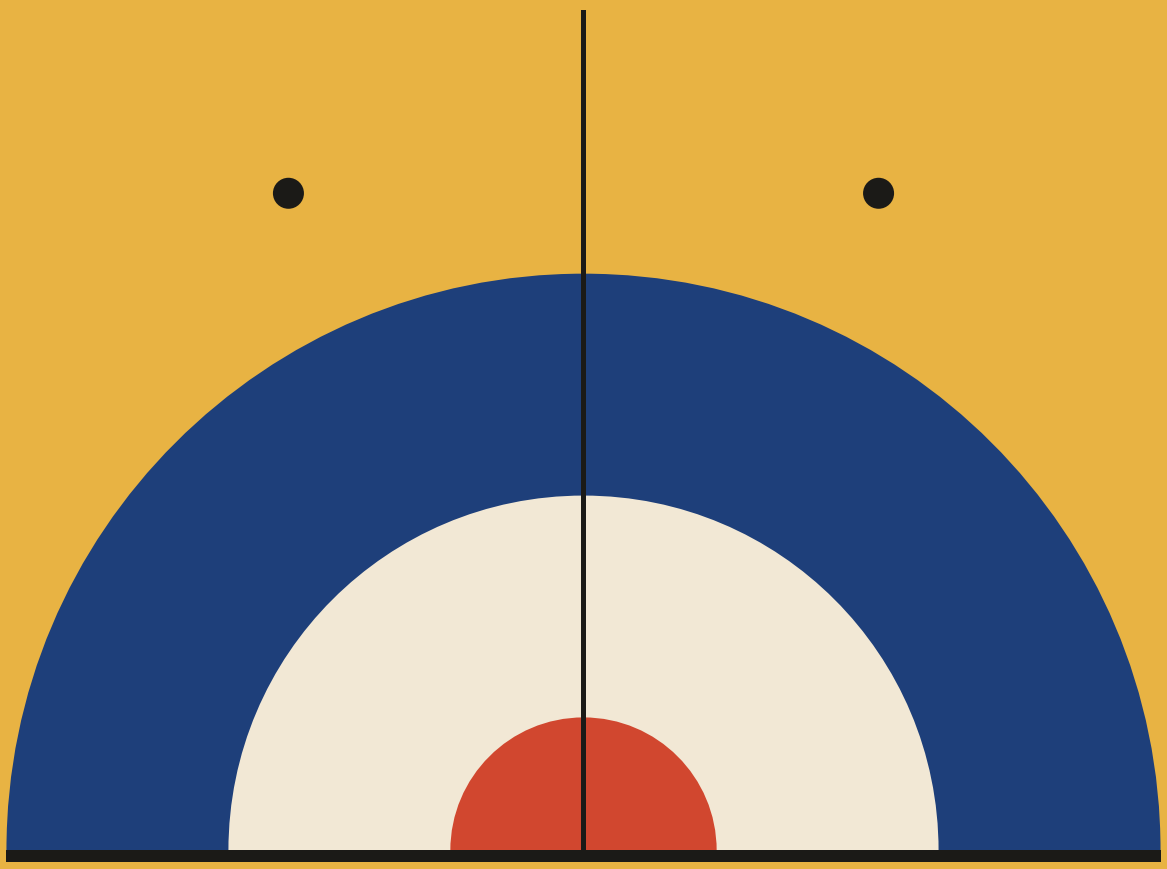
statistika

uvod u analizu podataka
za društvene znanosti

101

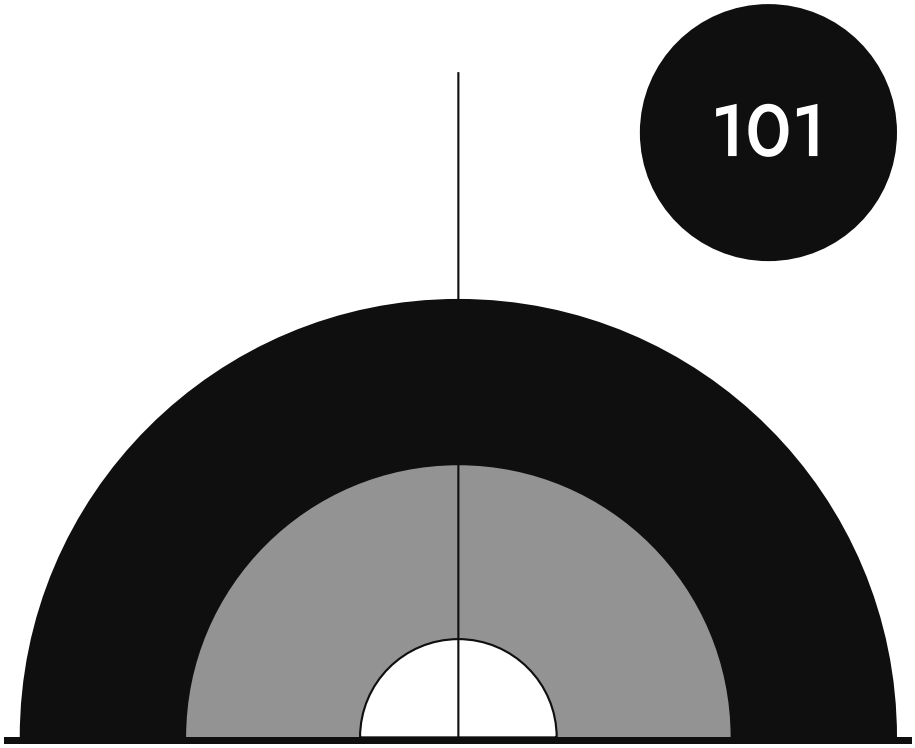


luka šikić · petra palić

UDŽBENIK

statistika 101

uvod u analizu podataka za društvene znanosti



Luka Šikić
Doc. dr. sc. · Sveučilišni odjel za komunikologiju

Petra Palić
Izv. prof. dr. sc. · Sveučilišni odjel za sociologiju

Hrvatsko katoličko sveučilište

28 listopada 2026

sadržaj

Predgovor 1
Put kroz knjigu 1
Čitanje i provjera 3

DIO I: STATISTIČKO MIŠLJENJE 5

Zašto statistika 7
Broj nije zaključak 9
Odluka koja se donosi bez dokaza 12
Životni ciklus podataka 13
Gdje intuicija popušta 16
Signal u promjenjivom svijetu 18
Razlika koju čini postupak 19
Zbirna slika i skrivena struktura 21
Simpsonov paradoks 25
Razrađeni primjer 28
Sažetak 29
Pojmovi 30
Zadaci 30

Mjerenje i istraživački dizajn 33
Od pojma do podatka 35
Prihvatljivost, isključenja i filtri 36
Razine mjerenja 38
Pouzdanost i valjanost 40
Treći čimbenik i logika eksperimenta 42
Opažačke studije i ankete 45
Doseg zaključka 47
Prikaz konfundirajuće varijable 48
Razrađeni primjer 51

Sažetak 56
Pojmovi 56
Zadaci 57

Kako brojke zavode **59**
Od točnoga broja do potpune tvrdnje 60
Nazivnik, os i rijedak ishod 63
Anketa, obuhvat i biranje trešanja 65
Istraživač margine pogreške 67
Podrijetlo ljudskog i strojnog broja 69
Razrađeni primjer 72
Protokol skeptičnoga čitanja 74
Sažetak 77
Pojmovi 78
Zadaci 78

DIO II: OPISIVANJE PODATAKA 81

Sažimanje podataka **83**
Od izvora do sažetka 84
Mjere raspršenosti 92
Standardna devijacija i kvartili 93
Oblik raspodjele i položaj opažanja 95
Oblikovanje distribucije 98
Razrađeni primjer 102
Sažetak 104
Pojmovi 105
Zadaci 105

Vizualizacija kao argument **109**
Gramatika grafike 111
Što oko može očitati 113
Prikaz prema tvrdnji 114
Dvije varijable u istom prostoru 117
Poštena ljestvica 118
Mala višestruka polja 120
Graf pred čitateljem 122
Isti podaci, četiri grafa 124
Razrađeni primjer 128
Sažetak 131
Pojmovi 132
Zadaci 132

Povezanost 135
Zajedničko kretanje . . . 137
Od standardizacije do matrice . . . 139
Matrica, kodovi i rangovi . . . 141
Ograničenje raspona i odabir . . . 143
Podskupine i obrat predznaka . . . 145
Država kao jedinica . . . 147
Pogodi korelaciju . . . 151
Razrađeni primjer . . . 155
Od odnosa do tvrdnje . . . 156
Sažetak . . . 161
Pojmovi . . . 161
Zadaci . . . 162

DIO III: OD UZORKA DO POPULACIJE 165

Vjerojatnost koliko treba 167
Što vjerojatnost opisuje . . . 169
Tri pravila i jedna pretpostavka . . . 171
Ponovljeni pokušaji s dva ishoda . . . 173
Simulator novčića i kampanje . . . 174
Zvonasta krivulja i njezino područje . . . 176
Kad podaci ne pristaju krivulji . . . 179
Nizovi koje slučajnost proizvodi . . . 181
Razrađeni primjer . . . 184
Sažetak . . . 186
Pojmovi . . . 186
Zadaci . . . 187

Uzorkovanje 189
Populacija, uzorak i pogreška uzorkovanja . . . 191
Standardna pogreška . . . 194
Oblik koji se pojavljuje . . . 196
CLT stroj . . . 198
Preciznost ankete od osamsto ljudi . . . 200
Nacrt uzorka i pristranost odabira . . . 202
Pokrivenost, odabir i odaziv . . . 204
Korpus kao uzorak . . . 206
Razrađeni primjer . . . 209
Sažetak . . . 211
Pojmovi . . . 212
Zadaci . . . 212

Procjena **215**
Od točke prema rasponu 217
Obećanje razine pouzdanosti 219
Hvatač intervala 220
Preciznost nasuprot pouzdanosti 223
Bootstrap kao vlastiti izum 226
Razrađeni primjer 231
Od procjene do tvrdnje 233
Sažetak 237
Pojmovi 237
Zadaci 238

DIO IV: ZAKLJUČIVANJE 241

Logika testiranja **243**
Sudnica i njezina asimetrija 245
Kako se gradi svijet bez učinka 245
Simulator p-vrijednosti 250
Dvije vrste pogreške 251
Prag je konvencija, a ne mjera 256
Što p-vrijednost nije 257
Drugo pitanje i drugi račun 258
Razrađeni primjer 260
Sažetak 262
Pojmovi 262
Zadaci 263

Veličina učinka i snaga **265**
Razlika koja nešto znači 267
Kad je značajno, a nije važno 269
Snaga kao svojstvo plana 271
Odabir značajnih nalaza pri niskoj snazi 273
Planiranje unatrag 275
Istraživač snage 276
Razrađeni primjer 280
Sažetak 282
Pojmovi 283
Zadaci 283

Kriza i obnova **285**
Od jedne tvrdnje do provjerljivoga pitanja 287
Analitička fleksibilnost prije modela 288
Replikacija kao kumulativno mjerenje 289

Cijeli zapis u forest plotu 290
Reforma redoslijeda i vidljivosti 292
Pješčanik p-hakiranja 294
Razrađeni primjer 297
Od testa do provjerljivog zaključka 300
Sažetak 303
Pojmovi 303
Zadaci 304

DIO V: MODELI 307

Kategorički podaci 309
Brojanje prije testiranja 311
Očekivano pod nezavisnošću 314
Očekivano i opaženo 315
Hi-kvadrat i njegova referentna raspodjela 317
Značajno nije isto što i snažno 319
Referentna raspodjela je istraživačka odluka 320
Kad je aproksimacija tanka 321
Razrađeni primjer 329
Sažetak 332
Pojmovi 333
Zadaci 333

Uspoređivanje dviju grupa 335
Tko nosi dva rezultata 337
Razlika prije oznake 338
Uzorkivač dviju grupa 340
Ista razlika, dva dizajna 342
Razlika skupina u linearnom modelu 344
Što razlika sama ne kaže 345
Pretpostavke i njihove granice 347
Razrađeni primjer 350
Sažetak 353
Pojmovi 353
Zadaci 354

Uspoređivanje više grupa 357
Cijena mnogih usporedbi 359
Varijanca između i unutar 361
Veličina i F-omjer 364
Isti model, više koeficijenata 366
Nakon ukupnog testa 368

Granice pretpostavki 371
Razrađeni primjer 374
Sažetak 375
Pojmovi 376
Zadaci 376

Regresija — opći okvir **379**
Pravac kao tvrdnja o prosjeku 381
Nagib koji nosi tuđu priču 384
Prilagodba i ono što je čini mogućom 386
Interakcija i regresijski pravac 390
Isti model iza ranijih poglavlja 393
Pristajanje i njegove granice 394
Objašnjenje i predviđanje 398
Granica prema uzroku 400
Razrađeni primjer 405
Sažetak 407
Pojmovi 408
Zadaci 408

Statistika u doba algoritama **411**
Odluka, korpus i jedinica 413
Nastanak oznake 414
Tri skupa, dva pitanja 415
Od vjerojatnosti do pravednosti 416
Istraživač pravednosti 418
Osporiva oznaka i postupak 420
Jezični modeli kao sustavi predikcije 421
Razrađeni primjer 424
Od modela do odluke 428
Sažetak 431
Pojmovi 432
Zadaci 432

ZAVRŠNICA 435

Vaše prvo istraživanje **437**
Pitanje prije podataka 439
Dizajn određuje granicu zaključka 441
Podaci i zapis odluka 442
Opis i graf prije modela 445
Model i promjena procjene 447
Zaključak koji uzorak podnosi 452

Provjera na stvarnim podacima 453
Rad s asistentom 458
Završna revizija paketa dokaza 462
Sažetak 466
Pojmovi 466
Zadaci 467

Putokaz, pojmovi i oznake 469
Putokaz kroz poglavlja 469
Primjeri koji se vraćaju 470
Hrvatsko-engleski rječnik 471
Dva skupa podataka za vježbu 482
Oznake u formulama 484

Provjere za samostalan rad 485
Naznake 485
Objašnjenja 491

Literatura 519

predgovor

Brojke susrećemo u vijestima, anketama i istraživanjima. Često uz njih dolazi i zaključak. Neka je pojava sve češća, jedna skupina prolazi bolje od druge ili neki postupak daje rezultate. Ova knjiga pomaže procijeniti takve tvrdnje. Uči nas provjeriti kako su podaci nastali, što pokazuju i dokle zaključak može ići.

Pisana je za studente društvenih znanosti. Za čitanje je dovoljno srednjoškolsko znanje matematike, a programiranje nije preduvjet. Cilj je razumjeti istraživanje, samostalno protumačiti njegove rezultate i prepoznati zaključak koji ide dalje od podataka. Knjiga pritom daje i osnovu za vlastitu analizu, od postavljanja pitanja do kratkog izvještaja.

Počinjemo jednostavnim pitanjima. Na koga se odnosi postotak i u odnosu na koji je broj izračunan? Jesu li skupine koje uspoređujemo doista usporedive? Što uzorak govori o populaciji i koliko je taj odgovor nesiguran? Račun pomaže odgovoriti, ali prvo trebamo znati što pitamo.

Put kroz knjigu

Prvi dio prati nastanak podataka i načine na koje brojke mogu zavesti. Zatim učimo sažeti podatke, odabrati graf i opisati povezanost. Na toj osnovi prelazimo na zaključivanje iz uzorka o populaciji.

Upoznajemo procjene, intervale i testove, uz pitanje kolika je razlika i što znači. Poglavlje o ponovljivosti pokazuje kako provjeravamo istraživačke rezultate. Slijede usporedbe skupina i regresija, a zatim predviđanje i društvene posljedice algoritama. U završnom poglavlju provodimo jedno istraživanje od pitanja do izvještaja.

Poglavlja se nadovezuju, pa ih vrijedi čitati redom. Novi pojam obično upoznajemo kroz primjer ili simulaciju, a tek zatim uvodimo naziv i potrebnu formulu.

Lakše je učiti novi postupak kad su nam podaci već poznati. Zato ćemo se kroz knjigu vraćati dvama skupovima podataka pripremljenima za vježbu. Možemo ih zamisliti kao dvije tablice u kojima svaki redak opisuje jednu osobu, a stupci njezina obilježja i odgovore. U prvoj, nazvanoj `populacija_medija`, upoznajemo medijske navike stanovnika zamišljenoga grada. Druga, `anketa_mreze`, sadrži odgovore zamišljenih ispitanika o društvenim mrežama. Obje smo tablice stvorili računalom, svaku zasebno. Takve podatke zovemo simuliranima. Znamo kako su nastali pa na njima možemo provjeriti što pojedini postupak otkriva, a što propušta. Rezultati tih vježbi odnose se na zamišljene osobe iz tablica.

Uz njih ćemo čitati i podatke iz stvarnih istraživanja i evidencija. Prijave na studij pomoći će nam razumjeti usporedbu skupina, izborni podaci značenje postotaka, a DigiKatovi podaci o medijskim objavama izbor dobrog sažetka i grafa. Susrest ćemo i Eurostatove pokazatelje, ponavljanja laboratorijskih pokusa te parlamentarne rečenice iz skupa ParlaSent. Uz svaki novi skup podataka najprije ćemo upoznati odakle dolazi, što predstavlja jedan redak i što iz njega možemo saznati. Kad mu se vratimo, nadovezat ćemo novo pitanje na ono što o tim podacima već znamo.

Čitanje i provjera

U ovom su izdanju prikazi statični. Slike i tablice donose odabrane rezultate simulacija, a tekst objašnjava što treba usporediti i što se iz usporedbe može zaključiti. Objašnjenja se mogu pratiti izravno uz tiskane prikaze. Razrađeni primjeri povezuju pitanje, podatke, račun i zaključak, a zadaci na kraju poglavlja omogućuju provjeru razumijevanja. Prije literature nalaze se odvojene naznake i objašnjenja odabranih zadataka. Naznaka pomaže započeti, a objašnjenje pokazuje račun, prihvatljiv zaključak i čestu stranputicu.

Potpune provjere rješenja[1], katalog podataka[2], R praktikum[3] i numerički podsjetnik[4] dostupni su u mrežnom izdanju. Adrese za pristup s papira navedene su uz poveznice. Za početni prolaz kroz razrađeni primjer dovoljno je pratiti pitanje, tablicu i tumačenje. Programski račun koji slijedi služi dodatnoj provjeri postupka.

Mrežni dodaci imaju stalne oznake. A je R praktikum, B put bez koda u jamoviju, C katalog podataka, D odabir postupka, E rječnik pojmova, F protokol rada s asistentom, a G numerički podsjetnik. Rubne oznake „Podsjetnik” vode na četiri kratka objašnjenja u dodatku G. Na kraju ove knjige tiskani rječnik povezuje hrvatske i engleske nazive s poglavljima, a popis oznaka pomaže pri ponovnom čitanju formula.

Kroz knjigu pratimo i rad s AI asistentom. On može predložiti postupak, izračunati rezultat ili objasniti nepoznat izraz. Njegov odgovor treba usporediti s podacima i izvorom. Osoba koja vodi analizu određuje pitanje, provjerava odgovara li mu postupak i preuzima odgovornost za zaključak. Primjeri pogrešnih odgovora služe vježbanju te provjere. Pisanje programskog koda ne traži se u zadacima.

[1] https://lusiki.github.io/statistika-knjiga/rjesenja.html
[2] https://lusiki.github.io/statistika-knjiga/podaci.html
[3] https://lusiki.github.io/statistika-knjiga/dodaci/a-praktikum.html
[4] https://lusiki.github.io/statistika-knjiga/dodaci/g-numericki-podsjetnik.html

Opseg je ograničen na temeljne postupke potrebne za čitanje istraživanja u društvenim znanostima. Vremenski nizovi, faktorska analiza, psihometrija i višerazinski modeli ostaju izvan njega. Bayesovsko zaključivanje predstavljamo kratkom usporedbom pristupa, a strojno učenje obrađujemo kroz njegove ideje i primjenu, bez matematičke izgradnje algoritama.

Nakon čitanja trebali biste moći protumačiti statistički rezultat, procijeniti koliko mu vjerujete i objasniti što još treba provjeriti. Te su vještine korisne i pri čitanju tuđeg istraživanja i pri prikazivanju vlastitih podataka.



DIO I

statističko mišljenje

Prije prvoga računa razjašnjavamo što broj predstavlja i koju tvrdnju može poduprijeti. Pratimo nastanak podataka, mjerenje i dizajn istraživanja, a zatim provjeravamo kako se točan broj može pretvoriti u pogrešan zaključak.

01

DIO I · STATISTIČKO MIŠLJENJE

zašto statistika

Zašto kontekst odlučuje što brojke znače

Podsjetnik.
Udjeli i stope

Jesu li poslijediplomski programi Sveučilišta Kalifornije u Berkeleyju diskriminirali žene pri upisu 1973. godine? Zbirne stope prijma muškaraca i žena toliko su se razlikovale da su potaknule sumnju u postupak upisa (Bickel i sur. 1975.).

Istraživači su zatim usporedili prijave unutar svakog odjela. Pokazalo se da su se žene češće prijavljivale na odjele koji su primali manji udio kandidata. Zbirna razlika zato nije izravno pokazivala kako su pojedini odjeli odlučivali o prijavama muškaraca i žena (Bickel i sur. 1975.).

Oba su izračuna točna i koriste iste prijave. Ukupne stope pokazuju veliku razliku, a stope po odjelima otkrivaju koliko na nju utječe raspored prijava. Koja usporedba odgovara na pitanje koje nas zanima?

Kad pročitamo da je jedna skupina uspješnija od druge, lako je preskočiti račun i odmah tražiti objašnjenje. Berkeleyjski slučaj poziva nas da zastanemo kod same usporedbe (Bickel i sur. 1975.). Zapisi su nastali u postupku prijave na studij i čuvaju stvarni sastav prijava po odjelima (Bickel i sur. 1975.). Prije odluke o tome što razlika znači trebamo znati koga uspoređujemo, što je zabilježeno i prema kojoj cjelini računamo postotak. To su pitanja od kojih ćemo krenuti, bez pretpostavke da već poznajemo statističke postupke.

U knjizi ćemo učiti pratiti put od istraživačkog pitanja do zaključka. Račun će nam pomoći opisati podatke, usporediti skupine i procijeniti nesigurnost, ali morat ćemo odlučiti odgovara li na pitanje koje smo postavili. Za početak su nam dovoljni poznati postoci, jednostavne tablice i spremnost da provjerimo što stoji iza broja. U središtu je sposobnost da objasnimo zašto smo nešto podijelili, koga smo izostavili i dokle dobiveni odgovor vrijedi. To možemo vježbati uz prikazane račune.

Uz stvarne prijave na studij koristit ćemo i skup podataka o medijskim navikama stanovnika zamišljenoga grada, nazvan `populacija_medija`. Stvorili smo ga računalnom simulacijom. Svaki redak njegove tablice opisuje jednu zamišljenu odraslu osobu. U prvom ćemo računu povezati njezin glavni izvor vijesti sa spremnošću na plaćanje vijesti. Usporedit ćemo broj osoba koje bi platile s njihovim udjelom unutar pojedinog izvora. Tako možemo pratiti svaki korak na populaciji čiji su odnosi unaprijed zadani.

Prijave na studij i spremnost na plaćanje vijesti različite su pojave, ali pred nas postavljaju sličan problem. Broj dobiva značenje tek kada znamo koje je ljude ili prijave obuhvatio i s čime ih uspoređujemo. Dok pratimo te usporedbe, razlikovat ćemo ono što podaci već pokazuju od onoga za što nam treba dodatno objašnjenje. Tu razliku tražimo i kada čitamo istraživanje, medijski naslov ili odgovor asistenta.

Broj nije zaključak

Zbirna stopa govori tko je češće primljen među svim prijavljenim kandidatima. Stope po odjelima govore kako su prolazili kandidati koji su se prijavili na isti odjel. Izbor usporedbe ovisi o pitanju koje istražujemo. Računom zatim provjeravamo rezultat te odabrane usporedbe.

Da bismo procijenili pravednost upisa, morali bismo poznavati i kriterije odlučivanja te okolnosti koje tablica ne bilježi. Zasad možemo provjeriti kako raspored prijava i veličina skupina mijenjaju značenje izračuna.

Statistika povezuje istraživačko pitanje s podacima i zaključkom koji iz njih možemo izvesti. Uz račun provjeravamo što je izmjereno i jesu li skupine koje uspoređujemo prikladne za naše pitanje. Pogreška u tom izboru može dovesti do pogrešnog zaključka i kada u računanju nema propusta.

Najprije utvrđujemo što predstavlja jedan zapis. U zbirnoj tablici prijava redak opisuje više prijava iste vrste. U tablici pojedinačnih odgovora redak može opisivati jednu osobu. Veza između broja redaka i broja ljudi, događaja ili neovisnih opažanja ovisi o načinu bilježenja podataka.

> **Jedinica analize** je ono o čemu donosimo zaključak, primjerice osoba, kućanstvo, organizacija ili objava. Jedan redak tablice ne mora predstavljati jednu neovisnu jedinicu.

Ako istu osobu pitamo za mišljenje svakoga od sedam dana, dobivamo sedam zapisa o jednoj osobi. Njezini odgovori mogu biti slični upravo zato što potječu od iste osobe. Pri zaključivanju zato tih sedam redaka promatramo kao ponovljena opažanja unutar iste jedinice.

U skupu `populacija_medija` imamo zapise za 50 000 zamišljenih odraslih stanovnika. Uz izvor vijesti i spremnost na plaćanje, za svaku su osobu zabilježeni dob, obrazovanje, povjerenje u medije

Podsjetnik. Postotci i postotni bodovi

na ljestvici od 1 do 10 i dnevno praćenje medija. Istom ćemo se populacijom služiti pri uzorkovanju, usporedbama skupina i regresiji, kako bismo nove postupke mogli pratiti na poznatom primjeru.

Računalno stvaranje tih zapisa slijedi zadana pravila. Najprije nastaju dob i izvor vijesti, čije se vjerojatnosti mijenjaju s dobi. Povjerenje ovisi o dobi i izvoru uz slučajno odstupanje, a spremnost na plaćanje o povjerenju i obrazovanju. Tako nastaju namjerno povezane varijable. Jedna osoba može odstupati od zadanoga obrasca, a cjelina od 50 000 osoba pokazuje obrazac koji je autor ugradio u generator. Ti zapisi ne opisuju stvarne stanovnike.

Prva tri retka pokazuju kako taj skup izgleda prije sažimanja. Svaki redak čuva više obilježja iste zamišljene osobe.

Tablica 1: Tri zapisa iz generirane populacije medijskih navika. Izrada autora.

Osoba	Dob	Izvor vijesti	Povjerenje (1–10)	Spremni platiti
1	59	portal	4	ne
2	50	radio	4	ne
3	45	tisak	6	da

Broj osobe služi povezivanju njezinih odgovora. Vrijednost u stupcu izvora vijesti jedna je kategorija, povjerenje je ocjena, a spremnost na plaćanje u ovoj je tablici prikazana odgovorom da ili ne. Grafom prikazujemo te zapise, a uzorkovanjem odabiremo dio njih. U oba postupka zadržavamo podatak o tome kako su nastali.

Iz takvih zapisa možemo prebrojiti osobe spremne platiti vijesti. Želimo li usporediti izvore s različitim brojem korisnika, trebamo i **nazivnik**, količinu s kojom dijelimo brojnik. Brojnik je broj slučajeva koji zadovoljavaju uvjet, ovdje osoba spremnih platiti vijesti. Među 4 855 osoba kojima je tisak glavni izvor vijesti njih 1 289 spremno je platiti vijesti. Među 15 101 korisnikom portala spremnih je 3 514. U većoj skupini takvih osoba ima više,

ali tek dijeljenjem možemo usporediti udjele. Jednako treba znati prema kojoj količini računamo stopu ili prosjek.

Provjeravamo i **podrijetlo podataka**. Treba znati tko ih je proizveo, kojim postupkom, koju inačicu koristimo i koja su njezina ograničenja. Za naš je nastavni skup poznat postupak stvaranja i njegovo sjeme, početna postavka koja omogućuje ponavljanje iste simulacije. Zato možemo provjeriti račun i tumačenje. Za tvrdnju o hrvatskoj publici trebali bismo podatke prikupljene od stvarnih ljudi.

Poznavanje podrijetla podataka pomaže nam procijeniti i što možemo doznati iz osobnog iskustva. Anegdota može upozoriti na problem, ponuditi moguće objašnjenje ili opisati posljedicu koja se iz tablice ne vidi. Sama ipak ne pokazuje koliko je pojava raširena ni bi li se ponovila u drugim okolnostima. Za to trebamo više opažanja, uz jasnu odluku o tome koga promatramo, što mjerimo i što uspoređujemo.

Iskaz ispitanice o tome zašto je prestala čitati vijesti, primjerice, otkriva okolnosti i redoslijed događaja koji bi mogli objasniti njezinu odluku. Taj iskaz upućuje na ono što treba provjeriti. Uređen skup podataka omogućuje da zatim ispitamo kako se pojava raspoređuje među promatranim ljudima. Želimo li zaključak proširiti na druge ljude, moramo opravdati i način odabira sudionika. Iz jednog svjedočenja ne možemo odrediti raširenost pojave, kao što iz prosjeka ne možemo iščitati iskustvo svake osobe.

Za takve odluke potrebno je poznavanje područja. Teorija i okolnosti slučaja pomažu nam odabrati usporedive skupine, predložiti objašnjenje i prosuditi koliko bi velika razlika bila važna. Izračun slijedi te odluke.

Dobar statistički postupak ostavlja jasan trag od opažanja do tvrdnje. Druga osoba mora moći pregledati te korake, ponoviti račun i provjeriti opravdavaju li podaci zaključak.

Odluka koja se donosi bez dokaza

Slične odluke donosimo i izvan istraživanja. Zamislimo portal kojem pada posjećenost. Urednik smatra da su naslovi dosadni i predlaže drukčiji stil. Možda je u pravu. Posjećenost se, međutim, mogla promijeniti i zbog vremena objave, duljine tekstova, nove konkurentske aplikacije ili godišnjeg doba u kojem publika ima druge navike.

Bez provjere tih objašnjenja lako prevagne mišljenje najupornijeg ili najutjecajnijeg sugovornika. Isto pitanje nastaje kada ministarstvo procjenjuje učinak mjere, udruga doseg kampanje ili stranački stožer promjenu u anketi. Statistički postupak omogućuje da svako objašnjenje usporedimo s podacima. Odgovor ponekad ostane neodređen, ali postaje jasnije što znamo i što bi još trebalo provjeriti.

Sociologija, politologija, psihologija, komunikologija i ekonomija empirijske su discipline. Njihove tvrdnje provjeravamo opažanjem. Tvrdnju da povjerenje u medije pada treba povezati s mjerenjima povjerenja, a tvrdnju da negativne vijesti privlače više pozornosti s odgovarajućom usporedbom. Isto vrijedi za moguću povezanost dohotka i sudjelovanja. U svakom slučaju moramo objasniti kako podaci podupiru upravo izrečenu tvrdnju.

Podacima se pritom služe i ljudi izvan istraživačkih ustanova. Odluke o sadržaju, oglasima i dosegu temelje se i na zapisima o klikovima, otvaranju poruka i vremenu gledanja. Razumijevanje tih zapisa omogućuje nam da sudjelujemo u raspravi o odlukama koje se na njima temelje.

Prije primjene formule zato procjenjujemo odgovara li postupak pitanju. Treba provjeriti jesu li skupine usporedive, što je doista izmjereno i koliko je rezultat nesiguran. Tek tada možemo reći koju odluku analiza podupire i dokle njezin zaključak vrijedi.

Izračun ne može nadoknaditi sve što nedostaje u podacima. Istraživač i dalje mora prosuditi koja je usporedba važna i jesu li mje-

renja dovoljna. Statistika pomaže izraziti nesigurnost, pa zaključak često mora biti oprezniji od početne pretpostavke. To omogućuje da nalaz upotrijebimo uz razumijevanje njegovih ograničenja.

Životni ciklus podataka

Da bismo provjerili urednikovo objašnjenje pada posjećenosti, morali bismo odabrati zapise, pregledati ih i tek potom usporediti naslove. I druge istraživačke tvrdnje nastaju kroz takav niz povezanih koraka, koji nazivamo **životnim ciklusom podataka**. On obuhvaća pribavljanje i analizu podataka, tumačenje rezultata te praćenje njihove uporabe nakon objave. Tablica pokazuje koje nas pitanje vodi u svakom koraku.

Tablica 2: Životni ciklus podataka i odluke koje vodi. Izrada autora.

Faza	Pitanje koje je vodi
pitanje	Koju odluku ili tvrdnju treba razjasniti?
pribavljanje	Odakle zapisi dolaze i tko može izostati?
provjera	Predstavlja li redak očekivanu jedinicu i jesu li vrijednosti moguće?
priprema	Koje su promjene napravljene prije analize?
istraživanje	Koji su obrasci, iznimke i alternativna objašnjenja vidljivi?
modeliranje	Koja usporedba ili model odgovara pitanju?
vrednovanje	Koliko je rezultat stabilan i gdje ne uspijeva?
komunikacija	Koju tvrdnju rezultat doista podupire?
praćenje	Mijenjaju li se podaci, uporaba ili posljedice nakon objave?

Tijekom analize često se vraćamo na ranije korake. Provjera može pokazati da smo preuzeli pogrešne zapise, a prvi pregled podataka da pitanje treba suziti. Takve promjene bilježimo kako bi se poslije moglo razumjeti zašto koristimo određenu datoteku ili odgovaramo na izmijenjeno pitanje.

Već pri pribavljanju podataka utvrđujemo čiji su zapisi uključeni, što predstavljaju i koji nazivnik trebamo. Nepoznato podrijetlo, dvostruko zabilježene jedinice ili mjerenje pogrešnog pojma treba razriješiti prije modeliranja. Složeniji model sam po sebi ne rješava te probleme.

Četiri djelatnosti i četiri pitanja

Statistika, podatkovna znanost, strojno učenje i sustavi umjetne inteligencije sudjeluju u različitim koracima tog postupka. Njihove se zadaće preklapaju, ali svaka od tih djelatnosti naglašava drugo pitanje.

Model ovdje znači pojednostavljen prikaz pravilnosti u podacima. Na primjer, može povezati dob osobe s njezinim očekivanim povjerenjem. Prikaz pomaže opisivanju ili predviđanju, ali sam ne pokazuje što uzrokuje odnos.

Tablica 3: Četiri djelatnosti prema pitanju koje postavljaju. Izrada autora.

Djelatnost	Vodeće pitanje
statistika	Što ovi podaci opravdavaju vjerovati?
podatkovna znanost	Kako stvarni izvori podataka mogu postati pouzdana i reproducibilna analiza?
strojno učenje	Hoće li naučeni obrazac raditi na doista novim opažanjima?
sustav umjetne inteligencije	Mogu li se naučeni obrasci upotrijebiti za generiranje, preporučivanje, klasificiranje ili djelovanje u institucionalnom okruženju?

Podatkovna znanost obuhvaća rad kojim od izvora podataka dolazimo do analize koju možemo ponoviti. **Strojno učenje** ispituje koliko dobro obrazac naučen iz jednih podataka djeluje na novim opažanjima. Takav pristup stavlja uspješnost predviđanja u prvi plan, dok je drugi pristup modeliranju više usmjeren na objašnjenje odnosa (Breiman 2001.). **Sustav umjetne inteligencije** uključuje model, sučelje i način njegove uporabe u određenoj ustanovi ili postupku odlučivanja. Pri procjeni

takvog sustava razmatramo tehnička svojstva, ali i pogreške te posljedice uporabe (Barocas i sur. 2023.).

Suvremena umjetna inteligencija ovisi o više vrsta rada. Statističko učenje i algoritmi strojnog učenja povezuju se s digitalnim podacima, optimizacijom i računalnom opremom. Ljudi označuju podatke, grade programsku infrastrukturu i provjeravaju rezultate. Podatkovna znanost organizira mnoge od tih koraka, pa je razlikujemo od samog strojnog učenja.

Svaki dio postupka treba zasebno provjeriti. Analizu možemo uspješno ponoviti, a da je njezin zaključak i dalje pogrešan. Model može dobro predviđati, a da ne objašnjava uzrok. Jednako tako, dobar statistički račun ostaje upitan ako ne znamo podrijetlo podataka, ako je ishod nehotice dospio među podatke za učenje ili ako nitko ne može ponoviti pripremu podataka. Ako među prediktorima već stoji odgovor koji navodno predviđamo, prividni uspjeh ne govori kako bismo predviđali novu osobu. Tom se problemu vraćamo u poglavljima o regresiji i algoritmima.

Gdje intuicija popušta

Zaključak se katkad čini toliko očitim da provjera izgleda suvišna. Ipak, pri procjeni slučajnosti i promjenjivosti lako uočimo obrazac i ondje gdje ga nema.

Uzmimo dva niza od 10 bacanja poštenog novčića. U prvom se pismo i glava pravilno izmjenjuju. U drugom nema pravilne izmjene, a pismo se pojavljuje i triput zaredom. Drugi niz može izgledati prirodnije, ali svaki od 1024 moguća niza ima jednaku vjerojatnost. Dojam nastaje zato što očekujemo da slučajan niz izgleda na određen način. Slučajnost može proizvesti i pravilnu izmjenu i nekoliko jednakih ishoda zaredom.

Slično zaključujemo iz poznatih primjera. Ako susjed i dvoje kolega hvale novu aplikaciju za vijesti, mogli bismo pomisliti da je ona općenito dobro prihvaćena. Troje poznanika ipak ne predstavlja cijelu publiku. Možda se vijestima bave više od drugih

pa su zato i isprobali aplikaciju. Sklonost da učestalost procjenjujemo prema tome koliko se lako prisjetimo nekog primjera poznata je kao heuristika dostupnosti (Tversky i Kahneman 1973.). Živopisan ili nedavan slučaj tako može ostaviti pretjerano snažan dojam.

Zamislimo i usporedbu u kojoj portali s više objavljenih tekstova imaju više posjeta. Iz nje se lako izvodi preporuka da treba više objavljivati. Veći portali, međutim, mogu imati i više novinara, veći proračun te drukčiju, stariju publiku. Treba razmotriti utječe li broj tekstova na posjete, potiče li veća posjećenost objavljivanje ili oboje ovisi o drugim obilježjima portala. Postupke kojima provjeravamo takva objašnjenja razradit ćemo u poglavljima o istraživačkom dizajnu i regresiji.

Za svaki od tih zaključaka trebamo dodatnu usporedbu. Niz bacanja uspoređujemo s ostalim mogućim nizovima, pohvale aplikacije s iskustvima šire publike, a posjete portalima uzimajući u obzir njihovu veličinu. Time postaje vidljivo na čemu se zaključak temelji.

Provjera obuhvata potrebna je i pri velikom broju zapisa. Milijuni interakcija na portalu opisuju zabilježene aktivnosti njegovih posjetitelja. Za upoznavanje ostale publike trebalo bi prikupiti dodatne podatke. Odgovori većega broja zasebno odabranih ljudi obično daju stabilniji postotak od odgovora male skupine. Korist je manja kada se isti odgovori ponavljaju, a sustavno izostavljen dio publike ostaje izvan obuhvata.

Unatoč tim ograničenjima možemo donositi korisne zaključke. Statistički postupci, kada su njihove pretpostavke opravdane, omogućuju da izračunamo dio nesigurnosti. Mjerenje, odabir sudionika i promjenu okolnosti moramo zasebno procijeniti. Tako možemo obrazložiti što zaključujemo i zašto zaključak ograničavamo na određene uvjete.

Signal u promjenjivom svijetu

I unutar dobro odabrane usporedbe opažanja se međusobno razlikuju. Dvije osobe mogu različito reagirati na istu poruku. Ista anketa na drugom uzorku dat će nešto drukčije postotke, a ponašanje jedne osobe može se mijenjati s vremenom. Statistika nam pomaže proučavati društvene pojave uz tu promjenjivost.

Pritom pokušavamo razlikovati **signal**, obrazac koji nas zanima, od **šuma**, ostalih izvora varijacije. Signal može biti razlika među skupinama ili povezanost dviju pojava. Šum otežava da taj obrazac razaberemo iz pojedinačnih opažanja. Što smatramo signalom ovisi o pitanju koje postavljamo.

Kad uspoređujemo povjerenje u institucije među dobnim skupinama, razlike među ljudima iste dobi mogu otežati uočavanje dobnog obrasca. Ako zatim počnemo istraživati zašto se ljudi iste dobi razlikuju, te razlike postaju ono što želimo objasniti. Isti podaci tako mogu poslužiti različitim pitanjima.

I usporedba političkog sudjelovanja može dati različite rezultate ovisno o tome razmatramo li dob, obrazovanje, dohodak ili mjesto stanovanja. Svaki od tih izbora izdvaja drugi dio promjenjivosti. Zbog toga se nalazi o istoj pojavi mogu razlikovati i kada su svi računi točni. Pri čitanju istraživanja treba provjeriti koja su obilježja uključena u usporedbu.

Povezanost korištenja društvenih mreža i praćenja vijesti procjenjujemo usporedbom većeg broja studenata. Navike pojedinog studenta mogu se uklapati u zajednički obrazac ili odstupati od njega. Osoba je jedinica analize, a obrazac opisuje odnos u promatranoj skupini.

Važni su i broj opažanja i način na koji su nastala. Uz isti postupak prikupljanja i dovoljno neovisne jedinice, razlika među desetak ljudi lakše može nastati slučajnim rasipanjem nego jednaka razlika među tisućama. Ponavljanje istih zapisa ili dodavanje snažno povezanih opažanja donosi manje nove informacije nego uključivanje dodatnih neovisnih jedinica.

Za tumačenje rasta, pada ili razlike trebamo znati s čime uspoređujemo rezultat. To može biti druga skupina, ranije razdoblje ili raspon vrijednosti koje bi nastale slučajnom promjenjivošću. Najprije treba jasno imenovati obje strane usporedbe.

Razlika koju čini postupak

Imenovanje obiju strana usporedbe pomaže nam odrediti i kakav zaključak tražimo. Opis govori što smo zabilježili, a povezanost kako se pojave mijenjaju zajedno. Generalizacijom nalaz prenosimo na jedinice koje nismo promatrali. Predviđanjem procjenjujemo nove ishode, dok uzročna tvrdnja govori što bi određena promjena proizvela. Odluka uz nalaz uključuje i procjenu mogućih postupaka te njihovih posljedica. Prijelaz od opisa podataka do uzročnog zaključka traži dodatno obrazloženje načina na koji je usporedba nastala.

Za isti skup podataka možemo postaviti različita pitanja. Sljedećim zamišljenim formulacijama ilustriramo vrste tvrdnji.

Tablica 4: Šest vrsta tvrdnji i pitanja koja ih razlikuju. Izrada autora.

Vrsta tvrdnje	Primjer pitanja
opis	Koliki udio zabilježenih osoba koristi portal?
povezanost	Razlikuje li se povjerenje među dobnim skupinama?
generalizacija	Što odgovori iz uzorka govore o svim odraslim stanovnicima?
predviđanje	Hoće li nova osoba biti spremna platiti vijesti?
uzročnost	Bi li promjena cijene promijenila spremnost iste publike na plaćanje?
odluka	Isplati li se promijeniti cijenu uz očekivanu korist i trošak pogreške?

Pri procjeni zaključaka koristimo **šest revizijskih pitanja**. Pitanja ostaju ista i u kasnijim poglavljima, a odgovori ovise o podacima i postupku.

Tablica 5: Šest revizijskih pitanja. Izrada autora.

Revizijsko pitanje	Što treba biti vidljivo
Koja je jedinica?	Što predstavlja zapis i na kojoj se razini izriče zaključak?
Tko ili što nedostaje?	Kako su jedinice ušle u podatke i tko je mogao izostati?
Koji je cilj tvrdnje?	Je li riječ o opisu, povezanosti, generalizaciji, predviđanju, uzročnosti ili odluci?
Gdje je neizvjesnost?	Dolazi li iz mjerenja, uzorkovanja, modela ili promjene okolnosti?
Koje drugo objašnjenje ostaje?	Koji bi još proces mogao proizvesti isti obrazac?
Tko snosi posljedice?	Tko dobiva korist, tko pogrešku i postoji li način osporavanja?

U Berkeleyjskim podacima promatramo prijave iz određenog razdoblja. Nemamo cijeli obrazovni put kandidata ni podatke o procesima koji su prethodili prijavi (Bickel i sur. 1975.). Iz stopa možemo opisati ishode tih prijava. Za motive ljudi koji su odlučivali ili posljedice drukčijih kriterija trebali bismo dodatne podatke.

Odgovori na šest pitanja pomažu odrediti što podaci podupiru i koja pretpostavka traži najviše provjere. Tako drugi mogu razumjeti na čemu zaključak počiva, provjeriti ga i uočiti moguću pogrešku.

Zbirna slika i skrivena struktura

Vratimo se sada razlici između ukupne stope prijma i stopa po odjelima. Ukupni prosjek ili stopa ovisi o rezultatima podsku-

pina i o tome koliko je svaka podskupina zastupljena. Brojnija podskupina više utječe na ukupni rezultat. Zato se zbirna stopa može promijeniti samo zbog drukčijeg sastava, čak i ako stope unutar podskupina ostanu jednake.

U Berkeleyju su se odjeli razlikovali po tome koliko je teško bilo upisati ih. Žene su češće birale selektivnije odjele, pa su ti odjeli više utjecali na njihovu ukupnu stopu prijma. Usporedba unutar pojedinih odjela uklonila je taj utjecaj rasporeda prijava. Razlika vidljiva u zbirnim podacima tada se smanjila ili preokrenula (Bickel i sur. 1975.).

Brojnosti pokazuju oba pogleda na Berkeley. Za cijelo sveučilište bilo je 8 442 prijave muškaraca, od kojih je prihvaćeno 3 738, te 4 321 prijava žena, od kojih je prihvaćeno 1 494 (Bickel i sur. 1975.). To daje 44,3 % i 34,6 %. Sljedeća tablica prikazuje šest najvećih odjela iz istoga slučaja, sačuvanih u skupu `UCBAdmissions`. Njihovih 4 526 prijava čini podskup, pa njegov zbirni postotak nije postotak cijeloga sveučilišta (Bickel i sur. 1975.).

Tablica 6: Prijave i prijam u šest najvećih odjela Berkeleyja 1973. Podaci UCBAdmissions prema Bickel i sur. (1975.). Postoci izračunati iz prikazanih brojnosti.

Odjel	Skupina	Prijave	Primljeni	Primljeni, %
A	Muškarci	825	512	62,1
A	Žene	108	89	82,4
B	Muškarci	560	353	63,0
B	Žene	25	17	68,0
C	Muškarci	325	120	36,9
C	Žene	593	202	34,1
D	Muškarci	417	138	33,1
D	Žene	375	131	34,9
E	Muškarci	191	53	27,7
E	Žene	393	94	23,9
F	Muškarci	373	22	5,9
F	Žene	341	24	7,0
Ukupno šest	Muškarci	2691	1198	44,5
Ukupno šest	Žene	1835	557	30,4

U odjele A i B, koji imaju visoke stope prijma, prijavilo se 1385 od 2691 muškaraca iz ovoga podskupa, a samo 133 od 1835 žena (Bickel i sur. 1975.). Razlikuju se i stope unutar odjela i udjeli prijava usmjerenih u pojedine odjele. Zbirni postotak uključuje oboje. Tablica sama ne objašnjava zašto je izbor odjela bio tako raspoređen.

> **Simpsonov paradoks** je obrazac u kojem se smjer povezanosti u združenim podacima promijeni ili preokrene kada se podaci razdvoje prema relevantnoj trećoj varijabli (Simpson 1951.).

Sve izračunate stope pritom mogu biti točne. Zaključak se mijenja zato što zbirna usporedba i usporedba po podskupinama odgovaraju na različita pitanja.

To možemo vidjeti na konstruiranom primjeru dvaju portala. Brojevi služe prikazu postupka i nisu podaci o stvarnim portalima. Portal A objavljuje 100 videozapisa s prosječno 200 reakcija i 10 tekstova s prosječno 50 reakcija. Portal B objavljuje 10 videozapisa s prosjekom od 220 reakcija i 100 tekstova s prosjekom od 60. Za A zato najprije računamo ukupne reakcije, $100 \times 200 + 10 \times 50 = 20\,500$, pa dijelimo sa 110 objava. Za B račun glasi $(10 \times 220 + 100 \times 60)/110$. Pri računanju svaki prosjek važemo brojem objava pripadajućeg formata. B postiže bolji rezultat u svakom formatu, ali njegov ukupni prosjek iznosi približno 75, naspram približno 186 reakcija po objavi na portalu A.

Ukupni prosjek računamo preko svih objava, pa format s više objava ima veću težinu. Videozapisi čine 91 % objava portala A i samo 9 % objava portala B. A zato ima veći udio formata koji privlači više reakcija. Viši ukupni prosjek ovdje proizlazi iz sastava objava, dok usporedbu uspješnosti po formatu čitamo iz zasebnih prosjeka.

Podjela na podskupine mora odgovarati istraživačkom pitanju. Ako podatke dijelimo na mnogo proizvoljnih načina, možemo naići na uvjerljiv obrazac koji se u drugim podacima neće ponoviti. Zato prije tumačenja obrazlažemo zašto smo odabrali baš tu podjelu.

Izbor prikaza ovisi o tvrdnji koju provjeravamo, načinu mjerenja, dizajnu i podrijetlu podataka. Ako nas zanima odlučivanje unutar odjela, uspoređujemo prijave unutar odjela. Ako želimo opisati ukupne ishode svih prijava, koristimo zbirnu stopu. Usporedba mora odgovarati pitanju.

Pri tumačenju Simpsonova paradoksa zato trebamo znati kako su podaci sastavljeni. Zaključak ograničavamo na usporedbu koju smo napravili.

Simpsonov paradoks

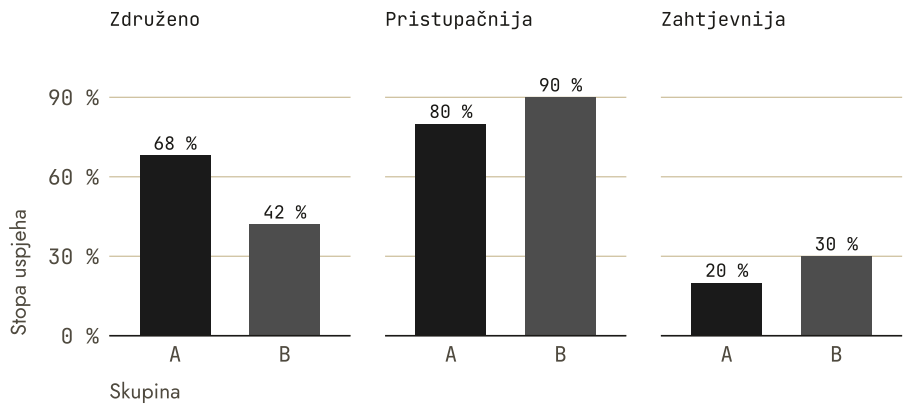
Za sljedeći prikaz namjerno zadajemo stope, bez izvlačenja slučajnih ishoda. U pristupačnijoj podskupini uspjeh je 80 % za A i 90 % za B, a u zahtjevnijoj 20 % i 30 %. U početnoj raspodjeli 80 % skupine A i 20 % skupine B pripada pristupačnijoj podskupini. Ukupno za A dobivamo $0{,}80 \times 0{,}80 + 0{,}20 \times 0{,}20 = 0{,}68$, a za B $0{,}20 \times 0{,}90 + 0{,}80 \times 0{,}30 = 0{,}42$. Promjenom rasporeda mijenjamo težine, dok četiri stope unutar podskupina ostaju zadane.

Zastupljenost podskupina određuje koliko koja stopa pridonosi zbirnom rezultatu. U tiskanom primjeru A ima četiri petine jedinica u pristupačnijoj podskupini, a B samo jednu petinu. To je mehanizam obrata, iako B postiže višu stopu u objema podskupinama.

Tablica 7: Sastav skupina u konstruiranom tiskanom primjeru. Izrada autora.

Skupina	Pristupačnija podskupina	Zahtjevnija podskupina
A	80 %	20 %
B	20 %	80 %

Sljedeći prikaz uspoređuje zbirne stope s usporedbama unutar podskupina. Promatramo kako njihov različit sastav mijenja smjer zbirne razlike.



Slika 1: Združeno prednjači A, a unutar obiju podskupina B. Različit sastav skupina iz prateće tablice objašnjava obrat. Izrada autora iz konstruiranoga primjera.

U združenom prikazu uspješnija je skupina A. Unutar svake podskupine višu stopu ima skupina B. Razlika nastaje zbog zastupljenosti skupina u pristupačnijoj i zahtjevnijoj podskupini.

Kada su skupine jednako raspoređene po podskupinama, zbirna usporedba zadržava smjer razlike unutar podskupina. Za obrat je potreban drukčiji sastav skupina. Razdvajanjem podataka možemo utvrditi koliko ukupnoj razlici pridonose stope uspjeha, a koliko zastupljenost podskupina.

● STATISTIKA U DIVLJINI

Zbirna stopa i sumnja u pristranost. Iza Berkeleyjske tablice stajalo je pitanje o postupanju prema kandidatima, a ne samo zadatak računanja postotaka. Niža ukupna stopa prijma žena potaknula je sumnju u upisni postupak (Bickel i sur. 1975.). Istraživanje se zato usmjerilo na pitanje kako su se prijave i odluke raspoređivale po odjelima, jer je svaki odjel zasebno odlučivao o prijmu (Bickel i sur. 1975.).

Vratimo se toj sumnji nakon prikaza obrata. Kada bismo sastavljali izvještaj za nekoga tko treba provjeriti upisni postupak, zbirna stopa bila bi važan početak. Pokazala bi razliku u ishodima svih zabilježenih prijava. No iz nje još ne bismo mogli utvrditi u kojem dijelu postupka nastaje razlika.

Za to trebamo uz svaku stopu vidjeti koliko je prijava upućeno pojedinom odjelu i koliko ih je prihvaćeno.

U izvornom slučaju žene su se češće prijavljivale na odjele koji su primali manji udio kandidata obaju spolova (Bickel i sur. 1975.). Zato ukupni postotak istodobno opisuje postupanje odjela i raspored prijava među njima. Razdvajanjem tih sastavnica mijenja se ono što možemo tvrditi o radu upisnih povjerenstava. Otvorenima ostaju pitanja zašto su prijave bile tako raspoređene i jesu li kandidati imali jednake mogućnosti prije prijave.

Dobar bi izvještaj zato zadržao oba pogleda. Zbirne stope pokazuju ukupnu razliku u ishodima, a odjelne usporedbe približavaju nas odluci koja nas zanima. Kad bismo prikazali samo prvi pogled, mogli bismo razliku neopravdano pripisati povjerenstvima. Kad bismo prikazali samo drugi, mogli bismo prerano zatvoriti pitanje nejednakosti.

Izbor usporedbe ovdje određuje što treba dalje istražiti. Prije nego što stopu uspjeha pretvorimo u ocjenu ustanove, trebamo znati koje su odluke ušle u njezin brojnik i kako su ljudi dospjeli u njezin nazivnik.

U našoj generiranoj populaciji isti se problem javlja pri usporedbi broja osoba spremnih platiti vijesti i njihova udjela unutar izvora.

NAĐITE GREŠKU

U usporedbi izvora vijesti zanima nas koliko je njihovih korisnika spremno platiti vijesti. Sljedeći izvadak iz poznate simulacije daje brojnosti potrebne za provjeru asistentove bilješke. Svaka zamišljena osoba pripada samo jednom retku.

Tablica 8: Spremnost na plaćanje u dvjema odvojenim skupinama izvora. Izrada autora.

Glavni izvor	Osobe	Spremni platiti
TV i radio zajedno	16 666	4 301
Tisak	4 855	1 289

> Među korisnicima tiska spremno je platiti 1 289 od 4 855 osoba, odnosno 26,55 %. Televizija i radio zajedno imaju 4 301 takvih osoba, više nego tisak. Prema tome, televizija i radio zajedno imaju veći udio osoba spremnih platiti.

Dopunite izvadak udjelom za televiziju i radio zajedno. Potom napišite dvije rečenice za izvještaj, jednu o usporedbi brojeva i drugu o usporedbi udjela. Označite koju rečenicu bilješke treba zamijeniti i objasnite zašto točan račun za tisak nije dovoljan za njezin zaključak.

Razrađeni primjer

Razliku između broja i udjela možemo provjeriti za sve izvore vijesti u istom simuliranom skupu podataka `populacija_medija`. Za svaki izvor pitamo koliki je udio njegovih korisnika spreman platiti vijesti. Jedinica analize ostaje zamišljena osoba, a veličina skupine daje nazivnik. Time povezujemo pitanje s računom koji na njega odgovara. Rezultat će opisivati tu generiranu populaciju, bez objašnjenja uzroka razlika ili zaključka o stvarnim ljudima.

Za svaki izvor tablica sadržava broj osoba, broj onih koji bi platili vijesti i njihov udio. Stupac `Osobe` daje nazivnik, a `Spremni platiti` brojnik. Redoslijed izvora određen je udjelom.

Tablica 9: Spremnost platiti prema glavnom izvoru vijesti u generiranoj populaciji. Izrada autora.

Izvor	Osobe	Spremni platiti	Udio
tisak	4 855	1 289	26,55 %
radio	5 839	1 528	26,17 %
TV	10 827	2 773	25,61 %
portal	15 101	3 514	23,27 %
društvene mreže	13 378	2 841	21,24 %

Među osobama kojima je portal glavni izvor vijesti njih 3 514 spremno je platiti vijesti, što je najveći broj među skupinama. No ta je skupina i najveća. Dijeljenjem s 15 101 dobivamo udio od 23,27 %. Za tisak dijelimo 1 289 s 4 855 i dobivamo 26,55 %. Portal ima više osoba spremnih platiti, a tisak njihov veći udio.

Pri tumačenju polazimo od cijele generirane populacije od 50 000 zapisa, pa udjele poznajemo točno za taj skup. Njihov se doseg završava opisom simulacije. Primjena na stvarnu publiku zahtijevala bi stvarne podatke i provjeru prikladnosti postupka, a objašnjenje uzroka odgovarajući istraživački dizajn. Nastavni primjer omogućuje da razmotrimo i posljedice tumačenja. U stvarnoj bi analizi zamjena broja udjelom mogla usmjeriti resurse prema većoj skupini samo zbog njezine veličine.

Zaključak stoga glasi da u generiranoj populaciji tisak ima najveći udio osoba spremnih platiti vijesti, a portal najveći broj takvih osoba. Pri usporedbi udjela uzimamo u obzir veličinu svake skupine. Pri usporedbi brojeva promatramo samo brojnike.

Sažetak

Za razumijevanje broja trebamo znati odakle podaci dolaze, što predstavlja jedinica analize i prema čemu računamo udio ili us-

poredbu. Te odluke pratimo od postavljanja pitanja do uporabe rezultata. Šest revizijskih pitanja pomaže prepoznati kada zaključak prelazi ono što smo doista provjerili. Simpsonov paradoks pokazuje kako dvije točne usporedbe mogu dati različite odgovore jer se odnose na različita pitanja. Statistika, podatkovna znanost, strojno učenje i sustavi umjetne inteligencije pridonose tom radu različitim postupcima. Koliko će zaključak biti koristan ovisi već o izboru onoga što mjerimo i o istraživačkom dizajnu.

Pojmovi

zbirna stopa (*aggregate rate*), signal (*signal*), šum (*noise*), brojnik (*numerator*), šest revizijskih pitanja (*six audit questions*), jedinica analize (*unit of analysis*), nazivnik (*denominator*), Simpsonov paradoks (*Simpson's paradox*), životni ciklus podataka (*data lifecycle*), statistika (*statistics*), podatkovna znanost (*data science*), strojno učenje (*machine learning*), sustav umjetne inteligencije (*artificial intelligence system*)

Zadaci

Konceptualni

Objasnite kako dvije računski točne stope mogu poduprijeti različite zaključke. U odgovoru razlikujte zbirnu usporedbu od usporedbe unutar podskupina te imenujte jedinicu i nazivnik svake usporedbe.

Odaberite dvije faze životnog ciklusa podataka. Za svaku u kratkom odlomku opišite jednu pogrešku koju kasniji izračun ne može sam popraviti. Nije potrebno predlagati programski kod.

Za svaki od sljedećih zadataka odredite pripada li ponajprije statistici, podatkovnoj znanosti, strojnom učenju ili radu sa sustavom umjetne inteligencije. Zadaci su provjera zaključka koji podaci podupiru, izrada ponovljivog postupka od izvora do tablice, provjera predviđanja na novim opažanjima i praćenje preporučivačkog sustava nakon uvođenja u ustanovu. Svaki izbor obrazložite jednom rečenicom.

Računski

Upotrijebite tablicu spremnosti na plaćanje. Za portal i tisak ručno podijelite vrijednost `Spremni platiti` s vrijednošću `Osobe` te rezultat pretvorite u postotak. Predajte dva razlomka, dva postotka i jednu rečenicu koja objašnjava zašto portal ima veći broj, a tisak veći udio. Zatim usporedite tisak s društvenim mrežama i razliku zaokružite na dvije decimale.

U simulaciji su stope skupine A u pristupačnijoj i zahtjevnijoj podskupini 80 % i 20 %, a stope skupine B 90 % i 30 %. Pretpostavite da obje skupine imaju po 50 % jedinica u svakoj podskupini. Izračunajte dvije zbirne stope i objasnite zašto obrata nema.

Kritički

Pronađite medijsku tvrdnju koja uspoređuje dvije skupine jednim zbirnim brojem. Navedite tvrdnju, izvor i nazivnik pa odgovorite na svih šest revizijskih pitanja, na svako jednom rečenicom. Označite i pitanja na koja objava ne daje odgovor. Na kraju procijenite podupire li broj opis, povezanost, generalizaciju, predviđanje, uzročnost ili odluku.

Revizija modela

Provjerite bilješku asistenta uz Tablica 8. Navedite točne brojnike i nazivnike, izdvojite pogrešnu rečenicu i napišite ispravnu.

Objasnite zašto se ni ispravljena analiza ne može primijeniti na stvarne stanovnike Hrvatske.

02

DIO I · STATISTIČKO MIŠLJENJE

mjerenje i istraživački dizajn

Od društvene pojave do opravdanog zaključka

Koliko ste vremena jučer proveli na mobitelu? Prije nego što pogledate zapis uređaja, pokušajte procijeniti. Možda se sjećate dugog razgovora, a teže prizivate kratka provjeravanja poruka između drugih obveza. I što biste ubrojili u odgovor, razgovor s prijateljem, slušanje glazbe sa zaključanim zaslonom ili samo vrijeme dok gledate u njega?

Sličan se problem pojavljuje u istraživanjima digitalnih medija. Sustavni pregled usporedio je samoprocjene uporabe sa zapisima uređaja i aplikacija, a pokazalo se da se ta dva načina mjerenja često razilaze (Parry i sur. 2021.). Za istraživača koji vrijeme uz ekran želi povezati s učenjem ili zadovoljstvom životom to je važno već prije prvog računa.

Ako se vaša procjena i zapis uređaja razlikuju, koji broj odgovara onome što zapravo želite saznati o svojem danu?

Broj minuta dobiva značenje tek kada odredimo koje aktivnosti obuhvaća. Povjerenje u medije možemo pokušati mjeriti anketnim odgovorom, a izloženost medijima vremenom provedenim uz sadržaj. Prije nego što usporedimo takve brojeve, moramo odlučiti što bi oni trebali predstavljati. Ako dvije osobe pod povjerenjem razumiju različite stvari, njihovi uredno zabilježeni odgovori mogu predstavljati različite pojave. U ovom ćemo poglavlju istražiti kako od društvenog pitanja dolazimo do opažanja koja na njega mogu odgovoriti.

U poglavlju o statističkom mišljenju tražili smo jedinicu analize, nazivnik i podrijetlo podataka. Sada se vraćamo korak prije gotove tablice. Odredit ćemo koje obilježje bilježimo, kojim pitanjem ili postupkom i prema kojim pravilima osoba ili zapis ulazi u istraživanje. Razlikovat ćemo mjerenje koje daje dosljedne rezultate od mjerenja koje dobro obuhvaća pojavu koja nas zanima. To će nam pomoći procijeniti što smijemo uspoređivati i prije nego što odaberemo račun.

Mjerenje ćemo proučiti na primjerima anketnih pitanja, bilježenja ponašanja i usporedbi eksperimentalnih uvjeta. U razrađenom primjeru sastavit ćemo mjeru povjerenja u lokalne institucije od konstruiranih tvrdnji i odgovora. Jedan redak toga nastavnog primjera predstavlja zamišljenog ispitanika, a stupci njegove odgovore. Provjerit ćemo znače li viši odgovori na svim tvrdnjama doista više povjerenja, prije nego što ih spojimo u zajednički rezultat.

Uz mjerenje ćemo ispitati i istraživački nacrt. Usporedba ljudi koji su sami odabrali neku aktivnost razlikuje se od pokusa u kojem im nasumično dodjeljujemo uvjete. Simulirani primjer pokazat će kako treći čimbenik može promijeniti tumačenje veze. Pri čitanju istraživanja moći ćemo obrazložiti što podaci govore o promatranim osobama, što bi opravdalo uzročni zaključak i na koga bi se on mogao odnositi.

Od pojma do podatka

Povjerenje, političku otuđenost ili osjećaj sigurnosti proučavamo putem odabranih opažanja. Zato najprije određujemo po čemu ćemo te pojave prepoznati. Isto vrijedi i za naizgled jednostavan pojam poput izloženosti medijima.

> **Operacionalizacija** je postupak kojim određujemo kako ćemo mjeriti teorijski pojam. Navodimo što ćemo opažati i po kojem ćemo pravilu svakom opažanju pridružiti vrijednost.

Za povjerenje u medije mogli bismo postaviti jedno anketno pitanje, sastaviti nekoliko tvrdnji ili promatrati izbor izvora vijesti. Time bismo dobili različite moguće pokazatelje, čiju vezu s povjerenjem tek treba opravdati. Uz odgovore i zabilježeno ponašanje možemo koristiti i procjenu promatrača, ali svaki od tih izbora obuhvaća neki dio pojave, a druge dijelove izostavlja.

Kada koristimo više tvrdnji, njihove odgovore možemo sažeti u jedan rezultat. Tako pojedina formulacija manje utječe na ukupnu mjeru, koja može postati dosljednija. Dobiveni rezultat tumačimo prema sadržaju svih uključenih tvrdnji, kao zajedničku mjeru.

Povjerenje u medije, primjerice, razlikuje se od odgovora na pitanje koliko netko vjeruje nekom izvoru na ljestvici od 1 do 10. Odgovor je nesavršena mjera povjerenja. Kvaliteta istraživanja ovisi o tome koliko dobro ta mjera odgovara pojmu koji proučavamo. I uz kvalitetan instrument mjera ostaje odabrani prikaz šireg pojma.

Zamislimo dva istraživanja društvenih mreža i političke polarizacije. Jedno polarizaciju mjeri razlikama u stavovima pristaša stranaka, a drugo učestalošću neprijateljskih izjava o protivnicima. Njihovi se nalazi mogu razlikovati jer prvo prati udaljenost mišljenja, a drugo ton rasprave. Prije usporedbe zaključaka treba provjeriti jesu li istraživači doista mjerili istu pojavu.

Ono što mjerimo bilježimo kao opažanje, a svojstvo koje može imati različite vrijednosti nazivamo **varijablom**. Ako 300 ljudi upitamo koliko minuta dnevno provode na mrežama, imamo 300 opažanja jedne varijable. Postavimo li svakome 10 pitanja, imamo 300 opažanja za svaku od 10 varijabli. Odabirom pitanja određujemo koje ćemo razlike među ljudima moći proučavati.

Te odluke donosimo prije prikupljanja podataka. Ljestvica s pet stupnjeva čuva odgovore u pet ponuđenih kategorija. Za razlikovanje deset stupnjeva trebalo bi primijeniti drukčiju ljestvicu. Usporedba grada i sela zahtijeva zabilježeno mjesto stanovanja. Mogućnosti analize ovise o onome što smo prikupili.

Pri čitanju istraživanja zato najprije tražimo opis mjerenja i jedinicu analize. Treba znati odnosi li se nalaz na osobu, objavu, kućanstvo ili, primjerice, sat gledanja. Metodološki odjeljak objašnjava i što odabrana mjera obuhvaća, a što izostavlja. Puno značenje varijable dobivamo iz toga opisa.

U anketi na mjerenje utječu formulacija, ponuđeni odgovori i redoslijed pitanja. Zamislimo osobu koja je upravo odgovorila na niz pitanja o nesigurnosti. Ona bi povjerenje u institucije mogla procijeniti drukčije nego da joj je to pitanje postavljeno prvo. Zato uz rezultate treba pročitati i upitnik.

Prihvatljivost, isključenja i filtri

Uz odluku što pitamo, trebamo odrediti i kome pitanje postavljamo. Ista mjera primijenjena na drukčiji krug ljudi može dati drukčiju sliku pojave. **Prihvatljivost** je pravilo po kojem jedinica pripada ciljanoj skupini. Isključenje je obrazložen razlog da neku inače prihvatljivu jedinicu ipak ne analiziramo. Filtrom ta pravila primjenjujemo na zapise. Time određujemo na koju će se populaciju moći odnositi zaključak.

Na turizmu možemo vidjeti kako nastaju različite mjere iste šire pojave. U lokalnom administrativnom izvatku Državnog zavoda za statistiku (DZS) za 2024. domaćim turistima pripada 2 867 339

dolazaka u komercijalni smještaj (Državni zavod za statistiku 2026.). Izvadak anketne tablice T03 za istu godinu daje procjenu od 3 585 264 privatnih putovanja s noćenjem unutar Hrvatske (Državni zavod za statistiku 2026.). Anketa obuhvaća stanovnike od 15 godina naviše u privatnim kućanstvima, a putovanja uključuju i nekomercijalni smještaj (Državni zavod za statistiku 2025.). Dolazak nastaje prijavom u smještaj, pa promjena objekta može istoga gosta ponovno ubrojiti (Državni zavod za statistiku 2026.). Brojevi zato opisuju različite jedinice i obuhvate. Prije usporedbe trebamo zapisati koji se događaj broji, koga izvor obuhvaća i kako podatak nastaje.

Zamislimo plan istraživanja televizijskih vijesti o prosvjedu. To je primjer planiranja, bez stvarnog nalaza. Ovisno o pitanju, možemo proučavati priloge, govorne iskaze ili pojedine riječi. U istraživanje bismo mogli uključiti sve priloge odabranih informativnih emisija u unaprijed određenom razdoblju. Ako isključimo reprize, priloge bez govora ili nečitljive zapise, trebamo zabilježiti koliko ih je isključeno i zašto. Zaključak će se odnositi na tako sužen skup.

Nedostajući odgovor razlikuje se od isključenja osobe. Sudionik koji preskoči jedno pitanje i dalje pripada istraživanju, ali za tu varijablu nemamo vrijednost. Ako zbog toga izbrišemo cijeli redak, uvodimo dodatni filtar. To može promijeniti rezultat, osobito kada preskakanje pitanja ovisi o iskustvu koje proučavamo. Zato zasebno bilježimo neprihvatljive jedinice, opravdana isključenja i nedostajuće vrijednosti uključenih jedinica.

Za analizu jezika također trebamo mjerno pravilo. **Kodiranje kao mjerenje** određuje kojoj izjavi pripada koja oznaka, što ne kodiramo i kako rješavamo dvosmislene slučajeve. U našem primjeru oznaku pripisivanja odgovornosti mogli bismo dati samo izjavi koja jasno imenuje aktera i radnju. Neizravna naznaka mogla bi dobiti oznaku nejasno. Isključimo li takve izjave, dobiveni udio opisivat će samo jednoznačne iskaze.

Slaganje dvaju kodera pokazuje koliko dosljedno primjenjuju pravilo, odnosno njegovu pouzdanost. Zasebno provjeravamo prepoznaje li to pravilo doista pripisivanje odgovornosti, što je

pitanje valjanosti. Dvosmisleni slučajevi pomažu uočiti ograničenja mjere. I etička pravila mogu suziti podatke kojima raspolažemo. Zaštita sudionika može opravdati izostavljanje osjetljivih zapisa, ali u izvještaju treba objasniti tko je time izostao i kako to ograničava zaključak.

Kad nekoga pozovemo u istraživanje, tražimo da nam povjeri dio svojega iskustva. Prije nego što pristane, treba razumjeti čemu istraživanje služi, što ćemo ga pitati, može li mu sudjelovanje biti neugodno i tko će vidjeti njegove odgovore. Objašnjavamo i kako može odustati. Ako poziv dolazi od nastavnika ili poslodavca, posebno pazimo da osoba može slobodno odbiti, bez osjećaja da mora sudjelovati. Prije početka provjeravamo i koja odobrenja i posebne uvjete traži naša ustanova.

Povjerenje te osobe trebamo sačuvati i nakon što ispuni upitnik. Uklonimo li ime, još uvijek može ostati prepoznatljiva. Zamislimo da u tablici pišu njezina dob, malo mjesto u kojem živi i rijetko zanimanje. Netko tko poznaje to mjesto mogao bi povezati odgovore s njom. Zato već pri sastavljanju upitnika razmišljamo koji su nam podaci doista potrebni, tko će im moći pristupiti, kako ćemo ih čuvati i kada izbrisati.

Ponekad ime zamijenimo šifrom, a zasebno čuvamo ključ koji ih povezuje. Taj postupak zovemo pseudonimizacijom. On skriva ime u radnoj tablici, ali onaj tko ima ključ i dalje može prepoznati osobu. I prije dijeljenja zbirne tablice trebamo pogledati krije li se iza neke male skupine samo jedna prepoznatljiva osoba. Ako je ne možemo dovoljno zaštititi, biramo sigurniji način dijeljenja ili prikazujemo podatke za veće skupine. Na to mislimo i kad tražimo pomoć asistenta, jer slanjem tablice također otvaramo pristup odgovorima koje nam je netko povjerio.

Razine mjerenja

Kada svakom opažanju pridružimo vrijednost, još trebamo znati što s tim vrijednostima možemo smisleno uspoređivati. Oz-

nakom izvora vijesti razlikujemo kategorije, a brojem minuta mjerimo trajanje uporabe. Uobičajena podjela na četiri razine mjerenja razlikuje oznake, poredak, jednake razmake i smislenu nulu (Stevens 1946.). Ta nam podjela pomaže razumjeti varijablu, premda sama ne određuje sve dopuštene analize.

Vrsta medija, država prebivališta, stranačka bliskost i status zaposlenosti imaju kategorije bez zadanog poretka. To je **nominalna razina**. Na **ordinalnoj razini** vrijednosti možemo poredati, ali razmaci među njima ne moraju biti jednaki. **Intervalna razina** ima jednake razmake uz proizvoljno ishodište, a **omjerna razina** uz to ima apsolutnu nulu. Zato možemo odrediti medijan, srednji odgovor po poretku, ali ne i smislen prosjek vrsta medija. Izjava da je nešto dvostruko veće odgovara omjernim varijablama poput vremena, broja dijeljenja ili dohotka.

Ljestvica slaganja ima poredak. Kada odgovore zbrajamo ili računamo prosjek, pretpostavljamo da razmake među stupnjevima možemo tumačiti brojčano. Spajanje više tvrdnji može povećati dosljednost, ali ne jamči jednake razmake. Ako ta pretpostavka utječe na zaključak, rezultat treba usporediti s prikazom koji koristi samo poredak. Razliku od dvije desetine boda tako tumačimo uz pretpostavku o ljestvici.

Varijable mogu biti i diskretne ili kontinuirane. Diskretna varijabla ima odvojene moguće vrijednosti. Broj komentara, primjerice, poprima cjelobrojne vrijednosti. Kontinuirana varijabla poput vremena na stranici može poprimiti bilo koju vrijednost unutar raspona. Ta razlika utječe na izbor grafa i analize. Diskretne varijable s mnogo mogućih vrijednosti, poput broja bodova na nekim testovima, u praksi se katkad analiziraju kao kontinuirane.

Sažimanjem podataka zadržavamo odabranu razinu pojedinosti. Dob u godinama možemo razvrstati u dobne skupine. Nakon takva razvrstavanja sama oznaka skupine govori samo raspon dobi. Isto vrijedi za pretvaranje minuta uporabe u kategorije čestih i rijetkih korisnika. Ako anketa od početka nudi samo razrede, točne vrijednosti ostaju nepoznate.

Mjerenje zato planiramo prema istraživačkom pitanju i čuvamo potrebne pojedinosti. Za izbor analize zatim uz razinu mjerenja razmatramo dizajn, raspodjelu podataka, način izgradnje ljestvice i pretpostavke postupka. Podjela na razine mjerenja pruža polazište za tu širu prosudbu.

Pouzdanost i valjanost

Uz razinu mjerenja procjenjujemo i kvalitetu mjere. Od nje očekujemo dosljednost i smislen sadržaj. Vaga koja pri uzastopnom vaganju iste osobe pokazuje različite težine upozorava na problem dosljednosti.

> **Pouzdanost** je stupanj u kojem instrument daje dosljedne rezultate pri ponovljenom mjerenju istog stanja u istim uvjetima.

Pouzdanost mjerenja razlikuje se od razine pouzdanosti intervala koju ćemo upoznati u poglavlju o procjeni. Prva opisuje dosljednost instrumenta, a druga svojstvo postupka intervalne procjene.

Pouzdanost provjeravamo na nekoliko načina. Pri ponovljenom mjerenju uspoređujemo rezultate istog instrumenta u dva trenutka, pod pretpostavkom da se mjereno stanje nije promijenilo. Kada ljudi kodiraju tekstove, provjeravamo njihovo međusobno slaganje. Kod ljestvice s više tvrdnji koje mjere isti pojam provjeravamo internu konzistentnost, odnosno koliko se odgovori na te tvrdnje slažu. Za nju se može navesti Cronbachov alfa, a za slaganje kodera Cohenova kappa ili Krippendorffova alfa. Uobičajene granice prihvatljivosti dogovorene su konvencije o kojima se raspravlja, pa ih treba tumačiti u kontekstu mjerenja.

U radu s višestavčanom ljestvicom zato tražimo prikladnu mjeru interne konzistentnosti. U analizi sadržaja tražimo mjeru slaganja kodera i broj jedinica na kojima je izračunata. Indeks koji namjerno spaja različite sastavnice može zahtijevati drukčiju

provjeru. Kad izvještaj izostavi te podatke, dosljednost mjerenja ostaje otvoreno pitanje koje traži dodatnu provjeru.

Razlika između ponovljenih mjerenja može nastati i zbog stvarne promjene. Nakon velikog medijskog skandala povjerenje u medije moglo bi pasti i uz posve pouzdan instrument. Da bismo razlikovali promjenu stavova od nestabilnosti mjere, moramo poznavati događaje između dvaju mjerenja.

Dosljedan rezultat vrijedi nam samo ako mjeri pojam koji istražujemo.

> **Valjanost** je stupanj u kojem instrument mjeri upravo onaj pojam o kojem se namjerava zaključivati, a ne neki drugi s kojim je povezan.

Kad bismo kvalitetu novinarstva mjerili brojem zareza, različiti bi brojači mogli dobiti isti rezultat. Mjera bi bila pouzdana, ali ne bi valjano mjerila kvalitetu novinarstva. Pouzdanost je zato potreban, ali nedovoljan uvjet dobre mjere.

Mjerna pogreška može utjecati i na procijenjenu povezanost. Ako su slučajne pogreške dviju mjera približno neovisne i ne ovise o skupini ni ishodu, opažena veza može biti slabija od veze među samim pojmovima. U tim uvjetima šum može prikriti obrazac.

Drukčije pogreške mogu imati drukčije posljedice. Sustavna pogreška, povezanost pogrešaka ili različita pogreška među skupinama mogu vezu oslabiti, pojačati, stvoriti ili preokrenuti. Zato nalaz ne postaje uvjerljiviji samo zato što se pojavio uz slabu pouzdanost. Moramo znati kako pogreška nastaje da bismo procijenili njezin mogući utjecaj.

Pri provjeri valjanosti pitamo obuhvaća li instrument važne dijelove pojma. Upitnik o medijskoj pismenosti koji izostavi digitalne sadržaje, primjerice, propušta dio predmeta. Provjeravamo i mjeri li odabrani pokazatelj baš ono što tvrdimo. Klik, vrijeme na stranici i komentar mogu upućivati na početni interes, čitanje ili želju za javnim očitovanjem. Uz instrument treba razmotriti

i uvjete istraživanja. Reakcija na neistinitu vijest na praznom zaslonu može se razlikovati od reakcije na istu vijest među porukama prijatelja. Pojmovi interne i eksterne valjanosti, koje ćemo uskoro uvesti, odnose se na taj dizajn i doseg zaključka.

Vratimo se jučerašnjoj uporabi mobitela. Za usporedbu samoprocjene i zapisa trebali bismo uskladiti dan, uređaj i aktivnosti koje brojimo. Ako nas zanima gledanje sadržaja, zapis trajanja uključenog zaslona može biti koristan pokazatelj. Ako nas zanima pažnja posvećena učenju, trebamo znati i što smo za to vrijeme radili. Zamislimo otvoren tekst pokraj kojeg razgovaramo s prijateljem. Uređaj može dosljedno bilježiti trajanje, dok naše pitanje traži nešto više od trajanja.

Neslaganje samoprocjene i zapisa tako postaje početak provjere mjerenja. Pregled istraživanja pokazao je i da su upitničke mjere problematične uporabe još slabije povezane sa zapisanim vremenom uporabe (Parry i sur. 2021.). Trajanje aktivnosti i teškoće koje osoba uz nju doživljava različita su obilježja. Izbor među njima ovisi o pitanju na koje želimo odgovoriti.

Treći čimbenik i logika eksperimenta

Čak i kada dobro izmjerimo dvije pojave, njihova povezanost ostavlja otvoreno pitanje uzroka. Trebamo razmotriti kako su nastale razlike koje uspoređujemo. U eksperimentu varijablu čiju vrijednost istraživač dodjeljuje nazivamo nezavisnom, a mjereni ishod zavisnom varijablom. U opažačkoj studiji prikladniji su nazivi prediktor i ishod jer vrijednosti samo opažamo. Dob ili obrazovanje sudionika, primjerice, ne dodjeljujemo.

Opaženu vezu može objasniti treća varijabla koja djeluje na obje promatrane pojave. Ona može stvoriti privid veze ili prikriti postojeći odnos.

> **Konfundirajuća varijabla** je zajednički uzrok koji prethodi pretpostavljenom uzroku i ishodu te utječe na oboje. Ako ga zanemarimo, opažena veza sadržava i njegov utjecaj.

Ako adolescenti koji više koriste Instagram imaju niže samopoštovanje, nekoliko objašnjenja ostaje moguće. Korištenje bi moglo snižavati samopoštovanje, niže samopoštovanje poticati korištenje, a društvena izoliranost mogla bi istodobno povećavati korištenje i snižavati samopoštovanje.

Ta objašnjenja traže različite provjere. Za obrnuti smjer djelovanja korisno je poznavati vremenski redoslijed mjerenja. Za pristranost odabira provjeravamo kako su sudionici ušli u uzorak i razlikuju li se od populacije. Kod konfundiranja razmatramo zajednički uzrok. Ako je izmjeren, njegov utjecaj možemo pokušati uzeti u obzir u analizi. Izbor varijabli za model zato povezujemo s provjerom svakoga od tih mogućih objašnjenja.

Ulogu treće varijable određujemo prema vremenskom redoslijedu i znanju o pojavi. **Medijator** nastaje djelovanjem pretpostavljenog uzroka i prenosi dio njegova djelovanja na ishod. Uključivanjem medijatora u prilagodbu mijenjamo pitanje o učinku. **Kolider** je u najjednostavnijem slučaju zajednička posljedica pretpostavljenog uzroka i ishoda. Prilagodba ili odabir prema njemu može stvoriti novu vezu. Izbor varijabli koje treba kontrolirati zato temeljimo i na poznavanju procesa koji proučavamo. Računski ćemo postupak učiti u poglavlju o regresiji, gdje ćemo i dalje morati obrazložiti izbor varijabli.

U zamišljenom primjeru izoliranost može prethoditi i uporabi mreža i samopoštovanju. Usporedba osoba slične izoliranosti pokušava razdvojiti taj zajednički uzrok. Drukčiji je pretpostavljeni put od uporabe mreža preko manje sna do slabijega rezultata na ispitu. San je tada medijator, pa njegovo izjednačavanje uklanja dio puta koji želimo proučiti. Za kolider zamislimo prijam na program na temelju ocjena ili sportskog uspjeha. Među primljenima slabije ocjene mogu biti povezane s boljim sportskim rezultatom jer je za prijam dovoljan jedan od dvaju razloga. Oda-

bir samo primljenih stvara vezu koja ne mora postojati među svim kandidatima.

Umjesto da naknadno pokušavamo razdvojiti utjecaje, možemo već pri nastanku skupina promijeniti način dodjele uvjeta. **Randomizacija** je nasumična dodjela jedinica uvjetima eksperimenta. Njome prekidamo sustavnu vezu između početnih obilježja sudionika i uvjeta koji će primiti. Istraživač dodjeljuje vrijednost nezavisne varijable i nastoji ostale postupke održati jednakima. Kroz mnogo ponavljanja takve dodjele početna bi obilježja, poznata i nepoznata, u prosjeku bila uravnotežena. U jednom pokusu ipak mogu ostati slučajne razlike, osobito uz male skupine. Randomizacija podupire usporedbu dodijeljenih uvjeta. Želimo li zaključivati o učinku primljenog tretmana, trebamo provjeriti jesu li se sudionici pridržavali dodjele. Učinak ponude tretmana može se razlikovati od učinka njegova primanja.

Zamislimo isti tekst s dvama naslovima, suzdržanim i zaoštrenim. Čitatelje nasumično rasporedimo tako da svatko vidi samo jedan naslov, a zatim usporedimo udjele otvaranja. Istraživač tako dodjeljuje naslov, a čitatelj ga ne bira. Tekst, vrijeme objave i položaj na stranici ostaju jednaki u obje skupine kako se učinak naslova ne bi pomiješao s tim okolnostima.

Kad bi čitatelji sami birali naslov, skupine bi se mogle razlikovati već po sklonosti zaoštrenom sadržaju. Razlika u otvaranju tada bi mogla odražavati i naslov i sastav publike. I uz veći broj zapisa ostala bi potreba da razdvojimo učinak naslova od razlika u sastavu publike. Za tumačenje je presudno kako su skupine nastale.

Za proučavanje obrazovanja, odrastanja u siromaštvu ili godina provedenih na platformi često koristimo opažački dizajn. Te životne okolnosti zatičemo kod sudionika, a mogućnosti istraživačke intervencije ograničene su i etičkim razlozima. Zaključke tada prilagođavamo načinu na koji su skupine nastale.

Interna valjanost odnosi se na to koliko uvjerljivo razliku u ishodu možemo pripisati ispitivanom uzroku. U eksperimentu je

mogu ugroziti događaji nakon dodjele. Tretman se može preliti na drugu skupinu, vanjski događaj različito zahvatiti uvjete ili se sudionici nejednako osipati. Ishod se može mjeriti drukčije među skupinama. Sudionici mogu i naslutiti svrhu istraživanja pa odgovarati prema očekivanjima, primjerice pokazivati veću kritičnost prema medijima nego u svakodnevici.

Opažačke studije i ankete

Opažačka studija prati postojeće vrijednosti i njihove odnose. Ako utvrdimo negativnu vezu korištenja mreža i političke informiranosti, sam taj nalaz ne govori smanjuju li mreže informiranost, koriste li ih slabije informirani više ili dob i obrazovanje utječu na oboje. Takva studija ipak omogućuje proučavanje svakodnevnih okolnosti i ljudi koje laboratorij teško može obuhvatiti.

U kvazieksperimentu intervencija, prag, pravilo ili vanjski događaj stvara usporedbu bez randomizacije. Treba obrazložiti zašto su nastale skupine dovoljno usporedive. Zamislimo da se stipendija dodjeljuje prema unaprijed određenom bodovnom pragu. Usporedba kandidata neposredno iznad i ispod praga može biti korisna ako ne mogu precizno namjestiti bodove i ako se na istom pragu ne mijenja još nešto važno. Bez takva obrazloženja razlika ostaje opažena povezanost. Studije slučaja i kvalitativne metode mogu samostalno odgovoriti na pitanja značenja, procesa i iskustva. Mogu također osporiti korištene kategorije ili predložiti hipotezu za drugi dizajn.

U anketi uz sadržaj upitnika treba provjeriti **okvir uzorkovanja**, skup iz kojeg se jedinice mogu odabrati. Ako on sustavno izostavlja dio populacije, i veći broj odgovora opisivat će samo obuhvaćeni dio. Taj ćemo problem razraditi u poglavlju o uzorkovanju.

Treba znati i tko se od pozvanih nije odazvao. Neodaziv može stvoriti pristranost ako vjerojatnost odgovora ovisi o ishodu koji

procjenjujemo. Sama stopa odaziva ne pokazuje ni veličinu ni smjer te pristranosti. Težina određuje koliko pojedini odgovor pridonosi procjeni, primjerice kako bi slabije zastupljena skupina dobila veći doprinos. Takvo vaganje prema zabilježenim obilježjima može pomoći ako ta obilježja objašnjavaju odaziv i ishod. Ne može ukloniti razlike o kojima podaci ništa ne bilježe. Zato tražimo podatke o pozvanima, postupku izrade težina i njegovim pretpostavkama.

Prije tumačenja postotaka i margine pogreške možemo podatke o anketi zabilježiti u sljedeću karticu. Neispunjeno polje pokazuje što ne možemo provjeriti, bez pretpostavke da je istraživanje na tom mjestu pogrešno.

Tablica 1: Kartica za čitanje ankete. Izrada autora.

Polje	Što treba pronaći u izvještaju
okvir i pozivanje	Tko je mogao biti izabran i kojim je putem pozvan?
odaziv i nedostajanje	Koliko je pozvanih sudjelovalo i za koja pitanja odgovori nedostaju?
procjena s težinama	Koje varijable ulaze u težine, što one pretpostavljaju o odazivu i koji mogući neodaziv ostaje?
mjerenje	Kako pitanje točno glasi, koji su odgovori ponuđeni i kojim je redom postavljeno?
provedba	Kada, kojim načinom i na kojim jezicima je anketa provedena?
doseg	Koji dio ciljane populacije okvir, odaziv ili filtri ne pokrivaju?

Tablica 1: (nastavak)

Polje	Što treba pronaći u izvještaju
nesigurnost	Što pripada uzoračkoj pogrešci, a što mjerenju, obuhvatu i neodazivu?

Uzoračku pogrešku odvajamo od ostalih izvora nesigurnosti. Veći uzorak može smanjiti rasipanje zbog slučajnog odabira. Različito razumijevanje pitanja, nepotpun okvir i sustavan neodaziv zahtijevaju zasebnu procjenu. Navedena margina pogreške opisuje samo dio nesigurnosti ankete.

Doseg zaključka

Okvir i odaziv ograničavaju na koga se anketni rezultat odnosi. Srodno pitanje postavljamo i drugim istraživanjima, čak i kada su dobro provedena. Eksterna valjanost odnosi se na mogućnost prenošenja nalaza na druge ljude, okolnosti i razdoblja. Nalaz dobiven na studentima jednog fakulteta ne mora vrijediti za druge skupine. Ponašanje pod nadzorom može se razlikovati od svakodnevnog, a promjene platformi i navika mogu ograničiti primjenu starijih rezultata. To vrijedi i za dobro proveden eksperiment.

Internu i eksternu valjanost zato provjeravamo zasebno. Laboratorijski pokus može dobro odvojiti uzrok od drugih objašnjenja, dok terenski pokus može zadržati nasumičnu dodjelu u svakodnevnim uvjetima. Za svaki dizajn pitamo što skupine čini usporedivima i kamo možemo prenijeti rezultat. Prikladan je onaj dizajn koji može odgovoriti na naše pitanje.

Metode možemo i povezati. Razgovori mogu pomoći odrediti važne kategorije, anketa opisati njihovu raširenost, a pokus ispitati uže uzročno pitanje. Kvalitativno istraživanje može biti i

cjelovit odgovor na pitanje o značenju ili procesu, ili pokazati da kvantitativni instrument pogrešno dijeli pojavu. Kombinaciju biramo prema pitanju.

Nesigurnost već sada možemo razmotriti, iako je još ne izražavamo jednim brojem. Za mjerenje pitamo bi li druga formulacija, koder ili instrument dali drukčiji rezultat. Za dizajn provjeravamo mogu li nalaz objasniti obrnuti smjer, neizmjeren zajednički uzrok, osipanje ili prelijevanje. Za doseg zaključka bilježimo koga izostavljaju okvir, pravila ulaska i isključenja te neodaziv.

Kasniji računi obuhvatit će dio nesigurnosti povezane s uzorkom i modelom. Uz interval ili test ipak ostaje provjera mjerenja, dizajna i odabira jedinica. Ograničenja izvan dosega izračuna ostaju važna i uz uži interval.

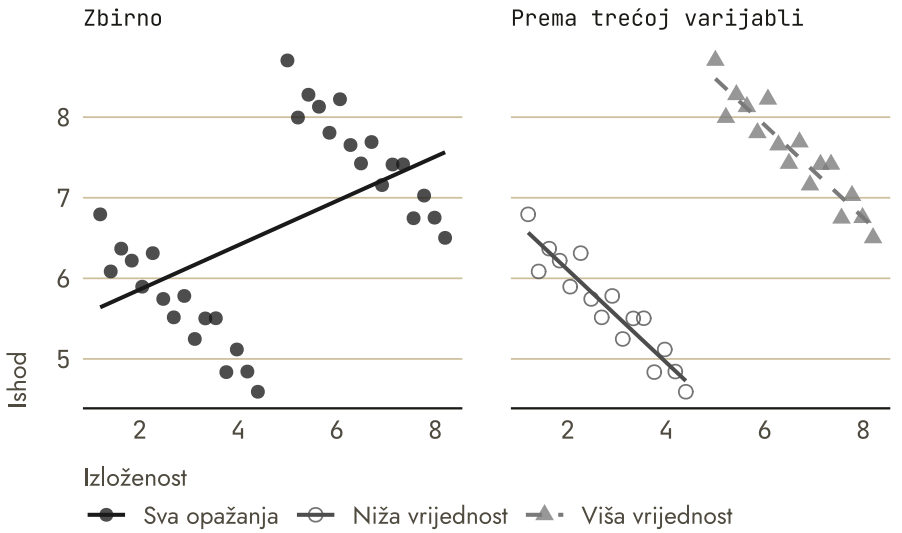
Prikaz konfundirajuće varijable

Ovdje prelazimo na zasebnu, namjerno konstruiranu ilustraciju s 32 točke, po 16 za svaku razinu treće varijable. Nižoj razini treće varijable pridružili smo niže izloženosti, a višoj više. Treća varijabla tako razdvaja obje promatrane veličine. Unutar obiju skupina ishod pada za 0,55 uz jedinicu izloženosti, uz male unaprijed zadane pomake točaka. Višoj skupini dodajemo zajednički pomak ishoda od četiri proizvoljne jedinice. Točke nisu odgovori iz ankete ni novi slučajni uzorci. Tiskani prikaz zadržava tu postavku.

Na prikazu uspoređujemo odnos dviju varijabli u svim opažanjima zajedno i unutar podskupina određenih trećom varijablom. Podaci su isti. Ako treća varijabla pomiče ishod, ukupna veza može imati suprotan smjer od veze unutar podskupina.

Izloženost i ishod u oba prikaza imaju isto značenje. Razlikuje se ono što uspoređujemo. Zbirni pravac koristi sva opažanja, a razdvojeni pravci jedinice iste razine treće varijable.

Moguću ulogu treće varijable kao zajedničkog uzroka procjenjujemo prema vremenskom redoslijedu i znanju o pojavi. Prikaz pokazuje moguću posljedicu zanemarivanja takvog čimbenika.



Slika 1: Ista opažanja prikazana zbirno i odvojeno prema trećoj varijabli. Izrada autora. Podaci su simulirani.

Lijevi panel pokazuje zbirnu vezu izloženosti i ishoda. U desnom je panelu odnos prikazan zasebno za svaku razinu treće varijable. Smjerovi se razlikuju jer zbirni pravac uključuje i razlike među tim podskupinama.

Konfundirajući čimbenik mora djelovati na obje promatrane veličine. Sama prisutnost treće varijable u tablici nije dovoljna. U ovom primjeru, bez dodatnog pomaka ishoda pri jednakoj izloženosti, zbirne i unutargrupne veze imaju isti smjer, iako strmine mogu ostati različite. Moguće konfundirajuće varijable zato biramo prema sadržajnom razlogu da utječu na obje pojave.

● STATISTIKA U DIVLJINI

Pravednost koju tablica može vidjeti. Za procjenu upisnog postupka možemo tražiti jednake ukupne ishode, jednako postupanje prema jednako

kvalificiranim kandidatima ili jednaku mogućnost pristupa različitim studijima. Ti zahtjevi zvuče blisko, ali svaki traži drukčije podatke. Berkeleyjska tablica sadržavala je ishode prijava, spol i odjel, bez potpunih podataka o kvalifikacijama kandidata i okolnostima njihova izbora (Bickel i sur. 1975.).

Zamislimo da iz takve tablice trebamo pripremiti prijedlog poboljšanja upisa. Ako nas zanima postupanje povjerenstva, trebali bismo znati koje je kriterije koristilo i jesu li uspoređene prijave bile slične po tim kriterijima. Ako nas zanima dostupnost studija, trebali bismo pratiti i ljude koji se nisu prijavili te razloge zbog kojih neki studij nisu razmatrali. Za njihov bismo put trebali zasebne podatke, izvan tablice podnesenih prijava.

Važna je i jedinica zapisa. Berkeleyjska analiza polazi od prijava koje su nakon povlačenja ili preusmjeravanja ostale u postupku do zabilježene odluke (Bickel i sur. 1975.). Takav obuhvat nije popis svih ljudi koji su u nekom trenutku željeli upisati studij. Za pitanje o njihovu konačnom uspjehu trebalo bi pratiti i one koji su odustali te utvrditi što se s njima dalje dogodilo.

Statistička teškoća ovdje prethodi izboru testa. Pojam pravednosti treba povezati s opažanjima koja ga mogu poduprijeti ili osporiti. U zamišljenom prijedlogu možemo opisati razlike koje tablica pokazuje i navesti koje bi dodatne informacije omogućile provjeru postupka. Praznine bilježimo kao otvorena pitanja za daljnje istraživanje nejednakosti.

Koristan rezultat takve analize može biti preciznije sljedeće pitanje. Treba li pregledati kriterije povjerenstva, savjetovanje prije prijave ili dostupnost prethodnog obrazovanja ovisi o tome koju prepreku želimo istražiti. Stopa prijma pomaže tek kada znamo koji dio toga puta mjeri.

Razlika između opažene povezanosti i uzročnog učinka vrijedi i za ankete o uporabi društvenih mreža.

NAĐITE GREŠKU

Nakon rasprave o opažačkim studijama pregledajte ovaj konstruirani istraživački zapis. Nije riječ o nalazu stvarne ankete.

Studenti su u jednom upitniku naveli dnevno vrijeme na društvenim mrežama, trenutačno povjerenje u institucije i dob. Istraživači nisu određivali njihovo vrijeme korištenja niti su ih pratili prije i poslije promjene navika. U ovom zadanom primjeru niže povjerenje prati dulje korištenje, a isti smjer ostaje i pri usporedbi studenata jednake dobi.

Asistent je za izvještaj pripremio sljedeći tekst.

> U opažačkoj anketi dulje korištenje mreža povezano je s nižim povjerenjem. Obje su varijable izmjerene istim upitnikom, a usporedba uzima u obzir dob. Rezultat zato pokazuje da dulje korištenje društvenih mreža smanjuje povjerenje.

Sastavite kratku bilješku istraživačkom timu. Sačuvajte opis opaženog odnosa, zamijenite nepoduprtu tvrdnju i navedite jedno drugo objašnjenje koje opisani postupak ne isključuje. Napišite koji bi dodatni podatak ili promjena dizajna pomogli razlikovati ta objašnjenja.

Razrađeni primjer

Dosad smo operacionalizaciju promatrali kao izbor opažanja koja predstavljaju pojam. U primjeru povjerenja u lokalne institucije pratimo kako odabrane odgovore spajamo u zajedničku mjeru. Koristimo četiri konstruirane tvrdnje s istom ljestvicom odgovora od 1 do 5, jer jedna tvrdnja ne obuhvaća cijeli pojam. Prije spajanja provjerit ćemo znači li viši odgovor na svakoj tvrdnji doista više povjerenja.

Tablica 2: Četiri tvrdnje konstruirane za nastavni primjer i njihov smjer. Izrada autora.

Oznaka	Konstruirana tvrdnja	Smjer višeg odgovora
t1	Mogu se osloniti na odluke lokalnih institucija.	više povjerenja
t2	Lokalne institucije postupaju pošteno prema građanima.	više povjerenja
t3	Lokalne institucije objavljuju informacije kojima mogu vjerovati.	više povjerenja
t4	Lokalne institucije često skrivaju važne informacije.	manje povjerenja

Prve tri tvrdnje izrečene su potvrdno, a četvrta niječno. U zamišljenom nacrtu niječna je formulacija trebala prekinuti mehaničko ponavljanje odgovora. Ona može i zbuniti ispitanika, a pri računanju zahtijeva posebno kodiranje.

Za račun smo ručno sastavili odgovore 12 zamišljenih osoba. Nema slučajnog uzorkovanja ni nedostajućih odgovora. Svih dvanaest redaka pokazuje vrijednosti prije ikakve promjene kodiranja. Odgovori služe provjeri kodiranja i opisuju samo ovaj nastavni primjer.

Tablica 3: Početni odgovori u namjerno konstruiranom instrumentu. Izrada autora.

Osoba	t1	t2	t3	t4
I01	4	5	4	2
I02	5	4	5	1
I03	3	3	3	3
I04	2	1	2	5
I05	1	2	1	4
I06	3	4	3	3
I07	5	5	4	1
I08	2	2	2	4
I09	4	3	3	2
I10	1	1	2	5
I11	3	2	4	3
I12	4	4	5	2

Za osobu I01 zbroj prvih triju odgovora iznosi 13, dok je odgovor na četvrtu tvrdnju 2. Taj par, zajedno s odgovarajućim parovima preostalih 11 osoba, ulazi u provjeru povezanosti četvrte tvrdnje s ostatkom instrumenta.

Smjer tvrdnje provjeravamo i u samim odgovorima. Ako osoba na jednoj tvrdnji iskazuje više povjerenja, očekujemo da će ga iskazati više i na ostalima koje mjere isti pojam. Zato svaku tvrdnju uspoređujemo sa zbrojem preostalih i tražimo pozitivnu povezanost. Računamo Pearsonovu korelaciju stavke i zbroja ostalih stavki. To je korigirana korelacija stavke s ukupnim rezultatom jer se stavka ne pojavljuje i u vlastitom usporednom zbroju. Njezina vrijednost može biti od −1 do 1. Pozitivan predznak znači kretanje u istom smjeru, a negativan u suprotnom. Postupak računanja upoznat ćemo u poglavlju o povezanosti.

Tablica 4: Povezanost svake tvrdnje s ostatkom instrumenta, prije i nakon okretanja niječne tvrdnje. Izrada autora.

Tvrdnja	Prije okretanja	Nakon okretanja
1 — potvrdna	0,75	0,95
2 — potvrdna	0,60	0,85
3 — potvrdna	0,76	0,83
4 — niječna	−0,95	0,95

Prve tri tvrdnje u tablici pozitivno su povezane s ostatkom instrumenta. Četvrta ima negativnu povezanost, −0,95. U ovom primjeru znamo razlog jer smo sami konstruirali tvrdnje. Visok odgovor na četvrtoj znači nisko povjerenje, dok na ostalima znači visoko povjerenje. U stvarnom istraživanju negativna povezanost sama ne bi bila dovoljna za takvu dijagnozu.

Za četvrtu tvrdnju okrećemo ljestvicu. Od najveće moguće vrijednosti uvećane za jedan oduzimamo zabilježeni odgovor. Na ljestvici od 1 do 5 to znači $6 -$ odgovor. Za I01 tako 2 postaje 4. Njezin prvotni prosjek $(4 + 5 + 4 + 2)/4 = 3{,}75$ postaje $(4+5+4+4)/4 = 4{,}25$. Isti postupak primjenjujemo na četvrti odgovor svih 12 osoba, čuvajući ostale odgovore. Povezanost s ostatkom tada iznosi 0,95. Sve tvrdnje sada imaju isti smjer, što omogućuje njihovo smisleno zbrajanje.

Tablica 5: Odgovori nakon okretanja niječne tvrdnje i izvedena mjera povjerenja. Izrada autora.

Ispitanik	Odgovori na tvrdnje 1 do 4	Povjerenje
I01	4 · 5 · 4 · 4	4,25
I02	5 · 4 · 5 · 5	4,75
I03	3 · 3 · 3 · 3	3,00
I04	2 · 1 · 2 · 1	1,50
I05	1 · 2 · 1 · 2	1,50
I06	3 · 4 · 3 · 3	3,25
I07	5 · 5 · 4 · 5	4,75
I08	2 · 2 · 2 · 2	2,00
I09	4 · 3 · 3 · 4	3,50
I10	1 · 1 · 2 · 1	1,25
I11	3 · 2 · 4 · 3	3,00
I12	4 · 4 · 5 · 4	4,25

Bez okretanja raspon prosječnih rezultata iznosio bi 1,75 boda. Nakon ispravka iznosi 3,50. Niječna tvrdnja prije ispravka poništavala je dio razlika koje su bilježile ostale tvrdnje. Iako bi sam račun bio točan, dobili bismo pogrešan dojam da se ispitanici vrlo malo razlikuju u povjerenju.

U stvarnim podacima prvo bismo provjerili ključ kodiranja. Negativna povezanost može nastati i zato što tvrdnja mjeri drugu dimenziju, prijevod mijenja značenje, formulacija je dvosmislena ili sudionici odgovaraju nepažljivo. Zato pregledavamo puni tekst tvrdnje i obrasce odgovora. Okretanje je opravdano tek kad odgovara sadržaju. Automatsko okretanje ili izbacivanje tvrdnje moglo bi sakriti drugi mjerni problem.

Nakon ispravka tvrdnje se u ovom konstruiranom skupu kreću zajedno, što podupire njihovu internu konzistentnost. Za valjanost trebamo dodatnu provjeru. Sve bi tvrdnje mogle bilježiti,

primjerice, sklonost slaganju umjesto povjerenja. Zato uz dosljednost razmatramo njihov sadržaj, način na koji ih ispitanici razumiju i odnos s drugim opravdanim mjerama.

Primjer se odnosi na jedno konstruirano mjerenje. Za primjenu u stvarnom istraživanju trebalo bi dodatno proučiti sudjelovanje, odbijanje poziva i razumijevanje tvrdnji. Te okolnosti određuju doseg zaključka uz prikazani račun. Zbog toga pri čitanju nalaza uvijek trebamo i opis postupka.

Sažetak

Mjerenjem pretvaramo teorijski pojam u opažanja, a pravilima ulaska, isključenja i postupanja s nedostajućim odgovorima određujemo koga ona predstavljaju. Na primjeru niječne tvrdnje vidjeli smo zašto prije sažimanja treba provjeriti značenje odgovora, pa tek zatim dosljednost i valjanost mjere. Uz dobro mjerenje, uzročni zaključak zahtijeva i odgovarajući dizajn. Randomizacija uravnotežuje skupine u očekivanju, ali i dalje treba provjeriti provedbu, pridržavanje dodjele, osipanje, prelijevanje i mjerenje. Ta pitanja ostaju važna i uz interval ili test. Već u sljedećem poglavlju poslužit će nam za čitanje tvrdnji koje prešućuju kako su nastali njihovi brojevi.

Pojmovi

operacionalizacija (*operationalization*), varijabla (*variable*), prihvatljivost (*eligibility*), kodiranje kao mjerenje (*coding as measurement*), razina mjerenja (*level of measurement*), pouzdanost (*reliability*), valjanost (*validity*), konfundirajuća varijabla (*confounder*), medijator (*mediator*), kolider (*collider*), randomizacija (*randomization*), interna valjanost (*internal validity*), okvir uzorkovanja (*sampling frame*), eksterna valjanost (*external validity*),

opažačka studija (*observational study*), kvazieksperiment (*quasi-experiment*), neodaziv (*nonresponse*), interna konzistentnost (*internal consistency*)

Zadaci

Konceptualni

Na vlastitom primjeru objasnite kako mjera može biti pouzdana, a nevaljana. Jasno navedite što je u mjerenju dosljedno i zašto rezultat ipak ne mjeri željeni pojam.

Odaberite pojam iz svojeg područja i predložite dva načina njegova mjerenja. Za svaki opišite što obuhvaća i što izostavlja te nalaz koji bi jedan način pokazao, a drugi ne bi. Za jedan prijedlog navedite jedinicu analize, ciljnu skupinu, pravilo prihvatljivosti, jedno isključenje i razinu mjerenja. Objasnite kako isključenje sužava doseg zaključka. Ako kodirate jezik, navedite pravilo za jednu oznaku i postupak za dvosmislen slučaj.

U zamišljenoj opažačkoj studiji pratite korištenje političkih vijesti i povjerenje u institucije. Razmotrite političko zanimanje prije praćenja, znanje stečeno nakon praćenja te ulazak u analizu koji ovisi o korištenju vijesti i povjerenju. Svaku od tih varijabli razvrstajte kao zajednički uzrok, medijator ili kolider i obrazložite izbor. Zatim objasnite što bi nasumična dodjela korištenja vijesti uravnotežila kroz ponavljanja te zašto u jednom malom pokusu skupine mogu ostati različite.

Računski

Upotrijebite tablicu povezanosti stavki iz razrađenog primjera. Za četvrtu tvrdnju ručno provjerite kako se svaki odgovor od 1 do 5 mijenja pri okretanju ljestvice i zapišite okrenute vrijednosti.

Zatim iz tablice s odgovorima uzmite prva tri ispitanika. Za svakoga izračunajte prosjek četiriju tvrdnji prije i nakon okretanja. Predajte svih šest prosjeka.

Jednom rečenicom objasnite zašto propušteno okretanje smanjuje razlike među ispitanicima. Postupak za provjeru cijelog skupa podataka nalazi se u praktikumu.

Kritički

Prosudite što se iz Berkeleyjskih podataka može zaključiti o ishodima, a što ne može o postupku odlučivanja (Bickel i sur. 1975.). Predajte dva stupca s dopuštenim i nedopuštenim zaključcima.

Pronađite objavljeno istraživanje iz svojeg područja. Ispunite karticu za čitanje ankete i napišite odlomak koji navodi dizajn, jedinicu analize, jedno pravilo ulaska ili isključenja te jedan izvor nedostajanja. Zasebno razmotrite nesigurnost mjerenja, dizajna i dosega. Za svaku u vlastitoj trorednoj tablici navedite jedan rizik koji se može provjeriti. Zatim navedite zaključak koji dizajn opravdava. Procijenite prelaze li autori taj doseg, uz primjer pretjerane tvrdnje ili obrazloženje zašto su ostali unutar njega.

Revizija modela

Provjerite analizu iz okvira iznad. Navedite dizajn, izdvojite neopravdanu tvrdnju i preoblikujte je prema onome što zadani primjer dopušta. Predložite dizajn koji bi bolje razlikovao uzrok od povezanosti i navedite jedno njegovo ograničenje. Razlikujte prijedlog budućeg istraživanja od onoga što je u opisanoj anketi već učinjeno.

03

DIO I · STATISTIČKO MIŠLJENJE

kako brojke zavode

Provjera točne brojke prije pogrešnoga zaključka

Pripremate kratku vijest o odazivu. U službenom izvješću o izborima za Hrvatski sabor 2024. nalazite 62,30 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Rečenica da je toliko birača glasovalo djeluje jasna. Podatak dolazi iz službenog izvora, a račun možete provjeriti.

No u istoj tablici broj pristupilih glasovanju nije jednak broju birača prema glasačkim listićima (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Obje biste veličine u razgovoru mogli nazvati brojem ljudi koji su glasovali. Za vijest ipak trebate odabrati riječi koje odgovaraju upravo onome što izvor broji.

Što zapravo potvrđuje provjera postotka? Može li naslov sadržavati točan broj, a čitatelju ipak prenijeti pogrešno značenje?

Iza postotka iz izbornog izvješća stoji nekoliko odluka o tome što se broji (Državno izborno povjerenstvo Republike Hrvatske

2024., 124). One lako nestanu kada tablicu skratimo u naslov. U prvim smo poglavljima naučili tražiti jedinicu i nazivnik te razlikovati pojavu od načina na koji je mjerimo. Sada ćemo tim pitanjima provjeravati tvrdnje koje već dolaze s gotovim brojem, grafom ili uvjerljivim objašnjenjem.

