

Statistička analiza podataka u eksperimentalnim istraživanjima usvajanja drugoga jezika

Miličević Petrović, Maja; Kraš, Tihana

Authored book / Autorska knjiga

Publication status / Verzija rada: **Published version / Objavljena verzija rada (izdavačev PDF)**

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:186:177297>

Rights / Prava: [Attribution-NonCommercial-NoDerivatives 4.0 International/Imenovanje-Nekomercijalno-Bez prerada 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-12-04**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)





Maja Miličević Petrović
Tihana Kraš

Statistička analiza podataka u eksperimentalnim istraživanjima usvajanja drugoga jezika



Maja Miličević Petrović i Tihana Kraš

STATISTIČKA ANALIZA PODATAKA U EKSPERIMENTALNIM
ISTRAŽIVANJIMA USVAJANJA DRUGOGA JEZIKA

Izdavač

Sveučilište u Rijeci, Filozofski fakultet

Za izdavača

prof. dr. sc. Aleksandar Mijatović

Autorice

izv. prof. dr. sc. Maja Miličević Petrović

izv. prof. dr. sc. Tihana Kraš

Recenzentice

izv. prof. dr. sc. Petra Anić

izv. prof. dr. sc. Mirjana Božić

prof. dr. sc. Luna Filipović

prof. dr. sc. Višnja Pavičić Takač

Lektura i korektura

izv. prof. dr. sc. Mihaela Matešić

Ilustracija na naslovnoj stranici

Freepik

Grafička priprema i prijelom

Grafika Helvetica d.o.o. za Centar za elektroničko nakladništvo Sveučilišne knjižnice Rijeka

Tisak

Grafika Helvetica d.o.o.

Naklada

200 primjeraka

Mjesto i godina objavljivanja

Rijeka, 2024.

ISBN 978-953-361-121-1 (e-izdanje)

ISBN 978-953-361-120-4 (tiskano izdanje)

CIP zapis dostupan u računalnome katalogu Sveučilišne knjižnice Rijeka pod brojem 150724018.

Odlukom Senata Sveučilišta u Rijeci (KLASA: 007-01/24-03/02, URBROJ: 2170-137-01-24-205) od 25. lipnja 2024. ovo se djelo objavljuje kao izdanje Sveučilišta u Rijeci. Sveučilište u Rijeci pokriva trošak e-izdanja koje obavlja Centar za elektroničko nakladništvo Sveučilišne knjižnice Rijeka.

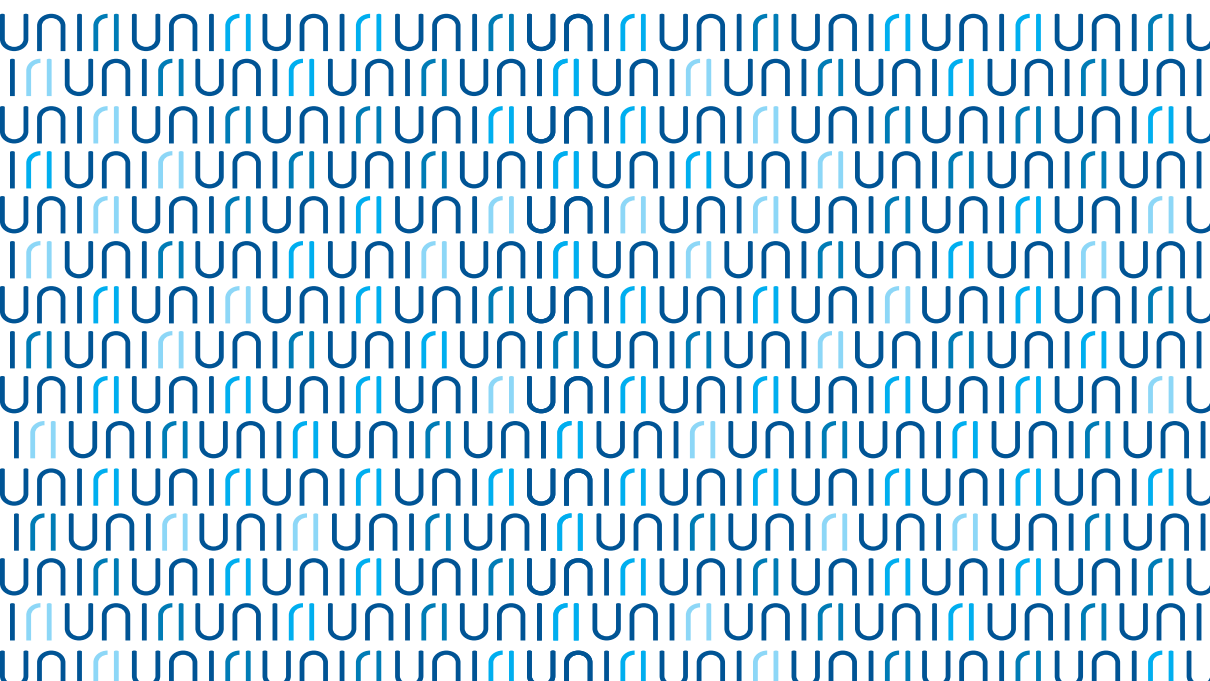
Ovu su knjigu sufinancirali Ministarstvo znanosti, obrazovanja i mladih Republike Hrvatske, Filozofski fakultet u Rijeci i Sveučilište u Rijeci projektom uniri-human-18-167.



Djelo je objavljeno pod uvjetima licence Creative Commons Autorstvo-Nekomercijalno-Bez prerada 4.0 Međunarodna (CC BY-NC-ND 4.0).

Maja Miličević Petrović i Tihana Kraš

Statistička analiza podataka u eksperimentalnim istraživanjima usvajanja drugoga jezika



UNIRI

FFRI Filozofski fakultet
Sveučilišta u Rijeci

Za tatu, koji bi se radovao ovoj knjizi više nego bilo ko drugi

M. M. P.

Tati, koji me naučio da nikad ne odustanem od ostvarenja svojih snova

T. K.

SADRŽAJ

Predgovor	XIII
1. Uloga statistike u eksperimentalnim istraživanjima usvajanja drugoga jezika	1
1.1 Što su eksperimentalna istraživanja usvajanja drugoga jezika?	1
1.2 Što je statistika?	4
1.3 Čemu služi statistika?	5
1.4 Što je potrebno za provedbu statističke analize?	8
1.5 Kako ovladati statistikom?	11
2. Koraci do statističke analize podataka	17
2.1 Osmišljavanje i priprema istraživanja	18
2.1.1 Od istraživačkih pitanja do hipoteza i varijabli	18
2.1.2 Mjerenje varijabli	22
2.1.3 Definiranje uzoraka ispitanika	28
2.1.4 Definiranje uzoraka podražaja	34
2.1.5 Izrada nacrtu istraživanja	41
2.1.6 Kontrola kvalitete istraživanja	50
2.2 Prikupljanje i priprema podataka	53
2.2.1 Eksperimentalni zadaci i podaci koji se u njima dobivaju	53
2.2.2 Kodiranje i organizacija podataka	55
2.2.3 Izbor programa za analizu	64
2.3 Provjera pouzdanosti instrumenata	70
3. Statističko opisivanje podataka	75
3.1 Izračunavanje deskriptivnih mjera	75
3.1.1 Frekvencijske mjere	77
3.1.2 Mjere centralne tendencije	82

3.1.3	Mjere varijabilnosti	87
3.1.4	Položajne mjere	92
3.1.5	Distribucija podataka	96
3.1.6	Dodatne operacije na podacima	104
3.2	Grafičko prikazivanje rezultata	111
3.2.1	Prikaz frekvencija kategorija	112
3.2.2	Prikaz mjera centralne tendencije i varijabilnosti	117
3.2.3	Prikaz povezanosti numeričkih varijabli	124
3.2.4	Drugi prikazi podataka i opće napomene	127
4.	Statističko zaključivanje	129
4.1	Osnove statističkoga zaključivanja	129
4.1.1	Testiranje hipoteza	130
4.1.2	Statistička značajnost i statistička snaga	133
4.1.3	Veličina efekta	140
4.1.4	Standardna pogreška i granice pouzdanosti	142
4.1.5	Analiza statističke snage testa	149
4.2	Statistički testovi	151
4.2.1	Zajedničke značajke statističkih testova	151
4.2.2	Primjer statističkoga testa	153
4.3	Kriteriji izbora testa	154
4.3.1	Nacrt istraživanja	156
4.3.2	Parametrijski uvjeti	158
5.	Često korišteni statistički testovi	163
5.1	Ispitivanje povezanosti između kategorijskih varijabli	166
5.1.1	Hi-kvadrat test	167
5.1.2	McNemarov test	173
5.2	Ispitivanje povezanosti između kategorijskih i numeričkih varijabli (dva uzorka)	175
5.2.1	T-testovi	177
5.2.2	Wilcoxonovi testovi	184
5.3	Ispitivanje povezanosti između kategorijskih i numeričkih varijabli (više od dva uzorka)	188
5.3.1	Analiza varijance	190
5.3.2	Kruskal-Wallisov i Friedmanov test	199
5.3.3	Analize prema ispitanicima i prema česticama	202

5.4 Ispitivanje povezanosti između numeričkih varijabli	204
5.4.1 Pearsonov koeficijent korelacije	208
5.4.2 Spearmanov koeficijent korelacije	210
5.5 Regresijska analiza i mješoviti modeli	212
5.6 Drugi testovi	217
6. Izvještavanje o eksperimentalnome istraživanju	221
6.1 Struktura izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika	222
6.2 Opće smjernice za izvještavanje o eksperimentalnome istraživanju	223
6.3 Uvodni dio izvještaja: odgovor na pitanje „zašto“?	225
6.4 Opis metode i postupka: odgovor na pitanje „kako“?	227
6.5 Opis rezultata: odgovor na pitanje „što“?	232
6.5.1 Rezultati deskriptivne statistike	233
6.5.2 Rezultati inferencijalne statistike	234
6.6 Rasprava o rezultatima i zaključci: odgovor na pitanje „pa što“?	236
7. Trendovi i perspektive u statističkoj analizi podataka	239
7.1 Dodatni aspekti kontrole kvalitete istraživanja	240
7.2 Metodološki zaokret ili teorijsko-metodološka revolucija?	245
7.3 Misli globalno, djeluj lokalno	248
Sažetak	251
Summary	253
Literatura	255
Kazalo pojmova	277
Kazalo autora	288
O autoricama	293

POPIS SLIKA

Slika 1. Ciljevi znanstvenih istraživanja i njihov odnos s vrstama statističke analize i teorijom	8
Slika 2. Tijek empirijskoga istraživačkog procesa	10
Slika 3. Analiza podataka i koraci koji joj prethode.	17
Slika 4. Shematski prikaz valjanosti i pouzdanosti (a – prisutnost objiju značajki, b – prisutnost samo pouzdanosti, c – prisutnost samo valjanosti, d – izostanak objiju značajki)	50
Slika 5. Primjeri podražaja iz istraživanja Kraš (2008b, 2016) te Miličević i Kraš (2017).	57
Slika 6. Dodatak za statističku analizu u programu Microsoft Excel	64
Slika 7. Dodatak za statističku analizu u programu LibreOffice Calc	65
Slika 8. <i>Web</i> -aplikacija VassarStats	66
Slika 9. Sučelje programa IBM SPSS	67
Slika 10. Sučelje programa R	68
Slika 11. Sučelje programa JASP	69
Slika 12. Sučelje programa Jamovi	69
Slika 13. Određivanje medijana u uzorku s neparnim brojem vrijednosti (a) i s parnim brojem vrijednosti (b).	85
Slika 14. Određivanje kvartila	94
Slika 15. Histogrami vremena reakcije (podaci prilagođeni iz Wang i Kaiser, 2022).	97
Slika 16. Histogram vremena reakcije s krivuljom gustoće (podaci prilagođeni iz Wang i Kaiser, 2022).	99
Slika 17. Poligon frekvencija vremena reakcije (podaci prilagođeni iz Wang i Kaiser, 2022).	100
Slika 18. Normalna distribucija	101
Slika 19. Primjeri normalnih distribucija s prosjekom 0 i različitim standardnim devijacijama	101
Slika 20. Primjeri distribucija koje odstupaju od normalne.	102
Slika 21. Grafikon kvantila (za podatke prilagođene iz Wang i Kaiser, 2022).	104
Slika 22. Kružni dijagram (prema Miličević, 2007)	113
Slika 23. Stupčasti grafikoni za prikaz frekvencija (prema Miličević, 2007) .	114
Slika 24. 100 % složeni stupčasti grafikon (prema Kraš, 2008c).	115
Slika 25. Mozaični grafikon (prema Robison, 1990)	116

Slika 26. Asocijativni grafikon (prema Robison, 1990)	117
Slika 27. Stupčasti grafikon za prikaz mjera centralne tendencije (prema Miličević, 2007).	118
Slika 28. Stupčasti grafikon s oznakama pogreške (prema Miličević, 2007)	119
Slika 29. Linijski grafikon (utemeljeni na podacima iz Jiang i sur., 2011)	120
Slika 30. Objašnjenje dijagrama s pravokutnikom	121
Slika 31. Dijagram s pravokutnikom, bez usjeka i s usjecima (prema Miličević, 2007).	122
Slika 32. Dijagram s pravokutnikom dopunjen prikazom pojedinačnih podataka (prema Miličević, 2007)	123
Slika 33. Piratski grafikon (prema Miličević, 2007).	124
Slika 34. Dijagram raspršenja (prema DeKeyser, 2000)	125
Slika 35. Dijagram raspršenja s dodanom Loessovom krivuljom (prema DeKeyser, 2000)	126
Slika 36. Primjer rezultata višestrukih bacanja novčića	134
Slika 37. Simulirane aritmetičke sredine i 95 %-tne granice pouzdanosti	145
Slika 38. Usporedba različitih crta koje označavaju intervale pogreške	147
Slika 39. Kritične regije binomne distribucije za $n=10$	154
Slika 40. Dijagram izbora statističkoga testa (utemeljen na Field, 2018)	155
Slika 41. Hi-kvadrat distribucija za nekoliko različitih stupnjeva slobode (df)	170
Slika 42. Hi-kvadrat distribucija s kritičnom regijom za $df=2$	170
Slika 43. T -distribucija za nekoliko različitih stupnjeva slobode (df)	180
Slika 44. Kritične regije t -distribucije ($df=15$).	180
Slika 45. Ukupna varijabilnost u situaciji s tri nezavisna uzorka	191
Slika 46. Varijabilnost između uzoraka u situaciji s tri nezavisna uzorka	192
Slika 47. Varijabilnost unutar uzoraka u situaciji s tri nezavisna uzorka	192
Slika 48. F -distribucija za nekoliko različitih stupnjeva slobode (df)	194
Slika 49. Varijabilnost između ispitanika u tri zavisna uzorka	196

POPIS TABLICA

Tablica 1. Usporedba mjernih ljestvica (prema Stevens, 1946)	26
Tablica 2. Primjer latinskoga kvadrata	48
Tablica 3. Primjer organizacije odgovora ispitanika (prema Miličević i Kraš, 2017).	58
Tablica 4. Prikaz podataka za zavisne uzorke u (a) širokome i (b) dugačkome formatu (prema Miličević i Kraš, 2017).	60
Tablica 5. Primjer kontingencijske tablice (prema Robison, 1990).	79
Tablica 6. Primjer zavisne kontingencijske tablice (prema Andringa, 2014)	80
Tablica 7. Primjer frekvencijske tablice sa zavisnim uzorcima i tri razine nezavisne varijable.	81
Tablica 8. Primjer prikaza različitih vrsta frekvencije	82
Tablica 9. Usporedba mjernih ljestvica prema dopuštenim deskriptivnim mjerama	95
Tablica 10. Pogreške u statističkome zaključivanju	138
Tablica 11. Prikaz podataka o dvjema kategorijskim varijablama	167
Tablica 12. Kontingencijska tablica opaženih frekvencija (prema Robison, 1990).	168
Tablica 13. Kontingencijska tablica teorijskih frekvencija (prema Robison, 1990).	168
Tablica 14. Dio kritičnih vrijednosti hi-kvadrata.	169
Tablica 15. Standardizirani reziduali za tablicu 12 (prema Robison, 1990)	171
Tablica 16. Kontingencijska tablica opaženih frekvencija (prema Andringa, 2014).	174
Tablica 17. Dio podataka iz istraživanja DeKeyser (2000, str. 525)	207

Predgovor

Ova je knjiga posvećena statističkoj analizi podataka prikupljenih u eksperimentalnim istraživanjima procesa usvajanja drugoga jezika. U knjizi se bavimo različitim fazama istraživanja važnima za analizu, od koraka koji prethode prikupljanju podataka i mogu se definirati kao priprema istraživanja, preko provjere podataka i mjernoga instrumenta te izračunavanja osnovnih mjera poput frekvencije i aritmetičke sredine, do provedbe statističkih testova, interpretacije njihovih rezultata i izvještavanja o njima. S tematske je strane naglasak na lingvističkim pristupima usvajanju drugoga jezika, odnosno istraživanjima koja za cilj imaju ispitivanje znanja neke jezične pojave, iako se velik dio sadržaja može primijeniti i na druge pristupe unutar istoga područja, pa i šire. Motivi za usmjeravanje na lingvističke pristupe bili su ponajprije naši osobni interesi i naša želja da se ova knjiga poveže s prethodnom zajedničkom knjigom istoga usmjerenja, *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika* (Kraš i Miličević, 2015). U prvotnim je planovima ta knjiga trebala dobiti drugo izdanje dopunjeno poglavljem posvećenim statističkoj analizi podataka. Međutim razdoblje od izlaska prve knjige naovamo bilo je toliko ispunjeno novom metodološkom i statističkom literaturom i u tolikoj su se mjeri počele uočavati promjene u načinu na koji se pristupa analizi podataka da jedno poglavlje nikako nije moglo biti dovoljno da bi to obuhvatilo.

Nova je knjiga dakle nastala kao odgovor na potrebu, uočenu i u literaturi i u praksi, za izvorima vezanima uz statističku analizu podataka u području usvajanja drugoga jezika. Kad je riječ o praksi, obje već dugi niz godina držimo sveučilišne kolegije i seminare, uživo i *online*, iz usvajanja drugoga jezika i srodnih područja te mentoriramo studente pri izradi završnih radova na različitim razinama studija, od prijediplomske do poslijediplomske, pri čemu redovito primjećujemo kod studenata potrebu za razrađenijom i temeljitijom metodološkom obukom u pogledu svih koraka provedbe istraživanja – od izrade nacrtu istraživanja do analize podataka, uz posebice prisutan strah od statistike. Uz to se zahtjevi vezani uz obuku za statističku analizu mogu često čuti i od kolega, kojima sve proširenija uporaba kvantitativnih metoda analize nameće potrebu da ih savladaju, a budući da se s njima obično nisu susreli tijekom studija, doživljavaju ih kao vrlo velik izazov.

Na takva se osobna zapažanja neposredno nadovezuje i to da je u trenutku nastanka knjige u međunarodnim krugovima istraživača usvajanja drugoga jezika vrlo naglašen trend objavljivanja članaka, uredničkih knjiga i monografija

posvećenih metodologiji istraživanja, a posebice statističkoj analizi. Taj trend proizlazi iz višegodišnje svijesti o metodološkim nedostacima u području i želje da se ti nedostaci uklone. Istraživačima usvajanja drugoga jezika dugo se zamjerala činjenica da, za razliku od istraživača u srodnim disciplinama, poput psihologije, ne prate aktualne trendove u analizi podataka i primijenjenoj statistici te da uporabom neprikladnih statističkih testova ponekad dolaze do nepouzdatih rezultata. Na te se zamjerke u međunarodnim krugovima sad vrlo intenzivno odgovara velikim brojem metaanaliza, statističkih priručnika i smjernica za provedbu statističke analize i izvještavanje o njoj. Međutim u znanstvenim sredinama koje su nama geografski najbliže – u Hrvatskoj i Srbiji¹ – istraživanja usvajanja drugoga jezika manje su proširena, pa se metodološki nedostaci i dalje ponekad očituju u kvantitativnim istraživanjima bez adekvatnoga nacrtu istraživanja ili onima u kojima se navode samo deskriptivne statističke mjere poput postotaka ili prosječnih vrijednosti, dok izostaju elementi koji bi omogućili uopćavanje rezultata izvan prikupljenih uzoraka.

Za osjetnije poboljšanje takve situacije potrebno je sustavno rješenje za metodološko obrazovanje na sveučilišnim studijima, koje je u ovome trenutku još vrlo različito od studija do studija i uglavnom prisutno na višim razinama studija (diplomskoj i poslijediplomskoj). Takvo rješenje zahtijeva vrijeme, te naša knjiga u tome kontekstu pokušava djelomično ispuniti postojeću prazninu, zajedno s općijim publikacijama poput udžbenika *Znanstvena istraživanja jezika i prevođenja* Nataše Pavlović i Mateusza-Milana Stanojevića, kao i *online-kolegijima* *Introduction to research methodology in linguistics*, *Introduction to experimental methods in linguistics* i *Introduction to statistics for language data* (Miličević, 2017a, 2017b, 2017c; usprkos naslovima sadržaj kolegija napisan je na srpskome jeziku). Koliko je nama poznato, na hrvatskome jeziku ili srpskome jeziku nije dosad objavljen nijedan udžbenik ili znanstvena monografija posvećen(a) statističkoj analizi u usvajanju drugoga jezika ili nekoj drugoj lingvističkoj disciplini. Studenti i istraživači jezika koji ne mogu ili ne žele upotrebljavati literaturu na engleskome jeziku oslanjaju se stoga najčešće na knjige i radove iz područja psihologije. To rješenje ne treba smatrati lošim jer su osnove statističke analize zajedničke različitim znanostima, od lingvistike do psihologije, od ekonomije do medicine. Pritom su postojeći udžbenici i priručnici vrlo kvalitetni, poput niza izdanja *Petzove statistike*, odnosno udžbenika *Osnovne statističke metode za nematematičare*

¹ Prva je autorica do 2020. radila na Sveučilištu u Beogradu, nakon čega je prešla na Sveučilište u Bologni, u Italiji, ali i dalje surađuje s kolegama u Srbiji i Hrvatskoj.

Borisa Petza, u Hrvatskoj ili knjiga *Metodologija psiholoških istraživanja* Dejana Todorovića i *Autostoperski vodič kroz statistiku* Olivera Toškovića u Srbiji. Međutim svako područje ima i svoje specifičnosti, a metode analize uz to se u posljednje vrijeme vrlo brzo mijenjaju pa se knjiga o statistici posvećena upravo usvajanju drugoga jezika čini vrlo potrebnom.

Imajući u vidu sve navedeno, cilj je naše knjige pružiti istraživačima, a posebice istraživačima-početnicima, uvid u osnovna načela statističke analize podataka o procesu usvajanja drugoga jezika, što bi im s jedne strane trebalo omogućiti ili olakšati kritičko čitanje postojeće literature, a s druge im strane pomoći u provedbi vlastitih istraživanja. Knjiga je napisana kao znanstvena monografija jer sadrži temeljitu analizu i sintezu metodološke situacije u području usvajanja drugoga jezika popraćenu mnogobrojnim primjerima iz objavljenih istraživanja te kritičkim osvrtom na postojeće prakse. Prema potrebi, može se koristiti i kao udžbenik, priručnik ili dopunska literatura na metodološki orijentiranim kolegijima vezanima uz usvajanje drugoga jezika, ali i druge lingvističke discipline. Knjigom se indirektno zauzimamo za to da se proučavanje usvajanja drugoga jezika i prateće metode analize podataka sustavnije uvedu kao kolegiji na razini ne samo diplomskoga i doktorskoga već i prijediplomskoga studija.

Knjiga je organizirana u sedam poglavlja. Prvo poglavlje govori o ulozi i važnosti koju statistika ima za područje usvajanja drugoga jezika. Drugo je poglavlje posvećeno pripremnoj fazi istraživanja i prvim koracima u analizi podataka, gdje se govori o osnovnim komponentama istraživanja poput varijabli, nacrti i uzorkovanja te o unosu podataka u programe za statističku analizu i provjeri pouzdanosti korištenih mjernih instrumenata. Slijedi treće poglavlje, posvećeno statističkomu opisivanju podataka, tj. izračunavanju mjera frekvencije, centralne tendencije i varijabilnosti te prikazu podataka na različitim vrstama grafikona. Četvrto poglavlje prikazuje analitičku paradigmu utemeljenu na testiranju statističke značajnosti, uz objašnjenja kako odabrati odgovarajući statistički test, kako protumačiti njegove rezultate i koje dodatne mjere izračunati kako bi se rezultati potpunije razumjeli; vrlo je važna komponenta toga poglavlja i kritički osvrt na tu paradigmu. Peto poglavlje sadrži opis nekih često korištenih statističkih testova, uz navođenje kad se oni koriste, kako se izračunavaju i kakve podatke daju. Šesto poglavlje posvećeno je izvještavanju o eksperimentalnome istraživanju usvajanja drugoga jezika. Naposljetku, sedmo poglavlje sadrži završna razmišljanja o mogućem daljnjem metodološkom razvoju lingvističkih disciplina, pa tako i usvajanja drugoga jezika, posebice s obzirom na statističku analizu podataka i u kontekstu kritika testiranja statističke značajnosti.

Čitatelji prema svojim potrebama mogu odlučiti hoće li knjigu čitati u cjelini (što preporučujemo početnicima u statističkoj analizi) ili će konzultirati njezine pojedine dijelove. Naše je mišljenje da je najbolji izbor čitanje „u krug“, tj. vraćanje na različite dijelove knjige nakon što se ona najprije pročita u cijelosti. U knjizi je naglasak stavljen na objašnjavanje i oprimjeravanje načela i pojmova važnih za statističku analizu podataka. Ona stoga ne uključuje zadatke ni upute za provedbu analize u nekome od statističkih programa. Organizacijom središnjega dijela ona djelomično podsjeća na ono što Sønning i Werner (2021) nazivaju „normativnim priručnikom“, koji daje stroge upute za odabir statističkih testova. Međutim naglašavamo i ovdje kao i u glavnome dijelu knjige da najvažnijim smatramo temeljito razumijevanje onoga što nam testovi mogu i ne mogu reći o rezultatima istraživanja. Statistički testovi nisu univerzalno rješenje koje treba mehanički provoditi samo zato što nam je netko rekao da to trebamo činiti, već metodološko sredstvo čiju uporabu trebamo pažljivo procijeniti u kontekstu teme koju istražujemo i teorijskoga okvira koji pretpostavljamo. Uz to smatramo da su koraci u istraživanju koji prethode analizi podataka (od postavljanja istraživačkih hipoteza do različitih aspekata pripreme podataka za analizu) iznimno važni za kvalitetnu analizu podataka, pa njima u knjizi posvećujemo pozornost više nego što je to u knjigama o statistici uobičajeno.

Od drugih značajki knjige koje treba istaknuti, napominjemo da smo u knjizi prilikom uvođenja većine termina, koji su pri prvoj uporabi otisnuti masnim slovima, navodile ekvivalent termina na engleskome jeziku kako bi čitatelji mogli uspostaviti poveznice između naše knjige i literature na engleskome jeziku. Nadalje, napominjemo da se u obraćanju „čitateljima“ i u referiranju na „istraživače“ i „autore“ služimo samo množinom muškoga roda smatrajući je inkluzivnom. Također svraćamo pozornost na to da knjiga zbog povezanosti tema ima određenih sličnosti s našom prethodnom knjigom spomenutom na početku ovoga Predgovora. Takva je sličnost bila neizbježna kako bi mogao biti objašnjen cijeli tijek statističke analize podataka, koja počinje mnogo prije nego što se podaci prikupe. Međutim, svi su sadržaji temeljito revidirani i dopunjeni novim temama, novim spoznajama i novim primjerima. Uz to se u tekstu eksplicitno pozivamo na prethodnu knjigu na mjestima sadržajnih preklapanja. Dijagrami i grafikoni uključeni u knjigu autorsko su djelo svugdje gdje nije napomenuto drugačije.ⁱⁱ

ⁱⁱ Slike 1–4, 13–14, 40, 45–47 i 49 izrađene su u programu Microsoft PowerPoint; slike 6–12 izrađene su kao snimke zaslona; grafikoni na slikama 15–29, 31, 33–36, 39, 41–44 i 48 izrađeni su u programu R.

Konačno, želimo izraziti zahvalnost svim institucijama i pojedincima koji su, na različite načine, pomogli da ova knjiga ugleda svjetlo dana. Knjiga je nastala u okviru znanstvenoistraživačkoga projekta Sveučilišta u Rijeci *Obrada rečenica u talijanskom jeziku*, u kojemu smo obje sudjelovale, a koji je vodila druga autorica. Uz sredstva toga projekta, tiskanje knjige financijski su poduprli Ministarstvo znanosti, obrazovanja i mladih Republike Hrvatske i Filozofski fakultet u Rijeci, posebice njegov Odsjek za anglistiku, na čemu smo zahvalne. Sa sadržajne strane veliku je ulogu u nastanku knjige imao interes naših studenata i kolega, posebice onih okupljenih oko mreže ReLDI, najprije kao projekta *Regional Linguistic Data Initiative*, a zatim udruženja *ReLDI centar za jezičke podatke*; kontakti s njima neiscrpan su izvor energije za bavljenje prijenosom metodoloških znanja u našoj „regionalnoj“ zajednici, a redovita pitanja nekih od njih o tome kad će knjiga biti gotova značajno su pridonijela njezinu završavanju. Posebnu zahvalnost dugujemo recenzenticama, koje su temeljito pročitale rukopis i dale nam vrlo korisne sugestije za njegovo poboljšanje, te lektorici, koja je uvelike unaprijedila tekst i s bezgraničnim strpljenjem otklanjala naše jezične, stilske i terminološke nedoumice. Velika hvala i predsjednici Povjerenstva za izdavačku djelatnost Filozofskoga fakulteta u Rijeci izv. prof. dr. sc. Barbari Kalebić Maglica na susretljivosti, fleksibilnosti i korisnim uputama. Uz sve navedeno, ističemo da ove knjige ne bi bilo bez velike podrške članova naših obitelji i prijatelja, koji su nas poticali, strpljivo čekali, pomagali nam na razne načine, pa čak i čitali rukopis i obogatili ga svojom perspektivom, različitom od naše.

Forlì/Rijeka, svibanj 2024.

1. Uloga statistike u eksperimentalnim istraživanjima usvajanja drugoga jezika

Usvajanje drugoga jezika (engl. *second language acquisition*) disciplina je suvremene lingvistike koja proučava prirodu znanja nematerinskoga jezika i procese kojima se to znanje stječe (Kraš i Miličević, 2015, str. 1). To je izrazito empirijska disciplina, u kojoj se do novih spoznaja dolazi na temelju podataka prikupljenih od neizvornih govornika, odnosno učenika drugoga jezika. Podaci se prikupljaju različitim metodama, ovisno o teorijskome pristupu, predmetu istraživanja, istraživačkim pitanjima, raspoloživim resursima i drugim čimbenicima. Eksperimentalne metode jedna su vrsta metoda koje se upotrebljavaju, a kako su usmjerene na testiranje unaprijed postavljenih hipoteza i na prikupljanje veće količine podataka, u njima važnu ulogu ima **statistička analiza** (engl. *statistical analysis*). U potpoglavljima koja slijede objašnjavamo što su eksperimentalna istraživanja usvajanja drugoga jezika, što je statistika, čemu ona služi te što je potrebno za provedbu statističke analize. Nadalje, objašnjavamo ulogu statistike u eksperimentalnim istraživanjima usvajanja drugoga jezika te dajemo prijedloge kako istraživači iz toga područja mogu (bolje) ovladati statistikom.

1.1 Što su eksperimentalna istraživanja usvajanja drugoga jezika?

U usvajanju drugoga jezika **eksperiment** (engl. *experiment*) je moguće definirati kao istraživanje u kojemu sudjeluju **ispitanici** ili **sudionici** (engl. *subjects, participants*) koji izvršavaju neki **jezični zadatak** (engl. *language task*). U lingvistički orijentiranim eksperimentima, u kojima se naglasak stavlja na proučavanje i opis sustavnoga znanja drugoga jezika i njegovih pravila na različitim razinama jezične strukture, jezični zadatak može ispitivati proizvodnju na drugome jeziku, ali i razumijevanje rečenica i tekstova ili prosuđivanje rečenica prema tome koliko su prihvatljive (detaljan pregled zadataka može se naći u Kraš i Miličević (2015; v. i potpoglavlje 2.2.1 ove knjige te Kraš i sur., 2023). Uobičajeno je da zadatak čini niz pojedinačnih **podražaja** (engl. *stimuli*) na koje ispitanici odgovaraju – takvi podražaji još se nazivaju **česticama** (engl. *items*). Naprimjer u Miličević (2007) ispitanici su u zadatku prosudbe slike ocjenjivali prihvatljivost rečenica u srpskome jeziku s različito obilježenim povratnim i obostrano-povratnim značenjem (*se : sebe : jedan drugog*) uz različite glagole; jedan primjer (negramatične)

eksperimentalne čestice bio je **Čarobnjak i veštica su se otrovali jedno drugo*. Izbor zadatka i čestica ovisi o predmetu istraživanja, istraživačkim pitanjima, značajkama ispitanika, raspoloživim resursima i drugim čimbenicima.

U pojedinim se slučajevima eksperimentalna istraživanja usvajanja drugoga jezika temelje na **eksperimentalnome tretmanu** (engl. *experimental treatment*). Takva su naprimjer istraživanja koja ispituju učinkovitost različitih strategija učenja vokabulara (npr. Webb i sur., 2013) ili različitih tipova korektivne povratne informacije (tj. ispravljanja pogrešaka: npr. Nguyen i sur., 2017); u tim se istraživanjima mjeri postignuće učenika prije i poslije njihova izlaganja tretmanu.

Eksperimenti predstavljaju tek jednu od dostupnih metoda prikupljanja podataka o lingvističkim aspektima usvajanja drugoga jezika. Primjenjuju se i opažanje, naprimjer putem učeničkih korpusa,¹ i rjeđe introspekcija (o odnosu eksperimentalnih i drugih metoda koje se primjenjuju u istraživanjima usvajanja drugoga jezika v. Chaudron, 2003, str. 764).² Općenito govoreći, jedan od ključnih kriterija za razlikovanje eksperimenata od drugih metoda prikupljanja jezičnih podataka jest razina kontrole postupka. Naime, dok se jezične pojave opažanjem proučavaju onako kako se spontano odvijaju, pri primjeni eksperimentalnih metoda istraživači imaju **visoku razinu kontrole** nad podacima. Točnije, istraživači u eksperimentima mogu mijenjati čimbenike koji ih zanimaju te kontrolirati one koji za istraživanje nisu relevantni, čime usmjeravaju istraživanje tako da dobivaju upravo one podatke koji su im potrebni. Kod jezičnih je istraživanja to posebice važno kod rjeđih pojava, poput pitanja s više upitnih zamjenica (npr. *Tko je kome što poklonio?*) – u korpusu može takvih pitanja biti nedovoljno da bi se temeljitije istražila njihova uporaba, dok se u eksperimentu zadatak može postaviti tako da ispitanici moraju upotrijebiti točno određene upitne zamjenice. U najkontroliranijem obliku eksperimenta istraživači izravno utječu na istraživane čimbenike i izazivaju pojave koje ih zanimaju. Naprimjer kako bi utvrdili ima li korektivna povratna informacija pozitivan učinak na usvajanje pitanja s više upitnih zamjenica u drugome jeziku, istraživači mogu formirati dvije skupine ispitanika i jednu izložiti korektivnoj povratnoj informaciji, a drugu ne.

¹ Opažanje podrazumijeva proučavanje jezične proizvodnje u njezinu spontanom obliku, u učionici ili izvan nje. Danas se u lingvistici, pa i području usvajanja drugoga jezika, opažanje najčešće provodi uporabom digitaliziranih skupina tekstova koji su nastali (relativno) spontano – jezičnih korpusa.

² Često korištenu metodu u istraživanjima usvajanja drugoga jezika predstavljaju i upitnici, ali njihova je uporaba vezana uz stavove i uz individualne čimbenike poput motivacije i straha od jezika, odnosno čimbenike koji nisu lingvističke prirode.

Eksperiment je zahvaljujući visokoj razini kontrole jedini pouzdan način utvrđivanja **uzročno-posljedičnih odnosa** između pojava, posebice u slučaju postojanja eksperimentalnoga tretmana. Treba međutim istaknuti da u društvenim znanostima, pa tako ni u usvajanju drugoga jezika, često nije moguće postići razinu kontrole jednaku onoj u prirodnim znanostima. Društvene je čimbenike zbog njihove složenosti teško održavati konstantnima, a u usvajanju drugoga jezika postoje i velike individualne razlike između učenika, kao i niz važnih čimbenika na koje nije moguće izravno utjecati (npr. materinski jezik učenika može se samo odabrati, a ne i promijeniti).³ Iako zbog toga treba biti oprezan prilikom tumačenja uzročno-posljedičnih odnosa (v. i Tošković, 2020, str. 65–72, za objašnjenje drugih mogućih odnosa – povezanosti, predikcije i ovisnosti), eksperimenti se u području usvajanja drugoga jezika uglavnom analiziraju jednako eksperimentima u drugim znanostima.

Prikupljanje podataka u eksperimentima najčešće je usmjereno ka testiranju određene **hipoteze** (engl. *hypothesis*) odnosno pretpostavke o vrijednostima nekoga čimbenika ili o odnosu između više čimbenika značajnih za danu pojavu (v. potpoglavlje 2.1.1, kao i Kraš i Miličević, 2015, str. 16–17; Wallis, 2021, str. 29). Naprimjer možemo pretpostaviti da u engleskome kao drugome jeziku izvorni govornici talijanskoga manje izostavljaju članove od izvornih govornika hrvatskoga (zbog razlika u značajkama materinskoga jezika) te tu pretpostavku provjeriti odgovarajućim eksperimentalnim zadatkom koji zahtijeva uporabu članova. Testiranje hipoteze prisutno je uz to i u pojedinim situacijama kad je ona manje precizna ili čak neiskazana. Naprimjer moguće je eksperimentalno usporediti učinkovitost dviju različitih strategija ispravljanja pogrešaka u uporabi članova izlažući jednu skupinu učenika poticanju (engl. *prompts*), a drugu preinačavanju (engl. *recasts*) čak i kada nemamo unaprijed postavljenu hipotezu o tome koja će strategija dati bolje rezultate. To je zato što će u tome slučaju ipak postojati pretpostavka da neka razlika postoji te će eksperiment biti usmjeren ka testiranju te pretpostavke. U svakome slučaju, hipoteze vode tome da se podaci u eksperimentu prikupljaju ciljano, tako da se odnose na odabrane čimbenike, dok se drugi čimbenici isključuju iz istraživanja ili kontroliraju (u našim se primjerima tako ne razmatra spol učenika, koji se, vjerojatno, može zanemariti, kao ni razina općega znanja drugoga jezika, koju bi se moralo kontrolirati tako da su učenici iz dviju skupina izjednačeni prema općemu znanju).

³ v. Todorović (2008, str. 48–49) te Kraš i Miličević (2015, str. 21), gdje se govori o razlici između „manipulativnih“ i „selektivnih“ varijabli, odnosno čimbenika, s obzirom na to određuju li istraživači sami vrijednosti ili ih biraju među prirodno zadanima.

Eksperimenti u pravilu predstavljaju **kvantitativna istraživanja** (engl. *quantitative research*) i oslanjaju se na neki vid mjerenja ili prebrojavanja, odnosno na prikupljanje i obradu numeričkih podataka. Prema svemu navedenom, oni pripadaju i široj **kvantitativnoj znanstvenoj paradigmi** (engl. *quantitative research paradigm*), koju odlikuje kontrolirano mjerenje pojava, objektivnost, usmjerenost pozornosti k ishodu istraživanja, mogućnost ponavljanja postupka, mogućnost uopćavanja rezultata, provjera unaprijed postavljenih hipoteza, pretežito deduktivni pristup te utemeljenost na statističkome zaključivanju (v. sljedeće potpoglavlje, kao i četvrto poglavlje, u cijelosti posvećeno statističkomu zaključivanju).⁴ Eksperimentalnim istraživanjima usvajanja drugoga jezika najbližija su istraživanja iz područja psihologije, posebice kognitivne, s kojima djelomično dijele metode, a u još većoj mjeri postupke u statističkoj analizi.

1.2 Što je statistika?

Kako navode Petz i sur. (2012, str. 1), **statistika** (engl. *statistics*) je „obrada brojčanih podataka radi jasnijeg prikazivanja“. Primjeri su brojčanih ili numeričkih podataka koji ulaze u područje statistike pojedinačan broj bodova u testu razine općega znanja drugoga jezika, prosječan broj bodova svih osoba koje su rješavale test, prosječan broj bodova jedne osobe koja je rješavala skupinu testova, postotak osoba koje su riješile test s određenom točnošću i slični podaci. Termin je prvi put upotrijebljen u njemačkome jeziku, gdje je imenica *Statistik* označavala analizu podataka o državi (od novolatinskoga *statisticum* ‘državni’).

Statistika je ujedno znanstvena disciplina koja se bavi prikupljanjem, organizacijom, analizom, predstavljanjem i interpretacijom podataka, pri čemu razlikujemo više vrsta statistike. **Teorijska** ili **matematička statistika** (engl. *theoretical/mathematical statistics*) grana je matematike koja se bavi izvođenjem statističkih formula, teorema i zakona. Unutar nje posebno su važne teorija statistike i teorija vjerojatnosti. **Primijenjena statistika** (engl. *applied statistics*) primjena je statističkih formula, teorema i zakona u rješavanju problema unutar neke znanstvene discipline izvan matematike, poput ekonomije, sociologije, psihologije, lingvistike ili u svakodnevnome životu.

⁴ S druge strane, značajke su **kvalitativnih istraživanja** (engl. *qualitative research*) opažanje u prirodnim uvjetima, određena doza subjektivnosti, usmjerenost pozornosti k procesu istraživanja, teže ostvarivo ponavljanje postupka, nemogućnost uopćavanja rezultata, izostanak pretpostavki formuliranih unaprijed, pretežito induktivni pristup te utemeljenost na tumačenju podataka.

Primijenjena statistika ima više uloga u analizi podataka s obzirom na to odnosi li se na opis podataka koji su nam dostupni ili uključuje i procjenu podataka u koje nemamo uvid. Taj se kriterij često poklapa s time odnosi li se statistička analiza na prikupljeni uzorak ili na cijelu populaciju o kojoj istraživači žele nešto doznati putem uzorka (o pojmovima uzorka i populacije detaljnije govorimo u potpoglavlju 2.1.3). Naime, važna je grana primijenjene statistike **deskriptivna statistika** (engl. *descriptive statistics*), koja opisuje podatke koji su nam dostupni putem numeričkih pokazatelja poput aritmetičke sredine ili frekvencije. Taj se opis može odnositi ili na cijelu populaciju ili na uzorak na koji se istraživanje koncentrira. Kad se odnosi na populaciju, statistička se analiza može završiti na opisu jer odmah znamo sve što nam je potrebno o skupini koju proučavamo (naprimjer ako populaciju predstavljaju svi učenici koji su polagali državnu maturu u nekome roku, njihov prosječan broj bodova iz matematike predstavljat će statistički podatak koji opisuje cijelu populaciju). Kad se analiziraju uzorci, što je u znanstvenim istraživanjima znatno češći slučaj, deskriptivna statistika pruža uvid u stanje u danome uzorku, ali ne omogućava da se zaključci uopće izvan njega. Naprimjer prosječan broj bodova iz matematike podskupine od 100 učenika koji su polagali državnu maturu predstavlja uzorački statistički podatak, iz kojega ne možemo izravno zaključiti da bi rezultat trebao biti sličan i u nekoj drugoj sličnoj podskupini te da takav rezultat opisuje cijelu populaciju. Nasuprot tome **inferencijalna statistika** ili **statistika zaključivanja** (engl. *inferential statistics*), koja se naziva još i **analitička statistika** (engl. *analytical statistics*), temelji se na statističkim testovima koji omogućavaju procjenu podataka koji nam nisu dostupni te uopćavanje rezultata i izvan uzorka, na cijelu populaciju (v. npr. Tošković, 2020, str. 10). Ne treba zaboraviti da uzoračka deskriptivna statistika i ovdje ima jako važnu ulogu u analizi jer inferencijalni testovi počivaju na njoj – značajke populacije procjenjuju se na temelju podataka dobivenih na uzorku. Inferencijalna statistika ima veliku važnost u znanosti zbog toga što su znanstvenici uglavnom zainteresirani za opća znanja i za populacije, ali su vrlo rijetko u mogućnosti ispitati ih u cijelosti, pa se najčešće moraju osloniti na uzorke – ova knjiga govori o takvim situacijama.

1.3 Čemu služi statistika?

U empirijskim istraživanjima općenito, pa tako i u eksperimentalnim istraživanjima usvajanja drugoga jezika, statistička analiza nije sama sebi svrhom i važno je razumjeti zbog čega je potrebna te kakav doprinos daje istraživanju. Na početku je još važno podsjetiti da jezik općenito, a posebice

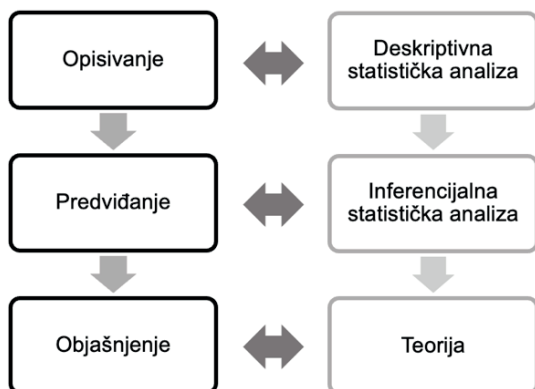
procesu vezani uz njegovo usvajanje, pripadaju pojavama koje se mogu opisati kao djelomično kaotične. Za razliku od velikoga broja situacija u fizici, gdje uvrštavanje poznatih vrijednosti uvijek daje točnu nepoznatu vrijednost, ne možemo lako pronaći formule kojima ćemo opisati jezične pojave i pojave vezane uz usvajanje (drugoga) jezika. Drugim riječima, dok će umnožak mase tijela i njegova ubrzanja uvijek dati vrijednost sile koja djeluje na tijelo (prisjetimo se drugoga Newtonova zakona i formule $F=ma$), u usvajanju drugoga jezika ne možemo biti sigurni da ćemo kod učenika koji imaju isti materinski jezik i na istoj su razini općega znanja drugoga jezika, kad ih izložimo jezičnome unosu (naprimjer vezanome uz članove u engleskome jeziku), uvijek dobiti isti rezultat u pogledu usvojenosti određene pojave (tj. članova, u navedenome primjeru). Čak i ako dodamo još čimbenika, poput dobi ili razine motivacije, gotovo je sigurno da nećemo doći do formule koja savršeno opisuje ishod. U takvim se slučajevima javlja potreba za statističkom analizom, koja je također zasnovana na formulama, ali formulama koje sadrže i elemente koje ne možemo potpuno precizno predvidjeti i koje počivaju na ideji što boljšega opisa pojave kroz čimbenike koji na nju utječu, ali uz neizbježan šum, koji ne možemo objasniti.

Statistička nam analiza pomaže tako što **modelira** pojave koje istražujemo. Dobiveni model nikad nije savršen i opća bi se formula mogla prikazati kao **ishod = model + pogreška**. Naprimjer mogli bismo u istraživanju dobiti da razina usvojenosti članova u engleskome jeziku (ishod) ovisi o materinskome jeziku i količini jezičnoga unosa kojemu su učenici izloženi (model), ali preostaje i jedan dio rezultata koji se ne može objasniti tim čimbenicima (i naziva se pogreškom). Jezik i njegovo usvajanje nisu po tome jedinstveni – veliku većinu pojava u psihologiji, ekonomiji, pa i medicini, također je vrlo teško obuhvatiti formulama poput Newtonove, te se umjesto njih moraju upotrebljavati statistički modeli. O modelima možemo govoriti i kod inferencijalne i kod deskriptivne statistike – već je i prosjek model jer njime predstavljamo veću skupinu podataka koji nemaju svi istu vrijednost.

Svrhu statistike posebno je važno istaknuti kada govorimo o inferencijalnoj statističkoj analizi, čija je uloga manje očigledna. Naime, naše iskustvo u nastavi govori da se prva glavna svrha statističke analize, **sažimanje značajki prikupljenoga uzorka** putem deskriptivne statistike, relativno lako razumije već na temelju iskustava iz svakodnevnoga života, dok inferencijalnu statistiku zbog njezine znatno veće apstraktnosti istraživači ponekad prihvate bez potpunoga razumijevanja, kao nešto što se mora provoditi jer čini istraživanje valjanim ili relevantnim. Međutim inferencijalne analize ne čine same po

sebi istraživanje ni dobro provedenim ni relevantnim (v. i potpoglavlje 2.1.6 o kontroli kvalitete istraživanja), već su jedan od metodoloških instrumenata koji, uz adekvatnu primjenu, pomaže u ispravnome izvođenju zaključaka o proučavanoj pojavi. Konkretnije, inferencijalna statistika ispunjava drugu glavnu svrhu statističke analize, a to je **uopćavanje** ili **generalizacija rezultata izvan ispitivanoga uzorka** (engl. *generalisation*).

Pitanje uopćavanja tijesno je povezano s različitim ciljevima kojima se u istraživanjima može težiti. Naime, iako je opći cilj svakoga znanstvenog istraživanja utvrđivanje činjenica o nekoj pojavi (naprimjer o nekome aspektu usvajanja drugoga jezika, poput utjecaja materinskoga jezika), konkretniji cilj može biti drukčiji u različitim istraživanjima ili čak različitim fazama istoga istraživanja. Kod empirijskih istraživanja **opisivanje** (engl. *description*) je nužan prvi cilj, povezan s korakom u kojemu se putem deskriptivne statističke analize daje pregled onoga što pokazuje ispitivani uzorak (tome je cilju posvećeno treće poglavlje). Budući da se u usvajanju drugoga jezika gotovo uvijek ispituju uzorci, a ne cijele populacije, ako se statistička analiza zaustavi na opisivanju, ostaje nam nepoznato što bi se moglo dogoditi izvan uzorka. **Predviđanje** (engl. *prediction*) je stoga jedan korak dalje od opisivanja – ono podrazumijeva pretpostavljanje novih podataka, tj. uopćavanje opisa dobivenoga iz uzorka na druge slične uzorke i čitavu populaciju koju ti uzorci predstavljaju. Predviđanje je cilj koji se može postići samo inferencijalnom statističkom analizom (kojom se u ovoj knjizi bavimo u četvrtome i petome poglavlju). Ono je iznimno važno u znanosti i često se može naići na tvrdnju da znanost bez mogućnosti predviđanja nije znanost (v. Fajgelj, 2014, str. 537). Konačno, **objašnjenje** (engl. *explanation*) nam omogućava ne samo da predvidimo nove podatke već i da razumijemo zbog čega su oni takvi. Objašnjenje bi trebalo biti krajnji cilj znanstvenih istraživanja. Međutim važno je istaknuti da je predviđanje vezano uz statističke postupke, dok je objašnjenje korak u kojemu je ključna teorijska osnova istraživanja i u kojemu istraživači povezuju svoje rezultate s prethodnim spoznajama u području (na tu ćemo se temu osvrnuti u drugome i šestome poglavlju, u kojima govorimo o istraživačkim pitanjima i o tome kako se na njih odgovara). Odnos između ciljeva znanstvenih istraživanja, vrsta statističke analize i teorijskoga okvira shematski je prikazan na slici 1.



Slika 1. Ciljevi znanstvenih istraživanja i njihov odnos s vrstama statističke analize i teorijom

Drugim riječima, statistička je analiza dio metodološke postavke istraživanja. Njezina deskriptivna strana pokazuje nam kakav je uzorak podataka koje smo prikupili, dok nam njezin inferencijalni aspekt može pomoći da iz podataka koje imamo u uzorku doznamo i više od onoga što izravno vidimo. No statistika nije svemoćna i ne oslobađa nas obveze interpretacije podataka u svjetlu teorijskoga okvira koji smo uzeli za polazište istraživanja.

1.4 Što je potrebno za provedbu statističke analize?

Kvantitativna obrada podataka iz istraživanja provodi se u više koraka, koji uključuju: provjeru, organizaciju i kodiranje prikupljenih podataka, izračunavanje osnovnih numeričkih mjera s grafičkim prikazivanjem rezultata i konačno provedbu testova koji istraživačima omogućavaju da dobivene rezultate uopće izvan ispitanoga uzorka (v. i Phakiti, 2014, str. 160, gdje se izdvaja čak devet faza analize). Kao što se može uočiti iz navedenih koraka, najveći dio obrade podataka čini statistička analiza.

Osnovni je preduvjet za provedbu statističke analize postojanje numeričkih podataka. Međutim, kod primjene inferencijalne statistike, uz takve je podatke potrebno i pridržavati se postavki kvantitativne znanstvene paradigme (usp. potpoglavlje 1.1), za koju je karakterističan **hipotetičko-inferencijalni pristup**: na temelju teorije postavlja se hipoteza, zatim se osmišljava metodologija, prikupljaju se i statistički analiziraju podaci te se na temelju rezultata analize prihvaća ili odbacuje hipoteza. Drugim riječima, kako ćemo vidjeti i u četvrtome i petome poglavlju, čija su tema statistički

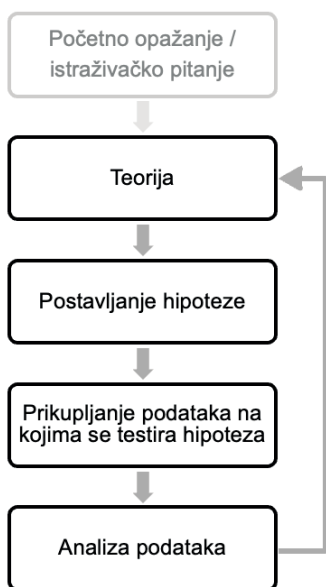
testovi, u kvantitativnim istraživanjima središnju ulogu ima hipoteza koja se u istraživanju provjerava – statistički testovi podrazumijevaju upravo testiranje hipoteza. Budući da većina istraživanja nije sveobuhvatna u pogledu podataka te ne uključuje cijelu populaciju, već se provodi na uzorcima, formalni okvir za provjeru hipoteza daje inferencijalna statistička analiza – kako je prethodno već navedeno, bez nje se rezultati ne mogu uopćiti izvan ispitanoga uzorka.

U istraživanjima jezika postoje i ona koja se temelje na kvantitativnim podacima, ali ne upotrebljavaju te podatke za provjeru hipoteza. Takva se istraživanja nazivaju **eksploracijskima** (engl. *exploratory*) jer polaze od širega skupa podataka o nekoj pojavi i pokušavaju uočiti pravilnosti u tim podacima. Za takve situacije postoje i posebni statistički postupci koji pomažu u pronalasku pravilnosti (posebice eksploracijska faktorska analiza, engl. *exploratory factor analysis* te analiza glavnih komponenti, engl. *principal component analysis*). Iako takvim induktivnim pristupom mogu dovesti do identificiranja novih čimbenika te postavljanja novih hipoteza, što ima posebno veliku važnost za razumijevanje pojava o kojima postoji malo prethodnih znanja, u pogledu mogućnosti uopćavanja takva su istraživanja vrlo ograničena te teško mogu ostvariti ciljeve dalje od (početnoga) opisivanja pojave.⁵ Suprotno njima, istraživanja koja sadrže značajke kvantitativne znanstvene paradigme i vode k predviđanju i objašnjenju nazivaju se **eksplanatornima** ili **objasnidbenima** (engl. *explanatory*). Ona ne polaze od podataka kako bi došla do hipoteza o pojavi koja se ispituje, već su im početna točka unaprijed postavljene hipoteze utemeljene na teoriji i/ili na već postojećim empirijskim spoznajama, dok nove podatke upotrebljavaju za provjeru hipoteza. Zbog kretanja od općega (hipoteza) k pojedinačnome (empirijski podaci) njihov se pristup označava kao deduktivan. To je pristup u većini često primjenjivanih statističkih postupaka.

Naravno, svaki zaključak o hipotezi povratno će utjecati na teorijske spoznaje i zatim voditi k novim istraživanjima, što zapravo čini hipotetičko-inferencijalni pristup cikličnim i iterativnim više nego linearnim. Naprimjer ako bismo u eksperimentalnome istraživanju dobili rezultat prema kojemu izvorni govornici hrvatskoga izostavljaju članove u engleskome kao drugome jeziku jednako kao i izvorni govornici talijanskoga, morali bismo preispitati teoriju za koju je materinski jezik najvažniji čimbenik u usvajanju i koja takav

⁵ Eksploracijski je pristup u lingvistici češće prisutan u korpusnim nego u eksperimentalnim istraživanjima. Korpusi se nerijetko izrađuju tako da sadrže veće količine tekstova na kojima se može ispitati velik broj pojava te istraživači mogu upotrebljavati podatke koji su već dostupni, dok eksperimenti zahtijevaju prikupljanje novih podataka i zbog visoke razine kontrole, kao i praktičnih ograničenja poput zamora ispitanika, po svojoj prirodi imaju užu fokus. Djelomičnu iznimku u pojedinim slučajevima mogu predstavljati eksperimentalni zadaci proizvodnje (usp. Kraš i Miličević, 2015, str. 65–70).

rezultat ne bi mogla objasniti (nešto se slično upravo i dogodilo prije nešto više od pola stoljeća s kontrastivnom analizom, čija su predviđanja o utjecaju materinskoga jezika na usvajanje drugoga jezika bila djelomično opovrgnuta u empirijskim istraživanjima). Ta je cikličnost prikazana na slici 2 (zasnovanoj na prikazima iz Field, 2018), koja u vrhu obuhvaća i generiranje teorija i hipoteza iz početnih opažanja (svakodnevnih, iz literature ili iz eksploracijskih istraživanja). „Teorija“, kao skupina načela kojima se pokušava objasniti određena pojava ili skupina pojava, može predstavljati cjelovit već postojeći okvir u koji pokušavamo smjestiti svoja opažanja (poput generativnoga ili funkcionalnoga pristupa usvajanju drugoga jezika), ali se može odnositi i na objašnjenja koja sami osmišljavamo za pojavu koju smo opazili, čime gradimo novi teorijski okvir. Ključna je značajka „teorije“ da može biti izvor konkretnijih hipoteza koje se mogu testirati na novim podacima.



Slika 2. Tijek empirijskoga istraživačkog procesa

Slika 2 istodobno pokazuje da statistička analiza, iako se kronološki odvija na kraju metodološkoga dijela istraživačkoga procesa, neposredno ovisi o prethodnim koracima, točnije o prirodi prikupljenih podataka i o hipotezi koja se testira. Taj nas složeni odnos dovodi i do pitanja znanja koja je potrebno posjedovati kako bi se statistička analiza provela. Prije nego što se u sljedećemu potpoglavlju detaljnije pozabavimo tom temom, dodajmo da

provedba statističke analize u najvećemu broju slučajeva zahtijeva primjenu nekoga **statističkog programa** za izračunavanje potrebnih mjera. U ovoj se knjizi tim aspektom analize nećemo detaljnije baviti jer smatramo da prvi korak u ovladavanju statistikom treba biti temeljito razumijevanje njezinih osnovnih pojmova. Na nekoliko često korištenih programa samo ćemo se kratko osvrnuti u potpoglavlju 2.2.3, u kontekstu pripreme podataka za analizu.

1.5 Kako ovladati statistikom?

Poznavanje statistike definirano je u literaturi uz pomoć više komponenata. Gonulal (2016, str. 2) navodi: a) sposobnost izbora statističkih metoda koje su primjerene za postavljeno istraživačko pitanje, b) sposobnost adekvatne provedbe statističke analize, c) sposobnost razumijevanja i interpretacije rezultata analize, d) sposobnost procjene solidnosti analize, e) sposobnost da se o rezultatima statističke analize adekvatno izvijeste drugi. Za usvajanje drugoga jezika relevantne su sve navedene komponente – s jedne strane, kako bi istraživači mogli prosuđivati kvalitetu znanstvenih radova u području, a s druge strane, kako bi mogli samostalno provoditi znanstvena istraživanja usvajanja drugoga jezika.

Gass (2009) navodi da je tijekom desetljeća od sredine 20. stoljeća usvajanje drugoga jezika preraslo iz pretežito deskriptivne u pretežito teorijsku disciplinu, u kojoj se primjenjuje velik broj metoda prikupljanja podataka i velik broj statističkih testova kojima se ti podaci analiziraju. U razdoblju od 1948. do 2006., koje je obuhvatila svojom analizom članaka objavljenih u četirima, prema njezinu mišljenju, najvažnijim časopisima u tome području (*Language Learning*, *Studies in Second Language Acquisition*, *Second Language Research* i *Applied Linguistics*), udio kvantitativnih istraživanja porastao je od 2 % na 61 % (Gass, 2009, str. 10). Porasla je i uporaba statistike – sve do 70-ih godina prošloga stoljeća ona je gotovo potpuno izostajala, dok je 2006. bila zastupljena u 70 % radova (str. 23). Uporaba statistike također je postala složenija, odnosno porastao je broj statističkih testova koji se upotrebljavaju. Na metodološki napredak u pogledu nacrtava istraživanja i statističkih testova upućuju i druge metaanalize provedene od 80-ih godina 20. stoljeća do danas (Gass i sur., 2021; Lazaraton, 2000; Loewen i Gass, 2009; Plonsky, 2013; Plonsky i Gass, 2011).

Međutim istraživači usvajanja drugoga jezika i dalje često smatraju da zaostaju za srodnim disciplinama, poput psihologije.⁶ Mnogobrojni radovi

⁶ Plonsky i Gass (2011, str. 326–327), između ostalih, ukazuju na često oslanjanje istraživača

svraćaju pozornost na problematične prakse koje se u području primjenjuju kako u samim analizama tako i u koracima koji im prethode (npr. u odabiru uzoraka) ili im slijede (posebice u izvještavanju o rezultatima). U sljedećim ćemo poglavljima često citirati takve radove i sugestije koje se u njima navode u cilju popravljanja situacije. Ovdje unaprijed upućujemo posebice na opći pregled metodološke kvalitete istraživanja usvajanja drugoga jezika (Plonsky, 2013), zatim na popis sedam glavnih statističkih „grijeha“ koji se susreću u području (Al-Hoorie i Vitta, 2019) te analizu istraživačkih praksi koje se smatraju „upitnima“ (engl. *questionable research practices*, Isbell i sur., 2022; Wood i sur., 2024). Problemi na koje takvi radovi ukazuju dijelom se podudaraju s onima koji se susreću i u drugim područjima, a koji često proizlaze iz općih razvojnih tijekova, a ponekad i pravih kriza ili revolucija, u primijenjenoj statistici (v. Howard, 2024, kao i potpoglavlja 4.1 i 5.6 u ovoj knjizi). U našem se slučaju moraju uzeti u obzir i relativna mladost područja,⁷ a možda još i više poslovična nesklonost istraživača matematici, tijesno povezana sa sustavno podupiranim shvaćanjem jezika kao predmeta isključivo umjetničke i humanističke prirode, te posljedični nedostatak obuke iz područja statistike (nasuprot primjerice studiju psihologije).⁸

Rezultati opsežnih anketnih istraživanja i intervjua potvrđuju da su istraživači usvajanja drugoga jezika pretežito nezadovoljni vlastitim poznavanjem statističke analize, kao i obrazovanjem koje su stekli iz područja statistike. Takav se trend bilježi od prvih istraživanja na ovu temu (Lazaraton i sur., 1987) do suvremenoga razdoblja (Crowther i sur., 2020; Gonulal, 2016; Gonulal, 2020; Gonulal i sur., 2017; Loewen i sur., 2014), podjednako među početnicima i među iskusnijim istraživačima. Slični su rezultati dobiveni u istraživanjima zasnovanima na testovima poznavanja statističkih pojmova (v. npr. Gonulal, 2018; Loewen i sur., 2020), te se sveukupno može reći da statistička pismenost unutar zajednice nije na visokoj razini. Dodatno treba istaknuti pretežito nisku statističku samoefikasnost, tj. nizak stupanj

usvajanja drugoga jezika, kao i časopisa u tome području, na metodologiju i standarde istraživanja u psihologiji. Jedan je od pokazatelja i česta primjena smjernica i preporuka sadržanih u *Priručniku za objavljivanje Američkoga psihološkog udruženja* (American Psychological Association, 2020).

⁷ Obično se smatra da se usvajanje drugoga jezika kao samostalna disciplina počelo razvijati potkraj 60-ih ili početkom 70-ih godina 20. stoljeća (Gass, 2009).

⁸ Kako navode Roever i Phakiti (2018, str. xviii), „Budući da istraživači u području primijenjene lingvistike često dolaze sa studija umjetničkih i humanističkih znanosti, odgoja i obrazovanja i/ili društvenih znanosti, oni obično slabo poznaju matematičke i statističke pojmove i postupke te doživljavaju statistiku kao strani jezik, plašeći se same pomisli da se trebaju uhvatiti ukoštac s kvantitativnim pojmovima i razvijati statističke vještine.“ (prijevod M. M. P. i T. K.)

uvjerenosti u vlastitu sposobnost provedbe statističke analize (v. npr. Gonulal i sur., 2017). Ipak, ohrabruje činjenica da je udio istraživača koji su stekli barem osnovno statističko obrazovanje u međuvremenu porastao te se nezadovoljstvo i nedostatak znanja izražavaju posebice u vezi s naprednijim pojmovima i analizama (v. Crowther i sur., 2020; Gonulal, 2016; Loewen i sur., 2020). Većina je navedenih radova posvećena engleskome govornom području, ali rijetki radovi koji se bave drugim sredinama, poput istraživanja Zhang i Han (u tisku) o situaciji u Kini, ukazuju na to da je opisana situacija doista karakteristična za znanstveno, a ne za geografsko područje. Naša osobna nastavna iskustva to potvrđuju za područja Hrvatske i Srbije (v. i Pavlović i Stanojević, 2020).

Kako bi se opća situacija poboljšala, istraživanja citirana u prethodnome odlomku, kao i brojna druga (među kojima Brown, 1991; Brown, 2015; Gass i sur., 2021), ističu potrebu temeljitije izobrazbe istraživača usvajanja drugoga jezika u području statistike, a posebice naprednije statističke analize. Već postoje naznake da povećanje broja odslušanih kolegija posvećenih statistici značajno pridonosi pozitivnijemu stavu prema njoj, kao i poboljšanju statističke pismenosti i samoefikasnosti; o tome detaljno izvještavaju Gonulal (2018), Loewen i sur. (2020) te Zhang i Han (u tisku). Može se stoga bez zadržke zaključiti da su u području usvajanja drugoga jezika potrebne sustavne inicijative da se u programe sveučilišnih studija uvedu kolegiji posvećeni statistici. Međutim treba istaknuti da su citirana istraživanja pronašla i pozitivan učinak samostalnoga učenja (ponajprije za uporabu udžbenika statistike i za sudjelovanje u radionicama posvećenima statističkoj analizi), dok je za oslanjanje na pomoć (su)stručnjaka utvrđeno da ima negativan učinak na razinu statističke pismenosti.⁹ Drugim riječima, premda tuđa pomoć zasigurno ne može škoditi u rješavanju pojedinih, konkretnih problema, svaki istraživač ipak najbolje poznaje svoje područje istraživanja i svoje podatke te je važno da veći dio koraka u analizi zna provesti sam (uz to treba imati na umu da čak i kada tražimo pomoć, moramo znati postaviti pitanje i razumjeti odgovor). Potpuno razumijevanje tuđih radova također je nemoguće bez znanja o statistici, zbog čega je jedino pravo dugoročno rješenje odgovarajuća naobrazba u području statistike.

S tim ciljem ovdje navodimo neke od mnogobrojnih resursa koji su dostupni za samostalan rad i mogu biti od pomoći istraživačima usvajanja drugoga jezika koji žele unaprijediti svoje poznavanje statistike. Među njima se

⁹ U skladu s napomenom 6, u usvajanju drugoga jezika posebice su česte situacije u kojima se istraživači obraćaju za pomoć stručnjacima iz područja psihologije.

nalaze knjige i radovi o jezičnim istraživanjima koje su napisali Arunachalam (2013), Baayen (2008, 2013), Brown (1991), Dörnyei (2007), Garcia (2021), Gries (2013a, 2013b), Johnson (2013), Larson-Hall (2010, 2016), Loerts i sur. (2020), Phakiti (2014), Rasinger (2013) i Roever i Phakiti (2018), kao i knjige namijenjene psiholozima, poput onih koje su objavili Field (2018 i ranija izdanja) te na hrvatskome jeziku ili srpskome jeziku Milas (2009), Petz i sur. (2012), Todorović (2008) i Tošković (2020).¹⁰ Može biti korisno upoznati se i s javno dostupnim resursima razvijenima unutar projekta i mreže istraživača ReLDI (*Regional Linguistic Data Initiative*, <https://reldi.spur.uzh.ch/>; v. posebice Miličević, 2017a, 2017b, 2017c) te projekta UPSKILLS (*UPgrading the SKills of Linguistics and Language Students*, <https://upskillsproject.eu/>, v. posebice Miličević Petrović, Đukanović, Budimirović i Gledić, 2023; Miličević Petrović, Ferraresi i Marty, 2023), čiji je cilj, između ostaloga, upravo razvijanje kompetencija studenata jezika i lingvistike u pogledu istraživačke metodologije i statističke analize podataka.

Također smatramo važnim podsjetiti i na to da je sama sposobnost provedbe statističke analize tek jedan dio onoga što je potrebno za provedbu te analize te da poznavanje matematike i programa za statističku analizu nije onoliko važno koliko bi se moglo pomisliti. Naime, da bi ostvarili ciljeve predviđanja i objašnjenja neke pojave, istraživači moraju na prvome mjestu poznavati prirodu podataka koje žele analizirati i imati jasnu zamisao o tome što analizom tih podataka žele utvrditi – to je često važnije od poznavanja matematičkih formula za statističke testove. Također, nasuprot uvriježenome mišljenju (popraćenome ili čak izazvanome strahom od matematike), teorijsko utemeljenje i osmišljavanje metodologije istraživanja često su zahtjevniji od provedbe analize. Statistička se analiza na te postupke nadovezuje i bez njih je nemoguća, a kad je istraživanje metodološki dobro osmišljeno, odnosno kad su hipoteze dobro postavljene i kad su prikupljeni podaci odgovarajući, analizu je relativno jednostavno provesti. Upravo su to pokazala istraživanja u kojima se ispitalo na temelju kojih se ekspertnih znanja odabiru statistički testovi – zaključeno je da najveću ulogu pritom ima dobro poznavanje nacrtu istraživanja (v. posebice Alacaci, 2004).

Imajući u vidu sve navedeno, cilj je ove knjige približiti osnove statističke analize podataka istraživačima u području usvajanja drugoga jezika, što

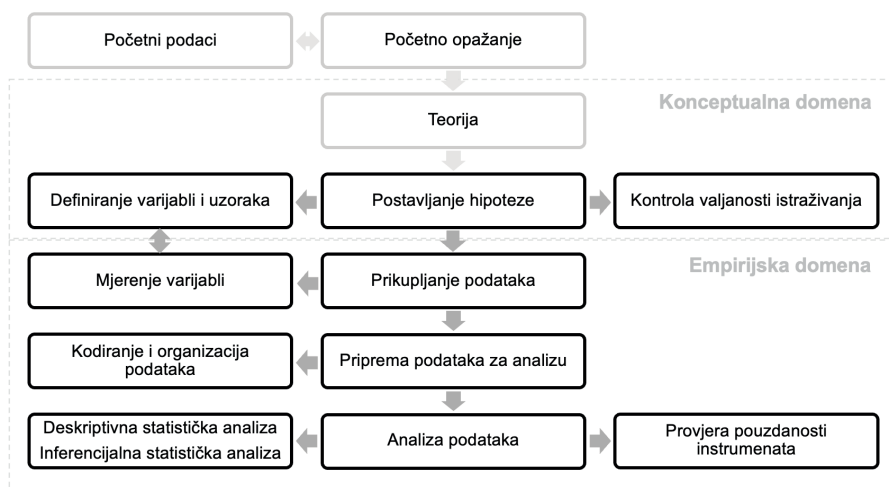
¹⁰ U knjizi i nadalje citiramo različite izvore koji čitateljima mogu koristiti u bavljenju statističkom analizom. Iscrpan anotirani popis relevantnih knjiga i radova objavljenih zaključno s 2008. daju Loewen i Gass (2009). Također, Luke Plonsky na svojoj stranici navodi bibliografiju radova na temu Bayesove statistike, koju u ovoj knjizi ne obuhvaćamo: <https://lukeplonsky.wordpress.com/bibliographies/bayes/> (posljednji pristup 17. 2. 2024.).

bi im trebalo olakšati daljnje samostalno učenje i usavršavanje. Trudimo se na prvome mjestu knjigom nadomjestiti nedostatak literature posebice u području usvajanja drugoga jezika na hrvatskome jeziku i na srpskome jeziku. Pored toga ova je knjiga posebna po pozornosti koja se u njoj posvećuje početnim koracima u analizi, koji se u statističkoj literaturi često podrazumijevaju, kao i po tome što se naglasak stavlja na vrlo detaljan prikaz osnovnih pojmova i postupaka te na podsjećanje da je statistička analiza usko vezana uz postavljanje hipoteza, izradu nacrtu istraživanja i druge elemente opće metodologije istraživanja, kao i uz teorijsku osnovu. Konačno, najbolji način usavršavanja u području statistike jest provedba vlastitih istraživanja i pronalazak rješenja za probleme koji se tijekom toga procesa mogu pojaviti, te se nadamo da će se istraživačima usvajanja drugoga jezika ova knjiga naći pri ruci posebice kao pomoć u takvim situacijama.

2. Koraci do statističke analize podataka

Moglo bi se reći da uspješna analiza podataka uvijek, a posebice u eksperimentalnome istraživanju, započinje puno prije nego što se podaci još i prikupe. Naime, odabir deskriptivnih mjera koje će biti izračunate i testova koji će biti provedeni velikim dijelom ovisi o postavci istraživanja, pa usporedno s definiranjem čimbenika koji će se proučavati, hipoteza koje će se testirati i podataka koji će se prikupiti treba razmišljati i o tome postoji li statistička analiza koja se na odabranu postavku može primijeniti – iznimno je teško pronaći odgovarajuću analizu ako su podaci prikupljeni ne uzimajući u obzir statističke aspekte. Drugim riječima, izostanak planiranja može dovesti do toga da se zbog poteškoća u statističkoj analizi istraživanje mora prilagoditi i ponoviti, a eksperimenti u kojima su sudionici ljudi specifični su po tome da je njihovo ponavljanje iznimno zahtjevno.

Provedbi uspješne analize podataka stoga prethodi pomno osmišljavanje i priprema istraživanja, pri čemu cijelo vrijeme treba razmišljati i o deskriptivnim mjerama i inferencijalnim statističkim testovima koji će se poslije provesti. Nakon planiranja te prikupljanja podataka, a prije glavnoga dijela statističke analize, također je važno ispravno pripremiti podatke za analizu i provjeriti pouzdanost instrumenata uz pomoć kojih su podaci prikupljeni. Tim su temama posvećena sljedeća potpoglavlja, a slika 3 prikazuje najvažnije korake, uz razradu onoga što je već prikazano na slici 2 (v. i Field, 2018; Milas, 2009).



Slika 3. Analiza podataka i koraci koji joj prethode

2.1 Osmišljavanje i priprema istraživanja

Kako navode, između ostalih, Field (2018) i Milas (2009), istraživački proces može se podijeliti u nekoliko faza, koje se mogu grupirati u dvije skupine: konceptualni okvir i provedba istraživanja. U konceptualni okvir istraživanja ulazi definiranje istraživačkoga problema (v. potpoglavlje 6.3), postavljanje hipoteze, izbor istraživačke metode i izrada nacrt istraživanja. Provedba istraživanja obuhvaća prikupljanje podataka, njihovu pripremu i analizu te nadalje izvještavanje, često u obliku pisanja izvještaja o istraživanju. U ovome se potpoglavlju zadržavamo na konceptualnome okviru te u nastavku dajemo pregled onih njegovih elemenata koji su najvažniji za statističku analizu podataka – formuliranje istraživačkih pitanja, postavljanje hipoteza, definiranje i mjerenje varijabli, uzorkovanje ispitanika i podražaja te izrada nacrt istraživanja. Neki su od tih elemenata već detaljno opisani u Kraš i Miličević (2015), ali zbog cjelovitosti prikaza njihov kraći pregled dajemo i ovdje, uz dopune i uz naglasak na aspektima na kojima počiva statistička analiza.¹¹

2.1.1 Od istraživačkih pitanja do hipoteza i varijabli

Kako je već istaknuto u potpoglavlju 1.4, ishodište je statističke analize podataka u eksperimentalnome istraživanju **istraživačka hipoteza** (engl. *research hypothesis*), s kojom je usko povezano **istraživačko pitanje** (engl. *research question*). Najjednostavnije rečeno, istraživačko je pitanje ono na koje se u istraživanju traži odgovor, dok je hipoteza pretpostavljeni odgovor na to pitanje. Oba elementa predstavljaju spojnicu između predmeta i metodologije istraživanja i ponekad će u izvještaju o istraživanju imati jednak sadržaj te se razlikovati samo po upitnome i izjavnome obliku (katkad se zbog toga u izvještaju navodi samo pitanje ili samo hipoteza). Ali pitanja mogu biti i šira od hipoteza, a uz to su i bliža teoriji – na gornjoj bi ih se slici moglo dodati i prije teorije i uz nju, ali i neposredno poslije nje¹² – dok hipoteze, iako izvedene iz teorije, predstavljaju korak dalje k metodologiji i trebaju biti usko fokusirane na čimbenike koji se istražuju, odnosno na pažljivo definirane **varijable** (engl. *variables*).

Varijable predstavljaju one značajke predmeta istraživanja kojima se u danome istraživanju posvećuje pozornost, a koje mogu poprimiti različite

¹¹ O tim se elementima govori i u nekoliko općih *online*-kolegija posvećenih metodologiji istraživanja i analizi podataka u jezičnim istraživanjima (v. Miličević, 2017a, 2017b, 2017c).

¹² Istraživačko se pitanje u prvome trenutku može razviti i iz početnih podataka i opažanja, ali najčešće se potom uobličuje i pobliže definira odabirom teorijske osnove i razmatranjem već postojećih znanja u danome području te u tome obliku pomaže u postavljanju hipoteze.

vrijednosti (Kraš i Miličević, 2015; Miličević, 2017a, 2017b, 2017c; Pavlović i Stanojević, 2020). U lingvistički orijentiranim eksperimentalnim istraživanjima usvajanja drugoga jezika najčešći su predmeti istraživanja s jedne strane učenici drugoga jezika, a s druge jezični sustav koji se usvaja. Stoga su neke od čestih varijabli materinski jezik, razina općega znanja drugoga jezika, dob na početku usvajanja drugoga jezika, vrsta jezične pojave koja se ispituje, razina usvojenosti te pojave, njezina gramatičnost i sl. – neke su od tih varijabli značajke učenika, a neke jezičnoga sustava (usp. Kraš i Miličević, 2015, str. 19; v. i potpoglavlja 2.1.3 i 2.1.4). Još konkretnije, kako je cilj eksperimenata gotovo uvijek ispitati međusobni odnos više čimbenika (poput razine općega znanja drugoga jezika i razine usvojenosti neke strukture), hipoteze izražavaju pretpostavke o tome kako na **zavisnu varijablu** (engl. *dependent variable*) (čimbenik čije vrijednosti želimo objasniti) utječu jedna ili više **nezavisnih varijabli** (engl. *independent variables*) (čimbenika za koje pretpostavljamo da bi mogli biti povezani s čimbenikom koji nam je u središtu pozornosti).¹³ Dodatno treba istaknuti **vanjske varijable** (engl. *extraneous variables*), koje nisu predmet istraživanja, ali mogu utjecati na rezultate; posebice je među njima važno kontrolirati **ometajuće varijable** (engl. *confounding variables*), kako se označavaju oni čimbenici koji mogu pružiti alternativno objašnjenje za rezultate te spriječiti jasan uvid u utjecaj istraživane nezavisne varijable (v. Kraš i Miličević, 2015, str. 23, kao i potpoglavlje 2.1.5 ove knjige).

Istraživačka pitanja, kako navodi Sunderland (2018, str. 14), ključan su dio svakoga empirijskog istraživanja jer daju smjernice istraživačima u pogledu potrebnih podataka, metoda prikupljanja podataka i njihove analize.¹⁴ Dobra istraživačka pitanja formulirana su jasno i precizno (v. npr. Milas, 2009, str. 65, 67). Prema Phakiti (2014), ona trebaju sadržavati elemente poput ispitanika, nezavisnih varijabli i zavisnih varijabli. Također su formulirana neutralno, odnosno tako da unaprijed ne upućuju na odgovor. Naprimjer Kim i McDonough (2008) u svojem istraživanju žele odgovoriti na pitanje: „Proizvode li korejski

¹³ Nazivi „zavisna“ i „nezavisna“ varijabla često se kritiziraju zbog navođenja na zaključak o uzročno-posljedičnoj prirodi njihova odnosa, koja nije nužno prisutna – varijable mogu biti međusobno povezane i kad jedna ne uzrokuje drugu. Budući da je utvrđivanje uzročnosti složen problem koji izlazi iz okvira statističke analize, nećemo se ovdje njime detaljnije baviti (više informacija može se pronaći u Todorović, 2008, str. 56–60, i Tošković, 2020, str. 65–72). U pogledu terminologije, zadržat ćemo te termine kao lakše za početno razumijevanje od alternativnih: **kriterijska varijabla** (engl. *criterion variable*, često i „ishodišna“, engl. *outcome*) i **prediktorska varijabla** (engl. *predictor variable*), koje neki autori upotrebljavaju u užemu značenju (v. Petz i sur., 2012, str. 344, i Todorović, 2008, str. 53, pri čemu se ti termini javljaju samo u kontekstu regresijske analize).

¹⁴ Izvještaji o istraživanjima ponekad navode **ciljeve istraživanja** (engl. *aims* ili *objectives*) umjesto istraživačkih pitanja – oni su najčešće istraživačka pitanja u izjavnoj formi.

učenici engleskoga kao stranoga jezika više pasiva kad ih potaknu glagoli u pasivnim rečenicama njihova sugovornika?“ (str. 150; prev. M. M. P. i T. K.). U tome pitanju možemo uočiti proizvodnju pasiva kod učenika kao zavisnu varijablu te glagole u pasivnim rečenicama njihova sugovornika kao nezavisnu varijablu; jasno je također da će ispitanici u istraživanju biti korejski učenici engleskoga kao stranoga jezika i njihovi sugovornici.

Hipoteze, osim što su formulirane u obliku izjavnih rečenica, također su uže, preciznije i manje eksploracijske od istraživačkih pitanja (Sunderland, 2018; Warne, 2017) te predstavljaju „tvrdnje o potencijalnom i(li) sugeriranome odnosu između barem dviju varijabli“ (Rasinger, 2013, str. 30; Rasinger, 2018, str. 120; prev. M. M. P. i T. K.). Primjer bi bila tvrdnja „Dob na početku usvajanja drugoga jezika utječe na razinu usvojenosti toga jezika“. Budući da je eksperimentalnim istraživanjima cilj testirati neku hipotezu te da hipoteze možemo definirati i kao „tvrdnje koje istraživanjem provjeravamo“ (Pavlović i Stanojević, 2020, str. 47), one moraju biti jasno definirane, mjerljive (usp. potpoglavlje 2.1.2) i, što je ključno, **opovrgljive** (engl. *falsifiable*), odnosno postavljene na način da dopuštaju odbacivanje (Milas, 2009, str. 69; Rasinger, 2013). Hipotezu koju je nemoguće opovrgnuti ne može se smatrati znanstvenom (Milas, 2009; Warne, 2017). Naprimjer hipoteza „Dob na početku usvajanja drugoga jezika mogla bi utjecati na razinu usvojenosti toga jezika“ problematična je jer će biti točna i u slučaju da dob utječe i u slučaju da dob ne utječe na razinu usvojenosti drugoga jezika.¹⁵

Primjere iz stvarnoga istraživanja uzimamo iz Heil i López (2020, str. 427). Jedno od dva istraživačka pitanja u tome istraživanju glasi:

Pokazuju li izvorni govornici engleskoga i izvorni govornici španjolskoga koji imaju visoku razinu znanja engleskoga kao drugoga jezika znakove usvojenosti struktura s objektom kontrolom i infinitiva s *for* ('za') u svojim mentalnim gramatikama engleskoga jezika, operacionalizirane kao statistički značajna razlika između rečeničnih dopuna s ličnim i neličnim glagolskim oblicima koja ide u prilog rečeničnih dopuna s neličnim glagolskim oblicima? (prev. M. M. P. i T. K.)¹⁶

¹⁵ Pojam je opovrgljivosti u znanost uveo Karl Popper u pokušaju rješavanja problema kako se od opažanja neke pojave može stići do zakona koji je definira. Poznat je njegov primjer s labudima: ako opazimo više bijelih labuda, možemo pomisliti da su svi labudi bijeli; kako bismo to dokazali, morali bismo vidjeti sve labude na svijetu, što je nemoguće; rješenje je u tome da potražimo crne labude jer bi već jedan crni labud opovrgnuo našu pretpostavku (Popper, 2010/1959).

¹⁶ Objektna kontrola: *Mary persuaded John to run the marathon.* (rečenična dopuna s neličnim glagolskim oblikom) – **Mary persuaded that John ran the marathon.* (rečenična dopuna s ličnim glagolskim oblikom); infinitiv s *for*: *Mary needed for John to run the marathon.* (rečenična dopuna

To je istraživačko pitanje popraćeno dvjema hipotezama (Heil i López, 2020, str. 427):

1. Izvorni govornici engleskoga jednako će tretirati strukture s objektnom kontrolom i infinitive s *for* u zadatku prosudbe prihvatljivosti rečenica.
2. Izvorni govornici španjolskoga koji imaju visoku razinu znanja engleskoga kao drugoga jezika drukčije će tretirati strukture s objektnom kontrolom i infinitive s *for* u zadatku prosudbe prihvatljivosti rečenica. (prev. M. M. P. i T. K.)

Navedene hipoteze, kao i druge opisane u ovome potpoglavlju, tipične su istraživačke hipoteze formulirane kao odgovor na istraživačko pitanje, u kojima se upotrebljavaju nazivi iz znanstvene discipline kojoj pripadaju (usp. Tošković, 2020, str. 306). U četvrtome ćemo poglavlju, posebice u potpoglavlju 4.1.1, govoriti i o nešto formalnijim statističkim hipotezama. Dodatni pojam koji je ovdje važno spomenuti jest i „predviđanje“. Ponekad se taj termin upotrebljava kao sinonim za hipotezu, kao naprimjer u istraživanju Mitterer i sur. (2020, str. 3), u kojemu je predviđanje bilo da „sudionici ocjenjuju svoj izmijenjeni glas boljim od tuđega glasa, čak i ako – zbog izmjene glasa – nisu svjesni da ocjenjuju i vlastiti glas“ (prev. M. M. P. i T. K.).¹⁷

Međutim ponekad se predviđanjima nazivaju konkretnije posljedice koje slijede ako je (nešto uopćenija) hipoteza točna. Za ilustraciju može poslužiti istraživanje Kim i sur. (2020). Cilj je toga istraživanja bio istražiti prenose li učenici sintaksu i semantiku prvoga jezika u svoje razumijevanje i proizvodnju engleskoga kao drugoga jezika. Jedna od hipoteza iz toga istraživanja, popraćena predviđanjima, bila je sljedeća (str. 127):

Pretpostavljamo da na razumijevanje engleskoga kao drugoga jezika kod učenika čiji je materinski jezik korejski utječe poznavanje semantike njihova prvoga jezika [...]. Stoga predviđamo da će se učenici suočiti sa znatnim poteškoćama u razumijevanju implicirane teličnosti engleskih rezultativnih konstrukcija u zadatku prosudbe prihvatljivosti rečenica, za razliku od izvornih govornika engleskoga. (prev. M. M. P. i T. K.)¹⁸

s neličnim glagolskim oblikom) – **Mary needed that John ran the marathon.* (rečenična dopuna s ličnim glagolskim oblikom)

¹⁷ Ispitanici su bili učenici engleskoga kao drugoga jezika, a izmjena se glasa sastojala u manipuliranju obilježjima vezanima uz spol na snimci (u programu za fonetsku analizu Praat).

¹⁸ „Pretpostavljamo“ je u ovome slučaju prijevod engleskoga „We hypothesize“. Primjer je rezultativne konstrukcije *John hammered the metal flat* ‘John je izravnao metal čekićem’, koja

U statističkoj se analizi podataka svakako polazi od najkonkretnijih pretpostavki o povezanosti između varijabli, pa se u tome smislu predviđanja mogu smatrati istraživačkim hipotezama. Važnost konkretnosti postaje dodatno jasna kad se postavljanje hipoteza promatra u širem kontekstu osmišljavanja zajedno s metodološko-analitičkim okvirom istraživanja, s kojim treba činiti kohezivnu i koherentnu cjelinu (v. npr. Rasinger, 2013, str. 12). Istraživačka pitanja i hipoteze trebaju se naime „konkretizirati kako bi se pojedine vrijednosti mogle izmjeriti“ (Pavlović i Stanojević, 2020, str. 51), što ponajprije podrazumijeva preciznije definiranje varijabli u istraživanju. Drugim riječima, istraživane varijable dobro je što eksplicitnije iskazati već u hipotezi ili predviđanju. Naprimjer u već spomenutome istraživanju Heil i López (2020) iz hipoteza je jasno da je zavisna varijabla rezultat zadatka prosudbe prihvatljivosti rečenica (dakle intuicije ispitanika o prihvatljivosti rečenica; engl. *acceptability judgements*), da je jedna nezavisna varijabla skupina ispitanika, tj. status s obzirom na engleski jezik (uspoređuju se izvorni i neizvorni govornici), a druga nezavisna varijabla vrsta strukture (uspoređuju se objektna kontrola i infinitivi *s for*). Iz istraživačkoga se pitanja dodatno vidi da je još jedna nezavisna varijabla gramatičnost rečenica, odnosno glagolski oblik (uspoređuju se rečenične dopune s ličnim i s neličnim glagolskim oblikom, od kojih su prve negramatične, a druge gramatične; v. primjere u napomeni 16). Konačno, u odlomku posvećenome opisu analize podataka (Heil i López, 2020, str. 433–434) objašnjeno je kako su prosudbe prihvatljivosti rečenica iz prvotne ljestvice od pet razina pretvorene u binarne (prihvaćeno/odbačeno), što je bio konačni način iskazivanja prihvatljivosti. Taj nam primjer pokazuje da je u istraživanjima jezika i njegova usvajanja definiranje varijabli i načina njihova mjerenja često vrlo složen zadatak. Toj je temi posvećeno sljedeće potpoglavlje.

2.1.2 Mjerenje varijabli

Pored toga što mogu poprimiti različite vrijednosti, ključna je značajka varijabli njihova **mjerljivost**, odnosno mogućnost da budu izmjerene. Kako navodi Fajgelj (2020, str. 34), **mjerenje** se najjednostavnije može definirati kao „dodjeljivanje brojeva objektima ili pojavama“, što dovodi do neposrednoga povezivanja mjerenja s kvantifikacijom, odnosno s time da rezultat mjerenja moraju biti količine nekih značajki koje objekti ili pojave posjeduju. U skladu

u engleskome jeziku sadrži glavni glagol koji iskazuje radnju (*hammered*) i implicitni sekundarni predikat koji iskazuje rezultat te radnje (*the metal [is] flat*). U engleskome su takve rečenice neizostavno telične, tj. izražavaju dostizanje cilja, dok u korejskome to nije slučaj.

s time, uobičajeni su primjeri mjerenja oni iz fizike, u kojima se mjere veličine poput mase ili brzine tijela. Znamo da čak i u takvim slučajevima mjerenje zahtijeva donošenje niza odluka, na prvome mjestu one o **jedinici mjerenja**, tj. standardu na koji ćemo se u mjerenju osloniti (sjetimo se britanskih ili američkih receptata za kolače s mjerama u uncama ili „šalicama“, engl. *cups*, te kuhanja uz preračunavanje u grame ako naša vaga samo njih pokazuje), a zatim i o **mjernome instrumentu** kojim će se mjerenje izvršiti (gdje nisu važne samo jedinice koje instrument pokazuje već i njegova točnost i preciznost – različite vage mogu nam dati različite informacije čak i kad upotrebljavaju iste jedinice). Međutim u društvenim je znanostima stanje još složenije jer mnogobrojne značajke koje želimo istražiti ne možemo neposredno opaziti.

U usvajanju drugoga jezika, kao i u društvenim znanostima općenito, neke su varijable relativno lako i izravno mjerljive (npr. dob na početku usvajanja drugoga jezika, brzina čitanja), dok druge nisu (npr. razina općega znanja drugoga jezika, prihvatljivost rečenica). Prve su vezane uz vrlo konkretne opazive pojave, dok se druge odnose na apstraktnije konstrukte te se u svrhu mjerenja moraju povezati s nekim opazivim indikatorom – u takvu slučaju razlikujemo između **teorijskih varijabli** i **empirijskih** ili **indikatorskih varijabli** (v. posebice Todorović, 2008, str. 81; usp. engl. *variable* : *indicator*). Razina općega znanja drugoga jezika može se naprimjer empirijski utvrditi testom čiji bi rezultat postao indikatorska varijabla ili se kao indikator može uzeti samoprocjena učenika (usp. Kraš i Miličević, 2015, str. 42–53). Nije dakle rijetko da jednoj te istoj teorijskoj varijabli odgovara više mogućih empirijskih, a isto tako pojedine indikatorske varijable mogu biti povezane s više teorijskih. Postupak kojim se utvrđuje način mjerenja neke varijable naziva se njezinom **operacionalizacijom** (engl. *operationalisation*) (v. npr. Miličević i Kraš, 2015, str. 18–19; Pavlović i Stanojević, 2020; Todorović, 2008).

Pored odnosa teorijskoga konstrukta i empirijskoga indikatora, važno je razumjeti i načine mjerenja empirijskih varijabli u pogledu numeričkih značajki, što je također vezano uz operacionalizaciju. Osnovna je podjela varijabli na **kvantitativne** (engl. *quantitative*) ili **numeričke** (engl. *numerical*), čije se vrijednosti mogu iskazati brojem, i **kvalitativne** (engl. *qualitative*) ili **kategorijske** (engl. *categorical*), čije se vrijednosti mogu razvrstati u kategorije (v. i Kraš i Miličević, 2015, str. 22; Tošković, 2020). Prvima pripada naprimjer broj točnih odgovora u nekome zadatku (gdje vrijednosti mogu biti samo cijeli brojevi, odnosno radi se o **diskretnim** varijablama; engl. *discrete variables*) ili brzina čitanja (koja može imati bilo koju vrijednost u određenome intervalu i time se definira kao **kontinuirana** varijabla; engl. *continuous variable*). Primjeri

su drugih, kategorijskih varijabli spol (koji u osnovnome biološkom smislu predstavlja **binarnu** varijablu (engl. *binary variable*) jer ima dvije moguće vrijednosti), prvi jezik, kontekst usvajanja jezika, vrsta riječi, vrsta jezične strukture i sl. U teoriji mjerenja ne postoji slaganje u pogledu toga mogu li se vrijednosti kategorijskih varijabli uopće smatrati mjerama jer ne mogu izraziti količinu (usp. Fajgelj, 2020, str. 35), ali kod objašnjenja značajki varijabli u empirijskim istraživanjima najčešće se na tome problemu ne inzistira, već se mjerenje shvaća šire (v. npr. Tošković, 2020, str. 76). U društvenim se znanostima u mjerenju varijabli zapravo najčešće upotrebljavaju četiri **mjerne ljestvice** (engl. *scales* ili *levels of measurement*) koje je definirao psihofizičar Stanley Stevens (Stevens, 1946), i koje opisuju prirodu onoga što vrijednosti neke varijable iskazuju, a uključuju i numeričke i kategorijske varijable.¹⁹

Prva, **nominalna ljestvica** (engl. *nominal scale*) podrazumijeva dodjeljivanje brojeva objektima tako da ti brojevi predstavljaju oznake kategorija kojima dani objekti pripadaju prema nekoj značajki. Nominalna ljestvica služi za mjerenje kategorijskih varijabli poput prvoga jezika, vrste riječi i vrste jezične strukture jer se takvim varijablama mogu pridružiti numeričke vrijednosti (naprimjer možemo hrvatski kao prvi jezik učenika označiti brojem 1, talijanski jezik brojem 2, a njemački brojem 3), ali te su vrijednosti posve proizvoljne (mogli smo odabrati i 6, 22 i 58) i jedina operacija (logička prije nego matematička) koja se nad njima može provesti jest utvrđivanje (ne)jednakosti: dva učenika mogu imati isti ili različit prvi jezik, ali hrvatski nije po rangu ispred talijanskoga i njemačkoga niti možemo uspoređivati udaljenost između hrvatskoga i talijanskoga ili onu između talijanskoga i njemačkoga. U praksi se u provedbi statističke analize zapravo često dodjeljuju izravno nazivi kategorija te se brojevi potpuno isključuju (ako ih program za statističku analizu ne zahtijeva).

Druga, **ordinalna ljestvica** (engl. *ordinal scale*) počiva na načelu da redosljed brojeva koji se dodjeljuju odražava redosljed u razini prisutnosti dane značajke. U svakodnevnome su životu primjer za nju ljestvice jačine potresa (npr. Mercallijeva ljestvica od 1 do 12, pri čemu 1 označava „nezamjetljiv potres“, a 12 „velik katastrofalan potres“).²⁰ U jezičnim je istraživanjima ordinalna ljestvica osobito česta u mjerenju razine prihvatljivosti rečenica, gdje ljestvice od 1 do 5 (ili od 1 do 7, ili od -2 do +2 itd.; v. Kraš i Miličević, 2015,

¹⁹ Steversona je podjela u vrlo širokoj uporabi, ali ne i potpuno prihvaćena od statističara. Među značajkama koje joj se zamjeraju nalaze se smatranje kategorija mjerenjem i izostanak postotaka kao posebne vrste ljestvice (v. i Fajgelj, 2020, str. 64–65).

²⁰ <https://enciklopedija.hr/clanak/mercalli-cancani-siebergova-ljestvica>, posljednji pristup 26. 4. 2024.

str. 84) označavaju razine od „posve neprihvatljivo“ do „posve prihvatljivo“. Ordinalna je ljestvica matematički informativnija od nominalne jer se na njoj vrijednosti mogu uspoređivati prema veličini (potres jačine 10 razorniji je od onoga jačine 6, a rečenica koja dobije ocjenu 5 prihvatljivija je od one koja je ocijenjena s 1). Međutim ta ljestvica nema početak (nulu koja znači odsutnost) ni jedinicu mjerenja, pa se udaljenost između pojedinih vrijednosti ne može numerički izraziti i ne može se smatrati jednakom (v. Fajgelj, 2020, str. 60). To možemo jasno vidjeti i na primjeru rang-lista, koje predstavljaju ordinalne ljestvice izvedene iz rezultata ostvarenih na nekome natjecanju, natječaju ili ispitu. Ako je naprimjer pobjednik na natjecanju za Pjesmu Eurovizije ostvario 591 bod, drugoplasirani 547 bodova, a trećeplasirani 453 boda te zbog tehničke pogreške naknadno pristigne informacija o još 55 glasova za trećeplasiranoga, njegov se rang neće promijeniti – usprkos smanjenoj razlici, dokle god ne pretekne natjecatelja ispred sebe, ostaje pod brojem 3 (usp. Tošković, 2020, str. 81).

Treća, **intervalna ljestvica** (engl. *interval scale*) dodjeljuje „prave“ brojeve, i to tako da razlika između njih pokazuje razliku u razini prisutnosti dane značajke, pri čemu postoji konstantna mjerna jedinica, tj. razmak između susjednih vrijednosti uvijek je jednak (Fajgelj, 2020, str. 61). Opći je primjer za to mjerenje temperature u stupnjevima Celzijusa. Podaci s te ljestvice jednako su udaljeni jedni od drugih (npr. udaljenost od 0 do 10 stupnjeva jednaka je udaljenosti od 10 do 20 stupnjeva), pa se vrijednosti mogu zbrajati i oduzimati. Međutim početak je te ljestvice arbitraran (tj. nula ne znači odsutnost značajke), pa nije smisleno računati omjere.²¹ Zbog toga se ne može reći da je temperatura od 20 stupnjeva Celzijusa dvostruko viša od temperature od 10 stupnjeva, što možemo provjeriti i usporedbom sa stupnjevima Farenheitove ljestvice, na kojoj 68 stupnjeva (=20 °C) nikako ne može biti dvostruko više od 50 stupnjeva (=10 °C). Međutim, kako je jedinica mjerenja konstantna (iako također arbitrarna), znamo za koliko je jedna vrijednost viša od druge, a možemo i reći da je razlika između 10 i 20 stupnjeva dvostruko veća od razlike između 10 i 15 stupnjeva. Pojedine vrste standardiziranih testova, ondje gdje nula ne znači odsutnost značajke, također odgovaraju toj ljestvici.

Konačno, **omjerna ljestvica** (engl. *ratio scale*) upotrebljava se za mjerenje numeričkih varijabli s nulom koja nije arbitrarna, već ima značenje. Točnije, vrijednost nula ovdje znači odsutnost neke vrijednosti, zbog čega brojevi

²¹ S gledišta fizike 0 °C predstavlja točku smrzavanja vode pa ta vrijednost nije arbitrarna u apsolutnome smislu. Međutim, za statistiku je važna matematička arbitrnost, odnosno to što u ovome slučaju 0 ne predstavlja odsutnost temperature.

na ljestvici izravno iskazuju omjer prisutnosti značajke. Velik broj fizikalnih veličina iskazuje se takvom ljestvicom – dužina, masa, vremenski intervali. U jezičnim se eksperimentima na takvoj ljestvici mjeri naprimjer dob ispitanika ili brzina čitanja (v. i potpoglavlje 2.2.1). Kao i kod intervalne ljestvice, vrijednosti su jednako udaljene jedna od druge, ali zahvaljujući apsolutnoj nuli kod te se ljestvice mogu računati i omjeri, odnosno može se reći da je brzina čitanja riječi od 200 ms upola manja nego brzina čitanja od 400 ms. Možemo primijetiti da su i u tome slučaju jedinice u kojima se mjerenje vrši arbitrarne (masu, kao u već spomenutom primjeru, možemo mjeriti u gramima ili uncama) te se kao posebna ponekad izdvaja i **apsolutna ljestvica** (engl. *absolute scale*). Ona se rabi kad se radi o prebrojavanju – broj ispitanika u eksperimentalnome istraživanju ili učestalost uporabe neke strukture neki su od primjera. Nije međutim široko prihvaćena (v. Tošković, 2020, str. 85) i često se konteksti prebrojavanja smatraju primjenom omjerne ljestvice. Jasno je da ta ljestvica daje najviše matematički korisnih informacija o predmetu istraživanja.

Najvažnije značajke četiri Stevensove ljestvice donosimo u tablici 1. Konkretno, navodimo koje se logičke i matematičke operacije mogu provoditi na kojemu tipu ljestvice te što je za svaku od njih glavna značajka u pogledu toga što njezini brojevi pokazuju (usp. i tablicu 9 u potpoglavlju 3.1, gdje su dodane odgovarajuće deskriptivne mjere koje se mogu izračunati kod različitih ljestvica – u toj se tablici vidi i sličnost intervalne i omjerne ljestvice). Naravno, sve operacije navedene za gornje ljestvice vrijede i za donje.

Tablica 1. Usporedba mjernih ljestvica (prema Stevens, 1946)

Mjerna ljestvica	Glavna značajka	Dopuštene operacije
Nominalna	popis kategorija	=, ≠
Ordinalna	poredak s nepoznatim razmacima	>, <
Intervalna	poredak s jednakim razmacima	+, -
Omjerna	poredak s jednakim omjerima	×, /

U eksperimentalnim istraživanjima usvajanja drugoga jezika zavisna je varijabla najčešće numerička (intervalna ili omjerna, u nekim slučajevima ordinalna), a nezavisna kategorijska (nominalna, u pojedinim slučajevima ordinalna). S gledišta statističke analize podataka, numeričke varijable sadrže više informacija o predmetu istraživanja i stoga su poželjnije od kategorijskih. Naravno, u nekim se slučajevima kategorije ne mogu izbjeći (neke su varijable po svojoj biti kategorijske, poput materinskoga jezika), ali u istraživanjima usvajanja drugoga jezika ima i varijabli koje se često iskazuju kategorijski i

kad to nije nužno. Naprimjer razina općega znanja drugoga jezika može se operacionalizirati putem podjele ispitanika u skupine poput početna/srednja/napredna, ali i putem točnoga broja bodova koji su ispitanici ostvarili na testu znanja. Praksa podjele u skupine sve se češće kritizira jer se time gubi dio informacija o ispitanicima (v. npr. Leal, 2018), a uz to i usporedba podataka iz različitih istraživanja može biti problematična jer isti nazivi skupina ne označavaju nužno usporedivu razinu znanja (v. Kraš i Miličević, 2015, str. 47–48). Nadalje, prihvatljivost rečenice može se operacionalizirati putem suda u binarnoj formi „prihvatljivo“ ili „neprihvatljivo“ ili na numeričkoj ljestvici od 1 do primjerice 7. Čak se i dob na početku usvajanja drugoga jezika može operacionalizirati na više načina – putem točnoga broja godina ili putem podjele ispitanika u skupine (npr. „učenik koji je počeo usvajati drugi jezik u djetinjstvu“ i „učenik koji je počeo usvajati drugi jezik u odrasloj dobi“).

Kako ćemo vidjeti u sljedećim poglavljima, numeričke značajke varijabli i ljestvica kojima su mjerene imaju veliku važnost za statističku analizu podataka i jedan su od ključnih čimbenika u odluci o tome što se iz podataka može izračunati, kako u deskriptivnoj tako i u inferencijalnoj analizi. Posebno još jednom vraćamo pozornost na ordinalne ljestvice jer one u praksi dovode do najviše dvojbi i ne analiziraju se uvijek na isti način. Naime, ordinalne ljestvice s niskim brojem vrijednosti često će se tretirati kao da su nominalne (naprimjer razina općega znanja drugoga jezika iskazana kao početna/srednja/napredna), dok se uz veći broj vrijednosti mogu smatrati „slabim“ kvantitativnim ljestvicama (v. Fajgelj, 2020, str. 60; Tošković, 2020, str. 81–82) te se analizirati kao brojevi (uključujući naprimjer izračunavanje koeficijenta korelacije, v. potpoglavlje 5.4.2). S druge strane, iako matematička razlika između intervalnih i omjernih ljestvica postoji, u statističkoj će se analizi podataka iz područja usvajanja drugoga jezika te dvije ljestvice najčešće tretirati jednako.

Konačno, u vezi s mjerenjem treba spomenuti i njegovu **točnost** (engl. *accuracy*). Mjerenje je točno kada pokaže vrijednost koja odgovara stvarnoj vrijednosti koju mjerimo. Iako bismo mogli pretpostaviti da se to podrazumijeva, zapravo je točnost potrebno istaknuti jer se mjerenje gotovo nikad ne događa u savršenim okolnostima i usprkos svim našim naporima može doći do **slučajne pogreške**, vezane uz mjerni instrument, značajke i ponašanje ispitanika ili neki vanjski utjecaj. To je jedan od razloga zbog kojih se u istraživanjima jezika u eksperimentima provodi **pilotiranje** (v. potpoglavlje 2.1.6) i u zadatke se uvijek uključuje više podražaja, koji omogućuju **višestruko mjerenje**. O tim temama detaljnije govorimo u sljedećim potpoglavljima, kao i u potpoglavlju

2.3, posvećenima uzorkovanju i kontroli kvalitete istraživanja (v. i Kraš i Miličević, 2015, str. 38–41).

2.1.3 Definiranje uzoraka ispitanika

Znanstvena istraživanja teže opisati i objasniti pojave u njihovoj cijelosti, pa su i eksperimentalna istraživanja usvajanja drugoga jezika usmjerena k donošenju zaključaka koji se odnose na čitavu **populaciju** (engl. *population*) koja je predmet istraživanja, odnosno na sve članove neke skupine koja ima određene značajke koje nas zanimaju. Kako navodi Tošković (2020, str. 35), značajke su članova jedne populacije takve da ih razlikuju od entiteta koji toj populaciji ne pripadaju. Svako istraživanje definira populaciju koja je za njega relevantna s obzirom na istraživačko pitanje na koje želi odgovoriti, odnosno s obzirom na značajke koje želi proučiti. U istraživanjima usvajanja drugoga jezika želi se objasniti jezično ponašanje svih neizvornih govornika nekoga jezika ili jezičnoga varijeteta, često uz specifikaciju dodatnih značajki poput toga da su govornici određenoga materinskog jezika ili da su na određenoj razini općega znanja drugoga jezika. Tako populaciju mogu činiti naprimjer učenici engleskoga jezika kojima je prvi jezik hrvatski, učenici njemačkoga jezika kojima je prvi jezik talijanski i koji su na početnoj razini općega znanja drugoga jezika i sl.

