent prediktora, a M_{xj} – aritmetička sredina prediktora.

8.3.1.4.3 Određivanje koeficijenta multiple korelacije

Koeficijent multiple korelacije nije ništa drugo do povezanost između $\bar{z}_Y$ prognoziranog rezultata u kriteriju na osnovu prediktora sa z_Y kao postignutog rezultata (već znamo da uvijek predviđamo aritmetičku sredinu u kriteriju na osnovu optimalno ponderiranih rezultata u prediktorima – princip najmanjeg kvadrata).

Koeficijent multiple korelacije možemo izračunati na različite načine. Ovdje ćemo se fokusirati na tehnike potrebne za razumijevanje kasnijih složenijih priča o multiploj regresiji u okviru multivarijatnih metoda.

1. Kada poznajemo β pondere svake varijable, koeficijent multiple korelacije jednostavno izračunavamo kao zbir umnožaka β pondera svakog prediktora s koeficijentom korelacije koju prediktori imaju s kriterijem, odnosno:

8-14

$$R^2 = \sum \beta_k r_{ky} ; R = \sqrt{R^2}$$

Primijetite da ovom formulom dobivamo kvadrirani koeficijent multiple korelacije, odnosno koeficijent multiple determinacije. Da bismo odredili koeficijent korelacije, potrebno je izračunati drugi korijen dobivenog rezultata. U terminima matričnog računa (kada imamo veliki broj prediktora) koeficijent multiple korelacije računa se kao umnožak transpona vektora korelacija prediktora s kriterijem (koji tako postaje retčani vektor), inverza matrice korelacija među prediktorima i originalnog stupčanog vektora korelacija prediktora s kriterijem, odnosno kao umnožak transpona vektora prediktora s kriterijem i stupčanog vektora β pondera:

8-15

$$R^2 = r' * R^{-1} * r \text{ ili } R^2 = r' * \beta$$

2. Drugi način za određivanje multiple korelacije (izuzetno važan za razumijevanje uključivanja prediktora u regresijsku analizu) je preko sume semiparcijalnih korelacija koje prediktori imaju s kriterijem. Dakle, koeficijent multiple korelacije bit će jednak:

$$R^2 = r_{y1}^2 + r_{y(2.1)}^2 + r_{y(3.21)}^2 + \cdots + r_{y(P.123..P-1)}^2$$

$$R = \sqrt{R^2}$$

Pri čemu je:

r_{y1}^2 – kvadrirana korelacija prvog prediktora s kriterijem

$r_{y(2.1)}^2$ – kvadrirana semiparcijalna korelacija drugog prediktora s kriterijem (nakon što je iz drugog prediktora parcijalizirana zajednička varijanca s prvim prediktorom)

$r_{y(P.123..P-1)}^2$ – semiprarcijalna korelacija P prediktora s kriterijem (kada su iz ovog prediktora parcijalizirane zajedničke varijance svih drugih prediktora).

U ovom trenutku pretpostavimo da nije važno koji je prediktor u grupi bio prvi u jednačini. Ono što treba da primijetite je da je prvi prediktor i jedini prediktor koji u jednačini određivanja multiple korelacije ima puni koeficijent korelacije. U dijelu 8.3.1.6 koji se odnosi na uključivanje prediktora u multiplu regresiju, postat će važan i poredak prediktora. Zašto koristimo semiparcijalnu korelaciju, a ne parcijalnu korelaciju? Zbog toga što nam je potrebno da parcijaliziramo povezanost već uključenog prediktora u jednačinu, ali ne i povezanost sa samim kriterijem.

3. Treći način smo spominjali kada smo govorili o linearnim kombinacijama. Rezultat na kriteriju predstavlja diferencijalno ponderiranu linearnu kombinaciju prediktora, odnosno

$$\bar{z}_Y = \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_k$$

Za razliku od r, beta ponderi mogu biti veći od 1, ali treba naći set beta pondera, takvih da se zadovolji princip najmanjeg kvadrata, odnosno da $\sum(z_y - \bar{z}_Y)^2$ =minimum, odnosno da $\sum(z_y - (\beta_1 z_1 + \cdots + \beta_k z_k)^2)$ = minimum. U slučaju sa dva prediktora ova korelacija između seta prediktora kao linearne kombinacije i kriterija iznosila bi:

8-16

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}$$

8.3.1.4.4 Određivanje značajnosti koeficijenta multiple korelacije

U testiranju značajnosti koeficijenta multiple korelacije koristimo F statistiku, odnosno F omjer koji se računa prema formuli:

8-17

$$F = \frac{\frac{R^2}{k}}{\frac{(1 - R^2)}{N - k - 1}}$$

pri čemu je k – broj prediktora, R^2 – kvadrirani koeficijent multiple korelacije, a N – broj ispitanika. Stepeni slodobe određeni su ovdje i brojem prediktora i brojem ispitanika, tako da je df= k/N-k-1.

8.3.1.4.5 Određivanje statističke značajnosti razlike između dva koeficijenta multiple korelacije

Pretpostavimo da smo u nekoj selekcijskoj situaciji koristili tri prediktora da bismo predvidjeli radno ponašanje. Nakon što smo u predviđanje rezultata uveli dva prediktora u regresijsku jednačinu, pitamo se: „Da li će treći prediktor, ukoliko bude uveden u jednačinu, značajno doprinijeti povećanju koeficijenta multiple korelacije?“ Da bismo ispitali da li postoji značajna razlika između dva koeficijenta multiple korelacije, pri čemu je jedan izračunat s manjim brojem prediktora i drugi izračunat s većim brojem prediktora, koristimo sljedeću formulu:

8-18

$$F = \frac{\frac{{R^2}_b - {R^2}_a}{k_b - k_a}}{\frac{(1 - {R^2}_b)}{N - k_b - 1}}$$

Svi parametri sa „b“ u supskriptu odnose na set s većim brojem prediktora, a parametri sa „a“ na set s manjim brojem prediktora. Vidimo da i ovdje koristimo F statistiku pri čemu su stepeni slobode:

df = $(k_b - k_a)$ / $(N - k_b - 1)$

Primjer: Pretpostavimo da matricu korelacija između tri prediktora i kriterija:

Tabela 8-1 Korelacije tri prediktora i kriterija

	1	**2**	**3**	**KRITERIJ**
1	1	0,56	0,4	0,47
2	0,56	1	0,6	0,3
3	0,4	0,6	1	0,46

Zadnja kolona predstavlja vektor korelacija između tri prediktora i kriterija, dok je lijevi dio ove matrice – matrica korelacija među prediktorima.

Odredit ćemo 1) inverz matrice, 2) vektor β pondera, 3) koeficijent multiple korelacije.

Korak 1: Određivanje inverza matrice

R = $\begin{bmatrix} 1 & 0{,}56 & 0{,}4 \\ 0{,}56 & 1 & 0{,}6 \\ 0{,}4 & 0{,}6 & 1 \end{bmatrix}$

1.1. Determinanta matrice:

D = $\begin{vmatrix} 1 & 0{,}56 & 0{,}4 \\ 0{,}56 & 1 & 0{,}6 \\ 0{,}4 & 0{,}6 & 1 \end{vmatrix}$ =1* $\begin{vmatrix} 1 & 0{,}6 \\ 0{,}6 & 1 \end{vmatrix}$ – 0,56* $\begin{vmatrix} 0{,}56 & 0{,}4 \\ 0{,}6 & 1 \end{vmatrix}$ + 0,4* $\begin{vmatrix} 0{,}56 & 0{,}4 \\ 1 & 0{,}6 \end{vmatrix}$ =

=0,64 – 0,179 – 0,026 = **0,435**

1.2. Inverz matrice:

1.2.1. Matrica minora M = $\begin{bmatrix} 0{,}64 & 0{,}32 & -0{,}064 \\ 0{,}32 & 0{,}84 & 0{,}376 \\ -0{,}064 & 0{,}376 & 0{,}686 \end{bmatrix}$

1.2.2. Matrica kofaktora K= $\begin{bmatrix} 0{,}64 & -0{,}32 & 0{,}064 \\ -0{,}32 & 0{,}84 & -0{,}376 \\ -0{,}064 & -0{,}376 & 0{,}686 \end{bmatrix}$

1.2.3. Transpon matrice K' = $\begin{bmatrix} 0{,}64 & -0{,}32 & -0{,}064 \\ -0{,}32 & 0{,}84 & -0{,}376 \\ 0{,}064 & -0{,}376 & 0{,}686 \end{bmatrix}$

1.2.4. Podijelimo matricu K' s determinantom:

R $^{-1}$ = $\left[\begin{bmatrix} 0{,}64 & -0{,}32 & -0{,}064 \\ -0{,}32 & 0{,}84 & -0{,}376 \\ 0{,}064 & -0{,}376 & 0{,}686 \end{bmatrix}\right]$ / 0,435 = $\begin{bmatrix} 1{,}47 & -0{,}73 & -0{,}14 \\ -0{,}74 & -1{,}93 & -0{,}86 \\ 0{,}15 & -0{,}86 & 1{,}58 \end{bmatrix}$

$$R^{-1} = \begin{bmatrix} 1{,}47 & -0{,}73 & -0{,}14 \\ -0{,}74 & -1{,}93 & -0{,}86 \\ 0{,}15 & -0{,}86 & 1{,}58 \end{bmatrix}$$

Korak 2: Određivanje vektora β pondera:

$$\beta = \begin{bmatrix} 1{,}47 & -0{,}73 & -0{,}14 \\ -0{,}74 & -1{,}93 & -0{,}86 \\ 0{,}15 & -0{,}86 & 1{,}58 \end{bmatrix} * \begin{pmatrix} 0{,}47 \\ 0{,}3 \\ 0{,}6 \end{pmatrix} = \begin{pmatrix} 0{,}40 \\ -0{,}16 \\ 0{,}54 \end{pmatrix}$$

Korak 3: Određivanje koeficijenta multiple korelacije

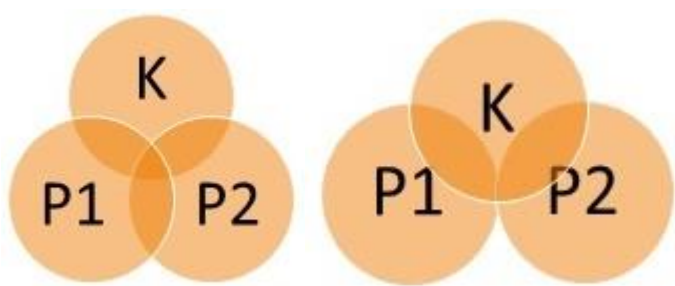
$$R^2 = r' * \beta = (0{,}47 \quad 0{,}3 \quad 0{,}6) * \begin{pmatrix} 0{,}40 \\ -0{,}16 \\ 0{,}54 \end{pmatrix} = 0{,}47 * 0{,}40 + 0{,}3 * (-0{,}16) + 0{,}6 * 0{,}54 = 0{,}386$$

8.3.1.5 Determinante koeficijenta multiple korelacije

Multipla korelacija ima osnovna načela važna za tumačenje. Osnovna svrha korištenja multiple korelacije da s jednim setom prediktora predvidimo sto je moguće više varijance u kriteriju. Osnovna načela su sljedeća:

1. Multipla korelacija će biti viša što su korelacije prediktora s kriterijem više, a korelacije među prediktorima niže. Iako već iz formula možemo vidjeti da je ovo jedno važno načelo, pogledajmo kako bi to izgledalo prikazano skupovima. Zamislimo da imamo dva prediktora i jedan kriterij pri čemu su prediktori na slici a u nultim korelacijama, a na slici b postoji korelacija:

Slika 8-5 Odnos između varijanci prediktora i kriterija kada su a) prediktori u korelaciji i b) prediktori ortogonalni



Na slici lijevo korelacije prediktora s kriterijem su jednake kao i na slici desno, odnosno njihove zajedničke varijance s kriterijem su jednake. Ali zbog korelacije među prediktorima, na slici lijevo dio zajedničke varijance „ulazi" u objašnjenje varijance prediktora samo jednom, dok na slici desno nema podskupa i povećava se proporcija objašnjene varijance za dupli dio zajedničke varijance prediktora. Ovo nije jedina prednost nekoreliranih prediktora. Druga prednost je što ne postoji dilema kojem prediktoru pripisati „zasluge" za objašnjenje varijance kriterija. Ovo će postati vrlo važno pitanje kod uključivanja prediktora u regresijsku jednačinu. Izuzetak od ovog načela su supresori i komplementarne varijable.

2. Koeficijent multiple korelacije ne može biti negativan. Statistički postupci koje koristimo za određivanje pondera podešavaju predznake tako da R bude 0 ili pozitivan (Nunnally i Bernstein, 1994, str. 187). Naravno, to znači da moramo obratiti pažnju na smjer odnosa među varijablama jer o njemu ne možemo zaključivati iz R.
3. Koeficijent multiple korelacije ne može biti niži od najviše vrijednosti korelacije koju neki prediktor ima s kriterijem. Na primjer, ako jedan prediktor korelira s kriterijem 0,6, onda nema smisla da koeficijent multiple korelacije bude niži. Logički, to bi značilo da dodavanje prediktora umanjuje mogućnost predviđanja rezultata na kriteriju.
4. Korelacije među prediktorima i beta pondere nezahvalno je određivati bez matematičkog izračuna oslanjanjem samo na grafičke prikaze. Ovo je posebno važno u situacijama kada imamo supresor varijable.
5. Kada su prediktori dobro izabrani, koeficijent multiple korelacije rijetko ima dramatičan porast dodavanjem prediktora. Osim toga, ne smijemo zaboraviti na princip parsimonije i trebamo voditi računa da ne dodajemo prediktore koji su praktično neznačajno povezani s kriterijem. Ponekada se psiholozi zaborave, neke im se varijable učine interesantnim, te ih uključuju u regresijske modele čime „kontaminiraju" objašnjenja dobivenih rezultata.
6. Multipla korelacija dobivena na uzorku sistematski je veća od korelacije u populaciji. Ova razlika se smanjuje s povećanjem uzorka ispitanika i povećanjem broja prediktora. Zašto se ovo dešava? Zbog toga što je sve u uzorku (kako smo govorili u teoriji generalizabilnosti)! Mi biramo iz veće grupe prediktora (s tim da sada imamo ograničeni broj prediktora koji su vezani za neko kriterijsko ponašanje) manji broj prediktora, i to tako da unaprijed znamo koji su prediktori najviše povezani s kriterijem. Ova „inflacija" korelacije je još veća kada prediktore biramo prema rezultatima istraživanja, a ne prema teoriji. Na pristranost odabira prediktora dolazi i greška uzorkovanja. Na primjer, u istraživanju organizacijskog kontraproduktivnog ponašanja u kojem je učestvovalo 500-600 ljudi odredili smo visinu korelacije između odabranih prediktora (savjesnosti, integriteta i lokusa kontrole) i kontraproduktivnih ponašanja (kao što je laganje, varanje, krađe, kašnjenja i slično). U selekciji nam je važno da predvidimo koliko bi kandidat bio sklon ovakvim ponašanjima. Ali mi smo izabrali neke prediktore iz čitavog skupa prediktora, te imali ispitanike koji su, koliko god mi vodili računa, samo uzorak veće populacije. Zato nam je potrebno da procijenimo kolika bi bila multipla korelacija u populaciji. Za procjenu populacijskog R koristimo formulu:

8-19

$$\widehat{R}^{\,2} = 1 - (1 - R^2)\frac{N-1}{N-k-1}$$

Pri čemu je:

$\widehat{R}^{\,2}$ – procjena populacijskog koeficijenta multiple korelacije

R – dobivena korelacija na uzorku, k – broj prediktora

8.3.1.5.1 Odnos između broja prediktora i broja ispitanika

Multipla korelacija je, kao i svaki drugi parametar, pod utjecajem uzorkovanja. Na njenu visinu utječu broj prediktora i broj ispitanika. U kakvom su odnosu broj prediktora koje uključujemo i potreban broj ispitanika da bismo mogli rezultate obrađivati koristeći multiplu regresiju? Odgovor je – broj ispitanika mora biti daleko veći od broja prediktora kako bismo dobili smislene podatke (Tabachnik i Fidell, 2007, str. 124). Kada bismo imali manje ispitanika nego prediktora dobili bismo savršeno predviđanje, ali samo zbog odnosa broja prediktora i veličine uzorka.

Green (1991, prema Tabachnik i Fidel, 2007) smatra da broj ispitanika treba biti osam puta veći ili jednak broju prediktora ($N \geq 50+ 8k$) za multiplu korelaciju. Koristan orijentir za određivanje veličine uzorka je da broj ispitanika treba biti 15 do 30 puta veći od broja prediktora. Ukoliko distribucija rezultata u kriteriju odstupa od normalne distribucije, te nisu provedene transformacije radi normalizacije, potrebno je još više ispitanika. Ako imamo nepouzdane mjere, potrebno nam je još više ispitanika. Ukoliko se koristi stepwise regresijska analiza, veličina uzorka treba biti još veća. Odnos bi trebao biti najmanje 40:1. Ukoliko namjeravamo raditi krosvalidaciju (dijeliti uzorak kako bismo na jednom dijelu uzorka odredili koeficijent korelacije, a onda provjeravali da li ćemo isti koeficijent dobiti i u drugoj polovini uzorka), jasno je da moramo imati zaista veliki uzorak.

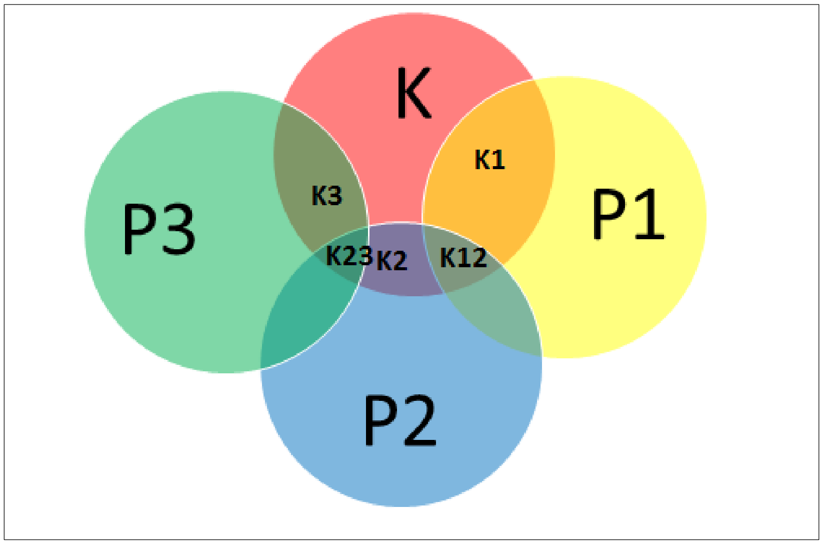
Može se desiti i da imamo previše ispitanika! Problem kod velikih uzoraka je da svaka korelacija malo veća od 0 biva značajna. Zbog toga je važno razlikovati **statističku** i **praktičnu** značajnost.

8.3.1.6 Osnovni tipovi multiple regresije

Razlikujemo tri osnovna tipa multiple regresije: standardna, hijerarhijska i statistička (stepwise) regresijska analiza. Osnovna razlika između njih je tretman zajedničke varijance prediktora u uključivanju u jednačinu. Vidjeli smo ranije da, kada imamo zajedničku varijancu prediktora (kada su prediktori u korelaciji), ne možemo tu varijancu „isparcelisati“ i dodijeliti je malo jednom, a malo drugom prediktoru. Također je ne možemo dva puta tretirati u regresijskoj analizi. Zato jedan prediktor mora biti „zakinut“ za komad varijance i u predviđanje se uključuje kroz semiparcijalnu korelaciju.

Na sljedećoj slici prikazan je odnos tri prediktora i jednog kriterija:

Slika 8-6 Odnos tri prediktora i kriterija u multiploj regresijskoj analizi

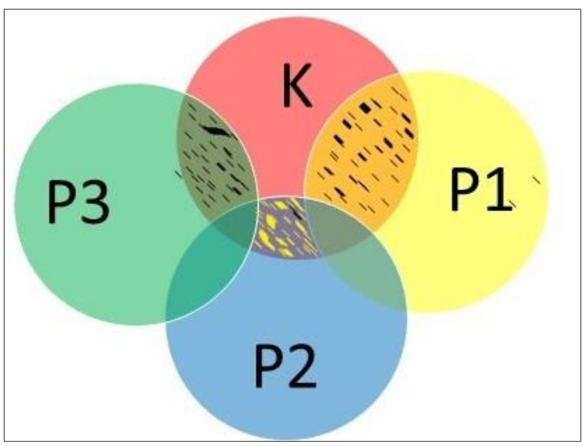


Na slici vidimo nekoliko dijelova zajedničkih varijanci. K1, K2 i K3 predstavljaju dijelove zajedničkih varijanci prediktora s kriterijem, a K12 i K23 su zajedničke varijance dva prediktora i kriterija.

8.3.1.6.1 Standardna regresijska analiza

U standardnoj regresijskoj analizi svi prediktori ulaze odjednom u regresijsku jednačinu pri čemu je svaki prediktor tretiran tako da su iz njega parcijalizirani utjecaji svih drugih prediktora (samo semiparcijalne korelacije svakog prediktora). Ako pogledate sliku 8-8, vidjet ćete da varijanca kriterija koju objasnimo standardnom regresijskom analizom predstavlja područja K1, K2 i K3, dok dijelovi K12 i K13 ostanu neobjašnjeni, odnosno koeficijent multiple determinacije odnosi se samo na „očišćene" zajedničke varijance prediktora s kriterijem.

Slika 8-7 Standardna regresijska analiza



Kod standardne regresijske analize koeficijent multiple korelacije predstavlja donju granicu procjene povezanosti rezultata prediktora i kriterija dobivenih ne nekom uzorku. Stoga se može desiti da rezultati pokažu neznačajnost prediktora P2, odnosno da dodatna količina objašnjene varijance s ovim prediktorom nije značajna, te da on ne treba ostati kao prediktor za predviđanje kriterija.

Pretpostavimo da smo željeli odrediti povezanost između tri dimenzije ličnosti (savjesnost, emocionalna stabilnost i otvorenost / intelekt) sa studentskim percepcijama efikasnosti u donošenju odluka važnih za karijeru (career decision self-efficacy ili CDSE). Također pretpostavimo da smo ispitali povezanost ova tri prediktora standardnom regresijskom analizom. Nakon što smo podatke obradili u SPSS-u, dobili smo da je koeficijent multiple regresije R=0,362. Dobivena je sljedeća tabela iz SPSS-a (ostavljena u originalnom obliku kako biste je mogli prepoznati u SPSS-ovim izvještajima):

Tabela 8-2 Koeficijent multiple korelacije u standardnoj regresijskoj analizi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.362[a]	0,13	0,12	11,56	0,13	19,99	3,00	397,00	0,00

U prvoj koloni je koeficijent korelacije R=0,362, a proporcija objašnjene varijance 0,13. F statistika nam pokazuje da je ovo značajan koeficijent korelacije, odnosno da su prediktori značajno povezani s kriterijem.

Druga tabela prikazuje β pondere (izvještaj sadrži i druge tabele, ali o njima više detalja u literaturi o statističkim programima):

Tabela 8-3 β ponderi u standardnoj regresijskoj analizi

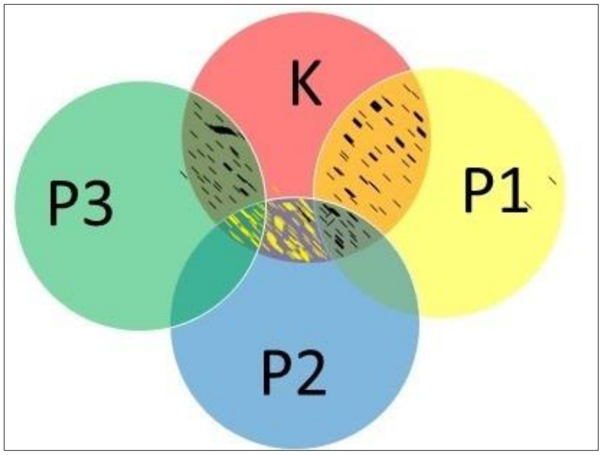
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	62,073	5,100		12,171	,000
	SAVJESNOST	,299	,097	,157	3,079	,002
	EMOCIONALNA_STABILNOST	,195	,092	,104	2,123	,034
	INTELEKT	,594	,117	,247	5,090	,000

U tabeli su dati nestandardizirani koeficijenti (b), standardna pogreška, te β ponderi i njihova značajnost. Vidimo da su svi naši prediktori značajni jer $p < 0,05$ za svaki od njih.

8.3.1.6.2 Hijerarhijska ili sekvencijalna regresijska analiza

Hijerarhijska analiza podrazumijeva uvođenje prediktora u regresijsku jednačinu prema nekom pravilu (jedan po jedan) pri čemu se koeficijent multiple korelacije određuje kao zbir semiparcijalnih korelacija prediktora s kriterijem. Semiparcijalna korelacija za svaki prediktor se računa na osnovu prediktora koji su ranije uvedeni u analizu. Da pogledamo kako bi hijerarhijska regresijska analiza izgledala u našem primjeru. Zamislimo da imamo jako teoretsko uporište prema kojem je savjesnost najvažnija za samoefikasnost u donošenju odluka u karijeri, potom emocionalna stabilnost, a posljednji po važnosti je intelekt / otvorenost za iskustvo. U hijerarhijskoj analizi prvo ćemo uvesti savjesnost (K1) pri čemu će biti korištena njegova puna korelacija sa CDSE.