Glavni primjer počiva na stvarnoj objavljenoj tablici Državnoga izbornog povjerenstva (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Njezine stavke predstavljaju nacionalne zbrojeve s različitim značenjima. Usporedit ćemo broj pristupilih glasovanju, broj prema listićima i broj birača koji služi kao nazivnik. Cilj je utvrditi koju rečenicu možemo napisati na temelju tih podataka. Razliku među evidencijama opisat ćemo u granicama onoga što je dokumentirano.

Od provjere omjera prijeći ćemo na načine prikazivanja. Ispitat ćemo kako promjena osnovice ili raspona osi mijenja dojam, zašto postotna promjena nije isto što i promjena u postotnim bodovima te kako zanemarivanje rijetkosti neke pojave može zavesti. Pri čitanju ankete povezat ćemo objavljenu marginu pogreške s onime što smo naučili o obuhvatu i odazivu. Konstruirani anketni primjeri poslužit će nam da odvojeno mijenjamo veličinu uzorka i sustavnu pristranost. Njihove rezultate nećemo čitati kao izbornu prognozu.

Provjera završava povratkom izvoru i usporedbom računa s riječima kojima je opisan. To vrijedi i kada analizu sastavi asistent. Tražit ćemo trag do tablice, provjeriti račun koji nosi tvrdnju i zabilježiti što nedostaje za njezino tumačenje. Tako ćemo sastaviti protokol skeptičnoga čitanja koji razlikuje potvrđenu brojku, nedostajući podatak i zaključak koji prelazi dokaz.

Od točnoga broja do potpune tvrdnje

Polazimo od službenog *Izvješća o provedenim izborima za zastupnike u Hrvatski sabor 2024.*, koje je objavilo Državno izborno

povjerenstvo (Državno izborno povjerenstvo Republike Hrvatske 2024.). Tablica „Odaziv birača" nalazi se na stranici 124 i provjerena je na službenom portalu 5. kolovoza 2026. (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Za provjeru koristimo objavljenu tablicu. Izborna datoteka nije preuzeta ni pohranjena uz knjigu, a dostupnost na portalu sama ne znači dopuštenje za njezino daljnje dijeljenje.

Naslov dokumenta, izdavač i stranica omogućuju drugoj osobi da pronađe broj. Bez tih podataka tvrdnja može biti točna, ali provjera zahtijeva dodatno traženje izvora.

U tablici je pet različitih veličina koje bismo u običnom govoru mogli nazvati brojem birača (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Za račun trebamo sačuvati njihove pune oznake.

Tablica 1: Nacionalne vrijednosti izbora za Hrvatski sabor 2024. Izrada autora prema Državno izborno povjerenstvo Republike Hrvatske (2024.), str. 124.

Službena veličina	Vrijednost	Uloga u čitanju
ukupno birača	3 558 089	nazivnik na obrađenim biračkim mjestima
pristupilo glasovanju	2 216 763	brojnik objavljenog odaziva
glasovalo prema glasačkim listićima	2 215 209	kontrola prema pronađenim listićima
važeći listići	2 154 733	dio broja prema listićima
nevažeći listići	60 476	drugi dio broja prema listićima

Objavljeni postotak slijedi iz prvoga i drugoga retka (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

$$\frac{2\,216\,763}{3\,558\,089} \times 100 = 62{,}30\,\%.$$

Za provjeru broja prema listićima trebamo drugu operaciju. Zbrajamo važeće i nevažeće listiće, jer zajedno čine sve pronađene listiće iz tablice (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

$$2\,154\,733 + 60\,476 = 2\,215\,209.$$

Oba su računa točna i odnose se na različite faze postupka. Razlika od 1 554 izražava odstupanje broja pristupilih od broja prema listićima (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Za objašnjenje uzroka, uključujući provjeru moguće pogreške u evidenciji, trebali bismo dodatne podatke. Značenje svake oznake određuje se pravilima evidentiranja, neovisno o tome jesu li zbrojevi jednaki ili različiti.

Pri pisanju naslova biramo podatke, računamo omjer i odlučujemo kako ćemo ga opisati. Na svakom koraku možemo promijeniti značenje, iako nismo promijenili brojeve. Zato uz dijeljenje provjeravamo i odgovaraju li riječi u tvrdnji oznakama iz izvora.

U ovom primjeru možemo povezati izvješće, njegove oznake, nacionalne zbrojeve i omjer s tvrdnjom o pristupanju glasovanju (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Ako asistent vrati samo 62,30 %, nedostaje objašnjenje što je podijelio (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Uz rezultat treba zadržati brojnik, nazivnik i njihov obuhvat kako bi se mogao provjeriti i račun i tumačenje.

Nazivnik, os i rijedak ishod

Uz svaki postotak pitamo na koju se cjelinu odnosi. Odaziv od 62,30 % zato čitamo zajedno s nazivnikom (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Graf koji prikazuje samo postotak izostavlja izvorne brojeve. Ako mu os počinje neposredno ispod opaženih vrijednosti, male razlike mogu izgledati velike. Skraćena os može biti korisna, ali njezin raspon mora biti vidljiv i odgovarati pitanju.

Značenje razlike čuvamo i izborom mjere promjene. Kada dvije stope oduzmemo, razliku izražavamo u **postotnim bodovima**. Za relativnu postotnu promjenu tu razliku još dijelimo početnom stopom. Prvi račun govori koliko su stope udaljene na istoj ljestvici, a drugi kolika je promjena u odnosu na polazište. U nastavku ćemo tu razliku provjeriti usporedbom dvaju obuhvata izbornog izvješća.

Nazivnik odlučuje o značenju postotka i kada procjenjujemo uspješnost nekog postupka. Sintetički zapis stvorio je program. Detektor je postupak koji pokušava prepoznati takvo podrijetlo, a upozorenje je njegova oznaka da bi zapis mogao biti sintetički.

Udio pronađenih sintetičkih zapisa, primjerice, nije isto što i udio ispravnih upozorenja. Uzmimo hipotetski skup od 10 000 zapisa. Program je stvorio njih 100, a ostali bilježe stvarne događaje. Detektor prepoznaje 90 od tih 100 sintetičkih zapisa i pogrešno označava 495 ostalih. Te brojeve namjerno zadajemo za ilustraciju rada zamišljenog sustava. Preostala mjesta dopunjavamo do 100 sintetičkih i 9 900 ostalih zapisa.

Tablica 2: Namjerno zadana brojanja za usporedbu nazivnika. Izrada autora.

Stvarno podrijetlo	Upozorenje	Bez upozorenja	Ukupno
sintetički zapis	90	10	100
stvarni događaj	495	9 405	9 900
ukupno	585	9 415	10 000

Prvi redak daje $90/100 = 90$ % pronađenih sintetičkih zapisa. Drugi daje $495/9\,900 = 5$ % pogrešnih upozorenja među ostalima. Za udio ispravnih upozorenja treba nam prvi stupac. Od ukupno 585 upozorenja ispravno je samo 90, približno 15,4 %. Detektor je pronašao 90 % sintetičkih zapisa, ali većina njegovih upozorenja ipak je pogrešna. Udio pronađenih zapisa razlikuje se od udjela ispravnih upozorenja.

> **Temeljna stopa** jest udio ishoda u relevantnoj populaciji prije nego što se uzme u obzir novi signal, test ili model.

Sintetički zapisi u tom primjeru čine 100 od 10 000 zapisa, odnosno 1 %. To je temeljna stopa. Pri tumačenju upozorenja treba uzeti u obzir koliko je tražena pojava uopće česta.

Ako je u skupu iste veličine 1 000 sintetičkih zapisa, isti detektor pronašao bi njih 900 i pogrešno označio 450 ostalih. Od 1 350 upozorenja sada bi bilo ispravno 900, odnosno dvije trećine. Promijenio se udio sintetičkih zapisa, dok je detektor zadržao istu sposobnost prepoznavanja i istu stopu pogrešnog označavanja.

Da bismo procijenili koliko vjerovati upozorenju, trebamo znati koliko detektor propušta traženih zapisa, koliko pogrešno označava ostale i kolika je temeljna stopa. Tu ćemo vezu ponovno razmotriti u poglavlju o algoritmima, kada klasifikacija vodi odlukama koje različito pogađaju ljude.

I pri čitanju izbornog izvješća rezultat ovisi o odabranom nazivniku. Nevažeće listiće možemo izraziti kao udio svih listića, pristupilih ili svih birača u objavljenom obuhvatu. Svaki omjer ima svoje značenje.

Anketa, obuhvat i biranje trešanja

DIP-ova tablica sažima administrativnu evidenciju izbora. Anketa, naprotiv, koristi odgovore dijela ljudi kako bi zaključivala o široj populaciji. Postotci iz tih izvora zahtijevaju različite provjere.

Pri čitanju ankete ponovno koristimo sedam polja kartice iz prethodnog poglavlja. Bilježimo ciljnu populaciju, okvir iz kojeg su ljudi odabrani i način regrutacije. Tražimo broj pozvanih i odgovora, datume, formulaciju pitanja, ponderiranje, odnosno primjenu težina, neodaziv, naručitelja te objavljenu marginu pogreške. Ono što nije navedeno ostavljamo označeno kao nepoznato.

Za administrativnu evidenciju provjeravamo koji događaji prema pravilima ulaze u nju, jesu li zapisi ispravni, oznake dosljedne i obrada uredna. Za anketu uz mjerenje provjeravamo izbor sudionika, okvir, regrutaciju i neodaziv. Anketa ima i promjenjivost koja nastaje odabirom uzorka. Margina uz znak ± odnosi se na određeni dio te nesigurnosti, pa ostale izvore procjenjujemo zasebno.

Dobrovoljna internetska anketa može prikupiti mnogo odgovora, a ipak izostaviti ljude koji se sustavno razlikuju u onome što mjerimo. Manji vjerojatnosni uzorak može imati veću slučajnu promjenjivost, ali jasnije pokazivati na koga možemo prenijeti nalaz. Veličinu uzorka zato uvijek čitamo uz način odabira.

Učinak odabira redaka možemo provjeriti i u izbornom izvješću. Za jedinice I.–X. zabilježeno je 2 140 824 pristupilih od 3 482 150 birača, što daje 61,48 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Uključimo li sve jedinice I.–XII., odaziv iznosi 62,30 %, odnosno 0,82 postotna boda više (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

Prije tumačenja razlike moramo provjeriti usporedivost nazivnika. Za XI. i XII. jedinicu tablica odaziva navodi jednake

brojeve birača i pristupilih, 41 120 i 34 819. Na prethodnoj stranici zasebni su registrirani brojevi 222 300 birača XI. jedinice i ukupno 275 521 pripadnik navedenih manjina (Državno izborno povjerenstvo Republike Hrvatske 2024., 123–24). Stopa od 100 % u tablici odaziva stoga ne znači da su glasovali svi iz tih registarskih populacija.

Zbroj I.–X. i zbroj I.–XII. opisuju različite obuhvate. Izbor užega nije sam po sebi biranje trešanja. To postaje kad ga biramo radi poželjnijeg rezultata, a prešutimo što izostavljamo i zašto se nazivnici razlikuju. Sljedeća tablica donosi brojnosti po izbornim jedinicama za provjeru zbrojeva iz izvješća (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

Tablica 3: Sastavnice odaziva za provjeru zbrojeva. Izrada autora prema Državno izborno povjerenstvo Republike Hrvatske (2024.), str. 124.

Jedinica	Birači u tablici odaziva	Pristupili
I.	341 023	235 854
II.	345 398	208 757
III.	349 058	218 626
IV.	351 860	208 606
V.	336 652	183 593
VI.	352 234	242 083
VII.	340 923	201 408
VIII.	355 951	211 123
IX.	350 941	208 648
X.	358 110	222 126
XI.	41 120	41 120
XII.	34 819	34 819

U odnosu na zaokruženu stopu I.–X. relativna razlika iznosi $(62{,}30 - 61{,}48)/61{,}48 \times 100 \approx 1{,}33$ %. Time uspoređujemo dva obuhvata istih izbora u istom razdoblju.

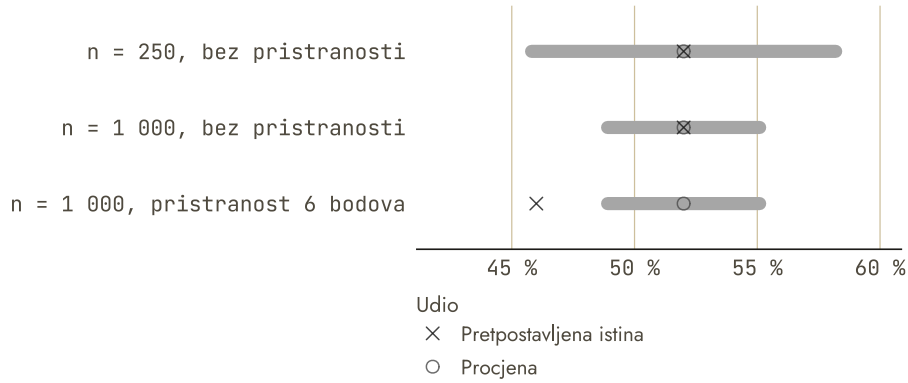
Razlika od 0,82 postotna boda izgledala bi mala na osi od 0 % do 100 %, a mnogo istaknutija na osi od 61 % do 63 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Broj i njegovo značenje ostaju isti. Mijenja se vizualni dojam, zbog čega raspon osi treba jasno označiti i opravdati.

Istraživač margine pogreške

U ovom prikazu udio i veličinu uzorka zadajemo izravno. Početnih 52 % može značiti 130 odgovora od 250 ili 520 od 1 000. Iz zadanoga udjela i veličine uzorka računamo približnu marginu

koju oduzimamo od procjene i dodajemo procjeni. Postupak radi izravno sa zadanim sažetkom ankete. Zadanom pristranošću odvajamo procjenu od pretpostavljene istine, pa se može vidjeti zašto uži interval sam ne popravlja sustavan pomak.

Prikazane su tri konstruirane anketne situacije s približnom 95-postotnom marginom za jednostavan neovisan uzorak. Primjer ne uključuje složen dizajn, ponderiranje, neodaziv ili mjernu nesigurnost. Prve dvije situacije razlikuju se po veličini uzorka. Treća pokazuje kako sustavna pristranost može pomaknuti uzak interval dalje od istine. Prikaz služi razumijevanju uzoračke nesigurnosti i ne izračunava pogrešku DIP-ove tablice. Margina je udaljenost od procjene do svakog kraja simetričnog intervala, prikazanog sivom crtom. Oznaka 95 % opisuje postupak koji bi, uz navedene pretpostavke i mnogo novih uzoraka, obuhvatio pravu vrijednost u približno 95 % intervala. Nije vjerojatnost da je pojedini anketni odgovor točan. Račun ćemo razraditi u poglavljima o uzorkovanju i procjeni.



Slika 1: Tri stanja pokazuju učinak veličine uzorka i sustavne pristranosti. Izrada autora. Podaci i postupak opisani su u tekstu.

U drugoj je situaciji uzorak veći nego u prvoj, pa je interval uži. U trećoj veličina uzorka i procijenjeni udio ostaju kao u drugoj, ali uz pretpostavljenu sustavnu pristranost istinita vrijednost leži izvan intervala.

Veći uzorak smanjuje slučajno rasipanje procjene. Pri sustavnom promašivanju istine veći uzorak daje uži interval oko pomaknute procjene. Preciznost tada prati pristranost.

Podrijetlo ljudskog i strojnog broja

● STATISTIKA U DIVLJINI

Kada „glasovalo" postane preširoka riječ. U zamišljenoj redakciji pripremamo kratku vijest o izlaznosti na izbore. Pred nama je stvarno izvješće DIP-a, u kojem tablica odaziva navodi 62,30 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Naslov „Glasovalo je 62,30 % birača" čini se prirodnim sažetkom tog podatka. Prije objave ipak trebamo utvrditi što glagol glasovalo prenosi iz tablice, a što mu čitatelj može pridodati.

U početnoj smo tablici već razlikovali broj pristupilih od broja birača prema glasačkim listićima (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). U naslovu svaku veličinu navodimo pod njezinim izvornim nazivom, a za tumačenje razlike tražimo dodatne podatke.

Redakcijska provjera zato uz dijeljenje obuhvaća i značenje upotrijebljenih pojmova. Kad bismo odaziv opisali kao udio birača koji su dali važeći glas nekoj listi, promijenili bismo značenje brojnika. Kad bismo nazivnik zamijenili svim punoljetnim stanovnicima, odgovorili bismo na drugo pitanje. Nijednu od tih promjena ne opravdava to što početni postotak dolazi iz službenog izvora.

Vijest zato možemo vezati uz imenovanu mjeru. Prema izvješću DIP-a, glasovanju je pristupilo 62,30 % birača na obrađenim biračkim mjestima (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Uz poveznicu na tablicu čitatelj može provjeriti brojnik, nazivnik i obuhvat. Uredništvo je sačuvalo glavnu informaciju i preciziralo na što se ona odnosi.

Razloge izlaska i političke stavove birača istražujemo zasebno od evidencije izlaznosti.

Asistent može zbrojiti vrijednosti, usporediti nazivnike i predložiti jasniju rečenicu. Svaki dobiveni broj ipak povezujemo s objavom i ulaznim podacima. Provjeravamo što je izračunato, koji je nazivnik korišten i vodi li citat do tog podatka.

NAĐITE GREŠKU

Prije objave vijesti provjeravate asistentov sažetak iste DIP-ove tablice koju smo već čitali. Izvadak zadržava službene oznake i brojnosti (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

Tablica 4: Izvadak službenih izbornih brojnosti. Izrada autora prema Državno izborno povjerenstvo Republike Hrvatske (2024.), str. 124.

Oznaka u izvješću	Broj
Pristupili glasovanju	2 216 763
Birači prema glasačkim listićima	2 215 209
Važeći glasački listići	2 154 733
Nevažeći glasački listići	60 476

> Važećih 2 154 733 i nevažećih 60 476 listića daju ukupno 2 215 209. To je 1 554 manje od broja pristupilih, odnosno približno 0,07 % broja pristupilih. Razlika je tako mala da u vijesti možemo na oba mjesta napisati „pristupili glasovanju”.

Na bilješci odvojeno označite provjerene račune i predloženu uredničku izmjenu. Napišite prihvatljivu zamjenu posljednje rečenice. Ako biste željeli objasniti nastanak razlike, navedite što biste još morali provjeriti u izvoru.

Provjera izvora potrebna je i za slike, zvuk i video. Uvjerljiv zapis može biti umjetno stvoren, a neobičan izgled ne mora značiti da je lažan. Zasebno provjeravamo odakle datoteka dolazi, je li izmijenjena i je li se prikazani događaj zbio. Za svako od tih pitanja potreban je zaseban dokaz.

Za tumačenje materijala razlikujemo sljedeće četiri oznake.

Tablica 5: Četiri vrste dokaznog predmeta i njihove granice. Izrada autora.

Oznaka	Što predmet jest	Što ne smije poduprijeti
simulacija	podaci proizvedeni poznatim mehanizmom radi učenja postupka	nalaz o stvarnoj populaciji
sintetički zapis	umjetno stvoren zapis koji oponaša oblik podatka ili medija	tvrdnju da se prikazani događaj doista zbio
hipotetski izlaz modela	uvjetni odgovor izrađen radi provjere zaključivanja	tvrdnju da je određeni sustav taj odgovor stvarno dao u zabilježenoj uporabi
izmišljeni dokaz	nepostojeći broj, opažanje ili izvor prikazan kao stvaran	bilo koju empirijsku tvrdnju

Primjer s rijetkim sintetičkim zapisima pokazao je zašto sama oznaka detektora nije dovoljna. Detektor sintetičkog sadržaja također proizvodi procjenu. Ona ovisi o podacima na kojima je učen, inačici, postavkama i pragu. Ako je sintetički sadržaj rijedak, čak i umjerena stopa lažnih upozorenja može dati više pogrešnih nego točnih oznaka. Rezultat detektora zato koristimo kao poticaj za daljnju provjeru podrijetla.

I autentičan zapis može imati nepotpun trag podrijetla. Snimka zaslona može izgubiti metapodatke, a izvorna datoteka postati nedostupna. S druge strane, uredni metapodaci ne jamče istinitost sadržaja. Neovisna potvrda događaja također ne dokazuje autentičnost određene snimke. Kad nema dovoljno dokaza,

bilježimo da podrijetlo nije provjereno. Tražimo prvu objavu, poznate izmjene i zasebnu potvrdu događaja.

Generirani podaci u računskom zadatku pripadaju simulaciji. Služe vježbi postupka, bez tvrdnje o stvarnim događajima. Pogrešni odgovor iz okvira Nađite grešku jest hipotetski izlaz modela. DIP-ova tablica službeni je administrativni izvor. Oznaka uz svaki materijal određuje za kakav ga zaključak možemo koristiti.

Razrađeni primjer

Početnu tvrdnju o odazivu sada možemo provjeriti od izvora do konačne rečenice. Već smo razlikovali službene oznake i njihove nazivnike. Preostaje provjeriti slažu li se zbrojevi pojedinih izbornih jedinica s nacionalnim retkom te sačuvati to značenje pri pisanju zaključka.

Vraćamo se tablici „Odaziv birača” na stranici 124 izvješća Državnog izbornog povjerenstva Republike Hrvatske o izborima za Hrvatski sabor 2024. (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Provjera na službenom portalu obavljena je 5. kolovoza 2026., pa uz račun zadržavamo i taj datum (Državno izborno povjerenstvo Republike Hrvatske 2024.). Uz knjigu nije pohranjena izborna datoteka niti se pretpostavlja dopuštenje za njezino daljnje dijeljenje.

Datum pregleda bilježimo kako bi se znalo kada smo provjerili dokument. Ne jamči da će dokument zauvijek ostati isti. Ako se sadržaj na adresi promijeni ili nestane, bez spremljene datoteke i njezina kontrolnog zbroja nećemo moći potvrditi istovjetnost s pregledanom inačicom.

Tablica sadržava dvanaest redaka izbornih jedinica I.–XII. i nacionalni zbroj (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Nacionalni redak služi provjeri zbrojeva, dok jedinice za usporedbu čine pojedini izborni redci. Nazivnik se odnosi na birače na obrađenim biračkim mjestima, pa ga ne

zamjenjujemo drugim ukupnim brojem birača bez provjere obuhvata (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

Zbrojeve za odaziv možemo ponovno provjeriti iz tablice izbornih jedinica. Zbrajanje dvanaest redaka daje 3 558 089 birača i 2 216 763 pristupilih, jednako nacionalnom retku izvješća (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Njihov omjer daje objavljenih 62,30 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124).

Za glasačke listiće ovdje koristimo nacionalne vrijednosti iz početne tablice (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Ondje smo već provjerili njihov zbroj i razliku prema broju pristupilih. Ovaj prikaz ne uključuje zasebnu provjeru svih izbornih i manjinskih redaka za listiće.

Doseg revidirane tvrdnje

Izbor jedinica I.–X. daje 61,48 %, a svih jedinica I.–XII. 62,30 %, uz razliku od 0,82 postotna boda (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Uz rezultat zato uvijek navodimo obuhvat. Uži skup koristimo samo ako imamo zasebno obrazloženo pitanje o tim jedinicama.

DIP-ova tablica omogućuje opis zabilježenog odaziva u navedenom obuhvatu. Redci izbornih jedinica mogu poslužiti za usporedbu ili analizu povezanosti na toj razini. Ne pokazuju zašto je pojedinac izašao, kojoj je listi dao potporu ni što će se dogoditi na budućim izborima. Ovdje ih koristimo da bismo odabrali točnu i provjerljivu formulaciju javne tvrdnje.

Administrativnom ukupnom broju ne pridružujemo marginu pogreške namijenjenu anketnoj procjeni. I dalje mogu postojati nesigurnosti u obuhvatu, definicijama, obradi i zapisima. Za svaku od njih ovo izvješće ne daje posebnu brojčanu mjeru.

Provjerena tvrdnja glasi da je prema izvješću DIP-a za izbore za Hrvatski sabor 2024. glasovanju pristupilo 2 216 763 od 3 558 089

birača na obrađenim biračkim mjestima, odnosno 62,30 % (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Zbroj važećih i nevažećih listića iznosi 2 215 209, što je 1 554 manje (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Oznake ostaju različite jer opisuju različite faze postupka, bez obzira na veličinu razlike.

Tvrdnju bismo ponovno provjerili kada bi se promijenio službeni izvor, obuhvat, značenje oznake ili slaganje sastavnica s ukupnim vrijednostima.

Protokol skeptičnoga čitanja

Provjeru odaziva možemo upotrijebiti i kao obrazac za druge javne tvrdnje. U tablici se vraćamo na šest pitanja iz poglavlja o statističkom mišljenju, ali sada uz svako možemo navesti što smo u izbornom izvješću doista provjerili.

Tablica 6: Šest revizijskih pitanja primijenjenih na isti javni slučaj. Izrada autora.

Revizijsko pitanje	Primjena na tvrdnju o odazivu
Koja je jedinica?	redak izborne jedinice. Nacionalni je redak kontrolni zbroj
Tko ili što nedostaje?	nema individualnih obilježja. Obuhvat su birači na obrađenim biračkim mjestima, a izbor I.–X. ili I.–XII. mora biti vidljiv
Koji je cilj tvrdnje?	opis zabilježenog odaziva, ne tvrdnja o pojedincu ili uzroku
Gdje je neizvjesnost?	aritmetika je provjerena. Obuhvat i moguća pogreška zapisa nisu svedeni na marginu ankete
Koje drugo objašnjenje ostaje?	broj prema listićima za drugo operativno pitanje. Uži obuhvat samo uz unaprijed obrazložen cilj
Tko snosi posljedice?	naslov može zamijeniti službene veličine i čitatelju pripisati zaključak koji tablica ne nosi

Odgovori se nadopunjuju. Provjera jedinice može promijeniti naš izbor nazivnika, a drugo moguće objašnjenje pokazati da je početna usporedba bila preuska. Koliko ćemo provjeravati ovisi i o posljedicama pogreške. Netočan naziv u naslovu i odluka koja nekome uskraćuje pravo imaju različite posljedice, ali u oba slučaja treba znati koga pogreška pogađa.

Zasebno određujemo vrstu tvrdnje i njezin doseg. Podaci mogu dobro opisati administrativno obuhvaćenu populaciju, a da ništa

ne govore o drugim izborima ili uzrocima ponašanja pojedinaca.

Tablica 7: Šest dimenzija tvrdnje na granici Dijela I. Izrada autora.

Dimenzija tvrdnje	Što DIP-ov dokaz dopušta u ovom primjeru
opis	poduprti su objavljeni odaziv i usklađenje službenih veličina
povezanost	samo na razini izbornih jedinica i tek nakon odgovarajuće analize. Nacionalni ukupni redak sam nije povezanost
generalizacija	ne izvan populacije i obuhvata službenog izvora
predviđanje	nije poduprto jer tablica nije model budućih izbora
uzročnost	nije poduprta jer iz tablice ne saznajemo zašto je netko pristupio
odluka	poduprta je omeđena urednička odluka o tome je li javna formulacija provjerljiva i poštena

Uzročno objašnjenje zahtijeva podatke o nastanku opaženih razlika. Za objašnjenje ponašanja stanovnika trebali bismo i podatke na razini pojedinaca. Za odluku uz račun trebamo razmotriti ciljeve i posljedice. Tablica je dovoljna za izbor provjerljive formulacije naslova, dok bi promjena izborne politike zahtijevala dodatne dokaze i vrijednosne kriterije.

Tablica 8: Samoprovjera Dijela I. Izrada autora.

Provjera	Pitanje
brojnik	Koja dva službena brojnika ne smijemo zamijeniti?
anketa	Zašto uzorak od nekoliko tisuća ljudi ne uklanja moguću sustavnu pristranost ankete?
temeljna stopa	Ako su udio pronađenih ciljanih zapisa i stopa pogrešnih upozorenja poznati, koji još podatak treba za udio ispravnih upozorenja?
podrijetlo	Navedite bilo koje dvije zajedničke provjere za broj i sintetičku sliku.

Sažetak

Javni postotak možemo provjeriti kada znamo njegov izvor, jedinicu, brojnik, nazivnik i obuhvat. Promjena bilo koje od tih sastavnica mijenja značenje rezultata, a za tumačenje upozorenja o rijetkom ishodu trebamo i temeljnu stopu. Veći uzorak može smanjiti slučajnu promjenjivost ankete, ali ne uklanja sustavnu pristranost. Račun margine pogreške slijedi u poglavljima o uzorkovanju i procjeni. Odgovore asistenta i sintetičke medije također pratimo do izvora i načina nastanka. Pritom razlikujemo simulaciju, sintetički zapis, hipotetski izlaz i izmišljeni dokaz. Šest revizijskih pitanja pomoći će nam i pri pripremi vlastite tablice te izboru sažetka u sljedećem poglavlju.

Pojmovi

postotak i postotni bod (*percentage and percentage point*), temeljna stopa (*base rate*), margina pogreške (*margin of error*), podrijetlo tvrdnje (*claim provenance*), sintetički zapis (*synthetic record*), protokol skeptičnoga čitanja (*skeptical reading protocol*)

Zadaci

Konceptualni

Objasnite zašto broj pristupilih ne možemo zamijeniti zbrojem važećih i nevažećih listića. Navedite što razlika od 1 554 pokazuje, a što ne možemo zaključiti o njezinu uzroku (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Zatim, bez novog računa, objasnite zašto se u dvama hipotetskim primjerima mijenja udio ispravnih upozorenja, iako detektor zadržava iste stope pronalaženja ciljanih zapisa i pogrešnog označavanja.

Računski

Koristite zbirnu tablicu generiranog skupa `populacija_medija` iz datoteke data/populacija-medija-agregat.csv[1]. U generiranom skupu portal je glavni izvor vijesti za 15 101 od 50 000 zamišljenih osoba, a televizija za 10 827. Izračunajte udjele, razliku u postotnim bodovima i relativnu razliku prema udjelu televizije. Uz rezultate napišite da opisuju simulaciju. Dopušten je kalkulator ili proračunska tablica. Ne predaje se kod.

[1] https://lusiki.github.io/statistika-knjiga/data/populacija-medija-agregat.csv

Kritički

Polazište je nacionalni redak službene tablice „Odaziv birača” (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Usporedite tri ponuđene uredničke prerade toga retka. To su „Na izborima je glasovalo 62,30 % birača”, „DIP bilježi da je glasovanju pristupilo 62,30 % birača na obrađenim biračkim mjestima” i „Većina građana podržala je pobjedničke liste” (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Za svaku navedite je li poduprta prikazanim dokazom. Najbolju preradu doradite tako da uključi brojnik, nazivnik, izvor i jednu važnu granicu.

Razvrstajte generiranu populaciju iz računskog zadatka, pogrešni odgovor modela iz okvira, umjetno stvorenu izbornu fotografiju označenu kao takvu i nepostojeću stranicu izvješća navedenu kao izvor. Svakom predmetu dodijelite oznaku simulacija, hipotetski izlaz modela, sintetički zapis ili izmišljeni dokaz. Objasnite koji zaključak ne možemo opravdati takvim predmetom.

Revizija modela

Provjerite odgovor modela iz okvira. Ponovno zbrojite listiće i izračunajte razliku prema broju pristupilih (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Napišite ispravljeni odgovor u najviše četiri rečenice. Jasno odvojite ono što izvor potvrđuje od objašnjenja za koje bismo trebali dodatni dokaz.



DIO II

opisivanje podataka

Podatke sada sažimamo sredinama, mjerama raspršenosti i grafovima. Svaki prikaz nešto čuva, a nešto izostavlja. Učimo odabrati sažetak prema pitanju i opisati povezanost bez preuranjenoga uzročnog tumačenja.

04

DIO II · OPISIVANJE PODATAKA

sažimanje podataka

Što sredina otkriva, a što raspodjela skriva

Zamislimo da pratite temu koja vas zanima i stalno nailazite na objave istih nekoliko portala. Znači li to da se tom temom bavi malen broj izvora ili da nekoliko velikih izvora zaklanja mnoštvo manjih? Za urednika koji želi opisati prisutnost neke teme na internetu ta razlika mijenja priču.

Takvo ćemo pitanje postaviti podacima projekta DigiKat. U odabrani korpus, odnosno skup tekstova za analizu, ulaze praćene objave s najmanje dva različita katolička pojma s projektnog popisa (Šikić i sur. 2026.). Zamišljeni urednik iz njihovih brojnosti želi doznati koliko objava ima tipičan izvor i gdje se objave najviše koncentriraju.

Kad bismo sve objave ravnomjerno podijelili među izvorima, dobili bismo jedan broj. Bi li nam taj broj pomogao zamisliti izvor na koji ćemo naići među mnogim manjima? I što prije računanja moramo znati o tablici koju smo dobili?

Urednikovo pitanje zahtijeva izbor mjere, ali i pripremu podataka iz kojih ćemo je izračunati. Prije usporedbe prosjeka s drugim sažecima provjerit ćemo usporedivost zapisa te razlikovati odsutno mjerenje od izmjerene nule.

Dosadašnja pitanja o jedinici, nazivniku i podrijetlu podataka sada primjenjujemo pri pripremi vlastite analize. Polazimo od stvarnih zbirnih tablica projekta DigiKat Hrvatskoga katoličkog sveučilišta (Šikić i sur. 2026.). U njima redak predstavlja platformu u određenom razdoblju ili internetsku domenu tijekom obuhvata datoteke. Broj objava, doseg i interakcije različite su veličine, a njihova dostupnost nije jednaka. Provjerit ćemo koji se zapisi mogu spojiti, što pritom mora ostati nepromijenjeno i koje vrijednosti možemo uključiti u sažetak.

Da bismo lakše pratili kako se računaju prosjek i druge mjere, poslužit ćemo se skupom podataka `anketa_mreze`. To je tablica koja izgleda kao rezultat ankete. Svaki redak opisuje zamišljenu osobu, njezinu dob, dnevne minute na društvenim mrežama i povjerenje u tamošnji sadržaj. Podatke smo stvorili računalnom simulacijom, zasebno od populacije medijskih navika iz ranijih poglavlja. Na istim ćemo odgovorima usporediti što nam pojedina mjera govori o tim zamišljenim osobama.

Sredina i medijan opisuju središte raspodjele, a mjere raspršenosti razlike među opažanjima. Njihovu prikladnost provjeravamo prema obliku podataka i reakciji na pomicanje krajnjih vrijednosti. Standardizacijom opisujemo položaj opažanja u odnosu na ostala. Na kraju se vraćamo DigiKatovim domenama i sastavljamo brojčani opis s jasno navedenim značenjem i granicom svake mjere.

Od izvora do sažetka

analysis table

Za vlastitu analizu trebamo *analitičku tablicu*. Pri njezinoj izradi određujemo što predstavlja redak, kojim ključem povezujemo zapise i kako tumačimo kodove. Bilježimo filtre i pravila za

nedostajuće vrijednosti kako bi druga osoba mogla ponoviti postupak.

Tri tablice, tri jedinice

DigiKatov izbor tekstova određuje i granicu našeg zaključka. Tablice opisuju objave obuhvaćene pravilima projekta, a ne slučajan uzorak hrvatskih medija, publike ili korisnika (Šikić i sur. 2026.). Raspolažemo već izračunatim zbrojevima, bez pojedinačnih objava. Prije računanja zato provjeravamo predstavljaju li retci različitih datoteka usporedive jedinice.

Tablica 1: Tri pogleda u paketu DigiKat imaju različite jedinice i ključeve. Izvor: DigiKat, agregatni izvadak, CC BY 4.0 (Šikić i sur. 2026.).

Datoteka	Što predstavlja redak	Ključ	Redaka
mjesečna	platforma u mjesecu	mjesec + platforma	438
godišnja	platforma u godini	godina + platforma	49
izvori	imenovana domena kroz cijeli raspon	izvor	3 604

Datoteka domena nema zajednički ključ s platformskim tablicama pa je analiziramo zasebno. Ona pokazuje raspored objava među domenama. Mjesečne i godišnje podatke o platformama možemo povezati parom `godina + platforma`. Izostavljanjem jednog dijela ključa spojili bismo zapise koji ne pripadaju istoj jedinici.

U mjesečnoj tablici, primjerice, jedan redak bilježi 11 701 obuhvaćenu objavu na webu u siječnju 2021. (Šikić i sur. 2026.). U tablici domena redak za `hkm.hr` bilježi 56 500 objava kroz cijeli obuhvaćeni raspon (Šikić i sur. 2026.). Prvi broj opisuje platformu i mjesec, a drugi jednu domenu kroz više godina. Oba broje objave. Za broj jedinstvenih čitatelja trebali bismo podatke o publici.

Ključ prije sažetka

Mjesečna tablica ima 438 redaka, svaki s jedinstvenim ključem `mjesec + platforma`. Iz godišnje joj tablice pridružujemo godišnji nazivnik i oznaku potpunosti. Nakon ispravnog spajanja broj redaka i ukupan broj objava moraju ostati isti. Ako spojimo samo po godini, svaki mjesečni redak dobit će par sa svim platformama te godine i pojavit će se više puta.

Za dva mjesečna zapisa weba vidimo što je spoj dodao. Mjesec i njegov broj objava dolaze iz mjesečne datoteke, a ukupni godišnji nazivnik iz godišnje. Tražimo podudaranje i godine i platforme. Oba mjesečna zapisa tako dobivaju isti godišnji nazivnik, bez ponovnog brojanja njihovih objava.

Tablica 2: Dva mjesečna retka nakon pridruživanja godišnjeg nazivnika. Izrada autora prema Šikić i sur. (2026.).

Mjesec	Platforma	Mjesečne objave	Godišnji nazivnik
2021-01	web	11 701	90 388
2021-02	web	6 619	90 388

U sljedećoj tablici uspoređujemo broj redaka, jedinstvene mjesečne ključeve i zbroj objava prije i nakon ispravnog spajanja.

Tablica 3: Kontrola retka, ključa i zbroja prije i nakon ispravnog spajanja DigiKatovih agregata. Izvor: DigiKat (Šikić i sur. 2026.).

Postupak	Redaka	Jedinstvenih mjesečnih ključeva	Zbroj objava
Mjesečna datoteka prije spajanja	438	438	710 307
Ispravno: godina + platforma	438	438	710 307

Provjeravamo i pojedine ćelije, jer jednak ukupan zbroj ne jamči njihovo slaganje. Mjesečni i godišnji izvadak razlikuju se u 17 od

49 kombinacija platforme i godine, a odstupanja se u ukupnom zbroju poništavaju. Zato ne zamjenjujemo mjesečne vrijednosti godišnjima i ne gradimo neprekinut trend. Za daljnju analizu koristimo potpunu 2025. godinu, uz isključenje 2024. zbog praznine i promjene metode. U 2024. nedostaju mjesečni retci od veljače do svibnja, siječanj je djelomičan, a lipanj označuje promjenu metode i obuhvata (Šikić i sur. 2026.).

Kada nula znači da mjera nije dostupna

U podacima za 2025. svih devet platformi ima zabilježen broj objava. Za tri od njih nema usporedivog podatka o dosegu i interakcijama. Ondje zapisana nula, uz oznaku `metrika_dostupna = ne`, znači da mjerenje nije dostupno. Izvorni stupac čuvamo, a za analizu izrađujemo novi u kojem te nule označujemo kao nedostajuće vrijednosti.

Primjerice, u mjesečnom retku za `reddit` izvorno je uz interakcije zapisana nula, ali oznaka dostupnosti glasi `ne`. Analitički zapis zato čuva broj objava, a interakcije pretvara u oznaku nedostajanja. Tako redak ostaje dostupan za analizu objava, a prosjek interakcija računamo samo iz dostupnih mjerenja.

To se odnosi na platforme `reddit`, `forum` i `comment`, koje zajedno imaju 7 648 objava. Metapodaci potvrđuju da metrika nedostaje zbog načina mjerenja. Broj objava zato možemo sažeti za svih devet platformi, a angažman samo za šest s dostupnom metrikom. Učinak sužavanja skupa provjeravamo na broju objava, koji poznajemo za sve platforme. Doseg i interakcije nedostajućih platformi ne možemo tako usporediti.

Tablica 4: Osjetljivost medijana broja objava na nepotrebno izbacivanje platformi kojima nedostaje druga metrika. Presjek je potpuna 2025. godina. Angažman i doseg ne uspoređuju se. Izvor: DigiKat (Šikić i sur. 2026.).

Pravilo	Platformi	Medijan objava
Sve platforme za broj objava	9	2 724,0
Samo platforme s metrikom angažmana	6	14 807,5

Medijan, središnja vrijednost u poredanom nizu, mijenja se s 2 724,0 na 14 807,5 objava. Usporedili smo svih devet vrsta platformi sa šest koje imaju metriku angažmana. Promijenio se skup na koji se rezultat odnosi. O dosegu i interakcijama preostalih triju platformi i dalje nemamo podatke.

Za ručnu provjeru poredani brojevi objava po platformama u 2025. iznose 231, 1 641, 1 779, 2 122, 2 724, 4 228, 26 891, 44 502 i 152 048 (Šikić i sur. 2026.). Izbacivanjem platformi bez metrike angažmana uklanjamo vrijednosti 1 641, 1 779 i 4 228 (Šikić i sur. 2026.).

Nazivnik je dio tvrdnje

Za udio trebamo znati i ukupnu količinu. Ako riječi brojimo prema razmacima, ne razlikujući velika i mala slova, *Analysis* se u naslovu *Exploratory Data Analysis* pojavljuje jednom među trima riječima (Tukey 1977.). U naslovu *Statistical Power Analysis for the Behavioral Sciences* također se pojavljuje jednom, ali među sedam riječi (Cohen 1988.). Učestalost je ista, dok se udjeli razlikuju zbog duljine naslova.

Tablica imenovanih domena ima 3 604 redaka i ukupno 551 712 objava. Za udjele objava među tim domenama zato koristimo nazivnik 551 712. Nazivnik 710 307 iz platformskih tablica odnosi se na drugi obuhvat.

Ponašanje sredine i drugih mjera lakše ćemo pratiti na tablici pojedinačnih odgovora. Zato prelazimo na skup `anketa_mreze`,

simuliranu anketu s 300 ispitanika. Jedan redak predstavlja jednu zamišljenu osobu. Za nju su zadani dob, dnevne minute na društvenim mrežama i povjerenje u tamošnji sadržaj na ljestvici od 1 do 10. Skup je generiran zasebno, neovisno o ranije uvedenoj populaciji medijskih navika.

Generator je zadao četiri dobne skupine s 90, 84, 66 i 60 osoba. Unutar svake dob je slučajno odabrana iz njezina raspona, a minute nastaju zasebnim izvlačenjem pozitivnih, desno asimetričnih vrijednosti. Tipična vrijednost minuta niža je u starijim skupinama. Zaokruživanjem zadržavamo barem jednu minutu. Povjerenju također zadajemo niže prosjeke u starijim skupinama, uz slučajna odstupanja i granice ljestvice od 1 do 10.

Nekoliko pojedinačnih zapisa omogućuje da prije prosjeka vidimo što se zbraja. Oznaka ispitanika služi razlikovanju redaka. Dnevne minute i ocjena povjerenja dvije su odvojene varijable.

Tablica 5: Tri zapisa iz zasebne simulirane ankete o društvenim mrežama. Izrada autora.

Ispitanik	Dob	Minute dnevno	Povjerenje (1–10)
I001	22	69	6
I002	22	53	8
I003	24	83	10

Na istoj ćemo anketi poslije usporediti grafove, izračunati povezanost i sastaviti završno istraživanje. Vrijednosti su stvorene tako da većina osoba provodi kraće vrijeme na mrežama, a manji dio znatno dulje. Takav oblik omogućuje da vidimo zašto sredina i medijan daju različite opise. Tumačenje zato ostaje unutar zadanog nastavnog primjera.

Dnevnik transformacija bilježi što smo promijenili između izvora i analize, prema kojem pravilu i s kakvim učinkom na broj redaka ili rezultat. Ovdje u njega pripadaju ključ spajanja, isključeno razdoblje i razlikovanje nedostupne metrike od nule.

Osjetljivost na odluku o podacima znači da provjerimo mijenja li se zaključak uz drugu opravdanu odluku, primjerice uz drukčiji obuhvat dostupnih platformi.

Poštena rečenica izvještaja imenuje jedinicu, obuhvat, mjeru i granicu zaključka. Rečenica o medijanu objava po domeni zato mora reći odnosi li se na sve domene ili samo one s dostupnom metrikom. Uz uzoračku procjenu dodaje se odgovarajući opis nesigurnosti. Precizan broj bez tih podataka ostavlja čitatelju da sam izmisli doseg tvrdnje.

Mjere središta

Prosjek dobivamo kada ukupno vrijeme ravnomjerno raspоredimo među ispitanicima. U formuli n označuje broj opažanja, a x_i vrijednost za pojedinog ispitanika. Crta iznad slova označuje aritmetičku sredinu tih vrijednosti.

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

> Aritmetička sredina je zbroj svih izmjerenih vrijednosti podijeljen njihovim brojem.

Zbroj minuta svih 300 osoba iznosi 15 019. Dijeljenjem toga zbroja s 300 dobivamo 50,06 minuta. Tri prikazana retka samo su početak toga zbrajanja. Svih 300 odgovora ovdje je dostupno.

Sredina iznosi 50,1 minuta dnevno, ali samo 40 % ljudi provodi više vremena od toga. Ukupno smo vrijeme ravnomjerno podijelili među svim ispitanicima, pa su i rijetke velike vrijednosti podigle dobiveni broj. Ako nas zanima osoba u sredini poretka, trebamo drukčiji sažetak. Do njega ćemo doći promatrajući što se događa kada krajnjim vrijednostima postupno smanjujemo utjecaj.

Utjecaj krajnjih vrijednosti možemo ublažiti **skraćenom sredinom** (*trimmed mean*). Poredamo opažanja, uklonimo unaprijed zadan jednak udio s oba kraja i izračunamo sredinu ostatka. U našem uzorku po 5 % znači 15 najmanjih i 15 najvećih vrijednosti, pa novi zbroj dijelimo s 270. Uz po 10 % ostaje 240 vrijednosti. Nakon uklanjanja po 5 % s oba kraja ovdje dobivamo 46,1 minuta, a uz po 10 % dobivamo 44,1.

Nastavimo li uklanjati vrijednosti s oba kraja, približavamo se medijanu, središtu poredanog niza. Različit odnos prema krajnjim vrijednostima možemo izraziti i računom odstupanja. Zbroj kvadriranih odstupanja najmanji je kada ih računamo od sredine. Kvadriranjem velika odstupanja dobivaju veću težinu. Medijan pak daje najmanji zbroj apsolutnih odstupanja, bez njihova kvadriranja. Zato te dvije mjere mogu dati različite vrijednosti središta podataka.

> **Medijan** je vrijednost koja poredani niz opažanja dijeli na dvije jednako velike polovine.

U poredanih 300 odgovora središnja su mjesta 150 i 151. Na njima su 39 i 39 minuta. Njihov prosjek daje medijan. Poredak nam omogućuje pronalaženje središnjih mjesta među zapisima istih osoba.

Medijan je 39 minuta, što je 11,1 minuta manje od sredine. Takav razmak potiče nas da pogledamo graf. Može nastati zbog repa prema velikim vrijednostima, ali i zbog mješavine skupina ili složenijeg oblika raspodjele. Još jedna mjera središta, mod, označuje najčešću vrijednost. Za kontinuirano vrijeme korištenja ovdje nam malo koristi, dok za cjelobrojne odgovore o povjerenju iznosi 5. Kod kategorija mod pokazuje koja je najčešća, bez računanja prosjeka kategorijskih oznaka.

Mjere raspršenosti

Uz središte trebamo opisati koliko se vrijednosti razlikuju. Zamislimo dva portala s istim prosjekom komentara po članku. Na prvom svaki članak ima između 45 i 55 komentara. Na drugom neki nemaju nijedan, a poneki imaju 200. Tu razliku u raspodjeli trebamo prikazati uz zajednički prosjek.

Raspon je razlika najveće i najmanje vrijednosti. Vrijeme u našem uzorku kreće se od 3 do 248 minuta, uz raspon od 245 minuta. Raspon ovisi samo o dvama krajnjim opažanjima. Za raspored ostalih vrijednosti trebamo dodatne mjere.

Za potpuniji opis razmatramo odstupanje svake vrijednosti od sredine. Njegova apsolutna vrijednost daje udaljenost bez obzira na smjer. Prosjek tih udaljenosti, prosječno apsolutno odstupanje, govori koliko se opažanja u prosjeku udaljavaju od sredine.

$$\text{PAO} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

Ovdje prosječno apsolutno odstupanje iznosi 29,4 minuta. Umjesto apsolutnih odstupanja mnogi postupci koriste kvadrirana jer ih je lakše povezati s kasnijim modelima. Tako dolazimo do varijance, koju je nešto teže izravno protumačiti.

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

Za varijancu uzorka zbroj kvadriranih odstupanja dijelimo s $n - 1$. To je Besselova korekcija. Povezana je s time što smo sredinu procijenili iz istog uzorka. U poglavlju o uzorkovanju simulacijom ćemo vidjeti zašto takvo dijeljenje uklanja dugoročnu pristranost procjene.

Uzoračke veličine zapisujemo latiničnim slovima. Sredina je $\bar{x}$, a varijanca s^2. Populacijsku sredinu označujemo grčkim slovom μ, a populacijsku varijancu oznakom σ^2. Veličina uzorka je n, dok N označuje veličinu populacije.

U toj ćemo simulaciji mnogo puta izvući uzorak iz iste poznate populacije i usporediti što se dugoročno događa s djeliteljima n i $n-1$.

Prva osoba ima 69 minuta. Odstupanje od sredine iznosi 18,94, a njegov kvadrat 358,60. Takav kvadrat računamo za svaku osobu, zbrajamo svih 300 vrijednosti i dijelimo s 299. Pri kvadriranju zadržavamo punu preciznost odstupanja. Prikaz na dvije decimale služi čitanju rezultata.

Varijanca iznosi 1 533,3 kvadriranih minuta. Kvadrira se i mjerna jedinica, zbog čega broj nije lako tumačiti u izvještaju. Kvadriranje također povećava utjecaj krajnjih vrijednosti. Dvostruko veće odstupanje pridonosi zbroju četverostruko više.

Standardna devijacija i kvartili

Vađenjem korijena iz varijance vraćamo se na izvornu mjernu jedinicu, ovdje minute.

> **Standardna devijacija** uzorka korijen je varijance. Varijancu računamo dijeljenjem zbroja kvadriranih odstupanja od sredine s $n-1$, a korijen vraća raspršenost u jedinice izvornih vrijednosti.

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

Standardna devijacija ovdje iznosi 39,2 minuta i veća je od prosječnog apsolutnog odstupanja od 29,4 minuta. Kvadriranje

daje veću težinu velikim odstupanjima. Zato standardnu devijaciju možemo samo približno opisati kao tipičnu udaljenost od sredine.

Za normalnu raspodjelu, koju ćemo razraditi u poglavlju o vjerojatnosti, postoji korisno pravilo. U približno normalnoj raspodjeli oko 68 % opažanja nalazi se unutar jedne standardne devijacije od sredine, oko 95 % unutar dviju, a oko 99,7 % unutar triju. U našem je uzorku unutar jedne devijacije 79,3 % ispitanika. Raspon dviju devijacija ide od minus 28,3 do 128,4 minuta. Budući da trajanje ne može biti negativno, taj raspon upozorava da trebamo provjeriti odgovara li normalno pravilo podacima. Sam po sebi ipak ne određuje oblik raspodjele.

Unutar dviju devijacija ovdje se nalazi 95,0 % ispitanika, gotovo kao prema normalnom pravilu. Procjena približne normalnosti traži pregled cijele raspodjele, uz ovo pojedinačno podudaranje.

Ako želimo opis raspršenosti koji manje ovisi o krajnjim vrijednostima, možemo se, kao kod medijana, osloniti na položaje u poredanom nizu. **Percentil** je vrijednost ispod koje se nalazi zadani postotak opažanja, pa je medijan pedeseti percentil. Prvi kvartil odvaja donjih 25 %, a treći gornjih 25 %. Njihova razlika obuhvaća raspon središnje polovine podataka.

Tablica 6: Percentili dnevnog vremena korištenja u simuliranoj anketi. Izrada autora.

Percentil	Minuta
5.	9,0
10.	12,0
25.	22,0
50.	39,0
75.	68,2
90.	103,1
95.	127,3

Taj interkvartilni raspon ovdje iznosi 46,2 minuta, od 22,0 do 68,2 minuta. Pomicanje opažanja koje već leži daleko izvan središnje polovine obično ga ne mijenja. Promjena poretka oko kvartila ipak može utjecati na rezultat. Interkvartilni raspon često navodimo uz medijan, a standardnu devijaciju uz sredinu. Druge kombinacije mogu biti korisne ako objasnimo čemu služe.

Oblik raspodjele i položaj opažanja

Graf nam pokazuje i oblik raspodjele. U simetričnoj raspodjeli dvije strane približno su zrcalne i sredina je često blizu medijana. Dugi rep prema većim vrijednostima može povući sredinu naviše. Njihov razmak ipak nije dovoljan opis oblika. *Asimetrija* uzima u obzir cijelu raspodjelu.

skewness

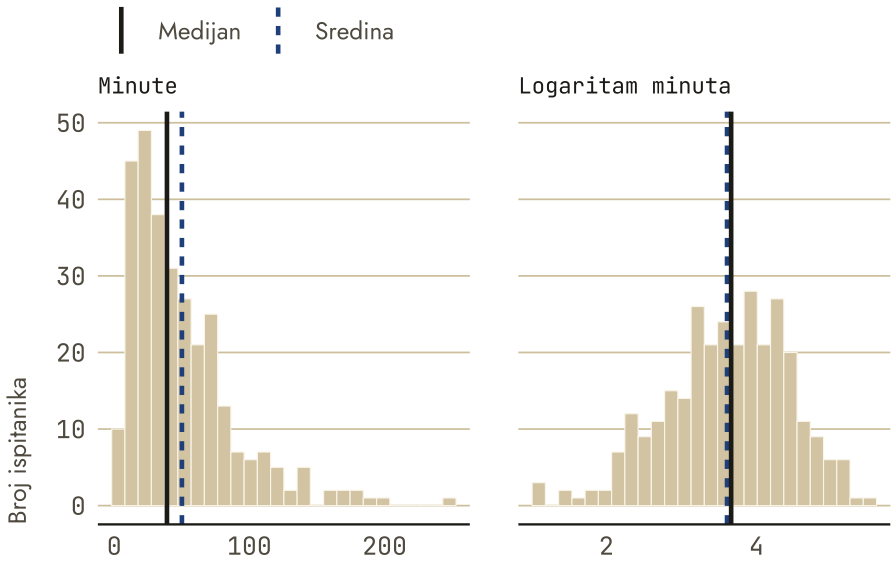
Dugi rep koji smo naslutili iz razmaka sredine i medijana želimo opisati i brojem koji razlikuje dvije strane raspodjele. Zato odstupanja najprije izražavamo u standardnim devijacijama, a zatim ih dižemo na treću potenciju. Neparna potencija čuva njihov predznak. Velika pozitivna odstupanja tako vuku zbroj prema pozitivnome, a velika negativna prema negativnome.

$$\text{asimetrija} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

Asimetrija simuliranog uzorka iznosi 1,61 i proizlazi iz načina generiranja podataka. Opisuje oblik koji smo zadali simulacijom. Ako u izrazu upotrijebimo četvrtu potenciju, dobivamo koeficijent zaobljenosti (*kurtosis*). Njegov višak nad referentnom normalnom vrijednošću ovdje je 3,36. On pomaže pri pregledu repova raspodjele, ali ne objašnjava njihov uzrok ni broj krajnjih slučajeva. Ovdje koristimo prosjek trećih odnosno četvrtih potencija odstupanja standardiziranih uzoračkom standardnom devijacijom, bez korekcije za mali uzorak. Od četvrtog momenta oduzimamo tri. Drugi programski postupci mogu koristiti drukčiju korekciju, pa uz usporedbu vrijednosti treba provjeriti i definiciju.

Podsjetnik. Logaritamska skala

Za strogo pozitivne vrijednosti možemo razmotriti logaritamsku ljestvicu. Ona sabija velike vrijednosti i jednake omjere prikazuje jednakim razmacima. Omjeri 20 prema 10 i 120 prema 60 zato imaju isti razmak. Nulu ne možemo logaritmirati. Dodavanje jedinice svim vrijednostima mijenja analizu i traži obrazloženje.



Slika 1: Dnevno vrijeme korištenja u izvornim jedinicama i na logaritamskoj ljestvici, sa sredinom i medijanom. Logaritamska ljestvica sažima dugi desni rep. Izrada autora.

Isti se odgovor sada zapisuje na drugoj ljestvici. Prvih 69 minuta postaje $\ln(69) = 4{,}23$. Sve su minute pozitivne, pa transformiramo svih 300 odgovora bez izostavljanja. Svaki transformirani zapis i dalje pripada istom izvornom odgovoru iste osobe.

Nakon logaritmiranja asimetrija iznosi −0,44, a sredina i medijan bliski su, 3,61 i 3,66. Sada, međutim, tumačimo drugu ljestvicu. Eksponenciranjem razlike logaritama dobivamo omjer izvornih vrijednosti. Rezultat stoga tumačimo kao višekratnik izvorne vrijednosti.

Uz oblik raspodjele zanima nas i položaj pojedine osobe u njoj. Dnevnih 100 minuta može biti mnogo u skupini s prosjekom od 15 minuta, a manje neobično u skupini s prosjekom od 80. Da bismo opisali položaj, udaljenost od sredine izražavamo brojem standardnih devijacija.

> **Standardizirana vrijednost** kaže koliko standardnih devijacija pojedino opažanje leži iznad ili ispod aritmetičke sredine.

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

Standardizirana varijabla ima sredinu nula i standardnu devijaciju jedan. Tako možemo usporediti relativni položaj osobe prema vremenu korištenja i prema povjerenju, svako unutar njegove raspodjele. Vrijeme i povjerenje pritom zadržavaju svako svoje značenje. Standardizacija čuva oblik raspodjele i ograničenja ljestvice, pa uz krajnje vrijednosti i dalje gledamo graf.

Moramo navesti i referentnu skupinu. Položaj u cijelom uzorku može se razlikovati od položaja unutar dobne skupine, čak i po predznaku. Za najstariju skupinu sredina iznosi 16,2333 minute, a standardna devijacija 9,4750 minuta. Račun koristi nezaokružene vrijednosti.

Tablica 7: Tri ispitanika iz najstarije skupine, standardizirani u cijelom uzorku i unutar vlastite skupine. Izrada autora.

Ispitanik	Minuta	U uzorku	U skupini
I288	39	−0,28	2,40
I251	37	−0,33	2,19
I280	37	−0,33	2,19

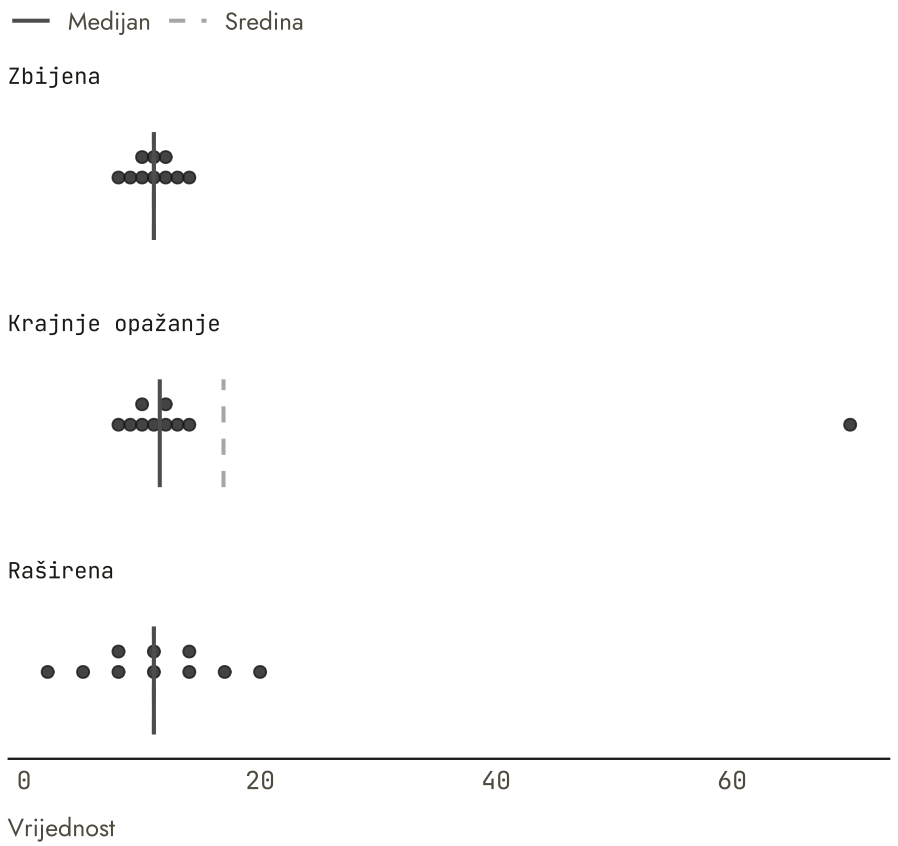
Osobe iz tablice provode manje vremena od prosjeka cijelog uzorka, a u svojoj su dobnoj skupini među najintenzivnijim korisnicima. Obje tvrdnje mogu biti točne jer koriste različite referentne skupine. U izvještaju zato navodimo prema kojoj je skupini položaj izračunat.

Oblikovanje distribucije

Tri raspodjele prikazuju utjecaj krajnjeg opažanja i raspršenosti ostalih vrijednosti. Usporedba sredine, medijana i mjera raspr-

nosti pokazuje koje mjere jače reagiraju na pojedinu vrijednost, a koje više ovise o poretku. Za tu smo usporedbu ručno zadali nizove vrijednosti. Svako je opažanje broj u nastavnoj konstrukciji, pa zaključak vrijedi za zadane nizove.

Svaka točka predstavlja jedno opažanje. Jednake vrijednosti slažemo okomito kako bi sve ostale vidljive. U zbijenom i raširenom stanju sredina i medijan podudaraju se, pa njihove crte leže na istome mjestu.



Slika 2: Tri raspodjele pokazuju odvojene učinke krajnjeg opažanja i raspršenosti. Izrada autora. Podaci i postupak opisani su u tekstu.

Raspršenost možemo usporediti i brojčano. Standardna devijacija koristi djelitelj n − 1, a kvartili linearnu interpolaciju R-ova postupka tipa 7.

Tablica 8: Središte i raspršenost istih triju prikazanih skupova. Izrada autora.

Stanje	Sredina	Medijan	SD	IQR
Zbijena	11,00	11,00	1,83	2,00
Krajnje opažanje	16,90	11,50	18,75	2,75
Raširena	11,00	11,00	5,48	6,00

Uz graf su u tablici navedene vrijednosti za sva tri stanja i njihovi sažeci. Na njima se račun može provjeriti ručno.

Tablica 9: Točne vrijednosti i sažeci triju prikazanih raspodjela. Izrada autora.

Stanje	Vrijednosti	Sredina	Medijan
Zbijena	8, 9, 10, 10, 11, 11, 12, 12, 13, 14	11,0	11,0
Krajnje opažanje	8, 9, 10, 10, 11, 12, 12, 13, 14, 70	16,9	11,5
Raširena	2, 5, 8, 8, 11, 11, 14, 14, 17, 20	11,0	11,0

● STATISTIKA U DIVLJINI

Prosječan izvor i tipična domena. Urednik iz uvodnog primjera želi opisati koliko se katolička tematika pojavljuje u izvorima koje DigiKat prati. Za naslov razmatra rečenicu da prosječan izvor ima 153 objave. Odluka i predloženi naslov zamišljeni su za ovaj primjer, a tablica domena preuzeta je iz stvarnog izvora.

Sredina tablice iznosi 153,1 objave, dok je medijan 4 (Šikić i sur. 2026.). Razlika proizlazi iz dvaju načina sažimanja iste tablice. Kada bi svaka domena nosila jednak dio svih objava, dobila bi vrijednost sredine. Kada domene poredamo prema broju objava i pogledamo sredinu toga niza, dobivamo medijan. Urednik treba odlučiti koju od tih dviju slika želi prenijeti čitatelju.

Raspored objava objašnjava zašto je odluka važna. Prvih deset domena ima 148 748 od ukupno 551 712 objava, odnosno 26,96 % (Šikić i sur. 2026.). Ako

naslov treba približiti tipičnu domenu, sama sredina mogla bi stvoriti dojam da je mnogo objava uobičajeno za većinu izvora. Ako istraživanje zanima gdje se objave koncentriraju, samo medijan ne bi pokazao ulogu najvećih domena.

Urednik zato može uz opis tipične domene prikazati koliko objava nose najveći izvori. Te se informacije dopunjuju. Pritom treba zadržati ograničenje da opisujemo 3 604 imenovane domene uključene prema pravilima DigiKatova korpusa, a ne sve hrvatske medije (Šikić i sur. 2026.). Broj objava također nije broj čitatelja. Iz ovih sažetaka ne zaključujemo koji izvor ima najveći utjecaj ili doseg.

Izbor mjere tako mijenja priču koju ista tablica podržava. Za pitanje o tipičnom izvoru tražimo njegov položaj u raspodjeli. Za pitanje o koncentraciji trebamo pogledati koliko ukupnog broja objava pripada malom dijelu izvora. Za ta dva pitanja biramo različite, jasno imenovane sažetke.

Prije računanja sažetaka trebamo provjeriti što jedan redak predstavlja nakon povezivanja tablica. Sljedeći zapis omogućuje takav pregled.

NAĐITE GREŠKU

Prije sažimanja objava po platformi pregledajte zapis spajanja DigiKatovih tablica (Šikić i sur. 2026.). Redak mjesečne tablice predstavlja platformu u mjesecu, a redak godišnje platformu u godini. U mjesečnoj je tablici jedinstvena kombinacija `mjesec + platforma`. U godišnjoj se ista godina pojavljuje uz više platformi.

Asistent je spojio mjesečnu i godišnju tablicu te priložio provjeru. Zbroj se odnosi na broj objava iz mjesečne tablice.

Tablica 10: Provjera ulazne i spojene tablice. Izrada autora prema Šikić i sur. (2026.).

Provjera	Prije spajanja	Nakon spajanja
Broj redaka	438	3 571
Različite kombinacije mjeseca i platforme	438	438
Zbroj objava	710 307	5 959 081

> Sve početne kombinacije mjeseca i platforme ostale su u rezultatu. Zbroj objava točno odgovara dobivenim redcima. Spoj je zato valjan i tablicu možemo sažeti po platformi.

Odlučite može li tablica u sljedeći korak analize. Pokažite koji podatak podupire vašu odluku, navedite potpuni ključ spajanja i zapišite broj redaka te zbroj objava koje treba dobiti nakon popravka. Zadržite ono što je u asistentovoj provjeri doista točno.

Razrađeni primjer

Na simuliranim minutama vidjeli smo zašto uz sredinu trebamo i druge opise raspodjele. Sada se vraćamo urednikovu pitanju o tipičnom izvoru u DigiKatovu izvatku. Redak predstavlja domenu tijekom cijelog obuhvaćenog razdoblja, a ključ je `izvor`. Datoteka već sadržava samo internetske domene zapisane malim slovima, bez stranica, kanala i osobnih računa. Najprije provjeravamo jedinstvenost domena i ukupan broj objava, a zatim biramo sažetke za opis ovog korpusa. Doseg procjene određen je uključenim domenama i razdobljem korpusa.

Sljedeći programski račun služi dodatnoj provjeri. Objekt `s4_izvori` sadržava retke datoteke data/digikat-izvori.csv[1], a funkcije za sažimanje pripadaju paketu dplyr. Priprema podataka opisana je u mrežnom R praktikumu. Račun i bez

[1] https://lusiki.github.io/statistika-knjiga/data/digikat-izvori.csv

pokretanja koda možemo pratiti prema brojčanim provjerama u tekstu.

```
s4_rezultat <- s4_izvori |>
  summarise(
    izvori = n(),
    jedinstveni_izvori = n_distinct(izvor),
    sredina = mean(objave),
    medijan = median(objave),
    prvih_deset = sum(sort(objave, decreasing = TRUE)[1:10]),
    objave = sum(objave)
  )

stopifnot(s4_rezultat$izvori == s4_rezultat$jedinstveni_izvori,
          s4_rezultat$objave == 551712)
```

Ako se domena ponovi ili zbroj odstupi od 551 712, zaustavljamo račun i provjeravamo razlog. Nakon uspješne provjere možemo sažeti 3 604 različite domene. Sredina i medijan opisuju različite aspekte raspodjele, a udio objava prvih deset domena računamo prema ukupnom broju objava u datoteci.

Tablica 11: Provjereni sažetak objava po imenovanoj domeni. Izvor: DigiKat, datoteka izvora, CC BY 4.0 (Šikić i sur. 2026.).

Provjera	Vrijednost
Jedinstvene domene	3 604
Objave u datoteci	551 712
Sredina objava po domeni	153,1
Medijan objava po domeni	4
Prvih deset domena	148 748 / 551 712 = 26,96 %

Sredina je više od 38 puta veća od medijana. To nije formalna mjera asimetrije, ali uz koncentraciju objava u prvih deset domena pokazuje zašto treba pregledati raspodjelu. Prosjek ne predstavlja tipičnu domenu. Doseg i interakcije izostavljamo jer njihove nule u ovoj datoteci ne možemo pouzdano protumačiti.

Rezultat možemo opisati ovako. U DigiKatovu filtriranom korpusu od 3 604 imenovane domene medijan je četiri objave po domeni. Prvih deset domena ima 26,96 % od 551 712 objava. Opis vrijedi za domene uključene prema pravilima korpusa i ne prenosimo ga na ostale hrvatske medije, objave ili publiku.

Sažetak

Sažetak može odgovoriti na naše pitanje tek kada znamo što predstavlja svaki redak, jesu li zapisi ispravno spojeni, kako su označene nedostajuće vrijednosti i koji nazivnik koristimo. Pogrešan spoj može umnožiti zapise, a isključivanje redaka mijenja skup koji opisujemo, pa ni točan račun ne može nadomjestiti te provjere. Na provjerenoj tablici biramo mjere središta i raspršenosti prema obliku raspodjele. Standardizacijom opisujemo relativni položaj opažanja, a transformacijom mijenjamo ljestvicu. Uz odabrane mjere zato zadržavamo jedinicu i obuhvat,

kao i ono što se jednim brojem ne može opisati. U poglavlju o vizualizaciji graf će nam pomoći vidjeti upravo te pojedinosti.

Pojmovi

analitička tablica (*analysis table*), spajanje (*join*), nedostajuća vrijednost (*missing value*), osjetljivost na odluku o podacima (*data-decision sensitivity*), nazivnik (*denominator*), dnevnik transformacija (*transformation log*), poštena rečenica izvještaja (*honest reporting sentence*), aritmetička sredina (*mean*), skraćena sredina (*trimmed mean*), medijan (*median*), mod (*mode*), varijanca (*variance*), standardna devijacija (*standard deviation*), interkvartilni raspon (*interquartile range*), standardizirana vrijednost (*z-score*), asimetrija (*skewness*), zaobljenost (*kurtosis*)

Zadaci

Za odabrane zadatke ovoga poglavlja na kraju knjige dostupne su naznake i odvojena objašnjenja. Najprije pokušajte samostalno odgovoriti.

Konceptualni

Mjesečna datoteka ima ključ `mjesec + platforma`, a godišnja `godina + platforma`. Objasnite zašto spoj samo po godini može zadržati sva popunjena polja, a ipak promijeniti jedinicu retka. Uz pomoć kontrolne tablice objasnite što provjeravaju broj redaka, broj jedinstvenih mjesečnih ključeva i zbroj objava. Zatim navedite koju od te tri provjere nijedna druga ne može zamijeniti.

Računski

Iz tablice zadanih vrijednosti uzmite stanja *Zbijena* i *Krajnje opažanje*. Iz svih deset vrijednosti ručno izračunajte sredine i medijane te ih usporedite s posljednjim dvama stupcima. U dvije rečenice objasnite zašto se sredina promijenila više od medijana. Za stanje *Zbijena* izračunajte i uzoračku standardnu devijaciju, koristeći djelitelj devet. U tablici raspršenosti usporedite promjene SD-a i IQR-a nakon zamjene krajnjeg opažanja. Koja se mjera više promijenila i zašto? Ako je standardizirana vrijednost neke osobe $z = 1{,}5$, opišite njezin položaj prema sredini bez pretvaranja te vrijednosti u postotak osoba. Zadatak je jednak u digitalnoj i tiskanoj inačici. Interakcija služi samo za dodatne pokuse.

Za drugu provjeru koristimo agregate simulirane ankete, bez dodatnog zaokruživanja. U svakom su retku uz prosjek navedeni brojnik, nazivnik i cjelobrojni zbroj.

Tablica 12: Kontrolni agregat simulirane ankete za ručnu i tiskanu provjeru. Izvor: data/anketa-mreze-agregat.csv, CC BY 4.0.

Šifra	Dobna skupina	Broj	Ukupno	Zbroj minuta	Prosjek minuta	Udio
1	18 do 24	90	300	7 339	81,5444	30 %
2	25 do 34	84	300	4 567	54,3690	28 %
3	35 do 44	66	300	2 139	32,4091	22 %
4	45 i više	60	300	974	16,2333	20 %

Za šifru 1 provjerite `7339 / 90` i `90 / 300`. Zatim zbrojite četiri brojnika i četiri zbroja minuta. Ako koristite datoteku data/anketa-mreze.csv[2], izdvojite šifru 1 i iz tih redaka ponovno izračunajte prvi redak zbirne tablice. Zadatak ne zahtijeva pisanje koda. U tisku su svi potrebni brojnici, nazivnici i provjereni odgovori već u tablici.

[2] https://lusiki.github.io/statistika-knjiga/data/anketa-mreze.csv

Kritički

Urednik objavljuje naslov „Prosječan izvor ima 153 objave”. Na temelju DigiKatova razrađenog primjera napišite tri rečenice. Prva objašnjava što medijan mijenja u čitanju naslova, druga imenuje 551 712 kao nazivnik udjela prvih deset domena, a treća je poštena rečenica s jedinicom, obuhvatom i granicom generalizacije. Ne dodajte tvrdnju o hrvatskim medijima, dosegu ili interakcijama.

Revizija modela

Provjerite modelsku analizu iz okvira. Najprije potvrdite zbroj u tablici koju je model proizveo. Zatim usporedite 3 571 redak s 438 jedinstvenih ključeva i navedite koji dio ključa nedostaje. Napišite ukupan broj objava koji bi dao ispravan spoj. Završite objašnjenjem zašto točan zbroj ne potvrđuje ispravno spajanje.

05

DIO II · OPISIVANJE PODATAKA

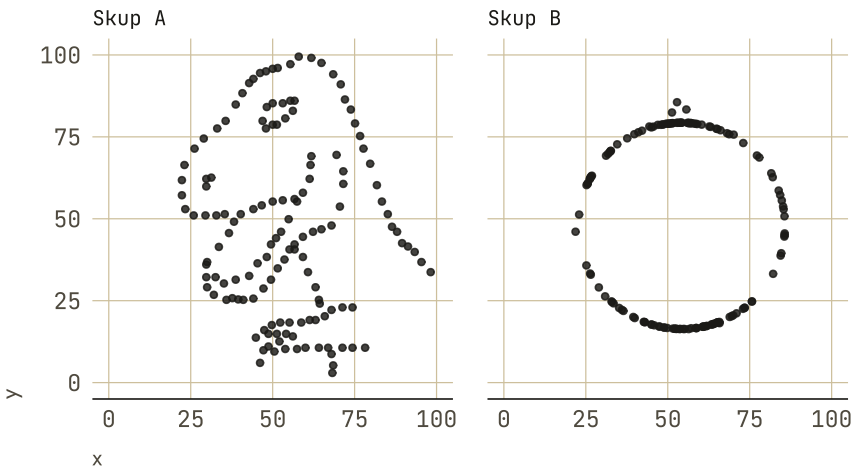
vizualizacija kao argument

Graf pokazuje ono što brojčani sažetak ne može

Netko vam pošalje sažetke dvaju skupova, svakog sa 142 para vrijednosti. Zaokružene na jednu decimalu, sredine su u oba skupa 54,3 za x i 47,8 za y, a standardne devijacije 16,8 i 26,9 (Matejka i Fitzmaurice 2017.). Upravo smo naučili čitati te mjere. Kakav raspored točaka zamišljate iza njih?

Brojevi dolaze iz primjera Datasaurus Dozen, u kojem su podaci namjerno konstruirani da bi ispitali koliko se možemo osloniti na sažetke (Matejka i Fitzmaurice 2017.). Nijedna točka ne predstavlja stvarnu osobu. Prije nego što pogledate graf, pokušajte skicirati dva rasporeda koja bi odgovarala tim brojevima. Koliko različiti smiju biti?

Na slici je svaki par prikazan jednom točkom, na jednakim osima u oba polja. Pogledajte što možemo prepoznati u rasporedu, a što nismo mogli naslutiti iz navedenih sredina i standardnih devijacija.



Slika 1: Dva konstruirana skupa iz primjera Datasaurus Dozen. Izrada autora prema Matejka i Fitzmaurice (2017.) i podacima paketa datasauRus, MIT licenca. Izvorni Datasaurus oblikovao je Alberto Cairo.

Dinosaur i prsten stanu iza gotovo istih brojčanih sažetaka. Podaci su namjerno oblikovani za tu demonstraciju (Matejka i Fitzmaurice 2017.), pa iz njih ne izvodimo tvrdnju o stvarnoj populaciji. U razrađenom primjeru vratit ćemo se istom pitanju kroz Anscombeov kvartet. Ondje ćemo vidjeti kako zakrivljenost i pojedino izdvojeno opažanje mogu promijeniti čitanje veze koju smo saželi brojevima (Anscombe 1973.).

U poglavlju o sažimanju birali smo mjeru prema obliku raspodjele i pitanju koje nas zanima. Pred grafom nas čeka slična odluka. Prikaz prosjeka može olakšati usporedbu skupina, ali sakriti razlike među osobama. Prikaz svih opažanja čuva te razlike, no može otežati čitanje. Želimo naučiti odabrati prikaz iz kojega se tražena usporedba doista može vidjeti.

Uz Datasaurus ćemo upoznati i Anscombeove skupove kao provjeru onoga što brojčani sažetak izostavlja (Anscombe 1973.). To su namjerno konstruirani parovi brojčanih vrijednosti. Jedna

točka predstavlja jedan takav konstruirani par. Služe nam za ispitivanje prikaza, a doseg zaključka ograničen je na taj nastavni primjer. Usporedbom tablica i grafova provjerit ćemo što možemo previdjeti kada preuzmemo samo gotove sažetke.

Vraćamo se poznatoj tablici simulirane ankete o društvenim mrežama, `anketa_mreze`, koju smo upoznali pri sažimanju podataka. Svaki redak i dalje opisuje jednu zamišljenu osobu. Sada ćemo njihove dnevne minute na mrežama prikazati po dobnim skupinama. Uz tu tablicu čitat ćemo i prikaze stvarnih zbirnih podataka projekta DigiKat, vodeći računa o njihovu obuhvatu (Šikić i sur. 2026.). Primjeri se mijenjaju jer nas zanima ista odluka u različitim situacijama. Prije čitanja oznake trebamo znati predstavlja li ona osobu, skupinski sažetak ili zbroj objava.

Izbor prikaza obrazlažemo tako da čitatelj može provjeriti graf i bez programa kojim je nacrtan. Ljestvica, oznake i opis trebaju pokazati obrazac koji podaci podupiru te upozoriti na ono što je sažeto ili izostavljeno.

Gramatika grafike

Grafom odlučujemo što ćemo uspoređivati položajem, duljinom, površinom ili bojom. Izbor zato počinje pitanjem. Za raspodjelu jedne brojčane varijable trebamo vidjeti oblik, za usporedbu kategorija zajedničku osnovicu, a za odnos dviju brojčanih varijabli pojedinačna opažanja. Uredan graf neće pomoći ako skriva ono što trebamo usporediti.

Nazivi poput stupčanog ili raspršenog dijagrama označuju uobičajene kombinacije odluka. Svaki graf možemo rastaviti na te odluke i zasebno ih provjeriti.

> **Gramatika grafike** opisuje od čega je graf sastavljen. Određuje što predstavlja oznaka, koja joj je varijabla pridružena, koji je račun prethodio crtanju i kako se vrijednosti prenose u vizualne veličine.

Iz te je ideje razvijen sustav za sastavljanje grafova (Wilkinson 2005.). Jedna od njegovih poznatih programskih izvedbi jest ggplot2 (Wickham 2016.). Istim se načinom razmišljanja možemo služiti pri čitanju tiskanog grafa, bez programiranja.

Najprije provjeravamo jedinicu. Točka može predstavljati ispitanika, državu, godinu ili stranku. Ako umjesto pojedinaca prikažemo skupinske prosjeke, graf odgovara na drugo pitanje. Zato promjenu jedinice treba jasno navesti.

Geometrija određuje oblik oznake, primjerice točku, stupac ili liniju. Varijable zatim povezujemo s položajem na osima, bojom, veličinom ili oblikom oznake.

> **Pridruživanje** (*aesthetic mapping*) je odluka o tome koja se varijabla prikazuje kojim vizualnim obilježjem. Time određujemo što će čitatelj moći neposredno usporediti.

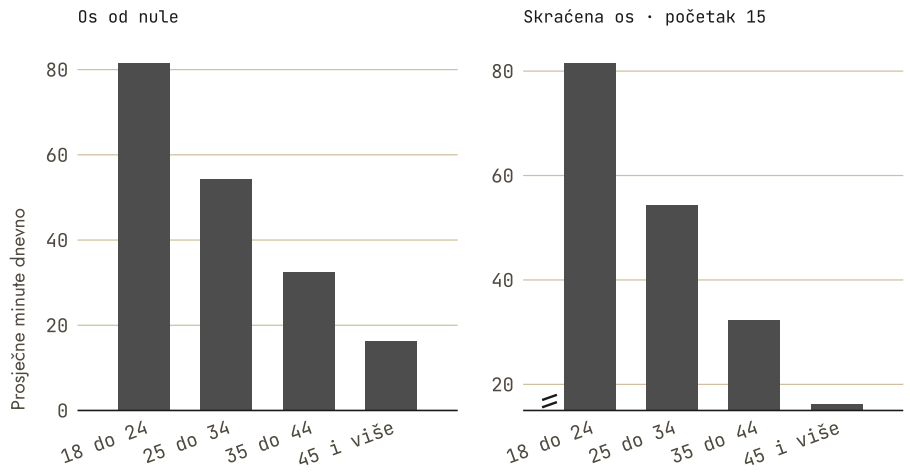
Ako skupine razlikujemo bojom u istom polju, lako ih uspoređujemo na zajedničkim osima. Ako ih rasporedimo u zasebna polja, lakše pratimo obrazac svake skupine. Različitim rasporedom istih podataka naglašavamo različite usporedbe.

Treba provjeriti i što je izračunato prije crtanja. Stupac koji prikazuje prosjek skriva raspodjelu. Okvir s brkovima ne pokazuje broj njezinih vrhova, a izglađena linija dodaje procijenjeni model. Graf zato može biti sažetak koji gubi dio podataka, baš kao brojčana mjera. Anscombeovi skupovi pokazuju zašto treba pogledati što je sažimanjem izostavljeno (Anscombe 1973.).

Ljestvica određuje kako se broj pretvara u vizualnu veličinu. Raspon osi, prekid, logaritamska transformacija ili sredina ljestvice boja mijenjaju prostor koji razlika zauzima. **Koordinatni sustav** određuje kako su položaji raspoređeni. Polarne koordinate, primjerice, uvode kut i mogu otežati usporedbu koju bismo na ravnoj osi lakše očitali.

Pri čitanju grafa zato tražimo jedinicu oznake, pridružene varijable, prethodni račun i pravilo ljestvice. Ta pitanja možemo postaviti svakom grafu, bez pristupa programu u kojem je izrađen.

Na sljedećem prikazu jedan stupac predstavlja dobnu skupinu iz simulirane ankete `anketa_mreze`. Vodoravni položaj označuje skupinu, a okomiti položaj vrha prosječne dnevne minute, izračunate prije crtanja. Stupac zato sažima više osoba i ne pokazuje raspodjelu njihovih odgovora. Lijevo os počinje od nule, a desno je skraćena. Uspoređujemo iste prosjeke, ali promjena ljestvice mijenja dojam njihovih razlika.



Slika 2: Isti prosjeci na dvjema osima. Lijevi prikaz počinje od nule, desni od 15 minuta, ispod najmanjeg prosjeka. Dvije kose crte označuju prekid desne osi. Vrijednosti su jednake u oba prikaza. Izrada autora.

Što oko može očitati

Grafička percepcija jest način na koji iz položaja, duljina i drugih oznaka očitavamo podatke. Neke vizualne usporedbe čitamo točnije od drugih. Pokusi s očitavanjem grafova potvrdili su prednost položaja pred duljinom i kutom (Cleveland i McGill 1984.). Povezivanjem tih rezultata s ranijim istraživanjima nastao je prijedlog šireg poretka vizualnih kanala (Cleveland i McGill 1984.). Prema tom se poretku najtočnije očitava položaj na zajedničkoj osi, zatim na istovjetnim, ali neporavnatim osima, pa duljina, kut, nagib i površina. Obujam, zakrivljenost i zasićenost boje nalaze se pri dnu (Cleveland i McGill 1984.). Cijeli

taj poredak predstavlja sintezu više izvora, s različitim stupnjem izravne pokusne provjere pojedinih usporedbi.

Za usporedbu koja određuje zaključak zato biramo kanal koji se lakše očitava. Boja ili površina mogu poslužiti za razlike kod kojih je dovoljna gruba procjena. Udjele ćemo točnije usporediti stupcima na zajedničkoj osi nego kutovima kružnog dijagrama (Cleveland i McGill 1984.). Ako trebamo samo prepoznati nekoliko izrazito različitih udjela, ta prednost može biti manje važna.

Trodimenzionalna perspektiva može otežati usporedbu ravnih podataka jer duljinu kombinira s površinom i prividnim obujmom, slabijim kanalima u predloženom poretku (Cleveland i McGill 1984.). Zato prije dodavanja treće dimenzije provjeravamo donosi li novu informaciju ili samo mijenja dojam.

Pri oblikovanju grafa korisno je za svaki element razmotriti koju informaciju prenosi (Tufte 2001.). Sjene, obrubi i preljevi mogu odvlačiti pozornost bez korisnog dodatka. Rešetka, razdjelna crta ili oznaka, s druge strane, mogu olakšati čitanje. Uklanjanjem elementa možemo provjeriti otežava li njegov izostanak usporedbu ili snalaženje na grafu.

Tufteov faktor laži uspoređuje relativnu promjenu prikazanu geometrijom s relativnom promjenom u podacima (Tufte 2001.). Kada duljina ili površina izravno predstavlja količinu, omjer blizu jedan pokazuje razmjerno kodiranje. Poštenost cijeloga grafa zahtijeva i provjeru ostalih odluka pri prikazu. Skraćena os, perspektiva ili zamjena duljine površinom mogu povećati dojam i uz točne brojčane oznake. Doseg te mjere ograničen je na određene geometrijske usporedbe.

Prikaz prema tvrdnji

Prednost zajedničke osi pomaže nam izabrati način usporedbe, ali još trebamo odlučiti koje podatke graf mora sačuvati. Za oblik

raspodjele trebamo drukčiji prikaz nego za usporedbu skupinskih prosjeka. Sljedeća tablica povezuje ta pitanja s brojem i vrstom varijabli te pokazuje što pojedini izbor izostavlja.

Tablica 1: Prikaz prema vrsti podataka i prema onome što svaki izbor žrtvuje. Izrada autora.

Što se prikazuje	Uobičajeni izbor	Što prikaz čuva, a što odbacuje
jedna brojčana varijabla	histogram, krivulja gustoće	čuva oblik cijele raspodjele, gubi pojedinačno opažanje
jedna kategorijalna varijabla	stupci na zajedničkoj osi	čuva učestalosti kategorija, izostavlja pojedinačne zapise i njihova druga obilježja
brojčana po skupinama	okvir s brkovima	čuva medijan, kvartile i izdvojena opažanja, odbacuje broj vrhova i položaj većine opažanja
dvije brojčane varijable	raspršeni dijagram	čuva svako opažanje, teško podnosi velik broj točaka
dvije kategorijalne varijable	grupirani ili složeni stupci	čuva odnos udjela, otežava usporedbu unutar složenih stupaca

Svaki prikaz zadržava dio informacija i izostavlja ostale. Okvir s brkovima sažima medijan i kvartile, pa može izgledati jednako

za raspodjelu s jednim ili dvama vrhovima. Histogram raspoređuje opažanja u razrede i prikazuje njihove učestalosti, zbog čega lakše otkriva oblik. Više skupina, međutim, može biti teže usporediti histogramima.

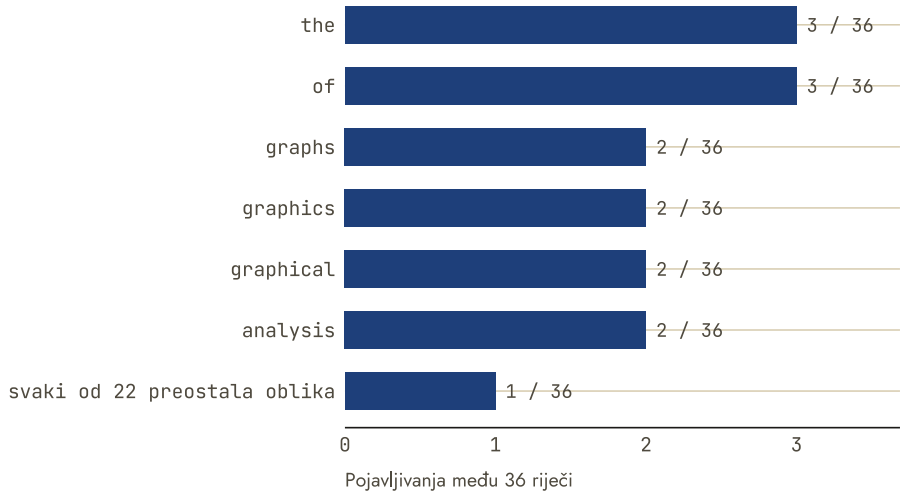
Za dvije kategorijalne varijable možemo izabrati stupce jedan uz drugi, složene stupce ili stupce jednake ukupne visine koji prikazuju udjele. Usporedni stupci pokazuju apsolutne brojeve. Stupci udjela pokazuju sastav, ali skrivaju razliku između skupine od dvadeset i skupine od dvjesto osoba. Za usporedbu brojnosti uz udjele trebamo navesti i veličine skupina.

Kod složenih stupaca samo najdonji segment ima zajedničku osnovicu. Ostali počinju na različitim visinama, ovisno o prethodnim segmentima. Njihove su duljine točne, ali ih je teže usporediti bez zajedničkog početka (Cleveland i McGill 1984.). Ako zaključak ovisi o jednom takvom segmentu, korisno ga je prikazati zasebno.

I riječi možemo uspoređivati kao kategorije, ali tek nakon što odredimo što ćemo smatrati istom riječju. Usporedit ćemo riječi u namjerno odabranim naslovima. Za širi zaključak o literaturi o vizualizaciji ili znanstvenom pisanju trebali bismo odgovarajući izbor tekstova. U šest odabranih naslova o vizualizaciji ukupno je 36 riječi i 28 različitih oblika (Anscombe 1973.; Cleveland i McGill 1984.; Tufte 2001.; Wilkinson 2005.; Wickham 2016.; Matejka i Fitzmaurice 2017.). Do tih brojeva dolazimo tako da sva slova pretvorimo u mala i uklonimo interpunkciju. Riječi ne svodimo na zajednički korijen, pa su `graphs`, `graphical` i `graphics` različite jedinice. Stupci će zato uspoređivati učestalost točnih oblika koje smo tako odredili.

Primjer pretvorbe daje naslov *The Grammar of Graphics* (Wilkinson 2005.). Nakon pretvorbe u mala slova i razdvajanja dobivamo četiri zapisa, `the`, `grammar`, `of` i `graphics`. Svaki dodaje jedan svome brojaču. Kad tako obradimo svih šest odabranih naslova, jedinica brojanja postaje pojedino pojavljivanje riječi. U Wickhamovu naslovu o paketu ggplot2 uklanjamo dvotočku, ali zadržavamo `ggplot2` kao jedan oblik (Wickham 2016.).

Graf izdvaja šest oblika koji se ponavljaju. U zadnjem retku navedena je učestalost koja vrijedi za svaki od preostala dvadeset dva oblika zasebno.



Slika 3: Učestalost točnih oblika u šest namjerno odabranih bibliografskih naslova. Izrada autora prema objavljenim naslovima (Anscombe 1973.; Cleveland i McGill 1984.; Tufte 2001.; Wilkinson 2005.; Wickham 2016.; Matejka i Fitzmaurice 2017.).

Pri tumačenju zato navodimo jedinicu, obradu teksta, nazivnik i obuhvat. Iste ćemo odluke ponovno razmotriti kada tekst postane ulaz algoritma.

Dvije varijable u istom prostoru

Stupcima smo uspoređivali količine po kategorijama. Kada želimo sačuvati odnos dviju brojčanih varijabli za svaku jedinicu, trebamo drugi raspored oznaka. Raspršeni dijagram može prikazati svako opažanje kao točku. Iz rasporeda čitamo smjer i zakrivljenost veze, podskupine i izdvojena opažanja. Zato njime počinjemo provjeru odnosa.

Ako se mnoge točke preklapaju, ne vidimo koliko opažanja stoji na istom mjestu. Djelomično prozirne točke mogu pokazati

gušća područja tamnijima. Kod diskretnih vrijednosti pomaže blago razmicanje točaka, a za vrlo mnogo opažanja možemo izravno prikazati procijenjenu gustoću.

Izglađenu liniju na raspršenom dijagramu dobivamo procjenom modela odnosa iz opaženih podataka. Njezin oblik ovisi o postupku procjene i može djelovati uvjerljivo čak i kad podaci ne podupiru stabilan obrazac. Uvijek je zato čitamo uz oblak točaka. U poglavlju o regresiji razradit ćemo kako se takve linije dobivaju i provjeravaju.

Poštena ljestvica

Vraćamo se simuliranoj anketi `anketa_mreze` iz poglavlja o sažimanju. Prikazujemo prosječne dnevne minute na društvenim mrežama po dobnim skupinama. Raspon osi može znatno promijeniti dojam. Prosjeci se ovdje razlikuju za 65,3 minuta, odnosno 80,1 % najvećeg prosjeka. Ta je razlika stalna, ali njezina veličina na stranici ovisi o osi.

Tablica 2: Broj ispitanika, prosjek i medijan dnevnih minuta u četirima dobnim skupinama simuliranoga nastavnog skupa `anketa_mreze`. Izrada autora.

Dobna skupina	Broj ispitanika	Prosjek	Medijan
18 do 24	90	81,5	68,5
25 do 34	84	54,4	46,0
35 do 44	66	32,4	26,0
45 i više	60	16,2	15,0

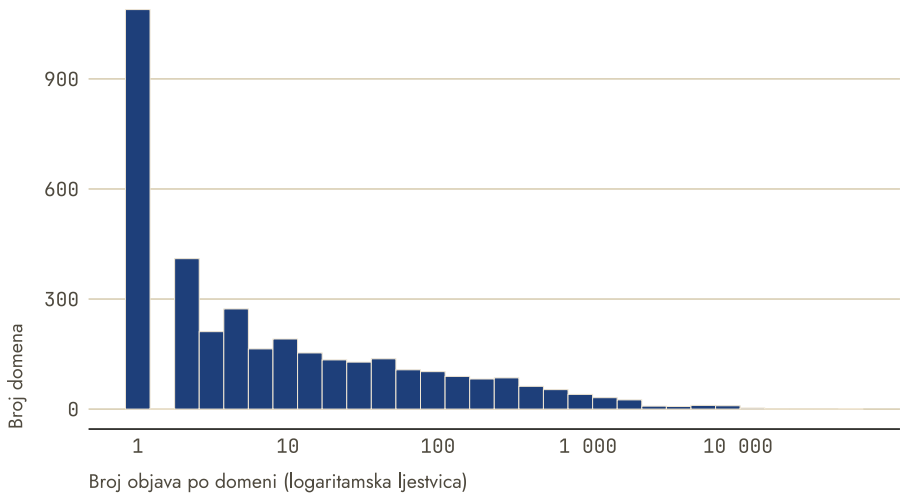
Tablica skupinskih sažetaka čuva vrijednosti za račun, a usporedba dviju osi pokazuje koliko dojam o istoj razlici ovisi o početku osi.

Oba prikaza koriste iste četiri vrijednosti. Desni im dodjeljuje uži raspon osi. Kod stupaca duljina predstavlja vrijednost, pa odsijecanje nule narušava tu razmjernost. U linijskom ili raspršenom

grafu uspoređujemo položaje, zbog čega os može pratiti raspon podataka ako je jasno označena.

Stupce zato crtamo od nule. Oznaka skraćene osi otkriva zahvat, ali ne vraća ispravan odnos duljina. Za linije i točke uži raspon može pomoći čitanju, uz jasno vidljiv početak i ljestvicu osi.

Za sljedeći prikaz vraćamo se istim zapisima iz DigiKatove tablice internetskih domena koju smo upoznali pri sažimanju podataka. Svaki redak sadrži ukupan broj objava jedne domene u odabranom korpusu. Opisujemo samo domene u ovom korpusu, bez zaključivanja o svim hrvatskim medijima, korisnicima ili pojedinačnim objavama. Tablica ima 3 604 domene i 551 712 objava (Šikić i sur. 2026.). Medijan je 4 objave po domeni, a najveći broj 56 500. Na linearnoj bi osi većina domena bila stisnuta uz početak. Logaritamska os omogućuje da se bolje vide i manje vrijednosti jer jednak prostor dodjeljuje jednakim omjerima.



Slika 4: Raspodjela broja objava među 3 604 imenovane domene na logaritamskoj osi. Izrada autora prema DigiKatu (Šikić i sur. 2026.).

Na logaritamskoj osi pomak od 10 do 100 jednak je pomaku od 100 do 1 000. Uspoređujemo omjere, dok apsolutne razlike nisu jednako velike. Promjena osi ne mijenja zapise, ali mijenja vizualni oblik raspodjele, pa je treba označiti.

Mala višestruka polja

Kada je na grafu mnogo skupina, linije se mogu preklapati i tražiti stalno provjeravanje legende. Tada isti oblik grafa možemo ponoviti u zasebnom polju za svaku skupinu.

> **Mala višestruka polja** (*small multiples*) niz su prikaza istoga oblika i iste ljestvice, po jedan za svaku skupinu, tako da se razlike među skupinama očitavaju usporedbom položaja između polja.

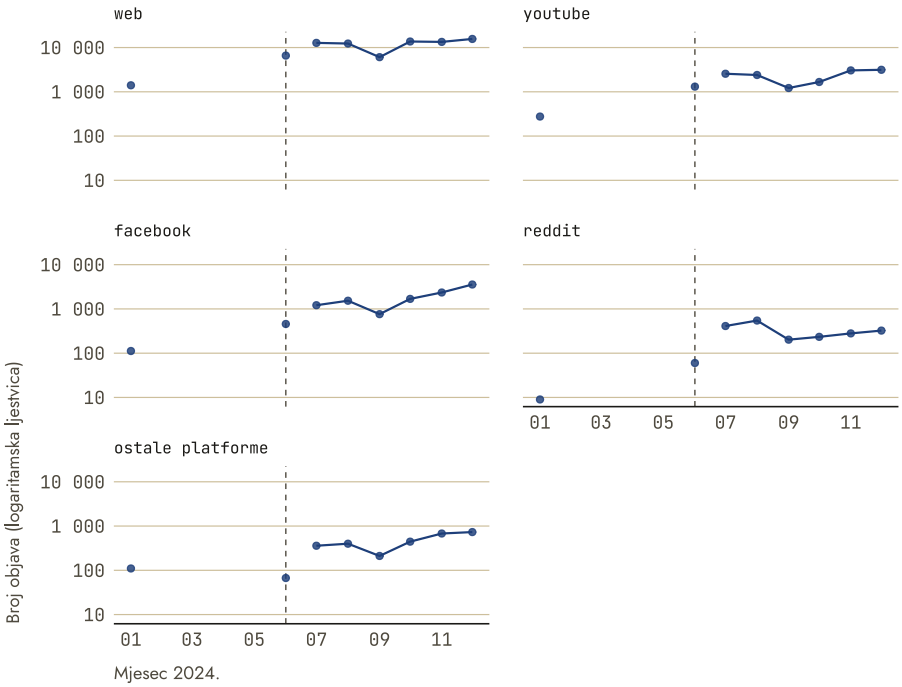
Za usporedbu razina polja trebaju zajedničku ljestvicu. Ako svako ima svoj raspon, male i velike razlike mogu izgledati jednako. Različite osi mogu biti prikladne za usporedbu oblika, ali tu odluku treba jasno navesti.

Iz iste mjesečne tablice projekta DigiKat izdvajamo zapise za promatranje prekida u mjerenju. U DigiKatovu mjesečnom izvatku za 2024. nedostaju retci od veljače do svibnja. Siječanj je djelomičan, a lipanj donosi promjenu metode i obuhvata (Šikić i sur. 2026.). Prazninu ne pretvaramo u nule niti linijom spajamo siječanj i lipanj. U tablici zadržavamo svih 12 mjeseci kako bi nedostajanje bilo vidljivo.

Tablica 3: Broj objava i udio weba u mjesečnom zbroju platformskoga izvatka za 2024. Duga crta označuje mjesec bez retka, ne nulu. Izrada autora prema DigiKatu (Šikić i sur. 2026.).

Mjesec	Objave	Udio weba	Status
01. 2024.	1 911	73,5 %	djelomičan mjesec
02. 2024.	—	—	nema retka
03. 2024.	—	—	nema retka
04. 2024.	—	—	nema retka
05. 2024.	—	—	nema retka
06. 2024.	8 477	77,7 %	lom metode
07. 2024.	17 291	73,7 %	nakon promjene obuhvata
08. 2024.	17 194	71,6 %	nakon promjene obuhvata
09. 2024.	8 458	71,7 %	nakon promjene obuhvata
10. 2024.	17 725	77,3 %	nakon promjene obuhvata
11. 2024.	19 771	67,9 %	nakon promjene obuhvata
12. 2024.	23 404	66,9 %	nakon promjene obuhvata

Brojem objava opisujemo mjesečnu količinu, a udjelom weba njezin sastav. Nazivnik u mjesečnoj tablici zbroj je svih platformi tog mjeseca. Ne koristimo 551 712 iz datoteke domena jer ona ima drugu jedinicu i nema zajednički ključ. U malim višestrukim poljima prikazujemo broj objava za četiri platforme s najviše objava u dostupnim mjesecima 2024. i za objedinjeni ostatak. Sva polja imaju istu logaritamsku os.



Slika 5: Mjesečni broj objava za četiri platforme s najvećim zbrojem u dostupnim mjesecima 2024. i objedinjene ostale platforme. Praznina od veljače do svibnja znači da nema redaka. Isprekidana crta i odvojene točke u lipnju označuju lom metode. Zajednička okomita os je logaritamska. Izrada autora prema DigiKatu (Šikić i sur. 2026.).

Linije ostavljaju četveromjesečnu prazninu, a lipanjska točka odvojena je od kasnijeg niza. Zbog nepotpunog siječnja i promjene metode ne zaključujemo o trendu, rastu ili učinku promjene nakon lipnja. Možemo usporediti dostupne brojeve uz vidljiva ograničenja zapisa.

Graf pred čitateljem

Zajednička ljestvica omogućuje usporedbu samo ako čitatelj može razaznati oznake i njihov smisao. Pri izradi grafa zato treba predvidjeti crno-bijeli tisak, čitače zaslona i različite sposobnosti razlikovanja boja.

Razliku među skupinama zato uz boju označujemo oblikom točke, vrstom linije ili izravnim nazivom. Tako značenje ostaje čitljivo i bez boje. Paleta knjige koristi i razlike u svjetlini kako bi se kategorije mogle razlikovati u sivim tonovima.

Alternativni tekst treba prenijeti nalaz osobi koja ne vidi graf. Navodimo varijable, smjer veze i najizraženije obilježje obrasca. Opis konkretizira odnos, primjerice njegov smjer i oblik, kako bi čitatelj mogao razumjeti nalaz.

Pristupačnost prikaza znači da podatke može pročitati i osoba kojoj boja ili sitan tekst nisu dostupni. Izravnim oznakama uz podatke smanjujemo potrebu za traženjem naziva u legendi. Pomaže i smislen redoslijed kategorija. Kada želimo usporediti veličine, poredak po vrijednosti olakšava čitanje više od abecednog.

Iste odluke provjeravamo i pri čitanju grafa. Primijenimo ih na desno polje u Slika 2.

Jedan stupac predstavlja dobnu skupinu i služi zaključivanju na toj razini. Za opis pojedinog ispitanika trebali bismo njegove odgovore. Skupina određuje vodoravni položaj, a prosjek duljinu stupca. Boja i širina su jednake jer ne prikazuju dodatnu varijablu.

Prije crtanja izračunat je prosjek svake skupine, čime je skriven oblik raspodjele i njezin dugi desni rep. Stupac od 81,5 minuta prikazuje prosjek, dok medijan iste skupine iznosi 68,5 minuta. Os uz to ne počinje od nule, što je u nastavnom primjeru označeno.

Prosjek skriva raspodjelu, a skraćena os prikazuje neprimjerene odnose duljina. Brojevi su točni, ali graf može navesti na pretjeran dojam razlike i pogrešno shvaćanje tipičnog ispitanika. Oznaka zahvata omogućuje da problem uočimo, ali ga ne uklanja.

Histogram grupira brojčane vrijednosti u razrede. Visina stupca ovdje broji opažanja u razredu jednake širine, a drukčija širina razreda može promijeniti dojam o obliku. Krivulja gustoće zaglađuje raspodjelu, pa njezina visina nije broj osoba. Kutijasti

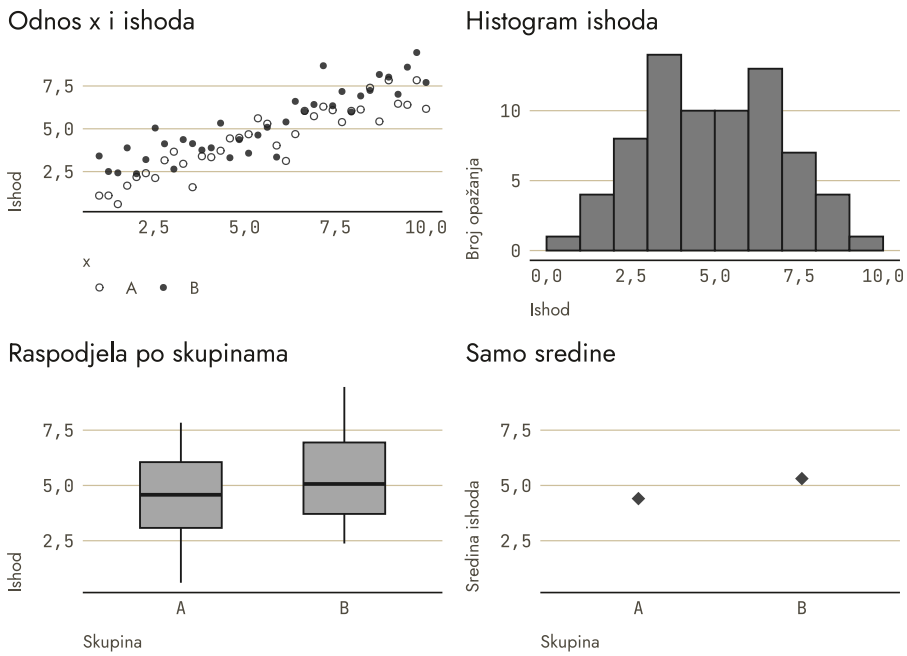
prikaz čuva medijan i srednjih 50 % opažanja između prvoga i trećega kvartila. U sljedećem prikazu brkovi sežu do najudaljenijih opažanja unutar 1,5 interkvartilnog raspona od kutije. Udaljenija opažanja prikazuju se zasebno. Takva točka nije automatski pogreška.

Sažimanje u dvije sredine skriva razlike među pojedincima. Zasebno pitanje jest nesigurnost procjene populacijske sredine. Nju procjenjujemo prema načinu odabira opažanja, što ćemo razraditi uz uzorkovanje i procjenu.

Isti podaci, četiri grafa

Za ovu ilustraciju stvaramo zaseban skup od 72 opažanja, po 36 u dvjema skupinama. U svakoj ravnomjerno rasporedimo vrijednosti x od 1 do 10. Polazište ishoda je $1{,}5 + 0{,}55x$, skupini B dodajemo 0,8, a svakoj točki neovisno normalno odstupanje sa sredinom nula i standardnom devijacijom 0,85. Na primjer, pri $x = 1$ polazište je 2,05 za A i 2,85 za B, prije slučajnog odstupanja. Varijable nemaju stvarnu mjernu jedinicu. Četiri grafa prikazuju isti skup od 72 zapisa.

Isti simulirani podaci prikazani su na četiri načina. Svaki graf olakšava jednu usporedbu i skriva dio drugih informacija. Raspršeni dijagram čuva odnos dviju varijabli, dok prikaz sredina sažima razliku među skupinama. Izbor ovisi o pitanju na koje želimo odgovoriti.



Slika 6: Ista 72 simulirana opažanja u četirima prikazima. Histogram odbacuje x i skupinu, kutijasti prikaz zadržava skupinu, a sredine skrivaju raspršenost. Izrada autora.

Tablica 4: Informacije koje čuva svaki od četiriju prikaza. Izrada autora.

Prikaz	Čuva	Odbacuje
raspršeni dijagram	svih 72 opažanja, pozitivan odnos x i ishoda, nešto viši položaj skupine B, pripadnost skupini	ništa od triju prikazanih varijabli
histogram ishoda	učestalost ishoda u razredima	vrijednost x i pripadnost skupini
raspodjela po skupinama	medijan, kvartile i izdvojena opažanja po skupini	odnos s vrijednošću x i točan položaj većine opažanja
sredine skupina	dvije aritmetičke sredine	raspodjelu, raspršenost, odnos s vrijednošću x i pojedinačna opažanja

STATISTIKA U DIVLJINI

Kružni dijagram pred konkretnim zadatkom. Pretpostavimo da kulturna ustanova priprema prikaz posjećenosti svojih programa. Treba pokazati koji programi okupljaju sličan udio publike, a koji se jasno izdvajaju. Netko predlaže kružni dijagram jer svaki program zauzima dio ukupne posjećenosti. Drugi ga odbacuje pozivajući se na pravilo da se takvi grafovi nikada ne koriste. Prije izbora valja razjasniti što čitatelj treba moći usporediti.

Točnost očitavanja može se provjeriti pokusom. U jednom takvom istraživanju sudionici su procjenjivali omjer dviju istaknutih vrijednosti na grafu (Cleveland i McGill 1984.). Položaj na zajedničkoj osi omogućavao je preciznije usporedbe od kuta (Cleveland i McGill 1984.). Nalaz daje razlog za izbor prikaza, ali taj je razlog vezan uz određeni zadatak.

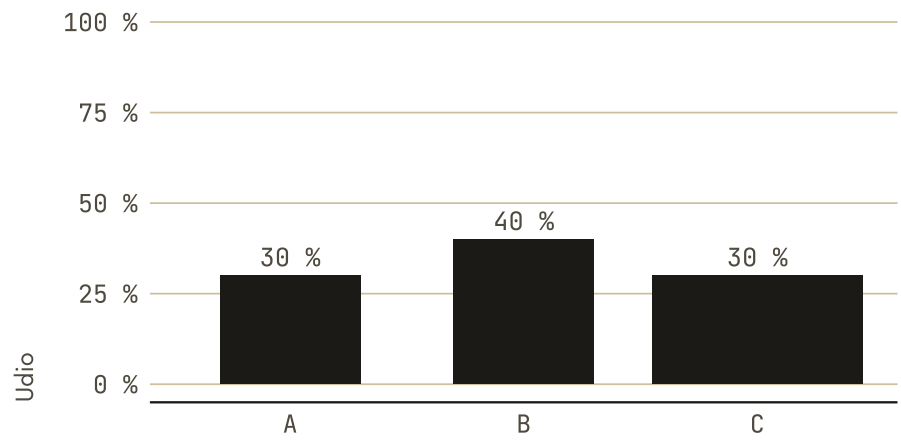
U našem zamišljenom izvještaju osobito nas zanimaju programi s bliskim udjelima. Na kružnici bismo uspoređivali slične kutove u različitim položajima. Na poredanim stupcima s istom početnom osi možemo usporediti njihove završetke, a uz oznake pročitati i točne udjele. Takav izbor olakšava upravo usporedbu zbog koje pripremamo izvještaj. Iz samoga grafa ipak ne saznajemo zašto neki program privlači više ljudi ni zaslužuje li zato više sredstava.

Kad bi nam namjera bila samo pokazati da jedan program zauzima približno polovinu cjeline, zahtjev za preciznošću bio bi drukčiji. Clevelandov i McGillov pokus ne predstavlja ispitivanje svake moguće uporabe, pamćenja ili razumijevanja grafa u stvarnom izvještaju (Cleveland i McGill 1984.). Zato iz njega ne izvodimo zabranu koja bi zamijenila prosudbu o publici i svrsi prikaza.