Kako istraživači gotovo nikad nemaju pristup cijeloj populaciji (najčešće jer je jednostavno prevelika), moraju odabrati odgovarajući **uzorak** (engl. *sample*), tj. podskup populacije koji će je u istraživanju predstavljati. Kao što je već istaknuto u potpoglavlju 1.3, inferencijalna statistička analiza upravo je zato i potrebna – ona nam pomaže da razumijemo značajke populacije na temelju značajki uzorka. Međutim, kako bi se zaključci mogli uopćiti, uzorak treba biti formiran na odgovarajući način, odnosno mora dobro oslikavati populaciju. Naprimjer ako je struktura populacije takva da je žena poprilično više nego muškaraca, adekvatan uzorak zadržat će takav odnos. Drugim riječima, dok je odabir ciljne populacije vezan uz predmet istraživanja, odabir uzorka koji će se istraživati jedan je od središnjih problema i u metodološkome smislu.

U eksperimentalnim istraživanjima usvajanja drugoga jezika posebna se pozornost posvećuje odabiru, tj. **uzorkovanju ispitanika** (engl. *participant sampling*), ali, kako ćemo nadalje vidjeti, i podražaji upotrijebljeni u eksperimentalnim zadacima predstavljaju uzorak te i njihov odabir treba pažljivo planirati. Ispitanici obično dobivaju više pozornosti zbog toga što u eksperimentalnim istraživanjima predstavljaju glavnu **jedinicu analize** (engl. *unit of analysis*), odnosno središnji entitet o kojemu se prikupljaju podaci

(v. Petz i sur., 2012, str. 427, kao i Tošković, 2020, str. 12, 35). U usvajanju drugoga jezika, konkretno, istraživače zanima kako učenici usvajaju ili u kojoj su mjeri usvojili različite aspekte drugoga jezika ovisno o tome koji jezik govore kao materinski, kad su počeli usvajati drugi jezik, na kojoj su razini općega znanja toga jezika i sl. Promatrano statistički, pojedinačni ispitanik u eksperimentima predstavlja i **jedinicu uzorkovanja** (engl. *sampling unit*), tj. jedinicu koja se smatra elementarnim sastavnim članom populacije – kad govorimo o veličini uzorka, zapravo govorimo o broju uzorkovanih jedinica.²²

Uzorkovanje ispitanika s praktične se i etičke strane tiče pitanja kako pronaći ispitanike i regrutirati ih za istraživanje te dobiti njihov pristanak (v. npr. Kraš i Miličević, 2015, str. 61–64), dok se konceptualno i statistički gledano odnosi na definiranje populacije te način prikupljanja i veličinu uzorka. Svaki uzorak trebao bi prije svega biti **reprezentativan** (engl. *representative*), odnosno dobro oslikavati populaciju iz koje je uzet (i koja se istražuje; Kraš i Miličević, 2015, str. 32). To znači da članovi uzorka moraju imati tipične značajke populacije, pri čemu se naglasak stavlja na one značajke koje mogu utjecati na rezultate istraživanja (Tošković, 2020, str. 210). Primjeri značajki koje mogu biti relevantne u istraživanjima usvajanja drugoga jezika jesu dob, spol i kontekst usvajanja drugoga jezika. Dakle, ako populaciju čine naprimjer odrasli učenici njemačkoga jezika kojima je prvi jezik talijanski i koji su na početnoj razini općega znanja drugoga jezika, uz to što svi članovi uzorka moraju odgovarati tome opisu, poželjno je da dobna struktura uzorka (postotak zastupljenosti različitih odraslih poddobi) odgovara strukturi populacije jer iz različitih istraživanja znamo da se kognitivne sposobnosti ljudi povezane s procesima učenja razlikuju prema dobi. S druge strane, ne moramo paziti na političko opredjeljenje ispitanika i strukturu uzorka prema tome kriteriju jer je vrlo malo vjerojatno da ono utječe na usvajanje lingvističkih značajki drugoga jezika (usp. Tošković, 2020, str. 210).

Kako najčešće ne postoji posve pouzdan način provjere reprezentativnosti, uzorak se smatra reprezentativnim ako se prikupi na odgovarajući način. Najreprezentativnijom vrstom smatra se **slučajni uzorak** (engl. *random sample*). Kako navodimo u Kraš i Miličević (2015, str. 32), takav se uzorak sastoji od slučajno odabranih članova populacije, što daje svakome članu jednaku priliku – točnije, jednaku vjerojatnost – da bude uključen u uzorak. U slučajnim uzorcima varijable koje se ne ispituju teže tome da se raspodijele

²² U drugim područjima lingvistike i drukčijim metodološkim pristupima jedinica uzorkovanja često nije pojedinačni govornik, već može biti naprimjer skupina govornika (u sociolingvistici), jezik (u jezičnoj tipologiji) ili tekst (u korpusnoj lingvistici).

među ispitanicima u uzorku podjednako kao u populaciji, što umanjuje mogućnost da utječu na ishod eksperimenta kao ometajuće varijable (v. Kraš i Miličević, 2015, str. 32, kao i potpoglavlje 2.1.1 ove knjige). Drugim riječima, slučajnim će se odabirom vrlo vjerojatno dobiti skupine ispitanika ujednačene prema dobi, spolu i drugim čimbenicima koje je nužno ili korisno adekvatno predstaviti (v. i Gabriel i Gyax, 2023).

Glavne podvrste slučajnih uzoraka jesu **jednostavni, sustavni i stratificirani** (v. Petz i sur., 2012, str. 119–123; Todorović, 2008, str. 40; Tošković, 2020, str. 213–215). Kod jednostavnih je slučajnih uzoraka (engl. *simple random samples*) proces odabira jednak „izvlačenju“ ispitanika na temelju slučajno odabranih brojeva (za koje je poželjno da su automatski generirani jer su ljudi skloni unošenju pravilnosti; v. Petz i sur., 2012, str. 119–120); preduvjet je za takve uzorke to da imamo popis svih članova populacije. Kod sustavnih je slučajnih uzoraka (engl. *systematic random samples*) također nužan popis članova populacije kako bi se zatim mogao za uzorak uzeti svaki *n*-ti član. Kod stratificiranih se uzoraka (engl. *stratified samples*) prije samoga uzorkovanja određuju „slojevi“, tj. podskupine unutar populacije, definirane nekom potencijalno ometajućom varijablom, kako bi se osigurala zastupljenost njezinih različitih vrijednosti. Naprimjer Hiver i Al-Hoorie (2020) u istraživanju posvećenome motivaciji u uzorak južnokorejskih učenika engleskoga jezika proporcionalno uključuju učenike iz škola iz dviju administrativnih regija, sjeverne i južne, kako bi bile zastupljene različite socioekonomske značajke jer je sjever pretežito urbaniziran, dok je jug ruralan. Takvo je uzorkovanje posebice važno ako je neka od podskupina rijetka te možda ne bi bila obuhvaćena slučajnim uzorkovanjem. Dodatna je vrsta reprezentativnoga uzorka **klusterski uzorak** (engl. *cluster sample*), kod kojega se uzorkuju cijele jedinice unutar populacije, najčešće onda kada nemamo popis pojedinačnih članova populacije, ali postoji popis jedinica u koje su članovi grupirani. Tada se slučajnim odabirom mogu izdvojiti jedinice, a zatim svi članovi tih jedinica postaju članovi uzorka – naprimjer svi učenici određenoga broja slučajno odabranih škola (Tošković, 2020, str. 215). Navedene je vrste uzoraka uz to moguće i kombinirati.

Budući da je u praksi često nemoguće formirati odgovarajući slučajan uzorak, u istraživanjima usvajanja drugoga jezika, kao i drugim bliskim disciplinama, upotrebljavaju se i druge vrste uzoraka, koje se mogu skupno definirati kao **neslučajni uzorci** (engl. *non-random samples*) (Petz i sur., 2012, str. 123–124; Todorović, 2008, str. 39; Tošković, 2020, str. 216–218). Vrlo je česta njihova vrsta **prigodan uzorak** (engl. *convenience sample*), koji se

sastoji od ispitanika koji su lako dostupni. To mogu biti studenti na instituciji gdje istraživač radi, razredi u školi ili vrtiću u kojima istraživač poznaje nastavnike ili odgojitelje, skupine s jezičnih tečajeva kojima ima pristup i sl. Kako navodimo u Kraš i Miličević (2015), ako se upotrebljavaju takve skupine, poželjno je primijeniti barem poluslučajni izbor tako što će skupinama biti nasumce dodijeljene vrijednosti nezavisne varijable ako priroda varijabli u istraživanju to dopušta (naprimjer ako je varijabla modalitet prikaza te postoje pisana i usmena verzija eksperimentalnoga zadatka; v. niže). Često se upotrebljava i **volonterski uzorak** (engl. *voluntary sample*), koji se sastoji od ispitanika koji su se sami javili istraživaču, odgovarajući na oglas ili poziv. Koliko god da su praktični i ponekad nužni, takvi su uzorci često nereprezentativni jer se na oglase najčešće javljaju pripadnici viših društvenih slojeva, obrazovaniji i radoznaliji pojedinci te oni koji su spremniji pomoći drugome (v. Milas, 2009, str. 407–408; Todorović, 2008, str. 39). Tim su uzorcima slični i **lančani uzorci** (engl. *snowball samples*), koji započinju kao prigodni ili volonterski, a zatim sami ispitanici preporučuju ili traže nove ispitanike. Iako nisu idealni, neslučajni su uzorci u usvajanju drugoga jezika u praksi vrlo česti.

Kako navode Vitta i sur. (2022), u istraživanjima usvajanja drugoga jezika slučajni su uzorci vrlo rijetki zbog praktičnih ograničenja poput nedostatka popisa članova populacije pa se i sama tema slučajnoga uzorkovanja rijetko javlja u metodološkim analizama i metaanalizama. Isti autori ističu da je u usvajanju drugoga jezika korisna i dostupna vrsta uzorka kojom se mogu u određenoj mjeri nadomjestiti manjkavosti neslučajnoga uzorkovanja **višelokacijski uzorak** (engl. *multisite sample*), kod kojega se prikupljaju poduzorci na različitim lokacijama (najčešće na više sveučilišta ili u više škola). Naprimjer Gyllstad i sur. (2023) istraživali su u višejezičnim razredima u Švedskoj učinkovitost u poučavanju engleskoga vokabulara (1) definicija riječi na engleskome, (2) prijevoda riječi na jezik škole (švedski) i (3) prijevoda riječi na jezik škole i materinski jezik učenika; podaci su prikupljeni u četirima gradskim školama u dvjema regijama Švedske. Takav se uzorak može opisati kao spoj klsterskoga i prigodnoga uzorka jer uključuje izbor jedinica unutar populacije, ali izbor jedinica (tj. lokacija) uglavnom nije slučajan, već prigodan. U pojedinim slučajevima takav uzorak može imati i značajke stratificiranoga uzorkovanja ako se jedinice biraju tako da predstave podskupine važne za istraživanje. Moranski i Ziegler (2021) ističu da višelokacijski uzorci imaju niz metodoloških prednosti, među kojima su uvećanje uzorka i veća raznovrsnost koja se unutar njega postiže (što je posebno važno za istraživanje tema u kojima na rezultat mogu utjecati socioekonomske značajke ispitanika). U

potpoglavlju 2.1.6 vidjet ćemo da te prednosti izravno utječu na poboljšanje valjanosti istraživanja, a u potpoglavlju 4.1.2 da mogu uvećati sposobnost statističkih testova da otkriju utjecaj nezavisnih varijabli.

U planiranju vlastitih istraživanja, kao i prilikom čitanja literature, treba međutim biti oprezan u pogledu toga što se definira kao višelokacijski uzorak. U usvajanju drugoga jezika mnogobrojna su istraživanja u kojima se uspoređuju učenici koji imaju različite materinske jezike te podatke može biti nužno prikupiti u različitim sredinama, često u različitim zemljama. Ako se na svakoj lokaciji uključi jedan uzorak ispitanika, ne može se govoriti o višelokacijskome uzorkovanju, iako je uzoraka ukupno više od jedan, jer svaka skupina ispitanika predstavlja poseban uzorak. Također treba imati na umu da višelokacijski uzorci donose i niz praktičnih i metodoloških problema za koje treba pronaći rješenje (v. Moranski i Ziegler, 2021, i Morgan-Short i sur., 2018, za pregled takvih problema i njihovih rješenja) te da statistički gledano i dalje nisu idealni jer najčešće nisu slučajni uzorci.

Budući da pretpostavka o slučajnome uzorkovanju stoji u temeljima inferencijalne statistike, neslužajni uzorci djeluju vrlo problematično. Međutim koliko oni to doista jesu, ovisi i o tome što istražujemo – pojedine su pojave dovoljno stabilne i homogene u populaciji da će ih uzorak ma kako on bio odabran, valjano predstavljati. U jezičnim se istraživanjima često susreće takva pretpostavka vezano uz intuicije o gramatičnosti rečenica u materinskome jeziku, gdje se jezična kompetencija smatra jednakom kod svih odraslih govornika koji nemaju jezične ili druge kognitivne poteškoće (v. npr. Arunachalam, 2013, str. 225). U istraživanjima usvajanja drugoga jezika stoga se pretpostavlja homogenost posebice kad je u pitanju **kontrolna skupina** (engl. *control group*), obično sastavljena od izvornih govornika, dok se više pozornosti posvećuje uzorkovanju ispitanika za **eksperimentalne skupine** (engl. *experimental groups*), koje čine učenici drugoga jezika (usp. Kraš i Miličević, 2015, str. 30–31). Međutim u obje je vrste skupina prisutna tendencija da ispitanici u istraživanjima pripadaju određenome sloju populacije, prije svega onome koji se u engleskome jeziku označava kao *WEIRD* – *White* (bijele rase), *Educated* (obrazovani), *Industrialised* (iz industrijaliziranih sredina), *Rich* (iz bogatih sredina), *Democratic* (iz demokratski uređenih sredina). Radovi Andringa i Godfroid (2020) i Plonsky (2023) neki su od novijih pregleda praksi uzorkovanja u istraživanjima usvajanja drugoga jezika i oni ukazuju na to da su uzorci u tome području iznimno često sastavljeni od studenata, posebice onih kojima je engleski ili prvi ili drugi jezik, dok je razina općega znanja drugoga jezika najčešće srednja. Ti i drugi autori pozivaju

da se situacija promijeni uključivanjem većega broja različitih uzoraka koji će adekvatnije predstaviti širu populaciju učenika drugoga jezika. Također, kako ističe npr. Larson-Hall (2010, str. 44–45), potrebno je prilikom tumačenja rezultata statističkih testova imati na umu izostanak slučajnoga uzorkovanja te uopćavati rezultate na širu populaciju s oprezom i prije svega ako su oni potvrđeni i u drugim istraživanjima (v. i potpoglavlja 2.1.6 i 7.1).

Drugo važno pitanje vezano uz uzorkovanje ispitanika jest potrebna **veličina uzorka** (engl. *sample size*). To je zapravo jedno od prvih pitanja koje istraživači u praksi postavljaju, ali se na njega, nažalost, ne može dati jednostavan i jedinstven odgovor. Uzorak naime mora biti dovoljno velik da omogući adekvatnu statističku analizu podataka i uopćavanje rezultata (tj. treba dati stabilne rezultate, u kojima se pokazuju eventualne pravilnosti u istraživanoj pojavi; Todorović, 2008, str. 43), a „dovoljno“ se može razlikovati ovisno o značajkama populacije i istraživane pojave, posebice o tome koliko su podaci varijabilni, tj. koliko se u njima može uočiti grupiranje oko neke vrijednosti te koliko su velika odstupanja (v. Petz i sur., 2012, str. 126). Načelno gledano, mali su uzorci skloni većoj varijabilnosti (ako zamislimo niz brojeva i njihov prosjek, što je niz kraći to će više svaka nova vrijednost utjecati na prosjek) i manje su reprezentativni od većih uzoraka (v. npr. Larson-Hall, 2012b), što dovodi do toga da u njima utjecaj nezavisnih varijabli može ostati prikriiven. Drugim riječima, veličina uzorka usko je povezana sa sposobnošću statističkih testova da „prepoznaju“ efekt neke nezavisne varijable kada je on prisutan (v. posebice potpoglavlje 4.1.2). Iako je u istraživanjima usvajanja drugoga jezika to glavni problem koji se susreće i o kojemu se govori, treba biti na oprezu i kod vrlo velikih uzoraka jer se u njima utjecaj nezavisnih varijabli može prikazati i kada on zapravo ne postoji, već proizlazi iz velike količine podataka. Dok je u slučaju malih uzoraka moguće djelomično umanjiti problem tako što će se prilagoditi statistička analiza ili upotrebljavati testovi koji su primjereni za manje uzorke (v. peto poglavlje), u slučaju velikih uzoraka važne su informacije o dodatnim mjerama koje uzimaju u obzir veličinu uzorka (v. posebice potpoglavlje 4.1.3).

Broj se ispitanika često određuje na temelju prethodnih istraživanja i/ili na temelju dostupnosti ispitanika. U pogledu nešto preciznijih općih smjernica, Fajgelj (2014, str. 543) navodi da se u statistici općenito uzorci s manje od 30 jedinica smatraju malima, uzorci s od 30 do 100 jedinica srednjima, a uzorci s više od 100 jedinica velikima. To bi značilo da je u eksperimentalnim istraživanjima usvajanja drugoga jezika poželjno imati najmanje 30 ispitanika u svakoj skupini – znamo međutim da to u praksi često nije slučaj jer je zbog

praktičnih ograničenja i vrlo specifičnih zahtjeva s obzirom na prvi jezik, razinu općega znanja drugoga jezika i druge značajke ponekad teško prikupiti toliko ispitanika. Statistički gledano, potrebnu veličinu uzorka moguće je preciznije odrediti uz pomoć formula koje uzimaju u obzir predviđene i željene statističke značajke istraživanja kao i procjenu varijabilnosti uzorka (v. potpoglavlje 3.1.3 i četvrto poglavlje). U ovoj se knjizi toj temi vraćamo posebice u potpoglavlju 4.1.5, a još više informacija daju Buchstaller i Khattab (2014, str. 81–82), Cohen (1988), Fajgelj (2014, str. 549–552), Norouzian (2020), Quené (2010, str. 276–277) i Todorović (2008, str. 43). Danas je, zahvaljujući razvoju računala i statističkih programa, također moguće osloniti se na simulacije podataka, naprimjer koristeći paket za program R (v. potpoglavlje 2.2.3) pod nazivom „faux“ (DeBruine, 2023).²³

Naposljetku, treba istaknuti da odgovarajući način formiranja uzorka i njegova veličina značajno pridonose tome da uzorak bude reprezentativan i pokaže vrijednosti bliske populacijskima, ali oni to, nažalost, ne mogu zajamčiti. Uzorci su po definiciji aproksimacija populacije iz koje su izdvojeni i čak i kad istraživači metodološki sve učine kako bi trebalo, gotovo uvijek će se javiti neka **pogreška uzorkovanja** (engl. *sampling error*) i vrijednosti izmjerene u uzorku neće savršeno predstaviti značajke populacije. Što je uzorak veći i što način njegova izbora bolje odgovara značajkama populacije, to je veća vjerojatnost da će pogreška biti mala, no treba imati na umu da će ona uvijek biti prisutna i da će različiti uzorci iz iste populacije pokazati različita odstupanja od populacije čak i ako se izdvajaju na isti način i jednake su veličine.

2.1.4 Definiranje uzoraka podražaja

Kao što smo navele u prethodnome potpoglavlju, u eksperimentalnim istraživanjima usvajanja drugoga jezika i podražaji uključeni u eksperiment najčešće predstavljaju uzorak mogućih podražaja te je i za njih potrebno razmišljati o načinu uzorkovanja i o veličini uzorka.

U pogledu načina uzorkovanja, u jezičnim je istraživanjima situacija zbog prirode predmeta istraživanja često složena pa je to slučaj i s lingvističkim pristupima usvajanju drugoga jezika. Pojedine se vrste podražaja mogu odabrati slučajnim uzorkovanjem iz dostupnih izvora koji predstavljaju populaciju. Naprimjer riječi koje će biti korištene kao podražaji u zadacima vezanima uz usvajanje i obradu vokabulara mogu se slučajno odabrati

²³ <https://debruine.github.io/faux/> (posljednji pristup 2. 4. 2024.)

iz rječnika.²⁴ Međutim kad se pozornost u eksperimentalnome zadatku usmjeri na rečenice, budući da je broj mogućih rečenica u svakome jeziku neograničen, ne postoji njihova konačna populacija te ne može postojati ni popis iz kojega bi se vršio slučajni odabir. Slučajan se odabir rečenica u određenoj mjeri može provesti uporabom postojećih jezičnih korpusa, ali treba imati u vidu da su i korpusi po svojoj prirodi uzorci danoga jezika ili varijeteta, a ne populacija. Također, zahtjevi su eksperimentalnih zadataka često takvi da je odgovarajuće rečenice teško pronaći u jezičnoj uporabi, naprimjer zato što je istraživana konstrukcija vrlo rijetka ili zato što je potrebno kontrolirati potencijalne ometajuće varijable poput dužine rečenice ili učestalosti riječi koje se u njoj pojavljuju. Zbog toga se u velikome broju slučajeva u eksperimentalnim zadacima usvajanja drugoga jezika ne radi o uzorku koji se uzima iz „opipljive“ populacije, već su podražaji rečenice koje su istraživači sami sastavili, odnosno dolaze iz teorijske populacije mogućih rečenica danoga jezika. To se osobito dobro može razumjeti ako se podsjetimo da u pojedinim eksperimentalnim zadacima rečenice mogu biti i negramatične te se i ne pojavljuju u jezičnoj uporabi izvornih govornika (v. primjere u potpoglavlju 2.1.1). Slično tome podražaji u istraživanjima vokabulara mogu biti i pseudoriječi (engl. *pseudowords*), tj. nizovi koji odgovaraju fonotaktičkim pravilima danoga jezika, ali se u njemu ne upotrebljavaju (usp. potpoglavlje 2.2.1). Naposljetku, određeni eksperimentalni zadaci kao podražaje uopće i ne upotrebljavaju jezični materijal već vizualna ili auditorna sredstva kojima se potiče jezična proizvodnja ispitanika (v. Kraš i Miličević, 2015, str. 35–36).

Zbog svega navedenog, u eksperimentalnim istraživanjima jezika i njegova usvajanja vrlo se često govori o **izradi podražaja** (engl. *stimulus design*; v. Keating i Jegerski, 2015, str. 5; Porte, 2002, str. 48), češće nego o njihovu uzorkovanju iz konkretne već postojeće populacije. U tome su ta istraživanja slična onima zasnovanima na testovima i upitnicima (naprimjer istraživanjima motivacije za učenje drugoga jezika). Naime, reprezentativnost upitnika i njihovih pojedinačnih čestica u većini se slučajeva ne postiže slučajnim odabirom već složenijim postupcima koji, s jedne strane, uključuju deduktivnu ekspertnu procjenu (točnije, analizu teorije i postojećih empirijskih podataka o danoj pojavi, na temelju čega se formuliraju čestice), a s druge, provjeru na podacima dobivenima primjenom upitnika, za što veliku važnost ima pilotiranje materijala prije njihove krajnje uporabe (Fajgelj, 2020, str. 257,

²⁴ U psiholingvističkim pristupima upotrebljavaju se i velike leksičke baze podataka i programi koji pomažu u odabiru podražaja uzimajući u obzir čimbenike koje je potrebno kontrolirati, poput učestalosti riječi, njezine konkretnosti/apstraktnosti i sl. (v. npr. Lintz i sur., 2021, kao i izvore koje oni navode).

387–388). Drugim riječima, slično tome kako istraživači motivacije polaze od njezine teorijske definicije i postojećih podataka o njezinim empirijskim manifestacijama te prema njima sastavljaju čestice za upitnike, istraživači u lingvistički orijentiranim pristupima usvajanju drugoga jezika najprije definiraju pojavu kojom se bave, a zatim osmišljavaju konkretne primjere te pojave koji će biti uključeni u eksperimentalni zadatak. Naprimjer čestica *Važno mi je učiti engleski jer će mi trebati za karijeru*, iz upitnika za mjerenje stavova i motivacije koji su sastavili Gardner i Mihaljević Djigunović (2003), odražava pretpostavku da je jedan oblik motivacije za učenje drugoga jezika onaj u kojemu se jezik percipira kao instrument za ostvarenje nekoga praktičnog cilja (razina motivacije mjeri se u upitniku na ljestvici od „uopće se ne slažem“ do „u potpunosti se slažem“). Slično tome, u lingvistički orijentiranome istraživanju usvajanja drugoga jezika Heil i López (2020) uključuju u istraživanje česticu *Mary persuaded John to run the marathon* ('Mary je uvjerala Johna da trči maraton') kao primjer rečenice s objektom kontrolom i rečeničnom dopunom s neličnim glagolskim oblikom (infinitivom) te vremenski ograničenim zadatkom prosudbe prihvatljivosti (na ljestvici od 1 do 5, tj. od „potpuno prirodno“ do „potpuno čudno“) pokušavaju utvrditi prepoznaju li učenici engleskoga kao drugoga jezika kojima je materinski jezik španjolski da je rečenica prihvatljiva, što bi prema teorijskome okviru koji prihvaćaju moglo ukazati na pristup univerzalnoj gramatici (v. i potpoglavlje 2.1.1).

U eksperimentalnim istraživanjima usvajanja drugoga jezika obično postoji više kategorija podražaja, odnosno više skupina podražaja koje predstavljaju različite razine neke nezavisne varijable vezane uz njihove značajke (usp. varijable poput dobi ili materinskoga jezika, koje su vezane uz značajke ispitanika). Naprimjer Heil i López (2020) su uz prethodno spomenute rečenice u eksperimentalni zadatak uključili i one s objektom kontrolom i rečeničnom dopunom s ličnim glagolskim oblikom (poput **Mary persuaded that John ran the marathon*), tj. 'vrsta dopune' u njihovu je istraživanju bila jedna od nezavisnih varijabli te su postojale dvije skupine podražaja, po jedna za svaku razinu te varijable ('nelična dopuna' i 'lična dopuna'). Slično, u istraživanju usvajanja engleskoga kao drugoga jezika koje su provele Roberts i Liszka (2013) u zadatku prosudbe prihvatljivosti rečenica i zadatku čitanja postojale su dvije skupine rečenica za varijablu 'glagolsko vrijeme' (jedna u kojoj je upotrijebljeno jednostavno prošlo vrijeme, *past simple*, i druga u kojoj je upotrijebljeno složeno prošlo vrijeme, *present perfect*) te dvije za varijablu 'vrsta rečenice' (jedna u kojoj je glagolsko vrijeme usklađeno s vremenskim prilogom i druga u kojemu oni nisu usklađeni). Podražaji poput

tih, koji predstavljaju različite razine neke istraživane nezavisne varijable (ili više njih), zovu se **eksperimentalni podražaji** (engl. *experimental items*). Uz njih se u eksperimentalne zadatke obično uključuju i **distraktori** (engl. *distractors, fillers*), podražaji nevezani uz predmet istraživanja koji imaju isključivu ulogu otežati ispitanicima njegovo uočavanje. Naime, budući da se u eksperimentalnim istraživanjima usvajanja drugoga jezika želi postići da ispitanici odgovaraju što spontanije, postoji potreba da im se odvrti pozornost s predmeta istraživanja kako razmišljanje o njemu (posebice o normativnim pravilima vezanima uz određenu jezičnu pojavu) ne bi utjecalo na njihove odgovore. Distraktori se, kao i eksperimentalni podražaji, mogu uzorkovati iz dostupnih populacija (poput rječnika ili jezičnih korpusa) ili izraditi.

Druge važne teme vezane uz podražaje odnose se na njihov broj i način analize. Na pitanje potrebne ukupne veličine uzorka, kao i kod uzorkovanja ispitanika, nemoguće je dati jednostavan odgovor jer ukupan potreban broj podražaja ovisi o prirodi ispitivane pojave i nacrtu istraživanja. S praktičnoga gledišta, treba pripaziti da podražaja ne bude previše kako ne bi došlo do zamora ili gubitka koncentracije ispitanika (što je dodatni razlog i za temeljito pilotiranje materijala). Sa statističkoga gledišta, potreban se broj podražaja određuje slično potrebnome broju ispitanika, odnosno prema željenim statističkim značajkama i procjeni varijabilnosti uzorka, kao i na temelju pokazatelja iz ranijih sličnih istraživanja. Kao što je kod ispitanika potrebno razmišljati o broju ispitanika u svakoj skupini, iznimno je važno uključiti više podražaja po razini nezavisne varijable ili kombinaciji varijabli, tj. po **eksperimentalnome uvjetu** (engl. *experimental condition*; usp. Milas, 2009, str. 179–180). Primjer eksperimentalnoga uvjeta može biti glagol u složenome prošlom vremenu u rečenici u kojoj je glagolsko vrijeme usklađeno s vremenskim prilogom (u istraživanju koje su provele Roberts i Liszka, 2013; npr. *Since last week, James has gone swimming every day. Now he's getting bored of it.* 'Od prošloga tjedna John svaki dan ide plivati. Sad mu je to već pomalo dosadilo.'). Uključivanje više podražaja za svaki uvjet, pri čemu bi u svakome uvjetu broj podražaja trebao biti jednak, uobičajeno je i u kognitivnoj psihologiji i proizlazi iz toga što jedan podražaj ne daje pouzdanu sliku o kognitivnim pojavama poput jezika ili percepcije (v. i potpoglavlje 2.1.2, gdje govorimo o višestrukim mjerenjima) te je nužno ispitati više odgovora i ocijeniti jesu li oni konzistentni ili pokazuju značajnu varijaciju.

Za odrasle je ispitanike uobičajeno uključiti najmanje šest podražaja po eksperimentalnome uvjetu – taj je broj odabran naprimjer u istraživanju Roberts i Liszka (2013), gdje su postojala ukupno 24 eksperimentalna podražaja raspoređena u četiri eksperimentalna uvjeta, koja su predstavljala kombinaciju glagolskoga vremena i vrste rečenice. Za djecu je zbog nižega kapaciteta pažnje uobičajen broj četiri, pri čemu se raspon kreće od dva do šest (Schmitt i Miller, 2010, str. 39). Jasno je dakle da je podražaja po eksperimentalnome uvjetu gotovo uvijek manje nego ispitanika po skupini (gdje preporučeni minimalni broj iznosi 30). Metodološke analize iz bliskih područja pokazuju da je takav pristup opravdan te da prepoznavanju učinka nezavisnih varijabli više pridonosi povećanje broja ispitanika nego povećanje broja podražaja. Naprimjer Miller (2024) simulacijama i djelomičnim repliciranjem različitih istraživanja iz područja psiholingvistike dolazi do zaključka da se u istraživanjima zasnovanim na vremenu reakcije s 5–10 podražaja po eksperimentalnome uvjetu dobivaju rezultati vrlo bliski onima dobivenima s većim brojem podražaja te sugerira da se pozornost ponajprije posveti povećanju uzorka ispitanika (osim u slučajevima kada je priroda istraživačkoga pitanja takva da zahtijeva veći fokus na broj podražaja).

S druge strane, broj distraktora treba prije svega ukupno biti tolik da oni ispune svoju ulogu prikrivatelja predmeta istraživanja. Kako navodimo u Kraš i Miličević (2015), distraktora treba biti barem toliko koliko i eksperimentalnih podražaja, pri čemu je poželjno da ih bude i više, posebice u istraživanjima s odraslima. U psiholingvistički orijentiranim istraživanjima često ih ima i dvostruko više od eksperimentalnih podražaja.²⁵ Schmitt i Miller (2010) navode da istraživanja s djecom obično sadrže manji broj distraktora jer je za djecu manje vjerojatno da će prozreti svrhu istraživanja. U istraživanjima u kojima distraktori obavljaju i dodatne funkcije to može utjecati i na njihov broj. Kako navode Schmitt i Miller (2010), budući da je do ispitanika u istraživanjima usvajanja drugoga jezika teško doći, može biti korisno da se dva ili tri eksperimenta kombiniraju u jedan tako da eksperimentalni podražaji jednoga eksperimenta čine distraktore drugoga. Naprimjer u Kraš (2008a) se uz usvajanje uporabe pomoćnih glagola i partitivne nenaglasnice *ne* putem distraktora ispitivalo i usvajanje položaja nenaglasnice u talijanskome jeziku kod izvornih govornika hrvatskoga.

²⁵ Naprimjer u Roberts i Liszka (2013) ima ih 60, dok je eksperimentalnih podražaja 24, u Cunnings i Fujita (2021) 56, u odnosu na 24 eksperimentalna podražaja, a u Felser i Roberts (2007) 60, uz 20 eksperimentalnih podražaja.

U prisutnosti više podražaja po eksperimentalnome uvjetu, u analizi podataka u području usvajanja drugoga jezika (i šire) vrlo je prošireno fokusiranje na ispitanike kao jedinicu analize, tako da se za svakoga ispitanika izračunava jedna konačna vrijednost iz mjerenja dobivenih za pojedinačne podražaje iz istoga eksperimentalnog uvjeta te da se nadalje u statističkim testovima razmatra samo ta vrijednost (najčešće aritmetička sredina, ali može biti naprimjer i postotak točnih odgovora). Tako dobivene vrijednosti nazivaju se **posrednim, indirektnim ili izvedenim mjerama** (Todorović, 2008, str. 135; engl. *derived measures*) i suprotne su **neposrednim ili direktnim mjerama** (engl. *raw measures*), koje se dobivaju jednim mjerenjem (što je slučaj s mnogobrojnim varijablama vezanima uz značajke ispitanika, recimo s dobi na početku usvajanja drugoga jezika). U pristupu zasnovanom na izvedenim mjerama prilikom njihova izračunavanja uzimaju se u obzir vrijednosti dobivene za pojedinačne podražaje, ali se zanemaruje razina njihove varijabilnosti. U psihologiji, a posebice psiholingvistici, takav je pristup već više od pola stoljeća izložen kritikama. Herbert Clark (1973), jedan od glavnih kritičara, ističe da se zbog nedostatka informacija o varijabilnosti na temelju toga pristupa ne može pouzdano znati vrijede li dobiveni rezultati za istraživani jezik općenito ili samo za konkretne podražaje uključene u istraživanje.

Clarkove i druge kritike dovele su do toga da se u velikome broju analiza razmatra i utjecaj podražaja, odnosno eksperimentalnih čestica, što znači da se sve pojedinačne vrijednosti dobivene u eksperimentu unose u analizu kao direktne mjere. Kako navode Cummings i Reeves (2022, str. 203), budući da se i podražaji odabiru uzorkovanjem, u eksperimentima u kojima su ispitanici izloženi podražajima uopćavanje uz pomoć inferencijalne statistike odnosi se podjednako na populaciju ispitanika i na populaciju podražaja. I ispitanici i podražaji stoga predstavljaju **slučajne efekte** (engl. *random effects*), tj. varijable koje za različita opažanja u jednoj skupini podataka imaju različite vrijednosti, za razliku od **fiksni efekata** (engl. *fixed effects*), kod kojih je vrijednost kod svih opažanja u jednoj skupini podataka jednaka i odgovara jednoj razini neke nezavisne varijable. Drugim riječima, dok su fiksni efekti vezani uz nezavisne varijable koje istraživači sami definiraju, slučajni su efekti proizvod uzoraka uključenih u istraživanje. Naprimjer materinski je jezik fiksni efekt ako su svi ispitanici u jednoj skupini izvorni govornici hrvatskoga jezika, a u drugoj izvorni govornici talijanskoga; ispitanik je za obje skupine slučajni efekt jer svaki član skupine pored materinskoga jezika ima i druge značajke koje mogu utjecati na rezultate istraživanja. Iako istraživači započinju istraživanja želeći proučiti fiksne efekte, za bolje razumijevanje podataka trebaju imati uvid i u slučajne.

Razmatranje slučajnih efekata u analizi donosi sa sobom niz statističkih pitanja, o kojima ćemo detaljnije govoriti u različitim dijelovima knjige koji slijede, a posebice u potpoglavljima 5.3.3 i 5.5. Čitateljima svakako savjetujemo da prate relevantnu literaturu jer su neka od tih pitanja i dalje predmet rasprave. Naprimjer Barr (2018), govoreći o psihologiji i psiholingvistici, ali sa širim implikacijama, predlaže da se efekt ispitanika i efekt podražaja ne razmatraju odvojeno, već kao efekt drukčije konceptualizirane jedinice analize – **susreta** (engl. *encounter*). Susreti predstavljaju spoj ispitanika i podražaja, tj. u praktičnome smislu odgovaraju onome što se naziva jednim eksperimentalnim **pokušajem** (engl. *trial*); to može biti naprimjer izlaganje jednoga neizvornog govornika engleskoga jezika eksperimentalnoj čestici *Since last week, James has gone swimming every day. Now he's getting bored of it.* u zadatku prosudbe prihvatljivosti rečenica (u primjeru istraživanja koje su provele Roberts i Liszka, 2013). Naime, Barr smatra da se uopćavanje kad je prisutno više ispitanika i više podražaja ne odnosi na uključene predmete istraživanja – ispitanike i podražaje – odvojeno, već na događaje koji nastaju njihovim susretom, jer ih uvijek razmatramo zajedno. Drugim riječima, potrebno je u analizi svratiti pozornost i na međusobni utjecaj različitih jedinica uzorkovanja, slično tome što se u prisutnosti više nezavisnih varijabli može istražiti njihova interakcija (o čemu govorimo u potpoglavlju 2.1.5). Takvo viđenje zapravo odražava širi plan istraživanja usvajanja drugoga jezika – naglasak je na tome kako učenici usvajaju i u kojoj su mjeri usvojili različite aspekte drugoga jezika, ali ne samo ovisno o čimbenicima vezanima uz same učenike (poput materinskoga jezika ili razine općega znanja drugoga jezika), već i ovisno o tome kakve su značajke danoga drugog jezika (ili čak tipa jezika koji dani jezik predstavlja).

Uz sve navedeno, za podražaje je važno istaknuti i to da ne predstavljaju nestrukturiranu skupinu uzorkovanih jedinica, već su organizirani u cjelovit instrument kojim se mjeri poznavanje određene značajke ili skupine značajki nekoga drugog jezika (usp. potpoglavlje 2.1.2). Organiziranje u instrument podrazumijeva povezivanje podražaja s eksperimentalnim uvjetima (ili s ulogom distraktora) i određivanje broja podražaja ukupno i po uvjetu, o čemu je već prethodno bilo riječi, ali uključuje i definiranje načina prikaza podražaja ispitanicima pri provedbi eksperimentalnoga zadatka. Potonja je tema u izvjesnoj mjeri vezana uz probleme uzorkovanja podražaja, ali seže i šire i obuhvaća različite načine kontrole ometajućih varijabli. Zajedno s nizom drugih razmatranja, kontrola ometajućih varijabli dio je plana za provedbu istraživanja. Dobrim osmišljavanjem toga plana nedostaci u jednome segmentu istraživanja mogu se barem djelomično kompenzirati u

drugim aspektima (naprimjer načinom raspodjele ispitanika ili podražaja po eksperimentalnim uvjetima mogu se kompenzirati nedostaci vezani uz njihovo neslučajno uzorkovanje). Tim je temama posvećeno sljedeće potpoglavlje.

2.1.5 Izrada nacrtu istraživanja

Svi elementi spomenuti u prethodnim potpoglavljima čine dio plana za provedbu istraživanja i način na koji se oni kombiniraju utječe na niz odluka vezanih uz statističku analizu. Ključni pojam koji objedinjuje različite elemente jest **nacrt istraživanja** (engl. *research design*). Nacrt sažima glavne značajke istraživanja: u njemu se određuje kakvi će se uzorci ispitanika i podražaja upotrijebiti, koje će se varijable ispitivati i kako će se mjeriti, kakav će se postupak upotrijebiti te kako će se rezultati analizirati (v. npr. Todorović, 2008, str. 117).²⁶ Unutar tih pitanja mogu se izdvojiti različite vrste nacrtu, i to na temelju više različitih kriterija.

S obzirom na razinu kontrole koju istraživači imaju u istraživanju, razlikuju se **korelacijski nacrti** (engl. *correlational designs*) i **eksperimentalni nacrti** (engl. *experimental designs*). U korelacijskim je nacrtima razina kontrole niska te se ne mogu utvrditi uzročno-posljedični odnosi već samo povezanost između varijabli (v. i potpoglavlje 5.4). Nasuprot tome u eksperimentalnim je nacrtima razina kontrole visoka i ti nacrti mogu utvrditi uzročno-posljedične odnose između zavisnih i nezavisnih varijabli. Iako se može činiti očiglednim da u svim eksperimentalnim istraživanjima treba biti prisutan eksperimentalni nacrt, u praksi je vrlo česta i njegova podvrsta **kvazieksperimentalni nacrt** (engl. *quasi-experimental design*). Takav se nacrt upotrebljava u eksperimentima zasnovanim na nezavisnim varijablama čijim se vrijednostima ne može manipulirati jer su unaprijed dane i mogu se samo odabrati. Drugim riječima, u kvazieksperimentalnome nacrtu prisutne su tzv. **selektivne varijable** (engl. *measured variables*), koje – za razliku od **manipulativnih varijabli** (engl. *manipulated variables*) – ne dopuštaju slučajnu raspodjelu ispitanika prema razinama (v. Kraš i Miličević, 2015, str. 21; Todorović, 2008, str. 48–49; kao i napomenu 3 u ovoj knjizi). To je najčešći tip nacrtu u lingvistički orijentiranim istraživanjima usvajanja drugoga jezika jer se ona iznimno često oslanjaju na podjele ispitanika prema kriterijima poput materinskoga jezika i razine općega znanja drugoga jezika, koji su prema svojoj prirodi selektivni – jasno je da ne možemo za potrebe istraživanja promijeniti nečiji materinski jezik. Najveću skupinu iznimaka čine istraživanja koja uključuju određenu

²⁶ Važnost tih značajki pokazuje i to što je upravo njih potrebno opisati i u postupku pretprijava istraživanja za objavljivanje (v. potpoglavlje 7.1).

nastavnu intervenciju (naprimjer davanje različitih vrsta povratne informacije o pogreškama) te istražuju njezin utjecaj na uspješnost učenika u usvajanju neke pojave u drugome jeziku. U takvim je istraživanjima intervencija manipulativna varijabla te je moguće ispitanike slučajno raspodijeliti prema njezinim različitim razinama.²⁷

Prema broju zavisnih varijabli, izdvajaju se **univarijatni nacrti** (engl. *univariate designs*), koji sadrže jednu zavisnu varijablu, **bivarijatni nacrti** (engl. *bivariate designs*), s dvjema zavisnim varijablama, te **multivarijatni nacrti** (engl. *multivariate designs*), s više zavisnih varijabli (v. npr. Abbuhl i sur., 2013; Field, 2018; Roever i Phakiti, 2018).²⁸ U ovoj knjizi nadalje govorimo samo o univarijatnim nacrtima, budući da su oni najprimjereniji za uvod u statističku analizu i da su u istraživanjima usvajanja drugoga jezika najčešći (v. pregled koji daje Plonsky, 2013, prema kojemu su multivarijatne analize višestruko rjeđe od univarijatnih). Međutim ne treba zaboraviti da postoje mnogobrojne situacije u kojima su multivarijatni nacrti primjereniji, npr. kad se razina usvojenosti neke pojave mjeri na više načina ili uz pomoć više pokazatelja. Konkretni su primjeri istraživanje koje je provela Hellman (2011), gdje su korištena tri različita testa poznavanja vokabulara, dva vezana uz širinu i jedan vezan uz dubinu znanja (*Peabody Picture Vocabulary Test*, *Self-Rated Vocabulary Test* i *Word Associates Test*), te istraživanje čiji je autor Holliday (2015), gdje je usvojenost značajki okluzivnih glasova u korejskome kao drugome jeziku kod govornika kineskoga mjerena u zadatku čitanja uz pomoć parametara vremena uključanja glasa (engl. *voice onset time*) i fundamentalne frekvencije; u oba su slučaja rezultati u različitim komponentama bili analizirani testom koji ih je uzimao u obzir zajedno.²⁹

²⁷ Temeljiti prikaz podvrsta nacrti za takva istraživanja i brojne primjere daje Phakiti (2014): dodatno se uz nastavnu intervenciju može odabrati između različitih pristupa u pogledu toga postoji li test prije intervencije, koliko testova postoji nakon intervencije i kad se oni primjenjuju, postoji li kontrolna skupina i slično.

²⁸ Treba međutim biti oprezan u tumačenju tih naziva budući da se ponekad upotrebljavaju i za ukupan broj varijabli u istraživanju ili za broj nezavisnih varijabli (usp. Baayen, 2008; Todorović, 2008, str. 120; Wallis, 2021, str. 31).

²⁹ Drukčiji se pristup susreće kad postoji više zavisnih varijabli koje su istodobno i u odnosu međusobne zavisnosti. Naprimjer u istraživanjima zasnovanima na mjerenju vremena reakcije često se kao varijabla uključuje i točnost odgovora te se i za vrijeme i za točnost analizira povezanost s nekom nezavisnom varijablom. Međutim kako se vrijeme reakcije obično analizira samo za točne odgovore, tj. za drukčiju skupinu čestica od one koja se razmatra kod točnosti odgovora, te se dvije zavisne varijable analiziraju odvojeno, kao dva univarijatna nacrti (v. npr. Wang i Kaiser, 2022) ili se točnost tretira samo kao kontrolna varijabla (kao u već navedenome istraživanju koje su provele Roberts i Liszka, 2013).

Prema broju nezavisnih varijabli razlikuju se **jednofaktorski nacrti** (engl. *single-factor designs, one-way designs*), s jednom nezavisnom varijablom, i **višefaktorski** ili **faktorijalni nacrti** (engl. *multi-factor designs, factorial designs*), s više nezavisnih varijabli. Faktorijalni se nacrti preciznije mogu opisati kao dvofaktorski, trofaktorski (engl. *two-way designs, three-way designs*) itd. U slučaju kategorijskih nezavisnih varijabli (što je najčešći slučaj), opisuju se i putem istodobnoga iskazivanja broja nezavisnih varijabli i broja njihovih razina, tj. broja kategorija koje se uspoređuju. Naprimjer 2 x 2 označava nacrt s dvjema nezavisnim varijablama od kojih svaka ima dvije razine, dok se nacrt u istraživanju koje su provele Roberts i Liszka (2013) može opisati kao 3 x 2 x 2 jer su u tome istraživanju postojale tri nezavisne varijable, od kojih je jedna imala tri razine (prvi jezik ispitanika – engleski, francuski i njemački), dok su druge dvije varijable imale po dvije razine (glagolsko vrijeme – jednostavno prošlo vrijeme i složeno prošlo vrijeme, te vrsta rečenice – ona u kojoj je glagolsko vrijeme usklađeno s vremenskim prilogom i ona u kojoj glagolsko vrijeme nije usklađeno s vremenskim prilogom).³⁰

Za faktorijalne je nacрте važno istaknuti da omogućuju ne samo ispitivanje samostalnoga utjecaja pojedinih nezavisnih varijabli, koji se naziva i **glavni efekt** (v. npr. Milas, 2009; engl. *main effect*) već i utvrđivanje njihovih **interakcija**, tj. **interakcijskoga efekta** (engl. *interaction*). Interakcijski efekt predstavlja zajednički utjecaj više nezavisnih varijabli, pri čemu utjecaj jedne ovisi o razini druge (Milas, 2009, str. 185–186). U konkretnoj statističkoj analizi istraživači ne moraju ispitati interakcije između svih varijabli, već mogu odabrati one koje smatraju važnima – u usvajanju drugoga jezika u nekim slučajevima mogu biti posebno važne interakcije između različitih varijabli vezanih uz podražaje, dok će u drugima veću važnost imati interakcija između značajki podražaja i značajki ispitanika. Roberts i Liszka (2013) naprimjer nalaze interakcijski efekt između skupine ispitanika i vrste rečenice, odnosno utvrđuju da izvorni govornici više reagiraju na negramatične rečenice od neizvornih govornika.

Nacrti se nadalje mogu podijeliti prema mjernim značajkama uključenih varijabli, pri čemu se u nazivima uzima u obzir ljestvica na kojoj se varijable mjere ili uobičajena deskriptivna mjera koja se računa. Kako navode Todorović (2008, str. 118) i Tošković (2020, str. 333–334), najčešći su tipovi unutar te podjele **frekvencijski nacrti**, u kojima su sve varijable kategorijske, **korelacijski** ili **korelacijsko-regresijski nacrti**,³¹ u kojima su sve varijable

³⁰ Točnije, Roberts i Liszka (2013) analizirale su podatke odvojeno za različito glagolsko vrijeme, te bi se istraživanje moglo opisati i kao zasnovano na dva paralelna nacrtu 3 x 2.

³¹ Nacrt koji je korelacijski u smislu prirode uključenih varijabli nije nužno korelacijski i u pogledu

numeričke, te **nacrti varijance**, u kojima je nezavisna varijabla kategorijska, a zavisna numerička – takvi se nacrti nazivaju još i **faktorijalni**, posebice ako je nezavisnih varijabli više (v. Baayen, 2004; Todorović, 2008, str. 118; Tošković, 2020, str. 333).³² Moguće su i složenije kombinacije, u kojima je naprimjer jedna nezavisna varijabla kategorijska, a druga numerička (pri čemu zavisna varijabla može biti ili numerička ili kategorijska). Također treba istaknuti djelomičnu terminološku nekonzistentnost, koja se odnosi na to da se isti nazivi upotrebljavaju za različite kontekste. Kod uporabe termina „faktorijalni“ naprimjer treba pripaziti na koji se aspekt nacрта odnosi, na broj nezavisnih varijabli, na numeričke značajke i zavisne i nezavisnih varijabli ili na oba kriterija istodobno, slično kao što „univarijlatni“ i „multivarijlatni“ mogu biti atributi nacрта s obzirom na zavisne varijable ili sve varijable zajedno.³³

Definiranje nacрта prema broju varijabli i prema njihovim mjernim značajkama ključno je za statističku analizu podataka. Kako ćemo detaljnije vidjeti u četvrtome i petome poglavlju, to su glavni kriteriji za izbor statističkoga testa i različite kombinacije značajki vode do različitih nacрта istraživanja te zahtijevaju uporabu različitih testova. Izmjena u bilo kojoj od značajki može dovesti do velikih izmjena u analizi.

U jezičnim se istraživanjima sve više ističe potreba da se upotrebljavaju složeni, višefaktorski i multivarijlatni nacrti jer su pojave koje se istražuju složene; kako navode Schmid i Dusseldorp (2010, str. 125), „[v]ječina jezičnih procesa – usvajanje, promjene, gubitak – odvija se unutar složene i višefaktorske mreže unutarjezičnih i izvanjezičnih utjecaja i određena je njome“ (prev. M. M. P. i T. K.). U literaturi se također naglašava važnost načina operacionalizacije varijabli, koji izravno utječe na nacrt i neovisno o broju varijabli. Posebice se ističe da varijable koje su po svojoj prirodi kontinuirane ne treba pretvarati u kategorijske kako bi se dobio nacrt koji se tradicionalno više upotrebljava ili ga je lakše analizirati. Naprimjer Baayen (2004) kritizira „dihotomizaciju“ i „faktorizaciju“, koja se susreće u psiholingvističkim

razine kontrole istraživanja.

³² U literaturi na engleskome jeziku nacrti se prema mjernim značajkama varijabli najčešće opisuju polazeći od prirode zavisne varijable (npr. *categorical outcome* 'kategorijski ishod' u Field, 2018) ili nabrojanjem svih uključenih varijabli (npr. *one dep. variable (nom./cat.) and one indep. variable (nom./cat.) (indep. samples)* 'jedna zavisna varijabla (nominalna/kategorijska) i jedna nezavisna varijabla (nominalna/kategorijska) (nezavisni uzorci)' u Gries, 2013b). Zbog toga u tekstu ne navodimo odgovarajuće engleske termine.

³³ Termin „faktorijalni“ potječe od uporabe naziva **faktor** (engl. *factor*) za kategorijske nezavisne varijable, dok je **varijat** (engl. *variate*) drugi naziv za varijablu koja s određenom vjerojatnošću može uzimati različite vrijednosti iz unaprijed definirane skupine mogućih vrijednosti. Međutim oba se naziva upotrebljavaju i u djelomično prilagođenim značenjima.

istraživanjima kad se podražaji kategoriziraju kao učestali i manje učestali (tj. dihotomiziraju se) umjesto da se u analizi zadrži numerička informacija o njihovoj točnoj učestalosti. Slično tome, kako je već spomenuto u potpoglavlju 2.1.2, razina općega znanja drugoga jezika često se izražava uz pomoć kategorija poput „početna“, „srednja“ i „napredna“, iako ispitanici u mnogim istraživanjima rješavaju test koji daje numerički pokazatelj (usp. Leal, 2018); drugim riječima, moguć korelacijsko-regresijski nacrt pretvara se u faktorijalni, u kojemu su dostupni podaci, statistički gledano, manje informativni.

Uz sve navedeno, važna je podjela nacrta i podjela prema međusobnome odnosu podataka u uzorcima koji se u istraživanju uspoređuju ili čija se povezanost ispituje. Ta se podjela u eksperimentalnim istraživanjima tiče toga sudjeluju li u različitim razinama nezavisne varijable različiti ili isti ispitanici. S obzirom na taj kriterij, možemo razlikovati **nacrte s nezavisnim uzorcima** (engl. *independent samples designs*), u kojima podaci u uzorcima koji se uspoređuju nisu međusobno povezani, te **nacrte sa zavisnim uzorcima** (engl. *dependent samples designs*), u kojima su podaci u uzorcima koji se uspoređuju povezani. Kod prve skupine nacrta, koji se još nazivaju **nacrtima na nezavisnim skupinama** ili **nacrtima između ispitanika** (engl. *between-subjects designs*), u različitim razinama nezavisne varijable sudjeluju različiti ispitanici, odnosno u istraživanju postoji više skupina ispitanika koje su definirane vrijednostima nezavisne varijable. Naprimjer često se prema varijabli 'materinski jezik' među ispitanicima razlikuju izvorni i neizvorni govornici određenoga jezika (npr. engleskoga u istraživanju koje su provele Roberts i Liszka, 2013). S druge strane, u nacrtima sa zavisnim skupinama, koji se još nazivaju **nacrtima unutar ispitanika** ili **nacrtima ponovljenih mjerenja** (engl. *within-subjects designs, repeated measures designs*), u različitim razinama nezavisne varijable sudjeluju isti ispitanici. Točnije, u toj su vrsti nacrta isti ispitanici izloženi različitim tipovima podražaja ili su izloženi istim podražajima u više različitih vremenskih točaka (Roberts i Liszka, 2013, uključuju i tu situaciju jer isti ispitanici odgovaraju na različite vrste rečenica). U istraživanjima s više nezavisnih varijabli česti su i **mješoviti nacrti** (engl. *mixed designs*), koji uključuju i nezavisne uzorke i ponovljena mjerenja. Velik broj lingvistički orijentiranih eksperimentalnih istraživanja usvajanja drugoga jezika upravo je toga tipa. Takvo je zapravo i prethodno spomenuto istraživanje Roberts i Liszka (2013) jer u njemu sudjeluje više skupina ispitanika, koje su sve izložene različitim vrstama podražaja (pri čemu su sve skupine izložene svim vrstama podražaja).

Treba napomenuti i da svaki od navedenih slučajeva podrazumijeva nezavisnost podataka različitih ispitanika unutar jednoga uzorka. Očigledan primjer kad je taj zahtjev neispunjen jest ako jedan ispitanik prepíše odgovore u eksperimentalnome zadatku od drugoga ispitanika. Međutim postoji i niz situacija u kojima neki manje očit čimbenik može dovesti do izostanka nezavisnosti; naprimjer kada bi u nekome uzorku bilo više ispitanika iz iste škole, pripadnost školi mogla bi utjecati na to da su njihovi rezultati međusobno slični, odnosno da nisu nezavisni. Takvim je čimbenicima potrebno posvetiti pozornost u osmišljavanju i analizi podataka iz istraživanja kako bi se ili kontrolirali ili uključili kao slučajni efekti (v. potpoglavlje 2.1.4; o problemima (ne)zavisnosti podataka u lingvistici detaljno govore Winter i Grice, 2021).

Izbor između nacrti s nezavisnim i zavisnim uzorcima nije uvijek slobodan jer se neke varijable mogu ispitati samo putem nezavisnih uzoraka. U istraživanjima usvajanja drugoga jezika takve su varijable naprimjer prvi jezik ispitanika ili dob na početku usvajanja drugoga jezika, kao i druge varijable vezane uz značajke ispitanika; kod takvih varijabli uzorci dolaze iz različitih populacija koje se žele usporediti. Stoga su u tim istraživanjima vrlo česti nacrti s nezavisnim uzorcima i mješoviti nacrti. Nacrt s nezavisnim uzorcima može podrazumijevati da različite skupine ispitanika rješavaju isti zadatak prema čijim se rezultatima uspoređuju, ali moguće je i da različite skupine ispitanika rješavaju različite zadatke definirane nekom nezavisnom varijablom vezanom uz podražaje (naprimjer da jedna skupina ispitanika ocjenjuje jedan tip rečenica, a druga drugi); u tome slučaju članovi nezavisnih uzoraka dolaze iz iste populacije. To je međutim iznimno rijedak slučaj u usvajanju drugoga jezika zbog toga što zahtijeva veći broj ispitanika, koje je u tome području često teško pronaći; izvan područja susreće se relativno često – klinička se ispitivanja u medicini često temelje na dvjema skupinama ispitanika iz iste populacije od kojih jedna dobiva lijek, a druga placebo. Takva situacija načelno zahtijeva veći broj ispitanika ne samo ukupno već i u svakoj skupini, jer uključuje više individualne varijabilnosti – kad različiti ispitanici odgovaraju na različite podražaje, razlike u njihovim odgovorima mogu biti vezane uz varijablu koja se proučava (naprimjer da lijek ublažava simptome bolesti, a placebo ne), ali postojat će i razlike između ispitanika koje nisu vezane uz značajke podražaja. Kako bi se utjecaj individualnih razlika sveo na najmanju moguću mjeru, ispitanici u različitim skupinama u takvim istraživanjima trebaju biti usporedivi po svim značajkama osim one koja se ispituje.

Nacrti sa zavisnim uzorcima u većoj mjeri kontroliraju individualne razlike. U jezičnim istraživanjima, kad isti ispitanici sudjeluju u različitim razinama

nezavisne varijable, eliminira se utjecaj individualnih razlika u čimbenicima za koje se zna da utječu na jezično ponašanje općenito, naprimjer u brzini čitanja, kapacitetu radne memorije, razini općega znanja drugoga jezika, kao i u općim čimbenicima poput spola ili vještina vezanih uz testiranje. Uporaba ponovljenih mjerenja stoga se smatra jednim oblikom kontrole u istraživanju, točnije jednim oblikom **kontrole vanjskih varijabli** (engl. *extraneous variable control*).³⁴ U pitanju je eksperimentalna tehnika kontrole jer se ponovljena mjerenja mogu primijeniti samo na manipulativne varijable (ako njihova priroda to dopušta; naprimjer u istraživanjima s nastavnom intervencijom najčešće nije moguće izložiti iste ispitanike dvjema intervencijama). Slično slučajnome uzorkovanju, ovim se putem može kontrolirati više ometajućih varijabli istodobno.

Zavisni nacrti međutim donose sa sobom i nekoliko problema. Oni zahtijevaju duže trajanje eksperimenta od nacrti s nezavisnim uzorcima, pa kod njih lakše dolazi do zamora i dosade ispitanika, kao i do pojave **učinka vježbe** ili **učinka učenja** (engl. *carryover effects*, v. npr. Abbuhl i sur., 2013, str. 120), gdje sudjelovanje u jednoj razini nezavisne varijable utječe na sljedeće razine.³⁵ Uobičajena dodatna tehnika kontrole u tome je slučaju **uravnoteživanje** (engl. *counterbalancing*). Ono podrazumijeva balansiranje poretka i postiže se tako da različiti ispitanici sudjeluju u različitim razinama nezavisne varijable različitim redoslijedom, tj. ispitanici se svrstavaju u podskupine ili blokove koji se razlikuju prema redoslijedu (više o tome govore Milas, 2009, str. 178–183; Todorović, 2008, str. 98–100). U slučaju kad je razina nezavisne varijable više od dvije, potrebno je izdvojiti veći broj podskupina ispitanika (naprimjer za četiri razine 24 podskupine), zbog čega je potpuno uravnoteživanje, praktično gledano, teško postići. Djelomično uravnoteživanje je nasuprot tome vrlo često, a posebno se često upotrebljava nacrt koji se naziva **latinski kvadrat** (engl. *Latin square*), u kojemu broj poredaka odgovara broju razina nezavisne varijable i svaka se razina nalazi u istome položaju jedanput. Shematski je prikaz toga tipa nacrti dan u tablici 2, gdje svako slovo odgovara jednoj razini nezavisne varijable, a svaki redak jednome poretku.

³⁴ Todorović (2008, str. 96) za taj način kontrole upotrebljava naziv „repeticija“, polazeći od toga da se testiranje ponavlja više puta na istome ispitaniku.

³⁵ Ti se učinci javljaju kao rezultat djelovanja ometajućih varijabli koje se odnose na redoslijed izloženosti ispitanika različitim eksperimentalnim uvjetima (usp. „sekvencijske“ varijable o kojima govori Todorović, 2008, str. 97–98).

Tablica 2. Primjer latinskoga kvadrata

a	b	c	d
b	a	d	c
c	d	a	b
d	c	b	a

Međutim, dok se u navedenom objašnjenju uravnoteživanje odnosi na ispitanike, u istraživanjima usvajanja drugoga jezika, kao i u drugim srodnim područjima u kojima je uobičajena prisutnost više podražaja po razini nezavisne varijable, još je uobičajenije da se postupci uravnoteživanja primjenjuju na podražaje, kako u nacrtima sa zavisnim tako i u onima s nezavisnim uzorcima. Naime, iako se učinak učenja nastoji izbjeći i uključivanjem distraktora, to se ne smatra dovoljnim – za pojedinačne je podražaje upravo kao i za razine nezavisne varijable iznimno važan redoslijed prikaza. Podražaji se rijetko prikazuju ispitanicima grupirani u kategorije, odnosno eksperimentalne uvjete jer to dovodi do učinka učenja i utječe na odgovore. Umjesto toga, podražaje je iz različitih kategorija, uključujući i distraktore, uobičajeno „izmiješati“, najčešće slučajno, odnosno nasumce, ili eventualno pseudoslučajno kako bi se izbjeglo da se podražaji iz istih eksperimentalnih uvjeta slučajno prikažu jedan iza drugoga (v. npr. Roberts i Liszka, 2013). Redoslijed pritom ne smije biti isti za sve ispitanike jer bi tako ponovno bilo nemoguće odvojiti eventualni učinak redoslijeda prikaza od učinka različitih struktura (Mackey i Gass, 2005, str. 143–145).

U okviru iste teme treba dodatno istaknuti **liste podražaja** (engl. *presentation lists*), koje podrazumijevaju da se isti leksički materijal u različitim listama pojavljuje u različitim eksperimentalnim uvjetima. To je naprimjer slučaj u istraživanju Roberts i Liszka (2013), gdje su ispitanici bili nasumce raspoređeni u skupine od kojih je svaka u zadatku dobila jednu od četiriju lista podražaja. Na taj su se način kontrolirali učinci leksičkoga materijala i interakcije između leksika i sintaktičkih konstrukcija, odnosno vrste rečenica. Kod lista podražaja, kao i kod ispitanika, može se primijeniti nacrt latinskoga kvadrata.

Uz već navedene elemene nacrta istraživanja, eksperimentalnim tehnikama kontrole pripada i **slučajna raspodjela ispitanika** prema eksperimentalnim uvjetima ili **randomizacija** (usp. Todorović, 2008, str. 96; engl. *random assignment*). Kako ističu Vitta i sur. (2022), za razliku od slučajnoga uzorkovanja, slučajna raspodjela dobiva dosta pozornosti u istraživanjima usvajanja drugoga jezika i glavno je sredstvo postizanja „slučajnosti“ u

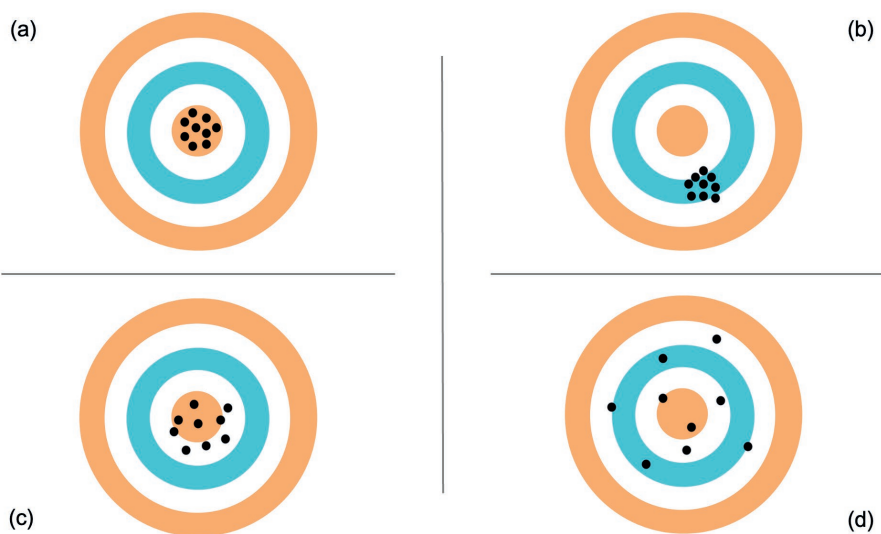
području. Međutim njezina je primjena ograničena na manipulativne varijable te je mnogobrojna istraživanja ne mogu primijeniti.

Naposljetku, u nacrtu istraživanja treba uzeti u obzir i neeksperimentalne tehnike kontrole, koje se jedine mogu primijeniti kod selektivnih varijabli, a mogu se uz to dodati i kod manipulativnih. Te se tehnike primjenjuju za svaku nezavisnu varijablu posebice kada se zna da postoje ometajuće varijable koje bi mogle utjecati na nju. Naprimjer kod podražaja je poznato da učestalost riječi/konstrukcija utječe na obradu, tj. da se učestalije jezične jedinice lakše i brže obrađuju od manje učestalih. Kako bi se neutralizirao mogući utjecaj učestalosti u istraživanjima u kojima ona nije nezavisna varijabla, ispitivane se jedinice mogu prema njoj ujednačiti, odnosno u istraživanje se mogu uključiti samo one riječi koje imaju određenu učestalost i međusobno su prema njoj podjednake; ta se tehnika naziva **homogenizacija** (Todorović, 2008, str. 101; engl. *matching*) jer se njome eliminira varijacija te se moguća ometajuća varijabla pretvara u konstantu. Nedostatak je toga pristupa to da se uzorak više ne uzima iz cijele populacije već iz jednoga njezina dijela, što ograničava i mogućnosti uopćavanja rezultata. U istraživanjima usvajanja drugoga jezika takav se pristup često primjenjuje i prema ispitanicima kad se u istraživanje uključi samo određena razina općega znanja drugoga jezika. Drugi česti načini neeksperimentalnoga kontroliranja jesu **kontroliranje blokovima** i **uprosječivanje** (v. Todorović, 2008, str. 101–104; engl. *blocking*; *averaging*). Kontroliranje blokovima podrazumijeva da unutar svake razine nezavisne varijable koja nas zanima postoje blokovi prema određenoj ometajućoj varijabli, naprimjer da i u kontrolnoj i u eksperimentalnoj skupini budu zastupljeni ispitanici obaju spolova, svuda u jednakim proporcijama. Uprosjekivanje znači naprimjer da se uzimaju riječi iz cijelog raspona učestalosti, ali tako da je prosječna učestalost jednaka u različitim eksperimentalnim uvjetima. Prva se strategija može primijeniti samo kod kategorijskih, a druga samo kod numeričkih varijabli.

Sve navedeno u ovome potpoglavlju važno je kako bi se omogućilo uopćavanje rezultata iz uzorka na populaciju, odnosno kako bi se moglo pouzdano tvrditi da su istraživane nezavisne varijable doista povezane s vrijednostima zavisne varijable koja je u središtu pozornosti. Kako će se vidjeti u sljedećim poglavljima, velik dio statističke analize vezan je uz analizu varijabilnosti u podacima i uz izdvajanje onoga njezina djela koji se može pripisati nezavisnim varijablama; teme iz ovoga potpoglavlja važne su kako bi to izdvajanje bilo moguće. Na već iznijeta razmatranja izravno se nadovezuju dodatna pitanja kontrole kvalitete istraživanja, kojima posvećujemo sljedeće potpoglavlje.

2.1.6 Kontrola kvalitete istraživanja

Sadržaj prethodnih potpoglavlja treba promatrati i u svjetlu kontrole kvalitete istraživanja, na čije središnje pojmove ovdje svraćamo pozornost. Kvaliteta eksperimentalnih istraživanja ponajprije ovisi o njihovoj **valjanosti** ili **validnosti** (engl. *validity*) i o njihovoj **pouzdanosti** (engl. *reliability*). U najkraćim crtama, valjanost znači da se u istraživanju doista mjeri ono što se tvrdi da se mjeri, dok pouzdanost znači da višestruka mjerenja daju uvijek iste rezultate, odnosno da korišteni mjerni instrument prilikom različitih mjerenja kod iste populacije i pod istim uvjetima konzistentno pokazuje (gotovo) isti podatak (v. Kraš i Miličević, 2015, str. 38). Te su značajke prikazane na slici 4, putem sheme mete u kojoj središnji krug predstavlja ono što u istraživanju želimo mjeriti, dok su točke („pogoci“) različita pojedinačna mjerenja koja pokazuju koliko smo u tome uspješni. Cilj je kojemu težimo grupiranost pogodaka u središnjem krugu (slika 4a) jer tako sva mjerenja doista mjere traženu pojavu te daju vrijednosti slične jedne drugima, dok su druge situacije manje ili više problematične – upućuju na izostanak valjanosti (slika 4b), izostanak pouzdanosti (slika 4c) ili čak izostanak obiju značajki kontrole kvalitete (slika 4d) (prilagođeno prema Puskas i sur., 2018).



Slika 4. Shematski prikaz valjanosti i pouzdanosti (a – prisutnost obiju značajki, b – prisutnost samo pouzdanosti, c – prisutnost samo valjanosti, d – izostanak obiju značajki)

U ovome ćemo potpoglavlju ponajprije govoriti o valjanosti („pogađanju središnjega kruga“), koja velikim dijelom potpada pod konceptualni okvir istraživanja, dok će pouzdanost biti glavna tema potpoglavlja 2.3, posvećenoga provjeri psihometrijskih značajki instrumenata korištenih u eksperimentu, te potpoglavlja 7.1, u kojemu će se govoriti o metodama za njezinu provjeru koje su dugoročne i povezane s pisanjem izvještaja o istraživanju. Kod pouzdanosti se naime mogu izdvojiti dvije vrste, čiji se načini provjere međusobno razlikuju (v. Kraš i Miličević, 2015, str. 39): **unutarnja pouzdanost** (engl. *internal reliability*) znači da se ponovnom analizom postojećih podatka dobivaju isti rezultati, što se može utvrditi iz prikupljenih podataka, dok **vanjska pouzdanost** (engl. *external reliability*) predstavlja **replikabilnost** (engl. *replicability*) i znači da se ponavljanjem cijeloga istraživanja u istome obliku dobivaju isti rezultati, za što je potrebno provesti nova istraživanja.

Kod valjanosti se također mogu razlikovati dvije osnovne vrste: **unutarnja valjanost** (engl. *internal validity*) i **vanjska valjanost** (engl. *external validity*). Unutarnja valjanost znači da se u istraživanju pouzdano utvrđuje uzročno-posljedični odnos između ispitivanih varijabli, tj. da se mogu isključiti druga objašnjenja za dobivene rezultate, a vanjska da se rezultati istraživanja mogu uopćiti, s uzorka na istraživanu populaciju, ali i na druge populacije, druge mjerne instrumente i druge okolnosti (Kraš i Miličević, 2015, str. 39). U eksperimentalnim istraživanjima usvajanja drugoga jezika veća se pozornost pridaje unutarnjoj valjanosti: ponajprije zbog toga što je ona preduvjet za vanjsku valjanost, a i zbog toga što je u eksperimentalnim istraživanjima usvajanja drugoga jezika vanjsku valjanost teže postići.

Unutarnju valjanost ugrožavaju prije svega vanjske varijable, koje mogu umjesto istraživanih nezavisnih varijabli (ili uz njih) izazvati promjenu u vrijednostima zavisne varijable. Kako navodimo u Kraš i Miličević (2015, str. 40), do neželjenih utjecaja mogu dovesti naprimjer osobine ispitanika, a u istraživanjima koja se odvijaju tijekom dužega razdoblja i promjene koje se u njima dogode neovisno o eksperimentu te učinak učenja unutar eksperimenta. Zbog toga se može reći da su tehnike kontrole vanjskih varijabli opisane u potpoglavlju 2.1.5 metode za zaštitu unutarnje valjanosti istraživanja. Među dodatnim tehnikama nalaze se uključivanje distraktora (v. potpoglavlje 2.1.4) te operacionalizacija iste varijable na više različitih načina i međusobna usporedba rezultata („višestruka operacionalizacija“; v. Kraš i Miličević, 2015, str. 40; Milas, 2009, str. 478–479).³⁶

³⁶ Naprimjer Kraš (2008a) istražuje usvajanje izbora pomoćnoga glagola s neprijelaznim glagolima u složenim glagolskim vremenima u talijanskome kao drugome jeziku kod izvornih

Vrsta valjanosti povezana s unutarnjom valjanošću jest i **valjanost statističkoga zaključka** (engl. *statistical conclusion validity*). Ona se odnosi na to jesu li nezavisna i zavisna varijabla uopće povezane, tj. predstavlja stupanj u kojemu je statističkim postupcima moguće pokazati takvu povezanost, što znači da predstavlja podvrstu unutarnje valjanosti ili preduvjet za nju (Milas, 2009, str. 119–120). Nju jednako kao i unutarnju valjanost ugrožavaju vanjske varijable, ali i nepouzdanost mjerenja, neadekvatnost nacrtu istraživanja i niz čimbenika vezanih uz statističke testove o kojima će biti riječi u četvrtome i petome poglavlju (prije svega niska statistička snaga, narušene pretpostavke testova i višestruke usporedbe istih uzoraka).

Za postizanje vanjske valjanosti, odnosno visokoga stupnja mogućnosti uopćavanja rezultata, pored ostvarene unutarnje valjanosti ključna je reprezentativnost uzoraka, s kojom u istraživanjima usvajanja drugoga jezika često ima poteškoća (v. posebice potpoglavlje 2.1.3, posvećeno uzorkovanju ispitanika). Stanje je u području bolje kad se uzme u obzir podvrsta valjanosti koja se naziva **konstruktna valjanost** (engl. *construct validity*), čija visoka razina znači da indikatorske varijable adekvatno predstavljaju teorijske varijable, odnosno da je operacionalizacija varijabli odgovarajuća (v. potpoglavlje 2.1.2 ove knjige; Kraš i Miličević, 2015, str. 39; Milas, 2009, str. 119–120). Višestruka operacionalizacija i ovdje predstavlja važan način postizanja valjanosti.

Konačno, jedan od općih načina provjere i postizanja valjanosti (kao i pouzdanosti) eksperimenta jest **pilotiranje** (engl. *piloting*), koje predstavlja provedbu probnoga istraživanja na manjemu uzorku ispitanika radi provjere različitih aspekata eksperimenta (ponajprije mjernih instrumenata i eksperimentalnoga postupka) i otkrivanja njihovih eventualnih nedostataka (v. Kraš i Miličević, 2015, str. 40). Kako navodimo u Kraš i Miličević (2015, str. 40–41), u eksperimentalnim istraživanjima usvajanja drugoga jezika upute za rješavanje zadataka mogu biti nejasne, podražaji mogu biti dvosmisleno formulirani, pojedini podražaji mogu odudarati od ostalih, ukupno trajanje eksperimenta može biti neadekvatno i sl. Slažemo se pritom s Mackey i Gass (2005, str. 138), koje ističu da je u fazi pripreme eksperimenta ključno voditi računa o tome omogućava li pripremljeni materijal odgovor na istraživačko pitanje, kao i s Blom i Unsworth (2010, str. 7), koje upozoravaju:

govornika hrvatskoga uz pomoć dva zadatka prosudbe prihvatljivosti rečenica – zadatka s tehnikom procjene veličine i ubrzanoga zadatka (za opis te vrste zadataka v. Kraš i Miličević, 2015, te Kraš i sur., 2023). Rezultati obaju zadataka, temeljeni na trima zavisnim varijablama (relativnim prosudbama, apsolutnim prosudbama i vremenu reakcije) upućuju na to da su učenici drugoga jezika usvojili i sintaktički i leksičko-semantički aspekt te pojave.

Nijedan eksperiment ne funkcionira savršeno otpočetak! Bolje je utrošiti više vremena na fazu planiranja, nego morati tumačiti nepotpune ili problematične podatke koji su rezultat eksperimenta koji nije bio pilotiran kako treba. (prev. M. M. P. i T. K.)

Ponekad je pogreške u eksperimentu nemoguće kompenzirati pa je jedino rješenje ponoviti istraživanje. Međutim ponavljanje istraživanja obično je i vremenski i financijski vrlo zahtjevno, a u nekim slučajevima i nemoguće, stoga je pogreške bolje otkriti pilotiranjem.

2.2 Prikupljanje i priprema podataka

Nakon što je istraživanje osmišljeno, a prije negoli se pristupi analizi podataka, potrebno je, naravno, podatke prikupiti. Kao što smo prethodno navele, podaci se prikupljaju kako bi se pokušalo odgovoriti na postavljena istraživačka pitanja i kako bi se dobile informacije koje govore u prilog postavljenoj istraživačkoj hipotezi ili protiv nje. Prikupljene podatke potrebno je uz to sačuvati na primjeren način, imajući u vidu analize koje se žele provesti. U sljedećim potpoglavljima govorimo najprije o prikupljanju, zatim o kodiranju i unosu podataka te o njihovoj provjeri i organizaciji, kao i o izboru programa za statističku analizu podataka.

2.2.1 Eksperimentalni zadaci i podaci koji se u njima dobivaju

Kako je već navedeno u potpoglavlju 1.1, u lingvistički orijentiranim eksperimentalnim istraživanjima usvajanja drugoga jezika ispitanici su izloženi jezičnim zadacima. Podaci koje treba analizirati njihovi su odgovori na zadatke, točnije na konkretne podražaje koji su u zadacima prisutni. Drugim riječima, podatak predstavlja vrijednost neke varijable za određenoga ispitanika i/ili određeni podražaj (v. Tošković, 2020, str. 73). Točna će priroda podataka izravno ovisiti o prirodi zadatka i načinu na koji su vrijednosti izmjerene.

Niz različitih eksperimentalnih zadataka koji se upotrebljavaju u istraživanjima usvajanja drugoga jezika temeljito smo opisale u četvrtome poglavlju Kraš i Miličević (2015), dok su nastavne materijale za pouku studenata njihovoj izradi sastavili Kraš i sur. (2023). Kao često korišteni zadaci razlikuju se: (1) **zadaci proizvodnje i oponašanja**, kojima se dobivaju podaci o tome koje riječi ili strukture učenici drugoga jezika mogu upotrijebiti u drugome jeziku (u to ulaze naprimjer zadatak potaknute proizvodnje ili zadatak povezivanja rečenica), (2) **zadaci razumijevanja i interpretacije**,

kojima se ispituju receptivne vještine u drugome jeziku, odnosno ispituje se tumače li učenici eksperimentalne čestice slično kao izvorni govornici (npr. takvi su zadatak prosudbe istinitosti rečenica i zadatak odabira slike) te (3) **zadaci prosudbe prihvatljivosti rečenica**, u kojima se od ispitanika traži da procijene jesu li zadane rečenice prihvatljive u danome drugom jeziku. Većina se zadataka može zadati u pisanome i u usmenome obliku. Podaci koji se dobivaju u tim zadacima mogu biti međusobno vrlo različiti i može se susresti sve od binarnih vrijednosti tipa „da“/„ne“ ili „točno“/„netočno“ (kako se često klasificiraju odgovori koje ispitanici daju u zadacima proizvodnje), preko odabira jedne od ponuđenih kategorija (naprimjer u zadacima proizvodnje u kojima je dopušteno više opcija ili u zadacima odabira slike), do prosudbi na ordinalnim ili intervalnim ljestvicama prihvatljivosti (u različitim varijantama zadataka prosudbe prihvatljivosti rečenica). U nekim slučajevima odgovori ispitanika izravno postaju podaci za analizu, dok je u drugima istraživač posrednik koji odgovorima ispitanika dodjeljuje vrijednosti (poput „točno“/„netočno“) koje će se poslije statistički analizirati.

Gornjim se zadacima mogu dodati i različiti **psiholingvistički zadaci** (v. posebice Covey i Gabriele, 2023; Jegerski i VanPatten, 2014; Marsden, Thompson i Plonsky, 2018; McDonough i Trofimovich, 2009). I ti zadaci mogu biti usredotočeni na jezičnu proizvodnju, poput vremenski kontroliranoga zadatka dopunjavanja rečenica (engl. *sentence completion task*, v. npr. Jackson i sur., 2018), ili razumijevanje jezika, kao u slučaju zadatka leksičke odluke, u kojemu ispitanici trebaju pritiskom tipke označiti smatraju li da niz znakova koji vide jest ili nije riječ u danome jeziku, uz udešavanje (engl. *priming*) ili bez njega, ili zadataka čitanja, posebice vlastitim tempom. Ključna je značajka većine psiholingvističkih zadataka prikupljanje podataka o brzini davanja odgovora ili brzini čitanja, tj. o vremenu reakcije ispitanika na različite eksperimentalne podražaje (engl. *reaction time*), najčešće izraženom u milisekundama. Kod nekih se pak zadataka ograničava trajanje prikaza podražaja. U svakome se slučaju želi izmjeriti ponašanje ispitanika u stvarnome vremenu te na temelju toga izvesti zaključke o njihovu implicitnom znanju drugoga jezika, a u većini se psiholingvističkih zadataka glavnim podacima za analizu smatraju upravo oni koji se odnose na vrijeme reakcije.

U daljnjim ćemo se poglavljima na različite zadatke vraćati putem konkretnih primjera provedenih statističkih analiza. Uz već citiranu literaturu, za detaljnije opise eksperimentalnih zadataka u istraživanjima usvajanja drugoga jezika upućujemo i na: Gass i Mackey (2007), Gudmestad i Edmonds (2018), Mackey i Gass (2005) te Mackey i Gass (2023).

2.2.2 Kodiranje i organizacija podataka

S obzirom na to da se analiza podataka gotovo uvijek provodi uz pomoć računala, prva faza u svakoj kvantitativnoj analizi nužno je prebacivanje u digitalni format ako podaci već izvorno nisu prikupljeni na računalu. Daljnji potrebni koraci ovisit će o prirodi prikupljenih podataka i o planiranim analizama.

Unos podataka u digitalni format zahtijeva njihovo odgovarajuće **kodiranje** i **organizaciju**, pri čemu se podrazumijeva i temeljita provjera podataka (jesu li svi ispitanici napravili sve što se od njih tražilo, jesu li svi dobiveni odgovori valjani i sl.).³⁷ Ti su koraci znatno olakšani u slučajevima kad su u eksperimentu korišteni *online*-upitnici poput Google obrazaca, ili softverskih/*online*-platformi za eksperimente, naprimjer PsychoPy, Open Sesame ili PsyToolkit,³⁸ koji bilježe odgovore automatski i imaju opciju njihova izvoza u programe za analizu. Podaci koji su zabilježeni *offline*, u pisanoj formi ili kao zvučni zapis, moraju se prenijeti u odabrani program za analizu, često ručno. S praktične strane, treba imati na umu da faza „sređivanja“ podataka može trajati dugo i može, s obzirom na koncentraciju koju zahtijeva, biti vrlo naporna za istraživače.

Dobra je praksa prilikom unosa podataka dodjela jedinstvenih šifri ispitanicima kako bi bilo jasno kome podaci pripadaju, a da se pritom ne oda identitet ispitanika. Također treba pazljivo imenovati varijable koje se analiziraju. Većina programa za statističku analizu nameće ograničenja naziva koji se mogu upotrebljavati (bilo u pogledu dužine naziva, uporabe razmaka ili prisutnosti znakova koji nisu alfanumerički), zbog čega se varijablama često moraju dodijeliti skraćeni nazivi. Međutim treba voditi računa da naziv bude dovoljno razumljiv ili zabilježen i objašnjen na dostupnome mjestu kako bi se podaci mogli bez poteškoća analizirati i naknadno.

Dvije napomene formalne prirode vezane uz unos podataka odnose se na decimalne oznake i pisanje nule u decimalnim brojevima manjima od jedan. Različiti jezici imaju različite konvencije u pogledu zapisivanja decimalnih brojeva. U hrvatskome jeziku uobičajena je decimalna oznaka zarez, dok je na engleskome govornom području u uporabi decimalna točka (uz pojedine iznimke, poput Južnoafričke Republike). To se očituje i u računalnome predstavljanju podataka, gdje različite lokalizacije računalnoga operativnog

³⁷ Početne faze u analizi podataka nazivaju se i „pretprocesiranjem“ (engl. *preprocessing*) (v. Schütze i Sprouse, 2013).

³⁸ <https://www.psychopy.org>, <https://osdoc.cogsci.nl>, <https://www.psychotoolkit.org> (posljednji pristup 26. 4. 2024.)

sustava automatski povlače različito prikazivanje decimalnih brojeva u statističkim i srodnim programima. O tome treba posebno voditi računa ako se radi na različitim računalima, koja imaju različite lokalizacije, ili ako se podaci kopiraju iz lokaliziranoga sustava u *web*-servis poput VassarStats (v. potpoglavlje 2.2.3), koji prihvaća samo zapis s decimalnom točkom. Također treba imati na umu da se u knjigama i radovima na jezicima koji upotrebljavaju decimalni zarez ipak može naići i na uporabu točke (to potvrđuju primjeri za hrvatski jezik i za srpski jezik: dok Milas, 2009, upotrebljava zarez, Petz i sur., 2012, upotrebljavaju decimalnu točku, kao i Todorović, 2008). Dodatno, u engleskome jeziku uobičajeno je da se prilikom prikaza decimalnih brojeva čija vrijednost mora biti između 0 i 1 (ili između -1 i 0), odnosno koji imaju tzv. „vodeću nulu“ (engl. *leading zero*), ta nula izostavi te da se naprimjer 0.5 predstavi kao .5 (tako nalažu primjerice preporuke Američkoga psihološkog udruženja, v. Lindstromberg, 2016a). U hrvatskome jeziku to je manje uobičajeno nego u engleskome, ali također prisutno. U ovoj ćemo se knjizi držati načela navođenja decimalnih brojeva uz uporabu decimalnoga zarez a i uz navođenje početne nule (0,5).

Kodiranje podataka podrazumijeva dodjeljivanje vrijednosti odgovorima koje su ispitanici dali u eksperimentalnome zadatku. Kodiranje nekad može značiti kopiranje ili doslovno prepisivanje odgovora (naprimjer u slučaju prosudbi prihvatljivosti na unaprijed definiranim ljestvicama), ali često obuhvaća i dodatne korake, poput dodjeljivanja numeričkih vrijednosti razinama kategorijskih varijabli (naprimjer skupina izvornih govornika može biti označena s 0, a skupina učenika s 1), što zahtijevaju pojedini programi za statističku analizu.³⁹ Također je vrlo često da se zbog balansiranja redosljeda prikaza različitih opcija u zadacima poput onih temeljenih na višestrukome izboru ista oznaka kod različitih podražaja odnosi na različite kategorije odgovora, što je potrebno urediti prilikom kodiranja. Naprimjer u istraživanjima koje su provele Kraš (2008b, 2016) te Miličević i Kraš (2017), posvećenima anaforičkim odnosima u talijanskome jeziku, korišten je zadatak odabira slike u kojemu je zavisna varijabla bila izbor referenta za izrečene i neizrečene subjekte zavisnih surečenica, ali izbor slike broj 1 nije uvijek imao isto značenje. Na slici 5, pod (a) vidimo da opcija 1 označava izbor subjekta glavne surečenice, opcija 2 izbor objekta glavne surečenice i opcija 3 izbor izvanlingvističkoga referenta. Međutim pod (b) raspored je: objekt glavne

³⁹ Među programima obuhvaćenima u potpoglavlju 2.2.3, IBM SPSS za izradu velikoga broja testova zahtijeva da vrijednosti budu prikazane numerički (v. npr. Field, 2018), dok za R to nije slučaj.

surećenice (1), subjekt glavne surećenice (2) i izvanlingvistički referent (3).⁴⁰

a. *La mamma rimprovera la figlia mentre lei lava i piatti.*

‘Majka kori kćer dok ona pere posuđe.’



1



2



3

b. *L'insegnante sorride alla bambina bionda mentre lei suona il violino.*

‘Nastavnica se smiješi djevojčici svijetle kose dok ona svira violinu.’



1



2



3

Slika 5. Primjeri podražaja iz istraživanja Kraš (2008b, 2016) te Miličević i Kraš (2017)

Dodatno, u nekim je slučajevima potrebna i standardizacija podataka, primjerice kad se upotrebljava tehnika procjene veličine (engl. *Magnitude Estimation*), u kojoj ispitanici sami određuju brojeve koje će upotrebljavati u prosudbi prihvatljivosti rečenica tako što prvoj rečenici (standardu) dodijele broj po želji, a daljnje brojeve određuju u odnosu na prvi ovisno o tome koliko im je rečenica manje ili više prihvatljiva nego prva (v. Bard i sur., 1996). Prije nego što se podaci prikupljeni tom tehnikom podvrgnu statističkoj analizi, potrebno je za svakoga ispitanika pojedinačno podijeliti brojeve dodijeljene pojedinim rečenicama brojem dodijeljenim standardu kako bi se rezultati različitih ispitanika sveli na zajedničku ljestvicu.

⁴⁰ Crteže na slici 5 izradila je Maja Dinić za potrebe istraživanja Kraš (2008b), i ovdje su uključeni uz njezinu suglasnost.

U vezi s organizacijom podataka važno je podsjetiti na razliku između neposrednih mjera s jedne i posrednih ili izvedenih mjera s druge strane (v. potpoglavlje 2.1.4). Prvima pripadaju one vrijednosti koje se dobivaju na temelju jednoga mjerenja, dok se druge izvode iz više neposrednih mjera računanjem njihove mjere centralne tendencije, zbroja ili postotka (v. npr. Todorović, 2008, str. 135). U području usvajanja drugoga jezika neposrednim mjerama pripada naprimjer dob ispitanika. Primjeri su posrednih mjera prosječne ili medijalne (v. potpoglavlje 3.1.2) prosudbe prihvatljivosti, gdje se rezultat svakoga ispitanika za određeni eksperimentalni uvjet dobiva kao mjera centralne tendencije iz većega broja pojedinačnih vrijednosti. Naprimjer tablica 3 pokazuje dio vrijednosti iz istraživanja koje su provele Miličević i Kraš (2017) – u pitanju su izbori referenta pet ispitanika za jedan eksperimentalni uvjet (anafora s izrečenom zamjenicom, kao u primjerima na slici 5), gdje odgovor 1 označava izbor subjekta glavne surečenice, odgovor 2 izbor objekta glavne surečenice, a odgovor 3 izbor izvanlingvističkoga referenta, dok se kao izvedena mjera izračunava postotak izbora subjekta glavne surečenice, koji se upotrebljava u daljnjoj analizi.

Tablica 3. Primjer organizacije odgovora ispitanika (prema Miličević i Kraš, 2017)

Ispitanik	Slika 1	Slika 2	Slika 3	Slika 4	Slika 5	Slika 6	% subjekt
TT1	2	2	2	2	2	2	0,00
TT2	2	1	2	2	2	2	16,67
TT5	2	2	2	2	2	2	0,00
TT26	2	3	3	1	2	2	16,67
TT31	2	3	1	1	2	3	33,33

Kako je već navedeno u prethodnim potpoglavljima, mjere izvedene iz više neposrednih mjerenja pouzdanije su od jednoga mjerenja i u lingvističkim je istraživanjima uobičajena praksa da se tako postupa, odnosno da se obavezno uključi više čestica u svakome eksperimentalnom uvjetu. Takvim višestrukim odgovorima osigurava se bolji uvid u istinsko znanje ili prosudbu ispitanika i umanjuje se opasnost od dobivanja odgovora koji ispitanici daju slučajno ili pogreškom, posebice u zadacima koji uključuju pritiskanje tipke poput zadatka leksičke odluke ili zadataka čitanja vlastitim tempom (v. npr. Marsden, Thompson i Plonsky, 2018; McDonough i Trofimovich, 2009). Slično je s rezultatima testova razine općega znanja drugoga jezika, gdje se pojedinačna pitanja ili čestice obično ne uzimaju u obzir zasebno, već je važan

zbroj pojedinačnih rezultata (naprimjer u testovima dopunjavanja teksta, v. npr. Miličević, 2007).

U programe za statističku analizu kakvi su IBM SPSS i R mogu se unijeti posredne mjere prethodno izračunate iz neposrednih (naprimjer uporabom Microsoftova Excela), ali dobro je sačuvati informacije i o neposrednim mjerama. Naime, ne samo da se posredne mjere mogu izračunavati i u tim programima⁴¹ već postoje i statistički testovi za koje je potrebno uključiti sve pojedinačne rezultate – takvi su prije svega mješoviti modeli (engl. *mixed models*), o kojima će kratko biti riječi u potpoglavlju 5.5. Njihova se uporaba vezuje uz razmatranja iz potpoglavlja 2.1.4, prema kojima izračunavanje posrednih mjera i zasnivanje statističkih testova na njima predstavlja gubitak informacija o varijabilnosti između čestica, zbog čega je bolje u analizi zadržati neposredne mjere kad god je to moguće.

Organizacija podataka svakako se mora prilagoditi zahtjevima odabranoga statističkog programa. Različiti programi mogu naime zahtijevati različite formate podataka, odnosno različite **obradne tablice** (engl. *data tables*). Tipičan je način unosa nezavisnih uzoraka takav da jedan redak u tablici predstavlja jednoga ispitanika, a jedan stupac jednu varijablu; razine varijable iskazane su različitim vrijednostima u stupcu (naprimjer oznakama poput „učenici“ i „izvorni_govornici“, ili brojevima 1 i 2 ako su prisutne dvije skupine ispitanika). Naravno, ako se unose višestruke neposredne mjere, pojedinačni stupci mogu predstavljati i vrijednosti za pojedinačne eksperimentalne čestice (kao u tablici 3).

Kod prikaza zavisnih uzoraka jedan je redak također jedan ispitanik, ali budući da postoji više podataka istoga ispitanika, različiti stupci mogu predstavljati različite razine nezavisne varijable. Među programima spomenutima u potpoglavlju 2.2.3, IBM SPSS uvijek primjenjuje to načelo unosa zavisnih uzoraka, poznato kao **široki format** (engl. *wide format*). R ovisno o analizi može zahtijevati i unos podataka koji je više nalik na onaj za nezavisne uzorke, poznat kao **dugački format** (engl. *long format*), gdje stupci predstavljaju varijable, a isti ispitanik zauzima više redaka, po jedan za svaku razinu nezavisne varijable. Kao ilustracija tih mogućnosti, u tablici 4 isti podaci, iz istraživanja koje su provele Miličević i Kraš (2017), prikazani su na dva različita načina (FAOP i FANP su kratice naziva eksperimentalnih uvjeta – engl. *forward anaphora with an overt subject* i *forward anaphora with a*

⁴¹ Izračunavanje podrazumijeva dodavanje novoga stupca u tablicu s podacima i zadavanje računalne operacije kojom će se izvesti odgovarajuća mjera; IBM SPSS za to upotrebljava opciju *Transform > Compute Variable*, dok R zahtijeva ispisivanje funkcije kojom će se dodati novi stupac i izračunati potrebne vrijednosti.

null subject, odnosno anafora s izrečenim i neizrečenim subjektom). Važnost izbora odgovarajućega formata podataka možda posebno treba istaknuti kako bi se usporedio s programima poput Microsoftova Excela i LibreOffice Calca, koji nemaju toliko strogo nametnut način unosa podataka, ali u kojima je u svrhu postizanja kompatibilnosti s drugim programima bolje pridržavati se istoga načela.

Tablica 4. Prikaz podataka za zavisne uzorke u (a) širokome i (b) dugačkome formatu (prema Miličević i Kraš, 2017)

(a)

Ispitanik	FAOP	FANP
TT1	0,00	100,00
TT2	16,67	66,67
TT5	0,00	66,67
TT26	16,67	50,00
TT31	33,33	66,67

(b)

Ispitanik	Uvjet	% subjekt
TT1	FAOP	0,00
TT1	FANP	100,00
TT2	FAOP	16,67
TT2	FANP	66,67
TT5	FAOP	0,00
TT5	FANP	66,67
TT26	FAOP	16,67
TT26	FANP	50,00
TT31	FAOP	33,33
TT31	FANP	66,67

Primjer dugačkoga formata podataka upotrebljavamo i kako bismo upozorili na to da se kategorijske varijable u pripremi podataka unose u tablice prema istim načelima kao i numeričke. Tablica s dvije kategorijske varijable izgledala bi isto kao tablica 4b, samo što bi u posljednjemu stupcu umjesto brojeva bile navedene kategorije.

Prilikom organizacije i kodiranja podataka posebnu vrstu problema predstavljaju **nedostajući** ili **nepostojeći podaci** (engl. *missing data*) i **netipične**, **ekstremne** ili **stršeće vrijednosti** (engl. *outliers*). U ovome

potpoglavlju govorimo o prvoj vrsti, koja se pronalazi i rješava upravo u fazi kodiranja i organizacije podataka, dok ćemo se netipičnim vrijednostima baviti u potpoglavlju 3.1.6 jer je za njihovo otkrivanje potrebna primjena deskriptivnih statističkih mjera, o kojima još nismo govorile.⁴²

Među nedostajućim podacima mogu se uočiti oni koji doslovce ne postoje, zato što ispitanici nisu odgovorili na neku eksperimentalnu česticu, i podaci koji se moraju isključiti iz analize zbog toga što zabilježeni odgovor nije relevantan (naprimjer u zadatku leksičke odluke brzina čitanja od 5 milisekundi jasno upućuje na nasumično pritiskanje tipke jer je za toliko vremena nemoguće doista pročitati riječ) ili je isključen kao netipična vrijednost (v. potpoglavlje 3.1.6). U pogledu najboljega pristupa nedostajućim podacima vrlo je teško dati jedinstvenu preporuku, stoga odluku, svakako, treba donijeti imajući u vidu uzrok izostanka podataka, ali i statističke testove koji se planiraju provesti i program koji će za to biti korišten.

U pogledu uzroka izostanka Lachaud i Renaud (2011) razlikuju **potpuno nasumično nedostajuće podatke** (engl. *data missing completely at random*), **nasumično nedostajuće podatke** (engl. *data missing at random*) i **nenasumično nedostajuće podatke** (engl. *data not missing at random*). Prvome tipu pripadaju podaci koji nedostaju zbog nekoga posve slučajnoga razloga (naprimjer ispitanik je bio rastresen i preskočio je jednu eksperimentalnu česticu). Drugome tipu pripadaju oni podaci za čiji je izostanak razlog samo prividno slučajan i na temelju skupa podataka moguće je objasniti zbog čega je određena vrijednost izostala (naprimjer jedna je čestica bila teža od ostalih, što je dovelo do većega broja netipičnih vrijednosti, koje su uklonjene; v. i Gries 2013b, str. 350–351, gdje je predložena regresijska analiza nedostajućih podataka kako bi se utvrdilo koji ih čimbenici predviđaju). Trećemu tipu pripadaju podaci kod kojih postoji prikriven razlog izostanka, odnosno razlog nije nasumičnost, ali se ne može ni pouzdano utvrditi iz podataka. U slučaju nasumično nedostajućih podataka može se razmotriti isključivanje ispitanika i/ili čestica kod kojih izostaju podaci. Ponekad se preporučuje i da se, ako je izostalih vrijednosti manje od 5 %, izbace kompletni slučajevi (tj. ispitanici), a ako je takvih podataka više od 5 %, da se oni tretiraju kao posebna kategorija u analizi, ili da se, posebice ako je izostalih podataka više od 15 %, čitava varijabla isključi iz analize (Fajgelj, 2014, str. 472). Kod nenasumičnih uzroka međutim takvo isključivanje nije opravdano. Tada se podaci mogu

⁴² Gries (2013b, str. 291, 350) izdvaja kao posebnu vrstu i **podatke s velikom težinom** (engl. *points with high leverage*), koji ne pripadaju netipičnim vrijednostima, ali mogu znatno utjecati na rezultate, no njima se u ovoj knjizi nećemo baviti.

„rekonstruirati“, odnosno praznine se mogu ispuniti relevantnim vrijednostima, čemu postoji više mogućih pristupa. Najjednostavniji je unos neke od mjera centralne tendencije (dobivene iz cijele skupine ispitanika, drugih relevantnih odgovora istoga ispitanika ili drugih odgovora na istu eksperimentalnu česticu).⁴³ Pouzdanijim pristupom smatra se primjena nekoga od algoritama koji iz preostalih podataka procjenjuju najbolju moguću zamjensku vrijednost (Lachaud i Renaud, 2011, spominju neke od algoritama; njima se ovdje nećemo baviti te pozivamo čitatelje da se samostalno o njima informiraju).

Konačna odluka ovisit će, naravno, i o konkretnome istraživanju i onome što se u njemu ispituje, a svakako je vrlo važno u izvještaju o istraživanju jasno navesti što je učinjeno s nedostajućim podacima (v. potpoglavlje 6.5.1). Naprimjer u istraživanju koje je provela Kraš (2008a), u odjeljku u kojemu se opisuje priprema podataka za statističku analizu, navedeno je da su u zadatku prosudbe prihvatljivosti rečenica s tehnikom procjene veličine nedostajući podaci nadomješteni mjerom centralne tendencije ispitanika za pojedini eksperimentalni uvjet. U istome je odjeljku navedeno da su u ubrzanome zadatku prosudbe prihvatljivosti rečenica nedostajući podaci nadomješteni na različite načine ovisno o vrsti podataka: kategorijske (binarne) vrijednosti prosudbi prihvatljivosti rečenica u jednome su aspektu statističke analize nadomještene mjerom centralne tendencije za eksperimentalni uvjet, dok su kontinuirane vrijednosti brzine odgovora najprije standardizirane (odnosno pretvorene u z-vrijednosti, v. potpoglavlje 3.1.4, za svaku skupinu eksperimentalnih čestica posebno) da bi onda mjera centralne tendencije skupine ispitanika za pojedini eksperimentalni uvjet bila upotrijebljena za nadomještanje nedostajućih podataka.

Sve tipove nedostajućih podataka potrebno je na odgovarajući način signalizirati programima za statističku analizu, zadajući im uputu u pogledu toga kako da ih tretiraju. IBM SPSS u analizama zanemaruje prazna polja, odnosno sustavno ih tretira kao nedostajuće podatke, koji ne ulaze u izračunavanje statističkih pokazatelja i jedino se prikazuju kao izdvojena stavka („*missing*“) u pregledu frekvencija različitih zabilježenih vrijednosti. Postojeće ali neodgovarajuće vrijednosti u tome je programu potrebno ručno definirati kao nedostajuće (v. Petz i sur., 2012, str. 439–441). S druge strane, R automatski dodjeljuje praznim poljima vrijednost „*NA*“ (od engl. *non-applicable*,

⁴³ Pichette i sur. (2015) navode da se u testiranju u području usvajanja drugoga jezika u upitnicima utemeljenima na *da/ne* odgovorima primjenjuje i praksa unosa 0 (u značenju „ne“) umjesto odgovora koji nedostaju. Međutim ti autori ističu da ta praksa, kao i isključivanje ispitanika i ostavljanje praznih polja, bitno utječe na koeficijente pouzdanosti te da je bolje upotrebljavati zamjene srednjim vrijednostima.

odnosno „nije primjenljivo“) i pruža više opcija za zanemarivanje tih polja prilikom izračunavanja (v. Levshina, 2015, str. 46–47); ako se ne upotrijebe te opcije, rezultat računalnih operacija nad podacima koji sadrže prazna polja korisniku će se prikazati kao „NA“. Nerelevantne podatke istraživač mora i tu sam označiti kao nedostajuće.

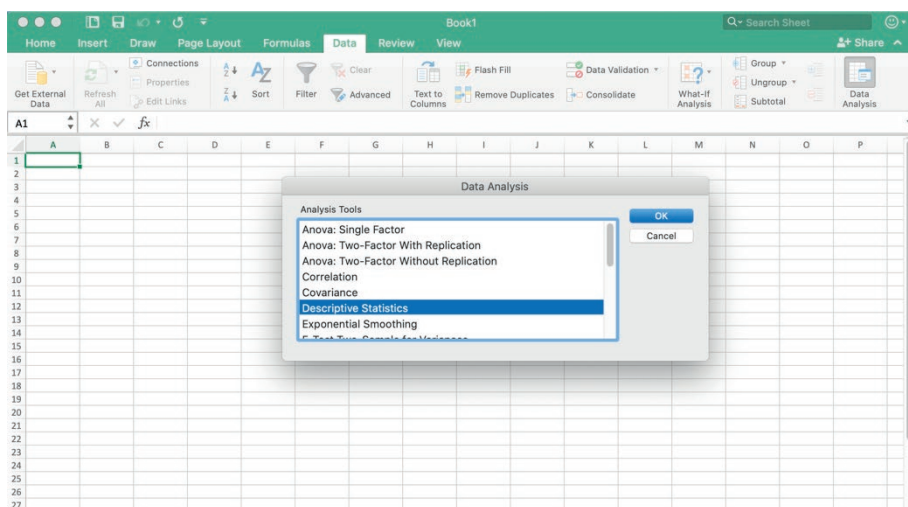
Konačno, kad su podaci u potpunosti pripremljeni za analizu, treba razmišljati ne samo o tome kako ćemo ih analizirati i rezultate opisati u izvještaju o istraživanju, već i o tome da podatke učinimo dostupnima znanstvenoj zajednici. Naime, kao i s instrumentima za prikupljanje podataka, kad god je to moguće, treba težiti objavljivanju cjelokupnih „sirovih“ podataka kako bi ih i drugi istraživači mogli upotrijebiti, u svrhe ponovne analize, replikacije istraživanja ili metaanalize.⁴⁴ U području usvajanja drugoga jezika ta je praksa novijega datuma, ali nedostupnost podataka iz ranijih istraživanja ne treba nas spriječiti u objavljivanju naših. Pritom treba shvatiti da svrha eventualnih ponovljenih analiza ili replikacija nije osobne prirode, već treba poslužiti kumulativnome napredovanju cijeloga područja. „Sirovi“ se podaci mogu učiniti dostupnima na mrežnim stranicama časopisa u kojemu se rad objavljuje (velik broj časopisa nudi takvu mogućnost), u namjenskim repozitorijima kakav je IRIS (*Instruments and Data for Research in Language Studies*) – baza jezičnih mjernih instrumenata, podražaja i podataka, na specijaliziranim platformama kakve su FigShare, Open Science Framework te u Hrvatskoj PUH – sustav za pohranu i upravljanje podacima (u Srbiji neki od repozitorija navedenih na Nacionalnom portalu otvorene nauke), ili na općim platformama za dijeljenje softvera i drugih materijala poput GitHub-a.⁴⁵ Za objavljivanje podataka u pojedinim časopisima, poput *Studies in Second Language Acquisition* i *The Modern Language Journal*, može se dobiti i nagrada u obliku značke, koja potvrđuje da autori slijede načela otvorene znanosti (<https://www.cos.io/initiatives/badges>). Više na tu temu uskoro će se moći pročitati u zborniku koji priređuje L. Plonsky (u pripremi); v. također Marsden i Plonsky (2018). U otvorenome pristupu mogu se naći, između ostaloga, i protokoli kodiranja podataka korišteni u objavljenim radovima (v. posebice repozitorij IRIS).

⁴⁴ Situacije u kojima nije moguće podijeliti podatke odnose se prije svega na privatnost i zaštitu podataka ispitanika. U najvećemu broju slučajeva međutim taj se problem može riješiti anonimiziranjem podataka i traženjem suglasnosti ispitanika za objavljivanje anonimiziranih podataka još prilikom provedbe eksperimenta.

⁴⁵ <https://www.iris-database.org>, <https://figshare.com>, <https://osf.io>, <https://www.srce.unizg.hr/puh>, <http://www.open.ac.rs/index.php/repozitorijumi>, <https://github.com> (posljednji pristup 26. 4. 2024.).