Slika 8-8 Hijerarhijska regresijska analiza



Za razliku od standardne regresijske analize, ovim prediktorom će biti objašnjen i dio varijance K12 (pogledajte sliku). Zatim će biti uveden prediktor 2, odnosno emocionalna stabilnost kao semiparcijalna korelacija u koju će biti uključen i dio K23 i na kraju će biti uveden prediktor otvorenost / intelekt kojem će ostati samo njegova semiparcijalna korelacija. Pogledajmo kako bi izgledao izvještaj u SPSS-u (samo osnove tabele potrebne da razumijemo razliku). Tabela s koeficijentom korelacije više nema samo jedan koeficijent nego tri, jer je u svakom narednom koraku ili bloku uvođenja novog prediktora algoritam odredio visinu novog koeficijenta korelacije, te značajnost povećanja koeficijenta u odnosu na koeficijent korelacije s manje prediktora (o čemu smo ranije govorili):

Tabela 8-4 Koeficijenti multiple korelacije u hijerarhijskoj regresijskoj analizi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.254[a]	,065	,062	11,96064	,065	27,531	1	399	,000
2	.273[b]	,075	,070	11,91167	,010	4,288	1	398	,039
3	.362[c]	,131	,125	11,55558	,057	25,906	1	397	,000

a. Predictors: (Constant), SAVJESNOST; b. Predictors: (Constant), SAVJESNOST, EMOCIONALNA_STABILNOST
c. Predictors: (Constant), SAVJESNOST, EMOCIONALNA_STABILNOST, INTELEKT; d. Dependent Variable: CDSE_total_v

Vidimo da je prvi prediktor u korelaciji s kriterijem r=0,254, te da je ovaj koeficijent značajan. Uvođenje dodatnog prediktora značajno je povećalo koeficijent korelacije (F omjer se promijenio za 4,288 i značajan je). Nakon uvođenja trećeg prediktora došlo je ponovo do značajnog povećanja korelacije. U tabeli sa β ponderima vidimo tri dijela koja se odnose na tri sekvence ili bloka:

Tabela 8-5 β ponderi u hijerarhijskoj regresijskoj analizi

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	82,922	3,533		23,471	,000
	SAVJESNOST	,486	,093	,254	5,247	,000
2	(Constant)	79,209	3,949		20,058	,000
	SAVJESNOST	,424	,097	,221	4,365	,000
	EMOCIONALNA_STABILNOST	,196	,095	,105	2,071	,039
3	(Constant)	62,073	5,100		12,171	,000
	SAVJESNOST	,299	,097	,157	3,079	,002
	EMOCIONALNA_STABILNOST	,195	,092	,104	2,123	,034
	INTELEKT	,594	,117	,247	5,090	,000

a. Predictors: (Constant), SAVJESNOST; b. Predictors: (Constant), SAVJESNOST, EMOCIONALNA_STABILNOST;c. Predictors: (Constant), SAVJESNOST, EMOCIONALNA_STABILNOST, INTELEKT
d. Dependent Variable: CDSE_total_v

Primijetite da smo u ovom didaktičkom primjeru u drugom bloku uveli emocionalnu stabilnost koja je manje povezana s efikasnošću u donošenju karijernih odluka nego intelekt. Razlog za to je što smo pretpostavili da su „rezultati do sada ukazivali na to da emocionalna stabilnost u odnosu na otvorenost više doprinosi samoefikasnosti u donošenju karijernih odluka".

8.3.1.6.3 Statistička ili stepwise metoda

Za razliku od sekvencijalne ili hijerarhijske regresijske metode, statistička ili stepwise metoda se oslanja na statističke parametre (visine bivarijatnih korelacija prediktora s kriterijem) i tako „odlučuje" o redoslijedu uvođenja prediktora u regresijsku analizu. Kada se oslanjamo na automatski postupak, algoritam određuje redoslijed prediktora prema bivarijatnim parametrima.

Tabela 8-6 Koeficijenti multiple korelacije u stepwise regresijskoj analizi

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.297[a]	,088	,086	11,80889	,088	38,564	1	399	,000
2	.348[b]	,121	,117	11,60636	,033	15,046	1	398	,000
3	.362[c]	,131	,125	11,55558	,010	4,505	1	397	,034

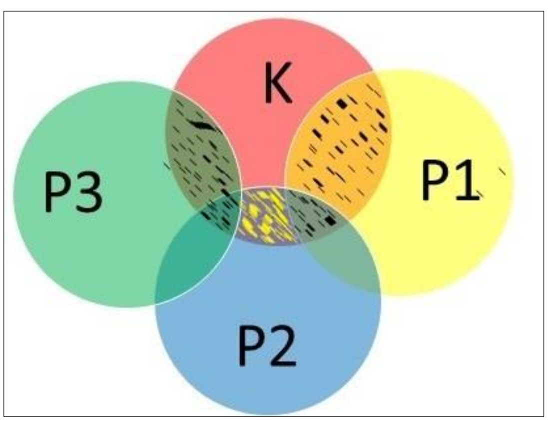
a. Predictors: (Constant), INTELEKT
b. Predictors: (Constant), INTELEKT, SAVJESNOST
c. Predictors: (Constant), INTELEKT, SAVJESNOST, EMOCIONALNA_STABILNOST
d. Dependent Variable: CDSE_total_v

Tabela 8-7 β ponderi u stepwise regresijskoj analizi

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	74,931	4,270		17,549	,000
	INTELEKT	,714	,115	,297	6,210	,000
2	(Constant)	65,749	4,818		13,646	,000
	INTELEKT	,594	,117	,247	5,073	,000
	SAVJESNOST	,361	,093	,189	3,879	,000
3	(Constant)	62,073	5,100		12,171	,000
	INTELEKT	,594	,117	,247	5,090	,000
	SAVJESNOST	,299	,097	,157	3,079	,002
	EMOCIONALNA_STABILNOST	,195	,092	,104	2,123	,034

a. Predictors: (Constant), INTELEKT
b. Predictors: (Constant), INTELEKT, SAVJESNOST
c. Predictors: (Constant), INTELEKT, SAVJESNOST, EMOCIONALNA_STABILNOST
d. Dependent Variable: CDSE_total_v

Slika 8-9 Stepwise regresijska analiza



Korištenje stepwise metode treba buditi dodatnu opreznost. Već smo govorili o pogreškama mjerenja i utjecaju pogrešaka na korelacije. Stoga, automatsko određivanje važnosti prediktora i, na osnovu toga, pripisivanje zajedničke varijance treba biti izvedeno s krajnjim oprezom.

8.3.1.6.4 Kako odlučiti koju metodu regresijske analize odabrati?

Standardna regresijska metoda je ateoretska (Tabachnik i Fidell, 2007) i korisna je kada ne znamo dovoljno o odnosima među varijablama jer je: teoretski model novi, prvi puta ga primjenjujemo u nekom kulturološkom kontekstu ili radimo provjere na novim grupama. Hijerarhijske analize su utemeljene na teoriji, te je određivanje redoslijeda uvođenja prediktora manje pod utjecajem podataka i pogreške. Stoga je hijerarhijska regresijska analiza preporučena metoda kada potvrđujemo teoretske modele. Statistička regresijska analiza je korisna kada želimo da iz istraživanja izbacimo redundantne prediktore. Kako je redoslijed uvođenja prediktora baziran na visini korelacija prediktora i kriterija i nije utemeljen na teoriji, mogli bismo reći da je ovo više eksploratorna metoda, odnosno metoda izgradnje modela.

8.3.1.7 *Važna pitanja o podacima i pripremi za multiplu regresijsku analizu*

Prije nego krenemo u analizu podataka linearnom regresijskom analizom, podaci moraju zadovoljiti neke kriterije kako bi analize bile valjane. Mada ovi kriteriji ne važe samo za multiplu regresiju, prodiskutirat ćemo o njima u okviru ovog poglavlja. U multivarijatnoj statistici odnosi više nisu tako vidljivi. Na primjer, multivarijatna regresija ne može biti vizuelno predstavljena kao bivarijatna korelacija jer podrazumijeva hiperprostor. Više se oslanjamo na brojeve i mogućnost apstrahiranja pojedinih odnosa. U tome se malo i zaboravimo, pa previdimo da multipla linearna regresija, baš kao i bivarijatna linearna regresija, zahtijeva normalitet distribucija, linearnost odnosa, homoscedascitet raspršenja, kontrolu iznimaka, i povrh svega, ne voli kolinearnost. Da vidimo koje kriterije podaci trebaju zadovoljiti da bi mogli biti korišteni u regresijskoj analizi.

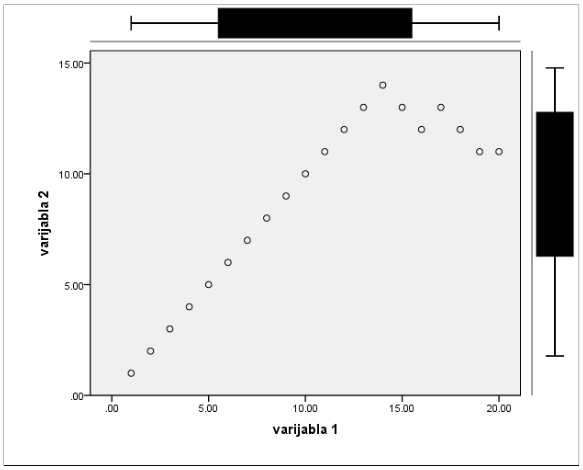
8.3.1.7.1 Normalitet, linearnost i homoscedascitet reziduala

U osnovi procedura za određivanje multiple korelacije leži pretpostavka o normalitetu distribucija. Pretpostavka o multivarijatnom normalitetu govori o tome da svaka varijabla samostalno, kao i njihove linearne kombinacije, poštuje normalnu distribuciju. Znamo da je bivarijatna korelacija vezana za parametre distribucije M i SD, a da se M i SD ne mogu računati na varijablama čija distribucija statistički značajno odstupa od normalne. Već na nivou univarijatne statistike moramo ispuniti zahtjev za normalitetom distribucija. Međutim, multivarijatni normalitet nije lako provjeravati jer bi to pretpostavljalo da provjeravamo svaku varijablu zasebno, kao i sve njihove linearne kombinacije. Stoga se oslanjamo na provjeravanje normaliteta svake pojedinačne varijable, linearnosti njihovih odnosa i homoscedasciteta (pri čemu trebamo znati da je to samo djelomična procjena).

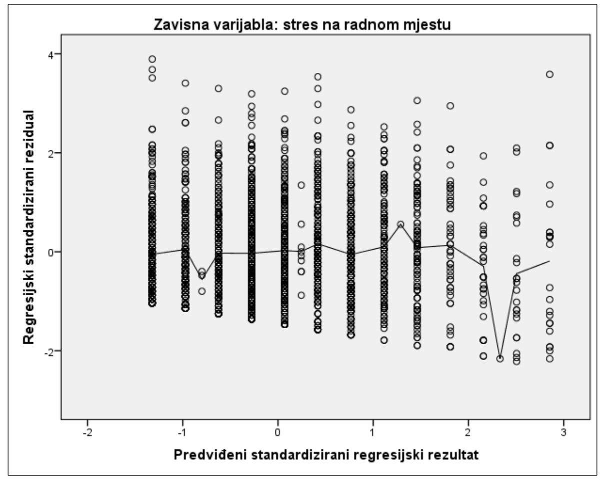
Linearni odnosi znače da su varijable tako povezane da je smjer promjene u jednoj varijabli praćen jednim smjerom promjene u drugoj varijabli, odnosno da njihov odnos možemo opisati pravom linijom. Pearsonov koeficijent korelacije ne može se računati ukoliko varijable nisu linearno povezane. Nelinearnost odnosa možemo provjeravati grafovima bivarijatnih odnosa među varijablama ili grafikonima reziduala.

Na slici 8-10 je didaktički primjer nelinearnog odnosa gdje vidimo da je porast u prvoj varijabli praćen porastom u drugoj varijabli, ali da u zoni viših rezultata na prvoj varijabli, ispitanici postižu niže rezultate u drugoj varijabli.

Slika 8-10 Primjer nelinearne povezanosti dvije varijable



Slika 8-11 Rezidualni plot



Na plotovima reziduala na X osi nalaze se standardizirani predviđeni rezultati u kriteriju, a na Y osi standardizirana odstupanja rezultata oko predviđenog rezultata (slika 8-11). Kada pretpostavka o linearnosti nije zadovoljena, tada na rezidualnom plotu vidimo da se većina reziduala nalazi ili iznad ili ispod 0. Homoscedascitet je povezan s normalitetom i on pretpostavlja da postoji podjednaka raspršenost rezultata oko crte regresije čitavom dužinom. Heteroscedascitet, kao suprotan pojam homoscedascitetu, ukazuje na to da neka od varijabli nije normalno distribuirana. Ako pogledamo sliku 8-11, vidimo da ovdje nije zadovoljena pretpostavka o homoscedascitetu. U područjima nižih rezultata reziduali se nalaze uglavnom smješteni iznad prosječne predviđene vrijednosti, o onda se taj trend mijenja u području srednjih vrijednosti.

8.3.1.7.2 Multikolinearnost i singularnost

Multikolinearnost znači visoku povezanost između prediktora, a singularnost znači potpunu povezanost među prediktorima. Multikolinearnost predstavlja i logički i matematički problem. Logički, to znači da imamo rezultate dva testa koja su mjerila istu stvar ili su mjerili varijable koje su jako visoko povezane. Postavlja se pitanje zašto bi u regresiji imali prediktore koji mjere istu stvar jer time ne doprinosimo kvaliteti predikcije. Matematički, vidjeli smo da je multipla korelacija produkt inverzne matrice korelacija, a inverz matrice je isto što i recipročna vrijednost matrice. Za singularnu matricu ne možemo izračunati recipročnu vrijednost, dok za matricu s visokom multikolinearnošću dobivamo nestabilne brojeve (Tabachnik i Fidell, 2007). Stoga je multikolinearnost nepoželjna za regresiju. Kao mjeru kolinearnosti koristimo tolerancu koja se računa kao

$$Tol_i = 1 - R^2_{i.1,.i-1,i+1,..p}$$

pri čemu je R – spuriozna korelacija između varijable za koju računamo multikolinearnost i ostalih prediktora. Mala toleranca ukazuje na visoku spurioznu korelaciju. Većina programa ima ugrađen algoritam za računanje tolerance koje je postavljena na 0,01 do 0,0001. R mora biti najmanje 0,99 da bi prediktor bio izbačen iz analize. Toleranca se može podešavati na strožije i blaže kriterije.

U svakom slučaju, potrebno je da riješimo pitanje multikolinearnog prediktora tako što neke od njih isključimo iz analize ili ih udružimo (grupiramo).

8.3.1.7.3 Iznimke

Iznimke su rezultati koji su numerički udaljeni od ostalih. To su, kako kažu Cohen, Cohen, West i Aiken (2002), atipični podaci koji nisu usklađeni s ostatkom podataka i djeluju kao da dolaze iz druge populacije. Iznimke se pojavljuju iz različitih razloga. Prva grupa razloga vezana je za pogreške u radu s podacima. Obično je to zbog pogrešnog unosa podatka ili pogrešnog kodiranja podatka, kao i pogrešne interpretacije. Na primjer, umjesto upisa 1 mi upišemo 11. Ukoliko nismo postavili zaštitu koja onemogućava unos ovakvih podataka, oni će biti registrirani i uzeti u obzir kod analiza. Druga grupa su stvarne statističke iznimke koje se pojave zbog pogreške mjerenja, ili zbog nelinearnog odnosa kojega mi nismo bili svjesni. Iznimke detektiramo pregledom grafova i statističkim određivanjem udaljenosti podatka od prosječne predviđene vrijednosti.

U jednostavnom bivarijatnom slučaju, iznimke mogu više ili manje utjecati na položaj regresijske crte. Pogledajmo neke didaktičke primjere.

Tabela 8-12 Iznimka na crti regresije

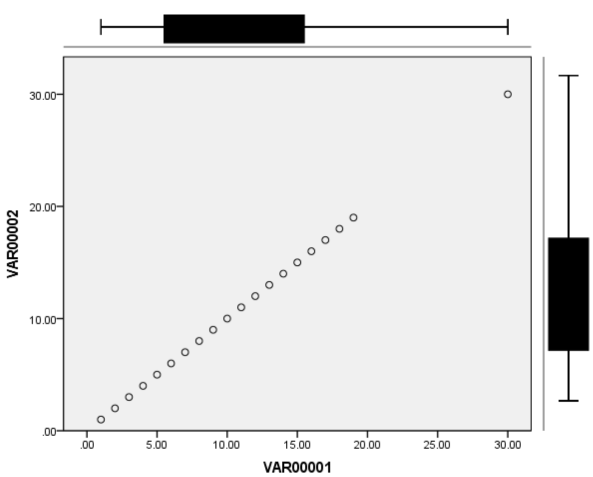
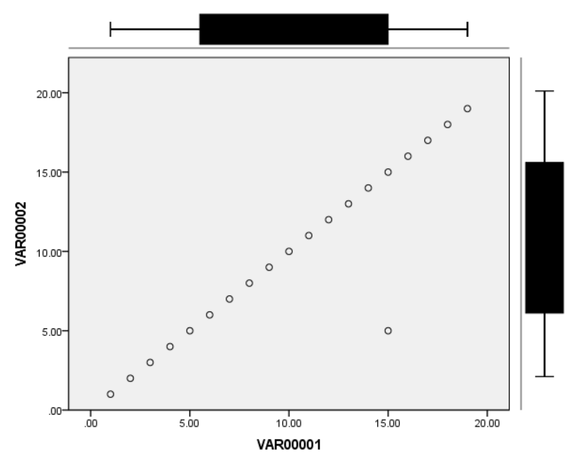


Tabela 8-13 Iznimka izvan crte regresije



Na slici 8-12 vidimo da je jedan ispitanik postigao visok rezultat i na prvoj i na drugoj varijabli, međutim njegov rezultat ne utječe na položaj crte regresije. Međutim, na slici 8-13 vidimo da je ispitanik postigao nizak rezultat na drugoj varijabli i očigledno nije „u trendu“ s ostalim rezultatima. Ova dva primjera nam pokazuju i razliku među važnim terminima u definiranju iznimaka – razlika (discrepancy), snaga (leverage) i utjecaj (influence) iznimke. Razlika pokazuje koliko je iznimka u liniji s drugima, snaga nam govori o tome koliko je iznimka daleko od ostalih ispitanika, a utjecaj je produkt razlike i snage. U prvom slučaju imamo primjer velike razlike, male snage i umjerenog utjecaja. U drugom primjeru imamo veliku razliku, umjerenu snagu i umjeren utjecaj.

Što radimo s iznimkama? Prva pomisao bi bila da ih izbrišemo iz baze podataka. Međutim prije donošenja ovako radikalne odluke, trebamo pogledati da li one tvore smislenu grupu ispitanika koja dijeli karakteristiku i da li ih ima mnogo. Prvo trebamo provjeriti da li smo pravilno unijeli podatke. Ukoliko jesmo i ukoliko su iznimke sporadični slučajevi, te nemamo objašnjenje za njihovo pojavljivanje, možemo ih brisati. Ali ukoliko ih ima više, onda je potrebno prikupiti još podataka kako bismo se uvjerili da su oni zaista iznimke. Ako se pokaže da oni tvore smislenu grupu ispitanika, trebamo ih posebno tretirati u obradama.

O iznimkama i drugim poteškoćama trebate dodatno saznati iz udžbenika multivarijatne statistike. Ovdje su date samo osnovne informacije za upoznavanje i upozoravanje da podatke moramo pregledati prije nego krenemo u obrade.

8.3.2 Povezanost s drugim varijablama – konvergentna i diskriminativna valjanost

Rekli smo na početku da se u modernom razumijevanju valjanost tretira kao konstruktna valjanost za koju se dokazi prikupljaju kroz pet tipova valjanosti. Važna grupa dokaza je povezanost s drugim varijablama. Također smo rekli da kriterijska valjanost u tradicionalnom modelu podrazumijeva povezanost s drugim varijablama, pri čemu sam konstrukt nije bio u fokusu. Kako u praksi često koristimo testove da bismo predviđali uspjeh u nekom ponašanju, tako smo u prethodnom dijelu obradili koeficijente korelacije relevantne za određivanje kriterijske valjanosti. Sada ćemo se vratiti na suvremenu priču o dokazu valjanosti kao povezanosti s drugim varijablama. Zamislimo da provjeravamo valjanost testnih rezultata za skalu Savjesnosti. Savjesnost je povezana s usmjerenošću na postignuća, uratkom u školi ili učinkom na poslu, te vođenju računa o zdravlju, a negativno je povezana s odgađanjem (prokrastinacijom). Ukoliko obrazac korelacija rezultata na testovima savjesnosti, ciljne orijentacije, zdravstvenih navika, odgađanja te mjere učinka na poslu odgovara teoretskom modelu, tada smo dobili dokaz koji podržava interpretaciju naših rezultata kao mjere savjesnosti. Ali ukoliko obrazac dobivenih korelacija ne pruži dokaz o valjanosti testnih interpretacija, tada moramo pokazati veliku dozu opreznosti u „proglašavanju" našeg upitnika – upitnikom savjesnosti. Ovdje se otvara jedno važno pitanje. Ukoliko koristimo testove koji su pripremljeni u drugim kulturama, koliko konceptualizacija konstrukta, sadržaj mjernog instrumenta i interpretacije rezultata doista odgovaraju kulturi za koju se test adaptira? Za ilustraciju ovog važnog pitanja iskoristit ćemo tzv. Paradoks japanskog samopouzdanja (Tafarodi, Shaughnessy, Yamaguchi, i Murakoshi, 2011). Jedan od najpoznatijih upitnika za mjerenje samopouzdanja je Rosenbergova skala samopouzdanja, nastala kao rezultat konceptualizacije konstrukta samopouzdanja kao odnosa prema sebi i vlastitoj vrijednosti. Pitanja u upitniku formulirana su tako da u američkoj kulturi ljudi lako odgovaraju na ova pitanja višim rezultatima. Primjeri pitanja su „Sve u svemu, zadovoljan sam sobom" ili „Mislim da imam veliki broj dobrih osobina". Kada su istraživači primijenili ovaj upitnik u Japanu, iznenadili su se kako Japanci imaju niske rezultate i kako su niski rezultati povezani sa sniženim rezultatima na samopromociji, nerealističnom optimizmu i atribucijskim pristranostima. Pri tome, zasluge za ono što im se dešava uglavnom pripisuju drugima (za razliku od Amerikanaca). Da li ovo treba biti interpretirano kako su Japanci „zakinuti" za samopouzdanje? Brown, jedan od kroskulturalnih istraživača samopouzdanja (Brown R. A., 2005) upravo je ukazao na problem valjanosti testova koji su nastali u jednoj kulturi i koriste se u drugim kulturama. Instrumenti ne mogu biti samo preneseni u drugu kulturu jer sam konstrukt ne mora imati istu konceptualizaciju i može zahtijevati drukčije instrumente i interpretacije.

U procjeni povezanosti između mjera, valjanost određujemo i kroz dokaze o konvergentnoj i diskriminativnoj valjanosti. Konvergentna valjanost je stepen u kojem rezultati testa koreliraju s rezultatima na testovima povezanih konstrukata, dok je diskriminativna valjanost stepen u kojem rezultati testa ne koreliraju s rezultatima testova konstrukata s kojima mjereni konstrukt ne bi trebao korelirati. Što to znači? Psihološki konstrukti utemeljeni su u teorijama i povezani su s nizom drugih psiholoških konstrukata. Te veze među konstruktima zovemo nomološke mreže, termin koji se odnosi na značenje koje okružuje neki konstrukt. Cronbach i Meehl (1955) u klasičnom tekstu „Construct validity in psychological tests" ističu da su osnovni principi nomološke mreže sljedeći:

1. Da bi bilo jasno što je neki konstrukt, potrebno je definirati pravila u okviru kojih se on pojavljuje. Povezani sistem pravila, koji povezuju teoriju u cjelinu, nazivamo nomološkom mrežom.
2. Pravila u nomološkoj mreži mogu se odnositi na povezanost manifestnih oblika ispoljavanja, povezanost konstrukta s manifestnim oblicima ispoljavanja ili na povezanost samih konstrukata.
3. Da bi konstrukt bio prihvaćen, mora biti dio nomološke mreže. On sam možda ne može biti viđen kroz manifestacije, ali u odnosu s drugim konstruktima omogućava predviđanje manifestnih pojava (ponašanja).
4. Upoznavanje konstrukta podrazumijeva elaboraciju nomološke mreže.
5. Obogaćivanje nomološke mreže dodavanjem konstrukata ili veza je opravdano ako generira elemente koji se mogu potvrditi kroz ponašanja ili ako će se smanjiti broj elemenata nomološke mreže potrebnih da se predviđaju ponašanja.
6. Kažemo da se dvije operacije, koje su kvalitativno različite, preklapaju onda kada ih njihova pozicija u nomološkoj mreži veže za isti konstrukt.

Dakle, dokaz o valjanosti testnih rezultata tražimo kroz njihovu povezanost s operacijama ili ponašanjima kojima su bliski (konvergencija), odnosno kroz njihovu nepovezanost s ponašanjima udaljenim u nomološkoj mreži (divergencija). Postoji više metoda za ispitivanje konvergentne i diskriminativne valjanosti. Mi ćemo naučiti jednu od njih – Matrica više crta – više metoda.

8.3.2.1 Matrica više crta – više metoda (MTMM matrica)

Jedan od najutjecajnijih članaka objavljenih u području mjerenja bio je „Konvergentna i diskriminativna validacija preko matrice više crta – više metoda" koji su objavili Campbell i Fiske (1959). U ovom članku oni su istaknuli sljedeće elemente procesa validacije:

1. Validacija je obično potvrda konvergencije neovisnih mjernih procedura. Upravo je neovisnost mjernih procedura zajednički nazivnik glavnih tipova valjanosti (osim sadržajne valjanosti).
2. Da bismo potvrdili novi konstrukt ili crtu (engl. trait), u validaciji testne interpretacije potrebno je dokazati diskriminativnu valjanost. Test ne može biti valjan ukoliko njegovi rezultati visoko koreliraju s rezultatima mjera s kojima ne bi trebali korelirati.

3. Svaki zadatak u mjerenju je jedna jedinica na koju utječu crta i metoda. Sadržaj zadatka određen je crtom koja se mjeri metodom, a koja nije specifična za sam sadržaj. Na primjer, pitanje u upitniku ekstraverzije ima sadržaj za koji se pretpostavlja da je interpretacija ekstraverzije, a metoda je skala od 1 do 5 koja se ne koristi samo za ekstraverziju nego i za druge konstrukte.
4. Da bismo ispitali diskriminativnu valjanost i procijenili relativni doprinos i crte i metode, mi u validaciji moramo koristiti više crta (radi razlikovanja crta) i više metoda (radi procjene utjecaja metode).

To nas vodi do formiranja matrice više crta – više metoda (skr. MTMM od engl. Multi trait – multi method).

