Izdavač
»NAKLADA SLAP«
10450 Jastrebarsko, dr. F. Tuđmana 33
nslap@nakladaslap.com
www.nakladaslap.com

Direktor
Biserka Matešić

Glavni urednik
dr. sc. Krunoslav Matešić

Lektor
Žarko Taraš

Recenzenti
doc. dr. sc. Mladen Havelka
prof. dr. sc. Vladimir Kolesarić
prof. dr. sc. Alija Kulenović

Izdavanje ove knjige pripomoglo je
Ministarstvo znanosti i tehnologije Republike Hrvatske.

Objavljivanje ovog udžbenika odobrio je Senat Sveučilišta u Zagrebu
rješenjem broj 02-1994/1-1996 od 25. veljače 1997.

CIP - Katalogizacija u publikaciji
Nacionalna i sveučilišna knjižnica, Zagreb

UDK 311(075.8)

PETZ, Boris
Osnovne statističke metode za nematematičare / Boris Petz. - 6. izd. - Jastrebarsko : Naklada Slap, 2007. - (Udžbenici sveučilišta u Zagrebu=Manualia Universitatis studiorum Zagrabiensis)
Bibliografija.
ISBN 978-953-191-058-3
I. Statistika -- Udžbenik
301109101

ISBN 978-953-191-058-3

prof. dr. sc. Boris Petz

OSNOVNE STATISTIČKE METODE ZA NEMATEMATIČARE

VI. izdanje

NAKLADA SLAP

PREDGOVOR TREĆEM IZDANJU

Ovo. treće, izdanje udžbenika "Osnovne statističke metode za nematematičare" u nekoliko se pojedinosti razlikuje od oba dosadašnja izdanja. Od važnijih promjena i dopuna valja spomenuti nešto proširenu diskusiju o nastanku normalne raspodjele, nadalje kratko poglavlje o Bayesovim principima u vezi s vjerojatnosti (a Bayes je tek vrlo kasno otkriveni autor, koji je dosta uzbudio modernu statističku publiku), prikazani su još neki postupci izračunavanja korelacija i prošireno je poglavlje o analizi varijance: dodani su postupci za većinu mogućih situacija, u kojima se analiza varijance koristi. Budući da analiza varijance spada među najomraženija poglavlja za nematematičare, ti dodatni dijelovi dani su u obliku "recepata korak po korak", pa s njima ne bi smjelo biti posebnih teškoća.

No, neizbježno je došlo i vrijeme da se nešto iz prethodnih izdanja ispusti: od prvog izdanja ove knjige prošlo je već šesnaest godina, i kroz to vrijeme došlo je upravo u području primjene osnovnih računskih postupaka do goleme promjene; dok smo početkom osamdesetih godina s oduševljenjem nabavljali prva džepna računala, koja su mogla pritiskom na tipku izračunati drugi korijen nekog broja, danas praktički i najjednostavnije džepno računalo ima mogućnosti koje daleko premašuju te zadatke. A budući da gotovo i nije više moguće naći studenta ili intelektualca bilo koje struke koji ne bi posjedovao barem jedno jednostavno džepno računalo, to je dodatak tablica s korijenima (a one su u ovom udžbeniku zauzimale preko 40 stranica!) postao potpuno suvišan, i te su tablice iz rukopisa izostavljene. (Uostalom, upute za njihovu upotrebu za neiskusnog su čitatelja bile ponešto komplicirane, pa je korist od njih bila možda i manja od očekivane).

U Zagrebu, na Novu godinu 1997.

B. Petz

SADRŽAJ

1. ZAŠTO STATISTIKA?

"Statistički način mišljenja jednog će dana za svakodnevni život građana postati jednako neophodan kao znanje čitanja i pisanja."

H. G. Wells (1866-1946)

Ako se upitamo "što je statistika", može nam pomoći misao jednog statističara na području medicine (Mainland), koji kaže: "Nijedna definicija ne znači mnogo tako dugo dok nismo proučili ono na čemu radimo — a tada je svaka definicija gotovo nepotrebna." A budući da ima više od 100 definicija statistike, zaista je bolje da se i ne pokušavamo upuštati u to da je definiramo. Uostalom, i bez definicije, manje-više ipak svi znamo što po prilici ona znači: obrada brojčanih podataka radi jasnijeg prikazivanja. (S ovim "jasnije" mnogi se od vas zasigurno ne slažu!)

Kao što će se kasnije iz teksta ove knjige vidjeti, počeci statističke metodologije nastali su iz potpuno praktičnih poticaja i problema svakodnevnog života (tj. iz problema na koje su nailazili profesionalni kockari i hazarderi!). No kada su se matematičari ozbiljno zainteresirali za te probleme, i kada su razradili statistiku kao granu primijenjene matematike, statistika je postala teško dostupna čovjeku, koji nije bio ili po profesiji ili po svom "hobiju" matematičar. Možda je to uzrok činjenice da se u različitim društvenim i prirodnim znanostima statistika još prije 70-80 godina tako malo primjenjivala. (Jedino jedan njezin dio — a taj je upravo i najmanje povezan sa statističkom metodologijom — unatrag nekoliko stoljeća poznat je u znanosti i u javnosti: to su bili različiti statistički pregledi broja rođenih i umrlih, broja oboljelih od neke bolesti i sl.)

Tako se, na primjer, u primijenjenoj psihologiji statistika prvi put pojavljuje oko 1920. godine, kada su postavljene neke hipoteze o nastanku nesreća, pa je predloženo da se te hipoteze provjere tako što bi se uspoređivalo ono što se pod vidom tih hipoteza može očekivati s onim što se zapravo događa.

No, ti prvi pokušaji "prodora" statistike u svakodnevnu praksu i u primijenjenu znanost, ostali su u početku bez većeg odjeka, i tek za Drugoga svjetskog rata, usporedno s "eksplozivnim" porastom znanstvenih spoznaja na različitim prirodnim i društvenim područjima naglo prodire i statistički način mišljenja, kako u primjenu znanstvenih disciplina, tako i u način razmišljanja u problemima biologije, ekonomije, poljoprivrede, psihologije, sociologije, tehnike, vojnih znanosti, prometa, trgovine i prodaje, športa — pa čak i lingvistike, književnosti i drugih područja znanosti.

Statistička metodologija postala je u suvremenom životu donekle čak i dio "općeg obrazovanja" i "opće kulture", jer je, na primjer, teško zamisliti danas čovjeka *bilo koje struke*, ako posjeduje više, a pogotovu visoko obrazovanje, da mu ne bi bili poznati, primjerice, pojmovi "aritmetičke sredine", "korelacije", "varijabiliteta" i tome slično.

Dakle, svaki je čovjek danas — kao što je to proročanski predvidio H. G. Wells, kojega smo citirali kao moto ovog poglavlja — izložen situacijama u kojima mu je potrebno poznavanje nekih osnovnih statističkih pojmova, a pogotovu statističkog "načina mišljenja".

Za suvremenog čovjeka, koji se bavi *znanstvenim* radom, postoje — mogli bismo reći — četiri "razine", na kojima on treba statistiku:

1. Poznavanje statistike potrebno je zbog *praćenja stručne i znanstvene literature*.

2. Poznavanje statistike potrebno je pri obradi rezultata, prikupljenih istraživanjem ili eksperimentom, radi *deskripcije i analize* tih rezultata.

3. Poznavanje statistike potrebno je u znanstvenom radu radi *zaključivanja* iz konkretnog slučaja na "opći zakon".

4. Poznavanje statistike potrebno je pri *planiranju* istraživanja i eksperimenata.

Razmotrimo neke primjere za svaku od navedenih "razina"!

Ad 1. *Praćenje literature.* Držimo da za ovo područje uopće nije potrebno dokazivati potrebu osnovnoga statističkog obrazovanja, jer je malone svaki stručnjak često bio u takvoj situaciji da nije mogao razumjeti neke osnovne rezultate istraživanja iz svoje struke, objavljene u različitim stručnim i znanstvenim časopisima: ti su rezultati sve češće izneseni u vrlo skraćenom obliku, ali zato uz pomoć niza statističkih termina i simbola.

Ad 2. *Statistika radi deskripcije i analize.* Područjem opisivanja konkretnih rezultata, dobivenih prilikom nekog ispitivanja ili mjerenja, bavi se tzv. "deskriptivna" statistika. Njezina je zadaća — što se dade zaključiti iz naziva — da opiše podatke, i to na taj način da ih *sredi i sažme*, kako bi bili što pregledniji. Bez takvog sređivanja mnogi podaci bili bi nepregledni, pa se stoga tako često i događa da početnik s velikim elanom i marljivošću skupi mnoštvo podataka prilikom nekog istraživanja, a nakon toga ne zna pravo što da s njima počne — te oni često ostaju u ladici!

Tako, na primjer, antropologa ili liječnika može zanimati visina djece određenog spola i dobi, ali kad izvrši mjerenja nekoliko stotina ili tisuća djece, iz mase postojećih podataka ne može se vidjeti praktički ništa tako dugo dok se ti podaci ne srede i "sažmu" kako bi ih se moglo grafički prikazati, ili dok se ne izračuna barem aritmetička sredina, kao jedna od vrijednosti koja "prezentira" te rezultate, tj. koja nam jednim brojem daje ono što je izvršenim mjerenjem za tu skupinu "najtipičnije".

Ili, jedan kliničar može biti zainteresiran za varijabilitet inkubacije za neku određenu bolest. *Prosječno* trajanje inkubacije možda mu je poznato, ali njega možda upravo zanima s koliko velikim *varijacijama* u inkubaciji može računati. Drugim riječima, može li kod neke bolesti, kod koje inkubacija u prosjeku traje, recimo, oko 10 dana, biti siguran da do oboljenja neće doći, ako je od momenta moguće zaraze prošlo mjesec dana. (Tako je poznato da za tetanus inkubacija može trajati vrlo dugo.) Kliničar će osnovnu informaciju te vrste dobiti iz jedne druge "reprezentativne" vrijednosti, npr. iz tzv. "standardne devijacije", koja pokazuje u kojem se rasponu mogu očekivati rezultati.

Psihologa ili fiziologa može zanimati da li pod utjecajem djelovanja alkohola dolazi kod čovjeka do povećanja *varijabiliteta* vremena reagiranja, pri čemu možda *prosječno* vrijeme reagiranja ostaje nepromijenjeno. I te podatke dobit će, naravno, na isti način, tj. tražeći neku mjeru varijabiliteta rezultata.

Vrlo se često u praksi pojavi pitanje relativnog položaja pojedinog rezultata među svim ostalim rezultatima. Na primjer, ako poliklinički liječnik saopći nekom roditelju da je njegovo dijete "vrlo visoko za svoje godine", to je doduše tome roditelju korisna informacija, ali nije dovoljno precizna, jer roditelja zanima točan podatak. No kada bi mu taj liječnik rekao da je njegovo dijete toliko visoko da među drugom školskom djecom istih godina i spola ima samo 2% djece koja su rastom viša, onda ta informacija predstavlja već vrlo detaljan opis položaja tog djeteta među drugom djecom. Ili, prilikom klasifikacijskih ispita za upis na neki fakultet primjenjuje se možda nekoliko testova znanja. Nekako je uobičajeno da se rezultati svih testova zbroje i suma bodova svih testova predstavlja konačan rezultat tog studenta. Je li to posve ispravno? Saznat ćemo da nije. Naime, ako netko, pretpostavimo, dobije 70 bodova u testu, u kojem ispitanici po rezultatima variraju od 10 do 110, taj mu dobiveni rezultat znači manje nego kada bi rezultati varirali od 40 do 80. U ovom, kao i u prijašnjem primjeru, potrebno je, dakle, točno odrediti položaj svakog rezultata među ostalim rezultatima, tj. potrebno je znati "koliko vrijedi pojedini rezultat" — a to se vrlo lako može postići primjenom nekih jednostavnih statističkih postupaka (izračunavanje "z-vrijednosti" ili izračunavanje "decila" i sl.).

Dalje, mnogo se puta u praksi pojavi pitanje kakav je *oblik* distribucije neke pojave koja nas zanima, tj. kako se ta pojava "raspoređuje" u prirodi. Kao što je poznato najčešća distribucija, s kojom se u životu susrećemo, jest tzv. "normalna distribucija", u kojoj nalazimo najviše srednjih, a najmanje ekstremno malih ili ekstremno velikih rezultata: istodobno, ta je distribucija simetrična. No, to ne mora uvijek biti tako, jer se mnoge pojave raspoređuju i na druge načine. Ako nam je poznato kako se neka pojava distribuira, možemo s većim razumijevanjem interpretirati pojedinačne rezultate. Tako je, na primjer, poznato da se bilirubin u krvi distribuira asimetrično, da se broj nesreća pri radu također distribuira asimetrično (jer najveći broj ljudi ili nema nesreća ili ima samo jednu, a vrlo mali broj ljudi ima mnogo nesreća), da promjer ljudskog srca daje nepravilnu krivulju s dva vrha, da upitnici za ispitivanje stavova obično daju takvu distribuciju u kojoj ima najmanje "srednjih" rezultata itd.

U deskriptivnu statistiku mogli bismo čak svrstati i jedno od inače vrlo kompleksnih pitanja, tj. pitanje što je "normalno" a što "nenormalno". O tom problemu mnogo je pisano i raspravljano, i problem zasigurno prelazi granice statistike, ali se uz pomoć statistike i statističkog načina mišljenja problem ipak može jasnije uočavati. Prvenstveno, izraz "normalno" znači ono što je prosječno, ili što je "najčešće", katkada "normalno" znači "bez bolesnih znakova", itd. Na primjer, povišeni tlak kod većeg broja starijih muškaraca nije "normalan" utoliko što je to znak nekih patoloških promjena u nekim organima, ali je, s druge strane, "normalan" utoliko što je "prosječni tlak starijih ljudi viši od tlaka mladih". No jedan od najtežih problema, osobito izražen u disciplinama koje se bave bolesnim stanjima organizma jest pitanje *granice* između "zdravog" i "bolesnog", tj. "normalnog" i "nenormalnog". Pitanje uopće nije rješivo na potpuno zadovoljavajući način (jer uvijek postoji značajno *preklapanje* simptoma između zdrave i bolesne grupe), ali neki grafički statistički postupci osiguravaju nam da s relativno najmanje pogrešaka odredimo gdje se ta granica nalazi.

Poznato je da mnoge pojave pokazuju neku manju ili veću međusobnu *zavisnost:* kada je jedna pojava u porastu, u porastu (ili u padu) je i druga, pa tako do određene granice možemo iz neke vrijednosti u jednoj od tih pojava zaključivati na odgovarajuću vrijednost u drugoj pojavi (na primjer iz podatka o visini nekog čovjeka možemo s određenom vjerojatnošću nagađati i o njegovoj vjerojatnoj težini). Ta se pojava naziva *korelacijom*, i nalazimo je svagdje oko sebe: u korelaciji su visina i težina, inteligencija i školski uspjeh, tjelesna temperatura i puls, intenzitet tjelesnog napora i krvni tlak, količina kiše i bujnost vegetacije, starost i tjelesna snaga, motivacija i radni učinak, itd., itd. Za psihologa, liječnika, sociologa i druge stručnjake pojam korelacije i te kako je važan, jer nam — kako je već rečeno — korelacija omogućuje postavljanje *prognoze* iz jedne pojave u drugu: što je povezanost (korelacija) viša, prognoza je točnija.

Ad 3. *Zaključivanje iz konkretnog slučaja na "opći zakon".* Ako smo izmjerili visinu, recimo, 100 desetogodišnje muške djece u Varaždinu i 100 desetogodišnje muške djece u Zagrebu, te smo našli da su u prosjeku zagrebačka djeca viša od varaždinske, onda — pod pretpostavkom da je mjerenje izvršeno bez pogrešaka — *nema nikakve sumnje* da su *izmjerena* varaždinska djeca u prosjeku niža od *izmjerene* zagrebačke djece. Ili, ako smo primijenili neku novu terapiju na određeni broj bolesnika, i usporedili trajanje bolesti s kontrolnom grupom bolesnika, liječenom na klasičan način, pa smo našli da su bolesnici, liječeni novim postupkom, ozdravili u prosjeku za kraće vrijeme — onda je to *kod promatranih bolesnika* sigurno tako. Također, ako smo našli da djeca, koja žive u teškim ekonomskim prilikama, u prosjeku imaju slabiji školski uspjeh od djece koja žive u obiteljima bez ekonomskih teškoća — također nema sumnje da je to točno kod djece na kojoj smo mjerenje izvršili. No, nas u ovim slučajevima zapravo toliko ne zanima jesu li *ova konkretna* zagrebačka djeca viša od varaždinske, ili jesu li ovi *konkretni* bolesnici ozdravili brže od druge grupe, ili jesu li *konkretno ispitivana* djeca iz ekonomski ugrožene sredine slabiji učenici. Ono što nas zanima jest ponajprije to *možemo li mi iz tih pojedinačnih slučajeva zaključivati općenito*, tj. možemo li na temelju izvršenog mjerenja zaključiti da su *općenito* zagrebački 10-godišnji dječaci viši od svojih varaždinskih vršnjaka, da nova metoda liječenja *općenito* u prosjeku skraćuje trajanje bolesti, kao i je li ekonomska ugroženost *općenito* povezana sa slabijim uspjehom djeteta u školi.

Kako vidimo, ovdje iz pojedinačnog slučaja želimo zaključiti na neku pojavu, koja je možda *opći zakon*. Drugim riječima, u svim takvim slučajevima (a to je gotovo redovita pojava pri istraživačkom radu!) mi zapravo želimo *iz pojedinačnog zaključiti na općenito*.

"Inferencijalna statistika", tj. onaj dio statističkih postupaka koji nam omogućuje stvaranje *zaključaka*, omogućuje nam da ustanovimo smijemo li ili ne smijemo nađenu pojavu smatrati generalnom, dakle općevažećom. Doduše, kao što će se vidjeti, potpuna, "100%-tna" sigurnost u statističkim je zaključivanjima praktički nemoguća, ali pri zaključivanju ipak možemo postići sigurnost koja je blizu sigurnosti od 100%.

Uostalom, o problemima zaključivanja u statistici bit će još govora, i tom je pitanju posvećeno jedno čitavo poglavlje u ovoj knjizi.

Kao što se logički i može očekivati, inferencijalna statistika je to "opreznija" u zaključivanju što smo na manjem 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. Ali sve to, naravno, pod pretpostavkom da je "uzorak" na kojem smo mjerenje izvršili *reprezentativan* za grupu kojoj taj uzorak pripada, tj. da uzorak nije "pristran". Kada bismo, na primjer, izabrali samo najbolje učenike iz nekog razreda pa im izmjerili znanje matematike, onda rezultat, dakako, ne bi bio "reprezentativan" za cijeli razred. Kada bismo, primjerice, željeli dobiti podatak o prosječnim osobnim dohocima zagrebačkih građana, pa kada bismo iz telefonskog imenika nazvali svakog 50. pretplatnika i tražili od njega podatke o dohocima, čak i onda kad bi nam svi upitani dali točne podatke, ***ne bi nam dobiveni rezultati dali ispravnu sliku,*** jer naš uzorak nije dovoljno "reprezentativan" za pojam "građanin Zagreba": u našem se uzorku nalaze samo vlasnici telefona, a možemo biti sigurni da prosječni dohoci tih ljudi nisu jednaki prosječnim dohocima onih građana koji nemaju telefona.

Kako u inferencijalnoj statistici problem uzorka i njegove reprezentativnosti ima primarno značenje (jer svi zaključci, koje iz dobivenih rezultata izvodimo, vrijede samo onda ako je uzorak pravilno izabran), mnogi statističari inferencijalnu statistiku nazivaju još i "statistikom uzoraka".

Ad 4. *Planiranje istraživanja i eksperimenata.* Poznavanje statističkih metoda može istraživaču dosta pomoći pri planiranju istraživanja odnosno izrade nacrta eksperimenta. Spomenut ćemo iz toga područja samo dva primjera:

a) pitanje eksperimentalne i kontrolne skupine i

b) pitanje veličine uzorka.

- Kao što je poznato, efekt nekoga određenog djelovanja možemo provjeriti samo onda ako smo istodobno ispitivali i slučajeve *bez* utjecaja tog djelovanja. Ako nas zanima ima li neki lijek određeni utjecaj na neku bolest, moramo njegovo djelovanje usporediti sa slučajevima kada taj lijek nije primijenjen. U eksperimentalnom radu skupinu na kojoj provjeravamo efekt neke metode nazivamo ***eksperimentalnom*** skupinom, a onu na koju metoda nije primijenjena, ***kontrolnom*** skupinom. No kontrolna i eksperimentalna skupina moraju biti međusobno što sličnije u svim onim svojstvima koja bi mogla imati utjecaj na ono što ispitujemo; na taj način jedina razlika između obiju skupina bit će to da smo na jednu primijenili, a na drugu nismo primijenili neki postupak. Na primjer, ispitujemo li neku novu metodu učenja stranog jezika, onda ćemo tu metodu usporediti s efektom neke druge (do sada upotrebljavane) metode, ali ispitanici eksperimentalne i kontrolne grupe moraju biti međusobno što sličniji po starosnoj dobi, školskom obrazovanju, predznanju iz tog jezika, itd. Jednako tako, ako ispitujemo neku novu kiruršku tehniku, ispitanici eksperimentalne i kontrolne grupe moraju biti međusobno "komparabilni" po težini oboljenja, po starosnoj dobi i drugim faktorima koji bi mogli biti u vezi s bolešću ili s faktorima rezistencije.

Dok je u laboratorijskim ispitivanjima (relativno) lakše provesti takvo izjednačavanje eksperimentalne i kontrolne skupine (jer pri laboratorijskim eksperimentima obično možemo prethodno pažljivo sastaviti kontrolnu i eksperimentalnu grupu od ispitanika sa sličnim karakteristikama), dotle je to često vrlo teško provesti npr. u *kliničkom* radu, u praktičnom radu industrijskog psihologa, i uopće u radu "na terenu", gdje ne možemo po volji sastavljati uzorak, nego smo prisiljeni raditi na tzv. "prigodnom" uzorku, tj. na uzorku kojim jedino raspolažemo.

Jedno od važnih pitanja pri planiranju nekog istraživanja — u kojem katkada dolazi do nesporazuma između početnika i iskusnog statističara — jest pitanje *veličine uzorka.* Početnik, očekujući od statističara gotove "recepte" za neke svoje postupke, obično se iskreno razočara kada mu statističar na njegovo pitanje koliko velik uzorak da uzme za neko istraživanje, odgovori da — ne zna!

Potrebna veličina uzorka zavisi naime od više faktora, od kojih neki ne moraju uopće biti poznati prije početka pokusa ili istraživanja. Evo glavnih faktora:

— *Željena preciznost rezultata:* što nam je veća preciznost potrebna u nekom istraživanju, to je potreban i veći uzorak.

— *Varijabilnost pojave, koju ispitujemo:* kada neka pojava uopće ne bi varirala (tj. kada bi svi članovi neke populacije imali identičnu količinu one karakteristike koja nas zanima), bio bi dovoljan uzorak veličine 1, jer bismo s tim jednim podatkom već znali sve o pojavi koju mjerimo. Iz toga slijedi da je potreban to veći uzorak što pojava, koju istražujemo, više varira. Kako, primjerice, čovjekova težina mnogo više varira od čovjekove tjelesne temperature (kao što ćete vidjeti u poglavlju "Mjere varijabilnosti", u težini ljudi variraju preko 30 puta više nego u tjelesnoj temperaturi!), iz toga slijedi da bi nam za dobivanje podataka o prosječnoj tjelesnoj temperaturi (zdravih) ljudi bio dovoljan znatno manji uzorak nego za dobivanje podataka o prosječnoj težini.

— *Frekvencija pojave:* što je pojava, koju ispitujemo, općenito *rjeđa*, to je veći uzorak potreban; kod vrlo rijetkih pojava potrebni su zato vrlo veliki uzorci. Evo primjera: ako bi netko želio ispitati neko novo cjepivo protiv poliomielitisa, pa bi, uzmimo, skupinu od 100 djece cijepio klasičnom vakcinom, a drugu, podjednako veliku skupinu, cijepio tim novim cjepivom, vrlo vjerojatno bi se dogodilo da ni u jednoj ni u drugoj grupi nijedno dijete ne bi oboljelo!

Slična se pojava može dogoditi praktičarima (psiholozima ili liječnicima) koji u nekoj maloj radnoj organizaciji ispituju, na primjer, razliku između mladih i starijih radnika u frekvenciji nesreća pri radu. Naprotiv, da smo u prvom slučaju

uzeli, pretpostavimo, u svaki uzorak oko 100 000 djece, onda bi se možda dogodilo da ih je u uzorku kontrolne grupe oboljelo 15-20, a u uzorku eksperimentalne grupe nijedno dijete — pa bismo na temelju toga, nakon potrebne statističke obrade — vjerojatno već mogli zaključiti nešto o korisnosti novog cjepiva.

Djelomično u vezi s pitanjem veličine uzorka jest i pitanje o razlikama u veličini između kontrolnog i eksperimentalnog uzorka. Još donedavno statističari nisu ovdje postavljali nikakve posebne zahtjeve, jer same formule dopuštaju da među uzorcima budu bilo koje razlike u veličini. No, u novije vrijeme utvrđeno je da je znatno bolje ako su oba uzorka približno jednake veličine (o tome će još biti govora u ovoj knjizi u poglavlju "Razlika između dvije aritmetičke sredine"), pa se u tom slučaju čak mogu kršiti neka pravila u računima, u vezi s uvjetima, potrebnim za primjenu neke metode, a da pri tome rezultati ipak budu pouzdani i upotrebljivi.

Sva ova 4 nivoa korisnosti poznavanja statistike rezultiraju *glavnom* koristi koju suvremeni čovjek ima od statistike, a to je ***usvajanje statističkog načina mišljenja,*** tj. usvajanje određene tehnike mišljenja i rada, bez koje zapravo nema znanstvenog mišljenja.

Često se mogu čuti izjave otprilike ovog sadržaja: "Jedino statistički obrađeni podaci mogu imati znanstvenu vrijednost." To je točno, ali to, dakako, ne znači da statistička obrada *garantira* znanstvenu vrijednost podataka. Ili — kako kaže Blalock — "statistika može biti samo pomoć, ali nikada zamjena za zdrav razum". Statistička je obrada samo *metoda* obrade podataka, ali nam ona ne može jamčiti da je skupljanje podataka ispravno provedeno, ili da je čitavo istraživanje dobro planirano. Bit će još dosta prilike u ovoj knjizi za primjere koji će pokazati čak *zloupotrebu* statistike radi nenamjernog ili čak namjernog zavaravanja javnosti. Jer, i potpuno izmišljeni podaci, mogu se statistički obraditi! No za to, naravno, ne treba optuživati statistiku, nego onoga tko je primjenjuje; ako netko kirurškim nožem izvrši zločin i nekoga ubije, nije nož za to odgovoran. Ili, ako netko uz pomoć statistike objavi neku laž, nije za to kriva statistika: statistika ne laže, nego lažu ljudi.

Katkada možemo čuti mišljenje da su fizikalne znanosti mnogo "egzaktnije" znanosti od znanosti koje se bave živim bićima (biologija, medicina, psihologija, sociologija, antropologija i dr.). Ove druge zbog toga se katkada zovu "statističke znanosti". Međutim — prema mišljenju Virginije Senders — razlika između jednih i drugih znanosti sastoji se u sljedećem:

1. *Varijabilnost* većine fizikalnih zakona obično je tako malena da ne maskira pojavu koju istražujemo. Naprotiv, varijabilnost životnih pojava, tj. živog materijala, katkada je toliko velika da otežava pronalaženje zakonitosti.

2. Druga se razlika sastoji u tome što se od onih znanosti koje se bave živim bićima, često zahtijeva prognoza o funkcioniranju i ponašanju *jednog* individuuma, dok se od fizike rijetko zahtijeva da dade prognozu o jednoj jedinoj molekuli ili atomu. Kada, na primjer, sociolozi i psiholozi daju prognozu o velikoj skupini ljudi, oni to čine uspješno, što pokazuju točna predviđanja o približnom broju prometnih nesreća za neke blagdane, o broju glasača na izborima i o ishodu glasovanja i sl. Ali te znanosti ne mogu dati (osim u nekim rijetkim slučajevima) sigurnu prognozu što će neki određeni *pojedinac* učiniti ili što će se s njime dogoditi: ako predvidimo

da će 80% glasača glasovati za nekog kandidata, mi, na žalost, ne možemo mnogo reći o tome kako će neki *određeni* pojedinac glasovati; ako znamo da neki novi lijek u prosjeku vrlo uspješno liječi neku bolest, mi ipak ne možemo garantirati da će i pojedinac N. N. biti pomoću tog lijeka izliječen. Međutim — misli Senders — razlika između tih dvaju područja znanosti samo je prividna, jer fizičar može s velikom preciznošću odrediti koliko će se od jednog grama radija raspasti u toku deset godina, ali bi se vrlo začudio kada bismo mu postavili pitanje neka točno kaže kada će se raspasti neki *određeni*, točno definirani atom. A upravo to se često zahtijeva od društvenih znanosti i od medicine! (Takav *individualni* pristup stoga i nazivamo "*kliničkim*" pristupom, za razliku od "*statističkog*" pristupa, tj. pristupa preko velikog broja.)

Kao što je poznato, u javnosti u velikoj mjeri postoji "*alergičnost*" prema statistici, ne samo među laicima nego često i među stručnjacima različitih struka. Na primjer, kada se studenti nekih društvenih ili prirodno-društvenih znanosti (sociologija, psihologija) iznenade što moraju slušati i polagati statistiku, oni to često komentiraju riječima da su taj studij upisali upravo zato što ne vole i ne znaju — matematiku. Dakle, poistovjećuju statistiku s matematikom.

No, pođimo redom, i pokušajmo nabrojiti glavne razloge toga negativnog stava velikog broja ljudi prema statistici.

1. *Iskrivljeno i odveć simplificirano značenje pojma "statistika"*; mnogi, naime, taj izraz upotrebljavaju za označavanje *tabličnog pregleda* nekih ispitanih podataka. Kada takvi ljudi kažu "imam statističke podatke", pokažu vam tablice, u kojima se pregledno vide sume u svakoj kategoriji! Naravno, takva je "statistika" do maksimuma suhoparna i dosadna, pa je ljudi izbjegavaju.

2. Mnogi su ljudi frustrirani pred statistikom zbog njezinog za njih nerazumljivog jezika, a naročito zbog *nepoznatih simbola*, na koje u statistici nailaze. No, *svaka* znanost ima svoje simbole, koji nisu ništa "ezoterično", nego predstavljaju skraćeni i praktičniji sustav komuniciranja, koji onome koji taj "jezik" poznaje, predstavlja jednako tako normalne riječi kao i sve ostale riječi. Kao što laika mogu impresionirati izrazi RR, SE, bilirubin, EEG, vitalni kapacitet ili, na primjer, G-faktor, satijacija, semantički diferencijal, IE-test (prvi od tih izraza koriste se u medicini, drugi u psihologiji), tako ga, dakako, mogu impresionirati i izrazi kao što su, primjerice, varijanca, $\overline{X}$, hi-kvadrat, $P < 0,05$, standardna pogreška i dr., a koji pripadaju statistici, i nisu ništa više "ezoterični" od onih drugih izraza. Treba, doduše, priznati da početnika u statistici može osobito obeshrabriti činjenica što svi statistički simboli nisu još uvijek unificirani pa za isti pojam različiti autori često još upotrebljavaju različite simbole (na primjer, aritmetičku sredinu neki pišu znakom M, neki A. M., neki $\overline{X}$, itd.). No sličnih pojava, iako vjerojatno u manjoj mjeri, ima i u drugim znanostima.

3. Mnogi ljudi — a to je vjerojatno i glavni razlog alergičnosti prema statistici — smatraju da je *statistiku nemoguće razumjeti i svladati bez znanja matematike*.

To, međutim, nije točno! Glavni se statistički *principi i način mišljenja* mogu usvojiti potpuno *logičkim* putem, a od "*matematike" je potrebno znati samo 4 osnovne operacije*, dakle zbrajanje, oduzimanje, množenje i dijeljenje. Doduše, u statističkim je računima često potrebno vaditi i drugi korijen, no za taj posao

gotovo u svim satističkim udžbenicima postojale su tablice u dodatku. A danas za takve slučajeve postoje elektronska džepna računala.

U jednoj anegdoti, koju spominje već prije citirani Mainland, neki matematičar, koji se specijalizirao u matematici vjerojatnosti, ustvrdio je da znade način kako da spriječi prehladu: kad bi osjetio da mu se sprema prehlada, on bi uveče prije spavanja uzeo jedan lijek, i drugog jutra prehlade nije bilo. Njegova supruga je međutim ustvrdila da to nije nikakav dokaz, jer da on zapravo ne zna što bi se dogodilo da taj lijek *nije* uzeo. Za razliku od svog supruga, ona je — kaže Mainland — primijenila statistički način rezoniranja i ilustrirala činjenicu da statistika i nije dio matematike.

Matematika je potrebna samo za *profesionalne* statističare, tj. za one koji statistiku i njezine metode stvaraju, jer oni neke statističke zakone moraju izraziti matematičkim putem i tako stvoriti formulu pomoću koje se nešto u statistici izračunava. No matematika nije potrebna za *primjenu* tih postupaka, dakle za one koji statistiku koriste samo kao jednu od metoda u svom radu. Slično kaže i statističar Levinson: "Nije moguće biti profesionalni statističar u doslovnom smislu te riječi bez dobrog poznavanja više matematike. No nije potrebno biti profesionalni statističar, pa da bi se znala razlika između dobrih i loših statističkih postupaka, i da bi se pametno koristile postojeće statističke metode". Budući da statistika koristi postupke koje su *izradili* matematičari, to je u statistici matematika upadljivija nego u nekim drugim disciplinama, i to početnika zavodi. Početnik često smatra da zato što ne može razumjeti kako je *nastala* neka formula koju on koristi, ne može razumjeti ni statistički način rezoniranja. No to, naravno, nije tako, što ćemo lako dokazati: automobil postoji zato da ga netko *vozi*; svi poznamo bolje i lošije vozače, a i one koji — kako kažemo — odlično voze. Vjerojatno nema nikoga tko bi ustvrdio da odličan vozač može biti samo onaj koji potpuno i detaljno razumije konstrukciju i funkciju motora! Jednako tako je i u situacijama u kojima se služimo npr. televizorom, radiom, elektrokardiografom, orto-raterom, mikroskopom, aparatom za mjerenje bazalnog metabolizma, itd., itd. Sve te aparate možemo vrlo korisno i mudro upotrebljavati a da ne znamo mnogo o mehanici, optici, elektronici, itd. Uostalom, koliko ljudi znade po kojem principu funkcionira i kako se konstruira neko PC računalo? Pa ako to ne znamo, zar se ne možemo tim računalom *vrlo uspješno služiti*?!