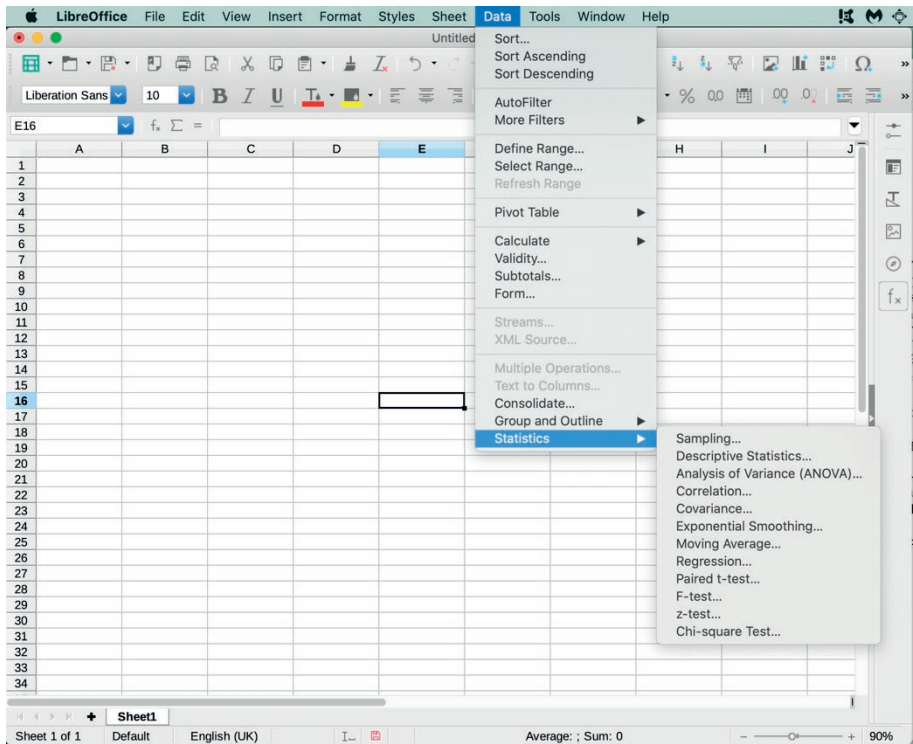
2.2.3 Izbor programa za analizu

Detaljnije bavljenje programima za statističku analizu izlazi iz okvira ove knjige te ćemo se u ovome potpoglavlju samo kratko osvrnuti na najčešće korištene resurse. Za unos i čuvanje podataka vrlo se često upotrebljava program **Microsoft Excel**, koji ujedno omogućuje i obavljanje velikoga broja matematičkih operacija na podacima (od jednostavnijih, poput zbrajanja i izračunavanja aritmetičke sredine, do složenijih, poput predviđanja novih vrijednosti na temelju postojećih), kao i grafičko prikazivanje podataka putem različitih vrsta grafikona. Uz to aktivacijom dodatka „Analysis ToolPak“ dobiva se mogućnost objedinjenoga izračunavanja većega broja deskriptivnih mjera, kao i izrade nekoliko često korištenih statističkih testova (v. sliku 6).⁴⁶ Slično vrijedi i za slobodno dostupan program **LibreOffice Calc**, koji ima unaprijed instaliran izbornik „Statistics“ za provedbu osnovnih statističkih analiza (v. sliku 7).



Slika 6. Dodatak za statističku analizu u programu Microsoft Excel

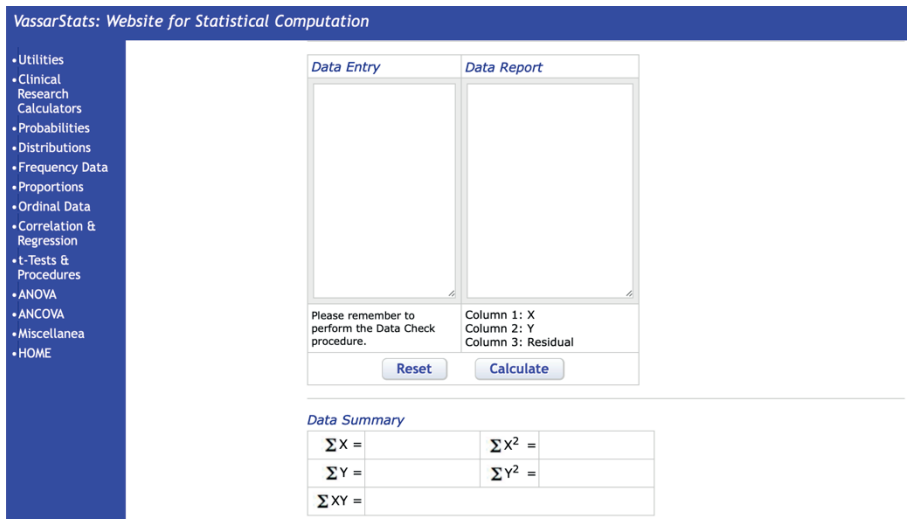
⁴⁶ Upute za aktiviranje toga dodatka u različitim verzijama programa dostupne su na službenim stranicama softverskoga paketa Microsoft Office: <https://support.microsoft.com/en-us/office/use-the-analysis-toolpak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6> (posljednji pristup 26. 4. 2024.).



Slika 7. Dodatak za statističku analizu u programu LibreOffice Calc

Nešto više mogućnosti u pogledu izrade statističkih testova pružaju *web*-aplikacije kakva je naprimjer **VassarStats: Website for Statistical Computation**.⁴⁷ Ta aplikacija ne pruža mogućnost čuvanja podataka, ali je u nju moguće ručno unijeti, kopirati ili uvesti podatke za analizu i na njima provesti niz testova (v. sliku 8). Aplikacija pruža i mogućnost izrade nekoliko osnovnih vrsta grafikona, a prati je i *online*-udžbenik, u kojemu se objašnjavaju testovi i pojmovi važni za njihovo razumijevanje (Lowry, bez dat.).

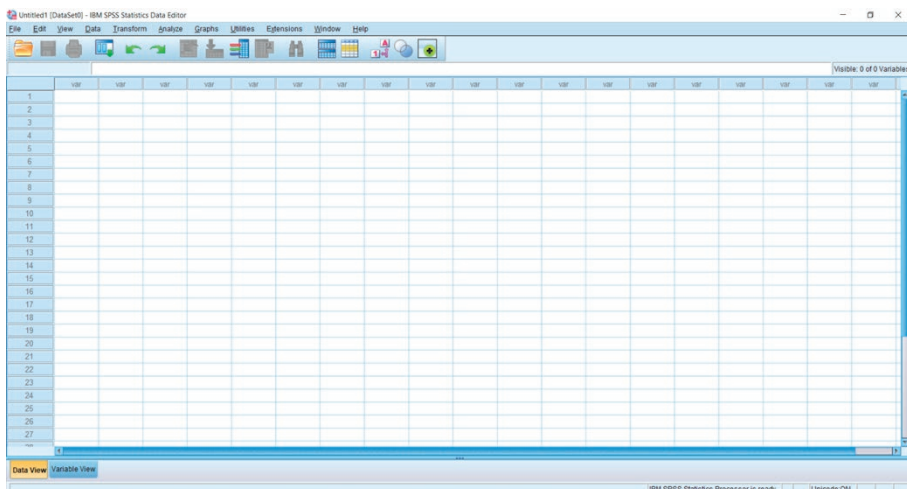
⁴⁷ <http://vassarstats.net/> (posljednji pristup 26. 4. 2024.)



Slika 8. Web-aplikacija VassarStats

Za iscrpnije analize, posebice u području statističkih testova, potrebno je upotrebljavati namjenske programe. U društvenim je znanostima vrlo proširena uporaba softvera **IBM SPSS**⁴⁸ (v. sliku 9). Prema podacima koje su među istraživačima usvajanja drugoga jezika prikupili Loewen i sur. (2014, str. 372), to je u jednome razdoblju bio najčešće korišten program za statističku analizu u istraživačkoj zajednici u tome području, s čak 69 % korisnika (ispred 56 % za Microsoft Excel, 17 % za ručno izračunavanje i 15 % za R, uz napomenu da je bilo moguće dati više odgovora). Mnogim istraživačima on može biti nedostupan zbog svoje komercijalne prirode, no treba provjeriti eventualnu dostupnost već postojeće institucijske licencije.

⁴⁸ <https://www.ibm.com/analytics/spss-statistics-software> (posljednji pristup 26. 4. 2024.)



Slika 9. Sučelje programa IBM SPSS

S druge strane, **R**⁴⁹ (v. sliku 10) je slobodno dostupan računalni program, koji kao veliku prednost ima i to što su u njegovu izradu u velikoj mjeri uključeni upravo lingvisti, zbog čega postoje posebni paketi namijenjeni lingvističkim analizama, kao i dokumentacija bogata lingvističkim primjerima. Otvorenost programskoga koda također dovodi do toga da postoje mnogobrojni *web*-forumi i drugi komunikacijski kanali putem kojih se može potražiti pomoć i u kojima se može pronaći kôd koji su sastavili drugi ili podijeliti vlastiti, čime se pridonosi sve prisutnijemu pokretu za otvorenu znanost.

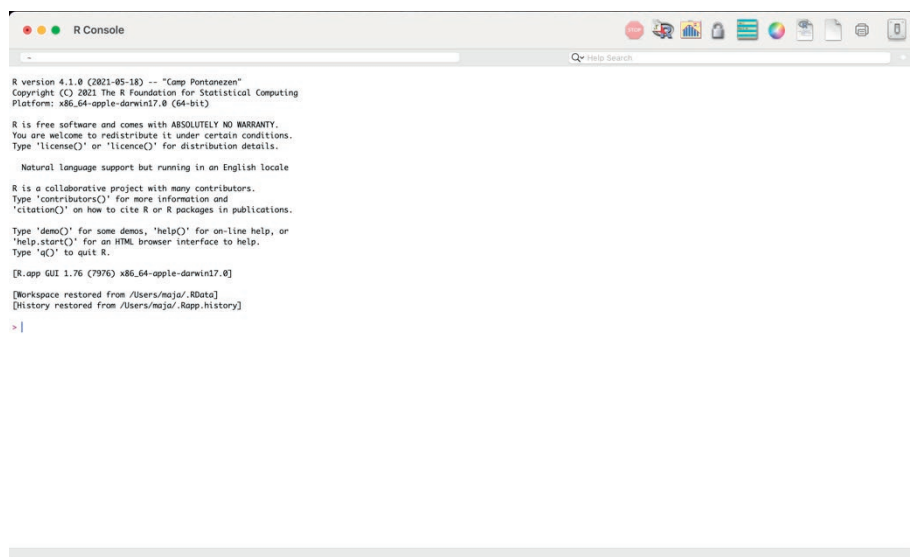
Mnogi se lingvisti teško odlučuju na uporabu R-a zbog nužnosti unosa naredaba u obliku računalnoga koda (što se samo donekle može izbjeći uporabom grafičkih sučelja).⁵⁰ Sudeći prema podacima iz već citiranoga istraživanja Loewen i sur. (2014), i u usvajanju drugoga jezika taj se program do prije 10 godina relativno rijetko upotrebljavao. Međutim već niz godina postoje važne inicijative da se on približi istraživačima i u tome području, ne samo isticanjem toga da nakon savladavanja početnih poteškoća pruža mnogobrojne pogodnosti i fleksibilniji je od drugih programa (v. Garcia, 2021; Larson-Hall i Mizumoto, 2020; Mizumoto i Plonsky, 2016), već i putem prilagođenih sučelja (v. npr. Mizumoto, 2015).⁵¹ Dok Mizumoto i Plonsky

⁴⁹ <http://www.r-project.org/> (posljednji pristup 26. 4. 2024.)

⁵⁰ Česta je naprimjer uporaba aplikacije RStudio (<https://posit.co>, posljednji pristup 26. 4. 2024.), koja se nešto više od osnovne instalacije oslanja na pristup putem grafičkoga sučelja.

⁵¹ Mizumoto (2015) predstavlja sučelje izrađeno u paketu „shiny“ (<https://shiny.rstudio.com>), dostupno na <https://langtest.jp> (posljednji pristup 26. 4. 2024.), koje dopušta izračunavanje

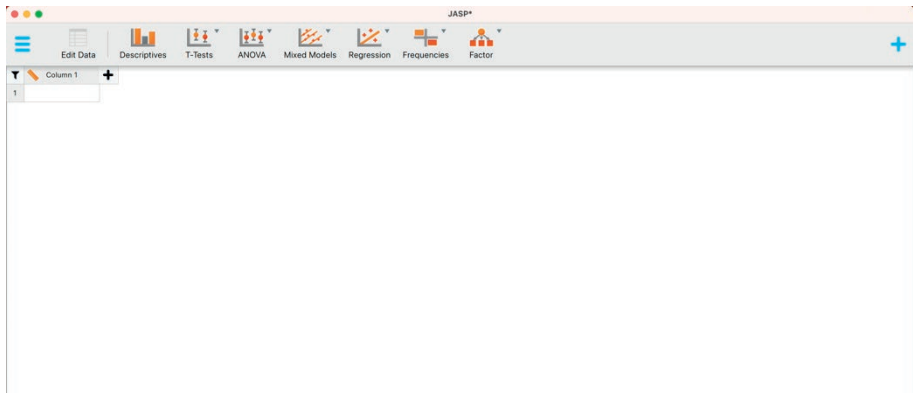
(2016) smatraju da taj program treba biti *lingua franca* kvantitativnih metoda, Larson-Hall i Mizumoto (2020, str. 386) uz to što je R besplatan ističu kako on dopušta izradu svih analiza koje dopušta SPSS, pa i više njih, i kako je izrada testova u njemu transparentnija za korisnike, koji pisanjem naredaba bolje upoznaju i bit statističke analize.



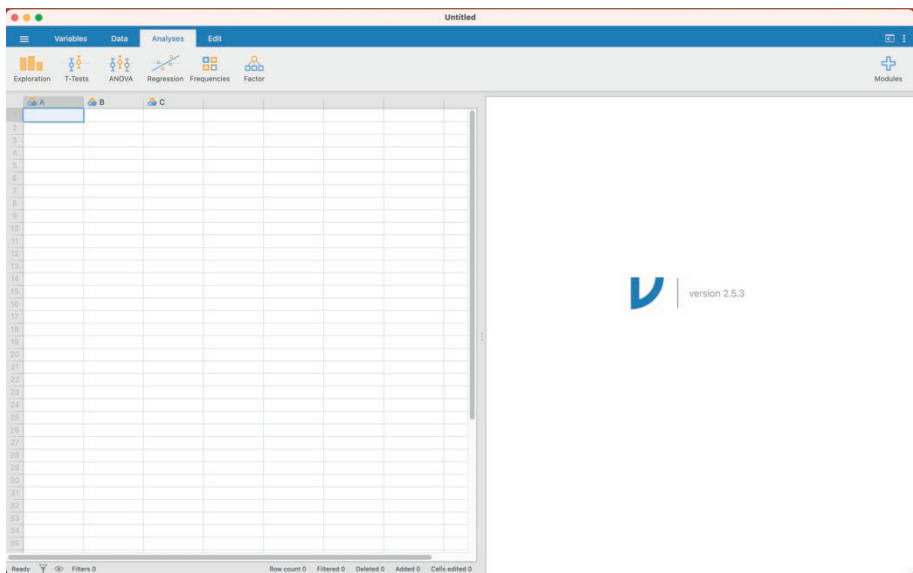
Slika 10. Sučelje programa R

U pojedinim je granama lingvistike, posebice psiholingvistici, sve prisutniji program **JASP** (v. sliku 11). Taj program dopušta izradu velikoga broja različitih analiza putem jednostavnih izbornika, ali se u njemu teže dobiva uvid u kôd na kojemu je zasnovan (koji potječe iz R-a), a uz to sadrži i manje opcija vezanih uz dodavanje i čuvanje podataka – praktički je potrebno sve operacije pripreme podataka provesti izvan JASP-a. Također se širi uporaba programa **Jamovi**, koji u pogledu analiza pruža manji broj mogućnosti, ali slično JASP-u dopušta rad u jednostavnome grafičkom sučelju nalik na IBM SPSS (v. sliku 12), iako se u pogledu provedbe analiza u pozadini oslanja na R.

osnovnih mjera i niza statističkih testova putem grafičkoga sučelja, ali ujedno u dnu stranice pokazuje i kôd koji se nalazi u pozadini procesa.



Slika 11. Sučelje programa JASP



Slika 12. Sučelje programa Jamovi

Mnogobrojne knjige navode detaljne upute za uporabu programa za statističku analizu. U područjima srodnima lingvistici, uz lako razumljive primjere, načini uporabe programa IBM SPSS i R prikazani su, zajedno s pristupačno objašnjenim osnovama statistike, u knjigama *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (Pallant, 2020, i ranija izdanja), *Discovering statistics using IBM SPSS Statistics* (Field, 2018, i ranija izdanja) i *Discovering statistics using R* (Field i sur., 2012). U području

psihologije korisna može biti i slobodno dostupna skripta *Learning statistics with R* (Navarro, bez dat.). Na hrvatskome jeziku uporaba IBM SPSS-a prikazana je u drugome dijelu knjige *Petzova statistika: Osnovne statističke metode za nematematičare* (Petz i sur., 2012; autor relevantnoga dijela je D. Ivanec). Uporabom programa za statističku analizu podataka u području usvajanja drugoga jezika bave se Larson-Hall u knjizi *A guide to doing statistics in second language research using SPSS and R* (Larson-Hall, 2016), Roever i Phakiti (2018) u knjizi *Quantitative methods for second language research: A problem-solving approach*, Loerts i sur. (2020) u knjizi *Essential statistics for applied linguistics: Using R or JASP*, kao i Garcia u knjizi *Data visualization and analysis in second language research* (Garcia, 2021). Korisne knjige koje nisu neposredno vezane uz usvajanje drugoga jezika, ali su posvećene analizi jezičnih podataka (u programu R), jesu *Analyzing linguistic data: A practical introduction to statistics using R* (Baayen, 2008), *Statistics for linguistics with R: A practical introduction* (Gries, 2013b), *How to do linguistics with R: Data exploration and statistical analysis* (Levshina, 2015) i *Statistics for linguists: An introduction using R* (Winter, 2020). Dostupni su i mnogobrojni online-tečajevi (na srpskome jeziku v. posebice Miličević, 2017c).

2.3 Provjera pouzdanosti instrumenata

Kad su podaci unijeti u odgovarajući računalni program, može se pristupiti izračunavanju statističkih mjera, ali i indikatora pouzdanosti korištenih instrumenata. Iako se o psihometrijskim značajkama instrumenata najčešće govori kod upitnika, odnosno kod istraživanja koja nisu eksperimentalnoga tipa (v. npr. Čolović i Milin, 2018), u posljednje se vrijeme u području usvajanja drugoga jezika o toj temi sve češće i općenitije raspravlja te se preporučuje detaljna provjera značajki svih vrsta mjernih instrumenata, uključujući testova razine općega znanja drugoga jezika i eksperimentalnih zadataka (v. npr. Grabowski i Oh, 2018; Tazik, 2019). Osnovne značajke koje su predmet zanimanja jesu njihova pouzdanost i valjanost, upravo kao i kod kontrole kvalitete istraživanja općenito (v. potpoglavlje 2.1.6, kao i Kraš i Miličević, 2015, str. 38–40). Al-Hoorie i Vitta (2019) naprimjer među sedam grijeha kvantitativnih istraživanja usvajanja drugoga jezika kao prva dva navode upravo izostanak navođenja indikatora pouzdanosti i valjanosti instrumenata.⁵²

⁵² Kako navode Plonsky i Derrick (2016, str. 539), iako je važnije da pouzdanost u istraživanju bude uistinu ostvarena nego eksplicitno opisana u izvještaju, navođenje relevantnih mjera umnogome pridonosi transparentnosti istraživanja i pomaže kako autorima tako i drugim istraživačima u interpretaciji cjelokupnih rezultata istraživanja, kao i u procjeni treba li isti

Pouzdanost instrumenata može se mjeriti iz više različitih perspektiva, pri čemu izbor ovisi o značajkama instrumenta i postupcima primijenjenima u istraživanju.⁵³ **Unutarnja konzistentnost** (engl. *internal consistency*) predstavlja stupanj u kojemu su opservacije dobivene uporabom danoga instrumenta stabilne, odnosno stupanj u kojemu se u ponovljenim mjerenjima ili u odgovorima na povezane stavke dobivaju isti ili slični odgovori (v. Čolović i Milin, 2018; Fajgelj, 2014; u području usvajanja drugoga jezika Plonsky i Derrick, 2016). Može se izraziti uz pomoć više indikatora, od kojih je najčešći u uporabi **Cronbachov koeficijent alfa** (engl. *Cronbach's alpha*), kojim se procjenjuje količina varijabilnosti u zabilježenim podacima koja se može pripisati varijabilnosti u stvarnome rezultatu (pri čemu se preostali postotak varijabilnosti tretira kao pogreška u mjerenju). Wu i Ortega (2013) naprimjer upotrebljavaju taj koeficijent u provjeri pouzdanosti novoga instrumenta temeljenoga na potaknutome oponašanju, koji su razvile u svrhu testiranja razine općega znanja kineskoga jezika. Autorice utvrđuju vrlo visoku vrijednost ($\alpha=0,97$), koju smatraju pokazateljem visoke pouzdanosti testa.⁵⁴

Iz druge perspektive, **stupanj slaganja između ocjenjivača** (engl. *interrater reliability/agreement*) izražava konzistentnost između odgovora različitih ocjenjivača, dok **stupanj slaganja unutar ocjenjivača** (engl. *intrarater reliability/agreement*) izražava slaganje između više ocjena koje je dao isti ocjenjivač (naprimjer u više vremenskih točaka). Najčešće korišten koeficijent u toj domeni jest **Cohenova kapa** (engl. *Cohen's kappa*).⁵⁵ Naprimjer Nahatame (2018) izračunava taj pokazatelj u svrhu usporedbe ocjena koje su dva nezavisna ocjenjivača dala rečenicama koje su učenici engleskoga jezika čiji je materinski jezik japanski proizveli u zadatku potaknutoga prisjećanja pročitanih parova rečenica povezanih uzročnim odnosima ili sličnošću u značenju. Ocjenjivači su imali zadatak da rečenicama dodijele 1 ako one predstavljaju ponavljanje ili parafrazu ciljnih rečenica, odnosno 0 ako taj uvjet nije ispunjen. U skladu s relativno čestom praksom u području

instrument primjenjivati u budućim istraživanjima. Drugim riječima, preporučuje se čitateljima da ne prihvate zdravo za gotovo da je dani instrument pouzdan.

⁵³ Ovdje govorimo o tzv. unutarnjoj pouzdanosti, dok se vanjska pouzdanost odnosi na mogućnost replikacije istraživanja sa sličnim rezultatima, o čemu će biti riječi u potpoglavlju 7.1.

⁵⁴ Al-Hoorie i Vitta (2019) svraćaju pozornost na to da prilikom izbora mjera pouzdanosti treba voditi računa o uvjetima koje podaci moraju ispuniti kako bi ih izračunate mjere adekvatno prikazale. Za Cronbachov koeficijent alfa potrebno je da konstrukt koji se mjeri bude jednodimenzionalan.

⁵⁵ Ta je mjera unijela podosta zbrke u svim poljima u kojima se primjenjuje jer je verzija koja se danas najčešće upotrebljava zapravo drukčija od one koju je izvorno predložio Cohen. Međutim naziv je zadržan.

usvajanja drugoga jezika, ali i u drugim područjima, dvostruko ocjenjivanje nije provedeno na svim podacima već na 30 % uzorka, nakon čega je na temelju visokoga koeficijenta slaganja ($\kappa=0,89$), uz razrješivanje zabilježenih neslaganja ocjenjivanje samostalno nastavio jedan ocjenjivač.

O mjerama slaganja između ocjenjivača još se više informacija može pronaći u računalnoj i korpusnoj lingvistici, koje se uvelike oslanjaju na anotaciju različitih pojava i upotrebljavaju te mjere za njezinu validaciju. Artstein i Poesio (2008) daju posebno koristan pregled navodeći da je za potrebe računalne lingvistike u nekim slučajevima od Cohenove kape primjereniji **Krippendorffov koeficijent alfa** (engl. *Krippendorff's alpha*), na koji Plonsky i Derrick (2016) u svojoj metaanalizi posvećenoj mjerama pouzdanosti u istraživanjima usvajanja drugoga jezika ne nailaze u 537 radova koje su uzeli u obzir. Drugi koristan savjet Artsteina i Poesia koji se može primijeniti i u usvajanju drugoga jezika jest uključivanje većega broja ocjenjivača. Dodatno, Norouzian (2021) vraća pozornost na važnost mjera slaganja kod metaanaliza i čak predlaže posebne mjere za tu namjenu.

Većina često korištenih statističkih programa ima ugrađene opcije za izračunavanje navedenih indikatora.⁵⁶ Vrijednosti koje se za njih dobivaju gotovo se uvijek – kao što ilustriraju i prethodni primjeri – kreću između 0 i 1. Tumačenje dobivene vrijednosti može se temeljiti na općim, unaprijed predloženim rasponima. Mogu se primijeniti smjernice koje je predložio Brown (2014, prema Plonsky i Derrick, 2016), sa sljedećim vrijednostima: 0,00–0,30 = gotovo nikakva pouzdanost; 0,31–0,50 = slaba pouzdanost; 0,51–0,70 = srednja pouzdanost; 0,71–0,89 = umjerena pouzdanost; 0,90–1,00 = velika pouzdanost. U području usvajanja drugoga jezika unutar se konzistentnost u istraživanjima koja su analizirali Plonsky i Derrick (2016, str. 544) kreće oko 0,82, odnosno može se okarakterizirati kao umjerena, dok je slaganje između i unutar ocjenjivača nešto više i iznosi 0,92–0,95. Međutim ti autori predlažu kao još bolju opciju to da se dobiveni koeficijenti usporede s onima iz istraživanja sličnih tema sa sličnim instrumentima i sličnim ispitanicima jer je za te čimbenike utvrđeno da utječu na razinu pouzdanosti (naprimjer kod naprednijih se učenika dobiva viša razina pouzdanosti nego kod početnika).

U vezi s pouzdanošću posebno je važno napomenuti da ona nije ni trajna ni nepromjenljiva značajka instrumenta, već se izračunava iz rezultata dobivenih u konkretnome istraživanju, s konkretnim ispitanicima, zbog čega je treba ponovno izmjeriti čak i ako je instrument već korišten u ranijim

⁵⁶ V. npr. implementaciju u *online*-sučelju programa R koje je razvio Mizumoto (2015), <http://langtest.jp/shiny/rel/> (posljednji pristup 26. 4. 2024.).

istraživanjima u kojima su provedene analize pouzdanosti (v. Larson-Hall i Plonsky, 2015, str. 140–141). Larson-Hall i Plonsky ističu i potrebu da se koeficijenti pouzdanosti navode za sve vrste eksperimentalnih mjera, dajući kao primjere prosudbe gramatičnosti, fonološke testove i upitnike na kojima se daju numerički odgovori. Kao motivaciju za to navode potrebu da se posebno u slučaju dobivanja rezultata koji nisu statistički značajni provjeri jesu li dobiveni rezultati možda posljedica nedostataka instrumenta jer niža pouzdanost instrumenta znači da su rezultati varijabilni i da vjerojatno ne mjere stvarno znanje i sposobnosti ispitanika u drugome jeziku.

S druge strane, valjanost je sposobnost instrumenta da adekvatno mjeri konstrukt koji operacionalizira, odnosno koji bi trebao mjeriti (Čolović i Milin, 2018). Budući da se može izdvojiti više različitih vrsta valjanosti (konstruktna, vanjska, unutarnja; v. Kraš i Miličević, 2015, str. 39), ona se ne može lako izraziti jedinstvenim indeksom. U području usvajanja drugoga jezika često se stoga preporučuje da se valjanost utvrdi i istraži putem konceptualnih pokazatelja, kao što su pilotiranje instrumenta ili njegova usporedba sa sličnim instrumentima koji su prethodno korišteni u relevantnim istraživanjima (v. Norris i sur., 2015), iako je u te svrhe moguće upotrijebiti i složenije statističke postupke poput faktorske analize (v. npr. Tabachnick i Fidell, 2013).

3. Statističko opisivanje podataka

U ovome poglavlju pozornost posvećujemo početnome opisu podataka različitim vrstama deskriptivnih statističkih mjera i vizualizacijom. Iako taj korak obuhvaća samo ispitivani uzorak i ne vodi izravno uopćavanju rezultata, uopćavanje izvan uzorka bez njega ne bi bilo moguće. Također, u novije se vrijeme sve više ističe važnost temeljitoga opisa podataka kao (djelomične) zamjene za statističko testiranje kao i važnost njegove kvalitete.

Opisivanje podataka može se podijeliti na dvije komponente, koje se međusobno nadopunjavaju: izračunavanje mjera koje sažimaju podatke dobivene u istraživanju i prikaz podataka grafikonima. U objema komponentama iznimno je važno odabrati pravi pristup na temelju prirode podataka koji se analiziraju. O tome se detaljno govori u sljedećim potpoglavljima.

3.1 Izračunavanje deskriptivnih mjera

Jedan od najvažnijih koraka u kvantitativnoj analizi podataka jest izračunavanje deskriptivnih mjera. Te mjere predstavljaju osnovni opis dobivenih podataka, tj. sažimaju značajke većega broja opservacija u svrhu prijenosa jedinstvenih informacija o njima (Upton i Cook, 1996, str. 36). Mogu se odnositi naprimjer na broj ili na postotak točne uporabe određene strukture, na učestalost izbora određene strukture (npr. određenoga referenta za subjektu zamjenu u prethodno navedenim primjerima iz Kraš, 2008b, 2008c, 2016, te Miličević i Kraš, 2017), na prosječna vremena reakcije za određene riječi itd. Ukratko, one predstavljaju način da se ukupni podaci dobiveni u istraživanju, koji su često opsežni i teško sagledivi kao cjelina, svedu na manju mjeru, koju je lakše shvatiti i prikazati (v. Johnson, 2013, str. 288). Deskriptivne statističke mjere izračunate iz uzorka opisuju ispitani uzorak i same po sebi ne govore ništa ni o značajkama drugih sličnih uzoraka ni populacije koju predstavljaju. Međutim te mjere ujedno čine temelj za provedbu statističkih testova i samim time su neophodne u svakome kvantitativnom istraživanju. Kako navode Larson-Hall i Plonsky (2015, str. 130), „deskriptivna statistika temelj je cjelokupnoga kvantitativnog razmišljanja i apsolutno je neophodno da je istraživački radovi navode“ (prev. M. M. P. i T. K.).

Izbor odgovarajućih deskriptivnih mjera koje će biti izračunate i navedene u izvještaju o istraživanju umnogome ovisi o korištenoj mjernoj ljestvici, kao i o nacrtu istraživanja i nekoliko dodatnih čimbenika (o kojima će u nastavku biti riječi). Mjere o kojima u ovome poglavlju govorimo opisane su u gotovo svim udžbenicima statistike, tako da se čitatelji lako mogu dodatno informirati o njima

(v. posebno Johnson, 2013; Larson-Hall, 2010, str. 62–68; Mackey i Gass, 2005, str. 250–261; Upton i Cook, 1996, drugo poglavlje; na hrvatskome jeziku Petz i sur., 2012, ponajprije poglavlja 3–6; na srpskome jeziku Todorović 2008, gdje je prikaz deskriptivnih mjera povezan s nacrtima uz koje se upotrebljavaju, te Tošković, 2020). Konkretno, osvrnut ćemo se na mjere kojima se izražava frekvencija različitih vrijednosti u podacima, mjere koje pokazuju relativni položaj pojedinačnih vrijednosti unutar skupa podataka, mjere kojima se izražava grupiranje podataka oko neke vrijednosti i mjere rasprišenja podataka oko te vrijednosti. Usredotočit ćemo se i na nekoliko elemenata opisa podataka koji imaju posebnu ulogu u izboru ili tumačenju inferencijalnih statističkih testova – distribuciju podataka i pokazatelje varijabilnosti koji imaju širu važnost od samoga uzorka i upotrebljavaju se i pri procjeni populacijskih vrijednosti, odnosno **parametara** (engl. *parameters*), na temelju mjera koje su izračunate za uzorak i koje nose zajednički naziv **statistik** (engl. *statistic*).

Statistiki koji se izračunavaju u deskriptivnome dijelu analize podataka odnose se ponajprije na vrijednosti zavisnih varijabli i na ispitanike kao jedinicu analize. Korelacijski i frekvencijski nacrti istraživanja u većoj mjeri uključuju u opis podataka i nezavisne varijable jer su u njima sve varijable jednake prirode, a analiza matematički gledano simetrična (v. potpoglavlja 5.1.1 o hi-kvadrat testu i 5.4 o korelaciji). Deskriptivna analiza prema česticama rijetko se susreće, čak i u istraživanjima u kojima se provodi inferencijalna statistička analiza i prema ispitanicima i prema česticama (usp. potpoglavlja 5.3.3 i 5.5). U istraživanjima usvajanja drugoga jezika deskriptivnim mjerama pripadaju i podaci o ispitanicima (poput njihova broja u istraživanju ili raspona njihove dobi), podaci o instrumentu (ukupan broj čestica u eksperimentu, broj čestica po eksperimentalnome uvjetu) i dodatni podaci o rezultatima (poput minimalnih i maksimalnih dobivenih vrijednosti i njihova raspona).⁵⁷ Te mjere nećemo detaljnije opisivati jer ne zahtijevaju posebna objašnjenja (njihov veliki dio čine jednostavna izravna mjerenja izvršena putem sociolingvističkoga ili demografskoga upitnika; v. Kraš i Miličević, 2015, str. 53–59), već slijedeći sugestije koje iznose Larson-Hall i Plonsky (2015, str. 130) ovdje želimo podsjetiti na važnost njihova navođenja.

⁵⁷ Svracamo pozornost čitatelja na višeznačnost naziva „raspon“, koji se može upotrebljavati u svojem svakodnevnom smislu razmaka *od-do* (npr. 20–24), što se često susreće u tabličnim prikazima deskriptivnih statističkih mjera, ali može označavati i mjeru varijabilnosti koja se dobiva kad se od najveće oduzme najmanja vrijednost u prikupljenim podacima (npr. 4 u slučaju 20 kao najniže i 24 kao najviše vrijednosti).

3.1.1 Frekvencijske mjere

Frekvencijske mjere, mjere učestalosti ili mjere prebrojavanja (engl. *frequency measures*) pokazuju nam s kakvom je frekvencijom neki rezultat zabilježen u istraživanju. Najveću važnost imaju za kategorijske varijable, izmjerene na nominalnoj ljestvici, jer za njih predstavljaju jedinu vrstu numeričkih podataka koju je moguće izračunati, dok su za numeričke varijable jedna od mogućih vrsta deskriptivnih mjera. Te su mjere rezultat prebrojavanja članova neke kategorije (naprimjer glagola upotrijebljenih u pasivu ili ispitanika koji nisu upotrijebili nijedan pasiv u zadatku proizvodnje) ili opažanja u uzorku za koje je zabilježena određena numerička vrijednost (naprimjer broja ispitanika s 25 bodova u testu dopunjavanja teksta koji mjeri razinu općega znanja drugoga jezika, ili broj izbora opcije „1 – potpuno neprihvatljivo“ u zadatku prosudbe prihvatljivosti rečenica).

Podaci o frekvenciji posebice su prisutni u korpusnim istraživanjima, gdje su česti nacrti istraživanja s više kategorijskih varijabli (frekvencijski nacrti, v. potpoglavlje 2.1.5, kao i 5.1) i gdje središnju vrstu podataka predstavljaju vrsta jezičnih jedinica (gramatičkih konstrukcija, leksičkih pojava i sl.) i vrsta teksta (proizvodnja izvornih govornika, proizvodnja učenika). Naprimjer Díaz-Negrillo i Valera (2010) proučavaju različite vrste pogrešaka u učeničkome korpusu engleskoga jezika (gdje je prvi jezik učenika španjolski) prema jezičnoj razini pogreške (interpunkcija, leksik itd.) i prirodi pogreške (interna ili eksterna, odnosno vezana uz oblik riječi ili uz njezinu uporabu). U eksperimentalnim se istraživanjima podaci o frekvenciji rjeđe susreću kao glavna vrsta podataka, ali važni su u okviru razmatranja distribucije podataka (v. potpoglavlje 3.1.5), a mogu se susresti i kao mjere izvedene iz višestrukih opservacija, kao u gornjim primjerima izbora referenta subjektne zamjenice (v. potpoglavlje 2.2.2).

Dvije glavne vrste frekvencije jesu **apsolutna frekvencija** (engl. *absolute frequency, raw frequency*) i **relativna frekvencija** (engl. *relative frequency*). Apsolutna je frekvencija ona na koju se najčešće najprije pomisli – to je broj pojavljivanja određene riječi ili konstrukcije u korpusu, broj izbora određenoga referenta zamjenice, broj ispitanika koji upotrebljavaju određenu konstrukciju i sl. Apsolutna je frekvencija rezultat prebrojavanja i po svojoj je prirodi uvijek cijeli broj (usp. apsolutnu ljestvicu opisanu u potpoglavlju 2.1.2). Ona može biti dovoljna kao mjera ako ispitujemo samo jedan uzorak ili uspoređujemo uzorke jednake veličine. Međutim u istraživanjima je vrlo česta situacija da se uzorci (bilo da ih čine korpusi, skupine ispitanika ili broj odgovora po ispitaniku) razlikuju u veličini, zbog čega je nemoguće osloniti se na apsolutnu frekvenciju, koja je u tome slučaju međusobno neusporediva za različite

uzorke. Naprimjer ako u zadatku proizvodnje ispitujemo upotrebljavaju li učenici s različitim materinskim jezicima različit broj članova u engleskome kao drugome jeziku i utvrdimo da su izvorni govornici hrvatskoga upotrijebili 220 određenih članova, a izvorni govornici nizozemskoga 150 određenih članova, ti podaci neće imati isto značenje ako u prvoj skupini ima 20, a u drugoj 10 ispitanika, ili ako ih je u prvoj skupini 20, a u drugoj 30.

Zbog toga je iznimno važna i relativna frekvencija, odnosno frekvencija izražena u **proporcijama** ili **postocima** (engl. *proportions, percentages*). Proporcije su matematičke veličine koje izražavaju udio koji određena kategorija ima u cjelini, odnosno u ukupnome ispitanom uzorku. Iskazuju se kao kvocijent frekvencije kategorije i frekvencije cjeline te se njihova veličina uvijek kreće u rasponu od 0 do 1. Naprimjer ako je 10 od 20 ispitanika u istraživanju engleskoga kao drugoga jezika ispravno upotrijebilo članove, njihova proporcija iznosi $10/20$, odnosno 0,5. Postoci predstavljaju jedan vid normalizacije proporcija – oni se izvedu iz proporcija množenjem sa 100, označavajući tako koliko članova neke kategorije dolazi na ukupno 100 članova uzorka. Postoci se bilježe dodavanjem simbola %, pa bi se prethodni primjer s proporcijom 0,5 kao postotak iskazao u obliku 50 %.

Kao primjer uporabe mjera relativne frekvencije možemo dati istraživanje koje su provele Montrul i sur. (2008). Te su autorice proučavale koliko učenici španjolskoga kao drugoga jezika čiji je materinski jezik engleski poznaju pomak upitne zamjenice u formulaciji pitanja na španjolskome u odnosu na nasljedne govornike španjolskoga i izvorne govornike španjolskoga u funkciji kontrolnih ispitanika. U individualnoj analizi podataka određivalo se koji ispitanik pokazuje kakav obrazac u svojim odgovorima u zadatku prosudbe prihvatljivosti rečenica – obrazac koji je u skladu s gramatičkim pravilima španjolskoga jezika (A), obrazac koji pokazuje utjecaj engleskoga jezika (B), obrazac koji upućuje na deficit u obradi rečenica (C) ili obrazac koji se može protumačiti na različite načine (D) – te je izračunata zastupljenost obrazaca u postocima u svakoj skupini ispitanika (jer su tri skupine sadržavale različit broj ispitanika). Dobiveno je, između ostaloga, da obrazac A pokazuje svega 12 od 70 učenika španjolskoga kao drugoga jezika, što odgovara relativnoj frekvenciji od 17 % te proporciji od 0,17, nasuprot 22 od 63 nasljednih govornika (35 %, odnosno 0,35) i 19 od 22 kontrolnih ispitanika (86 %, tj. 0,86).

Budući da se u svakodnevnome životu postoci upotrebljavaju češće od proporcija, nije rijetkost da istraživači i ne uoče da se radi o istoj mjeri, samo drukčije izraženoj. U statističkoj se analizi mogu podjednako upotrebljavati obje mjere, s time što su proporcije matematički jednostavnije (v. npr.

Todorović 2008, str. 197) i u tome smislu možda preporučljivije za unos podataka i izračunavanje eventualnih dodatnih mjera. Dodatni tehnički razlog zbog kojega bi se u kodiranju podataka mogla dati prednost proporcijama jest upravo simbol %, koji primjerice u programu Microsoft Excel podrazumijeva uporabu posebnoga formata polja u tablici.⁵⁸

Kad se u istraživanju analiziraju dvije varijable izmjerene na nominalnoj ljestvici, njihova se frekvencija uobičajeno prikazuje u **kontingencijskim tablicama** (engl. *contingency tables*), koje se još nazivaju i **unakrsne tablice** (engl. *crosstabs*). Takve tablice u najjednostavnijoj varijanti prikazuju dvije varijable, od kojih je jedna prikazana u recima, a druga u stupcima. Naprimjer Robison (1990) u okviru ispitivanja hipoteze prema kojoj učenici drugoga jezika ranije počinju upotrebljavati glagolske morfeme za iskazivanje vida nego za iskazivanje vremena ovako prikazuje podatke o tome kako jedan učenik engleskoga kao drugoga jezika upotrebljava osnovni oblik glagola, oblik s morfemom *-ing* i oblik prošloga vremena (označen kao oblik s morfemom *-EN*) s glagolima koji se prema leksičkome vidu mogu svrstati u statične i dinamične. Njegove podatke ovdje prikazujemo u tablici 5, koja je primjer tablice koja se može opisati kao 3 x 2 jer jedna varijabla ima tri, a druga dvije razine. Iz istih se podataka mogu izračunati i relativne frekvencije, koje se mogu prikazati u istoj vrsti tablice (v. Robison 1990, str. 325).

Tablica 5. Primjer kontingencijske tablice (prema Robison, 1990)

		Vrsta glagola prema leksičkome vidu	
		Statični	Dinamični
Morfološki oblik glagola	osnovni oblik	137	301
	<i>-ing</i>	39	48
	<i>-EN</i>	0	28

Uz tablicu 5 treba napomenuti da su uzorci koji sudjeluju u kategorijama nezavisne varijable nezavisni uzorci – jedinica je analize u tome slučaju glagolska pojavnica (tj. pojedinačna uporaba glagola) i ona može biti klasificirana ili kao statična ili kao dinamična, što znači da se unutar statične i unutar dinamične skupine nalaze različite i međusobno nezavisne pojavnice.⁵⁹

⁵⁸ Čitateljima je možda poznata situacija unošenja u Excel brojeva poput 15 ili 50 i pokušaja da im se mijenjanjem formata polja doda znak %, gdje se kao rezultat zapravo dobiva 1500 %, odnosno 5000 %. I to nam praktično iskustvo pokazuje da su postoci zapravo isto što i proporcije – da bi nam Excel pokazao željenih 15 % ili 50 %, u odgovarajuća polja zapravo moramo unijeti 0,15 odnosno 0,5.

⁵⁹ Kad bi jedinica analize bio ispitanik, mogli bismo imati naprimjer učenike na srednjoj i

Nezavisne uzorke u kontingencijskim tablicama moguće je prepoznati i po tome što su kategorije unutar dviju varijabli različite te se mogu lako dodati nove kategorije unutar obiju varijabli, pa tablice stoga mogu imati različite dimenzije.

Slične tablice mogu se susresti i u nekim situacijama kad su podaci međusobno zavisni. Andringa (2014) naprimjer ispituje utjecaj sastava kontrolne skupine na rezultate istraživanja hipoteze o kritičnome razdoblju. On uključuje u istraživanje dvije kontrolne skupine, jednu sastavljenu od izvornih govornika nizozemskoga različitih razina obrazovanja i jednu sastavljenu od osoba koje su završile diplomski studiji, te uspoređuje broj neizvornih govornika koji ostvaruju rezultate slične onima koje ostvaruju izvorni govornici u slučaju kad se uzimaju u obzir različiti govornici ili samo oni koji su visoko obrazovani (te uzorke označava kao „reprezentativni“ i „nereprezentativni“). U tablici 6 prikazujemo jedan dio podataka, koji se tiče gramatičke točnosti.⁶⁰ U toj se tablici vidi da 49 neizvornih govornika ostvaruje rezultat unutar norme bez obzira na to prema kojemu se uzorku izvornih govornika definira norma, 50 neizvornih govornika ni u jednome slučaju ne ostvaruje rezultat unutar norme, 10 neizvornih govornika ostvaruje rezultat unutar norme samo prema reprezentativnome uzorku, dok nema ispitanika koji ostvaruju normu samo prema nereprezentativnome uzorku.

Tablica 6. Primjer zavisne kontingencijske tablice (prema Andringa, 2014)

		Nereprezentativni uzorak	
		Gramatička točnost unutar norme	Gramatička točnost izvan norme
Reprezentativni uzorak	Gramatička točnost unutar norme	49	10
	Gramatička točnost izvan norme	0	50

naprednoj razini općega znanja engleskoga kao drugoga jezika koji jesu ili nisu položili ispit iz akademskoga pisanja, gdje bismo u skupinama definiranim prema varijabli razine općega znanja drugoga jezika imali različite ispitanike.

⁶⁰ Tablica 6 izrađena je na temelju nešto drukčijega formata podataka, koji navodi Andringa (2014, str. 587).

U ovome je istraživanju nezavisna varijabla vrsta uzorka, tj. skupina izvornih govornika prema kojoj se definira norma, te je jasno da su uzorci zavisni jer se svaki ispitanik kategorizira prema objema normama. U takvu je slučaju format tablice takav da varijable nisu ukrižane putem redaka i stupaca, već je nezavisna varijabla prikazana i u recima i u stupcima (tj. jedna njezina vrijednost je u recima, a druga u stupcima), dok se unutar nje na oba mjesta javljaju vrijednosti zavisne varijable. Zavisni se uzorci stoga mogu prepoznati u frekvencijskim tablicama i prema ponavljanju istih vrijednosti (zavisne varijable) u recima i stupcima. Također treba napomenuti da se kod tablica sa zavisnim uzorcima mogu dodati razine zavisne varijable (čime tablice postaju naprimjer 3 x 3 ili 4 x 4, tj. ostaju simetrične), dok dodavanje razina nezavisne varijable vodi drukčijemu formatu tablice, u kojemu je potrebno recima prikazati različite kombinacije vrijednosti zavisne varijable, uz vrijednosti nezavisne varijable u stupcima; takav je format ilustriran u tablici 7, na podacima kakvi bi se mogli dobiti proširenjem istraživanja koje je proveo Andringa (2014).

Tablica 7. Primjer frekvencijske tablice sa zavisnim uzorcima i tri razine nezavisne varijable

Reprezentativni uzorak	Nereprezentativni uzorak 1	Nereprezentativni uzorak 2	Frekvencija
unutar norme	unutar norme	unutar norme	35
unutar norme	unutar norme	izvan norme	14
unutar norme	izvan norme	unutar norme	0
unutar norme	izvan norme	izvan norme	10
izvan norme	unutar norme	unutar norme	0
izvan norme	unutar norme	izvan norme	0
izvan norme	izvan norme	unutar norme	0
izvan norme	izvan norme	izvan norme	50

Još je jedna posebna vrsta frekvencije **kumulativna frekvencija** (engl. *cumulative frequency*). Kod nje su u pitanju uzastopni zbrojevi frekvencija (bilo apsolutnih ili relativnih) pojedinih kategorija. Naprimjer ako trideset učenika engleskoga jezika koji su položili završni ispit grupiramo prema ocjenama koje su dobili, apsolutna i relativna frekvencija mogla bi izgledati kao u lijevom dijelu tablice 8, dok bi kumulativna frekvencija tada bila ona navedena u desnome dijelu te tablice (podaci su konstruirani po uzoru na primjer iz Woods i sur., 1986, str. 16). Kumulativna frekvencija rjeđe se susreće u istraživanjima usvajanja drugoga jezika, ali može biti korisna kad je potrebno utvrditi koliko se

opservacija u nekome skupu podataka nalazi ispod ili iznad neke vrijednosti, kao i za određivanje mjera relativnoga položaja (v. potpoglavlje 3.1.4).

Tablica 8. Primjer prikaza različitih vrsta frekvencije

Ocjena	Apsolutna frekvencija	Relativna frekvencija	Kumulativna apsolutna frekvencija	Kumulativna relativna frekvencija
1	0	0,000	0	0,000
2	6	0,200	6	0,200
3	9	0,300	15	0,500
4	10	0,333	25	0,833
5	5	0,167	30	1,000

Konačno, treba napomenuti da se različite vrste frekvencije ne moraju međusobno isključivati u istraživanjima. Upravo suprotno, istraživačima se često savjetuje da izračunaju i u prikaz rezultata uključe različite vrste kako bi uvid u podatke bio potpuniji. Naime, relativna će frekvencija, po svoj prilici, biti lakša za tumačenje, posebno ako je apsolutna frekvencija vrlo visoka, ali takav uvid nije potpun i bolje ga je dopuniti i neposrednim rezultatima prebrojavanja (Todorović, 2008, str. 198). Tako su naprimjer u istraživanju koje su provele Montrul i sur. (2008) dani podaci i za apsolutnu i za relativnu frekvenciju u prikazu rezultata individualne analize podataka – tablično za apsolutnu frekvenciju, a grafički za relativnu.

3.1.2 Mjere centralne tendencije

Sve se vrste varijabli mogu opisati pokazateljima koji izražavaju **centralnu tendenciju** (engl. *central tendency*), odnosno tendenciju grupiranja podataka oko neke vrijednosti koja se može upotrijebiti za sažeto prikazivanje situacije u čitavoj skupini. Te su mjere vrlo važne zbog toga što sažimaju veću količinu podataka u jednu vrijednost i lakše su za tumačenje od neuređene skupine pojedinačnih vrijednosti – mjera centralne tendencije uzima se kao tipična za cijelu skupinu vrijednosti. U svakodnevnome je životu dobro poznat primjer prosječne ocjene u školi ili na studiju, koja se uzima kao indikator općega uspjeha pojedinca i upotrebljava se za usporedbu uspjeha različitih pojedinaca, kao i primjer obitelji s dvoje djece kao tipičnih, zbog toga što je dvoje (bio) najčešći broj djece u obitelji.

U statističkoj analizi podataka kao mjere za izražavanje centralne tendencije najčešće se upotrebljavaju **aritmetička sredina** (engl. *arithmetic mean*),

medijan (engl. *median*) i **mod** (engl. *mode*). Izbor između tih mjera ovisi o korištenoj mjernoj ljestvici, ali i o distribuciji podataka. Uvjeti za uporabu najstroži su za aritmetičku sredinu, koju treba izračunavati samo za podatke s intervalne ili omjerne ljestvice, i to u slučaju kad su takvi podaci normalno distribuirani (o čemu će detaljnije biti riječi u potpoglavlju 3.1.5). Kad ti uvjeti nisu ispunjeni, aritmetička sredina ne može adekvatno predstaviti podatke. Sjetimo se naime iz potpoglavlja 2.1.2, o mjerenju varijabli, da je važna značajka intervalnih i omjernih ljestvica jednaka udaljenost između različitih vrijednosti na ljestvici – to je obavezan uvjet ako želimo da aritmetička sredina matematički ima smisla. Ako uvjeti za aritmetičku sredinu nisu ispunjeni, naprimjer kad su podaci izraženi na ordinalnoj ljestvici ili nemaju normalnu distribuciju, prikladna je mjera medijan. S druge strane, za podatke s nominalne ljestvice može se upotrebljavati samo mod. Naravno, medijan i mod mogu se izračunati i upotrijebiti kao dodatne mjere i u slučajevima kad je opravdano upotrijebiti aritmetičku sredinu.

Aritmetička sredina, koja se kraće naziva i **prosjeckom** (engl. *average, mean*), najčešće se upotrebljava i predstavlja ne samo najpoznatiju već i najpouzdaniju i najstabilniju od navedenih triju mjera centralne tendencije. Izračunava se tako što se zbroj svih relevantnih vrijednosti podijeli ukupnim brojem vrijednosti, što se može prikazati formulom u (1), u kojoj se aritmetička sredina označava oznakom M , X predstavlja pojedinačnu vrijednost iz skupa podataka, a N ukupan broj vrijednosti, odnosno veličinu uzorka. Napominjemo i da se oznaka M obično upotrebljava kad govorimo o prosjeku rezultata u uzorku, dok se odgovarajući parametar populacije označava malim grčkim slovom μ (μ).

$$(1) \quad M = \frac{\sum X}{N}$$

Budući da se u izračunavanju uzima u obzir točna vrijednost podataka, aritmetička je sredina vrlo osjetljiva na netipične podatke, pa njezinu uporabu u slučaju prisutnosti takvih podataka treba pažljivo razmotriti. Naprimjer ako u niz 3, 5, 5, 6, 8, 9 dodamo 55, aritmetička sredina iznositi će 13, dok bez te vrijednosti iznosi 6, što predstavlja vrlo veliku promjenu. Aritmetička se sredina zbog toga ponekad označava kao „nerobusna“ mjera (v. npr. Larson-Hall i Herrington, 2010). Ta njezina nerobusnost povezana je, dakako, i s veličinom uzorka jer će izmjene izazvane netipičnim vrijednostima imati bitno manji utjecaj u velikim uzorcima. To je upravo jedan od razloga zbog kojih pojedini autori predlažu određenu minimalnu veličinu uzorka kao dodatni uvjet za odluku o izračunavanju aritmetičke sredine. Naprimjer Petz i sur. (2012, str. 32) navode da bi za valjano izračunavanje aritmetičke sredine broj rezultata

trebao iznositi najmanje 30. U praksi se u područjima poput usvajanja drugoga jezika, za koje su karakteristični mali uzorci (v. npr. Plonsky, 2013), najčešće ne inzistira na poštovanju tako strogih uvjeta, ali svakako je preporučljivo da se veličina uzorka uzme u obzir prilikom odlučivanja.

Medijan, koji se još naziva i **centralna** ili **srednja vrijednost**, određuje se bitno drukčije od aritmetičke sredine i predstavlja onu vrijednost koja se u nizu vrijednosti poredanih po veličini nalazi točno u sredini. Drugim riječima, to je vrijednost koja dijeli skup podataka na dvije polovine – 50 % podataka manje je od medijana, dok je 50 % podataka veće. Budući da se određuje prema položaju u nizu vrijednosti, medijan spada u tzv. **položajne mjere** (engl. *positional measures*), o kojima detaljnije govorimo u potpoglavlju 3.1.4. Njegov položaj, odnosno redni broj u odgovarajućemu nizu podataka, određuje se formulom u (2), koja kao najvažniju vrijednost uzima veličinu uzorka, odnosno broj rezultata (N). Oznaku R_c ovdje preuzimamo iz Petz i sur. (2012, str. 39), gdje se kao oznaka za medijan upotrebljava C (od „centralna vrijednost“).⁶¹

$$(2) R_c = \frac{N+1}{2}$$

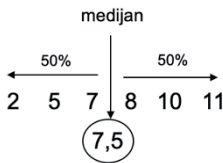
Naprimjer u nizu 2, 5, 7, 10, 11 medijan je vrijednost na trećoj poziciji, što dobivamo kao $(5+1)/2$, odnosno iznosi 7. Kad je broj vrijednosti paran, medijan se izračunava kao aritmetička sredina dviju vrijednosti u središtu niza – u nizu 2, 5, 7, 8, 10, 11 položaj bi bio određen kao 3,5 (odnosno $(6+1)/2$), tj. položaj između trećega i četvrtoga, gdje medijan iznosi 7,5, odnosno $(7+8)/2$. To je prikazano i na slici 13.

(a)



⁶¹ U literaturi na engleskome jeziku može se susresti M kao oznaka za medijan, ali i mod , te uvijek treba svratiti pozornost na što se oznaka u danome kontekstu odnosi.

(b)



Slika 13. Određivanje medijana u uzorku s neparnim brojem vrijednosti (a) i s parnim brojem vrijednosti (b)

Medijan se obavezno upotrebljava za podatke s ordinalne ljestvice, kao i za intervalne i omjerne podatke koji nisu normalno distribuirani. Naprimjer Cunningham-Andersson (1997) upotrebljava medijan za izražavanje centralne tendencije stavova prema izgovoru neizvornih govornika švedskoga u dvije skupine izvornih govornika švedskoga (srednjoškolaca i nastavnika švedskoga kao drugoga jezika), pri čemu se stavovi izražavaju na ljestvici od 1 do 5, pa se stoga radi o ordinalnim podacima. Za razliku od toga, u istraživanju proizvodnje pasiva u engleskome kao drugome jeziku kod učenika kojima je materinski jezik korejski tijekom interaktivnih aktivnosti s udešavanjem Kim i McDonough (2008) upotrebljavaju medijan umjesto postotka kako bi izrazili centralnu tendenciju proizvodnje pasiva kod triju skupina ispitanika (podijeljenih s obzirom na razinu općega znanja engleskoga jezika) u pojedinome eksperimentalnom uvjetu jer podaci nisu normalno distribuirani. Medijan se uz to može upotrebljavati kao dodatni podatak u slučajevima kad su ispunjeni uvjeti za uporabu aritmetičke sredine. U usporedbi s aritmetičkom sredinom, medijan je manje osjetljiv na netipične vrijednosti jer na njega ne utječe točna numerička vrijednost pojedinačnih rezultata. Ako se vratimo primjeru u kojemu u niz 3, 5, 5, 6, 8, 9 dodajemo 55, dok aritmetička sredina iznosi 13 (nakon što je prethodno iznosila 6), medijan iznosi 6, dok je prethodno bio 5,5 – u tome slučaju vrlo je očigledno da medijan bolje odražava centralnu tendenciju u podacima nego aritmetička sredina jer je bliži vrijednostima koje su se među rezultatima doista pojavile. Medijan se zbog toga smatra „robustnom“ mjerom.

Konačno, mod ili **dominantna vrijednost** ona je vrijednost koja se u danome nizu podataka javlja s najvećom frekvencijom, odnosno ponavlja se najviše puta. Naprimjer u nizu 2, 5, 7, 7, 10, 11 mod je 7, kao jedina vrijednost koja se javila dvaput. U slučaju jednake zastupljenosti više različitih vrijednosti, može se smatrati da modova ima više. Mod predstavlja najgrublju mjeru centralne tendencije jer ne ovisi ni o točnoj vrijednosti rezultata ni o njihovu ukupnom broju i upotrebljava se samostalno samo kada je nemoguće upotrijebiti prosjek i

medijan, naprimjer kod nominalnih varijabli. U analizi podataka iz eksperimenata najprisutniji je kad se govori o distribucijama frekvencija, a ponekad se može susresti i u slučajevima kad su navedene sve tri mjere centralne tendencije. Naprimjer Jensen i sur. (2020) za test razine općega znanja engleskoga kao drugoga jezika (*Oxford Proficiency*) za ispitanike čiji je materinski jezik norveški navode aritmetičku sredinu od 27,3, medijan od 29,5 i mod od 35 (na ljestvici čija je maksimalna vrijednost 40).

Za istraživanja usvajanja drugoga jezika treba istaknuti i to da se mjere centralne tendencije često izračunavaju u više koraka analize. Kao što je već navedeno u drugome poglavlju, u tome je području vrlo česta uporaba izvedenih mjera, koje zahtijevaju da se mjera centralne tendencije najprije izračuna za svakoga pojedinačnog ispitanika za svaki eksperimentalni uvjet (odnosno za više čestica koje predstavljaju taj uvjet), da bi se zatim iz individualnih mjera centralne tendencije računale mjere za cijelu skupinu, koje se uključuju u prikaz rezultata. Kad je mjera koja se upotrebljava aritmetička sredina, vrijednost koja se dobiva kao konačna jest **zajednička aritmetička sredina** (engl. *combined mean, weighted mean*). Ako je svaka od vrijednosti iz koje se ona izračunava (dakle, aritmetička sredina svakoga pojedinačnog ispitanika) dobivena iz jednakoga broja rezultata, ta se aritmetička sredina može izračunati prema uobičajenoj formuli navedenoj pod (1). Međutim ako postoje razlike u broju rezultata (naprimjer neki ispitanici nisu odgovorili na sve eksperimentalne čestice), potrebno je svaku individualnu aritmetičku sredinu pomnožiti pripadajućim brojem rezultata (N), a u nazivniku zbrojiti brojeve rezultata različitih pojedinaca, čime se osigurava da prosjeci izvedeni iz većega broja rezultata dobiju veću težinu. Formula za izračunavanje, preuzeta iz Petz i sur. (2012, str. 38), prikazana je u (3), a indeksi $1 \dots n$ označavaju ispitanike od 1 do n , tj. aritmetičke sredine pojedinačnih ispitanika i broj njihovih odgovora.

$$(3) \quad M_z = \frac{M_1 N_1 + M_2 N_2 + \dots + M_n N_n}{N_1 + N_2 + \dots + N_n}$$

Slično tome moguće je izračunati i **zajednički medijan** (engl. *weighted median*). Pritom se mora odrediti težina koju svaki individualni medijan nosi dijeleći broj rezultata iz kojega je dobiven ukupnim brojem rezultata, a zatim treba izračunati kumulativne zbrojeve težina; vrijednost koja odgovara prvome kumulativnom zbroju većem od 0,5 (50 %) predstavljat će u tome slučaju zajednički medijan. Ako se dogodi da jedan od zbrojeva iznosi točno 0,5, medijan će se računati kao prosjek vrijednosti koje odgovaraju tome

i sljedećemu zbroju.⁶² O toj se mjeri gotovo uopće ne govori u literaturi iz područja usvajanja drugoga jezika i uobičajenije je upotrebljavati „obični“ medijan. Međutim treba ga imati na umu kao mogućnost.

Iako postoje vrlo jasne smjernice u pogledu toga kad se u analizi podataka smije upotrebljavati aritmetička sredina, a kad je potrebno poslužiti se medijanom, u istraživanjima usvajanja drugoga jezika nerijetko se u praksi susreću oglašivanja o te smjernice, posebice u ranijim radovima. U istraživanjima koja se temelje na Likertovim ljestvicama, naprimjer u zadacima prosudbe prihvatljivosti rečenica, vrlo se često ocjene na toj ljestvici tretiraju kao intervalni podaci, pa se na njima izračunava aritmetička sredina i provode analize poput *t*-testova i analize varijance, koje nisu primjerene za ordinalne ljestvice (v. npr. analizu zadatka procjene slike na ljestvici od -3 do +3 u Montrul, 2001, ili analizu zadatka procjene veličine moralnoga prekršaja na materinskome ili drugome jeziku na ljestvici od 0 do 6 u Tonković i sur., 2020).⁶³ Al-Hoorie i Vitta (2019, str. 739) ističu da se ponekad, čak i u slučajevima kad je statistički test dobro izabran, susreću neadekvatne deskriptivne mjere (naprimjer prosjek umjesto medijana) te da treba paziti da do toga ne dođe.

3.1.3 Mjere varijabilnosti

Pored mjera prikazanih u prethodnome potpoglavlju, veliku važnost za opisivanje podataka imaju i **mjere varijabilnosti** (engl. *measures of variability*), odnosno **mjere raspršenosti** ili **disperzije** (engl. *measures of dispersion*) podataka u odnosu na mjeru centralne tendencije. Naime, iako podaci imaju tendenciju grupirati se oko nekoga težišta, u empirijskim su istraživanjima uvijek prisutna i određena odstupanja – iako su teorijski moguće, u praksi se gotovo nikad ne susreću situacije u kojima bi svi ispitanici u nekoj skupini ostvarili isti broj točnih odgovora ili potpuno jednako prosudili prihvatljivost rečenica u eksperimentalnome zadatku. Da bi se dobio uvid u to koliko su rezultati doista koncentrirani oko mjere centralne tendencije, odnosno koliko je dani uzorak homogen ili heterogen s obzirom na mjerenu pojavu, potrebno je izmjeriti količinu varijabilnosti. U svakodnevnome životu varijabilnost često procjenjujemo okvirno te ćemo sasvim sigurno primijetiti da dvije skupine učenika s prosječnim rezultatom od 50 bodova na testu engleskoga jezika

⁶² Primjer izračunavanja može se vidjeti naprimjer na stranici <https://real-statistics.com/descriptive-statistics/measures-central-tendency/weighted-mean-and-median/> (posljednji pristup 21. 4. 2024.).

⁶³ Taj se problem odnosi zapravo i na ocjene u sustavu obrazovanja, koje također obično predstavljaju rangove na ordinalnoj ljestvici jer ne možemo biti sigurni u jednaku udaljenost između različitih ocjena. Izračunavanje prosječne ocjene nekoga predmeta ili izračunavanje općega uspjeha ne treba uzimati kao nešto što je uvijek ispravno.

nisu potpuno jednake ako u jednoj većina učenika ima između 45 i 55 bodova, a u drugoj dominiraju rezultati od 35 i 65 bodova. Matematički gledano, varijabilnost možemo izmjeriti i izraziti uz pomoć više različitih mjera.

Najpoznatija mjera varijabilnosti je **standardna devijacija** (engl. *standard deviation*). Ona predstavlja mjeru koja prati aritmetičku sredinu, u smislu da se obavezno navodi uz nju i podliježe istim uvjetima uporabe kao i ona (intervalna ili omjerna ljestvica s normalnom distribucijom podataka, po mogućnosti velik uzorak bez mnogo ekstremnih vrijednosti). Izračunavanje standardne devijacije temelji se na razlikama između pojedinačnih rezultata i aritmetičke sredine, odnosno na odstupanju individualnih vrijednosti od centralne tendencije. Budući da neke razlike nužno moraju imati pozitivan, a neke negativan predznak, kako bi se izbjeglo njihovo međusobno poništavanje i dobivanje ukupne razlike koja je jednaka 0, odstupanja se kvadriraju, te se standardna devijacija računa prema formuli u (4), gdje kao i pri izračunavanju aritmetičke sredine u formuli u (1), X predstavlja pojedinačni rezultat, N ukupan broj rezultata, a M aritmetičku sredinu. Oznaka s predstavlja standardnu devijaciju u uzorku podataka, a izvan formula, u tablicama i tekstu opisa rezultata, česta je i kratica SD.

$$(4) \quad s = \sqrt{\frac{\sum(X-M)^2}{N-1}}$$

Dakle, u pitanju je korijen iz zbroja kvadriranih razlika između pojedinačnih vrijednosti i prosjeka podijeljenoga brojem rezultata. Točnije, zbog toga što govorimo o standardnoj devijaciji u uzorku, dijeljenje se vrši brojem rezultata umanjenim za jedan. Kad se govori o populaciji, kao oznaka za standardnu devijaciju upotrebljava se grčko slovo *sigma* (σ), a u nazivniku razlomka nalazi se N . Valja napomenuti da se N ne primjenjuje u slučaju uzoraka zbog toga što bi standardna devijacija uzorka izračunata na taj način podcijenila standardnu devijaciju populacije; točni razlozi vezani su uz značajke normalne distribucije i čitatelji ih mogu potražiti u Petz i sur. (2012, str. 143–146).

Iako je načelo izračunavanja standardne devijacije matematički vrlo jednostavno, početnicima u statističkoj analizi često je teško praktično protumačiti tu mjeru te ćemo je pokušati približiti uz pomoć primjera podataka iz istraživanja koje su proveli Berger i sur. (2019). Cilj je toga istraživanja bio da se u zadatku leksičke odluke utvrdi koje značajke riječi engleskoga jezika utječu na brzinu njihova prepoznavanja kod neizvornih govornika. Autori su između ostaloga utvrdili da je prosječna brzina koja je ispitanicima bila potrebna za odgovor iznosila 734,159 milisekundi, uz standardnu devijaciju

od 82,823 milisekunde, dok je prosječna točnost odgovora bila 0,940 (94 %) uz standardnu devijaciju od 0,078 (7,8 %; v. Berger i sur., 2019, str. 923).⁶⁴ Standardna devijacija pokazuje nam koliko su vrijeme reakcije i točnost odgovora u tome istraživanju raspršeni oko prosjeka.

U pogledu vremena reakcije, iz dobivenih rezultata možemo zaključiti da je većina ispitanika dala odgovore brzinom između 734,159±82,823 milisekundi, odnosno između 651,336 i 816,982 milisekunde (to ćemo pravilo dodatno objasniti u nastavku, a posebice u potpoglavlju 3.1.5). Budući da različite pojave odlikuje različita varijabilnost, ne postoji jedinstven način da se nedvosmisleno utvrdi predstavlja li dobiveni pokazatelj nisku ili visoku varijabilnost – potrebno je poznavati pojavu koja se ispituje i rezultate sličnih prethodnih istraživanja. Ako znamo da je uobičajeno da u zadatku leksičke odluke SD za neizvorne govornike iznosi više od 100 ms, zaključit ćemo da je varijabilnost u našim podacima niska, ali ako znamo da su druga istraživanja uglavnom pronalazila SD ispod 60 ms, tada ćemo svoje podatke ocijeniti kao nešto manje homogene. Dodatno se – samo kod omjernih ljestvica (v. Petz i sur., 2012, str. 95) – može izračunati **koeficijent varijabilnosti** (engl. *coefficient of variation*) dijeljenjem standardne devijacije aritmetičkom sredinom, čime se omogućava lakša usporedba rezultata kad su aritmetičke sredine različite (v. Petz i sur., 2012, str. 54–55). Ta se mjera u usvajanju drugoga jezika rijetko susreće; u područjima u kojima se upotrebljava, niskom se varijabilnošću obično smatra ona kod koje je taj koeficijent ispod 30 %. Može se primijeniti i slično grubo pravilo da se standardna devijacija smatra visokom ako iznosi više od 1/3 aritmetičke sredine (v. npr. <https://www.researchgate.net/post/When-is-a-standard-deviation-considered-as-high>, posljednji pristup 21. 4. 2024.), ali imajući u vidu varijabilnu prirodu učeničkoga jezika, zapravo je u području usvajanja drugoga jezika bolje ne oslanjati se previše na takve fiksne procjene.

Primjenjujući na točnost odgovora isto načelo tumačenja primijenjeno na vrijeme reakcije u istraživanju Berger i sur. (2019) možemo zaključiti da se točnost odgovora kod većine ispitanika kretala u rasponu 94 % ± 7,8 %, odnosno između 86,2 i 101,8 %. Ovdje, naravno, uočavamo da se zbrajanjem prosjeka i jedne standardne devijacije dobiva vrijednost viša od maksimalno moguće (100 %). Situacije u kojima se javljaju takvi rezultati mogu na prvi pogled djelovati neobično, ali zapravo nisu nimalo rijetke i uvelike su vezane uz prirodu podataka koji se analiziraju, a ponajprije uz njihovu distribuciju.

⁶⁴ Zainteresirani čitatelji mogu na stranici <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/using-lexical-features-to-investigate-second-language-lexical-decision-performance/1EE0B34CD6220D8AB3FBB2730A786D36#supplementary-materials> preuzeti cjelokupne podatke o vremenu reakcije (posljednji pristup 21. 4. 2024.).

Naime, takve „neobične“ vrijednosti dobit će se ponajprije kada su vrijednosti asimetrično grupirane na jednome kraju mogućega raspona. U takvim je situacijama zapravo upitno treba li izračunavati prosjek i standardnu devijaciju jer oni podrazumijevaju simetričnu distribuciju podataka (v. potpoglavlje 3.1.5). Također su moguće situacije u kojima standardna devijacija ima vrijednost veću od aritmetičke sredine. U istraživanju Berger i sur. (2019) to se dogodilo kod tzv. ortografskoga susjedstva riječi, gdje je dobiveno da riječi uključene u eksperimentalni zadatak u prosjeku imaju 4,978 ortografskih susjeda, uz standardnu devijaciju od 6,327. Vrlo je česta pojava da istraživači pri susretu s takvim rezultatima posumnjaju u pogrešku u izračunavanju, ali ni takvi podaci zapravo nisu rijetkost i samo upućuju na vrlo visoku varijabilnost (ili na vrlo asimetričnu distribuciju).

Standardna devijacija važna je i u definiranju i procjeni distribucije podataka. Kao što ćemo detaljnije vidjeti u potpoglavlju 3.1.5, u normalnoj distribuciji približno 68 % podataka nalazi se unutar prosjeka \pm jedna standardna devijacija, dok se 95 % podataka nalazi unutar prosjeka \pm dvije standardne devijacije. Dodatno je važno istaknuti da je standardna devijacija zapravo izvedena iz druge, osnovnije mjere varijabilnosti, odnosno predstavlja kvadratni korijen **varijance** (engl. *variance*), koja je jednaka prosjeku kvadriranih odstupanja (gdje je prosjek izračunat dijeljenjem s $N-1$). Varijanca je iznimno važna za statističko zaključivanje i inferencijalne statističke testove (jedna od najčešće korištenih skupina testova nosi naziv upravo „analiza varijance“, v. potpoglavlje 5.3.1, a kad se govori o količini varijabilnosti koja se može objasniti nekim čimbenikom, zapravo se misli na količinu varijance; takav je primjer već dan u potpoglavlju 2.3 kod objašnjenja mjera unutarnje konzistentnosti). U deskriptivnoj statistici varijanca se susreće rjeđe od standardne devijacije, koja se izražava u istim jedinicama kao i aritmetička sredina, pa je jednostavnija za tumačenje.

Još jedna vrlo važna mjera varijabilnosti, posebno često korištena kad podaci ne zadovoljavaju uvjete za izračunavanje standardne devijacije, jest **interkvartilno raspršenje** (engl. *interquartile range*), koje se još naziva i **interkvartilni raspon**, **interkvartilna razlika** ili jednostavno **interkvartil** (skraćeno IQR). U pitanju je mjera koja prati medijan kao mjeru centralne tendencije i koja se izračunava kao razlika između trećega i prvoga **kvartila** (engl. *quartile*), pri čemu se kvartili definiraju kao vrijednosti koje dijele skup podataka poredanih po veličini na četiri jednaka dijela. Medijan, koji dijeli podatke na dvije polovine, predstavlja drugi kvartil. Treći kvartil (uobičajeno označen kao Q_3) vrijednost je koja dijeli podatke tako da je 75 % podataka

ispod nje, a 25 % podataka iznad nje, dok je kod prvoga kvartila (Q_1) obrnuto – on je vrijednost koja dijeli podatke tako da je 25 % podataka ispod, a 75 % podataka iznad nje. O tim ćemo mjerama dodatno govoriti u sljedećemu potpoglavlju.

Budući da prvi i treći kvartil omeđuju središnjih 50 % podataka, interkvartilno nam raspršenje pokazuje koliko je ta središnja polovina podataka grupirana oko medijana – što je IQR veći, to su podaci raspršeniji. Prilikom tumačenja može biti korisno uzeti u obzir (i navesti u izvještaju o istraživanju) i vrijednosti prvoga i trećega kvartila kako bismo vidjeli između kojih je točno vrijednosti grupirano središnjih 50 % rezultata. Dodatno se dijeljenjem interkvartilnoga raspršenja s dva može izračunati i **poluinterkvartilno raspršenje** (engl. *semi-interquartile range*; v. Petz i sur., 2012, str. 54), koje nam se na prvi pogled može učiniti sličnijim standardnoj devijaciji. Međutim kako medijan ne mora biti točno na polovini središnjih 50 % podataka, tumačenje varijabilnosti u tome se slučaju ne može uvijek temeljiti na načelu medijan \pm poluinterkvartilno raspršenje. Ta se „prepolovljena“ mjera stoga ukupno gledano rjeđe susreće od interkvartilnoga raspršenja.

Primjer navođenja interkvartilnoga raspršenja nalazimo u istraživanju koje su provele Kim i McDonough (2008), gdje se medijan upotrebljava za izražavanje broja pasivnih oblika glagola koje učenici engleskoga jezika kojima je materinski jezik korejski upotrebljavaju tijekom interaktivnih aktivnosti s udešavanjem (u kojima istraživač upotrebljava pasive kao poticaj) u različitim eksperimentalnim situacijama. Autorice navode naprimjer da su učenici srednje razine općega znanja engleskoga jezika u situaciji gdje je glagol isti, upotrijebili medijalnu vrijednost od 6 pasiva uz IQR od 5, dok je na višoj razini znanja medijan iznosio 9, a IQR 3, što pokazuje veću varijabilnost kod skupine učenika na srednjoj razini znanja. Primjer navođenja vrijednosti i prvoga i trećega kvartila susrećemo i u metaanalizama (naprimjer kod Plonsky i Derrick, 2016, str. 544, u analizi vrijednosti koeficijenata pouzdanosti koji se tipično susreću u usvajanju drugoga jezika) i u usporedbi vlastitih rezultata s onima uobičajenima u području. Diependaele i sur. (2011) naprimjer proveli su zadatak leksičke odluke s udešavanjem i usporedili rezultate dobivene kod morfološki transparentnih (*viewer-view* 'gledatelj-pogled'), netransparentnih (*corner-corn* 'kut-kukuruz') i samo formalno povezanih parova riječi (*freezer-free* 'hladnjak-slobodan') kod izvornih i neizvornih govornika engleskoga jezika. Budući da su u dijelu posvećenome izvornim govornicima njihovi rezultati pokazali da nema statistički značajne razlike između brzine odgovora za netransparentne i formalno povezane parove riječi (gdje je prosječna razlika iznosila 15

milisekundi), autori su se oslonili na vrijednosti medijan=16, Q1=12, Q3=18, koje jesu bile slične ranijim istraživanjima, kako bi ustvrdili da njihov rezultat ipak ne odudara od rezultata ranijih istraživanja te da razlika između različitih parova riječi postoji. Naposljetku, Cunningham-Andersson (1997) navodi poluinterkvartilno raspršenje kako bi izrazila varijabilnost u stavovima dviju skupina izvornih govornika švedskoga prema izgovoru neizvornih govornika švedskoga, izraženih na ljestvici od 1 do 5, čija je centralna tendencija izražena putem medijana i moda.

Ako aritmetičku sredinu i medijan promatramo kao osnovne statističke modele kojima pokušavamo na najbolji mogući način predstaviti značajke skupine podataka, standardna devijacija i interkvartilna razlika mogu se promatrati kao **mjere pogreške** u modeliranju (engl. *error term*). Naime, kako navodi Field (2018), i kako je već spomenuto u potpoglavlju 1.3, statistička analiza u osnovi uvijek predstavlja pokušaj da se pojave koje istražujemo modeliraju na najbolji mogući način, pri čemu se ne mogu izbjeći pogreške (što se može predstaviti jednadžbom $ishod = model + pogreška$). Drugim riječima, aritmetička sredina predstavlja opću tendenciju u skupini rezultata, ali pojedinačni će rezultati odstupati od nje, pa standardna devijacija izražava uobičajenu mjeru odstupanja, odnosno pogreške. Field (2018) i Larson-Hall (2010, str. 64–66) ilustriraju kako su varijanca i standardna devijacija izvedene iz **sume kvadrata** (engl. *sum of squares*, gdje „kvadrat“ označava kvadirane razlike u odnosu na prosjek), koja je brojnik u razlomku u formuli za standardnu devijaciju navedenoj u (4) i koja predstavlja **najmanji kvadrat** među mogućima (engl. *least square*), zahvaljujući tome što prosjek predstavlja crtu koja prolazi kroz podatke tako da je udaljenost pojedinačnih vrijednosti od nje najmanja moguća – drugim riječima, kad su svi uvjeti za njegovu uporabu ispunjeni, predstavlja najbolji model za danu skupinu podataka.

Konačno, i standardna devijacija i interkvartilna razlika dodatno se mogu upotrebljavati u identificiranju netipičnih podataka. O tome detaljno govorimo u potpoglavlju 3.1.6.

3.1.4 Položajne mjere

Nekoliko pristupa podacima temelji se na relativnome položaju pojedinačnih vrijednosti u odnosu na ostatak uzorka. Takve se mjere nazivaju položajima ili **mjerama relativnoga položaja** (engl. *measures of relative standing*, v. Phakiti, 2014, str. 170). U prethodnim smo potpoglavljima već opisale neke od mjera koje pripadaju toj skupini, a koje odlikuje to što su u prvome koraku utemeljene na **rangiranju** (engl. *ranking*). Rangiranje podataka podrazumijeva

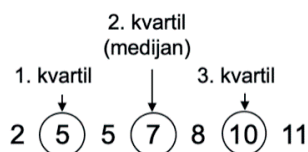
svrstavanje konkretnih zabilježenih vrijednosti u opadajući niz, od najveće do najmanje, i dodjeljivanje ranga od 1 do n svakoj pojedinačnoj vrijednosti, ili obrnuto u slučaju rastućega niza – izbor ovisi o prirodni varijabli. Opis podataka utemeljen na rangovima posebno je karakterističan za ordinalne ljestvice, koje već po svojoj prirodi predstavljaju rangove, a može se primijeniti i na intervalne i omjerne ljestvice.

Vrlo je česta pojava u rangiranju da se neka vrijednost pojavi više puta, odnosno da postoje **izjednačeni** ili **vezani rangovi** (engl. *tied ranks*). U takvim se slučajevima za potrebe statističke analize svako pojavljivanje određene vrijednosti rangira jednako, kao prosječna vrijednost dijeljenih mjesta na rang-ljestvici. Naprimjer u nizu 11, 10, 7, 5, 5, 2, u kojemu se vrijednost 5 javlja na 4. i na 5. mjestu, rangovi bi bili određeni kao 1, 2, 3, 4,5, 4,5, 6, tj. umjesto položaja 4 i 5 dvaput se javlja položaj 4,5.

Najvažnije mjere utemeljene na rangovima jesu **centili** (engl. *centiles*) (nazvani i **percentili**, engl. *percentiles*) i već spomenuti **kvartili**, koji se svi zajedničkim imenom nazivaju i **kvantili** (engl. *quantiles*) (v. npr. Levshina, 2015, str. 44). Radi se o mjerama koje dijele rangirane podatke na određeni broj jednakih dijelova. Percentili tako dijele niz podataka na 100 dijelova, te ih ima najviše smisla upotrebljavati u većim skupovima podataka (kakvi se u istraživanjima usvajanja drugoga jezika rjeđe susreću). Naprimjer u istraživanju odnosa između dubine vokabulara i (uspješnosti) uporabe strategija inferenciranja u drugome jeziku Nassaji (2006) dijeli učenike engleskoga kao drugoga jezika u dvije skupine na temelju njihovih rezultata u testu dubine vokabulara: rezultati leksički vještijih čitača bili su iznad 50. percentila, dok su rezultati leksički manje vještih čitača bili izjednačeni s 50. percentilom ili su bili ispod njega.

Kvartili, kao što je već navedeno u prethodnome potpoglavlju, dijele skup podataka na četiri jednaka dijela, pri čemu prvi kvartil (Q1) ostavlja ispod sebe 25 % podataka i odgovara 25. percentilu, drugi kvartil (Q2) dijeli podatke na dvije polovine, odnosno ostavlja ispod sebe 50 % podataka i predstavlja medijan, tj. 50. percentil, dok treći kvartil (Q3) ostavlja ispod sebe 75 % podataka, odnosno predstavlja 75. percentil. Takva podjela podataka prikazana je na slici 14, gdje se prvi i treći kvartil izračunavaju kao (pod)medijan donje i gornje polovine podataka nakon izdvajanja medijana cijeloga niza. Međutim takav je opis kvartila zapravo okviran i oni se u praksi mogu izračunati na više različitih načina, ponajprije s obzirom na to računa li se ili ne vrijednost koja predstavlja medijan među podatke iz kojih se izračunavaju prvi i treći kvartil. Slika 14 predstavlja tzv. „ekskluzivni“ primjer u kojemu se medijan u drugome

koraku ne razmatra; kad bi pristup bio „inkluzivan“, prvi bi kvartil opet iznosio 5 (kao prosjek dviju petica), dok bi treći iznosio 9, odnosno $(8+10)/2$. Statistički programi obično implementiraju više različitih načina izračunavanja kvartila, te može biti potrebno provjeriti koje se načelo primjenjuje i eventualno ga prilagoditi.⁶⁵



Slika 14. Određivanje kvartila

Primjer uporabe kvartila nalazimo u istraživanju percepcije naglaska u engleskome kao drugome jeziku kod izvornih govornika njemačkoga koje su proveli Mitterer i sur. (2020). U tome su istraživanju na temelju rangiranja ocjena njihova naglaska u drugome jeziku, koje su bile na ljestvici od 1 do 6, 24 ispitanika podijeljena u četiri kvartila, od kojih se onda formiralo šest skupina ispitanika s po jednim ispitanikom iz svakoga kvartila za sudjelovanje u glavnome eksperimentu.

Pored već opisane uporabe u istraživanjima, jedan praktični primjer uporabe kvartila može biti kad je u svrhu dodjeljivanja stipendija potrebno odrediti 5 % najboljih studenata, što znači da među ocjenama treba pronaći vrijednost iznad koje se nalazi 5 % podataka. Taj nam primjer također jasno ukazuje na vrlo važnu značajku mjera utemeljenih na rangovima – poslije rangiranja, kao prvoga koraka, zanemaruju se točne vrijednosti podataka i uzima se u obzir samo njihov redoslijed. U navedenome će primjeru sa stipendijama tako biti svejedno je li prvi student koji dobiva stipendiju imao ocjenu 4 ili 5, samo će biti važno da se njegova ocjena nalazi u gornjih 5 % unutar skupine. Ako su u pitanju podaci s intervalne ili omjerne ljestvice, na taj način jedan dio raspoloživih informacija ostaje neiskorišten, pa položajne mjere obično nisu jedine koje se računaju za takve podatke (naravno, pod uvjetom da su oni normalno distribuirani, v. sljedeće potpoglavlje). Mjere utemeljene na relativnim položajima jedine su prikladne za podatke s ordinalne ljestvice.

⁶⁵ Naprimjer R ima čak 9 načina izračunavanja kvartila, v. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/quantile> (posljednji pristup 21. 4. 2024.).

Kako bi uvjeti uporabe bili prikazani na jednome mjestu, navodimo ih u tablici 9. Za sve vrste ljestvica mogu se izračunavati i mjere navedene za ljestvice koje se nalaze iznad njih (naprimjer medijan za omjernu ljestvicu), ali ne i obrnuto (naprimjer medijan za nominalnu ljestvicu).

Tablica 9. Usporedba mjernih ljestvica prema dopuštenim deskriptivnim mjerama

Mjerna ljestvica	Dopuštene operacije	Dopuštene deskriptivne mjere
Nominalna	=, ≠	frekvencija, mod
Ordinalna	>, <	medijan, interkvartilni raspon
Intervalna	+, -	aritmetička sredina, standardna devijacija
Omjerna	×, /	aritmetička sredina, standardna devijacija, koeficijent varijabilnosti

Još jednu vrstu mjere položaja rezultata u skupini predstavljaju i **z-vrijednosti** (engl. *z-scores*), također zvane i **standardne vrijednosti** (engl. *standard scores*). One se izvode iz vrijednosti zabilježenih u rezultatima tako što se od konkretne vrijednosti oduzme aritmetička sredina skupine, a razlika se podijeli standardnom devijacijom, kao što je prikazano u formuli u (5).

$$(5) z = \frac{x-M}{s}$$

Z-vrijednost 0 označava vrijednost jednaku aritmetičkoj sredini skupine, vrijednost 1 je za jednu standardnu devijaciju viša od aritmetičke sredine, vrijednost -2 je za dvije standardne devijacije niža od aritmetičke sredine itd. Drugim riječima, na z-vrijednostima temelji se **standardna ljestvica** (engl. *standard scale*), odnosno ljestvica s aritmetičkom sredinom 0 i standardnom devijacijom 1. Tako dobivene standardne vrijednosti podliježu, naravno, svim uvjetima za izračunavanje aritmetičke sredine i standardne devijacije. Obično se z-vrijednosti ne izračunavaju (samo) u svrhu opisa, već kao jedan oblik transformacije podataka – korisne su kad je potrebno usporediti rezultate testova ili zadataka koji su upotrebljavali različitu ljestvicu ili različito bodovanje. Naprimjer Dallas i sur. (2013) prilikom mjerenja razine općega znanja engleskoga kao drugoga jezika kod svojih ispitanika (čiji je materinski jezik kineski) formiraju kompozitnu mjeru izvedenu iz više testova (rječničkoga dijela Wechslerove ljestvice inteligencije, Shipleyeva testa vokabulara i točnosti u zadatku leksičke odluke); kako svaki od tih testova upotrebljava drukčiju mjernu ljestvicu, oni najprije rezultate svakoga ispitanika na pojedinačnim testovima pretvaraju u z-vrijednosti, a zatim zbrajanjem

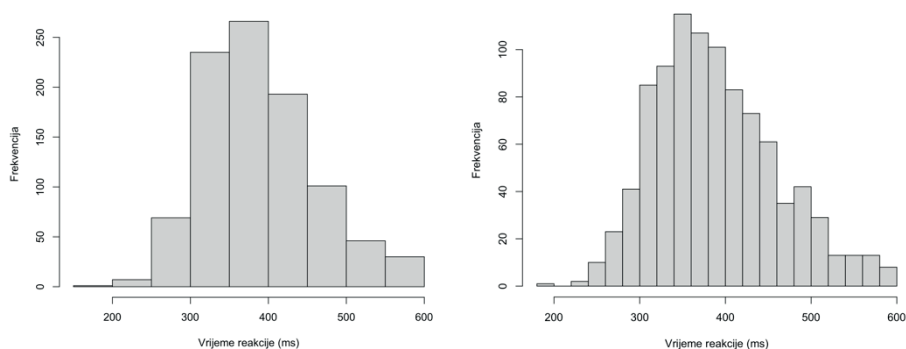
triju dobivenih z-vrijednosti izvode konačnu mjeru svakoga ispitanika (v. i Petz i sur., 2012, str. 103–109). Također se susreće uporaba z-vrijednosti za transformaciju postotaka; naime zbog ograničenosti raspona vrijednosti koje mogu poprimati, postoci mogu biti problematični za neke vrste analize (v. i primjere standardnih devijacija u potpoglavlju 3.1.3), pa se istraživači često služe njihovom transformacijom u z-vrijednosti (v. npr. Mayo i sur., 1997, koji tako transformiraju postotke točnih odgovora u zadatku u kojemu je trebalo slušati rečenice s bukom u pozadini). Budući da se temelje na standardnoj devijaciji, z-vrijednosti također se mogu upotrebljavati za utvrđivanje netipičnih vrijednosti, prema istome načelu opisanome u potpoglavlju 3.1.6, odnosno uzimajući kao graničnu vrijednost $z=\pm 2$, $z=\pm 2,5$ ili $z=\pm 3$ (v. Levshina, 2015, str. 59). Z-vrijednosti usko su povezane s normalnom distribucijom, o kojoj govorimo u sljedećemu potpoglavlju.

3.1.5 Distribucija podataka

Pojam distribucije podataka odnosi se na **distribuciju frekvencija** (engl. *frequency distribution*), odnosno **raspodjelu učestalosti**. Distribucija frekvencija predstavlja rezultat grupiranja mjera frekvencije prema vrijednostima, tj. pokazuje kolikom se frekvencijom u skupini podataka javljaju različite vrijednosti. Distribucije frekvencija najčešće se prikazuju grafički i iznimno su važne za početni uvid u podatke. Iako je grafički prikaz podataka tema posebnoga potpoglavlja, ovdje ćemo unaprijed uvesti nekoliko često korištenih vrsta grafikona jer su vrlo korisni u fazi upoznavanja s podacima i u fazi odabira odgovarajućih mjera centralne tendencije i varijabilnosti, kao i odgovarajućih statističkih testova.

Iako je čitateljima možda najpoznatiji način prikaza distribucije frekvencija kružni dijagram (popularna „torta“, odnosno „pita“, engl. *pie chart*), takav prikaz ima vrlo ograničenu korisnost u opisu podataka i zapravo se rijetko susreće u znanstvenim radovima (vratit ćemo se na njega u potpoglavlju 3.2.1). Distribucija frekvencija unutar statističke analize grafički se najčešće prikazuje uz pomoć **histograma** ili **poligona frekvencija** (v. npr. Dragičević, 2009, str. 40–50; Petz i sur., 2012, str. 85; engl. *histogram, frequency polygon*). Oni se vrlo često upotrebljavaju u području usvajanja drugoga jezika pa na slici 15 dajemo dva primjera za tu vrstu prikaza. Podaci su u oba slučaja isti i potječu iz istraživanja koje su proveli Wang i Kaiser (2022), koje se bavi sintaktičkim udešavanjem i implicitnim učenjem u kontekstu izloženosti zadatku koji nije jezični, nego uključuje statističke pravilnosti u redosljedu vizualnih podražaja (konkretno, slika glave psa i plavih krugova). Podaci su preuzeti sa stranice <https://osf.io/czy5h/> (posljednji pristup 21. 4. 2024.), a posebno napominjemo da naš prikaz

obuhvaća vremena reakcije iz dokumenta Rtdata_WindorizedADJ.csv, koji se odnosi na nejezični zadatak, bez izračunavanja izvedenih mjera za pojedinačne ispitanike i eksperimentalne uvjete, ali uz ograničavanje – iz praktičnih razloga – na uzorak od 948 rezultata kod kojih je vrijeme reakcije <600 ms. Zbog svega navedenog, naše grafikone treba shvatiti samo kao primjer prikaza, a ne kao pokazatelj rezultata rada, koji su mnogo složeniji (i ukupno pokazuju da učenje izvanjezičnih pravilnosti ne vodi izravnoj primjeni sličnih pravilnosti u jezičnome zadatku).



Slika 15. Histogrami vremena reakcije (podaci prilagođeni iz Wang i Kaiser, 2022)

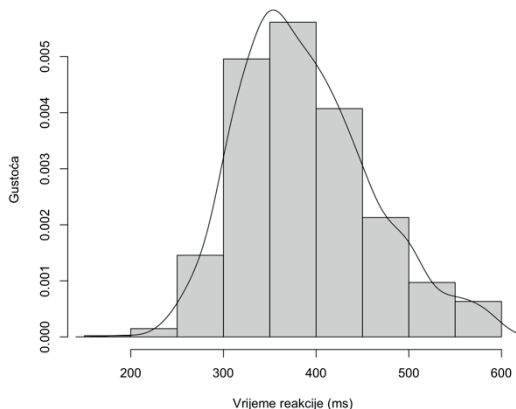
Histogrami su vizualno slični grafikonima s vertikalnim stupcima, s time što je kod njih na osi y obavezno prikazana neka mjera frekvencije (najuobičajenija je apsolutna frekvencija, ali je moguće prikazati i relativnu), a vrijednosti prikazane na osi x moraju biti numeričke i moraju pripadati ljestvici koja je ili kontinuirana (dopušta decimalne vrijednosti) ili diskretna (cijeli brojevi) s vrlo velikim brojem mogućih vrijednosti (čime se približava kontinuiranim podacima). Ako pak baratamo kategorijama, poput materinskoga jezika, ili malim brojem diskretnih vrijednosti, kao u slučaju broja jezika koje ispitanici govore, i želimo prikazati njihovu frekvenciju, za ispitivanje distribucije podataka upotrijebit ćemo klasične grafikone sa stupcima (v. potpoglavlje 3.2.1). Formalno gledano, ono što još razlikuje histograme i grafikone sa stupcima jest izostanak razmaka između stupaca kod histograma (Todorović, 2008, str. 207), koji odražava kontinuiranu prirodu mjerne ljestvice.⁶⁶

Kod izrade histograma posebnu pozornost treba posvetiti formiranju raspona vrijednosti („razreda“) u koje će podaci biti podijeljeni i koji će biti

⁶⁶ Pojedini programi izrađuju histograme s razmacima između stupaca (npr. dodatak Analysis ToolPak u Microsoftovu Excelu), ali takve je grafikone bolje urediti tako da se razmaci uklone.

predstavljani na osi x; tako se histogrami na slici 15 razlikuju upravo prema broju razreda, koji je na desnome histogramu veći. Programi u kojima se mogu izraditi histogrami određuju broj raspona automatski, na temelju značajki raspoloživih podataka, ali uvijek postoji i mogućnost izmjena i zadavanja raspona prema vlastitoj procjeni zasnovanoj na prirodi podataka. Važno je imati na umu da bilo da se na osi y prikazuje apsolutna ili relativna frekvencija, svi rasponi moraju biti jednake širine. Kad rasponi nisu jednake širine, na osi y mora se prikazati **gustoća frekvencije** (engl. *frequency density*), koja se dobiva tako što se frekvencija podijeli širinom razreda, što je potrebno učiniti kako bi površina stupca adekvatno predstavljala frekvenciju (usp. Petz i sur., 2012, str. 84, gdje se govori o prikazivanju apsolutne frekvencije koja je prilagođena širini razreda). To je bitno istaknuti zbog toga što je zbroj površina svih stupaca histograma uvijek jednak ukupnoj frekvenciji svih razreda.

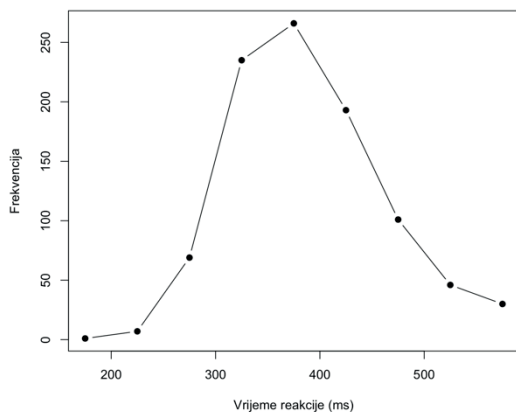
Dodatno se na histogramima može prikazati jedan drugi oblik gustoće, koji treba razlikovati od upravo opisanoga. U pitanju je **gustoća vjerojatnosti**, odnosno funkcija gustoće vjerojatnosti (u literaturi na engleskome jeziku naziva se *probability density*, ali često i jednostavno kao gustoća, *density*). U tome se slučaju uz stupce na histogram može dodati, ili samostalno prikazati, **krivulja gustoće** (eng. *density curve*, v. npr. Levshina, 2015, str. 51–52), koja uz određeno poravnanje (engl. *smoothing*) prati oblik koji ocrtavaju stupci histograma (što je više stupaca, to će i sam histogram više biti nalik na krivulju). Na slici 16 ponavljamo lijevi prethodno prikazani histogram uz dodatak krivulje gustoće. Način na koji možemo protumačiti prikazane rezultate jest taj da gustoća predstavlja mjeru vjerojatnosti pojavljivanja različitih raspona rezultata; naprimjer budući da je gustoća vjerojatnosti vremena reakcije koje iznosi između 300 i 350 ms približno jednaka 0,005, a raspon vrijednosti iznosi (približno) 50 ms, možemo zaključiti da je vjerojatnost pojavljivanja vrijednosti iz toga raspona jednaka $0,005 \cdot 50$, odnosno približno 25 %. Takvu procjenu možemo napraviti zahvaljujući tome što je površina ispod krivulje gustoće uvijek jednaka 1.



Slika 16. Histogram vremena reakcije s krivuljom gustoće (podaci prilagođeni iz Wang i Kaiser, 2022)

Poligon frekvencija temelji se na istome načelu prikaza, s time što on ne uključuje stupce već nalikuje linijskome grafikonu. Kako navode Petz i sur. (2012, str. 84), „[p]oligon se crta tako da iznad sredine svakog razreda označimo točke u visini ordinate, koja odgovara frekvenciji toga razreda.“⁶⁷ Označene točke zatim se spajaju crtama te je poligon frekvencija vizualno najbliži linijskim grafikonima (v. sliku 17). Takva vrsta vizualizacije može biti preglednija od histograma ako želimo zajedno prikazati distribuciju podataka u više uzoraka (naprimjer kod izvornih i neizvornih govornika, usp. primjer u Petz i sur., 2012, str. 87), ali se u općenitome slučaju prednost daje histogramima jer su precizniji u prikazu informacija, a mogu prikazati i rezultate više skupina zajedno. Načelo implementirano u poligonu frekvencija vrlo je korisno kad je potrebno prikazati kumulativne frekvencije, što se obično čini putem **krivulje kumulativnih frekvencija** (Petz i sur., 2012, str. 86 i 88; engl. *cumulative frequency curve*), koja se još naziva i **ogiva** (engl. *ogive*). Taj nam prikaz podataka pruža mogućnost da lako utvrdimo koliki se dio rezultata nalazi ispod ili iznad određene vrijednosti, što u istraživanjima usvajanja drugoga jezika može biti korisno naprimjer prilikom određivanja razine općega znanja drugoga jezika na temelju testa koji je istraživač sam izradio (v. Kraš i Miličević, 2015, str. 45–48). Takve je podatke moguće prikazati i na **histogramu kumulativnih frekvencija** (engl. *cumulative frequency histogram*).

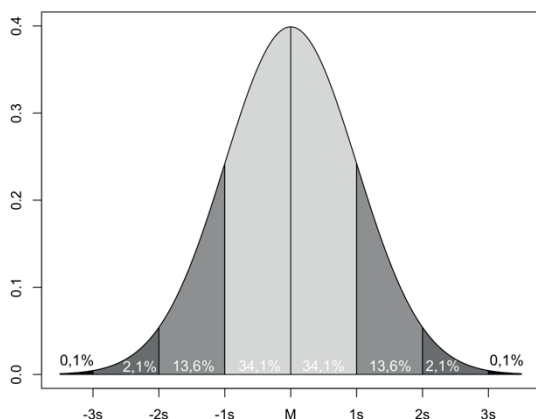
⁶⁷ Podsjećamo čitatelje da je ordinata naziv za os y, dok se os x još naziva i apscisa.



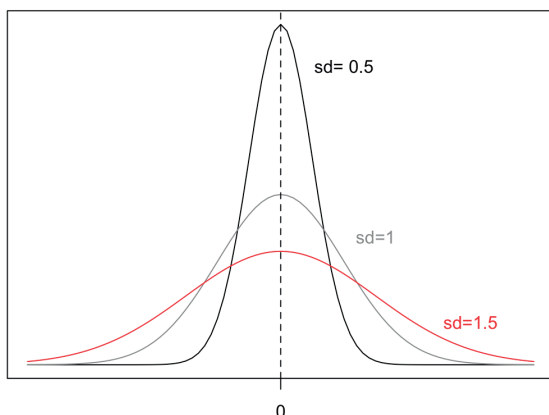
Slika 17. Poligon frekvencija vremena reakcije (podaci prilagođeni iz Wang i Kaiser, 2022)

Poligoni frekvencija i histogrami mogu nam dati početni uvid u to slijede li numerički podaci kojima raspolažemo **normalnu distribuciju** (engl. *normal distribution*). Normalna distribucija je ona u kojoj najveći dio skupa podataka ostvaruje vrijednost blisku aritmetičkoj sredini skupa, s odstupanjima koja su raspoređena simetrično u odnosu na prosjek i s relativno rijetkom pojavom vrijednosti koje su vrlo udaljene od prosjeka. Za normalnu distribuciju karakterističan je oblik **zvonaste** ili **Gaussove krivulje** (engl. *bell-shaped curve*, *Gaussian curve*), prikazan na slici 18 (os y prikazuje gustoću frekvencije). Ta je distribucija frekvencija u potpunosti opisana aritmetičkom sredinom i standardnom devijacijom. Kao što se može vidjeti na slici, kod podataka koji su normalno distribuirani, nešto više od 68 % podataka nalazi se u rasponu od $M \pm 1 \cdot SD$, približno 95 % podataka je u rasponu od $M \pm 2 \cdot SD$, dok je 99,7 % podataka u rasponu od $M \pm 3 \cdot SD$ (o čemu je već bilo riječi u potpoglavljima 3.1.2 i 3.1.3). Normalna je distribucija ujedno jedna od **distribucija vjerojatnosti** (engl. *probability distribution*), tj. matematička funkcija koja nam pokazuje koliko su različite vrijednosti vjerojatne.

Ovisno o tome koliko je velika standardna devijacija, normalna krivulja može imati djelomično drukčiji oblik („izduženiji“ kod male standardne devijacije, a „spljošteniji“ kod veće standardne devijacije, što predstavlja značajke s obzirom na **spljoštenost** ili **kurtičnost**, engl. *kurtosis*), ali su njezine osnovne značajke vezane uz frekvenciju podataka uvijek iste. To pokazuje slika 19 (generirana prilagođavanjem koda za R iz Larson-Hall, 2010, dostupnoga na stranici <https://routledge.com/textbooks/9781138024571/>, posljednji pristup 26. 4. 2024.).



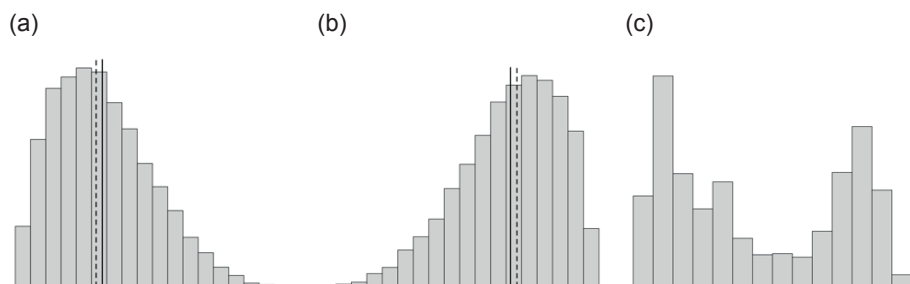
Slika 18. Normalna distribucija



Slika 19. Primjeri normalnih distribucija s prosjekom 0 i različitim standardnim devijacijama

U populacijama i dovoljno velikim uzorcima velik broj pojava teži tome da ima normalnu distribuciju. U svakodnevnom se životu kao primjer može navesti visina ljudi. U područjima bližima jeziku velik je broj pojava vezanih uz kognitivno funkcioniranje ljudi normalno distribuiran – vjerojatno najpoznatiji primjer predstavlja distribucija kvocijenta inteligencije u populaciji. Budući da se u eksperimentalnim istraživanjima jezika najčešće ispituje ponašanje ispitanika koje se može svrstati u domenu kognitivne obrade, normalno distribuirani podaci relativno su česti, iako u usvajanju drugoga jezika zbog varijabilne prirode učeničkoga jezika možemo očekivati više odstupanja od normalnosti kod neizvornih nego kod izvornih govornika.

Zbog svojih matematičkih značajki, normalna je distribucija najvažnija vrsta distribucije u statistici. Upoznavanjem njezinih značajki postaje jasnije zbog čega je ona uvjet za uporabu aritmetičke sredine i standardne devijacije. To bi trebalo biti još jasnije kad je usporedimo s distribucijama koje odstupaju od normalne. Na slici 20 prikazani su primjeri triju takvih distribucija. Primjeri pod (a) i (b) međusobno su zapravo vrlo slični jer ih odlikuje **asimetrična distribucija** (engl. *skewed distribution*), odnosno grupiranje podataka na jednome od krajeva ljestvice uz dugačak „rep“ na suprotnome kraju.⁶⁸ Vrsta distribucije pod (a) naziva se **pozitivno asimetričnom** (engl. *positively skewed, right-skewed*), dok je slika pod (b) ilustracija **negativno asimetrične** (engl. *negatively skewed, left-skewed*) distribucije. Takvi nazivi mogu zvučati neintuitivno jer se podaci grupiraju na strani suprotnoj od one označene nazivom, pa vrijedi napomenuti da se ti nazivi izvode iz indeksa asimetrije koji se izračunava kao u formuli u (6), prema Petz i sur. (2012, str. 75), odnosno uzimajući u obzir aritmetičku sredinu (M), medijan (C) i standardnu devijaciju (s). Indeks će imati pozitivan predznak kad je prosjek veći od medijana (što će se dogoditi kad su podaci „pomaknuti“ k nižim vrijednostima), a negativan predznak u suprotnome slučaju (puna crta na slikama 20a i 20b predstavlja aritmetičku sredinu, a isprekidana crta medijan). U svakodnevnome životu primjer pozitivno asimetrične distribucije u većini država predstavlja prosječna zarada – ona je obično viša od vrijednosti zarade koju ostvaruje većina pojedinaca zbog utjecaja manjega broja iznimno visokih zarada. U istraživanjima je u slučaju opažanja takvih distribucija nužno poslužiti se medijanom, interkvartilom i odgovarajućim daljnjim testovima jer oni adekvatnije opisuju podatke. Petz i sur. (2012, str. 74–75) dodatno svraćaju pozornost na provjeru korištenoga instrumenta i uvjeta provedbe eksperimenta ako smo na temelju prethodnih istraživanja očekivali normalno distribuirane podatke, a nismo ih dobili.



Slika 20. Primjeri distribucija koje odstupaju od normalne

⁶⁸ Field i Wilcox (2017, str. 22) navode da je u analizi debeo rep (engl. *heavy tail*) distribucije veći problem od asimetričnosti, a takav se rep može javiti i kod normalne distribucije.

$$(6) \alpha = \frac{3(M-C)}{s}$$

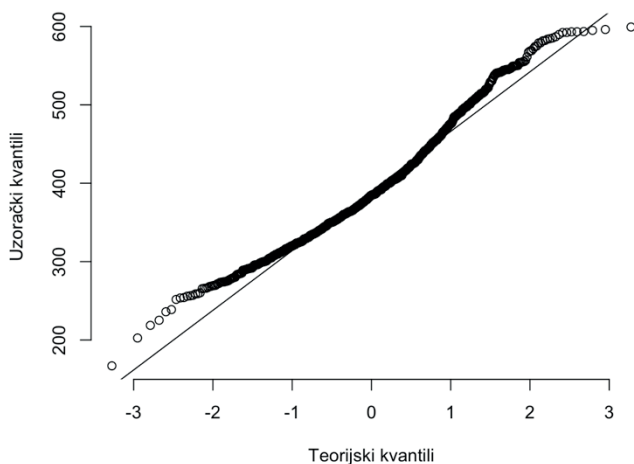
Distribucija pod (c) nešto je drukčije prirode i naziva se **bimodalnom** (engl. *bimodal*) jer sadrži podatke grupirane oko dviju dominantnih vrijednosti. Takvi se podaci ne susreću često u lingvistički orijentiranim istraživanjima usvajanja drugoga jezika i tipičniji su za istraživanja pojava poput stavova, gdje je česta prisutnost međusobno suprotstavljenih vrijednosti, ili naprimjer za istraživanja konkretnosti riječi (v. npr. Levshina, 2015, str. 105). U istraživanjima jezika pojava bimodalne distribucije može upućivati na prisutnost neke ometajuće varijable ili na problem s mjernim instrumentom te svakako treba temeljito provjeriti hipoteze i nacrt istraživanja. Kod takvih distribucija ni prosjek ni medijan ne predstavljaju dobar opis podataka jer oba padaju između dviju dominantnih vrijednosti. To je dakle jedna od situacija u kojima je mod prikladna mjera.

