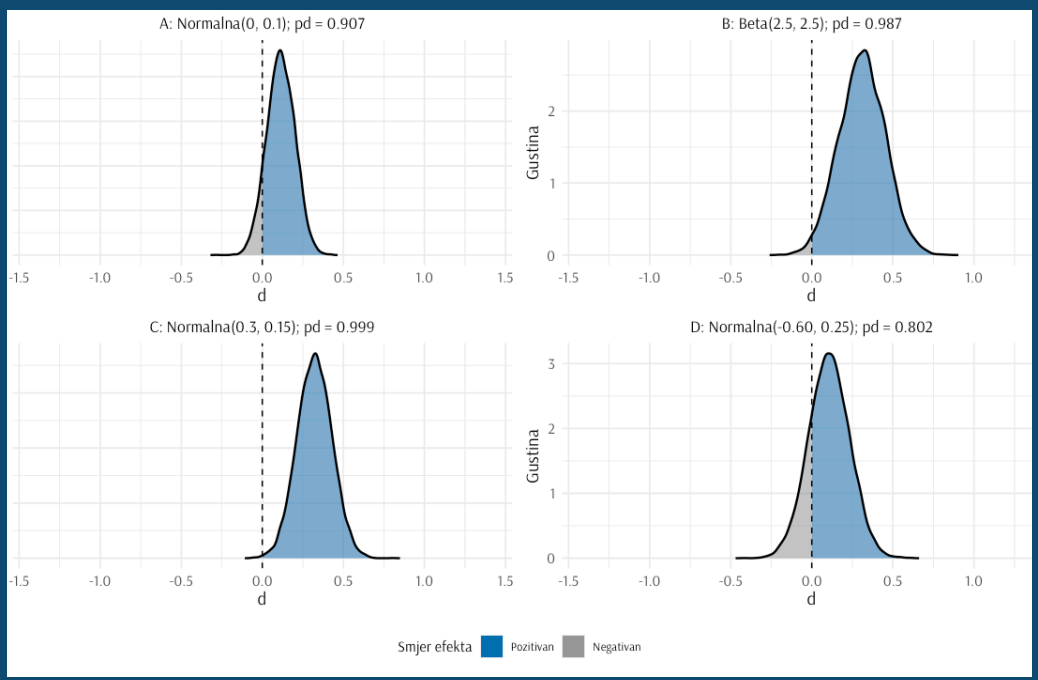
Siniša Lakić

STATISTIKA ZA PSIHOLOGE

Konceptualni uvod u kvantitativno zaključivanje



2025

Filozofski fakultet
Univerzitet u Banjoj Luci

Statistika za psihologe

Konceptualni uvod u kvantitativno zaključivanje

Siniša Lakić

Filozofski fakultet
Univerzitet u Banjoj Luci
Banja Luka, 2025.

Prof. dr Siniša Lakić

Statistika za psihologe: Konceptualni uvod u kvantitativno zaključivanje

Prvo izdanje

ISBN 978-99997-45-02-4

https://doi.org/10.63356/978-99997-45-02-4

Recenzenti

dr Nermin Đapo, redovni profesor (Filozofski fakultet Univerziteta u Sarajevu)

dr Vladimir Hedrih, redovni profesor (Filozofski fakultet Univerziteta u Nišu)

Lektor i korektor

dr Dragan Dragomirović

Izdavači

Univerzitet u Banjoj Luci

Filozofski fakultet Univerziteta u Banjoj Luci

Univerzitetski grad, Bulevar vojvode Petra Bojovića 1A, Banja Luka

Publikacija je odobrena kao univerzitetski udžbenik Odlukom Senata Univerziteta u Banjoj Luci broj 02/04-3.2253-77/25 od 23.10.2025. godine.

Za izdavače

Prof. dr Radoslav Gajanin, rektor

Prof. dr Srđan Dušanić, dekan

Sekretar za izdavačku djelatnost

Marko Aćić

Autor preloma i grafičke obrade publikacije

Siniša Lakić

Mjesto i godina izdavanja

Banja Luka, 2025.

Predgovor

Osnovni formalni razlog zašto je ovaj udžbenik pred vama jeste to što na Univerzitetu u Banjoj Luci predajem predmete *Statistika u psihologiji 1 i 2* sada već jednu deceniju, a njegov prvi proto-oblik (skripta) se pojavio već 2016. godine budući da nisam bio zadovoljan tadašnjom ponudom tekstova za studente. Mada je u međuvremenu objavljeno nekoliko zaista veoma kvalitetnih udžbenika na našim jezicima, tokom ovog podugačkog perioda je ipak ostalo dovoljno razloga da istrajem u njegovom pisanju. On je svakako drugačiji, ali ne znači nužno i bolji od postojećih udžbenika.[1] Kao i svaki autor, morao sam napraviti mnogo kompromisa pri odlukama o tome šta uvrstiti, a šta izostaviti, a da ostanem dosljedan načelima koje sam sebi postavio.

[1] Iskreno preporučujem i čitanje udžbenika drugih autora i poređenje, ako za to kao student ili "drugo zainteresovano lice" imate vremena.

Kao prvo, direktno iskustvo u nastavi sa brucošima koji dolaze iz srednjih škola u kojima su imali po samo jednu ili dvije godine matematike, te saznanja koja sam stekao u profesionalnom radu u području matematičke i čitalačke pismenosti, su me opredijelili ka tome da putovanje započnemo od samih osnova matematike, a da ton izlaganja bude minimalno formalan. Bilo mi je mnogo važnije uspostaviti i održavati živu vezu između psihičkih fenomena i kvantitativnog sistema rezonovanja nego li brzo se prebaciti u naprednu matematiku, tako da u ovom udžbeniku nećete naići na matrični račun.

Drugo, smatrao sam bitnim da pokušam zaraziti čitaoce fascinacijom detektivske igre na koju pozivaju podaci udruženi sa vizuelnim i numeričkim deskriptivnim tehnikama, nasuprot izražavanja strahopoštovanja s kojim se obično prikazuju formalne tehnike statističkog zaključivanja gdje se čitalac pretvara u bezličnog agenta koji uči i izvršava komande. Čini mi se da mnogi – naročito inostrani udžbenici prebrzo inženjerski nastupaju i prebacuju se na auto-put vjerovatnoće i statistike zaključivanja, pri čemu propuštaju da vide šta sve donose prizori lokalnih puteva sa raznovrsnim obrascima koji mogu biti mnogo interesantniji za psihologe u praksi.[2]

[2] Vjerovali ili ne, do ovih analogija sam došao sam, nisu proizvod AI četbotova.

Treće, trudio sam se da prikažem važne nedavne promjene

u psihološkoj statistici zaključivanja i epistemologiji, a što je dugogodišnji predmet mog profesionalnog interesovanja. Zato je detaljno – ali i vrlo kritički – izloženo standardno statističko zaključivanje na osnovu nulte hipoteze i p-vrijednosti, ali su prikazana i poglavlja o važnosti replikacija, izračunavanju statističke moći, te meta-analitičkom i bejzijanskom pristupu zaključivanju. I kada se pominju, tim temama nije dato dovoljno pažnje u tekstovima na našim jezicima.

Četvrto načelo koje sam pratio je da sam želio prikazati računarsko-tehnološki dio u dovoljnoj mjeri, ali da on ne stoji u fokusu. Nastojao sam da čitaocu pružim pogled "ispod haube" tako što tu i tamo pominjem softverska rješenja (redom open source alate), neophodne korake pri kreiranju baza podataka i čak sam začinio tekst sa nekoliko komandi u R programskom jeziku. Međutim, ovaj udžbenik nije namijenjen da samostalno uvede studente u rad na računarima, jer su uostalom i predmeti koje predajem osmišljeni tako da ih prate vježbe na kojima se uz pomoć posebnog priručnika prelaze konkretne tehnike na stvarnim podacima. Da tehnološki dio bude sporedan uticala je i sad svima dostupna sposobnost AI chatbotova da ispišu dobar kod za analizu u R-u ili Pythonu, te da čak samostalno izvedu kompletan izvještaj nakon što dobiju podatke i kvalitetan prompt.

Konačno, a to može izgledati posebno čudno onima koji su već čitali neke druge udžbenike, jeste to da knjiga završava regresionom analizom kao jedinom izloženom tehnikom statističke analize koja uzima u obzir više od dvije varijable. Takvu odluku sam donio zato što na internetu postoji ogroman broj tekstualnih i video materijala koji zainteresovane vode kroz provođenje različitih analiza, a u objašnjenjima teorijskih osnova se sve bolje snalaze i AI četbotovi. Međutim, da bi se došlo do te faze samoregulisanog učenja "prvo se mora naučiti ono što se naučiti mora". Prema mom viđenju stvari, to je fundamentalna logika opšteg linearnog modela u čiju strukturu se kasnije lako uklapaju ostale tehnike.

To su bile moje odluke pri pisanju udžbenika kakav bih volio da sam ga ja imao tokom svojih studija. Pritom, potpuno sam svjestan da moje izbore, preferencije i entuzijazam neće dijeliti svi čitaoci. Za ublažavanje količine kletvi koju bih inače dobio iskrenu zahvalnost dugujem onima koji su tokom proteklih godina ukazivali na tehničke greške i nejasne dijelove teksta. To su prije svega brojni studenti koji su silom prilika morali prolaziti kroz ranije verzije kako bi položili ispit, neke drage kolege

koji su se u odmaklijim godinama ohrabrili da se riješe straha od statistike, ali i moja djeca koja su - za manji honorar - ukazala na probleme sa čitljivošću za nove generacije.[3] Nadam se da ćete pročitavši ovu knjigu otkriti istovremeno zanimljiv i koristan univerzum ideja koji će vam olakšati sticanje novih znanja iz psihološke nauke, ali i omogućiti bolje rasuđivanje u profesionalnom i svakodnevnom životu.

[3] Honorar su dobili i Anthropic Claude i Google Gemini koji su mi uštedili mnoge sate pri kreiranju vizualizacija i snalaženju sa slaganjem knjige u RMarkdown / tufte / Latex okruženju.

Siniša Lakić
Banja Luka, juli 2025.

Sadržaj

Poglavlje 1

Uvod

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta je to zapravo statistika?
- Zašto je psihologiji i psiholozima potrebna statistika?
- Zašto je pogrešno pretjerano se osloniti na statistiku u psihologiji, čak i ako je ona vrlo korisna?
- Koje matematičke radnje je potrebno znati da bi se savladalo gradivo ove knjige?
- Kako izvesti osnovne i često neophodne matematičke radnje u računarskom okruženju?

Kakve asocijacije u vama izaziva riječ statistika? Nažalost, vjerujem da većini čitalaca prvenstveno naviru negativne asocijacije. Na primjer, dio matematike (*grozno!*) koji je toliko apstraktan i nejasan da su ga nastavnici preskakali u školi, jer i oni smatraju da to nije **prava matematika**? Podaci o zaraženim, vakcinisanim, hospitalizovanim i umrlim osobama tokom Covid-19 pandemije u periodu od 2020. do 2022. godine? Deprimirajući podaci o depopulaciji i emigraciji stanovništva u ekonomski stabilnije države, što prate podaci o niskim prosječnim platama u regionu? Podaci o stalnim mukama sa popisom stanovništva i borbama koje se tiču brojnosti etničkih, religijskih i jezičkih grupa na ovim prostorima? Uglavnom, vrlo vjerovatno je da te asocijacije nisu naročito vedre. Ovim tekstom ću pokušati da vas uvjerim da kada je upoznate, statistika može postati korisna alatka, pa i dobar drug, za snalaženje u svijetu pretrpanom informacijama.

Kako možemo definisati statistiku?

Logično bi bilo da upoznavanje počnemo definisanjem pojma statistika. Dvije stvari odmah napominjem. Prvo, pod statistikom ćemo podrazumijevati cijelu jednu disciplinu, a ne neke konkretne statističke podatke koji se mogu sresti u medijima (npr. statistiku nekog sportskog meča). Drugo, izuzetno je teško pronaći jedinstvenu definiciju koja bi uključila sve ono što obuhvatamo tom disciplinom. Zapravo, definicija ima mnogo, gotovo onoliko koliko i statističara. Ja sam osmislio jednu koja mi se čini dovoljno praktičnom da efikasno opisuje šta je njena suštinska uloga, te može da posluži kao prva pomoć:

"Statistika je pomoćna naučna disciplina čija je osnovna funkcija da pouči kako valjano prikupiti veći broj podataka i matematičkim postupcima svesti ih na manji broj relevantnih informacija, a kako bi se olakšalo opisivanje i razumijevanje pojave od interesa, te donijele što racionalnije odluke u vezi sa predvidivim ishodima." Ta definicija govori nešto o statusu statistike kao alata naučnog mišljenja, pominje matematičku osnovu, te određuje srž koja se sastoji u redukciji broja informacija kojima se lakše barata. Ono što mislim da je bitno jeste da je određuje i praktičnom namjenom koja je zajednička za svekoliku nauku: pojavu je potrebno prvo opisati, zatim razumjeti kako do nje dolazi, njeni oblici se onda mogu predvidjeti, a znajući sve navedeno potencijalno i kontrolisati.

Neke druge definicije s pravom naglašavaju da je statistika matematička poddisciplina koja se naročito oslanja na teoriju vjerovatnoće, dok druge detaljnije razrađuju različite aspekte koji predstavljaju sastavne dijelove statistike, npr. proučavanje načina kako adekvatno prikupiti podatke, kako podatke analizirati različitim tehnikama i kako ih nakon sprovedenih analiza prikazati i interpretirati. Ima onih koje dopunjavaju navedeno naglašavajući to da se statistika bavi pronalaženjem pravilnosti u svijetu sa neizvjesnim događajima pomoću teorije vjerovatnoće. Svi ovi elementi zaista su bili (i još uvijek jesu) sastavni dio onoga što zovemo statistikom, ali po mom mišljenju, nije nužno da se nađu u njenoj definiciji. Nemam problem sa tim ako zaključite da neka od alternativnih definicija bolje predstavlja statistiku.[1]

[1] Neke česte, alternativne definicije možete pročitati ovdje: https://sr.wikipedia.org/sr-el/statistika i https://www.enciklopedija.hr/clanak/statistika

Ima li statistika bliske rođake?

Osim dilema u vezi definisanja samog pojma statistike, potrebno je naglasiti da postoje neki manje ili više srodni termini koji su u

širokoj upotrebi. Jedan od takvih, konkurentnih termina je **nauka o podacima** (engl. *data science*). Grubo rečeno, pod naukom o podacima obično podrazumijevamo mješavinu upotrebe statističkih metoda, masivnih baza podataka, različitih programskih algoritama i konkretnih znanja o datom domenu kojima se rješavaju vrlo konkretne potrebe određene kompanije ili institucije. Klasičan primjer bi bila upotreba ogromnih baza podataka na kojima se uz pomoć mašinskog učenja kreira algoritam kojim se mejlovi koji vam pristižu na elektronsku adresu klasifikuju kao *spam* mejlovi ili sigurni mejlovi. Za razliku od te novije nauke, stara gospođa statistika bi se za to vrijeme bavila samo "uzvišenim" pitanjima naučne generalizacije.

Pojmova koji upadaju u širu kategoriju nauke o podacima ima još i moguće je da ste čuli za njih. Na primjer, **rudarenje podataka** (engl. *data mining*) opisuje procese kojima istraživači tragaju za novim obrascima među podacima, a već pomenuto **mašinsko učenje** (engl. *machine learning*) označava procese gdje se ljudi minimalno upliću u traženje postojećih obrazaca i puštaju već kreirane algoritme da sami dođu do njih na osnovu statističkih načela. Mašinsko učenje stoji u osnovi sveprisutnih alata tzv. vještačke inteligencije. Zapravo, sve je teže razdvojiti statistiku od brojnih novih pojmova, pa je možda najbolje nazvati ono čime se želimo baviti kvantitativno rezonovanje ili naučno zaključivanje zasnovano na kvantitativnim podacima. Dakle, imajte na umu da - od ovog mjesta nadalje - kad god spomenem termin statistika ja prvenstveno mislim na to, nešto šire značenje riječi. Iz toga slijedi da ćemo u ovom tekstu govoriti o tradicionalnim statističkim metodama, ali i o novim načinima zaključivanja koji ne proizilaze samo iz teorijske matematike, nego i iz dostignuća dostupnih zahvaljujući napretku računarske tehnologije.

Koje su prednosti upotrebe statistike u psihologiji?

Nego, funkcioniše li taj spoj statistike i psihologije? Nasuprot laičkim shvatanjima - sjetite se sebe prije nego što ste počeli studirati psihologiju, bilo da je to prošle sedmice ili prije dvadesetak godina - shvatanjima prema kojima su proučavanje duševnog života i statistika nekompatibilni, upotreba kvantitativnih metoda u psihologiji ima dugu istoriju i već odavno dominira njenim naučnim aparatom.[2]

Zašto je to tako? Postoji više razloga, a ja ću ovdje izdvojiti četiri. Naime, statistika donosi psihologiji:

[2] Na primjer, jedno istraživanje (Scholtz, de Klerk & de Beer, 2020, https://doi.org/10.3389/frma.2020.00001) pokazuje da kvantitativne studije čine čak 90% ukupnog broja psiholoških naučnih članaka u nekoliko zaista bitnih časopisa.

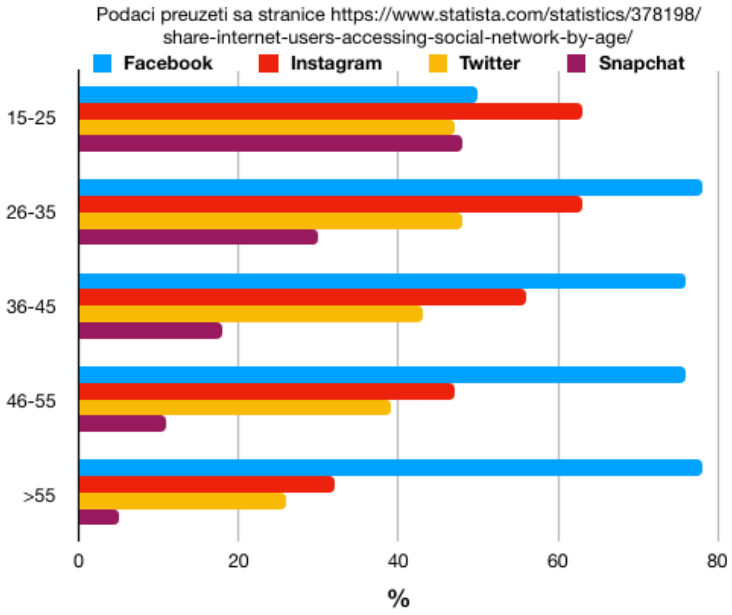
1. **objektivnost** Statističke tehnike omogućavaju da objektivno vrednujemo valjanost teorija i praktičnih intervencija, a sve to na osnovu prikupljenih empirijskih podataka. Mnogo su uvjerljiviji nalazi koje može provjeriti svako ko to želi ako ima pristup tim numeričkim podacima, nego ako se isključivo oslonimo na subjektivne iskaze, tumačenja i ubjeđivanja. Da li biste prije otišli na tretman depresije kod lokalne vračare koju hvali komšinica ili biste odabrali jedan od psihoterapijskih tretmana za koje su prikupljeni statistički dokazi da oni pozitivno djeluju na većinu klijenata?[3]

2. **preciznost** Statistika omogućava da se prikažu koliko su velike razlike među grupama ljudi, koliko snažne su povezanosti između različitih osobina i/ili koliki je intenzitet prisustva pojave unutar neke grupe, a ne zadržavamo se samo na tome da ustanovimo da li nešto postoji ili ne. Na primjer, poznato je da studenti psihologije imaju značajan strah od statistike. Međutim, jedna je stvar reći da takvih studenata ima, a druga da dvije trećine studenata psihologije doživljavaju umjereno snažnu anksioznost i da su metodološko-statistički predmeti prvi na listi studentske horor liste. Nas koji predajemo takve predmete ovakvi podaci opominju da se ne radi o pojedincima, nego o značajnoj većini, pa moramo da vas nježno "rasplašujemo" ovakvim tekstovima.

3. **efikasnost** Kao što smo već u definiciji to naglasili, statistika je sjajno sredstvo za konciznu deskripciju, kategorizaciju, generalizaciju, analizu, predikciju, modeliranje i komunikaciju rezultata. Danas je moguće u jedan grafikon smjestiti - bukvalno - nekoliko milijardi sirovih podataka. Koristeći grafikone i tabele na prikupljenim podacima možete, čak i ako ste laici, vrlo lako uočiti pravilnosti što ne bi bilo moguće bez novih tehnologija i statistike. Evo primjera koji nije striktno vezan za psihologiju, ali jeste za svakodnevni život. Grafikon prikazuje koliko često su internet korisnici u SAD-u tokom jednog kvartala 2020. godine pristupali četiri veoma popularnim društvenim mrežama, uzimajući u obzir starost ispitanika. Možete li da uočite neki trend?

4. **ugled** Zahvaljujući samim svojim prisustvom statistika čini psihologiju da izgleda naučnije, jer je time svrstava

[3] Primjera radi, jedan od repozitorija koji održava Američko psihološko udruženje (APA) nudi na uvid za koje su to psihološke tretmane prikupljeni naučni dokazi o djelotvornosti: https://div12.org/treatments/



Slika 1.1: Korišćenje društvenih mreža po godinama u SAD-u.

u društvo kvantitativnih nauka. Štaviše, istorija psihologije je značajno obilježena jednom debatom unutar Britanskog udruženja za unapređenje nauke koje je 1940. godine zaključilo da društvene nauke ne zadovoljavaju neophodna načela mjerenja. Kao reakciju na to, psiholog S.S. Stevens je osmislio posebnu teoriju mjerenja - koju detaljno obrađujemo u sljedećem poglavlju knjige - što je psihologiju gurnulo nekoliko koraka naprijed u naučnom statusu. I ne radi se samo o "šminci". Time što dijele zajedničku kvantitativnu metodologija sa mnoštvom drugih, "tradicionalnijih" nauka, psiholozi su osposobljeni da bolje razumiju druge discipline, kao i da lakše komuniciraju svoje rezultate opštoj naučnoj publici - što je od ogromne važnosti za sve prisutniju interdisciplinarnost naučnog znanja. Naime, danas smo svjesni da velika većina psihičkih fenomena paralelno imaju uzroke i posljedice u sociološkoj, istorijskoj, ekonomskoj, tjelesnoj, biohemijskoj, nutritivnoj, fiziološkoj, biološkoj, informacionoj ravni…

Gore navedeno govori u prilog tome da je ovaj predmet morao doći do vas već na prvoj godini studija psihologije. Svaki psiholog bi trebalo da može razumjeti rezultate dosad objavljenih naučnih istraživanja, te kreirati vlastita naučna istraživanja koristeći statističke principe. Uz to, neophodno je da kao odgovorni članovi društva znate ispravno tumačiti statističke informacije. Tim prije što smo tokom Covid-19 pandemije svjedočili poplavi

biblijskih razmjera kada su u pitanju informacije, te namjerna ili nenamjerna iskrivljavanja istine i različite zablude. Dakle, poznavanje statistike je sastavni dio opšte pismenosti u informacionoj eri (čiji smo stanovnici), gdje su podaci sveprisutni i korisni u svakoj oblasti ljudskog djelovanja, a naročito u nauci gdje predstavlja elegantan jezički sistem kojim se prenose saznanja o svijetu.

Koja su ograničenja upotrebe statistike u psihologiji?

Bez obzira na navedene prednosti treba uvijek imati na umu da upotreba statistike u psihologiji ima i svoja značajna ograničenja, od kojih su najznačnija sljedeća:

1. **mjerenje** Psihičke fenomene uopšte nije lako mjeriti. Zato se uvijek mora postaviti pitanje koliko su prikupljeni podaci na kojima se vrši statistička analiza smisleni. Važnost tog pitanja dobro oslikava načelo koje se na engleskom skraćeno naziva GIGO (Garbage In, Garbage Out), a na naš jezik se može prevesti kao SUSI: Smeće Uđe, Smeće Izađe. Drugim riječima, ako veoma apstraktne atribute od kojih vrvi u psihologiji - kao što su autoritarnost, strategije nošenja sa stresom, ekstraverzija, stepen moralnog razvoja, narcizam - ne izmjerimo kako treba, onda i zaključci koje izvedemo uz pomoć statističke analize neće baš biti smisleni. Usljed ogromnog značaja mjerenja razvijena je i zasebna oblast, **psihometrija**, koja se bavi osiguravanjem kvaliteta mjerenja u psihologiji. Uprkos velikim naprecima u toj oblasti, te napretku informacionih tehnologija i tehnologija mjerenja fizioloških i biohemijskih korelata ponašanja i mentalnih stanja, stanje je još daleko od savršenog, te morate kritički preispitivati svaku pročitanu naučnu informaciju time što ćete obratiti pažnju na to koliko je dobro neki atribut mjeren.

2. **prepojednostavljivanje** Iz praktičnih razloga smo u naučnim istraživanjima prinuđeni svesti ukupnost ličnosti, društveno-kulturološkog konteksta i ponašanja na mali broj izdvojenih varijabli. Time stvarnost psihičkog života pretjerano pojednostavljujemo. Brojna istraživanja imaju zaista minimalan broj karakteristika koje se dovode u vezu. Uzmimo za primjer istraživanja polnih razlika u pogledu inteligencije, motivacije postignuća ili stavova o seksualnosti gdje ih imamo samo po dvije: pol i tu drugu

karakteristiku od interesa. Slikovit primjer za studente psihologije je i nešto praktično što ste nedavno doživjeli, a tiče se selekcije. Na prijemnom ispitu na psihologiji na našem univerzitetu kandidate rangiramo na osnovu samo tri podatka: prosječne ocjene u srednjoj školi, rezultatu na testu znanja iz srednjoškolske psihologije i rezultata na testu kognitivnih sposobnosti. Istina jeste da smo na našem fakultetu statistički ustanovili da su na ranijim generacijama sve tri varijable pozitivno predviđale uspjeh u studiranju psihologije, a taj uspjeh u studiranju smo izrazili kao prosječnu ocjenu na kraju studija. Ali, može se postaviti pitanje da li smo izostavili da izmjerimo još neke bitne karakteristike koje su nam mogle pomoći da donesemo još bolje odluke pri izboru kandidata, kao što su empatičnost, introspektivnost, trenutna anksioznost na prijemnom ispitu, moralne vrijednosti, radoznalost, matematičko-statistička znanja :) ?

3. **uprosječavanje** Statističke tehnike su većinom usmjerene na prikazivanje tendencija prosjeka što u velikom broju slučajeva uopšte nije predmet interesovanja niti psihologa praktičara, niti laičke javnosti. Zapravo je nerijetko upravo suprotno slučaj, a to je da nas interesuju ekstremni pojedinci. Konkretno, prosječni sportisti, prosječni prekršioci zakona, prosječni umjetnici privlače manje pažnje, i javne i profesionalne. Više nas zanimaju karakteristike najboljih sportista, višestrukih ubica, izuzetnih umjetnika. Ovo ne znači da su statističke analize na prosjecima neupotrebljive, nego samo da treba imati na umu to ćemo mnogo puta dobijamo nalaze koji nam nešto govore o prosječnim tendencijama, a ako je nešto u prosjeku tačno, ne znači nužno da to važi i za one najposebnije. Doduše, u novije vrijeme se više pažnje posvećuje razvoju specijalizovanih statističkih tehnika kojima se istovremeno analizira uloga sadejstva pojedinačnih karakteristika (npr. inteligencija, marljivost, emocionalnost, traumatična iskustva, raniji uspjeh) koja u različitim kombinacijama dovodi do ciljanog ponašanja (npr. ekstremni uspjeh ili patologija).

4. **uzorkovanje** Iz praktičnih razloga, psihološka istraživanja rijetko poštuju principe idealnog odabira ispitanika. Istraživanje koje bi bilo bazirano na popisu stanovništva i koje bi adekvatno ispitalo različite slojeve stanovništva iziskuje velike novčane i resursne troškovi. Mnogo je jednostavnije uzeti za istraživanje nekog ko je lako dostupan, a na kome se može “obaviti posao”. Hm, ko bi to mogao biti? Pada li

uzorkovanje (glagol ”uzORkovati” znači ”uzimati uzorak”, dakle NE misli se na ”uzROkovanje”)

vam neka grupa na pamet? Naravno, studenti psihologije, pogotovo prve godine studija! Nema sumnje, biti ispitanik u mnogim istraživanjima jeste višestruko korisno za buduće psihologe. Osim što upoznajete sebe, time što učestvujete u nekom istraživačkom protokolu iznutra upoznajete i mehanizam naučnog istraživanja.

Ali avaj, možete i sami da pretpostavite da studenti psihologije nisu baš tipični predstavnici ljudskog roda. Studenti psihologije su, generalno gledajući, "fini" pripadnici ljudske vrste: prijatni, otvoreni za iskustva, natprosječno inteligentni, emocionalno labilni (jeste), generalno empatični prema patnjama drugih, većinom ženskog pola. A kad još shvatimo da jedna šira grupa "čudaka" (engl. *WEIRD*[4]) zna činiti i preko 95% uzoraka u psihološkim istraživanjima, a da istovremeno čine tek 12% ukupnog čovječanstva, moramo se zapitati da li nalazi većine istraživanja zaista dobro opisuju stvarnost.

[4] Akronim WEIRD dolazi od engleskih riječi Western, Educated, Industrialized, Rich, Democratic, što bi značilo da se radi o ljudima koji žive u zapadnim, obrazovanim, industrijalizovanim, bogatim i demokratskim društvima. Vrlo popularan članak koji je posebno detaljno istražio ovaj problem su napisali Henrich, Heine i Norenzayan, 2010, https://doi.org/10.1017/S0140525X0999152X.

Mimo navedenih, postoji još problema. Jedan se pojavljuje kada se – nećete vjerovati – sa statistikom u psihologiji baš pretjera. Dakle, ima među psiholozima i statističkih idolopoklonika. Nekritička upotreba statistike je dovela do svojevrsne tehnologizacije društvenih nauka i doprinijela pojavi nečega što se naziva scijentizam.

Scijentizam se može definisati kao stav nekritičnog glorifikovanja naučnog saznanja nastalog na osnovu metoda koje koriste „tvrde", prirodne nauke. U to spadaju i mjerenje kojim se dodjeljuju numeričke vrijednosti objektima istraživanja, te posljedična kvantitativna obrada podataka. Šta je zapravo tu problematično? Naime, često se dešava da istraživači psiholozi zadovolje formu naučnog rada tako što koriste standardizovanu recepturu za njegovo pisanje: poštuju uspostavljenu strukturu naučnih članaka, upotrebljavaju visokoparan „tehnički" jezik i prikazuju rezultate veoma komplikovanih statističkih metoda koji su izvedeni na kvantifikovanim podacima. Bez obzira na ispunjenost tih formalnih uslova i opažanje čitalačke publike da se radi o „pravoj nauci", takvi radovi mogu biti zapravo suštinski beskorisni ili čak da donose pogrešne zaključke usljed toga što su korištene kompleksne statističke procedure, ali na slabim podacima. To je kao kada bi vam neko dao da vozite putnički avion sa stotinama različitih komandi i instrumenata, a vi znate samo da vozite bicikl. Logično, da biste mogli razlikovati naučne istine od formalnih *scijentističkih* predstava morate se izuzetno dobro upoznati sa statistikom, otkloniti strepnje koje imate u pogledu njene nepristupačnosti i istinski se sprijateljiti s njom.

Nije zgoreg podsjetiti da postoji i još podmuklija zloupotreba statistike u javnoj sferi. Naime, manje ili veće "mračne sile" mogu namjenski odabirati i na iskrivljen način prikazavati statističke podatke kako bi među ciljanom publikom ostvarili željeni cilj iskorištavajuće kognitivne pristrasnosti kojima smo kao ljudi skloni. Nije rijetkost da plaćene istraživačke agencije naprave takva istraživanja ili ih kasnije "dorade" kako bi se njima pokazalo da određeni politički kandidat ima veće šanse na izborima (te time privuku glasove neopredijeljenih kako bi bili na pobjedničkoj strani ili kako bi pojačali misao da nema svrhe izaći na izbore), da neka marketinška ili farmaceutska reklama predstavlja izdvojene rezultate koji govore o zadovoljstvu kupaca proizvodom (a ne pomenuti brojeve vezane za pritužbe i loša iskustva), te da određeni politički akteri naglašavaju dobre ili loše društvene trendove tako što ne prikazuju čitavu sliku stanja (koristeći statističke tehnike koje ih prikazuju u željenom svjetlu). Dobro poznavanje statistike će vas imunizovati od takvih zloupotreba i opremiće vas da budete član društva kojeg je teže izmanipulisati. Nadam se da je i to dobar motivator da se zdušno posvetite učenju ove oblasti koja se možda čini apstraktnom, teškom i nepotrebnom. Ali prije nego što se konačno upustimo u izučavanje statistike, moramo da ponovimo neophodna matematička znanja što ćemo uraditi kroz upotrebu računara, budući da statistika bez upotrebe računara odavno već nije smislena

Koja matematička (i računarska znanja) su nam neophodna?

Očekujem da vam se pojavi osmijeh na licima kada pročitate da je preduslov za savladavanje gradiva samo da znate ono što ste već učili iz matematike još u osnovnoj školi. Neke naprednije pojmove, na primjer one vezane za teoriju vjerovatnoće, ćemo nanovo zajedno izučavati. Iskustvo u nastavi mi govori da je ipak potrebno da se zajedno podsjetimo i naj-naj-najosnovnijih matematičkih termina i operacija. Dakle, ne zamjerite mi ako izgleda kao da većinu vas potcjenjujem, jer će ipak nekima dobro doći sljedeći dio. Naime, na našem univerzitetu je nezanemariv broj studenata psihologije matematiku u srednjoj školi imao samo tokom dvije ili čak jedne školske godine što znači da su u međuvremenu izgubili potrebu i da je koriste. Istovremeno, zadatak sljedećeg dijela će biti da osvježimo i obogatimo znanje kompjuterskog označavanja (tj. notacije) osnovnih matematičkih

radnji. Pritom, koristiću komande otvorenog programskog jezika *R*, namjenski stvorenog za statistiku, programskog jezika koji će nas pratiti kroz ovaj tekst.

R

R (ne čitati po Vuku kao *r*, nego *er*, ili ako hoćete po engleski $a:^r$) je potpuno besplatan program čiji instalacioni fajl za tri velika operativna sistema (Linux, Windows, OSX) možete preuzeti sa stranice https://cran.r-project.org. Radi se o veoma moćnom kalkulatoru kojeg entuzijastično nadograđuje ogromna zajednica korisnika - gomila naučnika i nerdova - tako što programira specijalizovane programčiće (tzv. pakete). Te pakete možete učitati u memoriju i izvršavanjem jednostavnih komandi izvršiti teško zamislivo komplikovane analize.[5] Na internetu možete pronaći silne edukativne materijale i namjenski pisane knjige o R-u, pa ću u nastavku teksta pominjati samo ono što je neophodno za razumijevanje statistike koju obrađujemo zajedno. Kada god budem prikazivao nešto izvedeno u R okruženju, to će biti prikazano na sljedeći način.

[5] Štaviše, uz pomoć R-a možete i napisati knjigu, nalik na ovu pred vama, kao što sam ja uradio.

```
print("Hej vidi me, ja sam rečenica napisana u R-u!")
## [1] "Hej vidi me, ja sam rečenica napisana u R-u!"
```

Dakle, font će tada biti drugačije vrste, vrijednosti koje se unose u samu komande će biti i različite boje (pod uslovom da ovo ne čitate u crno-bijeloj štampi). Rezultati naredbe će se pojavljivati iza dvije, tradicionalno rečene "tarabe", ili vrlo moderno rečeno, "heštega" (##). Rezultati naredbe mogu sadržavati više elemenata i iz tog razloga će u uglastim zagradama biti navedeni redni brojevi početnog elementa u datoj liniji rezultata. Uz to, hešteg znak se koristi i unutar koda za njegovo komentarisanje, odnosno on kaže R-u "zanemari sve što ide nakon ovog znaka, to je komentar za ljude."

```
# Primjer sa većim brojem elemenata
1:30 # Ispiši brojeve od 1 do 30
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
 ↪  16 17 18 19 20 21 22 23 24 25
## [26] 26 27 28 29 30
```

Sad možemo da se vratimo matematici. Ekstremno pojednostavljeno rečeno, matematika se bavi veličinama i odnosima između njih, koje opisujemo unutar simboličkog sistema. U stvarnosti baratamo sa različitim konkretnim predmetima i vrijednostima (npr. jabuke, novac, stolovi za kojim sjedite), a brojni sistem smo konstruisali kao apstraktnu ravan u koju prenosimo odnose iz

prebrojavanje naspram mjerenja

realnosti. Brojeve nečemu dodjeljujemo ili na osnovu **prebrojavanja** elemenata iste vrste (npr. koliko jabuka imamo; koliko novca imamo) ili na osnovu **mjerenja** veličine dimenzija (npr. koliko kilograma jabuka kupujemo; kolika je dužina, kolika širina, a kolika visina stola za kojim sjedite).

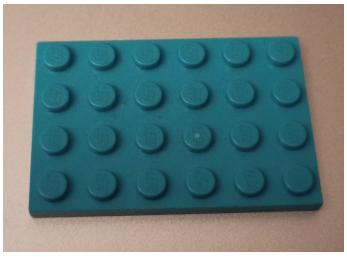
Odmah je neophodna napomena: potrebno je biti svjestan da čak i kada je u pitanju fizička stvarnost mi se trudimo da prebrojavamo i mjerimo savršeno, ali nam to ne uspijeva uvijek. Naime, kada na pijaci kupujemo jabuke, njih četiri mogu težiti 1kg ili 0.5 kg, pritom neke od njih mogu biti i gnjile i/ili crvave, pa ne znamo ni da li ih trebamo ubrojati u taj skup jabuka ako nam recept traži 1kg jabuka da napravimo štrudle. Uz to, nekad nemamo dobar instrument, odnosno vagu (npr. prodavac ju je malo "doradio"), te mjerimo ko zna kakve mase. Sve u svemu, broj koji dobijemo ne znači uvijek ono što želimo. Poenta je da je matematika idealistička, "presavršena" predstava realnosti, ali koja uprkos toga itekako može da koristi kada ste neko ko kupuje voće ili kada ste treba da naručite od stolara posebnu veličinu stola. Ono što je bitno da zapamtite jeste da nekada brojem izražena slika predstavlja apsolutno sve ono što nam treba, a nekad je to vrlo osiromašena slika stvarnosti. Sve u svemu, dosta toga zavisi od kvaliteta mjerenja, a tim pitanjem ćemo se malo pozabaviti u sljedećem poglavlju.

Četiri osnovne matematičke radnje su sabiranje, oduzimanje, množenje i dijeljenje. Iz njih su dodatno izvedene još dvije koje ćemo često koristiti, kvadriranje i korijenovanje. Sabiranje (dodavanje, uvećavanje) i oduzimanje (umanjivanje) neću ovdje teorijski dalje pojašnjavati, jer ipak nismo u prvom razredu osnovne škole. Ono što treba da znate, ako niste, jeste da se u računarskoj notaciji, kao i u "rukom pisanoj" matematici koriste znakovi + i -. Za decimalni znak se obično u statistici, pa tako i u ovoj knjizi (bez obzira na lokalne konvencije), umjesto zareza koristiti tačka, tako da ćemo sljedeći izraz ovako pisati:

```
5 + 3 - 1.5
## [1] 6.5
```

Kada je u pitanju množenje, umjesto uobičajenih znakova $\cdot$ ili $\times$ (ili označavanja množenja bez znaka za operaciju, kao $5(2+3)$) za operaciju množenja koristićemo tzv. asterisk *, pa će operacija množenja na računaru izgledati ovako:

```
4*6
## [1] 24
```



Slika 1.2: Lego pločica — ilustracija množenja kao sabiranja.

Ovdje je dobro podsjetiti da je množenje samo efikasnije prikazano sabiranje. Dakle, možemo posmatrati gornji izraz kao zbir vrijednosti za sve članova skupa od četiri elementa od kojih svaki element ima vrijednost šest. To se može ilustrovati sljedećom slikom jedne Lego pločice koju sam pozajmio od svoje djece - gdje se mogu prebrojati kružići.

Niz brojeva - ili još bolje *niz vrijednosti*, jer se koriste i slovni elementi umjesto brojeva - se u statističko-matematičkoj terminologiji često naziva **vektor**.

Radi lakšeg baratanja njima, vektore obično i imenujemo, pa ćemo ovaj vektor "krstiti" kao *lego_kruzici*.

```
lego_kruzici <- c(6, 6, 6, 6) # znak <- čitamo "ime
↪  za",
# c ispred zagrade od engleskog "concatenate", što
↪  znači "naniži, stavi u niz" - nizanje se izvodi
↪  pomoću zareza
```

Matematičari vole grčka slova, ima ih dosta i u statistici, a pretpostavljam da značenje jednog od njih već znate. Veliko grčko slovo sigma Σ označava zbir niza brojeva.

Obično ga dodatno opisuju još i oznake koje se pišu ispod i iznad tog slova, a označavaju sljedeće: i = indeks - slovo koje označava kako ćemo nazivati sve redne brojeve elemenata datog niza (dakle, u ovom slučaju ih zovemo i), m = redni broj početnog elementa u nizu, n = redni broj posljednjeg elementa u nizu. Prema tome, za zbir svih članova skupa bi kompletna oznaka bila $\sum_{i=m}^{n}$. Budući da na našem primjeru sa Lego pločicom imamo ukupno 4 elementa i da nas interesuje da saberemo sva četiri, od prvog do posljednjeg, odnosno četvrtog, onda ćemo to pisati ovako: $\sum_{i=1}^{4}$.

```
# komanda sum izvršava izračunavanje zbira svih
↪  elemenata vektora lego_kruzici, tj. 6+6+6+6
sum(lego_kruzici)
## [1] 24
```

Kvadriranje (x^2) je specijalni slučaj množenja, gdje se broj množi sa samim sobom.

Kada je u pitanju prebrojavanje kao rezultat se dobija kvadratna matrica sa jednakim brojem elemenata u redovima i kolonama, dok je za mjerene kontinuirane dimenzije to površina kvadrata koja se dobija na osnovu dužine ivice kvadrata. Evo i kvadratne Lego pločice koja to ilustruje, s tim da vas molim da zamislite da je dužina ivice zaista jednaka 4 centimetra.

Slika 1.3: Lego pločica — kvadriranje.

Računarskom notacijom kvadriranje možemo izraziti množenjem datog broja samim sobom (x*x), ali postoji i specijalno označavanje (x^2).

```
4^2 # jednako kao da smo naveli 4*4
## [1] 16
```

Više stepenovanje - odnosno množenje istog broja više puta - kao što je podizanje broja tri na treći stepen ($3 \cdot 3 \cdot 3$), tj. 3^3, se tako efikasnije izražava na sljedeći način:

```
3^3
## [1] 27
```

Dolazimo do dijeljenja. Ako se slučajno toga ne sjećate, dijeljenje je radnja direktno inverzna, odnosno obrnuta u odnosu na množenje. Ako neki broj n dijelite sa 4, onda ga u stvari množite sa njegovom inverzijom $\frac{1}{4}$. Recimo da želite podijeliti 24 na 4 jednaka dijela, to onda možete napisati i ovako $24 : 4 = 24 * \frac{1}{4} = \frac{24}{4}$.

Ovo znači da su dijeljenje i razlomci u praktičnom smislu jedna te ista stvar. Ono što nekad zbunjuje je da dijeljenje podrazumijeva operaciju sa dva zasebna broja, a razlomak predstavlja jedan broj koji se može napisati i uz pomoć decimalnih brojeva ako postoji neki ostatak. Kada je u pitanju računarska notacija za dijeljenje ćemo, umjesto kod nas upotrebljavanog školskog znaka :, koristiti znak /, odnosno kosu crtu koja počinje dole lijevo, a završava gore desno (engl. slash).

```
24/4
## [1] 6
```

Dijeljenje je komplikovanije od množenja usljed mogućnosti da kao rezultat operacije ne dobijete cijeli broj, iako ste za operaciju koristili dva cijela broja. Primjer:

```
10/3
## [1] 3.333333
```

Ovo je posebno značajno da se napomene, jer se tiče zaokruživanja brojeva, a zaokruživanje vas u statistici čeka iza svakog ćoška :). U praksi ćemo najčešće zaokruživati brojeve na jednu, dvije ili tri decimale, ali moguće je da ćete ih zaokruživati i na cijele brojeve ili na veći broj decimala. Od konteksta će zavisiti na koliko decimala biste to trebali radite, mada postoje neke konvencije o kojima će biti kasnije riječi. Evo kako to izvesti u R-u.

```
round(10/3, digits = 3) # komanda round, parametear
↪  digits određuje broj decimala
## [1] 3.333
round(10/3, digits = 2)
## [1] 3.33
round(10/3, digits = 1)
## [1] 3.3
round(10/3, digits = 0)
## [1] 3
```

statističko zaokruživanje

Međutim, postoji jedna bitna razlika u odnosu na ono na šta ste navikli u osnovnoj i srednjoj školi. Naime, polovine se ne zaokružuju automatski na veći broj. Drugim riječima, ako ste učestvovali u popularnim školskim pohodima na višu zaključnu ocjenu, onda ste bili sretni kada dobacite do polovine, jer biste i vi i nastavnici podrazumijevali da je $3.50 = 4$, a $4.50 = 5$. Uobičajeno statističko zaokruživanje je drugačije, jer se polovine zaokružuju na parni broj.

To bi značilo da bismo dobili da je $3.50 = 4$, ali i da je $4.50 = 4$. Zašto sad tako? Ovo se radi kako bi se eliminisalo potencijalno precjenjivanja prosječnih vrijednosti (pogotovo na velikim uzorcima gdje očekujete podjednak broj parnih i neparnih brojeva). Na primjer, ako imate u skupu samo dva broja (3.50 i 4.50) onda biste dobili prosjek 4.0 i prije zaokruživanja i poslije zaokruživanja na novi način, a ne 4.50 što biste dobili kao prosjek ako biste zaokruživali prema ranijoj navici. I R, kao zadrti statističar, koristi statistički način zaokruživanja što možemo da vidimo pomoću sljedeće komande.

```
round(3.50, digits = 0)
## [1] 4
round(4.50, digits = 0)
```

```
## [1] 4
```

množenje i dijeljenje brojevima između 0 i 1

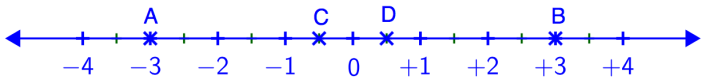
Budući da ćemo u statistici često koristiti brojeve koji su u rasponu između 0 i 1, važno je napomenuti još jednu pravilnost koju ste možda, jer je vaš matematički aparat zarđao, zaboravili. Ako neki broj pomnožite brojem koji je veći od 0, a manji od 1, kao rezultat ćete dobiti broj koji je manji od originalnog množioca. S druge strane, rezultat dijeljenja sa takvim brojem će biti veći broj. Zašto? Brojevi između 0 i 1 ukazuju na dijelove ili djeliće koji su manji od nečeg cijelog. Na primjer 0.1 označava jednu desetinu nečeg cijelog ($\frac{1}{10}$). To može biti "kockica", ili nešto tačnije rečeno kvadratić dobre čokolade na slici (Slika 1.4). Da biste kao rezultat gomilanja komadića imali na raspolaganju 1 cijelu čokoladu morate da skupite čak 10 takvih jednakih dijelova. Matematički izraženo: $10 * 0.1 = 10 * \frac{1}{10} = \frac{10}{10} = 1$. Kada je u pitanju dijeljenje onda zapravo sada prebrojavamo koliko takvih komadića ima u cjelini, a njih ima više nego cijelih predmeta.

Slika 1.4: Slatko-gorko dijeljenje.

Recimo da nam komadić od interesa predstavlja jedan red, odnosno dvije "kockice", odnosno dvije desetine, odnosno jedna petina, odnosno 0.2 cijele čokolade :). Kada želimo da cijelu čokoladu dijelimo na jednak broj tih petina, onda to možemo izraziti matematički putem ranije pomenute inverzije, pa umjesto dijeljenja možemo koristiti množenje: $1/0.2 = 1/\frac{2}{10} = 1 * \frac{10}{2} = 5$. Pomenute pravilnosti važe i onda kada je početni množilac veći od 1, kao i kada su oba broja veća od 0, a manja od 1:

```
5*0.1
## [1] 0.5
5/0.1
## [1] 50
0.2*0.1
## [1] 0.02
0.2/0.1
## [1] 2
```

Kao što je dijeljenje inverzno množenju, tako je i vađenje kvadratnog korijena inverzno kvadriranju. Konceptualno i geometrijski gledano, korijen nekog broja je zapravo dužina ivice kvadrata. Ako je površina kvadrata jednaka $9cm2$, onda je korijen, odnosno dužina njegove ivice jednaka 3 cm. Ukoliko je površina kvadrata jednaka $50cm2$, onda je korijen, odnosno dužina stranice jednaka, približno, 7.07 cm. U R-u koristimo



Slika 1.5: Brojevna prava.

komandu $sqrt(x)$ od engleskog "*square root*", odnosno **kvadratni korijen**.

```
sqrt(9)
## [1] 3
sqrt(50)
## [1] 7.071068
```

Međutim, neki od vas znaju da gore navedeni primjeri zapravo imaju dva tačna rješenja. Korijen iz 9 jeste zaista 3, ali tačno rješenje može biti i broj -3, jer je $(-3)^2 = -3 * -3 = 9$. Ovo zahtijeva dodatno pojašnjenje: mada ćemo u statistici često koristiti kvadriranje negativnih vrijednosti što će uvijek davati pozitivan rezultat, gotovo uvijek kada budemo vadili korijen podrazumijevaćemo da je rezultat apsolutna, nenegativna vrijednost broja. Taj pozitivni broj, koji je rezultat korijenovanja, nazivamo **glavna vrijednost korijena**.

Kad već spominjemo **apsolutne vrijednosti**, dobro bi bilo da se i njih podsjetimo s obzirom na to da se u statistici obilato koriste. Apsolutna vrijednost se matematički obično predstavlja kao veličina udaljenosti od tačke 0 na brojevnoj pravoj, pri čemu se zanemaruje smjer udaljenosti. Na dole prikazanoj slici se vidi da tačke A i B imaju jednaku apsolutnu vrijednost (3), baš kao i tačke C i D (0.5).

Kada notacijom iskazujemo da smo zainteresovani za apsolutnu vrijednost, onda taj broj ili izraz ogradimo sa dvije uspravne crte, npr. $|-5| = 5$ ili $|5 * 2 * -3| = 30$. U R-u se apsolutna vrijednost dobija jednostavno, komandom abs(x):

```
5*-2
## [1] -10
abs(5*-2)
## [1] 10
```

Retrospektiva bazične aritmetike ne bi bila potpuna bez prikaza redoslijeda operacija. Slijedićemo pravila koja su vas učili u školi. Prvo rješavamo računske operacije unutar zagrada.

```
redoslijed računskih operacija
```

Zatim dolaze na red operacije stepenovanja i korijenovanja. Treće u redu su operacije množenja i dijeljenja, pri čemu je konvencija da se rade prema redoslijedu s lijeva na desno. Na kraju dolaze sabiranje i oduzimanje, prateći isti princip s lijeva na desno.[6] Bitno je naglasiti da u R-u koristimo samo jednu vrsta zagrada, tzv. male zagrade, odnosno neće se koristiti srednje/uglaste i velike/vitičaste. Ako biste željeli riješiti sljedeći zadatak:

$$[4.70 + \frac{2}{5} \cdot \frac{3}{4} - 2^2 + (2.5 - \frac{3}{2}) \cdot 3] + 5 = ?$$

…u R-u biste napisali (obratite pažnju i na korištenje zagrada za definisanje razlomaka):

```
(4.7+(2/5)*(3/4)-2^2 + (2.5 - (3/2))*3) + 5
## [1] 9
```

Na samom kraju ovog brzog (je li baš bio brz?) pregleda, neophodno je istaći još jednu specifičnost R-a, ali i drugih statističkih softvera, a na koju vjerovatno niste obraćali pažnju ili se sa njom niste ni susretali. Radi se o prikazivanju izuzetno velikih i izuzetno malih brojeva koji će često izlijetati pred nas u ispisima rezultata. Zdrav razum nam kaže da je nepraktično pisati brojeve koje imaju desetak ili više cifara ispred decimalnog znaka (npr. 888 milijardi ili 888 000 000 000) ili brojeve koji su toliko mali da se tek nakon osam nula s desne strane decimalnog znaka pojavljuje neka druga brojka (npr. 1 milijarditi dio nečega ili 0.000000001). Tokom ranijeg školovanja ste učili da takve brojeve skraćeno pišete uz pomoć naučne notacije koja ima oblik $x \cdot 10^y$ za vrlo velike, odnosno $x \cdot 10^{-y}$ za minijaturne brojeve. Softveri - a i džepni kalkulatori - koriste blagu modifikaciju ovog sistema tzv. *naučnu **e** notaciju*.

naučna e notacija

U tom sistemu 888 milijardi pišemo kao 8.88e+11, a 1 "milijardnina" se piše 1e-9. Dakle, e+11 zamjenjuje $\cdot 10^{11}$, a e-9 zamjenjuje $\cdot 10^{-9}$. Broj koji stoji desno od **e+** nam govori za koliko mjesta treba da pomjerimo decimalni znak udesno, dok broj koji stoji desno od **e-** govori o pomjeranju decimalnog znaka ulijevo. Dakle, broj 500 bi se mogao napisati kao 5e+2, 515 kao 5.15e+2, a broj 0.05 kao 5e-2.

I to bi bilo dovoljno uvoda u statistiku i matematičkog osvježavanja. Vrijeme je da krenemo dalje. Sljedeće poglavlje je

[6] Za radoznale treba pomenuti da i ovdje postoje izvjesne dileme kada su u pitanju konvencije matematičke notacije koje se protežu kroz istoriju. Zainteresovani mogu da pročitaju diskusiju o jednostavnom aritmetičkom zadatku: 6 ÷ 2(1+2) = ? https://bit.ly/PEMDAS-problem

posvećeno upoznavanju najvažnijih pojmova koji čine gradivne elemente zdanja statistike. To su slova koja morate da naučite kako biste mogli uživati u ostatku ove lijepe debele knjige (izvinite ako sam subjektivan).

Poglavlje 2

Osnovni statistički pojmovi

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta zapravo podrazumijevamo pod mjerenjem i pratećim pojmovima (npr. operacionalizacija, varijabla, mjerni instrument)?
- Zašto je psihološko mjerenje specifično i koliko različitih nivoa mjerenja postoji?
- Šta su to istraživačke hipoteze?
- Šta označavamo pojmom populacija i na koji načine biramo uzorke za empirijska istraživanja?
- Šta su to parametar i statistik?
- Na koje tri oblasti se statistika obično dijeli?

Da ukratko sažmemo ono što smo naveli u uvodnom poglavlju: statistiku koristimo u psihologiji - baš kao i u drugim naukama - kako bismo na osnovu dostupnih podataka efikasno došli do što objektivnijih i preciznijih naučnih znanja o stvarnosti. Naučna znanja su pohranjena u jezičkoj formi i predstavljaju uvjerenja o opštim, apstrahovanim pravilnostima koje nadilaze pojedinačne slučajeve. Na primjer, iskaz "Tip koji je napisao ovaj tekst nas toliko maltretira statistikom da će to nanijeti teške posljedice na naše mentalno zdravlje." ne predstavlja nešto što je relevantno za nauku kao opšteljudsku potragu za znanjem. Nasuprot tome, sljedeća pretpostavka jeste naučno relevantna: "Ako neko kao dijete doživi veći broj ugrožavajućih iskustava - kao što su fizičko, seksualno ili psihičko nasilje nad njim ili bliskim osobama - ta osoba će imati veću sklonost ka problemima u mentalnom zdravlju u odrasloj dobi".

Dakle, zadatak statistike će biti da nam pomogne da donesemo sud o tome da li je ova jezička tvrdnja istinita, a na osnovu podataka koje smo adekvatno prikupili i obradili. Štaviše,

statistika će nam omogućiti da provjerimo **u kojoj mjeri** je ova pretpostavka tačna. Pretpostavka čiju istinitost ne znamo sa sigurnošću, ali je možemo empirijski provjeriti nazivamo **istraživačkom hipotezom** (engl. *research hypothesis*). Objedinjeni skup više hipoteziranih odnosa, a koji predstavlja pojednostavljenu sliku stvarnosti, nazivamo **modelom** (engl. *model*). Da bismo hipoteze i modele mogli provjeriti pomoću prikupljenih podataka potrebno je da sve relevantne termine iz uobičajenog jezika prevedemo na statistički jezik.

Jeste, baš kao i neki strani jezik, statistika predstavlja organizovan komunikacioni sistem sa vlastitim vokabularom. Nju su generacije - obično sijedih i bradatih - naučnika razvijale i dotjerivale. Vidjećete i sami da je taj vokabular obiman, ali savladiv. Početnike do panike naročito dovodi matematička notacija i glomazne formule pune mističnih oznaka kao što su ϕ, Ω, $\bar{x}$). Otom-potom, ta čudovišta ćemo pripitomljavati kada do njih dođemo, a ostatak ovog poglavlja ćemo posvetiti definicijama osnovnih pojmova nužnih da statistiku upregnemo u naučnu avanturu. Eh da, budući da engleski jezik danas predstavlja *de facto* jezik svekolike nauke, a ne samo psihologije, te da se u statističkim softverima mahom koristi engleski, uz svaki statistički termin će vam biti natovaren i njegov engleski naziv.

Šta su to podaci i kako ih dobijamo?

Počećemo od osnovnog elementa, takoreći atoma statističke realnosti. **Podatak** (engl. *datum*, mn. *data*) je pojedinačna vrijednost koja opisuje svojstvo nekog objekta / entiteta. U psihološkim istraživanjima najčešće prikupljamo podatke o pojedinačnim osobama (npr. inteligenciji *Jasne*, *Janka*, *Jadranka*), ali i o kolektivnim entitetima (npr. procenat žena koji studira na različitim *studijskim programima*, broj samoubistava u različitim *opštinama*), kao i neživim entitetima (npr. procjena rezonovanja različitih *AI modela*, karakteristike *riječi* koje upotrebljavamo kao draži: "riječ *statistika* je imenica i ima 10 slova"). Podatak može da opisuje entitet **kvalitativno** - njegovu kakvoću, odnosno "kakvo je nešto" (npr. Jasna je žensko; Jasna je plavokosa; Jasna studira psihologiju; Jasna ima vozačku dozvolu C kategorije) - ili **kvantitativno**, tj. količinski, odnosno "koliko nečeg" karakteriše taj objekat (npr. Jasna ima 20 godina; Jasna ima IQ 140; Jasnin prosjek ocjena u srednjoj školi je bio 4.75; Jasna ima 2 sestre; bilo je potrebno 2.3 sekunde da Jasna odgovori na postavljeno pitanje). Na ovom mjestu je bitno da zapazite da ćemo i po-

Nemojte da vas zbunjuje upotreba riječi entitet koju studenti obično asociraju sa 1. političko-pravnim uređenjem Bosne i Hercegovine i 2. bićima koja se javljaju tokom psihodeličnih iskustava. Pojam entitet označava zasebnu cjelinu - određenu nekim fizičkim ili pojmovnim granicama. Obično pod tim podrazumijevamo neku osobu ograničenu svojom kožom ili doživljajem ja-stva, ali to mogu biti i grupe ljudi (npr. muškarci ili žene, studenti psihologije ili medicine), brendovi ili organizacije za čiju privlačnost ili funkcionalnost smo zainteresovani, jezička pisma kao što su ćirilica i latinica čiju efikasnost želimo provjeriti, modeli vještačke inteligencije čije sposobnosti možemo uporediti, države za čije stanovnike procjenjujemo pravednost ili zadovoljstvo životom itd. Sitničaviji među vama će postaviti i pitanje šta je to zapravo informacija i kako se ona razlikuje od podatka. O definisanje informacije glavu lupaju mnogi filozofi, naučnici, informacioni stručnjaci. Definicija ima mnogo, a meni je najbliža funkcionalna koja kaže da je to elementarna saznajna čestica koja nastaje tumačenjem značenja podatka i kod primaoca mijenja predstavu o određenom sistemu. Takva definicija obuhvata i tačne i netačne informacije, kao i različite entitete koji informaciju prenose i/ili interpretiraju - od bioloških (npr. DNK, antitijela) do tehničkih (npr. podaci na hard-diskovima koje interpretiraju AI modeli).

datke kvalitativnog tipa moći statistički obrađivati. Naime, prebrojavanjem pojedinačnih slučajeva dobijamo kvantitativne vrijednosti koje nam govore o tome koliko nečeg imamo unutar skupa elemenata. Tako u jednom istraživanju u skupu od 35 ispitanika možemo imati 25 ženskih i 10 muških ispitanika ili 20 studenata psihologije i 15 studenata istorije. Ti podaci onda mogu biti iskorišteni za dalju statističku analizu, na primjer kako bismo utvrdili da li studij psihologije češće studiraju osobe ženskog pola, a istoriju muškog pola. Ali kako dobiti podatke za nešto tako apstraktno kao što je ekstraverzija?[1]

[1] Jedna popularna i prilično dobra definicija prezentovana i na engleskoj Wikipediji https://en.wikipedia.org/wiki/Extraversion_and_introversion kaže: "Ekstraverzija je opšta tendencija osobe da zadovoljstva primarno stiče u svijetu izvan svoje ličnosti."

Mehanizam dobijanja, ili tačnije rečeno, *stvaranja* podataka se naziva podužom, i u društvenom životu istraživača, vrlo popularnom riječju **operacionalizacija** (engl. *operationalization*). Za primjer operacionalizacije ćemo uzeti gore pomenutu hipotezu o vezi između izloženosti nasilju u djetinjstvu i lošijem mentalnom zdravlju u odrasloj dobi. I izloženost nasilju i kvalitet mentalnog zdravlja su apstraktni pojmovi sa izuzetno širokim obuhvatom. Razmislite kratko šta se sve može ubrojati u nasilje i šta sve računamo u mentalno zdravlje? Potrebno je konkretizovati te pojmove kako bismo mogli dobiti podatke o njima, a za to postoje različite opcije. Na primjer, o kvalitetu nečijeg mentalnog zdravlja bismo mogli da sudimo na osnovu dijagnostičke psihološke procjene odrasle osobe (npr. kada osobu procijeni psiholog koji će donijeti binarni sud da li osoba ima ili nema značajne probleme na planu mentalnog zdravlja). Druga mogućnost je da osoba ispuni neki poznati upitnik o mentalnom zdravlju (npr. u praksi popularni SCL-90-R[2] kojim se procjenjuje širok psihopatološki spektar), te da se na osnovu konačnog dobijenog rezultata procijeni ukupni stepen opterećenosti mentalnog zdravlja. O izloženosti nasilju u djetinjstvu možemo da sudimo na osnovu toga postoji li podatak iz službe socijalnog rada o tome da li je u nekom trenutku prijavljeno nasilje u porodici (prijavljeno ili nije prijavljeno). Međutim, podatke o količini izloženosti nasilju u porodici bismo mogli dobiti i pomoću samoizvještavajućeg upitnika koji sadrži relevantne tvrdnje kao što su "Da li se u Vašem djetinjstvu dešavalo da prisustvujete fizičkom nasilju među roditeljima/starateljima?" i "Da li vas je roditelj ili druga odrasla osoba iz domaćinstva često ili vrlo često psovala, vrijeđala, ponižavala ili posramljivala?". Na više takvih pitanja ispitanici mogu da odgovore sa *Da* ili *Ne* i da se na kraju sabere broj potvrdnih odgovora. Poenta je da se svakom ispitanik pripisuju dva podatka: jedan o stanju njegovog mentalnog zdravlja, drugi o izloženosti porodičnom nasilju u djetinjstvu. Time smo te teorijske pojmove preveli u empirijske podatke, što

[2] Symptom Checklist-90-Revised ili Revidirani kontrolni spisak simptoma koji od ispitanika traži da navede da li ga je u posljednje vrijeme mučilo nešto od 90 simptoma problema sa mentalnim zdravljem (npr. glavobolje, nesanica, osjećanje krivice, problemi sa koncentracijom)

nazivamo operacionalizacijom. Što se mene tiče, bolji termini bi bili konkretizacija, specifikacija ili radno definisanje pojmova.

Već smo u uvodu naveli da se konkretne vrijednosti dodjeljuju objektima ili na osnovu prebrojavanja ili na osnovu mjerenja. Dok prebrojavanje nije nužno dodatno definisati (ali ako baš hoćete, vrijednost koja se dobija prebrojavanjem predstavlja ukupan broj elemenata neke definisane vrste), **mjerenje** (engl. *measurement*) je već pojam koja ima nekoliko formalnih definicija. Neke od njih imaju istu količinu zajedničkog koliko zajedničke "pticosti" imaju pingvini i orlovi ili "psetosti" čivave i afganistanski hrtovi. Ovdje ću predstaviti dvije najprisutnije i najoprečnije verzije.

Klasična definicija mjerenja

Prema **klasičnoj definiciji mjerenja** - odnosno prema onome kako ste navikli da nešto mjerite prije nego ste upisali studij psihologije koristeći vagu, štopericu ili krojački metar - mjerenje je postupak pripisivanja broja objektu mjerenja na osnovu toga koliko nekog atributa/svojstva ima taj objekat u odnosu na definisanu jedinicu mjerenja. Tako kada stanete na vagu da izmjerite svoju tjelesnu masu, vi zapravo poredite koliko puta ste teži od jednog kilograma. Kilogram je univerzalno usvojena jedinica (mada ima nacija koje se uporno opiru njenom korištenju, kao što je SAD) koju su osmislili i koristili ljudi različitih naroda, prvo kao masu litre vode, zatim kao teg od kovine koji je kao etalon čuvan u Francuskom arhivu, dok je tek 2019. godine definisan putem Plankove konstante, odnosno fizičkog pojma, a ne primjerka direktno izvedenog iz fizičke realnosti.

Psihološka definicija mjerenja

Možete naslutiti da u psihologiji nije lako dogovoriti mjernu jedinicu koju bi najveći dio čovječanstva prihvatio. Iz tog razloga, a u svrhu odbrane statusa psihologije kao discipline dostojne riječi nauka, već pominjani psihofizičar S.S. Stevens je napravio "pakleni plan" i osmislio dosta relaksiraniju definiciju koja je nastanila i druge naučne oblasti. Najčešće se naziva **operacionalnom definicijom mjerenja**. Tako se naziva jer je njen glavni smisao da pojam operacionalizujemo, međutim možemo je zvati i definicijom numerizacije ili čak psihološkim mjerenjem. Ta definicija glasi: "Mjerenje je postupak pripisivanja brojke / cifre / znamenke objektu mjerenja ili događaju prema nekim pravilima".[3] Ono na šta bi trebalo da posebno obratite pažnju jeste to da definicija ne pominje brojeve (engl. *numbers*) nego brojke (engl. *numerals*). Brojke jesu simboli koje se koriste u brojevnom sistemu da iskažu količine i vrlo precizirane odnose među njima, ali one mogu da posluže i za druge svrhe. Naime, iz definicije slijedi da ćemo mjerenjem nazivati ne samo neko mjerenje vremena reakcije u nekom istraživanju psihologije opažanja (npr. za reakciju

[3] Stevens, 1946, https://doi.org/10.1126/science.103.2684.677

nakon prezentovanja draži je bilo potrebno 350 ms), nego i postupak kojim ćemo svakom muškom ispitaniku u istraživanju dodijeliti vrijednost 1, a svakom ženskom vrijednost 2. Navedena definicija mjerenja se može i dodatno razvući, pa budući da se u matematici uz brojke koriste i slovne oznake, mjerenjem bismo moglo nazvati i dodjeljivanje simbola (slova ili riječi) objektima mjerenja. Tako muški ispitanici mogu dobiti oznaku *m*, a ženski *ž*. Sve u svemu, dovoljno je da imamo neko pravilo koje smo definisali i kojeg se dosljedno pridržavamo, a pomoću kojeg karakteristike objekata u stvarnosti prenosimo u simbolički sistem radi donošenja zaključaka o pojavama od interesa.

Ako ste i vi na ovako popustljivu definiciju podigli obrve, niste jedini. Postoji mnogo kritičara koji smatraju Stevensovu definiciju u najbolju ruku nečim što se može nazvati **procjenjivanje** (engl. *assessment*), a da riječ mjerenje treba sačuvati za situacije kada imamo dogovorom utvrđenu jedinicu. Neki autori[4] idu toliko daleko da ovaj pristup smatraju prosto postupkom stvaranja podataka, što zapravo i jeste operacionalizacija. Taj postupak stvaranja podataka može biti manje ili više dobar ovisno o tome koliko je eksplicitno objašnjeno i koliko je logički dosljedno pravilo prema kojem se odnosi u stvarnosti povezuju sa odnosima u simboličkom sistemu. S druge strane,operacionalna definisano mjerenje je pragmatičan biznis i samo zahvaljujući njemu dobijamo podatke neophodne za provjeru naučnih hipoteza i modela. Ipak, veća pragmatičnost istovremeno znači i veću mogućnost grešaka, pa se zato ocjenjivanjem kvaliteta stvaranja podataka bavi čitava naučna oblast **psihometrija** (engl. *psychometrics*). Više iz konvencije nego iz ubjeđenja, u ostatku teksta kad god pominjem mjerenje i ostale izvedenice pod tim podrazumijevam operacionalnu definiciju. Da ne mrsimo dalje sa ovim teškim pitanjima, idemo na definicije drugih bitnih pojmova koji u sebi sadrže odrednicu mjerenja.

[4] …sa kojima sam saglasan, npr. Uher, J. (2018). https://doi.org/10.3389/fpsyg.2018.02599

Šta bi još trebalo da znamo kada je u pitanju mjerenje?

Predmet mjerenja (engl. *measured attribute* / *property* / *phenomenon*) je ono što želimo da mjerimo, odnosno teorijski definisano svojstvo / obilježje / atribut / pojava, čiji vid ili količinu izraženosti želimo da procijenimo. Dakle, pod predmetom mjerenja ne podrazumijevamo objekat na kojem se vrši mjerenje. Odnosno, predmet mjerenja mogu biti visina, depresivnost, znanje iz statistike, izloženost porodičnom nasilju,

ali predmet mjerenja nisu Jasna, Jasmina ili Jana. Druga bitna napomena je da kada nešto mjerimo pokušavamo da izdvojimo i mjerimo samo jedno svojstvo. Na primjer, kada želimo da dobijemo podatak o nečijoj visini ne želimo da taj podatak ovisi o tjelesnoj masi te osobe (tj. dvije osobe visoke 180cm bi trebalo da imaju isti podatak o visini, bez obzira na to što jedna ima 80kg, a druga 65kg). Isto tako kada mjerimo inteligenciju, ne želimo da podatak koji dobijemo sadrži u sebi i depresivnost ili ekstraverziju.

Na ovom mjestu se može skrenuti pažnja na to da su psihički predmeti mjerenja, odnosno psihička svojstva od interesa načelno difuznija od biofizičkih. Psihološki pojmovi kao što su ekstraverzija, inteligencija i sposobnost pamćenja, pa i obrazovni pojmovi kao što je znanje iz statistike, su složeni jer podrazumijevaju raznolike manifestacije. Takve uopštene složene pojmove, koji nisu direktno opazivi, nego o kojima sudimo na osnovu nekih pokazatelja nazivamo konstruktima (engl. constructs)

Mjerni instrument (*engl. measuring instrument*) predstavlja sredstvo pomoću kojeg stvaramo podatke. To bi za mjerenje tjelesne mase bila vaga. Međutim, šta bi bio mjerni instrument kada je u pitanju pol? Ne u svim, ali u najvećem broju slučajeva radi se o samoprocjeni ispitanika u nekoj anketi ili formularu, gdje osim muškog i ženskog pola mogu postojati i nebinarne opcije ili opcija da osoba odbija da navede kojeg je pola. Nekada čak ni ne koristimo instrument u užem smislu riječi, nego se kao istraživači oslanjamo na neposredno opažanje (npr. vidimo nečiji pol, ili vidimo da osoba gricka nokte dok očekuje da bude prozvana na ispitu što bi mogao biti pokazatelj anksioznosti). To nije slučaj samo u psihologiji, nego se koristi i u terenskim istraživanjima u biologiji, antropologiji, sociologiji, odnosno gdje god čulima dostupno ponašanje predstavlja predmet mjerenja. Sve u svemu, do podataka u psihološkim istraživanjima je, osim neposrednim opažanjem, moguće doći putem različitih sredstava: upitnika (npr. inventari ličnosti, upitnici o stavovima), testova postignuća (npr. test inteligencije, testovi znanja), projektivnih testova (npr. Rorschahove mrlje, dopunjavanje nedovršenih rečenica), individualnih i grupnih razgovora (npr. klinički intervju, diskusione fokus grupe), analize sadržaja (npr. analiza internet portala, dnevnika koje osoba vodi), bioloških i fizičkih mjerenja (npr. hormoni iz uzoraka krvi, mjerenje brzina reakcije), arhivisanih podataka (npr. svjedočanstva, ponašanja na mobilnom telefonu ili društvenim mrežama)…Koji god postupak da koristimo, on mora dovoljno valjano generisati podatke za pojam za koji smo zainteresovani u istraživanju.

Psihološki konstrukti su kompleksni fenomeni koji se vrlo rijetko mogu kvalitetno mjeriti "odjednom" ("bacanjem" jednog pogleda ili pojedinačnim zadatkom). Zato se u svrhu njihovog mjerenja najčešće koriste kompozitni / sastavni mjerni instrumenti. **Kompozitni mjerni instrumenti** (engl. *composite instruments*) su instrumenti sastavljeni iz više dijelova, odnosno kombinovanjem više pojedinačnih elemenata. Vi ste sa kompozitnim mjernim instrumentima odavno upoznati, jer je to bio

kompozitni instrumenti

svaki pismeni kontrolni rad sastavljen iz više zadataka. Vaše znanje (predmet mjerenja) je bilo mjereno putem ukupnog broja bodova.

Pojedinačni element mjernog instrumenta nazivamo **stavka**, **čestica** ili **ajtem** (od engl. *item*), što u testovima znanja ili u testovima inteligencije normalnim jezikom zovemo zadatak. Na primjer, neki inventar ličnosti u kojem ispitanici sami procjenjuju izraženost različitih crta može biti sastavljen od ukupno 50 stavki, od kojih 10 procjenjuje crtu savjesnosti koja bi bila predmet mjerenja, 10 ekstraverziju itd. Svrha svake stavke je da doprinese boljem mjerenju obilježja koje je u fokusu, a psihometrija je iznašla posebne procedure kojima se procjenjuje valjanost instrumenta i u ovom pogledu.

stavke

Bez obzira na to da li koristimo kompozitne ili mjerne instrumente koji se sastoje od samo jedne stavke (npr. "Koliko ste trenutno motivisani da nastavite čitati ovaj tekst na skali od 1 do 5?"), cilj svakog mjerenja jeste dobiti neku **mjeru** ili **skor** (engl. *score*). Mjera predstavlja vrijednost na mjernoj skali koja bi trebalo da ukaže na kakvoću ili količinu izraženosti obilježja. Mjere mogu biti vezane za pojedinačne ispitanike (npr. Jasna ima IQ 140) ili grupe ispitanika (npr. prosječna inteligencija studenata psihologije je 120).

Još jedan bitan pojam je **mjerna skala** (engl. *scale*). To je prostor definisan teoretski mogućim minimum i maksimumom unutar kojeg se putem mjerne procedure pozicioniraju ispitanici. Na primjer, u osnovnoj ili srednjoj školi na našem području ste dobijali ocjene koje nisu mogle biti niže od 1 ili više od 5. Svaki učenik na kraju godine iz određenog predmeta morao biti raspoređen na jedan od podioka untar te skale. U stručnoj literaturi se procedura određivanja pozicije ispitanika na mjernoj skali nekad naziva **skaliranje** (engl. *scaling*) - a to je opet ono isto što i operacionalno mjerenje.

Ako svi objekti istraživanja imaju jednaku vrijednost (npr. svi ispitanici imaju 19 godina ili - što je relativno često - da su svi ispitanici studenti psihologije 1. godine studija) onda imamo posla sa **konstantom**. Suprotnost konstanti je **varijabla** / **promjenjiva** (engl. *variable*). Varijablom zovemo svaki atribut / svojstvo za koji se dobijaju različite vrijednosti unutar skupa od interesa. Skup od interesa u kojem vrijednosti variraju se može odnositi na svojstva ispitanika (gdje na primjer razlikujemo ispitanike po postignutim mjerama na testu inteligencije), ali to mogu biti i svojstva draži (npr. neke riječi koje prikazujemo na ekranu u jednom eksperimentu su ispisane latinicom, a druge ćirilicom)

ili različite kontekstualne varijable (npr. u istraživanju efekta vremenskih prilika na depresivnost možemo razlikovati dane sa južnim vjetrom i dane sa sjevernim vjetrom). Kao istraživače će nas interesovati da nešto saznamo o svijetu na osnovu raspodjela pojedinačnih varijabli (npr. koliko studenata ima snažan strah od statistike ili kako su ti strahovi raspoređeni po intenzitetu od ležernosti do stanja panike) kao i na osnovu zajedničkog variranja dvije ili više varijabli (npr. postoji li veza između straha od statistike sa uspjehom u matematici u srednjoj školi).

Kakve sve vrste varijabli imamo?

Varijabli ima ovakvih i onakvih ("varijable su varijabilne") i potrebno je predstaviti različite podjele koje postoje. Prva od njih je istovremeno i bitna i problematična: ona ih dijeli na teorijske i operacionalne varijable. Teorijske ili konceptualne varijable predstavljaju pojmovno definisane osobine po kojima se članovi skupa razlikuju. Operacionalne varijable su skupovi podataka dobijeni preciziranim postupcima, a u svrhu empirijskog izražavanja teorijskih varijabli. Kao primjer možemo uzeti crtu ličnosti savjesnost. Ona se teorijski može definisati kao "spektar konstrukata koji opisuje individualne razlike u sklonosti da se bude samokontrolisan, odgovoran prema drugima, marljiv, uredan i sklon poštovanju pravila". S druge strane, primjer operacionalne definicije savjesnosti bi bio "prosječni skor koji osoba dobija tako što na ponuđenoj petostepenoj skali slaganja odgovara na 9 tvrdnji skale savjesnosti Big Five Inventory upitnika".

Ovdje treba imati na umu da istim imenom imenujemo tri različite stvari koje su različite: konstrukt $\neq$ teorijska varijabla $\neq$ operacionalna varijabla. Na primjer, pojmom visina označavamo 1) pojam razdaljine između najniže i najviše tačke tijela, 2) skup različitih vrijednosti tjelesne visine koju ispitanici imaju u idealnim uslovima (kada su bosi, oko podneva, uspravljenih leđa), 3) skup različitih vrijednosti koji se dobija konkretnim mjerenjem. To za visinu i možda nije neki presudan problem, ali može biti za inteligenciju, savjesnost, izloženost nasilju. Neki autori[5] primjećuju da time što i jednu i drugu i treću stvar nazivamo istim imenom otupljujemo svoju sposobnost rasuđivanja i zaboravljamo da su operacionalne varijable samo nesavršeni predstavnici teorijskih pojmova. Zato treba da imate na umu da u daljem tekstu, pod pojmom varijabla obično podrazumijevamo operacionalne varijable, tj. skupove podatke koji se nalaze unutar nekog tabelarnog softvera.

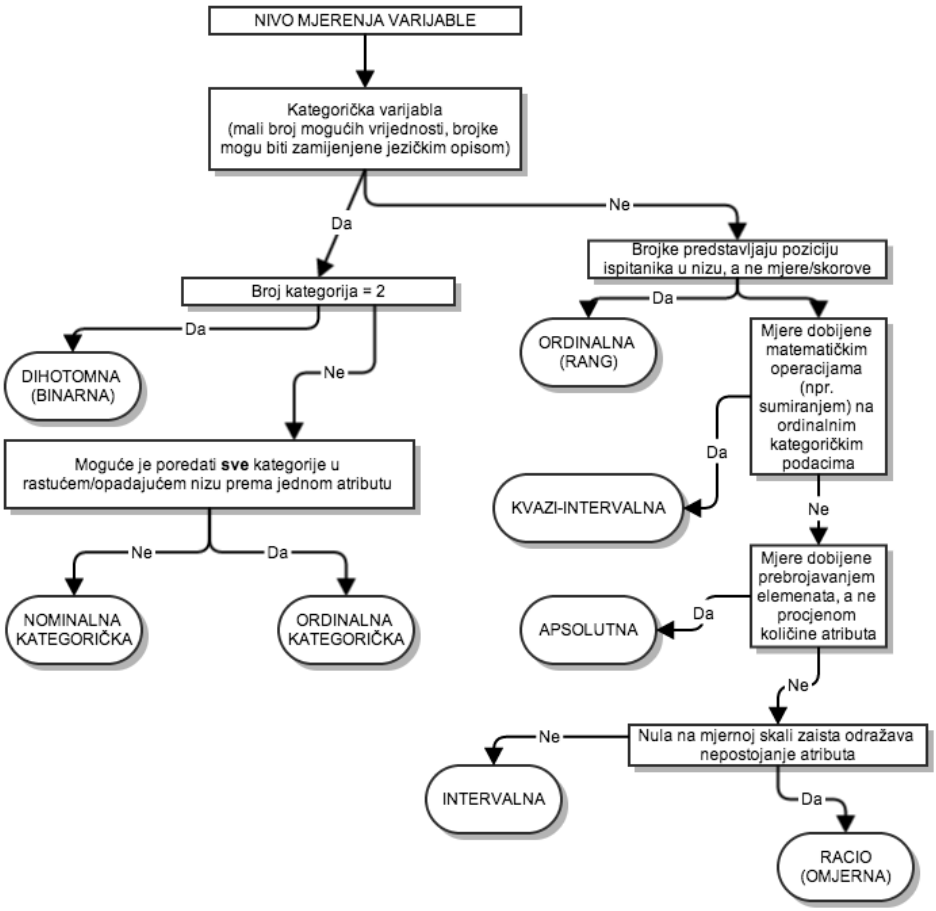
[5] npr. Uher, J. (2023). https://doi.org/10.1111/spc3.12740

Druga podjela varijabli se obično naziva podjelom prema nivou mjerenja varijabli (engl. *levels of measurement*). Ova podjela je ekstremno bitna iz dva razloga. Kao prvo, odabir statističkih procedura za analizu podataka će u velikoj mjeri zavisiti od toga u koju smo grupu svrstali neku varijablu. Drugo, svijest o tome koji nivo mjerenja je korišten u istraživanju će nam biti značajan u interpretaciji dobijenih rezultata nakon što smo izveli statističku obradu. Slika 2.1 prikazuje stablo odluke kojim možete da se pripomognete kako biste odredili status neke varijable. Primijetićete da se na vrhu nalazi glavna podjela na kategoričke (lijevo) i dimenzione varijable, te zato ovdje još jedna napomena. Naime, u literaturi se dimenzione varijable često nes(p)retno imenuju kao numeričke (srešćete i ezoteričnije nazive kao što su metričke ili skalarne varijable). Nes(p)retno, jer i kategoričke varijable mogu imati numeričke vrijednosti. Ono što je zapravo zajedničko za dimenzione varijable jeste da sve one koriste uobičajeni brojevni sistem za označavanje dimenzije duž koje se vrši skaliranje.

Kategoričke varijable

Kategorička varijabla (engl. *categorical variable*) je varijabla čije vrijednosti predstavljaju kvalitativno različita ispoljavanja nekog svojstva. Da ponovimo, kvalitet(a) se tiče toga kakvog je svojstva nešto, dok se kvantitet tiče količine svojstva. Pojedinačna klasa koju karakteriše ista ispoljenost svojstva (npr. *psihologija* za varijablu *studijski program*) se naziva kategorija, nivo ili valenca kategoričke varijable. Razlikovaćemo tri vida kategoričkih varijabli:

Nominalne varijable su varijable gdje se između različitih oblika ispoljavanja svojstva ne može uspostaviti potpuna hijerarhijska struktura veće ili manje ispoljenosti svojstva. Evo nekoliko primjera: studijski program kojeg student-ispitanik pohađa (kategorije: psihologija, elektrotehnika, biljna njega, medicina), oblik dijagnostikovane psihičke smetnje (depresija, fobija, bipolarni poremećaj, ovisnost), odjeljenje koje učenik neke osnovne škole pohađa (5a, 5b, 5c ili 5d), terapija koja ispitanici pohađaju (kognitivno-bihejvioralna, transakciona analiza, psihoanaliza, kontrolna grupa). Naziv nominalna dolazi od latinskog *nomen* što znači ime, budući da vrijednosti - bilo da su one brojčane ili drugačije - služe samo da označe imena različitih kategorija. Ta imena mogu biti potpuno proizvoljna, odnosno ona zavise od odluka istraživača i pragmatičnih karakteristika. Na primjer u svim navedenim slučajevima kategorije varijable mogu biti označene ili punim nazivima (npr. *psihologija*, *elektrotehnika*) ili skraćenicama (npr. *psi*, *elt*) ili brojkama (*1*, *2*…). Brojke se vrlo



Slika 2.1: Stablo odluke za određivanje nivoa mjerenja varijable.

često koriste jer se lakše unose u softverske baze i postoji manja šansa za nekompatibilnost unesenih podataka sa zahtjevima statističkih programa. Ako se odlučimo da koristimo brojke kao imena to što je elektrotehnika dobila brojku *2*, a psihologija *1* ili da je fobija = 1, a depresija = 2, to ne znači da je po nekom svojstvu elektrotehnika "viša" od psihologije, niti da je fobija problematičnije stanje od depresije. U pogledu odnosa koji postoji među objektima istraživanja imamo dvije mogućnosti: ili su po datom svojstvu jednaki (=) i pripadaju istoj kategoriji (npr. dva ispitanika studiraju isti studiji ili su obje riječi imenice) ili su nejednaki ($\neq$) i svrstavamo ih u različite kategorije (npr. jedan ispitanik studira psihologiju, a drugi elektrotehniku ili je jedna riječ imenica, a druga glagol).

nominalne (= ili $\neq$)

Nešto složenije od nominalnih varijabli su **ordinalne kategoričke varijable** gdje se između svih postojećih kategorija može ustanoviti stalni rast ili pad izraženosti svojstva koje definiše varijablu. Ne čudi to što na latinskom *orden* znači red ili poredak. Neki klasični primjeri ordinalnih kategoričkih varijablu su: stepen stručne spreme gdje osobe sa višom stručnom spremom imaju duži niz godina uspješno provedenih u formalnom obrazovanju (npr. osnovna škola, srednja škola, visoka škola), pojedinačna ocjena koja se dobija na provjeri znanja u osnovnim i srednjim školama (od 1 do 5 koje zapravo predstavljaju oznake za ocjene od "nedovoljan" do "odličan"), studijska godina na fakultetu (od 1. do 5. za studij psihologije) ili stepen depresivnosti (npr. nedepresivni, blago depresivni, umjereno depresivni, teško depresivni). Dakle, dva objekata možemo prema procjenjivanoj osobini ocijeniti kao jednaka (=) ili različita, ali ćemo za razliku od nominalnih varijabli ovdje pridodati podatak koji ima i značenje veći ili manji (> ili <). Ako je osoba dobila ocjenu odličan, ona je trebalo da pokaže više znanje nego osoba koja je dobila ocjenu vrlo dobar. Međutim, primijetite i ovdje da osoba koja je dobila ocjenu vrlo dobar (4), nije morala nužno da pokaže da posjeduje duplo veće znanje nego osoba koja je dobila ocjenu dovoljan (2). Kod ordinalnih kategoričkih varijabli veće brojke obično odražavaju veću izraženost svojstva, ali i ovdje one zapravo služe kao oznaka kategorije, a ne nešto što odražava odnose koji su zacrtani u sistemu realnih brojeva. Na primjer, u njemačkom obrazovnom sistemu je najviša ocjena 1, a ocjene 5 (u visokom obrazovanju) i 6 (u osnovnom i srednjem obrazovanju) predstavljaju ono što nikako ne biste voljeli da dobijete. Uprkos obrnutom značenju brojčanih vrijednosti to je i dalje ordinalna kategorička varijabla.

ordinalne kategoričke (>, < ili =)

Poseban oblik kategoričkih varijabli su **dihotomne ili binarne**

varijable (od grčkog *dikhótomos*, „presječen na pola, na 2 dijela", odnosno latinskog *binarius*, „sastavljen iz 2 dijela") koje imaju samo dvije moguće vrijednosti. Na primjer, status živog kada on ima kategorije živ ili mrtav, pitanju tačnost odgovora kada se ocjenjuje kao tačan ili netačan, pol kada su predviđene kategorije muški ili ženski, odgovor na pitanje "Plašite li se statistike?" ako su mogući samo odgovori Da ili Ne. Ono što omogućava svrstavanje podataka u jednu ili drugu grupu, odnosno kategoriju, jeste ponovo funkcija ekvivalentnosti. Na primjeru tačnih zadataka iz nekog testa mudrolija je sljedeća: svi tačni odgovori se grupišu kao tačni (i dobijaju npr. oznaku 1), a svi netačni su netačni (i dobijaju oznaku 0). Na primjeru pola, svi ispitanici koji odgovore da su muški imaju jednako obilježje (npr. 1 ili m), kao i sve ispitanice koje se smatraju ženskim (npr. 2 ili 0 ili ž).[6] Iako se u nekim klasifikacijama dihotomne varijable svrstavaju među nominalne postoje bitne razlike između njih zbog čega je potrebno identifikovati ih kao zasebnu vrstu. Naime, em se dihotomne varijable često analiziraju na drugačiji način od nominalnih varijabli što će biti prikazano u sljedećim poglavljima, em teorijski gledano u određenim slučajevima postoji i presudna kvantitativna razlika između dvije vrijednosti koja odražava prisustvo ili odsustvo nekog svojstva (npr. živ je življi nego mrtav, a tačan zadatak ukazuje na više znanja od netačnog).

[6] Iako se obično definiše i tretira kao dihotomna, pol može biti i nominalna varijabla ako se u istraživanju omogućavaju i nebinarne opcije ili ako se dopusti da ispitanik odbije da odgovori na pitanje.

Dimenzione varijable

Kako sam već naglasio **dimenzione varijable** (engl. *dimensional variable*) su one koje na nekoj brojevnoj dimenziji predstavljaju razlike u količini određenog svojstva. Ovdje ću prikazati podjelu na pet statistički relevantnih nivoa mjerenja dimenzione varijabli, mada neke podjele poznaju i manje i više njih.

Ordinalne rang-varijable (engl. *rank variables*) su one varijable gdje brojke označavaju poziciju u opadajućem ili rastućem nizu. Jedan tipičan, a nepsihološki, primjer bi bio rangiranje trkača u maratonu kada u obzir ne uzimamo postignuto vrijeme nego poziciju kada su trkači stigli na cilj u odnosu na čitavu grupu - a na osnovu njihove brzine tokom trke (ako smo imali 342 trkača: 1., 2., … 23., … 342.). Psihološkom kontekstu su bliže varijable kao što su rang kandidata nakon kvalifikacionog ispita, popularnost učenika u svom odjeljenju, rang univerziteta na nekoj od svjetskih lista. Iako ih na prvi pogled nema tako mnogo u psihologiji, zanimljivo je da se veliki broj statističkih tehnika zasniva na ordinalnim rang podacima i da se one upotrebljavaju čak i ako smo dobili podatke višeg nivoa mjerenja ("Zašto je to

tako saznaćete u nekoj sljedećoj epizodi, ostanite sa nama!").

Kao i kod ordinalnih kategoričkih varijabli osnovni matematički odnos između dva objekta koji se iz stvarnosti prenosi u sistem podataka jeste odnos manje (<) ili više (>). Međutim, za razliku od kategoričkih ordinalnih varijabli gdje odnos jednakosti (=) takođe spada u osnovne odnose - budući da on omogućava da se više objekata svrsta u istu kategoriju (npr. svi studenti koji dobiju ocjenu 6 na ispitu iz statistike) - on bi među ordinalnim rang-varijabla trebalo da rijetka i nepoželjna pojava. Takozvana vezanost rangova (engl. *rank tie*) predstavlja problem i za organizatore sportskog takmičenja koji treba da odluče šta da rade sa dvojicom takmičara koji su i isto vrijeme došli iza pobjednika utrke. Da li da im obojici dodijele srebro (odnosno drugo mjesto) ili bronzu (treće mjesto) ili da im naprave neku medalju od legure srebra i bronze (i proglase da su zauzeli drugo-i-po mjesto)?! Kako se ova poteškoća rješava biće riječi u jednom kasnijem poglavlju.

Sljedeće u nizu su **kvazi-intervalne varijable** (engl. *quasi-interval variables*) kod kojih postoji rastući / opadajući niz numeričkih vrijednosti, a skor se dobija matematičkim operacijama na skupu kategoričkih varijabli. Iz navedenog slijedi da su kvazi-intervalne varijable izvedene, kompozitne varijable. Najpoznatiji primjer je prosječna školska ocjena koja se dobija izračunavanjem aritmetičke sredine na ordinalnim kategoričkim varijablama. Jeste, unutar svakog predmeta ste dobijali pojedinačne ocjene koje predstavljaju znanje od nedovoljnog do odličnog, a brojčana ocjena tu služi samo kako bi se ta ocjena mogla kombinovati sa drugima i da se na kraju dobije mjera prosjeka. Ista stvar se dešava i sa procjenom ekstraverzije ako se skor za ekstraverziju dobije izračunavanjem aritmetičke sredina odgovora ispitanika na 10 različitih stavki koje za predmet mjerenja imaju ekstraverziju (npr. "Ja sam pričljiva osoba." i "Lako sklapam prijateljstva.", a na koje ispitanik odgovara u zavisnosti od stepena saglasnosti sa tvrdnjama od "nimalo" = 1 do "potpuno" = 5). Isti je slučaj i sa procjenom znanja iz statistike kada budete radili pismeni ispit gdje ćete skor dobiti na osnovu sume tačnih odgovora na ukupno 50 zadataka. Zašto se ovaj nivo mjerenja zove kvazi-intervalno? Zato što na mjernoj skali tih varijabli zaista postoje jednaki brojevni intervali (npr. razlika između 2.5 i 3.5 je jednaka kao i razlika između 3.5 i 4.5), međutim oni ne odražavaju nužno jednake razlike u posjedovanju svojstva koje se mjeri - odnosno predmeta mjerenja. Ako je osoba na skali zadovoljstva životom imala skor 3.5, to ne znači da je u istoj mjeri zadovoljnija od osobe koje ima skor 2.5 i manje zadovoljna od osobe koja ima

skor 4.5. Drugim riječima, brojčani sistem koji se koristi u takvim slučajevima je prigodni alat i za mnoge svrhe "radi posao". Ipak, pri interpretaciji podataka bismo uvijek trebali imati na umu ovu činjenicu.

Prave intervalne varijable (engl. *interval variables*) ne samo da imaju opadajući / rastući niz koji odgovara brojevnom sistemu, nego između jedinica mjerne skale postoji jednaka udaljenost koja odgovara promjeni ispoljenosti atributa. Pritom, minimalna vrijednost mjerne skale - tzv. **relativna nula** - ne označava nepostojanje atributa. Najčešći primjer intervalne varijable koji se u literaturi spominje je temperatura kada se ona izražava na Celzijusovoj ili Farenheitovoj skali. Na tim mjernim skalama razlike između podioka su jednake i odražavaju linearni (pravolinijski), ravnomjerni odnos u promjeni temperature koji postoji u fizičkim uslovima. Međutim, nula na ovim skalama ne znači da temperature - odnosno, kretanja molekula - nema. Nule su, suštinski govoreći, proizvoljno postavljene: na Celzijusovoj skali je to tačka kada se voda pretvara u led (zapravo je Celsius za života postavio inverznu skalu gdje je 0 bila tačka ključanja, a 100 mržnjenja vode), a na Farenhajtovoj skali je to $32^{\circ}F$.[7] Iz tog razloga besmisleno je tvrditi da se kaže da kada je napolju $20^{\circ}C$ taj dan dva puta topliji od dana kada je napolju $10^{\circ}C$. Dalje slijedi da matematičke operacije dijeljenja i množenja nemaju smisla na intervalnim varijablama, ali da ima smisla ih sabirati i oduzimati. Odnosno, kada se objekti međusobno porede možemo reći za koliko je nešto više ili manje izraženo, ali ne i koliko puta. Prave intervalne varijabli su zapravo rijetke u psihologiji, a veliki broj psihologa brka intervalne i kvazi-intervalne varijable iako se one suštinski razlikuju. Iako se identične matematičke procedure često primjenjuju na obje vrste varijabli, tumačenje rezultata mora biti različito jer pretpostavka o tome da mjerna skala linearno odražava razlike u izraženosti svojstva u najvećem broju slučajeva nije niti vjerovatna, niti dokazana.

[7] Prema jednoj verziji https://www.newworldencyclopedia.org/entry/Fahrenheit najniža vanjska temperatura koju je g. Fahrenheit mogao da izmjeri u prirodi je odredio kao 0, a za temperaturu 100 je odredio temperaturu vlastitog tijela?

Na najvišem mjernom nivou se nalaze **racio ili omjerne varijable** (engl. *ratio variables*). Karakterišu ih jednaki intervali među podiocima koji odražavaju linearnu promjenu atributa, ali i postojanje **apsolutne nule**: minimalne vrijednosti na mjernoj skale koja označava nepostojanje atributa. Za razliku od Celzijusove i Farenhajtove skale, Kelvinova nula ($-273,15^{\circ}C$) odražava teorijsko stanje na kojem nema toplote, odnosno „mrdanja" molekula. Bitno je da ona postoji, makar se znalo da u istraživanju ne može da se dosegne 0 (npr. kada se ispituje vrijeme reakcije na draž mjereno u milisekundama znamo da vrijednost za svakog ispitanika mora biti veća od 0, a znamo i

da veličina draži po jednoj dimenziji mora biti veća od 0 cm). U psihologiji se racio varijable skoro uvijek odnose na biofizička svojstva, ali pominjani S.S. Stevens je zagovarao i da se neka psihofizička svojstva mogu mjeriti na omjernoj skali (npr. skala glasnoće u son jedinicima). Što se tiče osnovnih matematičkih operacija sve su načelno dozvoljene, a izraženost svojstva među objektima je opravdano direktno upoređivati kako bi se dobili omjeri vrijednosti (otuda naziv *racio* ili *omjerni* nivo mjerenja), te je smisleno reći da je neko duplo brže reagovao ili da neka draž ima duplo veću masu.

Konačno imamo slučaj **apsolutnih ili zbirnih varijabli** (engl. *count*, *absolute variables*). Radi se o varijablama koje se dobijaju prebrojavanjem elemenata od interesa. To znači da je i nula apsolutna (gle čuda!?), pa je neki autori smatraju samo posebnim oblikom racio varijabli. Međutim, neke očite razlike postoje, a bitno ih je ilustrovati budući da postoje posebne statističke tehnike namijenjene za analizu apsolutnih varijabli. Ako za primjer racio mjere možemo uzeti 1.5kg jabuka, kao primjer apsolutne mjere bi to bilo 6, 7 ili 8 jabuka, ovisno o tome kolike su te jabuke. Drugim riječima na racio nivou predstavljamo "koliko nekog svojstva" ima, dok na apsolutnom nivou govorimo o tome "koliko komada" nečega postoji. U psihologiji su ove zbirne varijable relativno česte npr. osoba je 6 puta dotakla vrat tokom razgovora sa nepoznatom ili u tekstu se spominje 17 riječi koje označavaju strah. Bitno je naglasiti da često postoji nejasnoća oko toga da li je broj tačnih zadataka na nekom testu bolje tretirati kao apsolutnu ili kvazi-intervalnu varijablu. Kvazi-intervalni nivo je mnogo razumnija opcija. Svaki zadatak testa sam po sebi predstavlja zasebnu provjeru znanja ili sposobnosti, a ukupna suma predstavlja ukupnu procjenu specifičnog znanja ili sposobnosti (odnosno, procjenjuje se količina atributa!). Štaviše, postoje neke psihometrijske procedure koje nastoje da takav sumacioni skor posebnim transformacijama unaprijede u nešto što je blisko intervalnom nivou mjerenja.

Kakve sve još vrste varijabli imamo?

Avaj, još nismo završili sa podjelama varijabli! Sljedeća među njima je podjela numeričkih varijabli na prekidne (diskretne, tačkaste, engl. *discrete*) i neprekidne (kontinuirane, neprestane, engl. *continuous*). Ova podjela se ponovo dotiče podjele količina na mnoštvo (broj elemenata, množina, engl. *multitude*) i veličinu (količina svojstva, engl. *magnitude*). Diskretne varijable su

one koje su načelno prebrojive i imaju jasno odvojene tačke na mjernoj skali na kojoj se skaliraju objekti istraživanja. Dakle, uz apsolutne varijable i izvedenice gdje su na primjer polovine boda takođe moguće tačke - kao što je na primjer slučaj remija na šahovskim turnirima - tu podrazumijevamo i kategoričke varijable. Suprotno tome, neprekidne varijable u teoriji imaju beskonačno finu mjernu skalu i objekat istraživanja može imati tjelesnu masu od 70.1233333* kg. Možete već pretpostaviti da su u realnosti skoro sve mjere koje dobijamo u istraživanjima prekidne. Pobogu, zašto onda i ova podjela? Ona ima svoju važnost kada se uvedu specijalne matematičke distribucije koje predstavljaju idealizovanu sliku stvarnosti, jer će nam one omogućiti da koristimo statistiku za donošenje zaključaka.

Poštediću vas opisa dodatnih podjela varijabli, samo ću ih pomenuti. Dakle, varijable se mogu razvrstati i prema ulozi u nacrtu (npr. nezavisne, zavisne, prediktorske, kriterijske, moderatorske, medijatorske, kontrolne, spoljne), prema stepenu kontrole (registrovane, selektivne, manipulativne), prema objektu mjerenja (npr. subjekt-organizmičke, stimulus, situacione, kontekstualne). Budući da te podjele imaju primarno metodološku funkciju, očekujem da ste njihovo značenje već usvojili na drugom mjestu, a ako niste bićete u kasnijem dijelu knjige podsjećani na njihovo značenje - kada zato bude došlo vrijeme. Nakon što smo se "pozabavili" pitanjima mjerenja, ostalo je da se ma brzinu posvetimo i još nekim krupnim aspektima statističkog rezonovanja.

U čemu je razlika između deterministčkih i probabilističkih hipoteza?

Već smo se podsjetili da se naučna saznanja pohranjuju i prenose jezičkim iskazima, koji kada su naučno provjerljivi nazivamo hipotezama. Zadatak empirijske nauke je da hipotezirane veze među pojmovima iskustveno provjerimo tako što ćemo prikupiti podatke za predstavnike tih pojmova i onda ispitati da li postoji neka statistička pravilnost među njima. Etapu stvaranja podataka smo već definisali (određivanje predmeta mjerenja i varijabli, postupci prebrojavanja i mjerenja), a sad treba objasniti pojam statističke pravilnosti. Prvo ćemo definisati njenu suprotnost, determinističku pravilnost, kroz dvije **determinističke hipoteze**: "Ako je osoba bila izložena nasilju u djetinjstvu onda će ona imati nizak kvalitet mentalnog zdravlja u odrasloj dobi" ili "Žene (sve žene) su verbalno inteligentnije od muškaraca (svih

muškaraca)." Deterministička pravilnost je takva da iskazuje "ako-onda" pretpostavke i gdje nema prostora za vrdanje. Ako bi se pokazalo da postoji odrasla osoba koja je bila izložena nasilju u djetinjstvu i koja je istovremeno sjajnog mentalnog zdravlja (a takvih naravno ima mnogo), to bi automatski opovrgnulo istinutost hipoteze.

S druge strane, analogne **probabilističke hipoteze** bi glasile: "Što je osoba bila više izložena nasilju u djetinjstvu veća je vjerovatnoća da će imati slabiji kvalitet mentalnog zdravlja u odrasloj dobi" ili "U prosjeku gledano su žene verbalno inteligentnije od muškaraca". U ove rečenice su uvedeni pojmovi povećane vjerovatnoće i prosječnih tendencija. Razlog za njihovo uvođenje jeste to što shvatamo da su psihički ishodi (tj. kvalitet mentalnog zdravlja, verbalna inteligencija) multikauzalni, odnosno da imaju više uzročnika. Izloženost nasilju i pol nisu dovoljni da objasne svu raznolikost ishoda, te se ovakvom formulacijom ostavlja mogućnost da neke osobe koje su iskusile nasilje u djetinjstvu imaju odlično mentalno zdravlje, kao i da mnogi muškarci mogu biti verbalno inteligentniji od mnogih žena. S obzirom na to da su psihički ishodi zavisni od mnoštva uzročnika, gotovo sve naučne hipoteze u psihologiji su probabilističke, odnosno statističke.

Šta je to populacija?

Nakon što smo zajedno prošli osnove mjerenja i kada već znamo da smo u potrazi za statistički definisanim vezama ostalo je neodgovoreno još jedno, naizgled banalno, pitanje: Na koga se tačno odnose dati iskazi, odnosno kada ćemo moći da tvrdimo su date hipoteze istinite? Ako malo razmislimo shvatićemo da bi u idealnom slučaju statistička hipoteza o polnim razlikama u verbalnoj inteligenciji bila konačno potvrđena tek onda kada bismo imali podatke o svim muškarcima i ženama na planeti (i još par astronauta) i kada bi oni zaista pokazali da je prosječna vrijednost verbalne inteligencije viša za žene. S druge strane, za hipotezu o mentalnom zdravlju, bilo bi potrebno da ispitamo sve odrasle osobe i pokažemo da, u prosjeku gledano, najbolje mentalno zdravlje imaju oni koji uopšte nisu imali iskustva sa nasiljem u djetinjstvu, da nešto slabije mentalno zdravlje imaju oni koji su imali vrlo malo takvih iskustava, itd.

Iz naprijed navedenog slijedi da kada formulišemo neku hipotezu onda bi ona trebalo da važi unutar datog skupa od interesa kojeg čine svi njegovi članovi (npr. sve žene i svi

muškaraca, sve odrasle osobe). Skup svih elemenata od interesa - odnosno svih elemenata koji imaju definišuća obilježja - nazivamo **populacijom** (engl. *population*) ili rjeđe univerzumom (engl. *universe*). Ja mislim da je još bolje taj skup zvati **ukupnost** i mada ću radi konvencije nastaviti koristiti termin populacija, znajte da ja to sebi prevodim kao ukupnost. Ogromna većina populacija od interesa, tzv. **ciljnih ili teorijskih populacija** (engl. *target population*) su - baš kao i svi pojmovi - tek idealistički zamišljeni skupovi koji se mijenjaju iz sekunde u sekund jer je svaki univerzum dinamično mjesto gdje se rađa, mijenja i umire. To važi ne samo za ispitanike koji su živa bića, nego se može u određenoj mjeri odnositi i na populaciju stimulusa (npr. broj riječi u jeziku, budući da neke nove riječi ulaze u upotrebu kao "krindž"" i "susramlje", a druge odumiru kao što su "rabota", "polza"). Iz tog razloga se najčešće podrazumijeva da su populacije od interesa beskonačne. Statističkom notacijom se broj članova populacije obično označava velikim latiničnim slovom N, pa bi se ova pravilnost mogla napisati kao $N \to \infty$. Ipak, populacije mogu biti i konačne, naročito u praktičnijim istraživanjima gdje to mogu biti svi zaposleni u nekoj organizaciji (npr. $N = 100$) ili svi građani nekog grada ili države. U tim slučavima možda i uspijemo prikupiti podatke za sve članove populacije od interesa što onda zovemo *popis* (engl. *census*).

Usljed očite populacione dinamike i potrebe da jasno predstavimo sebi i drugima na koju tačno ciljnu populaciju se odnose naše tvrdnje, izuzetno je bitno precizno ih definisati. Obavezno to radimo *pojmovno* što po potrebi uključuje i *prostorne* (kulturno-geografske) i *vremenske* odrednice. Na primjer, u nekim istraživanjima je neophodno definisati geografski ili kulturološki kontekst za koji se nešto hipotezira, jer npr. pretpostavke o ponašanju žena u javnosti ne mogu biti iste za žene u Švedskoj i Saudijskoj Arabiji usljed različitih društvenih normi, kao što postoji značajna vjerovatnoća da neki drugačiji faktori mogu da odrede školsko postignuće učenika u Južnoj Koreji i na Zapadnom Balkanu. Opet, u drugim istraživanjima je potrebno definisati konkretan vremenski period na koji se odnosi hipoteza ili odrediti uzrast populacije, npr. stavovi prema seksualnim ponašanjima se mijenjaju kroz generacije i nisu isti čak ni unutar 10 godina, a u istraživanjima jezičkog razvoja razlika između djece sa 14, 16 i 18 mjeseci starosti može biti itekako relevantna. Iz tog razloga, neophodno je eksplicitno definisati populacione granice (engl. *population boundaries*), odnosno **kriterije uključenja** i **kriterije isključenja** ispitanika (još se zovu i inkluzioni i ekskluzioni kriteriji). Na primjer,

ciljnu populaciju je umjesto "djece jedinaca" eventualno moguće definisati kao "djeca jedinci uobičajenog psihičkog razvoja i kompletnog porodičnog statusa uzrasta između 7 i 10 godina", umjesto "trudnica koje su bile trudne tokom pandemije u Bosni i Hercegovini" kao "trudnice koje su bile trudne nakon proglašenja COVID-19 pandemije i koje su najveći dio svoje trudnoće od izbijanja pandemije provele u Bosni i Hercegovini" ili umjesto "depresivnih osoba" kao "odrasle osobe sa dijagnostikovanim depresivnim poremećajem koji nemaju istovremeno dijagnostikovane teže hronične fizičke bolesti (npr. maligna, neurološka i kardio-vaskularna oboljenja) i čija depresivnost nije uzrokovana psihološkom traumom u protekle 2 godine". Očito, konkretna definicija će uvijek zavisiti od onoga što kao istraživači imate na umu. A često ćete morati čitati i tuđe umove, s obzirom na to da veliki broj istraživača zaboravlja navesti jasne definicije populacija.

Kako iz populacije biramo istraživačke uzorke?

Dakle, kompletnu sliku stvarnosti dobijamo samo u idealnom slučaju, onda kada imamo podatke za sve članove teorijske populacije. Budući da ste već stigli do ovoga mjesta u knjizi, dovoljno ste pametni da shvatate da je to u realnosti nedostižno. Istraživanja vršimo na onome do čega možemo doći, a to su komadići stvarnosti koje nazivamo uzorci. Kada treba da provjerite svoju "krvnu sliku" medicinski radnik će vam uzeti uzorak krvi. Iz onoga što je skupljeno u epruveti, a nekad je i par kapi dovoljno, biće očitano vaše stanje u onome šta se dešava u kompletnom krvotoku, odnosno organizmu. Isto tako se o stanju u psihološki relevantnoj populaciji sudi na osnovu podataka prikupljenih na uzorcima. Formalno definisano, **uzorak** (engl. *sample*) je podskup populacije na kojem se provodi istraživanje. Za one koji to ne vide, uzorak nije svaki podskup populacije, nego onaj podskup na kojem se vrši istraživanje. E da, ako sam već populaciju nazvao ukupnost, onda bi uzorak bio istraživani komad (ili komadić, zavisno od njegove veličine).

Između postupaka konceptualnog definisanja ciljne populacije i odabira uzorka nailazimo na jednu međuetapu. Naime, potrebno je definisati **operacionalnu ili radnu populaciju** (engl. *operational population* ili *study population*), ili drugim nazivom, uzorački okvir (engl. *sampling frame*).[8] I u slučajevima kada se bavimo vrlo opštim psihološkim hipotezama za koje pretpostavljamo da geografsko-kulturološki i vremenski kontekst igraju tek

[8] Manji broj autora ide u sitna crijevca i pravi razliku između ova dva pojma, što ja ovdje odbijam da uradim jer uzorkovanje dotičem samo onoliko koliko je neophodno.

zanemarivu ulogu (npr. djeca jedinci su introvertniji od djece koja imaju siblinge, i u Istočnom Timoru i u Kanadi), vjerovatno nećemo imati istraživanje koje se provodi u svim državama svijeta, pa ni na čitavom teritoriju svog grada ili države. Dakle, operacionalna populacija je skup svih ispitanika koji je vama dostupan za istraživanje (operacionalna populacija) i redovno je manja od ciljne populacije na koju želite ekstrapolirati rezultate. Iz tog razloga je od ključnog značaja ispitati stepen poklapanja između ciljne i operacionalne populacije, jer upravo on određuje u kojoj mjeri možemo generalizovati dobijene nalaze na ciljnu populaciju.

Savršeno poklapanje je nemoguće i razumno je tolerisati određena odstupanja, ali je bitno da ona nisu tolika da ugrožavaju valjanost zaključaka. Konkretno, češći je slučaj da radnom populacijom nije pokriveno dovoljno ciljne populacije (engl. *undercoverage*). Na primjer, ako istražujemo stavove mladih o važnosti obrazovanja nije dovoljno u uzorak uzeti samo studente, budući da oni vrednuju obrazovanje više nego njihovi vršnjaci koji ne studiraju, te će rezultati dovesti do precjenjivanja. S druge strane, može se desiti i da istraživanje pokaže pristrasnost usljed prepokrivanja (engl. *overcoverage*). Dva su razloga za to. U jednom slučaju imamo veći broj duplih podataka, odnosno učesnika koji daju više od jednog odgovora. Za to su primjer razne ankete u klasičnim i onlajn medijima gdje "botovi" glasaju za određenu temu po 10 ili više puta. Drugi slučaj se tiče uključivanja u uzorak osoba koji nisu dio ciljne populacije. Na primjer, ako provodimo onlajn istraživanju o stavovima studenata na Zapadnom Balkanu o zadovoljstvu studiranjem onda je potrebno obratiti pažnju na to da se isključe odgovori onih studenata koji studiraju na nekom drugom geografskom području, bez obzira na to da li imaju državljanstvo neke od zemalja Zapadnog Balkana.

Uzorke koji sistematski netačno predstavljaju ciljnu populaciju nazivamo **pristrasnim uzorcima** (engl. *biased samples*). Oni mogu biti posljedica lošeg kvaliteta podataka, kao i neadekvatnog preklapanja ciljne i operacionalne populacije. Na primjer, (zlo)upotrebljavanje studenata psihologije u psihološkim istraživanjima baca sjenku na mnoge nalaze psiholoških istraživanja (a sjetite se i u prošlom poglavlju pomenutog WEIRD uzorkovanja). Iako studenti psihologije mogu biti u velikoj mjeri reprezentativni za neka istraživanja koja se tiču opšte populacije odraslih (npr. istraživanje opažanja kod kognitivno očuvanih mladih osoba), studenti psihologije su u odnosu na svoje vršnjake pretežno liberalne, nematerijalistički orijentisane, empatične, inteligentne, ali i povišeno emocionalne osobe, mahom ženskog pola. To

znači da će uzorak studenata psihologije obično biti slab izbor za istraživanja kojima se žele saznati društveni stavovi ukupne populacije mladih.

Terminom **uzorkovanje** (engl. *sampling*, *sampling procedure*) nazivamo proces odabira uzorka iz operacionalne populacije. Obično se pod ovim u psihologiji podrazumijeva odabir ispitanika (tj. osoba koje učestvuju u istraživanju), ali se uzorkovati mogu i stimulusi (npr. riječi iz populacije vokabulara nekog jezika). Odabir adekvatnog uzorka koji "u malom" vjerno zastupa populaciju od interesa predstavlja ključnu pretpostavku kako bi se donijeli adekvatni zaključci. Treba imati na umu da ćemo samo u rijetkim slučajevima imati spisak članova operacionalne populacije, a da u većini njih to možemo samo nagađati i pouzdati se u neku ekspertsku procjenu.

Uzorak za koji se može pretpostaviti da potpuno vjerno odražava karakteristike populacije se naziva **reprezentativnim uzorkom** (engl. representative sample). Nažalost, taj sveti gral uzorkovanja je još jedna teško uhvatljiva pojava. Štaviše, budući da obično ne znamo sve bitne karakteristike ciljne populacije koje stoje u vezi sa predmetom mjerenja, ne možemo nikad biti sigurni da je naš uzorak potpuno reprezentativan. Srećom, postoje neka zdravorazumska načela kojih se pridržavamo kako bismo se na dimenziji reprezentativnosti mogli približiti mističnom idealu.[9] Logičko-sadržinska ćemo pomenuti malo kasnije, a sada pažnju usmjeravamo na dva formalno-statistička aspekta koji se tiču 1) veličine uzorka i 2) načina njegovog odabira.

[9] Sve u svemu, pogrešno je govoriti o uzorcima kao reprezentativnim i nereprezentativnim, kao da se radi o binarnoj varijabli. Uzorci mogu biti manje ili više reprezentativni.

Kada je u pitanju veličina, stvar je naizgled jasna. Što je veći uzorak, veća je i šansa da je populacija adekvatno predstavljena. Veličina uzorka se označava malim latiničnim slovom n, pa je matematički rečeno poželjno ostvariti što veći omjer $\frac{n}{N}$. Populacija koja ima 1000 ljudi će obično biti tačnije predstavljena ako u uzorku imamo 900 osoba (0.90), nego kada u uzorku imamo samo 10 osoba (0.01). Međutim, u istraživanjima moramo obratiti pažnju i na efikasnost i troškove, jer niti smo vječni pa da imamo vremena napretek, niti su novci nepresušni pa da sve potrošimo na jednom istraživanju. Zato je osmišljena **analiza statističke moći** (engl. *statistical power*) kao posebna grana u statistici koja se bavi izračunavanjem dovoljnog broja ispitanika kako bi se došlo do adekvatnih zaključaka. Na tu granu ćemo se popeti dosta kasnije u ovoj knjizi. Ono što je zasad bitno da shvatite jeste da je moguće imati i prevelike uzorke, budući da se stabilni rezultati sa dovoljno preciznosti mogu dobiti na manjim uzorcima.

Probabilističko uzorkovanje

Način odabira uzorka je također kompleksnija rabota. U idealnom slučaju bi *svi članovi operacionalne populacije imali jednaku vjerovatnoću da budu izabrani u uzorak*. Ovo je osnova probabilističkog pristupa uzorkovanju. Ako je to moguće, ubjedljivo najbolje rješenje bi bilo da se istraživač posluži **prostim nasumičnim odabirom** (engl. simple random sampling). Šta je sad to? To je ono što zovemo sreća u smislu nasumičnosti događaja. Da bi istraživač mogao da napravi prosti nasumični odabir neophodno je da ima spisak svih elemenata operacionalne populacije, te da nasumičnim odabirom (npr. koristeći pseudo-randomne brojeve u nekom tabelarnom softveru ili algoritam koji koristi pseudo-randomne brojeve) izabere one koji će učestvovati u istraživanju. A kada ćete to moći da napravite? Skoro nikad. Dobro, možda onda kada se bavite izuzetno malim populacijama za koje ste obezbijedili potpun spisak, ali jako rijetko onda kada radite istraživanja zasnovana na popisu stanovništva i domaćinstava. Zato pribjegavamo drugim načinima odabira.

Ti drugi načini uzimaju istovremeno u obzir i formalno-statističke i logičko-sadržinske aspekte. Na primjer, poželjno bi bilo da svi relevantni **slojevi populacije** (nivoi, stratumi, engl. **stratum**, pl. **strata**) budu zastupljeni u uzorku proporcionalno njihovom udjelu u populaciji. Na primjer, ukoliko pretpostavljamo da je pol karakteristika koja je sistematski povezana sa izraženošću varijabli od interesa (npr. ispituje se veza između crta ličnosti i seksualne otvorenosti mladih, a znamo da mladići imaju seksualno otvorenije stavove), onda je potrebno da u uzorak uzmemo podjednak broj mladića i djevojaka ukoliko želimo da donesemo zaključak koji bi se odnosio na cijelu populaciju mladih. Ukoliko smatramo da je i neka druga demografska karakteristika bitna (npr. veličina mjesta stanovanja, obrazovni status, starost ispitanika) onda je potrebno da i te karakteristike budu proporcionalno zastupljene, pri čemu biste te karakteristike morali kombinovati sa polom. Ovakve odluke su očito vezane za fazu određivanja operacionalne populacije, pa možemo da shvatimo da je zapravo potrebno definisati više operacionalnih populacija (npr. studentkinje iz velikog grada, nezaposleni mladići koji žive na selu). Ako unutar takvih subpopulacija možemo napraviti nasumični odabir onda to zovemo **stratifikovanim nasumičnim odabirom** (engl. stratified random sampling). Dakle, ovdje se princip jednake vjerovatnoće odabira unutar populacije rasteže na jednaku vjerovatnoću unutar stratuma. A da biste mogli znati kolika je ta vjerovatnoća, morate znati koliki je udio određene

kategorije unutar populacije. Avaj, usljed komplikacija u praktičnom izvođenju i ovaj vid uzorkovanja se relativno rijetko ostvaruje i obavezno mi se pohvalite mejlom kada to napravite.

Uprkos pesimizmu kojeg izgleda propagiram, u praksi se ipak dešava da imamo elemente probabilističkog uzorkovanja i to onda kada su ispitanici grupisani u hijerarhijski nadređene skupove. Tipičan primjer se tiče školskog okruženja: učenici su u odjeljenjima, odjeljenja su u različitim školama, škole u različitim gradovima. Na primjer, dobili ste zadatak da uradite istraživanje sa učenicima šestih razreda osnovnih škola u jednom većem gradu koji ima 20 škola, ali nemate dovoljno resursa da ga izvedete na svim učenicima u svim školama. Kako biste dobili što reprezentativniji uzorak bilo bi dobro da škole odaberete nasumično. Nakon što tako izaberete, recimo 4 škole, u drugoj fazi se možete odlučiti da nasumično odaberete jedno od više odjeljenja u svakoj odabranoj školi. Ako vas nedovoljni resursi sprečavaju da ispitate sve učenike u odabranim odjeljenjima, onda je moguće u trećoj fazi da nasumično odaberete učenike iz tih odjeljenja. Ono što sam upravo opisao je postupak **klasterskog uzorkovanje** (engl. *cluster sampling*) koje može biti izvedeno u jednom koraku (npr. samo škole su nasumično odabrane, u njima svi ispitanici) ili biti **višefazno klastersko uzorkovanje** (engl. *multistage cluster sampling*) kao što bi to bilo u navedene tri faze. Ozbiljnija istraživanja pribjegavaju ovakvom uzorkovanju i koriste posebne statističke metode koje su nešto naprednije od onoga što ćemo stići proraditi u ovoj knjizi.

Neprobabilističko uzorkovanje

S druge strane, velika većina istraživanja u psihologiji, ali i drugim društvenim naukama koristi neprobabilističko uzorkovanje, mada postoje i kombinacije probabilističkog i neprobabilističkog uzorkovanja. Najpopularniji oblici neprobabilističkog uzorkovanja su prigodni i kvotni. **Prigodni uzorci** (engl. *convenience sample*) su oni gdje je osnovni kriterij za odabir ispitanika taj da su ispitanici lako dostupni i voljni da učestvuju u istraživanju (ako su studenti psihologije, ne moraju biti ni nešto voljni). Obično je cilj da se "nahvata" dovoljan broj ispitanika, odnosno da se zadovolji kriterij veličine uzorka. Neki ovo zovu i ad hoc uzorkovanjem, a često se prigodno uzorkovanje brka sa nasumičnim uzorkovanjem. Recimo da je potrebno da u vašem istraživanju imate 20 mladih osoba od 18 do 25 godina i da odlučite da će to biti prvih 20 koje sretnete unutar fakultetske zgrade. To svakako nije bilo nasumično u smislu teorije

vjerovatnoće, jer itekako postoji veća vjerovatnoća da srećete svoje kolege sa kojima provodite mnogo vremena i na istoj ste geografskoj lokaciji (npr. studenti psihologije, studenti Filozofskog fakulteta), nego neki drugi studenti koji su možda i na drugim lokacijama, a da ne govorimo o mladima koji ne studiraju i ne zalaze tako često u zgradu ili univerzitetski kampus.

Ako po istom prigodnom principu "lovimo" ispitanike kako bismo ispunili unaprijed zadane procentualne kvote - koje smo definisali na osnovu stratuma - onda zapravo imamo **kvotne uzorke** (engl. *quota sample*). Na primjer, određeno je da populaciju od interesa čini 50% žena i 50% muškaraca, te 60% osoba sa završenom srednjom školom i 40% sa završenim visokim obrazovanjem. Ako je potrebno da uzorak ima 100 ispitanika onda će biti potrebno po 30 žena i muškaraca sa srednjom stručnom spremom i po 20 žena i muškaraca sa visokom stručnom spremom, a ispitivači moraju da se snađu. Postavlja se pitanje: mogu li prigodni i kvotni uzorci biti dovoljno reprezentativni, odnosno mogu li oni uopšte valjano predstavljati teorijske populacije? Odgovor je da mogu, ali da to umnogome zavisi od dodatnih odluka istraživača. Da shvatimo tu vezu, pomoći će nam definicija još jednog neprobabilističkog načina uzorkovanja.

Radi se **namjernom uzorkovanju** (engl. *purposive sampling*), koje ima i druge nazive kao što su selektivno ili subjektivno uzorkovanje. Ova imena ukazuju na to da istraživač namjerno bira određene elemente u uzorak na osnovu svog suda, one za koje misli da će na najbolji način da predstave populaciju. Konkretno načelo koje će da stoji iza namjerne odluke istraživača može biti različito i ima ih mnogo (npr. heterogenosti, homogenosti, tipičnosti, ekstremnosti, ekspertize). Daću primjer ovdje samo jednog načela, jer je namjerno uzorkovanje opšta metodološka tema koja nije direktno vezana za statističke postupke. Na primjer, u istraživanju je potrebno da se odrede norme za neki psihološki test u populaciji osoba od 18 godina, a istraživači znaju da nemaju ni novca ni vremena da koriste neki od oblika probabilističkog uzorkovanja. Kako bi razumno predstavili raznoliku populaciju mladih od 18 godina ono što mogu da urade jeste da izvrše istraživanje u srednjim školama - budući da je to mjesto gdje se istraživanje može relativno lako uraditi - ali tako što će pažljivo izabrati škole koje bi predstavljale raznovrsne profile. Na primjer, uz nekoliko odjeljenja učenika gimnazije u uzorku bi se onda moralo naći i nekoliko odjeljenja iz trogodišnjih stručnih škola, ali većina učenika bi morala pohađati polno mješovite škole koje upisuju učenici prosječnog postignuća tokom ranijeg obrazovanja. To bi bila strategija obezbjeđivanja heterogenosti

(raznovrsnosti) uzorka, što bi trebalo da odražava i heterogenost teoretske populacije.

Važno je da zapazite dvije stvari. Kao prvo, isti broj ispitanika možete prikupiti provodeći istraživanje samo u gimnazijama, ali ukoliko akademsko postignuće ima efekat na rezultate testova, "čisto-prigodni-uzorak" će rezultirati u pristrasnijim nalazima koji netačnije odražavaju karakteristike populacije. Drugo, možemo zaključiti da za svaki prigodni uzorak istraživači mogu u različitom stepenu nametnuti svoje odluke, što onda utiče na količinu reprezentativnosti. I eto odgovora na pitanje o reprezentativnosti i neprobabilističkim uzorcima: logičko-sadržinsko poznavanje problema kojim se bavite može obezbjediti dobre odluke pri izboru uzorka.

Šta podrazumijevamo pod pojmovima parametar i statistik?

Sa prevashodno metodološkog polja se vraćamo sad na striktno statističko. Upoznaćete još dva izuzetno bitna pojma: parametar i statistik, te kako su oni vezani za različita polja statistike.

Parametar (engl. parameter) je mjera koja važi za populaciju i nju opisuje. Usljed fluktuiranja sastava i veličine populacija, parametar je vrlo rijetko mjerljiv u praksi. Drugim riječima, skoro nikad ne znamo njegovu stvarnu vrijednost, sem u slučaju ograničenih / konačnih populacija. Iako smo svjesni da ga u ogromnoj većini slučajeva ne možemo spoznati, želimo da procijenimo njegovu vrijednost. Taj mistično-idealistički, onostrani cilj se ogleda i u načinu na koji se parametri predstavljaju u statističkoj notaciji: kao simboli za predstavljanje parametara koriste se mala grčka slova. U zavisnosti od toga da li se radi o nekoj mjeri varijabilnosti, proporcije, povezanosti ili drugim bitnim statističkim karakteristikama koristićemo različita slova da označimo parametar (npr. σ, π, ρ). Štaviše, postoji i jedno slovo koje će da označava čitavu porodicu parametara, a to je slovo θ, koje čitamo teta.

Parametar = mjera populacije

Statistik (engl. statistic) je mjera za predmet mjerenja koju smo dobili na uzorku, odnosno to je neka konkretna vrijednost izračunata na podacima sa uzorka. Pazite: *statistik*, a ne *statistika*. Kako bi se razlikovali od parametara u statističkoj notaciji se kao simboli za predstavljanje statistika koriste latinična slova koja se u štampanim materijalima pišu kurzivnim slovima (npr. s, p, r). Na osnovu vrijednosti statistika i upotrebom teorije vjerovatnoće može se procijeniti traženi parametar, a u tom postupku se uzima

Statistik = mjera uzorka

u obzir i mogućnost greške pri procjeni. Kada za neki statistik želimo eksplicitno da naglasimo da ćemo ga koristiti za procjenu parametra koristimo i naziv **procjenjivač** (ili estimator od engl. estimator). U tim slučajevima se umjesto latiničnih slova vraćamo na grčka, samo što ćemo im staviti kapicu na glavu (npr. $\hat{\sigma}$, $\hat{\pi}$, $\hat{\rho}$). Kapica na glavi slova teta ($\hat{\theta}$) znači da je to procjenjivač parametra, uopšteno govoreći. Rekoh li da su statističari komplikovani?

Razliku između parametara i statistika možemo predstaviti na vrlo jednostavnom primjeru. Recimo da smo zainteresovani da saznamo prosječnu visinu odraslih muških osoba u Kini. Drugim riječima, želimo da procijenimo parametar prosječne visine koji ne znamo ($\theta = ?cm$). Svjesni smo da ga ni ne možemo saznati jer svakog dana izuzetno mnogo mladih puni 18. godina, ali i veliki broj umire. Iz tog razloga bi se istraživanje obavilo na uzorku od npr. 500 muškaraca starosti od 18 do 65 godina, pa bi za konačnu procjenu poslužio na tom uzorku dobijeni statistik (npr. $\hat{\theta} = 172cm$).

Na koje podoblasti se dijeli statistika?

Deskriptivna statistika vs. statistika zaključivanja

Veliki broj tekstova dijeli ukupnu oblast statistike na dva dijela. **Deskriptivna statistika** (engl. *descriptive statistics*) obuhvata one statističke postupke kojima numerički i grafički analiziramo i sažeto predstavljamo podatke sa uzorka. **Statistika zaključivanja** (engl. *inferential statistics*) obuhvata postupke kojima zaključujemo nešto o stanju u populaciji na osnovu rezultata prikupljenih na uzorcima. Ovo se izvodi primjenom opštih principa teorije vjerovatnoće i moćima intenzivnih proračuna koje obavljaju današnji računari. Nešto konkretnije, statistika zaključivanja obuhvata procjenu populacionih parametara, provjeru statističkih hipoteza, te provjeru statističkih modela. Sve u svemu, deskriptivna statistika se tiče statistika i njihovog prikazivanja, a statistika zaključivanja predviđanjem vrijednosti parametara.

Osim deskriptivne statistike i statistike zaključivanja, pominju se i još neke oblasti, a među njima najčešće **eksplorativna analiza podataka** (engl. *exploratory data analysis*, *EDA*). Radi se o postupcima na podacima sa uzoraka kojima se žele steći novi uvidi o pravilnostima koje bi važile u populaciji. Ako bi se baš morala napraviti razlika između eksplorativne analize podataka i, u prošlom poglavlju pomenutih, **rudarenja podataka** (engl. *data mining*) i **mašinskog učenja** (engl. *machine learning*), onda se kao kriterij može uzeti stepen automatizacije. EDA postupci se

u najvećoj mjeri zasnivaju na ekspertizi istraživača i manuelnim traganjem za obrascima, mašinsko učenje se u najvećoj mjeri zasniva na algoritmima koje u finalnoj formi istraživači uopšte ni ne razumiju ali su tu da ih navode i popravljaju, dok je rudarenje podataka negdje na pola puta. Naravno, granice ovih oblasti koje pogone podaci (a u manjoj mjeri teorija) su vrlo propusne i u modernoj statistici je teško odrediti gdje prestaje jedna, a počinje druga oblast. U ovom tekstu je naglasak na znanjima o klasičnim statističkim oblastima, što će vam svakako biti neophodno i ako u svojim avanturama pribjegnete mašinskom učenju.