Dakle, u svim navedenim slučajevima mi razumijemo možda samo neke osnovne principe konstrukcije i funkcioniranja pojedinog aparata, ali ono što treba da *potpuno razumijemo, to je način upotrebe i situacije u kojima takav uređaj treba upotrijebiti*. Takav jednaki zahtjev postavlja se i na "konzumenta" statistike, tj. on statističke metode treba s razumijevanjem primijeniti na pravom mjestu. Statističke su metode — kaže jedan statističar — za znanstvenog radnika ono što je alat za tesara. On prvo treba naučiti kako će svoj alat upotrebljavati, a nakon toga može — upotrebom tog alata — postići lijepe rezultate. Ali sam alat mora biti upotrijebljen na pravom mjestu. Obrtnik neće npr. upotrijebiti batić, dlijeto ili malu rezbarsku pilicu da bi blanjao dasku, niti će upotrijebiti čekić da bi uvrnuo neki vijak. Jednako tako treba i znanstveni radnik upotrebljavati statističke metode samo za ono za što su predviđene, i pri tome treba *znati što radi*. Upravo u tom zahtjevu katkada dolazi do nesporazuma: poneki student, koji je *pogriješio u računu*, buni se što mu ta pogreška snizuje ocjenu, jer — kako on kaže — ta pogreška nema veze sa znanjem i razumijevanjem statistike. Zaista, često i nema veze, i u takvim slučajevima ona se obično i ne penalizira strogo. No, ako ta pogreška dovede do rezultata koji je *besmislen* ili logički naprosto *nemoguć*, a onaj tko je pogrešku učinio, to ne primijeti — onda se više ne radi o "čistoj" računskoj pogrešci, nego o težoj pogrešci, a koja se sastoji u tome da zapravo ne znamo i ne razumijemo što radimo!

Početnika nematematičara donekle u statistici može obeshrabriti i činjenica — slična već spomenutoj pojavi upotrebe različitih simbola — da su pojedini statistički računski postupci u različitim statističkim udžbenicima prikazani "različitim" formulama. No, zapravo ne radi se o različitim formulama, nego samo o različitim transformacijama *iste* formule — što će matematičaru dakako odmah biti jasno, ali što nematematičara može frustrirati. Zaista, ne moramo nematematičara kriviti ako ne prepozna da je izraz

$$\sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$$

isto što i izraz

$$\frac{1}{N}\sqrt{N\Sigma X^2 - (\Sigma X)^2},$$

ili da $5^{1/2}$ znači $\sqrt{5}$, ili da 5^{-9} znači $\frac{1}{5^9}$.

U ovoj se knjizi pokušava osnovne statističke postupke i principe protumačiti što je više moguće logičkim putem. Čitaoci će katkada naići na slučajeve da se iza neke formule navodi neka druga, praktičnija formula, kojom se isti statistički postupak može provesti pa će opravdano postaviti pitanje čemu je prvu formulu trebalo uopće navoditi. Međutim, to je učinjeno u svim onim slučajevima kada prva formula logički slijedi iz dotadašnjih izlaganja. Tek kad je to shvaćeno, opravdano je prijeći na drugu, praktičniju formulu, koja zapravo predstavlja matematički ekvivalent prvoj formuli.

Ova je knjiga prvenstveno namijenjena onima kojima je "matematika" glavni razlog zbog kojeg ne vole statistiku. Kao što je već rečeno, u osnovnim statističkim postupcima, koji su prijeko potrebni svakom znanstvenom radniku bilo koje struke, matematike zapravo nema — iako bi je naravno moglo biti kada bi stručnjak želio uvijek do kraja shvatiti postanak neke formule.

Autor ovog udžbenika — i sam nematematičar — stjecao je svojedobno "metodom vlastite kože" iskustva s osnovnim teškoćama na koje nematematičari nailaze u statistici. A kasnije, imao je prilike i na drugim ljudima godinama pratiti, prilikom nastave statistike, različite vrste teškoća s kojima se sukobljuje nematematičar u susretu sa statistikom. I zato je teškoće ove vrste nastojao prebroditi na najbezbolniji način služeći se — što je više bilo moguće — logičkim zaključivanjem i praktičnim primjerima.

I, na kraju, još jedna važna primjedba: rekli smo da automobil možemo dobro voziti i bez nekog naročitog poznavanja mehanike i elektrike njegovih uređaja. Jednako tako statistiku možemo vrlo uspješno upotrebljavati i bez poznavanja matematike. Ali automobil nećemo naučiti uz pomoć *čitanja* o tome kako se automobil vozi ili uz pomoć *predavanja* o toj temi! Automobil treba *voziti*. Jednako tako svatko tko ima iskustva s praktičnim radom na nekom osobnom računalu, veoma dobro znade da se uspješna upotreba računala može postići *jedino* praksom, tj. čestim radom na računalu. Isto tako niti statistiku nećemo naučiti samo uz pomoć čitanja knjiga ili slušanja predavanja. I statistički rad, tj. *praktično provođenje* pravila i zakona koje smo logički mogli posve razumjeti, jest *vještina koju treba u praksi uvježbavati*.

U tu svrhu na kraju većine poglavlja dani su čitaocu *primjeri i zadaće za vježbu*. Ako zaista želite naučiti statističku metodologiju, onda nemojte zaboraviti da ti zadaci postoje upravo zato da biste stekli određenu "rutinu" u statističkim vještinama. Na kraju knjige naći ćete rješenja tih zadataka, pa ćete tako moći provjeriti do kojeg ste stupnja potrebne vještine usvojili.

2.

PROVJERITE SVOJE ZNANJE IZ OSNOVA RAČUNANJA

U prvom smo poglavlju spomenuli da je za praktičnu upotrebu statistike potrebno znati samo *četiri osnovne računske operacije:* zbrajanje, oduzimanje, množenje i dijeljenje. Događa se katkada da pojedinci to smatraju — matematikom! Međutim, degradirati matematiku na osnovne račune — koji se svladavaju već u početnim razredima osnovne škole — bilo bi neopravdano, i zato možemo ponoviti tvrdnju da barem za osnove, iznesene u ovoj knjizi, matematika nije potrebna.

Iz razumljivih razloga, koji su većinom uvjetovani našim načinom života i sustavom školovanja na fakultetima, poneki su ljudi, na žalost, zaboravili i neka osnovna *računanja*, pa tako, primjerice, imaju teškoća pri dijeljenju ili množenju decimalnih brojeva, u radu s razlomcima, pri vađenju korijena (iz tablica) ne mogu se snaći kamo treba staviti decimalni zarez, itd.

Da bi čitaoci, prije nego što počnu s proučavanjem materijala u ovoj knjizi, *provjerili* svoje sadašnje znanje osnova računanja, naveden je u ovom poglavlju niz zadataka koje treba riješiti. Mnogima će ti zadaci biti infantilno jednostavni. To bolje! No, ti zadaci nisu pisani za njih, nego za one koji će na nekima od njih zapeti.

Iza tih zadataka nalaze se rješenja. Provjerite jesu li vaša rješenja točna. Ako nisu, *nemojte početi proučavati materijal iz knjige tako dugo dok ne obnovite svoje znanje iz računanja,* što će predstavljati možda nekoliko desetaka minuta vježbanja na zadacima koje ćete sami postaviti iz područja koja vam stvaraju teškoće. Dosadašnja moja iskustva na usmenim i pismenim ispitima polaznika brojnih kurseva iz osnovnih statističkih metoda pokazuju da je *osnovni* razlog pogrešnim rezultatima i teškoćama u radu upravo u tomu što pojedinci učine neke osnovne računske pogreške u radu, a koje se ponajčešće sastoje u pogrešno postavljenom decimalnom zarezu, u teškoći oko dijeljenja dva decimalna broja i slično. U svim tim slučajevima takvi pojedinci ne mogu, na žalost, uspješno *primijeniti* svoje znanje i razumijevanje statističkih postupaka jer je konačan rezultat rada ipak negativan.

Iza rješenja zadataka navedena su u svim važnijim slučajevima kratka pravila koja se odnose na način računanja pojedinih vrsta zadataka. Ta će pravila poslužiti onima koji su ustanovili da u nekom od ovih područja nailaze na računske teškoće.

No prije nego što prijeđemo na zadatke, treba raščistiti s jednom dilemom. Danas kada "eksplozivni" razvoj elektronskih računala, a među njima i malih džepnih računala, napreduje — mogli bismo reći — gotovo iz dana u dan, možda će mnogi čitatelj postaviti pitanje *čemu treba ponavljati ove osnovne računske operacije kada se one uz pomoć računala mogu izvesti u trenu.*

Pitanje je opravdano, ali ne potpuno. Prvo, rijetko tko stalno sa sobom nosi elektronsko džepno računalo, i, prema tome, vrlo će se često naći u situaciji da će *morati* "pješice" izvesti neku računsku operaciju. A to neće moći ako je zaista zaboravio osnove računanja. Drugo, ako te osnove računanja poznajemo i *razumijemo*, moći ćemo logičkim zaključivanjem uočiti neke *pogreške*, koje su nam se u radu mogle potkrasti, pa makar račun izvodili i uz pomoć elektronskog računala (npr. postavljanje decimalnog zareza na pogrešno mjesto, jer ne postavljaju sva računala decimalni zarez automatski). Ako je netko gotovo posve zaboravio smisao kvadriranja, pa bi "pješice" ili uz pomoć elektronskog računala kvadrirao, na primjer, broj 0,53, i kao rezultat dobio 0,28, moglo bi mu se dogoditi da izgubi mnogo vremena tražeći pogrešku, jer "on se još iz školskih dana sjeća da svaki broj, koji kvadriramo, postaje veći nego što je bio prije". Ponavljanjem osnovnih pojmova i tehnika računanja on će spriječiti da mu se to dogodi.

Ako su nam logički principi računanja dobro poznati, onda zaista nema pravog smisla vaditi, na primjer, drugi korijen "pješice", čak niti uz pomoć tablica, ako rezultat možemo dobiti u trenu pritiskom jednog dugmeta na računalu. Isto je tako točno da je dijeljenje dvaju decimalnih brojeva neusporedivo jednostavnije uz pomoć računala. *Neke* formule, koje su u ovoj knjizi navedene (a koje će nematematičar odmah uočiti, jer će mu se činiti "zastrašujuće" i potpuno nerazumljive) navedene su upravo vodeći računa o tome da se one uz upotrebu kvalitetnijih džepnih kalkulatora mogu vrlo jednostavno koristiti.

2.1. ZADACI ZA PROVJERAVANJE ZNANJA

1. a) $-5-3=$ b) $-5+7=$ c) $9-16=$ d) $-8-(-4)=$ e) $5+(-4)=$ f) $-2-(-6)=$

2. Zaokružite ove decimalne brojeve:
   a) 37,8 na cijeli broj
   b) 115,4 na cijeli broj
   c) 0,762 na jednu decimalu
   d) 12,849 na jednu decimalu
   e) 6,98 na jednu decimalu
   f) 0,0329 na dvije decimale.

3. a) $6,35\cdot 10=$ b) $0,00209\cdot 10=$ c) $632\cdot 10=$ d) $8,92\cdot 100=$ e) $19,8\cdot 100=$ f) $0,00076\cdot 100=$ g) $23,247\cdot 1\,000=$ h) $0,0439\cdot 1\,000=$ i) $37968,3245\cdot 1\,000=$

4. a) 18,76 : 10 = b) 0,092 : 10 = c) 92,3 : 100 = d) 85,42 : 100 = e) 0,071 : 1 000 = f) 356,5 : 1 000 =

5. a) $\frac{182}{0,1}=$ b) $\frac{0,09}{0,1}=$ c) $\frac{12,85}{0,01}=$ d) $\frac{0,2046}{0,01}=$ e) $\frac{1705}{0,001}=$ f) $\frac{0,0092}{0,001}=$

6. Izvršite operaciju "kraćenja" u ovim razlomcima:
   a) $\frac{3}{12}$; b) $\frac{4\cdot 3\cdot 7}{2\cdot 6}$; c) $\frac{(2+3)\cdot 2}{4}$.

7. Riješite se razlomaka u ovim izrazima:
   a) $a=\frac{bc}{d}$; b) $1/2x=y$; c) $\frac{a}{b+d}=y$.

8. Izračunajte:
   a) $0,097\cdot 7=$
   b) $3,5\cdot 2,5=$
   c) $8,56\cdot 0,034=$ (izračunati na 3 decimale)

9. a) 32 : 1,6 = (izračunati na 1 decimalu)
   b) 2,8 : 3,92 = (izračunati na 2 decimale)
   c) 17,353 : 0,06 = (izračunati na 2 decimale)

10. a) $3,5^2=$ (na 2 decimale)
    b) $0,79^2=$ (na 3 decimale)
    c) $0,003^2=$ (na 6 decimala)
    d) $5^3=$
    e) $4^0=$

11. a) $5\cdot\frac{7}{6}=$ b) $8\cdot\frac{3}{6}=$ c) $\frac{1}{15}\cdot 4=$

12. a) $\frac{3}{4}\cdot\frac{5}{11}=$ b) $\frac{1}{7}\cdot\frac{1}{9}=$ c) $\frac{6}{19}\cdot\frac{1}{2}=$

13. a) $\frac{1}{2}:\frac{1}{2}=$ b) $\frac{3}{7}:\frac{6}{7}=$ c) $\frac{2}{3}:\frac{4}{5}=$

14. a) $4 : \frac{1}{2} =$ c) $\frac{1}{2} : 2 =$

b) $8 : \frac{3}{4} =$ d) $\frac{3}{8} : 5 =$

15. a) $1/3 + 2/4 =$ c) $2/9 - 1/3 =$
b) $7/8 + 2/16 =$ d) $3/7 - 2/8 =$

16. a) $\sqrt{16} =$ (na 2 decimale) d) $\sqrt{32{,}41} =$ (na 3 decimale)
b) $\sqrt{160} =$ (na 2 decimale) e) $\sqrt{1{,}60} =$ (na 2 decimale)
c) $\sqrt{6\,420} =$ (na 2 decimale) f) $\sqrt{0{,}052} =$ (na 3 decimale)

17. a) $16 + 8 \cdot 2 =$ c) $(16 + 8) \cdot 2 =$
b) $16 + 8 : 2 =$ d) $(16 + 8) : 2 =$

18. a) $(a + b)^2 =$
b) $(a - b)^2 =$

19. a) $\sqrt{16 + 9} =$ c) $\sqrt{25 - 9} =$
b) $\sqrt{16 \cdot 4} =$ d) $\sqrt{36 : 9} =$

20. Ako su rezultati nekog mjerenja ovi: 2, 3 i 4, pa ako te rezultate označimo slovom X, koliko iznosi:
a) ΣX^2; (Σ znači suma)
b) $(\Sigma X)^2$

21. Izračunajte:
a) 2!; b) 5!; c) 0!

2.2. RJEŠENJA I NEKA OSNOVNA PRAVILA

Stilizacije nekih pravila nisu naročito precizne u matematičkom smislu i treba da posluže samo kao podsjetnik.

1. a) −8; b) +2; c) −7; d) −4; e) +1; f) +4.

2. a) 38; b) 115; c) 0,8; d) 12,8; e) 7,0; f) 0,03.

Ako je znamenka koja slijedi iza posljednje koju želimo zadržati, veća od 5, posljednja zadržana znamenka povećava se za 1.

Ako je znamenka koja slijedi iza posljednje koju želimo zadržati, manja od 5, posljednja zadržana znamenka ostaje kakva jest.

Ako je znamenka, koja slijedi iza posljednje, koju želimo zadržati, točno 5 (bez ostatka, tj. sa samim nulama koje slijede), vrijedi ovaj običaj: ako je znamenka koju treba zaokružiti neparna, povećavamo je za 1, a ako je parna, ostaje kao što jest. Prema tome, u takvim slučajevima zadnja znamenka koju zadržavamo, mora biti parna.

3. a) 63,5; b) 0,0209; c) 6 320; d) 892; e) 1 980; f) 0,076; g) 23 247; h) 43,9; i) 37 968 324,5

Decimalni se broj množi s 10 tako da se decimalni zarez pomakne za jedno mjesto nadesno.

Pri množenju sa 100, decimalni se zarez pomakne za 2 mjesta nadesno, a pri množenju s 1 000, za tri mjesta nadesno, itd.

4. a) 1,876; b) 0,0092; c) 0,923; d) 0,8542; e) 0,000071; f) 0,3565.

Decimalni se broj dijeli s 10 tako da se decimalni zarez pomakne za jedno mjesto nalijevo.

Pri dijeljenju sa 100, decimalni se zarez pomakne za 2 mjesta nalijevo, pri dijeljenju s 1 000, za 3 mjesta nalijevo, itd.

5. a) 1 820; b) 0,9; c) 1 285; d) 20,46; e) 1,705 000; f) 9,2.

Svaki se broj dijeli s 0,1 tako da se pomnoži s 10, a s 0,01 dijeli se tako da se broj pomnoži sa 100, s 0,001 tako da se broj pomnoži s 1 000, itd.

6. a) 1/4; b) $\frac{2 \cdot 1 \cdot 7}{1 \cdot 2} = 14/2 = 7$; c) $\frac{5 \cdot 2}{4} = 5/2 = 2,5$.

Neki broj u brojniku ne može se kratiti nekim brojem u nazivniku, ako su brojevi brojnika vezani zbrajanjem ili oduzimanjem sa susjednim brojevima.

7. a) $ad = bc$; b) $x = 2y$; c) $a = y(b + d)$.

8. a) 0,679; b) 8,75; c) 0,291.

Kad se traži da nešto treba "izračunati na 2 decimale", to znači da treba računati na 3 decimale, i onda, ako je to potrebno (prema pravilu, navedenom pod 2), izvršiti korekturu druge decimale. Dakle, uvijek treba računati jednu decimalu više nego što nam je potrebno za rezultat.

Decimalni broj množi se s decimalnim brojem tako da se najprije oba broja međusobno pomnože, a u rezultatu se odbroji toliko decimalnih mjesta koliko ukupno decimalnih mjesta imaju oba broja.

9. a) 20,0; b) 0,71; c) 289,22.

Broj se dijeli decimalnim brojem tako da decimalni broj pretvorimo u cijeli broj, množeći ga s 10 ili 100 ili 1 000, itd. Ako smo jedan od brojeva pomnožili sa 100, da bi postao cijeli broj, moramo i drugi pomnožiti sa 100. Drugim riječima, uvijek oba broja množimo istim brojem (10, 100, 1 000, itd.). Kad je to učinjeno, dijelimo ih normalnim postupkom.

10. a) 12,25; b) 0,624; c) 0,000009; d) 125; e) 1.

Vrijede pravila navedena pod 6.

11. a) $\frac{35}{6} = 5\frac{5}{6}$; b) $\frac{24}{6} = 4$; c) $\frac{4}{15}$.

Razlomak se množi s cijelim brojem tako da se brojnik pomnoži s tim brojem, a nazivnik ostaje nepromijenjen.

12. a) $\frac{15}{44}$; b) $\frac{1}{63}$; c) $\frac{6}{38} = \frac{3}{19}$.

Razlomak se množi s razlomkom tako da se brojnik pomnoži s brojnikom, a nazivnik s nazivnikom.

13. a) $\frac{2}{2} = 1$; b) $\frac{21}{42} = \frac{1}{2}$; c) $\frac{10}{12} = \frac{5}{6}$.

Razlomak se dijeli s razlomkom tako da se brojnik prvog razlomka pomnoži s nazivnikom drugog (i to postaje brojnik rezultata), a nazivnik prvog razlomka pomnoži se s brojnikom drugog razlomka (i to postaje nazivnik rezultata).

14. a) $\frac{8}{1} = 8$; b) $\frac{32}{3} = 10\frac{2}{3}$; c) $\frac{1}{4}$; d) $\frac{3}{40}$.

Razlomak se dijeli s cijelim brojem (ili cijeli broj s razlomkom) tako da cijeli broj zamislimo kao da je razlomak, tj. da u nazivniku ima 1. Nakon toga dijelimo prema pravilu navedenom pod 13.

15. a) $\frac{4+6}{12} = \frac{10}{12} = \frac{5}{6}$; c) $\frac{2-3}{9} = -\frac{1}{9}$;

b) $\frac{14+2}{16} = \frac{16}{16} = 1$; d) $\frac{24-14}{56} = \frac{10}{56} = \frac{5}{28}$.

Razlomak se zbraja s razlomkom ili odbija od razlomka tako da se najprije nađe zajednički nazivnik, tj. najmanji broj koji je djeljiv s jednim i drugim nazivnikom. Na primjer, zajednički nazivnik od 3 i 4 je 12, zajednički nazivnik od 8 i 16 je 16, itd. Račun nakon toga izvedemo tako da u brojniku prijašnji brojnik pomnožimo istim brojem kojim smo morali pomnožiti prijašnji nazivnik.

16. a) 4,00; b) 12,65; c) 80,12; d) 5,693; e) 1,26; f) 0,228.

Osnovni postupak prije vađenja korijena (bilo računom bilo iz tablica) sastoji se u tom da broj pod korijenom podijelimo u grupice od po dvije znamenke, počevši od decimalnog zareza prema lijevo i prema desno. Na primjer, broj 139,36, koji je pod korijenom, podijelit ćemo ovako 1 | 39, | 36. To je vrlo važno učiniti ispravno, jer npr. $\sqrt{1,63}$ nalazimo u tablici pod brojem 163, dok $\sqrt{16,3}$ nalazimo u tablici pod brojem 1 630!

17. a) 32; b) 20; c) 48; d) 12.

Prilikom računanja uvijek treba najprije obaviti operacije "višeg stupnja" (dakle množenje, dijeljenje), a nakon toga operacije "nižeg stupnja" (zbrajanje, oduzimanje). Međutim, ako je zagradom označeno da neka operacija nižeg stupnja ide zajedno, npr. $(16 + 8) \cdot 2$, onda treba posebno obaviti tu operaciju, pa tek onda izvršiti operaciju višeg stupnja.

18. a) $a^2 + 2ab + b^2$; b) $a^2 - 2ab + b^2$.

19. a) $\sqrt{25} = 5$; b) $4 \cdot 2 = 8$ (ili $\sqrt{64} = 8$); c) $\sqrt{16} = 4$;
d) $6 : 3 = 2$ (ili $\sqrt{4} = 2$)

Prilikom vađenja drugog korijena iz niza brojeva povezanih operacijama, ne smijemo vaditi korijen iz svakog broja posebno ako su među brojevima znakovi + ili −. Naprotiv, ako među brojevima stoje znakovi · ili :, možemo vaditi korijen iz svakog broja posebno. Siguran način izbjegavanja pogrešaka ovog tipa jest taj da najprije završimo sve operacije pod korijenom, pa tek onda izvadimo iz rezultata drugi korijen.

20. $\Sigma X^2 = 29$. Kazano riječima, ovaj zadatak glasi: koliko iznosi suma *kvadriranih* X-ova? No u zadatku b) piše $(\Sigma X)^2$, a to znači: koliko iznosi suma X-ova na kvadrat? U našem primjeru to je $(2+3+4)^2 = = 81$. Brkanjem ovih pojmova nastaju *najčešće pogreške* pri statističkom računanju kod početnika.

21. a) 2; b) 120; c) 1.

Znak ! iza nekog broja zove se "faktorijel" i znači da moramo izvršiti operacije množenja, koje se sastoje u tome da taj broj najprije pomnožimo s brojem koji je za 1 manji od njega, pa nakon toga s brojem koji je za 2 manji, itd. Na primjer, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Valja zapamtiti: $0! = 1$.

3. O OSNOVNIM POJMOVIMA VJEROJATNOSTI

Dugo sam se kolebao da li da u svojedobno prerađeno izdanje "Osnovnih statističkih metoda" uključim ili ne poglavlje o osnovnim pojmovima vjerojatnosti. Glavni faktor, koji je govorio *protiv* uključivanja tog poglavlja jest činjenica da je pojam vjerojatnosti prvenstveno matematički pojam i da su ga matematičari rješavali na matematički način. A statistički način mišljenja može se posve dobro usvojiti i bez poznavanja nekih osnovnih zakona iz područja vjerojatnosti. Točnije rečeno, ona vrsta vjerojatnosti s kojom konzument statistike dolazi u svom radu u kontakt, to je vjerojatnost koja proizlazi iz površine krivulje normalne raspodjele, i koja — iako se, naravno, radi o istom pojmu "vjerojatnosti" — nije sasvim u vezi s onim što se obično u statistici obrađuje u poglavlju posvećenom osnovnim pojmovima vjerojatnosti.

No odluka o tome da to poglavlje ipak bude uključeno u knjigu napokon je bila potaknuta time što su pitanja vjerojatnosti ovog tipa vrlo praktična pitanja, česta u nekim područjima svakodnevnog života, a pogotovu se često pojavljuju u različitim igrama "na sreću", kao što su igre kartama, kockama, tombole, lutrije, rulet i sl. Kako i u tom području mnogi ljudi nemaju minimalno predznanje o onome što bi u takvim situacijama trebali znati, odlučeno je da poglavlje ipak ostane u udžbeniku.

Za one čitaoce koji su više za to zainteresirani, dane su i neke formule, no ako ste "alergični" na matematičke formule, nemojte se njima baviti, već ih jednostavno preskočite i pročitajte samo nekoliko primjera koji vas zanimaju. Pročitajte ujedno i o nekim tipičnim pogreškama što ih ljudi često čine u vezi s vjerojatnošću, a koje su opisane na kraju poglavlja.

Ovo je na početku bilo potrebno reći jer je čitaocu u prvom poglavlju obećano da osnovne statističke metode može svladati s 4 osnovna računa. To je točno, i — kao što će se vidjeti — više od toga zaista nije potrebno.

Vrlo je teško reći kada su zapravo započela znanstvena istraživanja vjerojatnosti iz kojih su se kasnije razvila glavna područja statističkog rezoniranja i statističke metodologije. Znade se da su stari Egipćani već 3500 godina prije naše ere igrali neke igre slične našim igrama kockom. Isto se tako znade da je čovječanstvo u stoljećima koja su slijedila imalo već mnogo iskustva s hazardnim igrama kockama; ali pravilo da će kod ispravne igraće kocke svaka njezina strana pasti otprilike jednako puta čini se da je relativno novijeg datuma, tj. da datira negdje oko 1560. godine, kada je talijanski liječnik, profesor geometrije i strastveni kockar Girolamo Cardano (1501-1576) u knjizi "Liber de ludo alea" ("Knjiga o igrama kockom") napisao, govoreći o igraćoj kocki: "Svaka polovica stranica kocke pojavljuje se jednako često ... Na primjer, ja mogu jednako lako baciti kockom 1, 3 ili 5, kao što mogu baciti 2, 4 ili 6." Cardano je i izračunao da je vjerojatnost svake strane kocke 1/6 i upozorio je da to vrijedi samo kod ispravne, tj. "poštene" igraće kocke. (N a p o m e n a. Cardano se inače bavio i astrologijom, u koju je čvrsto vjerovao. Između ostaloga, prorekao je i dan svoje smrti, i kad je taj dan došao — da bi potvrdio točnost svog proročanstva — Cardano je počinio samoubojstvo! I tako je poslužio kao "školski primjer" za ono što se u znanosti naziva "proročanstvo koje samo sebe ispunjava": najlakše je naime dokazati vrijednost proročanstva tako da učinimo točno ono što nam je proročanstvom predviđeno.)

Cardanovo otkriće o vjerojatnosti pojavljivanja svake strane kocke vrlo se brzo proširilo među matematičarima (iako mu je knjiga tiskana oko 85 godina nakon smrti), i Galileo Galilei, koji je objavio svoja "Razmišljanja o igrama kockom" 1620. godine, već je posvetio dio svog rukopisa vjerojatnosti različitih ishoda ako se igra *dvjema* kockama. On kaže ovako: "Budući da igraća kocka ima šest strana, kada je bacimo, ona može pasti na bilo koju stranu ... Ali ako zajedno s prvom kockom bacimo i drugu kocku, koja također ima šest strana, možemo dobiti 36 različitih ishoda, jer svaka strana prve kocke može se kombinirati sa svakom stranom druge kocke ... što čini 6 puta 6, tj. 36 kombinacija."

Nekako u slično vrijeme (oko g. 1655) francuski matematičari Blaise Pascal i Piere de Fermat započeli su pismenu diskusiju o zakonima vjerojatnosti kod igraće kocke, potaknuti pismom što ga je Pascalu uputio kockar Chevalier de Méré. Taj je, naime, kockar zaradio mnogo novaca kladeći se da će kod 4 bacanja igraće kocke šestica pasti barem jedanput. On je pri tome rezonirao ovako: ako kocku bacim jedanput, vjerojatnost šestice iznosi 1/6, ako je bacim dvaput, vjerojatnost je dva puta veća, dakle 2/6, ako je bacim triput, vjerojatnost iznosi 3/6 (50% : 50%), a ako je bacim četiri puta, vjerojatnost je 4/6. Iako njegov račun *nije bio točan* (jer po toj logici vjerojatnost kod 6 bacanja bila bi 6/6, dakle 1, što znači apsolutnu sigurnost!), on je ipak zaradio na toj igri veliku svotu novaca, jer *stvarna* vjerojatnost da će od četiri bacanja šestica pasti bar jedanput iznosi 0,52, dakle i u ovom slučaju vjerojatnost dobitka je u njegovu korist. Nakon toga on je smislio novu igru: kladio se da će pomoću dvije kocke u 24 bacanja baciti barem jedanput sumu 12 (dvije šestice). Rezonirao je na isti (pogrešan) način: ako je kod jednog bacanja dvije kocke vjerojatnost dvije šestice 1/36 (što je točno), onda će kod dva bacanja vjerojatnost biti 2/36 (što više nije točno), kod 3 bacanja 3/36 ... itd., a kod 24 bacanja 24/36 ili 2/3, pa to, dakle, znači da je vjerojatnije da ću dobiti, nego izgubiti. Na njegovu nesreću *stvarna* vjerojatnost da će se u 24 bacanja pojaviti barem jedna dvanaestica iznosi oko 0,49, pa je prema tome njegova šansa za dobitak nešto *ispod* 50%. I tako je — igrajući veliki broj ovih igara — izgubio golem imetak,

te se sav očajan obratio za pomoć Pascalu.

(N a p o m e n a. Računi za točno izračunavanje tih vjerojatnosti ne pripadaju među osnovne zakone, koji će ovdje biti spomenuti, i nisu suviše jednostavni. No, ako nekoga ipak zanima kako se to može izračunati, neka pogleda tumačenje na kraju ovog poglavlja).

U ovom ćemo poglavlju samo letimično proći neke osnovne pojmove vjerojatnosti. Zakoni vjerojatnosti nisu uvijek posve jednostavni i razumljivi, i svakodnevno iskustvo i logika, koju u životu koristimo, često nisu u skladu sa zakonima koje nam daje statistika. No, naravno, "subjektivna vjerojatnost" — kako je zovu pojedini statističari — pripada među psihološke, a ne statističke pojmove, i stoga sa statističkom (ili matematičkom) vjerojatnosti katkada nema mnogo zajedničkoga. Ljudi, na primjer, u većini slučajeva vjeruju da će možda ipak dobiti glavni zgoditak na lutriji, ali praktički *ne vjeruju* da će im se dogoditi neka prometna nesreća — iako je vjerojatnost za nesreću, na žalost, mnogo veća nego vjerojatnost dobitka glavnog zgoditka na lutriji!

U ovom, a i u nekim drugim poglavljima, čitalac će naići na više primjera koji pokazuju kako čovjek katkada nije sposoban da ispravno prosudi neku vjerojatnost. Na primjer, kada bismo postavili pitanje kolika je vjerojatnost da u skupini od 50 ljudi nađemo barem dvojicu koji su rođeni *istog dana i istog mjeseca*, gotovo svatko će odgovoriti da je ta vjerojatnost iznimno mala. Ali, istina je upravo obratna: vjerojatnost da će se to dogoditi vrlo je velika, ona iznosi oko 97%! (Već kod skupine veličine 23 vjerojatnost iznosi 50%!)

No takvi primjeri spadaju već u "finese" teorije vjerojatnosti, i mi se njima uglavnom u ovoj knjizi nećemo baviti. Na kraju ovog poglavlja pokušat ćemo ipak protumačiti zašto je šansa u maloprije navedenom primjeru u vezi s rođendanom tako velika.

Koliko su često "rafinirani" zakoni vjerojatnosti, lijepo pokazuje primjer što ga na početku knjižice "Vjerojatnost i slučaj" ("Wahrscheinlichkeit und Zufall", München, 1973) daje autor Max Woitschach. Ta je njegova knjižica mali programirani udžbenik za upoznavanje zakona vjerojatnosti, i na prvoj stranici autor se obraća čitaocu ovim riječima: "Dragi čitaoče, biste li smatrali korisnim kada biste odmah na početku neke knjige mogli ustanoviti da li vam se isplati da tu knjigu čitate? Zbog toga mi dopustite da započnem jednim malim pitanjem:

Molim vas da zamislite kako neki čovjek 60 puta zaredom baci igraću kocku i da pri tome izbroji koliko je puta dobio svaki od šest brojeva na kocki. Kad je završio bacanje, on konstatira jedno od ovoga dvoga:

1. Šest brojeva nisu se pojavili svaki jednako puta, *usprkos tome* što su svi brojevi imali jednaku vjerojatnost pojavljivanja.
2. Šest brojeva nisu se pojavili svaki jednako puta *zato što* su svi brojevi imali jednaku vjerojatnost pojavljivanja."

Na idućoj stranici Woitschach kaže:

"1. Ako mislite da se svi brojevi nisu jednako puta pojavili *usprkos tome* što su svi imali jednaku šansu — vi obavezno trebate ovu knjigu čitati dalje, jer ćete naučiti da je praktički nemoguće da se kod većeg broja bacanja svaki broj pojavi jednako puta. Potpuno jednaka šansa...upravo uzrokuje *nejednaku* raspodjelu brojeva ...

2. Ako ste smatrali da se svi brojevi nisu jednako puta pojavili *zato što* su svi imali jednaku šansu, tada vi očito pripadate malom broju onih kojima su osnove teorije vjerojatnosti poznate, i pitanje je hoće li iduće stranice išta novo pridonijeti vašem znanju...." Eto, tako pita Woitschach u svom udžbeniku. A isto pitanje mogao bih postaviti i ja vama, čitaocima ovog udžbenika. Kako ste odgovorili na Woitschachovo pitanje?

Vjerojatno ste odgovorili odgovorom 1. (Ako jeste, neka vas to ni najmanje ne deprimira, jer i mnogi od onih koji dosta dobro znadu statistiku, daju također odgovor br. 1.)

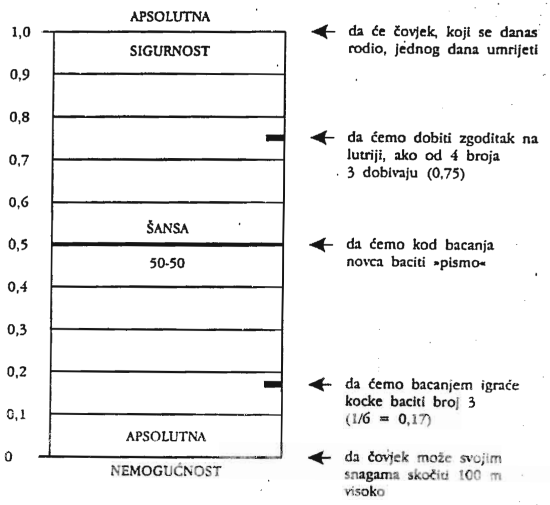
Prema tome, počnite s čitanjem koje će vas uvesti u zaista elementarne zakone vjerojatnosti.

3.1. NAJOSNOVNIJA PRAVILA

Evo nekoliko najosnovnijih pravila:

1. Ako je *potpuno sigurno* da će se nešto dogoditi, onda je vjerojatnost toga događaja maksimalna, i bilježi se s $p = 1$. Na primjer, potpuno je sigurno da će čovjek koji je danas rođen, jednog dana umrijeti. (Dolazi od latinskog "probabilitas" = vjerojatnost).