Pogledajmo primjer MTMM matrice. U ovoj matrici prikazani su rezultati mjerenja šest tipova profesionalnih interesa prema Hollandovoj teoriji (Holland, 1997). Holland je tipove profesionalnih interesa vidio kao tipove ličnosti i naziva ih profesionalnim ličnostima / identitetima (vocational personalities). Prema njegovoj teoriji, postoji 6 osnovnih tipova profesionalnih interesa: realistički (R), istraživački (I), artistički (A), socijalni (S), poduzetnički (S) i konvencionalni (C). Ovi tipovi u nomološkoj mreži formiraju heksagon. Međusobno su različiti, ali između njih postoje veze, i to tako da tipovi koji su bliži u heksagonu imaju i više povezanosti. Radi ilustracije, dva susjedna tipa tipu R su C i I, dok mu je suprotan tip A s kojim bi trebao imati najniže korelacije. Ovi tipovi su naše crte. Kako bismo mjerili ove tipove, koristili smo dvije metode. Jedna je procjenjivanje koliko se ispitanicima sviđaju zanimanja koja tipično pripadaju pojedinim tipovima, a druga je koliko se ispitanicima sviđaju određene aktivnosti koje su, opet, povezane s određenim tipovima.

Tabela 8-8 MTMM matrica skala Upitnika profesionalnih interesa

		Zanimanja						**Preferencije**					
		R	I	A	S	E	C	R	I	A	S	E	C
Zanimanja	R	*0,87*											
	I	**0,57**	*0,82*										
	A	**0,26**	**0,54**	*0,84*									
	S	**0,27**	**0,49**	**0,55**	*0,82*								
	E	**0,46**	**0,43**	**0,3**	**0,62**	*0,63*							
	C	**0,67**	**0,43**	**0,17**	**0,32**	**0,63**	*0,83*						
Sviđanje	R	**0,6**	0,4	0,18	0,18	0,34	0,62	*0,76*					
	I	0,36	**0,57**	0,48	0,41	0,28	0,22	**0,41**	*0,81*				
	A	0,21	0,45	**0,72**	0,44	0,22	0,1	**0,22**	**0,66**	*0,84*			
	S	0,16	0,4	0,53	**0,65**	0,36	0,08	**0,22**	**0,59**	**0,65**	*0,74*		
	E	0,38	0,42	0,39	0,52	**0,58**	0,37	**0,47**	**0,47**	**0,46**	**0,74**	*0,77*	
	C	0,56	0,38	0,24	0,32	0,48	**0,6**	**0,69**	**0,46**	**0,32**	**0,33**	**0,58**	*0,80*

MTMM matrica nije tipična korelacijska matrica iako su korelacije u njoj najviše zastupljene. U glavnu dijagonalu se upisuju koeficijenti pouzdanosti i to je razlikuje od tipične korelacijske matrice (u kojoj su u glavnoj dijagonali jedinice). MTMM matrica ima nekoliko osnovnih dijelova:

1. Dijagonala pouzdanosti – u dijagonali pouzdanosti MTMM matrice nalaze se koeficijenti pouzdanosti
2. Dijagonale konvergentne valjanosti – male dijagonale koje povezuju istu crtu mjerenu različitim metodama između mjera. Kada imamo više metoda, tada za svaki par metoda imamo po jednu dijagonalu konvergentne valjanosti.
3. Trokutovi različita crta – ista metoda – trokutovi diskriminativne valjanosti koji sadrže korelacije između različitih crta mjerenih istom metodom
4. Trokutovi različita crta – različita metoda – trokutovi diskriminativne valjanosti koji sadrže korelacije između različitih crta mjerenih različitom metodom. Ne očekujemo potpunu nezavisnost metoda tako da ove koeficijente čitamo u relativnom odnosu s dijagonalom valjanosti.

Kako čitamo MTMM matricu? Ukoliko su testovi dizajnirani da mjere istu crtu, onda bi njihovi rezultati trebali korelirati visoko, a ukoliko se testni rezultati odnose na druge konstrukte, onda rezultati ne bi smjeli visoko korelirati. Razmotrimo još pitanje nezavisnosti mjera. I pouzdanosti i valjanost kao metrijske karakteristike zahtijevaju da postoji slaganje između mjera. Međutim, razlika u interpretaciji slaganja između mjera kod pouzdanosti i valjanosti ogleda se u tome što kod pouzdanosti mjere nisu nezavisne (sjećate se, to su paralelne mjere), dok kod ispitivanja konvergentne valjanosti govorimo o slaganju nezavisnih mjera. Cronbach i Meehl (1955) kažu kako uspješnost predviđanja rezultata na nekoj neovisnoj mjeri daje veću snagu objašnjenju konstruktne valjanosti testnih rezultata nego što je to predviđanje nekog sličnog konstrukta. Međutim, nezavisnost mjera nije binarna, to je više stepen nezavisnosti, tako da i pouzdanost i valjanost trebamo posmatrati na kontinuumu (što smo već i spominjali – test nije valjan ili nevaljan). Same metode ne moraju biti potpuno nezavisne (ne traži se od nas da u testovima imamo potpuno različite načine bodovanja), tako da se očekuje da postoji jedan dio zajedničke varijance koja dolazi zbog veza između metoda (a ne crta per se).

Konvergentna valjanost očitava se u dijagonalama valjanosti i ukoliko testovi mjere isti konstrukt, onda očekujemo da u ovim dijagonalama imamo najviše koeficijente korelacije.

Diskriminativnu valjanost očitavamo u trokutovima različita crta – ista metoda i različita crta – različita metoda. Ukoliko su konstrukti različiti, onda bi u trokutovima različite crte – iste metode trebali dobiti korelacije koje su značajno niže od onih u dijagonalama valjanosti. Međutim, kako su to crte mjerene istom metodom, očekujemo da imamo nešto više korelacije nego u trokutovima različita crta – različita metoda zbog zajedničke varijance metode. U trokutovima različita crta – različita metoda očekujemo najniže korelacije u čitavoj matrici. Ovdje nam je važan i odnos među korelacijama. Interpretacija koeficijenata u dijagonali valjanosti u apsolutnom smislu bi imala svoje opravdanje samo u nekom slučaju da su i konstrukti i metode potpuno neovisne. Tada bismo u trokutovima imali sve nulte korelacije. Ali to se u praksi nikada ne dešava. Zbog toga, tumačenje visine konvergentne valjanosti je više u duhu relativnog odnosa koeficijenata dijagonale valjanosti s koeficijentima u trokutovima.

Iako se kao najčešći razlog za procjenu neadekvatne valjanosti uzimaju niske korelacije u dijagonali valjanosti, testovi mogu „proglašeni“ nevaljanim i zbog visokih korelacija koje imaju sa drugim mjerenim crtama. Na primjer, ako u trokutovima različita crta – različita metoda imamo korelacije slične kao onima u dijagonalama valjanosti ili ako su korelacije u trokutovima različita crta – ista metoda slične visinama pouzdanosti iz glavne dijagonale, možemo reći da naš test nije dovoljno valjan jer ne razlikuje dovoljno mjerene crte ili je sama metoda značajno utjecala na zajedničku varijancu.

Da sumiramo, ono što procjenjujemo u MTMM matrici su relativni odnosi između visina korelacija u pojedinim elementima, i to:

1. Procjenjujemo prvo da li imamo pouzdane mjere. Ako u glavnoj dijagonali, koja sadrži koeficijente pouzdanosti, imamo niske vrijednosti, onda znamo da je upliv greške u rezultate bio veliki. Kako je pouzdanost potreban (ali ne i dovoljan) uvjet za valjanost, svako daljnje tumačenje valjanosti mora biti uzeto s oprezom.
2. Procjenjujemo konvergentnu valjanost tako što pregledamo visine korelacija u malim dijagonalama valjanosti koje povezuju istu crtu mjerenu različitim metodama. Ove korelacije bi trebale biti najviše od svih korelacija u matrici. Da li su korelacije zadovoljavajuće procjenjujemo u odnosu na korelacije u trokutovima.
3. Diskriminativnu valjanost ocjenjujemo preko korelacija različite crte – ista metoda i različite crte – različite metode. Korelacije u prvim trokutovima (različite crte – ista metoda) mogu biti nešto više u odnosu na korelacije u drugom trokutovima (različite crte – različite metode) zbog uključivanja metodske varijance. Ipak, ni u ovim trokutovima nećemo imati nulte korelacije jer su konstrukti obično povezani u nomološkoj mreži i jer metode ne moraju biti potpuno nezavisne.
4. MTMM matricu možemo koristiti i za procjenu različitosti metoda. Na primjer, dvije mjere bazirane na samoposmatranju nisu različite metode, dok su mjere objektivnog opažanja i skala psihosomatskih simptoma – različite metode.

Pogledajmo sada naš primjer u tabeli 8-22. Vidimo da neke skale imaju vrlo visoku pouzdanost (uzimajući u obzir da su skale samoprocjene), dok su neke (kao što je E skala zanimanja) niske pouzdanosti. U dijagonali valjanosti vidimo da imamo umjerene korelacije među mjerama istih crta, dok su najzanimljiviji podaci u trokutovima ista crta – različita metoda. Sjetite se kako smo rekli da, prema Hollandovoj teoriji, susjedni tipovi visoko koreliraju, a suprotni tipovi nisko koreliraju. Ako pogledate korelacije između R tipa sa C i I vidjet ćete da su one više od ostalih, a da je korelacija R i A najniža, što govori u prilog heksagonalnoj strukturi profesionalnih interesa. Štaviše, korelacije između R i C su više nego što su sve korelacije u dijagonali valjanosti. To nam ukazuje da nismo uspjeli u potpunosti dokazati različitost konstrukata. Razlog tome mogu biti: sličnost sadržaja pitanja, nerazumijevanje pitanja od strane ispitanika, pristranost u uzorku, te niz drugih faktora koji ukazuju da je potrebno dalje raditi na validaciji ovog upitnika. Iako bismo mogli mnogo pričati o profesionalnim interesima, naš zadatak ovdje je da naučimo tumačiti MTMM matricu tako da se nećemo upuštati u elaboracije Hollandove teorije. Preporučuje se da sami pogledate matricu i pokušate još bolje razumjeti MTMM metodu.

MTMM matrica nam daje vrlo korisne smjernice u evaluaciji konstruktne valjanosti. Pregledom utjecaja varijance crte i varijance metode na korelacije među mjerama dobivamo bolji uvid u kvalitetu rezultata dobivenih mjernim instrumentom. Psihometričari su se nastavili baviti MTMM matricom i do danas su razvili metode statističkih analiza podataka u MTMM matrici. Možda najveća zamjerka MTMM matricama je ta da nemamo mogućnost neke „čvršće“ statistike za provjeru razlika umjesto da kažemo „generalno više“ ili „vidljivo manje“. Međutim, MTMM matrica je odličan prelaz između kriterijske i konstruktne valjanosti shvaćene u okviru tradicionalnog pogleda na valjanost, odnosno između dokaza o valjanosti preko povezanosti s drugim varijablama i unutarnje strukture testa.

8.4 DOKAZI O VALJANOSTI: UNUTARNJA STRUKTURA TESTA

Unutarnja struktura testa je pitanje dimenzionalnosti testa. Ovo je ključno pitanje u konstrukciji testa, njegovoj primjeni i procjeni kvalitete. Unutarnja struktura testa nam govori kako su dijelovi testa međusobno povezani. Neki testovi sadrže pitanja koja visoko koreliraju, dok drugi sadrže pitanja koja formiraju grupe pri čemu unutar grupe imamo visoku povezanost, a između grupa nisku povezanost. Pogledajmo jednu didaktičku matricu korelacija 6 pitanja:

Tabela 8-9 Didaktička matrica korelacija šest pitanja

	P1	*P2*	*P3*	*P4*	*P5*	*P6*
P1		1	1	0	0	0
P2			1	0	0	0
P3				0	0	0
P4					1	1
P5						1
P6						

U ovoj didaktičkoj matrici vidimo da imamo dvije grupe pitanja. Unutar dvije grupe pitanja su u potpunim korelacijama, dok između grupa među pitanjima nema korelacija. Možemo pretpostaviti da prva grupa pitanja mjeri jedan konstrukt, a druga grupa pitanja mjeri drugi konstrukt. Tri su osnovna pitanja koja sebi postavljamo u vezi s dimenzionalnošću testova:

1. Koliko dimenzija mjeri test?
2. Koje je psihološko značenje tih dimenzija?
3. Da li su dimenzije povezane, odnosno korelirane?

Da bismo razumjeli odgovore na ova pitanja, da krenemo od strukture bruto rezultata.

8.4.1 Multifaktorska struktura bruto rezultata

U poglavlju o pouzdanosti vidjeli smo da je bruto rezultat sastavljen od pravog rezultata i pogreške. Međutim, to nam ništa ne govori o tome da li je pitanje jednodimenzionalno (samo jedan konstrukt utječe na pravi rezultat) ili multidimenzionalno (više konstrukata utječe na pravi rezultat). Na rezultat ispitanika prvo utječu glavne dimenzije koje taj test mjeri. Osim toga, sama metoda ili neka druga sistematski prisutna varijabla (što se u KTT tretira kao pogreška) utječe na rezultate i ona se u teoriji generalizabilnosti ne vidi kao slučajna pogreška, nego kao sistematski faktor koji formira pravi rezultat.

Kada razvijemo formulu za pravi rezultat koju smo koristili u pouzdanosti ($X_b = X_t + X_e$), imajući na umu da pravi rezultat može biti posljedica djelovanja više konstrukata i specifičnih faktora, dobijemo sljedeću formulu pravog rezultata:

$$X_t = X_c + X_s$$

Pri čemu je Xc – dio pravog rezultata koji se javlja pod djelovanjem glavnih konstrukata ili zajedničkih faktora, a Xs – dio pravog rezultata koji nastaje pod djelovanjem specifičnih faktora (npr. metode).

Rekli smo da Xc može biti formiran pod djelovanjem više dimenzija. Na primjer, ako imamo problemni zadatak iz matematike, na uradak na zadatku djeluje numerički faktor, ali i razumijevanje verbalnog sadržaja. Tako možemo dalje razviti našu formulu pri čemu je:

$$Xc = X_A + X_B + X_C + ... + X_r$$

Kada sve uvrstimo u početnu formulu za bruto rezultat, dobijemo da je bruto rezultat jednak:

8-20

$$X_b = X_A + X_B + X_C + ... + X_r + X_s + X_e$$

Vidimo da je rezultat na pitanju linearna kombinacija dijelova rezultata uvjetovanih glavnim faktorima, dijela rezultata uvjetovanog specifičnim faktorom i dijela rezultata koji nastaje zbog pogreške. Ovo je opis moguće strukture bruto rezultata i, u zavisnosti od dimenzionalnosti pitanja, ova formula će imati više ili manje elemenata koji tvore dio rezultata uvjetovan glavnim dimenzijama. Isto tako, struktura rezultata zavisi i od koeficijenata učešća svakog pojedinog faktora tako da opaženi uradak na pitanju predstavlja diferencijalno ponderiranu linearnu kombinaciju dijelova rezultata koji nastaju pod utjecajem svih spomenutih faktora. Ukoliko su dimenzije u korelaciji, tada će struktura ukupnog rezultata ispitanika na pitanju (u svom standardiziranom obliku) biti:

8-21

$$z_X = \beta_A z_A + \beta_B z_B + \cdots + \beta_R z_R + \beta_S z_S + \beta_e e$$

Vidimo da koristimo logiku multiple regresije gdje predviđamo opaženi rezultat ispitanika na osnovu njegove povezanosti s latentnim dimenzijama. Ukoliko dimenzije nisu u korelaciji, tada su $\beta = r$ (što smo naučili u prethodnim dijelovima) te će onda opaženi rezultat biti jednak:

8-22

$$z_X = r_{AX} z_A + r_{BX} z_B + \cdots + r_{RX} z_R + r_{SX} z_S + r_{eX} z_e$$

Kod pouzdanosti, varijancu smo podijelili na varijancu bruto rezultata i varijancu pogreške. Sada, varijancu možemo dodatno podijeliti. Ukupna varijanca standardiziranih rezultata iznosi 1.

Ta varijanca sastavljena je od dijelova varijance glavnih dimenzija (zajedničkih faktora), dijela varijance specifičnog faktora i dijela varijance pogreške ili:

$$1 = {\beta_A}^2 + {\beta_B}^2 + \cdots + {\beta_R}^2 + {\beta_S}^2 + {\beta_e}^2$$

Ukoliko dimenzije ne koreliraju (ortogonalne su), onda je varijanca jednaka:

$$1 = {r_{AX}}^2 + {r_{BX}}^2 + \cdots + {r_{RX}}^2 + {r_{SX}}^2 + {r_{eX}}^2$$

Nazivi pojedinih dijelova varijanci i njihovih kombinacija koje ćemo koristiti u objašnjenju rezultata analize dimenzionalnosti testa su:

1. Dio varijance opažene varijable koji nastaje pod utjecajem glavnih dimenzija ili zajedničkih faktora naziva se **komunalitet** i označava se sa h^2
2. Dio varijance koji nastaje pod utjecajem specifičnih faktora naziva se **specificitet** i označava se sa s^2
3. Dio varijance pogreške naziva se (i dalje) **varijanca pogreške** i označava se sa e^2
4. Zbir specificiteta i varijance pogreške čine **unikvitet** $U^2 = s^2 + e^2$
5. Zbir komunaliteta i specificiteta je koeficijent pouzdanosti $r_{xx} = h^2 + s^2$
6. Karakteristični korijen je suma komunaliteta ili količina varijance koju jedna dimenzija objašnjava u svim analiziranim varijablama i označava se sa slovom λ (grčko slovo lamda).

8.4.2 Faktorska analiza

Kod MTMM matrice, o konstruktnoj valjanosti testa zaključivali smo na osnovu pregleda korelacija među manifestnim varijablama. Tako nismo mogli ništa saznati koliko je svaka od latentnih varijabli doprinijela objašnjavanju rezultata u manifestnoj varijabli, niti smo mogli nešto saznati o korelacijama među latentnim varijablama. **Faktorska analiza** predstavlja statističku tehniku kojom se identificira koje varijable formiraju koherentne grupe relativno nezavisne jedne od drugih (Tabachnik i Fidell, 2007, str. 607). Drukčije rečeno, FA je tehnika koja omogućava svođenje većeg seta manifestnih varijabli na manji set latentnih varijabli koje su u osnovi testnih rezultata. Osnovna pretpostavka od koje polazimo u faktorskoj analizi je – ako između dvije varijable postoji visoka korelacija, onda postoji zajednički faktor koji djeluje na uradak i u jednoj i u drugoj varijabli. Cilj faktorske analize je pronaći zajedničke faktore koji su u osnovi manifestnih varijabli. Faktorskom analizom nastojimo odgovoriti na tri ranije spomenuta pitanja – koliko zajedničkih faktora učestvuje u formiranju testnih rezultata, koje je psihološko značenje tih faktora i kakav je odnos među faktorima? Na prvo pitanje odgovaramo kroz ekstrakcije faktora iz manifestnih varijabli i pregledom objašnjene varijance, na drugo pitanje odgovaramo pregledom povezanosti faktora s manifestnim varijablama, a na treće pitanje odgovaramo rotacijom faktora. Faktorsku analizu dijelimo na eksploratornu i konfirmatornu. U eksploratornoj FA osnovni cilj je utvrđivanje zajedničkih latentnih faktora koji leže u osnovi korelacija što većeg broja manifestnih varijabli. Njen cilj je deskriptivan, tj. faktorski opis interkorelacija što većeg broja opaženih varijabli. Konfirmatorna FA polazi od unaprijed zadanog modela strukture rezultata i broja latentnih faktora, a glavni cilj je potvrđivanje modela.

8.4.2.1 *Utvrđivanje faktora ili dimenzija*

Utvrđivanje zajedničkih faktora radi se ekstrakcijom faktora. Faktori se uvijek izvode iz manifestnih varijabli, odnosno opaženih rezultata i definiraju se kao linearne kombinacije manifestnih varijabli, a prepoznajemo postojanje zajedničkih faktora na osnovu korelacija među varijablama. Nas zanima koji su to zajednički faktori u osnovi rezultata na manifestnim varijablama (pitanjima u testu).

Uzmimo primjer ispitivanja zadovoljstva na poslu. Upitnik zadovoljstva poslom (Warr, Cook i Wall, 1979). U matrici 8-10 prikazane su korelacije 10 pitanja od kojih svako pitanje mjeri neki aspekt zadovoljstva poslom. Interkorelacijska matrica pokazuje da su ova pitanja relativno visoko korelirana među sobom, te nas zanima koliko faktora objašnjava ovu povezanost.

Tabela 8-10 Korelacijska matrica 10 pitanja Upitnika zadovoljstva poslom

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
P1	1									
P2	0,65	1,00								
P3	0,38	0,42	1,00							
P4	0,42	0,50	0,40	1,00						
P5	0,42	0,51	0,46	0,60	1,00					
P6	0,42	0,48	0,40	0,55	0,59	1,00				
P7	0,39	0,41	0,26	0,55	0,44	0,42	1,00			
P8	0,46	0,56	0,36	0,55	0,51	0,55	0,51	1,00		
P9	0,41	0,48	0,30	0,63	0,49	0,51	0,56	0,67	1,00	
P10	0,43	0,51	0,36	0,58	0,57	0,52	0,50	0,57	0,62	1

Ulaz za faktorsku analizu može biti matrica bruto rezultata ili korelacijska matrica. Postoji jedna, pomalo paradoksalna, situacija koju trebamo poznavati da bismo mogli razumjeti razliku između osnovnih metoda faktorske analize. Kažemo da faktorskom analizom provjeravamo koliko zajedničkih faktora doprinosi objašnjavanju varijabli. Zanima nas struktura komunaliteta. Ali, mi ne znamo koliki je komunalitet manifestne varijable jer ne možemo parcijalizirati varijancu manifestne varijable (za koju znamo da je ukupno 1 kada je standardizirana i da je sastavljena od komunaliteta, specificiteta i varijance pogreške, a upisana je u glavnoj dijagonali korelacijske matrice). Kada u matrici ostavimo jedinice, mi smo, ustvari, ostavili čitavu varijancu manifestne varijable i dopustili da čitava bude analizirana. Na kraju analize dobijemo faktore od kojih su neki zajednički, a neki nisu, nego su specifični i pogreška. S druge strane, možemo pokušati procijeniti komunalitet kao spurioznu korelaciju svake manifestne varijable sa svim ostalim varijablama. To znači da ćemo u glavnoj dijagonali imati kovarijancu te varijable s ostalim varijablama. U prvom slučaju, kada u interkorelacijskoj matrici imamo jedinice, matricu zovemo **kompletnom matricom,** analiziramo varijancu i metoda ekstrakcije se naziva **metoda glavnih komponenti.** Cilj metode glavnih komponenti je da maksimalno ekstrahira varijancu iz seta podataka. U drugom slučaju, kada u matrici imamo procjene komunaliteta, matricu zovemo **reduciranom matricom,** a metoda ekstrakcije naziva se **metoda zajedničkih faktora.** Cilj metode zajedničkih faktora je da reproducira korelacijsku matricu s malim brojem ortogonalnih faktora (faktora koji nisu u korelaciji). Da vidimo još neke matrice koje će nam biti važne u faktorskoj analizi (tabela preuzeta iz Tabachnik i Fidell,2007):

Tabela 8-11 Vrste matrica u faktorskoj analizi

Oznaka	Matrica	Red matrice	Rotacija	Opis matrice
R	Korelacijska matrica	pxp	Pravokuta i kosokuta	Matrica opaženih korelacija među manifestnim varijablama
Z	Matrica rezultata	Nxp	Pravokuta i kosokuta	Matrica standardiziranih opaženih rezultata
F	Matrica faktorskih rezultata	Nxm	Pravokuta i kosokuta	Matrica standardiziranih rezultata varijabli na faktorima
A	Matrica faktorske strukture Matrica faktorskog obrasca	pxm	Pravokuta Kosokuta	Matrica regresijskih koeficijenata između varijable i faktora. Ukoliko su faktori ortogonalni. onda je ovo i matrica korelacija između varijabli i faktora
B	Matrica koeficijenata faktorskih rezultata	pxm	Pravokuta i kosokuta	Matrica regresijskih koeficijenata za određivanje faktorskih rezultata iz varijabli
C	Matrica strukture	pxm	kosokuta	Matrica korelacija između varijabli i koreliranih faktora
ϕ	Korelacijska matrica faktora	mxm	kosokuta	Matrica korelacija među faktorima

N – broj ispitanika, p – broj manifestnih varijabli, m – broj faktora / komponenti

Nakon što smo odredili da li ćemo koristiti metodu glavnih komponenti ili neku od analiza zajedničkih faktora, krećemo u ekstrakciju faktora. Bez obzira na to da li koristimo jedno ili drugo, algoritam je isti. Postupak ekstrakcije je iterativan i traje sve dok ima preostale varijance u matrici. Kod analize glavnih komponenti mi donosimo odluku koje od ekstrahiranih komponenti su zajednički faktori, dok kod analize zajedničkih faktora ekstrahiramo samo kovarijancu mada i tu moramo odlučiti da li su svi faktori zajednički. Na raspolaganju imamo više algoritama za ekstrakciju. Ovdje ćemo demonstrirati jednu metodu kako bismo shvatili što se, ustvari, dešava kod ekstrakcije faktora. Ova metoda naziva se **centroidna metoda,** a pokazat ćemo je na primjeru četiri pitanja koja mjere odanost organizaciji.

8.4.2.1.1 Koraci u centroidnoj metodi

Korak 1: Zadana je kompletna korelacijska matrica R:

Tabela 8-12 Kompletna korelacijska matrica 4 pitanja za faktorsku analizu

	P1	P2	P3	P4

P1	1,00	0,18	0,50	0,07
P2	0,18	1,00	0,02	0,51
P3	0,50	0,02	1,00	-0,07
P4	0,07	0,51	-0,07	1,00
C_1	1,75	1,71	1,45	1,51
r_{iF1}	0,69	0,67	0,57	0,60

- U redu C1 saberemo sve elemente pripadajuće kolone
- Potom saberemo sve vrijednosti po kolonama (odnosno odredimo ukupnu sumu svih elemenata korelacijske matrice) T

$$T = 1{,}75 + 1{,}71 + 1{,}45 + 1{,}51 = 6{,}42$$

- Korjenujemo vrijednost T

$$\sqrt{T} = 2{,}53$$

- Zatim odredimo količnik svakog C i drugog korijena T i tako dobijemo vrijednosti korelacija svih varijabli s prvim faktorom

Korak 2: ekstrakcija varijance varijabli objašnjene prvim faktorom iz matrice korelacija

- Odredimo reproduciranu matricu tako da pomnožimo korelacije manifestnih varijabli s prvim faktorom R_{rep}

	0,69	**0,67**	**0,57**	**0,57**
0,69	0,48	0,47	0,40	0,41
0,67	0,47	0,45	0,39	0,40
0,57	0,40	0,39	0,33	0,34
0,57	0,41	0,40	0,34	0,36

- Odredimo rezidualnu matricu R_{rez} tako da od početne matrice oduzmemo reproduciranu matricu R- R_{rep}

0,52	**-0,29**	**0,11**	**-0,35**
-0,29	0,55	-0,37	0,11
0,11	-0,37	0,67	-0,41
-0,35	0,11	-0,41	0,64

- U ovoj matrici imamo negativne znakove što nam u sljedećem koraku predstavlja problem kod sabiranja korelacija. Negativni znakovi bi nam umanjili ukupnu sumu elemenata. Stoga je potrebno ovu matricu prezrcaliti ili preslikati tako da redove i kolone nekih varijabli pomnožimo sa -1. Ako množimo varijablu sa -1, to radimo i za red i za kolonu.
  U našem slučaju to su varijable 2 i 4. Ove varijable, koje preslikamo ili prezrcalimo imat će negativnu korelaciju s narednim faktorom (u ovom slučaju drugim faktorom).