Nakon što smo datom podjelom iscrpili ovo teoretsko poglavlje, vrijeme je da se prebacimo na praktičnije poslove. U sljedećem poglavlju ćete saznati kako se prikupljeni podaci pripremaju za obradu - što je neophodan korak da bismo ostvarili plemenite ciljeve ispitivanja naučnih hipoteza i modela. Istina je da taj proces nije nešto zabavan i može biti dugotrajan, ali držite na umu da i svaki vrhunski sportista provede mnogo više vremena na treninzima nego u takmičarskim mečevima.

Poglavlje 3

Baze podataka u psihološkim istraživanjima

Nakon čitanja ovog poglavlja znaćete:

- kako se kreira baza podataka koja se koristi u psihološkim istraživanjima,
- kako se unose i snimaju podaci u bazi,
- postupke uređivanja sirovih podatke tako da se na njoj može vršiti statistička obrada.

Recimo da smo uspješno odredili problem istraživanja, formirali jezičku hipotezu i da smo prošli proceduru prikupljanja podataka potrebnih za njenu empirijsku provjeru. I šta sad kada imamo gomilu informacija i strastvenu želju da dođemo do epohalnog naučnog otkrića? Čeka nas ne tako uzbudljiva, ali odlučujuća faza pripreme podataka za njihovu obradu u statističkom softveru. Veliki poduhvati zahtijevaju dosadne pripremne poslove, a to je slučaj i sa naučnim avanturama. Ovaj dio avanture uključuje kreiranje baze, unos podataka, sređivanje baze i njeno arhiviranje.

Šta su baze podataka i koje oblike mogu imati?

Bazom podataka (engl. *database*) ćemo zvati sistematski organizovan skup podataka arhiviran u digitalnom obliku. U psihološkim istraživanjima se pod tim pojmom najčešće misli na jednostavne, dvodimenzionalno strukturisane **tabele podataka** (engl. *dataset*). One se najčešće definišu kao što je prikazano na Slici 3.1. Za početak primijetite da se u prvom redu nalaze imena varijabli (*rb*, *ime_studenta*, itd.). Ispod tog reda se nalazi matrica podataka u kojem je svaki sljedeći red rezervisan za pojedinačnog ispitanika (objekat ili entitet), a svaka kolona sadrži podatke o pojedinačnim varijablama. Bilo da je riječ o

nizu podataka za nekog ispitanika ili nizu podataka za neku varijablu taj niz predstavlja jedan vektor (vektori su već definisani u prvom poglavlju). Ovakve matrice u kojima se ukrštaju vektori za ispitanike i vektori za varijable se nazivaju ispitanici X (čitaj: puta) varijable matrice podataka, a čućete i naziv **matrice podataka širokog formata**.

	A	B	C	D	E
1	**rb**	**ime_studenta**	**pol**	**studij**	**strah_statistika**
2	1	Jasna	1	psihologija	8
3	2	Janko	2	pedagogija	7
4	3	Jasmina	1	sociologija	9
5	4	Jan	2	psihologija	4
6	5	Jadranka	1	psihologija	8

Slika 3.1: Baza podataka širokog formata kreirana u LibreOffice softveru.

Za nacrte sa ponovljenim mjerenjima, odnosno vezanim mjerama, u kojima se za jednog ispitanika dobija više podataka za isti predmet mjerenja (npr. ispitanik tri puta tokom semestra procjenjuje koliko mu je snažan strah od statistike), se može koristiti i tzv. **dugački format matrice podataka**. U tom dugačkom formatu redovi ne zastupaju pojedinačne ispitanike nego pojedinačna mjerenja varijable od interesa. Pogledajmo Sliku 3.2. Podaci se ponavljaju za neke varijable (redni broj ispitanika, ime studenta, pol ispitanika, studijski program koji se studira), jer njihove vrijednosti ostaju iste tokom istraživanja i odnose se na jedan te isti entitet. Međutim, strah od statistike se mijenjao tokom tri mjerenja koja su se mogla desiti 15. oktobra (mjerenje = 1), 1. decembra (2) i 15. januara (3). Kasnije u ovom poglavlju je opisani i postupak prevođenja baze iz jednog u drugi format.

	A	B	C	D	E	F
1	**rb**	**ime_studenta**	**pol**	**studij**	**strah_statistika**	**mjerenje**
2	1	Jasna	1	psihologija	8	1
3	1	Jasna	1	psihologija	6	2
4	1	Jasna	1	psihologija	3	3
5	2	Janko	2	pedagogija	7	1
6	2	Janko	2	pedagogija	5	2
7	2	Janko	2	pedagogija	5	3

Slika 3.2: Baza podataka dugačkog formata.

Koje korake moramo napraviti u fazi planiranja baze podataka?

Na prvi pogled može izgledati da redoslijed potrebnih koraka varira u zavisnosti od načina prikupljanja podataka, jer kada radimo onlajn istraživanja dobijamo i automatski generisanu bazu. Međutim, tako stvorene baze su obično tek preliminarne verzije koje je neophodno doraditi prije analize. Iz tog razloga, redoslijed koji dalje predstavljam se može smatrati univerzalnim.

Šta je to šifarnik i čemu služi?

Nakon što smo odlučili koje mjerne instrumente koristimo a prije nego što uopšte počnemo prikupljati podatke - potrebno je da načinimo šifarnik. **Šifarnik** (engl. *code sheet /codebook*) je dokument[1] koji detaljno opisuje bazu drugim ljudima (ili vještačkoj inteligenciji) koji će je potencijalno koristiti za obradu podataka. Šifarnik ima i druge bitne funkcije: služiće nam kao podsjetnik (vjerujte mi, problemi sa pamćenjem će biti sve učestaliji) i predstavljaće svojevrstan plan za kreiranje baze podataka. Dobar šifarnik minimalno sadrži sljedeće elemente:

[1] Šifarnik se može kreirati kao prosti tekstualni fajl (.txt), ali je poželjno da bude i čitljiv u tabelarnim softverima (npr. .csv). Kasnije u poglavlju je predstavljen opširan pregled opcija o preferiranom softveru i ekstenzijama dokumenata.

- imena varijabli
- opis svake od varijabli
- opseg / opis teorijski mogućih vrijednosti koje podatak može imati

Kako adekvatno imenovati varijable u bazi?

Imenovanje varijabli u bazi nije tako izazovno kao davanje imena svom djetetu, jer postoje neki standardi. Iako danas ne postoje bitnija ograničenja u pogledu dužine imena, preporučivo je da imena budu kratka. Koliko kratka? Onoliko koliko je dovoljno da pri obradi nemamo dvojbe o tome šta varijabla označava. Za neke varijable možemo lako osmisliti kratko i razumljivo ime (npr. *pol*, *starost*, *studij*). Za druge ćemo se morati malo potruditi, kao za pitanje o zaposlenosti trudnica u jednoj anketi gdje tvrdnja glasi "Nisam zaposlena jer…", a postoji pet ponuđenih odgovora: "Ne želim da radim." "Ne mogu da nađem posao." "Radila bih, ali nema ko da mi čuva djecu.", "Dobila sam otkaz." i "Nešto drugo (opisati šta).". Jedna mogućnost je da ovakvu varijablu nazovemo *nezaposlena_razlog*. Ako bismo je nazvali samo *nez_raz* ili ukoliko bismo je označili prema redoslijedu

pitanja u čitavoj anketi (npr. *p12_3*) može se desiti da pri obradi zaboravimo o čemu se tu radilo, te ćemo morati trošiti vrijeme na zagledanje šifarnika pri obradi svaki put kada želimo da analiziramo ovu varijablu.

S druge strane, za jedan tip varijabli obavezno koristimo vrlo jednostavan recept koji zanemaruje sadržaj stavke. Već ste saznali da se u psihološkim istraživanjima izuzetno često koriste kompozitni instrumenti koji su sastavljeni iz više stavki (npr. 21 stavka kojima se procjenjuje depresivnost ili 60 zadataka na testu inteligencije). U takvim slučajevima za imenovanje varijabli koristimo postojeće akronimne nazive tih instrumenata (npr. BDI od Beck's Depression Inventory ili RPM od Raven's Progressive Matrices), što prati redni broj stavke. U tim slučajevima bismo varijable nazvali *bdi_01, bdi_02…bdi_21* ili *rpm_01, rpm_02…rpm_60*). Primijetite da je preporuka da ako imate više od 9 stavki, onda za redne brojeve bi trebalo da koristite onoliki broj brojki koliko ih ima i zadnja stavka. Drugim riječima, ako imate 90 stavki, za prvu od njih unosite nulu ispred broja 1. Razlog za to je što izvjestan broj statističkih programa može da poremeti redoslijed pri izlaganju rezultata, tako što sortira stavke prema prvoj nastupajućoj brojci (npr. bdi_11 može doći prije bdi_2), a što onda pravi nepotrebnu konfuziju u prikazu.

Ipak, postoje određena ograničenja pri imenovanju stavki. Imena varijabli ne smiju sadržavati razmake, niti sljedeće interpunkcijske znakove (! , ? # @). Kako bi se omogućila puna kompatibilnost sa različitim kompjuterskim sistemima nije preporučivo ni da sadrže "naša" slova sa kvačicama (č, ć, đ, š, ž) ili da budu imenovana ćirilicom. Uz to, imena ne bi smjela počinjati niti interpunkcijskim znakovima, niti brojkama. Uobičajene konvencije su da se složeni nazivi varijabli (npr. *socijalna distanca* što predstavlja pojam iz oblasti socijalne psihologije) pišu razdvojeni na jedan od tri načina: ili donjom crtom (*soc_dist*) ili tačkom (*soc.dist*) ili kamiljim / grbavim stilom (*socDist*). Korištenje isključivo malih slova, brojeva i donjih crta je preferirano u R-u, a iskustvo mi govori da je to najbolji način kako bi sve bilo nedvosmisleno i pregledno.[2]

[2] Ako to nekoga interesuje, na internetu možete naći i smjernice za stilizaciju kodnog pisanja i baza podataka (npr. https://style.tidyverse.org/syntax.html ili https://google.github.io/styleguide/Rguide.html).

Kako opisati varijable u šifarniku?

Opis varijabli mora biti takav da nepoznati naučnik koji se odluči da nam zaviri u bazu može brzo shvatiti šta smo i kako smo to procjenjivali, jer sama imena varijabli neće biti dovoljna da se to zaključi. Usput, prva varijabla koji bismo uvijek morali

imati u bazi je varijabla koja označava redni broj, odnosno identifikacionu oznaku ispitanika. Skoro uvijek je imenujemo kao *rb* (skraćenica na našem jeziku) ili *id* (kada očekujemo da ćemo dijeliti rezultate van regiona), te bismo uz nju na ovom mjestu napisali opis "redni broj ispitanika". Ako bismo u našem istraživanju imali gore pomenutu varijablu *nezaposlena_razlog* onda bismo uz nju naveli kompletnu tvrdnju, te eventualno napomenuli da su na to pitanje odgovarale samo one ispitanice koje su u prethodnom pitanju navele da su trenutno nezaposlene. Gdje god je to moguće uraditi, naveščemo originalnu stavku (npr. to ne možete uraditi za stavke sa testa Ravenovih progresivnih matrica jer se radi o figuralnim zadacima). Pritom, obavezno imajte na umu da ukoliko planiramo da dijelimo šifarnik sa drugima moramo poštovati prava o intelektualnoj svojini. To znači da ako smo dobili prava za zadavanje nekog komercijalno zaštićenog instrumenta (npr. ranije pomenuti Beckov inventar depresivnosti) onda moramo da napravimo poseban šifarnik koji ćemo javno dijeliti, ali ćemo iz te verzije obrisati sadržaj tvrdnji. Naravno, ukoliko se radi o nekoj drugoj tehnici generisanja podataka, to ćemo ovdje i opisati (npr. *opaženi broj dodira nosa u eksperimentalnoj situaciji - posmatrač #1* , *vrijeme reakcije na crvenu draž u milisekundama* ili *slobodni kortizol u uzorcima pljuvačke (*$\mu g/dL$*)*).

Kako se definišu moguće vrijednosti u šifarniku?

Uz opis same varijable, neophodno je da popišemo vrijednosti koje podaci mogu da imaju - bez obzira na to da li su te vrijednosti zaista dobijene na našem uzorku ili nisu. Tek tada baza podataka postaje dovoljno razumljiva svim zainteresovanima, te će biti moguće provjeriti valjanost dobijenih podataka. Za kategoričke varijable ćemo navesti značenja korištenih oznaka za pojedine kategorije. Na primjer, ako smo odlučili da za vrijednosti varijable *nezaposlena_razlog* - koja je nominalnog nivoa mjerenja - koristimo brojke, onda bismo ovdje naveli korištene kodove: *1 = Ne želim da radim.; 2 = Ne mogu da nađem posao.;…* Za dimenzione varijable navodimo opseg mjerne skale, odnosno njihov teorijski minimum i maksimum. Na primjer, za varijablu naziva *ispit_bodovi* koja se dobija zbrajanjem broja tačnih zadataka od njih ukupno 30, naveščemo da je minimum 0, a maksimum 30.

Šta bi još trebalo da stoji u šifarniku?

Obično ćemo početnu verziju šifarnika dopuniti nakon što prikupimo podatke, budući da se mnoge varijable od interesa stvaraju iz već postojećih. To se pogotovo odnosi na pitanja sa tzv. otvorenim odgovorima (engl. *open-ended questions*) gdje su ispitanici slobodni da sami upišu narativni odgovor. Obično jedna varijabla služi kao posuda za takve odgovore, a zatim se definiše jedna ili više dodatnih varijabli za pretvaranje takvih odgovora u kategoričke. Kao ilustraciju, pretpostavimo da se u nekom onlajn istraživanju od ispitanika traži da upišu mjesto prebivališta. Ako smatramo da je za naše istraživanje veličina mjesta stanovanja bitna karakteristika, onda bi iz izvorne varijable gdje su ispitanici upisivali puno ime mjesta (npr. Banja Luka, Trebinje) bilo moguće stvoriti nove. Na primjer, mogli bismo stvoriti ili ordinalnu kategoričku varijablu sa kategorijama *veći grad*, *manji grad*, *selo* ili numeričku varijablu gdje bismo ime mjesta zamijenili brojčanom procjenom veličine mjesta sa posljednjeg popisa stanovništva (npr. za Banja Luku *138963*, za Trebinje *25589*). Bitno je da zapazite da je iz jedne izvorne varijable moguće izvesti više novih varijabli. Na primjer, na osnovu mjesta stanovanja možemo evidentirati i državu iz koje dolaze odgovori ukoliko smo radili regionalno ili međunarodno istraživanje.

Kad su u pitanju zatvoreni odgovori, broj varijabli u bazi se dopunjava putem postupaka koje ćemo nešto kasnije definisati (rekodiranje, transformisanje i izračunavanje novih varijabli), te je za sve njih potrebno da naknadno dodamo informacije u šifarnik nakon faze unošenja podataka u bazu ako to već nismo mogli napraviti na početku.

Konačno, poželjno je da kao dio šifarnika ili kao poseban tekstualni fajl imamo i sažeti opis ukupne baze (engl. *summary*) kojim se opisuje koja je bila svrha prikupljanja podataka, gdje i kada su podaci prikupljeni, kako su podaci prikupljeni i uneseni. To po potrebi obuhvata i opisivanje etičkih aspekata koji su vezani za proces prikupljanja podataka i njihovu kasniju distribuciju (npr. anonimnost podataka, dostupnost podataka za dalju analizu).

Kako kreiramo baze podataka u softveru

Nakon što smo kreirali i sačuvali početnu verziju šifarnika, prelazimo na fazu kreiranja baze. U tu svrhu moramo da izaberemo softver, a najefikasnije je koristiti programe namijenjene za

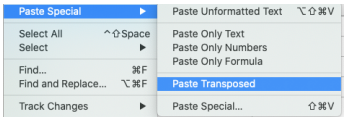
	A	B	C
1	**ime**	**opis**	**moguće vrijednosti**
2	rb	Redni broj ispitanika	1-250
3	ime_studenta	Ime ispitanika	
4	pol	Pol ispitanika	1=ženski, 2=muški
5	studij	Studij koji ispitanik studira	
6	strah_statistika	odgovor ispitanika na skali od 11 podioka na pitanje "Sveukupno gledano, u kojoj mjeri strahujete od predmeta statistika u [ime oblasti]?"	0-10
7	savjesnost	Savjesnost kao crta ličnosti procjenjivana HEXACO-60 upitnikom	Prosječna vrijednost odgovora na 10 stavki na skali 1 do 5.
8	religioznost	Oblik religioznosti kojeg ispitanik smatra sebi svojstvenim, a na osnovu vinjetnog opisa tipičnog uvjerenja.	Nominalna varijabla, 5 tipova: 1=institucionalno religiozan, 2 = neinstitucionalno religiozan, duhovan/spritualan, 4=agnostik, 5=ateista.

Slika 3.3: Primjer šifarnika kreiranog u LibreOffice-u.

tabelarne proračune. Tipični predstavnici takvih programa su *Microsoft Excel*, *Google Sheets* i *Apple Pages*, ali moja iskrena preporuka je da koristite open source rješenja kao što je *LibreOffice Calc*. LibreOffice je besplatna i pouzdana alternativa različitim Office paketima, može se instalirati na sve operativne sisteme i pri radu ne ovisimo o stabilnosti i brzini internet konekcije. Primjeri na prethodnim slikama su kreirani upravo u LibreOffice-u. Usput, preporučujem da i šifarnik kreirate u tabelarnom softveru tako što će tri kolone sadržavati: imena varijabli, opise varijabli, definicije mogućih vrijednosti (Slika 3.3.). To je preglednije nego da koristimo obične tekstualne fajlove koji se takođe sreću.

Ako ste koristili tabelarni softver za pravljenje šifarnika, postupak kreiranja baze podataka je jednostavan:

1. Otvorite novi prazan fajl u tabelarnom softveru
2. Kopirajte sva imena varijabli iz prve kolone šifarnika
3. Nalijepite transponovane vrijednosti u prvi red novog fajla



Slika 3.4: Transponovano znači da redovi postaju kolone, a kolone postaju redovi; prikazana je komanda transponovanog lijepljenja dostupna u LibreOffice-u.

Ako ste se kojim slučajem odlučili da se oglušite na moju preporuku, onda ćete morati pješke da upišete imena varijabli u prvi red istim redoslijedom kojim će i podaci biti unošeni. Za koji god se način odlučili obično ćete morati unijeti konačna imena varijabli za automatski generisane baze preuzete iz onlajn istraživanja. Nakon ovog prvog koraka inicijalnu bazu snimamo. I za to postoje neke preporuke.

Jedno od osnovnih načela nauke jeste biti transparentan što - kada je u pitanju statistička analiza - podrazumijeva i da drugi zainteresovani istraživači imaju pristup podacima. Iz tog razloga je preporučivo da baze snimamo u najprostijim, ljudski čitljivim formatima, odnosno da kada bazu snimimo možemo pročitati podatke bez obzira na to koji statistički softver koristimo. Primjeri

Slika 3.5: Primjer jedne baze snimljene kao .csv fajl i otvorene u tekstualnom softveru u kojem se nalaze imena svih varijabli i podaci za jednog ispitanika.

takvih datoteka / fajlova su datoteke sa ekstenzijama .csv, .dat, .tab i .txt koje svaki tabelarni softver ima mogućnosti da ih snimi. Preferiraćemo .csv format - čiji naziv dolazi od engleskog *comma separated values*, što u prevodu znači *vrijednosti razdvojene zarezom*.

To su i bukvalno vrijednosti razdvojene zarezom (Slika 3.4), te bi jedan vektor za ispitanika koji je odgovorio na tri pitanja o mjestu stanovanja, zanimanju i godinama starosti mogao biti: *Banja Luka, psiholog, 44*. Naravno, moguće je, a biće povremeno i vrlo poželjno da tokom procesa analize arhiviramo baze i u specifičnim formatima koje proizvode i čitaju određeni programi kao što su .xls, .xlsx, .ods, .sav, .jasp, .dta, .Rdata.

Uz sve prednosti .csv formata postoje i neki problematični aspekti kojih moramo biti svjesni kako bismo preduprijedili neželjene događaje. Kao prvo, na našem geografskom području se zarez koristi kao decimalni znak, pa postavke operativnog sistema utiču na to da li će računar "doživjeti" zarez kao decimalni znak ili kao znak za razdvajanje polja u koje se unose podaci (engl. *separator* ili *delimiter*). Ovo je moguće provjeriti na početku unosa podataka i po potrebi promijeniti postavke. Nadalje, kada ispitanici navode otvorene odgovore moguće je da se naiđe na zarez (npr. "*Tokom onlajn nastave sam koristio Zoom, Google Meet i BigBlueButton.*"), te to može dodatno zbuniti softver. Iz tog razloga je sigurnija alternativa da na našem području koristimo .csv format sa malom modifikacijom, gdje ćemo kao separator polja koristiti oznaku **;** (tačka-zarez) koja se rjeđe sreće u odgovorima ispitanika. Ovo je moguće podesiti u svakom tabelarnom softveru, npr. u LibreOffice Calc-u pri snimanju podataka softver nudi opciju kojim znakom će biti odvojena polja, pa čak i opcija kojima se mogu razgraničiti otvoreni odgovori (engl. *string delimiter*) kako bi računaru bilo jasno o čemu se radi. I konačno, pri snimanju podataka će nas softver pitati u kojem kodnom formatu želimo snimiti podatke, a mi ćemo izabrati UTF-8 (najčešće korišteni kodni format) kako bi softver mogao kako treba pročitati sva "naša slova" - uključujući i ćirilično pismo - kada i ako se ona pojave među podacima.

Ostaje još pitanje gdje da snimimo bazu podataka i kako da je imenujemo. Čak i za to postoje neke preporuke. Kako bismo što više optimizovali svoj radni proces, dobro je da podatke snimamo u zasebnom direktorijumu - koji će nositi ime *podaci* i koji će biti podređen direktorijumu koji nosi naziv našeg istraživanja. Ostali direktorijumi mogu nositi nazive *instrumenti* (u koji ćemo smjestiti datoteke o mjernim instrumentima), *rezultati* (u kojima ćemo snimati rezultate statističke analize), *komande* (u kojima

ćemo snimati obavljene komande ako koristite R ili neki drugi statistički jezik), a vjerovatno ćemo imati i posebne direktorijume koji će biti zaduženi za usmene prezentacije ili publikacije koje će nastati na osnovu rezultata. Kada je u pitanju imenovanje same datoteke većina eksperata se slaže da je potrebno da ime sadrži i prepoznatljiv naziv i datum (npr. iq_skole_23mar.csv ili 23-03-22-iqSkole.csv).

Očito je da ćemo bazu snimiti na svoj računar na kojem radimo. Međutim, uvijek se preporučuje da imamo arhivirane i rezervne kopije na drugim lokacijama budući da u istraživanja ulažemo dosta vremena i napora. Cloud servisi (od kojih su najpopularniji Google Drive, Microsoft OneDrive, Dropbox) su prvi izbor, a preporučivo je da sve to još jednom snimimo na neki prenosivi, arhivski medij koji koristite kao sigurnosnu kopiju. I ovdje jedna kratka napomena o etičkim aspektima o kojima će biti više riječi kasnije kada budemo govorili o objavljivanju podataka. Ako vaša baza sadrži neke osjetljive, intimne podatke koji bi se u teoriji mogli zloupotrijebiti, neophodno je da baza bude anonimizovana na način da se podaci ne mogu povezati sa ispitanicima ni u slabo vjerovatnom scenariju kada u bazi postoje neki indirektni identifikatori. Na primjer, na studiju psihologije na univerzitetu na kojem radim su rijetki muški studenti duge kose, te ako se podaci o polu i dužini kose pojave u bazi, potencijalno lako (ako neko baš hoće to da uradi) je moguće utvrditi identitet ispitanika bez obzira što mu ime nije navedeno u bazi. Dakle, ukoliko postoji ikakva prijetnja da bi u budućnosti podaci mogli biti zloupotrijebljeni, morate se dodatno raspitati o sigurnosnim mjerama kako oni ne bi bili dospjeli u javnost nakon nekog hakovanja vašeg računara, cloud naloga ili gubitka prenosivog diska.

Kako unosimo podatke u bazu?

Kako će podaci biti uneseni u bazu ovisi od načina registrovanja podataka. U sve većem broju istraživanja podatke jednostavno uvozimo direktno u bazu ako smo se za prikupljanje podataka služili nekim softverom (npr. onlajn anketni servisi, aplikacije na telefonima za vođenje dnevničkih istraživanja, namjenski softver za provođenje psihološkog eksperimenta). Kada koristimo papir-olovka format unos podataka može biti daleko brži ako smo sretnici koji imaju uređaj za optičko prepoznavanje znakova gdje nam softver automatski prepoznaje odgovore ispitanika na osnovu skeniranog papirnog primjerka. Ali, ova varijanta je relativno rijetka za istraživače početnike. Budući da je priku-

pljanje podataka putem papir-olovka instrumenata još uvijek čest pristup za studente psihologije koji zahtijeva "ukucavanje" podataka, potrebno je da dodatno opišem dobre prakse takvog unosa stečene mnogogodišnjim iskustvom.

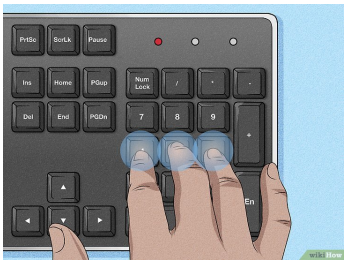
Prije nego što počnemo sa ukucavanjem podataka sa papira u računar, moraćemo da uradimo još nešto. Sjećate se da smo rekli da bi se u bazi prva definisana varijabla odnosila na redni broj ispitanika? Sada je potrebno da olovkom označimo svaki papirni primjerak rednim brojem, a onda ćemo te redne brojeve i unositi u bazu kako budemo unosili sve podatke. Ovo je neophodno kako bismo naknadno mogli korigovati tehničke greške prilikom unosa, a pomoći će nam i da se ne "pogubimo" pri unosu podataka u pogledu toga do kojeg papira smo stigli. Sebi ćemo dodatno olakšati rad ukoliko fiksiramo prvi red u kojem se nalaze imena varijabli kako bi nam uvijek bilo vidljivo o kojoj varijabli se radi. Za verziju LibreOffice Calc-a koju koristim tokom pisanja knjige (7.3) to se postiže biranjem opcije na padajućem meniju View › Freeze Cells › Freeze First Row, a za softver koji vi koristite ćete lako to pronaći internet pretragom.

I onda krećemo - u prvi sljedeći red unosimo slijeva na desno sve podatke koje sebi čitamo sa papira, konsultujući se sa šifarnikom u pogledu kodova. Ukoliko podatak nedostaje jer je, na primjer ispitanik preskočio pitanje, polje rezervisano za odgovor ćemo ostaviti praznim (postoje neke druge mogućnosti koje trenutno nisu bitne, ali pazite: gotovo nikada nećemo upisivati brojku 0 da označimo nedostajuću vrijednost!). Ukoliko naiđemo na nemoguće vrijednosti - npr. ispitanik se "pravi pametan" i zaokružuje dva odgovora na neko pitanje (2 i 3) ili upisuje nejasnu brojku - onda ćemo da kreiramo još jednu varijablu na kraju baze koju ćemo nazvati *napomene*, te ćemo opisati o čemu se tu radilo, a polje gdje smo trebali upisati vrijednost ćemo ostaviti prazno. Naknadno ćemo moći da donesemo informisaniju odluku; u najvećem broju slučajeva ćemo odgovor tretirati kao nedostajući podatak, a na osnovu specijalizovanih statističkih tehnika možemo taj nedostajući podatak zamijeniti najvjerovatnijim odgovorom.

Za pitanja otvorenog tipa ćemo u bazu unositi odgovore ispitanika onako kako su ih oni formulisali koliko god nam oni djelovali suludo - a ukoliko ste vi ujedno i glavni istraživač možete da parafrazirate odgovor ako bi kraća formulacija dosljedno odražavala ono što je ispitanik htio da navede. Ako smo već u softveru definisali da koristimo neki znak za razgraničenje odgovora za otvorena pitanja (npr. dvostruke navodnike " ili jednostruke navodnike ') moguće je to koristiti i pri unosu

(provjeriti da li je potrebno ovisno o softveru), pa bismo umjesto *Banja Luka* u dato polje upisali odgovor *"Banja Luka"* ili *'Banja Luka'*. Kao što sam već ranije pomenuo, ako namjeravamo da podatke iz tog pitanja kasnije kvantitativno obradimo izvodeći nove kategorije na osnovu kvalitativnog odgovora, onda ćemo u nekoj kasnijoj fazi kreirati dodatnu varijablu u bazi.

Unošenje numeričkih podataka će iziskivati manje napora, pogotovo zahvaljujući posebnoj numeričkoj tastaturi koja je i osmišljena da bismo mi (dešnjaci) mogli brže unositi podatke. Palcem ćemo pritiskati desnu strelicu koja nas vodi na sljedeće polje, a kažiprst, srednji prst i domali prst će dobiti novu životnu ulogu - njima ćemo unositi podatke bez gledanja u tastaturu (Slika 3.5).



Slika 3.6: Ilustracija preuzeta sa stranice https://www.wikihow.com/Ten-Key

Brzina i tačnost unošenja podataka je vještina koja se vježba i koja vam - uprkos sve većem oslanjanju na direktni uvoz podataka iz softvera - još uvijek može donijeti neki studentski džeparac (meni je taj posao finansirao odlaske na neke muzičke festivale).

S druge strane, odnos brzine i tačnosti unosa je obično negativan. Tehničke greške u unosu nastaju usljed žurbe da što prije dođemo do faze obrade podataka i / ili usljed kognitivnog zamora. Zato je u istraživanjima visokog uloga potrebno da dvije osobe paralelno unose iste podatke - to je ono što mi na našem studijskom programu radimo za svaki prijemni ispit. Nakon što se podaci na takav način unesu uporedićemo baze i za polja koja nisu saglasna provjerićemo šta je zapravo ispitanik odgovorio i korigovati problematični unos. Za sva ostala istraživanja je bitno da smo odmorni, pravimo pauze i da ako već konzumiramo psihoaktivne supstance to bi trebalo da budu one koje povećavaju pažnju i koje su legalne (dakle, mislim na kofeinske napitke). Bazu je preporučivo da snimamo manuelno nakon svakog unesenog reda, a također je mudro da omogućimo opciju automatskog snimanja svakih nekoliko minuta. Tako se nećemo naći u situaciji da proklinjemo sudbinu nakon iznenadnog nestanka struje ili zabagovanog programa.

Nažalost, u trenutku kada smo završili sa unosom podataka ili kada smo preuzeli gotovu bazu podataka iz softvera (i unijeli imena za našu bazu) nismo još ni blizu stigli do faze statističke obrade. Ono što u tom trenutku imamo su tzv. sirovi podaci (engl. *raw data*). Ovu bazu je potrebno da snimimo i zauvijek sačuvamo kao takvu, a da njene nove verzije prilagodimo za obradu. Tu sljedeću fazu najčešće nazivamo fazom uređivanja ili pospremanja baze podataka (engl. *data preparation*, *data cleaning* ili *data wrangling*).

Kako se osigurava da imamo kvalitetne podatke potrebne za obradu?

Relevantne postupke možemo podijeliti u tri grupe. Prvoj pripadaju postupci čija je svrha da osiguramo da su sami podaci kvalitetni, a u njih ubrajamo: provjeru ispravnosti i eventualnu korekciju već unesenih podataka, te evidentiranje i tretman sumnjivih obrazaca, štrčećih vrijednosti i nedostajućih podataka. Drugoj grupi pripadaju postupci dopunjavanja postojeće baze novim podacima u šta ubrajamo rekodiranje podataka, matematičke transformacije postojećih varijabli, izračunavanje novih varijabli iz više postojećih, te definisanje dodatnih obilježja varijabli. Treću grupu čine postupci restrukturisanja baza što podrazumijeva razdvajanje i spajanje varijabli ili čitavih baza, te konvertovanja oblika baze između širokog i dugačkog formata.

Descriptives

	godine	fakultet	burnout	dom4_anx
N	772	772	772	772
Missing	0	0	0	0
Minimum	18		0	0
Maximum	35		12	4

Slika 3.7: Pregled empirijski dobijenih vrijednosti za dimenzione podatke (softver Jamovi 2.3.12.0).

Frequencies of fakultet

fakultet	Counts	% of Total	Cumulative %
aggf	65	8.4%	8.4%
bez	29	3.8%	12.2%
eko	73	9.5%	21.6%
etf	79	10.2%	31.9%
fil	74	9.6%	41.5%
fpn	34	4.4%	45.9%
jez	63	8.2%	54.0%
mas	38	4.9%	58.9%
med	146	18.9%	77.8%
pmf	81	10.5%	88.3%
polj	21	2.7%	91.1%
pra	19	2.5%	93.5%
teh	30	3.9%	97.4%
umj	20	2.6%	100.0%

Slika 3.8: Pregled empirijski dobijenih vrijednosti za kategoričke podatke (softver Jamovi 2.3.12.0).

Kako da provjerimo da li su u bazu uneseni valjani podaci?

Prvo što ćemo željeti da uradimo jeste da provjerimo da li je tokom unosa podataka došlo do tehničkih grešaka. Ako smo podatke već prikupljali u digitalnoj formi, npr. putem onlajn upitnika i testova, onda greške pri unosu možemo da predupredimo tako što ćemo onemogućiti ispitanike da upisuju teoretski nedopustive vrijednosti. Međutim, tokom manuelnog unosa su česte greške tehničke prirode, jer nam senzorno-motorička veza zakazuje, pa u brzini umjesto vrijednosti 5 možemo ukucati 55 ili 54 (npr. ako je vrijednost upisana za sljedeću varijablu bila 4). Ono što moramo da uradimo jeste da pregledamo unos svih varijabli što se lako izvodi u svakom statističkom softveru. Potrebno je samo da pronađemo komandu koja omogućava da za numeričke varijable dobijemo minimalne i maksimalne empirijske vrijednosti, a za kategoričke varijable pregled svih biranih kategorija.

Slike 3.6 i 3.7 predstavljaju primjere takvih ispisa. Ukoliko uočimo neku "nevaljalu" vrijednost, potrebno je da lociramo u matrici podataka i korigujemo je konsultujući izvor odakle smo je unijeli. Postupak provjere je nakon korigovanja potrebno ponavljati sve dok ne budemo sigurni da sve varijable u bazi imaju tehnički opravdane vrijednosti. Ali, ovdje nećemo stati sa provjerom valjanosti podataka, jer je sasvim moguće da su neki podaci bili kvarni u ranijoj fazi, onda kada su nastajali.

Kao što možete i pretpostaviti, u psihološkim istraživanjima

nam je od ključnog značaja koliko su ispitanici motivisani da sarađuju, odnosno da pristupe ispitivanju iskreno i sa punom pažnjom. Uvijek će biti onih ispitanika koji su lijeni da rade sami i, ukoliko im to situacija dopušta, prepisaće odgovore drugih ispitanika. Zato je za početak potrebno provjeriti postoje li **duplirani nizovi odgovora**. Kada dobijemo dva ista obrasca odgovora na nizu od 200 varijabli gdje je ispitanik za svako pitanje imao pet ponuđenih opcija onda se veoma, veoma vjerovatno radilo o nekom obliku prepisivanju - ili možda o pogrešno iskopiranim podacima ili o nekom tehničkom problemu sa softverom gdje je softver dva puta zabilježio odgovore istog ispitanika. U takvim slučajevima je jasno da je potrebno barem jednog od ispitanika - ako ne i oba - izbrisati iz baze. S druge strane, očito je da će mnogo ispitanika u bazi imati identične odgovore na tri dihotomne varijable, te se postavlja pitanje koliko varijabli, te kakvih to varijabli, treba da čini niz pomoću kojeg detektujemo duplirane odgovore. Ne postoji jednoznačan odgovor, te ćete tek sa iskustvom steći jasniji osjećaj za to šta je slabo vjerovatno. Treba napomenuti da je moguće i izračunati neku konkretnu, objektivnu vjerovatnoću, ali ćemo raspravu o mehanizmu i potencijalnoj koristi ostaviti za neki napredniji susret. U svakom slučaju, ukoliko u bateriji instrumenata imamo više kompozitnih instrumenata i uočimo da dva ili više ispitanika imaju identične obrasce odgovora, trebalo bi da nam zazvoni alarm i da krenemo u pažljivu provjeru ostatka odgovora za takve ispitanike i prosudimo koliko je razumno očekivati da se dese identični obrasci.

Nadalje, nemotivisanost ispitanika ili slabo razumijevanje instrukcija može da dovede do nelogičnih obrazaca podataka kojeg pokazuju samo jedan ili manji broj pojedinaca. Na primjer, nije rijetkost da nam ispitanici žele podvaliti upitnik na kojem su iz pitanja u pitanje odabirali samo jedan odgovor (npr. brojku 5) ili da su pratili vizuelni cik-cak obrazac iako je iz sadržine pitanja jasno da je to skoro nevjerovatno da se desi ukoliko su savjesno odgovarali na pitanja. Za naprednije statističke softvere[3] su kreirani algoritmi koji detektuju broj uzastopno identičnih odgovora u nizu ili nelogične obrasce odgovora (npr. davanje suprotnih odgovora na značenjski sinonimna pitanja ili davanje identičnih odgovora na značenjski antonimna pitanja). Mora se napomenuti da i kada se koriste takve tehnike ponovo je istraživač taj koji mora da sam prosudi o vjerovatnosti odgovora, jer takve tehnike nisu savršene i mogu da lažno označe nemotivisanost odgovaranja. Praksa koja se pokazala kao relativno dobra u detekciji nemotivisanog odgovaranja jeste da

[3] Na primjer, u R-u te funkcije obavlja paket careless https://www.ryentes.com/careless/intro.html

se u samim istraživačkim instrumentima traži od ispitanika da bezuslovno označe određeni odgovor (npr. "Za ovo pitanje kao odgovor upišite broj 3.", kao i da na kraju ispitivanja odgovore na pitanje koliko su iskreno i pažljivo učestvovali u istraživanju. Uz nemotivisano ponašanje u istraživačkoj situaciji, problem često predstavlja i motivisano odgovaranje u smislu da ispitanici mogu imati određenu agendu, pa distorziraju odgovore kako bi ostavili određen utisak koji bi za njih imao poželjan ishod (npr. žele da se predstave kao izrazito fini ili veoma patologizirani). No, sve ove teme su dominantno psihometrijske, te se nećemo stići baviti njima u ovom tekstu.

Postoje ipak još dvije prijetnje valjanosti podataka na koje "čista" statistika obraća veliku pažnju, a to su problemi nedostajućih podataka i stršećih / ekstremnih / netipičnih mjera. Za oba problema je razvijen sijaset procedura kako za njihovo dijagnostikovanje, tako i za njihov tretman pri analizi. Budući da je za pametne odluke potrebno da posjedujete statistička znanja koja još uvijek nemate, na ovom mjestu dajem samo prikaz apsolutno najosnovnijih koraka kada su u pitanju nedostajući podaci, a detaljnije manevre za napast stršećih mjera ostavljam za kasnije.

Postupci tretmana nedostajućih vrijednosti su ogromna tema i predmet su zasebnih knjiga, te zainteresovani za njih mogu dobiti mnogo informacija konsultujući ih (npr. https://www.appliedmissingdata.com/). Osnovne informacije se mogu naći na jednostavnijim internet stranicama (npr. https://www.statswithr.com/foundational-statistics/handling-missing-data-in-statistics).

Dakle, još kada smo pregledavali ispravnost vrijednosti podataka vjerovatno smo u softveru dobili i podatak o broj nedostajućih podataka za svaku varijablu (vidjeti raniju Sliku 3.4a). Standardna oznaka u R-u za nedostajuće vrijednosti je *NA* (od engl. *not available*), te je softver i ne pitajući nas u međuvremenu pretvorio sva ona pri upisu prazno ostavljena polja u NA. Izuzev u slučaju da ste istraživanje radili putem onlajn formulara gdje je moguće ispitanika obavezati da mora dati odgovor na svaku stavku, veoma je vjerovatno da ćete u tabelarnom ispisu pronaći neke varijable koje imaju nedostajuće podatke. Veći broj nedostajućih podataka nam sugeriše da je stavka potencijalno problematična usljed svoje sadržine (npr. intimna pitanja, nerazumljiva pitanja), te je potrebno razmisliti o tome koliko uopšte možemo vjerovati podacima dobijenim za tu varijablu. Međutim, u ovaj fazi nas mnogo više interesuje koliko imamo nedostajućih podataka po redovima, odnosno ispitanicima. U softveru se taj podatak dobija kreiranjem nove varijable koja sadrži broj nedostajućih vrijednosti za svakog ispitanika. Veliki broj nedostajućih vrijednosti po ispitaniku ukazuje na nevalidnost odgovora usljed nemotivisanosti i nepažnje, te je takve redove obično potrebno izbrisati.

Ali šta je to veliki broj nedostajućih vrijednosti? "Mudrost gomile" pominje da bi se moglo tolerisati do 5% ili 10% ne-

dostajućih odgovora koji bi kasnije bili dodatno tretirani. Pitanje je zapravo dosta kompleksnije jer tretman zavisi i od vrste instrumenta koji je korišten, kao i od nacrta istraživanja (npr. nije isto radi li se o neponovljenim ili ponovljenim mjerenjima). Na primjer, ako smo u istraživanju zadali test sposobnosti ili test znanja, sva pitanja na koje ispitanik nije odgovorio predstavljaju zapravo netačne odgovore. Oni ne bi smjeli ulaziti u zbir nedostajućih vrijednosti, budući da ćemo za takve odgovore u naknadnom koraku uređivanja baze odrediti konkretne vrijednosti (npr. 0 ukoliko se radi o netačnom odgovoru koji ne dovodi do negativnih bodova na testu). Drugim riječima, prag dopustivog broja nedostajućih podataka je potrebno da posebno odredimo u svakom pojedinačnom istraživanju, a nakon toga ćemo odlučiti da li ispitanika brišemo iz baze podataka ili ga ostavljamo u njoj.

Na koje načine izvodimo nove varijable iz već postojećih?

Baza podataka sa – barem naizgled – valjanim podacima nas približava statističkoj analizi. Ipak, u najvećem broju slučajeva će nam još uvijek nedostajati neke od varijabli na koje smo računali, a koje se izvode iz varijabli za koje smo direktno prikupili podatke. Obično ćemo bazu podataka dopunjavati i prije početka analize, ali često ćemo i u toku obrade uočiti potrebu za novim varijablama. Slijedi opis tih postupaka.

Rekodiranje (engl. recoding) je termin pod kojim podrazumijevamo pretvaranje postojećih podataka u druge vrijednosti koristeći neko jednostavno pravilo zamjene koje određujemo kao istraživači. Najilustrativniji primjer su testovi znanja. Recimo da je kandidatima na prijemnom ispitu zadan test iz psihologije. U bazu su uneseni odgovori na pitanja onako kako su kandidati odgovarali, birajući opcije od 1 do 4. Ali, zadaci imaju različita tačna rješenja, npr. za 1. zadatak je to opcija 3, za 2. zadatak 1 itd. Očito je da iz podataka unesenih u bazu ne možemo odmah da dođemo do ukupnog skora. Moramo da načinimo međukorak, odnosno da kreiramo skup novih varijabli koje će iskazivati da li je kandidat tačno ili pogrešno odgovorio na pojedinačno pitanje, a gdje ćemo i nedostajuće podatke klasifikovati kao pogrešne odgovore. Drugim riječima, za varijablu *zad_01* ćemo samo onim kandidatima koji su odabrali opciju 3 dodijeliti vrijednost 1, a za ostale kandidate će to biti vrijednost 0 u novoj varijabli koju ćemo imenovati *bod_01* (usput, podrazumijevamo da na prijemnom ispitu nije dato pravilo da se za netačne odgovore dodjeljuju negativni bodovi). Za *zad_02* radimo to isto, ali imajući u vidu

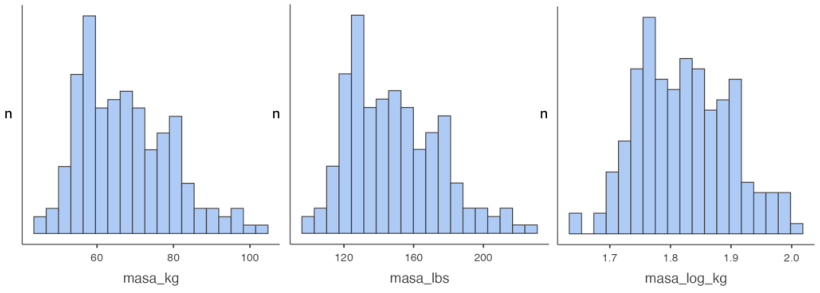
da je tačan odgovor bila samo opcija 1. Da bismo varijablu korektno i transparentno rekodirali, neophodno je da napravimo **ključ skorovanja** (engl. *scoring key*) ili kodnu šemu koju za sve rekodirane varijable dostavljamo ili u sklopu šifarnika ili kao posebnu datoteku iz čega će biti jasno šta smo to uradili da bismo došli do novih varijabli.

Rekodiranje vrlo često koristimo za inverzno sročene tvrdnje u upitnicima. Na primjer, u jednom inventaru ličnosti se procjenjuje ekstraverzija i tri tvrdnje na koje ispitanik odgovora birajući između odgovora DA i NE - koje su u bazu unošene kao vrijednosti 1 i 0 - su: "Da li mnogo volite da izlazite u provod?", "Da li generalno radite ili govorite stvari, a da se ne zaustavite kako biste promislili o tome?" i "Da li ste uglavnom ćutljivi kada ste okruženi drugim ljudima?". Za razliku od prve dvije tvrdnje gdje odgovor DA govori u prilog veće ekstravertiranosti, za treću tvrdnju je to odgovor NE. Iz tog razloga je na osnovu postojeće varijable *ext_3* potrebno napraviti novu varijablu, koju ćemo nazvati *ext_3_r*, gdje će vrijednosti 1 i 0 biti obrnute kako bismo prostim sabiranjem tri varijable mogli dobiti skor na skali ekstraverzije (dakle, formula će biti : *ekstraverzija = ext1 + ext_2 + ext_3_r*).

Još jedan čest povod za rekodiranje je potreba da neke dimenzione varijable kategorišemo ili da napravimo novu kategorizaciju već postojećih kategoričkih varijabli. Na primjer, ukoliko smo u istraživanju za školski uspjeh tražili da se upiše prosječna ocjena izražena na dvije decimale (npr. 4.22, 4.55) mi iz nje možemo kreirati kategoričku ordinalnu varijablu školski uspjeh koja će ispitanike kategorisati na osnovu pragova (npr. ocjena $\geq$ 3.50 -> 4, ocjena $\geq$ 4.50 -> 5). Ili ukoliko tokom analize shvatimo da je za varijablu *trenutni_bračni_status* dovoljno da razlikujemo ispitanike koje su u dužoj romantičnoj vezi onda bismo inicijalne kategorije "U dužoj vanbračnoj vezi" i "U braku" mogli spojiti u jednu kategoriju, a varijable "Udovac/ica", "Razveden/a", "U kraćoj vezi", "Bez stalnog partnera" spojiti u drugu kategoriju.

Treba primijetiti da se rekodiranje u bazi može izvršiti na dva načina. Može se kreirati nova varijabla, ali se može izvesti i prosta zamjena vrijednosti unutar već postojeće varijable. Međutim, druga opcija gotovo nikada nije preporučiva. Zašto? Pod jedan, mogu se desiti problemi u obradi usljed toga što dva ili više puta pokrenemo istu komandu rekodiranja, pa više ni sami ne znamo da li se radi o početnim ili modifikovanim vrijednostima. Pod dva, moguće je da će nam naknadno u analizi zatrebati početna varijabla, a ona nakon takvog rekodiranja više nije dostupna.

Transformacija (engl. transformation) varijabli, u užem smislu te riječi, podrazumijeva primjenu neke matematičke funkcije čime se ciljano mijenja mjerna skala postojeće varijable. Najčešće transformacije se postižu operacijama kvadriranja, korijenovanja, logaritmovanja, oduzimanja neke konstante od vrijednosti varijable, ali postoje i ezoteričnije varijante koje su dosta prisutne u psihologiji, kao što su, na primjer: standardizacija, normalizacija, konverzija u T-skorove, rangove i percentilne rangove. Upotrebom nekih transformacija će se promijeniti samo mjerna skala, ali se neće mijenjati odnosi među vrijednostima, a time ni izgled raspodjele vrijednosti. Takve transformacije nazivamo **linearnim transformacijama**. Na primjer, masa se može izraziti u kilogramima, ali i u funtama (lb). Ukoliko ste u istraživanju preuzeli podatke u kilogramima možete ih jednostavno transformisati u funte putem formule $mas_{lbs} = mas_{kg} * 2.2046$. Primjer nelinearnih transformacija bi bili rangovanje i logaritmovanje. Slika 3.8 prikazuje razliku efekata linearnih i nelinearnih transformacija. Na lijevom grafikonu su prikazane originalne vrijednosti tjelesne mase ispitanika koje su mjerenjem dobijane u kilogramima. Srednji grafikon prikazuje vrijednosti iskazane u funtama, dok su na desnom grafikonu date logaritmizovane kilogramske vrijednosti. Uporedite mjerne skale i oblike distribucija. Inače, o svrsi pojedinačnih oblika transformacija, njihovim prednostima i manama će biti više riječi kasnije u tekstu.



Slika 3.9: Poređenje efekata linearnih i nelinearnih transformacija.

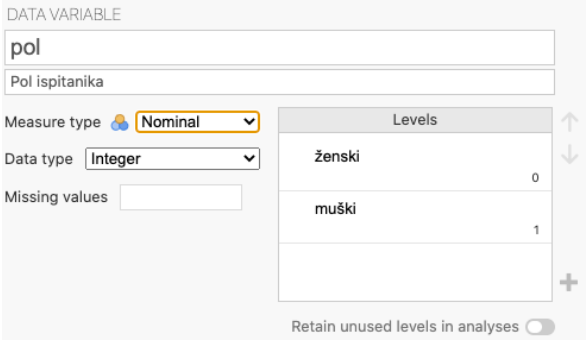
Izračunavanje novih varijabli (engl. *computing new variables*) na osnovu više drugih varijabli je posebna rabota. Kao što je već naglašeno, u psihologiji izuzetan značaj imaju kompozitni instrumenti koji sadrže više od jedne stavke kako bi se procijenila izraženost nekog svojstva. Kompozitne skorova najčešće dobijamo tako što izračunamo sumu za sve pripadajuće stavke (npr. iz ranijeg primjera *ekstraverzija = ext1 + ext_2 + ext_3_r*) ili izračunavanjem aritmetičke sredine tako što se suma odgovora koji pripadaju varijabli dijeli sa brojem stavki(npr. *ekstraverzija*

= *(ext_1 + ext_2 + ext_3_r)/3*). Ako možemo da biramo, gotovo uvijek treba da izaberemo aritmetičku sredinu, budući da ukoliko za nekog ispitanika imamo nedostajuću vrijednost na jednoj od stavki i to ne tretiramo adekvatno, sumacioni skor će svakako potcijeniti poziciju datog ispitanika što se ne bi desilo sa aritmetičkom sredinom budući da softver automatski koriguje broj valjanih stavki za svakog ispitanika. Iako su ova dva oblika izračunavanja novih varijabli zaista najčešća, u psihologiji postoji veliki broj statističkih procedura putem kojih se izvode skorovi za kompozitne mjere. Takve procedure mogu da koriste kompleksne formule u kojima se svaka varijabla različito vrednuje pri izračunavanju ukupnog skora.

Sljedeći postupak takođe može biti neophodan za početak obrade u softveru, ali za razliku od dosad opisanih procedura kojima se u bazu dodaju nove varijable, njime se samo dodaje **opis vrste varijable**. Drugim riječima ova modifikacija neće biti vezana za samu bazu podataka, nego za to kako bi određeni softver trebalo da tumači svaku varijablu, a tumačenje se odnosi na razlikovanje tipova varijabli. Podjela tipova varijabli u pogledu toga kako ih softver tretira stoji u bliskoj vezi, ali nije identična, sa ranijom podjelom varijabli na nivoe mjerenja. Tim prije što dodjeljivanja tipa varijabli ovisi o softverskom programu u kojem se planira izvršiti obrada; različiti softverski programi imaju različite sisteme klasifikacije (vidjeti Sliku 3.9 kao primjer).

Srećom, postoje i neke sličnosti. Za početak svi softveri razlikuju **tekstualne ili znakovne varijable** (engl. *string*, *character* ili *text*) od **numeričkih ili skalarnih varijabli** (engl. *numeric* ili *scale*). Nešto kompleksnija je situacija kada su u pitanju kategoričke varijable, koje softver, po pravilu, naziva **faktori** (engl. *factor*). Obično se razlikuju **nominalne kategoričke varijable** (engl. *unordered factor*, *nominal*) od **ordinalnih kategoričkih varijabli** (engl. *ordered factor*). Za kategorije kategoričkih varijabli je moguće dodati i njihove **opisne nazive** (engl. *labels*) kako bi ispis rezultata bio razumljiviji za onog ko obrađuje podatke, jer je lakše interpretirati tabelu sa rezultatima ukoliko u njoj stoje nazivi regija *Zapadni Balkan*, *Centralna Evropa* i *Skandinavija*, nego *1*, *2* i *3*. Dihotomne varijable su teže ukrotiv slučaj: mogu se definisati ili kao numeričke varijable ili kao faktori, te čak i kao tzv. **logičke varijable** (engl. *logical variables*) koje imaju samo predefinisane vrijednosti *Tačno* i *Netačno*. Ponekad ćemo za iste podatke morati napraviti dvije različite varijable sa različitim odrednicama (npr. *pol_kat*, *pol_dim*) kako bismo neke rigidne softvere mogli nadmudriti i poduzeti sve analize koje smo planirali. **Datumi** (engl. *date*) i drugi vremenski podaci su posebna vrsta varijabli

koju ćemo ponekad sretati u psihološkim istraživanjima. U svakom slučaju, od nas će se tražiti da u softveru dodijelimo odgovorajući opis svakoj od varijabli, a na Slici 3.9 vidite primjer kako se to može definisati u jednom softverskom rješenju.



Slika 3.10: Definisanje opisnih vrijednosti za varijablu pol (softver Jamovi 2.3.12.0).

Kako mijenjamo strukturu postojećih baza podataka?

Uz sve navedeno, u nekim slučajevima će biti potrebno da napravimo još neke strukturne promjene baze prije nego li bude sve spremno za obradu. Na primjer, recimo da smo u jednom istraživanju skupljali podatke i putem softvera i putem papir-olovka formata. Ukoliko smo imali iste stavke na obe vrste mjerenja, a različiti formati su nam bili potrebni da prikupimo više ispitanika, onda će biti neophodno da spojimo dvije baze, odnosno da u jednu od njih dodamo redove iz druge. Poželjno je da to uradimo pomoću specijalizovanih komandi unutar softvera za statističku obradu, ali istu stvar možemo postići i prostom *copy+paste* metodom u tabelarnom softveru. Niže je dat ekstremno jednostavan primjer R komande čime se objedinjuju tri baze od kojih svaka ima po 60 ispitanika i identično definisane varijable, a gdje broj u imenu baza označava početni redni broj ispitanika u tom istraživanju:

```
podaci_svi <- rbind.data.frame(podaci_01, podaci_61,
↪  podaci_121)
```

Nema potrebe da se plašite R komandi koje slijede, a niti da ih pamtite. Lijepo opišite nekom od AI četbotova vašu bazu i zamolite ga da vam ili navede prikladne komande ili da čak i obavi restrukturisanje za vas (pazite samo na to da se u bazi koju dijelite sa njim ne nalaze podaci koji bi mogli da naruše privatnost vaših ispitanika).

Nešto komplikovaniji je slučaj kada ne dodajemo nove ispitanike nego varijable, recimo onda kada nam je softver generisao bazu koja sadrži vrijeme reakcija na neki stimulus, a mi smo dodatni zadali papir-olovka upitnik kako bismo prikupili još neke podatke o ispitanicima. Tada je vrlo bitno da imamo neki identifikator za svakog ispitanika (npr. redni broj ili šifru) koji ćemo unijeti

u obje baze, a kako bismo odgovore iz njih mogli kasnije spariti. Napredniji softveri za statističku obradu obično imaju definisane komande kako bismo to postigli. Niže je predstavljane kako to jednostavno izgleda u R-u kada dvije baze treba da spojimo na osnovu rednog broja ispitanika (varijabla *rb*).

```
podaci_svi <- merge(podaci_digitalno, podaci_papir,
↪  by="rb")
```

Opet, u nekim slučajevima će biti potrebno odabrati unaprijed određen broj entiteta iz baze kako bi se izvršila željena obrada. Recimo da je neko istraživanje izvedeno u 10 svjetskih regija, a nas interesuju samo podaci iz naše regije i to podaci dobijeni od osoba ženskog pola. Ono što je potrebno da uradimo jeste da "iščistimo" ili "isfilterišemo" podatke - da, softver to obično naziva filterisanjem (engl. *filter*) - tako da pri obradi koristimo samo one redove koji sadrže odgovorajuće vrijednosti. Većina point-and-click statističkih softvera je opremljena opcijom kojom možete da filterišete podatke unutar programa sa par klikova miša. U R-u bismo se koristili logičkim operatorima: & za "i", | za "ili", te ! za "nije" (npr. != je znak za "nije jednako"), a primjer bi mogao biti sljedeći gdje koristimo paket `tidyverse` i komandu `filter`:

```
library(tidyverse) # učitavanje potrebnog paketa za
↪  komandu filter
podaci_filterisani <- filter(podaci, region == 3 & pol
↪  == 2)
# definisanje nove baze podataka uz definiciju da
↪  region mora
# biti jednak 3 (Zapadni Balkan), a pol 2 (ženski)
```

U drugim slučajevima će biti potrebno da vrijednosti za već unesene varijable ili razdvojimo ili spojimo. Razmotrimo dva slučaja u vezi sa podacima o datumu rođenja osoba. Recimo da nas interesuje samo godina rođenja osobe, a da smo podatke o datumu rođenja prikupili putem jedne varijable datum_rodjenja u formi dd.mm.gggg. Vrijednosti je potrebno razdvojiti u tri varijable (dan, mjesec, godina), a to se u R-u postiže veoma lako komandom `separate` u paketu `tidyverse`:

```
library(tidyverse) # učitavanje R paketa za komandu
↪  separate
baza_podataka %>%
```

```
  separate(datum_rodjenja,
           into = c("dan", "mjesec", "godina"),
           sep = ".")
```

U drugom slučaju, pretpostavimo da smo podatke o datumu rođenja dobili u tri različite varijable i da nam je potrebno da ih spojimo u jednu varijablu, razdvojene separatorom - (minus) i sa redoslijedom godina-mjesec-dan (npr. datum bi bio 2011-10-18, a u tom formatu nam R paket *DescTools* komandom `Zodiac` omogućava da izvedemo astrološki znak osobe). Opet, vrlo jednostavno u R-u, korištenjem paketa `tidyverse` i komande `unite`:

```
library(tidyverse)
baza_podataka %>%
  unite(col = "datum_rodjenja",
        c("godina", "mjesec", "dan"),
        sep = "-")
```

Konačno, u nekim slučajevima će postojati potreba da podatke iz širokog formata, koji su zgodni za unos, pretvorimo u matricu dugačkog formata kako bismo mogli izvesti prikladne analize za ponovljena mjerenja. Na Slici 3.10 se vidi kako bi baza širokog formata (lijevo) za jedan raniji primjer trebalo da izgleda kao baza dugačkog formata (desno).

	A	B	C	D	E	F	G
1	**rb**	**ime_studenta**	**pol**	**studij**	**strah_1**	**strah_2**	**strah_3**
2	1	Jasna	1	psihologija	8	6	3
3	2	Janko	2	pedagogija	7	5	5

	A	B	C	D	E	F
1	**rb**	**ime_studenta**	**pol**	**studij**	**strah_statistika**	**mjerenje**
2	1	Jasna	1	psihologija	8	1
3	1	Jasna	1	psihologija	6	2
4	1	Jasna	1	psihologija	3	3
5	2	Janko	2	pedagogija	7	1
6	2	Janko	2	pedagogija	5	2
7	2	Janko	2	pedagogija	5	3

Slika 3.11: Baza širokog formata (gore) restrukturisana u dugački format (dole)

Ovakvo restrukturisanje se može vrlo elegantno izvesti u R-u korištenjem `tidyverse` paketa i komande `pivot_longer` (naravno, postoji i inverzna komanda `pivot_wider`):

```
library(tidyverse)
baza_podataka %>%
  pivot_longer(c(strah_1, strah_2, strah_3),
```

```
  names_to = "mjerenje", names_prefix = "strah_",
  values_to = "strah_statistika")
# komanda kojom se vrijednosti za tri varijable
↪  prebacuju u
# jednu novu varijablu strah_statistika, dok nova
# varijabla mjerenje označava vrijeme mjerenja
```

Sve što smo uradili od trenutka kada smo imali sirovu bazu podataka do trenutka kada imamo pospremljenu bazu podataka, bi trebalo da dokumentujemo u protokolu obrade. Naše kolege iz istraživačkog tima, ali i drugi zainteresovani istraživači mogu onda da provjere da li smo sve korektno obavili. Protokol sa komandama se može snimiti čak i kao prosti .txt fajl, ali je jednostavnije koristiti opcije koje pružaju specijalizovana radna okruženja za statistički softver. Neka od njih (npr. RStudio kao okruženje za R) imaju vrlo praktične pogodnosti, kao što je integrisanost sa sistemima za upravljanje verzijama koda. Takvi sistemi (npr. GitHub) omogućavaju da snimamo izmjene tokom rada, na način da se svi koraci registruju, te po potrebi - ako nešto zabrljamo u procesu - možemo da se vratimo na ranije stanje. Korištenje integrisanih okruženja je napredna tema i ovisi od softvera do softvera, pa ukoliko ste zainteresovani za to, treba da sami proučite opcije koje se nude.

Nakon što smo uspješno pripremili podatke za obradu, stigli smo do mjesta od kojeg možemo početi da slušamo i sagledavamo šta nam to podaci govore o istraživačkom pitanju. Zato u sljedećem poglavlju učimo kako se opisuje raspodjela jedne kategoričke varijable - što predstavlja najlogičniji prvi korak u statističkoj analizi. Prikazani postupci će vam omogućiti da razumijete ko su vaši ispitanici, kako su distribuirani po različitim kategorijama, te da uočite prve obrasce u podacima prije nego što pređete na složenije analize. Da bismo sve to postigli, potrebno je da se upoznate sa prikazivanjem obrađenih podataka putem tabela i grafikona, te kako da njih učinite što informativnijim.

Poglavlje 4

Opisivanje raspodjele jedne kategoričke varijable

Nakon čitanja ovog poglavlja znaćete:

- kako pristupiti obuhvatnoj analizi distribucije jedne kategoričke varijable,
- kako se grafički, tabelarno i tekstualno prikazuju rezultati analize jedne kategoričke varijable, uz poštovanja usvojenih akademskih standarda,
- da donesete odluku o daljem postupanju sa slabo zastupljenim kategorijama.

Zamislite da ste upravo završili postupak prikupljanja podataka putem onlajn upitnika i da u bazi imate podatke za 300 ispitanika. Prirodno je da sebi prvo postavite pitanje: "Ko su zapravo ti ljudi?" Prije nego što ćemo se upustiti u razmatranje povezanosti među varijablama, moramo da razumijemo osnovnu strukturu našeg uzorka. Ovo poglavlje će vas naučiti kako da "upoznate" svoje ispitanike kroz analizu kategoričkih varijabli - one koje opisuju ko su, odakle dolaze, šta rade i kako se izjašnjavaju o različitim pitanjima.

Analiza kategoričkih varijabli je logičan uvod u statističku analizu jer nam omogućava da steknemo osjećaj za to s kim imamo posla u istraživanju. Nakon što sagledamo kakvog su pola, obrazovanja i osnovnih stavova naši ispitanici, možemo planirati naprednije analize i procijeniti koliko naši rezultati mogu biti reprezentativni. Zamislite da istražujete efikasnost kognitivno-bihejvioralne terapije za depresiju, a otkrivate da vam je 85% uzorka ženskog pola i da većina ima visoko obrazovanje. Ovi nalazi utiču na to kako ćete interpretirati rezultate i na koju populaciju možete da ih primjenite. Uz to, nekad će analize jedne kategoričke varijable biti suštinska analiza, na primjer onda

kada želimo da procijenimo rasprostranjenost (tj. prevalenciju) određenog poremećaja (npr. kliničke depresivnosti). Dakle, postupci koje predstavljam nisu statistička formalnost, nego su dobar temelj za sve ono što ćemo raditi kasnije.

Međutim, prije nego što se konačno bacimo na obradu podataka, dobro bi bilo da se upoznamo sa opštim načelima prikazivanja statističkih rezultata koji važe i za sve ostale analize.

Kako sebi i drugima predstavljamo rezultate statističke analize?

Grubo rečeno, prikazivanje obrađenih podataka ima dvije različite svrhe: **eksplorativnu** i **prezentacionu**. Naime, na početku istraživačkog procesa mi želimo da steknemo što bolji uvid u prirodu podataka koje smo prikupili na uzorku. Konkretno, u toj eksplorativnoj fazi najprije želimo da ustanovimo kakva nam je distribucija podataka i da potvrdimo da podaci ispunjavaju uslove za namjeravanu kasniju analizu. Nakon toga ćemo htjeti da neformalno provjerimo svoje početne hipoteze i eventualno postavimo neke nove hipoteze iz samih podataka. U takvom kontekstu, kada smo sa podacima u toj intimnoj atmosferi, takoreći u svom donjem vešu za računarom, ne moramo da posebno brinemo o estetici i formi u kojoj ćemo izložiti rezultate. Bitno nam je samo da tabelarni i grafički ispis kojeg softver izbacuje nudi priliku za neke dobre uvide.

S druge strane, kada je u pitanju prezentaciona faza svjesni smo da ćemo rezultate prikazivati drugima koji su manje upućeni u naše istraživanje. Iz tog razloga je vrlo bitno da koristimo što prikladnije forme izlaganja rezultata kako bi ti drugi ljudi što brže obradili informacije i bili ubijeđeni u iste zaključke do kojih smo i mi došli nakon napornih seansi analize. Dakle, u toj fazi se trudimo da informacije prevedemo sa statističkog na svakodnevni jezik.

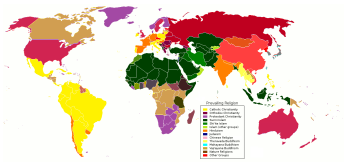
Rezultati statističke obrade se obično predstavljaju na tri načina: grafički, tabelarno i narativno (tekstualno, unutar rečenica).[1] Svaki od vidova prikazivanja ima svoje prednosti i mane koje treba da imamo na umu kada odlučujemo kako da predstavimo rezultate.

[1] Postoji i mogućnost dinamičkog prikaza pomoću animacija (tj. izmjene grafikona tokom vremena), ali time se ovdje ne bavimo.

Tabele su vrlo vrijedno sredstvo prikazivanja statističkih rezultata jer se njima na strukturisan način može prikazati veliki broj precizno izraženih numeričkih vrijednosti. Istovremeno, analiza velike količine brojeva unutar tabela nije najintuitivnija ljudska

aktivnost i brzina uvida se pogoršava što su tabele detaljnije i obimnije. Poznajem čak i ljude koji imaju svojevrsnu "tabelofobiju".

Upravo je najveća prednost **grafikona** mogućnost da brzo steknemo uvid u rezultate. Na primjer, na jednom grafikonu može biti izloženo na hiljade podataka koji nisu nekim statističkim postupkom uprosječeni, nego prikazuju rezultate pojedinačnih ispitanika - a mi bacajući jedan pogled možemo da uočimo trend koji je prisutan.



Slika 4.1: Mapa raspostranjenosti religija na svijetu; nekoliko milijardi ljudi predstavljeno na jednom grafikonu (commons.wikimedia.org).

Postoje naravno i mane grafikona. Kao prvo, grafikoni pružaju globalnu sliku, ali im nedostaje preciznost informacija koja je ponekad važna. Drugo, grafikonima smo često ograničeni na prikaz manjeg broja varijabli. Već kada su tri varijable u igri, stvari se komplikuju jer koristimo samo dvije dimenzije za prikaz.

Svrha **tekstualnog prikazivanja rezultata** je naglašavanje nalaza koji mogu biti, ali ne moraju, već prikazani u tabelama ili grafikonima. Narativom se prezentuju konkretni rezultati, ali se obično odgovara i na pitanja kao što su *Šta taj rezultat zapravo znači?*, *Zašto je taj rezultat bitan?*, *Koje implikacije imaju ovi rezultati?*. Ukoliko imamo samo 2 do 3 informacije koje treba da predstavimo (npr. podatak o raspodjeli pola ili o raspodjeli stručne spreme ispitanika koja je podijeljena u tri kategorije) onda je opis u jednoj rečenici dovoljno rješenje. Na primjer, "uzorkom je obuhvaćeno 200 ispitanika od kojih je 120, odnosno 60.0%, bilo muškog pola". S druge strane, tekst je slabo rješenje za predstavljanje većeg broja informacija jer se nizom rečenica "zagušuje smisao" nalaza. U takvim situacijama je potrebno koristiti tabele i grafikone koji na ograničenom prostoru prikazuju mnogo informacija.

Šta se prikazuje u analizama distribucije jedne kategoričke varijable?

Analiza distribucije jedne kategoričke varijable podrazumijeva prikazivanje apsolutnih i relativnih frekvencija za svaku od kategorija. Terminom **frekvencija** ili **učestalost** (*engl. frequency*) nazivaćemo broj članova nekog skupa koji imaju određeni vid atributa/svojstva ili broj javljanja nekog događaja. Konkretnije, razlikovaćemo **apsolutne frekvencije** koje se dobijaju prostim prebrojavanjem (npr. **4** studenta su imala ocjenu 10 iz statistike, a njih **20** ocjenu 9) i **relativne frekvencije** u koje spadaju **postoci / procenti** i **statističke proporcije**. Relativne frekvencije se tako

nazivaju jer uvijek ukazuju na *udio* određene kategorije unutar cjeline, a u odnosu na univerzalno definisani referentni sistem koji omogućuje poređenje različitih varijabli. Taj referentni sistem je za procente brojčana skala od 0 do 100, a za proporcije skala od 0 do 1.

Procenti (oznaka %) (engl. *percentages*) su vam sigurno poznatiji, jer su popularniji u medijskoj i svakodnevnoj komunikaciji. Predstavljaju udio date kategorije unutar ukupne cjeline izražen u stotim dijelovima. Na primjer, ako imamo 200 ispitanika od kojih je 120 muškaraca, to znači da je 60.0% uzorka muškog pola. Ovakav nalaz bi mogao biti očekivan u istraživanju agresivnog ponašanja u sportu, gdje muškarci tradicionalno čine većinu učesnika, ali bi bio neočekivan u istraživanju poremećaja ishrane gdje žene obično čine veći dio uzorka. Naravno, procentima se mogu izraziti i količine. Ako su za neko ljetnje druženje kupljene četiri litre džina, a popijene su tri litre to znači da je popijeno 75% džina (nadam se da se razumijemo: misli se na alkoholno piće koje se često koristi u koktelima i dobija maceracijom bobica kleka, a ne na neko ogromno mitsko biće u fluidnom stanju).

Proporcije (engl. *proportions*) su, pjesnički rečeno, procenti dok su još mali, odnosno vrijednosti koje se dobijaju dijeljenjem frekvencije određene kategorije sa ukupnom frekvencijom prije nego što se to pomnoži sa 100.[2] Dakle, cjelinu označava broj 1 (odnosno 1.0), polovinu 0.5, četvrtinu 0.25, a desetinu 0.1. Proporcije su neobično važne u naprednijoj statistici i daleko se češće koriste pri analizama nego procenti, te se tek prije saopštavanja rezultata konvertuju u procente ukoliko se smatra da će publika lakše razumjeti procente.

[2] Ovdje je riječ o statističkoj proporciji. Vi vjerovatno poznajete riječ proporcija u njenom drugom matematičkom izdanju, gdje ona znači jednakost omjera i može da bude direktna ili obrnuta: https://gradivo.hr/matematika/matematika-online-skripta-za-1-razred/omjeri-postoci-apsolutna-vrijednost-i-aritmeticka-sredina ili https://www.opsteobrazovanje.in.rs/matematika/osnovna-skola/proporcija

Kada je u pitanju statistička notacija, frekvencija ispitanika u uzorku se najčešće označava malim latiničnim slovom n, mada postoje i autori koji upotrebljavaju f. Veliko latinično slovo N se obično rezerviše za veličinu populacije, koju ćemo malo kasnije definisati u tekstu. Kada se želi navesti koliko je velika neka podgrupa (npr. broj muškaraca od ukupnog broja ispitanika) onda je najbolje upotrijebiti i indeks oznaku (npr. m za muško ili 1 ukoliko se ona smatra prvom grupom) pa bi se broj muškaraca mogao navesti kao $n_m = 120$ ili $n_1 = 120$ ili $f_m = 120$. Zgodno je ovdje napomenuti da je ova mnogostrukost izbora pri pisanju, odnosno nepostojanje saglasnosti, naznaka istine da su statističari međusobno svadljiva grupa ljudi. Notacija je neujednačena jer statističari vole da koriste po tri-četiri termina za jedan te isti pojam, što izluđuje. Moj zadatak je da vam prikažem najuobičajenije varijante, ali to istovremeno znači

da je moguće da u drugim tekstovima vidite drugačije oblike navođenja. Proporcije se najčešće označavaju malim slovom p, uz koje takođe mogu da stoje indeksi koji pobliže označavaju grupu. Budući da proporcije ne mogu da pređu vrijednost 1.0, obično se izostavlja 0 s lijeve strane decimalnog znaka, pa bi za gore navedeni primjer to bilo $p_m = .60$.

Iz svega naprijed navedenog možemo zaključiti da je formula za proporciju unutar neke grupe 1 jednaka $p_1 = \frac{n_1}{n}$, dok se procenti dobijaju kada odgovarajuću proporcije pomnožimo sa 100: $p_1 \cdot 100 = \frac{n_1}{n} \cdot 100$.

Šta se tačno i kako prikazuje u tabelama?

Da se vratimo sada na tabele. Osim prvog reda u kojem ćemo saopštiti nazive za svaku kolonu, tabele kojima se prikazuju distribucije jedne kategoričke varijable skoro uvijek u kolonama prikazuju sljedeće elemente:

- naziv kategorije
- apsolutnu frekvenciju
- relativnu frekvenciju, odnosno postotke ili proporcije čiji broj decimala zavisi od toga koliko je decimala smisleno prikazati za datu kategoriju (obično 1-3)

Uz navedeno je poželjno da se prikaže i red sa ukupnim vrijednostima (koji u koloni postotaka ima vrijednost 100%). Ponekad se sreće i kolona sa kumulativnim relativnim podacima pri čemu se u svakom novom redu uvrštava postotak (ili proporcija) sabran sa postocima prethodnih kategorija. Ako je bilo nedostajućih podataka, dodaje se red sa brojem i udjelom kategorije "nedostajući podaci". Uz nju se dodaje i kolona sa procentima za svaku kategoriju nakon što se od ukupnog broja ispitanika oduzme broj nedostajućih podataka za datu varijablu.

Za binarne varijable nećemo zabraniti softveru da prikaže tabele, ali ih mi nećemo prikazivati drugima (faza prezentacije). U najvećem broju slučajeva će biti dovoljno da opišemo raspodjelu u jednoj rečenici, na primjer: "uzorkom je obuhvaćeno 200 ispitanika od kojih je 120, odnosno 60.0%, bilo muškog pola". Izuzimno ćemo navesti i tabelarni prikaz ukoliko smatramo da je broj nedostajućih vrijednosti velik i da je potrebno da se uz ukupne procente, navedu i procenti valjanih vrijednosti.

Kada su u pitanju nominalne varijable sa više od dvije kategorije, prvo je potrebno da razmotrimo da li je zaista neophodno da

prikažemo tabelu. Ako se za to odlučimo, onda bi trebalo da slijedimo jedno prosto pravilo o redoslijedu prikazanih kategorija: redamo ih u opadajućem nizu od najčešće do najmanje zastupljene kategorije. Nakon što smo predstavili sve kategorije dolazi red koji informiše o broju nedostajućih podataka za tu varijablu. Takvim prezentovanjem čitaoci brže upijaju stanje stvari na uzorku. Na primjer, ako prikazujemo distribuciju različitih vrsta fobija u kliničkom uzorku, jasnije je ako počnemo od najčešće zastupljene (recimo, socijalna fobija sa 35% uzorka) pa idemo prema rjeđima (kao što je fobija od leta avionom koju pokazuje 3% uzorka). Na Slikama 4.2 i 4.3 su prikazani isti podaci o tome sa kojih fakulteta dolaze ispitanici u jednom istraživanju. Očito je u kojem slučaju lakše uočavate koje su tri najčešće kategorije i zašto je važan redoslijed u tabelama.

Fakultet	n	%	Validni %	Kumulativni %
AGGF	65	15.9	16.0	16.0
Ekonomija	73	17.8	18.0	34.0
ETF	79	19.3	19.5	53.4
Filozofski	74	18.0	18.2	71.7
FPN	34	8.3	8.4	80.0
PMF	81	19.8	20.0	100.0
Nedostajući	4	1.0		
Ukupno	410	100.0		

Slika 4.2: Primjer tabele distribucije jedne nominalne kategoričke varijable poredane prema abecedi.

Fakultet	n	%	Validni %	Kumulativni %
PMF	81	19.8	20.0	100.0
ETF	79	19.3	19.5	53.4
Filozofski	74	18.0	18.2	71.7
Ekonomija	73	17.8	18.0	34.0
AGGF	65	15.9	16.0	16.0
FPN	34	8.3	8.4	80.0
Nedostajući	4	1.0		
Ukupno	410	100.0		

Slika 4.3: Tabela distribucije nominalne varijable poredane prema učestalosti kategorija.

I kada su u pitanju ordinalne kategoričke varijable, prvo ćemo razmotriti da li su tabele uopšte potrebne, a ako se odlučimo za njih, onda ćemo se ipak držati onog redoslijeda kategorija

koji odražava originalni poredak kategorija prema količini svojstva procjene. Na primjer, ako kategorije predstavljaju stepen saglasnosti sa odgovorom na nekom upitniku od "nimalo" do "u potpunosti" onda ćemo se tog redoslijeda - ili eventualno inverznog, ako smatramo da je upečatljiviji - držati pri kreiranju tabele. Primjer tabele za takvu ordinalnu kategoričku varijablu vidimo na na Slici 4.4.

"Lako stupam u kontakt sa nepoznatima."	n	%	Validni %	Kumulativni %
Nimalo	10	6.9	6.9	6.9
U maloj mjeri	17	11.7	11.8	18.8
Umjereno	38	26.2	26.4	45.1
Uglavnom	56	38.6	38.9	84.0
U potpunosti	23	15.9	16.0	100.0
Nedostajući	1	0.7		
Ukupno	145	100.0		

Slika 4.4: Prikaz tabele distribucije jedne ordinalne varijable: prema redoslijedu kategorija.

Ekonomičnost je načelo kojeg se često moramo držati, naročito kada podatke prikazujemo u naučnim časopisima. Tada ćemo opis većeg broja pojedinačnih kategoričkih varijabli smjestiti u jednu tabelu. Držaćemo se gore navedenih pravila o redoslijedu kategorija, a ovdje ćemo biti slobodniji da prikažemo čak i dihotomne varijable. Slika 4.5 prikazuje "zgodnoću" ovakvog pristupa.

Varijabla	*n (%)*
Pol	
ženski	567 (75.9%)
muški	180 (24.1%)
Studenti psihologije	
ne	700 (93.7%)
da	47 (6.3%)
Depresivnost	
nimalo	58 (7.8%)
zanemarivo	154 (20.6%)
blago	206 (27.6%)
umjereno	209 (28.0%)
izuzetno	120 (16.1%)
Anksioznost	
nimalo	72 (9.6%)
zanemarivo	103 (13.8%)
blago	158 (21.2%)
umjereno	187 (25.0%)
izuzetno	227 (30.4%)
Zloupotreba supstanci	
nimalo	429 (57.4%)
zanemarivo	111 (14.9%)
blago	52 (7.0%)
umjereno	40 (5.4%)
izuzetno	115 (15.4%)

Slika 4.5: Prikaz tabele koja istovremeno prikazuje distribuciju više kategoričkih varijabli.

Već ste mogli da uočite kakav je dizajn tabela koji se uglavnom poštuje u psihološkoj naučnoj literaturi. Radi se o dizajnu propisanom u priručniku za objavljivanje Američkog psihološkog udruženja (Publication Manual of the APA; trenutna verzija je 7 izašla 2019. godine). Radi veće preglednosti, tabele formatiramo tako da prikazujemo samo vodoravne linije. Svaka Tabela ima svoj redni broj na osnovu redoslijeda javljanja u tekstu (npr. Tabela 1, Tabela 2), te se u tekstu pozivamo na tabelu tako što kažemo: "Kako se vidi iz Tabele 4, postoje velike razlike u pogledu…". Uz to svaka tabela mora da ima i svoj deskriptivni naslov, pa bi tabela prikazana na Slici 4.5 mogla biti nazvana *Osnovna deskriptivna statistika uzorka*. Jedno od važnijih pravila je da se napomene vezane za tabelu i pojašnjenja skraćenica daju netom ispod tabele.

Kako se grafički predstavljaju distribucije jedne kategoričke varijable?

Slijedi prikaz selekcije vidova grafikona koji se češće koriste za prikazivanje raspodjele jedne kategoričke varijable. Kratko ću predstaviti prednosti i mane svakog od njih, kao i pozadinski mehanizam na osnovu kojeg se generišu. Kada već to pominjem, moram naglasiti da ovdje ima prostora samo da se navedu osnovne karakteristike vizualizacije. Vizualizacija podataka je izuzetno dinamično i teorijski razvijeno područje kojem su posvećene mnoge knjige, a možete i čitavu karijeru izgraditi na tome.[3]

[3] Na našem jeziku je izuzetno dobro napisan interaktivni udžbenik Dejana Pajića, profesora sa Univerziteta u Novom Sadu: "Primena tehnika vizuelizacije u baziČnoj statistici". Apsolutno ga preporučujem, a besplatno je dostupan onlajn na stranici: http://psihologija.ff.uns.ac.rs/viz/.

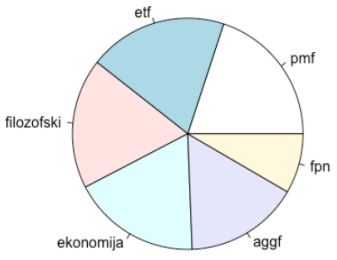
APA priručnik za objavljivanje, ali i prosta mudrost, nas obavezuju da svi grafikoni koji će da završe u naučnom ili stručnom izvještaju imaju sljedeće:

1. Tehnički naziv na koji ćete se referisati u tekstu. Već ste primijetili u ovom tekstu da koristim naziv *Slika*, a da numeracija ovisi o redoslijedu javljanja u tekstu, pa će grafikoni biti nazivani Slika 1, Slika 2 itd.

2. Opisni naziv koji sadrži nedvosmislen opis toga šta je prikazano na grafikonu. Na primjer, nećemo napisati samo *Dijagnoza*, nego *Raspodjela učestalosti različitih dijagnoza na uzorku (n = 220)*.

3. Tekstualna objašnjenja u legendi kojom se opisuje označavanje kategorija. Osim već pomenutog redoslijeda kojeg koristimo i za tabelarni prikaz kategoričkih varijabli, kategorije ćemo najčešće obilježavati različitim tonom, odnosno kvalitetom boje (nominalne), kombinacijom saturacije i tona boje (ordinalne), te u nekim slučajevima i / ili različitim znakovima.

4. Eventualno, radi bolje vidljivosti, mogu se dodati i numerički rezultati unutar grafikona.



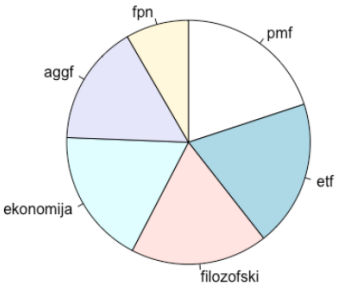
Slika 4.6: Kružni grafikon bez definicije početne tačke.

Kružni grafikon (engl. *pie chart*) se često sreće u popularnim medijima. Kod nas je poznat i pod manje formalnim nazivima kao torta, pita ili pizza grafikon - od kojih je pizza vjerovatno najprikladniji (osim u slučaju calzone verzije).

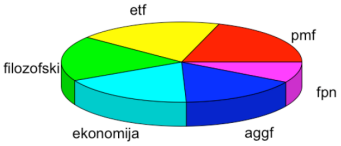
Veličina odsječka predstavlja udio kategorije u uzorku, a konkretna vrijednost se određuje tako što krećemo od određene

tačke i odmjeravamo ugao koji proporcionalno odgovara udjelu kategorije.

Budući da znamo da krug ima 360 stepeni, znamo da bi jedna četvrtina, odnosno kategorija koja predstavlja 25.0% uzorka bila odsječak od 90 stepeni ($360^\circ * \frac{1}{4} = 90^\circ$). Nažalost, uprkos enormnoj popularnosti danas znamo da su kružni grafikoni loš način prikazivanja raspodjele. Kao prvo, kružni grafikoni koriste previše prostora za prikazivanje relativnog udjela kategorija. Drugo, odsječci su neprirodan način opisivanja zastupljenosti kategorija. Zamislite, 1% je odsječak od 3.6 stepeni, a jednu petinu, odnosno 20% predstavljamo odsječkom od "okruglo" 72 stepena?! Stanje se pogoršava sa sve većim brojem kategorija koje varijabla sadrži, kao i kada se upotrebljavaju trodimenzionalni prikazi (Slika 4.8). Zaključak je da, uprkos površnoj ljepoti koju imaju, kružne grafikone nikada ne bismo trebali izrađivati.



Slika 4.7: Kružni grafikon sa definisanom početnom tačkom i redoslijedom kategorija prema učestalosti.

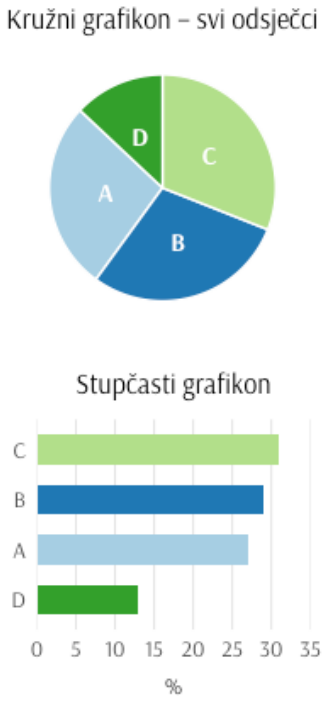


Slika 4.8: Trodimenzionalni kružni grafikon.

Osim kada vas pritisne viša sila (kompulzivna ljubav ka okruglom ili naredba nadređenog lica - šefa ili profesora), a i tada samo ako je izuzetno mali broj kategorija. Koristite isključivo dvodimenzionalni prikaz uz ucrtane numeričke pokazatelje. Za početnu tačku crtanja uzmite onu koja pokazuje sjever na kompasu i određujte veličinu ugla u smjeru kazaljke na satu.

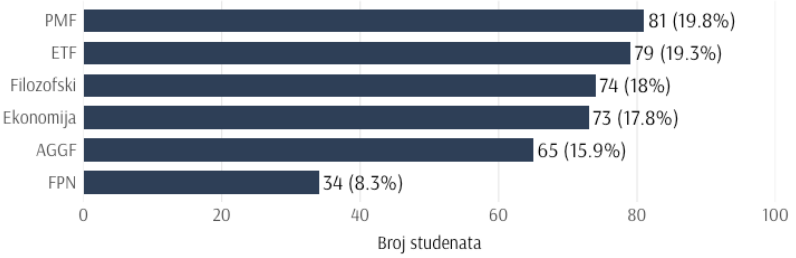
Stupčasti grafikon (neki ih nazivaju i stubičasti ili štapićasti, engl. *bar* ili *column chart/plot*) je uz kružni najpopularnija vrsta grafikona. Na jednoj od osa se nalazi oznaka kategorije, na drugoj osi su označene frekvencije (apsolutne ili relativne). Visina / dužina stupca označava zastupljenost određene kategorije, a njegova širina nema značenje. Iako su nešto češće zastupljeni uspravni stupčasti grafikoni, vodoravno položeni su zapravo pregledniji, jer omogućavaju da se navede puno ime i onih kategorija koje imaju dugačka imena. Kao i u slučaju tabela, kada su u pitanju nominalne varijable redoslijed kategorija bi trebalo da prati njihovu empirijsku zastupljenost (u uzlaznom ili silaznom nizu), dok za ordinalne varijable redoslijed bi trebalo da odražava teorijski rast (ili pad) prisutnosti atributa koji je svojstven kategorijama. Prednosti ovog tipa grafikona, definitivno boljeg od kružnog, su intuitivnost i dostupnost. Naime, uvid je intuitivan kada postoji logični slijed kategorija, a tada pratimo ista načela o redoslijedu kao za tabele.

Evo da provjerimo funkcionalnost kružnog grafikona naspram stupčastog. Pokušajte da procijenite i uporedite učestalost kategorija A, B i C na Slici 4.9. Možete li brzo reći koji je najveći od njih? Sada pokušajte to isto sa stupčastim grafikonom. Iako vam možda kružni ljepše izgleda, razlika u efikasnosti je očigledna.



Slika 4.9: Direktno poređenje kružnog i stupčastog grafikona.

Nisu nam čak potrebne ni boje da označimo različite kategorije nominalne varijable. To možemo vidjeti iz primjera na Slici 4.10 gdje stupčasti grafikon mnogo jasnije pokazuje koje su fakultete pohađali ispitanici u odnosu na gore prikazivane kružne varijante. A što se tiče dostupnosti praktično svi tabelarni i statistički softveri omogućavaju da kreiramo stupčaste grafikone.



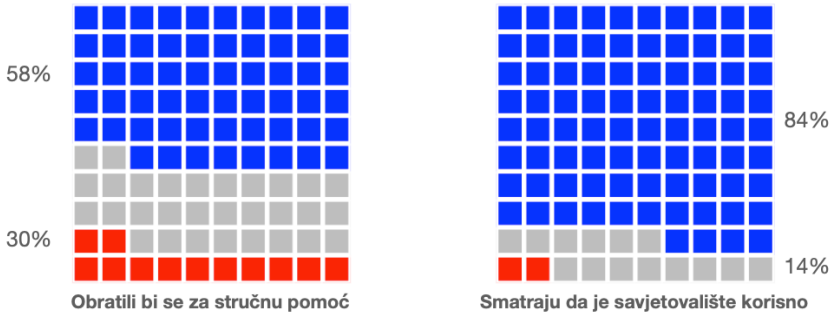
Slika 4.10: Stupčasti grafikon koji prikazuje podatke o fakultetima koje ispitanici pohađaju.

Ipak, čak i za ovaj vid grafikona pokazuje i neke mane: površina koja se koristi za izražavanje zastupljenosti kategorije nije baš efikasno iskorištena, širina stubaca je proizvoljna, a uz to je teško procijeniti kolika je zastupljenost kategorija unutar cjeline.

Upravo zato je osmišljena i nešto bolja zamjena za kružne grafikone. Radi se o grafikonu za koji još nemamo ni stabilno ime, ali se meni najviše dopada **rešetkasti grafikon** (engl. *grid chart*). Taj naziv mi se čini bolji nego što se povratnim prevodom sa engleskog dobijaju "neukusni" **vafl grafikon** (engl. *waffle plot*) ili **grafikon kvadratne pite** (engl. *square pie chart*)?!

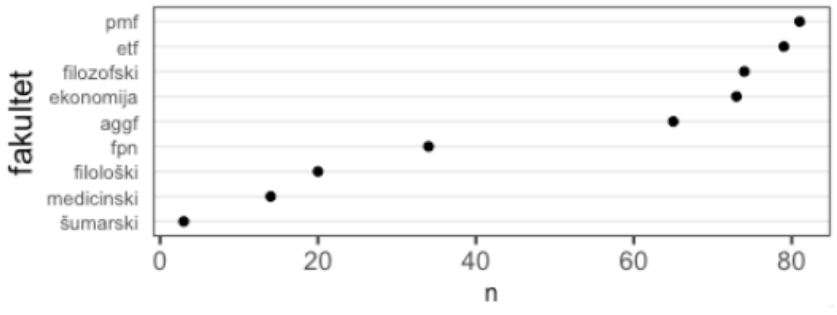
Zadana verzija rešetkastog grafikona ima 10 redova i 10 kolona, tako da imamo 100 kvadratića od kojih svaki predstavlja 1% varijable. To je izuzetno zgodno. Primjer se nalazi na Slici 4.11. Međutim, slobodni smo da odredimo i drugačiju veličinu, npr. 10 x 6 ili 6 x 10, ako smatramo da je to iz nekog razloga bolje. Dakle, prednost rešetkastog grafikona se sastoji u tome što kvadratići omogućuju intuitivno uočavanje razlike u zastupljenosti pojedinih kategorija unutar cjeline, pogotovo ako se radi o procentnoj skali. Mana se odnosi na to što ukoliko imamo veći broj kategorija moramo da koristimo i veliki broj boja koje bi morale biti distinktivne, te onda moramo da koristimo i spoljne legende. Takođe, slabo zastupljene kategorije možda neće biti ni predstavljene, niti vidljive na takvom grafikonu.

Clevelandov tačkasti grafikon (engl. *Cleveland dot plot*) je pokušaj da se prevaziđe i druga mana stupčastog grafikona, njegova neefikasnost. Na tom grafikonu se (obično) na Y osi



Slika 4.11: Rešetkasti grafikoni koji prikazuju odgovore na pitanja o spremnosti za obraćanje psihološkom savjetovalištu i mišljenje o korisnosti savjetovališta. Plavom bojom su označeni odgovori "Da", sivom "Ne znam", crvenom "Ne".

navodi ime kategorija, na X osi mjerna skala (apsolutne ili relativne frekvencije), dok se tačkom ili nekim drugim simbolom onda označava zastupljenost svake pojedinačne kategorije. Uobičajeno je da se taj simbol nanese na slabije naglašenu liniju koja se vodoravno proteže od imena kategorije duž čitave X dimenzije. Zahvaljujući uštedi prostora, Clevelandov tačkasti grafikon je naročito koristan za slučajeve kada postoji veći broj kategorija unutar nominalne kategoričke varijable. Ovaj tip grafikona je idealan kada prikazujemo rezultate istraživanja o zastupljenosti različitih problema sa mentalnim zdravljem, gdje se može pojaviti 15-20 različitih kategorija (od *anksioznosti* do *zloupotrebe psihoaktivnih supstanci*). Kao primjer može da posluži Slika 4.12 gdje sam za podatke već prikazane kružnim i stupčastim grafikonom dodao još neke kategorije koje su u uzorku bile slabije zastupljene. Doduše, u slučaju ordinalnih kategoričkih varijabli, kao i nominalnih kategoričkih varijabli sa malim brojem kategorija njegove prednosti nad stupčastim grafikonom splašnjavaju.



Slika 4.12: Primjer Clevelendovog tačkastog grafikona.

Kako strateški pristupiti deskriptivnoj analizi jedne kategoričke varijable?

Dakle, kad pred sobom imamo tabele i grafikone imamo dovoljno informacija da donesemo bitne sudove o stanju na uzorku. Ali na šta to konkretno treba da obratimo pažnju pri interpretaciji? Evo najvažnijih pitanja koja sebi treba da postavimo i da na njih odgovorimo:

1. Koja je kategorija dominantna - ako takva postoji, a koje su zanemarivo prisutne?
2. Da li smo sveukupno gledano očekivali ovakvu distribuciju relativnih frekvencija na osnovu naših pretpostavki o stanju u ciljnoj i radnoj populaciji? Postoje li neke iznenađujuće visoko ili slabo zastupljene kategorije? Sveukupno gledano, da li smo očekivali da dobijemo toliko visoku ili nisku raznovrsnost podataka?
3. Kako da nastavimo sa obradom podataka? Uočavamo li da postoje teorijsko-značenjski slične kategorije koje bismo mogli spojiti i kreirati novu varijablu? Šta da uradimo sa slabo zastupljenim kategorijama i nedostajućim podacima?
4. Na koji način ćemo najefektivnije prikazati podatke drugima? Da li je, ako je riječ o usmenoj prezentaciji potpomognutoj slajdovima, bolje ići na grafikone, a u pisanom izvještaju na tabele ili grafikoni mogu da posluže u oba slučaja? Možemo li upotrijebiti boju, oblike, dodatan tekst na grafikonima da potenciramo najbitnije informacije?
5. Ako čitamo tuđi izvještaj, uočavamo li možda neke namjerno distorzirane (npr. potkraćena mjerna skala koja naglašava razlike među kategorijama koje zapravo nisu velike) ili neprikazane aspekte (npr. kategorije koje uopšte nisu prikazane)? Ako da, šta bi mogao da bude motiv autora?

Šta da radimo ako imamo neke slabo zastupljene nivoe kategoričke varijable?

Među gore navedenim pitanjima je pomenut i poseban problem vezan za obradu kategoričkih podataka. Naime, u istraživanjima će neke kategorije unutar varijable biti slabo zastupljene. Ovo *slabo zastupljene* ima relativno značenje; nekad će to biti manje od 10 ispitanika po kategoriji, često manje od 25, a nekad čak i manje od 100 (ukoliko je ciljna populacija izuzetno velika i smatramo da 100 ispitanika ne može adekvatno teorijski

predstaviti populaciju). U svakom slučaju, slabo zastupljene kategorije ograničavaju našu moć da generalizujemo nalaze na ciljnu populaciju i svakako takve kategorije moramo tretirati prije predstavljanja rezultata, osim u onim slučajevima kada su te kategorije izuzetno zanimljive po nečemu. Na raspolaganju nam stoje tri postupka:

1. Možemo združiti takve kategorije sa nekim teorijski srodnim kategorijama, ukoliko one postoje. Na primjer, ukoliko se radi o varijabli *stručna sprema*, te imamo mali broj onih koji su završili magistrat nauka ili doktorat, vjerovatno je smisleno njih pridružiti ispitanicima koji imaju završenu visoku stručnu spremu. Slično tome, ako u istraživanju imamo samo po nekoliko ispitanika iz kategorija "udovac/udovica" i "rastavljen/a", možemo ih spojiti u kategoriju "bez partnera" zajedno sa kategorijom "neoženjen/neudata".
2. Članove rijetkih kategorija nekad možemo "strpati" u kategoriju "ostalo/i", ponovo pod uslovom da to ima teorijskog smisla.
3. Konačno, možemo odustati od tih kategorija, odnosno obrisati ih, ako zajedno one predstavljaju raznorodnu grupu čije spajanje u kategoriju *ostalo* nema posebnog smisla. U nedavnom istraživanju smo ispitivali mentalno zdravlje studenata sa različitih fakulteta. Na većini fakulteta smo imali dobar odziv ispitanika, ali je on bio slab na Filološkom, Medicinskom i Šumarskom (npr. Slika 4.12). Budući da pretpostavljamo da bi razlike u kvalitetu mentalnog zdravlja mogle biti vezane za prirodu studija, ne bi baš bilo smisleno spojiti ih u jednu kategoriju za kasnije analize jer studenti koji upisuju ta tri fakulteta imaju i različite motive, a i različite stresore tokom studija. Bez obzira na to koju smo od navedenih odluka donijeli, moramo je tekstualno obrazložiti.

Nakon čitanja ovog poglavlja trebalo bi da imate dobar temelj o standardima analize pojedinačnih kategoričkih varijabli. Nadam se da ste shvatili da je za dobru komunikaciju sa budućim čitaocima vaših radova potrebno posvetiti dužnu pažnju kreiranju tabela ili grafikona. Sljedećim poglavljem ćemo povećati arsenal analitičkih alata tako što ćemo se posvetiti dimenzionim varijablama. Ponovo ćemo koristiti tekst, tabele i grafikone, ali na nešto drugačiji način.

Poglavlje 5

Opisivanje raspodjele jedne dimenzione varijable

Nakon čitanja ovog poglavlja znaćete:

- kako pristupiti obuhvatnoj analizi distribucije jedne dimenzione varijable kombinovanjem različitih grafikona i statističkih mjera,
- izračunati i interpretirati mjere centralne tendencije, te donositi odluke o njihovoj prikladnosti na osnovu oblika distribucije i prisutnosti ekstremnih vrijednosti,
- izračunati i tumačiti mjere varijabilnosti kako biste opisali stepen raznovrsnosti podataka,
- tumačiti mjere pozicije u raspodjeli za pozicioniranje pojedinačnih skorova u kontekstu cjelokupne distribucije.

Analiza distribucije dimenzionih varijabli je kompleksnija od analize kategoričkih varijabli. Razlog je jednostavan: dimenzione varijable imaju više različitih vrijednosti. Zbog toga moramo odgovoriti na više pitanja koristeći i veći broj statističkih mjera i grafikona. Ta pitanja vam u ovom trenutku možda neće zvučati zanimljivo, ali hoće kada vi u svom istraživanju budete ispitivali distribuciju neke od dimenzionih varijabli koju ste procjenjivali - a to mogu biti: inteligencija, emocionalna kompetentnost, neka od mračnih crta ličnosti, postignuće na nekom testu znanja, zadovoljstvo porodičnim ili partnerskim odnosom, zadovoljstvo poslom…Pomoglo bi ako biste sada zaista zamislili da ste uključeni u ogroman istraživački projekat i da grčevito iščekujete odgovore na sljedeća pitanja:

1. Koja je to vrijednost varijable koja dobro opisuje cijelu grupu koju ispitujemo? Postoji li jedna ili više takvih vrijednosti?

2. Koliko su objekti međusobno različiti kada je u pitanju data dimenzija? Sveukupno gledano, da li smo očekivali da dobijemo toliko visoku ili nisku raznovrsnost podataka?
3. Da li smo očekivali ovakav oblik distribucije na osnovu naših pretpostavki o stanju u populaciji? Da li su na mjernoj skali neke lokacije iznenađujuće visoko ili slabo zastupljene? Da li nam se skorovi gomilaju pri krajevima distribucija, pa nemamo postepeni pad nego dolazi do naglog prekida (tzv. efekat poda i efekat plafona)?
4. Postoje li objekti koji toliko odstupaju od ostalih da se pitamo da li oni zaista pripadaju grupi ili se od nje suštinski razlikuju? Ako se razlikuju, da li postoji velika vjerovatnoća da se radilo o tehničkim greškama pri unosu ili namjernim distorzijama pri procjenjivanju, pa ih je potrebno eliminisati iz baze podataka?
5. Ako je došlo do nekih rezultata koji odstupaju od naših očekivanja, zašto je do njih moglo doći? Možemo li nekom dodatnom analizom provjeriti zašto je to tako? Sveukupno gledano, u kojoj mjeri vjerujemo da su rezultati reprezentativni za ciljnu populaciju? Šta nam rezultati govore o specifičnosti mjernog instrumenta i/ili uzorka na kojem je vršeno istraživanje?
6. Kako nastaviti sa obradom podataka? Postoji li potreba za statističkom korekcijom varijable (npr. matematičkom transformacijom, tretiranjem ekstremnih skorova, imputacijom nedostajućih podataka, kategorisanjem dimenzionih varijabli) kako bi se došlo do adekvatnijih uvida?
7. Ukoliko se radi o tuđem istraživanju, da li postoje neke naznake da su rezultati mogli biti i drugačije predstavljeni, npr. korištenjem prigodnije mjerne skale, prikazivanjem drugačijih statističkih mjera koje bi više odgovarale dobijenim podacima? Da li se čini da su razlozi za suboptimalni prikaz posljedica namjernih distorzija ili se radi o neznanju istraživača?

Znam, nije lako osjetiti taj istraživački žar samo čitajući udžbenik na 1. godini studija. Međutim, mnogi među vama ćete ga doživjeti čim se upustite u vlastito istraživanje. U svakom slučaju, da biste dobili dobre odgovore na gore navedenu seriju pitanja neophodno je da se upoznamo sa različitim vidovima grafičkih prikaza i statističkih mjera centralne tendencije, varijabilnosti i pozicije u nizu.

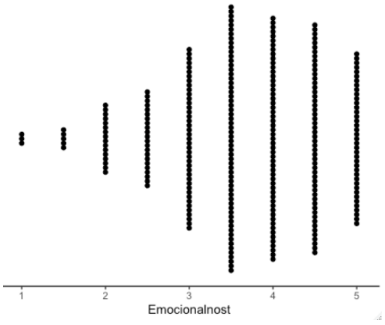
Kako grafički prikazati distribuciju jedne dimenzione varijable?

Postoji veći broj grafikona koji nam može poslužiti da predstavimo distribuciju jedne dimenzione varijable. Kako ništa na svijetu nije savršeno, tako neće biti ni neki pojedinačni tip grafikona, pa smo osuđeni na korištenje više njih kako bismo izvukli što više korisnih informacija. Štaviše, kada budemo prezentovali rezultate, često ćemo u jednom prikazu kombinovati dva ili više tipova grafikona. U ovom potpoglavlju opisujem jednostavnije grafikone, može se reći "dječije", jer oni ne zahtijevaju poznavanje statističkih mjera centralne tendencije, varijabilnosti i pozicije u nizu. Nakon što te pojmove savladamo, doći će na red i oni "odrasliji", s tim da to ne znači uopšte da se radi o boljim grafikonima.

Eh da, još jedna napomena: praktično sve što se navodi u ovom potpoglavlju važi za dimenzione varijable, osim za rang-ordinalne varijable. Predstavljanje pojedinačnih rang-varijabli putem grafikona je beskorisno, budući da su to varijable gdje su objekti pozicionirani podjednako duž mjerne skale bez obzir koji način prikazivanja mi koristili. Afirmativnije rečeno, sljedeći grafikoni su korisni za prikazivanje varijabli na kvazi-intervalnom, intervalnom, omjernom i apsolutnom nivou "mjerenja".

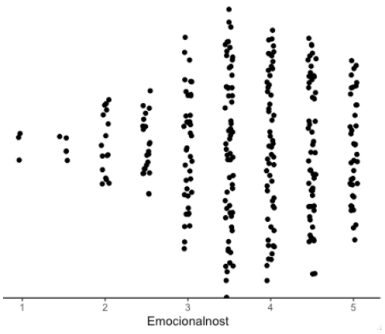
Tačkasti grafikon

Wilkinsonov tačkasti grafikon ili ponekad samo tačkasti grafikon (engl. *Wilkinson dot plot*, *strip plot*) ili 1d tačkasti grafikon je naprostiji način prikazivanja podataka jedne dimenzione varijable (Slika 5.1). Na jednoj osi - obično je to X-osa - nanosimo mjernu skalu za datu varijablu, dok druga osa nema značenje. (Za one koji to nisu primijetili, radi se o jednoj dimenzionoj varijabli, pa je u načelu dovoljno korištenje samo jedne dimenzije.) Unutar prostora za grafikon smještamo tačke ili neki druge simbole koji predstavljaju skorove pojedinačnih ispitanika. To je i uzvišena namjena tačkastog grafikona: da nikoga ne zaboravimo, odnosno da prikažemo sve pojedinačne podatke dobijene na uzorku.

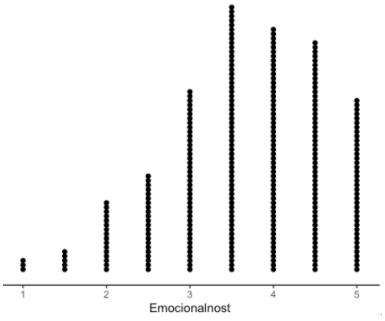


Slika 5.1: Distribucija varijable Emocionalnost prikazana putem tačkastog grafikona sa "rastrešenjem" samo po Y dimenziji.

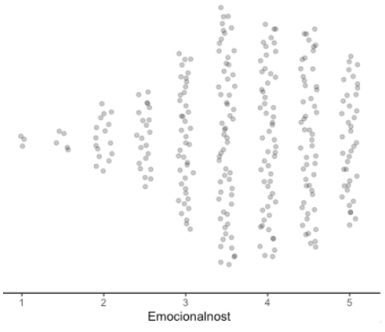
Međutim, dilema nastaje kada više ispitanika ima identičan skor. To se gotovo uvijek dešava. Šta uraditi sa pojedinačnim tačkama, a da se one ne preklapaju, odnosno šta da uradimo kako bismo postali svjesni nakon gledanja grafikona da više od jednog ispitanika ima određenu vrijednost? Postoje dva elegantna rješenja.



Slika 5.2: Distribucija varijable Emocionalnost prikazana putem tačkastog grafikona sa "rastrešenjem" po obje dimenzije.



Slika 5.3: Distribucija varijable Emocionalnost prikazana putem slaganog tačkastog grafikona.



Slika 5.4: Distribucija varijable Emocionalnost za koji je korištena kombinacija dvodimenzionalnog rastrešenja i variranja prozirnosti tačaka.

Prvo od njih je da se simboli za takve ispitanike rasprostru duž one druge, irelevantne dimenzije. Neki smatraju da je ova izmjena toliko velika da izmijenjeni grafikon zaslužuje drugi naziv. Zato ćete sresti i posebna imena za takve grafikone. Na primjer, rastresenim grafikonom (engl. *jitter plot*) se naziva grafikon gdje se simboli nasumično raspodjeljuju duž druge, dotad nekorištene dimenzije. U praksi se često neznatno raspršuje i raspored duž druge ose. To daje pregledniji prikaz, ali uz malo "laguckanja". Na primjer, ispitanik koji je ostvario skor 3.50 na grafikonu može izgledati kao da je ostvario 3.52 (Slika 5.2). Druga mogućnost je da simbole za ispitanike sa identičnim skorovima gomilamo jedan na drugi duž neupotrijebljene ose. To može da se radi poštujući princip gravitacije gdje se identične vrijednosti slažu uvis u odnosu na temelj, odnosno korištenu osu. Takođe, možete povući imaginarnu liniju paralelnu sa korištenom osom. U odnosu na tu liniju slažete identične vrijednosti ispod ili iznad. Ovaj oblik grafikona se naziva slagani tačkasti grafikon (engl. *stacked dot plot* ili *histodot plot*) (Slika 5.3).

Drugo rješenje, koje je naročito zgodno kada imamo veliki broj ispitanika, jeste da učinimo svaku pojedinačnu tačku prozirnijom. U softveru se zadaje neka proporcija zasićenosti boje - često se ova karakteristika naziva alpha - koja će biti manja od uobičajene vrijednosti 1 (minimalna vrijednost je 0). Time se postiže da se puna obojenost neke tačke dešava samo ukoliko postoji veći broj ispitanika sa istim skorom. Pomoću ove dvije intervencije, a najčešće njihovom kombinacijom - što prikazuje Slika 5.3 - dobijamo mnogo pregledniju sliku distribucije. Ali avaj, iako tačkasti grafikon treba cijeniti zbog njegove inkluzivnosti, preveliki broj pojedinačnih informacija nam ne daje baš jasan uvid u opšte tendencije distribucije.

Histogram

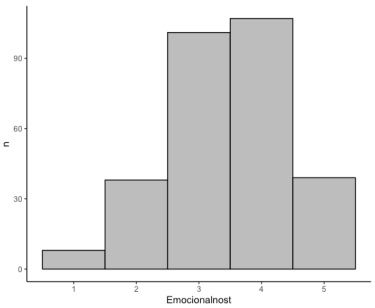
Upravo zbog sposobnosti da intuitivno predstavi karakteristične lokacije gomilanja i količinu asimetričnosti, **histogram** (engl. identično: *histogram*) je ipak najpopularniji oblik predstavljanja dimenzionih varijabli. Histogram je vrlo blizak rođak stupčastog grafikona, ali nije brat blizanac. Obično se na X-osi prikazuje mjerna skala, a visina stubaca - kao i u slučaju stupčastih grafikona predstavlja udio kategorije unutar uzorka. Otkud sad kategorije za dimenzionu varijablu? Evo otkud: da bismo dobili histogram dimenziona varijabla se prisilno kategoriše u neki broj ordinalnih kategorija koje zauzimaju jednake intervale duž mjerne skale. Treba naglasiti da se ta međuoperacija ne prikazuje

u softveru. Sve u svemu, ključna razlika u odnosu na stupčaste grafikone je to što između stubaca nemamo razmake, baš zato što se radi o kategorijama koje se naslanjaju jedna na drugu.

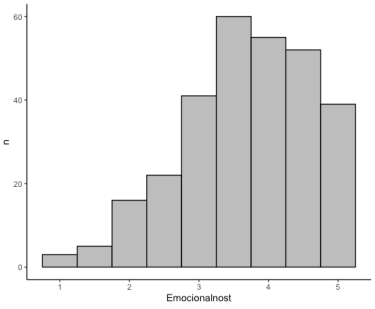
Iako histogrami slove za intuitivne grafikone koji efikasno prikazuju tendenciju grupe i asimetričnost distribucije postoji i jedan nezanemariv problem kada se oni konstruišu. Naime, izgled histograma u velikoj mjeri zavisi od toga koliki će broj intervalnih kategorija (engl. *bins*) biti kreiran, a time i stubaca [usput, kategorisanje intervala se nekad koristi i za tačkaste grafikone]. Statistički softveri koriste neki od gotovih algoritama za odabir optimalnog broja. Ti algoritmi su imenovani prema svojim tvorcima (npr. *Sturges*, *Freedman-Diaconis*) i u velikom broju slučajeva daju zadovoljavajući prikaz. Međutim, automatizavani algoritmi ne mogu garantovati da ćemo dobiti idealan prikaz distribucije podataka, te je pametno da razmotrimo više različitih rješenja kada smo u eksplorativnoj fazi kako bismo dobili adekvatne uvide, odnosno kako bismo se odlučili za najbolju opciju ako želimo naš histogram predstaviti i drugima.

Koliko broj intervalnih kategorija može da utiče na prikaz podataka vidimo ako uporedimo primjere na Slikama 5.5 i 5.6, odnosno na Slikama 5.7 i 5.8. Ako imamo premali broj kategorija možda ćemo propustiti da uočimo asimetričnost, postojanje više od jedne karakteristične vrijednosti ili prisutnost vrlo visokih ili niskih atipičnih skorova. S druge strane, ako imamo preveliki broj kategorija, baš kao što je slučaj i sa Wilkinsonovim tačkastim grafikonom, nećemo moći da uočimo neke opšte pravilnosti / tendencije distribucije.

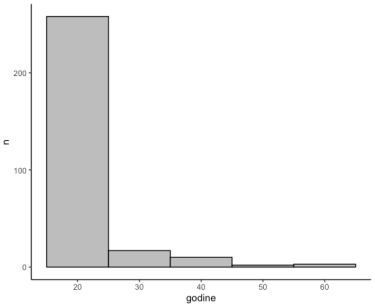
Kada eksperimentišete sa brojem kategorija, možete koristiti sljedeće smjernice: za uzorke manje od 50 ispitanika, rijetko je potrebno više od 10-12 kategorija. Za veće uzorke, možete ići do 20-30 kategorija, ali pažljivo posmatrajte da li se gube osnovni obrasci distribucije. Dobro je pravilo da testirate tri različita broja kategorija - konzervativniji pristup (manji broj), srednji pristup, i detaljniji pristup (veći broj) - te da odaberete onaj koji najbolje prikazuje ključne karakteristike vaših podataka.



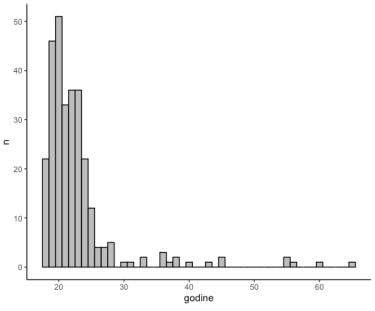
Slika 5.5: Distribucija varijable Emocionalnost prikazana histogramom sa malim brojem kategorija.



Slika 5.6: Distribucija varijable Emocionalnost prikazana histogramom sa većim brojem kategorija.



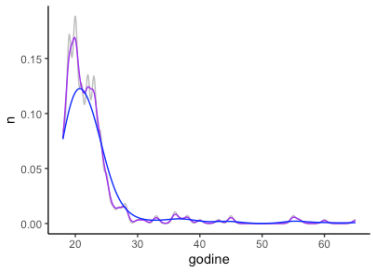
Slika 5.7: Distribucija varijable Godine starosti prikazana histogramom sa manjim brojem kategorija.



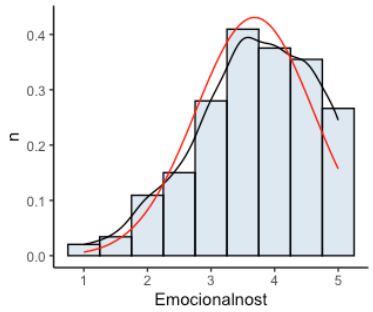
Slika 5.8: Distribucija varijable Godine starosti prikazana histogramom sa većim brojem kategorija.

Polirani grafikon gustine distribucije

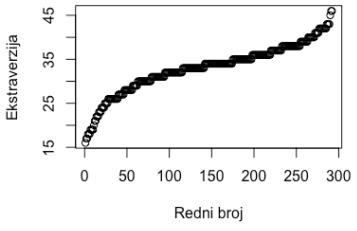
Za razliku od tačkastog grafikona, te u manjoj mjeri histograma, **polirani grafikon gustine distribucije** (engl. *(kernel) density plot*) uopšte nema za cilj prikazivanje konkretnih podataka iz uzorka, nego se njim upravo želi prikazati opšta tendencija distribucije date varijable. Umjesto mnoštva tačaka i stubaca ovaj grafikon



Slika 5.9: Distribucija varijable Godine starosti prikazana putem poligona frekvencije sa tri različita stepena poliranja.



Slika 5.10: Histogram i polirani grafikon proporcija za varijablu Emocionalnost, uz koje je dodana i očekivana normalna distribucija.



Slika 5.11: Indeks grafikon koji ne pokazuje izrazito stršeće mjere.

koristi samo jednu neprekidnu liniju. Tu liniju iscrtava softver na osnovu količine podataka na konkretnoj vrijednosti na mjernoj skali, ali i na susjednim lokacijama pri čemu se bliže lokacije više vrednuju, a vrednovanje podataka sa udaljenijih lokacija opada sve dok ne postanu potpuno nebitni. Načelo je dakle drugačije u odnosu na histogram gdje samo količina podataka unutar jednog intervala određuje dužinu jednog stupca. Kao i u slučaju histograma, postoje neka gotova rješenja o tome koliko bi interval trebalo da bude širok, ali nam je još isto tako dopušteno da ručno odredimo širinu intervala koji utiče na iscrtavanje linije.

Prednost predstavljanja podataka putem poliranog grafikona je što njime dobijamo življu predstavu o tome kako bi mogla da izgleda populaciona raspodjela varijable. Ali ponovo imamo problem koji važi i za histogram: stepen zaglađenosti linije, odnosno njene nazubljenosti, u velikoj mjeri mijenja izgled grafikona, te uvide koje možemo izvući iz njega. To je evidentno i na Slici 5.9 koja prikazuje iste podatke kao što su prikazani na Slikama 5.7 i 5.8. Dakle, pravilo isprobavanja različitih rješenja važi i za ovaj grafikon, a on je očito i dobar kandidat za kombinovanje sa drugim oblicima grafikona i naročito često ide ruku pod ruku sa histogramima.

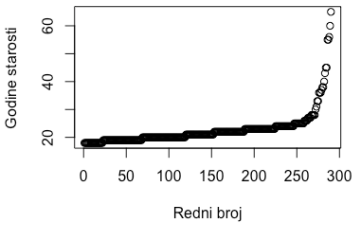
Pogledajmo sad Sliku 5.10 Na histogramu su nanesene dvije linije od kojih crvena predstavlja normalnu distribuciju, jednu od mnogobrojnih matematički definisanih raspodjela. Stepen podudarnosti, odnosno nepodudarnosti linija, nam ugrubo ukazuje na to kakve su šanse da naši podaci sa uzorka dolaze iz neke populacije u kojoj postoji normalna - ili neka druga - raspodjela vrijednosti. To može biti važna informacija za kasniju statističku obradu. Ipak, moram naglasiti da ove linije nisu polirani grafikoni gustine u istom smislu kao oni ranije prikazani, budući da se za njihovu konstrukciju koristi samo vrlo mali broj izvedenih empirijskih podataka sa uzorka (npr. mjere centralne tendencije i varijabilnosti), a ne i svi prikupljeni podaci.

Identifikacioni ili indeks grafikon

Identifikacioni ili indeks grafikoni (engl. *index plot*) su dosta neugledni i zanemarivani, mada predstavljaju vrlo korisne članove statističkog repertoara. Kao i u slučaju tačkastog grafikona pojedinačni element prikazuje skor pojedinačnog ispitanika. Međutim, za razliku od tačkastog grafikona, identifikacioni grafikon koristi i drugu dimenziju koordinatnog sistema. Na toj dimenziji se ispisuje redni broj ispitanika (tj. indeks ispitanika). Taj redni

broj može biti redni broj koji je ispitanik dobio pri unosu - sjećate se ID varijable - ili, još bolje, rang koji predstavlja poziciju ispitanika u rastućem ili opadajućem nizu na datoj varijabli.

Identifikacioni grafikon nam vrlo neposredno daje uvid u to da li postoje skorovi koji odskaču od ostatka uzorka (tj. stršeće mjere, odnosno sumnjivo niski ili visoki skorovi). Ovaj tip grafikona ni nema neku drugu funkciju - naime, njime je vrlo teško uočiti pravilnosti u distribuciji - ali ono što radi, radi kako treba. Ukoliko primijetimo prazan prostor, odnosno skok vrijednosti između dva ranga koji odstupa od niza, imamo itekako razloga da provjerimo da li je dati podatak valjan, te da odlučimo da li ga je potrebno tretirati prije dalje obrade podataka. Slike 5.11 i 5.12 prikazuju dva različita primjera gdje je očito da na jednom od njih imamo neke problematične vrijednosti koje treba istražiti prije nastavka analize.



Slika 5.12: Indeks grafikon koji ukazuje na evidentno stršeće mjere.

Mjere centralne tendencije

Šta je svrha mjera centralne tendencije?

Po samoj definiciji, varijable sadrže različite vrijednosti za jednu grupu od interesa. Na primjer, studenti psihologije bi se morali međusobno razlikovati u pogledu samoprocijenjene izraženosti opštih crta ličnosti kao što su ekstraverzija, prijatnost, savjesnost ili u pogledu nekog specifičnijeg konstrukta kao što je strah od statistike. Izradom grafikona možemo brzo steći uvid u raspodjelu varijable, ali jedna stvar nedostaje grafikonima: egzaktnost. Podsjetimo se da je opšti cilj statistike svesti veliku količinu izvornih informacija na mali broj bitnih informacija. Informaciona efikasnost bi bila maksimizovana kada bismo imali jedan broj kojim bismo mogli predstaviti cijelu grupu. Takvo nešto zaista i postoji. Statističke mjere čija je svrha da opišemo izraženost svojstva za grupu kao cjelinu nazivamo mjerama centralne tendencije. Zamislite koliko je ovo sredstvo efikasno: ako kažete da je srednja dob osobe u Kini 40.1 godina,[1] mi zapravo više od milijardu i četiri stotine i petnaest miliona podataka svodimo na jedan broj! Onda ovu informaciju možemo dovesti u kontekst sa podatkom o Japanom gdje u istom trenutku imamo procjenu da je srednja dob 49.8 godina (više od 120 miliona stanovnika) ili sa podatkom da je u Pakistanu procijenjena srednja dob 20.6 godina (više od 250 miliona stanovnika). Na osnovu samo ova tri podatka možemo već nešto da zaključimo o trenutnom demografskom stanju, a možda i o budućnosti tih

[1] https://www.worldometers.info/world-population/china-population/ pristupljeno 16.7.2025.

država. Efikasno, zar ne?

Postoji više mjera centralne tendencije. Najčešće korištene su: prosjek / aritmetička sredina, medijana / srednja vrijednost i mod / tipična vrijednost. Njima ćemo posvetiti posebnu pažnju, jer svaka od njih ima svoje uslove korištenja, kao i prednosti i mane koje će ovdje biti predstavljene. Uz to, upoznaćemo se i sa korisnim modifikacijama navedenih mjera kojima se služimo da bismo sebi i drugima predstavili opštu karakteristiku grupe.

Šta je mod i kako ga izračunavamo?

Počećemo od **moda**, modalne ili tipične vrijednosti (engl. *mode*), najjednostavnije mjere centralne tendencije za koju se u statističkoj notaciji obično koriste oznake *Mod* i *Mo*. Mod je najčešća mjera koja se pojavljuje u uzorku i nalazimo ga tako što jednostavno pregledamo tabelu javljanja svih mjera u uzorku (tabela frekvencija) i registrujemo onu koja se najčešće javlja. Na primjer, mod za niz od 10 mjera (1, 1, 3, 4, 2, 1, 4, 4, 5, 1) je 1, budući da se vrijednost 1 javlja četiri puta. Za razliku od drugih mjera centralne tendencije, jedan uzorak može imati više modova, jer skup podataka može imati dvije ili više mjera koje se "jednako najučestalije" javljaju! U takvim slučajevima je potrebno saopštiti sve takozvane *apsolutne modove*. Još je češće moguće da ćete imati više *relativnih* ili *lokalnih* modova, tj. više tačaka oko kojih se grupišu ostali podaci, pri čemu se one ne javljaju sa jednakom učestalošću, nego sa približno visokom učestalošću. Na primjer, u gore navedenom skupu, moglo bi se reći i da je vrijednost 4 relativni mod, jer se javlja tri puta. Usput, gore navedena distribucija bi predstavljala veoma pojednostavljeni primjer multimodalnih distribucija o kojima će biti više riječi kasnije u tekstu.

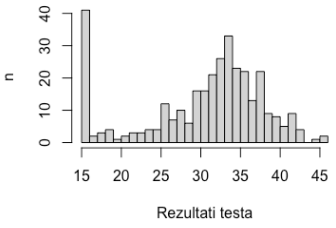
Koje su prednosti moda kao mjere centralne tendencije?

Mod može nekad biti zanimljiva mjera opisa varijable za veoma asimetrične distribucije, distribucije izrazito šiljatog oblika, varijable apsolutnog nivoa mjerenja, ordinalne kategoričke distribucije i distribucije sa više relativnih modova. Na primjer, ukoliko na grafikonu primijetimo da postoji više relativnih modova, to nam sugeriše da podaci sa uzorka predstavljaju različite poduzorke, odnosno različite populacije iz kojih dolaze. Takvu pojavu svakako treba istražiti, jer je korisno imati u vidu tipične vrijednosti koje karakterišu podgrupe. Prednost moda je i što uvijek

predstavlja vrijednost koja je realna, odnosno koju mogu imati objekti istraživanja što neće biti slučaj sa nekim drugim mjerama. U stručnim tekstovima se nekad navodi da je prednost to što mod donekle zadržava svoj smisao i u slučaju nominalnih kategoričkih podataka, ali to po meni nije nikakva prednost. Jednostavno, tada nema potrebe da govorimo o modalnoj vrijednosti, nego možemo prostije da kažemo koja je kategorija dominantna i da podatke prikažemo putem već prezentovanih tabela i grafikona.

Koje su mane moda?

Mod je izuzetno gruba deskriptivna mjera koja nema upotrebnu vrijednost u izvedenim statističkim formulama. Nadalje, kada je u pitanju velika većina dimenzionih varijabli koje imaju nijansirane mjerne skale sa mnogo teoretski mogućih vrijednosti, mod će da "skače" od uzorka do uzorka, te predstavlja nestabilnu karakterističnu mjeru. nekim slučajevima neka vrijednost može biti najčešća, a da pritom ne bude i najkarakterističnija, odnosno da se veća količina uzorka zapravo grupiše oko neke druge lokacije na uzorku (Slika 5.13). Sve u svemu, za opis empirijskih podataka uglavnom nećemo previše oslanjati na uvide koje dobijamo na osnovu modalne vrijednosti. Ponekad, mod će biti zanimljiva mjera za opis teorijski izvedenih distribucija.



Slika 5.13: Distribucija podataka gdje mod ne predstavlja grupu najčešću vrijednost.

Šta je medijana i kako je izračunavamo?

Medijana ili **središnja vrijednost** (engl. *median*) je isto tako vrlo jednostavna, zdravorazumska mjera centralne tendencije za koju se u statističkoj notaciji sreću različite oznake. Za vrijednosti na uzroku su one najčešće *Mdn*, *m* ili $\tilde{x}$ (*iks sa valovitom ili izvijenom crtom*), dok za parametar ne postoji standardna notacija, mada se nekad pojavljuje oznaka $\tilde{\mu}$ (*mi sa valovitom ili izvijenom crtom*). Vrijednost medijane je jednaka mjeri onog elementa koji se nalazi tačno u sredini niza podataka koji je prethodno poredan u rastućem ili opadajućem redoslijedu. Dakle, to je bukvalno srednja ili središnja vrijednost. Da bismo "ručno" odredili vrijednost medijane, neophodno je da izvedemo sljedeće korake:

1. Poredaćemo sve podatke iz skupa u rastućem ili opadajućem nizu (od najmanjem ka najvećem broju ili u suprotnom smjeru). U ovoj fazi ćemo uključiti svaki pojedinačan podatak, bez obzira da li je vrijednost unikatna u uzorku ili više objekata ima istu vrijednost. Dakle, ako sljedeći vektor

(2, 8, 1, 6, 1) koji ima 5 elemenata ćemo poredati kao (1, 1, 2, 6, 8).
2. Odredićemo poziciju elementa koji se nalazi tačno u sredini niza koristeći formulu $\frac{n+1}{2}$. Zapamtite, ovom formulom određujemo poziciju elementa koji predstavlja medijanu, a ne i samu vrijednost medijane!!! Ako varijabla ima samo 5 podataka, onda je pozicija traženog elementa $\frac{5+1}{2} = 3$.
3. Evidentiraćemo mjeru koju ima ispitanik koji se nalazi na središnjoj poziciji. U slučaju vektora (1, 1, 2, 6, 8) to je vrijednost 2, budući da je to treća vrijednost u nizu bilo da se krećemo od najniže ka više ili suprotnim smjerom. To je medijana.
4. Sve gore navedeno je važilo za slučaje kada imamo neparan broj mjera u varijabli. Kada imamo paran broj, onda ćemo pribjeći maloj korekciji. Recimo da smo prethodni vektor proširili za podatak 100, tako da je on sada (1, 1, 2, 6, 8, 100). I dalje ćemo poziciju središnjeg elementa tražiti korištenjem formule $\frac{n+1}{2}$. Međutim, sada ćemo kao rezultat dobiti vrijednost $\frac{6+1}{2} = 3.5$, što se nalazi na pola puta između 3. i 4. elementa u nizu. Medijanu ćemo tada izračunati tako što ćemo prvo identifikovati dvije mjere koje su neposredni susjedi središnje tačke: a u ovom slučaju to jesu 3. i 4. pozicija, odnosno vrijednosti 2 i 6. Zatim ćemo izračunati prosjek te dvije mjere $\frac{2+6}{2} = 4$, što će biti vrijednost medijane.

Koje su prednosti medijane?

Jednu od potencijalnih prednosti medijane smo već mogli da primijetimo na prethodnom primjeru gdje bismo mogli zaključiti da je medijana imuna na dejstvo ekstremnih skorova. Naime, ne bi igralo ulogu u njenom utvrđivanju to da li je najviša vrijednost varijable bila 10, 100 ili 1000, jer bi medijana uvijek bila jednaka vrijednosti 4. Ova otpornost ili robusnost je većem broju slučajeva njena prednost, pogotovo kada imamo male uzorke, te kada grafički prikazom već utvrdimo umjerenu ili snažniju asimetričnost distribucije. Nešto niže (na Slici 5.14) ćemo vidjeti zašto je medijana tada bolji reprezent grupe nego li je to mjera prosjeka, odnosno aritmetička sredina. Nadalje, u najvećem broju slučajeva - čak i kada na uzorku imamo paran broj ispitanika, medijana ima realne vrijednosti koje mjerimo na pojedinačnim ispitanicima. Uz to, postoji i nekoliko jednostavnijih tehnika statističkog zaključivanja koje se zasnivaju na medijani.