Ako je *potpuno sigurno* da se nešto *neće* dogoditi, vjerojatnost toga događaja nije nikakva, i bilježi se $p = 0$. Na primjer, potpuno je sigurno da čovjek ne može svojim snagama skočiti 100 metara visoko.



Slika 3.1. Svi slučajevi vjerojatnosti nalaze se između apsolutne sigurnosti ($p=1$) i apsolutne nemogućnosti ($p=0$)

sigurnosti ($p = 1$) i apsolutne nemogućnosti ($p = 0$)

Između apsolutne sigurnosti ($p=1$) i apsolutne nemogućnosti ($p=0$) nalaze se svi ostali slučajevi manje ili veće vjerojatnosti, kao što je to prikazano na slici 3.1.

2. Vjerojatnost da će se između N događaja, koji su jednako vjerojatni a međusobno nezavisni, dogoditi jedan određeni među njima jest $1/N$.

Na primjer, ako na igraćoj kocki postoje šest jednako mogućih rezultata, vjerojatnost da ćemo kockom baciti broj 3 je $1/6$ ili $p = 0,17$.

Ili, vjerojatnost da ćemo između 32 igraće karte izvući pikovu desetku $= 1/32$, tj. $p = 0,03125$ (nešto više od 3%).

3. Vjerojatnost da će se dogoditi *bilo koji od nekoliko* mogućih nezavisnih događaja *suma* je vjerojatnosti svakoga pojedinačnog događaja.

Na primjer: vjerojatnost p da će kod bacanja novca pasti glava *ili* pismo $= 1/2 + 1/2 = 1$. Razumljivo, jer nešto od toga dvoga *mora* pasti, i to je potpuno sigurno.

Ili, vjerojatnost da ćemo jednim bacanjem kocke baciti ili broj 5 ili broj 3 ili broj 2 je $1/6 + 1/6 + 1/6 = 3/6 = 0{,}5$.

4. Vjerojatnost da će se zajedno dogoditi dva ili više nezavisna događaja *produkt* je vjerojatnosti svakog od tih događaja.

Na primjer: vjerojatnost da ćemo kockom dva puta redom (ili da ćemo s dvije kocke istodobno) baciti broj 6 je $1/6 \cdot 1/6 = 1/36 = 0{,}028$ (ispod 3%).

Ili, vjerojatnost da ćemo pet puta redom baciti "pismo" (ili na 5 komada novca dobiti "pismo") $= 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 = 1/32 = 0,031$. To nam se, dakle, najvjerojatnije može dogoditi u prosjeku samo 3 puta u 100 pokušaja bacanja 5 komada novca (točnije: prosječno jedanput u 32 pokušaja).

Upravo na primjerima ove vrste može se lijepo uočiti čovjekova "pristranost" pri prosuđivanju određenih vjerojatnosti. Evo jednog zadatka koji će vas u to uvjeriti: Ako jedan komad novca bacimo 6 puta zaredom, koji je od donjih ishoda po vašem mišljenju najvjerojatniji? (G = "glava", P = "pismo"):

A:	G	P	P	G	P	G
B:	G	G	G	P	P	P
C:	P	P	P	P	P	P

Gotovo je sigurno da ste dali odgovor A. Taj odgovor dali ste zato, što iz iskustva znate da će se kod bacanja novčića najčešće nepravilno mijenjati bilo "glava", bilo "pismo". No točan odgovor na postavljeno pitanje glasi da je svaki od predložena tri ishoda jednako vjerojatan! Drugim riječima, za svaki od njih vrijedi navedeno pravilo da je vjerojatnost ishoda produkt vjerojatnosti svakog pojedinog događaja, pa prema tome vjerojatnost ishoda A, B, ili C iznosi $1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 = 1/64 = 0,0156$. (Ili, vjerojatnost je $0,5^6 = 0,0156$.)

Zašto ste se prevarili u odgovoru? (Ne zabrinjavajte se, jer će i mnogi iskusniji statističar "upasti u istu zamku"). Prvi nam se ishod čini najvjerojatnijim zato što će se kod šest bacanja novčića zaista najčešće događati da kao konačni ishod dobivamo čas "glavu" a čas "pismo", ali vjerojatnost nekog točno definiranog ishoda iznosi 0,0156. Kao što će se u poglavlju 7 vidjeti, kod bacanja jednog novčića 6 puta (ili 6 novčića jedanput, ali novčići u tom slučaju moraju biti numerirani) postoje 64 moguća ishoda. Samo jedan od njih 64 je ishod P P P P P P, i jedan je ishod G G G G G G, a svi ostali ishodi su ishodi sa barem jednim P među ostalim G ili barem jednim G među ostalim P. Najčešći i najvjerojatniji je ishod, u kojem će se pojaviti 3 puta P i tri puta G, ali to se može dogoditi u 20 različitih redoslijeda, a svaki od tih redoslijeda ima jednaku vjerojatnost, tj. vjerojatnost od P = 0,0156. Brojčano najviše ima ishoda sa "miješanim" G i P, i to nas dakako zavodi, pa kažemo da je ishod A najvjerojatniji. (U ovom poglavlju bit će još sličnih primjera).

Vjerojatnost da ćemo najprije kockom baciti broj 3, da ćemo nakon toga novčićem baciti "pismo" i napokon od 4 asa izvući "karo" asa, jednaka je:

$p = 1/6 \cdot 1/2 \cdot 1/4 = 1/48 = 0,02$.

Zakon multiplikacije može se lako protumačiti. Ako bacamo dvije kocke, onda one mogu pasti u ovih 36 kombinacija (I = prva kocka, II = druga kocka):

I	II	I	II	I	II	I	II	I	II	I	II
1	1	2	1	3	1	4	1	5	1	6	1
1	2	2	2	3	2	4	2	5	2	6	2
1	3	2	3	3	3	4	3	5	3	6	3
1	4	2	4	3	4	4	4	5	4	6	4
1	5	2	5	3	5	4	5	5	5	6	5
1	6	2	6	3	6	4	6	5	6	6	6

Kako vidimo, od ovih 36 kombinacija samo je jedna kombinacija 6 — 6; to znači da je njezina vjerojatnost 1/36, a to je $1/6 \cdot 1/6$.

N a p o m e n a. Do određene zbrke može doći i onda kada, na primjer, umjesto "bacanja 2 kocke" kažemo "bacanje jedne kocke dva puta". Malo prije smo izračunali da vjerojatnost za dobivanje dvije "šestice" iznosi $1/6 \cdot 1/6 = 1/36$. To znači da ćemo u 36 bacanja po *dvije kocke* u prosjeku jedanput dobiti dvije "šestice". Ali ako bacamo *jednu* kocku, moramo je baciti *dva puta* da bismo oponašali jednokratno bacanje dvije kocke; iz toga bi proizlazilo da kod bacanja jedne kocke možemo na 72 (a ne 36!) bacanja u prosjeku očekivati jedan slučaj pojave dvije "šestice" *uzastopce* (jer 72 bacanja jedne kocke je isto kao i 36 bacanja dvije kocke). Ali ipak nije tako, nego je jednu kocku dovoljno baciti u prosjeku 36 puta! No budući da se radi o jednom detalju, ako čitaoca baš naročito zanima zašto je to tako, neka pročita tumačenje na kraju ovog poglavlja.

Evo još jednog primjera: Vjerojatnost da će neki roditelji, koji planiraju četvoro djece, dobiti 4 kćeri, iznosi $1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 = 1/16$. To *ne znači* da ćemo u prosjeku na 16 brakova s djecom naći jedan brak s četvoro ženske djece, nego to naravno znači da ćemo na 16 *brakova s četvoro djece* naći u prosjeku 1 brak s 4 kćeri (a također i 1 brak s 4 sina, pa prema tome na 16 brakova s četvoro djece naći ćemo u prosjeku dva braka u kojima je sve četvoro djece istog spola.)

Nastavimo dalje našu diskusiju o različitim oblicima vjerojatnosti.

Oprez! Ako se zapitamo koja je vjerojatnost da će na jednoj kocki pasti broj 3, a na drugoj broj 4, čini se u prvi tren da je vjerojatnost također 1/36, tj. $1/6 \cdot 1/6$, ali to je tako samo onda ako *definiramo* točno situaciju: na *prvoj* kocki 3, a na *drugoj* kocki 4. Ako je *svejedno* na kojoj kocki padne 3, a na kojoj 4, onda — kako

vidimo iz tablice — postoje *dvije* mogućnosti na njih 36, i to: 3-4 i 4-3, pa, prema tome, vjerojatnost kombinacije 3-4 iznosi 1/36 + 1/36 = 1/18.

Koja je vjerojatnost da ćemo iz skupine od 52 igraće karte izvući karo kralja, srce desetku i tref asa? Oprez, to nije $1/52 \cdot 1/52 \cdot 1/52$, nego $1/52 \cdot 1/51 \cdot 1/50 = 1/132\,600$ — jer pošto smo izvukli prvu kartu, preostalo ih je još 51, itd. To će se dakle u prosjeku dogoditi 1 puta u 132 600 pokušaja!

5. Ako netko ima N mogućnosti da učini jedan zadatak, r mogućnosti da učini drugi zadatak, i p mogućnosti da učini treći zadatak, onda je broj svih mogućih kombinacija tih triju zadataka $N \cdot r \cdot p$.

Na primjer: ako jedna igraća kocka može pasti na 6 mogućih načina, a druga isto na 6 mogućih načina, onda obje mogu pasti u $6 \cdot 6 = 36$ različitih kombinacija.

Ili, ako netko može birati 10 vlakova na putu Zagreb-Rijeka, a pri povratku ponovno sve te vlakove, osim onoga kojim je stigao, onda on ima $10 \cdot 9 = 90$ mogućih kombinacija vlakova za odlazak i povratak zajedno.

Ili, ako u športskoj prognozi treba pogoditi rezultate 12 utakmica, od kojih svaka može imati tri ishoda (1, 2, 0), onda je mogući broj kombinacija:

$3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 531\,441$

Dakle, vjerojatnost da ćemo slučajno pogoditi iznosi $1/531\,441 = 0{,}0000019$ što praktički znači 2 na milijun.

Ili, ako se u jednom restoranu može dobiti ručak koji se sastoji od juhe, mesa, variva i kolača, a može se birati između 4 juhe, 3 vrste mesa, 5 vrsti variva i 4 vrste kolača, onda iz ovog menija možemo sastaviti $4 \cdot 3 \cdot 5 \cdot 4 = 240$ kombinacija.

Ako se brojevi automobila sastoje od 2 slova + 3 broja, a raspolažemo sa 22 slova i 10 znamenki (od 0 do 9), broj mogućih kombinacija iznosi $22 \cdot 22 \cdot 10 \cdot 10 \cdot 10 = 484\,000$. No budući da ne mogu sve tri znamenke biti 0, to taj ishod otpada, i to kod svih kombinacija, a njih ima $22 \cdot 22 = 484$. Dakle, $484\,000 - 484 = 483\,516$. — Možda je još i jednostavnije kazati ovako: postoji $22 \cdot 22 = 484$ kombinacije slova. Uz svaku kombinaciju može biti 999 brojeva (od 001 do 999). Dakle: $484 \cdot 999 = 483\,516$. — Budući da ovaj primjer približno odgovara našem načinu registriranja vozila, za naše prilike treba još uzeti u obzir mogućnost samo jednog slova uz 3 broja. Imamo dakle $22 \cdot 999 = 21\,978$. Iz toga proizlazi da postoji ukupno $483\,516 + 21\,978 = 505\,494$ moguće kombinacije. (Toliko za informaciju znatiželjnima, koji se pitaju koliko se automobila može registrirati na području Zagreba postojećim načinom registracije).

6. Ako imamo 4 broja, koliko postoji kombinacija kojima ih možemo poredati? Evo tih kombinacija (svaka kombinacija prikazana je u jednom stupcu):

1 1 1 1 1 1	2 2 2 2 2 2	3 3 3 3 3 3	4 4 4 4 4 4
2 2 3 3 4 4	1 1 3 3 4 4	1 1 2 2 4 4	1 1 2 2 3 3
3 4 2 4 2 3	3 4 1 4 1 3	2 4 1 4 1 2	2 3 1 3 1 2
4 3 4 2 3 2	4 3 4 1 3 1	4 2 4 1 2 1	3 2 3 1 2 1

Sve te moguće kombinacije mogu se izračunati ovako:

$4 \cdot 3 \cdot 2 \cdot 1 = 24$.

To se u statistici bilježi 4!, a čita se "četiri faktorijel". Prema tome, vjerojatnost da ćemo pogoditi neku od kombinacija, iznosi 1/24, ili $p = 0,042$.

Dakle, formula za izračunavanje broja svih mogućih kombinacija glasi jednostavno:

$$n! \tag{3.1}$$

Zamislimo da neka domaćica ima na večeri osmoro gostiju. Na koliko načina oni mogu sjediti oko stola? Prvi gost može se smjestiti na bilo koju od 8 stolica; kad je prvi gost smješten, drugi može sjesti na bilo koju od preostalih 7 stolica. Kada njih dva sjede, treći može sjesti na bilo koju od preostalih 6... itd., dakle za svih 8 gostiju postoji $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 8! = 40\,320$ kombinacija, što — kako kaže jedan autor — razjašnjava zašto to neku domaćicu može dovesti do potpune konfuzije!

7. Ako među n predmeta želimo ustanoviti koliko je mogućih permutacija za r tih predmeta, koristimo formulu:

$$\frac{n!}{(n-r)!} \tag{3.2}$$

Na primjer, ako je naša maloprije spomenuta domaćica zabrinuta samo kako će za stol smjestiti *četvoro* od ukupno 8 gostiju, onda ćemo to izračunati prema maloprije navedenoj formuli:

$$\frac{8!}{(8-4)!} = \frac{8!}{4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = 8 \cdot 7 \cdot 6 \cdot 5$$

(jer ostalo se krati!) $= 1\,680$.

Kako vidimo, sada ona ima mnogo manje posla nego prije.

Razumljivo je da se formula (3.2) može primijeniti i na prijašnji primjer, tj. kada je trebalo rasporediti svih 8 gostiju:

$$\frac{8!}{(8-8)!} = \frac{8!}{0!} = 8!,$$

i time smo ujedno dobili isto kao da smo računali prema formuli (3.1).

(N a p o m e n a. Neka vas ovaj račun ne zbuni. Ako ste pročitali prethodno poglavlje, onda znate da 0! iznosi 1.)

8. Ako nam međutim *nije važan redoslijed* ljudi ili stvari (u prijašnjim primjerima redoslijed je bio važan), račun je, naravno, opet drugačiji.

Primjer: Koja je vjerojatnost da između brojeva 1 do 10 izvučemo brojeve 3 i 5?

Kada bi se tražio točan redoslijed, onda bismo to računali prema formuli (3.2), tj. $\frac{10!}{(10-2)!} = 90$. Dakle, postoji 90 mogućih skupina od po dva broja, od kojih je samo jedna skupina koja se sastoji od 3 i 5.

Nama međutim redoslijed nije važan i time se vjerojatnost ishoda naravno povećava. Formula za izračunavanje za ove slučajeve glasi:

$$\frac{n!}{r!(n-r)!} \tag{3.3}$$

pri čemu n i r imaju ista značenja kao i u formuli (3.2). U našem slučaju s brojevima imamo dakle: $\frac{10!}{2!8!} = 45$.

Dakle, vjerojatnost da ćemo između brojeva 1 do 10 izvući brojeve 3 i 5 (ili bilo koja druga dva, unaprijed definirana broja) iznosi $1/45 = 0{,}022$.

(N a p o m e n a. Taj bismo primjer mogli izračunati i drugim načinom, tj. uz pomoć "zakona multiplikacije" i "zakona adicije" [vidi točke 3 i 4!]. Vjerojatnost da ćemo prvo izvući broj 3 iznosi 1/10; vjerojatnost da ćemo od preostalih 9 brojeva izvući broj 5 iznosi 1/9. Dakle, vjerojatnost za *oba* broja je $1/10 \cdot 1/9 = 1/90$. No isto takva vjerojatnost postoji i za drugi redoslijed, tj. prvo broj 5, a nakon toga broj 3. Stoga vjerojatnost da će [bez obzira na redoslijed] biti izvučeni brojevi 3 i 5 [točnije: 3 i 5 ili 5 i 3] iznosi $1/90 + 1/90 = 1/45$.)

Drugi primjer: U nekom gradu tramvajske karte nose 9 brojeva, a pri poništavanju tih karata (što vrše sami putnici u kolima), ponište se uvijek 4 broja. Budući da često valja mijenjati kombinaciju koja će 4 broja biti poništena (zbog kontrole putnika), koliko mogućih varijacija postoji?

Evo odgovora: $\frac{9!}{4!(9-4)!} = \frac{9!}{4!5!} = 126$.

(N a p o m e n a. U matematici se izraz, prikazan formulom [3.3] naziva "binomni koeficijent". Tablica T u Dodatku je tablica binomnih koeficijenata, iz koje se za određeni n i određeni x (u našoj formuli to su n i r) može odmah očitati gotov rezultat cijele formule. Također u Dodatku je i tablica S, tj. tablica faktorijela.)

Treći primjer: Koja je vjerojatnost da ćemo u nekoj igri pogađanja (npr. loto) od 49 brojeva pogoditi 6 brojeva (ne računajući "rezervni broj")?

Rješenje se sastoji u računanju svih mogućnosti od 6 brojeva među 49:

$$\frac{49!}{6!(49-6)!} = \frac{49!}{6!43!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6!} = 13\,983\,816.$$

Prema tome, vjerojatnost da ćemo pogoditi 6 brojeva između 49 više je nego neznatna, iznosi oko 1 : 14 milijuna!

Ovo su bili neki relativno jednostavni primjeri u vezi sa osnovnim zakonima vjerojatnosti. U životu katkada nailazimo na situacije u kojima je doista važno da što uspješnije prosudimo vjerojatnost nekog ishoda, a jedno od tih područja je *medicina*, i to u situacijama procjene vjerojatnosti neke *dijagnoze*.

Poslužit ćemo se jednim relativno jednostavnim primjerom, koji ćemo najprije rastumačiti bez ikakvih računa, tj. posve *logičkim* putem, a iza toga ćemo čitaoca, koji je ozbiljnije zainteresiran za to područje, uputiti na računski postupak. Pri tome će biti vidljivo, da se on *svodi na osnovne, nama već poznate zakone vjerojatnosti.*

Pretpostavimo da za neku određenu bolest postoji *potpuno siguran test*, tj. takav, koji u 100% slučajeva daje pozitivan rezultat, ako ta bolest zaista kod nekoga postoji. Međutim, taj isti test kod *zdravih* ljudi nije 100% pouzdan, već samo 99%, tj. kod zdravih ljudi on ipak u 1% slučajeva daje pozitivan rezultat, iako taj čovjek nije bolestan od te bolesti. Pretpostavimo nadalje, da se radi o nekoj rijetkoj bolesti, i da od nje oboljeva 1% populacije.

Ako se taj test rutinski primjenjuje na obaveznim pregledima pučanstva, i ako jednog dana pregledamo 1 000 stanovnika, prema učestalosti te bolesti u populaciji očekujemo da će među njima biti 1% bolesnih od te bolesti, a to je 10 ljudi. Od preostalih 990 pregledanih u 1% slučajeva taj će test lažno biti pozitivan, a to znači kod gotovo 10 njih (točnije: 9,9). Dakle, imat ćemo ukupno oko 20 *pozitivnih* dijagnoza. *Iz toga slijedi da će samo oko polovica dijagnoza biti točna — što zvuči dosta razočaravajuće, jer test nam na prvi pogled izgleda veoma siguran.*

A sada ćemo cijeli postupak provesti računski, i to uz pomoć naših osnovnih statističkih pravila o vjerojatnosti:

1. Vjerojantnost da će se istovremeno pojaviti oba slučaja, tj. vjerojatnost bolesti (ona iznosi 1%) i pozitivan rezultat testa (on je potpuno siguran kada bolest postoji, pa ta vjerojatnost iznosi 100%) (zakon multiplikacije):

$$0,01 \cdot 1,0 = 0,01$$

2. Vjerojatnost, da netko nije bolestan od te bolesti, ali ima pozitivan rezultat na tome testu (također zakon multiplikacije):

$$0,99 \cdot 0,01 = 0,0099$$

3. Vjerojatnost da bilo koja osoba (bez obzira ima li tu bolest ili ne) dobije pozitivan rezultat testa iznosi (zakon adicije):

$$0,01 + 0,0099 = 0,0199$$

4. Vjerojatnost da će neki pozitivni rezultat testa biti uzrokovan stvarnim oboljenjem od te bolesti, iznosi:

$$\frac{\text{vjerojatnost bolesti i istodobnog pozitivnog testa}}{\text{vjerojatnost pozitivnog testa}} = \frac{0,01}{0,0199} = 0,5025$$

Dakle, oko 50% je vjerojatno da osoba s pozitivnim testom ujedno boluje od te bolesti.

— — —

3.2. BAYESOVI STATISTIČKI PRINCIPI

U novije vrijeme u nekim se statističkim udžbenicima mogu naći odlomci o Bayesovim poučcima ili principima. Bayes je bio engleski svećenik, koji je živio pred 250 godina (!), i koji se pasionirano bavio problemima vjerojatnosti, te je pronašao neke zakone, koji se ponešto razlikuju od klasičnog pristupa pitanju vjerojatnosti. Konkretno, on je izradio matematičke postupke (formule) koji omogućuju *mijenjanje* vjerojatnosti nekog ishoda pod utjecajem *novih informacija*. Budući da Bayesov princip danas zauzima ključno mjesto u *teoriji odlučivanja*, postupak ćemo ukratko opisati na jednom jednostavnom primjeru.

Pretpostavimo da imamo skupinu od 50 muškaraca i 50 žena, i da znamo, da među muškarcima 30% njih ima svjetlu kosu, a među ženama 20%. (Vidi sliku 3.2 za daljnju diskusiju). Ako po čistom slučaju izaberemo bez vidne kontrole jednu od ovih 100 osoba, onda vjerojatnost da je ta osoba muškarac iznosi 50%. (To je tzv. "priorna" vjerojatnost). No ako nakon toga dobijemo novu informaciju, da izvučena osoba ima svjetlu kosu, vjerojatnost da je to muškarac *sada postaje nešto veća od* *50%*: jer muškaraca sa svjetlom kosom ima više od žena sa svjetlom kosom. To je "posteriorna" vjerojatnost, a njenu logiku možemo lako shvatiti prema slici 3.2, na kojoj vjerojatnosti izračunavamo iz *površine*: na slici vidimo da je vjerojatnost da između 100 ljudi slučajno izvučemo muškarca $P = 1/2$, a vjerojatnost da izvučemo muškarca svjetle kose je $1/2 \cdot 3/10 = 0,15$. Također, vjerojatnost da između 100 ljudi izaberemo ženu svjetle kose iznosi $1/2 \cdot 2/10 = 0,10$.

Među *svim* članovima grupe (muškarcima i ženama) vjerojatnost da se radi o osobi sa svjetlom kosom iznosi:

$$1/2 \cdot 3/10 + 1/2 \cdot 2/10 = 0,15 + 0,10 = 0,25.$$

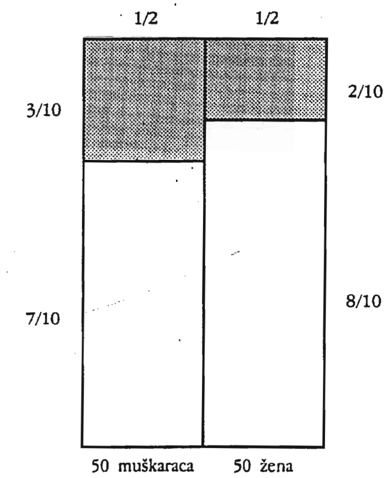
Dakle, vjerojatnost da je izabrana osoba sa svjetlom kosom *muškarac*, iznosi:

$$\frac{\text{vjerojatnost } \textit{muškarca} \text{ svjetle kose}}{\text{vjerojatnost } \textit{svih} \text{ učesnika svjetle kose}}$$

$$\frac{0,15}{0,15 + 0,10} = \frac{0,15}{0,25} = 0,60$$

Kako vidimo, "posteriorna" vjerojatnost sada je veća od "priorne", koja je iznosila 0,5.

Ako se nakon toga pojavi *nova* informacija, npr. da svi muškarci svjetle kose imaju modre oči, a od žena svjetle kose modre oči ima njih 50%, onda bi podatak da izvučena osoba ima modre oči, još više povisio vjerojatnost da se radi o muškarcu; prethodno izračunana "posteriorna" vjerojatnost sada naime postaje "priorna", pa računom, izvedenom po logici prethodnog računa, dobivamo da vjerojatnost da se radi o muškarcu, sada iznosi 75%. (Za podrobnije tumačenje Bayesova pristupa vidi B. Petz: Psihologijski aspekti teorije odlučivanja. — "Policija i sigurnost" br. 5, 1994.)



Slika 3.2. Primjer za Bayesovo izračunavanje vjerojatnosti (vidi tekst!)

3.3. PSIHOLOŠKI UZROCI NEKIM POGREŠKAMA KOD PROSUĐIVANJA VJEROJATNOSTI

Na kraju ćemo se osvrnuti na dvije zablude na koje možemo vrlo često naići čak i kod visokoobrazovanih ljudi ako nisu dovoljno upućeni u zakone vjerojatnosti:

1. Vrlo česta, gotovo univerzalna, zabluda u vezi s računom vjerojatnosti sastoji se u tom da — poznajući vjerojatnost pojavljivanja neke pojave — očekujemo da će se dugo nepojavljivanje te pojave "*nadoknaditi*" kasnijim pojavljivanjem. Tako npr. kod bacanja novca — ako netko deset puta zaredom igra na "pismo" i svih deset puta padne "glava" (a vjerojatnost za "pismo" iznosi $p = 1/2$), takav igrač ima intenzivan osjećaj da sada "*mora*" konačno doći "pismo", kad se tako dugo nije pojavilo! Međutim, novčić, rulet, kocka i slični rekviziti u "igrama slučaja" ***nisu živa bića i ne mogu pamtiti***, pa je, prema tome, vjerojatnost pojavljivanja nekog ishoda uvijek jednaka, ***potpuno bez obzira na to što je prethodilo.*** Dakle, u našem slučaju, da je i dvadeset puta prije toga pala "glava", kod 21. bacanja vjerojatnost pojavljivanja "pisma" jednaka je kao i kod prvog, tj. $p = 1/2$.

Godine 1913. u jednoj igračnici u Monte Carlu "crno" na ruletu (vjerojatnost da će izići "crno" iznosi 0,5) izišlo je 26 puta zaredom. Vjerojatnost takvog ishoda je $0,5^{26} = 0,0000000149$, ili — preračunato — jedanput u više od 67 milijuna okretaja.

Na pogrešan zaključak "nadoknađivanja" (ili — kako to u statistici nazivaju — "sazrijevanja šanse") možemo doći čak i kada *poznajemo* neke zakone vjerojatnosti (možda bi se čak moglo reći ***baš zato*** što poznajemo neke zakone vjerojatnosti). Evo primjera: jedan je čovjek imao sedmero sinova i očekivao je osmo dijete. On je dokazivao da će osmo dijete gotovo sigurno biti djevojčica, jer — rekao je — vjerojatnost da će netko imati 8 sinova zaredom je $0,5^8 = 0,0039$ (dakle oko 4 slučaja na 1000 brakova s osmero djece), pa je, prema tome, veća vjerojatnost da ću dobiti kćer, jer je kod nje vjerojatnost 0,5. No, međutim, ta vjerojatnost od 0,0039 vrijedi za svih 8 slučajeva zajedno; ali ako netko već ima 7 sinova, kod osmog djeteta vjerojatnost za muško ili žensko dijete jednaka je kao i kod svakog novog djeteta, tj. 0,5 za muško dijete, i 0,5 za žensko.

(N a p o m e n a. Ne zaboravimo da bi svaki drugi *unaprijed definirani* redoslijed, npr. sin, kći, sin, sin, kći, sin, kći, kći, imao jednako tako malu vjerojatnost, tj. 0,0039.)

Drugi vid te iste predrasude sastoji se u tom da vjerujemo da ishod koji se upravo dogodio nakon toga ima "neko vrijeme" manju vjerojatnost da se pojavi. Tako, na primjer, ako smo iz kutije sa žetonima izvukli, pretpostavimo, broj 45, stavili ga natrag i žetone dobro izmiješali, imamo "osjećaj" da je kod drugog izvlačenja vjerojatnost za svaki drugi pojedinačni broj veća nego ponovno za broj 45!. To naravno nije točno. Identična je pojava vjerovanje, ako se neki ishod nije dugo pojavio, da je sve veća šansa za njegovo pojavljivanje. U spomenutom slučaju u Monte Carlu 1913. godine, velik broj igrača počeo je stavljati velike svote na "crveno", očekujući da se "crveno" konačno mora pojaviti. (U toj igri kasino je zaradio nekoliko milijuna franaka!) Slična pojava opažena je kod pilota u 2. svjetskom ratu, kada su već imali određeni "staž" u zračnim borbama; izjavljivali su da je sada "na njih došao red" da budu oboreni, tj. da je sada vjerojatnost obaranja veća nego kad su kretali na prvi zadatak (iako je objektivna vjerojatnost bila veća kod prvih letova zbog njihova manjeg iskustva).

Slična je tome i predrasuda prema nekoj određenoj "pravilnosti" u rezultatima. Rijetko će tko u prognozi lota prognozirati recimo brojeve 1, 2, 3, 4, 5, 6, iako je za pojavljivanje te kombinacije *jednaka* vjerojatnost kao i za bilo koju drugu kombinaciju od 6 brojeva, npr. za kombinaciju 4, 13, 28, 32, 45, 47. Danas već postoji dosta literature o tzv. "subjektivnoj vjerojatnosti", tj. o tome kako se čovjek ponaša u različitim situacijama, u kojima jedino čisti slučaj igra ulogu, ili u kojima sudjeluje i slučaj i čovjekova vještina. Tako je npr. poznato da se većina ljudi kod igranja neke igre čistog slučaja ponaša kao da ima mogućnost utjecanja na rezultat: ljudi kod bacanja igraće kocke većinom unaprijed "prognoziraju" koji će broj baciti, a mnogo manje su skloni na oklade ove vrste nakon što su već bacili kocku (ali još nisu vidjeli rezultat). Ili, poznato je da ako želimo kockom baciti broj 1, da ćemo zamahnuti "mekše" i "nježnije" nego ako želimo baciti broj 6! Također je poznato da smo kod ishoda, koji ovise djelomično o slučaju, a djelomično o našoj vještini i sposobnosti, skloni da svoje uspjehe pripisujemo korištenju vještine, a neuspjehe utjecaju slučaja; neki psiholozi misle da su to glavni razlozi zašto su bridž i poker tako popularne i obljubljene kartaške igre. Ako imamo ispunjen jedan listić lota, i ako nas netko upita je li veća vjerojatnost da ćemo na lotu pogoditi svih 6 brojeva ili da ćemo stradati u automobilskoj nesreći, većina nas je sklona tvrditi da je šansa za dobitak na lotu znatno veća. (Čitalac, koji se zanima za ove subjektivne aspekte vjerojatnosti, neka prolista knjigu Dowiera i Leftera "Risk and chance", vidi u popisu literature).

2. Druga predrasuda sastoji se u tome da često smatramo da će se kod ***velikog broja pokušaja*** "jednake šanse" izjednačiti po broju ishoda. Ako, recimo, dvojica ljudi igraju igru bacanja novca, jedan igra na "pismo" a drugi na "glavu", onda je moguće — po tom pogrešnom stanovištu — da u 10 bacanja ne ispadne točno 5 puta glava i 5 puta pismo, ali u, na primjer, 10 000 bacanja broj glava i pisama morao bi se praktički izjednačiti.

No najvjerojatnije će se dogoditi ***upravo obratno***! Drugim riječima, ***apsolutna*** razlika između takmaca vjerojatno će postajati ***sve veća***, a smanjivat će se jedino razlika u ***proporciji***, tj. što je veći broj bacanja, to će postotak ishoda "glava" i "pismo" biti sve bliži odnosu 50% : 50%. Uvjerimo se u to na jednom primjeru: ako od 10 bacanja novčića četiri puta padne "glava", onda je proporcija tog ishoda $4/10 = 0,40$ (40%), a ***razlika*** od očekivanog ***broja*** je 1 (5-4). Ako novčić bacimo dvjesto puta, mogli bismo, recimo, dobiti sto petnaest puta "pismo", što u proporciji znači 57,5% (prema očekivanih 50), a apsolutna razlika iznosi 15 (115-100). Kod 1000 bacanja može se dogoditi da imamo 520 "pisama" i 480 "glava". Proporcija "pisma" sada iznosi 0,52, apsolutna razlika je 20. Dakle, povećavanjem broja pokušaja ***proporcija*** će se — uz povremene varijacije — postupno približavati teoretski očekivanom odnosu 50 : 50, ali ***apsolutne*** razlike od očekivane frekvencije mogu biti sve veće i veće. Iz toga proizlazi da onaj koji na kraju neke duge igre gubi, lako može izgubiti mnogo više, nego onaj koji gubi na početku igre.

— — —

I na kraju poglavlja dat ćemo najprije rješenje problema s istodobnim rođendanom. Među 50 ljudi, prvi od njih ima mogućnost 49 usporedbi s drugima u datumu rođenja, drugi ima 48 usporedbi, treći 47, četvrti 46 ... itd., dakle šansa za pronalaženje istog datuma rođenja je vrlo velika: $49 + 48 + 47 + 46 + \ldots + 1 =$ $= 1225$ usporedbi. U 1225 mogućih usporedbi *vrlo je vjerojatno* da će se naći isti datumi rođenja, jer godina ima samo 365 dana.

Inače, formula kojom se to izračunava glasi:

$$1 - 365/365 \cdot 364/365 \cdot 363/365 \ldots 316/365 = 1 - 0,03 = 0,97.$$

Dakle, vjerojatnost je 97%.

(N a p o m e n a. U formuli smo operaciju množenja proveli — kako se vidi — pedeset puta [tj. od 365 pa sve do 316], jer računamo za 50 ljudi. Ako je u grupi npr. 30 ljudi, onda treba množiti samo 30 puta.)

I evo, napokon, formule kojom se izračunava vjerojatnost da će se u nekoj igri slučaja kod različitog broja pokušaja neki određeni rezultat pojaviti barem jedanput:

$$1 - q^m. \tag{3.4}$$

U toj je formuli q = vjerojatnost da se taj rezultat *neće* pojaviti, a m = broj pokušaja.

U prvom slučaju s Méréovim kockama (barem jedna šestica u 4 bacanja) vjerojatnost iznosi:

$$1 - (5/6)^4 = 1 - 0,48 = 0,52.$$

U drugom slučaju (barem jedna dvanaestica u 24 bacanja po dvije kocke) vjerojatnost je:

$$1 - (35/36)^{24} = 1 - 0,51 = 0,49.$$

Ovaj jednostavan račun lako je primijeniti i na druge situacije. Ako se npr. zapitamo koja je vjerojatnost da ćemo između 52 igraće karte u 10 pokušaja izvući karo asa, onda račun pokazuje da vjerojatnost iznosi

$$1 - (51/52)^{10} = 0,18.$$

U 100 pokušaja vjerojatnost iznosi već 0,857, u 200 pokušaja 0,979, a u 300 pokušaja 0,997. Koliko treba pokušaja pa da vjerojatnost bude 1? *Praktički* možemo biti sigurni da bismo na nekoliko desetaka hiljada pokušaja izvukli barem jedanput karo asa, ali *teoretski* vjerojatnost postaje *sve veća*, ali nikada ne dostiže 1! (Kada biste to provjeravali računalom koje raspolaže mogućnošću golemog broja decimala, mogli biste se uvjeriti da biste dobivali rezultate sa samim devetkama, i to u beskonačnost).