	+	-	+	-
+	0,52	-0,29	0,11	-0,35
-	-0,29	0,55	-0,37	0,11
+	0,11	-0,37	0,67	-0,41
-	-0,35	0,11	-0,41	0,64

Kao rezultat dobit ćemo sljedeću korelacijsku matricu za računanje drugog faktora R_2:

	P1	**P2**	**P3**	**P4**
P1	0,52	0,29	0,11	0,35
P2	0,29	0,55	0,37	0,11
P3	0,11	0,37	0,67	0,41
P4	0,35	0,11	0,41	0,64
C_1	1,26	1,31	1,56	1,51
r_{iF1}	0,53	0,55	0,66	0,64

S ovom matricom ulazimo u novu iteraciju za računanje korelacija varijabli s drugim faktorom. U posljednje dvije kolone prikazane su vrijednosti C i korelacije varijabli s drugim faktorom. Ovaj postupak ponavljamo sve dok rezidualna matrica ne bude prazna. Osim centroidne metode postoje i druge metode koje nećemo proučavati u ovom priručniku. Centroidnu metodu smo iskoristili da razumijemo logiku ekstrakcije faktora. Pretpostavimo da smo sada „ispraznili" našu početnu matricu. Nakon što završimo s ekstrakcijom faktora, formiramo matricu faktorske strukture A. Matrica faktorske strukture je matrica regresijskih koeficijenata (ako su faktori ortogonalni onda korelacija) između varijabli i faktora.

Tabela 8-13 Matrica faktorske strukture

Pitanje	**F1**	**F2**
1	0,69	0,53
2	0,67	0,55
3	0,57	0,66
4	0,60	0,64

Za ova dva faktora možemo izračunati karakteristične korijene (u literaturi – eigenvalues) i komunalitete. Kako se u matrici nalaze korelacije varijabli s faktorima, kvadrat svake korelacije je proporcija varijance u manifestnoj varijabli objašnjena tim faktorom. Karakteristični korijen predstavlja sumu kvadriranih korelacija svih varijabli koje je objasnio taj faktor, odnosno:

$$\lambda_1 = 0{,}69^2 + 0{,}67^2 + 0{,}57^2 + 0{,}60^2 = 1{,}62$$

$$\lambda_2 = 0{,}53^2 + 0{,}55^2 + 0{,}66^2 + 0{,}64^2 = 1{,}43$$

Kako je ukupna varijanca koja je analizirana u ovom našem slučaju 4 (varijanca svake varijable je 1, imamo 4 varijable), možemo izračunati proporciju ili postotak objašnjene varijance u svim varijablama s faktorima. Prvi faktor objašnjava 1,62/4 = 0,40, odnosno 40 % varijance svih varijabli, a drugi faktor objašnjava 1,43/4 =0,36, odnosno 36 % ukupne varijance manifestnih varijabli. Kako znamo da je ekstrahirani faktor glavni ili zajednički? Prema dogovorenom standardu, ukoliko je karakteristični korijen veći od 1 (λ > 1), onda se taj faktor smatra značajnim. Ovaj kriterij zove se Kaiserov kriterij. Iako je ovaj kriterij često kritiziran (jer je arbitraran i ne mora uvijek biti jedini kriterij za određivanje koji su faktori zajednički), u standardnim paketima za obrade podataka, vrijednost karakterističnog korijena je postavljena na 1, tako da će program, ukoliko ne intervenirate, sve komponente čija je vrijednosti veća od 1 „prepoznati" kao zajedničke faktore. Ova vrijednost zaista je često neadekvatna za određivanje značajnosti faktora, posebno kada se radi o složenim, multidimenzionalnim mjerama i kod velikih uzoraka. Kako su Costelo i Osborne (2005) u svom jako korisnom članku istaknuli „..*eksploratorna faktorska analiza je složena procedura s malo apsolutnih smjernica i mnogo opcija*".

Osim toga, možemo izračunati i komunalitet varijabli kao zbir kvadriranih korelacija jedne varijable sa svim zajedničkim faktorima po principu:

$$h_1{}^2 = 0{,}69^2 + 0{,}53^2 = 0{,}76$$

Vidimo da ova dva faktora objašnjavaju 0,76 ili 76 % varijance naše prve manifestne varijable (našeg prvog pitanja).

8.4.2.1.2 Tumačenje rezultata ispitivanja dimenzionalnosti testova

Da vidimo kako izgleda faktorska analiza kada je radimo u SPSS-u na primjeru zadovoljstva poslom. Ulazna matrica je data na početku dijela o ekstrakciji faktora. U tekstu će biti prikazane najvažnije tabele kao u SPSS-u da bismo ih mogli prepoznati. Prva tabela je tabela komunaliteta:

Tabela 8-13 Komunaliteti 10 pitanja Upitnika zadovoljstva poslom

Communalities

	Initial	Extraction

zad1	1.000	.444
zad2	1.000	.561
zad3	1.000	.320
zad4	1.000	.631
zad5	1.000	.584
zad6	1.000	.555
zad7	1.000	.473
zad8	1.000	.631
zad9	1.000	.614
zad10	1.000	.605

Extraction Method: Principal Component Analysis.

U prvoj koloni vidimo da je korištena kompletna korelacijska matrica i da su kao komunaliteti korištene varijance pitanja, dok su u drugoj koloni dati komunaliteti procijenjeni na osnovu povezanosti sa zajedničkim faktorima.

Sljedeća važna tabela je pregled ukupne objašnjene varijance (engl. total variance explained) ili tabela koja prikazuje koliko komponenti je ekstrahirano iz podataka, te kolike su vrijednosti karakterističnih korjenova.

Tabela 8-14 Ukupna objašnjena varijanca za Upitnik zadovoljstva poslom

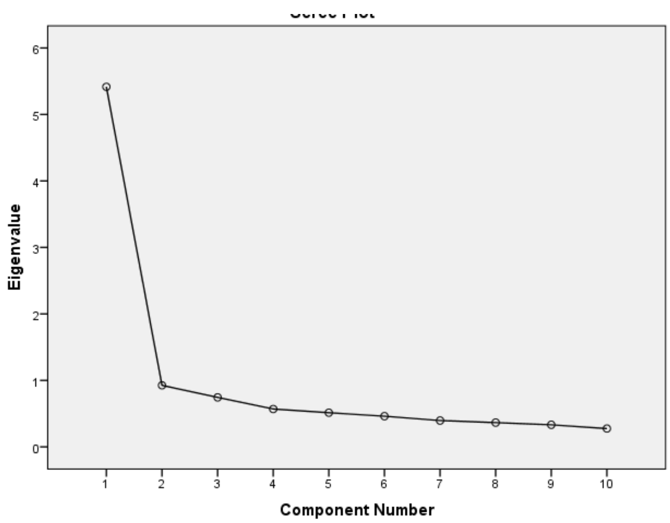
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.416	54.162	54.162	5.416	54.162	54.162
2	.926	9.258	63.419			
3	.744	7.443	70.863			
4	.570	5.697	76.560			
5	.514	5.142	81.702			
6	.461	4.612	86.314			
7	.396	3.958	90.273			
8	.365	3.651	93.924			
9	.332	3.321	97.245			
10	.276	2.755	100.000			

Extraction Method: Principal Component Analysis.

U ovom slučaju, iako je ekstrahirano 10 komponenti, samo jedna komponenta ima karakteristični korijen veći od 1. To je prva komponenta i njen karakteristični korijen je 5,416. Druga komponenta ima karakteristični korijen 0,926 i između njih je velika razlika. U drugoj koloni imamo postotak objašnjene varijance u svim varijablama pojedinim komponentama. Prva

Slika 8-14 Scree plot



komponenta objasnila je 54,162 % varijance rezultata, te možemo reći da je ovaj upitnik jednodimenzionalan. Pregled karakterističnih korjenova radimo i na osnovu Scree plota. Scree plot grafički prikazuje rezultate dobivene u gornjoj tabeli. Na Scree plotu tražimo tačku u kojoj postoji očigledan pad u vrijednosti karakterističnog korijena. Na našem Scree plotu vidimo da nakon prve komponente imamo veliki pad, što sugerira da druga komponenta nije više tako značajna. Ukoliko je pad nakon nekog drugog faktora (na primjer trećeg), onda kažemo da imamo tri zajednička faktora. Ali, kao što ni rezultati u tabli karakterističnih korjenova nisu uvijek jasni, tako i Scree plot nije uvijek jasan. U realnim situacijama dešava se da nema jasnog pada na grafikonu ili da imamo dva faktora od kojih jedan ima karakteristični korijen 1,1, a drugi 0,9. Kada znamo da su naši instrumenti uvijek opterećeni greškom, kako da napravimo razliku između ova dva faktora? Zbog ovih dilema, važno je dobro poznavati teoretski okvir i nomološku mrežu kako bismo sebi olakšali interpretaciju rezultata faktorske analize. Ukoliko ne možemo doći do jasnog odgovora na pitanje „Koliko imamo dimenzija?“ potrebno je da se dodatno angažiramo oko konstrukcije instrumenta. Na kraju, u našem didaktičkom primjeru imamo i matricu faktorske strukture:

Tabela 8-15 Matrica faktorske strukture Upitnika zadovoljstva poslom

Component Matrix[a]

	Component
	1
zad1	.666
zad2	.749
zad3	.566
zad4	.794
zad5	.764
zad6	.745
zad7	.687
zad8	.794
zad9	.783
zad10	.778

U ovoj matrici vidimo korelacije svakog pitanja s faktorom. Vidimo da je upitnik jednodimenzionalan, odnosno da mjeri samo jedan faktor.

Pogledajmo sada primjer gdje je upitnik dvodimenzionalan i gdje pretpostavljamo da faktori koreliraju međusobno. Koristit ćemo rezultate na Upitniku konflikta privatne i poslovne uloge (Carlson, Kacmar i Williams, 2000), odnosno šest pitanja iz ovog upitnika. Konflikt privatne i poslovne uloge podrazumijeva doživljaj konflikta uloga u kojem pritisak unutar radne ili porodične uloge nije kompatibilan s obavljanjem zadataka unutar druge uloge (Greenhaus i Beutell, 1985). Ovaj konflikt javlja se u dva smjera – od posla ka porodici i od porodice ka poslu, a domene konflikta su vrijeme, opterećenost i ponašanja. Pitanja koja ćemo iskoristiti u didaktičke svrhe su sljedeća:

1. Propuštam porodične aktivnosti zbog količine vremena koju moram provesti obavljajući poslovne zadatke.
2. Često sam toliko emocionalno iscrpljen kad dođem kući s posla da nisam u stanju učestvovati u porodičnim aktivnostima.
3. Aktivnosti koje me čine učinkovitim na poslu ne pomažu mi da budem bolji roditelj i suprug.
4. Propuštam aktivnosti na poslu zbog količine vremena koju moram provesti izvršavajući porodične obaveze.
5. Često sam toliko pod stresom zbog porodičnih obaveza da se ne mogu koncentrirati na radne zadatke.
6. Ponašanja koja su mi neophodna i učinkovita kod kuće ponekada su smetnja mojoj učinkovitosti na poslu.

Ako pažljivo pročitate pitanja, vidjet ćete da su prva tri namijenjena da mjere konflikt od posla prema porodici, a druga tri da mjere konflikt od porodice ka poslu. Korelacijska matrica naših pitanja izgleda ovako:

Tabela 8-16 Korelacijska matrica šest pitanja Upitnika konflikta privatne i poslovne uloge

	P1	*P2*	*P3*	*P4*	*P5*	*P6*
P1	1,00	0,59	0,39	0,22	0,21	0,19
P2	0,59	1,00	0,46	0,25	0,27	0,26
P3	0,39	0,46	1,00	0,36	0,31	0,32
P4	0,22	0,25	0,36	1,00	0,58	0,55
P5	0,21	0,27	0,31	0,58	1,00	0,70
P6	0,19	0,26	0,32	0,55	0,70	1,00

Iz korelacijske matrice vidimo da prva tri i druga tri pitanja između sebe relativno visoko koreliraju, dok su između njih niže korelacije. Međutim, to nam nije dovoljna informacija da bismo potvrdili dvodimenzionalnost upitnika, te ćemo podatke obraditi tako da ćemo koristiti faktorsku analizu, i to analizu zajedničkih faktora. Pogledajmo u nastavku rezultate:

Tabela 8-17 Komunaliteti za 6 pitanja konflikta privatne i poslovne uloge

Communalities

	Initial	Extraction
konfliktpp1	.365	.510
konfliktpp2.	.421	.673
konfliktpp3.	.301	.360
konfliktpp4	.402	.478
konfliktpp5	.546	.716
konfliktpp6	.524	.662

Extraction Method: Principal Axis Factoring.

Tabela 8-18 Ukupna objašnjena varijanca Skale konflikta privatne i poslovne uloge

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.897	48.283	48.283	2.480	41.338	41.338	2.201
2	1.320	22.000	70.283	.918	15.308	56.646	1.857
3	.611	10.186	80.468				
4	.480	7.999	88.467				
5	.392	6.538	95.005				
6	.300	4.995	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Slika 8-15 Scree plot faktora na Upitniku konflikta privatne i poslovne uloge

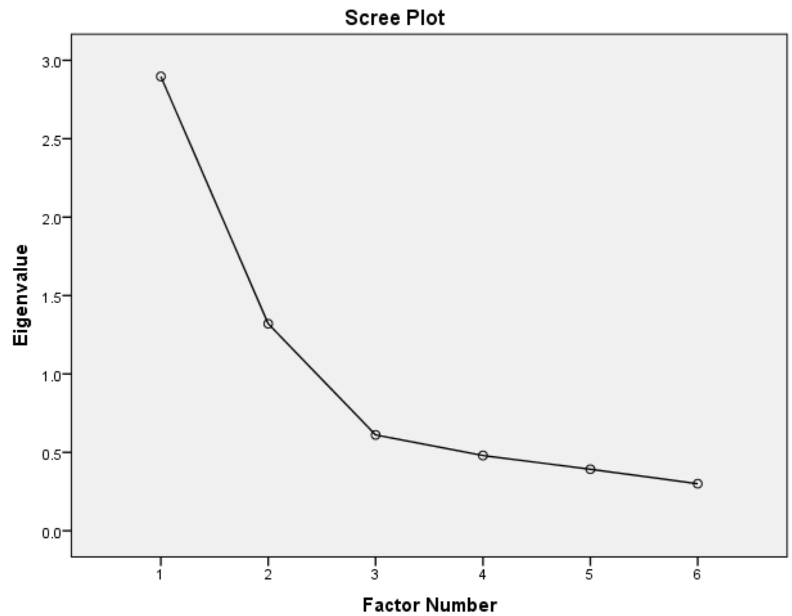


Tabela 8-19 Matrica faktorske strukture i obrasca Upitnika konflikta privatne i poslovne uloge

Factor Matrix[a]		
	Factor	
	1	2
P1	,516	,493
P2	,618	,540
P3	,560	,217
P4	,650	-,236
P5	,755	-,383
P6	,726	-,368
Extraction Method: Principal Axis Factoring.		
a. 2 factors extracted. 16 iterations required.		

Pattern Matrix[a]		
	Factor	
	1	2
P1	-,062	,739
P2	-,031	,833
P3	,206	,481
P4	,659	,068
P5	,859	-,031
P6	,826	-,029
Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization.		
a. Rotation converged in 5 iterations.		

Structure Matrix		
	Factor	
	1	2
P1	,259	,712
P2	,332	,820
P3	,415	,571
P4	,689	,355
P5	,846	,343
P6	,813	,330
Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization.		

Tabela 8-20 Matrica korelacija faktora na Upitniku konflikta privatne i poslovne uloge

Factor Correlation Matrix		
Factor	1	2
1	1.000	.435
2	.435	1.000

Extraction Method: Principal Axis Factoring; Rotation Method: Oblimin with Kaiser Normalization.

U ovom primjeru koristili smo analizu zajedničkih faktora. Analiza karakterističnih korijena pokazuje da su dva faktora značajna, te da oni zajedno objašnjavaju 70,283 % varijance manifestnih varijabli. Pregledom Scree plota vidimo da je najveći pad između prvog i drugog faktora, ali da nam Kaiserov kriterij, kombiniran s evidentnim padom između drugog i trećeg faktora omogućava da zaključimo kako se ovdje radi o dvodimenzionalnoj strukturi koju smo očekivali po teoretskom modelu.

8.4.2.2 Rotacija faktora

Kroz ovaj primjer došli smo do važnog pitanja na koje faktorska analiza treba odgovoriti – da li su faktori povezani? Povezanost između faktora često se očekuje, a koeficijente korelacija među faktorima određujemo rotacijom faktora. Nakon što smo ekstrahirali faktore i dobili matricu faktorske strukture, vidimo da imamo pitanja koja relativno visoko koreliraju s oba faktora. Pogledajmo pitanje 1. Ovo pitanje s prvim faktorom korelira 0,516, a s drugim 0,493. Ovakvi rezultati limitiraju interpretacije jer ne znamo koliki je udio svakog faktora u varijanci pravog rezultata, a ne možemo reći „Ispitanik je postigao 3 boda, od toga je 0,75 bodova za prvi faktor, a 0,70 bodova za drugi faktor“. Tako ne možemo formirati ukupne uratke u praksi niti možemo jednostavno interpretirati nalaze. Nama trebaju pitanja koja visoko koreliraju s jednim faktorom, odnosno jednodimenzionalna pitanja.

U geometrijskoj reprezentaciji, rotacija faktora podrazumijeva rotiranje sistema faktora tako da se dobije rješenje u kojem će svaki faktor zaklapati najmanji ugao sa grupom varijabli, a biti udaljen od drugih grupa varijabli. U svjetlu priče da je kosinus ugla kojeg vektori zaklapaju, u stvari, koeficijent korelacije, naš je cilj da u prostoru faktora, kao osi koordinatnog sistema, i varijabli, kao vektora, postignemo rješenje gdje će osi faktora nalijegati na grupe varijabli, sa njima zaklapati mali ugao, a istovremeno zaklapati ugao približan pravom uglu sa svim drugim grupama varijabli. To zovemo jednostavnom strukturom – pitanja visoko koreliraju s jednim, a nisko s ostalim faktorima.

Na raspolaganju imamo dvije vrste rotacija: pravokute (ortogonalne) i kosokute. Koju ćemo koristiti zavisi od teoretskog modela. Treba da znamo da sama rotacija neće popraviti osnovni aspekt faktorske analize, a to je proporcija objašnjene varijance. Pravokuta rotacija (varimax, quatrimax i equamax – tri rotacije koje su uobičajeno dostupne u programima) je rotacija koja pretpostavlja da faktori ne koreliraju među sobom, odnosno da su faktori prave koje međusobno zaklapaju pravi ugao. Iako ortogonalne rotacije imaju jednostavniju interpretaciju (jer koeficijenti u matrici faktorskog obrasca predstavljaju koeficijente korelacije), moramo voditi računa o pretpostavkama njihove povezanosti u nomološkoj mreži. Ako se povezanost očekuje, onda nema opravdanja za korištenje ortogonalne rotacije. U psihologiji, mi obično očekujemo da faktori koreliraju i korištenje ortogonalnih rotacija je gubitak dragocjene informacije o korelaciji faktora. Kosokute rotacije su malo kompleksnije od ortogonalnih jer se interpretacije vežu za matricu obrasca i matricu strukture. Postoji razne vrste kosokutih rotacija. Neke od njih su promax i oblimin.

Nakon što smo za naš primjer uradili kosokutu rotaciju (prema teoretskom okviru, ova dva tipa konflikta koreliraju), dobili smo matricu faktorskog obrasca (pattern matrica). Ovo je matrica koja prikazuje najjednostavnije rješenje odnosa faktora i varijabli koje se može postići datim podacima. Usporedite podatke u matrici faktorske strukture i matrici faktorskog obrasca (tabela 8-19) i vidjet ćete da se one razlikuju. Na primjer, u matrici faktorskog obrasca sada imamo pitanje 1 koje je u gotovo nultoj korelaciji s prvim faktorom (-0,062) i u visokoj korelaciji s drugim faktorom (0,739).

Osim pregleda podataka koji se odnose na povezanost varijabli s faktorima, dobili smo podatke i o povezanosti samih faktora. To je posljednja matrica koju označavamo sa ϕ. Vidimo da su faktori korelirani (0,435) iako ne tako visoko da bismo posumnjali u dvodimenzionalnost konstrukta.

8.4.2.3 Psihološka interpretacija faktora

Nakon što smo završili postupak ekstrakcije i rotacije faktora, ostaje nam da psihološki interpretiramo dobivene rezultate, odnosno damo imena faktorima. To radimo pregledom povezanosti varijabli s pojedinim faktorima. Ovo je ujedno i dio u kojem „svodimo" objašnjavanje većeg broja manifestnih varijabli na manji broj latentnih dimenzija.

U pregledu povezanosti pitanja i faktora obraćamo pažnju na nekoliko stvari. Na samom početku, pri konstrukciji skale, mi smo uzorkovali pitanja na osnovu toga koliko svojim sadržajem reprezentiraju određeni konstrukt. Faktorska analiza nam služi da potvrdimo da je naša interpretacija rezultata ispravna. Iako informacije o valjanosti testnih rezultata prikupljamo od samog početka, povezanost između pitanja i faktora je vrlo važan dio razumijevanja kvalitete testa.

Prvo trebamo utvrditi da li su pitanja za koja smo pretpostavljali da će pripadati istom faktoru zaista i povezana s tim faktorom. U našem primjeru vidimo da je većina pitanja povezana s pripadajućim faktorom, odnosno da su 1) zajedno grupirana i 2) vezana za isti faktor. Pregledom povezanosti pitanja s faktorima, posebno onih koja imaju najviše koeficijente, počinjemo razumijevati psihološko značenje faktora. U interpretaciji povezanosti pitanja s faktorom važne su nam dvije informacije. Prva informacija je visina koeficijenata koja se kreće od -1 do 1. Što je koeficijent zasićenja viši, to je snažnija veza pitanja i faktora. Faktorska zasićenja preko 0,3 ili 0,4 nisu zanemariva (ali ovaj nivo je relativan) i ne smijemo „zažmiriti" na činjenicu da neko pitanje ima faktorsko zasićenje na dva faktora preko 0,4. Faktorska zasićenja preko 0,7 su visoka i ta pitanja tretiraju se kao visoko povezana sa svojim faktorom. Druga važna informacija je predznak faktorskog zasićenja – da li je koeficijent pozitivan ili negativan. Ukoliko je predznak pozitivan, to ukazuje da ispitanici koji su imali visok rezultat na testu imaju i više izražen mjereni faktor, a negativan predznak znači da visok rezultat na testu indicira nisko izražen mjereni faktor. U našem primjeru, pogledamo visinu koeficijenata zasićenja i vidimo da većina naših pitanja ima visoke koeficijente zasićenja s jednim faktorom i niske s drugim, osim trećeg pitanja. Treće pitanje ima veće zasićenje na drugom faktoru (0,418) i, iako ovaj koeficijent nije zanemariv, ipak je nisko. Osim toga, ovo je jedino pitanje koje ima nešto viši koeficijent zasićenja s nepripadajućim faktorom (iako niže od 0,3).

Drugo, trebamo dati imena faktorima. Vidimo da su prva tri pitanja vezana za drugi faktor. Prema sadržaju pitanja vidimo da je to „Konflikt posao – porodica". Drugi faktor je „Konflikt porodica – posao".

Problemi na koje možemo naići u interpretaciji povezanosti pitanja i faktora su 1) da je pitanje vezano za jedan faktor, ali jako nisko ili 2) da ima visoku povezanost s više faktora. U realnom istraživanju, ovo su česte situacije i psiholog mora odlučiti da li će ovo pitanje ostaviti u skali i pribrajati njegov rezultat ostalim rezultatima ili će se dodatno „pozabaviti" kvalitetom ovog pitanja. Da li ovo pitanje pripada ovom upitniku? Da li možda ima neko drugo faktorsko rješenje, npr. da li ima samo jedan faktor koji bi mogao objasniti rezultate na svim pitanjima? Kada se suočimo s ovim dilemama u eksploratornoj faktorskoj analizi, potrebno je dodatno promisliti. Možemo probati neko drugo faktorsko rješenje, ali pri tome ne smijemo zaboraviti teoriju od koje smo pošli (bez obzira na to što je eksploratorna analiza). Ukoliko ne uspijemo postići bolje faktorsko rješenje, pitanje možemo isključiti iz testa jer – ili ne pripada ovom testu, ili je varijanca pogreške prevelika. Zašto se ovo desilo – odgovor možemo potražiti i u analizi sadržaja pitanja. Možda je pitanje bilo nejasno, možda su ga ispitanici drukčije razumjeli i slično. U svakom slučaju, naša preporuka je da se nastavi s dizajniranjem pitanja kako bi ono bilo što jasnije interpretirano.