Ustanova može izabrati stupce i objasniti zašto ih bira. Pritom treba imenovati broji li posjete ili različite posjetitelje, jer dobar oblik grafa ne popravlja nejasnu jedinicu. Prikladnost grafa procjenjujemo prema tome koliko jasno prikazuje usporedbu važnu za naše pitanje.

NAĐITE GREŠKU

Za izvještaj treba usporediti udjele triju kategorija. U ovoj konstruiranoj vježbi kategorijama A, B i C pripada 30 %, 40 % i 30 % iste cjeline. Asistent je iz tih podataka nacrtao sljedeći prijedlog. Pregledajte odnos između oznaka, visina i izgleda stupaca.



Slika 7: Prijedlog prikaza triju udjela za urednički pregled. Konstruirana nastavna vježba, izrada autora.

Odluka o širini vidi se i u dijelu poziva kojim je prijedlog izrađen.

```
ggplot(udjeli, aes(kategorija, udio)) +
  geom_col(width = c(0.6, 0.6, 0.9)) +
  geom_text(aes(label = paste0(100 * udio, " %")), vjust = -0.5) +
  scale_y_continuous(limits = c(0, 1),
    labels = function(x) paste0(100 * x, " %"))
```

Os počinje od nule, sve kategorije imaju oznaku i vrijednosti su prikazane. Šira treća kategorija samo popravlja optičku ravnotežu i ne mijenja dojam o odnosu udjela.

Usporedite posebno A i C. Pripremite uputu osobi koja uređuje graf, s jednom potrebnom izmjenom i dvjema odlukama koje treba zadržati. Objasnite na kojem dijelu prikaza temeljite izmjenu. Ne trebate pisati kod.

Razrađeni primjer

Datasaurus nas je iznenadio oblicima koje sažeci nisu otkrili. Sada ćemo istu pouku provjeriti na Anscombeovu kvartetu, če-

tirima konstruiranim skupovima s gotovo jednakim sažecima (Anscombe 1973.). Njihove ćemo točke rasporediti u mala višestruka polja i provjeriti što zajedničke osi otkrivaju. R-ovi ugrađeni podaci `anscombe` reproduciraju taj objavljeni primjer.

U svakom od četiriju skupova ima 11 parova. Prva tri para iz prvih dvaju skupova pokazuju da ista vrijednost x može biti povezana s drukčijim y.

Tablica 5: Prva tri para u dvama Anscombeovim skupovima iz ugrađenoga skupa u R-u prema Anscombe (1973.).

Redak	x I	y I	x II	y II
1	10	8,04	10	9,14
2	8	6,95	8	8,14
3	13	7,58	13	8,74

U prvom retku oba skupa imaju $x = 10$, ali ishode 8,04 i 9,14. Za sredinu prvoga skupa zbrajamo svih 11 ishoda, dobivamo 82,51 i dijelimo s 11. Isti račun ponavljamo za ostala tri skupa, a grafovima zadržavamo svaki izvorni par.

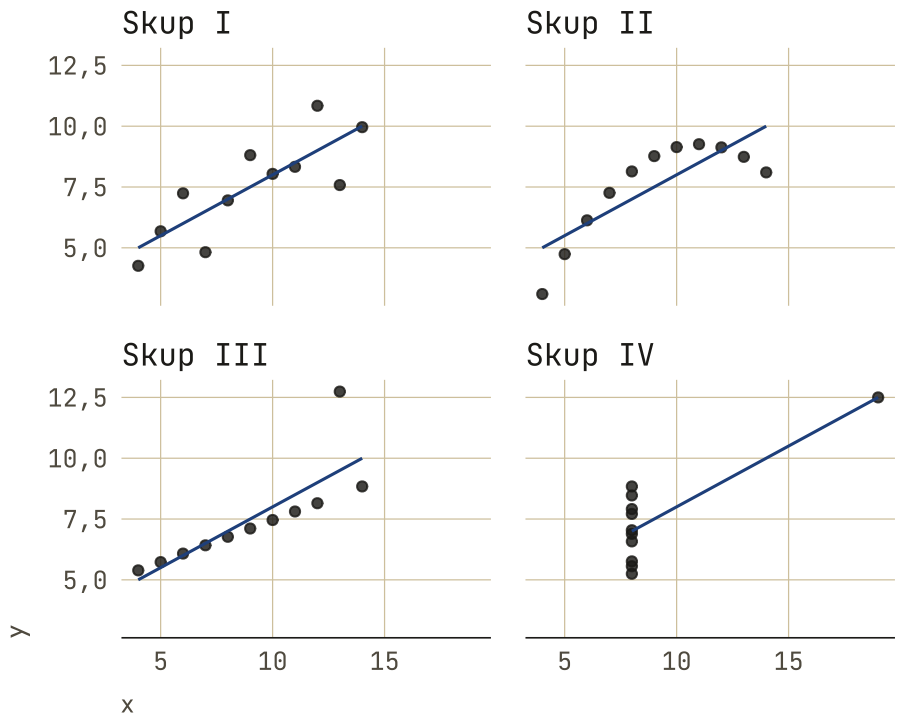
Kod u ovom primjeru koristi pakete ggplot2, dplyr i tibble, učitane u pripremi knjige. Njihova je priprema opisana u mrežnom R praktikumu. Najprije provjeravamo sažetke. Zaokružene na dvije decimale, sve četiri sredine ishoda iznose 7,50, a standardne devijacije 2,03. Iz tih se vrijednosti razlike u obliku ne mogu prepoznati.

```
anscombe_dugo <- bind_rows(
  tibble(skup = "Skup I", x = anscombe$x1, y = anscombe$y1),
  tibble(skup = "Skup II", x = anscombe$x2, y = anscombe$y2),
  tibble(skup = "Skup III", x = anscombe$x3, y = anscombe$y3),
  tibble(skup = "Skup IV", x = anscombe$x4, y = anscombe$y4)
)

p_anscombe <- ggplot(anscombe_dugo, aes(x, y)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(~ skup, ncol = 2)
```

U prvom dijelu postupka svaki par vrijednosti dobiva vlastiti redak, kako bi u drugome mogao postati jedna točka. Poziv `aes` određuje koja vrijednost ide na koju os, a `geom_point` crta te točke. Da bismo usporedili oblik podataka s njihovim linearnim sažetkom, `geom_smooth` dodaje procijenjeni odnos, dok `facet_wrap` odvaja skupove u zasebna polja. Upisom `p_anscombe` prikazujemo pohranjeni graf. Dodani pravci gotovo se podudaraju, kao u izvornom primjeru (Anscombe 1973.). U kasnijim je primjerima dovoljno prepoznati koju od tih odluka pojedini naziv provodi.

Uz gotovo jednake sažetke četiriju skupova (Anscombe 1973.), graf koristi jednake osi i gotovo podudarne pravce. Tako razlike među poljima proizlaze iz rasporeda opažanja.



Slika 8: Anscombeov kvartet s gotovo jednakim sažecima i različitim oblicima. Izrada autora prema Anscombe (1973.).

Prvi skup približno prati pravac, a drugi zakrivljenu vezu. U trećem i četvrtom izdvojeno opažanje snažno utječe na pravac. Brojčani sažetak zato čitamo zajedno s oblikom podataka koje sažima.

Sažetak

Graf počinje od usporedbe koju želimo omogućiti, pa jedinicu oznake, varijable, prethodni račun i ljestvicu biramo zajedno. Tim izborom dio podataka pokazujemo, a dio sažimamo ili izostavljamo, kao što su gotovo jednaki Anscombeovi sažeci skrivali različite oblike. Položaj i drugi vizualni kanali ne omogućuju jednako precizno očitanje (Cleveland i McGill 1984.), pa zajedničke osi, prikladno razdvajanje skupina i oznake uz boju pomažu

da tražena usporedba ostane čitljiva. Kada prikazujemo riječi, isto obrazloženje obuhvaća pravilo brojanja i izbor tekstova. U poglavlju o povezanosti krenut ćemo od raspršenog dijagrama i provjeriti koliko njegova oblika možemo sažeti jednim koeficijentom.

Pojmovi

gramatika grafike (*grammar of graphics*), pridruživanje (*aesthetic mapping*), geometrija grafa (*geom*), ljestvica (*scale*), grafička percepcija (*graphical perception*), prikaz učestalosti riječi (*word-frequency plot*), logaritamska ljestvica (*logarithmic scale*), skraćena os (*truncated axis*), mala višestruka polja (*small multiples*), pristupačnost prikaza (*visualization accessibility*), alternativni tekst (*alternative text*), Anscombeov kvartet (*Anscombe's quartet*), utjecajno opažanje (*influential observation*)

Zadaci

Konceptualni

Odaberite prikaz za raspodjelu jedne varijable, usporedbu kategorija i odnos dviju brojčanih varijabli. Za svaki obrazložite izbor i navedite što izostavlja. Za jedan graf napišite alternativni tekst koji prenosi nalaz te provjerite je li nalaz razumljiv i bez boje.

Računski

Iz tablice skupinskih sažetaka uzmite najveći i najmanji prosjek dnevnih minuta. Izračunajte njihovu razliku, a zatim je podijelite s većim prosjekom i pretvorite u postotak. Usporedite

dobiveni postotak s dvama prikazima osi i objasnite zašto se brojčana razlika nije promijenila, premda se promijenio vizualni dojam. Predajte račun, postotak i dvije rečenice prosudbe. Svi potrebni podaci nalaze se u tablici.

Kritički

Pronađite objavljeni graf sa skraćenom osi i prosudite je li odsjecanje opravdano. Odredite koliki bi udio visine osi zauzela razlika kada bi donja granica bila nula, a gornja ostala ista, je li prekid vidljivo označen i mijenja li se zaključak teksta uz graf. Predajte odlomak s presudom i s uvjetom pod kojim bi presuda bila suprotna.

Revizija modela

Provjerite prijedlog grafa iz okvira o pogrešci. Navedite ispravne grafičke odluke i izdvojite onu koja obmanjuje. Opišite što treba promijeniti da udio ponovno bude prikazan samo duljinom stupca. Opis predajte bez pisanja ili popravljanja koda.

06

povezanost

Sažetak odnosa bez cijele slike

Zamislimo večer u kojoj se nakon duljeg pregledavanja objava osjećamo lošije nego prije. Možda nas je uznemirio sadržaj. Možda smo već bili neraspoloženi pa smo zato ostali uz ekran. Kad bismo pitali prijatelje koliko vremena provode uz digitalne medije i kako se osjećaju, bismo li iz njihovih odgovora mogli razlučiti što se događa?

Analiza triju velikih skupova podataka o adolescentima povezala je uporabu digitalne tehnologije s pokazateljima njihove psihološke dobrobiti. Istraživači su usporedili niz opravdanih analitičkih izbora, uključujući različite mjere i varijable uključene u račun (Orben i Przybylski 2019.). Jedno široko pitanje tako je postalo skup konkretnih usporedbi, od kojih svaka nešto zadržava, a nešto izostavlja.

Što nam povezanost vremena uz tehnologiju i dobrobiti može reći o adolescentima, a što bismo još trebali znati da nekome savjetujemo promjenu navike?

Sada kada znamo pročitati raspršeni dijagram, možemo preciznije pitati kako se dvije varijable mijenjaju zajedno. U dosadašnjim smo sažecima dnevne minute i dob često promatrali odvojeno. Zanima nas i njihov odnos. Želimo ga opisati usporedivim brojem, a pritom sačuvati ono što smo iz grafova naučili o zakrivljenosti, podskupinama i izdvojenim opažanjima.

Glavni primjer ostaju isti zapisi iz simulirane ankete `anketa_mreze`, koje smo sažimali i prikazivali grafovima. Jedan redak predstavlja jednu zamišljenu osobu, pa njezinu dob povezujemo s njezinim dnevnim vremenom na društvenim mrežama. Za pregled više parova uključit ćemo i povjerenje u sadržaj. Podaci su konstruirani za nastavu. Na njima učimo opisivati odnos, bez zaključka o stvarnim navikama ili o tome što bi promjena vremena na mrežama izazvala.

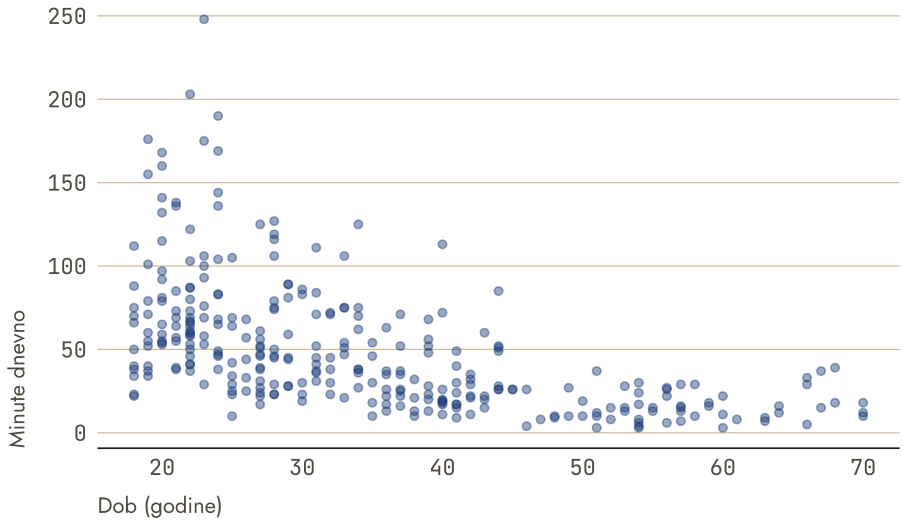
Mjeru zajedničkog kretanja gradimo od odstupanja od sredine i standardizacije, koje već poznajemo. Pearsonova korelacija sažima linearnu vezu, a Spearmanov koeficijent polazi od poretka vrijednosti. Izdvojena točka, uži raspon dobi ili spajanje podskupina mogu promijeniti koeficijent i njegov smisao. Usporedbom tih slučajeva vraćamo se problemu zbirne slike iz Berkeleyjskog slučaja, sada s dvjema brojčanim varijablama.

Provjeru ćemo proširiti na stvarne Eurostatove pokazatelje obrazovanja i uporabe interneta, gdje jedinica postaje država umjesto osobe (Eurostat 2026.d, 2026.c). Ta promjena određuje što povezanost smije značiti. Uočen odnos među državama nećemo pripisati njihovim pojedinim stanovnicima. Prikladnost koeficijenta provjeravamo uz graf i opis odabranih jedinica. Tako vidimo koliko dobro sažima odnos i što izostavlja. Za pitanje o uzroku trebamo odgovarajući nacrt i dodatne podatke, a počinjemo pregledom oblika odnosa koji već vidimo.

Zajedničko kretanje

Dvije su varijable povezane kada se vrijednosti jedne razlikuju ovisno o vrijednostima druge. Pri pozitivnoj vezi veće vrijednosti jedne češće dolaze uz veće vrijednosti druge. Pri negativnoj dolaze uz manje vrijednosti druge. Koeficijent blizu nule može značiti slabu linearnu vezu, ali pritom može postojati zakrivljen odnos.

Na grafu dobi i dnevnih minuta iz ankete `anketa_mreze` svaka točka povezuje vrijednosti iste zamišljene osobe. Točke tvore silazni oblak. Pad je strmiji među mlađima i poslije se izravnava. Graf nam pokazuje taj oblik prije računanja koeficijenta.



Slika 1: Dob i dnevno vrijeme korištenja u simuliranoj anketi. Oblak pokazuje silazan, zakrivljen obrazac. Pearsonov koeficijent sažima samo njegov linearni dio. Izrada autora. Podaci i postupak opisani su u tekstu.

Najprije uspoređujemo na koju stranu od prosjeka odstupa svaki ispitanik u objema varijablama. U poznatoj tablici simulirane ankete `anketa_mreze` s 300 ispitanika, 27 % odstupa na istu stranu u dobi i dnevnim minutama. Uz simetrične raspodjele i neovisna odstupanja očekivali bismo udio blizu polovine. Asimetrija ga

može promijeniti i bez povezanosti, pa sam udio nije mjera veze. U ovom primjeru češća su odstupanja na suprotne strane, što odgovara negativnom obrascu na grafu. Za jačinu veze trebamo uz smjer uzeti u obzir i veličinu odstupanja.

Zato za svaku osobu množimo odstupanja od dviju sredina. Umnožak je pozitivan ako oba odstupanja imaju isti predznak, a negativan ako imaju suprotne. Velika odstupanja pridonose više od malih. Zbrajanjem tih umnožaka i dijeljenjem dobivamo mjeru zajedničkog kretanja.

> **Kovarijanca** uzorka je zbroj umnožaka odstupanja dviju varijabli od njihovih sredina, podijeljen brojem opažanja umanjenim za jedan. Pozitivan predznak upućuje na prevladavanje odstupanja u istom smjeru, a negativan u suprotnom.

Zapisano simbolima, gdje x_i i y_i označavaju vrijednosti izmjerene kod istog opažanja, $\bar{x}$ i $\bar{y}$ njihove sredine, a n broj opažanja, kovarijanca glasi

$$\operatorname{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Djelitelj $n - 1$ isti je onaj iz poglavlja o sažimanju podataka, a formula je poopćenje varijance, jer varijanca nastaje kada se ista varijabla stavi na oba mjesta.

Za provjeru računa potrebne su obje sredine i standardne devijacije. U anketi je sredina dobi 34,517, a njezina standardna devijacija 13,384 godina. Za dnevne minute odgovarajuće vrijednosti iznose 50,063 i 39,157. Račun u nastavku koristi nezaokružene vrijednosti. Prikaz na tri decimale služi ručnoj približnoj provjeri.

Na prvoj osobi iz poznate ankete možemo pregledati jedan član zbroja. Dob je 22, a vrijeme 69 minuta. Odstupanja od dviju sredina iznose −12,52 godina i 18,94 minuta. Njihov je umnožak −237,02. Račun ponavljamo na svih 300 parova, zbrajamo um-

noške i dijelimo s 299. Svi su parovi potpuni, a dob jedne osobe uvijek ostaje uz njezine minute.

Veličina kovarijance ovisi o mjernim jedinicama. Za dob i dnevne minute ovdje iznosi −293,1. Ako iste minute pretvorimo u sate, dobivamo −4,9, šezdeset puta drukčiju vrijednost. Odnos je ostao isti. Kovarijancu zato moramo čitati uz jedinice i ne možemo njome izravno uspoređivati parove mjerene različitim jedinicama.

Za usporedbu različitih parova potrebna nam je mjera neovisna o jedinicama.

Od standardizacije do matrice

Standardizacijom vrijednosti pretvaramo u broj standardnih devijacija od sredine. Ako to učinimo s objema varijablama prije množenja, račun daje jednak rezultat za vrijeme izraženo u minutama i u satima.

> **Pearsonova korelacija** uzorka jest zbroj umnožaka uzorački standardiziranih vrijednosti dviju varijabli, podijeljen s $n - 1$. Mjeri smjer i jačinu njihove linearne veze na ljestvici od -1 do $+1$, neovisno o mjernim jedinicama.

Uzoračku korelaciju označujemo s r, a populacijsku s ρ. Standardizirane vrijednosti opažanja i jesu z_{x_i} i z_{y_i}, dok su s_x i s_y uzoračke standardne devijacije. Korelaciju možemo zapisati na dva jednaka načina

$$r = \frac{1}{n-1} \sum_{i=1}^{n} z_{x_i} \, z_{y_i} = \frac{\mathrm{Cov}(x, y)}{s_x \, s_y},$$

Dva zapisa daju isti broj. U prvome standardiziramo prije množenja, a u drugome kovarijancu dijelimo standardnim devijacijama. Računamo isto, samo drugim redoslijedom.

Nakon standardizacije prva osoba ima $z_x = -0{,}94$ i $z_y = 0{,}48$, pa u zbroj ulazi njihov umnožak −0,452. Svih 300 takvih umnožaka opet dijelimo s 299, jer smo standardizirali uzoračkim standardnim devijacijama.

U anketi račun preko standardiziranih vrijednosti daje −0,559, a ugrađeni izračun korelacije −0,559. Ako umjesto minuta koristimo sate, rezultat ostaje −0,559. Kovarijanca se pri toj promjeni jedinice promijenila šezdeset puta, a korelacija ostaje ista.

Kada je korelacija $+1$, sve točke leže na uzlaznom pravcu, a pri -1 na silaznom. Vrijednost nula znači da nema linearne veze. Međuvrijednosti opisuju koliko raspored točaka prati pravac. Nagib pravca odgovara na zasebno pitanje, za koliko se minuta vrijeme korištenja mijenja uz jednu godinu dobi. Uz korelaciju −0,56 za taj ćemo opis trebati i regresiju.

Kao gruba orijentacija u društvenim znanostima često se navode korelacije oko 0,10, 0,30 i 0,50 za male, srednje i velike odnose (Cohen 1988.). Te su vrijednosti predložene za situacije u kojima nemamo bolju osnovu u poznavanju područja (Cohen 1988.). Važnost koeficijenta zato procjenjujemo prema pojavi, načinu mjerenja, posljedicama i nalazima usporedivih istraživanja.

U jednostavnom linearnom opisu kvadrat korelacije pokazuje udio varijance povezan s tim opisom. Za korelaciju −0,56 to je 31 % varijance. Preostalih 69 % linearni opis ne obuhvaća. Primjerice, korelacija 0,30 daje devet posto. U poglavlju o regresiji tu ćemo veličinu koristiti za opis prilagodbe modela.

Uz veličinu veze treba procijeniti nesigurnost. Korelacija iz uzorka može odstupati od populacijske, a uz isti nacrt i usporedive neovisne jedinice veći uzorak obično daje precizniju procjenu. Vrijednost 0,40 na trideset i na tri tisuće osoba zato nema jednaku nesigurnost. Uz koeficijent uvijek navodimo broj opažanja, a račun nesigurnosti slijedi u poglavljima o uzorkovanju i procjeni.

U poznatoj anketi uz dob i dnevne minute imamo i povjerenje, pa možemo promatrati više odnosa. Umjesto da koeficijente zapisujemo odvojeno, slažemo ih u matricu u kojoj svaki redak i stupac

pripada jednoj varijabli. Isti se par pojavljuje s obje strane dijagonale, pa je matrica simetrična. Na dijagonali su jedinice jer je svaka varijabla savršeno povezana sama sa sobom.

Tablica 1: Korelacije triju brojčanih varijabli simulirane ankete. Izrada autora.

	dob	minute_dnevno	povjerenje
dob	1,00	−0,56	−0,33
minute_dnevno	−0,56	1,00	0,18
povjerenje	−0,33	0,18	1,00

Matrica, kodovi i rangovi

Matrica olakšava pregled, ali i traženje slučajno velikih koeficijenata. Tri varijable daju tri para, a deset ih daje četrdeset pet. Hipoteza nastala traženjem najvećeg broja ima istraživački status, različit od hipoteze postavljene prije pregleda rezultata. Posljedice za testiranje razmotrit ćemo u poglavlju o krizi i obnovi. Svaki par koji namjeravamo tumačiti najprije prikazujemo raspršenim dijagramom.

U usporedbu možemo uključiti i tekst pretvoren u brojčanu varijablu kodiranjem, uz pravila mjerenja koja smo već upoznali. **Kodirana kategorija teksta** ovisi o pravilu koje smo odabrali. Za šest namjerno izabranih naslova iz poglavlja o vizualizaciji dodjeljujemo jedan ako naslov sadržava `graphs`, `graphical` ili `graphics`, a inače nulu. Pet naslova dobiva jedan, a jedan nulu. Korelacija s godinom iznosi −0,05 (Anscombe 1973.; Cleveland i McGill 1984.; Tufte 2001.; Wilkinson 2005.; Wickham 2016.; Matejka i Fitzmaurice 2017.). Taj rezultat opisuje samo izabrane naslove i ne pokazuje opću promjenu znanstvenih naslova kroz vrijeme.

Ako pravilu dodamo oblik `visual`, svih šest naslova dobiva jedan i korelacija se više ne može izračunati jer nema razlika u kodu (Anscombe 1973.; Cleveland i McGill 1984.; Tufte 2001.; Wilkinson

2005.; Wickham 2016.; Matejka i Fitzmaurice 2017.). Tekstovi su isti, a mjera se promijenila. Autor poglavlja sastavio je pravilo za potrebe ovog primjera. Zato uz kod navodimo pravilo, osobu koja ga je odredila i tekstove na koje je primijenjen. Kasnije ćemo takve kategorije proučavati u tablicama i razmotriti što se mijenja kada ih dodjeljuje algoritam.

Izbor brojčanog prikaza važan je i kada izvorno imamo brojčana mjerenja. Pearsonova korelacija koristi njihove razmake i sažima linearnu vezu. No dob i minute u našoj anketi mogu se dosljedno kretati u suprotnim smjerovima, a da ne prate pravac. Ako želimo opisati tu dosljednost poretka, trebamo mjeru koja neće tražiti jednake promjene kroz cijeli raspon.

Vrijednosti zamjenjujemo rangovima. Najmanja dobiva prvi rang, sljedeća drugi, a jednake dijele prosječni rang. Zatim na rangovima računamo Pearsonovu korelaciju. **Spearmanov koeficijent ranga**, označen s r_s, tako mjeri dosljednost kretanja u jednom smjeru, odnosno monotonu vezu. Naziva se i Spearmanovom korelacijom. Rangovi čuvaju poredak, ali ne i izvorne razmake.

Pravilo za vezane rangove možemo vidjeti na namjerno zadanom nizu 10, 20, 20, 50. Rangovi su 1, 2,5, 2,5 i 4. Dvije dvadesetice dijele prosjek drugoga i trećega mjesta. U anketi tako zasebno rangiramo svih 300 dobi i svih 300 vremena, pa povezujemo rangove koji pripadaju istoj osobi.

Na grafu dobi i minuta vidjeli smo strmiji pad među mlađima i izravnavanje poslije, bez odvojene podskupine koja sama stvara odnos. Uz taj graf uspoređujemo Pearsonovu korelaciju −0,56 i Spearmanovu −0,68. Razlika odgovara vidljivoj zakrivljenosti, ali sama je ne bi dokazala. Pearsonova korelacija dobi i logaritma minuta iznosi −0,68 i bliža je Spearmanovoj. To je dodatna provjera ovog simuliranog odnosa, bez općeg pravila za izbor mjere.

Utjecajno opažanje je opažanje čije uključivanje ili uklanjanje znatno mijenja koeficijent. Ne mora svaka krajnja vrijednost biti utjecajna.

Sličnost Pearsonova i Spearmanova koeficijenta može odgovarati približno pravocrtnoj monotonoj vezi, ali ne jamči takav oblik ni otpornost nalaza. Oba koeficijenta može pomaknuti isti neobičan raspored. Njihova razlika također nema jedno objašnjenje. Može nastati zbog zakrivljenosti, utjecajnih točaka, vezanih rangova, podskupina ili suženog raspona. Zato najprije pregledavamo graf i način odabira podataka. Sama razlika nije dovoljan razlog da odaberemo Spearmanovu mjeru.

Ako veza nije monotona, oba koeficijenta mogu biti blizu nule. Zadovoljstvo, primjerice, može biti nisko pri vrlo malom i vrlo velikom opterećenju. Uzlazni i silazni dio tada se u koeficijentu poništavaju, iako graf pokazuje odnos.

Ograničenje raspona i odabir

Uz izbor koeficijenta važno je razmotriti i izbor opažanja jer oba utječu na opis povezanosti. U približno linearnoj vezi sa sličnom raspršenošću ishoda, odabir prema užem rasponu jedne varijable često smanjuje apsolutnu korelaciju. Raspon koji nosi signal postaje manji, dok preostalo rasipanje može ostati slično.

> **Ograničenje raspona** (*range restriction*) nastaje kada promatrani uzorak pokriva samo dio mogućih vrijednosti jedne ili obiju varijabli. Ono može oslabiti izmjerenu povezanost, ali smjer promjene ovisi o obliku odnosa i pravilu odabira.

Za čist prikaz slabljenja konstruirajmo 27 parova. Svaku vrijednost x od jedan do devet ponovimo triput, a y dobijemo dodavanjem −2, 0 i 2. Unutar svake vrijednosti x rasipanje je zato isto. Potom zadržimo samo x od četiri do šest. Tako sužavamo raspon uz očuvano pravilo nastanka y.

Tablica 2: Isti linearni mehanizam prije i nakon ograničenja raspona. Konstruirani nastavni primjer, izrada autora.

Uzorak	n	Najmanji x	Najveći x	r
Cijeli raspon	27	1	9	0,845
Od 4 do 6	9	4	6	0,447

Manji raspon x uz isto rasipanje oko pravca daje slabiju korelaciju. Sužavanje raspona ne mora uvijek smanjiti koeficijent. U zakrivljenoj vezi možemo izdvojiti dio koji bolje ili slabije prati pravac. Ako odabir ovisi o objema varijablama ili povezanoj trećoj, može se promijeniti i predznak. Zato uspoređujemo grafove prije i nakon odabira te provjeravamo pravilo uključivanja.

U cijeloj simuliranoj anketi korelacija dobi i minuta je −0,56. Najmlađa skupina ima 90 osoba sa sedam različitih dobi, od 18 do 24 godine. Njezina korelacija iznosi 0,18 i ima drugi predznak. To se ne može tumačiti samo kao slabljenje iste linearne veze. Generator razlikuje dobne skupine, ali unutar najmlađe minute ne ovise o točnoj dobi. Populacijska korelacija unutar nje zato je nula, a opaženih 0,18 proizlazi iz promjenjivosti uzorka od 90 osoba.

Nalaz za najmlađu skupinu i veza u cijelom skupu odgovaraju na pitanja različitog obuhvata. Slično, istraživanje jedne studentske generacije opisuje uzak raspon dobi, dok zaključak o cijeloj odrasloj populaciji traži širi obuhvat. Za primljene kandidate, zaposlene radnike ili preživjela poduzeća trebamo provjeriti kako je odabir promijenio raspon, oblik i sastav opažanja.

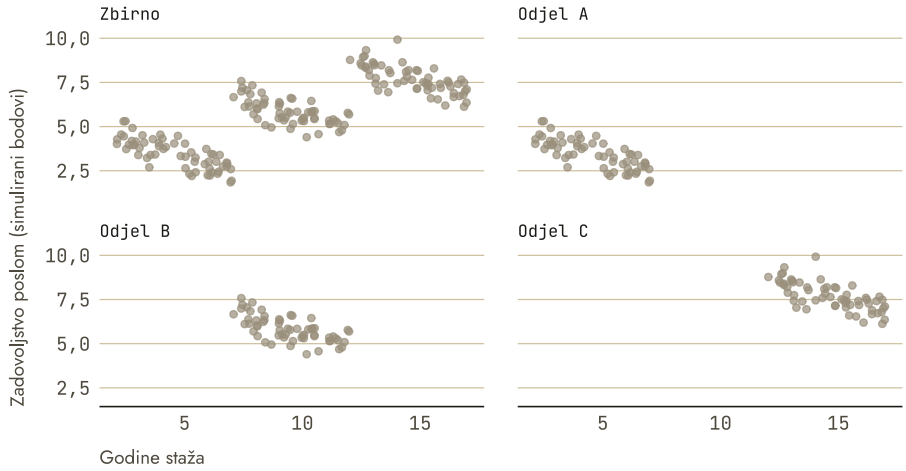
U primjeru dobivamo koeficijent različit od nule iako je populacijska korelacija unutar skupine nula. U usporedivim uvjetima takva su odstupanja češća u manjim uzorcima. Na koeficijent može snažno utjecati i jedna udaljena točka jer mijenja sredine, standardne devijacije i zajednička odstupanja. Zato uz broj opažanja pregledavamo graf, podskupine i način odabira.

Podskupine i obrat predznaka

Odabirom jedne dobne skupine promijenili smo odnos koji opisujemo. Sada ispitujemo obrnut potez, spajanje različitih skupina u jedan prikaz. Kada podskupine imaju različite razine obiju varijabli, zbirna korelacija može imati suprotan predznak od korelacija unutar podskupina.

To ćemo prikazati konstruiranim podacima o trima odjelima, bez tvrdnje o stvarnoj organizaciji. Unutar svakog odjela zadovoljstvo blago opada sa stažem. Odjeli s iskusnijim zaposlenicima ipak imaju više prosječno zadovoljstvo.

Za svaki odjel generirali smo 60 osoba. Staž je jednoliko raspoređen od 2 do 7 godina u A, od 7 do 12 u B i od 12 do 17 u C. U sredini tih raspona zadovoljstvu zadajemo 3,6, 5,7 i 7,7. Za svaku godinu iznad sredine odjela oduzimamo 0,42 i dodajemo neovisno normalno odstupanje standardne devijacije 0,5. Tako odjel istodobno određuje polazišta staža i zadovoljstva. Zadovoljstvo je ovdje konstruirana kontinuirana vrijednost bez unaprijed ograničenog raspona, a ne odgovor na stvarnu anketnu ljestvicu. Četiri polja sljedećeg grafa pregrupiraju istih 180 generiranih zapisa.



Slika 2: Konstruirani podaci u kojima zbirni oblak raste, a oblak unutar svakoga od triju odjela pada. To je konstruirani prikaz Simpsonova paradoksa. Četiri polja prikazuju iste točke bez regresijskih pravaca. Izrada autora. Podaci i postupak opisani su u tekstu.

Zbirna korelacija staža i zadovoljstva je pozitivna, 0,76. Unutar odjela iznosi −0,77, −0,66 i −0,71, pa su sve tri negativne. Svaki je račun točan. Zbirni rezultat uključuje i razlike među odjelima, koje su ovdje dovoljno velike da odrede pozitivan predznak.

To je pojava koju je Simpson opisao na frekvencijskim tablicama (Simpson 1951.). Već smo je susreli u Berkeleyjskim stopama prijma, gdje se zbirna razlika promijenila nakon razdvajanja odjela (Bickel i sur. 1975.). Grafovima po skupinama možemo odvojiti razlike među skupinama od odnosa unutar njih.

Poseban problem nastaje kada iz korelacije među skupinama zaključujemo o pojedincima. Zemlje, županije ili škole često su dostupne samo kao agregati. Njihovom korelacijom ne dobivamo odnos među ljudima unutar njih. Agregiranje uklanja dio unutarnjih razlika pa koeficijent može biti drukčiji, jači ili drugog smjera. Zaključivanje o pojedincima iz takvih odnosa nazivamo **ekološkom pogreškom** (*ecological fallacy*). Za individualni zaključak trebamo podatke i dizajn na individualnoj razini. Odnos među zemljama ostaje nalaz o zemljama.

Država kao jedinica

Na objavljenim pokazateljima možemo provjeriti kako promjena jedinice ograničava tumačenje povezanosti. S pojedinačnih osoba sada prelazimo na države. Eurostat, statistički ured Europske unije, objavljuje pokazatelje kojima uspoređujemo obrazovanje, zaposlenost, siromaštvo, uporabu interneta i dobnu strukturu stanovništva (Eurostat 2026.b, 2026.e, 2026.d, 2026.a, 2026.c, 2026.f). Analiza se ovdje oslanja na objavljene vrijednosti. U izvornoj tablici jedan redak pripada jednom pokazatelju za jednu državu i godinu. Kada pokazatelje spojimo za graf, jedna točka predstavlja državu.

Za Hrvatsku i Francusku u nastavku vidimo po dva izvorna retka. Obrazovanje se odnosi na osobe od 25 do 34 godine, a uporaba interneta na osobe od 16 do 74 godine (Eurostat 2026.d, 2026.c). Povezivanjem redaka iste države dobivamo par vrijednosti za jednu točku na grafu.

Tablica 3: Dva pokazatelja za dvije države u Eurostatovu izvatku za 2025. Izrada autora prema Eurostat (2026.d) i Eurostat (2026.c).

Država	Pokazatelj	Udio (%)
Francuska	Tercijarno obrazovanje, 25–34	55,8
Francuska	Uporaba interneta, 16–74	95,0
Hrvatska	Tercijarno obrazovanje, 25–34	42,0
Hrvatska	Uporaba interneta, 16–74	85,9

U tablici su udjeli za čitanje zaokruženi na jednu decimalu. Za graf i izračun povezanosti zadržavamo izvorne brojčane vrijednosti.

Eurostatov izvadak sadržava šest pokazatelja za 27 članica EU-a u 2025., odnosno 162 kombinacije `država + godina + pokazatelj` i 161 brojčanu vrijednost (Eurostat 2026.b, 2026.e, 2026.d, 2026.a, 2026.c, 2026.f). Nedostaje samo rano napuštanje obrazovanja za

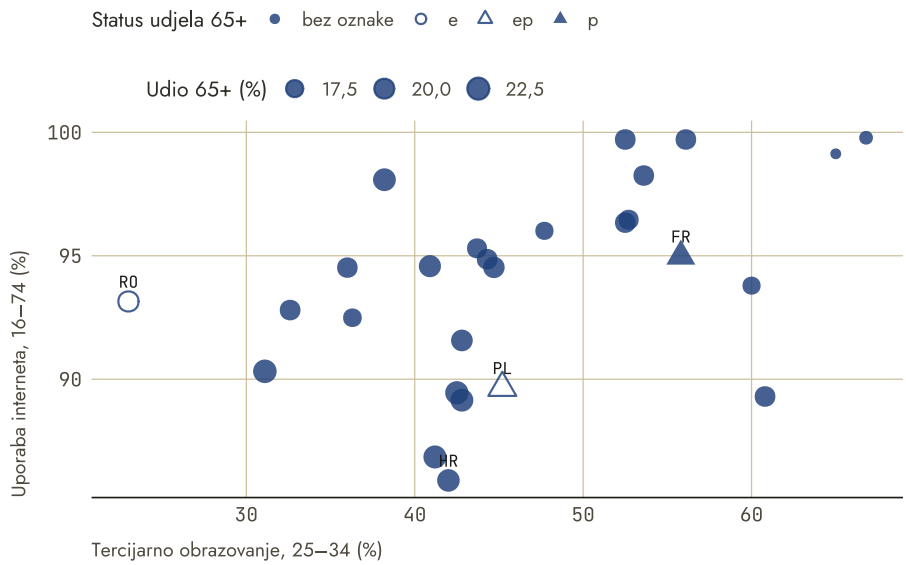
Luksemburg (Eurostat 2026.a). Zadržan je zapis : s izvornom oznakom u za nisku pouzdanost, bez oznake povjerljivosti (Eurostat 2026.a). Ne zamjenjujemo ga nulom ni vrijednošću druge godine.

Oznaka kvalitete razlikuje se od nedostajuće vrijednosti. Hrvatska vrijednost ranog napuštanja obrazovanja iznosi 2,1 i također nosi u, ali je objavljena kao broj (Eurostat 2026.a). Oznaku b, koja bilježi prekid vremenske serije, također čuvamo (Eurostat 2026.e, 2026.a). Analiza ipak ostaje presjek jedne godine.

Tablica 4: Izvorne statusne oznake u cijelom izvatku od 162 retka. Izrada autora prema šest Eurostatovih skupova za 2025. (Eurostat 2026.b, 2026.e, 2026.d, 2026.a, 2026.c, 2026.f).

Oznaka	Značenje	Redaka
nema	bez objavljene oznake	150
b	prekid u vremenskoj seriji	2
d	različita definicija; vidjeti metapodatke	4
e	procijenjena vrijednost	1
ep	procijenjena i privremena vrijednost	1
p	privremena vrijednost	2
u	niska pouzdanost	2

Za ilustraciju biramo tercijarno obrazovanje i uporabu interneta jer nas zanima odnos obrazovnog i digitalnog sudjelovanja, a oba pokazatelja pokrivaju EU-27 (Eurostat 2026.d, 2026.c). To je istraživački izbor, bez unaprijed registrirane hipoteze. Prvi pokazuje udio osoba od 25 do 34 godine s tercijarnim obrazovanjem, a drugi udio osoba od 16 do 74 godine koje su koristile internet u prethodna tri mjeseca (Eurostat 2026.d, 2026.c). Svaka točka predstavlja državu. Veličina joj prikazuje udio stanovništva od 65 godina naviše, a oblik čuva izvornu oznaku uz taj pokazatelj. Tako prije koeficijenta vidimo odnos i moguću ulogu dobnog sastava.



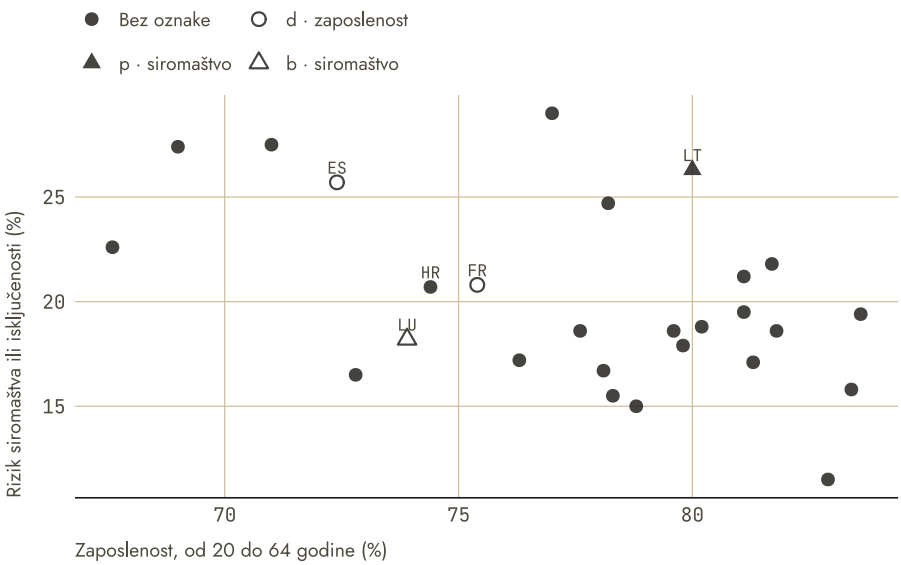
Slika 3: Tercijarno obrazovanje i uporaba interneta u 27 država članica EU-a 2025. Veličina točke pokazuje udio stanovništva od 65 godina naviše, a oblik prenosi njegovu statusnu oznaku. Oznaka e znači procjenu, p privremenu vrijednost, a ep procijenjenu i privremenu vrijednost. Izrada autora prema Eurostatu (Eurostat 2026.d, 2026.c, 2026.f).

Za 27 država Pearsonova korelacija iznosi 0,45, a Spearmanova 0,51 (Eurostat 2026.d, 2026.c). Bliske su i odgovaraju uglavnom uzlaznom oblaku, ali ne dokazuju linearnost ili otpornost nalaza. U EU-27 tijekom 2025. viši državni udjeli tercijarno obrazovanih osoba od 25 do 34 godine povezani su s višim udjelima uporabe interneta među osobama od 16 do 74 godine (Eurostat 2026.d, 2026.c). To ne pokazuje odnos obrazovanja i uporabe interneta kod pojedine osobe.

Veličina točaka potiče provjeru dobne strukture kao mogućeg trećeg čimbenika (Eurostat 2026.f, 2026.d, 2026.c). Za tu bismo provjeru trebali ispitati odnos dobne strukture s objema glavnim varijablama i razmotriti dizajn koji odvaja moguća objašnjenja. Ovaj presjek ne pokazuje trend, uzrok, odnos među pojedincima ili nalaz izvan EU-27 (Eurostat 2026.f, 2026.d, 2026.c).

Iste države možemo usporediti prema zaposlenosti te riziku siromaštva ili socijalne isključenosti. Zaposlenost obuhvaća osobe od 20 do 64 godine, a rizik stanovništvo svih dobi (Eurostat

2026.b, 2026.e). Sljedeći graf prikazuje 27 potpunih parova, po jednu točku za državu, uz izvorne oznake obaju pokazatelja (Eurostat 2026.b, 2026.e). Oblaci državnih udjela opisuju razlike među zemljama. Za vezu zaposlenosti i siromaštva pojedine osobe trebali bismo individualne podatke, a za uzročni zaključak odgovarajući dizajn.



Slika 4: Zaposlenost i rizik siromaštva ili socijalne isključenosti. Oznake d, p i b označavaju različitu definiciju, privremenu vrijednost i prekid serije (Eurostat 2026.b, 2026.e). Izrada autora.

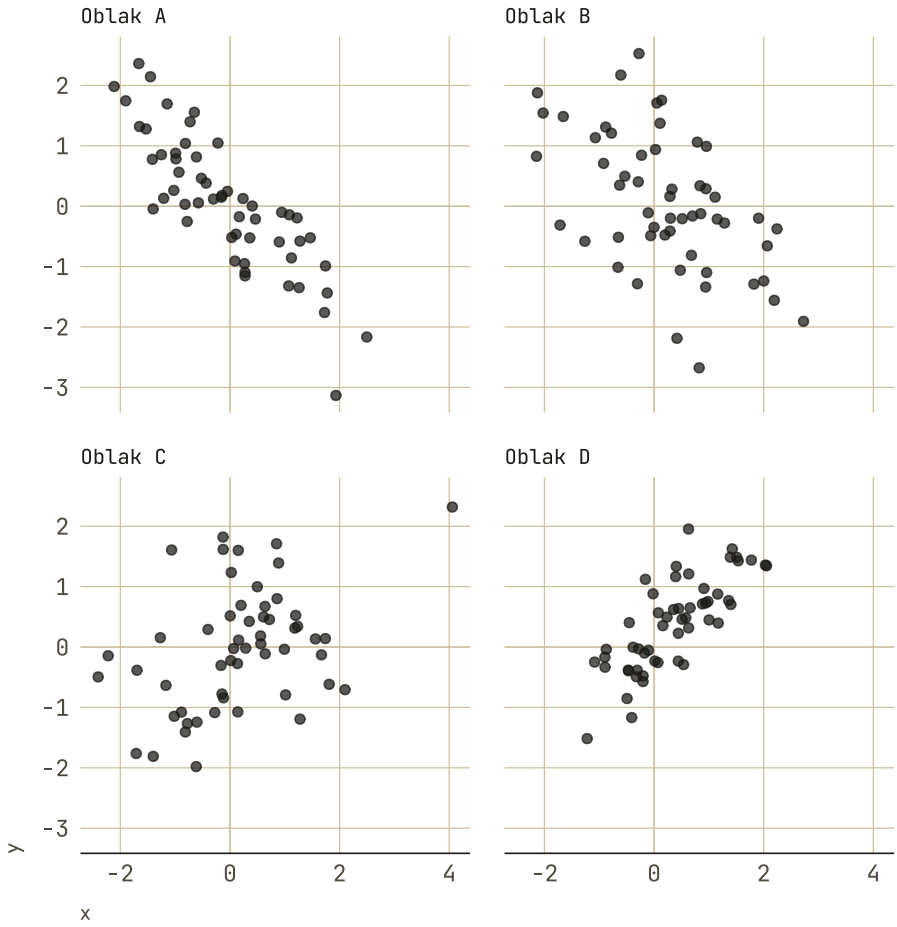
Moguća konfundirajuća varijabla mora prethoditi objema pojavama i utjecati na njih. To prosuđujemo prema sadržaju i dizajnu. U konstruiranom primjeru odjel tako oblikuje staž i zadovoljstvo. Razdvajanjem odjela vidimo promjenu povezanosti, ali sama ta promjena ne dokazuje uzročnost. O neizmjerenom čimbeniku koeficijent ne daje izravnu obavijest.

Pogodi korelaciju

Svaki oblak ovdje je novi nastavni skup s 54 para bez mjerne jedinice. Za jednu točku neovisno izvlačimo dvije normalne vrijednosti sa sredinom nula i standardnom devijacijom jedan. Prva postaje x, a y dobivamo miješanjem dijela prvoga izvlačenja i dijela drugoga. Više zajedničkoga dijela znači jaču vezu, a negativan predznak zajedničkoga dijela silaznu vezu. Generator zadaje četiri različite populacijske korelacije, dok koeficijent svakog konačnog oblaka računamo iz njegovih 54 para. Uzorački koeficijenti pritom mogu odstupati od zadanih populacijskih vrijednosti. Točke ostaju iste dok čitatelj iskušava vlastite procjene.

Četiri raspršena oblaka omogućuju usporedbu smjera i jačine povezanosti. Najprije promatramo oblik svakog oblaka. Tablica koja slijedi donosi Pearsonove korelacije i zadane procjene, pa se može provjeriti koliko su procjene blizu izračunanim vrijednostima.

Prije čitanja tablice zapišite vlastitu procjenu korelacije za svaki oblak. Tablica zatim pokazuje izračunate koeficijente i ogledne procjene zamišljenog čitatelja. Ogledne procjene nisu opažanja iz istraživanja.



Slika 5: Četiri simulirana oblaka prije otkrivanja Pearsonovih korelacija u tablici koja slijedi. Izrada autora. Podaci i postupak opisani su u tekstu.

Prije čitanja tablice zapišite vlastitu procjenu korelacije za svaki oblak. Stupac „Zadana procjena" prikazuje unaprijed odabrane odgovore zamišljenog čitatelja, kako bismo svi mogli provjeriti isti račun odstupanja.

Tablica 5: Pearsonovi koeficijenti i zadane procjene za tiskanu provjeru. Izrada autora.

Oblak	Izračunati r	Zadana procjena
A	−0,85	−0,70
B	−0,53	−0,20
C	0,40	0,20
D	0,77	0,50

Usporedite oblake A i D prema apsolutnoj vrijednosti korelacije. Za svaki oblak oduzmite zadanu procjenu od koeficijenta u tablici i uzmite apsolutnu vrijednost. Usporedite četiri odstupanja i obrazložite koji je oblak najteže procijeniti.

Predznak je lakše uočiti nego točnu jačinu veze. Pri usporedbi oblaka A i D trebamo i jednake ljestvice. Uz dobro procijenjen koeficijent potreban je i pregled oblika oblaka.

● STATISTIKA U DIVLJINI

Vrijeme uz ekran i dobrobit. U istraživanju s početka poglavlja povezanost uporabe digitalne tehnologije i dobrobiti adolescenata bila je u cjelini negativna i mala, a usporedba mnogih opravdanih analiza pokazala je zašto zaključak treba čitati uz odabrane mjere i postupke (Orben i Przybylski 2019.). To nam daje određeniji odgovor od tvrdnje da su „ekrani štetni". Govorimo o odnosu izmjerenih obilježja u obuhvaćenim skupovima podataka. Za razumijevanje nečije večeri trebali bismo znati i što je osoba radila, s kim je bila u kontaktu te kako se osjećala prije toga.

Takva povezanost sama ne govori koliko bi se dobrobit iste osobe promijenila kad bi smanjila uporabu. Za to trebamo usporedbu koja razdvaja promjenu navike od drugih mogućih objašnjenja. U poglavlju o dvjema grupama upoznat ćemo pokus s deaktivacijom Facebooka. On će postaviti uže pitanje o određenoj intervenciji, određenim korisnicima i određenom razdoblju. Razlika između tih pitanja pomoći će nam povezati opis navika s dokazom o njihovoj promjeni.

NAĐITE GREŠKU

U simuliranoj anketi od 300 osoba provjeravamo odnos dobi i dnevnih minuta na društvenim mrežama. Asistentu je postavljeno pitanje o cijelom uzorku. Vratio je ovaj račun.

```
anketa_mreze |>
  filter(dobna_skupina == "18 do 24") |>
  summarise(
    r = cor(dob, minute_dnevno),
    n = n()
  )
```

Za provjeru su dostupni isti sažeci koje smo usporedili uz raspršeni dijagram. Pearsonov koeficijent u tablici opisuje linearnu povezanost.

Tablica 6: Dob i dnevno vrijeme u dvama obuhvatima. Izrada autora.

Uključene osobe	Broj osoba	Pearsonov koeficijent
Cijeli simulirani uzorak	300	−0,56
Osobe od 18 do 24 godine	90	0,18

Račun daje korelaciju 0,18 uz 90 ispitanika. Prema tome, dob i vrijeme korištenja nisu negativno povezani u cijelom simuliranom uzorku.

Pridružite ispis jednom retku tablice. Označite mjesto na kojem se odgovor udaljio od postavljenog pitanja i napišite zamjenski zaključak koji opisuje oba obuhvata. Sačuvajte točan rezultat asistentova računa.

Razrađeni primjer

U odnosu dobi i dnevnog vremena korištenja već smo vidjeli zašto koeficijent trebamo čitati uz graf. Sada te provjere povezujemo u jedno izvješće o simuliranoj anketi. Graf pokazuje oblik odnosa. Pearsonov koeficijent sažima koliko odnos prati pravac, a Spearmanov koliko dosljedno slijedi silazni poredak. Zaključak se odnosi na te simulirane podatke.

```
sazetak <- anketa_mreze |>
  summarise(
    n = n(),
    pearson = cor(dob, minute_dnevno),
    spearman = cor(dob, minute_dnevno, method = "spearman")
  )

sazetak
```

```
# A tibble: 1 x 3
      n pearson spearman
  <int>   <dbl>    <dbl>
1   300  -0.559   -0.680
```

Za oba opisa koristimo iste parove dobi i minuta. Poziv `summarise`, koji već poznajemo, okuplja sažetke, a funkcija `cor` računa korelaciju dvaju stupaca. Zadana je Pearsonova korelacija. Odabirom Spearmanove u argumentu `method` mijenjamo način sažimanja odnosa, dok osobe i njihove vrijednosti ostaju iste.

Graf pokazuje da se pad vremena s dobi poslije izravnava. Pearsonova vrijednost −0,56 opisuje linearnu vezu, a Spearmanova −0,68 dosljednost silaznog poretka. Razlika je spojiva sa zakrivljenošću koju smo vidjeli, ali bez grafa ne bi bila dovoljna za taj zaključak.

Rezultat možemo navesti ovako. U simuliranom uzorku od 300 ispitanika dob i dnevno vrijeme korištenja negativno su

povezani. Graf pokazuje zakrivljen silazni odnos, uz Spearmanovu korelaciju −0,68 i Pearsonovu −0,56, manjeg apsolutnog iznosa. Ne zaključujemo o uzroku jer je skup simuliran i presječan, bez istraživačke intervencije ili zabilježenog vremenskog redoslijeda.

Od odnosa do tvrdnje

Na Eurostatovu presjeku pratimo kako od izvornog retka dolazimo do rečenice o povezanosti i što usput treba provjeriti.

Tablica 7: Šest revizijskih pitanja primijenjenih na povezanost službenih agregata. Izrada autora prema Eurostatu (Eurostat 2026.d, 2026.c, 2026.f).

Revizijsko pitanje	Primjena na Eurostatov odnos
Koja je jedinica?	nakon spajanja pokazatelja, jedna država članica EU-a u 2025.
Tko ili što nedostaje?	pojedinci unutar država, zemlje izvan EU-27 i druga razdoblja nisu jedinice ovoga presjeka
Koji je cilj tvrdnje?	Pearsonov i Spearmanov koeficijent opisuju povezanost dvaju državnih udjela
Gdje je neizvjesnost?	statusne oznake prenose upozorenja izvora. Koeficijenti ne mjere uzoračku ni mjernu nesigurnost, a presjek ne pokazuje promjenu kroz vrijeme
Koje drugo objašnjenje ostaje?	druga dobna skupina ili pokazatelj, rangovi umjesto vrijednosti ili drukčije opravdano postupanje sa statusnim oznakama
Tko snosi posljedice?	pogrešan opis stanovnika ili država može dovesti do neutemeljene uredničke ili javne odluke

Promjena jedinice mijenja i ono što procjenjujemo, moguća objašnjenja te posljedice zaključka. Statusna oznaka također traži prosudbu. O uključivanju retka odlučujemo prema značenju oznake. Pravilo unaprijed obrazlažemo i navodimo u izvještaju.

Granica tvrdnje o povezanosti ovisi o jedinici, obuhvatu i načinu mjerenja. Sljedeća tablica pokazuje koje vrste tvrdnji ti podaci podupiru.

Tablica 8: Šest dimenzija tvrdnje na granici Dijela II. Izrada autora prema Eurostatu (Eurostat 2026.b, 2026.e, 2026.d, 2026.a, 2026.c, 2026.f).

Dimenzija tvrdnje	Što ovaj dokaz dopušta
opis	opis vrijednosti i izvornih statusnih oznaka za šest pokazatelja u EU-27 2025.
povezanost	odnos državnih udjela tercijarnoga obrazovanja i uporabe interneta u istom presjeku
generalizacija	nije poduprta izvan država, godine i dobnih obuhvata koje izvori navode
predviđanje	nije poduprto jer nijedno prediktivno pravilo nije izgrađeno ni provjereno na doista novim opažanjima
uzročnost	nije poduprta jer vremenski redoslijed i konkurentska objašnjenja nisu razdvojeni
odluka	poduprta je omeđena urednička odluka o poštenoj formulaciji nalaza, ali ne obrazovna ili digitalna politika

Za ovaj presjek možemo reći da su dva državna udjela u EU-27 2025. pozitivno povezana (Eurostat 2026.d, 2026.c). Iz toga ne izvodimo zaključak o pojedincima, uzrocima, budućnosti ili drugim zemljama.

Za samoprovjeru možemo odgovoriti na četiri pitanja. Što predstavlja točka Eurostatova grafa? Zašto je Luksemburg uključen u glavni odnos, ali ne i u odnos s ranim napuštanjem obrazovanja? Zašto prije usporedbe koeficijenata pregledavamo graf? Koje tri vrste zaključka ovaj presjek ne podupire i što bi nam za njih trebalo?

Tri pogrešna tumačenja iz ovoga dijela knjige možemo provjeriti istim redom. Najprije ustanovimo što je doista u podacima, zatim usporedimo tvrdnju s odgovarajućim prikazom i na kraju odredimo granicu zaključka. Tako se provjere tablice, grafa i koeficijenta nadovezuju jedna na drugu.

Tablica 9: Tri provjere pogrešnih analiza u Dijelu II. Izrada autora.

Provjera	Sažetak u poglavlju o sažimanju	Graf u poglavlju o vizualizaciji	Povezanost u ovom poglavlju
Pogrešan zaključak	spoj po godini valjan je jer se zbroj prikazanih redaka može izračunati	treći stupac može biti širi radi optičke ravnoteže	povezanost u jednoj dobnoj podskupini određuje smjer veze u cijelom uzorku
Što gledamo	jedinicu retka, puni ključ, broj redaka i zbroj prije i nakon spajanja	jednake širine stupaca, os od nule i čitljivost bez boje	filtrirani uzorak, cijeli oblak točaka i oblik veze
Gdje je provjera	tablica kontrole spajanja i oznaka dostupnosti metrike	usporedba dviju osi i tablica izbora prikaza	raspršeni dijagram dobi i minuta te koeficijenti za cijeli uzorak
Granica	opis vrijedi za uključeni DigiKatov korpus	graf ne nadomješta provjeru podataka i uredničkoga pitanja	povezanost ne pokazuje uzrok ni populacijski doseg

Sažetak

Od zajedničkih odstupanja dviju varijabli izgradili smo kovarijancu, a standardizacijom Pearsonovu korelaciju, mjeru linearne veze između minus jedan i plus jedan koja ne ovisi o mjernim jedinicama. Kada vrijednosti zamijenimo rangovima, Spearmanovim koeficijentom opisujemo monotonost, pri čemu gubimo izvorne razmake. Graf nam pomaže protumačiti obje mjere jer pokazuje podskupine, utjecajna opažanja i oblik koji se može promijeniti sužavanjem raspona. Na rezultat utječu i pravilo kojim tekst pretvaramo u brojčane oznake te izbor jedinice, pa povezanost među državama ne prenosimo na pojedince. Za uzročni zaključak i dalje trebamo odgovarajući dizajn. U sljedećem dijelu vjerojatnost će nam pomoći razumjeti kakve obrasce može proizvesti sam slučajni postupak.

Pojmovi

kovarijanca (*covariance*), korelacija (*correlation*), Pearsonova korelacija (*Pearson correlation*), linearnost (*linearity*), oblik odnosa (*shape of a relationship*), matrica korelacija (*correlation matrix*), kodirana kategorija teksta (*coded text category*), Spearmanov koeficijent ranga (*Spearman rank correlation*), monotona veza (*monotonic relationship*), utjecajno opažanje (*influential observation*), ograničenje raspona (*range restriction*), ekološka pogreška (*ecological fallacy*), konfundirajuća varijabla (*confounder*), granica tvrdnje o povezanosti (*boundary of an association claim*)

Zadaci

Konceptualni

Vratite se u poglavlje o sažimanju podataka i primijenite ondje uvedena pravila o jedinici retka, ključu i odsutnoj vrijednosti na Eurostatov izvadak iz ovoga poglavlja. Objasnite zašto glavni odnos tercijarnoga obrazovanja i uporabe interneta zadržava Luksemburg i 27 država, zašto bi odnos koji uključuje rano napuštanje obrazovanja imao najviše 26 potpunih parova te zašto hrvatska vrijednost 2,1 sa statusom u ostaje broj, dok luksemburški zapis : s istim statusom ne postaje nula (Eurostat 2026.d, 2026.c, 2026.a). Predajte tablicu s tri retka, odlukom o uključivanju i razlogom, zatim jednu poštenu rečenicu o glavnom odnosu na razini država.

Računski

Upotrijebite tablicu korelacija triju varijabli simulirane ankete. Za svaki od triju parova zapišite smjer Pearsonove veze. Za par dobi i minuta zatim prema raspršenom dijagramu procijenite bi li se Spearmanova vrijednost razlikovala i u kojem smjeru. Za preostala dva para zapišite zašto se to ne može prosuditi bez njihovih dijagrama i što bi na njima trebalo pregledati. U tisku uzmite četiri zadane procjene i koeficijente iz tablice. Za svaki oblak izračunajte apsolutno odstupanje procjene. Predajte tablicu sa sedam redaka i jednom rečenicom obrazloženja u svakom. Postupak za ponavljanje izračuna nad cijelim skupom nalazi se u praktikumu.

Kritički

Usporedite objavljene Eurostatove tablice o tercijarnome obrazovanju i uporabi interneta s prikazom u ovom poglavlju (Eurostat 2026.d, 2026.c). Imenujte jedinicu analize i napišite dopuštenu

rečenicu na razini država. Zatim objasnite zašto iste tablice ne podupiru tvrdnju „Tercijarno obrazovanje povećava individualnu uporabu interneta". Pronađite ekološki i uzročni skok te postavite pitanje o dobnoj strukturi kao mogućem trećem čimbeniku. Navedite kakvi bi podaci i dizajn bili potrebni za tvrdnju o pojedincu i uzroku. Predajte odlomak s presudom i jednom poštenom zamjenom sporne rečenice.

Revizija modela

Ocijenite analizu iz okvira o pogrešci. Imenujte što je u pozivu ispravno izvedeno, redak koda koji mijenja ciljnu skupinu, razlog zbog kojeg dobiveni koeficijent ne odgovara na pitanje o cijelom rasponu dobi i napišite ispravljenu rečenicu izvještaja.



DIO III

od uzorka do populacije

Opis uzorka nije još tvrdnja o populaciji. Kroz slučajnost i ponovljeno uzorkovanje pratimo kako se procjene mijenjaju. Iz te promjenjivosti gradimo intervale, uz granice koje određuju odabir i mjerenje.

07

DIO III · OD UZORKA DO POPULACIJE

vjerojatnost koliko treba

Vjerojatnost povezuje model, ponavljanje i neizvjesnost

Košarkaši i navijači često govore o vrućoj ruci, uvjerenju da nakon nekoliko pogodaka raste vjerojatnost sljedećeg. Jedno je istraživanje provjerilo to vjerovanje kroz zapise utakmica i kontrolirani pokus. Nalazi su protumačeni kao znak da ljudi u slučajnim nizovima vide obrazac koji ne postoji (Gilovich i sur. 1985.).

Prema tom se tumačenju prividni niz uspjeha mogao objasniti slučajnošću (Gilovich i sur. 1985.). Kasnija ponovna analiza, međutim, otkrila je pristranost u samom načinu mjerenja tog obrasca (Miller i Sanjurjo 2018.).

Kako procijeniti je li niz posljedica promjene u procesu ili slučajnosti?

Niz pogodaka može izgledati uvjerljivije od raspršenih uspjeha. Da bismo procijenili taj dojam, trebamo najprije upoznati nizove

koje proizvodi proces sa stalnim izgledima za uspjeh. Nakon opisivanja raspodjela i povezanosti trebamo način da odredimo što bismo očekivali pod jasno zadanim uvjetima. Vjerojatnost će nam omogućiti da opaženi rezultat usporedimo s drugim rezultatima koje isti model može proizvesti.

Počet ćemo simuliranim bacanjima novčića. Jedno opažanje bit će ishod bacanja, a nizovi će pokazati kako se udio glava mijenja kada ponavljamo isti postupak. Budući da sami zadajemo izglede, možemo provjeriti koliko kratki niz odstupa od onoga što vrijedi za proces. Jednostavan model služi kao polazište za usporedbu. Tvrdnju o ponašanju košarkaša zatim bismo trebali provjeriti na stvarnim sportskim zapisima.

Za pitanja o preklapanju događaja vratit ćemo se simuliranoj populaciji `populacija_medija`, koristeći dob osobe i njezin glavni izvor vijesti. Tu možemo prebrojiti koliko osoba zadovoljava jedan uvjet, drugi uvjet ili oba. Tako ćemo provjeriti kada smijemo zbrajati ili množiti vjerojatnosti i zašto poznavanje jednog obilježja može promijeniti izglede za drugo. Populacija medijskih navika stvorena je za nastavu, s jednim retkom za jednu zamišljenu osobu, pa njezini obrasci ne opisuju stvarne stanovnike. U hipotetskoj kampanji e-biltena promatrat ćemo otvaranje poruke kao pokušaj s dvama mogućim ishodima.

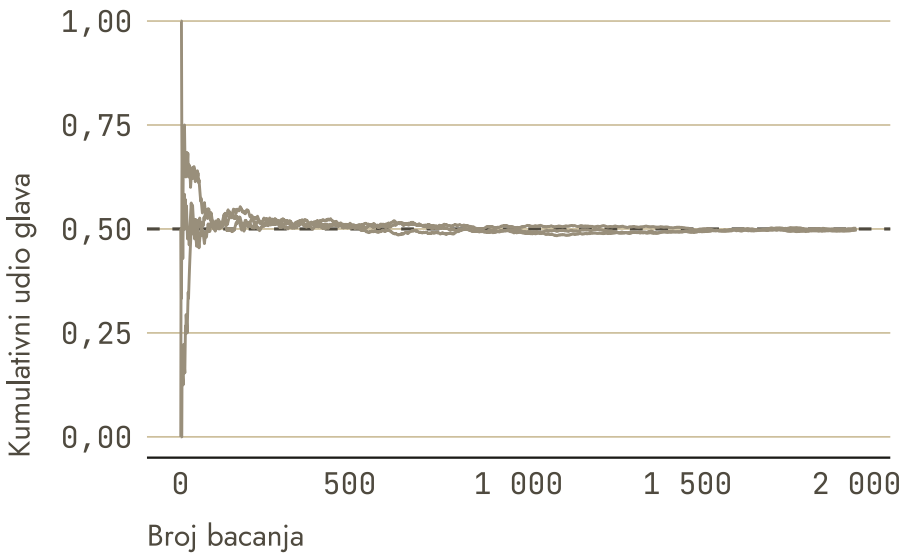
Raspodjele koje smo dosad čitali kao opise podataka sada ćemo koristiti i kao modele mogućih ishoda. Binomna raspodjela opisat će broj uspjeha pod određenim pretpostavkama, a normalna krivulja otvoriti pitanje kada je zvonasti oblik opravdan. Uz svaki ćemo račun zadržati uvjete koji ga omogućuju.

Usporedbom izračuna i ponovljenih simulacija iste kampanje utvrdit ćemo koliko je rezultat neobičan pod zadanim modelom. Taj račun treba razlikovati od provjere opisuje li model dobro stvarni svijet.

Što vjerojatnost opisuje

Računalni model svakom bacanju neovisno daje glavu s vjerojatnošću 0,5. U prvom generiranom nizu početnih pet ishoda, s jedinicom za glavu i nulom za pismo, glasi 1, 1, 1, 0, 0. Nakon petog bacanja imamo 3 glava od pet, pa je udio 0,60. Svaka sljedeća točka iste linije dodaje jedno bacanje i ponovno dijeli dotadašnji broj glava brojem svih dotadašnjih bacanja. Cijela linija ima 2 000 takvih koraka.

Simulirat ćemo bacanje poštenog novčića i nakon svakog bacanja zabilježiti udio glava. Ponavljamo postupak triput kako bismo usporedili moguće nizove. Kratki odsječci mogu izgledati neuravnoteženo i uz isti pošten novčić.



Slika 1: Kumulativni udio glava u trima neovisnim nizovima bacanja poštenog novčića. Vodoravna crta označuje vrijednost pola. Izrada autora. Podaci i postupak opisani su u tekstu.

Nakon dvadeset bacanja udjeli se kreću od 0,25 do 0,65. Nakon dvije tisuće između 0,498 i 0,502. Nizovi su različiti, ali su u du-

ljim nizovima njihovi udjeli bliži polovini. Sam kratki neuravnotežen niz zato nije dovoljan da novčić proglasimo nepoštenim.

> **Vjerojatnost** je broj od nule do jedinice koji događaju pridružuje razinu neizvjesnosti unutar jasno opisanoga modela, pri čemu veći broj označuje veće izglede za taj događaj.

Model može opisivati ponovljiv postupak ili jedinstven događaj. Za ponovljiv postupak vjerojatnost uspoređujemo s učestalošću kroz mnogo ponavljanja. Za jedinstven događaj, poput uspjeha reprezentacije na sljedećem prvenstvu, broj opisuje neizvjesnost uz dostupne informacije. U oba slučaja trebamo znati odakle procjena dolazi.

Zakon velikih brojeva opisuje približavanje relativne učestalosti vjerojatnosti kada raste broj neovisnih ponavljanja sa stabilnom vjerojatnošću.

Svako neovisno bacanje zadržava iste izglede. Približavanje udjela polovini objašnjavamo rastom nazivnika. Kako niz raste, početna neravnoteža čini sve manji dio ukupnog broja bacanja. Kasniji ishodi pritom nastaju po istom pravilu kao raniji.

Očekivanje da nakon četiri glave pošten novčić duguje pismo nazivamo kockarskom zabludom. U modelu neovisnih bacanja niz glava ne povećava ni izglede za novu glavu. Kod košarke je pitanje drukčije jer se mogu mijenjati forma, težina šuta, obrana ili odabir pokušaja. Fiksna neovisna vjerojatnost služi nam kao polazni model za usporedbu, bez pretpostavke da potpuno opisuje igru.

Razlikujemo modelnu vjerojatnost, opaženu učestalost i osobnu sigurnost. Modelna vjerojatnost proizlazi iz pretpostavki i informacija. Učestalošću kroz ponavljanja možemo provjeravati model. Za osobno uvjerenje izraženo postotkom tek treba pokazati na čemu se temelji i kako se slaže s ishodima.

Kalibrirana nesigurnost znači da se prognozirani izgledi slažu s učestalostima u mnogo usporedivih prognoza. Među događajima kojima smo pripisali sedamdeset posto izgleda, približno sedamdeset posto trebalo bi se ostvariti. Za pojedinu prognozu

navodimo njezine informacije i pretpostavke. Ocjena kalibracije stoga zahtijeva niz usporedivih prognoza i njihovih ishoda.

Tri pravila i jedna pretpostavka

Da bismo iz poznatih vjerojatnosti dobili novu, najprije razjašnjavamo koje ishode ubrajamo u događaj koji nas zanima. Vjerojatnost da se taj događaj ne dogodi dobivamo oduzimanjem njegove vjerojatnosti od jedan. To je pravilo komplementa. Kada povezujemo dva događaja riječju „ili", njihove vjerojatnosti zbrajamo, ali eventualni preklop oduzimamo kako iste ishode ne bismo prebrojili dvaput. Za zajedničko nastupanje množimo vjerojatnost prvog događaja uvjetnom vjerojatnošću drugog. Tek kada saznanje o prvome ne mijenja izglede drugoga, događaji su neovisni i možemo izravno pomnožiti njihove pojedinačne vjerojatnosti. Na izboru izvora vijesti vidjet ćemo zašto je taj uvjet važan.

Brojnosti u tablici omogućuju da nazivnike provjerimo prije računanja postotaka. Retci razlikuju dob do 29 godina od ostalih, a stupci glavni izvor vijesti u poznatoj simuliranoj populaciji.

Tablica 1: Dob i glavni izvor vijesti u simuliranoj populaciji. Izrada autora.

	Mreže	Ostalo	Portal	Ukupno
30 i više	7 710	18 767	11 163	37 640
Do 29 godina	5 668	2 754	3 938	12 360
Ukupno	13 378	21 521	15 101	50 000

U simuliranoj populaciji `populacija_medija` portal bira 30,2 % ljudi, a do dvadeset devet godina ima ih 24,7 %. Zbrajanjem dobivamo 54,9 %. No udio ljudi koji su mladi ili biraju portal jest 47,0 %. Razlika nastaje jer je 7,9 % ljudi istodobno mlado i bira portal, pa smo ih u zbroju računali dvaput.

Za pojednostavnjeno pravilo množenja trebamo opravdati neovisnost.

> **Neovisnost** dvaju događaja znači da saznanje o tome je li se jedan dogodio ne mijenja vjerojatnost drugoga.

Koristimo potpuno poznatu simuliranu populaciju od pedeset tisuća odraslih osoba. Društvene mreže kao glavni izvor vijesti bira 26,8 %, a do dvadeset devet godina ima 24,7 % ljudi. Uz neovisnost bismo množenjem dobili 6,6 % ljudi koji zadovoljavaju oba uvjeta. U generiranim podacima taj je udio 11,3 %, znatno veći.

Množenje je točno, ali pretpostavka neovisnosti ovdje ne vrijedi. Nju treba opravdati načinom nastanka podataka i dizajnom. Podaci mogu pokazati odstupanje od neovisnosti, dok odsutnost uočljive razlike sama ne potvrđuje pretpostavku. U ovom skupu trebamo razlikovati udio u cijeloj populaciji od udjela unutar određene skupine.

> **Uvjetna vjerojatnost** je vjerojatnost događaja izračunata unutar skupine u kojoj je neki drugi događaj već nastupio.

Među mlađima od trideset godina mreže bira 45,9 %, a među ostalima 20,5 %. Televiziju bira 21,7 % cijele populacije i 41,2 % onih od šezdeset godina naviše. Ukupne i uvjetne vrijednosti razlikuju se jer su dob i izbor izvora povezani.

Uvjetna vjerojatnost ovisi o smjeru pitanja. Udio upozorenja među ciljanim zapisima računa se s drugim nazivnikom od udjela ciljanih zapisa među upozorenjima. Taj smo problem susreli u poglavlju o zavaravanju brojkama. Za drugi smjer trebamo i temeljnu stopu te stopu pogrešnih upozorenja.

Udio konverzija među ljudima koji su kliknuli oglas ili udio otvaranja među poslanim porukama također su uvjetne vjerojatnosti. Vrijede za skupinu iz svojeg nazivnika. Pri usporedbi segmenata zato svaki put jasno određujemo na koga se udio odnosi.

Ponovljeni pokušaji s dva ishoda

Pravila za pojedine događaje možemo povezati kada se isti pokušaj ponavlja. Često pratimo unaprijed zadan broj pokušaja s dvama ishodima. Zanima nas, primjerice, koliko je ljudi kliknulo, odgovorilo na poziv ili otvorilo poruku. Za takve situacije možemo razmotriti binomni model.

> **Binomna raspodjela** opisuje broj uspjeha u zadanom broju pokušaja kad svaki pokušaj ima samo dva ishoda, jednaku vjerojatnost uspjeha i nikakvu vezu s ostalim pokušajima.

U hipotetskom modelu promatramo dvadeset objava. Svaka ima zadanu vjerojatnost od dva posto da u sedam dana prijeđe dogovoreni prag dijeljenja. Prag, razdoblje i broj izloženih korisnika dio su definicije viralnosti. Vjerojatnost smo zadali kao pretpostavku modela. Prema binomnom modelu, nijedna objava ne bi prešla prag u 66,8 % takvih zamišljenih mjeseci. Time opisujemo ponavljanja modela, bez jamstva za određeni mjesec.

Za pojedinu objavu izostanak viralnosti ima vjerojatnost $1 - 0{,}02 = 0{,}98$. Uz neovisnost dvadeset takvih izostanaka množimo, pa je vjerojatnost nijedne viralne objave $0{,}98^{20}$. Za barem jednu koristimo suprotni događaj i računamo $1 - 0{,}98^{20}$, odnosno 33,2 %.

Prije primjene provjeravamo uvjete. Različit doseg objava može značiti različite vjerojatnosti uspjeha. Ako jedno dijeljenje povećava vidljivost sljedeće objave, pokušaji nisu neovisni. Slično, odgovori dviju osoba iz istog kućanstva mogu biti povezani zbog zajedničkih okolnosti, premda ih zapisujemo u zasebne retke.

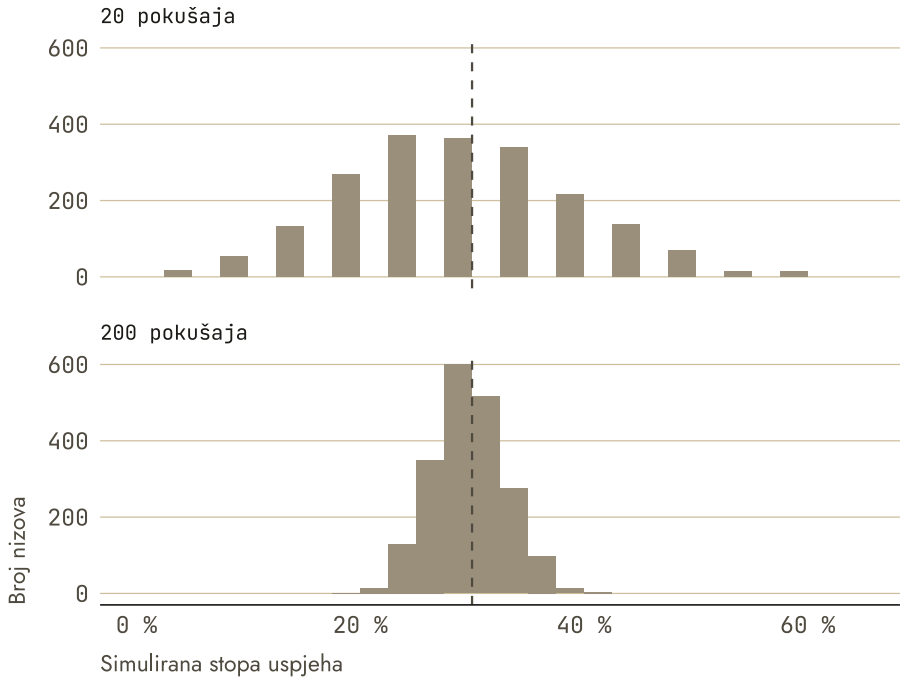
Kada uvjeti vrijede, očekivani broj uspjeha jednak je broju pokušaja pomnoženom vjerojatnošću uspjeha. S više pokušaja raste raspršenost broja uspjeha, ali sporije od samog broja pokušaja. Udio uspjeha zato postaje stabilniji. Veći raspon ukupnog broja ne znači nužno i nestabilniji udio.

Model pojednostavnjuje proces i jasno određuje što može nastati slučajno. Tako stvarni rezultat imamo s čime usporediti.

Simulator novčića i kampanje

Sada jednu vrijednost dobivamo tek nakon cijeloga niza. Ako u mogućem nizu od 20 pokušaja dobijemo šest uspjeha, u histogram unosimo jedan udio $6/20 = 0{,}30$. Taj smo niz odabrali radi ilustracije računa. Za svaki histogram ponavljamo cijele neovisne nizove, a vjerojatnost uspjeha zadržavamo. Visina stupca broji nizove čiji su udjeli u istom rasponu.

Prikaz uspoređuje ishode ponovljenih nizova uz istu zadanu vjerojatnost. Duljina niza mijenja raspon simuliranih udjela. Kratki nizovi daju raspršenije rezultate iako postupak koji ih stvara ostaje isti. U tiskanom prikazu vjerojatnost uspjeha iznosi 0,30, a za svaku duljinu ponavljamo cijeli niz 2 000 puta.



Slika 2: Simulirane stope uspjeha pri dvjema duljinama niza i istoj zadanoj vjerojatnosti modela. Dulji niz daje užu raspodjelu. Isprekidana okomita crta označuje zadanu vjerojatnost uspjeha od 0,30. Izrada autora. Podaci i postupak opisani su u tekstu.

Rezultat jedne kampanje uspoređujemo s cijelom raspodjelom koju proizvodi isti postupak. Ako je raspodjela široka, jedan ishod teže razlikuje moguća objašnjenja. Ishod daleko u repu može dovesti model u pitanje, ali ne govori sam što je uzrok odstupanja.

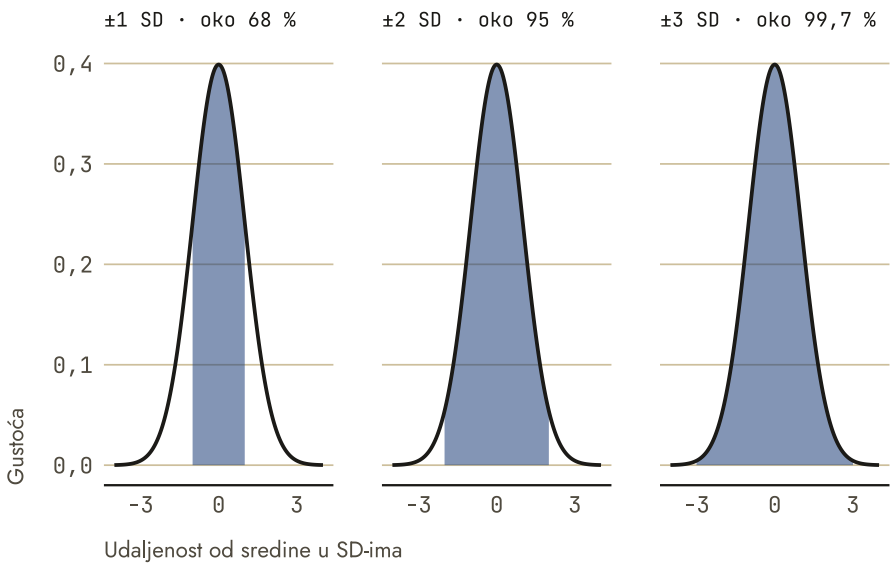
Uži histogram duljih nizova pokazuje promjenu raspršenosti. Za opis njegova oblika vratit ćemo se standardiziranoj ljestvici iz poglavlja o sažimanju podataka. Na njoj nula označuje sredinu, a jedan i minus jedan položaje po jednu standardnu devijaciju iznad i ispod nje. Taj će nam poznati način opisivanja položaja pomoći da raspodjelu usporedimo sa zvonastom krivuljom.

Zvonasta krivulja i njezino područje

Raspodjela mnogih ponovljenih udjela može poprimiti približno simetričan oblik s najviše vrijednosti oko sredine. Taj zvonasti oblik susrećemo u normalnoj raspodjeli.

> **Normalna raspodjela** je simetrična zvonasta raspodjela koju u cijelosti određuju njezino središte i njezina standardna devijacija.

Na nacrtanoj normalnoj krivulji udaljenost mjerimo standardnim devijacijama. Okomite granice omeđuju srednjih približno 68 %, 95 % i 99,7 % površine. Površina predstavlja udio mogućih vrijednosti, a visina krivulje gustoću.



Slika 3: Pravilo 68–95–99,7 za standardnu normalnu raspodjelu. Izrada autora iz matematičkog modela.

Standardizirana vrijednost određuje položaj na krivulji. Nula je sredina, pozitivan predznak označuje vrijednost iznad nje, a negativan ispod. Apsolutna vrijednost govori koliko je standardnih

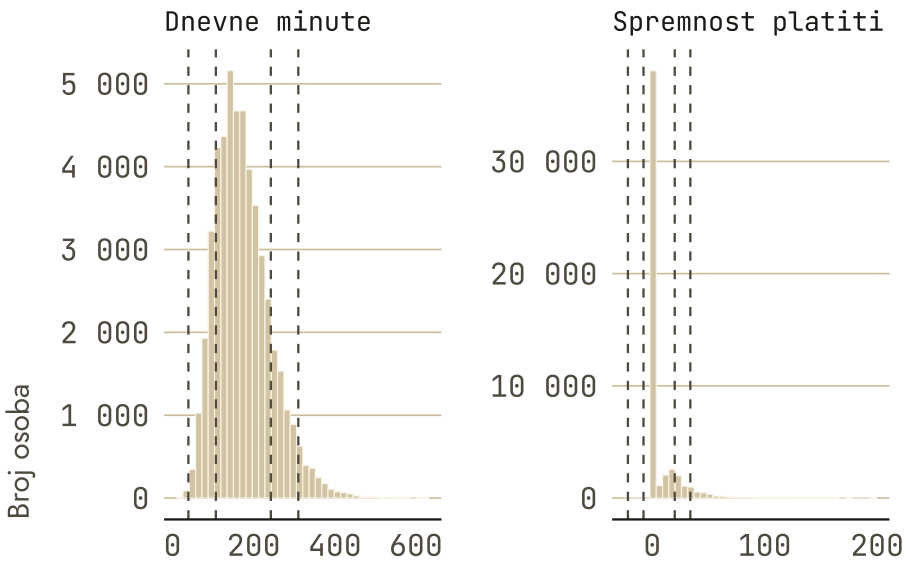
devijacija udaljena od sredine. Na normalnoj krivulji taj položaj možemo povezati s udjelom površine.

Svaki simulirani udio prosjek je nula za neuspjehe i jedinica za uspjehe u jednom nizu. Histogram stoga prikazuje raspodjelu udjela iz ponovljenih nizova. Jedna vrijednost u njemu sažima cijeli niz pojedinačnih nula i jedinica.

U našoj simulaciji pokušaji slijede isto stabilno pravilo i međusobno su neovisni. S duljim nizovima udjeli se manje rasipaju i njihov se histogram približava zvonastom obliku. Pojedinačni ishodi i dalje ostaju nule ili jedinice. Razlog tom obrascu i uvjete pod kojima vrijedi razradit ćemo u poglavlju o uzorkovanju, uz središnji granični teorem.

Sredina i standardna devijacija potpuno određuju položaj i širinu normalne raspodjele. Unutar jedne standardne devijacije od sredine leži oko 68 % vrijednosti, unutar dvije oko 95 %, a unutar tri oko 99,7 %.

S raspodjele prosjeka sada prelazimo na pojedinačne vrijednosti u istoj populaciji `populacija_medija`. Njihov oblik trebamo zasebno provjeriti. Pogledat ćemo dnevno vrijeme uz medije i iznos koji je osoba spremna platiti za sadržaj.



Slika 4: Dnevne minute uz medije i spremnost na plaćanje u simuliranoj populaciji, s isprekidanim crtama na jednoj i dvjema standardnim devijacijama od sredine. Izrada autora. Podaci i postupak opisani su u tekstu.

Za dnevne minute unutar jedne standardne devijacije nalazi se 70,1 % populacije, unutar dviju 95,9 %, a unutar triju 99,1 %. Te su vrijednosti blizu normalnom pravilu. Blaga asimetrija udesno stvara manja odstupanja.

Za spremnost na plaćanje odstupanja su mnogo veća. Unutar jedne devijacije nalazi se 87,7 %, a unutar triju 97,6 % populacije. Graf pokazuje da 76,1 % ljudi navodi nulu, uz dugi rep prema velikim iznosima. Jedna devijacija ispod sredine već daje negativan iznos, pa ispod te granice nema opažanja. Normalna raspodjela ne opisuje dobro te sirove vrijednosti.

Promjena ljestvice zahtijeva dva odvojena koraka. Iz istih 50 000 zapisa prvo izdvajamo 11945 pozitivnih iznosa, a zatim na njih primjenjujemo prirodni logaritam. Primjerice, za prvi pozitivan iznos od 55,0 eura nova vrijednost iznosi 4,007. Nule nemaju konačan logaritam i nisu dio toga izračuna. Nazivnik svakoga sljedećeg udjela zato su samo platitelji.

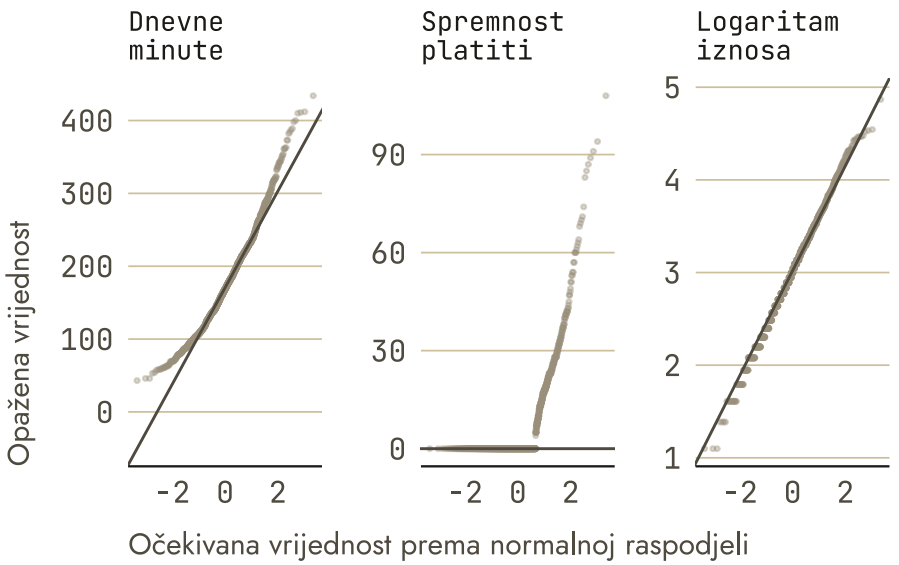
Ako izdvojimo osobe spremne nešto platiti i logaritmiramo njihove iznose, oblik postaje bliži zvonastom. Unutar odgova-

rajućih raspona sada je 68,0 %, 94,8 % i 99,7 %. Izdvajanjem smo, međutim, promijenili populaciju, a logaritmiranjem ljestvicu. Bolje slaganje s normalnom raspodjelom samo po sebi ne opravdava izostavljanje neplatitelja.

Kad podaci ne pristaju krivulji

Za sljedeći prikaz uzimamo po 1 200 vrijednosti iz svakoga od triju prethodno opisanih skupova. Svaki panel ima zaseban slučajni odabir bez vraćanja, pa se točke ne odnose nužno na iste osobe. Poredak unutar panela omogućuje usporedbu oblika raspodjele.

Postotci unutar raspona mogu sakriti mjesto odstupanja. QQ prikaz zato poreda opažene vrijednosti i usporedi ih s vrijednostima koje bismo na istim položajima očekivali iz normalne raspodjele. Ako se oblici podudaraju, točke prate pravac. Vodoravna os prikazuje kvantile standardne normalne raspodjele, a okomita poredane vrijednosti u jedinicama promatrane varijable. Referentni pravac ovdje prolazi kroz parove prvoga i trećega kvartila. Odstupanja od njega pokazuju gdje oblik podataka odstupa od normalnoga oblika.



Slika 5: Poredane vrijednosti triju varijabli nasuprot kvantilima standardne normalne raspodjele. Referentni pravac određen je prvim i trećim kvartilom. Izrada autora. Podaci i postupak opisani su u tekstu.

Dnevne minute uglavnom prate pravac, uz odizanje desnog kraja zbog repa prema višim vrijednostima. Za spremnost na plaćanje točke najprije stoje vodoravno zbog nula, a zatim skreću prema gore. Nakon izdvajanja platitelja i logaritmiranja iznosa opet su blizu pravca.

QQ prikaz pokazuje gdje podaci odstupaju od pretpostavke. Prikladnost analize procjenjujemo prema tome koliko su ta odstupanja važna za odabrani postupak. Posljedice blagog odstupanja u repu ovise o ciljnoj veličini, postupku i utjecajnim opažanjima. Velika skupina nula jasno pokazuje da normalna raspodjela nije dobar opis sirovih iznosa.

Poznata vjerojatnost i učenje o nepoznatoj stopi različita su pitanja. Ako smo sami zadali pošten novčić, pet glava ne mijenja sljedeću vjerojatnost 0,5. Ako ne znamo udio uspješnih objava, pet uspjeha postaje informacija za njegovu procjenu. Tada nesigurnost proizlazi i iz toga što pravilo nastanka podataka tek trebamo naučiti.

Nizovi koje slučajnost proizvodi

Usporedbom s normalnom krivuljom provjeravali smo odgovara li model obliku podataka. U pitanju o vrućoj ruci trebamo provjeriti i način na koji iz niza biramo opažanja. **Slučajni niz** nastaje procesom sa slučajnom sastavnicom i ne mora u kratkom odsječku izgledati pravilno. Za provjeru vruće ruke simuliramo neovisne pokušaje s fiksnom vjerojatnošću pogotka, kao kod poštenog novčića. Ne pretpostavljamo da je stvarna košarka takva. Iz svakog niza izdvajamo pokušaje neposredno nakon triju pogodaka. Želimo vidjeti mijenja li samo to pravilo odabira prosječan udio pogodaka među izdvojenim pokušajima.

Pravilo odabira možemo vidjeti na namjerno složenom nizu 1, 1, 1, 0, 1, 1, 1, 1. Nakon triju jedinica stoje četvrti i osmi pokušaj, s ishodima 0 i 1. Udio na odabranim mjestima zato je $1/2$, iako u cijelom nizu ima sedam jedinica od osam. Posljednja tri pogotka ne stvaraju dodatno opažanje jer deveti pokušaj ne postoji. U simulaciji umjesto ovoga ilustrativnog niza svaki put generiramo novi niz neovisnih pokušaja s vjerojatnošću 0,5.

Za svaku kombinaciju duljine i uvjeta generiramo 8 000 nizova. Najprije računamo udio odabranih pogodaka unutar svakog niza, a zatim prosjek tih udjela, dajući svakom uključenom nizu jednaku težinu. Niz bez ijednog prikladnog mjesta nema definiran udio i izostavljamo ga iz toga prosjeka. U stupcima „Bez mjesta” prikazujemo koliko je takvih nizova, pa je nazivnik prosjeka 8 000 umanjeno za taj broj.

Tablica 2: Prosječan udio pogodaka na pokušajima koji slijede zadani niz pogodaka, u simulaciji poštenog novčića bez ikakve memorije. Izrada autora.

Duljina niza pokušaja	Nakon jednog	Bez mjesta nakon jednog	Nakon tri	Bez mjesta nakon tri
20	0,476	0	0,367	1846
50	0,490	0	0,421	146
100	0,494	0	0,459	2

U nizovima od stotinu pokušaja taj prosječni udio iznosi 0,459, a u nizovima od dvadeset 0,367. Oba su ispod polovine, iako smo simulirali proces bez memorije. Odstupanje stvara odabir mjesta i računanje udjela po konačnim nizovima.

Mehanizam se vidi već u trima bacanjima, kada tražimo ishod nakon jedne jedinice. Od osam jednako vjerojatnih nizova šest ima takvo mjesto. To su 010, 011, 100, 101, 110 i 111. Njihovi udjeli uspjeha na odabranim mjestima redom su 0, 1, 0, 0, 1/2 i 1. Prosjek tih šest udjela iznosi 5/12, približno 0,417. Nizovi s jednim i dvama odabranim mjestima dobivaju jednaku težinu. Takvo pridavanje težina spušta prosjek ispod 1/2 uz stalne izglede pri svakom bacanju. Pristranost je veća u kraćim nizovima i uz dulji uvjet. Nakon samo jednog pogotka u nizovima od stotinu pokušaja prosječni udio je 0,494, blizu polovine.

Za niz od stotinu pokušaja odstupanje je oko četiri postotna boda. Takav je pomak usporediv s razlikom koju bi istraživač mogao tražiti kao dokaz promjene forme. Zato treba provjeriti koliki dio rezultata nastaje samim postupkom mjerenja.

Kada iz podataka odabiremo jedinice prema obilježjima tih istih podataka, pravilo odabira postaje dio analize. Njegov učinak možemo provjeriti na procesu čije je pravilo poznato, ali tada moramo simulirati i odabir jedinica i završni izračun.

Vruća ruka i njezin ispravak. Kada igrač nekoliko puta uzastopno pogodi, suigrač bi mu mogao ponovno dodati loptu jer očekuje nastavak niza. To je moguća odluka kojoj pitanje o vrućoj ruci daje smisao. Zanima nas govori li prethodni uspjeh nešto o sljedećem pokušaju ili promatramo niz kakav može nastati i bez promjene vjerojatnosti pogotka.

Izvorno istraživanje vruće ruke ponudilo je tumačenje da ljudi u nizovima vide promjenu koje nema (Gilovich i sur. 1985.). Za provjeru takve tvrdnje naizgled je dovoljno izdvojiti pokušaje nakon pogodaka i usporediti njihovu uspješnost s uspješnošću pokušaja nakon promašaja. No izdvajanjem mjesta unutar konačnih nizova već smo uveli postupak odabira koji može promijeniti očekivani rezultat.

Kasnija metodološka analiza pokazala je da je uobičajena mjera ovisnosti osjetljiva na tu pristranost odabira. Nakon ispravka, ponovna analiza izvornog kontroliranog pokusa dala je zaključak u prilog vrućoj ruci (Miller i Sanjurjo 2018.). Promijenio se referentni račun s kojim se uspoređuje opaženi niz.

Naša simulacija pokazuje kako postupak odabira i prosječenja proizvodi pomak iako se zadana vjerojatnost ne mijenja. Prije tumačenja stvarnih pokušaja zato treba provjeriti cijeli postupak na procesu čije je pravilo poznato. Tako znamo koliki pomak mogu proizvesti sam odabir i računanje.

Odluka na terenu traži dodatnu prosudbu. Uz moguću promjenu forme treba razmotriti težinu sljedećeg šuta, obranu i raspoložive suigrače. Zato provjeravamo i dojam o vrućoj ruci i mjerni postupak kojim ga pokušavamo osporiti.

Za planiranje kampanje zanima nas barem jedna objava koja prijeđe prag. Provjerimo koje informacije taj račun traži uz vjerojatnost jedne objave.

NAĐITE GREŠKU

U nastavnom modelu svakoj objavi zadajemo vjerojatnost od dva posto da u sedam dana prijeđe dogovoreni prag dijeljenja. To je pretpostavka vježbe, a ne izmjeren udio. Urednik pita kolika je vjerojatnost da barem jedna od pet objava iste kampanje prijeđe prag.

U opisu kampanje piše da objave mogu doći do istih pratitelja i da se dijele tijekom istog tjedna. Nema podataka o tome pojavljuju li se njihovi uspjesi zajedno. Asistent je predložio račun komplementom.

```
p_viral <- 0.02
p_barem_jedna <- 1 - (1 - p_viral)^5
```

> Za jednu objavu vjerojatnost ostanka ispod praga iznosi 0,98. Za svih pet ona iznosi 0,98 na petu potenciju, pa komplement daje približno 9,6 %. Budući da je svaka objava zasebna jedinica, taj broj možemo upotrijebiti kao vjerojatnost uspjeha kampanje.

Napišite uredniku može li broj preuzeti bez dodatnog uvjeta. U odgovoru zadržite valjan račun, imenujte pretpostavku koju traži jedan njegov korak i navedite koju bi informaciju o objavama trebalo pribaviti. Ne morate izračunati novu vjerojatnost ako podaci za nju nisu zadani.

Razrađeni primjer

Zamislimo da je e-bilten poslan na pedeset adresa i otvoren četrnaest puta. Zadajemo hipotetski model s poznatom vjerojatnošću otvaranja od dvadeset dva posto za svaku poruku i neovisnim otvaranjima. Koliko je pod tim modelom vjerojatno četrnaest ili više otvaranja? Rezultat ćemo izračunati binomnom raspodjelom i provjeriti kroz dvadeset tisuća simuliranih

kampanja. Primjer nije mjerenje stvarne kampanje. Za traženu vjerojatnost koristimo već poznato pravilo komplementa. Od svih mogućih ishoda izostavit ćemo one s najviše trinaest otvaranja, pa ostaju četrnaest ili više.

Jedna simulirana kampanja sadrži 50 neovisnih odgovora „otvoreno" ili „nije otvoreno", s istih 22 % izgleda za svaku poruku. Njihov zbroj daje broj otvaranja jedne kampanje. Prvih pet generiranih zbrojeva iznosi 12, 12, 15, 11, 14. Svaki uspoređujemo s 14, a među svih 20 000 kampanja taj prag doseže njih 3862. Dijeljenjem s 20 000 dobit ćemo simuliranu vjerojatnost. Sva ponavljanja imaju definiran rezultat i ulaze u nazivnik.

```
1 - pbinom(13, size = 50, prob = 0.22)
```

```
[1] 0,1941652
```

```
set.seed(709)
kampanje <- rbinom(20000, size = 50, prob = 0.22)
mean(kampanje >= 14)
```

```
[1] 0,1931
```

Poziv `pbinom` provodi prvi dio tog računa, zbraja vjerojatnosti od nula do trinaest otvaranja. Oduzimanje od jedan ostavlja traženi dio raspodjele. Poziv `rbinom` zatim simulira kampanje prema istim pretpostavkama. Udio kampanja s najmanje četrnaest otvaranja trebao bi biti blizu izračunatoj vjerojatnosti, pa možemo usporediti računski rezultat s ponavljanjem postupka.

Izračun daje 19,4 %, a simulacija 19,3 %. Razlikuju se malo jer smo simulirali konačan broj kampanja. Podudaranje provjerava račun u zadanom modelu, bez neovisne potvrde modela u stvarnosti.

Četrnaest ili više otvaranja nastalo bi otprilike u jednom od pet ponavljanja. Ishod zato nije posebno neobičan pod modelom, ali to ne potvrđuje sve njegove pretpostavke. Ni mnogo rjeđi ishod

sam ne bi dokazao učinak naslova jer bi publika, vrijeme slanja i doseg ostali moguća objašnjenja. Za učinak naslova trebali bismo usporediti nasumično dodijeljene inačice. Za uredničku odluku uz to trebamo cilj i troškove pogreške.

Izračunali smo vjerojatnost ishoda u jednom repu raspodjele pod izričitim modelom.

Sažetak

Vjerojatnost nam omogućuje da opaženi rezultat usporedimo s ishodima koje proizvodi jasno zadan model, a kalibraciju prognoza provjeravamo kroz mnogo usporedivih predviđanja. Prije zbrajanja i množenja određujemo događaje, njihove preklape i uvjete, kao što binomni model traži neovisne binarne pokušaje s jednakom vjerojatnošću. Ponavljanjem vidimo i zašto raspodjela prosjeka pod uvjetima središnjega graničnog teorema može postati približno normalna, iako pojedinačni podaci to nisu. Pravilo o 68, 95 i 99,7 % primjenjujemo na onu raspodjelu koja je doista približno normalna. Simulacija vruće ruke pokazala je da u provjeru treba uključiti i pravilo odabira, jer i ono može promijeniti rezultat. U poglavlju o uzorkovanju istim ćemo ponavljanjem proučiti procjene iz različitih uzoraka.

Pojmovi

vjerojatnost (*probability*), zakon velikih brojeva (*law of large numbers*), kalibrirana nesigurnost (*calibrated uncertainty*), neovisnost (*independence*), uvjetna vjerojatnost (*conditional probability*), binomna raspodjela (*binomial distribution*), normalna raspodjela (*normal distribution*), QQ prikaz (*Q–Q plot*), slučajni niz (*random sequence*)

Zadaci

Konceptualni

Usporedite niz od pet glava iz modela poštenoga novčića s poznatom vjerojatnošću i pet uspješnih objava kampanje čija stopa uspjeha nije poznata. Objasnite zašto zadana vjerojatnost sljedećeg bacanja ostaje ista, dok uspjesi mogu promijeniti uvjerenje o nepoznatoj stopi kampanje čak i kada uspjeh jedne objave ne mijenja izglede sljedeće. Odvojeno navedite koji bi dokaz bio potreban za tvrdnju da takva promjena ipak postoji. Na kraju prosudite smije li se pravilo 68–95–99,7 primijeniti na raspodjelu s velikom skupinom nula i navedite što bi u QQ prikazu poduprlo vaš odgovor.

Računski

Vratite se na hipotetsku provjeru zapisa iz poglavlja o tome kako brojke zavode. Od 10 000 zapisa njih 100 stvorio je program. Provjera označuje 90 od tih 100 i pogrešno označuje 495 od preostalih 9 900. Na tvrdnju da je provjera pouzdana jer pronalazi devedeset posto ciljanih zapisa primijenite svih šest revizijskih pitanja. Iz navedenih brojnosti izračunajte udio upozorenja među ciljanim zapisima, udio ciljanih zapisa među svim upozorenjima i temeljnu stopu. Predajte kratku revizijsku tablicu, račun i dvije rečenice o tome što rezultati dopuštaju zaključiti.

Vraćamo se hipotetskom modelu dvadeset neovisnih objava, od kojih svaka ima vjerojatnost uspjeha 0,02. Najprije izračunajte vjerojatnost da nijedna ne uspije, a zatim pravilom komplementa vjerojatnost barem jednoga uspjeha. Navedite zašto isti račun nije opravdan ako uspjeh jedne objave mijenja vjerojatnost uspjeha sljedeće.

Kritički

Prosudite kako je nalaz o vrućoj ruci prešao put od mjerenja do općenite tvrdnje o ljudskoj procjeni slučajnosti, i što se u toj tvrdnji promijenilo nakon ispravka mjere (Gilovich i sur. 1985.; Miller i Sanjurjo 2018.). Predajte jedan odlomak i imenujte rečenicu koju bi popularni prikaz smio zadržati.

Revizija modela

Provjerite analizu iz okvira. Izdvojite ispravne računske korake i označite račun koji vrijedi samo uz neovisnost. Zatim izdvojite rečenicu koja neovisnost neopravdano smatra potvrđenom. Napišite zamjenu koja jasno ograničava zaključak.

08

DIO III · OD UZORKA DO POPULACIJE

uzorkovanje

Jedan uzorak skriva raspodjelu mogućih rezultata

Kad uz internetsku anketu vidimo velik broj odgovora, lako joj pripišemo veću uvjerljivost. Zamislimo dvije prognoze istih izbora, jednu s nekoliko stotina sudionika i drugu s milijunima. Prije nego što pročitamo kako su prikupljene, koja nam ulijeva više povjerenja?

Časopis *Literary Digest* pred američke predsjedničke izbore 1936. poslao je milijune probnih listića i iz vraćenih odgovora predvidio pobjedu Alfa Landona. Pobijedio je Franklin Roosevelt (Squire 1988.). Pogrešna prognoza imala je iza sebe golemu količinu odgovora stvarnih ljudi.

Kako toliki broj odgovora može promašiti čak i pobjednika, a znatno manji uzorak ipak pružiti korisnu procjenu?

Da bismo razumjeli taj promašaj, trebamo odvojiti dvije stvari. Jedna je koliko bi se rezultat mijenjao pri ponavljanju odabira,

a druga kamo nas sam postupak odabira sustavno usmjerava. Slučajnost iz prethodnog poglavlja sada ulazi u sam odabir ljudi koje promatramo. Ako iz iste populacije uzmemo drugi uzorak, promijenit će se sastav sudionika, a s njime i izračunati prosjek. Uz razlike među ljudima sada istražujemo koliko bi se mijenjao naš odgovor na istraživačko pitanje kada bismo odabir mogli ponavljati.

Za tu ćemo provjeru iskoristiti prednost nastavne simulacije. Skup `populacija_medija` sadrži 50 000 računalno stvorenih odraslih osoba i omogućuje nam da poznajemo cijelu populaciju. Jedan redak predstavlja jednu osobu, a pratit ćemo povjerenje u medije, dnevno praćenje medija, izvor vijesti i spremnost na plaćanje. To je ista zamišljena populacija kojom smo se već služili, odvojena od manje ankete `anketa_mreze`. Iz nje možemo izvlačiti uzorke i njihove rezultate uspoređivati s poznatom ciljnom vrijednošću. Ti zapisi ne opisuju stvarne stanovnike. U stvarnom istraživanju ciljnu vrijednost obično ne vidimo.

Ponavljanjem odabira dobit ćemo raspodjelu uzoračkih rezultata, čije ćemo središte i širinu opisati sredinom i standardnom devijacijom. Standardnu pogrešku procjene pritom ćemo razlikovati od raspršenosti pojedinačnih odgovora. Mijenjanjem veličine uzorka i oblika populacije ispitat ćemo kada raspodjela sredina postaje približno zvonasta i zašto se njezina širina smanjuje sporije nego što raste broj sudionika.

Ta promjenjivost samo je dio problema ankete. Vratit ćemo se okviru uzorkovanja, odazivu i istraživačkom nacrtu iz poglavlja o mjerenju te provjeriti što veći uzorak može popraviti, a što ostaje izvan računa. Ista pitanja postavit ćemo i izboru tekstova u korpus. Pri procjeni uvjerljivosti istraživanja moći ćemo razlikovati slučajnu nesigurnost od sustavnog izostavljanja jedinica. Za to najprije trebamo jasno odvojiti populaciju koju želimo opisati od uzorka koji nam je dostupan.

Populacija, uzorak i pogreška uzorkovanja

Populacija je skup jedinica o kojima želimo zaključivati, a **uzorak** dio koji smo promatrali. Postupak odabira povezuje ih tako što određuje koje populacijske jedinice mogu ući u uzorak i s kojim izgledima.

Parametar je vrijednost koja opisuje populaciju. Populacijsku sredinu označujemo s μ, a standardnu devijaciju s σ. **Statistika** je vrijednost izračunata iz uzorka, primjerice sredina $\bar{x}$ ili standardna devijacija s. Veličine uzorka i populacije označujemo s n i N. Statistiku koristimo za procjenu parametra, uz očekivanje da se mogu razlikovati.

Prije izvlačenja uzoraka izračunajmo parametre cijele `populacija_medija`. Te će nam vrijednosti služiti kao cilj s kojim uspoređujemo svaku uzoračku procjenu.

Prosječno povjerenje u medije je 4,88 na ljestvici od 1 do 10, uz standardnu devijaciju 1,98. Prosječno dnevno praćenje medija iznosi 174,5 minuta, a portal kao glavni izvor bira 30,2 % osoba. To su svojstva naše simulirane populacije, bez tvrdnje o stvarnom gradu.

Za jedan od tih računa nasumično odabiremo 100 različitih osoba bez vraćanja. Početak izvučenog uzorka prikazan je u tablici. Prvi odgovor povjerenja iznosi 6, ali se sredina računa iz svih 100 odgovora. Njihov zbroj 500 dijelimo sa 100. Za udio portala umjesto zbroja ocjena brojimo 32 osoba s tom oznakom izvora, uz isti nazivnik.

Tablica 1: Prvih pet zapisa jednog uzorka od 100 osoba. Izrada autora.

Redak	Povjerenje	Izvor
1	6	radio
2	7	TV
3	6	tisak
4	3	društvene mreže
5	3	portal

U jednom uzorku od stotinu osoba prosječno povjerenje iznosi 5,00, a udio korisnika portala 32,0 %. Vrijednosti su bliske populacijskima, ali nisu jednake.

> **Pogreška uzorkovanja** je razlika između vrijednosti izračunate na uzorku i vrijednosti koju ima cijela populacija, a nastaje zato što je izmjeren dio umjesto cjeline.

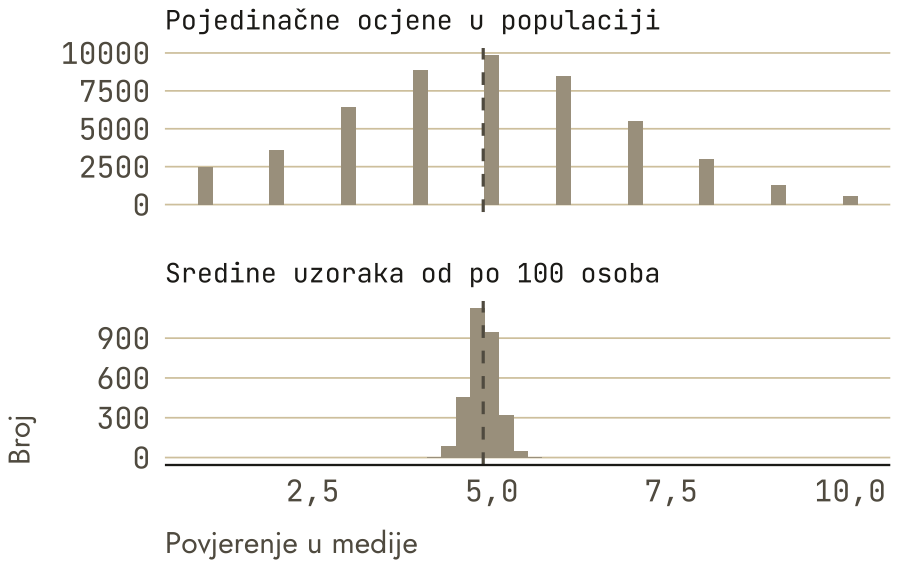
Pogreška ovdje označuje razliku koja nastaje zato što promatramo samo dio populacije. Moguća je i uz potpuno ispravan unos podataka i postavljanje pitanja. Tu slučajnu promjenjivost treba razlikovati od sustavnog odstupanja zbog pristranog odabira.