— — —

I na potpunom kraju ovog poglavlja, evo još jedne napomene, ali neka ju čitaju samo "cjepidlake". Napomena se odnosi na jedan dio primjera iz ovog poglavlja, tj. na problem "36 bacanja po 2 kocke je isto kao 72 bacanja jedne kocke. Iako je ta tvrdnja potpuno točna, ipak treba upozoriti da se "72 bacanja jedne kocke"

ne smije shvatiti *doslovno*, jer mi imamo 36 serija od po dva rezultata i sa — 36 bacanja *jedne* kocke! Zvuči malo čudno, zar ne?

No pogledajmo u čemu je stvar. Ako 36 puta bacimo jednu kocku, to je naravno samo 36 rezultata, ali to *nije* — kako bi se činilo — samo "18 puta po dva rezultata", već je to 36 puta po dva rezultata. Naime, u nekom nizu brojeva, koji se kreću recimo od 1 do 36, *parovi* rezultata su 1-2, 3-4, 5-6 itd, *ali također* i brojevi 2-3, 4-5, 6-7, itd. A ako izračunate sve moguće parove rezultata brojeva 1-36, lako ćete ustanoviti da sa 36 bacanja imamo točno 36 serija od po 2 bacanja. Pogledajmo ponovno tablicu o bacanju dvije kocke, koja je prethodno bila spomenuta: ako sve brojeve, koji su u tablici, ne gledamo kao skupine od po 2 broja, već ih napišemo redom, dobit ćemo:

1, 1, 1, 2, 1, 3, 1, 4, 1, 5, 1, 6, 2, 1, 2, 2, 2, 3, 2, 4, 2, 5, 2, 6, 3, 1, 3, 2, 3, 3, 3, 4, 3, 5, 3, 6, 4, 1, 4, 2, 4, 3, 4, 4, 4, 5, 4, 6, 5, 1, 5, 2, 5, 3, 5, 4, 5, 5, 5, *6*, *6*, 1, 6, 2, 6, 3, 6, 4, 6, 5, *6*, *6*.

i u tom nizu možemo otkriti *dva* slučaja *dviju šestica* među 72 broja (otisnuti su kurzivom). A to znači prosječno *jedan* slučaj na 36 brojeva.

Ova diskusija potrebna je zbog slučajeva u kojima se ide na proračunavanje vremena, prosječno potrebnog da bi se neki rijetki slučaj dogodio. Najpoznatiji primjer za to je već spomenut primjer iz 1913. godine kada je na jednom ruletu "crno" izašlo 26 puta zaredom. Lako je izračunati koliko je gotovo neizmjerno mala vjerojatnost da se to dogodi: budući da su mogući ishodi za "crno" ili "crveno" jednaki, to je vjerojatnost ishoda "crno" $p = 0,5$, pa iz toga slijedi da vjerojatnost 26 uzastopnih ishoda "crno" iznosi $0,5^{26} = 0,0000000149$, ili poprilici 1 prema više od 67 milijuna.

Iz toga nam se lako dogodi ovaj zaključak: ako imamo jedan rulet, onda bi se takav ishod (od serije od 26 "crno") mogao prosječno očekivati na oko 67 miliona serija od 26 okretanja ruleta. No upravo u takvom zaključku je pogreška. Kada bi to bilo točno, onda bi (prema izvještajima stručnjaka, koji tvrde da se u igračnici rulet prosječno dnevno zavrti oko 480 puta), to trajalo nekoliko tisuća godina! No kao što smo već pokazali u primjeru s bacanjem kocke, u 67 milijuna *pojedinačnih* okretanja ruleta mi već imamo 67 milijuna *serija* od po 26 ishoda: od 1. do 26. ishoda, od 2. do 27., od 3. do 28. itd. Prema tome, potrebno vrijeme je 26 puta kraće.

4. MJERE CENTRALNE TENDENCIJE

U svakodnevnom se životu često služimo izrazom "prosjek", "prosječan" i sl.: govorimo o čovjeku "prosječne" visine, o "prosječnoj" cijeni neke namirnice na tržnici, o "prosječnom" broju posjetilaca u nekom kinu ili na nekom stadionu, o "prosječnoj" plaći, itd. Tim izrazom obično mislimo na neku "srednju" vrijednost, koja najbolje "reprezentira" onu pojavu o kojoj dajemo mišljenje, tj. mislimo na vrijednost oko koje se obično kreće najviše rezultata u toj pojavi. Budući da svi uglavnom znamo da se takozvana aritimetička sredina računa tako da se svi rezultati zbroje i podijele ukupnim brojem rezultata, to — govoreći o "prosjeku" — obično mislimo na aritmetičku sredinu. No to ne mora uvijek biti tako, tj. ako nas npr. zanima prosječna cijena neke namirnice, onda imamo bolju informaciju ako znamo koja je ***najčešća*** cijena te namirnice na tržnici, a ako nas zanima prosječan broj djece u obiteljima nekog grada, onda je čak logičnije da to izrazimo *najčešćim* brojem djece, jer prosjek je često decimalni broj, i pomalo je i smiješno kazati da prosječna obitelj ima — recimo — 2,4 djece!

"Prosječan", dakle, ne znači uvijek prosječan u smislu aritmetičke sredine. Stoga ćemo u ovom poglavlju obraditi neke od najvažnijih mjera "prosjeka" i upozoriti na razlike među njima i na situacije u kojima je pojedina od njih relativno najprikladnija.

4.1. ARITMETIČKA SREDINA

Kao što smo već rekli, jedna od najčešćih i najpoznatijih mjera "prosjeka" je aritimetička sredina, i ona ujedno pripada u vjerojatno jedan od najčešće izvođenih računa za statističke potrebe.

Osnovna formula za izračunavanje aritmetičke sredine glasi:

$$\text{ARITMETIČKA SREDINA} = \frac{\text{SUMA SVIH REZULTATA}}{\text{BROJ REZULTATA}}$$

što se statističkim simbolima piše:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{N},$$

pri čemu $\overline{X}$ znači aritmetičku sredinu, X_1, X_2 ... itd., jesu rezultati od prvog do zadnjeg, a N = broj rezultata.

Skraćeno se ta formula piše:

$$\overline{X} = \frac{\Sigma X}{N}, \qquad (4.1)$$

pri čemu Σ (čitaj sigma) znači sumu.

Ako smo izvršili 10 mjerenja pulsa nekog čovjeka i dobili ove vrijednosti:

$$63, 60, 68, 62, 70, 62, 63, 61, 65, 64,$$

onda ćemo prema formuli (4.1) izračunati aritmetičku sredinu:

$$\overline{X} = \frac{63 + 60 + 68 + 62 + 70 + 62 + 63 + 61 + 65 + 64}{10} = \frac{638}{10} = 63,8.$$

Međutim, kad imamo velik broj rezultata, izračunavanje aritmetičke sredine na ovaj način bilo bi vrlo dugotrajno i naporno, i zato se u tom slučaju služimo drugim, kraćim postupcima. Osnova se tih postupaka sastoji u tome da se rezultati grupiraju u "razrede", te su svi rezultati unutar jednog razreda reprezentirani *jednim* rezultatom, tj. sredinom razreda. (Osim toga, grupiranje rezultata u "razrede" neophodno je kod *grafičkog* prikazivanja rezultata.)

Primjer. Uzmimo da smo na jednom čovjeku izvršili 50 mjerenja vremena reakcije i da smo dobili ove rezultate (u tisućinkama sekunde):

196	173	186	189	173	165	167	160	140	174	180	151
157	164	154	169	190	180	163	157	169	167	165	160
177	165	157	177	159	175	166	173	185	177	184	183
162	192	174	162	165	172	158	169	146	170	171	169
168	153.										

Ako zbrojimo sve te rezultate i podijelimo njihovim brojem (prema formuli 4.1), dobit ćemo:

$$\overline{X} = \frac{8\,458}{50} = 169,16.$$

Želimo li aritmetičku sredinu izračunati kraćim postupkom, treba najprije rezultate grupirati u tzv. "razrede". Najčešće se broj razreda kreće između 10 i 20, ali kao opće pravilo može poslužiti ovaj princip: što je broj mjerenja manji, treba i broj razreda biti manji i obratno.

Budući da razredi moraju biti *jednaki* po veličini, tj. svaki razred mora obuhvatiti jednako velik *interval*, to ćemo najlakše naći koliko nam otprilike treba iznositi interval razreda, ako "raspon" rezultata podijelimo sa željenim brojem razreda. Rasponom nazivamo razliku između najvećeg i najmanjeg rezultata. U našem slučaju raspon je $196 - 140 = 56$.

Uzmimo da smo odlučili raditi s 12 razreda. Dakle, interval će biti: $56/12 = 4{,}67$, što zaokružujemo na 5. Ako odlučimo započeti recimo sa 140 (a možemo započeti

i nekim drugim brojem koji je ispod najnižeg rezultata), naš 1. razred obuhvatit će sve rezultate od 140 do 144 (interval je 5, a ne 4, kako se u prvi trenutak čini, jer se u tom razredu nalaze rezultati 140, 141, 142, 143 i 144, a to je 5 jedinica), 2. razred rezultate od 145 do 149, itd., kao što je učinjeno na donjoj tablici.

Razred
140 - 144
145 - 149
150 - 154
155 - 159
160 - 164
165 - 169
170 - 174
175 - 179
180 - 184
185 - 189
190 - 194
195 - 199

Kao što se vidi, svaki idući razred počinje za jednu jedinicu više nego što prijašnji razred završava. To je potrebno da bi tako bilo točno definirano u koji razred spada pojedini rezultat. Jer, kad bismo npr. imali jedan razred 140 - 145, a idući razred 145 - 150, onda ne bismo znali u koji razred treba smjestiti rezultat 145! No kako u nekim statistikama nalazimo i tako raspoređene razrede, da donja granica nekog razreda počinje istim brojem kojim je označena gornja granica prethodnog razreda, treba ipak — da bi se izbjegla konfuzija — spomenuti ova dva pravila:

1. Ako pri određivanju granica razreda uzimamo *istu točnost* kojom su izvršena mjerenja (npr. cijeli brojevi, ako je mjerenje izraženo samo cijelim brojevima), onda idući razred treba započeti za jednu jedinicu mjerenja više nego što je prijašnji razred završio.

2. Ako granice razreda postavljamo na *veću točnost* od one kojom je izvršeno mjerenje (npr. mjerenje je izvršeno u cijelim brojevima, a granice razreda postavljamo u decimalama), onda je potrebno da donja granica idućeg razreda bude jednaka gornjoj granici prethodnog razreda.

Ako je mjerenje izvršeno na točnost od jedne jedinice, onda su primjeri za oba pravila ovi:

Prvo pravilo	Drugo pravilo
140 - 144	139,5 - 144,5
145 - 149	144,5 - 149,5
150 - 154	149,5 - 154,5
155 - 159	154,5 - 159,5
itd.	itd.

Ako je mjerenje provedeno na točnost od 0,5 jedinica, primjeri za oba pravila su ovi:

140 - 144,5	139,75 - 144,75
145 - 149,5	144,75 - 149,75
150 - 154,5	149,75 - 154,75
155 - 159,5	154,75 - 159,75
itd.	itd.

Kako vidimo, kod 1. pravila postavljali smo granice razreda na istu točnost kojom je izvršeno mjerenje, a kod 2. pravila na veću točnost.

Zapravo, "prave" granice razreda 140 - 144 su ustvari 139,5 - 144,5, jer već kod *mjerenja* svaki rezultat veći od 139,5 i manji od 144,5 registriramo kao neki rezultat između 140 i 144.

Kad smo na opisani način dobili 12 razreda, unijet ćemo svaki rezultat u odgovarajući razred, i to tako da ćemo rezultate bilježiti okomitim crticama, a kad se sakupe 4 takve crtice, petim ćemo rezultatom prekrižiti tu grupu. Na taj način lako ćemo poslije — brojanjem skupina od po 5 rezultata — izbrojiti sve rezultate.

Ako rezultate našeg primjera unesemo u postavljene razrede, dobivamo ovaj rezultat:

Razred		Frekvencija rezultata
140 - 144	|	1
145 - 149	|	1
150 - 154	|||	3
155 - 159	𝍸	5
160 - 164	𝍸 |	6
165 - 169	𝍸 𝍸 ||	12
170 - 174	𝍸 |||	8
175 - 179	||||	4
180 - 184	||||	4
185 - 189	|||	3
190 - 194	||	2
195 - 199	|	1
		Ukupno 50

Pri unošenju rezultata u tablice treba biti osobito savjestan, jer naknadno možemo kontrolirati samo jesmo li sve rezultate registrirali (suma frekvencija = N), ali ne možemo kontrolirati jesmo li svaki rezultat unijeli u *odgovarajući* razred.

Kako iz raspodjele rezultata razabiremo, postoji određena "tendencija" da se rezultati grupiraju oko jedne vrijednosti koja je nekako po sredini svih razreda (u našem primjeru razred 165 - 169), dok su rezultati prema jednom i drugom ekstremu ove razredne ljestvice sve rjeđi.

Pošto smo rezultate na ovaj način unijeli u razrede, možemo pristupiti izračunavanju aritmetičke sredine.

To možemo učiniti tako da sredinu (X) svakog razreda pomnožimo frekvencijom pojedinog razreda (f) i da sumu tih umnožaka podijelimo brojem rezultata $\left(\frac{\Sigma fX}{N}\right)$. Množeći sredinu svakog razreda s frekvencijom toga razreda, mi smo zapravo *zbrajali* sve rezultate u tom razredu. Na primjer, u razredu 155 - 159 imamo

5 rezultata. Kako vidimo, mi svih 5 rezultata u ovom slučaju identificiramo s *jednom* vrijednosti, tj. sa sredinom razreda. Prema tome, možemo zamisliti kao da u tom razredu imamo ovih 5 rezultata: 157, 157, 157, 157, 157. Sumu tih rezultata možemo dobiti tako da ih zbrojimo, a i tako da sredinu razreda pomnožimo sa frekvencijom: 157 puta 5.

Prema tome, formula za izračunavanje aritmetičke sredine po ovom postupku glasi:

$$\overline{X} = \frac{\Sigma fX}{N}. \tag{4.2}$$

Ta formula zapravo predstavlja isti princip kao i naša osnovna formula $\left(\overline{X} = \frac{\Sigma X}{N}\right)$, jer se i ona — kako vidimo — svodi na to da zbrojimo rezultate i podijelimo ih s brojem rezultata.

Naš bismo primjer prema tome izračunali ovako:

Sredina razreda (X)	frekvencija (f)	(fX)
142	1	142
147	1	147
152	3	456
157	5	785
162	6	972
167	12	2004
172	8	1376
177	4	708
182	4	728
187	3	561
192	2	384
197	1	197
		$\Sigma fX = 8460$

$$\overline{X} = \frac{\Sigma fX}{N} = 8460/50 = 169,2.$$

Međutim, i u ovom postupku moramo često računati s velikim brojevima, što možemo izbjeći ako u svrhu olakšavanja računanja odredimo jednu "provizornu" ili "privremenu" aritmetičku sredinu, pa računamo samo koliko ostali rezultati *odstupaju* od te sredine, i onda provizornoj aritmetičkoj sredini dodamo prosjek svih odstupanja. Na primjer, ako imamo ovih deset rezultata:

$$5,\ 6,\ 5,\ 5,\ 3,\ 4,\ 5,\ 7,\ 5,\ 10,$$

pa uzmemo 5 kao provizornu aritmetičku sredinu, onda imamo ova odstupanja od te sredine:

$$0,\ +1,\ 0,\ 0,\ -2,\ -1,\ 0,\ +2,\ 0,\ +5$$

Suma ovih odstupanja iznosi +5, a njihov prosjek +5/10 = +0,5.

Taj rezultat dodamo provizornoj aritmetičkoj sredini da bismo tako dobili pravu aritmetičku sredinu:

$$\overline{X} = 5 + 0,5 = 5,5.$$

Tom istom logikom možemo jednaki postupak primijeniti kod naših rezultata raspoređenih u razrede.

	X	f	odstupanja	odstupanja puta frekvencija
	142	1	-25	-25
	147	1	-20	-20
	152	3	-15	-45
	157	5	-10	-50
	162	6	-5	-30
$\overline{X}_{\text{pr}}$	$\boxed{167}$	12	0	-170
	172	8	5	40
	177	4	10	40
	182	4	15	60
	187	3	20	60
	192	2	25	50
	197	1	30	30
				+280
				280 - 170 = +110

Kao provizornu srednju vrijednost $\overline{X}_{\text{pr}}$ izabrali smo 167 (a mogli smo — naravno — izabrati i bilo koji drugi razred). Za svaki razred izračunali smo koliko odstupa od provizorne sredine, i ta smo odstupanja pomnožili s frekvencijama (jer u svakom razredu nemamo samo jedno odstupanje, već f odstupanja!). Suma svih odstupanja iznosi 280 - 170 = +110, a prosjek tih odstupanja je $+110/50 = +2,2$.

Prema tome, aritmetička sredina bit će 167 + 2,2 = 169,2. Iz toga slijedi da se $\overline{X}$ može izračunati i prema formuli:

$$\overline{X} = \overline{X}_{\text{pr}} + \frac{suma\ odstupanja}{N} \tag{4.3}$$

I, konačno, *najkraći*, zato i najčešće, upotrebljavni postupak sastoji se u tome da računanje *još više skratimo*: umjesto da odstupanja računamo u *apsolutnim* razlikama, mi ih računamo u *jedinicama intervala*, tj. računamo "odstupanje za 1 interval", "za 2 intervala", "za 3 intervala", itd., a sve ostalo računamo kao i u prošlom primjeru. Budući da smo odstupanja računali u *intervalima* (dakle svako odstupanje dijelili smo s intervalom:

$\frac{X - \overline{X}_{\text{pr}}}{i}$), moramo konačni prosjek odstupanja *pomnožiti s intervalom* da bismo dobili realni broj, koji treba dodati provizornoj aritmetičkoj sredini. Prema tome, formula za izračunavanje aritmetičke sredine, po skraćenom postupku, imat će ovaj oblik:

$$\overline{X} - \overline{X}_{\text{pr}} + \left(\frac{\Sigma f x'}{N} \cdot i\right), \tag{4.4}$$

pri čemu $\overline{X}_{\text{pr}}$ znači provizornu aritmetičku sredinu, a x' je inervalna udaljenost pojedinih razreda od provizorne aritmetičke sredine:

$$\left(x' = \frac{X - \overline{X}_{\text{pr}}}{i}\right)$$

Za lakšu orijentaciju, čitav ćemo "skraćeni postupak" izračunavanja aritmetičke sredine opisati u ovih 6 operacija:

1. Naći ćemo za svaki razred *sredinu* razreda (X). To ćemo dobiti tako da zbrojimo donju i gornju granicu razreda, i zbroj podijelimo sa 2. Npr. sredina našeg prvog razreda je (140 + 144) : 2 = 142.
2. Izabrat ćemo "privremenu aritmetičku sredinu" ($\overline{X}_{\text{pr}}$), a kao privremenu aritmetičku sredinu možemo uzeti sredinu bilo kojega razreda. Najpraktičnije je ako odaberemo onaj razred u kojem ima najviše rezultata.
3. Naći ćemo za koliko je intervala udaljena sredina pojedinog razreda od privremene aritmetičke sredine (x').
4. Dobivene brojeve (x') pomnožit ćemo s frekvencijom (f) pojedinih razreda (fx') i te ćemo rezultate zbrojiti s obzirom na predznak ($\Sigma fx'$).
5. Dobivenu sumu podijelit ćemo s brojem rezultata (N) i pomnožiti intervalom: $\frac{\Sigma fx'}{N} \cdot i$.
6. Vrijednost dobivenu pod točkom 5. algebarski ćemo pribrojiti privremenoj aritmetičkoj sredini, tj. ako je ta vrijednost pozitivna, onda ćemo njezin iznos dodati, a ako je negativna, tada ćemo njezin iznos oduzeti od privremene aritmetičke sredine.

Prema tome, čitav ćemo taj postupak unijeti u tablicu i izraditi kao što je prikazano:

Sredina razreda (X)	Frekvencija (f)	Intervalna udaljenost X od $\overline{X}_{\text{pr}}$ (x')	fx'
142	1	−5	−5
147	1	−4	−4
152	3	−3	−9
157	5	−2	−10
162	6	−1	−6
167 ($\overline{X}_{\text{pr}}$)	12	0	−34
172	8	1	8
177	4	2	8
182	4	3	12
187	3	4	12
192	2	5	10
197	1	6	6
	$N = 50$		+56
			$\Sigma fx' = +56 - 34 = +22$

$$\frac{\Sigma fx'}{N} \cdot i = +22/50 \cdot 5 = +2,2$$

$$\overline{X} = \overline{X}_{\text{pr}} + \left(\frac{\Sigma fx'}{N} \cdot i\right) = 167 + 2,2 = 169,2.$$

Kao što vidimo, rezultat se tek neznatno razlikuje od onoga koji smo dobili pomoću formule (4.1).

Male razlike između $\overline{X}$ izračunate prema formuli (4.1) i formuli (4.4), koje katkada nalazimo, uzrokovane su time što u skraćenom postupku *sve* rezultate unutar pojedinog razreda identificiramo s *jednim* rezultatom, tj. sa *sredinom* razreda. Kada je u razredu samo mali broj rezultata, onda njihova aritmetička sredina može znatnije odstupati od sredine razreda.

Kao i u svakom računu, tako je i pri izračunavanju aritmetičke sredine po skraćenom postupku potrebno provesti kontrolu računa. U ovom slučaju vrlo preporučljiv postupak kontrole sastoji se u tom da se kao privremena aritmetička sredina odabere sredina nekog drugog razreda. Na taj će način vrijednosti x', fx' i $\Sigma fx'$ biti promijenjene, a konačan rezultat — ako smo ispravno radili — mora biti potpuno jednak.

4.2. ZAJEDNIČKA ARITMETIČKA SREDINA

Često se u praksi događa da smo neku pojavu izmjerili nekoliko puta i svaki put smo izračunali aritmetičku sredinu iz više mjerenja. Ako konačno želimo dobiti *zajedničku* aritmetičku sredinu svih tih mjerenja, *ne smijemo* zbrojiti aritmetičke sredine i podijeliti ih njihovim brojem (osim u slučaju kad su sve aritmetičke sredine dobivene iz jednog broja mjerenja), jer aritmetička sredina je, kao ***težište*** rezultata, osjetljiva na *vrijednost i broj* rezultata. Prema tome, zajednička bi aritmetička sredina mogla zbog jedne ekstremne aritmetičke sredine biti značajno "pomaknuta". No ako je ta pojedinačna aritmetička sredina dobivena iz *malog* broja mjerenja, ona u ukupnom broju mjerenja ne bi smjela imati značajniji utjecaj. Kako je osnovna formula za aritmetičku sredinu $\overline{X} = \frac{\Sigma X}{N}$, slijedi da bi zajednička aritmetička sredina morala biti: $\frac{\text{suma svih rezultata}}{\text{ukupan broj mjerenja}}$. Ako je $\overline{X} = \frac{\Sigma X}{N}$, onda $N\overline{X} = \Sigma X$, pa ćemo prema tome zajedničku aritmetičku sredinu izračunati prema formuli:

$$\textit{Zajednička } \overline{X} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + \ldots + N_n\overline{X}_n}{N_1 + N_2 + \ldots + N_n} \tag{4.5}$$

Time smo postigli isti efekt kao da smo izračunali $\overline{X}$ iz *svih* pojedinačnih rezultata naših mjerenja. Primjer. Neko smo mjerenje ponavljali šest puta i dobili ove rezultate:

1. mjerenje	2. mjerenje	3. mjerenje	4. mjerenje	5. mjerenje	6. mjerenje
$\overline{X}_1 = 20,5$	$\overline{X}_2 = 22,0$	$\overline{X}_3 = 23,1$	$\overline{X}_4 = 22,2$	$\overline{X}_5 = 22,8$	$\overline{X}_6 = 22,6$
$N_1 = 5$	$N_2 = 40$	$N_3 = 17$	$N_4 = 35$	$N_5 = 19$	$N_6 = 25$

Zajednička $\overline{X}$ =

$$= \frac{5 \cdot 20,5 + 40 \cdot 22 + 17 \cdot 23,1 + 35 \cdot 22,2 + 19 \cdot 22,8 + 25 \cdot 22,6}{5 + 40 + 17 + 35 + 19 + 25}$$

$$= 3150,4/141 = 22,34.$$

4.3. NEKE DRUGE MJERE CENTRALNE TENDENCIJE

Osim aritmetičke sredine upotrebljavamo katkad i druge mjere centralne tendencije, i to prvenstveno onda kad zbog različitih razloga nije preporučljivo izračunavati aritmetičku sredinu. Jedan od glavnih razloga sastoji se u tom da katkada moramo uzeti u račun i neke *vrlo ekstremne* vrijednosti koje *bitno* mijenjaju aritmetičku sredinu. Kao što se već iz same osnovne formule za izračunavanje aritmetičke sredine vidi $\left(\overline{X} = \frac{\Sigma X}{N}\right)$, na aritmetičku sredinu djeluje *vrijednost* rezultata.

Na primjer:

Rezultati mjerenja (a)	Rezultati mjerenja (b)
6	6
5	5
5	5
5	5
3	3
4	4
5	5
7	7
5	5
5	15
$\overline{X} = 50/10 = 5,0$	$\overline{X} = 60/10 = 6,0$

Kako vidimo, zadnji rezultat od 15, koji je jedina razlika između mjerenja a) i b), "navukao" je aritmetičku sredinu s $\overline{X} = 5$ na $\overline{X} = 6$. Zbog toga što na aritmetičku sredinu djeluje *vrijednost* rezultata, ona je nazvana i ***težištem*** rezultata; drugim riječima, "težina" rezultata (koja se očituje u njihovom odstupanju od $\overline{X}$) iznad i ispod aritmetičke sredine uvijek je jednaka ili, kako kažu statističari, "aritmetička sredina je točka oko koje suma pozitivnih i negativnih odstupanja iznosi nula". (Vidi o tome sliku 6.7.)

Najpoznatije od tih drugih mjera su centralna vrijednost (ili medijan), dominantna vrijednost (ili modalna vrijednost), geometrijska sredina i harmonična sredina.

4.3.1. *Centralna vrijednost*

Centralna vrijednost (C) je vrijednost koja se u nizu rezultata, *poredanih po veličini*, nalazi točno u sredini.

Primjer: Ako u jednom mjerenju dobijemo ovih 11 rezultata:

7 9 4 7 8 7 10 6 6 9 8,

pa ih poredamo po veličini, dobivamo:

4 6 6 7 7 7 8 8 9 9 10.

Budući da imamo 11 rezultata, *srednji* rezultat je *šesti* rezultat (jer imamo 5 rezultata ispred i 5 rezultata iza njega), pa je, prema tome, u našem slučaju $C = 7$.

Kako se vidi, *položaj* rezultata koji zauzima centralna vrijednost, možemo odrediti pomoću formule:

$$\text{Položaj } C = (N+1)/2 \tag{4.6}$$

N a p o m e n a. Pazi, to nije *vrijednost* C, nego samo njezin *položaj* u rezultatima koji su poredani po veličini.

Ako je broj rezultata paran, centralna se vrijednost izračunava tako da zbrojimo dva srednja rezultata i taj zbroj podijelimo s 2.

Na primjer, kad bismo imali rezultate: 4 5 5 6 8 9,

$$C = (5+6)/2 = 5,5.$$

Prednost centralne vrijednosti pred aritmetičkom sredinom sastoji se u tome što na nju ne utječe *vrijednost* pojedinih rezultata, pa prema tome jedan vrlo ekstreman rezultat neće ništa promijeniti vrijednost C, koja je uvjetovana samo *brojem* rezultata.

Primjer: Ako 5 namještenika imaju ove bodove u platnoj listi:

1	2	3	4	5
750	800	850	900	5000,

onda bi "prosječni" bodovi, izračunati pomoću aritmetičke sredine bili: $\overline{X} = (750 + 800 + 850 + 900 + 5000)/5 = 1660$, a to je rezultat koji je — kako vidimo — daleko od bilo kakvog "prosjeka", pa prema tome i od stvarnog stanja. U tom slučaju najopravdanije je izračunati centralnu vrijednost, tj. bodove koji stoje po *sredini* čitave skale bodova. Drugim riječima, jednaki broj službenika ima manje odnosno više bodova od svih "srednjih" bodova. U našem slučaju $C = 850$.

Jedna od praktičnih upotreba centralne vrijednosti sastoji se u lociranju optimalnog položaja. Pretpostavimo da jedno građevinsko poduzeće u jednom nenaseljenom kraju ima uz glavnu prometnicu 5 postaja. Svaka od tih postaja mora svakodnevno svoje kamione opskrbljivati gorivom iz jednog centralnog skladišta. Gdje bi centralno skladište trebalo biti locirano da bi se minimalno smanjilo putovanje kamiona po gorivo? Odgovor glasi: na mjestu koje odgovara *centralnoj* vrijednosti.

Pretpostavimo da slika 4.1. predstavlja shemu ceste sa postajama A do E, i označenim kilometrima od točke 0. Centralna se vrijednost nalazi na postaji C. Mjerenjem udaljenosti lako se može ustanoviti da je centralna vrijednost ona točka od koje je *najmanja* suma svih odstupanja. Drugim riječima, ukupan broj potrebnih kilometara od svake postaje do centralnog skladišta najmanji je ako je centralno

skladište na poziciji centralne vrijednosti. (Ako je broj postaja *paran*, ukupna udaljenost bit će jednaka od centralne vrijednosti kao i od obiju postaja koje su s jedne i druge strane centralne vrijednosti.)

Slika 4.1. Shema ceste sa 5 postaja. Najbolja pozicija za centralno skladište je postaja C, tj. centralna vrijednost

4.3.2. *Dominantna vrijednost*

Dominantna vrijednost (D) je ona vrijednost koja je u nizu mjerenja najčešće postignuta (dakle koja "dominira"). U našem prijašnjem primjeru s 11 rezultata $D = 7$, jer je rezultat 7 najčešće (tri puta) postignut. Kad u tržnom izvještaju čujemo da se cijena nekoj namirnici na tržištu kretala od 4 do 9 kuna, ali da je "prosječna" cijena iznosila 6,5 kuna, to znači da je *najveći broj* prodavača namirnicu prodavao uz tu cijenu.

Ili, prikupljajući u jednom uzorku od 550 bračnih parova podatke o broju djece u obitelji, dobili smo, recimo, da je 550 obitelji imalo ukupno 1 660 djece, što daje aritmetičku sredinu od 3,02 djeteta po obitelji. No kada bismo, recimo, na temelju tih podataka počeli izgrađivati stanove za "prosječnu" obitelj s troje djece, učinili bismo dosta veliku pogrešku, jer, ako pogledamo na originalne rezultate, ustanovit ćemo da su najbrojnije obitelji s *dva djeteta i s jednim djetetom:*

Broj djece:	0	1	2	3	4	5	6	7	8	9	10	11	12
Broj bračnih parova s tim brojem djece:	70	90	108	86	70	47	30	20	15	5	4	3	2

Dakle, i u ovom slučaju najreprezentativnija je vrijednost ona koja je *najčešća*, dakle dominantna vrijednost: dvoje djece.

Prednost dominantne vrijednosti pred aritmetičkom sredinom je u tome što na nju ne utječe ni broj ni vrijednost rezultata, već samo frekvencija pojedinih rezultata. Ako imamo rezultate grupirane u razredu, aproksimativna dominantna vrijednost je sredina onog razreda koji ima najveću frekvenciju.

4.3.3. *Geometrijska sredina*

Geometrijska sredina (G). Prema definiciji geometrijska sredina je n-ti korijen iz umnožaka između N brojeva.

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \ldots x_N}. \tag{4.7}$$

Ona se pretežno koristi kao mjera prosječne brzine nekih promjena. Na primjer, ako je neko mjesto imalo g. 1960. 2000 stanovnika, g. 1961. 9000, a g. 1962. 18 000 stanovnika, onda je populacija g. 1961. bila 4,5 puta veća od populacije u g. 1960, a populacija g. 1962. dva puta veća od g. 1961. Postavimo li pitanje koliko je puta *prosječno* populacija svake godine porasla, izračunat ćemo to pomoću geometrijske sredine:

$$G = \sqrt{4,5 \cdot 2} = \sqrt{9} = 3 \text{ puta}.$$

Kako se vidi, to odgovara stvarnom stanju, jer se od g. 1960. do 1962. populacija ukupno povećala devet puta (2 000 · 3 · 3 = 18 000). Da smo, međutim, izračunali aritmetičku sredinu $\frac{4,5+2}{2}$, dobili bismo netočan podatak, tj. 3,25. G se ne može izračunavati ako je bilo koji broj negativan ili nula!

4.3.4. *Harmonična sredina*

Harmonična sredina (H) izračunava se prema formuli:

$$H = \frac{N}{\sum \frac{1}{x}}. \tag{4.8}$$

Harmoničnu sredinu valja upotrebljavati kad želimo dobiti prosjeke nekih odnosa (npr. prosječne kilometre na sat, prosječan broj slova u minuti i sl.). Uzet ćemo dva realna primjera iz svakodnevnog života:

Ako je automobilist udaljenost od 100 km vozio brzinom od 100 km/sat, a natrag je išao brzinom od 50 km/sat, kojom je prosječnom brzinom vozio?

Iako smo skloni da u prvi mah kažemo da mu je prosječna brzina bila 75 km/sat, brzo ćemo uvidjeti da to nije točno, jer je on put tamo i natrag prešao za 3 sata, pa je prema tome prosjek od 75 km/sat za 200 km puta — kako vidimo — previsok. Upotrebom harmonične sredine, dobit ćemo ispravan rezultat:

$$H = \frac{2}{\frac{1}{100} + \frac{1}{50}} = 66,7 \text{ km/sat}.$$

N a p o m e n a. Do varke dolazi zato što smo uzeli u obzir samo *brzinu*, a ne i *vrijeme* provedeno u nekoj brzini. Kad bi vrijeme bilo jednako, onda bi aritmetička sredina (a ne više harmonična sredina!) dala točan rezultat: vozeći 1 sat brzinom od 100 km, i 1 sat brzinom od 50 km, vozač je prešao ukupno 150 km, i imao je prosječnu brzinu od 75 km/sat.

Ako prilikom jednog tržišnog istraživanja anketiramo 3 domaćice i tražimo da nam kažu koliko dana u njihovoj obitelji traje staklenka od 1 kg marmelade, te ako dobijemo ove odgovore:

domaćica A: 5 dana
domaćica B: 10 dana
domaćica C: 15 dana,

koliko *prosječno* u ta tri domaćinstva traje 1 kg marmelade? Ne, *ne traje* 10 dana, iako se tako u prvi čas čini! Pogledajmo koliko staklenki ta 3 domaćinstva potroše u 30 dana:

domaćinstvo A: 6 staklenki
domaćinstvo B: 3 staklenke
domaćinstvo C: 2 staklenke
Ukupno: 11 staklenki.