Nekad se navodi da je prednost medijane i to što se može ko-

ristiti na ordinalnim kategoričkim podacima kada su oni kodirani brojkama u rastućem nizu. Međutim, u takvim slučajevima je bolje da izračunamo tzv. interpoliranu medijanu (engl. *interpolated median*) koja donekle rješava probleme sa asimetričnošću distribucije kategorija. Za razliku od proste medijane koja uopšte ne mari za to koliko se u kojoj kategoriji nalazi slučajeva nego joj je bitno da dođe do središnjeg slučaja, dodatnom korekcijom u formuli interpolirana medijana uzima u obzir ukupnu distribuciju, odnosno broj slučajeva koji se nalazi i sa lijeve i sa desne strane u odnosu na kategoriju koja predstavlja medijanu. Formula glasi:

$$IMdn = Mdn + \frac{(n_{>} - n_{<})}{(2 * n_{=})}$$

Šta možemo da saznamo iz ove formule u kojoj *n_=* predstavlja broj ispitanika u kategoriji koja sadrži srednji element, *n_>* broj ispitanika u kategorijama većim od medijane, te *n_<* broj ispitanika u kategorijama ispod medijane? Kao prvo, vidimo da u lijevom dijelu formule imamo medijanu, a da desni dio formule predstavlja korekciju proste medijane koju dobijamo tako što identifikujemo grupu u kojoj se nalazi srednji element (drugim riječima, medijana je u slučaju ordinalnih kategoričkih varijabli ona vrijednost u kojoj se nalazi srednji element). Drugo, korekcija je takva da će rezultat biti na dimenzionoj skali, odnosno može da bude između vrijednosti kategorija (npr. 4.23 ili 3.01). Treće, na smjer i količinu korekcije utiče gornji dio izraza na način da što je veća asimetričnost distribucije i korekcija će biti veća. Ukoliko je broj ispitanika u kategorijama iznad medijane jednak broju ispitanika u kategorijama ispod medijane, onda nećemo ni imati korekciju medijane, budući da će vrijednost izraza $n_{>} - n_{<}$ biti 0. S druge strane, ukoliko je broj ispitanika veći u kategorijama koji su manji od medijane, onda ćemo imati korekciju u negativnom smjeru, odnosno formula će "oboriti" konačnu vrijednost medijane i obratno. Konačno, količina korekcije inverzno ovisi i o brojnosti središnje kategorije; što je više ispitanika u toj kategoriji to će korekcija biti manja.

Šta su mane medijane?

Kada postoji veliki broj jednakih vrijednosti medijana može da izgubi svoje prepoznatljivo svojstvo da dijeli uzorak po brojnosti elemenata na dva jednaka dijela. Sa parnim brojem ispitanika se može desiti da medijana izgubi i svojstvo da predstavlja realnu vrijednost. Interpolacijska korekcija jeste plus koji proširuje upotrebljivost medijane, ali medijanu nema smisla koristiti na di-

hotomnim i nominalnim kategoričkim varijablama. Nema previše smisla ni koristiti je kada imamo multimodalne distribucije. Nadalje, to što se medijana zapravo bira preko identifikacije pozicije, a ne direktnim proračunom smanjuje njenu upotrebu u kompleksnijim statističkim procedurama. Konačno, procjena parametarske vrijednosti medijane na osnovu rezultata sa uzorka je u prosjeku manje pouzdana od procjene parametra kada se služimo aritmetičkom sredinom. Ali da bismo shvatili prethodne dvije mane, moramo da se bolje upoznamo sa aritmetičkom sredinom.

Šta je aritmetička sredina i kako je izračunavamo?

Aritmetička sredina ili **prosječna vrijednost** (engl. *arithmetic mean, average*) je najpoznatija i najčešće korištena mjera centralne tendencije. Dakle, kada u svakodnevnom govoru upotrebljavamo riječ prosjek, obično mislimo na aritmetičku sredinu. Aritmetička sredina populacije se označava grčkim slovom μ (čitamo to kao *mi*), dok se aritmetička sredina uzorka označava latiničnim M ili sa $\bar{x}$ (*iks sa crtom* ili *iks nadvučeno*). Da bismo dobili aritmetičku sredinu potrebno je da saberemo vrijednosti svih pojedinačnih članova skupa i da sumu podijelimo brojem članova skupa. Taj jednostavni proračun, poznat nam iz opšteg obrazovanja, možemo predstaviti i naizgled komplikovanom formulom:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Značenje elemenata u formuli smo već objasnili u prvom poglavlju, pa ako imate probleme sa tumačenjem iste, sada je prilika da se vratite i podsjetite značenja simbola. Ovdje još treba navesti da se aritmetička sredina ili prosječna vrijednost nekad još naziva srednjom vrijednosti, međutim trebalo bi izbjegavati taj načelno netačan izraz, jer je medijanu ta koja predstavlja srednju vrijednost.

Koje su prednosti aritmetičke sredine kao mjere centralne tendencije?

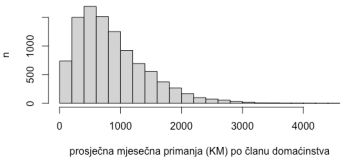
Kao prva prednost - ali koja se lako pretvara u manu - se obično navodi da je aritmetička sredina obuhvatna mjera. To znači da u njenom izračunavanju učestvuju baš svi podaci sa uzorka, a aritmetička sredina predstavlja tačku balansa svih mjera u

uzorku. Ovo *tačka balansa* znači da ukoliko saberemo sva pojedinačna odstupanja od aritmetičke sredine ($\Delta x_i = x_i - \bar{x}$; gdje veliko slovo delta označava odstupanje ili prirast) uvijek ćemo kao rezultat dobiti 0 ($\sum_{i=1}^{n} \Delta x_i = 0$). Direktno povezano sa tim svojstvom, a bitno za kasnije statističke tehnike, je i činjenica da je zbir kvadriranih odstupanja od aritmetičke sredine manji nego što se dobija za bilo koju drugu vrijednost ($\sum_{i=1}^{n}(x_i - \bar{x})^2 = min$). Uz to, u prosjeku gledano (ne radi se o igri riječima!), aritmetička sredina predstavlja najstabilniju popularnu mjeru centralne tendencije. Ovo znači da ćemo, kada uzmemo u obzir veliki broj slučajeva, napraviti manju grešku kada kao procjenu populacione centralne vrijednosti koristimo aritmetičku sredinu, u odnosu na to kada bismo koristili medijanu i mod. Povrh svega navedenog, mnogo drugih mjera i naprednih statističkih tehnika je direktno vezano za aritmetičku sredinu, jer se ona jednostavno algebarski definiše što nije slučaj sa medijanom i modom. To što je tako popularna mjera čini je najintuitivnijom za baratanje.

Šta su mane aritmetičke sredine?

Međutim, navedene krasote aritmetičke sredine važe ukoliko su ispunjeni određeni uslovi. Konkretno, dimenziona varijabla na kojoj je računamo bi morala biti u dovoljnoj mjeri simetrično distribuisana, trebalo bi da ima samo jednu izrazitu karakterističnu vrijednost, te ne bismo smjeli imati stršeće vrijednosti.

Na primjer, ukoliko ukoliko uvidimo da se podaci distribuiraju asimetrično na mjernoj skali, aritmetička sredina će ili potcijeniti ili precijeniti ono što je karakteristično za najveći broj pripadnika grupe. Pogledajte razlike između moda, medijane i aritmetičke sredine u čuvenom kontekstu izračunavanja uobičajenog mjesečnog primanja na Slici 5.14. Isto to će se desiti ukoliko u uzorku imamo jednu ili više ekstremnih mjera, budući da one mogu da distorziraju predstavu o čitavoj grupi, naročito kada su odstupanja od ostatka grupe visoka, te kada je uzorak mali. U takvim situacijama su bolji izbor medijana ili neka od korigovanih aritmetičkih sredina o kojima će biti riječi niže. Međutim, nema spasa za korištenje samo jedne aritmetičke sredine ukoliko smo na grafikonima primijetili da se podaci značajano grupišu na više od jedne lokacije, a što smo naveli da sugeriše postojanje više populacija. Ovo zaslužuje posebnu pažnju istraživača i ne dopušta nam da olako damo zaključke o jednoj grupi kao da ona zaista predstavlja relativno homogenu cjelinu.



Slika 5.14: Distribucija mjesečnih primanja za koju se dobijaju različite vrijednosti moda (670 KM), medijane (740 KM) i aritmetičke sredine (864 KM).

Uprkos uslova o dimenzionalnosti varijabli, treba navesti da iz

pragmatičnih razloga aritmetičku sredinu izračunavamo i za dihotomne i ordinalne kategoričke kada one imaju jednak broj kategorija (ali ne za nominalne i ordinalne kategoričke koje imaju različit broj kategorija!). Kada su u pitanju ordinalne kategoričke varijable sa jednakim brojem kategorija svi smo imali iskustva sa prosjecima školskih ocjena. Svaka pojedinačna ocjena je bila ordinalna kategorička, a kao rezultat smo dobijali kvazi-intervalnu varijablu koja je operacionalizovala našu savladanost gradiva - nakon čega nam je ocjena opet konvertovana u ordinalnu kategoričku varijablu, odnosno zaključnu ocjenu?! Što se tiče dihotomnih varijabli, ukoliko su njene kategorije kodirane sa 0 i 1, onda aritmetička sredina zapravo predstavlja proporciju (npr. proporciju tačnih odgovora). Ako je od 100 ispitanika njih 80 dalo tačan odgovor na neki zadatak, te ako smo tačne odgovore bodovali sa 1, a netačne sa 0, onda je prosječna vrijednost jednaka $80/100 = 0.8$ što predstavlja i proporciju tačnih odgovora.

Koje se još mjere centralne tendencije koriste u psihologiji?

Uz navedene mjere centralne tendencije, postoje još neke koje su više ili manje zanimljive za psihološku statistiku. Među njima se naročito izdvajaju tzv. otporne ili robusne aritmetičke sredine. Riječ otpornost se odnosi na otpornost na uticaj stršećih skorova, a ogleda se u tome da oni uopšte ne utiču na njeno izračunavanje. Budući da su po definiciji stršeći skorovi atipični i da mogu proizaći kao specifičnost datog uzorka, njihovim tretiranjem možemo dobiti vjerodostojniju procjenu populacijske vrijednosti.

Aritmetičku sredinu potkraćenog / potkresanog uzorka (engl. *truncated / trimmed mean*) računamo tako što ćemo najprije odbaciti jednak procenat uzorka sa svake strane rastućeg / opadajućeg niza, a onda na tako skraćenom uzorku izračunati aritmetičku sredinu. Najčešće ćemo odstranjivati između 5% i 20% vrijednosti sa svake strane. Treba napomenuti da je i medijana zapravo aritmetička sredina ekstremno potkraćenog uzorka od kojeg je odbačeno 50% krajnjih podataka sa obje strane distribucije. U svakom slučaju, ako se odlučimo za korekciju putem potkraćivanja uzorka, onda je pri prezentovanju rezultata neophodno navesti procenat ili proporciju odstranjenih slučajeva (npr. 5%-potkraćeni uzorak ili 0.05-potkraćeni uzorak). U notaciji se aritmetička sredina potkraćenog uzorka označava sa malim t u indeksu ($\bar{x}_t$), od engleskog *trimmed*.

Vinsorizovana aritmetička sredina (engl. *Winsorized mean*) je

zasebna varijanta potkraćene aritmetičke sredine. Pri njenom izračunavanju se, umjesto odbacivanja određenog postotka podataka koji se nalaze na krajevima distribucije, ti podaci zamjenjuju onim vrijednostima koji ostaju na njenim rubovima. Drugim riječima, tim podacima se dodjeljuju vrijednosti minimalne, odnosno maksimalne vrijednosti unutar skupa koji je ostao. Kako bismo naznačili da se radi o Vinsorizovanoj aritmetičkoj sredini u notaciji ćemo koristiti slovo w u indeksu ($\bar{x}_w$).

Na primjeru vektora sačinjenog od 10 različitih vrijednosti (1, 3, 3, 4, 5, 6, 6, 7, 7, 15), 10%-potkraćeni uzorak bi bio (3, 3, 4, 5, 6, 6, 7, 7), dok bi 10%-Vinsorizovani uzorak bio (3, 3, 3, 4, 5, 6, 6, 7, 7, 7). Dakle, 10%-potkraćeni uzorak izbacuje ukupno 20% krajnjih mjera, dok se u slučaju Vinsorizacije njihove vrijednosti zamjenjuju minimumom i maksimumom potkraćenog uzorka. Nakon što znamo čemu služe i kako se izvode nameću se logična pitanja: kojoj od ove dvije metode treba da damo prednost i koliki procenat vrijednosti treba da odstranimo? Odgovor je da sve zavisi od situacije. Oba postupka daju slične rezultate ukoliko koristimo isti procenat odbacivanja krajnjih rezultata. Aritmetička sredina potkraćenog uzorka je logičniji izbor, pogotovo u situacijama kada imamo veći broj atipičnih vrijednosti. S druge strane, Vinsorizacija se upotrebljava za neke druge proračune (mjere varijabilnosti i mjere povezanosti), pa je dobro biti upoznat sa njom. Promoteri ovih mjera (najpoznatiji je statističar Rand Wilcox), smatraju 20%-tne verzije veoma dobrim početnim izborom. Međutim, izuzetno bitno je prvo razmotriti prirodu ekstremnih skorova, čemu ćemo posvetiti dužnu pažnju kasnije u tekstu.

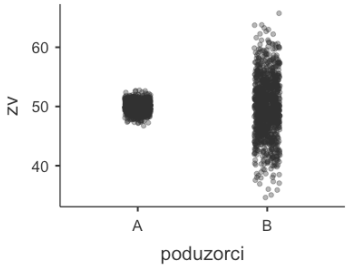
Usput, postoje i još neke, za psihologe mnogo manje interesantne mjere centralne tendencije. U poznatije spadaju geometrijska i harmonijska aritmetička sredina, te čitave porodice M- i R- mjera lokacije. Zasad je njihova upotreba toliko rijetka u psihologiji da je dovoljno i to što sam ih ovdje uopšte pomenuo.

Koji zaključak možemo donijeti o mjerama centralne tendencije?

Kada sebi i drugima opisujemo jednu dimenzionu varijablu neophodno je da najprije pogledamo relevantne grafikone, a zatim prikažemo sebi različite mjere centralne tendencije, prvenstveno aritmetičku sredinu i medijanu. Njihove vrijednosti bi morale biti saglasne ukoliko su raspodjele razumno simetrične i

ukoliko ne postoje ekstremni skorovi. Informacija o poklapanju ili vidnoj nesaglasnosti nam može dobro doći i kada čitamo tuđe izvještaje koji nam, najčešće radi uštede prostora, uopšte ne pružaju grafičke prikaze raspodjele varijabli. Ukoliko se njihove vrijednosti razlikuju, onda bismo trebali držati u vidu obje vrijednosti, detektovati šta je dovelo do razloga za to (asimetričnost i/ili ekstremni skorovi), te razmotriti upotrebu korigovanih aritmetičkih sredina ukoliko nam se čini da one mogu bolje da uopšte sliku uzorka. U kasnijem tekstu ćemo se upoznati sa dodatnim procedurama koje će nam pomoći da napravimo optimalne odluke prije nego se upustimo u statistiku zaključivanja. Kada su u pitanju ostali nivoi mjerenja, bilo bi dobro da za ordinalne kategoričke varijable izračunamo interpolirane medijane - ponovo nakon pregleda grafikona - te utvrdimo da li nam i aritmetičke sredine ili proste medijane mogu poslužiti za opis, ukoliko su distribucije simetrične. Nije potrebno uvijek posezati za komplikovanijom, manje poznatom procedurom, ukoliko su rezultati podjednaki.

Mjere varijabilnosti



Slika 5.15: Grafikon koji prikazuje dva poduzorka koji imaju jednake centralne vrijednosti ali različite varijabilnosti.

Mjere centralne tendencije pokazuju tipične vrijednosti za skup podataka, međutim to nije dovoljno da bismo adekvatno opisali varijable koju upoznajemo putem grafikonima. Navesti samo mjere centralne tendencije bi otprilike bilo isto kao da smo čitavu raspodjelu, umjesto histogramom, opisali samo jednim, najvišim stupcem. Koliko je to pogrešno možemo da shvatimo uz pomoć Slike 5.15. Za poduzorke A i B su dobijene jednake mjere centralne tendencije (M = Mdn = 50) , ali su te dvije grupe zapravo veoma različite - u pogledu širine distribucije. Konkretno, u jednoj grupi su ispitanici međusobno sličniji i imaju skorove bliske aritmetičkoj sredini, dok je druga grupa daleko "šarenija" gdje imamo i ispitanike koji imaju visoke, ali i niske skorove. Mjere varijabilnosti nam služe da ekonomično iskažemo stepen unutargrupne različitosti ili raznolikosti. U literaturi ćemo još sretati i nazive mjere raspršenja ili, visokoparnije, mjere disperzije.

Kao i mjera centralne tendencije, mjera varijabilnosti ima mnogo, a mogu se podijeliti na tri vrste: one koje se zasnivaju na rasponu između dvije tačke, one koje predstavljaju tipično odstupanja od mjere centralne tendencije, te opšte mjere raznovrsnosti podataka.

Sljedeće mjere zaslužuju nešto detaljniji opis:

- raspon
- interpercentilni rasponi
- standardna devijacija
- varijansa
- suma kvadrata odstupanja
- Vinsorizovana standardna devijacija
- srednje apsolutno odstupanje
- indeks raznovrsnosti

Raspon

Mjere **raspona / opsega** (engl. *range*) su najosnovnije mjere varijabilnosti. Među njih ubrajamo i empirijski **minimum** i **maksimum**, tačke na mjernoj skali do kojih doseže raspodjela varijable na uzorku. Kada znamo miminum i maksimum lako ćemo izračunati raspon, odnosno ono što pod tom riječju podrazumijevamo u užem smislu: $raspon = x_{max} - x_{min}$. Međutim, raspon je toliko slabo informativna mjera, da nema potrebe da je saopštavamo. Umjesto raspona, uvijek ćemo pažljivo razmotriti i saopštiti najnižu i najvišu vrijednost dobijenu na uzorku, a onaj ko to želi može jednostavno da izračuna o kolikom rasponu se radi.

Interpercentilni rasponi

Kada kažemo **interpercentilni rasponi** (engl. *interpercentile range*) mislićemo na više njih. Oni predstavljaju modifikaciju gore opšte mjere raspona, isto onako kao što aritmetičke sredine potkraćenih uzoraka modifikuju aritmetičku sredinu. Da bismo dobili neki interpercentilni raspon potrebno je da eliminišemo sa obje strane raspodjele fiksni procenat podataka. Ono što nas interesuje jesu tačke minimuma i maksimuma na tako potkraćenom uzorku koji omeđuju određen procenat podataka. Na primjer, 95%tni raspon će uključivati 95% podataka koji se nalaze u središnjem dijelu distribucije, nakon što je sa oba kraja odstranjeno po 2.5% podataka. Naravno, moguće bi bilo umjesto minimuma i maksimuma prikazati raspon putem formule $IPR_{95} = P_{97.5} - P_{2.5}$, ali mi to nećemo raditi, jer ako bismo htjeli da napravimo informativnu uštedu i izložimo jednu umjesto dvije informacije, smanjićemo uvid u prirodu podataka. Koliki procenat slučajeva je preporučivo obuhvatiti? Vidjećemo da se za potrebe statistike zaključivanja naročito praktikuju 90%tni, 95%tni i 99%tni interpercentilni rasponi, a kada su u pitanju bazični opisi distribucije varijabli najčešće se radi o 50%tnom

interpercentilnom rasponu, odnosno intervalu unutar kojeg se nalazi prosječnih 50% ispitanika. Taj raspon se još rogobatno naziva interkvartilni raspon i označava sa *IQR*. Na pitanje zašto sad ta nova riječ kvartili i šta su uopšte percentili i postoji li razlika u odnosu na procente dobićete odgovore malo kasnije u tekstu.

Standardna devijacija

Različite mjere raspona jesu jedan način da opišemo sebi i drugima količinu raznolikosti u skupu. Međutim, postoji i set drugih mjera čiji je cilj da varijabilnost opišu tako što se izračunava prosječno odstupanje od neke centralnih vrijednosti. Ako se malo zamislimo, ovo ne zvuči baš intuitivno: šta bi to bila prosječna razlika od prosjeka?! Kako god, u praksi funkcioniše. Istina je da nam takve mjere neće biti toliko korisne kada želimo da se upoznamo sa pojedinačnom varijablom, za tu svrhu su zaista mjere raspona slikovitije. Mjere prosječnih odstupanja će nam dobro doći onda kada želimo da uporedimo dvije ili više grupa, a neke od njih će dobiti vodeće uloge u statistici zaključivanja.

Među mjerama prosječnog odstupanja **standardna devijacija** (engl. *standard deviation*) je ubjedljivo najpopularnija. Simbol za njenu populacionu vrijednost je malo grčko slovo sigma (σ), dok se mjera uzorka označava malim latiničnim slovom *s* ili skraćenicom *SD*. U njenom imenu riječ standard stoji za prosjek, ali i za specifičnu ulogu koju ima kada više varijabli sa različitih skala poredimo na jednoj zajedničkoj mjerni skali - tada će standardna devijacija postati mjerna jedinica koja predstavlja standard. S druge strane, riječ devijacija znači odstupanje, a ovdje se odnosi na odstupanje od aritmetičke sredine. Moramo imati na umu da se prosječna odstupanja iskazuju na istoj mjernoj skali kao i sirovi skorovi iz kojih se izračunavaju. Odnosno, ako su podaci prikazani u centimetrima (ili IQ jedinicama) onda će i standardna devijacija biti u centimetrima (ili IQ jedinicama).

Postupak izračunavanja standardne devijacije Postoje dvije konceptualne formule za izračunavanje standardne devijacije koje bismo trebali znati. Prva među njima se koristi za izračunavanje standardne devijacije na populacionim podacima i vrlo rijetko ćemo je u praksi koristiti:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}}$$

Druga formula je zadana formula u statističkom softveru i koristimo je na podacima iz uzorka:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Već biste to morali znati, ali da ponovimo. U gore navedenim formulama x_i predstavlja pojedinačni podatak, dok veliko grčko slovo sigma Σ predstavlja zbir svih odstupanja (od prvog do n-tog člana skupa). Da bismo "pješke" izračunali standardnu devijaciju na podacima sa uzorka potrebno je da izvedemo sljedeće korake, a prateći prvenstvo matematičkih operacija kojeg smo se podsjetili u prvom poglavlju:

1. Za svaku mjeru u uzorku ćemo izračunati odstupanje od aritmetičke sredine (devijaciju): $\Delta x_i = x_i - \bar{x}$
2. Kvadriraćemo svako pojedinačno odstupanje: Δx_i^2
3. Sabraćemo sva kvadrirana odstupanja i dobićemo **sumu kvadriranih odstupanja** (engl. *Sum of Squares*, *SS*): $SS = \sum_{i=1}^{n} \Delta x_i$
4. Podijelićemo sumu kvadriranih odstupanja sa brojem ispitanika umanjenim za jedan kako bismo dobili **prosjek kvadriranih odstupanja** (engl. *mean squared deviation*), koji se još popularno naziva **varijansa** (engl. *variance*, s^2): $MSD = s^2 = \frac{SS}{n-1}$
5. Izvadićemo korijen iz količnika dobijenog u prethodnom koraku i dobićemo standardnu devijaciju: $s = \sqrt{s^2}$

Putujući kroz ovaj postupak smo nabasali na još dvije mjere varijabilnosti koje imaju važnu ulogu u statistici zaključivanja. Obje se izražavaju na kvadriranoj mjernoj skali i kao takve nam neće pružati direktne uvide u prirodu podataka, pa ćemo se njima vratiti tek kad se budemo bavili statistikom zaključivanja. Ostalo je još da shvatimo zašto je formula za standardnu devijaciju takva kakva jesta i zašto u praksi koristimo formulu koja ima $n-1$ u nazivniku.

Logika formule je sljedeća: kvadriranje, koje naknadno prati korijenovanje, nam služi da bismo uopšte mogli izračunati prosječno odstupanje. Podsjećanja radi, zbir svih odstupanja od aritmetičke sredine je jednak nula, tako da je neophodno da na neki način obezbijedimo da se odstupanja međusobno ne

ponište. Što se tiče razlike između formula, razlog za korištenje formule sa *n-1* jeste da ta formula daje, u prosjeku gledano (dobro ste pročitali!), tačnije procjene populacijske standardne devijacije. Konkretno, $n - 1$ će dati veće vrijednosti prosječnog odstupanja nego n u nazivniku. Drugim riječima, usvajajući tu korekciju očekujemo da bi raznolikost u stvarnosti (tj. populaciji) morala biti nešto malo veća nego što ćemo je uočiti na komadiću stvarnosti (tj. uzorku). Razlike između dobijenih vrijednosti pomoću dvije formule će se smanjivati sa povećanjem uzorka (npr. razlika između 5 i 6 u nazivniku je proporcionalno veća nego između 1005 i 1006), odnosno što je uzorak veći sve više sliči populaciji - reprezentativnije predstavlja njenu varijabilnost. I da, ako želite da se izdignete iznad statističkog plebsa možete $n - 1$ nazvati svojim stručnim nazivom Besselova korekcija, kako to rade nadobudni statističari.

Zainteresovanim među vama preporučujem da pogledate neki od YouTube instruktivnih videa koji o tome govore (npr. https://www.youtube.com/results?search_query=n-1+standard+deviation+) ili da koristitite sljedeći ilustrativni applet (programčić) ukoliko se nekako uspijete izborite sa podešavanjima Java zaštite na vašim kompjuterima: https://www.uvm.edu/~dhowell/fundamentals8/SeeingStatisticsApplets/N-1.html

Kada ima smisla koristiti standardnu devijaciju? S obzirom na to kako se izvodi, očito da je ima smisla koristiti samo onda kada ima smisla koristiti i aritmetičku sredinu. A to bi značilo da bi raspodjela podataka morala biti relativno simetrična - tolerisaćemo određenu asimetriju - te da nemamo atipične vrijednosti na uzorku. U slučajevima kada su oni prisutni, a stalo nam je do toga da koristimo standardnu devijaciju, treba da razmotrimo korištenje standardne devijacije izračunate na Vinsorizovanim podacima (s_w).

Srednje apsolutno odstupanje

Standardna devijacija nije jedina mjera prosječnog odstupanja. **Srednje apsolutno odstupanje** (engl. *median absolute deviation*) je mjera bazirana na medijani čija svrha je da se prikaže mjera odstupanja koja je otporna na ekstremne skorove. Formula joj je vrlo jednostavna $MAD = Mdn(|x_i - \tilde{x}|)$. U prevodu to je medijana apsolutnih odstupanja pojedinačnih skorova od medijane na uzorku. Devijacije se, dakle, odnose na odstupanja od medijane, a ne aritmetičke sredine. Problem poništavanja odstupanja - koje ionako ne bi bilo potpuno osim ukoliko medijana nije jednaka

aritmetičkoj sredini - se ovdje elegantno rješava apsolutnim vrijednostima, a srednja vrijednost po veličini poredanih odstupanja predstavlja traženu mjeru. Na vrlo kratkom primjeru, ako imamo tri mjere 1, 2 i 5 medijana je 2. Apsolutna odstupanja su 1 (1-2), 0 (2-2) i 3 (5-2). Njihova medijana je 1, budući da je niz apsolutnih odstupanja jednak 0, 1, 3.

Kao što je to bio slučaj i za standardnu devijaciju, za procjenu populacione vrijednosti se češće koristi verzija sa korekcijom. Korekcija se ogleda u tome što se MAD pomnoži sa nekom konstantom. Ukoliko pretpostavljamo da varijabla u populaciji ima zvonastu raspodjelu onda se za vrijednost pondera uzima 1.4826 kako bi vrijednost bila što bliža standardnoj devijaciji za istu distribuciju. Ta korigovana verzija će se u nekim statističkim programima nazivati robusnim srednjim apsolutnim odstupanjem (engl. *robust median absolute deviation*). Baš kao i za standardnu devijaciju, korekcijom se izražava razumno uvjerenje da će podaci sa uzorci potcijeniti ukupnu varijabilnost u populaciji. Sve u svemu, obje ove mjere nam mogu pomoći kada budemo upoređivali varijabilnost više grupa, ali su u odnosu na mjere raspona ove mjere preškrte kada želimo da shvatimo prirodu fenomena.

Koeficijent varijacije

Ostalo je još da razmotrimo mjere koje pokušavaju izraziti relativnu količinu raznovrsnosti. **Koeficijent varijacije** (engl. *coefficient of variation*, CV) je najpoznatija među njima i sreće se u različitim programima, ali ima mizernu upotrebljivost u psihologiji. Izračunava se kao omjer standardne devijacije i aritmetičke sredine ($CV = s/\bar{x}$). Ima ga smisla računati samo kada su u pitanju omjerni i apsolutni nivoi mjerenja koji imaju apsolutnu nulu i definisanu jedinicu mjerenja.

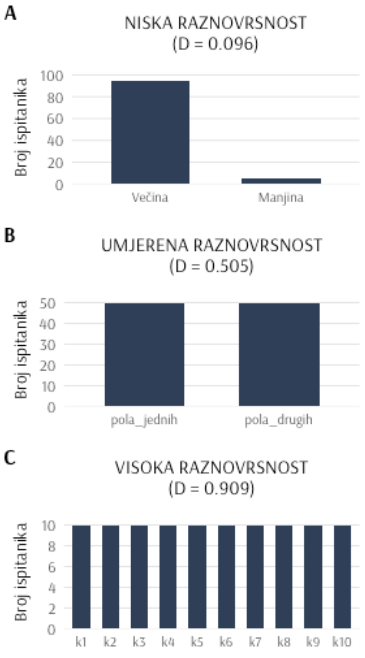
Simpsonov indeks raznovrsnosti

Postoji niz drugih mjera koje nemaju ovakva ograničenja, štaviše prevashodno su namijenjene za kategoričke varijable, ali se mogu koristiti i za dimenzione varijable. Jedna od jednostavnijih i najupotrebljavanijih u različitim naukama od biologije do lingvistike je M1 indeks raznovrsnosti. Zanimljivo je da ima vrlo raznovrsna imena - čak osam dodatnih se spominje na Wikipediji, od kojih je najpoznatije ime **Simpsonov indeks raznovrsnosti** (engl. *Simpson's index of diversity*, D). Formula je

sljedeća:

$$D = 1 - \sum_{i=1}^{j} \frac{n_j(n_j - 1)}{n(n-1)}$$

U ovoj formuli j označava jedinstvenu vrijednost na dimenzionoj skali, odnosno posebnu kategoriju kada su kategoričke varijable u pitanju. To znači da n_j može označavati broj osoba plave kose na uzorku ako je to jedna od kategorija, odnosno broj ispitanika koji ima $IQ = 120$ ako su u pitanju dimenzione varijable. Vrijednost D indeksa će se povećavati što je veća raznovrsnost, i smanjivati što je raznovrsnost manja. Mjerna skala za D-indeks se kreće od 0 do 1, a evo i ilustracije kada se mogu dobiti krajnji rezultati. Vrijednost će biti jednaka 0 ako svi ispitanici imaju identičnu vrijednosti, jer to znači da je $n_j = n$, iz čega slijedi da bi desni dio formule imao vrijednost 1, te da bismo onda imali $D = 1 - 1$. Teorijski maksimum 1 se dostiže kada je svaka vrijednost ili kategorija predstavljena samo jednim ispitanikom. U takvom slučaju će svako $n_j - 1$ biti jednako 0, što znači da će i suma desne strane biti jednaka 0. Logično je da je već u startu, sem u vrlo rijetkim slučajevima, raznovrsnost veća za dimenzione nego za kategoričke varijable jer postoji veći broj mogućih vrijednosti. Iz didaktičkih razloga, na Slici 5.16 su prikazane tipične vrijednosti niskih, srednjih i visokih D-indeksa za tri kategoričke varijable.



Slika 5.16: Ilustrativni primjer varijabli sa niskom, umjerenom i visokom raznovrsnošću.

Simpsonov indeks raznovrsnosti se može tumačiti kao vjerovatnoća da dva nasumično odabrana elementa skupa imaju različitu vrijednost, odnosno pripadaju različitim kategorijama. Ovo intuitivno tumačenje i činjenica da je mjerna skala jasno ograničena su značajne prednosti u odnosu na neke druge mjere koje se možda i češće spominju. Međutim, uprkos ovim relativnim prednostima nema mnogo situacija u kojima je ovaj indeks neophodan, sem u slučajevima kada imamo mnogo varijabli i želimo veoma brzo da detektujemo one koje imaju vrlo nisku varijabilnost.

Šta možemo zaključiti na osnovu mjera varijabilnosti?

Ponovo ću naglasiti koliko je važno izraditi grafikone i učiti iz njih, jer nam oni daju obuhvatan uvid u varijabilnost, čak i ako te informacije nisu precizno matematičko-jezički izražene. Simpsonov indeks nam može poslužiti za brzi pregled ukupne varijabilnosti u čitavoj bazi podataka, odnosno detektovanje snižene varijabilnost što može biti posebno bitno za kategoričke varijable. Za ordinalne kategoričke i dimenzione varijable je obavezan korak u analizi utvrđivanje empirijskog minimuma i maksimuma, jer je to na kraju krajeva i dio higijene vezane za

validaciju podataka. Ispitaćemo da li su minimum i maksimum očekivani, koliko su udaljeni od teorijski najmanjih i najvećih vrijednosti, te ako su ih dosegli, da li zapažamo takozvane efekte poda i plafona. Interkvartilni ili neki drugi interpercentilni raspon možemo izračunati kako bismo stekli uvid u to gdje nam se nalazi većina "prosječnih" ispitanika. Analogno poređenju aritmetičke sredine i medijane možemo da uporedimo standardnu devijaciju i medijanu apsolutnog odstupanja. Nesaglasnost će nam potvrditi probleme sa korištenjem proste standardne devijacije, te možemo razmotriti da li nam i koliko korigovana Vinsorizovana standardna devijacija daje saglasnije rezultate.

Mjere pozicije u raspodjeli

Postoji nekoliko mjera pozicije u raspodjeli koje opisuju položaj određenog skora u ukupnoj raspodjeli varijable. Kao što ćete vidjeti, ove mjere imaju niz funkcija. Osim što nam mogu poslužiti za opis raspodjele neke varijable, koristićemo ih pri psihološkom testiranju, a posebnu ulogu će imati u statistici zaključivanja.

Rang

Rang (engl. *rank*) je najjednostavniji pokazatelj pozicije u distribuciji. Rangiranje je prosta transformacija skorova koju izvodimo tako što ispitanike na uzorku poredamo po izraženosti varijable od interesa - u opadajućem ili rastućem nizu - i svakom od njih dodijelimo broj koji zauzima u takvom redoslijedu. S tim u vezi, dodjeljivanje brojeva možemo izvesti tako da broj 1 damo objektu sa najviše izraženim atributom - što bi bilo karakteristično za sportska takmičenjima - ili onome sa najniže izraženim atributom.

Prosti rangovi su lako shvatljiva mjera, ali postoji nekoliko mana njihovog korišćenja. Kao prvo, rangovi - kao i ostale mjere pozicije u nizu - nam ništa ne govore o veličini razlika između ispitanika. Na primjer, na testu znanja iz statistike se može desiti da je najbolje urađeni rad osvojio 40 bodova, a da su drugi i treći osvojili 21 i 20 bodova. Nadalje, ukoliko ne znate ukupan broj ispitanika u grupi ne možete adekvatno interpretirati značenje dobijenog ranga. Primjera radi, rang 30 na testu inteligencije znači dvije potpuno drugačije stvari ukoliko u uzorku imamo 30 ispitanika i ukoliko ih imamo 30 000. Takođe, možemo naići i na problem ukoliko nismo dobili jasnu instrukciju da li niži broj ranga označava viši ili niži intenzitet atributa. Konačno, postoji

i problem jednakih ili "vezanih" rangova. Naime, dva ili više ispitanika mogu postići isti skor i tada je potrebno odlučiti se po kojem osnovu će se dodijeliti rangovi tim ispitanicima. Najčešće su tada izračunava uprosječeni rang, ali može se i dodijeliti rang koji označava najniže ili najviše mjesto koje ispitanici dijele. Uprkos svih navedenih mana, rangiranje je veoma važno za statistiku jer predstavlja osnovu mnogih postupaka statistike zaključivanja koji su imuni na stršeće skorove i probleme u obliku distribucije.

Percentilni rang

Percentilni rang (engl. *percentile rank*) je mjera koja rješava neke probleme prostih rangova. Mjerna skala koja se koristi je procentna skala $(0-100)$ na kojoj se pozicioniraju objekti na osnovu postignutog skora. Percentilni rang je izuzetno važna mjera za psihologe u praksi. Naime, nakon testiranja normiranim psihološkim instrumentima, potrebno je da sebi i klijentima navedemo gdje se ispitanik nalazi u odnosu na referentnu grupu kojoj pripada, a procentna skala je pogodna za to. Na primjer, ako je ispitanik na skali emocionalne stabilnosti dobio skor 10.2 koji odgovara percentilnom rangu 90 (ovo bi se pisalo $PR_{(10.2)} = 90$), to bi značilo da samo 10% osoba iz ciljne grupe ima viši skor od te osobe. Kao i prosti rang, percentilni rang možemo dobiti na više načina ukoliko u uzorku imamo vezane rangove. Osnovna formula, kojom se kumulativna frekvencija konvertuje u percentilni rang je sljedeća:

$$PR_{x_i} = \frac{cf(x_i)}{n} * 100$$

Naglašavam, formula podrazumijeva da se radi o uzlaznom, kumulativnom rastu gdje viši rang označava veći intenzitet mjernog svojstva. Prema tome, maksimalni percentilni rang 100 može dobiti osoba koja je postigla najviši rezultat budući da je tada $rang(x) = n$. Kada više ispitanika ima identičan skor, moguće je koristiti i korigovanu formulu koja kumulativne frekvenciju za datu vrijednost x, umanjuje za polovinu javljanje te vrijednosti u uzorku:

$$PR_{x_i} = \frac{cf(x_i) - n_x/2}{n} * 100$$

Mala napomena: tačnija interpretacija percentilnog ranga bi bilo da on pokazuje od koliko drugih ispitanika osoba ima veći *ili jednak skor* na datoj varijabli, što pogotovo važi za prvu formulu.

Radi se o maloj korekciji, ali je možete imati na umu. Usput, psiholozi u praksi uz psihološke instrumente dobijaju već unaprijed izračunate percentilne rangove za moguće rezultate na psihološkom ispitivanju. Oni se izračunavaju u sklopu procedure normiranja psiholoških testova. Na tzv. normativnom uzorku, koji bi trebalo da je razumno reprezentativan za populaciju, se dobijeni rezultati prevode u percentilne rangove.

Linearno transformisani skorovi

Rangovi i percentilni rangovi nisu jedina transformacija skorova koja se koristi za određivanje pozicije ispitanika na novoj mjernoj skali. Za razliku od percentilnih rangova koji mijenjaju oblik distribucije varijable budući da se oslanjaju na kumulativnu učestalost, a ne na skorove, postoji niz popularnih transformacija gdje se mjerna skala mijenja, ali raspodjele ostaju identične. Takve transformacije nazivamo linearnim transformacijama. Već smo ranije vidjeli ilustraciju toga šta se dešava kada poduzmemo nelinearne i linearne transformacije.

Jedinično reskaliranje ili **jedinična normalizacija** (engl. *unity rescaling* / *normalization*) pretvara izvornu empirijsku mjernu skalu na skalu gdje je minimalna vrijednost 0, a maksimalna vrijednost 1. Ova vrsta transformacije omogućava poređenje među vrijednostima na različitim varijablama i u velikoj mjeri ujednačava konkretne vrijednosti prosječnog odstupanja za različite varijable, mada ukupna varijabilnost unutar varijable ostaje ista. Te osobine su bitne za neke napredne statističke metode kada se analizira više varijabli odjednom. Jedinično skaliranje se vrši pomoću formule:

$$x_i^1 = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Centriranje (engl. *centering*) ima nešto drugačiji cilj u odnosu na jedinično skaliranje. I prosječna i ukupna varijabilnost ostaju iste i skala se ne ograničava na krajevima. Jedino što se mijenja jeste pozicija ispitanika na brojevnoj pravoj. Sve vrijednosti se oduzimaju od aritmetičke sredine (mada je moguće centrirati ih i na drugim vrijednostima), tako da aritmetička sredina dobija vrijednost 0 za datu varijablu. Sjećate se da je ovo matematički isto što i devijacija, odnosno odstupanje od aritmetičke sredine. Poenta centriranja varijable jeste da sve ispodprosječne vrijednosti imaju negativni predznak, a iznadprosječne vrijednosti pozitivan. Centriranje je važan međukorak sljedeće transformacije koju opisujem, ali se i samostalno vrši u sklopu nekih statističkih procedura.

Formula je:

$$x_i^c = x_i - \bar{x}$$

Standardizovane vrijednosti ili **z-skorovi** (poznati i pod imenima standardni ili sigma skorovi; engl. *standardized values* ili *z-scores*) idu korak dalje od centiranja. Ne samo da će aritmetičke sredine biti jednake 0 (i time centrirane), nego će vrijednosti dobiti novu mjernu jedinicu. Jedinica nove skale će biti standardna devijacija za datu varijablu. Sjećate se da smo ovo najavili ranije? Konkretno, da bismo dobili z-skor od dobijenog skora ćemo oduzeti aritmetičku sredinu, a zatim ćemo dobijenu razliku podijeliti standardnom devijacijom uzorka:

$$z_i = \frac{x_i - \bar{x}}{s}$$

Na primjer, na originalnoj IQ mjernoj skali aritmetička sredina iznosti 100, a standardna devijacija je jednaka 15. Ukoliko ispitanik postigne skor 115, njegov z skor je jednak 1, jer je $(115 - 100)/15 = 1$. S druge strane, ispitanik koji postigne skor 85 ima standardizovani skor jednak -1, budući da je (85-100)/15 = -1. Da ponovimo, za z-skali uvijek važi da je $\bar{x} = 0$, kao i da je $s = 1$. Iako z-skala nije ograničena minimumom i maksimumom, ona ipak ima napredna svojstva u odnosu na jedinično skaliranje i centriranje, budući da su aritmetičke sredine i standardne devijacije savršeno uporedive među različitim varijablama. Međutim, vidjećemo da uloga z-skale postaje optimalna kada podaci u populacije raspodjeljuju normalno, odnosno prate zvonastu distribuciju.

T-skorovi Bliski srodnici, ili bolje rečeno, direktni potomci z-skorova su T-skorovi. Bitno je da ih znate jer se rezultati psiholoških testova često navode i u tom obliku. Zašto potomci, jasno je iz formule:

$$T_i = 50 + 10 * z_i$$

Dakle, radi se o linearnoj transformaciji z-skorova. Aritmetička sredina se pomiče sa 0 na 50, a standardna devijacija neće biti 1 nego 10. Drugim riječima $T = 50 \equiv z = 0$, dok je $T = 60 \equiv z = 1$. Inače, T-skorovi su navodno dobili ime po dva čuvena psihologa, Termanu i Thorndikeu, a razlog za njihovo uvođenje je bio to što su z-skorovi dati u obliku koji je potežak za tumačenje. Em imaju decimale, em mogu imati negativne vrijednosti, što je kako za laike, tako i za psihologe u praksi komplikovanije od skale centrirane na tačci 50 sa koracima odstupanja od 10 jedinica. Kao i z-skorovi, T-skorovi dobijaju na punom sjaju, sa dodatnim značenjima, onda kada imamo dokaze da je raspodjela skorova razumno zvonasta.

Kvantili

Za razliku od gore navedenih transformacija konkretnih skorova, imamo i posebne statističke mjere koje skup ispitanika dijele u grupe od kojih svaka ima jednaki broj elemenata. Sve zajedno ćemo ih zvati **kvantili** (engl. *quantile*), a pojedinačni elementi tog skupa su percentili, tercili, kvartili, decili, kvintili…

Počećemo od **percentila** (engl. *percentile*, *P*) koji su već bili pominjani ranije u tekstu. Percentil je mjesto na mjernoj skali ispod kojeg se (manji broj autora kaže, uključujući i to mjesto) nalazi određeni procenat distribucije. Teoretski raspon percentila se tako kreće od 1 do 100. Iz navedene definicije slijedi da je medijana načelno jednaka pedesetom percentilu. S obzirom na to da se percentili označavaju se velikim latiničnim slovom P pri čemu u indeksu stoji traženo mjesto na procentnoj skali, onda bismo to iskazali ovako: $Mdn = P_{50}$.

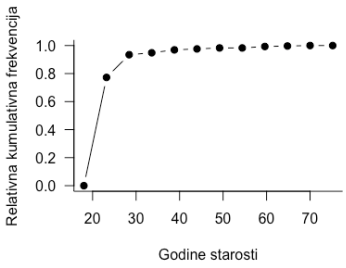
Ono što je problematično sa percentilima, pogotovo na manjim uzorcima, jeste da jedan te isti skor može istovremeno predstavljati više percentila (npr. i P_{33} i P_{50}i P_{67}), što je vidljivo u situaciji ako vektor čine sljedeće mjere: 1, 3, 3, 3, 5. Ovaj problem dijeljenja iste pozicije - kojeg imaju i rangovi i percentilni rangovi - se nekad može riješiti izračunavanjem percentila na različite načine (npr. pogledajte stranicu http://onlinestatbook.com/2/introduction/percentiles.html) Međutim, problem izračunavanja percentila nije od velikog praktičnog značaja i nije potrebno da se njime zamarate. Kada imate dovoljno veliki broj ispitanika (npr. više od 100) i kada u uzorku nemate mnogo jednakih vrijednosti, različite formule će vam davati iste ili približno iste vrijednosti. Zašto biste trebali znati nešto o percentilima? Ne samo da ćete nekad htjeti izračunati interpercentilne raspone, nego ćete u nekim radnim situacijama htjeti da postavite granicu od devedesetog ili nekog drugog percentila koju kandidati treba da prebace svojim postignućem kako biste ga dalje proslijedili (npr. na intervju za posao).

Osim percentila, upotrebljavaju se još i tercili (dijele distribuciju na tri dijela), kvartili i decili (dijele distribuciju na deset dijelova). Kvartili (engl. *quartile*) se označavaju latiničnim slovom *Q* i predstavljaju tačke na mjernoj skali koje dijele skup na 4 jednaka dijela. Postoje tri kvartila; donji kvartil Q_1 i gornji kvartil Q_3 su isto što i P_{25} i P_{75} , dok je srednji kvartil $Q_2 = P_{50} = Mdn$. Interkvartilni raspon smo već spominjali pod imenom 50%tni interpercentilni raspon: $IQR = Q_3 - Q_1$. Kvartili se koriste za konstrukciju posebne vrste grafikona, dosta prisutne u psihologiji, koji se nazivaju **kutijasti grafikoni** (engl. *boxplot*) koje ćemo pojasniti

odmah nakon što pojasnimo grafikon kumulativnih frekvencija koji je vezan za sve oblike kvantila i percentilne rangove.

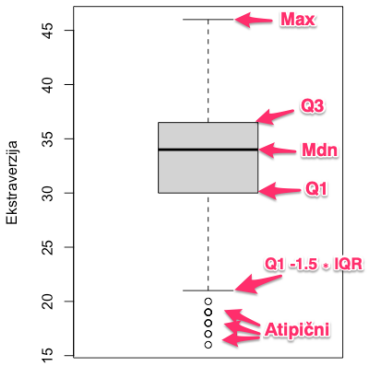
Grafički prikazi zasnovani na mjerama pozicije u nizu

Grafikon kumulativnih frekvencija



Slika 5.17: Grafikon kumulativnih frekvencija za varijablu Godine starosti.

Grafikon kumulativnih frekvencija, poznat i pod nazivom ogiva (engl. *cumulative frequency graph, ogive*) je još jedna vrsta grafikona koja ima vrlo specifičnu funkciju: on nam otkriva vezu između skorova na skali (obično datih na x-osi) i percentila / percentilnih rangova (na y-osi). Njegova primjena je relativno česta u psihometriji, a vidjećemo kasnije da će nam dobro doći i kada želimo da uporedimo dva ili više poduzoraka. Kada prezentujemo univarijatne distribucije možemo ga i kombinovati sa drugim vidovima grafikona kao što su histogrami ili tačkasti grafikoni. Na Slici 5.17 možemo lako da uočimo da značajnu većinu (nešto ispod 80%) ispitanika u datom uzorku čine ispitanici mlađi od 23-24 godine, a da ispitanici mlađi od 30 godina čine značajno više od 90% ispitanika.

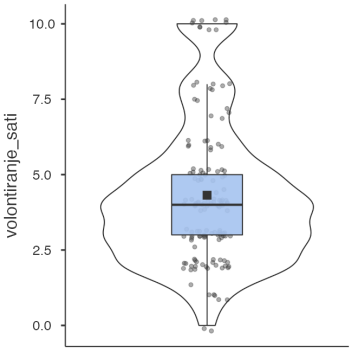


Slika 5.18: Kutijasti grafikon sa pojašnjenjem ključnih tačaka.

Kutijasti grafikon (engl. *boxplot* ili *box-and-whisker plot*) je grafikon koji ima za cilj da prikaže pet važnih tačaka empirijske distribucije, a to su najniža i najviša vrijednost na uzorku, te tri kvartila. Kao što vidite na Slici 5.18, pet ključnih tačaka se obilježava poprečnim linijama, s tim da su linije Q_1 i Q_3 povezane tako da čine pravougaona površinu koju presijeca Mdn, dok se između donjeg i gornjeg kvartila i krajnjih vrijednosti crta tzv. T-linija. Zapravo je ovo opis samo najpopularnije verzije kutijastog grafikona. Naime, ponekad se na grafikon dodaje i lokacija aritmetičke sredine (koristeći X ili neki drugi simbol), te se obično posebno prikazuju atipično visoke ili niske mjere (npr. simbolima **o** i/ili *). U kontekstu iscrtavanja kutijastog grafikona atipične mjere se definišu kao one koje su manje od $Q_1 - 1.5 * IQR$ ili veće od $Q_3 + 1.5 * IQR$. Dodatne modifikacije kutijastog grafikona ćemo spominjati kada budemo poredili dva ili više poduzoraka. Boxplot grafikoni imaju niz mana: ne prikazuju grupisanje podataka unutar distribucije, odnosno raspored pojedinačnih podataka - što znači da ne mogu da detektuju eventualnu višemodalnost distribucija, a nemamo ni uvid u ostale percentile mimo ovih glavnih pet tačaka. Njihova ekonomičnost - to što su sastavljeni iz prostih linija - doći će do većeg izražaja kada je potrebno da uporedimo više grupa podataka. Tada će nam, najčešće u kombinaciji sa rastresenim

tačkastim grafikonom, dobro poslužiti za detekciju glavnih tačaka i asimetričnosti distribucije.

Povremeno ćemo sretati **violinske grafikone** (engl. *violin plot*). To je zapravo vertikalno postavljeni polirani grafikon frekvencija zajedno sa svojim odrazom u ogledalu čime zaklapa zatvorenu krivu. Sam po sebi nije neka novost, ali ako se kombinuje sa kutijastim i/ili tačkastim grafikonom onda ima prednosti sve tri vrste (Slika 5.19); prikazuje pojedinačne skorove, ključne tačke distribucije, te tendenciju populacione distribucije.



Slika 5.19: Kombinacija violinskog, kutijastog i tačkastog grafikona.

Zaključni sa ovim grafikonom upoznali ste obuhvatn skup alata za analizu jedne dimenzione varijable. Naučili ste da računate i interpretirate mjere centralne tendencije, varijabilnosti i pozicije u nizu, te ste upoznali čitav spektar grafikona što će vam pomoći da dobro upoznate bilo koju dimenzionalnu varijablu u vašim budućim istraživanjima, da uočite atipične vrijednosti, procijenite oblik distribucije i komunicirate svoje nalaze drugim istraživačima. Prije nego što se upustimo u analizu odnosa između dvije ili više varijabli, u sljedećem poglavlju moramo da se osvrnemo na jedan poseban oblik raspodjele dimenzione varijable koji ima važnu ulogu u statistici: normalnu distribuciju.

Poglavlje 6

Normalna raspodjela: važnost, upotreba, mjere odstupanja

Nakon čitanja ovog poglavlja znaćete:

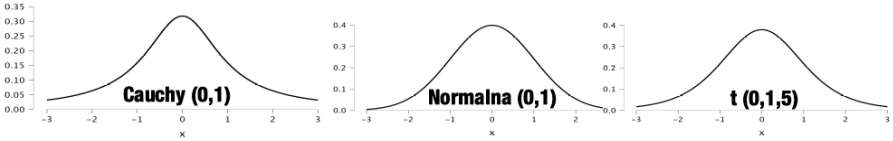
- objasniti značaj normalne distribucije za psihologiju i statistiku zaključivanja, uključujući centralnu graničnu teoremu i zakon velikih brojeva,
- koristiti z-skorove za izračunavanje vjerovatnoća i percentila u okviru normalne distribucije,
- procijeniti stepen odstupanja empirijske distribucije od normalne na osnovu mjera asimetričnosti i spljoštenosti.

Prethodna poglavlja su bila posvećena izračunavanju različitih mjera i konstruisanju grafikona koji nam pomažu da opišemo raspodjelu varijable na uzorku, kako god ta raspodjela izgledala. Bavili smo se "opipljivim", empirijski prikupljenim podacima. Ovo poglavlje je većinski posvećeno nečem drugačijem. Osnovna tema poglavlja je matematička raspodjela kontinuirane dimenzione varijable koju nikada nećemo empirijski opaziti u njenoj savršenosti, ali koja će vrlo često biti prisutna u istraživačkom radu kao ideal kojem treba težiti. Ona je jedna od mnogih distribucija vjerovatnoće i ima ogroman značaj za statistiku zaključivanja. Uz to, neke deskriptivne procedure gledaju na nju kao na uzornu, referentnu raspodjelu, te je potrebno da je natenane opišemo na ovom mjestu. Radi se o **normalnoj raspodjeli** (engl. *normal distribution*).

Zašto se normalna distribucija zove normalna?

Opis ćemo početi od njenog imena, koje je em mnogostruko, em zbunjujuće. Šta je tu uopšte *normalno*? Oni koji su dosad pažljivo

čitali primijetiće da sam za istu ovu distribuciju najprije koristio njen sinonimni naziv zvonasta raspodjela (engl. *bell curve*, od zvonasta kriva), jer sam pretpostavljao da će vam to biti asocijacija koju možete sebi življe da predočite. Zaista, pogled na Sliku 6.1 potvrđuje zašto se ona tako naziva. Međutim, iz iste slike možete vidjeti i da normalna distribucija nije jedina distribucija zvonastog oblika i da neiskusno oko teško razlikuje normalnu od druge dvije koje su također prikazane.



Slika 6.1: Normalna distribucija u društvu drugih zvonastih distribucija (copycats?).

Termin iz naslova, normalna raspodjela ćemo najčešće sretati. Navodno je (vidjeti izvor https://condor.depaul.edu/ntiourir/NormalOrigin.htm) usvajanju tog pojma u statističkoj zajednici najviše kumovao statističar Karl Pearson, tip koga ćemo još pominjati u ovoj knjizi. On je - kao i mnogi drugi u njegovo doba i prije njega - uvidio da ta distribucija opisuje ono što se opaža u variranju mnogih fenomena u stvarnosti, pa je insistirao da se to zove *normalnom* raspodjelom kako su već neki naučnici prije njega nazivali istu. Konkretno, normalna distribucija dobro opisuje kako tehnička odstupanja pri naučnom mjerenju (npr. greške u mjerenju u astronomiji), tako i neke biometrijske karakteristike (npr. visina unutar muškog ili ženskog pola[1]). Uz to, neke druge matematičke distribucije (npr. binomna, t-distribucija) "žele" da postanu normalna distribucija kad porastu, odnosno sa povećanjem broja događaja njihov izgled postaje sve sličniji normalnoj distribuciji. Ipak, potrebno je naglasiti da je usljed ovog epiteta *normalna* došlo do idolopoklonstva naspram drugih distribucija, koje također mogu da opišu prirodne procese generisanja varijabli, čega je i Pearson bio svjestan.[2]

[1] Zajednička visina ima oblik bimodalne distribucije!

[2] Goertzel & Fashing, 1981, https://doi.org/10.1177/016059768100500103

Treće ime koje se u literaturi vrlo često javlja je Gaussova raspodjela. Naime, J. C. F. Gauss, čuveni njemački matematičar, svojevremeno je izveo neke ključne matematičke dokaze za univerzalnost ove raspodjele. Međutim, istorija statistike - kakva uzbudljiva disciplina - sugeriše da ni on nije bio njen "pronalazač". Za autore bi se mogla proglasiti neka, vjerovatno manje poznata, imena kao što su de Moivre, Laplace, Legendre i Adrain.[3] Da ne komplikujemo dalje, držaćemo se odsad imena *normalna*, imajući na umu da će biti sasvim normalno i ako distribucija neke naše varijable odstupa od normalne raspodjele.

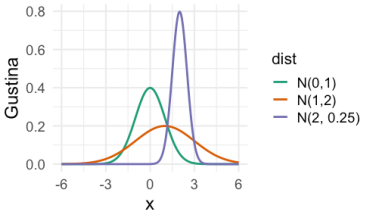
[3] Istorijski pregled daju Patel & Read, 1982, ISBN:9780824715410

Šta statistički karakteriše normalnu raspodjelu?

Nekoliko stvari. Kao prvo, očito je da se radi o unimodalnoj distribuciji dimenzione varijable. Taj mod je istovremeno i medijana i aritmetička sredina, s obzirom na to da se radi o savršeno simetričnoj raspodjeli. Automatski, ovo znači da se s jedne i druge strane od centralne vrijednosti nalazi po 50% gustine raspodjele. Nadalje, radi se o asimptotičnoj distribuciji. U prevodu na jezik razumljiviji brucošima psihologije: kriva linija nikad ne presijeca / dotiče X-osu, nego se prostire u beskonačnost, od $-\infty$ do $+\infty$ (čarobno, zar ne?). Ovo istovremeno objavljuje da imamo posla sa teorijsko-matematičkom distribucijom koja se ne sreće u stvarnosti u kojoj prikupljamo podatke. Iz tog slijedi i da njome predstavljamo zamišljene populacije, te ćemo koristiti grčka slova za opis njenih konkretnih mjera. A kad njih spominjemo, gustina normalne raspodjele duž x-ose je potpuno određena pomoću samo dvije mjere, μ i σ, kao što se vidi iz formule:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

U formuli vidimo poznate sastojke, konstante π i e, približno jednake 3.142 i 2.718. Dakle, ono što tvori različite normalne distribucije su vrijednosti aritmetičke sredine. Kako možete vidjeti sa Slike 6.2, one se naizgled razlikuju, ali samo zato što imaju različite pozicije na mjernoj skali, odnosno drugačija im je centralna tačka (aritmetička sredina) i širina (standardna devijacija). Eh da, možete odahnuti, jer formulu normalne distribucije nije potrebno da pamtite, nema praktičnu vrijednost.



Slika 6.2: Normalne distribucije sa različitim parametrima.

Postoje još neke zanimljive karakteristike normalne raspodjele, te ćemo se ovoj temi vratiti, ali sada malo više o njenom teorijskom značaju za psihologiju i posljedičnom značaju za upotrebu statistike u psihologiji.

Distribuiraju li se psihološki konstrukti normalno?

Odgovor na pitanje ćemo početi razbijanjem jedne iluzije. Naime, mnoge generacije psihologa su tokom prve godine svog studija gotovo nasilno uvjeravane da je normalna distribucija važna za psihologiju zato što bi izraženost najvećeg broja psiholoških svojstava morala biti normalno raspodijeljena. Zašto? Za veliku većinu psihologa, pa i laika koji se susreću sa psihologijom tokom svog obrazovanja, prva asocijacija na distribuciju psiholoških konstrukata jeste normalna raspodjela inteligencije, bolje rečeno

IQ skorova koji operacionalizuju inteligenciju. Uočićete da se i ja u ovoj knjizi prilično rigidno držim primjera koji uključuju inteligenciju, jer je ona dosta empirijski ispitivana i za svoje postojanje može zahvaliti statistici.

I zaista, prikupljeno je dosta dokaza da je normalna raspodjela jeste razuman model za predstavljanje raspodjele IQ skorova. Međutim, objašnjenje za to djelimično leži u činjenici da se testovi inteligencije kroje tako da odgovaraju modelu normalne raspodjele. Konkretno, autori testova u prvim fazama razvoja instrumenta provjeravaju koliko su pojedinačni zadaci teški za rješavanje, odnosno koji procenat ispitanika na uzorku može da riješi svaki od zadataka. Za konačnu verziju ostavljaju skup stavki koji u adekvatnoj mjeri zastupa vrlo lake i vrlo teške zadatke (najmanji broj njih), lake i teške zadatke (nešto veći broj), te srednje teške zadatke (najveći broj zadataka). Štaviše, trude se svim silama da distribucija ukupnog rezultata na takvom testu bude simetrična, a u idealnom slučaju, normalno raspodijeljena. Ali, kada bi neki test inteligencije sadržavao samo izuzetno lagana pitanja ili samo izuzetno teška pitanja dobili bismo potpuno drugačije, asimetrične distribucije. Pri nauk je da raspodjela operacionalizovane varijable ovisi o izboru stavki - nevezano za to kako se ciljno teorijsko svojstvo koje mjerimo distribuira u populaciji.

Zapravo, značajan broj naučnika danas tvrdi da su asimetrične distribucije u većini slučajeva adekvatniji opisi društvenih i psiholoških fenomena od interesa nego što je to normalna raspodjela.[4] Ilustracije radi, proširićemo diskusiju na još jedan primjer, jer inteligencija nije jedini zanimljiv psihološki konstrukt. Recimo da mene zanima kako je na samom početku semestra distribuisana zainteresovanost za statistiku među studentima prve godine psihologije. Da li i ovdje možemo očekivati simetričnu, pa i normalnu raspodjelu? Ako bih to ispitivao na skali samoprocjene zainteresovanosti od 0 (nimalo) do 10 (izuzetno zainteresovan), realističnije bi bilo zamisliti da ćemo dobiti asimetričnu distribuciju gdje najveći broj studenata nije nimalo zainteresovan za statistiku (skor 0 ili blizak njemu), dok je izuzetno mali broj studenata - ako iko od njih - fasciniran statistikom (ocjena 10) već na početku studija. Vjerujem da dijelite moje mišljenje da je populaciona distribucija tog konstrukta nenormalna.

Prikladnijim matematičkim distribucijima se posvećujemo kasnije kada se budemo bavili statistikom zaključivanja, a sad je potrebno najaviti da takvih distribucija ima tušta i tma. Među korisnijima su i one koje se odnose na diskretne varijable (npr.

[4] Neki članci koji govore o tome su: O'Boyle & Aguinis, 2012, https://doi.org/10.1111/j.1744-6570.2011.01239.x; Blanca et al.,2013, https://doi.org/10.1027/1614-2241/a000057; Cain, Zhang & Yuan, 2016, https://doi.org/10.3758/s13428-016-0814-1

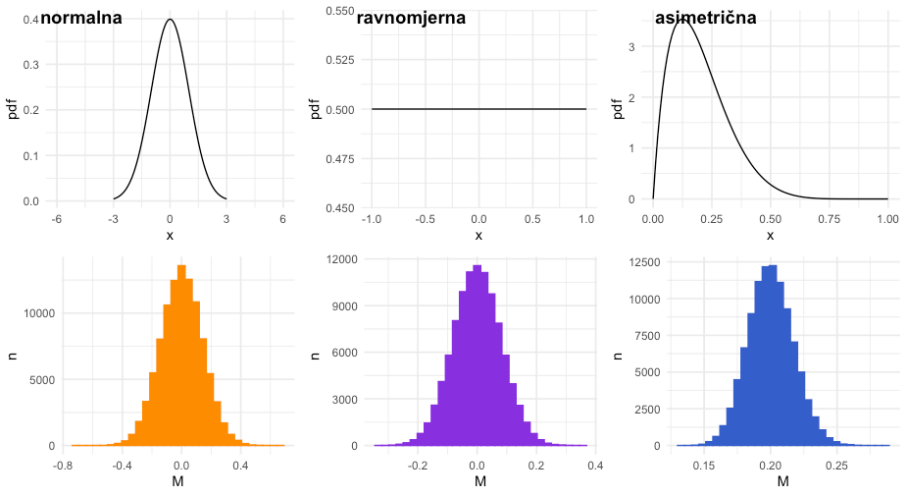
binomna, Poissonova, multinomna), ali i one koje se odnose na kontinuirane varijable (npr. beta, uniformna, Cauchyeva). Da zaključimo ovaj dio, poenta je da normalna distribucija može biti dobar opis populacione raspodjele operacionalizovane varijable - ili teorijske distribucije psihološkog atributa - ali niti to mora biti, niti se na tome mora insistirati.

Šta je to centralna granična teorema i čemu služi?

Dakle, oblik raspodjele psiholoških svojstava u populaciji nije glavni razlog zašto je normalna distribucija važna za psihologiju i zašto ja važna za vas koji sad učite statistiku. Pravi razlog ima mnogo više veze sa prirodom raspodjele statističkih mjera koje opisuju uzorak nego sa prirodom raspodjele pojedinačnih statističkih mjera / skorova. Naime, prirodu raspodjele mnogih statističkih mjera opisuje zakonitost koja je poznata pod imenom **Centralna granična teorema** (engl. *central limit theorem*). Njen autor bi mogao biti francuski matematičar P. S. Laplace (mogao jer se ponovo među originalnim autorima pominje zanemarivani de Moivre).

Centralna granična teorema spada u jedan od najbitnijih fenomena kada je u pitanju teorija vjerovatnoće i statistika zaključivanja. Konkretno, ona nam govori da bez obzira na to kakav je oblik izvorne populacione distribucije (tj. normalan ili neki drugi, simetričan ili asimetričan, bilo da je varijabla inteligencija ili stav prema statistici), što su uzorci veći, raspodjela očekivanih aritmetičkih sredina koji se dobijaju na tim uzorcima će sve više težiti normalnoj. Ova pravilnost postaje naročito vidljiva na uzorcima koji imaju 25-30 i više ispitanika. Vjerujem da sve ovo zvuči komplikovano, pa je moj savjet da sami uočite pravilnost koristeći sljedeće interaktivne applete: http://onlinestatbook.com/stat_sim/sampling_dist/ ili http://www.intuitor.com/statistics/CentralLim.html.

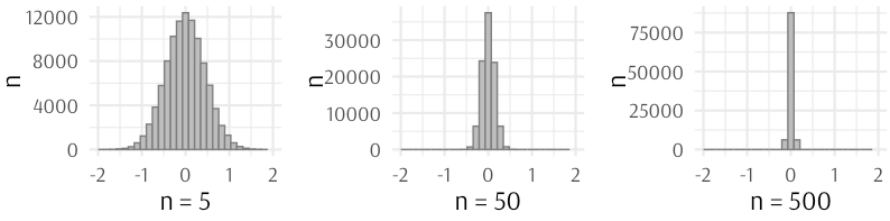
Za svaki slučaj, ja ovdje dajem (Slika 6.3) i fiksnu ilustraciju fenomena. U prvom redu se nalaze tri populacione distribucije iz kojih se uzimaju uzorci. Zatim je R-u data naredba da izvuče po 100 000 uzoraka od po 50 ispitanika i da za svaki od tih 100 000 uzoraka izračuna aritmetičku sredinu. U drugom redu je prikazana distribucija tako dobijenih aritmetičkih sredina. Postoje minimalna odstupanja za veoma asimetričnu populacionu distribuciju (konkretno, Beta(2,8) distribuciju), ali se i za nju dobijena distribucija aritmetičkih sredina može smatrati normalnom.



Slika 6.3: Ilustracija centralne granične teoreme.

Teorema istovremeno ima i dopunu kojom se utvrđuje da što je veći uzorak na kojem se dobija tražena mjera, mogućnost greške u procjeni parametra postaje sve niža (***zakon velikih brojeva***). Pogledajmo Sliku 6.4 na kojoj su prikazane raspodjele po 100 000 aritmetičkih sredina kada je uzorku bilo 5, 50 i 500 ispitanika. Naravno, krajnja situacija se dešava kada je $n = N$, jer tada uzorak potpuno tačno predstavlja populaciju, pa tada važi i $\hat{\theta} = \theta$. [5]

[5] Sjećate li se šta ova kapica znači?



Slika 6.4: Ilustracija zakona velikih brojeva.

Shvatanje centralne granične teoreme i zakona velikih brojeva je izuzetno značajno za shvatanje osnovnih principa statistike zaključivanja koji se baziraju na normalnoj distribuciji i nekim njoj vrlo srodnim distribucijama, pa ako niste sigurni da ste savladali njihovo značanje vraćajte se onlajn appletima do besvijesti.

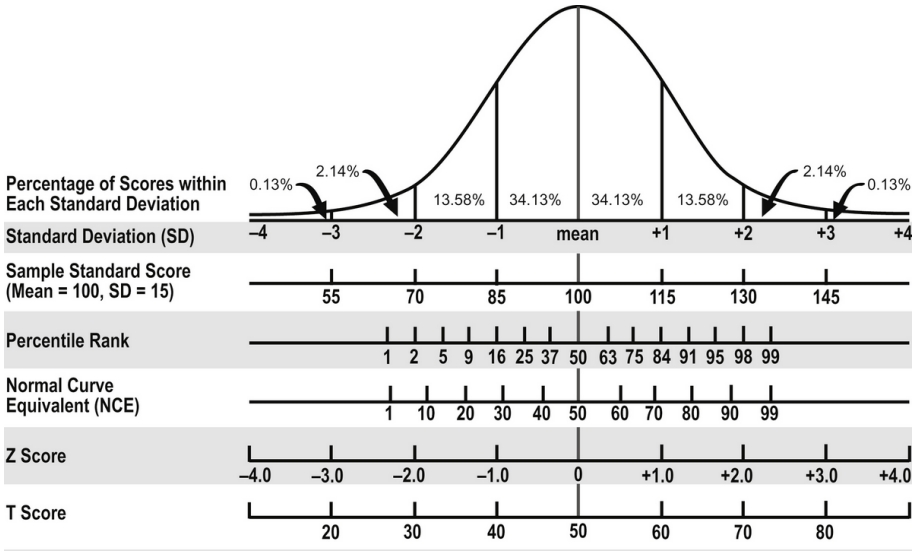
Ovo je pravo mjesto i da upozorim na dvije zabludu koju sam primijetio u pogledu centralne granične teoreme i zakona velikih brojeva. Prva je da ljudi često navode da što je veći uzorak postoje sve veće šanse da će distribucija varijable biti normalna. Ovo je tačno samo ukoliko je i izvorna populacija normalna, ali nije u drugim slučajevima. Veliki uzorak, na primjer onaj od 10000

ljudi i dalje neće davati normalnu raspodjelu odgovora ukoliko ljude pitate koliko vole statistiku. Distribucija će biti ispeglanija, ali neće postati simetrična. Druga zabluda je vezana za zakon velikih brojeva. Mnogi pomišljaju da ako imate veći uzorak da ćete garantovano imati bolji rezultat (statistik bliži parametru) nego ukoliko radite istraživanje na malom uzorku. To nije tačno, čak i onda kada bismo birali potpuno slučajne uzorke i imali optimalni kvalitet uzorkovanja. Naime, usljed greške uzorkovanja, očekujemo da će rezultati na uzorku uvijek odstupati od parametra - osim kada je $n = N$. Sa povećanjem uzorka se smanjuje opseg očekivane greške, ali se može desiti - doduše, za to postoje male vjerovatnoće - da na uzorku od 10 ispitanika dobijemo tačniju procjenu parametra nego na uzorku od 1000 ispitanika.

Kako određujemo površinu pod normalnom krivom pomoću z-skorova?

Sada kad smo razdvojili pitanje distribucije pojedinačnih skorova i distribucije mjera izračunatih na uzorku možemo da se vratimo još jednoj lijepoj osobini normalne distribucije ukoliko se skorovi u populaciji zaista raspodjeljuju normalno, a i u slučaju kada znamo da će distribucija traženih statistika biti normalna. Tada na značaju znatno dobijaju z-skorovi (kao i T-skorovi) koje smo upoznali u prethodnom poglavlju. Pogledajmo Sliku 6.5 na kojoj se nalaze oni zajedno sa nekim drugim mjerama koje smo već obrađivali.

Biće korisno da na početku odredimo neke pravilnosti. Ukupna gustina površine pod normalnom krivom (linija koja omeđuje raspodjelu) je jednaka 1.00 - odnosno jedno cijelo, odnosno 100.0% - uprkos činjenici o beskonačnosti prostiranja iste. Mislim da svi imate dovoljan IQ da zaključite da to znači da kako lijevo, tako i desno od centralne vrijednosti imamo po 50% raspodjele. Šta bi to značilo u pogledu raspodjele IQ skorova ako znamo da se normalna distribucija definiše tako da je prosječna populaciona vrijednost 100, a standardna devijacija 15 ($IQ \sim \mathcal{N}(100, 15^2)$? Stvari postaju posebno zanimljive kada shvatimo da je za svaki segment distribucije moguće odrediti količinu površine u odnosu na ukupno cijelu raspodjelu.



Slika 6.5: Površina pod normalnom distribucijom i različite mjerne skale. Preuzeto sa: http://www.bwgriffin.com/gsu/courses/edur8131/week-03.htm

Na primjer, Slika 6.5 nam govori da se između aritmetičke sredine i vrijednosti veće za jednu standardnu devijaciju od aritmetičke sredine, odnosno u intervalu između IQ skorova od 100 do 115, nalazi 34.1% distribucije. Drugačije rečeno, otprilike trećina populacije ima toliko procijenjenu inteligenciju. Ukoliko, interval proširimo simetrično i na lijevu stranu, onda ćemo shvatiti da se između IQ vrijednosti 85 i 115 nalazi nešto više od dvije trećine populacije, egzaktno 68.26%.

IQ skorovi su zanimljivi, ali su vezani samo za konstrukt inteligencije. Veća korist je od nečeg univerzalnog, a to su z-skorovi (kao i njihovi T-rođaci). Kao što znate njima možemo izražavati inteligenciju, ali i emocionalnu inteligenciju, savjesnost, fizičke karakteristike ili bilo šta što vam padne na pamet da bi u populaciji trebalo da bude normalno distribuisano. Svaka populaciona normalna distribucija $(X \sim \mathcal{N}(\mu, \sigma^2)$ se može linearno transformisati u tzv. standardnu normalnu distribuciju koja je izražena u z-skorovima $(z \sim \mathcal{N}(0, 1)$. To pretvaranje radimo jednostavno, uz pomoć formule za dobijanje z-skorova poznate od ranije koju smo samo blago modifikovali; naime, tada smo u formuli za aritmetičku sredinu i standardnu devijaciju imali statistike umjesto parametara:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Sve gore navedeno nam omogućava da odredimo koliko je vjerovatan određeni skor. Na primjer, nekad će nam zahtjevi

posla biti takvi da moramo postaviti granične skorove za selekciju kandidata. Možda će opis posla tražiti da kandidati spadaju među najboljih 20% kada je u pitanju inteligencija i da istovremeno budu među 20% emocionalno najstabilnijih (pod pretpostavkom da se emocionalnost distribuira normalno u populaciji). U tom slučaju moramo izračunati poziciju na z skali koja predstavlja P_{80} za inteligenciju, odnosno P_{20} za emocionalnost gdje veći skor znači manju emocionalnu stabilnost. Recimo da je za emocionalnu stabilnost aritmetička sredina bila 30, a standardna devijacija 5. Do željenih percentila možemo doći konsultujući posebno pripremljene tablice (možete ih lako naći internet pretragom za "z score table") ili koristeći onlajn softver (npr. https://www.psychometrica.de/normwertrechner_en.html ili https://www.calculator.net/z-score-calculator.html). Unoseći vrijednosti 0.20 i 0.80 (proporcije površine pod normalnom krivom) u kalkulator za ćeliju P(x<Z), odnosno za te percentile, lako ćemo doći do z-skorova -0.84 i +0.84. Sada kada imamo z skorove, kao i podatke o aritmetičkoj sredini i standardnoj devijaciji, onda je lako uvrstiti sve u formulu i izračunati traženi sirovi skor. Za IQ je to 112.6, a za emocionalnu stabilnost izračunajte sami. Obratite pažnju na to da taj skor mora biti manji od prosječnog skora.

Postojaće i slučajevi kada se od nas traži da odredimo granične skorove za površinu unutar distrubucije. Na primjer, u statistici zaključivanja su posebno značajna (reći ćemo unaprijed, statistički značajna) dva interpercentilna raspona. Prvi obuhvata srednjih 95% distribucije, a drugi 99% središnje distribucije. Dovoljno je da ponovo pogledamo na Sliku 6.5 i vidjećemo da se približno 95% distribucije nalazi između z-skorova -2 i +2. Zapravo je tačan opseg od 95%, koji će biti često pominjan u statistici zaključivanja, je omeđen z-skorovima -1.96 i +1.96. Slična situacija je sa opsegom od 99%; često uzimamo da je to otprilike raspon od -2.50 do +2.50, a tačne vrijednosti su -2.58 do +2.58. Dovoljno je da zasad zapamtite približne vrijednosti.

Kako se procjenjuju odstupanja od normalne distribucije?

Sve u svemu, bez obzira na to što se u psihologiji konstrukti ne raspodjeljuju uvijek normalno, nije loše imati normalnu distribuciju. Pogotovo je to korisno kada konstruišemo testove i kada za njih izrađujemo norme na velikim uzorcima. Zato je osmišljeno nekoliko načina da procijenimo koliko smo udaljeni od ideala kojem težimo. Upoznaću vas sa deskriptivnim mjerama odstupanja

po simetričnosti i spljoštenosti. Spomenuću samo da postoje i formalni statistički testovi koji uspoređuju empirijsku sa teorijskom normalnom distribucijom, te da postoje i posebni grafikoni za tu svrhu. Nećemo ih ovdje obrađivati budući da je za njihovo razumijevanje potrebno i da znate ponešto iz statistike zaključivanja, a pritom prikazane deskriptivne mjere sasvim "dobro rade posao", pogotovo ukoliko ih koristite zajedno sa već naučenim univarijatnim grafikonima za prikazivanje distribucije.

Zapravo, ovdje moram posebno naglasiti da na te mjere trebamo gledati samo kao na dopunski, brzi test koji nam sugeriše različite oblike distribucije, a da pravi sud o njima treba da odnosimo na osnovu odgovarajućih grafikona. Situacije u kojima su dole navedene mjere vrlo korisne su onda kada imamo veliki broj varijabli u našem istraživanju ili onda kada čitamo tuđe izvještaje u kojima uopšte nisu dati grafikoni (pa se bar možemo nadati da su autori saopštili te mjere kao neku zamjenu).

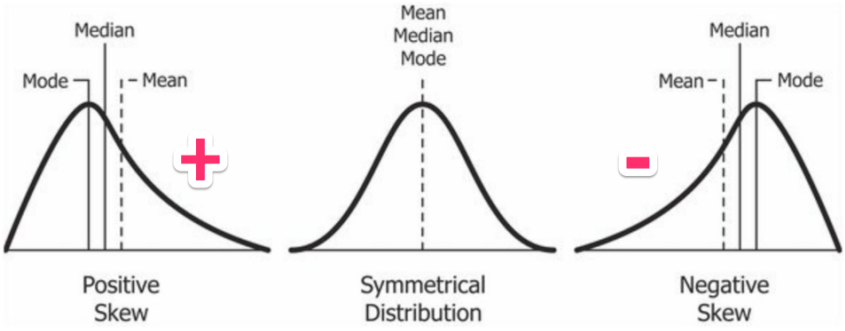
Odstupanja po simetričnosti

Za distribuciju podataka možemo da kažemo da je savršeno simetrična ukoliko se podaci raspodjeljuju jednako sa obe strane od mjere centralne tendencije. U takvom slučaju su i sve mjere centralne tendencije jednake. Takve distribucije nećemo sretati tako često u realnosti, odnosno viđaćemo manja ili veća odstupanja od savršene simetrije. Postoji više statističkih mjera koje operacionalizuju stepen asimetričnosti, a ubjedljivo najpopularniji - što se vidi i po zastupljenosti u statističkim softverima - je **Pearsonov G_1 skjunis koeficijent** (Sk; engl. *Pearson's skewness coefficient*). Njegov autor je - ponovo - onaj Karl Pearson i ova verzija je toliko popularna da se često u softverima identifikuje kao jedna i jedina mjera asimetričnosti i skraćeno se naziva samo skjunis (engl. skew se može prevesti kao nagnut, poprečan, iskošen). Formula koja se najčešće koristi za procjenu asimetričnosti na uzorku je sljedeća:

$$Sk = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s} \right)^3$$

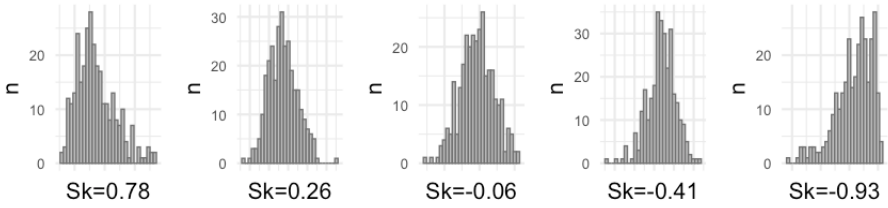
Lijeva strana formule $\frac{n}{(n-1)(n-2)}$ služi kao statistička korekcija kako bi se dobila tačnija procjena populacione vrijednosti, slično Besselovoj korekciji za standardnu devijaciju. Množilac sa desne strane je suma z-skorova na kub. Budući da se radi o stepenovanju na kub vrijednosti mogu biti i negativne (npr.

$-3 * -3 * -3 = -27$) i pozitivne $3 * 3 * 3 = 27$), pa tako i suma može biti negativna, nulta ili pozitivna. Iako postoje izuzeci, po pravilu će simetrična distribucija imati vrijednost $Sk = 0$. Na Slici 6.6 su prikazane tipične teorijske distribucije u pogledu (a)simetričnosti, uz očekivane pozicije aritmetičke sredine, medijane i moda.



Slika 6.6: Simetrične i asimetrične teorijske distribucije zajedno sa tipičnom raspodjelom osnovnih mjera centralne tendencije (preuzeto sa wikipedia.org).

S druge strane, na Slici 6.7 su prikazane tipične empirijske raspodjele za negativne, nulte i pozitivne vrijednosti skjunisa.



Slika 6.7: Simulirane empirijske distribucije koje oslikavaju različite vrijednosti skjunisa (n = 300).

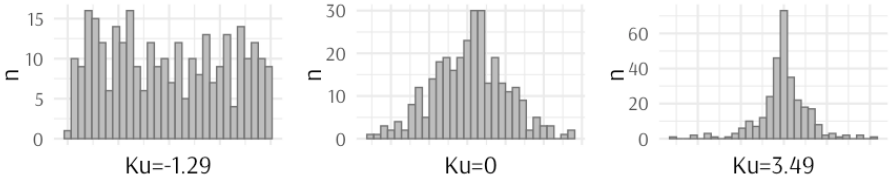
Slika 6.7 nam govori da je asimetričnost definitivno vidljiva kada je apsolutna vrijednost skjunis koeficijenta 0.50 ili veća. Za vrijednosti manje od toga potrebno je svakako da bacimo pogled na grafikon univarijatne distribucije.

Odstupanja u pogledu koncentracije skorova u sredini i na krajevima distribucije

Neka distribucija podataka može biti savršeno simetrična, a da ne bude normalna. Naime, distribucije se mogu razlikovati od normalne i u pogledu mase u njenoj sredini i na njenim krajevima, odnosno u pogledu spljoštenosti ili vršnosti distribucija.

Tipičan primjer "spljoštene" simetrične distribucije je uniformna ili ravnomjerna distribucija gdje svi skorovi duž X-ose imaju jednaku vjerovatnoću pojavljivanja. Na suprotnoj strani su distribucije u kojima značajno veći broj ispitanika ima skorove blizu mjere centralne tendencije u odnosu na normalnu distribuciju. Najčešća mjera koja se dovodi u vezu sa šiljatošću / spljoštenošću distribucije se zove **kurtozis** (Ku; engl. *kurtosis*).

Po pravilu, a izuzetaka ima, distribucije se mogu podijeliti na mezokurtične (one koje odgovaraju normalnoj, $Ku \approx 0$), platikurtične (spljoštenije u odnosu na normalnu, $Ku < 0$) i leptokurtične (šiljatije u odnosu na normalnu, $Ku > 0$). Tipične primjere možete vidjeti na Slici 6.8.



Slika 6.8: Simulirane empirijske distribucije koje oslikavaju različite vrijednosti kurtozisa (n = 300).

Za razliku od asimetričnosti koja statistički po sebi nije poželjna (iako može da odražava stvarnost i kao takvu je moramo prihvatiti), odstupanja po vršnosti / spljoštenosti predstavljaju prednost u nekim slučajevima. Na primjer, ako imamo neku mjeru sa malim brojem podioka, odnosno malo teorijskog prostora da rasporedimo ispitanike, spljoštene distribucije su korisnije od normalnih jer one imaju veću varijabilnost (a što nam onda olakšava dovođenje u vezu sa drugim varijablama). Ipak, ovo ovisi od istraživanja do istraživanja. I još jedna napomena, u velikom broju slučajeva odstupanja po simetričnosti i koncentraciji skorova dolaze zajedno.

Odstupanja u pogledu modalnosti distribucije

Sve gore navedeno za skjunis i kurtozis ima smisla dok je distribucija načelno unimodalna. U nekim slučajevima možemo dobiti njihove vrijednosti koje odgovaraju normalnoj, a da se radi o bimodalnoj distribuciji. Naravno, to jedino možemo provjeriti grafički, koristeći oštro oko i zdrav razum.

Ovim poglavljem smo završili sa opisivanjem distribucije jedne varijable, što je zaista neophodna osnova za razumijevanje ključnih statističkih principa. Međutim, psihologija kao nauka

postaje posebno zanimljiva kada pokušavamo da razumijemo kako se različiti psihološki fenomeni međusobno povezuju. Da li postoji veza između inteligencije i kreativnosti? Kako se anksioznost odnosi prema akademskom postignuću? Kako i koliko je samopoštovanje povezano sa društvenim stavovima? Odgovori na ovakva pitanja zahtijevaju da napravimo korak dalje, pa ćemo u narednom poglavlju istražiti kako se statistički opisuju povezanosti između dvije dimenzione varijable, što predstavlja srž empirijske psihologije i temelj za razumijevanje složenih psiholoških procesa koji nas okružuju.

Poglavlje 7

Analiza povezanosti dvije dimenzione varijable

Nakon čitanja ovog poglavlja znaćete:

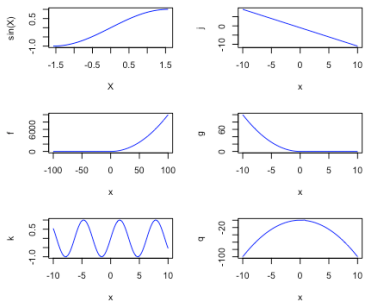
- definisati pojam statističke povezanosti i razlikovati njihove različite oblike,
- različite oblike grafičkog i numeričkog predstavljanja smjera i veličine povezanosti dvije dimenzione varijable,
- evidentirati probleme pri tumačenju povezanosti dobijenih na uzorcima.

Šta uopšte znači da su dvije varijable statistički povezane?

Sve do ovog poglavlja u fokusu smo imali pojedinačne varijable. Htjeli smo da utvrdimo njihove karakteristike tako što smo definisali centralne tačke grupisanja, opisivali količinu varijabilnost, ustanovljavali postoje li atipične mjere, te određivali kako ta distribucija izgleda - između ostalog, koliko je ona slična normalnoj. Dakle, može se reći da smo se u dosadašnjem dijelu teksta usredsredili na najfundamentalniji cilj nauke, a to je što bolje opisati pojedinačnu pojavu od interesa. Međutim, koliko god da je opis neke pojedinačne varijable koristan i zanimljiv, nauka uglavnom dobija na draži tek kada se dvije pojave dovedu u vezu. Na primjer, zanimljivo je istraživati samopoštovanje po sebi, ali je vjerovatno još zanimljivije istraživati odnos samopoštovanja i akademskog postignuća ili samopoštovanja i roditeljskih stilova vaspitanja. Analogno tome, za većinu ljudi je istraživati odnos količine stresa i fizičkog zdravlja nešto što je zanimljivije nego li istraživati stres po sebi, istraživati odnos fizičkog vježbanja i

kognitivnih sposobnosti zanimljivije nego istraživati puku distribuciju kognitivnih sposobnosti, a istraživati vezu između crta ličnosti i karijernog zadovoljstva zanimljive nego zasebno ispitavati distribuciju crta ličnosti.

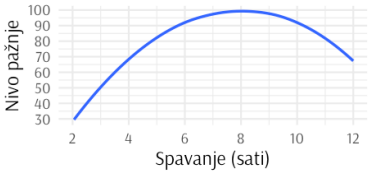
Ovim poglavljem ćemo zagaziti u teritoriju empirijskog ispitivanja povezanosti među varijablama. Da bismo znali kuda idemo neophodno je najprije odrediti šta podrazumijevamo pod povezanošću između dvije dimenzione varijable. U ogromnoj većini slučajeva ćemo pod tim smatrati statističko kovariranje, tj. pravilnost da promjene vrijednosti na jednoj dimenziji prati tendencija promjena vrijednosti na drugoj dimenzionoj varijabli. Već na početku je bitno naglasiti da pod tim mislimo na promjene u očekivanim vrijednostima, a ne nužno na ono što se u stvarnosti dešava u pojedinačnim slučajevima. Na primjer, (pozitivno) statističko kovariranje znači da ako za neku osobu registrujemo natprosječan skor na X varijabli onda bismo trebali očekivati da ona ima i natprosječan skor na Y varijabli. Dakle, to ne podrazumijeva da ako za Marka registrujemo natprosječni skor na X varijabli da Marko nužno mora imati i natprosječan skor na Y varijabli. Iz datoga slijedi da kada kažemo da su inteligencija i školsko postignuće povezani da mi tvrdimo da većina vrlo inteligentnih ljudi teži da postiže visoke ocjene, iako smo svjesni da postoje i "podbacivači", naši vrlo inteligentni drugari koji su loše prolazili u školi iz ovog ili onog razloga.



Slika 7.1: Tipične monotone i nemonotone funkcije.

Ova vrsta povezanosti - ako je X više onda će vjerovatno i Y biti više - se matematičkim jezikom zove stroga ili striktna monotonost, ali to nije jedini oblik statističke povezanosti. Za početak, riječ **monotona** (engl. *monotone* ili *monotonic*) povezanost se koristi u dva slučaja. U slučaju pomenute stroge monotonosti rast jedne varijable obavezno prati ili rast ili pad očekivanog skora na drugoj varijabli. Najjednostavniji oblik stroge monotonosti je pravolinijska funkcija koja se naziva i linearna funkcija - o kojoj će kasnije biti dosta riječi. U slučaju slabe monotonosti imamo uglavnom rastuću ili opadajuću funkciju, ali na nekom dijelu funkcije dolazi do platoa, odnosno promjena vrijednosti X varijabli nije povezana sa promjenom vrijednosti Y varijable. Drugim riječima, postoji interval na kojem je ta funkcija *nerastuća* ili *neopadajuća*. Suprotno tome, **nemonotone funkcije** (engl. *nonmonotonic*) su one kod kojih se smjenjuju trendovi kovariranja. Konkretno, za jedan interval X-vrijednosti očekuje se rast Y-vrijednost, a onda nakon neke tačke se sa rastom X-vrijednosti očekuje pad Y-vrijednosti. Sve gore navedeno je najlakše shvatiti vizuelno, pa su na Slici 7.1 dati ekstremni primjeri navedenih funkcija: strogo monotone su na

vrhu, slabo monotone u sredini, dok su u donjem redu prikazane dvije nemonotone funkcije.



Slika 7.2: Nemonotona funkcija koja opisuje sate spavanja i stepen pažnje.

Monotone funkcije su dosta lakše za razumijevanje, pa je potrebno dati primjer jedne realne nemonotone veze. To bi mogla biti povezanost između količine sna (varijabla X) i kognitivnog funkcionisanja (varijabla Y) (Slika 7.2). Iz sopstvenog iskustva znamo da, u prosjeku gledano, što smo više naspavani to smo bliži optimalnom kognitivnom funkcionisanju, u smislu da smo sposobniji više toga zapamtiti, usmjeriti svoju pažnju na zadatak i svrsishodno donositi odluke. Međutim, većina nas je doživjela stanje kada smo "preispavani", odnosno kada se po buđenju osjećamo letargično i potrebna nam je značajna količina kofeina ili tuširanje ledenom vodom da bismo vratili kognitivne kapacitete u normalno stanje. Drugim riječima, vjerovatno postoji tačka nakon koje dodatni minuti sna smanjuju kognitivnu funkcionalnost. I u takvim slučajevima ćemo govoriti o povezanosti X i Y varijable, samo što će opis odnosa biti nešto kompleksniji.

Ali, to nije sve. Dvije dimenzione varijable mogu biti povezane na jedan još čudniji, i zbog toga često zanemaren, način. Naime, rast jedne dimenzije uopšte ne mora da prati promjena u očekivanoj vrijednosti druge varijable, a mi i dalje možemo govorimo o njihovoj statističkoj povezanosti. To su slučajevi kada se sa rastom dimenzije X događa promjena količine varijabilnosti varijable Y. Mogući, ali zasad fiktivni primjer, bi bio odnos ekstraverzije i izraženosti negativnog afekta. Pretpostavimo da je tačno da su introvertne, ambivertne (na pola puta između introvertnih i ekstravertnih) i ekstravertne osobe u prosjeku podjednako sklone doživljavanju osjećanja neprijatnosti i napetosti. Ako već poznajemo pozadinske biofiziološke mehanizme ekstraverzije, moglo bi se zamisliti da ekstravertne osobe doživljaju širi dijapazon doživljaja negativnog afekta, tako da mogu biti i mnogo opuštenije i mnogo uznemirenije, a sve u zavisnosti od sredinskog uticaja kojem su podložnije od introvertnih osoba (ponavljam, primjer je fiktivan). Ako ste mislili da su monotonost i nemonotonost teški izrazi, čujte sad ovaj koji označava upravo opisani fenomen: heteroskedastičnost. **Heteroskedastičnost** (engl. *heteroskedasticity*, zapravo od grčkih riječi *hetero* što znači drugačiji i *skedannumi* što znači raspršiti) je relativno rjeđa pojava, ali kada se javi, može se javiti u kombinaciji sa monotonim ili nemonotonim povezanostima. Sve osnovne oblike povezanosti koje se javljaju u realnosti - uključujući tu i pitanja monotonosti i heteroskedastičnosti - najlakše je shvatiti putem grafikona, pa prelazimo zato na temu vizualizacije povezanosti

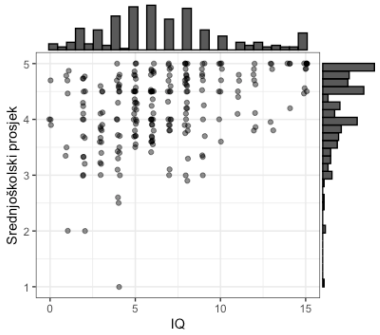
dvije dimenzione varijable.

Na koji način grafički predstavljamo povezanost dvije dimenzione varijable?

Siguran sam da se više ne iznenađujete kada najavim da postoje različite opcije kako vizuelno predstaviti odnos između dvije dimenzione varijable. Ići ćemo redom.

Grafikon raspršenja (engl. *scatterplot*) je osnovni i ubjedljivo najintuitivniji tip grafikona kojim se predstavlja povezanost između dvije dimenzione varijable. Njegova snaga leži u tome što unutar dvodimenzionalnog koordinatnog sistema prikazujemo svaki pojedinačni objekti pomoću tačkice ili nekog drugog simbola. Taj simbol predstavlja ukrštenu kombinaciju dvije mjere koje su registrovane za ispitanika (npr. za ispitanika A 133 za inteligenciju i 9.45 za ocjenu na fakultetu, te za ispitanika B 120 za inteligenciju i 7.87 za prosječnu ocjenu).

Slijedi nekoliko tehničkih detalja kreiranja grafikona raspršenja, a počinjemo sa pozicioniranjem varijabli na mjerne skale. Ukoliko teoretski možemo pretpostaviti iole plauzibilan uslovno-kauzalni odnos između dvije varijable, X-osu ćemo koristiti za prediktorsku varijablu, a Y-osu za kriterijsku. Ako je zaista teško domaštati moguću kauzalnu vezu između njih - ili kada imamo jake indicije da je uzročnik te povezanosti neka treća varijabla - onda je svejedno šta stavljamo na X, a šta na Y osu. Primjer za to bi moglo biti predstavljanje povezanosti rezultata kolokvija iz dva predmeta koji se tokom studija slušaju u istom semestru: Uvod u socijalnu psihologiju i Uvod u razvojnu psihologiju.

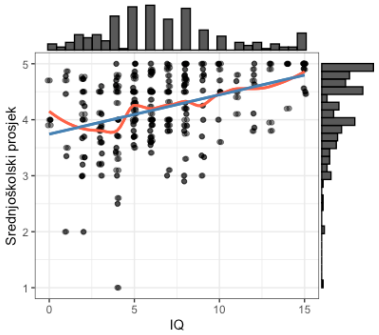


Slika 7.3: Grafikon raspršenja odnosa skora na kratkom testu neverbalne inteligencije i prosječne srednjoškolske ocjene (n=227).

Sa problemom preklapanja simbola koji označavaju mjere ispitanika smo se već sreli kada smo opisivali jednodimenzionalne tačkaste grafikone. Da bi skorovi baš svakog ispitanika bili vidljivi na grafikonu pribjegavaćemo neznatnoj randomizaciji pozicije simbola (engl. *jitter*), najčešće i po X i po Y osi. Opcionalno, naročito za prezentacione svrhe, zgodno je na rubovima osa dodati i grafikone distribucije pojedinačnih varijabli (najčešće histograme ili polirane grafikone proporcija). Primjer možete vidjeti na Slici 7.1a gdje su prikazani podaci iz stvarnog istraživanja na našim prostorima u kojem je, između ostaloga, ispitivana veza između opšte kognitivne sposobnosti (kratki test inteligencije sa ukupno 15 zadataka) i srednjoškolskog prosjeka.

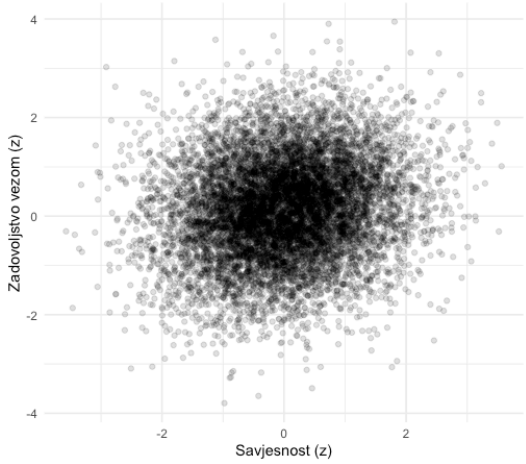
Za neistrenirano oko, Slika 7.3 može predstavljati neidentifikovanu nebulu na zvjezdanom nebu, odnosno teško mu je razaznati postoji li tu neka statistička povezanost. Ako je već

osnovni cilj uočiti obrazac povezanosti, onda bi trebalo uvijek iscrtati i jednu ili više linija koje nam pružaju uvid u opšti trend kovariranja. Takve linije za nas iscrtava softver, a na osnovu gustine pozicioniranja objekata unutar grafikona (detalji o proceduri slijede u ostatku teksta). U suštini, razlikovaćemo dvije mogućnosti: to mogu biti ili prave ili zakrivljene linije koje predstavljaju trend povezanosti i koje se vide na Slici 7.4. Iako zvuči banalno, izbor tih linija nas uvodi u jednu od najbitnijih tema statistike uopšte, temu kreiranja i provjeravanja statističkih modela. Ta tema je izuzetna bitna i predstavlja temelj statistike zaključivanje. U nastavku poglavlja ne možemo baš naširoko o tome, ali ćemo napraviti malu digresiju prije nego što se vratimo grafikonu raspršenja.



Slika 7.4: Grafikon raspršenja odnosa skora na kratkom testu neverbalne inteligencije i prosječne srednjoškolske ocjene sa prikazanim trendovima povezanosti (n=227).

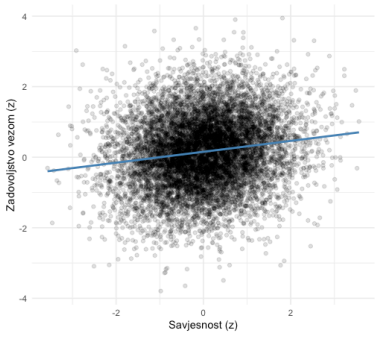
Kao što već znate modeli su pojednostavljeni prikazi stvarnosti. Potrebni su nam zato što je stvarnost prekompleksna. Da bismo stekli koristan uvid o pravilnostima života u nama i oko nas moramo da svedemo ogroman broj pristižućih informacija na manji broj relevantnih (sjećate li se još uvijek definicije statistike s početka knjige?). Grafikoni raspršenja zaista mogu izgledati kao zvjezdano nebo, potencijalno sa desetinama hiljada tačkica, čiji opšti trend nije lako uvidjeti. Pogledajmo primjer na Slici 7.5.



Slika 7.5: Grafikon raspršenja odnosa savjesnosti i zadovoljstva romantičnom vezom, prikazanim putem z-vrijednosti (n=10 000).

Recimo da taj grafikon (koji bi zaista trebalo da sadrži 10000 kružića! - nisam ih prebrojao) prikazuje podatke iz istraživanja u kojem je postavljena hipoteza da što je osoba savjesnija (tj. pouzdanija i odgovornija, obraća više pažnje na detalje u radu i okruženju, više planira i promišlja prije nego što donese odluku) teži da bude i zadovoljnija svojom romantičnom partnerskom vezom. Navedena hipoteza je verbalni modeli koji izuzetno

pojjednostavljuje stvarnost. Pritom se podrazumijeva da se radi o statističkoj povezanosti, odnosno još bolje bi bilo eksplicitno navesti: "U prosjeku gledano, savjesnije osobe su zadovoljnije romantičnim vezama." Naime, znamo da zadovoljstvo vezom ovisi i o mnogim drugim psihičkim karakteristikama partnera (npr. sličnost kognitivnih sposobnosti, nivoa fizičke privlačnosti, vrijednosnih orijentacija, te sličnost, ali i koplementarnost različitih crta ličnosti). To znači da ako je jedna osoba savjesna to ne može biti garant da će ona biti zadovoljna svojom vezom. Takođe bi nam svima već trebalo biti jasno da su verbalni modeli grubi i da nauka više voli matematički izražene modele. Oni su precizniji, a to istovremeno znači da se lakše mogu dobiti dokazi u prilog njihovog opovrgavanja ili potvrđivanja. Matematički izražen model koji bi preveo navedenu hipotezu u provjerljivi model bi bio onaj koji prosto kaže da zadovoljstvo vezom može da se objasni putem dvije komponente: savjesnosti i dijela kojeg ne možemo da objasnimo putem savjesnosti. Taj drugi dio ćemo zvati greškom modela. Najjednostavniji vid takvog modela bi tvrdio da sa sve većom savjesnošću proporcionalno raste zadovoljstvo vezom, pa bi konceptualna formula glasila: *zadovoljstvo = savjesnost + greška*.



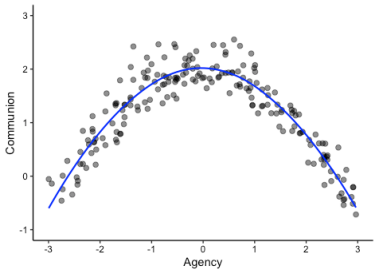
Slika 7.6: Grafikon raspršenja odnosa savjesnosti i zadovoljstva romantičnom vezom, prikazanim putem z-vrijednosti i sa prikazanom linearnom funkcijom (n=10 000).

Matematička magija leži u tome da se jednom običnom pravom linijom može predstaviti statistički model koji pretpostavlja jednak trend kovariranja duž cijele X varijable. Takvu funkciju nazivamo linearnom. Već smo je upoznali u nekim ranijim grafikonima i po obliku znamo da ona spada u stroge monotone funkcije. Dakle, linearna funkcija je najjednostavniji model odnosa između dvije varijable i - iznenađujuće - u zavidnom broju slučajeva predstavlja razumnu aproksimaciju odnosa između dva svojstva. Kako se tačno dobija pozicija i nagib te prave linije biće objašnjeno dosta kasnije kada se budemo bavili linearnim regresionim analizama. Zasad je dovoljno da znate da se ona automatski izvodi na podacima tako da se na najmanju moguću mjeru svodi greška modela koja je operacionalizovana odstupanjima svih tačkica od linije. Ako linija ima neki nagib, odnosno nije potpuno paralelna sa X-osom, tada je jasno da neka povezanost postoji. Na Slici 7.6 je prikazana dobijena linearna funkcija. Možete li sada da zaključite postoji li statistička povezanost između savjesnosti i zadovoljstva vezom?

A šta ako je povezanost među varijablama nelinearna?

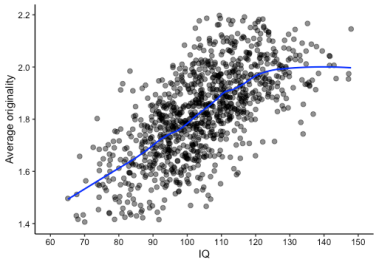
"Neprave", odnosno zakrivljene linije predstavljaju promjenjive trendove kovariranja X i Y varijable. Drugim riječima, funkci-

jama je dozvoljeno da zavijaju. Neke od njih su vam vjerovatno poznate iz gradiva srednjoškolske matematike - pripremite se za retraumatizaciju - kvadratna, logaritamska i eksponencijalna. Takve predefinisane funkcije imaju svoje mjesto u psihološkoj statistici, međutim za grafikone raspršenja se koristi jedna druga, vrlo fleksibilna vrsta funkcije koja ima relativno lako shvatljivu logiku iza sebe: lokalna regresiona funkcija. Za razliku od pravolinijske funkcije koja odražava ekstremno jednostavan linearan model, lokalna regresiona funkcija može izraziti kompleksnije, nelinearne i/ili nemonotone veze. Primjere takvih odnosa registrovanih u objavljenim istraživanjima možete vidjeti na Slikama 7.7 i 7.8. Šta vam govore funkcije prikazane iz dvije navedene studije?



Slika 7.7: Grafikon raspršenja percepcije ispitanika o vezi između aktivnog individualiteta (agency) i orijentacije ka zajedništvu (communion) koji pokazuje jasnu nemonotonu, obrnutu-U povezanost (rekreirano na osnovu rezultata Imhoff i Koch, 2017, <https://doi.org/10.1177/1745691616657334>).

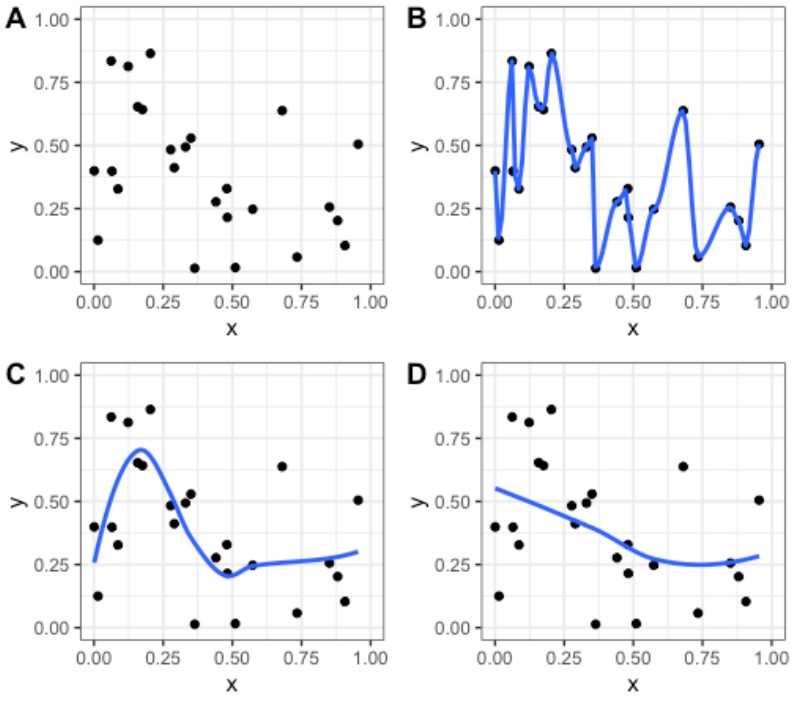
Mehanizam iscrtavanja lokalne regresije se može prikazati putem Slike 7.9. Grafikon A prikazuje grafikon raspršenja za dvije varijable koje su jedinično skalirane. Najdrastičniji i najkompleksniji vid lokalne regresije je onaj prikazan na grafikonu B gdje linija putuje tako što redom povezuje sve objekte na grafikonu. Ovakva funkcija savršeno objašnjava vezu između dvije varijable na uzorku. Međutim, postoje barem dva problema sa njom. Kao prvo, tehnički je izvodiva samo onda kada je broj ispitanika izuzetno mali i svaki od njih ima jedinstvenu X-vrijednost. Čim se desi da imamo dva ispitanika sa istom X-vrijednošću funkcija mora da uzme neku mjeru prosjeka između njih. Drugi i veći problem je što je uopštivost ovako nazubljene funkcije na druge uzorke i na populaciju gdje očekujemo drugačije podatke ravna nuli.



Slika 7.8: Grafikon raspršenja koji prikazuje nelinearnu, slabo monotonu povezanost inteligencije i originalnosti odgovora za neobičnu upotrebu poznatih predmeta (rekreirano na osnovu rezultata Jauk et al., 2013, <https://doi.org/10.1016/j.intell.2013.03.003>).

Iz tog razloga se u realnim situacijama pribjegava poliranju linija koje smo već sretali u slučaju poliranog grafikona proporcija. Proces kreiranja polirane lokalne regresione funkcije sjajno ilustruju animirane vizualizacije koje su dostupne na internetu[1], a sljedeći opis može da posluži kao nezamarajući uvod. Ukratko, algoritam ne uzima u obzir svaku pojedinačnu tačku pri iscrtavanju linije, nego samo one koje se nalaze unutar intervala na X-osi kojeg kao istraživači možemo odrediti. Za svaki interval se kreira linearna funkcija, te kako se interval pomiče na desnu stranu tako se linija povezanosti dopunjava. Koliko će linija biti zaglađena ili nazubljena ovisi i veličini intervala (**pojasna širina**, engl. *bandwidth*) koju određujemo kao istraživači - ako nećemo već da ga za nas odredi softver. Na Slici 7.9 C i D su primjeri povećanog poliranja. Kao istraživačima nam nije tako lako odrediti koja je funkcija bliža istini, jer ne znamo da li je nemonotonost prisutna na funkciji C plod slučajnog variranja na uzorku ili je to sistemski odraz stvarnosti.

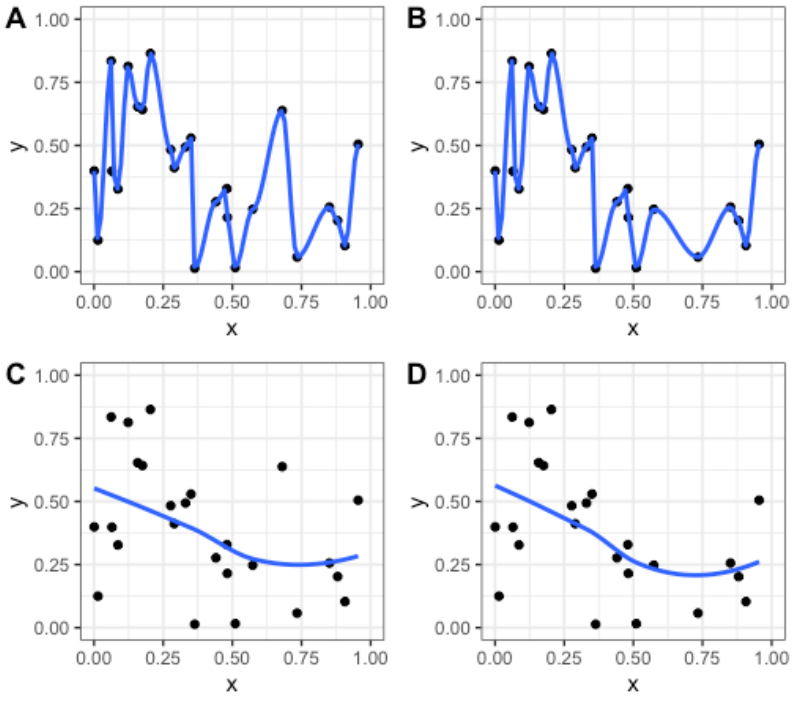
[1] npr. http://simplystatistics.org/posts/2014-02-13-loess-explained-in-a-gif/



Slika 7.9: Grafikon raspršenja koji prikazuje zavisnost oblika lokalne regresione funkcije od širine poliranja.

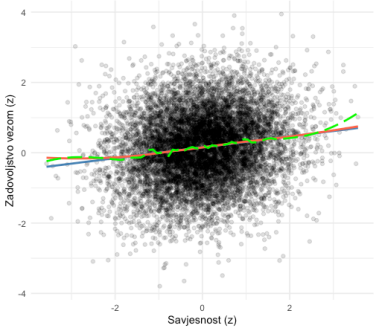
U većini slučajeva se možemo poslužiti načelom Okamovog brijača, odnosno načelom parsimoničnosti koje favorizuje manje kompleksne pretpostavke. Na Slici 7.10 možemo vidjeti da atipični pojedinačni podaci manje utiču na funkcije koje koriste šire intervale. Konkretno, gornji grafikoni (A i B) prikazuju nepolirane funkcije koje se međusobno značajno razlikuju iako je samo jedan ispitanik, čiji X skor je 0.70, izostavljen na grafikonu B. S druge strane, C i D funkcije koje prikuju polirane funkcije za A i B set podataka gotovo da se uopšte ne razlikuju. Naravno, potrebno je - obično iz više pokušaja - pronaći optimalnu vrijednost peglanja kako bi kriva adekvatno predstavljala stvarnu tendenciju kovariranja i bila otporna na slučajna variranja skorova.

U idealnom slučaju ćemo na istom grafikonu prikazati i pravu liniju i optimalnu krivu liniju kako bismo ocijenili da li je prava linija dovoljno dobar reprezent odnosa, jer je ona najparsimoničnija od svih. Preporučivo je da ih iscrtamo različitim bojama i/ili različitim načinom iscrtavanja (npr. puna linija i isprekidana linija). Ako se te dvije linije definitivno preklapaju dobili smo dokaz u prilog linearnog modela. Ako postoje značajnija razmimoilaženja potrebno je da promislimo da li su za to odgovorni samo atipični



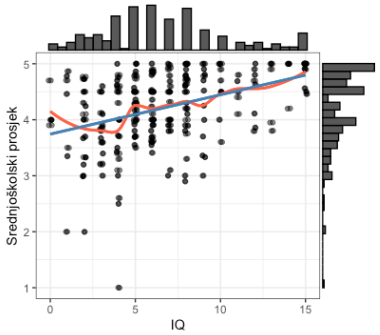
Slika 7.10: Grafikon raspršenja koji prikazuje osjetljivost lokalne regresione funkcije.

slučajevi ili se zaista radi o kompleksnijoj povezanosti. Analiza Slike 7.11 nas uvjerava da postoji načelno linearna povezanost između savjesnosti i zadovoljstva romantičnom vezom, budući da samo na krajevima funkcije postoje odstupanja. Takva odstupanja na krajevima često možemo očekivati budući da na krajevima distribucija obično imamo mali broj podataka, pa je funkcija podložna variranju.

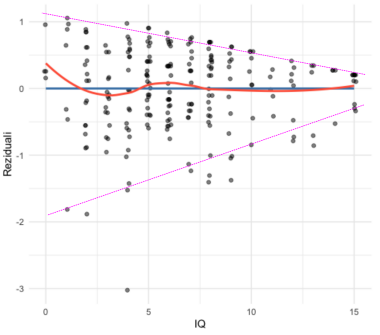


Slika 7.11: Grafikon raspršenja odnosa savjesnosti i zadovoljstva romantičnom vezom, prikazanim putem z-vrijednosti i sa prikazanom linearnom, te dvije lokalne regresione funkcije sa različitim stepenom poliranja (n=10 000).

Grafikon reziduala (engl. *residual plot*) je prilagođeni grafikon raspršenja koji se koristi da se ispita prisutnost linearnosti, ali i onog fenomena nakaradnog imena - heteroskedastičnosti. Inače, termin rezidual je jedan od omiljenih statističkih termina i jednostavno znači *ostatak*. Taj ostatak zapravo predstavlja razliku između modelom predviđene vrijednosti i stvarne vrijednosti koja je opažena na uzorku. Recimo da je osoba A imala IQ 140 i da je na osnovu linearnog modela za nju predviđena prosječna fakultetska ocjena 9.60. Ako je u stvarnosti ona zapravo bila savršen student i ostvarila prosjek 10.0, onda je rezidual za tu osobu jednak $res_A = 10.0 - 9.6 = 0.4$. S druge strane, ukoliko je osoba B imala IQ 140, a na kraju studija imala prosječnu ocjenu tek 8.60, onda je rezidual za nju jednak $res_B = 8.6 - 9.6 = -1.0$. Ovo znači da i reziduali mogu biti i pozitivni i negativni. Na grafikonu reziduala se na X-osi i dalje prikazuju vrijednosti za



Slika 7.12: Grafikon raspršenja odnosa skora na kratkom testu neverbalne inteligencije i prosječne srednjoškolske ocjene sa prikazanim trendovima povezanosti (n=227).

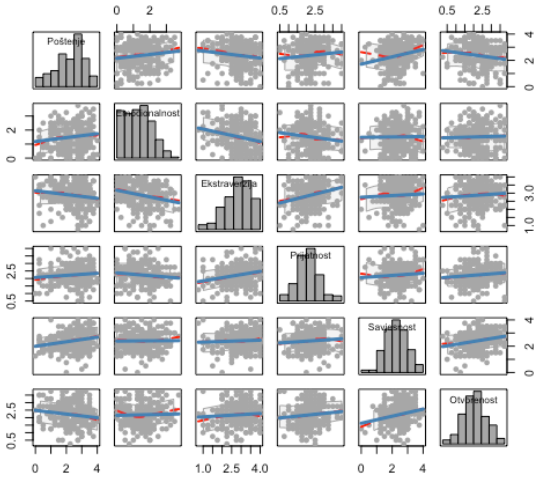


Slika 7.13: Grafikon reziduala za iste podatke (poliranje lokalne regresije je blago veće).

varijablu X (npr. IQ), ali se na Y-osi umjesto varijable Y (npr. prosjek ocjena) prikazuju reziduali koji mogu biti ili apsolutni ili standardizovani. To znači da će Y-osa uvijek sadržavati i vrijednost 0. U svakom slučaju, ukoliko je funkcija zaista linearna reziduali će se simetrično raspoređivati oko horizontalne linije koja prikazuje vrijednost 0, a u idealnom slučaju će oni biti normalno raspodijeljeni. Saglasnost te horizontalne linije i lokalne regresione linija koja se izvodi na ovim podacima će nam dodatno potvrditi ili opovrgnuti postojanje linearnosti.

Vratićemo se na primjer našeg istraživanja povezanosti inteligencije i prosječne srednjoškolske ocjene. Grafikon raspršenja smo već prikazali ranije, ali ga ponovo prikazujemo na Slici 7.12. Pogledaćemo sada šta nam to govori i on, ali i grafikon reziduala (Slika 7.13). Vidimo da se plava i crvena linija u velikoj mjeri preklapaju i da je razlog za krivudanje linije na početku X-ose vjerovatno vezan za atipične vrijednosti. Tu su sano jedan učenik koji je imao nedovoljan uspjeh prethodne godine pa je upisao prosječnu ocjenu 1, te četiri učenika koji su imali 0 bodova na testu pri čemu niko od njih nije imao prosječnu ocjenu značajno nižu od 4.0. Međutim, postoji nešto što nam još više privlači pažnju. Naime, isprekidane ružičaste linije nam pokazuje trend variranja i vidljivo je da se variranje sužava kako se krećemo duž X-ose, odnosno kako inteligencija raste. Drugim riječima, ukoliko bismo predviđali prosječnu ocjenu na osnovu ovog testa inteligencije tačnije bismo to mogli da uradimo za visoke inteligentne nego za one koji su postigli niske skorove. Konkretno, svi oni koji su na testu ostvarili svih 15 bodova su imali prosječne ocjene u intervalu od 4.5 do 5.0, dok su učenici koji su ostvarili 4 boda na testu imali prosjek koji je varirao od 1.0 do 5.0. Dakle, ovo je primjer linearne, heteroskedastične povezanosti.

Matrica grafikona raspršenja (engl. *scatterplot matrix*, *draftman's plot*) predstavlja matricu više grafikona raspršenja (npr. između svih dimenzionih varijabli ili koristeći izbor X i Y varijabli). Ovaj prikaz je koristan jer pruža efikasan uvid u međupovezanost velikog broja varijabli. Dijagonala matrice je takođe iskoristiva, pa u njoj možemo predstaviti grafikone univarijatne distribucije. Slika 7.14 prikazuje matricu grafikona raspršenja gdje su za odnose između svakog para varijabli prikazane i linearne funkcije (plave linije) i lokalne regresione funkcije (crvene linije). Naravno, efikasnost ovakvog prikazivanja se smanjuje sa povećanjem broja varijabli jer matrica i njeni elementi postaju nepregledni ukoliko je broj varijabli pretjeran, a to već počinju da budu sa više od 5-6 varijabli.



Slika 7.14: Matrica grafikona raspršenja koja prikazuje povezanost između šest crta ličnosti HEXACO modela (n=249).

Na koji način numerički prikazujemo povezanost dvije varijable?

Efikasnost prikazivanja je upravo jedan od ključnih razloga zašto su osmišljene mjere koje putem jednog broja sažimaju odnos između dvije varijable. Zamislimo da u bazi imamo 15 varijabli, te da želimo da procijenimo povezanosti između svake od njih. Koristeći formulu $\frac{n*(n-1)}{2}$, gdje n predstavlja broj varijabli, dolazimo do broja od 105 jedinstvenih parova. Za matricu grafikona raspršenja bi to bio problem predstaviti, dok su takve tabele sa numeričkim pokazateljima uobičajena pojava u naučnim radovima. Drugi ključni razlog za postojanje numeričkih mjera povezanosti jeste što su one građa za multivarijatne statističke postupke. Drugim riječima, ti brojevi mogu da se uvrste u formule koje omogućavaju uvide u simultane međuodnose 3, 30, 103 ili čak i više varijabli.

Šta su to koeficijenti korelacija?

Mada postoje različiti načini da se numerički iskaže povezanost između dvije varijable, daleko najprisutniji su **koeficijenti korelacija**. Koeficijenti korelacija predstavljaju porodicu standardizovanih mjera intenziteta povezanosti između varijabli. U ovom slučaju, standarizovani znači da oni izražavaju snagu povezanosti na istoj mjernoj skali od 0 do 1, bez obzira na to što sadržaj tih parova može biti različite prirode (npr. povezanost

visine i tjelesne mase osoba, inteligencije i prosječne fakultetske ocjene, broja crvenih krvnih zrnaca i ekstravertnosti). Velika većina koeficijenata povezanosti između dvije dimenzione varijable imaju i smjer povezanosti koji može biti ili pozitivan tako da porast vrijednosti varijable X prati porast vrijednosti varijable Y, ili negativan, tako da da porast vrijednosti varijable Y prati snižavanje vrijednosti varijable Y (npr. broj sati učenja korelira negativno sa intenzitetom ispitne anksioznosti). Sve u svemu, ova mjerna skala ima teorijski raspon od -1 (teorijski minimum) do +1 (teorijski maksimum). Apsolutni teorijski maksimum se u praksi dostižu samo onda kada varijable koreliraju same sa sobom ili sa svojom linearnom transformacijom (npr. korelacija visine izražene u centimetrima i visine izražene u inčima), dok se apsolutni teorijski minimum dobija kada varijablu koreliramo sa njenom inverznom verzijom. Budući da ne postoji teoretska mogućnost da se pređe apsolutna vrijednost 1, kada budemo saopštavali konkretne vrijednosti koeficijenata korelacija obično ćemo izostavljati brojku 0 koja bi došla sa lijeve strane decimalnog znaka. Uz to, korelacije se obično saopštavaju na dvije decimale tačnosti, a predznak se obično ne navodi kada je korelacija pozitivna (npr. za negativnu korelaciju -.34, ali za pozitivnu .25).

Uobičajen primjer tabele u APA formatu možete vidjeti niže (Slika 7.8) na primjeru povezanosti varijabli koje su prikazane na Slici 7.15.

Tabela 7.1

Aritmetičke sredine, standardne devijacije i interkorelacije HEXACO crta ličnosti

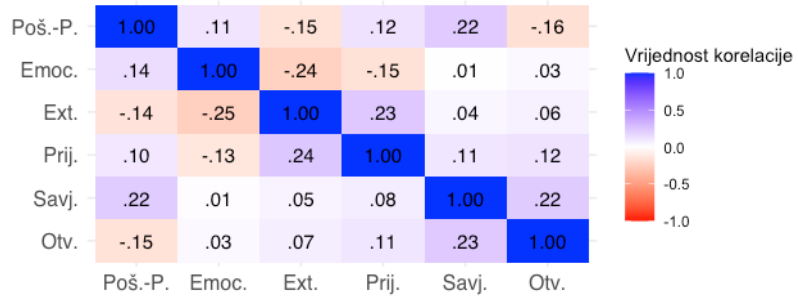
Varijabla	*M*	*SD*	1	2	3	4	5
1. Poštenje-poniznost	2.38	0.90					
2. Emocionalnost	1.51	0.87	.14				
3. Ekstraverzija	2.87	0.73	-.14	-.25			
4. Prijatnost	2.23	0.64	.10	-.13	.24		
5. Savjesnost	2.40	0.71	.22	.01	.05	.08	
6. Otvorenost	2.19	0.73	-.15	.03	.07	.11	.23

Slika 7.15: Tabela koja prikazuje koeficijente korelacija odnosa između HEXACO crta ličnosti (n=249).

Kao što možete vidjeti izostavljene su vrijednosti u dijagonalama (koje bi bile jednake 1.00 jer su to polja koja bi prikazivala korelacije varijable sa samom sobom), te vrijednosti iznad te dijagonale koje bi zapravo bile jednake vrijednostima ispod dijagonale. Drugim riječima, ako smo dobili korelaciju između crta

ličnosti *poštenje-poniznost* i *emocionalnost* koja iznosi .14, onda će i korelacija između *emocionalnosti* i *poštenja-poniznosti* iznositi .14, te nema potrebe da tabelu zasićujemo redundantnim informacijama.

Toplotna tabela (engl. *highlight table*) prikazuje iste ove podatke na nešto zanimljiviji način (Slika 7.16). Polja unutar koji su prikazane vrijednosti korelacija su obojena različito, zavisno od smjera i visine korelacija. Plava boja označava pozitivne, crvena negativne korelacije, a zasićenost bojom ukazuje na snagu veze. Na ovom prikazu su prikazane i jedinične vrijednosti na dijagonali, kao i vrijednosti sa obje strane dijagonale koje označavaju dvije različite vrste koeficijenata korelacija o kojima će najviše biti riječi u ostatku ovog poglavlja.



Slika 7.16: Toplotna tabela koja prikazuje koeficijente korelacija međuodnosa HEXACO crta ličnosti (n=249). Ispod dijagonale su date vrijednosti linearnog, a iznad dijagonale koeficijenta korelacije rangova.

Kakav je to linearni koeficijent korelacije?

Među velikim brojem različitih koeficijenata korelacija, najpoznatiji i najupotrebljavaniji linearni koeficijent korelacije. Linearni koeficijent korelacije je poznat i pod imenom Pearsonov koeficijent korelacije, prema svom glavnom propagatoru, čuvenom statističaru Karl Pearsonu koji je zaslužan i za neke druge statističke tehnike. (I opet tragovi nepravde u statističkom svijetu: prvi autor linearnog koeficijenta korelacija je zapravo francuski fizičar Auguste Bravais, pa se u njegovu čast koeficijent - doduše vrlo rijetko - naziva i Bravais-Pearsonov koeficijent korelacije.) Vrijednost linearnog koeficijenta korelacije nam sažeto opisuje smjer i snagu odnosa dvije varijable koji se izražava linearnom funkcijom, odnosno pravom linijom. Oznaka za linearni koeficijent korelacije na uzorku je latinično malo slovo r (odnosno r_{xy}, dok se traženi populacioni parametar označava malim grčkim slovom ρ (čita se: ro).

Šta nam formula za izračunavanje govori o prirodi linearnog koeficijenta korelacije?

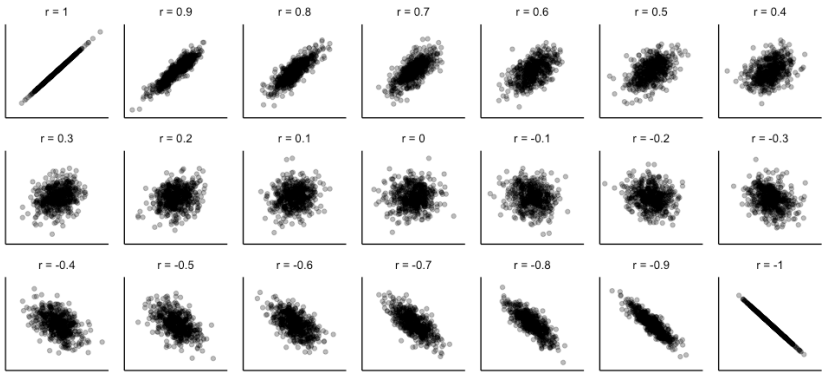
Postoji više formula kojima se dolazi do linearnog koeficijenta korelacije, ali je za studente psihologije 1. godine studija najbitnija konceptualna formula koja glasi:

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

Navedena formula pokazuje da je linearni koeficijent korelacije prosječan proizvod z-skorova na pojedinačnim varijablama. Formula odražava opštu logiku da ako u nekoj grupi prevladavaju osobe koje imaju saglasne parove skorova - tj. natprosječno visoke skorove na obje varijable ili natprosječno niske na obje varijable - onda će koeficijent korelacije biti pozitivan. Suprotno tome, ukoliko prevladavaju nesaglasni skorovi - gdje entiteti koji imaju natprosječan skor na jednoj varijabli teže da istovremeno imaju ispodprosječan na drugoj varijabli - onda će koeficijent korelacije biti negativan. Nulti koeficijent korelacije će se dobiti ukoliko je taj odnos uravnotežen, što bi značilo da statistička povezanost ne postoji.

Na slici ispod možete da uporedite izgled korelacija različitog smjera i intenziteta. Što je korelacija viša, to su manja odstupanja pojedinačnih skorova (tačkica) od prave linije koja predstavlja odnos varijabli, odnosno u tim slučajevima sama linija bolje predstavlja odnos. To istovremeno znači da je greška modela manja. Niže (Slika 7.17) je dat primjer različitih koeficijenata korelacija "okruglih brojeva".[2]

[2] Preporučujem da izoštrite oko vježbanjem uz neki od appleta na internetu koji vam daju zadatak da procijenite numeričku vrijednost na osnovu grafikona raspršenja: https://istics.net/Correlations/, https://www.rossmanchance.com/applets/GuessCorrelation.html, http://onlinestatbook.com/2/describing_bivariate_data/pearson_demo.html, https://www.geogebra.org/m/KE6JfuF9 ili http://guessthecorrelation.com/.



Slika 7.17: Grafikoni raspršenja koji prikazuju tipičnu bivarijatnu raspodjelu za "okrugle" koeficijente linearne korelacije.

Kako tumačiti visinu koeficijenata korelacije?

Većina studenata, kao i psihologa istraživača želi da prevede dobijene numeričke podatke u što konkretnije jezičke odrednice. To je potrebno i kako bismo saopštili drugima rezultate na neki smisleniji način, ali i za potrebe planiranja izvođenja istraživanja, odnosno planiranja toga koliko veliki uzorak nam je potreban da bismo ispitali hipoteze (broj ispitanika utiče na stabilnost veličine korelacija). Iz tih razloga je izvedeno nekoliko klasifikacija veličine korelacija. U društvenim naukama je duže vremena najpopularnija - ali time i najzloupotrebljavanija - bila Cohenova klasifikacija[3] koja podrazumijeva sljedeće tumačenje visina koeficijenata korelacija:

[3] Cohen, 1988, Statistical Power Analysis for the Behavioral Sciences

- .00 - .09 zanemariva, marginalna povezanost
- .10 - .29 niska povezanost
- .30 - .49 umjerena povezanost
- ≥ .50 visoka povezanost.