Pri novom odabiru ponovno je dostupna cijela populacija. Unutar jednog uzorka osoba ulazi najviše jednom, ali može ući u više različitih uzoraka. Svako novo izvlačenje ponovno kreće iz cijele populacije, a prethodnih 100 zapisa čini samo jedan od mogućih uzoraka. Populacija i pravilo odabira ostaju fiksni, mijenjaju se odabrane osobe. Tri neovisna uzorka od po stotinu osoba daju sredine 4,57, 4,47 i 4,84. Dobivene su iz iste populacije istim postupkom, uz raspon od 0,37 boda.

Nakon tri tisuće ponavljanja sredine možemo prikazati kao raspodjelu. Ona pokazuje gdje se procjene skupljaju i koliko se međusobno razlikuju. Njezine jedinice su procjene iz uzoraka.

> **Distribucija uzorkovanja** je raspodjela vrijednosti koje bi neka statistika poprimila kroz mnogo ponovljenih uzoraka iste veličine iz iste populacije.

Dvije raspodjele odgovaraju na različita pitanja. Raspodjela opažanja pokazuje koliko se ljudi razlikuju. Distribucija uzorkovanja pokazuje koliko se mijenja procjena kroz različite uzorke. Obje imaju sredinu i standardnu devijaciju, pa je važno svaki put navesti što raspodjela opisuje.



Slika 1: Raspodjela pojedinačnih ocjena povjerenja u simuliranoj populaciji i raspodjela sredina tri tisuće uzoraka od po sto osoba, na zajedničkoj osi. Izrada autora. Podaci i postupak opisani su u tekstu.

Uzoračke sredine ovdje su raspoređene oko populacijske sredine. Procjenitelj zato dugoročno ne promašuje sustavno naviše ili naniže i nazivamo ga **nepristranim**. Uži donji panel pokazuje manju promjenjivost sredina uzoraka. Pojedinačne ocjene ljudi ostaju jednako raspršene.

Standardna pogreška

Uži donji panel sada želimo opisati brojem. Raspršenost procjena također mjerimo standardnom devijacijom. Da bismo je razlikovali od raspršenosti pojedinačnih opažanja, nazivamo je standardnom pogreškom.

> **Standardna pogreška** je standardna devijacija distribucije uzorkovanja, pa opisuje koliko bi tipično varirala procjena kroz ponovljene uzorke.

Pojedinačne ocjene imaju standardnu devijaciju 1,98, dok standardna devijacija sredina iz tri tisuće uzoraka iznosi 0,198. Potonja je standardna pogreška, a prva je od nje 10,0 puta veća. Uzorci imaju po stotinu osoba, a korijen iz sto jest deset. U rezultatima se tako već nazire odnos koji će opisati formula.

U sredini malog uzorka svaka osoba ima veću težinu. Jedan odgovor u uzorku od deset ljudi čini desetinu računa, a u uzorku od tisuću ljudi tisućinu. U većem se uzorku neovisna pozitivna i negativna odstupanja više poništavaju. Zato sredina manje varira. Dobitak u preciznosti raste s korijenom broja opažanja.

Dosad koristimo **jednostavni slučajni uzorak** (*simple random sample*), u kojem svaki mogući skup od n jedinica ima jednaku vjerojatnost odabira. Jedinice iz konačne populacije biramo bez vraćanja. Sljedeća formula točna je za neovisne odabire s vraćanjem, a dobra je aproksimacija za odabir bez vraćanja ako je n malen prema N. Za druge nacrte slučajnog uzorkovanja može trebati drukčiji račun.

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Oznaka $SE_{\bar{x}}$ znači standardnu pogrešku sredine, σ populacijsku standardnu devijaciju, a n veličinu uzorka. Formula ovdje daje 0,1984, dok iz ponavljanja dobivamo 0,1981. Rezultati se slažu na tri decimale, iako su dobiveni različitim postupcima.

Odabir bez vraćanja dodatno smanjuje nesigurnost. Ako uključimo cijelu populaciju, pogreške uzorkovanja više nema. **Korekcija za konačnu populaciju** (*finite-population correction*) uzima u obzir taj dobitak množenjem početnog izraza korijenskim faktorom.

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Kada je uzorkovan malen dio populacije, drugi korijen blizu je jedan i malo mijenja rezultat. Kako se n približava N, faktor se smanjuje do nule. Simulacija bez vraćanja već uključuje taj učinak. U sljedećoj tablici stupac s formulom koristi aproksimaciju za veliku populaciju, bez te korekcije.

Svaka simulacijska standardna pogreška u tablici ispod nastaje iz 1 500 uzoračkih sredina za navedeni broj osoba. Računamo standardnu devijaciju tih 1 500 brojeva, bez izostavljenih ponavljanja. To mjeri preciznost, dok udaljenost njihova prosjeka od poznate populacijske sredine služi provjeri pristranosti. Konačan broj ponavljanja donosi i malo simulacijsko kolebanje, pa sitne razlike prema formuli ne tumačimo kao dokaz sustavnoga odstupanja.

U praksi nepoznatu σ procjenjujemo uzoračkom standardnom devijacijom s. Sada možemo provjeriti i raniji izbor djelitelja $n - 1$ za varijancu. U četiri tisuće uzoraka od po deset osoba dijeljenje kvadriranih odstupanja s deset daje prosječno 3,55, dok je populacijska varijanca 3,94. Dijeljenje s devet daje 3,94. Djelitelj n sustavno podcjenjuje varijancu jer odstupanja računamo od sredine istog uzorka, koja tim vrijednostima pristaje bliže od populacijske sredine.

Tablica pokazuje kako preciznost raste s veličinom uzorka. Za svako dodatno smanjenje standardne pogreške trebamo sve više opažanja.

Tablica 2: Standardna pogreška sredine povjerenja pri osam veličina uzorka, izračunata formulom i izmjerena na tisuću i petsto ponovljenih uzoraka. Izrada autora.

Veličina uzorka	Formula	Simulacija	Raspon ±1,96 SE
10	0,627	0,617	±1,23
25	0,397	0,405	±0,78
50	0,281	0,282	±0,55
100	0,198	0,194	±0,39
200	0,140	0,141	±0,27
500	0,089	0,088	±0,17
1 000	0,063	0,065	±0,12
2 000	0,044	0,043	±0,09

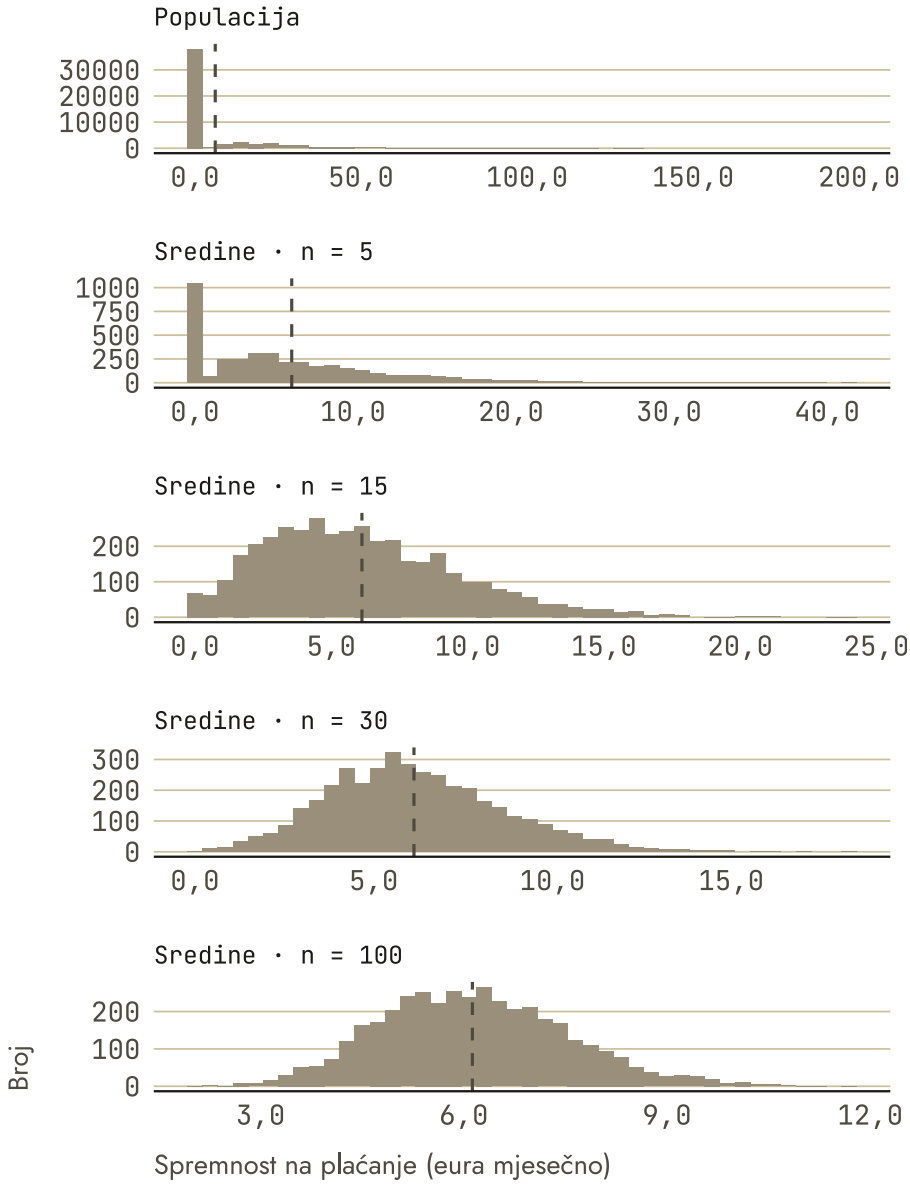
Svako deseterostruko povećanje uzorka smanjuje standardnu pogrešku za faktor $\sqrt{10}$, približno 3,16. Da bi se standardna pogreška prepolovila, uzorak se mora učetverostručiti. Za svako daljnje smanjenje slučajne raspršenosti zato treba prikupiti sve više opažanja.

Oblik koji se pojavljuje

Zadržavamo istu populaciju osoba, ali povjerenje zamjenjujemo iznosom spremnosti na plaćanje. Za svaku od veličina 5, 15, 30 i 100 izvučemo 4 000 novih uzoraka bez vraćanja i za svaki spremimo jednu sredinu. Nule su valjani odgovori i ostaju u svim računima. Tako promjena oblika koju ćemo gledati nastaje usrednjavanjem, bez uklanjanja neplatitelja.

Uz širinu treba pogledati i oblik distribucije uzorkovanja. Spremnost na plaćanje u populaciji ima 76,1 % nula i dug desni rep, uz asimetriju 3,2. Izvorne vrijednosti izrazito odstupaju od normalne raspodjele.

Za uzorke od pet osoba asimetrija raspodjele sredina je 1,47. Za petnaest pada na 0,76, za trideset na 0,52, a za sto na 0,32. Mijenja se raspodjela sredina, dok izvorna populacija ostaje ista.



Slika 2: Izrazito asimetrična populacijska varijabla i raspodjele njezinih uzoračkih sredina pri četiri veličine uzorka. Svaki panel ima vlastitu os. Izrada autora. Podaci i postupak opisani su u tekstu.

Središnji granični teorem (*central limit theorem*) u ovom modelu opisuje približavanje distribucije sredine normalnoj raspodjeli s rastom jednostavnog slučajnog uzorka. Potrebna je konačna varijanca i raspodjela u kojoj nekolicina jedinica ne nosi gotovo cijeli zbroj. Dovoljna veličina uzorka ovisi o obliku i repovima populacije. U našem je primjeru i pri trideset osoba ostalo dosta asimetrije, dok za gotovo simetričnu populaciju manji uzorak može biti dovoljan.

Mnogi se postupci zato oslanjaju na približnu normalnost raspodjele procjene, dok pojedinačna opažanja mogu imati drukčiji oblik. Primjena teorema i dalje zahtijeva odgovarajuće uvjete odabira i neovisnosti. Kod složenijeg uzorkovanja težine i zajednički odabir skupina mogu promijeniti i oblik i širinu raspodjele procjene. Račun mora pratiti taj nacrt.

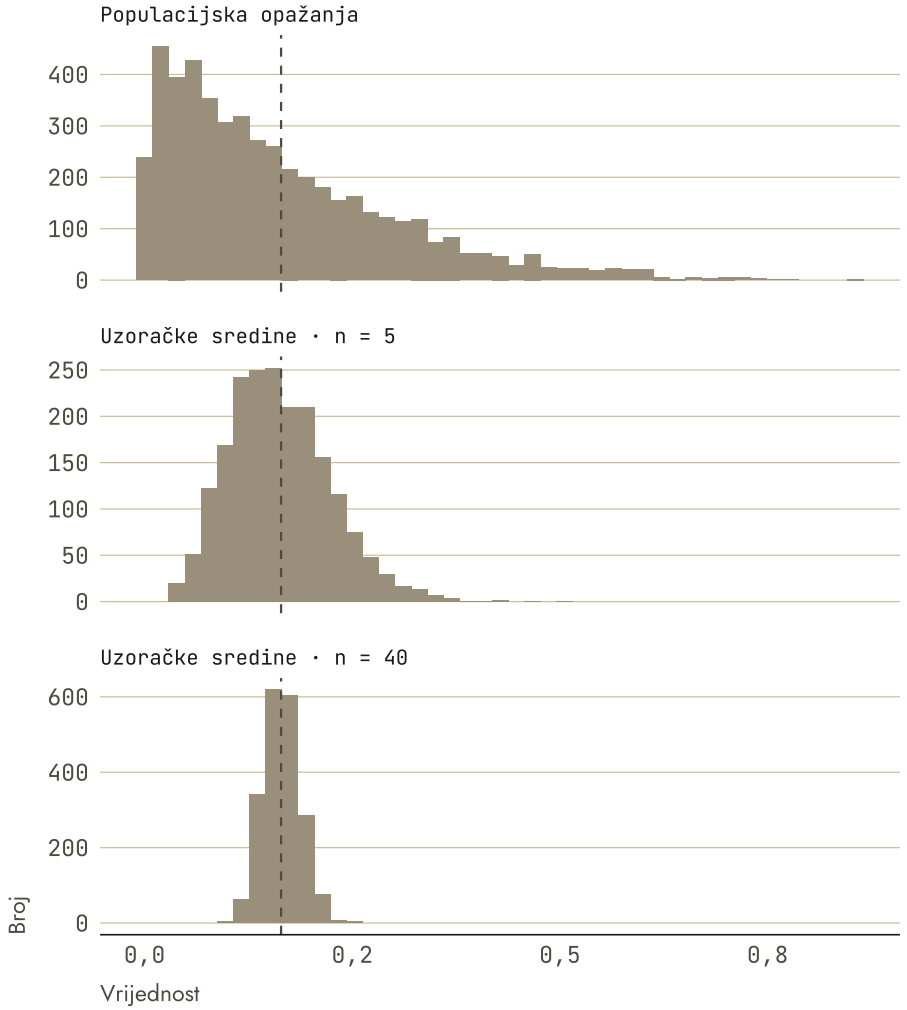
Dva prikaza empirijske standardne pogreške potječu iz zasebnih simulacija. Raspodjela sredina pri stotinu osoba ima 3 000 ponavljanja, a usporedna tablica za svaku veličinu uzorka koristi 1 500 novih ponavljanja. Male razlike njihovih procjena SE-a zato odražavaju i konačan broj simulacija.

CLT stroj

Za interaktivni pokus napuštamo mjerne jedinice medijskih podataka i koristimo zasebne umjetne raspodjele. Tiskani prikaz polazi od 5 000 generiranih vrijednosti između nule i jedinice, s mnogo malih vrijednosti i sve manje velikih. Iz tog fiksnog skupa izvlačimo s vraćanjem, po 2 000 uzoraka veličine pet i četrdeset. Jedno izvlačenje daje jedan broj, a prosjek cijelog uzorka jednu vrijednost donjeg histograma. Ovdje je vraćanje dio zadanoga modela uzorkovanja iz poznate populacije.

Uz populacijsku raspodjelu prikazane su raspodjele uzoračkih sredina za dvije veličine uzorka. Populacija je u oba slučaja

ista. Usporedba pokazuje kako veći uzorak mijenja oblik raspodjele sredina i smanjuje njezinu širinu. Pri fiksnoj populaciji standardna pogreška smanjuje se s korijenom veličine uzorka.



Slika 3: Desno asimetrična simulirana populacija i raspodjele sredina pri dvjema veličinama uzorka. Veći uzorak daje pravilniju i užu raspodjelu sredina. Izrada autora. Podaci i postupak opisani su u tekstu.

Preciznost ankete od osamsto ljudi

U anketi često želimo procijeniti udio, primjerice koliko ljudi bira portal, umjesto prosječne ocjene povjerenja. Za to možemo iskoristiti poznati račun. Izbor portala označimo jedinicom, a ostale izbore nulom, pa sredina tih vrijednosti daje **uzorački udio**, označen s $\hat{p}$. Njime procjenjujemo populacijski udio, uz istu logiku uzorkovanja sredine koju smo upravo upoznali. Raspršenost binarne varijable određena je njezinim udjelom. Najveća je kada je populacija podijeljena napola, a manja blizu nule ili jedan.

Standardnu pogrešku ukupnog udjela u velikoj populaciji procjenjujemo sljedećim izrazom za jednostavni slučajni uzorak. Nepoznati populacijski udio zamjenjujemo procjenom $\hat{p}$, pa je i standardna pogreška procijenjena.

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

Kod jednostavnoga slučajnog uzorka bez vraćanja izraz se množi istom korekcijom za konačnu populaciju kao i standardna pogreška sredine. Kod nejednakih vjerojatnosti odabira ili grupnog uzorkovanja ne smije se prenijeti bez prilagodbe nacrtu.

Margina pogreške (*margin of error*) polovica je širine intervala oko procjene. U prikazanom računu najveća je za udio od 50 %, jer je tada raspršenost najveća. Za razinu od 95 % i veliku populaciju približno iznosi jedan podijeljen korijenom veličine uzorka.

U sljedećem računu marginu dobivamo množenjem procijenjene standardne pogreške s 1,96. Obrnutim računom određujemo potreban uzorak za željenu marginu. Vrijede pretpostavke jednostavnog slučajnog uzorka i velike populacije te razumna normalna aproksimacija. Ona može biti neprikladna ako je očekivan broj jedne kategorije malen.

Tablica 3: Najveća margina pogreške za udio pri razini od 95 % i uzorak potreban za zadanu marginu. Izrada autora.

Veličina uzorka	Margina pogreške	Ciljana margina	Potreban uzorak
400	±4,9 %	±5 %	385
800	±3,5 %	±4 %	601
1 000	±3,1 %	±3 %	1 068
2 000	±2,2 %	±2 %	2 401

Tablica se odnosi na jedan ukupni udio, jednostavni slučajni uzorak i razinu od 95 %, bez korekcije za konačnu populaciju. Za osamsto ljudi najveća margina je približno ±3,5 %. Takva procjena razlikuje veličine udaljene desetak postotnih bodova bolje nego one udaljene dva ili tri. Drugi cilj procjene, nacrt ili zahtjev za preciznošću traži novi račun.

Kada izvještaj navodi da prva opcija ima 32 %, a druga 29 %, same dvije procjene i njihove pojedinačne margine nisu dovoljne za tvrdnju da je prva opcija ispred druge. Razlika dvaju udjela ima vlastitu nesigurnost, koju treba izračunati iz zajedničkoga nacrta.

Veličina populacije izostaje iz jednostavne formule jer smo zanemarili korekciju za konačnu populaciju. To je opravdano kada uzimamo malen dio populacije. Manju populaciju čini slučajno izdvojenih 5000 vrijednosti povjerenja iz iste populacije `populacija_medija`. Iz tako određenih cjelina zasebno izvlačimo uzorke. U simulaciji uzorak od osamsto osoba iz cijele populacije daje standardnu pogrešku 0,0700, a iz deset puta manje populacije 0,0640. Uzorak u drugom slučaju obuhvaća veći dio cjeline, pa je precizniji. Populacijske devijacije iznose 1,984 i 1,967. Formula s korekcijom uzima u obzir i tu malu razliku te daje 0,0696 i 0,0637, blizu simuliranim rezultatima.

Dovoljnu veličinu uzorka određujemo prema pitanju koje istražujemo. Trebamo znati procjenjujemo li ukupni udio, sredinu, rijetku pojavu ili razliku skupina. Trebamo i nacrt, način odabira

i odaziva te željenu preciznost. Preciznost za podskupinu ovisi o broju njezinih članova. Oko tisuću može biti korisna orijentacija za jedan ukupni udio u jednostavnom nacrtu. Za ostale ciljeve veličinu određujemo zasebno.

Margina pogreške opisuje promjenjivost zbog slučajnog odabira. Ne uključuje ljude izvan okvira, neodaziv ni učinak formulacije i redoslijeda pitanja. Oznaka ±3 % u tom kontekstu znači tri postotna boda margine za uzorkovnu promjenjivost. Tvrdnja da je razlika unutar margine zato ne potvrđuje kvalitetu svih ostalih dijelova istraživanja.

Nacrt uzorka i pristranost odabira

Dosadašnji račun polazio je od jednakih izgleda za odabir. U stvarnoj anketi neke osobe mogu imati veće izglede od drugih, pa njihov doprinos procjeni trebamo prilagoditi. U **vjerojatnosnom uzorkovanju** svaka jedinica ima poznatu pozitivnu vjerojatnost uključivanja, ali ona ne mora biti jednaka za sve. Jedinici s upola manjom vjerojatnošću dajemo približno dvostruk doprinos procjeni. Početna **težina uzorkovanja** računa se kao obrnuta vjerojatnost uključivanja. Bez težina bismo previše zastupali češće birane jedinice. Ni pravilne težine, međutim, ne vraćaju ljude koji su potpuno izostali iz okvira.

Učinak težina prikazat ćemo na šest odabranih jedinica iz zasebne sintetičke populacije od šesnaest osoba. U svakom od dvaju slojeva ima osam osoba. Jedinica označuje potporu zamišljenoj mjeri, a nula izostanak potpore. Vjerojatnosti uključivanja su 0,50 u prvom i 0,25 u drugom sloju, pa su težine dva i četiri.

Tablica 4: Opaženi uzorak iz sintetičke konačne populacije s poznatim vjerojatnostima uključivanja. Izrada autora.

Jedinica	Sloj	Vjerojatnost uključivanja	Odgovor	Težina uzorkovanja
A1	češće uključeni	0,50	1	2
A2	češće uključeni	0,50	1	2
A3	češće uključeni	0,50	1	2
A4	češće uključeni	0,50	0	2
B1	rjeđe uključeni	0,25	0	4
B2	rjeđe uključeni	0,25	0	4

Bez težina dijelimo brojnik 3 nazivnikom 6 i dobivamo 50,0 %. S težinama dobivamo brojnik 6, nazivnik 16 i udio 37,5 %. Procjena pada jer dvije nule iz rjeđe biranog sloja više pridonose nazivniku. Dobivenih 37,5 % i dalje je procjena, s mogućom pogreškom uzorkovanja. Primjer provjerava račun na vidljivoj sintetičkoj tablici, bez tvrdnje o stvarnim ljudima.

Složeni uzorci traže dodatne prilagodbe. **Kalibracija težina** usklađuje ponderirane zbrojeve s poznatim obilježjima populacije. Pri **grupnom uzorkovanju** biramo skupine, primjerice škole, pa osobe unutar njih. Sličnost tih osoba može smanjiti količinu neovisne informacije. **Učinak nacrta** uspoređuje varijancu procjene s onom jednostavnoga slučajnog uzorka, a **efektivna veličina uzorka** izražava istu preciznost brojem jedinica takva uzorka.

Pri čitanju izvještaja provjeravamo uključuje li račun nesigurnosti težine, skupine i slojeve. Neodaziv i potpunost okvira zah-

tijevaju zasebnu provjeru i obrazložene prilagodbe. Neobvezni odjeljak praktikuma[1] podrobnije objašnjava te prilagodbe.

Pokrivenost, odabir i odaziv

Težinama smo uzeli u obzir različite izglede uključenih jedinica. One ipak ne rješavaju problem ljudi koji uopće nisu mogli biti odabrani. **Pogreška pokrivenosti** nastaje kada okvir iz kojega se bira ne dopušta svim jedinicama ciljne populacije da uopće uđu u uzorak. Povećanje broja odabranih jedinica unutar istoga okvira ne doseže one koji u njemu nedostaju.

U prigodnom uzorku biramo dostupne ljude, bez poznatih vjerojatnosti odabira. U našoj populaciji mlađih od dvadeset pet godina s višim obrazovanjem ima 1 915. Njihovo prosječno povjerenje iznosi 4,09, dok je populacijsko 4,88. Razlika nadmašuje cijelu marginu ankete s tisuću ljudi. Uključivanjem svih članova te uske skupine točno bismo opisali upravo nju, uz istu razliku prema cijeloj populaciji.

U sljedećem modelu samoodabira ponovno razmatramo svih 50 000 osoba. Osobi mlađoj od 30 koja bira mreže dajemo 15 % izgleda za uključivanje, a ostalima te dobi 8 %. U dobi od 30 do 49 godina izgledi su 6 % za portal ili mreže i 3 % za ostale izvore. Od 50 godina naviše odgovarajući izgledi iznose 2 % i 0,5 %. Svakoj osobi zatim neovisno odredimo ulazi li u anketu. Ta smo pravila zadali za ovu ilustraciju samoodabira.

U dobrovoljnu mrežnu anketu ulaze ljudi koji su na internetu, vide poziv i odluče odgovoriti. Svaki korak može mijenjati sastav. Simulacija takvog odabira daje 2 676 odgovora, mnogo više od početnog slučajnog uzorka od sto osoba. Prosječna dob je 32,1, umjesto populacijskih 42,2 godina. Mreže bira 47,8 %, naspram 26,8 %, a povjerenje je 4,42, naspram 4,88.

[1] https://lusiki.github.io/statistika-knjiga/dodaci/a-praktikum.html#sec-a-slozeni-uzorci

Čak i uz dobar okvir i slučajni odabir dio pozvanih može izostati. **Neodgovor** smanjuje uzorak i preciznost, a može donijeti i pristranost. Ako se odazvani i neodazvani razlikuju u obilježju koje procjenjujemo, primjerice zbog interesa za temu ili povjerenja u istraživača, sam slučajni odabir nije dovoljan. Provjeravamo i odabir i odaziv.

Velik pristran uzorak može dati uzak interval oko sustavno pogrešne vrijednosti. Standardna pogreška prikazuje slučajnu promjenjivost pod nacrtom, ali ne i svu pristranost odabira. Povećanje uzorka uz isti pristrani postupak zato ne jamči točniju procjenu.

Sada možemo objasniti promašaj časopisa *Literary Digest*. Na temelju naknadne Gallupove ankete Squire je utvrdio da su pogrešnoj prognozi pridonijeli i sastav primatelja i njihov nejednak odaziv (Squire 1988.). Njegova analiza upućuje na to da bi potpuni odaziv izabranih primatelja, unatoč problemu njihova odabira, barem ispravno imenovao Roosevelta kao pobjednika (Squire 1988.). Zato popis onih kojima smo poslali pitanje i popis onih koji su odgovorili pregledavamo zasebno. Još odgovora prikupljenih na isti način može smanjiti slučajno kolebanje, dok sustavno odstupanje ostaje.

Zadržavamo sedam polja kartice za čitanje ankete iz poglavlja o mjerenju. Sada pitanja unutar tih polja preciziramo znanjem o procjeni i margini. Svako polje pokazuje dio postupka koji određuje na koga se rezultat može odnositi.

Tablica 5: Kartica za čitanje ankete nakon poglavlja o uzorkovanju. Izrada autora.

Polje kartice	Pitanje nakon ovoga poglavlja
okvir i pozivanje	Koja je ciljna populacija, tko je mogao ući u uzorak i je li odabir vjerojatnosan?
odaziv i nedostajanje	Koliko je pozvanih odgovorilo i kako bi se neodazvani mogli razlikovati?
procjena s težinama	Kako su uključivanje i sastav prilagođeni težinama i što ostaje neprilagođeno?
mjerenje	Kako glasi pitanje i koji odgovor ulazi u brojnik i nazivnik?
provedba	Kada i kako je anketa provedena te tko ju je naručio?
doseg	Na koju se podskupinu procjena odnosi i koga okvir, odaziv ili filtri izostavljaju?
nesigurnost	Uvažava li račun težine, slojeve i skupine stvarnoga nacrta?

Korpus kao uzorak

Iste granice odabira susrećemo kada umjesto ljudi prikupljamo njihove tekstove. I korpus tekstova zahtijeva ciljnu populaciju. Za opis političke rasprave trebamo navesti uključujemo li govore, intervjue ili objave, s kojih platformi ili arhiva, iz kojeg razdoblja, na kojim jezicima i od kojih govornika. Dostupnost tekstova tek je polazište. Populaciju određujemo prema sadržajnom pitanju i navedenom obuhvatu.

Obrisane i privatne objave mogu izostati iz arhiva, stariji sadržaj biti nedostupan, a odabrana ključna riječ propustiti relevantne tekstove. Veći broj riječi ne uklanja takve propuste ni činjenicu da su se neki govornici sami odlučili oglasiti. Zaključak zato ograničavamo na obuhvaćene vrste tekstova, platforme, datume, jezike i govornike.

Treba odrediti i jedinicu brojanja. Brojanje riječi daje više težine duljim tekstovima, a brojanje objava aktivnijim govornicima. Udio govornika odgovara na drugo pitanje. Uz zastupljenost zato navodimo brojimo li riječi, tekstove, istupe ili ljude, ponavljaju li se jedinice istog govornika i koji nazivnik koristimo.

Jedan dio korpusa možemo upotrijebiti za izradu pravila, a drugi sačuvati za provjeru na jedinicama koje pri izradi nismo vidjeli. Ista osoba ili gotovo isti tekst ne smije se naći s obiju strana takve provjere. Time provjeravamo predviđanje na novim jedinicama iz sličnog procesa prikupljanja. Ta podjela ne popravlja pokrivenost korpusa niti opravdava zaključak o izostavljenim tekstovima i govornicima. Prediktivna provjera i populacijsko uopćavanje zahtijevaju zasebna obrazloženja.

● STATISTIKA U DIVLJINI

Isti upitnik, različiti putovi do sudionika. Usporedba Pew Research Centera, objavljena 2023., koristila je isti upitnik u šest internetskih uzoraka odraslih stanovnika SAD-a. Tri su nastala iz panela s vjerojatnosnim početnim odabirom, a tri dobrovoljnim pristupanjem. Procjene su uspoređene s vanjskim referentnim vrijednostima, među kojima su bili službeni izborni rezultati i podaci kvalitetnih državnih anketa (Mercer i Lau 2023.). Ispunjavanje obrasca na internetu bilo je zajedničko svima. Put kojim je osoba došla do obrasca razlikovao se.

Procjene iz panela s vjerojatnosnim odabirom u toj su usporedbi u prosjeku manje odstupale od referentnih vrijednosti, uz isti postupak ponderiranja svih uzoraka (Mercer i Lau 2023.). Referentne vrijednosti također mogu imati pogrešku, pa ih čitamo kao najbolju dostupnu usporedbu (Mercer i Lau 2023.). Kad sljedeći put otvorimo mrežnu anketu, vrijedi pogledati

kako smo dobili poziv i tko ga uopće može dobiti. Oznaka „internetska anketa" sama po sebi još ništa ne govori o tom odabiru.

Pri objašnjavanju koristi većeg uzorka brojčani rezultat treba povezati s onim što je u postupku uzorkovanja doista postalo preciznije.

NAĐITE GREŠKU

Za objašnjenje standardne pogreške pripremamo nastavnu usporedbu dviju veličina uzorka. U zadanom modelu nasumično i neovisno izvlačimo osobe iz iste zamišljene populacije. Standardna devijacija njihova dnevnog vremena uz medije iznosi 20 minuta. Tablica sadrži teorijske vrijednosti prema tom modelu, a ne rezultate provedene ankete.

Tablica 6: Dvije veličine uzorka iz iste populacije. Konstruirana nastavna vježba, izrada autora.

Veličina uzorka	Standardna devijacija u populaciji	Standardna pogreška sredine
25 osoba	20 minuta	4 minute
100 osoba	20 minuta	2 minute

Asistent je ispod usporedbe napisao objašnjenje.

> Veći nasumični uzorak daje užu distribuciju uzoračkih sredina. Standardna pogreška pada s četiri na dvije minute, pa se i vremena pojedinaca u većem uzorku međusobno razlikuju upola manje.

Uz svaki od posljednjih dvaju stupaca dopišite čiju raspršenost opisuje. Provjerite račune standardne pogreške, a zatim preoblikujte drugu rečenicu objašnjenja tako da zadrži točan podatak iz tablice.

Razrađeni primjer

Za spremnost na plaćanje sada povezujemo dvije provjere koje smo dosad promatrali odvojeno. Želimo znati gdje su procjene usredotočene i koliko su raspršene, iako izvorna varijabla ima mnogo nula i dug desni rep. Odaberemo uzorak, izračunamo sredinu i postupak ponovimo mnogo puta. Mnoge kasnije formule opisivat će ili približno izražavati rezultat takvog ponavljanja unutar određenog modela.

Jedan korak možemo pratiti na zasebnoj, namjerno sastavljenoj ilustraciji s iznosima 0, 0, 10, 20 i 70 eura. Sredina je $100/5 = 20$ eura. U ponavljanju zamjenjujemo cijeli uzorak, uključujući moguće nule, novim odabirom iz svih 50 000 osoba. Svaki novi uzorak ostavlja samo jednu sredinu u raspodjeli koju uspoređujemo.

Isti ćemo postupak provesti s dvjema veličinama uzorka. Ako veći uzorak daje užu raspodjelu sredina, provjerit ćemo smanjuje li se njezina širina onako kako predviđa odnos s korijenom veličine uzorka. Zasebno ćemo pregledati je li se promijenio i njezin oblik.

```
uzmi_sredinu <- function(n) {
  mean(sample(populacija_medija$spremnost_platiti, n))
}

sredine_5 <- replicate(4000, uzmi_sredinu(5))
sredine_60 <- replicate(4000, uzmi_sredinu(60))

c(
  populacija = mean(populacija_medija$spremnost_platiti),
  sredina_5 = mean(sredine_5),
  sredina_60 = mean(sredine_60),
  se_5 = sd(sredine_5),
  se_60 = sd(sredine_60)
)
```

```
populacija  sredina_5 sredina_60       se_5      se_60
  6,126660   6,100150   6,144362   6,185099   1,790483
```

`replicate` ponavlja izraz i sprema rezultate. Prosjek četiri tisuće sredina iz uzoraka od pet osoba je 6,10 eura, a iz uzoraka od šezdeset 6,14 eura. Populacijska sredina je 6,13 eura. Obje veličine uzorka daju procjene usredotočene blizu nje, u skladu s nepristranošću.

Standardna pogreška pada s 6,19 za pet osoba na 1,79 za šezdeset, odnosno 3,5 puta. To je približno korijen iz dvanaest, koliko smo puta povećali uzorak. Raspodjela sredina za pet osoba ipak ostaje vidljivo asimetrična. Nepristranost procjenitelja zato treba razlikovati od normalnog oblika njegove raspodjele.

Provjeru smo mogli napraviti jer poznajemo cijelu generiranu `populacija_medija`. U stvarnom istraživanju obično raspolažemo samo jednim uzorkom i ne možemo ponavljati odabir iz potpuno poznate cjeline. Zato svojstva distribucije uzorkovanja procjenjujemo iz tog uzorka, uz pretpostavke. U poglavlju o procjeni uzorak će privremeno preuzeti ulogu populacije i omogućiti novo ponavljanje postupka.

Za računsku provjeru pretpostavimo neovisna opažanja i standardnu devijaciju od 12 minuta. Pri 36 opažanja standardna pogreška sredine iznosi $12/\sqrt{36} = 2$ minute, a pri 144 opažanja jednu minutu. Raspršenost pojedinačnih odgovora ostaje 12 minuta. Četverostruko veći uzorak prepolovio je nesigurnost sredine, uz istu raspršenost odgovora pojedinačnih ljudi.

Približenje $1/\sqrt{n}$ za 95-postotnu marginu udjela odnosi se na jednostavni slučajni uzorak, malu uzoračku frakciju i najnepovoljniji udio od 0,5. Nastaje iz $1{,}96\sqrt{0{,}25/n}$ uz zaokruživanje 0,98 na jedan. Nije opća formula za svaki udio i svaki nacrt. Kalibracija težina usklađuje ponderirane zbrojeve s poznatim obilježjima populacije. Razlikuje se od kalibracije prognoze, koja uspoređuje najavljene vjerojatnosti s ishodima.

Sažetak

Ponavljanjem odabira dobili smo distribuciju uzorkovanja i razdvojili promjenjivost procjene od razlika među pojedinačnim osobama. Standardna devijacija te distribucije jest standardna pogreška, koja za sredinu u jednostavnom slučajnom uzorku pada približno s korijenom veličine uzorka, uz dodatnu korekciju kada obuhvaćamo znatan dio populacije. Kad nacrt uključuje težine, skupine i slojeve, račun mora pratiti te odluke, a njihove posljedice za preciznost opisujemo učinkom nacrta i efektivnom veličinom uzorka. Takav račun ne obuhvaća automatski probleme pokrivenosti, neodaziva ili samoodabira, koje ni veći uzorak ni kalibracija sami ne uklanjaju. I za korpus tekstova zato odvojeno obrazlažemo odabir i procijenjenu nesigurnost. U poglavlju o procjeni iz te nesigurnosti gradimo interval za nepoznatu populacijsku vrijednost.

Distribucija i raspodjela ovdje su sinonimi. Statistika kao broj iz uzorka razlikuje se od statistike kao znanstvene discipline. Ponderiranje znači primjenu težina. Neodaziv, nazvan i neodgovor, znači da za pozvane jedinice nedostaju odgovori.

Pojmovi

populacija (*population*), uzorak (*sample*), parametar (*parameter*), statistika (*statistic*), pogreška uzorkovanja (*sampling error*), distribucija uzorkovanja (*sampling distribution*), nepristranost (*unbiasedness*), standardna pogreška (*standard error*), središnji granični teorem (*central limit theorem*), uzorački udio (*sample proportion*), margina pogreške (*margin of error*), jednostavni slučajni uzorak (*simple random sample*), korekcija za konačnu populaciju (*finite-population correction*), vjerojatnosno uzorkovanje (*probability sampling*), težina uzorkovanja (*sampling weight*), procjena s težinama (*weighted estimate*), kalibracija (*calibration*), grupno uzorkovanje (*cluster sampling*), učinak nacrta (*design effect*), efektivna veličina uzorka (*effective sample size*), pokrivenost (*coverage*), neodgovor (*nonresponse*), prigodni uzorak (*convenience sample*), samoodabir (*self-selection*)

Zadaci

Konceptualni

Razlikujte raspodjelu pojedinačnih opažanja od distribucije uzoračkih sredina. Za svaku navedite što joj je jedinica, što mjeri njezina širina i što se s njom događa kada uzorak naraste. Predajte skicu obiju raspodjela i dva popratna objašnjenja.

Računski

Upotrijebite tablicu opaženoga uzorka iz sintetičke konačne populacije. Bez programa izračunajte brojnik, nazivnik i postotak prvo bez težina, a zatim s težinama uzorkovanja. Objasnite zašto se procjena pomaknula prema dolje i navedite jednu pogrešku

koju taj pomak ne može ispraviti. Predajte oba računa i dvije rečenice tumačenja.

U zasebnom jednostavnom slučajnom uzorku iz velike populacije standardna devijacija obilježja iznosi 12. Izračunajte standardnu pogrešku sredine za uzorke od 64 i 256 osoba. Koliko se puta mora povećati uzorak da se pogreška prepolovi? Za udio 0,5 u uzorku od 400 osoba procijenite 95-postotnu marginu pravilom $1/\sqrt{n}$ i navedite pretpostavke toga pravila.

Za obje veličine uzorka izračunajte i 1,96 standardnih pogrešaka te navedite približan raspon odstupanja uzoračke sredine od populacijske sredine kada vrijedi normalno približenje. Riječima protumačite dobivenu marginu udjela.

Neobvezna nadogradnja dostupna je samo čitatelju s vlastitom službeno preuzetom kopijom ESS Round 11, edition 3.0. Slijedeći portalnu putovnicu iz kataloga, ograničite podatke na Hrvatsku, za `vote` primijenite njegov službeni valjani nazivnik te usporedite procjenu bez težina s procjenom uz zadani `anweight`. Navedite i jednu pogrešku odabira ili mjerenja koju ta težina ne može ukloniti. Rezultat tumačite kao samoprijavu, ne kao provjeru službene izlaznosti. ESS mikropodaci i rezultat te provjere nisu dio knjige ni obveznoga zadatka.

Kritički

Vratite se na karticu za čitanje ankete iz poglavlja o mjerenju i dizajnu i primijenite je na slučaj časopisa *Literary Digest* (Squire 1988.). Za ciljnu populaciju, okvir, regrutiranje, broj pozvanih i odgovora, datume, formulaciju pitanja, težine, naručitelja i marginu pogreške označite što je iz prikazanoga slučaja poznato, a što nije. Zatim odvojite pogrešku pokrivenosti i neodgovor od slučajne promjenjivosti te objasnite zašto veći broj poslanih listića ne rješava prva dva problema. Ne dopunjujte nepoznata polja pretpostavkama. Predajte ispunjenu karticu i jedan odlomak zaključka.

Revizija modela

Ocijenite analizu modela iz okvira o pogrešci. Izdvojite tvrdnju koja je točna, imenujte zamjenu dviju razina varijabilnosti i napišite ispravljenu verziju druge rečenice koja zadržava sve što je u njoj bilo točno.

09

DIO III · OD UZORKA DO POPULACIJE

procjena

Preciznost procjene u rasponu

Zamislimo da otvaramo ljestvicu školskih postignuća i odmah tražimo Hrvatsku. Iznad nje nekoliko poznatih zemalja, ispod nje neke druge. Takav je poredak lako pročitati kao ocjenu škole koju smo pohađali ili u koju ćemo upisati dijete. Svako mjesto izgleda određeno.

U istraživanju PISA 2022. postignuća petnaestogodišnjaka procijenjena su iz uzoraka učenika, pa OECD uz prosječne rezultate objavljuje intervale pouzdanosti i raspone mogućih mjesta na ljestvici (OECD 2023.). Zemlje s bliskim prosjecima tako mogu zauzimati više mogućih mjesta, ovisno o nesigurnosti procjene (OECD 2023.).

Koliko zapravo znamo o nečijem mjestu na ljestvici kad znamo samo njegov procijenjeni prosjek?

U simulacijama uzorkovanja mogli smo iznova odabirati ljude iz poznate populacije i pratiti kako se mijenjaju uzoračke sredine. Istraživač obično nema tu mogućnost. Pred njim je jedan skup odgovora, a treba navesti koliko bi se njegova procjena mogla promijeniti s drugim sudionicima. Standardna pogreška iz prethodnog poglavlja daje nam polazište za takav opis nesigurnosti.

Za razumijevanje intervala poslužit će nam jednostavniji nacrt od PISA-ina. Nastavljamo s nastavnom populacijom `populacija_medija`. Uzorak od 200 osoba izvlačimo iz nje zasebno od ranijih uzoraka, za procjenu prosječnog povjerenja u medije. Svako opažanje pripada jednoj računalno stvorenoj osobi i sadrži njezinu ocjenu povjerenja na ljestvici od jedan do deset. Zapisi ne opisuju stvarne stanovnike. Pri izgradnji procjene koristit ćemo uzorak, a poznatu populacijsku vrijednost zatim iskoristiti za provjeru postupka.

Oko procjene ćemo konstruirati interval i pratiti što se događa kroz ponovljena uzorkovanja. Pitanje je koliko često tako izrađen raspon obuhvaća ciljnu vrijednost. To će nam omogućiti da pouzdanost vežemo uz postupak, a širinu intervala uz preciznost. Usporedit ćemo što se mijenja kada tražimo višu pouzdanost i što dobivamo prikupljanjem više podataka. Posebno ćemo odvojiti nesigurnost procijenjenog prosjeka od razlika među pojedinačnim odgovorima.

Za procjenu medijana dnevnog praćenja medija poslužit će nam uzorak od 60 osoba, izvučen zasebno iz iste simulirane populacije. Njegove ćemo zapise ponovno izvlačiti, dopuštajući da ista osoba bude odabrana više puta. Taj postupak, bootstrap, omogućit će nam procijeniti promjenjivost i kada ne polazimo od gotove formule za odabranu mjeru. Razmotrit ćemo koliko takva provjera ovisi o uzorku kojim raspolažemo.

Uz procjenu u izvještaju navodimo raspon njezine nesigurnosti i objašnjavamo što bi precizniji odgovor zahtijevao. Interval pritom neće ukloniti ograničenja mjerenja ili odabira ljudi koja smo upoznali ranije.

Od točke prema rasponu

Uzorak od 200 osoba sada koristimo za procjenu povjerenja. Za svaku osobu koristimo povjerenje u medije na ljestvici od jedan do deset. Prosjek uzorka iznosi 4,82. Tu vrijednost zovemo **točkasta procjena** (*point estimate*). Ona sažima rezultat, a njegovu nesigurnost opisujemo zasebno. Isti prosjek možemo dobiti na dvadeset i na dvadeset tisuća ljudi, iako bi se pri ponovnom uzorkovanju te procjene mogle znatno različito mijenjati.

Za opis te promjenjivosti služi standardna pogreška, koju smo upoznali u prethodnom poglavlju. Ona procjenjuje koliko bi se rezultat tipično mijenjao kroz ponovljena uzorkovanja. U našem uzorku iznosi 0,148.

Račun intervala možemo pratiti od zbroja ocjena do njegovih granica. Zbroj svih ocjena u uzorku iznosi 963, pa dijeljenjem s 200 dobivamo 4,815. Uzoračka standardna devijacija 2,093 podijeljena korijenom iz 200 daje standardnu pogrešku 0,148. Umnožak te pogreške s 1,96 iznosi 0,290, što oduzimamo od sredine i dodajemo joj. Dobivene granice uspoređujemo s jednom fiksnom populacijskom sredinom, 4,881.

Provjerimo što iz toga možemo zaključiti u simulaciji s poznatom populacijom. Deset tisuća puta izvukli smo po dvjesto različitih osoba bez vraćanja, svaki put iz cijele populacije. Ponovno računamo i sredinu i standardnu pogrešku, pa se obje granice mijenjaju. Oko svake uzoračke sredine označili smo raspon od 1,96 njezinih standardnih pogrešaka na svaku stranu. Populacijska vrijednost nalazi se u 94,9 % raspona. Izvan raspona ostaje u 506 od 10 000 ponavljanja. Vidimo da pojedini raspon može promašiti cilj, ali i da možemo pratiti koliko ga često postupak obuhvaća. U račun uključujemo svih 10 000 intervala. Broj onih čije granice obuhvaćaju cilj dijelimo s 10 000. Dobiveni udio nazivamo obuhvatom intervala.

Središnji granični teorem pomaže objasniti taj rezultat. Pod njegovim uvjetima raspodjela uzoračkih sredina približno je

normalna. U normalnoj raspodjeli oko 95 % vrijednosti nalazi se unutar 1,96 standardnih devijacija od središta. Za raspodjelu procjena tu udaljenost mjerimo standardnom pogreškom.

Ako je uzoračka sredina dovoljno blizu populacijskoj vrijednosti, i populacijska vrijednost jednako je blizu uzoračkoj sredini. Zato raspon od 1,96 standardnih pogrešaka na svaku stranu od procjene obuhvaća populacijsku vrijednost u približno 95 % ponavljanja, kada vrijede pretpostavke postupka.

Oznaka z^* govori koliko standardnih pogrešaka odmjeravamo na svaku stranu od procjene. Za razinu pouzdanosti od 95 % ona iznosi 1,96.

Normalne kritične vrijednosti za obostrane intervale iznose 1,645 za 90 %, 1,960 za 95 % i 2,576 za 99 %. Koristimo ih kada je raspodjela procjenitelja približno normalna i standardna pogreška dovoljno dobro poznata. Za ovdje prikazane jednostavne sredine potrebna su neovisna opažanja iz odgovarajuće populacije, konačna varijanca i uzorak bez dominantnih krajnjih vrijednosti. Pri malim uzorcima i procjeni raspršenosti iz istih podataka prikladan t-postupak upoznat ćemo uz usporedbu skupina.

$$\bar{x} \pm z^* \cdot SE_{\bar{x}}$$

U našem uzorku dobivamo raspon od 4,52 do 5,11. Poznata populacijska vrijednost iznosi 4,88 i ovaj je raspon obuhvaća.

> **Interval pouzdanosti** je raspon oko procjene, izračunat postupkom koji kroz ponovljena uzorkovanja obuhvaća nepoznati parametar u unaprijed određenom udjelu slučajeva.

Razina pouzdanosti opisuje koliko često postupak obuhvaća cilj kada se uzorkovanje ponavlja. Tu razliku trebamo zadržati i kada tumačimo samo jedan izračunati interval.

Ovdje pouzdanost opisuje postupak intervalne procjene. U poglavlju o mjerenju ista se riječ odnosila na dosljednost mjerenja. Obuhvat intervala označuje koliko intervala sadrži cilj,

a obuhvat okvira uzorkovanja govori tko je uopće mogao biti odabran.

U simetričnom intervalu udaljenost od procjene do bilo koje granice jest **margina pogreške**. Ona je polovica ukupne širine i broj uz znak ± koji smo susreli uz anketne rezultate. U ovom postupku iznosi $z^* \cdot SE$. Asimetričan interval može imati različite udaljenosti od procjene do donje i gornje granice, pa ih treba navesti odvojeno. Opisuje promjenjivost zbog uzorkovanja pod pretpostavkama postupka. Ne uključuje pristran okvir uzorkovanja, neodgovor, formulaciju pitanja ni pogrešku mjerenja. Manja margina zato sama po sebi ne znači da je anketa kvalitetnija.

Obećanje razine pouzdanosti

Za provjeru granice postupka vratimo se izrazito asimetričnoj raspodjeli spremnosti na plaćanje iz prethodnog poglavlja. Iz nje smo 10 000 puta izvukli po 20 osoba i svaki put izračunali isti interval za sredinu. Poznatu populacijsku sredinu obuhvatilo je 8 622 intervala, odnosno 86,2 %. U malom uzorku iz takve raspodjele normalno približenje i procjena standardne pogreške nisu dovoljno pouzdani da bi postupak postigao zadanu razinu od 95 %.

U prethodnoj simulaciji povjerenja, s po dvjesto osoba, dobili smo obuhvat blizak zadanoj razini, ali podudaranje nije savršeno. Dobiveni obuhvat približan je iz dvaju razloga. U konačnom broju simulacija udio pogodaka slučajno se mijenja. Osim toga, nepoznatu populacijsku raspršenost procjenjujemo iz uzorka, što donosi dodatnu nesigurnost. Pri dvjesto osoba ona je mala, a pri dvadeset može biti važna. U poglavlju o dvjema grupama uvest ćemo postupak koji zbog toga širi interval u malim uzorcima.

Obuhvat od 94,9 % ne govori koji su pojedinačni intervali pogodili cilj. Svaki izračunati interval ili sadrži populacijsku

vrijednost ili je ne sadrži. U frekventističkom postupku parametar je fiksan, a slučajan je uzorak iz kojega računamo granice. Zato razinu pouzdanosti ne tumačimo kao vjerojatnost da fiksni parametar leži unutar upravo opaženih granica.

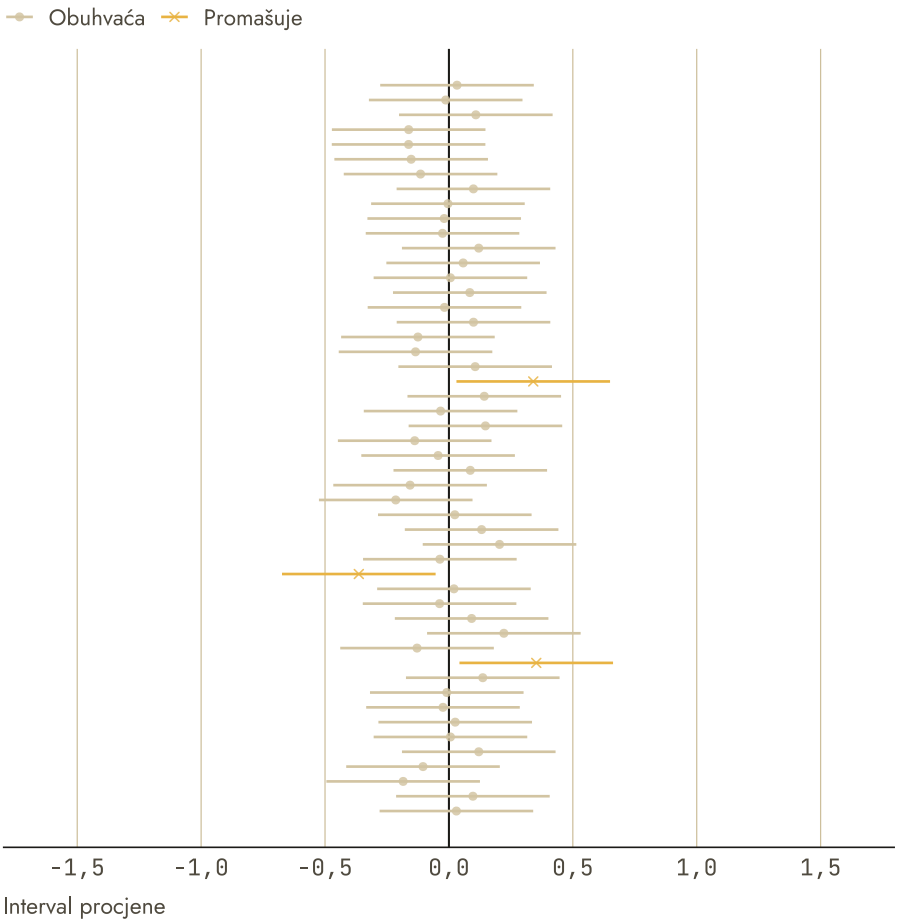
Možemo to zamisliti kao bacanje obruča na kolac. Kolac stoji na istom mjestu, a obruči padaju na različita mjesta. Razina pouzdanosti opisuje koliko često obruč obuhvati kolac kroz mnogo bacanja. Nakon pojedinoga bacanja pogodak je već određen, čak i ako ga ne možemo vidjeti.

Iz objavljenog intervala možemo čitati koje su vrijednosti uskladive s podacima pod pretpostavkama postupka. Njegova širina pokazuje i jesmo li postigli željenu preciznost, primjerice marginu manju od deset minuta. Za pitanje o razlici ili promjeni trebamo interval za samu razliku. Predviđanje rezultata sljedećeg istraživanja zaseban je zadatak jer će ono imati vlastiti uzorak i vlastiti interval.

Hvatač intervala

Za prikaz intervala prelazimo na zaseban normalni model sa sredinom nula i poznatom standardnom devijacijom jedan. Pojedinačno opažanje sada je umjetno generirani broj, bez jedinice povjerenja. Iz svakog skupa od 40 neovisnih brojeva računamo sredinu i oko nje raspon $1{,}96/\sqrt{40}$ na svaku stranu. Tiskani prikaz ponavlja to 50 puta. Ovdje je standardna pogreška poznata iz modela, za razliku od prethodnog računa u kojem smo je procjenjivali iz uzorka.

Na slici su intervali iz ponovljenih uzoraka poredani jedan ispod drugoga. Okomita crta označuje populacijsku vrijednost. Intervali koji je sijeku obuhvaćaju cilj. U tablici zatim uspoređujemo veličinu uzorka i razinu pouzdanosti. Obje utječu na širinu, a razina pouzdanosti određuje i dugoročnu učestalost obuhvata.



Slika 1: Pedeset intervala iz ponovljenih simuliranih uzoraka i fiksna populacijska sredina. Križići označuju intervale koji promašuju cilj. Izrada autora. Podaci i postupak opisani su u tekstu.

Tiskane postavke za vježbu

Tri prikazane postavke koriste normalnu populaciju sa sredinom nula i standardnom devijacijom jedan. Svaka sadrži pedeset intervala iz istoga zadanog simulacijskog niza. A i B razlikuju se samo po razini pouzdanosti, a A i C samo po veličini uzorka.

Tablica 1: Tri fiksne postavke za usporedbu širine i promašaja intervala u tiskanom izdanju. Izrada autora.

Mjera	A	B	C
Veličina uzorka	40	40	160
Razina pouzdanosti	95 %	99 %	95 %
Kritična vrijednost z*	1,960	2,576	1,960
Broj intervala	50	50	50
Širina intervala	0,620	0,815	0,310
Broj promašaja	3	0	1

Usporedite A i B, a zatim A i C. Tako možete odvojeno vidjeti učinak razine pouzdanosti i veličine uzorka. Unutar svakoga od 50 ponavljanja A i B koriste istih prvih 40 opažanja, a C svih 160 opažanja toga niza. Ponavljanja su međusobno različita.

Za zasebnu računsku provjeru na kraju poglavlja trebat će nam i sažetak cijele poznate sintetičke populacije po glavnom izvoru vijesti. Sljedeća tablica čuva točne cjelobrojne brojnosti i zbroj za portal, dok su udio i prosjek zaokruženi na pet decimala. Iz tih ćemo brojeva provjeriti dva dijeljenja. Provjera ne potvrđuje kako su pojedinačni zapisi zbrojeni. Taj je dodatni korak moguć u dobrovoljnoj mrežnoj vježbi. Podaci služe vježbi i ne opisuju stvarne stanovnike.

Tablica 2: Brojnici, nazivnici i agregatne vrijednosti za portal u sintetičkoj populaciji. Izrada autora prema javnom agregatnom prikazu sintetičkih podataka.

Mjera	Vrijednost
Izvor vijesti	portal
Broj osoba	15 101
Ukupno osoba	50 000
Udio	0,30202
Zbroj povjerenja	72 101
Prosjek povjerenja	4,77458

Preciznost nasuprot pouzdanosti

Za sljedeću usporedbu širina vraćamo se ocjenama povjerenja iz medijske populacije. Za svaku veličinu 50, 200 i 800 izvlačimo 400 novih uzoraka, računamo njihove intervale istim postupkom sa standardnom pogreškom iz uzorka i prosječimo 400 širina. Usporedba razina 90 %, 95 % i 99 % potom zadržava jedan isti početni uzorak od 200 osoba i mijenja samo množitelj.

Dvije usporedbe iz prikaza sada možemo razdvojiti i brojčano. Veći uzorak smanjuje standardnu pogrešku i sužava interval pri istoj razini pouzdanosti. Prosječna širina intervala pri pedeset osoba iznosi 1,09 boda, pri dvjesto 0,55, a pri osamsto 0,28. S više podataka dobivamo precizniju procjenu.

Povećanje razine pouzdanosti djeluje drukčije. Na istim podacima tražimo češći obuhvat, pa raspon mora biti širi. U našem uzorku interval od 90 % širok je 0,49 boda, interval od 95 % 0,58, a interval od 99 % 0,76. Viši zahtjev za obuhvat ovdje plaćamo širim rasponom, odnosno manjom preciznošću.

Preciznost opisuje širinu procjene, a **pouzdanost** odabranu dugoročnu učestalost obuhvata. Za isti postupak i razinu pouzdanosti više podataka obično daje uži interval. Za iste podatke viša

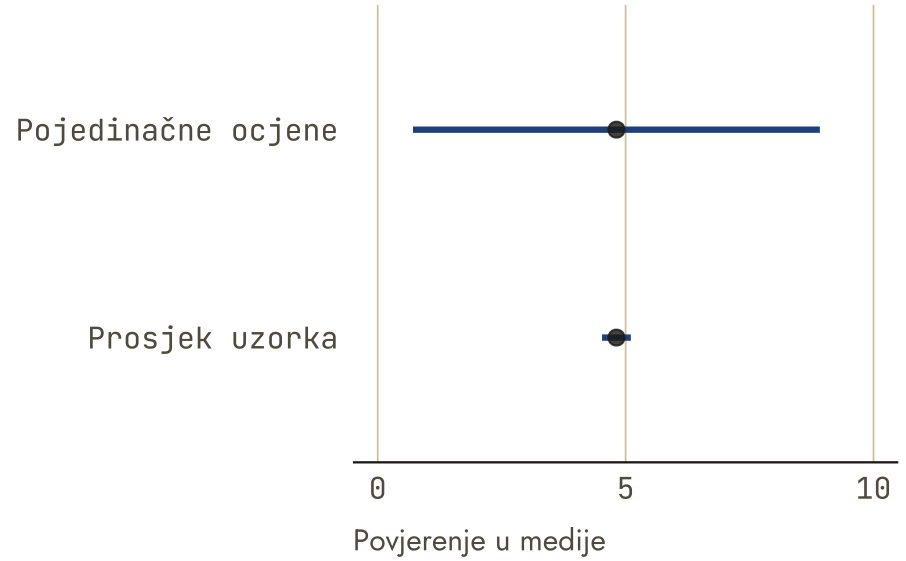
razina pouzdanosti daje širi interval. U knjizi najčešće koristimo 95 %, no to je uobičajen izbor koji treba navesti uz rezultat.

Širok interval može biti primjeren opis maloga uzorka ili jako raspršenih podataka. Uzak interval također treba provjeriti. Ako je uzorak pristran, interval može biti uzak oko pogrešne vrijednosti jer ne obuhvaća pristranost postupka odabira.

Na grafu s više skupina intervali pokazuju koliko su njihove procjene precizne. Širina ovisi i o broju opažanja i o njihovoj raspršenosti. Takav graf govori više od prikaza samih prosjeka, ali za tumačenje razlika među skupinama treba još jedan korak.

Nepreklapanje usporedivih intervala može upućivati na razliku, ali preklapanje ne dokazuje jednakost. Tumačenje ovisi i o nacrtu istraživanja te o tome jesu li procjene povezane. Najizravnije je izračunati interval za samu razliku. To ćemo učiniti u poglavlju o dvjema grupama. Ni iz takve usporedbe ne dobivamo vjerojatnost da je razlika „stvarna”.

Treba razlikovati i nesigurnost prosjeka od raspršenosti pojedinačnih vrijednosti. Interval za sredinu u uzorku od 200 osoba širok je 0,58 boda. Ako su ocjene približno normalno raspoređene, raspon sredine plus ili minus 1,96 uzoračkih standardnih devijacija opisivao bi oko 95 % pojedinačnih ocjena. Taj je raspon širok približno 8,2 boda. Normalno pravilo ne daje takav obuhvat za izrazito asimetrične, ograničene ili diskretne raspodjele. Taj opisni raspon nije ni interval predviđanja za novu osobu, jer ne uključuje nesigurnost procjene središta i raspršenosti niti određuje populaciju buduće osobe. Veći uzorak sužava interval za sredinu, dok razlike među ljudima ostaju svojstvo populacije.



Slika 2: Interval pouzdanosti za sredinu i opisni raspon po normalnom pravilu za pojedinačne ocjene na istom uzorku i istoj osi. Izrada autora. Podaci i postupak opisani su u tekstu.

U PISA-inoj tablici prirodoslovne pismenosti za Hrvatsku je objavljena procjena od 483 boda, uz 95-postotni interval od 478 do 487 bodova i raspon mjesta od osamnaestoga do trideset petoga među sudionicama istraživanja (OECD 2023.). Taj raspon mjesta uzima u obzir nesigurnost procjena i višestruke usporedbe zemalja sličnih rezultata (OECD 2023.). Iz njega možemo vidjeti zašto jedno precizno ispisano mjesto prenosi više sigurnosti nego što je imamo. Raspon se odnosi na položaj procijenjenoga nacionalnog prosjeka. Za iskustvo pojedinoga učenika trebali bismo pogledati i razlike unutar zemlje.

Kad bismo pripremali naslov o tim rezultatima, uz procjenu bismo zadržali njezinu nesigurnost i imenovali područje koje je ispitano. PISA pritom koristi složen uzorak škola i učenika te postupke procjene postignuća koje račun nesigurnosti mora uvažiti (OECD 2023.). Ovdje čitamo objavljene rezultate. Jednostavno ponovno izvlačenje pojedinačnih redaka, koje ćemo sada upoznati na našim simuliranim podacima, ne bi sačuvalo taj istraživački nacrt.

Bootstrap kao vlastiti izum

Za sredinu imamo jednostavnu formulu standardne pogreške. Za medijan, razliku percentila ili omjer račun može biti složeniji ili zahtijevati dodatne pretpostavke. Tada nam može pomoći ponovno uzorkovanje opaženih podataka.

Kad bismo mogli uzimati nove uzorke iz populacije, izračunali bismo medijan u svakome i promatrali koliko se mijenja. Budući da imamo samo jedan uzorak, pokušavamo ponoviti taj postupak izvlačenjem iz njegovih vrijednosti. Uzorak pritom mora razumno predstavljati ciljnu populaciju. Jednostavno ponovno uzorkovanje redaka pretpostavlja neovisne i međusobno zamjenjive jedinice na razini uzorkovanja. Ako su podaci upareni, grupirani ili ponovljeni, čuvamo tu strukturu i izvlačimo cijele parove, skupine ili osobe. Pod tim uvjetima novi uzorci iste veličine približno oponašaju promjene koje bi nastale ponovnim uzorkovanjem iz populacije.

Izvlačimo **s vraćanjem**. Bez vraćanja bismo iz uzorka od šezdeset osoba dobili istih šezdeset osoba u drugom redoslijedu, pa bi medijan uvijek bio jednak. S vraćanjem neke osobe ulaze više puta, a neke izostanu. Zbog promjene sastava mijenja se i procjena. Taj postupak zovemo **bootstrap**. Uveden je za procjenu nesigurnosti kada teorijski račun nije lako dostupan (Efron 1979.).

Percentilni bootstrap interval uzima odabrane percentile ponovljenih procjena. Njegova je primjena najjednostavnija kada bootstrap raspodjela dobro oponaša raspršenost procjenitelja, bez izražene pristranosti ili asimetrije. Postoje i drugi bootstrap intervali koji korigiraju pomak ili asimetriju. Ovdje učimo samo najjednostavniji percentilni postupak. Ni taj postupak ne jamči točno zadanu razinu obuhvata. Mali ili diskretni uzorci, pristranost procjenitelja i vrlo asimetrične raspodjele mogu oslabiti približenje. Ponavljanje odabira iz postojećih redaka ne dodaje ljude koji su sustavno izostali iz početnog uzorka.

Sada uzimamo novi početni uzorak od 60 osoba iz iste medijske populacije. Izvlačimo ga zasebno od prethodnih 200 osoba, a pratimo dnevne minute umjesto povjerenja. U lijevoj polovici tablice vidimo prvih osam zapisa toga uzorka, a u desnoj prvih osam izvlačenja iz prvoga bootstrap uzorka. Broj retka desno upućuje na zapis u cijelom početnom uzorku od 60 osoba. Svako ponavljanje broja retka upućuje na ponovljen odabir iste osobe.

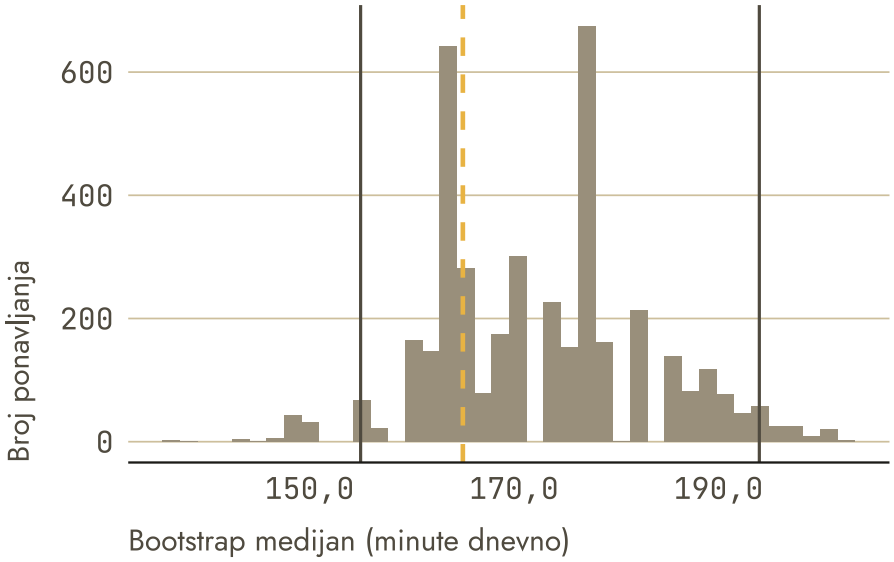
Tablica 3: Početak izvornog uzorka i prvoga ponovnog uzorka. Svaki puni uzorak ima 60 zapisa. Izrada autora.

Izvorni redak	Minute	Izvučeni redak	Minute
1	83	19	161
2	82	5	164
3	134	29	106
4	139	28	116
5	164	60	342
6	133	51	264
7	256	40	221
8	288	48	194

Prvo ponovno izvlačenje uzima redak 19, pa prenosi njegovih 161 minuta. Nakon svih 60 izvlačenja imamo 39 različitih osoba, a medijan cijeloga ponovnog uzorka iznosi 176,0 minuta. Taj jedan broj ulazi u sljedeći histogram. Postupak 4 000 puta počinje od istoga početnog uzorka, svaki put s novim izvlačenjem s vraćanjem. Sva su ponavljanja definirana. Poredamo 4 000 medijana i uzmemo percentile 2,5 i 97,5, ostavljajući približno 100 vrijednosti sa svake strane.

U uzorku od 60 osoba medijan dnevnog praćenja medija iznosi 173,5 minuta. Iz četiri tisuće bootstrap uzoraka dobivamo središnji percentilni raspon od 155,0 do 194,0 minuta. On obuhvaća poznati populacijski medijan od 165,0 minuta. Na slici vidimo

kako nastaje raspon iz jednoga uzorka. Za provjeru dugoročnog obuhvata trebali bismo ponavljati cijeli postupak s novim početnim uzorcima.



Slika 3: Raspodjela četiri tisuće bootstrap medijana iz jednog uzorka, s granicama središnjih 95 % vrijednosti i pravom populacijskom vrijednošću. Izrada autora. Podaci i postupak opisani su u tekstu.

Histogram je stubast jer medijan maloga, zaokruženog uzorka može poprimiti samo ograničen broj vrijednosti. Zato su i granice percentilnog raspona grube. Više bootstrap ponavljanja smanjuje računalno kolebanje percentila, ali ne nadoknađuje vrijednosti koje nedostaju u početnom uzorku.

Bootstrap ne zahtijeva da unaprijed imenujemo oblik populacijske raspodjele, no način izvlačenja mora odgovarati statistici i nacrtu istraživanja. Za korelaciju izvlačimo cijele parove vrijednosti. Za razliku neovisnih skupina uzorkujemo svaku skupinu zasebno. Kod ovisnih podataka izvlačimo cijele neovisne jedinice. Postupak ne može dodati rijetke ili krajnje vrijednosti kojih u uzorku nema. Zato su vrlo visoki percentili posebno osjetljivi na male uzorke i teške repove. Medijan manje ovisi o repovima, ali mnogo jednakih vrijednosti može dati grube granice.

Bootstrap također ne uklanja nereprezentativnost početnog uzorka.

Kodirani udio i dvije nesigurnosti

U poglavlju o povezanosti promjena pravila kodiranja promijenila je oznake istih naslova. Ponovno izvlačenje tih naslova ne bi samo po sebi provjerilo to pravilo. Kod udjela dobivenog kodiranjem teksta zato razlikujemo barem dva izvora nesigurnosti. **Uzoračka nesigurnost** odnosi se na odabir tekstova. **Nesigurnost kodiranja** odnosi se na pravila razvrstavanja, rubne slučajeve i odluke kodera ili klasifikatora. Ako kodirana kategorija samo približno odgovara istraživačkom pojmu, trebamo razmotriti i nesigurnost mjerenja.

Bootstrap redaka s gotovim oznakama mijenja sastav tekstova, ali zadržava njihove oznake. Uz ispravno odabranu jedinicu ponovnog uzorkovanja opisuje uzoračku nesigurnost. Valjanost kodiranja zahtijeva zasebnu provjeru oznaka i pravila. Drugo opravdano pravilo moglo bi promijeniti oznake rubnih tekstova i pomaknuti procijenjeni udio, čak i kada oba izračuna daju uske intervale. U izvještaju zato navodimo jedinicu teksta, ciljnu populaciju, način odabira i pravilo kodiranja te objašnjavamo koju nesigurnost interval uključuje.

● STATISTIKA U DIVLJINI

Šest tvrdnji o jednom intervalu. Istraživač može uredno izračunati interval, a ipak pogrešno objasniti što iz njega slijedi. U istraživanju razumijevanja intervala studenti i istraživači prosuđivali su šest tumačenja rezultata iz zamišljenog istraživanja (Hoekstra i sur. 2014.). Sve su tvrdnje bile netočne, ali su ih obje skupine u prosjeku prihvatile više od tri (Hoekstra i sur. 2014.). Pogreška se dogodila pri čitanju rezultata, nakon što je račun već bio zadan.

Razmotrimo vlastiti zamišljeni primjer. Studentska anketa treba procijeniti prosječno vrijeme putovanja na fakultet. U izvještaju stoji 95-postotni

interval od 30 do 40 minuta. Autor bi mogao napisati da se unutar njega nalazi vrijeme putovanja 95 % studenata. Time bi raspon nesigurnosti procijenjenog prosjeka pretvorio u opis razlika među ljudima. Podaci o pojedinačnim putovanjima mogli bi biti mnogo raspršeniji.

Drugo bi iskušenje bilo napisati da je nakon ovog računa vjerojatnost položaja prave sredine između granica jednaka 95 %. U frekvencijskom postupku koji smo učili razina pouzdanosti pripada postupku ponovljenog uzorkovanja. Ako njegove pretpostavke vrijede, približno taj udio tako dobivenih intervala obuhvaćao bi pravu sredinu. Ne raspoređujemo vjerojatnost nepoznate sredine unutar ovoga već izračunatog intervala.

Za čitatelja studentskog izvještaja ta razlika ima praktičnu posljedicu. Interval prosjeka može pomoći da procijeni tipično opterećenje putovanjem, ali ne govori koliko studenata putuje izrazito dugo. Za odluku o pomoći tim studentima trebala bi raspodjela pojedinačnih vremena i prag koji za tu odluku ima smisla.

Autor zato treba imenovati veličinu koju procjenjuje i objasniti koju nesigurnost interval opisuje. U našem primjeru to je nesigurnost procjene prosječnog vremena pod uvjetima odabranog postupka. Prije pisanja zaključka zato razdvajamo nesigurnost prosjeka od razlika među pojedinačnim putovanjima. Razinu pouzdanosti pritom vežemo uz dugoročni obuhvat postupka, uz fiksnu nepoznatu populacijsku sredinu.

NAĐITE GREŠKU

Vraćamo se uzorku od šezdeset osoba iz simuliranoga grada. Želimo procijeniti populacijski medijan dnevnog vremena uz medije. Vektor `minute` sadrži njihove pojedinačne vrijednosti, a osobe su uzorkovane neovisno. Asistent je vratio račun percentilnog bootstrap intervala.

```
B <- 4000
boot_medijan <- replicate(B, {
  median(sample(minute, size = length(minute), replace = TRUE))
})
quantile(boot_medijan, c(0.025, 0.975))
```

Izračunani medijan uzorka iznosi 173,5 minuta, a granice intervala 155 i 194 minuta. To su rezultati nastavne simulacije. Uz njih stoji asistentovo objašnjenje.

> U svakom ponavljanju uzorkujemo s vraćanjem isti broj osoba i računamo medijan. Granice izdvajaju središnjih 95 % bootstrap medijana. Zato postoji 95 % vjerojatnosti da je fiksni populacijski medijan između upravo dobivenih granica.

Napišite zamjensko objašnjenje za čitatelja izvještaja. Zadržite procjenu, granice i opis ponavljanja, a razinu od 95 % povežite s onim čemu ona pripada. Navedite što bi se mijenjalo kad bismo uzeli novi uzorak iz iste populacije.

Razrađeni primjer

Zamislimo naručitelja koji želi procijeniti koliko vremena dnevno odrasli stanovnici grada provode uz medije i koji traži marginu ne veću od deset minuta. Za vježbu tog zadatka koristimo već uvedeni uzorak od 60 osoba iz simuliranoga grada.

Trebamo povezati izbor mjere s njegovim pitanjem, a zatim provjeriti daje li uzorak traženu preciznost.

Najprije biramo mjeru koja odgovara pitanju. Dnevne minute imaju dug rep prema velikim vrijednostima, koji povlači prosjek prema gore. Za opis uobičajene osobe zato biramo medijan. Budući da za njegovu standardnu pogrešku nemamo jednako jednostavan račun kao za sredinu, nesigurnost ćemo procijeniti bootstrapom.

Postupak je sličan simulaciji iz prethodnog poglavlja. Ovaj put izvlačimo iz opaženog uzorka, s vraćanjem, i u svakom ponavljanju računamo medijan.

```
set.seed(505)
uzorak <- slice_sample(populacija_medija, n = 60)

boot_medijan <- replicate(
  4000,
  median(sample(uzorak$minute_medija, replace = TRUE))
)

c(
  procjena = median(uzorak$minute_medija),
  donja = quantile(boot_medijan, 0.025),
  gornja = quantile(boot_medijan, 0.975)
)
```

```
    procjena   donja.2,5% gornja.97,5%
       173,5        155,0        194,0
```

Poziv `sample` uz argument `replace` mijenja sastav ponovnog uzorka tako što dopušta višekratni izbor iste osobe. Nakon računanja medijana u svakom ponavljanju, funkcija `quantile` određuje granice odsijecanjem zadanog udjela raspodjele sa svake strane. Tako kod povezuje ponovno izvlačenje s trima već prikazanim vrijednostima, medijanom uzorka i granicama njegova bootstrap raspona.

Za odrasle stanovnike simuliranoga grada procjenjujemo medijan dnevnog praćenja medija na 173,5 minuta. Jedinica je jedna osoba. Percentilni bootstrap raspon od 95 % ide od 155,0 do 194,0 minuta i pod navedenim uvjetima opisuje uzoračku promjenjivost medijana. Ukupna širina iznosi 39 minuta. Udaljenost od procjene do donje granice iznosi 18,5, a do gornje 20,5 minuta. Veća od tih udaljenosti prelazi traženih deset minuta. Trebamo izvijestiti o toj nepreciznosti ili prikupiti dodatne informacije. Iz raspona ne zaključujemo o promjeni. Kada uzorak ne bi predstavljao ciljnu populaciju, rezultat bismo ograničili na opis uzorka.

U simulaciji znamo da populacijski medijan iznosi 165,0 minuta i da je unutar dobivenih granica. Jedan pogodak ipak ne provjerava pouzdanost postupka. Ranijih deset tisuća ponavljanja odnosilo se na normalni interval za sredinu. Za provjeru bootstrap raspona medijana morali bismo svaki put izvući novi uzorak od 60 osoba, iz njega izračunati novi bootstrap raspon i prebrojiti obuhvate populacijskog medijana. To ovdje nismo učinili, pa prikaz ne dokazuje zadanu razinu obuhvata.

U izvještaju najprije navodimo procjenu, njezin raspon nesigurnosti i ograničenja. Ovdje trebamo reći da su podaci simulirani, da procjenjujemo medijan i da raspolažemo sa šezdeset osoba. Isti oblik izvještaja koristimo u stvarnom istraživanju, iako ondje obično ne znamo parametar s kojim bismo usporedili rezultat.

Od procjene do tvrdnje

Procjenu i širinu raspona već smo izračunali, ali naručitelju trebamo objasniti i na koga se rezultat odnosi. Šest pitanja u tablici povezuje naš račun s odabirom osoba, mjerenjem minuta i zahtjevom za preciznošću. Tako vidimo koje bi ograničenje moglo promijeniti konačnu rečenicu.

Tablica 4: Šest revizijskih pitanja primijenjenih na procjenu iz simulirane populacije. Izrada autora.

Revizijsko pitanje	Primjena na bootstrap medijana
Koja je jedinica?	jedna zamišljena odrasla osoba s jednim zapisom dnevnih minuta
Tko ili što nedostaje?	uzorak je slučajno izvučen iz poznate sintetičke populacije, pa ne predstavlja stvarni grad ni osobe izvan te populacije
Koji je cilj tvrdnje?	populacijski medijan dnevnih minuta, opisan procjenom i rasponom uzoračke nesigurnosti
Gdje je neizvjesnost?	promjene sastava uzorka. Ne provjerava dugoročni obuhvat, mjerenje minuta ni prijenos na stvarne stanovnike
Koje drugo objašnjenje ostaje?	u stvarnoj anketi slab obuhvat, neodgovor ili pogrešno mjerenje mogli bi objasniti pomak procjene. Bootstrap takve izvore pogreške ne uklanja
Tko snosi posljedice?	ovdje nema stvarnih sudionika. U stvarnom projektu pretjerana preciznost mogla bi zavesti naručitelja i stanovnike

Ti odgovori ograničavaju zaključak na ono što smo doista procijenili. Sintetički podaci služe provjeri postupka u zadanom modelu. Uzorački interval opisuje promjenjivost zbog odabira, dok nesigurnost mjerenja ili kodiranja treba zasebno procije-

niti. Karta tvrdnji u nastavku pokazuje šest različitih dosega analize.

Tablica 5: Šest dimenzija tvrdnje na granici Dijela III. Izrada autora.

Dimenzija tvrdnje	Što ovaj dokaz dopušta
opis	opisuje medijan i raspodjelu minuta u opaženom sintetičkom uzorku
povezanost	nije poduprta jer primjer ne uspoređuje dvije varijable
generalizacija	usmjerena je na poznatu sintetičku populaciju pod postupkom slučajnoga odabira, bez prava prijenosa na stvarni grad
predviđanje	nije poduprto jer nije izgrađeno ni provjereno pravilo za nova opažanja
uzročnost	nije poduprta jer nema intervencije ni usporedbe mogućih ishoda
odluka	poduprta je odluka da raspon ne ispunjava unaprijed zadanu marginu od najviše deset minuta, ali ne tvrdnja o promjeni ni odluka o medijskoj politici

Na kraju ovoga dijela provjerite možete li odgovoriti na četiri pitanja. Zašto među pedeset simuliranih intervala ne mora biti točno pet posto promašaja? Zašto bootstrap gotovih oznaka ne uključuje nesigurnost kodiranja? Što trebamo navesti uz procjenu i zašto najprije imenujemo populaciju? Zašto povećanje prigodnog uzorka ne uklanja razliku između ljudi koji su u njega mogli ući i ciljne populacije?

Tablica provjere pokazuje što smo tražili od asistenta, što je vratio i kako smo provjerili rezultat. Možemo je pročitati i bez pisanja koda.

Tablica 6: Čitljiv račun provjere za procjenu iz Dijela III. Izrada autora.

Polje računa	Bootstrap medijana
Što je traženo	procijeniti populacijski medijan dnevnih minuta i njegovu uzoračku nesigurnost iz uzorka neovisnih osoba
Što je vraćeno	točkasta procjena te donja i gornja granica središnjega percentilnog raspona
Što je provjereno	puna veličina ponovnog uzorka, izvlačenje s vraćanjem, ista statistika u svakom ponavljanju i redoslijed izvještavanja
Kako je provjereno	čitanjem poziva `sample`, `median`, `replicate` i `quantile` te usporedbom ispisa s brojkama u prozi
Uloga AI-ja	instrument i pogrešiv analitičar
Što je ostalo neprovjereno	dugoročni obuhvat percentilnoga postupka za medijan, reprezentativnost u stvarnom istraživanju i valjanost mjerenja minuta
Odgovorna osoba	osoba koja odabire procjenjivanu količinu, provjerava račun i potpisuje zaključak

Sažetak

Točkasta procjena postaje informativnija kada uz nju prikažemo interval i objasnimo koju nesigurnost on obuhvaća. Razina pouzdanosti opisuje dugoročni obuhvat postupka, pa viša razina na istim podacima širi interval, dok ga veći uzorak obično sužava. Kada nemamo jednostavnu formulu, bootstrapom ponovno izvlačimo iz opaženoga uzorka i procjenjujemo uzoračku promjenjivost, čuvajući jedinicu i strukturu nacrta. Time ne popravljamo premalen ili pristran uzorak, a pri ponovnom izvlačenju gotovih oznaka teksta ne uključujemo nesigurnost kodiranja ili mjerenja. U izvještaju zato povezujemo ciljnu populaciju i jedinicu s procjenom, mjernom jedinicom, intervalom i ograničenjima. U poglavlju o testiranju toj procjeni pridružujemo drugo pitanje, koliko bi opaženi rezultat bio neobičan pod određenom pretpostavkom.

Pojmovi

točkasta procjena (*point estimate*), kritična vrijednost (*critical value*), obuhvat intervala (*coverage*),

procjena (*estimate*), interval pouzdanosti (*confidence interval*), razina pouzdanosti (*confidence level*), margina pogreške (*margin of error*), preciznost (*precision*), bootstrap (*bootstrap*), percentilni raspon (*percentile interval*), jedinica ponovnog uzorkovanja (*resampling unit*), nesigurnost kodiranja (*coding uncertainty*)

Zadaci

Za odabrane zadatke ovoga poglavlja na kraju knjige dostupne su naznake i odvojena objašnjenja. Najprije pokušajte samostalno odgovoriti.

Konceptualni

Objasnite zašto razina pouzdanosti pripada postupku, a ne nepoznatom parametru nakon izračuna intervala. U objašnjenju navedite što je u postupku slučajno, a što fiksno, zašto ta podjela isključuje rečenicu o vjerojatnosti da parametar leži unutar zadanih granica te zašto bootstrap gotovih oznaka teksta ne obuhvaća nesigurnost pravila kodiranja. Predajte jedan odlomak.

Računski

Najprije razlikujte točkastu procjenu, standardnu pogrešku i marginu pogreške. U ranijem uzorku od 200 osoba prosjek povjerenja iznosi 4,815, a standardna devijacija zaokruženo 2,093 boda. Iz tih vrijednosti izračunajte standardnu pogrešku, marginu $1{,}96 \cdot SE$ i obje granice približnoga 95-postotnog intervala. Jednom rečenicom protumačite čemu pripada razina od 95 %.

Za ručnu demonstraciju bootstrapa uzmite prvih pet već prikazanih vrijednosti dnevnih minuta, 83, 82, 134, 139 i 164. Zadajemo ponovno izvlačenje redaka 1, 1, 2, 4 i 5. Zapišite dobivene vrijednosti i usporedite njihov medijan s medijanom početnih pet vrijednosti. Objasnite zašto jedno takvo izvlačenje još ne daje interval pouzdanosti.

U tiskanoj tablici s trima postavkama najprije promotrite samo veličinu uzorka i razinu pouzdanosti. Predvidite poredak širina i dugoročni obuhvat. Izračunajte širine kao $2z^*/\sqrt{n}$ i usporedite ih s tablicom. Očitajte promašaje i objasnite zašto njihov manji broj u C nego u A ne znači višu razinu pouzdanosti. Iz agregatne

tablice podijelite broj korisnika portala ukupnim brojem osoba, a zbroj njihova povjerenja brojem korisnika portala. Predajte predviđanje, usporedbu s rezultatima, tri računa širine i dva dijeljenja uspoređena s objavljenim vrijednostima. Objasnite zašto iz agregata možemo provjeriti račun, ali ne potvrditi kako su pojedinačni zapisi zbrojeni.

Za dobrovoljnu provjeru pojedinačnih zapisa mrežni dodaci A i B pokazuju kako otvoriti data/populacija-medija.csv[1] i usporediti isti podskup s agregatom. Ta provjera nije uvjet tiskanoga zadatka.

Kritički

Vratite se prikazu Istraživač margine pogreške u poglavlju o tome kako brojke zavode. Usporedite stanje s uzorkom od 1000 osoba bez pristranosti i stanje s jednakim uzorkom te pristranošću od šest postotnih bodova. Oba pokazuju procjenu od 52 %. Primijenite karticu za čitanje ankete iz poglavlja o mjerenju i sadašnje znanje o intervalima. Objasnite zašto ista uska margina pogreške ne može obuhvatiti pristranost odabira, navedite koji je izvor nesigurnosti unutar, a koji izvan intervala te napišite ispravljenu tvrdnju za pristrano stanje. Predajte usporedbu dvaju stanja i ispravljenu tvrdnju.

Revizija modela

Pročitajte račun iz okvira o pogrešci. Za svaki od poziva `replicate`, `sample`, `median` i `quantile` jednom rečenicom povežite redak koda s pojmom koji provodi. Zatim izdvojite jedinu pogrešnu rečenicu i napišite frekventistički ispravnu zamjenu koja zadržava razinu od 95 %, procjenu i granice intervala. Kod ne treba prepisivati ni mijenjati.

[1] https://lusiki.github.io/statistika-knjiga/data/populacija-medija.csv



DIO IV

zaključivanje

Procjeni i intervalu pridružujemo usporedbu s nultim modelom. Proučavamo pogreške odluke, veličinu učinka i planiranje istraživanja, a zatim provjeravamo kako izbor analiza i objavljivanje oblikuju dokaze.

10

DIO IV · ZAKLJUČIVANJE

logika testiranja

Što p-vrijednost govori — i što ne govori

Šaljete prijavu za prvi posao i čekate odgovor. Ako poziv izostane, teško možete znati je li presudilo iskustvo, obrazovanje ili nešto što poslodavac zaključuje već iz vašeg imena. Vlastita prijava pokazuje samo jedan ishod. Ne možete vidjeti što bi se dogodilo s drukčijim imenom na usporedivu životopisu.

U terenskom pokusu u Bostonu i Chicagu istraživači su na oglase za posao slali izmišljene životopise. Nasumično su im dodjeljivali imena odabrana da zvuče bjelački ili afroamerički, a zatim bilježili pozive na razgovor (Bertrand i Mullainathan 2004.). Tako su mogli usporediti reakcije poslodavaca na nasumično dodijeljen signal percipirane rasne pripadnosti (Bertrand i Mullainathan 2004.).

Ako jednoj skupini prijava stigne više poziva, koliko bi takva razlika bila uvjerljiv dokaz da je ime utjecalo na odgovor?

Procjena s intervalom već nam govori koliku smo razliku našli i koliko smo je precizno odredili. U istraživačkim izvještajima uz nju često stoji i p-vrijednost. Da bismo taj broj mogli pročitati, trebamo promijeniti pitanje. Umjesto traženja raspona za nepoznatu veličinu, pitat ćemo kakve bi rezultate proizvodio određeni model u kojem nema povezanosti koju istražujemo.

Pitanje iz pokusa odnosi se na razliku udjela poziva. Logiku usporedbe s nultim modelom najprije ćemo pratiti na poznatim simuliranim podacima, gdje možemo vidjeti svaki korak premještanja oznaka. Pri procjeni povjerenja promatrali smo populaciju kao cjelinu. Sada nas zanima koliko se prosječno povjerenje razlikuje između čitatelja tiska i portala te koliko je ta procjena nesigurna.

Uzorak od 300 čitatelja tiska i portala izvlačimo zasebno iz nastavne populacije `populacija_medija`, nakon ograničavanja na ta dva izvora. Svaki redak predstavlja zamišljenu osobu s izvorom vijesti i povjerenjem na ljestvici od jedan do deset, bez tvrdnje o stvarnim stanovnicima. Najprije ćemo prikazati razliku i njezin interval. Potom ćemo pretpostaviti jednaku raspodjelu povjerenja među skupinama i provjeriti što bismo očekivali kada oznake izvora ne bi bile povezane s odgovorima.

I u ovoj usporedbi ponavljamo postupak, ali drukčije nego pri uzorkovanju ili bootstrapu. Ocjene ostaju zabilježene, a oznake skupina nasumično premještamo među osobama. Svaki takav raspored daje novu razliku. Iz njihove raspodjele možemo procijeniti koliko su česta odstupanja barem toliko velika kao opaženo, uz pretpostavke koje dopuštaju premještanje. Tek nakon te provjere imenovat ćemo nultu hipotezu, testnu statistiku i p-vrijednost.

Odluka na temelju testa može biti pogrešna na dva načina. Možemo prijaviti razliku koje nema ili propustiti razliku koja postoji. Prag odluke utječe na obje mogućnosti, a neodbacivanje hipoteze ne potvrđuje njezinu istinitost. Pitanje o vjerojatnosti hipoteze zahtijeva drugi račun i dodatne pretpostavke, kakve ćemo upoznati u kratkoj usporedbi s Bayesovim pristupom. Pri

čitanju nalaza zato odvajamo ono što test kaže o podacima od procjene veličine razlike, njezine važnosti i mogućeg uzroka.

Sudnica i njezina asimetrija

Prije testa trebamo prikazati veličinu opažene razlike i njezinu nesigurnost. Trebamo razmotriti i posljedice pogrešne odluke. Hoće li veći problem biti prijaviti razliku koje nema ili propustiti onu koja postoji? Ta pitanja određuju kako ćemo koristiti rezultat.

Usporedba podataka s početnom pretpostavkom ima sličnost sa suđenjem. Dokazi se procjenjuju u odnosu na pretpostavku nevinosti. Ako optužba nije dokazana, time ne mora biti dokazana nevinost. Slično vrijedi za test. Kada nemamo dovoljno dokaza za odbacivanje početne hipoteze, još nismo pokazali da je ona istinita.

Usporedba sa suđenjem ima granice. Statističke stope pogreške opisuju kako bi se postupak ponašao kroz mnogo ponavljanja. One nisu vjerojatnosti da je odluka u pojedinom istraživanju pogrešna.

Zato prije uvida u podatke zapisujemo nulti model, mjeru odstupanja i prag odluke. Za tako određen postupak možemo procijeniti stopu pogreške. Ako postupak prilagođavamo dobivenim rezultatima, ta stopa više nije zajamčena. Posljedice takvih prilagodbi razmotrit ćemo u poglavlju o krizi i obnovi.

Kako se gradi svijet bez učinka

Uzorak od 300 čitatelja tiska i portala zadržava izvorne oznake skupina. Njihove ocjene povjerenja najprije čitamo prije bilo kakva premještanja.

Odabir smo najprije ograničili na osobe kojima je glavni izvor tisak ili portal, a iz njih izabrali 300 zapisa bez vraćanja. Druge izvore nismo prekodirali u te dvije skupine. U uzorku je 73 čitatelja tiska i 227 korisnika portala. Ocjene zbrajamo zasebno i dijelimo odgovarajućim brojem osoba.

Tablica 1: Od ocjena do dviju sredina u istom uzorku. Izrada autora.

Izvor	Broj osoba	Zbroj ocjena	Sredina
portal	227	1117	4,921
tisak	73	406	5,562

Od sredine za tisak oduzimamo sredinu za portal. Račun zadržava izvorni raspored opažanja, pa razlika opisuje upravo te skupine.

Najprije navodimo procjenu. Čitatelji tiska imaju u prosjeku 0,64 boda više povjerenja, uz interval pouzdanosti od 0,17 do 1,11. Testom ćemo zatim provjeriti koliko su podaci uskladivi s određenim modelom bez razlike.

Taj 95-postotni interval koristimo kao normalnu približnu procjenu za razliku neovisnih sredina. Standardnu pogrešku računamo kao $\sqrt{s_T^2/n_T + s_P^2/n_P}$, gdje su s_T i s_P uzoračke standardne devijacije tiska i portala, a n_T i n_P njihove brojnosti. Razlici dodajemo i od nje oduzimamo $1{,}96$ standardnih pogrešaka. Račun traži neovisne osobe i dovoljno stabilnu približno normalnu raspodjelu razlike. Interval se temelji na tom približnom računu standardne pogreške. Premještanje oznaka upotrijebit ćemo za test. Za manje uzorke poglavlje o dvjema grupama razvija t-prilagodbu i Welchove stupnjeve slobode.

Pretpostavimo da je cijela raspodjela povjerenja jednaka među čitateljima tiska i portala. Pod tom pretpostavkom oznake izvora vijesti zamjenjive su u odnosu na ishod. Možemo ih nasumično premjestiti među osobama, ponovno izračunati razliku sredina i to ponoviti mnogo puta. Dobivene razlike pokazuju što bismo

mogli očekivati pod modelom u kojem oznaka nije povezana s ishodom.

U tablici je prvih šest osoba iz uzorka od 300 čitatelja tiska i portala prije i nakon prvoga nasumičnog preslagivanja svih 300 oznaka. Ocjena prvog retka ostaje 6, a uz nju sada stoji oznaka „portal". Za novi račun grupiramo ocjene prema desnom stupcu, pa se mijenjaju skupni zbrojevi, ali ne ukupan zbroj ocjena ni brojnost tiska i portala.

Tablica 2: Početak istih zapisa prije i nakon jedne permutacije oznaka. Izrada autora.

Redak	Povjerenje	Prije	Poslije
1	6	tisak	portal
2	10	portal	portal
3	2	portal	tisak
4	4	portal	tisak
5	3	portal	tisak
6	5	tisak	portal

Iz svih 300 redaka nakon toga preslagivanja dobivamo razliku 0,080 boda. To je jedna vrijednost nulte raspodjele. Za sljedeću ponovno preslažemo isti skup oznaka uz iste ocjene.

Početnu pretpostavku sada možemo imenovati.

> **Nulta hipoteza** je precizan model postupka koji je proizveo podatke, sastavljen tako da u njemu nema razlike ni veze koju istražujemo.

Alternativna hipoteza ovdje obuhvaća odstupanja od nultog modela. Nulti model daje raspodjelu s kojom uspoređujemo opaženi rezultat. Test stoga ne računa koja je od dviju hipoteza vjerojatnija.

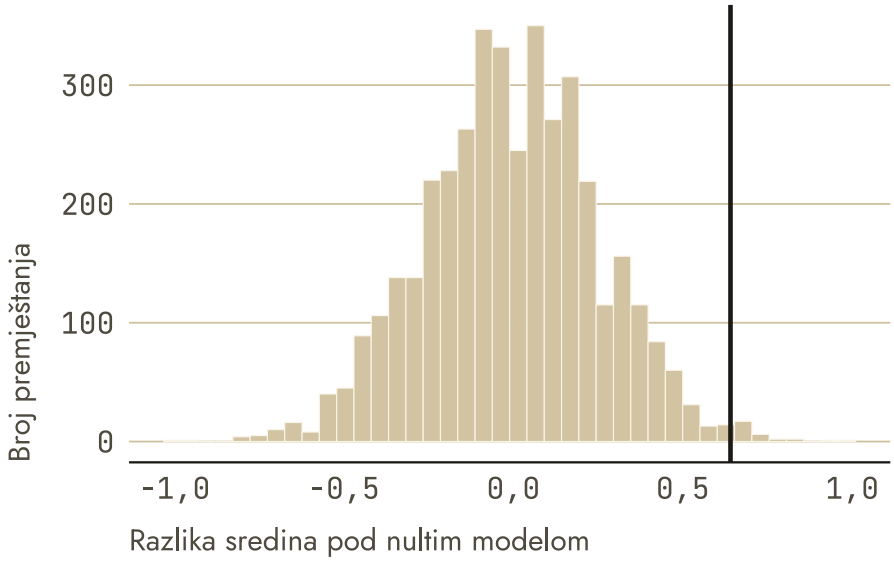
Takvo nasumično premještanje oznaka čini osnovu **permutacijskog testa**.

Nulti model ovog postupka pretpostavlja jednakost cijelih raspodjela, uključujući njihove sredine, raspršenosti i oblike. Pretpostavljamo i da su osobe neovisne jedinice opažanja. Izvor vijesti nije nasumično dodijeljen, pa zamjenjivost oznaka ovdje prihvaćamo kao pretpostavku modela. Iz rezultata ne možemo zaključiti da čitanje određenog izvora uzrokuje povjerenje.

Da bismo usporedili izvorni i premještene rasporede, trebamo isti sažetak za svaki od njih. Taj broj treba pratiti upravo odstupanje koje nas zanima.

> **Testna statistika** je jedan broj izračunat iz podataka koji sažima odstupanje od nultog modela u obliku usporedivom s odstupanjima koja taj model proizvodi.

Kao testnu statistiku koristimo razliku sredina. Zanimaju nas oba smjera, pa uspoređujemo apsolutnu opaženu razliku s apsolutnim razlikama nakon premještanja. Ova jednostavna permutacija ispituje opisani model jednakosti cijelih raspodjela, koji uključuje i njihove oblike i raspršenosti. Slika prikazuje razlike iz četiri tisuće premještanja i razliku iz izvornog rasporeda oznaka.



Slika 1: Razlike sredina kroz četiri tisuće nasumičnih premještanja oznaka skupine, uz okomitu crtu na razlici opaženoj u uzorku. Izrada autora. Podaci i postupak opisani su u tekstu.

Nulta raspodjela ima sredinu u nuli i standardnu devijaciju od 0,26 boda. U 95 % premještanja apsolutna razlika ne prelazi 0,50 boda. Opažena razlika od 0,64 boda veća je od te granice.

Prebrojimo sada premještanja u kojima je apsolutna razlika barem toliko velika kao opažena. Njihov udio koristimo za procjenu p-vrijednosti.

> **P-vrijednost** je udio rezultata koje nulti model proizvodi, a koji su s njime u najmanju ruku jednako neusklađeni kao opaženi rezultat.

U 4 000 preslagivanja nalazimo 64 apsolutnih razlika barem toliko velikih kao opažena. Dodavanjem izvornog rasporeda brojniku i nazivniku račun postaje $(64 + 1)/(4\,000 + 1)$. Sva su preslagivanja valjana jer se brojnost obiju skupina čuva.

U uzorku od 300 čitatelja tiska i portala p-vrijednost iznosi 0,016. To znači da, uz iste opažene ishode i veličine skupina, nulti model daje apsolutnu razliku ovoliku ili veću u otprilike 1,6 % simu-

liranih rasporeda oznaka. U tumačenju moramo zadržati uvjet da vrijedi nulti model.

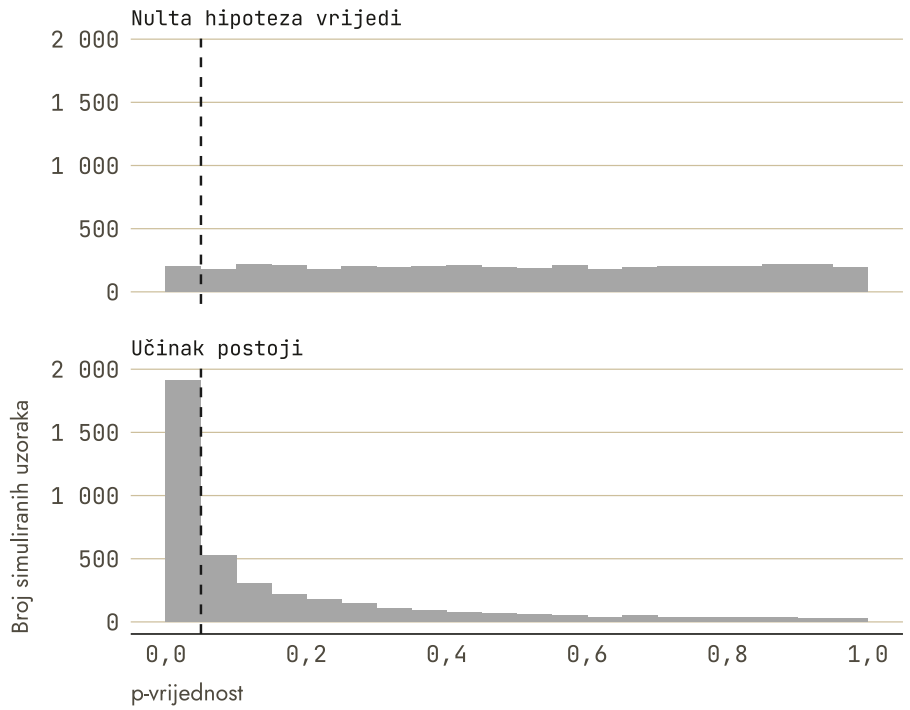
Kad bismo pregledali sva moguća premještanja, izračunali bismo točan udio, uključujući opaženi raspored. Ovdje uzimamo samo konačan broj nasumičnih premještanja. Ako njihov broj označimo s B, a broj apsolutnih razlika barem toliko velikih kao opažena s b, procjenjujemo p-vrijednost kao $(b + 1)/(B + 1)$. Dodavanje opaženog rasporeda sprječava rezultat nula i daje valjanu procjenu za nasumično uzorkovana premještanja.

Simulator p-vrijednosti

U početnim postavkama sljedećeg modela dvije skupine imaju po 60 osoba, a pojedinačni ishodi normalnu raspodjelu sa standardnom devijacijom jedan. Pod nulom obje sredine iznose nula. U scenariju s vezom druga je sredina viša za 0,35. Jedna studija uspoređuje dvije uzoračke sredine i dijeli njihovu razliku poznatom standardnom pogreškom $\sqrt{2/60}$. Simulacija izravno generira tako standardizirani rezultat, što je u ovom modelu jednako generiranju svih pojedinačnih ishoda i računanju iz njih. Digitalni prikaz počinje s 2 000 ponavljanja po scenariju i dopušta promjenu veličine skupina, učinka i broja ponavljanja. Tiskani prikaz zadržava po 60 osoba i razliku 0,35 te koristi 4 000 ponavljanja po scenariju. Svako ponavljanje daje jednu p-vrijednost. Svi rezultati ulaze u prikaz.

Dosad smo za jedan uzorak premještali oznake i promatrali raspodjelu razlika. Sada promatramo p-vrijednosti iz ponovljenih zamišljenih studija. U prvom scenariju učinka nema, a u drugome postoji. Koristimo normalne ishode s poznatom varijabilnošću i obostrani z-test. Neovisnost skupina znači da se varijance njihovih sredina zbrajaju i kada računamo razliku sredina. Standardna pogreška razlike korijen je toga zbroja. P-vrijednost obostranog testa zbroj je površina u oba repa standardne normalne raspodjele, izvan opaženog apsolutnog

z-rezultata. P-vrijednosti ovaj put računamo iz poznate normalne raspodjele, pa korekcija za konačan broj premještanja nije potrebna. Razliku izražavamo u standardnim devijacijama, što ćemo podrobnije objasniti u sljedećem poglavlju.



Slika 2: Raspodjele p-vrijednosti pod nultim modelom i uz standardiziranu razliku 0,35. Izrada autora. Podaci i postupak opisani su u tekstu.

Kada vrijedi nulti model ove simulacije, p-vrijednosti su ravnomjerno raspoređene između nule i jedinice. Jednako široki rasponi oko 0,03 i 0,53 zato su jednako vjerojatni. Prag od 0,05 izdvaja donjih pet posto raspodjele. Upravo taj dugoročni udio odbacivanja pod nultim modelom prag ograničava.

Dvije vrste pogreške

Prag je u prethodnom prikazu izdvojio i neke rezultate nastale bez učinka. Ako na temelju njih odbacimo istinitu nultu hipo-

tezu, činimo pogrešku prve vrste. U drugoj simulaciji učinak postoji, ali dio p-vrijednosti ostaje iznad praga. **Pogreška druge vrste** nastaje kada tako ne odbacimo nultu hipotezu iako je lažna. Isti postupak odlučivanja stoga trebamo procijeniti u oba uvjeta.

> Pogreška prve vrste je odbacivanje nulte hipoteze koja je istinita. Pod pretpostavkama modela postupak ograničava dugoročni udio takvih odbacivanja odabranim pragom.

Istraživački tim, primjerice, treba odlučiti hoće li poduprijeti tvrdnju da se skupine razlikuju. Pogreška prve vrste dovela bi do neutemeljene usporedbe skupina. Pogreška druge vrste mogla bi prikriti postojeću razliku i zaustaviti njezino daljnje istraživanje. Posljedice snose i čitatelji i osobe o kojima se zaključuje. Tim zato treba objasniti koju je pogrešku smatrao ozbiljnijom pri izboru praga.

Stopu pogreške prve vrste možemo provjeriti simulacijom u kojoj nulti model sigurno vrijedi. Iz poznate populacije svaki put uzimamo 300 ishoda, pa im nasumično dodjeljujemo jednako mnogo oznaka A i B. Oznaka je neovisna o ishodu. Tako obje skupine nastaju iz iste pune raspodjele povjerenja.

Sada jedno vanjsko ponavljanje sadrži cijeli test. Odaberemo novih 300 ishoda, nasumično im dodijelimo po 150 oznaka A i B, a zatim za taj uzorak provedemo 300 unutarnjih permutacija i izračunamo jednu p-vrijednost. Ponavljamo to 800 puta i broj p-vrijednosti ispod 0,05 dijelimo s 800. Drugi scenarij uzima nove uzorke iz podskupa tiska i portala s njihovim postojećim oznakama, u kojem veza postoji. To mijenja i sastav skupina, čije brojnosti nisu unaprijed izjednačene. U oba niza sva su ponavljanja definirana i nijedno nije izostavljeno.

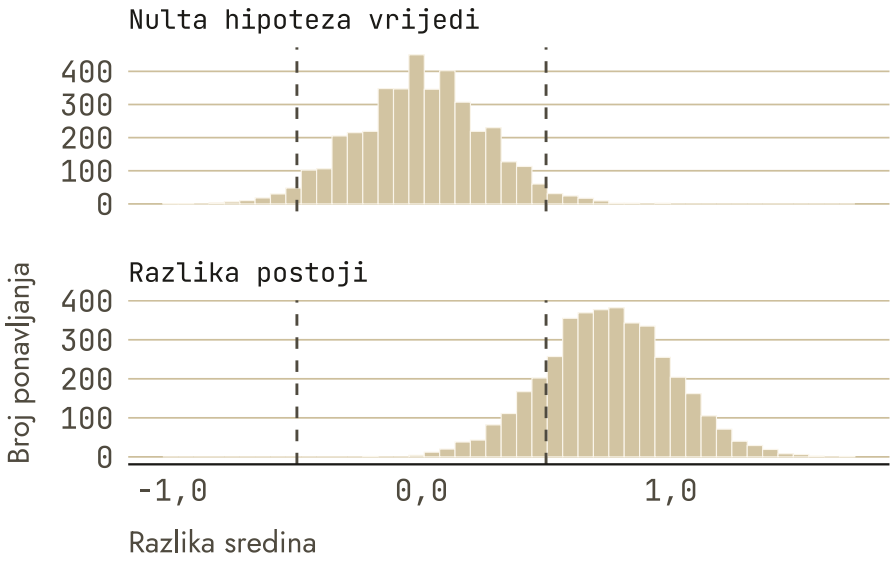
Na osamsto takvih uzoraka permutacijski postupak odbacuje nultu hipotezu uz prag 0,05 u 4,9 % slučajeva. Zbog konačnog broja premještanja i slučajnog kolebanja među uzorcima udio ne mora biti točno pet posto. Ovdje provjeravamo postupak u uvjetima koji zadovoljavaju njegovu pretpostavku zamjenjivosti.

Tablica 3: Udio uzoraka u kojima permutacijska p-vrijednost ne prelazi prag od 0,05, za dva pitanja s poznatim odgovorom. Osamsto uzoraka po pitanju, u svakome tristo premještanja. Izrada autora.

Pitanje	Stvarna razlika u populaciji	Broj odbacivanja	Udio odbacivanja
Nasumično dodijeljena oznaka	0,00	39	4,9 %
Razlika po izvoru vijesti	0,74	636	79,5 %

Drugi redak tablice pokazuje što se događa kada razlika postoji. U populaciji razlika prema izvoru vijesti iznosi 0,74 boda. Postupak je otkriva u 79,5 % uzoraka. U preostalih 20,5 % ne bismo imali dovoljno dokaza za odbacivanje nulte hipoteze, iako razlika postoji.

Na zajedničkoj osi prikazane su raspodjela pod nultim modelom i raspodjela procijenjene razlike kada učinak postoji. Prva premješta oznake u istom uzorku od 300 čitatelja tiska i portala, a druga nastaje novim izvlačenjima iz iste podpopulacije. Iste granične crte određuju odluku u oba slučaja.



Slika 3: Geometrija dviju pogrešaka — uvjetna nulta raspodjela iz glavnog uzorka i raspodjela procijenjene razlike kroz uzorke kad učinak postoji. Isprekidane crte prikazuju granične vrijednosti glavnog uzorka, pa plohe ne prikazuju stope iz tablice. Izrada autora. Podaci i postupak opisani su u tekstu.

Pomaknemo li granične crte prema van, rjeđe odbacujemo istinitu nultu hipotezu. Istodobno veći dio raspodjele s postojećim učinkom ostaje između crta, pa češće propuštamo razliku. Pri istom pragu veći uzorak ili preciznije mjerenje mogu smanjiti pogrešku druge vrste, uz zadržano ograničenje pogreške prve vrste. S više informacija možemo razmotriti i stroži prag, ali trebamo ponovno procijeniti odnos obiju pogrešaka.

Površine na slici objašnjavaju odnos pogrešaka, ali nisu izmjerene stope iz tablice. Gornja raspodjela prikazuje 4 000 premještanja oznaka iz glavnoga uzorka, a donja 4 000 novih uzoraka iz populacije u kojoj razlika postoji. Snaga postupka ovisi o veličini učinka, varijabilnosti, veličini i nacrtu uzorka, testnoj statistici i pragu odluke.