Dakle, 11 kg marmelade na 30 dana, što znači da 1 kg traje u prosjeku 30/11 = = 2,727 dana *za sva tri domaćinstva*. Za jedno domaćinstvo dakle 1 kg marmelade traje *dulje*, tj. u prosjeku traje 2,727 · 3 = 8,2 dana.

Provjerimo li tu logiku računom harmonične sredine, dobivamo:

$$H = \frac{3}{1/5 + 1/10 + 1/15} = 8,2.$$

Ni H se ne može izračunati ako je bilo koji broj negativan ili nula.

ZADACI ZA VJEŽBU

1. Izračunajte aritmetičku sredinu, centralnu vrijednost i dominantnu vrijednost za niže navedene podatke:
   a) 10, 8, 6, 0, 8, 3, 2, 2, 8, 0;
   b) 1, 3, 3, 5, 5, 5, 7, 7, 9;
   c) 120, 5, 4, 4, 4, 2, 1, 0.
   U kojem od prethodnih slučajeva aritimetička sredina predstavlja neprikladnu mjeru centralne tendencije i zašto?

2. Neki je novinar na putovanju kroz zemlju u 5 različitih gradova kupovao kasete za svoj kasetofon kojim je obavljao intervjue. Cijena pojedinoj kaseti i broj kupljenih kaseta u svakom od 5 gradova navedeni su dolje:

	Cijena kasete	Broj kupljenih kaseta
Grad 1	61 bod	17
Grad 2	55 bodova	21
Grad 3	58 bodova	16
Grad 4	65 bodova	11
Grad 5	57 bodova	19

   Koliko je tog novinara prosječno stajala svaka kaseta?

3. Prilikom jednog biometrijskog mjerenja zagrebačke srednjoškolske omladine izmjerena je svojedobno visina 135 20-godišnjih mladića. Razredi i frekvencije u pojedinim razredima uneseni su u donju tablicu. Izračunajte aritmetičku sredinu visine tih mladića:

Razred	f
157 — 159	1
160 — 162	2
163 — 165	9
166 — 168	15
169 — 171	25
172 — 174	28
175 — 177	20
178 — 180	16
181 — 183	13
184 — 186	5
187 — 189	1
	$N = 135$

5. MJERE VARIJABILNOSTI

Kod mjerenja mnogih pojava možemo opaziti da se rezultati grupiraju i skupljaju oko jedne srednje vrijednosti. Jedino pod tom pretpostavkom i imamo pravo računati neku srednju vrijednost, na primjer aritmetičku sredinu, jer želimo da nam ona na neki način *reprezentira* sve naše rezultate.

Ako su vrijednosti nekog niza mjerenja *gusto* grupirane oko srednje vrijednosti, onda nam ta srednja vrijednost dobro reprezentira te rezultate. Naprotiv, ako su vrijednosti mjerenja samo minimalno grupirane oko srednje vrijednosti, onda nam ona slabo reprezentira rezultate.

Kad bi — u ekstremnom slučaju — sve vrijednosti nekog niza mjerenja bile *jednake*, onda bi "srednja vrijednost" bila potpuno točan reprezentant *svih* rezultata. Naprotiv, u drugom ekstremnom slučaju — kad rezultati ne bi pokazivali *nikakvu* "centralnu tendenciju", tj. nikakvo grupiranje oko neke srednje vrijednosti — ta nam srednja vrijednost ne bi reprezentirala ništa, iako npr. aritmetičku sredinu i u tom slučaju možemo "izračunati"!

Naime, sama aritmetička sredina nije nam još nikakva garancija da se rezultati grupiraju oko te aritmetičke sredine i zato je uvijek potrebno znati kako i koliko se oni grupiraju, tj. je li nam dobivena aritmetička sredina dobar ili loš reprezentant naših rezultata.

5.1. RASPON

Najjednostavnija (ali i najnetočnija) mjera grupiranja rezultata oko neke srednje vrijednosti je tzv. "raspon", tj. razlika između najvećeg i najmanjeg rezultata.

Primjer. Prilikom dva puta po 10 mjerenja neke pojave, dobili smo ova dva niza rezultata (rezultati su poredani po veličini):

1. mjerenje:	8	8,5	8,5	9	9	9	9	9,5	9,5	10
2. mjerenje:	1	2	3	5	9	9	13	15	16	17.

U oba slučaja suma rezultata = 90, i aritmetička sredina $\overline{X} = 9,0$. No već površnim pregledom rezultata možemo ustanoviti da se rezultati prvog mjerenja znatno bolje grupiraju oko svoje aritmetičke sredine nego rezultati drugog mjerenja. U prvom je slučaju raspon $10-8=2$, a u drugom slučaju $17-1=16$. Prema tome, prva aritmetička sredina mnogo je "vrednija", jer ona znatno bolje reprezentira rezultate iz kojih je dobivena.

Međutim, "raspon" je vrlo nesigurna i varljiva mjera varijabilnosti rezultata, jer bilo koji osamljeni *ekstremni* rezultat znatno povećava raspon a da se grupacija rezultata oko aritmetičke sredine ipak nije bitno promijenila.

Osnovni se nedostatak raspona sastoji u tome što je on obično to veći što je veći broj mjerenja neke pojave. Na primjer, iz naših 50 rezultata vremena reakcije slučajnim biranjem ("naslijepo") izvadili smo 3 grupe rezultata: u prvoj grupi željeli smo samo 2 rezultata, u drugoj pet, a u trećoj deset rezultata. Dobili smo ovo:

I	II	III
163	173	164
165	173	170
Raspon = 2		
	165	190
	166	184
	146	187
	Raspon = 27	
		154
		192
		177
		189
		158
		Raspon = 38.

Kako se vidi, raspon je i ovoga puta bio to veći što je veći broj rezultata uzet u obzir.

Sa stajališta zakona vjerojatnosti tu je pojavu prilično lako razumjeti: uzmemo li sve rezultate u obzir, raspon je razlika između najvećeg i najmanjeg rezultata. Uzmemo li u obzir samo nekoliko rezultata, vrlo je mala vjerojatnost da će među njima biti upravo najveći i najmanji rezultat.

5.2. SREDNJE ODSTUPANJE

Zanima li nas prosječna veličina odstupanja pojedinačnih rezultata (bez obzira na smjer odstupanja), možemo izračunati srednje odstupanje prema formuli:

$$\text{Srednje odstupanje} = \frac{\Sigma\,|X-\overline{X}|}{N}, \tag{5.1}$$

gdje $|X-\overline{X}|$ znači ***apsolutnu*** veličinu odstupanja, dakle bez obzira na predznak. Npr., ako imamo ove rezultate:

Rezultati	Odstupanja
5	0
7	2
4	1
6	1
5	0
6	1
5	0
2	3
4	1
6	1
$\Sigma = 50$	$\Sigma = 10$

$\overline{X} = 50/10 = 5,0.$

Srednje odstupanje $= 10/10 = 1.$

Rezultati, prema tome, *prosječno* odstupaju od aritmetičke sredine za 1. Srednje odstupanje možemo izračunati uz aritmetičku sredinu, centralnu i dominantnu vrijednost, ali nam ono, nažalost, ne može dovoljno služiti ako želimo izvoditi daljnja računanja.

5.3. STANDARDNA DEVIJACIJA

Kada bismo prosječno odstupanje računali ***vodeći računa*** o predznaku, onda bismo uvijek kao sumu dobili ***nulu***. Razlog tome već nam je poznat: aritmetička sredina, kao ***težište*** rezultata, vrijednost je od koje suma odstupanja iznad i ispod nje uvijek iznosi 0.

Jedan od načina da se izbjegnu predznaci odstupanja jest taj da se odstupanja kvadriraju. Osim toga, što je odstupanje veće, to ono više dolazi do izražaja ako ga kvadriramo. Ako tako kvadrirana odstupanja zbrojimo i izračunamo im aritmetičku sredinu, dobit ćemo mjeru varijabiliteta koja se u statistici naziva "*varijanca*".

(N a p o m e n a. Zbog razloga koji će poslije biti spomenuti, aritmetička sredina kvadriranih odstupanja izračunava se s $N-1$, a ne s N u nazivniku.)

Iz definicije varijance proizlazi i njezina formula:

$$\text{varijanca} = s^2 = \frac{\Sigma\,(X-\overline{X})^2}{N-1}. \tag{5.2}$$

To je, dakle, "prosječna" suma kvadriranih odstupanja ("prosjek" je u ovom slučaju dobiven dijeljenjem s $N-1$ umjesto s N).

Dok matematičari i profesionalni statističari uglavnom koriste pojam varijance, taj je pojam grafički nemoguće predočiti, a osim toga za razumijevanje daljnjih statističkih pojmova, kao i za *razumijevanje* varijabiliteta nije prikladan. Međutim *korijen* iz varijance može se — kako ćemo vidjeti — prikazati kao potpuno definirani *razmak* na skali rezultata. Taj drugi korijen iz varijance nazvan je *standardna devijacija* (a bilježi se najčešće sa $s, S.D., \sigma$), i to zato što se ta mjera koristi kao *standard* za mjerenje varijabiliteta rezultata.

Dakle, standardna devijacija

$$s = \sqrt{\text{varijance}},$$

$$s = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N-1}}. \qquad (5.3)$$

N a p o m e n a. Neki udžbenici statistike u nazivniku formule za izračunavanje standardne devijacije navode N umjesto $N-1$. Za to postoje određeni razlozi, no za sada ćemo ovdje spomenuti samo ono što je osnovno. U tzv. "deskriptivnoj" statistici, tj. onoj gdje nas zanimaju samo karakteristike neke skupine podataka, bilo bi potpuno ispravno standardnu devijaciju uzorka računati s N umjesto $N-1$ u nazivniku. Međutim, mnogo češće od želje za samom deskripcijom rezultata mi smo zainteresirani za izvođenje nekih zaključaka iz naših podataka, i to zaključaka u vezi s populacijom iz koje su ti podaci uzeti (u 9. poglavlju bit će rastumačena razlika između "populacije" i "uzorka".) U tom slučaju — a zbog razloga koji će biti tek kasnije rastumačeni i čitaocu razumljivi (također u 9. poglavlju) — ispravan postupak sastoji se u računanju s $N-1$ u nazivniku. Iako je računanje s N u nazivniku jednostavnije (iako nećemo ozbiljno pogriješiti ako kod *velikog* N stavimo u nazivnik N umjesto $N-1$), odlučili smo se ipak da standardnu devijaciju računamo s $N-1$ u nazivniku. Tako smo izbjegli potrebu nekih potonjih korektura (tzv. Besselova korektura), koje je potrebno provoditi ako se standardna devijacija računa s N u nazivniku. Osim u posve deskriptivnoj analizi uzorka (kada nas dakle zanimaju karakteristike samo postojećih rezultata u uzorku), N u nazivniku dolazi i onda ako računamo standardnu devijaciju *populacije* (dakle svih mogućih slučajeva). No kako nam populacija gotovo nikada nije poznata, to gotovo nikada ne dolazi u obzir (osim kada *definicijom* ograničimo populaciju; npr: svi učenici 7.a razreda neke škole).

Pogledajmo kakve su standardne devijacije izračunate iz naša dva niza mjerenja koja smo maloprije spomenuli:

1. *mjerenje* Rezultati X_1	$X_1-\overline{X}_1$	$(X_1-\overline{X}_1)^2$	2. *mjerenje* Rezultati X_2	$X_2-\overline{X}_2$	$(X_2-\overline{X}_2)^2$
8	−1	1	1	−8	64
8,5	−0,5	0,25	2	−7	49
8,5	−0,5	0,25	3	−6	36
9	0	0	5	−4	16
9	0	0	9	0	0
9	0	0	9	0	0
9	0	0	13	4	16
9,5	0,5	0,25	15	6	36
9,5	0,5	0,25	16	7	49
10	1	1	17	8	64
$\Sigma = 90$	$\Sigma(X_1-\overline{X}_1)^2 = 3,00$		$\Sigma = 90$	$\Sigma(X_2-\overline{X}_2)^2 = 330$	
$\overline{X}_1 = 9,0$			$\overline{X}_2 = 9,0$		

$$s_1 = \sqrt{\frac{\Sigma(X_1-\overline{X}_1)^2}{N_1-1}} = \sqrt{\frac{3}{9}} = \sqrt{0,3333} = 0,58 \qquad s_2 = \sqrt{\frac{\Sigma(X_2-\overline{X}_2)^2}{N_2-1}} = \sqrt{\frac{330}{9}} = \sqrt{36,67} = 6,06.$$

Kako se vidi, standardna devijacija drugog mjerenja mnogo je veća od standardne devijacije prvog mjerenja.

Ako bismo u našem primjeru s 50 vremena reakcije izračunali na taj način standardnu devijaciju, dobili bismo — nakon dosta velikog posla — vrijednost $s = 11,85$.

Kada su rezultati pravilno, simetrično i "normalno" grupirani oko aritmetičke sredine (a što znači "normalna" raspodjela bit će ubrzo jasno), onda je u intervalu koji obuhvaća $\overline{X} \pm 1s$, 68,26% svih rezultata. Drugim riječima, to znači ovo: ako aritmetičkoj sredini na jednu i na drugu stranu "dodamo" vrijednost standardne devijacije (a to znači da od aritmetičke sredine *oduzmemo*, i također aritmetičkoj sredini *pribrojimo* standardnu devijaciju), obuhvatit ćemo oko 68% rezultata. Konkretno, u našem primjeru s 50 vremena reakcije aritmetička sredina iznosi 169,16, a standardna devijacija 11,85. Oduzmemo li 11,85 od aritmetičke sredine, dobivamo $169,16 - 11,85 = 157,31$. Ako aritmetičkoj sredini pribrojimo standardnu devijaciju, dobivamo 169,16 + 11,85 = 181,01. To znači da se oko 68% svih rezultata nalazi u intervalu između 157,31 i 181,01. Ako pogledamo originalne rezultate, pa ako izbrojimo koliko ih ima koji su jednaki ili veći od 158 (to je prvi rezultat po preciznosti mjerenja iznad 157) i jednaki ili manji od 181, izbrojit ćemo ih 35. To je 70% svih rezultata, što je vrlo dobra aproksimacija teoretskom broju od 68,26%, koji vrijedi samo za idealno pravilnu raspodjelu.

Ako aritmetičkoj sredini "dodamo" lijevo i desno *dvije* standardne devijacije, obuhvatit ćemo u idealnom slučaju 95,44% svih rezultata. A ako joj dodamo lijevo i desno *tri* standardne devijacije, obuhvatit ćemo 99,73% rezultata, dakle praktički sve rezultate. Upravo u *toj činjenici* nalazi se osnovna praktična prednost pojma standardne devijacije: uz njezinu pomoć možemo vrlo uspješno predvidjeti u kojem se rasponu kreću praktički svi rezultati (to je raspon $\overline{X} \pm 3s$), a naravno i u kojem se rasponu nalazi oko 95% rezultata ($\overline{X} \pm 2s$) i oko 68% rezultata ($\overline{X} \pm 1s$).

Standardna devijacija smije se računati samo uz *aritmetičku sredinu*, a ne i uz druge mjere centralne tendencije (dakle ne uz centralnu, ni dominantnu vrijednost, ni ostale "prosjeke").

Kako je aritmetička sredina gotovo uvijek decimalni broj, to bi traženje razlike između svakoga pojedinog rezultata i aritmetičke sredine i kvadriranje tih razlika predstavljalo naporan i dugotrajan (i dosadan!) posao. Zato su statističari izradili jednostavnije metode, koje nam omogućuju da standardnu devijaciju izračunamo iz *originalnih* rezultata (X), ne tražeći razlike između svakog rezultata i aritmetičke sredine. (Taj je način pogodan samo onda ako originalni rezultati nisu veliki brojevi.)

Standardnu devijaciju možemo izračunati i prema formuli:

$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}. \qquad (5.4)$$

U naša dva primjera dobili bismo u tom slučaju:

$$s_1 = \sqrt{\frac{813 - \frac{8100}{10}}{9}} = 0,58,$$

$$s_2 = \sqrt{\frac{1140 - \frac{8100}{10}}{9}} = 6,06.$$

Kadikad se može pri izračunavanju standardne devijacije potkrasti znatna pogreška na taj način što se, recimo, decimalni zarez postavi na pogrešno mjesto. Kao neka brza kontrola da se takva pogreška nije dogodila može nam poslužiti odnos između raspona i standardne devijacije: taj odnos nije gotovo nikad manji od 2 ili veći od 6,5. U naša dva primjera odnos raspon/standardna devijacija iznosi u prvom primjeru 2/0,57 = 3,5, a u drugom primjeru 16/6,06 = 2,6.

Kad su nam rezultati grupirani u razrede, standardnu ćemo devijaciju izračunati po "skraćenom postupku" na ovaj način:

1. Provest ćemo prve 4 operacije, navedene na str. 51, za skraćeno izračunavanje aritmetičke sredine.

2. Dobivenoj tablici dodat ćemo još jedan stupac, i to $f \cdot x'^2$. Taj ćemo stupac lako izračunati ako međusobno pomnožimo stupac fx' i x'. Dobivene rezultate zbrojit ćemo $(\Sigma f x'^2)$.

3. Izraz $\Sigma f x'$ (suma četvrtog stupca) kvadrirat ćemo: $(\Sigma f x')^2$.

Standardnu devijaciju izračunat ćemo tada prema formuli:

$$s = i \sqrt{\frac{\Sigma f x'^2 - \frac{(\Sigma f x')^2}{N}}{N-1}} \tag{5.5}$$

U našem primjeru s 50 vremena reakcije dobivamo dakle (samo zadnji stupac je nov):

X		f	x'	fx'	fx'^2	
142		1	−5	−5	25	
147		1	−4	−4	16	
152		3	−3	−9	27	
157		5	−2	−10	20	
162		6	−1	−6	6	
167	$(\overline{X}_{pr})$	12	0	−34		$i = 5$
172		8	1	8	8	
177		4	2	8	16	
182		4	3	12	36	
187		3	4	12	48	
192		2	5	10	50	
197		1	6	6	36	
				+56		
		$N = 50$		$\Sigma f x' = +22$	$\Sigma f x'^2 = 288$	

$$s = 5 \cdot \sqrt{\frac{288 - \frac{22^2}{50}}{49}} = 5 \cdot \sqrt{\frac{278,32}{49}} = 5 \cdot \sqrt{5,68} = 11,92.$$

Kako smo već spomenuli, standardna devijacija pokazuje nam donekle *koliko vrijedi dobivena aritmetička sredina*, tj. je li ona dobar ili loš reprezentant rezultata. Zbog toga se treba priviknuti na običaj *da se uz svaku aritmetičku sredinu navede i pripadna standardna devijacija*. U našem bismo dakle primjeru konačan rezultat izrazili ovako:

$$\overline{X} = 169,2$$

$$s = 11,92.$$

5.4. KOEFICIJENT VARIJABILNOSTI

Kad su nam poznate aritmetička sredina i standardna devijacija nekih rezultata, onda su ti rezultati potpuno definirani, i možemo ih uspoređivati s nekim drugim rezultatima. Npr., kad smo na strani 62 i 63 izračunali aritmetičke sredine i standardne devijacije od po 10 rezultata dvaju mjerenja, i kad smo ustanovili da je $\overline{X}_1 = 9,0$, $s_1 = 0,58$, a $\overline{X}_2 = 9,0$, $s = 6,06$, onda smo potpuno sigurni da je $\overline{X}_1$ znatno bolji reprezentant svojih rezultata nego $\overline{X}_2$ jer rezultati u prvom mjerenju znatno manje *variraju* od rezultata u drugom mjerenju.

Međutim, to je bilo lako zaključiti zato što smo imali dvije *jednake* aritmetičke sredine. Ali ako imamo dvije *različite* aritmetičke sredine, teško je na prvi pogled odmah ustanoviti koji rezultati relativno više variraju.

Primjer: $\overline{X}_1 = 100$ cm $\qquad \overline{X}_2 = 8$ cm
$s_1 = 10$ cm $\qquad s_2 = 2$ cm.

Iako se u prvi mah čini da je povoljniji drugi slučaj, budući da je $s_1 > s_2$, ipak je standardna devijacija od 10 uz *aritmetičku sredinu* $\overline{X}_1 = 100$ *relativno manja* od standardne devijacije 2 uz aritmetičku sredinu $\overline{X}_2 = 8$, jer s_1 po svojoj vrijednosti iznosi 10% pripadajuće aritmetičke sredine, a s_2 iznosi 25% pripadajuće aritmetičke sredine!

Da bismo mogli međusobno uspoređivati varijabilnost različitih pojava i svojstava, služimo se tzv. koeficijentom varijabilnosti (V) koji nam pokazuje kolik postotak vrijednosti aritmetičke sredine iznosi vrijednost standardne devijacije:

$$V = \frac{s \cdot 100}{\overline{X}}. \tag{5.6}$$

Koeficijent varijabilnosti vrlo je korisna mjera u svim onim slučajevima kad želimo znati:

a) u kojem svojstvu neka grupa varira više, a u kojem manje;

b) koja od grupa varira više, a koja manje u *istom* svojstvu.

Primjer ad a): Jednim mjerenjem zagrebačke školske omladine g. 1951. utvrđeno je da 10-godišnji dječaci ($N = 612$) imaju visinu $\overline{X}_v = 134,4$ cm, $s_v = 6,06$ cm, a težinu $\overline{X}_t = 29,2$ kg, $s_t = 3,89$ kg. Variraju li više dječaci u visini ili u težini?

$$V_v = \frac{6,06 \cdot 100}{134,4} = 4,51\%$$

$$V_t = \frac{3,89 \cdot 100}{29,2} = 13,32\%.$$

Prema tome, dječaci variraju u težini znatno više nego u visini.

Primjer ad b): Prilikom tog istog mjerenja utvrđeno je da 10-godišnje djevojčice ($N = 684$) imaju visinu $\overline{X}_v = 134,9$, $s_v = 6,43$, a težinu $\overline{X}_t = 29,7$, $s_t = 4,78$. Variraju li u visini i u težini više dječaci ili djevojčice?

Visina:

$$V \text{ za dječake} = 4{,}51\%$$

$$V \text{ za djevojčice} = \frac{6,43 \cdot 100}{134,9} = 4,77\%$$

Težina:

$$V \text{ za dječake} = 13{,}32\%$$

$$V \text{ za djevojčice} = \frac{4,78 \cdot 100}{29,7} = 16,09\%.$$

Prema tome, 10-godišnje djevojčice variraju u visini i u težini nešto više od 10-godišnjih dječaka.

Jedno od čovjekovih svojstava, koje varira *vrlo malo*, jest tjelesna temperatura (zdravih ljudi): s prosjekom negdje između 36,5 i 37,0 stupnjeva i standardnom devijacijom od oko 0,1 - 0,2 stupnja, koeficijent varijabiliteta tjelesne temperature iznosi približno oko 0,5%.

No katkad valja biti oprezan u izračunavanju, tj. interpretaciji koeficijenta varijabiliteta. U nekim situacijama, naime, nije dopušteno uspoređivati pojedine koeficijente varijabiliteta. No o tome ćemo nešto reći u poglavlju "Skale mjerenja".

ZADACI ZA VJEŽBU

1. Izračunajte standardnu devijaciju za podatke a), b) i c) iz zadatka 1 u poglavlju "Mjere centralne tendencije".
2. Izračunajte standardnu devijaciju podataka iz zadatka 3 u prethodnom poglavlju; nakon toga izračunajte koeficijent varijabilnosti.
3. Izračunajte koeficijente varijabiliteta za podatke iz zadatka 1.

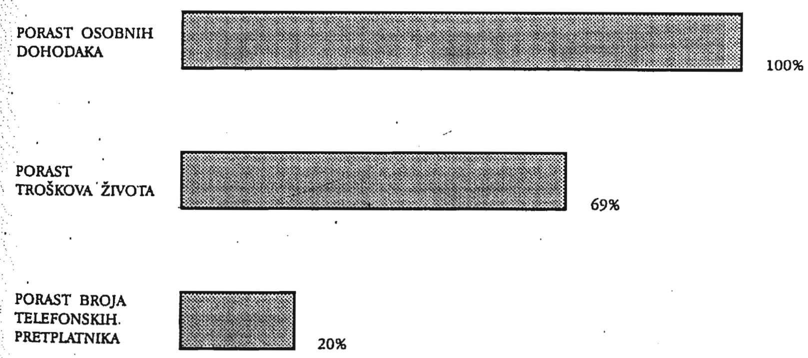
6. GRAFIČKO PRIKAZIVANJE REZULTATA

Jedna kineska poslovica kaže da slika (crtež) vrijedi koliko i tisuću riječi. Možda to nije nigdje toliko točno kao u statistici, gdje bismo poslovicu još mogli i proširiti i kazati da slika vrijedi kao tisuću riječi ili brojeva. U statistici naime često bez slike odnosno crteža ne možemo dobiti pregled nad rezultatima dok nam slika katkada u trenu otkrije neke osnovne karakteristike rezultata, ili zakonitosti, koje među samim rezultatima vladaju.

Ako "preskočimo" grafičko prikazivanje rezultata, izlažemo se dvostrukom riziku:

a. Postoji šansa da uopće nećemo uočiti neku posebnu i neočekivanu karakteristiku rezultata koju je gotovo nemoguće uočiti prilikom baratanja brojevima.

b. Neke teške računske pogreške također mogu ostati neotkrivene. Na primjer, pogrešno stavljen decimalni zarez pri izračunavanju aritmetičke sredine bit će odmah uočen kao pogreška ako su rezultati prikazani grafički.



Slika 6.1. Grafički prikaz porasta troškova

Navika da se rezultati prikazuju i grafički, a ne samo računom, dovodi do tzv. "grafičkog načina mišljenja", tj. do mogućnosti brzog i lakog razumijevanja nekih procesa, kao i do uspješnije komunikacije među stručnjacima. Iskustva sa studentima pokazuju da u tom pogledu postoje velike razlike među pojedincima, vjerojatno uzrokovane različitim obrazovanjem.

Rezultati se, kako je poznato, mogu grafički prikazati na mnogo načina.

Kad se radi o čistoj deskripciji međusobno nezavisnih ili slabo zavisnih podataka, najčešći način grafičkog prikazivanja je u horizontalnim ili vertikalnim *stupcima* i u *kružnom dijagramu* (koji se još naziva i "torta"-dijagram). Tako, na primjer, na slici 6.1. vidimo prikaz porasta nekih troškova u nekom određenom razdoblju.

TABLICA 6.1.

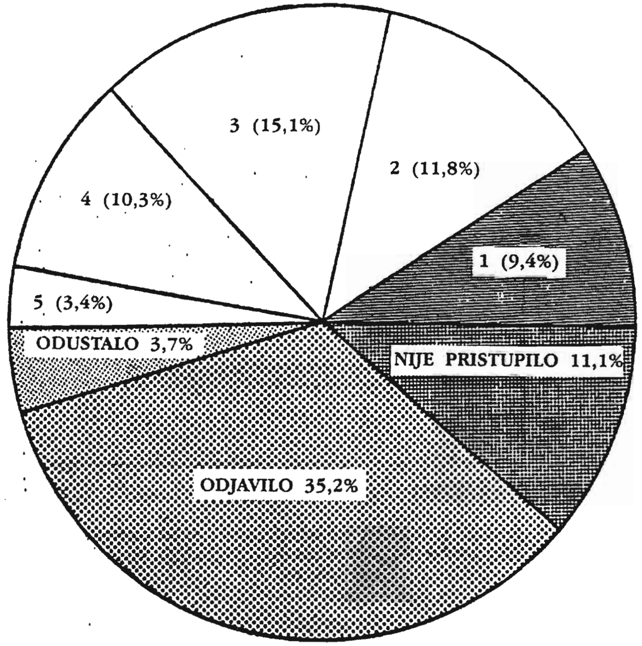
Ukupno prijavljeno	Položeno s ocjenom				Nije položeno	Nije pristupilo	Odjavljeno	Odustalo
	5	4	3	2				
4635	157	477	699	548	434	514	1634	172
Postotak 100,0	3,4	10,3	15,1	11,8	9,4	11,1	35,2	3,7

U tablici 6.1. prikazani su rezultati 4635 diplomskih ispita prijavljenih na Filozofskom fakultetu u Zagrebu u šk. g. 1967/68. Iako tablica daje posve jasne podatke, u *grafičkom* prikazu često možemo brže i jasnije ustanoviti neke upadljivije pojave. Stoga su rezultati iz tablice 6.1. prikazani i na kružnom dijagramu (vidi sliku 6.2). Na slici odmah upada u oči vrlo velik broj *odjavljenih* diplomskih ispita (više od 35%) kao i relativno velik broj ispita kojima kandidati — iako ispit nisu odjavili — uopće nisu pristupili (više od 11%). Na slici su položeni ispiti prikazani svijetlim, a nepoloženi tamnim površinama.

Kut koji u kružnom dijagramu zauzima pojedini postotak, računa se prema formuli:

$$\text{kut} = \frac{\text{frekvencija} \cdot 360}{N}. \tag{6.1}$$

U popularnim časopisima, dnevnoj štampi i sl. često se — zbog veće "plastičnosti" rezultata — daju *trodimenzionalni* grafički prikazi nekih odnosa veličina. Treba upozoriti da to *nije pogodan* način, što ćemo ubrzo dokazati. No prije toga jedno upozorenje i za *dvodimenzionalne* prikaze: ako se umjesto stupcima količine prikazuju *kvadratima* ili *krugovima*, to također nije pogodno, jer — kako znamo — kvadrat dvostruko dužih stranica po površini je *četiri puta veći*, a to isto vrijedi i za krug: krug dva puta većeg polumjera ima četiri puta veću površinu. Na primjer, ako jedan kvadrat ima stranicu dužine 10 mm, onda će kvadrat *dvostruke površine* imati stranicu dužine otprilike 14,2 mm, i *neće davati dojam dva puta većeg kvadrata*!

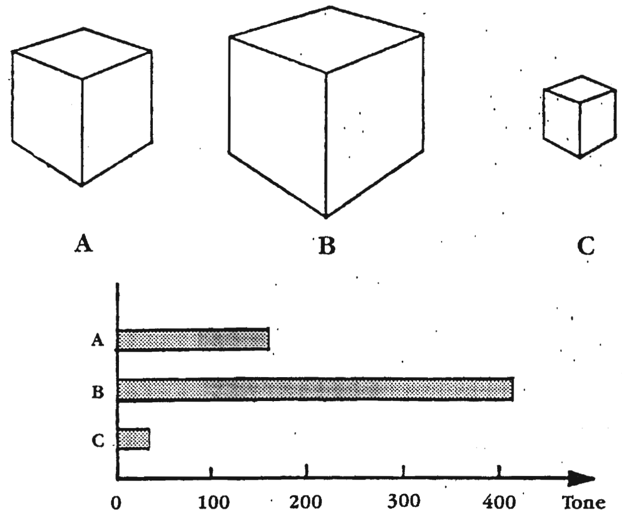


Slika 6.2. Kružni dijagram rezultata diplomskih ispita na Filozofskom fakultetu u Zagrebu za školsku godinu 1967/68

Još je gora situacija u trodimenzionalnom prikazu: kocka koja ima dva puta dužu stranicu od neke druge kocke, bit će po svom volumenu osam puta veća od prve kocke. Evo primjera: Pretpostavimo da proizvodnja nekoga industrijskog artikla u tri istorodna poduzeća u godini dana iznosi:

Poduzeće A:	160	tona
Poduzeće B:	420	tona
Poduzeće C:	30	tona.

Prikažemo li te podatke *stupcima*, dobit ćemo jasan i nepristran odnos (vidi sliku 6.3, dolje!). No ako te podatke prikažemo kockama, dakle trodimenzionalno, onda to izgleda kao gornji crtež na istoj slici. Iako su odnosi ispravni, *dojam* što ga slika daje ni približno nije realan, jer neupućeni promatrač slike neće smatrati da je kocka B po svom volumenu više od 2,6 puta veća od kocke A. Isto tako, čini li vam se zaista da je kocka C po volumenu više od pet puta manja od kocke A, a čak četrnaest puta manja od kocke B?!

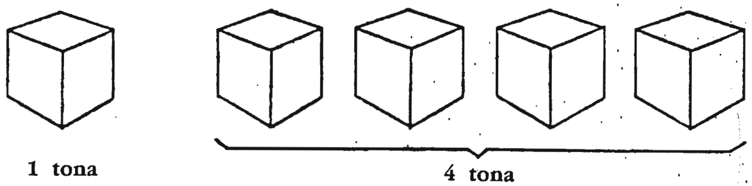


Slika 6.3. Isti podaci prikazani su stupcima i trodimenzionalno. Kao što se vidi, razlika među podacima mnogo je jasnija kada su prikazani stupcima

U praksi, tj. u različitim popularnim člancima, problem se obično "rješava" tako da se odnosi visina tih predmeta stave u stvarne brojčane odnose, što naravno dovodi do teškog nasilja nad rezultatima. Ako se to učini *namjerno* (a takvih slučajeva ima), onda to predstavlja falsificiranje rezultata. A ako se pak crtač pošteno pridržava onoga što smo rekli o odnosu volumena, onda postiže kod čitaoca *neopravdano mali efekt*!

Kako dakle treba postupiti?

Želimo li da podatke prikažemo trodimenzionalno, onda se to može jedino tako, a da pritom ne učinimo nikakvu pogrešku, da različite količine (vrijednosti) prikažemo sumom *jednakih* volumenskih jedinica. Na primjer, ako se produkcija od 100 tona prikaže kockom određene veličine, onda ćemo produkciju od 400 tona prikazati pomoću *četiri* jednako velike kocke (vidi sliku 6.4).



Slika 6.4. U trodimenzionalnim prikazima treba različite količine prikazivati sumom jedinica istog volumena

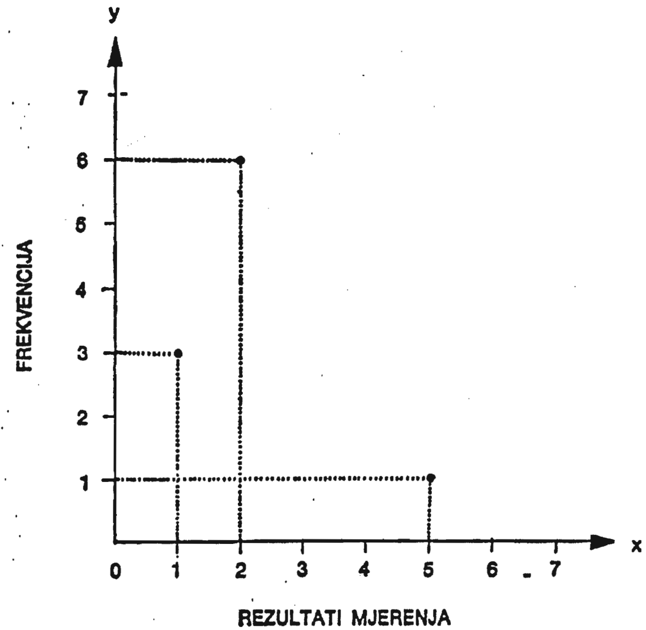
Kad imamo rezultate koji se razvijaju u vremenu (npr. krivulja dnevne temperature) ili koji se grupiraju oko neke reprezentativne vrijednosti, služimo se grafičkim sistemom *koordinatnog* sustava.

Položaj jedne točke u ravnini potpuno je određen s dvije koordinate koje su međusobno okomite. Horizontalna koordinata, koju nazivamo apscisa (ili os x), obično nam pritom služi za registraciju *vrijednosti* mjerenja, a vertikalna koordinata, koju zovemo *ordinata* (ili os y), za registraciju *frekvencija*.