Jacob Cohen (čuveni psiholog i statističar, a ne modni brend skupe džins odjeće) je svoju klasifikaciju zasnovao na temelju prostog gledanja u grafikone raspršenja. Najprije je odredio da bi granični skor za umjerenu povezanost trebalo da bude .30 jer je smatrao da većina ljudi može opaziti takvu povezanost golim okom.[4] Nakon toga je zaključio da kada se korelacija spusti na .10 povezanost postaje "vidno manja" od umjerene, a kada se popne na .50 ona postaje "vidno veća". Uprkos očitoj privlačnosti ovako jednostavne, čisto statističke klasifikacije, i sam Cohen je kukao na to što se ove "vidne" granice slijepo primjenjuju. Naglašavao je da je jedino ispravno da veličine koeficijenata korelacije u svakom pojedinačnom istraživanju tumačimo na osnovu nekih smislenijih kriterijuma koje navode i drugi eksperti:[5]

[4] Correl et al., 2020, https://doi.org/10.1016/j.tics.2019.12.009

[5] npr. Funder & Ozer, 2019, https://doi.org/10.1177/2515245919847202

- Kao prvo, trebalo bi da preispitamo postoje li neka teorijska očekivanja za dobijanje određenog koeficijenta korelacije između ispitivanih varijabli. Na primjer, ukoliko bismo izvodili novo istraživanje na temu povezanosti opšte kognitivne sposobnosti (čitaj: g-faktor ili IQ skor) i ocjena na nižim nivoima školovanja trebalo bi da očekujemo da dobijemo korelacije negdje u okviru $r = 0.40—0.50$, budući da su takvi nalazi dobijani u većini ranijih istraživanja. Ako bismo dobili značajno više ili značajno niže vrijednosti, onda bi trebalo da nam se uključe lampice

i alarmi i da krenemo da tragamo za razloge takve nedosljednosti. (Usput, u našem istraživanju o povezanosti inteligencije i srednjoškolskih ocjena za kojeg ste vidjeli grafikone raspršenja i reziduala smo dobili rezultat $r = .39$ što se uklapa u teorijski okvir, jer se minimalna odstupanja od navedenog intervala mogu očekivati usljed različitih faktora, a među njima je i korištenje kratkog instrumenta za testiranje inteligencije.)

- Drugi vid kriterija koji vrijedi konsultovati je nastao u posljednje vrijeme zahvaljujući lakoj dostupnosti bibliografskih podataka i mogućnosti da se naprave sekundarna istraživanja na već objavljenim naučnim radovima. Konkretno, izvedeno je već nekoliko istraživanja koja su za cilj imala da utvrde raspodjelu visina koeficijenata korelacija u naučnim člancima u pojedinim oblastima. Autori tih radova smatraju da su empirijski kvartili (srećom, znate već šta su to) tih distribucija razumne referentne tačke da se visine korelacije podijele u male, umjerene i visoke. Tako je izvedeno da bi u psihologiji individualnih razlika[6] granica malog efekta bila .10, srednjeg .20, a velikog .30, dok bi za oblast socijalne psihologije[7] granice mogle biti .12 za mali, .24 za srednji i .41 za veliki efekat. Očito je da ove referentne vrijednosti mogu da značajno variraju od oblasti do oblasti (eto, i podoblasti unutar jedne discipline), pa iako ovakvi kriteriji mogu biti zanimljiviji usljed empirijske pozadine, njihovo slijepo korištenje potencijalno vodi sudbini koju je doživjelo i nekritično korištenje Cohenovog prijedloga.
- Treći vid kriterija je najbitniji, a tiče se pragmatičnog, odnosno posljedičnog značaja na koje visina korelacije može ukazivati. To se najbolje vidi iz primjera koji pokazuju da statistički "zanemarive" visine korelacija mogu imati nezanemarive posljedice na odluke koje se mogu donijeti. Tako pomenuti Funder i Ozer navode primjer da statistički zanemariva korelacija od .05 između izraženosti crte ličnosti prijatnost (sklonost da se u odnosima bude ljubazan, saradljiv, empatičan, povjerljiv) i uspješnih društvenih interakcija može imati itekako značajan efekat na duge staze, budući da svaka osoba stupa u veliki broj društvenih interakcija tokom, na primjer, jedne godine. Drugi, redovno navođeni primjer je jedno starije medicinsko istraživanje o efektu aspirina na prevenciju srčanog udara. Naime, koeficijent korelacije utvrđen u velikom kontrolisanom randomizovanom eksperimentu (kontrolna grupa je dobijala placebo) je iznosio mizernih .03 na kraju

[6] Gignac & Szodorai, 2016, https://doi.org/10.1016/j.paid.2016.06.069

[7] Lovakov & Agadullina, 2021, https://doi.org/10.1002/ejsp.2752

istraživanja.[8] Ali kad te iste rezultate prevedemo, pa na primjer shvatimo da na 10 srčanih udara među onima koji su primali aspirin dolazi otprilike 18 srčanih udara u grupi koja je pila placebo, onda shvatamo da je rizik za srčani udar bio gotovo prepolovljen, odnosno skoro pa smanjen za 50% (konkretna računica: $\frac{18-10}{18} * 100 = 44.4\%$) ![9]

[8] Steering Comittee of the Physicians' Health Study Research Group, 1989, https://doi.org/10.1056/NEJM198907203210301

[9] Doduše, radilo se o istraživanju gdje su u pitanju bile dvije dihotomne varijable - sa kategorijama kontrolna/eksperimentalna grupa i doživio/nije doživio srčani udar - ali se isti principi mogu primijeniti jer je korišten fi-koeficijent, aritmetički identičan linearnom koeficijentu korelacije što će kasnije u tekstu biti pojašnjeno.

Zašto se uopšte povezanosti predstavljaju standardizovanim mjerama kao što su koeficijenti korelacija?

Osnovna prednost korištenja standardizovanih mjera kao što su koeficijenti korelacija jeste to što se na taj način mogu uporediti snage povezanosti između dvije varijable, što u velikom broju slučajeva znači i između različitih prediktora i jedne ishodišne varijable. Na primjer, u jednom istraživanju smo pronašli da različite crte ličnosti koreliraju sa uspjehom u studiranju psihologije, operacionalizovanom kao ocjena na kraju studija prvog ciklusa na Univerzitetu u Banjoj Luci.[10] Konkretno, i savjesnost i otvorenost za iskustva procjenjivani na prvoj godini studija su pozitivno korelirali sa ocjenama na kraju prvog ciklusa studija. Međutim, pronađeno je da je otvorenost za iskustva snažniji korelat, budući da je korelacija između te varijable i prosječne ocjene iznosila .33, dok je korelacija savjesnosti sa prosječnom ocjenom bila tek .09. Iz navedenog bi se moglo zaključiti da je za bolji uspjeh na datom studiju bitnije biti radoznao za nova, intelektualna iskustva, nego savjesno i redovno obavljati obaveze. Ipak savjesnost izgleda bitna i kod nas budući da postoji pozitivna korelacija.

[10] Lakić, Damjenić i Šain, 2017, Velikih pet kao prediktori uspjeha u studiranju psihologije na Univerzitetu u Banjoj Luci

Nadalje, standardizovana mjerna skala omogućava da uporedimo korelacije identičnih teorijskih varijabli kada koristimo različite instrumente za taj predmet mjerenja ili kada isti instrument koristimo na različitim uzorcima (npr. poređenje između različitih država). Primjera radi, u gore navedenom istraživanju su i prijatelji koji poznaju studenta barem 3 godine procjenjivali osobine ličnosti studenata psihologije. Za tako procijenjenu otvorenost za iskustva dobijena je nešto niža, ali snagom slična korelacija sa prosjekom ocjena (.30), dok je za savjesnost dobijena istovjetna korelacija (.09). Može se zaključiti da navedeni rezultati potvrđuju zaključak o relativnom značaju ove dvije crte za uspjeh u studiranju psihologije na Univerzitetu u Banjoj Luci. Međutim, obuhvatno istraživanje[11] na anglofonim studijima psihologije - gdje su skoro isključivo uzorci dolazili iz SAD, Velika Britanija i Australija - prosječne korelacije su zamijenile mjesta: korelacija prosječne ocjene sa savjesnošću je

[11] Vedel, 2014, https://doi.org/10.1016/j.paid.2014.07.011

bila viša (.32) nego korelacija sa otvorenošću za iskustva (.07). Sve u svemu, zaključak bi bio samo dijelom isti kao i naš: iako su obje crte ličnosti pozitivno povezane sa uspjehom u studiranju psihologije, postoje razlike u redoslijedu značaja, a zašto je to tako možete i vi ponuditi odgovore.

Postoje li nedostaci standardizovanog predstavljanja povezanosti putem koeficijenata korelacija?

Dosad ste morali shvatiti da korištenje standardizovanih mjera ima i svoje mane. Kao prvo, radi se o potpuno apstraktnom, matematičkom iskazivanju povezanosti koji je izgubilo vezu sa originalnim mjernim skalama. Dakle, nerijetko je problem shvatiti šta uopšte koeficijenti korelacije poručuju budući da se odnos između prediktorske i ishodišne varijable ne iskazuje u originalnim mjerama. Nešto što se zove linearna regresiona analiza - i što ostavljamo za kasnije - je postupak koji omogućava da se odnos varijabli iskaže upravo kroz njih. Drugi problem je da koeficijenti korelacija nisu aditivne mjere, što u prevodu znači da ih ne možete direktno sabirati i oduzimati, niti ih omjerno porediti tako da kažete da je korelacija od .20 duplo veća od .10 (jer to nije tačno). Da biste vršili takve operacije i poređenja, koeficijente korelacija morate transformisati na mjerne skale koje su za to primjerene. Konkretno, morate ili kvadrirati koeficijente korelacija pa tako stvoriti tzv. koeficijente determinacija (vidjeti sljedeći segment) ili koristiti tzv. *r*-u-*z* Fisher-ovu transformaciju pomoću softvera (ili ako ste ultra-tradicionalista, koristeći specijalizovane tablice koje možete naći u nekim starim udžbenicima).

Šta je to onda koeficijent determinacije?

Koeficijent determinacije (engl. *coefficient of determination*) je mjera količine određenosti varijabiliteta varijable nekom drugom varijablom. Da, dobro ste pročitali (pročitajte još jednom ako ne shvatate). Radi se o "matematičkom" principu tumačenja visine koeficijenta korelacije koji nema baš praktični značaj, ali se često sreće u naprednijim statističkim analizama i koristi za upoređivanje kvaliteta različitih statističkih modela. Vjerovatno je jednostavnije definisati koeficijent determinacije kao običnu kvadriranu vrijednost koeficijenta korelacije, a označava se sa r^2. S obzirom na način kako se računa ($r^2 = r * r$), lako ćete zaključiti da se koeficijent determinacije takođe može kretati u rasponu od 0 do 1, ali da nema smjer (uvijek je pozitivan, usljed

kvadriranja). Ono što njegova visina pokazuje jeste proporcija zajedničkog varijabiliteta dvije varijable (npr. prediktorske i ishodišne varijable). Iako se obično u statističkim izvještajima koeficijenti determinacije izražavaju kao proporcije, pri saopštavanju rezultata, odnosno njihovom čitanju, govorimo o procentima. Drugim riječima, u naučnim radovima možemo naići na podatak da je koeficijent korelacije iznosio .30, što znači da je 9.0% varijanse kriterijske varijable zajedničko sa variranjem prediktorske varijable (za svaki slučaj, evo kako je to dobijeno: $0.30 * 0.30 = 0.09 \rightarrow 0.09 * 100 = 9.0\%$).[12]

[12] Preporučujem da koristite sljedeći applet koja će vam pomoći da preciznije shvatite odnos između datih mjera i podataka na grafikonu raspršenja: http://rpsychologist.com/d3/correlation/

Na ovom mjestu dva upozorenja. Naime, često ćete moći pročitati da prediktorska varijabla objašnjava varijansu ishodišne varijable (npr. da prediktor objašnjava 9% varijanse kriterija). Ovaj termin "objašnjava" (engl. *to account for*) je jedan od najzavaravajućih termina u čitavoj statistici uopšte i treba ga oprezno koristiti. Naročito u korelacionim istraživanjima - ali čak i u eksperimentalnim koja nisu uzela u obzir dejstva svih spoljnih varijabli - je uvijek validno pitanje da li jedna varijabla zaista uzrokuje, odnosno objašnjava drugu. O korelaciji i kauzaciji ćemo malo više u jednom od narednih segmenata, ali ovdje je bitno da napomeno još nešto čega moramo biti svjesni.

Ukoliko dvije varijable dijele 9.0% varijabiliteta, time se istovremeno podrazumijeva da je ostalo čak 91.0% varijanse ishodišne varijable koje se "objašnjava" greškom modela, dakle činiocima koji nisu eksplicitno uključeni u model. Neobjašnjeni ostatak se naziva **koeficijent indeterminacije** i jednostavno se izračunava kao $1 - r^2$. U društvenim naukama je koeficijent indeterminacije redovno veći od koeficijenta determinacije. Na primjer, za koeficijent korelacije od .20, koeficijent determinacije iznosi tek .04 odnosno postoji samo 4.0% preklopljene varijanse među varijablama. U ostatak od 96.0% ulaze kako i greška modela (treće varijable koje nismo uzeli u model), greška uzorkovanja (to što smo u uzorak izabrali te ispitanike, a ne neke druge iz populacije) ali i greška mjerenja. Na primjer, jednom operacionalizacijom ne mjerimo samo konstrukt od interesa nego i neke sporedne, slabije relevantne aspekte. Na primjer, mjereći ekstraverziju kao konstrukt od interesa mi nekim stavkama mjerimo socijabilnost, drugima energičnost, trećima asertivnost, pritom stavke unutar faceta mogu biti vezane za specifičan kontekst (npr. neka pitanja o socijabilnosti mogu biti u vezi sa odnosima sa prijateljima, druga u odnosu sa nepoznatim osobama). Greške mjerenja mogu proizaći i iz drugih razloga kao što je, na primjer, korištenje isključivo metoda samoprocjene.

Koji uslovi treba da budu ispunjeni pa da linearni koeficijent korelacije daje suvislu procjenu povezanosti?

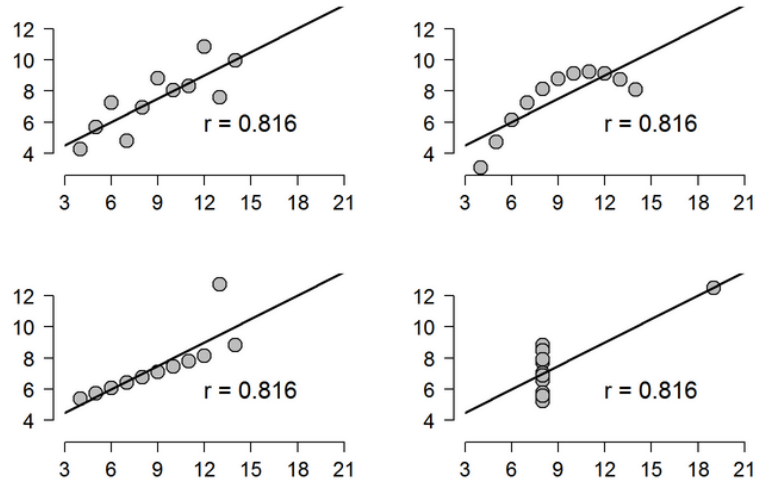
Postoje 3 osnovna uslova koje je potrebno ispuniti da bi linearni koeficijent korelacije "vršio" svoju ulogu kako treba:

- Obje varijable bi trebalo da su barem intervalnog nivoa mjerenja, odnosno da jedinice na mjernim skalama izražavaju ravnomjerni porast količine mjerenog svojstva. Međutim, znamo da su u psihologiji prave intervalne varijable manjina, jer u najvećem broju slučajeva moramo da koristimo kompozitno mjerenje, te kao rezultat imamo kvazi-intervalne varijable. Budući da između neke operacionalizovane varijable i izraženosti konstrukta koji leži u njenoj osnovi može postojati nelinearan odnos (baš tako, razmislite o tome ako imate vremena) linearni koeficijenti korelacija mogu promašiti u procjeni. Međutim, vjerovatno taj promašaj neće biti pretjeran. Iz tog razloga apsolutno niko ne "diže frku" kada se linearni koeficijent korelacija za procjenu odnosa između varijabli kada je barem jedna od njih kvazi-intervalna. Štaviše, u praksi ćete vidjeti da mnogi istraživači koriste linearni koeficijent korelacije i kada su jedna ili obje varijable ordinalne kategoričke. To nije nešto što bih preporučio jer bolje alternative postoje (vidjeti u kasnijem poglavlju), ali u iznimnim slučajevima može poslužiti kao brza procjena.
- Odnos između varijabli bi trebalo da je linearan. U praksi moramo ovo modifikovati na "linearan u dovoljnoj mjeri". Već smo se upoznali sa grafičkim načinima provjere linearnosti odnosa, ali se često postavlja pitanje u kojoj mjeri odnos može da odstupa od linearnog, pa da i dalje ostanemo pri korištenju linearnog koeficijent korelacije. Iako su razvijeni neki specijalizovani testovi linearnosti, odluku je najbolje donijeti na osnovu grafičke procjene i analize toga da li zaista postoji nešto što linearni model propušta da uzme u obzir.
- Ne bismo smjeli imati atipične skorove koji značajno utiču na procjenu snage i/ili smjera linearne povezanosti varijabli. Već smo se upoznali sa univarijatnim stršećim skorovima, odnosno sa skorovima koji vidno odstupaju od ostatka uzorka. Oni će biti problem kada izračunavamo linearni koeficijent korelacije, ali još veći problem predstavljaju bivarijatni stršeći skorovi, odnosno ispitanici koji na obje varijable postižu atipične vrijednosti. Kao i kada je u pitanju procjena "količine nelinearnosti" i u ovom slučaju

je odluku najbolje postići pratećom sadržinskom analizom.

Šta ako ti uslovi nisu ispunjeni?

Linearni koeficijent korelacije može lako da potcijeni ili precijeni stvarnu povezanost kada uslovi za njegovo računanje nisu ispunjeni. Negativan uticaj nelinearnosti i prisutnosti ekstremnih skorovi na vrijednost *r* koeficijenta se obično prikazuje putem Anscombeovog kvarteta (Slika 7.18). Naime, iako se prikazana četiri potpuno različita odnosa između dvije varijable, za sve njih je linearni koeficijent korelacije jednak ($r = .82$). Pritom, samo odnos prikazan u gornjem lijevom kvadrantu zadovoljava uslove korištenja linearnog koeficijenta. U gornjem desnom kvadrantu je prikazana jedna nemonotona povezanost u kojoj veza između varijabli na jednom mjestu prelazi iz pozitivne u negativnu. Ta povezanost bi se mogla baš savršeno predstaviti nelinearnom funkcijom (drugim riječima korelacija bi trebalo da iznosi 1). Na donjoj polovini su prikazani odnosi varijabli gdje postoji ekstremni skor na jednoj (lijevo) ili istovremeno na obje dimenzije (desno). Na ovim grafikonima samo jedan ispitanik čini razliku: da nema stršećih slučajeva u lijevom kvadrantu bi korelacija iznosila 1, a u desnom 0.



Slika 7.18: Prikaz Anskombovog kvarteta preuzet sa internet stranice: http://shinyapps.org/apps/RGraphCompendium/index.php

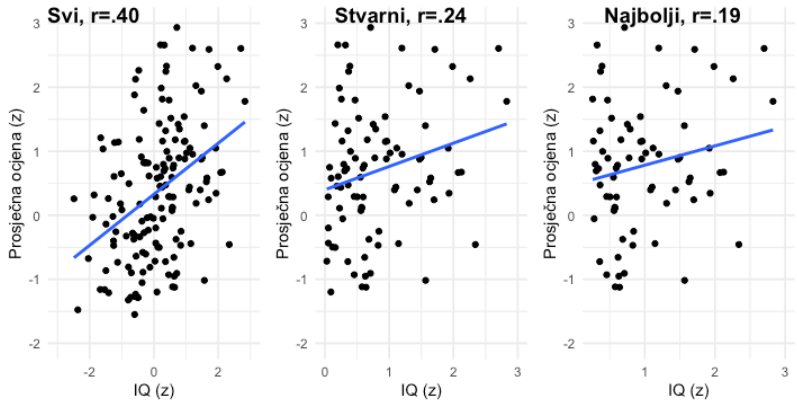
Navedena pitanja su specifični problemi linearnog koeficijenta korelacije. Prije nego što pređemo na alternativna rješenja, proći ćemo još neke važne dileme u tumačenju svih koeficijenata korelacija.

Kako ograničen obuhvat ispitanika može uticati na koeficijent korelacije?

Ograničenje opsega ili restrikcija ranga se odnosi na situaciju kada uzorkom nismo zahvatili čitav opseg varijabilnosti jedne ili obje varijable, te kao posljedicu toga dobijemo netačnu procjenu populacione vrijednosti koeficijenta korelacije. Na primjer, zamislimo realistični scenario kojim želimo da empirijski provjerimo postoji li povezanost između inteligencije (mjereno na prijemnom ispitu) i prosječne ocjenu na kraju prvog ciklusa studija psihologije. Recimo da smo na prijemnom ispitu imali 150 kandidata. Od tog broja njih 50 je primljeno na studij. Međutim, treba da imamo na umu da je na našem fakultetu test kognitivne sposobnosti sastavna komponenta prijemnog ispita rezultat na testu, što znači da rezultat na testu istovremeno utiče i na to ko će biti primljen na studij (odnosno, inteligentnije osobe imaju veću šansu da postanu studenti).

Slika 7.19 pokazuje tri grafikona raspršenja. Srednji grafikon opisuje ono što bismo mogli realistično dobiti na uzorku kandidata koji su primljeni na studij. Na tom uzorku korelacija između inteligencije i prosječne ocjene iznosi $.24$. Sa lijeve strane vidimo rezultate imaginarnog scenarija. Naime, ukoliko bismo na studij primili sve prijavljene kandidate, bez obzira na njihovu inteligenciju, na kraju studiranju bismo mogli da utvrdimo da bi stvarna korelacija iznosila $.40$, jer bi manje inteligentni ispitanici postizali niže ocjene. Zaključak je da usljed već selekcionisanog uzorka mi u praksi dobijamo nižu procjenu povezanosti ove dvije varijable nego što je to slučaj u populaciji. Konačno, treća slika pokazuje situaciju u kojoj bismo izabrali samo inteligentnije ispitanike - one koji imaju IQ veći za 0.25 standardnih devijacija u odnosu na ukupnu grupu. Korelacija na takvom poduzorku postaje još niža ($r = .19$. Kada bismo išli dalje mogli bismo zamisliti i situaciju u kojoj bismo na uzorku dobili nultu korelaciju, ili čak negativnu!

Drugim riječima, restrikcija opsega jedne (ili obje) varijable može pristrasno uticati na procjenu parametra. Intenzitet korelacije na uzorku može biti pojačan, smanjen, takav da korelacija bude nulta ili čak da korelacija okrene smjer u odnosu na stvarnu populacionu korelaciju. Nauk je da je uvijek neophodno da razmotrimo u kojoj mjeri uzorak dobro predstavlja varijabilnost u populaciji, te da budemo svjesni da može doći do ozbiljnog razmimoilaženja statistika i parametra ukoliko nismo zahvatili dovoljno raznovrstan uzorak.

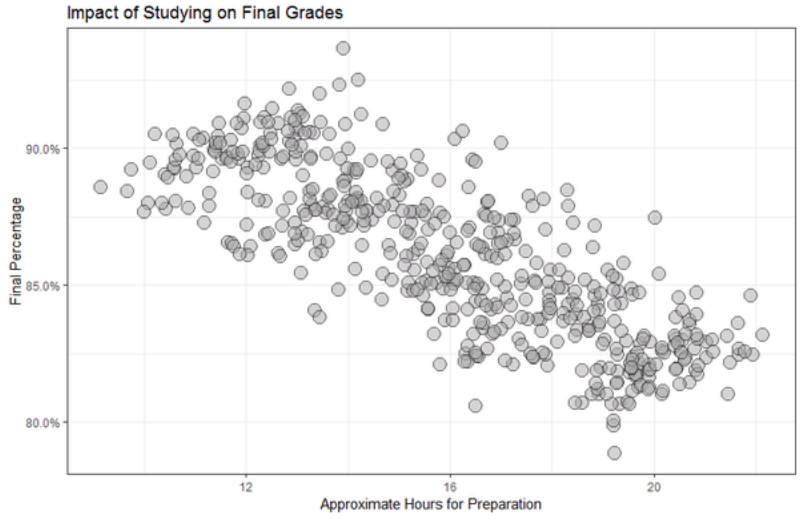


Slika 7.19: Ilustracija potencijalnog efekta ograničenog opsega na visinu korelacije.

Na koji način neka treća, prikrivena varijabla, može uticati na povezanost između dvije varijable?

Simpsonov paradoks (engl. *Simpson's paradox*) je često pominjana statistička enigma. Radi se o tome da povezanost između dvije varijable može imati jedan izgled na ukupnom uzorku (ili populaciji), a da se ispostavi da je potpuno drugačija na poduzorcima (ili podpopulacijama) koje čine taj ukupan uzorak (ili populaciju). Na Slici 7.13a nalazi se "neprirodna" negativna povezanost na cijelom uzorku između sati provedenih u učenju predmeta i dobijenog broja bodova za ocjenu (izraženog u postocima; podaci su fiktivni).[13]

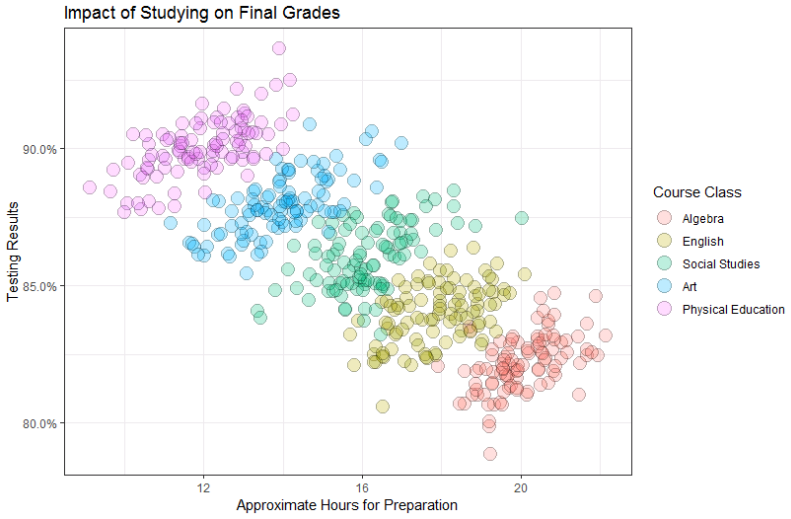
[13] Ilustracije su preuzete iz Quora odgovora Jon Waylanda sa https://www.quora.com/What-is-Simpsons-paradox koji tu navodi i R kod za grafikone.



Slika 7.20: Ilustracija Simpsonovog paradoksa - puni uzorak.

Na Slici 7.13b su prikazani isti ti podaci, samo što su sada bojama označeni podaci za pet predmeta (fizičko, umjetnost, društvene

nauke, engleski jezik, algebra). Na tom grafikonu se vidi da zapravo unutar svakog predmeta postoji pozitivna povezanost, odnosno da osobe koje su duže učile teže da dobiju više ocjene. Kako sad to? Razlog za ovaj paradoks leži u tome što je vrsta predmeta povezano sa obje varijable: i sa dužinom učenja i sa bodovima za ocjenu.



Slika 7.21: Ilustracija Simpsonovog paradoksa - poduzorci.

Ali šta je onda istina? Da li se može zaključiti da je dužina vremena učenja pozitivno ili negativno povezano sa ocjenama? Da bismo dobili odgovor na ovo pitanje moramo hrabro izaći van teritorije klasične statistike i ozbiljnije se pozabaviti pričom o kauzalnosti, odnosno uzročno-posljedičnim vezama. Rješavanje ovog, a i analognih paradoksa, zahtijeva da uključimo svoje logičko-racionalne sposobnosti i sadržinska znanja kako bismo odredili šta je čemu moglo biti (djelimični) uzrok. U ovom konkretnom slučaju će nam stvari olakšati ako u model uvedemo još jednu varijablu koja nije bila direktno mjerena, a to bi bila kompleksnost / zahtjevnost predmeta. Ta varijabla je skoro savršeno povezana sa varijablom vrsta predmeta za koju je rang očit - najzahtjevniji predmet je bila algebra, a najmanje zahtjevan fizičko. Istovremeno, ta varijabla uslovljava dužinu učenja (a ne obrnuto), te ista ta varijabla uslovljava savladanost gradiva koja se boduje. Dakle, zahtjevnost predmeta - preko vrste predmeta - uzrokuje obje varijable, pa da bi se otkrio stvarni odnos između njih, potrebno je eliminisati uticaj zahtjevnosti predmeta. To se radi tako što se analiza vrši po podgrupama, odnosno držeći zahtjevnost predmeta konstantnom. Kada to uradimo vidimo da je povezanost pozitivna, što je sasvim logično.

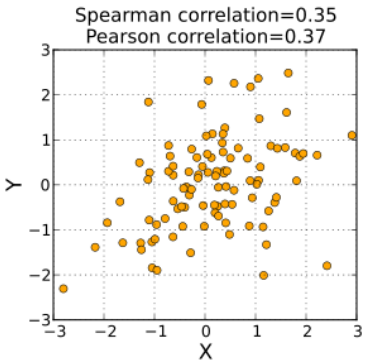
Šta nam onda koeficijenti korelacija govore o uzročnosti?

Simpsonov paradoks je samo jedan vid problematičnog zaključivanja o kauzalnim odnosima na osnovu uočene korelacije. Tačno je da je u ogromnoj većini slučajeva neophodno da postoji korelacija kako bismo govorili o kauzaciji, ali već vam je poznato iz opštih metodoloških principa, da ako uočimo statističku korelaciju to nije ni izbliza garant da se radi o kauzalnom odnosu između dvije varijable. Prikaz niza besmislenih korelacija koje očito ne stoje u direktnom uzročno-posljedičnom odnosu možete pronaći na sajtu http://tylervigen.com

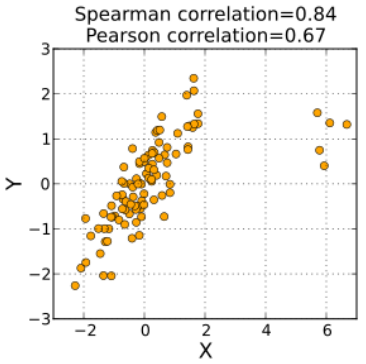
Ipak, korelacija zaista ukazuje na neku povezanost. Razlog za nju zaista može biti kauzalni odnos. Međutim, to može biti i sistematsko dejstvo neke treće varijable koja je njihov zajednički uzročnik, dok će se na mnogim uzorcima povezanost izroditi kao čista posljedica slučajnosti. Iz tog razloga nam ni osnovna, ali čak ni napredna statistika - koja obuhvata tehnike kao što su strukturalno modeliranje, analiza medijacije i moderacije i čitava nova oblast kauzalnog statističkog zaključivanja koja je u zamahu - ne mogu samostalno dokazati da kauzalni odnos postoji. Argumente za kauzaciju moramo prikupljati pomoću različitih pristupa koji uključuju ekspertska znanja i metodološke intervencije. Na primjer, situacija je nešto jasnija kada podaci dolaze iz eksperimentalnih studija gdje postoji manipulacija nezavisnom varijablom i aktivna kontrola relevantnih spoljnih varijabli. Međutim, čak i tada postoji šansa da su neke bitne spoljne varijable prikriveno djelovanje na rezultate na uzorku.

Situacija je daleko zamršenija kada su u pitanju neeksperimentalne studije. Relativno donedavno je u mainstream nauci postojala zabrana da se kauzalnost razmatra u tim slučajevima. Međutim, u posljednje vrijeme je opšte stanovište da je razložno argumentovati da se radi o kauzalnom odnosu ukoliko istraživač može da izloži sljedeće vrste ubjedljivih dokaza sa svoju tvrdnju:[14] direktne dokaze da postoji vremenski slijed između pretpostavljene uzročne i posljedične varijable, izložen razložan sadržinsko-logički mehanizam koji objašnjava vezu, paralelne dokaze iz više nezavisnih studija da je efekat postojan, kao i da su alternativna objašnjenja isključena. Sve u svemu, neophodno je da zapamtimo da "korelacija ne znači kauzaciju", ali da je potpuno legitimno da tragamo da li i zašto dolazi do nekog kauzalnog odnosa čak i u neeksperimentalnim studijama.

[14] prema Howick, 2011, https://doi.org/10.1002/9781444342673

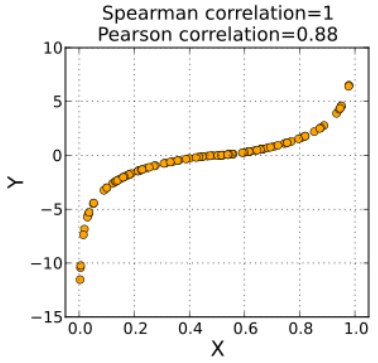


Slika 7.22: Poređenje Pearsonovog i Spearmanovog koeficijenta korelacije - zadovoljeni uslovi (ilustracije preuzete sa Wikipedia stranice o Spearmanovom koeficijentu).



Slika 7.23: Pearson vs. Spearman - atipični skorovi.

[15] de Winter et al., 2016, https://doi.org/10.1037/met0000079



Slika 7.24: Pearson vs. Spearman - nelinearna monotona veza

Koji su onda ostali koeficijenti korelacija koji se koriste za izražavanje povezanosti dvije dimenzione varijable?

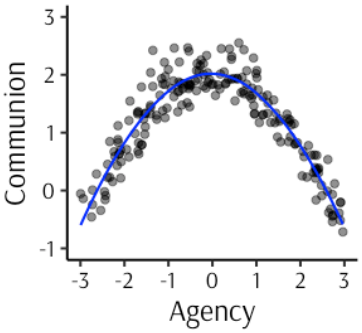
Ovdje obrađujemo samo dva najznačajnija među mnogim predloženim koeficijentima korelacija uz napomenu da ćemo se tek u jednom kasnijem poglavlju baviti i mjerama saglasnosti za ponovljene nacrte, pa ćemo ponovo pominjati linearni koeficijent korelacije.

Spearmanov koeficijent korelacije rangova (engl. *Spearman's rank correlation coefficient*) predstavlja direktnu i jednostavnu alternativu Pearsonovom linearnom koeficijentu.Kada je u pitanju notacija ranije se ovaj statistik označavao grčkim slovom ro (ρ), ali kako ne bi došlo do konfuzije sa mjerom parametra u savremenoj stručnoj literaturi se ovaj statistik označava sa r_s. Kalkulacija je jednostavna: sirovi skorovi za obje varijable se pretvore u odgovarajuće rangove (za svaku varijablu pojedinačno, jednakim skorovima se dodaje prosječni rang) i onda se na njima izračunava Pearsonov linearni koeficijent. Kada su uslovi za linearnu korelaciju zadovoljeni dobijaju se vrlo slični rezultati (Slika 7.22). Pritom rang-korelacija ne samo da značajno eliminiše pristrasne efekte ekstremnih skorova (Slika 7.23), nego eliminiše i negativan efekat nelinearnosti odnosa ukoliko je veza između dvije varijable monotono rastuća ili monotono opadajuća (Slika 7.24). Štaviše, postoji empirijski dokazi[15] da u većini slučajeva rang-korelacija na uzorku bolje procjenjuje parametar linearne korelacije nego što to radi r na uzorku!? Avaj, ni Spearmanov koeficijent ne pomaže kada postoji veza između varijabli koja do određene tačke raste, a zatim opada (nemonotona veza). Na primjer, rezultati za očito vrlo snažnu povezanost prikazanu na Slici 7.25 su $r = -.13$ i $r_s = -.10$.

Vinsorizovani linearni koeficijent korelacije (r_w, odnosno ρ_w) je još jedan koeficijent korelacije između dvije dimenzione varijable koji se može koristiti da zaliječi problem postojanja atipičnih vrijednosti. To je zapravo linearni koeficijent korelacije na podacima sa uzorka gdje su obje varijable prethodno vinsorizovane. Vinsorizovani koeficijent korelacija se pokazuje kao dobra mjera kada imamo simetrično raspoređene pojedinačne varijable, dok je Spearmanov koeficijent bolja opcija kada imamo barem jednu asimetrično raspoređenu varijablu ili kada je odnos nelinearan. Ni Vinsorizovani koeficijent korelacije ne pomaže u slučaju nemonotonih veza.

A šta da radimo kada imamo nemonotonu povezanost?

Ako naiđemo na nemonotone odnose i želimo da numerički izrazimo snagu povezanosti dvije varijable moramo koristiti naprednije tretmane koje nadilaze opseg ovog teksta. Posljednjih godina je smišljeno nekoliko novih koeficijenata korelacija koji kvantifikuju vezu između varijabli bez obzira na monotonost.[16] Kako još uvijek ne postoji jasan sud o njihovim prednostima i manama u realnim kontekstima nećemo dalje o njima.



Slika 7.25: Obrnuta U-kriva za koju ne pomažu uobičajeni koeficijenti korelacija.

[16] npr. maksimalni informacioni koeficijent, https://doi.org/10.1126/science.1205438; Chatarjee xi, https://arxiv.org/abs/2403.17670, koeficijent asimetrične zavisnosti, https://doi.org/10.1016/j.csda.2020.107058

Kako da analiziramo povezanost dvije dimenzione varijable?

Konačni savjeti za analizu povezanosti između dvije numeričke varijable bi bili sljedeći:

1. Uvijek bi trebalo da predstavimo podatke grafički, kako univarijatne, tako i bivarijatnu distribuciju pomoću grafikona raspršenja. Iscrtaćemo linearnu kao i loess funkcije sa različitim parametrima poliranja (ukoliko to softver dozvoljava). Trebalo bi da odgovorimo sebi na pitanja postoje li ekstremni skorovi i procijeniti trend kovariranja (linearnost, monotonost, heteroskedastičnost).
2. Izračunaćemo i Pearsonov i Spearmanov koeficijent korelacije (poželjno i Vinsorizovani) i direktno ćemo ih uporediti. Razlike u apsolutnim vrijednostima između linearnog i rang koeficijenta od .05 ili veće prilično će nam jasno potvrditi da postoje ili bitni stršeći slučajevi (univarijatni ili bivarijatni) ili izraziti nelinearni odnosi između varijabli. Ukoliko smo uočili nelinearnu ili nemonotonu funkciju razmotrićemo zašto je ona mogla nastati.
3. Procijenićemo u kojoj mjeri su dobijeni smjer i visina korelacije - ili njeno nepostojanje - saglasni sa onim što bismo teorijski očekivali i koliku pragmatičnu vrijednost ta korelacija može da ima.
4. Promislićemo o tome postoji li neka teorijska mogućnost da je opseg u uzorku na jednoj ili obje varijable značajno ograničen u odnosu na populacioni opseg i može li to da utiče na naše zaključke o povezanosti.
5. Razmotrićemo pitanje kauzalnosti i postojanja trećih varijabli koje bi mogle da objasne ili promijene dobijeni odnos. Ukoliko identifikujemo takvu mogućnost, onda ćemo napraviti hipoteze o mogućim medijacionim ili modera-

torskim odnosima koje bismo mogli testirati u ovom (samo probno) ili, kako treba, u narednim istraživanjima.

Poglavlje 8

Analiza povezanosti jedne dimenzione i jedne kategoričke varijable

Nakon čitanja ovog poglavlja znaćete:

- obrazložiti da su u nacrtima sa jednom dimenzionom i jednom kategoričkom varijablom razlike i povezanosti dva aspekta jedne pojave,
- koje grafičke i numeričke tehnike efikasno oslikavaju vezu između jedne dimenzione i jedne kategoričke varijable,
- odabrati adekvatne oblike analize na osnovu nivoa mjerenja kategoričke varijable.

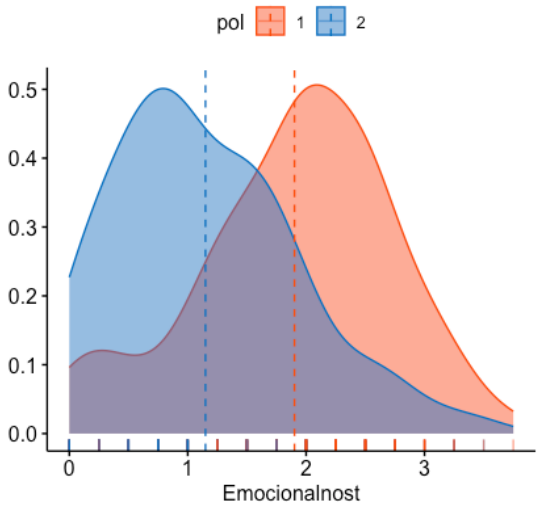
Ovim poglavljem se upoznajemo sa statističkom obradom jednostavnih nacrta koji u vezu dovode jednu dimenzionu i jednu kategoričku varijablu. Prije nego opišem specifične tehnike, naglasiću neke ključne principe koji će igrati ulogu u njihovom odabiru u konkretnim istraživačkim situacijama. Da počnemo od stvari koje neće uticati na taj odabir. Na primjer, nebitni aspekti su stepen kontrole pojedinačnih varijabli - odnosno to da li su one manipulativne, selektivne, ili registrovane - kao i uloga varijable u nacrtu - tj. koja od varijabli je prediktorska, a koja ishodišna. Nasuprot tome, značajnu ulogu u odabiru tehnike ima tip kategoričke varijable (binarna, nominalna ili ordinalna). Vidjećemo da postoji niz različitih koeficijenata korelacija koji se upotrebljavaju u zavisnosti od nivoa obje varijable, a shvatićemo i da se umjesto o povezanostima može **istovremeno** govoriti i o razlikama među grupama u pogledu dimenzione varijable.

Zapravo, odnosi jedne kategoričke i jedne dimenzione varijable se primarno gledaju iz pozicije razlika. A šta se to može razlikovati između grupa? Mnogo toga. U najvećem broju slučajeva su najinteresantnije razlike u mjerama centralne tendencije, jer

one predstavljaju tipičnosti grupa. Međutim, istraživače itekako mogu zanimati i razlike u varijabilnosti između grupa, obliku distribucije (asimetričnost, spljoštenost, modalnost), postojanju atipičnih vrijednosti (npr. da li su izuzetno visoki skorovi svojstveni samo za jednu grupu), kao i pozicijama percentilnih tačaka (npr. da li dvije grupe imaju identičan ili vrlo različit P_{90}). Dosta je bilo priče o principima; vrijeme je da krenemo u akciju i to sa grafičkim prikazivanjem podataka.

Kako možemo grafički prikazati povezanost jedne dimenzione i jedne kategoričke varijable?

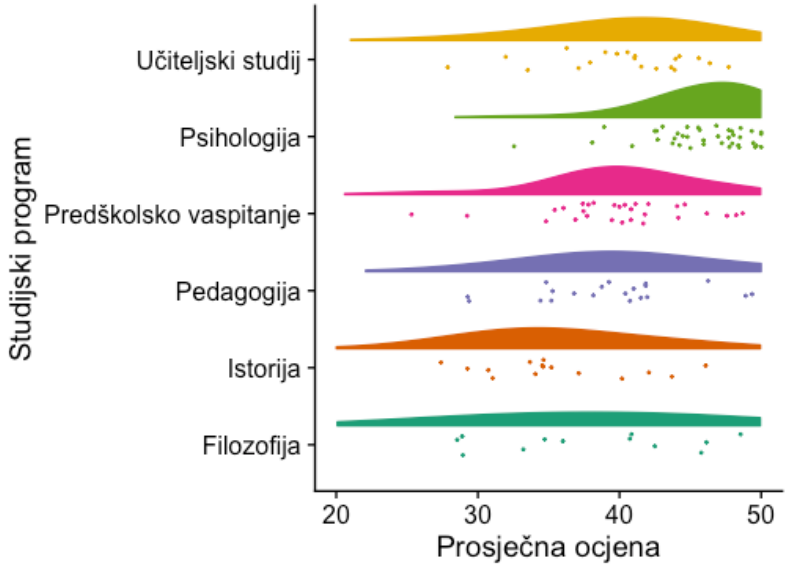
Načini na koji se mogu prikazati ove povezanosti - hm, zapravo htjedoh reći razlike između grupa - predstavljaju proširenje i/ili kombinaciju već prikazanih grafikona kojima smo oslikavali distribuciju pojedinačnih dimenzionih varijabli. Svi takvi ranije opisani grafikoni (npr. histogrami, polirani poligoni frekvencija, tačkasti) se mogu prikazati jedan pored drugog - ili jedan preko drugog u slučaju histograma i poliranih poligona frekvencija. Više pojedinačnih histograma ili poligona proporcija koji stoje zasebno možete da zamislite i bez konkretnog prikaza. Zato je na Slici 8.1 prikazan jedan *preklopljeni* grafikon. Isprekidane linije prikazuju aritmetičke sredine. Šta vam sve grafikon govori?



Slika 8.1: Poređenje polne distribucije varijable Emocionalnost (n=232) putem poliranog grafikona proporcija sa preklapanjem.

Kao bonus je prikazan jedan neprekloplјeni vid grafikona (Slika 8.2) koji se tmurno zove grafikon kišnih oblaka (engl. *raincloud plot*), mada suprotno svom nazivu izgleda najšarenije i

najenergičnije od svih. On na jednom mjestu okuplja polirane grafikone proporcija (oblaci) i tačkasti grafikon (kiša). Vedar grafikon, zar ne? Usput, moguće ga je dopuniti docrtavanjem kutijastih grafikona. Usput, kakve razlike uočavate među studentima različitih studijskih programa?

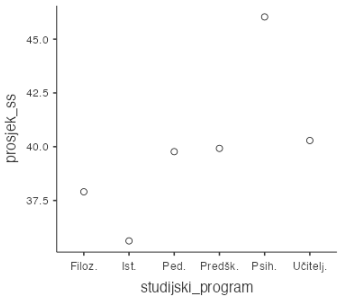


Slika 8.2: Grafikon kišnih oblaka na kojem su upoređene srednjoškolske ocjene brucoša 2022. godine studijskih programa na Filozofskom fakultetu u Banjoj Luci .

Iako grafikoni kišnih oblaka obiluju detaljima, obično im nedostaje znak za aritmetičku sredinu. S druge strane postoje i vrste grafikoni koji su opsjednuti isključivo aritmetičkim sredinama, pa samo njih prikazuju. Dvije verzije takvih grafikona se redovno pojavljuju u naučnim radovima, ali usljed te zaslijepljenosti obje ne daju dovoljan uvid u ostale podatke sa uzorka. Jedna verzija bahato troši previše prostora i boje na prikazivanje mjera prosjeka.[1] Pogledajte Sliku 8.3, grafikon A, gdje visina stubaca označava aritmetičke sredine. (T-linije označavaju nešto čime ćemo se tek baviti kada budemo radili statistiku zaključivanja). Te aritmetičke sredine su mogle proisteći iz bilo kojeg od prikazanih slučajeva B-E - a iza kojih se mogu kriti stršeći podaci, bimodalnosti raspodjela ili vrlo nejednak broj ispitanika. Budući da nam stupčasti grafikon ne otkriva dovoljno uvida u podatke, nećemo ga koristiti za prikazivanje dimenzionih varijabli.

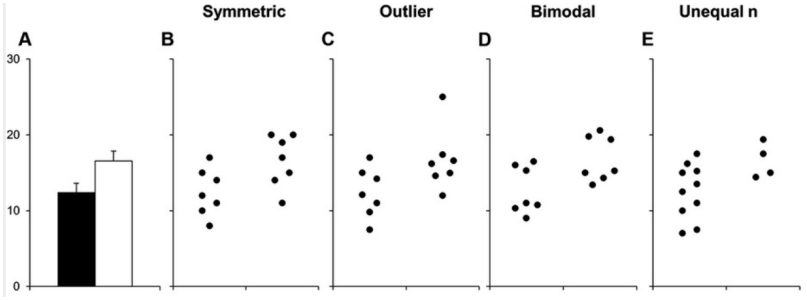
[1] vidjeti detaljnu kritiku u Weissgerber et al., 2015, https://doi.org/10.1371/journal.pbio.1002128

Nažalost, softverska rješenja će i dalje često nuditi još štedljiviju verziju koja koristi samo po jednu tačkicu da prikaže grupu (Slika 8.4).



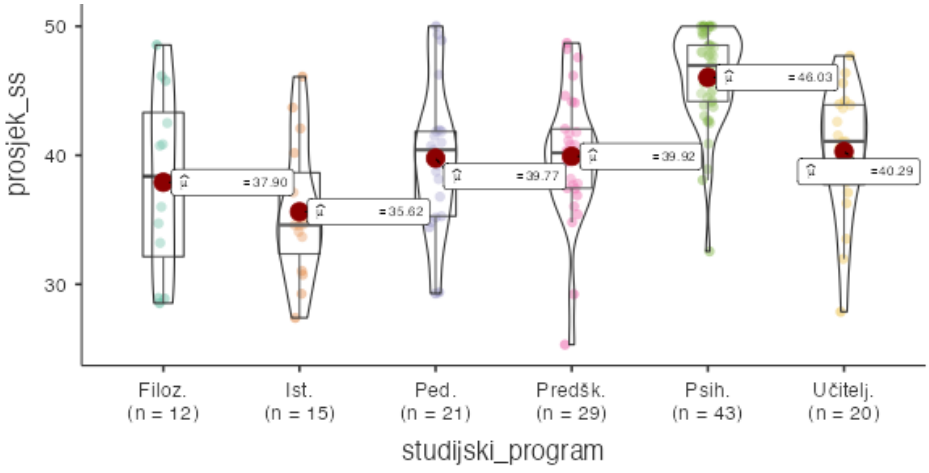
Slika 8.4: Grafikon aritmetičkih sredina za podatke prikazane na Slici 8.2

Srećom, postoje kombinovanih grafikoni koje omogućavaju da



Slika 8.3: Ilustracija problematične upotrebe stupčastog grafikona za prikazivanje aritmetičkih sredina (preuzeto iz članka Weissgerber et al., 2015).

se na jednom mjestu prikaže sve što nam treba: mjere centralne tendencije, varijabilnosti, oblika distribucija, te prisustva pojedinačnih i atipičnih skorova. Sami uporedite Sliku 8.4 gdje su grafikonom predstavljene samo aritmetičke sredine i Sliku 8.5 koja prikazuje podatke istog skupa podataka, ali gdje je grafikon dopunjen pojedinačnim podacima, violinskim i kutijastim grafikonom.



Slika 8.5: Slika 8.5. Obogaćeni grafikon koji uz grupne aritmetičke sredine prikazuje i pojedinačne skorove, violinske i kutijaste grafikone.

Kako da numerički izrazimo vezu između jedne dimenzione i jedne kategoričke varijable?

Iako sam naveo da nam zanimljive mogu biti i razlike u varijabilnosti uzoraka, lokacije percentilnih tačaka, količine odstupanja od normalne distribucije (npr. intenzitet asimetričnosti), ovdje ćemo razmatrati samo najpopularnije mjere razlika između aritmetičkih sredina i odgovarajuće koeficijente korelacije. Zainteresovane za više upućujem na članak u kojem se opisuje 65

različitih mjera razlika koje obuhvataju i razlike u medijanama, varijabilnosti, percentilima, veličine efekta izražene običnim jezikom.[2]

[2] Gyimesi et al., 2025, https://doi.org/10.1177/25152459251323186 sa pratećim onlajn kalkulatorom: https://marton-l-gy.shinyapps.io/StatCompare-Whiz/

Kako se izražavaju razlike između grupa na jednoj dimenzionoj varijabli?

Ubjedljivo najpopularnije mjere razlike se računaju kao razlike aritmetičkih sredina (engl. *difference in means* ili *means differences*). Iako je ranije naglašavano da aritmetičke sredine u mnogim slučajevima nisu najbolje mjere centralne tendencije, jednostavno se medijane i modovi rijetko koriste u direktnom poređenju, mada je naravno to moguće predstaviti. Kada su u pitanju razlike između aritmetičkih sredina dvije grupe, razlikujemo proste razlike i standardizovane razlike.

Kako se izračunava prosta razlika aritmetičkih sredina?

Zar da i to objašnjavam? Kada imamo dihotomnu kategoričku varijablu sa samo dvije kategorije (npr. pol) to je zaista najjednostavnija mjera. Međutim, treba imati u vidu da se stvari komplikuju kada imamo kategoričku varijablu sa šest različita nivoa - kao što je navedeno na primjeru sa brucošima različitih studijskih programa na Filozofskom fakultetu UNIBL. Na primjer, tada imamo petnaest parova[3] koji se mogu porediti.

[3] broj parova poređenja se izračunava na osnovu formule $\frac{n*(n-1)}{2}$)

Da bi označili razlike u populaciji statističari koriste u prefiksu veliko grčko slovo delta (Δ), dok se za statistik može koristiti oznaka latinično
$\bar{D}$. Prednost prostih razlika je to što se lako shvataju kada mjerna skala ima jasno značenje za istraživača i/ili zainteresovanu publiku. Primjera radi, svima je lako shvatljiva razlika u prosječnoj visini od 12cm između muškaraca i žena. Međutim, problemi nastaju ako mjerna skala nema jasno inherentno značenje, što je čest slučaj sa psihološkim mjerama koje nemaju univerzalno utvrđene mjerne jedinice. Uz to, pri procjeni psiholoških karakteristika često koristimo različite mjerne instrumente za procjenu istog konstrukta (npr. za procjenu nečije savjesnosti možemo koristiti različite upitnike crta ličnosti: NEO-PI-R, NEO-FFI, HEXACO-100, HEXACO-60, BFI, BFI-2, BHI, BFQ…). Na primjer, ako žene sebe opisuju kao da su savjesnije od muškaraca za 0.87 poena može predstavljati različit intenzitet razlika ovisno o korištenom instrumentu.

Kako onda računamo standardizovane razlike među aritmetičkim sredinama?

Usljed navedenog problema nepostojanja jasnog referentnog okvira - kao što su osmišljeni i z-skorovi i koeficijenti korelacija - kreirane su standardizovane mjere razlika između aritmetičkih sredina (engl. *standardized mean differences*, *SMDs*). Postoji više takvih mjera kojima je zajedničko to da se prosta razlika aritmetičkih sredina dijeli standardnom devijacijom. Drugim riječima, za sve te mjere je zajedničko da je jedinica razlike jednaka standardnoj devijaciji (vidite li sličnost sa z-skorovima?). Različite varijante standardizovanih razlika koriste različite standardne devijacije. Od svih varijanti najpoznatija mjera se naziva **Cohenovo *d***, a njena formula je:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s^*}$$

Tu s^*označava **objedinjenu** ili **združenu standardnu devijaciju** (engl. *pooled standard deviation*) koja se dobija putem sljedeće formule:

$$s^* = \frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}$$

Razlog za postojanje objedinjene standardne devijacije je to što poduzorci mogu biti imati različit broj ispitanika, a potrebno je da poduzorci bude proporcionalno zastupljeni pri izračunavanju te zajedničke standardne devijacije. U svakom slučaju, na uzorku se računa statistik d koji služi da se procijeni stvarni parametar, populaciona razlika, koja se označava malim grčkim slovom delta δ. Formula za populacionu vrijednost je onda:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma^*}$$

Tokom godina se pokazalo da za vrlo male uzorke ($n <= 20$) postoji nepristrasnija procjena parametra koja se naziva Hedgesovo g (ili d^*). Za njeno izračunavanje je potrebno d pomnožiti sa korektivnim faktorom, koji se najčešće definiše ovako:

$$d^* = d * (1 - \frac{3}{4(n_1 + n_2) - 9})$$

Kada se računa i Cohenovo d i Hedgesovo d^* pretpostavlja se da su varijabilnosti obje grupe podjednake. Kada je ova pretpostavka prekršena rezultati se teže interpretiraju jer se postavlja pitanje koja od dvije standardne devijacije bi bila referentna. S tim u vezi, u literaturi se pominje i varijanta poznata pod nazivom Glassovo d (d_Δ), i to naročito u kontekstu eksperimentalnih istraživanja. Tada se razlika među aritmetičkim sredinama dijeli standardnom devijacijom samo jedne od grupa - obično standardnom devijacijom kontrolne grupe (engl. *control*, (C)). Razlog za to je da standardna devijacija kontrolne grupe predstavlja "prirodnu" varijabilnost pojave, pa će dobijena vrijednost pokazivati koliko je tretman djelotvoran u odnosu na tu neutralnu varijabilnost.[4] Formula je sljedeća:

$$d_\Delta = \frac{\bar{x}_T - \bar{x}_C}{s_C}$$

Standardizovane razlike između aritmetičkih sredina imaju veoma široku upotrebu, koja jednim dijelom dolazi i iz zalaganja Cohena da za te mjere uspostavi neku opisnu klasifikaciju. Tako se na osnovu sljedećih graničnih skorova razlike mogu svrstati u:

- <.20 zanemarive, trivijalne
- .20 male
- .50 umjerene
- .80 velike

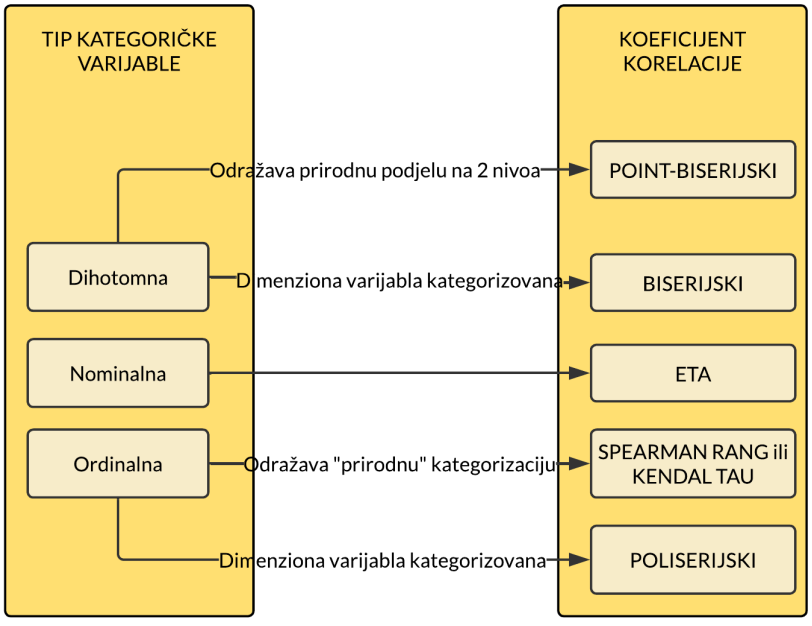
Kao i u pogledu interpretacije koeficijenata korelacija, ova klasifikacija je zapravo nesretni Cohenov proizvod. Naime, i sam Cohen je preporučivao da se ona koristi isključivo kao dopunsko sredstvo pri opisivanju veličine razlika i to samo u situacijama kada ne postoje drugi smisleniji načini (npr. teorijska očekivanja, ranija empirijska saznanja, praktične konotacije koja razlika ima). Iako sa striktno statističkog gledišta standardizovane razlike rješavaju proizvoljnost mjernih jedinica, one su daleko od toga da predstavljaju intuitivni opis snage razlika, pa bi zato uvijek trebalo prikazati ih uz saopštavanje prostih razlika izraženih na izvornoj mjernoj skali. Kako bi se stekla određena perspektiva o značenju d vrijednosti, vratićemo se na primjer polnih razlika u visini; npr. demografski podaci iz SAD-a pokazuju da razlike u prosječnoj visini između muškaraca i žena iznose 12cm, a to je kad se prevede u standardizovane vrijednosti $1.40d$. Kao kontrast možemo dati primjer procjene parametra polnih razlika u samoprocijenjenoj savjesnosti koja iznosi 0.20_d u korist žena.

[4] Zainteresovane za dileme u pogledu standardizovanih veličina efekata upućujem na onlajn knjigu: Guide to Effect Sizes and Confidence Intervals https://matthewbjane.quarto.pub/

A kako onda izražavamo povezanost između kategoričkih i dimenzionih varijabli?

Dosad smo govorili o razlikama između grupa, a naslovom poglavlja smo zapravo najavili da ćemo govoriti o povezanostima između kategoričkih i dimenzionih varijabli. Ponoviću, to je samo druga strana istog lica i u značajnom broju slučajeva će biti dovoljno da povezanost izrazite razlikama (baš tako kao što ste pročitali). Međutim, postoji i mnoštvo specijalizovanih koeficijenata korelacija, a koji od njih će se koristiti uglavnom ovisi o vrsti kategoričke varijable.

Slika 8.6 opisuje izbor odgovarajućih koeficijenata korelacije koji se najčešće javljaju u naučnoj literaturi. Nećemo ovdje ulaziti u detalje njihovog izračunavanja. Samo ću vam potvrditi da je za neke među njima proces računanja isti kao za one koje ste već upoznali. Konkretno, **point-biserijski koeficijent** r_{pb} (engl. *point-biserial coefficient*) je zapravo isto što i Pearsonov linearni koeficijent kada je jedna od varijabli dihotomna sa vrijednostima 0 i 1. Spearmanov rang-koeficijent smo već obrađivali u prethodnom poglavlju i računanje je identično, s tim da obje varijable moraju biti rangirane, odnosno potrebno je da se i za ordinalne kategorije dodijele brojčani rangovi. Nadalje, kada dimenziona varijabla prikazuje rangove, onda imamo i rangovanu point-biserijsku korelaciju.



Slika 8.6: Dijagram odabira koeficijenata korelacija u zavisnosti od tipa kategoričke varijable.

Kada je u pitanju snaga povezanosti, svi ovi koeficijenti su ograničeni mjernom skalom od 0 do 1. Eta (η) koeficijent je neusmjeren, a ostali mogu imati pozitivnu ili negativnu vrijednost. Međutim, da bismo smjer povezanosti mogli ispravno protumačiti potrebno je da znamo kako su varijable kodirane (npr. da li su kategorije rangovane tako da brojke idu od one koja posjeduje najmanje do najviše atributa od interesa ili je korišteno obrnuto kodiranje). Kako za eta koeficijent nije uslov da postoji monotona veza[5] između dvije varijable onda se njena kvadrirana varijanta koristi baš kao koeficijent determinacije u slučajevima kada imamo više nivoa kategoričke varijable i jednu dimenzionu varijablu. Dakle, η^2 nam govori o procentu preklopljenog varijabiliteta između kategoričke i dimenzione varijable. Što se tiče nekih drugih koeficijenata (biserijski, poliserijski i Kendallov τ koeficijent) ovdje ih samo pominjem imenom, a mnogo drugih nije našlo mjesto u datoj tabeli.

[5] Kada su u pitanju ordinalne kategoričke varijable, tada pretpostavljamo monotonost odnosa. Usput, ako se vratimo na primjer iz prethodnog poglavlja dat na Slici 7.3a - gdje je prikazana obrnuta U-korelacija - i podijelimo X-osu na 7 podjednakih dijelova (od -3 do +3), onda je $\eta = .89$. Takva snaga korelacije otprilike odgovara onome što bismo dobili kada bismo krivu liniju ispravili, a da tačkice ostanu na istoj razdaljini od linije. Ako je korelacija striktno linearna, onda će eta biti jednaka apsolutnoj vrijednosti linearnog koeficijenta korelacije

Kako i kada tabelarno i tekstualno prikazujemo odnos kategoričke i dimenzione varijable?

Kada imamo posla sa dihotomnom varijablom onda se podaci najčešće prikazuju u tekstu, unutar jedne ili dvije rečenice. Međutim, tabele su zgodnije kada moramo prikazati kategoričke varijable sa više kategorija ili porediti više varijabli. Na primjer, Slika 8.7 predstavlja tabelu formatiranu prema APA 7 preporukama[6] gdje su predočene polne razlike na HEXACO crtama ličnosti. Podaci su prikupljeni u sklopu jednog našeg istraživanja (Lakić i Perić, 2025), a rezultati za crtu Emocionalnost su već bili prikazani na Slici 8.1.

[6] U jednom naučnom radu bi ovakva tabela imala i mjere statistike značajnosti, ali mi do njih još nismo došli. Primijetite da u Napomenama (engl. *Note*) nisu data pojašnjenja za skraćenice *M*, *SD* i *n*, s obzirom na to da se one smatraju opštepoznatim.

Table xxx

Descriptive Sex Differences in HEXACO Traits using the Brief HEXACO Inventory

Trait	Sex	*n*	*M*	*SD*	*D*	*d*	r_{pb}
Honesty-Humility	Female	112	2.50	0.85	0.24	0.27	0.13
	Male	120	2.26	0.91			
Emotionality	Female	112	1.90	0.83	0.75	0.94	0.43
	Male	120	1.15	0.77			
Extraversion	Female	112	2.76	0.73	-0.21	-0.28	-0.14
	Male	120	2.97	0.74			
Agreeableness	Female	112	2.23	0.65	-0.01	-0.01	-0.01
	Male	120	2.24	0.61			
Conscientiousness	Female	112	2.46	0.71	0.11	0.15	0.08
	Male	120	2.35	0.71			
Openness	Female	112	2.32	0.72	0.22	0.31	0.15
	Male	120	2.10	0.72			

Note. *D* = difference between means; *d* = Cohen's d; r_{pb} = point-biserial correlation where females were coded 1 and males 0. Negative values indicate higher scores for males.

Slika 8.7: APA 7 formatirana tabela koja prikazuje polne razlike na HEXACO crtama ličnosti (n = 232).

Poglavlje 9

Analiza povezanosti dvije kategoričke varijable

Nakon čitanja ovog poglavlja znaćete:

- kako se tabelarno predstavljaju podaci kojima se dvije kategoričke varijable dovode u vezu,
- vidove grafikona i statističke mjere kojima se efikasno predstavlja povezanost dvije kategoričke varijable.

Već smo pominjali da kategoričke varijable koje koristimo u istraživanjima ne samo da mogu biti prirodno kategoričke nego u velikom broju slučajeva kao istraživači „prisilno" kategorišemo dimenzione varijable kako bismo izraženost predmeta mjerenja predstavili na razumljiviji način. Iz toga slijedi da broj kategoričkih varijabli - teorijski gledano - nadilazi broj dimenzionih varijabli, budući da dimenzione možemo kategorisati, a obrnuto ne ide. Ne samo to, treba naglasiti da je i naš saznajni aparat dosta opremljeniji za baratanje kategorijama nego dimenzijama. Vjerujem da je i vama lakše da sebi predstavite ideju da vrlo savjesne osobe (kategorija) češće doživljavaju duboku starost (npr. preko 85 godina; kategorija) nego nesavjesne osobe, u odnosu na to da zamislite „dimenzionu" ideju da postoji tendencija da što su ljudi savjesniji to duže žive. Prosti razlog za to je što možete lakše konkretizovati mentalnu sliku tipične vrlo savjesne osobe (npr. uvijek počešljana osoba sa krajnje rutinizovanim danom koji podrazumijeva ustajanje u isto vrijeme, čelično vođenje lične higijene, preciznost i tačnost u poslu, predan i bez straha odlazak na zdravstvene kontrolne preglede), kao i one nesavjesne (tj. neumjereni korisnik psihoaktivnih supstanci, svaka stvar u sobi može biti na bilo kojem mjestu, izvršavanje obaveza smatra grijehom), nego što možete da napravite konkretizovanu mentalnu sliku finog iznizane savjesnosti na jednoj dimenziji. Eto, i to je

razlog zašto je bitno ovladati efikasnim načinima prezentovanja odnosa između kategoričkih varijabli, čak i ako kategorije nisu uvijek najvjernija slika stvarnosti.

Kao i u prethodnom poglavlju, mada u naslovu imamo povezanost, dominantno ćemo biti usmjereni na otkrivanje razlika među grupama. S druge strane, korelacije među kategoričkim varijablama imaju veći značaj jer se dosta koriste u psihometriji, pa će o njima biti više riječi. Konačno, za razliku od prethodnog poglavlja, ovdje počinjemo sa tabelama, jer one imaju drugačiji format od onoga što smo dosad viđali.

Šta su to ukrštene tabele i čemu one služe?

Tabele kontingencije (engl. *contingency tables*) ili, još bolje, **ukrštene tabele** (engl. *cross tabulations* ili *crosstabs*) su osnovno sredstvo predstavljanja odnosa kategoričkih varijabli. Takvim tabelama istovremeno predstavljamo raspodjelu učestalosti pojedinačnih kategorija za dvije, pa čak i više varijabli (time se nećemo baviti ovdje). Raspodjela X-varijable se prikazuje u redovima, dok se Y-varijabla prikazuje u kolonama. Svaka ukrštena tabela dobija svoje dopunsko ime (nadimak?) na osnovu broja kategorija koje varijable imaju. Tako će tabela 2x2 označavati da se radi o odnosu dvije dihotomne varijable, dok nam tabela 5x4 govori da se prikazuje odnos jedne varijable koja ima pet i druge varijable koja ima četiri kategorije.

Pogledajmo zajedno prikaz jedne 2x5 ukrštene tabele koju smo dobili u jednom našem malom istraživanju ($n = 142$) (Slika 9.1). Tabele prikazuju vezu između varijable pol (kategorije: ženski i muški ispitanici) i ordinalne kategoričke varijable koja predstavlja jednu stavku upitnika ekstraverzije (tvrdnja "Lako stupam u kontakt sa nepoznatima." zajedno sa petostepenim odgovorima od "Nimalo" do "U potpunosti"). U različitim **ćelijama** (engl. *cells*) ili **poljima** unutar tabele su prikazani različiti podaci koji predstavljaju apsolutne i relativne frekvencije, pa je potrebno da ih pobliže pojasnim.

pol x Lako stupam u kontakt sa nepoznatima

			Lako stupam u kontakt sa nepoznatima.					
			0 Nimalo	1 U maloj mjeri	2 Umjereno	3 Uglavnom	4 U potpunosti	Ukupno
pol	ženski	n	7	7	14	24	10	62
		% unutar polne kat.	11.3%	11.3%	22.6%	38.7%	16.1%	100.0%
		% od ukupnog uzorka	4.9%	4.9%	9.9%	16.9%	7.0%	43.7%
	muški	n	3	10	23	31	13	80
		% unutar polne kat.	3.8%	12.5%	28.8%	38.8%	16.3%	100.0%
		% od ukupnog uzorka	2.1%	7.0%	16.2%	21.8%	9.2%	56.3%
Ukupno		n	10	17	37	55	23	142
		% od ukupnog uzorka	7.0%	12.0%	26.1%	38.7%	16.2%	100.0%

združeni procenti
marginalni procenti
uslovni procenti

Slika 9.1: Ukrštena tabela koja prikazuje distribuciju odgovora na tvrdnju "Lako stupam u kontakt sa nepoznatima" u zavisnosti od pola ispitanika.

Koje vrste frekvencija imamo unutar ukrštenih tabela?

Apsolutne frekvencije (neobojene ćelije) su jednostavno broj ispitanika koji pripada određenoj kategoriji ili kombinaciji kategorija X i Y varijable. Na primjer, bilo je ukupno 23 ispitanika koji su se u potpunosti saglasili sa navedenom tvrdnjom. Od tog broja 10 je bilo ženskog, a 13 muškog pola. Da li na osnovu te informacije, kao i podataka iz prethodne kolone gdje se vidi da je veći broj ispitanika muškog pola dao odgovor "Uglavnom" možemo da zaključimo da su muškarci skloniji ekstraverziji u ovom aspektu? Sa odgovorom moramo pričekati budući da primjećujemo da je ukupno gledano muškaraca bilo više u čitavom uzorku.

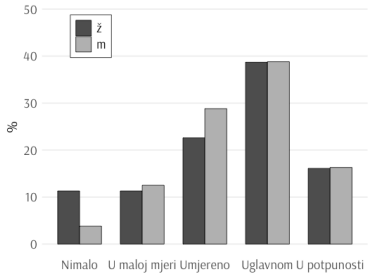
Marginalne/ivične/rubne (engl. *marginal*) relativne frekvencije nam mogu pomoći da utvrdimo omjer na standardizovanoj skali, budući da brojevi 62 i 80 (koliko je bilo ženskih i muških ispitanika) nisu najintuitivniji brojevi. Marginalne proporcije dobijamo kada podijelimo zbir apsolutnih frekvencija za određenu kategoriju sa ukupnim brojem ispitanika, a ukoliko želimo da dobijemo procente onda taj količnik množimo sa 100. Tako da za polnu raspodjelu dobijamo rezultat: $\% = \frac{62}{142} * 100 = 43.7\%$ odnosno $\%_m = \frac{80}{142} * 100 = 56.3\%$.

Združene (engl. *joint*) relativne frekvencije dobijamo kada apsolutne frekvencije za datu kombinaciju kategorija X i Y varijable podijelimo sa ukupnom veličinom uzorka (i naravno, pomnožimo taj rezultat sa 100 ukoliko želimo da podatak izrazimo kao postotak). Kada je u pitanju kategorija ispitanika koji su odgovorili „U potpunosti" vidimo da su ženski ispi-

tanici činili $\frac{10}{142} * 100\% = 7.0\%$ ukupnog uzorka, a muški $\frac{13}{142} * 100\% = 9.2\%$. Međutim, budući da smo ustanovili da je muškaraca ukupno više na uzorku to znači da nam ova informacija ne pomaže da vidimo da li postoje polne razlike.

Uslovne (engl. *conditional*) relativne frekvencije su ono što nam treba da dobijemo odgovor na pitanje da li pol stoji u vezi sa samoprocijenjenom lakoćom stupanja u kontakt sa nepoznatima. Uslovne frekvencije dobijamo tako što broj ispitanika unutar određene kombinacije varijabli X i Y dijelimo sa ukupnim brojem ispitanika za datu kategoriju X ili Y varijable. Drugim riječima, možemo dobiti dvije vrste uslovnih proporcija: unutar varijable X ili unutar varijable Y, a na Slici 9.1 je prikazana samo prva varijanta jer je ona potrebna da bismo došli do odgovora na na postavljeno istraživačko pitanje.[1] Najlakše ćemo shvatiti ovo kroz primjer. Ako sada uporedimo odgovore "u potpunosti" unutar različitih poduzoraka vidjećemo da su oni praktično jednaki: za žene je to $\frac{10}{62} * 100\% = 16.1\%$, dok je za muške $\frac{13}{80} * 100\% = 16.3\%$). Za odgovor "uglavnom" dobijeni su još sličniji rezultati s obzirom na to da je tako odgovorilo 38.7% ženskih, a 38.8% muških ispitanika. Zaključak je da su nas apsolutne frekvencije mogle zavarati. Međutim, istina je i to da postoje blage razlike među polovima kada se pogledaju odgovori koji ukazuju na introvernost sa nešto više takvih odgovora muških ispitanika. Vidite li to?

[1] Isto tako je moguće predstaviti uslovne relativne frekvencije po varijabli Y na sljedeći način. Od ukupnog broja ispitanika koji su odgovorili sa „nimalo" njih 7, odnosno 70% su djevojke, a 3 (30%) mladići. Međutim, u ovom primjeru bi nam takva kalkulacija dala pogrešan odgovor na naše pitanje o tome da li postoje polne razlike u odgovorima na tvrdnju „Lako stupam u kontakt sa nepoznatima." Ova analiza bi precijenila razlike. Nas zapravo interesuje, postoje li razlike kada bismo u uzorku imali jednak broj ispitanika muškog i ženskog pola - jer tek tada se polovi mogu porediti! Sami izračunajte uslovne postotke unutar odgovora „u potpunosti" i uporedite rezultate sa onima prikazanim u tabeli.



Slika 9.2: Distribucija odgovora na tvrdnju "Lako stupam u kontakt sa nepoznatima" u zavisnosti od pola ispitanika prikazana grupnim stupčastim grafikonom.

Mogu li se povezanosti dvije kategoričke varijable prikazati i grafički?

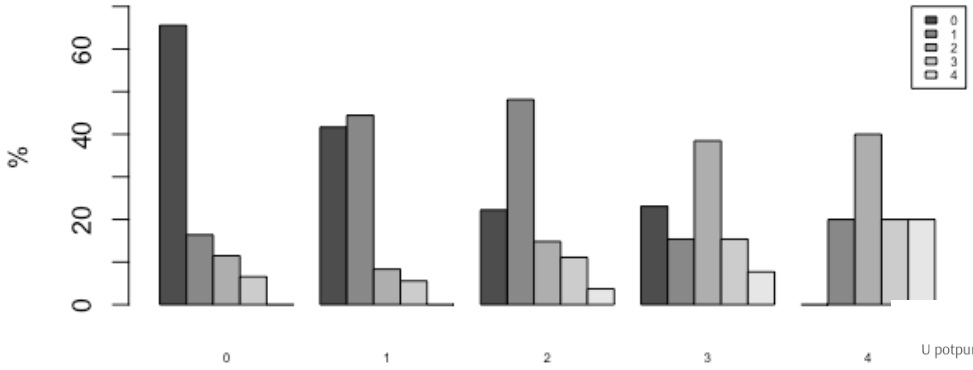
Naravno. I u slučajevima analize odnosa dvije kategoričke varijable grafički prikazi su zapravo kognitivno efikasniji metod. U tu svrhu se mogu preporučiti sljedeće vrste grafikona:

- grupni stupčasti grafikon
- grupni Clevelandov tačkasti grafikon
- toplotna mapa

Grupni stupčasti grafikon (engl. *grouped bar plot*) je stupčasti grafikon gdje se za kategoriju X-varijable docrtava po jedan stubić za svaku kategoriju Y varijable. Obično su ti stubići označeni drugačijom bojom, a opis kodnog sistema o tome koja boja prikazuje koju kategoriju se daje u legendi koja je pozicionirana negdje na grafikonu. Kao i u slučaju običnih stupčastih grafikona, visina grafikona označava frekvenciju. S tim u vezi, grupnim stupčastim grafikonom moramo prikazivati uslovne procente / propor-

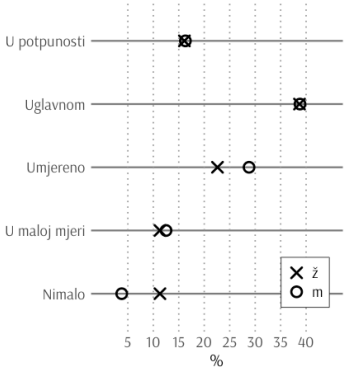
cije ukoliko želimo da dobijemo stvarnu sliku povezanosti. Na Slici 9.2. vidimo raspodjelu podataka koje smo ranije prikazali u tabeli.

Međutim, treba imati na umu da je grupni stupčasti grafikon dobar način prikazivanje odnosa kategoričkih varijabla samo onda kada barem jedna od varijabli ima mali broj kategorija (npr. četiri ili manje), a ni druga nema prevelik broj kategorija. Kada obje varijable imaju veći broj kategorija (npr. > 4) teško da će nam grupni stupčasti grafikon biti efikasno sredstvom prikaza povezanosti. Pogledajmo samo Sliku 9.3 koja prikazuje distribuciju odgovora na dvije stavke: "Lako stupam u odnos sa nepoznatima" i "Niko ne voli da razgovara sa mnom".



Slika 9.3: Grupni stupčasti grafikon koji prikazuje uslovne procente odgovora na tvrdnje "Lako stupam u kontakt sa nepoznatima" i "Niko ne voli da razgovara sa mnom".

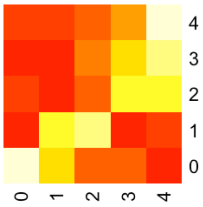
U slučajevima kada jedna od varijabli ima veliki broj kategorija, a druga vrlo mali broj (npr. do tri kategorije) **grupni Clevelandov tačkasti grafikon** (engl. *grouped Cleveland dot plot*) predstavlja štedljiviju alternativu koja može da prikaže iste podatke kao i grupni stupčasti grafikon (Slika 9.4). Na Y-osi se prezentuje lista kategorija varijable sa većim brojem kategorija, dok se različite kategorije druge varijable mogu prikazati i zasebnim znakovima (npr. X, O, *). I u slučaju ovog grafikona, prikazuju se uslovni procenti / proporcije. Iako efikasan grafikon, on prestaje to da bude ako obje varijable imaju veći broj kategorija (počevši od četiri, pa naviše) kada može da prouzrokuje popriličnu konfuziju.



Slika 9.4: Distribucija odgovora na tvrdnju "Lako stupam u kontakt sa nepoznatima" u zavisnosti od pola ispitanika prikazana grupnim Clevelendovim grupnim tačkastim grafikonom.

U takvim slučajevima, rješenje može da bude **toplotna mapa** (engl. *heat map*). Radi se o tipu grafikona koji zapravo direktno preslikava ukrštene tabele i umjesto brojeve koristi boje. Gušće „naseljene" ćelije će biti obojene sve zasićenijim bojama, dok sve rjeđe „naseljene" ćelije sve više blijede.

Slika 9.5 nam pruža neposredniji uvid u očekivanu negativnu



Slika 9.5: Toplotna mapa koja prikazuje frekvencije odgovora na tvrdnje "Lako stupam u kontakt sa nepoznatima" (X-osa) i "Niko ne voli da razgovara sa mnom" (Y-osa).

povezanost ranije navedenih stavki koje "gađaju" ekstravertnost. Naime, toplotna mapa efikasnije prenosi informaciju da su oni koji daju više odgovore na varijabli X, skloniji da da daju niže odgovore na varijabli Y, i obrnuto. Što je ćelija crvenija, ona je i više popunjena.

Osnovna mana toplotne mape je što se njome teže mogu procijeniti precizne mjere čije se vrijednosti ipak mogu bolje procijeniti korištenjem ranije navedenih vrsta grafikona. Međutim, postoji i djelimični lijek za to: u toplotnu mapu se mogu direktno upisati podaci iz tabele, pa onda imamo već pominjanu toplotnu tabelu. Djelimični lijek jer istraživač mora da izabere da li da prikaže uslovne frekvencije ili apsolutne frekvencije.

Na koji način se numerički izražava povezanost dvije kategoričke varijable?

Kao i slučaju povezanosti jedne dimenzione i jedne kategoričke varijable izbor koeficijenta korelacije ovisi o nivou mjerenja kategoričkih varijabli. Budući da ovdje imamo dvije kategoričke varijable od kojih svaka može biti dihotomna, ordinalna ili nominalna to znači da već postoji šest različitih kombinacija. Pritom, za kombinaciju istih varijabli (npr. obje varijable dihotomne) postoji čak i veliki broj različitih koeficijenata korelacija od kojih svaki ima svoje razloge postojanja. S druge strane, neki koeficijenti su podobni za više kombinacija.

Zašto je bitno poznavati toliko koeficijenata? Najviše zbog toga što je izbor odgovarajućeg koeficijenta bitan za psihometrijske postupke. Već smo pominjali da je većina psiholoških instrumenata, a pogotovo testova sposobnosti i znanja, kvazi-intervalne prirode i da predstavlja kompozitne instrumente. U procesu konstrukcije instrumenata, istraživači se redovno oslanjaju na statističke kriterije. Konkretno, kako bi odredili da li određena stavka adekvatno procjenjuje željeni predmet mjerenja, često koriste različite oblike posebnog statističkog postupka, faktorske analize. Faktorska analiza kao ulazne podatke najčešće uzima matrice korelacija, a vraća vam nove, latentne varijable koje su direktan proizvod korelacija među stavkama. Ako u takvu analizu unesete različite ulazne podatke (različite koeficijente korelacija) onda ćete dobiti i različite rezultate faktorske analize. Da zaključimo: sudovi o adekvatnosti ukupnog instrumenta ili pojedinačnih stavki mogu da ovise o izboru koeficijenta korelacije.

Ključni izvor razlika u rezultatima koji se dobijaju putem različi-

tih vrsta korelacija su raspodjele kategoričkih varijabli. Pod tim se podrazumijeva tzv. težina stavke (npr. koliko je ispitanika uspješno riješilo taj zadatak na testu), te simetričnost raspodjele odgovora što stoji i u vezi sa težinom stavke. Efekte koje raspodjela može da ima prikazaću koristeći najjednostavniju analizu, analizu povezanosti dvije dihotomne varijable. To je skoro uvijek glavni predmet analize pri konstrukciji testova sposobnosti i znanja kada imamo tačne i netačne odgovore.

Za primjer ćemo uzeti dva zadatka sa kolokvija iz statistike u psihologiji koji se odnose na izračunavanje medijane i standardne devijacije. Očito, oba zadatka bi trebalo da govore o znanju iz statistike, a time podrazumijevamo i da postoji pozitivna povezanost u uspješnosti rješavanja ovih zadataka. Drugim riječima, ako je neki student tačno izračunao medijanu, očekujemo da će postojati veća vjerovatnoća i da je tačno izračunao standardnu devijaciju. Pritom, kao što i sami možete da posvjedočite, lakše je izračunati medijanu nego standardnu devijaciju (vratite se na ranija poglavlja ako dosad niste naučili kako da ih izračunate!) . Izračunavanje standardne devijacije jeste kompleksniji posao za koji morate najprije tačno izračunati aritmetičku sredinu, ali onda pažljivo obaviti i ostale korake (oduzimanje pojedinačnih skorova od aritmetičke sredine, njihovo kvadriranje, sabiranje, dijeljenje sa n-1, te korijenovanje).

Pogledajmo sada 2x2 ukrštenu tabelu (Slika 9.6) koja prikazuju rezultate na ova dva pitanja na testu znanja kojeg je polagalo ukupno 90 studenata prve godine psihologije. Obratite pažnju na marginalne ćelije koje govore o težini svakog od zadataka, ali i na različito obojene ćelije koje govore o **saglasnim** i **nesaglasnim** odgovorima. Saglasnost se ovdje odnosi na jednakost odgovora na oba zadatka (oba tačna ili oba netačna), dok se nesaglasnim odgovorima smatra tačan odgovor na jedan zadatak i netačan na drugi.

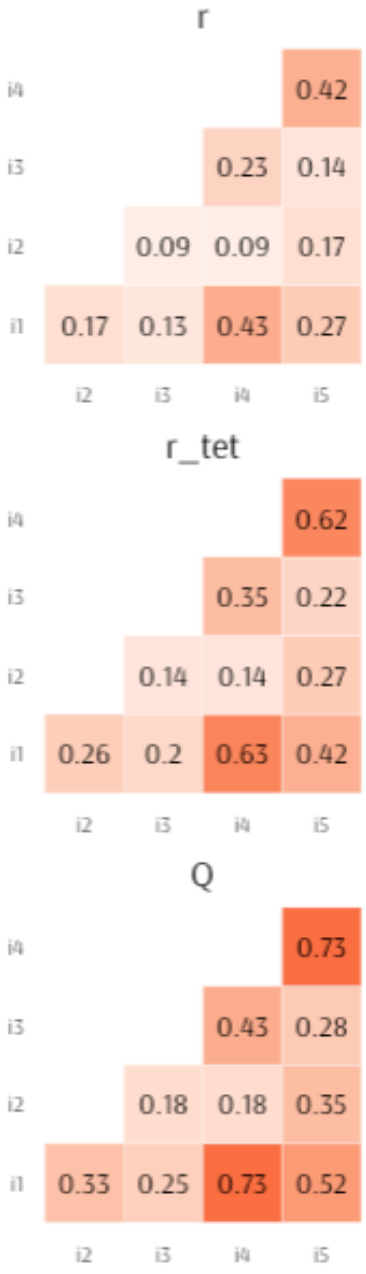
		Zadatak Mdn		ukupno
		+	-	
Zadatak	+	**30** (a)	**0** (b)	30
SD	-	**30** (c)	**30** (d)	60
ukupno		60	30	90

Slika 9.6: Kombinovana distribucija odgovara na dva zadatka.

Kao što se i očekivalo, izračunati medijanu (60 od 90, odnosno 66.7% tačnih odgovora) je bilo lakše nego izračunati standardnu devijaciju (30 od 90, 33.3% tačnih).[2] Međutim, potrebno je da ustanovimo da li su ove dvije varijable međusobno povezane, a to ne možemo samo na osnovu težine stavki.

Možete li da zamislite kako bi izgledala savršena povezanost? Tada bi svi podaci bili u zeleno obilježenim ćelijama, odnosno imali bi saglasne odgovore. To bi značilo da su svi ispitanici koji su tačno izračunali medijanu, tačno izračunali i standardnu devijaciju - te obratno, ukoliko nisu uspjeli izračunati medijanu da nisu uspjeli izračunati ni standardnu devijaciju.

[2] Geografski primjer može biti lakši za razumijevanje. Ako ispitujemo poznavanje glavnih gradova pretpostavljamo da više ljudi zna glavni grad Belgije nego glavni grad Mozambika. Pritom,vjerovatno će svi oni koji znaju glavni grad Mozambika znati i glavni grad Belgije.



Slika 9.7: Razlike u visini procijenjene korelacije među stavkama ovisno o odabranom koeficijentu.

Ali budući da znamo da je jedan od zadataka bio teži, logično je očekivati da imamo i neke nesaglasne podatke (crvene ćelije). Odnosno, logično je očekivati da postoje neki studenti koji su mogli tačno izračunati medijanu, ali nisu uspjeli tačno izračunati i standardnu devijaciju. Iz ovoga slijedi pitanje da li je u redu da smatramo da je povezanost savršena ako na osnovu tačnog odgovora na standardnu devijaciju možemo sa 100% tačnošću predvidjeti i da je ta osoba izračunala tačno i medijanu?

Odgovor je: Zavisi koji koeficijent korelacije pitate! Da je korelacija savršena smatraće Yuleov Q koeficijent korelacije ($Q = 1.00$). Yuleov Q koeficijent uzima u obzir hijerarhijsku uređenost varijabli, odnosno eksplicitno uzima u obzir postojeće razlike u težini stavki pri procjeni snage povezanost, te mu simetričnost odnosa nije bitna. Za razliku od njega, ϕ (fi-koeficijent) - koji koristi Pearsonovu formulu linearne korelacije samo sa vrijednostima 1 i 0 za kategorije *tačno* i *netačno* - će označiti ovu korelaciju kao vidno pozitivnu, ali daleko od toga da će je procijeniti kao savršenu. Razlog za dosta niži koeficijent ($\phi = .50$) je to što je osoba koja je tačno izračunala medijanu imala samo 50% šanse da je tačno izračunala i standardnu devijaciju. Fi-koeficijent je osjetljiv na težinu stavki i "kažnjava" asimetričnost povezanosti.

Formule i rješenja za Q i ϕ koeficijente su data ispod kako biste vidjeli da je to jednostavna matematika. Za ϕ koeficijent ovdje navodim formulu koja je mnogo lakša za izračunavanje "pješke".

$$Q = \frac{ad - bc}{ad + bc} = \frac{30 * 30 - 30 * 0}{30 * 30 + 30 * 0} = \frac{900}{900} = 1$$

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

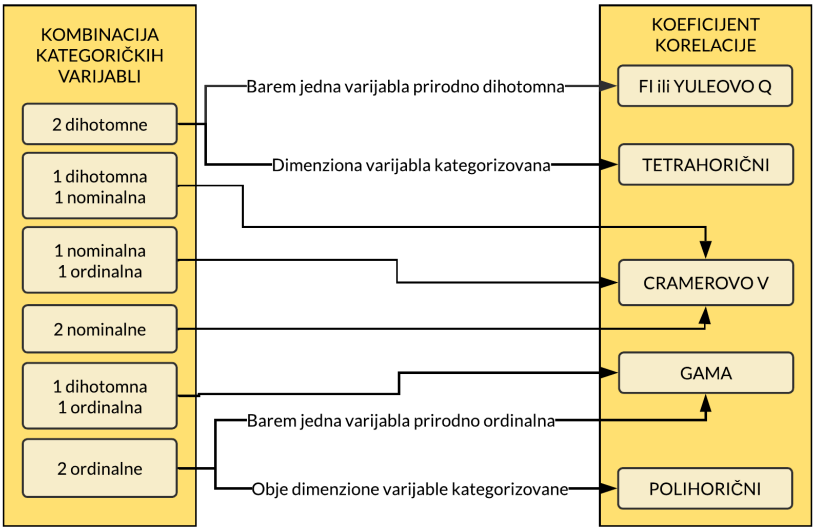
$$\phi = \frac{30 * 30 - 30 * 0}{\sqrt{(30+0) * (30+30) * (30+30) * (0+30)}} =$$

$$\frac{900}{\sqrt{30 * 60 * 60 * 30}} = \frac{900}{\sqrt{3240000}} = \frac{900}{1800} = 0.50$$

Međutim Q i ϕ koeficijent su samo dva krajnja primjera, a većina psihometrijske literature bi preporučila da se izračuna treća vrsta koeficijenata: tetrahorični koeficijent korelacije (r_{tet}). Njegove vrijednost su gotovo uvijek između vrijednosti Q i ϕ, ali ono što je nezgodno jeste da postoje barem četiri različita načini izračunavanja tog koeficijenta što daje različite rezultate (za konkretni primjer vrijednosti sežu od .52 do .91, a najčešće upotrebljavani metod daje vrijednost .86?!). Slika 9.7 prikazuje kako se za isti

set podataka (5 zadataka na jednom testu inteligencije) dobijaju različiti rezultati.

Na ovom mjestu neće biti dati ni potpuni spisak svih mogućih koeficijenata korelaciji, ni njihovi aritmetički detalji, niti podrobnija analiza snaga i ograničenja pojedinačnih koeficijenata. Samo za opis povezanosti između dvije dihotomne varijable postoji više od 10 različitih mjera koje su u upotrebi. Dakle, sljedeća klasifikacija (Slika 9.8) prikazuje izvod preporučenih koeficijenata korelacija, ali više o njima ćete vjerovatno saznati kada za to bude potrebe (npr. psihometrija).



Slika 9.8: Pregled koeficijenata korelacija u zavisnosti od nivoa mjerenja kategoričkih varijabli.

Kako se pomoću tabela mogu saopštiti rizici?

Gore smo se fokusirali na to kako da opišemo snagu povezanosti između dvije kategoričke varijable koristeći koeficijente korelacija. Međutim, u mnogim praktičnim situacijama nam je potrebnije da dobijemo neku indikaciju **praktičnog značaja** te povezanosti - odnosno, koliko jedna kategorija povećava ili smanjuje vjerovatnoću pojavljivanja druge kategorije. Ovo je posebno važno u medicini, zdravstvenoj i kliničkoj psihologiji, ali i svakoj drugoj oblasti u kojoj želimo da razumijemo **faktore rizika** koji stoje u vezi sa varijablom od interesa.

Ovdje je važno da razjasnimo terminologiju. U statistici se faktorima rizika smatraju svi činioci koji povećavaju vjerovatnoću nekog negativnog ishoda (npr. pušenje ili konzumacija alkohola

		Prošli ispit		ukupno
		Da	Ne	
Appleti	Da	50 (a)	50 (b)	100
	Ne	25 (c)	75 (d)	100
ukupno		75	125	200

Slika 9.9: 2x2 tabela kontingencije za varijable korištenje appleta i polaganje ispita.

koji povećavaju rizik oboljenja, smrti), ali ovisno o kontekstu istraživanja to mogu biti i činioci koji povećavaju vjerovatnoću uspjeha ili zaštitnih faktora (npr. ishrana sa mnogo povrća, redovna fizička aktivnost). Dakle, faktor rizika je sve ono što potencijalno povećava vjerovatnoću ishoda, bilo da je on povoljan ili nepovoljan.

Kao primjer ćemo koristiti tabelu sličnu onoj sa zadacima iz statistike. Slika 9.9 koristi slične brojeve, ali umjesto uspješnosti rješavanja zadataka imamo X-varijablu koja predstavlja korištenje statističkih appleta za razumijevanje statističkih koncepata (da/ne) i Y-varijablu koja predstavlja uspješnost polaganja ispita (prošao/pao). Studente koji su koristili statističke appleta možemo smatrati tretmanskom ili izloženom grupom, dok su ostali studenti u ovom primjeru kontrolna grupa.

Prva bitna mjera je **relativni rizik (RR)** (engl. *relative risk, risk ratio*) koja nam govori koliko je puta veća učestalost određenog ishoda u jednoj grupi u odnosu na drugu:

$$RR = \frac{P(\text{ishod}|\text{izložena grupa})}{P(\text{ishod}|\text{kontrolna grupa})} = \frac{a/(a+b)}{c/(c+d)}$$

Za našu tabelu će ishod od interesa biti polaganje ispita (vjerujem da je i većini čitaoca), a RR se onda računa na sljedeći način:

$$RR = \frac{50/(50+50)}{25/(25+75)} = \frac{0.5}{0.25} = 2.0$$

Rezultat $RR = 2.0$ nam govori da su studenti koji su koristili applete dvostruko češće polagali ispit.

Međutim, relativni rizik nam govori samo o tome koliko je puta veća ili manja učestalost ishoda u jednoj grupi u odnosu na drugu, ali nam ne govori koliko je ukupno poboljšanje. Za to su nam potrebne mjere apsolutnog i relativnog smanjenja rizika.

Apsolutno smanjenje rizika (engl. *absolute risk reduction*, *ARR*) nam govori za koliko se procenata smanjila učestalost neželjenog ishoda (ili povećala učestalost poželjnog ishoda):

$$ARR = P(\text{kontrolna}) - P(\text{izložena}) = \frac{c}{c+d} - \frac{a}{a+b}$$

U našem primjeru, ako posmatramo neuspjeh na ispitu kao neželjeni ishod:

$$ARR = \frac{75}{25+75} - \frac{50}{50+50} = 0.75 - 0.50 = 0.25 \text{ ili } 25\%$$

To znači da korištenje apleta apsolutno smanjuje rizik pada ispita za 25 procenata, tj. sa 75% na 50%.

Relativno smanjenje rizika (RRR) (engl. *relative risk reduction*) nam govori za koliko procenata se smanjio početni rizik:

$$RRR = \frac{P(\text{kontrolna}) - P(\text{izložena})}{P(\text{kontrolna})} = \frac{ARR}{P(\text{kontrolna})}$$

U našem primjer kalkulacije je:

$$RRR = \frac{0.75 - 0.50}{0.75} = \frac{0.25}{0.75} = 0.33 \text{ ili } 33\%$$

Ovo znači da korištenje apleta smanjuje rizik pada za jednu trećinu u odnosu na početni rizik - relativna snižavanje rizika je 33%.

Zašto su obje mjere važne?

Ove dvije mjere nam daju komplementarne informacije:

- **ARR** nam daje informaciju o tome koliko studenata konkretno profitira od intervencije
- **RRR** nam govori o relativnoj uspješnosti intervencije u odnosu na kontrolnu grupu

Kako se ove dvije mjere mogu razlikovati u praksi je moguće vidjeti iz sljedećih primjera. Zamislite da lijek smanjuje rizik rijetke bolesti ili da psihoterapija smanjuje rizik samoubistva sa 0.002% na 0.001% na 100 000 ljudi. U tom slučaju je:

- $RRR = 50$ (zvuči impresivno)
- $ARR = 0.001$ (zapravo samo 1 od 100.000 ljudi profitira)

Nasuprot tome, ako neki tretman smanjuje rizik sa 60% na 40% na 100 ljudi:

- $RRR = 33$ (zvuči skromnije…)
- $ARR = 20$ (20 od 100 ljudi je izbjeglo negativan ishod)

Ključna pouka je da RRR može da bude ima veliku vrijednost čak i kada je praktični uticaj mali, dok ARR uvijek odražava stvarnu korist na nivou populacije. Iz tog razlog je potrebno da kada pročitate informaciju o sniženju rizika zapitate se da li je saopšten samo relativni oblik ili su dati i apsolutni rezultati.

U ovom poglavlju smo se upoznali sa osnovnim načinima analize povezanosti između kategoričkih varijabli - od jednostavnih ukrštenih tabela do mjera procjene faktora rizika. Međutim, i ovim poglavljem se nismo dotakli istraživačkih situacija u kojima se registruje kako se isti ispitanici ponašaju u različitim uslovima ili kako se njihovi odgovori mijenjaju kroz vrijeme. Fokus sljedećeg poglavlja je upravo na tome - vidjećemo kako je moguće statistički analizirati ponašanja istih ili srodnih ispitanika u različitim vremenskim tačkama ili različitim eksperimentalnim uslovima.

Poglavlje 10

Opisivanje mjera vezanih entiteta

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta su to mjere vezanih entiteta (vezane mjere) i kako se razlikuju od do sada obrađivanih?
- Zašto je potrebno koristiti posebne mjere i grafičko predstavljanje, te o kojim analizama se radi?

Dosad obrađivane analize su podrazumijevale da se za jedan entitet (čitaj zasad: za jednog ispitanika) podatak o varijabli od interesa prikuplja u jednom navratu. Ne morate biti statistički guru da shvatite da to pravilo uvijek važi za nacrte koji imaju za cilj da opišu distribuciju samo jedne varijable. Ako napravimo analogiju sa hvatanjem vizuelnog zapisa, kada analiziramo jednu varijablu mi pravimo jednu statičnu fotografiju. S druge strane, nije tako drečeći uočljivo da statični snimak dobijamo i onda kada na dosad prikazane načine analiziramo da li postoji statistička povezanost dvije varijable. Pritom, nije važno da li izvodimo eksperimentalnu (npr. evaluacija psihoterapijskog tretmana na socijalnu anksioznost) ili neeksperimentalnu studiju (npr. inteligencija i uspjeh na testu znanja); snimak ostaje statičan jer protok vremena ne igra bitnu ulogu u statističkoj analizi.

Da to malo pojasnimo. Dok je vremenski redoslijed varijabli u eksperimentalnim studijama prost, gdje nezavisna prethodi zavisnoj, u neeksperimentalnim studijama imamo tri različite mogućnosti. Najčešći oblik koji se javlja jeste **jednokratni nacrt** ili **studija poprečnog presjeka** (engl. *cross-sectional design*, *transverse study*). U takvoj vrsti istraživanja podaci za sve varijable se prikupljaju istovremeno. Međutim, nije nužno da to bude tako. Na primjer, ako podatak o jednoj varijabli registrujemo danas, a o drugoj za pet godina, i dalje ćemo za analizu koristiti iste statističke postupke kao da smo ih prikupili istovremeno.

Grafikon raspršenja sa koeficijentom korelacije će biti prvi izbor da povežemo podatke sa testa inteligencije (pretpostavljena nezavisna varijabla) na prijemnom ispitu i sa uspjehom na testu znanja koji se zadaje u istom trenutku, i sa prosječnom ocjenom kandidata u prvom razredu srednje škole, kao i sa prosjekom ocjena na trećoj godini studija.

Naravno, vremenski slijed je itekako bitan iz metodoloških razloga, pa i ovi tipovi nacrta imaju drugačije nazive. **Prospektivne (prediktivne) studije** (engl. *prospective studies*) su one u kojima se prediktor (tj. inteligenciju) mjeri prije kriterija (npr. prosjek na fakultetu), a **retrospektivne (retroaktivne)** (engl. *retrospective studies*) one u kojima podaci za kriterij (npr. osnovnoškolska ocjena) potiču iz neke ranije vremenske tačke.

Razlog što pravimo razliku među njima je taj što se za kauzalne hipoteze nešto čvršći dokazi dobijaju pravim prediktivnim istraživanjima. Retroaktivne studije i studije poprečnog presjeka imaju veći broj alternativnih objašnjenja i spoljnih varijabli koje se ne mogu provjeriti i kontrolisati kada su podaci već prikupljeni. Na primjer, ako bismo iz jedne takve studije dobili podatke o povezanosti inteligencije i prosječne ocjene na ranijem nivou školovanja, ostaje razumno vjerovatna hipoteza da skor na testu inteligencije na prijemnom ispitu može biti proizvod intelektualnog rada koji je bio motivisan dobijanim visokim ocjenama. To bi značilo da bi zapravo prosječna ocjena bila - barem djelimični - prediktor inteligencije koja se testira na prijemnom, a ne obrnuto. Kako je pitanje vremenskog redoslijeda pri prikupljanju podataka primarno metodološka tema, njome se nećemo dalje ovdje baviti. Iz ovog uvoda je važno da uočimo da čak i u prospektivnim i retrospektivnim istraživanjima protok vremena ne dobija eksplicitnu ulogu nezavisne varijable.

Šta su to mjere vezanih entiteta?

Tek u **nacrtima sa ponovljenim mjerenjima** (engl. *repeated-measures designs*) se u fokusu istraživanja, kao nezavisna varijabla, može nalaziti protok vremena. Još tačnije bi bilo reći da nezavisnu varijablu predstavlja ono što se dešava između dvije tačke mjerenja. Zamislite da istraživanjem želimo da ustanovimo da li se crte ličnosti mijenjaju tokom studiranja psihologije, odnosno da li tokom godina studija studenti postaju savjesniji, emocionalno stabilniji i sl., te to procjenjujemo svake godine od upisa pa do završetka studija. Ono što pretpostavljamo da utiče na promjene su različiti psiho-socio-biološki procesi koji se dešavaju

Koliko sam mogao to da pretražim, termin "mjere vezanih entiteta" nećete naći u drugoj literaturi. Da se opiše ono što ja u nastavku opisujem, koriste se drugi termini kao što su: unutarsubjektni nacrti (engl. *within-subjects design*), zavisni uzorci (engl. *dependent samples*), sparena / uparena mjerenja (engl. *paired / matched measurements*), korelirane mjere (engl. *correlated measures*), ponovljena mjerenja (engl. *repeated measures*). Po mom mišljenju - očito ne tako skromnom - oni ili suštinski ne opisuju fenomen (npr. ponovljeni, unutarsubjektni) ili stvaraju potencijalnu konfuziju upotrebom termina koji imaju svoje ustaljenije značenje (npr. zavisni, korelirani). Sasvim je moguće da novim terminom doprinosim konfuziji, no...

kroz razvoj i iskustva, a ne činjenica da ste dali uslov i u indeksu imali potvrdu o upisu sljedeće godine. Naravno, ponovljena mjerenja imamo i u kvazi-eksperimentalnim istraživanjima sa pretestom i posttestom (npr. nivoi socijalne anksioznosti prije i nakon ciljane psihološke radionice, sposobnost razumijevanja naučnih radova prije i nakon završetka osnovnog kursa iz statistike).

Ali ponovljena mjerenja nisu samo ona u kojoj imamo vremenski razdvojene tačke mjerenja kao formalnu nezavisnu varijablu. To su sva ona istraživanja u kojima za neki entitet dva ili više puta prikupljamo vrijednost za istu varijablu. Dakle, među takva istraživanja spadaju i ona kojima želimo utvrditi da li određeni uslovi mijenjaju doživljaj i reakciju pojedinca. Na primjer, želimo da utvrdimo da li je podudarna tačnost slušne lokalizacije predmeta kada se koristi lijevo i desno uho (iste glave). Ponovljena mjerenja. Da li se brže prepoznaju latinične ili ćirilične pseudoriječi (ista glava, isti prsti koji pritiskaju tastere)? Ponovljena mjerenja. U takvim istraživanjima se podaci obično prikupljaju unutar jedne eksperimentalne seanse gdje se draži naizmjenično izlažu. Uz to, u sklopu istraživanja se može ponovljeno mjeriti i više varijabli (i nezavisnih i zavisnih) i to više od dva puta.

Kao što već znate nacrte sa ponovljenim mjerenjima koristimo zato što istovjetnost ispitanika predstavlja tehniku kontrole spoljnih varijabli. Na primjer, ako mjerimo sposobnost razumijevanja naučnih tekstova prije i poslije čitanja ove knjige (ZV), pretpostavljamo da su vaše trajne osobine koje bi kao spoljne varijable mogle uticati na razumijevanje (npr. inteligencija, strategije učenja, motivacione sklonosti, crte ličnosti, lična istorija), u velikoj mjeri slične prije i poslije njenog čitanja. Ako ste zaista nakon čitanja knjige (iskreno se nadam) u stanju da bolje razumijete naučnu literaturu onda se to u najvećoj mjeri desilo zahvaljujući čitanju (iz didaktičkih razloga zanemarićemo ostale moguće faktore kao što su: dolasci na predavanja, samostalno gledanje relevantnih video snimaka, čitanje drugih tekstualnih materijala). Da ponovim, vaša svojstva su slična na prvom i kasnijem mjerenju, te se zato može smatrati da se time kontroliše veliki broj spoljnih varijabli.

Da li mjere vezanih entiteta moraju biti ponovljena mjerenja iste osobe?

Ne moraju. Kao što je davno u tekstu navedeno, entitet ne mora da bude pojedinačna osoba. Ja koji sad pišem ovu rečenicu - a vi

koji je sada čitate - jesmo najsličniji sebi u prošlosti i budućnosti. Međutim, slični smo, odnosno imamo iste izvore psihičkih svojstava i dešavanja nekim drugim ljudskim bićima - više nego nekim trećim…Primjeri međusobne sličnosti uključuju bliske odnose poput: djece i roditelja, bračnih ili romantičnih partnera, te braće i sestara (bioloških siblinga). Naime, velika je vjerovatnoća da vi i vaši roditelji, kao i vi i vaša braća ili sestre (ako ih imate), dijelite bitne spoljne varijable. Te varijable ne proističu samo iz genetske prirode, već uključuju i zajednička iskustva. Na primjer, ako bi predmet istraživanja bilo zadovoljstvo porodičnim odnosima, vjerovatno biste imali međusobno sličnije ocjene u okviru porodice nego što biste ih imali u poređenju s kolegama sa studija. Slično tome, u romantične veze obično stupamo na temelju sličnosti u osobinama i vrijednostima (mada postoje i komplementarni razlozi, pa partner predstavlja našu dopunu), a tokom vremena prolazimo i zajednička iskustva. I zdravorazumski se onda može očekivati da je zadovoljstvo vezom u prosjeku sličnije unutar parova nego među osobama koje nisu u vezi. Iz tog razloga se porodične ili partnerske dijade (poput odnosa majka-dijete, brat-sestra ili muž-žena) mogu posmatrati kao posebne jedinice analize (entiteti). U tim slučajevima, razlike ili sličnosti između članova tih dijada postaju važniji predmet istraživanja od podataka za svakog pojedinačnog ispitanika.

Štaviše, uzimanje u obzir međusobne sličnosti nije relevantno samo onda kada razmatramo bliske interpersonalne odnose nego je relevantno i kada unutar uzorka postoje obuhvatnije grupe. Na primjer, članovi iste društvene klase, kulturne zajednice ili profesionalne organizacije često dijele obrasce mišljenja, vrijednosti i ponašanja. Ovo se dešava zbog kombinacije genetskih faktora, okruženja i selektivnih interakcija. Na primjer, ako bismo analizirali zadovoljstvo zaposlenih u istoj organizaciji, vjerovatno bismo pronašli viši stepen sličnosti među njima nego među zaposlenima iz drugih organizacija. Ovo proizilazi iz zajedničkih uslova rada i organizacione kulture. Slični obrasci javljaju se i u obrazovanju. Na primjer, učenici iz različitih gradova unutar jedne države, iz različitih škola u istom gradu, pa čak i iz različitih odjeljenja unutar jedne škole, mogu imati značajno različite prosječne akademske rezultate. Razlike mogu biti rezultat socioekonomskog statusa gradova ili dijelova gradova, motivisanosti i pedagoških kompetencija nastavnika koji vam predaju, pa čak i pojedinačnih učenika u svakom odjeljenju koji ili inspirišu ili ugrožavaju školsku klimu i motivaciju za akademsko postignuće.

Zbog ovih razloga se pri kreiranju nacrta i kasnijoj statističkoj

obradi entiteti nekad definišu kao pripadnost širim društvenim grupama ili institucionalnim kategorijama (npr. entitet u fokusu analize može biti određena škola ili preduzeće). To može ići čak do nivoa država kada se simultano ili longitudinalno - kroz duži vremenski period - mogu upoređivati obrazovne kompetencije i sociopsihološki fenomeni (npr. stavovi prema seksualnim manjinama, stopa suicida, prevalencija depresivnosti, zadovoljstvo životom).

Šta su prednosti korištenja mjera vezanih entiteta?

Kao prva prednost se najčešće navodi da istraživanja koja koriste iste ispitanike imaju veću statističku moć u odnosu na istraživanja sa nezavisnim uzorcima. To znači da je potreban manji broj ispitanika da biste došli do željenog statističkog zaključka. Da bismo to ilustrovali, zamislite da izvodite probnu evaluaciju novog psihoterapijskog tretmana anksioznosti. Istraživanje vas obavezuje da imate kontrolnu i eksperimentalnu grupu, ali da možete skupiti najviše četiri mjere (sveukupno). Dilema je sljedeća: da li biste se odlučili da imate po dva ispitanika u kontrolnoj i eksperimentalnoj grupi (nezavisne grupe) i da njihovu anksioznost izmjerite samo na kraju istraživanja ili biste radije po jednog ispitanika rasporedili u kontrolnu i eksperimentalnu grupu i izmjerili njihovu anksioznost prije i poslije psihoterapije (mjere vezanih entiteta)? Epistemička vrijednost prvog nacrta je niža jer kakve god rezultate da dobijete, s logičke strane ih je teško pripisati dejstvu terapije budući da ne znate kakva je bila anksioznost osoba prije početka tretmana. S druge strane, ako se pokaže da postoji neznatna promjena za ispitanika iz kontrolne grupe, a već neko umjereno snižavanje anksioznosti ispitanika koji je prošao tretman, onda biste već imali naznaku da bi terapija mogla biti djelotvorna. Razlika između ova dva vida nacrta se može uporediti sa razlikom između statičnog foto snimka i filma koji se sastoji iz mnogo statičnih slika koje se brzo smjenjuju. Znate i sami da nam u velikoj većini slučaja film daje više uvida.

Druga bitna prednost korištenja nacrta sa međupovezanim entitetima je ta što se njime mogu provjeravati hipoteze koje se ne mogu provjeravati drugim nacrtima. Radi se o hipotezama koje u obzir uzimaju interakciju između nivoa izraženosti varijable i protoka vremena. Na primjer, moguće je pretpostaviti da psihoterapeutske intervencije u većoj mjeri koriste osobama čija je anksioznost na pretestu visoka nego onima koji imaju

samo umjereno izraženu anksioznost. Nadalje, vezane mjere omogućavaju praćenje brzine i smjera promjena koja se odnosi na entitet, bilo da je to pojedinac ili neka grupa. Kako ćemo vidjeti, na osnovu ovih nacrta je moguće uočiti trendove za cijele grupe, ali i atipične obrasce promjena. Takvi uvidi mogu biti vrlo značajni za psihologe u praksi. Konačno, nacrti vezanih entiteta omogućavaju procjenu saglasnosti različitih instrumenata koji imaju isti predmet mjerenja. Ovo je vrlo bitno budući da se tako može utvrditi koji je od dva ili više mogućih instrumenata bolji za određenu svrhu. Za sve navedeno ćemo demonstrirati primjere u daljnjem tekstu.

Postoje li uopšte mane korištenja mjera vezanih entiteta?

Ako gledamo striktno iz ugla statistike, jedina bitna mana jeste to što sa mjerama vezanih entiteta dobijamo više međupovezanih informacija koje mogu postati prekompleksne za prikazivanje, analizu i razumijevanje. Ipak, postoji sijaset metodoloških ograničenja. Kako se ovaj tekst primarno bavi statističkim rezonovanjem, a ne metodologijom nacrta, samo ću podsjetiti na one osnovne koje ste vjerovatno već saznali na nekom drugom mjestu. Ta ograničenja se ne javljaju u svakom istraživanju vezanih entiteta, ali se u njima mogu javiti, ovisno o temi. Na rezultate tako mogu djelovati: prenos usljed učenja (npr. ako se koriste testovi koji se manje ili više ponavljaju ispitanici uče iz njih te postižu bolje rezultate), zamor (npr. ako se više puta koristi isti zadatak ispitanici ih mogu raditi sa manje pažnje usljed umora i/ili demotivisanosti), redoslijed mjerenja (npr. ranija mjerenja mogu da na različit način utiču na kasnija mjerenja tako što se pobuđuju različite asocijacije), te događaji između mjerenja koji možda nemaju vezu sa nezavisnom varijablom, a utiču na zavisnu varijablu. Na to sve, izvođenje nacrta sa vezanim entitetima je obično zahtjevnije u smislu da je potrebno uložiti više truda u logističke aranžmane kao što je obezbjeđivanje motivisanih ispitanika, njihovo praćenje kroz vrijeme, te povećane šanse da ispitanici odustanu potpuno ili da ne budu dostupni na nekom broju mjerenja. Uz to, statističke analize vezanih entiteta brzo postaju kompleksne i mnoge od njih nadilaze okvire ovog teksta. Iz tog razloga su ovakva istraživanja rjeđa i izvode se onda kada metodski problemi nemaju dominantno negativan efekat po internu valjanost.

Kako se definišu baze podataka za mjere vezanih entiteta?

Prije nego što krenemo na prikazivanje konkretnih tehnika, korisno bi bilo podsjetiti se načina strukturisanja baza podataka kada imamo vezane entitete. Naime, u trećem poglavlju sam najavio da se podaci za ponovljena mjerenja mogu javiti u tzv. dugačkom formatu. Dugački format zaista omogućava jasno strukturisanje podataka kada postoji više mjerenja po ispitaniku za istu varijablu, ali nije jedini način na koji se podaci mogu organizovati. Isti podaci se mogu predstaviti i u širokom formatu, pri čemu se svako mjerenje bilježi kao zasebna kolona (uporedite Slike 10.1 i 10.2). U softveru se obično većina deskriptivne statistike može obaviti na širokim bazama, dok obuhvatnije statističko modeliranje obično zahtijeva dugačke baze.

Svakako je korisno da se sami podsjetite ranije prikazanog primjera, a ovdje navodim novi koji se ne odnosi na ponovljena mjerenja na jednom ispitaniku, nego na situaciju gdje bi jedna država predstavljala entitet od interesa za koji se veže više mjera. Radi se o potpuno fiktivnim podacima koji se tiču incidencije (novih slučajeva) kliničke depresije u periodu između 2019. i 2023. godine unutar tri države.

drzava	2019	2020	2021	2022	2023
Belgija	305	428	512	489	401
Švedska	298	415	508	472	392
Portugal	287	402	495	458	375

Slika 10.1: Baza podataka u širokom formatu.

drzava	godina	novi_slucajevi
Belgija	2019	305
Belgija	2020	428
Belgija	2021	512
Belgija	2022	489
Belgija	2023	401
Švedska	2019	298
...	...	...

Slika 10.2: Identična baza u dugačkom formatu.

Kako se grafički predstavljaju vezane mjere?

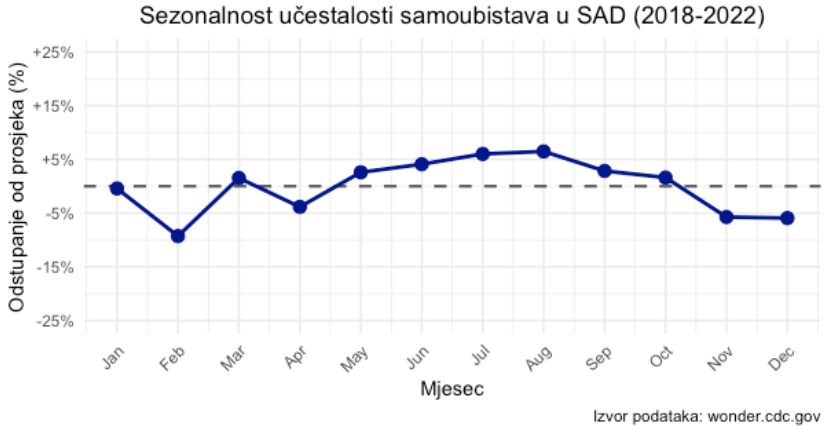
U mnogim slučajevima možemo, a ponekad i moramo, predstavljati podatke na neki od već prikazanih načina (npr. grafikoni raspršenja, toplotne mape, kišni oblaci, ostala meteorološka stanja :)). Vidjećemo da ipak postoje specifični načini prikazivanja vezanih mjera, koji baš insistiraju da prikažu te veze i vezice.

Najjednostavniji i najintuitivniji takav grafikon je **linijski grafikon** (engl. *line chart/graph*). Njegova osnovna namjena je prikazati promjenu neke ishodišne varijable (na Y-osi) kroz vrijeme (X-osa; godine, mjeseci, dani, sati, minuti…). Ishodišna varijabla može biti i kategorička i dimenziona. Za kategoričke varijable se na Y-osi navodi apsolutna ili relativna frekvencija, dok se za dimenzione navodi relevantna statistička mjera (npr. aritmetička sredina). Moguće je prikazati vrijednosti samo za jedan entitet ili više njih, a ako se poredi više grupa, onda boje i vrste linija mogu da pomognu u njihovom razlikovanju. Loša praksa je umjesto linijskog grafikona koristiti stupčasti grafikon

budući da se tada više površine troši na nebitne elemente.

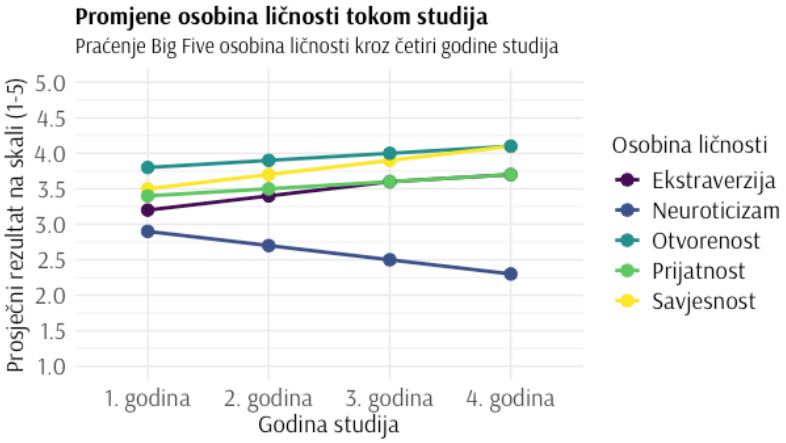
Ovdje navodim dva primjera. Slika 10.3 prikazuje stvarne podatke o sezonalnosti samoubistava u SAD na osnovu zvanično prikupljenih podataka sa stranice https://wonder.cdc.gov/ u periodu 2018-2022. godina. Iz grafikona se vidi da postoje razlike po mjesecima i da je učestalost samoubistava bila nešto veća u toplijim mjesecima, naročito u julu i avgustu.

Linijski grafikoni mogu dati zaista brz uvid u društvena kretanja tokom vremena i inspirišu na postavljanje hipoteza o mogućim uzrocima varijabilnosti. Dobar primjer za to je grafikon koji prikazuje standardizovanu stopu suicida od 1990. do 2017. godine na stranici: <https://www.csmonitor.com/World/Points-of-Progress/2019/0114/The-global-suicide-rate-has-seen-a-net-decline.-What-caused-it> Koje hipoteze biste vi izveli o razlozima zašto postoje različiti trendovi u kulturama koje imaju različitu istoriju i dominantne vrijednosti?



Slika 10.3: Sezonalnost samoubistava u SAD od 2020- do 2022. godine.

Drugi primjer je fiktivan i prikazuje poželjnu sliku onog što bi se moglo desiti tokom studiranja psihologije kada su u pitanju promjene Velikih pet crta ličnosti (Slika 10.4). Ovakav grafikon bi ukazivao na trend snižavanja emocionalne nestabilnosti, dok druge crte ličnosti pokazuju tendenciju rasta. Naravno, umjesto aritmetičke sredine bilo je moguće prikazati i druge mjere centralne tendencije, pa i mjere varijabilnosti (npr. "kretanje" standardnih devijacija).



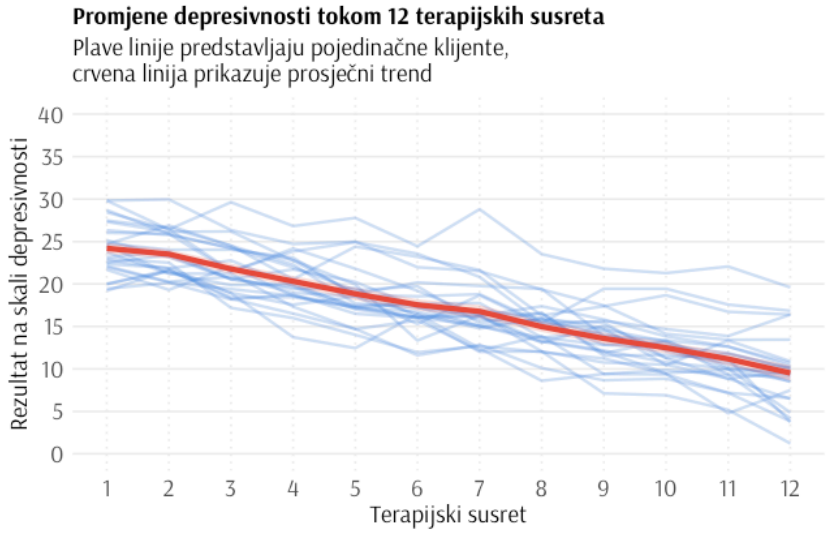
Slika 10.4: Razvoj Velikih pet kroz godine studiranja psihologije.

Kada tumačimo linijski grafikon potrebno je da obratimo pažnju na sljedeće elemente:

- trend (da li linija raste, opada ili stagnira)
- strminu nagiba (koliko brzo se promjena dešava)
- tačke "loma" (kada linija mijenja trend)
- varijabilnost (koliko je sveukupno gledano linija izlomljena ili glatka)

Špagetni grafikon (engl. *spaghetti plot*) je, narodnim jezikom rečeno, linijski grafikon sa baš velikim brojem linija. Štaviše, naučna zajednica je neusaglašena oko toga da li se time imenuje zasebna vrsta grafikona ili se samo radi o uvredljivom terminu za loše dizajniran linijski grafikon. Ako se posmatra kao zasebna vrsta, onda bi novost bila u tome da se pojedinačnim linijima prikazuju ponovljena mjerenja za individualne entitete, a nekom debljom linijom (linijčinom? ital. *bucatini*?) opšti trend. Dakle, datim grafikonom možemo uočiti individualne razlike u dinamici tokom vremena i da li i kako individualni trendovi odstupaju od opšteg.

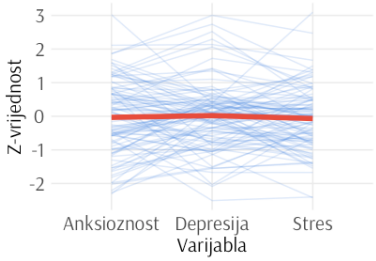
Špagetni grafikon se koriste i u slučajevima kada za svakog ispitanika nisu prikupljeni podaci u svim vremenskim tačkama. Npr. istraživanje Chen-a i saradnika (2022, doi.org/10.1016/j.neuroimage.2022.119276) je trajalo četiri godine, njegovi učesnici su imali između 20 i 89 godina na samom početku istraživanja, a istraživanjem su željeli proučiti faktore koji utiču na kognitivne promjene tokom vremena. Pogledajte Sliku 3 u tom članku i razmislite šta se sve može zaključiti iz nje.



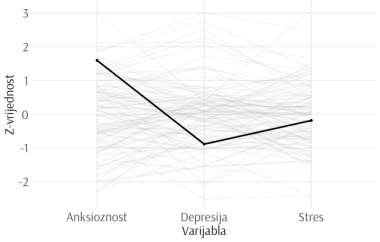
Slika 10.5: Špagetni grafikon.

Slika 10.5 prikazuje fiktivan i potpuno idealizovan grafikon promjena u simptomatologiji depresivnosti kroz 12 terapeutskih seansi. Inače, špagetne grafikone tumačimo kao i linijske, s tim da posebnu pažnju možemo posvetiti i individualnim putanjama, te eventualno pojavi nekih grupnih obrazaca.

Grafikon paralelnih koordinata (engl. *parallel coordinates plot*) je još jedna blaga varijacija na temu linijskog, odnosno špagetnog



Slika 10.6: Grafikon paralelnih koordinata.



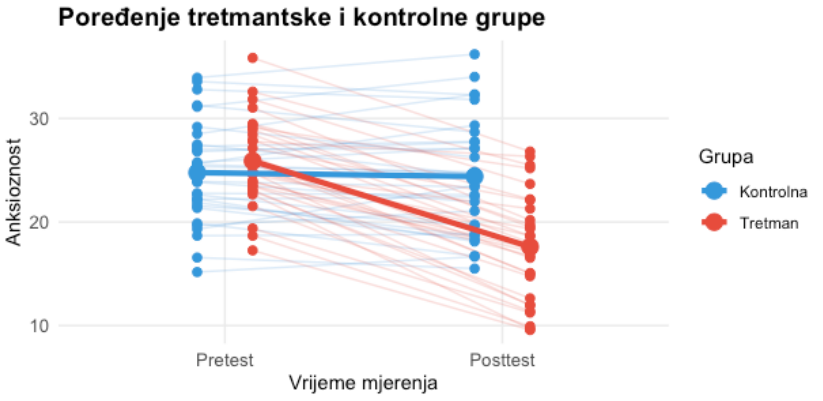
Slika 10.7: Grafikon paralelnih koordinata kojim je izdvojen atipični obrazac.

[1] Interaktivne grafikone ovog tipa je moguće kreirati npr. u Plotly okruženju \ https://plotly.com/python/parallel-coordinates-plot/

grafikona. Kao i u slučaju špagetnog grafikona predstavlja se mnoštvo linija, te obično i jasno označeni opšti trend. Razlika se sastoji u tome da se na X-osi predstavljaju različite varijable. One su definisane vertikalnim i paralelno postavljenim linijama (otuda naziv paralelne koordinate) koje predstavljaju mjerne skale. Te mjerne skale su često standardizovane, odnosno predstavljene kao z-skorovi. S obzirom na to i njihovo tumačenje drugačije od ranije opisanih grafikona. Pogledajte Sliku 10.6. Korelacije između prikazanih varijabli anksioznost, depresivnost i stres su pozitivne i kreću se u rasponu između .50 i .60. Međutim, to ne možete da uočite na osnovu linije koja prikazuje opšti trend. Ta linija samo prikazuje da razlika između prosjeka ovih varijabli nema, budući da sve aritmetičke sredine z-skala moraju biti jednake nuli. Ali oštrije oko ipak može nešto zaključiti o korelacijama na osnovu omjera paralelnih i ukrštenih linija: što je više paralelnih linija to će i korelacija biti veća.

Međutim, to nije osnovna namjena grafikona paralelnih koordinata. Njegova osnovna namjena je uočiti individualne rezultate i čudnovate obrasce, a to se može onda kada se izdvoji posebna linija ili više njih - kao što to prikazuje Slika 10.7. U idealnom slučaju se ovakvi grafikoni kreiraju kao interaktivni[1], gdje se linija podeblja kada korisnik klikne na nju.

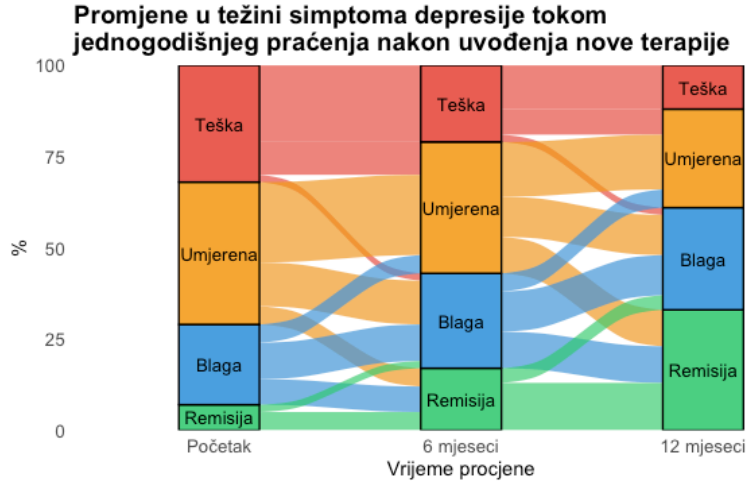
Poseban oblik grafikona paralelnih koordinata je mješavina linijskog i tačkastog i koristi se za prikazivanje razlika između pretesta i posttesta. Slika 10.8 prikazuje jedan takav grafikon koji daje mnogo bolji uvid u podatke nego grafikoni koji prikazuju samo aritmetičke sredine. Međutim, moramo imati na umu da bi tada bilo besmisleno koristiti standardizovane skorove, jer bi prosječna razlika morala biti jednaka nula.



Slika 10.8: Grafikon za prikazivanje pretest-posttest podataka.