Referentna oznaka nije isto što i istina

Stope pogrešaka mogli smo prebrojiti jer smo u simulaciji zadali kako podaci nastaju. Pri procjeni algoritama susrest ćemo srodno, ali drukčije pitanje, slaže li se predviđena kategorija s **referentnom oznakom**, primjerice ručno dodijeljenom kategorijom. Ta oznaka, za razliku od poznate istine simulacije, može biti pogrešna. Ovisi o mjerenju, uputama za označavanje i načinu rješavanja spornih slučajeva.

Neslaganje s referencom zato može potjecati i od postupka koji ispitujemo i od same reference. U izvještaju navodimo kako su referentne oznake nastale i jesu li neslaganja ponovno pregledana. To će biti posebno važno pri procjeni algoritama. Uz precizno izračunatu stopu neslaganja potrebna je i provjera kvalitete same reference.

Vratimo se prijavama za posao. U objavljenom pokusu životopisi s imenima prepoznatima kao bjelačka dobivali su više poziva na razgovor od životopisa s imenima prepoznatima kao afroamerička (Bertrand i Mullainathan 2004.). Nulta pretpostavka bez učinka dodijeljenog imena dopuštala bi razlike nastale samom nasumičnom dodjelom. Test bi pitao koliko su odstupanja barem toliko velika kao opaženo česta pod takvim postupkom. Račun mora pritom pratiti stvarnu dodjelu prijava i njihovo slanje na iste oglase, umjesto da sve retke automatski smatra neovisnima.

Nasumična dodjela daje osnovu da razliku u odgovorima povežemo s dodijeljenim imenima. Izmjereni ishod ipak je bio poziv na razgovor na obuhvaćenim oglasima, a ne kasnije zaposlenje ili plaća (Bertrand i Mullainathan 2004.). Za osobu koja traži posao i ta je prva prepreka važna. Zaključak o njoj zahtijeva zajedno opis pokusa, veličinu razlike i njezinu nesigurnost. Mala p-vrijednost sama ne bi opravdala tvrdnju o svakom poslodavcu ili o drugom tržištu rada.

Prag je konvencija, a ne mjera

Rezultat nazivamo statistički značajnim kada mu p-vrijednost ne prelazi unaprijed odabrani prag.

Koristimo **prag značajnosti** 0,05, ali moguće je odabrati i 0,01 ili 0,10. Izbor određuje dopuštenu dugoročnu stopu odbacivanja istinite nulte hipoteze. Treba ga obrazložiti stručnim običajima i posljedicama pogreške. Stroži prag ima smisla kada je lažni nalaz osobito skup, uz uvažavanje povećanog rizika da propustimo postojeći učinak.

Rezultati neposredno ispod i iznad praga daju gotovo jednaku količinu dokaza. P-vrijednosti 0,048 i 0,052 ne opravdavaju zaključke da u prvom slučaju učinak postoji, a u drugom ga nema. Zato Američko statističko udruženje traži da znanstveni zaključak ne ovisi samo o prelasku praga (Wasserstein i Lazar 2016.).

Izjava u šest načela objašnjava da p-vrijednost opisuje neusklađenost podataka s određenim modelom. Ne daje vjerojatnost hipoteze, vjerojatnost da je rezultat „nastao slučajno" ni veličinu ili važnost učinka. Za zaključivanje su potrebni potpuno izvještavanje, transparentnost i drugi znanstveni dokazi (Wasserstein i Lazar 2016.).

Kada izvještaj sve svede na oznaku „statistički značajan", čitatelj lako izgubi veličinu procjene, njezinu nesigurnost i posljedice odluke. Te informacije trebaju ostati vidljive bez obzira na položaj p-vrijednosti u odnosu na prag.

P-vrijednost pridonosi znanstvenom argumentu zajedno s procjenom učinka, kvalitetom nacrta i ostalim dokazima. U našem izvještaju zato najprije navodimo procjenu i interval, objašnjavamo pretpostavke i razmatramo posljedice obiju pogrešaka.

P-vrijednost se mijenja i kada ponavljamo potpuno isti postupak. Na osamsto uzoraka iz naše populacije, uz uvijek jednaku

stvarnu razliku, polovica p-vrijednosti bila je ispod 0,007, četvrtina iznad 0,037, a desetina iznad 0,137. Vrijednosti su se protezale preko dva reda veličine. U 12,9 % uzoraka bile su veće od 0,10.

Uzoračka promjenjivost ostaje prisutna i u rezultatu prikazanom na tri decimale. S uzorkom se mijenjaju i procjena i njezin interval. Stavljamo ih na prvo mjesto jer izravno odgovaraju na pitanje kolika je razlika i koliko smo je precizno procijenili.

Što p-vrijednost nije

Za kalibraciju zasebno uzimamo 300 vrijednosti povjerenja iz cijele simulirane populacije i nasumično im dodjeljujemo oznake A i B. Te oznake nisu izvori vijesti. Razlika iznosi −0,09 boda, interval ide od −0,54 do 0,36, a p-vrijednost je 0,70. U ovoj simulaciji znamo da nulti model vrijedi jer oznake nastaju neovisno o ishodu.

Ipak, sam rezultat testa ne dokazuje da razlike nema. Interval obuhvaća vrijednosti u oba smjera, među kojima neke mogu biti sadržajno važne. Velika p-vrijednost govori da podaci ne daju snažan razlog za odbacivanje nultog modela.

Kod male p-vrijednosti česta je obrnuta pogreška. Vrijednost 0,016 odnosi se na rezultate pod pretpostavljenim modelom. Ne daje vjerojatnost modela nakon što smo vidjeli podatke. To su različite uvjetne vjerojatnosti, kakve smo razdvojili u poglavlju o vjerojatnosti.

P-vrijednost ne daje ni vjerojatnost da je nalaz „slučajan”. Račun pretpostavlja određeni nulti model i ne utvrđuje uzrok opaženog rezultata. Također ne mjeri veličinu učinka. Uz dovoljno velik uzorak i mala, praktično nevažna razlika može dati malu p-vrijednost.

Drugo pitanje i drugi račun

Ponekad nas zanima koliko je određena veličina učinka vjerojatna nakon uvida u podatke. Test iz ovog poglavlja ne odgovara na to pitanje. On uspoređuje podatke s preciznim nultim modelom i omogućuje procjenu dugoročnog ponašanja unaprijed određenog postupka. Nulti model posebno je koristan kada odsutnost razlike ili zadana referentna vrijednost ima jasno značenje za odluku.

Bayesovski pristup povezuje početnu raspodjelu vjerojatnosti učinka s modelom koji opisuje podatke pri različitim vrijednostima učinka. Dobivamo ažuriranu raspodjelu vjerojatnosti. Iz nje možemo, uz prihvaćene pretpostavke, procijeniti vjerojatnost da je učinak veći od nule ili neke sadržajno važne granice.

U oba pristupa trebamo navesti pretpostavke. Bayesovski zaključak ovisi o početnoj raspodjeli i modelu podataka, pa izvještavamo koliko je osjetljiv na te izbore. Frekvencijski zaključak ovisi o modelu uzorkovanja ili dodjele, testnoj statistici i unaprijed zadanom postupku. Pristupi postavljaju različita pitanja i drukčije provjeravaju posljedice svojih pretpostavki.

U sljedećim poglavljima nastavljamo s frekvencijskim postupcima. Bayesovski pristup ovdje samo ukratko upoznajemo, a ponovno ćemo ga spomenuti uz regresiju. U oba pristupa veličina učinka i njezina nesigurnost ostaju važne za tumačenje rezultata.

● STATISTIKA U DIVLJINI

Mala p-vrijednost pred uredničkom odlukom. Pretpostavimo da redakcija dobije istraživački izvještaj o razlici u povjerenju između dviju skupina čitatelja. Autor navodi malu p-vrijednost i predlaže naslov prema kojem je dokazao svoje objašnjenje razlike. Urednik treba odlučiti što od toga može prenijeti publici. Uz točnost prepisanog broja urednik treba utvrditi što je njime provjereno.

Takve je prijelaze lakše prepoznati kad ih vidimo na konkretnim tumačenjima. Jedan pregled okuplja dvadeset pet pogrešnih tumačenja p-vrijednosti, intervala i snage, kao pomoć pri čitanju i pisanju istraživačkih zaključaka (Greenland i sur. 2016.). Među njima je zamjena vjerojatnosti podataka pod pretpostavljenim modelom vjerojatnošću da je sama hipoteza istinita. Mala p-vrijednost ne daje postotak sigurnosti u autorovo objašnjenje (Greenland i sur. 2016.).

Urednik bi zato u našem zamišljenom slučaju najprije trebao potražiti procijenjenu razliku i njezin interval. Zatim treba pročitati kako su čitatelji odabrani, što test uzima kao nulti model i jesu li skupine samo opažene ili je provedena nasumična dodjela. Ako je usporedba opažačka, uvjerljivo odstupanje od nultog modela još ne pokazuje što je uzrokovalo razliku.

Važan je i put do objavljenog rezultata. Autori pregleda upozoravaju da biranje analiza prema dobivenim p-vrijednostima može proizvesti male vrijednosti čak i kada provjeravana nulta hipoteza vrijedi (Greenland i sur. 2016.). Redakcija zato ima razlog zatražiti podatak je li prikazana usporedba bila unaprijed određena i koje su druge analize provedene. Takva provjera razjašnjava postupak odabira analize, uz otvoreno pitanje razloga autorovih odluka.

Nakon toga vijest može prenijeti opaženu razliku, njezinu nesigurnost i ograničenje objašnjenja. Ako te informacije nedostaju, urednik može odgoditi snažan naslov. Test je usporedio podatke s određenim modelom, dok predloženi naslov tvrdi nešto o tome zašto se ljudi razlikuju. Između te dvije rečenice mora stajati dokaz koji sama p-vrijednost ne sadrži.

NAĐITE GREŠKU

Provjerite odlomak o razlici povjerenja između čitatelja tiska i portala u poznatoj simuliranoj populaciji. U uzorku je procjena razlike tisak minus portal 0,64 boda, uz 95 % interval od 0,17 do 1,11. Osobe su zasebne jedinice opažanja.

Za test se zadaje nulti model jednake pune raspodjele povjerenja, pod kojim su oznake skupina zamjenjive. Asistentov zapis postupka glasi ovako.

```
razlika <- mean(tisak) - mean(portal)
B <- 4000
nulte <- replicate(B, {
  promijesano <- sample(skupina)
 mean(ishod[promijesano == "tisak"]) - mean(ishod[promijesano == "portal"])
})
b <- sum(abs(nulte) >= abs(razlika))
(b + 1) / (B + 1)
```

Od 4 000 premještanja njih 64 daje apsolutnu razliku barem toliko veliku kao opažena. Korigirani račun zato daje 65 / 4 001, približno 0,016. U izvještaju je asistent napisao sljedeće.

> Pod zadanim nultim modelom opažena je razlika neuobičajena. Skupine su promatračke, pa rezultat ne pokazuje uzrok razlike. P-vrijednost 0,016 znači da je vjerojatnost da između skupina nema razlike 1,6 %.

Povežite brojnik 65 s onim što je postupak prebrojio. Zatim uredite posljednju rečenicu, uz jasno imenovan uvjet pod kojim se vjerojatnost računa. Sačuvajte procjenu razlike, njezinu nesigurnost i ograničenje uzročnog tumačenja.

Razrađeni primjer

Na istom uzorku od 300 čitatelja tiska i portala usporedbu povjerenja provodimo od opažene razlike do njezina položaja u nultoj raspodjeli. Najprije računamo razliku sredina s izvornim oznakama. Zatim zadržavamo ocjene, premještamo oznake skupina i svaki put računamo istu razliku. Prebrojavanjem odstupanja barem toliko velikih kao opaženo procjenjujemo koliko je naš rezultat neobičan pod nultim modelom.

```
razlika <- mean(s10_ishod[s10_skupina == "tisak"]) -
  mean(s10_ishod[s10_skupina == "portal"])

B <- 4000
nulte <- replicate(B, {
  promijesano <- sample(s10_skupina)
  mean(s10_ishod[promijesano == "tisak"]) -
    mean(s10_ishod[promijesano == "portal"])
})

b <- sum(abs(nulte) >= abs(razlika))
razlika
```

```
[1] 0,640939
```

```
(b + 1) / (B + 1)
```

```
[1] 0,01624594
```

U ovom pozivu `sample` bez dodatnih argumenata nasumično mijenja poredak oznaka, dok `replicate` ponavlja račun i sprema razlike. Za razliku od bootstrap izvlačenja osoba, ovdje svako premještanje zadržava iste ocjene i iste veličine skupina. Ne zadajemo parametarski oblik raspodjele, ali i dalje pretpostavljamo neovisne jedinice i zamjenjivost oznaka pod nultim modelom jednake pune raspodjele. Brojimo odstupanja u oba repa jer je test obostran. Jedinicu dodajemo brojniku i nazivniku radi korekcije za konačan broj nasumičnih premještanja.

U izvještaju povezujemo veličinu razlike s rezultatom testa. Čitatelji tiska imaju u prosjeku 0,64 boda više povjerenja od čitatelja portala, uz interval od 0,17 do 1,11. Nulti model daje apsolutnu razliku ovoliku ili veću u 1,6 % simuliranih rasporeda opaženih oznaka. Te brojke opisuju samo naš simulirani uzorak. U stvarnoj anketi takva bi usporedba pokazivala razliku među skupinama. Za zaključak da čitanje tiska mijenja povjerenje trebali bismo i odgovarajući istraživački nacrt.

Uzročni zaključak zahtijevao bi istraživački nacrt koji razdvaja moguća objašnjenja razlike. Zasebno procjenjujemo koliko je razlika važna za istraživačko pitanje te opravdava li neku odluku.

Sažetak

Testom smo opaženi rezultat usporedili s raspodjelom nultog modela, koju smo ovdje izgradili premještanjem oznaka. P-vrijednost opisuje udio odstupanja barem toliko velikih kao opaženo, uz zadržan uvjet da vrijedi taj model, pa ne dobivamo vjerojatnost hipoteze ni važnost učinka. Izbor praga povezuje dvije moguće pogreške jer stroži prag smanjuje stopu prve, ali pri istim podacima povećava stopu druge, dok više informacija može smanjiti drugu pri istom pragu. Simulacijom s nasumičnim oznakama provjerili smo postupak kada nulti model vrijedi, a uz postojeću razliku od 0,74 boda vidjeli smo da je propušta u 20,5 % uzoraka. Ta nas propuštanja vode pitanju kako prije prikupljanja podataka planirati uzorak za učinak koji vrijedi uočiti. U sljedećem poglavlju odgovorit ćemo na njega pomoću veličine učinka i snage.

Pojmovi

nulta hipoteza (*null hypothesis*), nulti model (*null model*), obostrani test (*two-sided test*), testna statistika (*test statistic*), p-vrijednost (*p-value*), pogreška prve vrste (*Type I error*), pogreška druge vrste (*Type II error*), prag značajnosti (*significance threshold*), permutacijski test (*permutation test*), referentna oznaka (*reference label*)

Zadaci

Konceptualni

Objasnite u jednom odlomku zašto p-vrijednost nije vjerojatnost nulte hipoteze, tako da napišete dvije rečenice koje se razlikuju samo po tome što je uvjet, a što zaključak. Imenujte podatak koji bi bio potreban da se od prve dođe do druge.

Računski

Nulta raspodjela iz ovog poglavlja ima sredinu nula i standardnu devijaciju 0,26 boda. Primijenite pravilo područja iz poglavlja o vjerojatnosti i procijenite granice koje su dvije standardne devijacije udaljene od nule. Prije računa imenujte testnu statistiku i postupak kojim je nulta raspodjela izgrađena, pa procjenu usporedite s graničnom vrijednošću 0,50. Zatim pretpostavite da je isti uzorak prikupljen dobrovoljnom poveznicom na jednom portalu. Pozivajući se na razliku između slučajnosti uzorkovanja i selekcije iz poglavlja o uzorkovanju, odredite koju tvrdnju o populaciji ni mala p-vrijednost ne bi mogla opravdati. U tiskanom ili dokumentnom izdanju na kraju očitajte udio iz prvoga retka tablice stopa odbacivanja, u kojem stvarne razlike nema.

Kritički

Prosudite zašto je jednom strukovnom udruženju trebalo objaviti izjavu o pojedinačnom statističkom postupku, a skupini statističara popis od dvadeset pet pogrešnih tumačenja (Wasserstein i Lazar 2016.; Greenland i sur. 2016.). Predajte kratku uredničku bilješku i navedite jedno pravilo izvještavanja koje bi uklonilo najviše pogrešaka odjednom. Dodajte tko u glavnom primjeru snosi posljedice svake vrste pogreške i zašto procijenjena stopa pogreške prema referentnim oznakama ne dokazuje da su te oznake nepogrešive.

Revizija modela

Ocijenite analizu iz okvira o pogrešci. Imenujte korake koji su provedeni ispravno, redak koda iz kojeg izlazi izvještajna brojka, rečenicu koja iz nje ne slijedi i njezinu ispravljenu inačicu.

11

DIO IV · ZAKLJUČIVANJE

veličina učinka i snaga

Koliki je učinak i koliko ga pouzdano vidimo

Otvaramo društvenu mrežu i uz poziv na izbore ugledamo lica prijatelja koji su označili da su već glasovali. Možda smo ionako namjeravali izaći, a možda nas baš taj podsjetnik pokrene. Kako bismo prepoznali njegov doprinos među svim razlozima zbog kojih ljudi odlaze na biralište?

U velikom pokusu tijekom američkih izbora istraživači su nasumično dodjeljivali različite poruke korisnicima mreže (Bond i sur. 2012.). Neke su prikazivale fotografije prijatelja koji su označili da su glasovali, a druge istu poruku bez tih fotografija (Bond i sur. 2012.). Glasovanje su mogli provjeriti u biračkim evidencijama samo za upareni podskup korisnika (Bond i sur. 2012.).

Za odluku o takvoj poruci trebamo procijeniti veličinu učinka i što bi on značio pri širokoj primjeni, uz granice skupine za koju je ishod provjeren. Koliki bi pomak bio sadržajno važan i koliko podataka trebamo da bismo ga pouzdano uočili?

Pri planiranju istraživanja polazimo od razlike koju bi vrijedilo uočiti. Zatim provjeravamo možemo li je planiranim postupkom dovoljno pouzdano razlikovati od slučajne promjenjivosti. Time pitanje iz prethodnog poglavlja okrećemo prema planiranju. Umjesto da tek nakon analize čitamo rezultat testa, unaprijed provjeravamo kakav bi nam odgovor istraživanje moglo dati.

Razliku u povjerenju čitatelja tiska i portala već smo procijenili s intervalom. Sada nas zanima njezina veličina i koliko je često uzorak može uočiti. Za poznatu referencu služe svi čitatelji tiska i portala u populaciji `populacija_medija`, umjesto prethodnog uzorka od 300 čitatelja tiska i portala.

Svaki redak predstavlja jednu zamišljenu osobu s izvorom vijesti i povjerenjem na ljestvici od jedan do deset. Podaci su stvoreni za nastavu i ne opisuju stvarne stanovnike. Poznatu razliku uspoređujemo s rezultatima ponovljenih izvlačenja kako bismo prebrojili koliko je često postupak uočava.

Razliku najprije izražavamo u bodovima povjerenja, a zatim je uspoređujemo s raspršenošću odgovora i opisujemo Cohenovim d. Sredina i standardna devijacija iz opisne statistike tako zajedno daju standardiziranu veličinu učinka. Važnost pomaka ipak trebamo obrazložiti sadržajem pitanja i mogućim posljedicama odluke. Vrlo malen pomak može postati statistički značajan uz dovoljno opažanja, što ćemo provjeriti simulacijom dviju naslovnica.

Snagu ćemo proučavati mijenjanjem učinka, veličine uzorka, raspršenosti i praga odluke. Provjerit ćemo i što se događa kada iz malih istraživanja zadržimo samo značajne rezultate.

Planirati možemo i unatrag, od razlike na koju bi zamišljeno uredništvo reagiralo prema potrebnom broju osoba. Drugi je mogući cilj prihvatljiva širina intervala. Usporedba tih ciljeva pomaže nam obrazložiti što od uzorka očekujemo prije nego što vidimo nalaz i prepoznati kada je nesigurnost prevelika za namjeravanu odluku.

Razlika koja nešto znači

Veličina učinka brojčano opisuje razmjer razlike ili odnosa. Možemo je izraziti u izvornim jedinicama, kao razliku bodova, ili bez njih, kao standardiziranu razliku, korelaciju ili omjer. Njezinu sadržajnu važnost procjenjujemo prema istraživačkom pitanju i posljedicama odluke.

Povjerenje u medije bilježimo bodovima na ljestvici, pa razliku među čitateljima tiska i portala prvo čitamo u tim jedinicama.

Za ovu referencu služe svi čitatelji tiska i portala u populaciji, ukupno 19956 osoba. Zbroj ocjena među 4855 čitatelja tiska iznosi 26791, a među 15101 korisnika portala 72101. Dijeljenjem svakog zbroja njegovim brojem osoba dobivamo sredine 5,518 i 4,775, pa ih možemo oduzeti.

Razlika iznosi 0,74 boda na ljestvici od jedan do deset. Navodimo je u izvornim jedinicama kako bismo je mogli povezati s odlukom. Uredništvu koje razmatra preraspodjelu sadržaja nije svejedno razlikuju li se skupine za tri četvrtine boda ili za tri boda.

Razliku možemo usporediti i s raspršenošću odgovora. Jedan bod izgleda drukčije kada su odgovori zbijeni oko sredine nego kada zauzimaju cijelu ljestvicu. Standardizirana mjera izražava taj odnos.

Standardizirana razlika je razlika dviju sredina podijeljena združenom standardnom devijacijom skupina, pa se izražava u standardnim devijacijama umjesto u izvornim jedinicama.

$$d = \frac{\bar{x}_2 - \bar{x}_1}{s_{\text{zdr}}}$$

Oznakom s_{zdr} označavamo združenu standardnu devijaciju, zajedničku mjeru raspršenosti unutar dviju skupina.

Za zajedničku raspršenost povezujemo dvije varijance. Svaka dobiva težinu jednaku broju osoba u skupini umanjenom za jedan, pa veća skupina više pridonosi. Tek nakon toga vadimo korijen. Skupina 1 ovdje je portal, a skupina 2 tisak. Brojeve osoba označavamo s n_1 i n_2, a standardne devijacije s s_1 i s_2. Za dva uzorka račun glasi

$$s_{\text{zdr}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

U ovom primjeru račun obavljamo na cijeloj poznatoj sintetičkoj populaciji. Zato dobivena razlika i njezina standardizacija opisuju te zapise bez uzoračke nesigurnosti. Uzoračka procjena iste veličine zahtijevala bi i interval. U brojnik ulazi $(15101 - 1) \times 3{,}641 + (4855 - 1) \times 3{,}705$, a nazivnik je 19954. Račun koristi nezaokružene varijance.

Združena standardna devijacija u ovoj populaciji iznosi 1,91 boda. Razlika od 0,74 boda zato daje standardiziranu razliku od 0,39. Skupine dijeli nešto više od trećine standardne devijacije, a njihove se raspodjele velikim dijelom preklapaju. Ista ocjena povjerenja može pripadati korisnicima obaju izvora, pa ona sama ne određuje izvor vijesti pojedine osobe.

Preklapanje raspodjela možemo približiti i pitanjem o nasumično izabranom paru osoba. Čitatelj tiska ima strogo više povjerenja od čitatelja portala u 53,6 % parova, dok je 13,9 % parova izjednačeno. Kada polovicu izjednačenja pripišemo svakoj strani, prilagođena vjerojatnost nadmoći iznosi 60,6 %. Za jednake raspodjele iznosila bi pedeset posto. Razliku među skupinama tako opisujemo usporedbom ljudi, što može biti razumljivo i čitatelju koji ne poznaje standardnu devijaciju.

Sadržajnu važnost učinka procjenjujemo prema ranijim nalazima iz istog područja i posljedicama odluke. Standardizacija olakšava usporedbu različitih ljestvica, a prosudba važnosti ostaje vezana uz sadržaj istraživanja.

Kad je značajno, a nije važno

Rezultat nazivamo statistički značajnim kada njegova p-vrijednost ne prelazi unaprijed odabrani prag odluke. Uz statističku značajnost navodimo i procjenu veličine učinka kako bismo opisali razmjer razlike. To se posebno jasno vidi na vrlo velikom uzorku.

Zamislimo pokus u kojem portal uspoređuje dvije naslovnice. Stvarna stopa klika prve iznosi 2,00 %, a druge 2,05 %. Svaku prikazuje pola milijuna puta. Podatke u nastavku simuliramo za tu zamišljenu usporedbu.

Jedno opažanje predstavlja jedan hipotetski prikaz naslovnice i njegov ishod, klik ili izostanak klika. Pretpostavljamo neovisne prikaze s istom vjerojatnošću unutar naslovnice. Račun generira zbroj klikova u svakoj od dviju skupina od 500 000 prikaza. Za prvu dobivamo 9879 klikova, a za drugu 10382. Dijeljenjem svakog broja s 500 000 nastaju stope koje uspoređujemo. Ovo je jedan simulirani pokus, ne procjena snage preko ponovljenih pokusa.

Prikazani 95-postotni interval i p-vrijednost koriste normalni Waldov račun za razliku dvaju neovisnih udjela, s procijenjenim varijancama obiju skupina. Ostvarena razlika veća je od zadane razlike od 0,05 postotnih bodova. To je slučajno precjenjivanje u ovome pokusu, a ne promjena poznatoga učinka. Značajnost ovog jednog rezultata ne mjeri snagu nacrta.

Uz zadane stope i broj prikaza normalno približenje daje snagu od oko 43 %. Otkrivanje stvarnoga pomaka zato je u ovom nacrtu rjeđe od njegova propuštanja.

U simulaciji prva naslovnica postiže 1,976 %, a druga 2,076 % klikova. Procijenjena razlika iznosi 0,101 postotnog boda, uz interval od 0,045 do 0,156. P-vrijednost iznosi 0,00036. Pod nultim modelom ovako velika ili veća odstupanja bila bi vrlo rijetka.

Dobivena razlika znači otprilike jedan dodatni klik na tisuću prikaza. Opravdava li promjenu naslovnice, ovisi o trošku promjene i vrijednosti klika. Trošak i vrijednost klika trebamo utvrditi iz dodatnih izvora. Velik uzorak omogućio nam je razlikovati mali pomak od nule, a te će informacije pomoći u obrazlaganju odluke.

U malom uzorku možemo propustiti i razliku koja bi bila važna za uredničku odluku. Izostanak značajnosti zato ne dokazuje da je učinak nevažan. Procjenu i interval izvještavamo prije testa kako bi čitatelj mogao razmotriti i veličinu i nesigurnost rezultata.

Poruka s fotografijama prijatelja iz početnog pokusa pokazuje zašto i malen pomak zaslužuje sadržajno tumačenje. Među korisnicima čije je glasovanje provjereno u evidenciji procijenjena razlika prema poruci bez fotografija iznosila je 0,39 postotnog boda, uz standardnu pogrešku 0,17 postotnog boda (Bond i sur. 2012.). Normalnim približenjem $0{,}39 \pm 1{,}96 \times 0{,}17$ iz tih objavljenih vrijednosti dobivamo približan 95-postotni interval od 0,06 do 0,72 postotnog boda (Bond i sur. 2012.). Procjena je mala za pojedinca, a njezina nesigurnost ostaje važna i uz vrlo mnogo korisnika.

U zamišljenoj primjeni na milijun usporedivih ljudi pomak od 0,39 postotnog boda odgovarao bi približno 3 900 dodatnih izlazaka. To je ilustracija razmjera, uz pretpostavku da se učinak može prenijeti na tu skupinu. Za stvarnu odluku trebalo bi provjeriti taj prijenos, trošak poruke i prihvatljivost korištenja prijateljskih veza za poticanje ponašanja. U samom su pokusu istraživači zasebno proučavali i širenje učinka kroz mrežu prijatelja (Bond i sur. 2012.). Jednostavno množenje izravnog učinka zato još ne daje ukupan društveni učinak.

Snaga kao svojstvo plana

Prije prikupljanja podataka možemo procijeniti koliko bi često naš plan otkrio zadani učinak. Za to trebamo odrediti postupak, veličinu učinka, uzorak, raspršenost i prag odluke.

Ako nakon istraživanja opaženi učinak proglasimo stvarnim i iz njega izračunamo snagu, račun je kružan. Uz isti uzorak i prag takva snaga prati već opaženu testnu statistiku i ne daje neovisnu provjeru je li dizajn mogao otkriti unaprijed važan učinak.

> **Statistička snaga** je vjerojatnost da postupak odbaci nultu hipotezu kad je ona lažna, dakle udio ponavljanja u kojima bi učinak zadane veličine bio otkriven.

U prethodnom poglavlju postupak je otkrio stvarnu razliku u otprilike četiri od pet uzoraka od tristo osoba. Taj udio procjenjuje snagu. Sada ćemo ga usporediti za nekoliko veličina uzorka.

Zadržavamo permutacijski test iz prethodnog poglavlja. Nulti model pretpostavlja jednaku cijelu raspodjelu povjerenja u objema skupinama, pa su oznake zamjenjive u odnosu na ishod. Statistika je apsolutna razlika sredina, a osobe smatramo neovisnim jedinicama. Iz podskupa koji čine svi čitatelji tiska i portala u populaciji izvlačimo uzorke bez vraćanja. Za svaku veličinu provodimo tristo zamišljenih studija. U svakoj koristimo dvjesto nasumičnih premještanja i korekciju $(b + 1)/(B + 1)$, gdje je B broj premještanja, a b broj barem toliko ekstremnih rezultata kao opaženi. Krivulja procjenjuje snagu za taj postupak i način nastanka podataka.

Veličina uzorka u ovoj tablici jest ukupan broj osoba iz obiju skupina, čije brojnosti nisu unaprijed izjednačene. Kasniji plan s jednakim skupinama navodi broj po skupini. Uz 200 permutacija moguće p-vrijednosti razlikuju se za 1/201, približno 0,005. Odluka neposredno uz prag zato može ovisiti o slučajnom izboru premještanja.

U svakom retku brojimo studije s p-vrijednošću koja ne prelazi 0,05 i dijelimo brojem valjanih studija. Ako bi uzorak sadržavao samo jednu skupinu, test ne bi bio definiran i studiju bismo isključili, uz prikazan broj isključenja. U ovoj izvedbi takvih studija ima 0. Tristo ponavljanja daje razmjerno grubu procjenu snage. Oko snage od 80 % simulacijska standardna pogreška iznosi približno 2,3 postotna boda.

Tablica 1: Udio uzoraka u kojima permutacijska p-vrijednost ne prelazi prag od 0,05 pri stvarnoj razlici od 0,74 boda. Tristo simuliranih studija po retku. Izrada autora.

Ukupno osoba	Otkriveno	Valjane studije	Izostavljene	Snaga
60	68	300	0	22,7 %
100	107	300	0	35,7 %
200	170	300	0	56,7 %
300	239	300	0	79,7 %
500	281	300	0	93,7 %
800	299	300	0	99,7 %

Kada povećamo uzorak sa šezdeset na tristo osoba, procijenjena snaga raste s 22,7 % na 79,7 %. Rast se postupno usporava kako se približavamo gornjoj granici. Sa šezdeset osoba ovaj bi postupak propustio stvarnu razliku u približno tri od četiri studije.

Snaga raste s većim učinkom i većim uzorkom. Pomažu i manja raspršenost ishoda te blaži prag odluke. Istraživač može planirati uzorak i prag, a nacrtom i kvalitetom mjerenja utjecati na raspršenost. Veličina stvarnog učinka ostaje nepoznata, pa snagu treba razmotriti za više mogućih veličina.

Često se planira snaga od 80 %, ali taj izbor treba opravdati (Cohen 1988.). Ako bi propuštena razlika imala ozbiljne posljedice, možda trebamo višu snagu. Istraživanje izvedivosti može imati drukčiji cilj i zahtjeve.

Odabir značajnih nalaza pri niskoj snazi

Slaba snaga povećava broj propuštenih učinaka. Može utjecati i na veličinu procjena koje objavljujemo ako zadržavamo samo značajne rezultate. To možemo provjeriti u simulaciji s poznatom populacijom.

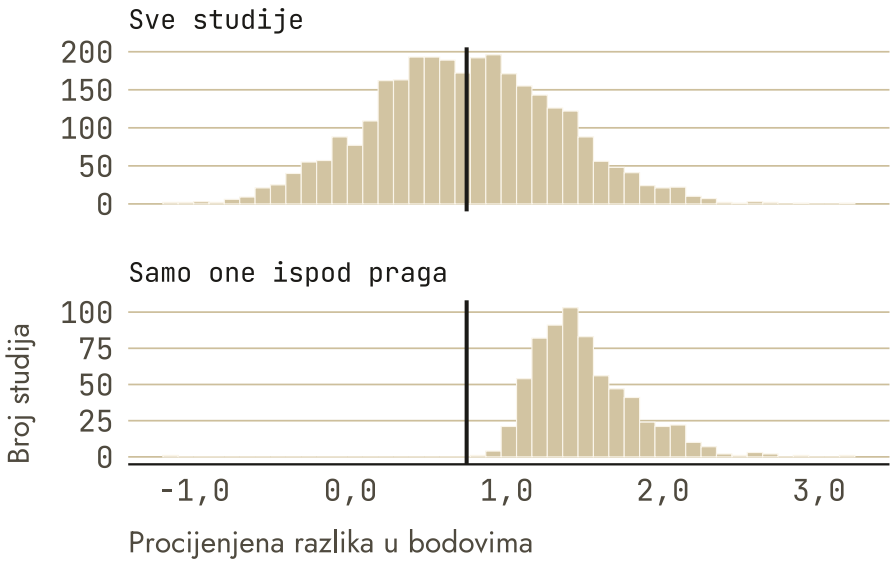
Za sljedeću usporedbu ponovno izvlačimo uzorke iz istoga podskupa tiska i portala, svaki put po 60 osoba. Svaka studija daje jednu procjenu razlike i jednu permutacijsku p-vrijednost. Tablica donosi prvih pet valjanih studija u redoslijedu generiranja. Oznaka u posljednjem stupcu pokazuje bi li studija bila objavljena kada bismo zadržali samo rezultate ispod 0,05.

Tablica 2: Pet simuliranih studija prije odabira prema pragu. Izrada autora.

Studija	Razlika	p	Zadržana
1	0,098	0,9254	ne
2	1,104	0,1095	ne
3	1,444	0,0498	da
4	1,140	0,0498	da
5	0,085	0,9104	ne

Prva procjena iznosi 0,098 boda i ulazi u prosjek svih studija. U prosjek odabranih ulazi samo ako zadovolji prag. Od 3 000 pokušaja neprimjenjivo je 0, pa prosjek svih računamo na 3000 valjanih studija, a prosjek značajnih na 678. Isti manji nazivnik koristimo za udio značajnih procjena s pogrešnim predznakom. Kod zasebnih 1 000 studija s po 500 osoba izostavljeno je 0, a prag prolazi 961 studija.

Proveli smo tri tisuće zamišljenih studija sa šezdeset osoba. Prosjek svih procjena iznosi 0,74 boda, blizu stvarne razlike od 0,74. Prije odabira prema pragu procjene u ovoj simulaciji ne pokazuju pristranost.



Slika 1: Procijenjena razlika u tri tisuće studija sa šezdeset osoba, sve zajedno i samo one koje su prešle prag. Okomita crta označuje stvarnu razliku u populaciji. Izrada autora. Podaci i postupak opisani su u tekstu.

Donji panel zadržava samo rezultate koji su prešli prag značajnosti. Njihov prosjek iznosi 1,50 boda, odnosno 2,0 puta onoliko koliko iznosi stvarna razlika. U malom uzorku prag najčešće prelaze procjene koje je slučajnost povećala. Najmanja apsolutna vrijednost značajne procjene u ovoj simulaciji iznosi 0,84 boda. Sam odabir rezultata prema pragu može stvoriti to iskrivljenje, čak i kada istraživač postupa u dobroj vjeri.

Među 678 značajnih procjena njih 1 ima suprotan predznak od stvarne razlike. Račun može biti točan, a pojedinačni zaključak o smjeru učinka pogrešan.

S pet stotina osoba značajne procjene premašuju stvarnu razliku za faktor 1,03, pa je iskrivljenje mnogo manje. Ti brojevi vrijede za ovu populaciju, uzorke, postupak i prag. Ne možemo ih prenijeti na svaki nalaz iz maloga uzorka.

Pri čitanju rada zato provjeravamo širinu intervala, način odabira rezultata i odnos prema ranijim dokazima. Velik učinak uz p-vrijednost blizu praga sam po sebi ne dokazuje precjenjivanje. Sličan rezultat iz neovisne replikacije daje dodatni razlog za

povjerenje u procjenu. Niska snaga upozorava na nesigurnost i moguću selekciju nalaza.

Planiranje unatrag

Veličinu uzorka planiramo prema pitanju koje želimo riješiti. Najprije trebamo odrediti kolika bi razlika bila dovoljna da promijeni postupanje.

> **Najmanji važan učinak** je najmanja razlika koja bi promijenila zaključak, odluku ili postupanje, određena sadržajno i prije prikupljanja podataka.

Tu granicu obrazlažemo troškovima odluke, ozbiljnošću ishoda i ranijim nalazima iz područja. Tako obrazložena granica postaje ulaz u račun potrebnog broja opažanja za željenu snagu.

Polazimo od važne razlike i željene snage, pa računamo potreban broj jedinica. Zatim provjeravamo je li takav uzorak izvediv. Ako nije, ograničenje znamo prije prikupljanja podataka i možemo prilagoditi plan ili doseg istraživanja.

Veličinu uzorka možemo odrediti i vraćanjem na pitanje iz poglavlja o procjeni. Koliko uzak interval trebamo da bi procjena bila korisna? Uz raspršenost iz ove populacije interval razlike širok je 1,06 boda pri stotinu osoba po skupini, 0,61 pri tristo i 0,37 pri osamsto. Margina je polovica ukupne širine, pa za marginu do pola boda tražimo širinu do jednoga boda. Ulazi u taj plan jesu željena širina intervala, razina pouzdanosti i pretpostavljena raspršenost.

Uz iste pretpostavke kao u završnom planu, poznatu standardnu devijaciju 1,9 i dvije jednake neovisne skupine, margina do 0,5 traži $n \geq 2(1{,}96 \times 1{,}9/0{,}5)^2$, odnosno 111 osoba po skupini. Za otkrivanje razlike od 0,5 uz snagu oko 80 % normalno približenje daje $n \approx 2[(1{,}96 + 0{,}842)1{,}9/0{,}5]^2$, približno 227 osoba po skupini. Broj 0,842 jest kvantil standardne normalne raspodjele

ispod kojega leži 80 % površine. U ovoj približnoj planskoj formuli odgovara ciljanoj snazi od 80 %, dok 1,96 dolazi iz obostranoga testa na razini 5 %. Ta dva plana odgovaraju na različita pitanja pod jednakim pretpostavkama.

Planiranje prema snazi zahtijeva pretpostavljeni učinak, koji još ne znamo. U ovom modelu planiranje prema preciznosti polazi od razine pouzdanosti, pretpostavljene raspršenosti i prihvatljive širine intervala. Cilj je dovoljno precizna procjena, bez obzira na to gdje će se nalaziti. Ostvarena preciznost ipak ovisi o tome koliko su planske pretpostavke odgovarale podacima.

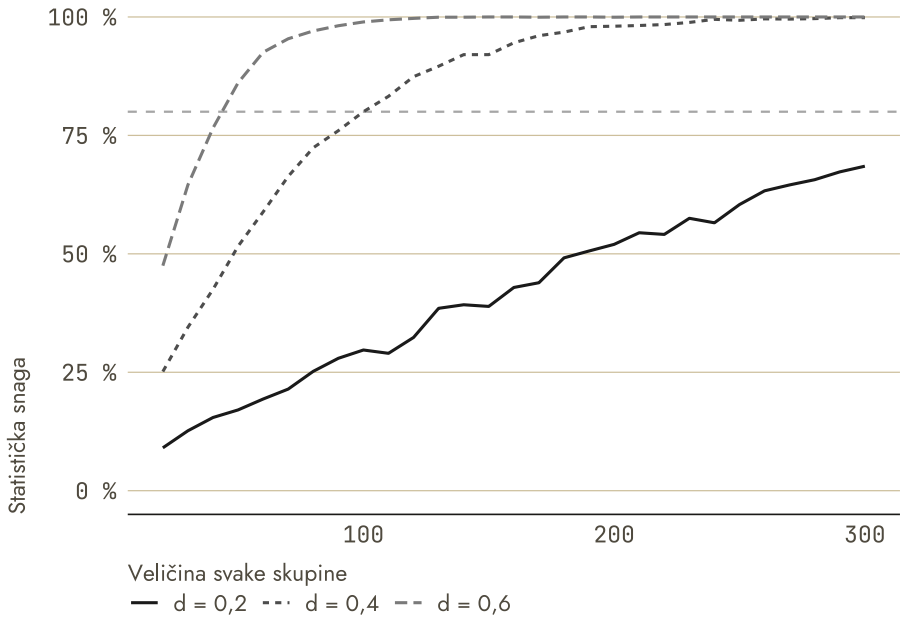
Istraživač snage

Dosad smo mijenjali veličinu uzorka iz populacije sa zadanom razlikom. Sada želimo vidjeti i što se događa kada pretpostavimo manji ili veći učinak. Zato prelazimo na jednostavan model u kojem učinak možemo mijenjati, a raspršenost zadržati jednakom. Krivulje prikazuju snagu za različite uzorke i zadane učinke. Za isti učinak veći uzorak povećava vjerojatnost otkrivanja, dok je uz isti uzorak manji učinak teže otkriti.

Jedna točka krivulje sažima zamišljene studije za zadani učinak i broj osoba po skupini. Tiskani prikaz koristi 2 000 ponavljanja po točki i učinke 0,2, 0,4 i 0,6. Primjerice, za razliku 0,4 i po 100 osoba standardna je pogreška razlike $\sqrt{2/100}$ kada je poznata standardna devijacija ishoda jedan. Svaka studija daje razliku podijeljenu tom pogreškom. Broj rezultata kojima je apsolutna vrijednost iznad 1,96 dijelimo s brojem ponavljanja za tu točku. Svi rezultati ulaze u nazivnik. Krivulja ne broji pojedinačne osobe, a njezina mala kolebanja potječu od simulacije.

Prikaz pretpostavlja dvije neovisne skupine jednake veličine, neovisne normalne ishode i zajedničku poznatu standardnu devijaciju. Nulta hipoteza zadaje standardiziranu razliku nula. Obostrani test koristi apsolutnu z-vrijednost. Za svaku točku simuliramo z-vrijednosti iz odgovarajuće normalne raspodjele.

Više ponavljanja daje manju simulacijsku nesigurnost procjene snage.



Slika 2: Simulirana snaga kroz veličine uzorka pri trima standardiziranim razlikama. Izrada autora. Podaci i postupak opisani su u tekstu.

Za računski zadatak na kraju poglavlja trebaju nam i vrijednosti iz populacije koju smo dosad uspoređivali. Tablica donosi dva retka iz datoteke data/populacija-medija-agregat.csv. Brojevi osoba i zbrojevi preneseni su točno, a prosjeci su zaokruženi na pet decimala. Dijeljenjem zbroja brojem osoba možemo provjeriti prikazani prosjek.

Tablica 3: Brojnosti, zbrojevi i zaokruženi prosjeci za dvije uspoređene skupine. Izrada autora.

Skupina	Broj	Zbroj povjerenja	Prosjek povjerenja
portal	15 101	72 101	4,77458
tisak	4 855	26 791	5,51823

Srednja krivulja odgovara standardiziranoj razlici 0,4, najbližoj ponuđenoj vrijednosti učinku u našoj simuliranoj populaciji. Četiri vrijednosti u nastavku omogućuju usporedbu i bez očitavanja grafa.

Tablica 4: Tiskana postavka snage za standardiziranu razliku 0,4 i prag 0,05. Izrada autora.

Jedinica po skupini	Snaga
40	42,5 %
80	72,4 %
160	94,6 %
300	99,9 %

Za najmanju prikazanu standardiziranu razliku, 0,2, tristo jedinica po skupini još ne daje snagu od 80 %. Trebao bi veći uzorak. Zato pri planiranju moramo jasno reći koju veličinu učinka želimo moći otkriti.

● STATISTIKA U DIVLJINI

Velik objavljeni učinak i plan male studije. Za novi istraživački projekt privlačno je uzeti učinak iz prethodnog rada i prema njemu odrediti potreban broj sudionika. Ako je objavljena razlika velika, plan može izgledati izvedivo i uz skroman uzorak. No prije nego što tu brojku pretvorimo u obećanje budućeg istraživanja, trebamo razmotriti kako je dospjela u literaturu.

Analiza istraživanja u neuroznanosti pokazala je i drugu posljedicu niske snage. Uz propuštanje postojećih učinaka, mala istraživanja mogu ostaviti i pretjeran dojam o njihovoj veličini ako se nalazi odabiru za objavu prema značajnosti (Button i sur. 2013.). U takvom izboru lakše prolaze rezultati koji su, uz postojeći učinak, dobili i povoljan slučajni pomak.

Zamislimo tim koji na temelju jednoga takvog nalaza planira ponavljanje. Ako precijenjenu razliku prihvati kao realan cilj, mogao bi odabrati uzorak koji ima premalu snagu za skromniji, ali sadržajno važan učinak. Tada neodlučan rezultat ne bi jasno razlikovao izostanak učinka od nedostatka informacija. Velik broj u prvom radu tako može dovesti do premalog plana u drugome. Zamišljeni tim ilustrira moguću posljedicu takvog planiranja.

Odgovor je pregledati interval iz prethodnog istraživanja, neovisna ponavljanja i širi skup dostupnih nalaza. Uz njih tim treba odrediti najmanju razliku koja bi promijenila njegov sadržajni zaključak te ispitati koliko sudionika treba za nekoliko uvjerljivih veličina učinka. Ako izvediv uzorak ne može dovoljno precizno odgovoriti na pitanje, korisno je to znati prije prikupljanja.

Naša simulacija pokazala je mehanizam odabira, ali njezin faktor precjenjivanja vrijedi za zadanu populaciju i postupak. Ne možemo njime korigirati sve objavljene radove. Za planiranje je korisnija promjena polaznog pitanja. Umjesto da pitamo koliko je ljudi dovoljno za objavljenu brojku, pitamo kakav nam je uzorak potreban da razlučimo mogućnosti koje su važne za naše istraživanje.

NAĐITE GREŠKU

Pregledavate završni izvještaj male studije. Za ovu je vježbu konstruiran nalaz standardizirane razlike od 0,78 uz trideset osoba po skupini. Prije prikupljanja tim je kao najmanju važnu razliku odredio 0,5 boda i za planiranje pretpostavio standardnu devijaciju od 1,9 boda. Te su vrijednosti pretpostavke nastavnog primjera, a ne objavljen nalaz.

Asistent je nakon uvida u rezultat izračunao sljedeće.

```
d_opazeni <- 0.78
power.t.test(n = 30, delta = d_opazeni, sd = 1,
             sig.level = 0.05, type = "two.sample")
```

Za zadane ulaze poziv vraća snagu približno 0,84. Uz ispis stoji obrazloženje.

> Opaženu standardiziranu razliku unijeli smo kao `delta`, a standardnu devijaciju postavili na jedan jer računamo na standardiziranoj skali. Dobivena snaga veća je od 0,80, pa potvrđuje da je završena studija bila dovoljno velika za naše pitanje.

Sastavite odgovor timu. Objasnite što račun od 0,84 vrijedi uz zadane ulaze, a što iz njega ne slijedi. Predložite vrijednost učinka za planiranje nove studije na temelju prethodno određenog cilja i navedite koji biste podatak iz završene analize još tražili za prosudbu preciznosti.

Razrađeni primjer

Planiranje unatrag sada možemo provesti za konkretnu uredničku odluku. Pretpostavimo da bi uredništvo reagiralo na razliku od pola boda, uz očekivanu raspršenost sličnu onoj u našoj populaciji. Ta je granica pretpostavka primjera koju bi u stvarnom istraživanju trebalo opravdati. Za nekoliko veličina

uzorka procijenit ćemo koliko bi često simulirana studija otkrila zadanu razliku.

Nulta hipoteza zadaje razliku sredina nula, a testna statistika dijeli opaženu razliku poznatom standardnom pogreškom. Pretpostavljamo dvije neovisne skupine, neovisne normalne ishode i jednaku poznatu standardnu devijaciju od 1,9 boda. Koristimo obostrani z-test s pragom 0,05. Za svaku veličinu uzorka simuliramo dvije tisuće parova skupina. Dobivena snaga vrijedi za te uvjete i zadanu razliku od pola boda.

Prva studija s po 100 opažanja počinje vrijednostima −0,45, 1,11, −1,65 u prvoj skupini i 2,36, −0,49, −2,33 u drugoj. To su generirani ishodi na centriranoj, neograničenoj normalnoj ljestvici, pa nisu doslovne ocjene od jedan do deset. Iz svih 100 vrijednosti po skupini dobivamo sredine 0,027 i 0,309. Njihovu razliku dijelimo s $1{,}9\sqrt{2/100}$ i uspoređujemo apsolutni rezultat s 1,96. Svako novo ponavljanje generira svih 200 ishoda iznova. Broj otkrivenih razlika dijelimo s 2 000, bez isključenih studija.

Za svaku ponuđenu veličinu uzorka ponavljamo studiju dvije tisuće puta. Granica 1,96 dolazi iz standardne normalne raspodjele. Pripada ovom obostranom z-testu, a ne permutacijskom testu iz prvog prikaza snage.

Tablica 5: Udio simuliranih studija u kojima bi razlika od pola boda bila otkrivena, pri četirima veličinama uzorka po skupini. Izrada autora.

Jedinica po skupini	Otkrivanja od 2 000	Snaga
100	934	46,7 %
150	1226	61,3 %
200	1533	76,6 %
250	1663	83,2 %

Za snagu od približno 80 % pri razlici od pola boda potrebno je između dvjesto i dvjesto pedeset osoba po skupini. Ako je izvedivo samo stotinu, istraživanje možemo provesti uz jasno ograničenje. Pod planskim pretpostavkama propustilo bi takvu razliku u više od polovice ponavljanja.

Račun povezuje planirani uzorak s učinkom koji želimo otkriti. Granicu od pola boda trebamo opravdati sadržajno, a očekivanu raspršenost ranijim mjerenjima. Simulacija ovdje ne uključuje odustajanje ispitanika ni probleme mjerenja, pa i njih treba razmotriti pri konačnom planiranju.

Sažetak

Da bismo planirali istraživanje, trebamo odrediti koliku bi razliku vrijedilo uočiti, jer statistička značajnost sama ne govori je li ona važna. Standardizacija pomaže usporediti veličine učinka na različitim ljestvicama, a njihovo sadržajno značenje prosuđujemo zasebno. Snaga povezuje zadani učinak, uzorak, raspršenost i prag, dok planiranje prema preciznosti pita koliko nam uzak interval treba. Velik uzorak može otkriti i vrlo malu razliku, a u našoj simulaciji maloga uzorka procjene odabrane prema značajnosti bile su u prosjeku dvostruko veće od stvarne razlike. Taj faktor opisuje odabir prema pragu u prikazanom scenariju i ne prenosimo ga na druga istraživanja. U poglavlju o

krizi i obnovi vidjet ćemo kako istraživačke odluke i objavljivanje dodatno oblikuju nalaze koje čitatelj uopće može vidjeti.

Pojmovi

veličina učinka (*effect size*), standardizirana razlika (*Cohen's d*), sadržajna važnost (*practical significance*), statistička snaga (*statistical power*), precjenjivanje pri odabiru značajnih nalaza (*selection-induced effect-size exaggeration*), najmanji važan učinak (*smallest effect size of interest*), planiranje veličine uzorka (*sample size planning*), širina intervala (*interval width*)

Zadaci

Konceptualni

Objasnite u jednom odlomku zašto u simulaciji ovoga poglavlja podskup procjena koje su prešle prag u prosjeku premašuje poznatu razliku, iako prosjek svih procjena ostaje blizu istini. Imenujte korak u kojem nastaje iskrivljenje i objasnite zašto se dobiveni faktor ne smije prenijeti na svako područje.

Računski

Iz upravljanoga agregata data/populacija-medija-agregat.csv uzmite retke za portal i tisak. U tiskanom ili dokumentnom izdanju iste su vrijednosti u tablici uz statični prikaz. Za svaki redak podijelite zbroj povjerenja brojem osoba, usporedite rezultat s pohranjenim prosjekom i izračunajte razliku prosjeka. Zatim za standardiziranu razliku 0,4 zabilježite snagu pri 40, 80, 160 i

300 jedinica po skupini. U tiskanom izdanju tablicu sa zadanom postavkom upotrijebite kao izvor vrijednosti.

Vratite se objašnjenju standardne pogreške i širine intervala u poglavlju o procjeni. Objasnite zašto veći uzorak istodobno sužava interval i povećava snagu, ali ne mijenja unaprijed zadanu veličinu učinka. Predajte dva računa prosjeka, njihovu razliku, četiri vrijednosti snage i jedan odlomak koji povezuje preciznost sa snagom. Ocjenjuje se račun i tumačenje, ne pisanje koda.

Kritički

Prosudite tvrdnju da niska prosječna snaga nekog područja znači samo da istraživanja propuštaju stvarne učinke (Button i sur. 2013.). Predajte kratku uredničku bilješku i navedite podatak koji bi vam trebao da procijenite koliko je objavljena veličina učinka precijenjena.

Revizija modela

Ocijenite provjeru iz okvira o pogrešci. Imenujte što je u pozivu ispravno, argument u kojem stoji kružnost, redak koda u kojem ona ulazi u račun, i napišite čime bi taj argument trebalo zamijeniti.

12

DIO IV · ZAKLJUČIVANJE

kriza i obnova

Kako se znanstveni zaključci mogu popraviti

Možemo li se oraspoložiti tako da najprije promijenimo izraz lica? Ta je zamisao privlačna jer nam daje nešto jednostavno što možemo pokušati u lošem danu. Za njezinu znanstvenu provjeru ipak treba smisliti kako promijeniti izraz lica, a da sudionicima ne sugeriramo kako bi se trebali osjećati.

U poznatom pokusu sudionici su ocjenjivali stripove držeći olovku zubima ili usnama. Razlika u ocjenama protumačena je kao potpora tvrdnji da položaj mišića lica može utjecati na doživljaj zabave (Wagenmakers i sur. 2016.). Gotovo tri desetljeća poslije 17 laboratorija ponovilo je postupak prema zajedničkom protokolu, s planom i analitičkim kodom pripremljenima prije uvida u podatke (Wagenmakers i sur. 2016.).

Što bismo trebali promijeniti u svojem uvjerenju ako takvo ponavljanje pruži drukčiju sliku? Možemo li iz njega naučiti nešto određenije od toga da je jedna studija bila „uspješna", a druga „neuspješna"?

Procjenu i njezin interval sada znamo čitati uz veličinu učinka, snagu i pogreške testiranja. Pri usporedbi izvornog nalaza s replikacijom otvara se i novo pitanje. Vidimo li sve što je istraživač mogao analizirati ili samo rezultat koji je na kraju odabran za objavu? Razloge odabira rezultata tražimo u zapisima istraživačkih odluka, uz provjeru samoga računa.

Pokus iz vinjete pratit ćemo kroz objavljene rezultate laboratorija koji su ponovili postupak (Wagenmakers i sur. 2016.). U tablicama s kojima radimo jedan redak sažima rezultate jednog laboratorija. Uspoređivat ćemo razlike u ocjenama zabavnosti između uvjeta držanja olovke, njihove intervale i brojnosti skupina. Ti sažeci pokazuju koliko se procjene razlikuju među stvarnim provedbama zajedničkog postupka (Wagenmakers i sur. 2016.). Od njih ćemo jasno odvojiti nastavne simulacije u kojima sami određujemo koliko se analiza iskušava.

Mijenjanje ishoda, podskupina ili trenutka zaustavljanja utječe na mogućnost lažnog nalaza, što možemo provjeriti simulacijom. Pritom razlikujemo namjerno traženje poželjnog rezultata od odluka koje ovise o viđenim podacima čak i kada istraživač provede samo jednu analizu. Izbor se nastavlja među studijama koje postaju dostupne čitateljima. Zato pogreške pojedinog testa razmatramo zajedno s postupkom kojim je od više mogućih nalaza odabrana objavljena tvrdnja.

Obnovu istraživačke prakse zato povezujemo s dostupnošću plana, podataka i analitičkih odluka. Pri čitanju laboratorijskih procjena oslanjamo se na intervale i veličine učinka iz prethodnih poglavlja, a u razrađenom primjeru uspoređujemo sirovu i standardiziranu sintezu. Doseg novoga rezultata određujemo prema onome što je točno ponovljeno, preciznosti procjena i razlikama u postupku koje još treba istražiti. Na toj osnovi procjenjujemo nalaz i predlažemo sljedeću provjeru, uz dovoljno jasan zapis da je druga osoba može ponoviti.

Od jedne tvrdnje do provjerljivoga pitanja

Istraživanje uključuje mnogo odluka prije završne tablice. Određujemo tko može sudjelovati, kako provodimo zadatak, što mjerimo, kako postupamo s nepotpunim odgovorima i koju usporedbu smatramo glavnom. Čitatelj treba znati koje smo odluke donijeli unaprijed, a koje nakon pregleda podataka. Inače ne može procijeniti koliko je rezultat utjecao na izbor analize.

p-hacking

P-hakiranje nastaje kada istraživač iskušava ishode, podskupine, pravila isključivanja ili trenutke zaustavljanja kako bi dobio poželjan nalaz, pa prikaže samo odabrani postupak. Metodološko istraživanje pokazalo je da takva neprikazana fleksibilnost povećava udio lažno pozitivnih zaključaka (Simmons i sur. 2011.). Uobičajena p-vrijednost vrijedi za unaprijed određen postupak. Nakon pretraživanja ne opisuje stopu pogreške cijelog procesa odabira.

garden of forking paths

Vrt račvajućih putova može nastati i bez svjesnog traženja poželjnog nalaza. Istraživač provede jednu analizu, ali bi odabrao drugu da su podaci izgledali drukčije. Odluka je zato ovisila o podacima čak i ako ostale mogućnosti nisu izračunate (Gelman i Loken 2013.). P-hakiranje opisuje namjerno ponašanje, a račvajući putovi obilježje postupka koje treba prepoznati i dokumentirati.

publication bias

Izbor ne završava analizom unutar jedne studije. *Publikacijska pristranost* nastaje pri odabiru onoga što postaje dostupno čitateljima. Čak i korektno provedene studije mogu dati iskrivljenu sliku ako objavljivanje ovisi o rezultatu. Open Science Collaboration pokazao je važnost sustavnog proučavanja replikabilnosti većeg skupa nalaza (Open Science Collaboration 2015.). Za razumijevanje literature trebamo razmotriti i pokušaje koji su ostali neobjavljeni.

Analitička fleksibilnost prije modela

U jednom je istraživanju 29 timova analiziralo iste početne podatke i dobilo različite procjene povezanosti boje kože nogometaša i crvenih kartona (Silberzahn i sur. 2018.). Razlike među rezultatima sada nastaju promjenom analize, dok smo pri uzorkovanju mijenjali uzorke. Uspoređujemo veličinu procijenjene veze, njezinu nesigurnost i opravdanost postupaka.

> **Analitička fleksibilnost** je prostor više obranjivih odluka duž cijeloga puta podataka, od prikupljanja, uključivanja i spajanja do rekodiranja, postupanja s nedostajućim vrijednostima i izbora analize.

U registriranom replikacijskom izvještaju (RRR) laboratoriji su unaprijed dogovorili kriterije uključivanja, računanje ocjene i glavnu sirovu razliku među skupinama. Plan i kod pripremili su prije stvarnih podataka (Wagenmakers i sur. 2016.). Lokalne prilagodbe i dalje su bile moguće, ali trebalo ih je zabilježiti i usporediti s planom. Tako čitatelj može provjeriti kada je odluka donesena i što je promijenila.

> **Reproducibilnost** je mogućnost da druga osoba iz istih podataka, materijala i opisanoga postupka ponovno dobije iste rezultate i utvrdi gdje su nastali.

replication

Pri reproduciranju ponovno izvodimo račun iz istih podataka i materijala. *Replikacija* jest nova studija koja istu ciljnu tvrdnju provjerava novim podacima i dovoljno usporedivim postupkom. Reproduciranjem možemo ponoviti i račun temeljen na lošem mjerenju. Replikacija pak može dati drukčiji rezultat zbog uzoračke promjenjivosti ili konteksta. Obje provjere pridonose prosudbi zaključka, zajedno s procjenom mjerenja, nacrta i konteksta.

Replikacija kao kumulativno mjerenje

Da bismo procijenili što smo naučili iz replikacije, trebamo usporediti rezultate laboratorija na istoj ljestvici. Služimo se podacima iz objavljenog izvještaja, u kojem svaki laboratorij daje procjenu razlike između dvaju načina držanja olovke i njezin interval (Wagenmakers i sur. 2016.). Laboratorij tako postaje jedinica usporedbe.

Svaki od 17 laboratorija daje procjenu i interval na istoj ljestvici (Wagenmakers i sur. 2016.). Devet procjena bilo je pozitivno, ali nijedan interval nije bio u cijelosti iznad nule (Wagenmakers i sur. 2016.). Samo dva intervala obuhvatila su izvornu razliku od 0,82 boda (Wagenmakers i sur. 2016.). Podjela na devet „uspješnih" i osam „neuspješnih" replikacija zanemarila bi veličinu i nesigurnost tih procjena.

Objedinjena procjena sažima prosječnu razliku u replikacijama provedenima prema zajedničkom protokolu. Vrijednost 0,03 uz interval od −0,11 do 0,16 uskladiva je s malim razlikama u oba smjera (Wagenmakers i sur. 2016.). Interval ne obuhvaća vrijednosti blizu izvornih 0,82 boda. To ograničava povjerenje u prijenos izvorne veličine na ovaj postupak, ali ne isključuje svaki mali učinak ni druge načine mjerenja facijalne povratne sprege (Wagenmakers i sur. 2016.).

Najprije zato čitamo istraživačko pitanje i nacrt, zatim pojedinačne procjene i intervale, pa objedinjeni rezultat. Razlike među laboratorijima mogu potjecati od uzorkovanja, lokalnog konteksta, mjerenja ili provedbe. Ovdje možemo opisati rezultate i oprezno ih prenijeti na usporedive postupke. Utvrđivanje uzroka razlika zahtijevalo bi dodatno istraživanje.

Cijeli zapis u forest plotu

forest plot

Procjene i intervale svih laboratorija lakše ćemo usporediti ako ih vidimo zajedno. Na prikazu *forest plot* svaki redak predstavlja jedan laboratorij. Točka označuje procijenjenu razliku, a vodoravna crta njezin 95-postotni interval. Puna okomita crta označuje nulu, odnosno jednake prosječne ocjene. Isprekidana crta na 0,82 prikazuje izvornu procjenu za usporedbu (Wagenmakers i sur. 2016.). Služi kao vanjska referenca uz laboratorije uključene u objedinjavanje.

Oznaka `smile` u izvornim podacima znači držanje olovke zubima, a `pout` držanje usnama (Wagenmakers i sur. 2016.). Pozitivna razlika `smile - pout` znači da je skupina s olovkom među zubima dala višu prosječnu ocjenu stripovima.

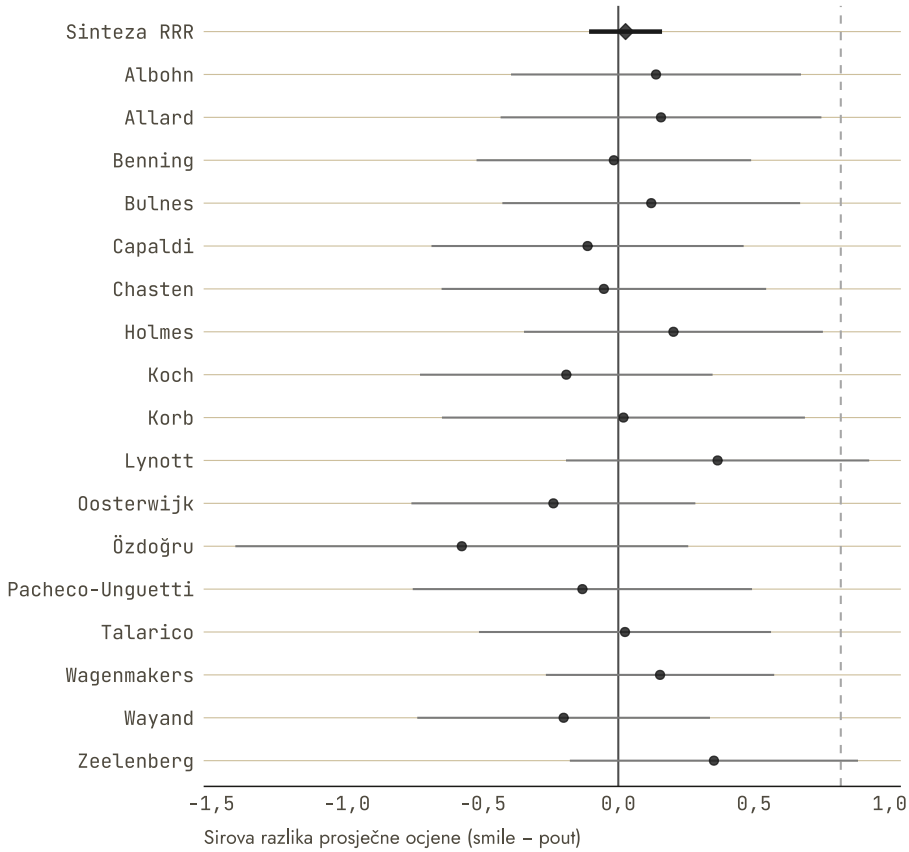
Pogledajmo prvo kako nastaje jedna točka. Tablica izdvaja prve tri laboratorijske usporedbe iz provjerenog izvornog poretka. Brojnosti su u osobama, a sredine na ljestvici ocjene stripova od 0 do 9 (Wagenmakers i sur. 2016.).

`Tablica 1:` Skupne sredine prije laboratorijske razlike. Izrada autora iz lokalne rekonstrukcije prema Wagenmakers i sur. (2016.).

Laboratorij	n zubi	n usne	Sredina zubi	Sredina usne	Razlika
Albohn	67	72	4,20	4,06	0,14
Allard	55	70	5,05	4,89	0,16
Benning	58	57	4,69	4,70	−0,02

U prvom laboratoriju od 4,201 oduzimamo 4,062 i dobivamo približno 0,139 boda (Wagenmakers i sur. 2016.). Razlika postaje položaj točke. Raspršenost ocjena i broj osoba u objema skupinama određuju njezinu standardnu pogrešku i interval. Za račun čuvamo nezaokružene sažetke. Razlika dviju prikazanih zaokruženih sredina zato može odstupati od zaokružene razlike.

Metaanaliza je objedinjavanje usporedivih rezultata više istraživanja. Ovdje odgovara na drugo pitanje, kolika je prosječna razlika kroz laboratorije. Njihove procjene množimo težinama i zbrajamo. Težine ovise o nesigurnosti pojedine procjene i procijenjenoj varijabilnosti među laboratorijima, pa svaki sudionik i svaki laboratorij nemaju automatski jednaku težinu. U razrađenom primjeru pokazat ćemo kako model slučajnih učinaka određuje te težine. U našoj rekonstrukciji sirove sinteze prvi laboratorij nosi 6,3 % ukupne težine (Wagenmakers i sur. 2016.). Svih 17 težina zbraja se na 100 %, a izvorni pokus iz 1988. nije među njima (Wagenmakers i sur. 2016.). Slika povezuje tako dobivene procjene i intervale.



Slika 1: Procjene sirove razlike smile – pout i 95-postotni intervali u 17 laboratorija. Objedinjena procjena RRR-a iznosi 0,03 [–0,11; 0,16], a isprekidana crta označuje izvornu procjenu 0,82, koja ne ulazi u sintezu. Izrada autora iz lokalne rekonstrukcije službenoga arhiva prema Wagenmakers i sur. (2016.).

Točke se nalaze oko nule i svi je intervali presijecaju (Wagenmakers i sur. 2016.). Objedinjena procjena ima uži interval jer koristi informacije iz više laboratorija. I dalje se odnosi na laboratorije i postupak uključene u ovaj RRR. Značenje prikaza možemo čitati iz položaja, duljine crta i oznake objedinjene procjene i u crno-bijelom tisku.

Razlike smo dosad izrazili u bodovima ocjene stripova. Iz prethodnog poglavlja znamo da ih možemo izraziti i u jedinicama standardne devijacije, pa želimo provjeriti ovisi li tumačenje o tom izboru. Provjera *osjetljivosti* uspoređuje sirovu razliku s Cohenovim *d*. U našoj standardiziranoj rekonstrukciji iz objavljenih laboratorijskih sažetaka procjena iznosi 0,01, uz interval od –0,08 do 0,10. Objavljena sirova procjena iznosila je 0,03, uz interval od –0,11 do 0,16 (Wagenmakers i sur. 2016.). Razlikujemo objavljeni rezultat od vlastitoga preračuna. Obje analize koriste iste laboratorije, ali standardizacija mijenja ciljanu veličinu i relativni doprinos laboratorija. Sadržajni zaključak ostaje sličan. Procjene su blizu nule i intervali obuhvaćaju male vrijednosti u oba smjera. Ovom usporedbom provjeravamo ovisi li zaključak o izboru mjerila.

sensitivity analysis

Reforma redoslijeda i vidljivosti

preregistration

Predregistracija jest vremenski označen zapis pitanja, ishoda i glavne analize prije uvida u podatke koji bi mogli utjecati na te odluke. Omogućuje razlikovanje planiranih analiza od onih potaknutih podacima. Neočekivane obrasce možemo istraživati, ali ih trebamo označiti kao naknadne, istraživačke analize. Potvrđujuća (konfirmatorna) analiza provjerava unaprijed određenu tvrdnju. Istraživačka (eksploratorna) analiza traži obrasce i oblikuje nova pitanja. U ovom RRR-u unaprijed su pripremljeni i kod i ogledni prikazi. Dobiveni rezultat zato možemo usporediti s postupkom koji je bio vidljiv prije podataka (Wagenmakers i sur. 2016.).

registered report

Registrirani izvještaj mijenja redoslijed uredničke procjene. Urednik i recenzenti razmatraju pitanje i metodu prije nego što znaju je li nalaz velik, nov ili značajan (Wagenmakers i sur. 2016.). U ovom slučaju povezani su službeni članak, projekt, plan, kod i podaci. To je omogućilo neovisnu provjeru laboratorijskih procjena (Wagenmakers i sur. 2016.).

Kad želimo da drugi provjere istraživanje, trebamo im pokazati kako smo došli do rezultata, a pritom sačuvati privatnost ljudi koji su u njemu sudjelovali i poštovati uvjete uporabe materijala. Za naše je usporedbe dovoljno imati rezultate pojedinih laboratorija. U knjigu smo zato uključili samo izvedene sažetke na toj razini, iz kojih se sudionici ne mogu prepoznati. Pojedinačne odgovore, službeni kod, plan i izdavačev graf nismo preuzeli. Uz prikaz razlikujemo što možemo provjeriti ovdje, a za što se trebamo vratiti službenom izvoru.

Tvrdnje razlikujemo prema tome opisuju li, povezuju, uopćavaju, predviđaju, objašnjavaju uzrok ili podupiru odluku. Ovaj RRR omogućuje opis i usporedbu eksperimentalnih uvjeta, ograničen prijenos na provedene postupke te procjenu težine izvornog nalaza. Ne daje alat za predviđanje ni objašnjenje uzroka razlike između izvornog i replikacijskog rezultata.

Prije prihvaćanja zaključka provjeravamo jedinicu opažanja, izostavljene jedinice, ciljanu veličinu i vrstu tvrdnje. Pitamo koju nesigurnost interval uključuje, mijenja li razumna alternativa zaključak i tko snosi posljedice pogreške. Reformirani postupci trebaju te odgovore učiniti dostupnima za provjeru, a njihovu kvalitetu zatim procjenjujemo na temelju priloženih dokaza.

Priča o izrazu lica nastavila se i nakon replikacije s olovkom. Suradnja Many Smiles okupila je istraživače različitih polazišta u unaprijed registriranom istraživanju s 3 878 sudionika iz 19 zemalja (Coles i sur. 2022.). Usporedili su više načina izazivanja izraza lica. Oponašanje osmijeha i namjerno pomicanje mišića lica pružili su potporu utjecaju na osjećaj sreće, dok su nalazi za olovku u ustima bili manje uvjerljivi (Coles i sur. 2022.). Različiti postupci nisu dali jednak odgovor.

Na početno pitanje sada možemo odgovoriti preciznije. Slabija potpora učinku pri ponavljanju traži promjenu uvjerenja o tom postupku i veličini njegova učinka. Drukčija manipulacija otvara zasebnu provjeru, s vlastitim ishodom i ograničenjima. Ocjena zabavnosti stripa i procjena vlastite sreće nisu ista mjera. Istraživanje napreduje kada tu razliku pretvorimo u jasno pitanje, unaprijed dogovorimo provjeru i dopustimo da nas rezultat iznenadi.

Pješčanik p-hakiranja

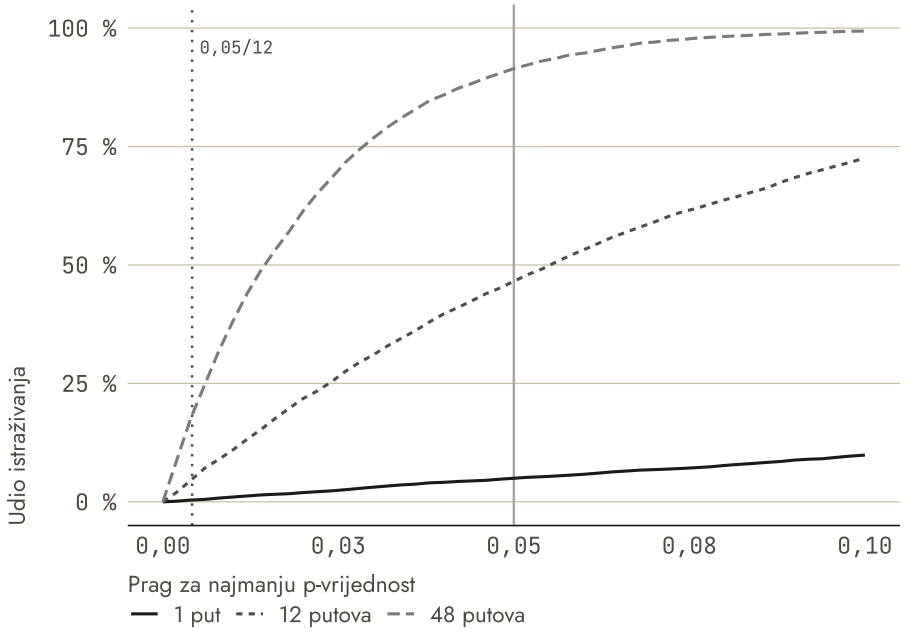
Sljedeći prikaz pokazuje što se događa kada, uz odsutnost učinka, povećavamo broj analitičkih putova. Više ishoda, podskupina ili trenutaka provjere daje više prilika za rezultat ispod praga. Usporedba s korigiranim pragom pokazuje kako uzeti u obzir cijelu obitelj analiza.

Za ovu zasebnu nastavnu simulaciju izravno generiramo p-vrijednosti. Pod idealiziranim nultim modelom svaki put daje neovisan broj jednoliko raspoređen između nule i jedinice. Jedno zamišljeno istraživanje s tri puta moglo bi dati 0,21, 0,03 i 0,54. To je namjerno sastavljen primjer. Ako prijavimo najmanji broj, 0,03 prolazi prag 0,05 iako u modelu učinka nema. Za unaprijed zadana tri puta korigirani prag bio bi $0{,}05/3$, koji taj broj ne prolazi.

Prikaz provodi isti odabir najmanje vrijednosti za 1, 12 i 48 putova. Za svaki broj putova generiramo 5 000 cijelih istraživanja. Na svakom pragu brojimo istraživanja čija je najmanja p-vrijednost ispod ili na pragu i dijelimo s 5 000. Svi rezultati su definirani i ulaze u račun. Visina krivulje zato prikazuje rizik barem jednoga lažnog nalaza po istraživanju, a ne postotak pogrešnih testova među svim provedenim testovima.

Simulacija pretpostavlja neovisne putove, pravilno kalibrirane p-vrijednosti pod nultim modelom i unaprijed određen skup analiza. Stvarne analize često koriste iste podatke, pa krivulja

može izgledati drukčije. Korekcija za poznati skup analiza ne rješava skrivene odluke ni neprijavljeno prilagođavanje postupka podacima.



Slika 2: Vjerojatnost barem jednog rezultata ispod praga raste s brojem neovisnih analitičkih putova. Puna okomita crta označuje nominalni prag 0,05, a točkasta prag 0,05/12 za zadano stanje od 12 putova. Izrada autora. Podaci i postupak opisani su u tekstu.

Ako unaprijed određeni broj putova označimo s m, prag $0{,}05/m$ dijeli dopušteni ukupni rizik među njima. Svaka analiza mora zadovoljiti stroži prag. To je Bonferronijeva korekcija. Razinu 0,05 dijelimo brojem usporedbi u unaprijed određenoj obitelji. O obitelji testova više ćemo govoriti u poglavlju Više grupa. Takva korekcija ograničava rizik lažnih nalaza u unaprijed određenom skupu analiza, ali ne uklanja sve probleme analitičke fleksibilnosti.

Na pragu 0,05 očitajte visinu krivulje za jedan put. Na istome pragu usporedite krivulje za 12 i 48 putova. Na krivulji za 12 putova usporedite prag 0,05 s točkastim pragom 0,05/12. Objasnite što prikaz ne može reći kada analitički putovi dijele podatke.

Kada odaberemo najmanju p-vrijednost nakon pretraživanja, trebamo opisati kako smo do nje došli. Tumačenje mora uzeti u obzir cijeli postupak odabira i moguće analize.

STATISTIKA U DIVLJINI

Sedamnaest procjena nije referendum. Nakon ponavljanja pokusa s olovkom i procjenom duhovitosti crteža trebalo je odlučiti što zajednički rezultati govore o izvornom nalazu. Prema izvještaju o replikaciji, devet je laboratorija dalo pozitivnu, a osam negativnu točkastu procjenu (Wagenmakers i sur. 2016.). Kad bismo prebrojili samo predznake, mogli bismo zaključiti da su dokazi gotovo ravnomjerno podijeljeni.

Takvo prebrojavanje svakoj procjeni daje ulogu glasa za ili protiv. Mala razlika neposredno iznad nule tada vrijedi jednako kao velika, precizno procijenjena razlika. Nestaju i veličina i nesigurnost nalaza, upravo informacije zbog kojih je ponavljanje više laboratorija korisno.

U ovom slučaju nijedan laboratorijski interval nije bio u cijelosti iznad nule, a samo su dva obuhvaćala izvornu razliku od 0,82 boda (Wagenmakers i sur. 2016.). Objedinjena procjena iznosila je 0,03 boda, uz 95-postotni interval od −0,11 do 0,16 (Wagenmakers i sur. 2016.). Te brojke govore nešto određenije od raspodjele pozitivnih i negativnih predznaka. Pod uvjetima zajedničkog postupka ne podupiru ponavljanje izvorno velike razlike, a ostavljaju nesigurnost oko malih razlika u oba smjera.

Za istraživača koji odlučuje na čemu graditi sljedeći pokus to znači da treba promijeniti očekivanje o veličini nalaza. Polaznu vrijednost za planiranje treba procijeniti u svjetlu novih, znatno manjih procjena. Rezultat pritom pripada određenoj manipulaciji, mjerenju i protokolu. Drugi načini povezivanja izraza lica s doživljajem ostaju predmet zasebnih provjera.

Pojedinačne procjene zato čitamo uz objedinjenu, a razlike među laboratorijima ne pretvaramo u objašnjenja bez dodatnih dokaza. Uži zajednički interval nije prognoza točne vrijednosti sljedeće replikacije. Korist ponavljanja ovdje je u tome što sužava raspon uvjerljivih tvrdnji. Umjesto pitanja koliko je laboratorija glasovalo za učinak, možemo pitati koliko velik učinak njihov zajednički rad još dopušta.

Unaprijed vidljiv plan omogućuje nam provjeru istraživačkih odluka. U sljedećem obrazloženju procijenimo što na temelju samog postojanja takva plana možemo zaključiti o valjanosti nalaza.

NAĐITE GREŠKU

Za bilješku o ponavljanju pokusa s izrazom lica asistent je izdvojio sljedeće podatke. RRR je unaprijed odredio glavni ishod, pravila uključivanja i primarnu analizu (Wagenmakers i sur. 2016.). Objedinjena razlika u ocjenama stripova iznosi 0,03 boda, uz 95-postotni interval od −0,11 do 0,16 (Wagenmakers i sur. 2016.).

Urednički zapis navodi da su plan i završni izvještaj dostupni, ali ih urednik još nije usporedio. To je zamišljena situacija pregleda ovoga rada, a ne tvrdnja o propustu njegovih autora. Asistent predlaže bilješku.

> Glavni ishod i analiza određeni su prije uvida u rezultate. Objedinjena procjena blizu je nule, uz interval koji dopušta male razlike u oba smjera. Predregistracija je zato dovoljan razlog da prihvatimo zaključak bez usporedbe provedene analize s planom.

Napišite uredniku što iz bilješke može zadržati, a koji korak pregleda još treba obaviti. Predložite jednu konkretnu usporedbu plana i završnog izvještaja te napišite zamjenu posljednje rečenice. Ne pretpostavljajte da provjera mora otkriti odstupanje.

Razrađeni primjer

Dosadašnji zaključak počiva na razlici izraženoj u bodovima ocjene stripova. Provjerit ćemo ostaje li sadržajno sličan kada iste laboratorijske rezultate izrazimo standardiziranom razlikom. Prije te usporedbe trebamo potvrditi da obje analize polaze od istih provjerenih zapisa. Službeni arhiv sadrži sudioničke

zapise iz 17 laboratorija, koje knjiga ne redistribuira (Wagenmakers i sur. 2016.). Iz njih je neovisnom provjerom izvedena tablica grupnih sažetaka, procjena i intervala. Redci se zbrajaju na 1 894 sudionika i odgovaraju objavljenim sažetcima, uz jedno odstupanje u zaokruživanju standardne devijacije koje ne koristimo u tvrdnjama knjige (Wagenmakers i sur. 2016.).

Tablicu možemo provjeriti bez pisanja koda. Svaki redak predstavlja laboratorij, a `n` broj uključenih sudionika. Interval pripada sirovoj razlici prosječnih ocjena između skupina koje su olovku držale zubima (`smile`) i usnama (`pout`).

Tablica 2: Sirove laboratorijske procjene iz RRR-a. Izrada autora iz lokalne rekonstrukcije službenoga arhiva prema Wagenmakers i sur. (2016.).

Laboratorij	n	Razlika	95-postotni interval
Albohn	139	0,14	[−0,40; 0,67]
Allard	125	0,16	[−0,43; 0,75]
Benning	115	−0,02	[−0,52; 0,49]
Bulnes	101	0,12	[−0,43; 0,67]
Capaldi	117	−0,11	[−0,69; 0,46]
Chasten	94	−0,05	[−0,65; 0,54]
Holmes	99	0,20	[−0,35; 0,75]
Koch	100	−0,19	[−0,73; 0,35]
Korb	101	0,02	[−0,65; 0,69]
Lynott	126	0,37	[−0,19; 0,92]
Oosterwijk	110	−0,24	[−0,76; 0,28]
Özdoğru	87	−0,58	[−1,41; 0,26]
Pacheco-Unguetti	120	−0,13	[−0,76; 0,49]
Talarico	112	0,02	[−0,51; 0,56]
Wagenmakers	130	0,15	[−0,27; 0,57]
Wayand	110	−0,20	[−0,74; 0,34]
Zeelenberg	108	0,35	[−0,18; 0,88]

Pri standardizaciji svaku laboratorijsku razliku dijelimo združenom standardnom devijacijom dviju skupina u tom laboratoriju. U prvom laboratoriju 0,139 boda tako postaje $d = 0{,}087$. Mijenjaju se mjerna jedinica, standardna pogreška i težina, a sudionici i smjer oduzimanja ostaju isti. U obje analize zadržavamo svih 17 laboratorija.

Za objedinjavanje koristimo model slučajnih učinaka. Težina laboratorija prije normalizacije iznosi $w_i = 1/(SE_i^2 + \tau^2)$, gdje SE_i opisuje nesigurnost njegove razlike, a τ^2 procijenjenu varijancu razlika među laboratorijima. Postupak REML procjenjuje tu drugu varijancu iz svih laboratorijskih sažetaka. Veći laboratorij zato ne mora dobiti težinu razmjernu samo broju sudionika. Objavljena sirova analiza i naša standardizirana rekonstrukcija imaju različite jedinice i težine.

Sljedeća tablica prikazuje obje rekonstrukcije iz istih 17 laboratorija. Intervali su normalni Waldovi intervali. Procjeni dodajemo i od nje oduzimamo $1{,}96\sqrt{1/\sum_i w_i}$. Taj račun tretira procijenjenu među-laboratorijsku varijancu kao zadanu, pa ne obuhvaća svu njezinu nesigurnost. U ovome poglavlju učimo čitati sintezu. Procjena REML-a nije zadatak za ručno računanje.

Tablica 3: Autorska rekonstrukcija dviju sinteza iz laboratorijskih sažetaka prema Wagenmakers i sur. (2016.). Standardizirani redak nije zasebno objavljeni nalaz.

Mjera	Procjena	Donja granica	Gornja granica
Sirova razlika	0,027	−0,108	0,161
Standardizirana razlika	0,014	−0,076	0,104

Sirova procjena iznosi 0,03, uz interval od −0,11 do 0,16. Standardizirana rekonstrukcija iznosi 0,01, uz interval od −0,08 do 0,10. U laboratorijskoj tablici ima 9 pozitivnih procjena, 0 intervala u cijelosti iznad nule i 2 intervala koja obuhvaćaju 0,82. Ta brojanja opisuju raspored procjena. Zaključak temeljimo na njihovoj veličini i nesigurnosti.

Prema zajedničkom protokolu prosječna razlika u replikacijama bila je blizu nule i mnogo manja od izvornih 0,82 boda (Wagenmakers i sur. 2016.). Sirovi i standardizirani interval iz naše rekonstrukcije obuhvaćaju male razlike u oba smjera. Zaključak se odnosi na ispitani način mjerenja facijalne povratne sprege. Uzroci razlika među procjenama i eventualni utjecaj predregistracije zahtijevali bi zasebne usporedbe osmišljene upravo za ta pitanja.

Takav zaključak povezuje opaženi rezultat s provedenim postupkom. Pri odluci o težini izvornog nalaza uzimamo u obzir obje analize, uz navedene granice prijenosa.

Od testa do provjerljivog zaključka

Sada možemo zajedno razmotriti nulti model, veličinu učinka, interval, posljedice pogreške i način odabira analize. Dokumentirani postupak omogućuje provjeru tih dijelova. I dalje trebamo posebno opravdati uzročni zaključak, predviđanje ili prijenos na drugo okruženje.

U tablici su odgovori na šest pitanja za višelaboratorijski RRR i njegove dvije analize (Wagenmakers i sur. 2016.). Povezuju rezultat s jedinicom, postupkom i dopuštenim zaključkom.

Tablica 4: Šest revizijskih pitanja primijenjenih na provjereni RRR. Izrada autora prema Wagenmakers i sur. (2016.).

Revizijsko pitanje	Primjena na provjereni zapis
Koja je jedinica?	zapis za jedan laboratorij sa sažetkom razlike između eksperimentalnih uvjeta i njezine nesigurnosti
Tko ili što nedostaje?	ljudi izvan uključenih uzoraka, drugi laboratoriji i drukčije operacionalizacije facijalne povratne sprege
Koji je cilj tvrdnje?	prosječna razlika ocjena između uvjeta u zajedničkom postupku, opisana sirovom i standardiziranom sintezom
Gdje je neizvjesnost?	preciznost laboratorijskih i objedinjenih procjena. Ne obuhvaćaju sve probleme mjerenja, odabira i provedbe
Koje drugo objašnjenje ostaje?	standardizacija mijenja ciljanu veličinu i doprinos laboratorija, pa rezultate uspoređujemo sadržajno
Tko snosi posljedice?	istraživači, urednici i čitatelji koji procjenjuju koliko vjerovati izvornom nalazu

Karta tvrdnji u nastavku pokazuje što iz ove analize možemo prenijeti i gdje su granice pojedine tvrdnje.

Tablica 5: Šest dimenzija tvrdnje na granici Dijela IV. Izrada autora prema Wagenmakers i sur. (2016.).

Dimenzija tvrdnje	Što ovaj dokaz dopušta
opis	opisuje laboratorijske procjene, intervale i dvije objedinjene analitičke grane
povezanost	povezuje prosječnu ocjenu s eksperimentalnim uvjetom u provedenom postupku
generalizacija	oprezno prenosi zaključak na usporedive provedene postupke, ne na svaku populaciju ili operacionalizaciju
predviđanje	nije poduprto jer analiza ne gradi ni provjerava pravilo za novi laboratorij ili novoga sudionika
uzročnost	ne podupire uzročno objašnjenje razlike između izvorne i replikacijskih procjena
odluka	podupire opreznije odmjeravanje težine izvorne procjene uz obje grane i njihove intervale

Na kraju dijela provjerite četiri povezane teme. Zašto p-vrijednost ne daje vjerojatnost nulte hipoteze, a nalaz blizu nule ne dokazuje njezinu istinitost? Kako čitamo učinak i interval te zašto opaženi učinak nije neovisan ulaz za naknadnu provjeru snage? Kako pretraživanje analiza mijenja tumačenje najmanje p-vrijednosti? Koji dokumenti omogućuju provjeru rezultata i gdje bi zaključak prekoračio doseg nacrta?

U idućim analizama uvijek ćemo navesti ciljanu tvrdnju, primarni postupak, jednu razumnu alternativu, podrijetlo podataka i ograničenja. Počinjemo kategoričkim tablicama. Prije hi-kvadrat testa trebamo utvrditi koga smo prebrojili, što predstavlja ćelija, koji je nazivnik svakog postotka i kakvu razliku smatramo sadržajno važnom.

Sažetak

Nalaz možemo procijeniti tek kada uz završni račun vidimo kako su podaci prikupljeni, pripremljeni i analizirani te kako je rezultat odabran za objavu. P-hakiranje i račvajući putovi upozoravaju na različite oblike ovisnosti analize o podacima, zbog kojih je važan zapis trenutka i razloga odluke. Reproduciranjem provjeravamo račun na istom materijalu, a replikacijom prikupljamo nove podatke za usporedbu. U višelaboratorijskom primjeru čitanje procjena i intervala dalo je više od brojanja predznaka, a sirova i standardizirana analiza poduprle su sličan, ograničen zaključak o razlici blizu nule. Predregistracija, registrirani izvještaji i otvoreni materijali čine takvu provjeru dostupnijom, iako sami ne jamče dobro mjerenje ni valjan zaključak. U sljedećim analizama zato ćemo uz metodu pratiti i odluke koje su nas do nje dovele.

Pojmovi

potvrđujuća analiza (*confirmatory analysis*), istraživačka analiza (*exploratory analysis*), p-hakiranje (*p-hacking*), vrt račvajućih putova (*garden of forking paths*), publikacijska pristranost (*publication bias*), analitička fleksibilnost (*analytical flexibility*), reproducibilnost (*reproducibility*), replikacija (*replication*), forest plot (*fo-*

rest plot), provjera osjetljivosti (*sensitivity analysis*), predregistracija (*preregistration*), registrirani izvještaj (*registered report*)

Zadaci

Konceptualni

Istraživač je nakon pregleda podataka iskušao pet pravila isključivanja i objavio jedino ono koje je dalo poželjan rezultat. Druga je istraživačica prije analize odlučila „upotrijebit ću medijan ako je raspodjela jako asimetrična, a sredinu inače" i zatim provela samo odabranu granu. Objasnite koji primjer prikazuje p-hakiranje, a koji vrt račvajućih putova. Za svaki navedite koji bi trag postupka čitatelju omogućio provjeru i pokažite gdje se u primjeru nalazi analitička fleksibilnost (Simmons i sur. 2011.; Gelman i Loken 2013.). Na kraju jednom rečenicom razlikujte reproducibilnost takve provjere od replikacije koja bi prikupila nove podatke.

Računski

Koristite samo tiskanu tablicu laboratorijskih procjena, bez pisanja koda. Prebrojite koliko je od 17 točkastih procjena veće od nule i izračunajte taj udio u postocima (Wagenmakers i sur. 2016.). Zatim prebrojite koliko je 95-postotnih intervala u cijelosti iznad nule i koliko ih obuhvaća izvornu procjenu 0,82 (Wagenmakers i sur. 2016.). Napišite dvije rečenice koje objašnjavaju zašto prvi udio nije stopa „uspjelih replikacija".

Kritički

Vratite se definiciji intervala iz poglavlja Procjena i logici nultoga modela iz poglavlja Logika testiranja. Forest plot temelji se na objavljenoj replikaciji (Wagenmakers i sur. 2016.). Zamislite

sljedeći pogrešan komentar. „Niti jedan laboratorij nije dobio značajan rezultat, pa je dokazano da učinak ne postoji.” Revidirajte tu tvrdnju u tri rečenice. Uključite objedinjenu procjenu s intervalom, objasnite što nula u prikazu znači, usporedite primarnu analizu sa standardiziranom granom provjere osjetljivosti te navedite jednu populacijsku ili kontekstualnu granicu koju forest plot ne uklanja. Dodajte jednu rečenicu o tome kako bi publikacijska pristranost promijenila tumačenje sinteze kada nevidljive studije ovise o rezultatu.

Revizija modela

Revidirajte analizu iz okvira o pogrešci. Predajte zapis s četiri polja. U prvome prepišite provjerene brojke, u drugome označite jedinu pogrešnu tvrdnju, u trećemu napišite ispravljeni zaključak, a u četvrtome navedite koji biste dokument ili redak postupka provjerili prije prihvaćanja analize i objasnite koju dodatnu vidljivost daje registrirani izvještaj. Ne traži se novi kod.



DIO V

modeli

Ista pitanja o procjeni, nesigurnosti i dosegu primjenjujemo na tablice kategorija, usporedbe skupina i regresiju. Modelom povezujemo više obilježja, a pri predviđanju provjeravamo kako model radi na novim podacima i kakve posljedice odluke imaju za ljude.

13

DIO V · MODELI

kategorički podaci

Od frekvencija u ćelijama do obrazaca povezanosti

Sumnja u pravednost Berkeleyjskih upisa tražila je usporedbu prijava i ishoda za muškarce i žene (Bickel i sur. 1975.). Zbirne i odjelne stope otvorile su različita pitanja, iako su nastale iz istih zapisa. Izbor usporedbe određivao je što se iz tablice moglo zaključiti.

Kad jednom odaberemo usporedbu, preostaje procijeniti koliko se opaženi raspored razlikuje od onoga koji bismo očekivali bez povezanosti. Bi li sam nalaz povezanosti bio dovoljan za zaključak o pravednosti upisa?

Isti zapisi prijava i odluka o prijmu na Sveučilište Berkeley, prema spolu i odjelu, pokazali su koliko tumačenje postotka ovisi o nazivniku i okolnostima nastanka podataka (Bickel i sur. 1975.). Za analizu kategorija trebamo povezati taj uvid s opisom

raspodjele i povezanosti, procjenom nesigurnosti te razlikom između statističke značajnosti i veličine učinka.

Takve podatke često susrećemo u društvenim znanostima. Zanima nas koji izvor vijesti osoba bira, kojoj obrazovnoj skupini pripada ili podržava li neki prijedlog. Brojanjem dobivamo polazište za usporedbe koje odgovaraju na istraživačko pitanje. Ako mlađi i stariji biraju različite izvore vijesti, kako ćemo opisati tu razliku i procijeniti što nam uzorak govori o populaciji?

Logika zaključivanja koju smo dosad usvojili ostaje ista. Najprije postavljamo nulti model u kojem dvije varijable nisu povezane. Zatim uspoređujemo opažene podatke s onime što bismo pod tim modelom očekivali. Ovdje ćemo tu usporedbu provesti na frekvencijama u kontingencijskoj tablici. Uz rezultat testa i dalje trebamo pokazati veličinu razlike, njezinu nesigurnost i nazivnike na kojima se temelje prikazani udjeli.

Usporedbom tiska i portala procjenjivali smo razliku u povjerenju i njezinu nesigurnost. Sada nas zanima kako se sam izbor izvora vijesti razlikuje s dobi.

Za tu usporedbu zasebno izvlačimo uzorak od 800 osoba iz poznate nastavne populacije `populacija_medija`. Svaki zapis pripada zamišljenoj osobi s dobnom skupinom i glavnim izvorom vijesti. Nalaz ostaje opis računalno stvorene populacije, bez zaključka o stvarnim stanovnicima.

Najprije ćemo prebrojiti odgovore i usporediti postotke unutar skupina. Potom ćemo izračunati očekivane frekvencije pod pretpostavkom nezavisnosti i hi-kvadrat testom procijeniti ukupno odstupanje od toga modela. Reziduali će nam pomoći pronaći ćelije s izraženim odstupanjima, a Cramérovo V opisati jačinu povezanosti. Razmotrit ćemo i kako se analiza mijenja kada raspodjelu uspoređujemo s unaprijed zadanim udjelima ili kada su očekivani brojevi u ćelijama mali.

Izbor kategorija i usporedbe oblikuje zaključak. Pri opisu povezanosti dobi i izvora vijesti zato ćemo navesti i što nam još treba za njezino objašnjenje.

Brojanje prije testiranja

Kategoričke varijable razvrstavaju jedinice u skupine. Dobna skupina, izvor vijesti, obrazovanje i regija primjeri su takvih varijabli. Analiza počinje brojanjem jedinica u svakoj kategoriji, pa prvo trebamo provjeriti što smo i koga prebrojili.

Frekvencija je broj jedinica u kategoriji. Za udio trebamo i **uvjetni nazivnik**, skupinu unutar koje računamo postotak. Tvrdnja da četrdeset posto ispitanika bira društvene mreže može se odnositi na cijeli uzorak, samo najmlađu dobnu skupinu ili samo osobe koje su odgovorile. Jasno naveden nazivnik omogućuje čitatelju da provjeri na koju se skupinu postotak odnosi.

Prije brojanja određujemo kategorije. Dob možemo podijeliti na različite skupine, a portal i društvene mreže promatrati zajedno ili odvojeno. To je dio operacionalizacije. Test zatim uspoređuje raspored unutar tih kategorija. Njihovu sadržajnu prikladnost obrazlažemo prema pojavi koju istražujemo.

> Kontingencijska tablica je tablica frekvencija u kojoj svaka ćelija sadrži broj jedinica koje istovremeno pripadaju jednoj kategoriji retka i jednoj kategoriji stupca.

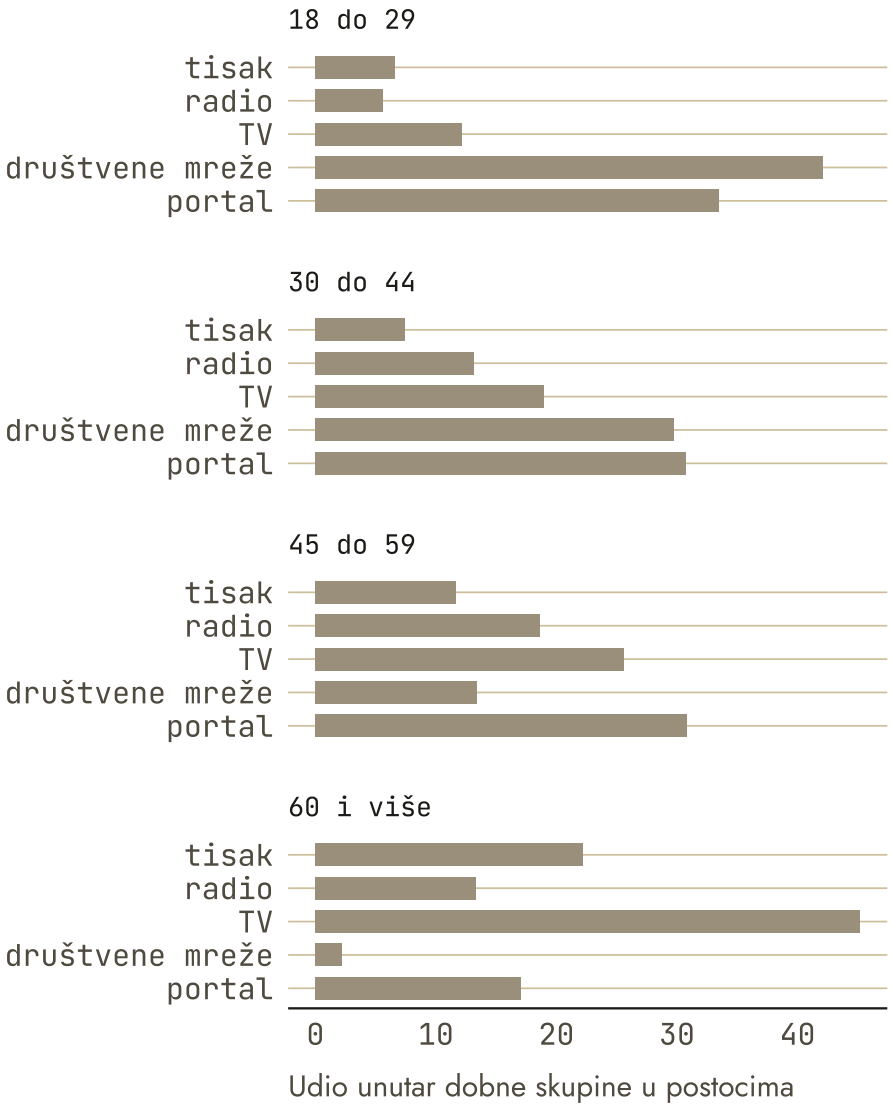
U našoj će tablici retci predstavljati dobne skupine, a stupci izvore vijesti. Ćelija na sjecištu najmlađe skupine i televizije pokazat će koliko najmlađih bira televiziju. Zbroj cijelog retka obuhvatit će sve osobe te dobi, neovisno o njihovu izvoru, a zbroj stupca sve korisnike televizije, neovisno o dobi. Takvi rubni zbrojevi daju raspodjelu svake varijable zasebno, dok unutarnje ćelije čuvaju njihove kombinacije.

Prijelaz od zapisa do tablice možemo provjeriti na najmlađoj skupini. Svaka od njezinih 197 osoba ulazi u točno jednu ćeliju. Među njima 83 bira društvene mreže, pa njihov udio računamo kao 83 podijeljeno s 197. Ostale retke gradimo istim prebrojavanjem.

Tablica 1: Dob i glavni izvor vijesti u istom simuliranom uzorku. Izrada autora.

Dob	Portal	Mreže	TV	Radio	Tisak	Ukupno
18 do 29	66	83	24	11	13	197
30 do 44	91	88	56	39	22	296
45 do 59	53	23	44	32	20	172
60 i više	23	3	61	18	30	135
Ukupno	233	197	185	100	85	800

Postoci po retku pokazuju kako su izvori vijesti raspoređeni unutar dobne skupine. Postoci po stupcu pokazuju dobni sastav korisnika pojedinog izvora. Ovdje koristimo postotke po retku jer želimo usporediti izbor izvora među dobnim skupinama različitih veličina. Kad bi nas umjesto toga zanimalo koji dio televizijske publike čine najstariji, koji bismo nazivnik trebali odabrati? Sada bismo broj najstarijih korisnika televizije dijelili brojem svih korisnika televizije. Promijenilo se pitanje, pa isti zapisi traže postotak po stupcu. Duljine stupaca na sljedećoj slici ipak prikazuju prvo pitanje, udjele izvora unutar svake dobne skupine.



Slika 1: Udio pojedinog izvora vijesti unutar svake dobne skupine u simuliranom uzorku od osamsto osoba. Raspodjele izvora razlikuju se među dobnim skupinama. Udjeli se unutar svake skupine zbrajaju do 100 %. Izrada autora.

Društvene mreže bira 42,1 % osoba od 18 do 29 godina, a 2,2 % osoba od 60 i više godina. Televiziju bira 12,2 % najmlađih i 45,2 % najstarijih. Obrazac je jasan u uzorku. Sada trebamo provjeriti koliko bi takav raspored bio neobičan kada u populaciji ne bi bilo veze.

Očekivano pod nezavisnošću

Opažene razlike među dobnim skupinama sada trebamo usporediti s rasporedom u kojem dobi i izvoru vijesti ne bismo pripisali povezanost. Pritom zadržavamo broj osoba u svakoj dobnoj skupini i ukupan broj korisnika svakog izvora. Iz tih rubnih zbrojeva možemo izračunati očekivane frekvencije pod pretpostavkom nezavisnosti.

Ako televiziju bira petina cijelog uzorka, pod nezavisnošću bismo očekivali isti udio u svakoj dobnoj skupini. Očekivani broj u ćeliji dobivamo primjenom udjela stupca na broj osoba u retku.

> **Očekivana frekvencija** ćelije je umnožak zbroja njezina retka i zbroja njezina stupca podijeljen ukupnim brojem jedinica, dakle broj koji bi se u toj ćeliji nalazio kad bi dvije varijable bile nezavisne uz nepromijenjene rubne zbrojeve.
> Oznaka E_{ij} predstavlja očekivanu frekvenciju u i-tom retku i j-tom stupcu. Oznaka $n_{i\cdot}$ predstavlja zbroj i-tog retka, $n_{\cdot j}$ zbroj j-tog stupca, a n ukupan broj jedinica.
>
> $$E_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{n}$$

U istoj tablici televiziju bira ukupno 185 od 800 osoba. Za najmlađe zato očekujemo $197 \times 185/800 = 45{,}56$ korisnika televizije, a opažamo ih 24. Očekivani broj smije imati decimale jer nastaje iz modela kao referenca za usporedbu s opaženim brojem osoba.

Prije nastavka možemo dopuniti isti račun za najstariju skupinu. U njoj je 135 osoba, a zajednički udio korisnika televizije i dalje iznosi 185 podijeljeno s 800. Koji broj dobivamo kada taj udio primijenimo na veličinu najstarije skupine i zašto u nazivnik ne stavljamo samo broj najstarijih? Vrijedi zapisati račun prije čitanja objašnjenja u sljedećem odlomku.

Dobivamo 31,22 očekivanih korisnika televizije. Udio iz cijelog uzorka određuje referencu pod nezavisnošću, a broj najstarijih određuje na koliko osoba primjenjujemo tu referencu. Kad bismo umjesto toga uzeli opaženi udio televizije među najstarijima, samo bismo rekonstruirali opaženi broj i izgubili usporedbu s nultim modelom.

Nezavisnost znači da se raspodjela izvora vijesti ne mijenja s dobnom skupinom. To privremeno pretpostavljamo kako bismo izračunali očekivanja i usporedili ih s podacima.

Najmanja očekivana frekvencija iznosi 14,3. U našoj tablici nema vrlo rijetkih očekivanih ćelija. Očekivana frekvencija manja od pet uobičajen je znak za dodatnu provjeru hi-kvadrat aproksimacije (Cochran 1954.). Taj prag služi kao upozorenje, a ne kao jamstvo valjanosti ili najava smjera pogreške. Poslije ćemo provjeriti kako se postupak ponaša uz rijetke ćelije.

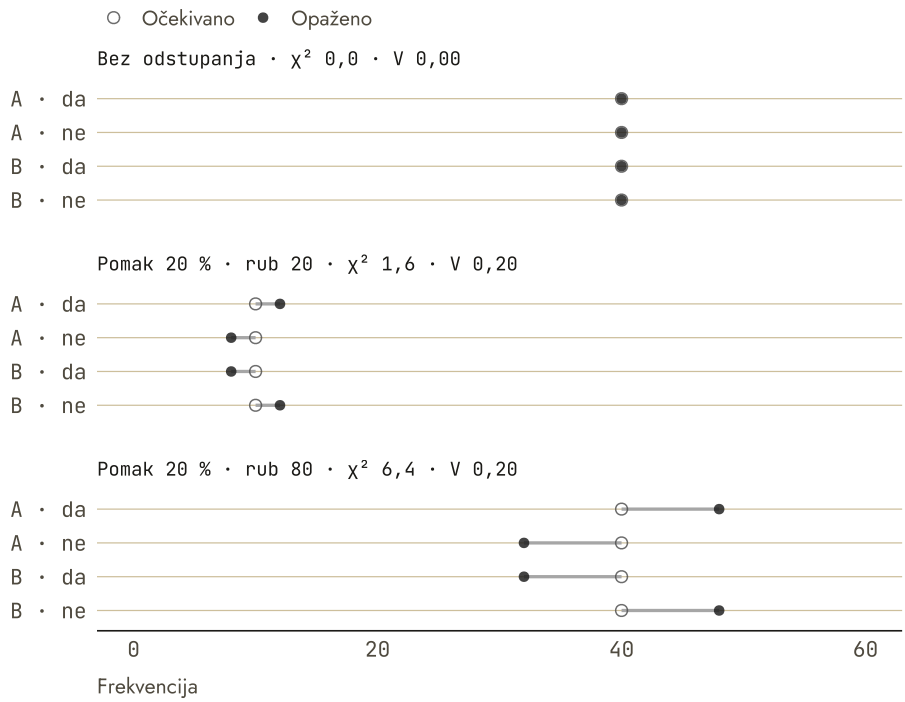
Za sljedeći prikaz trebamo dvije sažete mjere. Hi-kvadrat statistika zbraja doprinose $(O - E)^2/E$ svih ćelija, pa veći broj označuje veće odstupanje od nezavisnosti uz zadane rubove. Cramérovo V dijeli tu statistiku veličinom uzorka i dimenzijom tablice prije vađenja korijena. Za tablicu dva puta dva vrijedi $V = \sqrt{\chi^2/n}$. Tako razdvajamo količinu dokaza od veličine povezanosti. Nakon prikaza razvijamo opće formule i njihove pretpostavke.

Očekivano i opaženo

Za ovaj prikaz konstruiramo novu tablicu dviju skupina A i B s odgovorima „da” i „ne”. Ona ne nastaje izvlačenjem osoba iz prethodnog uzorka. Svaka skupina ima zadani broj osoba, a ukupno je jednako odgovora „da” i „ne”. Ako svaka skupina ima 20 osoba, nezavisnost daje po 10 u svakoj ćeliji. Pomak od 20 % dodaje dva odgovora ćelijama A–da i B–ne, a oduzima dva preostalima. Dobivamo retke 12 i 8 te 8 i 12, uz nepromijenjene rubne zbrojeve. Njihov doprinos računamo kao $(12 - 10)^2/10 = 0{,}4$ za prvu

ćeliju, a isto pravilo vrijedi za ostale tri. Zbroj je $4 \times 0{,}4 = 1{,}6$, uz $n = 40$, pa je $V = \sqrt{1{,}6/40} = 0{,}20$. Ako sve frekvencije pomnožimo s četiri, doprinos svake ćelije i hi-kvadrat postaju četiri puta veći, dok V ostaje 0,20. Prikaz uspoređuje unaprijed zadane tablice i ne procjenjuje stopu pogreške kroz slučajne uzorke.

Najprije u svakom retku uspoređujemo opaženu i očekivanu frekvenciju. Puni krug označuje opaženu vrijednost. Prazni krug označuje očekivanu vrijednost. Prvi panel pokazuje podudaranje opaženoga s očekivanim. U drugome svaka skupina ima 20 osoba, a u trećemu 80. Treći panel zato ima četiri puta veće frekvencije od drugoga uz jednake postotke. Uz naslove su ukupni hi-kvadrat i Cramérovo V.



Slika 2: Opažene i očekivane frekvencije bez odstupanja te pri jednakom relativnom pomaku uz dva rubna zbroja. Izrada autora. Podaci i postupak opisani su u tekstu.

Prije usporedbe rezultata možemo predvidjeti što bi se dogodilo kada bismo učetverostručili sve četiri frekvencije, a zadržali iste postotke. Bi li hi-kvadrat ostao isti i bi li se jednako ponašalo

Cramérovo V? Ovdje mijenjamo brojnosti zamišljene tablice radi usporedbe. Kopiranje zapisa stvarnog istraživanja ne bi stvorilo dodatna neovisna opažanja.

Isti relativni pomak daje veće ukupno odstupanje kada imamo više opažanja. Pri pomaku od 20 % i povećanju rubnog zbroja svake skupine s 20 na 80 hi-kvadrat raste s 1,6 na 6,4, a Cramérovo V ostaje 0,20. Relativna jačina veze nije se promijenila. Promijenila se količina podataka s kojom uspoređujemo opaženi obrazac i model nezavisnosti. Pri učetverostručenju frekvencija i očekivanja svaka razlika postaje četiri puta veća. Njezin kvadrat postaje šesnaest puta veći, a očekivanje u nazivniku četiri puta veće, pa se svaki doprinos i njihov zbroj učetverostručuju. U računu za V hi-kvadrat i broj osoba rastu zajedno, zato V ostaje isti.