Te spoznaje o normalnoj i drugim vrstama distribucija možemo iskoristiti prilikom analize vlastitih podataka tako što ćemo za naše podatke nacrtati histogram i na njemu pogledati nalikuje li distribucija normalnoj ili nekoj drugoj vrsti. Pritom ne treba u rezultatima očekivati idealnu normalnost distribucije, koja se u praksi vrlo rijetko susreće. No manja odstupanja ne moraju nužno narušiti normalnost. Budući da nam vizualizacija može pomoći u procjeni, ali često ne i dati sasvim jasan odgovor, u četvrtome ćemo poglavlju govoriti i o statističkim testovima kojima se može provjeriti odgovara li distribucija dobivenih podataka normalnoj distribuciji; tu ćemo objasniti i da u statističkoj analizi zapravo nije ključna normalna distribucija podataka u uzorku, već to dolazi li uzorak iz populacije čije su vrijednosti normalno distribuirane (v. potpoglavljje 4.3.2).⁶⁹

Prije nego zaključimo ovo potpoglavljje, ističemo još jedan način prikaza distribucije frekvencija u svrhu provjere normalnosti, a to je **grafikon kvantila** ili **Q-Q grafikon** (engl. *quantile-quantile* ili *Q-Q plot*). Takav je grafikon prikazan na slici 21. Na njemu se na jednoj osi prikazuju vrijednosti zabilježene u uzorku (u našem primjeru na osi *y*), a na drugoj (kod nas osi *x*) očekivane vrijednosti standardne normalne distribucije (koja predstavlja normalnu distribuciju izraženu *z*-vrijednostima, što je objašnjeno u potpoglavljju 3.1.4). Ako su pojedinačni podaci (prikazani kružićima) distribuirani u obliku koji je približno jednak ravnoj crti, radi se o normalnoj distribuciji. Što više podataka odstupa od crte, to je vjerojatnije da distribucija ne odgovara normalnoj distribuciji (v. npr.

⁶⁹ Linstromberg (2020) vraća pozornost na to da je odnos između distribucije u uzorcima i populaciji vrlo složen. Normalno distribuiran uzorak može se dobiti i iz populacije koja nije normalno distribuirana, kao i obrnuto.

Larson-Hall, 2010, str. 82–84; Levshina, 2015, str. 53). Naziv grafikona potječe od toga što točke na grafikonu predstavljaju uparene kvantile dviju distribucija – broj kvantila koji se prikazuju može biti različit, uz tendenciju da bude velik. Kad su uzorci na osima x i y jednake veličine, uobičajeno je da se umjesto kvantila jednostavno prikažu sortirane pojedinačne vrijednosti.⁷⁰



Slika 21. Grafikon kvantila (za podatke prilagođene iz Wang i Kaiser, 2022)

U potpoglavlju 3.2.3 vidjet ćemo da je taj grafikon vrlo sličan tzv. dijagramu raspšenja, odnosno predstavlja jednu njegovu podvrstu.

3.1.6 Dodatne operacije na podacima

Više smo puta u ovome poglavlju odgađale raspravu o netipičnim podacima (v. posebno potpoglavlje 2.2.2). Razlog je tome povezanost takvih podataka s oblikom distribucije frekvencija. Budući da nam je sad i distribucija frekvencija poznat pojam, te smo uvidjeli značaj normalne distribucije, u ovome se potpoglavlju možemo opsežnije pozabaviti mogućnostima tretiranja „neobičnih“ podataka.

Podsjećamo da „netipičnima“ nazivamo podatke čija je vrijednost uočljivo udaljena od većine ostalih podataka i od mjere centralne tendencije skupine. Takve vrijednosti značajno povećavaju varijabilnost u podacima i time negativno

⁷⁰ Kada se usporedba vrši sa standardnom normalnom distribucijom, koja je teorijski konstrukt, podrazumijeva se uzorak jednake veličine kao onaj koji sadrži podatke iz istraživanja. Niz primjera za različite oblike distribucija može se vidjeti na stranici [https://www.ucd.ie/ecomodel/Resources/ QQplots_WebVersion.html](https://www.ucd.ie/ecomodel/Resources/QQplots_WebVersion.html) (posljednji pristup 26. 4. 2024.).

utječu na niz drugih mjera, uključujući i preciznost statističkih testova. One se mogu javiti u svakome istraživanju, a u istraživanjima usvajanja drugoga jezika možemo ih očekivati još i više nego u nekim drugim područjima zbog toga što je učeničkome jeziku posebno svojstvena varijabilnost (v. npr. Nicklin i Plonsky, 2020, str. 27).⁷¹

Jedno moguće načelo za otkrivanje takvih podataka temelji se na značajkama normalne distribucije, te se kao netipične definiraju one vrijednosti koje su više od dvije, više od dvije i pol ili više od tri standardne devijacije udaljene od aritmetičke sredine. Izbor udaljenosti ovisi o uobičajenoj praksi u području i o temi koja se istražuje, a dijelom i o procjeni istraživača. Druga veličina koja se može upotrijebiti u svrhu određivanja granice ispod ili iznad koje se podaci smatraju netipičnima jest interkvartilna razlika. Posebno je uobičajen pristup onaj prema kojemu se kao netipične klasificiraju vrijednosti manje od $Q1 - 1,5 * IQR$ ili veće od $Q3 + 1,5 * IQR$.⁷² Deniz (2022) naprimjer upotrebljava taj kriterij u istraživanju uporabe sintaktičkih i semantičkih tragova u obradi složenih rečenica u engleskome kao prvome i drugome jeziku te u turskome kao prvome jeziku tako što isključuje iz analize brzine čitanja pojedinih rečeničnih segmenata u zadatku čitanja vlastitim tempom one vrijednosti koje su veće ili manje od $Q1/Q3 \pm 1,5 * IQR$. To se načelo primjenjuje i pri crtanju često korištenih dijagrama s pravokutnikom (engl. *boxplots*), o kojima ćemo detaljno govoriti u potpoglavlju 3.2.2 (v. i Lindstromberg, 2020, gdje se govori o „pravilu dijagrama s pravokutnikom“ u određivanju netipičnih vrijednosti). Treba međutim istaknuti i jednu nelogičnu stranu određivanja netipičnih podataka na temelju mjera varijabilnosti – te su mjere izračunate uzimajući u obzir vrijednosti koje zatim želimo isključiti, a jasno je da su one znatno utjecale na veličinu mjera varijabilnosti, posebno standardne devijacije (v. Field, 2018; Nicklin i Plonsky, 2020). Dodatne mogućnosti uključuju odbacivanje određenoga postotka najnižih i najviših vrijednosti bez obzira na to jesu li identificirane kao netipične prema mjerama varijabilnosti ili nisu.

Pored netipičnih vrijednosti, za koje smo vidjeli da mogu bitno utjecati na mjere centralne tendencije (posebice aritmetičku sredinu), problem u analizi podataka može stvoriti i distribucija podataka koja odstupa od normalne.

⁷¹ Među brojnim metodološkim kritikama koje se istraživačima usvajanja drugoga jezika upućuju jest i ona na račun nedostatka transparentnosti u radovima u pogledu toga kako su tretirane netipične vrijednosti (v. npr. Marsden, Thompson i Plonsky, 2018; Nicklin i Plonsky, 2020; Plonsky i Ghanbar, 2018).

⁷² Mogu se međutim susresti i drukčije granice, posebice u istraživanjima koja upotrebljavaju vrijeme reakcije ili dužinu zadržavanja pogleda. Naprimjer Felser i sur. (2009) govore o „ekstremnim netipičnim vrijednostima“ i primjenjuju pravilo od 3 puta IQR; v. i Diependaele i sur. (2011).

Već smo prije napomenule da se kod takvih podataka ne bi smjele računati aritmetička sredina i standardna devijacija, kao niti provoditi određeni statistički testovi, što istraživačima bitno sužava izbor analiza. Istraživači stoga ponekad pokušavaju pronaći način da distribuciju koju su dobili u podacima pretvore u normalnu, a jedan je od načina i uklanjanje netipičnih vrijednosti jer su u nekim slučajevima upravo one krive za distribuciju koja odstupa od normalne. U ovome potpoglavlju govorimo o različitim postupcima koji se mogu primijeniti kad želimo eliminirati netipične podatke i pretvoriti distribuciju u normalnu. Međutim prije nego što spomenemo relevantne strategije važno je naglasiti da prethodno opisane načine prepoznavanja netipičnih vrijednosti treba pažljivo primjenjivati u svjetlu distribucije podataka. Naime, kako ističe Lindstromberg (2020), moramo paziti na simetričnost distribucije jer se s asimetričnim distribucijama može dogoditi da podaci budu identificirani kao netipični i onda kada oni to zapravo nisu ako uzmemo u obzir oblik distribucije – ako distribucija odgovara log-normalnoj, odnosno logaritamskom transformacijom daje normalnu distribuciju, bolje ju je transformirati (v. u nastavku) nego tražiti netipične vrijednosti putem mjera varijabilnosti. Dodatni je oprez potreban kod manjih uzoraka, kakvi su u istraživanjima usvajanja drugoga jezika vrlo česti, ako ne i potpuno dominantni (v. npr. Plonsky, 2013 – medijalna veličina uzorka iznosi 19 ispitanika), jer su takvi uzorci podložniji svim vrstama odstupanja (Linstromberg, 2020).

Kao što je u potpoglavlju 2.2.2 navedeno za nedostajuće podatke, i netipične se vrijednosti mogu u nekim slučajevima isključiti iz analize, što se naziva **podrezivanjem** (engl. *trimming*).⁷³ Kao i za popriličan broj drugih odluka, i ovdje je važno upoznati se s postupcima koji su uobičajeni u srodnim istraživanjima. Naprimjer u zadacima leksičke odluke, zadacima čitanja vlastitim tempom i drugim zadacima utemeljenima na mjerenju vremena reakcije uobičajeno je da se iz analize isključe odgovori koji su prebrzi ili prespori. Kao što se može vidjeti iz primjera koje daju McDonough i Trofimovich (2009, str. 152–153) te Nicklin i Plonsky (2020, str. 28), uobičajeno je da se odrede granice valjanih odgovora u milisekundama, bilo u apsolutnome smislu (naprimjer 100, 200 ili 300 ms kao minimum, a 1400, 2000, 3000 ili 5000 ms kao maksimum kod zadatka leksičke odluke) ili relativno s obzirom na dobivene podatke (prosjek ± 2 , 2,5 ili $3 \cdot SD$). Ahn (2021) kombinira ta dva moguća pristupa u istraživanju obrade članova u engleskome kao drugome jeziku kod izvornih govornika korejskoga, gdje iz analize podataka prikupljenih zadatkom čitanja vlastitim tempom isključuje sve brzine čitanja kraće od 200 ms i duže od 2.500 ms kao i one koje su tri

⁷³ Naravno, sva se izračunavanja potom ponavljaju bez isključenih vrijednosti.

standardne devijacije ili više od toga udaljene od prosječne brzine čitanja. Tim je postupkom isključeno 2,2 % podataka iz analize.

Pristup utemeljen na relativnome određivanju valjanih odgovora obično podrazumijeva standardnu devijaciju u odnosu na ukupni prosjek svih rezultata, što može dovesti do uklanjanja vrijednosti koje unutar svojih eksperimentalnih uvjeta zapravo ne odstupaju previše od mjere centralne tendencije, kao i do zadržavanja vrijednosti koje u svojim eksperimentalnim uvjetima jesu netipične; detaljnije o tim problemima, uz primjere, govore Lachaud i Reanud (2011), koji nazivaju takvo odbacivanje podataka „slijepom eliminacijom“ i ukazuju na brojne probleme koje ono može izazvati. Ti autori predlažu oprez prilikom primjene toga postupka, a smatraju da uz njega u najmanju ruku treba provesti **analizu osjetljivosti** (engl. *sensitivity analysis*), u kojoj će se usporediti rezultati koji se dobivaju bez uklanjanja podataka i s njime.

Uobičajeno je i da se u potpunosti odbacuju cjelokupni podaci ispitanika kojima je više od određenoga postotka odgovora bilo pogrešno (naprimjer Duffield i White, 1999, postavljaju granicu od 15 %). Odbacivanje veće količine netipičnih podataka trebalo bi se ipak primjenivati samo ako postoji opravdan razlog za to, naprimjer ako brzina rješavanja zadatka jasno upućuje na to da ga neki ispitanik nije rješavao na predviđeni način ili ako se naknadno utvrdi da prema biografskim varijablama neki ispitanik odudara od ostalih ispitanika. Kao što je već navedeno u potpoglavlju 2.2.2, Lachaud i Renaud (2011) svraćaju pozornost da je uklanjanje ispitanika opravdano samo ako smo sigurni da su podaci izostali ili su neodgovarajući bez ikakva sustavnog razloga. Larson-Hall i Herrington (2010, str. 377–378) također posvećuju podosta pozornosti razmatranju mogućnosti uklanjanja netipičnih podataka. Ti se autori s jedne strane slažu da uklanjanje može pomoći naprimjer u ostvarivanju normalne distribucije u podacima, koja omogućuje izračunavanje aritmetičke sredine i standardne devijacije, kao i u provedbi čitavoga niza statističkih testova (usp. potpoglavlje 4.3.2). S druge strane međutim ističu da netipične vrijednosti nikako ne treba uklanjati *ad hoc* jer to dovodi do povećane subjektivnosti u istraživanju i čini preostale podatke međusobno ovisnima.⁷⁴

Drugi mogući način rješavanja problema netipičnih vrijednosti ne uključuje njihovo isključivanje, već njihovo zamjenjivanje drugim vrijednostima, pri čemu se zamjena može provesti na različite načine. Jedan uključuje zamjenu najnižom ili najvišom vrijednošću koja ulazi u dopušteni raspon. Naprimjer McDonough i Trofimovich (2009, str. 153) za vrijeme reakcije navode

⁷⁴ Međusobna se ovisnost podataka ovdje ne odnosi na prirodu uzorka, već na to što uklanjanje jedne pojedinačne vrijednosti utječe na sve ostale – naprimjer utjecat će na promjenu medijana.

mogućnost vrijednosti koja je jednaka $M+3*SD$ kao zamjenu za vrlo visoke vrijednosti, a $M-3*SD$ za vrlo niske vrijednosti (ako je, naravno, granica određena kao $\pm 3*SD$); taj postupak često dovodi do toga da se u zamjeni upotrijebe vrijednosti koje se nisu niti javile u podacima, a karakteristično je za njega i to što se istim vrijednostima zamjenjuju podaci svih ispitanika (odnosno, gledaju se rezultati skupine, a ne pojedinačni), uglavnom promatrajući sve prikupljene podatke zajedno, odnosno ne tretirajući odvojeno različite eksperimentalne uvjete, kao i kod uklanjanja netipičnih podataka (usp. Lachaud i Reanud, 2011). Druga je proširena praksa **winsoriziranje** (engl. *winsorizing*, prema Charles P. Winsoru), gdje se uklonjeni netipični podaci zamjenjuju posljednjom zabilježenom vrijednošću koja ulazi u dopušteni raspon. Tu je uobičajeno određivanje netipičnih vrijednosti kao određenoga postotka podataka (naprimjer 10 %) sa svake strane. Važno je istaknuti i da se taj postupak primjenjuje na pojedinačnim podacima svakoga ispitanika, za svaki eksperimentalni uvjet posebno. Barem kod nekih vrsta podataka (naprimjer brzina čitanja u zadacima čitanja vlastitim tempom ili brzina odgovora u zadatku leksičke odluke), winsoriziranje daje nešto bolje rezultate od podrezivanja (v. Nicklin i Plonsky, 2020), najvjerojatnije zahvaljujući zasebnome tretiranju individualnih podataka i eksperimentalnih uvjeta. Taj je postupak primijenjen u istraživanju koje su proveli Hulstijn i sur. (2009) u analizi rezultata vizualnoga zadatka leksičke odluke provedenoga unutar eksperimenta učenja riječi u engleskome kao drugome jeziku, u kojemu su sudjelovali izvorni govornici nizozemskoga. Za svakoga ispitanika tri najbrže i tri najsporije brzine odgovora zamijenjene su četvrtom najbržom i četvrtom najsporijom brzinom odgovora. U individualnim slučajevima u kojima se jedna od triju najbržih brzina čitanja ponovila dvaput, zamijenjene su četiri, a ne tri najbrže brzine odgovora.

Pored uklanjanja podataka iz analize i zamjene netipičnih vrijednosti drugim vrijednostima, postoji i mogućnost, često korištena naprimjer u psiholingvistici, **transformacije podataka** (engl. *data transformation*). Transformacija se provodi ne samo zbog netipičnih podataka već i zbog odstupanja distribucije podataka od normalne (v. Larson-Hall 2010, str. 92–93; McDonough i Trofimovich, 2009, str. 151–153).⁷⁵ Postupci transformacije za razliku od prethodno opisanih postupaka ne uključuju pročišćavanje

⁷⁵ McDonough i Trofimovich (2009) ističu da su vremena reakcije u zadacima utemeljenima naprimjer na udešavanju često pozitivno asimetrična (što povlači za sobom netipične vrijednosti u obliku vrlo sporih odgovora), što objašnjava česte transformacije podataka u psiholingvistici i psiholingvističkim zadacima u području usvajanja drugoga jezika (v. i Nicklin i Plonsky, 2020, koji ističu osjetljivost mjera utemeljenih na vremenu reakcije).

podataka (engl. *data cleaning*, v. Nicklin i Plonsky, 2020, str. 27), odnosno uklanjanje ili zamjenu pojedinih vrijednosti, već podrazumijevaju zamjenu svih vrijednosti novima. Larson-Hall (2010) ističe da je transformacija podataka u istraživanjima usvajanja drugoga jezika relativno rijetka, navodeći kao mogući razlog to što je istraživači doživljavaju kao vrlo složenu i teško razumljivu. Ta autorica ujedno ističe da je provedba transformacije u statističkim programima relativno jednostavna. Međutim istraživačima usvajanja drugoga jezika veći problem može predstavljati odluka o tome kakvu transformaciju odabrati s obzirom na značajke podataka, daljnju analizu i mogućnost razumijevanja transformiranih podataka.

Larson-Hall (2010, str. 92) preuzima od Tabachnick i Fidell (2001) prijedloge za najbolji oblik transformacije kod različitih asimetričnih distribucija. Ovdje nećemo ulaziti u sve ondje navedene detalje (koje prepuštamo čitateljima za samostalno proučavanje), već samo svratiti pozornost na to da se operacije transformacije kreću od zaokretanja ($1/X$) i vađenja kvadratnoga korijena do različitih logaritamskih operacija s konstantama koje se određuju iz podataka. U literaturi iz područja psiholingvistike najčešće su korištene vrste transformacije logaritamskoga tipa. Logaritmi su povezani s operacijama stupnjevanja i korjenovanja jer je logaritam stupanj na koji treba podići odabranu osnovu kako bi se dobila izvorna vrijednost ($\log_b(X)=Y$ ako je $X=b^Y$).⁷⁶ Najčešća je transformacija putem prirodnoga logaritma, u kojemu bazu (b) predstavlja matematička konstanta e (Eulerov broj, odnosno 2,718282), ali ovisno o vrsti podataka dobar izbor mogu biti i neke druge osnove. Transformirane podatke svakako prije daljnje analize treba provjeriti jer rezultat ne mora nužno biti normalna distribucija. Ako novodobiveni podaci i dalje nemaju normalnu distribuciju, netipični se podaci u njima mogu pojedinačno ispitati (v. Nicklin i Plonsky, 2020), u suprotnome ih nema previše svrhe zadržati u analizi, odnosno može biti bolje vratiti se izvornim podacima i tražiti (neparametrijsko) rješenje za njihovu analizu. Prije analize podataka dobivenih zadatkom čitanja vlastitim tempom Ahn (2021) transformira podatke putem prirodnoga logaritma te navodi da vrijednosti asimetričnosti i spljoštenosti upućuju na to da je normalnost distribucije poboljšana transformacijom.

Nicklin i Plonsky (2020) za zadatke čitanja u kojima se mjeri vrijeme reakcije na temelju isprobavanja različitih mogućnosti predlažu prirodnu logaritamsku transformaciju kao bolje rješenje od uklanjanja i zamjene podataka, navodeći da ona ima najveći utjecaj na „normalizaciju“ distribucije. To ne znači nužno

⁷⁶ Naprimjer ako je osnova 10, a vrijeme reakcije ili broj bodova u testu iznosi 100 (ms/bodova), transformirana vrijednost bit će 2 jer je 10^2 jednako 100.

da je isti postupak najpogodniji za druge vrste podataka u usvajanju drugoga jezika, ali svakako treba napomenuti da, ako se transformirani podaci nakon povoljnoga ishoda (dobivanja normalne distribucije) zadrže u analizi, potrebna je pozornost pri njihovu tumačenju jer više ne postoji izravan uvid u iste jedinice koje su dobivene prvotnim mjerenjem. Treba imati na umu i kako pojedini autori ističu da transformacije podataka utječu na to kakvim se veličinama barata u analizi (naprimjer njima se dolazi do uporabe geometrijske umjesto aritmetičke sredine, čime se mijenja priroda hipoteze koja se testira), ali i na to koji se konstrukti zapravo ispituju (usp. potpoglavlja 2.1.2 i 2.1.6, gdje se govori o operacionalizaciji varijabli) te da korisnost transformacija i opravdanost njihove uporabe treba pažljivo procijeniti (v. npr. Field i Wilcox, 2017, str. 22).

Konačna će odluka o netipičnim podacima i prilagođavanju distribucije normalnoj u velikoj mjeri ovisiti o konkretnoj situaciji u istraživanju i o tome jesu li određene vrijednosti relevantne za istraživanje ili ne, odnosno potječu li naprimjer od pomanjkanja pozornosti ispitanika (među istraživačima nisu rijetke anegdote o ispitanicima koji su zaspali tijekom istraživanja), u kojemu slučaju ne pridonose našem znanju o ispitivanoj pojavi, ili su istinski pokazatelji poznavanja neke jezične pojave koji iz nepoznatih razloga „odudaraju“. Moguće je i jednostavno zadržati netipične vrijednosti i poslužiti se izračunavanjem medijana, interkvartilnoga raspršenja i odgovarajućih inferencijalnih testova, ili se osloniti na postupke tzv. **robustne statistike** (engl. *robust statistics*), gdje se primjerice izračunavaju **podrezane aritmetičke sredine** (engl. *trimmed means*), koje predstavljaju svojevrsan kompromis između prosjeka i medijana (Larson-Hall i Herrington, 2010; v. i Lachaud i Renaud, 2011). Field i Wilcox (2017) također daju sličnu preporuku izvan područja usvajanja drugoga jezika, upućujući na postupke koji su otporni na netipične podatke, kakvi su postupci **uzorkovanja s povratom** (engl. *bootstrapping*; v. i Plonsky i sur., 2015). S obzirom na nedostatak jasnih smjernica (na što ukazuju i Nicklin i Plonsky, 2020), treba svakako razmotriti više rješenja. I svakako treba paziti da se u istraživanjima usvajanja drugoga jezika ne primijene previše doslovno kriteriji koji se upotrebljavaju za izvorne govornike jer učenike drugoga jezika obično odlikuje varijabilnije jezično ponašanje, rjeđa uporaba ekstremnih vrijednosti ljestvice u zadacima prosudbe prihvatljivosti, kao i sporije vrijeme reakcije u psiholingvističkim zadacima.⁷⁷ Konačno, Nicklin i Plonsky (2020) svraćaju

⁷⁷ Nicklin i Plonsky (2020) posebno svraćaju pozornost na apsolutno određivanje granica kod vremena reakcije, gdje nije prikladno preuzeti vrijednosti iz istraživanja s izvornim govornicima. To stvara dodatni problem kad u istraživanju sudjeluju i izvorni i neizvorni govornici jer bi za svake od njih trebalo odrediti drukčije apsolutne vrijednosti.

pozornost da se u izvještajima o istraživanju ne treba pozivati samo na „ustaljenu praksu“ u području, već objasniti kako su točno netipične vrijednosti identificirane i tretirane i zbog čega upravo tako; u suprotnome bi moglo djelovati, čak i kad to nije istina, da je postupak arbitraran i da je istraživač možda pokušao i više načina pa prihvatio onaj koji je dao željene rezultate (naprimjer statistički značajan rezultat testa). Idealno bi bilo unaprijed definirati postupak koji će se primijeniti (v. i potpoglavlje 7.1 o pretprijavi istraživanja za objavljivanje), a po mogućnosti čak i predvidjeti analize bez netipičnih vrijednosti i s njima, koje će zatim u radu biti prikazane i uspoređene (v. npr. Baayen i Milin, 2010).

3.2 Grafičko prikazivanje rezultata

Grafički prikaz prikupljenih podataka vrlo je važan dio njihova opisa jer se značajke podataka tim putem mogu najizravnije i na najpristupačniji način sažeti. Kako navodi Larson-Hall (2017, str. 245), „[g]rafikoni su mnogim ljudima inherentno privlačniji od numeričkih statističkih pokazatelja; oni olakšavaju uočavanje tendencija u svrhe početne procjene i analize podataka, i otkrivaju tendencije koje se inače ne bi mogle uočiti u sumarnim pokazateljima“ (prev. M. M. P. i T. K.). Tufte (2001, prema Larson-Hall i Plonsky, 2015) također ističe da grafički prikaz može biti precizniji i otkriti nam više o podacima od same statističke analize. To je u kontekstu istraživanja usvajanja drugoga jezika posebno važno istaknuti zbog toga što nekoliko metaanaliza pokazuje da u tome području istraživači smatraju vizualizaciju podataka prije poželjnom nego nužnom te se grafički prikaz može pronaći tek u polovini ili čak i manje radova (v. informacije koje o tome daju Larson-Hall i Plonsky, 2015).

Podaci se grafički mogu prikazati na velik broj različitih načina. Premda ne postoje stroga pravila kad koju vrstu grafikona treba upotrebljavati, moguće je izdvojiti tipične prikaze za različite vrste podataka. Dodatno je moguće napraviti razliku između grafikona koji nam ponajprije služe kako bismo bolje upoznali podatke i onih koje planiramo uključiti u izvještaj o istraživanju – prvi imaju kao glavni cilj omogućiti nam da uočimo tendencije u rezultatima (naprimjer normalnu ili neku drukčiju distribuciju, usp. potpoglavlje 3.1.5), dok je drugima svrha jasno predstaviti podatke ciljnoj publici. Prvi stoga često predstavljaju korak u analizi koji prethodi izračunavanju deskriptivnih mjera i pomaže u njihovu odabiru (v. npr. Johnson, 2013, str. 288, kao i Wilkinson i Task Force on Statistical Inference, 1999, p. 597, prema Larson-Hall i Plonsky, 2015). Međutim ti se grafikoni rjeđe uključuju u izvještaje, stoga je poželjno i u drugome tipu, onome namijenjenom izvještavanju, predočiti čitateljima

što više informacija, uključujući i distribuciju podataka (v. npr. Larson-Hall i Herrington, 2010, str. 369).

U ovoj su knjizi neki grafikoni koji pripadaju prvome tipu, poput histograma i Q-Q grafikona, već opisani u potpoglavlju 3.1.5, dok se ovdje bavimo drugim tipom, tj. grafikonima koji se uključuju u izvještaj o istraživanju (v. šesto poglavlje). U nastavku pobliže opisujemo nekoliko često korištenih vrsta, grupirajući ih prema tome koje deskriptivne mjere prikazuju. Dodatno uzimamo u obzir kriterije koje u svojim preporukama vezanim uz vizualizaciju podataka u području usvajanja drugoga jezika navodi Larson-Hall (2017). Naime ta autorica razlikuje grafičke prikaze podataka koji su „sveobuhvatni“ (engl. *data accountable graphics*) i one koji su „bogati“ (engl. *data rich graphics*), pri čemu je prva vrsta utemeljena na prikazivanju pojedinačnih podataka i njihove distribucije, eventualno uključujući i mjere centralne tendencije i mjere varijabilnosti, dok druga vrsta isključuje individualne podatke, ali uključuje distribuciju, a često i mjere centralne tendencije te mjere varijabilnosti (v. Larson-Hall, 2017, str. 244). Od već spomenutih grafikona prvoj vrsti pripadaju Q-Q grafikoni, a drugoj histogrami.

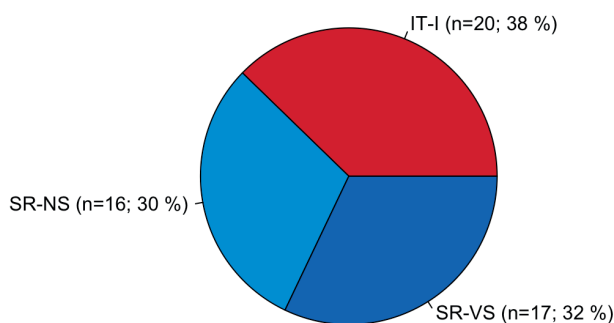
Grafikoni se mogu izraditi u bilo kojemu programu za statističku analizu, kao i Microsoftovu Wordu i PowerPointu. Među programima koje smo spomenule u potpoglavlju 2.2.3, najnaprednije grafičke funkcije implementirane su u programskome okružju R. Kako je već ranije navedeno, i ovdje se kao primjeri daju grafikoni koje su autorice izradile u R-u ili njegovu sučelju RStudio.

3.2.1 Prikaz frekvencija kategorija

Pregled različitih vrsta grafikona započinjemo onima koji se ne mogu sasvim doslovno uklopiti u navedenu podjelu koju predlaže Larson-Hall (2017) zbog toga što se upotrebljavaju u nacrtima kod kojih pojedinačni podaci iskazuju pripadnost kategoriji, pa ih ne bi imalo previše smisla pojedinačno vizualizirati. Umjesto toga grafički se prikazuje frekvencija različitih kategorija ili kombinacija kategorija.⁷⁸ U potpoglavlju 3.1.5 već je bilo riječi o histogramima i poligonima frekvencija kao prikazima koji se često upotrebljavaju kad želimo utvrditi distribuciju frekvencija u numeričkim podacima. U ovome se potpoglavlju usredotočujemo na prikaz frekvencija kod kategorijskih varijabli, s posebnim naglaskom na situacije koje uključuju dvije takve varijable.

⁷⁸ Larson-Hall (2017, str. 254) u tome slučaju navodi da su grafikoni sveobuhvatni kad su obuhvaćene sve frekvencije, ističući da frekvencije tu predstavljaju individualne podatke („*all of the data points in the data set*“). Mi smatramo da o frekvencijama, koje su ovdje deskriptivni pokazatelji izvedeni iz kategorijskih individualnih vrijednosti, ne treba govoriti kao o individualnim podacima, ali se slažemo s time da grafikoni o kojima govorimo sadrže sve najvažnije podatke.

Općenito gledano, kod prisutnosti samo jedne kategorijske varijable vjerojatno je najpoznatiji prikaz frekvencija **kružni dijagram** ili **strukturni krug**, poznatiji pod popularnim nazivom „torta“ (kako je već navedeno u 3.1.5). Taj prikaz predstavlja krug podijeljen na isječke čija veličina odražava frekvenciju pojedinačnih vrijednosti u nekome uzorku. Na slici 22 naprimjer možemo vidjeti distribuciju ispitanika u trima skupinama uključenima u istraživanje koje je provela Miličević (2007); kratice označavaju izvorne govornike talijanskoga jezika (IT-I) te učenike više srednje (SR-VS) i niže srednje (SR-NS) razine općega znanja talijanskoga jezika čiji je materinski jezik srpski. Kao što se na slici može uočiti, kružni dijagrami posebno su pogodni za prikazivanje relativne frekvencije, pri čemu isječki u zbroju daju 100 %.

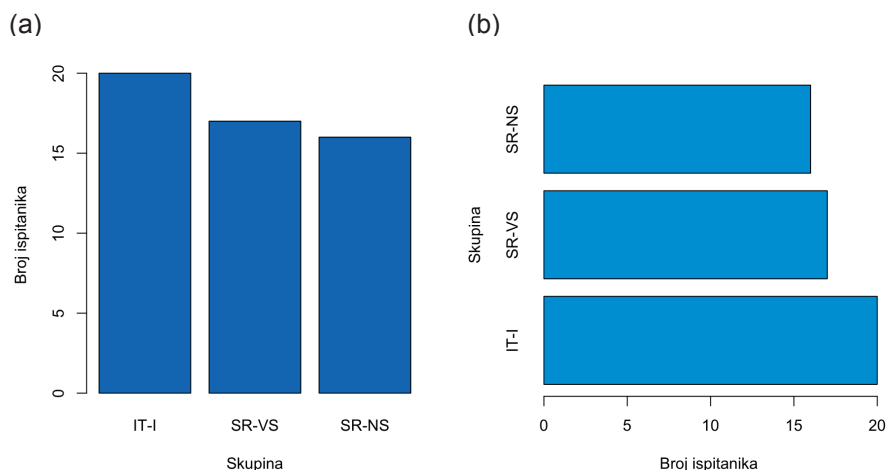


Slika 22. Kružni dijagram (prema Miličević, 2007)

Kao što je već spomenuto, iako popularan u neznanstvenim krugovima, u znanstvenim je istraživanjima kružni dijagram vrlo rijetko poželjan način prikaza podataka. U istraživanjima usvajanja drugoga jezika podatke poput onih na gornjoj slici najčešće možemo susresti u opisu ispitanika, a budući da taj opis obično sadrži čitav niz informacija, one se češće prikazuju u obliku tablice (v. šesto poglavlje). U analizi glavnoga eksperimentalnog zadatka čak i ako imamo kategorije u zavisnoj varijabli, najčešće će nacrt biti presložen da bi se mogao prikazati kružnim dijagramom.

Druga je česta vrsta prikaza frekvencija **stupčasti grafikoni** (engl. *column chart*, *bar chart*), na kojemu se na jednoj osi mogu prikazati kategorije, a na drugoj frekvencija, kao u primjerima na slici 23 (gdje se upotrebljavaju isti podaci kao i na slici 22). Izbor horizontalnih ili vertikalnih stupaca odredit će koja os mora predstavljati kategorije, a koja frekvenciju. Na engleskome se jeziku ponekad razlikuju *column chart* i *bar chart* kako bi se naglasio smjer grafikona (okomiti u prvome slučaju, kao na slici 23a, a vodoravni u drugome,

kao na slici 23b), ali je vrlo uobičajeno i da se *bar chart* upotrebljava kao opći naziv. U području usvajanja drugoga jezika okomiti se prikaz češće upotrebljava.

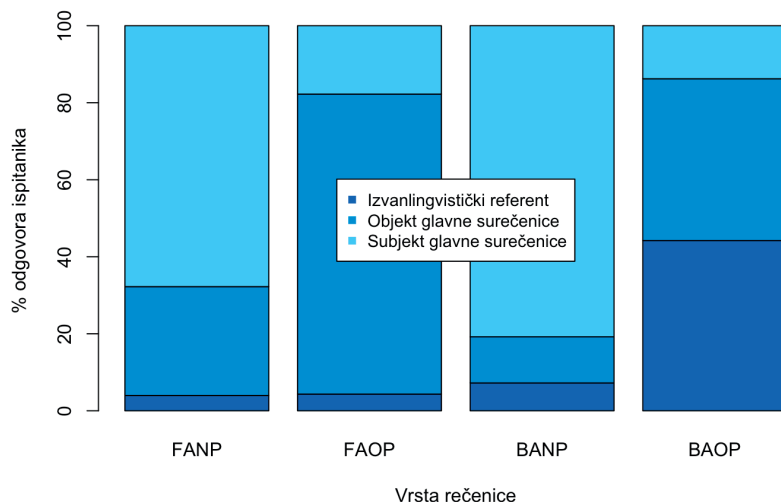


Slika 23. Stupčasti grafikoni za prikaz frekvencija (prema Miličević, 2007)

Uporaba stupčastih grafikona u svrhu prikazivanja frekvencija možda je jedina koja se danas smatra potpuno adekvatnom za podatke iz znanstvenih istraživanja (v. Larson-Hall, 2017, str. 253, kao i potpoglavlje 3.2.4, gdje se govori o manjkavostima takvih grafikona kao prikaza mjera poput aritmetičke sredine i medijana). Ako usporedimo takav prikaz s histogramima, vidimo da su stupci međusobno odvojeni razmakom te da naziv osi y ne mora biti „Frekvencija“ (iako mjera koja se prikazuje uvijek mora biti frekvencijska mjera), već se može prilagoditi sadržaju istraživanja.

Posebna je podvrsta takozvani **složeni stupčasti grafikon** (engl. *stacked bar chart*), u kojemu se relativne frekvencije prikazuju uz pomoć „složenih“ stupaca, u kojima se različite kategorije označavaju različitim bojama ili šarama unutar istoga stupca, dok različiti stupci mogu predstavljati naprimjer različite skupine ispitanika ili različite konstrukcije; na osi y prikazana je frekvencija. Posebno se često upotrebljava varijanta toga grafikona u kojoj svi stupci dosežu 100 %. Primjer takva grafikona dan je na slici 24, koja prikazuje dio podataka iz istraživanja Kraš (2008c) o izboru različitih referenata za izrečene i neizrečene subjektne zamjenice (subjekt glavne surečenice, objekt glavne surečenice, izvanlingvistički referent) u različitim vrstama rečenica u hrvatskome jeziku kod

izvornih govornika.⁷⁹ Kao što je već rečeno u potpoglavlju 2.2.2, gdje smo se referirale na istraživanja u kojima se upotrebljavao istovrsni eksperimentalni zadatak (Kraš, 2008b, 2016; Miličević i Kraš, 2017), FAOP i FANP su nazivi za eksperimentalne uvjete, kao i BAOP i BANP (engl. *forward/backward anaphora with an overt subject* i *forward/backward anaphora with a null subject*, odnosno anafora/katafora s izrečenim i neizrečenim subjektom).



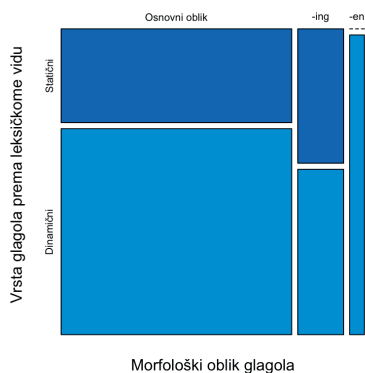
Slika 24. 100 % složeni stupčasti grafikon (prema Kraš, 2008c)

Još jedna korisna vrsta grafikona za situacije kad ispitujemo odnos dviju ukriženih kategorijskih varijabli jest **mozaični grafikon** (engl. *mosaic plot*). Putem mozaičnoga grafikona možemo naprimjer prikazati podatke iz tablice 5, iz istraživanja koje je proveo Robison (1990) i koje je prethodno kratko prikazano u potpoglavlju 3.1.1. U pitanju su podaci o uporabi glagola s različitim leksičkim vidom u različitim morfološkim oblicima, tj. različitim glagolskim vremenima, u engleskome kao drugome jeziku. Tako na slici 25a vidimo da je svaka varijabla (vrsta glagola prema leksičkome vidu i morfološki oblik glagola) prikazana na jednoj osi, pa svaki pravokutnik predstavlja jednu kategoriju nastalu križanjem vrijednosti dviju varijabli (naprimjer dinamični glagoli upotrijebljeni s nastavkom *-ing*). Dimenzije pravokutnika pokazuju veličinu različitih kategorija (vidimo naprimjer da dinamičnih glagola ima

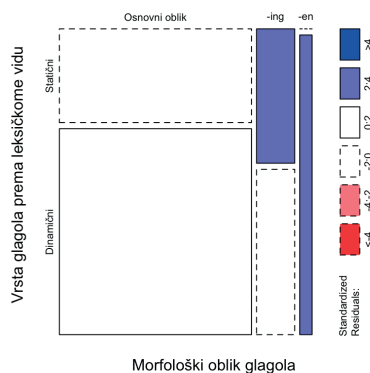
⁷⁹ Iako su u tome istraživanju sudjelovali izvorni, a ne neizvorni govornici, navodimo taj primjer jer je istovrsni eksperimentalni zadatak i istovrsni prikaz podataka bio upotrijebljen u nizu istraživanja usvajanja drugoga jezika povezanih s njime.

više nego statičnih), pa je zbog toga ta vrsta grafikona još informativnija od stupčastoga grafikona, na kojemu bi se također mogli prikazati ti podaci.

(a)

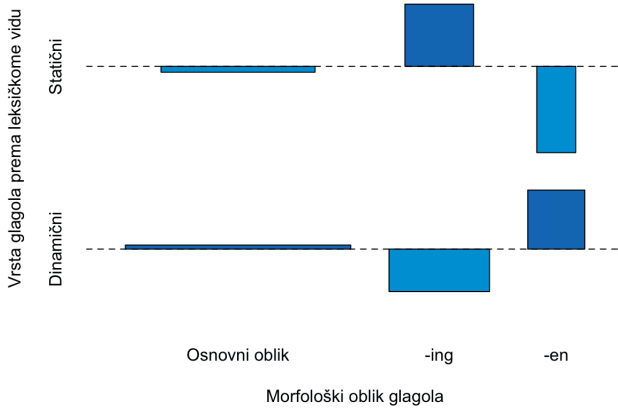


(b)



Slika 25. Mozaični grafikon (prema Robison, 1990)

Mozaični je grafikon moguće nacrtati i složenije, uzimajući u obzir i inferencijalne statističke pokazatelje, odnosno označavajući kod kojih je kategorija opažena frekvencija viša od teorijske (tj. one koja bi se mogla očekivati ako između prikazanih kategorija ne postoji povezanost), a kod kojih niža (o teorijskim frekvencijama detaljnije govorimo u potpoglavlju 5.1.1). Primjer vidimo na slici 25b, gdje pozitivna „standardizirana rezidualna odstupanja“ (engl. *standardized residuals*) znače višu opaženu frekvenciju, a negativna nižu. Tome je prikazu blizak i **asocijativni grafikon** (engl. *association plot*), koji također pokazuje gdje je opažena frekvencija viša od teorijske, a gdje niža, s time što su te situacije prikazane s različite strane osi. Podatke iz tablice 5 moguće je prikazati i tom vrstom grafikona, kao na slici 26.



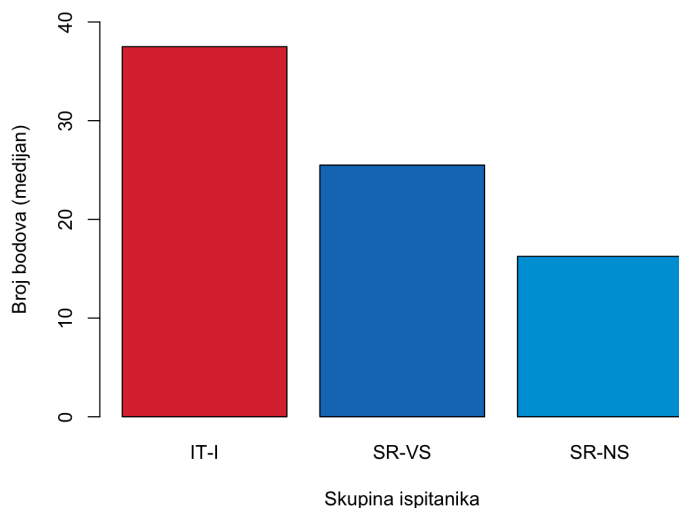
Slika 26. Asocijativni grafikon (prema Robison, 1990)

Na kraju ovoga potpoglavlja vraćamo pozornost da su navedeni primjeri vezani uz situacije u kojima imamo jednu ili dvije kategorijske varijable i frekvencija predstavlja jedinu deskriptivnu statističku mjeru koja se može izračunati. U nekim drugim slučajevima frekvencija može biti uključena u istraživanje ne kao deskriptivna mjera već kao numerička varijabla (naprimjer ako bismo analizirali koliko su ispravnih glagolskih oblika različite skupine učenika engleskoga kao drugoga jezika upotrijebile u zadatku proizvodnje) – tada za analizu vrijedi ono što je rečeno u potpoglavlju koje slijedi.

3.2.2 Prikaz mjera centralne tendencije i varijabilnosti

U istraživanjima usvajanja drugoga jezika vrlo se često susreću situacije u kojima se u nekome eksperimentalnom zadatku žele usporediti odgovori različitih skupina ispitanika, odnosno centralne tendencije njihovih odgovora. U takvim je situacijama dugo dominirao prikaz mjera centralne tendencije putem stupčastih grafikona poput onoga na slici 27 (gdje su prikazane medijalne vrijednosti bodova koje su ispitanici iz različitih skupina ostvarili na testu dopunjavanja teksta, kojim je u istraživanju Miličević, 2007, mjerena razina općega znanja talijanskoga jezika). Larson-Hall i Herrington (2010) nalaze takve grafikone u 46 od 110 studija koje su analizirali, a Larson-Hall (2017) u više od 70 % radova u pojedinim časopisima. Razlog je tome, po svoj prilici, jednostavnost toga prikaza, kao i lakoća predstavljanja na jednome grafikonu većega broja rezultata (naprimjer rezultata više skupina ispitanika). Međutim takva je uporaba stupčastih grafikona već neko vrijeme predmet vrlo oštih kritika u području usvajanja drugoga jezika kao i izvan njega (v. npr. Larson-

-Hall, 2017; Larson-Hall i Herrington, 2010; Larson-Hall i Plonsky, 2015). Na meti je kritika ponajprije to što svaki stupac prikazuje samo jednu informaciju o uzorku (mjeru centralne tendencije, poput medijana ili aritmetičke sredine), ne pružajući uvid ni u distribuciju podataka ni u eventualne netipične vrijednosti. Nije rijetkost ni tvrdnja da takvi grafikoni samo oduzimaju prostor koji bi se mogao iskoristiti za prikaz daleko veće količine podataka.⁸⁰

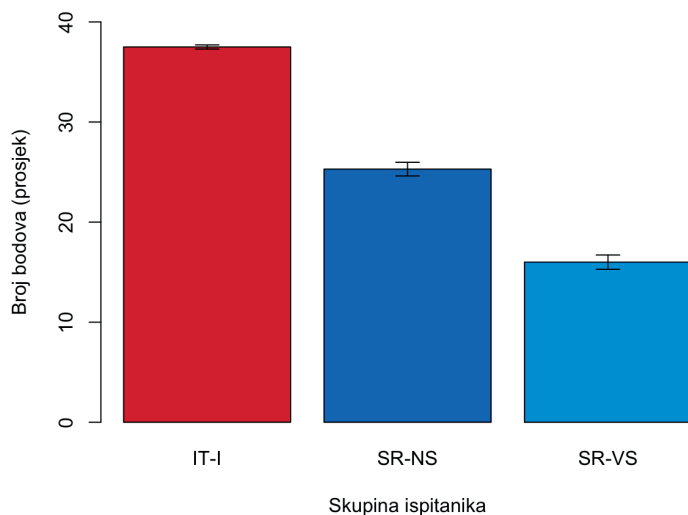


Slika 27. Stupčasti grafikon za prikaz mjera centralne tendencije (prema Miličević, 2007)

Nešto bolja mogućnost, koja sadrži informacije o varijabilnosti, jest **grafikon s oznakama pogreške**, čijim su stupcima dodane **crte koje označavaju intervale pogreške** (engl. *error bars*). Te crte mogu pokazivati standardnu devijaciju (obično ± 1 ili samo $+1$), ali još češće standardnu pogrešku (također najčešće ± 1 , iako je ± 2 po nekim autorima informativnije; v. Cumming, 2014, str. 18) ili granice pouzdanosti (Cumming, 2014, naprimjer predlaže da je najinformativnije prikazati granice pouzdanosti jer standardna pogreška nije uvijek izravno povezana s njima, a sama po sebi nema inferencijalni značaj). O standardnoj pogrešci i granicama pouzdanosti govorimo u četvrtome poglavlju, gdje će njihovo značenje i tumačenje biti razjašnjeno – one se naime mogu upotrebljavati za procjenu statističke značajnosti razlike između

⁸⁰ Naprimjer E. Tufte u forumskome komentaru navodi kako stupčasti grafikoni trate prostor te kako bi se u prostoru gdje se na njima prikazuje jedan broj moglo prikazati najmanje stotinu brojeva. (http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR, posljednji pristup 26. 4. 2024.)

uzoraka. Tu vrstu grafikona prikazujemo na slici 28 (za svaku su skupinu prikazani prosjek i standardna pogreška prosjeka), koristeći iste podatke iz Miličević (2007) kao na prethodnoj slici, uz napomenu da se i taj grafikon smatra manje informativnim od onih koji će biti uvedeni u drugoj polovini ovoga potpoglavlja.



Slika 28. Stupčasti grafikon s oznakama pogreške (prema Miličević, 2007)

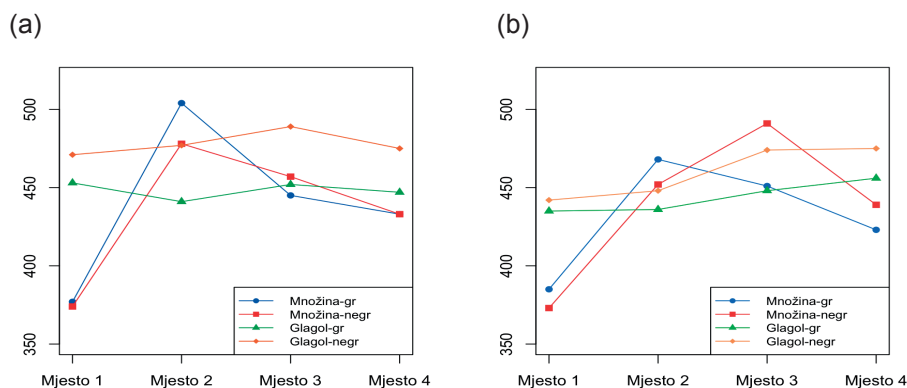
Na sličan se način gleda na **linijski grafikon** (engl. *line chart*), na kojemu je mjera centralne tendencije obično predstavljena točkom (ili nekim sličnim simbolom) umjesto stupcem, a točke su zatim povezane crtama (v. sliku 29); uz simbol je moguće dodati i oznake pogreške.⁸¹ Toj se vrsti grafikona priznaje ipak veća korisnost nego grafikonima sa stupcima jer se na njima može preglednije prikazati veća količina podataka kao i interakcija između varijabli (ili njezin izostanak; o tome će biti riječi u potpoglavlju 5.3.1).⁸² Dodatno, takvi su grafikoni korisni kad je potrebno prikazati podatke u vremenskome ili nekome drugom slijedu (naprimjer poznavanje određene jezične pojave na različitim razinama općega znanja drugoga jezika) jer mogu jasno vizualizirati promjene (Larson-Hall i Plonsky, 2015). U nekim se slučajevima na njima mogu prikazati i individualni podaci uz posebnu crtu za centralnu tendenciju skupine, naprimjer kada imamo rezultat istih ispitanika prije i poslije nastavne

⁸¹ Moguće je zapravo i izostaviti simbol i upotrebljavati samo crte, podrazumijevajući da njihovi krajevi označavaju relevantne vrijednosti.

⁸² Pojedini autori, čak i u slučaju prikaza interakcija, ističu da možda prikaz nekoliko informacija nije vrijedan prostora koji ti grafikoni zauzimaju (v. npr. Larson-Hall i Plonsky, 2015).

intervencije (v. Larson-Hall, 2017, str. 260).⁸³ Prikaz izmjena tijekom vremena zapravo je i njihova glavna namjena, i osim u posebnim slučajevima poput prikaza interakcije trebali bi se upotrebljavati samo za podatke koji su poredani prema vremenskome ili nekome sličnom kriteriju.

Kao primjer navodimo dio podataka iz istraživanja Jiang i sur. (2011; usp. i Jiang 2012, str. 176–177), posvećenoga čitanju vlastitim tempom. Ti su autori mjerili brzinu čitanja u različitim dijelovima rečenica, za različite vrste rečenica, kod izvornih govornika engleskoga jezika i učenika čiji je materinski jezik ruski ili japanski. Te je govornike bilo zanimljivo usporediti jer se ruski i japanski razlikuju u pogledu označavanja množine, što je dovelo i do razlika u osjetljivosti na negramatičnost kod tih skupina ispitanika. Na slici 29 prikazujemo njihove podatke – vidimo brzine čitanja na četirima različitim mjestima u rečenici za četiri različite vrste rečenica, od kojih su dvije gramatične, a dvije negramatične (slika lijevo prikazuje rezultate govornika japanskoga, a slika desno govornika ruskoga jezika kao materinskoga).



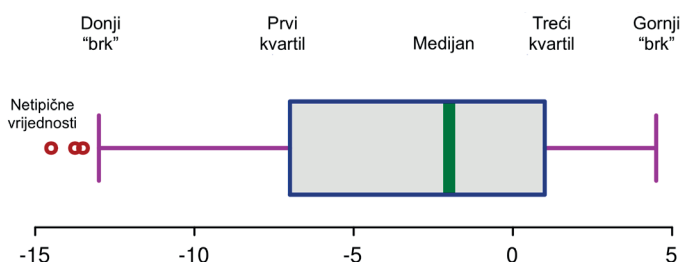
Slika 29. Linijski grafikoni (utemeljeni na podacima iz Jiang i sur., 2011)

Kad prijeđemo u domenu grafikona koji se preporučuju kao informativni, tj. bogati podacima, svakako među prvima treba navesti **grafikon s okvirima i brkovima** (engl. *box and whisker plot*, odnosno „kutija s brkovima“ ili *boxplot*), u hrvatskome jeziku poznat i kao **dijagram s pravokutnikom**, koji je predložio još Tukey (1977, prema Larson-Hall, 2017).⁸⁴ Shematski prikaz toga dijagrama

⁸³ Larson-Hall i Plonsky (2015) i Larson-Hall (2017) spominju i **paralelni dijagram** (engl. *parallel coordinate plot*) kao jedan od grafikona koji mogu biti korisni za prikaz ponovljenih mjerenja za različite skupine. No čini se da se tu ne radi o posebnoj vrsti grafikona, već o jednostavnome kombiniranju više linijskih grafikona na jednoj većoj slici.

⁸⁴ „Brkovi“ u nazivu grafikona potječu od njegove vodoravno okrenute varijante (slika 30; usp.

može se vidjeti na slici 30, gdje se uočava da on prikazuje čitav raspon rezultata, s označenim medijanom i netipičnim vrijednostima obilježenima kružićima (v. npr. Larson-Hall, 2012a, str. 253; Larson-Hall, 2017, str. 256; Larson-Hall i Herrington, 2010, str. 369–371). Ta vrsta grafikona temelji se na položajnim mjerama, točnije kvartilima: pored medijana označenoga središnjom debelom crtom, vidimo da pravokutnik obuhvaća središnjih 50 % podataka, odnosno ima prvi kvartil kao donju granicu i treći kvartil kao gornju; crte koje se protežu izvan kutija („brkovi“) označavaju granice izvan kojih se podaci smatraju netipičnima – kao što je već objašnjeno u potpoglavlju 3.1.3, te granice temelje se na interkvartilnoj razlici – s donje strane granicu određuje vrijednost koja je jednaka $Q1 - IQR * 1,5$, a s gornje $Q3 + IQR * 1,5$. Ako u podacima nema netipičnih vrijednosti, crte obuhvaćaju sve podatke do maksimalne, odnosno minimalne vrijednosti.



Slika 30. Objašnjenje dijagrama s pravokutnikom⁸⁵

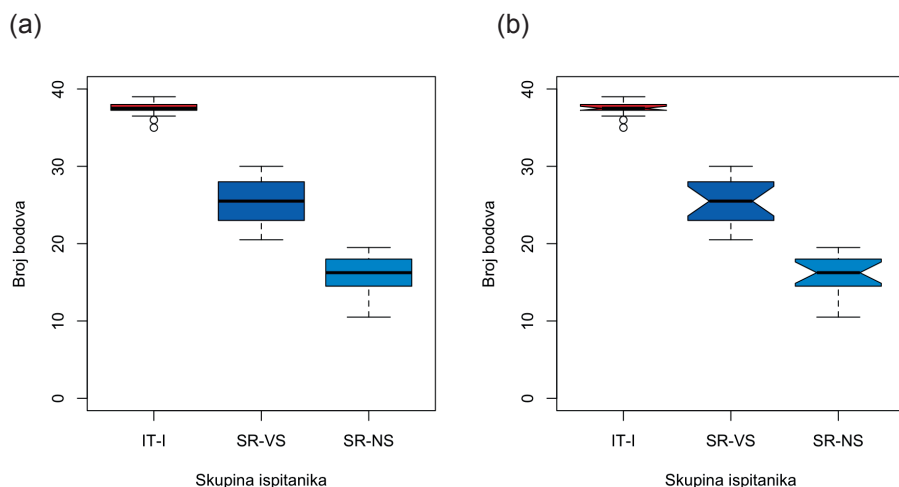
Dijagram s pravokutnikom vrlo je koristan za grafičko ispitivanje distribucije podataka jer veličinom pravokutnika pokazuje koliko su podaci grupirani oko centralne vrijednosti, odnosno kolika je njihova varijabilnost, a položajem medijana pokazuje i (a)simetričnost distribucije. Na dijagramima s pravokutnikom može se uočiti i „usjek“ (engl. *notch*), koji nam poput oznaka pogreške može pomoći u procjeni statističke značajnosti razlike između skupina. Usjek se proteže od medijana ka krajevima pravokutnika u širini

razliku između dviju vrsta grafikona sa stupcima), u kojoj se crte nalaze s lijeve i desne strane. U hrvatskome se jeziku često zadržava i engleski naziv *boxplot* (v. npr. Petz i sur., 2012, str. 472), dok naziv „grafikon s okvirima i brkovima“ nalazimo u lokaliziranim verzijama softverskoga paketa Microsoft Office (<https://support.microsoft.com/hr-hr/office/dostupne-vrste-grafikona-u-sustavu-office-a6187218-807e-4103-9e0a-27cdb19afb90#bmcolumncharts>, posljednji pristup 26. 4. 2024.), a „dijagram s pravokutnikom“ u Rječniku statističkog nazivlja Državnoga zavoda za statistiku (https://web.dzs.hr/app/rss/rjecnik_en-hr.html, posljednji pristup 26. 4. 2024.).

⁸⁵ Slika je preuzeta i (jezično) prilagođena sa stranice https://commons.wikimedia.org/wiki/File:Elements_of_a_boxplot.svg; RobSeb, CC BY-SA 3.0, putem Wikimedia Commons (posljednji pristup 26. 4. 2024.).

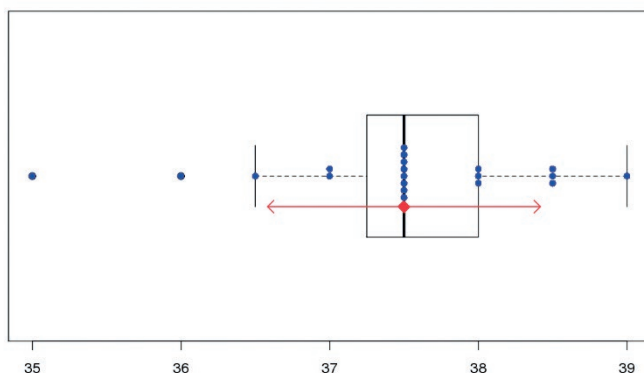
(odnosno visini) od $\pm 1,57 * IQR / \sqrt{n}$, što odgovara intervalu pouzdanosti medijana od 95 % (v. McGill i sur., 1978, str. 16). Ako se usjeci dviju skupina rezultata ne preklapaju, medijani tih skupina značajno se razlikuju (v. potpoglavlje 5.2). Treba primijetiti kako je moguće da se pojave usjeci koji se protežu izvan rubova kutije, koji se prikazuju kao „zavrnuti“ kutovi i upućuju na asimetričnu distribuciju podataka. Prednost je dijagrama s pravokutnikom i u tome što se mogu primijeniti kod svih numeričkih podataka, uključujući ordinalne, kao i kod različitih vrsta distribucija.

Kako bismo sažele značajke te vrste grafikona, na slici 31 prikazujemo rezultate koje su na testu razine općega znanja talijanskoga jezika (testu dopunjavanja teksta s maksimalnim brojem bodova 40) ostvarili izvorni govornici (I) i dvije skupine učenika čiji je materinski jezik srpski (NS – niža srednja, VS – viša srednja; Miličević, 2007); lijevo se vidi osnovni prikaz, a desno prikaz s usjecima. Visina pravokutnika pokazuje količinu varijabilnosti u podacima – očigledno je naprimjer da su rezultati izvornih govornika koncentriraniji i manje varijabilni od rezultata učenika. Također se vidi da su u pravokutniku rezultati učeničkih skupina ravnomjerno raspoređeni oko medijana, dok su rezultati izvornih govornika u većoj mjeri grupirani, odnosno manje variraju ispod njega. Na temelju toga što nema nikakva preklapanja, ne samo između usjeka nego ni između pravokutnika različitih skupina ispitanika, može se pretpostaviti da se njihovi rezultati značajno razlikuju (što ipak svakako treba provjeriti i odgovarajućim testom). Vidimo i dvije netipične vrijednosti, označene kružićima, koje se javljaju u skupini izvornih govornika.



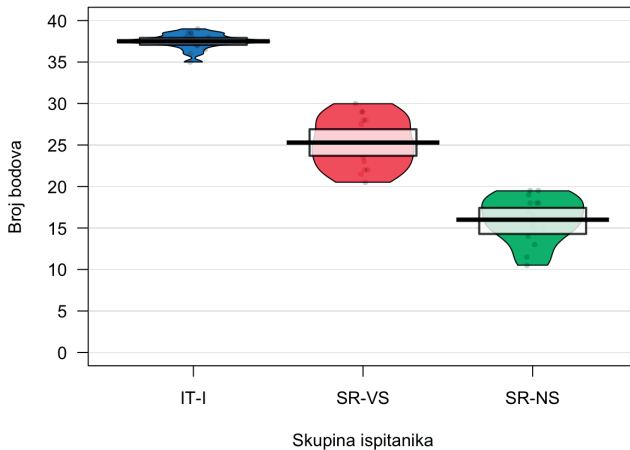
Slika 31. Dijagram s pravokutnikom, bez usjeka i s usjecima (prema Miličević, 2007)

Dijagram s pravokutnikom predstavlja vrstu grafikona na kojoj znanstveni časopisi danas često inzistiraju zato što daje detaljniji prikaz podataka od grafikona koji prikazuju samo mjeru centralne tendencije. Dijagram s pravokutnikom može se i dodatno obogatiti prikazom individualnih podataka, kao i dodavanjem – ako su ispunjeni uvjeti za to – oznaka aritmetičke sredine i granica pouzdanosti (v. Larson-Hall, 2017; Larson-Hall i Plonsky, 2015). To ilustriramo grafikonom na slici 32, generiranim na stranici <https://langtest.jp/shiny/bs/> (posljednji pristup 26. 4. 2024.) iz podatka o bodovima skupine IT-I (iz Miličević, 2007). Crveni kvadratić označava aritmetičku sredinu, crvene strelice standardnu devijaciju, a plave točkice pojedinačne rezultate.



Slika 32. Dijagram s pravokutnikom dopunjen prikazom pojedinačnih podataka (prema Miličević, 2007)

Posljednja vrsta grafikona koju ćemo spomenuti u ovome potpoglavlju naziva se **piratskim grafikonom** (engl. *pirate plot*). Primjer je dan na slici 33, uz uporabu istih podataka kao u dijagramu s pravokutnikom na slici 31.



Slika 33. Piratski grafikon (prema Miličević, 2007)

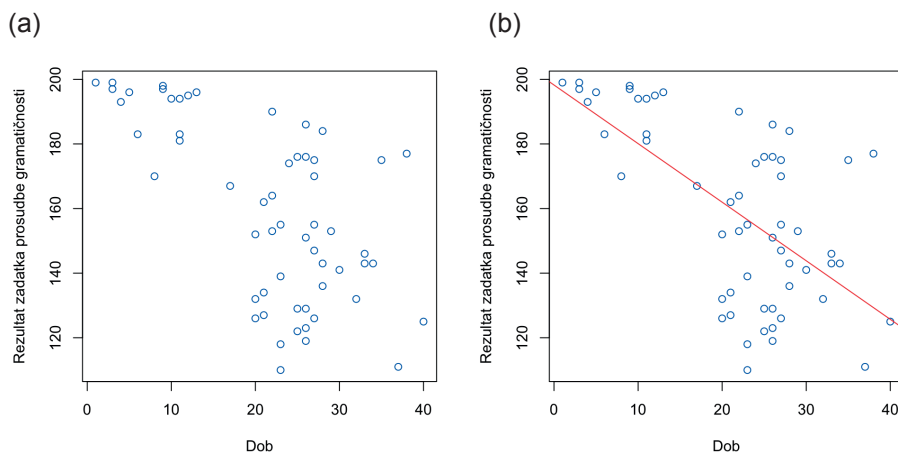
Na tome prikazu možemo vidjeti individualne podatke (sive točkice), kao i njihovu distribuciju, koja je prikazana krivuljom gustoće vjerojatnosti (nacrtanom simetrično, s obje strane). Dodatno je prikazana aritmetička sredina (vodoravna crta) s granicama pouzdanosti (osjenčano polje oko crte). Tu je vrstu grafikona osmislio Phillips (2017) upotrebljavajući elemente iz nekoliko drugih vrsta grafikona (krivulja gustoće upotrebljava se primjerice i kod tzv. „violinskih“ grafikona, engl. *violin/beans plots*). Taj se sveobuhvatni grafički prikaz definira kao grafikon koji prikazuje pojedinačne podatke, deskriptivne podatke i inferencijalne podatke (engl. *RDI – Raw data, Descriptive and Inferential statistics*).⁸⁶

3.2.3 Prikaz povezanosti numeričkih varijabli

Još jedna vrsta grafikona koja se temelji na sveobuhvatnome prikazu podataka i uključuje sve pojedinačne vrijednosti jest **dijagram raspršenja** ili **rasipanja** (engl. *scatter plot*). Tu vrstu grafikona često hvale kao vrlo informativnu (v. npr. Tufte, 2001, prema Larson-Hall, 2017), a upotrebljava se za prikazivanje dviju skupina numeričkih podataka čija se (ne)povezanost istražuje. Na dvjema osima grafikona prikazane su dvije numeričke varijable, a svaka pojedinačna točka (ili drugi simbol) predstavlja jedan pojedinačni podatak čije vrijednosti za dvije varijable možemo očitati s odgovarajućih osi. Naprimjer

⁸⁶ Više informacija može se pročitati u opisu paketa koji omogućava njegovu izradu u programu R, <https://www.rdocumentation.org/packages/yarr/versions/0.1.5/topics/pirateplot> (posljednji pristup 26. 4. 2024.).

slika 34 prikazuje podatke iz DeKeyser (2000), dostupne u Dodatku A toga rada, gdje svaka točka predstavlja jednoga učenika engleskoga kao drugoga jezika (materinski je jezik svih učenika mađarski, a u pitanju su osobe koje su u trenutku provedbe istraživanja živjele u Sjedinjenim Američkim Državama). Za svakoga učenika možemo vidjeti u kojoj je dobi počeo usvajati engleski jezik i koji je rezultat ostvario u zadatku prosudbe gramatičnosti rečenica, koji je sadržavao 200 rečenica te dopuštao maksimalni rezultat od isto toliko bodova.



Slika 34. Dijagram raspršenja (prema DeKeyser, 2000)

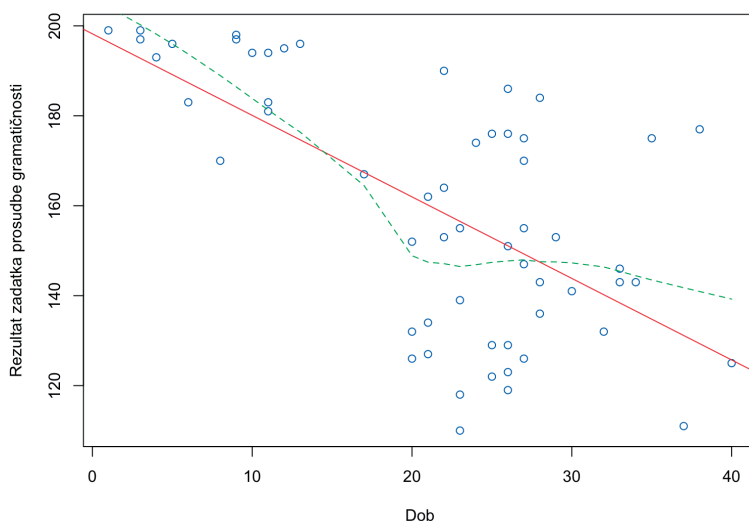
Posebno je važno uočiti na takvim grafikonima odnos između točaka. Ako su one grupirane tako da približno prate crtu koja se dijagonalno pruža kroz grafikon, možemo zaključiti da su dvije varijable međusobno usko povezane. Što su podaci raspršeniji, to je povezanost slabija. Također možemo pretpostaviti izostanak povezanosti ako se crta umjesto dijagonalno proteže vodoravno ili okomito jer takav položaj pokazuje da jedna od varijabli uvijek ima sličnu vrijednost. S druge strane, za jačinu povezanosti nije bitan smjer u kojemu se crta pruža (iz lijevoga gornjeg kuta naniže, kao na slici 34, što nam pokazuje da boljemu rezultatu odgovara niža dob na početku usvajanja, ili iz lijevoga donjega kuta naviše, što bi bila suprotna tendencija). O tim ćemo temama detaljnije govoriti u potpoglavlju 5.4, posvećenome korelaciji.

Kao što pokazuje slika 34b, na dijagram je zapravo moguće dodati crtu koja će nam olakšati tumačenje. Takva se crta naziva **regresijski pravac** (engl. *regression line*). Iako o regresiji u ovoj knjizi ne govorimo detaljno, uvest ćemo osnovne pojmove u potpoglavlju 5.5, a korisno je već ovdje uvesti

formulu prema kojoj se pravac određuje, koju vidimo u (7). U toj formuli X i Y predstavljaju vrijednosti točaka koje leže na pravcu (odnosno vrijednosti dvije varijable). Ako nam je poznata samo vrijednost X , iz nje možemo odrediti vrijednost Y upotrebljavajući **nagib pravca** (engl. *slope*) (b) i njegov **intercept** ili **odsječak** (engl. *intercept*), odnosno vrijednost osi y u kojoj je pravac siječe (a), gdje a i b predstavljaju koeficijente koji su izračunati na temelju parova vrijednosti koje su nam poznate. Budući da na pravcu ne leže sve vrijednosti, on je zapravo pokušaj da se najbolje moguće opišu (tj. modeliraju) podaci, slično tome kako se prosjek upotrebljava kao model značajki skupine podataka (v. potpoglavlje 3.1.2, kao i 1.3, gdje uvodimo pojam modeliranja).

$$(7) Y = bX + a$$

Larson-Hall (2010) i Larson-Hall i Herrington (2010) preporučuju da se pored regresijskoga pravca na dijagram raspršenja doda i **Loessova krivulja** (engl. *Loess curve*), koja pokazuje je li odnos između varijabli linearan (što je uvjet za test linearne korelacije), a može nam u nekim slučajevima pomoći i u otkrivanju netipičnih vrijednosti. Način crtanja takve krivulje podrazumijeva veći broj krivulja definiranih lokalno, na manjim skupinama podataka. Na slici 35 prikazujemo takvu krivulju ucrtanu (isprekidano) na podatke iz DeKeyser (2000), gdje vidimo da odnos dviju varijabli nije u potpunosti linearan.



Slika 35. Dijagram raspršenja s dodanom Loessovom krivuljom (prema DeKeyser, 2000)

Unutar dijagrama raspršenja podaci se mogu i grupirati u više skupina te se za svaku posebno može prikazati regresijski pravac (v. Larson-Hall, 2010, str. 157). Također, više dijagrama raspršenja može se kombinirati u korelacijsku matricu ako želimo prikazati odnose između različitih parova varijabli u istraživanjima u kojima imamo više od dviju varijabli.

3.2.4 Drugi prikazi podataka i opće napomene

Naš pregled iz ovoga potpoglavlja ne predstavlja, naravno, ni izbliza iscrpan popis mogućnosti za prikaz podataka. Takav bi popis teško bilo moguće napraviti, te nam je cilj ovdje bio uputiti čitatelje na često korištene vrste grafikona i svratiti pozornost na njihove prednosti i mane. Dodatno je preporučljivo informirati se kako o drugim vrstama grafikona tako i o podvrstama i dodatnim mogućnostima za grafikone koji su ovdje opisani. Pored Larson-Hall (2017), Gries (2013a) i Johnson (2013), korisne se informacije mogu pronaći i u Gries (2013b), Levshina (2015), Winter (2020) i Garcia (2021).

Ono što u zaključku potpoglavlja treba istaknuti kao bitno za sve vrste grafikona jest da ih treba jasno obilježiti, navodeći nazive i legende iz kojih se vidi što je prikazano na grafikonu kao cjelini, na svakoj osi i putem obilježja poput različitih boja ili različitih simbola. Također uvijek treba paziti da se podaci prikažu vjerodostojno te da se ciljni korisnici grafikona ne navode na neopravdane zaključke. Naprimjer pozornost je potrebna kod određivanja omjera osi x i y (tako da grafikon ne bude ni suviše visok i uzak, ni suviše nizak i spljošten), kao i kod izbora raspona vrijednosti koje će biti prikazane na osi y (gdje bi najčešće trebalo predstaviti cijeli raspon mogućih vrijednosti, a ne samo jedan njihov dio; naprimjer 0–40 u primjerima s brojem bodova iz Miličević, 2007). Ako je koji od tih zahtjeva iz praktičnih razloga teško ispuniti (naprimjer prikazivanje cijeloga raspona mogućih vrijednosti zauzelo bi previše mjesta), to treba jasno napomenuti. U suprotnome bi mogao nastati **obmanjujući grafikon** (engl. *misleading graph*).

Konačno, završavamo ovo poglavlje napomenom preuzetom iz Lindstromberg (2016a, str. 5) da prilikom izvještavanja o istraživanju treba pokušati dati čitateljima sve informacije koje im mogu biti potrebne za razumijevanje istraživanja, ali se istodobno i suzdržati od davanja previše informacija, koje bi lako mogle ugroziti čitljivost.

4. Statističko zaključivanje

Prikupljanje podataka u eksperimentalnim istraživanjima usvajanja drugoga jezika podrazumijeva utvrđivanje vrijednosti relevantnih zavisnih varijabli za odabrani uzorak. Kako je objašnjeno u drugome poglavlju (v. i Kraš i Miličević, 2015, str. 18–24), vrijednosti nezavisnih varijabli utvrđuju se prije prikupljanja podataka, čime se definiraju uzorci na kojima će prikupljanje biti provedeno. Da bi se od prikupljenih podataka došlo do zaključaka o postavljenoj istraživačkoj hipotezi, deskriptivne mjere prikazane u trećemu poglavlju treba dopuniti ispitivanjem mogućnosti uopćavanja rezultata na druge uzorke i cijelu populaciju, u čemu ključnu ulogu imaju inferencijalni statistički testovi.

Prije nego što prijeđemo na prikaz najčešće korištenih testova, čemu je posvećeno peto poglavlje, u ovome poglavlju iznosimo osnove statističkoga zaključivanja i statističkoga testiranja hipoteza – smatramo ih posebno važnima jer ni najjednostavnije testove nije moguće razumjeti bez dobrog razumijevanja tih osnova. Dodatno ovdje objašnjavamo kriterije koje je potrebno primijeniti prilikom izbora statističkoga testa te podsjećamo da se izbor ne vrši u fazi analize podataka, već o njemu treba razmišljati još prilikom planiranja istraživanja. Ističemo i da se pojedini kriteriji za izbor konačnoga testa i sami moraju statistički testirati. Naposljetku, osvrnut ćemo se i na statističko planiranje potrebne veličine uzorka, kao i na nekoliko pratećih tema.

4.1 Osnove statističkoga zaključivanja

Kako je već navedeno u potpoglavlju 1.2, u uopćavanju rezultata istraživanja ključnu ulogu ima inferencijalna statistika, koja se naziva i statistika zaključivanja. Pojam **statističko zaključivanje** (engl. *statistical inference*) odnosi se na proces kojim se utvrđuje stupanj u kojemu deskriptivne mjere izračunate za uzorak dobro predstavljaju slične uzorke i populaciju (Brown, 1991, str. 575). Drugim riječima, svrha su ovoga dijela statističke analize predviđanja vezana uz podatke koji nisu već zabilježeni i stoga se ne mogu opisati. Statističkom terminologijom rečeno, najčešće se radi o procjeni parametara na temelju statistika, odnosno o procjeni populacijskih vrijednosti iz vrijednosti dobivenih u uzorku. Iako inferencijalna statistika u kronološkome smislu predstavlja jedan od posljednjih koraka u analizi podataka, znanja vezana uz nju potrebno je imati već u fazi planiranja i osmišljavanja istraživanja.

Statistika zaključivanja počiva na elementima teorije vjerojatnosti.⁸⁷ U knjizi govorimo gotovo isključivo o frekvencijskome poimanju vjerojatnosti, odnosno o **frekvencijskome statističkom zaključivanju** (engl. *frequentist inference*), koje nije jedina, ali je trenutačno dominantna paradigma u istraživanjima usvajanja drugoga jezika. Ono podrazumijeva izvođenje zaključaka o populaciji na temelju podataka prikupljenih na uzorcima bez razmatranja čimbenika izvan provedenoga eksperimenta. Svaki se pojedinačni eksperiment promatra kao slučajan događaj i smatra se jednim u nizu ponavljanja istoga eksperimenta (v. npr. Cumming, 2014). Pored frekvencijskoga važno je i **Bayesovo statističko zaključivanje** (engl. *Bayesian inference*), kod kojega se u obzir uzimaju i čimbenici izvan eksperimenta, ponajprije prethodna i naknadna vjerojatnost različitih ishoda (v. npr. Mackey i Ross, 2015). Na taj ćemo se pristup kratko osvrnuti u petome poglavlju (točnije, u potpoglavlju 5.6), ali se njime nećemo detaljnije baviti.

Frekvencijsko statističko zaključivanje oslanja se na **testiranje statističke značajnosti** (engl. *significance testing*). To se testiranje provodi putem inferencijalnih statističkih testova, čiji je ukupni broj velik, a čija mogućnost i primjerenost uporabe ovise o vrsti prikupljenih podataka, ali i o hipotezama koje se žele testirati te o nacrtu istraživanja. Bez obzira na ponekad velike razlike u kontekstima uporabe i načinu izračunavanja, svi testovi počivaju na istim načelima. U okviru ovoga potpoglavlja objašnjavamo upravo pojmove koji su potrebni za razumijevanje svih testova koji će poslije biti prikazani.

4.1.1 Testiranje hipoteza

U ranijim je poglavljima već bilo riječi o istraživačkoj hipotezi kao iznimno važnome elementu kvantitativnih istraživanja (v. posebice potpoglavlje 2.1.1). Istraživačke su hipoteze po svojoj prirodi opisne, tj. navode konkretne čimbenike koji se istražuju. Iako je krajnji cilj istraživača ispitati neku takvu hipotezu (npr. „Dob na početku usvajanja drugoga jezika utječe na razinu usvojenosti toga jezika“ ili „Usvojenost članova ovisi o razini općega znanja engleskoga jezika“), iz perspektive kvantitativne analize podataka važnije su **statističke hipoteze** (engl. *statistical hypotheses*). Statističke hipoteze odražavaju način operacionalizacije varijabli u danome istraživanju i formulirane su s obzirom na statističke vrijednosti. Naprimjer opća pretpostavka o utjecaju razine općega znanja engleskoga jezika na usvojenost članova neće imati istu statističku postavku ako je razina znanja iskazana kao kategorijska varijabla (npr. uključeni su ispitanici razine znanja B2 i C2) ili kao kontinuirana varijabla (npr. putem broja

⁸⁷ Osnove teorije vjerojatnosti na pristupačan način opisuju Upton i Cook (1996).

bodova ostvarenih na testu; usp. Gries, 2013a, str. 317). U prvome ćemo se slučaju baviti razlikama između skupina ispitanika u pogledu uporabe članova u nekome eksperimentalnom zadatku i statistička bi hipoteza mogla glasiti „aritmetičke se sredine dvaju uzoraka razlikuju“ (gdje bi se aritmetičke sredine dobile iz rezultata eksperimenta i mogle bi odražavati naprimjer broj točnih uporaba članova). U drugome se slučaju bavimo povezanošću između broja bodova i usvojenosti članova te bi statistička hipoteza (u pojednostavljenoj obliku) mogla glasiti „vrijednosti varijable X koreliraju s vrijednostima varijable Y “. Drugim riječima, statističke su hipoteze formalne tvrdnje o nekoj pojavi, odnosno iskazi o vrijednosti određenih parametara u populaciji (Tošković, 2020, str. 306–307). Kako navode Milas (2009) i Tošković (2020), statističke hipoteze mogu biti izražene i kao matematičke funkcije. One predstavljaju temelj statističke analize podataka jer se u analizi provjerava njihova istinitost.

Moguće je izdvojiti dvije glavne vrste statističkih hipoteza. **Nulta hipoteza** (engl. *null hypothesis*), često označena kao H_0 , neutralna je pretpostavka, tj. pretpostavka da između istraživanih varijabli ili skupina ne postoji veza ili razlika – točan način na koji je formulirana ovisi o istraživanome kontekstu i statističkome testu koji se primjenjuje. Nasuprot tome, **alternativna hipoteza** (engl. *alternative hypothesis*), označena kao H_a ili H_1 , pretpostavlja da između istraživanih varijabli ili skupina postoji veza ili razlika.⁸⁸ Iako se u istraživanjima najčešće polazi od pretpostavke o postojanju određene veze ili razlike, za statističku analizu posebno je važna nulta hipoteza jer se statistički testovi neposredno temelje na njoj – čitavo se testiranje statističke značajnosti još naziva i **testiranjem nulte hipoteze** (engl. *null hypothesis statistical testing*).⁸⁹ Točnije, prilikom statističkoga testiranja procjenjuje se koliko je vjerojatno da se zabilježeni rezultati (ili rezultati ekstremniji od njih) dobiju u slučaju kad je nulta hipoteza točna – što je vjerojatnost za to manja, to je veća mogućnost da je dobiveni rezultat povezan s ispitivanim varijablama (McDonough i Trofimovich, 2009, str. 156; Todorović, 2008, str. 215). Drugim riječima, istraživačka se hipoteza može testirati samo posredno, putem testiranja nulte hipoteze.⁹⁰

⁸⁸ Alternativna se hipoteza ponekad naziva i „istraživačkom“, ali zapravo je točnije reći da je ona statistička verzija istraživačke hipoteze (koja je opisna, a ne formalna; v. potpoglavlje 2.1.1).

⁸⁹ Moguće je, iako rijetko, da se istraživačka hipoteza poklapa s nultom hipotezom. Heil i López (2020) naprimjer postavljaju kao jednu od hipoteza predviđanje da će izvorni govornici engleskoga jezika jednako ocjenjivati strukture s objektivnom kontrolom i infinitive s *for* (v. i potpoglavlje 2.1.1).

⁹⁰ Takvo je testiranje dio širega pristupa znanosti utemeljenoga na načelu opovrgavanja, koji je utemeljio Karl Popper, ističući da znanstvene teorije i hipoteze mora prije svega biti moguće opovrgnuti te da se one mogu (tentativno) smatrati ispravnima ako nisu opovrgnute. O tome govorimo i u potpoglavlju 2.1.1.

Tavakoli (2013, str. 266) taj pristup sažima tako što definira testiranje hipoteza kao znanstveni pristup testiranju naših uvjerenja o nekoj pojavi, gdje odluku donosimo statističkom procjenom sličnom onoj koju činimo u svakodnevnim situacijama kad moramo napraviti izbor, iako nam okolnosti ne dopuštaju da zasigurno znamo je li taj izbor najbolji (naprimjer kad po oblačnome danu odlučujemo hoćemo li ponijeti kišobran).

Alternativne hipoteze (pa time posredno i istraživačke) nadalje se mogu podijeliti na **jednosmjerne hipoteze** (engl. *directional/one-way/one-tailed hypotheses*), kod kojih se pretpostavlja smjer veze ili razlike, i **dvosmjerne hipoteze** (engl. *non-directional/two-way/two-tailed hypotheses*), kod kojih se pretpostavlja samo postojanje veze ili razlike, ali ne i njezin smjer. Jensen i sur. (2020) naprimjer jednosmjerno predviđaju da je u usvajanju engleskoga kao drugoga jezika u kombinaciji s norveškim kao prvim jezikom uočavanje negramatičnoga reda riječi (s predikatom ispred subjekta) manje problematično od uočavanja negramatičnoga morfološkog slaganja (u broju, između subjekta i predikata). S druge strane, hipoteza koju postavljaju Heil i López (2020), spomenuta u potpoglavlju 2.1.1, koja predviđa da će izvorni govornici španjolskoga koji imaju visoku razinu općega znanja engleskoga kao drugoga jezika drukčije tretirati strukture s objektnom kontrolom i infinitive s *for*, dvosmjerna je jer ne izražava očekivani smjer razlike, odnosno ne precizira koje će strukture dobiti više ocjene u zadatku prosudbe prihvatljivosti rečenica. Izbor između tih dviju vrsta hipoteza ovisit će o prethodnome poznavanju teme, nalazima prethodnih istraživanja i nizu povezanih kriterija. Matematički gledano, jednosmjerne hipoteze sužavaju prostor variranja (što ćemo detaljnije objasniti u potpoglavljima 4.2.2 i 5.2.1), te je kod njih dovoljan manji utjecaj nezavisne varijable kako bi se on smatrao značajnim. Međutim uz jednosmjernu hipotezu postaje vrlo teško protumačiti rezultate ako se dobije efekt suprotnoga smjera od predviđenoga (primjerice da učenici u istraživanju Jensen i sur., 2020, imaju više problema s redom riječi nego s morfološkim slaganjem). To je jedan od razloga zbog kojih statističari preporučuju oslanjanje na dvosmjerne statističke hipoteze čak i u onim slučajevima kad je istraživačka hipoteza jednosmjerna (v. npr. Field i sur., 2012, str. 55–56).

Treba podsjetiti i da tvrdnje koje se opisno mogu označiti kao predviđanja (v. potpoglavlje 2.1.1) za statističku analizu predstavljaju hipoteze, odnosno jednako se testiraju. Tako primjerice Jensen i sur. (2020) govore o konkretnijim predviđanjima koja proizlaze iz šire teorijske „hipoteze uskoga grla“ (engl. *Bottleneck Hypothesis*). Hipoteza uskoga grla vrlo je široka pretpostavka prema kojoj je najizazovniji aspekt usvajanja drugoga jezika – njegovo „usko

grlo“ – funkcijska morfologija, što se u pojedinačnim istraživanjima ispituje na konkretnim pojavama iz različitih jezika. U statističkoj analizi uža i konkretnija predviđanja odgovaraju statističkim hipotezama koje se testiraju.

Konačno, treba odmah istaknuti da prilikom zaključivanja o nultoj hipotezi uvijek postoji vjerojatnost pojave pogreške, odnosno izvođenja pogrešnoga zaključka. Ta je situacija slična mnogima iz svakodnevnoga života – naprimjer kad odlučujemo hoćemo li prilikom izlaska iz kuće ponijeti kišobran, razmatramo činjenice koje su nam dostupne, poput vremenske prognoze, ali znamo da postoji mogućnost da se odluka pokaže pogrešnom. Moguće su dvije vrste pogrešnih zaključaka: **pogreška tipa I** (engl. *type I error*) podrazumijeva da je nulta hipoteza odbačena u slučaju u kojemu je zapravo točna (npr. nismo ponijeli kišobran, a kiša je pala), dok **pogreška tipa II** (*type II error*) znači da je nulta hipoteza prihvaćena, iako zapravo nije točna (npr. ponijeli smo kišobran, a kiša nije pala). Mogućnost pojavljivanja tih pogrešaka ne može se eliminirati, ali se vjerojatnost njihova javljanja može umanjiti odgovarajućim nacrtom istraživanja, kao i primjenom odgovarajućih statističkih testova. U daljnjem ćemo se tekstu više puta vraćati na te pogreške te pobliže definirati i njih same i njihovu važnost za statističku analizu.

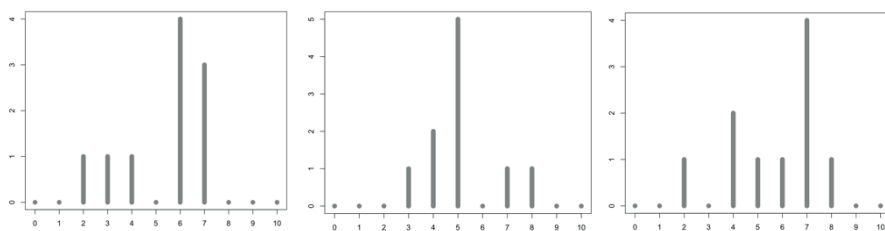
4.1.2 Statistička značajnost i statistička snaga

Rezultati statističkih testova imaju oblik **iskaza o vjerojatnosti** (engl. *probability statement*). Vrijednost ***p*** koja se dobiva kao jedan od glavnih pokazatelja u statističkim testovima pokazuje kolika je vjerojatnost dobivanja zabilježenoga rezultata (ili drugih rezultata koji su još ekstremniji) u slučaju kad je nulta hipoteza točna. To znači da se ni nulta hipoteza ni istraživačka hipoteza ne mogu statistički dokazati, nego se samo mogu dobiti naznake o njihovoj vjerojatnosti (v. npr. Larson-Hall, 2012a, str. 246–247).

Općenito govoreći, **vjerojatnost** (engl. *probability*) je šansa pojavljivanja određenoga događaja, odnosno određenoga ishoda. Ona se može izraziti numerički i uvijek ima vrijednost između 0 i 1, odnosno između 0 % i 100 %. Ako uzmemo kao primjer bacanje novčića, koje se može smatrati jednostavnim **statističkim eksperimentom** ili **eksperimentalnom situacijom** (engl. *statistical experiment* ili *trial*; „eksperimentalna situacija“ označava pojedinačne događaje, dok se „eksperiment“ može odnositi i na niz ponovljenih situacija), prilikom bacanja postoje dva moguća ishoda – pismo i glava – i svaki je od njih podjednako vjerojatan, te je vjerojatnost svakoga od njih jednaka 0,5 (odnosno 50 % ili 1/2 jer decimalni brojevi, postoci i razlomci predstavljaju prikaze iste veličine). Kad su ishodi podjednako vjerojatni, vjerojatnost svakoga

pojedinačnog ishoda izračunavamo uzimajući u obzir ukupan broj mogućih ishoda (u slučaju bacanja novčića, dva) i broj načina na koje se može dobiti određeni ishod (u slučaju novčića imamo jedan način dobivanja pisma i jedan način dobivanja glave).

Slično je načelo u pozadini testiranja nulte hipoteze, iako je ta situacija nužno složenija od bacanja novčića. Možemo pokušati približiti te situacije tako što ćemo opisati statistički eksperiment u kojemu bacamo novčić više puta, tj. više puta ponavljamo isti jednostavni eksperiment. Zamislimo primjerice deset bacanja. U svakome bacanju vjerojatnost za svaku od dviju mogućnosti iznosi 50 %. Ako bismo u deset bacanja brojili koliko smo puta dobili pismo, moguće je da rezultat bude pet (što bi idealno odgovaralo početnoj vjerojatnosti od 50 %), ali je moguće i da bude drukčiji – bilo koji broj između nula i deset mogući je rezultat. Ako bismo pak dodali još jedan stupanj složenosti i deset puta ponovili po deset bacanja, bilježeći koliko se puta pismo javilo u svakoj od serija od deset bacanja, mogli bismo dobiti da su najčešće serije u kojima se pismo javilo pet puta, ali i taj bi rezultat mogao biti i drukčiji. To pokazuje slika 36, na kojoj vidimo rezultate triju eksperimenata s deset nizova od po deset bacanja novčića (računalno simuliranih; os x na slici prikazuje broj pisama dobivenih u jednoj seriji od deset bacanja, a os y broj pojavljivanja svakoga od tih brojeva u nizu od deset serija). Možemo uočiti da je u prvome eksperimentu bilo najviše serija sa šest pisama (ukupno četiri), u drugome najviše serija s pet pisama (ukupno pet), a u trećemu najviše serija sa sedam pisama (ukupno četiri). Eksperimenti su se razlikovali i po broju serija u kojima se pismo javilo neki drugi broj puta.⁹¹



Slika 36. Primjer rezultata višestrukih bacanja novčića

Drugim riječima, i kad je ishod potpuno slučajnan, kao kod bacanja „fer“ novčića, ako eksperiment ponovimo više puta, nužno dolazi do određene varijabilnosti u ishodima. Takvu varijabilnost možemo nazvati prirodnom jer

⁹¹ Ti nam jednostavni podaci pokazuju i važnost veličine uzorka za statističku analizu – razlike između eksperimenata bile bi zasigurno manje da je bilo više serija bacanja novčića.

do nje ne dovodi utjecaj ikakva čimbenika, već se ona jednostavno događa. Zbog postojanja takve varijabilnosti ne postoji statistička formula koja može potpuno pouzdano predvidjeti ishod slučajnoga statističkog eksperimenta. Međutim, ako znamo koji su sve ishodi mogući i na koliko se načina svaki od njih može ostvariti, možemo izračunati koliko je koji ishod vjerojatan i koji su ishodi vjerojatniji od drugih. Naprimjer na slici 36 možemo vidjeti da se nijednom nije javila serija s deset pisama. To je očekivano jer deset pisama možemo dobiti na samo jedan način – sva bacanja moraju dati pismo. S druge strane, rezultat od pet pisama možemo dobiti tako što će se u prvih pet bacanja javiti pismo, a u drugih pet glava, tako što ćemo u svakome drugom bacanju dobiti pismo, kao i na druge načine. Ne iznenađuje stoga što se takav ishod češće javlja.⁹²

Prilikom izrade inferencijalnih statističkih testova primjenjuje se isto načelo, s time što se vjerojatnost izračunava za pojavljivanje različitih vrijednosti rezultata statističkoga testa. Točnije, statistički testovi daju kao osnovni rezultat neki testni statistik izračunat iz podataka i počivaju na provjeri toga kolika je vjerojatnost „slučajnoga“ dobivanja toga statistika, odnosno njegova dobivanja u slučaju kad je nulta hipoteza (koja podrazumijeva da ne postoje dodatni čimbenici koji utječu na rezultat) točna. Ako se pokaže da je dobiveni statistik vrlo malo vjerojatan pod nultom hipotezom (što nam pokazuje vrijednost p), pretpostavit ćemo da je za nju odgovoran utjecaj nezavisne varijable koju ispituje.

To možemo ilustrirati na jednostavnome povijesnom primjeru. Jedna od najpoznatijih priča iz područja statistike jest ona o eksperimentu s damom koja je kušala čaj (engl. *the lady tasting tea experiment*; v. npr. Richardson, 2021). Poznati britanski statističar Ronald Fisher našao se jednoga dana 1920-ih godina u vrijeme popodnevnoga čaja u društvu psihologinje Muriel Bristol, koja je tvrdila da može prepoznati je li u šalicu čaja najprije uliven čaj ili mlijeko. Fisher je to poželio eksperimentalno testirati, pa je pozvao dotičnu damu da kuša osam nasumce poredanih šalica među kojima je u četiri najprije uliven čaj, a u četiri mlijeko. Zadatak je bio ispravno prepoznati sve četiri šalice u koje je najprije uliveno mlijeko (što automatski znači ispravno prepoznavanje

⁹² Bacanje novčića zanimljiv je primjer, na kojemu se mogu ilustrirati i različita shvaćanja vjerojatnosti. Dok je prema klasičnome shvaćanju vjerojatnost dvaju ishoda 50 % zato što je novčić simetričan i objektivna, odnosno sam po sebi ima značajke koje određuju jednaku vjerojatnost dobivanja pisma i glave, frekvencijsko je viđenje da vrijednost od 50 % potječe otuda što se u vrlo velikome broju ponovljenih bacanja dobiva distribucija koja (približno) iznosi 50 % pismo – 50 % glava. Frekvencijsko se statističko zaključivanje naziva tako upravo zbog toga što počiva na viđenju vjerojatnosti kao relativne frekvencije ishoda u velikome broju ponavljanja istoga eksperimenta.

svih šalica). Fisher je primjenjujući načela kombinatorike izračunao da je vjerojatnost slučajnoga pogađanja točnoga odgovora 1/70, odnosno 1,4 %, te je smatrao da bi uz tako malu šansu za slučajno pogađanje točan izbor šalica doista potvrdio iznimnu sposobnost prepoznavanja redosljeda ulijevanja čaja i mlijeka. Iako, nažalost, ne znamo pouzdano kakav je bio konačni ishod toga eksperimenta (postoje nepotvrđene informacije o točnome izboru tri šalice od četiri), način na koji ga je Fisher osmislio pokazuje nam na koji način i danas u istraživanjima izvodimo zaključke o istraživačkim hipotezama i zbog čega uvijek moramo najprije pretpostaviti nultu hipotezu.

Polazeći od toga da statistički testovi pokazuju vjerojatnost dobivanja danoga rezultata u slučaju da nulta hipoteza vrijedi (odnosno, u prisutnosti prirodnoga variranja kao jedinoga izvora variranja), inferencijalna statistika kao važnu mjeru kod različitih statističkih testova daje podatak o **statističkoj značajnosti** (engl. *statistical significance*) rezultata. Statistička značajnost tijesno je povezana s vrijednošću p , koja pokazuje upravo vjerojatnost rezultata kad vrijedi nulta hipoteza, i koja se dobiva kao rezultat svih testova. Ta se vrijednost zatim uspoređuje s unaprijed definiranom **razinom statističke značajnosti** (engl. *significance level*), koja se označava kao **alfa (α)**, i koja se odnosi na razinu rizika pojave pogreške tipa I koju smatramo prihvatljivom.⁹³ U društvenim znanostima, pa i u području usvajanja drugoga jezika, prihvatljivim se načelno smatra rizik od 5 %, tj. rezultat se smatra statistički značajnim ako se dobije $p < 0,05$. Razina značajnosti koja se u nekome području smatra prihvatljivom predstavlja konvenciju, a ne matematički utvrđenu vrijednost. Naprimjer pojedini autori smatraju da je u područjima poput primijenjene lingvistike adekvatnije podići prag na $\alpha = 0,1$ (v. Larson-Hall, 2010, str. 102; Mackey i Gass, 2005, str. 268), pa se uz $\alpha = 0,05$ za rezultate u rasponu 0,06–0,09 često kaže da se približavaju statističkoj značajnosti (v. npr. Gass, 2009). Nasuprot tome razina statističke značajnosti u disciplinama poput medicine obično se postavlja na $\alpha = 0,01$.⁹⁴ To je jedan od razloga zbog kojih je u radovima dobro eksplicitno navesti odabranu razinu značajnosti (v. npr. Larson-Hall i Plonsky, 2015, str. 130).

Važno je naglasiti da vrijednost p ne izražava vjerojatnost da je nulta hipoteza točna (v. npr. Larson-Hall, 2010, str. 97). Odnosno, ona ne pokazuje

⁹³ Petz i sur. (2012, str. 154) smatraju primjerenijim naziv „razina rizika“ jer se radi o veličini rizika pogreške koju prihvaćamo. Međutim u većemu je dijelu literature i dalje uobičajeno govoriti o statističkoj značajnosti.

⁹⁴ Takve razlike jasno upućuju na to da je u pitanju konvencija utemeljena ne na matematici već na procjeni prihvatljivoga rizika, odnosno procjeni težine posljedica u slučaju izvođenja pogrešnoga zaključka.

vjerojatnost točnosti nulte hipoteze za određene podatke već upravo suprotno: vjerojatnost određenih podataka za slučaj važenja nulte hipoteze. Također treba napomenuti da vrijednost p koja se dobiva kao rezultat statističkoga testa nije izravna mjera vjerojatnosti donošenja pogrešne odluke (tj. odbacivanja nulte hipoteze kad je ona točna). Kako navode Tenjović i Smederevac (2011, str. 319–320), α je apriorna vjerojatnost pogrešnoga odbacivanja nulte hipoteze, koja se određuje prije nego što se podaci prikupe i koja služi kao kriterij za odluku o odbacivanju nulte hipoteze, dok je p posteriorna vjerojatnost koja se dobiva iz podataka, tj. naknadno, i predstavlja vjerojatnost javljanja dobivenoga (ili većega) testnog statistika onda kada je nulta hipoteza točna (v. i Porte, 2002, str. 106).

Ako postoji vjerojatnost niža od 5 % da smo određeni rezultat dobili uz točnost nulte hipoteze, na temelju toga podatka posredno možemo pretpostaviti da su do rezultata doveli čimbenici koje smo ispitivali (npr. razina općega znanja drugoga jezika ili materinski jezik ispitanika; usp. i Fisherov primjer s čajem i mlijekom). Kako navodi Todorović (2008, str. 212), alternativna je hipoteza alternativa koja se prihvaća samo ako se testom utvrdi da se nulta hipoteza može odbaciti. U tome postupku ne možemo međutim zaključiti da postoji vjerojatnost od 95 % (točnije, $1-p$) da je alternativna hipoteza točna, budući da $1-p$ predstavlja vjerojatnost dobivanja manje ekstremnih rezultata od onih dobivenih u slučaju kad nulta hipoteza jest točna. Drugim riječima, odbacivanje nulte hipoteze ne znači nužno dokazivanje alternativne hipoteze, a prihvaćanje alternativne hipoteze uz to predstavlja tek statistički korak, dok se povezivanje s istraživačkom hipotezom odvija na konceptualnoj razini.

Posebno treba još jednom naglasiti da uz svaku razinu statističke značajnosti ostaje određeni rizik od pogrešnoga zaključivanja. Kao što smo već spomenule u prethodnome potpoglavlju, u statistici se izdvajaju dvije vrste pogrešaka, pogreška tipa I i pogreška tipa II. Detaljnije opisujemo te pogreške u tablici 10, gdje se vidi da je vjerojatnost pogreške tipa I vezana uz razinu statističke značajnosti (α). Vjerojatnost pogreške tipa II označava se grčkim slovom beta (β). Prilikom statističkoga testiranja općenito se traži ravnoteža između vjerojatnosti pojave tih dviju pogrešaka jer niža vjerojatnost jedne dovodi po porasta vjerojatnosti druge (iako taj odnos nije izravno proporcionalan). Pritom se pogreška tipa I najčešće smatra problematičnijom i obično joj se pridaje veća važnost – usp. već navedenih 5% nasuprot uobičajenih maksimalnih 20 % za pogrešku tipa II (v. npr. Petz i sur., 2012, str. 366).⁹⁵

⁹⁵ Slično načelu na kojemu počiva većina pravnih sustava, gdje se kažnjavanje nedužnih smatra

Tablica 10. Pogreške u statističkome zaključivanju

	Odbacujemo nultu hipotezu	Prihvaćamo nultu hipotezu
Nulta hipoteza je točna	Pogreška tipa I (vjerojatnost α)	Nema pogreške (statistička snaga, $1-\beta$)
Nulta hipoteza nije točna	Nema pogreške (vjerojatnost $1-\alpha$)	Pogreška tipa II (vjerojatnost β)

Uz pogrešku tipa II vezuje se i pojam **statističke snage** ili **snage statističkoga testa** (engl. *power of a statistical test*), koja predstavlja vjerojatnost ispravnoga odbacivanja nulte hipoteze, odnosno njezina odbacivanja u slučaju kad doista nije točna. Ta se vjerojatnost dobiva oduzimanjem od 1 vrijednosti β i najčešće je u istraživanjima cilj da ona ne bude manja od 80 % (budući da β uglavnom iznosi 20 %). U usporedbi sa statističkom značajnosti, snazi statističkih testova tradicionalno se posvećivalo relativno malo pozornosti. Međutim ta se situacija danas sve više mijenja te se upravo u područjima poput usvajanja drugoga jezika ističe da je potrebno raditi na povećanju statističke snage, naprimjer povećanjem uzoraka ispitanika (v. npr. Plonsky, 2013). Ta je mjera uz to važna i kod definiranja potrebne veličine uzorka (o čemu govorimo u potpoglavlju 4.1.5).

Na kraju ovoga potpoglavlja vrlo je važno napomenuti da se testiranje utemeljeno na nultoj hipotezi već duže vrijeme u statističkim, pa i širim znanstvenim krugovima, kritizira kao nepouzdana, nedovoljno informativno i arbitrarno; Plonsky (2013, str. 673) te Tenjović i Smederevac (2011, str. 318–319) navode niz radova koji se bave tom temom (v. i Larson-Hall, 2010, str. 96–97). U posljednje se vrijeme sve češće govori i o općoj krizi te postupnoj promjeni paradigme u znanstvenim istraživanjima (Isbell i sur., 2022; Plonsky i sur., 2024; Sønning i Werner, 2021; v. i zaključno poglavlje ove knjige).⁹⁶ Najveći broj kritika odnosi se na preveliko oslanjanje na pojam statističke značajnosti i vrijednost p (v. npr. Lindstromberg, 2016a, 2016b; Plonsky i Gass, 2011, za napomene na tu temu u području usvajanja drugoga jezika). Preveliko oslanjanje na vrijednost p smatra se problematičnim ponajprije zbog arbitramosti u određivanju granice

problematičnijim od propuštanja da se kazne krivi, opće je načelo da je bolje prihvatiti višu razinu rizika za onu pogrešku koja ima manje štetne posljedice, što je najčešće pogreška tipa II (Petz i sur., 2012, str. 155).

⁹⁶ Pojedini časopisi čak eksplicitno zabranjuju taj pristup u radovima (v. <https://www.sciencebasedmedicine.org/psychology-journal-bans-significance-testing/>, posljednji pristup 26. 4. 2024.), ističući kao moguće alternative temeljitiju deskriptivnu analizu ili Bayesovo statističko zaključivanje, kao i povećanje uzoraka.

statističke značajnosti, iako se i samo postojanje takve vrste crno-bijela praga također često ocjenjuje kao problematično, posebice u širem kontekstu tendencija vezanih uz objavljivanje radova. Poznato je naime da znanstveni časopisi prilikom odabira radova za objavljivanje prednost daju onim radovima koji ukazuju na značajan utjecaj nezavisnih varijabli na zavisne, odnosno koji sadrže statistički značajne rezultate, zbog čega se istraživači prilikom analize ponekad i previše koncentriraju na rezultate statističkih testova i čak provode više testova ili uređuju podatke na više načina kako bi dobili statistički značajne rezultate.⁹⁷

S time povezan i nešto teže uočljiv problem jest i to što istraživači ponekad biraju nacрте istraživanja koji su prejednostavni za istraživanje teme kojom se bave (naprimjer uključuju jednu nezavisnu varijablu, iako je pojava jasno višedimenzionalna), ali kojima se lako mogu dobiti statistički značajni rezultati (v. Tenjović i Smederevac, 2011, str. 319). Konačno, još jedan velik problem s pristupom utemeljenim na testiranju nulte hipoteze jest često pogrešno tumačenje dobivenih rezultata i donošenje neutemeljenih zaključaka o „dokazivanju“ hipoteza, što je dijelom rezultat složenosti samoga pristupa, a dijelom proizlazi iz nedostatka temeljitije obuke istraživača (što se, kako smo vidjeli u potpoglavlju 1.5, čini vrlo relevantnim za discipline koje se bave istraživanjima jezika). Svi ti problemi zajedno dovode do toga da spoznaje koje dospiju u znanstvenu i širu javnost nisu nužno stečene posve objektivnim postupcima te da niz novih i moguće dragocjenih spoznaja nikad ne ugleda svjetlo dana samo zato što ne ispunjava arbitrarno određen kriterij statističke značajnosti.

Naravno, takva je situacija iznimno problematična za čitavu akademsku zajednicu te se danas njome zajednice iz različitih područja intenzivno bave. No ti problemi ne dovode nužno do potpunoga otklona od paradigme zasnovane na testiranju nulte hipoteze (koja je u mnogim područjima, uključujući usvajanje drugoga jezika, i dalje dominantna). Umjesto toga sve se češće čuju prijedlozi o ukidanju strogih graničnih vrijednosti za statističku značajnost te za fleksibilnije tumačenje vrijednosti p , zatim se poziva na oslanjanje na detaljnije opisivanje i vizualiziranje podataka (usp. treće poglavlje) te na izračunavanje dodatnih vrijednosti uz podatak o statističkoj značajnosti, poput veličine efekta (v. posebice potpoglavlja 4.1.3 i 4.1.4). Predlažu se i novi testovi koji ne izlaze posve iz okvira

⁹⁷ U engleskome se jeziku često upotrebljava izraz *p-hacking*, koji označava „jurenje“ za niskom vrijednosti p , kao i niz izraza koji izravno upućuju na manipulaciju podataka, npr. *data butchery* „kasapljenje podataka“ ili *torturing the data until they confess* „mučenje podataka dok ne priznaju“. Wikipedia na mrežnoj stranici https://en.wikipedia.org/wiki/Data_dredging (posljednji pristup 26. 4. 2024.) sadrži niz objašnjenja te pojave, kao i poveznice za dodatne tekstove na tu temu. Na mrežnoj stranici <https://xkcd.com/1478/> (posljednji pristup 26. 4. 2024.) može se vidjeti slikovit stripovni prikaz te prakse.

testiranja statističke značajnosti, ali upotrebljavaju pristupe u izračunavanju koji daju pouzdanije rezultate (tzv. robusna statistika, kojoj pripadaju metode utemeljene na višestrukim analizama poduzoraka, takozvano uzorkovanje s povratom, na koje ćemo se kratko osvrnuti u potpoglavlju 5.6, v. i potpoglavlje 3.1.6). Sve se češće susreće i Bayesovo statističko zaključivanje. Osim toga u potpoglavlju 7.1. bit će riječi o problemu pristranosti u objavljivanju radova te će se navesti replikacija, metaanaliza i pretprijava istraživanja kao oblici borbe protiv postojećih praksi, ali i kao oblici kontrole nad pojedinačnim istraživanjima. Naravno, uza sve se to dodatno inzistira i na preciznim objašnjenjima značenja vrijednosti p i zaključaka koji se iz nje mogu izvesti, čemu želimo pridonijeti i ovom knjigom.