8.4.3 Konfirmatorna faktorska analiza

Eksploratorna faktorska analiza o kojoj smo govorili u prethodnom dijelu predstavlja analizu unutarnje strukture pri čemu, iako imamo teoretske pretpostavke o broju faktora koji utječu na rezultate, dopuštamo da u faktorskom rješenju može biti neki drugi broj faktora i da pitanja mogu biti povezana s bilo kojim faktorom. Glavni cilj eksploratorne FA je utvrditi broj faktora i povezanost pitanja s faktorima. U konfirmatornoj faktorskoj analizi cilj je drukčiji. Konfirmatornu analizu koristimo kada imamo jasne pretpostavke o dimenzionalnosti testa, a cilj je potvrditi da podaci odgovaraju pretpostavljenom modelu. U provjeri unutarnje strukture konfirmatornim modelom unaprijed specificiramo broj faktora i povezanost pitanja s faktorima. Osim toga, unaprijed specificiramo i kovarijance među pitanjima, kao i kovarijance među latentnim dimenzijama (Brown T. A., 2006). Za konfirmatornu FA moramo imati vrlo jasnu empirijsku ili konceptualnu bazu na osnovu koje tvorimo model. S obzirom na kompleksnost obrade podataka u konfirmatornoj FA koja prevazilazi namjere ovog priručnika, mi ćemo se baviti logikom i razumijevanjem, dok za daljnje učenje preporučujemo knjige kao što su Brown (2006) „*Confirmatory Factor Analysis for Applied Research*" i udžebnike o strukturalnom modeliranju (Byrne, 2010).

Pretpostavljeni model odnosa latentnih i manifestnih varijabli u konfirmatornoj FA nazivamo mjerni model. Mjerni model može biti zadan ukoliko koristimo test koji je već ranije konstruiran, a mi želimo provjeriti da li naši podaci odgovaraju modelu, a može ga pripremiti psiholog koji konstruira test i može biti baziran na teoriji koja se koristi za konstrukciju.

Na primjer, želimo ispitati valjanost postojeće Skale odanosti organizaciji (Meyer i Allen, 1997) koja prema teoretskom okviru ima tri facete: afilijativnu, instrumentalnu i normativnu[1]. Isto tako, vođeni teoretskim modelom mogli smo odlučiti da sami pripremimo skalu. Ukoliko se odlučimo za korištenje postojećeg testa, znamo da je upitnik zamišljen da mjeri tri dimenzije, te da je svaka dimenzija mjerena sa 8 pitanja. Nakon pripreme pitanja i prikupljanja podataka, daljnji koraci u

[1] Preporučujemo članak u kojem se upravo debatira na temu valjanosti ove skale, a koji je napisao Jaros (2007).

konfirmatornoj analizi su određivanje kovarijančne matrice podataka, procjena parametara u modelu, izračunavanje indikatora slaganja podataka s postavljenim modelom, pregled indeksa modifikacije, donošenje odluke o modificiranju početnog modela i (povratak na početnu tačku) postavljanje modificiranog modela za provjeru. Vrlo kratko ćemo spomenuti što svaki od ovih koraka uključuje.

8.4.3.1 Specifikacija modela

Za konfirmatornu analizu koriste se statistički paketi kao što su AMOS, R, LISREL i drugi. Neki od njih, kao što je AMOS, omogućavaju da se model grafički specificira tako što latentne varijable prikazujemo elipsama, a manifestne varijable pravougaonicima. Između njih se definiraju veze (regresijske linije i kovarijance).

Prilikom specificiranja modela vodimo se načelom kongeneričnosti – da je pitanje vezano samo za jedan faktor i da svaka dimenzija utječe na makar jedno pitanje (manifestnu varijablu). Ukoliko pretpostavljamo korelaciju među latentnim dimenzijama, moramo je specificirati. Važno je napomenuti da će sve što smo specificirali biti provjeravano, ali i ono što nije specificirano bit će testirano. Ukoliko rezultati analize pokažu da bi podaci bolje odgovarali modelu s dodatnim vezama, program će to prijaviti.

8.4.3.2 Određivanje usklađenosti modela

Proces provjere usklađenosti baziran je na tome da algoritam provjerava koliko podaci koje smo prikupili podržavaju postavljeni model. Proces procjene usklađenosti modela najjednostavnije je objasniti uz pomoć matrične algebre. Matrica S predstavlja kovarijančnu matricu u uzorku (kovarijančna matrica naših opaženih rezultata), ∑ je kovarijančna matrica u populaciji, dok je Θ (teta) vektor parametara modela. U skalu s tim ∑(Θ) reducirana je kovarijančna matrica koju model pretpostavlja (specifična struktura koju model predlaže). U provjeri se polazi od nul-hipoteze (H0) koja glasi da pretpostavljeni model odgovara populaciji, odnosno može se generalizirati na populaciju (∑=∑(Θ)). Za razliku od tradicionalnih modela, u kojima istraživač očekuje da će odbaciti nul-hipotezu, ovdje se nadamo da ćemo potvrditi nul-hipotezu, odnosno da će naši rezultati biti u skladu s pretpostavljenim modelom. Osnovni fokus u procesu procjenjivanja (iterativnom postupku procjene modela) je da se dođe do vrijednosti parametara tako da diskrepanca (rezidual) između kovarijančne matrice dobivene na uzroku S i kovarijančne matrice populacije koju model pretpostavlja ∑(Θ) bude minimalna (Byrne, 2010).

Nakon što se završi postupak provjere usklađenosti, program nas izvještava o rezultatima. U tom izvještaju nalaze se:

1. Indikatori usklađenosti modela s podacima – indikatori usklađenosti govore o tome koliko su naši podaci u skladu s modelom. Ukoliko su indikatori usklađenosti u vrijednostima koje su propisane, onda kažemo da je model „usklađen". Ukoliko nisu, to znači da postavljeni model nije konzistentan s podacima. Ovdje je važno imati na umu kako se konstruktna valjanost odnosi na testne rezultate i njihove interpretacije. Ako se pokaže slaba usklađenost podataka s modelom, ne možemo sebi dopustiti da mijenjamo model

dok ne dobijemo dobru usklađenost jer bi to značilo da su važniji podaci nego teorija (a mi znamo kako podaci mogu biti opterećeni greškom i pristranošću).

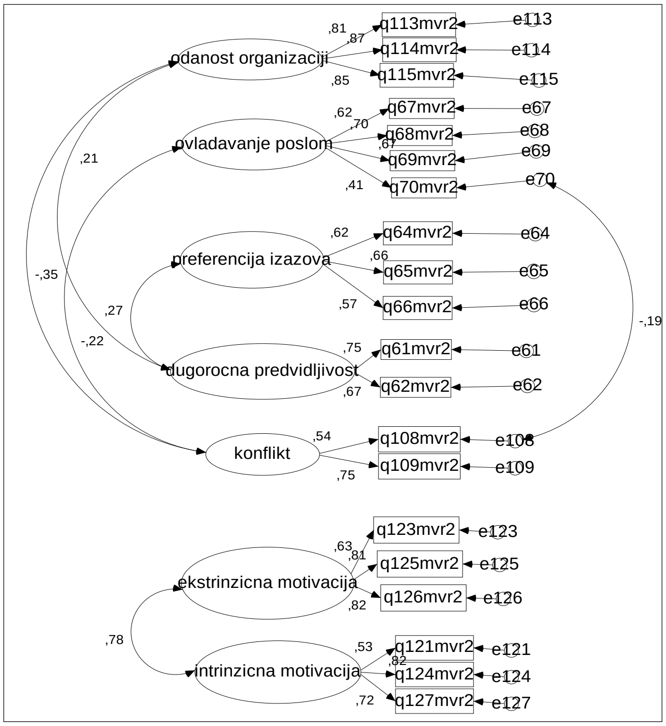
2. Procjene parametara koji su testirani – vrlo bitna informacija je vrijednost svih parametara koje smo pretpostavili u modelu. U ovom dijelu izvještaja saznajemo koliki su regresijski koeficijenti pitanja s latentnim dimenzijama, te kolike su korelacije među latentnim dimenzijama.
3. Modifikacijski indikatori – osim što procijenimo sve zadane parametre i usklađenost modela s podacima, program nam ponudi i spisak mogućih modifikacija koje bi popravile model. Na primjer, ukoliko je neko pitanje multifaktorsko, program „ponudi" da to pitanje povežemo i sa drugim faktorom kako bismo dobili veću usklađenost. Da li ćemo neke od ovih indikatora iskoristiti zavisi od toga da li za njih možemo naći teoretsko opravdanje. Ukoliko nemamo teoretskog opravdanja, onda je bolje provjeriti da li bi neki od alternativnih modela bolje opisao podatke (na primjer, definirali smo trodimenzionalni model, a možda bi uvođenje dodatne dimenzije popravilo usklađenost). Ipak, u donošenju odluka moramo biti vođeni teorijom jer program može uraditi što god zamislimo, ali da li je to što smo zamislili i valjano – to je pravo pitanje.
4. Nakon što pregledamo sve dobivene rezultate, odlučujemo se o respecifikaciji modela i vraćamo se na početnu tačku.

Radi ilustracije pogledajmo jedan primjer malo složenijeg modela kojim se određivala faktorska struktura nekoliko skala individualnih dimenzija koje utječu na radnike na poslu. U konfirmatornu faktorsku analizu uključeno je 7 skala: odanost organizaciji, percepcija ovladavanja poslom, preferencija za izazovima na poslu, predvidljivost posla, konflikt uloga, ekstrinzična i intrinzična motivacija.

Na desnoj strani se nalaze pravougaonici i to su pitanja. Svaki pravougaonik je jedno pitanje.

Na desnoj strani su elipse i to su latentne dimenzije. Latentne dimenzije su povezane s pitanjima. Primijetite smjer strelica između latentnih dimenzija i pitanja – one idu od dimenzije prema pitanju sugerirajući upravo da nije pitanje to koje tvori dimenziju, nego dimenzija tvori odgovor na pitanje.

U krajnjoj desnoj strani modela nalaze se mali krugovi – to su greške mjerenja i ovo je jedna od najvećih prednosti konfirmatorne analize – možemo da vidimo grešku i da računamo varijancu greške kao i njihove kovarijance. Između latentnih varijabli prikazane su veze, odnosno kovarijance. One su uvijek prikazane strelicama dvostrukog smjera. Nakon završenog izračuna u modelu možemo vidjeti i visinu svih koeficijenata. Na našem primjeru vidimo da sva pitanja imaju relativno visoke regresijske koeficijente. Također vidimo i korelacije među latentnim dimenzijama koje ukazuju da su one povezane, ali ipak različite. Na ovom modelu možemo vidjeti još jednu kovarijancu koju ne možemo vidjeti u eksploratornoj FA – kovarijancu među greškama. To je upravo ono o čemu smo pričali u teoriji generalizabilnosti – nije svaka greška slučajna, jedan dio greške se javlja i pod utjecajem neke sistematske varijable, kao što je metoda. Konfirmatorna FA nam omogućava mnogo kvalitetnije analize, ali to ne znači da isključuje eksploratornu FA. I jedna i druga imaju svoje mjesto, samo je važno znati zašto služi jedna, a zašto druga.



Slika 8-16 Mjerni model Skala psihosocijalnih faktora na radnom mjestu – individualni nivo

Da sumiramo, unutarnja struktura testa je jedan od najvažnijih dokaza konstruktne valjanosti. Postupcima za ispitivanje unutarnje strukture, kao što su MTMM matrica i različiti oblici faktorskih analiza, dokazujemo da je interpretacija testnih rezultata u skladu s latentnim dimenzijama za koje smo namijenili mjere. Bez dokaza o unutarnjoj strukturi testa, mi ne možemo reći koje dimenzije se mjere testom niti ga možemo upotrebljavati u praksi za dijagnostiku i prognoziranje rezultata na drugim varijablama.

8.5 Dokaz o valjanosti: psihološki procesi pri odgovaranju

Dokazi o psihološkim procesima ušli su 1999. u Standarde za pedagoško i psihološko testiranje kao vrsta dokaza da su testni rezultati valjani. Što to znači? Interpretacije konstrukata često sadrže eksplicitne izjave o psihološkim procesima koji leže u osnovi ponašanja. Tako, na primjer, kažemo da test matematičkog rezoniranja mjeri korištenje matematičkih modela u rješavanju zadataka. Međutim, konstruktori testa moraju empirijski pokazati da u osnovi rješavanja testa zaista leže procesi matematizacije, a ne korištenje naučene sheme po kojoj se zadaci rješavaju. Ovo je dobar primjer problema s kojim se srećemo u našem obrazovnom procesu kada se testovi pripreme tako da se vjeruje da ispituju razumijevanje matematičkih modela, a oni samo mjere vještost učenika da primjene već rutinski obrazac rješavanja zadataka. Jasno nam je da je prikupljanje dokaza o psihološkom procesu dosta kompleksan posao, ali jako važan kako bismo bili sigurni da latentne dimenzije nazivamo pravim imenima. Prikupljanje dokaza o psihološkim procesima obično se radi zahtjevnim analizama individualnih rezultata. Jedna od metoda je da uzorku ispitanika, koji je sličan populaciji za koju je test namijenjen, damo test, a onda od njih tražimo da „razmišljaju naglas“, odnosno da pričaju sve što im pada na pamet kada rješavaju zadatak ili odgovaraju na pitanje. Pri tome se monolog snima i kasnije se analiziraju procesi na osnovu kojih je ispitanik došao do odgovora. Za neke konstrukte, posebno u eksperimentalnim istraživanjima, relevantna informacija može biti i vrijeme latencije (vrijeme od zadavanja pitanja do produciranja odgovora) ili pokreti očiju. Ispitivanje psiholoških procesa može otkriti razlike u testnim interpretacijama između grupa ispitanika. Na primjer, na Zungovoj skali depresivnosti DPR-S (Zung, 1965) u kliničkoj populaciji autori smatraju da mjeri jedan faktor. Depresivni simptomi se pojavljuju i kao ishodi nepovoljne radne klime, tako da ovaj upitnik koriste organizacijski psiholozi da bi utvrdili prisustvo depresivnih simptoma kod radnika. Ali, na faktorizaciji upitnika mi smo dobili dva faktora. Što bi moglo objasniti ovakav nalaz? Pa upravo različiti psihološki procesi koji se nalaze iza odgovaranja u kliničkoj i nekliničkoj populaciji. U kliničkoj populaciji osnova odgovaranja na pitanja je depresija kao poremećaj raspoloženja koji zahvata sve aspekte života. U nekliničkoj populaciji prema Schonfeldu (1992), skale depresivnih simptoma mjere dva konstrukta: nespecifični psihološki stres i kliničku depresiju. Dohrenwend i autori (prema Schonfeld, 1992) su u svojim istraživanjima pokazali da depresivni simptomi koreliraju s nizom skala koje mjere druge simptome, i to onoliko koliko im koeficijenti pouzdanosti dopuštaju, te u slučaju nekliničkih uzoraka skale depresije vjerovatno mjere konstrukt nazvan nespecifični psihološki stres ili demoralizacija. Nespecifični stres je vrsta psihološkog stanja koje bi moglo motivirati osobe bez klasične depresivne simptomatologije da potraže pomoć psihoterapeuta. To su osobe koje imaju problema u psihološkom funkcioniranju, ali ne zadovoljavaju DSM-IV kriterije za depresiju. Da smo prije studije na nekliničkoj populaciji proveli ispitivanje psiholoških procesa, utvrdili bismo vjerovatno da radnici drukčije rezonuju kada odgovaraju na pitanja koja su se odnosila na nespecifični psihološki stres. Psihološki proces kao dokaz valjanosti nije nužno pripisan ispitanicima. Ukoliko imamo situaciju da sudije (nastavnici, procjenjivači) ocjenjuju uradak ispitanika, tada nam je važnije da vidimo koji su psihološki procesi u pozadini kada procjenjivač ocjenjuje ispitanika.

Psihološki procesi su važna vrsta dokaza o valjanosti instrumenta kada ispitanici daju odgovore na pitanja otvorenog tipa, odnosno kada na testovima nema ponuđenog odgovora. U testovima s ponuđenim odgovorima ovaj dokaz ne mora biti toliko važan koliko je važan sam produkt (tačan odgovor). Ali, i dalje ostaje važno kako je ispitanik razmišljao dok je tražio rješenje, posebno u slučaju pogrešnih odgovora.

8.6 DOKAZ O VALJANOSTI: POSLJEDICE TESTIRANJA

Od 1999. godine kada su izašli pretposljednji Standardi za pedagoško i psihološko testiranje (kod nas prevedeni 2006), posljedice testiranja postale su važan dio ispitivanja kvalitete testova. Međutim, i prije ovih standarda, psihozi su upozoravali da testni rezultati ne smiju štetiti bilo kojoj grupi ili individui. Tako je Cronbach (1988) istaknuo da svaki korisnik testa ima obavezu da provjerava da li njegova praksa ima adekvatne posljedice po osobe i institucije kako bi se spriječili negativni utjecaji testiranja. Ovaj kriterij izazvao je mnogo polemika. Znanstvenici su postavljali pitanja što je to negativni utjecaj i na osnovu čega određujemo da je nešto negativan utjecaj? Pitanje posljedica testiranja, naravno, uključuje vrijednosne sudove i zato je važno **razlikovati pravednost (nepristranost) i posljedice** (Furr i Bachanbach, 2013). Pretpostavimo da u testiranju profesionalnih interesa muškarci tendiraju da postižu više rezultate na skali Poduzetničkog tipa. Zamislimo sada da se stipendije za programe dodjeljuju na osnovu profesionalnih interesa i da je važno imati poduzetničkog duha da bi dobili stipendiju (jer se pretpostavlja da ćete biti samoinicijativniji, samostalniji, da ćete željeti nešto uraditi nakon stipendije). Ako bismo se oslonili na rezultate testiranja, za pretpostaviti je da bi stipendije dobilo više studenata nego studentica. Ili, ako socioekonomski status porodice utječe na postignuća na verbalnom testu općih sposobnosti, a u neku srednju školu se biraju učenici na osnovu njihovog postignuća upravo na tom testu, da li je to pravično i koje su posljedice? U prvom slučaju, posljedice su negativne po studentice, ali test (ako je konstruiran kako treba) može biti nepristran (pravedan). Za drugi slučaj ostavljamo vama da promislite što je pravično, nepravično, pozitivno ili negativno.

Da sumiramo – valjanost je kompleksna metrijska karakteristika testa koja govori o tome koliko su interpretacije testnih rezultata u skladu s teorijskim konceptima za koje pretpostavljamo da su mjereni testom. Valjanost nije karakteristika testa, nego testnih rezultata. Test nije valjan ili nevaljan, ali naše interpretacije mogu biti valjane i manje valjane. Bez poznavanja valjanosti testa nemoguće ga je odgovorno primjenjivati. Stoga je zadatak svih psihologa da promišljaju o valjanosti testnih rezultata jer samo tako mogu biti sigurni da su njihovi dijagnostički zaključci, preporuke i predviđanja ispravni.

9 PRISTRANOST U TESTIRANJU

Šta ja pristranost? Kakva je razlika između pristranosti i pravednosti? Kako znamo da je test pristran? Šta nije pristranost u testiranju?

Pristranost u testiranju (engl. test bias) predstavlja jedan od fundamentalnih problema o kojem razmišljamo u svim koracima, počevši od konceptualnih definicija do interpretacije testnih rezultata i donošenja odluka. Pristranost u testiranju je metrijsko pitanje, suprotno je od jednakosti tretiranja, a povezano je i s pravednošću kao socijalnim konstruktom. Pristran test je test na kojem ispitanici sistematski dobivaju precijenjene ili potcijenjene bruto rezultate varijable koju smo namjeravali mjeriti.

Testiranje regruta u SAD-u: Army Alpha i Army Beta kao radikalni primjer pristranosti u testiranju

Kada je izbio Prvi svjetski rat, javila se velika potreba u Sjedinjenim Američkim Državama za regrutacijom vojnika i njihovim pozicioniranjem na različite poslove u vojsci. Robert Yerks, koji je 1917. bio predsjednik Američkog društva psihologa (APA), okupio je vrhunske psihometričare da pripreme test mentalnih sposobnosti za vojsku. Autori su pripremili tri testa – Army Alpha, Army Beta i test za individualnu primjenu. Army Alpha je bio verbalni test, a Army Beta slikovni. Regruti su bili testirani tako što se prvo provjeravalo jesu li pismeni. Ako su bili pismeni, radili bi Army Alphu. Ukoliko bi postigli loše rezultate, uradili bi Army Betu. Ukoliko bi i tu pokazali loše rezultate, išli bi na individualno testiranje mentalnih sposobnosti (s nekom od verzija Binet-Simonove skale). Nepismeni kandidati su radili Army Betu. Testiranje je provedeno na oko 1 750 000 regruta. Rezultati testiranja nisu objavljeni do 1921. kada je izašla knjiga „Psyschology Examining in the United States Army“ čiji je editor bio sam Yerks (Yerks, 1921). Testiranja su rađena u prepunim prostorijama slabog osvjetljenja i s mnogo buke (Gould, 1981, str. 194-232). Edwin Garrigues Boring, koji je kasnije postao poznati psiholog, u to vrijeme Yerksov podređeni, izabrao je sam 160 000 ispitanika iz čitave baze prema naslijeđenim osobinama (rasa, zemlja porijekla...). Na osnovu rezultata Army Alphe, Army Bete i individualnog testiranja (ne zaboravite da je jako mali broj Afroamerikanaca tada bio opismenjen i u uzorku su bile velike razlike u proporciji ispitanika) izvršeno je uspoređivanje „inteligencije“ između grupa. Tri glavna zaključka su bila (kako ih Gould opisuje iz Yerksove knjige):

1. *Prosječna mentalna dob muškaraca bijelaca bila je malo iznad nivoa „morona“ (na nivou 13-godišnjaka). Terman je ranije postavio standard na 16 godina, ali nova brojka je postala početna tačka tadašnjih eugenista koji su kukali nad sunovratom nacije i inteligencije uslijed „nekontroliranog uzgoja siromašnih i slaboumnih“ i širenja „krvi crnaca uslijed miješanja rasa“ te preplavljivanje inteligentnog domaćeg stanovništva imigrantskim talogom koji dolazi iz južne i istočne Evrope. Yerks (1921) je čak dao i objašnjenje za ovaj nevjerovatni rezultat od tri godine razlike u prosječnoj inteligenciji koju je predložio Terman i koju je on „izračunao“. Rekao je da je Terman svoje rezultate bazirao na malom uzorku probranih ispitanika (srednjoškolaca i zaposlenika), dok je njegov uzorak mnogo veći.*
2. *Evropski imigranti se razlikuju prema zemlji porijekla. Prosječan građanin svake zemlje je „moron“. Tamniji ljudi južne Evrope i slavenskog porijekla iz istočne Evrope su manje*

inteligentni od svijetlih ljudi zapadne i sjeverne Evrope. Nordijska nadmoćnost nije nikakva ultranacionalistička predrasuda. Prosječan Rus ima mentalnu dob 11,34 godine, Italijan 11,01, a Poljak 10,74.

3. *Crnci su na dnu skale prema mentalnoj dobi od 10,41 godinu.*

(Gould, 1981, str. 227)

Reakcije na ovaj izvještaj pokazuju da su osviješteni psiholozi i tada promišljali o pristranosti. Prvo, pitanja u Army Alpha bila su potpuno zasićena znanjima koja nisu bila svima dostupna, posebno ne imigrantima različitih obrazovnih iskustava. Neka od pitanja odnosila su se na igrače bejzbola i tipičnu američku hranu. Army Beta nije bila ništa bolja. Početna pitanja bila su lagana (npr. da se docrta nedostajući dio tijela čovjeku ili životinji), ali „teža“ pitanja su se odnosila na stvari s kojima većina regruta nije imala iskustvo (npr. mreža na teniskom igralištu ili lopta u ruci igrača bejzbola). Testovi su bili vremenski ograničeni, ali je ova informacija data samo regrutima koji su polagali Army Alpha test. Regruti koji su odmah bili klasificirani za Army Beta test nisu imali informaciju o vremenskom ograničenju. Mjesta testiranja jako su varirala u uvjetima u kojima se radio test i svako mjesto je donekle mijenjalo pravila ko polaže koji test. Sam način kako je Boring standardizirao i uspoređivao rezultate unio je dodatni element pristranosti.[2] Kada se sve sumira, ovo je bilo čisto gubljenje vremena. Ali ova „studija“ dala je krila tadašnjim zagovornicima hereditarne teorije, posebno protestantima anglosaksonskog porijekla za zagovaranje uvođenja restrikcija imigracije iz pojedinih zemalja Evrope. Tako je 1924. godine u SAD-u izglasan Akt o restrikciji koji je doveo do toga da je samo malom postotku ljudi iz južne i istočne Evrope (2 %) bilo dozvoljeno da imigriraju u SAD naspram visokih procenata imigranata iz sjeverozapadne Evrope i Velike Britanije.

Za završetak priče o „testiranju“ regruta – jedna trivija: da li znate da je jedan od Yerksovih pomoćnika, Carl C. Brigham, postao sekretar Odbora za prijemne ispite (College Entrance Examination Board) i da je prema modelu Army testova razvio Scholastic Aptitude Test (poznat kao SAT) koji se danas primjenjuje u SAD-u kao jednoobrazni test za sve srednjoškolce za prijeme na fakultete?

Danas se nadamo da se ovakva pristranost ne može desiti zahvaljujući nizu rigoroznih kontrola koje moraju proći nacrti istraživanja. Pristranost u testiranju više nije ovako otvorena. Međutim, zamislimo sljedeću situaciju – želimo ispitati koliko djeca u trećem razredu osnovne škole poznaju neke geografske pojmove. Jedno od pitanja u testu odnosi se na rijeke koje protiču kroz velike gradove:

[2] Za više o ovom mjerenju, ali i drugim „mjerenjima“ u psihologiji koja su nas naučila kako ne mjeriti, preporučujemo čitanje knjige Stevena Jaya Goulda (1981) „The Mismeasure of Man“.

Koja rijeka protiče kroz Banju Luku?   a. Bosna   b. Vrbas   c. Una

Djeca koja žive u Banjoj Luci vrlo lako će odgovoriti na ovo pitanje. Ali, zamislimo da želimo ispitati djecu u čitavoj BiH na istom testu s ciljem ispitivanja postignuća iz Moje okoline. Da li će ovo pitanje biti jednako lagano i za dijete u Sarajevu ili Mostaru? Naravno da neće. Ovo pitanje će za njih biti mnogo teže. Da li onda imamo pravo, na osnovu pristranog pitanja, zaključivati o razlikama u znanju između naših ispitanika? Naravno da nemamo.

9.1 VRSTE PRISTRANOSTI U TESTIRANJU

Prema Van de Vijveru i Tanzeru (2004) možemo razlikovati tri tipa pristranosti u testiranju. Za svaki od ovih tipova vezani su izvori pristranosti. U tabeli su prikazani osnovni tipovi i izvori pristranosti o kojima ćemo govoriti u nastavku teksta.