Aluvijalni grafikon, koji se još zove i Sankeyev grafikon ili

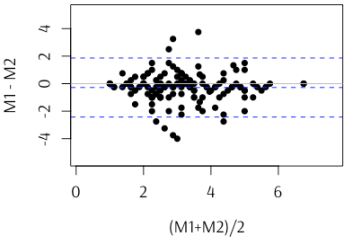
grafikon paralelnih skupova (engl. *alluvial, Sankey, parallel sets plot*) se koristi za prikazivanje promjena u kategoričkim varijablama. Konkretno, njime se prikazuju promjene u relativnoj zastupljenosti kategorija unutar čitavog uzorka na različitim mjerenjima kojih može biti i više od dva. Primjer je dat na Slici 10.9 gdje se vidi koliko ispitanika se tokom vremena "prebacuje" iz jedne u drugu kategoriju.



Slika 10.9: Promjene nivoa depresije tokom godinu dana od početka nove terapije.

Bland-Altmanov grafikon saglasnosti (engl. *Bland-Altman plot*) je posebna vrsta grafikona namijenjena za procjenu apsolutne saglasnosti između dva mjerenja. Tehnički gledano, sa njegovim rođakom - grafikonom reziduala - smo se već sreli (vidjeti poglavlje o povezanosti dvije dimenzione varijable). Bland-Altmanov grafikon je nešto sofisticiranija verzija čija je svrha upoređivanje slaganja dva mjerna instrumenta koji mjere isti predmet mjerenja. To radi tako što na Y-osi prikazuje razlike između dva mjerenja ($M_1 - M_2$), a na X-osi njihov prosjek ($(M_1 + M_2)/2$).

Na Slici 10.10 su prikazani podaci iz istraživanja 138 parova roditelja (majke i očevi) koji su procjenjivali svoju djecu starosti između 3 i 7 godina na različitim dimenzijama temperamenta - a između ostalog i stidljivosti djeteta (Lakić, Miščević & Tutnjević, 2014). Bland-Altmanov grafikon nam pokazuje kolika je saglasnost roditeljskih procjena. Da je svaki par majki i očeva imao istu procjenu onda bi sve tačke bile navedene na liniji koja označava 0 na Y-osi. Vidimo da takvih ima - kao što su otac i majka koji su dali maksimalan odgovor 7 na svim pitanjima kojima je procjenjivana stidljivost. Ipak, postoje i roditelji koji su imali veće razlike u procjeni - to je 8 partnera čiji parovi vrijednosti su



Slika 10.10: Bland-Altman grafikon procjene stidljivosti djece starosti 3 do 7 godina od strane majki (1) i očeva (2)

izvan spoljnih plavih linija - u ovom slučaju veći od vrijednosti 2. Ono što se takođe može vidjeti jeste da je centralna plava isprekidana linija - koja predstavlja prosjek razlika - nešto ispod vrijednosti nula. Izraženo sirovim skorovima, očevi teže da procjenju da su im djeca stidljivija nego što to procjenjuju majke ($M_o = 3.30, M_m = 3.02$).

Dakle, s jedne strane, Bland-Altmanov grafikon nam pomažemo da uvidimo tendenciju da li je za jednog procjenjivača dobijamo dosljedno više mjere, kao i to da vidimo u kojem opsegu možemo očekivati razlike ako dobijemo mjere iz samo jedne tačke (npr. samo procjene majke ili samo oca). Iz toga slijedi da se Bland-Altmanov grafikon uglavnom koristi u psihometriji za procjenu pristrasnosti, saglasnosti različitih formi instrumenta i/ili različitih procjenjivača.

Koje statistike koristimo da opišemo vezane mjere?

Prethodna poglavlja koja su prikazivala odnose između dvije (i više) varijabli su koristila i mjere razlika i mjere povezanosti. Ta dva tipa mjera se sreću i kada se obrađuju nacrti sa vezanim entitetima. Svi tada opisani koeficijenti povezanosti, kao i mjere prostih i standardizovanih razlika se koriste i za opis odnosa vezanih mjera. Međutim, specifičnosti vezanih mjera traže i dodatne analize. Slijedi njihov prikaz i upoređivanje sa dosad poznatim.

Kako izraziti razliku između dvije vezane dimenzione varijable?

Kada smo dvije grupe upoređivali po nekoj dimenzionoj varijabli, naveli smo da se najčešće koriste proste i standardizovane razlike aritmetičkih sredina (npr. Cohenovo *d*). To bi bili prvi izbori i u istraživanjima u kojima koristimo vezane mjere. U gore navedenom primjeru sa procjenama očeva i majki uz aritmetičke sredine je dovoljno da imamo i podatke o standardnim devijacijama ($SD_o = 1.33, SD_m = 1.32$). Kalkulacija pokazuje da se radi o malim razlikama među grupama ($d = -0.21$).

rb	majka	otac	razlika m-o
1	2	2.5	-0.5
2	5.75	4.25	1.50
3	4.5	4.75	-0.25
4	4.75	5	-0.25

Slika 10.11: Tabela koja prikazuje sparene procjene stidljivosti za majke i očeve.

Međutim, u slučaju vezanih entiteta postoji mogućnost za još jednu mjeru prosjeka. Na Slici 10.11 su prikazana prva četiri para procjena za majke i očeve. Između procjena majki i procjena očeva je moguće izračunati razliku, a onda je naravno moguće

izračunati i mjere centralne tendencije i varijabilnosti za sve dobijene razlike.

Ako aritmetičku sredinu kolone sa razlikama podijelimo njenom standardnom devijacijom, dobijamo standardizovanu prosječnu razliku unutar pojedinca. Ona se označava sa d_z - upravo zato što je to standardizovana razlika, izražena kao z-skor.

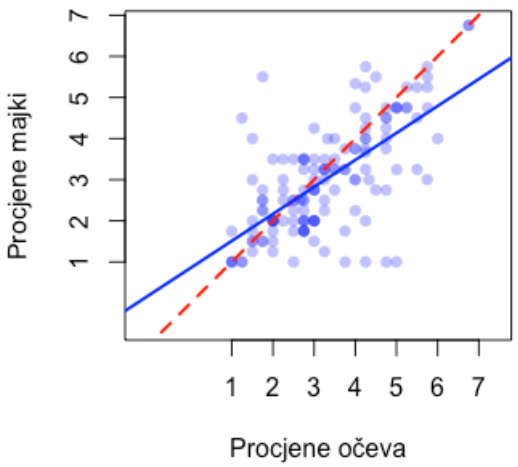
$$d_z = \frac{M_{\text{razlike}}}{SD_{\text{razlike}}}$$

d_z ima gotovo uvijek različitu vrijednost od vrijednosti Cohenovog d (drugim riječima, d_z nije d). I to je sasvim logično jer operacionalizuju dvije različite stvari: d označava prosječnu razliku između grupa mjerenja (npr. pretest grupa i posttest grupa), a d_z prosječnu promjenu unutar entiteta. Postoje čak i neke pravilnosti koje opisuju njihov odnos: ako je Cohenovo d jednako 0, onda znamo da ni d_z neće biti bitno različito od 0. S druge strane, d_z vrijednost je obično viša od d, a vrlo rijetko se dešava da Cohenovo d ima višu vrijednost. To se događa onda kada su korelacije između prvog i drugog mjerenja blizu vrijednosti 0 ili ako stoje u negativnom odnosu. U našem slučaju je vrijednost d_z bila blago viša ($M_{razlika} = -0.28, SD_{razlika} = 1.09, d_z = -0.26$).

Kako izraziti saglasnost dvije dimenzione varijable vezane entitetima?

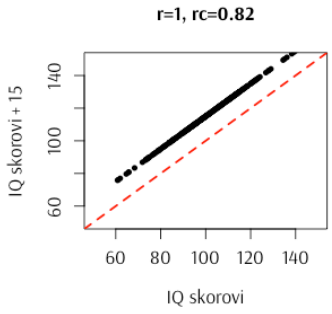
Toliko o razlikama, a sada prelazimo na izražavanje povezanosti, ali putem mjera povezanosti i mjera saglasnosti (da, dobro ste pročitali, razlikuju se). Na osnovu podatke iz istog istraživanja izveden je grafikon raspršenja (Slika 10.12). Uz postojeću linearnu korelaciju (puna plava linija) vidimo i još jednu dijagonalnu liniju koja prikazuje savršenu saglasnost mjera (isprekidana crvena linija). Da su majke i očevi identično procjenjivali stidljivost svoje djece onda bi sve tačke bile na toj isprekidanoj liniji. Očito saglasnost nije potpuna, uprkos tome što vidimo da postoji visoka povezanost između procjena ($r = .66$).

Linov koeficijent saglasnosti (engl. *Lin's concordance correlation coefficient*) je mjera kojom se procjenjuje **apsolutna linearna saglasnost** između dvije dimenzione procjene. Ona se u ovom primjeru koristi za procjenu saglasnosti različitih procjenjivača na istoj operacionalizaciji, a inače se može koristiti i za procjenu saglasnosti različitih operacionalizacija. Primjer bi bio da se zadaju dva testa inteligencije na jednom uzorku. U literaturi se najčešće označava



Slika 10.12: Povezanost procjene stidljivosti djece od strane očeva i majki.

kao r_c ili CCC. Kao i za koeficijent linearne korelacije, njen teoretski raspon se kreće od -1 do +1. No, dok je za izuzetno visoku i gotovo savršenu linearnu korelaciju dovoljno da redoslijed mjera na drugoj varijabli ostane isti, Linov koeficijent saglasnosti uzima u obzir i apsolutnu vrijednost mjera na mjernoj skali. Kako je pristrasnost relativno mala (očevi i majke su se relativno malo razlikovali), te kako su odstupanja od savršene saglasnosti simetrična i ujednačena (što nam pokazuje i Bland-Altman plot), tako je i razlika između dobijenog koeficijenta linearne korelacije ($r = .66$) i Linovog koeficijenta saglasnosti faktički neprimjetna ($r_c = .65$).



Slika 10.13: Korelacija skorova sa istim skorovima uvećanim za 1SD.

Međutim, u velikom broju slučaja između njih može doći do značajnog razmimoilaženja. Na Slici 10.13 je prikazana korelacija skora na testu inteligencije sa tim istim skorovima uvećanim za 15 (to je, da vas podsjetim, linearna transformacija). Iako je linearna korelacija savršena $r = 1.00$, mjera saglasnosti je opala na $r_c = .82$. Naravno, iznosila bi 1.00 ako bi svi parovi mjera bili na iscrtanoj dijagonali.

Linov koeficijent u formuli izračunavanja već sadrži linearni koeficijent korelacije i zapravo se može izraziti kao koeficijent linearne korelacije pomnožen korektivnim faktorom C_b.

$$r_c = r \cdot C_b = r \cdot \frac{2SD_1SD_2}{SD_1^2 + SD_2^2 + (M_1 - M_2)^2}$$

Sumnjam da će vam ova formula pomoći da utvrdite smisao svog života, ali iz nje se može vidjeti da se korekcija bazira na varija-

bilnosti oba mjerenja, ali i na razlici između prosjeka prvog i drugog mjerenja. Dakle, korekcija će biti veća u dva slučaja: (1) što je veća razlika između X i Y prosjeka, te (2) što je veća razlika u standardnim devijacijama X i Y varijable. Ako su M_1 i M_2 jednake, te ako su SD_1 i SD_2 jednake onda je C_b jednako 1, pa će $r_c = r$, odnosno korekcije neće biti.

U literaturi se nekad navodi da se vrijednost 0.90 može smatrati pragom zadovoljavajuće saglasnosti. Međutim, uvijek je bolje pogledati grafikon saglasnosti i uočiti postoji li neki specifičan trend gdje su saglasnosti slabije[2], te analizirati postoji li pristrasnost mjerenja (tj. postoji li razlika u prosjecima).

[2] Sjećate li se šta ono znači heteroskedastičnost?

Za kraj, samo jedna usputna napomena: Linov koeficijent konkordantnosti je bliski rođak nečeg što se zove **intraklasni koeficijent korelacije (ICC)**. To je prilično komplikovana porodica mjera čija je svrha da se procijeni sličnost skorova unutar postojećih klasa (npr. sličnosti postignuća na nekom obrazovnom testu unutar jedne škole u odnosu na druge škole). Ima svoju funkciju u naprednoj statistici i psihometriji, ali je zasad ne obrađujemo.

Kako izraziti saglasnost dvije kategoričke varijable vezane entitetima?

Ostalo nam je još da posvetimo malo pažnje vezanim kategoričkim podacima. Oni se često susreću kada se uspoređuju sudovi dva ili više procjenjivača. Zamislimo situaciju u kojoj dva stručnjaka treba da procijene da li bi za 100 djece sa indikacijama poremećaja hiperaktivnosti i deficita pažnje (engl. *ADHD*) trebalo propisati lijek Ritalin. Slika 10.14 prikazuje tabelu u kojoj su dati fiktivni podaci.

Da li je potrebno farmakološki tretirati suspektni PHDP?

		Ekspert B		
		Ne	Da	
Ekspert A	Ne	30	5	35
	Da	25	40	65
		55	45	100

Slika 10.14: Tabela saglasnosti eksperata o potrebi za farmakološkim tretmanom.

U idealnom svijetu bi saglasnost stručnjaka bila savršena - što misli jedan, misli i drugi - međutim to se vrlo rijetko dešava u stvarnom svijetu. Vrijednosti koeficijenata korelacije ($\phi = .45$ i

$\gamma = .81$) pokazuju da postoji povezanost između ovih mjera. Velike razlike u njihovim vrijednostima sugerišu da postoji razlika u liberalnosti / konzervativnosti propisivanja Ritalina za ova dva eksperta. Ekspert A bi u 65% slučajeva dao Ritalin, a Ekspert B tek u 45% slučajeva.

Da bismo procijenili stepen saglasnosti eksperata moramo pribjeći drugim postupcima. Prvo ćemo izračunati **proste relativne frekvencije saglasnosti** (proporcije ili postoci) (engl. *proportion / percentage of agreement*). Formula je vrlo jednostavna: broj slučajeva kada su eksperti saglasni u svojoj preporuci (oba DA ili oba NE) dijeli se sa ukupnim brojem slučajeva. U datom primjeru je to $(30 + 40)/100 = 0.70$, odnosno $0.70 * 100\% = 70.0\%$ ako želimo da izrazimo rezultat na procentnoj mjernoj skali. Sedamdeset posto može zazvučati dobro, ali ako bolje razmislimo em postoji 30% nesaglasnosti, em bismo očekivali saglasnost od oko 50% ako su oba eksperta nasumice davali preporuke.

Zato se kao dopunske mjere koriste mjere sa korekcijom, među kojima je najpoznatija **Cohenov kappa koeficijent** (engl. *Cohen's kappa coefficient*). [Jeste opet onaj Cohen, i opet neko ko nije prvi predložio tu mjeru - bio je to Francis Galton.] Ovaj koeficijent se označava grčkim slovom κ (kappa), izražava se na proporcionoj mjernoj skali (od 0 do 1) i formula uzima u obzir proporciju saglasnosti koja se mogla desiti pukom slučajnošću:

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

U prikazanoj formuli p_o predstavlja opaženu proporciju saglasnih slučajava (o u indeksu dolazi od engl. *observed*), a p_e proporcije slučajeva za koje bismo očekivali da budu saglasne (e od engl. *expected*). Što više od slučajnosti odstupaju opaženi saglasni sudovi, to je κ veće. Ako je $\kappa = 0$ znači da imamo situaciju koju je mogla proizvesti i prosta slučajnost, odnosno da su oba eksperta "lupala" sudove - ali postoji čak i mogućnost da κ koeficijent bude negativan (kakvi bi to onda bili eksperti?)! Izneseno je više preporuka o tome kako kvalitativno tumačiti visinu κ koeficijenata, a ako se u nečem u velikoj mjeri one saglasne ("*pun intended*") onda je to za minimalni prag od 0.40 da bi se saglasnost ocijena kao dovoljna, dok vrijednost 0.75 razdvaja relativno dobru od odlične saglasnosti. Potrebno je da definišemo slučajno očekivane saglasne proporcije. Iz niže navedene formule se vidi da se one dobijaju zbrajanjem proizvoda svih marginalnih frekvencija za X i Y mjeru i njihovim dijeljenjem sa kvadriranom veličinom uzorka:

$$p_e = \sum_{i=1}^{k} \left(\frac{f_{i.}}{n} \cdot \frac{f_{.i}}{n} \right)$$

U datoj formuli i označava broj kategorije, a u našem slučaju maksimalan broj kategorija je $k = 2$. Tačkice označavaju da se u prvom slučaju radi o marginalnoj frekvenciji u redovima ($i.$), a u drugom u kolonama ($.i$), za tu istu kategoriju i (npr. za kategoriju *ne preporučuje*, a zatim za kategoriju *preporučuje*). Proračun je sljedeći:

$$p_e = (\frac{35}{100} * \frac{55}{100}) + (\frac{65}{100} * \frac{45}{100}) = 0.1925 + 0.2925 = 0.485$$

Dakle, uzimajući u obzir razlike između eksperata, moglo se očekivati da bi saglasnost mogla doseći 48.5% čak i ako su oni propisivali lijek nasumično. Uvrštavanjem te vrijednosti u kompletnu formulu, vidimo da je κ jedva prebacila minimalnu preporučeni prag od 0.40. Mogli bismo zaključiti da postoji neka saglasnost između ova dva eksperta, ali je ona daleko od idealne, te da postoji još dug put do usaglašavanja preporuka za davanje lijeka.

$$\kappa = \frac{0.70 - 0.485}{1 - 0.485} = \frac{0.215}{0.515} = 0.42$$

Trebamo ipak biti svjesni da se Cohenovom kappa koeficijentu upućuju i opravdane kritike. Ako je jedna od kategorija dominantna (npr. oba eksperta su u 95% slučajeva izabrali jednu opciju) onda njegova vrijednost može biti artificijelno snižena, čak i u slučajevima kada postoji veliko poklapanje procjena. Takođe, veći broj kategorija može artificijelno da poveća vrijednost ovog koeficijenta. Nadalje, za druge kombinacije mjera postoje bolje mjere saglasnosti: ponderisani Cohenov kappa koeficijent za ordinalne kategoričke podatke, Fleissova kappa kada postoji više od dva procjenjivača, Krippendorphova alpha kao obuhvatna mjera saglasnosti i još neke druge kombinacije prezimena i grčkih slova. Tokom COVID-19 pandemije su na popularnosti dobile mjere senzitivnosti, specifičnosti i druge mjere procjene dijagnostičkog testiranja. No, opis svih njih je bolje ostaviti za psihometriju.

Poglavlje 11

Uvod u statistiku zaključivanja

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta je statistika zaključivanja i u kakvoj vezi stoji sa istraživačkim problemima i hipotezama?
- Zašto ljudi često donose iracionalne odluke u probabilističkim situacijama i koje kognitivne pristrasnosti utiču na naše zaključivanje?
- Šta je vjerovatnoća, kako se formalno izražava koje su glavne razlike između objektivističkog i subjektivističkog pristupa vjerovatnoći?

Relativno rano u ovom tekstu smo saznali da se statistika najčešće dijeli na dvije velike oblasti: deskriptivnu statistiku i statistiku zaključivanja. Dosad smo se bavili deskriptivnom statistikom, odnosno bavili smo se sa numeričkim i grafičkim postupcima pomoću kojih se analiziraju i sažeto predstavljaju podaci koji su empirijski prikupljeni na uzorku. Svrha takvih postupaka je da sa što manje - a ipak dovoljno - informacija, pronicljivo predstavimo sebi i drugima komadić stvarnosti koji smo zahvatili istraživanjem.

Mada njihova konačna funkcija jeste da se otkriju postojeći trendovi i razlike, nismo se usuđivali da izvodimo formalne zaključke o stanju stvari u stvarnosti, odnosno o populaciji iz koje je uzorak uzet. Moglo bi se čak reći da smo se dosadašnjim analizama bavili utvrđivanjem prošlosti, onoga što je na uzorku opaženo, mada bi svako od nas – barem potajno – želio da govori o sadašnjosti i da proriče budućnost. Kvalitetno razumijevanje sadašnjosti i realistično proricanje budućnosti (predviđanje, predikcija) je neophodno da bismo došli do krajnjeg cilja nauke, a to je da budućnost kontrolišemo i mijenjamo. Ono što slijedi će omogućiti da steknemo takve moći.

Pretpostavljam da znate da je ravan mogućeg i vjerovatnog predmet naučnog interesovanja jedne posebne matematičke oblasti: teorije vjerovatnoće. Udruživanjem matematičke teorije vjerovatnoće sa intenzivnim proračunima za koje su sposobni današnji računari, te novim pristupima saznanju kao što je mašinsko učenje, smo dobili savremenu statistiku zaključivanja.

Ciljevi statistike zaključivanja

Statistika zaključivanja se može definisati kao domen statistike koji za ciljeve ima da na formalizovan, principijelan i objektivan način: 1) opiše stanje pojave u populaciji (tj. da se procijene populacioni parametri), 2) formira sudove o vjerovatnosti pojedinačnih hipoteza, te 3) provjerava i unapređuje naučne modele kojima se obično podrazumijeva da se konstrukti nalaze u složenim vezama. Uvidjećete da statistika zaključivanja predstavlja skup različitih tehnika i pristupa koji su komplementarni, ali nekad mogu izgledati i međusobno suprotstavljeni.

Krajnja svrha statistike zaključivanja

Ono što je zajedničko za sve pristupe jeste njihova krajnja svrha, a to je da se izmijene ljudska uvjerenja o objektivnoj stvarnosti na osnovu prikupljenih empirijskih podataka. Kako bi se to postiglo slijedi se jednostavan recept: stvori se zamišljeni teorijski model stvarnosti koji se poredi sa komadićama stvarne stvarnosti (upravo tako), te se ocjenjuje koliko je zamišljena slika saglasna sa dostupnim podacima, odnosno koliko se one razlikuju. Ovaj recept je operacionalizovan sveobuhvatnom konceptualnom formulom: podaci = model + greška. Cilj nauke je smanjiti grešku, odnosno neizvjesnost koju imamo o događajima u stvarnosti.

Vjerujem da ste saglasni sa tim da prelazak sa "opipljivih" podataka na zaključivanje o neviđenom zvuči kao izazovan epistemološki skok. Tim prije što postavke probabilističkog zaključivanja - odnosno zaključivanja koje nije crno-bijelo - djeluju strano svakodnevnom razmišljanju, pa za mnoge ljude ("normalce"?) je i teško shvatljivo. Mnogi zato ignorišu probabilističko zaključivanje ili bježe od njega, pogotovo u situacijama kada se potreba za takvim zaključivanjem nameće po prirodi stvari (npr. donošenje odluka o većim finansijskim ulaganjima; praktikovanje ponašanja koja nemaju neposredne efekte, ali imaju posljedice po zdravlje ili dobrobit na duge staze; kontinuirano igranje igara na sreću). Postoji nekoliko slikovitih primjera koji dokazuju da smo skloni da izvodimo zaključke suprotne probabilističkim načelima. Mada je već dosta popularizovana, Monty Hall dilema je i dalje idealan primjer usađenog opiranja probabilističkom zaključivanju.

Monty Hall dilema

Postoji nekoliko verzija Monty Hall dileme, a ovdje predstavljam pojednostavljenu verziju koja sadrži sva relevantna svojstva originala. Zamislite da učestvujete u igri u kojoj vam osoba nudi da

izaberete jednu od tri zatvorene koverte koje su pred vama. U jednoj od koverata se nalazi 1000 EUR (ili drugih važećih valuta), a sadržaj druge dvije koverte čini 1000 "para" iz društvene igre Monopol. Možete zadržati sadržinu samo jedne koverte. U prvom koraku vam osoba koja vodi igru ostavlja slobodan izbor da izaberete jednu kovertu. Nakon što odaberete jednu od njih (npr. kovertu broj 1), voditelj igre koji zna šta se u kojoj koverti nalazi je onda obavezan da otvori jednu od preostale dvije koverte u kojoj se nalaze lažne pare (npr. broj 3) i otkrije vam njen sadržaj. Zatim vam mora izreći sljedeću ponudu: "Želite li ipak da izaberete broj 2?". Da li je pametnije, odnosno racionalnije, prihvatiti tu ponudu i promijeniti izbor (tj. izabrati broj 2) ili ostati na vašem početnom izboru (tj. koverti broj 1)? Šta biste vi uradili?

Mnogobrojna istraživanja su pokazala da velika većina ljudi – čak i profesora matematike (jeste, i oni su ljudi) – odlučuju da ostanu pri svom prvobitnom izboru, pri čemu mogu još da tvrde da nije važno da li bi promijenili prvobitni izbor ili ostali na njemu. Međutim, relativno jednostavna analiza svih opcija kada se izbor promijeni (Slika 11.1), kao i simulacije putem appleta (npr. https://www.rossmanchance.com/applets/2021/montyhall/Monty.html ili https://www.mathwarehouse.com/monty-hall-simulation-online/) jasno pokazuju da postoji veća vjerovatnoća da ćete osvojiti pravi novac ako promijenite izbor (66.7% šanse) nego ako ostanete pri početnom (33.3% šanse). Možda će i vama biti potrebno više vremena da **izmijenite svoje uvjerenje** nakon što proučite navedeno, ali vjerujem da ćete na kraju doći do istog zaključka.[1]

[1] Ako vam ni tabela, ni simulacije ne pomognu da shvatite, možda će vam pomoći da razmišljate ovako: ako se zaista novac nalazi u koverti broj 1, onda je šansa da pogodim iz prve ⅓ ako sam izabrao/la baš tu kovertu. Međutim, ako je novac u koverti broj 1, a ja sam izabrao/la kovertu 2 ili 3 na početku pa mi se otkrije da u preostaloj koverti nije novac, onda promjenom izbora vjerovatnoća pogotka postaje ⅔

Pravi novac	Igrač odabira	Voditelj otvara	Konačan izbor	Rezultat
A	A	B ili C	C ili B	-
A	B	C	A	+
A	C	B	A	+
B	A	C	B	+
B	B	A ili C	C ili A	-
B	C	A	B	+
C	A	B	C	+
C	B	A	C	+
C	C	A ili B	B ili A	-

Slika 11.1: Prikaz svih ishoda Monty Hall problema kada igrač izmijeni svoj prvobitni izbor.

Ali zaboga, zašto smo kao vrsta slabi u probabilističkom zaključivanju? Postoji više konkretnijih objašnjenja za to, ali su sva ona uglavnom obuhvaćena teorijom ograničene racionalnosti (engl.

bounded rationality). Koliko ta teorija predstavlja značajan doprinos znanju pokazuje i činjenica da je autor teorije, Herbert Simon, bio dobitnik Nobelove nagrade iz ekonomije, a da je drugu Nobelovu nagradu iz ekonomije na osnovu istraživanja u kontekstu te teorije dobio psiholog Daniel Kahneman (jedini psiholog koji je zaradio Nobelovu). Sažeto opisano, teorijom i pratećim istraživanjima se pokazuje da naša vrsta jednostavno nije razvijana tako da postanemo savršeno razumni organizmi. Savršeno razumni organizmi bi se uvijek ponašali u skladu sa proračunima koji nam donose maksimalnu dobit. Međutim, da bismo izveli savršene proračune - što se u kontekstu date teorije naziva optimizacija izbora - neophodno je da saznamo sve bitne informacije koje mogu uticati na ishod, da ih natenane procijenimo i onda izaberemo onu odluku koja ima najbolji omjer željene dobiti i neželjene štete. Niti mi na raspolaganju možemo imati sve potrebne informacije, niti imamo toliko vremena da ih pažljivo razlažemo.

Drugim riječima, naše odluke su tek djelimično razumne, onoliko koliko je potrebno da budu da bismo mogli izvući dovoljno zadovoljavajući ishod (na engleskom se to zove *satisficing*, što predstavlja kontrast optimizaciji). Umjesto da se služimo savršenim statističkim modelima mi odluke donosimo koristeći **heuristike**. Heuristici su misleće prečice na koje smo naviknuti i koje nam omogućavaju da pod vremenskim i energetskim pritiskom relativno brzo reagujemo. U heuristike spadaju i evolutivni odgovor na prijetnju koji su pod uticajem nagona, emocija i socijalnih draži. Na primjer, kada bi se naši davni preci sretali sa divljim zvijerima ili sa humanoidom jačim od sebe onda bi uradili nešto od navedenog: dali bi se u bijeg ili bi se borili ili bi se pokušali sakriti - ili bi u slučaju jačih primjeraka iste vrste pokazali poniznost i ponudili saradnju (koju bi napadač možda prihvatio). U mnogim situacijama su heuristici efikasni i primjereni - te su mnogima u davnoj prošlosti, ali i danas, spašavali glavu. Ipak, isto tako su u mnogim situacijama bili neumjesni i vodili su u katastrofalne greške u zaključivanju.

Psiholozi i filozofi su posebno posvećeni istraživanju heuristika koje predstavljaju kognitivne pristrasnosti, tako da postoje i kompletne taksonomije dokumentovane u vrlo popularnim knjigama. Primjeri svjetskih bestselera su "The Art of Thinking Clearly" čiji je autor Ralf Dobelli, "Thinking Fast and Slow" čiji je jedan od koautora pomenuti Kahneman, a izvorno na našem jeziku je pisana i knjiga "50 kognitivnih pristrasnosti" Predraga Stojadinovića.

Suboptimalnu odluku u Monty Hall dilemi može objasniti nekoliko kognitivnih pristrasnosti koje po svemu sudeći imaju kumulativan efekat.[2] Prva, i vjerovatno najuticajnija, se tiče uvjerenja o magičnoj posebnosti svakog od nas i naziva se **iluzija kontrole** (engl. *illusion of control*). Iluzija kontrole predstavlja tendenciju ljudi da vjeruju da mogu uticati na ishode događaja koji su zapravo sasvim slučajni ili na koje objektivno nemaju uticaja. U kontekstu Monty Hall dileme, iluzija kontrole se manifestuje kroz osjećaj da smo pomoću nekog intuitivnog znanja ili 'predosjećaja' inicijalno izabrali pravu kovertu, iako statistički gledano, taj iz-

[2] Neke empirijske dokaze za mehanizme našeg "moronizma" pružaju radovi A. Moronea i saradnika: https://doi.org/10.1007/s11238-021-09809-0 i https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5129596

bor ima samo 33.3% šanse da bude tačan. Ova pristrasnost nas vodi ka tome da precjenjujemo naše sposobnosti donošenja odluka i da se držimo prvobitnog izbora čak i kada postoji racionalniji pristup.

Druga pristrasnost je problem koji se tiče pretpostavke da nas drugi želi prevariti, odnosno nepovjerenja kojeg imamo u čin drugog. Naime, ponudu da nam pruža alternativu doživljavamo kao namjeran mamac da promijenimo ispravni ishod iako voditelj igre uopšte nema tu namjeru. Ovakav obrazac razmišljanja potpada pod pristrasnosti atribucije (puno ime: **temeljna atributivna pristrasnost**, engl. *fundamental attribution error*), gdje tuđe postupke težimo da tumačimo kroz prizmu mogućih skrivenih namjera i ličnosnih dispozicija koje im učitavamo, što je evolutivno korisno u socijalnim interakcijama, ali može voditi pogrešnim zaključcima u situacijama gdje postoje situaciona pravila koja voditelj jednostavno mora da poštuje. Drugim riječima, mi učitavamo tuđe loše motive jer nas priroda uči da budemo oprezni kada su ograničeni resursi u pitanju.

Treća se tiče **dejstva oduzetog** (engl. *endowment effect*) i vezuje se za strah od gubitka nečeg što je bilo naše. Konkretno, kada jednom donesemo odluku, ona postaje "naša" i stvara osjećaj vlasništva, što nas čini nevoljnim da je mijenjamo. Ako ostanemo kratkih rukava više ćemo se kajati ako smo promijenili odluku nego ako smo zadržali prvobitni izbor, budući da nam emocionalno teže pada da smo u jednom trenutku nešto "imali" pa "izgubili", nego da ni u jednom trenutku to nismo imali.

Konačno, tu je i jedan opštiji problem koji nas sprječava da se prebacimo na racionalnije odlučivanje u budućim situacijama, pogotovo ako je razlika u vjerovatnoćama mala. Ta pristrasnost se zove **pristrasnost ishoda** (engl. *outcome bias*). Ona se odnosi na to da o kvalitetu odluka težimo da sudimo na osnovu rezultata događaja, pa makar odluka bila pogrešna, odnosno iracionalna. Ako neko u konkretnoj situaciji izabere kovertu 1 i ostane pri svom izboru te zaista tako dođe do pravog novca onda su nam džaba sva ubjeđivanja da to nije bio racionalan izbor. I u mnogim drugim situacijama se ljudi vode ovom pristrasnošću, ne razumijući principe vjerovatnoće (npr. kude vakcinisanje ako ipak dobiju bolest ne razumijući da se vakcinacijom pojačava imunitet i smanjuje vjerovatnoća ozbiljnije bolesti, ali da to po sebi ne garantuje potpuni imunitet, pa čak ni to da neće doći do fatalnog ishoda). Sve u svemu, potrebno je biti svjestan da kada u privatnim situacijama imamo puno nepoznanica jednostavno moramo odlučiti intuitivno, pa tada naš sistem odlučivanja

uzima u obzir i informacije kojih možda i nismo svjesni.

Međutim, kada su krupne stvari u pitanju, odnosno kada želimo da stvorimo čvrst korpus znanja koji nam pomaže da donosimo važne društvene odluke u zdravstvu, obrazovanju i društvenim promjenama dobro je biti svjestan ogromnih prednosti logičko-matematičkog rezonovanja. Logika i matematika predstavljaju alate koje bi svako od nas mogao da primijeni na podjednak način. Iako je svaka ljudska djelatnost podložna korupciji, logičko-matematička ravan nam omogućava da se u velikoj mjeri eliminišu pristrasnosti koje su gore navedene, kao i mnoge druge koje postoje. Nauka zato primjenjuje formalni sistem za testiranje hipoteza koji minimizira uticaj ličnih uvjerenja i heuristika. Vrijeme je zato da se upoznamo sa konkretnim principima probabilističkog zaključivanja, uz podsjećanje da je za naučno zaključivanje potrebno i da imamo valjane operacionalizacije pojmova, odnosno kvalitetne podatke i metodološku kontrolu koji eliminišu alternativna objašnjenja.

Šta je istraživački problem, a šta hipoteza?

Istraživački problem

Svako naučno istraživanje započinje definisanjem **istraživačkog problema** (*engl. research problem*). Prosto rečeno, problem je neugodna situacija za koju nemamo gotovo, spremno rješenje. Dakle, istraživački problem je otvoreno pitanje na koje ne možemo dati adekvatan odgovor (čitaj: dovoljno objektivan i pouzdan), odnosno to je rupa u trenutnom znanju koju želimo popuniti provođenjem konkretnog istraživanja. U kontekstu naučnog rada, problem može biti takav da muči čitavo čovječanstvo (npr. pandemija, globalno zatopljenje), naučno-stručnu zajednicu u nekoj oblasti (npr. mogući negativni efekti korištenja društvenih mreža među adolescentima, nepridržavanje zdravih stilova života među osobama sa hroničnim bolestima) ili neki bitan segment ljudske zajednice (npr. postojanje neetičkih ponašanja pri pisanju naučnih radova).

Ako čitate ovaj tekst vrlo vjerovatno je da ste student. Vaš (veliki) problem, a i problem mnogih kolega studenata, širom svijeta, je kako lakše usvajati gradivo. Zar ne bi bilo divno da postoji nešto što ne predstavlja opasnost po zdravlje (npr. nije farmakološko sredstvo sa nuspojavama), a što vam omogućava da lakše pamtite ono što čitate? U potrazi za čarolijom se moglo desiti da nabasate na nešto što se zove *sugestopedija*. Sugestopedija je svojevrsni pedagoški pokret čiji je autor bugarski psihijatar Georgi Lozanov. Postulati sugestopedije tvrde da je jednostavnim intervencijama

moguće kreirati bolje uslove učenja što onda vodi uspješnijem pamćenju, pogotovo pri učenju stranih jezika. Postaćete još zainteresovaniji za sugestopediju ako predstavite sebi statistički vokabular kao strani jezik, pa će vam se učiniti zgodnim da provjerite da li bi neka načela radila i za učenje statistike. Na primjer, pristalice sugestopedije navode da je pri učenju korisno slušati nenapadnu instrumentalnu muziku (tj. muziku bez vokala) relativno sporog tempa koja stvara ugodnu atmosferu. Konkretno se tu najčešće spominju barokna djela kompozitora kao što su Bach, Vivaldi, Pachelbel, Corelli i Händel. Istraživački problem je to što nemamo kvalitetne empirijske podatke, a ni jasne zdravorazumske mehanizme, koji bi nas uvjerili da bi slušanje takve muzike pospješilo usvajanje gradiva. Sve nas navodi na to da bismo mogli poduzeti jedno istraživanje i ispitati **istraživačku hipotezu** (*engl. research hypothesis*), odnosno provjerljivu pretpostavku o stanju stvari u populaciji koja bi glasila: "Slušanje instrumentalne barokne muzike sporog tempa tokom učenja doprinosi boljem zapamćivanju gradiva iz statistike."

Istraživačke hipoteze

Ovakva, kao i skoro sve druge naučne hipoteze, predstavljaju jezički model stvarnosti. Mora se naglasiti da se radi o ekstremno pojednostavljenom modelu gdje dovodimo u vezu samo dvije pojave. No, od problema smo došli do nešto uže teorijske hipoteze, ali ostalo je još nekoliko koraka dok dođemo do statistike zaključivanja. Sljedeći korak je da osvijestimo da je teorijsku hipotezu potrebno operacionalizovati (npr. kako ćemo mi to procjenjivati "bolje zapamćivanje gradiva iz statistike"). Na primjer, za jedno konkretno istraživanje bismo mogli imati sljedeću radnu hipotezu koja kaže: "Studenti 1. godine psihologije koji tokom 30 minuta čitaju tekst o njima nepoznatim statističkim metodama i slušaju odabrana Vivaldijeva, Bachova, Pachelbelova, Corellijeva i Handelova djela u largo tempu (oko 60 taktova u minuti) ostvaruju bolji uspjeh na testu znanja iz datog gradiva neposredno nakon čitanja u odnosu na studente koji čitaju u tišini." Čak i ovom primjeru zapravo nedostaju mnogi detalji koji bi činili punu radnu hipotezu (npr. koja tačno djela, kako izgleda taj test, koliko bodova nosi u teoriji, kako će biti osigurano da ispitanici o toj metodi ništa ne znaju, koliko će ispitanika biti, kako će oni biti razvrstavani u grupe, kolika je glasnoća muzike, sluša li se ona na slušalicama ili zvučnicima, da li se gradivo i testovi zadaju na računaru ili u papir-olovka formatu). Tek kad sve to odredimo možemo preći na ravan statističkih hipoteza.

Pretpostavimo da smo odlučili da izvedemo randomizovani eksperiment u kontrolisanim uslovima (npr. koristićemo slušal-

ice za izlaganje muzike, ista muzička djela će slušati svi ispitanici u eksperimentalnoj grupi). U eksperimentu ćemo imati ukupno 20 ispitanika koji nemaju iskustvo niti sa statistikom niti sa baroknom muzikom koje ćemo nasumično rasporediti u dvije grupe od po 10 ispitanika. Eksperimentalna grupa će čitati tekst sa namjerom da ga nauči uz slušanje muzike, a kontrolna bez muzike, odnosno u tišini. Za svakog ispitanika ćemo imati dva prikupljena podatka: podatak o tome kojoj grupi pripada (npr. eksperimentalna = 1 ili kontrolna = 0), te podatak o tome koliko je bodova osvojio na testu znanja koji ima 20 zadatak, na teorijskoj skali od 0 do 20. Nakon što smo sve pripremili postavlja se logično pitanje: Kakve to rezultate treba da dobijemo da bismo bili ubijeđeni da je istraživačka hipoteza tačna, odnosno da slušanje takve muzike doprinosi boljem učenju?

Već vidim kako odobravate klimanjem glave ako bih rekao da bi dovoljan dokaz u prilog istraživačke hipoteze bio nalaz da su svi ispitanici iz eksperimentalne grupe imali bolje rezultate od svih ispitanika iz kontrolne grupe. Ako bi ta razlika bila velika - npr. svi ispitanici iz eksperimentalne grupe ostvare 20 bodova, a svi iz kontrolne 10 bodova, to bi bio snažniji dokaz nego kada bi razlika bila mala (npr. 11 naspram 10 bodova za svakog iz dvije grupe). Međutim, koliko je realistično da ćemo tako nešto dobiti? Učenje statističkog gradiva ne zavisi samo od toga prati li proces pogodna muzika ili ne. Postoji mnoštvo spoljnih varijabli koje utiču na postignuće na takvom testu kao što su: motivisanost, inteligencija, anksioznost, fiziološko stanje organizma, strategije učenja. Uz to, već znamo da ne ispitujemo determinističku, nego statističku hipotezu. Sve u svemu, očekivaćemo da postoji i varijabilnost rezultata unutar grupa.

Šta onda možemo uzeti kao dokaz u prilog istraživačkoj hipotezi? Ako umjesto pojedinačnih rezultata, za relevantnu mjeru uzmemo razliku u centralnim tendencijama (npr. aritmetičkim sredinama, budući da one predstavljaju grupe) koliko onda treba tačno da iznosi ta razlika, pa da je smatramo dovoljno ubjedljivom? Da li je potrebno da ta razlika bude velika - na primjer 5 bodova razlike, ili ona može biti i mala, ali primjetna, kao što je jedan bod razlike? Nadalje, ako je neka od tih razlika ubjedljiva za vas, da li će ona biti ubjedljiva i za nekog drugog?

Kada se ovakvo pitanje temeljno razmatra počinjemo da uviđamo da je mnogo odgovora moguće, te da je nužno da nekako ograničimo subjektivnost prizivajući u pomoć objektivnije intelektualne alate i standarde. Tokom istorije psihologije, ali i drugih empirijskih nauka, pokazalo se da se najviše povjerenja

stiče ako napravimo intersubjektivni dogovor u pogledu dva ključna elementa: na koji način definišemo vjerovatnoću i kako ćemo konkretizovati statističku hipotezu koju želimo provjeriti. Tim elementima moramo posvetiti dužnu pažnju za šta će vam biti potreban viši stepen koncentracije, pa se nekako spremite na to (npr. kafa, šetnja, meditacija, ili možda barokna muzika?).

Šta je to vjerovatnoća i koji su pristupi njenoj definiciji?

Prvo ključno pitanje koje bi trebalo razriješiti jeste šta je to zapravo vjerovatnoća, odnosno kako se ona može definisati. Nevjerovatno je - što kažu na engleskom "pun intended" - da su definicije vjerovatnoće vrlo raznolike i da potpunog slaganja baš i nema.

Dakle, postoje dva oprečna pristupa vjerovatnoći: objektivistički i subjektivistički. Objektivistički daje primarnost objektivnoj stvarnosti i smatra da su istinite vjerovatnoće samo one koje izražavaju karakteristike stvarnosti nezavisne od posmatrača. Tako imamo klasičnu teoriju vjerovatnoće gdje unaprijed znamo da kada palcem izbacimo običan novčić visoko u vazduh da imamo jednaku vjerovatnoću da će novčić sletjeti na jednu od dvije strane. Drugim riječima, za neke pojave unaprijed znamo parametre koji su dio stvarnosti, te na osnovu njih možemo predviđati kolike su šanse da se nešto desi.

Klasična vjerovatnoća

Drugi objektivistički pravac je frekvencionistički kojim se pretpostavlja da će se objektivna stvarnost empirijski otkriti ako beskrajno dugo ponavljamo bacanje novčića. Frekvencionistički pristup je dobro ilustrovan kompjuterskim simulacijima[3] ubrzanog "bacanja" novčića čiji su uobičajeni rezultati prikazani i na Slici 11.2. Kao što možete vidjeti više bacanja novčića će učiniti da se proporcija rezultata približi tačno 50%, odnosno onom što je očekivano klasičnom vjerovatnoćom kada znamo da postoje dva moguća ishoda koja su podjednako vjerovatna. Vidljivo je da na početku postoje značajna odstupanja, ali da će se do istine doći ako smo dovoljno strpljivi.

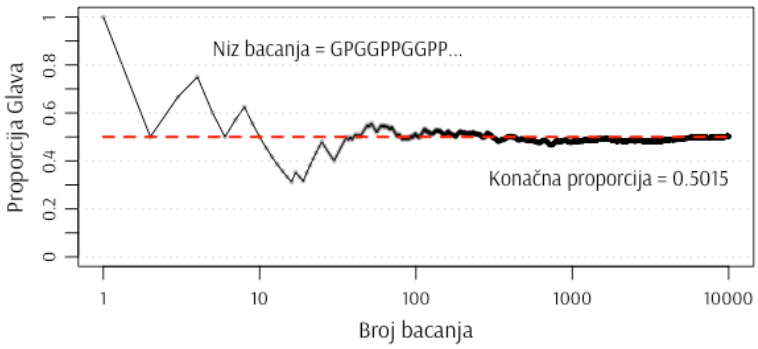
Frekvencionistička vjerovatnoća

[3] npr. https://digitalfirst.bfwpub.com/stats_applet/stats_applet_10_prob.html ili https://demonstrations.wolfram.com/SimulatedCoinTossingExperimentsAndTheLawOfLargeNumbers/

S dijametralno druge strane se nalazi subjektivistički pristup vjerovatnoći. Eksperti koji zauzimaju taj stav mogu čak otvoreno tvrditi da vjerovatnoće jednostavno ne postoje u spoljnoj stvarnosti[4]. Ti eksperti nisu u deluziji nego pod tim samo objavljuju činjenicu da je vjerovatnoća poimanje koje ljudi koriste kako bi se lakše snalazili u svijetu. Drugim riječima, vjerovatnoće su proizvod ljudskog mišljenja i postoje u umovima, a mogu – ali ne moraju – odražavati neke pravilnosti fizičke

Subjektivistička vjerovatnoća

[4] npr. de Finetti (1970), https://doi.org/10.1016/0001-6918(70)90012-0 i Nau (2001). de Finetti was right: Probability does not exist, https://doi.org/10.1023/A:1015525808214



Slika 11.2: Konvergencija frekvencijske vjerovatnoće ka poznatom parametru (0.50) u simulaciji bacanja novčića (n=10 000).

stvarnosti. Izuzetno korisno je ako se poklapaju sa stvarnošću i tome treba težiti, jer ćemo tako češće donositi racionalne odluke, ali stvarnost ne haje mnogo za to u šta smo mi uvjereni. U fokusu subjektivističkog pristupa su pitanja kako konkretno izraziti unutrašnja uvjerenje putem formalne vjerovatnoće i kako principijelno (sistematsko i formalizovano) ažurirati ta uvjerenja na osnovu empirijskih podataka. Subjektivistički pristup čak zauzima stav da su parametri kao konačne istine teško uhvatljivi, odnosno da čak i ako postoje ili nisu potpuno saznatljivi ili se mogu saznati ali su promjenjivi kroz vrijeme (kao što se i sav svijet mijenja). Nesigurnost u procjenu je ugrađena u subjektivistički pristup vjerovatnoći koji se često identifikuje i kao bejzijanski (o tome ćemo naveliko kasnije u tekstu).

Srećom, postoje neke polazne tačke oko kojih se slažu svi razumni agensi (čitaj: većina trezvenih ljudi) i što nam može pomoći da definišemo pojam vjerovatnoće. Na primjer, kada govorimo o vjerovatnoćama govorimo o pretpostavkama o tome da će se nešto desiti u budućnosti. To mogu biti pretpostavke o tome: (1) kako ćete proći na pismenom ispitu ako to procjenjujete prije nego što vidite pitanja, (2) da je slučajno odabrani student sa spiska studenata psihologije muškog pola, ili (3) da ćete biti zadovoljni romantičnim partnerskim odnosom sa osobom za koju ste se odlučili prvenstveno na osnovu njenog imovinskog stanja. Sve u svemu, vjerovatnoćom ćemo zvati stepen uvjerenja koje neko ima o određenom ishodu (npr. da će dobiti stranu sa brojem 6 prilikom bacanja kockice sa šest stranica) ili grupi ishoda (npr. da će bacanjem kockice dobiti neki paran broj, odnosno 2, 4 ili 6). Kako bismo izrazili svoj stepen uvjerenja u običnom i kafanskom jeziku koristimo odrednice kao što su *nimalo*, *slabo*, *umjereno*, *veoma* vjerovatno, pa i *skoro sigurno*, *sigurno*

Potrebno je naglasiti da na osnovu vjerovatnoća zaključujemo i o prošlim događajima ili o trenutnom stanju stvari. Zanimljivo je da na engleskom postoje dvije riječi koje se ne odnose nužno na temporalnost nego na to da li se rasuđivanje zasniva na parametrima ili statisticima. Riječ probability označava rasuđivanje od pretpostavljenih parametara do mogućih statistika, dok riječ likelihood označava rasuđivanje unazad od opaženih statistika ka mogućim parametrima. Likelihood se nekad prevodi kao inverzna vjerovatnoća ili vjerodostojnost. O tom aspektu će biti više riječi u poglavlju u bejzijanskoj statistici.

ili odrednice kao *moguće je*, *možda*, *ima šanse*, *fifti-fifti*. Upravo je ova šarolikost i neodređenost navela matematičare da stvari urede koliko mogu i jezičke priloge zamijene nekim brojkama.

E sad, matematičari su precizni ljudi, ali siguran sam (još jednom, *pun intended*) da među njima ima i nespretnih u komunikaciji. Prema mom mišljenju, generacije matematičara su koristili terminološki košmarna rješenja pa se čitajući matematičke tekstove krećemo kroz aksiome i teoreme koje počinju sa "neka je…", nakon čega se roje različita čudna slova koja mijenjanju konkretne primjere. To sve prate i nova značenja izraza eksperiment, opit, ogled, događaj, ishod, slučajnost, varijabla i njihove kombinacije. Budući da sam dugo godina i sam lupao glavu oko razumijevanja tih tekstova, smatram da takav pristup niti odgovara duhu jezika, niti je didaktičan. Zato u nastavku nudim nešto drugačiju terminologiju sa dubokom nadom da će vam biti lakše da razumijete suštinu iza tih riječi.

Slučajnost, ishodi, događaji?

Počećemo od onoga što se u literaturi nesretno naziva slučajnim eksperimentom (engl. *random experiment*). Za početak riječ slučajnost (engl. *randomness*) ima dva značenja: epistemičko i ontološko, a na ovom mjestu ćemo razmatrati samo epistemičko značenje, odnosno ono vezano za naše znanje. U tom značenju, riječ *slučajno* označava da nešto nije u potpunosti predvidivo. Mislim da bi zato bilo razumljivije da *slučajni eksperiment* zovemo **radnja sa neizvjesnim ishodom**. To je svaka radnja koja je barem u teoriji ponovljiva i ima različite moguće ishode koje ne možemo predvidjeti sa potpunom sigurnošću. Jako često navođena ponovljiva radnja sa neizvjesnim ishodom je bacanje igraće kockice sa 6 stranica, mada to može biti i rezultat polaganja ispita (ne morate ga vi ponovo ponavljati, nego je dovoljno to što više studenata izlazi na njega) ili testiranje inteligencije. Dakle, **ishod** (engl. *outcome*) je ono što se dobija kao rezultat takve radnje (npr. dobijete šesticu bacanjem kockice ili šesticu na ispitu, ili ostvarite skor na testu inteligencije od 125).

Skup svih mogućih ishoda ćemo nazivati **ukupnošću ishoda** (engl. *sample space*), a označavaćemo ga velikim grčkim slovom omega (Ω). Neki takvi skupovi su vrlo ograničeni, pa je na primjer za bacanje kockice to skup od samo šest elemenata ($\Omega = \{1, 2, 3, 4, 5, 6\}$), pri čemu "nož" (kada se kockica zadrži na ivici, a ne na jednoj od šest strana) obično izostavljamo, jer ako se on i desi, tražimo od igrača da ponovo baci kockicu.

(Matematika je model stvarnosti, nisam li to napomenuo.) Ukupnost ishoda se može sastojati od samo dva elementa (npr. bacanje novčića) pa tada govorimo o binarnoj slučajnoj varijabli / promjenjivoj. S druge strane može imati i veliki broj ishoda kada koristimo kontinuiranu mjernu skalu. Na primjer, na IQ skali više od 99.8% skorova bi trebalo da bude između 55 i 145, te tad imamo posla sa kontinuiranom slučajnom varijablom.

Konačno, treba da definišimo još jedan pojam koji se navodi u literaturi o teoriji vjerovatnoće. To je **događaj od interesa** (engl. *event*) kojem pripisujemo vjerovatnoću. Taj događaj može da bude jedan određeni ishod za koji smo zainteresovani (npr. 6 na kockici; tada je *outcome* = *event*). Međutim, može da bude i skup izdvojenih ishoda date radnje (npr. ako smo zainteresovani za dobijanje parnog broja onda je događaj od interesa dobijanje bilo kojeg broja iz skupa ishoda 2, 4 i 6).

Kako se vjerovatnoća izražava?

Sad kad znamo neophodne pojmove, možemo preći na numeričko izražavanje vjerovatnoće nekog događaja. Postoji više načina za koje sam uvjeren da ste ih već koristili. To su razlomci, omjeri, statističke proporcije i procenti.

Počinjemo sa primjerima iz **klasične teorije vjerovatnoće** kada poznajemo fizička svojstva predmeta. Najklasičniji primjeri klasične teorije su oni u kojima znamo dvije stvari: tačan broj mogućih ishoda, te to da svaki ishod ima jednaku vjerovatnoću. Kockari su bili oni koji su započeli ozbiljan rad na teoriji vjerovatnoće, pa se vraćamo kockicama (iako su se oni često bavili i kartaškim špilovima). Recimo da ste se dogovorili sa prijateljem da će nas on častiti ručkom koji želimo ako bacajući "poštenu" kockicu na pošten način dobijemo paran broj (događaj od interesa je paran broj, a iza toga je besplatan ručak). U tom slučaju ispravno je da vjerujete da je vjerovatnoća tog događaja $3/6 = 1/2$, jer postoje tri izdvojena ishoda (2, 4, 6) koji se mogu desiti od ukupno šest mogućih koji imaju istu vjerovatnoću. U proporcijama izraženo to znači da je vjerovatnoća jednaka $3/6 = ½ = 0.50$, a u procentima je to 50% ($0.50 * 100$).

Četvrta mjera su **omjeri** (engl. *odds*), koji su kod nas poznati još pod nazivima kvota, šansa, prilike, izgledi, a možda ima čak još neki koji sam izostavio. Nju je najlakše opisati ako imamo i suprotni događaj koji razmatramo. Recimo da ako onaj ko baca klasičnu šestostranu kockicu dobije neparan broj, onda on plaća piće svom prijatelju (i obrnuto). Vjerovatnoća da dobijete paran

broj je jednaka vjerovatnoći da dobijete neparan broj (tj. $3/6 = 3/6$). Ako prvo skratimo nazivnike onda vidimo da je broj ishoda za oba različita događaja 3, pa se to obično navodi 3:3. Izgledi se obično svode na to da saopštite kolika je šansa da se desi događaj od interesa izražen u jedinicama (1) u odnosu na šansu da se dogodi nešto drugo, pa bismo zapravo onda to izrazili kao 1:1 (1 naprema 1).

Ali hajde da vidimo šta se dešava ako nam je prijatelj obećao piće samo ako na kockici dobijemo šesticu. Tada je vjerovatnoća izražena u razlomcima $1/6$, u proporcijama 0.167 (zaokruženo na treću decimalu), u procentima 16.7%, a u omjerima 1:5 budući da postoji 5 drugih opcija koje nam ne odgovaraju. Šta se dešava ako umjesto kockice bacamo novčić (ponovo nož ne računamo)? Konvertibilne marke (valuta u BiH) imaju na jednoj strani broj, a na drugoj neki znak (npr. golub). Ako piće dobijamo kada novčić padne na broj onda vjerovatnoću iskazujemo na sljedeći način: $1/2$, 0.50, 50%, 1:1.

Ako ste prethodno shvatili, onda je to dobar trenutak da stvari zakomplikujemo formalnom notacijom. Za označavanje vjerovatnoće se svašta može naći, ali najčešće se koristi veliko latinično slovo P. Najpopularnije ne znači i najkvalitetnije. Mislim da je daleko bolji prijedlog koristiti Pr (od engl. *probability*) kako bismo razdvojili oznaku za vjerovatnoću od nekih drugih statistika koje ćemo kasnije koristiti. Takođe, potrebno je da znate da se u matematici, pa i teorijskoj i primijenjoj statistici, od četiri gore navedena načina najčešće koriste statističke proporcije,[5] pa se gore navedeno za bacanje novčića može izraziti kao: $Pr(X = broj) = 0.50$ ili za bacanje kockice i dobijanja broja šest: $Pr(X = 6) = 0.167$.

Iz gore navedenog se može zaključiti već da se vjerovatnoća izračunava kao mogućnost određenog događaja u odnosu na ukupan broj mogućih ishoda. Idemo dalje sa objašnjavanjem aksioma vjerovatnoće koje je uspostavio ruski matematičar Andrej Kolmogorov, a sva tri su zapravo laka za shvatanje. Napomena: u ovom kontekstu koristimo proporcije za izražavanje vjerovatnoće.

[5] Na korištenju proporcija se istrajava iako postoji jasni dokazi da su one priličan promašaj. Mada statističke proporcije predstavljaju matematički elegantno rješenje istraživanja pokazuju se da tačniji sudovi lakše donose kada koristimo apsolutne, a ne relativne frekvencije. Vidjeti npr. članak Gigerenzer et al. (2007), Helping doctors and patients make sense of health statistics, https://doi.org/10.1111/j.1539-6053.2008.00033.x Uz to, za laike su postoci mnogo popularniji nego proporcije pa se gore navedeno često čita 50% ili 16.7% vjerovatnoće.

Koja tri pravila ćemo poštovati pri brojčanom izražavanju vjerovatnoća?

Da se podsjetimo, aksiomi su osnovna pravila igre u matematici koje moramo prihvatiti bez dodatnog dokazivanja. Prvi aksiom kaže da svaki pojedinačni mogući ishod (malo omega: ω) mora

imati ne-negativnu vjerovatnoću ($Pr(\omega) \geq 0$). To znači da vjerovatnoća nekog pojedinačnog ishoda mora biti ili 0 ako je taj događaj nemoguć (da dobijete 7 na kockici sa 6 stranica) ili veći od 0, odnosno da mu pridate neku vjerovatnoću.

Drugi aksiom kaže da zbir vjerovatnoća svih ishoda mora biti 1, odnosno 100% ($Pr(\Omega) = 1$). To znači da moramo uglaviti ukupnu količinu uvjerenja na skalu od 0 do 1 i raspodjeliti logički dosljedno tu vjerovatnoću na pojedinačne ishode da bi korištenje matematike imalo ikakvog smisla.

Treći aksiom kaže da vjerovatnoća događaja koji obuhvataju međusobno isključive ishode mora biti jednaka zbiru tih ishoda. Drugim riječima, bacajući jednu kockicu ne možete istovremeno dobiti brojeve i 5 i 6, jer su ti ishodi međusobno isključivi. Iz toga slijedi da je vjerovatnoća događaja kojim bismo dobili ili 5 ili 6 jednaka zbiru vjerovatnoće pojedinačnih ishoda: $Pr(X = 5 \vee 6) = 0.167 + 0.167 = 0.33$. Ako se odlučimo da to pretvorimo u omjere onda dobijamo da je vjerovatnoća jednaka: $2/6 : 4/6 = 2 : 4 = 1 : 2$. To znači da imamo dva puta veća šansu da u jednom bacanju NE dobijemo brojke 5 ili 6 nego da ih dobijemo.

Sve u svemu, vjerovatnoća nekog jednostavnog događaja X se - pojednostavljenom - formulom može izraziti na sljedeći način:

$$Pr(X) = \frac{N_X}{N_\Omega}$$

…gdje bi X predstavljao događaj (npr. paran broj na kockici), N_X predstavljalo broj željenih ishoda (tj. 3 - odnosno stranice kockice 2, 4, 6), a N_Ω broj ukupno mogućih ishoda (tj. 6 - što predstavlja sve stranice kockice).

Kako se izračunavaju "spojene" vjerovatnoće?

Gore navedeni primjer sa sabiranjem vjerovatnoća podrazumijeva da se radi o jednom događaju od interesa koji ima više mogućih, međusobno isključivih ishoda. U tom slučaju koristimo principe **aditivne / zbrajajuće** vjerovatnoće. Na primjer, vjerovatnoća da ćemo pri bacanju šestostrane kockice dobiti broj 5 ili 6 iznosi $P(5 \text{ ili } 6) = Pr(5) + Pr(6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \approx 0.33$.

Aditivna vjerovatnoća

Kada izračunavamo vjerovatnoću dva međusobno nezavisna događaja koristimo principe **multiplikativne / presječne vjerovatnoća**. Aditivna vjerovatnoća se bavila *ili* mogućnostima (ili 5 ili 6), dok se multiplikativna bavi *i* mogućnostima (npr. dobiti 6 na

Multiplikativna vjerovatnoća

jednoj kockici i dobiti 6 na drugoj kockici). Formula presječne vjerovatnoće kada su događaji međusobno nezavisni je:

$$Pr(X\&Y) = Pr(X) * Pr(Y)$$

.....gdje bi $X\&Y$ predstavljalo kombinaciju dva događaja (npr. dvije šestice na dvije kockice), $Pr(X)$ predstavljalo vjerovatnoću jednog događaja, a $Pr(Y)$ vjerovatnoću drugog događaja. Za dobijanje te dvije šestice rezultat bi bio:

$$Pr(6\&6) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36} = 0.028$$

Dobijeni rezultat nam govori da je razumno očekivati da ćemo otprilike u tri od 100 bacanja ($100 * 0.028 = 2.8$) dobiti dvije šestice na dvije kockice. Logiku ovog proračuna nam pokazuje i prikaz ukupnosti ishoda na Slici 11.3 gdje su u prvoj koloni prikazani mogući ishodi bacanja za prvu kockicu, dok su u ostalim kolonama date sve kombinacije koje je moguće dobiti sa obje kockice.

k1	k2=1	k2=2	k2=3	k2=4	k2=5	k2=6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Slika 11.3: Tabela ukupnosti ishoda za bacanje dvije igraće kockice koje imaju po šest stranica.

Ali šta ako događaji nisu međusobno nezavisni?

Dosadašnji primjeri su prikazivali slučajeve kada vjerovatnoća jednog događaja ne stoji u vezi sa vjerovatnoćom drugog događaja. Na primjer, ono što ćemo dobiti na jednoj kockici nije povezano sa onim što ćemo dobiti na drugoj kockici. Međutim, u mnogo slučajeva postoji određena povezanost događaja. Zapravo je to ono što nas najviše zanima u nauci, jer nam ta povezanost pomaže da na osnovu nekog događaja X sa većom vjerovatnoćom predvidimo događaj Y. Već su nam poznati koncepti korelacije / povezanosti, pa na primjer znamo da ako neko

Uslovna vjerovatnoća

ima vrlo visoku inteligenciju i savjesnost, povećava se vjerovatnoća da će taj student imati iznadprosječne ocjene. Vjerovatnoće koje uzimaju u obzir međusobno povezane događaje nazivamo **uslovne vjerovatnoće** (engl. *conditional probabilites*).

Za primjer uslovnih vjerovatnoća ćemo uzeti primjer pola studenata i studiranja psihologije. Na univerzitetu na kojem radim studenti psihologije su većinom djevojke. Taj procenat varira od generacije do generacije, ali se u nekoliko zadnjih godina kreće oko 80% što je zgodno da uzmemo za parametar iz didaktičkih razloga. Dakle, ako bismo nasumično odabrali jednu dvadesetogodišnju osobu iz populacije i dodatno znali da se radi o osobi koja studira psihologiju imamo veću šansu da pogodimo njen pol ako pretpostavimo da se je ta osoba žensko.

Uslovne vjerovatnoće se prikazuju putem sljedeće notacije $Pr(Y|X)$ što se obično čita "vjerovatnoća da se desi događaj Y, ako se već desio događaj X". Zapravo je to bolje čitati kao "vjerovatnoća da se desi događaj Y, ako ZNAMO događaj X".

Koristeći formalnu notaciju proći ćemo sada kroz neke kombinacije, a počećemo sa međusobno nezavisnim događajima. Pol i starost od 20 godina su međusobno nezavisne karakteristike, zato što je vjerovatnoća da je neka slučajno odabrana osoba žensko, podjednaka vjerovatnoći da je osoba uopšte žensko (zanemarićemo razlike koje postoje za starije do 70 godina, gdje imamo nešto više žena).

$$Pr(\check{\text{Z}}|god. = 20) = Pr(\check{\text{Z}}) = 0.50$$

Kao primjer gdje ne važi jednakost između proste i uslovne vjerovatnoće navodio gore prikazani primjer sa brojem žena na studiju psihologije:

$$Pr(\check{\text{Z}}|St_{\psi}) = 0.80 \neq Pr(\check{\text{Z}}) = 0.50$$

Zapravo je gore navedeno pojednostavljeni rezultat sljedeće opšte formule uslovnih vjerovatnoća:

$$Pr(Y|X) = \frac{Pr(X \cap Y)}{Pr(X)}$$

Znam da izgleda komplikovano, ali nije. Znak presjeka skupova $\cap$ znate još iz osnovne škole. To znači da je uslovna vjerovatnoća jednaka količniku presjeka dva događaja i vjerovatnoće događaja za koji smo saznali. Nećemo na ovom mjestu ići u komplikovanu proceduru računanja (ostavićemo to za kasnije),

nego ćemo prikazati sve na najjednostavniji mogući način putem apsolutnih frekvencija. Pretpostavimo da populacija ima milion ljudi (1 000 000) i da je pola njih ženskog pola (500 000). Istovremeno, pretpostavimo da znamo da u ukupnoj populaciji imamo trenutno 500 studenata psihologije, te također da znamo da je 400 njih istovremeno i student psihologije i ženskog pola. U ovom slučaju nam podaci o veličini populacije nije potreban. Formula je jednostavna:

$$Pr(\check{Z}|St_{\psi}) = \frac{N(St_{\psi} \cap \check{Z})}{N(St_{\psi})} = \frac{400}{500} = 0.8$$

Ono što je još bitno da shvatimo prije sljedećeg poglavlja jeste da uslovne vjerovatnoće nisu simetrične, odnosno da po pravilu neće važiti: $Pr(Y|X) = Pr(X|Y)$. I sami možete da razmislite da li je vjerovatnoća da je student psihologije osoba ženskog pola jednaka vjerovatnoći da je osoba ženskog pola student psihologije. Ako to hoćemo da izračunamo onda se moramo poslužiti ranije navedenim podacima o veličini populacije. Vjerovatnoća da je osoba ženskog pola student psihologije je:

$$Pr(St_{\psi}|\check{Z}) = \frac{N(\check{Z} \cap St_{\psi})}{N(\check{Z})} = \frac{400}{500000} = 0.0008$$

…a to svakako nije jednako 0.8. Dakle, ovo bi bio očit dokaz zašto $Pr(St_{\psi}|\check{Z}) \neq Pr(\check{Z}|St_{\psi})$.

Kratko ćemo se još vratiti na kockice kako bismo predstavili neke primjere u kojima uslovna vjerovatnoća drastično mijenja predviđanja. Mogu se činiti banalni, ali veoma su značajni za kasnija razmatranja. Na primjer, vjerovatnoća da dobijemo broj 6 na kockici neće više biti 0.17 ako znamo da je broj koji smo dobili paran broj: $Pr(6|paran) = \frac{1}{3} = 0.33$. S druge strane, ako znamo da smo dobili neparan broj onda je vjerovatnoća da smo dobili šesticu bukvalno nepostojeća: $Pr(6|neparan) = 0$. Konačno, vjerovatnoća da smo dobili šesticu ako znamo da smo dobili broj veći od pet je 1, odnosno to je siguran događaj: $Pr(6|X > 5) = 1$.

Ovim smo postavili odlične temelje da sa čiste teorije vjerovatnoće pređemo u primijenjenu statistiku zaključivanja. Sljedeće poglavlje prikazuje ubjedljivo najpopularnije sredstvo koje ćete sresti u mnogim psihološkim člancima. Što bi se reklo: "must-read". Međutim, prije nego što počnete osigurajte da vam je sve iz ovog poglavlja jasno.

Poglavlje 12

Testiranje nulte hipoteze korištenjem p-vrijednosti

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Zašto je testiranje nulte hipoteze postalo najčešći način ispitivanja istraživačkih hipoteza?
- Šta su to p-vrijednosti?
- Na koji način veličina uzorka i veličina efekta utiču na p-vrijednosti i zašto je to logično?
- Kako se izračunavaju p-vrijednosti za hipoteze koje sadrže kategoričke varijable, dimenzione varijable i/ili njihovu kombinaciju?

U prethodnom poglavlju smo se upoznali sa osnovama teorije vjerovatnoće, kao matematičkog alata koji će nam omogućiti da objektivno suočimo prikupljene podatke i teorijska uvjerenja koje imamo o stvarnosti. Ovim poglavljem ćemo konkretizovati taj postupak. Kao primjer ćemo koristiti hipotezu o efektima slušanja barokne muzike na uspješnost učenja. Da se podsjetimo, na višem teorijskom nivou hipoteza je glasila: "Slušanje instrumentalne barokne muzike sporog tempa tokom učenja doprinosi boljem zapamćivanju gradiva iz statistike." Operacionalizovana hipoteza je glasila: "Studenti 1. godine psihologije koji tokom 30 minuta čitaju tekst o njima nepoznatim statističkim metodama i slušaju odabrana Vivaldijeva, Bachova, Pachelbelova, Corellijeva i Handelova djela u largo tempu (oko 60 taktova u minuti) ostvaruju bolji uspjeh na testu znanja iz datog gradiva neposredno nakon čitanja u odnosu na studente koji čitaju u tišini." Postavlja se pitanje kakvi rezultati istraživanja bi bili dovoljno ubjedljivi pa da se razumni ljudi usaglase da je hipoteza istinita. Drugim riječima, u potrazi smo za standardom kojeg bismo svi poštovali: istraživači koji provode istraživanje, recenzenti koji daju

zeleno svjetlo da se naučni rad objavi, naučna zajednica koju predstavljaju zainteresovani čitaoci koji su eksperti u tom polju, pa i javnost odnosnu čitaoci laici koji nova saznanja mogu na ovaj ili onaj način iskoristiti. Ostatak poglavlja je posvećen predstavljanju takvog standarda, a radi se o testiranju nulte hipoteze putem p-vrijednosti.

Još davno u ovom tekstu smo usvojili postavku da se u psihologiji bavimo probabilističkim hipotezama, pa ćemo za početak hipotezu blago modifikovati i dodati "…ostvaruju **u prosjeku** bolji uspjeh…". Time smo omogućili prebacivanje hipoteze u statističku ravan gdje je parametar za koji smo zainteresovani razlika hipotetičkih aritmetičkih sredina (δ) između onih koji bi učili uz muziku i onih koji bi učili u tišini. Novi problem je to što smo time dobili beskonačno mnogo statistički iskazanih hipoteza razlike, jer δ (parametar razlike) može biti bilo gdje u rasponu od vrlo malih vrijednosti koje su > 0 pa sve do $+\infty$. Na prvu se čini da je taj problem nerješiv, jer je nemoguće ispitati beskonačan broj hipoteza.

Elegantno rješenje ovog problema je ponudio Ronald Aylmer Fisher (1890-1962), statističar i genetički biolog. U svojim radovima objavljenim od 1920-ih do 1950-ih Fisher je odlučio da obrne perspektivu predlažući testiranje nulte hipoteze kao osnovni alat empirijskog rezonovanja. Upitao se: "Umjesto da pokazujemo da neki efekat postoji i biramo iz beskonačnog broja mogućih hipoteza koje govore o tom efektu, zar ne bi bilo dovoljno samo da pokažemo da postoji banalno nisko vjerovatnoća da efekta uopšte nema?". Istraživač koji želi pokazati da neki efekat (razlika ili povezanost) postoji u populaciji bi trebalo zapravo da nam pokaže da je suprotna hipoteza - hipoteza koja tvrdi da ta razlika ili povezanost NE postoji! - toliko slabo vjerovatna da izgleda apsurdna. Drugim riječima, budući da je teško na osnovu jednog istraživanja pretpostaviti tačan parametar koji postoji u populaciji, ono što možemo razumno uraditi je da pokažemo da podaci sa uzorka jasno govore protiv nulte hipoteze.

Princip testiranja nulte hipoteze je vrlo sličan Karl Popperovom principu opovrgljivosti / falsifikabilnosti koji podrazumijeva da predmet nauke mogu biti samo empirijski opovrgljive hipoteze, kao i da se istinitost naučne hipoteze nikada ne može potpuno dokazati, nego da se samo može uvjerljivo ukazati na to da je ona netačna. Ključna razlika je što Popper stavlja stalni naglasak na pokušaje odbacivanja istraživačke hipoteze, dok Fisher u fokus stavlja nultu hipotezu. Pošteno je navesti i da je Fisher objavio svoj prijedlog prije Poppera, odnosno da ga nije namjerno distorzirao.

Dakle, **nultom hipotezom** (engl. *null hypothesis*) se tvrdi da je sistemsko dejstvo ravno nuli $\delta = 0$ (za hipoteze korelacije to bi bilo $\rho = 0$).[1] Ako ta hipoteza oslikava stvarnost onda se eventualne razlike na uzorku mogu objasniti tzv. greškom uzorkovanja. **Greška uzorkovanja** (engl. *sampling error*) proizilazi iz toga što u istraživanju imamo baš te ispitanike koji su slučajno odabrani, a ne čitavu populaciju.

[1] Vidjećemo da je tehnički, ali i u praksi, nekad moguće definisati nultu hipotezu (engl. null hypothesis) tako da zastupa druge vrijednosti, pa se posebno označava hipoteza nule (engl. nill hypothesis) za koju pretpostavljeni parametar mora biti jednak tačno 0.

Šta je tajna "uspjeha" nulte hipoteze?

Ogromna prednost nulte hipoteze jeste što se statistički ona lako definiše, a zahvaljujući teoriji vjerovatnoće i postojećim matematičkim distribucijama, vidjećemo da je moguće izračunati konkretne vjerovatnoće za različite rezultate dobijene na uzorku kada je pretpostavljena nulta hipoteza tačna u populaciji. Zahvaljujući tim vjerovatnoćama, moguće je donijeti i odluku da li je ta hipoteza "apsurdna". Kako je hipoteza o efektima barokne muzike nešto složenija jer imamo dvije populacije, tehničke principe testiranja nulte hipoteze ćemo prvo predstaviti na jednostavnijem primjeru.

Nema jednostavnijeg nacrta od onoga sa jednom binarnom kategoričkom varijablom, pa se vraćamo na primjer polnog sastava studenata psihologije. Recimo da smo željeli ispitati hipotezu da psihologiju češće studiraju žene nego muškarci. Shodno tome, nulta hipoteza koju bismo provjeravili bi glasila: "Psihologiju studira jednak broj žena i muškaraca". Prije nego što krenemo sa testiranjem ovakve nulte hipoteze, naglasićemo nekoliko stvari. Kao prvo, ova hipotezu se optimalno provjerava analizom spiskova upisanih studenata, međutim mi ćemo koristiti analizu na uzorku iz didaktičkih razloga. Drugo, potrebno je naglasiti da ovu hipotezu moramo ograničiti vremenski i geografski, pa bismo je vezali za aktuelno stanje na području država bivše Jugoslavije (npr. pretpostavljamo da je trenutno u Avganistanu drugačija polna raspodjela studenata psihologije, a i kod nas se kroz istoriju mijenjao polni sastav studenata). Treće, ovdje uzimamo u obzir samo dva pola (muški i ženski), odnosno ne uzimamo u obzir nebinarne rodne odrednice ili nedostajuće podatke o polu.

Koji konkretne korake poduzimamo pri testiranju nulte hipoteze?

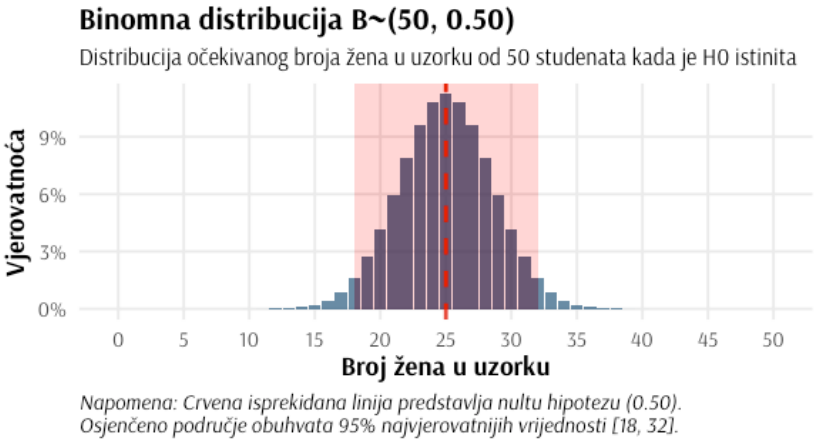
Može se reći da algoritam testiranja nulte hipoteze obuhvata četiri koraka od kojih će svaki - uključujući i nove pojmove koji se pominju - biti podrobno opisan u nastavku:

1. Statističko definisanje nulte hipoteze (H_0)
2. Kreiranje hipotetičke distribucije relevantnog statistika (rezultata sa uzorka) ako je H_0 istinita u populaciji
3. Izračunavanje $p - vrijednosti$ za dati statistik
4. Donošenje odluke o H_0 na osnovu poređenja $p - vrijednosti$ i nivoa statističke značajnosti (α)

Statistička operacionalizacija nulte hipoteze predstavlja prvi korak u njenom testiranju. U našem primjeru bi se nulta hipoteza statističkom notacijom izrazila ovako: $H_0 : \pi(\check{\text{Z}}) = 0.50$, ali imajte na umu da zapravo podrazumijevamo da se hipoteza odnosi na studij psihologije pa bi kompletan izraz bio: $(\pi(\check{\text{Z}}|St_{_\psi}) = 0.50)$. Iz postavljene hipoteze, važi i sljedeće: $H_0 : \pi(\text{M}) = 0.50$, odnosno $H_0 : \pi(\check{\text{Z}}) - \pi(\text{M}) = 0$. Bitno je da shvatite da postoji više načina kako statistički operacionalizovati jednu te istu hipotezu, budući da ćete i u literaturi sretati različite formate.

U drugom koraku moramo kreirati hipotetičku distribuciju relevantnog statistika (čitaj: rezultata sa uzorka) onda kada je nulta hipoteza istinita u populaciji. Dakle, ono što se traži od nas kao istraživača je da zamislimo svijet u kojem studij psihologije zaista upisuje jednak broj žena i muškaraca. Za takav svijet je moguće opisati distribuciju očekivanih rezultata na uzorcima. **Binomna distribucija** (engl. *binomial distribution*) je matematička distribucija koja to za nas radi kada je varijabla od interesa binarna / dihotomna. Konkretna binomna distribucija se definiše na osnovu pretpostavljenog parametra zastupljenosti kategoriju u populaciji (tj. statistička proporcija žena) i broja ispitanika u uzorku. Dakle, u našem primjeru gdje pretpostavljamo jednak broj muškaraca i žena na studiju psihologije parametar proporcije je $\pi = 0.50$, a broj ispitanika određujemo na osnovu toga koliko nam je zaista velik uzorak. Budući da na mom studijskom programu prvu godinu obično upisuje 50 studenata, uzećemo u ovom primjeru da je $n = 50$. Statističkom notacijom se binomna distribucija označava na sljedeći način: $X \sim B(n, \pi)$. U konkretnom primjeru imamo već vrijednosti populacione proporcije i veličine uzorka, pa bi to bila binomna distribucija: $X \sim B(50, 0.50)$. Ta distribucija je prikazana na Slici 12.1.

Šta možemo da uvidimo sa Slike 12.1? Na primjer, vidimo da ako zaista jednak broj muškaraca i žena upisuje psihologiju, da će od svih mogućih rezultata na uzorku najvjerovatniji rezultat biti da u uzorku bude 25 djevojaka (i 25 mladića). Međutim, istovremeno vidimo da taj rezultat nije pretjerano vjerovatan u apsolutnom smislu, odnosno možemo očekivati da ga dobijemo otprilike u tek jednom od devet slučajeva ($Pr(\check{\text{Z}} = 25) = 0.112$). Jedva nešto malo niža vjerovatnoća je da ćemo u uzorku imati 24 ili 26 djevojaka ($Pr(\check{\text{Z}} = 26) = 0.108$), još nešto manja da će ih biti 23 ili 27 ($Pr(\check{\text{Z}} = 27) = 0.096$), a vjerovatnoće sve više opadaju što se više udaljavamo od broja 25, odnosno od tačke 50%. Konačno, bitno je da uvidimo da nam binomna distribucija omogućava da za svaki mogući ishod na uzorku - a to znači od 0 do 50 - možemo dobiti vjerovatnoću njegovog javljanja u tom



Slika 12.1: Binomna distribucija B~(50, 0.50).

zamišljenom svijetu.[2]

[2] Sjećate li se koliko mora da iznosi zbir vjerovatnoća svih mogućih ishoda? Naravno: 100%, odnosno 1.0 iskazano statističkim proporcijama.

U trećem koraku se upoređuje distribuciju rezultata iz zamišljenog svijeta sa podatkom koji smo uočili u stvarnom svijetu (statistik), odnosno u našem istraživanju. Potrebno je da prosudimo koliko li su ta dva svijeta saglasna, a alat koji koristimo jeste vjerovatnoća. Kao što smo već upoznali Fisherov prijedlog, ako je vjerovatnoća suviše mala da podatak iz stvarnog svijeta dolazi iz zamišljenog svijeta nulte hipoteze, onda bismo uvjerenje da je nulta hipoteza istinita morali oslabiti ili čak odbaciti. Ono što još nije razjašnjeno jeste o kojoj vrsti događaja od interesa govorimo i koliko niska ta vjerovatnoća mora biti pa da nulta hipoteza bude odbačena (četvrti korak).

Slika 12.1 već prikazuje vjerovatnoće svih pojedinačnih ishoda[3]. Gore sam prikazao i konkretne vrijednosti ako se na uzorku opazi 25, 24 ili 26, te 23 ili 27 djevojaka (istim redom: 0.112, 0.108, 0.096). Takođe, primijetili smo da je svaka od ovih pojedinačnih vjerovatnoća bila relativno mala. Šta se dešava ako kombinujemo ove vjerovatnoće? Šansa da dobijemo između 23 i 27 žena u našem uzorku — pod pretpostavkom da je naša nulta hipoteza o jednakoj polnoj distribuciji u populaciji studenata psihologije istinita — iznosi 0.52, odnosno približno 50%. To znači da praktično u svakom drugom istraživanju koje koristi slučajni uzorak od 50 učesnika iz populacije sa zaista jednakim brojem muškaraca i žena, možemo očekivati da vidimo u njemu bude između 23 i 27 žena.

[3] Sjećate li se razlika između ishoda i događaja?

Šta su te p-vrijednosti i kako ih izračunavamo?

Definicija p-vrijednosti

Ali šta ako u uzorku prebrojimo 28 žena? Na ovom mjestu možemo konačno uvesti pojam p-vrijednosti. Ta **p-vrijednost** (engl. *p-value*) je posebna vrsta vjerovatnoće koja se definiše kao **vjerovatnoća da se (na uzorku iste veličine) dobije data ili ekstremnija vrijednost statistika ako je nulta hipoteza tačna u populaciji**. Malo drugačijim riječima rečeno, a kako bi se ova bitna definicija urezala u vašu neuronsku mrežu: p-vrijednost je uslovna vjerovatnoća da se u istraživanju dobije toliki ili ekstremniji rezultat ako je nulta hipoteza zaista istinita.[4]

[4] Usput, evo zašto smo se odlučili da koristimo Pr kao opštu oznaku za vjerovatnoću: p-vrijednost se u literaturi označava najčešće sa malim p, ali i sa velikim P, dok ćete ponegdje vidjeti i *Sig.* od engleske riječi *significance* što znači značajnost.

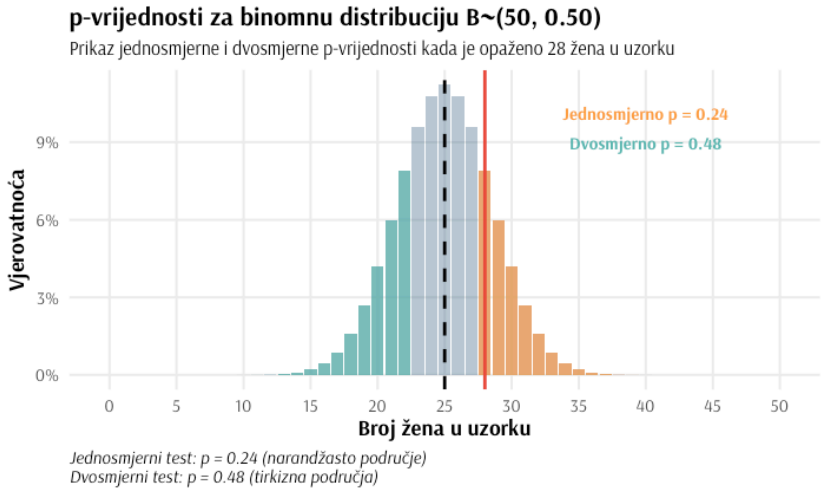
Također, primijetite da se u navedenoj definiciji pominje riječ ekstremniji. Ovdje ekstremniji znači udaljeniji od rezultata koji su najvjerovatniji, odnosno rezultati bliži krajevima distribucije. Ako pogledamo Sliku 12.2, vidjećemo da su ekstremnije vrijednosti $29, 30, 31...48, 49, 50$, tj. one koje se kreću ka krajevima distribucije od najvjerovatnije vrijednosti 25.

Idemo sad da izračunamo konkretnu p-vrijednost za $X = 28$, odnosno 56.0% žena na uzorku od 50 ispitanika. Ako znamo da se u intervalu između 23 i 27 nalazi 52% date binomne distribucije, onda to automatski znači da je ostalih 48% van tog intervala ($100\% - 52\% = 48\%$). Znajući da je ova binomna distribucija simetrična[5], 24% distribucije se odnosi na vrijednosti $\hat{\pi} \geq 28$, a drugih 24% distribucije na lijevi kraj, gdje je $\hat{\pi} \leq 22$. Slika 12.2 prikazuje da se mogu izračunati dvije različite p-vrijednosti: $p = .48$ (tzv. dvosmjerni test) i $p = .24$ (tzv. jednosmjerni test). Imajte u vidu da s obzirom na to da su p-vrijednosti omeđene vrijednostima između 0 i 1, obično se ne navodi brojka 0 sa lijeve strane decimalnog znaka.

[5] Binomne distribucije mogu itekako biti asimetrične što će biti prikazano kasnije u tekstu!

Avaj, zašto sada dvije p-vrijednosti, zar situacija nije već dovoljno komplikovana? Pojednostavićemo zato stvari. Gotovo uvijek se saopštavaju **dvostrane** ili **dvosmjerne** (engl. *two-tailed*) p-vrijednosti koje uzimaju u obzir oba kraja distribucije. Razlog za to je to što je po svojoj suštini opovrgavanje nulte hipoteze neusmjereno. Naime, ako već prihvatamo nultu hipotezu kao startnu epistemičku poziciju - odnosno, pravimo se da "nemamo pojma" - onda je principijelno da se ponašamo tako kao da zaista ne znamo u kojem smjeru postoje razlike. **Jednostrane** (engl. *one-tailed*) p-vrijednosti se mogu dobiti u softveru, ali je vrlo teško odbraniti njihovu upotrebu, čak i kada ste uvjereni da je istraživačka hipoteza tačna (o principijelnijim pristupima statističkom zaključivanju kasnije u tekstu).

Ovdje možemo predstaviti i zaguljenu matematičku definiciju p-



Slika 12.2: p-vrijednosti za binomnu distribuciju $B(50, 0.50)$ kada je X = 28.

vrijednosti koja samo izražava gore navedenu značenjsku definiciju:

$$p = Pr(|\hat{\theta}| \geq |\hat{\theta}_X| \mid \theta)$$

U navedenoj formuli imamo tri jajeta koji liče jedno na drugo, ali ne označavaju isto:

- $\hat{\theta}_X$ predstavlja statistik dobijen na uzorku,
- $\hat{\theta}$ predstavlja različite moguće vrijednosti statistika,
- θ predstavlja pretpostavljenu vrijednost parametra.

U našem slučaju je $\theta = 0.50$ ako koristimo proporciju žena kao parametar, a može biti i $\theta = 0$ ako bismo koristili razliku između proporcije žena i muškaraca. Takođe, u formuli su navedene apsolutne vrijednosti kako bi se ukazalo na dvosmjernu prirodu p-vrijednosti.

Kako odlučujemo da li da formalno odbacimo nultu hipotezu usljed njene slabe vjerovatnoće?

U redu, ali da li je $p = .48$ dovoljan dokaz u prilog alternativne hipoteze? Nije, jer rezultat koji se pojavljuje u skoro polovini istraživanja na populaciji u kojoj je nulta hipoteza istinita nije posebno neobičan. Potrebne su nam dosta niže vjerovatnoće događaja, pa da ga proglasimo čudnim. Koliko bi niska ta vjerovatnoća morala biti je još uvijek tema velikih rasprava unutar statistike i nauke, te ćemo se tom pitanju vraćati u sljedećim poglavljima. Na ovom mjestu je dovoljno da zapamtimo da

Nivo statističke značajnosti

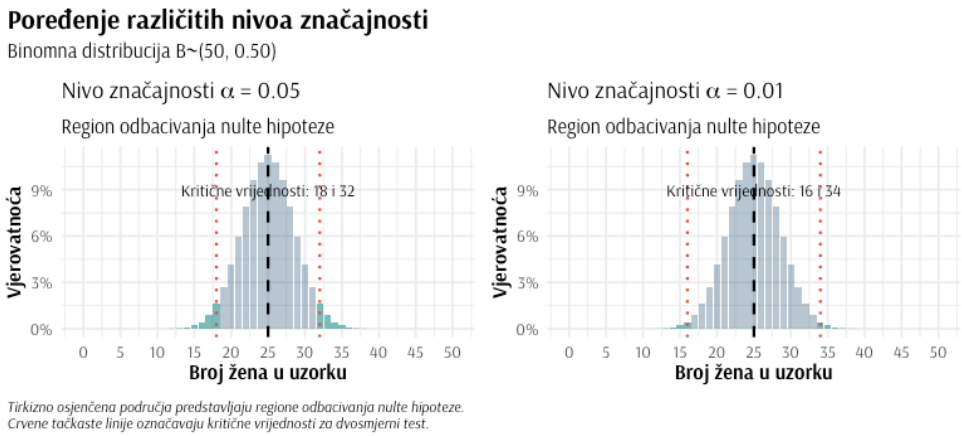
se – zahvaljujući prvenstveno gore pomenutom R.A. Fisheru – u većini oblasti ona ustoličila na 5%, odnosno 0.05 izraženo statističkim proporcijama. Dakle, vjerovatnoća dobijanja datog ili ekstremnijeg statistika mora biti manja od 5% kako bi se rezultat istraživanja smatrao formalnim dokazom protiv nulte hipoteze. Predefinisana granica vjerovatnoće koja se koristi za u ovoj proceduri se naziva **nivo značajnosti** (engl. *significance level*), a u notaciji se označava grčkim slovom alfa (α). Ako je $p \geq \alpha$ – a u našem slučaju jeste jer je $0.48 > 0.05$ – onda nultu hipotezu ne bi trebalo odbaciti. Ona ostaje dovoljno vjerovatna, mada to ne znači automatski da je i istinita. S druge strane, ako je $p < \alpha$ onda govorimo o **statistički značajnom** (engl. *statistically significant*) rezultatu, te bi se u datom postupku odbacila nulta hipoteza. Primijetite da se radi o robotizovani algoritmu "ako -> onda" tipa koji se izvršava bez pogovora.

Da vidimo sada šta se dešava ako imamo 40 žena na uzorku od 50 studenata, što je obično slučaj na studijskom programu na kojem radim. Uz pomoć binomne distribucije dobijamo da je $p = 0.00002386$. To znači da je vjerovatnoća da u uzorku imamo 40 ili više žena – kao i 10 ili manje prateći logiku dvosmjernog testiranja — daleko manja ne samo od 0.05 nego i od jednog promila ($p < .001$). To se već smatra dokazom u prilog odbacivanja nulte hipoteze. Slika 12.3 pokazuje kritične vrijednosti (engl. *critical values*) statistika za koje bismo odbacili nultu hipotezu na nivoima značajnosti $\alpha = .05$ i $\alpha = .01$. Obrnutom logikom bismo onda zaključili da je to istovremeno dokaz u prilog alternativne hipoteze. Sve u svemu, dobijena p-vrijednost nam govori da postoji apsurdno mala vjerovatnoća da su podaci proistekli iz populacije u kojoj jednak broj muškaraca i žena studira psihologiju, te da bi bilo racionalno vjerovati u to da žene češće studiraju psihologiju. Primijetite da smo time dobili odgovor na jednu vrlo uopštenu hipotezu.

Zadržaćemo se još na primjerima binomne distribucije kako bismo zacementirali još dvije bitne pravilnosti o p-vrijednostima. One se mogu iskaziti sljedećim izrazom: $p \propto \frac{1}{\delta \times n}$, koji se čita: p-vrijednost je obrnuto srazmjerna veličini efekta (apsolutnom odstupanju od nule bilo da je ona razlika ili povezanost) i veličini uzorka.

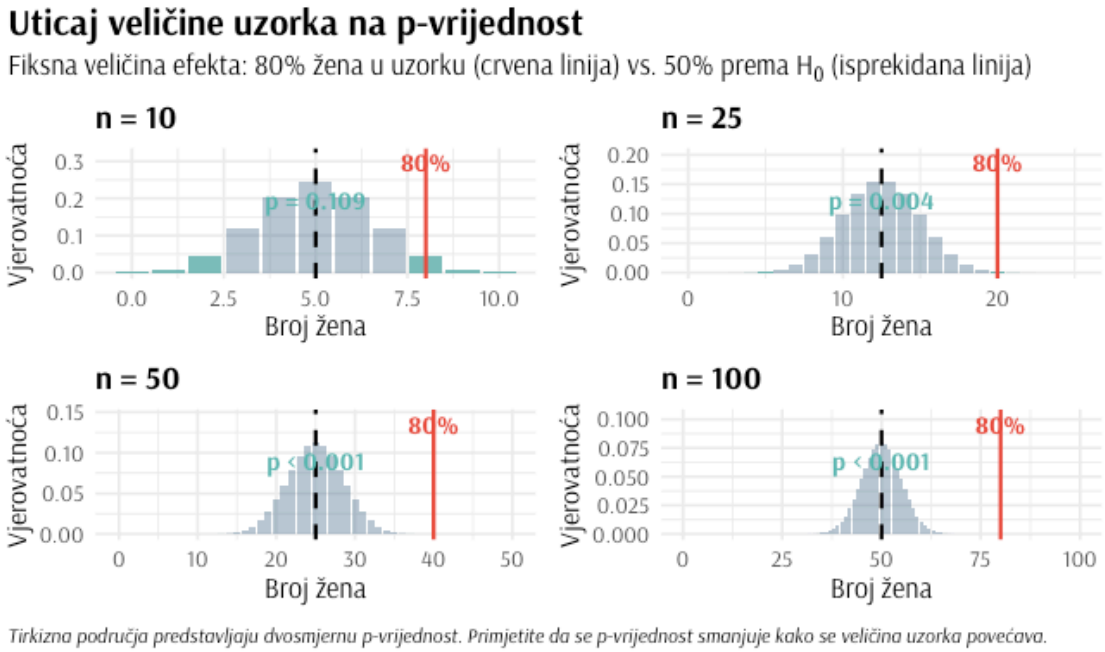
Kako veličina uzorka utiče na p-vrijednosti?

Na Slici 12.4 vidimo efekat povećanja veličine uzorka kada se veličina efekta drži konstantnom. U svim prikazanim situacijama imamo 80% žena u uzorku, a uzorci sadrže 10, 25, 50 i 100



Slika 12.3: Kritične vrijednosti za binomnu distribuciju B~(50, 0.50) za nivoe značajnosti .05 i .01.

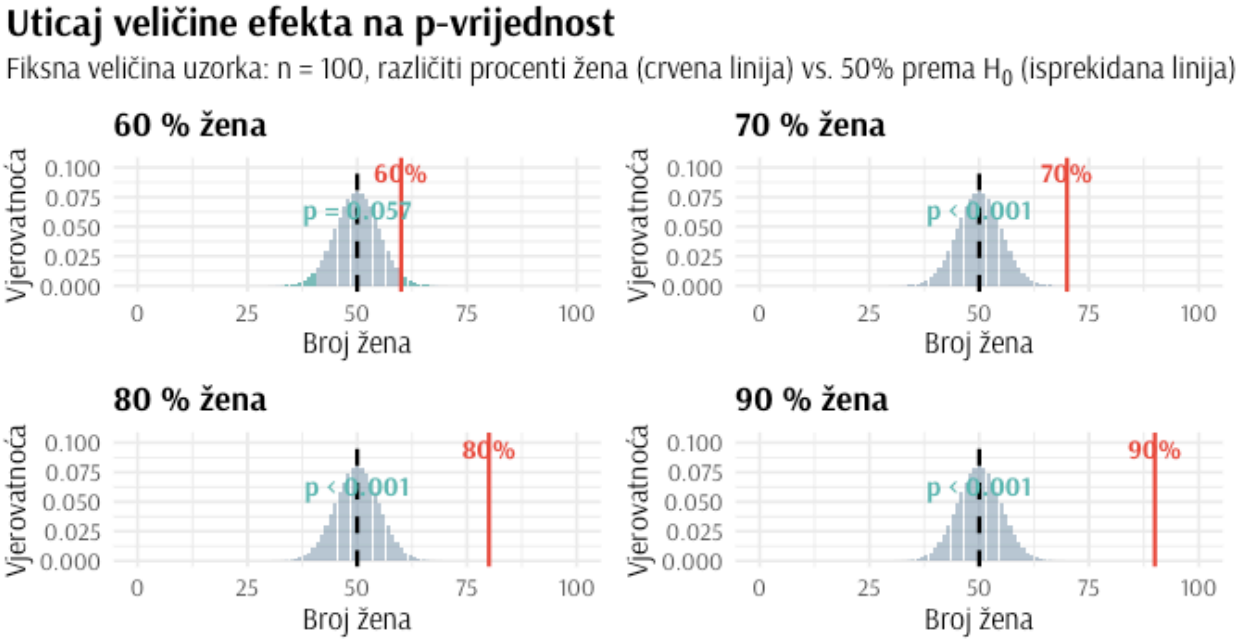
ispitanika. Kao što možete vidjeti *p*-vrijednost se sa povećanjem uzorka smanjuje, odnosno vjerovatnoća da se takav ili ekstremniji rezultat desi je sve manja ako je nulta hipoteza istinita u populaciji. Ova pravilnost potpuno odgovara načelima koje smo već naučili: što je uzorak veći, on teži da tačnije predstavlja populaciju. Na uzorku od 10 ispitanika postoje veće šanse da ćemo dobiti veća odstupanja od nule kao plod greške uzorkovanja, manje je vjerovatno da će se to desiti kada imamo 100 ispitanika, a još manje kada imamo 1000.



Slika 12.4: Dejstvo veličine uzorka na *p*-vrijednosti dobijene putem binomnih distribucija.

Kako veličina efekta utiče na p-vrijednosti?

Slika 12.5 prikazuje dejstvo veličine efekta na p-vrijednosti na uzorcima od kojih svaki ima 100 ispitanika. Ono što je varirano jeste procenat žena na uzorku: 60%, 70%, 80% i 90%. I tu prikazani rezultati su potpuno logični: snažniji dokazi protiv nulte hipoteze se dobijaju sa sve većim veličinama efekata. Sve u svemu, i veličina efekta i veličina uzorka utiču na p-vrijednosti, a kasnije ćemo vidjeti da uticaj mogu izvršiti još neke karakteristike podataka i analiza.

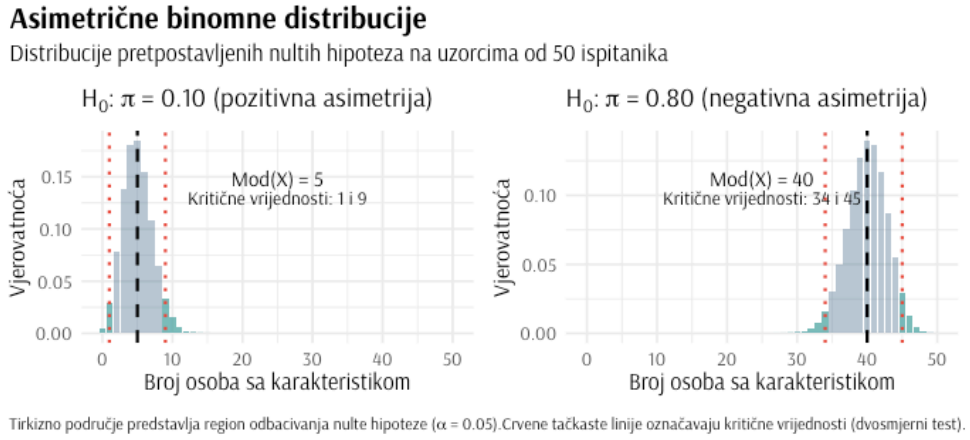


Slika 12.5: Dejstvo veličine efekta na p-vrijednosti dobijene putem binomnih distribucija.

Kako izgledaju asimetrične binomne distribucije?

Izlaganje o binomnim distribucijama ne bi bilo kompletno ako ne bismo pokazali i njihovu raznolikost po pitanju populacionog parametra. Naime, nije nužno da nulta hipoteza bude takva da u populaciji imamo jednak udio dvije kategorije. Na primjer, distribucija psihičkih poremećaja je pozitivno asimetrična, pa tako obično očekujemo da manje od 50% neke populacije ima simptomatologiju izraženu u klinički relevantnoj mjeri. Alternativnom hipotezom prevalencije možemo tvrditi da je više od 10% studenata klinički anksiozno. Ako za nultu hipotezu postavimo vrijednost 10% ($H_0 : \pi(Anx) = 0.10$) onda možemo tu hipotezu provjeriti podacima, odnosno potencijalno je opovrgnuti. S

druge strane, određene varijable se distribuiraju negativno asimetrično, pa se može podrazumijevati da je kategorija onih koji zadovoljavaju neki kriterijum daleko iznad 50%. Na primjer, može se ispitati hipoteza kojom se tvrdi da je više od 80% studenata zadovoljno studiranjem na datom univerzitetu. Tada će nulta hipoteza glasiti: ($H_0 : \pi(Zad) = 0.80$). Na Slici 12.6 vidimo kako izgledaju pretpostavljene distribucije rezultata na uzorcima ako bi date nulte hipoteze bile tačne. I za ovako asimetrične distribucije možemo da postavimo nivoe značajnosti, izračunamo kritične vrijednosti statistika i izračunamo konkretne p-vrijednosti.



Slika 12.6: Primjeri asimetričnih binomnih distribucija.

Kako testiramo nulte hipoteze za dimenzione varijable?

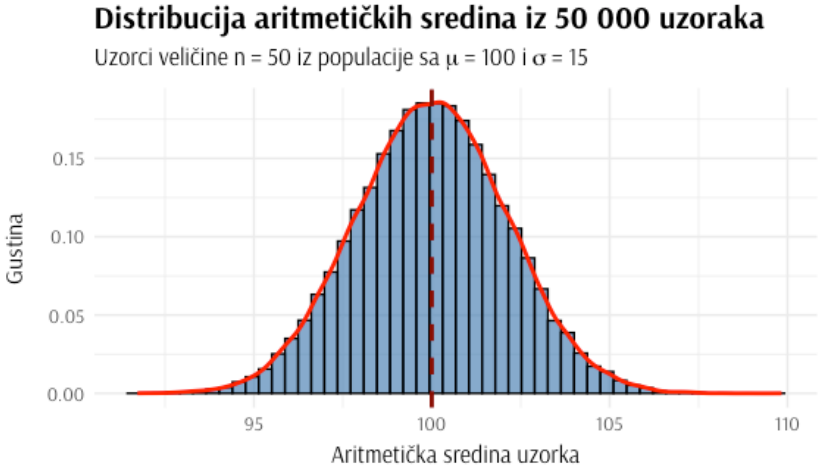
Dosad je prikazano računanje p-vrijednosti za kategoričke varijable. Postupak testiranja dimenzionih varijabli je analogan dosad prikazanom. Razlika se sastoji samo u tome što nećemo koristiti prekidne distribucije[6] kao što je binomna, nego ćemo se poslužiti kontinuiranim matematičkim distribucijama. Najnormalnije je da za primjer uzmemo normalnu distribuciju.

[6] Sjećate li se, zvali smo ih još tačkastim i diskretnim?

Pretpostavimo da na uzorku od 50 studenata želimo ispitati hipotezu da su studenti psihologije u prosjeku inteligentniji od opšte populacije. Za tu svrhu bismo koristili neki standardizovani test inteligencije čiji skorovi su - kao što znamo iz ranijeg poglavlja - normalno distribuisani u opštoj populaciji i čija je aritmetička sredina jednaka 100 izražena u IQ jedinicama (μ_O). Nultu hipotezu ćemo statistički definisati na sljedeći način:

$H_0 : \mu_\psi = 100$ ili $H_0 : \mu_O - \mu_\psi = 0$.

U drugom koraku je potrebno kreirati hipotetičku distribuciju aritmetičkih sredina na slučajno odabranim uzorcima od 50 ispitanika iz opšte populacije. Kako li će izgledati takva distribucija? Jedan od načina da to saznamo jeste da uz pomoć R-a simuliramo veliki broj aritmetičkih sredina na uzorcima od 50 ispitanika koji dolaze iz opšte populacije za koju su nam poznate vrijednosti aritmetičke sredine ($\mu = 100$), ali i standardne devijacije ($\sigma = 15$). Slika 12.7 prikazuje takvu distribuciju dobijenu nakon što je izabrano 50 000 uzoraka koji imaju po 50 ispitanika, te nakon što je za svaki uzorak izračunata aritmetička sredina. Ne bi trebalo da vas čudi što je ta distribucija također normalna, budući da poznajete **centralnu graničnu teoremu**.



Slika 12.7: Raspodjela aritmetičkih sredina uzoraka n = 50, za $\mu = 100$ i $\sigma = 15$.

Ova normalna distribucija ima još neke svoje pravilnosti. Za razliku od populacione raspodjele sirovih skorova čija je standardna devijacija jednaka 15, ova normalna distribucija je značajno uža. Ona se može izračunati putem sljedeće formule:

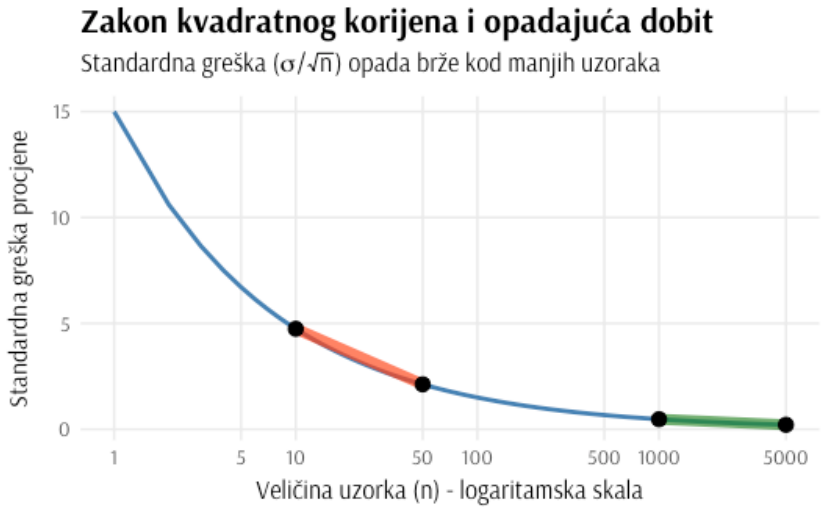
$$SD_M = SE_M = \frac{\sigma_X}{\sqrt{n}} = \frac{15}{\sqrt{50}} = \frac{15}{7.07} = 2.12$$

Budući da se radi o standardnoj devijaciji, to znači da je 2.12 prosječno očekivano odstupanje aritmetičkih sredina od vrijednosti 100 za uzorke od 50 ispitanika. Ovu standardnu devijaciju zovemo posebnim imenom: standardna greška aritmetičke sredine. U literaturi ćete vidjeti i oznake (SG_M od standardna greška, odnosno SE_M od engl. *standard error*). Dakle, **standardna greška** je mjera očekivanog odstupanja statistika od parametra usljed greške uzorkovanja. Standardna greška će biti

Definicija standardne greške

sve veća što je varijabilnost na uzorku veća, ali će se i smanjivati sa povećanjem uzorka. I ovo je potpuno logično s obzirom da smo ranije naglasili da što je raznolikost u grupi veća (koju operacionalizujemo putem standardne devijacije) to su mjere centralne tendencije slabiji predstavnici čitave grupe. Ako bi svi članovi populacije bili jednaki, njihova standardna devijacija bi bila jednaka 0. Aritmetička sredina bi tada savršeno predstavljala čitavu grupu, a na svim uzorcima bismo dobijali jednaku aritmetičku sredinu, odnosno statistik bi bio savršena procjena parametra ($SG_M = 0$).

Efekat veličine uzorka je također logičan. Što je veći uzorak to procjena parametra teži da bude tačnija, odnosno očekivana greška procjene se smanjuje. U nazivniku se ipak ne nalazi "gola" veličina uzorka, nego njen kvadratni korijen. Taj matematički izraz $\frac{1}{\sqrt{n}}$ je dobio ni manje ni više nego status "zakona"; naziva se **zakon kvadratnog korijena** (engl. *the square-root law*) i opisuje pravilnost opadajuće dobiti. Drugim riječima, povećanje malih uzoraka ima veći relativni uticaj na povećanje preciznosti procjene; npr. porast sa 10 na 50 ispitanika ima veći relativan uticaj na preciznost procjene nego kada uzorak povećavamo sa 1000 na 5000 (Slika 12.8).



Slika 12.8: Stopa povećanja preciznosti procjene u zavisnosti od veličine uzorka.

Elem, to što znamo da imamo posla sa normalnom distribucijom nam omogućava da iskoristimo njene razne pravilnosti pri izračunavanju p-vrijednosti. Na primjer, od ranije znamo da se unutar opsega od -2SD do +2SD nalazi otprilike 95% populacije pod normalnom krivom [konkretna brojka je 1.96]. Time istovremeno podrazumijevamo da je van tog opsega 5% distribucije. To

znači da možemo da izračunamo konkretne kritične ili granične vrijednosti na IQ skali za $\alpha = .05$.

$$\begin{aligned} KV_{1,2} &= 100 \pm 1.96 \cdot 2.12 \\ &= 100 \pm 4.16 \\ \Rightarrow \quad KV_1 &= 95.84, \quad KV_2 = 104.16 \end{aligned}$$

Iz toga slijedi da svaka artimetička sredina sa uzorka koja je manja od 95.84, odnosno veća od 104.16, ima vjerovatnoću nižu od 5% da se desi ako je nulta hipoteza u populaciji istinita. Naravno, za svaku konkretnu aritmetičku sredinu možemo izračunati odgovarajuću p-vrijednost. Međutim, umjesto da koristimo sirove skorove koji se razlikuju od istraživanja do istraživanja u zavisnosti od mjerne skale određenog instrumenta, možemo koristiti standardizovanu normalnu distribuciju putem već poznate formule za z-skorove. U novoj formuli ćemo sirovi skor (X) zamijeniti aritmetičkom sredinom uzorka:

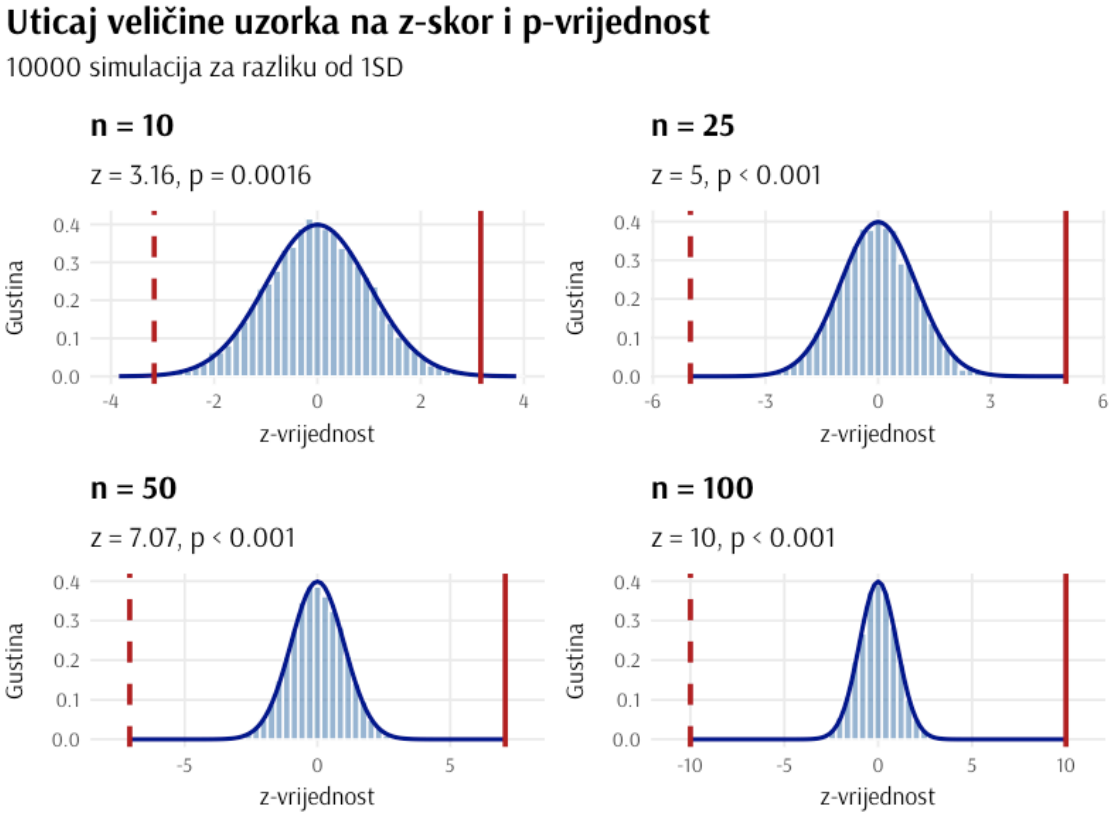
$$z = \frac{X - M}{SD} \Rightarrow z = \frac{M - \mu_0}{SD_M}$$

Tako ćemo u našem slučaju za vrijednost $M = 100$ dobiti $z = 0$, dok bi vrijednost $M = 102.12$ bila jednaka standardizovanoj vrijednosti $z = 1$ budući da je $SD_M = 2.12$. Transformacija u standardizovane skorove nam omogućava lakše snalaženje sa vjerovatnoćama. Kao što bi vam već trebalo biti poznato iz ranijeg poglavlja o normalnoj distribuciji, ako je $z = 1.96$ to znači da je $p = .05$ (dvosmjerni test). Nadalje, korištenjem tablica o površini ispod normalne krive možemo otkriti da kada je $z = 1$ onda je $p \approx 0.32$ zato što se otprilike 68% površine pod normalnom krivom nalazi između $-1z$ i $+1z$). Analogno tome, za $z = 3$ je odgovarajuće $p = .003$. Sve ovo navodim zato što se ovakva provjera nulte hipoteze u statističkoj literaturi vezuje za naziv z-test za koji statistički softver izbacuje rezultate u standardizovanim skorovima.

Dakle, statistički smo definisali nultu hipotezu i izgradili smo distribuciju vjerovatnoća aritmetičkih sredina na uzorcima kada je nulta hipoteza istinita. Naredni korak je izračunati vjerovatnoću dobijanja određene aritmetičke sredine ili ekstremnijeg rezultata na uzorku ako je nulta hipoteza istinita. Recimo da smo na uzorku dobili vrijednost $M_\psi = 115$. Na osnovu gore navedenih formula, odgovarajući z-skor bi bio:

$$z = \frac{M_\psi - \mu_0}{\frac{\sigma}{\sqrt{n}}} \Rightarrow z = \frac{115 - 100}{\frac{15}{\sqrt{50}}} = \frac{15}{2.12} = 7.08$$

Dobijeni z-skor odgovara krajnje niskoj p-vrijednosti ($p < .0001$), što znači da bismo definitivno odbacili nultu hipotezu. Kao i u slučaju binomnih distribucija, a da bismo bolje utvrdili gradivo izvešćemo i ovdje variranje veličine uzorka i veličine efekta. Slika 12.9 pokazuje razlike u z i p-vrijednostima kada uzorak variramo tako što imamo 10, 25, 50 i 100 ispitanika.

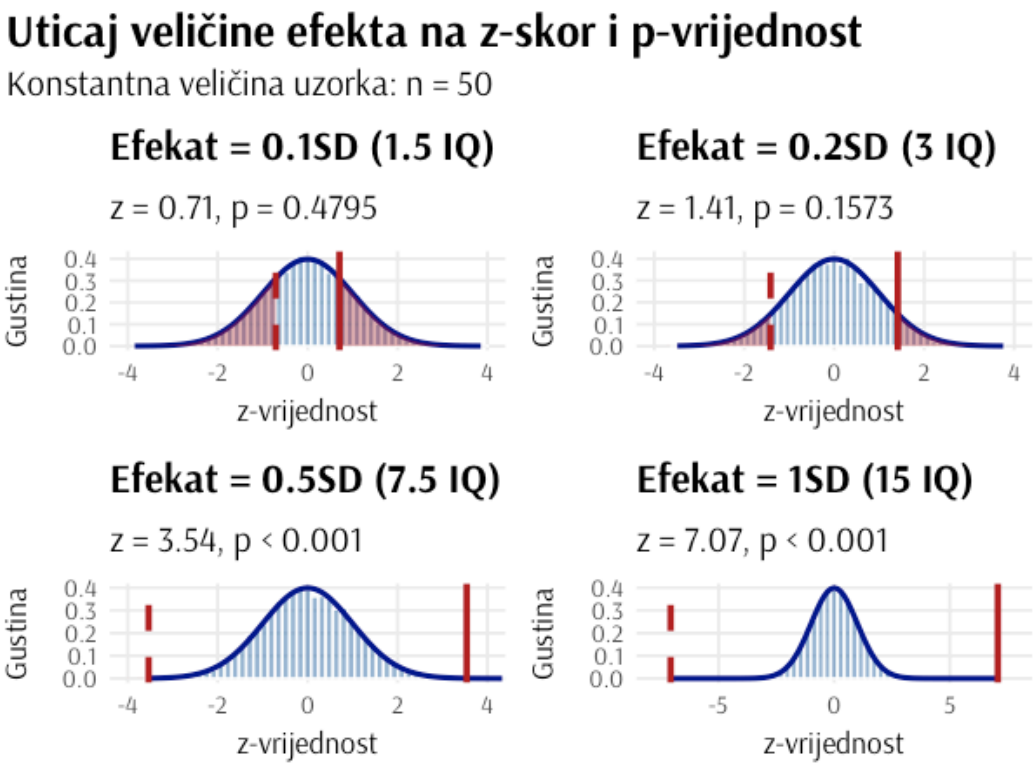


Slika 12.9: Uticaj veličine uzorka na z-skorove i p-vrijednosti.

Možemo da vidimo da se sa porastom veličine uzorka smanjuje širina distribucije, odnosno da je standardna greška manja[7]. To znači da iako imamo iste razliku $d = 1.0$ u sva četiri slučajeva, z-skorovi rastu, a p-vrijednosti postaju sve manje.

[7] Sjećate se Zakona velikih brojeva?

S druge strane Slika 12.10 pokazuje šta se dešava kada variramo veličinu efekta na uzorcima od 50 ispitanika. Kao što ste mogli pretpostaviti, sa sve većim razlikama, dobijaju se sve manje p-vrijednosti. Razlike od $d = 0.10$, kao i $d = 0.20$ nisu dovoljne da se formalno odbaci nulta hipoteza na nivou $\alpha = .05$. Međutim, razlika od pola standardne devijacije to već jeste. Vidimo da sa povećanjem efekta rastu i z-skorovi, odnosno p-vrijednosti bivaju sve manje. Drugim riječima, što su razlike među grupama veće, to rezultat ima veću dokaznu snagu protiv nulte hipoteze.



Slika 12.10: Uticaj veličine efekta na z-skorove i p-vrijednosti.

Kako ćemo izračunati p-vrijednosti kada ne znamo populacione karakteristike?

Međutim, u prošlom primjeru smo imali zadovoljene dvije pretpostavke koje uopšte ne važe u većini istraživačkih situacija sa dimenzionim varijablama. Prva pretpostavka je bila da je izvorna populaciona distribucija normalna, a druga da znamo vrijednost populacione standardne devijacije. Kako da izračunamo p-vrijednosti ako distribucija nije normalna i/ili ako ne znamo standardnu devijaciju populacije?

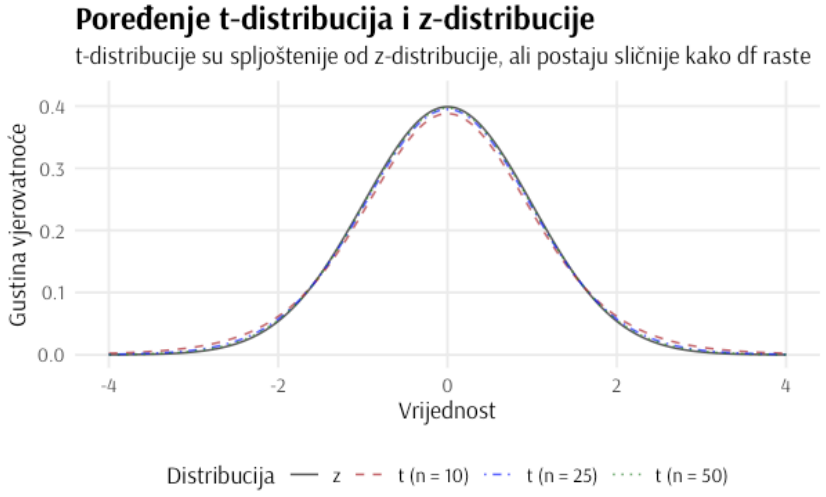
Ako imamo posla sa varijablom sa nekom nenormalnom izvornom distribucijom u pomoć će nam najprije priskočiti naša stara drugarica, *centralna granična teorema*. Tješiće nas kako ne bi trebalo da brinemo ako imamo relativno velike uzorke jer se aritmetičke sredine ionako raspodjeljuju normalno, bez obzira na to kakvog je oblika populaciona distribucija pojedinačnih skorova. U statistici uvijek postoje neko "ali", pa da biste se opušteno oslonili na centralnu graničnu teoremu morate da pretpostavite kako bi trebalo da izgleda ta distribucija. Za relativno simetrične distribucije je dovoljno da je $n \geq 30$, za umjereno

asimetrične preporučeno je da je $n \geq 50$, dok bi za izrazito asimetrične trebalo imati $n \geq 80$. Ne brinite ni ako nemate dovoljno ispitanika; umjesto anksiolitika možete koristiti i tzv. **neparametrijske testove** koje ćemo kasnije obrađivati.

Problem sa standardnom devijacijom rješavamo na drugačiji način. Kao prvo, koristićemo drugačiju formulu za z-skor. Nećemo je puno izmijeniti, samo ćemo umjesto populacione standardne devijacije koristiti standardnu devijaciju dobijenu na uzorku. Prema tome, formula standardne greške aritmetičkih sredina će tada glasiti:

$$SE_M = \frac{SD_X}{\sqrt{n}}$$

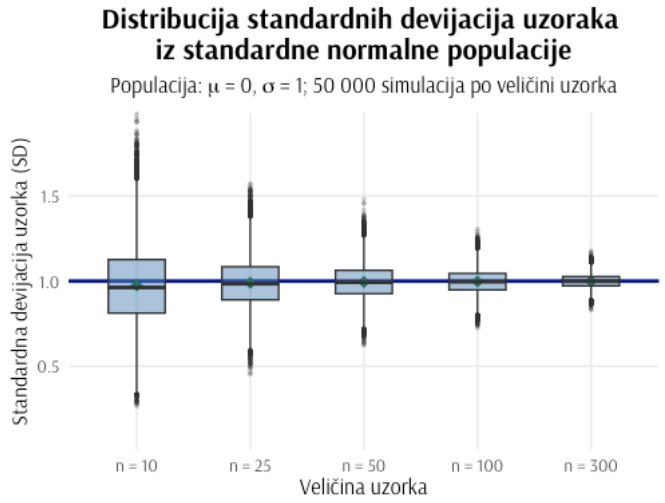
Drugo, nećemo koristiti normalnu distribuciju za izračunavanje p-vrijednosti, tj. njenu standardizovanu varijantu z-distribuciju, nego ćemo se pouzdati u njene bliske rođakinje, t-distribucije. Rođakinje u množini jer ih ima više, a nekoliko njih možete vidjeti na Slici 12.11 Od z-distribucija se razlikuju tako što su blago spljoštenije, a sa sve većim uzorcima izgledaju sve utegnutije, toliko da ih nećete moći razlikovati od z već kada se n približi 30.



Slika 12.11: Razlike i sličnosti z i t-distribucija.

Neko se može pitati, pa zašto su spljoštenije? Kao prvo, zato što na malim uzorcima očekujemo, u prosjeku gledano, blago manju varijabilnost nego na čitavoj populaciji. Ako bismo imali samo uzorak od 10 ispitanika, teško je da bi se u njemu istovremeno našle i neke veoma inteligentne i veoma neinteligentne osobe kojih ima u populaciji. Drugim riječima, mali uzorci teže da u

prosjeku budu manje raznoliki. Istovremeno, varijabilnost varijabilnosti je veća - odnosno na manjim uzorcima možemo dobiti i potcjenjivanje i precjenjivanje varijabilnosti u populaciji, te je logično da t-skor bude "oprezniji" u zaključivanju. Slika 12.12 oslikava ove dvije pravilnosti. Dok su za manje uzorke medijane standardnih devijacija jedva niže, varijabilnost standardnih devijacija je očito veća.



Slika 12.12: Distribucija standardnih devijacija u zavisnosti od veličine uzorka.

Da u našem primjeru nismo znali da je parametar standardne devijacije 15, morali bismo posegnuti za standardnom devijacijom uzorka. Recimo da je ona iznosila 10, odnosno da je uzorak potcijenio standardnu devijaciju opšte populacije. Tada bismo odgovarajući t-skor izračunali na sljedeći način:

$$t = \frac{M_\psi - \mu_0}{\frac{SD_\psi}{\sqrt{n}}} \Rightarrow \frac{115 - 100}{\frac{10}{7.07}} = \frac{15}{1.41} = 10.64$$

Konkretna t-distribucija koja nam je potrebna da dobijemo p-vrijednost ima ime $t_{(n-1)}$. Kako smo u našem uzorku imali 50 ispitanika, onda bi se ova zvala $t_{(49)}$. Dobijena p-vrijednost je nešto viša od one koja se dobija z-testom, ali i dalje ispod α nivoa 0.05 ($p < .00001$). Dakle, i u tom slučaju bi nam p-vrijednost govorila da bi trebalo odbaciti nultu hipotezu.

A šta je sa p-vrijednostima za hipoteze korelacije?

Kada već pominjemo t-distribucije uvod u p-vrijednosti bio bi nekompletan ako bismo propustili da pokažemo da se p-vrijednosti za testiranje statističke značajnosti korelacija mogu

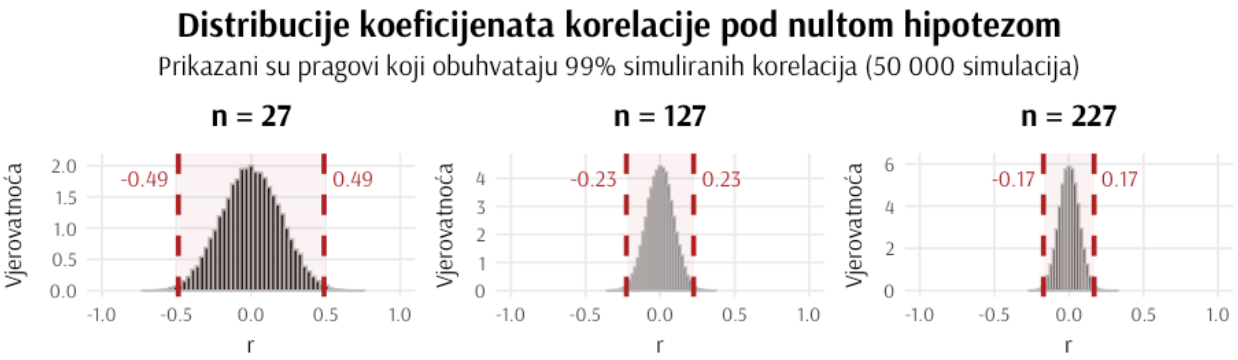
dobiti putem t-skorova. Za primjer hipoteze ćemo uzeti sad već davno prikazanu korelaciju između inteligencije i prosjeka ocjena. Da li se tada dobijena vrijednost $r = .40$ na uzorku od 227 ispitanika može smatrati statistički značajnom?

Nultom hipotezom se pretpostavljalo da u populaciji nema povezanosti ($H_0 : \rho = 0$). Hipotetički koeficijenti korelacija za takvu hipotezu se distribuiraju kao $t_{(n-2)}$ distribucija, pri čemu se koristi formula kojom se r (ili r_s kada je u pitanju Spearmanova rang korelacija) konvertuje u t:

$$t = r * \sqrt{\frac{n-2}{1-r^2}}$$

Iz formule se može vidjeti da se t povećava sa većim uzorkom i sa višim korelacijama, što smo mogli i pretpostaviti. Sa tako velikim uzorkom gdje je $n = 227$ koristili bismo t_{225} distribuciju koja je gotovo neraspoznatljiva od z-distribucije. Tako kritična vrijednost za $\alpha = .05$ iznosi $t_{225} = 1.97$, dok bi za z-distribuciju to bilo $z = 1.96$. Međutim, lakše bi nam bilo da umjesto t-skorova vidimo konkretne vrijednosti koeficijenata korelacija na x-osi. Da bismo to izveli, uvešćemo ovdje i još jedan način izračunavanja p-vrijednosti koji se zasniva na simulaciji podataka.

Naime, računari nam dopuštaju da na kvazi-slučajan način generišemo parove podataka za 227 ispitanika koji bi proisticali iz nulte hipoteze. Drugim riječima, nasumičnim putem se kreira 227 parova skorova i izračuna se korelacija između tih skorova. I to se radi iznova i iznova, na primjer 50 000 puta kao što je to urađeno za potrebe prikazivanja Slike 12.13. Na Slici 12.13 se još vidi da se distribucija očekivanih korelacija razlikuje u zavisnosti od veličine uzorka. Što su uzorci veći, to je manja varijabilnost dobijenih korelacija. Za naš uzorak od 227 ispitanika očekivalo bi se da se 99% korelacija nađe između vrijednosti -.17 i +.17 ako je nulta hipoteza u populaciji istinita. Za vrijednost koju smo dobili na uzorku ($r = .40$) dobijeno je $p < .0001$, pa bi bilo logično odbaciti nultu hipotezu.



Slika 12.13: Očekivane korelacije pod nultom hipotezom u zavisnosti od veličine uzorka.

Znači da postoji više načina da izračunamo p-vrijednosti?

Upravo tako. Ovo zahtjevno poglavlje ćemo i dovršiti vrlo kratkim predstavljanjem tri različita načina na koji se dolazi do p-vrijednosti. Najprije ćemo ponoviti prva dva koja smo gore upoznali: parametrijsko i simulacijsko.

Parametrijski pristup koristi već postojeće mape. Matematičari su već izračunali koliko su česte različite vrijednosti statistika za definisane distribucije kada važi nulta hipoteza (npr. binomne, z, t i još neke koje ćemo tek upoznati). Da bismo kreirali te distribucije potrebni su nam samo pretpostavljeni parametri (npr. μ, σ, N), pa se takvo zaključivanje naziva parametrijsko. Parametrijski pristup je zapravo tehnički najlakše izvesti i najčešće je korišten u psihološkim istraživanjima. Međutim, ovaj pristup počiva na određenim pretpostavkama o distribuciji podataka koje nisu uvijek ispunjene u stvarnim psihološkim istraživanjima.