4.1.3 Veličina efekta

Među dodatnim statističkim pokazateljima čije se navođenje preporučuje uz informaciju o statističkoj značajnosti posebna se važnost pridaje mjerama **veličine efekta** ili **veličine učinka** (engl. *effect size*), koje pokazuju koliki je efekt nezavisne varijable, a ne samo ima li ga ili ne, te time ukazuju na razinu praktične značajnosti rezultata (v. npr. Petz i dr. 2012, str. 353). O mjerama veličine efekta dodatno će biti riječi u petome poglavlju jer se uz različite testove izračunavaju različite mjere. Ovdje želimo ponajprije objasniti njihovu opću ulogu u analizi podataka i njihovo uobičajeno tumačenje te naglasiti to da se u području usvajanja drugoga jezika, kao i u drugim znanstvenim disciplinama, vrlo aktivno potiče njihovo izračunavanje i navođenje u radovima (v. npr. Plonsky i Gass, 2011).

Veličina efekta tiče se razlike između statističke značajnosti i **praktične značajnosti** (engl. *practical significance*) te ukazuje na to je li statistički značajan rezultat ujedno i smisljeno značajan u kontekstu danoga istraživanja. Drugim riječima, statistička značajnost i praktična relevantnost rezultata ne moraju se nužno poklapati, zbog čega vrijednost p nije neposredni pokazatelj važnosti dobivenih rezultata (Tenjović i Smederevac, 2011, str. 321). Osnovna je značajka mjera veličine efekta ta da prikazuju rezultate statističkoga testa u svjetlu veličine uzoraka koji su analizirani. Naime manji su uzorci skloni većoj varijabilnosti pa se relativno lako može dogoditi da rezultati statističkoga testa za takve uzorke ne budu značajni unatoč postojanju efekta nezavisne varijable, dok je kod velikih uzoraka često suprotno, tj. statistički značajni rezultati dobivaju se i zbog same veličine uzorka, zbog koje i vrlo mali efekt može

djelovati značajno.⁹⁸ Drugim riječima, isključivo oslanjanje na vrijednost p u statističkome zaključivanju nedovoljno je jer se u većim uzorcima statistički značajni rezultati dobivaju mnogo lakše nego u malim uzorcima (v. Larson-Hall i Plonsky, 2015, str. 135).

Kako je već napomenuto, uz različite statističke testove mogu se izračunati različite mjere veličine efekta. Među najjednostavnijim mjerama nalazi se razlika između aritmetičkih sredina dobivenih za dvije skupine ispitanika. Međutim, budući da takva mjera tijesno ovisi o veličini jedinica koje se koriste (usp. primjerice vrijeme reakcije u milisekundama, koje uobičajeno iznosi nekoliko stotina, s brojem bodova u testu dopunjavanja teksta s maksimalnim brojem bodova od 40), uobičajenije je upotrebljavati standardizirane mjere koje olakšavaju usporedbu između različitih istraživanja. Lindstromberg (2016a) navodi da se u literaturi mogu identificirati četiri glavna tipa mjera veličine efekta. Jednome pripadaju mjere razlike između aritmetičkih sredina ili medijana dvaju uzoraka, najčešće izražene kao standardizirana razlika dobivena dijeljenjem razlike standardnom devijacijom, što predstavlja mjeru **Cohenov d** (engl. *Cohen's d*). Često se upotrebljavaju i mjere povezanosti poput **Pearsonova** ili **Spearmanova koeficijenta korelacije** (engl. *Pearson's/Spearman's correlation coefficient*; v. potpoglavlje 5.4), zatim mjere utemeljene na dijeljenju količine varijabilnosti u podacima koja se može objasniti nezavisnom varijablom ukupnom varijabilnošću u zavisnoj varijabli (**kvadrirana eta**, engl. *eta squared*, η^2). Konačno, novijega je datuma **opća mjera veličine efekta** (engl. *common language effect size*), utemeljena na izračunavanju vjerojatnosti da slučajno odabrani podatak iz jedne skupine ima veću vrijednost od slučajno odabranoga podatka iz druge skupine. O tim ćemo mjerama još govoriti u petome poglavlju, a dodatno upućujemo čitatelje na rad Tenjovića i Smederevac (2011), u kojemu je objašnjeno još nekoliko mjera, kao i uvjeti njihove uporabe te mogućnosti izračunavanja u nekima od često korištenih statističkih programa. Larson-Hall (2012a) također daje korisne informacije o mogućnostima računanja mjera veličine efekta za najčešće korištene statističke testove u području usvajanja drugoga jezika.

Veličina efekta obično se klasificira kao mala, srednja ili velika. Naprimjer Plonsky i Oswald (2011, str. 286) za Cohenov d predlažu da se u području usvajanja drugoga jezika d od 0,40 smatra malim efektom, 0,70 srednjim efektom, a 1,00 velikim efektom, uz napomenu da se prema tim vrijednostima

⁹⁸ Ekstremnu ilustraciju utjecaja veličine uzorka daju Bennett i sur. (2009), koji su pokazali da fMRI skener uz dovoljno velik broj snimanja detektira moždanu aktivnost i kod mrtvog lososa.

ne treba odnositi nekritički i nužno ih primjenjivati na istraživanje svih tema unutar vrlo širokoga i raznovrsnoga područja (v. i Plonsky i Oswald 2014, gdje se поближе govori i o različitim vrijednostima za testove s nezavisnim i zavisnim uzorcima).⁹⁹ Larson-Hall i Plonsky (2015) dodatno ističu da se u interpretaciji ne treba zaustaviti na grupiranju veličine efekta u jednu od triju kategorija, slično dihotomnome tumačenju statističke značajnosti, već konkretno dobivenu vrijednost treba interpretirati u svjetlu rezultata drugih istraživanja na slične teme ili u sličnim kontekstima. Boers i sur. (2021) primjerice ističu da je uobičajeno da se u istraživanjima koja uključuju testiranje ispitanika prije i poslije određene intervencije zabilježi veći efekt nego kad se usporedi napredak dviju skupina izloženih različitim intervencijama.

Metaanalize upućuju na relativno nisku prisutnost mjera veličine efekta u objavljenim radovima iz područja usvajanja drugoga jezika (uz već citirane radove, v. i Norris, 2015; Plonsky, 2013). Autori koji to opažaju ujedno potiču istraživače da promijene tu praksu i da obavezno izračunavaju i navode ove mjere među rezultatima. Al-Hoorie i Vitta (2019) čak navode kako je bez proratne mjere veličine efekta podatak o statističkoj značajnosti „trivijalan“. Možemo uz takve poticaje samo još istaknuti da je danas izračunavanje mjera veličine efekta lako dostupno u programima za statističku analizu te da nema ni praktičnih razloga za njihovo izostavljanje iz analize podataka.

4.1.4 Standardna pogreška i granice pouzdanosti

Još dvije dodatne mjere koje su vrlo korisne u procjeni dobivenih rezultata statističke analize, a za koje se često navodi da nisu dovoljno prisutne u istraživanjima usvajanja drugoga jezika, jesu **standardna pogreška** (engl. *standard error*, SE) i **granice pouzdanosti**, koje se još nazivaju i **granice povjerenja** ili **granice sigurnosti** (engl. *confidence intervals*, CI). Te su mjere već kratko spomenute u potpoglavlju 3.2.2, gdje se govorilo o njihovu prikazu na stupčastim grafikonima putem crta koje označavaju intervale pogreške. Kako ćemo ubrzo vidjeti, one na prvi pogled mogu nalikovati deskriptivnim mjerama (doista i jesu povezane sa standardnom devijacijom kao mjerom varijabilnosti), ali zapravo pripadaju inferencijalnoj statistici jer nisu vezane samo uz podatke iz uzorka koji se analizira. Obje se mjere mogu izračunati za različite statističke pokazatelje, ali se najčešće izračunavaju za aritmetičku sredinu, te ćemo ih u ovome potpoglavlju tako ilustrirati.

⁹⁹ Plonsky i Oswald (2014) prilagođavaju, na temelju analize određenoga broja objavljenih radova, izvorno predložene vrijednosti (v. prikaz izvornih vrijednosti u Tenjović i Smederevac, 2011, str. 327).

Standardna pogreška aritmetičke sredine dobiva se dijeljenjem standardne devijacije uzorka (s) kvadratnim korijenom broja opažanja, odnosno broja ispitanika (N); formula je prikazana u (8), gdje indeks M u s_M označava aritmetičku sredinu. Iz formule je jasno da će uz istu standardnu devijaciju standardna pogreška biti manja ako je uzorak veći.

$$(8) s_M = \frac{s}{\sqrt{N}}$$

Naprimjer ako razmotrimo podatke iz istraživanja Berger i sur. (2019) o vremenu reakcije neizvornih govornika engleskoga jezika za različite riječi (v. potpoglavlje 3.1.3), uz aritmetičku sredinu od 734,159 milisekunde, standardna devijacija iznosila je 82,823 milisekunde, dok je broj ispitanika bio 1315. Iz druga dva podatka dobivamo standardnu pogrešku od 2,284.

Iako se izračunava vrlo jednostavno, standardna pogreška kao pojam i njezina uporaba u statističkoj analizi zapravo su vrlo složene.¹⁰⁰ Kako bismo ih razumjeli, trebamo imati na umu da je uzorak iz kojega smo izračunali aritmetičku sredinu samo jedan od niza mogućih uzoraka iz iste populacije. Kad bismo iz populacije izdvojili još uzoraka jednake veličine (naprimjer još skupina od 1315 neizvornih govornika engleskoga jezika), gotovo je sigurno da bi svaki uzorak imao nešto drukčiju aritmetičku sredinu te bismo mogli izračunati i aritmetičku sredinu aritmetičkih sredina različitih uzoraka, kao i pripadajuću standardnu devijaciju. Standardna pogreška predstavlja procjenu standardne devijacije koja bi se dobila kad bismo doista imali na raspolaganju aritmetičke sredine svih, ili barem vrlo velikoga broja uzoraka (iste veličine) iz neke populacije. Drugim riječima, ona je mjera varijabilnosti u distribuciji aritmetičkih sredina koje bismo dobili višestrukim ponavljanjima istoga eksperimenta. Kao takva, predstavlja dodatnu mjeru konzistentnosti uzorka, odnosno pomaže nam da procijenimo koliko dobro naš uzorak predstavlja populaciju iz koje je uzet. Kod Berger i sur. (2019) naprimjer mala standardna pogreška govori nam da je aritmetička sredina populacije vrlo vjerojatno bliska vrijednosti od 734,159 milisekunde, koja je dobivena na uzorku. Način na koji se standardna pogreška izračunava ujedno nam iznova ilustrira jedan od temelja frekvencijskoga pristupa statističkoj analizi – pretpostavku višestrukoga ponavljanja istoga eksperimenta.

Standardna pogreška može se nadalje iskoristiti u izračunavanju granica pouzdanosti. One nam daju dodatnu informaciju o procjeni populacijskih parametara poput aritmetičke sredine na temelju statistika dobivenih na

¹⁰⁰ Petz i sur. (2012, str. 143) navode da je standardna pogreška „jedini ozbiljan ‘kamen smutnje’ u razumijevanju statističkoga načina rezoniranja“.

uzorku, te o razini povjerenja koje možemo imati u te procjene (Tenjović i Smederevac, 2011, str. 322). Granice pouzdanosti naime ograničavaju raspon vrijednosti za koji se očekuje da će vjerojatno sadržavati aritmetičku sredinu populacije. Točnije, ako u velikome broju uzoraka jednake veličine uzetih iz iste populacije odredimo granice pouzdanosti, tako dobiveni raspon trebao bi u određenome postotku uzoraka sadržavati pravu vrijednost populacije. Najčešći postotak koji se razmatra jest 95 %, u kojemu slučaju možemo reći da bismo ako iz populacije uzmemo veći broj uzoraka jednake veličine, očekivali da rasponi omeđeni njihovim granicama pouzdanosti u 95 % slučajeva sadrže aritmetičku sredinu populacije (a da je u 5 % uzoraka ne sadrže). Izbor 95 % povezan je s načinom određivanja praga statističke značajnosti prethodno opisanim u potpoglavlju 4.1.2.

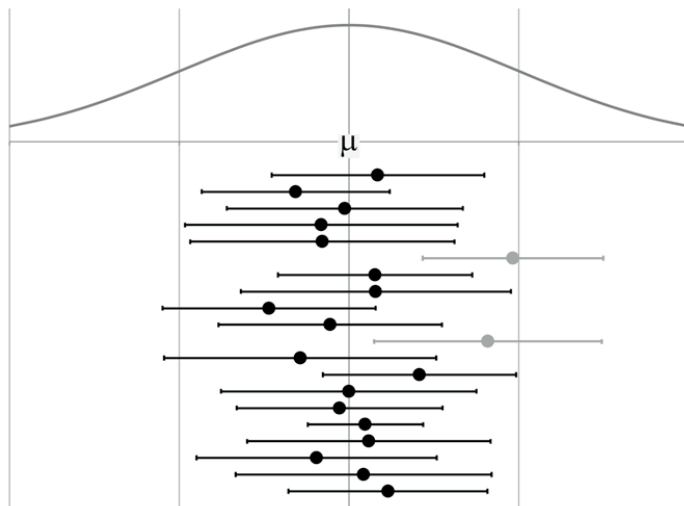
Granice pouzdanosti (*GP*) izračunavaju se iz aritmetičke sredine dobivene iz uzorka i iz standardne pogreške aritmetičke sredine. Čine ih dvije vrijednosti simetrično distribuirane oko aritmetičke sredine, koje se dobivaju tako što od aritmetičke sredine s jedne strane oduzmemo, a s druge strane njoj dodamo standardnu pogrešku pomnoženu odgovarajućim koeficijentom. Vrijednost će koeficijenta ovisiti o tome koji postotak sigurnosti odaberemo, a u slučaju 95 % on iznosi 1,96, što je prikazano i u formuli u (9). Koeficijent 1,96 dobiven je iz standardne normalne distribucije (v. potpoglavlje 3.1.5) jer se kod takve distribucije 95 % podataka nalazi u rasponu $M \pm 1,96 \cdot SD$.

$$(9) \quad GP = M \pm 1,96s_M$$

Naprimjer u istraživanju Berger i sur. (2019) uz aritmetičku sredinu od 734,159 milisekunde i standardnu pogrešku od 2,284 milisekunde granice pouzdanosti od 95 % iznosit će $734,159 \pm 1,96 \cdot 2,284$, odnosno 729,679–738,639. Ako bi eksperiment bio ponovljen s jednakim brojem ispitanika, dobile bi se drukčija aritmetička sredina i drukčija standardna pogreška te bi i granice pouzdanosti posljedično imale drukčiju vrijednost. U stotinu bismo takvih ponavljanja očekivali da ćemo dobiti 95 granica pouzdanosti koje sadrže populacijsku vrijednost i pet koje je ne sadrže. Na sličan bismo način mogli izračunati i granice pouzdanosti od 99 %, s time što bi tada koeficijent u formuli iznosio 2,58 i raspon vrijednosti bio bi nešto širi, ali mogli bismo s nešto većom sigurnošću zaključiti da se „prava“ (populacijska) aritmetička sredina vjerojatno nalazi u tome rasponu (v. i Petz i sur., 2012, str. 142).

Kako bi pojam granica pouzdanosti bio još jasniji, upotrebljavamo grafikon izrađen pomoću dokumenta u programu Microsoft Excel koji su pripremili

Krzywinski i Altman (2013).¹⁰¹ Na tome se grafikonu, na slici 37, može vidjeti rezultat simulacije uzimanja 20 različitih uzoraka ($n=10$) iz iste normalno distribuirane populacije čija je aritmetička sredina (μ) poznata. U tome slučaju vidimo da većina uzoraka (crna crta) ima granice pouzdanosti aritmetičke sredine koje obuhvaćaju aritmetičku sredinu uzorka, ali vidimo i da u dva uzorka (sive crte) nije tako.¹⁰² Dakle, kad se kaže da granice pouzdanosti u određenome postotku slučajeva obuhvaćaju vrijednost populacije, to ne podrazumijeva jedan konkretan raspon dobiven u jednome eksperimentu.



Slika 37. Simulirane aritmetičke sredine i 95 %-tne granice pouzdanosti

Treba također napomenuti da je opisani način izračunavanja primjeren za veće uzorke (u pravilu veće od 30 ispitanika), dok se za manje uzorke upotrebljava drukčija formula (v. npr. Levshina, 2015, str. 99; usp. Cumming, 2014, str. 18).¹⁰³ Tom se formulom nećemo ovdje baviti jer nam je cilj objasniti načelo koje je u temeljima određivanja granica pouzdanosti, ali je potrebno napomenuti da su te granice od vrlo ograničene praktične koristi za vrlo male uzorke, u kojima je varijabilnost nužno veća (v. Cumming 2014, str. 16). O formuli za manje uzorke najčešće nije nužno previše brinuti jer se izračunavanje gotovo

¹⁰¹ <https://www.nature.com/articles/nmeth.2659#Sec1> (posljednji pristup 26. 4. 2024.)

¹⁰² Prema načelu 95 % očekivao bi se jedan takav uzorak, ali empirijska je situacija rijetko idealna, posebice u manjim uzorcima.

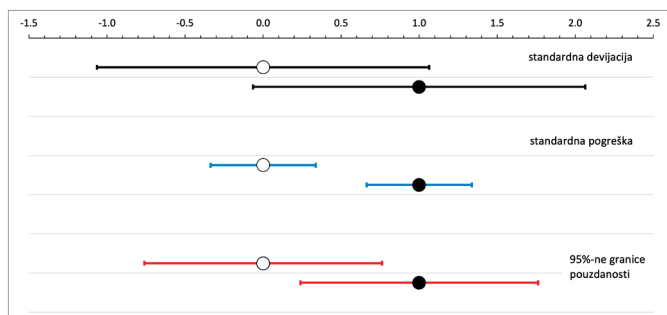
¹⁰³ Odnos između standardne pogreške i granica pouzdanosti također je manje izravan kod pojedinih mjera poput koeficijenta korelacije (Cumming 2014, str. 18).

uvijek provodi u računalnim programima za statističku analizu. Ti programi automatski uključuju standardnu pogrešku i 95 %-tne granice pouzdanosti u rezultate velikoga broja statističkih analiza za koje su relevantni. Naprimjer IBM SPSS u okviru osnovnih deskriptivnih pokazatelja daje podatak o standardnoj pogrešci aritmetičke sredine, koeficijenta simetričnosti distribucije i koeficijenta spljoštenosti krivulje, a prilikom uspoređivanja aritmetičkih sredina dvaju uzoraka (v. potpoglavlje 5.2.1) daje podatak o standardnoj pogrešci i granicama pouzdanosti razlike, dok u okviru opisnoga dijela analize može dati i podatke o standardnoj pogrešci i granicama pouzdanosti aritmetičke sredine za svaki pojedini uzorak. Slične podatke daje i R, ponajprije u okviru rezultata statističkih testova.

Granice pouzdanosti, kao i standardna pogreška, ali i standardna devijacija uzorka, često se mogu vidjeti na grafičkome prikazu podataka, posebice kod stupčastih i linijskih grafikona, kao crte koje označavaju intervale pogreške (v. potpoglavlje 3.2.2). Sve tri mjere predstavljaju mjere nesigurnosti u pogledu aritmetičke sredine, ali sad bi trebalo biti jasno da imaju različita značenja te se trebaju različito i tumačiti na grafikonima: standardna devijacija pokazuje raspršenost podataka u uzorku, standardna pogreška procjenjuje veličinu pogreške u mjerenju aritmetičke sredine, dok granice pouzdanosti omeđuju interval koji vjerojatno sadrži aritmetičku sredinu populacije.

Sve se te tri mjere barem donekle mogu upotrijebiti kako bi se već „odoka“ vidjelo je li razlika između dviju aritmetičkih sredina koje potječu iz nezavisnih skupina značajna. Kako navode Krzywinski i Altman (2013; v. i Belia i sur., 2005; Larson-Hall, 2017), procjenu je najteže napraviti kod standardne devijacije jer ona izražava raspršenost podataka u uzorku i moguće je da razlika između dvaju prosjeka bude statistički značajna, iako se crte velikim dijelom preklapaju (odnosno, vrijednosti se podataka iz dvaju uzoraka u većoj mjeri poklapaju). Standardna pogreška drukčije je prirode jer ne predstavlja mjeru raspršenosti već pogreške u mjerenju – dvije se aritmetičke sredine statistički značajno razlikuju samo ako se crte koje označavaju intervale pogreške utemeljene na s_M uopće ne dodiruju i udaljene su jedna od druge približno jednu standardnu pogrešku. To postaje još jasnije kad znamo da crte utemeljene na standardnim pogreškama približno odgovaraju granicama pouzdanosti od 67 % (barem u većim uzorcima). Ako crte pogreške prikazuju granice pouzdanosti, one se mogu djelomično preklapati (25–50 %, ovisno o veličini uzoraka) i ipak dati statistički značajnu razliku. Na slici 38 još jednom upotrebljavamo dokument koji su podijelili Krzywinski i Altman (2013), ovoga puta za usporedni prikaz različitih crta koje označavaju intervale pogreške.

Slika pokazuje odnos crta zasnovanih na standardnoj devijaciji, standardnoj pogrešci i granicama pouzdanosti pri postojanju statistički značajne razlike razine $p < 0,05$ između dviju aritmetičkih sredina (označenih kružićima).



Slika 38. Usporedba različitih crta koje označavaju intervale pogreške

Posljednjih se godina u istraživanjima usvajanja drugoga jezika, kao i u drugim područjima, podosta inzistira na obaveznome navođenju granica pouzdanosti u radovima (v. Larson-Hall, 2010; Larson-Hall i Plonsky, 2015; Lindstromberg, 2016a; kao i Cumming, 2014, za područje psihologije). Njihovo navođenje preporučuju i smjernice Američkoga psihološkog udruženja, na koje se lingvisti često oslanjaju. Primjerice, Cunnings i Fujita (2021) ih dosljedno navode u opisu rezultata analize podataka prikupljenih zadatkom čitanja vlastitim tempom u istraživanju individualnih razlika u obradi „rečenica stranputica“ (engl. *garden path sentences*) na engleskome jeziku kod izvornih i neizvornih govornika. Razloga za takve preporuke ima više i one se ne odnose samo na granice pouzdanosti aritmetičke sredine već i na slične granice za druge mjere. Prvi je razlog dopuna, ili čak zamjena (v. npr. Cumming, 2014; Larson-Hall, 2010, str. 57), za uobičajeni način procjene rezultata putem statističke značajnosti, za koji je u prethodnim potpoglavljima već navedeno da ima određenih nedostataka. Naime autori koji se zalažu za širu uporabu granica pouzdanosti često smatraju da se one trebaju upotrebljavati za procjenu značajnosti razlike između dvaju uzoraka. Naprimjer Larson-Hall i Plonsky (2015, str. 136–137) navode da kad usporedimo dva uzorka i utvrdimo da granice pouzdanosti razlike između njihovih prosjeka ne obuhvaćaju vrijednost 0, možemo zaključiti da je razlika između uzoraka značajna. Levshina (2015, str. 103) ističe da nepreklapanje granica pouzdanosti aritmetičkih sredina dvaju uzoraka također ukazuje na postojanje statistički značajne razlike, veliko preklapanje na izostanak statistički značajne razlike, dok je u slučaju manjega preklapanja potrebno okrenuti se statističkome testu. Cumming (2014) ističe

kao prednost temeljenja odluka o značajnosti na granicama pouzdanosti to što se kod njih radi o rasponu vrijednosti i procjeni (engl. *estimation*), a ne o dihotomnome odlučivanju kao kod statističke značajnosti.¹⁰⁴ Kako ističe taj autor, interpretacija rezultata na temelju granica pouzdanosti (i veličine efekta) predstavlja „informiranu procjenu u danome kontekstu“ (engl. *knowledgeable judgment in context*), odnosno manje je doslovna i manje utemeljena na strogim crno-bijelim smjernicama nego vrijednost p .

Ulogu granica pouzdanosti pojedini istraživači vide i u predviđanju ishoda replikacijskih studija te naprimjer Cumming (2014, str. 13, 17) navodi da 95 %-tne granice pouzdanosti predstavljaju 83 %-tni interval predviđanja dane vrijednosti (naprimjer razlike između dviju aritmetičkih sredina) u repliciranim eksperimentima, odnosno da se može očekivati da procijenjena vrijednost bude u okviru granica pouzdanosti okvirno 5 od 6 puta. Istodobno se granice pouzdanosti ističu kao koristan pokazatelj preciznosti procjene odgovarajućega parametra, odnosno smatra se da uži interval upućuje na bolju procjenu vrijednosti populacije od širega intervala (v. npr. Larson-Hall i Plonsky, 2015; Lindstromberg, 2016a).

Međutim, iako granice pouzdanosti nesporno mogu ukazati na to koliko podaci u uzorku variraju, brojni istraživači savjetuju oprez u njihovoj uporabi i svraćaju pozornost na pogrešna tumačenja koja im se ponekad pripisuju. Jedno je od čestih pogrešnih tumačenja, na koje svraćaju pozornost naprimjer Levshina (2015, str. 103), da granice pouzdanosti označavaju vjerojatnost (od 95 % ili neku drugu) da se u njihovim okvirima nalazi prava populacijska vrijednost. Još je složenije pitanje izvođenja zaključaka o pouzdanosti, odnosno tretiranje granica pouzdanosti kao indikatora preciznosti rezultata dobivenih u uzorku. Granice pouzdanosti nisu izvorno definirane kao mjera preciznosti i pogrešno im je pripisati takvo značenje. Naime prema tome kako su prvotno definirane, granice pouzdanosti odnose se isključivo na „dugoročnu“ (u većemu broju ponovljenih eksperimenata) vjerojatnost pojavljivanja rezultata iz određenoga raspona (Al-Hoorie i Vitta 2019, str. 12). Morey i sur. (2016) među autorima su koji su se osobito mnogo bavili tom problematikom, u području psihologije i šire, detaljno objašnjavajući da granice pouzdanosti ne mogu imati značenje preciznosti koje im se često pripisuje, te da nije utemeljeno upotrebljavati ih umjesto testiranja statističke značajnosti u svrhu izvođenja zaključaka o populaciji (v. i Greenland i sur., 2016).

¹⁰⁴ Cumming (2014, str. 14) predlaže i formuliranje istraživačkih pitanja u obliku procjena. To bi značilo da bi pitanja trebala imati oblik poput „U kojoj mjeri varijabla X utječe...“ umjesto „Utječe li varijabla X ...“.

Možemo stoga zaključiti da je u radu s granicama pouzdanosti potreban velik oprez. One mogu biti korisne kao dodatak na grafičkim prikazima podataka, gdje nam u nekim situacijama mogu doista pomoći u formuliranju pretpostavke o statističkoj značajnosti razlike (usp. i usjeke na dijagramu s pravokutnikom, opisane u potpoglavlju 3.2.2). Možemo se pokušati osloniti na neke savjete proizašle iz njihove uporabe, naprimjer da formuliramo pitanja i hipoteze manje dihotomno. Međutim, u svjetlu neslaganja prisutnih u literaturi, ne treba ih nužno prihvatiti kao potpunu zamjenu za inferencijalne statističke testove jer njihove značajke ne dopuštaju nedvosmisleno takvo tumačenje.

4.1.5 Analiza statističke snage testa

Pojam snage statističkoga testa uvele smo u početnim potpoglavljima ovoga poglavlja i više smo se puta na njega vraćale. On se, podsjećamo, odnosi na moć statističkoga testa da ispravno detektira statistički značajnu razliku onda kada ona doista postoji (odnosno da odbaci nultu hipotezu kad ona nije točna). Taj je pojam povezan i s veličinom efekta, kao i s utvrđivanjem veličine uzorka koji je potrebno obuhvatiti istraživanjem.

Istraživači usvajanja drugoga jezika, posebice početnici, vrlo često žele dobiti jednostavan odgovor na pitanje koliko je ispitanika potrebno uključiti u istraživanje. Iako postoji vrlo općenito pravilo da uzorci trebaju biti „što veći“,¹⁰⁵ a dodatno je moguće formulirati opće smjernice na temelju prethodnih istraživanja i spoznaja iz bliskih područja (usp. potpoglavlje 2.1.3), za precizno planiranje svakoga pojedinog eksperimenta potrebno je prije prikupljanja podataka provesti **analizu statističke snage testa** (engl. *power analysis*), čiji je dio veličina uzorka. Mnogobrojni autori preporučuju taj korak kao obavezan i inzistiraju na njemu (kao i na detaljnome izvještavanju o njemu) upravo zato što je on danas i dalje rijedak u istraživanjima usvajanja drugoga jezika (v. npr. Larson-Hall, 2016; Norris i sur., 2015; Plonsky, 2015). Istodobno međutim postoji i svijest o tome da takvu analizu nije nimalo jednostavno provesti (v. npr. Plonsky, 2013).

Analiza statističke snage testa temelji se na više komponenata: (1) ciljnoj snazi testa, (2) odabranoj razini statističke značajnosti (α), (3) varijabilnosti podataka, (4) očekivanoj veličini efekta i (5) veličini uzorka. Te su komponente međusobno usko povezane, i to je ponajprije snaga testa povezana s ostalim komponentama. Naime snaga statističkoga testa povećava se kad je razina

¹⁰⁵ Petz i sur. (2012, str. 5) navode „Kao što se logički i može očekivati, inferencijalna je statistika to „opreznija“ u zaključivanju što smo na manjemu broju podataka neke rezultate dobili. Naprotiv, što je broj podataka veći, to možemo više biti sigurni da dobiveni rezultati vrijede i „inače“, a ne samo u slučaju koji smo izmjerili.“

statističke značajnosti viša, kad je varijabilnost podataka niža, kad je veličina efekta viša i kad je uzorak veći.

Za ciljnu snagu testa, koja predstavlja mjeru željene preciznosti rezultata (v. Petz i sur., 2012, str. 7), obično se podrazumijeva da treba iznositi 80 %, odnosno 0,80, što se obično uparuje s graničnom razinom statističke značajnosti od 5 %, odnosno 0,05 (v. potpoglavlje 4.1.2). Varijabilnost podataka ne može nam biti poznata unaprijed i ne možemo je odrediti proizvoljno, ali možemo se osloniti na prethodna istraživanja na slične teme ili na vlastito probno istraživanje (tj. pilotiranje, v. potpoglavlje 2.1.6) i odatle izvesti pretpostavku o očekivanoj varijabilnosti, što vrijedi i za očekivanu veličinu efekta. Konačno, u slučaju kad analizu snage upotrebljavamo u svrhu određivanja potrebne veličine uzorka, ta je veličina nepoznata, tj. predstavlja rezultat koji želimo dobiti iz analize. Ovdje nećemo ulaziti u formule koje se prilikom analize snage koriste. Mnogobrojni kalkulatori analize snage dostupni su izravno putem interneta ili se mogu jednostavno preuzeti s njega. U istraživanjima jezika najčešće se citira program G*Power (Faul i sur., 2009; <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>, posljednji pristup 26. 4. 2024.). Nešto je novijega datuma *online*-kalkulator za *t*-testove dostupan na mrežnoj stranici <https://rezanorouzian.shinyapps.io/power/> (posljednji pristup 26. 4. 2024.), koji je razvio Reza Norouzian.¹⁰⁶ U takve se programe unose vrijednosti koje poznajemo (ili ih pretpostavljamo), a na temelju njih se zatim dobiva podatak o vrijednosti koju želimo doznati (najčešće je to potreban broj ispitanika). Sučelje koje je razvio Norouzian čak ne zahtijeva unošenje podatka o varijabilnosti, te ako pretpostavimo nezavisni *t*-test (v. potpoglavlje 5.2.1) s pragom statističke značajnosti od 0,05, s veličinom efekta Cohenov $d=0,6$, sa snagom testa od 80 %, dobit ćemo da je potrebno uključiti 45 ispitanika u svaku od dviju skupina (odnosno 90 ispitanika ukupno).

Iako se može učiniti da se analiza statističke snage može provesti i nakon prikupljanja podataka, kad već poznajemo potrebne parametre, to zapravo ne treba činiti. Analiza snage predstavlja postupak koji je dio pripreme istraživanja, a nakon prikupljanja podataka ne govori se više o (priprema) analizi snage, već o izračunavanju snage koja je doista dobivena u danome istraživanju. Također, ono što prilikom takvih analiza ne treba zaboraviti jest da je određivanje veličine uzorka vezano i uz broj podražaja, a ne samo uz broj ispitanika (v. potpoglavlje 2.1.4). Kalkulatori navedeni u ovome potpoglavlju podrazumijevaju ponajprije situaciju u kojoj se kod svakoga ispitanika vrši

¹⁰⁶ Kalkulator se temelji na sučelju za R koje je izradio Mizumoto, a koje smo spomenule u trećemu poglavlju.

jedno ili mali broj mjerenja, što u istraživanjima jezika najčešće nije slučaj. Kao što smo vidjeli u ranijim poglavljima (v. i Kraš i Miličević, 2015), u istraživanjima usvajanja drugoga jezika koriste se zadaci u kojima se jezične pojave gotovo uvijek predstavljaju putem većega broja čestica kako bi se osiguralo dobivanje pouzdanijih odgovora. Postojanje većega broja čestica može u nekoj mjeri nadomjestiti manjak ispitanika.

Plonsky (2013) ističe da je praksa provedbe analize snage vrlo rijetka u istraživanjima usvajanja drugoga jezika, ali da ujedno može biti i teško provediva. U kontekstu planiranja broja ispitanika važno je stoga imati na umu da analiza statističke snage nije jedini način za to. Problem veličine uzorka detaljno razmatraju Buchstaller i Khattab (2014), Norouzian (2020) te Nicklin i Vitta (2021), pa čitatelje koji se žele upoznati s drugim metodama upućujemo na te radove. Uz to Lindstromberg i Eyckmans (2017) ističu da postoje tri alternativna puta povećavanja snage statističkoga testa osim povećavanja uzorka i prilagođavanja drugih kriterija iz analize snage. U pitanju su uporaba optimiziranih eksperimentalnih nacrti,¹⁰⁷ robusnih postupaka poput uzorkovanja s povratom i Bayesova statističkog zaključivanja (v. potpoglavlje 5.6).

4.2 Statistički testovi

Provedba statističkih testova posljednji je korak u kvantitativnome dijelu analize podataka te nakon njega preostaje još interpretirati dobivene rezultate i zaključiti daju li oni potporu istraživačkoj hipotezi ili takva potpora izostaje. Odabir testa koji odgovara podacima i nacrtu istraživanja smatra se jednom od najvažnijih odluka u procesu analize. U ovome potpoglavlju govorimo o čimbenicima koji utječu na tu odluku, ali prije toga ističemo načela koja su zajednička svim statističkim testovima te dajemo primjer jednostavnoga testa kao uvod u način razmišljanja koji stoji u pozadini i složenijih testova.

4.2.1 Zajedničke značajke statističkih testova

Kao opći temelj za tumačenje svih testova kojima ćemo se baviti potrebno je uz statističku značajnost, detaljno objašnjenu na početku ovoga poglavlja, poznavati još nekoliko pojmova. Kao prvo, svi statistički testovi kao rezultat

¹⁰⁷ Radi se o nacrtima koji su optimizirani prema nekome statističkom kriteriju, odnosno nacrti u kojima se eksperiment prilagođava tako da može pouzdano istražiti efekte varijabli uz što manji broj ispitanika ili mjerenja. To može uključivati složena statistička razmatranja (v. npr. McClelland, 1997), ali i odabir između nezavisnih i zavisnih skupina te druge odluke o kontroli istraživanja o kojima je bilo riječi u drugome poglavlju.

daju odgovarajući **testni statistik** (engl. *test statistic*) izračunat iz deskriptivnih mjera. Naziv „statistik“ za mjere izračunate iz uzorka spomenut je već nekoliko puta u prethodnome dijelu knjige (npr. aritmetička sredina uzorka predstavlja jedan statistik). Načelno je pojam „statistik“ suprotstavljen pojmu „parametar“, koji se odnosi na vrijednosti u populaciji (v. i objašnjenje parametrijskih testova u potpoglavlju 4.3.2). U kontekstu inferencijalnih testova pod pojmom „statistik“ ponajprije se podrazumijeva vrijednost izračunata određenim testom. Svaki test dakle kao rezultat daje odgovarajući testni statistik, a utvrđivanje statističke značajnosti nadalje se temelji na poznavanju distribucije toga statistika i na tome koliko je prema toj distribuciji vjerojatno da se dobivena vrijednost, ili vrijednosti ekstremnije od nje, javi u slučaju točnosti nulte hipoteze. Konkretno, značajnost se utvrđuje na temelju **kritične vrijednosti** (engl. *critical value*), od koje testni statistik najčešće treba biti veći (samo kod nekolicine testova manji) kako bi se moglo govoriti o značajnome rezultatu. To znači da se vrijednost testnoga statistika uspoređuje s kritičnom vrijednošću, pri čemu se odgovarajuća kritična vrijednost pronalazi na temelju distribucije statistika za odgovarajuću veličinu uzorka, odnosno za odgovarajuće stupnjeve slobode.

Stupnjevi slobode (engl. *degrees of freedom*) načelno pokazuju koliki broj vrijednosti može varirati, a da se rezultat pritom ne promijeni. Jednostavan primjer izvan inferencijalnih testova imali bismo kad bismo htjeli odabrati tri broja čiji zbroj iznosi 20 – prva dva broja u tome slučaju možemo varirati (naprimjer odabrati 3 pa 4, ili 8 pa 11), ali kod trećega nemamo izbora (morali bismo odabrati 13 u prvome, a 1 u drugome slučaju) – zato što možemo odabrati dva broja, u tome se slučaju može govoriti o dvama stupnjevima slobode. Stupnjevi slobode u statističkim testovima uvijek upućuju na veličinu uzoraka i/ili broj uzoraka koji se uspoređuju, ali se uz svaki test izračunavaju na drukčiji način. To je izračunavanje osobito kompleksno kod analize varijance (v. potpoglavlje 5.3.1), koja uključuje više vrsta stupnjeva slobode. Stupnjevi slobode nužni su kod većine testova za pronalazak odgovarajuće kritične vrijednosti u odgovarajućoj tablici.

Pored toga što su organizirane prema stupnjevima slobode, statističke tablice u kojima su prikazane kritične vrijednosti različitih testnih statistika uglavnom su organizirane i prema razinama statističke značajnosti te se u njima obično mogu pronaći kritične vrijednosti danoga statistika za $p < 0,05$, $p < 0,01$ i $p < 0,001$; u petome ćemo poglavlju vidjeti primjer. Još detaljnija objašnjenja, s primjerima, daju Mackey i Gass (2005, str. 270, 280–282), dok se jasno ilustriran primjer pojma stupnjeva slobode može naći kod Larson-Hall (2010, str. 47). Statistički programi u pravilu automatski daju sve potrebne podatke o testnome statistiku, stupnjevima slobode i statističkoj značajnosti,

te konzultiranje tablica u tome slučaju nije potrebno, samo je dobro poznavati postupak na kojemu se različiti koraci u statističkoj analizi temelje. Dodatna je prednost statističkih programa u tome što daju i točnu vrijednost p , a ne samo razinu značajnosti (osim u rijetkim slučajevima kad je vrijednost p iznimno mala).

4.2.2 Primjer statističkoga testa

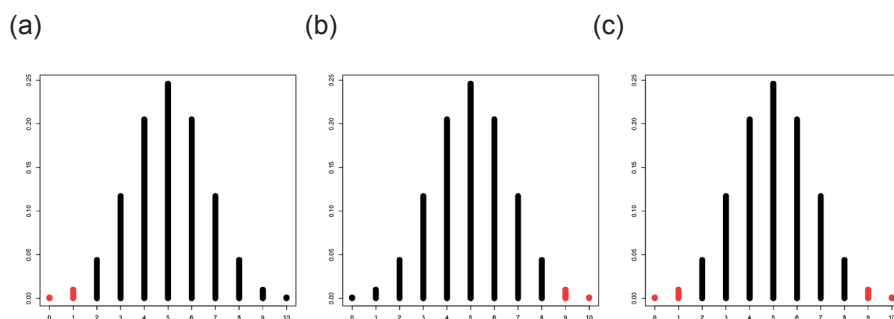
Prije nego što prijeđemo na konkretne testove treba podsjetiti da je cilj inferencijalnih statističkih testova utvrditi s kolikom sigurnošću možemo tvrditi da je razlika ili veza ustanovljena u nekome istraživanju doista vezana uz nezavisnu varijablu koja se ispituje, a da se nije pojavila kao rezultat prirodne varijabilnosti u uzorcima (v. Brown, 1991, str. 576). Kao što smo već objasnile kad smo uvele pojam pogreške tipa I (v. potpoglavlje 4.1.2), određeni rizik dobivanja statistički značajnoga rezultata i kad efekta zavisne varijable nema ne može se izbjeći, pa se statističkim testovima zapravo ispituje koliko je taj rizik velik. Drugim riječima, kod statističke je analize nužno u pitanju „igra vjerojatnosti“, koju možemo pobliže ilustrirati na jednostavnome primjeru bacanja novčića i na još jednostavnijemu testu kojim se rezultati bacanja mogu analizirati.

Već je prethodno objašnjeno kako vjerojatnost dobivanja pisma u jednome bacanju novčića iznosi 50 % ako je taj novčić „fer“ te nije iskrivljen ili na neki drugi način oštećen (v. potpoglavlje 4.1.2). Ako bismo prilikom više ponovljenih bacanja često dobivali pismo, mogli bismo posumnjati da novčić nije fer, već iskrivljen tako da češće pada na stranu pisma. Ako bismo tu pretpostavku poželjeli statistički testirati, mogli bismo izračunati koliko je za fer novčić vjerojatno da se u određenome broju bacanja pismo dobije određeni broj puta. Naprimjer ako smo dobili pismo osam puta u deset bacanja, trebamo najprije izračunati koliko je mogućih načina da se dobije osam istih ishoda iz deset bacanja, a zatim bi trebalo taj broj podijeliti ukupnim brojem mogućih ishoda – time se dobiva tražena vjerojatnost. Taj je postupak prikazan u formuli u (10), gdje n predstavlja broj bacanja, a r broj puta koliko je dobiveno pismo (u pitanju je izračunavanje broja kombinacija bez ponavljanja).

$$(10) p = \frac{\text{broj ciljnih ishoda}}{\text{ukupan broj mogućih ishoda}} = \frac{n!}{r!(n-r)!} = \frac{10!}{8!(10-8)!} = \frac{3628800}{80640} = \frac{45}{1024} = 0,04395$$

Tim putem doznajemo da je vjerojatnost dobivanja pisma točno osam puta u deset bacanja nešto niža od 0,05. Međutim, kad bismo dobili pismo devet ili deset puta, to bi također išlo u prilog pretpostavci da je novčić iskrivljen u korist

pisma te za konačni rezultat ovoga jednostavnog testa (u pitanju je **binomni test**, engl. *binomial test*) na vrijednost od 0,04395 moramo dodati vjerojatnosti dobivanja devet od deset i deset od deset pisama, čime vrijednost p dostiže 0,05469, odnosno viša je od uobičajenoga praga statističke značajnosti. Iz konačnoga rezultata moramo zaključiti da naš rezultat od osam pisama nema dovoljno malu vjerojatnost pod nultom hipotezom da bismo je odbacili. Da bi rezultat bio statistički značajan, potrebno je dakle dobiti pismo devet ili deset puta u deset bacanja. Te vrijednosti stoga predstavljaju kritičnu regiju **binomne distribucije** (engl. *binomial distribution*), što je na slici 39 pod (a) označeno crvenom bojom s desne strane distribucije. Budući da je distribucija simetrična, za nešto drukčiju hipotezu, da je novčić iskrivljen u korist glave, bila bi važna lijeva strana iste distribucije, označena pod (b). Konačno, ako bi hipoteza bila dvosmjerna – da je novčić iskrivljen, bez preciziranja na koju stranu – onda bismo morali uzeti u obzir obje strane distribucije (kritična regija za dvosmjerni test predstavlja zbroj regija za dva jednosmjerna testa te će kritična vrijednost biti viša nego za jednosmjerni test), što je prikazano pod (c).



Slika 39. Kritične regije binomne distribucije za $n=10$

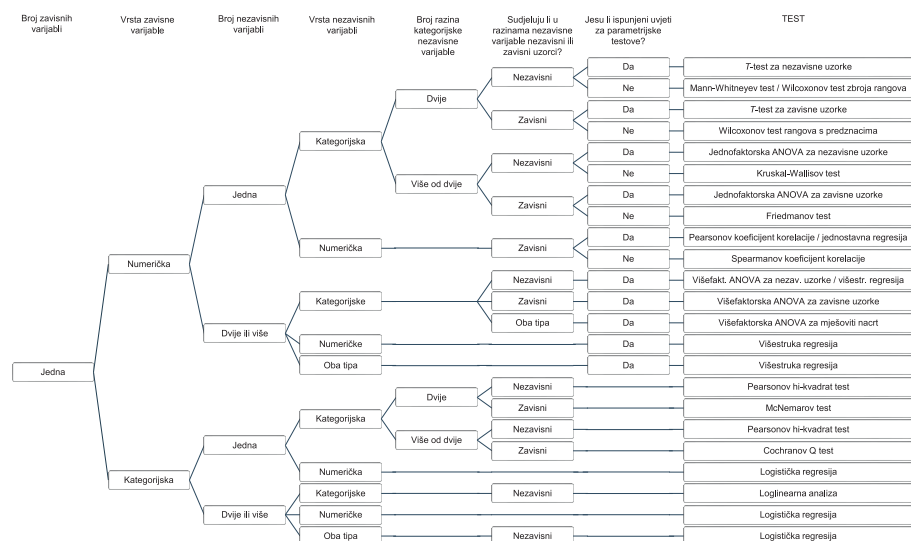
Binomni se test rijetko upotrebljava u istraživanjima usvajanja drugoga jezika te ga ovdje navodimo samo kao jednostavan uvodni primjer, dok ćemo se u sljedećemu poglavlju baviti drugim, složenijim testovima.

4.3 Kriteriji izbora testa

Odluka o tome koji će statistički test biti upotrijebljen za analizu podataka ovisi o više čimbenika vezanih uz prirodu ispitivanih varijabli i nacrt istraživanja, kao i uz značajke prikupljenih podataka. Na temelju cjelokupne literature, u ovome potpoglavlju navodimo najvažnije kriterije, za čije su nam razumijevanje nužni elementi metodologije istraživanja, čiji smo pregled dali u drugome poglavlju.

O načinu izbora testa detaljno piše Gries (2013a, 2013b), a pregledne dijagrame izbora testa, na kojima ćemo utemeljiti i naš dijagram, navode Field i sur. (2012) i Field (2018).

Kao sažetak kriterija, prije njihova detaljnijeg objašnjenja, na slici 40 prikazujemo dijagram izbora testa koji se može upotrebljavati prilikom odlučivanja (njegov format zapravo predstavlja stablo odluke). Napominjemo da dijagram nije sveobuhvatan, već obuhvaća jedan dio često korištenih testova. U petome poglavlju spomenut ćemo sve testove iz dijagrama, a dodat ćemo i neke testove kojih na dijagramu nema (prije svega testove za nacрте bez nezavisne varijable). Također neće svim testovima biti posvećena podjednaka pozornost – detaljnije ćemo se zaustaviti na onim testovima koji se vrlo često koriste i onima koji su najjednostavniji za razumijevanje i provedbu (a vidjet ćemo i da se ta dva kriterija često poklapaju). Smatramo naime da je u upoznavanju sa statističkom analizom najteže razumjeti osnove te želimo upravo tim osnovama posvetiti najviše pozornosti, dok očekujemo da je naprednije testove nakon toga lakše razumjeti i samostalno.



Slika 40. Dijagram izbora statističkoga testa (utemeljen na Field, 2018)

Elementi dijagrama objašnjeni su u sljedećim potpoglavljima. Oni se s jedne strane odnose na nacрте istraživanja i njegove značajke u pogledu broja i vrste varijabli, kao i vrste uzoraka, dok su s druge strane vezani uz značajke dobivenih podataka u pogledu mjerne ljestvice i distribucije.

4.3.1 Nacrt istraživanja

Prvi čimbenik koji je potrebno razmotriti u izboru statističkoga testa jest **broj zavisnih varijabli** koje imamo u istraživanju. Konkretnije, ključno je utvrditi postoji li jedna zavisna varijabla ili više njih (što odgovara univarijatnim nasuprot multivarijatnim nacrtima). U ovoj ćemo se knjizi baviti samo prvim slučajem, tj. nacrtima u koje je uključena jedna zavisna varijabla.¹⁰⁸ U pogledu toga može biti korisno podsjetiti se da se u području usvajanja drugoga jezika često susreću istraživanja u kojima je spomenuto više zavisnih varijabli, ali se one ne obrađuju unutar jedne jedinstvene analize. To je primjerice vrlo česta pojava kod primjene psiholingvističkih metoda kojima se istodobno mjeri i točnost odgovora i vrijeme reakcije, kao u više puta spomenutome istraživanju Berger i sur. (2019), gdje se i deskriptivne mjere i statistički testovi izračunavaju odvojeno za te dvije mjere zbog toga što se vrijeme reakcije analizira samo za točne odgovore (v. potpoglavlja 2.1.5 i 3.1.2).

Sljedeći pokazatelj koji moramo uzeti u obzir pri izboru testa jest **vrsta zavisne varijable**, odnosno njezina mjerna ljestvica. U ovome slučaju ne moramo odmah razmišljati o tome na kojoj je od četiri postojeće ljestvice varijabli izmjerena, već je dovoljno uočiti je li ona numeričke ili kategorijske prirode. Ovdje treba imati na umu da ni ta razlika ne mora uvijek biti lako uočljiva te da se ponekad i numeričke vrijednosti, ako ih je vrlo malo, moraju analizirati kao kategorije (to bi naprimjer vjerojatno bio slučaj s prosudbama prihvatljivosti rečenica koje bi bile iskazane na ljestvici od svega tri razine).

Zatim razmatramo **broj nezavisnih varijabli**, gdje je glavni kriterij podjele jednak onome kod broja zavisnih varijabli, tj. jedna varijabli nasuprot dvjema ili više (što odgovara jednofaktorskim nasuprot višefaktorskim nacrtima). Treba međutim dodati da u tome slučaju vrijednost može biti i „nijedna“. Naime zavisna varijabla može u rijetkim slučajevima biti i jedina varijabla u analizi. Najčešće se takva situacija odnosi na jedan dio unutar šire analize podataka, u kojemu je potrebno odvojeno razmotriti zavisnu varijablu. U istraživanjima usvajanja drugoga jezika to nije često jer nas najčešće zanima kako na ishod procesa usvajanja utječu čimbenici poput materinskoga jezika, dobi na početku usvajanja ili vrste konstrukcije. Međutim moguće je da nas zanima primjerice možemo li neki rezultat učenika drugoga jezika pripisati slučajnome pogađanju (engl. *chance*). U tome ćemo slučaju morati odabrati neki od

¹⁰⁸ Među analizama koje nisu uključene u dijagram na slici 40 i koje uključuju više zavisnih varijabli nalazimo naprimjer multivarijatnu analizu varijance (engl. *multivariate analysis of variance*, MANOVA), kanoničku korelacijsku analizu (engl. *canonical correlation analysis*), analizu glavnih komponenti (v. potpoglavlje 1.4) i mnogobrojne druge.

testova koji pripada skupini **testova za jedan uzorak** (engl. *one sample tests*), u kojima se provjerava razlikuje li se primjerice aritmetička sredina eksperimentalne skupine od neke unaprijed zadane vrijednosti (npr. točnosti od 50 %, koja se može uzeti kao pokazatelj slučajnoga pogađanja) ili **testova prikladnosti modela** (engl. *goodness of fit*), gdje se provjerava odgovara li distribucija dobivenih podataka nekoj unaprijed definiranoj distribuciji (npr. normalnoj, distribuciji utvrđenoj u kojemu ranijem sličnom istraživanju ili teorijski pretpostavljenoj distribuciji). Takve je testove važno istaknuti i zato što se na njima temelji provjera ispunjenosti uvjeta normalnosti distribucije (v. potpoglavlje 4.3.2). Oni u dijagram sa slike 40 nisu uključeni zbog rijetke uporabe izvan konteksta testiranja uvjeta.

Ako su prisutne nezavisne varijable, što je u većini istraživanja slučaj, u sljedećemu je koraku važna **vrsta nezavisnih varijabli** – jesu li one numeričke, kategorijske ili su prisutna oba tipa (u višefaktorskim nacrtima). Naprimjer test koji ćemo odabrati bit će različit ako smo razinu općega znanja drugoga jezika izrazili putem kategorija (početna, srednja, visoka ili A1–C2 Zajedničkoga europskog referentnog okvira za jezike) i ako smo istu varijablu odlučili prikazati kontinuirano, putem broja bodova na odgovarajućemu testu. Ako je u istraživanje uključena jedna kategorijska nezavisna varijabla, potrebno je uzeti u obzir i **broj razina nezavisne varijable**, pri čemu je u tome slučaju važno postoje li dvije razine ili više njih. Izbor je testa dakle drukčiji kad uspoređujemo samo jednu skupinu učenika i skupinu izvornih govornika i kad uz izvorne govornike uključimo dvije ili tri skupine učenika.

Nakon svih koraka vezanih uz broj i vrstu varijabli i broj njihovih razina, posljednji važan čimbenik u izboru testa na koji možemo u potpunosti utjecati unaprijed, prilikom pripreme istraživanja, jest vrsta nacrti s obzirom na **vrste uzoraka**, odnosno na to jesu li uzorci koje treba analizirati nezavisni ili zavisni. U jednofaktorska istraživanja može biti uključena samo jedna od tih dviju vrsta uzoraka, ali je kod višefaktorskih nacrti moguće imati i obje istodobno (usp. potpoglavlje 2.1.5). Ta je podjela uzoraka i testova matematički vrlo važna zbog razlike u varijabilnosti – u zavisnim uzorcima jedan dio varijabilnosti možemo lako identificirati i ukloniti iz analize zahvaljujući tome što isti ispitanici sudjeluju u mjerenjima za različite razine nezavisne varijable, dok je u nezavisnim uzorcima vrlo prisutna i mora se uzeti u obzir i individualna varijabilnost.¹⁰⁹ Detaljnije ćemo o tome problemu govoriti u potpoglavlju 5.2.

¹⁰⁹ Ne treba zaboraviti da zavisni uzorci nose sa sobom drukčije probleme, vezane uz efekte učenja tijekom eksperimenta (v. Kraš i Miličević, 2015, str. 27–28).

Povezujući različite kriterije, možemo uočiti da na dijagramu sa slike 40, posebice u njegovu gornjem dijelu, dominiraju nacrti varijance te da za njih postoji najviše različitih statističkih testova (v. potpoglavlja 5.2 i 5.3). U središnjemu dijelu kod kombinacija numeričkih varijabli nalazimo korelacijske nacрте, s numeričkim varijablama (v. potpoglavlje 5.4), a nešto niže i frekvencijske, kod kojih su i zavisna i nezavisna varijabla kategorijske prirode (v. potpoglavlje 5.1). Možemo također na više mjesta uočiti i regresiju, koja ima najviše podvrsta prikladnih za različite kombinacije zavisnih i nezavisnih varijabli, kako u pogledu broja, tako i u pogledu vrste varijabli (v. potpoglavlje 5.5). Kao u potpoglavlju 3.2 kod grafičkih prikaza, u petome ćemo poglavlju strukturirati prikaz testova upravo prema nacrtima za koje se koriste.

4.3.2 Parametrijski uvjeti

Dodatni korak u izboru statističkoga testa tiče se izbora između **parametrijskih** i **neparametrijskih testova** (engl. *parametric vs. non-parametric tests*). Taj je izbor vezan uz značajke podataka koji se analiziraju, s jedne strane s obzirom na mjernu ljestvicu na kojoj su izražene vrijednosti zavisne varijable (što znamo i prije provedbe analize, a tiče se prije svega toga je li varijabla ordinalne prirode ili je intervalna ili omjerna), a s druge i s obzirom na to jesu li podaci koji su izmjereni na intervalnoj ili omjernoj ljestvici normalno distribuirani.¹¹⁰ Kako možemo vidjeti na slici 40, za neke je testove to nužan uvjet, dok u nekim slučajevima imamo izbor – za velik broj nacрта postoje paralelni parametrijski i neparametrijski testovi. Ti su nam uvjeti također već poznati iz trećega poglavlja, gdje je više puta naglašeno da oni moraju biti ispunjeni ako izračunavamo aritmetičku sredinu, standardnu devijaciju i druge s njima povezane mjere. Kako se vidi iz naslova ovoga potpoglavlja, možemo o njima skupno govoriti kao o **parametrijskim uvjetima** (engl. *parametric assumptions*). Testovi za nacрте koji uključuju kategorijsku zavisnu varijablu načelno se smatraju neparametrijskima ili testovima za koje podjela na parametrijske i neparametrijske testove nije relevantna (može se primijetiti da na slici 40 za te nacрте nije dan odgovor na pitanje ispunjenosti parametrijskih uvjeta). Situacija je zapravo složenija, ali se ovdje njome nećemo dalje baviti jer odstupa od razmatranja zasnovanih na normalnoj distribuciji.

Parametrijski statistički testovi nazivaju se tako jer se temelje na procjeni parametara, tj. (nepoznatih) vrijednosti različitih značajki populacije, na temelju uzorka. Naprimjer uspoređivanjem prosječnih rezultata dviju skupina učenika

¹¹⁰ Gries (2013a, str. 319) i Johnson (2013, str. 294–295) navode da su dopušteni još neki tipovi distribucije podataka, slično normalnoj distribuciji (t , F i χ^2 , v. peto poglavlje).

želimo utvrditi razlikuju li se populacijski prosjeci. Kako procjena parametara ovisi o značajkama distribucije podataka, potrebno je poznavati tu distribuciju, što najčešće znači da treba postojati normalna distribucija. Za parametrijske se testove ponekad kaže da su relativno otporni na odstupanja od normalnosti distribucije, posebice ako ta odstupanja nisu prevelika. Međutim treba biti vrlo oprezan kod odlučivanja o takvim ustupcima jer se jednako često tvrdi da parametrijski testovi nisu robusni. Između ostaloga, uporaba parametrijskih testova kod podataka koji nisu normalno distribuirani, isto kao i primjena neparametrijskih testova kad podaci jesu normalno distribuirani, dovodi do umanjene statističke snage testa (Larson-Hall 2012b; v. i potpoglavlje 4.1.5).¹¹¹

Neparametrijski se testovi u pravilu ne temelje na procjeni populacijskih parametara, već testiraju hipoteze o populacijama (naprimjer pretpostavku da su rezultati dviju populacija učenika različiti). Oni se često nazivaju i „slobodnim od distribucije“ (engl. *distribution free*) jer nisu ograničeni normalnošću distribucije. Upotrebljavaju se u analizi podataka koji su ordinalne prirode, kao i u analizi intervalnih i omjernih podataka čija distribucija nije normalna. Njihova se primjena također često savjetuje u slučajevima kad se analiziraju mali uzorci i uzorci koji sadrže netipične vrijednosti. Neparametrijski testovi ostvaruju manju statističku snagu od parametrijskih, tj. kad se može primijeniti i jedna i druga vrsta testova, neparametrijski u manjemu broju slučajeva otkrivaju efekte istraživanih varijabli. No kad uvjeti za parametrijske testove nisu ispunjeni, te prednosti više nema, pa se svakako moraju upotrebljavati neparametrijske opcije – kod jezičnih je podataka to vrlo čest slučaj (za detaljniju raspravu i smjernice u pogledu neparametrijskih testova v. Egbert i LaFlair, 2018).

Osnovna razlika u postupku izračunavanja odgovarajućih testnih statistika kod parametrijskih i neparametrijskih testova u tome je što neparametrijski testovi nisu neposredno utemeljeni na konkretnim vrijednostima podataka i mjerama poput aritmetičke sredine, varijance i standardne devijacije, već na **rangovima** (engl. *ranks*).¹¹² Njihov se način izračunavanja izravno nadovezuje na način određivanja položajnih mjera koje se koriste za opis podataka čija analiza zahtijeva neparametrijske testove. Razlog za

¹¹¹ Lindstromberg (2020) u usvajanju drugoga jezika, a Field i Wilcox (2017) šire, ističu kako u slučajevima odstupanja od uvjeta normalnosti distribucije parametrijski testovi bolje funkcioniraju kod distribucija koje su (blago) asimetrične, ali nemaju debeo rep (gdje rep čine vrijednosti udaljene od prosjeka), nego kod distribucija s debelim repom.

¹¹² Pojedini neparametrijski testovi čak su utemeljeni na istim ili sličnim matematičkim operacijama kao i njihovi parametrijski parnjaci, ali se te operacije ne provode na izvornim vrijednostima podataka već na njihovim rangovima.

manju informativnost tih testova, pa time i razlog za njihovo izbjegavanje u slučajevima kad nisu nužan izbor, upravo je u tome što zanemaruju točne vrijednosti dobivenih rezultata i uzimaju u obzir samo njihov redoslijed, čime jedan dio raspoloživih matematičkih informacija ostaje neiskorišten. Međutim ti testovi iz istoga razloga predstavljaju i dragocjeno sredstvo analize kada uvjeti za parametrijske testove nisu ispunjeni i mjere poput aritmetičke sredine i varijance nisu primjerene za opis uzorka.¹¹³ Rangiranje podataka, kao što je već objašnjeno u trećemu poglavlju, podrazumijeva svrstavanje konkretnih vrijednosti u uređeni niz i dodjeljivanje ranga od 1 do n svakoj pojedinačnoj vrijednosti. U slučaju ordinalnih ljestvica podaci su često već po svojoj prirodi rangovi te dodatne konkretne vrijednosti i ne postoje. Važne mjere koje se (ovisno o testu) mogu upotrebljavati u daljnjoj analizi, pored medijana, jesu i zbrojevi rangova i prosjeci rangova.

Kad su podaci izmjereni na intervalnoj ili omjernoj ljestvici, kako bi se mogao napraviti izbor između dviju vrsta testova, potrebno je, naravno, provjeriti jesu li ti podaci normalno distribuirani. Okvirno se distribucija podataka može procijeniti putem histograma ili drugih grafikona opisanih u trećemu poglavlju, ali u takvu procjenu ne možemo biti posve sigurni te je potrebno na podacima napraviti namjenski test. Dva neparametrijska testa koja se najčešće upotrebljavaju u tu svrhu jesu **Kolmogorov-Smirnov test** (čiji je statistik KS) i **Shapiro-Wilk test** (čiji je statistik W). Ti testovi pripadaju skupini testova prikladnosti modela jer uspoređuju distribuciju podataka s normalnom, odnosno testiraju nultu hipotezu da uzorak dolazi iz normalno distribuirane populacije. Kad oni daju statistički značajan rezultat, to znači da distribucija vjerojatno nije normalna (tj. značajno se razlikuje od normalne), te je potrebno primijeniti neparametrijske testove (ili eventualno transformirati podatke, v. potpoglavlje 3.1.6). Ponekad se može dogoditi da ti testovi daju statistički značajan rezultat i onda kada distribucija podataka na histogramu djeluje normalno – u takvim situacijama treba biti posebno oprezan. Načelno, za velike uzorke često će biti opravdano popratiti test analizom histograma te konačnu odluku donijeti na temelju obaju izvora.¹¹⁴ Kod manjih uzoraka

¹¹³ U analizi ordinalnih podataka ili podataka koji nisu normalno distribuirani nije rijetko da se uz neparametrijske testove kao deskriptivni podatak navedu aritmetička sredina i standardna devijacija, ponekad radi dobivanja „granularnijih“ podataka (jer bi za podatke poput prosudbi prihvatljivosti rečenice medijan mogao dovesti do velikoga broja ponavljanja istih vrijednosti ako ljestvica prosudbe ima relativno nizak broj razina, poput četiri ili pet). Međutim tu praksu treba što više izbjegavati jer te mjere ne pokazuju centralnu tendenciju i varijabilnost adekvatno ako nisu ispunjeni uvjeti za njihovo izračunavanje (v. npr. Al-Hoorie i Vitta, 2019).

¹¹⁴ To posebice vrijedi za Kolmogorov-Smirnov test, koji može biti odviše strog za veće uzorke te proglasiti i vrlo mala odstupanja od normalne distribucije značajnima.

takav pristup može biti rizičan zbog opće tendencije malih uzoraka da budu vrlo varijabilni. Naime naglasak je u ispitivanju normalnosti distribucije na distribuciji populacije, a odnos između normalnosti uzoraka i populacija nije nužno neposredan. Točnije, iz normalno se distribuiranih populacija mogu dobiti uzorci čija distribucija nije normalna, kao i obrnuto (Lindstromberg, 2020). Što je uzorak manji, to se manje možemo osloniti na distribuciju koju vidimo na histogramu kao procjenu populacijske distribucije.

Dodatno, ne treba zaboraviti ni važnost vizualizacije podataka u procjeni ispunjenosti uvjeta. Hu i Plonsky (2021, str. 179) naprimjer ističu da testovi normalnosti distribucije nisu savršeni te da im se lako može dogoditi pogrešna procjena (bilo da su previše konzervativni ili previše „popustljivi“). Stoga je zapravo najbolje uvijek zajedno sagledati grafički prikaz podataka i rezultat odgovarajućega testa.

Navedeni testovi konceptualno pripadaju široj skupini **testova ispunjenosti uvjeta** (engl. *assumption tests*). Hu i Plonsky (2021, str. 173) ističu da je upravo uvjet normalnosti distribucije onaj koji se najčešće testira. Međutim pojedini inferencijalni testovi, kako parametrijski tako i neparametrijski, imaju i druge specifične uvjete čiju ispunjenost zahtijevaju. O tim ćemo uvjetima više govoriti u petome poglavlju, kad budemo opisivali pojedine testove. Ovdje treba napomenuti da je vrlo važno provjeriti ispunjenost svih uvjeta (i navesti rezultate provjere u izvještaju o istraživanju, v. šesto poglavlje), kao i da kod nekih uvjeta u slučaju neispunjenosti nije nužno u potpunosti mijenjati test, već je moguće primijeniti odgovarajuću korekciju u načinu izračunavanja testnoga statistika (v. posebice potpoglavlja 5.1.1 i 5.2.1). Provjera normalnosti distribucije u tome je smislu jedan od najstrožih uvjeta.

Konačno, uz navedene korake, koji se obično prikazuju u dijagramima odlučivanja o statističkim testovima, potrebno je u odlučivanju voditi računa i o broju ispitanika. Još su Hatch i Lazaraton (1991) preporučile za područje primijenjene lingvistike da se parametrijski testovi koriste samo na uzorcima od najmanje 30 ispitanika, a slično navodi i Gries (2013a, str. 322). Larson-Hall (2012b) čak navodi da je teško procijeniti normalnost distribucije ako uzorak sadrži manje od 50 ispitanika. Larson-Hall (2016) te Hu i Plonsky (2021) također su među autorima koji naglašavaju da je kod malih uzoraka vrlo teško povući jasnu granicu između podataka koji jesu i nisu normalno distribuirani. Za manje uzorke načelno se dakle adekvatnijima smatraju neparametrijski testovi, koji daju pouzdane rezultate i s manjim brojem ispitanika. Međutim, budući da je u praksi u istraživanjima usvajanja drugoga jezika često vrlo teško ili čak nemoguće prikupiti 30 ili više ispitanika, literatura u tome području

obiluje primjerima istraživanja u kojima se parametrijski testovi upotrebljavaju u analizi manjih uzoraka (nekoliko metaanaliza navodi da su parametrijski testovi u području usvajanja drugoga jezika najkorišteniji testovi, v. Hu i Plonsky, 2021; Khany i Tazik, 2019; Lindstromberg, 2016b; Plonsky, 2013; Plonsky i Gass, 2011). Sve više autora stoga kao kompromis predlaže češću uporabu robusnih testova utemeljenih na uzorkovanju s povratom jer su ti testovi otporniji na odstupanja od normalnosti (v. potpoglavlje 5.6).

Za kraj ovoga poglavlja možemo još jednom istaknuti kao najvažnije da provjeru statističkih pretpostavki obavezno treba provesti i o njoj izvijestiti u izvještaju o istraživanju (v. potpoglavlja 6.4 i 6.5.2). Al-Hoorie i Vitta (2019) te Hu i Plonsky (2021) među autorima su koji ističu da ta praksa često izostaje u području usvajanja drugoga jezika i da to nikako ne bi smio biti slučaj.

5. Često korišteni statistički testovi

Radovi koji daju metodološki pregled područja usvajanja drugoga jezika (v. ponajprije Gass, 2009; Khany i Tazik, 2019; Lazaraton, 2000; Loewen i Gass, 2009; Plonsky, 2013) vrlo konzistentno pokazuju da su najčešće korišteni statistički testovi oni kojima se uspoređuju mjere centralne tendencije različitih uzoraka (najčešće različitih skupina ispitanika), pri čemu je na prvome mjestu analiza varijance, a na drugome *t*-testovi. U daleko su manjoj mjeri zastupljene korelacija i (linearna) regresija za kontinuirane varijable, kao i hi-kvadrat test za kategorijske varijable. Svi navedeni testovi ujedno pripadaju skupini „osnovnih“ testova, zajedno s deskriptivnom statistikom. Naime Khany i Tazik (2019) upotrebljavajući podjelu statističkih analiza (preuzetu iz Goodwin i Goodwin, 1985) na „osnovne“, „srednje“ i „napredne“, utvrđuju da se u području primijenjene lingvistike (točnije, u radovima u časopisima posvećenima usvajanju i poučavanju drugoga jezika), usprkos porastu zastupljenosti srednjih i naprednih analiza, i dalje najviše upotrebljavaju one osnovne (u približno 61 % slučajeva). Česta je i uporaba više različitih statističkih testova u istome istraživanju, ali su i tada testovi uglavnom osnovnoga tipa. Nekoliko metaanaliza (Gass, 2009; Plonsky, 2013; Plonsky i Gass, 2011) upućuje na porast broja istraživanja u kojima se upotrebljava više od jednoga testa kao na negativnu tendenciju. Plonsky (2013) naprimjer utvrđuje da je medijan broja testova u jednome objavljenom radu iz područja usvajanja drugoga jezika 18, što se može negativno odraziti na ukupnu statističku snagu, koja bi bila veća kad bi podaci bili objedinjeni u manjemu broju složenijih analiza.

Pozivi da se trend pretežite primjene najjednostavnijih testova promijeni, sve su češći, između ostaloga i zbog opće težnje k povećanju uporabe robusnih testova i testova koji nisu utemeljeni na testiranju nulte hipoteze.¹¹⁵ Usprkos tome, poznavanje tih testova smatramo i dalje veoma važnim ne samo zbog njihove dosadašnje česte uporabe te mogućnosti razumijevanja postojeće literature već i zato što je njihovo razumijevanje korisno, ako ne i nužno, za razumijevanje naprednijih statističkih metoda. U ovome se poglavlju stoga usredotočujemo na često korištene „tradicionalne“ vrste testova, nakon čega dodajemo kraći prikaz nekih od naprednijih testova, koji se i u području usvajanja drugoga jezika sve češće upotrebljavaju. Svojevrsan „most“ između dobro utemeljenih i nešto novijih testova predstavlja regresija, koja pripada

¹¹⁵ U ovome poglavlju osvrtaćemo se više puta i na značajke podataka u istraživanjima usvajanja drugoga jezika zbog kojih je dodatno poželjno u tome području više se oslanjati na složenije testove koji uzimaju u obzir više čimbenika istodobno.

ustaljenim analizama, ali ima velik broj varijanti za različite nacрте istraživanja i može se proširiti u mješovite modele, koji dopuštaju analizu slučajnih efekata ispitanika i čestica te time pružaju više informacija od većine drugih testova.

U prikazu grupiramo odabrane testove prema osnovnim značajkama nacрта istraživanja za koje se upotrebljavaju. Polazimo od testova koji se primjenjuju kod frekvencijskih nacрта s dvjema kategorijskim varijablama, zatim govorimo o nacrtima varijance ili faktorijalnim nacrtima, s dvije ili više razina nezavisne varijable (tj. dva ili više uzorka koji se uspoređuju) i s jednom ili više različitih nezavisnih varijabli,¹¹⁶ te o korelacijskim nacrtima s dvjema numeričkim varijablama, na koje dodajemo i uvodna razmatranja regresijskih nacрта i mješovitih modela. Ondje gdje je to primjereno obuhvaćamo i parametrijske i neparametrijske testove za isti nacrt. Drugim riječima, kao u dijagramu izbora testa prikazanome na slici 40 (v. potpoglavlje 4.3), polazimo od nacрта istraživanja definiranih prema vrsti varijabli kao od glavnoga kriterija za podjelu testova, unutar čega zatim odvajamo parametrijske i neparametrijske testove.

U naslovima se potpoglavlja svugdje do regresije pozivamo na „povezanost“ varijabli jer to smatramo zajedničkom značajkom kojom se mogu opisati svi nacrti, kao i dobrim opisom situacija kakve se u lingvističkim pristupima usvajanju drugoga jezika proučavaju. Naime, kako je već navedeno u potpoglavlju 2.1.5, prema razini kontrole, nacrti istraživanja u tome području najčešće su kvaziekperimentalni, tj. zasnovani na selektivnim varijablama (a ne na manipulativnim). U njima stoga nije lako utvrditi uzročno-posljedične veze između nezavisnih i zavisnih varijabli, već se radije govori o tome kako su varijable povezane (eventualno kako jedne predviđaju druge). To posebno napominjemo jer je u statističkim priručnicima uobičajeno da se „povezanost“ odnosi prije svega na analizu frekvencijskih i korelacijskih nacрта, koji su osobito česti kod podataka prikupljenih neeksperimentalnim metodama (v. npr. Levshina, 2015).¹¹⁷ O analizi nacрта varijance uglavnom se govori kao o testiranju razlika između skupina (npr. Levshina, 2015) ili kao o uspoređivanju aritmetičkih sredina (npr. Petz i sur., 2012), gdje drugi način referiranja

¹¹⁶ Todorović (2008, str. 136) prema broju razina definira kategorijske nezavisne varijable, pa time i pripadajuće nacрте, kao „bivalentne“ i „multivalentne“. Iako nam se takvi nazivi čine vrlo praktičnima, proučavanje literature pokazalo je da su oni ograničeni na područje psihologije i literaturu na srpskome jeziku, zbog čega zadržavamo nazive u kojima se eksplicitno govori o broju razina.

¹¹⁷ Na engleskome se jeziku susreće i razlika između povezanosti kao „*relationship*“ za numeričke i „*association*“ za kategorijske varijable (v. npr. Levshina, 2015). Čini se međutim da ne postoji jasna definicija razlike između tih dvaju pojmova (v. npr. raspravu na stranici <https://stats.stackexchange.com/questions/371747/association-relationship-and-correlation>, posljednji pristup 26. 4. 2024.).

podrazumijeva prethodno odvajanje parametrijskih i neparametrijskih testova, tj. grupiranje testova najprije prema kriteriju (ne)ispunjenosti parametrijskih uvjeta, pa tek zatim prema nacrtu; među iznimkama je Todorović (2008), koji govori uopćenije o analizi nacrtu varijance. Nama se čini poželjnim konzistentno govoriti o povezanosti iz više razloga, među kojima je glavni taj da nju podrazumijevaju čak i uzročno-posljedični odnosi te da ona neposredno opisuje metodološku bit istraživanja, koja je vezana upravo uz odnos nezavisnih i zavisnih varijabli (usp. potpoglavlje 2.1.6, gdje se govori o valjanosti statističkoga zaključka, koja se odnosi upravo na povezanost varijabli kao preduvjet uzročno-posljedičnih odnosa između njih). Dodatno je na odabir utjecalo to što testove grupiramo primarno prema nacrtu i pod zajedničkim naslovom pokrивamo parametrijske i neparametrijske testove, zbog čega želimo izbjeći referiranje na isključivo parametrijske pojmove (poput aritmetičke sredine i varijance).

Testove opisujemo objašnjavajući njihov kontekst uporabe i način izračunavanja. Iako za istraživače usvajanja drugoga jezika najčešće nije nužno poznavati matematičke formule koje su u temeljima statističkih analiza – za njih su od veće važnosti poznavanje prirode podataka koje žele analizirati i jasna ideja o tome što analizom podataka žele utvrditi¹¹⁸ – uvid u to koje su vrijednosti uključene u izračunavanje rezultata testova doprinosi boljem razumijevanju kako podataka tako i rezultata. Opisi različitih testova nisu podjednako detaljni, zbog toga što primarni cilj knjige nije udžbenički prikaz već analiza i kritičko promišljanje osnovnih općih pojmova i načela u statističkoj analizi, zbog čega nema razloga da pojmovi i načela objašnjeni u vezi s jednim testom dobiju jednaku pozornost i u kontekstu drugih testova. S druge strane, ne objašnjavamo provedbu testova u različitim programima za statističku analizu jer su ti postupci više tehnički nego znanstveni i već im je posvećen vrlo velik broj namjenskih knjiga i/ili članaka (niz korisnih izvora o različitim programima citiran je u ranijim poglavljima, posebno prvome i drugome). Napominjemo i da se u dijelovima teksta u kojima se opisuju testovi i njihove formule vrlo rijetko referiramo na izvore jer se najvećim dijelom radi o općim znanjima, kakva se pronalaze u (gotovo) svim udžbenicima i priručnicima iz područja statistike, te bi ih bez temeljitoga proučavanja povijesti teško bilo pripisati pojedinim autorima i djelima. Naša se opća znanja iznesena u ovome poglavlju pretežito zasnivaju na sljedećim knjigama: Field (2018); Field i sur. (2012); Gries (2013b); Levshina (2015); Petz i sur. (2012); Rasinger (2013);

¹¹⁸ Kako navode Loewen i Gass (2009, str. 181), u području usvajanja drugoga jezika „statistika je sredstvo, a ne predmet istraživanja“.

Todorović (2008). Pozivanja na literaturu prisutna su ondje gdje se radi o specifičnim ili posebno aktualnim razmatranjima.

5.1 Ispitivanje povezanosti između kategorijskih varijabli

Ovo potpoglavlje govori o frekvencijskim nacrtima istraživanja. Glavna je značajka takvih nacrti to da su sve varijable koje se u njima javljaju kategorijske, dok je frekvencija osnovna statistička mjera. Takvi su nacrti češći u korpusnim nego u eksperimentalnim istraživanjima, ponajprije zbog toga što se u eksperimentima bilježe individualni odgovori ispitanika, koji tipično dovode do primjene nacrti varijance. Međutim oni se mogu javiti i u nekome aspektu analize eksperimentalnih podataka.

Frekvencijski nacrti u istraživanjima jezika najčešće uključuju dvije varijable, od kojih svaka ima dvije ili više razina. Kako smo vidjeli u potpoglavlju 3.1.1 (i ponovno ćemo vidjeti u potpoglavlju 5.1.1), podaci takve prirode obično se prikazuju u kontingencijskim tablicama, u kojima se dvije varijable križaju i izračunava se frekvencija svake od dobivenih situacija. U pogledu kodiranja i čuvanja podataka, u pojedinim su slučajevima takve tablice format u koji se frekvencijski podaci unose neposredno. Tako bi bilo naprimjer kad bismo prebrojili frekvencije dviju konstrukcija (poput aktiva i pasiva) u dvama različitim korpusima (poput korpusa proizvodnje učenika drugoga jezika i korpusa proizvodnje izvornih govornika), pri čemu bi korpus predstavljao jedinicu analize. To je čest slučaj kad se u istraživanju upotrebljavaju korpusi u kojima nemamo uvid u manje jedinice, poput pojedinačnih tekstova ili tekstova pojedinačnih autora. Druga je situacija ona u kojoj je jedinica analize pojedinačni tekst, pojedinačni ispitanik ili pojedinačno opažanje te za svako opažanje imamo izmjerene vrijednosti varijabli koje se ispituju, a kontingencijska tablica se zatim dobiva naknadnim brojenjem ukrižanih frekvencija tih vrijednosti. Naprimjer možemo pretpostaviti da je u već citiranome istraživanju leksičkoga vida glagola u engleskome jeziku i njihove uporabe u različitim morfološkim oblicima kod jednoga učenika početna tablica izgledala slično tablici 11 (izrađenoj na temelju primjera rečenica u radu Robison, 1990; v. i potpoglavlje 3.1.1); tj. da je za svaki glagol koji je učenik upotrijebio bilo označeno je li statični ili dinamični i je li upotrijebljen s *-ing*, s *-EN* (prošlo vrijeme) ili u osnovnome obliku. Podaci iz kontingencijskih tablica mogu se uz to prikazati putem mozaičnoga grafikona, gdje se vide i distribucija frekvencija po različitim kategorijama i eventualne razlike u veličini uzoraka (v. potpoglavlje 3.2.1).

Tablica 11. Prikaz podataka o dvjema kategorijskim varijablama

Glagolski oblik	Leksički vid	Morfološki oblik
<i>told</i>	dinamični	-EN
<i>want</i>	statični	osnovni
<i>studied</i>	dinamični	-EN
<i>tried</i>	dinamični	-EN
<i>havin'</i>	statični	-ing
...

Dodatno se može izdvojiti frekvencijski nacrt sa samo jednom varijablom. Takav bismo nacrt mogli imati naprimjer ako bismo htjeli provjeriti upotrebljavaju li učenici engleskoga aktiv i pasiv u jednakoj mjeri. Tu ne bismo dakle uspoređivali učenike s izvornim govornicima, odnosno izostavili bismo nezavisnu varijablu, već bismo se fokusirali samo na konstrukciju kao zavisnu varijablu te učestalost aktiva i pasiva u proizvodnji učenika usporedili s teorijskom distribucijom tih konstrukcija od 50 % naprama 50 %. Ta situacija, kao ni drugi nacrti bez nezavisne varijable, nije obuhvaćena na dijagramu na slici 40.

5.1.1 Hi-kvadrat test

Kad su uzorci nezavisni, za analizu oba prethodno navedena nacrtu upotrebljava se test koji se uopćeno može nazvati **Pearsonov hi-kvadrat test** (engl. *Pearson's chi-square(d) test*). Točnije, u pitanju je skupina srodnih testova, od kojih se **hi-kvadrat test prikladnosti modela** (engl. *chi-square goodness of fit test*) upotrebljava za nacрте s jednom, a **hi-kvadrat test nezavisnosti** (engl. *chi-square test of independence*) za nacрте s dvije varijable. Testom nezavisnosti mogu se ispitati hipoteze o povezanosti dviju kategorijskih varijabli, dok se naziv testa odnosi na to što su dvije varijable u slučaju da vrijedi nulta hipoteza međusobno nezavisne. U oba se slučaja izračunava isti testni statistik, hi-kvadrat (označen simbolom grčkoga slova χ^2), prema formuli navedenoj u (11), u kojoj F_o označava opažene frekvencije, dok F_t predstavlja teorijske frekvencije, odnosno frekvencije koje su očekivane pod nultom hipotezom.

$$(11) \chi^2 = \sum \frac{(F_o - F_t)^2}{F_t}$$

Hi-kvadrat test jedan je od statističkih testova koji se vrlo lako mogu provesti i ručno – potrebno je samo izračunati teorijske frekvencije, odnosno frekvencije koje bi odgovarale nultoj hipotezi, i usporediti s njima frekvencije doista opažene u istraživanju. Kod testa nezavisnosti, na temelju opaženih frekvencija mogu se izračunati teorijske frekvencije prema formuli **(zbroj retka * zbroj stupca)/ukupni zbroj**. To najbolje možemo pokazati na primjeru, te u donjim tablicama 12 i 13 ponavljamo kontingencijsku tablicu prikazanu u trećemu poglavlju (tablica 5; v. potpoglavlje 3.1.1) i dopunjavamo je **marginalnim frekvencijama** (engl. *marginal frequencies*), koje vidimo pod „Ukupno“ i koje su nam nužne za izračunavanje teorijskih frekvencija.

Tablica 12. Kontingencijska tablica opaženih frekvencija (prema Robison, 1990)

		Vrsta glagola prema leksičkome vidu		UKUPNO
		Statični	Dinamični	
Morfološki oblik glagola	osnovni oblik	137	301	438
	-ing	39	48	87
	-EN	0	28	28
UKUPNO		176	377	553

Tablica 13. Kontingencijska tablica teorijskih frekvencija (prema Robison, 1990)

		Vrsta glagola prema leksičkome vidu		UKUPNO
		Statični	Dinamični	
Morfološki oblik glagola	osnovni oblik	$\frac{438 * 176}{553} = 139,4$	$\frac{438 * 377}{553} = 298,6$	438
	-ing	$\frac{87 * 176}{553} = 27,7$	$\frac{87 * 377}{553} = 59,3$	87
	-EN	$\frac{28 * 176}{553} = 8,9$	$\frac{28 * 377}{553} = 19,1$	28
UKUPNO		176	377	553

Zatim možemo sve vrijednosti unijeti u formulu kao u (12) te izračunati hi-kvadrat, koji iznosi $\chi^2=19,9$:

$$(12) \chi^2 = \frac{(137-139,4)^2}{139,4} + \frac{(301-298,6)^2}{298,6} + \frac{(39-27,7)^2}{27,7} + \frac{(48-59,3)^2}{59,3} + \frac{(0-8,9)^2}{8,9} + \frac{(28-19,1)^2}{19,1}$$

Na sličan se način može izračunati i vrijednost hi-kvadrata za test prikladnosti modela. U tome se slučaju teorijske vrijednosti izračunavaju na temelju teorijskih pretpostavki ili prethodne literature. Naprimjer želimo li doznati upotrebljava li učenik drugoga jezika iz istraživanja Robison (1990) uz dinamične glagole češće glagolske oblike na *-EN* nego oblike na *-ing*, pretpostavit ćemo za teorijske vrijednosti koje odgovaraju nultoj hipotezi jednaku frekvenciju za oba oblika, tj. relativnu frekvenciju od 0,50.

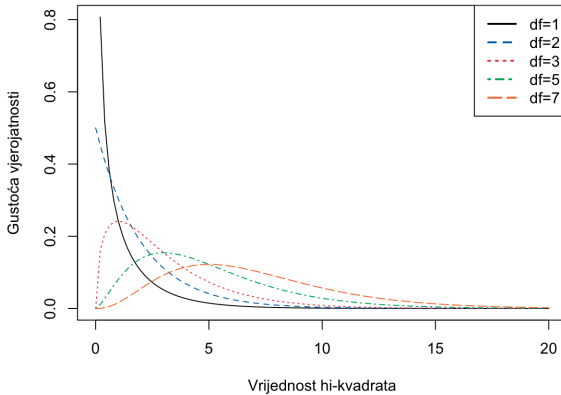
Za utvrđivanje je li dobiveni rezultat statistički značajan, dodatno je uz vrijednost hi-kvadrata potrebno utvrditi broj stupnjeva slobode, prema formuli **(broj redaka–1)*(broj stupaca–1)**. U iznesenom primjeru dakle taj bi broj iznosio $2*1$, tj. $df=2$.¹¹⁹ Kad poznamo te dvije vrijednosti, možemo utvrditi je li rezultat koji smo dobili statistički značajan tako što ćemo potražiti odgovarajuću kritičnu vrijednost u tablici kritičnih vrijednosti hi-kvadrata. Ako je dobivena vrijednost hi-kvadrata veća od kritične, rezultat je statistički značajan. Za prethodno navedenu vrijednost, koja iznosi 19,9, možemo uočiti da za 2 stupnja slobode predstavlja statistički značajan rezultat.

Tablica 14. Dio kritičnih vrijednosti hi-kvadrata

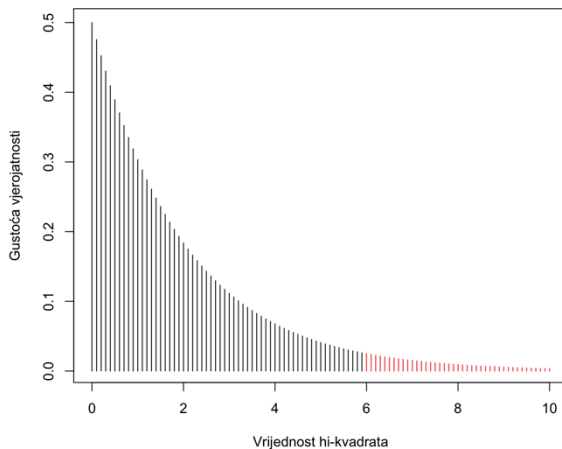
	$\alpha=0,05$	$\alpha=0,01$
$df=1$	3,84	6,63
$df=2$	5,99	9,21
$df=3$	7,81	11,34
$df=4$	9,49	13,28
$df=5$	11,07	15,09
$df=6$	12,59	16,81
$df=7$	14,07	18,48

Tablica 14, koja je samo jedan isječak opsežnije tablice koja obuhvaća više vrijednosti stupnjeva slobode i više razina statističke značajnosti, temelji se na hi-kvadrat distribuciji vjerojatnosti, koja je u potpunosti definirana brojem stupnjeva slobode. Na slici 41 može se vidjeti usporedni prikaz distribucija za 1, 2, 3, 5 i 7 stupnjeva slobode, a na slici 42 distribucija za dva stupnja, s označenom kritičnom regijom. Iako je ta regija u potpunosti s desne strane distribucije, hi-kvadrat test uvijek se tumači kao dvosmjernan.

¹¹⁹ Engleski je termin za stupnjeve slobode *degrees of freedom* (v. i potpoglavlje 4.2.1), što se obično skraćuje u *df*. Kako je ta kratica vrlo raširena u statističkoj literaturi, a ujedno i jednoznačna u odnosu na *ss*, što bi se moglo izvesti iz prijevodnoga ekvivalenta „stupnjevi slobode“, u daljnjem ćemo je tekstu nastaviti upotrebljavati.



Slika 41. Hi-kvadrat distribucija za nekoliko različitih stupnjeva slobode (df)



Slika 42. Hi-kvadrat distribucija s kritičnom regijom za $df=2$

Pri radu u statističkim programima svi navedeni podaci dobivaju se automatski. Dodatno, statistički programi obično daju podatak o standardiziranim rezidualnim odstupanjima, koja pokazuju koja polja u tablici posebice utječu na dobivene rezultate. Ta su odstupanja već uvedena u potpoglavlju 3.2.1 kod prikaza mozaičnih i asocijativnih grafikona. Za njihovo bolje razumijevanje dodajemo da polje ne pridonosi značajno rezultatima ako se dobivena vrijednost reziduala kreće između -2 i $+2$. Kod vrijednosti koje ne pripadaju tome opsegu, polja imaju značajnoga utjecaja na rezultate. U tablici 15 vidimo reziduale za podatke iz istraživanja koje je proveo Robison (1990). Možemo uočiti koji se oblici upotrebljavaju više no što je očekivano

pod nultom hipotezom (pozitivni predznak), a koji manje (negativni predznak), te koja imaju posebno značajan utjecaj na rezultat hi-kvadrat testa: oblici s *-ing* češći su no što je očekivano uz statične glagole, oblici s *-EN* češći su od očekivanoga uz dinamične, a rjeđi od očekivanoga uz statične glagole. Ti su reziduali prikazani i na slikama 25 i 26 u potpoglavlju 3.2.1.

Tablica 15. Standardizirani reziduali za tablicu 12 (prema Robison, 1990)

		Vrsta glagola prema leksičkome vidu	
		Statični	Dinamični
Morfološki oblik glagola	osnovni oblik	-0,2	+0,14
	<i>-ing</i>	+2,15	-1,47
	<i>-EN</i>	-2,99	+2,04

Uz hi-kvadrat test kao mjera veličine efekta izračunava se **koeficijent phi** (engl. *phi coefficient*, ϕ , samo za tablice 2 x 2) ili **Cramérov koeficijent V** (engl. *Cramér's V*, za sve ostale tablice).¹²⁰ Formule za izračunavanje tih koeficijenata prikazane su u (13) i (14). U njima se pojavljuje vrijednost hi-kvadrata dobivena u testu, kao i N – ukupni broj opažanja u istraživanju (553 u gornjim tablicama), k – broj stupaca u tablici i r – broj redaka u tablici.

$$(13) \phi = \sqrt{\frac{\chi^2}{N}}$$

$$(14) V = \sqrt{\frac{\frac{\chi^2}{N}}{\min(k-1, r-1)}}$$

Uobičajena je procjena dobivenoga rezultata da koeficijenti čija je vrijednost niska i iznosi približno 0,10 označavaju mali efekt, koeficijenti vrijednosti približno 0,30 srednji efekt, a koeficijenti vrijednosti oko 0,50 jak efekt. Međutim, kako je već istaknuto u potpoglavlju 4.1.3, u različitim znanstvenim područjima, pa i užim temama unutar istoga područja, tumačenja se mogu razlikovati zbog razlika u vrsti podataka i u tipičnim rezultatima koji se dobivaju, stoga treba dodatno potražiti smjernice za tumačenje dobivenih vrijednosti u već objavljenim radovima.

¹²⁰ Za tablice 2 x 2 Cramérov koeficijent V daje isti rezultat kao i koeficijent phi, kojemu se daje prednost jer se lakše izračunava.

U vezi s hi-kvadrat testom treba napomenuti da se on smatra doista pouzdanim ako u tablici nema teorijskih frekvencija nižih od pet i ako ukupni zbroj frekvencija nije manji od 30. Također, tablice formata 2 x 2 smatraju se problematičnima za uobičajenu formulu (precjenjuju značajnost rezultata), te se za njihovu analizu upotrebljava i tzv. **Yatesova korekcija** (engl. *Yates' correction*), posebice u slučajevima kad su prisutne niske teorijske frekvencije. Ta korekcija predstavlja posebno prilagođavanje formule oduzimanjem 0,5 od svake razlike između opažene i teorijske frekvencije (većina statističkih programa automatski primjenjuje tu korekciju). Budući da njezina primjena često dovodi do suprotnoga problema, tj. podcjenjivanja značajnosti rezultata, ponekad se savjetuje da se umjesto hi-kvadrat testa na tablicama 2 x 2 provede **Fisherov egzaktni test** (engl. *Fisher's exact test*), koji je nastao u kontekstu eksperimenta s mlijekom i čajem koji smo opisali u potpoglavlju 4.1.2. Taj je test posebno prikladan za niske frekvencije.¹²¹ Među drugim testovima o kojima nećemo detaljnije govoriti, ali koji mogu biti korisni u situacijama kad hi-kvadrat test nije odgovarajući izbor, možemo spomenuti **G-testove** (engl. *G-tests*) iz skupine **testova vjerodostojnosti** (engl. *likelihood-ratio tests*). Također, kad je nezavisnih varijabli više, provodi se **loglinearna analiza** (engl. *log-linear analysis*).

Kao rezultat hi-kvadrat testa obavezno se u izvještaju o istraživanju navode vrijednost hi-kvadrata, stupnjevi slobode (obično u zagradi uz χ^2) i vrijednost p , kao i veličina efekta. Naprimjer navedeni bi se rezultat mogao prikazati ovako: „Hi-kvadrat testom utvrđeno je da postoji statistički značajna povezanost između leksičkoga vida glagola i morfoloških oblika u kojima ih upotrebljava istraživani učenik engleskoga kao drugoga jezika, $\chi^2(2)=19,9$, $p<0,01$. Veličina efekta je relativno mala, $V=0,1897$, pa možemo zaključiti da efekt nema veliku praktičnu važnost.“¹²² Kao dio rezultata treba, svakako, uključiti i opis podataka (kontingencijsku tablicu) i njihov grafički prikaz, naprimjer na mozaičnome ili asocijativnome grafikonu.

Za kraj treba posebno istaknuti da se hi-kvadrat test ne može provoditi na relativnim frekvencijama, za koje postoje posebni postupci, kao ni na normaliziranim frekvencijama, već isključivo na apsolutnima. Također

¹²¹ Fisherov se test naziva „egzaktnim“ zbog toga što pokazuje točnu vrijednost p izračunatu primjenom pravila kombinatorike, a ne temelji se na aproksimaciji kao većina drugih statističkih testova. Zbog toga je kod visokih frekvencija vrlo zahtjevan za izračunavanje.

¹²² U tome primjeru ostavljamo $p<0,01$ umjesto navođenja točne vrijednosti p . Razlog je taj što smo ovdje test proveli ručno i utvrdili razinu značajnosti iz tablice kritičnih vrijednosti. U slučaju rezultata dobivenih putem statističkoga softvera jedini slučajevi u kojima je opravdano navesti samo razinu značajnosti jesu oni u kojima je p iznimno nizak, kad se uobičajeno navodi $p<0,001$.

podsjećamo još jedanput da je uvjet za odabir toga testa to da opažanja budu međusobno nezavisna, odnosno da uzorci koji sudjeluju u različitim razinama nezavisne varijable budu nezavisni (poput statičnih i dinamičnih glagola u navedenom primjeru; usp. potpoglavlje 3.1.1, kao i dijagram izbora testa na slici 40). Čak i kad je taj uvjet ispunjen, ono što dodatno treba razmotriti prije nego se hi-kvadrat test provede jesu i drugi mogući izvori zavisnosti u podacima. Naime primjena toga testa u korpusnoj lingvistici podosta se kritizira jer se u istraživanjima često izdvaja više riječi ili konstrukcija dobivenih iz istoga teksta, koje samim tim što su dio vezanoga teksta ne mogu biti potpuno međusobno neovisne (v. posebice Winter i Grice, 2021). I s obzirom na primjer koji je u ovome potpoglavlju prikazan moglo bi se postaviti slično pitanje: jesu li istraživane glagolske pojavnice doista međusobno neovisne? To što sve dolaze od istoga učenika može se u ovome slučaju zanemariti ako uopćavanje rezultata ograničimo na toga učenika. Međutim, budući da je uzorak prikupljen u intervjuu (Robison, 1990, str. 321), između različitih glagolskih pojavnica vrlo vjerojatno ne postoji potpuna nezavisnost, te je moguće da nam hi-kvadrat test ne daje potpuno pouzdan odgovor. U eksperimentalnim istraživanjima nacrti na kojima se provodi hi-kvadrat test u pravilu su rijetki, ali uz to što je u pitanju jedan od osnovnih statističkih testova, vidjet ćemo da se problem (ne)zavisnosti podataka u jezičnim istraživanjima pojavljuje i u opisu drugih testova.

5.1.2 McNemarov test

Ako su u istraživanju prisutni zavisni uzorci (s obzirom na vrijednosti zavisne varijable), odgovarajući je test **McNemarov test** (engl. *McNemar's test*). Njegov se testni statistik također označava kao χ^2 jer u nekim slučajevima njegova distribucija odgovara hi-kvadrat distribuciji. On se može primijeniti na tablice 2 x 2, poput one koju smo vidjeli kao primjer u trećemu poglavlju, iz istraživanja koje je proveo Andringa (2014), a koju ponavljamo u nastavku kao tablicu 16, uz dodavanje marginalnih frekvencija. U pitanju su podaci o tome koliko je učenika nizozemskoga jezika u gramatičkoj točnosti dostiglo normu određenu prema reprezentativnome i prema nereprezentativnome uzorku izvornih govornika – uzorci su zavisni jer su isti učenici procjenjivani u odnosu na obje norme (v. i potpoglavlje 3.1.1). Test ispituje jesu li marginalne frekvencije redaka i stupaca jednake (ovdje: ukupna frekvencija u retku „Gramatička točnost unutar norme“ i stupcu istoga naslova te ukupna frekvencija u retku „Gramatička točnost izvan norme“ i stupcu istoga naslova).

Tablica 16. Kontingencijska tablica opaženih frekvencija (prema Andringa, 2014)

		Nereprezentativni uzorak		UKUPNO
		Gramatička točnost unutar norme	Gramatička točnost izvan norme	
Reprezentativni uzorak	Gramatička točnost unutar norme	49	10	59
	Gramatička točnost izvan norme	0	50	50
UKUPNO		49	60	109

Uobičajena je formula za McNemarov test, utemeljena na hi-kvadrat distribuciji (iz koje se dobivaju kritične vrijednosti), ona u (15), gdje vidimo da se neposredno uzimaju u obzir samo frekvencije slučajeva u kojima postoji neslaganje između ishoda u dvjema različitim situacijama (engl. naziv za takve slučajeve je *discordant*; u našem primjeru, neslaganje znači gramatičku točnost unutar jedne norme, a izvan druge). U formuli *b* i *c* označavaju polja u kojima su prikazane frekvencije slučajeva neslaganja (u našem primjeru polja u kojima su 10 i 0).

$$(15) \chi^2 = \frac{(b-c)^2}{b+c}$$

Formula u (15) ne može se međutim primijeniti kad je broj slučajeva neslaganja mali te kad vrijedi $b+c < 25$; tada se primjenjuje egzaktni binomni test sličan testu koji je prikazan u četvrtome poglavlju (v. potpoglavlje 4.2.2), a uz to postoji još nekoliko varijanti i mogućih korekcija McNemarova testa. Za statističke programe uobičajeno je da automatski primijene neku od tih opcija. U našem primjeru iz Andringa (2014) R uz primjenu korekcije pokazuje $\chi^2=8,1$, $p=0,004$. U izvještaju o istraživanju može se navesti naprimjer „McNemarov test pokazao je da se kod gramatičke točnosti rezultati neizvornih govornika značajno razlikuju kad se usporede s reprezentativnim i nereprezentativnim uzorkom izvornih govornika ($\chi^2(1)=8,1$, $p=0.002$)“.

Naposlijetku, korisno je spomenuti i testove slične ovome koji se mogu provesti u nešto složenijim situacijama. Ako zavisna varijabla ima više od dvije razine te tablica postane naprimjer 3 x 3 (što bi se u navedenom primjeru dogodilo kad bismo dodali još jedan referentni uzorak izvornih govornika),

potrebno je provesti **McNemar-Bowkerov test** (engl. *McNemar-Bowker test*). Ako pak imamo nezavisnu varijablu s više razina (usp. tablicu 7), onda je odgovarajući statistički test **Cochranov Q test** (engl. *Cochran's Q test*).

5.2 Ispitivanje povezanosti između kategorijskih i numeričkih varijabli (dva uzorka)

U ovome i potpoglavlju 5.3 govorimo o statističkoj analizi u situacijama u kojima postoji jedna numerička zavisna varijabla i jedna ili više kategorijskih nezavisnih varijabli. Kategorijske nezavisne varijable u takvim nacrtima često vode k izdvajanju skupina ispitanika, gdje jedna razina varijable odgovara jednoj skupini. Naprimjer u istraživanjima usvajanja drugoga jezika vrlo se često prema varijabli 'jezični status' definira skupina koju čine 'izvorni govornici' i skupina koju čine 'učenici drugoga jezika' ili čak više skupina učenika u kojima su oni definirani i putem značajki kakve su primjerice razina općega znanja drugoga jezika ili materinski jezik.¹²³ Tako Roberts i Liszka (2013), čije smo istraživanje o obradi jednostavnoga i složenoga prošlog vremena u engleskome jeziku više puta citirale u prethodnim poglavljima, govore o skupinama 'izvorni govornici engleskoga jezika' (engl. *English native speakers*), 'učenici čiji je materinski jezik francuski' (engl. *French L2 learners*) i 'učenici čiji je materinski jezik njemački' (engl. *German L2 learners*). Također je česta situacija da se u jednome istraživanju ista skupina ispitanika izloži različitim razinama nezavisne varijable koja je vezana uz podražaje, tj. uz neku jezičnu pojavu koju oni predstavljaju. U istraživanju Roberts i Liszka (2013) svi su ispitanici čitali i rečenice u kojima je glagolsko vrijeme bilo usklađeno s vremenskim prilogom i rečenice u kojima to nije bio slučaj; autorice varijablu nazivaju 'vrsta (rečenice)' (engl. *(sentence) type*), a njezine razine 'usklađena' (engl. *match*) i 'neusklađena' (engl. *mismatch*; usp. i potpoglavlja 2.1.4 i 2.1.5). U tome se slučaju ne može govoriti o postojanju različitih skupina ispitanika već samo o različitim skupinama odgovora.

Ti su primjeri ponovljena ilustracija podjele uzoraka na nezavisne i zavisne (o istome je primjeru i istoj podjeli već bilo riječi u potpoglavlju 2.1.5, kao i na drugim mjestima u knjizi). „Uzorak“ upravo i jest jedini termin koji obuhvaća obje situacije te se opisi testova najčešće na njega referiraju (uz nazive koji

¹²³ Pod „jezičnim statusom“ podrazumijevamo to jesu li ispitanici izvorni govornici ili učenici nekoga jezika kao drugoga jezika. Ta se varijabla u istraživanjima često naziva jednostavno „skupina ispitanika“ ili samo „skupina“, dijelom i zbog njezina čestoga kombiniranja s drugim značajkama učenika (koje se na izvorne govornike najčešće ne mogu primijeniti pa ne predstavljaju varijablu koja je potpuno ukrižana s jezičnim statusom).

u središte pozornosti stavljaju ispitanike – „nacrt između ispitanika“ i „nacrt unutar ispitanika“ – i koji su također spomenuti u potpoglavlju 2.1.5 te vrlo prisutni u sljedećemu potpoglavlju). Treba međutim uočiti da „uzorak“ nema uvijek jednako značenje. Kod nezavisnih je uzoraka osnovno značenje toga termina isto kao i ono kojemu je bilo posvećeno potpoglavlje 2.1.3, tj. takvi su uzorci rezultat postupka uzorkovanja ispitanika, dok u slučaju zavisnih uzoraka često nije tako. Točnije, tako jest kad su u zavisne uzorke uključeni različiti ispitanici koji su prema svojim značajkama vrlo slični, tj. „upareni“; u istraživanjima usvajanja drugoga jezika takvi su uzorci rijetki, ali susreću se naprimjer u sociolingvistici, gdje se mogu uparivati članovi iste obitelji ili iste društvene skupine. S druge strane, u situacijama poput istraživanja Roberts i Liszka (2013), gdje postoje ponovljena mjerenja, uzorak ispitanika je jedan, a različite uzorke koji se (testom) uspoređuju čine različiti odgovori tih ispitanika, tj. njihovi podaci u različitim mjerenjima. Pri provedbi statističkih testova i kod nezavisnih se uzoraka može zauzeti takva perspektiva te govoriti o dvjema skupinama podataka i njihovoj međusobnoj nezavisnosti (koja potječe od međusobne nezavisnosti ispitanika).

U oba se slučaja kod testova govori o uspoređivanju uzoraka, i hipoteze koje se testiraju eksplicitno predviđaju razlike između uzoraka (posebice kod nezavisnih, gdje se može govoriti o razlici između skupina). Treba međutim imati na umu da takva razlika (ili njezin izostanak) zapravo odražava povezanost (ili izostanak povezanosti) između istraživanih varijabli, slično tome što se i kod hi-kvadrat testa umjesto o povezanosti dviju kategorijskih varijabli može govoriti o razlici između njihovih razina (u prethodno navedenom primjeru iz Robison, 1990, naprimjer statični i dinamični glagoli razlikuju se po tome u kojim se morfološkim oblicima najčešće javljaju).

Zbog čestoga oslanjanja na varijable koje su po svojoj prirodi kategorijske, u mnogim lingvističkim disciplinama, pa i u usvajanju drugoga jezika, to je najproširenija vrsta nacrt. Kako je već prethodno objašnjeno (v. posebice potpoglavlje 2.1.5), prema broju nezavisnih varijabli razlikuju se jednofaktorski i višefaktorski (ili faktorijalni) nacrti, a za statističku analizu jednofaktorskih nacrti važan je i broj razina nezavisne varijable, pri čemu treba svratiti pozornost na to ima li ih dvije ili više. U slučajevima kad su ispunjeni svi zahtjevi za uporabu parametrijskih testova, analiza tih nacrti temelji se na uspoređivanju aritmetičkih sredina uključenih uzoraka radi procjene toga razlikuju li se međusobno aritmetičke sredine odgovarajućih populacija. Kad parametrijski uvjeti nisu ispunjeni, može se govoriti o usporedbi medijana, ali najčešće ne izravno – većina neparametrijskih testova na temelju uzoraka

procjenjuje jesu li distribucije populacija jednake, što tek posredno može značiti i procjenu (ne)jednakosti njihovih medijana. Varianca je kao važna mjera prisutna samo u parametrijskim testovima, iako se naziv „nacrt varijance“ upotrebljava i šire, kao krovni naziv. Dva ključna dodatna kriterija za izbor testa jesu i već spomenuta međusobna (ne)zavisnost uzoraka i broj uzoraka koji se uspoređuju.