Tabela 9-1 Tipovi i izvori pristranosti u testiranju (prema Van de Vijver i Tanzer, 2004)

Tip pristranosti	Izvori pristranosti
1. Konstruktna pristranost	Parcijalno preklapanje definicije konstrukta između grupa
	Različitost ponašanja povezanih s konstruktom
	Neadekvatno uzorkovanje svih relevantnih ponašanja (npr. kratki instrumenti)
	Neadekvatna pokrivenost domena konstrukta
2. Pristranost metode	
a. Pristranost uzorka	Nekompatibilnost uzorka (razlike u obrazovanju, motivaciji...)
b. Pristranost procedure testiranja	Razlike u uvjetima administriranja – fizičke i socijalne razlike
	Nejasne instrukcije za ispitanike ili za ispitivače
	Različit stepen obučenosti ispitivača
	Efekat opažača
	Problemi u komunikaciji između ispitivača i ispitanika
c. Pristranost instrumenta	Različita upoznatost sa stimulusom
	Različita upoznatost s procedurama odgovaranja
	Različiti stilovi odgovaranja
3. Pristranost pitanja	Neadekvatni prevodi ili nejasna pitanja
	Faktori šuma (pitanje potiče još neke faktore)
	Kulturalne specifičnosti u pitanju

9.1.1 Konstruktna pristranost

Konstruktna pristranost ogleda su u parcijalnom preklapanju definicija konstrukta među različitim grupama (rodnim, klasnim, kulturalnim). Česte primjere konstruktne pristranosti nalazimo između zapadnih i dalekoistočnih kultura. Na primjer, u istraživanju dimenzija ličnosti (Cheung, i dr., 1996) pokazalo se da Kinezi imaju „domaće“ osobine ličnosti koje nisu opisane u uobičajenim, nama poznatim teorijama. Tako se kod Kineza pojavljuje dimenzija „Obraz“ koju

stanovnici zapadne civilizacije ne poznaju na taj način. Obraz je poštovanje u zajednici i osnova je uvezanosti s društvom. Gubitak poštivanja je velika prijetnja odnosima i vrlo je teško povratiti „obraz“ jednom kada se izgubi. U studijama vrijednosti u Kini (Connection, 1987) rezultati su pokazali da se kod Kineza izdvaja faktor nazvan Konfučijanski radni dinamizam koji je visoko korelirao s ekonomskim rastom, a nije bio pokriven zapadnim konceptom vrijednosnog sistema.
U različitim grupama i kulturama nisu sva ponašanja jednako važna i relevantna za ispitivanje konstrukta. Na primjer, pitanje u skali Autonomije na poslu u Upitniku psihosocijalnih faktora na radnom mjestu glasi: „Da li možete sami odrediti kada ćete doći na posao?“ Kako je ova skala pripremljena u skandinavskim državama, gdje je fleksibilno radno vrijeme odavno opcija za radnike, ovo je važna čestica i njen rezultat je visoko povezan s faktorom Autonomija na poslu, te dobro predviđa zadovoljstvo na poslu. Međutim, u našim uvjetima, gdje većina radnika ne može birati kada će doći na posao, ova čestica ne mora imati jasnu faktorsku strukturu i biti povezana sa drugim česticama kojima se mjeri Autonomija.
Konstruktna pristranost je posljedica i neadekvatnog uzorkovanja ponašanja povezanih s konstruktom. Kada su konstrukti vrlo varijabilni i zahvataju veliki broj domena, trebamo mnogo pitanja da bismo postigli reprezentativnost domena i ponašanja. Desi se da zbog ekonomičnosti u upitnicima primijetimo nereprezentativnost uzoraka ponašanja. Ponekad, nakon prikupljanja podataka, moramo izbaciti neka pitanja upravo zbog konstruktne pristranosti.
Pronađite i vi primjer konstruktne pristranosti – naučit ćete i psihometriju i kroskulturalnu psihologiju.

9.1.2 Pristranost metode

Drugi tip pristranosti je **pristranost metode**. Pristranost metode može se podijeliti na tri podtipa: nekompatibilnost uzoraka, pristranost mjernog instrumenta i pristranost procedure testiranja.

9.1.2.1 Nekompatibilnost uzoraka

Nekompatibilnost uzoraka odnosi se na varijable u uzorcima koje nisu predmet mjerenja. Na primjer, proces regrutiranja ispitanika može se jako razlikovati. U nekim zemljama, ispitanicima plaćaju da učestvuju u istraživanju. Njihova motivacija je sigurno jednim dijelom određena džeparcem. U državama kao što je BiH, ispitanici obično nisu plaćeni za učešće i imaju veću motivaciju, ali i veće ego učešće. Zamislimo sada da je primijenjen test inteligencije. Za ispitanika koji je plaćen rezultat ne mora biti toliko bitan (jer će ionako biti plaćen), ali za ispitanika koji nije plaćen ovo je veliko ego ulaganje i situacija može isprovocirati misli koje ugrožavaju samopouzdanje, povećavaju ispitnu anksioznost i dovode do slabijeg učinka na testu. Primjer razlike u motivaciji imamo i kada poredimo grupe od kojih je prva često izložena testiranju (pa zbog toga može biti manje motivirana) u odnosu na drugu grupu koja je rijetko učestvovala u istraživanjima. Razlike u obrazovnom nivou mogu biti razlogom nekompatibilnosti uzorka. Ako primijenimo test inteligencije koji je zasićen školskim znanjem, zbog razlika u obrazovnom iskustvu između dvije grupe mi ne možemo tek tako potvrditi da su se razlike pojavile zbog inteligencije.

9.1.2.2 Pristranost mjernog instrumenta

Pristranost mjernog instrumenta najčešće je uzrokovana razlikama u poznavanju stimulusa, poznavanju procedura odgovaranja i stilovima odgovaranja.

Kada koristimo instrument čiji je sadržaj poznat jednoj, a manje poznat drugoj grupi, tada razlike koje dobijemo između grupa ne mogu biti interpretirane kao stvarne razlike (poznatost stimulusa). Na primjer, ako djeci u Bosni i Hercegovini damo da sklope sliku predmeta s kojim nisu imali iskustvo jer predmet potječe iz Indije, onda je razlika u brzini sklapanja slike više mjera nepoznatosti nego spretnosti. Poznati primjer pristranosti zbog poznatosti stimulusa dao je Piswanger (prema Van de Vijver i Tanzer, 2004). On je identificirao razlike na testu inteligencije između djece u Nigeriji i Togu (dvije škole u kojima su djeca bila educirana prema kurikulumu na arapskom jeziku) i djece iz Australije. Najveće razlike u postignuću bile su na zadacima koji su tražili primjenu pravila u horizontalnom smjeru od lijevo ka desno. Na tim zadacima djeca iz Nigerije i Toga bila su mnogo slabija. Razlike su se pojavile zbog toga što se identifikacija trebala raditi s lijeva na desno, a oni su školovani na pismu koje se piše s desna na lijevo.

Razlike u poznavanju procedura odgovaranja također spadaju u pristranost instrumenta. Zamislimo da ispitujemo psihomotorne sposobnosti dvije grupe djece, od kojih je jedna grupa imala mnogo iskustva u radu s maketama, a druga grupa nije nikada slagala makete. Test se sastoji od slaganja malih dijelova auta i drugih predmeta, a vrijeme potrebno da se složi maketa je mjera psihomotorne spretnosti. Možemo pretpostaviti da će grupa s iskustvom slaganja maketa imati bolji rezultat, ali ne zato što su motorno spretniji nego im je poznat stimulus.

9.1.2.3 Pristranost procedure testiranja

Treći podtip metodske pristranosti je vezan za fizičke i socijalne uvjete testiranja. Testiranje može biti pristrano zbog korištenja opreme nepoznate ispitaniku. Takva oprema kod njh može buditi radoznalost i skretati mu pažnju s testiranja. Doba dana, zagušljivost prostorije, temperatura u prostoriji... sve su to faktori koji mogu dovesti do pristranosti u testiranju.

Primjer socijalnih faktora je broj ispitanika u prostoriji za vrijeme testiranja. Nije isto da li se grupno testiranje provodi sa pet ili sa dvadeset ispitanika. Ovo je čest izvor pristranosti u školskom i radnom okruženju. Na primjer, u jednom odjeljenju je 10 učenika, a u nekom drugom odjeljenju je 30 učenika koji sjede jedan do drugog, mogu da vire u testove i guraju se u klupama. Ovi faktori ozbiljno narušavaju objektivnost testiranja i doprinose pristranosti. Osim toga, u slučaju da radimo analize na nivou razreda kao grupa, tada prosječne vrijednosti razreda sa 10 i 30 učenika ne mogu biti jednako tretirane.

Procedura testiranja može dovesti do pristranosti kada ne postoje jasne instrukcije za ispitanike i jasne upute za ispitivače. Na primjer, procedura testiranja na vanjskoj maturi može dovesti do pristranosti ukoliko dežuraju nastavnici različitog nivoa ekspertize u smislu općeg iskustva s ovakvim vrstama testiranja, upoznatosti s materijalom i poznavanjem dozvoljenih i nedozvoljenih obrazaca interakcije s ispitanicima.

Efekat ispitivača je proučavan u različitim situacijama, posebno u studijama u kojima su se poredile različite kulture. Samo prisustvo ispitivača ili opažača utječe na ponašanje ispitanika.

Njegova očekivanja i ponašanje mijenjaju odgovore ispitanika. Na primjer, niz studija je pokazalo da su ispitanici spremniji pokazati pozitivan stav prema nekoj grupi kada ih intervjuira neko ko dolazi iz te grupe (teorija priklanjanja).

Posljednji, ali ne i manje važan, izvor pristranosti u proceduri testiranja je komunikacija između ispitivača i ispitanika. Ispitanici ne moraju poznavati značenje riječi koje koristi ispitivač. Na primjer, ako ispitivač pita ispitanika: „Da li ste anksiozni?" ispitanik ne mora znati tu riječ. Ali, „anksioznost" mu može zvučati kao neki „poremećaj". Kako on sebe ne vidi „poremećenim", na pitanje odgovara: „Nikada", ali ne zato što on zaista nije nikada anksiozan, nego zato jer nije razumio riječ i pripisao joj je značenje prema svojoj preferenciji. Ovo je još veći izvor pristranosti u pisanim testovima (bilo papir-olovka bilo elektronski) kada ispitanik nema priliku da pita što znači neka riječ ili fraza.

Metodska pristranost može imati ozbiljne posljedice po valjanost istraživanja jer vodi ka promjeni prosječnih rezultata.

9.1.2.4 Pristranost pitanja

Treći tip pristranosti je pristranost pitanja. Često se naziva i diferencijalno funkcioniranje pitanja (engl. differential item functioning). Ovaj naziv ukazuje na smisao pristranosti pitanja – to je situacija kada pitanje drugačije funkcionira u različitim grupama. Pristranost pitanja je najviše istraživana vrsta pristranosti.

Mi često koristimo instrumente koji su pripremljeni u drugoj kulturi i na drugom jeziku (najčešće engleskom). U postupku adaptacije, pokušavamo da ih pripremimo tako da budu podjednako valjani i pouzdani za korištenje u našoj kulturi. Međutim, to nije tako jednostavno. Prijevodi, iako mogu biti lingvistički potpuno korektni, ponekada ne uspijevaju uhvatiti psihološko značenje. Neke riječi ili fraze jednostavno ne mogu biti doslovno prevedene zbog toga što imaju metaforičko značenje. Na primjer, u engleskom jeziku postoji fraza „cold feet" koja ukazuje na strah od nekog događaja i želju osobe da odustane od daljnje aktivnosti (na primjer, pred udaju mlada se prepadne i kaže se da je „got the cold feet"). U našem jeziku bi bilo smiješno da ovo prevedemo kao „Da li vam se desi da imate hladne noge pred neki važan događaj?" Ako pitanje nije jasno, ispitanici će ga interpretirati na različite načine.

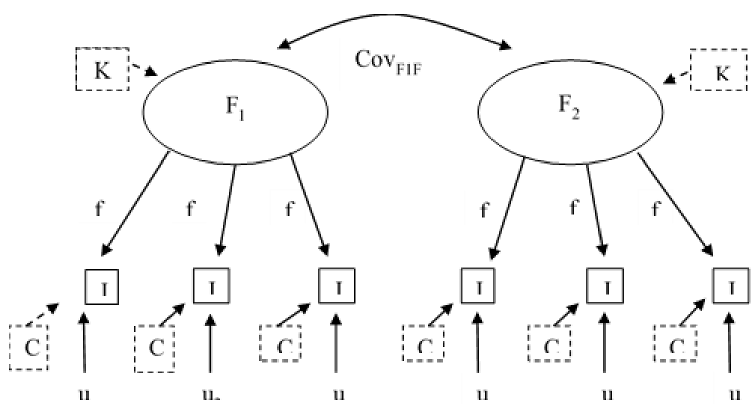
Poseban problem u konstrukciji testova je kada pripremamo test za konstrukt koji „ne postoji", odnosno nije se izdiferencirao u kulturi ili društvu. Ponekad, društvo ne prepoznaje set ponašanja kao povezan s fenomenom jer još uvijek ne vidi postojanje tog fenomena. Na primjer, fenomen pod engleskim nazivom „entitlment mentality" podrazumijeva uvjerenje osobe da joj nešto pripada ili da joj društvo nešto duguje zbog zasluga dobivenih rođenjem ili završavanjem škole. Primjer ovog fenomena su izjave: „Ja sam završio fakultet i društvo je dužno da mi osigura radno mjesto." Ili „Ja sam platio ovo školovanje, a vi mene obarate na ispitu". Entitlment mentality povezan je s vanjskim lokusom kontrole i pasivnim odnosom prema životu. U našem društvu ovaj fenomen se još uvijek ne prepoznaje u dovoljnoj mjeri da bi bio važna varijabla u istraživanjima. Prvo pitanje za istraživače ovog fenomena bi bilo: „Kako prevesti entitlment mentality?" Literarni prijevod bi bio „mentalitet prava", ali ovaj prijevod je potpuno neadekvatan za psihološko značenje. U ovakvim situacijama, istraživači se suočavaju s velikim izazovom, počevši od toga kako izbjeći konstruktnu pristranost do toga kako adekvatno pripremiti pitanja u upitniku koja bi najbolje opisala sam fenomen (npr. kako prevesti pitanje „I feel entitled to more of everything“).

9.2 MJERNA EKVIVALENCIJA

Mjerna ekvivalencija je termin koji se odnosi na nepostojanje razlika između konstrukata, mjera i njihovih odnosa, a koji u literaturi iz multivarijatne statistike poznajemo i kao mjernu invarijancu. Ako postoji invarijatnost između dvije mjere, onda su one ekvivalentne. Dvije mjere su invarijatne / ekvivalentne onda kada ispitanici iz različitih populacija, koji imaju isto razumijevanje konstrukta, postižu isti rezultat na testu. Dakle, dvije osobe iz dvije različite populacije koje isto razumiju konstrukt i koje imaju jednako izražen konstrukt imat će isti rezultat na testu. Ekvivalencija ne podrazumijeva da dvije populacije moraju imati isti prosječni rezultat jer ekvivalencija nije jednakost distribucije neke karakteristike između populacija. Prosječni rezultati mogu biti različiti, a da je mjera ekvivalentna.

I kod ekvivalencije imamo nekoliko nivoa. Prikazat ćemo nivoe koje su predložili Van de Vijver i Leung (van de Vijver i Leung, 1997), te autori koji se bave multivarijatnim metodama (Byrne, 2004). Iskoristit ćemo sliku da vidimo na kojim se nivoima pojavljuje mjerna ekvivalencija (Schmitt i Kuljanin, 2008):

Slika 9-1 Dvodimenzionalni mjerni model



Svako pitanje objašnjeno je jednačinom (faktorska jednačina bruto rezultata):

$$I = c + fF + u$$

Na ovoj slici vidimo sljedeće:

1. F ovale – dva konstrukta međusobno povezana (COV_{F1F2})
2. I kvadrate – pitanja koja mjere svaki od faktora
3. C – konstanta u regresijskoj jednačini – intercept pitanja
4. u – rezidual u regresijskoj analizi (odstupanje bruto rezultata na pitanju od predviđenog rezultata na osnovu povezanosti s faktorom)

9.2.1 Konstruktna ekvivalencija

Konstruktna ekvivalencija odnosi se na univerzalnost konstruktne valjanosti instrumenta. Konstruktna ekvivalencija pretpostavlja da se testni rezultati mogu jednako interpretirati u smislu konstrukata koje mjere. Ako smo u jednoj populaciji dobili rezultate koji ukazuju na dvodimenzionalnu mjeru pri čemu su konstrukti povezani u nomološkoj mreži kao na slici, tri pitanja mjere prvi konstrukt, druga tri pitanja drugi konstrukt, onda bismo očekivali da i u drugoj populaciji dobijemo isti obrazac – dva povezana faktora sa po tri pitanja. Konstruktna ekvivalencija je vrlo važna, ali često u opasnosti jer instrument možda ne mjeri isti konstrukt u dvije grupe ili je konstrukt povezan s različitim ponašanjima u različitim grupama. Na primjer, motiv za postignućem se u zapadnoj civilizaciji veže najviše za lično postignuće, dok se u istočnim civilizacijama osjećaj postignuća veže i za dobrobit zajednice. Ukoliko se ustanovi da postoje velike razlike u ponašanjima koja objašnjavaju konstrukt, tada se mora pripremiti adekvatan mjerni instrument.

9.2.2 Ekvivalencija mjernih jedinica

Ova ekvivalencija odnosi se na to da postoje jednake metričke jedinice, ali sama skala ne mora biti ista. Tipičan primjer je mjerenje temperature u kelvinima i celzijima. One imaju istu mjernu jedinicu, ali im je početak različit (skala Kelvina započinje na -273°C). Kod ispitivanja metričke invarijatnosti provjeravamo da li su faktorska zasićenja pitanja jednaka kroz grupe. Metrička ekvivalencija nam omogućava da uspoređujemo razlike koje pronađemo između podgrupa unutar velikih grupa (npr. da li se muškarci i žene razlikuju i u jednoj i u drugoj kulturi), ali zbog razlika u mjernim skalama ne možemo direktno uspoređivati rezultate.

9.2.3 Skalarna ekvivalencija ili ekvivalencija skale

Ovaj nivo ekvivalencije podrazumijeva da su rezultati u dvije grupe ne samo interpretirani istim faktorima i da su zasićenja pitanja na faktorima ista, nego i da je skala mjerenja ista. Kada testiramo skalarnu invarijancu, gledamo da li su tačke presjeka (intercepti) isti za sve grupe. Kada je postignuta invarijanca i za faktorska zasićenja i za intercept, to znači da različite grupe imaju istu jedinicu mjerenja, kao i istu polaznu tačku, odnosno da faktorski prosjeci mogu biti uspoređivani kroz grupe. Ukoliko ovo ne stoji, onda nije sigurno da li se razlike u faktorskim prosjecima mogu pripisati kulturalnim razlikama ili mjernim pogreškama.

9.3 MJERE NA DRUGIM JEZICIMA – PREVOĐENJE

Pitanje pristranosti je jedno od najvažnijih pitanja u mjerenju. Danas, kada imamo veliku potrebu da provodimo internacionalna komparativna istraživanja, kada ne možemo znati gdje smo ako se ne usporedimo s drugima, pitanje pristranosti postalo je pitanje visokog prioriteta. Vidimo da na pristranost djeluje jako mnogo faktora – kultura, jezik, običaji, školovanje, životni stilovi, socijalizacija... mnogo elemenata koje moramo razmotriti i kontrolirati, ali ih ne možemo mijenjati da bismo postigli veću ekvivalenciju. Ipak, ima jedna stvar koju možemo i moramo kontrolirati, a to je prevođenje.

Prevođenje psiholoških mjernih instrumenata s jednog na drugi jezik daleko je od pukog lingvističkog usklađivanja. U svom članku Bracken i Barona (1991) naveli su da je neophodno provesti proceduru prijevoda koja će osigurati da prevedeni instrument ima psihološko, lingvističko i kulturalno značenje. Dvije uobičajene procedure za prevođenje instrumenata su prijevod s ponovnim prijevodom na originalni jezik i grupni prijevod. U prvom slučaju, mjerni instrument s originalnog na ciljani jezik prevede jedna osoba, a zatim se prijevod da drugoj stručnoj osobi koja prevede s ciljanog na originalni jezik. Tačnost prijevoda ustanovljava se usporedbom originalne verzije instrumenta i njegovog prijevoda „nakon prijevoda“ (engl. back-translation).

Druga procedura je grupno ili komisijsko prevođenje. Grupa relevantnih stručnjaka (psiholozi, jezičari i ostali potrebni) zajedničkim trudom traže najbolje rješenje. Odabir procedure zavisi od toga da li se razvija novi instrument ili se priprema prijevod starog. U prvom slučaju, razvoj instrumenta zovemo simultani razvoj, a u drugom slučaju sukcesivni razvoj. Primjer simultanog razvoja bio je razvoj instrumenta za mjerenje karijerne adaptabilnosti (Savickas i Porfeli, 2012) kada je tim psihologa iz 18 država radio zajedno na lingvističkom usklađivanju i operacionalnom definiranju novog konstrukta zvanog „karijerna adaptabilnost“.

Kod sukcesivnog razvoja situacija je drukčija. Konstrukt je već operacionalno definiran, umrežen u nomološkoj mreži i za njega je pripremljen instrument na originalnom jeziku. Istraživačima koji žele koristiti taj instrument date su različite opcije pristupa prevođenju.

Prvi pristup bi bio čisto prevođenje. Korištenje ovog pristupa opravdano je onda kada smo sigurni da ne postoji nikakva razlika u pretpostavljenim dimenzijama koje su mjerene testnim pitanjima. Prijevod se radi tako da se čestice doslovno prevedu (i eventualno provjeri njihova razumljivost, te za svaki slučaj uradi i ponovni prijevod na originalni jezik).

Drugi pristup je adaptacija instrumenta. Adaptacija više nije samo jednostavno prevođenje, nego prilagođavanje riječi, fraza, pa čak i čitavih rečenica kako bi se postigla lingvistička jasnoća, dalo pravo psihološko značenje i pokrile sve domene konstrukta. Na primjer, nakon početnog prijevoda primijetimo da dvije čestice u našem jeziku imaju isto značenje, kao da smo poduplali česticu. Zbog toga moramo jednu česticu izmijeniti da bismo zahvatili što je predviđeno originalnim upitnikom.

Ponekada moramo dodati pitanja jer u originalnom upitniku nije bilo pitanja koja su za drugu kulturu jako važna. Međutim, ukoliko se instrument jako razlikuje od onoga što bi bilo potrebno za mjerenje u drugoj kulturi (zbog toga što je malo preklapanje u sadržaju konstrukta ili što originalna ponašanja nisu reprezentativna za novu kulturu), tada je potrebno pristupiti sklapanju novog instrumenta. Već smo ranije spominjali kako u Kini postoje specifične dimenzije ličnosti kao što je „obraz“ te je za tu dimenziju bilo potrebno sklopiti novu skalu.

9.4 KAKO IDENTIFICIRAMO PROBLEM PRISTRANOSTI U TESTIRANJU

Identifikacija pristranosti u testiranju i razdvajanja pristranosti od pravednosti je vrlo važna. Ako pojednostavimo sve gore rečeno, mogli bismo reći da pristranost identificiramo ili kada ne možemo jednako interpretirati konstruktnu valjanost ili kada ispitanici postižu različite rezultate na pitanjima. Shodno tome, metode identifikacije pristranosti testa odnose se ili na analizu pitanja ili na analizu unutarnje strukture testa, ali i na pitanje povezanosti s drugim varijablama (u smislu provjere prediktivne pristranosti).

Za provjeru konstruktne pristranosti na nivou pitanja i na nivou unutarnje strukture testa imamo na raspolaganju najmanje 4 metode, od kojih smo većinu ranije spomenuli.

1. Jednakost indeksa diskriminativnosti pitanja
2. Rangiranje pitanja po težini
3. Diferencijalna funkcija pitanja
4. Faktorska analiza

9.4.1 Jednakost indeksa diskriminativnosti pitanja

O indeksu diskriminativnosti govorili smo u dijelu analize pitanja, odnosno određivanja diskriminativne valjanosti pitanja. Zapamtite, nepristranost pitanja ne znači da ga obje grupe isto rade. Jedna grupa može biti uspješnija, ali ako je pitanje nepristrano, onda će dvije grupe imati sličnu diskriminativnu valjanost, odnosno oni koji su bolji na testu bolje rade pitanje, slabiji na testu slabije rade pitanje i ta tendencija je slična u obje grupe. Na primjer, ako su žene i muškarci popunjavali Upitnik profesionalnih interesa, supskalu realističnog faktora, onda će kod obje grupe postojati slična tendencija – oni koji postižu viši rezultat na realističkoj skali, imat će viši rezultat i na pitanju. Da se podsjetimo, indeks diskriminativnosti određuje se tako da se odrede ekstremne grupe, za obje se izračuna indeks težine pitanja, te se od indeksa težine gornje grupe oduzme indeks težine donje grupe i tako odredi indeks diskriminativnosti. Ukoliko se desi da se indeksi diskriminativnosti razlikuju, to znači da pitanje ne odražava mjereni konstrukt na isti način u ove dvije grupe i da je pitanje pristrano. Ukoliko odlučimo da ovo pitanje ostane u upitniku, trebamo znati da testni rezultati nisu u potpunosti usporedivi što nam pravi problem u zaključivanju o psihološkim dimenzijama među grupama. Kako bismo riješili ovaj problem, pitanje treba biti revidirano ili uklonjeno iz testa. Ova analiza treba biti provedena za svako pitanje u testu. Kao razlika između indeksa težine dvije grupe, indeks diskriminativnosti ne ovisi o težini zadatka, nego samo o razlikama među grupama, te na osnovu njega možemo reći da pitanje jednako funkcionira za muškarce i za žene, ali to ne znači da su im aritmetičke sredine jednake.