Drugi gore prikazan način traži više računskog napora, gdje računar mora za nas stvoriti mnogo fiktivnih podataka pod nultom hipotezom. Da bismo to postigli računarski algoritam nahranimo pretpostavljenim parametrima, a on nam na osnovu njih proizvodi kvazi-slučajne uzorke. Algoritmu treba još da pojasnimo koliko tačno uzoraka želimo da dobijemo kako bismo dobili što tačniju procjenu zamišljenog svijeta. Taj broj bi morao biti velik, pa je opšta preporuka da se izvede barem 10 000 uzorkovanja. Za današnje procesore to obično nije veliki problem, te nam taj broj kao rezultat donosi glatke distribucije podataka koje se minimalno razlikuju od onoga što bismo dobili parametrijskim pristupom. Proponenti simulacionog pristupa su u James Bond stilu nazvali ovaj način Monte Carlo simulacijama budući da je Monte Carlo važio za grad sa čuvenim kazinima u

kojima slučajni podaci određuju sudbine (usput, nekih velikih para u teorijskoj statistici nema).

Treći način se koristi za ispitivanje veze između barem dvije varijable. To izračunavanje koristi princip kojeg nastavnici često žrtvuju ako je potrebno skratiti program srednjoškolske matematike: *permutacije*. Nakon što se izračuna statistik za stvarne podatke, računarski algoritam drži vrijednosti jedne varijable istim, a za drugu nasumično premeće mjesta originalnih vrijednosti, tj. permutuje ih. Na primjer, ako je Ana imala vrijednosti $IQ = 120$ i $Ocj = 4.80$, u permutovanom uzorku Ani ostaje vrijednost za inteligenciju $IQ = 120$, ali dobija skor za ocjenu nekog drugog učenika (npr. $Ocj = 3.35$), a u trećem slučaju neku treću ocjenu (npr. $Ocj = 5.00$) Budući da je permutovanje unošenje haosa u podatke, time se zapravo operacionalizuje nulta hipoteza. Kada se za sve ispitanike uradi ispremetačina onda se napravi lista svih dobijenih statistika. Oni se poredaju po apsolutnoj veličini (kako bismo dobili dvosmjernu p-vrijednost), te se izračuna percentilni rang vrijednosti koja je stvarno dobijena na uzorku. Ako je dobijeni percentilni rang 96, to znači da je postoji 4% jednakih ili ekstermnijih statistika (odnosno, $p = .04$). Kako biste to bolje shvatili Slika 12.14 ilustruje primjer ovog pristupa na izvodu iz analize gdje su od 24 moguće situacije ($4! = 4 * 3 * 2 * 1 = 24$) prikazane samo originalna sa uzorka i tri ispremetane vrijednosti za ocjenu, uz odgovarajuće koeficijente korelacija. Treba napomenuti da kada imamo prevelik broj podataka koji opterećuje čak i moderne procesore, moguće je napraviti parcijalne permutacione testove gdje se u obzir takođe uzima dovoljno velik broj uzoraka (npr. 10 000).

Student	IQ	Ocjena	Permutacija1	Permutacija2	Permutacija3
student #1	120	4.80	3.35	5.00	4.30
student #2	118	3.35	4.30	4.80	5.00
student #3	115	4.30	5.00	3.35	4.80
student #4	122	5.00	4.80	4.30	3.35
r =		.55	-.33	.62	-.83

Slika 12.14: Primjer permutacija.

Kada se koristi koji pristup za izračunavanje p-vrijednosti?

Parametrijski pristup, koji je obično zadan u statističkom softveru, je najbolji izbor kada:

- imamo dovoljno velike uzorke koji zadovoljavaju uslove

centralne granične teoreme
- su zadovoljene dodatne pretpostavke pojedinačnih analiza (o kojima će više riječi biti kasnije)
- želimo brzo dobiti rezultate

Simulacije podataka su najbolji izbor:

- imamo indicije da su populacione distribucije drastično različite od uobičajenih
- analiziramo složene odnose većeg broja varijabli

Permutacioni testovi bi bili prvi izbor kada:

- analiziramo vrlo male uzorke
- postoji mogući uticaj stršećih slučajeva
- podaci pokazuju snažniju asimetričnost distribucija

Konačno, najbolje postupanje - ako za njega imamo vremena i tehničke mogućnosti - bi bilo da provjerimo robusnost zaključaka tako što bismo ispitali da li sva tri pristupa daju saglasne zaključke. Sjajno je ako se rezultati podudaraju, ali ako ne, onda je potrebno detaljnije razmotriti zašto bi to bio slučaj.

Ovim smo završili tehnički opis dolaska do p-vrijednosti. Kao što možete da pretpostavite testiranje nulte hipoteze putem p-vrijednosti posjeduje niz poželjnih osobina za kojima smo tragali od početka knjige, ali postoje i neka značajna ograničenja. Kako se radi o ubjedljivo najpopularnijem i najzloupotrebljavanijem pristupu u psihološkoj statistici zaključivanja, cijelo sljedeće poglavlje je posvećeno prikazu prednosti i mana testiranja nulte hipoteze putem p-vrijednosti.

Poglavlje 13

Prednosti i mane testiranja nulte hipoteze putem p-vrijednosti (NHST)

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Koje su glavne prednosti testiranja nulte hipoteze putem p-vrijednosti koje su dovele do njegove široke upotrebe u psihologiji?
- Koji su ključni problemi i ograničenja p-vrijednosti koji su doprinijeli krizi ponovljivosti rezultata u psihologiji?
- Kako se p-vrijednosti najčešće pogrešno tumače i zašto takva tumačenja dovode do netačnih naučnih zaključaka?
- Na koji način p-vrijednosti mogu biti korisne u naučnom zaključivanju uprkos svojim ograničenjima?

Svrha prethodnog poglavlja je bila da shvatimo kako se dolazi do p-vrijednosti, te zašto koristimo **testiranje značajnosti nulte hipoteze** (engl. *Null Hypothesis Significance Testing, NHST*) u kvantitativnim istraživanjima. Vidjeli smo da se u epistemološkom smislu radi o opovrgavanju nulte hipoteze na osnovu opaženih podatak i teorije vjerovatnoće. Također, opisan je algoritam testiranja koji se sastoji iz četiri koraka:

1. Statistički se definiše nulta hipoteza
2. Generiše se hipotetička distribucija statistika koja se očekuje pod nultom hipotezom (na osnovu gotove matematičke distribucije, računarskom simulacijom podataka ili korištenjem permutacija)
3. Utvrđuje se vjerovatnoća dobijanja datog ili ekstremnijeg statistika na uzorku ako nulta hipoteza zaista važi u populaciji (p-vrijednost)

4. Ako je dobijena p-vrijednost niža od postavljenog praga za nivo statističke značajnosti (α) onda se nulta hipoteza odbacuje. U suprotnom se ne odbacuje.

Ova mehanička jednostavnost zaključivanja koja se zasniva na objektivnim matematičkim principima je svakako glavni razlog što je testiranje nulte hipoteze putem p-vrijednosti postalo ubjedljivo najpopularniji alat za testiranje istraživačkih hipoteza. Ako se dohvatite naučnih časopisa iz 1970ih, 1990ih, pa i 2010ih, shvatićete da su u njima p-vrijednosti razasute kao zvijezde na zvjezdanom nebu, odnosno skoro svaki rad ih pominje. Barem naizgled, u postupku testiranja je uloga istraživača svedena na najnižu mjeru: njihovo je da ubace podatke u softver, a on će da izbaci svoj sud.

Postoje li još neki razlozi što je NHST pristup stekao toliku popularnost?

NHST je primjenjiv na najraznovrsnije kvantitativne nacrte

Naravno. Jedan od njih je taj što se p-vrijednosti mogu izračunati za najrazličitije nacrte. Vidjeli smo da se njim mogu ispitati jednostavne hipoteze prevelencije kategoričkih i dimenzionih varijabli, razlika u mjerama centralnih tendencija i varijabilnosti (npr. aritmetičkih sredina, medijana, proporcija, standardnih devijacija), kao i povezanosti koje se procjenjuju najrazličitijim koeficijentima korelacija (npr. linearne, rang, poliserijalne, tetrahorične, fi). Ali p-vrijednosti se mogu dobiti i za različite kompleksne hipoteze (npr. multivarijatni modeli, medijacione i interakcijsko-moderacijske hipoteze, višeslojno linearno strukturalno modeliranje) koje sadrže i procjenu ukupnog modela, kao i pojedinačnih elemenata unutar njih. I ne radi se samo o psihologiji, nego i medicini, biologiji, ekonomiji, sociologiji… Dakle, p-vrijednosti su univerzalno sredstvo, baš kao i multinamjenski WD-40 sprej (koji služi za čišćenje, podmazivanje, dehidraciju, zaštitu i dekoroziranje metala), što omogućava i interdisciplinarnu komunikaciju.

NHST daje jednostavne odgovore na kompleksna pitanja

Još jedan razlog za popularnost je to što nam testiranje nulte hipoteze na gore opisani način daje binarni odgovor na postavljeno pitanje o hipotezi. Ovakvim testiranjem se donosi sud: ona ili jeste ili nije istinita. Kao vrsta smo skloni crno-bijelom razmišljanju zato što je i praktično odlučivanje binarno: ili ćemo nešto poduzeti ili nećemo.[1] Testiranje na osnovu poređenja p i α je kao pritiskanje prekidača kojim palite svjetlo. Ako je $p < \alpha$, popaliće se svjetla, a ništa se ne dešava ako je $p \geq \alpha$. To olakšava

[1] vidjeti npr. Fisher & Keil, 2018, https://doi.org/10.1177/0956797618792256

mijenjanje uvjerenja, kako recenzentima i urednicima, tako i opštoj javnosti koja bi trebalo da poštuje taj standard.

U ranom predračunarskom razvoju je razlog bio i to što je p-vrijednosti bilo moguće relativno jednostavno izračunati. Naime, zahvaljujući postojećim matematičkim distribucijama kreirane su i tablice za različite oblike distribucija pomoću kojih je lako bilo moguće procijeniti p-vrijednost za dobijeni statistik. Dolaskom računara je i sama procedura izračunavanja p-vrijednosti ekstremno pojednostavljena zahvaljujući statističkom softveru. Sa dva klika mišem učitate podatke, a onda još njih pet ili koji više da dođete do željene p-vrijednosti. Naučno zaključivanje je postalo dostupno svima. Ne morate čak ni završiti čitanje ove knjige…

Izračunavanje p-vrijednosti je dostupno svima

Konačno, razlog za sveopštu prisutnost p-vrijednosti je i pedagoška tradicija, odnosno to što značajna većina studijskih programa, kao i udžbenika (i na našim jezicima) predstavlja testiranje nulte hipoteze kao jedinu tehniku statističkog zaključivanja. Time je stvoren gotovo unificirani stav psihologa da su p-vrijednosti isto što i statistika zaključivanja. Drugim riječima, obezbijeđeno je da svi psiholozi znaju za p-vrijednost, odnosno koriste isti jezik.

NHST je bila jedina metoda koja se podučava

Postoje li neki nedostaci NHST pristupa?

Nažalost, uprkos velikom broju prednosti, istorija nauke je pokazala da pretjerano oslanjanje na p-vrijednosti dovodi do zanemarivanja temeljnih načela nauke i da može da čak ugrozi njenu reputaciju. Kroz istoriju psihologije su postojali glasovi[2] koji su ukazivali na nedostatke i nerazumijevanje NHST pristupa u naučnoj zajednici. Međutim, tek je jedno veliko istraživanje iz 2015. godine pokazalo o kolikim razmjerama se radi. Radi se o članku *Estimating the reproducibility of psychological science* objavljenom u prestižnom časopisu *Science*.[3] U njemu su prikazani rezultati ogromnog poduhvata gdje je učestvovalo 100 timova istraživača širom svijeta. Svaki tim je dobio zadatak da prateći originalne nacrte i opisane postupke prikupljanja podataka i statističke analize ispita da li će na novim uzorcima dobiti $p < \alpha$ kao što je to bio slučaj u originalnim radovima objavljenim u četiri prestižna časopisa. Rezultati su bili poražavajući. U originalnim istraživanjima je 97% radova imalo $p < \alpha$, ali je u ponovljenim istraživanjima to bio slučaj za samo 36% radova. Veliki broj radova je imao drastično niže veličine efekata nego u originalnim studijama. Ti rezultati su glasno odjeknuli i u

[2] npr. vama već dobro poznati Jacob Cohen u članku The earth is round ($p < .05$) https://doi.org/10.1037/0003-066X.49.12.997, kao i Paul Meehl, npr. https://doi.org/10.1086/288135

[3] Open Science Collaboration, 2015, https://doi.org/10.1126/science.aac4716

Kriza replikabilnosti

opštim javnim medijima koji su iz toga zaključili da je tek svaki treći objavljeni rad u psihologiji istinit, nakon čega je proglašena **kriza ponovljivosti rezultata** (engl. *replication crisis*).

Postoji niz identifikovanih propusta koji su doveli do te krize, a koje ćemo sada upoznati. Nisam advokat, ali važno je naglasiti da za njih nije uopšte kriva p-vrijednost, nego njena nekritička upotreba. To je problem svake životne situacije u kojoj nastojimo sve probleme riješiti istim alatom, a što je poznato kroz izjavu[4]: "Ako je čekić jedini alat koji imate, doći ćete u iskušenje da sve probleme tretirate kao eksere."

[4] Inače, ona se pripisuje čuvenom psihologu Abrahamu Maslowu: https://en.wikipedia.org/wiki/Law_of_the_instrument

Nepotpuni spisak problema testiranja nulte hipoteze putem p-vrijednosti je sljedeći[5] i svakim ćemo se kratko pozabaviti:

[5] Dužu listu daju npr. Greenland et al., 2016 https://doi.org/10.1007/s10654-016-0149-3

1. Nulte hipoteze su banalne
2. Statistički značajni rezultati ne znače i da rezultati imaju praktičan značaj
3. Dobijaju se automatizovani crno-bijeli odgovori za kompleksnu stvarnost
4. Obično se p-vrijednosti pogrešno tumače
5. Nulta hipoteza se ovim putem ne može potvrditi iako je to nekad cilj istraživanja
6. Previše fokusa se stavlja na rezultate pojedinačnih studija
7. Zanemaruje se uloga veličine uzorka u dobijanju p-vrijednosti
8. Višestruka testiranja povećava šansu da se dobiju statistički značajni rezultati
9. Ljudske odluke imaju veliku ulogu u izračunavaju p-vrijednosti
10. Fokusom na p-vrijednosti se zanemaruju druge metodološke pretpostavke

Banalnost nultih hipoteza

Nulte hipoteze su po sebi banalne. Naime, nulte hipoteze, pa i one o efektu slušanja barokne muzike na uspješnost zapamćivanja gradiva, obično tvrde da je populacioni efekat tačno nula: $\delta = 0.0000000$. Međutim, gotovo uvijek postoji barem neka razlika u populaciji, ali je najčešće nama istraživačima efekat $\delta = 0.03$, a nekad i $\delta = 0.12$, podjednako nezanimljiv kao i 0.00. Drugim riječima, testiramo hipotezu za koju u konačnici znamo da je neistinita.

Statistički značajno ne znači i praktično značajno

Iz navedenog slijedi da ako dobijemo statistički značajan rezultat to ne znači automatski da on ima neki praktični značaj. Nekad ćemo dobiti $p < \alpha$ kada je efekat nebitan u praktičnom smislu. Na primjer, na uzorku od 10000 ispitanika možemo dobiti da $d = 0.03$ bude statistički značajno, iako to zapravo znači da je prepoklapanje skorova dvije grupe iznosi čak 99%. U tom slučaju bismo imali tek 51% šanse da slučajno odabrani ispitanik iz eksperimentalne grupe (koji sluša baroknu muziku) ima bolji rezultat od slučajno odabranog ispitanika iz kontrolne grupe.

Automatizovani crno-bijeli odgovori o kompleksnoj stvarnosti

Svođenjem kompletne statističke analize na esenciju koju čine *Da* ili *Ne* odgovori dobijamo ekstremno pojednostavljenu sliku stvarnosti. Ako se nulta hipoteza i odbaci u opisivanom istraživanju efekta barokne muzike na učenje šta to tačno znači i za koga to važi? Na sva životna pitanja (npr. s kim ću izaći na romantični sastanak) postoji spektar pozitivnih i negativnih odgovora i stvarnost nije nikad tako jednostavna. Sama p-vrijednost nam neće dati odgovore na pitanja kod kojeg procenta ispitanih efekat funkcioniše, u kojim slučajevima i s kojim ispunjenim preduslovima. Nažalost, binarno odlučivanje o hipotezama na osnovu p-vrijednosti je otišlo tako daleko da je postalo bezumni ritual[6].

[6] Tako ga baš naziva poznati kritičar NHST postupka, psiholog Gerd Gigerenzer, 2004, https://doi.org/10.1016/j.socec.2004.09.033

Distorzirana tumačenja p-vrijednosti

Veliki izvor problema predstavljaju i dibidus pogrešna tumačenja p-vrijednosti. Već smo definisali p kao uslovnu vjerovatnoću događaja, ako je nulta hipoteza istinita u populaciji. Iz p-vrijednosti se ne može zaključiti ništa o vjerovatnoći nulte hipoteze, ali ni o vjerovatnoći alternativne hipoteze, o čemu govore i pravila uslovnih vjerovatnoća iz ranijeg poglavlja: $Pr(Podaci \mid H_0) \neq Pr(H_0 \mid Podaci) \neq Pr(H_a \mid Podaci)$. Ako je $p = .02$ to ne znači da postoji samo 2% vjerovatnoće da je nulta hipoteza istinita, niti da postoji 98% vjerovatnoće da se rezultati dobijaju pod dejstvom slučajnosti. Ako je nulta hipoteza tačna, istina je da postoji 2% vjerovatnoće da dobijemo takav ili ekstremniji rezultat, ali je istovremeno moguće da je vjerovatnoća da je nulta hipoteza istinita veća od vjerovatnoće da je alternativna hipoteza istinita[7].

[7] Poznati primjer je tzv. Lindleyev paradoks, https://en.wikipedia.org/wiki/Lindley%27s_paradox

Nepotvrdivost nulte hipoteze

Dok u slučaju $p < \alpha$ formalno odbacujemo nultu hipotezu, to ne znači da u slučaju $p \geq \alpha$ možemo prihvatiti nultu hipotezu. Takav rezultat može samo opravdati odluku da je ne odbacimo. Međutim, u nekim istraživanjima mi zaista želimo da pokažemo da između dva pristupa nema razlike. Drugim riječima, nekad želimo da dobijemo dokaz u prilog nulte hipoteze kako bismo možda pokazali da jedan psihoterapijski tretman koji je jeftiniji i zahtijeva manje seansi nego drugi daje iste rezultate kao drugi tretman. Podsjećam da ako dobijemo $p = .90$ to ne znači da postoji 90% vjerovatnoće da je nulta hipoteza istinita, niti da postoji 10% vjerovatnoće da je alternativna hipoteza istinita. Dakle, problem je to što nam $p \geq \alpha$ govori da nemamo dokaz o efektu, a ne govori da imamo dokaz da efekat NE postoji (savjet: pročitajte ovu rečenicu barem tri puta).

Pretjerana fokusiranost na pojedinačne studije

Naizgled, statistički značajne p-vrijednosti iz jednog istraživanja daju jasan odgovor na istraživačko pitanje. Ali već znamo da ako je $\alpha = .05$ onda očekujemo da u barem 5% ponavljanih istraživanja dobijemo $p < .05$ čak i kada je nulta hipoteza istinita. Kao što smo saznali iz krize ponovljivosti, pogotovo kada su studije sa malim uzorcima u pitanju, tek kada se isti rezultat ponovi više puta u različitim kontekstima ispravno je značajno povećati uvjerenje da je alternativna hipoteza tačna. Sve u svemu, zaključivanje o stvarnosti na osnovu jedne studije je obično preuranjeno.

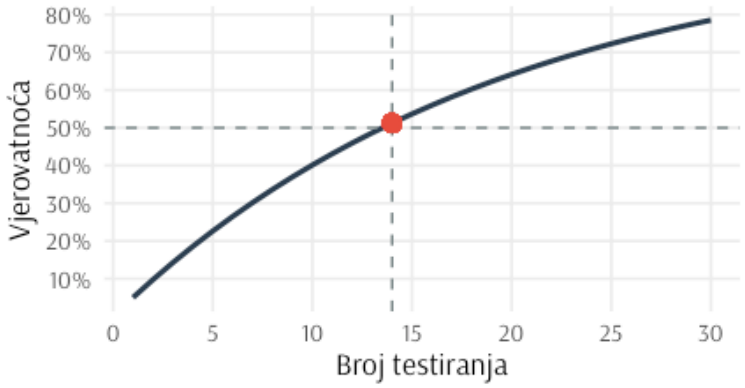
Ovisnost p-vrijednosti o interakciji veličine uzorka i efekta

Upravo je najčešći uzrok tzv. **lažno pozitivnih rezultata** (još poznatih i kao **Greška tip 1**) zanemarivanje uloge veličine uzorka pri izračunavanju p-vrijednosti. Naime, na malim uzorcima je moguće da snažniji uticaj na zaključke ostvare netretirani stršeći slučajavi ili neke treće varijable koje nisu kontrolisane nacrtom. Istovremeno, mali uzorci su i najčešći uzrok suprotne greške u zaključivanju poznate pod nazivom **greška promašaja** (**Greška tip 2**). Drugim riječima, p-vrijednost može biti iznad postavljene α vrijednosti zato što smo imali premalo ispitanika za datu veličinu efekta, čemu se mora posvetiti posebna pažnja

u pripremi istraživanja - te će zato i jedno poglavlje ovog teksta biti tome posvećeno.

Problemi višestrukog testiranja

Ono što posebno iznenađuje jeste da – sem veličine uzorka – i broj izvođenja testiranja stoji u vezi sa mogućnošću dobijanja lažno pozitivnih rezultata. Evo kako. Kada provodimo samo jedan test nulte hipoteze na osnovu p-vrijednosti i kada za granicu statističke značajnosti uzmemo nivo $\alpha = 0.05$, onda smo zapravo spremni da u 5% slučajeva odbacimo nultu hipotezu čak i kada ona važi u populaciji. To istovremeno znači da ćemo u 95% slučajeva donijeti *ispravnu odluku o neodbacivanju nulte hipoteze kada je ona istinita*. Međutim, kada imamo dva testiranja, vjerovatnoća dobijanja ispravne odluke u oba slučaja se mijenja. Formula kojom se izračunava šansa da za oba rezultata donesemo ispravnu odluku je: $0.95 * 0.95 = 0.9025$. To znači da postoji skoro 10% vjerovatnoće da se u jednom od dva testiranja dobije $p < .05$ čak i kada imamo potpuno nasumične podatke. Formula za izračunavanje mogućnosti da donesemo sve ispravne odluke o neodbacivanju nulte hipoteze kada je ona istinita glasi: $(1 - \alpha)^t$, gdje je t broj testiranja. Tako je za već 14 testova - što nije rijedak slučaj u psihološkim istraživanjima - vjerovatnoća da dobijete barem jedan lažno pozitivni rezultat već 50% (vidjeti Sliku 13.1).



Slika 13.1: Vjerovatnoća dobijanja barem jednog $p < .05$ kada je nulta hipoteza istinita u zavisnosti od broja testova.

Problemi ljudske prirode

Znanje o tome da povećavanje broja testiranja vodi većoj šansi da se napravi greška nije nužno korišteno za odvraćanje istraživača od pogrešne upotrebe, nego upravo suprotno. Neke istraživače je upravo ovo motivisalo da konačno dođu do statistički značajnih rezultata na ovaj ili onaj način. Naime, donedavno je važilo pravilo da ako želite da objavite članak, on mora istovremeno prenositi inovativnu ideju i pokazati statistički značajne rezultate. A objavljivanje članaka je neophodno da bi istraživači zadržali svoje poslove, dobili nove istraživačke projekte ili putovali na naučne skupove na egzotičnim destinacijama. Zato se nemali broj istraživača odlučivao da se kreativno poigra sa podacima, pa ako i ne dobije $p < .05$ za svoju glavnu hipotezu, onda ili malo dotjera neke podatke ili uključi u hipotezu i neke dodatne, neočekivane moderatorske varijable. Iz mnogobrojnih testiranja će sigurno nešto ispasti statistički značajno, te se onda narativ u uvodu i diskusiji prilagođava dobijenim rezultatima. Ove prakse su poznate kao lov na p (engl. *p-fishing* ili *p-hacking*), HARKovanje (od engl. *Hypothesizing After the Results are Known*) i upitne istraživačke prakse (engl. *Questionable Research Practices*).

Zanemarivanje teorije, metodologije, psihometrije

Postavljanjem primarnog fokusa na p-vrijednosti svi ostali koraci u analizi ostaju u drugom planu[8]. Nažalost, ovome su doprinijeli mnogi udžbenici statistike u psihologiji koji su zanemarivali deskriptivnu statistiku posvećujući im obično tek nekoliko stranica i brzo prelazeći na teoriju vjerovatnoće i objašnjavanje p-vrijednosti. No, kamo sreće da je problem samo u tome što se deskriptivnoj statistici i upoznavanju sa podacima daje premalo značaja. Ono što je još opasnije jeste što time sekundarni postaju i pitanja valjanosti mjerenja, kvaliteta nacrta i kontrole spoljnih varijabli, da ne spominjemo pitanje uzorkovanja koje gotovo nikad ne poštuje principe slučajnog odabira. Štaviše, ateoretski automatizovani proces testiranja nulte hipoteza se čak krivi i za teorijsku krizu koja se dogodila psihologiji. Razlog je što je postalo bitno samo dobiti statistički značajnu p-vrijednost putem statističkih tehnika koje psihološku teoriju vrlo skromno modeliraju i rijetko nude uvid u psihičku dinamiku.

[8] Posebno ilustrativan je članak autora Leek & Peng, 2015, https://doi.org/10.1038/520612a

Čemu onda služe p-vrijednosti?

Nakon svega navedenog, normalno je postaviti sebi pitanje da li p-vrijednosti mogu poslužiti ičem korisnom? Za neke eksperte je odgovor kategorično NE. Na primjer, još 2015. godine je časopis Basic and Advanced Social Psychology zabranio upotrebu p-vrijednosti[9]. U tamošnjim člancima se uglavnom naglasak stavlja na deskriptivnu statistiku, uz neke alternativne pristupe. Tako da ako ništa niste razumjeli iz prethodna tri poglavlja, ipak ima nade da možete objaviti naučni članak koji koristi statistiku.

[9] Urednici su to obrazložili ovdje: Trafimow & Marks, 2015, https://doi.org/10.1080/01973533.2015.1012991

Većina eksperata ipak tvrdi da problem nisu p-vrijednosti nego kako se one koriste. Ako smo svjesni njihovih ograničenja, onda ih možemo doživijeti kao vrijednu "prvu pomoć" u interpretaciji rezultata. Eksperti sa kojima sam saglasan zato savjetuju sljedeće pri korištenju p-vrijednosti u naučnom radu[10]:

[10] Sljedeći radovi su posebno ilustrativni: Dick & Tevaearai, 2015, https://doi.org/10.1016/j.ejvs.2015.07.026; Greenland et al., 2016, https://doi.org/10.1007/s10654-016-0149-3; Wasserstein et al., 2019, https://doi.org/10.1080/00031305.2019.1583913

- Uvijek saopštiti tačne p-vrijednost, a ne samo odluka o statističkoj značajnosti,
- Odluke o hipotezama nikada ne zasnivati isključivo na arbitrarnim pragovima kao što je $\alpha = .05$,
- Uvjerenja o praktičnoj značajnosti i uzročno-posljedičnom odnosu se ne smiju mijenjati samo na osnovu p-vrijednosti,
- Visoke p-vrijednosti ($p \to 1$) se ne mogu po sebi uzimati kao dokaz u prilog istinitosti nulte hipoteze,
- Gradacijski sve niže p-vrijednosti ($p \to 0$) se zaista mogu posmatrati kao korelati neistinitosti nulte hipoteze,
- Uz p-vrijednost je neophodno uvijek uzimati u obzir relevantne teorijski-metodološko-psihometrijski-statistički kriterije.

Iako je postojala apsolutna dominacije NHST pristupa tokom prethodnog vijeka godina, pojavilo se i nekoliko alternativnih postupaka koji adresiraju navedene kritike. Zadatak sljedećih poglavlja da se upoznate sa širokim repertoarom alata naučno-statističkog rasuđivanja. Počećemo sa onim što je temeljem svake zdrave nauke, a to je da nalazi moraju biti ponovljeni u više istraživanja kako bi bili razmatrani za uvrštavanje u korpus stvarnog naučnog saznanja.

Poglavlje 14

Uloga replikacionih studija u stvaranju naučnog znanja

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Koji oblici replikacionih studija postoje?
- Zašto su replikaciona istraživanja važna, a istovremeno rijetka u psihologiji i kako to utiče na pouzdanost naučnih zaključaka?
- Kako tumačiti situacije kada originalna studija pokazuje statistički značajan rezultat, a replikacija ne?
- Na koji način prakse otvorene nauke (preregistracija, dijeljenje podataka, kros-validacija) doprinose povećanju replikabilnosti i ukupnom kvalitetu psiholoških istraživanja?

Zamislite da smo naišli na dva eksperimentalna istraživanja o uticaju slušanja barokne instrumentalne muzike na učenje statistike[1]. U prvoj studiji, 22 studenta je u laboratoriji tokom 20 minuta učilo statističke pojmove dok im je u pozadini neprestano svirao Pachelbelov Kanon u d-duru[2]. Na testu znanja, ovi studenti su postigli značajno bolje rezultate od kontrolne grupe (takođe $n = 22$) koja je učila u tišini: $D = +3.6, SD_D = 6.97, t(42) = 2.43, p = .02$. Autori tog rada su zaključili da slušanje barokne muzike pospješuje učenje.

U drugom istraživanju, eksperimentalna i kontrolna grupa su imale po 18 studenata. Korištena je ista muzička kompozicija, isti materijali za učenje i isti test znanja, ali na uzorku studenata sa drugog univerziteta. Rezultati te studije su bili sljedeći: $D = +2.2, SD_D = 7.59, t(34) = 1.25, p = .22$. S obzirom na to da su i u toj studiji autori postavili nivo statističke značajnosti na

[1] Suština opisa i brojke koje slijede su preuzete iz knjige Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

[2] Poznato barokno djelo: https://www.youtube.com/results?search_query=pachelbel+canon+in+d

$\alpha = .05$, nulta hipoteza nije mogla biti odbačena i zaključak tih autora je bio da ne postoje dokazi u prilog alternativne hipoteze.

Gore navedeno oslikava situaciju koja se desila u skoro dvije trećine slučajeva u poduhvatu sa 100 nezavisnih istraživačkih timova, što je detaljnije opisano u prethodnom poglavlju[3]. U jednom istraživanju je dobijena statistički značajna razlika, a u drugom nije. Jesu li ovakvi rezultati kontradiktorni, odnosno da li je druga studija dala potpuno drugačije rezultate od prve? Šta možemo zaključiti iz prikazanog? U šta biste vi bili uvjereni i s kolikim stepenom uvjerenosti? I kako od te tačke da na principijelan način povećamo uvjerenje u jednu ili drugu stranu? Zdrav razum, ali i zdrava nauka, nam nalažu da bi trebalo da uradimo još jedno istraživanje. I još, i još jedno, i onoliko mnogo njih koliko je potrebno da vidimo da li je nešto istina ili nije, jer *repetitio est mater studiorum*.

[3] Open Science Collaboration, 2015, https://doi.org/10.1126/science.aac4716

Istraživanje koje se izvodi sa primarnom svrhom da se provjeri da li se dobijaju isti nalazi kao u nekom prethodnom istraživanju datog problema nazivamo **replikativno istraživanje** ili **replikaciona studija** (engl. *replication study*). Dok je zadatak svakog empirijskog istraživanja da provjeri vrijednost teorijske hipoteze, zadatak replikativnih istraživanja je da provjere pouzdanost tih empirijskih dokaza.

Mada je jasno da je taj zadatak od enormnog značaja za svaku nauku, ustanovljena su dva problema koja muče psihologiju. Kao prvo, neki pregledni radovi pokazuju da je tek svako stoto objavljeno istraživanje replikacionog karaktera![4] Drugo, čak i kada se replikacija obavi – kao što je slučaj sa primjerom s kojim smo započeli poglavlje ili sa značajnom većinom replikacija objavljenih u pominjanom Open Science Collabaration članku – stanje znanja nakon toga može postaje još više zbunjujuće nego što je bilo. Dakle, replikaciona istraživanja su armaturni materijal nauke, međutim i njima je potrebno pristupati s dužnom pažnjom. U ostatku poglavlja, ali i čitave knjige, razmatraćemo koje to oblike replikacionih studija imamo, koja je njihova funkcija, kako omogućiti njihovo češće provođenje, te na koji način je potrebno integrisati rezultate kako bi se došlo do čvršćih uvjerenja.

[4] Vidjeti Martin & Clarke, 2017, https://doi.org/10.3389/fpsyg.2017.00523

Kakve vrste replikacionih studija postoje?

Replikativna istraživanja se obično dijele na direktna (engl. *direct* / *close* / *exact*) i konceptualna (engl. *conceptual*). Međutim, kao i druge slične podjele ni ova nije u praksi binarna, već je

tačnije govoriti o dimenziji duž koje se mogu smjestiti konkretne manifestacije. Na jednom polu se nalaze **direktna replikaciona istraživanja** koja su u idealnom slučaju metodski identična originalnom istraživanju. To znači da bi sastavni dijelovi, kao što su postupak prikupljanja podataka, istraživački materijali, mjerni instrumenti, karakteristike uzorka, te postupak statističke analize morali biti isti kao u originalnom istraživanju. Naravno da je to nemoguće stoprocentno postići, ako ni zbog čeg drugog, zato što su uzorci ispitanika drugačiji. Na primjer, čak ako smo koristili i isti materijal za učenje, testove znanja, iste slušalice i glasnoću reprodukcije muzike, ako je originalno istraživanje o efektima barokne muzike provedeno na studentima u Japanu, a replikacija na Zapadnom Balkanu postoji velika šansa da kulturološke razlike utiču na rezultate. U svakom slučaju, da bismo određenu studiju zvali direktnom replikacijom trebalo bi da imamo što manje razlike u metodološkim karakteristikama.

Direktne replikacije

Kad se pomene termin replikaciona studija većina pomisli na direktne replikacije, čime se zanemaruje drugi pol dimenzije na kojem se nalaze konceptualne replikacije. I **konceptualne replikacije** imaju za cilj da prikupe dodatne dokaze za hipotezu originalnog istraživanja, ali se u njima namjerno koriste drugačije metode. Odnosno, replikacija postaje sve više konceptualna što više ima različitih metodskih karakteristika u odnosu na originalno istraživanje. Za primjer sa baroknom muzikom intervencije koje bi vodile sve više ka konceptualnoj replikaciji bi mogle biti da umjesto statističkog vokabulara dali ispitanicima da uče vokabular nekog nepoznatog jezika (npr. finskog), umjesto studenata bismo u uzorak uzeli osnovnoškolce ili sredovječne osobe, Vivaldijevu muziku umjesto Pachelbela, umjesto 20 minuta učili bi 5 minuta, tražili bi se otvoreni odgovori umjesto zatvorenih, znanje bi bilo provjeravano za 3 dana, a ne odmah nakon završetka sesije itd.

Konceptualne replikacije

Konačno, uz pojam replikabilnosti (tj. mogućnost da se dobiju istoznačni rezultati u ponovljenom istraživanju) potrebno je da definišemo još dva srodna pojma: reproducibilnost i robusnost[5]. Mada još uvijek postoje određena neslaganja u definicijama većina se slaže u tome da se pojam **reproducibilnost** odnosi na ponovljivost rezultata koje obrađuje neko drugi na osnovu dostavljenih podataka iz datog istraživanja pri čemu se koriste iste statističke procedure kao u originalnoj studiji. Konačno, **robusnost** se odnosi na nepromjenjivost zaključaka kada se na istim podacima koriste druge primjerene tehnike statističke analize. Dakle, rezultati bi bili reproducibilni ako bismo dobili iste rezultate kao i autori originalnog istraživanja kada koristimo

[5] Detaljno razgraničavanje pojmova prikazuju LeBel et al., 2018 https://doi.org/10.1177/2515245918787489 i Nosek et al., 2022, https://doi.org/10.1146/annurev-psych-020821-114157

njihove podatke i softver po našem izboru (npr. Jamovi, Jasp, R), te istu statističku tehniku (npr. t-test). Nalazi su robusni ako dobijemo istoznačne rezultate (npr. odluke o statističkoj značajnosti, slične veličine efekta) kao i autori originalnog istraživanja, ali umjesto linearnog koeficijenta korelacije koristimo rang koeficijent korelacije koji je također primjeren za analizu tih podataka.

Šta je svrha različitih oblika replikacionih studija?

Osnovna svrha direktnih replikacija jeste da se provjeri da li je dokaz za neku hipotezu postojan, dok je svrha konceptualnih replikacija je da se ispita uopštivost rezultata, odnosno granice njihove primjenjivosti. Jednostavnijim jezikom direktne replikacije odgovaraju na pitanje: "Možemo li mi to opet vidjeti?", a konceptualne replikacije na pitanje "Da li se to dešava u drugim uslovima?".

Obuhvatniji pogled nam sugeriše da replikacione studije ipak imaju više funkcija. Tako se u jednom članku[6] prikazuje njih pet, a ja ovdje predstavljam u blago modifikovanoj formi klasifikaciju sedam različitih funkcija:

[6] Schmidt, 2009, https://doi.org/10.1037/a0015108

Kontrola greške uzorkovanja

Primarni razlog postojanja replikacije jeste otkrivanje toga da li je rezultat u originalnoj studiji nastao usljed posebnosti uzorka na kojem je izvedeno istraživanje (greška uzorkovanja) ili zaista ukazuje na neku istinu. Pod rezultatom mislimo najviše na provjeru eventualnog pravljenje **Greške tipa I**[7], ali se zapravo jednako mogu koristiti i za provjeru eventualnog činjenja **Greške tipa II**[8] Za ovu svrhu su najpodobnija replikativna istraživanja u kojima su svi aspekti jednaki kao u originalnoj studiji osim što se regrutuju drugačiji ispitanici, te je poželjno i da ih ima više nego u originalnoj studiji.

[7] Greška tipa I: Lažno pozitivni rezultat kada se u analizi dobije statistički značajna p-vrijednost iako je nulta hipoteza zaista istinita u populaciji.

[8] Greška tipa II: Lažno negativni rezultat kada se u analizi dobije statistički neznačajan rezultat iako je alternativna hipoteza istinita u populaciji.

Kontrola istraživačkih artefakata u istraživanju

Istraživački artefakti su neželjeni efekti koji proizilaze iz nedostataka u nacrtu ili samom postupku provođenja istraživanja. Oni su najčešće rezultat grešaka mjerenja usljed tehničkih specifičnosti korištene aparature (npr. felerični ekrani na kojima se

izlažu stimulusi ili greške u programima za mjerenje reakcije), neodgovarajuće analize (npr. greške u podacima, suboptimalne procedure) ili dodatnih spoljnih varijabli na koje autori originalne studije nisu obratili pažnju (npr. uvježbavanje ispitanika, prisutnost eksperimentatora, placebo efekat). Za ovu funkciju su najpodobnije direktne replikacione studije u kojima se varira određeni aspekat protokola istraživanja koji nije direktno vezan za teoriju, ali bi trebalo provjeriti i reproducibilnosti i robusnost analiza na originalnom istraživanju. Drugim riječima, moraju se eliminisati i mogućnosti grešaka koja nastaju tokom obrade podataka.

Provjera istraživačkih prevara

Iako su rijetke, istraživačke prevare (npr. svjesno friziranje ili kompletno izmišljanje podataka i rezultata) predstavljaju ozbiljan problem za nauku budući da gadno urušavaju reputaciju discipline. Ako je originalno istraživanje zasnovano na izmišljenim podacima, nezavisne replikacije će vrlo vjerovatno otkriti da se efekti ne mogu reprodukovati. Poznati slučajevi naučnih prevara[9] su otkriveni onda kada drugi istraživači nisu mogli replikovati prikazane inovativne nalaze. Važno je naglasiti da ako rezultat originalne studije nije replikovan, to ne znači automatski da su autori originalne studije šarlatani. Kao što u ovom dijelu vidite, mnogi faktori mogu dovesti do određenih rezultata - ali višestruki neuspjesi da se iznenađujući rezultat reprodukuje vodi ka takvoj sumnji. U razotkrivanju mogućih prevara, direktne replikacije jesu najvažniji prvi korak, ali često nisu dovoljne same po sebi. One se moraju kombinovati sa analizom reproducibilnosti što u ovom slučaju predstavlja statističku provjeru vjerodostojnosti originalnih podataka, njihovih obrazaca i provedene statističke analize.

[9] npr. slučaj Diederika Stapela za kojeg se pokazalo da je za barem 58 radova izmislio podatke ili doradio rezultate https://retractionwatch.com/category/by-author/diederik-stapel/ što opisuje u svojoj autobiografiji *Faking Science: A True Story of Academic Fraud* https://forrt.org/curated_resources/faking-science-a-true-story-of-academic/

Provjera generalizacije nalaza na različite populacije

Replikacije su naročito korisno sredstvo za ispitivanje eksterne valjanosti zaključaka, odnosno pitanja da li se efekti generalizuju na druge populacije osim originalno ispitivane. Kao što je već navedeno, ako je originalna studija o uticaju barokne muzike izvedena sa studentima psihologije u Japanu, replikacije mogu ispitati da li isti efekti postoji kod srednjoškolaca, starijih osoba, ili osoba iz drugačijih kultura. Ova funkcija je posebno važna kada se ima u vidu saznanje sa početka knjige, da veliki dio psiholoških istraživanja koristi tzv. WEIRD uzorke (Western,

Educated, Industrialized, Rich, Democratic), što ograničava uopštivost nalaza. Direktne replikacije uz sistematsko variranje ciljne populacije (ispitanika, ali i materijala!) pomažu da se utvrdi koji su psihološki fenomeni univerzalni, a koji specifični kada se u obzir uzme i sociokulturalna perspektiva.

Provjera teorijskog obuhvata hipoteze korištenjem drugačijih procedura

Replikacije omogućavaju provjeru teorijskog obuhvata određenog fenomena. Dakle, ako zaista ima istine u tome da slušanje konkretnog djela barokne muzike pospješuje učenje, onda bi slični efekti trebalo da budu vidljivi i kada se koriste različite operacionalizacije varijabli. Kao što je već navedeno, potrebno je provjeriti i djelotvornost drugih baroknih djela, a zatim i muzike klasicističkog perioda ili djela drugih žanrova, te osim efekta na pamćenje kao zavisne varijable, eventualno provjeriti i efekte na druge kognitivne procese kao što su pažnja ili rezonovanje. Ako se efekti pojavljuju i pod izmijenjenim uslovima, to značajno jača teorijsku osnovu fenomena. Odmak od originalne recepture nam pomaže u prikupljanju dokaza o granicama obuhvata efekta.

Procjena preciznosti i pouzdanosti veličine efekta

Osim što je primijećena slaba replikabilnost statističke značajnosti rezultata, Open Science Collaboration projekat je pokazao i da se veličine efekata koje se dobijaju u replikacionim studijama u prosjeku gledano upola niže nego veličine efekata u originalnim istraživanjima. Ova pojava, da prve objavljene studije o nekoj pojavi teže da precijene veličinu efekta je odranije poznata i ima svoje filmsko ime "prokletstvo prve studije"[10]. Kombinacija faktora doprinosi toj pojavi, ali najbitniji je to da istraživači svjesno odabiraju da saopšte one rezultate koji će zadovoljiti uredničku politiku najvećeg broja naučnih časopisa, a to opet znači sljedeće: hipoteze bi morale biti što inovativnije, rezultati moraju biti statistički značajni, a to opet na manjim uzorcima zahtijeva veće veličine efekata. Slika 14.1 upoređuje motivaciju koja krasi "agresivnog inovativnog istraživača" i "savjesnog replikatora". Više replikacionih studija, kako direktnih, tako i konceptualnih, pruža iskreniji uvid u stvarnu snagu efekta.

[10] Ioannidis, 2008, https://doi.org/10.1097/EDE.0b013e31818131e7

	Agresivni pronalazač	*Promišljajući replikator*
Ono što je bitno jeste...	Novo otkriće	Ponavljanje ranijeg nalaza
Baze podataka su...	Privatni rudnici zlata koji se ne dijele	Javno dobro
Smatra da je dobar istraživač onaj ko...	Smišlja i kreativno ispituje nove ideje	Čvrsto se pridržava nacrta i plana analize
U naučnim radovima izvještava...	Ono što je zanimljivo	Sve
Objavljuje naučne radove tako što...	Svaki uočeni efekat objavljuje u zasebnom radu	Objedinjuje sve efekte i prikazuje ih u jednom naučnom radu
Nakon objavljivanja svog rada...	Promoviše nalaze što je snažnije moguće	Kritično i s oprezom govori o svojim nalazima

Slika 14.1: Razlike između "agresivnog pronalazača" i "savjesnog replikatora" (prilagođeno iz Ioannidis, 2008).

Pedagoško-obrazovna funkcija

Konačno, direktne replikacione studije imaju veoma važnu ulogu u obrazovanju budućih naučnika. Izvođenjem replikacionih studija studenti iz prve ruke uče naučnu metodologiju, razvijaju kritičko mišljenje i stiču bitna praktična iskustva u istraživačkom procesu. Replikacije predstavljaju idealne studentske projekte jer imaju jasno definisane postupke, a istovremeno su važan doprinos naučnom saznanju. Pedagoška vrijednost replikacija leži i u tome što one podstiču naučni skepticizam - studenti uviđaju da pojedinačna istraživanja ne treba smatrati konačnim dokazom, već dijelom kumulativnog procesa izgradnje znanja. Ova funkcija replikacija je rjeđe predmet rasprave, ali je izuzetno važna za razvoj naučne pismenosti i istraživačke kulture. Ako ste se umorili od čitanja, pravi trenutak je da razmislite koju studiju biste željeli replikovati.

Kakve istraživačke prakse omogućuju veću prisutnost replikacionih studija?

Zanemarivanje replikativnih istraživanja kroz istoriju psihologije je velikim dijelom uzrokovan politikama naučnih časopisa koje favorizuju originalnost. Za istraživače, a posebno one na početku karijere, replikacije su manje privlačna opcija u poređenju sa novim otkrićima koja donose više prestiža i citata. Naučne karijere su vrlo kompetetivne i u njima vlada kultura preživljavanja poznata pod imenom "objavi ili nestani" (engl. *"publish or perish"*). U takvom okruženju istraživači sebi dopuštaju mnogo stepeni slobode pri analizi podataka - od selektivnog izbacivanja netipičnih rezultata, do isprobavanja najrazličitijih akrobacija sve dok se iz podataka ne iscijedi neki statistički značajan rezultat

(već pomenuti lov na p-vrijednosti). Kao šlag na torti, mnogi istraživači su bili nevoljni da dijele svoje protokole i podatke, što je dodatno otežavalo kvalitetne replikacije, jer sam opis metodske sekcije obično ne pruža dovoljno informacija da se uradi direktna replikacija.

Srećom, svaka kriza je prilika za rast i razvoj, pa je kriza replikabilnosti dovela tokom posljednjih godina do povećanog insistiranja na praksama **otvorene nauke** (engl. *open science*):

Veća transparentnost metoda

U važnijim časopisima se sve detaljnije izvještava o metodskim aspektima istraživanja. Za to su zaslužne inicijative poput smjernica definisanih u preporukama kao što je **Transparency and Openness Promotion**[11] kojima se časopisi boduju u zavisnosti od ispunjenosti različitih uslova. Da bi se dostigao najviši nivo transparentnosti autori se obavezuju da prilože detaljne protokole istraživanja i liste materijala. U idealnim slučajevim se na stranici časopisa zajedno uz članak objavljuje i dopunski materijal koji je dovoljan da i sami obavite replikaciju, te provjeru robusnosti i reproducibilnosti rezultata.

[11] tzv. TOP smjernice: Nosek et al., 2015, ažurirano 2025, https://www.cos.io/initiatives/top-guidelines

Preregistracija studija

Među TOP smjernicama se nalaze i zahtjevi za definisanjem hipoteza, uzorka i statističko-analitičkih postupaka mnogo prije nego se pristupi prikupljanju podataka, što značajno smanjuje selektivno izvještavanje samo onoga što odgovara istraživačima. Platforme poput OSF (Open Science Framework, https://osf.io/) olakšavaju preregistraciju, a mnogo časopisa nudi posebne oznake za preregistrovane studije što čitaocu povećava povjerenje u rezultate studije. Na primjer, preregistracijom se traži da se jasno navede šta ćemo uraditi sa ekstremnim podacima ako se oni pojave u uzorku. Tu moramo navesti ili da ćemo isključiti iz analize mjere koje odstupaju više od 2.5 standardne devijacije od prosjeka ili da ćemo koristiti neparametrijske testove. Ono što je bitno jeste da se odlučimo za jednu opciju koje ćemo se držati kada podaci stignu. Ovakva transparentnost smanjuje mogućnost naknadnog prilagođavanja analiza što inače dovodi do povećanja Greške tipa I.

Veća dostupnost baza podataka

Dijeljenje podataka omogućava drugim istraživačima da nezavisno provjere reproducibilnost i robusnost nalaza, ali dosad je to obično bio mukotrpan proces gdje ste morali da kontaktirate istraživače mejlom i ovisite o njihovoj dobroj volji. Repozitoriji (onlajn skladišta) poput Dataverse projekta i OSF su danas postali standardne platforme za dijeljenje podataka, a u mnogim časopisima je dostavljanje baze podataka na kojoj je vršena analiza nužan uslov za objavljivanje rada. Štaviše, postoji i specijalizovani časopis Journal of Open Psychology Data u kojem se nakon recenzija objavljuju baze podataka sa pratećim vrlo detaljnim opisom načina prikupljanja podataka. Ako imate slobodnog vremena na raspolaganju eto šta možete da radite.

Dostupnost koda za statističku analizu

Da bi statistička analiza bila reproducibilna preporučeno je da imamo ispis komandi koje su korištene pri analizi. Ranije je bilo uobičajeno da se analize provode u komercijalnim statističkim softverima (npr. SPSS, Stata) klikanjem miša, ali je reproducibilnost takvih analiza u načelu niska. Često se sami istraživači ne mogu ni sjetiti kako su tačno došli do rezultata. Upravo zato posljednjih godina postoji naglasak na korištenju besplatnih statističkih jezika kao što su R ili Python. Oni su svima dostupni, omogućavaju ispis komandi koje se mogu sačuvati kao jednostavni tekstualni fajlovi i podijeliti sa drugim istraživačima koji mogu da procijene adekvatnost analiza i reprodukuju nalaze. Dostupni kodovi sa statističkim komandama imaju i pedagošku svrhu, jer omogućuju motivisanim studentima da steknu napredne kompetencije u pogledu statističkih analiza.

Ispitivanje robusnosti analiza u zavisnosti od statističkih tehnika

Ako su podaci dostupni (a može pomoći i kod), onda nezavisni istraživači mogu da provjere da li nalazi ostaju isti ako se koriste različite statističke tehnike. Za najjednostavniji primjer možemo uzeti pitanje da li se može reprodukovati saopšteni nalaz o linearnoj korelaciji između dvije varijabli, ako koristimo rang koeficijent korelacije, odnosno da li se mogu reprodukovati različite p-vrijednosti u zavisnosti od korištene tehnike. Stručnim jezikom se ovo naziva **analiza osjetljivosti** (engl. *sensitivity analyses*), a u posljednje vrijeme postoje i studije koje pokazuju koliko

istraživački zaključci mogu varirati ako se za kompleksnije nacrte koriste alternativni setovi analiza (engl. *multiverse analyses*.[12] Na primjer, studija o efektu barokne muzike na učenje mogla bi da ispita kako bi rezultati izgledali ako se: (a) uključe svi ispitanici ili samo oni koji su naveli da nisu imali probleme sa spavanjem prethodnog dana; (b) koriste parametrijski ili neparametrijski testovi za procjenu statističke značajnosti; (c) kontrolišu različiti kovarijati poput prethodnog muzičkog iskustva ili pola. Ako se zaključak održi kroz većinu ovih analitičkih odluka, to značajno povećava povjerenje u njegovu robusnost.

[12] Vidjeti npr. Götz et al., 2024, https://doi.org/10.1002/ijop.13229 ili Heyman & Vanpaemel, 2022, https://doi.org/10.15626/MP.2020.2718

Pojava kolaborativnih replikacionih projekata

Pored individualnih replikacija, posljednjih godina se sve češće javljaju i kolaborativni replikacioni projekti koji uključuju više istraživačkih timova. Takvi projekti[13], koordinišu istraživače širom svijeta u simultanom provođenju jednog istraživanja u mnogo različitih država. Ovakav pristup ne samo da značajno povećava veličinu uzorka, nego istovremeno omogućava ispitivanje kulturoloških i kontekstualnih faktora koji mogu uticati na replikabilnost[14].

[13] npr. Many Labs https://osf.io/wx7ck/, Many Babies https://manybabies.org/, Psychological Science Accelerator https://psysciacc.org/

[14] Yours truly, autor ove knjige, je do trenutka pisanja knjige učestvovao u tri ovakva projekta: 1) The Global Temperament Project: Parent-reported temperament in infants, toddlers, and children from 59 nations, https://doi.org/10.1037/dev0001732, 2) Psychological Science Accelerator *PSA-JTF2 Dignity Honor Face https://psysciacc.org/projects/psajtf002.html, 3) Cross-cultural TCAM/iNAR study on questionable health behaviors: https://osf.io/z7598*

Dakle, sve gore navedeno su replikativne prakse koje povećavaju povjerenje u psihologiju kao nauku uopšte i nalaze pojedinačnih studija. Uz njih, postoji još jedan inventivni pristup koji je u psihologiju došao relativno skoro, a radi se o kros-validacionim studijama.

Šta su to kros-validacione studije?

Kros-validacija ili **ukrštena validacija** (engl. *cross-validation*) je statistički postupak kojim se baza podataka prikupljenih u jednom istraživanju dijeli na dva ili više podskupova kako bi se na njima nezavisno provele statističke analize i uporedili rezultati. Za razliku od klasičnih replikacionih studija koje podrazumijevaju prikupljanje novih podataka, kros-validacijom obavljamo internu replikaciju nalaza. S obzirom na to, zapravo se radi o najdirektnijoj mogućoj replikaciji, budući da se provjera vrši unutar istog uzorka, te se time eliminiše faktor različitosti uzoraka (npr. ispitanici iz različitih gradova ili država). Zato kros-validacije imaju status replikacionih studija koje maksimizuju internu valjanost zaključaka. Međutim, iz istog razloga, radi se o studijama koje nemaju za cilj proširenje eksterne valjanosti. Izuzetak predstavljaju slučajevi kada se podjela baza na

podskupove vrši ciljano na osnovu neke bitne spoljne varijable (npr. pol, škola ili grad iz kojeg su ispitanici). Ali da vidimo kako se kros-validacione studije izvode u praksi, te koje su im prednosti i mane.

Kako se izvode kros-validacione studije?

U najjednostavnijem i najčešćem obliku (engl. *holdout* ili **"ostavi-sa-strane"** pristup), kros-validacija se izvodi kroz dva koraka:

1. Na slučajan način se izdvaja jedan dio uzorka (obično 70-80%, ali ne manje od 50%) na kojem se procjenjuju parametri, testiraju hipoteze i/ili razvijaju statistički modeli. Taj dio uzorka se naziva **trening ili kalibracioni poduzorak** (engl. *training dataset*).
2. Na preostalom dijelu uzorka – koji se naziva **validacioni ili testni poduzorak** (engl. *test dataset*) – se onda koriste identične statističke procedure kako bi se utvrdilo koliko blisko se rezultati sa kalibracione baze replikuju na ostatku podataka.

Osnovna prednost ovakvog dijeljenja uzorka jeste jednostavnost. Potrebno je samo da softverom odredimo kako će na slučajan način podaci biti raspoređeni u dvije grupe. Nedostatak je to što dosta toga ovisi o tome kako će se izvršiti to raspoređivanje. Naime, ako bi se uzorak podijelio na drugačiji način možda ne bismo dobili upravo takve rezultate. Logično, mana ovog pristupa je i to što je potrebno da imamo dovoljno veliki startni uzorak tako da i ovaj manji ima dovoljan broj ispitanika (npr. barem 200 u studijama poprečnog presjeka). Ovaj vid kros-validacije se često koristi u eksploratornim istraživanjima kada imamo dovoljno velike, ali ne i BAŠ velike uzorke, a želimo doći do nekih stabilnost nalaza. Posebno je čest u psihometrijskim istraživanjima u kojima se analizira interna struktura instrumenta pa se na jednom poduzorku kreira model, a na drugom on provjerava. Konačno, treba napomenuti da podjela na podskupove može služiti i drugoj svrsi. Iako se može voditi debata o tome da li je to kros-validacija u pravom smislu, baza podataka se može dijeliti prema postojećim kategorijama u uzorku (npr. pol, škola, grad) kako bi se ispitao uticaj potencijalno moderatorskih varijabli na rezultate. Što je veća sličnost između rezultata uzorka, time na principijelan način povećavamo uvjerenje o stabilnosti nalaza.

Kako bi se prebrodili efekti proizvoljnosti do kojih može da dovede ta jedna podjela, pribjegava se i automatizovanoj **k-strukoj kros-validaciji** (engl. *k-fold cross-validation*). Taj pristup podrazumijeva da se kompletni uzorak nasumično dijeli na k-dijelova, što je obično 5 do 10 podskupova koji imaju jednak broj ispitanika. Za trening se koristi objedinjeni poduzorak $k-1$ dijelova (npr. objedine se poduzorci 1, 2, 3 i 4), a preostali dio se koristi kao testni podskup (tj. podskup 5). U sljedećim koracima se izvode permutacije uloga (što već znate šta je). Kroz iteracije (a eto još jedne nove riječi koja označava ponavljane procedure) se onda mijenjaju uloge: npr. poduzorci 1, 2, 3 i 5 se objedine i koriste za trening, a na 4. podskupu se provjerava model. To se sve ponavlja dok svaki poduzorak ne odigra ulogu testnog podskupa, dakle k puta. Dobijeni rezultati iz svih k-iteracija se uprosječuju, a uz te prosječne mjere se prikazuju i neka mjera varijabilnosti (npr. SD, SE) koje operacionalizuju nestabilnost modela. Na primjer, ako koristimo 5-struku kros-validaciju za predviđanje akademskog uspjeha na osnovu osobina ličnosti, dobićemo pet različitih procjena uspješnosti našeg modela. Prosječna tačnost će služiti kao naša najbolja procjena performansi modela, a standardna greška ovih pet vrijednosti će nam ukazivati na to koliko je model osjetljiv na promjene u sastavu uzorka. Ova varijanta kros-validacije je jedan od standardnih postupaka u mašinskom učenju, gdje algoritam samostalno kreira napredne modele koje može da istestira i gdje možemo na kraju da izaberemo najpouzdaniji među njima. Iako se ovim pristupom u velikoj mjeri rješava problem proizvoljnosti podjele na kalibracioni i validacioni poduzorak, očito je da je za njegovu primjenu neophodno da imamo velike početne uzorke. Dakle, radi se o epistemički snažnom pristupu koji se može koristiti na masovnim istraživačkim projektima.

Kako biste shvatili inovativnost kros-validacionog pristupa replikaciji, zgodno je opisati i najekstremniju varijantu među njima. Ona se se naziva **"izostavi-jednu-opservaciju"** (engl. *Leave-One-Trial-Out*), a najčešća je njena podverzija kreirana za vezana mjerenja **"izostavi-jednog-ispitanika"** (engl. *Leave-One-Subject-Out*). Statistička analiza se vrši n puta – odnosno na onoliko uzoraka koliko ima opservacija ili ispitanika ako oni imaju više od jednog mjerenja. U svakoj iteraciji se samo jedan ispitanik koristi za testiranje modela, a svi ostali za treniranje modela. Za svaku iteraciju se zatim izračunava greška modela, odnosno koliko taj jedan ispitanik odstupa od onog što model predviđa. Dobijene mjere greške, odnosno odstupanja mjere pojedinca od predviđene mjere koja se dobija na svim ostalim

ispitanicima u uzorku, služe da se zaključi da li neki ispitanici predstavljaju stršeće slučajeve, kao i koliko stabilnost modela zavisi od pojedinačnih slučajeva. Za razliku od drugih varijanti, ova verzija kros-validacije se koriste kada u uzorku imamo mali broj ispitanika (često između 30 i 40), a veliki broj varijabli. To je pogotovo slučaj u kliničkim, psihofiziološkim i neurokognitivnim studijama, kao i u drugim oblastima, u kojima se koriste višestruka mjerenja tokom jedne sesije ili više dana.

Može li jedan primjer kros-validacione studije?

Kao primjer prikazujem jednu našu studiju (Lakić, Damjenić i Grahovac, 2018, https://osf.io/preprints/psyarxiv/t94gv_v1) koja koristi ultrajednostavnu primjenu uz korištenje krajnje jednostavne statističke tehnike. Naime, istraživačkim projektom smo željeli da saznamo koje to strategije učenja pouzdano doprinose akademskom postignuću. U tu svrhu smo kreirali iscrpan inventar strategija učenja, koji je uz dodatne afektivne i motivacione aspekte, sadržavao opise 40 konkretnijih ponašanja (npr. "Pišem bilješke na marginama udžbenika iz kojih učim." i "Kada učim, imam posebno mjesto na kojem to radim (npr. moj radni sto, čitaonica)." Prikupili smo podatke za 402 učenika srednjih škola, a s obzirom na veliki broj različitih ponašanja, željeli smo sa snažnijim stepenom uvjerenja izdvojiti manji broj onih koje stoje u stabilnoj vezi sa prosječnom ocjenom. Zato smo na slučajan način prepustili softveru da podijeli kompletan uzorak na dva podskupa (svaki $n = 201$). Za prag selekcije finalnih strategija učenja smo uzeli vrijednost korelacije sa prosječnom ocjenom $|r| \geq 0.19$, što je na uzorku od 201 ispitanika odgovoralo $p \leq .007$ (konkretan granični skor smo zapravo odabrali na osnovu vrijednosti drugog statistika kojeg obrađujemo kasnije u knjizi). Dakle, kako bismo određenu strategiju učenja ocijenili kao stabilan korelat prosječne srednjoškolske ocjene, ona je morala pokazati korelaciju $|r| \geq 0.19$ na oba poduzorka. Slika 14.2 prikazuje rezultate za 11 izdvojenih strategija. Primijetićete da postoje i vidne varijacije u koeficijentima korelacija među poduzorcima. Međutim, vjerovatno ćete se složiti sa tim da smo - dobijajući dokaze u dvije studije, umjesto samo na jednoj - dobili čvršće indicije da se radi o stabilnim korelatima akademskog postignuća.

Šta smo naučili o značaju replikacija?

Temelji naučne psihologije su objašnjeni i u Popperovim i Fisherovim tekstovima u kojima su jasno davali do znanja da

Stavka	r_1	r_2	r
Kada treba da naučim neku listu pojmova, smišljam sopstvene asocijacije kako bih ih lakše zapamtio/la (npr. kreiram akronime/skraćenice, vizualiziram odnose među pojmovima).	.42	.23	.32
Tokom učenja procjenjujem (tj. preispitujem se) koliko dobro sam usvojio/la gradivo koje sam učio/la.	.21	.23	.21
Precizno planiram periode i/ili količinu gradiva koje ću da učim tokom jedne "seanse" učenja.	.24	.22	.22
Redovno učim tokom školske godine, a ne samo pred provjere znanja.	.36	.29	.33
Nakon što završim sa "seansom" učenja, procijenim koliko sam ostvario/la od planiranog.	.29	.24	.26
Ako za predmet postoji skraćena verzija teksta (skripta), najviše učim iz nje.	-.19	-.29	-.24
Kada učim koristim šire izvore informacija od onoga što je obavezna literatura (npr.druge udžbenike i tekstove; napomena: ovim se ne misli na skraćene skripte).	.20	.26	.24
Trudim se da gradivo pamtim od riječi do riječi.	-.22	-.19	-.21
Tokom nastave postavljam pitanja nastavnicima kada se izlože meni nejasni dijelovi gradiva.	.24	.27	.25
Tokom nastave pravim pismene bilješke.	.43	.28	.36
Koristim druge digitalne resurse koji nisu uključeni kao obavezni sadržaji (npr. softver, edukativne sajtove, online kurseve, blogove, video materijale).	.20	.23	.22

Slika 14.2: Korelacije između 11 izdvojenih strategija učenja i prosječne srednjoškolske ocjene na ukupnom uzorku (r) i dva poduzorka (r1 i r2).

samo ponovljena istraživanja mogu da dokažu da li je neki istraživački nalaz bio plod puke koincidencije ili on zaista odražava određenu pravilnost u populaciji. Mada smo vidjeli da je istorija psihologije prilično mračna kako po pitanju prevalencije replikacionih studija, tako i po ukupnom broju replikovanih nalaza, postoje naznake da se situacija poboljšava u posljednje vrijeme. Replikativna istraživanja, ali i različite prateće mjere koje to omogućavaju (npr. naglasak na transparentnijem opisu metoda, preregistracije istraživanja, dostupnost podataka i kodova statističke analize, saradnja nezavisnih istraživačkih timova) su u porastu, mada je situacija još uvijek daleko od idealne[15].

Konačno, potrebno je imati u vidu da psihologija kao nauka nije homogena po pitanju uspješnosti replikacija. Pregledna studija[16] koja je analizirala radove objavljene u prethodnih 20 godina u šest najcitiranijih psiholoških časopisa je ukazala na vrlo zanimljive nalaze. Iako je procjena prosječnog broja replikovanih nalaza ispod 50% što sugeriše da je svaki drugi rad lažno pozitivan, pokazalo se da postoje značajne razlike među područjima psihologije. Konkretno, studije iz psihologije ličnosti, a zatim organizacione psihologije pokazuju višu stopu replikabilnosti (tj. 50% i više) nego socijalna psihologija i naročito razvojna psihologija koja se nalazile u donjoj grupi prema replikabilnosti (ispod 40%). Posebno značajan nalaz je bio da tip istraživanja ima još veći efekat na replikabilnost nego

[15] npr. Nosek et al., 2022, https://doi.org/10.1146/annurev-psych-020821-114157 i Torka et al., 2023, https://doi.org/10.1080/1359432X.2023.2206571

[16] Youyou et al., 2023, https://doi.org/10.1073/pnas.2208863120

oblast psihologije. Konkretno, korelaciona neeksperimentalna istraživanja su pokazala značajnu višu stopu replikabilnosti od eksperimentalnih istraživanja[17]. Međutim, i ta je podjela neraskidivo povezana sa veličinom uzoraka, jer poznato je da su eksperimentalne studije zahtjevnije za izvođenje, te po pravilu imaju manje uzorke. S obzirom na to, prikaz repertoara za zaključivanje o stvarnosti na osnovu istraživanja na uzorcima nastavljamo stavljanjem u fokus baš toga faktora. Sljedeće poglavlje je u potpunosti posvećeno izračunavanju potrebne veličine uzorka kako bi se postigla zadovoljavajuća pouzdanost nalaza.

[17] Pritom, u psihologiji ličnosti je procentualno mnogo više korelacionih istraživanja nego u socijalnoj psihologiji

Poglavlje 15

Statistička moć testiranja u zavisnosti od veličine uzorka

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta je statistička moć i zašto su psihološka istraživanja često statistički "nejaka"?
- Koji sastojci su potrebni za izračunavanje statističke moći i koji dodatni faktori utiču na nju?
- Kako se konkretizuje alternativna hipoteza i šta je "najmanji efekat od interesa"?

Prethodno poglavlje nam je otkrilo da psihologija kao nauka ima problematičnu reputaciju usljed relativno slabe replikabilnosti rezultata. Takođe smo saznali i glavne krivce za takvo stanje. U motivacionom smislu su to kultura i politike naučnog objavljivanja koje forsiraju inovativnost naspram pouzdanosti (za naučnike nije dovoljno da budu samo visoko otvoreni za iskustva, trebalo bi da budu i vrlo savjesni). S tehnološke strane se za mnogo toga mogu okriviti istraživanja na malim uzorcima koja daju nepouzdane rezultate. Ali ne samo da nedovoljno veliki uzorci doprinose pojavi lažno pozitivnih rezultata, nego dovode i do situacija u kojima je $p > \alpha$ čak i kada je alternativna hipoteza istinita, jer veličina uzorka direktno utiče na p-vrijednost.

Šta je to statistička moć?

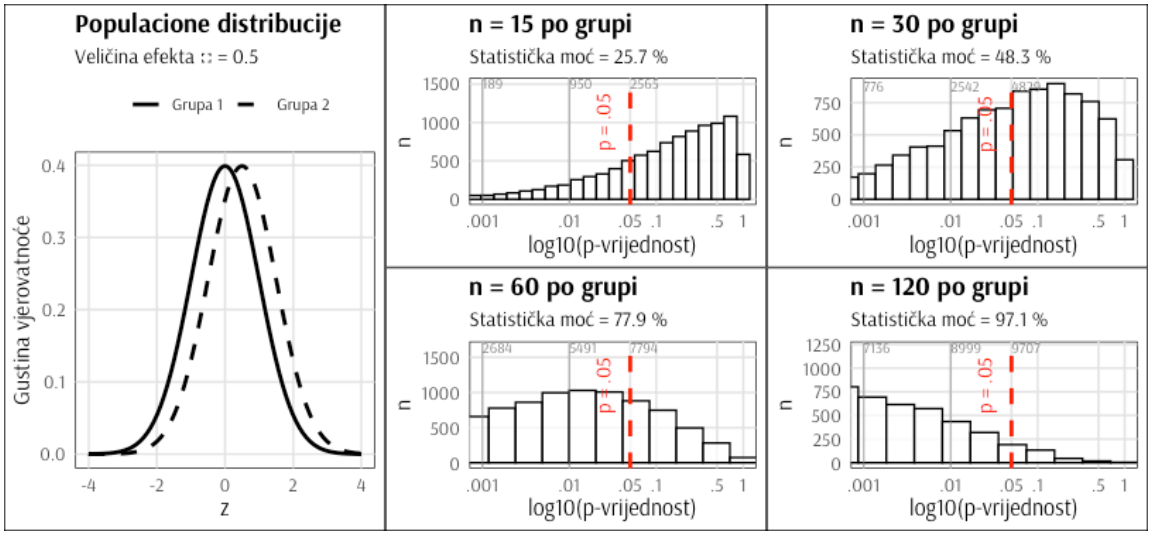
Statističari su odavno bili svjesni problema koje mali uzorci uzrokuju (10x brzo ponovite "uzorci uzrokuju"). Još 1930-ih su Jerzy Neyman i Egon Pearson (sin Karla Pearsona kojeg već znate) uspostavili temelje za koncept **statističke moći** (engl. *statistical power*) kojim se adresira pitanje adekvatne veličine uzorka

Definicija statističke moći

za testiranje neke hipoteze. Statistička moć je mjera vjerovatnoće da se donese odluka o odbacivanju nulte hipoteze pod uslovom da je alternativna hipoteza zaista istinita u populaciji. Drugim riječima, radi se o numerički izraženoj vjerovatnoći da se NE napravi Greška tipa 2, poznata i kao greška promašaja. Kako se Greška tipa 1 (lažno pozitivna odluka) vezuje za nivo statističke značajnosti α, pravedno je bilo i dodijeliti Greškama tipa 2 neko grčko slovo, pa budući da se ta greška obično smatra malo manje važnom, odlučeno je da to bude β. U pitanju je vjerovatnoća, pa se ona izražava na mjernoj skali od 0 do 1. Statistička moć određene analize se matematičkom notacijom računa kao $1 - \beta$. Dakle, ako je određen nivo Greške tipa 2 od $\beta = .05$, onda je statistička moć jednaka $1 - 0.05 = 0.95$ (što se čita: 95%).

Kao i p-vrijednost, statistička moć je primarno funkcija veličine efekta na koju kao istraživači ne utičemo direktno, nego je samo mjerimo, te veličine uzorka na koju možemo uticati. Slika 15.1[1] prikazuje kako izgledaju distribucije p-vrijednosti u zavisnosti od veličine uzorka, kada u populaciji postoji efekat $\delta = 0.50$.

[1] Inspirisana radom Halsey et al., 2015, https://doi.org/10.1038/nmeth.3288



Slika 15.1: Distribucija p-vrijednosti rezultata u zavisnosti od veličine uzorka kada je delta = 0.50 (10000 simulacija za svaku veličinu uzorka).

Međutim, psiholozi su još od 1960-ih počeli da uviđaju koliko štete prave potkapacitirane (engl. *underpowered*) studije, te koliko su one zaista sveprisutne. Opet onaj Jacob Cohen je ranih 1960-ih objavio pregledni članak[2] u kojem je pokazao da su studije objavljene u časopisu *Journal of Abnormal and Social Psychology* tokom 1960. godine imale nešto manje od 50% moći (tačno 0.46) da detektuju statistički značajne rezultate ako u populaciji postoji stvarni efekat umjerenog intenziteta. Drugim riječima, izvesti studiju i očekivati $p < .05$ je bilo isto vjerovatno kao i

[2] Originalni članak: Cohen, 1962, https://doi.org/10.1037/h0045186, kasnija replikacija: Sedlmeier & Gigerenzer, 1989, https://doi.org/10.1037/0033-2909.105.2.309

baciti novčić u vazduh i očekivati da padne na "pravu" stranu. Pritom, gotovo svi objavljeni radovi su imali statistički značajne rezultate?! Situacija se nije popravila ni više od dvadeset godina kasnije kada su u tom istom (samo preimenovanom časopisu koji se sad zove Journal of Abnormal Psychology) drugi istraživači otkrili da je srednja vrijednost statističke moći za umjerene efekte još niža (.37)! Štaviše, drugi radovi procjenjuju da je situacija gora u drugim disciplinama kao što neuronauke ($8\% - 31\%$)[3], ekonomija (18%)[4] i biomedicinska istraživanja ($17\% - 20\%$)[5].

[3] Button et al., 2013, https://doi.org/10.1038/nrn3475

[4] Ioannidis et al., 2017, https://doi.org/10.1111/ecoj.12461

[5] Dumas-Mallet, 2017, https://doi.org/10.1098/rsos.160254

Čemu služi statistička moć?

Izračunavanje statističke moći pomaže na nekoliko načina. Kao prvo, ako shvatimo da ne možemo postići dovoljno veliku statističku moć onda je bolje da istraživanje i ne izvodimo. Uštedićemo svoje vrijeme i energiju, novac finansijera, dobru volju i vrijeme ispitanika. Odustajanje od potkapacitiranog istraživanja će nas spasiti od iskušenja da bjesomučno lovimo niske p-vrijednosti analitičkim akrobacijama i time doprinosimo urušavanju povjerenja u nauku. S druge strane, izračunavanje statističke moći nas tjera da ozbiljnije shvatimo svoje istraživanje i ako možemo da dođemo do željenog nivoa, trudićemo se da se držimo tog plana. Prilaganje dokaza o unaprijed izračunatoj statističkoj moći potencijalnim finansijerima projekta ili akademskim supervizorima ostavlja utisak profesionalnosti. Konačno, kao što smo vidjeli, statistička moć se može analizirati i za već obavljena istraživanja. Dobijeni rezultati nam omogućavaju da zaključimo koliko pouzdanja možemo imati u njihove nalaze.

Kolika bi trebalo da bude ta moć?

Cohen, kao čuveni postavljač granica, je još davno utvrdio da ne bi trebalo ni izlaziti na teren ako nemamo barem 80% statističke moći. To bi značilo da je maksimalna $\beta = .20$, odnosno da očekujemo da bi, na duge staze, u četiri od pet izvedenih istraživanja dobili statistički značajne rezultate. U naučnoj zajednici se ovo obično smatra najnižom granicom. Postoje zagovornici striktnijeg ograničenja Greške tipa 2, pa se u literaturi pojavljuju i granice od 90% i 95%. U idealnom slučaju se statistička moć približava 100%, ali mnogo faktora utiče na to koliko ispitanika je moguće ispitati (npr. ekonomski i drugi troškovi istraživanja, dostupnost populacije, motivisanost ispitanika), te postoje i argumentacije da bi za svako pojedinačno istraživanje trebalo definisati odgovarajuće α i β nivoe, a ne da se po automatizmu preuzimaju zadane

vrijednosti[6]. Za početak, držite se mete od 80%.

[6] Lakens et al., 2018 https://doi.org/10.1038/s41562-018-0311-x

Možemo li mi to izračunati i kako?

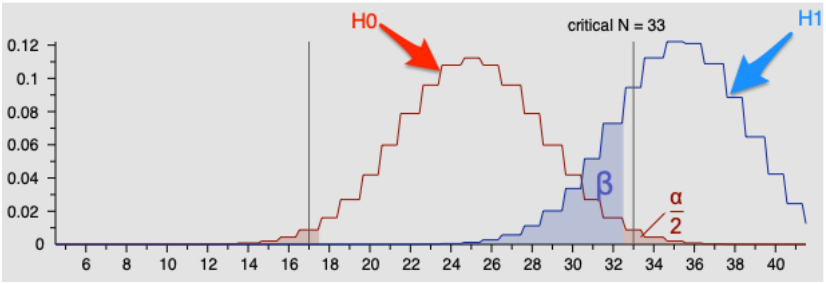
Ako sam se već o tome raspisao onda je odgovor očito potvrdan. Međutim, postoji začkoljica koju moramo riješiti. Sjećate se već da smo pominjali da je nulta hipoteza jedinstvena, a da postoji beskonačno mnoštvo alternativnih hipoteza. E pa, moraćemo se odlučiti za jednu od njih, odnosno operacionalizovati konkretnu alternativnu hipotezu s kojom direktno suočavamo nultu hipotezu. Dakle, izračunavanje slijedi četiri koraka od kojih su vam prva dva poznata:

1. Kreira se hipotetička distribucija statistika pod H_0
2. Odredi se nivo značajnosti (α) i na osnovu njega kritični skor
3. Kreira se hipotetička distribucija statistika pod H_1
4. Izračuna se procenat / proporcija površine distribucije za H_1 koji je viši od kritičnog skora dobijenog za H_0

Zvuči apstraktno, pa je stvari najbolje nacrtati. A kad već pojednostavljujemo onda ćemo se vratiti i najprimitivnijem vidu nacrta sa samo jednom binarnom kategoričkom varijablom. Vjerujem da se sjećate da se u tom slučaju koristi binomna distribuciju kako bi se predstavila hipotetička distribuciju odabrane kategorije (npr. žena među studentima psihologije) na uzorcima, ako je nulta hipoteza istinita u populaciji. Za njeno iscrtavanje je dovoljno da znamo veličinu uzorka (držaćemo se tadašnjeg primjera: $n = 50$) i populacioni parametar ($H_0 : \pi = .50$). Zasad ćemo uzeti proizvoljnu vrijednost za alternativnu hipotezu $H_1 : \pi = .70$ (a kasnije ćemo se vratiti pitanju njenog odabira). To je njen parametar, a veličina uzorka je logično jednaka kao i za nultu hipotezu ($n = 50$). Time smo dobili sve što je potrebno da i za nju iscrtamo binomnu distribuciju očekivanih statistika.

Pogledajmo sada Sliku 15.2. Crvenom bojom je iscrtana očekivana distribucija pod nultom hipotezom, identična onoj koju smo prikazivali kada smo učili kako se izračunavaju p-vrijednosti. Kao i tada, mod distribucije je 25 (što je 50% kada uzorak ima 50 ispitanika), distribucija je simetrična i vidno se rasteže od 14 do 36, nakon čega vjerovatnoće postaju tako male da prestaju biti vidljive. Primijetite i dva crveno osjenčena područja koji predstavljaju regione statističke značajnosti ($\alpha/2$) koji predstavljaju krajnjih 2.5% distribucije sa obje strane distribucije, odnosno ukupno 5%. Jasno je označena i konkretna kritična vrijednost ($n_X = 33$) kada je $p < \alpha$, konkretno $p = .033$.

(Ako se pitate zašto nije tačno .05, razlog je što se ovdje bavimo diskretnom, tačkastom distribucijom gdje zaokruživanje ne može biti savršeno. Možete da primijetite i malo odstupanje u sjenčenju, jer postoji i crveno osjenčen dio koji se nalazi sa lijeve strane kritične vrijednost).



Slika 15.2: Računanje statističke moći za nacrt sa jednom binarnom varijablom.

Ono što je na Slici 15.2 novost u odnosu na ranije poglavlje jeste da smo plavom linijom predstavili i distribucija statistika pod uslovom da je alternativna hipoteza istinita ($H_1 : \pi = .70$). Ta distribucija je asimetrična, kakve su i sve binomne distribucije koje imaju parametar različit od $\pi = .50$. Njen mod se nalazi na $n_X = 35$ što je logično s obzirom na to da je $35/50 = 0.70$, tačno koliko iznosi pretpostavljeni parametar. Za razliku od distribucije statistika pod H_0, osjenčeni plavi dio koji predstavlja nivo Greške tip 2 (β) se nalazi samo sa strane koja se suočava sa nultom hipotezom. Nas i interesuje samo preklapanje suprotstavljenih hipoteza. U ovom konkretnom slučaju osjenčeni dio predstavlja 22% distribucije pod H_1 raspodjelom statistika.

Ali šta tačno označava taj plavo osjenčeni dio? On govori o procentu H_1 distribucije statistika koja se preklapa sa srednjih 95% distribucije pod H_0. Drugim riječima, to je procenat rezultata koji se mogu desiti ako je istinita $H_1 : \pi = .70$, a da p-vrijednosti - koje se dobijaju na osnovu H_0 - budu veće od $\alpha = .05$. Na primjer, ako bismo na uzorku od 50 ispitanika imali 32 ženu, p-vrijednost bi bila .065. Mada je vjerovatnije da je taj rezultat proistekao pod H_1 nego H_0 (o tome vam govore visine distribucija kada je $n_X = 32$), ne bismo mogli da odbacimo H_0. Međutim, ako je ova H_1 zaista istinita, očekivali bismo da u 78% slučajava odbacimo H_0, odnosno da u nešto više od tri četvrtine slučajeve na uzorku dobijemo $n_X \geq 33$ i $p < .05$.

Kako se obično konkretizuje alternativna hipoteza?

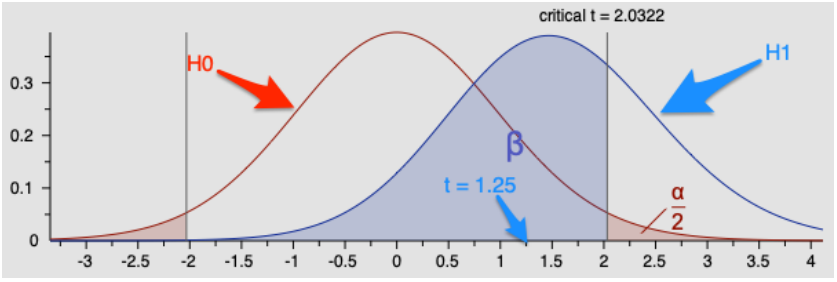
U prošlom primjeru sam potpuno proizvoljno postavio alternativnu hipotezu. To tako ne ide u nauci. Postoji nekoliko načina kako se istraživači odlučuju za konkretnu alternativnu hipotezu. Neki od njih su lošiji (i daleko popularniji), a neki bolji (ali dosta zahtjevniji). Međutim, potrebno je naglasiti da je i najlošiji način izračunavanja statistički moći bolji od neizračunavanja, jer ipak time stičemo bitne uvide. Dakle, ubjedljivo najpopularniji način je da se poslužimo Cohenovim granicama malog, umjerenog i velikog efekta, odnosno da se jedan od njih postavi kao parametar alternativne hipoteze, u zavisnosti od toga šta mislimo da je bitno istražiti. E sad, ono što je bitno da shvatite jeste da odabirom određenog parametra ne mislimo da je on tačan (to ne možemo uopšte da znamo), nego time odabiramo najmanji efekat za koji mislimo da ga je vrijedno ispitivati datim istraživanjem. Dobro ste shvatili, odabirom parametra alternativne hipoteze mi definišemo najmanju veličinu efekta od interesa (engl. *smallest effect size of interest* ili *minimal important difference*) i postavljamo je kao referentnu tačku koju provjeravamo. To znači da ako smatrate praktično značajnom razlikom u polnom udjelu studenata već razliku od 55% naspram 45%, onda bi trebalo da postavite $H_1 : \pi = .55$; a ako mislite da bitna razlika tek počinje kada imate odnos 2:1 onda odredite da je $H_1 : \pi = .66$.

Sad ćemo na primjeru sa sugestopedijom isprobati kako to ide sa Cohenovim pragovima razlika između aritmetičkih sredina. Recimo da smo pretpostavili da praktičan značaj imaju barem umjereni efekti dejstva barokne muzike na učenje, odnosno da bi u populaciji najmanji efekat od interesa bio pola standardne devijacije ($\delta = 0.50$). Željeli bismo da naknadno, odnosno *post hoc*, provjerimo koliku smo moć imali u pomenutom istraživanju za koje nisu dobijeni statistički značajni rezultati. Da vas podsjetim, u obje grupe smo imali po 18 ispitanika, a konačan rezultat je bio: $t(34) = 1.25, p = .22$. Budući da znamo veličinu uzorka obje grupe (odnosno broj stepeni slobode t-distribucije) i da standardne devijacije možemo procijeniti na osnovu rezultata sa uzorka, možemo konstruisati hipotetičke distribucije statistika za H_0 i H_1. Radi se o t-distribucijama, a finalni statistik koji je dobijen je t-skor. Inače, ako to nekog zanima, formula za t-skor za dva nezavisna uzorka, kao i formula koja konvertuje d u t (a i obrnuto je moguće) je sljedeća:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}} = d \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

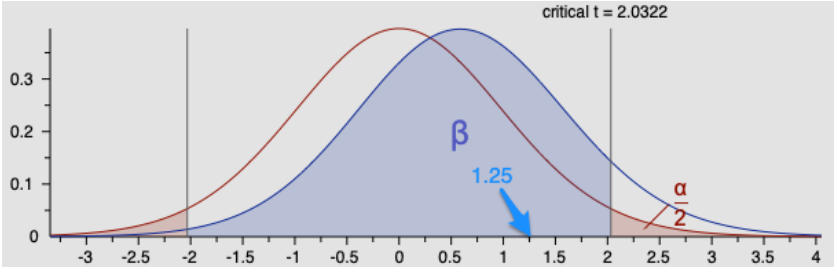
Iz prve formule možete da primijetite da se – baš kao i u slučaju sa t-testom za jednu dimenzionu varijablu – razlika aritmetičkih sredina dijeli standardnom greškom te razlike. U slučaju sa samo jednim uzorkom je standardna greška bila samo $\frac{SD}{\sqrt{n}}$, što je zapravo jednako $\sqrt{\frac{SD^2}{n}}$. Budući da ovdje imamo dvije grupe koje imaju potencijalno drugačiju veličinu uzorka i standardne devijacije, to se mora uzeti u obzir u formuli. U drugom dijelu formule više nemamo posebne standardne devijacije po grupama, zato što se one već uzimaju u obzir pri izračunavanju d-statistika.

Slika 15.3 pokazuje situaciju kada u uzorku imamo ukupno 36 ispitanika, podjednako raspoređenih u dvije grupe i kada procjenjujemo statističku moć t-testa za nezavisne uzorke da detektuje populacioni parametar razlike od ($\delta = 0.50$). Možete da vidite da je više od pola plave distribucije osjenčeno, odnosno da zalazi u Grešku tip 2. Konkretno, radi se o 69% distribucije, dok je statistička moć tek 31%. Takođe je označen i dobijeni skor na uzorku $t(34) = 1.25$ za koji je dobijeno $p = .22$. I u ovom slučaju vidimo da je dobijeni rezultat vjerovatniji ako je istinita $H_1 : \delta = 0.50$, nego ako je istinita $H_0 : \delta = 0.00$. Drugim riječima, shvatili smo da smo jednostavno imali premalo ispitanika ako smo htjeli ispitivati ovu hipotezu.



Slika 15.3: Statistička moć za t-test za nezavisne uzorke kada obje grupe imaju po 18 ispitanika, a veličina efekta u populaciji je $\delta = 0.50$.

Da uvijek može gore, pokazuje Slika 15.4. Na njoj je prikazana situacija sa istim testom i veličinom uzorka, ali pod pretpostavkom da alternativna hipoteza predstavlja mali efekat, tj. $\delta = 0.20$. U tom scenariju je statistička moć pišljivih 9%, odnosno kada bismo ponavljali iznova istraživanje na slučajnim uzorcima iste veličine, tek svaki jedanaesti put bismo dobili statistički značajnu razliku. I ponovo je dobijeni skor na uzorku $t(34) = 1.25$ vjerovatniji za definisanu H_1 nego za H_0. Sve u svemu, ako zapravo postoji mali efekat, istraživanje sa ovako malim uzorkom nam gotovo sigurno neće to pokazati putem p-vrijednosti.



Slika 15.4: Statistička moć za t-test za nezavisne uzorke kada obje grupe imaju po 18 ispitanika, a veličina efekta u populaciji je $\delta = 0.20$.

Čvršći uvid u mehanizam statističke moći i njenu zavisnost od veličine efekta i veličine uzorka možete steći ako se poigrate sa onlajn dostupnim interaktivnim appletima. Preporučujem vrlo preglednu kreaciju rpsychologist.com/d3/nhst Kristof Magnussena, kao i interaktivnu R aplikaciju istats.shinyapps.io/power/ napravljenu uz udžbenik The Art and Science of Learning from Data (autori Agresti i saradnici, 2021).

Koji su drugi načini za konkretizovanje alternativne hipoteze?

Gore opisani bi bili slučajevi *post hoc* izračunavanja statističke moći što je daleko od optimalnog. Mnogo više smisla ima isplanirati statističku moć prije izvođenja istraživanja, odnosno *a priori*. I u tom slučaju najveći broj istraživača pribjegava Cohenovim graničnim skorovima ili empirijski utvrđenim veličinama efekata unutar određene oblasti istraživanja. Međutim, idealno bi bilo definisati prag razlike u efektu koji ima neki praktičan značaj. Postoje radovi koji detaljnu izlažu kako odrediti minimalno važnu razliku.[7] Zanimljivo je da je motivacija poticala iz medicinskih istraživanja gdje su pacijenti zainteresovani da odrede osjetan stepen razlike između stanja prije ili poslije određenog tretmana (npr. smanjivanje stepena boli, smanjivanje vidljivosti dermatološkog problema, subjektivan osjećaj poboljšanja funkcije).

[7] Vidjeti npr. Anvari & Lakens, 2021, https://doi.org/10.1016/j.jesp.2021.104159 i King, 2011, https://doi.org/10.1586/erp.11.9

Među različitim pristupima je najpopularnija tehnika koja se naziva **metoda globalne procjene promjene** (engl. *global rating of change*). U pilot studiji se ispitanicima u dva vremenska navrata zada određeni instrument kojim se procjenje predmet mjerenja (npr. samoprocjena depresivnosti, anksioznosti, fokusiranosti pažnje, samopoštovanja). U drugom mjerenju se zadaje i stepen globalne evaluacije vlastitog stanja kojim se ispitanik pita u kojoj mjeri se osjeća drugačije u odnosu na prvo mjerenje. Najčešće se nude odgovori na to pitanje na petostepenoj ordinalnoj skali sa opcijama: "dosta gore", "nešto gore", "isto", "nešto bolje" i "dosta bolje". Unutar svake kategorije promjena (npr. za sve ispitanike koji su naveli da se osjećaju "nešto bolje") se onda može izračunati prosječna ili srednja razlika između dva mjerenja na zadanom instrumentu što onda određuje najmanju veličinu efekta od interesa. Naravno, postupak koji koristi evaluaciju subjektivnog osjećanja je prikladan za samo neke predmete mjerenja, ali se mogu zamisliti i njegove adaptacije za predmete mjerenja koji nemaju jasnu valencu za ispitanika,

odnosno kada ispitanici ne saopštavaju svoj osjećaj ugodnosti ili neugodnosti. Postoji mogućnost da se umjesto od ispitanika informacije uzimaju od eksperata u oblasti, ali o tom pristupu ćemo nešto kasnije u knjizi.

Od kojih još sastojaka zavisi statistička moć?

Daleko najveći efekat na statističku moć imaju veličina uzorka i pretpostavljena veličine efekta u populaciji. Ipak, ona ovisi i od sljedećih elemenata[8]: nivoa statističke značajnosti (α), varijabilnosti rezultata, vrste nacrta (ponovljena ili neponovljena mjerenja) i izbora konkretnog statističkog testa.

[8] Na statističku moć utiču i psihometrijski faktori kao što su valjanost i pouzdanost mjerenja.

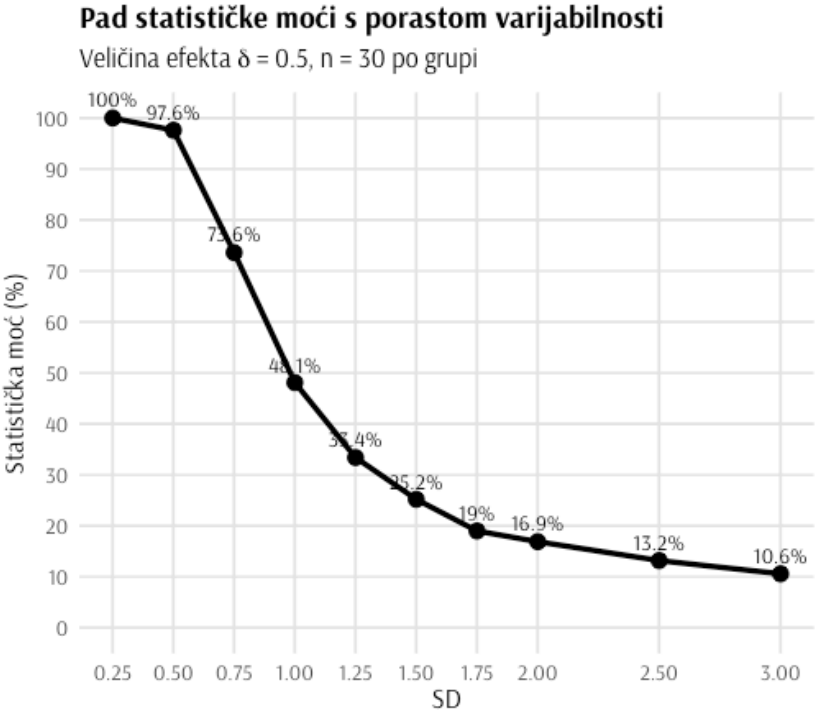
Kada je u pitanju statistička značajnost već smo pomenuli da taj nivo ne mora biti uvijek .05. Često se postavljaju oštrije granice, kao što su .01 ili .005, ali smo saznali da postoje zalaganja da se za svako istraživanje argumentuje nivo značajnosti, koji bi u teoriji mogao biti i opušteniji (samo ću napomenuti da povećana opuštenost teško prolazi kod većine recenzenata koji ocjenjuju kvalitet rada). Ovo je idealno mjesto da se pozabavimo problemom višestrukog testiranja. Nadam se da se sjećate da sa porastom broja statističkih testiranja, šansa da barem jedan rezultat bude lažno pozitivan raste iznad postavljenog nivoa statističke značajnosti. Drugim riječima, već kada imamo dva testiranja, ako je H_0 istinita, postojaće vjerovatnoća veća od nominalne da se dobije statistički značajan rezultat. Kako bi se prevenirali lažno pozitivni rezultati statističari su domislili korekcije nivoa značajnosti. Najpoznatija i najbrutalnija među njima je Bonferroni korekcija gdje se postavljeni nivo značajnosti dijeli sa brojem statističkih testova koji će biti izračunat. To znači da ako imamo petnaest hipoteza koje testiramo u istraživanju, korekcija zaoštrava nivo značajnosti: $\alpha = 0.05/15 = 0.0033$. Odnosno, ako se obavežete na Bonferroni korekciju u takvom slučaju onda će rezultati biti formalno proglašeni statistički značajnim tek ako je $p < .003$. Logično, u želji da se kontroliše Greška tipa I istovremeno se povećava mogućnost Greške tipa II, tj. smanjuje se statistička moć testiranja. Postoji cijeli dijapazon korektivnih procedura koje pokušavaju da naprave bolji balans između kontrolisanja vjerovatnoće lažno pozitivnih i lažno negativnih ishoda, ali to nadilazi svrhu ovog teksta.

Uloga nivoa statističke značajnosti

Što se tiče varijabilnosti, stvar je vrlo jednostavna: što je varijabilnost veća, statistička moć je manja. Ako se sjetite izlaganja iz prvog dijela knjige, veća raznolikost znači da mjere centralne tendencije slabije rade posao predstavljanja grupe, odnosno da

Uloga varijabilnosti

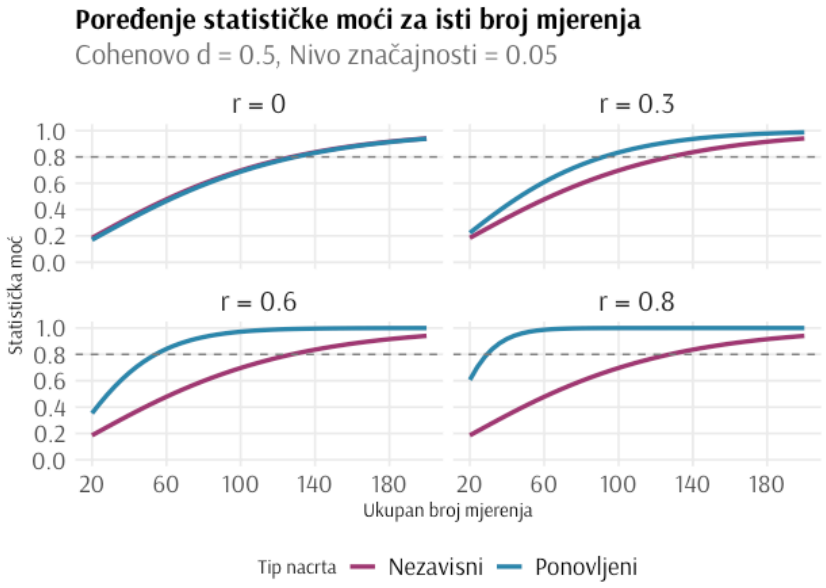
u podacima imamo manje pravog signala, a više šuma ili buke. Čim imamo više šuma, onda imamo i manje moći da detektujemo stvarni efekat. Slika 15.5 ilustruje pad statističke moći sa povećanjem standardne devijacije, a kada se broj ispitanika i veličina efekta drže konstantnim.



Slika 15.5: Promjene u statističkoj moći u zavisnosti od standardne devijacije.

Inače, jedan od najefikasnijih načina za smanjenje neželjene varijabilnosti jeste izbor nacrta sa ponovljenim mjerenjima. Budući da isti ispitanici učestvuju u svim uslovima, individualne razlike među njima se statistički kontrolišu i eliminišu kao izvor šuma. Slika 15.6 prikazuje zavisnost statističke moći testa od tipa nacrta u kojima imamo jednak broj mjerenja (npr. po 30 ispitanika u eksperimentalnoj i kontrolnog grupi ili 30 ispitanika koji prolaze i kontrolni i eksperimentalni uslov). Primijetićete da razlike u statističkoj moći nema ako nema korelacije između prvog i drugog mjerenja, dok sa povećanjem korelacije raste i statistička moć testova ponovljenih mjerenja. Dakle, zaključak je da su nacrti sa ponovljenim mjerenjima zaista statistički moćniji od nacrta sa nezavisnim grupama ako pretpostavljamo korelaciju između mjerenja istih ispitanika. Međutim, ponovljeni nacrt nije uvijek praktičan ili moguć, te donosi sa sobom i dodatne metodološke probleme (npr. efekte prenosa, vježbe i umora, probleme sa osipanjem uzorka), pa konačan izbor zavisi od prirode konkretnog

istraživačkog problema i dostupnosti ispitanika.



Slika 15.6: Statistička moć za t-test u zavisnosti od tipa nacrta.

Uloga vrste statističkog testa

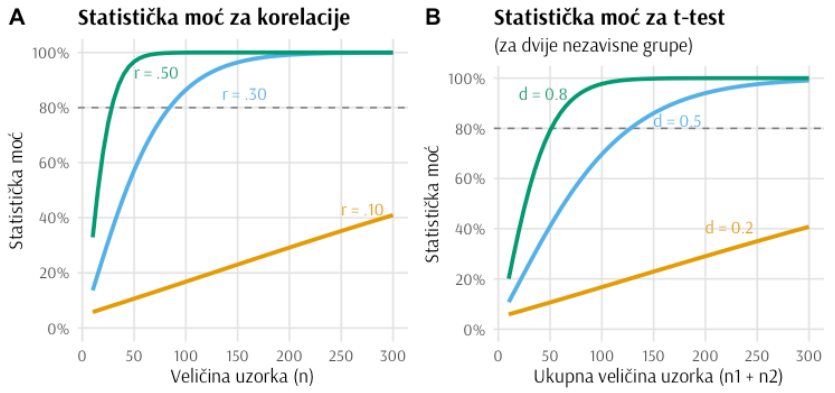
Kada su u pitanju statistički testovi koji se koriste u analizi, dosta davno (kada sam ja bio student) se izričito tvrdilo da su parametrijski testovi (npr. testovi za linearne korelacije, t-test na nezavisnim uzorcima) moćniji od svojih parnjaka koji koriste analogne procedure na rangovanim podacima (neparametrijski testovi). Međutim, pokazalo se da su neparametrijski testovi vidno moćniji ako imamo posla sa asimetričnim distribucijama i ekstremnim slučajevima, dok je prednost parametrijskih testova u idealnim uslovima (npr. normalnost raspodjela) zanemariva. Uzimajući u obzir i druge prednosti rangovanih podataka, i ovo možete doživjeti kao argument da se kada god je moguće ispita senzitivnost rezultata u pogledu korištene statističke tehnike, odnosno da se provjerava robusnost rezultata.

U kojem softveru da izračunamo statističku moć?

Danas postoji relativno veliki broj softverskih rješenja koja omogućavaju da vrlo jednostavno možete izračunati statističku moć za vlastita istraživanja. Već dugo vremena je najpopularniji i najcitiraniji softver *G*Power*.[9] Besplatan je, vrlo jednostavan za korištenje i daje iscrpne izvještaje za obavljene proračune. Uz to, daje i vizuelni prikaz distribucija koje ste u ovom poglavlju već vidjeli kada su u pitanju bili binomni i t-test za nezavisne uzorke. Tu je naravno i *R* koji ima sve što se tiče statistike, pa tako i niz

[9] https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

paketa koji su specijalizovani za izračunavanje statističke moći. Neki od tih paketa su ugrađeni u izvedenice *R*-a sa grafičkim interfejsom: *Jamovi* (jamovi.org) i *JASP* (jasp-stats.org). Kada se u nekoj dalekoj budućnosti susretnete sa kompleksnim modelima onda biste mogli ipak posegnuti za naprednijim *R* paketima jer oni mogu izvesti simulacije kompleksnih modela na osnovu vaših specifikacija. U *R*-u je izvedena i ilustracije na Slici 15.7 koja daje grubi uvid u zavisnost statističke moći od veličine uzorka za ispitivanje statističke značajnosti korelacija i za t-testove za nezavisne uzorke.



Slika 15.7: Odnos statističke moći i veličine uzorka za linearne korelacije i t-test za nezavisne uzorke.

Postoje li ograničenja upotrebe statističke moći?

Iako je očito da postoje jasne prednosti izračunavanja statističke moći i da se zahvaljujući tome mogu povećati replikabilnost i pouzdanost nalaza, naravno da postoje određena ograničenja. Kao prvo, samo mali procenat istraživanja unaprijed procjenjuje minimalno važan efekat, tako da se većina alternativnih hipoteza definiše, kako da kažem, *ofrlje*… Drugo, u jednom istraživanju se obično testira više različitih hipoteza. Čak i kada se parametri porede prema jednoj od njih, pitanje je da li se time pokriva i moć za druga testiranja koja se izvode. Potrebno je dakle unaprijed utvrditi koja je to "najslabija karika", odnosno za koju hipotezu je potrebno najviše ispitanika - što se obično ne radi. Treće, nekad jednostavno nemamo dovoljno resursa da postignemo odgovarajuću moć. Međutim, ipak ima smisla da obavimo istraživanje što će postati jasnije kada pročitate poglavlje o meta-analizama. Konačno, i najbitnije, izračunavanje statističke moći nam osigurava kontrolu grešaka pri zaključivanju o p-vrijednostima, ali se i dalje radi o dihotomnim zaključcima. Nešto ili jeste ili nije.

Ali to nije u skladu sa stvarnošću, kao ni sa jednom od osnovnih prednosti korištenja statistike u psihologiji, a to je *preciznost*. Sljedeće poglavlje je zato posvećeno upravo tom pitanju.

Poglavlje 16

Intervalne procjene parametara

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Kako se ispravno tumači interval povjerenja, koje su najčešće zablude i koji faktori utiču na njegovu širinu?
- Šta je margina greške i kako možemo unaprijed isplanirati potrebnu veličinu uzorka da bismo postigli željenu preciznost procjene?
- Šta je to intervalna nulta hipoteza i kako se pomoću nje može zaključivati o praktičnoj ekvivalentnosti, superiornosti ili inferiornosti nekog efekta?
- Koje su ključne prednosti korištenja intervala povjerenja u odnosu na klasično testiranje tačkaste nulte hipoteze (p-vrijednost)?

Otkako smo zamakli u teritoriju statističkog zaključivanja, bavili smo se isključivo dihotomnim ispitivanjem hipoteza. Time smo zanemarili jednu od temeljnih prednosti korištenja statistike u psihologiji navedenu na samom početku knjige, njenu sposobnost da se vrlo precizno izražava. Mada smo podsjećani da veličina efekta igra bitnu ulogu u zaključivanju, dosad izložene procedure statistike zaključivanja ih samo pominju kao sastojak za recept kojim se dobija binarni ishod u NHST okviru. Zaključci tih procedura nam govore da li je vjerovatno da postoji sistemsko dejstvo nezavisne varijable (npr. slušanja barokne muzike, psihoterapije) na zavisnu varijablu (npr. količinu naučenog, smanjenje simptomatologije), ali ne i kolika je ta uloga u populaciji. Začuđujuće je da je izuzetno dug period istorije naučne psihologije obilježen testiranjem nulte hipoteze putem p-vrijednosti uprkos činjenici da su bila dostupna znanja o tome

kako iskazati precizne procjene parametara. Šta je konkretno potrebno? Vrlo malo toga: procjenjivač populacije dobijen na uzorku ($\hat{\theta}$), odgovarajuća hipotetička distribucija statistika, te procjena standardne greške. To sve znate, zar ne? Da provjerimo.

Kako postižemo da na osnovu rezultata sa uzorka procijenimo parametar?

Da vas podstaknem na razmišljanje uzmimo za primjer ispitivanje toga koliko su inteligentni studenti psihologije. Recimo da smo na uzorku od 50 studenata koristili standardizovani test inteligencije i ustanovili prosječnu vrijednost $M_{psi} = 115$. Ako pretpostavimo da standardna devijacija $\sigma_{IQ} = 15$ važi i za subpopulaciju studenata psihologije, onda se standardna greška aritmetičke sredine u našem primjeru izračunava na sljedeći način: $\sigma_M = 15/\sqrt{50} = 2.12$.

Šta nam ovaj broj znači? Standardna greška nam govori da bi 2.12 IQ bodova trebalo biti prosječno odstupanje statistika (procjenjivača parametra, $\hat{\theta}$) od parametra (θ). S obzirom da znamo i da se kvocijenti inteligencije raspodjeljuju normalno, te da imamo i dovoljno veliki uzorak ($n = 50 > 25$) možemo da se oslonimo na **centralnu graničnu teoremu**. Dakle, ne znamo parametar (μ_ψ), ali imamo jednu od beskonačno mnogo mogućih procjena (M_ψ). Približno 95% takvih procjena se nalazi unutar dvije standardne devijacije od parametra, standardna devijacija te raspodjele je zapravo standardna greška aritmetičke sredine (σ_M).

Sad dolazi finalni dio: iako ne možemo znati da li je naša procjena sa uzorka veća ili manja od parametra, znamo da 95% aritmetičkih sredina uzoraka (različitih M_ψ) neće biti udaljenije od $1.96 * \sigma_M$ od aritmetičke sredine populacije (μ_ψ). Slikovitije izraženo, ako raširimo ruke na obje strane od 115 u širini $1.96 * \sigma_M$ postoji velika šansa da smo dodirnuli i populacionu aritmetičku sredinu, a da je i ne vidimo i da zapravo ne znamo gdje se tačno nalazi. Ovo "širenje ruku" na mjernoj skali podrazumijeva da smo ispod i iznad statistika postavili neke granične skorove, te time kreirali određeni interval ili raspon vrijednosti kojeg nazivamo interval povjerenja.

Šta su intervali povjerenja i kako se izračunavaju?

Formalna definicija kaže da **interval povjerenja** ili **interval pouzdanosti** (engl. *confidence interval*, *CI*) predstavlja raspon

na mjernoj skali koji je ograničen donjom i gornjom granicom, a unutar kojeg se, sa određenom tačnošću u dugom nizu uzorkovanja, obuhvata populacioni parametar. Kao istraživači mi možemo da podešavamo tačnost te procedure, a gore je opisan primjer najčešćeg, 95%-tnog intervala povjerenja. Zašto je on najčešći? Zato što je on komplementaran najčešćem nivou statističke značajnosti $\alpha = .05$. Drugim riječima, 95% interval povjerenja je $(1 - \alpha) * 100$ interval povjerenja kojeg izračunavamo na sljedeći način kada koristimo z-distribuciju[1]. Donju granicu dobijamo tako što od statistika oduzmemo standardnu grešku pomnoženu sa 1.96 (odnosno, odračunamo približno dvije standardne devijacije):

Definicija intervala povjerenja

[1] Principi su jednaki za računanje intervala povjerenja za druge nacrte od kojih neke prikazujem kasnije.

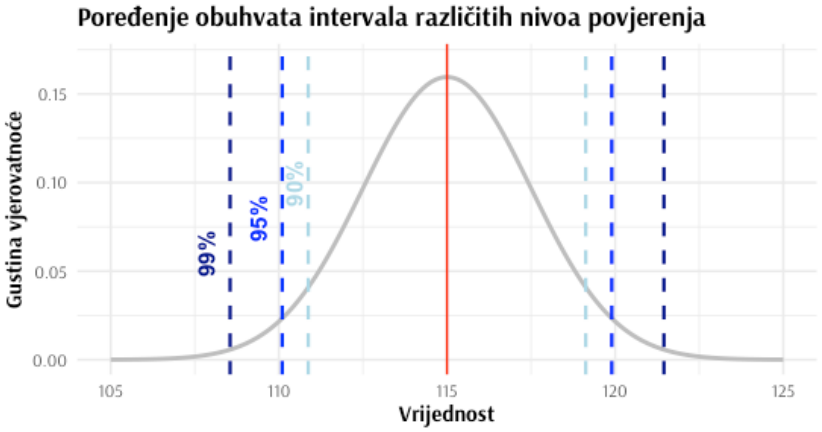
$$95\%CI_{d.g.} : M - (z_{95\%} * \frac{\sigma}{\sqrt{n}}) = 115 - (1.96 * \frac{15}{\sqrt{50}}) = 115 - 4.16 = 110.84$$

Analogno tome, izračunavamo i gornju granicu gdje na statistik dodajemo standardnu grešku pomnoženu sa 1.96:

$$95\%CI_{g.g.} : M + (z_{95\%} * \frac{\sigma}{\sqrt{n}}) = 115 + (1.96 * \frac{15}{\sqrt{50}}) = 115 + 4.16 = 119.16$$

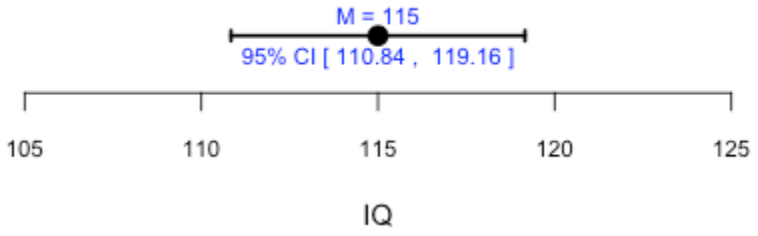
Rezultat se saopštava tako što uz statistik dodajemo – obično u uglastim zagradama – dobijeni interval povjerenja, pa bi to izgledalo ovako: $M = 115$, $95\%CI[110.84, 119.16]$. Pretpostavljam da već shvatate da je ovo već informativnije nego samo saopštiti p-vrijednost koja nam kaže da je izuzetno malo vjerovatno da je IQ populacioni parametar studenata psihologije jednak 100. Nakon što smo dobili ovakav rezultat znamo da ne samo da je slabo vjerovatno da je parametar 100, nego je slabo vjerovatno i da je 105, pa i 110, ali da je isto tako teško da je on 130, 125, pa i 120. Pravilno tumačenje intervala povjerenja ćemo ostaviti za malo kasnije, a sad ćemo se vratiti pitanju preciznosti procjene.

Objašnjeno je zašto se najčešće izračunavaju 95%-tni intervali povjerenja, ali treba naglasiti da se relativno često izračunavaju i 90%-tna kao i 99%-tna verzija. Da biste ih pješke izračunali za gornji primjer potrebno je samo da zamijenimo kritičnu vrijednost $z_{95\%} = 1.96$ sa odgovarajućim kritičnim vrijednostima $z_{90\%} = 1.65$, odnosno $z_{99\%} = 2.58$. U ovom slučaju bi rezultati bili: $90\%CI[111.51, 118.49]$ i $99\%CI[109.54, 120.46]$. Dakle, što je nivo povjerenja veći, to su veće i širine intervala. Odnosno, sve veći stepen pouzdanosti traži da obuhvatimo više mogućih rješenja (Slika 16.1).



Slika 16.1: Poređenje obuhvata intervala različitih nivoa povjerenja.

Kada se intervali povjerenja prikazuju grafički obično se statistik prikazuje nekim marker znakom (najčešće kružić ili kvadratić), a ručice obuhvata sa dvije tzv. T-linije (Slika 16.2).



Slika 16.2: Najčešći oblik grafičkog prikazivanja intervala povjerenja.

Koje još pravilnosti važe za intervale povjerenja?

Ako niste kristalno jasno shvatili mehanizam intervala povjerenja, ako sumnjate da bi takva magija mogla da funkcioniše i/ili ako želite da shvatite šta još utiče na širinu i tačnost intervala povjerenja, ovo je pravi trenutak da isprobate neki od appleta koji simuliraju dugoročno uzorkovanje aritmetičkih sredina. Na internetu ih ima mnogo, a ovdje navodim moje omiljene. Najpreglednija, ali i sa najmanje opcija koje vi možete mijenjati, je interaktivna animacija rpsychologist.com/d3/ci/ koju je kreirao Kristofer Magnusson. Meni najsimpatičnija je ribolovna simulacija zoology.ubc.ca/~whitlock/Kingfisher/CIMean.htm koja će vam omogućiti da shvatite kako na širinu intervala utiče i varijabilnost (veličina SD), dok će vam simulacija istats.shinyapps.io/ExploreCoverage/ kreirana uz udžbenik *The*

Art and Science of Learning from Data omogućiti uvid u efekat asimetričnosti distribucija na tačnost obuhvata. Igranje sa ovim appletima bi trebalo da vam pomogne da shvatite sljedeće važne zakonitosti:

1. **Što je uzorak veći, intervali povjerenja su uži**. Drugim riječima, veći uzorci teže da daju sigurniju procjenu, pa se to manifestuje i u širini obuhvata. Ako imamo mali uzorak imamo i veliku nesigurnost u parametar i ovo upada u oči čak i onda kada na osnovu p-vrijednosti odbacimo nultu hipotezu. Što nas dovodi do sljedećeg:
2. **Intervali povjerenja mogu da posluže kao direktna zamjena p-vrijednostima**. Već je naglašeno da su α i $(1-\alpha)*100\%$-tni intervali povjerenja komplementarni. To znači da ako je postavljen nivo značajnosti $\alpha = .05$, rezultat će biti proglašen statistički značajnim ako 95%-tni interval povjerenja ne obuhvata vrijednost 0[2]. Mnogi se zato pitaju šta će nam uopšte p-vrijednosti?
3. **Što je veća varijabilnost, širi su intervali povjerenja**. Moguće je da već sami izgovarate: "Veća raznolikost unutar uzorka podrazumijeva da bi trebalo da imamo manje povjerenje u mjere centralne tendencije, jer one slabije rade svoj posao predstavljanja grupe." Da, što je veća varijabilnost, veća je i standardna greška, te se povećava i vjerovatna udaljenost između statistika i parametra. Zato se intervali povjerenja rastežu kako bi mogli da pokriju šarenilo uzorka.
4. **Broj ispitanika i procijenjena varijabilnost ne utiču na postotak tačnosti procedure.** Čudno? Nije, ako shvatite da se intervali povjerenja prilagođavaju broju ispitanika i količini varijabilnosti tako da svaki 95%-tni interval povjerenja, u dugom nizu ponavljanja, hvata procjenjivani parametar. Naravno, veći uzorci i manja varijabilnost rezultiraju preciznijom (užom) procjenom parametra, ali omjer pogrešnih u odnosu na tačne zaključke je isti kao kada imate manje uzorke i veću varijabilnost.

[2] Vrijednost 0 ili neku drugu vrijednost koja je postavljena za nultu hipotezu. Na ranijem primjeru sa brojem žena na studiju psihologije parametar je bio 50%. Ako 95%-tni interval povjerenja ne obuhvata tu vrijednost (npr. 95\%CI [55.0%, 75.0%]), onda možemo zaključiti da postoji statistički značajna razlika na nivou .05.

Postoje li problemi sa upotrebom intervala povjerenja?

Naravno da postoje, pa ćemo sada upoznati i te neke mračnije aspekte. Oni se tiču prvenstveno zabluda u tumačenju toga šta su intervali povjerenja. Takođe, vidjećemo da u nekim slučajevima u te intervale povjerenja ne možemo imati baš toliko povjerenje koliko se nominalno tvrdi.

Koja je najčešća zabluda kada je u pitanju tumačenje intervala povjerenja?

Za početak, pogrešno je reći da možemo biti uvjereni sa 95% vjerovatnoće ili sigurnosti da se traženi parametar nalazi unutar 95%-tnog intervala povjerenja. Zašto? Najprije zato što mi ne znamo da li je na našem uzorku dobijen interval koji uključuje ili ne uključuje parametar. Prisjetite se toga da je vjerovatnoću potrebno doživljavati kao nivo uvjerenja. Primjer bacanja novčića zgodno ilustruje problem pogrešnog shvatanja intervala povjerenja.[3] Recimo da vaš drugar baci novčić 100 puta i da 61 put novčić padne na jednu od strana. Za taj ishod p-vrijednost je .035, a 95%-tni interval povjerenja seže od 50.7% do 70.6%. Da li bi nakon toga trebalo da budemo uvjereni sa 95% vjerovatnoće da je stvarni parametar u tom regionu, a da nije okruglo 50.0%? Naravno da ne, jer još uvijek snažnije vjerujemo da je parametar 50.0% što smo vjerovali i prije bacanja novčića, pa pretpostavljamo da smo svjedočili slabije vjerovatnom događaju pod nultom hipotezom. Ako se naš stepen uvjerenja i promijenio na osnovu drugarovih bacanja, on ne bi trebalo da dostigne baš toliko visok nivo od 95%. Mijenjanje stepena uvjerenosti će biti posebna tema jednog kasnijeg poglavlja.

[3] Primjer preuzet iz Lakensovog udžbenika *Improving your statistical inferences*, 2022, https://lakens.github.io/statistical_inferences/07-CI.html

Šta je onda ispravno tumačenje intervala povjerenja?

Ispravno tumačenje intervala povjerenja

Ispravno tumačenje kaže da će – ako su zadovoljene sve pretpostavke (tj. slučajnost odabira uzorka, nepristrasnost procjene standardne greške) – **95%-tni interval povjerenja u 95% slučajeva beskonačno ponavljanog uzorkovanja sadržavati populacioni parametar**. Dakle, stepen tačnosti se odnosi na proceduru izračunavanja intervala, a ne i na konkretni interval koji je dobijen. Hm, to i ne zvuči više tako sjajno kako smo u početku zamišljali. Uprkos tome entuzijazam za intervale povjerenja ne bio smio potpuno da splasne, budući da ipak dobijamo dodatne informacije iz te procedure. Možda nećemo vjerovati sa tako visokim nivoom u dobijeni interval, ali bi ipak trebalo da izmijenimo stepen uvjerenosti o stvarnoj populacionoj vrijednosti. Ako smo prosudili da je uzorak relativno reprezentativan, onda možemo pretpostaviti da je parametar "tu negdje", odnosno da se ne nalazi predaleko od procjene, čak i ako je zaista izvan intervala povjerenja.

Koje još zablude postoje kada je u pitanju tumačenje intervala povjerenja?

Imajte na umu da postoji još pogrešnih tumačenja intervala povjerenja. Jedno – totalno pogrešno, ni pod kojim uslovom ne pomišljajte na to – je da će se 95% budućih pojedinačnih skorova nalaziti unutar takvog intervala povjerenja. To su zapravo **intervali predviđenih skorova** (engl. *prediction interval*). Oni se mogu izračunati, blago komplikovanijom procedurom koja rezultuje dosta širim intervalima. Intervali predviđanja su širi jer postupak mora da uzme u obzir dvije vrste nesigurnosti: 1) nesigurnost u procjenu lokacije populacione sredine (što i jeste interval povjerenja) i 2) dodatnu nesigurnost koja potiče iz varijabilnosti pojedinačnih skorova oko te sredine. Drugim riječima, dosta je teže predvidjeti gdje će pasti jedan budući, nasumični skor, nego procijeniti gdje se nalazi prosjek svih rezultata.

Drugo fališno tumačenje je manje opasno. Njime se tvrdi da će se budućih 95% aritmetičkih sredina uzoraka nalaziti unutar tog intervala. Ovo i nije tako katastrofalno jer simulacije pokazuju da se to dešava u otprilike 83% slučajeva, ali na duge staze, odnosno u prosjeku.[4] Tako da je najbolje da potpuno zaboravite na ova tumačenja (mada pasus neću obrisati).

[4] Dokazi za tih 83%, kao i za rasprostranjenost ovog pogrešnog shvatanja prikazani su u Cumming, 2010, https://doi.org/10.1207/s15328031us0304_5; dok Morey et al., 2016, https://doi.org/10.3758/s13423-015-0947-8 u svom članku pokazuju i da autori prethodnog članka ne znaju tačno da tumače intervale povjerenja!?

Postoje li neki tehnički problemi sa izračunavanjem intervala povjerenja?

Postavlja se pitanje da li intervali povjerenja mogu griješiti i u tehničkom smislu, tj. da se njima navodi da su to 95%-tni intervali, ali da u idealnom slučaju u manjem procentu hvataju parametar. Iznad smo prolazili primjer u kojem smo koristili z-distribuciju jer smo znali standardnu devijaciju populacije. Kao što već znate, u ogromnom broju slučajeva poznavanje standardne devijacije populacije je luksuz, pa i nju procjenjujemo pomoću uzorka. To znači da imamo više nepoznatih, te da moramo da radimo sa većim oprezom. Zato u slučajevima kada su aritmetičke sredine u fokusu (npr. kada procjenjujemo parametar za jedan uzorak μ_1, procjenjujemo parametar razlike između dvije grupe δ) koristimo t-distribucije koje imaju blago "deblje repove" od z-distribucija. To čini naše intervale nešto širim, naročito kada su u pitanju vrlo mali uzorci ($n < 20$). Statistika tako uzima u obzir dodatnu neizvjesnost i kaže nam "budući da moramo pogađati i standardnu devijaciju, moramo biti malo oprezniji s našim zaključcima." Srećom, simulacije podataka pokazuju da tačnost tako korigovane procedure odgo-

vara ciljanim nivoima, ali ne treba zaboraviti da smo ponovo pretpostavili normalnu distribuciju inteligencije u populaciji.

Pokazalo se da intervali povjerenja podbacuju onda kada uz relativno male uzorke istovremeno postoji i značajnu asimetrija skorova u populaciji. Dobro je podsjetiti se da su mnoge popularne psihološke mjere značajno asimetrične: klinička simptomatologija, samopoštovanje i zadovoljstvo životom, socijalno poželjne crte ličnosti. Ako bi se neke preporuke mogle dati onda su to sljedeće. Uopšte ne bi trebalo vjerovati intervalima povjerenja ako su oni izvedeni na uzorcima sa manje od 25 ispitanika. Na uzorcima koji imaju između 25 i 100 ispitanika asimetrične distribucije obaraju tačnost obuhvata, ali ne tako strašno i obuhvat je najčešće između 90% i 94%. Međutim, asimetričnost populacione distribucije može ozbiljnije oboriti tačnost procedure kada se intervali povjerenja računaju za univarijatne kategoričke procjene, korelacije i razlike između grupa. Postoje lijekovi za ovaj problem čiji sastav nećemo analizirati jer to pitanje nadilazi predviđeni nivo knjige (a ionako bolje da čuvam vaše neurone za dileme sa kojima ćete se češće susretati).[5] Ako vam ipak u životu zatreba da izračunate intervale povjerenja za asimetrične distribucije ($Sk > 0.5$ je već indikativno) konsultujte Sliku 16.3 i potražite odgovore na internetu.

[5] Često se preporučuju bootstrap metode koje koriste računarsku moć da iz postojećeg uzorka ponavljanim uzorkovanjem kreiraju hiljade novih vještačkih uzoraka na kojima se izračunavaju intervali povjerenja.

Parametar	Preporučeni način izračunavanja intervala povjerenja
μ	Ponavljani uzorci sa zamjenom (10000 uzoraka) sa dodatnim korekcijama za pristrasnost i porast greške (engl. bias-corrected and accelerated bootstrap)
δ	
ρ	Izračunati intervale povjerenja za Spearmanovu rang korelaciju (F metod)
π	Koristiti Wilsonovu formula za intervale povjerenja

Slika 16.3: Preporučene metode izračunavanja intervala povjerenja u slučaju asimetričnih distribucija i manjih uzoraka.

Sve u svemu, uprkos nekih ograničenja, intervali povjerenja su vrijedna alternativa p-vrijednostima. Mogu da se izračunaju kako za sve jednostavne, tako i za mnoge kompleksne parametre, kako za originalne skale, tako i za standardizovane skale. Na sve to, intervali povjerenja omogućavaju još dvije bitne procedure u statističkom zaključivanju. Prva je planiranje statističke moći u kontekstu određivanja preciznosti mjerenja, a druga omogućavanje testiranja intervalne nulte hipoteze.

Kako možemo planirati preciznost mjerenja?

Zamislimo da želite da osnujete psihološko savjetovalište, ali da potencijalni finansijer tog projekta želi da mu pokažete jasne brojke o procentu studenata koji ima za to potrebu i koji bi bio motivisan da dolazi u savjetovalište. Finansijeru je potrebna ta informacija jer želi da zna koliko sredstava će biti potrebno da plaća savjetnike, a uz to zna nešto o statistici i intervalima povjerenja. Traži od vas da sa 95% vjerovatnoće svedete procjenu tako da ona dopušta maksimalno 5% greške sa obje strane onog što dobijete na uzorku. To znači da ako ste dobili na uzorku da je 20% studenata motivisano, da sa 95% vjerovatnoće tvrdite da pravi parametar leži između 15% i 25%. Finansijer očito ne zna pravo tumačenja intervala povjerenja (khm, primijetili ste grešku?), ali se pitate da li je uz to lud, jer kako da napravimo procjenu greške procjene kada ne znamo koliki je taj parametar? E pa, zahvaljujući tome što znamo šta sve ide u recept za intervale povjerenja možemo izračunati približno ono što je on zamislio. Dakle, kao što smo za statističku moć testiranja nulte hipoteze mogli da unaprijed izračunamo potrebnu veličinu uzorka, isto tako možemo da izračunamo koliko ispitanika nam je potrebno pa da suzimo interval povjerenja do željenog nivoa.

Kako bismo obavili dobar posao za situaciju sa binarnom kategoričkom varijablom (npr. ima / nema kliničku izraženu anksioznost, voljan / nije voljan doći u savjetovalište) bitan nam je samo jedan sastojak: broj ispitanika na uzorku (n). Veći uzorak = veća preciznost = uži interval povjerenja. Finansijer je tražio od nas da dopustimo po maksimalno 5% greške u procjeni sa obje strane statistika "sa 95% vjerovatnoće", iz čega možemo zaključiti da je mislio na 95%-tni interval povjerenja koji je ukupno širok 10% na mjernoj skali (5 sa dvije strane). On uz to vjerovatno ne zna da su procjene procenata asimetrične, osim za tačku 50% koja je tačka najveće varijabilnosti[6]. Budući da ne znamo stvarni parametar moramo izabrati 50% kao najgori scenario i tražiti od softvera da nam izračuna potreban broj ispitanika. Ipak ćemo ispitati i mogućnost da je to samo 10%, poređenja radi. U R-u (dugo ga nismo pominjali) je dovoljna da instalirate paket *presize* i po jedna linija koda je dovoljna da dobijemo tražene podatke:

[6] Najveća varijabilnost zato što je to najraznovrsnija populacija gdje je podjednako nekih jednih (motivisanih) i nekih drugih (nemotivisanih), ne postoji većina koja dominira.

```
# install.packages("presize")
presize::prec_prop(p =   0.50, conf.width =   0.10,
                   conf.level =   0.95, method =
                    ↪  "wilson")
##
```

```
##      sample size for a proportion with Wilson
↪  confidence interval.
##
##     p padj        n
## 1 0.5  0.5 380.3044
##   conf.width conf.level  lwr
## 1        0.1       0.95 0.45
##    upr
## 1 0.55
##
## NOTE: padj is the adjusted proportion, from which
↪  the ci is calculated.
presize::prec_prop(p =   0.10, conf.width =   0.10,
                   conf.level =   0.95, method =
                   ↪  "wilson")
##
##      sample size for a proportion with Wilson
↪  confidence interval.
##
##     p      padj        n
## 1 0.1 0.1106107 140.9728
##   conf.width conf.level
## 1        0.1       0.95
##          lwr       upr
## 1 0.06061072 0.1606107
##
## NOTE: padj is the adjusted proportion, from which
↪  the ci is calculated.
```

Dakle, u najgorem scenariju je potrebno da ispitamo ukupno 380 studenata, a u scenariju kada je prevalencija samo 10% bilo bi dovoljno samo 141 pa da dobijemo interval povjerenja kao što je $95\% CI [6.0\%, 16.0\%]$[7]. Naravno, pri uzorkovanju bismo se morali potruditi da imamo što reprezentativniji uzorak - po mogućnosti stratifikovani slučajni odabir kojim bismo pazili na barem one osnovne varijable koje bi mogle uticati na prevalenciju, kao što su: pol, godina studija, studijski program. Uz to, u istraživanjima redovno dobijamo i nevalidno popunjene upitnike, pa je moj savjet da uvijek imamo 10% više prikupljenih podataka od onoga što nam savjetuje formula o statističkoj moći, odnosno preciznosti procjene. U našem slučaju bi trebalo onda da uradimo istraživanje na 418 studenata ($380 + 0.1 * 380$).

[7] Ako se neko slučajno pita, Wilsonova formula koriguje proporciju prevalencije kako bi se spriječile pogrešne procjene usljed asimetričnih distribucija kada su parametri blizu vrijednosti 0% ili 100%.

Šta je to margina greške?

Kada je finansijer od nas tražio da svedemo grešku na maksimalno 5% sa obje strane procjene, on je mislio na nešto što se zove margina greške. **Margina greške** (engl. *Margin of Error*, *MoE*) je polovina širine intervala povjerenja. Formalno definisano, to je maksimalna očekivana razlika između procjene sa uzorka i vrijednosti parametra u populaciji, s određenim nivoom tačnosti procedure izračunavanja intervala povjerenja (obično 95%). Za različite procedure se izračunava različito, ali zapravo se to svodi na standardnu grešku koja se množi određenim skorom teoretske distribucije. Na primjer, već je to dobro poznato za z-distribuciju kada želimo 95%-tni interval povjerenja: $z_{95\%} = 1.96$).