Hi-kvadrat i njegova referentna raspodjela

Doprinose ćelija računamo tako da oduzmemo očekivanu frekvenciju od opažene, razliku kvadriramo i podijelimo očekivanom frekvencijom. Kvadriranjem sprječavamo poništavanje pozitivnih i negativnih odstupanja. Zbroj doprinosa sažima odstupanje cijele tablice.

Hi-kvadrat statistika je zbroj kvadriranih odstupanja opaženih od očekivanih frekvencija, pri čemu je svako odstupanje podijeljeno očekivanom frekvencijom svoje ćelije.
Zbroj označujemo s χ^2. Oznaka O_{ij} predstavlja opaženu, a E_{ij} očekivanu frekvenciju u i-tom retku i j-tom stupcu.

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Za našu tablicu statistika iznosi 149,8, uz 12 stupnjeva slobode. Približna p-vrijednost iz hi-kvadrat referentne raspodjele iznosi

<0,001. Stupnjeve slobode računamo množenjem broja redaka umanjenog za jedan i broja stupaca umanjenog za jedan.

Stupnjevi slobode proizlaze iz ograničenja rubnih zbrojeva. U tablici dva puta dva, uz zadane rubove, odabir jedne ćelije određuje preostale tri. U našoj tablici s četiri retka i pet stupaca možemo slobodno zadati 12 ćelija. Ostalih osam određuju rubovi. Broj stupnjeva slobode zato utječe na referentnu raspodjelu odstupanja.

U poglavlju o testiranju opaženu smo statistiku uspoređivali s vrijednostima koje nastaju ponavljanjem pod nultim modelom. Ovdje istu ulogu ima teorijska hi-kvadrat raspodjela odgovarajućih stupnjeva slobode, kada je to približenje primjereno. Usporedbom naše statistike s tom raspodjelom dobivamo p-vrijednost. Za jedan stupanj slobode granica na razini 0,05 iznosi 3,84. S više stupnjeva slobode granica je viša.

Hi-kvadrat sažima odstupanje cijele tablice. Da bismo vidjeli gdje se ono pojavljuje, vraćamo se konkretnim dobnim skupinama i izvorima te računamo rezidual svake ćelije. Budući da rubne udjele procjenjujemo iz iste tablice, oni ograničavaju promjenjivost ćelija. Pri standardiziranju odstupanja zato uz očekivanu frekvenciju uzimamo u obzir i udjele retka i stupca. Ako u tablici 2 × 2 uz zadane rubove povećamo jednu ćeliju, preostale tri moraju se odgovarajuće promijeniti. Odstupanja su stoga međusobno vezana zadanim rubnim zbrojevima.

Prilagođeni standardizirani rezidual ćelije je razlika opažene i očekivane frekvencije podijeljena procijenjenom standardnom devijacijom te razlike, koja uzima u obzir rubne udjele retka i stupca.
Oznaka e označuje ostatak, $p_{i\cdot}$ rubni udio retka, a $p_{\cdot j}$ rubni udio stupca. Slovo r u knjizi zadržavamo za korelaciju.

$$e_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}(1 - p_{i\cdot})(1 - p_{\cdot j})}}$$

Pozitivan rezidual pokazuje više opažanja nego što model očekuje, a negativan manje. Kada je aproksimacija primjerena, apsolutna vrijednost oko dva može upozoriti na neuobičajeno odstupanje. To je orijentir za čitanje. Pregled mnogo reziduala nakon ukupnog testa zahtijeva oprez zbog višestrukih usporedbi.

Najveći pozitivan rezidual pripada televiziji među osobama od 60 i više godina i iznosi 6,67. Najveći negativan pripada društvenim mrežama u istoj skupini i iznosi −6,63. Među najmlađima predznaci su obrnuti, s 6,57 za mreže i −4,20 za televiziju. Tako ukupno odstupanje možemo povezati s konkretnim skupinama i izvorima.

Značajno nije isto što i snažno

Hi-kvadrat raste s veličinom uzorka. Uz iste postotke tablica na osam tisuća osoba daje deset puta veću statistiku nego tablica na osamsto. Zato p-vrijednost ne može mjeriti jačinu veze. Mijenja se i kada relativni obrazac ostane isti.

Cramérovo V prilagođava hi-kvadrat veličini uzorka i dimenzijama tablice. Kreće se između nule i jedinice. Nula označuje nezavisnost u tablici, a jedinica najjaču povezanost koju ta mjera može prikazati.

> **Cramérovo V** je korijen hi-kvadrat statistike podijeljene umnoškom veličine uzorka i manje dimenzije tablice umanjene za jedan.
> Oznakom V označujemo tu mjeru. U formuli je n ukupan broj jedinica, a k manji od broja redaka i broja stupaca.
>
> $$V = \sqrt{\frac{\chi^2}{n\,(k-1)}}$$

U našoj tablici V iznosi 0,25. Jednaka vrijednost V-a ne nosi univerzalnu oznaku male ili velike veze za tablice različitih dimen-

zija. Važnost veze trebamo procijeniti prema pitanju i ranijim nalazima iz istog područja.

Za odluku su često jasniji postoci. Televiziju bira 12,2 % najmlađih i 45,2 % najstarijih ispitanika. Uredništvo može tu razliku povezati s raspodjelom resursa. Uz V kao sažetak jačine veze čitamo pojedine udjele kako bismo razumjeli obrazac.

Referentna raspodjela je istraživačka odluka

Dosad smo uspoređivali dob i izvor vijesti. Na istom uzorku od 800 osoba možemo postaviti drukčije pitanje, odgovara li sastav uzorka prema obrazovanju poznatom sastavu naše simulirane populacije. Zadržavamo osobe, ali sada brojimo samo razine obrazovanja. Za svaku razinu očekivani broj dobivamo množenjem 800 s njezinim udjelom u populaciji. U tablici je uz taj broj i druga moguća referenca, 200 osoba po razini, koja bi slijedila iz pretpostavke o četiri jednako zastupljene kategorije.

Tablica 2: Ista opažanja prema dvjema referencama. Izrada autora.

Obrazovanje	Opaženo	Prema populaciji	Jednaki udjeli
osnovna	157	142,9	200
srednja	366	377,5	200
viša	188	189,4	200
diplomska	89	90,2	200

Prvi redak sada možemo usporediti s dvama različitim očekivanjima, a zatim isti račun ponoviti za preostale razine. Taj postupak zovemo **testom prilagodbe**. Zbrajanje kvadriranih odstupanja podijeljenih očekivanim brojevima ostaje isto kao prije. Uz četiri izvana zadane kategorijske vjerojatnosti raspodjela ima tri stupnja slobode, jedan manje od broja kategorija. Očekivanja ovdje zadajemo izvana, umjesto da ih računamo iz rubnih zbrojeva tablice dviju varijabli.

Ako pretpostavimo jednaku zastupljenost četiriju razina obrazovanja, statistika iznosi 209,3, uz p-vrijednost manju od 0,001. To pokazuje velik nesklad. Ako koristimo poznate udjele simulirane populacije, statistika iznosi 1,77, uz p-vrijednost 0,62. Drugi test ne pokazuje jasan nesklad, ali ne dokazuje jednakost. Najveća opažena razlika jednog udjela prema populacijskom iznosi 1,8 postotnih bodova.

Oba računa mogu biti točna jer ispituju različite tvrdnje. Jednaki udjeli ovdje nemaju sadržajno opravdanje, dok poznata populacijska raspodjela ima. U stvarnom istraživanju referenca obično nije tako sigurno poznata. Trebamo je odabrati i obrazložiti prije testa.

Ista pitanja o kategorijama i nazivnicima pojavljuju se kada brojimo kodirane tekstove. Jedinica tada može biti dokument, objava ili govor. Prije tablice navodimo tko je odredio kodnu knjigu i pravila razvrstavanja, a nazivnik usklađujemo s jedinicama koje zadovoljavaju uvjete za kodiranje. Posebno prikazujemo nekodirane jedinice i one kojima je dodijeljeno više oznaka.

Ako retci predstavljaju izvor teksta, a stupci kodiranu kategoriju, nalaz ovisi i o tekstovima i o pravilima kodiranja. U poglavlju o algoritmima istu ćemo logiku primijeniti na tablicu zabune. I ondje moramo znati kako su kategorije nastale i koji je nazivnik svakog udjela.

Kad je aproksimacija tanka

U uzorku od 800 osoba najmanja očekivana frekvencija iznosi 14,3 pa nijedna ćelija nije osobito rijetka. Da bismo vidjeli što se događa kada očekivani brojevi postanu mali, privremeno napuštamo taj uzorak i pojednostavljujemo pitanje. U zasebnom nastavnom pokusu imamo dvije hipotetske dobne skupine i samo odgovor koristi li osoba određenu društvenu mrežu.

To nisu novi podaci o Hrvatskoj niti podskup prethodnih 800 zapisa.

Najprije nas zanima koliko često test prijavljuje vezu koje u modelu nema. Svakoj zamišljenoj osobi neovisno dodjeljujemo odgovor „da” s unaprijed zadanom vjerojatnošću, a inače odgovor „ne”. Jednake vjerojatnosti u objema skupinama stvaraju nulti model. U većem uzorku generiramo po 200 osoba, a u manjem po 20. Donja dva retka tablice mijenjaju pitanje i poslužit će poslije za procjenu snage, kada je veza ugrađena u model.

Tablica 3: Pravila zasebnoga nastavnog pokusa. Izrada autora.

Svrha i uzorak	Osoba po skupini	„Da” među mlađima	„Da” među starijima
Nula, veći	200	50 %	50 %
Nula, manji	20	10 %	10 %
Snaga, veći	200	60 %	40 %
Snaga, manji	20	16 %	4 %

Prema prvom nultom modelu u svakoj skupini očekujemo $200 \times 0{,}50 = 100$ odgovora „da” i 100 odgovora „ne”. U manjem modelu očekujemo $20 \times 0{,}10 = 2$ odgovora „da” i 18 odgovora „ne”. To su očekivanja iz pravila generatora. U pojedinoj izvučenoj tablici ukupan broj korisnika mreže varira, pa test očekivane frekvencije ponovno računa iz njezinih ostvarenih rubnih zbrojeva.

Jedno ponavljanje manjega pokusa znači generirati svih 40 odgovora, prebrojiti ih u tablici i izračunati jednu Pearsonovu statistiku i njezinu p-vrijednost. Sljedeću smo moguću tablicu namjerno odabrali radi preglednog računa. Ona nije nasumično izdvojeno ponavljanje iz simulacije.

Tablica 4: Mogući ishod manjega pokusa, odabran za ilustraciju. Izrada autora.

Dobna skupina	Koristi mrežu	Ne koristi mrežu	Ukupno
Mlađi	4	16	20
Stariji	0	20	20
Ukupno	4	36	40

Mrežu koristi 4 od 20 mlađih, odnosno 20 %, i 0 od 20 starijih. Pod nezavisnošću primjenjujemo zajednički udio $4/40$ na svaku skupinu od 20 osoba. Očekivani brojevi zato su 2 i 18 u oba retka. Ovdje se slučajno podudaraju s očekivanjima generatora. Da smo dobili ukupno šest korisnika, test bi očekivao po tri u svakoj skupini. Za prikazanu tablicu četiri doprinosa Pearsonovoj statistici iznose $4/2$, $4/18$, $4/2$ i $4/18$, jer je svako opažanje od svoga očekivanja udaljeno za dva. Njihov je zbroj 4,44.

Učestalost pogrešaka procjenjujemo ponavljanjem cijelog postupka, pa svaki scenarij ponavljamo 4 000 puta. Veličine skupina i vjerojatnosti ostaju zadane, a svaki put generiramo nove neovisne odgovore. Brojimo koliko puta Pearsonov test, odnosno test sa zbrojem (opaženo − očekivano)^2/očekivano, bez korekcije kontinuiteta daje p-vrijednost manju od 0,05. Broj odbacivanja dijelimo s 4 000, uključujući ponavljanja bez ijednog korisnika mreže ili bez ijednog nekorisnika. U njima test nije definiran i ovdje ih računamo kao neodbacivanja. Tablica ih pokazuje zasebno kako bi nazivnik ostao vidljiv.

Tablica 5: Odbacivanja u po 4 000 ponavljanja. Neprimjenjiva ponavljanja ostaju u nazivniku kao neodbacivanja. Izrada autora.

Scenarij	Odbacivanja	Postotak	Neprimjenjiva
Nula, veći uzorak	218	5,4 %	0
Nula, manji uzorak	162	4,0 %	66
Veza, veći uzorak	3928	98,2 %	0
Veza, manji uzorak	819	20,5 %	64

Prva dva retka procjenjuju lažna odbacivanja jer u njihovim modelima nema veze. Dobivenih 5,4 % i 4,0 % treba čitati uz slučajnu promjenjivost simulacije. Oko stope od 5 % standardna pogreška na 4 000 ponavljanja iznosi oko 0,34 postotna boda. Male razlike u postocima zato nisu same po sebi dokaz razlike u ponašanju postupaka. Rezultat manjega scenarija odnosi se i na naše pravilo o neprimjenjivim tablicama. U ovom scenariju s malim očekivanim frekvencijama stopa je ispod 5 %, pa je postupak konzervativan. Ovaj pokus zato ne pokazuje povećanu pogrešku prve vrste. Male očekivane frekvencije upozoravaju na nepouzdanu aproksimaciju, ali ne određuju unaprijed smjer odstupanja.

Vratimo se tablici s četiri korisnika među mlađima i nijednim među starijima. Želimo provjeriti koliko je taj raspored neobičan bez oslanjanja na glatku hi-kvadrat krivulju. Zadržimo po 20 osoba u skupinama i ukupno četiri korisnika mreže. Tada su mogući samo sljedeći rasporedi korisnika, dok brojeve nekorisnika dobivamo oduzimanjem od 20.

Tablica 6: Rasporedi četiriju korisnika uz fiksne rubne zbrojeve. Izrada autora.

Među mlađima	Među starijima	Uvjetna vjerojatnost
0	4	5,30 %
1	3	24,95 %
2	2	39,50 %
3	1	24,95 %
4	0	5,30 %

Tablice nisu jednako vjerojatne. Dva korisnika u svakoj skupini mogu nastati na više načina nego sva četiri u jednoj, kada svaki izbor četiriju osoba među njih 40 ima jednaku vjerojatnost. Primjerice, vjerojatnost posljednjeg retka dobivamo dijeljenjem broja izbora četiriju među 20 mlađih brojem izbora četiriju među svih 40 osoba. Za srednji redak biramo dvije među 20 mlađih i dvije među 20 starijih. Tako nastaju uvjetne vjerojatnosti u tablici.

Na tom računu počiva **Fisherov egzaktni test**. Za dvostranu p-vrijednost ovdje zbrajamo vjerojatnosti svih tablica čija je uvjetna vjerojatnost manja ili jednaka vjerojatnosti opažene tablice. Posljednjem retku zato pridružujemo prvi i dobivamo približno 0,106. Pearsonov račun bez korekcije za istu tablicu daje približno 0,035. Razlika pokazuje da način računanja referentne vjerojatnosti može promijeniti odluku pri pragu 0,05. Ne pokazuje da je Fisherova p-vrijednost uvijek veća, niti jedna odabrana tablica utvrđuje opće ponašanje tih testova. Egzaktnost se odnosi na račun uz zadane rubove, a ne uklanja nesigurnost malog uzorka, pogreške mjerenja ili ograničenja nacrta.

Preostaje drugo pitanje, koliko često otkrivamo vezu koja postoji u modelu. Za **snagu** zato trebamo promijeniti pravila generiranja. Donja dva retka nacrta daju mlađima i starijima različite vjerojatnosti odgovora „da”. Uz jednako velike skupine obje razlike odgovaraju populacijskom Cramérovu V od 0,20. U većem scenariju razlika udjela iznosi 0,20 uz

zajednički udio 0,50, a u manjem 0,12 uz zajednički udio 0,10. Standardiziranje tih razlika daje istu vrijednost V, iako su apsolutne razlike različite. Za dvije jednako velike skupine vrijedi $V = |p_1 - p_2|/[2\sqrt{\bar{p}(1-\bar{p})}]$. Zato su računi $0{,}20/[2\sqrt{0{,}50 \cdot 0{,}50}] = 0{,}20$ i $0{,}12/[2\sqrt{0{,}10 \cdot 0{,}90}] = 0{,}20$.

Sada svaki odbačeni nulti model znači otkrivenu ugrađenu vezu. Od po 4 000 ponavljanja odbacujemo u 98,2 % većih i 20,5 % manjih uzoraka. Između scenarija mijenjaju se i broj osoba i učestalost korištenja mreže, pa cijelu razliku u snazi ne možemo pripisati samo veličini uzorka. Ta dva retka ne procjenjuju pogrešku prve vrste niti pouzdanost pojedine odluke.

U početnoj analizi dobi i izvora vijesti ove provjere usmjeravaju čitanje tablice. Najprije provjeravamo brojanja i očekivane frekvencije, zatim primjerenost referentnog računa, a uz test i dalje navodimo udjele, njihovu razliku i nesigurnost. Uz pouzdaniji račun p-vrijednosti i dalje trebamo obrazložiti značenje kategorija i doseg uzorka.

Odluka o spajanju kategorija traži i sadržajno opravdanje. Sljedeći primjer pokazuje što se pritom može izgubiti iz istraživačkog pitanja.

● STATISTIKA U DIVLJINI

Tri odgovora u jednoj kategoriji. Pew Research Center u izvještaju o odraslim stanovnicima SAD-a bez religijske pripadnosti zajedno prikazuje ateiste, agnostike i osobe koje na pitanje o svojoj religiji biraju odgovor „ništa posebno” (Pew Research Center 2024.). Zajednička kategorija omogućuje usporedbu s osobama koje navode neku religiju. Izvorni odgovori pokazuju i razlike unutar nje. Prema procjeni s težinama iz ankete provedene 2023., u Boga ili neku višu silu vjeruje 67 % agnostika, uz objavljenu 95-postotnu marginu uzoračke pogreške za tu skupinu od ±6,0 postotnih bodova (Pew Research Center 2024.). Za osobe s odgovorom „ništa posebno” procjena iznosi 83 %, a odgovarajuća margina za tu skupinu ±3,5 postotnih bodova (Pew Research Center 2024.). Svaki se postotak odnosi na vlastitu skupinu. Jedan zajednički udio prikrio bi tu razliku.

Ako nas zanimaju uvjerenja agnostika i osoba koje su odabrale odgovor „ništa posebno”, trebamo sačuvati zasebne kategorije i podatak o vjerovanju. Izbor podjele određuje koje ćemo razlike moći opisati. O tome vodimo računa i kada u hi-kvadrat analizi razmatramo spajanje rijetkih kategorija. Veći očekivani brojevi mogu poboljšati hi-kvadrat aproksimaciju, ali ne vraćaju razlike koje smo spajanjem uklonili. Zato uz provjeru prikladnosti postupka čuvamo izvorne frekvencije i udjele te objašnjavamo na koje pitanje odgovara spojena tablica.

NAĐITE GREŠKU

Može li ovaj zaključak u izvještaj?

Upravo smo vidjeli da pri tumačenju testa trebamo provjeriti očekivane frekvencije i razlikovati statističku značajnost od jačine povezanosti. Vratimo se sada uzorku od 800 zamišljenih osoba i pogledajmo kako se te provjere koriste pri pisanju izvještaja.

Kolegica je analizirala povezanost dobne skupine i glavnog izvora vijesti. Prije predaje rada zamolila vas je da pregledate njezin zaključak. Poslala vam je sažetak rezultata.

Tablica 7: Sažetak analize simuliranog uzorka. Izrada autora.

Što je provjereno ili izračunano	Rezultat
Broj osoba u simuliranom uzorku	800
Broj dobnih skupina i izvora vijesti	4 skupine i 5 izvora
Najmanja očekivana frekvencija	14,3
Hi-kvadrat statistika	149,8
Stupnjevi slobode	12
p-vrijednost	< 0,001
Cramérovo V	0,25

U izvještaju je zadržala i dio opisa koji čitatelju pokazuje kako se skupine razlikuju.

Tablica 8: Dio postotaka po dobnim skupinama. Izrada autora.

Dobna skupina	Televizija kao glavni izvor	Društvene mreže kao glavni izvor
Od 18 do 29 godina	12,2 %	42,1 %
60 i više godina	45,2 %	2,2 %

Postoci su izračunani unutar svake dobne skupine. Prikazana su dva od pet izvora, pa se postoci u retku ne zbrajaju do sto.

Ispod tablica napisala je sljedeće.

> Sve očekivane frekvencije veće su od pet. Hi-kvadrat test daje p-vrijednost manju od 0,001, pa podaci snažno odstupaju od modela nezavisnosti. Budući da je p-vrijednost tako mala, povezanost dobi i izvora vijesti vrlo je snažna.

Prije nego što joj odgovorite, povežite svaku rečenicu s rezultatom na koji se oslanja.

1. Koja rečenica izvodi zaključak koji navedeni pokazatelj ne opravdava?
2. Koji rezultat trebate koristiti za opis jačine povezanosti, a koji dio prikaza pomaže objasniti njezino praktično značenje?
3. Prepravite završnu rečenicu. Zadržite ono što rezultati pokazuju, ali nemojte veličinu povezanosti zaključivati iz p-vrijednosti.

Razrađeni primjer

Spajanje kategorija možemo razmotriti na pitanju koje smo sadržajno odredili prije računa. Umjesto razlika među svih pet izvora sada nas zanima razlikuje li se izbor digitalnih i tradicionalnih izvora među dobnim skupinama. Portal i društvene mreže svrstavamo u digitalne, a televiziju, tisak i radio u tradicionalne izvore. Ta podjela odgovara novom pitanju i stvara novu varijablu. Najprije ćemo pročitati njezine udjele, a zatim ukupno odstupanje od nezavisnosti.

Vraćamo se istom uzorku od 800 osoba, a zasebni pokus s dvjema skupinama ovdje završava. Sve osobe ostaju uključene. Primjerice, u najmlađem retku zbrajamo 66 korisnika portala i 83 korisnika mreža, što daje 149 digitalnih korisnika. Preostali izvori daju 48 tradicionalnih korisnika. Nazivnik retka ostaje 197.

Tablica 9: Iste osobe nakon spajanja izvora u dvije kategorije. Izrada autora.

Dob	Digitalni	Tradicionalni	Ukupno
18 do 29	149	48	197
30 do 44	179	117	296
45 do 59	76	96	172
60 i više	26	109	135

Za konkretnu usporedbu najmlađih i najstarijih čitamo udjele digitalnih izvora. Bira ih 75,6 % najmlađih i 19,3 % najstarijih. Razlika iznosi 56,4 postotnih bodova, uz 95-postotni interval od 47,4 do 65,3. Ukupni test, koji obuhvaća sve dobne skupine, odgovara na šire pitanje o povezanosti u cijeloj tablici. Hi-kvadrat iznosi 114,3, uz 3 stupnja slobode, a V iznosi 0,38. Za pet kategorija V je bio 0,25. Prikladnost podjele prosuđujemo prema pitanju na koje želimo odgovoriti. Prva čuva razlike među pojedinačnim izvorima, a druga opisuje digitalno-tradicionalnu podjelu. Usporedba pokazuje koliko zaključak ovisi o operacionalizaciji. Interval razlike udjela računan je normalnom aproksimacijom s odvojenim procjenama varijance dvaju udjela, bez korekcije kontinuiteta. Pretpostavlja neovisne osobe i skupine te dovoljno velike brojnosti u svakoj od četiriju ćelija.

Tablica opisuje povezanost dobi i izbora izvora. Među mogućim objašnjenjima jesu ranije stečene navike, dostupnost uređaja i životno razdoblje. Za njihovo razdvajanje trebali bismo dodatne varijable i prikladan nacrt. U poglavlju o regresiji naučit ćemo zajedno razmatrati više varijabli.

Iz dobivenih rezultata sada možemo sastaviti kratak izvještaj.

> U slučajnom uzorku od 800 osoba iz nastavne simulacije usporedili smo izbor digitalnih i tradicionalnih izvora među dobnim skupinama. Digitalne izvore bira 75,6 % osoba od 18 do 29 godina i 19,3 % osoba od 60 i više godina. Razlika iznosi 56,4 postotnih bodova uz

> 95-postotni interval od 47,4 do 65,3. Za cijelu tablicu hi-kvadrat iznosi 114,3, uz 3 stupnja slobode i p-vrijednost manju od 0,001, a Cramérovo V 0,38. Rezultat opisuje zamišljene osobe i ne daje procjenu medijskih navika stanovnika Hrvatske niti dokaz da dob uzrokuje izbor izvora.

Prva rečenica određuje pitanje i podatke. Sljedeće daju opaženi obrazac i preciznost usporedbe najmlađih s najstarijima. Ukupni test i V odnose se na sve četiri dobne skupine, pa ne zamjenjuju taj konkretni kontrast. Završna rečenica određuje doseg nalaza. Za uredničku publiku mogli bismo u prvi plan staviti razliku u postotnim bodovima, a tehničke pokazatelje ostaviti u pratećoj tablici. Oznaka simulacije, nazivnici i granice zaključka moraju ostati i u kraćoj verziji.

Za dodatnu provjeru slijedi programski račun prethodne analize.

```
uzorak <- s13_uzorak |>
  mutate(skupina_medija = if_else(
    izvor_vijesti %in% c("portal", "društvene mreže"),
    "digitalni",
    "tradicionalni"
  ))

tablica <- table(uzorak$dobna_skupina, uzorak$skupina_medija)

round(prop.table(tablica, 1) * 100, 1)
chisq.test(tablica)
```

Kod preuzima uzorak i učitane pakete iz već pripremljene analize poglavlja. U računu pratimo isti put kojim smo čitali prethodnu tablicu. Poziv `table` prebroji kombinacije dobne skupine i nove podjele izvora, a `prop.table` pretvori frekvencije u udjele prema zadanom smjeru. Tek potom `chisq.test` sažima odstupanje od nezavisnosti u hi-kvadrat, određuje stupnjeve slobode i računa p-vrijednost prema referentnoj raspodjeli. Za

tablicu 2 × 2 R zadano primjenjuje korekciju kontinuiteta, koja ublažava prijelaz s diskretnih brojanja na kontinuiranu krivulju. Za ovdje opisani nekorigirani račun treba navesti `correct = FALSE`.

Vratimo se istim berkeleyjskim prijavama. Zbirna tablica pokazivala je povezanost spola i ishoda prijave (Bickel i sur. 1975.). Hi-kvadrat, V i reziduali mogli bi opisati taj obrazac, ali ne i objasniti ga. Nedostajao je odjel prijave. Muškarci i žene nisu se jednako raspoređivali po odjelima, a odjeli su imali različite stope prijma. U poglavlju Zašto statistika vidjeli smo Berkeleyjev primjer, a u poglavlju o povezanosti pokazali smo kako dodavanje takva sloja može promijeniti zbirni obrazac (Simpson 1951.). Statistički opis dviju varijabli zato treba razlikovati od objašnjenja mehanizma.

Sažetak

Od brojanja kategorija prelazimo na usporedbu tek kada odredimo nazivnik svakog udjela. Model nezavisnosti tada iz rubnih zbrojeva daje očekivane frekvencije s kojima uspoređujemo opaženu tablicu. Hi-kvadrat sažima ukupno odstupanje, reziduali ga vraćaju konkretnim ćelijama, a Cramérovo V pomaže odvojiti jačinu veze od količine podataka. Ako umjesto nezavisnosti provjeravamo podudaranje sa zadanom raspodjelom, trebamo obrazložiti tu referencu. Male očekivane frekvencije mogu oslabiti aproksimaciju, pa razmatramo Fisherov postupak ili sadržajno opravdanu novu podjelu kategorija, svjesni da potonjom mijenjamo pitanje. U sljedećem poglavlju zadržat ćemo usporedbu skupina, ali ćemo umjesto rasporeda kategorija promatrati brojčani ishod.

Nazivi neovisnost i nezavisnost u ovom probabilističkom značenju označavaju isti odnos. Neovisnost znači da zajedničku vjerojatnost možemo dobiti množenjem pojedinačnih vjerojat-

nosti. Zato očekivane frekvencije u tablici nastaju umnoškom rubnih udjela i ukupnog broja osoba.

Pojmovi

stupnjevi slobode (*degrees of freedom*), doprinos ćelije (*cell contribution*), kontingencijska tablica (*contingency table*), uvjetni nazivnik (*conditional denominator*), očekivana frekvencija (*expected frequency*), hi-kvadrat statistika (*chi-squared statistic*), prilagođeni standardizirani rezidual (*adjusted standardized residual*), test prilagodbe (*goodness-of-fit test*), referentna raspodjela (*reference distribution*), Cramérovo V (*Cramér's V*), Fisherov egzaktni test (*Fisher's exact test*)

Zadaci

Za odabrane zadatke ovoga poglavlja na kraju knjige dostupne su naznake i odvojena objašnjenja. Najprije pokušajte samostalno odgovoriti.

Konceptualni

Objasnite razliku između testa nezavisnosti i mjere jačine veze tako da napišete dvije rečenice koje bi se mogle pojaviti u istom izvještaju, jednu o tome što test kaže i jednu o tome što V kaže.

Zatim izvještaj iz razrađenog primjera preoblikujte u tri rečenice za novinara. Zadržite oznaku simulacije, dobne skupine, nazivnike postotaka, razliku s intervalom i granicu zaključka. Obrazložite koji biste tehnički pokazatelj ostavili u pratećoj tablici i zašto njegovo izostavljanje iz glavnog teksta ne mijenja značenje nalaza.

Računski

Tablica ima dva retka s po sto ispitanika i dva stupca s rubnim zbrojevima stodvadeset i osamdeset. Izračunajte sve četiri očekivane frekvencije, a zatim za opaženu tablicu sa sedamdeset u prvoj ćeliji izračunajte doprinos svake ćelije i njihov zbroj. Usporedite ga s graničnom vrijednošću 3,84.

Kritički

Vratite se na Simpsonov paradoks iz poglavlja Zašto statistika i na Berkeleyjev slučaj (Bickel i sur. 1975.). Skicirajte zbirnu tablicu spola i ishoda prijave te skup odjelskih tablica koje vraćaju izostavljeni sloj. Objasnite zašto hi-kvadrat test prve tablice ne može razlikovati razliku u sastavu prijava od razlike u odlučivanju unutar odjela. Predajte obje sheme, jedan odlomak usporedbe i naziv varijable bez koje se ta dva objašnjenja ne mogu razdvojiti.

Revizija modela

Vratite se izvještaju iz okvira o pogrešci. Izdvojite dijagnostičke korake koji su provedeni ispravno i jedan pogrešan zaključak. Napišite zamjenski odlomak koji rabi već navedeno Cramérovo V i uspoređuje udjele unutar dobnih skupina. Objasnite što provjera očekivanih frekvencija potvrđuje, a koje pretpostavke njome još nisu provjerene. Ne dodjeljujte jačini univerzalnu oznaku bez sadržajnog obrazloženja.

14

DIO V · MODELI

uspoređivanje dviju grupa

Ista razlika, različiti dizajni i zaključci

Što biste dobili, a što propustili kada biste nakratko isključili Facebook? Možda biste oslobodili vrijeme, ali biste drukčije dolazili do vijesti i obavijesti ljudi koje pratite. U poglavlju o povezanosti razmatrali smo odnos digitalnih navika i dobrobiti. Ovdje nas zanima posljedica promjene navike.

Deaktivacija računa tijekom četiri tjedna istražena je pokusom prije američkih izbora 2018. godine (Allcott i sur. 2020.). Dobrovoljci koji su pristali na ponuđenu naknadu za deaktivaciju računa nasumično su raspoređeni u skupine (Allcott i sur. 2020.). Istraživači su pratili dobrobit i informiranost, uz usporedbu sa skupinom koja je nastavila koristiti Facebook (Allcott i sur. 2020.).

U odnosu na istraživanje adolescenata iz poglavlja o povezanosti ovdje se mijenjaju i sudionici i pitanje. Pratimo određenu skupinu korisnika i privremenu deaktivaciju jedne platforme.

Za svaki ishod trebalo je procijeniti razliku među skupinama i njezinu nesigurnost. Randomizacija je oblikovala skupine, ali sudionici nisu bili slučajan uzorak svih korisnika (Allcott i sur. 2020.). Kako usporedbu uskladiti s istraživačkim nacrtom i obrazložiti na koga se zaključak odnosi?

Oznaka glavnoga izvora vijesti u prethodnom nam je poglavlju služila za brojanje osoba u kategorijama. Sada uz oznaku skupine promatramo brojčani odgovor. Zanima nas koliko se prosječno povjerenje u medije razlikuje među skupinama i koliko smo tu razliku precizno procijenili. Poznato pitanje o povezanosti time dobiva drukčiji oblik podataka i drukčiji račun.

Usporedba tiska i portala pokazala je razliku u povjerenju, a dobne skupine različit sastav publike izvora. Sada procjenjujemo razliku između korisnika televizije i društvenih mreža s njezinim intervalom.

Iz iste nastavne populacije `populacija_medija` zasebno izvlačimo uzorak od 120 korisnika televizije i društvenih mreža. Svaki redak predstavlja jednu zamišljenu osobu, njezin izvor i povjerenje na ljestvici od jedan do deset. Rezultati se odnose na tu simulaciju, bez zaključka o stvarnoj publici.

Za izbor postupka vratit ćemo se nacrtu istraživanja. Razlika između rezultata različitih ljudi i promjena izmjerena na istim ljudima traže različite račune. Zasebno ćemo razmotriti i usporedbu jedne sredine s unaprijed zadanom vrijednošću. Na simuliranim ponovljenim mjerenjima provjerit ćemo kako očuvanje parova mijenja procjenu nesigurnosti. To nas vodi do različitih oblika t-testa, uz isto polazište u procjeni razlike.

Intervale, veličinu učinka i pogreške odluke već poznajemo, pa možemo usmjeriti pozornost na to koji račun odgovara kojim opažanjima. Welchov postupak usporedit ćemo s linearnim

modelom u kojem oznaka skupine postaje prediktor, odnosno varijabla kojom model opisuje razlike u ishodu. Provjerit ćemo zašto jednaka procjena razlike ne mora značiti jednaku standardnu pogrešku. Grafovima ćemo pregledati pretpostavke i razmotriti što mijenja postupak temeljen na rangovima. Naš će izvještaj povezati nacrt, procjenu i njezine granice. Sama razlika u povjerenju neće nam reći što bi se dogodilo kada bi osoba promijenila izvor vijesti.

Tko nosi dva rezultata

Prije nego što izaberemo račun za razliku, trebamo razjasniti što uspoređujemo. Jesu li to odgovori različitih ljudi, dva mjerenja istih ljudi ili rezultat jedne skupine prema zadanoj vrijednosti? Ta tri pitanja vode različitim postupcima.

Možemo usporediti sredinu jedne skupine s unaprijed zadanom vrijednošću, primjerice sredinom ljestvice ili pragom odluke. Procjena iz ranije ankete nije takva konstanta. Ima vlastitu nesigurnost, a možda i drukčiju ciljnu populaciju ili mjerenje, pa pri usporedbi trebamo uvažiti obje procjene. Druga situacija obuhvaća dvije skupine različitih ljudi, primjerice korisnike televizije i društvenih mreža. U trećoj iste osobe mjerimo dvaput, prije i poslije događaja.

> **Jedinica neovisnosti** je entitet koji se u istraživanju mogao pojaviti ili izostati neovisno o ostalima, pa njegove vrijednosti nisu unaprijed vezane uz vrijednosti bilo koje druge jedinice u istom skupu.

U našoj usporedbi dviju neovisnih skupina jedinica je osoba. Kod ponovljenog mjerenja osoba ostaje jedinica, ali daje dva povezana rezultata. Za dvadeset osoba mjerenih dvaput računamo dvadeset razlika, po jednu za svaku osobu. Ako zanemarimo povezanost unutar para, pogrešno računamo nesigurnost. Uz

pozitivnu povezanost iz primjera koji slijedi interval postaje nepotrebno širok.

Najprije zato imenujemo jedinicu i opisujemo veze među mjerenjima. Tek potom biramo postupak usporedbe.

Ovisnost može nastati i među različitim osobama. Učenici istog razreda ili članovi istog kućanstva mogu dijeliti uvjete koji utječu na rezultat. Povezani mogu biti i ljudi u društvenoj mreži. Ovisnost među njihovim zapisima trebamo zabilježiti i odabrati postupak koji uvažava tu strukturu. Takvi modeli prelaze opseg ovog poglavlja.

Razlika prije oznake

U usporedbi televizije i društvenih mreža zadržavamo osobe koje navode jedan od ta dva glavna izvora. Osobe s drugim izvorom ostaju izvan analize, pa se procjena odnosi samo na odabranu usporedbu unutar simulirane populacije.

U izvještaju prvo navodimo razliku u izvornim jedinicama i njezin interval. Zatim prikazujemo test i standardiziranu razliku. Tako čitatelj prije odluke o nuli može procijeniti veličinu i nesigurnost opaženog odnosa.

Uzorak od 120 korisnika televizije i društvenih mreža izvučen je bez vraćanja iz populacije ograničene na ta dva izvora, zasebno od uzorka od 800 osoba. Broj korisnika svakog izvora nije unaprijed zadan. Zbrajanjem njihovih odgovora i dijeljenjem odgovarajućim brojem osoba dobivamo dvije sredine.

Tablica 1: Od odgovora do sredina u novom uzorku od 120 osoba. Izrada autora.

Izvor	Osoba	Zbroj	Sredina	SD
društvene mreže	70	316	4,51	1,95
TV	50	285	5,70	2,05

U retku društvenih mreža dijelimo 316 s 70, a u retku televizije 285 s 50. Zatim od televizijske sredine oduzimamo sredinu mreža. Kada raspršenost procjenjujemo iz istog uzorka, njezina nesigurnost također ulazi u zaključak. **t-raspodjela** zvonasta je referentna krivulja s debljim repovima od normalne. Pri manje podataka zato traži veću udaljenost od nule. Kako raste broj stupnjeva slobode, približava se normalnoj krivulji. **Kritična t-vrijednost** omeđuje središnjih 95 % te referentne raspodjele za obostrani 95-postotni interval.

Stupnjevi slobode opisuju koliko neovisnih odstupanja ostaje nakon procjene sredina. U jednoj skupini procijenjena sredina veže zbroj odstupanja na nulu, pa ostaje n − 1. Združeni postupak za dvije skupine koristi $n_1 + n_2 - 2$ uz pretpostavku zajedničke varijance. Welchov postupak umjesto toga računa efektivnu vrijednost prema doprinosu nesigurnosti svake skupine. Ona može biti razlomljena i ne predstavlja broj osoba.

Za interval trebamo i nesigurnost te razlike. Welchov račun zbraja dvije procijenjene varijance sredina, $s_T^2/n_T + s_M^2/n_M$, i iz njihova zbroja vadi korijen. Dobivenu standardnu pogrešku 0,373 množimo pripadnom kritičnom t-vrijednošću i oduzimamo od razlike te joj dodajemo. Uključeni su svi odgovori.

U ovome računu Welchovi stupnjevi slobode iznose 102,471, a pripadna kritična vrijednost 1,983. Za test nulte razlike dijelimo opaženu razliku njezinom standardnom pogreškom. Dobivenu t-statistiku uspoređujemo s oba repa iste t-raspodjele. Taj račun pretpostavlja neovisne skupine i jedinice te približno normalne sredine. Za male skupine posebno je važan oblik raspodjele bez dominantnih krajnjih vrijednosti.

U uzorku od 120 korisnika televizije i društvenih mreža prosječno povjerenje iznosi 5,70 među onima koji navode TV kao glavni izvor vijesti i 4,51 među onima koji navode društvene mreže. Razlika iznosi 1,19 boda uz interval pouzdanosti od 0,45 do 1,92.

Koristimo Welchov interval, koji zasebno procjenjuje varijancu svake skupine. Pripadni obostrani test ispituje je li populacijska sredina TV skupine minus sredina skupine društvenih mreža jednaka nuli. Ovdje uspoređujemo sredine. Ne testiramo jednakost cijelih raspodjela.

Interval obuhvaća razlike ispod pola boda i blizu dva boda. Te vrijednosti mogu voditi različitim praktičnim odlukama. Širina intervala čini tu nepreciznost vidljivom uz oznaku značajnosti.

U simulaciji znamo da stvarna razlika iznosi 1,30 boda i da je interval obuhvaća. Dugoročni obuhvat provjeravamo ponavljanjima kakva smo koristili u poglavlju o procjeni. Ovaj je pogodak samo jedno od mogućih ponavljanja.

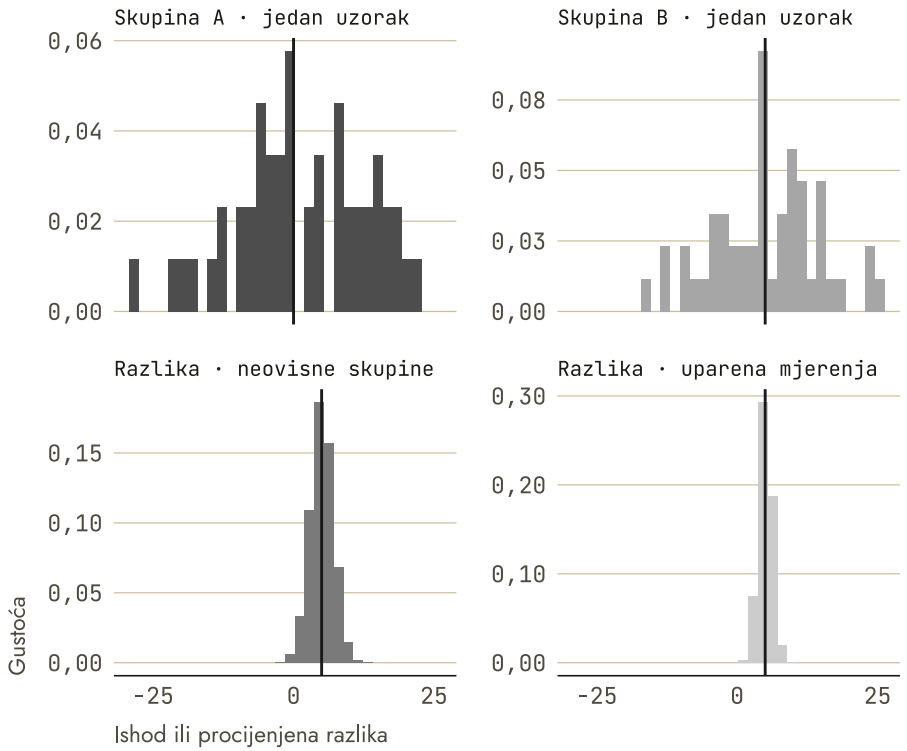
Uzorkivač dviju grupa

Ovo je zaseban normalni model bez ljestvice povjerenja. U prikazanoj postavci svaka skupina ima 50 ishoda, sredine su 0 i 5, a standardna devijacija 10. U neovisnom nacrtu svaki ishod izvlačimo zasebno. U uparenom dva ishoda iste osobe dijele dio slučajnog odstupanja, tako da je zadana korelacija 0,65. Za jedan skup oduzimamo sredinu A od sredine B, odnosno računamo prosjek promjena unutar parova.

Raspodjelu tih razlika ovdje stvaramo kraćim računom. Za zadani normalni model poznajemo njezinu sredinu i standardnu pogrešku, pa razlike izvlačimo izravno iz te raspodjele, bez stvaranja svih pojedinaca svakoga novog uzorka. Digitalni prikaz koristi 1 500 takvih razlika, a tiskani 4 000 po nacrtu. Uključena su sva ponavljanja. Njihov raspon pokazuje preciznost procjene. Za snagu testa ili obuhvat intervala trebali bismo u

svakom ponavljanju provesti odgovarajući dodatni račun. Oblik histograma može se malo promijeniti pri novom izvlačenju.

Na slici uspoređujemo raspršenost pojedinačnih rezultata i preciznost procijenjene razlike. Uz ishode skupina prikazane su raspodjele razlika za neovisna i uparena mjerenja. Veličina uzorka, stvarna razlika i raspršenost ishoda jednake su, pa možemo vidjeti učinak uparivanja. Okomite crte označuju zadane sredine ishoda u prva dva panela i stvarnu razliku u preostala dva. Okomite ljestvice prilagođene su pojedinom panelu, pa uspoređujemo širinu raspodjela na zajedničkoj vodoravnoj osi.



Slika 1: Preklapanje ishoda u jednom uzorku te raspodjele procijenjene razlike u neovisnom i uparenom dizajnu. Izrada autora. Podaci i postupak opisani su u tekstu.

U ovom primjeru uparena mjerenja daju užu raspodjelu procijenjenih razlika. Kada računamo promjenu unutar osobe, uklanjamo dio varijacije koji potječe od trajnih razlika među

osobama. Taj dobitak zadržavamo samo ako u analizi sačuvamo parove.

Ista razlika, dva dizajna

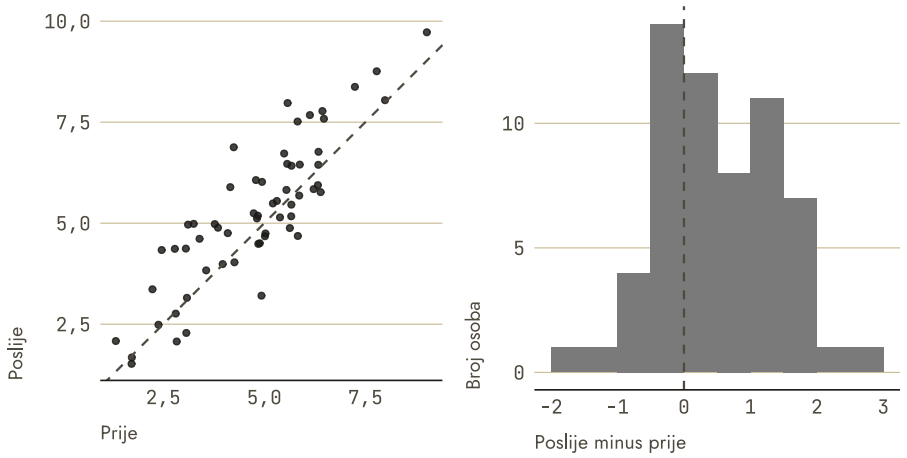
Pogledajmo to na 60 simuliranih osoba izmjerenih prije i poslije događaja. Pomak, varijance i povezanost zadani su kako bismo uvježbali račun za povezana mjerenja. Podaci ne opisuju stvarne ljude. Ni u stvarnom istraživanju takva promjena sama, bez kontrolne skupine ili randomizacije, ne bi dokazivala učinak događaja.

Vraćamo se računu iz pojedinačnih zapisa, ali stvaramo nove osobe. Svakoj od 60 osoba najprije generiramo zajedničku osnovu iz normalne raspodjele sa sredinom 5 i standardnom devijacijom 1,6. Početnom mjerenju dodajemo neovisno odstupanje standardne devijacije 0,7. Završnom dodajemo zadani pomak 0,45 i drugo takvo odstupanje. Prva tri para prikazuju ostvarene vrijednosti i njihovo oduzimanje. Ishodi ovdje nisu zaokruženi ni ograničeni na ljestvicu od 1 do 10.

Tablica 2: Tri od 60 zasebno simuliranih parova prije računanja prosječne promjene. Izrada autora.

Osoba	Prije	Poslije	Promjena
1	6,48	7,59	1,11
2	6,13	7,68	1,55
3	4,89	4,50	−0,39

Za pregled uparenih podataka potrebna su dva pogleda. Na prvome svaka točka povezuje dva mjerenja iste osobe, a dijagonala znači jednaka mjerenja. Na drugome pregledavamo raspodjelu razlika, jer se na njih odnosi upareni postupak. Ne možemo je zamijeniti zasebnim histogramima prije i poslije.



Slika 2: Parovi i njihove razlike u zasebno konstruiranoj nastavnoj simulaciji. Izrada autora.

Prva promjena iznosi 7,59 minus 6,48, odnosno 1,11. Zbroj svih 60 promjena dijelimo sa 60, a njihovu standardnu devijaciju s $\sqrt{60}$ za standardnu pogrešku sredine. Svi parovi ostaju uključeni.

Uparena t-statistika zato je sredina promjena podijeljena sa $s_{\text{promjena}}/\sqrt{60}$. Referentna t-raspodjela ima 59 stupnjeva slobode. P-vrijednost zbraja oba repa iza apsolutne opažene statistike. Neovisnost se ovdje traži među osobama, dok se povezanost dvaju mjerenja iste osobe čuva izračunavanjem njezine promjene.

Korelacija mjerenja iznosi 0,86. Osobe s višim početnim rezultatom uglavnom imaju viši i završni rezultat. Velik dio raspršenosti zato odražava razlike među osobama, dok nas zanima njihova promjena. Za svaku osobu najprije oduzmemo početni rezultat od završnoga. Dobivene razlike čine jedan uzorak čiju sredinu možemo usporediti s nulom, pa upareni račun prepoznajemo kao jednouzoračni postupak primijenjen na promjene.

Prosječna promjena iznosi 0,48, a standardna devijacija pojedinačnih razlika 0,90. Uparena analiza daje interval od 0,25 do 0,71, uz p-vrijednost 0,0001.

Ako iste podatke pogrešno obradimo kao neovisne skupine, dobivamo interval od −0,14 do 1,10 i p-vrijednost 0,126. Prosječna promjena ostaje ista, ali procijenjena nesigurnost znatno se mijenja.

Neovisna analiza koristi raspršenost među osobama, koja je ovdje velika. Uparena koristi raspršenost promjena unutar osoba, koja je manja. Budući da su iste osobe izmjerene dvaput, upareni postupak odgovara nacrtu istraživanja. Nacrt stoga određuje izbor postupka prije uvida u p-vrijednost.

Dobitak ovisi o varijanci razlika, a ona o raspršenosti obaju mjerenja i njihovoj povezanosti. Uz slične varijance veća pozitivna povezanost obično daje uži interval. Slaba povezanost daje manji dobitak, a pogrešno spojeni parovi mogu ga ukloniti. Parove zato određujemo prema identitetu jedinica i nacrtu.

Razlika skupina u linearnom modelu

Razliku dviju neovisnih skupina dosad smo dobivali oduzimanjem njihovih sredina. Istu procjenu možemo zapisati modelom koji svakoj osobi pridružuje sredinu njezine skupine. Za to je dovoljna oznaka nula ili jedan, a koeficijent uz nju postaje razlika sredina. Pritom čuvamo razliku među trima opisanim nacrtima i njihovim postupcima.

$$\text{povjerenje} = \beta_0 + \beta_1 \cdot \text{skupina} + \varepsilon$$

Oznaka β_0 predstavlja očekivani ishod u polaznoj skupini, a β_1 razliku druge skupine prema njoj. Oznaka ε obuhvaća odstupanje pojedine osobe od modela. Varijabla skupine ima vrijednost nula za polaznu i jedan za drugu skupinu.

> **Referentna skupina** je kategorija prema kojoj se izražavaju sve ostale kategorije istog prediktora, pa njezina sredina postaje

> polazna vrijednost modela, a koeficijenti ostalih kategorija odstupanja od nje.

Osobi iz referentne skupine model daje predviđanje β_0, jer se koeficijent razlike množi nulom. Osobi iz druge daje $\beta_0 + \beta_1$. Svi ljudi iste skupine dobivaju isti predviđeni prosjek, ali zadržavaju vlastite opažene odgovore i različite reziduale, odnosno razlike između opaženih i modelom predviđenih odgovora.

Kao referencu uzimamo društvene mreže. Zato β_0 iznosi 4,51 i jednak je sredini te skupine. Koeficijent β_1 iznosi 1,19 i jednak je prethodno izračunatoj razlici. Promjena reference mijenja odsječak i predznak koeficijenta, ali ostavlja iste procijenjene sredine skupina.

Jednouzoračni postupak uspoređuje sredinu s vanjskom vrijednošću, a upareni analizira sredinu razlika unutar parova. Svi postupci daju procjenu, nesigurnost i usporedbu s nulom, ali njihove pretpostavke nisu iste.

Uzročno tumačenje koeficijenta zahtijeva opravdanje iz nacrta i načina nastanka skupina. U našem primjeru ljudi sami biraju izvor vijesti, pa koeficijent opisuje razliku među njima.

Što razlika sama ne kaže

Model nam je dao drugi zapis iste razlike, ali još trebamo procijeniti njezin razmjer. Tu se vraćamo raspršenosti unutar skupina. Jedan bod predstavlja veće razdvajanje kada su odgovori zbijeni nego kada su široko raspršeni.

Standardiziranu razliku računamo dijeljenjem razlike sredina združenom standardnom devijacijom. Tako rezultat izražavamo u standardnim devijacijama, kako smo učinili u poglavlju o veličini učinka.

$$d = \frac{\bar{x}_2 - \bar{x}_1}{s_{\text{zdr}}}$$

Združena varijanca ponderirani je prosjek dviju varijanci, s težinama $n_1 - 1$ i $n_2 - 1$. Nakon korjenovanja dobivamo združenu standardnu devijaciju. Vraćamo se istom uzorku od 120 korisnika televizije i društvenih mreža. Za 70 korisnika mreža i 50 korisnika televizije račun iz nezaokruženih podataka daje 1,9957. Razlika sredina iznosi 1,1857, pa njihov omjer daje $d \approx 0{,}5941$, odnosno 0,59. Račun iz ranije tablice može se malo razlikovati jer su njezine standardne devijacije zaokružene.

Sadržajnu važnost procjenjujemo prema usporedivim nalazima i posljedicama razlike u izvornim jedinicama. U uparenom primjeru dijelimo standardnom devijacijom razlika i dobivamo 0,53. Budući da nazivnici nisu isti, te dvije standardizirane mjere nisu izravno usporedive.

Pri tumačenju standardizirane razlike trebamo razmotriti i ostala obilježja po kojima se skupine razlikuju. U našoj populaciji korisnici društvenih mreža imaju prosječno 33,4 godine, a korisnici televizije 50,2 godine. Povjerenje pritom raste s dobi. Dob je moguća konfundirajuća varijabla, u smislu objašnjenom u poglavlju o mjerenju i dizajnu.

Provjerimo zato osjetljivost usporedbe na dobni raspon. U cijeloj simuliranoj populaciji razlika iznosi 1,30 boda. Među osobama od trideset do četrdeset devet godina, gdje su dobne razlike skupina manje, iznosi 0,90 boda. Time smo promijenili ciljnu populaciju. Nismo izmjerili udio razlike koji je uzrokovan dobi niti odvojili uzročne učinke dobi i izvora vijesti.

U uzorku od 120 korisnika televizije i društvenih mreža tom dobnom rasponu pripada 55 od 120 osoba. Razlika sredina iznosi 0,77 boda. Interval ide od −0,34 do 1,88 i obuhvaća nulu. Ciljana razlika je manja, a zbog manjeg broja osoba procjena je manje precizna. Za planiranje uzorka potrebnog za takvu razliku koristimo znanje o snazi i preciznosti.

Zaključak vrijedi za ovdje korištene mjere i populaciju koju nacrt može dosegnuti. Za odluku trebamo i sadržajni kriterij. Interval od nepunih pola do gotovo dva boda može biti dovoljno informativan kada nas zanima bilo koja pozitivna razlika, a nedovoljan

kada tražimo najmanje dva boda. Taj kriterij moramo obrazložiti prije čitanja rezultata.

Što se onda promijenilo nakon deaktivacije Facebooka? Objavljena analiza pokusa s početka poglavlja pokazala je porast subjektivne dobrobiti, uz smanjenje poznavanja činjenica iz vijesti (Allcott i sur. 2020.). Ista je intervencija, dakle, imala različite posljedice za dva ishoda do kojih nam može biti stalo. Savjet osobi koja razmatra prekid uporabe ovisio bi i o tome koji joj je ishod važniji te kako bi nadomjestila ono što gubi.

Procjena se odnosi na sudionike uključene u taj pokus i razdoblje deaktivacije, uz usporedbu omogućenu nasumičnom dodjelom (Allcott i sur. 2020.). Prosječna razlika ne obećava jednaku promjenu svakom korisniku. Za pitanje kome je korist bila najveća trebala bi unaprijed obrazložena usporedba učinaka među podskupinama i njezina nesigurnost. U našim simuliranim medijskim podacima korisnici su sami odabrali izvor vijesti. Zato dvije analize mogu obje uspoređivati sredine, a pružati različitu osnovu za zaključak o tome što bi promjena navike izazvala.

Pretpostavke i njihove granice

Za neovisnu usporedbu provjeravamo da jedinice nisu povezane i da svaka osoba pripada samo jednoj skupini. Ishod treba biti ispravno kodiran, referentna skupina jasno određena, a uzorkovanje primjereno ciljnoj populaciji. Te provjere određuju koliko se možemo pouzdati u rezultat odabranog testa.

U malim uzorcima t-aproksimacija ovisi o približnoj normalnosti ishoda unutar skupina. U velikima su važne konačne varijance i odsutnost opažanja koje samo određuje rezultat. Kod uparenog postupka pregledavamo raspodjelu razlika unutar parova.

Welchov postupak zasebno koristi varijancu i veličinu svake skupine te ne zahtijeva jednake varijance. Uobičajena inferen-

cija običnog linearnog modela koristi zajedničku rezidualnu varijancu i pretpostavlja jednake uvjetne rezidualne varijance.

Budući da veličine i varijance naših skupina nisu jednake, koristimo Welchov postupak. Obični linearni model daje istu procjenu razlike, ali nesigurnost računa drukčije. U razrađenom primjeru usporedit ćemo standardne pogreške, stupnjeve slobode i intervale te jasno navesti pretpostavku zajedničke rezidualne varijance.

Raspodjelu najprije pregledavamo grafički. Test normalnosti ne treba sam određivati izbor analize. U velikom uzorku može otkriti nevažno odstupanje, a u malom propustiti ozbiljno. To vrijedi i za Shapiro-Wilkov test.

Krajnje vrijednosti pregledavamo pojedinačno. Mogu biti pogrešno unesene, rijetke ali stvarne ili znak da sredina loše opisuje ishod. Opravdanu provjeru s njima i bez njih treba prikazati s oba rezultata. Opažanja ne uklanjamo tiho prema tome kako mijenjaju p-vrijednost.

Možemo se pitati nose li zaključak velike brojčane udaljenosti ili dosljednost smjera promjene. Za tu provjeru koristimo Wilcoxonov postupak predznačenih rangova. Uzimamo iste promjene 60 osoba, poredamo njihove apsolutne iznose i vratimo svakom rangu predznak promjene. Prva osoba dobiva rang 41 i pozitivan predznak. Zatim uspoređujemo zbrojeve rangova pozitivnih i negativnih promjena. U ovom skupu nema nultih razlika ni vezanih rangova. Za tumačenje kao provjere središta promjena potrebno je i opravdanje simetričnosti njihove raspodjele.

Uz pretpostavku simetrične raspodjele postupak provjerava je li središte promjena jednako nuli. To nije isto pitanje kao provjera prosječne promjene. U našem primjeru daje p-vrijednost 0,0005. Zaključci su slični, ali odnose se na različite veličine.

● STATISTIKA U DIVLJINI

Preklapanje crtica. Na grafu su dvije sredine i oko svake crtice pogreške. Čitatelj želi znati razlikuju li se skupine, pa pogleda dodiruju li se crtice.

Takav brz postupak može djelovati uvjerljivo dok ne saznamo što crtice prikazuju i kako su mjerenja nastala. Ista slika ne mora predstavljati istu usporedbu.

Problem se pokazao i u istraživanju u kojem su grafove čitali autori znanstvenih radova. Internetski zadatak dovršila su 473 autora iz psihologije, bihevioralne neuroznanosti i medicine (Belia i sur. 2005.). Trebali su pomaknuti prikazane sredine do razlike koju smatraju upravo statistički značajnom (Belia i sur. 2005.). Mnogi nisu dovoljno razlikovali standardne pogreške od intervala pouzdanosti ni neovisne skupine od ponovljenih mjerenja (Belia i sur. 2005.).

Posljedicu možemo razmotriti na zamišljenoj procjeni radionice. Ako graf prikazuje rezultate istih polaznika prije i poslije radionice, zanima nas promjena svake osobe. Interval za tu promjenu ovisi i o povezanosti njezinih dvaju mjerenja. Odvojene crtice oko sredine prije i sredine poslije tu informaciju ne prenose. Zaključak da radionica nije povezana s promjenom, izveden samo iz preklapanja crtica, zato preskače potrebnu uparenu analizu.

Kod dviju neovisnih skupina trebamo drukčiji račun nesigurnosti razlike. I tu opis crtica prethodi zaključku. Standardna devijacija opisuje raspršenost opažanja, standardna pogreška nesigurnost procjene sredine, a interval pouzdanosti određen je postupkom s odabranom razinom pouzdanosti. Svaku oznaku zato čitamo prema njezinu imenovanom značenju.

Autor izvještaja može pomoći čitatelju tako da uz graf navede nacrt i značenje crtica te izravno prikaže procjenu razlike s njezinim intervalom. Ako je radionica procijenjena samo prije i poslije, ni dobro izračunata promjena sama ne isključuje druga objašnjenja. Prije zaključka o dvjema sredinama zato procjenjujemo nesigurnost njihove razlike. Preklapanje dvaju zasebnih raspona nije dovoljno.

NAĐITE GREŠKU

Za izvještaj o televiziji i društvenim mrežama pregledajte usporedbu među osobama od 30 do 49 godina iz našeg simuliranog uzorka. Ishod je povjerenje u medije na ljestvici od jedan do deset, a razlika se računa kao televizija minus društvene mreže. Svaka osoba pripada jednoj skupini.

Tablica 3: Usporedba dviju skupina u simuliranom uzorku. Izrada autora.

Rezultat u dobnoj podskupini	Vrijednost
Ukupan broj osoba	55
Procijenjena razlika sredina	0,77 boda
Welchov 95-postotni interval razlike	od −0,34 do 1,88 boda

Asistentov osvrt na postupak i rezultat glasi ovako.

> Jedinica neovisnosti je osoba. Referentna skupina su društvene mreže, pa pozitivan koeficijent uz TV znači višu sredinu u TV skupini. Raspodjele i neobična opažanja pregledani su, a Welchov test ne traži jednake varijance. Interval razlike obuhvaća nulu, pa u toj dobnoj skupini glavni izvor vijesti nema veze s povjerenjem.

Uz nulu navedite još dvije različite vrijednosti razlike koje pripadaju prikazanom intervalu. Upotrijebite ih u obrazloženju je li posljednja rečenica prihvatljiva. Zatim napišite zamjenu koja zadržava smjer usporedbe, procjenu i granice intervala, bez odbacivanja valjanih provjera postupka.

Razrađeni primjer

U uzorku od 120 korisnika televizije i društvenih mreža zadržavamo iste osobe, isto povjerenje i istu ciljanu razliku. Najprije

ćemo je procijeniti Welchovim postupkom, a zatim zapisati kao koeficijent uz binarnu oznaku skupine. Budući da se točkasta procjena time ne mijenja, možemo se usredotočiti na pretpostavke i račun njezine nesigurnosti.

U tablici uspoređujemo procjenu i njezin interval prije čitanja programskog računa. Sve su vrijednosti u bodovima povjerenja, a stupci Donja i Gornja granice su 95-postotnog intervala. Razlika je u oba retka televizija minus društvene mreže. Za glavni zaključak zadržavamo Welchov postupak, dok drugi redak pokazuje što se mijenja kada model pretpostavi zajedničku rezidualnu varijancu.

Tablica 4: Ista procjena razlike i dva računa nesigurnosti u simuliranom uzorku. Izrada autora.

Postupak	Razlika	SE	df	Donja	Gornja
Welch	1,186	0,373	102,471	0,447	1,925
Linearni model	1,186	0,370	118,000	0,454	1,917

U izvještaju možemo navesti da u uzorku od 120 korisnika televizije i društvenih mreža korisnici televizije imaju prosječno 1,19 boda više povjerenja od korisnika društvenih mreža, uz Welchov 95-postotni interval od 0,45 do 1,92 boda. To je razlika među skupinama koje same biraju izvor vijesti. Nije procjena promjene povjerenja koju bi izazvalo prebacivanje iste osobe na televiziju. Cijeli interval ostaje ispod dva boda, pa podaci ne podupiru odluku koja zahtijeva najmanje toliku razliku.

Welchov postupak daje razliku 1,186 boda, standardnu pogrešku 0,373 i 102,471 stupnja slobode. Koeficijent uz TV jednak je toj razlici jer su društvene mreže referenca. Obična standardna pogreška modela iznosi 0,370, uz 118 stupnjeva slobode. Njegov interval ide od 0,454 do 1,917, dok Welchov ide od 0,447 do 1,925. Procjene su jednake, ali njihova nesigurnost nije. Ovdje su intervali bliski jer su skupne standardne devijacije slične. Uz izrazito različite varijance i brojnosti razlika u standardnim pogreškama

može biti mnogo veća. I pri proširenju na više skupina moramo zasebno odabrati način računanja nesigurnosti.

Uz brojke navodimo da ljudi sami biraju izvor vijesti i da se skupine razlikuju po dobi. Time čitatelju objašnjavamo zašto rezultat opisuje razliku među skupinama i ne utvrđuje učinak medija.

Za provjeru računa slijedi isti postupak u R-u. Prvi prolaz kroz primjer može završiti prethodnim tumačenjem. Ovdje provjeravamo kako program dobiva procjenu i zašto se njegova dva intervala razlikuju.

```
tv <- filter(dvije_skupine, izvor == "TV") |>
  pull(povjerenje_medijima)
mreze <- filter(dvije_skupine, izvor == "društvene mreže") |>
  pull(povjerenje_medijima)
welch <- t.test(tv, mreze)
model <- lm(povjerenje_medijima ~ izvor, data = dvije_skupine)

welch
coef(model)
summary(model)$coefficients["izvorTV", ]
df.residual(model)
confint(model)["izvorTV", ]
```

Oba računa uspoređuju isto povjerenje u istim skupinama. Poziv `t.test` prima njihove vrijednosti odvojeno i u zadanoj postavci provodi Welchov postupak. U pozivu `lm` zapis `povjerenje_medijima ~ izvor` povezuje brojčani ishod s oznakom skupine. Uobičajeni ispis toga modela računa nesigurnost uz pretpostavku jednake rezidualne varijance u skupinama, odnosno homoskedastičnosti. Pozivom `df.residual` izdvajamo i rezidualne stupnjeve slobode kako bismo mogli usporediti način računanja intervala. Kod je izvadak iz analize poglavlja i pretpostavlja već pripremljen objekt `dvije_skupine`. Vrijednost `welch$stderr` daje Welchovu standardnu pogrešku.

Sažetak

Ista brojčana razlika može potjecati iz različitih nacrta, pa usporedbu počinjemo jedinicom neovisnosti i odnosom među mjerenjima. Tako razlikujemo jednouzoračni, neovisni i upareni postupak te vidimo zašto zanemarivanje parova može znatno promijeniti procijenjenu nesigurnost. Razliku neovisnih sredina možemo zapisati i koeficijentom uz binarnu oznaku skupine, ali jednak koeficijent ne jamči jednak interval u Welchovu i običnom postupku najmanjih kvadrata (OLS). Standardizirana razlika pomaže usporediti razmjer, dok druga obilježja po kojima se skupine razlikuju i dalje ograničavaju tumačenje. U sljedećem poglavlju isti model proširujemo na više skupina, pa ćemo uz pojedinu razliku morati razmotriti i cijeli skup usporedbi.

Prediktor je varijabla kojom model opisuje razlike u ishodu. Homoskedastičnost označuje jednaku rezidualnu varijancu, pretpostavku obične modelne standardne pogreške o kojoj smo govorili uz usporedbu s Welchovim postupkom.

Pojmovi

homoskedastičnost (*homoskedasticity*), Wilcoxonov postupak predznačenih rangova (*Wilcoxon signed-rank test*), prediktor (*predictor*), jedinica neovisnosti (*unit of independence*), neovisne skupine (*independent groups*), upareni podaci (*paired data*), razlika sredina (*difference in means*), Welchov postupak (*Welch's t-test*), t-raspodjela (*t distribution*), kritična t-vrijednost (*critical t value*), stupnjevi slobode (*degrees of freedom*), referentna skupina (*reference category*), standardizirana razlika (*Cohen's d*), ovisnost opažanja (*dependence among observations*)

Zadaci

Za odabrane zadatke ovoga poglavlja na kraju knjige dostupne su naznake i odvojena objašnjenja. Najprije pokušajte samostalno odgovoriti.

Konceptualni

Za tri istraživačke situacije imenujte jedinicu neovisnosti i odgovarajući dizajn. Prva uspoređuje prosječno povjerenje u uzorku s unaprijed zadanim pragom od pet bodova, druga uspoređuje dvije skupine prema primarnom izvoru vijesti, a treća mjeri iste ispitanike prije i poslije kampanje.

Računski

Iz upravljanoga agregata data/populacija-medija-agregat.csv uzmite retke za TV i društvene mreže. Za svaki redak podijelite `zbroj_povjerenja` stupcem `broj`, provjerite pohranjeni prosjek i izračunajte razliku aritmetičkih sredina. Zatim pretpostavite jednake standardne devijacije 1,6 pa 3,2 i za obje izračunajte standardiziranu razliku. Objasnite zašto se prva brojka nije promijenila, a druga jest. Za rad bez datoteke upotrijebite izvadak iz istoga agregata.

Tablica 5: Izvadak upravljanoga agregata za obvezni zadatak.

Izvor	Broj	Zbroj povjerenja	Pohranjeni prosjek
društvene mreže	13378	54432	4,068770
TV	10827	58098	5,366029

Kritički

Vratite se operacionalizaciji iz poglavlja o mjerenju i istraživačkom dizajnu i tumačenju intervala iz poglavlja o procjeni. Za glavni primjer najprije napišite što ljestvica povjerenja i kategorija izvora vijesti ne mjere. Zatim za isti interval usporedite odluku kojoj je važna bilo koja pozitivna razlika s odlukom koja traži najmanje dva boda. Predajte kratku bilješku recenzentu i objasnite zašto sama oznaka značajnosti ne može riješiti ni problem mjerenja ni izbor praga (Cumming 2014.).

Revizija modela

Ocijenite usporedbu dviju skupina iz okvira. Imenujte referentnu skupinu i objasnite što bi se promijenilo, a što ostalo isto kada bi se referenca zamijenila. Zatim provjerite jedinicu neovisnosti, vrstu ishoda, odnos prema varijancama i doseg ciljne populacije. Izdvojite jedinu tvrdnju koja iz rezultata ne slijedi i napišite rečenicu kojom bi je trebalo zamijeniti.

15

DIO V · MODELI

uspoređivanje više grupa

Jedan model za razlike među skupinama

Pred vama je tekst koji trebate razumjeti i zapamtiti za ispit. Možete ga ponovno pročitati, nacrtati veze među pojmovima ili zatvoriti knjigu i pokušati zapisati čega se sjećate. Posljednji je pokušaj možda najmanje ugodan jer odmah otkriva praznine. Ipak, osjećaj lakoće tijekom učenja ne mora nam reći koliko ćemo znati poslije.

U jednom pokusu studenti su učili prirodoznanstveni tekst u četiri uvjeta. Jedni su ga pročitali jednom, drugi više puta, treći su uz otvoren tekst izrađivali pojmovnu mapu, a četvrti izmjenjivali čitanje i prisjećanje bez teksta. Nakon tjedan dana svi su rješavali provjeru znanja (Karpicke i Blunt 2011.). Zastanimo i predvidimo koja bi skupina najbolje odgovorila na pitanja koja traže povezivanje ideja.

Kako usporediti sve te načine učenja i utvrditi između kojih je skupina razlika dovoljno uvjerljiva da bismo zbog nje promijenili vlastitu naviku?

Četiri načina učenja otvaraju šest parnih usporedbi. S većim brojem skupina broj pitanja brzo raste. Ista se poteškoća pojavljuje kad proširimo poznati primjer medijskih navika. Usporedba televizije i društvenih mreža ostavila je izvan analize korisnike ostalih izvora vijesti. Ako želimo obuhvatiti sve skupine, mogli bismo ponavljati postupak iz prethodnog poglavlja za svaki par. Tada donosimo niz povezanih odluka, pa moramo razmotriti koliko prilika za lažni nalaz nastaje kroz sve usporedbe koje ćemo pregledati. Tu ćemo poteškoću najprije ispitati u simulaciji u kojoj skupine dolaze iz iste raspodjele povjerenja. Svaku pronađenu razliku tada možemo prepoznati kao lažni nalaz. Povezat ćemo tu provjeru s onime što smo u poglavlju o krizi i obnovi naučili o odabiru rezultata. Ovdje je skup usporedbi unaprijed poznat, pa možemo tražiti postupak koji uvažava cijeli skup odluka.

Usporedba televizije i društvenih mreža opisala je jednu razliku u povjerenju. Sada nas zanimaju razlike među svim izvorima i nesigurnost tih procjena.

Uzorak od 300 osoba s pet izvora vijesti zasebno izvlačimo iz nastavne populacije `populacija_medija`. To je novo izvlačenje, a ne podskup ranijeg uzorka. Jedan redak predstavlja zamišljenu osobu s glavnim izvorom vijesti i povjerenjem na ljestvici od jedan do deset. Rezultati opisuju taj nastavni skup, bez tvrdnje o stvarnoj publici.

Umjesto jedne razlike sredina sada ispitujemo jesu li sve skupinske sredine jednake. Razlaganjem varijabilnosti usporedit ćemo razmake među sredinama s raspršenošću osoba unutar skupina. Iz te ćemo usporedbe razviti analizu varijance i F-test, uz mjeru veličine učinka. Ukupni nalaz još neće odrediti koji se parovi razlikuju. Za njih ćemo prikazati procjene i intervale koji uzimaju u obzir višestruke usporedbe, pa razmotriti i granice pretpostavki.

Tako možemo sastaviti odgovor koji razlikuje ukupni obrazac, konkretne razlike i nesigurnost svake opravdane usporedbe.

Cijena mnogih usporedbi

Među pet skupina postoji deset parova. Mogli bismo svaki usporediti zasebnim testom iz prethodnog poglavlja. Simulacijom ćemo provjeriti koliko često taj niz testova daje barem jedan lažni nalaz.

Iz poznate simulirane populacije medijskih navika nasumično sastavljamo pet skupina po četrdeset osoba. Pri tome zanemarujemo njihov izvor vijesti. Sve skupine nastaju odabirom iz iste raspodjele povjerenja, pa među njima nema unaprijed zadane razlike. U svakom ponavljanju provodimo deset usporedbi i bilježimo je li barem jedna značajna. Svako takvo odbacivanje ovdje je pogrešno. Poslije ćemo se vratiti skupinama prema izvoru vijesti, među kojima razlike mogu postojati.

Jedno ponavljanje stvara pet skupina po 40 odgovora. Unutar svake skupine biramo bez vraćanja, ali svaku novu skupinu ponovno izvlačimo iz pune populacije. Ista osoba zato može biti odabrana u različitim skupinama. U prvom stvarno izvedenom ponavljanju sredine su 5,47, 5,08, 4,85, 5,50, 4,85. Među deset obostranih Welchovih usporedbi najmanja p-vrijednost iznosi 0,164, a prag 0,05 prelazi u smjeru odbacivanja 0 usporedbi. Za zbirni rezultat zapisujemo samo ima li barem jedne takve usporedbe. Dva lažna nalaza u istom ponavljanju i dalje znače jedno ponavljanje s pogreškom.

Ponovno izvlačimo svih pet skupina 2 000 puta, zadržavajući istu populaciju, veličine i pravila testiranja. Barem jedan lažni nalaz dobivamo u 548 od 2 000 ponavljanja. Svi su testovi definirani i sva ponavljanja ulaze u račun.

U dvije tisuće ponavljanja barem jedna usporedba bila je značajna u 27,4 % slučajeva. Svaki test koristi prag pet posto, ali

cijeli skup daje lažni nalaz u više od četvrtine ponavljanja. Prikaz samo značajnog para sakrio bi tu veću stopu pogreške.

> **Stopa obiteljske pogreške** je vjerojatnost da će barem jedan test u unaprijed određenom skupu testova dati lažno pozitivan rezultat, odnosno pogrešno odbaciti barem jednu istinitu nultu hipotezu.

U našoj su simulaciji sve nulte hipoteze istinite. Kada neke razlike postoje, pratimo vjerojatnost da pogrešno odbacimo barem jednu od preostalih istinitih nultih hipoteza.

Postupak koji kontrolira pogrešku samo kada su sve nulte hipoteze istinite ima slabu kontrolu. Jaka kontrola mora vrijediti i kada dio razlika doista postoji. Tukeyjev postupak u odgovarajućem normalnom modelu sa zajedničkom varijancom štiti obitelj parnih usporedbi i u tom drugom slučaju. Primjerice, ako se samo prva od pet skupina razlikuje od ostalih, ukupni test može lako odbaciti jednakost svih sredina. Među preostalim četirima ipak postoji šest usporedbi jednakih populacijskih sredina. Njihovo zasebno testiranje bez korekcije i dalje otvara mogućnost barem jednoga lažno pozitivnog zaključka.

Za deset neovisnih testova vjerojatnost da nijedan ne pogriješi iznosi $0{,}95^{10}$. Suprotni događaj, barem jedna pogreška, zato ima vjerojatnost $1 - 0{,}95^{10}$, približno 0,401.

Za deset neovisnih testova račun daje stopu 40,1 %. U parnim usporedbama pet skupina ista se opažanja koriste više puta, pa usporedbe nisu neovisne i u našoj simulaciji daju nižu stopu. Zasebna simulacija deset doista neovisnih usporedbi daje 40,4 %, blizu računu.

U zasebnom pokusu svaki od deset testova dobiva svoja dva nova uzorka po 40 osoba, pa testovi više ne dijele iste uzorke. Od 2 000 ponavljanja 808 sadrži barem jedno odbacivanje, bez isključenih ponavljanja. Uspoređujemo stope pogreške dviju organizacija testiranja pod nulom. Za usporedbu snage trebali bismo zadati model s razlikom. Novi niz izvlačenja dao bi malo drukčije udjele.

Formula za neovisne testove nije opća gornja granica za ovisne usporedbe. Ipak, oba simulirana skupa daleko prelaze pet posto. Zato uz najmanju p-vrijednost trebamo znati koliko je usporedbi provedeno i kako su povezane.

Za stopu blizu 5 % uz 2 000 ponavljanja standardna pogreška samoga simulacijskog postotka iznosi oko 0,49 postotnih bodova. Mala odstupanja od 5 % zato nisu sama po sebi dokaz pogrešno podešena postupka.

Možemo postaviti i jedno ukupno pitanje o jednakosti svih pet sredina. U istoj simulaciji takav test odbacuje u 4,2 % slučajeva. Uz dvije tisuće ponavljanja taj je udio u skladu s planiranih pet posto.

Varijanca između i unutar

Ukupni test treba prepoznati jesu li razmaci među skupnim sredinama veliki u odnosu na raspršenost koju već imamo među ljudima. Taj smo odnos susreli pri usporedbi dviju skupina. Razlika od jednoga boda jače ih statistički razdvaja kada su osobe unutar skupina slične nego kada su jako raspršene. Sada istu logiku primjenjujemo na sve skupine zajedno.

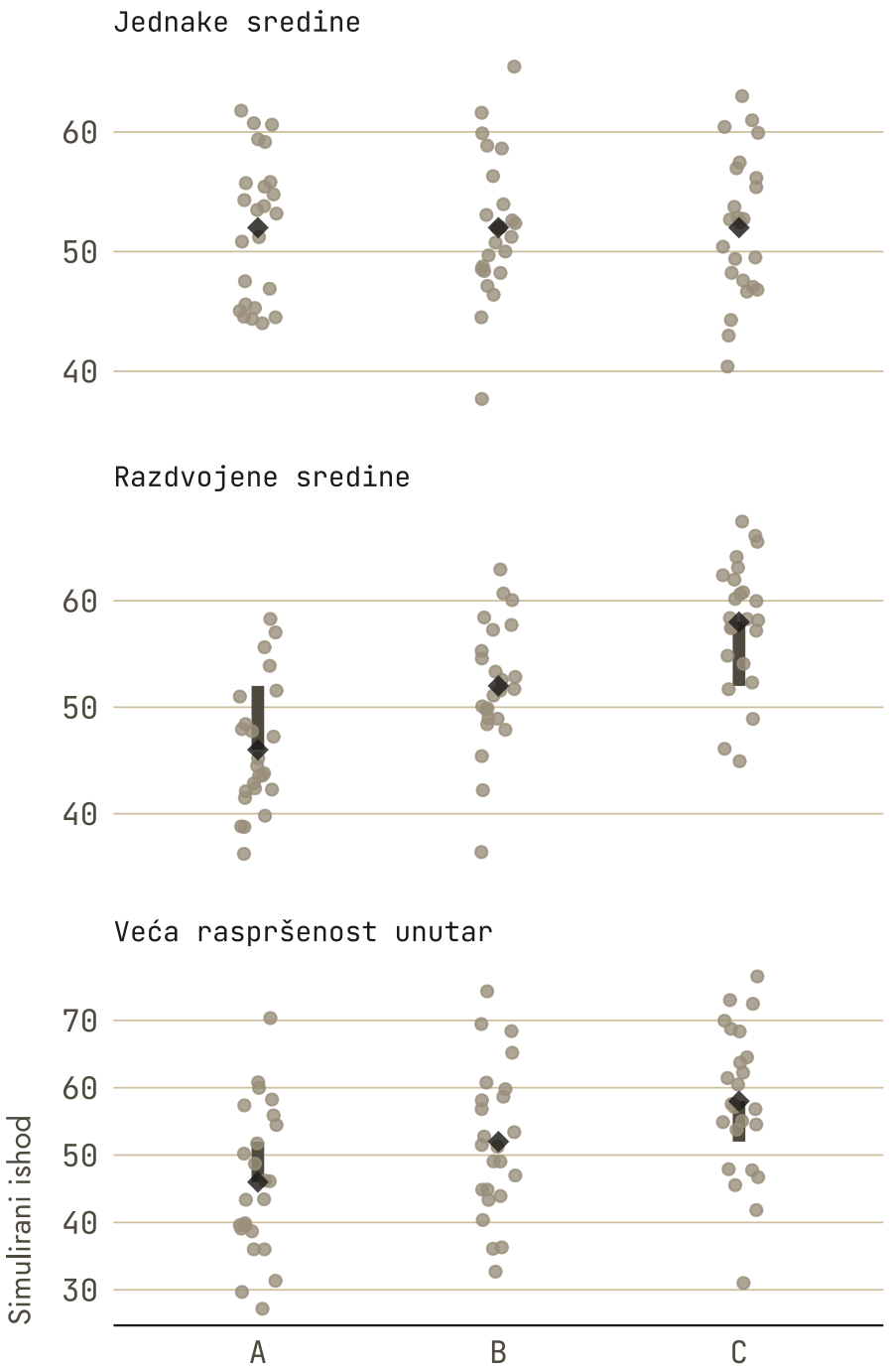
Ukupnu varijabilnost zato razdvajamo na odstupanja među skupnim sredinama i odstupanja pojedinaca unutar skupina. Model sa skupinama opisuje prvi dio, dok drugi ostaje neobjašnjen.

Da bismo razlaganje vidjeli na manjem prikazu, prelazimo na tri skupine po 24 točke bez stvarne mjerne jedinice. Počinjemo normalnim slučajnim vrijednostima, ali ih zatim u svakoj skupini centriramo i skaliramo tako da njihova ostvarena sredina i standardna devijacija budu točno zadane. Tako dobivamo ilustraciju razlaganja s točno zadanim sažecima. Pri sredinama 46, 52 i 58 zajednička je sredina 52. Skupina A pridonosi zbroju između skupina s $24(46 - 52)^2 = 864$. Isti račun radimo za B i C, dok unutar skupina zbrajamo kvadrate udaljenosti svake od 72 točke od njezine skupne sredine.

Zbroj između skupina iznosi $864 + 0 + 864 = 1\,728$. Dijeljenjem s $3 - 1 = 2$ dobivamo srednji kvadrat 864. Uz standardnu devijaciju 6, zbroj unutar skupina iznosi $3 \times 23 \times 6^2 = 2\,484$. Nakon dijeljenja sa $72 - 3 = 69$ srednji kvadrat iznosi 36. Omjer je zato $F = 864/36 = 24$. Srednji kvadrat u nazivniku ima istu mjernu jedinicu kao onaj u brojniku, pa je F bez mjerne jedinice.

Kad su sve tri sredine 52, zbroj između skupina iznosi nula, zbroj unutar ostaje $2\,484$ i F je nula. U trećem stanju zbroj između ostaje $1\,728$, a uz standardnu devijaciju 11 zbroj unutar iznosi $3 \times 23 \times 11^2 = 8\,349$. Srednji kvadrat unutar skupina tada je 121, pa je $F = 864/121 \approx 7{,}14$.

Okomite trake prikazuju udaljenost skupnih sredina od zajedničke sredine. Točke prikazuju raspršenost osoba oko sredine njihove skupine. Usporedba stanja pokazuje kako isti razmak sredina izgleda uz različitu raspršenost unutar skupina.



Slika 1: Dekompozicija varijance — razmak sredina povećava omjer raspršenosti između i unutar skupina, a veća raspršenost unutar skupina smanjuje ga. Izrada autora. Podaci i postupak opisani su u tekstu.

U trećem stanju sredine su iste kao u drugome, ali je raspršenost unutar skupina veća, pa omjer pada. Uz standardnu devijaciju 11 njegova je vrijednost $864/121 \approx 7{,}14$, umjesto $864/36 = 24$ uz standardnu devijaciju šest. Za procjenu odstupanja od nultog modela trebamo oba izvora varijacije.

Veličina i F-omjer

Zbroj kvadriranih odstupanja označujemo s SS. Dio $SS_{\text{između}}$ sažima odstupanja skupnih sredina od zajedničke, a SS_{unutar} odstupanja pojedinaca od njihove skupne sredine. Ukupni zbroj SS_{ukupno} jednak je zbroju tih dijelova. Udio prvog dijela u ukupnom pokazuje koliko je varijabilnosti povezano sa skupinama.

> **Eta-kvadrat** je udio ukupne varijabilnosti ishoda koji otpada na razlike među skupnim sredinama.
>
> $$\eta^2 = \frac{SS_{\text{između}}}{SS_{\text{ukupno}}}$$

Sada prelazimo s početne simulacije nasumično sastavljenih skupina na uzorak iz poznate simulirane populacije medijskih navika. Zadržavamo oznake pet izvora vijesti zadane pri stvaranju podataka i uspoređujemo povjerenje njihovih korisnika. Skupine zato mogu imati različite prosjeke i prije uzorkovanja.

Uzorak od 300 osoba s pet izvora vijesti izvukli smo bez vraćanja iz pune populacije, zasebno od uzorka od 120 korisnika televizije i društvenih mreža. Tablica 1 pokazuje broj odgovora i sredinu koju model pridružuje svakoj osobi iste skupine.

Tablica 1: Pet skupina u glavnom uzorku od 300 osoba. Izrada autora.

Izvor	Broj osoba	Sredina	SD
tisak	32	5,47	1,72
TV	63	5,43	1,78
radio	36	5,36	1,90
portal	99	4,82	1,73
društvene mreže	70	3,84	2,06

Prva uzorkovana osoba navodi tisak i povjerenje 5. Njezina skupna sredina je 5,469, pa u zbroj unutar skupina ulazi $(5 - 5{,}469)^2$. Za zbroj između skupina kvadriramo razliku svake skupne i zajedničke sredine te je množimo brojem osoba u toj skupini. Tako dobivamo 113,85 između i 1 001,70 unutar. Njihov zbroj, 1 115,55, nazivnik je eta-kvadrata. Svih 300 odgovora ulazi u oba opisa.

U našem uzorku eta-kvadrat iznosi 0,102. Razlike među izvorima vijesti opisuju oko 10 % varijabilnosti povjerenja. Većina razlika među ljudima ostaje izvan modela. Za tu mjeru ovdje ne prikazujemo interval, pa njezina uzoračka nesigurnost nije kvantificirana.

Je li takav udio važan, ovisi o pitanju. Može biti koristan za razumijevanje društvenog stava, a nedovoljan za precizno predviđanje pojedinca. Procjenu zato povezujemo s ranijim nalazima i unaprijed obrazloženim najmanjim važnim učinkom.

Eta-kvadrat opisao je koliki dio ukupne varijabilnosti otpada na razlike među skupinama. Za test jednakosti sredina trebamo još usporediti raspršenost između skupina s onom koja ostaje unutar njih. Sirove zbrojeve zato najprije pretvaramo u usporedive srednje kvadrate. Za k skupina i n opažanja dijelimo $SS_{\text{između}}$ s $k - 1$, a SS_{unutar} s $n - k$. Ti djelitelji predstavljaju stupnjeve slobode. Dobivamo $MS_{\text{između}}$ i MS_{unutar}, gdje MS označuje prosječnu kvadriranu raspršenost.

F-statistika je omjer prosječne raspršenosti među skupnim sredinama i prosječne raspršenosti opažanja oko njihovih skupnih sredina.

$$F = \frac{MS_{\text{između}}}{MS_{\text{unutar}}}$$

Pod nultim modelom jednakih populacijskih sredina i zajedničke varijance, uz neovisna opažanja, oba srednja kvadrata procjenjuju tu varijancu, pa se njihov omjer F obično nalazi blizu jedinice. Stupnjevi slobode $k - 1$ i $n - k$ određuju referentnu F-raspodjelu. Veći omjer daje više razloga za preispitivanje jednakosti sredina.

Referentna F-raspodjela opisuje koje bismo omjere dobivali pod tim nultim modelom uz normalne pogreške. Njezina dva parametra ovdje iznose 4 i 295. Gornjih 5 % raspodjele počinje kod 2,40, dok naš opaženi omjer iznosi 8,38. Tu usporedbu koristimo za test. Veličinu i važnost razlika i dalje čitamo iz sredina, eta-kvadrata i intervala.

Postupak se zove analiza varijance jer uspoređuje izvore varijabilnosti. Njime ipak ispitujemo pitanje o jednakosti skupnih sredina.

Isti model, više koeficijenata

U prethodnom poglavlju binarna oznaka skupine imala je jedan koeficijent razlike. Za pet skupina trebamo četiri takva koeficijenta. Jedna skupina ostaje referentna, a ostale uspoređujemo s njom. U našem je računu to portal.

Pokazatelji x_1 do x_4 označuju četiri nereferentne skupine. Svaki ima vrijednost jedan za svoju skupinu i nula za ostale. Koeficijent β_0 daje sredinu reference, a koeficijenti β_1 do β_4 razlike prema njoj. Populacijska pogreška ε opisuje odstupanje od modelne

skupne sredine. Nakon prilagodbe opaženo odstupanje zovemo rezidualom.

$$\text{povjerenje} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Svaka osoba pripada samo jednoj skupini. Pet koeficijenata tako opisuje pet skupnih sredina.

Klasični ukupni test zajednički ispituje četiri koeficijenta razlike. Uspoređuje model sa skupnim sredinama i model sa samo jednom zajedničkom sredinom. To je jedno globalno pitanje, različito od deset zasebnih parnih usporedbi.

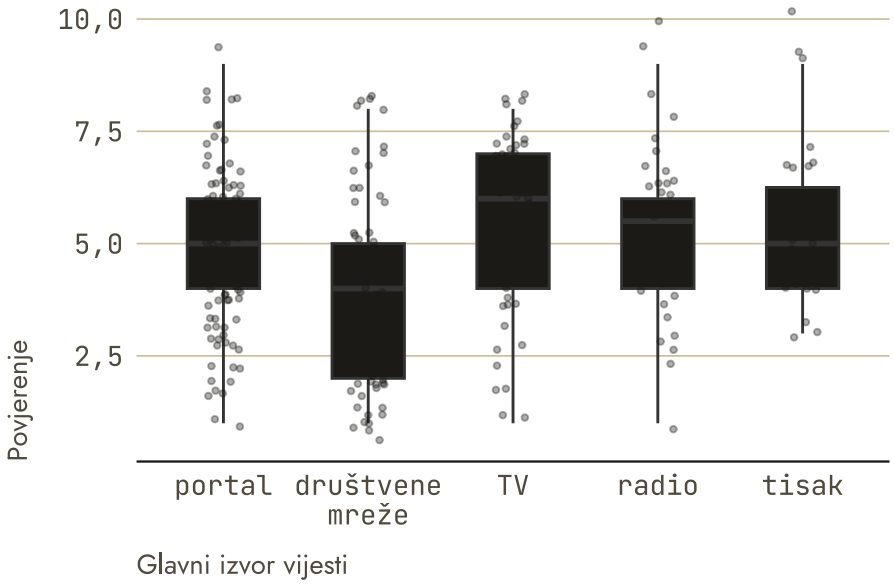
Klasični i Welchov postupak koriste iste točkaste procjene sredina, ali drukčije računaju nesigurnost. Klasična analiza varijance, kao obični homoskedastični linearni model, procjenjuje zajedničku rezidualnu varijancu. Welch zadržava zasebne skupne varijance i prilagođava stupnjeve slobode. Zato se njihovi intervali i p-vrijednosti mogu razlikovati.

Taj model primjenjujemo na isti uzorak od 300 osoba s pet izvora vijesti. Svakoj osobi pridružujemo sredinu njezina izvora vijesti, pa odstupanja oko tih pet sredina uspoređujemo s odstupanjima oko jedne zajedničke sredine.

Uzorak od 300 osoba dijelimo prema pet izvora vijesti. Ukupni test daje F od 8,38, uz 4 i 295 stupnjeva slobode te p-vrijednost manju od 0,001. Srednji kvadrat između skupina iznosi 28,5, a unutar njih 3,4. Opaženi rezultat teško se uklapa u pretpostavku jednakih populacijskih sredina. Još trebamo procijeniti pojedine razlike i njihovu nesigurnost.

Ljudi sami biraju izvor vijesti, pa se skupine mogu razlikovati i po drugim obilježjima. Model opisuje njihove razlike u povjerenju, ali ne utvrđuje učinak izbora izvora.

Oblik razlika provjeravamo na samim odgovorima. Kutije prikazuju srednjih 50 % vrijednosti, vodoravna crta medijan, a točke pojedinačne odgovore. Raspršenost i preklapanje ostaju vidljivi uz razlike sredina.



Slika 2: Povjerenje po glavnom izvoru vijesti u istom nastavnom uzorku. Izrada autora.

Nakon ukupnog testa

Za odgovor o konkretnim skupinama trebamo procjene njihovih razlika s intervalima i jasan popis pitanja koja smo postavili. Ukupni test odgovara na šire pitanje o jednakosti svih sredina. Unaprijed planiranu usporedbu možemo provesti izravno, neovisno o njegovu rezultatu.

Usporedbu određenu prije pregleda rezultata zovemo **planirana usporedba**. Za naknadne usporedbe navodimo **post-hoc postupak** i obitelj pitanja koju štiti. Tukeyjev postupak u klasičnom modelu sa zajedničkom rezidualnom varijancom obuhvaća sve parove. Daje istodobne intervale koji u toj obitelji kontroliraju stopu pogreške.

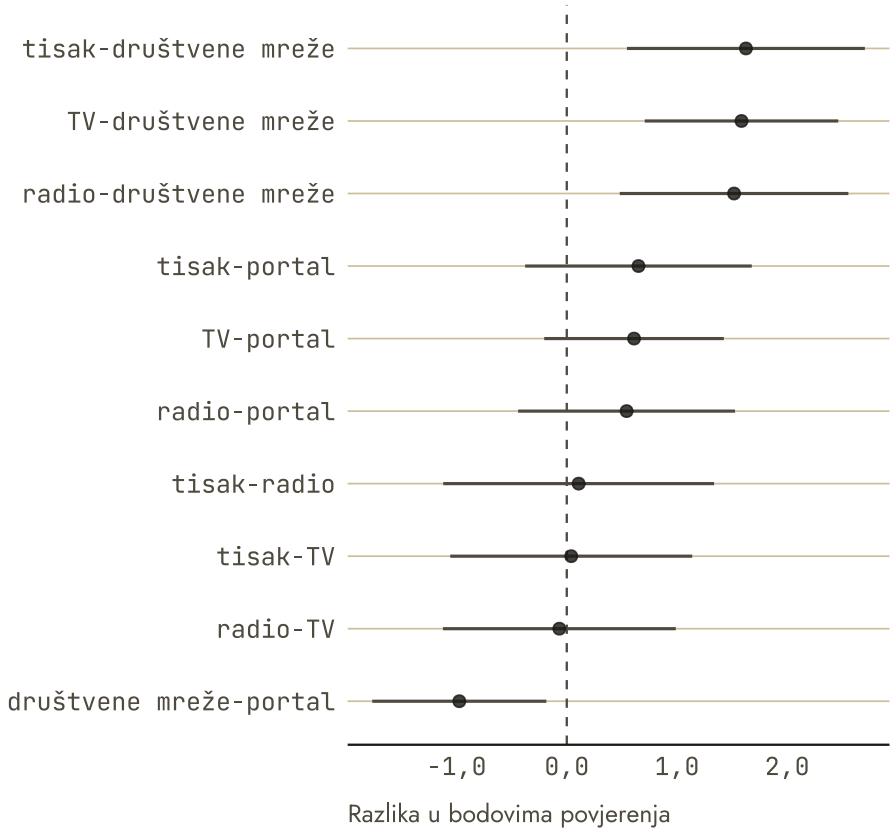
Zašto tim razlikama pridružujemo šire intervale nego jednoj zasebnoj usporedbi? Promatramo ih više, pa najveća među deset razlika može biti velika i kada nijednu nismo unaprijed odabrali.

Tukeyjev račun uzima u obzir taj izbor koristeći raspodjelu raspona među skupnim sredinama. Širina istodobnih intervala zato ovisi o broju skupina i povezanosti usporedbi.

Za jedan par sada oduzmemo dvije sredine iz tablice. Razlici televizije prema mrežama pridružujemo marginu 0,88 boda, izračunatu zajedničkom rezidualnom varijancom, brojnostima tih dviju skupina i Tukeyjevim pravilom za svih pet skupina. Oduzimanjem i dodavanjem te margine nastaju granice intervala. Svih deset usporedbi temelji se na odgovorima iz uzorka od 300 osoba s pet izvora vijesti.

Među 10 parova statistički su razlučive četiri usporedbe s društvenim mrežama. Televizija se od njih razlikuje za 1,59 boda, uz istodobni 95-postotni interval od 0,71 do 2,46. Razlika tiska prema televiziji iznosi 0,04 boda, uz interval od −1,06 do 1,14. Taj par ovim uzorkom ne razlučujemo.

Na zajedničkoj osi možemo pregledati svih deset razlika. U nazivu para prvu skupinu uspoređujemo s drugom, pa pozitivna vrijednost znači višu sredinu prve skupine. Crte su istodobni 95-postotni Tukeyjevi intervali. Okomita crta označuje izostanak razlike.



Slika 3: Svih deset parnih razlika iz istoga modela, uz istodobne intervale. Izrada autora.

Interval za televiziju i tisak širok je više od dva boda. Takav rezultat ostavlja otvoren raspon razlika u oba smjera, uključujući nulu. Za njihovo razlučivanje trebala bi nam preciznija procjena.

Ako nas unaprijed zanima određena usporedba, primjerice tradicionalnih i digitalnih izvora, možemo je izravno planirati. Tada ne štitimo obitelj od deset naknadno odabranih parova. Interval može biti uži, ali samo ako je pitanje doista određeno prije pregleda sredina. Naknadni izbor ne smijemo prikazati kao unaprijed planiran.

U pokusu s početka poglavlja najbolji rezultat na provjeri nakon tjedan dana postigla je skupina koja je vježbala prisjećanje. Pred-

nost se pojavila i na pitanjima koja su tražila povezivanje ideja iz teksta (Karpicke i Blunt 2011.). Studenti su tijekom učenja, međutim, predviđali najveću korist od ponovljenog čitanja, a najmanju od prisjećanja (Karpicke i Blunt 2011.). Vrijedi usporediti taj nalaz s vlastitim početnim predviđanjem.

Sada preciznije vidimo i što bi bilo potrebno za savjet o učenju. Ukupni test jednakosti četiriju sredina još ne bi pokazao da je prisjećanje bolje od izrade pojmovne mape. Ta tvrdnja traži upravo njihovu usporedbu, uz veličinu razlike i njezinu nesigurnost. U objavljenom pokusu vrijeme učenja bilo je izjednačeno za te dvije skupine (Karpicke i Blunt 2011.). Pri primjeni nalaza na vlastiti ispit dodatno razmatramo gradivo, način provjere i vrijeme do ispita. Možemo iskušati prisjećanje, a njegovu korist prosuđivati prema onome što kasnije uspijemo objasniti.

Granice pretpostavki

Najprije provjeravamo neovisnost jedinica. Ponovljena mjerenja istih ljudi i zajedničko kućanstvo, razred, škola, grad, organizacija ili mrežna veza mogu stvarati ovisnost. Obični klasični, Welchov i rangovni postupak takvu strukturu ne rješavaju. Modeli za ponovljene i ugniježđene podatke prelaze opseg knjige.

Zatim pregledavamo ostatke i skupne raspršenosti. Pretpostavka približne normalnosti odnosi se na ostatke. Klasična inačica pretpostavlja jednake varijance, što trebamo procijeniti iz više informacija, a ne iz jednog pokazatelja.

Omjer najveće i najmanje skupne varijance iznosi 1,43. Koristimo ga kao orijentir uz grafove, veličine skupina i način nastanka podataka. Valjanost klasičnog testa procjenjujemo iz svih tih informacija zajedno. Ako nejednake varijance izgledaju uvjerljivo, Welchov ukupni test zadržava njihove zasebne procjene.

Welchov test daje F od 7,32, uz četiri stupnja slobode u brojniku i 112,4 u nazivniku. Procijenjene sredine i razlike ostaju iste. Provjerili smo osjetljivost ukupnog zaključka na račun nesigurnosti, ali time nismo dokazali jednake varijance ni opravdali Tukeyjeve intervale. Ako varijance nisu jednake, za parove možemo provesti Welchove usporedbe uz Bonferronijevu korekciju. Svaku uspoređujemo s pragom 0,05 podijeljenim brojem planiranih parova. To je zasebno od Tukeyjeva postupka.

Kruskal-Wallisov postupak rangira vrijednosti i uspoređuje raspodjele rangova. Utjecaj krajnjeg opažanja zato više ne ovisi izravno o njegovoj brojčanoj udaljenosti, ali postupak i dalje ima pretpostavke i ispituje drugo pitanje. U našem uzorku statistika iznosi 29,8. Uspoređujemo je s približnom hi-kvadrat raspodjelom uz 4 stupnjeva slobode, jedan manje od broja skupina. Na razini 0,05 kritična vrijednost te hi-kvadrat raspodjele iznosi 9,49. To nije F-raspodjela s dvama brojevima stupnjeva slobode. Za tumačenje razlike u položaju potrebni su usporedivi oblici raspodjela. Rangiranje također ne uklanja ovisnost redaka.

● STATISTIKA U DIVLJINI

Značajno ovdje, neznačajno ondje. Pretpostavimo da istražujemo program pomoći pri učenju u dvjema dobnim skupinama. U svakoj uspoređujemo sudionike programa s odgovarajućom kontrolnom skupinom. Prvi rezultat označen je kao značajan, a drugi nije. U sažetku zato napišemo da program djeluje samo na mlađe i predložimo da se ubuduće nudi samo njima. Ta je odluka ovdje zamišljena, ali pogreška na kojoj počiva dokumentirana je u istraživačkoj literaturi.

Nieuwenhuis, Forstmann i Wagenmakers pregledali su 513 neuroznanstvenih radova iz pet vodećih časopisa. Među njima su pronašli 78 koji su izravno testirali razliku dvaju učinaka i 79 koji su je zaključili iz toga što je jedan učinak bio značajan, a drugi nije (Nieuwenhuis i sur. 2011.). Zasebni testovi odgovaraju razlikuje li se svaki učinak od svoje nulte vrijednosti. Ne odgovaraju razlikuju li se učinci međusobno.

U našem bi primjeru procjene koristi programa mogle biti slične, dok bi jedna imala širi interval zbog manjeg uzorka ili veće raspršenosti. Tada

bi različite oznake značajnosti odražavale različitu preciznost, a ne dokaz da dob mijenja korist programa. Kad bismo prema tim oznakama suzili buduću primjenu, sadržajnu bismo odluku oslonili na usporedbu koja još nije provedena.

Potrebno je izravno procijeniti razliku između dvaju učinaka i njezinu nesigurnost. Kao i pri razlikovanju ukupnog testa od pojedinačnih usporedbi, prvo određujemo koju razliku tvrdnja traži. Pritom treba opravdati nacrt unutar svake dobne skupine. Smislenost dodatnog računa ovisi o kvaliteti tako osmišljene usporedbe.

Ako je interval razlike učinaka širok, možda još ne možemo odlučiti treba li program prilagoditi dobi. To je korisnije znati nego jednu skupinu proglasiti neprikladnom za program. Razlika između dviju oznaka značajnosti može biti povod za pitanje, ali odgovor traži procjenu razlike o kojoj stvarno odlučujemo.

NAĐITE GREŠKU

Nakon ukupnog testa razlika u povjerenju među korisnicima pet izvora vijesti tim želi pregledati sve parove. Podaci su iz poznate simulacije. U ovoj nastavnoj provjeri pretpostavke o neovisnosti, raspodjelama i jednakim varijancama smatramo zadovoljenima. Zadatak je procijeniti postupak donošenja odluka o parovima.

Asistent je pripremio ovaj zapis, koji treba čitati, a ne izvršavati.

```
model <- aov(povjerenje_medijima ~ izvor_vijesti,
data = pet_izvora)
pairwise.t.test(
  pet_izvora$povjerenje_medijima,
  pet_izvora$izvor_vijesti,
  p.adjust.method = "none"
)
```

Ukupni test ispituje jesu li sve skupne sredine jednake. S pet skupina imamo deset parova. Nakon značajnog ukupnog testa

svaki par možemo prosuditi na razini 0,05 bez prilagodbe, a da vjerojatnost barem jedne pogrešne tvrdnje o parovima ostane zaštićena na toj razini.

Odvojite pitanje ukupnog testa od pitanja pojedinačnih parova. Navedite koju tvrdnju o zaštiti treba popraviti i predložite postupak iz poglavlja kojim bi tim izvijestio parne razlike uz odgovarajuću nesigurnost. Sačuvajte ispravno određene skupine i broj usporedbi. Ne pišite zamjenski kod.

Razrađeni primjer

Na istom uzorku od 300 osoba s pet izvora vijesti želimo razdvojiti dva pitanja o povjerenju. Prvo se odnosi na ukupan obrazac skupnih sredina, a drugo na veličinu konkretnih razlika među izvorima. Najprije ćemo prikazati veličinu učinka i ukupni test, a zatim parne razlike s istodobnim intervalima. Tako će uz svaku usporedbu ostati vidljivo da je dio šireg skupa odluka.

```
model <- aov(povjerenje_medijima ~ izvor_vijesti,
             data = pet_izvora)
tablica <- summary(model)[[1]]
eta2 <- tablica$`Sum Sq`[1] / sum(tablica$`Sum Sq`)

eta2
```

```
[1] 0,1020532
```

```
tablica
```

```
               Df  Sum Sq Mean Sq F value      Pr(>F)
izvor_vijesti   4  113,85 28,4613  8,3818 0,000002042 ***
Residuals     295 1001,70  3,3956
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
```

```
parovi <- TukeyHSD(model)$izvor_vijesti
```

Račun počinje modelom skupnih sredina. Poziv `aov` procjenjuje isti model kao `lm`, ali prikazuje razlaganje varijance iz kojega prvo računamo eta-kvadrat. Uobičajeni F-test zatim uspoređuje srednje kvadrate uz pretpostavku zajedničke rezidualne varijance. Pod istom pretpostavkom `TukeyHSD` daje parne razlike, istodobne intervale i korigirane p-vrijednosti. Zato te rezultate čitamo kao klasičnu analizu, uz ranije objašnjenu razliku prema Welchovu postupku.

Zbrojevi kvadrata 113,8 između skupina i 1 001,7 unutar njih daju opisani udio varijabilnosti. F-test provjerava ukupno odstupanje od jednakih sredina. Na prethodnom prikazu parova vidimo da se društvene mreže odvajaju od svakog preostalog izvora, dok ostale parove ne razlučujemo.

U izvještaju možemo navesti da skupine u simuliranom uzorku opisuju oko desetinu varijabilnosti, uz napomenu da nesigurnost toga udjela ovdje nije procijenjena. Razlika televizije i društvenih mreža iznosi oko 1,59 bodova na ljestvici od deset, uz istodobni 95-postotni interval približno od 0,71 do 2,46. Sve četiri usporedbe s društvenim mrežama razlučive su pod pretpostavkama prikazanog postupka. Podaci ipak ne opravdavaju uzročni zaključak.

Sažetak

Kada usporedbu proširimo na više skupina, trebamo pratiti i veću mogućnost barem jednoga lažnog nalaza. Razlaganje

varijabilnosti između skupina i unutar njih omogućuje jedno ukupno pitanje o jednakosti sredina, na koje odgovara F-test. Eta-kvadrat opisuje udio varijabilnosti povezan sa skupinama, uz ograničenje da njegovu uzoračku nesigurnost u ovom primjeru nismo kvantificirali. Za odgovor o konkretnoj razlici ipak trebamo planiranu usporedbu ili imenovani post-hoc postupak s odgovarajućim intervalima. Njihove pretpostavke procjenjujemo prema nacrtu, grafovima i skupnim podacima, a za ovisna opažanja trebamo drukčiji postupak. Dosad je model povezivao povjerenje s oznakom skupine, dok ćemo ga u sljedećem poglavlju proširiti na brojčane prediktore i više varijabli zajedno.

Pojmovi

Tukeyjev postupak (*Tukey HSD*), Kruskal-Wallisov postupak (*Kruskal-Wallis test*), analiza varijance (*analysis of variance*), stopa obiteljske pogreške (*familywise error rate*), F-statistika (*F-statistic*), planirana usporedba (*planned comparison*), post-hoc postupak (*post-hoc procedure*), eta-kvadrat (*eta squared*), varijanca između skupina (*between-group variance*), varijanca unutar skupina (*within-group variance*)

Zadaci

Konceptualni

Objasnite zašto značajan ukupni test ne znači da se svaki par skupina razlikuje. Skicirajte tri skupne sredine i raspršenosti uz koje bi ukupni test bio značajan, a samo jedan par razlučiv. Zatim pretpostavite da ste taj par odabrali tek nakon pregleda svih sredina. Odredite je li usporedba planirana ili post-hoc,

imenujte obitelj koju treba zaštititi i objasnite zašto je ne smijete prikazati kao jedno unaprijed postavljeno pitanje.

Računski

Razlaganje daje 120 na razlike među četirima skupinama i 480 na razlike unutar njih, uz ukupno sto opažanja. Izračunajte obje prosječne raspršenosti, njihov omjer i eta-kvadrat. Objasnite što bi se s omjerom dogodilo da je drugi broj dvostruko veći.

Kritički

Vratite se poglavlju o mjerenju i dizajnu radi razlike između opisne i uzročne tvrdnje te poglavlju o veličini učinka i snazi radi najmanjega važnog učinka. U agregiranoj tablici data/populacija-medija-agregat.csv pronađite retke za TV i društvene mreže. Prije računanja zapišite najmanju razliku na ljestvici od 1 do 10 koju biste smatrali sadržajno važnom. Izračunajte populacijsku razliku iz stupaca `zbroj_povjerenja` i `broj`, usporedite je sa svojim pragom i objasnite zašto rezultat ne dokazuje da odabir izvora uzrokuje povjerenje.

Revizija modela

Ocijenite postupak i obrazloženje iz okvira. Imenujte jedinu pogrešku, odredite koju stopu pogreške ugrožava i razdvojite taj prigovor od dijelova analize koji su još korisni. Predložite način izvještavanja parnih razlika koji odgovara cijeloj obitelji usporedbi. Ne pokrećite kod i ne pišite njegovu zamjenu.

16

DIO V · MODELI

regresija — opći okvir

Od pravca do općeg linearnog modela

Kad razmišljamo o vlastitoj budućnosti, želimo vjerovati da naš trud otvara mogućnosti. Ipak, obitelji se razlikuju po onome što mogu pružiti djeci, a odrastanje u različitim mjestima donosi različite prilike. Koliko se ekonomski položaj roditelja prepoznaje u položaju njihove odrasle djece?

U istraživanju međugeneracijske mobilnosti u Sjedinjenim Državama istraživači su povezali porezne zapise roditelja i djece te usporedili njihove položaje u raspodjeli prihoda u različitim područjima (Chetty i sur. 2014.). Tako se pitanje koje često promatramo kroz vlastitu obitelj moglo postaviti za mnoge obitelji zajedno.

Zamislimo da u jednom gradu djeca iz obitelji s nižim prihodima kao odrasli stoje bolje nego u drugome. Što bismo trebali znati prije nego što obitelji preporučimo preseljenje?