Ovo je potpoglavlje posvećeno testovima za nacрте varijance u kojima postoji jedna nezavisna varijabla s dvije razine. Testovi za takve kontekste pripadaju skupini najčešće korištenih ne samo u području usvajanja drugoga jezika već i mnogo šire (v. npr. Lindstromberg, 2016b). Podsjećamo ujedno da oni, kao i testovi opisani u sljedećemu potpoglavlju, nisu važni samo za testiranje glavnih hipoteza u istraživanju već se vrlo često upotrebljavaju i u analizi podataka o razini općega znanja drugoga jezika. Naime, ako razina znanja predstavlja varijablu u istraživanju, izmjerena je instrumentom koji su istraživači sami izradili (poput testa dopunjavanja teksta ili riječi, v. Kraš i Miličević, 2015, str. 45–48) i operacionalizirana putem kategorija (tako da se dobivaju dvije ili više skupina, npr. niža, srednja i viša), prije nego što se usporede rezultati tih skupina u glavnome eksperimentalnom zadatku, potrebno je utvrditi postoji li između skupina doista značajna razlika u razini znanja; bez takve provjere razina bi znanja mogla postati ometajuća varijabla.

5.2.1 T-testovi

Najjednostavniji je slučaj parametrijske analize nacрте varijance uspoređivanje dviju aritmetičkih sredina. Test koji se u tu svrhu upotrebljava jest **t-test**, također poznat pod nazivom **Studentov t-test** (engl. *Student's t-test*) zbog toga što ga je autor, William Sealy Gosset, prvotno objavio pod pseudonimom Student. Naime Gosset je bio kemičar koji je početkom dvadesetoga stoljeća radio za tvornicu piva Guinness i zbog potrebe čuvanja poslovnih tajni nije mogao objaviti rezultate svojih istraživanja pod pravim imenom. On je u Guinnessu radio na problemu kontrole kvalitete piva, gdje je trebalo kvalitetu održavati konstantnom usprkos tome što značajke sastojaka od kojih se pivo proizvodi nužno variraju. Jedan od problema kojima se bavio bilo je održavanje željene razine alkohola u pivu. Ta razina ovisi o sadržaju šećera u ječmenome sladu, koji prirodno varira zbog čimbenika poput kvalitete zemljišta na kojemu je ječam uzgajan, količine sunca kojoj je bio izložen i sl. Kada bi slad stigao u Guinnessova postrojenja, nije bilo moguće mjeriti sadržaj šećera za cijelu količinu, već je Gosset tražio način da taj sadržaj što pouzdanije procijeni iz uzoraka, računajući njihov prosječan sadržaj šećera. U to je vrijeme već bila

dobro poznata normalna distribucija, koju je Gosset namjeravao upotrebljavati kako bi ocijenio pouzdanost takve procjene, ali imao je na raspolaganju samo mali broj uzoraka i ubrzo je primijetio da je distribucija njihovih aritmetičkih sredina odstupala od normalne, čime je ostao bez temelja za ocjenu pouzdanosti dobivenih aritmetičkih sredina. *T*-test i distribucija vjerojatnosti na kojoj počiva nastali su kao rješenje za taj problem (v. detaljnije u Brown, 2008).

Kao i za hi-kvadrat, za *t*-test je s vremenom razvijeno nekoliko specifičnih verzija, za nekoliko različitih situacija. Ovisno o vrsti uzoraka, razlikuju se ***t*-test za nezavisne uzorke** (engl. *independent samples t-test*) i ***t*-test za zavisne ili uparene uzorke** (engl. *dependent/paired samples t-test*). U oba se slučaja mogu primijeniti jednosmjerna i dvosmjerna varijanta, gdje se odluka donosi prema istraživačkoj hipotezi ili slijedeći opću preporuku da se treba oslanjati na dvosmjerne statističke hipoteze čak i kad je istraživačka hipoteza jednosmjerna (usp. potpoglavlje 4.1.1). Kako bi se ikoja verzija toga testa mogla upotrebljavati, potrebno je da zavisna varijabla bude izmjerena na intervalnoj ili omjernoj ljestvici, kao i da distribucija podataka bude normalna, točnije da podaci dolaze iz normalno distribuirane populacije (pri čemu se test normalnosti provodi za svaki uzorak posebno i potrebno je da oba puta ukaže na normalnost). Testni statistik koji se uvijek dobiva jest ***t*-vrijednost** (engl. *t-value, t-statistic*).

T-test za nezavisne uzorke jedan je od najčešće korištenih statističkih testova u dosadašnjim analizama u mnogobrojnim područjima znanosti. Kako su uzorci koji se analiziraju nezavisni, oni mogu biti jednake ili različite veličine (usp. istraživanje Miličević, 2007, za koje smo u trećemu poglavlju vidjeli da je uključivalo skupine s različitim brojem ispitanika; v. posebice slike 22 i 23), s time što postoje različite testne formule za te dvije opcije. Također se može dogoditi da je u jednome od uzoraka varijabilnost u podacima veća, a u drugome manja. Izvorni je *t*-test podrazumijevao da to nije slučaj te da će uzorci imati sličnu varijabilnost, odnosno sličnu varijancu. Međutim u međuvremenu je razvijena i varijanta koja dopušta razlike u varijancama, poznata pod nazivom **Welchov *t*-test** (engl. *Welch's t-test*). Kako bi se odabrala odgovarajuća varijanta *t*-testa, potrebno je najprije usporediti varijance dvaju uzoraka namjenskim testom jednakosti varijanci, za što se upotrebljavaju **Leveneov test** (engl. *Levene's test*) ili ***F*-test**.¹²⁴ Poput testova

¹²⁴ Dodatno je dostupan i **Ansari-Bradleyev test** (engl. *Ansari-Bradley test*), ali je njega moguće upotrijebiti samo kad dva uzorka koja se uspoređuju imaju jednake medijane; u suprotnome slučaju daje pogrešne rezultate.

za provjeru normalnosti distribucije, kad ti testovi pokažu statistički značajan rezultat, moramo zaključiti da se uzorci međusobno razlikuju te da varijance nisu jednake; tj. uvjet jednakosti varijanci ispunjen je onda kada se dobije $p > 0,05$.

Iako se t -vrijednost kod t -testa za nezavisne uzorke u različitim situacijama izračunava prema različitim formulama, one sve uzimaju u obzir aritmetičke sredine zavisne varijable za uzorke koji se analiziraju (m), standardne devijacije uzoraka (s) i njihovu veličinu (n). U (16) vidimo formulu za uzorke jednake veličine s jednakim varijancama, u (17) za uzorke različite veličine s jednakim varijancama, a u (18) formulu za uzorke s različitim varijancama bez obzira na veličinu uzoraka (indeksi 1 i 2 odnose se na prvi i drugi uzorak).

$$(16) \quad t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$$

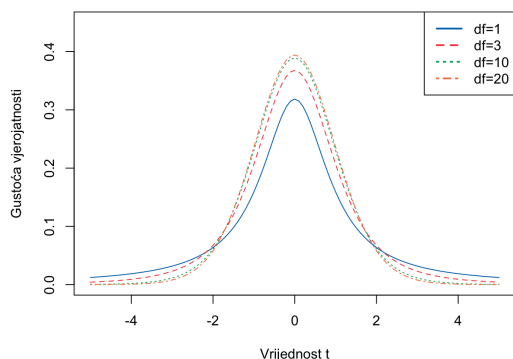
$$(17) \quad t = \frac{m_1 - m_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$(18) \quad t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Nakon što se primjenom odgovarajuće formule izračuna t -vrijednost i odrede stupnjevi slobode (tako što se zbroje veličine uzoraka koji se uspoređuju te se od toga broja oduzme dva, $df = n_1 + n_2 - 2$), potrebno je – slično kao kod hi-kvadrat testa – utvrditi je li dobivena vrijednost veća od kritične vrijednosti te je li rezultat statistički značajan (i na kojoj razini). Tablica kritičnih t -vrijednosti za različite stupnjeve slobode može se pronaći u mnogim statističkim knjigama (v. npr. Petz i sur., 2012), kao i na brojnim mrežnim stranicama (poput <http://vassarstats.net/tsamp.html>; posljednji pristup 26. 4. 2024.). Ta tablica sadrži apsolutne t -vrijednosti – iako će dobiveni rezultat nekad imati pozitivan, a nekad negativan predznak, taj predznak samo oslikava redoslijed unosa varijabli u analizu, a ne govori ništa o hipotezi koja se testira.

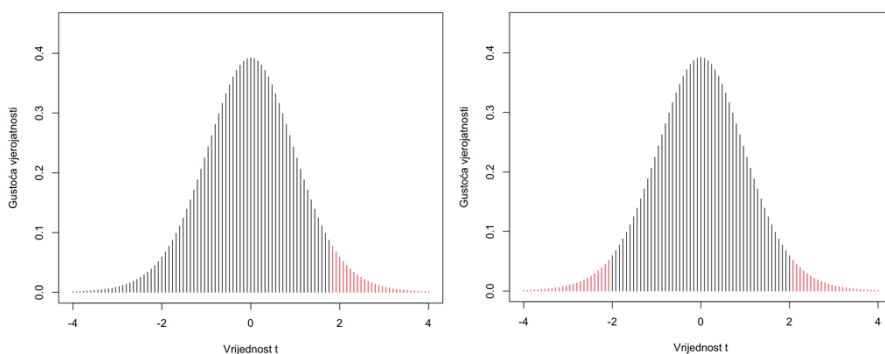
Temelj tablice kritičnih t -vrijednosti predstavlja t -distribucija, koju prikazuje slika 43. Ta je distribucija definirana brojem stupnjeva slobode kao jedinim parametrom i primjenljiva je na procjenu aritmetičke sredine populacije koja je normalno distribuirana u slučajevima kad je uzorak mali, a standardna devijacija nije poznata (kao u provjeri kvalitete piva *Guinness*). Kako se može

vidjeti na slici, ta je distribucija slična normalnoj, između ostaloga prema simetričnosti oblika, i utoliko joj je bliža kako broj stupnjeva slobode raste.



Slika 43. *T*-distribucija za nekoliko različitih stupnjeva slobode (*df*)

Kao i kod hi-kvadrat testa, putem krivulje možemo pronaći kritičnu regiju statističke značajnosti za različite stupnjeve slobode. Tako slika 44 pokazuje *t*-distribuciju za 15 stupnjeva slobode, gdje je kritična vrijednost za jednosmjerni test $t=1,7531$, a kritična vrijednost za dvosmjernu varijantu testa $t=2,1315$ – kako je već napomenuto kod binomnoga testa u potpoglavlju 4.2.2, uz dvosmjerni test potrebno je razmotriti oba kraja distribucije. Kritične su vrijednosti stoga u pravilu veće kod dvosmjernoga testa, a ako za određenu *t*-vrijednost dobivenu u testu potražimo u tablici ili rezultatima statističkoga programa vrijednost *p* za dvosmjernu i jednosmjernu varijantu testa (statistički programi često prikazuju obje), prva će uvijek biti viša.



Slika 44. Kritične regije *t*-distribucije (*df*=15)

Naravno, kao i kod drugih testova, vjerojatnost da ćemo u današnje vrijeme t -test raditi ručno i morati upotrebljavati tablice kritičnih vrijednosti vrlo je mala. Statistički softver poput programa IBM SPSS i R vrlo brzo dat će nam sve potrebne podatke, uključujući i točnu vrijednost p , ponekad i oba različita izračuna s obzirom na (ne)jednakost varijanci. Međutim, kako smo već više puta naglasile, ipak je korisno poznavati temelj onoga što vidimo kao konačni rezultat na ekranu jer nam takve spoznaje pomažu da interpretiramo i podatke i dobivene rezultate statističkih testova.

Slično se načelo koje je prethodno opisano primjenjuje i na t -test za zavisne uzorke, s time što je priroda zavisnih uzoraka takva da su vrijednosti uparene te uzorci uvijek moraju biti jednake veličine. Formula za izračunavanje t -vrijednosti kod te varijante testa osim veličine uzorka uzima u obzir razlike između parova rezultata – u formuli u (19) predstavlja aritmetičku sredinu razlika između parova rezultata, a s je u ovome slučaju standardna devijacija razlika; n predstavlja broj razlika, tj. broj uparenih vrijednosti.

$$(19) \quad t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}$$

Stupnjevi slobode jednaki su broju parova vrijednosti umanjenom za jedan ($df=n-1$). Drugim riječima, dok je kod nezavisnih uzoraka bitan ukupni broj opažanja (tj. vrijednosti koje se analiziraju), kod zavisnih je uzoraka bitan broj uparenih rezultata, te će za uzorke jednake veličine broj stupnjeva slobode biti niži za zavisne nego za nezavisne uzorke.

Razlike u izračunavanju t -testa za nezavisne i zavisne uzorke potječu otuda što je kod zavisnih uzoraka eliminiran jedan izvor varijabilnosti koji u istraživanjima sam po sebi nije predmet zanimanja – individualne razlike između ispitanika u dvjema skupinama. Prisutnost istih ili uparenih ispitanika u oba uzorka pridonosi smanjivanju nesustavne varijabilnosti u podacima, što nam olakšava prepoznavanje utjecaja nezavisne varijable. Analize utemeljene na t -testu za zavisne uzorke imaju stoga veću statističku snagu od onih utemeljenih na t -testu za nezavisne uzorke, odnosno mogu lakše prepoznati efekt varijabli (mada ne treba zaboraviti da zavisni uzorci često donose sa sobom opasnost od bilježenja učinka učenja, koji se također može odraziti na rezultate; v. potpoglavlje 2.1). Uz to, u istraživanjima usvajanja drugoga jezika često ne postoji mogućnost provedbe t -testa za zavisne uzorke zato što su istraživačka pitanja i hipoteze takvi da zahtijevaju nezavisne uzorke.

Ono što je obama testovima i svim njihovim varijantama zajedničko jest to da kroz različite formule pokušavaju otkriti isto – koliki se dio razlike između dvaju uzoraka može objasniti efektom istraživane nezavisne varijable. Naime u svim prethodno navedenim formulama brojnik iskazuje efekt koji se može pripisati nezavisnoj varijabli i koji se izračunava kao neki vid razlike između aritmetičkih sredina njezinih razina (bilo doslovce kao razlika između prosjeka dvaju uzoraka ili kao prosjek razlika između svih uparenih podataka), dok je nazivnik mjera varijabilnosti koja postoji unutar dvaju uzoraka (i predstavlja standardnu pogrešku razlike između aritmetičkih sredina). Na još se općenitiji način to može opisati kao odnos između **signala** (engl. *signal*), koji nas zanima, i **šuma** (engl. *noise*), koji moramo uzeti u obzir. Što je varijabilnost unutar uzoraka veća, to će šum biti veći (shematski se prikaz toga može vidjeti naprimjer u *online*-udžbeniku *Research methods knowledge base*, Trochim, bez dat., <https://conjointly.com/kb/statistical-student-t-test/>, posljednji pristup 26. 4. 2024.) O pojmovima signala i šuma može se govoriti u kontekstu većine inferencijalnih statističkih testova.

Kao primjer istraživanja u kojemu su se upotrebljavali *t*-testovi, i to obje vrste – i onaj za nezavisne uzorke i onaj za zavisne uzorke – možemo uzeti istraživanje učinka glosa (morfoloških oznaka) na prvome i drugome jeziku na spontano učenje vokabulara u engleskome kao drugome jeziku, koje su proveli Öztürk i Yorganci (2017). U tome su longitudinalnom istraživanju dvije skupine izvornih govornika turskoga čitale dvije vrste tekstova – jedna skupina s glosama na turskome, a druga s glosama na engleskome jeziku. Prije početka čitanja ispitanici su rješavali predtest, a nakon završetka čitanja posttest bez vremenske odgode i posttest s vremenskom odgodom. Za usporedbu rezultata dviju skupina ispitanika u posttestu bez vremenske odgode i posttestu s vremenskom odgodom korišten je *t*-test za nezavisne uzorke, dok je za usporedbu rezultata dviju vrsta posttesta unutar svake skupine ispitanika korišten *t*-test za zavisne uzorke. Možemo dodatno navesti i jednu moguću usporedbu iz podataka koji su već predstavljani u knjizi. *T*-test može se provesti naprimjer na dvjema skupinama učenika iz istraživanja Miličević (2007). U potpoglavlju 3.2 vidjeli smo na više grafikona njihove rezultate na testu razine općega znanja talijanskoga jezika i premda razlika djeluje kao očigledna, uvijek ju je dobro testirati. Primjenom Shapiro-Wilkova testa dobivamo naznake normalne distribucije (za SR-VS $p=0,402$, za SR-NS $p=0,518$), pa prosječne bodove dviju skupina (SR-VS=25,294, SR-NS=16,000) možemo usporediti *t*-testom za nezavisne uzorke. *F*-test pokazuje da su varijance dvaju uzoraka jednake ($p=0,785$),

stupnjeva slobode ima 31 (budući da ispitanika ima 17 u prvoj i 16 u drugoj skupini), a t -vrijednost od 9,377 ukazuje na visoko značajnu razliku ($p < 0,001$) i uz dvosmjerni test.

Kao treća verzija postoji i **t -test za jedan uzorak** (engl. *one-sample t-test*). Taj se test u istraživanjima usvajanja drugoga jezika rjeđe upotrebljava. Pogodan je za slučajeve u kojima u istraživanju postoji samo jedna zavisna varijabla, čiju aritmetičku sredinu uspoređujemo s nekom teorijskom ili na drugi način izvedenom aritmetičkom sredinom. Naprimjer, izvan eksperimentalnih istraživanja, Biedroń (2011) je tim testom utvrđivala postoji li statistički značajna razlika između prosječne inteligencije darovitih učenika drugoga jezika koji žive u Poljskoj i norme u vidu već poznatoga prosječnog rezultata „normalne“ poljske populacije, na različitim ljestvicama poljske verzije Wechslerova testa inteligencije za odrasle.

Veličina efekta za t -testove najčešće se iskazuje kao **Cohenov d** (gdje $d=0,4$ predstavlja mali efekt, $d=0,7$ srednji efekt, a $d=1,0$ veliki efekt, v. npr. Larson-Hall, 2012a, str. 252; treba primijetiti da vrijednost može biti i veća od 1, naprimjer uz t -test pokazan na podacima iz Miličević, 2007, dobiva se Cohenov $d=3,269$).¹²⁵ Cohenov d izračunava se kao razlika između aritmetičkih sredina uzoraka podijeljena standardnom devijacijom bilo koje skupine (ako je varijanca skupine homogena) ili kombiniranom standardnom devijacijom, koja predstavlja korijen prosjeka varijanci dviju skupina (v. Tenjović i Smederevac, 2011, str. 323–324). Ta mjera stoga zapravo pokazuje razliku između dviju skupina izraženu standardiziranim standardnim devijacijama (što omogućava izravno uspoređivanje razlika dobivenih u različitim istraživanjima). Postoje i skraćene formule za izračunavanje d , prikazane u (20) za nezavisne uzorke i u (21) za zavisne uzorke, gdje t predstavlja t -vrijednost dobivenu u testu, df stupnjeve slobode, a n veličinu uzorka. Treba upamtiti i da je formula u (20) adekvatna za uzorke koji su barem približno jednake veličine.

$$(20) d = \frac{2t}{\sqrt{df}}$$

$$(21) d = \frac{t}{\sqrt{n}}$$

¹²⁵ Susreću se i tumačenja prema kojima je 0,2 oznaka maloga efekta, 0,5 srednjega, a 0,8 velikoga efekta. Treba imati na umu i da su različita tumačenja relevantna za uspoređivanja nezavisnih i zavisnih uzoraka (v. Plonsky i Oswald, 2014).

Prilikom izvještavanja o rezultatima, za obje varijante testa navode se (apsolutna) t -vrijednost, stupnjevi slobode i vrijednost p , uz mjeru veličine efekta, na način koji je vrlo sličan kao i za hi-kvadrat test. Naprimjer za test za nezavisne uzorke ilustriran u ovome potpoglavlju navodimo: „ T -test za nezavisne uzorke pokazao je da se dvije skupine učenika čiji je materinski jezik srpski značajno razlikuju prema razini općega znanja talijanskoga pokazanoga na testu dopunjavanja teksta ($t(31)=9,377$, $p<0.001$; veličina efekta $d=3,269$).“ Kao grafički prikaz podataka preporučuju se dijagrami s pravokutnikom, piratski grafikoni ili neki drugi tip grafikona koji se može definirati kao sveobuhvatan ili bogat podacima (v. potpoglavlje 3.2.2), dok se deskriptivne mjere, prije svega aritmetička sredina i standardna devijacija, navode u tekstu izvještaja o istraživanju ili u tablicama, kako je u danoj situaciji preglednije s obzirom na količinu podataka koje treba prikazati (v. i šesto poglavlje).

5.2.2 Wilcoxonovi testovi

Među neparametrijskim testovima koji se upotrebljavaju onda kada nisu ostvareni uvjeti za primjenu t -testova, spomenut ćemo **Wilcoxonove testove** (engl. *Wilcoxon tests*). Kao i t -testovi, ti testovi imaju posebne verzije za nezavisne i zavisne uzorke. Za njih je važno istaknuti da nisu prema pristupu u potpunosti paralelni t -testovima. Naime, dok t -testovi ispituju razliku između aritmetičkih sredina dvaju uzoraka, Wilcoxonovi testovi ne ispituju izravno razliku između dvaju medijana. Lindstromberg (2020) govori o tome navodeći da neparametrijski testovi ispituju razliku između distribucija dvaju uzoraka, te da se o razlici između medijana može govoriti samo ako su distribucije jednake po obliku, dok rezultat testa može biti značajan i kad su dva medijana jednaka. Kada se dvije distribucije razlikuju i u obliku i u medijanu, statistički značajan rezultat može biti teško interpretirati.

Za nezavisne se uzorke upotrebljava **Wilcoxonov test zbroja rangova** (engl. *Wilcoxon rank sum test*). Tome je testu prema svojoj biti jednak **Mann-Whitneyev test** (engl. *Mann-Whitney test*),¹²⁶ s time što različiti statistički programi odabiru jednu ili drugu varijantu i nešto ih drukčije prikazuju, upotrebljavajući i različite oznake za testni statistik. Mi ćemo prikazati postupak karakterističan za Mann-Whitneyev test i njegov testni **statistik U** (engl. *U-statistic*). Uz to što se mora upotrebljavati za ordinalne podatke i intervalne/omjerne podatke koji nisu normalno distribuirani, taj je test načelno bolji izbor od t -testa za analizu malih uzoraka.

¹²⁶ Ponekad se kao naziv upotrebljava i Mann-Whitney-Wilcoxonov test, skraćeno M-W-W test.

Formule koje se upotrebljavaju za izračunavanje statistika U prikazane su u (22). U se najprije izračunava za svaki uzorak posebno na temelju zbroja rangova (R) danoga uzorka i na temelju broja opažanja (n) u obama uzorcima (indeksi 1 i 2 odnose se na prvi i drugi uzorak). Rangiranje iz kojega se poslije formiraju zbrojevi ne vrši se zasebno za svaku skupinu već za sve podatke zajedno (u suprotnome se ne bi mogao odrediti odnos između uzoraka jer bi svaki sadržavao iste rangove). Rangirani se podaci zatim ponovno raspoređuju u odgovarajuće nezavisne skupine i na njima se temelji daljnji proračun (zbroj rangova). Rangiranje se može provesti u rastućem ili u padajućem redosljedju, a konačni je rezultat u oba slučaja isti.

$$(22) U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Manja od dvije dobivene vrijednosti (U_1 ili U_2) uzima se kao konačna vrijednost statistika U , prema kojoj se određuje statistička značajnost na temelju odgovarajuće tablice kritičnih vrijednosti, u kojoj je potrebno potražiti polje u kojemu se križaju veličine dvaju uzoraka (v. primjerice <https://real-statistics.com/statistics-tables/mann-whitney-table/>, posljednji pristup 26. 4. 2024.). Za taj se test ne računaju stupnjevi slobode (u tumačenju se tablice uzima u obzir veličina uzoraka). Tablica kritičnih vrijednosti pritom obuhvaća samo manje uzorke – za veće je uzorke izračunavanje nešto drukčije i prepustit ćemo ga statističkim programima (aproksimira se normalna distribucija i umjesto statistika U javlja se standardna vrijednost z). Treba naglasiti i da je za razliku od prethodnih testova o kojima smo govorili, rezultat ovdje statistički značajan ako je dobiveni statistik manji od kritične vrijednosti. Također treba istaknuti da statistički programi koji primjenjuju samo verziju Wilcoxonova testa zbroja rangova kao testni statistik ne daju U već W i obično primjenjuju nešto drukčiji način njegova izračunavanja, koji međutim daje istu vrijednost p . U navođenju rezultata podjednako se mogu upotrebljavati obje verzije te je načelno najbolje držati se onoga što pokaže program u kojemu smo radili analizu.

Provedbom Mann-Whitneyeva testa na podacima iz Miličević (2007) dobiva se statistik $U_1=277$, $U_2=5$, što znači da se uzima $U=5$, koje je manje od potrebne kritične vrijednosti 47 za razinu $\alpha=0,001$, upućujući na visoko značajnu razliku, odnosno visoko značajnu povezanost između varijabli ‘skupina’ i ‘rezultat na testu’. Ako se test pak ponovi u programu R kao Wilcoxonov test zbroja rangova, rezultat je statistik $W=272$ te $p<0,001$. U ovome slučaju vidimo podudaranje s rezultatom t -testa iz prethodnoga potpoglavlja, što međutim neće uvijek biti slučaj – kada je razlika između uzoraka manja

nego u našem slučaju, lako se može dogoditi da samo parametrijski test prepozna efekt nezavisne varijable.

Dopunjujemo taj poznati primjer i jednim novim. Mann-Whitneyev test korišten je u istraživanju koje su provele Margaza i Gavarró (2020) o uporabi izrečene i neizrečene subjektne zamjenice u španjolskome i grčkome. U tome su istraživanju sudjelovale tri skupine ispitanika za svaki jezik, učenici drugoga jezika na srednjoj razini općega znanja (tzv. „srednji“ učenici), učenici drugoga jezika na naprednoj razini (tzv. „napredni“ učenici) i izvorni govornici. Svi su ispitanici rješavali pisani test popunjavanja praznina u kojemu su za svaku prazninu trebali odabrati izrečenu ili neizrečenu subjektivnu zamjenicu koja im je bila ponuđena. U testu su se pojavljivale subjektivne zamjenice za različita lica, dok su rečenice sadržavale dvije vrste konteksta – onoga u kojemu se referent mijenja i onoga u kojemu se referent ne mijenja. Mann-Whitneyev test upotrijebljen je za uspoređivanje odgovora parova skupina ispitanika, za svaki test (posebno na španjolskome i grčkome) i za različite eksperimentalne uvjete: srednjih i naprednih skupina, srednjih i kontrolnih skupina te naprednih i kontrolnih skupina; nezavisna je varijabla dakle bila ‘skupina’, a zavisna ‘izbor referenta’ (koji je bio iskazan kroz prosječni postotak izbora). Može se primijetiti i da je varijabla ‘skupina’ ovdje zapravo imala tri razine te da su autorice uspoređivale svaku skupinu sa svakom. Iako su autorice to uzele u obzir i primijenile odgovarajuću korekciju testa, u sljedećemu ćemo potpoglavlju detaljno govoriti o toj korekciji i o tome zbog čega takvu praksu treba izbjegavati te umjesto nje odabrati testove za više uzoraka.

Za uspoređivanje zavisnih uzoraka upotrebljava se **Wilcoxonov test rangova s predznacima** (engl. *Wilcoxon signed ranks test*). Kao rezultat toga testa dobiva se **statistik W** (koji se u nekim programima označava i kao V). Budući da je u pitanju test za zavisne uzorke, njegovo se izračunavanje kao i kod uparenoga t -testa temelji na parovima rezultata. U postupku izračunavanja rangiraju se apsolutne vrijednosti razlika između uparenih podataka, pri čemu je potrebno zabilježiti i predznak razlike (+ ili –) i izostaviti iz daljnje analize razlike jednake nuli. Po završenome rangiranju rangovi se povezuju sa svojim predznacima i posebno se izračunavaju zbrojevi pozitivnih i negativnih rangova. Manja od dviju dobivenih vrijednosti uzima se kao vrijednost statistika W . Za manje uzorke, statistička značajnost dobivenoga rezultata za danu veličinu uzorka (tj. broj parova uspoređenih vrijednosti) može se provjeriti u odgovarajućoj tablici kritičnih vrijednosti (v. npr. <https://real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/>, posljednji pristup 26. 4. 2024.). Rezultat je i u tome slučaju statistički značajan ako je dobiveni statistik W

manji od kritične vrijednosti i također se ne računaju stupnjevi slobode. Treba svratiti posebnu pozornost na to da kod toga testa n koje se traži u tablici nije nužno jednako početnome broju uparenih vrijednosti – od početnoga se broja oduzima broj razlika koje su bile jednake nuli, te su zbog toga isključene iz daljega izračunavanja. Ako je broj nula bio visok, to treba pažljivo razmotriti jer test proveden na podacima iz kojih je velik dio izostavljen gubi na valjanosti. Konačno, kao i kod testa za nezavisne uzorke, za veće uzorke treba se osloniti na statističke programe jer se procjene zasnivaju na aproksimaciji normalne distribucije i na standardiziranome testnom statistiku z . Ponekad za velike uzorke može biti dopuštena i provedba t -testa, koji se u tome slučaju smatra relativno otpornim na narušavanje parametrijskih uvjeta.

Wilcoxonov test rangova s predznacima korišten je naprimjer u istraživanju usvajanja korejskoga honorifičkog afiksa (*u*)*si* kod naprednih učenika drugoga jezika čiji je materinski jezik engleski (Mueller i Jiang, 2013). Dvije skupine ispitanika – učenici drugoga jezika i izvorni govornici korejskoga – sudjelovale su u zadatku čitanja vlastitim tempom. Zavisna varijabla bila je brzina čitanja, a nezavisne varijable prisutnost (ili odsutnost) honorifičkoga afiksa (*u*)*si* u rečenici te gramatičnost (ili negramatičnost) rečenice. Kombinacijom tih dviju nezavisnih varijabli s dvjema razinama dobivena su četiri eksperimentalna uvjeta. Zbog izostanka normalnosti distribucije podataka o brzini čitanja upotrijebljen je neparametrijski test. U četirima kritičnim regijama u rečenicama Wilcoxonovim testom rangova s predznacima testirana je razlika između gramatičnih i negramatičnih rečenica u kojima je bio upotrijebljen honorifički afiks (*u*)*si* i u onima u kojima nije; za svaku skupinu ispitanika provedeno je ukupno osam testova. Zanimljiva je značajka toga istraživanja i to da je tih osam testova provedeno u analizi prema ispitanicima i u analizi prema česticama, koja se s neparametrijskim testovima, posebno jednostavnijima poput Wilcoxonovih, rijetko provodi. Tim se dvjema različitim vrstama analize nastoji uzeti u obzir i varijabilnost između različitih ispitanika i varijabilnost između različitih podražaja u istome eksperimentalnom uvjetu (o čemu detaljnije govorimo u potpoglavljima 5.3.3 i 5.5; usp. i potpoglavlje 2.1.4).

Veličina efekta za Wilcoxonove testove može se iskazati kao koeficijent korelacije r (usp. potpoglavlje 5.4.), uzimajući u obzir standardizirani testni statistik z i veličinu uzorka n , tj. dijeleći z kvadratnim korijenom iz n , koje je jednako zbroju veličina pojedinačnih uzoraka kod nezavisnih, a broju parova vrijednosti kod zavisnih uzoraka (v. npr. Field i sur., 2012); tumačenje je da $r=0,1$ predstavlja mali efekt, $r=0,3$ srednji efekt, a $r=0,5$ veliki efekt. Rezultati se pri izvještavanju navode gotovo jednako kao za hi-kvadrat test i t -testove,

uz manjak informacije o stupnjevima slobode. Obavezno je dakle navođenje vrijednosti odgovarajućega testnog statistika i vrijednosti p , kojima se dodaje mjera veličine efekta. Naprimjer „Mann-Whitneyev test za nezavisne uzorke pokazao je da se dvije skupine učenika čiji je materinski jezik srpski značajno razlikuju prema razini općega znanja talijanskoga pokazanoga na *cloze* testu ($U=5, p<0.001$). Statistička značajnost popraćena je i praktičnom značajnošću, veličina efekta $r=0,854$.“ Deskriptivne mjere koje se uz neparametrijske testove poput Wilcoxonovih navode jesu medijan i interkvartilno raspršenje, dok su grafikoni isti kao kod t -testova, najčešće dijagrami s pravokutnikom.

Prikaz nam jedne parametrijske i jedne neparametrijske skupine testova pokazuje da ti testovi počivaju na istim općim načelima statističkoga zaključivanja zasnovanoga na testiranju nulte hipoteze, ali i na vrlo različitim načinima izračunavanja testnih statistika. Način se određivanja njihovih kritičnih vrijednosti također razlikuje jer dok za parametrijske testove postoje definirane distribucije vjerojatnosti (poput t -distribucije), kod neparametrijskih se testova barem za manje uzorke radi o nešto naprednijem „prebrojavanju“ vjerojatnosti različitih ishoda. U nastavku poglavlja vratit ćemo se i na problem zajednički objema skupinama testova, a taj je da su oni najpogodniji za jednostavnija uspoređivanja poput onih prema razini općega znanja drugoga jezika kada postoje dvije skupine ispitanika i za razinu znanja je upotrijebljena jedna mjera (poput broja bodova na testu, kao u istraživanju Miličević, 2007). Njihova primjena u analizi rezultata eksperimentalnih zadatka najčešće zahtijeva da se najprije iz rezultata za pojedinačne podražaje izračunaju izvedene mjere za svaki eksperimentalni uvjet, pri čemu se zanemaruje varijabilnost između podražaja (usp. potpoglavljja 5.3.3 i 5.5). Na kraju ovoga potpoglavljja možemo zaključiti da i primjeri iz postojećih istraživanja jasno pokazuju da su pojave i konteksti koji se istražuju u lingvističkim pristupima usvajanju drugoga jezika vrlo složeni te da se jednostavni testovi poput t -testova i Wilcoxonovih testova u njima moraju provoditi više puta – to je statistički gledano lošiji izbor od složenijih testova u kojima se istodobno analizira više uzoraka i više varijabli; već su testovi opisani u sljedećemu potpoglavljju korak k njima.

5.3 Ispitivanje povezanosti između kategorijskih i numeričkih varijabli (više od dva uzorka)

U istraživanjima usvajanja drugoga jezika vrlo često želimo usporediti više od dva uzorka, tj. više od dviju skupina ispitanika ili više od dvaju eksperimentalnih uvjeta. Može se učiniti da za takve slučajeve postoji jednostavno rješenje –

uporaba više *t*-testova ili više Wilcoxonovih testova, koje ćemo provesti na različitim parovima skupina ili eksperimentalnih uvjeta. Međutim, što je veći broj uzoraka koje uspoređujemo, to je veći broj testova potreban da bi se usporedili svi parovi, a kako podaci u različitim testovima nisu međusobno nezavisni (ista se skupina ili uvjet javlja u više testova), s porastom broja testova raste i vjerojatnost da ćemo pogrešno odbaciti neku od nultih hipoteza – u takvu se slučaju može govoriti o visokoj **ukupnoj vjerojatnosti pogreške tipa I** (engl. *family-wise error rate*).

Ako primjerice provedemo tri povezana *t*-testa ili Wilcoxonova testa (kako je bio slučaj naprimjer u istraživanju o subjektivnim zamjenicama koje su provele Margaza i Gavarró, 2020, a koje je prikazano u prethodnome potpoglavlju), ukupno će za sve njih vjerojatnost pojave pogreške I tipa iznositi ne 0,05, već 0,14, odnosno $1-0,95^3$ (prema formuli $1-(1-\alpha)^n$, gdje n predstavlja broj usporedbi) te će rizik pogrešnoga odbacivanja nulte hipoteze bitno porasti. Jedan od načina da se spriječi takvo uvećavanje rizika jest primjena neke od mogućih korekcija. Najproširenija je **Bonferronijeva korekcija** (engl. *Bonferroni correction*), kojom se granična razina statističke značajnosti prilagođava broju usporedbi. Za tri bi se usporedbe razina značajnosti prema toj korekciji morala postaviti na 0,05/3 (ako pretpostavimo $\alpha=0,05$), tj. da bi se test smatrao značajnim, trebalo bi u njemu dobiti $p<0,01667$; takvu korekciju primijenile su naprimjer Margaza i Gavarró (2020). Veći broj usporedbi nameće još strože zahtjeve – Mueller i Jiang (2013), čije je istraživanje također citirano u prethodnome potpoglavlju, morali su zbog osam različitih usporedbi uparenih eksperimentalnih uvjeta spustiti traženu vrijednost p na $p<0,001$.

Iako je u mnogim situacijama korisna ili čak nužna (posebice kad moramo upotrebljavati neparametrijske testove), poznato je da je Bonferronijeva korekcija vrlo konzervativna, odnosno restriktivna, i da loše utječe na statističku snagu testova. Zbog toga su za situacije koje bi zahtijevale usporedbu više parova uzoraka bolje rješenje analize koje su prilagođene upravo takvim nacrtima istraživanja.¹²⁷ U ovome potpoglavlju govorimo o nekima od njih, navodeći kako parametrijske testove, u potpoglavlju 5.3.1, tako i neparametrijske testove, u potpoglavlju 5.3.2. Dodatno nakon tih testova razmatramo i status eksperimentalnih čestica u statističkoj analizi, što nije pitanje vezano samo uz testove kojima je posvećeno ovo potpoglavlje, ali se kod njih najčešće susreće njegovo razmatranje (v. potpoglavlje 5.3.3).

¹²⁷ Norris i suradnici (Norris, 2015; Norris i sur., 2015) upućuju i na nekoliko manje konzervativnih korekcija za višestruke usporedbe, poput korekcija Holm–Bonferroni i Benjamini–Hochberg (v. i Al-Hoorie i Vitta, 2019; Larson-Hall, 2016).

5.3.1 Analiza varijance

Vrlo raširena statistička analiza kojom se istraživači mogu poslužiti kad trebaju usporediti više aritmetičkih sredina jest **analiza varijance**, skraćeno **ANOVA** (engl. *ANalysis Of VAriance*). Točnije, u pitanju je skupina srodnih testova koji u analizi povezanosti zavisne i nezavisnih varijabli pokušavaju odvojiti signal od šuma na temelju različitih vrsta varijance te kao rezultat daju **F-omjer** (engl. *F-statistic*). Uz *t*-testove ta skupina analiza pripada najčešće korištenim testovima u području usvajanja drugoga jezika (usp. Plonsky i Oswald, 2017, gdje se istraživače poziva da se manje oslanjaju na njih). Uz mogućnost usporedbe više od dviju aritmetičkih sredina, analiza varijance ima i niz složenijih varijanata koje dopuštaju istraživanje dviju ili više nezavisnih varijabli.

Dva najjednostavnija slučaja, usporediva s *t*-testom za nezavisne i zavisne uzorke, jesu **jednofaktorska ANOVA za nezavisne uzorke** ili **nezavisne skupine**, koja se naziva i **jednostavna** ili **jednosmjerna ANOVA za nezavisne uzorke** (engl. *one-way ANOVA for independent samples*),¹²⁸ i **jednofaktorska ANOVA za zavisne uzorke** ili **za ponovljena mjerenja**, koja se naziva i **jednostavna** ili **jednosmjerna ANOVA za zavisne uzorke** (engl. *one-way ANOVA for repeated measures*). Među terminima, kao što se može vidjeti, u uporabi je više mogućnosti. Mi smatramo da je najvažnije razumjeti da se pod faktorima podrazumijevaju kategorijske nezavisne varijable (kako je već spomenuto u potpoglavlju 2.1.5) te da se „smjer“ u nazivu „jednosmjerni“ ne odnosi na smjer hipoteze ili statističkoga testa već na broj uključenih nezavisnih varijabli.¹²⁹ Čini nam se da je u kontekstu ove knjige, u kojoj je naglasak na nacrtima istraživanja i razumijevanju njihove uloge u analizi, jasniji termin „jednofaktorski“ te ćemo pretežito njega upotrebljavati.

Krenut ćemo od prikaza najjednostavnijih slučajeva, tj. dviju jednofaktorskih analiza varijance.¹³⁰ U jednofaktorskoj se analizi varijance uspoređuju aritmetičke sredine dobivene iz vrijednosti zavisne varijable za više skupina ispitanika ili za više eksperimentalnih uvjeta. Ako pođemo od usporedbe nezavisnih skupina, ona se slično kao i kod *t*-testova temelji na uspoređivanju dviju vrsta varijabilnosti, one koja nas zanima (razlikâ u

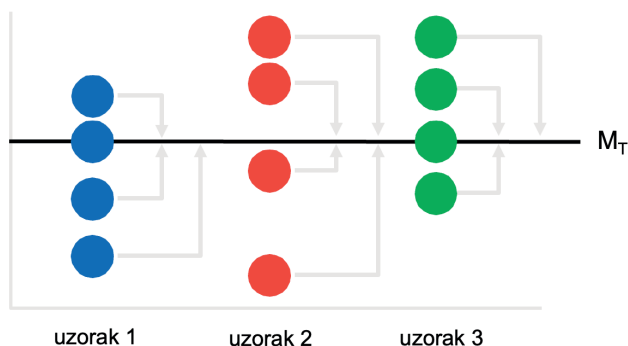
¹²⁸ Kad se izostavi kvalifikacija uzoraka – te kažemo primjerice „jednofaktorska ANOVA“ – pod time obično podrazumijevamo test za nezavisne uzorke.

¹²⁹ Usp. engl. *one-tailed* za smjer hipoteza i testova (v. potpoglavlja 4.1.1 i 4.2.2 te testove u ovome poglavlju), što bi se doslovce moglo prevesti kao „jednorepa/i“ i odnosi se na „repove“, tj. rubove distribucija vjerojatnosti, i *one-way* za jednosmjerne varijante analize varijance.

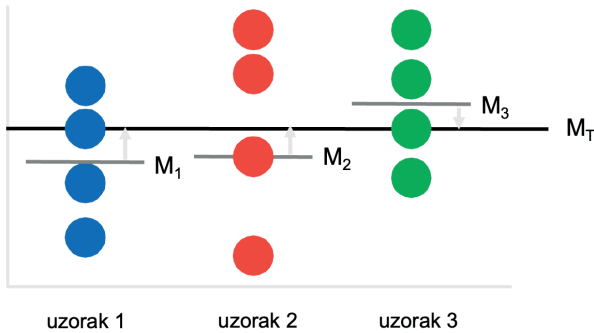
¹³⁰ Među radovima koji se detaljno bave tim analizama upućujemo posebice na Amoroso (2018).

aritmetičkim sredinama različitih skupina) i one koju su izazvali čimbenici ili faktori koji su nam nepoznati (varijabilnost unutar pojedinačnih uzoraka). Drugim riječima, ukupna varijabilnost koja je prisutna u podacima raščlanjuje se na dva glavna izvora – razliku između skupina i razlike unutar skupina. Razlika između skupina moguća je posljedica utjecaja nezavisne varijable, odnosno faktora koji se ispituje; ta je razlika zato objašnjiva i ona je indikator povezanosti nezavisne i zavisne varijable. Razlike unutar skupina predstavljaju neobjašnjivu varijabilnost; budući da nisu vezane uz nezavisnu varijablu koja nas zanima, za te se razlike upotrebljava i naziv **pogreška** ili **rezidual** (tj. ostatak; engl. *error, residual*).

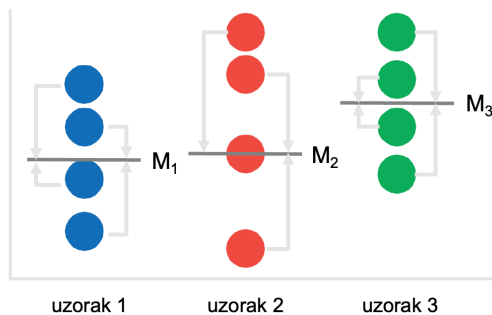
Možemo to shematski prikazati kao na slikama 45–47, koje opisuju varijabilnost u nezavisnim uzorcima. Slika 45 prikazuje ukupnu varijabilnost, koja ovisi o udaljenosti pojedinačnih vrijednosti od ukupnoga prosjeka (izračunatoga bez obzira na to kojemu uzorku koja vrijednost pripada; svaka točka predstavlja jednoga ispitanika, a M_T ukupni prosjek, od T kao „total“). Slika 46 prikaz je varijabilnosti između skupina, koja ovisi o udaljenosti prosjeka svake od skupina od ukupnoga prosjeka (M_T ponovo predstavlja sveukupni prosjek, dok su M_1 , M_2 i M_3 prosjeci pojedinačnih uzoraka); to je varijabilnost koja nas u istraživanju zanima, budući da skupine predstavljaju različite razine nezavisne varijable. Naposljetku, slika 47 prikaz je varijabilnosti unutar skupina (M_1 , M_2 i M_3 predstavljaju prosjeke pojedinačnih uzoraka); ona se mjeri udaljenošću pojedinačnih vrijednosti unutar svake od skupina od prosjeka te skupine – iako nije naš primarni predmet proučavanja, ona nužno postoji, pa je moramo uzeti u obzir. Ukupna je varijabilnost uvijek jednaka zbroju varijabilnosti između uzoraka i unutar njih.



Slika 45. Ukupna varijabilnost u situaciji s tri nezavisna uzorka



Slika 46. Varijabilnost između uzoraka u situaciji s tri nezavisna uzorka



Slika 47. Varijabilnost unutar uzoraka u situaciji s tri nezavisna uzorka

Testni statistik, F -omjer, izračunava se usporedbom varijabilnosti između skupina i varijabilnosti unutar skupina; točnije, radi se o dijeljenju odgovarajućih varijanaca – analiza varijance nosi takav naziv upravo zato što pokušava izdvojiti različite komponente koje sudjeluju u ukupnoj varijanci.¹³¹ Obje varijance dobivaju se dijeljenjem odgovarajuće sume kvadrata odgovarajućim stupnjevima slobode, koji se kod analize varijance izračunavaju zasebno između skupina i unutar njih. Suma kvadrata pojam je koji smo već upotrijebile kod izračunavanja standardne devijacije (usp. potpoglavlje 3.1.3); kod različitih se analiza ona može izračunavati na djelomično različite načine, ali njezina je bit u onome što smo objasnile u trećemu poglavlju – njome se mjeri odstupanje pojedinačnih vrijednosti od centralne tendencije skupine podataka (gdje je kvadrat nužan kako se odstupanja suprotnoga smjera ne

¹³¹ Dok je „varijanca“ termin za statističku mjeru koja se izračunava na točno određeni način, „varijabilnost“ je općenitiji naziv i upotrebljavaju se za situacije u kojima su vrijednosti neke značajke različite, tj. variraju (usp. i naslov potpoglavlja 3.1.3).

bi međusobno poništila). Točne formule za izračunavanje sume kvadrata između i unutar skupina, kao i način izračunavanja odgovarajućih stupnjeva slobode, navodimo u (23) i (24). SK predstavlja sumu kvadrata, T je total (zbroj svih vrijednosti u jednoj skupini, k), n broj ispitanika u skupini, N ukupan broj ispitanika u svim skupinama, a m broj skupina; x predstavlja pojedinačnu vrijednost. Indeksi i i u označavaju vrijednosti između i unutar skupina.

$$(23) SK_i = \sum \frac{T_k^2}{n_k} - \frac{(\sum x)^2}{N} \quad df_i = m - 1$$

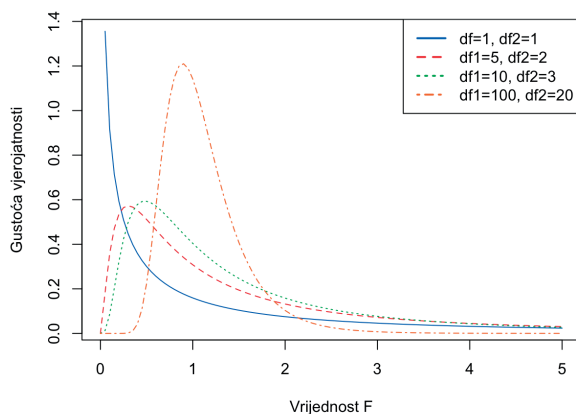
$$(24) SK_u = \sum x^2 - \sum \frac{T_k^2}{n_k} \quad df_u = N - m$$

Nakon primjene formula za sume kvadrata i stupnjeve slobode treba izračunati varijance, koje predstavljaju **srednje sume kvadrata** (engl. *mean sum of squares*), dijeleći SK/df_i (što daje **srednju sumu kvadrata između skupina**, MSK_i ; engl. *mean square between*) te SK_u/df_u (što daje **srednju sumu kvadrata unutar skupina**, tj. sredinu kvadrata pogreške, MSK_u ; engl. *mean square within*). Naposljetku, potrebno je podijeliti dvije sredine kvadrata međusobno, čime se konačno dobiva vrijednost F -omjera. Taj je dio postupka prikazan u formuli u (25).

$$(25) F = \frac{MSK_i}{MSK_u} = \frac{\frac{SK_i}{df_i}}{\frac{SK_u}{df_u}}$$

(Ne)postojanje i razina statističke značajnosti utvrđuju se na uobičajeni način iz tablice kritičnih vrijednosti. Važna je razlika u usporedbi s t -testovima ta da je za F -omjer potrebno ukrižati dvije vrijednosti za stupnjeve slobode, df_i za varijancu između skupina i df_u za varijancu unutar skupina. Na intuitivnoj razini, slično kao i kod t -testa, što je vrijednost F -omjera bliža jedan, to manje nezavisna varijabla objašnjava ukupnu varijabilnost, odnosno manje je vjerojatno da imamo temelja za odbacivanje nulte hipoteze; što je vrijednost F -omjera veća, to je utjecaj nezavisne varijable veći i to je vjerojatnije da se nulta hipoteza može utemeljeno odbaciti. Dodatno treba napomenuti da je za nezavisnu analizu varijance, kao i za nezavisni t -test, potrebno provjeriti imaju li različiti uzorci približno jednake varijance (v. npr. Hu i Plonsky, 2021). Većina statističkih programa pruža zatim mogućnost izbora između opcije izračunavanja utemeljene na ispunjenosti toga uvjeta i primjene korekcija onda kada taj uvjet nije ispunjen.

Kritične vrijednosti iz tablice izvedene su iz F -distribucije. Na slici 48 prikazane su krivulje ove distribucije za nekoliko različitih vrijednosti stupnjeva slobode.



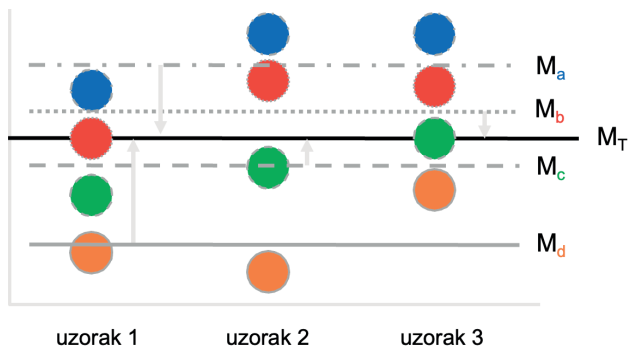
Slika 48. F -distribucija za nekoliko različitih stupnjeva slobode (df)

Također treba naglasiti da kod analize varijance ne postoje jednosmjerno i dvosmjerno testiranje, već je test uvijek dvosmjernan, s kritičnom regijom s desne strane. To je povezano i s time što ANOVA predstavlja „omnibus“ test koji daje samo podatak o (ne)postojanju ukupnoga efekta nezavisne varijable, bez naznake o tome koje se točno skupine međusobno razlikuju. Ako se pokaže da je ukupna razlika između skupina statistički značajna, najčešće će nas zanimati i od kojih skupina ta razlika potječe, te parove skupina možemo usporediti putem takozvanih **post hoc testova** (engl. *post hoc tests*), tj. testova višestrukih usporedbi parova; u tome se slučaju ne radi o višestrukim t -testovima, već o posebnoj skupini testova prilagođenih za tu namjenu. Jedan je primjer **Tukeyev test** (engl. *Tukey's test*), koji se još naziva i „Tukeyev test istinske značajnosti razlika“ (engl. *Tukey's honestly significant difference test*), te skraćeno „Tukeyev HSD test“ (engl. *Tukey's HSD test*). Taj je test vrlo sličan nezavisnome t -testu, ali dodatno uzima u obzir i broj usporedbi. Naime, on uspoređuje parove skupina oduzimajući njihove aritmetičke sredine jednu od druge (kao u brojniku t -testa, usp. formule u 16–18 u potpoglavlju 5.2.1) te dijeleći apsolutnu vrijednost razlike standardnom pogreškom izračunatom iz sredine kvadrata pogreške (koja uzima u obzir sve uzorke). Rezultati toga testa lako se mogu dobiti u statističkim programima i iako u postupku izračunavanja postoji testni statistik, on se rjeđe navodi, a naglasak se stavlja na prisutnost ili odsutnost statističke značajnosti.

Nezavisna jednofaktorska ANOVA popraćena Tukeyevim *post hoc* testom upotrijebljena je u Kraš (2008a), u istraživanju izbora pomoćnoga glagola pri tvorbi složenoga prošlog vremena *passato prossimo* s neprijelaznim glagolima u talijanskome kao drugome jeziku kod izvornih govornika hrvatskoga. U tome su istraživanju četiri skupine ispitanika – učenici drugoga jezika dječje dobi, učenici drugoga jezika odrasle dobi, izvorni govornici talijanskoga dječje dobi i izvorni govornici talijanskoga odrasle dobi – rješavali test dopunjavanja riječi kao mjeru razine općega znanja drugoga jezika. Nezavisnom jednofaktorskom analizom varijance utvrđivalo se razlikuju li se skupine ispitanika (nezavisna varijabla) u postotku točnosti (zavisna varijabla), dok se Tukeyevim *post hoc* testovima utvrđivalo koje se skupine ispitanika razlikuju jedna od druge.

Sličnu analizu možemo provesti i na podacima iz Miličević (2007) jer se već pokazalo da su ti podaci normalno distribuirani (v. potpoglavlje 5.2.1). Ako onome što otprije znamo o tome istraživanju dodamo to da se i u skupini izvornih govornika pronalazi normalna distribucija (Shapiro-Wilk test $p=0,054$), možemo na rezultatima triju skupina provesti nezavisnu jednofaktorsku analizu varijance, koja pokazuje visoku razinu značajnosti razlike ($F(2, 50)=392,28$, $p<0,001$), uz značajnu razliku u svim *post hoc* testovima ($p<0,01$). Ako pak provedemo isti test umjesto *t*-testa provedenoga u potpoglavlju 5.2.1, upotrebljavajući podatke dviju skupina učenika, vidjet ćemo da je testni statistik drukčiji, $F(1, 31)=87,92$, ali vrijednost p je ista (u ovome je primjeru to teško jasno pokazati jer je p izuzetno mali i u rezultatima se vidi kao $p<0,001$).

Kod jednofaktorske analize varijance za ponovljena mjerenja pored elemenata prikazanih na prethodnim slikama možemo zahvaljujući zavisnosti uzoraka dodatno izračunati i varijabilnost, tj. varijancu između ispitanika. Kao što se vidi na slici 49 (gdje M_T predstavlja ukupni prosjek, dok su M_a , M_b , M_c i M_d prosjeci individualnih ispitanika u različitim eksperimentalnim uvjetima – svaki „uzorak“ ovdje označava jedan uvjet), ta je varijabilnost odraz individualnih razlika i oduzima se od ukupne varijabilnosti unutar uzoraka jer objašnjava jedan njezin dio. Time se smanjuje pogreška pa veći dio varijance postaje objašnjiv.



Slika 49. Varijabilnost između ispitanika u tri zavisna uzorka

Izračunavanje F -omjera u biti je ovdje isto kao i kod analize varijance za nezavisne skupine, odnosno zasniva se na dijeljenju varijance koja nas zanima onom koja predstavlja pogrešku. Međutim formula za nazivnik drukčija je od navedenih formula jer od ukupne varijance unutar skupina treba oduzeti varijancu između ispitanika. Formulu za njezino izračunavanje navodimo u (26), zajedno s formulom za stupnjeve slobode unutar skupina; M_k predstavlja prosjek pojedinačnoga ispitanika, a M_T ukupni prosjek.

$$(26) SK_{is} = m \sum (M_k - M_T)^2 \quad df_p = (n - 1)(m - 1)$$

Kod analize varijance za zavisne uzorke također je potrebno paziti na uvjet jednakosti varijanci. Međutim, kako se radi o ponovljenim mjerenjima, uzimaju se u obzir varijance razlika između svih parova eksperimentalnih uvjeta; takva se jednakost varijanci naziva i **sferičnost** (engl. *sphericity*) i uobičajeno se provjerava primjenom **Mauchlyeva testa sferičnosti** (engl. *Mauchly's sphericity test*).

Primjer uporabe obiju opisanih vrsta analize varijance, popraćenih Tukeyevim *post hoc* testom, nalazimo u istraživanju koje su provele Kim i Lakshmanan (2009), u kojemu su proučavale obradu i prihvatljivost članova u engleskome kao drugome jeziku kod izvornih govornika korejskoga. Tri skupine ispitanika – učenici drugoga jezika na naprednoj razini znanja, učenici drugoga jezika na srednjoj razini znanja i izvorni govornici u funkciji kontrolnih ispitanika – rješavali su dva zadatka: zadatak čitanja vlastitim tempom i *offline* zadatak prosudbe semantičke prihvatljivosti rečenica. U oba su se zadatka upotrebljavale iste eksperimentalne rečenice, u kojima su nezavisne varijable bile kontekst (specifičan i nespecifičan) i član (određeni i neodređeni). Zavisna

varijabla u zadatku čitanja vlastitim tempom bila je brzina čitanja, a u zadatku prosudbe semantičke prihvatljivosti rečenica semantička prosudba izražena na ljestvici od 1 do 4. U svakome od dvaju zadataka dvjema analizama varijance s ponovljenim mjerenjima, jednom u specifičnome, a drugom u nespecifičnome kontekstu, utvrđivalo se postoji li statistički značajna razlika između brzine čitanja riječi ili prosudbi semantičke prihvatljivosti u rečenicama s određenim i neodređenim članom u svakoj od triju skupina ispitanika. Također, u svakome od dvaju zadataka dvjema analizama varijance za nezavisne skupine, popraćenima Tukeyevim *post hoc* testom, utvrđivalo se postoji li statistički značajna razlika između triju skupina ispitanika u brzini čitanja riječi ili prosudbi semantičke prihvatljivosti u specifičnome kontekstu i u nespecifičnome kontekstu.

Analiza varijance uz to ima i složenije varijante, koje se mogu upotrebljavati za višefaktorske nacрте istraživanja, u kojima se ispituje istodobni utjecaj više nezavisnih varijabli. I tu se može načiniti razlika između više vrsta analize varijance ovisno o vrsti uzoraka, ponajprije s obzirom na to jesu li nezavisni ili zavisni: (1) **višefaktorska ili faktorijalna ANOVA za nezavisne uzorke ili nezavisne skupine**, koja se još naziva i **složena ili višesmjerna ANOVA za nezavisne uzorke** (engl. *factorial ANOVA for independent samples*), uključuje dvije ili više nezavisnih varijabli koje se ispituju na nezavisnim uzorcima; (2) **višefaktorska ili faktorijalna ANOVA za zavisne uzorke ili za ponovljena mjerenja**, koja se još naziva i **složena ili višesmjerna ANOVA za zavisne uzorke** (engl. *factorial ANOVA for repeated measures*), uključuje dvije ili više nezavisnih varijabli koje se ispituju na zavisnim uzorcima; (3) **višefaktorska ili faktorijalna ANOVA za mješoviti nacrt**, koja se još naziva i **složena ili višesmjerna ANOVA za mješoviti nacrt** (eng. *mixed factorial ANOVA*), uključuje dvije ili više nezavisnih varijabli koje se ispituju dijelom na nezavisnim uzorcima, a dijelom na zavisnima – to je vjerojatno najčešća situacija u istraživanjima usvajanja drugoga jezika.¹³² Što je više faktora obuhvaćeno, to je analiza složenija jer se mora uzeti u obzir veći broj izvora varijabilnosti.

Osnovna je značajka složenih analiza varijance to da osim podataka o glavnome efektu pojedinačnih faktora daje i podatak o interakciji između njih, tj. o interakcijskim efektima (usp. potpoglavlje 2.1.5). Glavni efekt predstavlja onaj efekt koji promjena u danome faktoru ima na zavisnu varijablu bez

¹³² Višefaktorske se analize varijance mogu preciznije opisati prema broju faktora, naprimjer dvofaktorska ANOVA za nezavisne skupine, dvofaktorska ANOVA za ponovljena mjerenja, trofaktorska ANOVA za mješoviti nacrt i sl.

obzira na druge faktore. Interakcija između dvaju ili više faktora, nasuprot tome, predstavlja utjecaj koji kombiniranje dvaju ili više faktora ima na zavisnu varijablu. Interakciju smo već prethodno spomenule i u potpoglavlju 3.2.2, gdje je bilo riječi o različitim vrstama grafikona te je linijski grafikon spomenut kao pogodan prikaz interakcije varijabli jer se interakcijski efekt može uočiti prema međusobnome položaju linija.

Primjer uporabe faktorijalne analize varijance za mješoviti nacrt nalazimo u već prethodno spomenutom istraživanju usvajanja i obrade glagolskoga vremena i vida u engleskome kao drugome jeziku kod izvornih govornika francuskoga i njemačkoga koje su provele Roberts i Liszka (2013) (v. potpoglavlja 2.1.4 i 2.1.5). U istraživanju je upotrijebljen zadatak prosudbe prihvatljivosti rečenica i zadatak čitanja vlastitim tempom. Oba su zadatka sadržavala iste eksperimentalne rečenice, koje su bile u različitim vremenima (*past simple* i *present perfect*); dvije su vrste rečenica analizirane zasebno. U obje je vrste rečenica nezavisna varijabla bila (ne)podudarnost između vremenskoga priloga i glagolskoga vremena. U zadatku prosudbe prihvatljivosti rečenica zavisna je varijabla bila prosudba izražena na ljestvici od 1 do 6, pri čemu je brojka 1 izražavala najniži stupanj prihvatljivosti, a brojka 6 najviši stupanj prihvatljivosti, dok je u zadatku čitanja vlastitim tempom zavisna varijabla bila brzina čitanja. U analizi varijance, faktor unutar ispitanika, s dvije razine, bila je (ne)podudarnost između vremenskoga priloga i glagolskoga vremena, dok je faktor između skupina bila skupina ispitanika, s tri razine (izvorni govornici engleskoga, izvorni govornici francuskoga i izvorni govornici njemačkoga). Između ostaloga, tom je analizom utvrđen interakcijski efekt između skupine ispitanika i vrste rečenice (tj. (ne)podudarnosti) jer su izvorni govornici više reagirali na negramatične rečenice od neizvornih govornika.

Kao mjera veličine efekta uz analizu varijance posebno se često upotrebljava kvadrirana eta (η^2), koja se dobiva dijeljenjem varijance obuhvaćene određenim izvorom varijabilnosti i ukupne varijance u zavisnoj varijabli, što je slično Cohenovom d , a još sličnije koeficijentu determinacije, koji se izračunava kod parametrijske korelacije (v. potpoglavlje 5.4.1), budući da primjerice vrijednost $\eta^2=0,53$ znači da ispitivani faktor objašnjava 53 % varijance u rezultatima. Detaljnije o toj mjeri govore Norouzian i Plonsky (2018b). U pogledu navođenja rezultata načelo je isto kao i kod prethodno opisanih testova – navode se vrijednost testnoga statistika, sa stupnjevima slobode, vrijednost p te mjera veličine efekta; dodatno se uključuju informacije o *post hoc* testovima. U objašnjenju rezultata i/ili metoda analize navodi se i podatak o testiranju parametrijskih i drugih relevantnih uvjeta.

Konačno, treba napomenuti da je analiza varijance u ovome potpoglavlju, kao i t -testovi u prethodnome, opisana na način na koji se ona najčešće predstavlja u udžbenicima i priručnicima. Međutim u posljednje se vrijeme sve više svraća pozornost na to da se ti testovi mogu sagledati i iz druge perspektive, prema kojoj predstavljaju podvrste **općih linearnih modela** (engl. *general linear models*, GLM) i gdje se t -vrijednost i F -omjer računaju putem opće formule $ishod_i = model_i + pogreška_i$, već navedene u prethodnim poglavljima knjige, uz potrebna prilagođavanja formule za svaki pojedinačni test. To znači da su t -testovi i analiza varijance zapravo po svojoj prirodi vrlo bliski linearnoj regresiji (v. potpoglavlje 5.5), koja je također vrsta općega linearnog modela (Field, 2018; Plonsky, 2013, str. 675), što nadalje znači da se uz prilagođavanje načina na koji su varijable u istraživanju operacionalizirane u određenim slučajevima umjesto t -testa ili analize varijance može upotrebljavati linearna regresija. To se posebice odnosi na ispitivanja razine općega znanja drugoga jezika učenika, dobi na početku usvajanja drugoga jezika ili varijabli vezanih uz trajanje učenja drugoga jezika ili dužinu boravka u sredini u kojoj se drugi jezik govori – te varijable, posebice prva među njima, često se operacionaliziraju kao kategorijske, a zapravo su numeričke prirode i njihovim se svođenjem na kategorije gubi jedan dio podataka te smanjuje statistička snaga testa (v. npr. Plonsky, 2013, str 676; Tenjović i Smederevac, 2011, str. 328). U kontekstu usvajanja drugoga jezika, o prednostima razmatranja analize varijance iz perspektive regresijske analize posebno detaljno govore Plonsky i Oswald (2017). Mi ćemo se toj temi nakratko vratiti u potpoglavlju 5.5.

5.3.2 Kruskal-Wallisov i Friedmanov test

Kad je uzoraka koji se žele usporediti više od dva, a nisu ispunjeni parametrijski uvjeti, odgovarajući testovi postoje ponajprije za jednofaktorske nacрте. Nezavisnoj jednofaktorskoj analizi varijance odgovara **Kruskal-Wallisov test** (engl. *Kruskal-Wallis test*). Kao rezultat toga testa dobiva se **statistik H** , koji se izračunava prema formuli u (27). R u toj formuli predstavlja zbroj rangova pojedinačnoga uzorka, n je veličina pojedinačnoga uzorka, a N veličina svih uzoraka zajedno. Rangiranje, koje je nužno kao početni korak, vrši se na isti način kao i za Mann-Whitneyev test, odnosno Wilcoxonov test zbroja rangova, tj. rangiraju se podaci iz svih uzoraka zajedno, a nakon toga svi se rangovi „vraćaju“ u svoje skupine te zbrajaju.

$$(27) H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N - 1)$$

Formula u (27) zapravo je izvedena jednostavnija verzija osnovne formule, koja uzima u obzir kako zbrojeve rangova tako i prosječne rangove pojedinačnih uzoraka te kvadrate devijacija rangova. Može se stoga reći da je izračunavanje Kruskal-Wallisova testa doista blisko onome koje smo vidjeli u analizi varijance.¹³³

Stupnjevi slobode kod toga se testa izračunavaju kao $df=k-1$ (gdje k predstavlja broj skupina koje se uspoređuju). Vrijednost statistika H provjerava se zatim u tablici kritičnih vrijednosti – za statističku značajnost statistik treba biti viši od kritične vrijednosti. Kad je svaki od uzoraka veći od pet članova, distribucija statistika H približava se hi-kvadrat distribuciji, tako da se za kritične vrijednosti Kruskal-Wallisova testa upotrebljava ista tablica kao i za hi-kvadrat test. Treba napomenuti i da postoji poseban način izračunavanja, prikazan u formuli u (28), koji se primjenjuje kad u podacima postoje vezani rangovi. U toj formuli t označava broj vezanih, tj. izjednačenih vrijednosti za svaki pojedinačni rang. Treba međutim imati na umu da tu korekciju ne primjenjuju dosljedno svi statistički programi, pa u slučaju dobivanja različitih rezultata u različitim programima treba najprije provjeriti prisutnost vezanih rangova i način izračunavanja testnoga statistika u njihovoj prisutnosti.

$$(28) H' = \frac{H}{1 - \frac{\sum(t_i^2 - t_i)}{N^3 - N}}$$

S druge strane, jednofaktorskoj analizi varijance za ponovljena mjerenja odgovara **Friedmanov test** (engl. *Friedman's test*), za koji se ponekad upotrebljava i naziv **Friedmanova ANOVA** (engl. *Friedman's ANOVA*). Rezultat je toga testa statistik koji se ponekad označava kao F_r , a češće jednostavno kao χ^2 jer se i kod njega distribucija približava hi-kvadrat distribuciji. U tome se slučaju rangiranje podataka vrši posebno za svakoga ispitanika, odnosno rangiraju se ponovljene mjere, nakon čega se izračunava zbroj rangova za svaki uzorak (tj. za svaku razinu nezavisne varijable, što najčešće odgovara različitim eksperimentalnim uvjetima). Konačno, χ^2 dobiva se formulom u (29); i ovdje je, kao i za Kruskal-Wallisov test, dana jednostavnija formula.¹³⁴ Stupnjevi slobode ponovno iznose $df=k-1$ i vrijednost χ^2 mora biti veća od kritične vrijednosti iz tablice da bi se rezultat smatrao značajnim. Kako bi se distribucija statistika dobivenoga Friedmanovim testom približila

¹³³ Više detalja o načinu izračunavanja Kruskal-Wallisova testa može se pronaći na stranici <http://vassarstats.net/textbook/ch14a.html> (posljednji pristup 26. 4. 2024.).

¹³⁴ Više detalja može se pronaći na stranici <http://vassarstats.net/textbook/ch15a.html> (posljednji pristup 26. 4. 2024.).

hi-kvadrat distribuciji, za tri uzorka ($df=2$), potrebno je $n \geq 7$, a za četiri uzorka ($df=3$), $n \geq 5$. I u tome slučaju postoji posebna formula za uzorke s vezanim rangovima, ali tu formulu zbog njezine složenosti ovdje nećemo navoditi.

$$(29) \chi^2 = \frac{12}{nk(k+1)} \sum R_i^2 - 3n(k+1)$$

Neparametrijska uspoređivanja većega broja uzoraka, bilo nezavisnih bilo zavisnih, pružaju mogućnost izrade *post hoc* testova, ali u vezi s tim mogućnostima postoji nešto manje slaganje nego oko analize varijance. Jedna je mogućnost uporaba višestrukih testova za dva uzorka, poput Wilcoxonovih testova. Uz Kruskal-Wallisov test susreće se i **Dunnov test** (engl. *Dunn test*; v. Dinno, 2015), dok je uz Friedmanov test jedna od dodatnih mogućnosti **Conoverov test** (engl. *Conover test*; v. npr. Pereira i sur., 2015). Kako bi se izbjeglo povećavanje vjerojatnosti pojave pogreške I tipa, kod tih se testova primjenjuje neka od korekcija koje smo spomenuli na početku potpoglavlja 5.3 (v. i McDonough i Trofimovich, 2009, str. 165). Prijedlozi za Friedmanov test uključuju i jedan pravi *post hoc* test, za koji nisu potrebni naknadni konzervativni ispravci – **Nemenyiev test** (engl. *Nemenyi test*), koji je nastao prilagođavanjem Tukeyeva testa (usp. Pereira i sur., 2015).

Oba neparametrijska testa za usporedbu više uzoraka – i Kruskal-Wallisov test i Friedmanov test – upotrebila je Jaensch (2009) u istraživanju uporabe i ispuštanja člana u njemačkome kao trećemu jeziku kod izvornih govornika japanskoga kojima je drugi jezik engleski. Dvama zadacima potaknute usmene proizvodnje željelo se potaknuti ispitanike, podijeljene u tri skupine s obzirom na razinu općega znanja njemačkoga – nižu srednju, višu srednju i naprednu – na proizvodnju članova u nominativu, akuzativu i dativu. U analizi rezultata Kruskal-Wallisovim testom utvrđivalo se razlikuju li se tri skupine ispitanika u uporabi i ispuštanju člana, dok se Mann-Whitneyevim testom utvrđivalo koje se točno skupine razlikuju jedna od druge. Također, Friedmanovim testom za zavisne uzorke utvrđivalo se ima li gramatički rod imenice utjecaja na ispuštanje člana, dok se Wilcoxonovim testom rangova s predznacima utvrđivalo koji se rod (muški, ženski i srednji) razlikuje od kojega po vršenju toga utjecaja; sličan je postupak primijenjen i kako bi se utvrdilo ima li padež imenice (nominativ, akuzativ i dativ) utjecaja na ispuštanje člana. Kao dodatni primjer navodimo da ako provedemo Kruskal-Wallisov test na podacima triju skupina ispitanika iz već višestruko spominjanoga istraživanja koje je provela Miličević (2007), rezultat se podudara s onim dobivenim u analizi varijance – skupine se međusobno vrlo značajno razlikuju ($H(2)=46,08$, $p<0,001$).

Neparametrijski se testovi opisani u ovome i prethodnome potpoglavlju upotrebljavaju rjeđe od parametrijskih u području usvajanja drugoga jezika (o neparametrijskim opcijama u kontekstu primijenjene lingvistike vrlo detaljno pišu Egbert i LaFlair, 2018). Jednim je dijelom to posljedica izostanka testiranja pretpostavke normalnosti distribucije te moguća neopravdana provedba parametrijskih testova. Drugim je dijelom to povezano s nedostupnošću neparametrijskih testova za složenije nacрте, zbog čega se na sve načine teži normalnim distribucijama (između ostaloga, transformacijom podataka; v. potpoglavlje 3.1.6); alternative uključuju s jedne strane podjelu analize na manje komponente i odvojeno ispitivanje svakoga faktora, čime se ne dobiva potpuna slika ispitivane pojave i iznova se povećava rizik od pogreške tipa I, ili s druge strane oslanjanje na analize poput regresijskih i mješovitih modela, koji međutim također imaju niz uvjeta uporabe među kojima je i normalna distribucija (v. potpoglavlje 5.5).

5.3.3 Analize prema ispitanicima i prema česticama

U ovome se potpoglavlju osvrćemo na jedan više puta spomenuti problem, koji nije tijesno povezan samo s nacртima za koje je prikladan test analiza varijance ili Kruskal-Wallisov/Friedmanov test. Naime, više smo puta istaknule kako u eksperimentima unutar lingvističkih pristupa usvajanju drugoga jezika svaki eksperimentalni uvjet mora biti predstavljen većim brojem čestica (v. posebice potpoglavlje 2.1.4, kao i Kraš i Miličević, 2015). Kao što se prilikom formiranja skupina ispitanika pokušava postići da individualni ispitanici budu međusobno slični u svemu što nije predmet istraživanja, trebalo bi i različite čestice u istome eksperimentalnom uvjetu učiniti međusobno što sličnijima kako bi odgovori ispitanika na njih bili konzistentni (v. i potpoglavlje 2.1.4). Međutim u analizama koje su opisane u prethodnim potpoglavljima (5.2 te 5.3.1 i 5.3.2), rezultati za pojedinačne čestice promatraju se kao neposredne mjere iz kojih se izvode posredne (koje su aritmetičke sredine ili medijani neposrednih mjera) bez analize varijabilnosti između čestica; tj. ono što se analizira kao vrijednosti zavisne varijable jest izvedena mjera, u čiju varijabilnost nemamo uvid. Razlog je tome stavljanje naglaska na ispitanike kao jedinicu analize, gdje se aritmetičke sredine i standardne devijacije, ili medijani i druge odgovarajuće mjere, a potom i testni statistici utemeljeni na njima, izračunavaju iz rezultata pojedinačnih ispitanika, dok se čestice ostavljaju po strani. U potpoglavlju 2.1.4, u kojemu govorimo o uzorkovanju podražaja, već smo dosta napisale o tome zbog čega je takav pristup problematičan, a ovdje se želimo osvrnuti na neke – iako ne sve – mogućnosti za umanjenje takve problematičnosti.

Ako želimo imati uvid i u varijabilnost između čestica, jedan od načina na koji ga možemo steći jest putem pilotiranja istraživanja, koje se provodi između ostaloga upravo kako bi se prilikom analize mogle uočiti čestice na koje se dobivaju netipični odgovori (i kako bi se te čestice eliminirale ili popravile; usp. potpoglavlje 2.1.6 te Kraš i Miličević, 2015, str. 40–41). Međutim taj pristup ne jamči da će izmjene doista dati željene rezultate niti da će novi ispitanici odgovarati na isti način kao oni iz probnoga istraživanja. Dobro je zato i analizu glavnoga istraživanja dopuniti detaljnijom provjerom čestica kako bismo bili sigurni da izvedene mjere dobro predstavljaju različite eksperimentalne uvjete te da ne postoje čestice koje dovode do drukčijih odgovora od ostalih u istome eksperimentalnom uvjetu (zato što su lakše ili teže, zato što sadrže neku problematičnu riječ i sl.). Takav je pristup posebno zastupljen u okviru psiholingvističkih istraživačkih metoda, gdje postoji duga tradicija da se provodi s jedne strane **analiza prema ispitanicima** (engl. *by-subject analysis*), čiji su primjeri većina istraživanja o kojima govorimo u ovoj knjizi, a s druge **analiza prema česticama** (engl. *by-item analysis*).

Kako navode McDonough i Trofimovich (2009, str. 154–155), svrha je analize prema česticama provjeriti da rezultate dobivene u analizi prema ispitanicima nije uzrokovao manji broj „anormalnih“ čestica. Kako bi se takva analiza provela, potrebno je podatke reorganizirati tako da pojedinačna čestica postane jedinica analize (tj. da predstavlja jedan redak u tablici), gdje se zatim za svaku česticu računa odgovarajuća mjera centralne tendencije (uz odgovarajuću mjeru varijabilnosti) iz odgovora različitih ispitanika. Naprimjer u istraživanju Berger i sur. (2019), koje smo više puta spomenule u prethodnim poglavljima, mogli bismo u zadatku leksičke odluke izračunati prosječnu ili medijalnu brzinu reakcije i postotak točnosti odgovora za svaku pojedinačnu riječ prikazanu u eksperimentu te zatim grupirati rezultate prema eksperimentalnim uvjetima. Na tako pripremljenim podacima mogli bismo provesti i odgovarajući statistički test te provjeriti jesu li rezultati slični onima koji su dobiveni u analizi prema ispitanicima. Samo ako postoji podudaranje, možemo biti sigurni u „ispravnost“ eksperimentalnih čestica i u efekt istraživane varijable ili varijabli.

Takva se dvostruka analiza, koja uključuje i analizu prema česticama, najčešće provodi(la) uz analizu varijance kao statistički test, ali nije uvjetovana njome i može se provesti i kod drugih testova. Među primjerima istraživanja u kojima je provedena kad je test bila ANOVA, možemo navesti Roberts i Liszka (2013) te Kraš (2008a), dok je primjer analize prema česticama uz drukčiji test, točnije Wilcoxonov test rangova s predznacima, istraživanje Mueller i Jiang

(2013). Budući da se u analizi u takvu kontekstu dobivaju dva testna statistika za iste varijable, uobičajeno je da se u rezultatima oni označe indeksima 1 i 2, naprimjer F_1 i F_2 kod analize varijance ili t_1 i t_2 kod t -testova, gdje 1 označava analizu prema ispitanicima, a 2 analizu prema česticama. Iako su takve analize svakako bolji izbor od oslanjanja na izvedene mjere bez analize čestica, taj je pristup već dugo na meti kritika zbog toga što promatra ispitanike i čestice odvojeno; još jednom podsjećamo da je o tome bilo riječi u potpoglavlju 2.1.4, te ponovo upućujemo posebice na Barr i sur. (2013) te Barr (2018). Kod novijih analiza poput mješovitih modela (v. potpoglavlje 5.5) problem utjecaja pojedinačnih čestica promatra se drukčije – one se mogu unijeti u model kao slučajni faktori čineći zasebnu analizu prema česticama suvišnom i dajući pouzdanije podatke.

5.4 Ispitivanje povezanosti između numeričkih varijabli

Kako je najavljeno na početku poglavlja, prikaz i kritičko promišljanje testova u ovoj su knjizi i posebice ovome poglavlju organizirani s obzirom na nacrt istraživanja, ali u naslovima uzimamo u obzir i to što svi testovi na neki način ispituju povezanost nezavisnih i zavisnih varijabli. Ovdje možemo dodati da se kod numeričkih varijabli, kojima je posvećeno ovo potpoglavlje, radi o povezanosti na kakvu se obično najprije pomišlja kada se upotrijebi ta riječ – to je kontekst u kojemu se govori o povezanosti kao **korelaciji** (engl. *correlation*). Iako je moguće istraživati korelaciju i između većega broja varijabli, najčešći su i najosnovniji korelacijski nacrti oni koji uključuju dvije varijable te je ovo potpoglavlje posvećeno njima. Treba istaknuti i da zapravo nije nužno da obje varijable budu numeričke, ali je to daleko najčešći slučaj te ćemo ga ovdje uzeti kao fiksnu značajku situacija o kojima govorimo.¹³⁵ Još jedna značajka koja nije nužno uvijek prisutna, ali je mi upotrebljavamo da nadalje ograničimo izbor slučajeva, jest priroda numeričkoga odnosa između varijabli. Govorimo samo o **linearnoj korelaciji** (engl. *linear correlation*), koja predstavlja odnos u kojemu se s povećanjem jedne varijable druga varijabla povećava ili smanjuje na izravno proporcionalan način (primjeri su linearnih odnosa $y=3x$, $y=0.5x+10$ i $y=x+(2/3)$), dok stupnjevanje i korjenovanje nisu linearni procesi).

U korelacijskim nacrtima kakvi nas ovdje zanimaju susrećemo dakle dvije numeričke varijable i iako često imamo razloga da jednu od njih smatramo

¹³⁵ Primjer je korelacijskoga konteksta s jednom nenumeričkom varijablom onaj u kojemu je jedna varijabla kategorijska, točnije dihotomna, kad se govori o **biserijalnoj korelaciji** (engl. *point-biserial correlation*).

nezavisnom, a drugu zavisnom (usp. i dijagram izbora testa na slici 40), matematički gledano obje se varijable tretiraju jednako. S time je, između ostaloga, povezano i to da korelacijski nacrti ne daju informacije o uzročnosti, odnosno da **korelacija ne implicira kauzalnost**. To se najbolje može vidjeti ako pomislimo na razne numeričke vrijednosti koje mogu korelirati posve slučajno. Na stranici *Spurious Correlations*, posvećenoj „prividnim korelacijama“ (<https://www.tylervigen.com/spurious-correlations>, posljednji pristup 26. 4. 2024.), koju održava pravnik i korelacijski entuzijast Tyler Vigen, mogu se vidjeti mnogobrojni primjeri korelacija između varijabli koje teško mogu biti smisleno povezane jedna s drugom; dva su već naširoko poznata primjera prividno povezanih varijabli s te stranice potrošnja margarina po glavi stanovnika u Sjedinjenim Američkim Državama i stopa razvoda u američkoj saveznoj državi Maine te broj smrti davljenjem u bazenu i broj filmova koje je snimio Nicolas Cage. U istraživačkim kontekstima situacija je, naravno, drukčija od takvih primjera, ali oprez nije naodmet (v. i strip na tu temu na stranici <https://xkcd.com/552/>, posljednji pristup 26. 4. 2024.).

Dvije su glavne značajke korelacije. Jedna je ta da joj se može izmjeriti **jačina**, koja se uobičajeno iskazuje **koeficijentom korelacije** (engl. *correlation coefficient*). Koeficijenta korelacije ima više i odabir se između njih vrši prema ispunjenosti parametrijskih uvjeta i prema drugim značajkama uzoraka čija se povezanost istražuje; u nastavku ćemo potpoglavlja vidjeti dva često korištena koeficijenta, parametrijski i neparametrijski. Iako im se postupci izračunavanja razlikuju, svi koeficijenti korelacije na neki način pokazuju u kojoj mjeri dvije varijable zajednički variraju. Vrijednosti koeficijenta korelacije kreću se **između -1 i +1**; što je koeficijent bliži (apsolutnoj) vrijednosti od jedan, to je korelacija jača; vrijednost koeficijenta koja je blizu nule označava izostanak korelacije. Načelno se jačina korelacije ovako tumači iz (apsolutne) vrijednosti koeficijenta: 0,0–0,2 vrlo slaba korelacija, 0,2–0,4 slaba korelacija, 0,4–0,6 umjerena korelacija, 0,6–0,8 jaka korelacija, 0,8–1,0 vrlo jaka korelacija (vrijednosti su okvirne pa se može zanemariti pojavljivanje graničnih vrijednosti u dva raspona; usp. Phakiti, 2014, str. 211). Predznak uz vrijednosti koeficijenta vezan je uz drugu bitnu značajku, tu da korelacija može imati različit **smjer**: ako porast u vrijednostima jedne varijable prati porast u vrijednostima druge, u pitanju je **pozitivna korelacija** (engl. *positive correlation*; naprimjer broj godina učenja određenoga jezika i rezultat na testu razine općega znanja toga jezika obično će pozitivno korelirati), a ako porast jedne varijable prati opadanje vrijednosti druge, govori se o **negativnoj korelaciji** (engl. *negative correlation*; naprimjer broj godina učenja određenoga jezika vrlo će vjerojatno negativno korelirati s brojem pogrešaka na testu općega znanja toga jezika).

Korelacijski se nacrti u lingvistici, pa i u usvajanju drugoga jezika, češće susreću u korpusnim nego u eksperimentalnim istraživanjima, u skladu s time što se češće upotrebljavaju kod neeksperimentalnih metoda. Možemo povući paralelu i s onime što je rečeno o hi-kvadrat testu, koji je također češći u neeksperimentalnim istraživanjima i kod kojega su varijable u pogledu izračunavanja testnoga statistika također „simetrične“; međutim, dok hi-kvadrat može pokazati vjerojatnost postojanja povezanosti između varijabli (posredno, odbacivanjem nulte hipoteze o njihovoj međusobnoj nezavisnosti), korelacija pokazuje stupanj povezanosti (Petz i sur., 2012, str. 189, 211). Za korelaciju treba istaknuti i to da ima širu uporabu u statističkoj analizi: kao statistički postupak često je dio analize složenijih nacрта, upotrebljava se za izražavanje veličine efekta (usp. Wilcoxonove testove u potpoglavlju 5.2.2), a nalazi se u temelju i barem jednoga dijela mjera pouzdanosti instrumenta (v. npr. Norouzian i Plonsky, 2018a; usp. i potpoglavlje 2.3). Dodatno je od hi-kvadrat testa razlikuje to što se upotrebljava kad su uzorci zavisni – korelacija se zasniva na uparenim podacima te se vrijednosti obiju uključenih varijabli moraju odnositi na istu jedinicu analize.¹³⁶

Kao primjer numeričkih podataka na kojima se izračunava koeficijent korelacije u tablici 17 navodimo, u prilagođenome formatu, dio podataka iz istraživanja koje je proveo DeKeyser (2000); kompletni su podaci već prikazani u trećemu poglavlju kao primjer dijagrama raspršenja (v. slike 34 i 35) – u pitanju su podaci o dobi na početku usvajanja engleskoga kao drugoga jezika i o rezultatu (istih ispitanika) u zadatku prosudbe gramatičnosti rečenica, gdje se željelo izmjeriti odgovara li bolji uspjeh u zadatku prosudbe gramatičnosti nižoj dobi na početku usvajanja. Dijagram raspršenja predstavlja prikladan izbor jer su kod korelacije bitni odnos parova vrijednosti i uvid u individualne podatke, odnosno u distribuciju individualnih parova vrijednosti. Kako smo napomenule i u trećemu poglavlju, već iz ovoga dijagrama možemo mnogo doznati o jačini korelacije i o njezinu smjeru – što je distribucija točaka na grafikonu bliža ravnoj crti koja se pruža kroz dijagram dijagonalno, to je povezanost veća, dok smjer crte oko koje se podaci grupiraju upućuje na to je li korelacija pozitivna ili negativna.

¹³⁶ U nekim to situacijama može biti manje očigledno. Nije naprimjer lako uvidjeti zavisnost između podataka o potrošnji margarina u SAD-u i razvodima u Maineu – i prema opisu izvora podataka na stranici *Spurious Correlations* (https://www.tylervigen.com/spurious/correlation/5920_per-capita-consumption-of-margarine_correlates-with_the-divorce-rate-in-maine, posljednji pristup 26. 4. 2024.) reklo bi se da uzorci teško mogu imati dodirnih točaka. Situacija postaje jasnija kad vidimo da su podaci organizirani prema godinama, koje u metodološkome i statističkome smislu predstavljaju jedinicu analize – o svakoj godini u uzorku dostupna su dva podatka.

Tablica 17. Dio podataka iz istraživanja DeKeyser (2000, str. 525)

Ispitanik	Dob	Rezultat
I1	28	136
I2	8	170
I3	11	181
I4	20	132
I5	23	139
I6	22	153
I7	30	141
I8	27	147

Deskriptivne mjere koje se mogu izračunati za takve podatke jesu aritmetička sredina i standardna devijacija, ili medijan i interkvartilna razlika svake varijable zasebno (ovisno o (ne)ispunjenosti uvjeta za parametrijske mjere). Međutim te nam mjere same po sebi ne mogu reći mnogo o povezanosti dviju varijabli te često zapravo mogu biti i suvišne. Nasuprot tome deskriptivna mjera koja jest vrlo informativna u ovome slučaju jest koeficijent korelacije. On je jedna od mjera koje u statistici imaju pomalo podvojen status između deskriptivne i inferencijalne mjere. Točnije, koeficijent korelacije je deskriptivna mjera, izračunata iz podataka u uzorku, ali vidjet ćemo u nastavku da se nadalje može provesti i inferencijalni test kojim će se procijeniti je li korelacija statistički značajna. U ovome potpoglavlju pokušat ćemo na primjerima konkretnih koeficijenata razgraničiti jačinu korelacije izraženu koeficijentom korelacije i inferencijalni **test korelacije** (engl. *correlation test*).

Već kod odabira koeficijenata treba svratiti pozornost na uvjete za njihovo izračunavanje. Kako govorimo o linearnoj korelaciji, linearnost je potrebno prvu provjeriti. Za to je najčešće dovoljan grafički prikaz podataka na dijagramu raspršenja, posebice ako su dodane i Loessove krivulje (v. potpoglavlje 3.2.3), jer linearnost vizualno vrlo doslovno podrazumijeva uočavanje linije u podacima, odnosno grupiranje podataka oko zamišljene (ili ucrtane) linije (nasuprot primjerice krivulje). Nadalje je potrebno provjeriti ispunjenost parametrijskih uvjeta te na temelju rezultata provjera odabrati parametrijski ili neparametrijski pristup.¹³⁷ Dvije su od dostupnih opcija prikazane u sljedećim potpoglavljima, najprije jedna parametrijska, a potom jedna neparametrijska.

¹³⁷ Ako se u jednome od uzoraka utvrdi normalna distribucija, a u drugome distribucija koja nije normalna, bolje je okrenuti se neparametrijskoj metodi, posebice kod manjih uzoraka.

5.4.1 Pearsonov koeficijent korelacije

Kad su parametrijski uvjeti ispunjeni, u svrhu ispitivanja povezanosti numeričkih varijabli izračunava se **Pearsonov koeficijent korelacije** (engl. *Pearson's correlation coefficient*), koji se označava malim slovom r . Taj je koeficijent jedna od najčešće korištenih statističkih mjera, u različitim područjima i kontekstima. On konkretno mjeri količinu zajedničkoga variranja dviju varijabli čija se povezanost ispituje i temelji se na uobičajenim parametrijskim deskriptivnim mjerama, ponajprije na aritmetičkim sredinama i standardnim devijacijama dvaju uzoraka.

Pearsonov se koeficijent korelacije r može izračunati primjenom više različitih formula ili verzija iste formule. U (30) navodimo jednu često korištenu formulu, u dužoj i skraćenoj verziji, gdje X i Y predstavljaju dvije varijable čija se korelacija istražuje. Prilikom izračunavanja uzimaju se u obzir devijacije pojedinačnih vrijednosti od aritmetičke sredine uzorka (označene kao d , zasebno za varijablu X i varijablu Y) i broj parova vrijednosti (n). U početnoj verziji brojnik ukupno gledano predstavlja kovarijancu varijabli X i Y , tj. izražava njihovo zajedničko variranje, dok se u nazivniku množe mjere varijabilnosti (standardne devijacije) utvrđene za svaki uzorak posebno. Postupak, sveukupno gledano, ponovno odgovara pronalaženju različitih izvora i različitih vrsta varijabilnosti u podacima te načelu traženja signala i šuma.

$$(30) \quad r_{XY} = \frac{\frac{\sum d_X d_Y}{n-1}}{\sqrt{\frac{\sum d_X^2}{n-1}} \sqrt{\frac{\sum d_Y^2}{n-1}}} = \frac{\sum d_X d_Y}{\sqrt{\sum d_X^2} \sqrt{\sum d_Y^2}}$$

Kako je već istaknuto, vrijednost Pearsonova koeficijenta korelacije mora biti između -1 i $+1$ i ta vrijednost već sama po sebi pokazuje jačinu povezanosti dviju varijabli, ali u okvirima analiziranoga uzorka. Ako želimo provjeriti je li slična povezanost očekivana i u drugim sličnim uzorcima, možemo provesti i inferencijalni korelacijski test i dopuniti izračunavanje koeficijenta r utvrđivanjem statističke značajnosti. Nulta je hipoteza ta da između varijabli ne postoji povezanost, tj. da je koeficijent korelacije jednak nuli (v. npr. Brown, 1991, str. 577–578). Kako bismo utvrdili statističku značajnost, dovoljno je odgovarajućim postupkom izračunati koeficijent korelacije i usporediti njegovu vrijednost s tablicom referentnih kritičnih vrijednosti, pri čemu treba uzeti u obzir broj parova vrijednosti koje razmatramo, ili stupnjeve slobode (koji se u ovome slučaju izračunavaju kao $df=n-2$, gdje je n broj parova vrijednosti koje se uspoređuju); kod tablica treba provjeriti jesu li organizirane prema

stupnjevima slobode (df) ili prema broju parova vrijednosti (n) – vrlo su proširene obje opcije. Tablica se može pronaći naprimjer na stranici <http://vassarstats.net/rsamp.html> (posljednji pristup 26. 4. 2024.). Vrijednost koeficijenta r dobivena u istraživanju odgovara statistički značajnome rezultatu ako je njezina apsolutna vrijednost veća od odgovarajuće kritične vrijednosti. Dodatno, pozornost treba posvetiti i različitim vrijednostima za jednosmjerni i dvosmjerni test (odnosno test u kojemu se pretpostavlja i smjer povezanosti i test u kojemu se pretpostavlja samo postojanje povezanosti bez očekivanja o njezinu smjeru).

Iako utvrđivanje kritičnih vrijednosti Pearsonova koeficijenta koje su potrebne za odbacivanje nulte hipoteze djeluje jednako već opisanim postupcima, ono zapravo nije potpuno neposredno. Kako bi se testirala značajnost korelacije, provodi se posebna vrsta t -testa kojom se testira nulta hipoteza da je dobivena korelacija jednaka nuli. Formula za taj test navedena je u (31), gdje se vidi da se uzima u obzir vrijednost r , ali i n (veličina uzorka, tj. broj parova vrijednosti). Taj je dio postupka važno naglasiti jer nije očigledan – tablice radi lakše uporabe navode izravno kritične vrijednosti r , ali izračun uključuje i t -vrijednost i vezan je uz t -distribuciju (usp. potpoglavlje 5.2.1). To pomaže i da se bolje uvidi razlika između jačine korelacije i statističke značajnosti – u manjim se uzorcima može dogoditi da korelacija ne dostigne statističku značajnost usprkos vrijednosti r koja upućuje na srednju ili jaku povezanost, dok u velikim uzorcima i relativno nizak r može pokazati značajnu korelaciju. Vrijednost r pokazuje jačinu korelacije u uzorku, dok statistički test kaže možemo li očekivati sličnu povezanost u drugim uzorcima i populaciji.

$$(31) \quad t = r \sqrt{\frac{n-2}{1-r^2}}$$

Uz Pearsonov koeficijent korelacije ne postoje propratne mjere veličine efekta jer on već i sam pokazuje jačinu korelacije. Obično se dodatno izračunava **koeficijent determinacije** (engl. *coefficient of determination*), koji se označava kao R^2 . Taj je koeficijent doslovno jednak Pearsonovu koeficijentu korelacije podignutom na kvadrat. Zbog kvadriranja se kod njega gubi informacija o smjeru povezanosti, ali zato taj koeficijent upućuje na to koliki se dio varijabilnosti u varijabli Y može objasniti njezinom vezom s varijablom X , zbog čega se upotrebljava i u regresijskoj analizi. Kao rezultat testa korelacije navode se vrijednost koeficijenta korelacije r i vrijednost p , kao i broj ispitanika ili stupnjevi slobode, a može se navesti i koeficijent determinacije R^2 ; nije posve uobičajeno da se navodi i t -vrijednost iz inferencijalnoga testa (što je još jedan od razloga zbog kojih postupak nije sasvim transparentan).

Pearsonov koeficijent korelacije računala je naprimjer Biedroń (2011) u prethodno spomenutom neeksperimentalnom istraživanju inteligencije kod darovitih učenika drugoga jezika koji žive u Poljskoj. Za mjerenje jezične darovitosti učenika drugoga jezika upotrijebljen je test *Modern Language Aptitude Test* (MLAT), a za mjerenje njihove inteligencije poljska verzija Wechslerova testa inteligencije za odrasle. Izračunavanjem Pearsonova koeficijenta korelacije za svaku od pet ljestvica MLAT-a i njegovu opću ljestvicu te šesnaest ljestvica Wechslerova testa inteligencije za odrasle kao i njegovu opću ljestvicu željelo se utvrditi postoji li opća povezanost između jezične darovitosti i inteligencije te povezanost između pojedinih aspekata jezične darovitosti i pojedinih aspekata inteligencije.

5.4.2 Spearmanov koeficijent korelacije

Kada su podaci koje želimo korelirati ordinalne prirode ili su intervalni/omjerni, ali nisu normalno distribuirani, može se izračunati **Spearmanov koeficijent korelacije** (engl. *Spearman rank (order) correlation coefficient*). Oznaka je toga koeficijenta r_s . Da bi se izračunala njegova vrijednost, potrebno je, kao i kod drugih neparametrijskih postupaka koji su ranije spomenuti, najprije rangirati podatke. Zatim se izračunavaju razlike između rangova za pojedinačne vrijednosti varijabli X i Y te suma njihovih kvadrata. Konačno, uzimajući u obzir i broj parova vrijednosti, koeficijent se dobiva formulom u (32); detaljnije objašnjenje te formule (naprimjer odakle broj 6 u njoj) može se naći na stranici <http://vassarstats.net/textbook/ch3b.html> (posljednji pristup 26. 4. 2024.).

$$(32) r_s = 1 - \frac{6 \sum D^2}{n(n^2-1)}$$

Vrijednost Spearmanova koeficijenta dobivena navedenom formulom u nekim se slučajevima podudara s Pearsonovim koeficijentom korelacije izračunatim iz rangova (umjesto iz konkretnih vrijednosti). Međutim to vrijedi samo ako u podacima ne postoje vezani rangovi (iako neki programi za statističku analizu računaju Spearmanov koeficijent korelacije kao Pearsonov koeficijent iz rangova i onda kada postoje vezani rangovi). Na vezane rangove treba svratiti pozornost i zato što oni mogu dovesti do umjetnoga povećanja Spearmanova koeficijenta, odnosno do toga da povezanost varijabli bude precijenjena. Taj problem nije ozbiljan kad je broj vezanih rangova mali, ali kako taj broj raste, i problem postaje veći te postaje nužno primijeniti odgovarajuću korekciju (tj. izračunati koeficijent prema korigiranoj formuli).

Tumačenje Spearmanova koeficijenta korelacije jednako je tumačenju Pearsonova koeficijenta, s time što se za Spearmanov koeficijent ne izračunava koeficijent determinacije. Razlika je i u tome što se na Pearsonov koeficijent korelacije može nadovezati linearna regresija (v. potpoglavlja 3.2.3 i 5.5), dok to sa Spearmanovim koeficijentom nije slučaj. Stupnjevi slobode izračunavaju se na isti način kao kod parametrijske varijante, kao $df=n-2$. Tablica kritičnih vrijednosti odvojena je od one za Pearsonov koeficijent i dostupna je naprimjer na mrežnoj stranici <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118342978.app2> (posljednji pristup 26. 4. 2024.). Slično primjeni t -testa u testiranju značajnosti Pearsonova koeficijenta, i Spearmanov se koeficijent korelacije dodatnim postupkom testira spram nulte hipoteze. U tu se svrhu može upotrijebiti više različitih testova (statistički programi obično navode informacije o tome kako izračunavaju p), među kojima je i aproksimacija t -testa.

Spearmanov koeficijent korelacije računali su naprimjer Bulté i Roothoof (2020) u istraživanju povezanosti između razine općega znanja drugoga jezika i lingvističke kompleksnosti govora na istome jeziku. Na 71 uzorku spontanoga govora dobivenome u govornoj komponenti standardiziranoga testa razine općega znanja engleskoga jezika *International English Language Testing System* (IELTS), koji potječu od učenika engleskoga kao drugoga jezika na pet razina općega znanja prema IELTS-u, izračunato je devet mjera lingvističke kompleksnosti – pet mjera sintaktičke, tri mjere leksičke i jedna mjera morfološke kompleksnosti. Spearmanov koeficijent korelacije izračunat je između vrijednosti mjera lingvističke kompleksnosti i razine općega znanja drugoga jezika kako bi se utvrdilo postoji li povezanost između njih.

Kao primjer analize i navođenja rezultata možemo navesti korelaciju iz istraživanja DeKeyser (2000, str. 510), gdje se navodi „Korelacija između rezultata u zadatku prosudbe gramatičnosti i dobi na početku usvajanja bila je $[-0,546]$ ($p < 0,001$) za cijelu skupinu ($n=57$). Za ispitanike koji su se doselili kao odrasli ($n=42$), korelacija je bila $[-0,02]$ (nz); za ispitanike koji su se doselili u dobi mlađoj od 16 godina ($n=15$), korelacija je bila $[-0,408]$ (nz).“ (prev. M. M. P. i T. K.).¹³⁸

¹³⁸ Vrijednosti koeficijenta korelacije stavljene su u zagrade zbog toga što su izmijenjene u odnosu na izvorni tekst. Naime autor je proveo parametrijski test zasnovan na Pearsonovu koeficijentu korelacije, dok testiranje normalnosti pokazuje da je adekvatniji izbor Spearmanov test. Kratica „nz“ znači „nije značajno“; danas je uobičajeno umjesto nje navesti vrijednost p .

5.5 Regresijska analiza i mješoviti modeli

Naziv **regresija** (engl. *regression*) razlikuje se od naziva drugih do sada prikazanih testova – on ne opisuje što u podacima tražimo (kao korelacija) ni mjere na koje se oslanjamo (kao analiza varijance), niti se referira na testni statistik (kao *t*-testovi) ili na osobe važne za nastanak testa (kao neparametrijski testovi). Kod regresije se radi o tome da je kao analiza nastala u kontekstu u kojemu je istraživana „regresija ka prosjeku“, pojava koju je potkraj 19. stoljeća definirao Francis Galton kad je u proučavanju biološkoga nasljeđivanja uočio da izrazito visoki roditelji najčešće imaju djecu koja su niža od njih, a izrazito niski roditelji djecu koja su viša od njih, tj. da se ekstremne značajke često ne nasljeđuju, već se u sljedećoj generaciji pomiču bliže prosjeku populacije, odnosno regrediraju. Termin se „regresija“ otuda zadržao ne samo kao naziv za analizu koju je razvio Galton (v. Stanton, 2001) već i za niz drugih povezanih analiza koje danas čine široku skupinu, a o kojima kratko govorimo u ovome potpoglavlju.

Razlog je kratkoće našega prikaza taj što je cilj knjige osvrnuti se na manji broj osnovnih analiza, pojmove koji su za njih bitni te njihovu povezanost sa širim metodološkim značajkama istraživanja. Regresija, uza svu svoju složenost i prilagodljivost različitim nacrtima istraživanja, donosi više novih matematičkih nego osnovnih metodoloških pojmova. Među pojmovima širega značenja koji se kod nje čine prisutniji nego kod drugih statističkih analiza mogu se izdvojiti „model“ i „predviđanje“. Prvi pojam opisuje sve inferencijalne testove jer se njima uvijek (nesavršeno) modeliraju pojave, ali je to kod regresije očiglednije – polazište su analize regresijska jednadžba i pravac (uvedeni u potpoglavlju 3.2.3), kojima se kao najboljom mogućom aproksimacijom modeliraju manje ili više raspršeni podaci (te se javlja i termin **prilagodba modela**, engl. *model fitting*). Predviđanje, s druge strane, može biti širi cilj znanstvenih istraživanja ili konkretnija pretpostavka vezana uz neku općenitiju hipotezu (v. potpoglavlja 1.3 i 2.1.1), ali se može odnositi i na predviđanje vrijednosti zavisne varijable iz nezavisne; ponovno je zahvaljujući regresijskome pravcu to očiglednije kod regresije nego kod drugih testova koje smo spomenuli. U svojem najosnovnijem obliku regresija se može shvatiti kao produžetak korelacije, i to njezine parametrijske varijante. Ona omogućava da se na temelju poznatih uparenih vrijednosti dviju varijabli predvidi vrijednost zavisne varijable (*Y*) onda kada nam je poznata samo vrijednost nezavisne varijable (*X*). Taj se njezin najjednostavniji oblik, koji je razvio upravo Galton, naziva **jednostavna linearna regresija** (engl. *simple linear regression*). U njemu se nepoznate vrijednosti zavisne varijable predviđaju iz poznatih vrijednosti nezavisne varijable putem

regresijskoga pravca definiranoga jednadžbom $Y = bX + a$, tj. koeficijentima a – odsječak na osi y i b – nagib pravca (v. formulu u 7 u potpoglavlju 3.2.3). To se može odnositi na predviđanje konkretnih pojedinačnih vrijednosti koje se „očitavaju“ s pravca (naprimjer ako znamo s koliko se godina neka osoba kojoj je engleski drugi jezik doselila u Sjedinjene Američke Države, možemo uz pomoć podataka iz istraživanja DeKeyser, 2000, s određenom pouzdanošću predvidjeti njezin rezultat na testu poznavanja gramatike kakav je korišten u tome istraživanju; v. prethodno i potpoglavlje 3.2.3). Još se češće radi o općenitijem utvrđivanju toga predviđaju li i u kolikoj mjeri (gdje mjera može biti koeficijent determinacije R^2) vrijednosti nezavisne varijable vrijednosti zavisne varijable (kao kad bismo rekli da dob na početku usvajanja drugoga jezika s određenom pouzdanošću predviđa uspješnost u njegovu usvajanju).

Uz jednostavnu postoji i **višestruka regresija** (engl. *multiple regression*), kojom se ispituje odnos dviju ili više nezavisnih varijabli i jedne zavisne varijable. Ona nam omogućava da odgovorimo na pitanje: što najbolje predviđa varijablu Y (npr. usvojenost određene gramatičke pojave). Višestruka se regresija može provoditi kad su sve nezavisne varijable numeričke, ali i kad su neke od njih kategorijske (usp. sliku 40 u potpoglavlju 4.3), što omogućava analizu različitih nacrtu istraživanja. Kao i složene analize varijance, ona uz efekte pojedinačnih varijabli dodatno omogućava istraživanje interakcijskih efekata. Zahtijeva da između nezavisnih varijabli ne postoji **multikolinearnost** (engl. *multicollinearity*), odnosno međusobna korelacija;¹³⁹ razlog je za taj uvjet to što međusobno povezane nezavisne varijable nisu neovisne jedna o drugoj, pa se ne bi mogli jasno izdvojiti njihovi pojedinačni efekti (v. npr. Hu i Plonsky, 2021).

I jednostavna i višestruka linearna regresija pripadaju skupini općih linearnih modela, koji su već spomenuti u potpoglavlju 5.3 (budući da im pripadaju i analize varijance, a zapravo i t -testovi). Ti modeli svi počivaju na pronalaženju (regresijskoga) pravca koji najbolje opisuje podatke, za što se uobičajeno upotrebljava metoda najmanjih kvadrata, tj. najmanjih mogućih odstupanja skupine podataka od regresijskoga pravca (usp. potpoglavlje 3.1.3, gdje je pojam najmanjih kvadrata spomenut u objašnjenju mjera varijabilnosti).