Margina greške se smanjuje, odnosno intervali postaju uži sa sve većim uzorkom, ali se i tu srećemo sa problemom opadajuće dobiti. Intervali se mnogo više sužavaju na početku, ali kad želimo da preciznost stisnemo sa 2% greške na 1% greške moramo povećati uzorak sa skoro 2400 ispitanika na njih 9600! Ovdje treba napomenuti da je ova kalkulacija izvedena pod pretpostavkom da je naša populacija beskonačna. Međutim, na univerzitetu na kojem radim ima ukupno oko 10000 aktivnih studenata. U takvim slučajevima je korisno primijeniti korekcije za konačne populacije[8], pa bi nam skok išao sa 1937 studenata za 2% greške na "samo" 4900 za 1%! Ovi proračuni važe za onaj najgori scenario gdje se pretpostavlja parametar prevalencije traženog fenomena od 50%, a ako je to manje ili više onda može da se uštedi na hiljadi ili dvije ispitanika…Sljedeće R komande to potvrđuju:

```
# install.packages("samplingbook")
samplingbook::sample.size.prop(e = 0.01, # MoE
                               P = 0.50, N = Inf, level
                                ↪  = 0.95)
##
## sample.size.prop object: Sample size for proportion
 ↪  estimate
## Without finite population correction: N=Inf,
 ↪  precision e=0.01 and expected proportion P=0.5
##
## Sample size needed: 9604
samplingbook::sample.size.prop(e = 0.01, # MoE
                               P = 0.50, N = 10000,
                                ↪  level = 0.95)
```

[8] Za najradoznalije među vama, margina greške se koriguje tako što se dodatno množi sa *Finite Population correction* faktorom $FPC = \sqrt{\frac{N-n}{N-1}}$ koji sužava marginu greške srazmjerno tome koliko se veličina uzorka približava veličini populacije. Ako bi svi članovi populacije bili u uzorku, onda bi faktor korekcije iznosio 0 (jer $N - n = 0$), a što znači da margine greške uopšte nema, što je logično jer bi u tom slučaju parametar bio direktno izmjeren.

```
##
## sample.size.prop object: Sample size for proportion
↪  estimate
## With finite population correction: N=10000,
↪  precision e=0.01 and expected proportion P=0.5
##
## Sample size needed: 4899
```

Za procjenu prevalencije binarnih ishoda su nam dovoljni broj ispitanika i uzimanje u obzir pretpostavljene asimetrije. Kada su u pitanju dimenzione varijable moramo uzeti u obzir i varijabilnost. Ako želimo da procijenimo marginu grešku procjene aritmetičke sredine, onda je potrebno da softveru zadamo neku, što realističniju vrijednost. Na Slici 16.4 su navedeni podaci o potrebnom broju ispitanika u zavisnosti od varirane margine greške i standardne devijacije. Očekivano, veće standardne devijacije traže veći broj ispitanika.

Pretpostavljeni parametar	**A**	**B**	**C**	**D**
Standardna devijacija (SD)	15	15	10	10
Margina greške (MoE)	3	5	3	5
Ukupna širina intervala povjerenja	6	10	6	10
Potreban broj ispitanika (n)	98	37	45	18

Slika 16.4: Potrebna veličina uzorka za preciznost procjene aritmetičke sredine (95% povjerenja).

Kao što možete pretpostaviti, sve ovo je moguće napraviti i za procjenu drugih parametara (npr. koeficijenta korelacije, parametra razlike među aritmetičkim sredinama).

Koliku marginu greške bi trebalo da izaberemo?

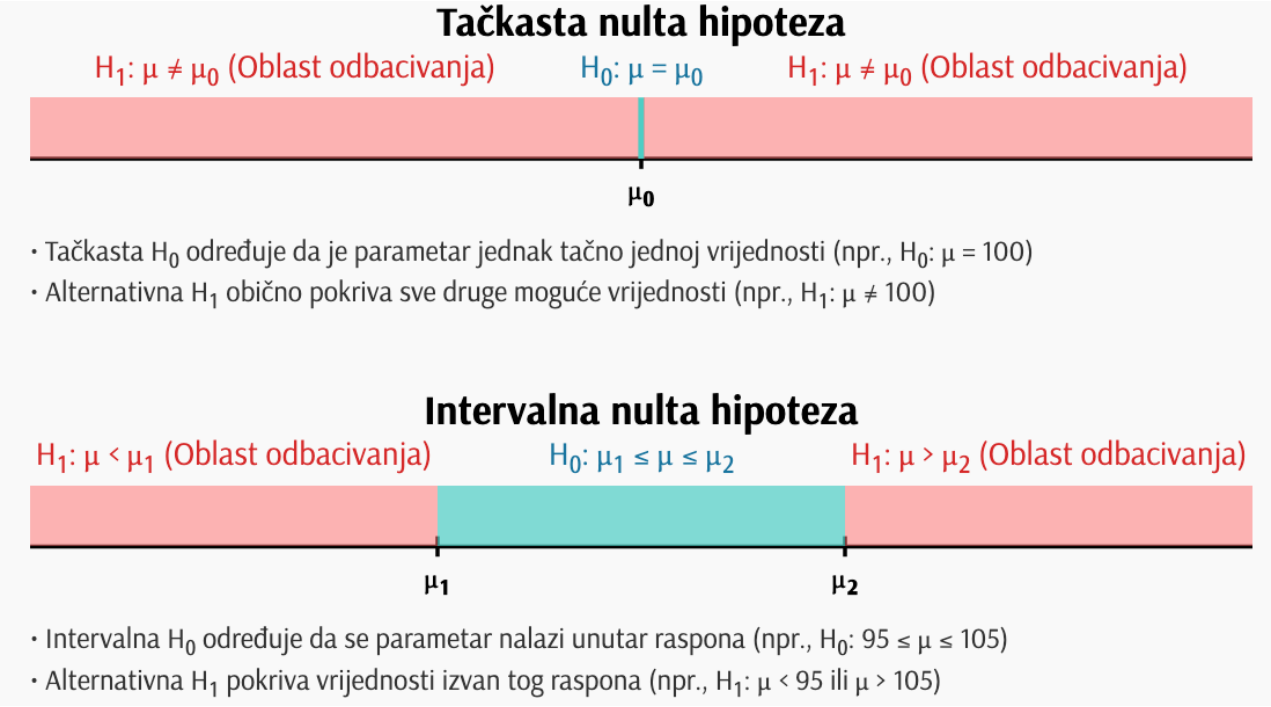
Na ovo pitanje se ne može dati jedan tačan odgovor. Odgovor ovisi o konkretnom istraživanju, o resursima koje imate, o zahtjevima finansijera. Nekad će biti dovoljno da imate šire intervale povjerenja, a nekada - pogotovo ako radite poslove za političke partije koje žele da procijenite kotiranje njihovog kandidata prije izbora - biće potrebno da svedete marginu na što manju mjeru (ako se upustite u takve rabote).

Šta je to intervalna nulta hipoteza?

Osim mogućnosti da unaprijed izračunamo preciznost procjene, intervali povjerenja nam dopuštaju da proširimo testiranje nulte hipoteze na viši nivo. Kao što znate, nultu hipotezu smo statistički operacionalizovali kao jednu specifičnu vrijednost. Onda smo u poglavlju o manama testiranja nulte hipoteze naglasili da je takva hipoteza banalna, odnosno da ima sljedeća ograničenja. Kao prvo, suštinski je nerealna budući da parametar efekta gotovo nikada nije tačno nula (da je npr. $H_0 : \delta = 0.0000$ ili $\rho = 0.000$), odnosno gotovo nikad nije jednaka nekom konkretnom parametru prevalencije (npr. $H_0 : \pi = 0.500000$). Drugo, odbacivanje H_0 je slabo informativno u pogledu praktične značajnosti nalaza. Čak i mizerni efekti će biti statistički značajni ako skupimo dovoljno velik uzorak ili ako se istraživači nakane i dovoljno potrude da statističkim igrarijama dođu do $p < \alpha$. Treće, testiranje tačkaste H_0 nam ne dopušta da prihvatimo nultu hipotezu, nego samo da je odbacimo ili ne odbacimo. Rekli smo da je nekad bitno da možemo da prihvatimo nultu hipotezu, na primjer kada poredimo dva psihoterapijska tretmana, pa želimo da zaključimo da nema razlike u njihovoj djelotvornosti.

Sva tri navedena ograničenja mogu da se prevaziđu ako umjesto da testiramo da li je parametar jednak *tačno* određenoj vrijednosti, testiramo intervalnu nultu hipotezu kojom provjeravamo da li parametar leži unutar intervala vrijednosti kojeg smo odredili na osnovu teorijskih i praktičnih kriterija. Dakle, istraživači mogu da unaprijed odrede raspon koji sadrži suštinski nebitne razlike u odnosu na nulti efekat. Definisaće ga nekom donjom i gornjom granicom i reći će: "ako je parametar ovoliko nizak (θ_D) ili ovoliko visok (θ_G), to je toliko praktično nebitno da možemo smatrati da efekat i ne postoji". Kako se određuju te granice? Isto onako kao što se određuje *minimalno važna razlika* (vidjeti prethodno poglavlje). Slika 16.5 ilustruje razliku između testiranja klasične, tačkaste nulte hipoteze i intervalne nulte hipoteze. Formalnom notacijom iskazano, intervalna nulta hipoteza se može definisati ovako:

$$H_0 : \theta_D \leq \theta \leq \theta_G$$

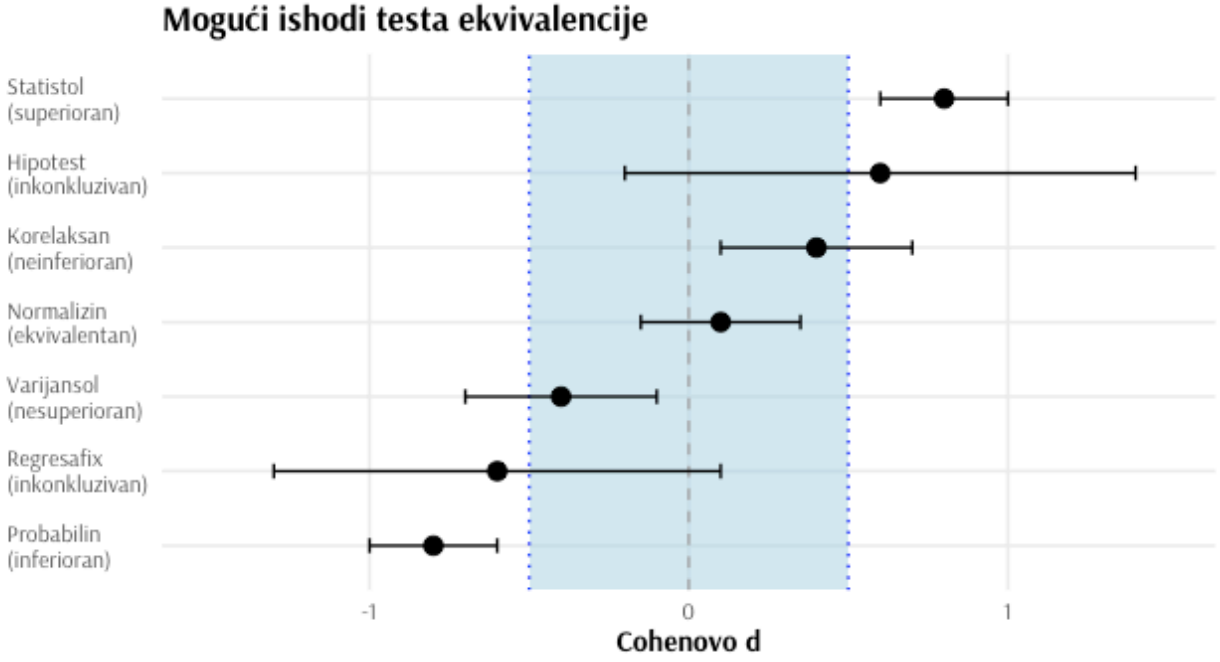


Slika 16.5: Poređenje tačkaste i intervalne nulte hipoteze.

Koje odnose i kako testiramo intervalnom nultom hipotezom?

Testiranjem intervalne nulte hipoteze možemo testirati više različitih odnosa, a to je najbolje ilustrovati primjerom. Recimo da je istraživanjem ispitivana efektivnost sedam novih biohemijskih sredstava za učenje statistike. Njihova efektivnost je operacionalizovana poređenjem sa kontrolnom grupom (k) koja uči bez farmakološke pomoći. Inače, umjesto kontrolne grupe u ovakvim nacrtima poredbenu grupu često čine neki zlatni standardi koji su ustaljeni, a naspram kojih se testiraju nove snage (npr. ekonomičnije terapije, lijekovi nove generacije, inovativni protokoli). Dakle, tačkasta nulta hipoteza bi bila $H_0 : \mu_x - \mu_k = 0$, odnosno $H_0 : \delta = 0$. Recimo da su se nakon velikog vijećanja i burne ekspertske debate istraživači odlučili za Cohenove granicu umjerenog efekta, odnosno da se zanemarivim proglase svi efekti razlike između učenja "na suvo" i potpomognutog učenja koji su manji od pola standardne devijacije. Tako definisana intervalna nulta hipoteza se matematičkom notacijom može napisati na sljedeći način: $H_0 : \delta \in I[-0.5, 0.5]$. Ovdje I služi kao oznaka za interval, dok se ovog čudnog E ($\in$) možda i sjećate iz rane matematike: ono znači "je element u skupu" ili "pripada". Analogno tome,

precrtano E ($\notin$) označava da nešto nije dio skupa, pa bi najopštija neusmjerena alternativna hipoteza glasila $H_1 : \delta \notin I[-0.5, 0.5]$. Kao što ćete vidjeti na Slici 16.6 postoji zapravo više hipoteza koje se mogu testirati.



Slika 16.6: Moguće interpretacije rezultata testiranja intervalne nulte hipoteze.

Slika 16.6 prikazuje vrlo komplikovanom terminologijom šest različitih odnosa za sedam farmakoloških sredstava. U zagradama su označeni odnosi inferiornosti ("lošije od nečeg") i superiornosti ("bolje od nečega"), kao i njihove inverzije neinferiornost i nesuperiornost, a tu je i ekvivalentnost (u smislu "podjednakosti"), kao i inkonkluzivnost ("nedovoljna količina informacija za donošenje zaključka")[9]. Vidimo i da je isprekidanom linijom označena tačkasta nulta hipoteza, dok je intervalna nulta hipoteza obilježena osjenčenim prostorom. Rezultati su prikazani u standardizovanim jedinicama, ali su mogli biti prikazani i na originalnoj skali (npr. razlika u broju bodova na testu), što se čak i preporučuje. Konačno, za sve razlike su prikazani 90%-tni intervali povjerenja koji su češći izbor za testiranje intervalne nulte hipoteze. Naime, oni održavaju nivo značajnosti od $\alpha = .05$, budući da smo obično zainteresovani da vidimo da li se dobijeni interval procjene parametra preklapa sa intervalnom nultom hipotezom na onoj strani koja je bliža nultoj tački (dakle, obično nas ne interesuje drugih 5%).

[9] Susramljen sam zbog ovako komplikovane terminologije.

Prođimo sada detaljnije kroz rezultate naših sedam statističkih farmakoloških supstanci. *Statistol* pokazuje jasan **superioran**

efekat, sa intervalom povjerenja koji je u potpunosti iznad gornje granice ekvivalencije ($+0.5d$), što znači da je bolji od kontrolne grupe. *Korelaksan* je **neinferioran** u odnosu na uobičajeno učenje - njegov donji interval povjerenja ne pada ispod vrijednosti $-0.5d$, ali ne možemo tvrditi da je superioran jer njegov interval dijelom ulazi u zonu ekvivalencije. *Hipotest* pokazuje **inkonkluzivan rezultat** - njegov široki interval povjerenja obuhvata i zonu ekvivalencije i zone superiornosti, što znači da nemamo dovoljno preciznu procjenu i možemo samo da nagađamo stvarnu korist od njega. *Normalizin* je **ekvivalentan** kontrolnoj grupi - njegov interval povjerenja je u potpunosti unutar zone ekvivalencije ($-0.5d$ do $+0.5d$). *Regresafix* daje takođe inkonkluzivan rezultat, ali sa opaženom tendencijom ka inferiornosti. *Varijansol* je **nesuperioran** - njegov gornji interval povjerenja ulazi u zonu ekvivalencije, ali ne prelazi tačku $+0.5d$. Konačno, *Probabilin* je jasno **inferioran**, sa intervalom povjerenja koji je u potpunosti ispod donje granice ekvivalencije ($-0.5d$), što ukazuje da je značajno lošiji od kontrolne grupe.

Kako mi možemo da uradimo testiranje intervalne nulte hipoteze?

R, *Jamovi* i *JASP* omogućavaju jednostavno testiranje razlika između dvije grupe gdje uz intervale povjerenja dobijate i p za klasični t-test. Obično ćete dobiti i posebne p-vrijednosti koje testiraju razliku o odnosu na postavljene granice, ali njih možete slobodno zanemariti ako ste shvatili logiku gore napisanog. Osim t-testova (za jedan uzorak, sparena mjerenja i nezavisne uzorke) lako ćete pronaći i testove ekvivalentnosti korelacija i proporcija. Takođe, postoje i specijalizovani R paketi (npr. *TOSTER*) putem kojih možete jednostavno da izračunate potrebnu statističku moć za buduća istraživanja.

Kako saopštavati rezultate testiranja intervalne nulte hipoteze?

S obzirom na to da se radi o relativno novijoj proceduri koja se ne sreće tako često u naučnim radovima, evo kako bi se u skraćenoj verziji mogli navesti rezultati za dva ispitivana sredstva:

"Istraživanjem smo provjeravali djelotvornost *Normalizina* i *Statistola* u odnosu na standardno učenje statistike. Testirali smo intervalnu nultu hipotezu sa granicama praktične ekvivalentnosti postavljenim na $\pm 0.5d$, što se tradicionalno smatra umjerenom veličinom efekta. Rezultati pokazuju da je *Normalizin* ekvivalentan standardnoj metodi ($d = 0.10, 90\% CI[-0.15, 0.35]$),

budući da se cijeli interval povjerenja nalazi unutar postavljenih granica praktične jednakosti. S druge strane, *Statistol* se pokazao kao superioran u odnosu na standardnu metodu ($d = 0.80, 90\%CI[0.60, 1.00]$)."

Uz to, kad god je moguće, bilo bi dobro prikazati i grafikon budući da u ovakvim testiranjima postoji više granica koje se mogu preći, te se čitaoci mogu izgubiti u narativu.

Koja su ograničenja intervalnog testiranja nultih hipoteza?

Mada su prednosti testiranja intervalnih hipoteza jasno uočljive, ni ova procedura nije savršena. Ključni izazov kod intervalnog testiranja je definisanje odgovarajućih granica ekvivalencije, koje bi morale biti teorijski opravdane i postavljene prije analize podataka (npr. u fazi preregistracije istraživanja). Preširoke granice mogu dovesti do zaključka o ekvivalentnosti tamo gdje postoje praktično značajne razlike, dok preuske granice mogu spriječiti dokazivanje ekvivalencije i kada su tretmani suštinski jednaki. Na primjer, da su u gore opisanom istraživanju postavljene granice na $\pm 0.3d$ umjesto $\pm 0.5d$ *Normalizin* ne bi bio klasifikovan kao ekvivalentan kontrolnoj grupi. Međutim, definisanje granica nakon uvida u rezultate predstavljalo kvarnu naučnu praksu, jer se tako vještački rastežu kriterijumi kako bi se dobio željeni zaključak.

Drugo bitno ograničenje je ograničenje praktične prirode. Za vrlo precizne odgovore moramo imati izuzetno velike uzorke. Tu su i zablude o tumačenjima intervala povjerenja. Naime, i dalje se radi o mehaničkoj proceduri koja u određenom postotku slučajeva obuhvata parametar, što nam ne omogućava da tvrdimo da postoji neka razlika ili ekvivalentnost sa određenim stepenom vjerovatnoće. Problemu stepena vjerovatnoće ćemo se uskoro, ali sljedeće poglavlje posvećujemo logičnom iskorištavanju prednosti koje omogućavaju intervali povjerenja. Radi se o krunskom statističkom pristupu koji se naziva meta-analiza.

Poglavlje 17

Meta-analize

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta su meta-analitičke studije i zašto se smatraju najsnažnijim oblikom kvantitativnih istraživanja?
- Koja je razlika između modela fiksnog efekta i modela slučajnih efekata, te kako se izračunavaju i tumače rezultati meta-analize?
- Šta je heterogenost u kontekstu meta-analiza i kako se ona mjeri?
- Koji su osnovni problemi meta-analiza i kako se procjenjuje moguća pristrasnost procjena?

Prešli smo dobar komad puta kroz statistiku zaključivanja. Naučili smo koristiti različite alatke za stvaranje uvjerljivih modela stvarnosti. Sada smo došli do tačke kada je moguće udružiti njihove funkcije. Kako se to radi ćemo uvidjeti na starom primjeru sugestopedije. Ako se sjećate poglavlja o replikacijama, tamo sam opisao situaciju u kojoj su izvedene dvije studije koje su dale kontradiktorne rezultate iako su korišteni isti materijali (tj. muzička kompozicija koju je slušala eksperimentalna grupa, tekst iz kojeg obje grupe uče statistiku, test provjere znanja). U prvom istraživanju je dobijena vrijednost $p = .02$, a u drugom $p = .22$. Ove p-vrijednosti su oprečne ako se želi donijeti odluka o nultoj hipotezi, a takvi su i odgovarajući intervali povjerenja budući da u prvoj studiji 95%-tni interval ne obuhvata vrijednost 0 ($95\%CI[0.59, 6.63]$), dok se u drugoj studiji 0 nalazi unutar izračunatog intervala ($95\%CI[-1.38, 5.84]$). Dosad naučeno o replikacijama i statističkoj moći nam govori da bi izlaz iz ove kontradiktorne situacije bio da izvedemo dodatnu studiju sa znatno većim uzorkom koji bi otklonio sumnje o tome da li pozitivan efekat muzike uopšte postoji. Ali šta ako nam zapravo postojeće studije već nude sve sastojke za novu? Drugim

riječima, možemo li iskoristiti već postojeće rezultate, pa da podatke obradimo na drugačiji način i kao rezultat dobijemo kombinovanu procjenu efekta? Upravo to nam omogućavaju meta-analitičke studije.

Šta su to meta-analize?

Definicija meta-analize

Meta-analitička studija (ili kraće **meta-analiza**, engl. *meta-analysis*) je studija koja statistički kombinuje rezultate više empirijskih studija koje testiraju istu hipotezu, te ih prezentuje kao pojedinačnu procjenu parametra uz odgovarajuće intervale povjerenja. Kao takve, meta-analize su po svojoj suštini replikacione studije visoke statističke moći koje proizvode sužene intervale povjerenja. Ako se pitate kako je moguće direktno statistički uporediti studije koje imaju različite materijale, procedure i/ili mjerne skale, odgovor je: tako što ćemo koristiti još nešto što smo ranije upoznali, standardizovane mjere veličine efekta.

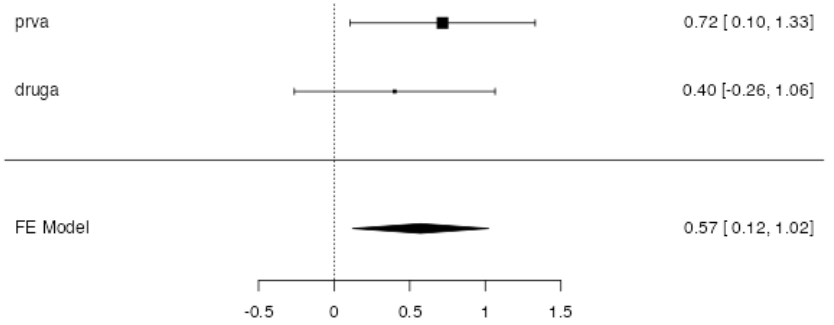
Kako se prikazuju rezultati meta-analize?

Mada se meta-analize najčešće provode na većem broju studija, za početnu demonstraciju ćemo uzeti ekstremno jednostavan primjer mini meta-analize koja opisuje sad već proganjajući problem sa baroknom muzikom. Na Slici 17.1 vidimo da su prikazani intervali povjerenja za obje studije. Pritom vidimo da su sirove vrijednosti konvertovane u Cohenovo d tako da je, npr. za Studiju 1 $95\%CI[0.59, 6.63]$ to postalo $95\%CI[0.10, 1.33]$. Slikovni prikaz nam jasno pokazuje da za razliku od prve studije, u drugoj studiji interval povjerenja obuhvata vrijednost 0, odnosno ostavlja razumnu mogućnost da je nulta hipoteza istinita. Na datom grafikonu - koji se inače naziva **grafikon šume**[1] (engl. *forest plot*) - vidimo i jedan novi oblik na dnu. Romb ili dijamant-marker predstavlja na svom najvišem dijelu kombinovani finalni statistik zajedno sa intervalom povjerenja koji se širi na dvije strane. Oblik romba se koristi za finalni rezultat, jer ga možemo razlikovati od rezultata pojedinačnih studija, ali i zato što taj razliveniji oblik slabije naglašava pojedinačne vrijednosti unutar intervala. Drugim riječima, takav oblik nam sugeriše da ne usmjeravamo previše pažnje na neku centralnu tačku nego na ukupni obuhvat, odnosno krajeve.[2]

Ali šta nam taj oblik i numerički rezultati prikazani desno od njega govore u našem slučaju o efektu barokne muzike?

[1] Grafikon šume se naziva tako jer umjesto jednog "drveta" (pojedinačne studije) slikovito prikazuje čitavu "šumu" istraživanja. Samo u ovom slučaju morate da umjesto dva drvceta zamislite da ih je stotinjak.

[2] Primjećujete i da je jedan kvadrat koji označava dobijeni statistik značajno veći od drugog. Zašto je to tako saznaćete kasnije u poglavlju.



Slika 17.1: Grafikon šume koji prikazuje veličinu efekata pojedinačnih studija i kombinovanu veličinu efekta.

[Ponoviću još jednom da se radi o fiktivnim primjerima, pa baroknu muziku tokom učenja slušajte na svoju odgovornost.] Za početak, možemo da vidimo da se donja granica intervala povjerenja nalazi desno od nulte tačke. Nadalje, gornja granica intervala povjerenja nam sugeriše da bi parametar mogao iznositi i čak jednu standardnu devijaciju. Uočavamo i da je konačni interval povjerenja nešto uži od oba intervala za pojedinačne studije. Kako sam već naglasio, to i jeste funkcija meta-analize: da dobijemo precizniju procjenu traženog parametra. Ipak, postojaće rijetki slučajevi kada kombinovani efekat ima šire intervale povjerenja nego individualne studije, a razlozi za to će biti pojašnjeni kasnije. Sve u svemu, meta-analitički pristup nam je na elegantan način pomogao da prevaziđemo kontradiktornost rezultata pojedinačnih studija.

Zašto se meta-analize smatraju najbitnijim kvantitativnim studijama?

Valuta kojom se najčešće mjeri bitnost rada u nauci je njegova citiranost, odnosno broj puta koliko se naučnici pozivaju na taj rad. Podatak o citiranosti pokazuje koliko određeni rad živi u mentalnim modelima drugih naučnika. U prosjeku gledano, citiranost meta-analiza je daleko veća nego citiranost pojedinačnih studija, te se one zato mogu smatrati najvrednijim oblikom naučnih publikacija. Zahvaljujući tome što sintetišu nalaze ranijih radova, njihovi rezultati odjekuju kroz nauku i tako što služe kao referentni okvir za buduće studije. Ako dobijete neki naučni ili profesionalni zadatak gdje vam može pomoći neko naučno saznanje, po pravilu bi trebalo da u pretragu unesete i termin "meta-analiza". Visoki status meta-analiza je potpuno zaslužen jer su utemeljene na načelima kvalitetne i kumulativne nauke:

proističu iz višestrukih replikacija i imaju visoku statističku moć koja omogućava precizniju procjenu parametra.

Evo dva primjera. Gigantska i uticajna meta-analiza je analiza rodnih razlika u seksualnosti[3]. Broj ukupno uključenih studija čiji su rezultati kombinovani je bio 730, a sveukupan uzorak ispitanika je dostigao broj od čak 1 419 807 (da, skoro milion i po). Kada dobijete nalaz na takvom uzorku nema baš puno mjesta za sumnje. Druga, vjerovatno za vas praktičnija meta-analiza (mada možete iskoristite neka saznanja i iz prethodno pomenute?) se bavila procjenom važnosti psiholoških korelata fakultetskih ocjena.[4]. Autori su iz 217 objavljenih radova izvukli korelacije između ocjena i pedesetak razmatranih konstrukata gdje su uzorci za svaku korelaciju bili veliki (između 933 i 75000 ispitanika, ovisno o konstruktu). Naravno da je preporučivo da ako ste student pročitate ovu studiju.[5]

[3] Petersen & Hyde, 2010, https://doi.org/10.1037/a0017504

[4] Richardson et al., 2012, https://doi.org/10.1037/a0026838

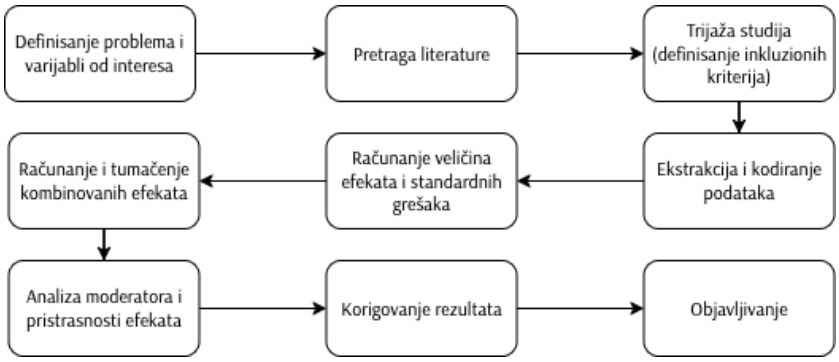
[5] Nekima će biti interesantno to da su obje ove studije objavljene u izuzetno uticajnom časopisu *Psychological Bulletin*, koji prvenstveno objavljuje meta-analize.

Meta-analize imaju poseban značaj za oblasti u kojima - iz logističkih razloga - studije sa malim uzorcima predstavljaju pravilo. To obično važi za kliničke studije u kojima se provjerava djelotvornost psihoterapija zato što je teško regrutovati veliki broj ispitanika i u tretmanskoj i u kontrolnoj grupi. Kao primjer može poslužiti meta-analitička studija kojom su direktno poređeni efekti psihoterapije i farmakoterapije u tretiranju različitih oblika depresivnosti i anksioznosti.[6] U nju su uključeni podaci iz 67 studija gdje su ukupno 3142 pacijenta prošli psihoterapiju, a 2851 farmakoterapijski tretman. Ako se zna da je srednja vrijednost veličine uzoraka u tim studijama bila $n = 30$ onda se vidi koliko je ovo drastičan skok u odnosu na uobičajenu sliku koja se stvara na osnovu pojedinačnih studija. Pojava meta-analiza time omogućava i već izvedenim, statistički "nemoćnim" studijama da ne završe u korpama za smeće, nego da doprinesu izgradnji kumulativnog naučnog znanja.

[6] Cuijpers et al., 2013, https://doi.org/10.1002/wps.20038

Kako se izvode meta-analize?

Iako su konačni rezultati meta-analitičkih studija statistički podaci koji se dobijaju posebnim statističkim tehnikama, njihovo izvođenje podrazumijeva više koraka različitog karaktera koje biste trebalo da znate. Uobičajeni protokol je šematski prikazan na Slici 17.2.



Slika 17.2: Uobičajeni protokol izrade meta-analiza.

Koji su to nestatistički koraci koje je potrebno izvesti?

Kao i u svakoj drugoj vrsti istraživanja, na početku se moraju precizno definisati istraživački problem i varijable od interesa. To je najčešće narativno kao na primjeru rada koji je procjenjivao efekte psihoterapije depresije među djecom i adolescentima.[7] Njihova dugačka definicija kaže: "*psihoterapija depresije* se definiše kao intervencija osmišljena da olakša depresivni poremećaj ili povišene nivoe depresivne simptomatologije putem strukturisanih ili nestrukturisanih interakcija ili trening programa koje provode jedan ili više pojedinaca obučenih za intervenciju ili kroz program koji je osmišljen tako da se samostalno provodi." Iz definicije vidimo da su autori uzimali u obzir kako tretmane koje je vodio terapeut prateći manje ili više strukturisan protokol, tako i tretmane koji su kreirani kao oblik samopomoći. Nekad se ipak konstrukti od interesa moraju pobrojati u formi tabele kao što je to slučaj u gore pomenutoj meta-analizi Richardsona i saradnika koja sadrži pedesetak psiholoških korelata uspješnosti u studiranju.

[7] Weisz et al., 2006, https://doi.org/10.1037/0033-2909.132.1.132

Nakon definisanja problema i varijabli se prelazi na pretragu onlajn bibliografskih baza kao što su PsycINFO, Google Scholar, MEDLINE, ERIC, Cochrane. Na primjer, Weisz i saradnici u istom članku navode: imena pretraživanih bibliografskih baza, podatak da su pretraživali sve radove koji su objavljeni do 2004. godine, opasku da su čak dodatno pješke pretresali časopise od 1965. do 2004. godine u kojima je objavljeno barem pet psihoterapijskih studija, te konačno da su lično komunicirali sa autorima relevantnih studija pitajući ih da li znaju za dodatna relevantna istraživanja. Kada se opisuje pretraga onlajn bibliografskih baza moraju se navesti korištene ključne riječi, pa su ovi autori naveli kako su koristili tri dijagnostička pojma ("depresija", "distimija", "velika depresija" - naravno, na engleskom jeziku), te da su još

ograničili studije na one koji su u uzorku imali djecu ili adolescente. U poznatoj meta-analizi o povezanosti crta ličnosti i uspjeha na fakultetu, autorka studije[8], navodi da je pretraga baza uključivala sljedeće termine i operatore:"(academic OR education OR university OR school) AND (grade OR GPA OR performance OR achievement) AND (personality OR temperament)". Dakle, u ovom koraku je vrlo bitno odrediti koje sve ključne riječi su uzimane u obzir što može da ukaže i na temeljnost pretrage, ali i na moguće propuste.

[8] Vedel, 2014, https://doi.org/10.1016/j.paid.2014.07.011

Na osnovu početne bibliografske pretrage se dobije višestruko veći broj radova nego što je broj onih koji će biti uzeti u obzir za statističku obradu. Tako je u studiji Vedelove od njih početnih 826 u obradu uključeno svega 20, dok je u studiji Cuijpersa i saradnika konačan broj iznosio 67, iako su na početku imali čak 21 729 reference koje su odgovarale kriterijima pretrage. Naime, u statističku obradu mogu da uđu samo one studije koje zadovoljavaju ranije definisane kriterije uključivanja (tzv. inkluzioni i ekskluzioni kriteriji). Na primjer Cuijpers i saradnici su nakon isključivanja duplih rezultata iz različitih baza, isključili i radove u kojima nisu eksplicitno navođene dijagnoze tretiranog poremećaja ili u kojima nije bilo kontrolne grupe. Često se postavljaju rigorozni kriteriji, pa su tako Weisz i saradnici dopustili da u finalni uzorak uđu studije koje su - uz to što jasno opisuju prisustvo depresivnosti među ispitanicima - morale biti eksperimentalne studije sa randomizovanom alokacijom ispitanika u barem jednu tretmansku i jednu kontrolnu grupu (klasična kontrolna, na listi čekanja, minimalno tretirani pacijenti ili aktivni placebo), kao i da je prosječna dob ispitanika morala biti manja od 19 s obzirom na to da je istraživana populacija djece i adolescenata. Ovaj proces trijaže radova je najmukotrpniji dio meta-analitičke avanture, jer se svi radovi moraju detaljno pregledati, ali je presudan za njen konačni kvalitet.

Koje podatke je potrebno ekstrahovati iz originalnih studija?

Tek nakon što je evidentirano koje studije su zadovoljavajućeg kvaliteta, iz njih se vade podaci relevantni za statističku obradu. U te podatke uvijek spadaju broj ispitanika (ukupno n, ali i po grupama ako se radi o poredbenim studijama), te osnovna deskriptivna statistika (npr. M i SD za svaku grupu u poredbenim studijama, r za studije povezanosti ili frekvencije unutar tabela kontingencije ako su sve varijable bile kategoričke). Kao varijabla identifikator se obično navodi ime prvog autora i godina

objavljivanja studije. Uz to, registruju se i vrijednosti svih značajnijih varijabli koje bi mogle da modifikuju / mijenjaju jačinu povezanosti između ispitivanih varijabli (tzv. moderatori). Na primjer, više od 10 godina nakon studije Weisz i saradnika, drugi tim autora[9] je htio još detaljnije ispitati efekte psihoterapije na depresivnost djece i adolescenata. Iz tog razloga su uključili u studiju uključili i veliki broj moderatorskih varijabli za koje su uzimali podatke i upisivali ih u tabelarni fajl. Neke od njih su: informant koji je saopštavao o simptomatologiji (sam pacijent, roditelj ili terapeut), oblik tretmana (individualna, grupna ili mješovita terapija), vrsta psihoterapije i protokola, vrsta kontrolne grupe, godina studije, etnička pripadnost, polni sastav, mjerni instrumenti. Nakon što se kompletira baza podataka - bez ili sa moderatorskim varijablama - kreće se u statističku obradu.

[9] Eckshtain et al., 2020, https://doi.org/10.1016/j.jaac.2019.04.002

Kako se provodi statistička analiza meta-analitičkih studija?

Prvi korak u statističkoj obradi jeste obezbijediti da rezultati svih izabranih studija imaju jednaku metriku, odnosno da su izraženi na istoj mjernoj skali. Ovo je neophodan korak ako su podaci dobijani putem različitih mjernih instrumenata - što je gotovo uvijek slučaj. Kao što znate to se radi tako što se podaci svode na standardizovane veličine efekata. Skup svih standardizovanih veličina efekata se u statistici označava sa Y, te se matematičkom notacijom to navodi ovako: $Y = \{d, g, \Delta, r, OR, RR, RD, ...\}$. Današnji softveri koji su sposobni da izvedu meta-analitičke proračune (osim *R* paketa koji se podrazumijevaju, to rade i *JASP* i *Jamovi*) automatski transformišu unesene deskriptivne podatke u standardizovane veličine efekta.

Slika 17.3 prikazuje tabelu koja prikazuje tipični izgled tako obračunatih veličina efekta. U koloni "*yi*" su date vrijednosti standardizovanih razlika aritmetičkih sredina.[10] Ovdje je korisno da se napomene da u slučaju da su rezultati u originalnoj studiji dati u jednoj metrici (npr. r), njih je moguće konvertovati u drugu veličinu efekta (npr. d) - za šta postoje formule, a i *R* paketi.

[10] Slovo i predstavlja indeks kojim se označava broj studije - Y_i. Na primjer, Y_1 je veličina efekta prve studije, Y_2 druge studije, itd.

rb	autori	n1	m1	sd1	n2	m2	sd2	yi	vi
1	Prvić (2020)	22	51.8	4.93	22	48.2	4.93	0.72	0.10
2	Drugoje (2022)	18	51.1	5.37	18	48.9	5.37	0.40	0.11
3	Trećko (2024)	100	51.5	10	100	48.5	10	0.30	0.02

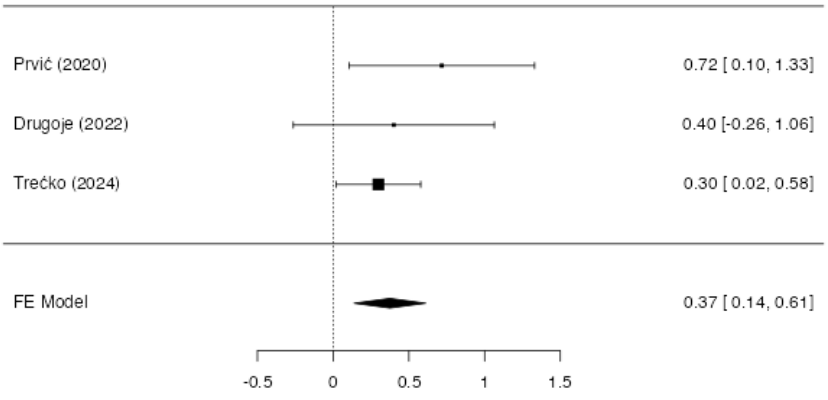
Slika 17.3: Tipična baza osnovnih podataka za meta-analizu.

Konačno, na Slici 17.3 primjećujemo i kolonu ("*vi*"), a evo i šta se u njoj nalazi. Naime, već smo naučili da se rezultati pojedinačnih studija izražavaju uz intervale povjerenja, a to podrazumijeva da moramo imati standardnu greška datog statistika. Slovo V označava kvadriranu standardnu grešku, odnosno varijansu dobijenog statistika (označava se još i sa Var, ali i sa σ^2 za populaciju, odnosno SD^2 kada je u pitanju uzorak s obzirom na to da je varijansa kvadrirana standardna devijacija). Kasnije ćemo saznati zašto se u kontekstu meta-analiza preferira kvadrirana vrijednost standardne greške umjesto proste standardne greške, ali zasad je bitno da imate na umu da je suština ista - što je lako vidljivo iz formule za varijansu mjere prosjeka[11]:

[11] Za druge mjere, kao što su koeficijenti korelacija, proporcije, omjeri rizika, postoje druge formule za izračunavanje standardnih grešaka.

$$V_i = \left(\frac{SD_i}{\sqrt{n_i}}\right)^2 = \frac{SD_i^2}{n_i}$$

Ova formula ponovo pokazuje ono što biste morali već znati: greška procjene se povećava sa porastom varijabilnosti, odnosno smanjuje se sa sve većim uzorcima. Pogledajmo primjer na Slici 17.4 Uz dva ranije predstavljena istraživanja dodao sam i još jedno, koje ima znatno više ispitanika (po 100 ispitanika po grupi) nego prva dva. Budući da rezultati na većim uzorcima - i sa manje varijabiliteta - daju pouzdanije procjene parametra, postavlja se pitanje da li je ta studija vrednija od druge dvije. Odnosno, zar ne bi trebalo da pri izračunavanju kombinovanog efekta uzmemo u obzir i statističku moć različitih studija?



Slika 17.4: Grafikon šume sa dodatnom studijom koja ima značajno veći uzorak.

Kako se statistička moć studija uzima u obzir pri izračunavanju kombinovanog efekta?

Naravno da su statističari svjesni toga da bi studije sa manjom očekivanom greškom trebalo da imaju veći uticaj na konačni

rezultat. Istraživanje provedeno na samo 10 ispitanika ne može da ima istu informativnu vrijednost kao i istraživanje na 1000 ispitanika. Iz tog razloga, prije izračunavanja kombinovanog efekta, svakoj studiji se dodjeljuje poseban **množilac** (**koeficijent**, **ponder**, engl. *coefficient* ili *weight*) koji uvrštava informativnost svake studije u konačnu procjenu. Vrijednost tog pondera bi morala biti direktno proporcionalna veličini uzorka i obrnuto proporcionalna varijabilitetu unutar studije. Ako razmislite o tome, shvatićete da je to bukvalno inverzno od onoga što predstavlja već prikazana vrijednost kvadrirane standardne greške, pa je formula za taj množilac sljedeća[12]:

$$W_i = \frac{1}{V_{Y_i}} = \frac{n_i}{SD_i^2}$$

[12] Desni dio formule važi u slučaju da se radi o razlikama između aritmetičkih sredina.

Oznaka W dolazi od engleskog *weight* što znači **opterećenje** (da, to bi bio još jedan sinonim za množilac). Malo slovo i u indeksu ponovo označava da se zaseban ponder dobija za svaku pojedinačnu studiju. I još jednom: studije sa većim brojem ispitanika i manjim varijabilitetom dobijaju veće pondere, odnosno imaju veći uticaj na konačnu kombinovanu vrijednost. Eto konačno i odgovora na pitanje zašto su neki kvadrati na grafikonu šume veći od drugih. Njihova veličina je proporcionalna njihovom ponderu unutar tog skupa pojedinačnih studija. Uporedite ponovo Sliku 17.1 i Sliku 17.4. U prvom slučaju studija sa 44 ispitanika je imala mnogo veći značaj nego u drugom slučaju kada je naišla studija sa 200 ispitanika.

Kako se onda izračunava kombinovani efekat?

Kako bih lakše prikazao aritmetiku koja leži iza izračunavanja kombinovanog efekta prikazaću ekstremno pojednostavljen slučaj.[13] Recimo da su u dva istraživanja dobijene razlike u pogledu zadovoljstva vlastitim mentalnim zdravljem (dimenziona varijabla) između studenata koji žive sa roditeljima i onih koji ne žive sa roditeljima (dihotomna varijabla). U prvoj studiji, provedenoj na Zapadnom Balkanu, na samo 16 studenata, neki Lakić je dobio rezultate u kojima studenti koji žive sa roditeljima prikazuju za jednu standardnu devijaciju lošije zdravlje nego oni koji žive odvojeno od njih ($d = -1.0, SD = 2$). U drugoj studiji, u Japanu na 160 ispitanika, Yasui dobija suprotne rezultate: studenti koji još uvijek žive zajedno sa roditeljima tvrde da imaju za jednu standardnu devijaciju bolje mentalno zdravlje $d = 1.0, SD = 2$. Da bismo mogli izračunati kombinovani efekat, prvo ćemo izračunati pondere za pojedinačne studije. Za

[13] Ponovo se radi o fiktivnom istraživanju; zaključujte o prikazanim rezultatima na vlastitu odgovornost.

balkansku studiju to bi bilo:

$$W_{Lakic} = \frac{n_{Lakic}}{SD^2_{Lakic}} = \frac{16}{4} = 4$$

…a za japansku:

$$W_{Yasui} = \frac{n_{Yasui}}{SD^2_{Yasui}} = \frac{160}{4} = 40$$

Konačni korak predstavlja kombinovanje efekata pojedinačnih studija (Y_i) u jedan statistik. Da su obje studije imale istu preciznost (isti broj ispitanika i istu varijabilnost), onda bismo ih jednostavno kombinovali tako što bismo prosto izračunali prosječnu vrijednost. Međutim, budući da je preciznost 10 puta veća u slučaju japanske studije - jer je u toj studiji 10 puta više ispitanika - računa se **ponderisana ili opterećena aritmetička sredina** putem sljedeće formule:

$$M_Y = \frac{\sum(W_i * Y_i)}{\sum W_i}$$

U našem konkretnom slučaju bi značilo da kombinovani efekat (M_Y) iznosi:

$$M_Y = \frac{(4 * -1d) + (40 * 1d)}{4 + 40} = \frac{-4d + 40d}{44} = 0.82d$$

Vidimo da je kombinovani efekat pozitivan i vrlo blizu onoga što je dobijeno u japanskoj studiji. Deset puta manja studija je tek blago uticala na kombinovani efekat povlačeći ga u svoju stranu.

Kako se izračunavaju intervali povjerenja za kombinovanu veličinu efekta?

Da bismo izračunali intervale povjerenje potrebna nam je vrijednost standardne greške. Znamo da je $SE_{M_Y} = \sqrt{V_M}$, što zapravo nije ništa drugo nego $SD = \sqrt{SD^2}$. Za pojedinačni interval povjerenja je lako: njegov obuhvat se gradi oko izračunatog statistika tako što se standardna greška pomnoži sa kritičnim skorom za određeni stepen povjerenja (npr. $z = 1.96$ za 95%-tni interval povjerenja ako je normalna distribucija odgovarajuća). Međutim, kako da standardnu grešku izračunamo za kombinovanu veličinu efekta sastavljenu od više studija? Kakva bi uopšte trebalo da bude standardna greška ako imamo više studija - veća ili manja nego što je za pojedinačne studije?

Naravno, manja. Što imamo više studija imamo i više podataka. Drugim riječima, sa više podataka imamo i više preciznosti, te posljedično manji prostor za grešku. Upravo nam to i govori formula za kombinovani efekat koja kaže da je kvadrirana standardna greška obrnuto proporcionalna sumi preciznosti pojedinačnih studija:

$$V_M = \frac{1}{\sum W_i}$$

Iz formule možemo da zaključimo da svaka dodatna studija doprinosi smanjenju greške, odnosno da se interval povjerenja za kombinovani efekat smanjuje sa svakom dodatnom studijom i to u zavisnosti od količine njene preciznosti. U našem banalno jednostavnom primjeru ta varijansa se izračunava na sljedeći način:

$$V_M = \frac{1}{W_1 + W_2} = \frac{1}{4 + 40} = \frac{1}{44} = 0.022$$

Da bismo došli do standardne greške, ostalo je još samo da izvadimo kvadratni korijen:

$$SE_M = \sqrt{V_M} = \sqrt{0.022} = 0.15$$

Konačno možemo da izračunamo i intervale povjerenja koristeći poznatu formulu:

$$M_Y \pm z_{95\%} \cdot SE_M = 0.82d \pm 1.96 \cdot 0.15 = 0.82d \pm 0.29 = [0.53d, 1.11d]$$

S obzirom da konačni interval ne obuhvata nultu vrijednost kao mogući parametar (tj. $\delta = 0$), možemo da odbacimo jedinstvenu nultu hipotezu ako neko za to mari. Čak možemo i izračunati p-vrijednost putem formule za z-skor:

$$z_{M_Y} = \frac{M_Y}{SE_{M_Y}} = \frac{0.82}{0.15} = 5.47$$

S obzirom na tako veliku z vrijednost koja je mnogo veća od graničnog skora 1.96, slijedi da je $p(z = 5.47 \mid \delta = 0) < .001$.

"Heh, pa ova meta-analiza je bila baš jednostavna" - možda ste sebi rekli. Ako ste to pomislili, prevarili ste se. Nije baš sve tako jednostavno ni u stvarnosti, pa smo zapravo tek na pola puta. Kada je u pitanju statistička obrada meta-analiza sve što smo dosad napravili je vezano za tzv. model fiksnog efekta.

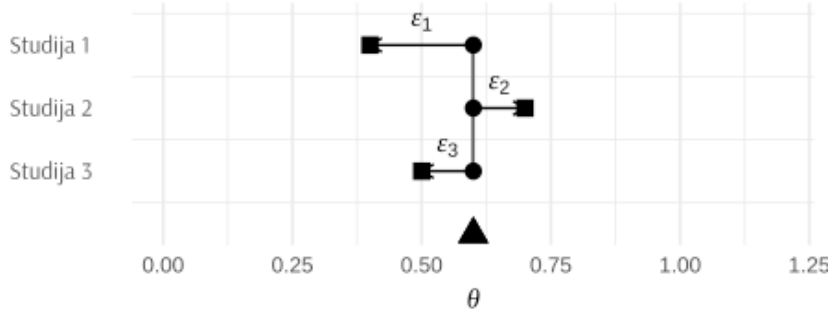
Šta je model fiksnog efekta?

Model **fiksnog, fiksiranog ili jedinstveno određenog efekta** (engl. *fixed-effect model*) je statistički model kojim se pretpostavlja da se u pozadini svih pojedinačnih studija obuhvaćenih meta-analizom nalazi samo jedan skriveni parametar (θ). To je pretjerano naivan model, a evo i zašto. Takvim modelom se pretpostavlja da jedan te isti parametar važi za: različite instrumente koji su korišteni da mjere zavisnu varijablu, sve države u kojima je izvedeno istraživanje, različite dužine trajanja tretmana u različitim studijama, različite tipove i podtipove intervencija, drugačiju starosnu i polnu kompoziciju itd. Drugim riječima, modelom se pretpostavlja da su sve studije nesavršeni izrazi mističnog, idealnog, jedinstvenog parametra koji postoji negdje u dubokoj istini univerzuma. Taj parametar je jedinstven i fiksan, odnosno jasno određen.

Prevedeno u matematički jezik, rezultat svake pojedinačne studije se može izraziti kao: parametar + greška uzorkovanja, pri čemu se pretpostavlja da greška uzorkovanja može biti pozitivna i negativna, tačnije da je normalno distribuisana oko parametra sa već definisanom varijansom. Evo to isto izrečeno hard-core notacijom:

$$Y_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Fiksni model se može i grafički prikazati (Slika 17.5). Na slici vidimo tri oblika. Kvadrati predstavljaju rezultate dobijene na uzorcima, trougao stvarnu vrijednost parametra, a krugovi ono što bi se dobilo na uzorku da ne postoji greška uzorkovanja.



Slika 17.5: Prikaz modela jedinstvenog / fiksnog efekta (analogno vizualizaciji Borenstein et al., 2021, *Introduction to meta-analysis*).

Šta je model slučajnih efekata?

Međutim, iz gore navedenog balkansko-japanskog poređenja se može zaključiti da bi bilo realističnije pretpostaviti da usljed snažnih kulturoloških razlika postoje dva različita parametra: jedan za Balkan, drugi za Japan. Ta pretpostavka podrazumijeva da ove dvije studije lutaju oko svojih parametara, a da mi u najboljem slučaju možemo kombinovanim efektom uhvatiti *prosjek tih parametara*. Štaviše, znajući da postoje populacije različitih slojeva - ne samo one koje se razlikuju po državama i regionima, nego i populacije koje su definisane različitim mjernim instrumentima (jer instrumenti koji mjere mentalno zdravlje imaju različite karakteristike), tipova i trajanja intervencija, polnih i starosnih kompozicija - znamo i da imamo toliko mnogo parametara. Ipak, u suštini smo zainteresovani da utvrdimo njihov prosjek - koji ima svoju varijabilnost - a efekti različitih slučajnih varijabli nam nisu u fokusu, mada ih priznajemo da postoje. Zato se ovakav pristup meta-analitičkom proračunu naziva model slučajnih efekata (engl. *random-effects model*). Uzbudljivim matematičkim jezikom se on iskazuje na sljedeći način:

$$Y_i = \mu_\theta + \zeta_i + \varepsilon_i, \quad \zeta_i \sim \mathcal{N}(0, \tau^2), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

Šta ovdje imamo? Za početak umjesto jedne greške uzorkovanja sada imamo dvije. Radujete li se novim grčkim slovima? Čudno izvijeno slovo zeta (ζ_i) predstavlja **razliku između stvarnog parametra određene grupe studija** (npr. onih sa Balkana koji koriste taj i taj instrument) **i prosjeka svih parametara** (μ_θ), dok ε_i (grčko slovo epsilon) i dalje označava **odstupanje pojedinačne studije od parametra određene grupe studija** (tj. balkanskih). Odnosno, u teoriji je moguće da se provede više studija na Balkanu koje koriste isti instrument, za koje se očekuje da imaju istu zetu ζ_i, ali i različite epsilone ε_i. (Kakav bajkovit jezik.) Ako se ograničimo samo na razlike u regionu i zanemarimo druge moguće izvore sistemske varijanse, onda se konkretni rezultati dobijeni u našim studijama mogu izraziti na sljedeći način:

$$Y_{Laki} = \mu_{Svijet} + \zeta_{Balkan} + \varepsilon_{Lakic}, \quad Y_{Yasui} = \mu_{Svijet} + \zeta_{Japan} + \varepsilon_{Yasui}$$

Dakle, na konkretan rezultat uticaj ima ono što se dešava u opštem čovječanstvu, ali i kulturološko-geografska regija, te specifičnosti nekog pojedinačnog uzorka. U teoriji biste mogli zamisliti još mnogo zeta (ζ) - kakav divan prizor - koje se mogu

odnositi na specifične instrumente koji se koriste za mjerenje, polnu distribuciju i ko zna šta sve ne.

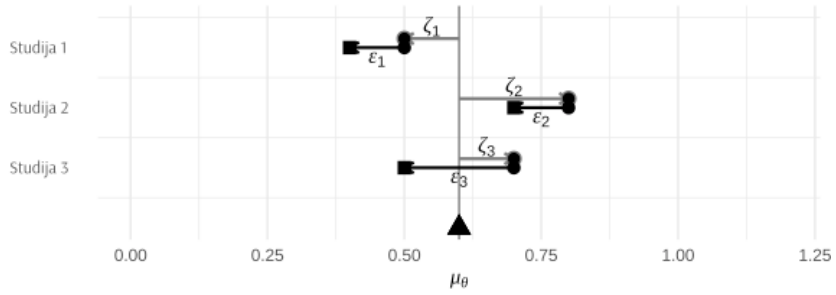
Nego, zašto vas uopšte ovim mučim? Zato što je razumijevanje konceptualizacije ovog modela slučajnih efekata bitno iz tri razloga. Prvo, ovom formulom smo zakoračili u važnu statističku oblast **dekompozicije varijanse**, odnosno razdvajanja onog što vidimo na uzorku na sistematske i nesistematske izvore varijabilnosti. To je pristup kojeg prate mnoge statističke tehnike čemu ćemo se više posvetiti u poglavlju o regresionoj analizi. Drugo, s obzirom da imamo mnoge izvore varijabilnosti postajemo svjesni da kada procjenjujemo taj kombinovani efekat zapravo procjenjujemo neki *prosjek prosjeka*. Dodatni izvori raznolikosti rezultata na uzorcima - odnosno njihove **heterogenosti** (engl. *heterogeneity*) - nam mogu biti interesantni, ali moguće je isto tako da ih nikada ne saznamo. Treće, ovaj model savjesno priznaje da postoje dodatni izvori varijabilnosti između studija koji ne ovise o veličini uzorka i varijabilnosti unutar studije. Čak i kada bismo ispitali čitavu populaciju Balkana ne bismo mogli procijeniti parametar za čovječanstvo, zato što je Balkan samo jedan član šire populacije (čovječanstva).

U vezi s tim, još jedno grčko slovo je ostalo neobjašnjeno (τ, tau), pa ćemo se sada i sa njim porvati. Kao i ostala grčka slova, ono označava parametar, a ovdje je taj parametar standardna greška parametara (standardna devijacija odstupanja geografskih regiona, θ_i) oko centralnog parametra (čovječanstva, μ_{θ_i}). Logično, zar ne? U kontekstu meta-analiza smo rekli da koristimo kvadrirane standardne greške, pa se zato distribucija odstupanja prikazuje ovako: $\zeta_i \sim \mathcal{N}(0, \tau^2)$. Za aritmetičku sredine te distribucije se pretpostavlja da je jednaka nula, a to znači da kada bismo mogli izmjeriti sve geografske subpopulacije, njihov prosjek bi bio jednak prosjeku čovječanstva.

Ovaj komplikovani dio je moguće lakše razumjeti grafičkim prikazom modela slučajnih efekata (Slika 17.6). Kvadrati ponovo predstavljaju rezultate dobijene na uzorcima, trougao aritmetičku sredinu parametara, a krugovi ono što bi se dobilo na uzorku da ne postoji greška uzorkovanja unutar subpopulacije. Drugim riječima krugovi pokazuju parametre tih subpopulacija.

Kakve su praktične posljedice razlike između ova dva modela?

U praksi se razlike između modela fiksnog efekta i modela slučajnih efekata svode na razlike u načinu izračunavanja pondera ko-



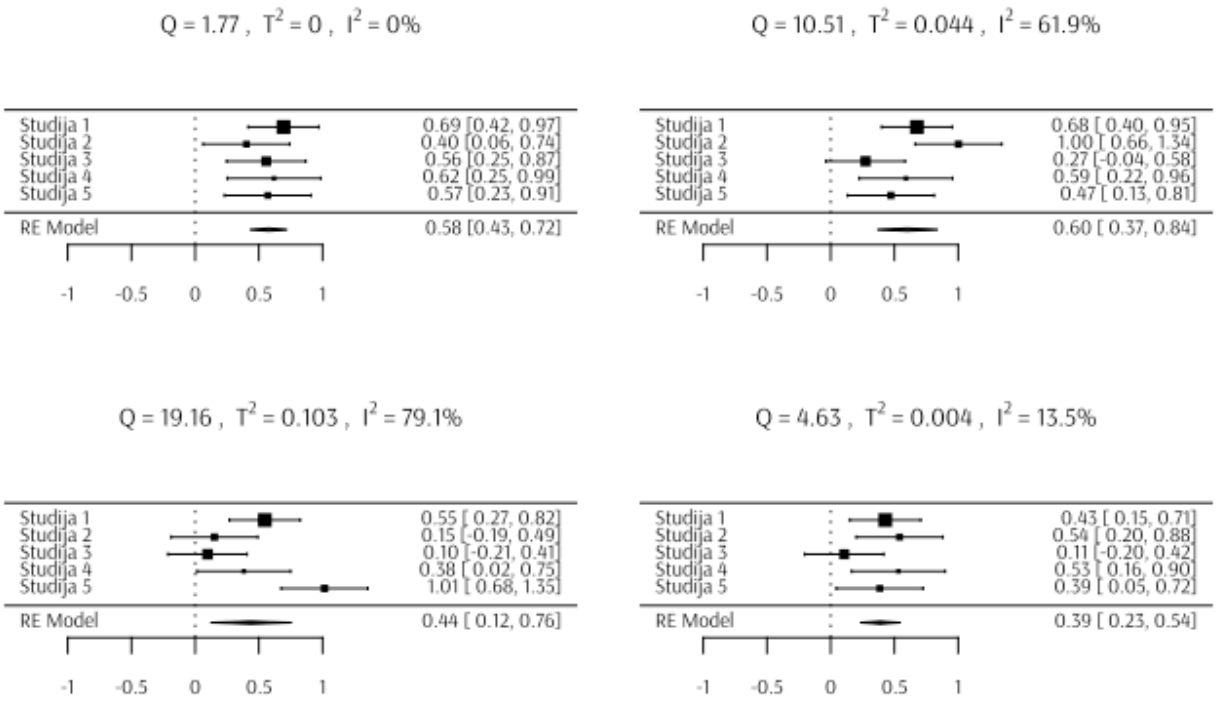
Slika 17.6: Prikaz modela slučajnih efekata (Inspirisano prikazom u Borenstein et al., 2021).

jima se množe efekti pojedinačnih studija pri izračunavanju kombinovanog efekta. Već smo vidjeli da je ponder za fiksni efekat vrlo prost i predstavlja inverznu varijansu greške izračunate na uzorku, $\frac{1}{V_i}$ (tj. inverznu kvadriranu standardnu grešku). Ponder za slučajne efekte (W^*) ima malu modifikaciju, odnosno prilјepak:

$$W_i^* = \frac{1}{V_i^* + T^2}$$

Ono što je novina (osim zvjezdice / asteriska $*$ koji je tu da razlikuje dvije metode) jeste ovo T^2. To T^2 je uzoračka procjena populacione varijanse greške između subpopulacija (τ^2) i predstavlja mjeru korekcije pondera. Ona se izračunava na osnovu ukupne razlike između pojedinačnih studija i prosjeka (tzv. Q mjera; ali kako bih vas malo poštedio, formule stavljam u fusnotu).[14] Ono što je bitno i što možete da vidite iz formule za ponder je da što je T^2 veće, to će ponder biti manji. Šta to zapravo znači? To znači da što je veća raznolikost rezultata između studija, to će ponderi pojedinačnih studija biti manji. A šta ako sve studije imaju isti rezultat? U tom slučaju je $T^2 = 0$, te će veličina pondera za model slučajnih efekata biti jednaka veličini pondera za model fiksnog efekta. Štaviše, nije nužno da statistici budu potpuno jednaki za različite studije, dovoljno je da se intervali povjerenja svih studija međusobno barem malo preklapaju pa da T^2 bude zanemarivo. Ovo se može lijepo uočiti na grafičkom primjeru na Slici 17.7, gdje donji desni grafikon prikazuje baš takav slučaj.

[14] Za one najhrabrije, $Q = \sum_{i=1}^{k} W_i \left(Y_i - M\right)^2$, a $T^2 \propto Q - (k-1)$. Ovo znači da je Q ponderisana suma kvadrata odstupanja pojedinačnih veličina efekta od njihovog prosjeka. Drugim riječima, Q direktno operacionalizuje raspršenje oko mjere prosjeka, dok T^2 dovodi u vezu to raspršenje sa onim što bi se očekivalo ako su ta raspršenja plod proste šanse, odnosno greške uzorkovanja. Da biste shvatili zašto $k-1$ označava količinu varijabilnosti koja bi bila uobičajena, odnosno očekivana, morate sačekati neku kasniju lekciju.



Slika 17.7: Poređenje indeksa heterogenosti.

Kako se iskazuje heterogenost rezultata u meta-analizi?

Već sam naglasio da se količina raznolikosti rezultata pojedinačnih studija u kontekstu meta-analiza naziva heterogenost. Bitno je nekako saznati količinu heterogenosti jer nam to govori o tome u kojoj mjeri različite studije - koje koriste različite manje ili više različite metode - procjenjuju jedan ti isti zajednički parametar. Upoznali smo već dvije mjere koje nam na to ukazuju (T^2 i Q), ali obje su razumljivije za tumačenje, odnosno rade svoj posao, u situacijama kada heterogenosti nema. Konkretno, kada je $T^2 \approx 0$, odnosno $Q \leq k - 1$, znamo da su rezultati studija u velikoj mjeri homogeni. Nažalost, ove mjere nemaju intuitivno značenje u izražavanju količine heterogenosti kada ona zaista postoji, budući da njihove apsolutne vrijednosti zavise od preciznosti procjene pojedinačnih studija (T^2) ili broja uključenih studija (Q). U naučnim člancima ćete ipak ponegdje sresti T^2, a još češće Q sa pripadajućom p-vrijednošću koja se može izračunati. Međutim, mjera koja se ubjedljivo najčešće pojavljuje kao mjera heterogenosti je I^2. I^2 možete zapamtiti

kao *Indeks Intuitivne Interpretacije* meta-analitičke heterogenosti (mada je onda trebalo da se nazove I^3?!). Indeks zato što se radi o jednom broju koji predstavlja pokazatelj, a "intuitivan" zato što je smještena na standardizovanu procentnu skalu od 0% do 100%. Matematički se izvodi iz Q (a teorijski i iz T^2):

$$I^2 = \frac{Q-(k-1)}{Q} * 100\% = \frac{T^2}{T^2+V_Y} * 100\%$$

Drugi dio formule nam je važan, jer pruža uvid u to kako se u statistici realizuje pomenuti princip *dekompozicije varijanse*. Naime, I^2 predstavlja omjer 1) varijabiliteta između studija (T^2) i 2) ukupnog varijabiliteta kojeg čine raznolikost između studija i unutar studija ($T^2 + V_Y$). Ako bi sve pojedinačne studije imale istu vrijednost onda je $T^2 = 0$, pa je $I^2 = 0$ čak i ako imamo jednu studiju sa 1000 ispitanika i drugu sa 10 - što znači da bi imale drastično različite intervale poverenja. To znači da je $I^2 = 0\%$ kada su nalazi pojedinačnih studija homogeni. S druge strane, zamislite da imamo dvije studije sa veoma velikim uzorcima (npr. 100 000 ispitanika) koji zbog toga imaju vrlo niske vrijednosti kvadrirane standardne greške V_Y, ali da između njih postoji razlika u procjeni zajedničkog parametra. Tada se osnovni dio formule svodi na T^2/T^2, pa će $I^2 \approx 100\%$. To su krajnji slučajevi (mada se 0% javlja u gotovo polovini studija), a autori ove mjere[15] su uz oprez preporučili okvirnu semantičku klasifikaciju heterogenosti na one sa malom ($I^2 \approx 25\%$), umjerenom ($I^2 \approx 50\%$) i velikom ($I^2 \approx 75\%$) heterogenošću. Uz oprez u tumačenju tog odnosa varijanse pravog efekta i greške uzorkovanja, treba imati na umu i to da sve gore navedene mjere heterogenosti imaju više smisla kada meta-analiza broji barem desetak studija.

[15] Higgins et al., 2003, https://doi.org/10.1136/bmj.327.7414.557

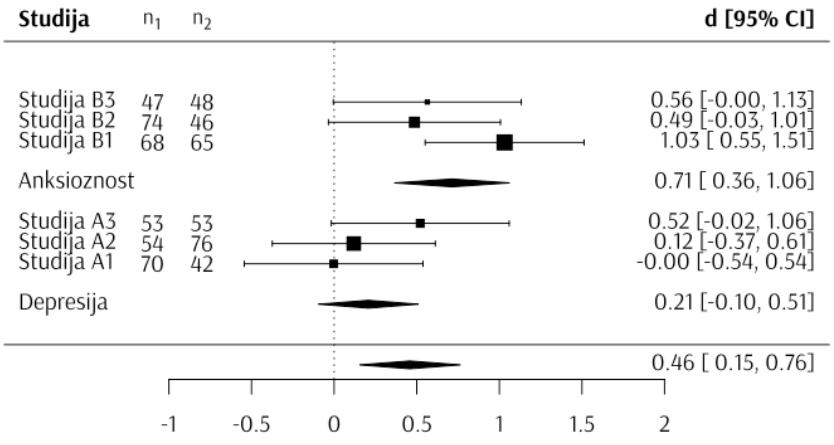
Šta su posljedice opažene heterogenosti rezultata?

Homogeni rezultati nam ulivaju više povjerenja u parametar, pogotovo ako su se studije značajno razlikovale u metodskim odlukama. Ali šta kada opazimo heterogene rezultate, čemu nas to dalje vodi? Za početak ću naglasiti da većina eksperata smatra da bez obzira na to kolika je ta heterogenost, trebalo bi uvijek koristiti metod slučajnih efekata za izračunavanje kombinovanog efekta. Kao što smo vidjeli, ako su rezultati potpuno homogeni dobićemo jednake rezultate kao i za metod slučajnih efekata, a svaka porast heterogenosti će biti adresirana ovom metodom. Međutim, šta nam govori postojanje heterogenosti?

Tada znamo da imamo posla sa moderatorskim varijablama. **Moderatori** (engl. *moderators*) u kontekstu meta-analize su varijable koje potencijalno mijenjaju intenzitet efekta između ispitivanih varijabli. Dakle, kada ustanovimo da raznolikost postoji možemo ispitati i njihov eventualni statistički uticaj, tako što ćemo ih uvesti u dodatne analize. Naravno, da bi se izvela analiza moderatora, potrebno je da iz originalnih studija izvučemo neophodne podatke za tu provjeru. Kategorički moderatori su često pol, vrsta mjernog instrumenta, tip intervencije, tip studije (npr. eksperimentalne naspram korelativnih), geografski region u kojem je istraživanje izvedeno. Najčešći dimenzioni moderatori su godine starosti, godina publikacije rada, prosječno trajanje tretmana. Pomalo paradoksalno, pol može biti operacionalno definisan kao dimenzioni moderator ako lista studija ne sadrži posebne studije samo sa muškim, samo sa ženskim i sa mješovitim uzorcima, nego kada kao mjeru uzimamo procentualnu zastupljenost jednog od njih (npr. žena u svakom uzorku). Isto tako starost može biti tretirana kao kategorički moderator ako su u nekim studijama ispitanici bili isključivo odrasle osobe, a u drugim samo adolescenti.

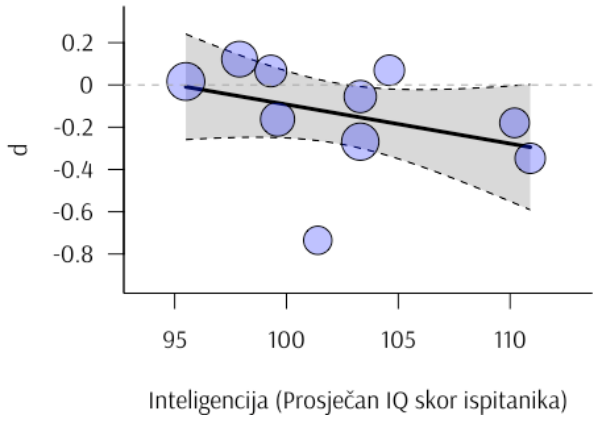
Najčešći kategorički i dimenzioni moderatori

Nakon što se identifikuju potencijalni moderatori, provodi se ili **analiza podgrupa** (za kategoričke varijable) ili **meta-regresija** (za dimenzione varijable). Analizom podgrupa se jednostavno uporede kombinovani efekti između grupa, dok je meta-regresija postupak kojim se procjenjuje postoji li korelacija između vrijednosti moderatora unutar studija i veličine efekta. U svakom slučaju, cilj ovih analiza je da se utvrdi da li se određenim moderatorima može objasniti zašto različite studije pokazuju različite veličine efekta, tj. da li svojstva studija ili uzoraka sistematski utiču na rezultate.



Slika 17.8: Tipičan primjer meta-analize podgrupa.

Slika 17.8 prikazuje fiktivne rezultate meta-analize efekta psihoterapije na tretman anksioznosti i depresivnosti. Na dnu vidimo ukupan efekat, ali nam analiza podgrupa pokazuje da bi tip mentalnog problema bio moderator, jer izgleda da je ova vrsta psihoterapije uspješnija u tretiranju anksioznosti nego depresivnosti (veći skor označava snažniji efekat).



Slika 17.9: Tipičan primjer meta-regresije.

Slika 17.9 pokazuje kako se grafički saopštava jedna meta-regresija. Na ovom fiktivnom primjeru efekta psihoterapije na nivoe anksioznosti svaki prikazani krug označava d vrijednost za datu studiju, a njihova veličina odgovora njihovom ponderu. U ovoj meta-analizi je snižavanje anksioznosti, odnosno uspješan tretman, bilo operacionalizovano negativnim vrijednostim što vidimo da se dešavalo u većini studija. Na x-osi se navodi potencijalni moderator, a on je u ovoj studiji inteligencija ispitanika. Grafikon nam otkriva da postoji blagi efekat visine inteligencije na uspješnost ovog psihoterapijskog tretmana, s obzirom na to da postoji negativan trend veze između prosječne inteligencije u uzorku i uspješnosti tretmana što je bilo operacionalizovano negativnim d-vrijednostima.

Da li pristrasnost u objavljivanju utiče na rezultate meta-analiza?

Volio bih da sada vaš tok misli izgleda ovako: "Dobro, ubijeđeni smo da su meta-analize vrhunski ljudski pronalazak i sjajno je to što postoji razrađeni sistem provjere moderatora. Ali zar nismo već pričali o tome da su nalazi koji se objavljuju u originalnim člancima često pristrasni, odnosno da je mnogo njih nereplikabilno i da precjenjuju snagu efekata? Šta je sad sa

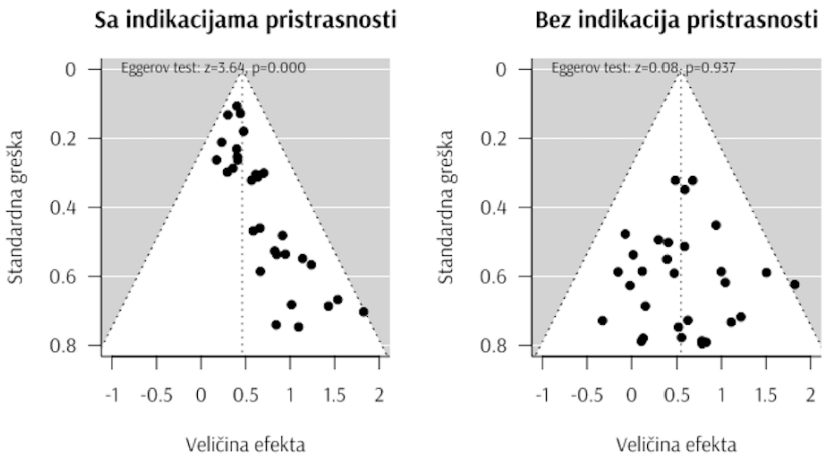
onim principom sa početka knjige 'ako smeće uđe, smeće će i izaći'?" Čak i ako nisam pogodio vaš tok misli, srećom su tako razmišljali eksperti koji su itekako bili svjesni gorkog problema **pristrasnosti u objavljivanju** (engl. *publication bias*). Zamislite zaista da je 100 istraživačkih timova provjeravalo efekat slušanja barokne muzike na učenje statistike i da je njih pet dobilo statistički značajan pozitivan efekat (baš onoliko koliko bi i očekivali za Grešku tipa 1). I upravo tih 5 studija onda završi u naučnim časopisima, a podaci iz ostalih 95 istraživanja i dalje nesretno tavore u ladicama ili po penzionisanim elektronskim memorijskim uređajima. Kada bismo uradili meta-analizu koristeći samo objavljene studije, dobili bismo jasne rezultate efekta slušanja barokne muzike na poboljšavanje učenje statistike. Time bi meta-analiza obavila potpuno suprotan zadatak i zapečatila neistinu na duže vrijeme. Ovaj problem selektivnosti rezultata i distorzirane slike koju usljed toga dobijamo ima i svoje ime: **problem efekta ladice** (engl. *file drawer problem*).[16]

[16] Termin slikovito opisuje sudbinu studija sa statistički neznačajnim rezultatima koje, umjesto da budu objavljene, ostaju zaboravljene u ladicama radnih stolova istraživača. Očito je ovaj naziv smišljen u pred-digitalno doba.

Možemo li nekako otkriti tu pristrasnost?

Da, postoji elegantan vizuelni način da u određenoj meta-analizi detektivski ispitamo da li možda svjedočimo takvoj situaciji. Grafikon lijevka (engl. *funnel plot*) je zapravo dijagram raspršenja gdje na X-osi imamo veličinu efekta studije (bez intervala povjerenja), a na Y-osi neku mjeru preciznosti procjene (najčešće invertovanu standardnu grešku, odnosno ponder, mada to može biti i broj ispitanika u uzorku). Zašto baš ime "lijevak"? Zato što bi u idealnom svijetu bez pristrasnih objava, rezultati prikazanih studija morali formirati oblik lijevka - na dnu bi se naširoko šetali rezultati malih studija, a pri vrhu bi se nalazili rezultati velikih, preciznih studija.

Dakle, logika je prosta: očekuje se da efekti manje variraju u studijama sa većim uzorcima. Ako bismo mogli ispitati čitavu populaciju (i sve moguće subpopulacije), dobili bismo pravi parametar. Što se više približavamo toj veličini uzorka, približavamo se i pravom efektu. Zato bi varijabilnost veličine efekata morala biti sve veća što su uzorci manji, ali ono što je posebno bitno jeste da bi to širenje moralo biti simetrično oko procjene parametra. Kada je grafikon asimetričan - a to je slučaj kada nedostaju studije sa malim uzorcima koje pokazuju negativne efekte - to nam ukazuje da nešto "smrdi". Slika 17.10 prikazuju dva tipična, a potpuno različita slučaja. Statistički softveri nude i dodatne testove statističke značajnosti te asimetrije od kojih je jedan (tzv. Eggerov test) prikazan na istoj slici.



Slika 17.10: Grafikoni lijevka koji prikazuju tipične slučajeve u pogledu pristrasnosti u objavljivanju.

Lijevi grafikon sugeriše snažnu pristrasnost, budući da statistički manje snažne studije čije su standardne greške sve veće što su više pri dnu y-ose, pokazuju mnogo veće veličine efekta koje su označene pojedinačnim kružićima. Desna slika pokazuje ono što bismo očekivali kada pristrasnosti ne bi bilo. Skoro sve studije su unutar lijevka, samo jedna je vani, ali je i to u redu, budući da tu i tamo očekujemo poneku studiju da iskoči jer lijevak prikazuje intervale povjerenja sa određenom greškom obuhvata (trougao ukazuje na 95%-tne intervale povjerenja koji se očito šire sa sve većom standardnom greškom). Ono što je bitno jeste da su sve studije simetrično raspoređene oko srednje procjene. Naravno, ovo su idealni slučajevi, a u stvarnosti se gotovo uvijek javlja nešto između ova dva primjera.

Šta ako se ustanovi pristrasnost?

Predloženo je nekoliko načina nošenja sa pristrasnošću, odnosno procjenom njene količine i potrebnom korekcijom kada se procjenjuje stvarni efekat. Ovdje opisujem najpopularniju proceduru, a ostale su opisane u referentnim radovima.[17] Radi se o **skrati i popuni postupku** (engl. *trim-and-fill procedure*) koji koristi grafikon lijevka.[18]

Operacija potkraćivanja (trimovanja) i popunjavanja (filovanja) se izvodi ovako:

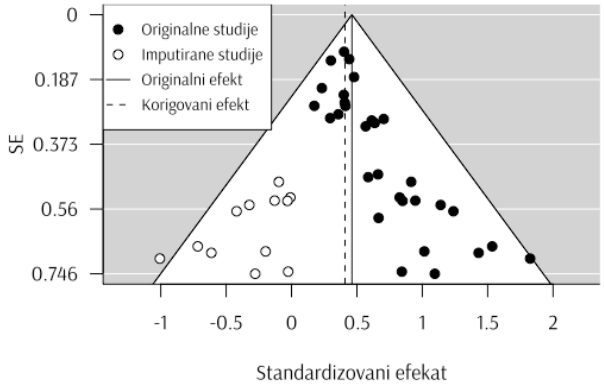
1. Prvo se uklone studije koje pokazuju pristrasnost.
2. Zatim se procijeni "stvarni" efekat na osnovu preostalih, "nepristrasnih" studija.

[17] npr. Afonso et al., 2024, https://doi.org/10.1007/s40279-023-01927-9

[18] Duval & Tweedie, 2000, https://doi.org/10.1111/j.0006-341X.2000.00455.x

3. Slijedi kreativna faza popunjavanja kada se dodaju hipotetičke studije koje bi imale potpuno iste veličine efekta i standardne greške samo sa druge strane od centrirane vrijednosti kako bi se stvorila ravnoteža.

Rezultat kompletne procedure je vidljiv na Slici 17.11. Treba napomenuti da je moguće i izračunati novi interval povjerenja za kombinovani efekat koji se prikazuje zajedno sa starim i koji pokazuju u kojoj mjeri je pristrasnost mogla uticati na dobijene rezultate.



Slika 17.11: Prikaz procedure potkraćivanja i popunjavanja radi korekcije pristrasnosti.

Međutim, neophodno je naglasiti da se ovim procedurama može tek ukazati na stepen rizika od pristrasnosti, ali da prilagođeni efekat nije nužno bolja procjena stvarnosti. Oprez je pogotovo preporučen kada se radi o meta-analizama sa malim brojem uključenih studija, pogotovo ako one ne odražavaju dovoljnu raznolikost po mjeri preciznosti koja se koristi (tj. nema dovoljno studija i sa malim i sa velikim uzorcima). Trebalo bi biti i svjestan da se pristrasnost objavljivanja možda dešavala i u drugom smjeru, odnosno da u nekom trenutku urednici nisu više htjeli objavljivati studije koje potvrđuju efekat, budući da to više nije bilo inovativno. Uz to, sasvim je moguće da se radi o stvarnoj vezi između veličine efekta i veličine uzorka[19] jer je za manje studije neophodno provoditi intenzivnije intervencije, te se u njima obično i regrutuju osobe koje su u većoj potrebi i podložnije efektima intervencijama.

[19] Vidjeti npr. Harrer et al., 2021, https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pub-bias.html

Za kraj, postoje li još neka ograničenja meta-analiza?

Osim činjenice da je vrlo teško razaznati u kojoj mjeri su rezultati nekih meta-analiza proizvod pristrasnosti u objavljivanju, postoje još neki pominjani problemi koje bi trebalo da uzimamo

u obzir. Jedan od njih jeste problem "krušaka i (japanskih) jabuka" budući da je razložno postaviti pitanje koliko smisla ima kombinovati studije koje koriste različite instrumente, tretmane i karakteristike ispitanika. Da, korištenje metoda slučajnih efekata adresira taj problem i još uz to možemo da ispitamo moderatore, ali će često postojati skriveni moderatori koji nisu otvoreno opisani u originalnim studijama. Drugi veliki problem je vezan za kvalitet originalnih studija i trenutak je da se opet naglasi deviza "*garbage in, garbage out*". Koliko god bile sofisticirane statističke metode koje meta-analize koriste (a samo smo načeli njihove varijante), njen kvalitet ultimativno ovisi o kvalitetu - ali i o broju - pojedinačnih studija na kojima se zasniva. Zato je potrebno pažljivo čitati dio o selekciji studija, odnosno rigoroznosti u njihovom odabiru i transparentnosti samih studija, odnosno dostupnosti podataka iz tih studija. Kako bi uprkos ovih ograničenja meta-analize održale status našeg najboljeg alata za sintezu naučnog znanja razvijene su i čekliste koje vode prvenstveno autore - ali i čitaoce - kroz preporučeni protokol. Na internet stranici www.prisma-statement.org se nalaze Preporučene stavke za izvještavanje sistematskih pregleda i meta-analiza (engl. *Preferred Reporting Items for Systematic reviews and Meta-Analyses*; *PRISMA*) koje predstavljaju standard pri izradi ovakvih studija. Svakako preporučujem čitanje tih preporuka, uz ostalu preporučenu literaturu za zainteresovane.[20]

U sljedećem poglavlju upoznaćemo se sa bejzijanskom statistikom – pristupom koji poput meta-analiza takođe naglašava kumulativnu prirodu naučnog znanja - ali kroz potpuno drugačiji filozofski pristup vjerovatnoćama. Vidjećemo da bejzijanska statistika nudi elegantan okvir za racionalno ažuriranje naših prethodnih uvjerenja u svjetlu novih dokaza i prirodno integrisanje ranijeg znanja u statističku analizu. Budući da ovaj pristup postaje sve popularniji u cjelokupnoj nauci, izuzetno bitno je da razumijete njegove osnove kako biste mogli pratiti savremenu literaturu.

[20] Knjige koje detaljno i na jednostavan način tretiraju meta-analize su: 1. Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2021). Introduction to meta-analysis (2nd ed.). Wiley. 2. Cumming, G., & Calin-Jageman, R. (2024). Introduction to the New Statistics: Estimation, Open Science, and Beyond (2nd ed.). Routledge. 3. Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL and London: Chapmann & Hall/CRC Press.

Poglavlje 18

Uvod u bejzijansku statistiku

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Šta je suštinska razlika između bejzijanskog i frekvencionističkog pristupa zaključivanju?
- Šta su apriorne, posteriorne i distribucije izglednosti i kako se one kombinuju u bejzijanskoj analizi?
- Kako se interpretiraju ključni bejzijanski testovi hipoteza kao što su vjerovatnoća smjera (pd), ROPE i Bayesov faktor?
- Koje su glavne prednosti i izazovi primjene bejzijanskih metoda?

Već nekoliko puta smo kroz tekst pomenuli subjektivističku vjerovatnoću kao i bejzijansku statistiku, pa je došao i trenutak da to konačno obradimo. Ovo poglavlje predstavlja uvod koji prikazuje najbitnije elemente tog revolucionarnog pristupa statističkom zaključivanju koji neke koncepte obrće naglavačke. Krenućemo sa primjerom.

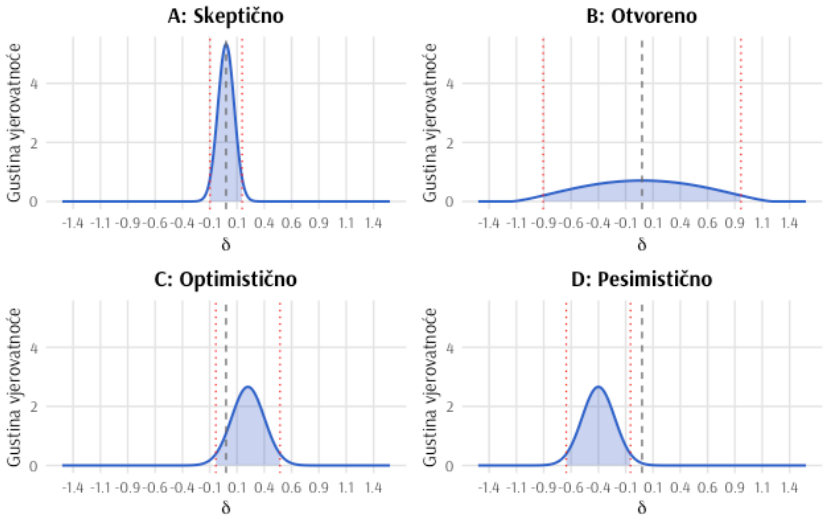
Već smo toliko puta pomenuli baroknu muziku i njen mogući efekat na učenje statistike da se pitam da li neki od vas čitaju ove redove uz zvukove iste. Ali sad jedno ozbiljno pitanje: koliko ste zaista uvjereni da bi slušanje barokne muzike mogla da pomogne u shvatanju ovog gradiva bez obzira na to što sam savjesno naglašavao da su prikazivani fiktivni rezultati? Neki od vas će razmisliti i odgovoriti A: *"Nema šanse da to ima ikakav efekat"*, drugi B: *"Možda tu nečeg ima, ali stvarno nemam pojma, čak ni u kom smjeru bi dejstvo bilo."*, treći C: *"Vjerujem da bi moglo malo pomoći."*, a četvrti D: *"To samo može značajno pogoršati ionako gadnu situaciju"*. Sve navedene izjave ispoljavaju određena uvjerenja koja ste možda stekli. Moguće je čak da ste ih mijenjali tokom

čitanja knjige. Međutim, sva ova uvjerenja su iskazana semantički grubo. Mogli bismo ih okarakterisati na sljedeći način: A = *zadrto skeptično*, B = *nemarno otvoreno*, C = *blago optimistično*, D = *ogorčeno pesimistično*. Ali nismo se uludo dosad bavili preciznim matematičkim jezikom da ga sad počnemo ignorisati. Zar ne bismo mogli snagu uvjerenja iskazati statistički?

Kako se snaga uvjerenja prevodi u statističku ravan?

E, upravo to možemo da uradimo ako se odlučimo da ova uvjerenja izrazimo u obliku njihove statističke raspodjele. Njih ćemo iscrtati tako što ćemo na x-osi predstaviti moguće vrijednosti parametara, a na y-osi snagu uvjerenja koja se mijenja ovisno o kojem parametru je riječ. Slika 18.1 prikazuje četiri gore navedena slučaja pretvorena u distribucije.[1] Hajde da vidimo šta nam sve govore ovi prikazi.

[1] Možda nekog zanima: A, C i D su normalne distribucije, a B je linearno transformisana beta distribucija (Beta*).



Slika 18.1: Prikaz tipičnih prethodnih / apriornih uvjerenja.

Kao što je rečeno, na x-osi se nalazi prostor mogućeg efekta u populaciji. U principu, parametar se može nalaziti bilo gdje između $-\infty$ do $+\infty$. Ove četiri osobe su prilično razumne, odnosno niko nije toliko lud da pretpostavi da bi efekat slušanja barokne muzike mogao prelaziti apsolutnu vrijednost od $1.4d$, pa su grafici ograničeni na taj raspon. Na y-osi se nalazi snaga uvjerenja. Tačkaste uspravne linije ukazuju na interval unutar kojeg se nalaze srednjih 95% uvjerenja. Podsjeća li vas to na nešto? Naravno, na intervale povjerenja / povezanosti. Razlika je u tome što nam oni ovdje ne pokazuju jedan od mogućih intervala koji sa određenim procentom tačnosti uključuje parametar,

nego nam oni govore baš o tome u šta određena osoba vjeruje sa određenim procentom uvjerenja. Zato se oni i zovu drugačije: **intervali uvjerenja** ili **uvjerljivi intervali** (engl. *credible intervals*). Dakle, intervali uvjerenja predstavljaju interval vrijednosti na mjernoj skali koji sa određenim stepenom vjerovatnoće obuhvata parametar.

Definicija intervala uvjerenja

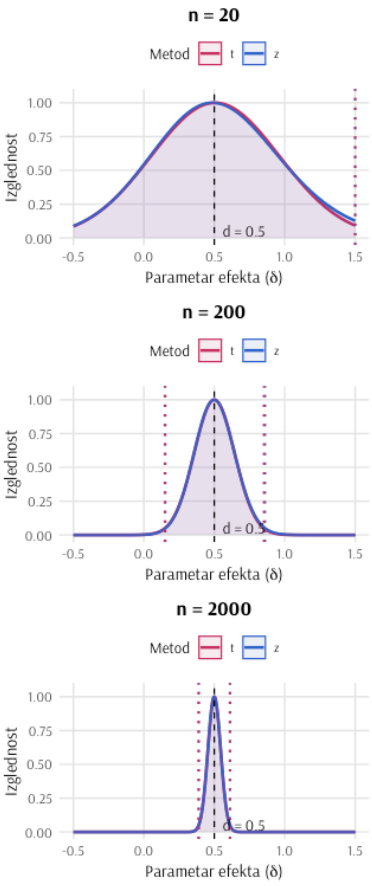
Vidimo da se na prikazanom primjeru razlikuju ta četiri intervala. Skeptik (A) ima vrlo zadrta uvjerenja na šta ukazuju visina distribucije i vrlo uski intervali uvjerenja od -0.14 do $+0.14$ koji ostaju u okviru onoga što bi Cohen nazvao zanemarivim efektom. Bezbrižno otvoreni istraživač (B) ima radikalno drugačiju distribuciju uvjerenja i dopušta sa 95% vjerovatnoće da se parametar nalazi negdje između -0.91 i $+0.91$, dakle dopušta čak i mogućnost velikog efekta (i sve između). Pažljivo oko će otkriti i jednu sličnost A i B istraživača; oba imaju centrirane distribucije na tački 0. Ipak, skeptik se drži nje kao dobro pijan bandere, a otvorenjak ne haje puno baš za tu vrijednost. Blagi optimista (C) je izjavio da bi barokna muzika mogla imati blagi efekat pa je distribucija centrira na vrijednost na $+0.20$ (što je granični skor za Cohenov mali efekat), ali gornja granica intervala uvjerenja seže do $+0.49$ (0.50 je već, nominalno gledano, umjereni efekat). Ipak, nije previše ubijeđen da efekat mora biti pozitivan, pa čak dozvoljava da efekat može biti nikakav ili blago neutralan (donji interval uvjerenja je -0.09). Konačno, ogorčenog pesimistu nose neke averzivne emocije kada pomisli na baroknu muziku te bi efekat morao biti negativan: interval uvjerenja se kreće od -0.69 do -0.11, a najvjerovatnije vrijednosti parametra su oko -0.40.

Dakle, ono što vidite su matematički iskazane distribucije snage uvjerenja, odnosno vjerovatnoće da je određeni parametar istinit. Dobar trenutak je i da se podsjetimo da subjektivistički orijentisani pristup tako i definiše vjerovatnoću: **vjerovatnoća jeste stepen uvjerenja koje neko ima o određenom ishodu**. Budući da u ovom slučaju imamo uvjerenja koja su prisutna prije nego li smo izveli istraživanje, u bejzijanskom pristupu ove vjerovatnoće zovemo **prethodnim, početnim ili apriornim vjerovatnoćama** (engl. *prior probabilities*). Da bi takva početna ili apriorna uvjerenja bila korisna ona moraju biti formalizovana kako bi se mogla kombinovati sa onim što nam podaci sa uzorka govore.

Subjektivistička definicija vjerovatnoće

Kako se u bejzijanskoj statistici razmatraju rezultati sa uzorka?

Recimo - nažalost, opet fiktivno - da je izvedeno jedno istraživanje i da je dobijen rezultat $d = +0.50$, odnosno da je "barokna grupa" imala za pola standardne devijacije bolji rezultat nego kontrolna. Kada se pojavi ovakav dokaz, šta bi trebalo da se dogodi sa početnim uvjerenjima racionalnih ljudi? Naravno, ona bi trebalo bi da se ažuriraju, odnosno izmjene u skladu sa dobijenim informacijama i to onoliko koliko su ti dokazi "teški". Kako se mjeri težina dokaza u kontekstu naučnih istraživanja? Kao što ste mogli i pretpostaviti, osnovnu ulogu ponosno igra veličina uzorka. Nisu jednako ubjedljiva istraživanja koja imaju ukupno 20 ispitanika (po 10 u kontrolnoj i eksperimentalnoj grupi), 200 ispitanika i 2000 ispitanika. Kao što znate sa većim brojem ispitanika se smanjuje mogućnost greške u procjeni parametra.



Slika 18.2: Distribucije izglednosti parametra u zavisnosti od veličine uzorka.

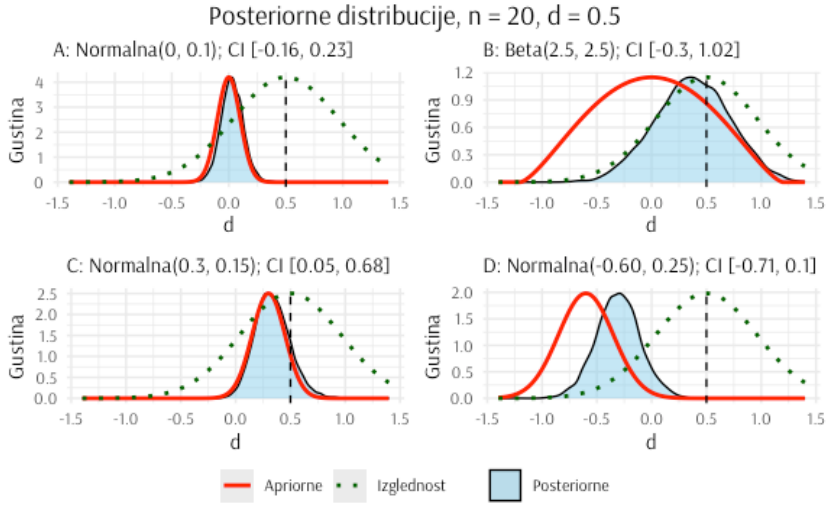
Probaćemo sada da razmišljamo obrnuto od onoga što smo navikli dosad (najavio sam već obrtanje stvari naglavačke). Ako znamo rezultat na uzorku - a to je zapravo jedino što znamo - onda možemo da pretpostavimo kakve su šanse da su različite vrijednosti parametra doveli do tog rezultata. Drugim riječima, ovdje eksplicitno pretpostavljamo da je parametar varijabilan, odnosno da su moguće njegove različite vrijednosti (tzv. prostor parametara). Pogledajte Sliku 18.2 koja prikazuje upravo to, što se na engleskom naziva *likelihood* funkcija, a na našem funkcija **izglednosti** ili **vjerodostojnosti** parametra (i ponekad inverzna vjerovatnoća).

Šta nam to ova tri grafikona izvedena za istu veličinu efekta prikazuju? Za početak za sve tri veličine uzorka je zajedničko da je najizglednije dobiti rezultat $d = 0.50$ na uzorku ako je $\delta = 0.50$. Međutim, vidimo i razlike: na uzorku od 20 ispitanika 95%-tni interval obuhvata moguće vrijednosti parametra od $\delta = -0.59$ do $\delta = 1.65$. S druge strane za uzorak od 2000 ispitanika obuhvata moguće vrijednosti parametra od 95%-tni interval obuhvata vrijednosti od $\delta = 0.33$ do $\delta = 0.67$.

Kako se prethodna uvjerenja ažuriraju kombinovanjem sa rezultatima sa uzorka?

Da vidimo sad kako bi empirijski nalazi sa uzorka trebalo da utiču na distribucije i intervale uvjerenja ako ih integrišemo sa prethodnim vjerovatnoćama parametra. Ta uvjerenja, odnosno distribucije vjerovatnoće parametra nakon što su empirijski podaci inte-

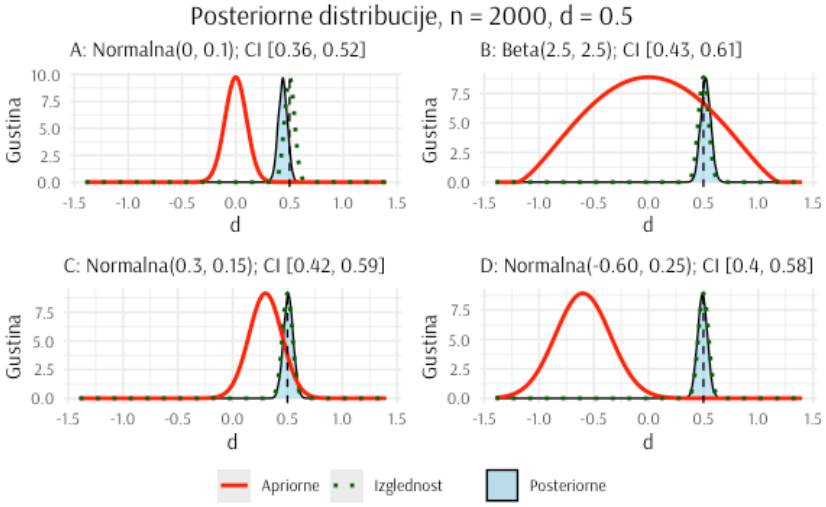
grisani u prethodne vjerovatnoće nazivamo **naknadne, ažurirane ili posteriorne distribucije vjerovatnoća** (engl. *posterior probability distributions*). Ako smo razumni, tako bi trebalo uvijek da funkcionišemo. Imamo neka uvjerenja i onda - u svjetlu novih podaka - korigujemo ista. Hajde da sad pogledamo na Slici 18.3 šta bi tačno trebalo da se desi sa naših četvoro tipičnih istraživača (A-D) predstavljenih na početku poglavlja ako smo na uzorku od ukupno 20 ispitanika dobili $d = 0.50$.



Slika 18.3: Posteriorne distribucije u zavisnosti od prethodnih vjerovatnoća, $n = 20$.

Uz početna uvjerenja i distribuciju vjerodostojnosti parametara uzimajući u obzir dobijeni efekat na uzorku od 20 ispitanika, za sva četiri istraživača su prikazana njihova ažurirana uvjerenja. Do njih se dolazi matematičkim putem, poštujući tzv. **Bejzovo pravilo** (engl. *Bayes rule*) kojim ćemo se uskoro pozabaviti. Kada pogledamo 95%-tni interval uvjerenja, primjećujemo da je *zadrti skeptik* (A) i dalje ostao jednako skeptičan jer se njegovo uvjerenje gotovo uopšte nije pomjerilo (apriorna i posteriorna distribucija uvjerenja se preklapa). Možemo čuti njegove misli: "Da, vidim rezultat, ali to je samo jedno mini-istraživanjce čiji je rezultat sigurno plod slučajnosti (ili ko zna čega drugog)". *Bezbrižno otvoreni* (B) već pokazuje drugačija uvjerenja i čujemo ga kako sebi kaže: "Ok, ovo mi već nešto govori. Mada i dalje sam otvoren za svašta nešto od malog negativnog efekta, do izuzetno velikog pozitivnog." *Blagi optimista* (C) ostaje pri onome što je ranije mislio: "Eto, ovo malo istraživanjce je samo pokazalo ono što sam i prije pretpostavljao. A, možda je čak efekat nešto veći." Konačno, tu je *ogorčeni pesimista* (D) koji kaže: "U redu, barokna muzika možda nije toooliko kontraproduktivno loša kako sam mislio, ali i dalje vjerujem da u najboljem slučaju ima

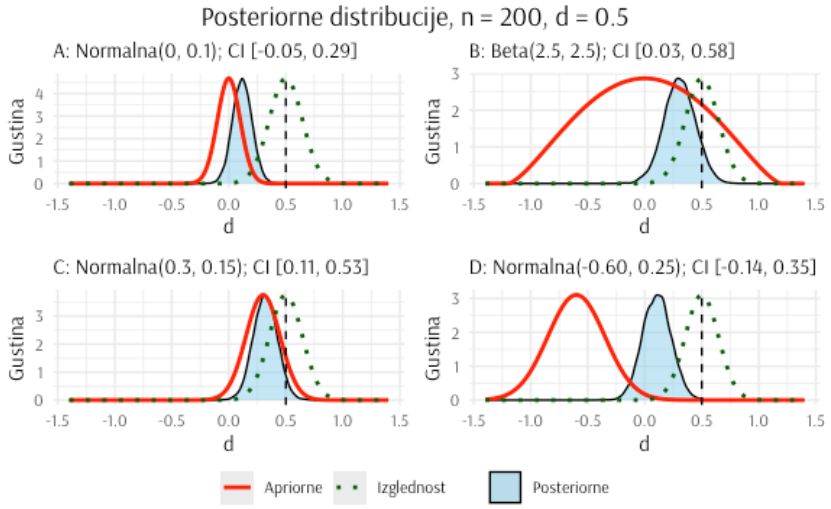
nulti efekat". Ako bolje razmislimo, svi ovi likovi principijelno razmišljaju s obzirom na ranije formirane stavove prije nego što su saznali za rezultate ovog malog istraživanja. Uzorci su mali i zato bi snaga mijenjanja uvjerenja trebalo da bude slaba. Ali šta ako su umjesto istraživanja sa 20 ispitanika saznali za efekat od $d = 0.50$ na uzorku od 2000 ispitanika?



Slika 18.4: Posteriorne distribucije u zavisnosti od prethodnih vjerovatnoća, $n = 2000$.

Na Slici 18.4 vidimo potpuno drugačiji prizor. Snaga empirijskog dokaza je toliko ubjedljiva da sad sve četvoro vjeruju u praktično isto! Sve četvoro bi trebalo da dođu do uvjerenja da se parametar nalazi u blizini vrijednosti 0.50, sa vrlo malim odstupanjima ovisno o početnim uvjerenjima. To znači da bi snažni empirijski dokazi morali pregaziti ranija uvjerenja - što bismo očekivali od racionalnih ljudi. A nauka je racionalan posao. Kada se pojave kvalitetni dokazi, naučnici moraju da mijenjaju svoja uvjerenja u skladu s njima, bez obzira da li su optimisti, zadrti skeptici ili ogorčeni pesimisti. Sve ovo gore opisano je logika bejzijanskog statističkog zaključivanja koji je sve prisutniji u sveopštoj nauci, ali koji vlada i mašinskim učenjem: imamo neka početna uvjerenja - > dolazimo do iskustvenih dokaza -> naša uvjerenja se ažuriraju. Prije nego što predstavim nešto više teorije, te saznamo i otkud ovo ime *bejzijansko*, bio bi red da vidimo šta bi se desilo sa posteriornim uvjerenjima ako bi uzorak bio srednje velik ($n = 200$) na kojem je isto tako dobijeno $d = 0.50$ u korist grupe koja je slušala baroknu muziku.

Kako ste mogli i pretpostaviti, na Slici 18.5 vidimo da su promjene uvjerenja bile "na pola puta" između onoga što se desilo sa $n = 20$ i sa $n = 2000$. Iako nisu potpuno identične, distribucije posteriornih uvjerenja su vrlo slične za *skeptika* i *pesimistu*



Slika 18.5: Posteriorne distribucije u zavisnosti od prethodnih vjerovatnoća, $n = 200$.

($Mdn \approx 0.10$), kao i za *otvorenog* i *optimistu* ($Mdn \approx 0.30$). Kao što smo vidjeli, sa sve većim uzorcima bi i uvjerenja različitih polaznih stajališta sve više konvergirala. Nadam se da je ovaj dio jasan, pa da možemo da pređemo na malo tehničkih stvari.

Šta je to onda bejzijanska statistika i otkud joj ime?

Bejzijanska (ili bayesijanska, bejsijanska) statistika je ime dobila po Thomas Bayesu (originalni izgovor: /beIz/, otuda preferencija za bejZijanska statistika). Bayes je bio protestantski sveštenik čiji rad o vjerovatnoćama je posthumno objavljen davne 1763. godine zahvaljujući prijatelju Richard Priceu koji ga je značajno unaprijedio. Kada već pominjem neka imena - evo još jedno: Pierre-Simon Laplace je nešto kasnije nezavisno od Bayesa formulisao ono što se u statistici obično naziva **Bayesovim pravilom ili teoremom** koja leži u osnovi bejzijanske statistike. Međutim, gotovo čevrt milenijuma je ovaj pristup ostao u zapećku statistike, pa je tek nedavni razvoj informacione tehnologije i otvorenije epistemologije doveo do njegovog procvata. Bejzijanska statistika trenutno predstavlja vodeći statistički pristup u mnogim granama teorijske i primijenjene nauke (npr. inženjerstvo), te čini osnovu mašinskog učenja.

Sve počiva na tom famoznom Bayesovom pravilu. Ono se tiče izvrnutog pristupa uslovnim vjerovatnoćama, pa bi bilo dobro da prije prelaska na njegovo pojašnjenje ponovite šta su to uslovne vjerovatnoće. Tada smo računali vjerovatnoće da je neki student psihologije žena (npr. $Pr(Y|X)$), kao i da je neka

žena student psihologije (npr. $Pr(X|Y)$). Konkretno, Bayesovo pravilo nam omogućava da izračunamo uslovnu vjerovatnoća nekog događaja $Pr(Y|X)$, ako znamo tri stvari: vjerovatnoću tog uslova $Pr(X)$, vjerovatnoću tog događaja $Pr(Y)$ i obrnutu uslovnu vjerovatnoću $Pr(X|Y)$. Ne zvuči Bog-zna-kako korisno, zar ne?![2] Evo notacije:

$$\Pr(Y|X) = \frac{\Pr(X|Y) * \Pr(Y)}{\Pr(X)}$$

[2] Mada su Bayes i Price ovom teoremom zapravo htjeli dokazati postojanje Boga: https://blogs.cornell.edu/info2040/2018/11/28/bayes-theorem-and-the-existence-of-god/

Da bismo shvatili logiku Bayesovog pravila pozabavićemo se konkretnim primjerom. Recimo da je izvedeno istraživanje u kojem je 100 ispitanika u trajanju od jednog minuta prepričavalo neku neugodnoj situaciju koju su doživjeli, te da su njihove izjave snimane kamerom gledano s lica, oči u oči (an fas). Istraživači su željeli da provjere da li je skretanje pogleda tokom prepričavanja (X, koje je imalo vrijednosti *da* i *ne*) indikator laganja (Y, takođe *da* i *ne*). Od 100 ispitanika, njih 90 je dobilo instrukciju da govore istinu, a njih 10 da lažu: $Pr(Y) = 10/100 = 0.10$. U grupi *lažova* 8 od 10 je vidno skretalo pogled tokom izvještavanja: $Pr(X|Y) = 8/10 = 0.80$. U ovom trenutku možete pomisliti: 80% onih koji lažu skreću pogled tako da je stvar prilično jasna? Međutim, i u grupi *iskrenih* je bilo onih koji su vidno skretali pogled - njih 9 od 90, što predstavlja samo 10% tog poduzorka. Sveukupno, 17 od 100 ispitanika je skretalo pogled $Pr(X) = (8 + 9)/100 = 17/100 = 0.17$. Pitanje glasi: ako je osoba skretala pogled koja je vjerovatnoća da je lagala[3], odnosno koliko je $Pr(Y|X)$? Račun je sljedeći:

[3] Obratite pažnju zašto je ovo drugačije od pitanja "Ako je lagala, da li je skretala pogled?" i zašto mnogo ljudi upada u zablude kada neki znak uzima kao signal latentnog ponašanja.

$$\Pr(Y|X) = \frac{0.80 * 0.10}{0.17} = \frac{0.08}{0.17} = 0.47$$

Ovo znači da je vjerovatnoća neznatno manja od 50% da je osoba lagala ako je skretala pogled tokom izvještavanja. Ako se pogleda omjer vjerovatnoća, dobija se da je $(1 - 0.47)/0.47 = 0.53/0.47 = 1.13$ puta veća vjerovatnoća da je osoba koja je skretala pogled govorila istinu nego da je lagala. S obzirom na to, možemo da zaključimo da skretanje pogleda nije pouzdan znak kojim možemo da utvrdimo da osoba laže. Drugim riječima, Bayesova teorema nam omogućuje da prosudimo kolika je vjerovatnoća da se opaženi događaj desi ako je određena hipoteza tačna, i čak je možemo uporediti sa vjerovatnoćom da je neka druga hipoteza tačna. Ovo drugo, gdje je rezultat bio 1.13, se naziva **posteriorni omjer** (engl. *posterior odds*) - koji predstavlja omjer vjerovatnoća dvije hipoteze koje se porede. Zar to i nismo htjeli od naših podataka sa uzorka, da nam ukažu na to koja od postavljenih hipoteza je vjerovatnija ako smo već

Posteriorni omjer

dobili određene podatke na uzorku? Ovo su ultra-jednostavni primjeri, pa ćemo proći i neke komplikovanije primjere. Ali prije toga bi bilo dobro da osvijestimo u čemu se sastoji ključna razlika između bejzijanskog pristupa i ranije prikazivane statistike zaključivanja (koja se naziva i frekvencionistička)?

Kako se bejzijanska statistika razlikuje od frekvencionističke?

Da bismo shvatili razilku pomoći će nam da formulu Bayesovog pravila svedemo na jednostavniju konceptualnu formulu, koju smo implicitno već uveli na primjeru sa četiri različita seta uvjerenja. Ta formula kaže da su naknadne ili posteriorne vjerovatnoće proporcionalne proizvodu podataka sa uzorka i prethodnih vjerovatnoća:

Posteriorna vjerovatnoća $\propto$ Izglednost $\times$ Apriorna vjerovatnoća

Ovo znači da je **apriorna vjerovatnoća** = $Pr(Y)$, da je **izglednost** = $Pr(X|Y)$, te da je **posteriorna vjerovatnoća** - ono $Pr(X|Y)$. Ako se neko pita, izostavljena komponenta formule Bayesovog pravila $Pr(X)$ služi da normalizuje vjerovatnoću na skalu od 0 do 1, te nema suštinsku ulogu.

Da se vratimo sada razlici između dosadašnje statistike i bejzijanskog pristupa. Ono čega u ranijim poglavljima nije bilo su apriorne vjerovatnoće. Dakle, dosadašnji pristup je podrazumijevao da u procesu potrage za parametrom prethodna uvjerenja nemaju nikakvu ulogu, odnosno da podaci iz datog istraživanja - pa bila to i meta-analiza - govore sve što treba da se kaže. Ako bismo željeli da asimiliramo ranije obrađenu statistiku zaključivanja u ovaj novi model, možemo reći da su sve moguće vrijednosti parametara imale jednaku apriornu vjerovatnoću. Ovo je najbolje pokazati na primjeru.

Recimo da je istraživače interesovalo da li je na uzrastu od 16 mjeseci većina djece, odnosno njih više od 50%, ($H_0 : \pi_+ = .50$)[4] sposobno da svrsta njima potpuno nepoznate predmete koji imaju neku vizuelnu sličnost u zajedničku grupu ako se ti predmeti nazovu istim imenom (tzv. zadatak kategorizacije). Na uzorku od 100 ispitanika, njih 58 je tačno uradilo zadatak $P_+ = 0.58$.[5] Za taj rezultat se dobijaju p-vrijednosti $p = .133$ (dvosmjerni test) i $p = .067$ (jednosmjerni test), ali ono što nas još više interesuje je 95%-tni interval povjerenja koji obuhvata vrijednosti 0.50: $95\%CI[0.48, 0.68]$. Dakle, dobijeni rezultati bi

[4] Možete li predložiti još neke načine kako se nulta hipoteza može definisati?

[5] Fiktivni podaci, ali je istraživanje sa manjim uzorkom na ovu temu objavljeno: Tutnjević & Lakić, 2018, https://doi.org/10.1080/17405629.2017.1325733

nam samo kazali da ne možemo da odbacimo nultu hipotezu. Sada ćemo da pogledamo bejzijanski pristup koji bi najviše odgovarao ovom proračunu. Rekli smo da bi u frekvencijskom pristupu apriorna vjerovatnoća bila jednaka za sve moguće vrijednosti parametra. Budući da ovdje pominjemo mjernu skalu od 0 do 1 kako bi distribucija tih vjerovatnoća izgledala? Očito, kao na Slici 18.6.

Slika 18.6: Neinformativna uniformna prethodna vjerovatnoća.

Na grafikonu je prikazana tzv. uniformna ili ravnomjerna apriorna distribucija koja operacionalizuje istraživačko potpuno neznanje. Ta distribucija je jedna od Beta distribucija, konkretno Beta(1,1). Grafikon prikazuje i 95%-tni interval uvjerenja koji seže od 0.48 do 0.67. Taj interval je gotovo identičan intervalu povjerenja - a zanemarive razlike dolaze usljed činjenice da čak i u ovom slučaju ipak unosimo neke informacije sa datom Beta distribucijom. Kao prvo korištenjem ove Beta distribucije smo ograničanili teorijski opseg parametra na mjernu skalu od 0% do 100% što znači da on ne može biti -50% ili +300% što osnovna frekvencionistička statistika dozvoljava, ma koliko to bilo apsurdno. Zanimljivo, za Beta distribuciju se broj informacija može pretvoriti u veličinu uzorka i ona predstavlja sumu njenih definišućih vrijednosti ($n = a + b = 1 + 1 = 2$). Odnosno, korištena Beta distribucija je imala jednaku težinu kao da je u kalkulaciju ugrađen nalaz istraživanja proveden na ukupno 2 ispitanika. U praktičnom smislu su ove razlike potpuno zanemarive. Međutim, ovo je pravi trenutak da se posvetimo pitanju različitih apriornih vjerovatnoća, njihove funkcije i smislenosti.

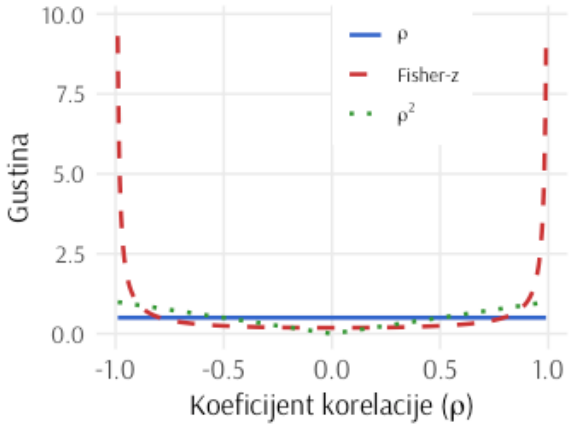
Kako se dijele prethodne vjerovatnoće i kakav je njihov efekat na naknadne distribucije ?

Vrlo gruba podjela kategoriše apriorne vjerovatnoće u tri grupe ovisno o količini informacija koje se ugrađuju u procjenu parametra: **neinformativne**, **slabo informativne** i **informativne**. Neinformativne prethodne vjerovatnoće se još nazivaju i objektivne, budući da se njima pretpostavlja (gotovo) apsolutno neuplitanje istraživača u procjenu. Primjer smo upravo vidjeli - a radi se o uniformnim distribucijama. Uniformne / ravnomjerne distribucije pretpostavljaju jednaku vjerovatnoću parametra duž svih teorijski mogućih vrijednosti. Za ispitivanje proporcija je to skala od 0 do 1, a za ispitivanje korelacija to je skala od -1.0 do +1.0. Laicima i onima koji su skeptični u uvođenje subjektivnih informacija u naučna istraživanja se ovo čini idealnim, jer smatraju da nauka mora zadržati punu objektivnost. Međutim, postoje dva pravca kritike "objektivne" prethodne vjerovatnoće.

Zašto neinformativne prethodne vjerovatnoće nisu potpuno neinformativne?

Prvi pravac se tiče formalnih aspekata budući da se ističe da ne postoji jedinstven način da postavimo uniformne distribucije. To možemo ilustrovati primjerom za pomenute koeficijente korelacije. Naime, trebalo bi da već znate da skala za koeficijente korelacije nelinearna, odnosno da korelacija od .20 nije duplo snažnija od .10 budući da korelacija od .20 predstavlja 4% zajedničkog varijabiliteta ($r^2 = .20 * .20 = .04$), dok korelacija od .10 predstavlja samo 1% zajedničkog varijabiliteta ($r^2 = .10 * .10 = .01$). To znači da postavljanje jednakih vjerovatnoća za proste koeficijente korelacije možda i nije najbolje rješenje. S tim u vezi, možda bi trebalo da razmotrimo distribuciju simetrično raspoređenih kvadrirane vrijednosti ili čak ranije pominjanu Fisherovu r-u-z transformaciju koja drugačije izgleda? Slika 18.7 prikazuje da se te tri linije razlikuju. Dakle, ovdje ne postoji jedna "objektivna" apriorna distribucija, mada se mora priznati da su razlike u rezultatima koje se dobijaju nakon odabira različitih varijanti obično zanemarive. Uz to, nekad je nemoguće postaviti teorijski neinformativnu distribuciju, kao što je slučaj za hipoteze razlika kada potencijalni efekti sežu od $-\infty$ do $+\infty$.

Suštinski bitnija kritika se odnosi na to da kada koristimo "neinformativne" vjerovatnoće mi zapravo u proces zaključivanja



Slika 18.7: Tri objektivne apriorne distribucije za koeficijente korelacije.

"unosimo informaciju" o apsolutnom neznanju što itekako ima posljedice po konačnu procjenu. Kao prvo, time ignorišemo potencijalno važna saznanja koja procjenu mogu učiniti efikasnijom. Na primjer, ako ispitujemo – po ko zna koji put – povezanost između opšte kognitivne sposobnosti i prosječne ocjene na fakultetu, potpuno je nerazumno očekivati da postoje jednake šanse da korelacije budu -1, -0.50, 0, +.50 ili +1, kao i uopšte da bi $\rho \approx |1.0|$ budući da smo svjesni da postoje različiti faktori koji utiču na ocjene. Nerezonska specifikacija parametara nije neki problem kada imamo ogromne uzorke jer će njihova snaga da prevlada sve apriorne vjerovatnoće. Međutim, kada imamo manje uzorke iz kojih želimo da iscijedimo što više zdravih informacija to predstavlja rasipništvo. Druga suštinska zamjerka je uvid da kada istraživači biraju neinformativne vjerovatnoće time zapravo ne pokazuje svoju neutralnost nego intelektualnu ljenost i neodgovornost, budući da se ne trude da adekvatno istraže i argumentuju svoja prethodna uvjerenja. Treća zamjerka se tiče situacija u kojima se u okviru bejzijanskog pristupa porede vjerovatnoće nulte i alternativne hipoteze – o čemu će riječi biti malo kasnije. Razlijevanje prethodnih vjerovatnoća i na praktično nemoguće vrijednosti ima za direktnu posljedicu da na vještački način nulta hipoteza povećava svoju vjerovatnoću, s obzirom na to da se teorijski nerealne vrijednosti moraju ugraditi u alternativnu hipotezu koja time postaje slabije vjerovatna.

Kako se definišu informativne apriorne vjerovatnoće?

Nasuprot neinformativnim vjerovatnoćama stoje informativne vjerovatnoće. To su distribucije vjerovatnoća koje istraživači

postavljaju na osnovu svojih manje ili više izraženih uvjerenja i koje smo mogli vidjeti u primjeru sa baroknom muzikom u slučajevima *skeptika*, *optimiste* i *pesimiste*. Vidjeli smo da su oni imali potpuno različita uvjerenja i da su u skladu s tim i posteriorne procjene parametara bile potpuno različite kada je uzorak bio mali. I pomisao na tako jak uticaj subjektivnosti na konačne rezultate užasava objektiviste u nauci. Međutim, postoje barem tri argumenta za njihovu upotrebe koje navode pristalice informativnih vjerovatnoća.

Prvi razlog je taj da je u naučnom radu vrlo teško proturiti nerazumne apriorne vjerovatnoće, odnosno da bi recenzenti radova, eksperti u oblasti, morali uočiti iste. Štaviše, informisane vjerovatnoće u radovima zahtijevaju temeljna obrazloženja, te se ona postavljaju na osnovu ekspertskih procjeni. Postoje i formalno definisane procedure kako se do ekspertskih procjena distribucije prethodnih vjerovatnoća dolazi (tzv. elicitacija uvjerenja). U procesu elicitacije se koriste i pomoćni softveri čija je svrha da se što vjernije ekspertsko uvjerenje - koje je često i prosjek procjena više eksperata - izrazi kako matematička distribucija. Internet stranica koja sadrži sjajne edukativne materijale i potreban softver je https://shelf.sites.sheffield.ac.uk.

Drugi argument je sljedeći. Da, može se desiti da eksperti drastično pogriješe ili da istraživači uspiju da namjerno proguraju nerealistične procjene kako bi proveli svoju agendu i objavili rad. Međutim, to će biti rjeđe nego što će rezultati sa nekog nereprezentativnog uzorka dovesti do statistički značajnih rezultata uz pomoć "masaže" podataka. Štaviše, uvođenjem skeptičnih uvjerenja za hipoteze koje se doživljavaju kao slabo uvjerljive[6] zahtijevaju se jači empirijski dokazi nego "objektivističkim" pristupom. Deviza kaže: "Izuzetne tvrdnje traže izuzetne dokaze". Vidjećemo da je moguće izračunati omjer snage dokaza u prilog jedne ili druge hipoteze, te da je i to dodatna informacija koja služi nezavisnim čitaocima da kreiraju svoja uvjerenja o snazi efekata.

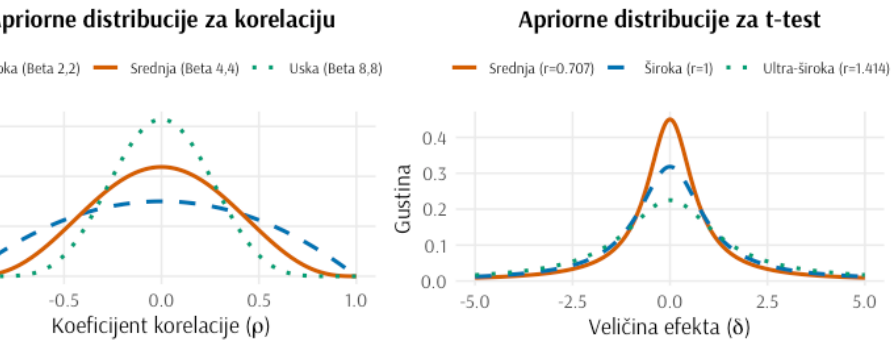
[6] npr. tvrdnje o dokazima za paranormalne psihičke pojave, Wagenmakers et al., 2011, https://doi.org/10.1037/a0022790

Treći argument se tiče filozofskog pogleda na esencijalnu prirodu i svrhu nauke. Kao i objektivisti, bejzijanska škola priznaje da je nauka kao i svaka ljudska djelatnost podložna greškama. Međutim, ona ima i praktičnu stranu, odnosno ne može se temeljiti isključivo na stalnom oprezu i odbacivanju hipoteza, nego se suprotstavljena gledišta moraju uzeti u obzir i nekom od njih se moramo u datom trenutku prikloniti na osnovu trenutno prisutnog znanja. U dalekoj budućnosti je moguće da će se doći do krajnje istine zahvaljujući tome što se prikuplja

sve više empirijskih podataka koji poništavaju efekte uvjerenja pojedinačnih naučnika ili dominantne naučne zajednice (princip autokorektivnosti nauke). U međuvremenu se mora djelovati na osnovu dostupnog znanja i nadograđivati ga, a informativne apriorne vjerovatnoće to omogućavaju. Iz tog razloga je bejzijanska statistika daleko prisutnija i u primijenjenim, inženjerskim područjima nauke.

Postoji li nešto između neinformativnih i informativnih apriornih vjerovatnoća?

Postoji. Dakle, ako dijelite stav onih istraživača koji nisu ubijeđeni u navedene argumente i ako strahujete da bi zloupotreba informisanih vjerovatnoća mogla ugroziti objektivnost nauke, postoji srednji put kojeg omogućavaju slabo informativne prethodne vjerovatnoće. One predstavljaju kompromis između izričite subjektivnosti informativnih i naivne objektivnosti neinformativnih pristupa. Slabo informativne vjerovatnoće su distribucije koje su dovoljno široke da dozvoljavaju podacima da "govore sami za sebe" kada su dokazi jaki, ali istovremeno blago usmjeravaju zaključivanje ka realističnim vrijednostima parametara. Na primjer, umjesto da postavimo uniformnu distribuciju za koeficijent korelacije, možemo koristiti centriranu beta distribuciju koja favorizuje umjerene vrijednosti, ali i dalje dozvoljava ekstremne vrijednosti ako podaci snažno upućuju na njih. U psihološkim istraživanjima, slabo informativne vjerovatnoće su posebno korisne kada znamo da su određeni efekti rijetko izrazito veliki (npr. korelacije rijetko prelaze .50 u istraživanjima ličnosti), ali ne želimo da strogo ograničimo mogućnost njihovog pojavljivanja. Ove distribucije takođe pomažu u stabilizaciji procjena kada su uzorci mali, što je čest problem u psihološkim istraživanjima. Čak i ako smatramo da ne posjedujemo dovoljno znanja o fenomenu koji istražujemo, u velikoj većini slučajeva se mogu razumno isključiti ekstremne vrijednosti parametara, i to je upravo ono što slabo informativne vjerovatnoće omogućavaju. Konačno, slabo informativne distribucije su jedina zamjena za neinformativne vjerovatnoće u slučaju hipoteza razlika. Na Slici 18.8 se vidi primjer često korištenih slabo informativnih distribucija za korelacije i poređenje razlika između dvije grupe.



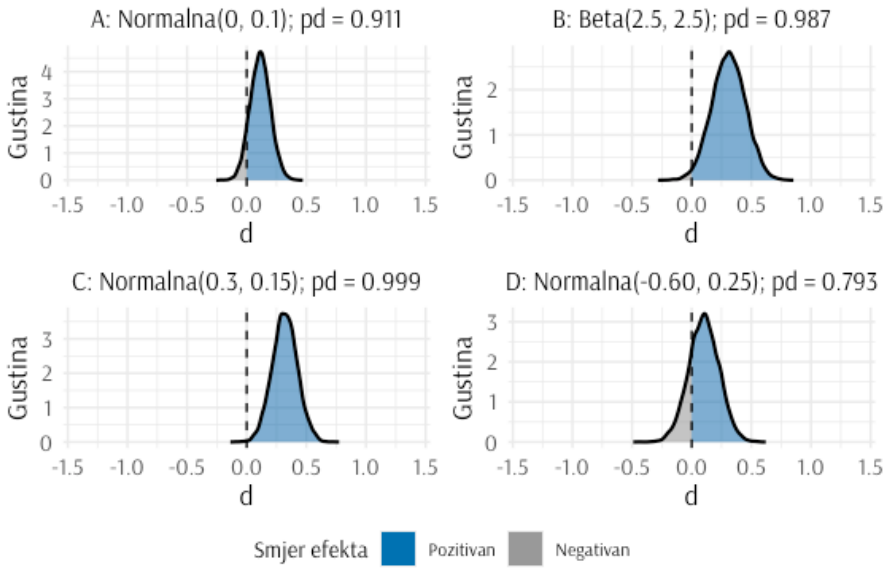
esto korištene slabo informativne distribucije za hipoteze korelacija i razlika aritemtičkih sredina

Kako bejzijanci testiraju hipoteze?

Osim integrisanja prethodnih vjerovatnoća u procjenu parametru, velika prednost bejzijanskog pristupa je i mogućnost da se kvantifikuje vjerovatnoća kompetitivnih modela. Postoji veći broj tehnika koje su analogne frekvencionističkim postupcima testiranja nulte hipoteze, a ovdje navodim one najosnovnije.

Vjerovatnoća smjera efekta

Najosnovnija među najosnovnijim se naziva **vjerovatnoća smjera** (engl. *probability of direction*, *pd*) i predstavlja jednostavni pandan p-vrijednosti. Vjerovatnoća smjera se direktno izvodi iz posteriorne vjerovatnoće i izražava kolika je površina većeg od dva dijela pod krivom od tačke nulte hipoteze (Slika 18.9). Za razliku od p-vrijednosti ona se može tumačiti kao vjerovatnoća postojanja efekta (uzimajući u obzir početne vjerovatnoće i rezultate sa uzorka). Vrijednosti su na skali od .50 do 1.00. Najniža vrijednost ($pd = .50$) se dobija kada je medijana posteriorne distrubucije jednaka vrijednosti nulte hipoteze (npr. $Mdn = 0$), odnosno kada postoji jednaka vjerovatnoća da je efekat pozitivan i da je efekat negativan. Što je posteriorna distribucija udaljenija od nultog efekta vrijednosti se više približavaju vrijednosti 1.00. Vrijednost $pd = .975$ bi tehnički odgovarala dvosmjernoj p-vrijednosti od .05. Potrebno je još napomenuti da kao ni p-vrijednost, ni pd-vrijednost ne govori ništa o veličini efekta.

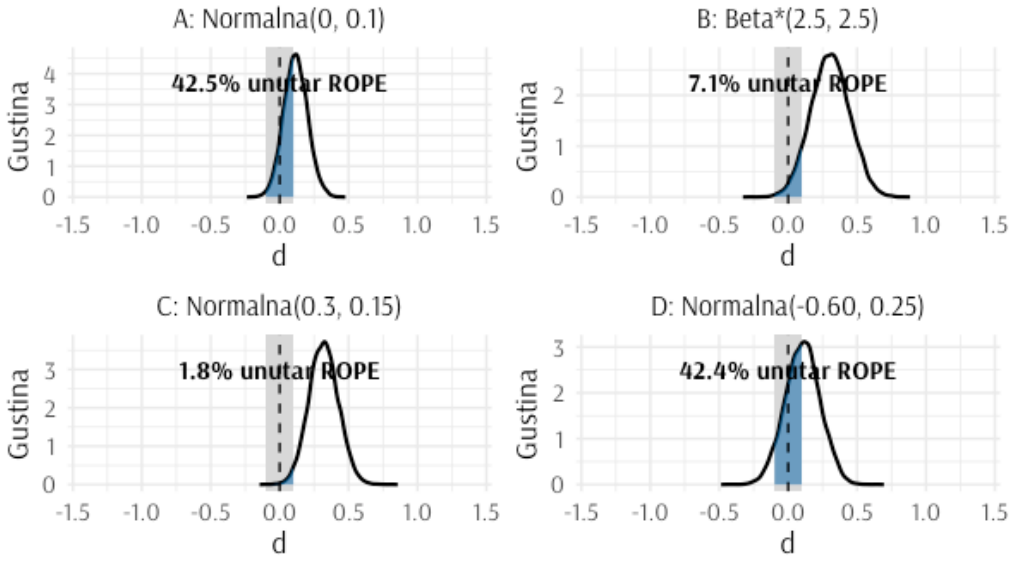


Slika 18.9: Vjerovatnoća smjera efekta prikazana za raniji primjer kada je $n = 200$.