Uspoređivali smo prosječno povjerenje među korisnicima različitih izvora vijesti. Sada nas zanima kako se ono razlikuje s dobi, bez razvrstavanja ljudi u dobne skupine. Možemo postaviti pravac koji za svaku dob daje predviđeni prosjek i zatim provjeriti što takav opis propušta.

Regresija povezuje nekoliko postupaka koje već poznajemo. Korelacija nas je naučila čitati smjer i jačinu linearne veze, a usporedbe dviju i više skupina pokazale su kako oznake skupina ulaze u linearni model. Sada ćemo brojčani i kategorički prediktor promatrati zajedno. Pitanje iz poglavlja o mjerenju i dizajnu pritom ostaje otvoreno. Trebamo procijeniti opisuje li model odnos ili imamo i razloge za uzročno tumačenje.

Usporedbe su pokazale razlike u povjerenju među izvorima, a dobne skupine različit sastav njihove publike. Dob i izvor sada promatramo zajedno. Polazimo od svih 50 000 zamišljenih odraslih osoba u nastavnom skupu `populacija_medija`, koristeći dob, glavni izvor vijesti i povjerenje u medije.

Jedan zapis pripada jednoj osobi, a računalno stvoreni odgovori ne opisuju stvarne stanovnike. Koristimo cijelu konačnu simuliranu populaciju, pa ne procjenjujemo nepoznatu vrijednost iz uzorka. Koeficijenti će sažeti odgovore, a od njih ćemo odvojiti pravila kojima su stvoreni. Pri povratku ranijim uzorcima ponovno ćemo razmotriti i njihovu nesigurnost.

Odabrani pravac usporedit ćemo s opažanjima preko njihovih odstupanja, reziduala. Tako ćemo razumjeti metodu najmanjih kvadrata prije tumačenja koeficijenata. Dodavanjem izvora vijesti usporedit ćemo zbirni odnos dobi i povjerenja s odnosom pri jednakom izvoru, a interakcijom dopustiti da nagib ovisi o drugoj varijabli. Grafovi reziduala pomoći će nam provjeriti prilagodbu.

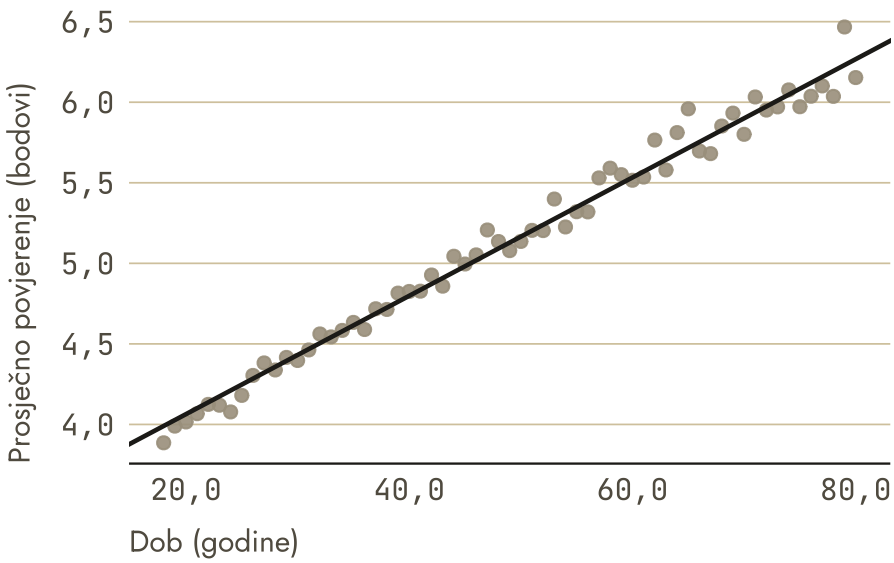
Model ćemo na kraju prosuđivati prema svrsi. Dobar opis opaženih odgovora, uspješno predviđanje nove osobe i objašnjenje uzroka traže različite provjere. Trebamo znati što koeficijent uspoređuje, što je ostalo izvan modela i koji bi dodatni dokaz podupro jači zaključak.

Pravac kao tvrdnja o prosjeku

Povjerenje u medije bilježimo na ljestvici od jedan do deset, a dob u godinama. Korelacija bi sažela jačinu linearne veze. Regresijski pravac omogućuje da za odabranu razliku u godinama očitamo modelom predviđenu razliku u bodovima povjerenja.

Podsjetnik. Nagib

Cilj analize je linearni sažetak zabilježenih odgovora svih pedeset tisuća osoba. Budući da analiziramo cijelu zadanu populaciju, nema nesigurnosti zbog uzorkovanja. Pravilo kojim su podaci generirani odnosi se na vrijednosti prije zaokruživanja i ograničavanja ljestvice. Njegove koeficijente trebamo razlikovati od koeficijenata za zabilježene odgovore. Procjenjivane veličine u ovom su poglavlju koeficijenti najmanjih kvadrata za cijelu konačnu populaciju.



Slika 1: Prosječno povjerenje po godini dobi u simuliranoj populaciji, s pravcem najmanjih kvadrata izračunatim na svih pedeset tisuća pojedinačnih odgovora. Izrada autora. Podaci i postupak opisani su u tekstu.

Na slici jedna točka sažima sve osobe iste cijele godine dobi. Za nju zbrajamo njihove ocjene i dijelimo njihovim brojem. Pravac

ipak računamo iz svih pojedinačnih odgovora, pa dob s više osoba nosi veću težinu. Ne prilagođavamo ga dobnim prosjecima kojima bismo dali jednaku težinu.

Za svaki odabrani odsječak i nagib možemo izračunati odstupanje svakog opažanja od pravca. Tražimo kombinaciju koja daje najmanji zbroj kvadrata tih odstupanja.

> **Rezidual** je razlika između opažene vrijednosti ishoda kod pojedine jedinice i vrijednosti koju za nju predviđa model, dakle onaj dio ishoda koji model nije objasnio.
> Rezidual osobe i označujemo s e_i, opaženi ishod s y_i, a modelom predviđenu vrijednost s $\hat{y}_i$. Kapica označuje vrijednost dobivenu prilagodbom modela. U konačnoj populaciji možemo je izračunati bez uzorkovanja. Rezidual e_i opaženo je odstupanje od prilagođenog pravca, dok ε_i označuje neopaženo odstupanje od pretpostavljene populacijske regresijske sredine.

$$e_i = y_i - \hat{y}_i$$

Sami reziduali mogu se poništiti jer su jedni pozitivni, a drugi negativni. Zato ih kvadriramo prije zbrajanja. Kvadrat daje veću težinu velikim odstupanjima, pa krajnja opažanja mogu snažno utjecati na odabrani pravac.

> **Metoda najmanjih kvadrata** bira koeficijente modela tako da zbroj kvadriranih reziduala bude najmanji mogući za zadane podatke.

Ovdje računamo koeficijente najmanjih kvadrata za cijelu konačnu populaciju. Model ima poznati oblik iz usporedbe skupina. Ishod y_i opisujemo odsječkom β_0, nagibom β_1 pomnoženim s prediktorom x_i i odstupanjem ε_i.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Model s jednim brojčanim prediktorom zovemo **jednostavna linearna regresija**.

Odsječak β_0 daje modelsku vrijednost ishoda kada je prediktor nula. Ovdje β_0 i β_1 pripadaju cijeloj fiksnoj populaciji, a e_i je izračunati rezidual osobe. Kada bismo koristili uzorak, procjene ciljnih koeficijenata označili bismo kapicama i prikazali njihovu nesigurnost. Za razliku od binarne usporedbe skupina, x_i sada predstavlja izmjeren broj.

Pogledajmo tri pojedinačna zapisa prije tumačenja koeficijenata. Stupac predviđanja dobivamo uvrštavanjem dobi u izračunati pravac, a rezidual oduzimanjem toga predviđanja od zabilježenoga odgovora.

Tablica 1: Prva tri opažanja i njihovi reziduali u zbirnom modelu dobi. Izrada autora.

Osoba	Dob	Opaženo	Predviđeno	Rezidual
1	59	4	5,496	−1,496
2	50	4	5,166	−1,166
3	45	6	4,982	1,018

Za prvu osobu račun glasi $3{,}3294 + 0{,}0367 \times 59$, približno 5,496 boda. Kvadrat njezina reziduala, 2,238, jedan je od 50 000 članova zbroja koji smo minimizirali. Za račun koristimo nezaokružene koeficijente.

Nagib iznosi 0,0367 boda po godini. Deset godina veća dob povezana je s modelskom razlikom od 0,37 boda povjerenja. Odsječak 3,33 odnosio bi se na dob nula, izvan naših podataka. Možemo ga učiniti razumljivijim ako od dobi oduzmemo njezinu sredinu. Tada odsječak opisuje osobu prosječne dobi, a nagib ostaje isti.

Pravac sažima prosječan odnos. Točke na slici predstavljaju prosjeke nekoliko stotina ljudi iste dobi i leže blizu pravca. Pojedinačni odgovori ipak odstupaju po nekoliko bodova. Viši predviđeni prosjek za pedesetogodišnjake nego za tridesetogodišnjake ne određuje odgovor konkretne osobe.

Nagib koji nosi tuđu priču

Koeficijent za zabilježeno povjerenje iznosi 0,0367 boda po godini. Znamo ga bez uzoračne nesigurnosti jer imamo cijelu populaciju. Generator je prije zaokruživanja i ograničavanja ishoda koristio latentni, neposredno neopaženi nagib 0,028. Razlika proizlazi iz toga što brojevi opisuju različite veličine.

U generatoru se s dobi mijenja i izvor vijesti. Prosječna dob kreće se od 33,4 godine među korisnicima društvenih mreža do 47,9 među slušateljima radija. Povjerenje ovisi o oba obilježja. Zbirni nagib uz dob zato uključuje i razlike povezane s izborom izvora.

Provjerimo odnos unutar svakog izvora. Generator zadaje isti latentni dobni nagib svim skupinama. Očekujemo da će njihovi zasebni nagibi biti manji od zbirnoga jer unutar skupine izvor ostaje isti.

Tablica 2: Nagib uz dob izračunat zasebno među korisnicima svakog izvora vijesti, uz zbirni konačnopopulacijski nagib i latentno pravilo prije mjerenja. Izrada autora.

Izvor vijesti	Broj ljudi	Nagib uz dob
portal	15 101	0,0269
društvene mreže	13 378	0,0258
TV	10 827	0,0281
radio	5 839	0,0268
tisak	4 855	0,0255
Svi zajedno	50 000	0,0367
Latentno pravilo prije mjerenja	—	0,028

Skupni nagibi kreću se od 0,0255 do 0,0281. Zbirni nagib veći je od svakoga. Blizina skupnih nagiba latentnom pravilu pomaže provjeriti generator, ali to su i dalje koeficijenti zabilježenih odgovora. Kao u poglavlju o povezanosti, odnos unutar skupina razlikuje se od ukupnog odnosa.

Izvor ovdje nije konfundirajuća varijabla u strogom uzročnom smislu. U poznatom generatoru nastaje nakon dobi i prije povjerenja, pa nalikuje posredniku. Zbirni koeficijent opisuje ukupni dobni obrazac, a koeficijent pri jednakom izvoru uvjetni obrazac koji isključuje taj put. Nijedan ne pokazuje kako bi se mijenjalo povjerenje iste osobe tijekom starenja.

Različite odgovore dobivamo i uz svih pedeset tisuća osoba jer smo postavili različita pitanja. Kada bismo uzeli uzorak, procjenjivali bismo iste ciljane koeficijente uz odgovarajuću nesigurnost. Interval ni tada ne bi opisivao udaljenost od latentnog pravila prije mjerenja.

Prilagodba i ono što je čini mogućom

Zasebni nagibi pokazali su nam odnos dobi i povjerenja unutar svakog izvora. Ako dob i izvor uključimo u isti model, možemo zajedno izračunati dobni nagib pri jednakom izvoru i razlike među izvorima pri jednakoj dobi. Time pitanje o zbirnom odnosu pretvaramo u skup uvjetnih usporedbi.

> **Višestruka regresija** procjenjuje modelnu povezanost svakog prediktora s ishodom pri zadanim jednakim vrijednostima svih ostalih prediktora u modelu. Koeficijent je uvjetna usporedba koju zadaje model. Kao usporedba podataka obranjiv je samo ondje gdje postoji dovoljno potpore i preklapanja, pa ne mora označavati stvarno uparene jedinice.

Na primjer, usporedbu četrdesetogodišnjaka koji prate televiziju i portale možemo poduprijeti podacima ako obje skupine sadrže dovoljno ljudi blizu te dobi. Kad jedna skupina sadrži samo mnogo starije osobe, procjena za četrdesetogodišnjake produžuje pravac izvan opaženoga raspona. To zovemo ekstrapolacijom. Zato provjeravamo sadrže li podaci osobe s odgovarajućom kombinacijom dobi i izvora.

Za ovaj model napišimo cijelu jednadžbu. Oznaka $I_{\text{TV},i}$ ima vrijednost jedan za korisnika televizije i nula za ostale, a jednako definiramo indikatore radija, tiska i društvenih mreža. Portal je referenca za koju su sva četiri indikatora nula.

$$y_i = \beta_0 + \beta_1\,\text{dob}_i + \beta_2 I_{\text{TV},i} + \beta_3 I_{\text{radio},i} + \beta_4 I_{\text{tisak},i} + \beta_5 I_{\text{mreže},i} + \varepsilon_i.$$

Koeficijenti daju konkretan račun za svih pedeset tisuća simuliranih osoba. Ovdje su opis konačne populacije, pa im ne pridružujemo interval uzorkovanja.

Tablica 3: Svi koeficijenti modela s dobi i glavnim izvorom vijesti. Portal je referenca. Izrada autora.

Dio	Koeficijent
Odsječak	3,7035
Dob	0,0269
Društvene mreže	−0,5335
TV	0,3116
Radio	0,5951
Tisak	0,4864

Svaka osoba ostaje u istom retku, ali pravac sada dobiva i dodatak za njezin izvor. Za portal je dodatak nula. Za korisnika televizije od 40 godina zbrajamo odsječak 3,704, dobni dio $40 \times 0{,}0269$ i televizijski dodatak 0,312. Rezultat je 5,091 boda. To predviđanje opisuje modelnu sredinu za osobe sa zadanim obilježjima. Za korisnika portala iste dobi mijenja se samo dodatak izvora.

U modelu s dobi i izvorom nagib uz dob iznosi 0,0269. To je uvjetni linearni sažetak ove populacije. Blizak je latentnom pravilu 0,028, ali ne procjenjuje istu veličinu. Slaganje koristimo samo za zasebnu provjeru poznatog generatora.

Tablica 4: Konačnopopulacijski koeficijenti izvora vijesti uz jednaku dob, uz latentna pravila prije mjerenja. Portal je referentna skupina. Izrada autora.

Izvor vijesti	Latentno prije mjerenja	Koeficijent zabilježenog ishoda
društvene mreže	−0,55	−0,53
TV	0,35	0,31
radio	0,60	0,60
tisak	0,50	0,49

Stupci tablice razlikuju latentno pravilo i koeficijente zabilježene ljestvice. Potonji su ovdje blago manji zbog zaokruživanja i ograničavanja vrijednosti na jedan do deset. Zaokruživanje i ograničavanje ljestvice tako ulaze i u tumačenje koeficijenata.

Uvid u generator ovdje nam omogućuje usporedbu koju u stvarnom istraživanju obično nemamo. Bez zabilježenog izvora vijesti nagib uz dob ostao bi 0,0367. Ispis sam ne bi objasnio da taj broj odgovara zbirnom, a ne uvjetnom pitanju.

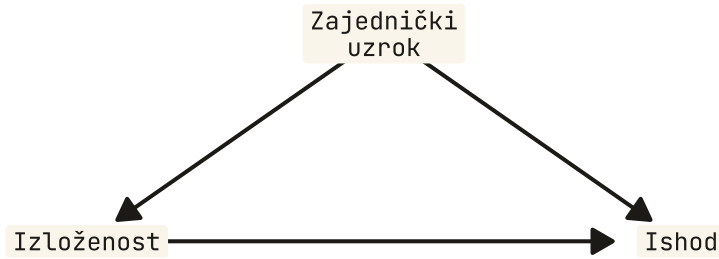
„Kontrola" u regresiji znači usporedbu pri jednakim vrijednostima uključenih varijabli. Ona nije zamjena za randomizaciju. U pokusu randomizacija čini dodjelu neovisnom o početnim obilježjima i uravnotežuje skupine u očekivanju. U konkretnom uzorku mogu ostati slučajne razlike, a provedba i osipanje mogu narušiti usporedbu.

Koeficijent možemo opisati ovako. Ondje gdje podaci podupiru usporedbu, deset godina veća dob uz jednak izvor vijesti povezana je s modelskim prosjekom povjerenja višim za 0,27 boda. Naveli smo jedinicu prediktora, jedinicu ishoda i uvjet usporedbe. Za kategorički prediktor navodimo i referentnu skupinu.

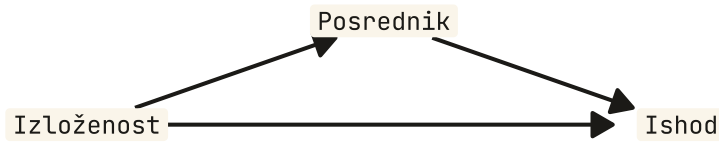
Mogli bismo sada pomisliti da će uključivanje svake dostupne varijable dati bolju usporedbu. No već je izvor vijesti promijenio

pitanje na koje odgovara dobni koeficijent. Prije sljedeće prilagodbe zato trebamo razmotriti ulogu te varijable u nastanku podataka. Dijagram razlikuje tri uloge zbog kojih ista računska radnja može imati različite posljedice.

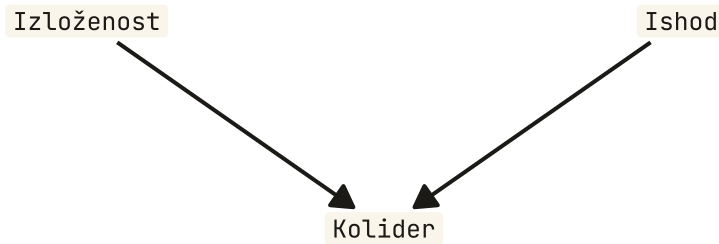
Zajednički uzrok



Posrednik



Kolider



Slika 2: Tri uloge treće varijable u uzročnom dijagramu. Prilagodba za zajednički uzrok može zatvoriti neuzročni put, prilagodba za posrednika mijenja cilj s ukupne na izravnu vezu, a prilagodba za kolider može otvoriti novu pristranost. Izrada autora.

Zajednički uzrok prethodi izloženosti i ishodu te može opravdati prilagodbu. Posrednik nastaje nakon izloženosti, pa njegovim uključivanjem mijenjamo ukupnu usporedbu. Kolider je posljedica obiju varijabli. Uvjetovanje na njega može stvoriti povezanost među inače nepovezanim uzrocima. Prije izbora modela zato razmatramo vremenski redoslijed, kvalitetu mjerenja i ulogu svake varijable.

Odsječak sada daje predviđanje za dob nula i izvor portal, izvan opaženog raspona odraslih osoba. Ovdje služi matematičkom zapisu pravca, dok sadržajno tumačimo predviđanja unutar poduprtog raspona.

Kada dva prediktora nose gotovo istu informaciju, teško je razdvojiti njihove pojedinačne koeficijente. Nedostaju usporedbe u kojima se jedan mijenja, a drugi ostaje isti. Male promjene podataka tada mogu znatno promijeniti koeficijente, iako njihova zajednička predviđanja ostaju slična.

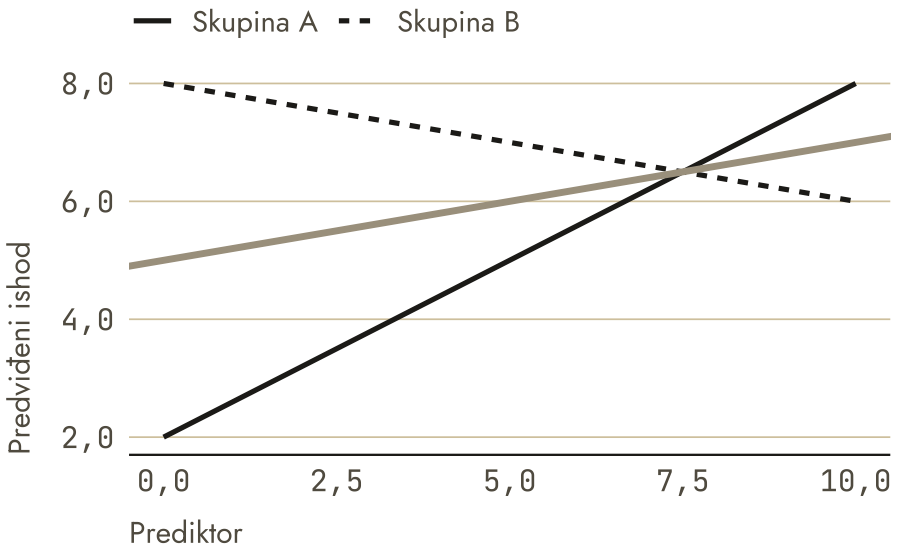
Interakcija i regresijski pravac

U dosadašnjem modelu s dobi i izvorom vijesti nagib uz dob jednak je za sve izvore. Ako nas zanima razlikuje li se nagib među skupinama, trebamo modelu dopustiti i tu mogućnost. To čini interakcija, kojom nagib jednog prediktora može ovisiti o drugome. Za planiranu provjeru unaprijed navodimo modifikator, kodiranje, kontrast i postupak zaključivanja, a kod usmjerene hipoteze i očekivani smjer. Naknadno uočena razlika među podskupinama može dati novo pitanje, ali zahtijeva neovisnu provjeru. Modifikator je varijabla prema kojoj se nagib razlikuje. Kodiranje određuje njezin brojčani zapis, a kontrast konkretnu razliku koju uspoređujemo.

Sljedeće dvije linije namjerno konstruiramo izvan medijske populacije. Za svaku cjelobrojnu vrijednost x od 0 do 10 upisujemo $2 + 0{,}6x$ u A i $8 - 0{,}2x$ u B, bez slučajnog odstupanja. Pri $x = 5$ vrijednosti su 5 i 7. Od $x = 5$ do $x = 6$ prva raste za 0,6, a druga

pada za 0,2. Jednako mnogo točaka na istim vrijednostima x daje zbirnom pravcu prosjek tih nagiba.

U prikazu smo skupini A zadali nagib 0,6, a skupini B −0,2. Zbirni nagib iznosi 0,2 i ne opisuje nijednu skupinu zasebno. Dvije linije prikazuju razliku koju bi zajednički koeficijent sakrio. Ako je $B = 1$ za drugu skupinu, a $B = 0$ za prvu, obje zapisujemo jednadžbom $\hat{y} = 2 + 0{,}6x + 6B - 0{,}8xB$. Član xB mijenja nagib iz 0,6 u −0,2. Taj namjerno izražen primjer služi učenju interakcije. Ne tvrdi da su takve razlike nagiba prisutne u prethodnoj medijskoj populaciji.

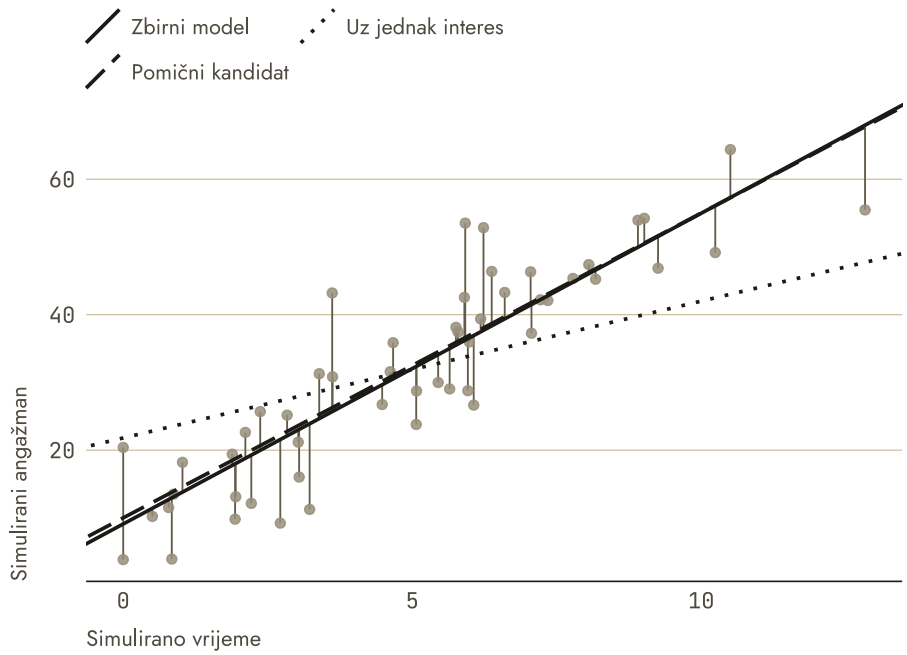


Slika 3: Predviđene vrijednosti u simuliranom primjeru s različitim skupnim nagibima. Deblja zbirna linija opisuje prosjek koji ne vrijedi ni za jednu skupinu. Izrada autora.

Interakciju čitamo kroz skupne nagibe ili predviđene vrijednosti. Značajan nagib u jednoj i neznačajan u drugoj skupini ne dokazuje da se nagibi razlikuju. Potrebna je izravna procjena interakcije s nesigurnošću (Nieuwenhuis i sur. 2011.). Ovaj deterministički primjer pokazuje oblik odnosa. U stvarnoj analizi trebalo bi prikazati i interval za razliku nagiba.

Za pomicanje pravca stvaramo drugi nastavni skup od 52 zapisa, s vremenom, angažmanom i prethodnim interesom. Interes generiramo normalno sa sredinom nula i standardnom devijacijom jedan. Vrijeme nastaje kao $5 + 1{,}8$ puta interes, uz neovisno odstupanje standardne devijacije 1,5 i donju granicu nula. Angažman je $20 + 2{,}3$ puta vrijeme $+7$ puta interes, uz drugo neovisno odstupanje standardne devijacije 4,5. Zato prethodni interes ovdje prethodi objema varijablama i utječe na obje. To je drukčije pravilo od dobnog puta preko izvora vijesti u glavnom primjeru.

Sljedeći prikaz povezuje položaj pravca i duljinu reziduala. Zadani pravac uspoređujemo s pravcem najmanjih kvadrata i s pravcem prilagođenim za prethodni interes. Koristimo manji simulirani skup kako bismo mogli vidjeti pojedinačna odstupanja.



Slika 4: Regresijski pravac — reziduali zbirnog modela, jedan neoptimalni kandidat i blaži prilagođeni nagib pri jednakom prethodnom interesu. Crtkani kandidat gotovo se podudara s punim pravcem najmanjih kvadrata. Izrada autora. Podaci i postupak opisani su u tekstu.

Crtkani kandidat s odsječkom 10 i nagibom 4,5 daje zbroj kvadrata 2 575,9. Puni zbirni pravac postiže minimum 2 561,1. Pravac pri jednakom prethodnom interesu blaži je od zbirnoga.

Blizu minimuma manji pomaci pravca mogu samo malo povećati zbroj kvadrata. Kod uzorka je taj odnos dio računa nesigurnosti koeficijenata. Ovaj prikaz sam ne daje interval ni procjenu uzoračne nesigurnosti.

Prije nastavka vratite se glavnom primjeru s povjerenjem. Pokušajte bez ponovnog čitanja objasniti koju veličinu ondje računamo, što mjeri rezidual, kako biramo pravac i što bi bilo potrebno za uzročni zaključak.

Isti model iza ranijih poglavlja

Model s binarnim prediktorom daje razliku dviju sredina. Model s više kategorija daje razlike prema referentnoj skupini. Sada smo dodali brojčani prediktor. Provjerimo vezu s ranijim računima.

Iste osobe iz uzorka od 120 korisnika televizije i društvenih mreža, iz poglavlja o dvjema grupama, daju razliku sredina 1,19 boda. Koeficijent uz izvor u modelu iznosi 1,19. To su dva zapisa iste točkaste procjene.

Omjer koeficijenta i obične homoskedastične standardne pogreške iznosi 3,21. Jednak je združenom Studentovu t-testu s postavkom `var.equal = TRUE`. Oba koriste zajedničku rezidualnu varijancu i 118 stupnjeva slobode. Welchov postupak dopušta različite varijance te daje t od 3,18 i 102,471 stupnja slobode. Procijenjena razlika ostaje jednaka, ali račun nesigurnosti se mijenja.

Na istom uzorku od 300 osoba s pet izvora vijesti iz poglavlja o više skupina klasični model daje F od 8,38, uz 4 i 295 stupnjeva slobode. To je isti homoskedastični postupak kao `aov` iz poglavlja Više grupa. Welch daje F od 7,32, uz 4 i 112,4 stupnja slobode.

Zajednički model sredina ne znači jednak test. Objašnjeni udio varijabilnosti u klasičnom modelu iznosi 10,2 %, koliko i prethodno izračunati eta-kvadrat.

T-test, analiza varijance i regresija zato mogu dijeliti model sredina. U svakom izvještaju pitamo koja je jedinica, što se uspoređuje i pod kojim uvjetima. Zasebno provjeravamo pretpostavke i način računanja nesigurnosti.

Tablica za odabir postupka u mrežnom dodatku pomaže povezati vrstu podataka i istraživačko pitanje. Pri tom odabiru i dalje provjeravamo ishod, prediktore, neovisnost i nesigurnost. Welchovu ili običnu homoskedastičnu inferenciju biramo prema pretpostavkama o varijancama i nacrtu.

Ako jedna osoba daje više redaka ili osobe dijele razred, ustanovu ili grad, ponovno provjeravamo jedinicu i ciljanu veličinu. Više redaka može nekim osobama dati veću težinu, a uobičajena nesigurnost više nema opravdanje neovisnosti. Trebamo postupak koji uvažava tu strukturu. Takve modele ovdje ne procjenjujemo.

Pristajanje i njegove granice

Vezu regresije s analizom varijance možemo iskoristiti i za procjenu pristajanja. Ondje nas je zanimalo koliki dio varijabilnosti opisuju skupne sredine. Sada isto pitanje postavljamo za vrijednosti koje predviđa cijeli model, neovisno o tome sadrži li oznake skupina ili brojčane prediktore.

> **Koeficijent determinacije** je udio ukupne varijabilnosti ishoda u promatranom skupu podataka koji je model objasnio, izračunat kao jedan minus omjer zbroja kvadrata reziduala i zbroja kvadrata odstupanja od zajedničke sredine.
> Mjeru označujemo s R^2. U formuli je e_i rezidual, y_i opaženi ishod, a $\bar{y}$ njegova zajednička sredina.

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

Za model s dobi zbroj kvadrata 50 000 reziduala iznosi 179 511,3. Bez prediktora svakoj bismo osobi dodijelili zajednički prosjek. Zbroj kvadrata odstupanja od njega iznosi 196 832,6. Dijeljenjem prvoga zbroja drugim dobivamo preostali dio varijabilnosti, pa ga oduzimamo od jedan. Usporedba koristi iste osobe i isti ishod.

Model s dobi opisuje 8,8 % varijabilnosti povjerenja, a s dobi i izvorom 12,2 %. Nizak udio može odražavati razlike među ljudima, mjernu pogrešku ili neprikladan oblik modela. Za razlikovanje tih mogućnosti trebamo provjeriti mjerenje i dijagnostičke prikaze modela. Blizina prilagođenog koeficijenta latentnom pravilu također ne mijenja činjenicu da su to različite veličine.

Za sljedeću provjeru uzimamo novi slučajni uzorak od 200 osoba bez vraćanja iz te populacije. Svakoj dodajemo pet neovisnih normalnih brojeva sa sredinom nula i standardnom devijacijom jedan. Ti stupci po pravilu nastanka nemaju vezu s povjerenjem, iako je u konačnom uzorku mogu slučajno pokazati. Uspoređujemo dva modela na istih 200 potpunih redaka.

U toj usporedbi osnovni model sadrži dob i glavni izvor vijesti. Prošireni model istim prediktorima dodaje pet neovisno generiranih normalnih slučajnih varijabli, označenih `sum1` do `sum5`. Oba se prilagođavaju istim dvjema stotinama osoba. Dodatne varijable nemaju sadržajno opravdanje.

Na istim redcima i istom ishodu dodavanje prediktora ne smanjuje obični objašnjeni udio. U uzorku od dvjesto osoba pet slučajnih stupaca povećava ga s 8,1 % na 9,1 %. Prilagođeni koeficijent determinacije uzima u obzir broj prediktora i pada s 5,8 % na 4,3 %. Ta prilagodba pomaže pri usporedbi, ali sama ne provjerava predviđanje, uzročnost ni sadržajnu valjanost.

Za provjeru ograničenja raspona vraćamo se cijeloj simuliranoj populaciji, zasebno od prethodnog uzorka od 200 osoba. Među

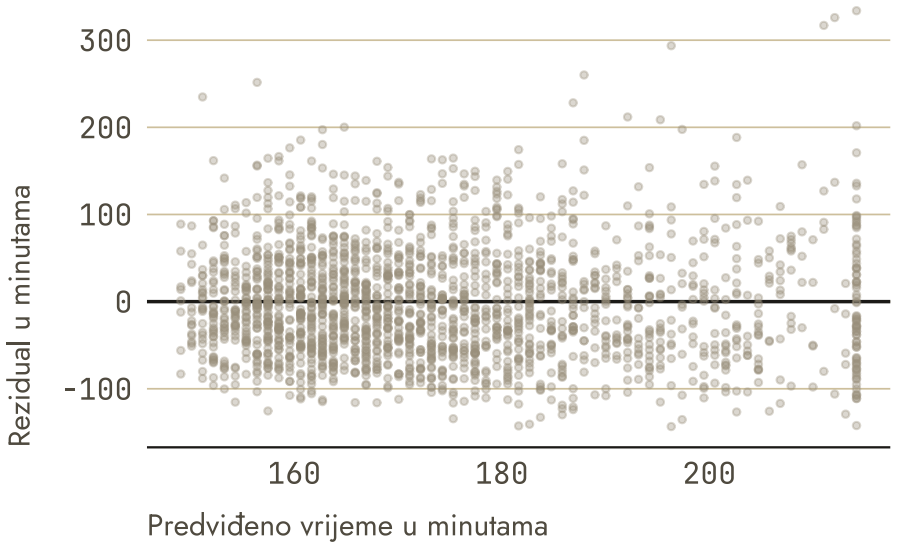
osobama od trideset do pedeset godina nagib iznosi 0,0385, gotovo kao prije. Objašnjeni udio ipak pada s 8,8 % na 1,4 % u skupu od 24 282 ljudi. Koeficijent determinacije zato ovisi o populaciji i rasponu podataka.

Slični nagibi u različitim populacijama mogu imati vrlo različite objašnjene udjele. Za usporedbu koeficijenata trebamo provjeriti ishod, ljestvicu, populaciju, kodiranje i uključene prilagodbe. Kod uzoraka trebamo i intervale koji odgovaraju nacrtu.

Iza zajedničkog naziva mogu stajati različiti konstrukti i jedinice. U našoj simulaciji „medijska aktivnost" označuje dnevne minute osobe iz konačne populacije odraslih. DigiKat obuhvaća platformske objave prema pravilima korpusa, s jedinicom objave ili agregata (Šikić i sur. 2026.). Osobne minute ne prikazuju svako privatno ponašanje, a korpus ne obuhvaća sadržaj izvan svojeg izbora. Regresija objava zato ne opisuje osobnu uporabu, a skupni obrazac ne možemo pripisati pojedincu. Trebamo usporediti konstrukt, jedinicu, populaciju i obuhvat mjerenja.

Za sljedeći dijagnostički primjer ponovno koristimo svih 50 000 osoba, ali mijenjamo ishod iz povjerenja u dnevne minute. Te su minute generirane tako da s dobi rastu i očekivana vrijednost i raspršenost.

Mjesta na kojima model slabije odgovara podacima tražimo pregledom pojedinačnih odstupanja, reziduala. Tako dopunjujemo ukupni opis objašnjenog udjela. U našoj populaciji linearni odnos dnevnog praćenja medija i dobi ima nagib 1,05 minute po godini. Reziduali oko tog pravca nisu jednako raspršeni duž cijelog raspona, što možemo vidjeti na njihovu prikazu. Model računamo na cijeloj populaciji, a slika prikazuje slučajni podskup od 2 000 osoba.



Slika 5: Reziduali modela za dnevno vrijeme uz medije nasuprot vrijednostima koje model predviđa, na slučajnom podskupu od dvije tisuće osoba. Izrada autora. Podaci i postupak opisani su u tekstu.

Od šestine s najnižim do šestine s najvišim predviđanjima raspršenost reziduala raste s 58 na 77 minuta. Pogrešku zato trebamo opisati duž raspona predviđanja, uz ukupnu mjeru. Kada bismo imali uzorak, račun nesigurnosti trebao bi dopustiti promjenjivu varijancu. Graf otkriva problem, a sljedeći korak biramo prema cilju analize.

Vrijeme ne može pasti ispod nule. Zbog te granice donji rub ostataka slijedi kosu crtu, a osobe s nulom gomilaju se na njoj. Zakrivljenost oblaka i širenje ostatka prema većim predviđanjima upozoravaju da jedan pravac sa stalnom varijancom ne opisuje sve dijelove raspodjele jednako dobro. Promjena računa standardne pogreške može ublažiti problem nesigurnosti, ali ne ispravlja pogrešan oblik sredine ni predviđanja izvan mogućeg raspona.

Takav pregled modela potreban je uz svaki odabrani način izražavanja nesigurnosti. Bayesovski pristup, koji smo kratko upoznali uz testiranje, može istom modelu pridružiti prethodne raspodjele i dati raspodjelu nesigurnosti koeficijenata ili predviđanja.

I tada trebamo dobro mjerenje, prikladne jedinice i opravdane pretpostavke. Takav račun sam ne rješava curenje informacija ni uzročno zaključivanje, a ovdje ostajemo na tom kratkom osvrtu.

Objašnjenje i predviđanje

Za opis odnosa tražimo sadržajno smislen i razumljiv model. Za predviđanje tražimo malu pogrešku na novim jedinicama. Uzročno objašnjenje zahtijeva dodatno opravdanje iz nacrta i identifikacijskih pretpostavki, uvjeta pod kojima opažene odnose možemo pripisati pretpostavljenom uzroku. Shmueli pokazuje da se ti ciljevi razlikuju već pri izboru mjera i varijabli (Shmueli 2010.). Model dobar za jedan cilj može biti loš za drugi.

U sljedećim primjerima povjerenje predviđamo neposredno prije nego što nova osoba odgovori na pitanje o povjerenju. Tada znamo dob i izvor vijesti. Slučajni stupci kojima ćemo pokazati preprilagodbu također su dostupni prije odgovora i jednako definirani u oba skupa. Informacije koje nastaju poslije ishoda ne smijemo koristiti.

Spremnost na plaćanje vijesti u generatoru nastaje nakon povjerenja i pod njegovim utjecajem. Ako je dodamo modelu povjerenja, objašnjeni udio raste s 12,2 % na 13,2 %. Za naš trenutak primjene to je curenje informacija. Prediktor tada još nije dostupan, pa poboljšanje na postojećim podacima nije valjan prediktivni dobitak.

Sada izvlačimo zasebnih 150 osoba za učenje, a iz preostalih osoba 2 000 za provjeru. Skupovi se ne preklapaju. Svakom skupu dodajemo 25 neovisnih normalnih stupaca sa sredinom nula i standardnom devijacijom jedan. To je novi pokus, različit od prethodnih pet stupaca na 200 osoba. Složeniji model koristi svih 25 dodataka, a jednostavniji samo dob i izvor. Oba procjenjujemo na 150 osoba. Dobivene koeficijente zatim zamrzavamo prije predviđanja za 2 000 odvojenih osoba.

Predviđanje provjeravamo na podacima odvojenima prije procjene modela. Pogrešku mjerimo srodno ranijem računu reziduala. Odstupanja opaženih odgovora od predviđanja kvadriramo, izračunamo njihovu sredinu i izvadimo korijen kako bismo se vratili u bodove povjerenja. Tako možemo usporediti pogrešku na podacima za učenje s pogreškom na odvojenim osobama. Ovdje je to interna provjera istog simuliranog mehanizma koja ne dokazuje prijenos u drugo vrijeme, sustav ili populaciju. Oba uspoređena modela koriste samo informacije dostupne do zadanog trenutka.

Prva tri zapisa iz skupa za provjeru pokazuju kako nastaje njegova pogreška. Svako predviđanje dobiveno je bez uporabe odgovora te osobe pri procjeni koeficijenata. Takvo predviđanje za osobe koje nisu sudjelovale u prilagodbi modela zovemo predviđanjem izvan uzorka.

Tablica 5: Odvojene osobe, predviđanja jednostavnijeg modela i kvadrirane pogreške. Izrada autora.

Osoba	Opaženo	Predviđeno	Kvadrat pogreške
30952	2	5,188	10,162
31037	9	5,662	11,143
36739	3	3,999	0,997

Prvi kvadrat ulazi u zbroj s još 1 999 takvih kvadrata. Zbroj dijelimo s 2 000 i vadimo korijen. Za učenje je nazivnik 150. Svi su odgovori i predviđanja definirani, pa nema izostavljanja. Tablica koja slijedi sažima iste račune za oba modela i za predviđanje prosjekom skupa za učenje.

Tablica 6: Prosječna pogreška predviđanja dvaju modela na podacima na kojima su procijenjeni i na odvojenom skupu, u bodovima povjerenja. Izrada autora.

Model	Na skupu za učenje	Na odvojenom skupu
Dob i izvor vijesti	1,78	1,88
Dob, izvor i 25 slučajnih brojeva	1,65	2,06
Sama sredina učenja	—	1,99

Na skupu za učenje od 150 osoba složeniji model opisuje 25,4 % varijabilnosti, a jednostavniji 13,1 %. Na tih istih podacima složeniji griješi manje. Na odvojenih 2 000 osoba griješi više od jednostavnijeg modela, pa i od predviđanja zajedničkim prosjekom skupa za učenje. Prilagodio se slučajnim obrascima koji se ne ponavljaju u novim podacima.

Jednostavniji model bolji je na ovom odvojenom skupu. Standardna devijacija povjerenja u populaciji iznosi 1,98 boda, a rezidualna raspršenost nakon uključivanja dobi i izvora 1,86. Ostatak uključuje slučajnu sastavnicu generatora, ali možda i neprimjerenost oblika modela. Za procjenu koliko bi se ta pogreška još mogla smanjiti trebalo bi ispitati i druge opravdane modele.

Valjana provjera predviđanja zato zahtijeva odvojene podatke i prediktore dostupne u trenutku primjene.

Granica prema uzroku

Konfundirajuća varijabla prethodi pretpostavljenom uzroku i ishodu. Regresijom možemo prilagoditi usporedbu za tu varijablu. Njezinu ulogu pritom opravdavamo sadržajnim znanjem, vremenskim redoslijedom i nacrtom istraživanja.

U našem generatoru izvor vijesti nastaje poslije dobi i prije povjerenja. Prilagodba za izvor zato zatvara mogući posrednički put i mijenja cilj sa zbirnog na uvjetni odnos. Blizinu prilagođenog koeficijenta latentnom izravnom pravilu razumijemo jer poznajemo generator. U stvarnom istraživanju takvo slaganje ne bismo mogli pretpostaviti.

Prilagođena povezanost sama ne dokazuje uzrok. Trebamo opravdati vremenski red, izostanak neizmjerenih zajedničkih uzroka i izbor prilagodbi. Važni su i valjano mjerenje, dovoljno preklapanja usporedivih jedinica, jasno definirana intervencija, selekcija i oblik modela. Za svaki uvjet trebamo pokazati na kojim dokazima počiva opravdanje.

Dob je posebno zahtjevna jer nema jednostavno intervencijsko tumačenje. U stvarnom istraživanju ulogu dobi i izvora trebali bismo razmotriti prije računanja. Modelni ispis ne može potvrditi te pretpostavke.

Zato u izvještaju opisujemo usporedbu koju smo doista napravili. Tvrdnja o tome što bi se dogodilo kada bismo nešto promijenili traži nacrt i pretpostavke koji opravdavaju takvu promjenu.

Prije riječi „učinak” korisno je zapisati koja bi druga objašnjenja proizvela isti obrazac i kako ih nacrt isključuje. Ako se ključna pretpostavka ne može provjeriti ovim podacima, trebamo je jasno navesti uz zaključak.

Sada možemo preciznije pročitati priču o ekonomskim prilikama s početka poglavlja. Istraživanje međugeneracijske mobilnosti pokazalo je da se odnos između položaja roditelja i njihove odrasle djece razlikuje među područjima (Chetty i sur. 2014.). Nagib pravca koji povezuje njihove položaje u raspodjeli prihoda opisuje koliko je položaj djece vezan uz položaj roditelja. Za obitelj s nižim prihodima važno je i na kojoj se visini nalazi taj pravac pri njezinu položaju. Dva područja možemo uspoređivati prema jačini veze, ali i prema predviđenom položaju djece iz obitelji s jednakim početnim položajem.

Autori su veću mobilnost povezali, među ostalim, s manjom segregacijom i boljim osnovnim školama, uz izričito ograniče-

nje da njihova opisna analiza ne utvrđuje uzročne mehanizme (Chetty i sur. 2014.). Obitelji koje žive u tim područjima mogu se razlikovati i po onome što nismo izmjerili. Za preporuku o preseljenju trebalo bi istražiti što se događa usporedivim obiteljima nakon promjene sredine i obrazložiti zašto razliku pripisujemo toj promjeni. Karta mobilnosti može pokazati gdje tražiti objašnjenje, a regresijski pravac sažeti obrazac. Nijedno od toga samo po sebi ne određuje budućnost pojedinog djeteta.

● STATISTIKA U DIVLJINI

Od omjera do dopuštene rečenice. Školski stručni tim mogao bi iz istraživanja o društvenim mrežama pokušati zaključiti bi li ograničenje vremena uporabe smanjilo simptome depresije. U toj zamišljenoj primjeni nalaz treba pročitati prije nego što ga pretvorimo u preporuku. Analiza se oslanja na presječne podatke istraživanja Ungdata iz 2018., u kojem je 12 353 norveških adolescenata od 15 do 16 godina dalo podatke o vlastitim navikama i simptomima, uz ukupni odaziv od 85 % (Kleppang i sur. 2021.). Ti odgovori omogućuju proučavanje povezanosti uporabe medija i simptoma. Učinak provedene zabrane zahtijevao bi drukčije istraživanje.

Ishod u modelu razlikuje rezultat na mjeri simptoma depresije na ili iznad 80. percentila od nižih rezultata (Kleppang i sur. 2021.). To nije isto što i klinička dijagnoza. Sljedeći izvadak prikazuje kako se omjeri mijenjaju kada se društvene mreže i igranje uključe zasebno ili zajedno.

Tablica 7: Prilagođeni omjeri izgleda za simptome depresije prema uporabi društvenih mreža i igranju. Sadržajno skraćena, preoblikovana i na hrvatski prevedena prilagodba Tablice 3 iz istraživanja povezanosti uporabe medija i simptoma depresije (Kleppang i sur. 2021.).

Prediktor i kategorija	Model 1, društvene mreže AOR (95 % interval)	Model 2, igranje AOR (95 % interval)	Model 3, oba prediktora AOR (95 % interval)
Društvene mreže, do 3 sata	1 (ref.)	nije u modelu	1 (ref.)
Društvene mreže, više od 3 sata	1,60 (1,43–1,80)	nije u modelu	1,51 (1,34–1,70)
Igranje, do 3 sata	nije u modelu	1 (ref.)	1 (ref.)
Igranje, više od 3 sata	nije u modelu	1,57 (1,36–1,80)	1,38 (1,19–1,59)

Binarni ishod u ovome radu analizira se logističkom regresijom. Taj model povezuje prediktore s logaritmom izgleda, a ne s neograničenom sredinom binarnoga ishoda. Izgledi su omjer vjerojatnosti ishoda i vjerojatnosti njegova izostanka. Eksponiranjem odgovarajućega koeficijenta dobiva se omjer izgleda uz ostale prediktore u modelu. Ovdje učimo čitati objavljeni rezultat. Prilagodba i provjera logističkog modela izvan su opsega knjige (Kleppang i sur. 2021.).

AOR označuje prilagođeni omjer izgleda. Sva tri modela uključuju rod ili spol, imanje prijatelja, pušenje, visoko obrazovanje roditelja i obiteljsko materijalno stanje (Kleppang i sur. 2021.). Model 3 istodobno uključuje obje aktivnosti, pa je omjer za mreže prilagođen i za igranje, a omjer za igranje i za uporabu mreža. Referentne su skupine označene jedinicom. Prije čitanja nalaza treba razlikovati izglede od vjerojatnosti. U zamišljenoj skupini u kojoj ishod ima 20 od 100 osoba, vjerojatnost je 0,20, a izgledi su 20 prema 80, odnosno 0,25. Za vjerojatnost p isti račun zapisujemo kao $p/(1-p)$. Omjer izgleda uspoređuje te izglede, a omjer rizika uspoređuje vjerojatnosti. Nijedan nije razlika u postotnim bodovima.

U Modelu 1 procijenjeni izgledi za ishod kod uporabe mreža dulje od tri sata iznose 1,60 puta izglede referentne skupine, uz interval od 1,43 do 1,80 (Kleppang i sur. 2021.). Vrijednosti iznad jedan znače veće, a ispod jedan manje izglede. Interval koji obuhvaća jedan dopušta i jednake izglede. Tim stoga ne može nalaz prepričati kao povećanje vjerojatnosti za 60 postotnih bodova ili kao dokaz da će ograničenje uporabe ukloniti odgovarajući dio simptoma.

Za predviđenu vjerojatnost trebaju odsječak, ostali koeficijenti i vrijednosti prediktora. Izvadak ne daje nazivnike modela, mjere pristajanja ni dijagnostiku, pa ne omogućuje provjeru gubitka redaka i pristajanja. Koeficijenti prilagodbenih varijabli nisu automatski njihovi uzročni učinci (Westreich i Greenland 2013.).

Školskom timu nalaz pokazuje povezanost vrijednu daljnjeg proučavanja. Presječni nacrt, samoprijava i mogući neizmjereni čimbenici ne dopuštaju uzročni zaključak (Kleppang i sur. 2021.). Tablica tako čuva opaženi nalaz, a pokazuje koji dokaz još nedostaje za odluku o ograničavanju uporabe.

Izvor su Annette Løvheim Kleppang, Anne Mari Steigen, Li Ma, Hanne Søberg Finbråten i Curt Hagquist, „Electronic media use and symptoms of depression among adolescents in Norway”, PLOS ONE 16(7), e0254197, 2021, članak i izvorna Tablica 3. Licenca je CC BY 4.0. Tablica je skraćena, preoblikovana i prevedena na hrvatski. Brojevi i granice intervala sačuvani su. Prilagodba ne podrazumijeva odobrenje autora ni PLOS-a.

NAĐITE GREŠKU

Tim želi provjeriti predviđa li model povjerenje kod novih korisnika uz korijen srednje kvadratne pogreške manji od dva boda. To je unaprijed zadan kriterij ove nastavne vježbe. Za prilagodbu je uzeto prvih 200 zapisa simulirane populacije. U zamišljenoj primjeni svi prediktori dostupni su prije odgovora o povjerenju.

Asistentov račun rabi naziv `uzorak` za tih istih 200 zapisa.

```
model <- lm(povjerenje_medijima ~ dob + izvor_vijesti + obrazovanje +
              minute_medija + spol, data = uzorak)
pogreska <- sqrt(mean((uzorak$povjerenje_medijima - fitted(model))^2))
c(r2 = summary(model)$r.squared, pogreska = pogreska)
```

Ispis daje $R^2 = 0{,}12$ i korijen srednje kvadratne pogreške 1,78 boda na ljestvici povjerenja od jedan do deset. U zapisu nema drugog skupa podataka. Asistent zaključuje sljedeće.

> Izračunane vrijednosti opisuju pristajanje modela navedenim zapisima. Pogreška je manja od dva boda, pa je zadani kriterij za nove korisnike zadovoljen i model možemo preporučiti za primjenu.

Pratite iz kojih redaka dolaze koeficijenti, a iz kojih pogreške. Napišite što ovaj ispis već podupire i koju provjeru treba provesti prije odluke o primjeni. Vaša preporuka treba odrediti kako bi se novi rezultat usporedio s unaprijed zadanim kriterijem.

Razrađeni primjer

Vratimo se našim simuliranim podacima. Koliko se povjerenje razlikuje s dobi i što se u tom odgovoru promijeni kada uspoređujemo osobe s jednakim izvorom vijesti? Izračunat ćemo oba

nagiba na istim zabilježenim odgovorima svih pedeset tisuća osoba. Tako razliku među koeficijentima možemo povezati s promjenom pitanja, bez promjene analizirane populacije.

```
zbirni <- lm(povjerenje_medijima ~ dob, data = populacija_medija)
prilagodeni <- lm(
  povjerenje_medijima ~ dob + izvor_vijesti,
  data = populacija_medija
)
c(zbirni = coef(zbirni)[["dob"]],
  prilagodeni = coef(prilagodeni)[["dob"]])
```

```
     zbirni prilagodeni
 0,03672382  0,02688517
```

Funkcija `lm` ovdje računa opis cijele konačne populacije. Zato ne koristimo `confint` za uzoračne intervale. Uobičajeni interval te funkcije odnosio bi se na drugi modelni ili nadpopulacijski cilj. To znači zaključivanje o zamišljenom procesu koji bi stvarao i druge populacije, a taj cilj ovdje nismo odabrali.

Tablica 8: Konačnopopulacijski nagib uz dob u dvama modelima za zabilježeni ishod. Izrada autora.

Model	Konačnopopulacijski nagib	Uzoračna nesigurnost
Samo dob	0,0367	nema
Dob i izvor vijesti	0,0269	nema

Oba nagiba znamo točno za zadanu populaciju. Zbirni uključuje dobni obrazac povezan s izborom izvora, dok prilagođeni uspoređuje pri jednakom izvoru. Različite brojke proizlaze iz različitih pitanja. Izvještaj mora pokazati koje je pitanje postavljeno.

U zbirnom modelu deset godina veća dob povezana je s predviđenim prosjekom povjerenja višim za 0,37 boda. Uz jednak izvor razlika iznosi 0,27 boda. Prva rečenica opisuje ukupni, a druga uvjetni linearni odnos.

Ovim smo koeficijentima opisali cijelu zadanu populaciju, pa im ne pridružujemo uzoračni interval. Kada bismo istu usporedbu procjenjivali iz uzorka, trebali bismo odgovarajuću nesigurnost. Za predviđanje odgovora nove osobe trebali bismo i provjeru na odvojenim podacima, uz prediktore dostupne u trenutku primjene. Uzročna tvrdnja tražila bi dodatno opravdanje izvan ovog računa.

Granica našeg opisa ovisi i o onome što smo uključili u model. Uz dva odabrana prediktora u stvarnom bismo istraživanju razmotrili, primjerice, obrazovanje, iskustvo s redakcijama ili političke sklonosti. Njihovo uključivanje moglo bi promijeniti ciljanu uvjetnu usporedbu i koeficijent.

Prikazani nagibi opisuju razlike među osobama u populaciji u jednom trenutku. Za proučavanje promjene povjerenja iste osobe tijekom starenja ili promjene izvora vijesti trebali bismo je pratiti kroz vrijeme, uz nacrt primjeren tom pitanju.

Sažetak

Regresiju možemo čitati kao nastavak ranijih usporedbi, u kojem najmanji kvadrati biraju koeficijente prema odstupanjima opažanja od predviđenih vrijednosti. Dodavanjem prediktora prelazimo na uvjetne usporedbe, a interakcijom dopuštamo da se nagibi razlikuju. Naš je primjer pokazao zašto koeficijente zabilježenih odgovora, nakon zaokruživanja i ograničavanja ljestvice, trebamo razlikovati od pravila kojima smo podatke stvorili. Reziduali pomažu provjeriti pristajanje, a veći objašnjeni udio ne jamči bolje predviđanje novih osoba jer ga mogu povećati i slučajni prediktori. Zato predviđanje provjeravamo na odvojenim jedinicama uz informacije dostupne u trenutku primjene, dok uzročno tumačenje prilagođene povezanosti traži dodatno opravdanje. S tom razlikom između opisa, predviđanja i uzroka možemo u sljedećem poglavlju prosuditi kako se modeli koriste u odlukama o ljudima.

Pojmovi

linearna regresija (*linear regression*), procjenjivana veličina (*estimand*), rezidual (*residual*), metoda najmanjih kvadrata (*least squares*), višestruka regresija (*multiple regression*), prilagođena povezanost (*adjusted association*), koeficijent determinacije (R-*squared*), interakcija (*interaction*), curenje informacija (*information leakage*), predviđanje izvan uzorka (*out-of-sample prediction*), izgledi (*odds*), omjer izgleda (*odds ratio*)

Zadaci

Konceptualni

Objasnite u jednom odlomku zašto se nagib uz dob mijenja kada u model uđe izvor vijesti, a nijedan podatak o dobi pritom nije promijenjen. Imenujte pitanje na koje odgovara svaki od dvaju nagiba. Na prikazu interakcije zatim odredite koji nagib opisuje skupinu A, koji skupinu B i zašto zbirni nagib ne opisuje nijednu.

Računski

Zbirni nagib iznosi 0,0367 boda po godini. Izračunajte modelom predviđenu razliku prosjeka između dobi od 25 i 55 godina, a zatim isti račun ponovite s prilagođenim nagibom 0,0269 uz jednak izvor vijesti. Predajte obje brojke i objasnite zašto nijedna ne određuje ostvarenu razliku dviju konkretnih osoba.

Kritički

Vratite se poglavlju o mjerenju i dizajnu radi operacionalizacije i vremenskog reda te poglavlju o tome kako brojke zavode radi razlike između točne brojke i preširoke tvrdnje. U prilagođenoj tablici iz okvira odredite što je mjereno, koja je referentna skupina, koju količinu podupire ćelija 1,60 i koja bi dodatna količina trebala za tvrdnju o apsolutnoj vjerojatnosti. Zatim preuredite naslov „Više od tri sata na društvenim mrežama povećava depresiju za 60 %" u jednu rečenicu koju presječni nacrt i prikazani omjer izgleda doista podupiru.

Revizija modela

Ocijenite račun i zaključak iz okvira o pogrešci. Imenujte što je u kodu ispravno i povežite obje prijavljene veličine sa skupom na kojem su dobivene. Napišite zaključak koji bi isti ispis podnio te opišite provjeru potrebnu za nove korisnike. Objasnite zašto unaprijed dostupni prediktori sami ne potvrđuju uspješnost predviđanja na novim podacima.

17

DIO V · MODELI

statistika u doba algoritama

Kako predviđanja postaju društvene odluke

„Da li je pošteno da se ukida prethodna stopa i da se povećava u odnosu na prethodnu?" Ta je rečenica iz hrvatske parlamentarne rasprave zabilježena u skupu ParlaSent (Mochtak i sur. 2023.). Zastanite prije nastavka i procijenite njezin ton. Čujete li neutralno pitanje, kritiku ili nešto između? Zapišite što je u rečenici presudilo za vaš izbor.

Zamislimo da urednički sustav šalje negativno označene rečenice na ljudski pregled prije mogućeg javnog sažetka. Oznaka se ne objavljuje automatski, ali određuje koji tekst dobiva dodatnu pozornost. Pogrešno uključivanje može neopravdano istaknuti rečenicu, a pogrešno izostavljanje sakriti važan dio govora.

Treba li početnu rečenicu poslati na pregled? Ako se dvoje ljudi oko toga ne složi, čiju ćemo prosudbu pretvoriti u odgovor iz kojega model uči?

Kod regresije smo razlikovali opis postojećih podataka od predviđanja novih ishoda. Sada nas zanima što se događa kada predviđanje pokrene postupak prema nekoj osobi ili tekstu. U zamišljenom uredničkom sustavu iz vinjete procijenjeni ton odlučuje koja će rečenica dobiti dodatni pregled. Za procjenu takva sustava trebamo pratiti i nastanak oznake i posljedice pogrešne odluke.

U hrvatskom dijelu skupa podataka ParlaSent 1.0 nalaze se stvarne parlamentarne rečenice čiji su ton procijenili ljudi. Skup je objavljen u repozitoriju Instituta Jožef Stefan (Mochtak i sur. 2023.). Rečenica je jedinica na koju se odnose zabilježene oznake i naši zaključci. Pratit ćemo tekst, dostupne oznake kodera i podjelu podataka. U nastavnom primjeru iz tih ćemo oznaka izvesti jednostavan rezultat za usporedbu pravila slanja na pregled. Time provjeravamo kako izbor praga mijenja odluku. Uspješnost budućega uvježbanog klasifikatora zahtijevala bi zasebnu provjeru njegovih predviđanja. Stvarne oznake omogućuju nam da pregledamo neslaganje ljudskih kodera i pogrešivost reference prema kojoj prosuđujemo odluku.

U tom se zadatku sastaju ranija pitanja o mjerenju, uzorkovanju i pogreškama. Ljudske oznake tona trebamo provjeriti prema pravilima kodiranja, a doseg korpusa prema načinu odabira rečenica. Razmotrit ćemo zašto podatke razdvajamo za učenje, provjeru i ispitivanje te kako prilagođavanje poznatim primjerima može prikriti slabo predviđanje novih. Za svaku procjenu trebamo znati na kojim je podacima nastala i što je već moglo utjecati na odluke o modelu.

Promjenom praga usporedit ćemo pogrešna uključivanja i propuštanja. U zasebnoj simulaciji s dvjema skupinama istražit ćemo zašto se mjerila pravednosti mogu razilaziti kada se razlikuju

učestalosti ishoda. ParlaSentu pritom nećemo pripisivati skupinska obilježja koja u paketu ne postoje. Provjeru ćemo proširiti na mogućnost osporavanja oznake, odgovornost za postupak i uporabu jezičnih modela. Umjesto samog postotka točnosti moći ćemo tražiti opis toga što sustav predviđa, prema kojoj oznaci mjeri uspjeh i tko snosi posljedice. Te odluke počinju izborom korpusa i jedinice koju sustav razvrstava.

Odluka, korpus i jedinica

Klasifikator može procijeniti vjerojatnost pripadnosti kategoriji. Institucija zatim bira prag koji pokreće pregled i određuje daljnje postupanje. Procjena, odluka i njezine posljedice zasebni su dijelovi sustava.

Predviđanje procjenjujemo na novim podacima, a objašnjenje prema sadržajnom pitanju, teoriji i nacrtu (Breiman 2001.; Shmueli 2010.). Dobro predviđen ton ne otkriva namjeru govornika ni uzrok njegova izražavanja.

text unit
corpus boundary

Naša *tekstna jedinica* jest rečenica. Ona ne predstavlja cijeli govor, osobu ili sjednicu. *Granica korpusa* ovdje obuhvaća retke ParlaSenta s oznakom `country = HR`. Paket ne daje govornika, stranku, datum, položaj vlasti ili oporbe ni poveznicu na cijeli govor.

Nedostajući kontekst ne smijemo nadomjestiti nagađanjem. Izvorni autori izostavili su moderatorske rečenice i zadržali duljine između prvog i trećeg kvartila (Mochtak i sur. 2024.). Dio za učenje stratificirano su odabrali prema prisutnosti pozitivnih, negativnih ili nijedne leksikonske riječi. Ispitni dio odabrali su slučajno, uz iste filtre duljine i moderatora (Mochtak i sur. 2023., 2024.). Zbog takvog odabira udio negativnih oznaka ne procjenjuje udio negativnog tona u Hrvatskom saboru. Iz rečenice također ne zaključujemo o osobini ili namjeri govornika.

Nastavna prilagodba ParlaSenta 1.0 sadrži stvarne parlamentarne rečenice s oznakom zemlje `HR`, pod licencom CC BY-SA 4.0

(Mochtak i sur. 2023.). Ima 2 698 redaka. Sve rečenice istog dokumenta zadržane su u jednom izvedenom skupu.

Sljedeća tablica prikazuje broj redaka u svakom skupu. Račun čita pripremljenu datoteku, pa brojnosti možemo reproducirati.

Tablica 1: Oznake i brojnosti u nastavnom paketu ParlaSent. Izrada autora prema Mochtak i sur. (2023.).

Skup	Ukupno	Negativno	Neutralno	Pozitivno
Učenje	1 090	530	343	217
Provjera	272	122	90	60
Ispitivanje	1 336	560	546	230

Za učenje je izdvojeno 1 090 redaka, za provjeru 272, a za ispitivanje 1 336. Zadržani su svi izvorni ispitni redci. Iz izvornog dijela za učenje uklonjeno je 25 redaka iz 20 dokumenata koji su se pojavljivali i u ispitnom dijelu. Uklanjanje nije ovisilo o oznaci. Koraci i kontrolne sume zapisani su u podatkovnoj putovnici data/parlament_tekst/PUTOVNICA.md. Te brojnosti opisuju nastavni paket.

Nastanak oznake

coding frame

Nakon izbora rečenica trebamo razjasniti što njihove oznake tona znače i kako su nastale. *Okvir kodiranja* određuje značenje oznaka `Negative`, `Neutral` i `Positive`. Dodijeljena oznaka ovisi o uputama, prosudbi kodera i načinu rješavanja neslaganja.

Izvorni dio za učenje sadrži dvije pojedinačne oznake i usklađenu oznaku. U ispitnom dijelu dostupna je samo oznaka jednog uvježbanog kodera. Iako oba dijela imaju stupac `recorded_label`, oznake nisu nastale jednakim postupkom.

> **Zabilježeni referentni ishod** jest ishod ili oznaka zapisana određenim postupkom mjerenja, prema kojoj se vrednuje klasifikator, ali koja može sadržavati pogrešku.

Vratimo se vašoj prvoj procjeni. U nastavnoj datoteci početnu je rečenicu jedan koder označio neutralnom uz negativan ton, a drugi negativnom. Nakon usklađenja zabilježena je oznaka `Negative` (Mochtak i sur. 2023.). Usporedite to sa svojim razlogom. Je li vas na kritiku navela riječ „pošteno", oblik pitanja ili pretpostavka o okolnom govoru koji ovdje nemamo? Slaganje s konačnom oznakom samo po sebi još ne objašnjava zašto je ona opravdana.

Neslaganje kodera može upozoriti na nejasan pojam ili teško primjenjivo pravilo. Drukčije upute ili usklađivanje mogu promijeniti referentnu oznaku. Time se mijenja i ocjena klasifikatora, čak i ako njegova predviđanja ostanu ista.

Tri skupa, dva pitanja

Jedno je pitanje kojim ćemo modelom i pragom razvrstavati rečenice. Drugo je koliko taj zaključani postupak uspješno radi na novim rečenicama.

U regresiji smo odvojili podatke na kojima ćemo provjeriti predviđanje novih osoba. Ako sada na izdvojenim podacima biramo model ili prag, i oni utječu na konačni postupak. Za njegovu završnu provjeru zato trebamo sačuvati još jedan skup. *Razdvajanje na skup za učenje, provjeru i ispitivanje* odvaja tri zadatka. Model prilagođavamo na učenju, pravilo i prag biramo na provjeri, a završnu uspješnost jednom procjenjujemo na sačuvanom ispitnom skupu. Ovdje ispitni dio ima drukčiji odabir i način označavanja, pa ga ne možemo koristiti za izravno usporedivu završnu provjeru.

train-validation-test split

Paket razdvaja cijele dokumente. Dokumente iz izvorne ispitne datoteke uklonili smo iz dijela za učenje, a preostale prema po-

novljivom pravilu podijelili na učenje i provjeru. Tako rečenice istog dokumenta ne prelaze granice skupova i ne stvaraju privid provjere na neovisnom tekstu. Takvo preklapanje dokumenata bilo bi još jedan oblik curenja informacija.

overfitting

Preprilagodba nastaje kada model nauči posebnosti viđenih podataka pa njima bolje pristaje, ali lošije predviđa nove jedinice. Zato ga provjeravamo na odvojenim jedinicama. Populacijski doseg zasebno određujemo prema odabiru podataka. Dobar rezultat na pravilno izdvojenom skupu podržava prediktivnu uspješnost u obuhvaćenim uvjetima. Prijenos na druge tekstove zahtijeva dodatnu provjeru. Ovdje različit odabir i označavanje ispitnog dijela dodatno ograničavaju usporedbu.

Od vjerojatnosti do pravednosti

Uz prag 0,60 rečenica s procjenom 0,59 ne ulazi u pregled, a ona s 0,61 ulazi. Mala razlika u broju tako mijenja postupanje. Izbor praga treba povezati s posljedicama obiju vrsta pogreške.

> **Klasifikacijski prag** jest unaprijed određena vrijednost koja predviđenu vjerojatnost pretvara u kategoričku odluku. Njegovim se pomicanjem mijenja odnos vrsta pogrešaka.

Svaka rečenica sada završava među poslanima na pregled ili među izostavljenima. Tu odluku možemo usporediti s njezinom zabilježenom oznakom, kao što smo u poglavlju o kategoričkim podacima križali dob i izvor vijesti. Pozitivna odluka ovdje znači slanje na pregled, a tražena je oznaka negativnog tona. *Tablica zabune* jest kontingencijska tablica odluka i referentnih oznaka.

confusion matrix

Stopa lažno pozitivnih odluka računa se među zabilježenim nenegativnim rečenicama. Stopa lažno negativnih računa se među zabilježenim negativnima. Pozitivna prediktivna vrijednost pita koliko je zabilježenih negativnih među svima poslanima na pregled. Ukupna točnost spaja uspjeh u obje skupine prema nji-

hovim brojnostima, pa može sakriti neravnomjerno raspoređene pogreške.

Nazivnik određuje značenje svake stope. Osjetljivost, odnosno odziv (*recall*), računa se kao $TP/(TP + FN)$, a specifičnost kao $TN/(TN + FP)$. Pozitivna prediktivna vrijednost, u klasifikaciji zvana i preciznost (*precision*), računa se kao $TP/(TP + FP)$. Točnost (*accuracy*) jest $(TP + TN)/n$. Kratice TP i TN znače točno pozitivne i točno negativne odluke, a FP i FN lažno pozitivne i lažno negativne. Klasifikacijska preciznost zato nije isto što i uski interval procjene.

Brojeve ćemo u razrađenom primjeru dobiti na istim rečenicama izvedenog skupa za provjeru, prebrojavanjem ljudskih glasova i usporedbom pravila koja se na njima temelje. Za prag od barem jednoga negativnog glasa tablica daje $TP = 122$, $FP = 16$, $FN = 0$ i $TN = 134$. Osjetljivost je $122/122$, specifičnost $134/150$, pozitivna prediktivna vrijednost $122/138$, a točnost $256/272$. Svaki nazivnik izdvaja drugu skupinu rečenica, pa zajedno daju potpuniji opis odluke. Lažna uzbuna podsjeća na pogrešku prve vrste iz testiranja, ali ovdje brojimo odluke prema referentnoj oznaci pojedinih rečenica, a ne odbacivanja statističke hipoteze kroz ponovljene uzorke.

Temeljna stopa ovdje je udio zabilježenih negativnih ishoda prije odluke modela. Ako se razlikuje među skupinama, jednake stope lažno pozitivnih i lažno negativnih odluka mogu dati različite pozitivne prediktivne vrijednosti. Zato se neka poželjna mjerila pravednosti ne mogu uvijek istodobno izjednačiti (Chouldechova 2017.; Barocas i sur. 2023.). Uz broj pogrešaka trebamo razmotriti i njihove posljedice.

subgroup errors

Za *pogreške po podskupinama* trebamo pouzdanu skupnu varijablu. Nastavni paket takve podatke nema. Ne izvodimo ih nagađanjem iz teksta ili imena. Zato za prikaz koristimo izmišljene skupine A i B, a empirijski primjer ostaje bez podskupinske analize. Raspored pogrešaka među stvarnim podskupinama zato ostaje otvoreno pitanje.

algorithmic fairness

Algoritamska pravednost zahtijeva više od jedne stope. Uz nazivnike i posljedice pogrešaka razmatramo obavijest, obrazloženje te mogućnost prigovora i žalbe.

Istraživač pravednosti

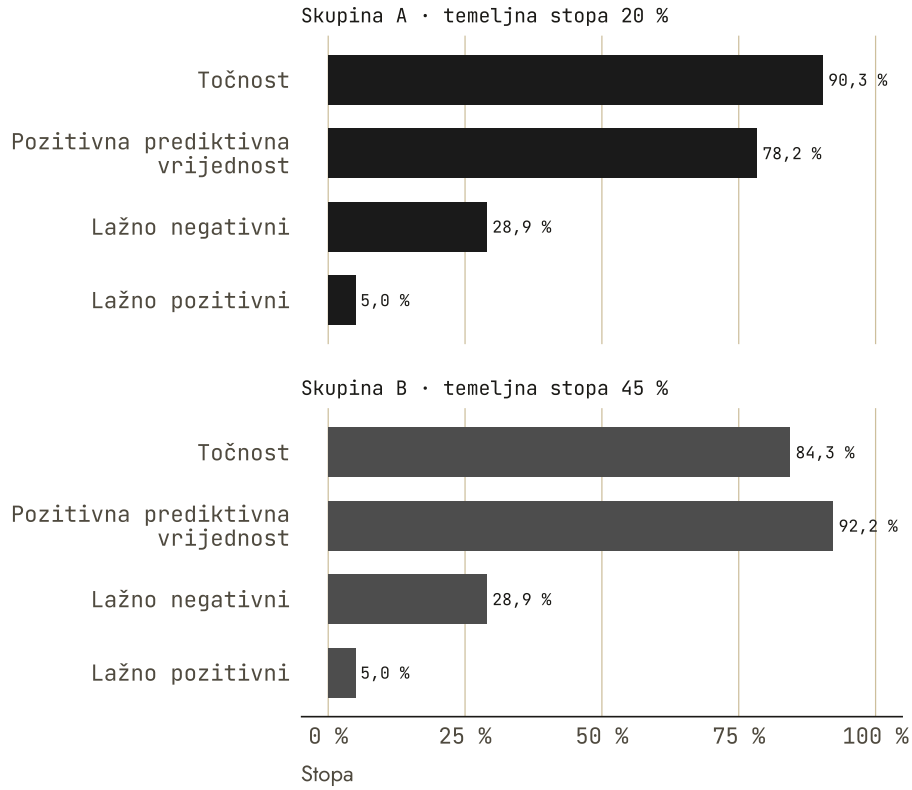
Za ovaj pokus zadajemo referentne ishode 0 i 1, ne ton stvarnih rečenica. Za ishod 0 generiramo 6 000 rezultata iz normalne raspodjele sa sredinom 0,30 i standardnom devijacijom 0,18, a za ishod 1 drugih 6 000 sa sredinom 0,70 i istom standardnom devijacijom. Vrijednosti ispod nule zamjenjujemo nulom, a iznad jedan jedinicom. Tako dobiveni brojevi služe razvrstavanju, bez pretpostavke da su kalibrirane vjerojatnosti. Kalibracija iz poglavlja o vjerojatnosti tražila bi slaganje najavljenih izgleda s učestalošću ishoda među usporedivim prognozama. Ovdje to nismo provjerili.

Pri pragu 0,60 rezultat 0,65 vodi pozitivnoj odluci. Ako je referentni ishod 0, taj zapis ulazi u brojnik lažno pozitivnih. Taj broj dijelimo sa svih 6 000 zapisa ishoda 0. Analogno među 6 000 zapisa ishoda 1 brojimo rezultate ispod praga za stopu lažno negativnih. Sve vrijednosti zadržavamo, uključujući one ograničene na nulu i jedan.

Skupine A i B zatim slažemo ponderiranjem istih dviju uvjetnih raspodjela, u prikazanoj postavci s udjelima ishoda 1 od 20 % i 45 %. Obje se skupine temelje na istim generiranim rezultatima, uz različite težine. Primjerice, u A udio točno pozitivnih među svima iznosi $0{,}20(1 - \text{stopa lažno negativnih})$, a udio lažno pozitivnih $0{,}80 \times$ stopa lažno pozitivnih. Njihov zbroj nazivnik je pozitivne prediktivne vrijednosti. U B mijenjaju se težine u 0,45 i 0,55. Uvjetne stope zato su u prikazu točno jednake po konstrukciji. Novi niz simuliranih rezultata malo bi promijenio njihove vrijednosti.

Na slici uspoređujemo dvije simulirane skupine s različitim temeljnim stopama. Za isti referentni ishod raspodjele rezultata

jednake su u obje skupine. Zajednički prag zato može dati jednake uvjetne stope pogreške, ali različitu točnost i prediktivnu vrijednost. Zaključci prikaza odnose se na taj zaseban simulirani model.



Slika 1: Istraživač pravednosti — jednake uvjetne stope pogreške ne daju jednaku prediktivnu vrijednost kada se temeljne stope razlikuju. Izrada autora. Podaci i postupak opisani su u tekstu.

Kada su jednake i temeljne stope, sva četiri mjerila poklapaju se. Kada temeljne stope promijenimo, razlikuju se pozitivna prediktivna vrijednost i točnost, iako uvjetne stope pogreške ostaju jednake. Prikaz objašnjava matematički odnos koji smo susreli u poglavlju o zavaravajućim brojkama. Sam ne dokazuje nejednako postupanje u stvarnom sustavu.

Osporiva oznaka i postupak

U simulaciji smo točno znali koji je referentni ishod. Vratimo se sada početnoj rečenici, čija je oznaka nastala ljudskom prosudbom. Nakon usklađenja nosi negativnu oznaku, dok je prvi koder odabrao neutralnu kategoriju s negativnim tonom. Trebamo pregledati upute i obrazloženje te procijeniti bi li drugo opravdano pravilo dalo drukčiji rezultat. Takva provjera može dovesti do ispravka reference prema kojoj ocjenjujemo odluke.

Osoba koja pregledava rečenicu treba vidjeti izvorni tekst, pravilo i razlog slanja na pregled. Treba moći zabilježiti obrazloženje i osporiti oznaku. Ako odluka utječe na govornika ili javni prikaz njegova govora, potreban je i postupak obavijesti i žalbe. Paket bez identiteta govornika to sam ne omogućuje. Ljudski pregled također treba jasne upute, odgovornost i nadzor.

Promjena referentne oznake mijenja ćelije tablice zabune. Prethodno lažno pozitivna odluka može postati točno pozitivna bez promjene predviđanja. Zato čuvamo inačicu kodnih pravila, datum i razlog ispravka. Samo tako možemo uspoređivati mjerila kroz vrijeme.

U primjeni pratimo i podatke, prag, sučelje, pregled, prigovore i daljnje učenje. Ako nove oznake dobivaju samo pregledane rečenice, sustav uči uglavnom iz teksta koji je već odabrao. Pogreške među izostavljenim rečenicama teže će biti otkrivene.

Tako nastaje povratna sprega između odluka i budućih podataka. Slično je kod preporuka. Rangiranje određuje što korisnici vide, a njihovo ponašanje postaje podatak za novo rangiranje. Vrijeme zadržavanja pritom ne mjeri automatski zadovoljstvo, informiranost ni javnu vrijednost (Barocas i sur. 2023.).

distribution shift

Pomak distribucije može nastati zbog novih tema, jezika, odabira tekstova ili načina označavanja. Nakon primjene pratimo pregledane i nepregledane tekstove, sporne oznake, vrijeme ispravka i promjene korpusa. Za procjenu lažno negativnih odluka trebamo zasebno i usporedivo označiti slučajni ili unaprijed

stratificirani uzorak nepregledanih rečenica. Ranija uspješnost ne jamči trajno dobar rad ni valjano mjerenje tona.

Posljedice takvih odluka lakše je zamisliti kada ocjena određuje vlastitu priliku za nastavak školovanja. Nakon otkazivanja ispita u Engleskoj 2020. Ofqual je primijenio statistički postupak standardizacije ocjena, oslonjen na nastavničke procjene i poredak učenika unutar predmeta (Ofqual 2020.). U izvještaju je posebno razmotrio učenike čije se prethodno postignuće znatno razlikovalo od postignuća ostalih u njihovoj školi, zbog čega im izračunana ocjena može biti nepouzdana (Ofqual 2020.). Zamislite da upravo vama ocjena zatvori željeni studij. Slaganje ukupne raspodjele s očekivanjem ne odgovara na pitanje je li vaša ocjena opravdana.

Nakon gubitka javnog povjerenja konačno se dodjeljivala viša od nastavničke procjene i standardizirane izračunane ocjene (Ofqual 2020.). Taj slučaj statističke standardizacije pokazuje zašto uz rad algoritma trebamo razmotriti mogućnost obrazloženja, ispravka i žalbe. Budući da ispiti nisu održani, ne možemo jednostavno usporediti svaku dodijeljenu ocjenu s ocjenom koju bi učenik na njima ostvario. U parlamentarnom primjeru također trebamo provjeriti što referenca doista bilježi. Posljedice su drukčije, ali u oba slučaja kvaliteta zajedničkog pravila uključuje i postupak za osobu koju ono pogrešno opisuje.

Jezični modeli kao sustavi predikcije

language model as a prediction system

I asistent kojim se služimo proizvodi predviđanja. *Jezični model kao sustav predikcije* možemo razumjeti na primjeru autoregresivnog GPT-3 iz 2020., koji generira nastavke uvjetovane prethodnim tekstom (Brown i sur. 2020.). Tečan odgovor zato ne potvrđuje izvor, brojku ni zaključak. Predviđanje teksta samo po sebi ne daje ni uzročno objašnjenje odgovora.

U odgovoru asistenta provjeravamo podatke, jedinicu, planiranu odluku i način otkrivanja pogreške. Asistent može predložiti ra-

čun ili tablicu. Ne smije nadomjestiti nedostajuća predviđanja izmišljenim stupcem, nagađati kontekst ili tretirati referentnu oznaku kao nepogrešivu.

Odnos temeljnih stopa i mjerila susrećemo i u drugim primjenama klasifikacije.

STATISTIKA U DIVLJINI

Jednaka ocjena, različite pogreške. Procjena da je osoba visokorizična može utjecati na odluku koja se na nju odnosi. Chouldechova je u istraživanju instrumenata za procjenu ponavljanja kaznenog djela razmotrila kako različita mjerila pravednosti mogu dati različite ocjene istog sustava (Chouldechova 2017.). Procjena sustava zato zahtijeva pregled vrsta pogrešaka i njihovih posljedica uz ukupnu točnost.

Jedan zahtjev polazi od značenja ocjene. Među ljudima s jednakom ocjenom rizika zabilježeni bi ishod trebao biti podjednako čest u različitim skupinama. Drugi zahtjev polazi od pogrešno označenih ljudi. Među onima kod kojih ishod nije zabilježen želimo usporediti koliko ih je ipak označeno visokorizičnima. Ti zahtjevi koriste različite nazivnike, pa jednakost prema jednome ne jamči jednakost prema drugome.

Za cijelu skupinu označenu visokorizičnom možemo tražiti jednak udio zabilježenih pozitivnih ishoda. Chouldechova je pokazala zašto taj zahtjev pri različitim temeljnim stopama u pravilu dolazi u sukob s jednakošću obiju stopa pogreške (Chouldechova 2017.). Takav sukob ne javlja se kod savršenog predviđanja, ali njega ne možemo pretpostaviti u stvarnoj primjeni. Taj matematički nalaz ne određuje koju bi pogrešku društvo trebalo smatrati prihvatljivom niti opravdava njezine posljedice za osobu.

Zato bi prije uporabe takve procjene trebalo razjasniti što referentni ishod bilježi, kome pogrešna oznaka šteti i kako se odluka može osporiti. Promjena praga može promijeniti raspored pogrešaka, ali ne može sama dati odgovor na ta pitanja. Ni dobro slaganje s postojećim zapisima ne potvrđuje da oni potpuno opisuju pojavu koju želimo procjenjivati.

U našem parlamentarnom primjeru posljedice i jedinice drukčije su. Pogrešno uključena rečenica dobiva dodatni pregled, a pogrešno izostavljena

ostaje izvan njega. Kaznenopravne skupine i njihove temeljne stope ne prenosimo u taj korpus. Prenosimo način postavljanja pitanja. Tvrdnja da je sustav pravedan treba imenovati ishod, skupine, prag i pogrešku čije posljedice procjenjujemo. Bez toga jednaka ocjena modela može prikrivati različita iskustva ljudi na koje se odnosi.

Plan vrednovanja treba odrediti što će se provjeravati i koju će odluku dobiveni rezultat poduprijeti. Sljedeći prijedlog povezuje te dvije zadaće.

NAĐITE GREŠKU

Tim priprema plan vrednovanja klasifikatora parlamentarnih rečenica. U nastavnom paketu ParlaSent rečenice su razdvojene po dokumentima. Izvorni dio za učenje ima dvije pojedinačne i usklađenu oznaku, dok ispitni dio ima oznaku jednog uvježbanog kodera (Mochtak i sur. 2023., 2024.). To su dva različita puta nastanka oznake koje smo već upoznali.

Predloženi plan predviđa zaključavanje modela i praga prije otvaranja ispitnog skupa. Točnost bi se računala kao udio predviđanja koja se podudaraju sa zabilježenim oznakama. Tim još nije proveo tu provjeru ni neovisno ponovno kodiranje ispitnih rečenica. Asistent je predložio rečenicu za plan.

> Izdvojeni skup omogućit će provjeru slaganja predviđanja sa zabilježenim oznakama, uz ograničenja odabira i različitog kodiranja. Ako model na njemu postigne visoku točnost, time ćemo dokazati da ispravno mjeri ton rečenica.

Uredite plan tako da jasno kaže što će se tom provjerom doznati. Navedite jednu dodatnu provjeru mjerenja i objasnite što bi ona mogla otkriti čak i kada bi podudarnost modela s postojećim oznakama bila visoka. Ne pripisujte klasifikatoru rezultat koji još nije izmjeren.

Razrađeni primjer

Neslaganje kodera iz početne rečenice sada pretvaramo u pitanje o pravilu pregleda. Je li dovoljan jedan negativan glas ili ćemo zahtijevati oba? Usporedbu provodimo samo na 272 retka skupa za provjeru, s dvije pojedinačne i usklađenom oznakom. Iz glasova kodera računamo rezultat 0, 0,5 ili 1. Oznake `Negative` i `M_Negative` računamo kao negativne, a `Positive`, `M_Positive`, `N_Neutral` i `P_Neutral` kao nenegativne. Ta nastavna podjela zahtijeva provjeru osjetljivosti. Rezultat nije vjerojatnost klasifikatora ni neovisno predviđanje. Usklađena referenca nastala je iz istih ulaza, pa primjer pokazuje kako način označavanja oblikuje mjerila.

To objašnjava i nulu u ćeliji lažno negativnih pri blažem pravilu. Svaka referentno negativna rečenica u ovom paketu ima barem jedan negativan glas. Nula stoga proizlazi iz odnosa glasova i njihove usklađene oznake, a nije dokaz savršenog prepoznavanja novih rečenica. Pri strožem pravilu postoji jedna rečenica s dva negativna glasa i nenegativnom usklađenom oznakom, `train-2089`. To je stvarno neslaganje zapisa, koje čuvamo vidljivim.

Od svih 2 698 zapisa sada zadržavamo samo 272 s oznakom `provjera`, bez daljnjeg izostavljanja. Za mali prikaz namjerno biramo prvi redak sa svakom od triju mogućih vrijednosti glasovanja. Broj retka odnosi se na cijelu nastavnu datoteku, pa se zapis može ponovno pronaći.

Tablica 2: Tri odabrana koderska zapisa prije tablice zabune. Izrada autora prema Mochtak i sur. (2023.).

Redak	Glasovi	Rezultat	Referenca	Pregled pri 0,5
3	0 + 0	0,0	nenegativna	ne
38	1 + 0	0,5	negativna	da
4	1 + 1	1,0	negativna	da

Dva negativna glasa daju $(1 + 1)/2 = 1$, a jedan $(1 + 0)/2 = 0{,}5$. U svakom retku odluku uspoređujemo s referencom. Slanje uz negativnu referencu dodaje jedan točno pozitivnima, a slanje uz nenegativnu jedan lažno pozitivnima. Neslanje analogno puni preostale dvije ćelije. Tako prebrojavamo svih 272 zapisa, bez procjene klasifikatora.

Prvo pravilo postavlja prag na 0,5, pa rečenica ulazi na pregled ako je barem jedan koder dao negativan glas. Taj prag primjenjujemo na glasove, za razliku od ranijeg praga modelom predviđene vjerojatnosti. Poziv `mutate()` donosi odluku za svaki redak, a `count()` zatim prebroji njezine kombinacije s referentnom oznakom u četiri ćelije tablice.

```
provjera_glasovi_c17 |>
  mutate(u_red = negativni_glasovi >= 0.5) |>
  count(referentno_negativno, u_red) |>
  tidyr::complete(referentno_negativno = c(FALSE, TRUE),
    u_red = c(FALSE, TRUE), fill = list(n = 0))
```

Brojanja pri prvom pragu možemo pročitati kao običnu kontingencijsku tablicu. Njezini se redci odnose na referentnu oznaku, a stupci na odluku.

Tablica 3: Brojanja za prag 0,5 na istom skupu za provjeru. Izrada autora prema Mochtak i sur. (2023.).

Referentna oznaka	Poslano na pregled	Nije poslano	Ukupno
negativna	122	0	122
nenegativna	16	134	150
ukupno	138	134	272

Stopa lažno pozitivnih je $16/150$, a pozitivna prediktivna vrijednost $122/138$. Ista ćelija zato može sudjelovati u različitim pitanjima, uz drukčiji nazivnik.

Da bismo vidjeli što traženje suglasnosti obaju kodera mijenja, na iste rečenice primjenjujemo i prag 1. Sljedeća tablica uspoređuje ga s pragom 0,5, uz isti zabilježeni referentni ishod. Sve su vrijednosti izračunate iz pripremljenog paketa.

Tablica 4: Dvije odluke o ljudskom pregledu prema istom zabilježenom referentnom ishodu u skupu za provjeru. Izrada autora prema Mochtak i sur. (2023.).

Mjera	Barem jedan glas	Dva glasa
TP	122	100
FP	16	1
FN	0	22
TN	134	149
FP / (FP + TN)	10,7 %	0,7 %
FN / (FN + TP)	0,0 %	18,0 %
TP / (TP + FP)	88,4 %	99,0 %
(TP + TN) / N	94,1 %	91,5 %

U tablici prve četiri brojnosti koriste upravo definirane kratice, a sljedeća tri retka pokazuju nazivnike stopa. Posljednji redak daje ukupnu točnost, udio ispravnih odluka među svim rečenicama. Jedna lažno pozitivna odluka pri strožem pragu znači $1/150$, odnosno 0,7 %. Još jedna takva odluka udvostručila bi taj postotak. Taj mali broj ne opravdava preciznu tvrdnju o stopi u budućim tekstovima.

Pri nižem pragu na pregled ulaze sve 122 rečenice sa zabilježenim negativnim ishodom i 16 od 150 nenegativnih. Stopa lažno pozitivnih odluka iznosi 10,7 %, a lažno negativnih 0,0 %. Pri strožem pragu izostaju 22 od 122 negativne rečenice, a ulazi jedna od 150 nenegativnih. Stope tada iznose 0,7 % i 18,0 %.

Za izbor pravila trebamo znati kapacitet pregleda, posljedice nepotrebnog uključivanja i izostavljanja te mogućnost prigovora. Prije toga uredništvo mora opravdati izbor negativnog tona kao kriterija. Treba ga usporediti sa slučajnim pregledom ili kriterijem relevantnosti. Tablica pokazuje posljedice praga, a uredništvo na toj osnovi i prema svrsi pregleda obrazlaže odluku.

U ispitnom dijelu nemamo drugog kodera ni usklađenje, pa pravilo dvaju glasova ne možemo jednako primijeniti i ocijeniti.

Završnu uspješnost zato ne izvještavamo. Za budući klasifikator trebali bismo unaprijed zaključati model, prag i usporediv referentni postupak, spremiti predviđanja i jednom provesti završnu provjeru. Ovdje provjeravamo podatke i pravilo odluke, bez tvrdnje o izvedbi modela koji nije izgrađen.

Sada znamo zašto je oznaka početne rečenice sporna i što bi značilo uključivanje u pregled. Još ne znamo namjeru govornika. Odluku treba vezati uz izvorni tekst, obrazloženje, mogućnost ispravka i praćenje posljedica praga.

Od modela do odluke

U tablici provjeravamo jedinicu i obuhvat podataka, vrstu tvrdnje, nesigurnost, alternative i osobe na koje odluka može utjecati. Tako određujemo doseg ovog primjera.

Tablica 5: Šest revizijskih pitanja za odluku o ljudskom pregledu. Izrada autora prema Mochtak i sur. (2023.) i Mochtak i sur. (2024.).

Revizijsko pitanje	Primjena na odluku o ljudskom pregledu
Koja je jedinica?	jedna odabrana parlamentarna rečenica s putom nastanka zabilježene oznake, ne govor, govornik ili sjednica
Tko ili što nedostaje?	rečenice moderatora i rečenice izvan srednjega raspona duljine isključene su uzvodno. Paket nema govornika, cijeli govor ni obranjivu skupnu varijablu
Koji je cilj tvrdnje?	tablica uspoređuje dva pravila slanja u pregled prema zabilježenom negativnom ishodu. Riječ je o opisu i provjeri odluke, ne izvedbi postojećega klasifikatora
Gdje je neizvjesnost?	osjetljivost na prag. Ne provjeravamo reprezentativnost, valjanost oznake, pogrešku kodiranja ni budući pomak distribucije
Koje drugo objašnjenje ostaje?	stroži prag mijenja pogreške. Slučajni pregled ili kriterij relevantnosti mijenjaju i cilj odabira
Tko snosi posljedice?	pregledavatelji troše vrijeme, važan tekst može izostati, a govornik i javnost mogu dobiti iskrivljen prikaz

Sljedeća tablica razlikuje šest vrsta tvrdnji. Doseg analize određujemo prema dokazima potrebnima za svaku od njih.

Tablica 6: Šest dimenzija tvrdnje na završetku dijela o modelima. Izrada autora prema Mochtak i sur. (2023.).

Dimenzija tvrdnje	Što ovaj dokaz dopušta
opis	opisuje brojnosti, put oznake i posljedice dvaju pravila unutar pripremljenoga paketa
povezanost	pokazuje povezanost dvaju koderskih glasova s usklađenom oznakom, uz izričitu ovisnost jer su isti glasovi sudjelovali u njezinu nastanku
generalizacija	ne podupire prijenos na sav parlamentarni govor, govornike ili tekstove izvan odabranih rečenica
predviđanje	ne ocjenjuje postojeći model. Buduća bi procjena tražila zaključan model, prag, usporediv referentni postupak i jedan izdvojeni ispitni skup
uzročnost	ne pokazuje što uzrokuje ton, uredničku važnost ili bilo koju posljedicu za govornika
odluka	podupire provjeru tereta dvaju pragova, ali ne dokazuje da je negativan sentiment legitiman cilj usmjeravanja pozornosti

Zašto razdvajanje na skup za učenje, provjeru i ispitivanje ne stvara populacijsku reprezentativnost? Kako izbor rečenica i način njihova označavanja ograničavaju ono što možemo zaključiti

iz tablice zabune? Zašto jednake uvjetne stope pogreške ne jamče jednaku pozitivnu prediktivnu vrijednost? Nakon primjene sustava, koje rečenice moramo dodatno pregledati da bismo mogli procijeniti koliko je negativnih slučajeva sustav propustio?

U završnom poglavlju sastavit ćemo vlastiti istraživački protokol. U njega uključujemo pragove odluke, posljedice pogreške, nadzor i prigovor. Asistentu možemo povjeriti račun i provjeru dosljednosti, dok cilj, prosudba dokaza i odgovornost za zaključak ostaju naši.

Zaštitu osoba povezujemo s pravilima iz poglavlja o mjerenju i dizajnu. Javna dostupnost zapisa sama ne uklanja mogućnost identifikacije ni štetu od njegova povezivanja s drugim podacima. Prije dijeljenja provjeravamo svrhu, nužnost pojedinih obilježja i prikladnost alata za osjetljivost podataka. Pravila određenoga kolegija dodatni su uvjet u toj nastavi, a nisu univerzalna zamjena za odgovornost prema sudionicima.

Sažetak

Ocjenu algoritma počinjemo pitanjem što model može naučiti iz odabranih podataka i što možemo provjeriti prema referentnim oznakama. Odvajanje skupova pomaže provjeri predviđanja, ali samo ne stvara reprezentativnost ni valjano mjerenje. Kada rezultat pretvorimo u odluku, prag mijenja raspodjelu pogrešaka, a različite temeljne stope mogu dovesti mjerila pravednosti u sukob. Nakon primjene ostajemo odgovorni za pregled, ispravke, prigovor i nadzor, uključujući provjeru izostavljenih slučajeva. Izvore i zaključke provjeravamo i u odgovoru jezičnog modela, koliko god bio tečan.

Pojmovi

osjetljivost klasifikatora (*sensitivity*), specifičnost (*specificity*), preciznost odnosno pozitivna prediktivna vrijednost (*precision*, *positive predictive value*), ukupna točnost (*accuracy*),

tekstna jedinica (*text unit*), granica korpusa (*corpus boundary*), okvir kodiranja (*coding frame*), zabilježeni referentni ishod (*recorded reference outcome*), razdvajanje na skup za učenje, provjeru i ispitivanje (*train-validation-test split*), preprilagodba (*overfitting*), klasifikacijski prag (*classification threshold*), tablica zabune (*confusion matrix*), pogreške po podskupinama (*subgroup errors*), algoritamska pravednost (*algorithmic fairness*), pomak distribucije (*distribution shift*), jezični model kao sustav predikcije (*language model as a prediction system*)

Zadaci

Konceptualni

Objasnite kako razdvajanje na skup za učenje, provjeru i ispitivanje otkriva preprilagodbu te zašto dokument koji prijeđe među skupovima stvara curenje informacija. Zatim povežite tablicu zabune s kontingencijskom tablicom iz poglavlja o kategoričkim podatcima i objasnite zašto predviđena vjerojatnost iz poglavlja o regresiji još nije odluka. Imenujte oba uvjetna nazivnika pogreške.

Računski

U pripremljenoj tablici odluka nalazi se 90 točno pozitivnih, 30 lažno pozitivnih, 10 lažno negativnih i 170 točno negativnih slučajeva. Izračunajte stopu lažno pozitivnih odluka, stopu

lažno negativnih odluka, pozitivnu prediktivnu vrijednost i točnost. Uz svaki rezultat napišite njegov nazivnik i jednu moguću posljedicu za red ljudskoga pregleda.

Kritički

Osporite zabilježenu oznaku početne rečenice. Napišite koje biste dijelove okvira kodiranja, izvornih oznaka i usklađenja trebali vidjeti prije odluke. Zatim objasnite zašto promjena oznake može promijeniti mjerila pravednosti bez ijedne promjene modela.

Revizija modela

Ocijenite tvrdnju iz okvira o pogrešci. Navedite jednu prihvatljivu tvrdnju, jedan pogrešan prijelaz i dokaz koji pokazuje razliku. Zatim u sažetom odgovoru revidirajte tvrdnju tako da obuhvati jedinicu, granicu korpusa, put oznake, curenje, prag, dostupne podskupine, pomak distribucije te postupak obavijesti, prigovora i žalbe. Objasnite zašto tečan odgovor jezičnoga modela ne dokazuje nijednu nedostupnu sastavnicu.



završnica

Povezujemo pitanje, podatke, model i izvještaj u jednu provjerljivu cjelinu. Pratimo jedan regresijski primjer i bilježimo odluke koje određuju njegov doseg. Asistent pomaže u radu, dok provjera i odgovornost ostaju naše.

18

ZAVRŠNICA

vaše prvo istraživanje

Od pitanja do provjerljivog zaključka

Tim ima tri tjedna za završni projekt, zanimljivu temu i datoteku koju još treba pripremiti. Jedan član želi odmah procijeniti model, drugi prvo provjeriti mjerenje, a treći predlaže da posao ubrzaju slanjem datoteke javnom asistentu.

Prije toga trebaju razjasniti što predstavlja redak, odakle podaci dolaze i koje pitanje žele postaviti. Trebaju provjeriti i sadrži li datoteka informacije po kojima se ljudi mogu prepoznati.

Kako provesti prvo istraživanje tako da druga osoba može pratiti odluke i provjeriti zaključak?

U završnom projektu iskustvo čitanja tuđih zaključaka primjenjujemo na vlastito istraživanje. Postupak trebamo sastaviti tako da ga druga osoba može pratiti. Pitanje, mjerenje, priprema

podataka, graf i model moraju se odnositi na istu istraživačku namjeru. Te odluke povezujemo u jednom istraživanju i ostavljamo zapis koji omogućuje provjeru rezultata.

Vraćamo se istim zapisima svih 300 zamišljenih osoba iz skupa `anketa_mreze`, koje poznajemo iz sažimanja, vizualizacije i povezanosti. Svaki redak predstavlja jednu zamišljenu osobu s dobi, dnevnim minutama na društvenim mrežama i povjerenjem u tamošnji sadržaj. To je zaseban skup od populacije iz koje smo izvlačili uzorke u inferencijalnim poglavljima. Istražit ćemo jesu li duže dnevno korištenje mreža i veće povjerenje povezani te kako se procjena mijenja kada uvažimo dob. Time vježbamo istraživanje odnosa, bez tvrdnje da smo prikupili dokaze o stvarnim korisnicima ili uzroku njihova povjerenja.

Za svaki poznati postupak sada trebamo objasniti zašto ga koristimo. Odredit ćemo pitanje i granicu nacrta, pregledati zapise i dokumentirati pripremu, a zatim opisati raspodjele i nacrtati odnos. Regresijsku procjenu s intervalom usporedit ćemo s opravdanom alternativom. Razlike među rezultatima vratit će nas na odluke o tome što model uspoređuje. Iz tog ćemo rada sastaviti izvještaj koji se može razumjeti bez otvaranja računalnog zapisa, uz sačuvan trag za ponavljanje računa.

Postupak ćemo potom provjeriti na stvarnim parlamentarnim rečenicama iz ParlaSenta, gdje ćemo umjesto osobnih odgovora pratiti izbor tekstova i nastanak oznaka (Mochtak i sur. 2023., 2024.). Za nove podatke oblikovat ćemo pitanje koje odgovara njihovoj jedinici i načinu nastanka.

U oba primjera odlučivat ćemo što povjeriti asistentu, kako provjeriti njegov rad i što smijemo dijeliti. Na kraju treba biti jasno kako smo od pitanja i podataka došli do zaključka i za koje odluke odgovaramo. Zato rad počinjemo od pitanja koje zaista možemo istražiti.

Pitanje prije podataka

Tema određuje područje interesa, a pitanje određuje što ćemo istražiti. Medijske navike, angažman na mrežama ili formati naslova tek su početak. Za istraživačko pitanje trebamo imenovati populaciju, jedinicu, mjere i usporedbu.

U tablici su tri moguća smjera semestralnog rada. Razlikuju se prema izvoru podataka i zaključcima koje nacrt dopušta.

Tablica 1: Tri izvediva smjera prvog istraživanja, prema izvoru podataka i granici zaključka. Izrada autora.

Smjer rada	Odakle podaci	Što dizajn dopušta
Anketa među kolegama	vlastito prikupljanje	opis skupine i povezanosti unutar nje
Analiza javno dostupnih objava	postojeći zapis	opis i predviđanje, bez uzročne tvrdnje
Usporedba dviju verzija poruke	nasumična dodjela uvjeta	usporedba koja uz očuvanu dodjelu podnosi ograničen uzročni jezik

Prva dva smjera promatraju postojeće podatke, a treći nasumično dodjeljuje uvjete. Kasnija analiza ne može promijeniti njihov dokazni doseg. I kod randomizacije uzročni zaključak ovisi o očuvanoj dodjeli, mjerenju, pridržavanju uvjeta, osipanju i mogućem utjecaju jedne skupine na drugu.

U vođenoj analizi skupa `anketa_mreze` ishod je povjerenje u sadržaj na društvenim mrežama na ljestvici od jedan do deset, a prediktor dnevno vrijeme uporabe u minutama. Dob ćemo uključiti kako bismo usporedili osobe jednakih godina. Time naše pitanje o vezi vremena i povjerenja dobiva određene varijable i usporedbu koju možemo izračunati.

Naše pitanje odnosi se na povezanost. Razlikujemo ga od opisa podataka, prijenosa na populaciju, predviđanja novih opažanja, procjene uzroka i odluke o postupanju. Regresiju ćemo tumačiti u okviru odabranog pitanja o povezanosti.

Već na početku dogovaramo što će drugoj osobi trebati da razumije i provjeri naš rad. Samostalan izvještaj treba prenijeti pitanje i zaključak, a ponovljiv zapis postupka pokazati kako je nastala svaka tablica i graf. Uz njih predajemo podatke koje smijemo dijeliti, kratko izlaganje i osvrt. Zapis može biti skripta ili dokumentirana povijest rada s izvozom iz alata bez koda. Izlaganjem ćemo dodatno provjeriti možemo li rezultat objasniti vlastitim riječima.

> **Paket dokaza** je provjerljiva cjelina koja povezuje putovnicu skupa podataka, zapis izvora i verzije, dnevnik transformacija, analizu, provjeru osjetljivosti, granicu tvrdnje, zapis o uporabi asistenta i objavu.

dataset passport

Svaki dio te cjeline odgovara na određeno pitanje. Najprije nas *putovnica skupa podataka* vraća izvoru, verziji i obuhvatu podataka kako bismo znali od čega je istraživanje počelo. Put od tog izvornika do redaka koje analiziramo čuva *dnevnik transformacija*. U njemu možemo slijediti svaku promjenu i ponoviti pripremu tablice.

transformation log

decision trail

confirmatory analysis

Razloge zbog kojih smo odabrali upravo te korake povezuje *trag odluka*. Važan je i trenutak u kojem je odluka donesena. Za *potvrđujuću analizu* pitanje i glavne postupke zapisujemo prije uvida u ishode. U izvještaju zato možemo odvojiti unaprijed planirane odluke od onoga što su nam tek podaci sugerirali. Kada nas opaženi obrazac potakne na novo pitanje, provodimo *istraživačku analizu* i jasno je tako označujemo.

exploratory analysis

sensitivity analysis

claim boundary

Da bismo vidjeli ovisi li zaključak o pojedinoj odluci, dodajemo *provjeru osjetljivosti*. Primarnu analizu uspoređujemo s obranjivom alternativom i bilježimo što se promijenilo u procjeni, intervalu i tumačenju. Uz te rezultate navodimo i *granicu tvrdnje*,

kako čitatelj ne bi morao nagađati što podaci i nacrt podupiru, a što ostaje nedostupno.

Kada druga osoba može od istoga izvora ponoviti promjene i dobiti naše rezultate, ostvarili smo *reproducibilan tijek rada*. Time joj otvaramo račun za provjeru, ali još trebamo objasniti njegovo značenje. To čini *pošten izvještaj*, u kojem uz procjenu i interval navodimo jedinice, dosegnutu populaciju i ograničenje koje bi moglo promijeniti zaključak.

reproducible workflow

honest report

Provjerljivost rada traži i dogovor o podacima i pomoći koju ćemo koristiti. U projektu zato primjenjujemo *minimizaciju podataka*, prema kojoj koristimo najmanji opseg podataka potreban za opravdanu svrhu. Da bi bilo jasno tko je pomogao u kojem koraku i kako je njegov rad provjeren, prilažemo i *objavu uporabe asistenta*. To su dijelovi datirane politike kolegija, a ne opća pravna tumačenja.

data minimization

AI-use disclosure

Dizajn određuje granicu zaključka

Presječna anketa bilježi vrijeme uporabe i povjerenje u istom trenutku, pa njihov vremenski redoslijed ostaje otvoren. Ista povezanost mogla bi nastati zato što uporaba povećava povjerenje, zato što povjerenje povećava uporabu ili zato što obje varijable ovise o trećoj.

Način uključivanja ispitanika određuje na koga se rezultat može odnositi. Prigodna anketa među kolegama sa studija nema sama po sebi osnovu za zaključak o svim studentima ili mladima. Veći broj odgovora iz istog ograničenog kruga ne rješava selekciju.

Mjere biramo prije prikupljanja. Dnevnu uporabu možemo mjeriti samoprocjenom, očitati iz izvještaja uređaja ili približno opisati brojem otvaranja aplikacije. Svaka mjera zahvaća uporabu na drukčiji način. Samoprocjenu je lako prikupiti, a veličinu i obrazac njezine pogreške trebamo zasebno provjeriti. U izvještaju zato jasno opisujemo način mjerenja.

Veličinu uzorka planiramo prema željenoj preciznosti i pretpostavkama o podacima. Potom provjeravamo izvedivost. Ako traženi broj nije dostupan, prilagodimo plan ili doseg pitanja prije prikupljanja.

U jednostavnom hipotetskom planu za srednju ocjenu povjerenja pretpostavimo standardnu devijaciju od 2 boda i ciljnu marginu od 0,5 boda za 95-postotni interval. Normalna aproksimacija daje $n \approx (1{,}96 \times 2/0{,}5)^2 = 61{,}47$, pa početni plan traži najmanje 62 potpuna odgovora. Taj račun pretpostavlja neovisni slučajni uzorak i zadanu standardnu devijaciju. Neodaziv i prigodni odabir traže zasebnu procjenu.

U planu unaprijed zapisujemo glavno pitanje, model i postupanje s nedostajućim odgovorima. Naknadne odluke također su dopuštene, ali ih u izvještaju označujemo kao istraživačke.

Možemo isprobati i analizu koju nismo planirali. Zapis nam omogućuje da je jasno razlikujemo od prvotnog postupka, kako smo naučili u poglavlju o krizi i obnovi.

Podaci i zapis odluka

Uz podatke bilježimo izvor, verziju, licencu, datum preuzimanja ili prikupljanja i način uključivanja jedinica. Taj zapis omogućuje provjeru podrijetla i obuhvata rezultata.

Prije procjene modela spremamo konačnu analitičku tablicu i opis njezina nastanka. Za naš simulirani primjer zapis izgleda ovako.

Tablica 2: Revizija konstrukcije analitičke tablice prije gledanja procjene. Izrada autora.

Odluka	Provedba u glavnoj studiji	Na što bi drukčija odluka utjecala
jedinica	jedna simulirana osoba	nazivnike, ovisnost redaka i smisao procjene
prihvatljivost	svih 300 jedinica iz generatora	skup ljudi koje rezultat opisuje
spajanja	nema spajanja tablica	broj i identitet redaka
prekodiranja	dob je sažeta u četiri skupine samo za opis i prikaz	položaj skupina na grafu, ne primarni model s dobi
transformacije	nagib po minuti prikazuje se kao razlika povezana s 30 dodatnih minuta	mjerilo, ali ne i prilagodbu modela
isključenja	nema isključenih redaka	doseg uzorka i moguću procjenu
nedostajuće vrijednosti	nema ih. U stvarnom radu naveli bismo broj po varijabli i pravilo postupanja	uzorak koji ulazi u svaki račun

Izvornu datoteku čuvamo neizmijenjenu. Popravke, preimenovanja i isključenja provodimo u odvojenim, dokumentiranim koracima koje možemo ponoviti. Tako se svaka vrijednost može pratiti do izvora.

Na početku pregledavamo imena, kontakte, adrese, otvorene odgovore i kombinacije obilježja koje mogu otkriti osobu. Odlu-

čujemo jesu li potrebni i kako ih smijemo koristiti. Pravila ovog projekta povezana su s poglavljem o mjerenju i Dodatkom F[1].

Pri čišćenju za svaku promjenu zapisujemo razlog.

1. Provjeravamo nedostajuće vrijednosti i bilježimo koliko ih je i gdje su.
2. Tražimo nemoguće vrijednosti, dakle dob od dvjesto godina ili udio veći od sto posto.
3. Za svaku izvedenu varijablu zapisujemo iz čega je nastala i po kojem pravilu.
4. Za svako isključeno opažanje zapisujemo razlog prije nego što smo vidjeli kako to isključenje mijenja rezultat.

Vraćamo se na istih 300 zapisa iz opisnih poglavlja, bez novog izvlačenja. Generator određuje četiri dobne skupine i unutar njih zasebno stvara minute i povjerenje. Pri povratku na istraživačko pitanje zadržavamo izvorne zapise. Provjera pripreme zato uspoređuje izvorne vrijednosti s dopuštenim rasponima i potvrđuje da je u analizu ušlo svih 300 osoba.

Simulirana anketa ima 300 ispitanika, bez nedostajućih vrijednosti. Svi su odgovori u dopuštenim rasponima. Dnevno vrijeme ide od 3 do 248 minuta. Provjeru bilježimo i kada ne pronađemo problem.

Za nedostajuće odgovore unaprijed određujemo postupak. Brisanje svih nepotpunih redaka smanjuje uzorak i može promijeniti njegov sastav. Ne pretpostavljamo da odgovori nedostaju slučajno. Ako analiziramo samo potpune retke, prikazujemo koliko vrijednosti nedostaje po varijabli i navodimo tu odluku. Takva analiza može ostati pristrana i opisivati uži skup, što treba prenijeti u zaključak.