9.4.2 Rangiranje

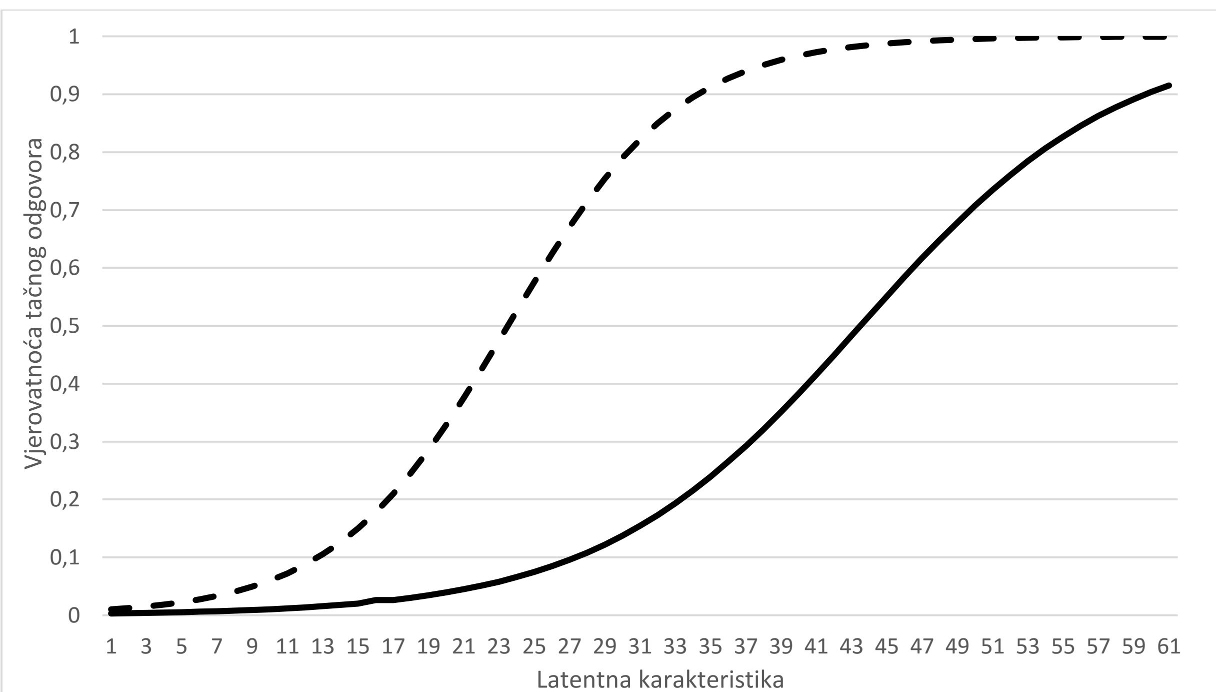
Jedna jednostavna i lagana metoda je rangiranje pitanja prema indeksima težine, te uspoređivanje rang-liste pitanja između dvije grupe. Ukoliko se rangovi pitanja razlikuju između grupa, moguće je da „imamo posla“ s pristranošću pitanja. Za ustanovljavanje sličnosti rangova koristimo Spearmanov rho (ρ)koeficijent korelacije. Ukoliko ustanovimo da je koeficijent korelacije nezadovoljavajući, potrebno je provesti dodatne analize da bismo utvrdili izvore pristranosti.

9.4.3 Diferencijalna funkcija pitanja

Diferencijalna funkcija pitanja (Item discrimination function) predstavlja najbolji način da se utvrdi postojanje pristranosti na nivou pitanja. Diferencijalna funkcija zadatka je osnovni koncept Teorije odgovora na zadatak (Item response theory ili IRT). Prije diferencijalne funkcije, da objasnimo što je funkcija zadatka u jednoj grupi. Kada povežemo rezultate na testu i rezultate na zadatku, dobijemo karakterističnu krivulju zadatka (engl. item characteristic curve ili ICC). ICC je krivulja koja na apscisi ima standardizirane vrijednosti testnih rezultata, a na ordinati indekse *p*, odnosno vjerovatnoću da će ispitanik koji postiže neki testni rezultat dobiti jedan bod ili zaokružiti jedan u tom pitanju. Na primjer, ako u grupi ispitanika koji postižu testni rezultat jednu standardnu devijaciju iznad prosjeka odredimo indeks težine 0,5, to znači da neko s ovim testnim rezultatom ima vjerovatnoću od 0,5 (ili 50 %) da zaokruži 1 u tom pitanju. Kada imamo dvije grupe, onda za svaku možemo nacrtati njenu karakterističnu krivulju zadatka. Da bismo procijenili prisustvo pristranosti, usporedimo karakteristične krivulje dvije grupe. One bi trebale biti vrlo slične, odnosno vjerovatnoća odgovora 1 trebala bi biti slična za ispitanike iz dvije grupe koji imaju isti nivo osobine (testni rezultat). Ako je pitanje pristrano, tada karakteristične krivulje neće biti slične. Na primjer, ako muškarac i žena imaju isti testni rezultat, ali za njega je vjerovatnoća odgovora 1 – 0,3, a za nju 0,5, onda je pitanje pristrano.

Pogledajmo jedan primjer na kojem imamo dvije karakteristične krivulje zadatka.

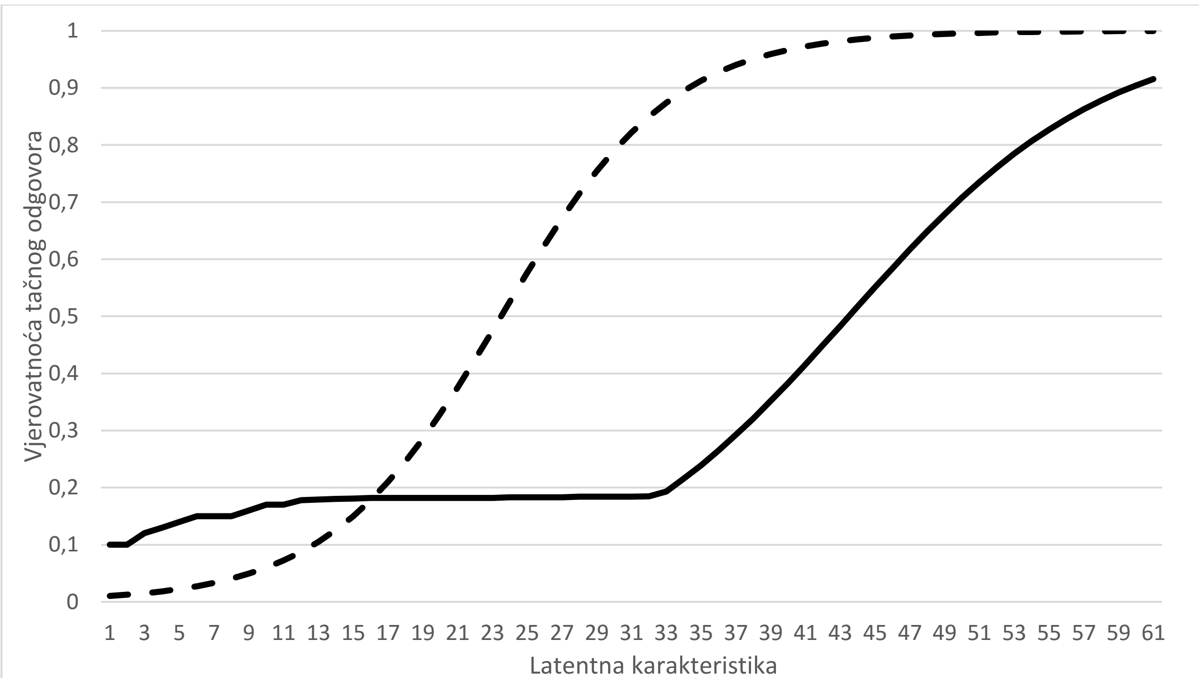
Slika 9-2 Diferencijalna krivulja pitanja za muškarce i žene – uniformna pristranost



Pretpostavimo da je puna krivulja za žene, a isprekidana za muškarce. Vidimo da je ovaj zadatak ženama teži. Kada imaju isti rezultat s muškarcima, vjerovatnoća uratka ovog zadatka je niža. Pogledajmo rezultat 29 na testu – vjerovatnoća uratka ovog zadatka (ili zaokruživanja odgovora 1) za žene je 0,11, a za muškarce 0,8. Dakle, ovo pitanje ne mjeri konstrukt na isti način u ove dvije grupe. Ovo pitanje različito funkcionira, odnosno ima diferencijalnu funkciju kod žena i muškaraca. Ovaj obrazac pristranosti, kada je trend razlike jednak duž cijele krivulje zadatka, zovemo uniformna pristranost.

U slučaju kada postoji drukčiji obrazac razlika između grupa, pristranost zovemo neuniformnom pristranošću. Na slici 9-3 žene u području nižih rezultata imaju veću vjerovatnoću zaokruživanja jedinice, ali kasnije pitanja postaju teža, odnosno za žene su teža nego za muškarce.

Slika 9-3 Diferencijalna krivulja pitanja za muškarce i žene – neuniformna pristranost



9.4.4 Faktorska analiza

Faktorska analiza je najbolji način provjere unutarnje strukture testa i razlika u unutarnjoj strukturi među grupama. Posebno nam pomaže konfirmatorna faktorska analiza i tzv. multigrupna faktorska analiza. To je analiza unutarnje strukture koja nam omogućava i poređenje razlika u dobivenim regresijskim koeficijentima, varijancama i kovarijancama kako bismo potvrdili da model postavljen za jednu grupu važi za drugu grupu. O konfirmatornoj analizi smo govorili u ranijim poglavljima, a u dijelu o ekvivalenciji smo opisali i koje osnovne vrste invarijatnosti određujemo (konfiguralnu, metričku i skalarnu).

9.4.5 Ustanovljavanje prediktivne pristranosti

Kao posljedica pristranosti konstrukta na nivou unutarnje strukture i na nivou pitanja, možemo imati ozbiljan problem u povezanosti testnih rezultata s vanjskim kriterijskim varijablama. Često čujemo „Svi imaju iste šanse za zapošljavanje, svi rade iste testove pa ko osvoji najbolje rezultate, dobit će i poziciju“. Međutim, jednom kada osvijestimo pitanje pristranosti u testiranju i vidimo koliko je izvora pristranosti i kako nije u redu koristiti iste kriterije u donošenju odluka, onda nam postane jasno kako pristranost kao metrijska karakteristika mora biti uzeta ozbiljno jer je povezana s nepravednošću.

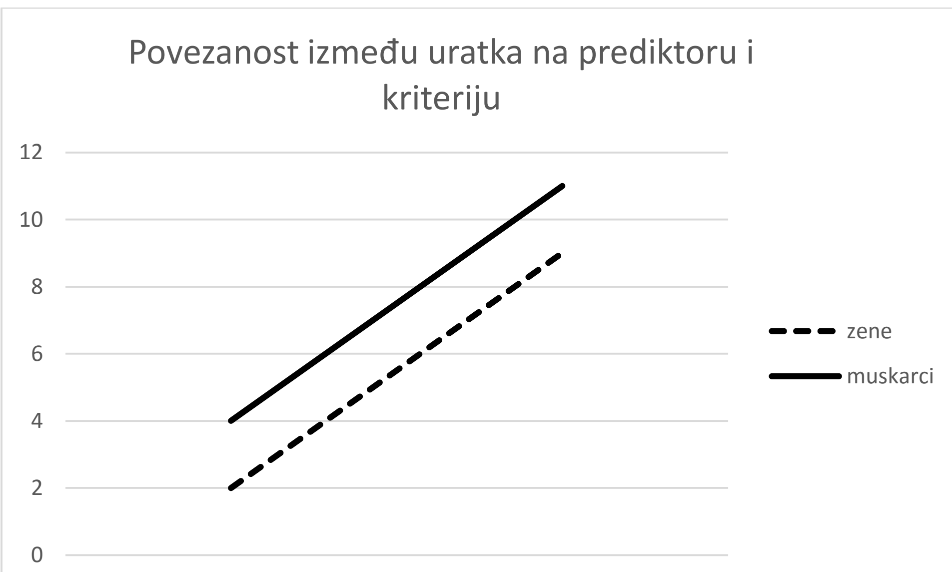
U prethodnim poglavljima govorili smo o regresijskoj analizi kao osnovi predviđanja rezultata. Ako pretpostavimo jednostavni slučaj regresije gdje na osnovu rezultata u jednoj varijabli predviđamo rezultat u drugoj varijabli, onda znamo da je Y = a + bX.

Postoji nekoliko mogućnosti razlikovanja ili pristranosti testiranja – pristranost intercepta, pristranost nagiba i pristranost intercept x nagib.

9.4.5.1 Pristranost intercepta (odsječka na Y osi)

Zamislimo da smo odredili regresijsku jednačinu odvojeno za muškarce i žene i ustanovili da su im nagibi pravca jednaki, ali da postoji sistematska razlika u aritmetičkim sredinama na kriterijskoj varijabli, kao što je prikazano na slici:

Slika 9-4 Pristranost intercepta

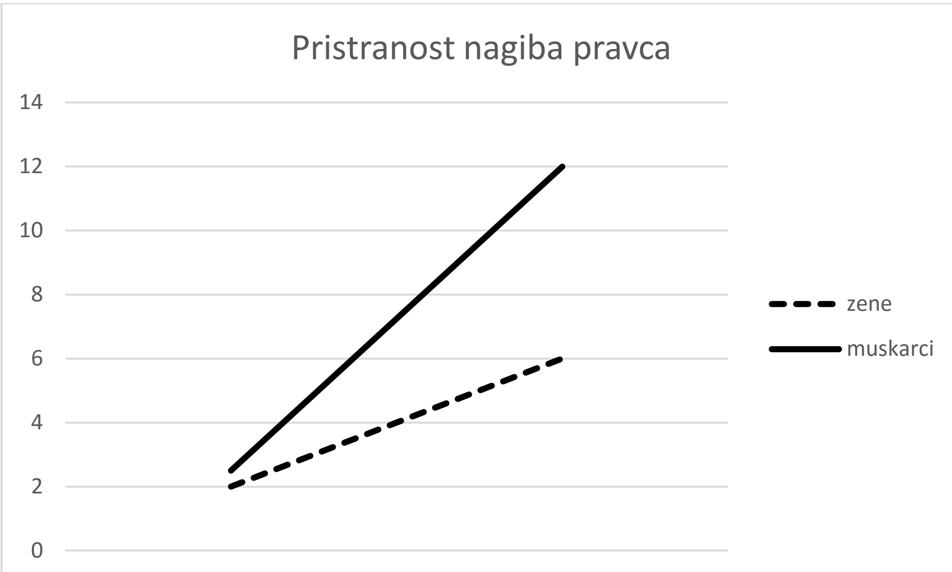


Ovo se naziva pristranost intercepta. Zamislimo da je prediktor neka objektivna mjera učinka muškaraca i žena u timskom radu, a da je kriterij procjena nadređenog o kvaliteti njihovog rada u timu. Vidimo da muškarac i žena koji imaju isti učinak imaju različitu procjenu supervizora i ta se razlika zadržava čitavom regresijskom linijom. Žene i muškarci imaju isti trend povećanja ocjene supervizora, ali žene imaju sistematski nižu ocjenu u svakoj tački povećanja učinka.

9.4.5.2 Pristranost nagiba pravca

U ovom slučaju grupe polaze od iste tačke, odnosno imaju isti odsječak na Y osi na početku, ali za jednu grupu porast u kriterijskoj varijabli je veći nego za prvu grupu.

Slika 9-5 Pristranost nagiba pravca

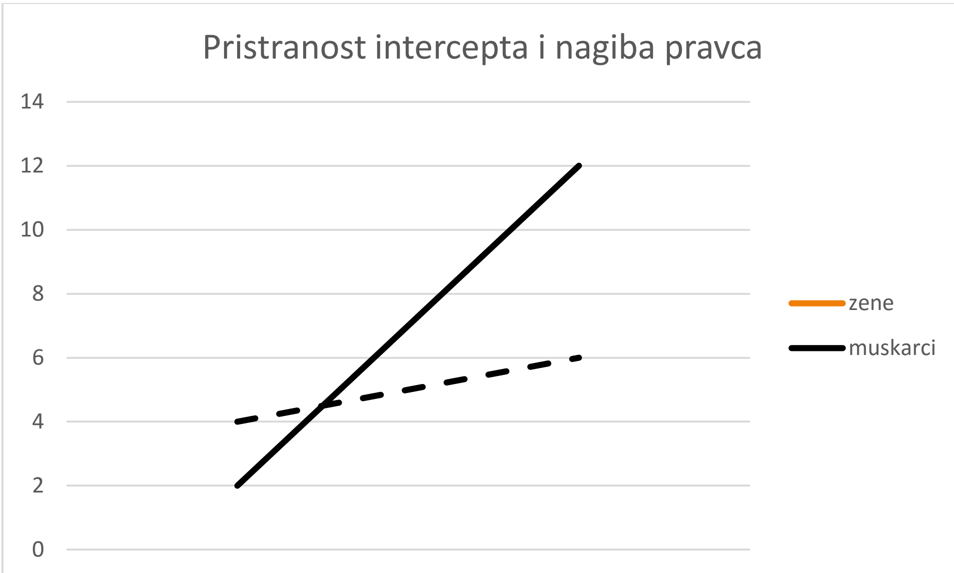


Zamislimo sada istu situaciju – prediktor je objektivna mjera učinka u timskom radu, a kriterij je procjena supervizora. Sada oni koji su objektivno najlošiji bivaju podjednako procijenjeni od nadređenog, ali što je objektivni uradak bolji, to muškarci dobivaju sve bolje ocjene u odnosu na žene.

9.4.5.3 Pristranost intercepta i nagiba pravca

Pristranost intercepta i nagiba pravca značila bi npr. da zaposlenike koji su objektivno najlošiji u timskom radu nadređeni procjenjuje tako da žene procjenjuje kao uspješnije, ali kako rezultati objektivne mjere rastu, tako se mijenja i razlika u ocjenama nadređenog. Kod najviših rezultata objektivne mjere, muškarci su daleko više procijenjeni nego žene.

Slika 9-6 Pristranost intercepta i nagiba pravca



Nakon što smo identificirali postojanje pristranosti, trebamo naći rješenje za problem. Rješenja je mnogo, spomenut ćemo samo neka.

9.5 METODE ZA MINIMIZIRANJE PRISTRANOSTI

O minimiziranju pristranosti treba razmišljati od samog početka konstrukcije testa. Ukoliko pretpostavljamo da bi test mogao biti korišten za različite grupe ili kulture tada je vrlo važno izbjegavati riječi ili fraze karakteristične za uže grupe. Ova metoda naziva se kulturalno decentriranje (Werner i Campbell, prema Van de Vijver i Tanzer, 2004). Koristi se u kroskulturalnim istraživanjima i u svim drugim situacijama gdje očekujemo da bi razumijevanje konstrukta moglo biti „zamagljeno“ razumijevanjem riječi i izraza. Na primjer, prilikom biranja riječi i izraza moramo voditi računa da ih mogu razumjeti ispitanici različiti po porijeklu, obrazovnom statusu, radnom mjestu i životnim ulogama.

Jedna od metoda je korištenje eksperata za lokalnu kulturu i jezik. Nakon što se instrument pripremi, ponudi se lokalnim ekspertima koji procjenjuju kvalitetu pitanja prema razumljivosti i obuhvatnosti unutar specifične kulture, te predlažu izmjene kako bi se ispunili kriteriji sadržajne valjanosti i omogućilo postizanje valjane strukture testa.

Dobra taktika za postizanje univerzalnosti mjerenja konstrukta je prikupljanje podataka o konstruktu na lokalnim razinama, a potom konstrukcija instrumenta bazirana na zajedničkim elementima. Na primjer, želimo ispitivati ispitnu anksioznost kod studenata u nekom komparativnom istraživanju. Za početak, u svim zajednicama pitamo studente koje su to situacije koje kod njih izazivaju ispitnu anksioznost, te termine koje koriste da opišu stanje. Analizom njihovih odgovora dolazimo do zajedničkih situacija i termina, te na osnovu njih pripremamo upitnik. Ukoliko postoje velike razlike među grupama, verzije upitnika mogu sadržavati i specifična pitanja, ali se poređenje grupa radi na osnovu zajedničkih karakteristika.

Za nas je jako bitno što uraditi kada se neki fenomen ispituje prvi put, odnosno kada fenomen još uvijek nije dovoljno prepoznat kod šire populacije ili kada test koristimo prvi put. U ovim slučajevima treba provesti pilot ispitivanje s ciljem provjeravanja shvatanja fenomena i njegove interpretacije, te razumljivosti pitanja u instrumentu. U ovoj fazi provjeravamo i psihološke procese uključene u odgovaranje tako što zamolimo ispitanike da „razmišljaju naglas“, odnosno da naglas govore sve što im pada na pamet dok odgovaraju. Analizom snimaka dolazimo do informacija da li su psihološki procesi uključeni u odgovaranje kompatibilni s teoretskim pretpostavkama na kojima je konstruiran test.

Provjeravanje pristranosti na nivou pitanja uključuje lingvističke, psihološke i psihometrijske provjere o kojima smo mnogo govorili. Ukoliko se pokaže da je pitanje pristrano, ono mora biti preformulirano ili zamijenjeno adekvatnim pitanjem. U test se mogu dodati specifična pitanja, koja se u priručniku za korištenje testa označavaju kao „dodatna pitanja“ i koja se daju na raspolaganje ispitivačima ako im trebaju.

Ako se pokaže da postoji pristranost u predviđanju rezultata na kriterijskoj mjeri, onda donošenje odluka mora uzeti u obzir razlike u povezanosti mjera. Na primjer, ukoliko se pokaže da postoji razlika između žena i muškaraca u povezanosti objektivne mjere postignuća i procjene supervizora, sve daljnje odluke – od nagrađivanja do otpuštanja – moraju uzimati u obzir ovu razliku i ne smije se dopustiti razmišljanje po principu: „Svi su radili isto, pa ko bude bolji“ . Oni nisu bolji, nego je test bio pristran.

9.6 ŠTO NIJE PRISTRANOST U TESTIRANJU

Priča o pristranosti u testiranju, osim što je znanstveno relevantna, ona je društveno i politički važna. Kao znanstvenici koji kontraintuitivno razmišljaju, ne smijemo potpasti pod utjecaj predrasuda i nedokazanih tvrdnji koje se čuju u javnom diskursu. Jensen (1980, str. 370) kaže da postoje tri pogrešne pretpostavke koje trebamo imati na umu kako ne bismo pogrešno zaključivali o pristranosti testova. To su pogreška o jednakosti (engl. egalitarian falacy), kulturalna pogreška (engl. culture-bond falacy) i pogreška standardizacije (engl. standardization falacy). Jensen je u svojoj knjizi sjajno opisao što ove pogreške znače. U nastavku teksta upoznat ćemo se s ovim pogreškama kako ih Jensen opisuje.

9.6.1 Pogreška o jednakosti

Ova pogreška bazirana je na pretpostavci da je čitava ljudska populacija jednaka po bilo kojoj osobini. Shodno tome, svaka razlika između populacija u distribuciji testnih rezultata (aritmetička sredina, standardna devijacija) trebala bi biti pripisana pristranosti testa. Potraga za manje pristranim testom značila bi traganje za kriterijem koji minimizira ili eliminira razlike u aritmetičkim sredinama i varijancama. U krajnjoj liniji, to bi bio test koji pouzdano identificira individualne razlike, ali na nivou grupa ne bi bilo razlika. Ovo jedino ne bi važilo za predselekcionirane grupe. Na primjer, ako tražimo razliku između grupe bez tretmana i grupe s tretmanom, tada očekujemo razlike. Između grupa koje nisu bile pod tretmanom ne očekujemo razlike bez obzira na to kojoj kulturi ili prostoru pripadaju. Danas znamo da ovaj koncept jednakosti jednostavno nije opstojiv. Ako ništa, imamo dovoljno dokaza da se grupe razlikuju prema psihomotornim i senzornim sposobnostima. Također, imamo dovoljno dokaza da se razlike u školskim postignućima ne mogu pripisati samo pristranosti testiranja, nego kvaliteti obrazovnog procesa i varijablama porodičnog konteksta. Pogreška jednakosti dosta je česta u tretiranju inteligencije, jer je politički i socijalno korektno smatrati da je distribucija inteligencije jednaka za sve grupe. Pristranost je i najčešća kritika testova inteligencije. Za razumjeti je zašto je ova pogreška tako snažno inkorporirana u tretiranje pristranosti testova inteligencije. U današnjem svijetu, kada su predrasude i netrpeljivost prema etničkim grupama jako izraženi, znanstvena priča o pogreški jednakosti mogla bi lako biti zloupotrebljena u svrhe segregacionističkih politika. Pogreška jednakosti i potreba da budemo „socijalno osviješteni“ dovodi i do toga da vjerujemo kako treba imati i zajedničke forme testova za različite grupe. Na primjer, profesionalni interesi muškaraca i žena jesu isti, ali se područja njihovog manifestiranja razlikuju i samo zanimanje ne mora imati isto značenje za muškarce i žene. Ali, zbog koncepta ravnopravnosti, u mjerenju insistiramo na jednakosti iako smo svjesni razlika.

Prema pogreški jednakosti, izmjerene razlike među grupama su indikator pristranosti, a ne razlika među grupama u mjerenoj osobini, pa je, prema ovome, nemoguće potvrditi ili odbaciti hipotezu o razlikama. Ali znanost ne funkcionira na ovaj način. Znanost jeste znanost zato što se hipoteze mogu provjeravati. Stoga je ova hipoteza o jednakosti znanstveno beskorisna, koliko god bila društveno i politički promovirana. Pitanje razlika je znanstveno pitanje, a pitanje jednakosti i ravnopravnosti je društveno i političko pitanje.

9.6.2 Pogreška kulturalne uključenosti

Ova pogreška je povezana sa sadržajnom valjanošću testa. Kada se desi da neko primijeti da je sadržaj pitanja „zasićen kulturalnim elementima“, test se proglasi kulturalno pristranim i nepravednim prema drugim kulturama. Pitanja koja najčešće dožive ovu osudu su ona koja se prepoznaju kao „pitanja iz knjiga“, ili pitanja koja se odnose na neke moralne, etičke ili estetske prosudbe. Jedno takvo pitanje može biti: „Ko je napisao knjigu Na Drini ćuprija?“ ili „Poznatu sliku Mona Lisa je naslikao...“? Problem nije u tome što ova pitanja odražavaju različito iskustvo učenja, nego u tome što se kritičari često oslanjaju samo na vlastitu prosudbu i površni pregled, a koji također mogu biti utemeljeni na njihovim predrasudama prema drugim grupama. Istraživanja su pokazala da neka pitanja, koja su bila snažno kritizirana da favoriziraju određenu grupu, uopće nisu bila tako teška kao druga pitanja za koja su „stručnjaci“ rekli da su oslobođena od kulture. Dakle, površna inspekcija i subjektivni sud ne mogu biti razlog da se neki test ili pitanja u testu proglase kulturalno zasićenim.