Primjer: U nizu nekog mjerenja dobili smo ovih deset rezultata:

1 2 1 1 5 2 2 2 2 2.

Kako se vidi, vrijednost "1" postignuta je tri puta, vrijednost "2" šest puta, a vrijednost "5" jedanput. Prikazani u koordinatnom sustavu, ti bi rezultati, dakle, dali situaciju prikazanu na sl. 6.5.



Slika 6.5. Položaj pojedinih rezultata u koordinatnom sustavu

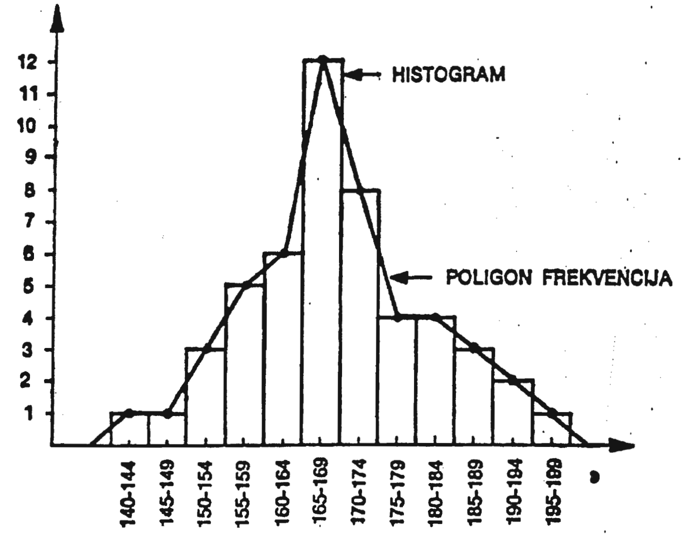
Ako su rezultati grupirani u razrede, onda na apscisu unesemo razrede, ili sredinu razreda, već prema tome želimo li rezultate prikazati histogramom ili poligonom frekvencija.

Histogram se sastoji od niza pačetvorina kojima površina (i visina) odgovara frekvenciji pojedinog razreda, a suma površina svih pačetvorina odgovara ukupnoj frekvenciji (ukupna frekvencija = N) svih razreda.

Poligon se crta tako, da iznad *sredine* svakog razreda označimo točke u visini ordinate, koja odgovara frekvenciji toga razreda. (Ako imamo nacrtani histogram, te točke su na sredini svakog "krovića" pojedinih pačetvorina.) No poligon treba "uzemljiti", tj. na lijevoj i desnoj strani krivulje dovesti ga na nultu frekvenciju, tj. na apscisu.

Poligon je više ili manje pravilna krivulja, kojoj totalna površina odgovara ukupnoj frekvenciji svih rezultata, ali površina iznad pojedinog razreda ne odgovara frekvenciji toga razreda, već frekvenciju razreda označuje samo *visina* poligona točno iznad sredine razreda.

Prikažemo li rezultate našeg mjerenja vremena reakcije pomoću histograma i poligona frekvencija, dobit ćemo rezultat prikazan na sl. 6.6.

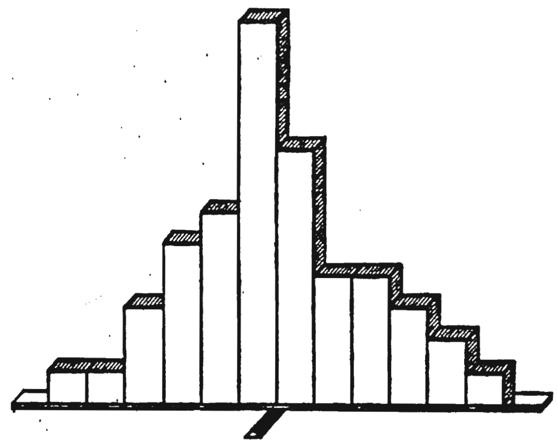


Slika 6.6. Histogram i poligon frekvencija rezultata iz primjera u tekstu

Na ovom mjestu, međutim, treba upozoriti na jednu pogrešku pri crtanju histograma koja se u praksi često događa. U praksi se, naime — zbog različitih razloga — katkada 2 ili više razreda želi spojiti zajedno. Recimo, kada bismo u našim rezultatima s 50 vremena reakcije željeli spojiti razrede "155-159" i "160-164", u prvom od ta dva razreda imamo 5 rezultata, a u drugom 6. Zajedno, dakle, imamo 11 rezultata. Pogreška se, koju smo spomenuli, sastoji u tome da se u poligon unese dvostruko široka pačetvorina (jer zauzima dva razreda), koja ima visinu 11, a to je pogrešno! Rekli smo da površina poligona predstavlja točnu reprodukciju frekvencije u pojedinom razredu. Iz te slike slijedilo bi dakle da mi u razredu 155-159 imamo 11, i u razredu 160-164 također 11 rezultata, a to — kako znamo — ne odgovara istini. *Ispravno* crtanje stupca iznad oba spojena razreda moralo bi dakle

izgledati tako da visina stupca ide do 5,5 — jer to znači za oba razreda zajedno 11, što odgovara realnom stanju.

Mi smo već spominjali da je aritmetička sredina "težište" rezultata. To se može uz pomoć histograma i dokazati. Ako naš histogram sa slike 6.6. izrežemo iz kakvoga debelog i čvrstog materijala (npr. iz daske), pa ako na apscisi označimo vrijednost aritmetičke sredine, moći ćemo na tom mjestu histograma *postaviti u ravnotežu*, kao što je to prikazano na slici 6.7.



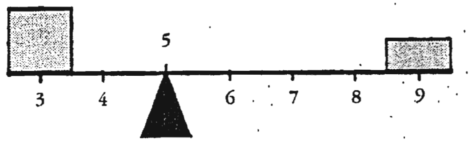
Slika 6.7. Aritmetička sredina je "težište" rezultata

Da je aritmetička sredina zaista pravo težište rezultata, i da se može usporediti s *vagom*, lako je dokazati i na još jedan, grafički način. No prije toga jedno pitanje čitaocu: Zamislite da imate običnu kuhinjsku metlu, i da je postavite horizontalno preko ispruženog prsta, te je smjestite na ono mjesto gdje je metla u potpunoj *ravnoteži*. Nakon toga na tom mjestu prerežete držak metle i vagnete svaki komad posebno. Što mislite, hoće li biti jednako teški? *Neće*, nego će teži biti kraći komad (tj. onaj, gdje se nalazi sama metla); duži komad (dio držala metle), upravo zato što je duži, ali lakši, po zakonu poluge bit će u ravnoteži s kraćim težim komadom.

A sada uzmimo da imamo ova 3 rezultata nekog mjerenja:

$$3, 3, 9.$$

Njihova aritmetička sredina iznosi $15/3 = 5$. Ako svaki od tih rezultata zamislimo kao uteg jednake težine na jednoj vagi (vidi sliku (6.8), vidimo da će vaga biti u ravnoteži onda ako su umnošci udaljenosti od težišta s težinom utega jednaki za lijevu i desnu stranu vage: na lijevoj strani vage imamo dva utega, pa ako njihovu težinu (primjerice 1 kg) pomnožimo s udaljenosti od težišta koja iznosi $5 - 3 = 2$, imamo $2 \cdot 2 = 4$. Na desnoj strani imamo *jedan* uteg, udaljen 4 jedinice $(9 - 5 = 4)$, pa umnožak težine i udaljenosti $(1 \cdot 4)$ iznosi 4.



Slika 6.8. Dokaz da je aritmetička sredina težište rezultata

Prema tome, težište vage, koja je u ovom slučaju u ravnoteži, jest točka označena vrijednošću 5.

Zamislite, gdje bi bilo težište vage kada bismo desni uteg pomaknuli, recimo, za 9 jedinica udesno, tj. na 18. Vaga bi bila u ravnoteži ako bismo joj i težište pomaknuli desno, i to točno na oznaku 8, jer u tom slučaju imali bismo s lijeve strane vage umnožak $2 \cdot (8-3) = 10$, a s desne strane umnožak $1 \cdot (18-8) = 10$.

Aritmetička sredina, naravno, također iznosi 8:

$$\overline{X} = \frac{3+3+18}{3} = \frac{24}{3} = 8.$$

Iz ovog je primjera naročito jasno vidljivo koliko je aritmetička sredina osjetljiva na *vrijednost* (veličinu) pojedinog rezultata.

Vratimo se još malo razmatranju histograma i poligona. Histogram je *najtočniji* prikaz distribucije frekvencije nekih rezultata, ali ipak u praksi se najčešće za grafičko prikazivanje koristi poligon frekvencija. Glavni razlog tome je činjenica da je poligon obično *pregledniji* način prikazivanja (on je sličniji "teoretskoj" krivulji, o kojoj će biti riječi u idućem poglavlju), a osobito je pogodan u slučajevima kada na istoj slici želimo prikazati *dvije ili više* distribucija.

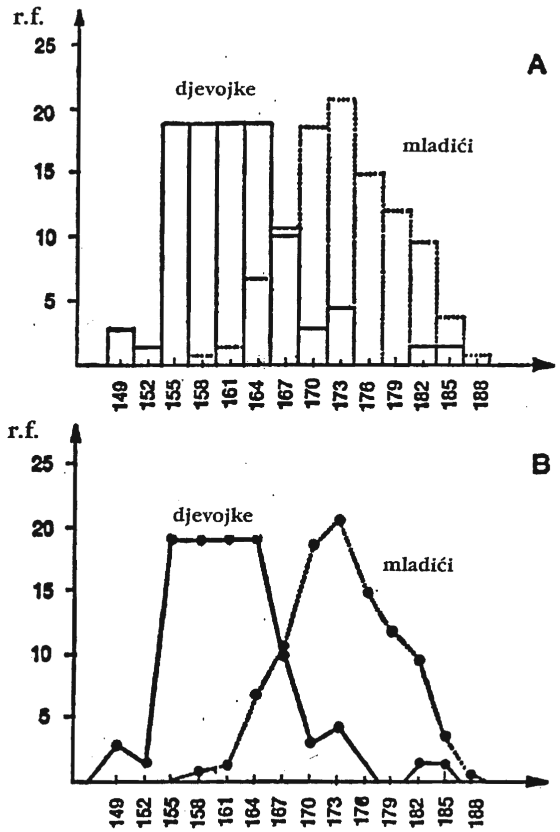
To ćemo ilustrirati jednim primjerom. Mjereći svojedobno visinu zagrebačke omladine, za skupinu od 135 20-godišnjih mladića i 69 20-godišnjih djevojaka, dobiveni su rezultati prikazani u tablici 6.2.

TABLICA 6.2.

VISINA GRUPE MLADIĆA I DJEVOJAKA

Mladići ($N = 135$)			Djevojke (N=69)		
Razred	Frekvencija	Rel. frekv. (%)	Razred	Frekvencija	Rel. frekv. (%)
			148-150	2	2,90
			151-153	1	1,45
			154-156	13	18,84
157-159	1	0,74	157-159	13	18,84
160-162	2	1,48	160-162	13	18,84
163-165	9	6,67	163-165	13	18,84
166-168	15	11,11	166-168	7	10,14
169-171	25	18,52	169-171	2	2,90
172-174	28	20,74	172-174	3	4,35
175-177	20	14,82	175-177	—	—
178-180	16	11,85	178-180	—	—
181-183	13	9,63	181-183	1	1,45
184-186	5	3,70	184-186	1	1,45
187-189	1	0,74		$N = 69$	100,00
	$N = 135$	100,00			

Kako broj izmjerenih mladića i djevojaka nije jednak, moramo prethodno obje distribucije svesti na "zajedničku mjeru" (zbog lakšeg uspoređivanja). To se postiže tako da se za svaki razred izračunaju *relativne frekvencije*, tj. svaka se frekvencija prikaže u postotku ukupnog broja.



Slika 6.9. Prikažemo li jednom slikom dvije distribucije uz pomoć poligona frekvencija (B), preglednost slike bit će mnogo bolja nego ako distribucije prikažemo histogramima (A)

Kada bismo sada obje distribucije prikazali histogramom, dobili bismo rezultat prikazan gornjim crtežem na slici 6.9. Kao što se vidi, stranice pojedinih pravokutnika kod mladića i djevojaka prekrivaju se te je gotovo nemoguće dobiti jasnu sliku o svakoj distribuciji posebno. Naprotiv, prikažemo li relativne frekvencije pomoću *poligona* (vidi crtež B), dobivamo sliku koja je toliko jasnija i preglednija od crteža A da više nije i potrebno govoriti o prednostima ovog načina prikazivanja!

Iz histograma i poligona frekvencija dade se izravno očitati jedino dominantna vrijednost: to je u histogramu razred s najvišim stupcem, a u poligonu razred iznad kojega se nalazi vrh krivulje.

Postoji još jedan, i to vrlo koristan i praktičan način grafičkog prikazivanja rezultata, koji je naročito pogodan kada želimo brzi pregled koliko rezultata, ili koliki se postotak rezultata nalazi ispod ili iznad nekog konkretnog rezultata, kao i podatak o tome gdje se otprilike nalazi centralna vrijednost. (O tome načinu prikazivanja i nekim računima u vezi s njime bit će još govora, i to u poglavlju "Položaj pojedinog rezultata u grupi".) Taj način prikazivanja zove se *krivulja kumulativne frekvencije* (neki je zovu i "Galtonova oživa").

Postupak je vrlo jednostavan: na apscisu se nanesu *prave* gornje granice razreda, a na ordinati se nalazi kumulativna (ili relativna kumulativna) frekvencija. No prethodno treba reći nekoliko riječi o tzv. "pravoj" gornjoj granici razreda. Uzmimo da smo pri nekom mjerenju razrede u koje smo smještavali pojedine rezultate rasporedili u intervale po 3, i to upravo tako kako je to na tablici 6.2, dakle ovako:

157-159
160-162
163-165
itd.

Već prije je rastumačeno zašto razrede sastavljamo na ovaj način (vidi stranicu 47-48). Prilikom samog *mjerenja,* iako smo mjerili na preciznost od 1 cm točnosti, mi smo svaki rezultat, koji je, recimo, u realnosti bio iznad 159, ali ispod 159,5 (npr. 159,4 cm), registrirali kao 159, a rezultat koji je bio veći od 159,5, ali ispod 160 (npr. 159,6), registrirali kao 160 cm. Iz toga proizlazi da *prava* gornja granica razreda 157-159 iznosi zapravo 159,5, prava gornja granica razreda 160-162 iznosi 162,5, itd. Prema tome, prave gornje granice razreda računski dobijemo tako da uzmemo sredinu između gornje vrijednosti jednog razreda i donje vrijednosti idućeg razreda.

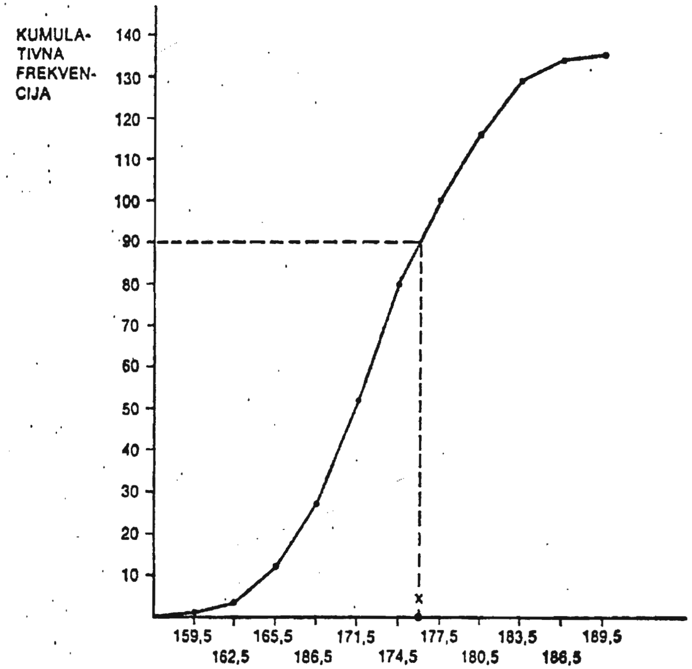
Evo, i te prave gornje granice razreda nanijet ćemo — kako rekosmo — na apscisu, a na ordinatu kumulativne frekvencije, tj. zbrojene frekvencije od najnižeg razreda nadalje.

Uzmimo za primjer maloprije prikazane (vidi tablicu 6.2) rezultate mjerenja visine 20-godišnjih mladića. Ako zbrajamo frekvencije u svakom razredu sa sumom frekvencija u razredima ispod njega, dobit ćemo rezultate prikazane u tablici 6.3.

TABLICA 6.3.
KUMULATIVNE FREKVENCIJE REZULTATA 135 MLADIĆA, PRIKAZANIH U TABLICI 6.2.

Razred	Kumul. frekv.
157-159	1
160-162	3
163-165	12
166-168	27
169-171	52
172-174	80
175-177	100
178-180	116
181-183	129
184-186	134
187-189	135

Prikažemo li te rezultate grafički na opisani način, dobit ćemo krivulju prikazanu na slici 6.10.



Slika 6.10. Krivulja kumulativne frekvencije rezultata prikazanih u tablici 6.3.

Iz slike možemo brzo čitati koliko ima rezultata ispod ili iznad (ili jednakih) kao neki određeni rezultat. Na primjer, ako nas zanima neki određeni rezultat X (vidi sliku), pa ako s tog rezultata povučemo s apscise paralelu s ordinatom sve do krivulje, a od krivulje paralelu s apscisom sve do ordinate, dobivamo željeni odgovor: u toj konkretnoj izmjerenoj grupi ima 90 rezultata, koji su manji (ili jednaki) od rezultata X, a 45 rezultata koji su veći od njega.

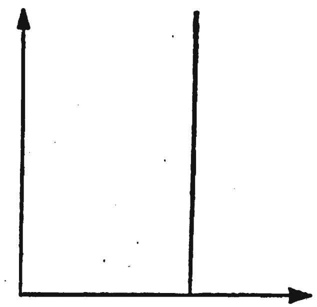
Da smo frekvencije imali u *relativnim* frekvencijama, dakle u postocima, onda bismo te iste podatke mogli davati u postocima. A kako je centralna vrijednost ona vrijednost iznad koje i ispod koje se nalazi točno po 50% rezultata, to bismo aproksimativno centralnu vrijednost mogli očitati sa krivulje relativne kumulativne frekvencije tako da s ordinate, kod oznake 50%, povučemo paralelu s apscisom do krivulje, i odande spustimo okomicu na apscisu, te tamo očitamo vrijednost C.

ZADACI ZA VJEŽBU

Izradite krivulju kumulativne frekvencije za 50 rezultata mjerenja vremena reakcije prikazanih na stranici 40.

7. NORMALNA RASPODJELA, NEKE DRUGE RASPODJELE
(i još ponešto o računanju vjerojatnosti)

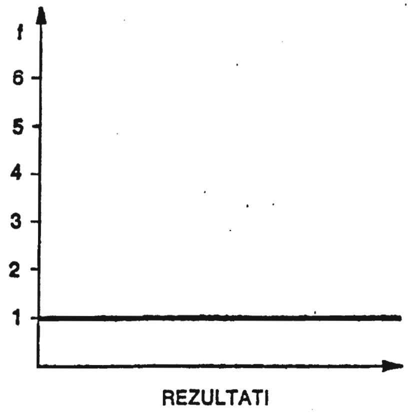
Kada bi *svi* rezultati nekog mjerenja bili potpuno istovjetni, onda bi to u grafičkom prikazu izgledalo kao na slici 7.1. Tendencija grupiranja rezultata ovdje je maksimalna.



Slika 7.1. Kada bi pri nekom mjerenju svi rezultati bili potpuno jednaki, onda bi to grafički izgledalo ovako

Naprotiv, kada bi kod nekog mjerenja svaki put dobili drukčiji rezultat, točnije rečeno kada bi *svaki rezultat* na nekoj skali bio dobiven samo *jedanput,* onda ne bi bilo nikakve tendencije grupiranja rezultata, pa bi oni grafički izgledali kao na slici 7.2.

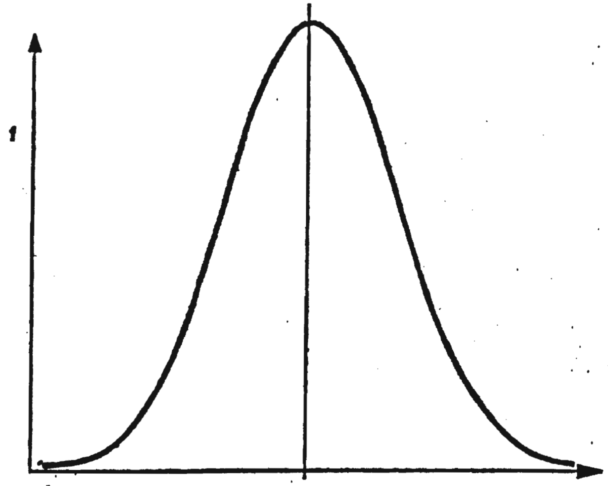
Međutim, mjereći različite pojave i karakteristike, mi nikada ne nailazimo ni na jedan od tih ekstrema, nego većinom dobivamo rezultate koji pokazuju *obje* ove tendencije, tj. i tendenciju *grupiranja* oko neke centralne vrijednosti i tendenciju *raspršenja* oko te srednje vrijednosti. Ako zamislite komad neke mekane mase koju



Slika 7.2. Kada bi pri nekom mjerenju svaki put dobili drugi rezultat, onda bi to grafički izgledalo ovako

istodobno razvlačite vertikalno (tendencija koncentriranja) i horizontalno (tendencija raspršivanja), onda bi se iz te mase stvorilo nešto slično onome što zapravo većinom i dobivamo kod mjerenja različitih pojava, tj. stvorila bi se raspodjela prikazana na slici 7.3.

Takva se raspodjela naziva *normalna raspodjela.* Krivulja koja prikazuje takvu raspodjelu naziva se *normalna krivulja*, a neki je — po matematičaru Gaussu, koji ju



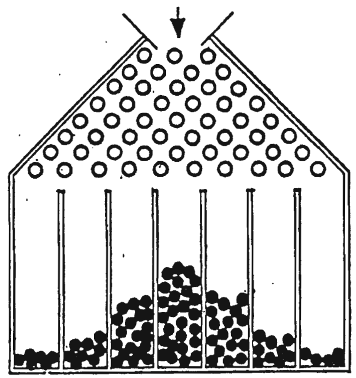
Slika 7.3. Često prikazom uzastopnog mjerenja neke pojave dobivamo rezultate koji pokazuju i tendenciju grupiranja oko jedne središnje vrijednosti i tendenciju raspršivanja oko te vrijednosti. Takva se raspodjela naziva normalna raspodjela

je, uz neke druge matematičare, matematički definirao — nazivaju i Gaussovom krivuljom, ili pak — prema njezinu karakterističnom obliku — "zvonastom" krivuljom.

Normalna raspodjela predstavlja jedan od osnovnih pojmova statističkog rezoniranja i statističkog "načina mišljenja" jer je ona *osnova za razumijevanje glavnih statističkih pojmova vjerojatnosti* (na primjer: vjerojatnost da ćemo slučajno dobiti neki rezultat koji je veći od aritmetičke sredine jednaka je vjerojatnosti da ćemo dobiti rezultat koji je manji od aritmetičke sredine; to proizlazi iz slike 7.3).

Iskustva iz nastave statistike redovno potvrđuju da netko tko *ne razumije* smisao normalne raspodjele (i sve ono što iz toga proizlazi), *ne može razumjeti ni mnoge druge statističke principe, a osobito ne one koji se izravno logički izvode iz normalne raspodjele.*

Stoga se vrijedi malo više pozabaviti normalnom raspodjelom i nekim zakonima njezinog postanka. U praktičnoj nastavi statistike u tu svrhu postoje različita zorna pomagala, a jedno od najpoznatijih sastoji se iz jedne koso položene plitke kutije, koja na gornjoj strani u sredini ima lijevak u koji se sipaju kuglice; ispod toga nalaze se ukucani čavlići koji sprečavaju da se kuglice bez smetnje spuštaju prema dnu kutije, a na dnu kutije su ogradice koje nam omogućuju kontrolu kamo je pojedina kuglica pala. (Vidi sliku 7.4.)



Slika 7.4. Galtonova "daska s čavlima" za demonstraciju postanka normalne raspodjele

Demonstracija postanka normalne raspodjele sastoji se u tome da se kuglice sipaju kroz lijevak, i one, udarajući putem u zapreke (čavliće), putuju prema dnu uređaja, gdje se slažu na karakterističan način, tj. u obliku normalne raspodjele: najviše ih ima u sredini, a prema krajevima u svakoj pregradici ima ih sve manje i manje.

To što kuglice sipamo u *sredini* gornjeg ruba kutije — to je tendencija *grupiranja* rezultata oko sredine; a čavlići koji ometaju kuglice, i u koje kuglice udaraju, predstavljaju tendenciju *raspršenja* rezultata.

Promatrajući ovaj demonstracijski uređaj za vrijeme jednog pokusa sipanja kuglica, promatraču postaje jasno da se u *najviše slučajeva* "sile" koje "odvlače" pojedinu kuglicu od njezina puta prema sredini *međusobno ukidaju:* neki čavlići skrenuli su put kuglice ulijevo, a neki udesno, pa tako kuglica ipak konačno pada u sredinu dna uređaja. Takvih kuglica — kod kojih su se faktori djelovanja različitog smjera ukinuli — ima najviše. No kod nekih kuglica udaranje u prepreke završilo je tako što je kuglica konačno pretežno otišla u jednom ili u drugom smjeru. I, napokon, samo kod *malog* broja kuglica dogodilo se to da je *svaki udarac u prepreku doveo do odbijanja kuglice u istom smjeru* — i to su one kuglice koje nalazimo u ekstremnim lijevim ili desnim pregradicama na dnu.

Normalna je raspodjela, dakle, rezultanta istodobnog djelovanja golemog broja faktora koji djeluju u jednom ili u drugom smjeru, po slučaju se najčešće manje ili više ukidaju, ali — isto tako po slučaju — kadšto se i zbrajaju, te tako dolazi i do značajno aberantnih rezultata.

Glavni su uvjeti da kod nekog mjerenja dobijemo normalnu raspodjelu ovi:

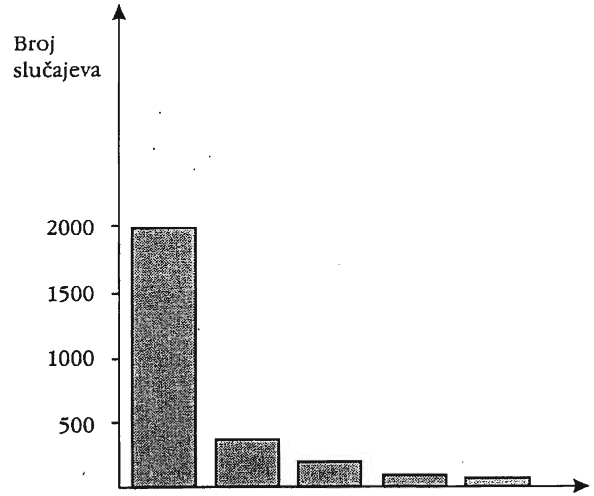
1. Da se ono što mjerimo, stvarno raspoređuje po normalnoj raspodjeli. Prema prilično raširenom mišljenju sve, ili gotovo sve, što u prirodi mjerimo raspoređuje se prema normalnoj raspodjeli. To, međutim, nije točno.
2. Da imamo velik broj rezultata (mjerenja).
3. Da su sva mjerenja provedena jednakom metodom i u što sličnijim vanjskim prilikama.
4. Skupina na kojoj obavljamo mjerenja mora biti *homogena* po ostalim svojstvima, a *heterogena* (neselekcionirana) po onom svojstvu koje mjerimo. Na primjer, mjerimo li visinu ljudi, skupina koju mjerimo treba da bude izjednačena (homogena) po spolu, godinama i eventualno još nekim drugim svojstvima, ali ne smije biti selekcionirana po visini, tj. treba izmjeriti *sve* članove grupe, a ne samo one visoke ili one niske (dakle skupina po visini mora biti heterogena).

Što se tiče prvog uvjeta (tj. da ono što mjerimo mora i u prirodi biti raspodijeljeno po normalnoj raspodjeli), već je u knjizi spomenuto da npr. bilirubin u krvi daje asimetričnu raspodjelu, da dijametar ljudskog srca daje "bimodalnu" raspodjelu (bi-modalnu, dva moda, dakle krivulja s dva vrha). Evo još jednog primjera. Na slici 7.5. prikazana je distribucija vremena dolaženja na posao izmjerena na 50 radnika jednog poduzeća u ukupno 2545 mjerenja.

Kako se iz slike vidi, u najviše slučajeva (78,6%) radnici su stizali na posao na vrijeme, u 13,8% kasnili su u rasponu od 1/2 sata, itd., sve do 2 sata zakašnjenja, gdje je nađeno 0,1% slučajeva.

Među pojave koje se ne distribuiraju po normalnoj raspodjeli, pripada donekle i *težina ljudi:* distribucija ljudske težine lagano je asimetrična nadesno (što znači da je desni kraj distribucije lagano "razvučen"); ali ta je asimetričnost toliko blaga da se ipak s distribucijom težine može postupati kao s normalnom raspodjelom. No zbog jednoga drugog razloga vrijedno je težinu spomenuti na ovom mjestu: zanimljivo je da se u nekim statističkim udžbenicima u kojima se spominje ta asimetrična distribucija težine, mogu naći dosta kompleksna tumačenja zbog kojih bioloških razloga je tome tako.

Slika 7.5.



	TOČNO	do 1/2 s	1/2-1 s	1-1,5	1,5-2
Broj	2001	351	150	38	5
%	78,6	13,8	5,9	1,5	0,1

Slika 7.5. Distribucija vremena dolaska na posao

Međutim, razlog asimetričnoj distribuciji težine mnogo je jednostavniji i sastoji se u tome: ako se *visina* ljudi distribuira po normalnoj raspodjeli, onda se težina *ne može* distribuirati po normalnoj raspodjeli!

Evo dokaza. Zamislimo jednu skupinu od, recimo, 128 kocaka, koje su po dužini bridova raspodijeljene po normalnoj raspodjeli prikazanoj na slici 7.6.

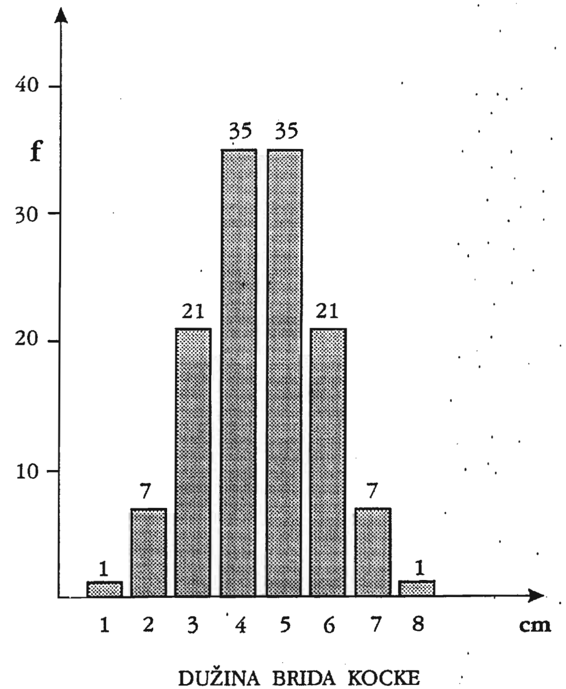
Ako nas zanima *volumen* tih kocaka, dobit ćemo distribuciju prikazanu na slici 7.7.

Kako je volumen kocke ujedno i njezina težina, to znači da se težina kocaka, kojima je visina brida distribuirana po normalnoj raspodjeli, ne može distribuirati normalno.

Čovjek naravno nije kocka, ali je ipak u osnovi *proporcionalno* građen, što će reći da su viši ljudi u prosjeku i krupniji (deblji) od niskih, pa će prema tome i čovjekova težina (budući da mu se visina distribuira po normalnoj raspodjeli) dati laganu asimetričnu distribuciju.

Drugi uvjet, tj. da broj mjerenja bude dovoljno velik, posve je razumljiv: kod malog broja mjerenja neke pojave, pa bila ona i idealno normalno distribuirana u nekoj populaciji, pukim slučajem možemo dobiti rezultate koji znatno odstupaju od normalne raspodjele. Uostalom, pokušajte iz 50 naših rezultata mjerenja vremena reakcije nasumce izabrati, na primjer, samo 10, pa će se vrlo lako dogoditi da distribucija dobivenih rezultata jedva nalikuje normalnoj distribuciji.

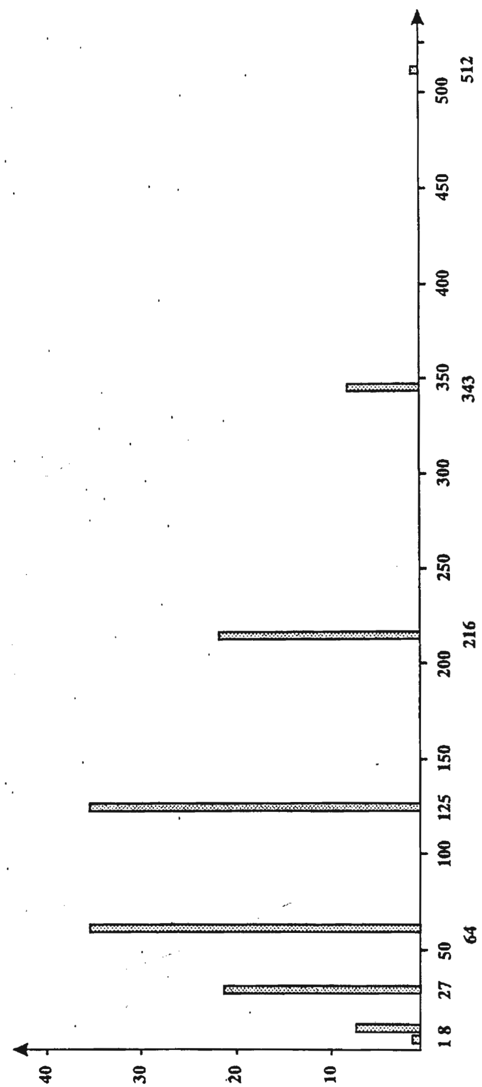
Slika 7.6.



Slika 7.6. Hipotetička raspodjela dužine bridova 128 kocki

Treći uvjet, tj. da su sva mjerenja provedena istom metodom i u što sličnijim vanjskim prilikama, također je lako shvatljiv. Prilikom jednoga biometrijskog mjerenja djece jedna od mjerilačkih ekipa nije bila dovoljno upućena u postupak te je — za razliku od ostalih ekipa koje su djecu vagale bez odjeće — obavila mjerenje težine djece u donjem rublju i cipelama. Naravno da je posljedica bila — bimodalna raspodjela, koja je nastala zbrajanjem ("superpozicijom") dviju normalnih raspodjela, koje nisu bile na *istom mjestu:* jedna je normalna raspodjela bila na nižim vrijednostima (djeca bez odjeće), a druga na nešto višim vrijednostima (djeca u donjem rublju i cipelama). Ili, pretpostavimo da želimo na većoj skupini ljudi izmjeriti vrijeme reakcije na slušne podražaje, ali se pri mjerenju služimo dvama uređajima, od kojih jedan daje podražaje jakog, a drugi podražaje slabog intenziteta. Budući da vrijeme reagiranja ovisi i o intenzitetu podražaja, to ćemo kao rezultat za cijelu skupinu dobiti krivulju koja neće biti normalna raspodjela, nego vjerojatno opet neka bimodalna krivulja.