[1] https://lusiki.github.io/statistika-knjiga/dodaci/f-ai-protokol.html#sec-protokol-asistenta

Opis i graf prije modela

Nakon provjere zapisa najprije opisujemo sastav simuliranog uzorka te raspodjele vremena i povjerenja. Taj će nam pregled pomoći protumačiti odnos vremena i povjerenja prije nego što ga sažmemo jednim koeficijentom. Primjerice, za minute zbrajamo svih 300 odgovora i dobivamo 15 019. Dijeljenjem s 300 nastaje sredina, dok medijan dobivamo iz 150. i 151. vrijednosti nakon poredanja. U tablici po dobi ponavljamo isti postupak unutar svake skupine, koristeći njezin broj osoba kao nazivnik.

Prosječna dob iznosi 34,5 godina. Dnevno vrijeme ima prosjek 50,1 minuta i medijan 39. Prosječno povjerenje iznosi 5,45 boda, uz standardnu devijaciju 1,79.

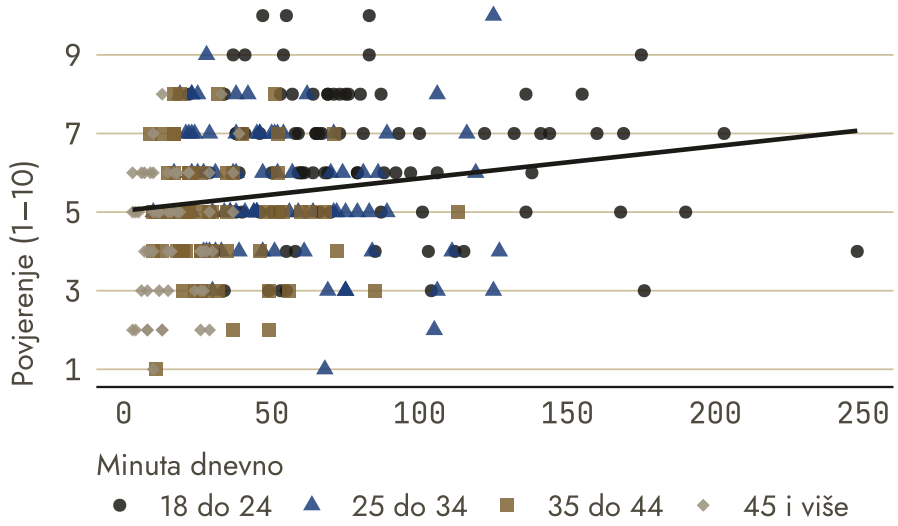
Prosjek vremena veći je od medijana jer je raspodjela desno asimetrična. Manjina visokih vrijednosti podiže prosjek. Zato prikazujemo obje mjere.

Tablica 3: Opis simuliranog uzorka po dobnim skupinama, s prosječnim dnevnim vremenom i prosječnim povjerenjem. Izrada autora.

Dobna skupina	Ispitanika	Minuta dnevno	Povjerenje (bodovi)
18 do 24	90	81,5	6,38
25 do 34	84	54,4	5,56
35 do 44	66	32,4	4,89
45 i više	60	16,2	4,50

Po dobnim skupinama vidi se dodatni obrazac. Dnevno vrijeme pada s 81,5 minuta među najmlađima na 16,2 među najstarijima. Povjerenje pada s 6,38 na 4,50 boda. Obje mjere povezane su s dobi.

Grafom provjeravamo oblik odnosa, utjecaj krajnjih opažanja i promjene raspršenosti. Time započinjemo dijagnostiku, ali još nismo provjerili sve pretpostavke višestrukog modela.

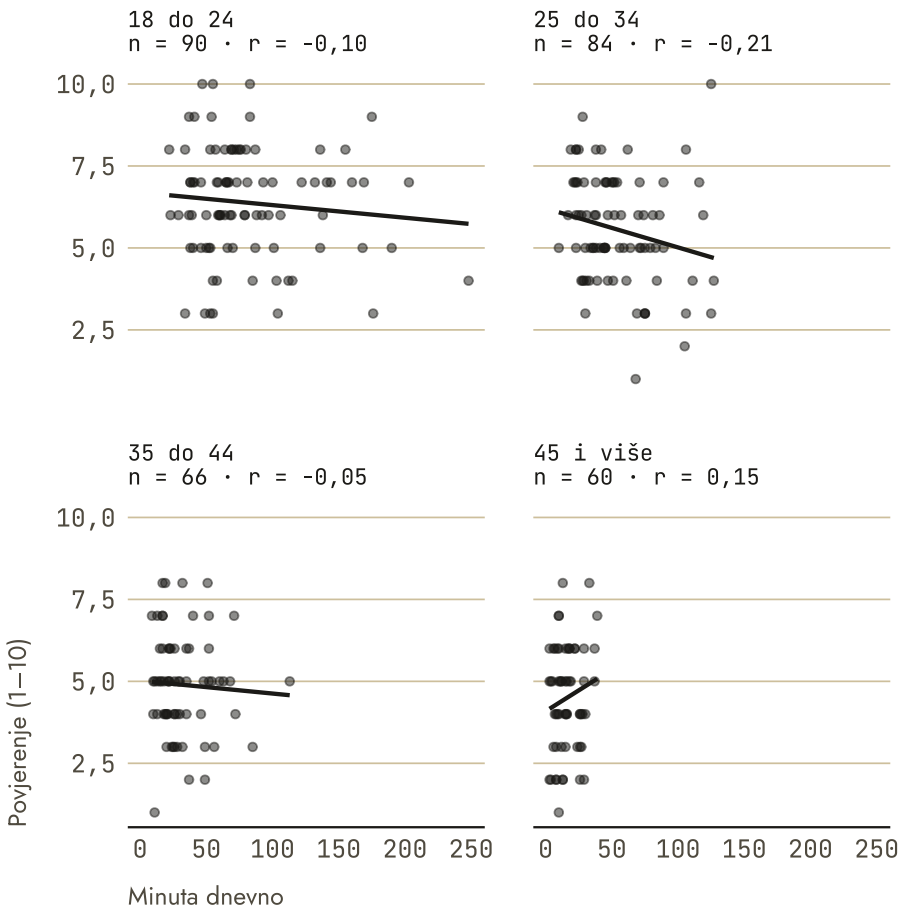


Slika 1: Dnevno vrijeme na mrežama i izjavljeno povjerenje u simuliranom uzorku, s pravcem najmanjih kvadrata i točkama razdvojenima po dobnoj skupini. Izrada autora. Podaci i postupak opisani su u tekstu.

Zbirni pravac blago raste. Dobne oznake pokazuju da su mlađi uglavnom desno i više, a stariji lijevo i niže. Unutar dobnih skupina nema istog uzlaznog obrasca. Zasebne korelacije kreću se od −0,21 do 0,15.

Sličan smo odnos unutar skupina i ukupno vidjeli u poglavljima o povezanosti i regresiji. Ovdje ga graf otkriva prije modela i pomaže nam odabrati smislenu usporedbu.

Zasebni paneli omogućuju da zbirni obrazac usporedimo s odnosom unutar svake skupine. Koristimo iste osi, a naslov svakog panela daje broj osoba i Pearsonovu korelaciju. Pravac je procijenjen zasebno u svakom panelu.



Slika 2: Odnos vremena i povjerenja unutar dobnih skupina istoga simuliranog uzorka. Izrada autora.

Model i promjena procjene

Na grafu smo vidjeli zašto zbirni odnos vremena i povjerenja treba usporediti s odnosom pri jednakoj dobi. U primarnom modelu zato zadržavamo ranije postavljeno pitanje, s povjerenjem kao ishodom i dnevnim vremenom kao glavnim prediktorom. Dob uključujemo prema unaprijed obrazloženoj pretpostavci da je povezana s objema mjerama. Tako ćemo moći odvojeno pročitati zbirnu i dobno prilagođenu procjenu.

```
primarni <- lm(povjerenje ~ minute_dnevno + dob,
               data = anketa_mreze)
bez_prilagodbe <- lm(povjerenje ~ minute_dnevno,
                     data = anketa_mreze)

rbind(
  primarni = c(
    Procjena = coef(primarni)[["minute_dnevno"]],
    confint(primarni)["minute_dnevno", ]
  ),
  osjetljivost = c(
    Procjena = coef(bez_prilagodbe)[["minute_dnevno"]],
    confint(bez_prilagodbe)["minute_dnevno", ]
  )
)
```

```
                Procjena        2,5 %      97,5 %
primarni     -0,0002774402 -0,006211828 0,005656947
osjetljivost  0,0082048533  0,003088870 0,013320837
```

Prije sažetka nagiba možemo pogledati predviđanja za tri postojeća retka simulirane ankete. Predviđanja obaju modela dobivamo uvrštavanjem istih minuta, a u primarnom modelu i dobi. Predviđanja su rezultat modela pridružen postojećim odgovorima ispitanika.

Tablica 4: Ista opažanja u zbirnom i dobno prilagođenom modelu. Izrada autora.

Osoba	Dob	Minute	Opaženo	Zbirni	Uz dob
I001	22	69	6	5,60	6,00
I002	22	53	8	5,47	6,00
I003	24	83	10	5,72	5,90

Primarni model trećoj osobi pridružuje 5,90 umjesto opaženih 10 bodova. Opaženo minus predviđeno daje rezidual 4,10. Taj rezidual i preostalih 299 reziduala ulaze u račun prilagodbe i nesigurnosti koeficijenata. Interval za nagib po minuti nastaje dodavanjem i oduzimanjem standardne pogreške 0,0030 pomnožene kritičnom t-vrijednošću uz 297 stupnjeva slobode. Množitelj t čitamo kao u poglavlju o dvjema grupama. On određuje granicu središnjih 95 % referentne raspodjele pri tim stupnjevima slobode. Taj račun pretpostavlja prikladan linearni model i zajedničku rezidualnu varijancu. Njihovu primjerenost procjenjujemo zasebnim dijagnostičkim pregledom.

Prvi model uključuje dob, a drugi pokazuje kako se procjena mijenja kada je izostavimo. Poziv `rbind` spaja njihove rezultate u dva retka, dok `confint` daje granice intervala. Taj ispis izražava razliku po jednoj dodatnoj minuti. U sljedećoj tablici i procjenu i obje granice množimo s trideset kako bismo ih čitali za pola sata više dnevnog vremena, bez promjene samih modela.

Tablica 5: Procijenjena razlika u povjerenju uz trideset minuta više dnevnog vremena, u dvama modelima, s 95 % intervalima pouzdanosti. Izrada autora.

Model	Razlika u bodovima	95 % interval	R^2, udio varijance
Primarni model s linearnom prilagodbom za dob	−0,01	−0,19 do 0,17	10,8 %
Provjera osjetljivosti samo s vremenom	0,25	0,09 do 0,40	3,2 %

Primarna analiza uspoređuje modelom predviđeno povjerenje pri jednakoj dobi, uz linearnu i aditivnu prilagodbu. Kao što smo vidjeli u regresiji, takav zapis ne uparuje ljude niti nasumično dodjeljuje uporabu mreža. Uz trideset minuta više dnevnog vremena procijenjena razlika iznosi −0,01 boda, uz 95 % interval od −0,19 do 0,17. Interval dopušta male razlike u oba smjera.

Za provjeru osjetljivosti izostavljamo dob i vraćamo se zbirnoj usporedbi, u kojoj se s vremenom na mrežama može mijenjati i dobni sastav. Procjena tada iznosi 0,25 boda, uz 95 % interval od 0,09 do 0,40, bez nule. Isti podaci o 300 osoba daju drukčiju usporedbu. Alternativa pokazuje važnost uključivanja dobi za tumačenje. Zadržavamo je uz primarnu analizu kako bi čitatelj mogao usporediti oba pitanja i rezultata.

Ovdje opisujemo povezanost uz prilagodbu za dob. Predviđanje novih osoba tražilo bi zasebnu provjeru na odvojenim podacima, a uzročno tumačenje nacrt koji razdvaja moguća objašnjenja. Naša analiza ostaje u okviru presječne povezanosti.

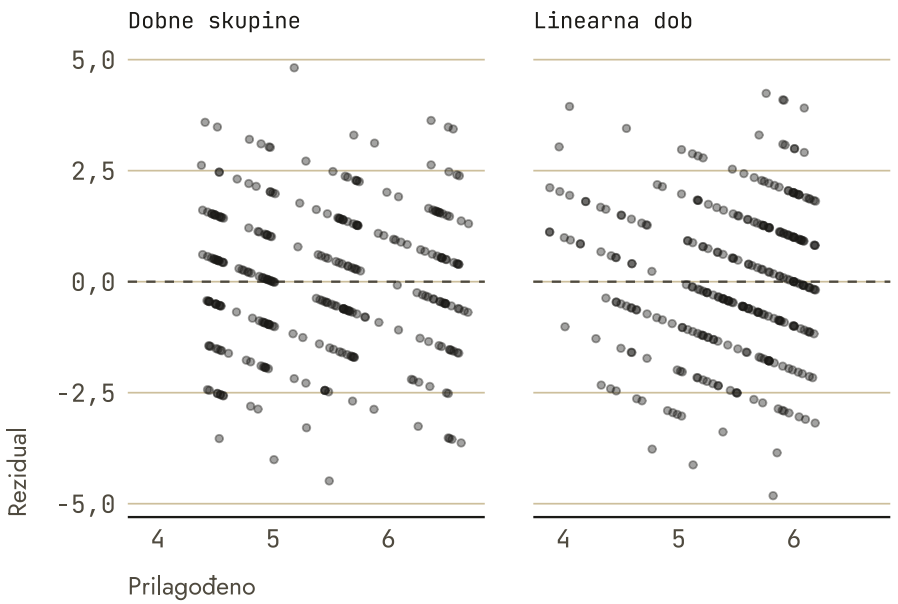
Izostavljanjem dobi ispitali smo njezinu važnost. Sada dodajemo ciljanu provjeru prikladnosti linearnoga dobnog člana. Dob zamjenjujemo dobnim skupinama koje koristi generator. Svaka skupina dobiva vlastiti odsječak, a nagib uz minute ostaje zajednički. Ta provjera koristi poznato pravilo nastavnih podataka. U stvarnom istraživanju takvo pravilo ne bismo znali unaprijed.

Sljedeća tablica zadržava iste osobe i pokazuje promjenu procjene pri zamjeni linearnoga dobnog člana skupinama. Oba intervala i dalje počivaju na modelu sa zajedničkom rezidualnom varijancom.

Tablica 6: Provjera oblika prilagodbe za dob. Razlike i intervali odnose se na 30 minuta više dnevno. Izrada autora.

Prilagodba	Procjena	Donja	Gornja
Linearna dob	−0,008	−0,186	0,170
Dobne skupine	−0,160	−0,343	0,024

Reziduale obaju modela također uspoređujemo. Sustavno zakrivljenje ili promjena raspršenosti upozorili bi da sama zamjena dobnog člana nije dovoljna. Prikaz je dijagnostička provjera, a ne dokaz da su pretpostavke ispunjene.



Slika 3: Reziduali dviju prilagodbi za dob. Izrada autora.

Koeficijent R^2 iz ranije tablice izračunat je kao $1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$. Brojnik je zbroj kvadrata reziduala, a nazivnik

zbroj kvadrata odstupanja od ukupne sredine. On opisuje prilagodbu ovim podacima, ne uzročni udio ni uspjeh predviđanja novih osoba. Ovaj završni primjer namjerno prati jedan regresijski model i njegovu dokumentaciju. Ne pokušava primijeniti svaki postupak iz knjige.

U dijagramima ostataka kose usporedne trake nastaju zato što povjerenje ima cjelobrojne vrijednosti od jedan do deset. Za svaku od njih ostatak je ta vrijednost umanjena za predviđanje. Dobne skupine bolje prate poznato pravilo nastanka ovih simuliranih podataka. Razlika dviju procjena za minute pokazuje koliko linearna prilagodba dobi može ostaviti dio skupnog obrasca u tom koeficijentu. Ni jedan model time nije postao uzročna analiza.

Zaključak koji uzorak podnosi

Zaključak povezuje oba računa. U zbirnom modelu trideset minuta više dnevne uporabe povezano je s 0,25 boda višim povjerenjem, uz 95 % interval od 0,09 do 0,40. Uz linearnu prilagodbu za dob razlika iznosi −0,01 boda, uz interval od −0,19 do 0,17.

Istu linearnu procjenu drugdje prikazujemo na tri decimale, −0,008. Prikazi −0,01 i −0,008 potječu iz istog računa, uz različit broj decimala pri zaokruživanju.

Nesigurnost navodimo uz brojku. Prilagođeni interval isključuje razlike veličine zbirne procjene u oba smjera, ali dopušta male pozitivne i negativne vrijednosti. Uzorak od 300 osoba zato ne dokazuje potpunu odsutnost povezanosti.

Ograničenje treba napisati jasno. Umjesto niza riječi „možda” i „eventualno”, navodimo procjenu, interval i pretpostavku o kojoj zaključak ovisi.

Procjena blizu nule može biti koristan rezultat. Uzak interval može isključiti razlike veće od unaprijed opravdanoga sadržajnog praga, dok širok može pokazati da preciznost za tu odluku

nije dovoljna. U ovome primjeru takav prag nismo zadali, pa blizinu nule ne proglašavamo sadržajnom nevažnošću. Ta dva ishoda razlikujemo prema veličinama koje interval dopušta.

U simulaciji znamo da generator nema izravnu vezu vremena i povjerenja. Obje mjere ovise o dobi. Prilagođeni interval uskladiv je s tim pravilom. Zbirni odnos postoji, ali njegovo tumačenje kao izravne veze zanemarilo bi dob.

Znanje o generatoru dobili smo izvan samog računa. Model je dao raspon uskladivih vrijednosti, bez dokaza da je koeficijent točno nula. U stvarnom radu zato uz rezultat trebamo nacrt, teoriju i prethodna mjerenja.

Izvještaj počinjemo pitanjem i njegovom vrstom. Opisujemo uzorak, prikupljanje i mjere, zatim podatke i analizu s procjenama i intervalima. Završavamo tumačenjem i granicama prijenosa. Tablice i grafovi trebaju jasne oznake, jedinice, izvore i broj opažanja na koji se odnose. Skraćenu os izričito označujemo.

Uz izvještaj predajemo dopuštene podatke, njihovo podrijetlo, ponovljiv postupak za tablice i grafove te zapis o asistentu. Reprodukciju provjeravamo u čistoj R sesiji ili novom projektu alata bez koda, po mogućnosti na drugom računalu.

Za petnaestominutno izlaganje izdvajamo pitanje, način prikupljanja, jedan važan graf, procjenu s intervalom i glavno ograničenje. Detaljne koeficijente i čišćenje ostavljamo u izvještaju. Cilj je da publika razumije rezultat i bez čitanja cijelog rada.

Na kraju zapisujemo što bismo promijenili u sljedećem istraživanju. Korisno je imenovati konkretnu odluku, primjerice prebrzo odabranu mjeru ili prekasno planiran uzorak, i objasniti kako bismo je drukčije riješili.

Provjera na stvarnim podacima

Simulacija nam je dopustila da zaključak usporedimo s poznatim pravilom nastanka podataka. Kako ćemo provjeriti vlastiti

postupak kada to pravilo nemamo? Vraćamo se stvarnim rečenicama iz hrvatskog dijela skupa podataka ParlaSent 1.0, dostupnog u repozitoriju Instituta Jožef Stefan (Mochtak i sur. 2023., 2024.). Svaki redak predstavlja jednu parlamentarnu rečenicu s oznakom tona, pa trebamo drukčije pitanje od prethodnog odnosa vremena i povjerenja. Provjerit ćemo kako su rečenice odabrane i označene, opisati raspodjelu oznaka i odrediti što iz njih možemo zaključiti.

Tablica 7: Podaci i njihovo podrijetlo u primjeru s ParlaSentom. Izrada autora prema Mochtak i sur. (2023.) i Mochtak i sur. (2024.).

Polje putovnice	Opis podataka
izvor i verzija	ParlaSent 1.0, hrvatski izvadak pripremljen 18. kolovoza 2026.
prava	CC BY-SA 4.0. Atribucija, oznaka promjena i obveza dijeljenja pod istim uvjetima putuju uz izvedenu datoteku
jedinica	jedna odabrana i označena parlamentarna rečenica
obuhvat	2 698 redaka iz 2 499 dokumenata. Nije uzorak svih parlamentarnih rečenica, govornika ni stavova
put oznake	1 362 redaka ima dva koderska glasa i usklađenje, a 1 336 jedan uvježbani koderski zapis
razdvajanje	1 090 redaka za učenje, 272 za provjeru i 1 336 za ispitivanje. Preklopljeni dokumenti uklonjeni su s ulaznih podataka prije razdvajanja

Tablica 7: (nastavak)

Polje putovnice	Opis podataka
nedostajanje	nedostupni drugi koder, usklađenje i izvorna podjela u ispitnim redcima nose izričit marker, a ne prazninu ili nulu
provjerljivi put	data/parlament_oznake.csv, data/parlament_tekst/PUTOVNICA.md i data/parlament-oznake.LICENCA.md vode do izvora. Ista CSV datoteka služi radu u R-u i bez koda te pripremi tiskanog izdanja

Dnevnik bilježi izbor redaka `HR`, uklanjanje 25 redaka iz 20 dokumenata koji se pojavljuju i u ispitnom dijelu te ponovljivu podjelu preostalih dokumenata. Zadržano je svih 1 336 izvornih ispitnih redaka. Paket ne spaja podatke o govornicima, strankama, mandatima ili datumima. Sadrži rečenice, izvorne ključeve, oznake i njihov nastanak, pa zaključke ograničavamo na te podatke.

Za ovaj opis ponovno koristimo svih 2 698 pripremljenih rečenica, a ne samo 272 iz usporedbe pragova u prethodnom poglavlju o algoritmima. Svaka oznaka `Negative` povećava brojač negativnih za jedan, neovisno o podjeli za učenje ili provjeru. Podjela prema putu oznake poslije razdvaja ta ista brojanja prema tome je li oznaku dao jedan koder ili dva kodera uz usklađenje.

Brojimo 1 212 negativnih, 979 neutralnih i 507 pozitivnih oznaka. Nazivnik je svih 2 698 redaka paketa. Ti brojevi opisuju oznake odabranih rečenica. Za zaključak o raspoloženju parlamenta trebali bismo opravdati drukčiji cilj mjerenja i odgovarajući odabir građe.

Kao provjeru osjetljivosti razdvajamo načine označavanja. U dijelu s dva kodera i usklađenjem negativno je 652 od 1 362 redaka, odnosno 47,9 % redaka. U dijelu s jednim uvježbanim koderom negativno je 560 od 1 336 redaka, odnosno 41,9 %. Zbirni udio tako objedinjuje dva mjerna postupka. Njihova usporedba istodobno uključuje razlike u broju kodera, dokumentima i odabiru, pa doprinos samog broja kodera ostaje nerazdvojen.

Ovo je opis konačnog odabranog paketa. Nismo zadali model slučajnog odabira neovisnih rečenica, pa uz te postotke ne prikazujemo binomne intervale niti procijenjenu uzoračku nesigurnost za širu populaciju.

Možemo provjeriti podrijetlo, nastanak oznaka i prelazak dokumenata među skupovima. Ne možemo ocijeniti pogreške klasifikatora, njegove podskupine, promjene izvedbe ni posljedice praga jer nemamo takav model, predviđanja i odluke iz primjene. Za budući red ljudskog pregleda trebali bismo unaprijed odrediti prag, posljedice pogreške, nadzor novih i izostavljenih rečenica te prigovor i ispravak. Asistent može provjeriti brojnosti i dosljednost, ali ne određuje opravdani cilj ili prihvatljive posljedice.

Zaključak se ograničava na oznake, njihove nazivnike i dva načina nastanka unutar odabranog paketa. Ne prenosi se na sav parlamentarni govor i ne provjerava novi klasifikator, uzrok tona ni institucionalnu odluku. Uz putovnicu, dnevnik, alternativnu analizu i zapis o asistentu druga osoba može pregledati dokaze i ponoviti račun u R-u ili alatu bez koda.

U ovom empirijskom primjeru asistentu su dane samo javne zbirne brojnosti i polja putovnice. Pojedinačne rečenice nisu poslane. Zapis uporabe prilažemo putovnici i dnevniku. Podatkovnom trakom označujemo koje okruženje smije primiti podatke s obzirom na njihovu osjetljivost. Prvo procjenjujemo podatke, a zatim biramo alat koji ih smije primiti.

Tablica 8: Objava uporabe asistenta u primjeru s ParlaSentom. Izrada autora.

Polje objave	Zapis za primjer s ParlaSentom
sustav, inačica i datum	Codex, GPT-5, 19. kolovoza 2026.
delegirani zadatak	provjera zbrojeva po oznaci i putu označivanja te nacrt granice tvrdnje
podatkovna traka	javni alat. Samo objavljeni agregati iz paketa s provjerenim pravima uporabe
zadržani rezultat	aritmetička provjera zbrojeva. Granicu tvrdnje odgovorna je osoba ponovno napisala
neovisna provjera	lokalni račun iz data/parlament_oznake.csv, usporedba s putovnicom i otvaranje obaju citiranih izvora
što je ostalo neprovjereno	valjanost oznake, uzvodna selekcija rečenica i prijenos izvan pripremljenoga paketa
odgovorna osoba i mjesto objave	autor koji potpisuje paket. Zapis uz putovnicu, dnevnik i izvještaj

U oba primjera završni izvještaj treba jasno pokazati što je procijenjeno i pod kojim uvjetima.

● STATISTIKA U DIVLJINI

Izvještaj s procjenom u prvom planu. Na kraju našega nastavnog istraživanja trebamo napisati sažetak koji čitatelj može razumjeti bez otvaranja računalnog ispisa. Mogli bismo ga svesti na rečenicu da je zbirna veza bila

značajna, a nakon prilagodbe za dob više nije. Time bismo izostavili ono što smo istraživanjem doznali o veličini razlike i usporedbi koju pojedini model provodi.

Pristup poznat kao nova statistika stavlja procjenu učinka, njezin interval i povezivanje dokaza u središte izvještavanja (Cumming 2014.). Primijenimo taj prijedlog na naš simulirani skup, bez tvrdnje da smo otkrili ponašanje stvarnih korisnika. Za dodatnih pola sata dnevnog korištenja mreža zbirni model procjenjuje razliku od 0,25 boda povjerenja. Nakon prilagodbe za dob procjena je −0,008 boda, uz 95 % interval od −0,186 do 0,170 boda.

Drugi račun uspoređuje osobe jednake dobi pod pretpostavkama modela. Njegov interval obuhvaća nulu, ali je dovoljno uzak da isključi razlike veličine zbirne procjene u oba smjera. Zaključak zato temeljimo na veličinama koje interval dopušta. Male razlike u oba smjera ostaju spojive s podacima i modelom.

To mijenja i odgovor koji bismo dali zamišljenom naručitelju zainteresiranom za povjerenje. Iz ove analize ne slijedi da bi skraćivanje vremena na mrežama promijenilo povjerenje. Za takvu odluku nedostaje uzročno opravdanje, a naš skup povrh toga služi nastavi. Usmjerenost na procjenu pomaže prikazati ono što smo procijenili. Interval ne može nadoknaditi ograničenje podataka i nacrta.

Završni izvještaj zato povezuje veličinu u razumljivim jedinicama, nesigurnost, usporedbu i granicu primjene. Čitatelj može vidjeti zašto smo promijenili početno tumačenje, a sljedeći istraživač što bi još trebalo provjeriti. To je više informacija od završne oznake značajnosti, čak i kada konačna odluka ostaje otvorena.

Rad s asistentom

Račun i neke provjere možemo delegirati, ali za predani rad odgovara autor. Koristimo sljedećih šest koraka, povezanih s pravilima u Dodatku F.

1. Specificiramo pitanje, vrstu tvrdnje, jedinicu, granicu podataka, dopušten zaključak i provjerljivi izlaz.
2. Delegiramo samo omeđeni račun, nacrt koda ili kritiku za koju su ulaz i očekivani oblik izlaza poznati.
3. Reproduciramo svaki račun u čistoj lokalnoj sesiji ili novom projektu alata bez koda, od učitavanja dopuštenih podataka do posljednje tablice.
4. Osporavamo rezultat jednom obranjivom alternativom i uspoređujemo procjenu, interval, sadržajni zaključak i dosegnutu populaciju.
5. Dokumentiramo što je asistent vratio, što je provjereno, kako je provjereno i što je ostalo neprovjereno.
6. Objavljujemo sustav, inačicu, datum, zadatak, podatkovnu traku, zadržani rezultat i odgovornu osobu, sve prema pravilima kolegija ili časopisa.

Zapis uporabe treba pokazati što je alat napravio i kako smo to provjerili. Potvrda u nastavku čitljiva je i bez koda, neovisno o korištenom alatu.

Tablica 9: Čitljiva potvrda provjere rada s asistentom. Izrada autora.

Polje potvrde	Zapis za ovaj projekt
što je zatraženo	prijedlog analize i nacrt izvještaja za presječnu povezanost vremena i povjerenja
što je vraćeno	kod za dva modela i nacrt zaključka
što je provjereno	broj redaka, procjene, intervali, podrijetlo podataka i granica tvrdnje
kako je provjereno	čista lokalna sesija, usporedba s tablicom i otvaranje svakoga navedenog izvora
uloga asistenta	pomaže u računu i nacrtu uz ljudsku provjeru. Predmet je istraživanja samo kada proučavamo sam sustav
što je ostalo neprovjereno	valjanost mjere izvan generatora i prijenos na ljude izvan simuliranoga skupa
odgovorna osoba	osoba koja potpisuje pitanje, provjeru i objavljenu tvrdnju

Dodajemo ime sustava, dostupnu inačicu ili napomenu da nije poznata, datum, podatkovnu traku i dio rezultata koji smo zadržali.

Alat biramo prema osjetljivosti podataka i dopuštenim uvjetima uporabe. Kao primjer služi politika kolegija uz ovaj udžbenik, inačica 1.0 od 4. kolovoza 2026. Puni uvjeti nalaze se u Dodatku F[2].

[2] https://lusiki.github.io/statistika-knjiga/dodaci/f-ai-protokol.html#sec-politika-kolegija

Prije slanja podataka asistentu zastanemo i provjerimo što smijemo podijeliti. U ovom primjeru javnom alatu možemo poslati dopuštene javne, simulirane ili zbirne podatke. Ako radimo s dopuštenim podacima u kojima su imena zamijenjena šiframa, trebamo alat s ugovorenom zaštitom podataka. Ograničeni podaci ostaju u lokalnom okruženju koje je odobrila ustanova. Odabir zato vežemo uz podatke koje imamo i pravila svojega zadatka i ustanove.

Ta datirana politika nije propis Hrvatskoga katoličkog sveučilišta ni pravno tumačenje. Uklanjanje imena samo po sebi ne potvrđuje da osobu nije moguće prepoznati.

Prije predaje provjerimo kako asistent povezuje rezultate s pitanjem istraživanja. Završni odlomak treba ostati provjerljiv prema istom paketu podataka, odluka i izračuna.

NAĐITE GREŠKU

Prije predaje završnog izvještaja usporedite asistentov sažetak s rezultatima dvaju modela na istih 300 simuliranih osoba. Oba procjenjuju razliku u povjerenju povezanu s trideset minuta više dnevnog vremena na društvenim mrežama. Prilagodba za dob bila je unaprijed određena kao primarna analiza.

Tablica 10: Rezultati za završnu provjeru izvještaja. Izrada autora.

Analiza	Procjena u bodovima	95 % interval
Bez dobi	0,25	od 0,09 do 0,40
Uz linearnu prilagodbu za dob	−0,01	od −0,19 do 0,17

Asistent je vratio sljedeći sažetak.

> U simuliranom uzorku, nakon linearne prilagodbe za dob, trideset minuta više dnevnog vremena povezano je s procijenjenom razlikom od −0,01 boda u povjerenju. Zaključak ovisi o linearnoj, aditivnoj prilagodbi za dob. Drukčiji oblik dobnoga odnosa mogao bi promijeniti procjenu. 95 % interval od −0,19 do 0,17 obuhvaća nulu, pa u ovom skupu nema povezanosti vremena i povjerenja.

Pripremite konačan odlomak koji možete potpisati. Zadržite točne brojke i ograničenje oblika modela, objasnite što usporedba dviju analiza pokazuje i zamijenite nepoduprtu tvrdnju. Dodajte jednu provjeru koju biste proveli prije predaje, uz objašnjenje zašto je tražite.

Završna revizija paketa dokaza

U završnoj provjeri razlikujemo svih šest vrsta tvrdnji. Za svaku navodimo postoji li opravdanje ili ostaje izvan dosega projekta.

Tablica 11: Šest dimenzija tvrdnje u završnom paketu dokaza. Izrada autora.

Dimenzija	Što završni paket podupire ili omeđuje
opis	opisuje 300 simuliranih osoba te brojnosti i put oznake 2 698 odabranih rečenica
povezanost	procjenjuje prilagođenu povezanost vremena i povjerenja. Razliku udjela po putu oznake ne tumači kao njegov učinak
generalizacija	ne prelazi sa simuliranoga skupa na stvarne korisnike ni s odabranoga korpusa na sav parlamentarni govor
predviđanje	nema zaključanoga prediktivnog pravila ni rezultata na doista novim opažanjima
uzročnost	presječna anketa i odabrani korpus ne podupiru intervencijsku tvrdnju
odluka	ne opravdava postavljeni sustav. Buduća bi odluka morala imenovati prag, teret pogrešaka, nadzor i put prigovora

Zatim odgovaramo na šest revizijskih pitanja.

Tablica 12: Šest revizijskih pitanja prije predaje. Izrada autora.

Revizijsko pitanje	Gdje je odgovor
Koja je jedinica?	osoba u glavnoj studiji. Odabrana rečenica u ParlaSentu
Tko ili što nedostaje?	stvarni ljudi i simulirane jedinice izvan zadanoga skupa te rečenice izvan odabira i bez hrvatske oznake
Koji je cilj tvrdnje?	dobno prilagođena razlika u povjerenju povezana s 30 minuta više dnevnog vremena. Opis oznaka i puta mjerenja unutar paketa
Gdje je neizvjesnost?	interval kvantificira uzoračku varijaciju pod pretpostavljenim linearnim modelom. Oblik modela, mjerenje, selekcija, kodiranje i odluka ostaju zasebni
Koje drugo objašnjenje ostaje?	izostavljanje dobi. Odvojeno čitanje dvaju putova nastanka oznake
Tko snosi posljedice?	na ljude o kojima bi se neopravdano generaliziralo te na autore rečenica koje bi se slale u ljudski pregled

U putovnici i analitičkoj tablici navodimo jedinicu, način odabira i izostavljene podatke. Uz udjele navodimo nazivnike, a uz procjene nesigurnost i njezina ograničenja. Posljedice pogreške povezujemo s odlukom i prigovorom, podrijetlo s izvorom i dnevnikom, a tumačenje s izvještajem i provjerom asistentova nacrta.

Dokumentacija omogućuje drugoj osobi da prati istraživanje od početnog pitanja i pripreme podataka do analize i izvještaja. Naša studija ne uključuje sustav koji donosi odluke u praksi. Kada bismo uveli sustav za odabir rečenica za pregled, trebali bismo unaprijed odrediti kako ćemo pratiti nove tekstove, pogreške i prigovore.

Sljedeća tablica razlikuje četiri aktivnosti i pitanja na koja odgovaraju.

Tablica 13: Granice četiriju aktivnosti u završnom paketu. Izrada autora.

Aktivnost	Pitanje koje postavlja ovaj paket
statistika	što dizajn, procjena i interval opravdavaju vjerovati?
znanost o podacima	može li se od izvora i verzije ponoviti svaka transformacija i tablica?
strojno učenje	postoji li zaključan obrazac provjeren na doista novim opažanjima? Ovdje ne postoji
sustav umjetne inteligencije	kako bismo pratili odluke, pogreške i posljedice nakon primjene? Ovdje takav sustav još ne postoji

U završnom projektu koristimo četiri vještine razvijene kroz knjigu. Provjeravamo tuđu tvrdnju, opisujemo i prikazujemo vlastite podatke, tumačimo i reproduciramo procjene te surađujemo s asistentom uz provjeru. Ne trebamo poznavati svaki statistički postupak. Trebamo znati što rezultat znači, koje su mu granice i kako ga druga osoba može provjeriti.

Prije dijeljenja paketa s parlamentarnim rečenicama primjenjujemo pravila zaštite osoba iz poglavlja o mjerenju i dizajnu. Javna dostupnost zapisa ne uklanja mogućnost identifikacije ni

štetu od povezivanja s drugim podacima. Provjeravamo svrhu dijeljenja, nužnost pojedinih obilježja i prikladnost alata za osjetljivost podataka. Pravila kolegija dodatni su uvjet u nastavi. Odgovornost prema osobama o kojima podaci govore ostaje na nama.

Sažetak

Prvo istraživanje završavamo izvještajem i dokumentima koji drugoj osobi omogućuju da od našeg pitanja ponovno dođe do rezultata. Pitanje i nacrt određuju doseg zaključka, pa prije modela provjeravamo jesmo li analitičku tablicu, opis i grafove oblikovali tako da možemo odgovoriti na postavljeno pitanje. Primarnu analizu zatim uspoređujemo s opravdanom alternativom po procjeni, intervalu i značenju, kao što smo učinili pri uključivanju i izostavljanju dobi. Simulacija nam je omogućila provjeru poznatog pravila, a stvarni podaci pokazali su kako izbor rečenica i nastanak oznaka ograničavaju njihovo tumačenje. Asistent može ubrzati dio posla, dok provjera, objava uporabe i odgovornost ostaju na autoru. Uz svaku tvrdnju zato ostavljamo podrijetlo, nesigurnost i granicu dokaza kako bi čitatelj mogao prosuditi i rezultat i put kojim smo do njega došli.

Pojmovi

paket dokaza (*evidence package*), putovnica skupa podataka (*dataset passport*), dnevnik transformacija (*transformation log*), trag odluka (*decision trail*), potvrđujuća analiza (*confirmatory analysis*), istraživačka analiza (*exploratory analysis*), provjera osjetljivosti (*sensitivity analysis*), granica tvrdnje (*claim boundary*), reproducibilan tijek rada (*reproducible workflow*), pošten

izvještaj (*honest report*), minimizacija podataka (*data minimization*), objava uporabe asistenta (*AI-use disclosure*)

Zadaci

Konceptualni

Vratite se temama zavaravajućih brojki, uzorkovanja, krize i obnove te statistike u doba algoritama. Za primjer s hrvatskim rečenicama iz ParlaSenta odgovorite na šest revizijskih pitanja. U odgovorima upotrijebite barem po jedan dokaz o zavaravajućem nazivniku, selekciji, analitičkoj fleksibilnosti i teretu odluke. Pokažite kako nazivnik, izbor rečenica, analitičke odluke i posljedice pogreške ograničavaju zaključak. Zatim napišite jednu tvrdnju koju paket podupire i jednu koju ne podupire. Sam regresijski model iz ovoga poglavlja nije dovoljan za te odgovore.

Računski

Iz tablice opisa po dobnim skupinama izračunajte razliku u prosječnom dnevnom vremenu i razliku u prosječnom povjerenju između najmlađe i najstarije skupine. Zatim iz neprilagođenoga nagiba u provjeri osjetljivosti od 0,0082 boda po minuti izračunajte koliko bi razlike u povjerenju taj model pripisao dobivenoj razlici u vremenu. Usporedite taj broj s opaženom razlikom u povjerenju među skupinama i predajte oba rezultata s jednom rečenicom o tome zašto račun ne zamjenjuje primarnu analizu uz dob.

Kritički

Pronađite studentski ili novinski izvještaj o anketi koji navodi postotke bez broja ispitanika ili bez opisa načina prikupljanja.

Prosudite ga prema redoslijedu izvještavanja koji procjenu i njezin raspon stavlja u prvi plan (Cumming 2014.). Predajte popis podataka koji nedostaju i jednu rečenicu o tome koji je zaključak zbog toga neprovjerljiv.

Revizija modela

Ocijenite asistentov kratki sažetak iz okvira o pogrešci prema standardu poštene rečenice. Prepišite točne brojke, imenujte jedini pogrešan prijelaz, napišite ispravljeni zaključak te ispunite sedam polja potvrde provjere. Na kraju navedite što biste delegirali u sljedećem krugu, a koju odluku ne biste.

putokaz, pojmovi i oznake

Broj poglavlja u putokazu odgovara oznaci na njegovoj početnoj stranici. U rječniku su nazivi iz završnih popisa pojmova. Uz svaki pojam navodimo poglavlje i stranicu završnog popisa, odakle se možemo vratiti njegovu objašnjenju u tekstu.

Putokaz kroz poglavlja

Poglavlje	Tema	Stranica
1	Zašto statistika	7
2	Mjerenje i istraživački dizajn	33
3	Kako brojke zavode	59
4	Sažimanje podataka	83
5	Vizualizacija kao argument	109
6	Povezanost	135
7	Vjerojatnost koliko treba	167
8	Uzorkovanje	189
9	Procjena	215
10	Logika testiranja	243

Poglavlje	Tema	Stranica
11	Veličina učinka i snaga	265
12	Kriza i obnova	285
13	Kategorički podaci	309
14	Uspoređivanje dviju grupa	335
15	Uspoređivanje više grupa	357
16	Regresija — opći okvir	379
17	Statistika u doba algoritama	411
18	Vaše prvo istraživanje	437

Primjeri koji se vraćaju

Oznake poglavlja i stranice vode na početak poglavlja u kojem je primjer razvijen. Tablica pomaže povezati njegove različite uporabe u knjizi.

Primjer	Gdje ga tražiti	Poglavlje i stranica
Berkeley	zbirni i odjelni podaci; povratak uz kategoričke podatke	1, str. 7; 13, str. 309
Anscombeov kvartet	isti brojčani sažeci i različiti oblici odnosa	5, str. 109
DigiKat	sažimanje digitalnih tragova i izbor ljestvice prikaza	4, str. 83; 5, str. 109

Primjer	Gdje ga tražiti	Poglavlje i stranica
ParlaSent	oznake parlamentarnih rečenica i provjera klasifikatora	17, str. 411; 18, str. 437
Registrirana replikacija (RRR)	usporedba laboratorija, procjena učinka i obnova istraživačke prakse	12, str. 285

Hrvatsko-engleski rječnik

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
algoritamska pravednost	algorithmic fairness	17, str. 432
alternativni tekst	alternative text	5, str. 132
analitička fleksibilnost	analytical flexibility	12, str. 303
analitička tablica	analysis table	4, str. 105
analiza varijance	analysis of variance	15, str. 376
Anscombeov kvartet	Anscombe's quartet	5, str. 132
aritmetička sredina	mean	4, str. 105
asimetrija	skewness	4, str. 105
binomna raspodjela	binomial distribution	7, str. 186
bootstrap	bootstrap	9, str. 237
brojnik	numerator	1, str. 30
Cramérovo V	Cramér's V	13, str. 333

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
curenje informacija	information leakage	16, str. 408
distribucija uzorkovanja	sampling distribution	8, str. 212
dnevnik transformacija	transformation log	4, str. 105; 18, str. 466
doprinos ćelije	cell contribution	13, str. 333
efektivna veličina uzorka	effective sample size	8, str. 212
ekološka pogreška	ecological fallacy	6, str. 161
eksterna valjanost	external validity	2, str. 56
eta-kvadrat	eta squared	15, str. 376
Fisherov egzaktni test	Fisher's exact test	13, str. 333
forest plot	forest plot	12, str. 303
F-statistika	F-statistic	15, str. 376
geometrija grafa	geom	5, str. 132
grafička percepcija	graphical perception	5, str. 132
gramatika grafike	grammar of graphics	5, str. 132
granica korpusa	corpus boundary	17, str. 432
granica tvrdnje	claim boundary	18, str. 466
granica tvrdnje o povezanosti	boundary of an association claim	6, str. 161
grupno uzorkovanje	cluster sampling	8, str. 212
hi-kvadrat statistika	chi-squared statistic	13, str. 333
homoskedastičnost	homoskedasticity	14, str. 353
interakcija	interaction	16, str. 408

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
interkvartilni raspon	interquartile range	4, str. 105
interna konzistentnost	internal consistency	2, str. 56
interna valjanost	internal validity	2, str. 56
interval pouzdanosti	confidence interval	9, str. 237
istraživačka analiza	exploratory analysis	12, str. 303; 18, str. 466
izgledi	odds	16, str. 408
jedinica analize	unit of analysis	1, str. 30
jedinica neovisnosti	unit of independence	14, str. 353
jedinica ponovnog uzorkovanja	resampling unit	9, str. 237
jednostavni slučajni uzorak	simple random sample	8, str. 212
jezični model kao sustav predikcije	language model as a prediction system	17, str. 432
kalibracija	calibration	8, str. 212
kalibrirana nesigurnost	calibrated uncertainty	7, str. 186
klasifikacijski prag	classification threshold	17, str. 432
kodirana kategorija teksta	coded text category	6, str. 161
kodiranje kao mjerenje	coding as measurement	2, str. 56
koeficijent determinacije	R-squared	16, str. 408
kolider	collider	2, str. 56

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
konfundirajuća varijabla	confounder	2, str. 56; 6, str. 161
kontingencijska tablica	contingency table	13, str. 333
korekcija za konačnu populaciju	finite-population correction	8, str. 212
korelacija	correlation	6, str. 161
kovarijanca	covariance	6, str. 161
kritična t-vrijednost	critical t value	14, str. 353
kritična vrijednost	critical value	9, str. 237
Kruskal-Wallisov postupak	Kruskal-Wallis test	15, str. 376
kvazieksperiment	quasi-experiment	2, str. 56
linearna regresija	linear regression	16, str. 408
linearnost	linearity	6, str. 161
logaritamska ljestvica	logarithmic scale	5, str. 132
ljestvica	scale	5, str. 132
mala višestruka polja	small multiples	5, str. 132
margina pogreške	margin of error	3, str. 78; 8, str. 212; 9, str. 237
matrica korelacija	correlation matrix	6, str. 161
medijan	median	4, str. 105
medijator	mediator	2, str. 56
metoda najmanjih kvadrata	least squares	16, str. 408
minimizacija podataka	data minimization	18, str. 466
mod	mode	4, str. 105

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
monotona veza	monotonic relationship	6, str. 161
najmanji važan učinak	smallest effect size of interest	11, str. 283
nazivnik	denominator	1, str. 30; 4, str. 105
nedostajuća vrijednost	missing value	4, str. 105
neodaziv	nonresponse	2, str. 56
neodgovor	nonresponse	8, str. 212
neovisne skupine	independent groups	14, str. 353
neovisnost	independence	7, str. 186
nepristranost	unbiasedness	8, str. 212
nesigurnost kodiranja	coding uncertainty	9, str. 237
normalna raspodjela	normal distribution	7, str. 186
nulta hipoteza	null hypothesis	10, str. 262
nulti model	null model	10, str. 262
objava uporabe asistenta	AI-use disclosure	18, str. 466
oblik odnosa	shape of a relationship	6, str. 161
obostrani test	two-sided test	10, str. 262
obuhvat intervala	coverage	9, str. 237
očekivana frekvencija	expected frequency	13, str. 333
ograničenje raspona	range restriction	6, str. 161
okvir kodiranja	coding frame	17, str. 432
okvir uzorkovanja	sampling frame	2, str. 56
omjer izgleda	odds ratio	16, str. 408

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
opažačka studija	observational study	2, str. 56
operacionalizacija	operationalization	2, str. 56
osjetljivost klasifikatora	sensitivity	17, str. 432
osjetljivost na odluku o podacima	data-decision sensitivity	4, str. 105
ovisnost opažanja	dependence among observations	14, str. 353
paket dokaza	evidence package	18, str. 466
parametar	parameter	8, str. 212
Pearsonova korelacija	Pearson correlation	6, str. 161
percentilni raspon	percentile interval	9, str. 237
permutacijski test	permutation test	10, str. 262
p-hakiranje	p-hacking	12, str. 303
planirana usporedba	planned comparison	15, str. 376
planiranje veličine uzorka	sample size planning	11, str. 283
podatkovna znanost	data science	1, str. 30
podrijetlo tvrdnje	claim provenance	3, str. 78
pogreška druge vrste	Type II error	10, str. 262
pogreška prve vrste	Type I error	10, str. 262
pogreška uzorkovanja	sampling error	8, str. 212
pogreške po podskupinama	subgroup errors	17, str. 432
pokrivenost	coverage	8, str. 212
pomak distribucije	distribution shift	17, str. 432

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
populacija	population	8, str. 212
post-hoc postupak	post-hoc procedure	15, str. 376
postotak i postotni bod	percentage and percentage point	3, str. 78
poštena rečenica izvještaja	honest reporting sentence	4, str. 105
pošten izvještaj	honest report	18, str. 466
potvrđujuća analiza	confirmatory analysis	12, str. 303; 18, str. 466
pouzdanost	reliability	2, str. 56
prag značajnosti	significance threshold	10, str. 262
preciznost	precision	9, str. 237
preciznost odnosno pozitivna prediktivna vrijednost	precision, positive predictive value	17, str. 432
precjenjivanje pri odabiru značajnih nalaza	selection-induced effect-size exaggeration	11, str. 283
prediktor	predictor	14, str. 353
predregistracija	preregistration	12, str. 303
predviđanje izvan uzorka	out-of-sample prediction	16, str. 408
preprilagodba	overfitting	17, str. 432
pridruživanje	aesthetic mapping	5, str. 132
prigodni uzorak	convenience sample	8, str. 212
prihvatljivost	eligibility	2, str. 56
prikaz učestalosti riječi	word-frequency plot	5, str. 132

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
prilagođena povezanost	adjusted association	16, str. 408
prilagođeni standardizirani rezidual	adjusted standardized residual	13, str. 333
pristupačnost prikaza	visualization accessibility	5, str. 132
procjena	estimate	9, str. 237
procjena s težinama	weighted estimate	8, str. 212
procjenjivana veličina	estimand	16, str. 408
protokol skeptičnoga čitanja	skeptical reading protocol	3, str. 78
provjera osjetljivosti	sensitivity analysis	12, str. 303; 18, str. 466
publikacijska pristranost	publication bias	12, str. 303
putovnica skupa podataka	dataset passport	18, str. 466
p-vrijednost	p-value	10, str. 262
QQ prikaz	Q–Q plot	7, str. 186
randomizacija	randomization	2, str. 56
razdvajanje na skup za učenje, provjeru i ispitivanje	train-validation-test split	17, str. 432
razina mjerenja	level of measurement	2, str. 56
razina pouzdanosti	confidence level	9, str. 237
razlika sredina	difference in means	14, str. 353
referentna oznaka	reference label	10, str. 262

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
referentna raspodjela	reference distribution	13, str. 333
referentna skupina	reference category	14, str. 353
registrirani izvještaj	registered report	12, str. 303
replikacija	replication	12, str. 303
reproducibilan tijek rada	reproducible workflow	18, str. 466
reproducibilnost	reproducibility	12, str. 303
rezidual	residual	16, str. 408
sadržajna važnost	practical significance	11, str. 283
samoodabir	self-selection	8, str. 212
signal	signal	1, str. 30
Simpsonov paradoks	Simpson's paradox	1, str. 30
sintetički zapis	synthetic record	3, str. 78
skraćena os	truncated axis	5, str. 132
skraćena sredina	trimmed mean	4, str. 105
slučajni niz	random sequence	7, str. 186
spajanje	join	4, str. 105
Spearmanov koeficijent ranga	Spearman rank correlation	6, str. 161
specifičnost	specificity	17, str. 432
središnji granični teorem	central limit theorem	8, str. 212
standardizirana razlika	Cohen's d	11, str. 283; 14, str. 353
standardizirana vrijednost	z-score	4, str. 105

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
standardna devijacija	standard deviation	4, str. 105
standardna pogreška	standard error	8, str. 212
statistička snaga	statistical power	11, str. 283
statistika	statistics	1, str. 30
statistika	statistic	8, str. 212
stopa obiteljske pogreške	familywise error rate	15, str. 376
strojno učenje	machine learning	1, str. 30
stupnjevi slobode	degrees of freedom	13, str. 333; 14, str. 353
sustav umjetne inteligencije	artificial intelligence system	1, str. 30
šest revizijskih pitanja	six audit questions	1, str. 30
širina intervala	interval width	11, str. 283
šum	noise	1, str. 30
tablica zabune	confusion matrix	17, str. 432
tekstna jedinica	text unit	17, str. 432
temeljna stopa	base rate	3, str. 78
testna statistika	test statistic	10, str. 262
test prilagodbe	goodness-of-fit test	13, str. 333
težina uzorkovanja	sampling weight	8, str. 212
točkasta procjena	point estimate	9, str. 237
trag odluka	decision trail	18, str. 466
t-raspodjela	t distribution	14, str. 353
Tukeyjev postupak	Tukey HSD	15, str. 376
učinak nacrta	design effect	8, str. 212
ukupna točnost	accuracy	17, str. 432

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
upareni podaci	paired data	14, str. 353
utjecajno opažanje	influential observation	5, str. 132; 6, str. 161
uvjetna vjerojatnost	conditional probability	7, str. 186
uvjetni nazivnik	conditional denominator	13, str. 333
uzorački udio	sample proportion	8, str. 212
uzorak	sample	8, str. 212
valjanost	validity	2, str. 56
varijabla	variable	2, str. 56
varijanca	variance	4, str. 105
varijanca između skupina	between-group variance	15, str. 376
varijanca unutar skupina	within-group variance	15, str. 376
veličina učinka	effect size	11, str. 283
višestruka regresija	multiple regression	16, str. 408
vjerojatnosno uzorkovanje	probability sampling	8, str. 212
vjerojatnost	probability	7, str. 186
vrt račvajućih putova	garden of forking paths	12, str. 303
Welchov postupak	Welch's t-test	14, str. 353
Wilcoxonov postupak predznačenih rangova	Wilcoxon signed-rank test	14, str. 353
zabilježeni referentni ishod	recorded reference outcome	17, str. 432

Hrvatski naziv	Engleski naziv	Poglavlje i stranica
zakon velikih brojeva	law of large numbers	7, str. 186
zaobljenost	kurtosis	4, str. 105
zbirna stopa	aggregate rate	1, str. 30
životni ciklus podataka	data lifecycle	1, str. 30

Dva skupa podataka za vježbu

Oba skupa podataka stvorili smo računalnom simulacijom, svaki zasebno. Jedan redak predstavlja jednu zamišljenu osobu. Isto ime varijable ne znači da oba skupa mjere isto obilježje ili obuhvaćaju iste osobe. Brojčani kodovi kategorija služe razmjeni datoteka i ne pretvaraju kategorije u brojčane mjere.

Anketa o društvenim mrežama ima 300 zapisa u datoteci `data/anketa-mreze.csv`.

Varijabla	Značenje
`ispitanik`	jedinstvena oznaka zamišljene osobe
`dob`	dob u godinama
`dobna_skupina`, `dobna_skupina_sifra`	dobna kategorija i njezin brojčani kod
`minute_dnevno`	dnevne minute na društvenim mrežama
`povjerenje`	generirana ocjena povjerenja od 1 do 10

Populacija medijskih navika ima 50 000 zapisa u datoteci `data/populacija-medija.csv`. Iz nje iznova izvlačimo uzorke jer znamo vrijednosti za cijelu nastavnu populaciju.

Varijabla	Značenje
`osoba`	jedinstveni broj zamišljene osobe
`dob`, `spol`	dob u godinama i generirana kategorija spola
`obrazovanje`, `obrazovanje_sifra`	kategorija obrazovanja i njezin kod
`izvor_vijesti`, `izvor_vijesti_sifra`	glavni izvor vijesti i njegov kod
`povjerenje_medijima`	generirana ocjena povjerenja od 1 do 10
`minute_medija`	dnevne minute praćenja medija
`spremnost_platiti`	generirani iznos; nula označuje nespremnost na plaćanje

U popratnim agregatima jedan redak predstavlja dobnu skupinu ili izvor vijesti. Brojnosti, zbrojevi i prosjeci omogućuju provjeru računa, ali ne zamjenjuju pojedinačne zapise pri provjeri njihova nastanka.

Rezultati tih dvaju skupova ne potkrepljuju tvrdnje o stvarnim stanovnicima Hrvatske. Opisi stvarnih izvora i njihovih ograničenja nalaze se uz primjere i u mrežnom katalogu podataka.

Oznake u formulama

Ista oznaka čita se uz cilj i jedinicu navedene uz konkretni račun.

Oznaka	Značenje
n, N	broj jedinica u uzorku i u konačnoj populaciji
$\bar{x}$, μ	uzoračka i populacijska sredina
s, s^2	uzoračka standardna devijacija i varijanca
σ, σ^2	populacijska standardna devijacija i varijanca
$\hat{p}$	procijenjeni udio; razlikuje se od p-vrijednosti testa
SE	standardna pogreška procjenitelja
z^*, t^*	kritične vrijednosti normalne i t-raspodjele
α	unaprijed odabrana razina testa
d, s_{zdr}	standardizirana razlika i združena standardna devijacija
O_{ij}, E_{ij}	opažena i očekivana frekvencija ćelije
χ^2, V	hi-kvadrat statistika i Cramérova mjera povezanosti
df	stupnjevi slobode referentne raspodjele
F, η^2	omjer srednjih kvadrata i udio objašnjene varijabilnosti
r, R^2	Pearsonova korelacija i koeficijent determinacije
β_0, β_j	odsječak i koeficijent prediktora u populacijskom modelu
ε_i	odstupanje ishoda od populacijske regresijske sredine
AOR	prilagođeni omjer izgleda u logističkom modelu

provjere za samostalan rad

Ovdje je barem jedna provjera iz svakoga od osamnaest poglavlja, uz dodatne odabrane računske provjere. Najprije pokušajte riješiti zadatak. Ako zastanete, pročitajte samo njegovu naznaku. Objašnjenja su odvojena i počinju na novoj stranici. Usporedite postupak i granice svojega zaključka s primjerom, a zatim odgovorite ponovno bez gledanja. Drukčije sročen odgovor može biti jednako dobar.

Naznake

Poglavlje 1 — Konceptualni zadatak

Za svaku stopu odredite skupinu i nazivnik. Zatim povežite odabrane faze rada s pogreškama koje u njima mogu nastati, a svaku djelatnost s pitanjem na koje odgovara.

Poglavlje 2 — Konceptualni zadatak

Odvojite dosljednost mjerenja od onoga što mjera zahvaća. Za svaku varijablu u primjeru razmotrite kada nastaje i na što može

djelovati. Razlikujte svojstvo randomizacije kroz ponovljene dodjele od ishoda jedne dodjele.

Poglavlje 3 — Konceptualni zadatak

Najprije pročitajte što svaki ukupni broj označuje. Razliku među brojevima odvojite od objašnjenja njezina uzroka. U primjerima upozorenja provjerite koja se veličina promijenila, a koje su ostale iste.

Poglavlje 4 — Konceptualni zadatak

Pratite što jedan redak predstavlja prije i nakon spajanja. Zasebno provjerite broj redaka, broj različitih ključeva i zbroj objava. Razmotrite što svaka provjera može otkriti, a što joj može promaknuti.

Poglavlje 4 — Računski zadatak

Za sredinu trebate svih deset vrijednosti. Za medijan poredajte niz i pronađite dva središnja mjesta. U drugom dijelu zadatka razlikujte broj osoba u dobnoj skupini od broja osoba u cijeloj anketi.

Poglavlje 5 — Konceptualni zadatak

Najprije odredite usporedbu koju graf treba sačuvati. Navedite što bi čitatelj izgubio bez položaja točaka, oznaka osi i izvora.

Poglavlje 6 — Konceptualni zadatak

U izvornom zapisu zasebno pročitajte brojčanu vrijednost, oznaku dostupnosti i status pouzdanosti. Zatim odredite koji zapisi tvore potpune parove i na koju se jedinicu zaključak smije odnositi.

Poglavlje 7 — Konceptualni zadatak

Odvojite poznatu vjerojatnost iz zadanoga modela od stope koju tek procjenjujemo. Razmotrite kakav bi dokaz pokazao promjenu procesa. Prije primjene pravila za normalnu raspodjelu pogledajte oblik podataka.

Poglavlje 7 — Računski zadatak

Najprije utvrdite nazivnik svakog od triju udjela. Za dvadeset objava krenite od vjerojatnosti da nijedna ne uspije i tek zatim upotrijebite komplement. Provjerite pretpostavku neovisnosti.

Poglavlje 8 — Konceptualni zadatak

Odredite što predstavlja jedna točka u svakoj raspodjeli. Zatim razmotrite opisuje li njezina širina razlike među osobama ili promjene sredine kroz ponovljene uzorke.

Poglavlje 8 — Računski zadatak

Najprije zbrojite odgovore i prebrojite retke. Zatim svaki odgovor pomnožite pripadnom težinom dva ili četiri, zasebno zbrojite ponderirane odgovore i sve težine te usporedite doprinose dvaju slojeva nazivniku. Račun pretpostavlja jednostavni slučajni uzorak, malu uzoračku frakciju i udio 0,5. Ne obuhvaća pogreške

pokrivenosti ili mjerenja. Za raspon oko populacijske sredine pomnožite svaku standardnu pogrešku s 1,96.

Poglavlje 9 — Konceptualni zadatak

Razdvojite što se mijenja između uzoraka od onoga što ostaje fiksno. Provjerite na što se odnosi postotak pouzdanosti.

Poglavlje 9 — Računski zadatak

Za prvi interval odvojite prosjek, standardnu pogrešku i marginu. Za ručni bootstrap prepišite zadane retke, uključujući ponovljeni redak, pa ponovno poredajte vrijednosti. Prije računanja predvidite poredak širina i obuhvat. Usporedite predviđanje s rezultatima i razdvojite dugoročni obuhvat od promašaja u prikazanom nizu. Za portal zadržite ukupan broj redaka kao nazivnik udjela, a broj korisnika portala kao nazivnik njihova prosjeka.

Poglavlje 10 — Konceptualni zadatak

U svakoj tvrdnji odvojite ono što pretpostavljamo od događaja čiju vjerojatnost računamo. Provjerite mijenja li se smjer uvjetovanja i koje bi dodatne informacije trebale za drugi zaključak.

Poglavlje 11 — Konceptualni zadatak

Usporedite sve simulirane studije s podskupom koji ostaje nakon praga značajnosti.

Poglavlje 12 — Konceptualni zadatak

Pratite kada je pravilo odabrano i jesu li iskušana druga pravila analize. Zatim razmotrite koji bi zapis omogućio provjeru odluke. Kod ponavljanja istraživanja provjerite ostaju li podaci isti ili se prikupljaju novi.

Poglavlje 13 — Konceptualni zadatak

Prva rečenica treba govoriti o odstupanju od nezavisnosti, a druga o jačini obrasca. U verziji za novinara najprije odlučite koji je konkretni kontrast razumljiv bez naziva testa, a zatim sačuvajte podatke potrebne za provjeru tog kontrasta.

Poglavlje 13 — Računski zadatak

Najprije izračunajte ukupan broj osoba. Očekivana ćelija primjenjuje udio pripadnog stupca na veličinu retka. Preostale opažene ćelije dovršite iz zadanih rubova prije računanja odstupanja.

Poglavlje 14 — Konceptualni zadatak

Za svaki slučaj odredite tko daje opažanje i odakle dolazi drugi broj. On može biti unaprijed zadana konstanta, rezultat drugih osoba ili drugo mjerenje istih osoba.

Poglavlje 14 — Kritički zadatak

Odvojite pitanje što ljestvica mjeri od pitanja koliko je procjena precizna. Zatim isti interval usporedite zasebno s nulom i s pragom od dva boda. Ne mijenjajte prag nakon gledanja rezultata.

Poglavlje 15 — Konceptualni zadatak

Razlikujte tvrdnju o svim sredinama od tvrdnje o pojedinom paru. U skici usporedite razmake sredina s granicom razlučivosti. Za odabrani par provjerite kada je izabran i među kojim je parovima mogao biti odabran.

Poglavlje 16 — Konceptualni zadatak

Provjerite jesu li se promijenili podaci ili pitanje koje model postavlja. Usporedite zbirni nagib s nagibom pri jednakom izvoru vijesti, a zatim razmotrite mogu li skupine imati različite nagibe.

Poglavlje 17 — Konceptualni zadatak

Pratite kada se svaki skup smije upotrijebiti i može li isti dokument prenijeti informacije preko granice skupova. Za svaku stopu pogreške odredite nazivnik, a za odluku pravilo kojim nastaje iz predviđene vjerojatnosti.

Poglavlje 18 — Konceptualni zadatak

Krenite od jednog retka i odredite njegov odnos prema dokumentu i cijelom paketu. Provjerite odabir rečenica, postupak označavanja i nazivnik udjela. Razdvojite opis pripremljenih podataka od tvrdnji o široj populaciji ili budućoj primjeni.

Objašnjenja

Poglavlje 1 — Konceptualni zadatak

Zbirna stopa opisuje ukupni ishod mješavine podskupina, a stopa unutar podskupine uspoređuje skupine pri istoj vrijednosti relevantne treće varijable. Ta je razlika objašnjena uz Simpsonov paradoks u poglavlju. U odgovoru treba imenovati jedinicu analize i nazivnik zbirne stope te stope unutar podskupine i pokazati da nije riječ o istoj usporedbi. Jedinica ostaje opažana jedinica, dok nazivnik prelazi s cijele skupine na dio skupine unutar podskupine.

Za dvije odabrane faze navodimo po jednu konkretnu pogrešku koju kasniji račun ne može automatski popraviti. Primjeri mogu uključiti pogrešan izvor, izostale jedinice, duplicirane retke, pogrešno mjerenje, nedokumentiranu transformaciju ili tvrdnju jaču od dizajna.

Statistici pripada provjera onoga što podaci opravdavaju vjerovati, što je njezino vodeće pitanje u tablici Četiri djelatnosti i četiri pitanja. Podatkovna znanost bavi se izradom ponovljiva puta od izvora do tablice, s ciljem pouzdane i reproducibilne analize stvarnih izvora. Strojno učenje vrednuje predviđanja i ispituje kako obrazac radi na novim opažanjima. Praćenje preporučivačkoga sustava nakon uvođenja u ustanovu pripada sustavu umjetne inteligencije jer uključuje institucionalnu uporabu te praćenje pogrešaka i posljedica uvedenoga sustava.

Odgovor razlikuje nazivnike dviju stopa, povezuje pogreške s fazama rada i razvrstava djelatnosti prema pitanju koje rješavaju. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Dvije su stope proturječne jer iz istih podataka daju različite zaključke.

Stope nisu proturječne nego odgovaraju na različita pitanja s različitim nazivnicima i razinama usporedbe.

Poglavlje 2 — Konceptualni zadatak

Primjer pokazuje instrument koji daje dosljedan rezultat, ali taj rezultat ne zahvaća ciljani pojam. Odvojeno su obrazložene postojanost mjerenja i promašaj sadržaja, primjerice dosljedno brojanje zareza kao nevaljana mjera kvalitete novinarstva. Za dvije operacionalizacije istoga pojma treba navesti različita pravila opažanja i objasniti što svaka zahvaća, a što izostavlja. Primjer može usporediti samoprijavljene minute izloženosti vijestima s digitalno zabilježenim otvaranjima, uz imenovane razlike u pamćenju, platformi i značenju izloženosti. Odgovor opisuje nalaz koji bi jedna operacionalizacija mogla proizvesti, a druga ne, bez tvrdnje da jedna brojka predstavlja cijeli pojam. Razlika u mogućem nalazu izravno slijedi iz onoga što pojedina mjera bilježi ili izostavlja.

Za jednu operacionalizaciju imenovani su jedinica, ciljna skupina, pravilo prihvatljivosti, opravdano isključenje, učinak isključenja na doseg i razina mjerenja. Svaki element ima sadržajno pravilo, a suženje dosega odgovara jedinicama koje je isključenje uklonilo. Ako odabrana operacionalizacija kodira jezik, odgovor daje jednoznačno pravilo oznake i čuva dvosmislen slučaj kao zasebnu odluku. Pravilo imenuje opažljiv tekstualni znak, a ne dojam kodera, te navodi postupak poput oznake ‘nejasno’ ili dvostrukoga kodiranja.

Političko zanimanje izmjereno prije praćenja jest mogući zajednički uzrok korištenja političkih vijesti i povjerenja. Varijabla vremenski prethodi objema promatranim veličinama i može djelovati na obje. Znanje stečeno nakon praćenja jest mogući medijator između korištenja vijesti i povjerenja. Praćenje prethodi znanju, a znanje može prenijeti dio djelovanja prema povjerenju. Ulazak u analizu koji ovisi i o korištenju vijesti i o povjerenju jest kolider. Ulazak je zajednička posljedica dviju promatranih veličina, pa uvjetovanje na nj može stvoriti vezu. Randomizacija uravnotežuje poznata i nepoznata početna obilježja u očekivanju, ali ne jamči jednaku početnu skupinu u jednom malom pokusu.

Objašnjenje razlikuje svojstvo postupka kroz ponovljene dodjele od slučajne ostvarene neravnoteže jednoga uzorka.

Odgovor razdvaja dosljednost i valjanost mjerenja, obrazlaže pravila opažanja i obuhvat te razlikuje uloge varijabli i granice randomizacije. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Pouzdana mjera automatski je valjana jer svaki put daje isti rezultat.

Dosljednost ne pokazuje da instrument zahvaća pojam o kojem se želi zaključivati. Mjera može postojano promašivati cilj.

Poglavlje 3 — Konceptualni zadatak

Broj birača koji su pristupili bilježi jednu operativnu fazu, a zbroj važećih i nevažećih listića broj birača prema glasačkim listićima. Te su veličine zasebno označene i iskazane u službenom izvješću (Državno izborno povjerenstvo Republike Hrvatske 2024., 124). Razlika od 1 554 pokazuje veličinu nepodudaranja dvaju ukupnih brojeva, ali sama ne dokazuje kvar, uzrok ni dopuštenje za zamjenu oznaka. U zadanom dokazu možemo provjeriti izračunatu razliku, ali nema izvornog objašnjenja zašto je nastala.

Veća temeljna stopa povećava broj stvarnih ciljanih zapisa među svim zapisima, pa uz iste stope provjere raste udio točnih među izdanim upozorenjima. U prvom primjeru 90 točnih upozorenja stoji nasuprot 495 pogrešnih, a u drugom 900 nasuprot 450. Promjena udjela ispravnih upozorenja ne smije se opisati kao tehničko poboljšanje provjere jer su njezine dvije zadane stope ostale jednake. Jedina promijenjena veličina u hipotetskim primjerima jest temeljna stopa ciljanoga ishoda.

Odgovor razlikuje nepodudaranje brojeva od objašnjenja njegova uzroka te promjenu temeljne stope od promjene svojstava provjere. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Brojnici su zamjenjivi jer dolaze iz iste tablice.

Zajednički izvor ne briše razliku među operativnim fazama koje tablica zasebno označava.

Poglavlje 4 — Konceptualni zadatak

Mjesečna tablica prije spajanja ima 438 redaka, 438 jedinstvenih kombinacija mjesec plus platforma i zbroj 710 307 objava, što potvrđuje neovisno prebrojavanje i zbrajanje mjesečnoga DigiKatova CSV-a (Šikić i sur. 2026.). Spajanje samo po godini daje 3 571 redak i zbroj 5 959 081, ali i dalje samo 438 različitih mjesečnih ključeva. Svaki se mjesečni redak ponavlja za broj platformi u svojoj godini, bez stvaranja nove kombinacije mjesec plus izvorna platforma. Sva polja mogu ostati popunjena i svi izvorni ključevi prisutni, a jedinica se ipak mijenja jer se isti mjesečni ključ pojavljuje više puta. To pokazuje razlika između 3 571 ukupnih redaka i 438 različitih mjesečnih ključeva. Potrebne su sve tri provjere jer broj redaka, jedinstvenost ključa i zbroj objava čuvaju različita svojstva analitičke tablice. U pogrešnom spoju skup jedinstvenih ključeva ostaje nepromijenjen dok se druge dvije vrijednosti mijenjaju.

Odgovor zajedno provjerava broj redaka, jedinstvenost mjesečnog ključa i zbroj objava te pokazuje zašto popunjena polja nisu dokaz valjanog spoja. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Ako nema praznih polja, broj redaka mora ostati isti.

Potpuno popunjen many-to-many spoj može pridružiti više godišnjih redaka svakom mjesečnom retku.

Poglavlje 4 — Računski zadatak

Zbijeni niz ima zbroj 110, pa je sredina 110/10 = 11. Peto i šesto mjesto imaju vrijednost 11, pa je i medijan 11. U nizu s krajnjim opažanjem zbroj je 169, sredina 16,9, a medijan (11 + 12)/2 = 11,5. Sredina se povećala za 5,9, a medijan za 0,5 jer krajnja vrijednost ulazi u puni zbroj, dok medijan određuju središnja mjesta.

U agregatu za prvu dobnu skupinu prosjek je 7 339/90 = 81,5444 minute, a udio 90/300 = 0,30. Zbroj broja osoba kroz četiri skupine jest 300, a zbroj minuta 15 019. Prosjek cijele ankete zato je 15 019/300, približno 50,0633 minute. Prosjek skupine i njezin udio imaju različite nazivnike.

Za zbijeni niz SD iznosi 1,82574, a IQR 2. Nakon zamjene opažanja SD iznosi 18,74656, a IQR 2,75. Vrijednost z = 1,5 znači položaj 1,5 standardnih devijacija iznad sredine referentne skupine. Ne znači 1,5 % osoba. Zbroj kvadrata odstupanja zbijenog niza jest 30, pa SD = korijen iz 30/9. Kvartile računamo navedenom linearnom interpolacijom tipa 7.

Odgovor pokazuje račun, objašnjava razliku sredine i medijana te zasebno provjerava dva nazivnika. Prijepis gotovog prosjeka iz tablice ne bi pokazao kako je broj provjeren.

Česta stranputica je sljedeći odgovor.

> Oba stanja imaju sredinu 11 jer se devet vrijednosti grupira oko 11.

Krajnje opažanje 70 ulazi u puni zbroj i podiže sredinu na 16,9.

Poglavlje 5 — Konceptualni zadatak

Za oblik jedne brojčane raspodjele prikladan je histogram ili krivulja gustoće, uz svijest da razredi ili izglađivanje utječu na vidljivi oblik. Prikaz čuva približan raspored i repove, ali odbacuje točan položaj pojedinoga opažanja i druge varijable. Za usporedbu kategorija prikladni su stupci na zajedničkoj osi koja počinje od nule. Položaj i duljina omogućuju usporedbu učestalosti ili drugoga imenovanog sažetka, ali ne prikazuju raspodjelu unutar kategorije. Za odnos dviju brojčanih varijabli prikladan je raspršeni dijagram. Svaka oznaka čuva jedan par vrijednosti, dok se gusto preklapanje i neprikazane dodatne varijable navode kao granice.

Alternativni tekst imenuje varijable, jedinicu, smjer ili razliku i najizraženiju osobinu odabranoga prikaza. Opis prenosi nalaz koji nije već sadržan u naslovu ili pukom imenu vrste grafa. Glavni nalaz ostaje čitljiv bez boje samo ako ga nosi i položaj, oblik, vrsta linije ili izravna oznaka. Provjera bez boje razlikuje dekorativnu boju od boje koja je jedini nositelj podatka. Primjer alternativnog teksta glasi ovako. Raspršeni dijagram prikazuje 72 generirana opažanja u skupinama A i B. Na vodoravnoj je osi x, a na okomitoj y. U objema skupinama y uglavnom raste s x, uz raspršenost i preklapanje skupina. Različiti oblici točaka označuju skupine.

Odgovor povezuje vrstu grafa s pitanjem, imenuje što prikaz izostavlja i prenosi glavni nalaz riječima te bez oslanjanja samo na boju. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Za sva tri pitanja najbolji je stupčani dijagram jer ga je najlakše čitati.

Stupci odbacuju oblik jedne brojčane raspodjele i pojedinačne parove dviju brojčanih varijabli, pa ne odgovaraju dvama zadanim pitanjima.

Poglavlje 6 — Konceptualni zadatak

Izvorni jedinstveni ključ jest država + godina + pokazatelj, a nakon širenja se parovi povezuju po državi i godini. Svaka od 162 kombinacije u izvatku jedinstvena je unutar 27 država, 2025. godine i šest pokazatelja. Luksemburg ulazi u glavni odnos, koji ima 27 država, jer su oba tražena pokazatelja brojčana. Redci za tercijarno obrazovanje i uporabu interneta u Luksemburgu nose dostupne brojčane vrijednosti. Dvotočka u luksemburškom zapisu ne ulazi kao nula u par s ranim napuštanjem obrazovanja, pa taj odnos ima najviše 26 potpunih parova. Vrijednost je označena kao nedostupna, uz status u, dok redak ostaje sačuvan u datoteci. Hrvatskih 2,1 ostaje brojčana vrijednost, a status u ostaje uz nju kao upozorenje o niskoj pouzdanosti. Polje vrijednosti sadrži 2,1 i dostupnost da. Status ne zamjenjuje vrijednost.

Dopuštena rečenica opisuje pozitivnu povezanost dvaju državnih udjela u 27 članica EU-a 2025. i ne pripisuje je pojedincima ili uzroku. Jedna točka i jedan redak nakon spajanja predstavljaju državu, a izvorni par ima potpun obuhvat EU-27.

Odgovor razlikuje brojčanu vrijednost, nedostupnost i status pouzdanosti te zadržava državu kao jedinicu usporedbe. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Svaki redak sa statusom u treba ukloniti.

Status niske pouzdanosti nije isto što i odsutna brojčana vrijednost. Hrvatskih 2,1 ostaje dostupno uz sačuvano upozorenje.

Poglavlje 7 — Konceptualni zadatak

U modelu poštenoga novčića s neovisnim bacanjima sljedeće bacanje i nakon pet glava ima vjerojatnost 0,5 za glavu. Poznata stabilna vjerojatnost i neovisnost dio su zadanoga modela, pa prethodni ishodi ne mijenjaju sljedeći. Pet uspješnih objava može promijeniti procjenu nepoznate stope kampanje čak i ako su ishodi uvjetno neovisni uz jednu stabilnu stopu. Novi podaci omogućuju novu procjenu nepoznate stope. To nije isto što i međusobni utjecaj objava. Tvrdnja da se stopa tijekom procesa mijenja traži dokaz koji razlikuje vrijeme, uvjete ili intervenciju, a ne samo niz uspjeha. Prihvatljivi su unaprijed određen vremenski kontrast, ponovljeni segmenti, nasumično dodijeljeni uvjeti ili procesni podaci koji otkrivaju ovisnost.

Pravilo 68–95–99,7 nije opravdano za raspodjelu s velikom skupinom nula i dugim repom. Histogram pokazuje nagomilavanje na nuli, a QQ prikaz vodoravan odsječak i sustavno savijanje od pravca.

Odgovor odvaja poznati model, procjenu nepoznate stope i dokaz promjene procesa te provjerava oblik raspodjele prije primjene pravila za normalnu raspodjelu. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Pet glava dokazuje da novčić nije pošten, a pet uspješnih objava dokazuje da kampanja postaje bolja.

Kratak niz može nastati pod poznatim modelom, a uspjesi pri nepoznatoj stopi mijenjaju procjenu stope bez dokaza da se proces mijenja kroz vrijeme.

Poglavlje 7 — Računski zadatak

90/100 = 90 %. 90/(90 + 495) = 90/585 = 0,153846..., odnosno 15,4 %. 100/10 000 = 1 %. Preostale ćelije tablice jesu 10 propuštenih ciljeva i 9 405 ispravno neupozorenih zapisa. Vjerojatnost bez uspjeha iznosi $0{,}98^{20}$ = 0,667608, pa je vjerojatnost barem jednoga uspjeha 1 – $0{,}98^{20}$ = 0,332392, približno 33,2 %.

Odgovor razlikuje nazivnike i navodi pretpostavku koja opravdava množenje vjerojatnosti. Umnožak pretpostavlja neovisne pokušaje i jednaku zadanu vjerojatnost. Kod ovisnih objava trebali bismo znati uvjetne vjerojatnosti.

Česta stranputica je sljedeći odgovor.

> Provjera je pouzdana 90 % jer nalazi 90 od 100 ciljanih zapisa.

To je udio upozorenja među ciljanima, a ne udio ciljanih među svim upozorenjima. Drugi nazivnik uključuje i 495 pogrešnih upozorenja.

Poglavlje 8 — Konceptualni zadatak

Jedinica raspodjele pojedinačnih opažanja jest jedna osoba ili druga promatrana jedinica. Svaka vrijednost dolazi izravno iz populacije ili uzorka, prije izračuna sredine. Širina raspodjele opažanja opisuje heterogenost pojedinaca i mjeri se standardnom devijacijom. Uz fiksnu populaciju očekivana širina pojedinačnih vrijednosti ne pada samo zato što je uzorak veći.

Jedinica distribucije uzorkovanja jest uzoračka sredina iz jednoga ponovljenog uzorka. Svaka točka sažima cijeli uzorak, a mnoge točke nastaju ponavljanjem istoga postupka odabira. Širinu distribucije sredina mjeri standardna pogreška, koja se u jednostavnom modelu smanjuje približno s korijenom iz n. Veći n smanjuje utjecaj pojedinačnih odstupanja na sredinu, pa sredine kroz ponovljene uzorke postaju zbijenije.

Odgovor razlikuje opažanja od uzoračkih sredina i objašnjava zašto se njihove raspodjele različito ponašaju pri povećanju uzorka. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Obje raspodjele prikazuju ljude, samo je druga nacrtana iz većega uzorka.

Distribucija uzorkovanja ne prikazuje ljude nego po jednu statistiku iz svakoga ponovljenog uzorka.

Poglavlje 8 — Računski zadatak

Bez težina 3/6 = 0,50, odnosno 50 %. S težinama ponderirani je brojnik 2 + 2 + 2 = 6, ponderirani nazivnik 2 + 2 + 2 + 2 + 4 + 4 = 16, pa je 6/16 = 0,375, odnosno 37,5 %. Razlika iznosi −12,5 postotnih bodova. Za n = 64 standardna pogreška jest 12/8 = 1,5. Za n = 256 jest 12/16 = 0,75. Prepolovljenje traži četverostruki uzorak. Za n = 400 približna margina udjela iznosi 1/20 = 0,05, odnosno pet postotnih bodova.

Za 64 osobe dobivamo 1,96 × 1,5 = 2,94, a za 256 osoba 1,96 × 0,75 = 1,47. Približno 95 % uzoračkih sredina tada bi bilo unutar ±2,94, odnosno ±1,47 jedinica od populacijske sredine.

Prikazani su računi, njihove pretpostavke i granica zaključka. Raspon pretpostavlja normalno približenje. Margina udjela od pet postotnih bodova opisuje uzoračku nesigurnost pod zadanim nacrtom. Ne uključuje pristranost odabira ni pogrešku mjerenja.

Česta stranputica je sljedeći odgovor.

> Procjena s težinama iznosi 6/6 = 100 % jer je ponderirani brojnik šest.

Ponderiranju podliježe i nazivnik. Zbroj svih težina iznosi šesnaest.

Poglavlje 9 — Konceptualni zadatak

Slučajni su uzorak i interval koji bi novi uzorak proizveo. Kroz ponovljena uzorkovanja mijenjaju se opažanja, procjena i granice. Populacijski parametar ostaje fiksan kroz sva ponavljanja. Svaki novi interval uspoređujemo s istom vrijednošću parametra. Razina pouzdanosti opisuje dugoročni udio intervala postupka koji obuhvaćaju parametar. Nakon opažanja granica pojedini interval parametar obuhvaća ili ne obuhvaća.

Bootstrap gotovih oznaka ne obuhvaća nesigurnost pravila kodiranja. Ponovno uzorkovanje mijenja zastupljenost već označenih tekstova, ali ne prekodira rubne slučajeve drugim obranjivim pravilom.

Odgovor vezuje pouzdanost uz ponovljeni postupak te razlikuje nesigurnost uzorkovanja od nesigurnosti kodiranja. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Parametar je slučajan jer ga ne poznajemo.

Nepoznatost nije isto što i slučajnost u ovom frekventističkom postupku. Populacijski parametar ostaje fiksan kroz ponavljanja.

Poglavlje 9 — Računski zadatak

Predviđamo poredak širina B > A > C, a račun daje približno 0,814603, 0,619806 i 0,309903. Dugoročni obuhvat iznosi 99 % za B te 95 % za A i C. Četverostruki uzorak u C prepolovljuje širinu u odnosu na A. U prikazanom nizu A ima tri promašaja, B nijedan, a C jedan. Taj jedan niz ne utvrđuje dugoročni obuhvat postupka.

Tiskana tablica zadaje ukupno 50 000 osoba, 15 101 korisnika portala i njihov zbroj povjerenja od 72 101. Udio je 15 101/50 000 = 0,30202, odnosno 30,202 %. Prosjek povjerenja korisnika portala jest 72 101/15 101, približno 4,774584 boda. Ti brojevi odgovaraju agregatu. Ako radimo samo s tiskanom tablicom, možemo provjeriti dijeljenja, ali ne tvrdimo da smo provjerili retke analitičke datoteke.

Točkasta procjena ovdje je prosjek 4,815. Standardna pogreška iznosi 2,093/√200 ≈ 0,148, a margina približno 0,290 boda. Granice su približno 4,525 i 5,105. Razina od 95 % odnosi se na dugoročni obuhvat postupka kada vrijede njegove pretpostavke. Početni medijan pet vrijednosti jest 134 minute. Ponovni uzorak glasi 83, 83, 82, 139, 164 i ima medijan 83 minute.

Odgovor odvaja promjenu preciznosti od obuhvata i provjerava brojnik i nazivnik svakog sažetka. Usporedba s agregatom čuva istu skupinu i istu definiciju povjerenja. Poredani početni niz jest 82, 83, 134, 139, 164. Ponavljanje i izostavljanje zapisa mijenjaju medijan. Za percentilni interval trebali bismo mnogo takvih ponovnih uzoraka. Ova demonstracija s pet vrijednosti ne zamjenjuje raniji račun na svih 60 osoba.

Česta stranputica je sljedeći odgovor.

> Postavka B uža je od A jer ima višu razinu pouzdanosti.

Pri istom n viša kritična vrijednost širi interval. B je široka približno 0,815, a A približno 0,620.

Poglavlje 10 — Konceptualni zadatak

Ako nulti model vrijedi, p-vrijednost je vjerojatnost rezultata barem toliko neusklađenoga s njime kao opaženi. Uvjet je da vrijedi H0, a događaj je opažena ili ekstremnija testna statistika. Vjerojatnost da nulti model vrijedi nakon podataka jest druga uvjetna vjerojatnost. Uvjet su opaženi podaci, a događaj je H0. Iz prve uvjetne vjerojatnosti ne možemo izravno izračunati drugu bez priora i specificiranih modela podataka pod nultom i konkurentskim hipotezama. Bayesovo pravilo traži početne izglede i vjerojatnosni doprinos podataka. Repna p-vrijednost ne sadrži te sastavnice.

Odgovor jasno imenuje uvjet i događaj u objema vjerojatnostima te informacije potrebne za prijelaz na tvrdnju o nultoj hipotezi. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> P-vrijednost nije vjerojatnost nulte hipoteze jer je obično mala.

Veličina broja ne objašnjava smjer uvjetovanja niti imenuje informacije potrebne za posteriornu tvrdnju.

Poglavlje 11 — Konceptualni zadatak

Prosjek svih procjena ostaje blizu poznate razlike, dok se prosjek mijenja nakon odabira samo procjena koje su prešle prag. Gornja raspodjela obuhvaća sve studije, a donja je uvjetovana na p-vrijednost ispod 0,05. Iskrivljenje nastaje u koraku filtriranja prema pragu. Pri malom uzorku prag češće prelaze procjene koje je slučajnost udaljila od nule, uglavnom povećavajući stvarnu razliku.

Omjer prosjeka odabranih procjena i istine nije univerzalna konstanta. Poglavlje ga veže uz jednu simuliranu populaciju, uzorak od šezdeset, jedan permutacijski postupak i jedan prag. Niska snaga usmjerava pozornost na intervale, selekciju i neovisne replikacije, ali sama ne presuđuje pojedinom nalazu. Velik učinak tik ispod praga traži dodatne provjere, a nije automatski lažan ili precijenjen.

Odgovor pronalazi iskrivljenje u odabiru značajnih rezultata i ne prenosi obrazac iz simulacije na svaki pojedini nalaz. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Male studije uvijek daju prevelike učinke.

Prosjek svih malih studija u simulaciji ostaje blizu istini. Pretjerivanje nastaje tek u odabranom podskupu koji je prešao prag.

Poglavlje 12 — Konceptualni zadatak

Naknadno iskušavanje pet pravila i selektivna objava poželjnoga ishoda primjer su p-hakiranja. Odluke su donesene nakon pregleda podataka, više je rezultata izračunano, a vidljiv je samo poželjni put. Uvjetno pravilo za medijan ili sredinu primjer je vrta račvajućih putova. Provedena je jedna grana, ali njezin izbor ovisi o opaženoj asimetriji i bio bi drukčiji pod drukčijim podacima. Prvi primjer traži trag svih iskušanih pravila i rezultata, a drugi vremenski označeno uvjetno pravilo, kriterij asimetrije i zapis odabrane grane. Tek cjelovit trag pokazuje skup razmotrenih putova, vrijeme odluka i razlog konačnoga izbora.

Reproducibilnost provjeravamo ponavljanjem istoga računa iz istih podataka i materijala, a pri replikaciji prikupljamo nove podatke za usporedivu ciljnu tvrdnju. Prvi postupak provjerava trag izračuna, a drugi donosi podatke iz novog istraživanja.

Odgovor razlikuje pretraživanje rezultata od uvjetnog izbora analize i navodi zapis potreban za njihovu provjeru. Pritom odvaja ponavljanje računa od istraživanja na novim podacima. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Oba su primjera p-hakiranje jer u oba postoji više mogućih analiza.

Drugi primjer ne pretražuje rezultate nakon analize. Njegova je fleksibilnost u podatkovno uvjetovanom pravilu koje bi odabralo drugu granu pod drugim uzorkom.

Poglavlje 13 — Konceptualni zadatak

Dvije rečenice za tablicu s pet izvora mogu glasiti ovako. Hi-kvadrat test uz p-vrijednost manju od 0,001 pokazuje da simulirana tablica snažno odstupa od modela nezavisnosti. Cramérovo V od 0,25 sažima jačinu povezanosti, dok usporedbe udjela pokazuju njezino sadržajno značenje.

Za novinara, nakon spajanja izvora, prihvatljiva je sljedeća verzija u tri rečenice. U nastavnoj simulaciji na slučajnom uzorku od 800 zamišljenih osoba digitalne izvore bira 75,6 % osoba od 18 do 29 godina i 19,3 % osoba od 60 i više godina, pri čemu je svaki postotak izračunan unutar svoje dobne skupine. Razlika je 56,4 postotna boda uz 95-postotni interval od 47,4 do 65,3. To opisuje simulirane osobe i ne daje procjenu navika stanovnika Hrvatske niti dokaz da dob uzrokuje izbor izvora.

Prvi dio razdvaja p-vrijednost i veličinu povezanosti. Drugi dio čuva pitanje, populaciju, nazivnike, procjenu, interval i granicu zaključka. Ukupni hi-kvadrat i V mogu ostati u pratećoj tablici jer se glavni tekst usredotočuje na imenovani dobni kontrast. V od 0,25 odnosi se na pet izvora, a ne na spojenu tablicu.

Česta stranputica je sljedeći odgovor.

> Test je značajan, pa je veza snažna.

Jedna rečenica spaja dva pitanja i pogrešno pretvara p-vrijednost u mjeru jačine.

Poglavlje 13 — Računski zadatak

Ukupno je 200 osoba. Za prvi stupac očekivanje u svakom retku jest 100 × 120/200 = 60, a za drugi 100 × 80/200 = 40. Opažena prva ćelija ima 70 osoba, pa druga u istom retku mora imati 30. Prvi stupac ukupno ima 120, pa je donja lijeva ćelija 50, a posljednja također 50.

Svako odstupanje ima apsolutni iznos 10. Doprinosi su zato 100/60, 100/40, 100/60 i 100/40, odnosno približno 1,67, 2,50, 1,67 i 2,50. Zbroj bez međukoraka zaokruživanja jest 25/3, približno 8,33. On premašuje zadanu granicu 3,84 uz jedan stupanj slobode, pa na toj razini odbacujemo model nezavisnosti. Iz toga još ne slijedi da je povezanost snažna ili uzročna.

Odgovor čuva oba rubna zbroja, pokazuje sve doprinose i izračun tumači u odnosu na nulti model. Broj 8,33 sam bez očekivane i opažene tablice ne bi bio dovoljan trag računa.

Česta stranputica je sljedeći odgovor.

> Sve četiri očekivane frekvencije iznose 50 jer svaki redak ima sto ispitanika.

Stupci nisu jednaki. Očekivanje čuva stupčane zbrojeve 120 i 80, pa svaki redak očekuje 60 i 40.

Poglavlje 14 — Konceptualni zadatak

U prvoj situaciji osoba je jedinica neovisnosti, a dizajn je jednouzorački jer uspoređujemo sredinu jednog uzorka s fiksnim pragom od pet bodova. U drugoj su jedinice različite osobe u dvjema neovisnim skupinama prema glavnom izvoru vijesti. U trećoj su jedinice iste osobe mjerene prije i poslije, pa je dizajn uparen i analiziramo razlike unutar osoba.

Odgovor povezuje naziv postupka s nacrtom. Dva stupca ili dva prosjeka sama ne znače neovisne skupine. Kada bi usporedna vrijednost u prvoj situaciji bila procjena iz druge ankete, nosila bi vlastitu nesigurnost i ne bismo je smjeli tretirati kao fiksni prag.

Česta stranputica je sljedeći odgovor.

> Prva situacija ima dvije skupine jer uspoređuje uzorak i broj pet.

Broj pet nije druga skupina ni druga procjena, nego unaprijed zadana referentna vrijednost.

Poglavlje 14 — Kritički zadatak

Ocjena povjerenja opisuje odgovor na zadanu ljestvicu, a ne objektivnu istinitost medija. Glavni izvor vijesti ne obuhvaća svu izloženost i nije nasumično dodijeljen. U simuliranom uzorku razlika televizija minus društvene mreže iznosi približno 1,19 boda, uz Welchov 95-postotni interval od 0,45 do 1,92 boda. Interval u cijelosti leži iznad nule, pa podupire zaključak o pozitivnoj razlici. U cijelosti je i ispod dva boda, pa ne podupire zahtjev da razlika iznosi najmanje dva boda. Ni jedan od tih zaključaka ne utvrđuje učinak promjene izvora vijesti.

Bilješka daje koristan odgovor za oba unaprijed zadana kriterija te zasebno navodi granice mjerenja i nacrta. Oznaka statističke značajnosti ne bi odredila praktični prag niti opravdala uzročnu tvrdnju.

Česta stranputica je sljedeći odgovor.

> Rezultat je značajan, pa su obje odluke riješene u korist TV skupine.

Pragovi nisu isti. Interval isključuje nulu, ali ne doseže zahtjev od najmanje dva boda.

Poglavlje 15 — Konceptualni zadatak

Značajan ukupni test pokazuje da nisu sve populacijske sredine jednake, ali ne određuje koliko se parova ni koji parovi razlikuju. Globalna nulta tvrdnja odnosi se na sve sredine i može biti odbačena zbog jedne dovoljno velike razlike. Valjana skica ima tri skupine kod kojih je razmak krajnjih sredina velik prema raspršenosti, a oba susjedna razmaka manja od istodobne granice. Primjer sredina 0, 0,5 i 1 uz zajedničku granicu razlučivosti između 0,5 i 1 ostavlja razlučiv samo krajnji par.

Par odabran tek nakon pregleda sredina jest post-hoc, a obitelj koju treba zaštititi obuhvaća sve parove među kojima je mogao biti izabran. Za pet skupina puna obitelj ima 10 parova. Unaprijed nije postojao jedan izdvojeni par.

Odgovor razlikuje ukupnu tvrdnju od usporedbi parova i određuje obitelj usporedbi prema tome kada je par odabran. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Značajan F znači da je svaka skupna sredina različita od svake druge.

Ukupna hipoteza pada čim najmanje jedna sredina odstupa. F ne razvrstava svih deset parova.

Poglavlje 16 — Konceptualni zadatak

Zbirni nagib dijelom odražava činjenicu da se dobna struktura razlikuje među izvorima koji imaju različite razine povjerenja. Prosječna dob ide od 33,4297 godine među korisnicima društvenih mreža do 47,8674 među slušateljima radija, a skupine se razlikuju i po povjerenju. Zbirni nagib opisuje kako se prosjek mijenja s dobi u cijeloj populaciji, a prilagođeni koliko se mijenja s dobi među osobama jednakoga izvora. Prvi model sadrži samo dob, dok drugi uz dob drži kategoriju izvora jednakom.

Skupina A ima nagib 0,6, skupina B −0,2, a zbirni nagib 0,2 ne opisuje nijednu skupinu. Predviđene su linije 2 + 0,6x i 8 − 0,2x. Njihov uravnoteženi zbirni sažetak ima nagib 0,2.

Odgovor objašnjava kako se promjenom modela mijenja usporedba na istim podacima i zašto zbirni nagib može prikriti različite nagibe skupina. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Nagib se promijenio zato što su se promijenile dobi u podacima.

Redci i dobi ostali su isti. Promijenilo se pitanje modela dodavanjem izvora vijesti.

Poglavlje 17 — Konceptualni zadatak

Podacima iz skupa za učenje prilagođavamo model, pomoću skupa za provjeru biramo postupak i prag, a skup za ispitivanje čuvamo za jednu završnu procjenu. Ispitni ishodi ne smiju sudjelovati u učenju ni izboru modela ili praga. Ako rečenice istoga dokumenta rasporedimo u različite skupove, one mogu prenijeti temu, stil i kontekst između njih te uljepšati procjenu na navodno novom tekstu. Jedinica razdvajanja mora biti cijeli povezani dokument, ne pojedinačna rečenica.

Tablica zabune jest kontingencijska tablica koja križa odluku i zabilježeni referentni ishod. Četiri ćelije TP, FP, FN i TN nastaju križanjem dviju binarnih varijabli. FPR dijeli FP s FP + TN, a FNR dijeli FN s FN + TP. Prvi nazivnik obuhvaća sve zabilježeno nenegativne, a drugi sve zabilježeno negativne slučajeve.

Predviđena vjerojatnost postaje odluka tek nakon unaprijed obrazložena klasifikacijskog praga i pravila postupanja. Vrijednosti tik ispod i iznad praga mogu proizvesti različito postupanje iako su numerički vrlo bliske.

Odgovor razdvaja uloge skupova, poštuje dokument kao jedinicu razdvajanja, imenuje nazivnike stopa pogreške i razlikuje vjerojatnost od odluke. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

> Tri skupa služe tome da svaki sadržava približno jednak broj redaka.

Uloge skupova određuje trenutak uporabe, a ne jednaka veličina.

Poglavlje 18 — Konceptualni zadatak

Jedan redak predstavlja jednu odabranu i označenu parlamentarnu rečenicu, a dokument je jedinica grupiranoga razdvajanja. Paket ima 2 698 redaka iz 2 499 dokumenata i ne spaja govornike, stranke, mandate ni datume. Izvan podataka ostaju rečenice izvan izvornoga odabira, bez hrvatske oznake i uklonjeni preklopljeni redci za učenje. Dnevnik bilježi HR filtar te uklanjanje 25 redaka za učenje iz 20 dokumenata prisutnih u ispitnom izvoru.

Cilj je opisati brojnost zabilježenih oznaka i dva postupka kojima su nastale unutar pripremljenoga paketa. Brojnosti 1 212 negativnih, 979 neutralnih i 507 pozitivnih oznaka imaju nazivnik 2 698 odabranih redaka. Paket nema uzorački interval za parlamentarnu populaciju, a selekcija, kodiranje i valjanost oznake ostaju izvan jednostavnih udjela. Pripremljeni korpus nije vjerojatnosni uzorak svih rečenica, govornika ili stavova. Obranjiva alternativa razdvaja dva puta nastanka oznake umjesto da ih čita kao jedan mjerni postupak. Negativna je oznaka zabilježena u 47,9 % redaka s dvama koderima i usklađenjem te u 41,9 % redaka s jednim uvježbanim koderom.

Pogrešna generalizacija može etiketirati govor i autore izvan dokaza, a budući prag može nepotrebno opteretiti pregled ili propustiti relevantne rečenice. Paket nema zaključan klasifikator, prag, podskupine, nadzor ni put prigovora koji bi opravdali postavljenu odluku. Paket podupire opis oznaka i njihova puta u odabranim redcima, ali ne podupire tvrdnju o tonu svega parlamentarnog govora, učinku broja kodera ili opravdanosti sustava odluke. Putovnica i dnevnik određuju jedinicu i obuhvat, a provjera osjetljivosti ne izolira uzrok razlike.

Odgovor povezuje nazivnik s odabranim rečenicama, razlikuje postupke označavanja i ograničava zaključak na ono što paket doista podupire. Formulacija može biti drukčija ako ostaju isti račun i obrazloženje.

Česta stranputica je sljedeći odgovor.

Paket podupire tvrdnju da je 44,9 % hrvatskoga parlamentarnog govora negativno.

Nazivnik čini 2 698 odabranih redaka u pripremljenom paketu, a ne sav hrvatski parlamentarni govor.

literatura

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, i Matthew Gentzkow. 2020. „The Welfare Effects of Social Media". *American Economic Review* 110 (3): 629–76. https://doi.org/10.1257/aer.20190658.

Anscombe, F. J. 1973. „Graphs in Statistical Analysis". *The American Statistician* 27 (1): 17–21. https://doi.org/10.1080/00031305.1973.10478966.

Barocas, Solon, Moritz Hardt, i Arvind Narayanan. 2023. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. https://fairmlbook.org/.

Belia, Sarah, Fiona Fidler, Jennifer Williams, i Geoff Cumming. 2005. „Researchers Misunderstand Confidence Intervals and Standard Error Bars". *Psychological Methods* 10 (4): 389–96. https://doi.org/10.1037/1082-989X.10.4.389.

Bertrand, Marianne, i Sendhil Mullainathan. 2004. „Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination". *American Economic Review* 94 (4): 991–1013. https://doi.org/10.1257/0002828042002561.

Bickel, P. J., E. A. Hammel, i J. W. O'Connell. 1975. „Sex Bias in Graduate Admissions: Data from Berkeley". *Science* 187 (4175): 398–404. https://doi.org/10.1126/science.187.4175.398.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, i sur. 2012. „A 61-million-person experiment in social influence and political mobilization". *Nature* 489 (7415): 295–98. https://doi.org/10.1038/nature11421.

Breiman, Leo. 2001. „Statistical Modeling: The Two Cultures". *Statistical Science* 16 (3): 199–231. https://doi.org/10.1214/ss/1009213726.

Brown, Tom B. i sur. 2020. „Language Models Are Few-Shot Learners". https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, i sur. 2013. „Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience". *Nature Reviews Neuroscience* 14 (5): 365–76. https://doi.org/10.1038/nrn3475.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, i Emmanuel Saez. 2014. „Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States". *The Quarterly Journal of Economics* 129 (4): 1553–623. https://doi.org/10.1093/qje/qju022.

Chouldechova, Alexandra. 2017. „Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments". *Big Data* 5 (2): 153–63. https://doi.org/10.1089/big.2016.0047.

Cleveland, William S., i Robert McGill. 1984. „Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods". *Journal of the*

American Statistical Association 79 (387): 531–54. https://doi.org/10.1080/01621459.1984.10478080.

Cochran, William G. 1954. „Some Methods for Strengthening the Common χ^2 Tests“. *Biometrics* 10 (4): 417–51. https://doi.org/10.2307/3001616.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2. izd. Lawrence Erlbaum Associates.

Coles, Nicholas A., David S. March, Fernando Marmolejo-Ramos, i sur. 2022. „A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration“. *Nature Human Behaviour* 6 (12): 1731–42. https://doi.org/10.1038/s41562-022-01458-9.

Cumming, Geoff. 2014. „The New Statistics: Why and How“. *Psychological Science* 25 (1): 7–29. https://doi.org/10.1177/0956797613504966.

Državni zavod za statistiku. 2025. *Turistička aktivnost stanovništva Republike Hrvatske u 2024*. TUR-2025-4-1. Državni zavod za statistiku. https://podaci.dzs.hr/2025/hr/97089.

Državni zavod za statistiku. 2026. *Turizam*. PxWeb, tablice BS_TU11 i T03. https://web.dzs.hr/PxWeb/pxweb/hr/Turizam/.

Državno izborno povjerenstvo Republike Hrvatske. 2024. „Izvješće o provedenim izborima za zastupnike u Hrvatski sabor 2024.“ https://www.izbori.hr/site/UserDocsImages/2024/Izbori_za_zastupnike_u_Hrvatski_sabor/Rezultati/Izvje%C5%A1%C4%87e%200%20provedenim%20izborima%20za%20zastupnike%20u%20Hrvatski%20sabor%202024.pdf.

Efron, B. 1979. „Bootstrap Methods: Another Look at the Jackknife“. *The Annals of Statistics* 7 (1): 1–26. https://doi.org/10.1214/aos/1176344552.

Eurostat. 2026.a. „Early leavers from education and training by labour status“. https://doi.org/10.2908/EDAT_LFSE_14.

Eurostat. 2026.b. „Employment and labour force by sex and age - annual data“. https://doi.org/10.2908/LFSI_EMP_A.

Eurostat. 2026.c. „Individuals - internet use“. https://doi.org/10.2908/ISOC_CI_IFP_IU.

Eurostat. 2026.d. „Persons aged 25–34 with tertiary educational attainment level“. https://doi.org/10.2908/SDG_04_20.

Eurostat. 2026.e. „Persons at risk of poverty or social exclusion by age and sex“. https://doi.org/10.2908/ILC_PEPS01N.

Eurostat. 2026.f. „Population structure indicators at national level“. https://doi.org/10.2908/DEMO_PJANIND.

Gelman, Andrew, i Eric Loken. 2013. „The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No ‚Fishing Expedition‘ or ‚p-Hacking‘ and the Research Hypothesis Was Posited Ahead of Time“. https://sites.stat.columbia.edu/gelman/research/unpublished/forking.pdf.

Gilovich, Thomas, Robert Vallone, i Amos Tversky. 1985. „The Hot Hand in Basketball: On the Misperception of Random Sequences“. *Cognitive Psychology* 17 (3): 295–314. https://doi.org/10.1016/0010-0285(85)90010-6.

Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, i sur. 2016. „Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations“. *European Journal of Epidemiology* 31 (4): 337–50. https://doi.org/10.1007/s10654-016-0149-3.

Hoekstra, Rink, Richard D. Morey, Jeffrey N. Rouder, i Eric-Jan Wagenmakers. 2014. „Robust Misinterpretation of Confidence Intervals“. *Psychonomic Bulletin & Review* 21 (5): 1157–64. https://doi.org/10.3758/s13423-013-0572-3.

Karpicke, Jeffrey D., i Janell R. Blunt. 2011. „Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping“. *Science* 331 (6018): 772–75. https://doi.org/10.1126/science.1199327.

Kleppang, Annette Løvheim, Anne Mari Steigen, Li Ma, Hanne Søberg Finbråten, i Curt Hagquist. 2021. „Electronic Media Use and Symptoms of Depression among Adolescents in Norway“. *PLOS ONE* 16 (7): e0254197. https://doi.org/10.1371/journal.pone.0254197.

Matejka, Justin, i George Fitzmaurice. 2017. „Same Stats, Different Graphs“. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1290–94. https://doi.org/10.1145/3025453.3025912.

Mercer, Andrew, i Arnold Lau. 2023. *Comparing Two Types of Online Survey Samples*. Pew Research Center. https://www.pewresearch.org/methods/2023/09/07/comparing-two-types-of-online-survey-samples/.

Miller, Joshua B., i Adam Sanjurjo. 2018. „Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers“. *Econometrica* 86 (6): 2019–47. https://doi.org/10.3982/ECTA14943.

Mochtak, Michal, Peter Rupnik, i Nikola Ljubešić. 2024. „The ParlaSent Multilingual Training Dataset for Sentiment Identification in Parliamentary Proceedings“. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024) (Torino, Italia), svibanj, 16024–36. https://aclanthology.org/2024.lrec-main.1393/.

Mochtak, Michal, Peter Rupnik, Katja Meden, i Nikola Ljubešić. 2023. *The Multilingual Sentiment Dataset of Parliamentary Debates ParlaSent* 1.0. http://hdl.handle.net/11356/1868.

Nieuwenhuis, Sander, Birte U. Forstmann, i Eric-Jan Wagenmakers. 2011. „Erroneous Analyses of Interactions in Neuroscience: A Problem of Significance". *Nature Neuroscience* 14 (9): 1105–7. https://doi.org/10.1038/nn.2886.

OECD. 2023. *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. https://doi.org/10.1787/53f23881-en.

Ofqual. 2020. *Standardisation of grades in general qualifications in summer 2020: outliers*. Office of Qualifications; Examinations Regulation. https://www.gov.uk/government/publications/standardisation-of-grades-in-general-qualifications-in-summer-2020-outliers.

Open Science Collaboration. 2015. „Estimating the Reproducibility of Psychological Science". *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

Orben, Amy, i Andrew K. Przybylski. 2019. „The association between adolescent well-being and digital technology use". *Nature Human Behaviour* 3 (2): 173–82. https://doi.org/10.1038/s41562-018-0506-1.

Parry, Douglas A., Brittany I. Davidson, Craig J. R. Sewall, Jacob T. Fisher, Hannah Mieczkowski, i Daniel S. Quintana. 2021. „A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use". *Nature Human Behaviour* 5 (11): 1535–47. https://doi.org/10.1038/s41562-021-01117-5.

Pew Research Center. 2024. *Religious 'Nones' in America: Who They Are and What They Believe*. Pew Research Center. https:

//www.pewresearch.org/religion/wp-content/uploads/sites/7/2024/01/PR_2024.01.24_religious-nones_REPORT.pdf.

Shmueli, Galit. 2010. „To Explain or to Predict?“ *Statistical Science* 25 (3): 289–310. https://doi.org/10.1214/10-STS330.

Silberzahn, R., E. L. Uhlmann, D. P. Martin, i sur. 2018. „Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results“. *Advances in Methods and Practices in Psychological Science* 1 (3): 337–56. https://doi.org/10.1177/2515245917747646.

Simmons, Joseph P., Leif D. Nelson, i Uri Simonsohn. 2011. „False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant“. *Psychological Science* 22 (11): 1359–66. https://doi.org/10.1177/0956797611417632.

Simpson, E. H. 1951. „The Interpretation of Interaction in Contingency Tables“. *Journal of the Royal Statistical Society, Series B* 13 (2): 238–41. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x.

Squire, Peverill. 1988. „Why the 1936 Literary Digest Poll Failed“. *Public Opinion Quarterly* 52 (1): 125–33. https://doi.org/10.1086/269085.

Stevens, S. S. 1946. „On the Theory of Scales of Measurement“. *Science* 103 (2684): 677–80. https://doi.org/10.1126/science.103.2684.677.

Šikić, Luka i sur. 2026. „Prikaz i analiza katoličke tematike u digitalnom medijskom prostoru“. Hrvatsko katoličko sveučilište. https://github.com/lusiki/DigiKat/tree/278a127f9170c1aca82035a4a8357b8a995f91d8.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2. izd. Graphics Press.

Tukey, John W. 1977. *Exploratory Data Analysis*. Addison-Wesley.

Tversky, Amos, i Daniel Kahneman. 1973. „Availability: A Heuristic for Judging Frequency and Probability". *Cognitive Psychology* 5 (2): 207–32. https://doi.org/10.1016/0010-0285(73)90033-9.

Wagenmakers, E.-J., T. Beek, L. Dijkhoff, Q. F. Gronau, i sur. 2016. „Registered Replication Report: Strack, Martin, & Stepper (1988)". *Perspectives on Psychological Science* 11 (6): 917–28. https://doi.org/10.1177/1745691616674458.

Wasserstein, Ronald L., i Nicole A. Lazar. 2016. „The ASA's Statement on p-Values: Context, Process, and Purpose". *The American Statistician* 70 (2): 129–33. https://doi.org/10.1080/00031305.2016.1154108.

Westreich, Daniel, i Sander Greenland. 2013. „The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients". *American Journal of Epidemiology* 177 (4): 292–98. https://doi.org/10.1093/aje/kws412.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2. izd. Springer. https://doi.org/10.1007/978-3-319-24277-4.

Wilkinson, Leland. 2005. *The Grammar of Graphics*. 2. izd. Springer. https://doi.org/10.1007/0-387-28695-0.

KOLOFON

statistika 101: uvod u analizu podataka za društvene znanosti

Luka Šikić · Petra Palić

lusiki.github.io/statistika-knjiga

Predobjavni nacrt. Izdanje, inačica, datum i trajni identifikator nisu zamrznuti. Za navod nacrta treba zabilježiti točan Git commit i datum pristupa. Konačni bibliografski podaci bit će navedeni u objavljenom izdanju.