9.6.3 Pogreška standardizacije

Pogreška standardizacije odnosi se na tretiranje rezultata grupa koje nisu bile dio normativnog uzorka. Zagovornici ideje kažu: „Ako skupina nije bila dio normativnog uzorka i ako za tu skupinu ne postoje norme, onda je test pristran i nepravedan prema njima“. Ovo je bila dosta česta kritika testova inteligencije koji su bili normirani na uzorcima bijelaca, ali su onda korišteni i na drugim rasnim grupama. Kod nas bi ovo mogla biti kritika korištenja testova na romskoj populaciji, a norme su pripremane na neromskoj populaciji. Ovdje također trebamo promisliti da li je test pristran ili nije. Prije nego koristimo test na populaciji na kojoj nije normiran, moramo procijeniti da li je taj test odgovarajući za njih. Mi ionako moramo odrediti pouzdanost i valjanost testa na toj populaciji. Ne trebamo unaprijed reći da je test neiskoristiv jer za njega nemamo norme. U procesu standardizacije, dva aspekta su najvažnija: 1) odabir pitanja i 2) skaliranje pitanja. Za pristranost je važan odabir pitanja. Odabir pitanja za test baziran je na provjeri kvalitete pitanja, njihovih težina, koeficijenata korelacije između pitanja, te korelacija s ukupnim uratkom ili nekim drugim kriterijem. Sve metrijske karakteristike testa zavise od kvalitete pitanja u testu. Ukoliko procedura odabira pitanja za završnu formu zaista pokaže da za različite grupe imamo i različita pitanja, onda postoji osnova za tvrdnju da test može biti pristran, ali prije statističke provjere pristranosti ne možemo ništa tvrditi. Ukoliko test jeste pristran, tada nam reskaliranje neće značajno pomoći. Šta to znači? Ako bismo odlučili da rezultatima nove grupe samo dodamo neku konstantu, na pojavnoj razini dobit ćemo ujednačavanje s normativnom grupom. Ali ono što smo uradili jeste da smo transformirali rezultate.

U samom procesu standardizacije postavlja se pitanje da li je dovoljno samo uključiti određeni broj ispitanika iz manjinske grupe u normativni uzorak? Na primjer, da li je dovoljno samo uključiti određeni broj romske djece u Bosni i Hercegovini za normiranje testa? Vrlo lako bismo na ovo pitanje dali potvrdan odgovor, ali nam ni ovo ne garantira da će test biti nepristran, posebno ako je njihov procent u ukupnom uzorku mali. Statistika može pokazati i da su kriteriji za odabir pitanja različiti za grupe u normativnom uzorku, ali ovo nećemo saznati ukoliko statističke procedure ne provedemo odvojeno za grupe. Ispravan postupak standardizacije podrazumijevao bi uključivanje jednakog broja ispitanika iz svake grupe ili barem dovoljno velik uzorak manjinske grupe da se mogu provesti statističke analize, a tek u finalnom normiranju (određivanju standardiziranih skorova) za čitav uzorak možemo kombinirati grupe u proporcijama zastupljenosti kao u općoj populaciji.

Namjera standardizacije unutar grupa nema za cilj izjednačavanje aritmetičkih sredina, nego postizanje slične pouzdanosti, unutarnje strukture, dimenzionalnosti i raspona težine pitanja pri čemu rangovi pitanja po težini trebaju biti isti u svakoj grupi.

Dakle, pristranost testa mora biti jedna od naših primarnih briga, baš kao i sve druge metrijske karakteristike. Pri tome moramo voditi računa da pristranost testa ne smije biti opterećena pretpostavkama koje onemogućavaju znanstveni pristup ispitivanju psiholoških karakteristika.

Za kraj...

Već dugo razmišljam ko je „izvrstan psiholog". Da li je to osoba koja je pročitala sve što se moglo pročitati iz svog područja? Da li je to osoba koja je topla i nesebična? Osoba koja je ubjedljiva, elokventna..? Upoznala sam mnogo psihologa, neki su djelovali toplo, neki hladno, bili su puni znanja ili puni prazne priče... ali ono što sam zaključila o izvrsnim psiholozima je da se oni ničemu ne čude. Izvrsni psiholozi koriste se mjerenjima i teorijama da bi razumjeli i prestali se čuditi. Tek kada se prestanemo čuditi, možemo razumjeti. A razumjeti možemo samo ako mjerimo pažljivo, strpljivo i kritički. Izvrsni psiholozi su otvoreni za sve perspektive i sve interpretacije. I nikada ne prestaju da se dive ljudima. Mjerenja nam pomažu da se uvijek i iznova divimo ljudima i da ih razumijemo. Nadam se da će ovaj skromni priručnik doprinijeti vašem razvoju kao psihologa koji razumije da nema apsolutne istine i da je ljudska vrsta veličanstvena u svojoj različitosti koju otkrivamo upravo kroz mjerenja.

„A great many people think they are thinking

when they are merely rearranging their predjudices"

William James, 1842-1910

`

10 LITERATURA

AERA, APA, NCME. (2006). *Standardi za pedagoško i psihološko testiranje.* Jastrebarsko: Naklada Slap.

Allen , M. J., Yen, W. M. (1979). *Introduction to Measurement Theory.* Brooks/Cole Pub.

Bennet, G. K., Seashore, H. G., Wesman, A. G. (2006). Diferencijalni testovi sposobnosti (DAT) za selekciju – Baterija općih sposobnosti (BOS). Jastrebarsko: Naklada Slap.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*(32), 1-18.

Bracken, B. A., Barona, A. (1991). State of the art procedures for translating,validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*, 119-132.

Brown, R. A. (2005). *The Paradox of Japanese Self-Esteem.* Preuzeto od http://www.bunkyo.ac.jp/

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research* (1st izd.). The Guilford Press.

Browne, M. (2000). Psychometrics. *Journal of American Statistical Association, 95*, 450-661.

Bucik, V. (1997). *Osnove psihološkega testiranja.* Ljubljana: Filozofska fakulteta Univerze v Ljubljani.

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS:Basic Concepts, Applications, and Programming.* New York, NY: Routledge Taylor & Francis Group.

Campbell, D. T., Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Carlson, D. S., Kacmar, M. K., Williams, L. J. (2000). Construction and Initial Validation of a Multidimensional Measure of Work–Family Conflict. *Journal of Vocational Behaviour*(56), 249–276.

Cattell, R. B., Cattell, K. A., Cattell, H. E. (2000). *16 faktora ličnosti – 16PF.* Naklada Slap.

Cheung, F. M., Leung, K., Fan, R. M., Song, W. Z., Zhang, J. X., Chang, J. P. (1996). Development of the Chinese personality assessment inventory. *Journal of Cross-Cultural Psycholgy, 18*, 143-164.

Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2002). *Applied Multiple Regression/Correlation Analysis for Behavioral Sciences* (3rd izd.). Routledge.

Connection, C. C. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology, 18*, 143-164.

Costello, A. B., Osborne, J. W. (2005). Best practices in exloratory factor analysis: Four recommentadions for getting the most from your analysis. *Practical Assesment, Research and Evaluation, 10*(7), 1-9.

Cronbach , L. J., Gleser, G. C., Nanda, H., Rajaratnam, N. (1972). *The dependability of behavioural measurements: Theory of generalizability for scores and profiles.* New York: Willey .

Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika*(4), 297-334.

Cronbach, L. J. (1971). Test Validation. U R. L. Thorndike (Ur.), *Educational Measurement* (str. 443-507).

Cronbach, L. J. (1988). Five perspectives on the validity argument. U H. Wainer, & H. I. Braun (Ur.), *Test validity* (str. 3-18). NJ: Hillsdale.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th izd.). New York: HarperCollins Publishers Inc.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Cronbach, L. J., Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement, 64*, 391-418.

Ebel, R. L. (1965). *Measuring educational achievement.* Prentice-Hall education series .

Ekstrom, J. (2009). *The phi-cefficient, the tetrahoric correlation coefficient and the Pearson – Yule debate.* Preuzeto od statistics.ucla.edu

Fajgelj, S. (2003). *Psihometrija: metod i teorija psihološkog mjerenja.* Beograd: Centar za primjenjenu psihologiju.

Ferguson, G. A. (1949). On the theory of test discrimination. *Psychometrika*(14), 61-68.

Finn, S. E., Tonsager, M. E. (1997). Information-gathering and therapeutic models of assesment: Complementary paradigms. *Psychological Assesment, 9*(4), 374-385.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinetely large population. *BIometrika, 4*(10), 507-521.

Fulgosi, A. (1988). *Faktorska analiza.* Zagreb: Školska knjiga.

Furr , M. R., Bacharach, V. R. (2013). *Psychometrics: An Introduction.* SAGE Publications.

Ghiselli, E. E., Campbell, J. P., Zedeck, S. (1981). *Measurment theory for the behavioral sciences.* San Francisco: W.H.Freeman.

Gould, S. J. (1981). *The Mismeasure of Man.* New York: W.W.Norton.

Greenhaus, J. H., Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review, 10*(1), 76-88.

Gregory, R. J. (1992). *Psychological testing; History, principles and applications.* Boston: Allyn & Bacon.

Guilford, J. P. (1965). *Fundamental Statistics in psychology and education.* New York: McGraw-Hill.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd izd.). Odessa: Psychological Assesment Resources.

Hoyt, C. J. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153-160.

Hubbard, D. W. (2014). *How to Measure Anything: Finding the Value of "Intangibles" in Business.* Hoboken, New Jersey: John Wiley and Sons, Inc.

Janović, J. (17. 11 2012). *Osnovne trigonometrijske funkcije.* Preuzeto od On-line učionica: https://profesorka.wordpress.com

Jaros, S. (2007). Meyer and Allen Model of Organizational Commitment: Measurement Issues. *The Icfai Journal of Organizational Behavior, VI*(4), 7-25.

Jensen, A. R. (1980). *Bias in Mental Testing.* New York: The Free Press A Division of Macmillan Publishing Co., Inc.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of reliability. *Psychometrika*(2), 151-160.

Leahey, T. H. (2001). *The History of Modern Psychology.* Prentice Hall PTR.

Lord, F. M., Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Addison-Wesley Pub. Co.
Lord, F. M., Novick, M. R. (2008). *Statistical Theories of Mental Test Scores.* Information Age Publishing.
Marcoulides, G. A. (1996). Estimating variance komponents in generalizability theory: The covariance structure analysis approach. *Structural Eqation Modeling*(3), 290-299.
Maslach, C., Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of occupational behaviour, 2*, 99-113.
McDonald, R. P. (1999). *Test Theory: A Unified Treatment.* Mahwah, New Jersey: Lawrence Erlbaum Assoc.
Messick, S. (1989). Validity. U R. L. Linn (Ur.), *Educational Measurement* (3rd izd., str. 13-103). New York: Macmillan.
Meyer, J. P., Allen, N. J. (1997). *Commitment in the workplace: Theory,research, and application .* Thousand Oaks, CA: SAGE Publications.
Norman, G. R. (2008). Discrimination and reliability: equal partners? *Health and Quality of Life Outcomes, 6*(81).
Novick, M. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*, 1-18.
Nunnally, J. C., Bernstein, I. H. (1994). *Psychometric Theory* (III izd.). New York: McGraw-Hill Series in Psychology.
Plumlee, L. B. (1952). The effect of difficulty and chance success on item – test correlation and on test reliability. *Psychometrika*, 69-86.
Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P. (2013). Sources of Method Bias in Social Science Research and recommendations on how to control It. *Annual Review of Psychology*(63), 539-569.
Reese, T. W. (1943). The Application of the Theory of Physical Measurement to the Measurement of psychological Magnitudes, sith Three Experimental Examples. *Psychological MOnographs, 55:3*(251), 6-20.
Rencher, A. C., Christensen, W. F. (2012). *Methods of Multivarate Analysis* (3rd izd.). Hoboken , New Jersey: John Wiley & Sons.
Savickas, M. L., Porfeli, E. J. (2012). The Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behaviour, 80*, 661-673.
Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assesment, 8*(4), 350-353.
Schmitt, N., Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210-222.
Schonfeld, I. S. (1992). Assesing Stress in Teachers: Depressive Symptoms Scales and Neutral Self-Reports of the Work Environment. U J. Campbell Quick, L. R. Murphy, & J. J. Hurrell (Ur.), *Stress and Well-Being at Work: Assesment and Interventions for Occupational Mental Health* (str. 270-285).
Shavelson, R. J., Webb, N. M. (1991). *Generalizability Theory: A Primer.* SAGE Publication.
Smith, P., Whetton, C. (1999). *Testovi općih sposobnosti TOS.* Jastrebarsko: Naklada Slap.
Stevens, S. (1946). On the Theory of Scales of Measurement. *Science*, 677-680.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. U S. S. Stevens, *Handbook of Experimental Psychology* (str. 1-49). New York: Wiley.

Tabachnik, B. G., Fidell, L. S. (2007). *Using Multivariate Statistics* (5th izd.). Pearson.

Tafarodi, R. W., Shaughnessy, S. C., Yamaguchi, S., Murakoshi, A. (2011). The reporting of selfesteem in Japan and Canada. *Journal of Cross-Cultural Psychology, 42*(1), 155-164.

Thompson, B. (2002). *Score Reliability: Contemporary Thinking on Reliability Issues.* SAGE Publications, Inc.

Thurlow, W. (1950). Direct measures of discriminations among individuals performed by psychological tests. *Journal of Psychology*(29), 281-314.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 1*(11), 1-13.

Urbina, S. (2004). *Essentials of Psychological Testing.* John Wiley & Sons, Inc.

van de Vijver, F. J., Leung, K. (1997). *Methods and Data Analysis for Cross-cultural Research.* Newbury Park, CA: Sage.

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue européenne de psychologie appliquée, 54*, 119-135.

Warr, P., Cook, J., Wall, T. (1979). Scales for the Measurement of Social Work Attitudes and Aspects of Psychological Well-Being. *Journal od Occupational Psychology, 52*, 129-148.

Wikipedia. (17. 1 2016). *Cosine similarity*. Preuzeto od https://en.wikipedia.org

Yerks, R. (1921). *Psychology Examination in the United States Army* (Tom. XV). Memoirs National Academy of Sciences.

Zung, W. (1965). A Self-Rating Depression Scale. 12: Arch Gen Psychiatry 63-70. 1965. *Archives of General Psychiatry, 12*, 63-70.

Nermin Đapo, van. prof.

Odsjek za psihologiju, Filozofski fakultet u Sarajevu

Recenzija priručnika *Osnove psihometrije za studente psihologije* autorice Dženane Husremović

Psihometrija je jedna od temeljnih psiholoških disciplina koja se bavi teorijskim i praktičnim problemima mjerenja psihičkih procesa, osobina i stanja. S obzirom na to da je mjerenje od ključne važnosti za razvoj psihologije kao naučne discipline, psihometrija je izuzetno važna, ne samo u akademskom okruženju već i u svakodnevnom praktičnom radu. Stoga je studentima psihologije u njihovom akademskom obrazovanju psihometrija jedan od ključnih nastavnih predmeta nužan za usvajanje kompetencija potrebnih u psihološkim istraživanjima i stručnoj praksi.

Dženana Husremović, nastavnik na predmetima Psihometrija 1, Psihometrija 2 i Multivarijatne statističke metode na Odsjeku za psihologiju Filozofskog fakulteta Univerziteta u Sarajevu, prihvatila se izazova pisanja priručnika iz psihometrije za studente psihologije. Rezultat je rukopis *Osnove psihometrije za studente psihologije*, koji uz obaveznu i dodatnu literaturu iz psihometrije studentima treba biti pri ruci, kao pomoćno nastavno štivo u praćenju nastave, izučavanju literature i pripremanju ispita.

Tekst je organiziran u devet poglavlja koja svojim sadržajem obuhvaćaju najvažnije teme koje studenta uvode u kompleksnu materiju psihometrije. Svako poglavlje započinje uvodnim tekstom u kojem su data osnovna i kratka objašnjenja koja se u nastavku detaljno i jasno prezentiraju. Rukopis počinje određenjem predmeta psihometrije, zatim razmatranjem različitih pitanja mjerenja u psihologiji, te definicijama i vrstama psiholoških testova i razloga njihovog korištenja. U nastavku autorica, vođena izrekom da je ponavljanje majka znanja, podsjeća studente na najvažnija znanja iz statistike i matričnog računa neophodna za razumijevanje psihometrijskih koncepata opisanih u priručniku. Slijedi poglavlje o transformacijama rezultata, u kojem se, između ostalog, obrađuje formiranje ukupnog uratka na testu. Problem pouzdanosti testa najprije je obrađen kroz Klasičnu teoriju testova, postupke empirijskog određivanja pouzdanosti i faktora od kojih zavisi pouzdanost testa. Nakon što su prezentirane i objašnjene kritike Klasičnoj teoriji testova, elaborirana je teorija generalizabilnosti (G teorija) i postupak koji se koristi u donošenju odluke za buduća mjerenja (tzv. D studije). Premda je osjetljivost testa direktno povezana s pouzdanošću, iz didaktičkih razloga ova tema obrađena je u zasebnom poglavlju. Na sistematičan i pregledan način prezentirana je i objašnjena analiza zadataka (item analiza) i osjetljivost testa. U poglavlju o valjanosti, najvažnijoj karakteristici mjernog instrumenta, najprije se definira koncept valjanosti. Na temelju novijih pogleda na pitanje valjanosti, prema kojim je konstruktna valjanost pozicionirana kao centralni koncept koji povezuje sve druge koncepte, tj. dokaze za identificiranje valjanosti rezultata, ostatak poglavlja organiziran je u pet potpoglavlja o dokazima o valjanosti: sadržaj pitanja, unutrašnja struktura testa, psihološki procesi uključeni u odgovaranje, povezanost sa drugim varijablama i posljedice odgovaranja. U ovom poglavlju na jasan i pristupačan način obrađeni su postupci koji se koriste u provjeravanju valjanosti: matrica više crta – više metoda, faktorska analiza i konfirmatorna faktorska analiza. Priručnik se završava poglavljem o jednom od osnovnih problema u testiranju, problemom pristranosti. Objašnjena su tri tipa pristranosti: konstruktna, pristranost metode i pristranost pitanja. Prezentirani su i načini identifikacije problema pristranosti. Važnim se čini činjenica da je u ovom poglavlju razmatrano pitanje prevođenja s drugih jezika kao izvora pristranosti koji se može kontrolirati, a kojem se ne posvećuje dovoljno pažnje u istraživačkoj i stručnoj praksi.

Napisati rukopis namijenjen podučavanju psihometrije, koji zadovoljava neophodne pedagoške kriterije, nije nimalo lak zadatak. U ovom pogledu, autorica je ostvarila puni uspjeh. Rukopis je napisan jasnim i preciznim jezikom na zavidnom stručnom nivou. U tekstu se prepoznaje bogato predavačko iskustvo koje je autorici omogućilo da pripremi sadržajno i tehnički optimalno uređen tekst, na pristupačan, sistematičan i pregledan način, pri čemu se niti u jednom trenutku ne zaboravlja na rigoroznost koju zahtijeva psihometrija. Uvjeren sam da će studentima biti od velike koristi i da će im poslužiti kao siguran i sistematičan vodič kroz izučavanje psihometrije. Priručnik će poslužiti prvenstveno studentima, ali bit će od koristi i drugima koji se iz različitih razloga zanimaju za pitanja mjerenja u psihologiji i drugim društvenim znanostima.

Đapo

Dr. Nermin Đapo, vanredni profesor

Recenzija rukopisa dr. Dženane Husremović „Osnove psihometrije za studente psihologije“

Recenziju napisao: Doc. dr. Đorđe Čekrlija

Knjiga pod naslovom „Osnove psihometrije za studente psihologije“ predstavlja vrlo važan doprinos promovisanju i unapređenju istraživačke djelatnosti. Imajući u vidu da sva istraživanja u društvenim znanostima imaju potrebu za razvijanjem i usavršavanjem mjernih instrumenata na osnovu kojih se donose procjene i generišu zaključci, jasno je kolika je potreba za „zanatskim vještinama“ koje se postavljaju pred psihologe i stručnjake srodnih oblasti. Autorica Husremović svojim predstavljanjem i razradom psihometrijskih koncepata direktno ukazuje na takvu potrebu. Pored toga, autorica daje i vrlo jasnu sugestiju kojim strategijama i postupcima je potrebno ovladati da bi bilo moguće odgovoriti na opšte psihometrijske kriterije.

Priručnik se sastoji od 188 stranica koje su podijeljene u devet sekcija, a koje se grubo mogu razvrstati u dvije cjeline. Prvom se čitaoca nastoji uvesti u oblast psihometrije i predstaviti osnovne pojmove i načine zaključivanja, koristeći pojmove poznate iz ranijih statističko-metodoloških kolegija. U njima se koristeći saznanja deskriptivne statistike i statistike zaključivanja, predstavljaju i upoznaju koncepti mjerenja u psihologiji, psihološki test kao mjerni instrument i osnovni psihometrijski postulati. Glavna odlika prvog dijela je vrlo precizno opisivanje i razrađivanje koncepata pri kome autorica uspijeva u dva vrlo važna zadatka. Kao prvo – da koristeći ranije stečena znanja rasplaši studente i ukaže na mogućnost neproblematičnog bavljenja psihometrijom i statistikom. Drugo – koristeći ranija saznanja studenata omogućava povezivanje matematičko-statističkih aspekata društvenih nauka sa psihološkim konceptima za čije su se izučavanje studenti opredijelili. Jezik koji pri tome autorica koristi je vrlo razumljiv i ne ostavlja prostora nedoumicama. Na ovakav način, prva cjelina knjige daje odličnu podlogu za konkretnije psihometrijske analize sadržane u dijelu koji slijedi. Druga cjelina se sastoji od sekcija koje u užem smislu spadaju u psihometrijske karakteristike testa: analiza učinaka na različitim testovima kao mjera razmatranih psiholoških pojava, pouzdanost, valjanost i diskriminativnost. Sva četiri poglavlja odlikuju se vrlo jasnim i preciznim definicijama psihometrijskih karakteristika i njihovih aspekata. Predstavljanje psihometrijskih karakteristika testova i statističkih parametara se direktno nadovezuje na osnove date u prvoj cjelini knjige, tako da čitalac ima priliku vraćati se na osnovne pojmove i postulate na kojima se psihometrija zasniva i čije je poznavanje potrebno.

Ovaj priručnik predstavlja unikatno djelo u približavanju psihometrije i psiholoških mjerenja studentima. Kao prvo, ova vrsta knjiga je u Bosni i Hercegovini, kao i u zemljama regiona, vrlo rijetka. Autorica time daje doprinos, ne samo popularizaciji i promociji psihometrije već i psihologijske nauke uopšte. Pri tome, treba naglasiti da knjiga „Osnove psihometrije za studente psihologije“ ne pretenduje da bude najobuhvatniji ili najbolji udžbenik psihometrije. Autorica je vrlo vješto preskočila prepreku da svojim širim saznanjima optereti priručnik dodatnim teorijskim aspektima koji jesu bitni, ali koji u prvom susretu studenata sa psihometrijom mogu biti samo varijable koje će im otežati usvajanje novih istraživačkih znanja ili faktori zbog kojih se rađa animozitet prema matematičko-statističkim postupcima u društvenim naukama.

Korištenje ovog priručnika će kod studenata svakako smanjiti otpor prema psihometriji i olakšati potrebe nastavnika koji predaju ovaj predmet. Pored toga, njegova dodatna vrijednost se ogleda i u tome što ga mogu koristiti i studenti pedagogije, sociologije, socijalnog rada ili bilo koje profesije iz domena društveno-humanističkih nauka.

Imajući sve navedeno u vidu, uz zadovoljstvo što imam priliku biti recenzent najtoplije preporučujem knjigu „Osnove psihometrije za studente psihologije“ za štampanje.

S poštovanjem,

Doc. dr. Đorđe Čekrlija

Biografija

Dženana Husremović rođena je 14. januara 1971. u Zenici gdje je završila osnovnu školu i gimnaziju. Prije 1992. studirala je na Medicinskom fakultetu u Sarajevu. U periodu od 1992. do 1995. radila je na projektima psihosocijalne pomoći izbjeglicama i raseljenim licima i to iskustvo je uvelike odredilo njen daljnji karijerni put. Godine 1997. upisala je studij psihologije na Odsjeku za psihologiju gdje je diplomirala 2001. na temu iz područja profesionalne selekcije, a 2006. je magistrirala na temu „Validacijska studija upitnika profesionalnih interesa za učenike osnovnih škola“. Doktorirala je 14. januara 2011. pod mentorstvom prof. dr. Valentina Bucika na temu „Model odnosa psihosocijalnih faktora na radnom mjestu i njihovog utjecaja na dobrobit radnika”. Za vrijeme studija boravila je na usavršavanju na Institutu za psihologiju Ludvig Maksimilijan Univerziteta u Minhenu. U toku postdiplomskog studija boravila je mjesec dana na naučno-stručnom usavršavanju na Institutu za psihologiju Univerziteta u Geteborgu. Godine 2013. bila je stipendista US State Departmenta i kao JFDP stipendista bila na školovanju na Michigan State University, SAD. 2014. kao stipendista Britanske Ambasade boravila je na usavršavanju iz područja upravljanja i javnog zagovaranja na Queen Mary University u Londonu. Od oktobra 2001. zaposlena je na Odsjeku za psihologiju Filozofskog fakulteta u Sarajevu. Osim rada na matičnom fakultetu, Dženana Husremović predaje na Likovnoj akademiji predmet Ergonomija. Predavač je na postdiplomskim studijama pri Centru za interdisciplinarne studije Univerziteta u Sarajevu na predmetu Metodologija društvenih istraživanja. Glavna područja interesiranja Dženane Husremović su individualni i grupni fenomeni vezani za obrazovanje, rad, radne uvjete i razvoj karijere, te povezivanje istraživanja s javnim politikama u području obrazovanja i karijere. Bavi se evaluacijom projekata kojima je cilj unapređenje civilnog društva i kvalitete života građana. Predsjednica je Društva psihologa u Federaciji Bosne i Hercegovine i članica Upravnog odbora Fonda otvoreno društvo u Bosni i Hercegovini. Njena vizija je graditi vlastite kompetencije kako bi pomogla studentima u njihovom razvoju, te djelovati kao znanstvenik u podizanju kvaliteta obrazovanja i rada građana Bosne i Hercegovine.