Nešto je drukčije prirode **logistička regresija** (engl. *logistic regression*), kojom se predviđa ishod kategorijske zavisne varijable na temelju jedne ili više nezavisnih varijabli, koje mogu biti kategorijske ili kontinuirane. Ta nam analiza omogućava da predvidimo kojoj će kategoriji netko ili nešto pripadati (naprimjer

¹³⁹ Testiranje ispunjenosti toga uvjeta predstavlja dodatni primjer uporabe korelacije u statističkoj analizi podataka.

kakva će subjektna zamjenica biti odabrana, izrečena ili neizrečena). Ona je najčešće **binarna** (engl. *binary*), tj. uključuje dva moguća ishoda, odnosno zavisnu varijablu s dvije razine, ali može biti i **politomna** (engl. *polytomous*) te uključivati izbor između više od dviju mogućnosti (usp. izbor između tri referenta za subjektnu zamjenicu u Kraš, 2008b, 2008c, 2016, te Miličević i Kraš, 2017; v. i potpoglavlje 2.2.2 te primjere na slici 5). Logistička regresija pripada skupini **uopćenih linearnih modela** (engl. *generalized linear models*) jer predstavlja adaptaciju linearne regresije za kategorijske zavisne varijable putem posebne matematičke funkcije (kod logističke regresije u pitanju je logistička funkcija); logistička je regresija stoga parametrijska metoda, kao i linearna regresija.¹⁴⁰

Uz prethodno nabrojene vrste regresije posljednjih se godina sve više u lingvistici (posebice psiholingvistici, ali i u drugim područjima) upotrebljavaju **mješoviti modeli** (engl. *mixed models*). Oni su posebna vrsta regresijskih modela koja dopušta uključivanje fiksnih faktora i slučajnih faktora, koje smo već spomenule u potpoglavlju 2.1.4 – podsjećamo da se fiksnim faktorima nazivaju varijable koje istraživače zanimaju (i čije su sve razine koje se proučavaju uključene u istraživanje), dok su slučajni faktori oni koji se ne istražuju, ali su prisutni i mogu utjecati na rezultate kao dodatni izvor varijabilnosti (poput individualnih ispitanika, individualnih čestica i slično); kod slučajnih su faktora u istraživanju zastupljene slučajne vrijednosti (jer bi izbor ispitanika i podražaja trebao biti utemeljen na slučajnome uzorkovanju, v. npr. Baayen, 2013, str. 349; usp. i potpoglavlja 2.1.3 i 2.1.4). Također, dok su linearna i logistička regresija najpogodnije za analizu nezavisnih uzoraka, mješoviti su modeli jednako pogodni i za ponovljena mjerenja.

Razloge u korist uporabi mješovitih modela navode, između ostalih, Baayen i sur. (2008) za lingvistiku općenito, a Cunnings (2012) te Linck i Cunnings (2015) za područje usvajanja drugoga jezika. Ti autori ističu da uzimanje u obzir slučajnih faktora pridonosi preciznosti modela i boljem uvidu u utjecaj fiksnih faktora. Jaeger (2008) se posebno osvrće na analize kategorijskih zavisnih varijabli, za koje su istraživači jezičnih pojava često izračunavali relativne frekvencije (naprimjer ako je varijabla bila točnost odgovora, izračunao bi se postotak točnih odgovora za svakoga ispitanika i/ili za svaki eksperimentalni uvjet) te na takvim frekvencijama provodili analizu varijance, koja u tome slučaju ne daje pouzdane podatke. Svi navedeni autori ističu i to da su mješoviti modeli statistički pouzdaniji i učinkovitiji način uzimanja u

¹⁴⁰ Regresijske su analize gotovo uvijek parametrijske prirode. Neparimetrijske varijante postoje, ali nisu u široj uporabi (v. npr. Fox, 2005).

obzir slučajne varijabilnosti od paralelnih analiza prema ispitanicima i prema podražajima (usp. potpoglavlje 5.3.3), prema općemu načelu da se pouzdaniji rezultati dobivaju kad uključimo sve što nas zanima u jednu složenu analizu, nego kad provodimo više odvojenih manjih analiza (usp. potpoglavlje 5.3). Preostale rasprave i daljnja razrada ovih modela odnose se prije svega na to koje sve slučajne faktore uključiti i koje njihove aspekte. Te teme izlaze iz okvira ove knjige te upućujemo čitatelje na radove Barr i sur. (2013) i Barr (2018), koji nisu vezani uz analize jezika, te u domeni lingvistike na rad Winter i Grice (2021).

U istraživanjima usvajanja drugoga jezika regresija se dugo koristila rjeđe nego naprimjer u psihologiji (Plonsky, 2013, str. 676), što se u posljednje vrijeme sve češće ističe kao pogrešno. Naime, dok će u jednostavnijim nacrtima analiza varijance (koja je u području dominirala) i regresijska analiza dati podjednake rezultate (vrijednost F -omjera može se osim iz formula koje smo naveli u potpoglavlju 5.3.1 izračunati i primjenom regresijske analize, pri čemu se dobiva identična vrijednost p ; v. npr. Field, 2018; Field i sur., 2012), u prisutnosti većega broja nezavisnih varijabli regresija ima veću statističku snagu i jednostavnija je jer zajedno može ispitati više varijabli od analize varijance (v. posebice Plonsky i Ghanbar, 2018; višestruku regresiju, kao i logističku, detaljno obrađuje i Baayen, 2013). Čini se međutim da se trendovi mijenjaju, pa se u novijim istraživanjima regresijske analize, uključujući i mješovite modele, upotrebljavaju sve češće.

Kako bismo ilustrirale kontekste uporabe regresijskih analiza u području usvajanja drugoga jezika, navodimo nekoliko primjera empirijskih istraživanja u kojima su ovakve analize provedene. Linearnu je regresiju primijenio Testa (2019) u istraživanju uporabe članova u španjolskome kao trećemu jeziku kod izvornih govornika poljskoga kojima je drugi jezik engleski. Podaci su prikupljeni iz 72 sastavka na španjolskome učenika drugoga jezika koji su bili na razini općega znanja španjolskoga B1 i B2. Za svaku razinu općega znanja drugoga jezika izračunate su tri jednostavne linearne regresije kojima se željelo utvrditi postoji li statistički značajna povezanost između broja pogrešaka s članovima i broja članova, broja pogrešaka s članovima i broja imenica te broja članova i broja imenica. Statistički značajni rezultati upućivali bi na to da ti odnosi mogu, u prva dva slučaja, predvidjeti broj pogrešaka s članovima bolje od prosječnoga broja pogrešaka te, u trećemu slučaju, broj članova bolje od prosječnoga broja članova. Višestruka se linearna regresija izračunavala u već citiranome istraživanju Biedroń (2011) kako bi se utvrdilo koji aspekti opće inteligencije i u kojoj mjeri predviđaju različite aspekte jezične

darovitosti (pri čemu su aspekti inteligencije tretirani kao nezavisne varijable, a aspekti jezične darovitosti kao zavisna varijabla).

Kao primjer za binarnu logističku regresiju može poslužiti istraživanje koje je proveo Martín Gómez (2023), posvećeno usvajanju penjanja nenaglasnica u španjolskome jeziku kod nasljednih govornika španjolskoga čiji je drugi jezik engleski i kod učenika španjolskoga kao drugoga jezika čiji je materinski jezik engleski. U tome su istraživanju tri skupine ispitanika – nasljedni govornici španjolskoga, učenici španjolskoga kao drugoga jezika i izvorni (nenasljedni) govornici španjolskoga u funkciji kontrolnih ispitanika – rješavale usmeni zadatak dopunjavanja rečenica koji je sadržavao četiri eksperimentalna uvjeta, koji su se razlikovali u vrsti glagola (pomoćni glagol, modalni glagol, bezlični glagol, glagol u zavisnoj surečenici u niječnome obliku). Od dvije binarne logističke regresije koje su provedene, u prvoj se utvrđivalo koja je vjerojatnost da će ispitanici upotrijebiti proklizu, odnosno penjanje nenaglasnice, umjesto enklize, s obzirom na skupinu kojoj su pripadali i s obzirom na vrstu glagola koji je bio upotrijebljen u rečenici. U toj je analizi zavisna varijabla bila položaj nenaglasnice (prokliza ili enkliza), dok su nezavisne varijable bile skupina ispitanika i vrsta glagola. U drugoj se logističkoj regresiji određivalo koja je vjerojatnost da će ispitanici upotrijebiti proklizu umjesto enklize s obzirom na skupinu kojoj su pripadali (u analizu nije bila uključena kontrolna skupina), vrstu glagola i razinu općega znanja španjolskoga jezika. U toj su analizi varijable bile iste kao u prethodnoj, s time što je bila uključena još jedna nezavisna, kategorijska, varijabla – razina općega znanja španjolskoga jezika.

Primjer uporabe politomne logističke regresije nalazimo u Kraš (2016), istraživanju interpretacije izrečene i neizrečene subjektne zamjenice u talijanskome jeziku kod učenika drugoga jezika čiji je materinski jezik hrvatski, a koji su počeli usvajati talijanski u djetinjstvu (v. također Kraš, 2008b, 2008c; Miličević i Kraš, 2017). U tome su istraživanju dvije skupine adolescenata – učenici drugoga jezika i izvorni govornici talijanskoga – rješavali zadatak odabira slike u kojemu su u osam vrsta rečenica – četiri eksperimentalne i četiri kontrolne, pri čemu su eksperimentalne rečenice sadržavale anaforu ili kataforu – trebali identificirati referenta izrečene ili neizrečene zamjenice koja je bila upotrijebljena u rečenici vršeći odabir između triju mogućih referenata: subjekta glavne surečenice, objekta glavne surečenice i izvanlingvističkoga referenta. U politomnoj se logističkoj regresiji utvrđivalo koliko je izgledno da će koji referent biti odabran na temelju vrste zamjenice (izrečena ili neizrečena), vrste anafore (anafora ili katafora) i skupine ispitanika (učenici drugoga jezika ili izvorni govornici).

Naposljetku, mješoviti model primijenili su Perek i Hilpert (2014) u istraživanju u kojemu su testirali hipotezu da se razina tolerancije prema uporabi glagola u više različitih konstrukcija prenosi iz prvoga u drugi jezik (u ovome slučaju iz engleskoga i francuskoga u njemački). Fiksni su faktori u istraživanju bili skupina i konstrukcija, a slučajni faktori ispitanik i eksperimentalna čestica.

5.6 Drugi testovi

Kako je u prethodnim poglavljima i potpoglavljima više puta spomenuto, u posljednje se vrijeme uz već uvriježene ili „tradicionalne“ testove često savjetuje i uporaba novih statističkih testova.¹⁴¹ Kao i u svakome drugom području, promjene i napredak u primijenjenoj statistici i metodologiji istraživanja uobičajena su pojava i poželjno je upoznati se s novostima. Međutim ne treba zaboraviti da novi testovi ne čine nužno stare testove zastarjelima – „tradicionalno“ je testiranje i dalje važno poznavati; između ostaloga, novije se metode često izravno nadovezuju na već postojeće i predstavljaju njihovu adaptaciju te se ne mogu protumačiti bez njihova poznavanja. Prije nego što zaključimo ovo poglavlje isticanjem značajki koje nam se čine temeljnima za sve vrste statističke analize, osvrćemo se na neke od prijedloga za promjene u načinu provedbe statističke analize o kojima se govori u području usvajanja drugoga jezika.

Na nekoliko smo mjesta već spomenule da jednu od novijih opcija o kojima se dosta govori predstavlja Bayesovo statističko zaključivanje (v. npr. potpoglavlje 4.1), s odgovarajućim testovima. Ovdje se nećemo detaljnije baviti tim testovima, a čitatelje koji o njima žele doznati više upućujemo na radove Ross i Mackey (2015) i Vasishth (2023), kao i na radove Norouzian (2018) te Norouzian i sur. (2018, 2019), gdje je predstavljena i aplikacija u kojoj se rezultati uobičajenih *t*-testova mogu ponovno analizirati uporabom Bayesove metode.¹⁴² Druga je opcija robusna statistika, također više puta spomenuta u prethodnome tekstu (v. npr. potpoglavlje 3.1.6, gdje se govori o podrezanoj aritmetičkoj sredini kao robusnoj mjeri). Naziv potječe iz zapažanja da su testovi opisani u prethodnim potpoglavljima nerobusni, odnosno neotporni na male uzorke, a posebice na narušavanja uvjeta poput normalnosti distribucije

¹⁴¹ Atribut „nov“ ovdje ne treba shvatiti odviše doslovno. Prije svega, treba ga protumačiti kao „novije uporabe“. Često je riječ o metodama zasnovanim na odavno poznatim spoznajama koje su – metode – ranije bile ili manje dominantne ili prekomplikirane za praktičnu primjenu.

¹⁴² Aplikacija je dostupna na stranici <https://morouzian.shinyapps.io/bayesian-t-tests> (posljednji pristup 26. 4. 2024.).

ili jednakosti varijanci. Kako navodi Lindstromberg (2016a, str. 5), tradicionalni testovi u takvim slučajevima daju vrijednosti p koje su netočne (bilo da su previsoke ili preniske), i to se događa čak i kod neparametrijskih testova. Lindstromberg (2020) pobliže objašnjava da robusnost podrazumijeva održavanje razine pogreške tipa I na 5 % čak i kada su podaci na neki način problematični, dok se istodobno održava i visoka razina statističke snage testa, čak i u prisutnosti netipičnih vrijednosti; oba su aspekta vrlo važna, a posebice se može istaknuti prednost robusnih mjera za statističku snagu testova, koja je u istraživanjima usvajanja drugoga jezika zbog čestih malih uzoraka obično niska (Plonsky, 2013).

Larson-Hall (2012b) ističe da robusna statistika uključuje niz statističkih tehnika koje se bolje prilagođavaju stvarnim podacima čak i kad oni ne ispunjavaju uvjete u pogledu distribucije i sličnih kriterija. Jedna je od takvih tehnika uzorkovanje s povratom (v. i potpoglavlje 3.1.6), prvi puta preporučeno za uporabu u području usvajanja drugoga jezika (točnije, primijenjene lingvistike) u radu Larson-Hall i Herrington (2010). U pitanju je postupak u kojemu se iz prikupljenih podataka putem takozvane metode Monte Carlo (tj. računalnoga algoritma koji je primjenjuje) izdvaja velik broj slučajnih uzoraka kako bi se simulirala distribucija podataka koja bi se uobičajenim putem mogla dobiti samo na temelju bitno većega uzorka. Deskriptivne mjere i inferencijalni testovi izračunavaju se zatim iz procijenjene uzoračke distribucije. Rezultati koji se dobivaju približno su jednaki tradicionalnim testovima prema točnosti i snazi u slučaju kad su svi uvjeti ispunjeni, ali robusni testovi postižu znatno bolje rezultate u pogledu snage u slučajevima kad uvjeti nisu ispunjeni (Plonsky i sur., 2015).

To da robusni testovi ulaze u širu uporabu upravo sada, povezano je s metodološkim promjenama koje se događaju u usvajanju drugoga jezika kao i u drugim područjima (v. sedmo poglavlje), ali i s time što za njihovo izračunavanje danas postoje tehnički preduvjeti koje starija računala nisu mogla ispuniti (a ne radi se o postupcima koji se mogu lako provesti ručno). Danas se ti testovi mogu provoditi u velikome broju statističkih programa. LaFlair i sur. (2015) daju praktične upute za provedbu robusnih analiza u programima IBM SPSS i R, rad Egbert i Plonsky (2020) sadrži i opći uvod i praktične upute (iako ponajprije za korpusna istraživanja), dok koristan za obuku može biti i rad Egbert i LaFlair (2018). Međutim treba istaknuti da se ni uz novije i naprednije testove ne smiju zaboraviti osnovni metodološki zahtjevi i zahtjevi statističke analize – veličina i reprezentativnost uzoraka ispitanika i podražaja važni su za svaki test, kao što je uvijek važno temeljito

upoznavanje sa značajkama dobivenih podataka te provjera svih uvjeta za izračunavanje deskriptivnih mjera i testnih statistika. Naposljetku, sve informacije iz ovoga poglavlja upućuju na to da su preduvjeti za adekvatnu inferencijalnu statističku analizu: pozorno razmatranje odnosa zavisnosti i nezavisnosti (između varijabli, između uzoraka i unutar uzoraka) te analiza mogućih izvora varijabilnosti koje bi trebalo uzeti u obzir.

6. Izvještavanje o eksperimentalnome istraživanju

U trećemu smo poglavlju rekly da prilikom izvještavanja o istraživanju treba pokušati dati čitateljima sve informacije koje im mogu biti potrebne za razumijevanje istraživanja, ali se istodobno i suzdržati od davanja previše informacija, koje bi lako mogle ugroziti čitljivost (usp. Lindstromberg, 2016a, str. 5). U ovome ćemo poglavlju pokušati precizirati koje su to informacije te dati opće smjernice i upute za navođenje i interpretaciju rezultata statističke analize podataka u kontekstu izvještavanja o eksperimentalnim istraživanjima usvajanja drugoga jezika.

O rezultatima statističke analize podataka može se izvjestiti u različitim vrstama tekstova. Woodrow (2014, str. 3) razlikuje šest temeljnih vrsta tekstova o kvantitativnim istraživanjima, od kojih je za izvještavanje o rezultatima statističke analize podataka relevantno pet: kraći studentski radovi, neobjavljeni diplomski radovi ili doktorske disertacije, neobjavljeni izvještaji o istraživanju, znanstveni članci objavljeni u časopisima te znanstvene monografije i poglavlja u uredničkim knjigama.¹⁴³ Svaka se od tih vrsta razlikuje s obzirom na svoju svrhu, ciljno čitateljstvo, očekivanja koja čitateljstvo ima od njih (odnosno, svojevrstne kriterije za procjenu kvalitete) te norme i konvencije prema kojima se pišu, što sve zajedno utječe na njihov sadržaj i stil.

U ovome ćemo se poglavlju usmjeriti na znanstvene članke namijenjene objavi u znanstvenim časopisima i uredničkim knjigama, i to u području usvajanja drugoga jezika. Kako navodi Woodrow (2014, str. 6), temeljna je svrha znanstvenih članaka informirati čitatelje o novim znanstvenim spoznajama. Ciljno su čitateljstvo tih tekstova znanstvenici koji i sami provode istraživanja u danome području. Od te vrste tekstova čitatelji očekuju znanstveni doprinos području istraživanja te da budu fokusirani i relativno kratki. Što se tiče normi i konvencija prema kojima se pišu, one se u određenoj mjeri razlikuju od časopisa do časopisa, ali su, općenito govoreći, struktura i stil takvih radova poprilično konvencionalizirani. U nastavku poglavlja najprije opisujemo uobičajenu strukturu izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika te ukratko navodimo što sadrži svaki njegov dio. Nakon toga dajemo opće smjernice za izvještavanje o eksperimentalnome istraživanju, koje vrijede i za područje usvajanja drugoga jezika i za druge srodne discipline. Nadalje, detaljnije opisujemo pojedine dijelove izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika, počevši od

¹⁴³ Sinopsis istraživanja (engl. *research proposal*) nije relevantan jer se on piše prije provedbe istraživanja, odnosno prije prikupljanja podataka.

uvodnoga dijela, preko dijela u kojemu se opisuju metoda i postupak te dijela u kojemu se opisuju rezultati, do dijela u kojemu se raspravlja o rezultatima i iznose zaključci.

6.1 Struktura izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika

Četiri su temeljna dijela svakoga izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika: (1) uvod, (2) metoda i postupak, (3) rezultati, (4) rasprava i zaključci (v. Porte, 2002). Ta je struktura općenito svojstvena znanstvenim člancima, kako u prirodnim tako i u društvenim znanostima. U engleskome se jeziku često označava kraticama IMRD ili IMRaD, koje stoje za *Introduction, Methods, Results, and Discussion* (Uvod, Metode, Rezultati, Rasprava) (v. npr. Pavlović i Stanojević, 2020, str. 158). Prema Field i Hole (2003, str. 288), koji se ponajprije bave izvještajima o eksperimentalnome istraživanju u području psihologije, ti dijelovi redom odgovaraju na četiri temeljna pitanja u vezi s istraživanjem o kojemu se izvještava: (1) zašto?, (2) kako?, (3) što?, (4) pa što?. Kako autori objašnjavaju, u uvodu istraživači razlažu zašto su proveli istraživanje te što su očekivali da će u njemu otkriti, u dijelu posvećenome metodi i postupku opisuju kako su proveli istraživanje, odnosno kojim su se instrumentima i postupkom koristili, u dijelu koji se bavi rezultatima navode što su otkrili, a u dijelu koji sadrži raspravu i zaključke objašnjavaju što to sve znači, odnosno u kakvu su odnosu njihovi rezultati s rezultatima prethodnih istraživanja na istu temu, koje su njihove teorijske i, ako je primjenjivo, praktične implikacije te što se na temelju svega toga može zaključiti.

Uz te temeljne dijelove važni su dijelovi izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika (kao i u području psihologije, o čemu govore Field i Hole, 2003) također naslov, sažetak i popis literature. Svrha je naslova dati čitateljima osnovnu ideju o čemu se u radu radi kako bi na temelju toga mogli odlučiti zanima li ih rad i je li relevantan za njih. U dobru se naslovu spominju zavisne i nezavisne varijable iz istraživanja (Field i Hole, 2003). Sažetak je „minijatura verzija rada“, odnosno sažeti prikaz četiriju temeljnih dijelova izvještaja – uvoda, metode i postupka, rezultata te rasprave i zaključaka (Day i Gastel, 2014, str. 53). Drugim riječima, u sažetku je ukratko objašnjeno zašto i kako je provedeno istraživanje, što se u njemu otkrilo i što to znači, odnosno zašto je to važno. Njegova je funkcija omogućiti čitateljima da odluče hoće li pročitati cjeloviti rad te da stvore očekivanja u vezi s time što se u radu nalazi (Porte, 2002). Popis literature sadrži sve (i samo) radove na koje

su se autori rada referirali u tekstu, a njegova je svrha dati čitateljima potpune informacije o tim radovima kako bi ih oni mogli potražiti i provjeriti jesu li ih autori rada na vjerodostojan način predstavili te se detaljnije upoznati s temom rada. O tim dijelovima rada neće biti riječi u nastavku poglavlja; za više informacija o njima upućujemo čitatelje na: Field i Hole (2003), Gastel i Day (2022), Pavlović i Stanojević (2020) i Porte (2002).

Struktura izvještaja o eksperimentalnome istraživanju u kojemu se provelo više pojedinačnih eksperimenata složenija je utoliko što se u njegovu središnjem dijelu višestruko ponavljaju dijelovi posvećeni metodi i postupku, rezultatima i raspravi (onoliko puta koliko je eksperimenata provedeno), nakon čega slijedi opća rasprava, u kojoj se raspravlja o rezultatima svih eksperimenata zajedno, prije iznošenja zaključaka. Primjer je za to znanstveni članak u kojemu se izvještava o istraživanju Felser i sur. (2009), u kojemu su provedena dva eksperimenta – jedan u kojemu je upotrijebljen ubrzani zadatak prosudbe prihvatljivosti rečenica i drugi u kojemu je primijenjena tehnika praćenja pokreta očiju.

6.2 Opće smjernice za izvještavanje o eksperimentalnome istraživanju

Korisne opće smjernice za izvještavanje o eksperimentalnome istraživanju daju Field i Hole (2003). Njihove se smjernice ponajprije odnose na studentske izvještaje u području psihologije pisane po uzoru na znanstvene članke iz časopisa, ali su primjenjive i na druge vrste izvještaja iz srodnih znanstvenih disciplina pa tako i na izvještaje iz usvajanja drugoga jezika. U ovome potpoglavlju sažimamo sadržaj tih smjernica referirajući se, prema potrebi, i na druge izvore.

Ponajprije, svrha je izvještaja o istraživanju prenijeti znanstvene spoznaje na jasan i sustavan način. Ti izvještaji imaju standardizirani format, kojega se pri pisanju izvještaja treba što vjernije pridržavati. Svrha je toga omogućiti čitateljima da što brže pronađu informacije koje ih zanimaju jer točno znaju u kojemu se dijelu izvještaja što navodi.

Nadalje, pri pisanju izvještaja treba stalno voditi računa o ciljnome čitateljstvu i zamišljenim čitateljima (v. i Woodrow, 2014). Zamišljeni čitatelji trebali bi biti zainteresirane, obrazovane osobe koje znaju nešto o našem području istraživanja (npr. usvajanju drugoga jezika), ali nisu upoznate s temom našega istraživanja (npr. usvajanjem članova u engleskome kao drugome jeziku). Pri pisanju treba imati na umu da čitatelji znaju samo ono što je rečeno, odnosno napisano do toga trenutka, pa od njih ne bi trebalo očekivati

da čitaju dijelove izvještaja koji slijede kako bi razumijeli dijelove izvještaja koje čitaju u nekome trenutku. Sve relevantne informacije u izvještaju treba dakle navoditi postupno.

Uz to sve tvrdnje koje se iznose treba potkrijepiti. To se može učiniti referiranjem na tuđe radove (koji se zatim uključuju u popis literature), ali i na aspekte vlastitoga istraživanja. Naprimjer pri opisu rezultata treba navoditi numeričke rezultate statističkih testova, dok se pri izvođenju zaključaka treba referirati na prethodno opisane rezultate vlastitoga istraživanja, uz tuđe rezultate.

Također, izvještaj treba biti napisan formalnim stilom, takozvanim znanstvenim funkcionalnim stilom, čije su značajke jasnoća, preciznost i objektivnost (v. i Pavlović i Stanojević, 2020, str. 169). Poželjno je upotrebljavati kraće i jednostavnije rečenice te izbjegavati suvišnu uporabu termina (stručnih pojmova), nepreciznu uporabu riječi i redundanciju, odnosno suvišna ponavljanja. U hrvatskome se jeziku nameće dilema treba li u terminologiji dati prednost uvrježenijoj riječi stranoga podrijetla, poput „distribucija“, „frekvencija“ i „faktor“ ili manje uvrježenoj riječi domaćega podrijetla poput „raspodjela“, „učestalost“ i „čimbenik“. U odluci o tome hoćemo li upotrijebiti domaći termin ili internacionalizam odlučuju čimbenici poput prevladavajuće uporabe u struci, mogućnosti postizanja jednoznačnosti domaćega termina u određenome kontekstu, tvorbene mogućnosti domaće riječi i internacionalizma, semantičke implikacije u tvorbenim postupcima i sl. Nadalje, u izvještaju treba izbjegavati diskriminirajući jezik, odnosno uporabu izraza koji bi nekoga mogli uvrijediti ili omalovažiti. Pri uporabi kratica ne treba pretjerivati, pri čemu se kraticama treba koristiti dosljedno, odnosno kad se kratica jednom uvede (a treba je uvesti pri prvoj uporabi izraza izvan sažetka ili naslova), treba je svugdje upotrebljavati. Treba se pridržavati konvencija u pogledu pisanja brojeva, pri čemu je u engleskome jeziku najosnovnija ta da se brojevi do 10 pišu riječju, dok se oni veći od 10 pišu brojkom (uz pojedine iznimke). U hrvatskome se jeziku u znanstvenome tekstu mogu svi brojevi pisati brojkom.

Za upoznavanje s jezičnim i stilskim konvencijama u području psihologije Field i Hole (2003) preporučuju uporabu *Priručnika za objavljivanje Američkoga psihološkog udruženja*, koji je u ovome trenutku dostupan u svojem sedmom izdanju (American Psychological Association, 2020). Taj je priručnik uvrježen i u području usvajanja drugoga jezika, iako ne do te mjere kao u području psihologije, pa ga i mi preporučujemo za pomoć u pisanju izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika, posebice na engleskome jeziku. Uz nužne prilagodbe priručnik se

može primijeniti i na pisanje izvještaja na hrvatskome jeziku – primjer je za to način formatiranja popisa literature u ovoj knjizi. Za pomoć u pisanju izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika mogu poslužiti i opći priručnici za pisanje znanstvenih (i stručnih) radova, poput onoga na engleskome jeziku Gastel i Daya, trenutno dostupnoga u svojem devetom izdanju (Gastel i Day, 2022), i onih na hrvatskome jeziku Milice Gačić (Gačić, 2012, 2017). Uza svu pomoć koju takvi priručnici mogu pružiti pisanje dobrog izvještaja o istraživanju u pravilu je dugotrajan proces, koji uključuje proizvodnju višestrukih inačica izvještaja i temeljito dorađivanje i uređivanje teksta s ciljem postizanja što veće jasnoće i ekonomičnosti izražavanja.

U nastavku poglavlja dajemo detaljnije informacije o tome što treba sadržavati svaki od četiriju temeljnih dijelova izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika navedenih u prethodnome potpoglavlju, podsjećajući na temeljna pitanja na koja oni odgovaraju prema Field i Hole (2003, str. 288) i oslanjajući se ponajprije na taj izvor. U području usvajanja drugoga jezika ne postoji apsolutni konsenzus u pogledu sadržaja i strukture tih dijelova, odnosno mnogi objavljeni radovi odstupaju u većoj ili manjoj mjeri od onoga što opisujemo, no naš se pristup temelji na onome što smatramo da je dobra (a ujedno u dovoljnoj mjeri zastupljena) praksa.

6.3 Uvodni dio izvještaja: odgovor na pitanje „zašto“?

Svrha je uvoda uvesti čitatelje u temu eksperimentalnoga istraživanja i objasniti im zašto je istraživanje provedeno (Field i Hole, 2003). Iz toga proizlazi da uvod ima dvojaku svrhu. Kao prvo, omogućava čitateljima razumijevanje istraživanja u svim njegovim segmentima, odnosno pruža im adekvatnu teorijsku i empirijsku osnovu za razumijevanje istraživačkih pitanja i(li) hipoteza te evaluaciju dijelova izvještaja u kojima se opisuju i interpretiraju rezultati istraživanja. Kao drugo, objašnjava čitateljima zašto je istraživanje bilo vrijedno provedbe, odnosno na koji je način unaprijedilo znanstvene spoznaje o temi istraživanja.

Uvodni se dio izvještaja o eksperimentalnome istraživanju usvajanja drugoga jezika obično sastoji od dvaju glavnih dijelova: onoga koji možemo nazvati „uvod u užemu smislu riječi“ i pregleda teorijske i empirijske literature. Međutim ta dva dijela izvještaja najčešće samo u konceptualnome smislu čine jedinstvenu cjelinu. U strukturalnome su smislu oni u većini slučajeva odvojeni dijelovi izvještaja, među kojima samo prvi nosi naslov „Uvod“. Taj se dio izvještaja obično sastoji od četiriju dijelova: (1) opisa pozadine istraživačkoga

problema,¹⁴⁴ (2) definicije istraživačkoga problema, (3) pregleda provedenoga istraživanja i (4) pregleda strukture izvještaja.

Opis pozadine istraživačkoga problema sadrži kratak opis relevantnih teorijskih i empirijskih spoznaja, koje su proizašle iz prethodnih istraživanja. Definicija istraživačkoga problema sadrži opis konkretne „praznine“ u prethodnim (dotad u izvještaju opisanim) znanstvenim spoznajama koju provedeno istraživanje popunjava te objašnjenje zašto je popunjavanje te praznine važno. Popunjavanje „praznine“ u prethodnim spoznajama može poprimiti četiri oblika, koji odgovaraju četirima temeljnim razlozima zbog kojih Field i Hole (2003, str. 313) tvrde da se provode sva istraživanja: (1) kako bi testirala neku teoriju, (2) kako bi replicirala, odnosno potvrdila, postojeće nalaze, (3) kako bi proširila postojeće nalaze i (4) kako bi razriješla anomaliju, odnosno konfliktnu situaciju, koja je proizašla iz prethodnih istraživanja. Eksperimentalna istraživanja usvajanja drugoga jezika ne predstavljaju u tome iznimku. Iz rješavanja istraživačkoga problema proizlazi znanstveni doprinos istraživanja, pa je istraživački problem potrebno jasno i eksplicitno definirati. Iz definicije istraživačkoga problema trebaju biti vidljive zavisna/e i nezavisna/e varijabla/e u istraživanju (v. Porte, 2002).¹⁴⁵ Pregled provedenoga istraživanja, koji slijedi, može znatno varirati u opsegu i sadržaju, od toga da se u njemu iznose istraživačka pitanja i(li) hipoteze, glavne značajke metodologije, glavni rezultati, zaključci i njihove teorijske implikacije do toga da se u njemu navode samo ciljevi istraživanja, koji mogu, ali ne moraju biti popraćeni glavnim značajkama metodologije i(li) predviđanjima u pogledu rezultata. Neovisno o opsegu i sadržaju, u tome se dijelu izvještaja definiraju razine nezavisne varijable ili nezavisnih varijabli, dok se njihova operacionalizacija (v. potpoglavlje 2.1.2) navodi tek poslije. Koji god da se pristup primijeni, važno je čitateljima pružiti dovoljno informacija kako bi mogli pratiti ostatak izvještaja te kako bi ih se ujedno motiviralo za nastavak čitanja. Pregled strukture izvještaja sadrži kratak opis dijelova izvještaja koji slijede.

¹⁴⁴ U ovoj se knjizi ne bavimo istraživačkim problemom jer on nije izravno relevantan za statističku analizu podataka – za nju su relevantniji istraživačko pitanje i istraživačka hipoteza (v. potpoglavlje 2.1.1), koji proizlaze iz istraživačkoga problema. **Istraživački problem** (engl. *research problem*) jest „praznina“ u postojećim znanstvenim spoznajama koju istraživanje popunjava. Za detaljnije objašnjenje toga pojma v. npr. Phakiti (2014, str. 47, 326).

¹⁴⁵ Primjerice, u istraživanju koje su proveli Webb i sur. (2013), istraživački problem definiran je na sljedeći način: „Relativno se malo zna o tome kako se kolokacije uče. Istraživanja su pokazala da se kolokacije mogu učinkovito naučiti putem eksplicitnoga poučavanja [...]. Međutim, koliko nam je poznato, malo je istraživanja objavljeno o slučajnome učenju kolokacija putem unosa usredotočenoga na značenje.“ (prev. M. M. P. i T. K.) Iz te je definicije jasno da je zavisna varijabla u istraživanju slučajno učenje kolokacija, a nezavisna varijabla unos usredotočen na značenje (operacionaliziran kao broj pojavljivanja kolokacija u čitankama), te da se istraživanje provodi kako bi se proširili postojeći nalazi o učenju kolokacija.

Pregled teorijske i empirijske literature, koji slijedi, razrađuje teorijsku i empirijsku pozadinu istraživanja, koja je u kratkim crtama već opisana u „uvodu u užemu smislu riječi“. Dok je u „uvodu u užemu smislu riječi“ ta pozadina opisana samo onoliko koliko je potrebno da se definira istraživački problem, u ovome se dijelu ta pozadina opisuje onoliko detaljno koliko je potrebno za razumijevanje svih aspekata provedenoga istraživanja. To uključuje ne samo opis rezultata i osnovnih elemenata metodologije prethodnih relevantnih istraživanja već i definiranje svih ključnih pojmova u istraživanju (v. Porte, 2002). Pregled je obično strukturiran u više dijelova, ovisno o temi istraživanja, no u izvještajima o lingvistički orijentiranim eksperimentalnim istraživanjima usvajanja drugoga jezika, kojima se ovdje bavimo, obično sadrži dva temeljna dijela: dio posvećen opisu i teorijskoj analizi jezične pojave čije se usvajanje proučava, kako u ciljnome jeziku tako i u prvome jeziku ispitanika, te dio posvećen opisu prethodnih empirijskih spoznaja o usvajanju te pojave u ciljnome kao drugome jeziku, odnosno pregledu rezultata prethodnih istraživanja. Uz ta dva temeljna dijela pregled ponekad sadrži i opis empirijskih spoznaja o usvajanju jezične pojave u ciljnome kao prvome jeziku, koji obično prethodi opisu spoznaja koje se odnose na drugi jezik. Ti dijelovi pregleda mogu biti objedinjeni u jednu cjelinu nazvanu „Teorijska pozadina“, „Pregled literature“ ili „Prethodna istraživanja“, mogu samo neki od njih biti objedinjeni u cjelinu ili mogu svi stajati zasebno. U sadržajnome smislu, pregled teorijske i empirijske literature treba biti selektivan i usredotočen na istraživački problem, odnosno iz njega „praznina“ u prethodnim spoznajama treba postati još očiglednija, kao i potreba za provođenjem istraživanja o kojemu se izvještava (Porte, 2002). Prema Porte (2002), dobar pregled teorijske i empirijske literature predstavlja poveznicu između istraživačkoga problema i njegove pozadine s jedne strane i istraživačkih pitanja i(li) hipoteza istraživanja o kojemu se izvještava s druge. K tome dobar pregled ne sadrži samo opis prethodnih teorijskih i empirijskih spoznaja već i mjestimični kritički osvrt na njih.

6.4 Opis metode i postupka: odgovor na pitanje „kako“?

Svrha je dijela izvještaja posvećenoga metodi i postupku dati čitateljima jasnu ideju o tome kako je istraživanje provedeno (Field i Hole, 2003). To omogućava čitateljima da procijene prikladnost metodologije te pouzdanost i valjanost rezultata, kao i da, ako poželev, repliciraju istraživanje (Porte, 2002; za replikaciju v. potpoglavlje 7.1). U istraživanju postoji toliko prijetnji pouzdanosti i valjanosti (v. potpoglavlje 2.1.6) da ih istraživači ne mogu sve kontrolirati; ono što realno mogu učiniti jest pokazati da su svjesni postojanja

tih prijetnji te ih pokušati svesti na najmanju moguću mjeru (Porte, 2002, str. 36, 49). To ponajprije treba biti vidljivo u ovome dijelu izvještaja. U ovome dijelu izvještaja istraživači trebaju pokazati svoju osviještenost i za etička pitanja u provedbi istraživanja (v. npr. potpoglavlje 3.3. u Kraš i Miličević, 2015) pa, uz to što moraju poštovati etička načela u pogledu izvještavanja o istraživanju (v. u nastavku), poželjno je da navedu i da je istraživanje odobrilo etičko povjerenstvo za znanstvena istraživanja (ili neko ekvivalentno tijelo) njihove ustanove ako je to bio slučaj.

Taj se dio izvještaja sastoji od više dijelova: (1) dijela u kojemu se formuliraju istraživačka pitanja i(li) postavljaju hipoteze,¹⁴⁶ (2) opisa uzorka ispitanika, (3) opisa materijala i nacrt istraživanja, (4) opisa istraživačkoga postupka i (5) opisa analize podataka. Svaki od tih dijelova izvještaja obično nosi svoj naslov – redom navođenja „Istraživačka pitanja i hipoteze“, „Sudionici“, „Materijali“, „Postupak“ i „Analiza podataka“ – a svi su obično objedinjeni u cjelinu nazvanu „Istraživanje“, „Provedeno istraživanje“ ili „Eksperiment“. Takva je struktura znanstvenoga članka u kojemu se izvještava o istraživanju Heil i López (2020). Moguća su i drukčija rješenja, poput znanstvenoga članka u kojemu se izvještava o istraživanju Ahn (2021), gdje je dio s istraživačkim pitanjima i predviđanjima pripojen (prethodnome) dijelu članka u kojemu se daje pregled teorijske i empirijske literature, dok su dijelovi članka posvećeni ispitanicima, materijalima i postupku te analizi podataka objedinjeni u zasebni dio članka, u kojemu su materijali i postupak prikazani zajedno. Moguće je i razdvajanje prikaza materijala i nacrt istraživanja, kao u znanstvenome članku u kojemu se izvještava o istraživanju Mitterer i sur. (2020). Dio izvještaja pod nazivom „Istraživanje“, „Provedeno istraživanje“ ili „Eksperiment“ ponekad sadrži ne samo dio posvećen metodi i postupku već i dio posvećen rezultatima, opisan u sljedećemu potpoglavlju, kao u znanstvenome članku u kojemu se izvještava o istraživanju Margaza i Gavarró (2020).

Istraživačka pitanja i(li) hipoteze ključni su element za razumijevanje veze između istraživačkoga problema i načina na koji su ga istraživači odlučili riješiti; oni daju istraživanju smjer i oblik (Porte, 2002). U tome su smislu posebno važne hipoteze, odnosno tvrdnje o postojanju povezanosti ili razlika između varijabli, koje se istraživanjem provjeravaju. Istraživačka se pitanja također tiču povezanosti ili razlika između varijabli. Iznimno je važno da istraživačka pitanja i hipoteze budu jasno, precizno i nedvosmisleno formulirani. Iz njihove formulacije treba biti jasno koje su varijable zavisne, a koje nezavisne te

¹⁴⁶ Za razliku od područja usvajanja drugoga jezika, u području psihologije istraživačka pitanja i hipoteze obično se navode u uvodnome dijelu izvještaja.

koliko razina imaju; u nekim je slučajevima također jasno kako su varijable operacionalizirane, odnosno kako će biti mjerene. Kao što je objašnjeno u prethodnim poglavljima, te informacije imaju važne implikacije za statističku analizu podataka. Primjeri formulacije istraživačkih pitanja i(li) hipoteza, preuzetih iz istraživanja Kim i McDonough (2008), Heil i López (2020) te Kim i sur. (2020), dani su u potpoglavlju 2.1.1, u kojemu su ti dijelovi istraživačkoga procesa i objašnjeni.

U opisu uzorka ispitanika navode se značajke ispitanika, prvenstveno demografske, poput spola, dobi, stupnja obrazovanja i obrazovnoga profila, ali i one koje imaju veze s njihovom jezičnom biografijom, poput prvoga jezika, dobi na početku usvajanja drugoga jezika, konteksta usvajanja jezika i ostalih jezika koje govore (za prikupljanje podataka o tim značajkama ispitanika v. npr. potpoglavlje 3.2 u Kraš i Miličević, 2015). U velikome se broju istraživanja navode podaci o razini općega znanja drugoga jezika ispitanika jer se na temelju tih podataka ispitanici dijele u skupine (kad je razina znanja jedna od nezavisnih varijabli) ili uključuju u uzorak (ako se razina znanja kontrolira). Uz te se podatke navodi i način na koji je razina općega znanja drugoga jezika bila izmjerena (za određivanje razine općega znanja drugoga jezika ispitanika v. npr. potpoglavlje 3.1 u Kraš i Miličević, 2015). Početnici u izvještavanju o eksperimentalnome istraživanju usvajanja drugoga jezika često podatke vezane uz značajke ispitanika navode u dijelu izvještaja posvećenome rezultatima – bolje ih je međutim navoditi ovdje. Koje se sve značajke ispitanika navode, ovisi o značajkama istraživanja. Naprimjer u istraživanjima u kojima se upotrebljava tehnika praćenja pokreta očiju važno je navesti da su svi ispitanici imali normalan ili korigiran vid, kao što je navedeno u istraživanju Felser i sur. (2009).

Obavezno se navodi broj ispitanika, kako u cijelome uzorku tako i u pojedinim skupinama (ako postoje). Ako je tijekom provedbe istraživanja došlo do smanjenja broja ispitanika, to se također navodi, uz objašnjenje zašto je do toga došlo. Opisuje se i postupak odabira, tj. uzorkovanja, ispitanika (je li bio slučajan ili nije), njihove regrutacije (jesu li ispitanici regrutirani putem oglasa, poziva, postupkom lančanoga uzorkovanja i sl.), kao i njihova raspoređivanja u eventualne skupine (je li bilo slučajno ili s obzirom na neku varijablu, poput prvoga jezika, razine općega znanja drugoga jezika, dobi na početku usvajanja drugoga jezika i sl.) (za sva ta pitanja v. potpoglavlje 2.1.3). Uobičajeno je navesti i jesu li ispitanici primili novčanu ili neku drugu nagradu za sudjelovanje u eksperimentu. Zbog zaštite privatnosti ispitanika, odnosno poštovanja etičkih načela provedbe istraživanja, podaci se o nazivu ustanove

u kojoj su ispitanici regrutirani (ako je primjenjivo) ne navode (osim ako ustanova to ne želi), kao ni drugi podaci na temelju kojih bi se ispitanike moglo identificirati. Numerički podaci o ispitanicima obično se prikazuju tablično.

U opisu materijala i nacrtu istraživanja opisuju se mjerni instrumenti kojima su podaci prikupljeni, nacrt istraživanja (v. potpoglavlje 2.1.5) i podražaji (v. potpoglavlje 2.1.4). Iz opisa mjernoga instrumenta proizlazi kako su varijable u istraživanju bile operacionalizirane, odnosno kako su bile izmjerene (v. potpoglavlje 2.1.2). Mjerni instrumenti u lingvistički orijentiranim istraživanjima usvajanja drugoga jezika jesu eksperimentalni zadaci koje su ispitanici rješavali (za opis eksperimentalnih zadataka v. potpoglavlje 2.2.1, kao i četvrto poglavlje u Kraš i Miličević, 2015). Uz navođenje vrste eksperimentalnoga zadatka u izvještaju se može navesti i izvor, odnosno rad u kojemu je ta vrsta zadatka opisana, uvedena ili upotrijebljena, posebice ako se radi o inovativnoj ili manje proširenoj vrsti zadatka (npr. ubrzani zadatak prosudbe prihvatljivosti rečenica ili test dopunjavanja riječi).

Kao što smo već spomenule u potpoglavlju 2.3, danas se u području usvajanja drugoga jezika zagovara transparentnost u navođenju značajki mjernih instrumenata (v. npr. Al-Hoorie i Vitta, 2019; Derrick, 2016; Plonsky i Derrick, 2016; Woodrow, 2014). Točnije, preporučuje se navođenje podataka o podrijetlu mjernoga instrumenta (jesu li ga istraživači izradili sami ili su ga preuzeli od nekoga drugog i eventualno prilagodili), o njegovu pilotiranju (v. potpoglavlje 2.1.6), o njegovim psihometrijskim značajkama, tj. o njegovoj pouzdanosti (uz navođenje metode kojom je izračunata; v. potpoglavlje 2.3), a ujedno se preporučuje prilaganje radovima cjelokupnih instrumenata (ako nisu standardizirani), u dodacima radova, na internetu i na druge načine.¹⁴⁷ Rezultati psihometrijske analize ponekad se navode u dijelu izvještaja posvećenome metodi i postupku, a ponekad u dijelu izvještaja posvećenome rezultatima (v. sljedeće potpoglavlje).

Što se tiče opisa nacrtu istraživanja, radi se o opisu strukture eksperimenta s obzirom na broj, vrstu varijabli i njihove razine. U njemu se navodi koliko je varijabli sadržavao eksperiment, koje su bile zavisne, a koje nezavisne (iako se taj podatak često implicitno iščitava iz opisa, a ne navodi eksplicitno) te koliko je razina imala koja varijabla. Navode se i ilustriraju svi eksperimentalni uvjeti u zadatku (v. potpoglavlje 2.1.4), odnosno daje se primjer podražaja za sve eksperimentalne uvjete. Primjer podražaja za jedan eksperimentalni uvjet u istraživanju Roberts i Liszka (2013) naveden je u potpoglavlju 2.1.4, dok

¹⁴⁷ Jedna je od mogućnosti i njihovo objavljivanje u javno dostupnoj bazi mjernih instrumenata, podražaja i podataka IRIS (v. potpoglavlje 2.2.2, posebice napomenu 45).

su primjeri za sve eksperimentalne uvjete u istraživanju dani na 421. stranici znanstvenoga članka u kojemu se izvještava o tome istraživanju. Poželjno je ilustrirati i distraktore (v. potpoglavlje 2.1.4), posebice ako je i vrsta distraktora u eksperimentu bila sustavno manipulirana, kao npr. u istraživanju Heil i López (2020); na stranicama 430–432 znanstvenoga članka posvećenoga tome istraživanju autori daju primjere svih vrsta eksperimentalnih podražaja i distraktora u istraživanju. Broj podražaja u zadatku ne smije se izostaviti u ovome dijelu izvještaja, kako eksperimentalnih tako i distraktora. Potpuni popis podražaja poželjno je priložiti u dodatku radu. Navodi se i je li se radilo o nacrtu s nezavisnim uzorcima, nacrtu sa zavisnim uzorcima ili mješovitome nacrtu (v. potpoglavlje 2.1.5), iako te nazive nije nužno izrijekom spomenuti. U opisu nacrtu istraživanja najvažnije je uspostaviti jasnu vezu između istraživačkih pitanja i(li) hipoteza, varijabli te načina mjerenja varijabli (Woodrow, 2014).

U opisu istraživačkoga postupka opisuje se način na koji su podaci prikupljeni, odnosno koraci koji su poduzeti ili faze u kojima se to dogodilo i njihov redosljed. Ako se prikupljanje podataka vršilo u više navrata ili primjenom više mjernih instrumenata, navodi se koliko je svaka faza ili primjena svakoga od instrumenata približno trajala. Važno je spomenuti i jesu li svi ispitanici sudjelovali u različitim fazama eksperimenta istim redosljedom ili se redosljed mijenjao za različite ispitanike. Navodi se i jesu li ispitanici testirani grupno ili individualno te kakve su upute dobili za rješavanje zadataka, odnosno što se od njih točno očekivalo.¹⁴⁸ Opisuju se i prostorni uvjeti u kojima su se podaci prikupili te relevantne značajke ambijenta (npr. da je prostorija bila tiha ili zvučno izolirana ako se u eksperimentu mjerila brzina reakcije). Iz etičkih razloga, kao i kod opisa postupka regrutacije ispitanika, podaci o nazivu ustanove u kojoj su ispitanici testirani (ako je primjenjivo) ne navode se, već se navode njezine značajke, npr. „javna visokoškolska ustanova u većemu hrvatskom gradu“. Ponekad se spominje i godina ili razdoblje u kojemu su podaci prikupljeni. Navode se i podaci o načinu prikupljanja podataka te eventualnoj opremi koja se upotrebljavala prilikom njihova prikupljanja (npr. je li se zadatak dopunjavanja rečenica rješavao na papiru ili na računalu; ako se rješavao na računalu, koja je bila marka računala). Daju se i tehničke specifikacije zadataka koji su se rješavali na računalu (npr. u ubrzanome zadatku prosudbe prihvatljivosti rečenica kojim su fontom i kojom veličinom fonta riječi bile prikazane na ekranu te koliko se dugo svaka riječ zadržavala na ekranu). U tim je zadacima važno spomenuti i jesu li svi ispitanici bili izlagani podražajima istim redosljedom ili je redosljed podražaja bio različit (i na koji

¹⁴⁸ Ako su upute duže, mogu se priložiti u dodatku radu.

način) za različite ispitanike. Ako su se podaci prikupljali putem interneta, to se također navodi, uz objašnjenje kako. Naposljetku, navodi se i je li podatke prikupljalo više osoba.

U opisu analize podataka istraživači objašnjavaju kako su podatke pripremili za analizu, koje su statističke testove proveli i kojim su se statističkim softverom koristili. Kao što smo već navele, priprema podataka za analizu uključuje različite načine postupanja s nedostajućim podacima (v. potpoglavlje 2.2.2) kao i različite načine identifikacije i tretmana netipičnih vrijednosti (v. potpoglavlje 3.1.6), što se sve opisuje u ovome dijelu izvještaja. U okviru opisa pripreme podataka za analizu navodi se i način bodovanja rezultata eksperimentalnoga zadatka (ako je primjenjivo; npr. zadatka dopunjavanja teksta ili riječi). Konačno, u ovome se dijelu izvještaja navode rezultati provjere parametrijskih uvjeta u svrhu dokaza da su pretpostavke za uporabu pojedinih statističkih testova ispunjene (v. potpoglavlje 4.3.2), uz napomenu da se ti rezultati ponekad iznose u dijelu izvještaja posvećenome rezultatima (v. potpoglavlje 6.5.2). Kao što smo već navele, važnost navođenja rezultata provjere parametrijskih uvjeta ističu npr. Al-Hoorie i Vitta (2019) te Hu i Plonsky (2021). Opis analize podataka treba biti dovoljno iscrpan da omogući čitateljima prosudbu o kvaliteti analize i replikaciju; manjkavi opis analize podataka Wood i sur. (2024) uključuju u popis dvojbeneh istraživačkih praksi.

6.5 Opis rezultata: odgovor na pitanje „što“?

Svrha je dijela izvještaja posvećenoga rezultatima opisati čitateljima što se otkrilo u istraživanju (Field i Hole, 2003). To je stožerni dio rada (Day i Gastel, 2014), u kojemu se navode rezultati statističke analize podataka, u većini slučajeva prethodno opisane u dijelu izvještaja posvećenome opisu metode i postupka (v. prethodno potpoglavlje) i daje inicijalna interpretacija tih rezultata. Rezultate treba shvatiti kao posljedicu istraživačkoga konteksta, ciljeva istraživanja, ispitanika koji su odabrani za sudjelovanje u istraživanju i instrumenata koju su upotrijebljeni za prikupljanje i obradu podataka.

Navođenje rezultata statističke analize podataka obuhvaća rezultate i deskriptivne i inferencijalne statistike. Kao što je rečeno u prvome poglavlju, deskriptivna statistika opisuje rezultate dobivene u ispitanome uzorku (v. i treće poglavlje), dok inferencijalna statistika omogućava da se iz rezultata uzorka izvedu zaključci o razmatranoj populaciji (v. i četvrto i peto poglavlje). Prikaz rezultata obuhvaća i rezultate psihometrijske analize, koja pokazuje stupanj pouzdanosti korištenoga instrumenta (v. potpoglavlje 2.3). Kao što je već rečeno u prethodnome potpoglavlju, ti se rezultati ponekad navode u

dijelu izvještaja posvećenome metodi i postupku, a ponekad u dijelu izvještaja posvećenome rezultatima.

Temeljno je načelo navođenja rezultata statističke analize podataka **transparentnost** (v. Larson-Hall i Plonsky, 2015; Lindstromberg, 2016a). To znači da rezultati trebaju biti navedeni jasno i precizno, a po mogućnosti i iscrpno.¹⁴⁹ Jedna je od svrha transparentnoga navođenja rezultata omogućavanje replikacije istraživanja te provjere dobivenih rezultata (v. potpoglavlje 7.1). Nedovoljno iscrpno navođenje rezultata analize podataka također je prema Wood i sur. (2024) dvojbena istraživačka praksa. Pri navođenju rezultata treba ipak biti selektivan kako se čitatelje ne bi opteretilo suvišnim informacijama. Potrebno je iznijeti samo rezultate koji su izravno relevantni za istraživački problem i za istraživačka pitanja i(li) hipoteze (Field i Hole, 2003). Rezultate treba iznijeti logičnim slijedom, naprimjer od općenitijih prema manje općenitima (Pavlović i Stanojević, 2020). Pri navođenju rezultata treba se služiti grafičkim prikazom podataka (v. potpoglavlje 3.2) kako bi čitatelji uvidjeli odnose te sličnosti i razlike između varijabli. Također, treba se koristiti uvriježenim statističkim simbolima i kraticama (v. npr. Field i Hole, 2003, str. 335, za njihov selektivni popis).

6.5.1 Rezultati deskriptivne statistike

Kao što je prethodno rečeno, rezultati deskriptivne statističke analize opisuju podatke dobivene u uzorku. Oni predstavljaju temelj za navođenje rezultata inferencijalne analize podataka te omogućavaju replikaciju istraživanja (Woodrow, 2014). U eksperimentalnim istraživanjima usvajanja drugoga jezika ti se podaci mogu odnositi na demografske značajke ispitanika prikupljene sociolingvističkim upitnikom, na rezultate testa razine općega znanja drugoga jezika ili pak na jezično ponašanje ispitanika ispitano eksperimentalnim zadatkom (v. potpoglavlje 2.2.1). Rezultati deskriptivne analize navode se u dijelu izvještaja posvećenome metodi i postupku te u dijelu izvještaja posvećenome rezultatima. U dijelu izvještaja posvećenome metodi i postupku rezultati se obično odnose na demografske značajke ispitanika i veličinu uzorka te na razinu općega znanja drugoga jezika (v. potpoglavlje 6.4), dok se u dijelu rada (eksplicitno) posvećenome rezultatima oni obično odnose na jezično ponašanje ispitanika.

¹⁴⁹ Mjera u kojoj će rezultati biti navedeni iscrpno ovisi o vrsti izvještaja o istraživanju jer različite vrste izvještaja postavljaju različita prostorna ograničenja. Tako će naprimjer doktorska disertacija ili monografija omogućiti detaljnije i opširnije navođenje rezultata od članka u znanstvenome časopisu, u kojemu istraživači imaju manje prostora na raspolaganju. Dio rezultata može biti naveden i u dodatku radu, posebice ako se on može objaviti na internetu.

Rezultati deskriptivne statistike mogu obuhvaćati frekvencijske mjere, mjere centralne tendencije, mjere varijabilnosti i položajne mjere (v. potpoglavlje 3.1). Istraživači odlučuju o tome koje će rezultate deskriptivne statistike iznijeti na temelju svrhe rada, ciljnoga čitateljstva, kompleksnosti inferencijalne statističke analize i distribucije frekvencija u uzorku (Woodrow, 2014). Što se tiče distribucije frekvencija, ako su podaci normalno distribuirani, obično je dovoljno navesti aritmetičku sredinu (v. potpoglavlje 3.1.2) i standardnu devijaciju (v. potpoglavlje 3.1.3). Ako podaci nisu normalno distribuirani ili sadrže netipične vrijednosti, treba navesti mjere poput medijana, moda (za obje mjere v. potpoglavlje 3.1.2), interkvartilnoga raspršenja (v. potpoglavlje 3.1.3), asimetričnosti i spljoštenosti (v. potpoglavlje 3.1.5).

Što se tiče samoga prikaza rezultata, navođenje mjera centralne tendencije (aritmetičke sredine, medijana i moda) mora uvijek biti popraćeno navođenjem mjera varijabilnosti (standardne devijacije ili interkvartilnoga raspršenja). Pri navođenju numeričkih vrijednosti obično se navode dva ili tri decimalna mjesta.

Rezultati se mogu iznijeti u tekstu, u tablici ili putem grafikona. Glavna je svrha tablice ili grafikona olakšati razumijevanje rezultata analize. Način prikaza rezultata ovisi o količini podataka i uvriježenoj praksi u navođenju određene vrste rezultata. Ako je podataka relativno malo, iznose se u tekstu; ako ih je više, prikazuju se u tablici ili putem grafikona. Uobičajeno je da se demografski podaci o ispitanicima prikazuju tablično, a rezultati jezičnoga ponašanja ispitanika grafički. Ako se rezultati analize iznose u tekstu, nije ih potrebno iznijeti i u tablici ili putem grafikona. U tekstu međutim treba navesti da su određeni rezultati prikazani u tablici ili putem grafikona te čitatelje uputiti na njih. Tablice i grafikoni moraju biti jasno numerirani, redoslijedom kojim se spominju u tekstu. Tablice i grafikoni numeriraju se odvojeno.

Kao što je već rečeno u potpoglavlju 3.2, grafikone treba jasno obilježiti (tu se ponajprije misli na označavanje varijabli) te pažljivo odabrati raspon veličina koje će biti prikazane na osi y kako bi eventualne razlike bile vjerodostojno prikazane. Razmatranja vezana uz odabir vrste grafikona te smjernice za njihovu izradu navedeni su u potpoglavlju 3.2.

6.5.2 Rezultati inferencijalne statistike

Rezultati inferencijalne statističke analize omogućavaju nam da iz rezultata uzorka izvedemo opće zaključke, koji se odnose na cijelu populaciju. U toj se analizi provjerava postoje li statistički značajne razlike između skupina ispitanika ili eksperimentalnih uvjeta (v. npr. Field i Hole, 2003). Kao što je

navedeno u potpoglavlju 4.1.2, statistički značajne razlike jesu one za koje se može pretpostaviti da nisu nastale slučajno, odnosno zbog netipičnosti uzorka (Porte, 2002).

Rezultati inferencijalne statistike mogu se odnositi na rezultate parametrijskih i neparametrijskih testova (v. potpoglavlje 4.3.2). Rezultati tih dviju vrsta testova navode se na isti način. Jedina je razlika u tome što navođenju rezultata neparametrijskih testova prethodi objašnjenje za njihovu uporabu. Prije navođenja rezultata obiju vrsta testova navode se rezultati provjere parametrijskih uvjeta. To se čini u ovome dijelu izvještaja ili u dijelu izvještaja posvećenome metodi i postupku, i to u onome dijelu u kojemu se opisuje analiza podataka (v. potpoglavlje 6.4). Izvor u kojemu je neki statistički test uveden ili opisan ne navodi se, osim kad se radi o rijetkim i neuobičajenim testovima ili kad je sam test u središtu interesa istraživanja (Field i Hole, 2003).

Kao i rezultati deskriptivne statističke analize, i rezultati inferencijalne statističke analize mogu se navesti u tekstu, u tablici ili putem grafikona. Field i Hole (2003) prenose preporuku Američkoga psihološkog udruženja prema kojoj se do tri broja navode u tekstu, od četiri do 20 u tablici, a više od 20 putem grafikona. Bez obzira na to na koji su način prikazani numerički rezultati inferencijalne statistike, u tekstu se mora objasniti što oni pokazuju.

Navođenje numeričkoga rezultata svakoga inferencijalnog statističkog testa osim vrijednosti p , koja predstavlja podatak o statističkoj značajnosti rezultata (v. potpoglavlje 4.1.2), uključuje još niz podataka koji se obavezno navode u izvještajima o istraživanju. Za svaki test postoji koeficijent ili testni statistik, označen uvriježenom kraticom, koji se dobiva odgovarajućom formulom (naprimjer koeficijent r za korelaciju ili t -vrijednost za t -test). Za većinu testova navode se i stupnjevi slobode (v. potpoglavlje 4.2.1), a za neke veličina uzorka (npr. korelaciju). Za sve testove treba navesti i veličinu efekta (v. potpoglavlje 4.1.3). Ona se donedavno nije (sustavno) navodila u izvještajima o eksperimentalnim istraživanjima usvajanja drugoga jezika, no njezino je navođenje vrlo važno (v. npr. Larson-Hall i Plonsky, 2015). Također se preporučuje navođenje granica pouzdanosti (v. potpoglavlje 4.1.4) (v. npr. Larson-Hall i Plonsky, 2015; Woordow, 2014). U petome poglavlju navodimo koje sve podatke treba dati pri navođenju rezultata pojedinih inferencijalnih statističkih testova.

Što se tiče vrijednosti p , preporučuje se navesti točnu numeričku vrijednost (npr. $p=0,034$), osim za jako male vrijednosti, kada se navodi $p<0,001$ (v. npr. Larson-Hall i Plonsky, 2015). Prethodna je praksa bila usporediti vrijednost p s prihvaćenom razinom statističke značajnosti, koja u području usvajanja

drugoga jezika iznosi $\alpha=0,05$ (v. potpoglavlje 4.1.2), bez navođenja točne numeričke vrijednosti (npr. $p<0,05$). Kao što smo već navele u potpoglavlju 4.1.2, preporučuje se i eksplicitno navesti odabranu razinu statističke značajnosti (v. npr. Larson-Hall i Plonsky, 2015).

Važno je naglasiti da se rezultati inferencijalne statističke analize ne navode samo onda kada testovi daju statistički značajne rezultate već i onda kada to nije slučaj. Nenavođenje rezultata testova koji ne daju statistički značajne rezultate Wood i sur. (2024) smatraju dvojbena istraživačkom praksom.

Što se tiče načina zapisivanja decimalnih brojeva u numeričkim rezultatima inferencijalne statistike, u potpoglavlju 2.2.2 već smo objasnile da različiti jezici imaju različite konvencije u pogledu uporabe decimalne oznake i pisanja nule u decimalnim brojevima manjima od jedan. Ondje smo navele da se u hrvatskome kao decimalna oznaka upotrebljava zarez (ali i točka), a u engleskome decimalna točka. Također smo navele da je u engleskome uobičajeno izostaviti nulu u decimalnim brojevima kod vrijednosti koje se kreću između 0 i 1 ili između -1 i 0, kao naprimjer kod koeficijenta korelacije (koji se kreće od -1 do $+1$) ili kod vrijednosti p (koja ne može biti veća od 1) (npr. $p<0,05$), dok se u hrvatskome nula ponekad izostavlja, a ponekad ne. Što se tiče broja decimalnih mjesta, Field i Hole (2003) navode preporuku Američkoga psihološkog udruženja prema kojoj je broj decimalnih mjesta poželjno zaokružiti na dva, no ističu da je u nekim situacijama prikladno navesti i više decimalnih mjesta.

6.6 Rasprava o rezultatima i zaključci: odgovor na pitanje „pa što“?

Svrha je dijela izvještaja koji sadrži raspravu i zaključke objasniti čitateljima što dobiveni rezultati zapravo znače, u odnosu na istraživačka pitanja i(li) hipoteze te prethodne teorijske i empirijske spoznaje (Field i Hole, 2003). Day i Gastel (2014, str. 73) navode da uvodni dio rada funkcionira u paru s raspravom i zaključcima, kao što dio posvećen metodi i postupku odgovara onom s rezultatima. To je tako jer rasprava i zaključci, koji su ponekad spojeni u jednu cjelinu (kao u npr. Heil i López, 2020), a ponekad odvojeni (kao u npr. Ahn, 2021), trebaju odgovoriti na pitanja koja su postavljena u uvodu. Napominjemo međutim da se, kao što smo navele u potpoglavlju 6.4, u izvještaju o eksperimentalnom istraživanju usvajanja drugoga jezika istraživačka pitanja i hipoteze obično navode u dijelu rada u kojemu se opisuju metoda i postupak. Povezanost rasprave i zaključaka s uvodom proizlazi i iz toga što rasprava i zaključak trebaju objasniti čitateljima na koji se način

istraživanjem pridonijelo rješavanju istraživačkoga problema, koji je definiran u uvodu (v. npr. Porte, 2002; Pavlović i Stanojević, 2020). Povezanost rasprave i zaključaka s uvodom ističe cikličku prirodu istraživanja (Porte, 2002, str. 139).

Kako navode Field i Hole (2003), na početku rasprave uobičajeno je (još jednom) sažeti najvažnije rezultate istraživanja, ne upotrebljavajući statističke nazive i način izražavanja. Nakon toga ti se rezultati interpretiraju, odnosno tumače. Rezultati se povezuju s istraživačkim pitanjima i(li) hipotezama – daju se odgovori na istraživačka pitanja i(li) navodi jesu li hipoteze potvrđene ili ne. Nadalje, rezultati se povezuju s rezultatima prethodnih istraživanja, spomenutih u uvodnome dijelu izvještaja. Razmatra se jesu li rezultati u skladu ili u suprotnosti s njima. Ako su u suprotnosti, predlažu se objašnjenja. Prethodni i vlastiti nalazi pokušavaju se povezati u koherentnu „priču“. U ovome se dijelu izvještaja također navode i tentativno objašnjavaju eventualni neočekivani rezultati, kao naprimjer ako neka varijabla nije utjecala na rezultate na očekivani način. Uz to rezultati se tumače u svjetlu teorijskih spoznaja. Razmatra se do koje mjere rezultati potvrđuju ili opovrgavaju jednu ili više teorija, odnosno koje su njihove teorijske implikacije. Ponekad se također upućuje na praktične implikacije rezultata, a u izvještaju o istraživanju usvajanja drugoga jezika to su, najčešće, pedagoške implikacije. To znači da se objašnjava zašto i kako su rezultati važni za učenje i poučavanje jezika. U raspravi istraživači zapravo objašnjavaju koji je doprinos njihova istraživanja – do kojih su novih spoznaja došli, kako se one uklapaju s prethodnim spoznajama i kako to sve zajedno produbljuje naše razumijevanje proučavane pojave (Field i Hole, 2003).

Interpretacija rezultata (eksperimentalnih) istraživanja usvajanja drugoga jezika poprilično je kompleksna iz više razloga. Jezični su podaci kompleksni za analizu jer podliježu različitim kontekstualnim i individualnim izvorima varijabilnosti i mogu imati vrlo različite distribucije. Dodatna razina kompleksnosti proizlazi iz varijabilnosti učeničkoga jezika, posebno u kombinaciji s praktičnim poteškoćama u pronalasku većega broja ispitanika.

U (eksperimentalnim) istraživanjima usvajanja drugoga jezika statistički značajne razlike između kontrolne skupine i eksperimentalnih skupina mogu se javiti iz različitih razloga. Pri interpretaciji tih rezultata treba svratiti pozornost na to razlikuju li se odgovori skupina samo u kvantitativnome pogledu ili i kvalitativno. Naprimjer u zadacima prosudbe prihvatljivosti rečenica učenici drugoga jezika često pokazuju tendenciju da rjeđe od izvornih govornika upotrebljavaju ekstremne vrijednosti na ljestvici, vjerojatno zbog nesigurnosti u svoje znanje (v. npr. Chaudron, 2003, str. 806–807). To može dovesti do statistički značajnih razlika između skupine učenika i izvornih govornika.

Međutim, dokle god se odgovori kreću u istome smjeru, odnosno obje skupine ispitanika prihvaćaju prihvatljive i odbacuju neprihvatljive rečenice, može se govoriti o razlici koja, iako statistički značajna, ne mora nužno biti teorijski i praktično značajna. Kod prosudbi prihvatljivosti rečenica učenici mogu pokazivati i tendenciju ka konzervativnosti, odnosno k davanju više „negramatičnih“ ocjena (v. Paolillo, 2000; Norris i Ortega, 2003, str. 746–747). U svakome slučaju, u istraživanjima usvajanja drugoga jezika potreban je poseban oprez prilikom interpretacije rezultata i nužno je uvijek imati na umu značajke učeničkoga jezika koje mogu utjecati na rezultate.

Pri kraju svakoga izvještaja o istraživanju, najčešće u dijelu posvećenome zaključku (ako je izdvojen), obično se navode nedostaci istraživanja, kao i njihov utjecaj na nalaze. Nedostaci mogu biti metodološke ili koncepcijske prirode, a svrha je njihova navođenja predvidjeti i unaprijed odgovoriti na moguće kritike čitatelja, odnosno uvjeriti ih da je usprkos nedostacima istraživanje bilo vrijedno provedbe (Field i Hole, 2003). Nakon navođenja nedostataka, na temelju njih ili neovisno o njima, uobičajeno je dati smjernice za buduća istraživanja. U njima se navode neriješena pitanja i predlažu načini njihova rješavanja. Field i Hole (2003) predlažu da rasprava završi zaključcima o tome što je istraživanje pokazalo, ali bez ponavljanja onoga što je već rečeno i uporabe klišeiziranih fraza. Zaključci trebaju biti usklađeni s rezultatima, odnosno proizlaziti iz njih.

7. Trendovi i perspektive u statističkoj analizi podataka

U šest poglavlja ove knjige prikazane su različite faze procesa analize podataka u eksperimentalnim istraživanjima usvajanja drugoga jezika. Nakon uvodnih razmatranja u prvome poglavlju u drugome smo poglavlju podsjetile na opće metodološke postavke lingvistički orijentiranih eksperimenata kojima se ispituje usvajanje drugoga jezika, osvrćući se na formuliranje istraživačkih pitanja i hipoteza, definiranje i mjerenje varijabli, uzorkovanje ispitanika i podražaja, izradu nacrtu istraživanja i druge važne elemente osmišljavanja istraživanja. Pozornost smo usmjerile i na pripremu podataka za analizu i na početne korake u provedbi analize, u okviru provjere pouzdanosti korištenoga mjernog instrumenta. Treće je poglavlje bilo posvećeno izračunavanju frekvencijskih mjera, mjera centralne tendencije poput aritmetičke sredine, i mjera varijabilnosti. Zadržale smo se i na grafičkome prikazu podataka, da bismo potom u četvrtome i petome poglavlju prešle na inferencijalne statističke testove. Objasnile smo osnove frekvencijskoga pristupa testiranju statističke značajnosti, detaljnije predstavile nekoliko često korištenih testova, kao i dodatne mjere koje se uz njih izračunavaju i navode. Konačno, šesto je poglavlje bilo posvećeno izvještavanju o eksperimentalnim istraživanjima usvajanja drugoga jezika. Cilj je takva izbora tema bio postizanje zaokružene slike procesa statističke analize podataka, s isticanjem veza između njegovih različitih elemenata.

Naš je naglasak bio na osnovnim statističkim testovima koji pripadaju paradigmi utemeljenoj na testiranju nulte hipoteze i određivanju statističke značajnosti, ali su u knjigu uključena i kritička razmatranja te paradigme. Takav pristup poprilično vjerno oslikava trenutnu situaciju u istraživanjima usvajanja drugoga jezika, pa i u empirijskim znanostima šire gledano. Testiranje statističke značajnosti i dalje je vrlo prisutno, ali prisutne su i mnogobrojne njegove kritike i pokušaji da se taj model statističke analize unaprijedi uključivanjem novih vrsta testova i inzistiranjem na navođenju dodatnih mjera. U ovome ćemo se zaključnom poglavlju dodatno pozabaviti tom problematikom. Cilj nam je pokušati dati kratak prikaz toga kamo u širem kontekstu empirijskih znanstvenih istraživanja idu lingvistika i njezina poddisciplina usvajanje drugoga jezika, kao i navesti nekoliko mogućih smjernica za buduće korake. U takvu kontekstu razmatramo i nekoliko dodatnih metoda kontrole kvalitete istraživanja, koje su namjenski uključene ovamo umjesto u drugo poglavlje jer se odnose upravo na trendove i perspektive.

7.1 Dodatni aspekti kontrole kvalitete istraživanja

U cijeloj knjizi, a posebice u potpoglavljima 2.1.6 i 2.3, posvećenima različitim aspektima kontrole kvalitete istraživanja, naglašena su razmatranja i postupci kojima se unutar jednoga istraživanja mogu osigurati njegova valjanost i pouzdanost. Ovdje vraćamo pozornost na postupke koji izlaze iz okvira onoga što individualni istraživači trebaju i mogu učiniti jer zahtijeva suradnju šire znanstvene zajednice.

Kako je već spomenuto u potpoglavlju 2.1.6, jedna je od važnih komponenata kontrole kvalitete istraživanja vanjska valjanost, koja se odnosi na mogućnost uopćavanja rezultata izvan ispitanoga uzorka. Vanjska se valjanost eksperimenta može provjeriti **replikacijom** (engl. *replication*; v. npr. Abbuhl, 2011; Abbuhl i sur., 2013, str. 116), odnosno ponavljanjem istraživanja u identičnome ili djelomično izmijenjenome obliku. Naprimjer Porte (2002, str. 35–36) navodi da uopćavanje rezultata s uzorka na populaciju postaje opravdanije kad se u repliciranim istraživanjima, provedenima s drugim ispitanicima i/ili u drugim kontekstima, dobiju slični rezultati kao u izvornome istraživanju. Kako navodimo i u Kraš i Miličević (2015, str. 41), važnost replikacije u području usvajanja drugoga jezika vrlo je velika jer tu vanjska valjanost često nije osigurana načinom uzorkovanja ispitanika, budući da su slučajni uzorci rijetki (usp. potpoglavlje 2.1.3), a dijelom i zato što su u području uobičajeni mali uzorci (Lindstromberg i Eyckmans, 2017).

U društvenim znanostima, pa i u usvajanju drugoga jezika, iznimno je teško stvoriti identične eksperimentalne uvjete u više navrata i u različitim kontekstima, pa replikacija u usporedbi s izvornim istraživanjem gotovo nužno podrazumijeva izmjene u uzorcima i kontekstu. S obzirom na to što se prilikom replikacije događa s izvornim nacrtom istraživanja, Morgan-Short i sur. (2018, str. 394) kao podvrste razlikuju **izravnu replikaciju** (engl. *direct replication*), u kojoj se ne mijenja ništa u nacrtu izvornoga istraživanja, **djelomičnu replikaciju** (engl. *partial replication*), u kojoj se uvodi jedna jasno definirana izmjena u jednoj od glavnih varijabli u izvornome istraživanju, i **konceptualnu replikaciju** (engl. *conceptual replication*), kod koje izmjena može biti više jer je uz provjeru valjanosti cilj i proširenje izvornoga istraživanja. Kao posebna se (pod)vrsta dodaje **višelokacijska replikacija** (engl. *multisite replication*), odnosno replikacija provedena s uzorcima na više različitih lokacija, čime se ukupno dobiva veći uzorak i lakše se može razlikovati ono što je stvarni utjecaj neke varijable od šuma nastaloga zbog slučajnih pogrešaka u formiranju pojedinačnih uzoraka (usp. potpoglavlja 2.1.3 i 4.1.2). Ukratko, replikaciju je moguće provesti na više različitih načina, koji će nužno dati djelomično različite

informacije o rezultatima izvornoga istraživanja, ali sve su te informacije vrlo dragocjene i pomažu nam da procijenimo koliko široku primjenu imaju zaključci izvornih istraživanja.

Usprkos svojoj važnosti replikacijska su istraživanja u području usvajanja drugoga jezika dugo bila relativno rijetka. Međutim to nije posve krivica istraživača već i časopisa koji su (bili) skloniji objavljivati izvorna nego replicirana istraživanja te sustava koji replikacije manje vrednuje prilikom evaluacije istraživačkoga rada pojedinaca i institucija. Tek su se u novije vrijeme stekli potrebni uvjeti kako bi se sa svih strana počelo više inzistirati na replikacijskim istraživanjima i više isticati njihova važnost (v. npr. Mackey, 2013; Porte i McManus, 2019). Pojačano zanimanje za replikaciju u znanosti općenito velikim dijelom potječe iz „replikacijske krize“ zabilježene ranih 2000-ih godina u psihologiji (i drugim kvantitativno orijentiranim znanostima), gdje se kod većega broja pojava koje su se smatrale temeljito istraženima dogodilo da replikacija pokazuje drugačije rezultate od izvornih istraživanja, čime je poljuljano povjerenje u uobičajene statističke prakse.¹⁵⁰ Sønning i Werner (2021) u tematskome broju časopisa *Linguistics* navode da je očekivano da takva kriza bude još dublja u istraživanjima jezika, u kojima je tradicija uporabe statističke analize novijega datuma, nego u psihologiji.

O važnosti replikacije i rastućem zanimanju za nju u području usvajanja drugoga jezika govori to što u pojedinim časopisima, poput *Studies in Second Language Acquisition*, danas postoji posebna rubrika posvećena replikacijskim istraživanjima, a pojedini drugi časopisi također eksplicitno govore o prihvaćanju radova o replikacijama (naprimjer *Language Learning*).¹⁵¹ Primjer je jednoga rada iz te rubrike istraživanje koje je provela Feng (2019), u kojemu autorica replicira ranije istraživanje, čija je autorica Cho (2017), posvećeno usvajanju članova u engleskome kao drugome jeziku kod izvornih govornika korejskoga jezika, koristeći isti zadatak prosudbe prihvatljivosti rečenica. Za kontrolnu skupinu upotrebljava podatke iz izvornoga istraživanja, dok skupine učenika prilagođava tako što prikuplja podatke od izvornih govornika kineskoga jezika, koji se od korejskoga razlikuje u značajkama važnima za aspekt usvajanja članova koji istražuje. Rad pritom sadrži uobičajene elemente, od pregleda literature do opisa i interpretacije rezultata. Još jedan koristan primjer replikacije, točnije njezina temeljitog planiranja u kontekstu

¹⁵⁰ Pojedini znanstvenici radije govore o novim izazovima i novim prilikama, nego o krizi i poljuljanome povjerenju (v. npr. Fanelli, 2018, a u području lingvistike Roettger, 2021).

¹⁵¹ Postojanje takve rubrike ima dodatno veliku važnost u svjetlu već spomenute višedesetljezne prakse časopisa u gotovo svim područjima da prednost daju izvornim istraživanjima.

više ciljnih jezika, jest istraživanje Watorek i sur. (2020), koje se bavi samim početkom procesa usvajanja drugoga jezika.

Uz replikaciju je kao sve prisutniji oblik kontrole kvalitete istraživanja važno spomenuti i **sinteze istraživanja** (engl. *research syntheses*; v. npr. Macaro, 2020) te **metaanalize** (engl. *metaanalyses*; v. npr. In'nami i sur., 2020). One predstavljaju posebnu vrstu istraživanja i posebnu vrstu radova, u kojima se analiziraju već objavljena istraživanja na određenu temu radi boljšega sagledavanja dobivenih rezultata, utvrđivanja eventualnih problematičnih aspekata i davanja mogućih smjernica za buduća istraživanja. Naglasak je najčešće na metodološkim aspektima, ali i na načinu definiranja pojmova koji se mjere, kao i na kvaliteti i transparentnosti analize podataka. U daljnjem ćemo tekstu radi jednostavnosti upotrebljavati naziv „metaanaliza“ kao zajednički, budući da se taj naziv često koristi i u širem smislu sinteze istraživanja (iako primarno označava izračunavanje statističkih mjera poput aritmetičke sredine veličine efekta u većemu broju istraživanja na određenu temu; v. Plonsky i Oswald, 2011, kao i potpoglavlje 4.1.3). Treba međutim napomenuti da ni u širem smislu nije u pitanju „narrativni“ prikaz, već sustavna analiza utemeljena na objektivnim metodama i kvantitativnim pokazateljima (Plonsky i Oswald, 2011, str. 275–276).

Kao primjer metaanalize možemo navesti onu koju su proveli Yang i sur. (2017) na temu međujezičnoga transfera između kineskoga i engleskoga jezika. Autori kao motiv za provedbu metaanalize navode varijabilnost i nekonzistentnost među istraživanjima u pogledu pojave koju istražuju te u pogledu rezultata. Njihova se analiza usredotočuje na fonološku svijest, vještine dekodiranja, leksik i morfološku svijest te uzima u obzir 33 ranije objavljena članka. Zaključci do kojih dolaze jesu da na rezultate utječu dob ispitanika i mjesto provedbe istraživanja (u geografskome smislu – primjerice Hong Kong, Kina ili Sjedinjene Američke Države) te da je, sveukupno gledano, razina transfera niska do srednja, ali da je on ipak moguć, tj. dva ga jezika usprkos velikim tipološkim razlikama dopuštaju.

U pitanju je dakle temeljit pregled prethodnih istraživanja putem kojega se doznaje koliko su konzistentno istraživači određene teme dobivali određeni rezultat, kakve su analize upotrebljavali, na koji su način navodili rezultate statističkih postupaka i što se iz svega navedenog u konačnici može zaključiti. Kako navode Boers i sur. (2021), metaanalize predstavljaju prečac do spoznaja iz prethodnih istraživanja.¹⁵² Njihova se važnost ogleda između ostaloga u

¹⁵² Ti autori navode da istraživači prilikom prvoga bavljenja nekom temom često započnu pregled literature metaanalizama, a ne pojedinačnim primarnim istraživanjima.

tome što pridonose akumuliranju znanja tijekom vremena te time i izgradnji „kumulativne kvantitativne discipline“ (Cumming, 2014). Drugim riječima, metaanalize nam mogu ukazati na postojanje utjecaja određenoga čimbenika pouzdanije od ikojega individualnog istraživanja (za koje se uvijek možemo zapitati je li se neki čimbenik pokazao važnim zbog utjecaja konteksta ili neke druge značajke toga konkretnog istraživanja). Takva vrsta akumuliranja znanja posebice je važna kad se istraživanja provode na malim uzorcima; Lindstromberg i Eyckmans (2017) kažu da je, ako se u takvim slučajevima želi doći do pouzdanih rezultata, zapravo nužno kombinirati replikativne studije i metaanalize. Ukupno gledano, premda nam ni metaanalize ne mogu dati idealnu sliku trenutačne situacije u istraživanjima (v. Plonsky i Brown, 2015, kao i Boers i sur., 2021, koji ističu da su i metaanalize podložne metodološkim nekonzistentnostima, da treba biti oprezan prilikom oslanjanja na njihove rezultate i da ih kao i izvorna istraživanja treba replicirati), one predstavljaju vrlo važan kontrolni mehanizam na razini istraživačke zajednice.

U području usvajanja drugoga jezika broj je metaanaliza posebno velik kod pedagoški orijentiranih istraživanja, gdje se ispituje učinkovitost različitih pristupa poučavanju, različitih vrsta korektivne povratne informacije i slično (v. npr. Nguyen i sur., 2017; Plonsky, 2011; Plonsky i Kim, 2016). Međutim posljednjih se godina one sve češće susreću i u lingvističkim pristupima, te postoje metaanalize različitih često korištenih eksperimentalnih zadataka, poput zadataka prosudbe prihvatljivosti rečenica (Plonsky i sur., 2020) i zadataka čitanja vlastitim tempom (Marsden, Thompson i Plonsky, 2018), kao i temeljite sinteze načina mjerenja razine općega znanja drugoga jezika (Park i sur., 2022; Tremblay, 2011). Velik je i broj metaanaliza posvećenih različitim aspektima analize podataka – naprimjer Plonsky i Derrick (2016) te Norouzian (2021) ispituju korištene mjere pouzdanosti instrumenata (v. potpoglavlje 2.3), dok Hu i Plonsky (2021) pobliže proučavaju testiranje ispunjenosti uvjeta za provedbu odabranih analiza (v. potpoglavlje 4.3.2). Metaanalize također ispituju kvalitetu analize podataka i izvještavanja o njoj u cijelosti (Larson-Hall i Herrington, 2010; Larson-Hall i Plonsky, 2015). Vrlo detaljna bibliografija metaanaliza vezanih uz istraživanja usvajanja drugoga jezika može se pronaći na mrežnoj stranici koju održava Luke Plonsky, (su)autor vjerojatno najvećega broja metaanaliza u širem području primijenjene lingvistike: <https://lukeplonsky.wordpress.com/bibliographies/meta-analysis/> (posljednji pristup 26. 4. 2024.). Članak koji su napisali Plonsky i Oswald (2011) može se preporučiti kao posebno praktičan vodič za razumijevanje, ali i provedbu metaanaliza.

Konačno, dijelom kontrole kvalitete istraživanja može se smatrati i **pretprijava istraživanja** (engl. *preregistration*), koja predstavlja praksu da se časopisu najprije preda preliminarna verzija rada – **prijavljeni izvještaj** (engl. *registered report*) – koja sadrži uvod, teorijsku pozadinu, pregled literature i opis planiranoga empirijskog istraživanja (uključujući i prikaz metoda, cjeloviti instrument koji će se koristiti za prikupljanje podataka, kao i opis planirane analize podataka i eventualne podatke iz faze pilotiranja). Takav se rad predaje časopisu prije prikupljanja podataka te u postupku recenzije može biti prihvaćen uvjetno (engl. *in principle* 'načelno'), čime se ostavlja prostora za dodatnu recenziju konačne verzije, koja mora uključiti i analizu i interpretaciju podataka. Uz opravdane razloge moguće je djelomično izmijeniti izvorni plan, kao i prijaviti istraživanje kad su podaci već prikupljeni (što je rjeđi slučaj u eksperimentima, a češći u korpusnim istraživanjima). Podvrsta je te vrste izvještaja i **prijavljeni izvještaj o replikaciji** (engl. *registered replication report*, v. Morgan-Short i sur., 2018).

Pretprijava se smatra korisnom kako bi se izbjegle subjektivnost i različite upitne prakse istraživača vezane primjerice uz prilagođavanje analize kako bi se dobili statistički značajni rezultati (Isbell i sur., 2022; Plonsky i sur., 2024). S druge strane, ona donosi bitnu prednost istraživačima kao pojedincima i cijeloj zajednici time što može u ranome stadiju upozoriti na moguće probleme u metodološkoj postavci istraživanja, a uz to uvjetnim prihvaćanjem jamči objavljivanje rezultata bez obzira na to hoće li oni biti statistički značajni ili ne, što ranije često nije bilo slučaj – opće je poznata sklonost časopisa da objavljuju ponajprije statistički značajne rezultate (o čemu je bilo riječi u četvrtome poglavlju). Konačno, smatra se da pretprijava pomaže i u razgraničenju eksploracijskih i eksplanatornih istraživanja (usp. potpoglavlje 1.4), tj. onih u kojima se generiraju i onih u kojima se testiraju hipoteze (v. npr. Roettger, 2021).

Među časopisima koji su uveli tu praksu od 2018. nalazimo *Language Learning* (v. Marsden, Morgan-Short, Trofimovich i Ellis, 2018).¹⁵³ Pored korisnih informacija koje se mogu naći na mrežnim stranicama toga časopisa, pozivamo čitatelje da pogledaju i resurse dostupne na stranicama Centra za otvorenu znanost (*Center for Open Science*, <https://www.cos.io/initiatives/prereg>, posljednji pristup 26. 4. 2024.), kao i popis prijavljenih istraživanja na mrežnoj stranici <https://www.zotero.org/groups/479248/osf/collections/KEJP68G9> (posljednji pristup 26. 4. 2024.), gdje se mogu pronaći i primjeri

¹⁵³ https://onlinelibrary.wiley.com/page/journal/14679922/homepage/registered_reports.htm (posljednji pristup 26. 4. 2024.).

prijavljenih replikacijskih istraživanja.¹⁵⁴ Koliko nam je poznato, pretprijava istraživanja u ovome trenutku nije dostupna u časopisima koji se objavljuju u Hrvatskoj i Srbiji, barem ne u području usvajanja drugoga jezika. Nadamo se da će zajedno s drugim praksama usmjerenima k poboljšanju kvalitete znanstvenih istraživanja u ovome području i ova uskoro biti uvedena.

7.2 Metodološki zaokret ili teorijsko-metodološka revolucija?

Kako je u ovoj knjizi već više puta istaknuto, u različitim područjima koja temelje svoja istraživanja na empirijskim podacima, danas su u statističkoj analizi, kao i u metodologiji istraživanja općenito, u tijeku intenzivne promjene. Te su promjene u velikoj mjeri izazvane već spomenutom replikacijskom krizom (u psihologiji, medicini i šire). Kako je već kratko spomenuto u prethodnome potpoglavlju, kriza je vezana uz to da se sve češće počelo uočavati kako se čimbenici prethodno jasno identificirani kao (statistički) značajni ne pokazuju nužno takvima u ponovljenim istraživanjima na istu temu, zbog čega se javio val sumnje u pouzdanost tvrdnji koje su izvedene iz empirijskih podataka primjenom statističkoga zaključivanja. Jedan je od najupečatljivijih pokazatelja te sumnje naslov iznimno citiranoga rada koji je napisao liječnik i istraživač John Ioannidis: „Zašto je većina objavljenih rezultata istraživanja netočna?“ (engl. *Why most published research findings are false*; Ioannidis, 2005). Velikoga broja identificiranih uzročnika takvih izostalih potvrda rezultata već smo se dotakle u ovoj knjizi – prenaplašeno oslanjanje na testiranje nulte hipoteze, slabo razumijevanje načela statističkoga zaključivanja, izostanak transparentnosti u pogledu korištenih instrumenata i postupaka, tendencija da časopisi objavljuju ponajprije izvorna istraživanja koja se temelje na statistički značajnim rezultatima, samo su neki među njima. Vidjelo se već u prethodnim poglavljima i da pokušaji popravljivanja situacije uključuju kako izmjene u pristupu statističkoj analizi tako i šire intervencije na razini znanstvene zajednice (naprimjer izmjene u postupcima objavljivanja radova).

Iz uže perspektive, unutar statističke analize usmjerene na uopćavanje rezultata, tekuće promjene slikovito opisuje Field (2018). Taj istraživač, psiholog i autor niza knjiga i radova o primijenjenoj statistici, govori o promjeni statističke paradigme koja se sastoji u prelasku s pristupa u kojemu su *SPINE*, tj. „kralježnicu“, statistike činili *S* – standardna pogreška, *P* – parametar, *I* – procjena intervala (tj., granica pouzdanosti), *N* – testiranje nulte hipoteze, i *E* – procjena (engl. *estimation*), na pristup utemeljen na *EMBERS*, tj. „žaru“,

¹⁵⁴ Na toj se mrežnoj stranici nalaze prijavljena istraživanja iz različitih znanstvenih disciplina.

gdje je E – veličina efekta, M – metaanaliza, BE – Bayesova procjena (engl. *Bayesian estimation*), R – pretprijava (engl. *registration*), a S – zdrav razum (engl. *sense*), koji se treba primijeniti prilikom evaluiranja rezultata testova utemeljenih na nultoj hipotezi. Pojedini drugi autori, poput Cumminga (2014), također psihologa, daleko su drastičniji i dramatičniji u tvrdnjama, navodeći da je testiranje statističke značajnosti kao paradigma „fatalno pogrešno“ te da je potrebno okrenuti se „novoj statistici“ i potpuno ukinuti testiranje nulte hipoteze.¹⁵⁵ U široj perspektivi, u čitavoj su istraživačkoj zajednici, od pojedinaca do časopisa i sveučilišta, sve zastupljenije inicijative okrenute institucionalnome rješavanju problema. Naglašava se potreba za većom zastupljenošću statistike u sveučilišnome obrazovanju, potiče se provedba i objavljivanje replikacijskih istraživanja i metaanaliza, postupno se uvodi praksa pretprijave istraživanja, a posebno treba istaknuti i razvoj pokreta za otvorenu znanost, koji doprinosi široj dostupnosti svih relevantnih informacija (v. npr. Marsden i Plonsky, 2018).

Kad pomislimo na područje usvajanja drugoga jezika u ovome kontekstu, jasno možemo uočiti prisutnost svih navedenih „popravnih“ strategija. Posljednje je desetljeće posebno obilježeno inicijativama za izmjene toga tipa. Već je Byrnes (2013) govorila o „metodološkome zaokretu“ (engl. *methodological turn*) prisutnome u ovome području, a otada se u njemu još mnogo toga dogodilo pa se danas metodološki orijentirane knjige i radovi toliko često objavljuju da ih je postalo doista teško pratiti. Među novijim radovima, Gass i sur. (2021) u svojem povijesnom pregledu razvoja kvantitativnih istraživanja usvajanja drugoga jezika ističu napredak u postizanju znanstvene rigoroznosti (engl. *scientific rigor*) putem primjene sve naprednijih statističkih testova i putem sve veće metodološke transparentnosti. I doista, usprkos čestoj samokritičnosti istraživača usvajanja drugoga jezika (koja se odražava u vrlo kritičnim i barem djelomično negativno intoniranim metodološkim pregledima unutar područja), u široj slici lingvističkih istraživanja usvajanje drugoga jezika može se izdvojiti kao jedno od naprednijih područja i u pogledu analiza koje se provode i u pogledu osviještenosti o problemima s replikacijskom krizom i paradigmom utemeljenom na statističkoj značajnosti (v. posebice Sönning i Werner, 2021, str. 1183).

¹⁵⁵ Cumming (2014) smatra da testiranje statističke značajnosti opstaje kao dominantan pristup zahvaljujući duboko ukorijenjenoj ljudskoj potrebi za jasnim klasifikacijama i zahvaljujući sklonosti ljudi da čak i vrlo složene pojave promatraju crno-bijelo, što testiranje nulte hipoteze i oslanjanje na vrijednost p potiču. S druge strane, Sönning i Werner (2021, str. 1195) ističu da je vjerojatnost kakvu izražava vrijednost p strana ljudskoj kogniciji i uobičajenome svakodnevnom poimanju vjerojatnosti, te da nije nimalo neobično da se u kontekstu testiranja utemeljenoga na nultoj hipotezi javlja toliko pogrešnih tumačenja.

Što onda možemo očekivati u sljedećemu razdoblju? S jedne strane, iako je statistička analiza utemeljena na testiranju nulte hipoteze još uvijek vrlo prisutna, jasno je da je povjerenje u nju ozbiljno poljuljano. Kroz replikacijsku krizu pokazalo se naime da ona ne uspijeva u potpunosti osigurati upravo ono zbog čega je toliko dugo bila tako široko prihvaćena – objektivnost u istraživanjima i znanstvenu rigoroznost. Zbog toga je vrlo izvjesno da će se ili morati zamijeniti drukčijim pristupom ili dopuniti dodatnim metodološkim sredstvima. Trenutačno se nalazimo u prijelaznome razdoblju u kojemu druga od navedenih opcija djeluje vjerojatnije. Sønning i Werner (2021) primjerice ističu da će vrijednost p uskoro proslaviti 100. rođendan (2025, usp. Fisher, 1925), te da ga dočekuje u dobrome društvu, zajedno s mjerama veličine efekta i drugim informacijama koje pridonose boljemu razumijevanju inferencijalnoga zaključivanja. Također, čini se da u znanstvenoj zajednici postoji relativno visoka razina slaganja u pogledu toga da vrijednost p ne može sama po sebi biti „kriva“ za toliko problema. Tako se i za područje usvajanja drugoga jezika ističe (v. posebice Gass i sur., 2021) kako jest važno poboljšati kvalitetu statističkih analiza, naprimjer fokusirajući se u većoj mjeri na statističku snagu testova i veličinu efekta, ali je istodobno potrebno raditi i na institucionalnome okviru koji omogućava otvorenu znanost i potiče transparentnost istraživanja i replikabilnost, kako bi se zajedničkim djelovanjem doista stiglo do kumulativne znanosti, u kojoj naglasak neće biti toliko na pojedinačnim istraživanjima koliko na njihovu doprinosu izgradnji širega skupa spoznaja.

S druge strane, nakon svih rasprava o metodološkim problemima treba se prisjetiti i toga da koliko god one bile važne za razvoj empirijskih znanstvenih disciplina, naprednije statističke metode nisu jedini pokazatelj znanstvene rigoroznosti, već se moraju sagledati u kontekstu tema koje se istražuju i njihovih teorijskih okvira. U istraživanjima jezika u tome je smislu došlo do dvostruko složene situacije, a može se reći i do istodobne pojave pritiska s dviju strana. S jedne, primjena kvantitativnih metoda u lingvistici relativno je novijega datuma, što već samo po sebi donosi mnogo nesigurnosti i posuđivanja analiza iz drugih područja (što se zbog nedostatka neposrednoga iskustva zna provoditi i nedovoljno kritično, v. npr. Sønning i Werner, 2021). S druge, prodor kvantitativnih metoda u lingvistiku intenzivira se upravo u trenutku kad dominantna statistička paradigma prolazi kroz najveću krizu u posljednjemu stoljeću, donoseći sa sobom novi sloj nesigurnosti.¹⁵⁶ Nije

¹⁵⁶ U istraživanjima usvajanja drugoga jezika može se svemu tome pridodati još i relativna mladost čitavoga područja, čijim se počecima smatraju 60-e i 70-e godine 20. stoljeća (Gass, 2009; Gass i sur., 2021).

neobično ni neočekivano da se pod tolikim pritiskom metodološkim pitanjima počne pridavati iznimno velika pozornost te da ona počnu u nekoj mjeri i određivati teme koje će se istraživati, usprkos tome što bi – uz svu svoju važnost – metodologija trebala biti pomoćno sredstvo analize koja je primarno lingvistička (Sønning i Werner, 2021, str. 1190–1191). I autorice ove knjige mogu posvjedočiti kako su se više puta tijekom njezina pisanja osjetile preplavljene količinom metodoloških informacija te na trenutke gotovo zaboravljale da se ispod svih mjera i statističkih testova krije učenički jezik sa svim svojim zanimljivim značajkama.

Jedna važna komponenta metodološkoga zaokreta u istraživanjima jezika trebala bi stoga biti teorijske prirode. Neadekvatnost teorija zapravo se i navodi kao jedan od uzročnika replikacijske krize (v. npr. Smaldino, 2019) pa bi temeljenje analize podataka na teorijskim spoznajama (umjesto obratnoga) zapravo moglo pridonijeti i poboljšanju metodologije (usp. Sønning i Werner, 2021). Takav pristup nije nimalo lako postići u uvodnim prikazima ili udžbenicima i priručnicima za statističku analizu koji se obraćaju širim skupinama istraživača u nekome području, a ne samo istraživačima koji poznaju određeni teorijski okvir (kakvih posebice u lingvističkome pristupu usvajanju drugoga jezika ima više). Ali i to može biti jedna od tema za razmišljanje u okviru metodoloških promjena kojima svjedočimo.

7.3 Misli globalno, djeluj lokalno

Čitatelji su svakako primijetili da je u ovoj knjizi fokus vrlo naglašeno stavljen na istraživanja usvajanja drugoga jezika u međunarodnim okvirima. S iznimkom nekoliko radova i statističkih priručnika i udžbenika iz područja psihologije, najveći dio literature na koju smo se oslonile na engleskome je jeziku, a prikaz metodološke situacije koji smo dale odnosi se na vrlo široko koncipiranu međunarodnu istraživačku zajednicu. Kako bi se takav prikaz dopunio i lokalnom komponentom, u ovome ćemo se odjeljku osvrnuti na situaciju na području Hrvatske, Srbije i drugih zemalja koje se mogu smatrati dijelom iste regije. Također ćemo dati nekoliko dodatnih smjernica za daljnje individualno usavršavanje.

Kao smo već istaknule u Kraš i Miličević (2015), dojam koji se može steći iz objavljenih radova i izlaganja na znanstvenim skupovima jest da eksperimentalna istraživanja usvajanja drugoga jezika u regiji u pogledu analize podataka za nekoliko koraka zaostaju za onima na međunarodnoj razini. Ne postoji mnogo sistematiziranih podataka o tome, ali možemo za uži kontekst Hrvatske kao primjer navesti istraživanje koje su provele Bagarić

Medve i Pavičić Takač (2014). Te su autorice analizirale 55 znanstvenih članaka objavljenih u časopisima u Hrvatskoj (ponajprije u časopisu *Strani jezici* i zborniku Hrvatskoga društva za primijenjenu lingvistiku) i 337 znanstvenih članaka objavljenih u relevantnim međunarodnim časopisima u razdoblju od 2007. do 2012. te zaključile da se neke značajke u velikoj mjeri poklapaju (primjerice prosječan broj ispitanika), ali da se u radovima u hrvatskim časopisima ukupno gledano manje vodi računa o odabiru statističkih testova i izvještavanju o njima nego što je to slučaj s međunarodnim časopisima. Autorice navode (Bagarić Medve i Pavičić Takač, 2014, str. 99): „Kada je riječ o metodama analize i prikazima rezultata, hrvatska istraživanja povremeno ostavljaju dojam prigodnih istraživanja, provedenih *ad hoc* i istraživanja u kojima se rabe instrumenti koji se izrađuju za potrebe danoga istraživanja, ali psihometrijski nisu provjerena ili se o tome ne izvješćuje. Iznimka su istraživanja koja su dijelom većih projekata.“

Ta analiza potvrđuje postojanje potrebe za temeljitijom lokalnom metodološkom obukom istraživača usvajanja drugoga jezika, moguće i uz dodatno poticanje na provedbu eksperimentalnih istraživanja – citirana analiza pokazuje i to da su u hrvatskim časopisima manje zastupljena istraživanja usvajanja jezičnih pojava nego u međunarodnim časopisima (više se pozornosti posvećuje proučavanju individualnih razlika, gdje su istraživanja rjeđe eksperimentalnoga tipa). Kao što je već naglašeno u Predgovoru, jedan je od glavnih motiva za nastanak ove knjige bilo upravo to da se lokalnoj istraživačkoj zajednici ponudi djelo koje, iako temi pristupa monografski (uz mnogo analize, kritičkih razmatranja i citiranja relevantne literature), može poslužiti kao uvodni konceptualni priručnik koji će barem djelomično ispuniti prazninu u literaturi o usvajanju drugoga jezika napisanoj na jezicima koji nisu engleski.

Naravno, jedna knjiga ne može riješiti vrlo širok problem kakav je metodološko obrazovanje istraživača. Nasreću, istraživačima je dostupan velik broj drugih izvora putem kojih se spoznaje o statističkoj analizi podataka mogu proširiti i produbiti. Uz već navedene izvore u ovoj knjizi, posebice u prvome i drugome poglavlju, upućujemo na članke Sönning i Werner (2021), Winter (2022) te Gass i sur. (2021) kao na vrlo prikladna polazišta za daljnje upoznavanje s literaturom. Sva su tri članka novijega datuma i referiraju se na velik broj metodološki orijentiranih knjiga i drugih vrsta izvora znanja (poput znanstvenih skupova, radionica i ljetnih škola), u lingvistici općenito (prvi i drugi rad) i u usvajanju drugoga jezika (treći rad). U lokalnoj zajednici svakako treba pratiti aktivnosti velikih udruženja, poput Hrvatskoga društva za primijenjenu lingvistiku (<http://www.hdpl.hdpl.hr/home>, posljednji pristup 26. 4.

2024.) i Društva za primijenjenu lingvistiku Srbije (<http://dpls-alas.rs>, posljednji pristup 26. 4. 2024.) te njihove publikacije i skupove.

U nadi da će ova knjiga, uz sve mane koje zasigurno ima, doista biti od koristi istraživačima usvajanja drugoga jezika pozivamo ih i da izađu iz njezinih okvira i informiraju se o statističkoj analizi iz drugih raspoloživih izvora te da naučeno primijene u svojim istraživanjima i time daju dodatni poticaj i lokalnoj i široj istraživačkoj zajednici. Za sam kraj posuđujemo ideju za završni pozdrav iz rada Jamesa Deana Browna (Brown, 2015), u kojemu autor nizom primjera odgovara na pitanje zašto učiti napredne kvantitativne metode i time također potiče istraživače na daljnje učenje. Brown je poželio svojim čitateljima sretno istraživanje, a mi tome dodajemo i – sretna analiza podataka!

Sažetak

Ova knjiga govori o primjeni statističke analize u eksperimentalnim istraživanjima usvajanja drugoga jezika. U pogledu usvajanja drugoga jezika usredotočuje se na lingvistički orijentirana istraživanja, u kojima se naglasak stavlja na proučavanje sustavnoga znanja drugoga jezika i njegovih značajki na različitim razinama jezične strukture. U pogledu statistike najviše se pozornosti posvećuje tumačenju onih pojmova i analitičkih koraka koji su važni za sve ili velik broj statističkih testova koji se provode u istraživanjima usvajanja drugoga jezika. Naglasak se također stavlja na odnos deskriptivne i inferencijalne statistike te na važnost nacrtu istraživanja za sve aspekte analize podataka. Knjiga je napisana kao znanstvena monografija koja se može upotrebljavati i kao priručnik za početnike koji se žele upoznati s temeljnim statističkim postupcima i pretpostavkama, primijenjenima na proučavanje lingvističkih značajki učeničkoga jezika. Obuhvaćene teme ne zahtijevaju prethodno poznavanje statistike ili temeljito poznavanje usvajanja drugoga jezika. Pristup je konceptualan, a ne praktičan: ne daju se upute kako provesti analize u kojemu statističkom softveru niti su uključeni zadaci.

Knjiga je strukturirana na sljedeći način. Prvo poglavlje objašnjava motivaciju za nastanak knjige, usredotočujući se na ulogu statističke analize u istraživanjima usvajanja drugoga jezika. Drugo se poglavlje bavi koracima u istraživačkome procesu koji prethode statističkoj analizi podataka. Točnije, sažima temeljna metodološka pitanja o kojima treba voditi računa prilikom planiranja eksperimentalnoga istraživanja i analize podataka, a koja su vezana uz istraživačke hipoteze, mjerenje varijabli, odabir ispitanika i podražaja te nacrt istraživanja. Uz to u tome se poglavlju opisuju koraci u pripremi podataka za statističku analizu te daju smjernice za provjeru pouzdanosti mjernoga instrumenta. Treće je poglavlje posvećeno deskriptivnoj statistici, odnosno početnome opisu podataka pomoću različitih deskriptivnih statističkih mjera i vizualizacije podataka različitim vrstama grafikona. Četvrto se poglavlje nadovezuje na deskriptivni temelj trećega baveći se statističkim zaključivanjem. Ono uvodi pojmove statističke značajnosti i testiranja nulte hipoteze, objašnjava logiku statističkoga testiranja i navodi kriterije izbora statističkoga testa. Peto poglavlje objašnjava postupke izračunavanja nekoliko često korištenih „tradicionalnih“ testova (poput korelacije, t -testa, analize varijance te njihovih neparametrijskih inačica) i daje kraći prikaz nekih naprednijih testova, koji se sve češće upotrebljavaju i u istraživanjima jezika. U tome se poglavlju također raspravlja o aktualnim pitanjima vezanima uz

pouzdanost testova koji se temelje na statističkoj značajnosti te se upućuje na moguća rješenja, od pozornije interpretacije rezultata testa i izračunavanja veličine efekta do oslanjanja na robusnu statistiku. U šestome se poglavlju govori o izvještavanju o eksperimentalnim istraživanjima usvajanja drugoga jezika – daju se opće smjernice za izvještavanje te opisuju dijelovi izvještaja, ponajprije znanstvenoga članka za objavu u znanstvenome časopisu i uredničkoj knjizi. Konačno, sedmo poglavlje sadrži razmatranja o mogućim budućim trendovima u uporabi statistike u empirijskim disciplinama poput usvajanja drugoga jezika kao i poziv na zajedničko promišljanje statističke analize i teorijskih spoznaja u tome području te na temeljitiju metodološku obuku istraživača usvajanja drugoga jezika u našem lokalnom kontekstu.

Summary

This book is dedicated to the application of statistical analysis in experimental studies of second language acquisition. In terms of second language acquisition, the focus is on linguistically oriented research, where emphasis is placed on the study of systematic knowledge of the second language and its properties at different levels of the language structure. When it comes to statistics, most attention is dedicated to those concepts and analytical steps that are important for all or most of statistical tests used in research on second language acquisition. Emphasis is also placed on the relationship between descriptive and inferential statistics and the importance of research design for all aspects of data analysis. The book is written as a scientific monograph that can also be used as a manual for beginners who wish to familiarise themselves with basic statistical procedures and the assumptions behind them, applied to the study of linguistic properties of learner language. The topics covered do not require prior knowledge of statistics or in-depth familiarity with issues in second language acquisition. The approach is conceptual rather than practical: no instructions are given on how to perform the analyses in any statistical software and no exercises are included.