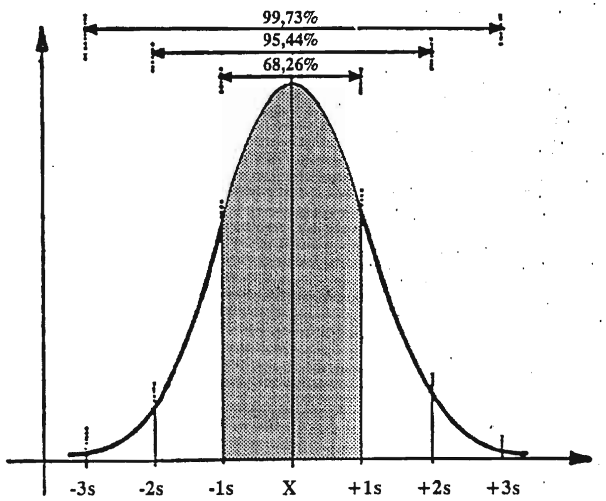
Četvrti uvjet čini se na prvi pogled najkompleksniji, jer on zahtijeva da skupina, koju mjerimo, istodobno bude i *homogena* i *heterogena*. No ti zahtjevi *ne odnose se na iste stvari:* ako, primjerice, želimo mjeriti slušnu osjetljivost, onda unaprijed moramo odlučiti hoćemo li mjeriti mlade ili starije ljude, jer ako ih mjerimo zajedno,



Slika 7.7. Distribucija volumena (težine) 128 kocaka iz slike 7.6.

nećemo dobiti normalnu raspodjelu jednostavno zato što se mladi od starih ispitanika *značajno razlikuju* po slušnoj osjetljivosti, pa će dobivena raspodjela opet biti rezultanta dviju (normalnih) raspodjela koje se razlikuju po poziciji, tj. po položaju aritmetičke sredine; dakle u svim onim svojstvima, koja *mogu ili bi mogla* imati utjecaja na ono što mjerimo, skupina treba biti homogena. (Zato se u eksperimentalnom radu toliko pazi na to da eksperimentalna i kontrolna skupina budu "izjednačene" u svim ostalim faktorima, osim u onom koji upravo istražujemo.) No, istodobno, kada bi nas zanimalo prosječno znanje učenika neke škole, pa bismo uzorak tih učenika podvrgnuli ispitivanju pomoću nekog testa znanja — a onaj tko je učenike birao, izabrao je u uzorak, pretpostavimo, samo najbolje đake — distribucija koju bismo dobili, vjerojatno također ne bi bila normalna, jer *selekcija* uzorka obično narušava normalitet distribucije.

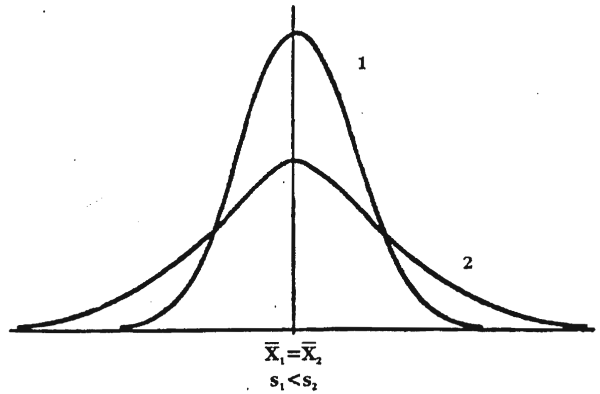
Kao što smo već spomenuli, kod normalne raspodjele u intervalu $\overline{X} \pm 1s$ nalazi se 68,26% svih rezultata, a u intervalu $\overline{X} \pm 2s$ 95,44% rezultata, a u intervalu $\overline{X} \pm 3s$ 99,73% svih rezultata. Sada kada imamo normalnu distribuciju i grafički prikazanu, to je mnogo razumljivije nego što je to bilo prije kada smo to spomenuli u poglavlju o mjerama varijabilnosti. Intervali koji su obuhvaćeni, ako aritmetičkoj sredini na jednu i drugu stranu dodamo jednu, dvije ili tri standardne devijacije, prikazani su na slici 7.8.



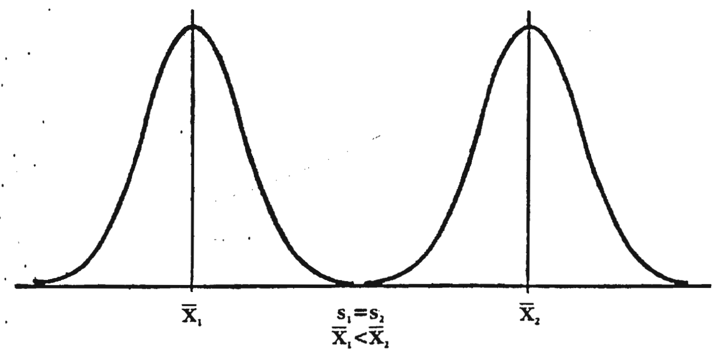
Slika 7.8. Normalna raspodjela i intervali koji su obuhvaćeni ako aritmetičkoj sredini dodamo lijevo i desno jednu, dvije ili tri standardne devijacije

Već smo rekli da je normalna raspodjela (kao uostalom i sve druge raspodjele) matematički posve točno definirana. No mi u formulu normalne raspodjele nećemo

ulaziti, jer ona nije baš jako jednostavna, a osim toga za naše svrhe nije nam niti potrebna. Reći ćemo samo to da je jedna od osnovnih karakteristika normalne raspodjele to da se tzv. "mjesto infleksije" (tj. mjesto gdje krivulja iz konveksne prelazi u konkavnu) nalazi iznad $\pm 1s$, kao i to da je normalna raspodjela potpuno definirana ako joj znamo aritmetičku sredinu i standardnu devijaciju. Iz toga slijedi da postoje normalne raspodjele vrlo različite "širine": od uskih ("leptokurtičnih") do vrlo širokih ("platikurtičnih"). Na slici 7.9. prikazan je slučaj dviju normalnih raspodjela koje imaju jednake aritmetičke sredine, ali se razlikuju po standardnoj devijaciji, a slika 7.10. prikazuje dvije normalne raspodjele s različitim aritmetičkim sredinama, ali jednakim standardnim devijacijama.



Slika 7.9. Aritmetičke sredine se ne razlikuju, ali se razlikuju standardne devijacije



Slika 7.10. Standardne devijacije se ne razlikuju, ali se razlikuju aritmetičke sredine

A sada razmotrimo na jednom drugom primjeru, koji je jednostavniji od upotrebe Galtonove "daske s čavlićima", nastajanje normalne raspodjele. (U strogom smislu to je zapravo jedna druga raspodjela, koju ćemo malo kasnije imenovati, ali je ona vrlo dobra aproksimacija normalnoj raspodjeli.)

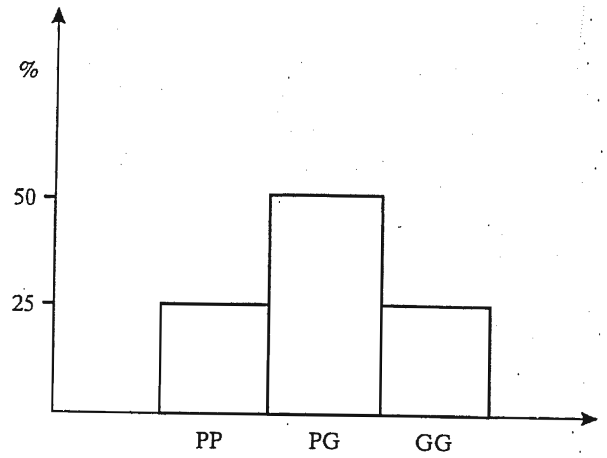
Ako bacamo 2 komada novca, pa gledamo na koju su stranu pali, postoje praktički tri mogućnosti ishoda bacanja:

1. mogućnost: na oba novca "pismo",
2. mogućnost: na jednom novcu "pismo", a na drugom "glava",
3. mogućnost: na oba novca "glava".

Zašto onda kod mnogo bacanja ipak najčešće dobivamo kombinaciju br. 2, tj. "pismo-glava"? To je lako protumačiti ako pažljivo pogledamo koje se sve kombinacije zapravo mogu dogoditi. To su ove kombinacije:

1. kombinacija:	na I. novcu	PISMO,	na II. novcu	PISMO
2. "	" " "	PISMO,	" " "	GLAVA
3. "	" " "	GLAVA,	" " "	PISMO
4. "	" " "	GLAVA,	" " "	GLAVA

Svaka od tih kombinacija jednako je vjerojatna pa prema tome, kako ih imamo 4, vjerojatnost svake kombinacije je 25%. Budući da 2. i 3. kombinacija ***praktički*** znače isto, to ćemo kombinacije PISMO — GLAVA (ili GLAVA — PISMO) dobiti otprilike u 50% slučajeva, a kombinaciju PISMO — PISMO, i kombinaciju GLAVA — GLAVA, svaku otprilike u 25% slučajeva, kao što nam to pokazuje slika 7.11.



Slika 7.11. Vjerojatnost pojavljivanja pojedinih ishoda kod bacanja dva komada novčića

Ako — umjesto dva novčića — bacamo 4 komada, postoji šesnaest mogućih ishoda bacanja. Ti su ishodi prikazani u tablici 7.1. i na slici 7.12.

TABLICA 7.1.
SVI MOGUĆI ISHODI KOD BACANJA 4 KOMADA NOVČIĆA

	I. novac	II. novac	III. novac	IV. novac	
1.	P	P	P	P	(4P)
2.	P	P	P	G	
3.	P	P	G	P	
4.	P	G	P	P	(3 P, 1 G)
5.	G	P	P	P	
6.	G	G	P	P	
7.	G	P	P	G	
8.	P	P	G	G	
9.	P	G	G	P	(2 P, 2 G)
10.	P	G	P	G	
11.	G	P	G	P	
12.	P	G	G	G	
13.	G	P	G	G	
14.	G	G	P	G	(1 P, 3 G)
15.	G	G	G	P	
16.	G	G	G	G	(4 G)

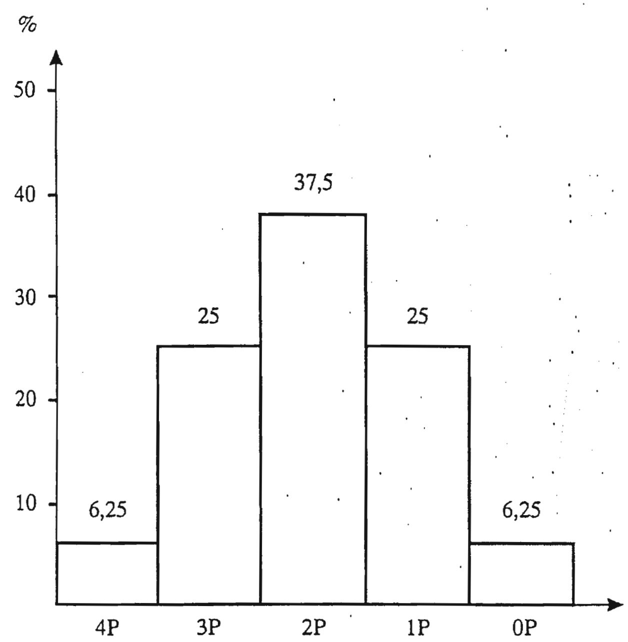
Budući da svaka od tih 16 kombinacija ima jednaku vjerojatnost (u tom slučaju vjerojatnost $p = 1/16$), to ćemo kombinaciju 4 P dobiti otprilike u 6,25% slučajeva, kombinaciju 3 P, 1 G u oko 25% slučajeva, kombinaciju 2 P, 2 G u oko 37,5%, 1 P, 3 G u oko 25%, i, konačno, 4 G u oko 6,25% slučajeva, kao što je to prikazano na sl. 7.12.

Vjerojatnost pojedinih kombinacija možemo izračunati pomoću tzv. binomne raspodjele,

$$(p+q)^n, \tag{7.1}$$

pri čemu p = vjerojatnost da će se nešto dogoditi (na primjer "glava"), q = vjerojatnost da se to neće dogoditi (dakle dogodit će se "ne glava", tj. "pismo"), a eksponent n je broj faktora (u našem slučaju broj komada novca). Vjerojatnost da će se nešto dogoditi, plus vjerojatnost da se to neće dogoditi, uvijek je 100% (ili vjerojatnost = $P = 1$), pa je prema tome $(p+q)$ uvijek = 1.

Ako u našem primjeru s 2 komada novca postavimo jednadžbu $(p+q)^n$, pa p i q zamijenimo izrazima G (glava) i P (pismo), dobivamo $(G+P)^2 = G^2 + 2GP + P^2$, što "prevedeno" znači: jedanput 2 "glave" + dva puta "glava-pismo" + jedanput 2 "pisma".



Slika 7.12. Vjerojatnost pojavljivanja pojedinih ishoda kod bacanja četiri komada novčića

U primjeru s 4 komada novca imamo analogno: $(p+q)^4 = (G+P)^4 = G^4+4G^3P+$ $+6G^2P^2+4GP^3+P^4$, što znači: jedanput 4 "glave", + četiri puta 3 "glave" i 1 "pismo", + šest puta 2 "glave" i 2 "pisma", + četiri puta 1 "glava" i 3 "pisma" + jedanput 4 "pisma".

Uz pomoć tzv. "Pascalova trokuta" moguće je utvrditi ove različite kombinacije i bez računanja. Pascalov trokut za vrijednost od N = 1 do 10 prikazan je u tablici 7.2. (Po samom imenu tog trokuta jasno je da je nazvan po matematičaru Blaise Pascalu, no danas je poznato da su već matematičari u staroj Kini posjedovali sistem tog trokuta.)

Ako dobro proučimo Pascalov trokut, vidjet ćemo da je u biti vrlo jednostavan: svaki idući red dobiven je tako da su sumirana dva broja koji su lijevo i desno iznad njega. Prema tome, možemo i sami izraditi nastavak Pascalova trokuta za N koji je veći od 10.

TABLICA 7.2.
PASCALOV TROKUT

N	
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1
8	1 8 28 56 70 56 28 8 1
9	1 9 36 84 126 126 84 36 9 1
10	1 10 45 120 210 252 210 120 45 10 1

Iz tablice čitamo da su npr. kod 10 komada novca, ako ih bacamo 1024 puta (1024 je suma svih brojeva u tablici uz N = 10) očekivane frekvencije ove:

Kombinacija		Očekivane frekvencije	Vjerojatnost
10 G	0 P	1	1/1024
9 G	1 P	10	10/1024
8 G	2 P	45	45/1024
7 G	3 P	120	120/1024
6 G	4 P	210	210/1024
5 G	5 P	252	252/1024
4 G	6 P	210	210/1024
3 G	7 P	120	120/1024
2 G	8 P	45	45/1024
1 G	9 P	10	10/1024
0 G	10 P	1	1/1024

(N a p o m e n a: Uz pomoć Pascalova trokuta mogu se lako izračunati već spominjani binomni koeficijenti [str. 37 i tablica T]: ako n očitamo u redu Pascalova trokuta, a x na položaju $x+1$ u tom istom redu, dobivamo jednak rezultat, kao i iz tablice T.

Naprimjer, binomni koeficijent $\binom{8}{5}$ (što je isto kao i $\frac{8!}{5!(8-5)!}$) očitavamo u osmom redu trokuta, na *položaju* 5 + 1, a rezultat iznosi 56, što je dakako isti broj kao i u tablici T).

Uz pomoć Pascalova trokuta možemo na trenutak pogledati još neke pogreške koje nam se katkad događaju pri procjeni pojedinih vjerojatnosti. Na primjer, kada bi netko postavio pitanje koja je vjerojatnost da ćemo kod bacanja 10 komada novčića dobiti kao rezultat 5 "glava" i 5 "pisama", znatan broj ljudi odgovorit će da je vjerojatnost 50%. To međutim *nije točno*, a pogreška koja je pri tome učinjena, proistječe iz toga što ljudi katkada "miješaju" pojam "najvjerojatniji ishod" s pojmom "vjerojatnost 50 : 50". Ljudi otprilike ovako rezoniraju: "Ako bacim 10 novčića, polovica će pasti na 'glavu' na polovica na 'pismo', jer su šanse za 'glavu' i 'pismo' jednake." Zvuči logično, zar ne?

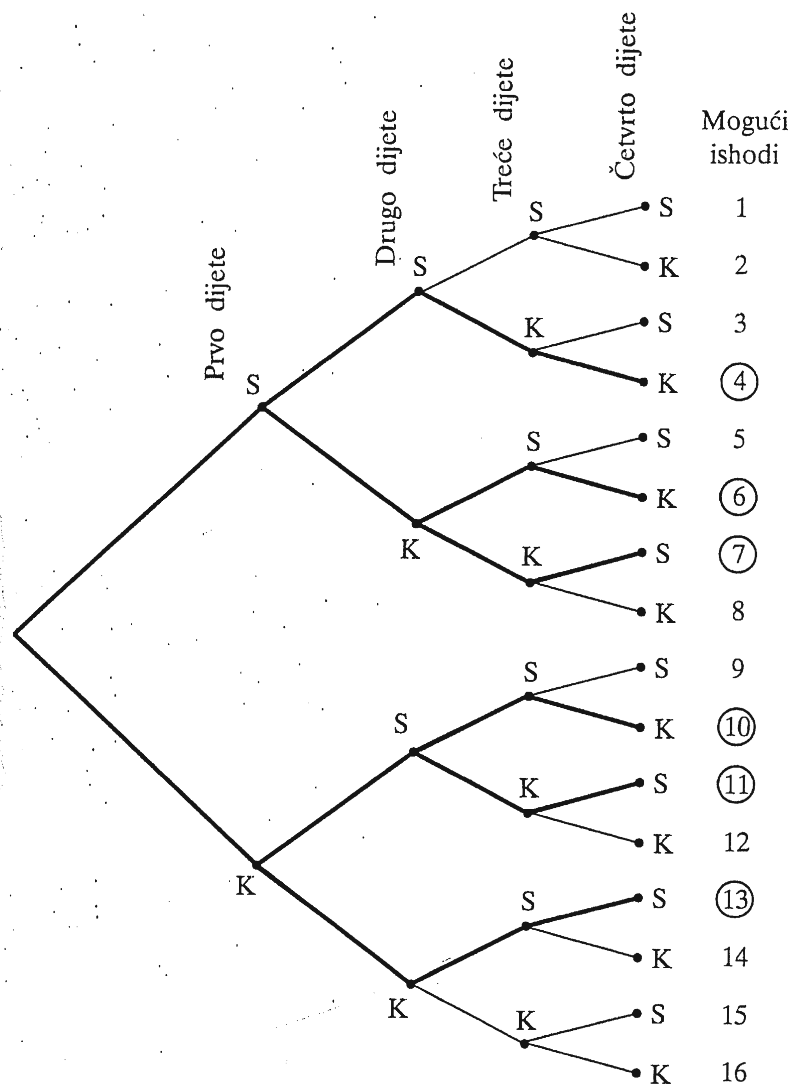
No, činjenica je da šansa 50 : 50 vrijedi samo kod bacanja *dva* novčića. Već kod bacanja 4 novčića, kako se to vidi iz slike 7.12. i tablice 7.1, vjerojatnost da će pasti 2 "glave" i 2 "pisma" nije 50%, već 37,5%. Pa tako i kod bacanja 10 komada novčića, vjerojatnost da će pasti 5 "glava" i 5 "pisama" jest doduše *najveća* među svim ostalim ishodima, ali ona ne iznosi 50% ($p = 0,5$), nego (vidi Pascalov trokut!) oko 25% ($p = 252/1024 = 0,246$).

Dakle, ako bacamo 2 komada novčića, pa smo se kladili da će jedan pasti na "glavu" a drugi na "pismo", šansa da ćemo dobiti okladu iznosi 0,5. Ali ako pet puta zaredom bacamo 2 novčića (što je isto kao da jedanput bacimo 10 komada!), pa ako se također kladimo da ćemo dobiti pet puta "glavu" i pet puta "pismo", svi su izgledi da ćemo izgubiti okladu, jer nam je šansa ispod 25%, tj. točnije $p = 0,246$ (iz 252/1024)!

Postoji još jedan vrlo jednostavan i slikovit način za izračunavanje različitih ishoda jednake vjerojatnosti, a to je tzv. ***"stablo vjerojatnosti"***, prikazano na slici 7.13. Slično prijašnjem primjeru s bacanjem novčića, pretpostavimo da jedna obitelj želi imati *četvero djece*. Koja je vjerojatnost da će imati dva sina i dvije kćeri?

Sada naravno već znate da to nije 50% a to se vrlo lako (kao i kod prethodnog primjera) može ustanoviti iz slike 7.13: prvo dijete može biti ili sin ili kći; ako je prvo dijete sin, drugo može ponovno biti ili sin ili kći; to isto vrijedi i u slučaju da je prvo dijete bila kći. Razrađujući tim načinom "stablo vjerojatnosti" sve dalje, dolazimo do 4. djeteta. Kako iz slike vidimo, tu već postoji *16 mogućih ishoda* (vidi i zbroj 4. reda u Pascalovu trokutu!), a sve mogućnosti za 2 sina i 2 kćeri u crtežu su podebljane, pa vidimo da od 16 mogućih ishoda postoji samo šest njih u kojima je taj uvjet postignut; to su ovi ishodi: S(in) S(in) K(ći) K(ći), nadalje S K S K, S K K S, K S S K, K S K S, i K K S S. (To naravno možemo odčitati i iz Pascalova trokuta). Vjerojatnost za taj ishod (2S + 2K) iznosi dakle 6/16, ili 0,375. Ako bismo tražili točno određeni *redoslijed* (npr. S K K S), onda samo *jedan* od postojećih 16 mogućih ishoda udovoljava tom uvjetu (vidi sliku). To naravno točno odgovara nama već poznatom "zakonu multiplikacije" (vidi str. 33), tj. $1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 = 1/16$. A u primjeru u kojem redoslijed nije važan, vrijedi naravno zakonitost, opisana na str. 33, tj. treba *zbrajati* vjerojatnosti svakog povoljnog ishoda: S S S K = 1/16 plus S K S K = 1/16 plus.......itd = 6/16 (= 0,375).

Osim spomenute formule (7.1) postoji i druga, naoko kompleksnija formula za izračunavanje vjerojatnosti pojavljivanja nekog ishoda kod binomne raspodjele, ali je ona ipak znatno jednostavnija pod pretpostavkom da posjedujemo džepno elektronsko računalo.



Slika 7.13. Prikazivanje vjerojatnosti pojedinog ishoda uz pomoć "stabla vjerojatnosti". Postoje 6 od 16 mogućih ishoda da neka obitelj s četvero djece ima 2 sina i 2 kćeri. Vjerojatnost toga ishoda iznosi 6/16 = 0,375. Za jedan određeni *redoslijed* (npr. k-s-s-k) vjerojatnost je 1/16.

Ta formula glasi:

$$P_k = \frac{n!}{k!(n-k)!} \cdot p^k \cdot q^{n-q}, \qquad (7.2)$$

pri čemu

P = vjerojatnost da će se nešto dogoditi,
q = vjerojatnost da se to neće dogoditi,
k = broj "pogodaka" (npr. 2 "šestice" kod bacanja kocke),
n = veličina uzorka (npr. 3 igraće kocke).

(N a p o m e n a. Čitalac koji je u međuvremenu zaboravio što znači znak ! u formulama, neka pogleda stranicu 28).

Evo primjera: Kolika je vjerojatnost da će kod bacanja 3 igraće kocke pasti 2 "šestice" (ili — što je isto — kod tri bacanja jedne kocke pasti 2 "šestice")?

$$P_2 = \frac{3!}{2!(3-2)!} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^{(3-2)} =$$

$$= \frac{3\cdot 2\cdot 1}{(2\cdot 1)\cdot 1} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} =$$

$$= \frac{6}{2\cdot 1} \cdot \frac{1}{36} \cdot \frac{5}{6} = \frac{30}{432} = 0,069.$$

Dakle, vjerojatnost da će kod bacanja 3 igraće kocke pasti 2 "šestice" iznosi oko 6,9%.

Postoje naravno i tablice binomne raspodjele iz kojih se to izravno može očitati, ali su binomne tablice vrlo opsežne jer moraju uzimati u obzir tri elementa: općenitu vjerojatnost pojavljivanja neke pojave (npr. kod bacanja novca vjerojatnost "glave" je 0,5), veličinu uzorka (npr. 4 komada novca) i željeni ishod (npr 2 "glave"). Zbog mnoštva mogućih vjerojatnosti, tablice sadrže samo neke od njih, tj. 5%, 10%, 20% ... 90% i 95% pa *neke* probleme ne možemo pomoću njih rješavati: vjerojatnost nekog ishoda na igraćoj kocki iznosi 1/6 = 16,7%, pa probleme kocke nije moguće izravno očitavati iz binomne tablice. No razmotrit ćemo dva druga primjera, kako bismo samo segmente tablice mogli protumačiti (tako da onaj čitalac kome su one možda u nekoj situaciji prijeko potrebne, znade kako da se njima služi).

Prvi primjer. Ako u nekoj populaciji produkata ima 10% defektnih komada, koja je vjerojatnost da ćemo, uzevši uzorak veličine 4, dobiti 3 komada defektna?

Dio tablice binomne raspodjele prikazan je u tablici 7.3. Iz tablice čitamo da je vjerojatnost da su 3 komada u uzorku defektna, relativno mala, tj. 0,4% (dakle kod oko 1 000 uzoraka veličine 4 našli bi u prosjeku četiri uzorka u kojima bi 3 komada bila defektna).

(N a p o m e n a. To isto izračunali bismo pomoću formule (7.2) ovako:

$$P_3 = \frac{4!}{3!(4-3)!} \cdot 0,1^3 \cdot 0,9^1 = \frac{24}{6\cdot 1} \cdot 0,001 \cdot 0,9 = 0,0036, \text{ dakle oko } 0,4\%)$$

TABLICA 7.3.
DIO TABLICE BINOMNE RASPODJELE (RADI DEMONSTRACIJE)

Veličina uzorka n	Broj pogodaka r	Vjerojatnosti 0,05	0,10	0,50 .....
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
4	0	.	0,656	0,062
.	1	.	0,292	0,250
.	2	.	0,049	0,375
.	3	.	0,004	0,250
.	4	.		0,062
.	.	.	.	.
.	.	.	.	.

Drugi primjer. Koja je vjerojatnost da ćemo kod bacanja 4 komada novčića dobiti dva puta "glavu"?

Upotrebom tablice čitamo da je vjerojatnost $p = 0,375$ (tj. 37,5%), a točno toliko imamo napisano i na srednjem stupcu histograma na slici 7.12.

Iz tablice 7.3. ujedno vidimo da je vjerojatnost da nećemo dobiti nijednu "glavu" nešto više od 6%, i da ćemo dobiti samo jednu "glavu" 25%.

Vratimo se ponovno Pascalovu trokutu! Povećavajući broj novčića koje bacamo u zrak, dobili bismo konačno praktički potpuno pravilnu zvonastu "krivulju slučaja", dakle normalnu raspodjelu. No razlika između binomne i normalne raspodjele (koja je više teoretske nego praktične prirode) jest u tome što binomna raspodjela nastaje kombinacijom faktora, kojih je pojavljivanje uvijek jednako vjerojatno (npr. 50% ili 10% ili bilo koja druga vjerojatnost), dok je kod normalne raspodjele situacija nešto drukčija. Iako smo njezin postanak već pokušali opisati, uzmimo još jedan primjer, ovaj put primjer sličan binomnoj raspodjeli, dakle primjer bacanja novčića.

Zamislimo da imamo mnogo komada novčića koji *nisu ispravni*, nego je svaki novčić po slučaju *svinut*, pa oko polovice novčića ima veću vjerojatnost (svaki novčić drukčiju) da padne na "glavu", a također oko polovice da padne na "pismo". Ako takve novčiće bacamo u zrak i promatramo kako su pali, ponovno ćemo dobiti krivulju rezultata koja će biti jednaka krivulji binomne raspodjele kada je N veliki broj!

— — —

Osim normalne i binomne raspodjele postoji još mnogo drugih raspodjela. Jednu od njih još ćemo u ovoj knjizi spominjati: to je Poissonova raspodjela, koja je također slučajna raspodjela, samo — za razliku od normalne raspodjele — slučajna raspodjela *vrlo rijetkih događaj* (dok je normalna raspodjela raspodjela događaja, kojima se vjerojatnost kreće u blizini 50%). Spominjat ćemo još i F-raspodjelu, hi-kvadrat raspodjelu i druge, ali nema smisla time sada čitaoca opterećivati. Kada dođe na njih red, te će raspodjele biti razumljive.

Recimo na kraju još nekoliko riječi o zanimljivoj povijesti nastanka spoznaja o o normalnoj raspodjeli. Posebno je za te podatke zaslužna američka statističarka Helen Walker, koja je publicirala nekoliko radova o toj temi, i evo njezinih glavnih podataka (iz Walker H. and Lev J. Elementary statistical methods, vidi literaturu).

Matematičar Abraham De Moivre (1667-1754), koji je kao izbjeglica živio u Londonu, prehranjujući se djelomice rješavanjem različitih problema zakona vjerojatnosti za bogate kockare, ustanovio je da distribucija rezultata bacanja novčića, kada je broj novčića velik (kako vidimo, počelo je od binomne raspodjele!), poprima jedan definitivni oblik, koji je on 1733. uspio izraziti jednadžbom, što ju je demonstrirao svojim prijateljima, ali o čemu nisu objavljeni službeni radovi.

Nekako istovremeno jedan švicarski matematičar, Jacques Bernoulli (1654-1705) upozorio je da bi teorija vjerojatnosti mogla biti od koristi u ekonomiji i društvenim znanostima, no smrt ga je spriječila da radi na tom pitanju, pa je tek 1713, dakle 8 godina nakon njegove smrti, objavljena (zaslugom njegova nećaka) njegova knjiga "Ars conjectandi", što bismo mogli prevesti kao "Umjetnost hazarda", no ta publikacija ipak nije pobudila interes ondašnjih matematičara.

Prošlo je stoga gotovo puno stoljeće do vremena, kada su Laplace (1749-1827) u Francuskoj i Gauss (1777-1858) u Njemačkoj nezavisno jedan od drugoga, a vjerojatno i ne znajući za De Moivreove rezultate, matematički definirali "krivulju vjerojatnosti" ili "krivulju pogrešaka". Ta je krivulja kasnije nazvana *Gaussovom krivuljom*, jer se u početku smatralo da ju je Gauss prvi opisao i matematički definirao. No 1924. godine je Karl Pearson, poznati statističar, otkrio dotad nepoznate rukopise De Moivrea.

Bernoullijevu misao da se ta krivulja može primijeniti ne samo na "pogreške mjerenja" u fizici, već i na ostale podatke mjerenja, izgleda da je definitivno formulirao i objavio belgijski statističar Adolphe Quetelet (1796-1874), koji je prvi upozorio da se statistika može koristiti i u bilo kojem drugom području mjerenja, dakle i kod mjerenja "mentalnih i moralnih" svojstava.

ZADACI ZA VJEŽBU

1. Upotrebom Pascalova trokuta nađite kolika je vjerojatnost da će kod bacanja 9 komada novčića pasti 2 "pisma" (ili 7 "glava").
2. Promatrajući sliku 7.8, ustanovite koja je vjerojatnost da ćemo slučajnim izvlačenjem rezultata iz kojih se sastoji nacrtana normalna distribucija, izvući rezultat koji je po vrijednosti:
   a) negdje između aritmetičke sredine i $+1\,s$,
   b) manji od $-2\,s$,
   c) veći od $+1\,s$ ali manji od $+2\,s$.
3. Koristeći tablicu 7.3, očitajte koja je vjerojatnost da u uzorku veličine $N = 4$, a koji uzimamo iz populacije u kojoj je 10% defektnih komada, ne nađemo ni jedan defektni komad.

8. POLOŽAJ POJEDINOG REZULTATA U GRUPI

8.1. z-VRIJEDNOSTI

Budući da nam aritmetička sredina i standardna devijacija potpuno definiraju raspodjelu nekih rezultata, to je za svaki rezultat moguće izračunati *na koji dio standardne devijacije* on pada, a kad znamo taj podatak, znamo odmah i koliko imamo rezultata *većih* i koliko *manjih* od toga rezultata. Na taj način možemo točno odrediti *položaj* pojedinca u grupi.

Primjer. Mjereći visinu veće grupe odraslih ljudi, dobili smo ove rezultate:

$$\overline{X} = 170,0 \text{ cm.}$$
$$s = 10,0 \text{ cm}$$

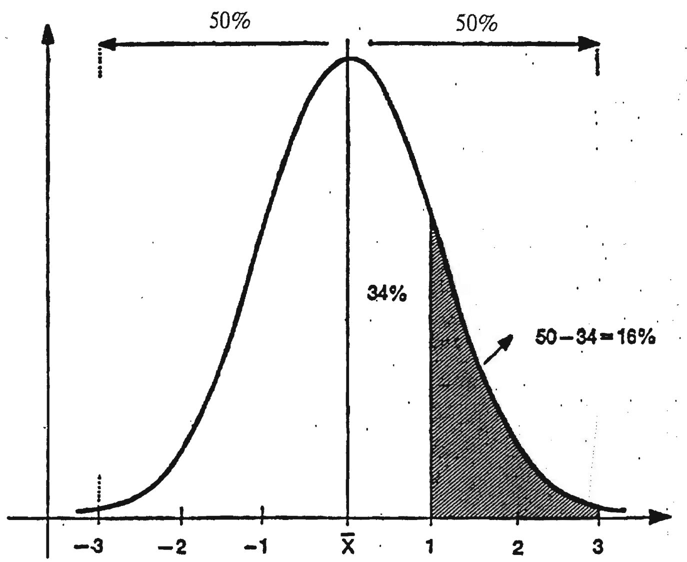
Kamo pripada ispitanik X koji je visok 180 cm? Drugim riječima, koliki postotak tih ljudi su viši od njega? Najprije moramo ustanoviti na koji dio standardne devijacije pada rezultat $X = 180$ cm. To ćemo ustanovit tako da najprije nađemo razliku između toga rezultata i aritmetičke sredine, i da onda tu razliku podijelimo sa standardnom devijacijom. Drugim riječima, ovim ćemo postupkom rezultat izraziti u *dijelovima* standardne devijacije:

$$z = \frac{X - \overline{X}}{s}. \tag{8.1}$$

U našem slučaju dakle imamo: $\dfrac{180 - 170}{10} = \dfrac{10}{10} = 1,0.$

Naš rezultat od 180 pada dakle točno na $+1\,s$. Budući da znamo da $\overline{X}$ *plus* $1\,s$ obuhvaća oko 34% rezultata, to nam do kraja krivulje na desnoj strani ostaje još 50% - 34% = oko 16% rezultata. Dakle, 16 posto ispitanika ove grupe viši su od 180 cm. (Vidi sl. 8.1.)