ROPE testiranje

Druga, popularnija tehnika testiranja hipoteza se naziva **ROPE testiranje** (engl. *ROPE testing*) i kombinuje posteriornu distribuciju sa testiranjem intervalne nulte hipoteze. S obzirom da se testira intervalna nulta hipoteza, prvi neophodni korak je da se definiše **region praktične ekvivalentnosti** (engl. *Region of Practical Equivalence*; *ROPE*). ROPE se definiše određivanjem donje i gornje granice za koje važi tvrdnja: "ako je parametar ovoliko nizak (θ_D) ili ovoliko visok (θ_G), to je toliko praktično nebitno da možemo smatrati kao da efekat i ne postoji". Nakon što se dobije posteriorna distribucija saopštava se koliki njen procenat se nalazi unutar regiona praktične ekvivalentnosti. Ako se baš želi izvršiti binarna odluka o prihvatanju ili odbacivanju nulte hipoteze[7] onda se provjerava da li se manje od 2.5% posteriorne distribucije nalazi unutar regiona praktične ekvivalentnosti, odnosno da je 97.5% van njega. Otkud 2.5% i 97.5%? Zato što se ovdje pravi direktna analogija sa dvosmjernim testiranjem hipoteza u frekvencionističkoj statistici. Kao što se tamo koristi nivo značajnosti od 5% koji se dijeli na 2.5% za svaku stranu (rep) distribucije, tako se i ovdje postavlja prag koji bi bio uporediv sa tom konvencijom. Slika 18.10 jasnije pokazuje odluke koje se donose na osnovu ovog ROPE testiranja.

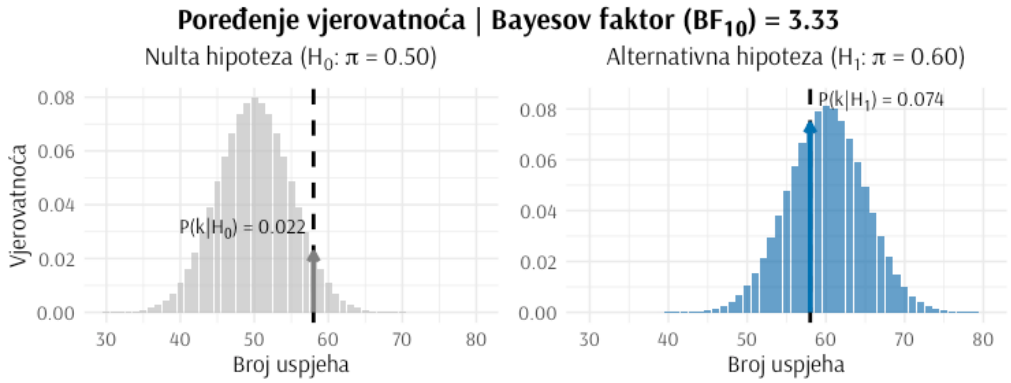
[7] Binarno odlučivanje se baš ne preporučuje: Etz et al., 2024, https://doi.org/10.1037/met0000660



Slika 18.10: ROPE testiranje prikazano za raniji primjer kada je $n = 200$.

Treći, i najpopularniji oblik bejzijanskog testiranja se naziva **Bayesov faktor** (ili **Bejzov faktor** engl. *Bayes factor*). Bayesov faktor je ponder koji pokazuje kojom snagom je potrebno promijeniti uvjerenja nakon što se dokazi iz istraživanja uzmu o obzir. Druga definicija kaže da je to *omjer vjerovatnoća* da se dobiju baš ti podaci na uzorku ako je tačna hipoteza A u odnosu na pretpostavku da je tačna hipoteza B - pod uslovom da su hipoteze A i B bile uzete kao jednako vjerovatne na samom početku. Biće mnogo jasnije ako to vidite na ekstremno jednostavnom primjeru. Vratićemo se istraživanju u kojem je istraživače zanimalo da li na uzrastu od 16 mjeseci, većina djece, odnosno njih više od 50% djece ($H_0 : \pi_+ = .50$) može da kategoriše predmete na osnovu imena kada ti predmeti imaju vizuelnu sličnost (npr. oba su okrugla). Na uzorku od 100 malih ispitanika je njih 58 ($P = .58$) tačno uradilo zadatak odnosno ubacilo predmetić u traženu korpicu. Recimo za trenutak da je istraživač – prije istraživanja – precizirao da je uvjeren da je tačno 60% djece tog doba sposobno da uradi dati zadatak - što predstavlja alternativnu hipotezu ($H_A : \pi_+ = .60$). Postavlja se pitanje da li je dobijeni rezultat vjerovatniji ako pretpostavimo da je alternativna hipoteza tačna ili ako je nulta hipoteza tačna (i da obje hipoteze imaju jednaku početnu vjerovatnoću). S obzirom da znamo da je binomna distribucija odgovorajuća za prikazivanje vjerovatnoća događaja sa binarnim ishodom (jeste ili nije riješen zadatak), možemo da kreiramo pretpostavljene distribucije za ove dvije hipoteze i da prosudimo za koju od njih je veća vjerovatnoća da dobijemo rezultat 0.58.

Bayesov faktor



Slika 18.11: Poređenje vjerovatnoća rezultata sa uzorka za dvije konkurentne hipoteze.

Na Slici 18.11 vidimo da je vjerovatnoća da baš 58 od 100 ispitanika tačno uradi zadatak jednaka 0.0223 ako je nulta hipoteza istinita, te da je ona veća za postavljenu alternativnu hipotezu 0.0742. Ne samo da vidimo da je jedan stubić viši od drugoga, nego možemo da izračunamo i njihov omjer: $0.0742/0.0223 = 3.33$. Drugim riječima, šansa da 58% ispitanika tačno riješi zadatak je 3.33 puta veća ako je u populaciji istinita alternativna hipoteza ($H_A : \pi_+ = .60$) nego ako je u populaciji istinita nulta hipoteza ($H_0 : \pi_+ = .50$). To i nije nešto neočekivano jer je broj 58 bliži 60 nego 50. Međutim, trebalo bi da priznate da je zgodno da imamo jedan broj koji direktno poredi vjerovatnoće. Ali avaj, alternativna hipoteza gotovo nikada nije tačkasta hipoteza. Zar upravo bejzijanski pristup naglašava da alternativna hipoteza, odnosno prostor vjerovatnoće parametra, ima neku distribuciju? Drugim riječima, zašto ne bi bilo moguće da je parametar .59, .61 ili .80?

Kako se izračunava Bayesov faktor u stvarnim situacijama?

Prvo da razjasnimo notaciju. Bayesov faktor se u radovima navodi oznakom **BF** ili rjeđe samo kao **B**. U indeksu se navodi redoslijed modela koji se porede, pa BF_{10} označava omjer alternativna naspram nulte, dok BF_{01} označava omjer nulta naspram alternativne. S obzirom da se radi o omjerima, BF_{01} predstavlja inverznu vrijednost BF_{10}, odnosno $BF_{01} = 1/BF_{10}$

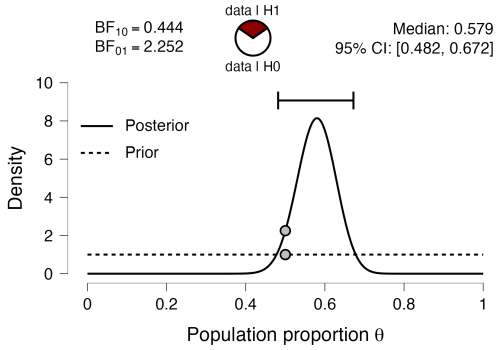
Zamislimo sad da imamo dva istraživača koji imaju dvije alternativne hipoteze u koje podjednako vjeruju. Jedan vjeruje da je parametar 0.60, a drugi da je parametar 0.70. Kako objediniti ovo u jednu alternativnu hipotezu i uporediti je sa nultom hipotezom? Rješenje je jednostavno: uprosječićemo ih. Konkretna kalkulacija je tada jednaka sljedećem:

$$\left(\frac{p(58|\pi = .60)}{p(58|\pi = .50)} + \frac{p(58|\pi = .70)}{p(58|\pi = .50)}\right)/2 = (3.33 + 0.14)/2 = 1.74$$

U ovom slučaju vidimo da kombinovana alternativna hipoteza i dalje nešto bolje objašnjava dobijene podatke nego nulta. Omjer je ublažen jer je 0.70 manje vjerovatno od 0.50 (s obzirom na to da je 58 bliže 50 nego 70). Međutim, i ova nova alternativna hipoteza je očito vještačka. Zašto ne bi i druge vrijednosti parametra bile moguće? Idemo zato odmah na krajnji slučaj gdje su sve vrijednosti duž x-ose jednako vjerovatne, odnosno na uniformnu alternativnu hipotezu. Stvarno uprosječavanja zahtijeva upotrebu integralnog računa budući da između između 0.60 i 0.70 postoji beskonačno fin prostor (npr. da je parametar tačno 0.6342781), ali se možemo poslužiti pojednostavljenim modelom koji izvrsno opisuje princip izračunavanja Bayesovog faktora u tom slučaju. Dakle, zamislite da umjesto dva istraživača imamo njih 101 koji imaju jednaku snagu uvjerenja za vrijednosti parametra koje su: 0.00, 0.01, 0.02…0.99, 1.00. Za svaku od tih pretpostavljenih vrijednosti se izračuna omjer vjerovatnoće u odnosu na nultu hipotezu i onda se uzima njihov prosjek, baš kao i u slučaju da smo imali samo dva istraživača. Rezultirajuća vrijednost je $BF_{10} = 0.444$ i s obzirom da je vrijednost ovog uprosječenog omjera manja od 1, ona nam govori da postoji veća vjerovatnoća da je $P = 0.58$ na uzorku proistekla iz nulte hipoteze nego iz uniformne alternativne hipoteze. Kao što smo ranije naveli, omjere manje od 1 je lakše tumačiti ako se koristi invertovana vrijednost. Dobijena vrijednost $BF_{01} = 1/0.444 = 2.25$ nam sugeriše da je vjerovatnoća da se dobiju dati podaci na uzorku 2.25 puta veća ako je nulta hipoteza tačna nego ako je tačna uniformna alternativna hipoteza.[8] Dobra ilustracija tog odnosa je prikazana na Slici 18.12 gdje možemo da vidimo i posteriornu distribuciju koja pokazuje da je vjerovatnoća da je $\theta = \pi = 0.50$ porasla nakon što se podaci sa uzorka integrišu sa prethodnim vjerovatnoćama. Gustina apriorna vjerovatnoća za tu vrijednost je bila 1 prije istraživanja, a nakon istraživanja ona ima vrijednost 2.25.

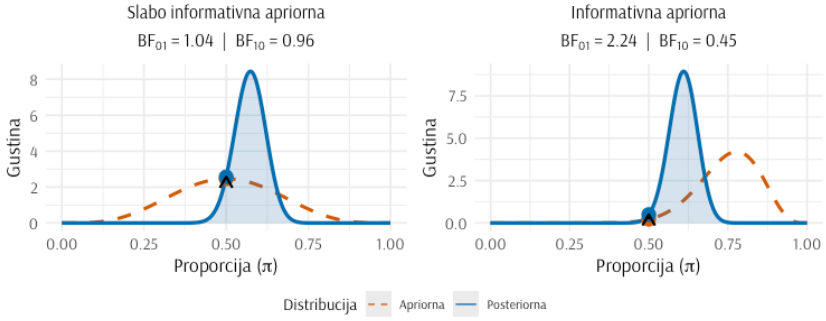
[8] Približnom procjenom sa 101 vrijednošću parametra dobijamo vrijednost 2.27.

Naravno, postavlja se logično pitanje da li je uniformna distribucija zaista bila dobra operacionalizacija alternativne hipoteze. Na Slici 18.13 vidimo vjerovatno realističnije apriorne opcije - slabo informativnu distribuciju centriranu na vrijednosti 0.50 kojom se ekstremne vrijednosti smatraju nevjerovatnim i informativnu distribuciju kojom su predviđane vrijednosti parametra preko 50%, ali ne i veće od 95%. Grafikoni ukazuju na to da u oba



Slika 18.12: Bayesov faktor i posteriorna distribucija prikazani u JASP softveru.

slučaja nulta hipoteza ($\pi = .50$) postaju blago vjerovatnije nego što su to bile prije prikupljanja podataka. Međutim, istovremeno je vidljivo da su neke druge vrijednosti iznad $\pi = .50$ bitno vjerovatnije, kao i to da apriorne distribucije nisu mnogo uticale na posteriornu distribuciju koje su gotovo jednake. Sve u svemu, i ovdje uviđamo da je grafički prikaz sadržajniji nego oslanjanje na jednu statističku mjeru - u ovom slučaju Bayesov faktor.



Slika 18.13: Bayesov faktor i posteriorna distribucija za slabo informativnu i informativnu hipotezu.

Ali kako da interpretiramo vrijednosti Bayesovog faktora?

Kada je u pitanju tumačenje intenziteta datog omjera, potrebno je najprije da se podsjetimo da omjeri predstavljaju asimetričnu, nelinearnu skalu koja je zbunjujuća za većinu čitaoca (koji su ionako zbunjeni bejzijanskom statistikom). Neutralna vrijednost za omjere je vrijednost 1, koja nam u slučaju Bayesovog faktora saopštava da je rezultat na uzorku sa jednakom šansom mogao proisteći iz obje hipoteze. To znači da je $BF_{10} = 1$ isto što i $BF_{01} = 1$. Međutim, ako je vrijednost $BF_{10} = 10$ onda je $BF_{01} = 0.10$, a ako je vrijednost $BF_{01} = 2$ onda je $BF_{10} = 0.50$. Iz tog razloga,

kad god je moguće biramo da saopštimo onu verziju čija je vrijednost veća od 1. To i ovdje napominjem jer prijedlog kategorizacije kojom se brojčane vrijednosti prevode u jezičke kvantifikatore na osnovu graničnih skorova (Slika 18.14) podrazumijeva saopštavanje tako izraženih vrijednosti.

Korisno je pomenuti da vrijednosti $BF = 1-3$ često odgovaraju $p = .01-.05$ vrijednostima[9], te se zato povremeno eksperti zagovaraju više vrijednosti - npr. $BF = 6$ - kako bi se rezultati kolikotoliko uzeli zaozbiljno[10]. Iako izgledaju smisleno – baš kao i za različite Cohenove klasifikacije – ovakve preporuke o graničnim vrijednostima je neophodno razmatrati unutar specifičnog konteksta istraživanja i primjene rezultata. Zato jedan čuveni autor mudro preporučuje: "Postavi sebi takvu granicu BF nakon koje bi bio dovoljno uvjeren da ono što hipoteza implicira upotrijebiš sam u praksi ili da impliciranoj intervenciji podvrgneš bliske osobe (npr. partnera, dijete, roditelja)".[11] Štaviše, u okviru bejzijanskog pristupa postoje i formalizovani načini da se u formule uvrste i razmatranja koristi i moguće štete koje donose odluke, ali je to van dometa ovog teksta.

BF_{10}	Snaga dokaza u prilog H_1
> 100	Ekstremni dokazi
30 - 100	Veoma snažni dokazi
10 - 30	Snažni dokazi
3 - 10	Umjereni dokazi
1 - 3	Anegdotalni dokazi
1	Nema dokaza
1/3 - 1	Anegdotalni dokazi (za H_0)
1/10 - 1/3	Umjereni dokazi (za H_0)
1/30 - 1/10	Snažni dokazi (za H_0)
1/100 - 1/30	Veoma snažni dokazi (za H_0)
< 1/100	Ekstremni dokazi (za H_0)

Slika 18.14: Interpretativna klasifikacija visine Bayesovog faktora prezueta iz Wagenmakers et al., 2018, https://doi.org/10.3758/s13423-017-1323-7.

[9] Wetzels et al., 2011, https://doi.org/10.1177/1745691611406923

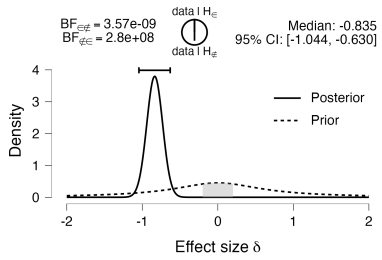
[10] Schönbrodt & Wagenmakers, 2018, https://doi.org/10.3758/s13423-017-1230-y

[11] Lakić, 2019, https://www.psihologijanis.rs/arhiva-godisnjak/godisnjak-2019/Godisnjak-Vol-16-%282019%29-39-58.pdf

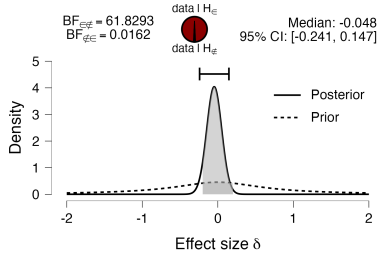
Možemo li odrediti Bayesov faktor i za intervalne hipoteze?

Da, i to je moguće. Nakon što se definiše region praktične ekvivalentnosti i odredi apriorna distribucija za alternativnu hipotezu, upoređuje se gustina vjerovatnoće unutar intervala ekvivalentnosti pod posteriornom distribucijom i apriornom distribucijom kako unutar regiona ekvivalentnosti, tako i van njega. Komplikovano? Hajde da na primjeru vidimo tri tipična ishoda.

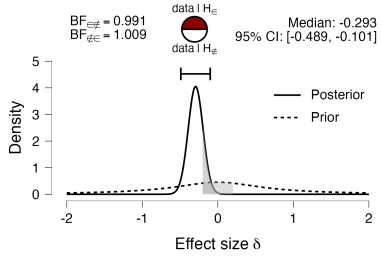
Vidjećemo rezultate provjera intervalnih hipoteza o polnim razlikama za tri crte ličnosti iz HEXACO modela: Emocionalnost, Otvorenost za iskustva i Ekstraverzija. Uzorak je sačinjavalo 492 studenata iz Bosne i Hercegovine i Srbije, a za samoprocjenu crta ličnosti je korišten The Brief HEXACO Inventory (BHI) (i konačno se radi o stvarnim podacima iz jednog našeg neobjavljenog istraživanja!). Za region praktične ekvivalentnosti je postavljen interval od -0.2δ do $+0.2\delta$, odnosno ako se parametar nalazi u datom intervalu smatra se praktično zanemarivim. Apriorna distribucija vjerovatnoća parametra – koju je neophodno postaviti – je u sva tri slučaja definisana kao slabo informativna Cauchy distribucija sa širinom obuhvata 0.707 što je zadana vrijednost u softveru JASP kada se dvije grupe upoređuju u pogledu izraženosti dimenzione varijable.



Slika 18.15: BF za intervalnu hipotezu za crtu Emocionalnost.



Slika 18.16: BF za intervalnu hipotezu za crtu Otvorenost.



Slika 18.17: BF za intervalnu hipotezu za crtu Ekstraverzija.

Za Emocionalnost (Slika 18.15) je na uzorku dobijen rezultat $d = -0.84$, što znači da su – u prosjeku gledano – djevojke sebe procjenjivale dosta emocionalnijim bićima (vau, kakvo otkriće!). Na grafikonu možemo vidjeti da je zasivljeni prostor pod apriornom krivom praktično nepostojeći pod posteriornom distribucijom. Posteriorni 95%-tni interval uvjerenja se prostire od -1.04 do -0.63, pa zato ne čudi zato da Bayesov faktor sugeriše ogromnu podršku podataka alternativnoj hipotezu kojom se tvrdi da parametar NE pripada regionu od -0.20 do $+0.20$ u poređenju sa nultom hipotezom kojom se pretpostavlja da parametar jeste u tom intervalu: $BF_{\notin\in} = 2.8 \times 10^8$.

Za Otvorenost (Slika 18.16) je na uzorku dobijen rezultat vrlo blizak nuli $d = -0.05$. Ako uporedimo dvije sive regije, vidimo da je ona pod posteriornom distribucijom zahvata vidno veći prostor nego ona pod apriornom (5.2 puta), što nam već ukazuje na to da podaci govore u prilog nulte hipoteze. Međutim, da bi se dobio odgovarajući Bayesov faktor $BF_{\in\notin}$ moraju se istovremeno uzeti u obzir i oblasti van regiona ekvivalentnosti (čitaj: dio koji nije siv, apsolutne vrijednosti veće od 0.20). Taj dio zahvata 11.6 puta manje prostora pod posteriornom nego pod apriornom distribucijom, što ne iznenađuje budući da se pod posteriornom distribucijom 95%-tni interval uvjerenja već završava na -0.24. Kada se napravi omjer ovih promjena (apriorna $\rightarrow$ posteriorna) pokazuje se da su podaci izuzetno pojačali uvjerenje o istinosti nulte intervalne hipoteze, konkretno: $BF_{\in\notin} = \frac{5.2}{1/11} = \frac{5.2}{0.086} = 61.8$ puta.

Konačno, studentkinje su u prosjeku sebe procjenjivale ekstravertnijim nego muške kolege: $d = -0.29$. Ali grafikon (Slika 18.17) nam pokazuje da je siva površina pod apriornom distribucijom otprilike jednaka kao ona pod posteriornom. Shodno tome je i distribucija uvjerenja van regiona ekvivalentnosti podjednako raspoređena prije i nakon što su se podaci pojavili (u procjenjivanju vam može pomoći da zakrenete glavu pod 90 stepeni). Dobijeni Bayesov faktor $BF_{\notin\in} = 1.009$ nam upravo govori da nije došlo do promjene uvjerenja, odnosno da je rezultat na uzorku imao jednaku vjerovatnoću da se desi ako se parametar nalazi unutar regiona ekvivalentnosti i ako se parametar nalazi izvan tog regiona - naravno, uzimajući u obzir prethodna uvjerenja.

Sve u svemu, vidjeli smo situaciju u kojoj BF jasno govori u prilog alternativne i nulte hipoteze, kao i situaciju u kojoj nam otkriva da nam podaci nisu baš nimalo pomogli u rasvjetljavanju postojanja razlika.

Koje su prednosti bejzijanskog pristupa statističkom zaključivanju?

Već sam predočio bitne epistemološke argumente za uvođenje ekspertskih uvjerenja u proces statističkog zaključivanja. Naime, ranije prikupljena znanja nisu za bacanje, jer nauka mora biti kumulativni poduhvat. Na primjer, vrlo razumno bi bilo koristiti rezultate meta-analiza za definisanje apriornih uvjerenja o distribuciji parametara u novim istraživanjama. Također, suštinski je smisleno kontinuirano ažurirati uvjerenja, odnosno ne oslanjati se na stara znanja. Naime, i mnogi psihološki fenomeni, pogotovi oni u oblasti socijalne, ali i drugih oblasti psihologije, su podložni promjenama. Današnja prevalencija određenih društvenih stavova, ali i poremećaja mentalnog zdravlja, ne može biti ista kao u pred-digitalno doba u kojem čovječanstvo nije bilo globalno uvezano internetom i gdje društvene mreže nisu činile najčešći komunikacioni medij. Prednost bejzijanskih pristupa jeste to što priznaje fluidnost parametara, ali i našu ograničenost da ih u potpunosti saznamo.

Drugo, bejzijanske vjerovatnoće i intervali vjerovatnoće imaju direktno značenje za istraživača: kada navedemo da postoji 95% vjerojatnoće da se parametar nalazi unutar određenih granica, upravo to i mislimo (sa nekim malim ogradama – u smislu da smo vjerovali i u apriorne vjerovatnoće). Takva interpretativna transparentnost nedostaje klasičnim intervalima pouzdanosti koji se odnose na hipotetički beskonačan broj replikacija i koji se najčešće pogrešno tumače. Intuitivna interpretacija intervala uvjerenja je posebno korisna u istraživanjima gdje je potrebno iskomunicirati neizvjesnost nalaza širokoj publici, a neizvjesnost uvijek postoji.

Konačno, posebnu prednost pružaju metode direktnog poređenja statističkih modela, među kojima je najkoherentniji i najelegantniji Bayesov faktor. Za razliku od p-vrijednosti, Bayesov faktor ne zahtijeva arbitrarne pragove odlučivanja i omogućava punu kvantitativnu procjenu relativne vjerodostojnosti konkurentskih hipoteza. Posebno je korisno što se bejzijanskim pristupom omogućava principijelno prihvatanje nulte hipoteze - nešto što frekvencionistička statistika ne radi adekvatno. U praksi ovo znači da možemo razlikovati odsustvo dokaza efekta od dokaza o odsustvu efekta, što je suštinski bitno za teorijsko napredovanje u psihologiji i za praktične primjene gdje je potrebno demonstrirati da određena intervencija nema štetne efekte. Dostupnost otvorenog softvera – kao što je *JASP* koji je bukvalno stvoren kako bi popularizovao Bayesov faktor – dodatno spušta tehničku barijeru i podstiče transparentno dijeljenje kompletnih fajlova

sa sačuvanim postavkama, čime se olakšava replikacija nalaza i provjera analiza. Ukupno gledano, bejzijanski pristup nudi zdravorazumski koherentan i metodološki prilagodljiv okvir koji odgovara savremenim potrebama psihologije kao empirijske i kumulativne nauke.

A postoje li neke mane bejzijanskog pristupa?

Itekako. Iako je bejzijanski okvir izuzetno moćan, kao i sa svim moćnim alatima, njegova primjena podrazumijeva znatno viši nivo metodološke sofisticiranosti nego što ga zahtijevaju standardne frekvencionističke procedure. U "full-on" modu – kojeg nisam ovdje predstavljao – istraživač mora eksplicitno definisati strukturu cijelog modela, uključujući oblik i parametre svih relevantnih distribucija (npr. distribuciju parametra standardne devijacije).[12] Čitav taj posao predstavlja poseban izazov za manje iskusne istraživače koji nisu sigurni kako odabrati odgovarajuće apriorne vjerovatnoće - da li koristiti neinformativne, zadane slabo informativne ili ako se ipak odluče da se unesu informativne, kako do njih doći. Uz to, ogromna raznovrsnost dostupnih distribucija (pomenuli smo samo njih par, kao što su Beta i Cauchy, koji već zvuče ezoterično većini psihologa) i načina njihovog kombinovanja zahtijeva duboko statističko znanje koje prelazi tipičnu obuku psihologa-istraživača.

[12] Preporučio bih sljedeće udžbenike posvećene bejzijanskoj analizi: 1. Wagenmakers & Metzke, 2024, https://www.bayesianspectacles.org/free-course-book/; 2. Kruschke, 2015, ISBN: 978-0-12-405888-0, dostupno na https://shop.elsevier.com/books/doing-bayesian-data-analysis/kruschke/978-0-12-405888-0; 3. McElreath, 2020, https://doi.org/10.1201/9780429029608

Drugo, uz sadržinsku ekspertizu, ozbiljnije bejzijanske analize zahtijevaju visoku tehničku opremljenost. Analiza kompleksnijih nacrta obično zahtijeva poseban oblik simulacije podataka za procjenu posteriornih vjerovatnoća (tzv. Markovljevi Monte Karlo lanci - što već zvuči jezivo), što može biti izuzetno vremenski zahtjevno i računarski intenzivno. Za razliku od brzih frekventističkih procedura koji se i za kompleksne nacrte izvode za nekoliko sekundi, bejzijanska analiza može zahtijevati sate, pa i dane kalkulacija na moćnim procesorima. Srećom, za većinu normalnih analiza koje se obavljaju u softverima kao što je *JASP* to nije slučaj i rezultati se dobijaju unutar par sekundi.

Na kraju, postoje i značajne prepreke u pogledu akademske prijemčivosti za bejzijanski pristup. Iako on dobija na popularnosti, većina udžbenika psihološke statistike u regionu se i dalje mahom fokusira na frekventističke metode (dakle, osjećajte se privilegovano), ostavljajući tako istraživače bez adekvatne obuke. Zbog toga još uvijek postoji značajno nepovjerenje ili čak potpuno neznanje o bejzijanskom pristupu među većinom recenzenata i urednika časopisa, a time je i broj objavljenih istraživanja

relativno mali. Standardizacija izvještavanja bejzijanskih analiza se tek razvija, a različite prakse među istraživačima mogu otežavati reproduktibilnost rezultata, pa se zato mnogi još drže tradicionalnih p-vrijednosti kao zlatnog standarda statističkog zaključivanja.

Sve u svemu, bejzijanski pristup zaključivanju će biti sve više prisutan, budući da se njegova popularnost širi i da se sve više razvijaju bejzijanske alternative klasičnim procedurama. Činjenica da je u *JASP*-u možete da izvedete bejzijansku meta-analizu dosta toga govori. Međutim, nakon što smo prošli različite strateške pristupe postupcima zaključivanja, došao je trenutak da objasnimo neke osnovne statističke principe za nacrte koji zahtijevaju više od dvije varijable. Kao što ćete vidjeti, za njih je moguće koristiti i frekvencionističke i bejzijanske pristupe, ali razvićete i znanja o nekim bitnim pravilnostima neophodnim da se snađete u kompleksnijim modelima koji bolje predstavljaju stvarnost. Sljedeće poglavlje, i posljednje u ovoj knjizi, posvećeno je kraljici statističkih tehnika: regresionoj analizi.

Poglavlje 19

Uvod u opšti linearni model: Linearna regresiona analiza

Nakon čitanja ovog poglavlja moći ćete odgovoriti na sljedeća pitanja:

- Kako se koncept linearne regresije nadovezuje na koeficijent korelacije i šta predstavljaju nestandardizovani i standardizovani regresioni koeficijenti u kontekstu višestruke regresione analize?
- Na koji način se vrši poređenje kvaliteta različitih regresionih modela?
- Šta je svrha statističke kontrole i kako se tehnički izvodi parcijalizacija?
- Koje su ključne pretpostavke opšteg linearnog modela i kako njihovo kršenje utiče na ispravnost interpretacije rezultata?

Ovaj udžbenik je specifičan i po tome što će u njemu detaljno biti predstavljena samo jedna statistička tehnika koja istovremeno tretira više od dvije varijable. Za tu svrhu sam odabrao regresionu analizu jer ona u sebi sadrži čitav svijet statističkih tehnika. Principi koji važe za regresione analize važe i za mnoge druge striktno statističke modele bez obzira na to da li se radi o latentnim ili manifestnim varijablama, kategoričkim ili dimenzionim ili njihovim kombinacijama, bivarijatnim ili multivarijatnim analizama. Zato će poznavanje regresionih analiza biti dovoljno da značajno proširite svoj analitički repertoar u odnosu na dosad prikazane bivarijatne tehnike, ali što je još važnije kroz njih ćete shvatiti veliki broj novih koncepata koje bi svaki kvantitativno orijentisani istraživač morao znati. Kako bismo lakše došli do tog cilja počećemo sa nečim što smo već obrađivali, a to je linearni koeficijent korelacije.

Šta ono bi linearni koeficijent korelacije?

Linearni koeficijent korelacije predstavlja broj koji nam ukazuje na smjer i snagu linearne povezanosti dvije dimenzione varijable. U tom kontekstu riječ **linearno** označava vrlo jednostavan model - prostu pravu liniju koja bi trebalo da opiše postojeću statističku povezanost između dvije varijable. Datim modelom se pretpostavlja da se sa sve većim vrijednostima duž vrijednosti x-varijabli *opaža* i ravnomjerna promjena vrijednosti y-varijabli.

Formulaciju "opaža se promjena" sam pažljivo odabrao. Naime, znate već da korelacija ne znači kauzaciju, pa je bolje da se ne zapletemo u jezičke zamke kojima bismo implicirali neke više odnose (npr. već granično bi bilo i "sa promjenom x-varijable dolazi i do promjene y-varijable", a kamoli "sa promjenom x-varijable možemo predvidjeti i promjenu y-varijable" ili najgore od svega "mijenjanjem x-varijable doći će do promjene y-varijable"). Riječ **ko-relacija** znači zajednički odnos ("*sa-odnos*") i ne pretenduje da ukaže na neko predviđanje ili kauzaciju.

Konačno, tu je i riječ koeficijent koju smo već sretali. Riječ **koeficijent** označava konstantni množilac kojim se množi neka varijabla da bi se dobila neka druga, transformisana varijabla. Ako je r zaista neki pravi koeficijent onda bi moralo da važi da za neku varijablu y postoji x gdje je: $y = r * x$. Budući da znamo da je veza između varijabli statistička (tj. probabilistička), a ne deterministička – odnosno, budući da znamo da model ne pogađa savršeno vrijednosti y-varijable – odgovarajuća konceptualna formula mora biti: $\hat{y} = r * x$ gdje je $\hat{y}$ modelom procijenjena vrijednost. Kada su x i y standardizovane varijable tada zaista važi formula $z_{\hat{y}} = r * z_x$. Na ovom mjestu bi bilo lijepo i da se podsjetite i formule za linearni koeficijent korelacije koji koristi z-vrijednosti i koji pokazuje da je r standardizovana mjera:

Osnovna konceptualna formula

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

Kako se izračunava linearni koeficijent korelacije?

Dosad sam vas u velikoj mjeri poštedio računanja konkretnih vrijednosti, ali se sada zaista moramo prihvatiti toga kako biste shvatili suštinu postupka. Bez brige, koristićemo megajednostavan primjer računanja za pet skorova na testu inteligencije i prosječne ocjene u srednjoj školi (inače, u Bosni i Hercegovini se ocjene u osnovnoj i srednjoj školi daju na skali od 1=nedovoljan do

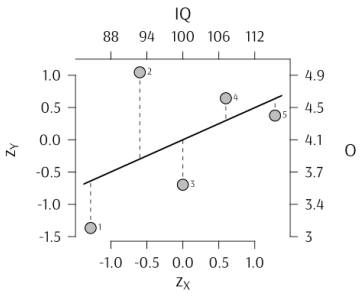
5=odličan). U ovom slučaju, istovremeno fiktivnom i po veličini korelacije teorijski mogućem, je $r = .50$ (tačnije, $r = .498$). Podaci iz tabele (Slika 19.1) nam pokazaju kako se pješice dolazi do te vrijednosti.

id	x	y	z_x	z_y	$z_x * z_y$	r	$z_{\hat{y}}$	$z_y - z_{\hat{y}}$	$(z_y - z_{\hat{y}})^2$	z_y^2
1	85	3.1	-1.43	-1.53	2.19	0.50	-0.71	-0.82	0.67	2.34
2	93	4.9	-0.67	1.17	-0.78	0.50	-0.33	1.50	2.26	1.37
3	100	3.6	0.00	-0.78	0.00	0.50	0.00	-0.78	0.61	0.61
4	107	4.6	0.67	0.72	0.48	0.50	0.33	0.39	0.15	0.52
5	115	4.4	1.43	0.42	0.60	0.50	0.71	-0.29	0.09	0.18
Σ	**500**	**20.6**	**0**	**0**	**2.49**	**-**	**0**	**0**	**3.77**	**5.01**

Slika 19.1: Izračunavanje linearnog koeficijenta korelacije i sume kvadrata odstupanja (Mx = 100, SDx = 11.7; My = 4.12, SDy = 0.75).

U prvoj koloni (id) se nalaze redni brojevi ispitanika, nakon čega slijede IQ skorovi (x) i prosječne školske ocjene (y). Sljedeće dvije kolone prikazuju standardizovane vrijednosti za IQ (z_x) i standardizovane prosječne ocjene (z_y). Shodno formuli za izračunavanje korelacija, u šestoj koloni se nalazi proizvod standardizovanih vrijednosti za te dvije varijable ($z_x * z_y$). Da bi se dobio koeficijent linearne korelacije, suma tih proizvoda (2.49) se dijeli sa brojem ispitanika ($n = 5$), pa nam tako r vrijednost predstavlja prosječan proizvod standardizovanih vrijednosti: $r = 2.49/5 = 0.498 \approx 0.50$.

Kolona $\hat{z}_y$ nam prikazuje y-vrijednosti koje se očekuju na osnovu linearnog modela. Pogled na Sliku 19.2 će nam otkriti zašto se očekuju baš te vrijednosti koje su prikazane u tabeli.



Slika 19.2: Linearna regresija sa označenim rezidualnim vrijednostima.

Naime, radi se o y-vrijednostima koje se nalaze tačno na liniji modela za ispitanike koji su dobili određeni IQ skor. Na primjer, za prosječan IQ skor 100 (tj. $z_x = 0$) bismo prema linearnom modelu očekivali da osoba ima i prosječnu ocjenu 4.12 (tj. $\hat{z}_y = 0$). Međutim, vidimo da je ispitanik #3 imao znatno nižu ocjenu $y = 3.6$, što pretvoreno u standardizovane skorove iznosi $z_y = -0.78$ (negativna standardizovana vrijednost označava da se radi o ispodprosječnom skoru). Odstupanje stvarnog y skora od modelom predviđenog skora ($\hat{y}$) predstavlja rezidual, odnosno grešku modela što je označeno isprekidanim linijama na grafikonu.

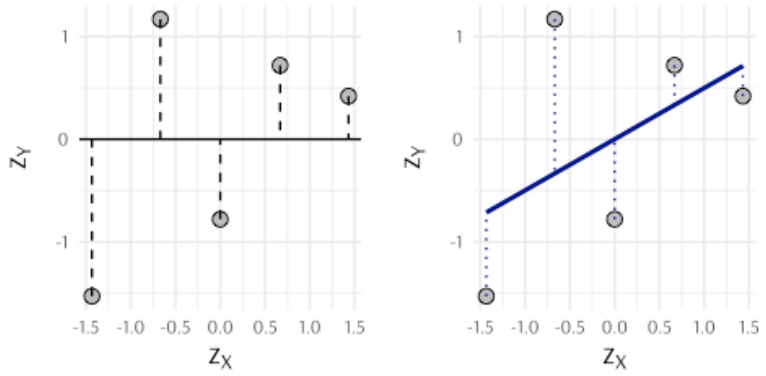
Inače, linearni model jeste jedna vrsta promjenjive aritmetičke sredine koja se mijenja u zavisnosti od vrijednosti x-varijable. Naziva se nekad i uslovnom aritmetičkom sredinom, a kao i najobičnija aritmetička sredina zbir reziduala mora biti jednak 0. To je i dokazano kolonom $z_y - \hat{z}_y$. Kao i u slučaju izračunavanja

standardne devijacije, da bi zbir odstupanja imao neku smislenu vrijednost reziduali se zato moraju kvadrirati.[1]

[1] Umjesto kvadriranja je moguće koristiti apsolutne vrijednosti, ali je to kroz istoriju statistike bila tehnički zahtjevnija opcija koja se tek u posljednje vrijeme popularizuje.

Čemu služi zbir kvadrata odstupanja?

Ovaj **zbir / suma kvadrata odstupanja** (engl. *sum of squares*, SS) operacionalizuje ukupnu grešku procjene linearnog modela. Ona ovdje iznosi $SS_{LM} = 3.77$, ali taj broj nam sam po sebi ništa ne govori. Da bismo utvrdili da li je i koliko linearni model koristan u smanjivanju greške procjene, moramo ga uporediti sa početnim stanjem, odnosno početnim ili nultim modelom. Taj početni model je model koji predstavlja ukupnu raznolikost y-varijable, a operacionalizovan je sumom kvadrata odstupanja od prosječne vrijednosti SS_T (T - od engl. *total*). Podsjećam vas da za aritmetičku sredinu (u ovom kontekstu se označava sa $\bar{y}$) važi da je to mjera za koju zbir kvadrata odstupanja ima najmanju moguću vrijednost za sve moguće mjere na y skali: $SS_T = \sum(y_i - \bar{y})^2 \rightarrow min$. Kako smo vidjeli, linearni model je uslovna aritmetička sredina, pa i za njega važi da predstavlja pravu liniju za koju se dobija najmanji zbir kvadrata odstupanja od svih mogućih linija koje se mogu povući u prostoru kojeg definišu x i y varijable: $SS_{LM} = \sum(y_i - \hat{y})^2 \rightarrow min$.



Slika 19.3: Odstupanja pojedinačnih podataka od aritmetičke sredine i linearnog modela.

Na Slici 19.3 uočavamo da su apsolutne vrijednosti reziduala smanjene ako koristimo linearni model. Za ispitanika #3 odstupanje je isto u oba slučaja, odstupanje je povećano za ispitanika #2, ali je za tri ostala ispitanika ono smanjeno. Dakle, stvari naoko izgledaju bolje, ali je još preciznije da numerički uporedimo dva modela. Jedan način da to uradimo jeste da izračunamo proporciju odnosa između dva zbira kvadrata odstupanja:

$SS_{LM}/SS_T = 3.77/5.01 = 0.75$. Drugim riječima, greška linearnog modela predstavlja 75.0% ili tri četvrtine ukupne raznolikosti podataka. Obrnuto rečeno – a ovo je gledište vrlo bitno – to znači da smo linearnim modelom smanjili početnu grešku procjene za jednu četvrtinu. "Ček, ček, ovo mi se čini poznato?!" Da, kvadriranjem koeficijenta korelacije bismo također dobili: $r^2 = 0.50 * 0.50 = 0.25$ (sve je ovo bilo zaokruživano na dvije decimale, inače je tačan rezultat $r^2 = 0.248$). To je već pominjani **koeficijent determinacije**, dok je koeficijent indeterminacije – koji označava proporciju "neobjašnjene" varijanse – jednak: $1 - r^2 = 1 - 0.25 = 0.75$.

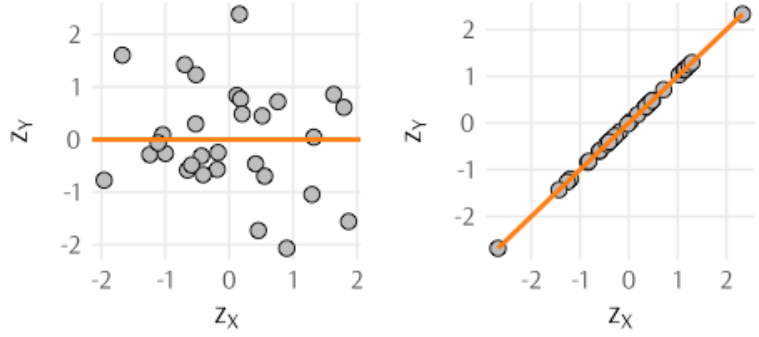
Koeficijent determinacije

Navedenom kalkulacijom smo došli do preciznog odgovora na pitanje koliko nam neki statistički model smanjuje neizvjesnost procjene. Sjećate li se da smo na samom početku knjige naveli da je preciznost jedna od ključnih prednosti korištenja statistike u nauci i psihologiji uopšte? Umjesto da kažemo "čini mi se da ovaj model objašnjava varijabilitet prosječne ocjene bolje nego slučajnost" možemo navesti egzaktnu brojku koja govori o tome koliko model pomaže u objašnjavanju varijabilnosti prosječne ocjene.

Bio bi red i da utvrdimo u kojim slučajevima dobijamo krajnje vrijednosti na skali r^2. Ako reziduala uopšte ne bi bilo, odnosno ako bi na grafikonu raspršenja sve tačke koje predstavljaju ispitanike ležale tačno na liniji modela, tada bi proporcija zbira kvadriranih odstupanja iznosila 0. Slijedilo bi da je $r^2 = 1 - \frac{SS_{LM}}{SS_T} = 1 - 0 = 1$, odnosno linearni model bi mogao da "objasni" svu varijabilnost y-varijable. S druge strane, ako je $SS_{LM} = SS_T$, tada bi proporcija odnosa zbira kvadriranih odstupanja bila 1, što bi značilo da je $r^2 = 1 - \frac{SS_T}{SS_T} = 1 - 1 = 0$. Takav model bi bio predstavljen linijom koja ja paralelna x-osi i siječe y-osu na prosječnoj vrijednosti y-varijable. Slika 19.4 prikazuje ova dva ekstremna slučaja.

Koje mjere koristimo za procjenu smanjenja neizvjesnosti?

Gore navedeno nam pruža krupniji nauk o ulozi eksplicitnih statističkih modela u statistici značajnosti. Funkcija eksplicitnih statističkih modela jeste da na principijelan način procijenimo smanjivanje neizvjesnosti (engl. *uncertainty*). Neizvjesnost je uvijek prisutna u svijetu jer je budućnost ispunjena slučajnim, probabilističkim događajima koji se ne mogu u potpunosti predvidjeti. Količinu smanjenja neizvjesnosti možemo izraziti, a tu informaciju komuniciramo različitim statističkim mjerama. Sad ćemo utvrditi koje su to mjere.



Slika 19.4: Linearni modeli koji prikazuju 0% i 100% "objašnjenosti" y-varijable.

U gore navedenom primjeru smo najprije koristili **zbir kvadrata odstupanja**. Zbog toga se linearna regresiona analiza u stručnim tekstovima često naziva i **uobičajena regresija najmanjih kvadrata odstupanja** (engl. *ordinary least squares regression*, OLS). Druge dvije mjere koje smo dosta ranije upoznali i koje se lakše tumače jer se izlažu na standardizovanoj skali od 0 do 1 su **koeficijent indeterminacije**, te iz njega izveden **koeficijent determinacije**. Uz njih postoje još neke mjere koje ćete sretati u statističkom softveru i u naučnim radovima, a koje su zasnovane na istom principu kvadriranih odstupanja od aritmetičke sredine. Mjera analogna varijansi koja se izračunava na ishodišnoj varijabli je **prosječni kvadrat odstupanja** (engl. *mean squared error*, *MSE*) koji predstavlja varijansu reziduala. Tu je i **korijenovani prosječni kvadrat odstupanja** (engl. *root mean square error*, *RMSE*) koji predstavlja standardnu devijaciju reziduala tj. standardnu grešku procjene i direktno je uporediv sa početnom standardnom devijacijom. Ako imamo početnu $SD = 15$ i dobijemo $RMSE = 15$ znamo da model ništa nije napravio, ali ako je $RMSE = 10$ ili još bolje $RMSE = 5$ znamo da model "radi".

Sve nabrojane **mjere saglasnosti / poklapanja opaženih podataka i pretpostavljenog modela** (engl. *goodness-of-fit indices*) operacionalizuju smanjenje greške procjene na uzorku. Budući da od uzorka do uzorka situacija može značajno da varira, a da statistikom zaključivanja želimo nalaze generalizovati na populaciju, prikazane mjere saglasnosti ne možemo uzimati zdravo-za-gotovo nego ih obično kombinujemo sa još nekim sastojcima kako bi se dobila razumna ocjena kvaliteta modela. Kako su to već drugi autori pametno rekli: "za procjenu statističkih modela bi prvi prioritet morao biti uopštivost modela na buduće

/ još neopažene podatke, a ne dobra saglasnost sa opaženim podacima."[2] Drugim riječima, iako su korisni vidjećemo da r^2 i družina nisu nešto na šta bismo se smjeli potpuno oslanjati. Zgodno je još jednom naglasiti: sve gore navedene mjere pripadaju deskriptivnoj statistici, budući da su rezultati ograničeni na samo jedan uzorak.

[2] npr. Hitchcock & Sober, 2004, https://doi.org/10.1093/bjps/55.1.1; Pitt & Myung, 2002, https://doi.org/10.1037/0033-295X.109.3.472

Moramo li u analizama koristiti standardizovane podatke?

Pri saopštavanju rezultata istraživanja je nekad smislenije koristiti podatke na originalnim mjernim skalama kao što su IQ i skala školskih ocjena. Već se na Slici 19.2 vidi da to nije problem jer se radi o prostoj transformaciji koordinatnog sistem, pa model isto izgleda sem što se na x i y osi koriste originalne mjerne skale. Iako ova linearna transformacija ne dovodi do suštinskih promjena odnosa, ona ipak mijenja matematički izraz regresione funkcije. Razlog za to je što centralno mjesto funkcije više neće biti tačka standardizovane aritmetičke sredine za obje varijable $(0, 0)$. Na našem primjeru, centralnu vrijednost unutar prostora koordinatnog sistema će sada imati prosječne vrijednosti na originalnim skalama $(100, 4.12)$. Posljedično, model se više ne može izraziti pomoću samo jednog broja koji pokazuje intenzitet nagiba funkcije u odnosu na x-osu i koja neminovno prolazi kroz tačku $(0, 0)$. Umjesto toga imamo novu formulu koja predstavlja klasičnu formulu regresione analize i glasi:

Formula linearne regresione analize

$$\hat{y} = a + b * x$$

Formula izgleda nešto drugačije od prethodne, ali prije nego što objasnim razlike, napominjem da smo konačno došli do termina: **regresiona analiza** (engl. *regression analysis*). Riječ regresija u statističkom kontekstu znači svođenje, odnosno mogućnost da se vrijednosti jedne varijable svedu na jednu ili više drugih varijabli. U našem primjeru se razlike u školskom prosjeku pokušavaju svesti na razlike u inteligenciji.

Da se vratimo na objašnjenje formule. Razlika se odnosi na broju konstanti koje se u njoj nalaze. Umjesto jedne stalne vrijednosti sada ih imamo dvije koje se najčešće označavaju sa a i b. Komponenta a se naziva **tačka sjecišta** zato što predstavlja y vrijednost na kojoj linearna funkcija siječe y-osu kada je $x = 0$. Komponenta b se naziva **nestandardizovani regresioni koeficijent** i ona nam govori o nagibu linearne funkcije. Ako b ima pozitivnu vrijednost onda imamo nagib nagore kao i u slučaju pozitivne korelacije, a ako je b negativno nagib nadole. Konačno, ako je $b = 0$ onda se

od tačke a povlači linija paralelna sa x-osom koja u tom slučaju predstavlja i aritmetičku sredinu za y varijablu (tada je korelacije jednaka 0).

Konkretna regresiona formula za odnos inteligencije (IQ) i prosječne ocjene (O) za naših pet ispitanika bi glasila:[3]

$$\hat{O} = 0.9448 + 0.0318 * IQ$$

[3] Prikazujem ovdje podatke na čak četiri decimale kako bi najsavjesniji među vama uspjeli izračunati baš te brojeve koji su dati u tabeli. Inače, ovo računar obavlja za nas, te nije potrebno biti tolika cjepidlaka.

Da malo proučimo ovu formulu. Ako bi osoba imala $IQ = 0$ onda bi imala i školsku prosječnu ocjenu 0.95. Znamo da je i jedno i drugo teorijski nemoguće, te se ova potencijalna nelogičnost nekad predstavlja kao zamjerka nestandardizovanih formula. Nadalje, vidimo da je b vrijednost pozitivna, što znači da sa sve većim rastom IQ skora očekujemo da opazimo i veću ocjenu. Nestandardizovani regresioni koeficijent b je direktno povezan sa koeficijentom korelacije r formulom $b = r * \frac{SD_y}{SD_x}$, što nam omogućava prelazak iz nestandardizovanog u standardizovani model i obratno. Ali da ipak prođemo konkretne kalkulacije sa nestandardizovanim koeficijentima (Slika 19.5).

id	x	y	a	b	b * x	ŷ	e
1	85	3.1	0.9448	0.0318	2.7030	3.6	-0.5
2	93	4.9	0.9448	0.0318	2.9574	3.9	1.0
3	100	3.6	0.9448	0.0318	3.1800	4.1	-0.5
4	107	4.6	0.9448	0.0318	3.4026	4.3	0.3
5	115	4.4	0.9448	0.0318	3.6570	4.6	-0.2

Slika 19.5: Tabela sa nestandardizovanim regresionim koeficijentima.

U tabeli vidimo već poznate simbole. U posljednjoj koloni se nalaze i reziduali koji se označavaju sa e (od engl. *error* što znači greška, a shodno formuli $e = y - \hat{y}$). S obzirom na to da se za pojedinačne ispitanike razlikuju i y i e, u indeksu stoji i oznaka i koja označava da za svakog ispitanika važe drugačiji brojevi, pa kompletirana formula glasi:

$$y_i = a + b * x + e_i$$

I ponovićemo još jednom šta je šta, jer je ovo važna konceptualna formula:

- y_i je stvarna vrijednost zavisne varijable za ispitanika i
- a je tačka sjecišta (y vrijednost kada je $x = 0$)
- b je nestandardizovani regresioni koeficijent (nagib linije)

- x_i je vrijednost x-varijable za ispitanika i
- e_i je rezidual ili greška modela za ispitanika i

Na osnovu ovoga slijedi da za ispitanika #3 možemo rastaviti njegov finalni skor ($y_3 = 3.60$) na dio koji je određen modelom i dio koji predstavlja grešku, odnosno druge izvore varijabilnosti.

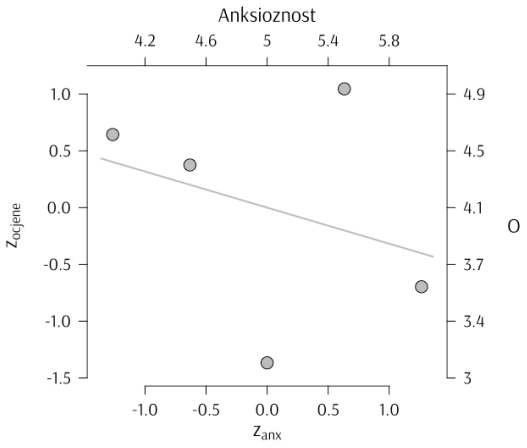
$$3.6 = 0.9448 + 0.0318 * 100 - 0.52$$

Iz ovog konkretnog slučaja vidimo da je formula precijenila školski uspjeh učenika #3 za pola ocjene (0.52). Drugim riječima, kada bismo se oslonili na ovaj model, očekivali bismo da osoba koja je postigla prosječnu mjeru inteligencije ($IQ = 100$) ima prosječnu ocjenu u srednjoj školi 4.12. Jedno objašnjenje za opaženo odstupanje, odnosno grešku modela, je da postoje varijable mimo inteligencije koje nismo uzeli u obzir, a koje su mogle uticati da ovaj učenik ostvaruje niži skor. To su snižena akademska motivacija, povećana školska anksioznost, nedjelotvorne strategije učenja koje koristi, nizak socioekonomski status, itd. Drugo objašnjenje odstupanja - kojeg bi trebalo da držimo u vidu - je mogućnost da model koji smo dobili (tj. linija fita koja je povučena) važi za dati uzorak, ali da ona predstavlja samo jednu od mogućih linija koje odražavaju vezu između x i y varijable i koje se mijenjaju u zavisnosti od toga koji ispitanici su u uzorku. Ovo drugo objašnjenje nam predočava zaista bitno ograničenje koje se tehnički ublažava izračunavanjem i iscrtavanjem intervala povjerenja / uvjerenja oko linije, ali zasad ćemo se fokusirati na prvo objašnjenje.

U kontekstu regresione analize postoji veći broj imena za y i x varijable. Na primjer, y-varijabla se naziva ishodišnom, kriterijskom,zavisnom, ciljnom ili endogenom, dok se x-varijable nazivaju prediktorima, eksplanatorima, nezavisnim, kovarijatnim ili egzogenim varijablama. Problem leži u tome što dosta datih naziva podrazumijeva ili temporalni slijed gdje bi x varijabla morala prethoditi ili čak kauzalnost. Iskustvo mi pokazuje da to lako dovodi do semantičke konfuzije gdje se onda podrazumijevaju uloge u modelu koje ove varijable u stvarnosti nemaju. Naime, nekada se dešava da i x i y varijabla zavise od treće varijable, nekada da imaju uzajamni odnos jedna na drugu, nekad se dešava da zapravo x zavisi od y, a može se desiti da ako u modelu imamo više x-varijabli one imaju različit temporalni i logički slijed. Čak i za relativno očitu vezu koju sam uzeo za primjer, onu između inteligencije i školskog uspjeha – pogotovo ako se one simultano mjere – postoji mogućnost uzajamnog mehanizma, gdje se sa sve boljim školskim uspjehom učenik ulaže više u učenje, čime angažuje mentalne reprezentacije i prepoznavanje apstraktnih obrazaca, odnosno dalje razvija inteligenciju. Nazivi kao što su nezavisna i zavisna varijabla, prediktori i ishodišne varijable mogu biti itekako opravdani, ali to zavisi od vjerodostojnosti odnosa varijabli i nacrta. Iz tog razloga se u ostatku teksta držim što neutralnijih imena koja samo opisuju statistički postupak.

Šta se dešava kada u formulu uključimo i još jednu x-varijablu?

Dakle, da vidimo šta se događa kad y-varijablu koja je u fokusu pokušamo dovesti u vezu sa dva ili više x-regresora. U dosadašnji primjer ćemo uključiti i anksioznost. Ako pogledamo odgovarajući grafikon raspršenja (Slika 19.6), vidi se da je anksioznost negativno povezana sa prosječnom školskom ocjenom. Vrijednost koeficijenta korelacije je $r = -.32$, što znači da je $r^2 = 0.10$. Postavlja se pitanje da li je ispravno sabrati koeficijente determinacije za ova dva regresora, odnosno da li je varijabilitet koje inteligencija i anksioznost zajednički dijele sa školskom ocjenom jednaka $0.25 + 0.10 = 0.35$? Odgovor je **NE**, a u nastavku ćemo saznati zašto.



Slika 19.6: Grafikon raspršenja anksioznosti i prosječnih ocjena.

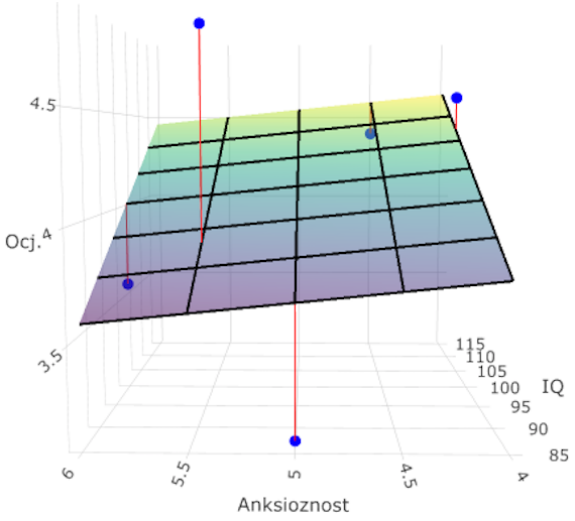
Zašto ne možemo jednostavno da združimo efekte pojedinačnih varijabli?

Razlog za to jeste što u ovom slučaju postoji korelacija između dvije x-varijable, te se njihovi veza sa y-varijablom dijelom preklapa. Naime, inteligencija (x_1) i anksioznost (x_2) su u ovom slučaju negativno povezane $r = -0.49$. Negativna korelacija između njih se i teorijski očekuje budući da anksioznost odvlači pažnju i radnu memoriju koje su bitne za kognitivno procesiranje, a time i za uspjeh na testovima inteligencije. Ako dva regresora međusobno dijele varijabilitet, onda oni ne mogu potpuno nezavisno jedan od drugog predviđati raznolikost školske ocjene.

Šta su glavne karakteristike višestruke linearne regresione analize?

Dobrodošli u čudesni svijet **višestruke linearne regresione analize** (engl. *multiple linear regression analysis*)! Za početak bi trebalo da shvatite kako multipla linearna regresiona analiza izgleda. Kao što ste vidjeli, modeli za pojedinačne linearne analize koje se zovu i bivarijatne - gdje imamo jednu x-varijablu i jednu y-varijablu - su prave linije prikazane u dvodimenzionalnom prostoru. To bi značilo da bismo odnos tri dimenzione varijable morali predstaviti u trodimenzionalnom prostoru? Baš tako. Ali da li ovo znači da ako imamo sedam dimenzionih varijabli onda bismo morali imati sedmodimenzionalni univerzum u kojem se predstavljaju odnosi? Upravo tako! Čarobnost matematike leži

u tome da ona zaista podrazumijeva postojanje tzv. multivarijatnog prostora, mada mi to sebi ne možemo da predstavimo u fizičkom obliku jer smo kao bića opremljeni da vizuelno percipiramo samo trodimenzionlani fizički svijet. Međutim, principi koji će biti prikazani važe za takve n-dimenzionalne svjetove. Pogledajmo sada prikaz 3d modela multiple regresione analize (naravno, putem dvodimenzionalne Slike 19.7).



Slika 19.7: Linearni regresioni model sa dva regresora.

Dakle, regresioni model je se od jedne *prave linije* pretvorio u jednu *ravan*. Možete takođe da vidite kako je ta ravan postavljena unutar prostora određenog inteligencijom, anksioznošću i ocjenama. Primjećujemo da bismo očekivali da bi najvišu ocjenu trebalo da ima učenik koji ima najvišu inteligenciju i istovremeno najnižu anksioznost. Iz grafikona se može zaključiti i da je sa ocjenom inteligencija snažnije povezana nego anksioznost, tako da bi učenik sa visokom inteligencijom i visokom anksioznošću trebalo da ima višu prosječnu ocjenu nego učenik sa niskom anksioznošću, ali i niskom inteligencijom.

Takođe možemo da uočimo da nijedan od opaženih skorova (plavi kružići) ne leži tačno na ravni. Za svakog ispitanika su crvenom bojom označena odstupanja od ravni, odnosno reziduali: najmanji su za dva ispitanika sa desne strane grafikona. Zahvaljujući svim dobijenim rezidualima možemo da izračunamo grešku modela i uporedimo ga sa početnom varijabilnošću, ali i sa modelima zasnovanim na po samo jednoj x-varijabli. Procedura izračunavanja zbira kvadrata odstupanja je ista, tj. dužina svakog odstupanja od modela se kvadrira i onda sumira. Moguće bi bilo izračunati to geometrijski, ali je ovo prilika i da

naučite konceptualnu formulu koja postavlja svaki opaženi skor u trodimenzionalnu poziciju:

$$y_i = a + b_1 * x_1 + b_2 * x_2 + e_i$$

Novost u odnosu na model sa samo jednom x-varijablom jeste da sad imamo dva regresiona koeficijenta za nagibe (b_1 i b_2). To je sasvim logično jer imamo dva regresora, a logično je i to da tačka sjecišta a sada predstavlja vrijednost y-varijable kada su obje x vrijednosti jednake 0. Komponente y_i i e_i imaju isto značenje kao u slučaju sa samo jednim regresorom: y_i je opaženi skor na uzorku za ispitanika i, a e_i označava razdaljinu između skora procijenjenog regresionom funkcijom $\hat{y}_i$ i opaženog skora y_i. Naprijed navedeno znači da je formula za dobijanje procijenjenog, modelom pretpostavljenog skora jednaka:

$$\hat{y} = a + b_1 * x_1 + b_2 * x_2$$

Nova varijabla $\hat{y}$ ("ipsilon sa kapicom") spada u **sintetičke varijable** (engl. *synthetic variable*), budući da je dobijena fuzionisanjem dvije ili više postojećih varijabli. Kao i u slučaju bivarijatne regresione analize, kada saznamo njenu vrijednost za svakog ispitanika, možemo izračunati i vrijednost reziduala, jer znamo da je $e_i = y_i - \hat{y}_i$. Naravno, algoritam će za nas izračunati tačne vrijednosti a i b komponenata za koje je minimizirana suma kvadrata odstupanja. Postoji samo jedno moguće rješenje za takvo izračunavanje a i b komponenti, a konkretna formula u našem primjeru glasi:

$$\hat{O} = 1.7194 + 0.0287 * IQ - 0.0935 * Anx$$

Služeći se podacima iz tabele (Slika 19.8) možemo izračunati odnos sume kvadrata odstupanja modela ($\Sigma(e^2)$) i ukupnog kvadrata odstupanja od aritmetičke sredine $\Sigma((y - M(y))^2$, odnosno izračunati koeficijent determinacije.

IQ	Anx	y	ŷ	e	e²	y-M(y)	(y-M(y))²
85	5	3.1	3.69	-0.59	0.35	-1.02	1.0404
93	5.5	4.9	3.87	1.03	1.06	0.78	0.6084
100	6	3.6	4.03	-0.43	0.18	-0.52	0.2704
107	4	4.6	4.41	0.19	0.03	0.48	0.2304
115	4.5	4.4	4.60	-0.20	0.04	0.28	0.0784
	M(y)=	4.12		**Σ(e²) =**	1.659	**Σ(y-M(y))²=**	2.228

Slika 19.8: Podaci za primjer multiple regresione analize.

Prema formuli za koeficijent determinacije, rezultat je: $1 - (1.659/2.228) = 1 - 0.745 = 0.255$. Uporedimo to sa modelom

u kojem je jedina x-varijabla bila inteligencija gdje je $r^2 = 0.248$. Uvjeren sam da ste se iznenadili! Iako anksioznost samostalno smanjuje grešku procjene za 10.1%, njen doprinos onome što je već "objasnila" inteligencija je tek 0.7%. Razlog smo već pomenuli, a to je postojanje korelacije između inteligencije i anksioznosti. U slučaju da njih dvije nisu bile u korelaciji, onda bi zbir objašnjenosti varijanse bio 25%.

Da li ova nova mjera povezanosti ima svoje ime?

Da. Korelacija između sintetičkog skora ($\hat{y}$) sastavljenog od dvije ili više varijabli i opaženog skora (y) se naziva **koeficijent višestruke korelacije** (engl. *coefficient of multiple correlation*) i označava se velikim slovom R. Shodno tome, multipli koeficijent determinacije je takođe zaslužio veliko slovo, pa se označava sa R^2. Za R i R^2 je moguće izračunati p-vrijednost, ali istraživači su obično više zainteresovani da uporede različite modele prema pokrivenosti varijabiliteta y-varijable. Najčešće se za to koristi prosta razlika koeficijenata determinacija ΔR^2. Dakle, ako bismo htjeli uporediti Model 1 ($\mathcal{M}_1$) koji se sastoji samo od anksioznosti i Model 2 ($\mathcal{M}_2$) koji je uključio i inteligenciju i anksioznost, onda bismo dobili da je $\Delta R^2 = 0.255 - 0.101 = 0.154$. I obratno, ako bismo poredili model sa obje x-varijable sa onim koji sadrži samo inteligenciju dobili bismo da je $\Delta R^2 = 0.255 - 0.248 = 0.007$. U ovom drugom slučaju bismo zaključili da je doprinos anksioznosti objašnjenju prosječne srednjoškolske ocjene povrh onoga što objašnjava inteligencija tek 0.7%.

Da li regresije koje postepeno nadograđuju model novim regresorima imaju posebno ime?

Ako $\mathcal{M}_2$ sadrži sve elemente koje ima $\mathcal{M}_1$ i još barem jedan novi, onda govorimo o posebnom obliku regresione analize koji se zove hijerarhijska regresiona analiza. **Hijerarhijska regresiona analiza** (engl. *hierarchical regression analysis*) analizira promjene funkcionalnosti regresionih modela kako se oni postepeno nadograđuju dodavanjem x-varijabli. Na primjer, u našem slučaju smo imali jedan regresor na početku, a u drugom koraku je dodan i drugi. U sljedećem koraku bismo mogli dodati još pet varijabli (npr. crte ličnosti Velikih pet), pa empirijski provjeriti da li crte ličnosti dodatno doprinose objašnjenju varijabiliteta školske ocjene povrh onoga što već rade anksioznost i inteligencija. Takve hijerarhijske regresione analize predstavljaju osnovnu tehniku

Hijerarhijska regresiona analiza

provjere **inkrementalne valjanosti** konstrukata, odnosno osnovnu tehniku kojom se ispituje da li je smisleno uključiti neki novi konstrukt u već utvrđenu teorijsku šemu x i y varijabli. U hijerarhijskim analizama se obično prvo uvode "vremenski i logički starije" varijable (npr. pol, godine starosti), nakon čega nastupaju psihološke varijable za koje već znamo da pokazuju empirijsku vezu sa kriterijskom varijablom, te se na samom kraju uključuje i varijabla koja bi predstavljala novost i čiji se doprinos želi ispitati.

Kako se utvrđuje da li dodavanje x-varijabli bitno poboljšava statistički model?

Imate li drugove koji u svemu nastoje da nađu nešto dobro i pozitivno interpretiraju šta god da im se dešava? E upravo takva je i priroda linearne regresione analize. Koju god x-varijablu ubacite u nju, regresiona analiza će pokušati da se adaptira i da iz nje izvuče najbolje, odnosno da podesi model tako da se dobije još manji kvadrat odstupanja. Drugim riječima, najgore što se modelu može desiti kada u njega ubacite još jednu varijablu je da R^2 i srodne mjere ostanu iste, one se u svakom slučaju neće smanjivati. Zato je potrebno biti svjestan da se model koji je izveden / istreniran na jednom uzorku i koji pokazuje neku količinu saglasnost sa podacima NE MOŽE smatrati provjerenim modelom koliko god on izgledao optimistično. Kao što ćemo ubrzo vidjeti, rast vrijednosti R^2 koeficijenta, pa čak ni statistička značajnost te promjene, se ne smatraju dovoljno dobrim dokazima da dodavanje varijabli popravlja model.

Prije nego li se upoznamo sa konkretnim povećalima uvida u dokazni materijal, kratko ćemo se osvrnuti na to zašto su štedljivost i uopštivost odlike dobrih naučnih modela, budući da se dodatne metode zasnivaju na štedljivosti i, s njom povezanom, uopštivosti. Preferiranje štedljivih modela u objašnjenju pojava "fancy" naučnim jezikom naziva **parsimoničnost**. Princip parsimoničnosti kaže da je smisleno odlučiti se za najjednostavniji model ako biramo između više njih koji približno uspješno opisuju stvarnost. Jednostavnost, odnosno komplikovanost modela se u statistici odražava u broju uključenih varijabli i hipoteziranih odnosa između njih. Ako u modelu imamo samo 2 varijable (npr. 1 nezavisnu i 1 zavisnu) jasno je šta ispitujemo. Ako imamo tri varijable, broj odnosa raste na 3, a sa četiri varijable to je već 6 međusobnih odnosa. Pitanje uopštivosti modela smo već pominjali kada smo se bavili lokalnom regresionom funkcijom, odnosno kada smo naučili zašto je

jednostavni linearni model uopštiviji nego krivudava liniju koja ima potencijalno mnogo parametara - mjesta gdje se funkcija lomi.

Dobro je podsjetiti se da poređenje različitih statističkih modela predstavlja jednu od najvažnijih funkcija statistike – što je već naglašeno na početku knjige. Kada raspravljamo o tome koja teorija je tačna, od nadglasavanja u kafanskoj prepirci bi mnogo veću epistemološku vrijednost trebalo da ima poređenje statističkih mjera za dva modela nakon što su prikupljeni konkretni podaci. Sljedeći primjeri u tekstu koji prikazuju konkretne tehnike su iz didaktičkih razloga vezani za hijerarhijske regresione analize, ali bi trebalo da znate da je pomoću sličnih principa moguće statistički porediti i modele koji nisu nastajali kao evolucija nekog proto-modela.

Zašto koeficijent determinacije nije idealno sredstvo za poređenje statističkih modela?

A sada konačno: kako se modeli konkretno porede? Najprimitivnije sredstvo – za koje sam već rekao da po sebi nije dovoljno ubjedljivo – jeste razlika koeficijenata determinacije (ΔR^2) između dva modela. Istina je da taj podatak nešto govori, ali potrebno ga je konzumirati sa oprezom. Znamo da se R^2 povećava sa svakom pridodatom x-varijablom – makar neznatno – a to može da se desi i u slučajevima kada je korelacija između nekog regresora i kriterija jednaka okruglo 0. Dovoljno je da taj regresor korelira sa drugim regresorom, pa da dođe do većeg R^2. Tehnički razlog zašto se to dešava jeste da novi regresor skida šum sa već postojećeg, odnosno u njegovoj vezi sa y-varijablom. Ponavljam jer je to jako bitno: regresiona funkcija je po svojoj prirodi podešena tako da se uvijek dobije najbolje moguće rješenje (tzv. minimizacija sume kvadrata odstupanja), pa procedura prisilno iscjeđuje iz novog regresora sve što je potencijalno korisno. U rijetkim slučajevima dodavanje regresora koji nema korelacija sa kriterijskom varijablom ima teorijskog smisla, ali mnogo češće takav rezultat predstavlja specifičnost datog uzorka i vrlo je sporan za interpretaciju. Kao ilustraciju ćemo uzeti naš primjer u kojem modelu u kojem inteligencija "objašnjava" ocjene ($M1$) dodajemu još jednu, namjerno generisanu varijablu ($M2$). Recimo da je to sposobnost divergentnog mišljenja, bitan aspekt kreativnosti. Ta sposobnost korelira relativno visoko sa inteligencijom ($r = .40$), ali je korelacija praktično 0 sa ocjenama ($r = -.002$). Na Slici 19.9 su predstavljene različite mjere poređenja modela.

Model	R	R^2	Adjusted R^2	AIC	BIC	RMSE
1	0.498	0.248	−0.00270	14.7	13.6	0.579
2	0.546	0.298	−0.40454	16.4	14.8	0.559

Slika 19.9: Mjere poređenja regresionih modela (ispis iz Jamovi softvera).

Kao prvo, vidimo da je uvođenje sposobnosti divergentnog mišljenja u model dovelo do rasta koeficijenta determinacije $\Delta R^2 = 0.298 - 0 - 248 = .05$, iako divergentno mišljenje i ocjena nisu u korelaciji. Šta nam ovakav rezultat može sugerisati? Na primjer, da ako bismo nekako izdvojili kreativnost iz veze inteligencije i ocjene – odnosno odstranili višak varijabilnosti koji je zajednički između kreativnosti i inteligencije – onda bismo mogli još preciznije dovesti u vezu inteligenciju i ocjene. Da biste ovo uklopili u neki teorijski narativ potrebna vam je dobra doza divergentnog mišljenja. Mi ćemo se vratiti ovoj tehnici izdvajanja varijanse koji omogućava fantastične uvide, ali na ovom mjestu se možemo složiti da svjedočimo paradoksu koji zaslužuje svoje ime.

Supresorske varijable

Varijable koje nimalo ili slabo koreliraju sa y-varijablom, ali značajno povećavaju R^2 se nazivaju **supresorske varijable** (engl. *supressor variables*). Idealni uslovi da se nagledate supresorskih varijabli su da kreirate model sa ogromnim brojem regresora. Kada se supresorski efekti ukažu u rezultatima, časni istraživači se hvataju za glavu jer ne znaju šta će s njima, odnosno kako da ih protumače. Uzroci njihove pojave mogu biti smisleni moderatori, ali njih je potrebno potkrijepiti teorijskim argumentima i dokazati kros-validacionim procedurama ili replikacijama. Ako ne postoji takva mogućnost topla preporuka je da se te varijable isključe iz modela i procedura ponovi bez njih.

"Dobro, ubijeđeni smo da nije pametno krenuti u lov na dobar model samo sa prostom razlikom koeficijenata determinacije (ΔR^2). Šta se još nudi?" Kao što ste mogli pretpostaviti, moguće je izračunati p-vrijednost za nultu hipotezu koja kaže da je razlika između dva koeficijenta determinacije hijerarhijski vezanih modela jednaka okruglo nula, tj. $H_0 : \Delta\rho^2 = 0.00$. Tom nultom hipotezom se pretpostavlja da se statistici na uzorcima distribuišu pozitivno asimetrično kao F-distribucija (ovo je prvi put da je srećemo, ali je vrlo prisutna u klasičnoj statistici zaključivanja) koja svoj finalni oblik dobija kad se uzme u obzir broj x-varijabli i broj ispitanika $\mathcal{M}_1$ i $\mathcal{M}_2$. U ovom konkretnom slučaju rezultat je: $F(1, 2) = 0.142, p = .743$.

No, znate da p-vrijednosti imaju gomilu mana. Na primjer,

posebno loš recept bi bio osloniti se na p-vrijednosti ako imamo veliki broj regresora. Veliki broj regresora povećava šansu Greške tipa I još brže nego što je to slučaju bivarijatnih analiza, upravo zbog mogućnosti supresorskih uticaja. Blagoslov da se mnogo varijabli može strpati u jednu jednačinu neki istraživači zloupotrebljavaju, pa su nekorumpirani istraživači morali da domisle načine kako da im stanu u kraj. U opticaju je nekoliko pristupa koji imaju za cilj da se poveća uopštivost pobjedničkog modela. Svi oni balansiraju između kriterijuma uspješnosti fita (tj. stepena saglasnosti modela i opaženih podataka) i zahtjeva da se na što manju mjeru svede uključivanje elemenata koji su važni u uzorku, ali potencijalno nevažni u populaciji.

Šta je to prilagođeni koeficijent determinacije?

Prvi među njima predstavlja mjeru koja se redovno nalazi u statističkim softverima. **Prilagođeni koeficijent determinacije** (engl. *adjusted R-squared*, R^2_{adj}) predstavlja R^2 dobijen na uzorku umanjen nakon korektivnog zahvata koji se obično zasniva na tri sastojka:

- veličini uzorka = n (veći uzorak $\rightarrow$ manja korekcija)
- broju regresora u modelu = p (više x-varijabli $\rightarrow$ veća korekcija)
- dobijenoj veličini efekta na uzorku = R^2 (veće $R^2 \rightarrow$ manja korekcija)

Zadatak prilagođenog koeficijenta determinacije je da prorekne najvjerovatniji parametar koeficijenta determinacije (P^2).[4] Najčešće korištena formula se pripisuje Ezekielu (ne biblijskom nego statističarskom proroku):

[4] Da, dobro ste pročitali P, jer veliko grčko slovo ρ odgovara našem ćiriličnom R, odnosno latiničnom P.

$$R^2_{\text{Ezekiel}} \;=\; 1 - (1 - R^2)\,\frac{n-1}{n-p-1}$$

Istraživač tako može uporediti prilagođene R^2 vrijednosti za dva modela. Da vidimo šta radi ta korekcija na našem primjeru. Kada kao regresor imamo samo inteligenciju onda je $R^2 = 0.25$, ali je $R^2_{adj} = -0.002$, odnosno praktično nula! Vjerovatno mislite da ne može gore, ali može. Dovoljno je da dodamo anksioznost u model. Tada je $R^2 = 0.255$, ali prilagođena vrijednost postaje ubjedljivo negativna $R^2_{adj} = -0.49$! Dakle, dok su vrijednosti R^2 ograničene donjom granicom na 0, model koji ima dvije x-varijable i samo pet ispitanika je korištenjem ove

mjere posramljujuće kažnjen. I treba da bude, jer je svaka analiza sa dva regresora i pet ispitanika vrijedna izrugivanja. "Drevna" statistička tradicija je zabranjivala da se regresija izvodi bez najmanje 10 ispitanika po regresorskoj varijabli, ali modernija shvatanja kažu da bi trebalo uspostaviti barem neki minimum od ukupno 50 ispitanika nakon čega se za svaki uvedeni regresor dodaje još fiksni broj ispitanika. Naravno, još bolje je za regresione analize unaprijed izračunati statističku moć ako je nekom do testiranja nulte hipoteze i razumnog planiranja Greške tip 2.

Dakle, prilagođeni koeficijent determinacije koriguje statistik uzimajući u obzir i kompleksnost modela (broj varijabli) i uopštivost (broj ispitanika). Da li onda možemo zaključiti da najbolji model mora biti onaj koji ima najveću vrijednost R^2_{adj}? Avaj, i R^2_{adj} ima svoje mane. Zanemariva je to što može da ima teorijski nemoguće negativne vrijednosti, budući da se to lako rješava zaokruživanjem svakog negativnog rezultata na 0. Stvarna mana je što svoj posao stražara R^2_{adj} slabije obavlja na sve većim uzorcima. Kada je n veliko tada i modeli sa velikim brojem prediktora dobro prolaze, te se dešava da se favorizuju kompleksniji modeli sa puno šuma / buke koji precjenjuju "objašnjeni" y-varijabilitet. Čestu pojavu da su model i podaci saglasni iznad onoga što bismo našli u populaciji, zovemo *pretjerano uklapanje* ili *pretreniranost* modela (zapravo ga najčešće zovemo *overfit* što je postalo odomaćeno iz engleskog).

Drugo važno ograničenje leži u tome što je R^2_{adj} tačkasta procjena populacionog parametra. Tačkaste procjene parametra liče na preambiciozne osobe niskih kompetencija. Pokušavaju da kažu mnogo više nego što im kapaciteti dozvoljavaju i ispadaju smiješni ako ih uzimamo zaozbiljno. To je očito na našim primjerima sa nultim i negativnim vrijednostima parametara, pa je potrebno imati na umu da je R^2_{adj} prilično primitivno sredstvo korekcije koje može biti i pregrubo i prenježno, te se može koristiti za komparaciju modela, ali ne i za stvarnu procjenu parametra. Treba napomenuti da je moguće izračunati i intervale povjerenja kako za samu procjenu R^2_{adj}, tako i za razlike između modela ΔR^2_{adj}, mada se to rijetko radi i tek nedavano je takva mogućnost postala opcija u nekim komercijalnim softverima.[5]

[5] npr. Raykov & DiStefano, 2025, https://doi.org/10.1177/00131644251329178

Šta su to informacioni kriteriji?

Usljed toga se pri poređenju statističkih modela često koristi druga grupa mjera koja je zasnovana na teoriji informacija. One se zato nazivaju **informacioni kriteriji** (engl. *information criteria*).

Njihov cilj je da direktno uporede različite modele uzimajući u obzir stepen saglasnosti podataka i modela (njihov fit), ali da ipak oštrije kazne složene modele. Njihova velika prednost je što nisu ograničeni na linearne regresione analizama (za razliku od R^2_{adj}) nego su prilagođene za ogroman broj naprednijih analiza kojima imena nećemo ni spominjati, da se ne uplašite.

Dva tradicionalna informaciona kriterija koja se pojavljuju i u softverima i u naučnim člancima, su AIC (*Akaike information criterion*) i BIC (*Bayesian information criterion*), mada je potrebno napomenuti da se pojavljuju i noviji, bolji i ljepši.[6] Stari znanci su uključeni u AIC i BIC formule koje se koriste za linearnu regresiju: veličina uzorka (n), broj regresora (p) i mjera fita modela (SS_{LM}). Formula za AIC je:

$$AIC = n \log\left(\frac{SS_{LM}}{n}\right) + 2(p+1)$$

BIC ima vrlo sličnu formulu, pa sam se čak pitao da li da kandidujem neku novčanu nagradu ako možete sami da uočite koje su razlike među njima i kakve posljedice to ima.

$$BIC = n \log\left(\frac{SS_{LM}}{n}\right) + (p+1)\log(n)$$

Odgovor za milion uzbekistanskih somova (nešto manje od sedamdeset eura u trenutku pisanja knjige) bi bio da BIC teži da favorizuje jednostavnije modele, pogotovo ako su uzorci veći. Ako se ove dvije mjere uporede, onda se može reći da je za BIC još bit(c)nije da svaki dodani regresor doprinese smanjenju greške modela. Zato BIC lakše izbacuje iz optimalnog modela sve što liči na višak, što kao lošu posljedicu može imati da izabere prejednostavne modele (crkvenoslovenski rečeno: da *anderfituje*). S druge strane, AIC je blaži prema dodatnim prediktorima pa se u finalni model mogu uvući i varijable koje zapravo predstavljaju šum. Inače AIC je osmišljen sa plemenitim ciljem da među različitim kandidatima izabere onaj model koji bi trebalo da ima najmanju grešku odstupanja u budućim uzorcima. S druge strane, smisao života BIC-a je da identifikuje pravi pravcati istiniti model među kandidatima, ako se on uopšte nalazi među njima. Ako vam oboje zvuči prepoetično – da sad parafraziram McElreatha – to znači da ste normalni.

[6] Nećemo zalaziti pretjerano u dekonstrukciju informacionih kriterija, jer potpuno dijelim upozorenje R. McElreatha koji najavljuje opis istih na tridesetak stranica svoje knjige *Statistical Rethinking* ovim riječima: "…gradivo koje slijedi je teško. Ako budete zbunjeni na nekom mjestu, onda ste zapravo normalni. Osjećanje zbunjenosti pokazuje da se mozak navikava na ono što se izučava…".

Ali kako se uopšte donose odluke o modelima na osnovu informacionih kriterija?

Dobro pitanje zar ne, jer nije baš jasno šta nam grozomorne formule govore? Pomalo je smiješno (ako ste ljubitelj ironije) to što dobijene vrijednosti nemaju nikakvo inherentno značenje. AIC i BIC su relativne mjere, odnosno ništa vam ne znači da ih izračunate samo za jedan model. Istraživačima su one od pomoći kada direktno upoređuju njihove vrijednosti za različite modele (tj. $\mathcal{M}_1$, $\mathcal{M}_2$ i/ili $\mathcal{M}_n$). Načelno, najbolji model će biti onaj koji ima najmanju vrijednost informacionih kriterija.

U našem ranijem primjeru (Slika 19.9) su za oba pokazatelja dobijene niže vrijednosti za modele u kojima je inteligencije jedini regresor. Pogledajmo i Sliku 19.10 gdje drugi model sadrži inteligenciju i anksioznost kao x-varijable.

Model	R	R^2	Adjusted R^2	AIC	BIC	RMSE
1	0.498	0.248	−0.00270	14.7	13.6	0.579
2	0.505	0.255	−0.48908	16.7	15.1	0.576

Slika 19.10: Mjere poređenja regresionih modela (ispis iz Jamovi softvera) za modele sa inteligencijom ($M1$) i inteligencijom i anksioznošću ($M2$).

Dakle, vidimo da su i uvom slučaju oba pokazatelja $\mathcal{M}_1$: $AIC = 14.7, BIC = 13.6$ niža nego za model u kojem je dodana i anksioznost $\mathcal{M}_2$: $AIC = 16.7, BIC = 15.1$. Ovdje je potrebno navesti da postoje i preporuka za poređenje različitih modela tako što se model sa minimalnom vrijednošću – koji je načelno najbolji – poredi sa drugim modelima putem jednostavne razlike: $\Delta AIC = AIC_n - AIC_{min}$. Preporuke koji su niže predočene – i koje bi trebalo koristiti sa oprezom – se zasnivaju na činjenici da ubacivanjem svake nove x-varijable u model automatski povećavamo AIC vrijednost za 2 (jer u formuli imamo $2 * p$):

- $\Delta AIC \leq 2$ novi model je jednako vjerovatan ili bolji
- $4 \leq \Delta AIC \leq 7$ novi model ima značajno slabiju podršku
- $\Delta AIC \geq 10$ novi model je gotovo definitivno slabije vjerovatan od modela sa kojim se poredi

Za BIC takođe postoje neke preporuke koje glase:

- $\Delta BIC = 0$–2: Slabi dokazi protiv modela s višim BIC-om (tj. model s višim BIC-om je gotovo jednako dobar kao BIC min model).

- $\Delta BIC = 2$–6: Umjereni dokazi protiv modela s višim BIC-om.
- $\Delta BIC = 6$–10: Snažni dokazi protiv modela s višim BIC-om.
- $\Delta BIC > 10$: Vrlo snažni dokazi protiv modela s višim BIC-om

Sve u svemu, AIC i BIC nam govore da nemamo baš jake argumente da tvrdimo da je jedan model vjerovatniji od drugog. Kad pominjemo vjerovatnoću modela, pominjanje velečasnog Bayesa u imenu BIC-a nije bilo bez razloga. BIC je zapravo približna procjena Bayesovog faktora (ako nekog zanima: na logaritamskoj skali sa uniformnim apriornim vjerovatnoćama). Ali zašto bismo išli naokolo, ako bi postojala mogućnost da direktno izračunamo Bayesove faktore?

Možemo li izračunati Bayesove faktore za različite regresione modele i porediti ih na osnovu toga?

Nego šta. Izvođenje Bayesovog faktora u izboru kompetitivnih modela je posljednjih godina postalo vrlo popularno (možda se snimi i Netflix serija). Kao što znate Bayesov faktor predstavlja omjer vjerovatnoća da se podaci na uzorku dobiju u zavisnosti od dva modela koji se porede. Za regresione analize – ali i mnoge druge postupke – istraživači mogu da dobiju broj koji ne samo da poredi nultu hipotezu sa jednim alternativnim modelom, nego koji poredi taj model i sa drugim, pa i trećim, četvrtim, petim alternativnim modelom. Popularnosti je pridonio i softver JASP u kojem se vrlo lako izvodi automatska procjena mnoštva modela, nakon čega se oni u ispisu lijepo poredaju po veličini.

Naravno, bejzijanski pristup sa svojim prednostima nosi i glavobolje u vidu podešavanju apriornih vjerovatnoća. Dodatna komplikacijica leži u tome da je za regresije potrebno definisati dvije vrste apriornih vjerovatnoća: početnu vjerovatnoća svakog od modela i vjerovatnoću parametara unutar modela. Vjerovatnoća modela se može podesiti tako da svi modeli imaju jednaku vjerovatnoću, ali postoje i druge mogućnosti koje mogu biti opravdanije. Na primjer, za nultu hipotezu se može postaviti vjerovatnoća od 50%, a da se preostalih 50% "bratski" podijeli na tri alternativna modela (tj. 16.7%). Moguće je podesiti i da sve kompleksniji modeli sa više parametara dobijaju nižu apriornu vjerovatnoću kako bi se u startu favorizovali jednostavniji modeli sa manje regresora. Kada je u pitanju postavljanje apriornih vjerovatnoća parametara unutar modela njih je takođe

moguće podesiti i informativno, ali postoje i mnogobrojne slabo informativne varijante koje softveri omogućavaju. Naravno da odabir obje vrsta vjerovatnoća može uticati na konačni rezultate, te je zdrava praksa procijeniti robusnost nalaza, odnosno koliko rezultati variraju zavisno od oblika prethodnih distribucija.

Da vidimo sada kako izgleda kada se Bayesovi faktori izračunaju za naš primjere. Slika 19.11 prikazuje rezultate.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	0.250	0.343	1.563	1.000	0.000
iq	0.250	0.258	1.046	0.755	0.248
anx	0.250	0.221	0.851	0.645	0.101
iq + anx	0.250	0.178	0.650	0.520	0.255

Slika 19.11: Bayesovi faktori za regresioni model izvedene u softveru Jamovi.

Kolona *P(M)* nam prikazuje da su postavljene jednake apriorne vjerovatnoće za sve modele. Kolona *P(M | data)* nam prikazuje izmijenjene vjerovatnoće modela nakon integracije podataka. Vidimo da su za dva modela (nulti i model sa inteligencijom) vjerovatnoće povećane, a za druga dva smanjene. Kolona BF_M pokazuje omjer datog modela nakon integrisanja podataka u odnosu na prosjek svih ostalih modela. Na primjer, za nulti model je kalkulacija:

$$BF_{M_0} = \frac{0.343}{(0.258 + 0.221 + 0.178)/3} = 1.56$$

Direktni Bayesovi faktor koji porede dva modela jedan naspram drugoga su dati u koloni BF_{10}. Na primjer, kada se model u kojem se nalaze i inteligencija i anksioznost uporede sa modelom u kojem se nultim modelom dobija se $BF_{10} = 0.178/0.343 = 0.520$. Ovo znači da je nulti model dva puta vjerovatniji od modela sa obje x-varijable, jer je inverzna vrijednost $BF_{01} = 1/0.52 = 1.92$, iako on obuhvata 0% varijanse na uzorku, dok drugi model obuhvata 25.5%. Naravno, vrijednosti BF važe pod uslovima koji su podešeni u softveru - u ovom slučaju uniformne apriorne vjerovatnoće modela i zadane apriorne vrijednosti parametara (JZS, $rsc = 0.354$). Zahvaljujući principu tranzitivnosti iz tabele još možemo izvući omjere vjerovatnoća različitih alternativnih modela. Model koji sadrži samo inteligenciju vjerovatniji je od modela sa inteligencijom i anksioznošću 1.45 puta, jer je $BF_{12} = 0.755/0.520 = 1.45$.

Šta nam svi ovi Bayesovi faktori govore o našem primjeru? Samo to da su nam podaci sa uzorka nedovoljno informativni, odnosno da nemaju snagu da bitnije promijene prethodna uvjerenja. To je zapravo i jedini smislen zaključak na uzorku od 5 ispitanika ako već nismo u proračun uključili vrlo snažne ekspertske pretpostavke. Za dati zadatak smo ih možda i mogli uključiti jer iz mnogih istraživanja znamo da su inteligencija i anksioznost povezani sa školskom ocjenom, ali zamislite situaciju u kojoj se po prvi put neke dvije x-varijable (npr. dužina butne kosti i broj treptaja u sekundi) stavljaju u odnos sa y-varijablom (ocjenom). Slaboinformativni Bayesovi faktori bi nam tad brutalno saopštili da smo džaba izvodili tako nejako istraživanje.

Da od malih uzoraka nema neke vajde smo znali i ranije. Ali kada uradimo istraživanje na ozbiljnijem uzorku i između više kandidata se odlučimo za najbolji model (ili nekoliko najboljih) potrebno je ocijeniti relativni značaj pojedinačnih regresora.

Kako se ocjenjuje relativni značaj pojedinačnih regresora u modelu?

Kako bih odgovorio na ovo pitanje koristićemo primjer istraživanja za master tezu koju sam mentorisao (S. Simić, 2017, *Poređenje uloge kognitivne sposobnosti i emocionalne kompetencije pri prepoznavanju facijalnih ekspresija*). U fokusu rada je bio konstrukt emocionalne kompetentosti, poznatiji pod svojim estradnim nazivom emocionalna inteligencija. Emocionalna inteligencija je od sredine 1990-ih privukla ogromnu pažnju naučne zajednice, ali i opšte javnosti. S jedne strane su proponenti smiono tvrdili da je emocionalna inteligencija prediktor različitih životnih ishoda snažniji i od opšte inteligencije, empirijski najdokazanijeg psihološkog prediktora. S druge strane, kritičari su sugerisali da je emocionalna inteligencija u stvari samo preobučena opšta inteligencija sa nekim dodatnim crtama ličnosti i/ili posebnim domenima znanja. Problem je predstavljalo i to da je većina istraživanja mjerila emocionalnu inteligenciju ili metodom samoprocjene ili testovima objektivnog postignuća, ali vrlo rijetko koristeći oba pristupa. Mi smo željeli da provjerimo oba pristupa u teorijski relevantnom zadatku (prepoznavanje facijalnih ekspresija), te da pritom statistički kontrolišemo efekat opšte inteligencije.

Konkretno, randomizovanim redoslijedom smo ispitanicima izlagali niz slika facijalnih ekspresija (ukupno 28 slika koje prikazuju šest osnovnih emocija i neutralno lice) u blic trajanju

od 200ms. U istoj seansi smo zadali i matrični neverbalni test opšte inteligencije, objektivni test razumijevanja emocija gdje zadaci traže od ispitanika da procijene najvjerovatniju emociju u datoj situaciji, te upitnik samoprocjene emocionalne kompetentnosti. S obzirom na zahtjevnost istraživanja mi smo uspjeli skupiti 100 ispitanika, ali sam za potrebe daljnjeg prikazivanja tehnika simulirao podatke za 500 ispitanika na osnovu matrice korelacije dobijene u stvarnom istraživanju. To znači da su simulirani podaci stoje u gotovo identičnom odnosu kao u našem istraživanju, samo ćemo u ovoj fiktivnoj verziji imati veću statističku moć.

Dakle, kriterijska ili y-varijabla u ovom primjeru je broj tačno prepoznatih facijalnih ekspresija, dok su regresori uspješnost na testu opšte inteligencija, samoprocijenjena emocionalna kompetencija i objektivno procjenjivana emocionalna kompetencija. Logično je da bismo u prvom koraku procjene relativne važnosti pojedinačnih varijabli morali pogledati proste korelacije. Na Slici 19.12 vidimo da objektivno mjerena emocionalna kompetencija ostvaruje višu korelaciju nego li opšta inteligencija (obje statistički značajne na nivou .001), dok samoprocjena ostvaruje nultu korelaciju sa uspješnošću prepoznavanja facijalnih ekspresija.

	fac_exp	iq_test	ei_samop	ei_test
fac_exp	—			
iq_test	0.174 ***	—		
ei_samop	0.002	–0.117 **	—	
ei_test	0.290 ***	0.227 ***	0.092 *	—

Note. * p < .05, ** p < .01, *** p < .001

Slika 19.12: Matrica korelacija (Jamovi).

No shvatili smo već da proste linearne korelacije ne mogu ispričati čitavu priču, te da je potrebno da vidimo šta se događa kada sva tri regresora uključimo u jednačinu. U sljedećoj tabeli (Slika 19.13) se vidi da su za svaku komponentu (sem za tačku sjecišta, engl. *Intercept*) uz proste korelacije navedene još i vrijednosti nestandardizovanih (b) i standardizovanih regresionih koeficijenata ($\hat{\beta}$), standardne greške za nestandardizovani koeficijente (SE_b), njihovi 95%-tni intervali povjerenja, t-vrijednosti, te na osnovu njih izračunate odgovarajuće p-vrijednosti.

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_0	(Intercept)	40.236	0.371		108.484	0.000	39.507	40.965
H_1	(Intercept)	22.719	3.188		7.126	3.666e−12	16.455	28.983
	iq	0.063	0.025	0.113	2.553	0.011	0.015	0.112
	ei_samop	−0.004	0.018	−0.009	−0.215	0.830	−0.039	0.031
	ei_test	0.451	0.075	0.265	5.994	3.935e−9	0.303	0.598

Slika 19.13: Tabela regresionih koeficijenata (ispis iz JASP softvera).

Kada je u pitanju međusobno poređenje važnosti regresora nestandardizovani koeficijenti nam nisu od neke koristi, budući da su mjerne skale različite. Zato su u psihologiji tradicionalni prvi izbor standardizovani regresioni koeficijenti i p-vrijednosti. U slučaju bivarijatne regresije – gdje postoji samo jedan regresor – standardizovani regresioni koeficijent je jednak koeficijentu linearne korelacije. Standardizovani regresioni koeficijent će biti identičan koeficijentu korelacije i u multiploj regresiji pod uslovom da je korelacija između regresora jednaka tačno nula. To se u praksi ekstremno rijetko dešava. Iz tabele se vidi da su standardizovani koeficijenti nešto niži od koeficijenata korelacija, te da je to naročito primjetno smršao koeficijent za opštu inteligenciju. Interesantno je i da je p-vrijednost za opštu inteligenciju je ostala statistički značajna, ali sada na nivou .05. Ranija redovna praksa istraživača je bila da sude o vrijednosti regresora na osnovu p-vrijednosti, pa bi mnogi već otpisali opštu inteligenciju kao vrijednu pomena i izbacili je iz modela ako bi se p-vrijednost prešla nivo .05, dok bi eventualno ostavili u modelu neki supresorski efekat koji je postao statistički značajan. Ukratko, to ne radite nego razmotrite obavezno teorijske razloge zašto je do takve razlike između prostih korelacija i standardizovanih koeficijenata moglo doći.

Ali šta nam zapravo govore regresioni koeficijenti?

Prema definiciji, standardizovani regresioni koeficijent nam pokazuje koliku promjenu y-varijable možemo da očekujemo ako se regresor promijeni za jednu standardnu devijaciju – dok se svi drugi regresori u modelu drže konstantnim. To znači da bismo za osobu koja je ostvarila za 1SD bolji skor od prosjeka na testu emocionalne kompetencije, mogli očekivati da ima i za 0.26SD bolji rezultat u pogađanju facijalnih ekspresija od onoga koji bi postigao prosječan ispitanik – pod uslovom da se ne mijenjaju skorovi na opštoj inteligenciji (i samoprocijenjenoj

emocionalnoj kompetenciji, mada je taj efekat mizeran). Dakle, standardizovani regresioni koeficijent nam iskazuje destilovani očekivani efekat regresora na kriterijsku varijablu. Nažalost, zvuči mnogo bolje nego što u stvarnosti jeste, budući da se u stvarnosti obično dešava da regresori stoji u komplikovanim međuodnosima. Ali sada je vrijeme da se usmjerimo pažnju na dugo najavljivani fenomen parcijalizacije.

Šta označava termin parcijalizacija?

Sam termin **parcijalizacija** (engl. *partialization*) označava postupak kojim se uklanja, odnosno statistički kontroliše efekat jedne ili više dodatnih varijabli na odnos između dvije varijable (x i y) koji stoji u fokusu proučavanja. Na primjer, želimo da ispitamo korelaciju između između srednjoškolske ocjene i fakultetske ocjene možemo procijeniti na dva načina. Jedan način je da jednostavno na uzorku procijenimo korelaciju. Drugi način je da procijenimo njihov odnos ako iz njega izuzmemo potencijalne efekte "trećih" varijabli kao što su inteligencija, crte ličnosti, pol, strategije učenja i akademska motivacija. Zašto bismo to htjeli uraditi? Na primjer, ako bi inteligencija već objašnjavala u velikoj mjeri korelaciju između srednjoškolske i fakultetske ocjene, onda nije potrebno da koristimo i srednjoškolsku ocjenu i test inteligencije za rangiranje kandidata. Od toga šta istraživači žele tačno da postignu zavisi i da li će i koje će varijable držati pod kontrolom.

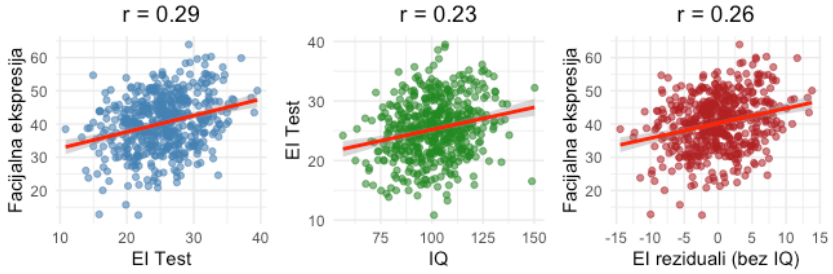
Vratimo se sad na primjer sa prepoznavanjem facijalnih ekspresija. Željeli smo da vidimo koliko su zaista povezane sposobnost prepoznavanja izraza lica i objektivno testirana emocionalna inteligencija, onda kada se iz tog odnosa isključi ono što objašnjava opšta inteligencija. Taj podatak će nam pokazati da li zaista postoji neki jedinstven doprinos emocionalne kompetentnosti.

Kako se, tehnički gledano, izvodi čarolija parcijalizacije?

Tehnika statističke kontrole je začuđujuće jednostavna, jer koristi samo ono što vam je već poznato: reziduale. Dakle, želimo da testiramo hipotezu kritičara koncepta emocionalne inteligencije koji kažu da svi bitni efekti emocionalne inteligencije potiču zapravo iz opšte inteligencije. Zato bismo morali nekako pročistiti opštu inteligenciju iz emocionalne inteligencije. To se radi u dva koraka. U prvom koraku ćemo izvesti linearnu regresiju gdje će y-varijabla biti emocionalna kompetencija, a x-varijabla opšta inteligencija. Dobijene reziduale ($y - \widehat{y}$) ćemo onda

koristiti umjesto originalnih skorova za emocionalnu kompetenciju i izračunaćemo njihovu korelaciju sa prepoznavanjem facijalnih ekspresija. Da podsjetim, ako bi korelacija između opšte inteligencije i emocionalne kompetentnosti bila jednaka 0, onda bi reziduali bili samo centrirano transformisani skorovi y-varijable. S druge strane, ako bi korelacija između inteligencije i emocionalne kompetentnosti bila savršena, onda bi svi reziduali iznosili 0 (postala bi konstanta), te bi iz toga slijedilo da je korelacija između pročišćene emocionalne kompetentnosti i sposobnosti prepoznavanja facijalne ekspresije jednaka 0. Ovo je suština toliko puta pominjane statističke kontrole.

Na Slici 19.14 se vidi da u našem primjeru nije došlo do znatne promjene visine korelacije nakon kontrole za opštu inteligenciju. Korelacija nakon hirurškog odstranjivanja inteligencije iz emocionalne kompetentnosti ($sr = .26$, desni grafikon) je tek nešto niža u odnosu na originalnu korelaciju između emocionalne kompetentnosti i sposobnosti pogađanja emocionalnog izraza ($r = .29$, lijevi grafikon).



Slika 19.14: Grafikoni raspršenja korelacija nultog reda i semi-parcijalne korelacije.

Primjerom je prikazana *semi-parcijalizacija* (engl. *semi-partialization*, notacija: sr), odnosno polu-odstranjenje varijabilnosti budući da smo samo iz x-varijable odstranili efekat opšte inteligencije. Inače, u kontekstu multiplih regresionih analiza sa još većim brojem x-varijabli semi-parcijalne korelacije se dobija tako što se ciljana varijabla istovremeno regresira na sve ostale x-varijable (npr. i na samoprocjenu emocionalne kompetencije u našem primjeru). Konačna svrha semi-parcijalizacije je da utvrdimo koliki je jedinstveni, unikatni doprinos određenog regresora u objašnjenju y-raznolikosti kada se odstrane veze sa svim ostalim regresorima. U kontekstu regresionih analiza je još smislenije kvadrirati datu vrijednost kako bismo dobili sr^2, odnosno komad koeficijenta determinacije kojeg samo ta varijabla može da objasni. Takvi podaci su posebno bitni za selekcionu praksu kada se u baterijama zadaje mnogo instrumenata, pa putem ovog procesa možemo da prosudimo koje komponente su neza-

Svrha semi-parcijalnih korelacija

mjenjive drugima. U našem primjeru je na početku procenat preklapajućeg varijabiliteta između emocionalne kompetentnosti i sposobnosti prepoznavanja facijalnih ekspresija bio $r^2 = 0.084$, od čega je $sr^2 = 0.066$ ili 78.6% predstavljalo jedinstveni doprinos emocionalne kompetentnosti (dok bi se 21.4% moglo pripisati i varijabilitetu opšte inteligencije).

Ako postoji polu-parcijalizacija, onda vjerovatno postoji i puna parcijalizacija?

Naravno. Kada govorimo o **punoj parcijalizaciji** (engl. *full partialization*), mislimo na postupak uklanjanja efekta trećih varijabli iz obje varijable čiju korelaciju ispitujemo. Drugim riječima, dok semi-parcijalna korelacija uklanja varijabilnost kontrolne varijable samo iz jednog člana korelacije, puna parcijalizacija to čini za oba člana.[7] Postupak je isti kao što sam naveo za semi-parcijalizacije, samo što se on ponavlja i na y-varijabli. Rezultat je parcijalna korelacija (pr) koja nam sugeriše kakva bi bila veza između dvije varijable kada bi se uklonio uticaj svih ostalih varijabli u modelu. Puna parcijalna korelacija je posebno korisna kada želimo da utvrdimo da li veza između dvije varijable *opstaje* i kada se iz nje potpuno iščisti neka treća varijabla ili više njih.

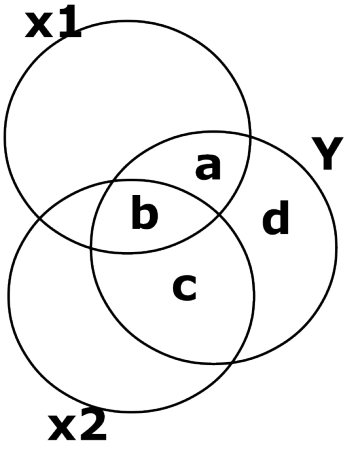
Konačno, parcijalizacija i semi-parcijalizacija se često šematski i vrlo uprošćeno prikazuju pomoću Venovih dijagrama, a ne bih želio da ostanete uskraćeni za tako lijep i poučan prikaz. Slika 19.15 prikazuje standardnu sliku (prilagođenu u odnosu na onu koja se nalazi za Wikipediji). Evo šta označavaju slova $a - d$ u pogledu varijabiliteta koji se dijeli sa y-varijablom:

- a = dio varijabilnosti koji se dijele samo sa $x1$
- b = dio varijabiliteta koji dijeli i sa $x1$ i sa $x2$
- c = dio varijabilnosti koji se dijele samo sa $x2$
- d = dio varijabilnosti koji ne dijeli sa varijablama u modelu jednak koeficijentu indeterminacije
- e = dio varijabiliteta koji dijele $x$varijable koji nije zajednički sa y-varijablom

Iz navedenog se mogu izvesti konceptualne formule za sr^2 i pr^2, a time i za sr i pr. Formula koja ukazuje na jedinstveno preklapanje $x1$-varijable sa y varijablom je:

$$sr^2 = \frac{a}{a + b + c + d}$$

[7] Parcijalne korelacije žive novu mladost u psihologiji zahvaljajući odnedavno ekstremno popularnom **psihometrijskom mrežnom modeliranju** (engl. *psychometric network analysis, PNA*). Mrežno modeliranje ima za cilj da predstavi međusobne veze velikog broja konstrukata, a snage parcijalnih korelacija su idealne da pokažu između kojih varijabli opstaju veze što se sugerišu potencijalne puteve uzročno-posljedičnog djelovanja. No, to je tema za neki psihometrijski tekst.



Slika 19.15: Prikaz dekompozicije objašnjenosti varijanse y-varijable putem Venovog dijagrama.

Kada je u pitanju parcijalna korelaciju potpuno se izbacuje dio koji se preklapa sa $x2$-varijablom, pa je formula:

$$pr^2 = \frac{a}{a+d}$$

Budući da je nazivnik u formuli za parcijalnu korelaciju manji po pravilu je parcijalna korelacija veća nego semi-parcijalna korelacija.

A koje pretpostavke se sve moraju ispuniti kako bi se regresione analize pravilno interpretirale?

Pretpostavki ima mnogo. U načelu, one važe ne samo za regresione analize, nego i za **opšti linearni model** (engl. *general linear model, GLM*) čiji je regresiona analiza najbolji predstavnik. U ovu široku porodicu statističkih postupaka spadaju i druge tehnike koje se često koriste u psihologiji, kao što su analiza varijanse (ANOVA) za poređenje grupnih prosjeka, analiza kovarijanse (ANCOVA) koja u model uključuje i kontinuirane i kategoričke prediktore, pa čak i osnovni t-test s kojim smo se već susretali. Pravilno razumijevanje i provjera ovih pretpostavki je ono što razdvaja mehanicističko korištenje softverskih ispisa od smislene statističke analize.

Pretpostavke se mogu podijeliti u dvije grupe:

- Pretpostavke o distribuciji podataka i reziduala
- Pretpostavke o specifikaciji modela i odabiru varijabli

Koje su to pretpostavke u pogledu distribucije podataka i reziduala?

Zapravo ove pretpostavke već znate:

- **Linearnost.** Korelacije između svih varijabli bi morale biti dovoljno linearne. Provjera se vrši pregledom grafikona raspršenja. Ako pretpostavka nije ispunjena, ali jeste pretpostavka o monotonom odnosu onda nam može pomoći konvertovanje varijabli u rangovane vrijednosti. Za složenije nelinearne odnose, potrebno je koristiti polinomne regresije ili druge nelinearne metode koje ne obrađujemo ovdje.

- **Homoskedastičnost.** Siguran sam da (ni)ste zapamtili ovu riječ pominjanu kada smo obrađivali korelacije. Radi se o tome da bi se reziduali morali raspršivati simetrično i podjednakim intenzitetom duž regresione linije. Grafikon raspršenja nam pokazuje da li je to tako. Ako postoji neki drugačiji vidljivi obrazac onda je potrebno to uzeti u obzir pri tumačenju rezultata.

- **Nepostojanje stršećih vrijednosti.** Kao i za linearne korelacije, s tim da u višestrukim regresionim analizama imamo višestruke kombinacije potencijalno čudnih skorova. Dvije mjere koje uzimaju u obzir odnos sa y-varijablom su posebno popularne. Cookova distanca (engl. Cook's distance) je mjera uticaja pojedinačnog ispitanika na čitav model. Vrijednosti Cookove distance veće od 0,5, a posebno one koje prelaze vrijednost 1, ukazuju na značajan uticaj ispitanika na model i zahtijevaju dodatnu provjeru. S druge strane, mjera DFBETAS (engl. *difference in standardized betas*) procjenjuje uticaj pojedinačnog ispitanika na određeni regresioni koeficijent u modelu. Pragovi za označavanje sumnjivih vrijednosti su apsolutne vrijednosti DFBETAS-a veće od 1 ili one koje prelaze vrijednost $\frac{2}{\sqrt{n}}$, gdje je n broj ispitanika.

- **Normalnost distribucije reziduala.** Ova pretpostavka je bitna za manje uzorke i kada izvodimo zaključke o p-vrijednostima. Reziduali bi trebalo da su približno normalno distribuirani ili relativno simetrični. Tehnike provjere su histogram ili Q-Q grafikon. Ako se uoči veliko odstupanje rangiranje varijabli ili neka druga transformacija mogu da pomognu, pogotovo ako to sve prate ekstremne vrijednosti. Važno je napomenuti da ova pretpostavka ne podrazumijeva da distribucija varijabli mora biti normalna ili simetrična, mada to povećava vjerovatnoću da će i reziduali biti takvi.

Koje su to pretpostavke o specifikaciji modela i odabiru varijabli?

Ove pretpostavke su raznorodne:

- **Ne postoje previsoke korelacije među regresorima.** Kada među regresorima ima više vrlo sličnih regresora (npr. skorovi na testovima verbalnih analogija, informativnog

znanja, usmenog razumijevanja) koji međusobno ostvaruju visoke korelacije onda model teško može da pravedno razdvoji doprinose pojedinačnih varijabli. Pojava previsokih korelacija se naziva (multi)kolinearnost (engl. *(multi)collinearity*). Konkretno, previsokim se već smatraju korelacije veće od .70 – a sigurno one veće od .90. Postoje i specijalizovane mjere multikolinearnosti kao što je faktor inflacije varijance (engl. *Variance Inflation Factor, VIF*) čija vrijednost veća do 4 ili 5 predstavlja alarm. Multikolinearnost se tretira tako što će u modelu ostati ili samo jedna, teorijski najsmislenija, od visoko koreliranih varijabli ili će se iz više takvih varijabli izvesti jedna sintetička varijabla (npr. izračunavanjem prosjeka standardizovanih vrijednosti ili postupkom koji se naziva analiza glavnih komponenti).

- **Nezavisnost mjerenja.** Sjećate se vezanih mjera? Klasična regresiona analiza podrazumijeva da su opservacije međusobno nezavisne. Kršenje pretpostavke može nastati usljed načina uzorkovanja, nacrta studije ili procedure prikupljanja podataka; na primjer, ako anketiramo više osoba iz istog domaćinstva, ljude koji žive u istoj ulici ili kada prikupljamo podatke u već formiranim grupama kao što su školska odjeljenja, sportski timovi, preduzeća. Narušavanje pretpostavke dovodi do potcjenjivanja standardnih grešaka regresionih koeficijenata, što rezultuje preuskim intervalima povjerenja i malim p-vrijednostima. Ako već znate da na uzorku imate međusobno vezane mjere (npr. ispitanike iz različitih država), preporučivo je koristiti napredni regresioni oblik: višestruku regresionu analizu na višestrukim nivoima (engl. *multilevel regression analysis*).

- **U model su uključene sve relevantne varijable.** Naravno, već sama ova pretpostavka zvuči nemoguće. Ipak, korisno ju je imati na umu iz dva važna razloga. Prvo, ako iz modela izostavimo neku bitnu kontrolnu varijablu, procijenjeni regresioni koeficijenti za ostale varijable će biti pristrasni, jer će implicitno sadržavati uticaj i tih izostavljenih, teorijski „starijih" varijabli. Na primjer, prilikom istraživanja odnosa između crta ličnosti (X) i školskog postignuća (Y) često se dobija rezultat da savjesnost i emocionalnost pokazuju pozitivne regresione koeficijente. Međutim, kada se u model uvede kontrolna varijabla pol ispitanika, savjesnost obično ostaje značajan prediktor, dok se efekat emocionalnosti primjetno smanjuje. Razlog ove promjene je što djevojke istovremeno ostvaruju bolje školsko postignuće, pokazuju

nešto viši nivo savjesnosti, ali i značajno viši nivo emocionalnosti. Izostavljanje varijable pol iz modela bi dovelo do precjenjivanja stvarne veze između emocionalnosti i školskog uspjeha. Drugi problem je pretjerana kontrola, što se može desiti kada se u model uvode irelevantne varijable i s njima nepotreban šum ili kada se blokiraju medijacioni putevi djelovanja - a nije postojala namjera da se testira medijacioni efekat. Na primjer, ako se procjenjuje efekat pohađanja psihoterapije (X) na kvalitet života (Y), a kao dodatni regresor se uključi anksioznost (Z), onda će se koeficijent za psihoterapiju smanjiti proporcionalno snazi veze anksioznost-kvalitet života i stepenu u kojem je ta veza glavni put djelovanja terapije. Dakle, tu se radi o medijacionom odnosu $X \rightarrow Z \rightarrow Y$ koji se mora testirati iz više koraka, budući da pohađanje psihoterapije snižava anksioznost, što je onda povezano sa višim kvalitetom života.

Zapravo smo posljednjom tačkom zagrebali suštinske izazove ispitivanja teorijskih veza putem opšteg linearnog, ali i drugih statističkih modela. Na primjer, osim odnosa **medijatora**, moguće je da unutar modela postoje **moderatorski** efekti.[8] Oni se unutar opšteg linearnog okvira modeliraju kao proizvodi između regresora (npr. Pol * Godina studija), te predstavljaju temelj tzv. faktorskih analiza varijansi u kojima svi regresori mogu biti kategoričke varijable. Nadalje, na uzorku se mogu ukazati prividne korelacije ili da se ne ukažu stvarnu korelacije između dvije varijable usljed selekcionisanog uzorka iako je na populaciji istina drugačija. Na primjer, poznato je da su savjesnost i inteligencija u populaciji odraslih imaju nultu do zanemarivu korelaciju. Ali i inteligencija i savjesnost su nešto što doprinosi upisu prestižnog studija (npr. psihologije) i to na sinergijski način. Međutim, veza između njih i upisa studija je nezavisna i asimetrična s obzirom da je često potrebno da budete ili vrlo inteligentni ili vrlo savjesni da biste upisali studij: *visoka inteligencija* $\rightarrow$ *prijem na fakultet* $\leftarrow$*visoka savjesnost*. Tako se na uzorku primljenih studenata može uočiti bitnija negativna korelacija između inteligencije i savjesnosti koja ne postoji u opštoj populaciji. Ovaj odnos se naziva odnosa sudarača (engl. *collider*), budući da maskira prave efekte obostranim usmjerenjem na y-varijablu. Sve gore navedene odnose (medijatorski, moderatorski, sudarači) je bitno uzeti u obzir pri tumačenju rezultata i tu vam striktno statistički postupci nekad ne mogu pomoći bez dobrog teorijskog i metodološkog rezonovanja.

[8] Usput, statistički dio medijacionih i moderacionih analiza se lako izvode u Jamovi i JASP softveru: https://blog.jamovi.org/2017/09/25/medmod.html ili https://jasp-stats.org/2020/03/12/mediation-and-moderation-analysis-in-jasp/. Ako nekog interesuje da podigne znanja o opštem linearnom nivou na visok stepen, preporučujem knjigu: Fox, 2016, https://www.john-fox.ca/AppliedRegression/index.html

Zar je "već" kraj?

Da, napokon smo došli do kraja ovog teksta. Nadam se da su vam načete teme otvorile apetit za nastavak istraživanja statističkih tehnika. Ovo što ste saznali o linearnoj regresiji vam nadohvat ruke dovodi mnogo drugih tema koje ćete relativno lako savladati: analizu varijanse, medijatorske i moderatorske analize, logističke regresije, modeliranje latentnih varijabli (tj. eksploratorne i konfirmativne faktorske analize), mrežno modeliranje, modeliranje višestrukih nivoa. Iako zvuče kao čudovišta, zapravo su to sve alati koji samo nadograđuju ovo što ste upravo shvatili. Postoji mnogo onlajn izvora, raznih tutorijala i dobrih knjiga koji predstavljaju konkretne tehnike koje su ovdje izostavljene, te koje vam mogu pomoći da nastavite da sebi krčite put kroz teren naprednije statistike ili softversku primjenu naučenog. Želim vam puno uspjeha u daljnjim psihološkim avanturama sa nadom da je ova etapa u kvantitativnom opismenjavanju prošla manje bolno nego što ste očekivali. Pitanja, komentare i uočene greške šaljite mejlom na sinisa.lakic@ff.unibl.org kako bi sljedeće generacije polagale ispite iz statistike u stanju nirvane.

Spisak literature

Afonso, J., Ramirez-Campillo, R., Clemente, F. M., Büttner, F. C., & Andrade, R. (2023). The Perils of Misinterpreting and Misusing "Publication Bias" in Meta-analyses: An Education Review on Funnel Plot-Based Methods. *Sports Medicine, 54*(2), 257–269. https://doi.org/10.1007/s40279-023-01927-9

Anvari, F., & Lakens, D. (2019). The SAM: A simplified approach to null hypothesis testing. *Advances in Methods and Practices in Psychological Science, 2*(2), 159-170. https://doi.org/10.1177/2515245919847202

Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology, 96*, 104159. https://doi.org/10.1016/j.jesp.2021.104159

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2021). *Introduction to meta-analysis (2nd ed.)*. Wiley.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2016). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods, 49*(5), 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology, 65*(3), 145–153. https://doi.org/10.1037/h0045186

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist, 49*(12), 997–1003. https://doi.org/10.1037/0003-066x.49.12.997

Correll, J., Mellinger, C., McClelland, G. H., & Judd, C. M. (2020). Avoid Cohen's 'Small', 'Medium', and 'Large' for Power Analysis. *Trends in Cognitive Sciences, 24*(3), 200–207. https://doi.org/10.1016/j.tics.2019.12.009

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry, 12*(2), 137–148. Portico. https://doi.org/10.1002/wps.20038

Cumming, G., & Calin-Jageman, R. (2024). *Introduction to the New Statistics: Estimation, Open Science, and Beyond (2nd ed.)*. Routledge.

Cumming, G., Williams, J., & Fidler, F. (2004). Replication and Researchers' Understanding of Confidence Intervals and Standard Error Bars. *Understanding Statistics, 3*(4), 299–311. https://doi.org/10.1207/s15328031us0304_5

de Finetti, B. (1970). Logical foundations and measurement of subjective probability. *Acta Psychologica, 34*, 129–145. https://doi.org/10.1016/0001-6918(70)90012-0

de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods, 21*(3), 273–290. https://doi.org/10.1037/met0000079

Dick, F., & Tevaearai, H. (2015). Significance and Limitations of the p Value. *European Journal of Vascular and Endovascular Surgery, 50*(6), 815. https://doi.org/10.1016/j.ejvs.2015.07.026

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in biomedical science: a review of three human research domains. *Royal Society Open Science, 4*(2), 160254. https://doi.org/10.1098/rsos.160254

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics, 56*(2), 455–463. Portico. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Eckshtain, D., Kuppens, S., Ugueto, A., Ng, M. Y., Vaughn-Coaxum, R., Corteselli, K., & Weisz, J. R. (2020). Meta-Analysis: 13-Year Follow-up of Psychotherapy Effects on Youth Depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 59*(1), 45–63. https://doi.org/10.1016/j.jaac.2019.04.002

Etz, A., Chávez de la Peña, A. F., Baroja, L., Medriano, K., & Vandekerckhove, J. (2024). The HDI + ROPE decision rule is logically incoherent but we can fix it. *Psychological Methods*. https://doi.org/10.1037/met0000660

Fashing, J., & Goertzel, T. (1981). The Myth of the Normal Curve a Theoretical Critique and Examination of its Role in Teaching and Research. *Humanity & Society, 5*(1), 14–31. https://doi.org/10.1177/016059768100500103

Fisher, M., & Keil, F. C. (2018). The Binary Bias: A Systematic Distortion in the Integration of Information. *Psychological Science, 29*(11), 1846–1858. https://doi.org/10.1177/0956797618792256

Fox, J. (2016). *Applied regression analysis and generalized linear models (3rd ed.)*. Sage Publications.

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gao, M. & Li, Q. (2024). A family of Chatterjee's correlation coefficients and their properties. arXiv:2403.17670. https://doi.org/10.48550/arXiv.2403.17670

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest, 8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Götz, M., Sarma, A., & O'Boyle, E. H. (2024). The multiverse of universes: A tutorial to plan, execute and interpret multiverses analyses using the R package multiverse. *International Journal of Psychology, 59*(6), 1003–1014. Portico. https://doi.org/10.1002/ijop.13229

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Gyimesi, M. L., Webersberger, V., Del Giudice, M., Voracek, M., & Tran, U. S. (2025). One App to Rule Them All: A One-Stop Calculator and Guide for 95 Effect-Size Variants for Two-Group Comparisons of Central Tendency, Variability, Overlap, Dominance, and Distributional Tails. *Advances in Methods and Practices in Psychological Science, 8*(2). https://doi.org/10.1177/25152459251323186

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods, 12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapmann & Hall/CRC Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Heyman, T., & Vanpaemel, W. (2022). Multiverse analyses in the classroom. *Meta-Psychology, 6*. https://doi.org/10.15626/mp.2020.2718

Higgins, J. P. T. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hitchcock, C., & Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. *The British Journal for the Philosophy of Science, 55*(1), 1–34. https://doi.org/10.1093/bjps/55.1.1

Howick, J. (2011). The Philosophy of Evidence-Based Medicine. https://doi.org/10.1002/9781444342673

Imhoff, R., & Koch, A. (2017). How Orthogonal Are the Big Two of Social Perception? On the Curvilinear Relation Between Agency and Communion. *Perspectives on Psychological Science, 12*(1), 122–137. https://doi.org/10.1177/1745691616657334

Ioannidis, J. P. A. (2008). Why Most Discovered True Associations Are Inflated. *Epidemiology, 19*(5), 640–648. https://doi.org/10.1097/ede.0b013e31818131e7

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal, 127*(605), F236–F265. https://doi.org/10.1111/ecoj.12461

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence, 41*(4), 212–221. https://doi.org/10.1016/j.intell.2013.03.003

Junker, R. R., Griessenberger, F., & Trutschnig, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Computational Statistics & Data Analysis, 153*, 107058. https://doi.org/10.1016/j.csda.2020.107058

King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(2), 171–184. https://doi.org/10.1586/erp.11.9

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis (2nd ed.)*. Academic Press.

Lakens, D. (2022). *Improving Your Statistical Inferences*. Retrieved from https://lakens.github.io/statistical_inferences/. https://doi.org/10.5281/zenodo.6409077

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S.-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., … Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour, 2*(3), 168–171. https://doi.org/10.1038/s41562-018-0311-x

Lakić, S. (2019). Bayesov faktor: Opis i razlozi za upotrebu u psihološkim istraživanjima. *Godišnjak za psihologiju, 16*, 39-58. https://doi.org/10.46630/gpsi.18.2019.03

Lakić, S., Damjenić, M. i Šain, D. (2017). Velikih pet kao prediktori uspjeha u studiranju psihologije na Univerzitetu u Banjoj Luci. U *IV Kongres psihologa BiH – Zbornik radova (str. 325-338)*. Društvo psihologa Republike Srpske i Društvo psihologa Brčko Distrikta.

Lakić S. & Perić, A. (2025). Comparing the Effects of Intelligence, Learning Strategies, and HEXACO Personality Traits on High School Academic Achievement. U S. Borojević (ur.) *Zbornik radova Banjalučki novembarski susreti 2024*. Univerzitet u Banjoj Luci.

Lakić, S., Miščević, S. i Tutnjević, S. (2014). Adaptacija kratke forme Children's Behavior Questionnaire (CBQ-SF) za procjenu temperamenta kod djece od tri do šest godina. *Empirijska istraživanja u psihologiji - Zbornik Radova*. Univerzitet u Beogradu.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389–402. https://doi.org/10.1177/2515245918787489

Leek, J. T., & Peng, R. D. (2015). Statistics: P values are just the tip of the iceberg. *Nature, 520*(7549), 612–612. https://doi.org/10.1038/520612a

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology, 51*(3), 485–504. Portico. https://doi.org/10.1002/ejsp.2752

Martin, G. N., & Clarke, R. M. (2017). Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.00523

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN (2nd ed.)*. CRC Press.

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science, 34*(2), 103–115. https://doi.org/10.1086/288135

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Morone, A., Caferra, R., Casamassima, A., Cascavilla, A., & Tiranzoni, P. (2021). Three doors anomaly, "should I stay, or should I go": an artefactual field experiment. *Theory and Decision, 91*(3), 357–376. https://doi.org/10.1007/s11238-021-09809-0

Nau, R. F. (2001). De Finetti was right: Probability does not exist. *Theory and Decision, 51*(2/4), 89–124. https://doi.org/10.1023/a:1015525808214

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology, 73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

O'Boyle Jr., E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65*(1), 79–119. https://doi.org/10.1111/j.1744-6570.2011.01239.x

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716

Patel, J. K., & Read, C. B. (1982). *Handbook of the Normal Distribution*. Marcel Dekker, Inc.

Petersen, J. L., & Hyde, J. S. (2010). A meta-analytic review of research on gender differences in sexuality, 1993–2007. *Psychological Bulletin, 136*(1), 21–38. https://doi.org/10.1037/a0017504

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472–491. https://doi.org/10.1037/0033-295x.109.3.472

Putnam, S. P., Sehic, E., French, B. F., Gartstein, M. A., & Lira Luttges, B. (2024). The Global Temperament Project: Parent-reported temperament in infants, toddlers, and children from 59 nations. *Developmental Psychology, 60*(5), 916–941. https://doi.org/10.1037/dev0001732

Raykov, T., & DiStefano, C. (2025). Evaluating Change in Adjusted R-Square and R-Square Indices: A Latent Variable

Method Application. *Educational and Psychological Measurement*. https://doi.org/10.1177/00131644251329178

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting Novel Associations in Large Data Sets. *Science, 334*(6062), 1518–1524. https://doi.org/10.1126/science.1205438

Richardson, M., Abraham, C., & Bond, R. (2022). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Schmidt, S. (2009). Shall we Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences. *Review of General Psychology, 13*(2), 90–100. https://doi.org/10.1037/a0015108

Scholtz, S. E., de Klerk, W., & de Beer, L. T. (2020). The Use of Research Methods in Psychological Research: A Systematised Review. *Frontiers in Research Metrics and Analytics, 5*. https://doi.org/10.3389/frma.2020.00001

Schönbrodt, F. D., & Wagenmakers, E.-J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*(2), 309–316. https://doi.org/10.1037/0033-2909.105.2.309

Simić, S. (2017). *Poređenje uloge kognitivne sposobnosti i emocionalne kompetencije pri prepoznavanju facijalnih ekspresija* [Master teza, Filozofski fakultet Univerziteta u Banjoj Luci].

Steering Committee of the Physicians' Health Study Research Group. (1989). Final report on the aspirin component of the ongoing physicians' health study. *New England Journal of Medicine, 321*(3), 129–135. https://doi.org/10.1056/NEJM198907203210301

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, 103*(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Torka, A.-K., Mazei, J., Bosco, F. A., Cortina, J. M., Götz, M., Kepes, S., O'Boyle, E. H., & Hüffmeier, J. (2023). How well are open science practices implemented in industrial and organizational psychology and management? *European Journal of Work and Organizational Psychology, 32*(4), 461–475. https://doi.org/10.1080/1359432x.2023.2206571

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology, 37*(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991

Tutnjević, S., & Lakić, S. (2017). Language-mediated object categorization: A longitudinal study with 16- to 20-month-old Serbian-speaking children. *European Journal of Developmental Psychology, 15*(5), 608–622. https://doi.org/10.1080/17405629.2017.1325733

Uher, J. (2018). Quantitative Data From Rating Scales: An Epistemological and Methodological Enquiry. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.02599

Uher, J. (2023). What's wrong with rating scales? Psychology's replication and confidence crisis cannot be solved without transparency in data generation. *Social and Personality Psychology Compass, 17*(5). Portico. https://doi.org/10.1111/spc3.12740

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wagenmakers & Metzke (2024). *Bayesian Inference from the Ground Up: The Theory of Common Sense*. https://www.bayesianspectacles.org/free-course-book/

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < .05." *The American Statistician,*

73(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology, 13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132*(1), 132–149. https://doi.org/10.1037/0033-2909.132.1.132

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology. *Perspectives on Psychological Science, 6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Youyou, W., Yang, Y., & Uzzi, B. (2023). A discipline-wide investigation of the replicability of Psychology papers over the past two decades. *Proceedings of the National Academy of Sciences, 120*(6). https://doi.org/10.1073/pnas.2208863120

CIP - Каталогизација у публикацији
Народна и универзитетска библиотека
Републике Српске, Бања Лука

159.9:311(075.8)(0.034.2)

ЛАКИЋ, Синиша, 1978-
Statistika za psihologe [Elektronski izvor] : konceptualni uvod u kvantitativno zaključivanje / Siniša Lakić. - Onlajn izd. - Ел. књига. - Banja Luka : Univerzitet u Banjoj Luci : Filozofski fakultet, 2025

Системски захтјеви нису наведени. - Način pristupa (URL): Način pristupa (URL): https://blupress.unibl.org/index.php/ff. - Насл. са насл. екрана. - Опис извора дана 30.10.2025. - Ел. публикација у ПДФ формату опсега 386. стр. - Библиографија: стр. 377-386.

ISBN 978-99997-45-02-4

COBISS.RS-ID 143427329