The book is structured as follows. Chapter 1 explains the motivation behind the book, focusing on the role of statistical analysis in second language acquisition research. Chapter 2 deals with those steps in the research process that precede statistical data analysis. More precisely, it summarises the fundamental methodological matters that should be taken into account when planning an experimental study and associated data analysis, which are related to research hypotheses, variable measurement, the selection of participants and stimuli, and research design. This chapter also describes the steps involved in the preparation of data for statistical analysis and provides guidelines for checking the reliability of the data collection instrument. Chapter 3 is dedicated to descriptive statistics, that is, to the initial description of data via different kinds of descriptive statistical measures, and data visualization using different types of graphs. Chapter 4 builds on the descriptive background of Chapter 3, dealing with inferential statistics. It introduces the notions of statistical significance and null hypothesis testing, explains the logic behind statistical testing and outlines the criteria for choosing a statistical test. Chapter 5 explains the calculation procedures of several frequently used “traditional” tests (such as correlation, t -tests and analysis of variance, as well as their non-parametric equivalents) and it gives a brief overview of some of the

more advanced tests, which are increasingly used in language research. This chapter also discusses current issues related to the reliability of significance-based statistical tests and suggests possible solutions, from more careful interpretation of test results and effect size calculations to reliance on robust statistics. Chapter 6 discusses the reporting of experimental studies of second language acquisition. It provides general guidelines for reporting experimental research and describes the structure of an experimental research report on second language acquisition, primarily focusing on scientific articles for publication in a scientific journal or an edited volume. Finally, Chapter 7 outlines possible future trends in the use of statistics in empirical disciplines such as second language acquisition, and puts forward a call for an integrated consideration of statistical analysis and the field's theoretical knowledge, and for more thorough methodological training of second language acquisition researchers in our local context.

Literatura

- Abbuhl, R. (2011). Why, when, and how to replicate research. U A. Mackey i S. M. Gass (ur.), *Research methods in second language acquisition* (str. 296–312). Wiley. <https://doi.org/10.1002/9781444347340.ch15>
- Abbuhl, R., Gass, S. M. i Mackey, A. (2013). Experimental research design. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 337–372). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.018>
- Ahn, H. (2021). From Interlanguage grammar to target grammar in L2 processing of definiteness as uniqueness. *Second Language Research*, 37(1), 91–119. <https://doi.org/10.1177/0267658319868003>
- Alacaci, C. (2004). Inferential statistics: Understanding expert knowledge and its implications for statistics education. *Journal of Statistics Education*, 12(2), <https://doi.org/10.1080/10691898.2004.11910737>
- Al-Hoorie, A. H. i Vitta, J. P. (2019). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR Impact Factors. *Language Teaching Research*, 23(6), 727–744. <https://doi.org/10.1177/1362168818767191>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association: The official guide to APA style* (7. izdanje). American Psychological Association.
- Amoroso, L. W. (2018). Analyzing group differences. U A. Phakiti, P. De Costa, L. Plonsky i S. Starfield (ur.), *The Palgrave handbook of applied linguistics research methodology* (str. 501–521). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1_22
- Andringa, S. (2014). The use of native speaker norms in critical period hypothesis research. *Studies in Second Language Acquisition*, 36(3), 565–596. <https://doi.org/10.1017/S0272263113000600>
- Andringa, S. i Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. <https://doi.org/10.1017/S0267190520000033>
- Artstein, R. i Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Arunachalam, S. (2013). Experimental methods for linguists. *Language and Linguistics Compass*, 7(4), 221–232. <https://doi.org/10.1111/lnc3.12021>
- Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers*, 1, 1–45.

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Baayen, R. H. (2013). Multivariate statistics. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 337–372). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.018>
- Baayen, R. H., Davidson, D. J. i Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, H. R. i Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. <https://doi.org/10.21500/20112084.807>
- Bagarić Medve, V. i Pavičić Takač, V. (2014). Istraživanja poučavanja inih jezika i ovladavanja njima. *Strani jezici*, 43(2), 81–103.
- Bard, E. G., Robertson, D. i Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32–68. <https://doi.org/10.2307/416793>
- Barr, D. J. (2018). Generalizing over encounters: Statistical and theoretical considerations. U S.-A. Rueschemeyer i M. G. Gaskell (ur.), *The Oxford handbook of psycholinguistics* (str. 916–929). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198786825.013.39>
- Barr, D. J., Levy, R., Scheepers, C. i Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Belia, S., Fidler, F., Williams, J. i Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10(4), 389–396. <https://doi.org/10.1037/1082-989X.10.4.389>
- Bennett, C., Miller, M. i Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *NeuroImage*, 47, S125. [https://doi.org/10.1016/S1053-8119\(09\)71202-9s](https://doi.org/10.1016/S1053-8119(09)71202-9s)
- Berger, C., Crossley, S. i Skalicky, S. (2019). Using lexical features to investigate second language lexical decision performance. *Studies in Second Language Acquisition*, 41(5), 911–935. <https://doi.org/10.1017/S0272263119000019>
- Biedroń, A. (2011). Intelligence in gifted L2 learners. U M. Pawlak (ur.), *Extending the boundaries of research on second language learning and teaching* (str. 129–142). Springer.
- Blom, E. i Unsworth, S. (2010). Introduction. U E. Blom i S. Unsworth (ur.), *Experimental methods in language acquisition research* (str. 1–10). John Benjamins.

- Boers, F., Bryfonski, L., Faez, F. i McKay, T. (2021). A call for cautious interpretation of meta-analytic reviews. *Studies in Second Language Acquisition*, 43(1), 2–24. <https://doi.org/10.1017/S0272263120000327>
- Brown, A. (2008). The strange origins of the Student's *t*-test. *Physiology News*, 71, 13–16. <https://doi.org/10.36866/pn.71.13>
- Brown, J. D. (1991). Statistics as a foreign language, part 1: What to look for in reading statistical language studies. *TESOL Quarterly*, 25(4), 569–586. <https://doi.org/10.2307/3587077>
- Brown, J. D. (2014). Classical theory reliability. U A. J. Kunnan (ur.), *The companion to language assessment* (str. 1165–1181). Wiley–Blackwell.
- Brown, J. D. (2015). Why bother learning advanced quantitative methods in L2 research? U L. Plonsky (ur.), *Advancing quantitative methods in second language research* (str. 9–20). Routledge.
- Buchstaller, I. i Khattab, G. (2014). Population samples. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 74–95). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.006>
- Bulté, B. i Roothoof, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, 91, 102246. <https://doi.org/10.1016/j.system.2020.102246>
- Byrnes, H. (2013). Notes from the editor. *The Modern Language Journal*, 97(4), 825–827. <https://doi.org/10.1111/j.1540-4781.2013.12051.x>
- Chaudron, C. (2003). Data collection in SLA research. U C. J. Doughty i M. H. Long (ur.), *The handbook of second language acquisition* (str. 762–830). Blackwell. <https://doi.org/10.1002/9780470756492.ch22>
- Cho, J. (2017). The acquisition of different types of definite noun phrases in L2-English. *International Journal of Bilingualism*, 21(3), 367–382. <https://doi.org/10.1177/1367006916629577>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. izdanje). Lawrence Erlbaum Associates.
- Covey, L. i Gabriele, A. (2023). Psycholinguistic methods in second language research. U A. Mackey i S. M. Gass (ur.), *Current approaches in second language acquisition research* (str. 170–194). Wiley. <https://doi.org/10.1002/9781394259670.ch8>
- Crowther, D., Tigchelaar, M., Maloney, J. i Loewen, S. (2020). Statistical knowledge in second language acquisition research: A researchers' perspective. *Second Language Studies*, 38(1), 19–39.

- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- Cummings, J. J. i Reeves, B. (2022). Stimulus sampling and research integrity. U J. J. Cummings i B. Reeves (ur.), *Research integrity* (str. 203–223). Oxford University Press. <https://doi.org/10.1093/oso/9780190938550.003.0008>
- Cunningham-Andersson, U. (1997). Native speaker reactions to non-native speech. U A. James i J. Leather (ur.), *Second-language speech* (str. 133–144). Mouton De Gruyter. <https://doi.org/10.1515/9783110882933.133>
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382. <https://doi.org/10.1177/0267658312443651>
- Cunnings, I. i Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, 42(3), 579–599. <https://doi.org/10.1017/S0142716420000648>
- Čolović, P. i Milin, P. (2018). *Korelacioni istraživački nacrti*. Filozofski fakultet. <http://digitalna.ff.uns.ac.rs/sites/default/files/db/books/978-86-6065-426-9.pdf#overlay-context=sadrzaj/2018/978-86-6065-426-9>
- Dallas, A., DeDe, G. i Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, 63(4), 766–799. <https://doi.org/10.1111/lang.12026>
- Day, R. A. i Gastel, B. (2014). *How to write and publish a scientific paper* (7. izdanje). Cambridge University Press.
- DeBruine, L. (2023). *faux: Simulation for factorial designs* (1.2.0) [Softver]. <https://doi.org/10.5281/ZENODO.2669586>
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499–533. <https://doi.org/10.1017/S0272263100004022>
- Deniz, N. D. (2022). Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2. *Bilingualism: Language and Cognition*, 25(5), 713–725. <https://doi.org/10.1017/S1366728921001140>
- Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, 50(1), 132–153. <https://doi.org/10.1002/tesq.217>
- Díaz-Negrillo, A. i Valera, S. (2010). A learner corpus-based study on error associations. *Procedia - Social and Behavioral Sciences*, 3, 72–82. <https://doi.org/10.1016/j.sbspro.2010.07.014>

- Diependaele, K., Duñabeitia, J. A., Morris, J. i Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, 64(4), 344–358. <https://doi.org/10.1016/j.jml.2011.01.003>
- Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal: Promoting Communications on Statistics and Stata*, 15(1), 292–300. <https://doi.org/10.1177/1536867X1501500117>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Dragičević, Č. (2009). *Statistika za psihologe* (2. izdanje). Centar za primenjenu psihologiju.
- Duffield, N. i White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research*, 15(2), 133–160. <https://doi.org/10.1191/02676589968237583>
- Egbert, J. i LaFlair, G. T. (2018). Statistics for categorical, nonparametric, and distribution-free data. U A. Phakiti, P. De Costa, L. Plonsky i S. Starfield (ur.), *The Palgrave handbook of applied linguistics research methodology* (str. 523–539). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1_23
- Egbert, J. i Plonsky, L. (2020). Bootstrapping techniques. U M. Paquot i S. Th. Gries (ur.), *A practical handbook of corpus linguistics* (str. 593–610). Springer. https://doi.org/10.1007/978-3-030-46216-1_24
- Fajgelj, S. (2014). *Metode istraživanja ponašanja* (6. dopunjeno izdanje). Centar za primenjenu psihologiju.
- Fajgelj, S. (2020). *Psihometrija* (5. izdanje). Centar za primenjenu psihologiju.
- Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115(11), 2628–2631. <https://doi.org/10.1073/pnas.1708272114>
- Faul, F., Erdfelder, E., Buchner, A. i Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Felser, C. i Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research*, 23(1), 9–36. <https://doi.org/10.1177/0267658307071600>
- Felser, C., Sato, M. i Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12(4), 485–502. <https://doi.org/10.1017/S1366728909990228>

- Feng, S. (2019). The acquisition of English definite noun phrases by Mandarin Chinese speakers. *Studies in Second Language Acquisition*, 41(4), 881–896. <https://doi.org/10.1017/S0272263118000323>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5. izdanje). Sage.
- Field, A. i Hole, G. (2003). *How to design and report experiments*. Sage.
- Field, A., Miles, J. i Field, Z. (2012). *Discovering statistics using R*. Sage.
- Field, A. i Wilcox, R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98, 19–38. <https://doi.org/10.1016/j.brat.2017.05.013>
- Fisher, R. (1925). *Statistical methods for research workers*. Oliver & Boyd.
- Fox, J. (2005). Nonparametric regression. U B. S. Everitt i D. C. Howell (ur.), *Encyclopedia of statistics in behavioral science* (str. 1426–1430). Wiley. <https://doi.org/10.1002/0470013192.bsa446>
- Gabriel, U. i Gygax, P. (2023). Dealing with participant variability in experimental linguistics. U S. Zufferey i P. Gygax (ur.), *The Routledge handbook of experimental linguistics* (str. 345–355). Routledge. <https://doi.org/10.4324/9781003392972-25>
- Gačić, M. (2012). *Pisanje znanstvenih i stručnih radova*. Školska knjiga.
- Gačić, M. (2017). *Pisanje u znanosti i struci*. Narodne novine.
- Garcia, G. D. (2021). *Data visualization and analysis in second language research*. Routledge.
- Gardner, R. C. i Mihaljević Djigunović, J. (2003). *Međunarodni projekt AMTB (hrvatsko-engleska verzija)*. https://anglist.ffzg.unizg.hr/wp-content/uploads/2017/03/AMTB_UPITNIK_MIHALJEVIC.pdf
- Gass, S. M. (2009). A historical survey of SLA research. U W. C. Ritchie i T. K. Bhatia (ur.), *The new handbook of second language acquisition* (str. 3–28). Emerald.
- Gass, S. M., Loewen, S. i Plonsky, L. (2021). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching*, 54(2), 245–258. <https://doi.org/10.1017/S0261444819000430>
- Gass, S. M. i Mackey, A. (2007). *Data elicitation for second and foreign language research*. Routledge.
- Gastel, B. i Day, R. A. (2022). *How to write and publish a scientific paper* (9. izdanje). Greenwood.
- Gonulal, T. (2016). *Statistical literacy among second language acquisition graduate students*. [Doktorska disertacija]. Michigan State University.

- Gonulal, T. (2018). An investigation of the predictors of statistical literacy in second language acquisition. *Eurasian Journal of Applied Linguistics*, 4(1), 49–70.
- Gonulal, T. (2020). Statistical knowledge and training in second language acquisition: The case of doctoral students. *ITL - International Journal of Applied Linguistics*, 171(1), 62–89. <https://doi.org/10.1075/itl.18031.gon>
- Gonulal, T., Loewen, S. i Plonsky, L. (2017). The development of statistical literacy in applied linguistics graduate students. *ITL - International Journal of Applied Linguistics*, 168(1), 4–32. <https://doi.org/10.1075/itl.168.1.01gon>
- Grabowski, K. C. i Oh, S. (2018). Reliability analysis of instruments and data coding. U A. Phakiti, P. De Costa, L. Plonsky i S. Starfield (ur.), *The Palgrave handbook of applied linguistics research methodology* (str. 541–565). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1_24
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N. i Altman, D. G. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337–350. <https://doi.org/10.1007/s10654-016-0149-3>
- Gries, S. T. (2013a). Basic significance testing. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 316–336). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.017>
- Gries, S. T. (2013b). *Statistics for linguistics with R: A practical introduction* (2. izmijenjeno izdanje). De Gruyter Mouton.
- Gudmestad, A. i Edmonds, A. (ur.). (2018). *Critical reflections on data in second language acquisition*. John Benjamins. <https://doi.org/10.1075/llt.51>
- Gyllstad, H., Sundqvist, P., Sandlund, E. i Källkvist, M. (2023). Effects of word definitions on meaning recall: A multisite intervention in language-diverse second language English classrooms. *Language Learning*, 73(2), 403–444. <https://doi.org/10.1111/lang.12527>
- Hatch, E. i Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Heinle & Heinle.
- Heil, J. i López, L. (2020). Acquisition without evidence: English infinitives and poverty of stimulus in adult second language acquisition. *Second Language Research*, 36(4), 415–443. <https://doi.org/10.1177/0267658319850611>
- Hellman, A. B. (2011). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics*, 21(2), 162–182. <https://doi.org/10.1111/j.1473-4192.2010.00265.x>

- Hiver, P. i Al-Hoorie, A. H. (2020). Reexamining the role of vision in second language motivation: A preregistered conceptual replication of You, Dörnyei, and Csizér (2016). *Language Learning*, 70(1), 48–102. <https://doi.org/10.1111/lang.12371>
- Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics*, 50, 1–14. <https://doi.org/10.1016/j.wocn.2015.01.004>
- Howard, A. L. (2024). Graduate students need more quantitative methods support. *Nature Reviews Psychology*, 3, 140–141. <https://doi.org/10.1038/s44159-024-00288-y>
- Hu, Y. i Plonsky, L. (2021). Statistical assumptions in L2 research: A systematic review. *Second Language Research*, 37(1), 171–184. <https://doi.org/10.1177/0267658319877433>
- Hulstijn, J. H., Van Gelderen, A. i Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, 30(4), 555–582. <https://doi.org/10.1017/S0142716409990014>
- In'nami, Y., Koizumi, R. i Tomita, Y. (2020). Meta-analysis in applied linguistics. U J. McKinley i H. Rose (ur.), *The Routledge handbook of research methods in applied linguistics* (str. 240–252). Routledge.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Isbell, D. R., Brown, D., Chen, M., Derrick, D. J., Ghanem, R., Arvizu, M. N. G., Schnur, E., Zhang, M. i Plonsky, L. (2022). Misconduct and questionable research practices: The ethics of quantitative data handling and reporting in applied linguistics. *The Modern Language Journal*, 106(1), 172–195. <https://doi.org/10.1111/modl.12760>
- Jackson, C. N., Morner, E. i Brehm, L. (2018). The production of subject-verb agreement among Swedish and Chinese second language speakers of English. *Studies in Second Language Acquisition*, 40(4), 907–921. <https://doi.org/10.1017/S0272263118000025>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jaensch, C. (2009). Article choice and article omission in the L3 German of native speakers of Japanese with L2 English. U M. D. P. García Mayo i R. Hawkins (ur.), *Second language acquisition of articles: Empirical findings and theoretical implications* (str. 233–263). John Benjamins. <https://doi.org/10.1075/lald.49.14jae>

- Jegerski, J. i VanPatten, B. (ur.). (2014). *Research methods in second language psycholinguistics*. Routledge.
- Jensen, I. N., Slabakova, R., Westergaard, M. i Lundquist, B. (2020). The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*, 36(1), 3–29. <https://doi.org/10.1177/0267658318825067>
- Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.
- Jiang, N., Novokshanova, E., Masuda, K. i Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61(3), 940–967. <https://doi.org/10.1111/j.1467-9922.2010.00627.x>
- Johnson, D. E. (2013). Descriptive statistics. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 288–315). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.016>
- Keating, G. D. i Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 37(1), 1–32. <https://doi.org/10.1017/S0272263114000187>
- Khany, R. i Tazik, K. (2019). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics*, 26(1), 48–65. <https://doi.org/10.1080/09296174.2017.1421498>
- Kim, L. K. i Lakshmanan, U. (2009). The processing role of the Article Choice Parameter: Evidence from L2 learners of English. U M. D. P. García Mayo i R. Hawkins (ur.), *Second language acquisition of articles: Empirical findings and theoretical implications* (str. 87–113). John Benjamins. <https://doi.org/10.1075/lald.49.08kim>
- Kim, S., Ko, H. i Yang, H.-K. (2020). Telicity and mode of merge in L2 acquisition of resultatives. *Language Acquisition*, 27(2), 117–159. <https://doi.org/10.1080/10489223.2019.1659275>
- Kim, Y. i McDonough, K. (2008). Learners' production of passives during syntactic priming activities. *Applied Linguistics*, 29(1), 149–154. <https://doi.org/10.1093/applin/amn004>
- Kraš, T. (2008a). *L2 acquisition of the lexicon-syntax interface and narrow syntax by child and adult Croatian learners of Italian*. [Doktorska disertacija]. University of Cambridge.
- Kraš, T. (2008b). Anaphora resolution in near-native Italian grammars: Evidence from native speakers of Croatian. *EUROSLA Yearbook*, 8, 107–134. <https://doi.org/10.1075/eurosla.8.08kra>

- Kraš, T. (2008c). Anaphora resolution in Croatian: Psycholinguistic evidence from native speakers. U M. Tadić, M. Dimitrova-Vulchanova i S. Koeva (ur.), *Proceedings of the Sixth International Conference Formal Approaches to South Slavic and Balkan Languages* (str. 67–72). Hrvatsko društvo za jezične tehnologije – Filozofski fakultet Sveučilišta u Zagrebu.
- Kraš, T. (2016). Cross-linguistic influence at the discourse–syntax interface: Insights from anaphora resolution in child second language learners of Italian. *International Journal of Bilingualism*, 20(4), 369–385. <https://doi.org/10.1177/1367006915609239>
- Kraš, T. i Miličević, M. (2015). *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika*. Filozofski fakultet Sveučilišta u Rijeci.
- Kraš, T., Podboj, M. i Simonović, M. (2023). *Collecting language data from human participants* [Online-kolegij]. <https://upskills.fil.bg.ac.rs/login/index.php>
- Krzywinski, M. i Altman, N. (2013). Error bars. *Nature Methods*, 10(10), 921–922. <https://doi.org/10.1038/nmeth.2659>
- Lachaud, C. M. i Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, 32(2), 389–416. <https://doi.org/10.1017/S0142716410000457>
- LaFlair, G. T., Egbert, J. i Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, *t*-tests, and ANOVAs. U L. Plonsky (ur.), *Advancing quantitative methods in second language research* (str. 46–77). Routledge.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.
- Larson-Hall, J. (2012a). How to run statistical analyses. U A. Mackey i S. Gass (ur.), *Research methods in second language acquisition* (str. 245–274). Wiley-Blackwell. <https://doi.org/10.1002/9781444347340.ch13>
- Larson-Hall, J. (2012b). Our statistical intuitions may be misleading us: Why we need robust statistics. *Language Teaching*, 45(4), 460–474. <https://doi.org/10.1017/S0261444811000127>
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2. izdanje). Routledge.
- Larson-Hall, J. (2017). Moving beyond the bar plot and the line graph to create informative and attractive graphics. *The Modern Language Journal*, 101(1), 244–270. <https://doi.org/10.1111/modl.12386>

- Larson-Hall, J. i Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, 31(3), 368–390. <https://doi.org/10.1093/applin/amp038>
- Larson-Hall, J. i Mizumoto, A. (2020). Using statistical analysis software (R, SPSS). U J. McKinley i H. Rose (ur.), *The Routledge handbook of research methods in applied linguistics* (str. 385–397). Routledge.
- Larson-Hall, J. i Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(S1), 127–159. <https://doi.org/10.1111/lang.12115>
- Lazaraton, A. (2000). Current trends in research methodology and statistics in applied linguistics. *TESOL Quarterly*, 34(1), 175–181. <https://doi.org/10.2307/3588103>
- Lazaraton, A., Riggenbach, H. i Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly*, 21(2), 263–277. <https://doi.org/10.2307/3586735>
- Leal, T. (2018). Data analysis and sampling: Methodological issues concerning proficiency in SLA research. U A. Gudmestad i A. Edmonds (ur.), *Critical reflections on data in second language acquisition* (str. 63–88). John Benjamins. <https://doi.org/10.1075/llt.51.04lea>
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins. <https://doi.org/10.1075/z.195>
- Linck, J. A. i Cunnings, I. (2015). The utility and application of mixed-effects models in second language research: Mixed-effects models. *Language Learning*, 65(S1), 185–207. <https://doi.org/10.1111/lang.12117>
- Lindstromberg, S. (2016a). Guidelines, recommendations, and supplementary discussion. *Language Teaching Research*, 20(6), supplementary material 1–27. <https://doi.org/10.1177/1362168816651895>
- Lindstromberg, S. (2016b). Inferential statistics in *Language Teaching Research: A review and ways forward*. *Language Teaching Research*, 20(6), 741–768. <https://doi.org/10.1177/1362168816649979>
- Lindstromberg, S. (2020). *The assumptions of normality and similar distributions in small-scale quantitative research: Diagnostic graphs & choosing a significance test*. <https://doi.org/10.13140/RG.2.2.34876.59527>
- Lindstromberg, S. i Eyckmans, J. (2017). The particular need for replication in the quantitative study of SLA: A case study of the mnemonic effect of assonance in collocations. *Journal of the European Second Language Association*, 1(1), 126–136. <https://doi.org/10.22599/jesla.26>

- Lintz, E. N., Lim, P. C. i Johnson, M. R. (2021). A new tool for equating lexical stimuli across experimental conditions. *MethodsX*, 8, 101545. <https://doi.org/10.1016/j.mex.2021.101545>
- Loerts, H., Lowie, W. i Seton, B. (2020). *Essential statistics for applied linguistics: Using R or JASP* (2. izdanje). Bloomsbury.
- Loewen, S. i Gass, S. (2009). The use of statistics in L2 acquisition research. *Language Teaching*, 42(2), 181–196. <https://doi.org/10.1017/S0261444808005624>
- Loewen, S., Gönülal, T., Isbell, D. R., Ballard, L., Crowther, D., Lim, J., Maloney, J. i Tigchelaar, M. (2020). How knowledgeable are applied linguistics and SLA researchers about basic statistics? Data from North America and Europe. *Studies in Second Language Acquisition*, 42(4), 871–890. <https://doi.org/10.1017/S0272263119000548>
- Loewen, S., Lavolette, E., Spino, L. A., Papi, M., Schmidtke, J., Sterling, S. i Wolff, D. (2014). Statistical literacy among applied linguists and second language acquisition researchers. *TESOL Quarterly*, 48(2), 360–388. <https://doi.org/10.1002/tesq.128>
- Lowry, R. (bez dat.). *Concepts and applications of inferential statistics*. [Online-udžbenik]. <http://vassarstats.net/textbook/>
- Macaro, E. (2020). Systematic reviews in applied linguistics. U J. McKinley i H. Rose (ur.), *The Routledge handbook of research methods in applied linguistics* (str. 230–239). Routledge.
- Mackey, A. i Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates.
- Mackey, A. i Gass, S. M. (ur.). (2023). *Current approaches in second language acquisition research: A practical guide*. Wiley Blackwell.
- Mackey, B. i Ross, S. J. (2015). Bayesian informative hypothesis testing. U L. Plonsky (ur.), *Advancing quantitative methods in second language research* (str. 329–345). Routledge.
- Margaza, P. i Gavarró, A. (2020). Null/overt subject alternations in L2 Spanish and L2 Greek. *Glossa: A Journal of General Linguistics*, 5(1), 55. <https://doi.org/10.5334/gjgl.969>
- Marsden, E., Morgan-Short, K., Trofimovich, P. i Ellis, N. C. (2018). Introducing registered reports at *Language Learning: Promoting transparency, replication, and a synthetic ethic in the language sciences*. *Language Learning*, 68(2), 309–320. <https://doi.org/10.1111/lang.12284>

- Marsden, E. i Plonsky, L. (2018). Data, open science, and methodological reform in second language acquisition research. U A. Gudmestad i A. Edmonds (ur.), *Critical reflections on data in second language acquisition* (str. 219–228). John Benjamins. <https://doi.org/10.1075/lllt.51.10mar>
- Marsden, E., Thompson, S. i Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 39(5), 861–904. <https://doi.org/10.1017/S0142716418000036>
- Martín Gómez, A. (2023). The effects of age and experience with input in the acquisition of clitic climbing in heritage and L2 Spanish. *International Journal of Bilingualism*, 27(4), 432–461. <https://doi.org/10.1177/13670069221093937>
- Mayo, L. H., Florentine, M. i Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40(3), 686–693. <https://doi.org/10.1044/jslhr.4003.686>
- McClelland, G. H. (1997). Optimal design in psychological research. *Psychological Methods*, 2(1), 3–19. <https://doi.org/10.1037/1082-989X.2.1.3>
- McDonough, K. i Trofimovich, P. (2009). *Using priming methods in second language research*. Routledge.
- McGill, R., Tukey, J. W. i Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1), 12–16. <https://doi.org/10.2307/2683468>
- Milas, G. (2009). *Istraživačke metode u psihologiji i drugim društvenim znanostima*. Slap.
- Miličević, M. (2007). *The acquisition of reflexives and reciprocals in L2 Italian, Serbian and English* [Doktorska disertacija]. University of Cambridge.
- Miličević, M. (2017a). *Introduction to experimental methods in linguistics* [Online-kolegij]. <https://elearn.mnf.uzh.ch/courses/course-v1:ReLDI-Project+ReLDI103+2022/about>
- Miličević, M. (2017b). *Introduction to research methodology in linguistics* [Online-kolegij]. <https://elearn.mnf.uzh.ch/courses/course-v1:ReLDI-Project+ReLDI101+2022/about>
- Miličević, M. (2017c). *Introduction to statistics for language data* [Online-kolegij]. <https://elearn.mnf.uzh.ch/courses/course-v1:ReLDI-Project+ReLDI104+2022/about>
- Miličević, M. i Kraš, T. (2017). Translation between L2 acquisition and L1 attrition: Anaphora resolution in Italian by English–Italian trainee translators. *Applied Linguistics*, 38(1), 21–42. <https://doi.org/10.1093/applin/amu070>

- Miličević Petrović, M., Đukanović, M., Budimirović, J. i Gledić, J. (2023). *First steps into scientific research* [Online-kolegij]. <https://upskills.fil.bg.ac.rs/course/view.php?id=9>
- Miličević Petrović, M., Ferraresi, A. i Marty, P. (2023). *A glimpse into language data science* [Online-kolegij]. <https://upskills.fil.bg.ac.rs/course/view.php?id=8>
- Miller, J. (2024). How many participants? How many trials? Maximizing the power of reaction time studies. *Behavior Research Methods*, 56(3), 2398–2421. <https://doi.org/10.3758/s13428-023-02155-9>
- Mitterer, H., Eger, N. A. i Reinisch, E. (2020). My English sounds better than yours: Second-language learners perceive their own accent as better than that of their peers. *PLOS ONE*, 15(2), e0227643. <https://doi.org/10.1371/journal.pone.0227643>
- Mizumoto, A. (2015). Langtest (verzija 1.0) [Web-aplikacija]. <http://langtest.jp>
- Mizumoto, A. i Plonsky, L. (2016). R as a lingua franca: Advantages of using R for quantitative research in applied linguistics. *Applied Linguistics*, 37(2), 284–291. <https://doi.org/10.1093/applin/amv025>
- Montrul, S. (2001). The acquisition of causative/inchoative verbs in L2 Turkish. *Language Acquisition*, 9(1), 1–58. https://doi.org/10.1207/S15327817LA0901_01
- Montrul, S., Foote, R. i Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58(3), 503–553. <https://doi.org/10.1111/j.1467-9922.2008.00449.x>
- Moranski, K. i Ziegler, N. (2021). A case for multisite second language acquisition research: Challenges, risks, and rewards. *Language Learning*, 71(1), 204–242. <https://doi.org/10.1111/lang.12434>
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D. i Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103–123. <https://doi.org/10.3758/s13423-015-0947-8>
- Morgan-Short, K., Marsden, E., Heil, J., Issa II, B. I., Leow, R. P., Mikhaylova, A., Mikołajczak, S., Moreno, N., Slabakova, R. i Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and reading comprehension. *Language Learning*, 68(2), 392–437. <https://doi.org/10.1111/lang.12292>
- Mueller, J. i Jiang, N. (2013). The acquisition of the Korean honorific affix (*u*)si by advanced L2 learners. *The Modern Language Journal*, 97(2), 318–339. <https://doi.org/10.1111/j.1540-4781.2013.12005.x>

- Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, 90(3), 387–401. <https://doi.org/10.1111/j.1540-4781.2006.00431.x>
- Nahatame, S. (2018). Comprehension and processing of paired sentences in second language reading: A comparison of causal and semantic relatedness. *The Modern Language Journal*, 102(2), 392–415. <https://doi.org/10.1111/modl.12466>
- Navarro, D. (bez dat.). *Learning statistics with R*. [Online-udžbenik]. <https://learningstatisticswithr.com>
- Nguyen, M. T. T., Do, H. T., Pham, T. T. i Nguyen, A. T. (2017). The effectiveness of corrective feedback for the acquisition of L2 pragmatics: An eight month investigation. *International Review of Applied Linguistics in Language Teaching*, 56(3), 345–375. <https://doi.org/10.1515/iral-2015-0059>
- Nicklin, C. i Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics*, 40, 26–55. <https://doi.org/10.1017/S0267190520000057>
- Nicklin, C. i Vitta, J. P. (2021). Effect-driven sample sizes in second language instructed vocabulary acquisition research. *The Modern Language Journal*, 105(1), 218–236. <https://doi.org/10.1111/modl.12692>
- Norouzian, R. (2018). *Methodological reform in quantitative second language research: Effect sizes, Bayesian hypothesis testing, and Bayesian estimation of effect sizes* [Doktorska disertacija]. Texas A&M University.
- Norouzian, R. (2020). Sample size planning in quantitative L2 research: A pragmatic approach. *Studies in Second Language Acquisition*, 42(4), 849–870. <https://doi.org/10.1017/S0272263120000017>
- Norouzian, R. (2021). Interrater reliability in second language meta-analyses: The case of categorical moderators. *Studies in Second Language Acquisition*, 43(4), 896–915. <https://doi.org/10.1017/S0272263121000061>
- Norouzian, R., De Miranda, M. i Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, 68(4), 1032–1075. <https://doi.org/10.1111/lang.12310>
- Norouzian, R., De Miranda, M. i Plonsky, L. (2019). A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal*, 103(1), 248–261. <https://doi.org/10.1111/modl.12543>
- Norouzian, R. i Plonsky, L. (2018a). Correlation and simple linear regression in applied linguistics. U A. Phakiti, P. De Costa, L. Plonsky i S. Starfield (ur.), *The Palgrave handbook of applied linguistics research methodology* (str. 395–421). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1_19

- Norouzian, R. i Plonsky, L. (2018b). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, 34(2), 257–271. <https://doi.org/10.1177/0267658316684904>
- Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning*, 65(S1), 97–126. <https://doi.org/10.1111/lang.12114>
- Norris, J. M. i Ortega, L. (2003). Defining and measuring SLA. U C. J. Doughty i M. H. Long (ur.), *The handbook of second language acquisition* (str. 716–761). Blackwell. <https://doi.org/10.1002/9780470756492.ch21>
- Norris, J. M., Plonsky, L., Ross, S. J. i Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65(2), 470–476. <https://doi.org/10.1111/lang.12104>
- Öztürk, M. S. i Jorganci, M. (2017). Effects of L1 and L2 glosses on incidental vocabulary learning of EFL prep students. *Turkish Studies*, 12(6), 635–656. <https://doi.org/10.7827/TurkishStudies.11432>
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7. izdanje). Routledge.
- Paolillo, J. C. (2000). Asymmetries in Universal Grammar: The role of method and statistics. *Studies in Second Language Acquisition*, 22(2), 209–228. <https://doi.org/10.1017/S0272263100002035>
- Park, H. I., Solon, M., Dehghan-Chaleshtori, M. i Ghanbar, H. (2022). Proficiency reporting practices in research on second language acquisition: Have we made any progress? *Language Learning*, 72(1), 198–236. <https://doi.org/10.1111/lang.12475>
- Pavlović, N. i Stanojević, M.-M. (2020). *Znanstvena istraživanja jezika i prevođenja*. Filozofski fakultet Sveučilišta u Zagrebu – FF Press. <https://doi.org/10.17234/9789531759014>
- Pereira, D. G., Afonso, A. i Medeiros, F. M. (2015). Overview of Friedman's test and post-hoc analysis. *Communications in Statistics - Simulation and Computation*, 44(10), 2636–2653. <https://doi.org/10.1080/03610918.2014.931971>
- Perek, F. i Hilpert, M. (2014). Constructional tolerance: Cross-linguistic differences in the acceptability of non-conventional uses of constructions. *Constructions and Frames*, 6(2), 266–304. <https://doi.org/10.1075/cf.6.2.06per>
- Petz, B., Kolesarić, V. i Ivanec, D. (2012). *Petzova statistika: Osnovne statističke metode za nematematičare*. Naklada Slap.
- Phakiti, A. (2014). *Experimental research methods in language learning*. Bloomsbury.

- Phillips, N. D. (2017). YaRrr! The pirate's guide to R. *March Methodology Madness*, 30(3), 22–23.
- Pichette, F., Béland, S., Jolani, S. i Leśniewska, J. (2015). The handling of missing binary data in language research. *Studies in Second Language Learning and Teaching*, 5(1), 153–169. <https://doi.org/10.14746/ssl.t.2015.5.1.8>
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61(4), 993–1038. <https://doi.org/10.1111/j.1467-9922.2011.00663.x>
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35(4), 655–687. <https://doi.org/10.1017/S0272263113000399>
- Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes. A “back-to-basics” approach to advancing quantitative methods in L2 research. U L. Plonsky (ur.), *Advancing quantitative methods in second language research* (str. 23–45). Routledge.
- Plonsky, L. (2023). Sampling and generalizability in Lx research: A second-order synthesis. *Languages*, 8(1), 75. <https://doi.org/10.3390/languages8010075>
- Plonsky, L. (ur.). (u pripremi). *Open science in applied linguistics*. John Benjamins.
- Plonsky, L. i Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, 31(2), 267–278. <https://doi.org/10.1177/0267658314536436>
- Plonsky, L., Brown, D., Chen, M., Ghanem, R., Arvizu, M. N. G., Isbell, D. R. i Zhang, M. (2024). “Significance sells”: Applied linguists' views on questionable research practices. *Research Methods in Applied Linguistics*, 3(1), 100099. <https://doi.org/10.1016/j.rmal.2024.100099>
- Plonsky, L. i Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, 100(2), 538–553. <https://doi.org/10.1111/modl.12335>
- Plonsky, L., Egbert, J. i Laflair, G. T. (2015). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics*, 36(5), 591–610. <https://doi.org/10.1093/applin/amu001>
- Plonsky, L. i Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61(2), 325–366. <https://doi.org/10.1111/j.1467-9922.2011.00640.x>

- Plonsky, L. i Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*, 102(4), 713–731. <https://doi.org/10.1111/modl.12509>
- Plonsky, L. i Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97. <https://doi.org/10.1017/S0267190516000015>
- Plonsky, L., Marsden, E., Crowther, D., Gass, S. M. i Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, 36(4), 583–621. <https://doi.org/10.1177/0267658319828413>
- Plonsky, L. i Oswald, F. L. (2011). How to do a meta-analysis. U A. Mackey i S. M. Gass (ur.), *Research methods in second language acquisition* (str. 275–295). Wiley. <https://doi.org/10.1002/9781444347340.ch14>
- Plonsky, L. i Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Plonsky, L. i Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, 39(3), 579–592. <https://doi.org/10.1017/S0272263116000231>
- Popper, K. (2010/1959). *The logic of scientific discovery*. Routledge.
- Porte, G. K. (2002). *Appraising research in second language learning: A practical approach to critical analysis of quantitative research*. John Benjamins.
- Porte, G. K. i McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.
- Puskas, G., Miličević Petrović, M. i Samardžić, T. (2019). *Introduction to research in linguistics: Theory, logic, method* [Online-kolegij]. <https://elearn.mnf.uzh.ch/courses/course-v1:PHIL+Movetia101+2022/about>
- Quené, H. (2010). How to design and analyze language acquisition studies. U E. Blom i S. Unsworth (ur.), *Experimental methods in language acquisition research* (str. 269–284). John Benjamins. <https://doi.org/10.1075/llt.27.15que>
- Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction* (2. izdanje). Bloomsbury.
- Rasinger, S. M. (2018). Quantitative methods: Concepts, frameworks and issues. U L. Litosseliti (ur.), *Research methods in linguistics* (2. izdanje) (str. 117–138). Bloomsbury Academic.
- Richardson, J. T. E. (2021). A closer look at the lady tasting tea. *Significance*, 18(5), 34–37. <https://doi.org/10.1111/1740-9713.01572>

- Roberts, L. i Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29(4), 413–439. <https://doi.org/10.1177/0267658313503171>
- Robison, R. E. (1990). The primacy of aspect: Aspectual marking in English interlanguage. *Studies in Second Language Acquisition*, 12(3), 315–330. <https://doi.org/10.1017/S0272263100009190>
- Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics*, 59(5), 1227–1249. <https://doi.org/10.1515/ling-2019-0048>
- Roever, C. i Phakiti, A. (2018). *Quantitative methods for second language research: A problem-solving approach*. Routledge.
- Ross, S. J. i Mackey, B. (2015). Bayesian approaches to imputation, hypothesis testing, and parameter estimation: Bayesian approaches. *Language Learning*, 65(S1), 208–227. <https://doi.org/10.1111/lang.12118>
- Schmid, M. S. i Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, 26(1), 125–160. <https://doi.org/10.1177/0267658309337641>
- Schmitt, C. i Miller, K. (2010). Using comprehension methods in language acquisition research. U E. Blom i S. Unsworth (ur.), *Experimental methods in language acquisition research* (str. 35–56). John Benjamins. <https://doi.org/10.1075/llt.27.04sch>
- Schütze, C. T. i Sprouse, J. (2013). Judgment data. U R. J. Podesva i D. Sharma (ur.), *Research methods in linguistics* (str. 27–50). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734.004>
- Smaldino, P. (2019). Better methods can't make up for mediocre theory. *Nature*, 575, 9. <https://doi.org/10.1038/d41586-019-03350-5>
- Sönning, L. i Werner, V. (2021). The replication crisis, scientific revolutions, and linguistics. *Linguistics*, 59(5), 1179–1206. <https://doi.org/10.1515/ling-2019-0045>
- Stanton, J. M. (2001). Galton, Pearson, and the Peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3). <https://doi.org/10.1080/10691898.2001.11910537>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Sunderland, J. (2018). Research questions in linguistics. U L. Litosseliti (ur.), *Research methods in linguistics* (2. izdanje) (str. 13–34). Bloomsbury Academic.

- Tabachnick, B. G. i Fidell, L. S. (2001). *Using multivariate statistics* (4. izdanje). Allyn & Bacon.
- Tabachnick, B. G. i Fidell, L. S. (2013). *Using multivariate statistics* (6. izdanje). Pearson Education.
- Tavakoli, H. (2013). *A dictionary of research methodology and statistics in applied linguistics*. Rahnamā.
- Tazik, K. (2019). Validity and reliability reports in applied linguistics research articles: The case of tests and questionnaire. *International Journal of Foreign Language Teaching & Research*, 7(28), 121–134.
- Tenjović, L. i Smederevac, S. (2011). Mala reforma u statističkoj analizi podataka u psihologiji: Malo p nije dovoljno, potrebna je i veličina efekta. *Primenjena psihologija*, 4(4), 317–333. <https://doi.org/10.19090/pp.2011.4.317-333>
- Testa, M. (2019). The acquisition of L3 Spanish articles: What can be learned from a simple linear regression analysis? *Research in Corpus Linguistics*, 7, 84–112. <https://doi.org/10.32714/ricl.07.05>
- Todorović, D. (2008). *Metodologija psiholoških istraživanja*. Centar za primenjenu psihologiju.
- Tonković, M., Dumančić, F. i Anđel, M. (2020). Efekt stranog jezika i odlučivanje o moralu. *Suvremena lingvistika*, 45(88), 213–230. <https://doi.org/10.22210/suvlin.2019.088.06>
- Tošković, O. (2020). *Autostoperski vodič kroz statistiku*. Centar za primenjenu psihologiju.
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, 33(3), 339–372. <https://doi.org/10.1017/S0272263111000015>
- Trochim, W. M. K. (bez dat.). *The research methods knowledge base*. [Online-udžbenik. <https://conjointly.com/kb/>
- Tufte, E. R. (2001). *The visual display of quantitative information* (2. izdanje). Graphics Press.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.
- Upton, G. i Cook, I. (1996). *Understanding statistics*. Oxford University Press.
- Vasishth, S. (2023). New directions in statistical analysis for experimental linguistics. U S. Zufferey i P. Gygax (ur.), *The Routledge handbook of experimental linguistics* (str. 313–329). Routledge. <https://doi.org/10.4324/9781003392972-23>
- Vitta, J. P., Nicklin, C. i McLean, S. (2022). Effect size–driven sample-size planning, randomization, and multisite use in L2 instructed vocabulary acquisition experimental samples. *Studies in Second Language Acquisition*, 44(5), 1424–1448.

- Wallis, S. (2021). *Statistics in corpus linguistics: A new approach*. Routledge.
- Wang, F. H. i Kaiser, E. (2022). Linguistic priming and learning adjacent and nonadjacent dependencies in serial reaction time tasks. *Language Learning*, 72(3), 695–727. <https://doi.org/10.1111/lang.12491>
- Warne, R. T. (2017). *Statistics for the social sciences: A general linear model approach*. Cambridge University Press. <https://doi.org/10.1017/9781316442715>
- Watorek, M., Rast, R., Yu, X. C., Trévisiol, P., Majdoub, H., Guan, Q. i Huang, X. (2020). Replication: Measuring the influence of typologically diverse target language properties on input processing at the initial stages of acquisition. U A. Edmonds, P. Leclercq i A. Gudmestad (ur.), *Interpreting language-learning data* (str. 71–110). Language Science Press. <https://doi.org/10.5281/ZENODO.4032284>
- Webb, S., Newton, J. i Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91–120. <https://doi.org/10.1111/j.1467-9922.2012.00729.x>
- Wilkinson, L., i Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>
- Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.
- Winter, B. (2022). Mapping the landscape of exploratory and confirmatory data analysis in linguistics. U D. Tay i M. X. Pan (ur.), *Data analytics in cognitive linguistics* (str. 13–48). De Gruyter. <https://doi.org/10.1515/9783110687279-002>
- Winter, B., i Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, 59(5), 1251–1277. <https://doi.org/10.1515/ling-2019-0049>
- Wood, M., Larsson, T., Plonsky, L., Sterling, S., Kytö, M. i Yaw, K. (2024). *Addressing questionable research practices in applied linguistics: A practical guide*. Applied Linguistics Press.
- Woodrow, L. (2014). *Writing about quantitative research in applied linguistics*. Palgrave Macmillan. <https://doi.org/10.1057/9780230369955>
- Woods, A., Fletcher, P. i Hughes, A. (1986). *Statistics in language studies*. Cambridge University Press.
- Wu, S. i Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46(4), 680–704. <https://doi.org/10.1111/flan.12063>

- Yang, M., Cooc, N. i Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1), 15. <https://doi.org/10.1186/s40862-017-0036-9>
- Zhang, P. i Han, C. (2023). Examining statistical literacy, attitudes toward statistics, and statistics self-efficacy among applied linguistics research students in China. *International Journal of Applied Linguistics*, 34(2), 433–449. <https://doi.org/10.1111/ijal.12500>

Kazalo pojmova

A

alfa razina (v. *i* pogreška tipa I) 136–138, 149, 169, 185, 189, 236

Američko psihološko udruženje 12, 56, 147, 224, 235–236

analiza

analiza glavnih komponenti 9, 156

analiza osjetljivosti 107

analiza pouzdanosti 73, 91

analiza prema česticama 76, 187, 202–204, 215

analiza prema ispitanicima 76, 187, 202–204, 215

analiza statističke snage testa 149–151

analiza varijance, v. ANOVA
faktorska analiza 9, 73

ANOVA 87, 90, 152, 156, 163, 190, 192–204, 212–215, 251

ANOVA za nezavisne uzorke 190–193, 195–197, 199

ANOVA za nezavisne skupine, v.

ANOVA za nezavisne uzorke

ANOVA za zavisne uzorke 190, 195–197, 200

ANOVA za ponovljena mjerenja, v.

ANOVA za zavisne uzorke

ANOVA za mješoviti nacrt 197–198

faktorijska ANOVA, v. višefaktorska ANOVA

jednofaktorska ANOVA 190, 195, 199–200

jednosmjerna ANOVA, v.

jednofaktorska ANOVA

jednostavna ANOVA, v.

jednofaktorska ANOVA

multivarijatna ANOVA 156

složena ANOVA, v. višefaktorska ANOVA

višefaktorska ANOVA 197

Ansari-Bradleyev test 178

aritmetička sredina (v. *i* prosjek) XIII,

5, 39, 64, 82–90, 92, 95, 100, 102,

105–107, 110, 114, 118, 123–124, 131,

141–148, 152, 157–160, 165, 176–179,

181–184, 190–191, 194, 202, 207–208,

217, 234, 239, 242,

podrezana aritmetička sredina 110, 217

zajednička aritmetička sredina 86

B

Bayesovo statističko zaključivanje, v. statističko zaključivanje

binomni test 154, 174, 180

C

centil 93

centralna tendencija, v. mjera centralne tendencije (pod mjera)

centralna vrijednost, v. medijan

Cochranov Q test 175

Cohenov *d* 141, 150, 183, 198

Cohenova kapa 71–72

Conoverov test 201

Cramérov koeficijent *V* 171

Cronbachov koeficijent alfa 71

Č

čestica (v. *i* podražaj) 1–2, 35–36, 39–40, 42, 58–59, 61, 76, 86, 151, 164, 187, 202–204, 214

eksperimentalna čestica 2, 39–40,

54, 59, 61–62, 86, 189, 217

analiza prema česticama, v. analiza

D

decimalna oznaka 55

decimalna točka 55–56, 236

decimalni zarez 55–56, 236

deduktivni pristup (v. *i* hipotetičko-

inferencijalni pristup) 4, 9, 35

dijagram (v. *i* grafikon) XVI, 120

dijagram raspršenja 104, 123–127, 206–207

dijagram rasipanja, v. dijagram

raspršenja

dijagram s pravokutnikom 105, 120–

123, 149, 184, 188

kružni dijagram 96, 113

paralelni dijagram 120

distraktor (v. *i* podražaj) 37–38, 40, 48, 51, 231

distribucija

asimetrična distribucija 90, 102, 106, 109, 121–122, 124, 159

bimodalna distribucija 103

binomna distribucija 154

distribucija frekvencija 86, 96, 100,

103–104, 112, 118, 121, 166, 224,

234

distribucija podataka 76–77, 83, 86, 88, 90, 96–97, 99, 105–106, 108, 111–112, 155, 157, 159–160, 178, 184, 218, 237
distribucija statistika 152, 200
distribucija vjerojatnosti 100, 178, 188, 190
F-distribucija 194
hi-kvadrat distribucija 169–170, 173–174, 200–201
log-normalna distribucija 106
negativno asimetrična distribucija 102
normalna distribucija 83, 88, 90, 96, 100–107, 109–111, 144, 157–161, 178–179, 182, 185, 187, 195, 202, 207, 218
pozitivno asimetrična distribucija 102
rep distribucije 102, 159
simetrična distribucija 90, 154
t-distribucija 180, 188, 209
dominantna vrijednost, v. mod
Dunnov test 201

E

efekt 33, 40, 132, 151, 153, 157, 159, 171–172, 181–183, 186–187, 194, 203, 213
glavni efekt 43, 197
interakcijski efekt (v. *i* interakcija) 43, 197–198, 213
fiksni efekt 39
slučajni efekt 39–40, 46, 164
veličina efekta 139–142, 148–150, 171–172, 183–184, 187–188, 198, 206, 209, 235, 242, 246, 247, 252
eksperiment 1–4, 9, 17, 26–27, 29–30, 34, 38–39, 47, 51–53, 55, 63, 76, 102, 130–131, 143–145, 148–149, 151, 166, 202–203, 223, 228, 230–231, 240, 244
eksperimentalna čestica, v. čestica
eksperimentalna metoda 1–2
eksperimentalna skupina, v. skupina
eksperimentalni nacrt, v. nacrt
istraživanja
eksperimentalni podražaj, v. podražaj
eksperimentalni pokušaj 40
eksperimentalni postupak 52
eksperimentalni tretman 2–3
eksperimentalni uvjet 37–39, 40–41, 47–48, 58, 62, 76, 85–86, 94, 97, 107–108, 115, 186–190, 195–196,

200, 202–203, 214, 216, 230–231, 234
eksperimentalni zadatak, v. zadatak
neeksperimentalna metoda 164, 206
statistički eksperiment 133–135
ekstremna vrijednost (v. *i* netipična vrijednost) 60, 88, 105, 110

F

F-omjer 190, 192–193, 196, 199, 215
F-test 178, 182
faktor 44, 190–191, 197–198, 202, 224
fiksni faktor 214, 217
slučajni faktor 204, 214–215, 217
faktorska analiza, v. analiza
Fisherov egzaktni test 172
format (prikaza zavisnih uzoraka)
dugački format 59–60
široki format 59–60
frekvencija XIII, XV, 5, 62, 76, 77–79, 81–82, 85–86, 95–97, 100, 103–104, 112–114, 117, 166, 169, 172–174, 214, 224, 234
apsolutna frekvencija 77, 81–82, 97–99
gustoća frekvencije 98
kumulativna frekvencija 81, 99
marginalna frekvencija 168, 173
opažena frekvencija 116, 167–168, 172, 174
relativna frekvencija 77–79, 81–82, 97–98, 113, 135, 169, 214
teorijska frekvencija 116–168, 172
Friedmanov test 199–202
Friedmanova ANOVA, v. Friedmanov test

G

G-test 172
Gaussova krivulja, v. krivulja
generalizacija (v. *i* uopćavanje) 7
grafikon (v. *i* dijagram) XV–XVI, 64–65, 75, 96–97, 111–116, 122, 124–125, 127, 198, 206, 234–235, 251
100 % složeni stupčasti grafikon 114–115
asocijativni grafikon 116–117, 170, 172
grafikon kvantila (v. *i* Q-Q grafikon) 103, 112
grafikon sa stupcima, v. stupčasti grafikon

grafikon s okvirima i brkovima 120–121, 123, 184
grafikon s oznakom pogreške 118–119
linijski grafikon 99, 119–120, 146, 198
mozaični grafikon 115–116, 166, 170, 172
piratski grafikon 123–124, 184
Q-Q grafikon (v. *i* grafikon kvantila) 103–104, 112
složeni stupčasti grafikon 114
stupčasti grafikon 97, 113–114, 116–118, 121, 142, 146
violinski grafikon 124
granice pouzdanosti 118, 123–124, 142–149, 235, 245
granice povjerenja, v. granice pouzdanosti
granice sigurnosti, v. granice pouzdanosti
gustoća frekvencije, v. frekvencija
gustoća vjerojatnosti, v. vjerojatnost

H

hi-kvadrat test 76, 163, 167–173, 176, 178–180, 184, 187, 200, 206
 hi-kvadrat test prikladnosti modela 167, 169
 hi-kvadrat test nezavisnosti 167–168
 Pearsonov hi-kvadrat test 167
hipotetičko-inferencijalni pristup (v. *i* deduktivni pristup) 8–9
hipoteza XVI, 1, 3–4, 8–10, 14–15, 17–22, 79, 80, 103, 132, 139, 149, 159, 167, 176–177, 179, 181, 190, 212, 217, 225, 226, 228–229, 231, 233, 236–237, 239, 244
 alternativna hipoteza 131, 137
 dvosmjerna hipoteza 132, 154, 176
 istraživačka hipoteza 18, 22, 53, 129–132, 136–137, 151, 176, 226, 227, 251
 jednosmjerna hipoteza 132, 176
 nulta hipoteza 131, 133–139, 152, 154, 160, 163, 167–169, 171, 188–189, 193, 206, 208–209, 211, 239, 245–247, 251
 statistička hipoteza 130–133, 176
histogram 96–100, 103, 112, 114, 160–161
 histogram kumulativnih frekvencija 99

homogenizacija 49

I

IBM SPSS 56, 59, 62, 66–70, 146, 181, 218
implikacije (istraživanja)
 pedagoške implikacije 237
 praktične implikacije 222, 237
 teorijske implikacije 226, 237
IMRaD 222
induktivni pristup 4, 9
instrument XIII, 23, 40, 53, 70–73, 76, 102, 177, 222, 232, 244, 249
 mjerni instrument 23, 27, 50–51, 103, 230, 251
 psihometrijske značajke instrumenta 51, 70, 230, 232, 249
 stupanj pouzdanosti instrumenta 206, 230, 232, 239
interakcija (v. *i* interakcijski efekt) 40, 43, 48, 119–120, 197–198, 213
intercept, v. odsječak
interkvartil, v. interkvartilno raspršenje
interkvartilna razlika, v. interkvartilno raspršenje
interkvartilni raspon, v. interkvartilno raspršenje
interkvartilno raspršenje (v. *i* poluinterkvartilno raspršenje) 90–92, 95, 102, 105, 110, 121, 188, 207, 234
interval pogreške, v. pogreška
IRIS (*Instruments and Data for Research in Language Studies*) 63, 230
ispitanik 2, 9, 19, 22, 26–43, 46–48, 51, 53–55, 57–63, 73, 76–78, 80–81, 86, 89, 94–97, 101, 106–108, 110, 113–114, 117, 119, 137, 142, 149, 151, 164, 166, 176, 181, 183, 187, 191, 195–196, 198, 200–204, 206–207, 209, 211, 214, 216–218, 227–234, 237, 242, 251
 broj ispitanika 26, 33–34, 37–38, 46, 106, 138, 143–145, 149–151, 161, 178, 193, 209, 237, 249
 uzorkovanje ispitanika 18, 28–33, 37–38, 41, 52, 176, 183, 228–229, 240
 varijabilnost između ispitanika, v. varijabilnost
 varijabilnost unutar ispitanika, v. varijabilnost
istraživački problem 18, 226–228, 233, 237

istraživačko pitanje 1–2, 7, 11, 18–22, 28, 38, 52–53, 148, 181, 226–229, 231, 233, 236–237, 239

istraživanje

ciljevi istraživanja (v. *i* istraživačko pitanje) 7–9, 19, 212, 226, 232
eksplanatorno istraživanje 9, 244
eksploracijsko istraživanje 9–10, 20, 244
faza istraživanja XIII, XV, 7–8, 18, 52–53, 55, 61, 96, 129, 231, 239
kvalitativno istraživanje 4
kvantitativno istraživanje XIII–XIV, 4, 8–9, 11, 70, 75, 130, 221, 246
objasnidbeno istraživanje, v. eksplanatorno istraživanje
predmet istraživanja 1–2, 18–19, 26, 28, 34, 37–38, 40, 165, 202
priprema istraživanja XIII, XV, 17–18, 52, 150, 157

izvještaj o istraživanju (struktura)

metoda i postupak 222, 227, 232, 236
naslov 222, 224, 228
pregled literature 225, 227–228, 241–242, 244
popis literature 222, 224
rasprava i zaključci 222–223, 236–238
rezultati 222–223, 225–226, 228–230, 232–238
sažetak 222, 224
smjernice za buduća istraživanja 238
uvod 222, 225, 227–228, 236–237, 244

J

jedinica 30–31, 33, 49, 77, 90, 110, 166
jedinica analize 28, 39–40, 76, 79–80, 166, 202–203, 206
jedinica mjerenja 23, 25–26, 141
jedinica uzorkovanja 29, 40

K

kodiranje podataka, v. podatak
koeficijent 71–72, 126, 144, 171, 213, 235
koeficijent determinacije 198, 209, 211, 213
koeficijent korelacije, v. korelacija
koeficijent phi 171
koeficijent pouzdanosti 62, 73, 91

koeficijent simetričnosti distribucije 146

koeficijent slaganja 72

koeficijent spljoštenosti krivulje 146

koeficijent varijabilnosti 89, 95

Kolmogorov-Smirnov test 160

kontingencijska tablica, v. tablica

kontrola 2–3, 9, 41, 44, 164, 227

eksperimentalne tehnike kontrole 47–48

kontrola kvalitete istraživanja 7, 28, 50, 70, 140, 151, 239–240, 242–244

kontrola vanjskih varijabli 19, 35, 40, 46–48, 51, 229

neeksperimentalne tehnike kontrole 49

kontroliranje blokovima 49

korekcija 161, 174, 186, 189, 193, 200–201, 210

Benjamini–Hochberg korekcija 189

Bonferronijeva korekcija 189

Holm–Bonferroni korekcija 189

Yatesova korekcija 172

korelacija 76, 125, 163, 198, 204–212

biserijalna korelacija 204

jačina korelacije 205, 207, 209

kanonička korelacijska analiza 156

koeficijent korelacije 27, 141, 145,

187, 205–211, 235–236

korelacijska matrica 127

korelacijski nacrt, v. nacrt istraživanja

korelacijsko-regresijski nacrt, v. nacrt istraživanja

linearna korelacija 126, 204, 207

negativna korelacija 205–206

Pearsonov koeficijent korelacije 141, 208–211

pozitivna korelacija 205–206

smjer korelacije 205–206

Spearmanov koeficijent korelacije 141, 210–211

test korelacije 207–209

kovarijanca 208

Krippendorffov koeficijent alfa 72

kritična regija 154, 169–170, 180, 194

kritična vrijednost 152, 154, 169, 172,

174, 179–181, 185–188, 193–194, 200, 208–209, 211

krivulja 146, 180, 194, 207

Gaussova krivulja 100

krivulja gustoće 98–99, 124

krivulja kumulativnih frekvencija 99

Loessova krivulja 126, 207
zvonasta krivulja, v. Gaussova
krivulja
Kruskall-Wallisov test 199–202
kurtičnost, v. spljoštenost
kvadrat 92, 192–194, 200
 najmanji kvadrat 92, 213
 srednja suma kvadrata 193
 suma kvadrata 92, 192–193, 210
kvadrirana eta 141, 198
kvantil 93–94, 103, 104, 278
kvantitativna znanstvena paradigma 4,
8–9
kvartil 90–95, 121

L

latinski kvadrat 47–48
Leveneov test 178
LibreOffice Calc 60, 64–65
Loessova krivulja, v. krivulja
logaritamska transformacija 106, 109
loglinearna analiza 172

LJ

ljestvica, v. mjerna ljestvica

M

Mann-Whitneyev test 184–186, 188,
199, 201
Mauchlyev test sferičnosti 196
McNemarov test 173–175
McNemar-Bowkerov test 175
medijan 83–87, 90–95, 102–103, 107,
110, 114, 118, 121–122, 141, 160, 163,
176–178, 184, 188, 202, 207, 234
 zajednički medijan 86
metaanaliza XIV, 11, 31, 63, 72, 91, 111,
140, 142, 162–163, 242–243, 246
metoda analize
 kvantitativna metoda analize XIII, 68,
 75, 130, 247, 250
Microsoft Excel 59–60, 64, 66, 79, 97,
144
mjera XIII, XV, 8, 11, 24, 59, 61, 68,
70–73, 78–79, 103, 105, 108, 114, 138,
140–142, 145, 147–148, 150, 152,
158–160, 171, 183–184, 188, 198, 202,
207–208, 242–243, 247
 deskriptivna mjera XIV, 17, 26, 43,
 64, 75, 77, 87, 95, 111–112, 117,
 129, 142, 184, 188, 207–209, 212–
 213, 218–219

direktna mjera 39
frekvencijska mjera 77–79, 96–97,
114, 234, 239
indirektna mjera 39
izvedena mjera 39, 58, 77, 86, 95,
97, 203–204
mjera centralne tendencije 58, 62,
82–83, 86–87, 96, 104–105, 107,
112, 117–119, 123, 163, 203, 234,
239
mjera disperzije, v. mjera
varijabilnosti
mjera pogreške 92, 146
mjera prebrojavanja, v. frekvencijska
mjera
mjera raspšenja podataka 76
mjera raspšenosti, v. mjera
varijabilnosti
mjera relativnoga položaja, v.
položajna mjera
mjera učestalosti, v. frekvencijska
mjera
mjera varijabilnosti 87–91, 105–106,
112, 117, 142–143, 146, 176, 182,
203, 213, 234, 239
neposredna mjera (v. *i* direktna
mjera) 39, 58–59, 202
nerobusna mjera 83
položajna mjera 82, 84, 92–95, 121,
234
posredna mjera (v. *i* indirektna mjera;
izvedena mjera) 39, 58–59, 202
robusna mjera 85, 217–218
mjerjenje 4, 18, 22–27, 36–37, 39, 42,
50, 52, 58, 76, 83, 95, 106, 110, 146,
151, 157, 210, 231, 239, 243, 251
 jedinica mjerenja, v. jedinica
 ponovljena mjerenja 45, 47, 71, 120,
 176, 190, 195–197, 200, 214
 točnost mjerenja 27
 višestruko mjerjenje 27, 37, 50,
 58–59, 77
mjerljivost 20, 22–23
mjerna ljestvica 22, 24–27, 36, 43, 56–
57, 75, 83, 85–87, 92, 94–95, 97, 110,
155–156, 158, 160, 197–198, 237
 apsolutna ljestvica 26
 diskretna ljestvica 97
 intervalna ljestvica 25–27, 54, 83, 85,
 87–88, 93–95, 158–160, 178, 184,
 210
 kontinuirana ljestvica 97

nominalna ljestvica 24–27, 77, 79, 83, 95
omjerna ljestvica 25–27, 83, 85, 88–89, 93–95, 158–160, 178, 184, 210
ordinalna ljestvica 24–27, 54, 83, 85, 87, 93–95, 122, 158–160, 184, 210
mjerni instrument, v. instrument
mod 83–86, 92, 95, 103
model 6, 92, 126, 157, 160, 167, 169, 199, 212–215
 mješoviti model 59, 164, 202, 204, 214–215, 217
 statistički model 92
modeliranje (v. *i* model) 92, 126
Monte Carlo metoda 218
multikolinearnost 213

N

nacrtr istraživanja 11, 14–15, 18, 37, 41–49, 52, 75–77, 103, 112–113, 130, 133, 139, 151, 154–158, 164–167, 173, 175, 189–190, 197, 202, 204, 206, 212–213, 215, 228, 230–231, 239–240, 251
 bivarijantni nacrt 42
 eksperimentalni nacrt 41, 151
 faktorijalni nacrt (v. *i* višefaktorski nacrt) 43–45, 164, 176
 frekvencijski nacrt 43, 76–77, 164, 166
 jednofaktorski nacrt 43, 199
 korelacijski nacrt (v. *i* korelacijsko-regresijski nacrt) 41, 43, 76, 158, 164, 204–206
 korelacijsko-regresijski nacrt (v. *i* korelacijski nacrt) 43, 45, 164
 kvazieksperimentalni nacrt 41, 164
 mješoviti nacrt 45, 197–198, 231
 multivarijantni nacrt 42, 156
 nacrt između ispitanika (v. *i* nacrt na nezavisnim skupinama, nacrt s nezavisnim uzorcima) 45, 176
 nacrt na nezavisnim skupinama (v. *i* nacrt između ispitanika, nacrt s nezavisnim uzorcima) 45
 nacrt ponovljenih mjerenja (v. *i* nacrt sa zavisnim uzorcima, nacrt unutar ispitanika) 45
 nacrt s nezavisnim uzorcima 45–48, 157, 231
 nacrt sa zavisnim uzorcima 45–48, 157, 231
 nacrt unutar ispitanika (v. *i* nacrt

ponovljenih mjerenja, nacrt sa zavisnim uzorcima) 45, 176
 nacrt varijance 44, 158, 164–166, 177
 univarijantni nacrt 42, 156
 višefaktorski nacrt (v. *i* faktorijalni nacrt) 43–44, 156–157, 176, 197
 nagib (regresijskog pravca) 126, 213
 nedostajući podatak 60–63, 106, 232
 nasumično nedostajući podatak 61
 nenasumično nedostajući podatak 61
 potpuno nasumično nedostajući podatak 61
Nemenyiev test 201
nepostojeći podatak, v. nedostajući podatak
nerobusnost 83

 nerobusna mjera, v. mjera
 nerobusna statistika, v. statistika
 nerobusni test 217

netipična vrijednost 60–61, 83, 85, 92, 96, 104–111, 118, 121–122, 126, 159, 218, 232, 234
normalizacija (v. *i* standardizacija) 78, 109

O

objasnidbena istraživanja, v. eksplanatorna istraživanja
objašnjenje (kao cilj istraživanja) 7, 9, 14
odsječak 126, 213
ogiva 99
opći linearni model 199, 213
operacionalizacija varijabli 23, 44, 51–52, 110, 130, 226
 višestruka operacionalizacija 51–52
opisivanje (kao cilj istraživanja) 7, 9
opovrgljivost (hipoteze) 20
opservacija 71, 75, 77, 82
organizacija podataka 4, 8, 53, 55, 58–61
ovisnost 3, 107

P

parametar 76, 83, 129, 131, 143, 148, 150, 152, 158–159, 179, 245
 neparametrijski test 109, 158–159, 161, 165, 176, 184, 187–188, 202, 205, 207, 210, 212, 218, 235, 251
 parametrijski test 152, 158–162, 164–165, 176–177, 186–189, 198, 202, 207, 211–212, 214, 235

parametrijski uvjet 158, 165, 176, 187, 198–199, 205, 207–208, 232, 235
 Pearsonov hi-kvadrat test, v. hi-kvadrat test
 Pearsonov koeficijent korelacije, v. korelacija
 percentil 93
 pilotiranje 27, 35, 37, 52–53, 73, 150, 203, 230, 244
 podatak
 kodiranje podataka 8, 53, 55–56, 60–61, 63, 79, 166,
 obrada podataka 4, 8, 55, 232
 podatak s velikom težinom 61
 prikupljanje podataka XIII, 1–4, 9, 11, 17–19, 29, 53–54, 63, 129, 149–150, 221, 229, 231–232, 244
 pročišćavanje podataka 108–109
 podražaj (v. *i* čestica) 1, 27–28, 34–41, 43, 45–46, 48–49, 52–54, 56–57, 63, 96, 175, 187–188, 215, 218, 230–231
 broj podražaja 37–40, 150, 218, 231
 eksperimentalni podražaj 37–38, 54, 231
 izrada podražaja 35
 kategorija podražaja 36, 45
 lista podražaja 48
 redosljed prikaza podražaja 48, 231
 skupina podražaja 36
 uzorkovanje podražaja 18, 34–35, 39–41, 202, 214, 218, 251
 podrezivanje 106, 108
 pogreška (v. *i* rezidual) 6, 53, 71, 92, 118–119, 121, 133, 136–138, 146, 191, 193–196, 199, 245
 interval pogreške 118, 121, 142, 146–147
 pogreška tipa I 133, 136–138, 153, 189, 201–202, 218
 pogreška tipa II 133, 137–138
 pogreška uzorkovanja 34
 slučajna pogreška 27
 standardna pogreška 118–119, 142–146, 182, 194
 ukupna vjerojatnost pogreške tipa I 189
 poligon frekvencija 96, 99–100, 112
 poluinterkvartilno raspršenje (v. *i* interkvartilno raspršenje) 91–92
 populacija 5, 7, 9, 28–35, 37, 39, 46, 49–51, 75–76, 83, 88, 101, 103, 129–131, 143–148, 152, 158–161, 176–179, 183, 209, 212, 232, 234, 240
 postotak XIV, 4, 24, 29, 39, 58, 71, 75, 78, 85, 96, 105, 107–108, 144–145, 186, 195, 203, 214
 pouzdanost 50–52, 62, 72–73, 178, 227, 240, 252
 mjera pouzdanosti 72, 91, 206
 pouzdanost instrumenta XV, 17, 70–71, 73, 206, 230, 232, 239, 243, 251
 unutarnja pouzdanost 51, 71
 vanjska pouzdanost 51, 71
 predikcija (v. *i* predviđanje) 3
 predviđanje 7, 9, 14, 21–22, 64, 129, 131–133, 148, 212–213, 226, 228
 pretprijava istraživanja 41, 111, 140, 244–246
 pretprocesiranje podataka 55
 prijavljeni izvještaj 244–245
 prijavljeni izvještaj o replikaciji 244–245
 prilagodba modela 212
 prirodno variranje 134, 136, 135, 177
 psihometrijska analiza 51, 70, 230, 232, 249
 procjena 5, 34, 37, 62, 70, 76, 89, 90, 98, 103, 129, 131–132, 142, 143–144, 146–148, 158–161, 171, 176–179, 187, 245–246
 program za statističku analizu XVI, 11, 14, 24, 34, 53, 55–56, 59–62, 64–70, 72, 79, 94, 97–98, 109, 112, 124, 141–142, 144, 146, 150, 152–153, 165, 170, 172, 174, 180–181, 184–187, 193–194, 200, 210–211, 218
 proporcija 78–79
 prosjek (v. *i* aritmetička sredina) XIV, 4–6, 33, 49, 58, 75, 82–83, 85–90, 92–94, 100–103, 106–107, 110, 119, 126, 146–147, 158–160, 177, 182–183, 186, 191, 195–196, 200, 203, 212, 215, 249

R

R 56, 59, 62, 66–70, 124, 146, 181, 218
 randomizacija 48
 rang 24–25, 87, 93–94, 159–160, 184–187, 199–201, 203, 210
 izjednačeni rang, v. vezani rang
 vezani rang 93, 200–201, 210
 rangiranje 93–94, 160, 185–186, 199–200
 raspodjela (v. *i* distribucija) 41, 48, 96, 224

raspon 38, 49, 72, 76, 78, 89–90, 95–98, 100, 107–108, 121, 127, 136, 144–145, 148, 205, 234
 razina statističke značajnosti (v. *i* alfa razina; *p*-vrijednost) 136–137, 149–150, 153, 169, 189, 193, 195, 235–236
 regresija 19, 61, 125, 158, 163–164, 199, 202, 209, 212–215
 binarna logistička regresija 214, 216
 jednostavna linearna regresija 212–213, 215
 linearna regresija 199, 211–215
 logistička regresija 213–214
 politomna logistička regresija 214, 216
 regresijska jednadžba 212
 regresijski pravac 125–127, 212–213
 višestruka regresija 213, 215
 replikabilnost 51, 247
 replikacija 63, 71, 140, 148, 227, 232–233, 240–242, 244–248
 djelomična replikacija 240
 izravna replikacija 240
 konceptualna replikacija 240
 višelokacijska replikacija 240
 reprezentativnost 29–31, 33–35, 52, 80–81, 173–174, 218
 rezidual (v. *i* pogreška) 170–171, 191
 standardizirani rezidual 116
 robusnost 218
 robusna analiza 151, 218
 robusna mjera, v. mjera
 robusna statistika, v. statistika
 robusni test 159, 162–163, 218

S

sferičnost 196
 Shapiro-Wilk test 160, 182, 195
 signal 182, 190, 208
 sinteza istraživanja XV, 242–243
 skupina ispitanika
 kontrolna skupina 32, 42, 49, 78, 80, 186, 196, 216, 237, 241
 eksperimentalna skupina 32
 skupina podražaja, v. podražaj
 slučajna raspodjela ispitanika, v. randomizacija
 snaga statističkog testa, v. statistička snaga testa
 Spearmanov koeficijent korelacije, v. korelacija
 spljoštenost 100, 109, 146, 234

srednja vrijednost, v. medijan
 standardizacija (v. *i* normalizacija) 57
 standardna devijacija 88–92, 95–96, 100–102, 105–107, 118, 123, 141–143, 146–147, 158–160, 179, 181, 183–184, 192, 202, 207–208, 234
 standardna ljestvica 95
 standardna vrijednost (v. *i* *z*-vrijednost) 95
 statistička snaga testa 52, 133, 138, 149–151, 159, 163, 181, 189, 199, 215, 218, 247
 statistička značajnost, v. značajnost
 statistički eksperiment, v. eksperiment
 statistički program, v. program za statističku analizu
 statistički test, v. test
 statističko zaključivanje 4, 90, 129–130, 138, 141, 188, 245, 247, 251
 Bayesovo statističko zaključivanje 130, 138, 140, 151, 217
 frekvencijsko statističko zaključivanje 130, 135, 143
 statistik 76, 135, 143, 152, 185
 statistik *H* 199–200
 statistik *KS* 160
 statistik *U* 184–185
 statistik *W* 160, 186–187
 testni statistik 135, 137, 152, 159, 161, 167, 173, 178, 184–185, 187–188, 192, 194–195, 198, 200, 204, 206, 212, 219, 235
 statistika
 analitička statistika 5
 Bayesova statistika 14
 deskriptivna statistika 5–6, 75, 163, 232–234
 inferencijalna statistika (v. *i* statistika zaključivanja) 5–8, 32, 39, 129, 136, 149, 232, 234–236, 246
 matematička statistika 4
 primijenjena statistika 4
 robusna statistika 110, 140, 217–218, 252
 statistika zaključivanja (v. *i* inferencijalna statistika) 5, 129–130
 teorijska statistika 4
 stršeća vrijednost, v. netipična vrijednost
 strukturni krug, v. kružni dijagram (pod dijagram)
 Studentov *t*-test, v. *t*-test
 stupanj slaganja između ocjenjivača 71–72

stupanj slaganja unutar ocjenjivača
71–72
stupanj slobode 152, 169–170, 172,
179–181, 183–185, 187–188, 192–194,
196, 198, 200, 208–209, 211, 235
sudionik (v. *i* ispitanik) 1, 17, 21, 228
susret (ispitanika i podražaja) 40

Š

šum 6, 182, 190, 208, 240

T

t-vrijednost 178–181, 183–184, 199,
209, 235
t-test 87, 150, 163, 177–178, 181–185,
187–190, 193–194, 209, 211–213, 217,
235, 251
 Studentov *t*-test 177
 t-test za jedan uzorak 183
 t-test za nezavisne uzorke 178–179,
 181–182, 190, 194–195, 199, 204
 t-test za zavisne uzorke 178, 181,
 186, 190
 Welchov *t*-test 178
tablica 81, 88, 113, 172, 184, 203,
234–235
 frekvencijska tablica 81
 kontingencijska tablica 79–80, 166,
 168, 170–174
 obradna tablica 59
 statistička tablica 152–153
 tablica kritičnih vrijednosti 152, 169,
 172, 179–181, 185–187, 193–194,
 200, 208–209, 211
 unakrsna tablica, v. kontingencijska
 tablica
tehnika praćenja pokreta očiju 223, 229
test
 dvosmjerni test 154, 178, 180, 183,
 184, 209
 inferencijalni test 5, 17, 90, 110,
 129–130, 135, 152–153, 161, 182,
 207–209, 212, 218, 235, 239
 jednosmjerni test 154, 178, 180, 194,
 209
 post hoc test 194–198, 201
 test ispunjenosti uvjeta 157, 161,
 165, 193, 207, 213, 243
 test korelacije, v. korelacija
 test prikladnosti modela 157, 160,
 167, 169
 test vjerodostojnosti 172

testiranje

 statističko testiranje 75
 testiranje hipoteze 1, 3, 9, 130, 132,
 177, 194, 217
 testiranje ispunjenosti uvjeta 211,
 213, 243
 testiranje nulte hipoteze 131, 138,
 245–246
 testiranje statističke značajnosti
 130–131, 239, 246
transformacija podataka 95–96, 106,
108–110, 160, 202
transparentnost 70, 209, 230, 242,
245–247
Tukeyev test 194–197, 201

U

učestalost (v. *i* frekvencija) 26, 35, 45,
49, 75, 77, 96, 167, 224
učinak (v. *i* efekt)
 učinak učenja 47–48, 51, 157, 181
 učinak vježbe 47
unakrsna tablica, v. tablica
unutarnja konzistentnost 71–72, 90
uopćavanje (v. *i* generalizacija) XIV, 4–5,
7, 9, 33, 39–40, 49, 52, 75, 129, 173,
240, 245
uopćeni linearni model 214
upitnik 35–36, 70, 73, 76, 233
uprosječivanje 49
uravnoteživanje 47–48
usjek (na dijagramu s pravokutnikom)
121–122, 149
uzorak
 jednostavni slučajni uzorak 30
 klasterski uzorak 30–31
 lančani uzorak 31, 229
 neslučajni uzorak 30–32
 nezavisni uzorci 45–48, 59, 79–80,
 142, 146, 151, 157, 175–179, 181–
 185, 187–188, 190–193, 197, 201,
 219, 231
 prigodni uzorak 30–31
 slučajni uzorak 29–35, 218, 240
 stratificirani uzorak 30–31
 sustavni slučajni uzorak 30
 veličina uzorka 29, 33–34, 37, 77,
 83–84, 106, 129, 134, 138, 143–144,
 146, 149–151, 152, 166, 179, 181,
 183, 185–187, 199, 209, 218, 233,
 235
 višelokacijski uzorak 32

volonterski uzorak 31
WEIRD uzorak 32
 zavisni uzorci 45–48, 59–60, 80–81, 142, 151, 157, 173, 175–179, 181–184, 186–187, 190–191, 195–197, 201, 206, 219, 231
 uzorkovanje XV, 28, 30, 32–34, 40, 229, 239
 neslučajno uzorkovanje 31, 41
 pogreška uzorkovanja, v. pogreška slučajno uzorkovanje 30–34, 47–48, 214
 uzorkovanje ispitanika 18, 28–29, 32, 37, 52, 176, 229, 240
 uzorkovanje podražaja 18, 35, 37, 39–40, 202
 uzorkovanje s povratom 110, 140, 151, 162, 218
 višestruko uzorkovanje 32
 uzročno-posljedični odnos 3, 19, 41, 51, 71, 164–165, 205

V

validnost, v. valjanost
 valjanost 32, 50–51, 70, 73, 187, 227, 240
 konstruktna valjanost 52
 unutarnja valjanost 51–52
 valjanost instrumenta 70
 valjanost statističkoga zaključka 52, 165
 vanjska valjanost 51–52, 240
 varijabilnost XV, 33–34, 37, 39, 49, 71, 76, 87–92, 96, 104–106, 112, 117–118, 121–122, 134–135, 140–143, 145, 149–150, 153, 157, 160, 178, 181–182, 190–191, 193, 195, 197, 202–203, 208–209, 213, 215, 219, 234, 237, 239, 242
 individualna varijabilnost 46
 mjera varijabilnosti, v. mjera varijabilnost između čestica/
 podražaja 59, 187–188, 203, 214
 varijabilnost između ispitanika 187, 196, 214
 varijabilnost između uzoraka 191–192
 varijabilnost unutar uzoraka 191–192, 195
 varijabla XV, 18–20, 22–24, 26–27, 29, 31, 36, 39, 41, 43–46, 53, 55, 59, 61, 76, 79–80, 82–83, 93, 107, 110, 115, 119, 125–126, 130–131, 148, 151, 154–155,

157, 159, 165–167, 175–176, 179, 181, 185–186, 188, 198–199, 203–210, 212–216, 219, 228–231, 233–234, 237, 239–240, 251
 binarna varijabla 24
 diskretna varijabla 23
 empirijska varijabla (v. *i* indikatorska varijabla) 23
 indikatorska varijabla (v. *i* empirijska varijabla) 23, 52
 kategorijska varijabla 23–24, 26, 56, 60, 77, 80, 86, 112–113, 115, 117, 130, 157–158, 163–164, 166–167, 175–176, 204, 213, 216
 kontinuirana varijabla 23, 130, 157, 163
 kontrolna varijabla 42
 kriterijska varijabla 19
 kvalitativna varijabla 23
 kvantitativna varijabla 23
 manipulativna varijabla 3, 41–42, 47, 49, 164
 nezavisna varijabla 19–20, 22, 31–33, 36–49, 51–52, 59, 76, 79, 81, 129, 132, 135, 139–141, 153, 155–158, 164, 167, 172–173, 175–177, 181–182, 186–187, 190–191, 193–198, 200, 204, 212–213, 215–216, 222, 226, 228–230
 nominalna varijabla, v. kategorijska varijabla
 numerička varijabla 23–26, 77, 117, 124, 157–158, 164, 175, 188, 204, 208, 213
 ometajuća varijabla 19, 30, 35, 40, 47, 49, 103, 177
 prediktorska varijabla 19
 razina varijable 36–37, 39, 45–49, 56, 59, 79, 81, 157, 164, 166, 173, 175, 177, 187, 191, 198, 200, 214, 226, 229–230
 selektivna varijabla 3, 41, 49, 164
 teorijska varijabla 23, 52
 vanjska varijabla 19, 47, 51–52
 zavisna varijabla 19–20, 22, 26, 41–42, 44, 49, 51–52, 56, 76, 81, 113, 129, 141, 153, 156, 158, 164–165, 173–175, 178–179, 183, 187, 190–191, 195–198, 202, 204, 212–214, 216, 222, 226, 228, 230
 varijanca 90, 92, 159–160, 165, 177–178, 183, 190, 192–193, 195–196, 198

analiza varijance, v. ANOVA
homogene (jednake) varijance 178–
179, 181–183, 193, 196, 218
nacrt varijance, v. nacrt istraživanja
nehomogene (različite) varijance
178–179, 181
varijanca između uzoraka 193, 195
varijanca unutar uzoraka 193
varijant 44
veličina efekta 139–142, 148–150, 172,
183–184, 187–188, 206, 235, 242,
246–247, 252
 mjera veličine efekta 140–142, 171,
 184, 188, 198, 209
 opća mjera veličine efekta 141
veličina učinka, v. veličina efekta
vjerojatnost 29, 34, 44, 98, 101, 130–
131, 133–138, 141, 148, 153–154, 169,
178, 181, 188–190, 201, 206, 216, 236,
246
 gustoća vjerojatnosti 98, 124
 iskaz o vjerojatnosti 133
 teorija vjerojatnosti 4, 130
vrijednost p 133–141, 148, 152–154,
172, 180, 184–185, 188–189, 195, 198,
211, 215, 218, 235, 246–247

W

WEIRD, v. uzorak
Welchov t -test, v. t -test
Wilcoxonov test 184–189, 201, 206
 Wilcoxonov test rangova s
 predznacima 186–187, 201, 203
 Wilcoxonov test zbroja rangova
 184–185, 199
winsoriziranje 108

Z

z-vrijednost 62, 95–96, 103
zadatak (jezički, eksperimentalni) 1–3,
21–23, 27, 31, 35–37, 40, 42, 46, 48,
52–54, 56, 58, 61–62, 70–71, 77–78,
87–91, 95–97, 105–109, 113, 115, 117,
125, 131–132, 135, 147, 177, 187–188,
196–198, 203, 206, 211, 216, 223, 230–
233, 241, 243
značajnost
 praktična značajnost 140, 172, 188,
 238
 statistička značajnost XV, 118, 121,
 130–131, 133, 136–140, 142, 144,
 147–154, 169, 172, 180, 185–186,

188–189, 193–195, 200, 208–209,
211, 235–236, 239, 246, 251–252
znanstveni funkcionalni stil 224

Kazalo autora

A

- Abbuhl, R. 42, 47, 240, 255
Afonso, A. 201, 270
Ahn, H. 106, 109, 228, 236, 255
Al-Hoorie, A. H. 12, 30, 70–71, 87, 142, 148, 160, 162, 189, 230, 232, 255, 262
Alacaci, C. 14, 255
Altman, D. G. 148, 261
Altman, N. 145–146, 264
American Psychological Association 12, 224, 255
Amoroso, L. W. 190, 255
Andel, M. 87, 274
Andringa, S. XII, 32, 80–81, 173–174, 255
Artstein, R. 72, 255
Arunachalam, S. 14, 32, 255
Arvizu, M. N. G. 12, 138, 244, 262, 271

B

- Baayen, R. H. 14, 42, 44, 70, 111, 214–215, 255–256
Bagarić Medve, V. 248–249, 256
Ballard, L. 12, 13, 266
Bard, E. G. 57, 256
Barr, D. J. 40, 204, 215, 256
Bates, D. M. 214, 256
Béland, S. 62, 271
Belia, S. 146, 256
Bennett, C. 141, 256
Berger, C. 88–90, 143–144, 156, 203, 256
Bertenshaw, N. 105, 223, 229, 259
Biedroń, A. 183, 210, 215, 256
Blom, E. 52, 256
Boers, F. 142, 242–243, 257
Brehm, L. 54, 262
Brown, A. 178, 257
Brown, D. 12, 138, 243–244, 262, 271
Brown, J. D. 13–14, 72, 129, 153, 208, 250, 257
Bryfonski, L. 142, 242–243, 257
Buchner, A. 150, 259
Buchstaller, I. 34, 151, 257
Budimirović, J. 14, 268
Bulté, B. 211, 257
Buus, S. 96, 267
Byrnes, H. 246, 257

C

- Carlin, J. B. 148, 261
Chang, A. 2, 226, 275
Chaudron, C. 2, 237, 257
Chen, M. 12, 138, 244, 262, 271
Cho, J. 241, 257
Clark, H. H. 39, 257
Cohen, J. 34, 257
Cooc, N. 242, 276
Cook, I. 75–76, 130, 275
Covey, L. 54, 257
Crossley, S. 88–90, 143–144, 156, 203, 256
Crowther, D. 12–13, 243, 257, 266, 272
Cumming, G. 118, 130, 145–148, 243, 246, 256, 258
Cummings, J. J. 39, 258
Cunningham-Andersson, U. 85, 92, 258
Cunnings, I. 38, 147, 214, 258, 265

Č

- Čolović, P. 70–71, 73, 258

D

- Dallas, A. 95, 258
Davidson, D. J. 214, 256
Day, R. A. 222–223, 225, 232, 236, 258, 260
De Miranda, M. 217, 269–270
DeBruine, L. 34, 258
DeDe, G. 95, 258
Dehghan-Chaleshtori, M. 243, 270
DeKeyser, R. M. XI–XII, 125–126, 206–207, 211, 213, 258
Deniz, N. D. 105, 258
Derrick, D. J. 12, 70–72, 91, 138, 230, 243–244, 258, 262, 272
Díaz-Negrillo, A. 77, 258
Diependaele, K. 91, 105, 259
Dinno, A. 201, 259
Do, H. T. 2, 243, 269
Dörnyei, Z. 14, 259
Dragičević, Č. 96, 259
Duffield, N. 107, 259
Dumančić, F. 87, 274
Duñabeitia, J. A. 91, 105, 259
Dusseldorp, E. 44, 273

Đ

- Đukanović, M. 14, 268

E

Ediger, A. 12, 265
Edmonds, A. 54, 261
Egbert, J. 110, 159, 202, 218, 259, 264, 272
Eger, N. A. 21, 94, 228, 268
Ellis, N. C. 244, 266
Erdfelder, E. 150, 259
Eyckmans, J. 151, 240, 243, 265

F

Faez, F. 142, 242–243, 257
Fajgelj, S. 7, 22, 24–25, 27, 33–35, 61, 71, 259
Fanelli, D. 241, 259
Faul, F. 150, 259
Felser, C. 38, 105, 223, 229, 259
Feng, S. 241, 260
Ferraresi, A. 14, 268
Fidell, L. S. 73, 109, 274
Fidler, F. 146, 256
Field, A. XI, 10, 14, 17–18, 42, 44, 56, 69, 92, 102, 105, 110, 132, 155, 159, 165–166, 187, 199, 215, 222–227, 232–238, 245, 260
Field, Z. 69, 132, 155, 165–166, 187, 215, 260
Fisher, R. 247, 260
Fletcher, P. 81, 276
Florentine, M. 96, 267
Foote, R. 78, 82, 268
Fox, J. 214, 260
Fujita, H. 38, 147, 258

G

Gabriel, U. 30, 260
Gabriele, A. 54, 257
Gačić, M. 225, 260
Garcia, G. D. 14, 67, 70, 127, 260
Gardner, R. C. 36, 260
Gass, S. M. 11–14, 42, 47–48, 52, 54, 76, 136, 138, 140, 152, 162–163, 165, 240, 243, 246–247, 249, 260, 266, 272
Gastel, B. 222–223, 225, 232, 236, 258, 260
Gavarró, A. 186, 189, 228, 266
Ghanbar, H. 105, 215, 243, 270, 272
Ghanem, R. 12, 138, 244, 262, 271
Gledić, J. 14, 268
Godfroid, A. 32, 255
Gonulal, T. (Gönülal, T.) 11–13, 260–261, 266

Goodman, S. N. 148, 261
Grabowski, K. C. 70, 261
Greenland, S. 148, 261
Grice, M. 46, 173, 215, 276
Gries, S. T. 14, 44, 61, 70, 127, 131, 155, 158, 161, 165, 261
Guan, Q. 242, 275
Gudmestad, A. 54, 261
Gygax, P. 30, 260
Gyllstad, H. 31, 261

H

Han, C. 13, 276
Hatch, E. 161, 261
Heil, J. 20–22, 32, 36, 131–132, 228–229, 231, 236, 240, 244, 261, 268
Hellman, A. B. 42, 261
Herrington, R. 83, 107, 110, 112, 117–118, 121, 126, 218, 243, 265
Hilpert, M. 217, 271
Hiver, P. 30, 262
Hoekstra, R. 148, 268
Hole, G. 222–227, 232–238, 260
Holliday, J. J. 42, 262
Howard, A. L. 12, 262
Hu, Y. 161, 162, 193, 213, 232, 243, 262
Huang, X. 242, 275
Hughes, A. 81, 276
Hulstijn, J. H. 108, 262

I

In'nami, Y. 242, 262
Ioannidis, J. P. A. 245, 262
Isbell, D. R. 12–13, 138, 244, 262, 266, 271
Issa II, B. I. 32, 240, 244, 268
Ivanec, D. XIV, 4, 14, 19, 29–30, 33, 56, 62, 70, 76, 83, 84, 86, 88–89, 91, 96, 98–99, 102, 121, 136–138, 140, 143–144, 149–150, 164–165, 179, 206, 271

J

Jackson, C. N. 54, 262
Jaeger, T. F. 214, 262
Jaensch, C. 201, 262
Jegerski, J. 35, 54, 263
Jensen, I. N. 86, 132, 263
Jiang, N. XI, 120, 187, 189, 203–204, 263, 269
Johnson, D. E. 14, 75–76, 111, 127, 158, 263
Johnson, M. R. 35, 266

Jolani, S. 62, 271

K

Kaiser, E. X. 42, 96–97, 99–100, 104, 275

Källkvist, M. 31, 261

Keating, G. D. 35, 263

Keuleers, E. 91, 105, 259

Khany, R. 162–163, 263

Khattab, G. 34, 151, 257

Kim, L. K. 196, 263

Kim, S. 229, 263

Kim, Y. 19, 85, 91, 229, 263

Ko, H. 229, 263

Koizumi, R. 242, 262

Kolesarić, V. XIV, 4, 14, 19, 29–30, 33, 56, 62, 70, 76, 83–84, 86, 88–89, 91, 96, 98–99, 102, 121, 136–138, 140, 143–144, 149–150, 164–165, 179, 206, 271

Kraš, T. X, XII–XIII, 1, 3, 9, 18–19, 23–24, 27–32, 35, 38, 41, 50–53, 56–60, 62, 70, 73, 75–76, 99, 114–115, 129, 151, 157, 177, 195, 202–203, 214, 216, 228–230, 240, 248, 263–264, 267

Krzywinski, M. 145–146, 264

Kytö, M. 12, 232–233, 236, 276

L

Lachaud, C. M. 61–62, 107–108, 110, 264

LaFlair, G. T. 110, 159, 202, 218, 259, 264, 272

Lakshmanan, U. 196, 263

Lang, A.-G. 150, 259

Larsen, W. A. 122, 267

Larson-Hall, J. 14, 33, 67–68, 70, 73, 75–76, 83, 92, 100, 104, 107–112, 114, 117–121, 123–124, 126–127, 133, 136, 138, 141–142, 146–149, 152, 159, 161, 183, 189, 218, 233, 235–236, 243, 264–265

Larsson, T. 12, 232–233, 236, 276

Lavolette, E. 12, 66–67, 266

Lazaraton, A. 11–12, 161, 163, 261, 265

Leal, T. 27, 45, 265

Lee, M. D. 148, 268

Leow, R. P. 32, 240, 244, 268

Leśniewska, J. 62, 271

Levshina, N. 63, 70, 93, 96, 98, 103–104, 127, 145, 147–148, 164–165, 265

Levy, R. 204, 215, 256

Lim, J. 12–13, 266

Lim, P. C. 35, 266

Linck, J. A. 214, 265

Lindstromberg, S. 56, 105–106, 127, 138, 141, 147–148, 151, 159, 161–162, 177, 184, 218, 221, 233, 240, 243, 265

Lintz, E. N. 35, 266

Liszka, S. A. 36–38, 40, 42–43, 45, 48, 175–176, 198, 203, 230, 273

Loerts, H. 14, 70, 266

Loewen, S. 11–14, 66–67, 163, 165, 247, 249, 257, 260–261, 266

López, L. 20–22, 36, 131–132, 228–229, 231, 236, 261

Lowie, W. 14, 70, 266

Lowry, R. 65, 266

Lundquist, B. 86, 132, 263

M

Macaro, E. 242, 266

Mackey, A. 48, 52, 54, 76, 130, 136, 152, 217, 241, 255, 260, 266

Mackey, B. 130, 217, 266, 273

Majdoub, H. 242, 275

Maloney, J. 12–13, 257, 266

Margaza, P. 186, 189, 228, 266

Marsden, E. 32, 54, 58, 63, 105, 240, 243–244, 246, 266–268, 272

Martín Gómez, A. 216, 267

Marty, P. 14, 268

Masuda, K. XI, 120, 263

Mayo, L. H. 96, 267

McClelland, G. H. 151, 267

McDonough, K. 19, 54, 58, 85, 91, 106–108, 131, 201, 203, 229, 263, 267

McGill, R. 122, 267

McKay, T. 142, 242, 243, 257

McLean, S. 31, 48, 275

McManus, K. 241, 272

Medeiros, F. M. 201, 270

Mihaljević Djigunović, J. 36, 260

Mikhaylova, A. 32, 240, 244, 268

Mikołajczak, S. 32, 240, 244, 268

Milas, G. 14, 17–20, 31, 37, 43, 47, 51–52, 56, 131, 267

Miles, J. 69, 132, 155, 165–166, 187, 215, 260

Miličević, M. X–XIV, 1, 3, 9, 14, 18–19, 23–24, 27–32, 35, 38, 41, 50–53, 56–60, 70, 73, 75–76, 99, 113–115, 117–119, 122–124, 127, 129, 151, 157, 177–178, 182–183, 185, 188, 195, 201–203, 214, 216, 228–230, 240, 248, 264, 267

Miličević Petrović, M. 14, 50, 268, 272

Milin, P. 70, 71, 73, 111, 256, 258
Miller, J. 38, 268
Miller, K. 38, 273
Miller, M. 141, 256
Mitterer, H. 21, 94, 228, 268
Mizumoto, A. 67–68, 72, 150, 265, 268
Montrul, S. 78, 82, 87, 268
Moranski, K. 31–32, 268
Moreno, N. 32, 240, 244, 268
Morey, R. D. 148, 268
Morgan-Short, K. 32, 240, 244, 266, 268
Mormer, E. 54, 262
Morris, J. 91, 105, 259
Mueller, J. 187, 189, 203, 269

N

Nahatame, S. 71, 269
Nassaji, H. 93, 269
Navarro, D. 70, 269
Newton, J. 2, 226, 275
Nguyen, A. T. 2, 243, 269
Nguyen, M. T. T. 2, 243, 269
Nicklin, C. 31, 48, 105–106, 108–110,
151, 269, 275
Nicol, J. 95, 258
Norouzian, R. 34, 72, 150–151, 198, 206,
217, 243, 269–270
Norris, J. M. 73, 142, 149, 189, 238, 270
Novokshanova, E. XI, 120, 263

O

Oh, S. 70, 261
Ortega, L. 71, 238, 270, 276
Oswald, F. L. 141–142, 183, 190, 199,
242–243, 272

Ö

Öztürk, M. S. 182, 270

P

Pallant, J. 69, 270
Paolillo, J. C. 238, 270
Papi, M. 12, 66–67, 266
Park, H. I. 243, 270
Pavičić Takač, V. 249, 256
Pavlović, N. XIV, 13, 19–20, 22–23, 222–
224, 233, 237, 270
Pereira, D. G. 201, 270
Perek, F. 217, 271
Perpiñán, S. 78, 82, 268
Petz, B. XIV, XV, 4, 14, 19, 29–30, 33,
56, 62, 70, 76, 83–84, 86, 88–89, 91,
96, 98–99, 102, 121, 136–138, 140,
143–144, 149–150, 164–165, 179,

206, 271

Phakiti, A. 8, 12, 14, 19, 42, 70, 92, 205,
226, 271, 273
Pham, T. T. 2, 243, 269
Phillips, N. D. 124, 271
Pichette, F. 62, 271
Plonsky, L. 11–12, 32, 42, 54, 58, 63,
67, 70–73, 75–76, 84, 91, 105–106,
108–111, 118–120, 123, 136, 138,
140–142, 147–149, 151, 161–163,
183, 189–190, 193, 198–199, 206,
213, 215, 217–218, 230, 232–233,
235–236, 242–244, 246, 260–262,
264–265, 267–272, 276
Podboj, M. 1, 52–53, 264
Poesio, M. 72, 255
Poole, C. 148, 261
Popper, K. 20, 272
Porte, G. K. 35, 137, 222–223, 226–228,
235, 237, 240–241, 272
Puskas, G. 50, 272

Q

Quené, H. 34, 273

R

Rasinger, S. M. 14, 20, 22, 165, 273
Rast, R. 242, 275
Reeves, B. 39, 258
Reinisch, E. 21, 94, 228, 268
Renaud, O. 61–62, 107–108, 110, 264
Richardson, J. T. E. 135, 273
Riggenbach, H. 12, 265
Roberts, L. 36–38, 40, 42–43, 45, 48,
175–176, 198, 203, 230, 273
Robertson, D. 57, 256
Robison, R. E. X, XII, 79, 115–117, 166,
168–171, 173, 176, 273
Roettger, T. B. 241, 244, 273
Roever, C. 12, 14, 42, 70, 273
Roothoof, H. 211, 257
Ross, S. J. 73, 130, 149, 189, 217, 266,
270, 273
Rothman, K. J. 148, 261
Rouder, J. N. 148, 268

S

Samardžić, T. 50, 272
Sandlund, E. 31, 261
Sato, M. 105, 223, 229, 259
Scheepers, C. 204, 215, 256
Schmid, M. S. 44, 273
Schmidtke, J. 12, 66–67, 266
Schmitt, C. 38, 273

Schnur, E. 12, 138, 244, 262
Schoonen, R. 73, 108, 149, 189, 262, 270
Schütze, C. T. 55, 273
Senn, S. J. 148, 261
Seton, B. 14, 70, 266
Sheng, L. 242, 276
Simonović, M. 1, 52–53, 264
Skalicky, S. 88–90, 143–144, 156, 203, 256
Slabakova, R. 32, 86, 132, 240, 244, 263, 268
Smaldino, P. 248, 274
Smederevac, S. 137–142, 144, 183, 199, 274
Solon, M. 243, 270
Sönning, L. XVI, 138, 241, 246–249, 274
Sorace, A. 57, 256
Spinner, P. 243, 272
Spino, L. A. 12, 66–67, 266
Sprouse, J. 55, 273
Stanojević, M.-M. XIV, 13, 19–20, 22–23, 222–224, 233, 237, 270
Stanton, J. M. 212, 274
Sterling, S. 12, 66–67, 232–233, 236, 266, 276
Stevens, S. S. XII, 24, 26, 274
Sunderland, J. 19–20, 274
Sundqvist, P. 31, 261
Szudarski, P. 32, 240, 244, 268

T
Tabachnick, B. G. 73, 109, 274
Tavakoli, H. 132, 274
Tazik, K. 70, 162, 163, 263, 274
Tenjović, L. 137–142, 144, 183, 199, 274
Testa, M. 215, 274
Thompson, S. 54, 58, 105, 243, 267
Tigchelaar, M. 12–13, 257, 266
Tily, H. J. 204, 215, 256
Todorović, D. XV, 3, 14, 19, 23, 30–31, 33–34, 39, 41–44, 47–49, 56, 58, 76, 79, 82, 97, 131, 137, 164–166, 274
Tomita, Y. 242, 262
Tonković, M. 87, 274
Tošković, O. XV, 3, 5, 14, 19, 21, 23–30, 43–44, 53, 76, 131, 274
Tremblay, A. 243, 274
Trévisiol, P. 242, 275
Trochim, W. M. K. 182, 275
Trofimovich, P. 54, 58, 106–108, 131, 201, 203, 244, 266, 267
Tufte, E. R. 111, 118, 124, 275

Tukey, J. W. 120, 122, 267

U

Unsworth, S. 52, 256
Upton, G. 75, 76, 130, 275

V

Valera, S. 77, 258
Van Gelderen, A. 108, 262
VanPatten, B. 54, 263
Vasishth, S. 217, 275
Vitta, J. P. 12, 31, 48, 70–71, 87, 142, 148, 151, 160, 162, 189, 230, 232, 255, 269, 275

W

Wagenmakers, E.-J. 148, 268
Wallis, S. 3, 42, 275
Wang, F. H. X, 42, 96–97, 99, 100, 104, 275
Wang, X. XI, 120, 263
Warne, R. T. 20, 275
Watorek, M. 242, 275
Webb, S. 2, 226, 275
Werner, V. XVI, 138, 241, 246–249, 274
Westergaard, M. 86, 132, 263
White, L. 107, 259
Wilcox, R. 102, 110, 159, 260
Williams, J. 146, 256
Winter, B. 46, 70, 127, 173, 215, 249, 275–276
Woldorf, G. 141, 256
Wolff, D. 12, 66–67, 266
Wood, M. 12, 232–233, 236, 276
Woodrow, L. 221, 223, 230–231, 233–234, 276
Woods, A. 81, 276
Wu, S. 71, 276

Y

Yang, H.-K. 229, 263
Yang, M. 242, 276
Yaw, K. 12, 232–233, 236, 276
Yorganci, M. 182, 270
Yu, X. C. 242, 275

Z

Zhang, M. 12, 138, 244, 262
Zhang, P. 13, 276
Ziegler, N. 31–32, 268

O autoricama

Maja Miličević Petrović izvanredna je profesorica na Odsjeku za usmeno i pisano prevođenje Sveučilišta u Bologni u Italiji. Do 2020. radila je na Odsjeku za opću lingvistiku Filološkoga fakulteta Sveučilišta u Beogradu. Diplomirala je talijanski jezik i književnost na Sveučilištu u Beogradu, a magisterij iz engleskoga jezika i primijenjene lingvistike te doktorat iz usvajanja drugoga jezika stekla je u *Research Centre for English and Applied Linguistics* na Sveučilištu u Cambridgeu u Ujedinjenome Kraljevstvu. U svojem se znanstvenoistraživačkom radu bavi morfosintaksom, usvajanjem drugoga jezika, lingvističkim značajkama prijevoda i korpusnom lingvistikom, usredotočujući se na srpski, talijanski i engleski jezik. Suautorica je znanstvene monografije *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika* (Filozofski fakultet Sveučilišta u Rijeci, 2015). Članica je uredničkoga odbora časopisa *Italica Belgradensia* (Filološki fakultet Sveučilišta u Beogradu) i časopisa *Strani jezici* (Hrvatsko filološko društvo). Bila je među idejnim tvorcima strateškoga partnerstva Erasmus+ Europske unije *Upgrading the SKills of Linguistics and Language Students (UPSKILLS)*, na kojemu je sudjelovala u definiranju novih ciljnih profila za studente jezika i lingvistike te u izradi nastavnih materijala. Bila je voditeljica srpskoga tima projekta *(Dis-)entangling Traditions on the Central Balkans: Performance and Perception – TraCeBa* u potprogramu ERA.Net RUS Plus u okviru programa FP7/Obzor 2020 Europske unije, kao i srpske strane institucijskoga projekta *Regional Linguistic Data Initiative (ReLDI)*, posvećenoga unapređenju metodologije i resursa za istraživanje srpskoga jezika i hrvatskoga jezika, koji je financirala Švicarska nacionalna zaklada za znanost. Drugi je projekt u Beogradu prerastao u udruženje građana *ReLDI centar za jezičke podatke*, čija je predsjednica. Trenutačno sudjeluje u jednome nacionalnom projektu u Italiji (*UNITE – Universally Inclusive Technologies to Practice English*) i jednome u Srbiji (*VLingS – Vulnerable Languages and Linguistic Varieties in Serbia*). Drži nastavu iz područja opće lingvistike i usvajanja drugoga jezika te kvantitativne obrade jezičnih podataka. Autorica je ili suautorica sedam *online*-kolegija (četiri na srpskome i tri na engleskome jeziku) posvećenih različitim područjima metodologije istraživanja u lingvistici (<https://elearn.mnf.uzh.ch>, <https://upskills.fil.bg.ac.rs>).

Tihana Kraš izvanredna je profesorica na Odsjeku za anglistiku Filozofskoga fakulteta Sveučilišta u Rijeci. Diplomirala je engleski jezik i književnost te talijanski jezik i književnost na Filozofskome fakultetu Sveučilišta u Zagrebu, gdje je završila i poslijediplomski stručni prevoditeljski studij (smjer: engleski jezik). Magisterij iz engleskoga jezika i primijenjene lingvistike te doktorat iz usvajanja drugoga jezika stekla je u *Research Centre for English and Applied Linguistics* na Sveučilištu u Cambridgeu u Ujedinjenome Kraljevstvu. Na poslijedoktorske usavršavanju na Sveučilištu u Edinburghu u Ujedinjenome Kraljevstvu vodila je znanstvenoistraživački projekt Hrvatske zaklade za znanost *Procesiranje jezika kod dvojezičnih govornika*. U svojemu se znanstvenoistraživačkome radu bavi usvajanjem prvoga i drugoga jezika, dvojezičnošću i psiholingvistikom, usredotočujući se na gramatičke pojave u talijanskome, hrvatskom i engleskom jeziku. Suautorica je znanstvene monografije *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika* (Filozofski fakultet Sveučilišta u Rijeci, 2015). Suurednica je serije znanstvenih knjiga *Studies in Bilingualism* (John Benjamins) te članica uredničkoga odbora časopisa *Linguistic Approaches to Bilingualism* (John Benjamins) i časopisa *Strani jezici* (Hrvatsko filološko društvo). Vodila je tim s Filozofskoga fakulteta u Rijeci na znanstvenoistraživačkome projektu FP7 Europske unije *Advancing the European Multilingual Experience (AThEME)*, *Projekt razvoja karijera mladih istraživača – izobrazba novih doktora znanosti* Hrvatske zaklade za znanost, znanstvenoistraživačke projekte Sveučilišta u Rijeci *Procesiranje zamjenica u talijanskome jeziku* i *Obrada rečenica u talijanskom jeziku* te tim s Filozofskoga fakulteta u Rijeci u dvama strateškim partnerstvima Erasmus+ Europske unije – *Upgrading the SKills of Linguistics and Language Students (UPSKILLS)* i *Teacher Education About Multilingualism (TEAM)*. Na Filozofskome fakultetu u Rijeci vodi ogranak neprofitne organizacije *Bilingualism Matters* sa sjedištem u Edinburghu u Ujedinjenome Kraljevstvu, pod nazivom *Bilingualism Matters@Rijeka*, koji se bavi popularizacijom znanstvenih spoznaja o višejezičnosti. Za vođenje ogranka dobila je 2019. nagradu Filozofskoga fakulteta u Rijeci za popularizaciju znanosti u području humanističkih znanosti. Na istoj ustanovi drži nastavu na diplomskoj razini studija iz usvajanja drugoga jezika te jednojezičnoga i dvojezičnoga usvajanja prvoga jezika.



9 789533 611211