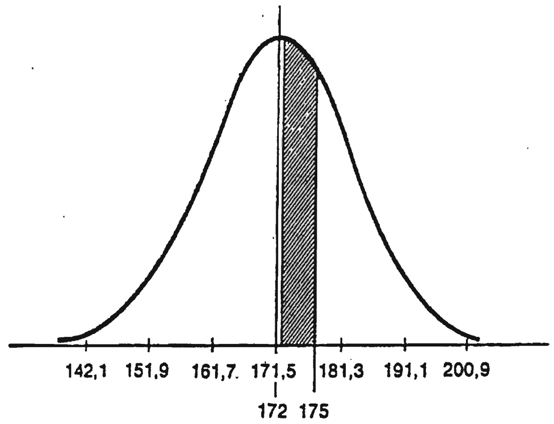
Kako je normalna raspodjela potpuno točno definirana, to mi *za svaki rezultat,* ako izračunamo na koji dio standardne devijacije pada, možemo potpuno točno ustanoviti koliki postotak rezultata je ispod, a koliki iznad njega. Dakle, izražavanje nekog rezultata *u terminima standardne devijacije,* tj. izračunavanje na koji dio standardne devijacije taj rezultat pada, naziva se pretvaranje rezultata u z-vrijednosti. Tablica A u Dodatku je tablica koja nam pokazuje *površinu* normalne distribucije od nekog definiranog rezultata X pa sve do *bližeg* kraja krivulje.



Slika 8.1. Ako neki rezultat pada točno na + 1 s, onda u čitavoj populaciji ima oko 84% slabijih i oko 16% boljih rezultata

Budući da je u svim statističkim tablicama cijela površina neke krivulje prikazana brojem 1 (što znači 100%), to su dijelovi površine prikazani brojevima koji se kreću od 0 pa sve do blizu 0,5 (0,5 je površina *jedne strane* normalne raspodjele). *Veličina površine ujedno znači i vjerojatnost,* kao što se to već vidjelo iz dosadašnjih primjera. Evo još jednog: U skupini od 1 000 mladića nađena je prosječna visina $\overline{X} = 171,5$cm, i $s = 9,8$ cm. Koliko ima približno mladića koji su visoki između 172 i 175 cm?

Kod svih takvih zadataka najbolje je *nacrtati* površinu koja nas zanima. Na slici 8.2. prikazana je krivulja normalne raspodjele, koja ima aritmetičku sredinu 171,5, a standardnu devijaciju 9,8 cm. Površina krivulje između 172 i 175 cm je šrafirana (zatamnjena), i to je upravo površina koja nas zanima.



Slika 8.2. U normalnoj raspodjeli unutar zatamnjene površine nalazi se onaj postotak rezultata koliki postotak je zatamnjena površina od cijele površine krivulje. U našem slučaju taj postotak iznosi oko 12% (vidi tekst)

Da bismo to izračunali, treba izračunati površinu krivulje od 172 cm pa do kraja, potom površinu od 175 cm do kraja, i, konačno, drugu površinu oduzeti od prve:

$$z_1 = \frac{172 - 171,5}{9,8} = 0,05\ , \quad z_2 = \frac{175 - 171,5}{9,8} = 0,36.$$

Prema tome, rezultat 172 nalazi se na + 0,05 z (dakle jedva nešto malo iznad prosjeka), a rezultat 175 na + 0,36 z. Iz tablice A vidimo da se između 0,05 z i kraja krivulje nalazi 48,40% rezultata, a između 0,36 z i kraja krivulje 35,94% rezultata. Zbog jednostavnosti, pretpostavimo da su to postoci 48% i 36%. Prema tome, u intervalu između 172 i 175 (to je zatamnjena površina) nalazi se 48% — 36% = oko 12% svih rezultata, dakle u našem slučaju oko 120 mladića.

U skladu s već rečenim tih 12% znači i ovo: kada bismo nasumce izvadili jedan rezultat iz cijele populacije rezultata, onda je 12% vjerojatno ($p = 0,12$) da ćemo izvući rezultat ne manji od 172 niti veći od 175.

Pomoću dijelova standardne devijacije, dakle pomoću z vrijednosti, mogu se lakše *uspoređivati* rezultati različitih mjerenja kod istog čovjeka, a također i među pojedinim ljudima. Tako se npr. češće događa da pojedinim ljudima iz neke skupine želimo dati jednu *skupnu* ocjenu za njihov rezultat u niz disciplina, a ako su mjerne jedinice tih disciplina različite (npr. sekunde, centimetri, kilogrami itd.), nailazimo na velike teškoće. Jedno od ispravnih rješenja toga problema sastoji se u pretvaranju originalnih vrijednosti rezultata u z-vrijednosti.

Primjer. Skupina od 100 mladića istih godina natječe se u ovim športskim disciplinama: a) trčanje 100 metara, b) skok uvis, c) skok udalj, d) bacanje kugle i e) bacanje diska. Prosječne vrijednosti i standardne devijacije čitave skupine za te discipline su:

a) $\overline{X}_a = 12,8$ sek	$s_a = 2,0$ sek
b) $\overline{X}_b = 145,0$ cm	$s_b = 21,1$ cm
c) $\overline{X}_c = 485$ cm	$s_c = 50$ cm
d) $\overline{X}_d = 813$cm	$s_d = 103$ cm
e) $\overline{X}_e = 2560$ cm	$s_e = 400$ cm

Dva mladića (A i B) postigli su ove rezultate:

Mladić A	Mladić B
a) 12,2 sek	a) 13,0 sek
b) 140 cm	b) 136,5 cm
c) 580 cm	c) 490 cm
d) 804 cm	d) 920 cm
e) 2400 cm	e) 2980 cm.

Koji je od ovih mladića u *svim disciplinama zajedno* bolji? Ako svaki rezultat svedemo na zajedničku mjeru, tj. na z-vrijednost, dobit ćemo za svakog mladića ove "ocjene" $\left(\frac{X-\overline{X}}{s}\right)$:[1]

Mladić A	Mladić B
a) + 0,30	a) −0,10
b) −0,24	b) −0,40
c) + 1,90	c) + 0,10
d) −0,09	d) +1,04
e) −0,40	e) + 1,05
Skupna ocjena = + 1,47	Skupna ocjena = + 1,69
Prosječna ocjena = + 0,29	Prosječna ocjena = + 0,34

Prema tome, prosječan uspjeh mladića B u tim disciplinama veći je od uspjeha mladića A.

[1] Treba paziti da u disciplini trčanja na 100 m vrijednost ispod 12,8 sek svrstavamo u plus, a vrijednost iznad 12,8 sek u minus z-vrijednosti, vrijednosti iznad 12,8 znače *sporije* trčanje.

Takvo ocjenjivanje u z-vrijednostima vrlo je potrebno i u onim slučajevima kad tražimo skupnu ili prosječnu ocjenu iz niza mjerenja koja su na prvi pogled provedena u jednakim "mjernim jedinicama" (npr. u bodovima), ali pomoću testova, koji imaju različit varijabilitet rezultata.

Primjer: Skupinu ljudi podvrgavamo ispitivanju psihomotorike pomoću 2 testa (T_1 i T_2) koji imaju ove karakteristike:

T_1	T_2
Minimalni broj bodova = 0	Minimalni broj bodova = 0
Maksimalni broj bodova = 15	Maksimalni broj bodova = 120
Prosjek (izmjeren na većoj skupini) = $\overline{X}_1 = 7,0$	Prosjek (izmjeren na većoj skupini) = $\overline{X}_2 = 60,0$
$s_1 = 1,0$	$s_2 = 14,0$

Ispitanik A postigao je u testu T_1 9 bodova, a u testu T_2 74 boda; ispitanik B postigao je u testu T_1 6 bodova, a u testu T_2 90 bodova. Dakle, *ukupno* je ispitanik A postigao 83 boda, a ispitanik B 96 bodova. Je li zaista ispitanik B u ukupnoj ocjeni bolji od ispitanika A?

Ako bodove pretvorimo u z-vrijednosti, dobivamo za oba ispitanika:

Ispitanik A		Ispitanik B	
Test T_1	+2,0	Test T_1	−1,0
Test T_2	+1,0	Test T_2	+2,14
Ukupno	+3,0	Ukupno	+1,14

Prema tome, znatno bolji u ukupnom uspjehu je ispitanik A, a ne ispitanik B!

Dakle, potrebno je pretvaranje u z-vrijednosti, jer bi jednostavnim zbrajanjem bruto-rezultata dvaju ili više mjerenja u ukupnom zbroju imali veću "težinu", tj. veću važnost rezultati iz onih mjerenja u kojima je veća standardna devijacija. Na primjer, rezultati iz mjerenja sa standardnom devijacijom 10 imali bi dva puta veću težinu u ukupnom rezultatu od rezultata mjerenja sa standardnom devijacijom 5. Često se pogrešno misli da glavnu "težinu" rezultatima daje njihova apsolutna veličina: na primjer, ako u jednom testu ispitanici postižu vrlo visok broj bodova, a u drugim testovima znatno manji, onda će, po tom mišljenju, u ukupnom rezultatu iz svih testova "prevagnuti" rezultat iz toga testa s najvećim brojem postignutih bodova. Da to nije točno, vidi se iz ovog primjera: pretpostavimo da u tom jednom testu *svi* ispitanici postignu mnogo, ali *svi jednako*, recimo 500 bodova svaki ispitanik. U ukupnom rezultatu taj broj, iako velik, neće ***ništa*** utjecati na položaj pojedinog ispitanika, tj. njihov rang bit će potpuno jednak bez toga testa kao i s njim — jer smo svakom ispitaniku dodali *jednaku* vrijednost. Dakle, kako se vidi, taj test ne pridonosi ništa ukupnom rezultatu, jer taj test nema *varijabiliteta.*

Da bi se izbjegle *negativne* vrijednosti kod takvog preračunavanja originalnih rezultata u z-vrijednosti, može se svakoj z- vrijednosti dodati neki broj, tako da nam svi rezultati postanu pozitivni brojevi. Ako z-vrijednostima dodamo broj 5, dobivamo tzv. *standardne* vrijednosti, koje idu od 2 do 8, s prosjekom 5.

Računski se taj broj (u ovom slučaju 5) samo pribroji već poznatom računu kod z-vrijednosti.

Primjer: $\overline{X} = 25, s = 6$. Koju standardnu vrijednost ima rezultat $X = 32$ u skali standardnih vrijednosti kojima je aritmetička sredina 5?

$$\text{Standardna vrijednost} = \frac{X - \overline{X}}{s} + 5 = \frac{32 - 25}{6} + 5 = 6,17.$$

8.2. CENTILI

Položaj pojedinca u grupi može se izraziti i tzv. *centilom* (percentilom): prvi centil obuhvaća 1% najslabijih, drugi centil 1% idućih najslabijih,.. deseti centil obuhvaća onaj jedan posto koji je na desetom mjestu od najslabijeg,.. sedamdeseti centil znači da je 70% rezultata jednako ili slabije, a 30% bolje; ako neki rezultat pada u devedeset drugi centil, to znači da je samo 8% rezultata bolje od njega, a 92% rezultata su jednaki ili slabiji, itd. Mi smo zapravo i kod z-vrijednosti već baratali centilima: ako smo na str. 97. rekli da je 16% ispitanika rastom viša od ispitanika X, to zapravo znači da on prema svojoj visini pripada u 84. centil: $100 - 16 = 84$. Prema tome, iz tablice A možemo za svaki rezultat očitati i njegov položaj u centilima, ali prethodno treba rezultat pretvoriti u z-vrijednost. Na primjer, ako neki rezultat pada na 1,75 z, znači da se taj rezultat nalazi u 96. centilu, jer ima samo oko 4% rezultata koji su bolji od njega.

Postoji računski postupak za dobivanje centila pojedinog rezultata, a da pritom nije potrebno rezultat najprije pretvoriti u z-vrijednost, ali ćemo najprije ukratko prikazati *grafičku* metodu kojom možemo s priličnom točnošću odrediti položaj pojedinog rezultata u centilima.

Postupak se sastoji u neznatnoj modifikaciji grafičkog prikaza kumulativne frekvencije (vidi sliku 6.10), tj. sastoji se u sljedećem:

1. Frekvencije neke distribucije rezultata pretvorimo najprije u *relativne* frekvencije, a nakon toga u relativne *kumulativne* frekvencije (tj. zbrajamo relativne frekvencije [postotke] iz razreda u razred). Ti nam brojevi u postotku pokazuju koliko je ukupno rezultata do prave gornje granice svakog razreda.
2. Na apscisu unesemo prave gornje granice razreda (koje smo već upoznali u 6. poglavlju), a na ordinatu unesemo relativne kumulativne frekvencije, tj. kumulativne frekvencije u postotku ukupnog N.

Primjer: Nama već poznat primjer s visinom 135 20-godišnjih mladića, koji smo na slici 6.10. prikazali krivuljom kumulativne frekvencije, pretvorit ćemo u primjer s *relativnim* kumulativnim frekvencijama. Postupak je prikazan u tablici 8.1.

TABLICA 8.1.
PRETVARANJE FREKVENCIJE U RELATIVNU KUMULATIVNU FREKVENCIJU

Razred	Frekven-cija	Kumul. frekv.	Relativna frekven-cija	Relativna kumul. frekv.
157-159	1	1	0,74	0,74
160-162	2	3	1,48	2,22
163-165	9	12	6,67	8,89
166-168	15	27	11,11	20,00
169-171	25	52	18,52	38,52
172-174	28	80	20,74	59,26
175-177	20	100	14,82	74,08
178-180	16	116	11,85	85,93
181-183	13	129	9,63	95,56
184-186	5	134	3,70	99,26
187-189	1	135	0,74	100,00
	135		100,00	

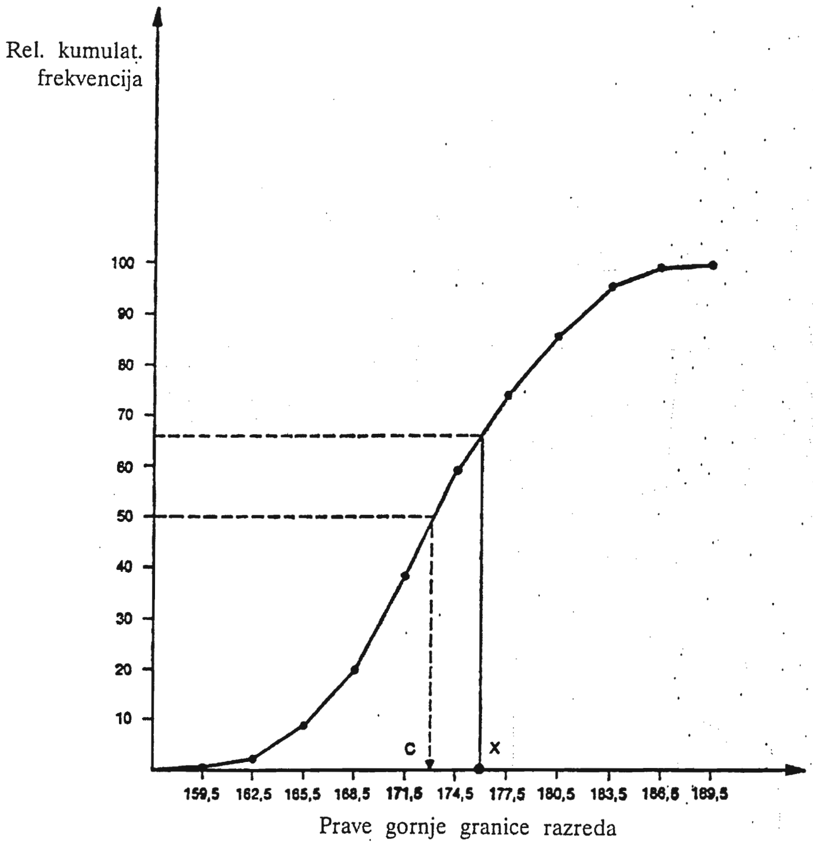
Kada tako dobivene vrijednosti unesemo u grafikon, dobivamo krivulju relativne kumulativne frekvencije, prikazanu na slici 8.3.

Za aproksimativno očitavanje centila ili centralne vrijednosti ova nam krivulja može odlično poslužiti. Tako, recimo, iz slike 8.3. vidimo da neki rezultat X, veličine 176 cm, pada otprilike u 66. centil, tj. da su 66% rezultata jednaki ili niži od njega, a 34% rezultata su jednaki ili viši. Centralna se vrijednost iz ove slike očitava tako da se s relativne kumulativne frekvencije 50% povuče paralela s apscisom sve do krivulje i odanle se spusti okomica na apscisu. Na mjestu gdje ta okomica siječe apscisu nalazi se centralna vrijednost.

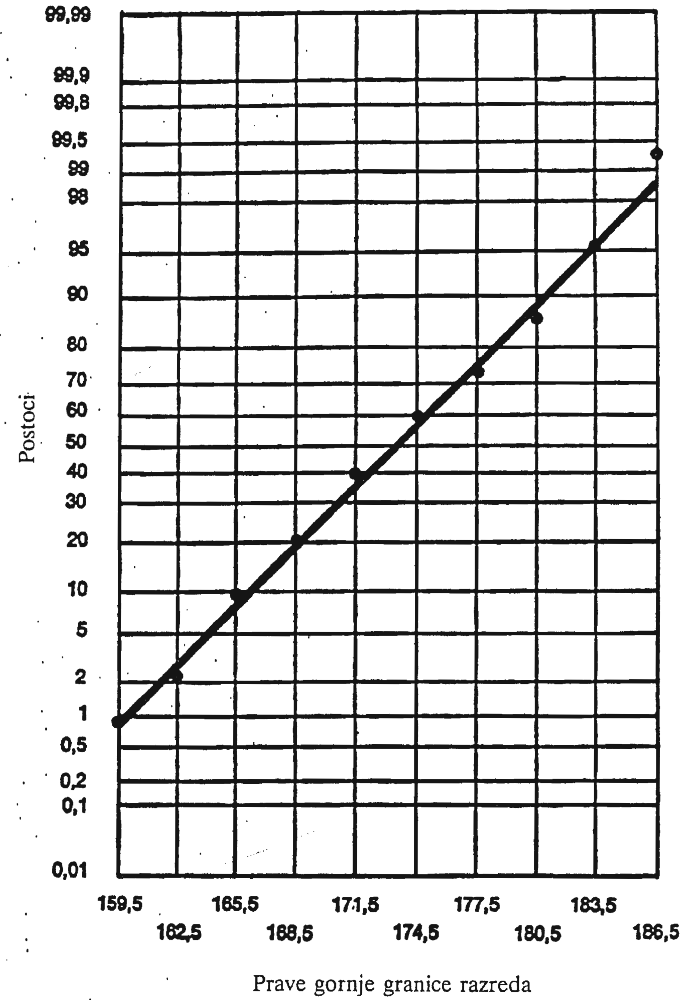
(N a p o m e n a. Poneki čitaoci možda su čuli i za izraz "decil". Za razliku od "centila", "decili" su frakcije od po 10% rezultata: 1. decil obuhvaća najslabijih (ili najboljih) 10%, 2. decil idućih 10%, itd., i, konačno, 10. decil predstavlja 10% najboljih (ili najslabijih) rezultata: 1. centil je početna vrijednost za 1. decil, 10. centil je početna vrijednost za 2. decil, itd. Naravno da se i decili, jednako kao i centili, mogu posve jednostavno očitati s krivulje relativne kumulativne frekvencije. Treba reći da su u *praktične svrhe,* kao što je to npr. baždarenje nekog testa, decili koriste češće nego centili.)

Jedan po izgledu nešto drukčiji, ali u biti jednak postupak sastoji se u tome da se relativne kumulativne frekvencije nanesu na ordinatu tzv. "papira vjerojatnosti". Na slici 8.4. prikazani su isti rezultati na papiru vjerojatnosti. (Relativna kumulativna frekvencija od 100% ne može se ucrtati jer na papiru ne postoje vrijednosti 0 niti 100.)

Sve ono što smo očitavali iz slike 8.3., tj. s obične krivulje relativne kumulativne frekvencije, možemo očitati i sa slike 8.4, samo što nam ona pruža još jednu prednost: ako je distribucija, koju na ovaj način prikazujemo, približno *normalna,* točke na papiru vjerojatnosti bit će manje-više na *pravcu.*

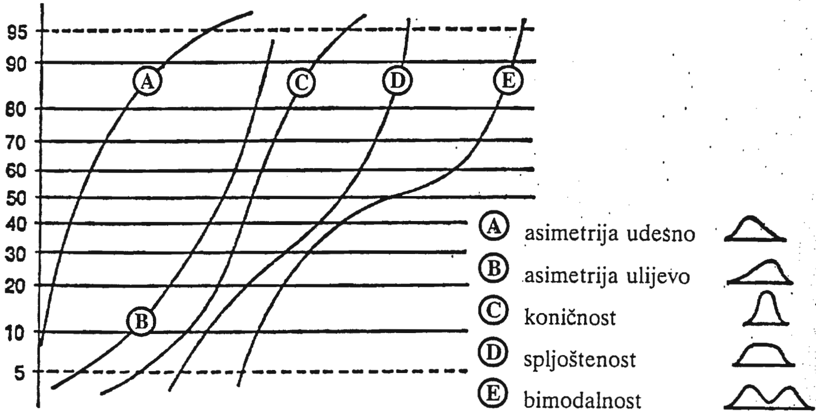


Slika 8.3. Krivulja relativne kumulativne frekvencije



Slika 8.4. Relativne kumulativne frekvencije prikazane na "papiru vjerojatnosti". Normalna raspodjela daje u tom slučaju pravac

(Neki autori crtom spajaju sve točke, a neki povlače među točkama pravac ili krivulju, koji točkama relativno najbolje pristaju. Taj drugi postupak učinjen je i na slici 8.4.) Ako je distribucija asimetrična na jednu ili drugu stranu, ili ako na bilo koji način značajnije odstupa od normaliteta, kroz njih se neće moći povući pravac. Prema tome, prikazivanje relativne kumulativne frekvencije na papiru vjerojatnosti omogućuje nam još i *približnu kontrolu normaliteta distribucije*; (o preciznoj *računskoj* kontroli bit će riječi u poglavlju "Hi-kvadrat test"). Na slici 8.4. vidimo da naši rezultati uglavnom daju normalnu distribuciju, jer "odoka" nacrtan pravac dobro prati položaje točaka.



Slika 8.5. Neke distribucije koje odstupaju od normalne distribucije, prikazane na papiru vjerojatnosti

Na slici 8.5. prikazano je kako na papiru vjerojatnosti izgledaju neke distribucije koje značajnije odstupaju od normalnosti. No, naravno, takva odstupanja ne narušavaju mogućnost očitavanja decila, centila, centralne vrijednosti, pa čak donekle i aritmetičke sredine i standardne devijacije: budući da znamo da aritmetička sredina plus i minus jedna standardna devijacija obuhvaćaju oko 68% rezultata, možemo od vrijednosti 50 na ordinati (što je zapravo položaj centralne vrijednosti, ali ako je raspodjela pravilna, onda se i aritmetička sredina nalazi uglavnom na istom mjestu) dodati po 34% prema gore i dolje, i to očitati. Ako ta očitanja pokušamo provesti, očitat ćemo aproksimativno aritmetičku sredinu na oko 173,5 (ona zapravo iznosi isto toliko!), a vrijednost jedne standardne devijacije iznad aritmetičke sredine na oko 179 (zapravo 179,4) i jedne standardne devijacije ispod aritmetičke sredine oko 167,5 (zapravo 167,6). Kako se vidi, vrlo dobra aproksimacija!

Računski možemo centil nekog rezultata izračunati pomoću formule:

$$\text{Centil nekog rezultata} = \frac{\text{Rang rezultata}}{N} \cdot 100. \qquad (8.2)$$

Na primjer, ako je među 50 rezultata neki rezultat 43. po redu (odozdo!), onda se taj rezultat nalazi u $43/50 \cdot 100 = 86$. centilu.

No kod velikih skupina, kada su nam rezultati svrstani u razrede, bio bi taj postupak previše kompliciran, i zato postoji mnogo praktičnija formula za pronalaženje centila nekog rezultata. Ona glasi:

$$\text{Centil} = RKF_{\mathrm{D}} + \left(\frac{X - D}{i} \cdot RF_{\mathrm{R}}\right), \qquad (8.3)$$

pri čemu:

- X = rezultat za koji tražimo centil,
- RKF_{D} = relativna kumulativna frekvencija (suma postotaka) rezultata *ispod* razreda u kojem je rezultat X,
- D = prava *donja* granica razreda u kojem se nalazi rezultat X,
- i = interval,
- RF_{R} = relativna frekvencija (postotak) rezultata u razredu u kojem se nalazi rezultat X.

Na primjer, kada bi nas u našem primjeru zanimao centil rezultata 176, postupili bismo ovako:

$$\begin{aligned} \text{Centil} &= 59,26 + \frac{176 - 174,5}{3} \cdot 14,82, \\ &= 59,26 + (0,5 \cdot 14,82), \\ &= 59,26 + 7,41 = 66,67, \text{ili praktički } 67. \end{aligned}$$

Naš se rezultat nalazi dakle u 67. centilu. Da smo ga očitavali sa slike 8.3, očitali bismo otprilike 66. centil.

Često nas međutim može zanimati *obratni* postupak: dok smo upravo sada našli koji centil odgovara nekom određenom rezultatu, nas može zanimati i to koji *rezultat* odgovara nekom određenom centilu. Na primjer, u našem slučaju mogli bismo postaviti pitanje: Koja visina odgovara 50. centilu (dakle koja visina predstavlja centralnu vrijednost ili medijan)?

U tu svrhu treba najprije željeni centil pomnožiti s N i podijeliti sa 100 te ćemo tako dobiti *rang* rezultata koji je u 50. centilu. U našem slučaju, dakle, imamo

$$\frac{\text{centil} \cdot N}{100} = \frac{50 \cdot 135}{100} = 67,5.$$

Nakon toga treba u stupcu kumulativne frekvencije naći razred u kojem se taj rang rezultata nalazi. U našem slučaju rezultat koji zauzima 67,5. rang pada u razred 172-174 (jer se u tom razredu, kako vidimo iz tablice 8.1, nalaze rezultati iznad 52. ranga pa sve do 80. ranga).

Kada smo našli rang rezultata i razred u kojem se on nalazi, primijenit ćemo ovu formulu:

$$X_{\mathrm{C}} = D + \frac{\frac{C \cdot N}{100} - f_{\mathrm{D}}}{f_{\mathrm{R}}} \cdot i, \qquad (8.4)$$

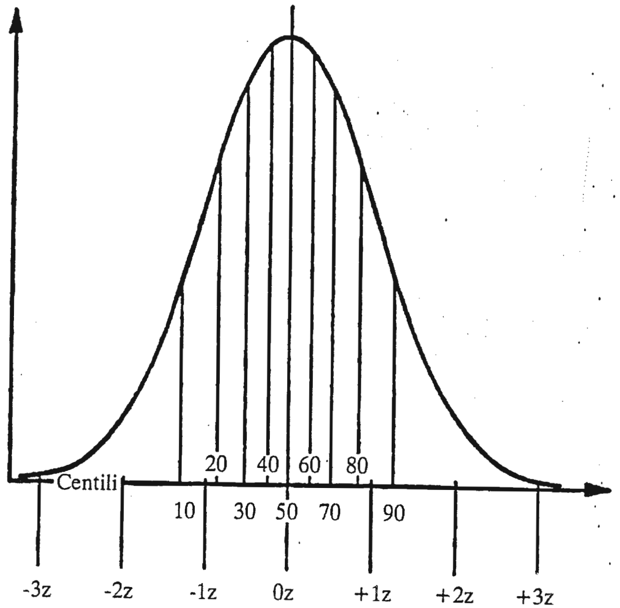
pri čemu:

X_c = rezultat koji pada u željeni centil,
D = prava donja granica razreda u kojem je rezultat X,
C = zadani centil (u našem slučaju to je 50),
f_D = ukupni broj rezultata *ispod* razreda u kojem je rezultat X,
f_R = broj rezultata u razredu u kojem se nalazi rezultat X,
i = interval.

Ako sada primijenimo formulu (8.4), dobivamo:

$$X_{50} = 171,5 + \left(\frac{67,5 - 52}{28} \cdot 3\right),$$
$$= 171,5 - (0,55 \cdot 3)$$
$$= 171,5 + 1,65 = 173,15.$$

Prema tome, rezultat koji odgovara 50. centilu, iznosi oko 173. Da smo centralnu vrijednost očitavali s pravca na slici 8.4, dobili bismo otprilike jednaku vrijednost.



Slika 8.6. Odnos između z-vrijednosti i centila: za razliku od centila z-vrijednosti predstavljaju ekvidistantnu skalu na apscisi

I na kraju još nekoliko riječi o usporedbi između z-vrijednosti i centila.

Prednost z-vrijednosti pred centilima sastoji se u tome što, kako smo vidjeli, z-vrijednosti možemo zbrajati i tražiti njihov prosjek (jer su to "ekvidistantne" jedinice, tj. jedinice s jednakim međusobnim razmakom), dok kod centila takvo zbrajanje i traženje "prosječnog" centila nije dopušteno, jer centili u normalnoj raspodjeli ne predstavljaju ekvidistantne jedinice. To se može lijepo vidjeti iz slike 8.6. na kojoj su označene z- vrijednosti i centili (a ista se stvar može vidjeti i na tablici 8.2).

S druge strane, prednost centila pred z-vrijednostima sastoji se u tome da centili *ne zahtijevaju normalnu distribuciju:* mi možemo iz *bilo kakve* distribucije odrediti koji rezultat postiže 10% najboljih, ili 35% najlošijih ispitanika, itd. Naprotiv, z-vrijednosti predstavljaju — kako znamo — dijelove standardne devijacije, a ona je vezana uz normalnu raspodjelu.

TABLICA 8.2.
CENTILI KOJI ODGOVARAJU POJEDINIM z-VRIJEDNOSTIMA

z-vrijednost	Centil	z-vrijednost	Centil
5,0		0,0	C 51
4,0		−0,1	C 47
3,0	C 100	−0,2	C 45
2,5		−0,3	C 39
2,33		−0,4	C 35
2,3	C 99	−0,5	C 31
2,2	C 99	−0,6	C 28
2,1	C 99	−0,7	C 25
2,0	C 98	−0,8	C 22
1,9	C 98	−0,9	C 19
1,8	C 97	−1,0	C 16
1,7	C 96	−1,1	C 14
1,6	C 95	−1,2	C 12
1,5	C 94	−1,3	C 10
1,4	C 92	−1,4	C 9
1,3	C 91	−1,5	C 7
1,2	C 89	−1,6	C 6
1,1	C 87	−1,7	C 5
1,0	C 85	−1,8	C 4
0,9	C 82	−1,9	C 3
0,8	C 79	−2,0	C 3
0,7	C 76	−2,1	C 2
0,6	C 73	−2,2	C 2
0,5	C 70	−2,3	C 2
0,4	C 66		
0,3	C 62	−2,33	
0,2	C 58	−2,5	
0,1	C 54	−3,0	C 1
		−4,0	
		−5,0	

ZADACI ZA VJEŽBU

1. Skupina od 500 ispitanika u nekom je testu dobila normalno distribuirane rezultate s ovim karakteristikama: aritmetička sredina iznosi 80 bodova, a standardna devijacija 10 bodova. Izračunajte uz pomoć tablice A u Dodatku koliko otprilike ima ispitanika (ne postotak!) koji imaju rezultat:

   a. veći od 80
   b. veći od 85
   c. veći od 101
   d. manji od 71
   e. manji od 83
   f. manji od 56
   g. veći od 78
   h. između 60 i 65
   i. između 73 i 84
   j. između 52 i 108.

2. Koja je vjerojatnost da će student A, koji se nalazi u skupini od 500 studenata iz zadatka 1, postići bolji rezultat od 95 bodova?
3. Jedan proizvođač tvrdi da je testirao trajnost svojih žarulja, i da trajnost — uz normalnu upotrebu — u prosjeku iznosi 300 dana, sa standardnom devijacijom 90 dana. Vaša žarulja, koju ste kupili od tog proizvođača, izgori nakon 60 dana. Koja je šansa postojala da će se to dogoditi?
4. U jednom gradu u kojem ima ukupno 500 muških učenika završnih razreda, njihova visina je $\overline{X} = 175$ cm, a $s = 11$ cm. Prilikom jedne športske manifestacije nastavnik tjelesnog odgoja namjerava izvesti neku vježbu s tim učenicima, ali za vježbu dolaze u obzir samo oni koji su viši od 190 cm. Za vježbu ih je potrebno 80. Hoće li ih nastavnik uspjeti toliko pronaći?
5. U Bennettovu testu mehaničkog razumijevanja 75 učenika postiglo je rezultate prikazane u tablici:

Razred	f
5 — 9	2
10 — 14	—
15 — 19	3
20 — 24	10
25 — 29	10
30 — 34	14
35 — 39	17
40 — 44	12
45 — 49	2
50 — 54	5

   U koji centil spada rezultat 46?

6. Uz pomoć tablice A u Dodatku odgovorite na pitanje: u koji centil spadaju rezultati koji imaju ove z-vrijednosti:

   a. 0
   b. 1
   c. — 0,5
   d. — 1,2
   e. — 2
   f. — 2,5

9. RAZLIKA IZMEĐU DVIJE ARITMETIČKE SREDINE

9.1. POPULACIJA I UZORAK

Vrlo često u statistici čujemo pojmove "populacija" i "uzorak". Pod *populacijom* (ili "univerzumom" ili "osnovnim skupom") razumijevamo sve članove neke skupine s određenom karakteristikom koju mjerimo. Na primjer, ako mjerimo visinu 16-godišnjih mladića u Zagrebu, onda populaciju predstavljaju svi 16-godišnji mladići u Zagrebu. Ako je "grupa" neograničena, onda populaciju predstavlja *neizmjerni* (ili "vrlo velik") broj mjerenja (npr. ako mjerimo utjecaj fenamina na bazalni metabolizam, populaciju bi predstavljao "neizmjerni", tj. vrlo velik broj mjerenja).

Poneki je put populaciju *nemoguće* izmjeriti (npr. ako je ona beskonačno velika); poneki put to je *vrlo skupo i komplicirano* (npr. kad bismo htjeli doznati prosječnu zaradu svih odraslih stanovnika jedne zemlje, morali bismo stvarno pronaći *svakoga* odraslog stanovnika, a to ne bi bilo ni lako ni jeftino); katkada bi mjerenje cijele populacije bilo *besmisleno* (na primjer, ako se prilikom mjerenja *uništava* ono što se mjeri — kao što je to, recimo, slučaj kod testiranja konzervi s prehrambenim artiklima).

Zato većinom mjerimo samo *ograničeni broj* slučajeva, koji nazivamo *uzorkom*, a aritmetičku sredinu, standardnu devijaciju, itd., tih uzoraka nazivamo "procjenama" tih istih parametara. Na primjer, ako iz neke populacije uzmemo uzorak od 30 mjerenja, aritmetička sredina tih mjerenja je *procjena* prave aritmetičke sredine, standardna devijacija je *procjena* prave standardne devijacije, itd. Pravu aritmetičku sredinu (koja se bilježi sa μ; *čitaj* mi) ili pravu standardnu devijaciju (koja se bilježi sa σ; *čitaj* sigma) dobili bismo tek onda kad bismo izmjerili čitavu populaciju. Prema tome, kada radimo s uzorcima, izračunate vrijednosti, nisu uvijek zapravo prave vrijednosti, nego samo približne vrijednosti (*"procjene"*).

Kako je za bolje *razumijevanje* sadržaja, koji slijedi, potrebno biti dobro upoznat s pojmom uzorka, te s vezom između uzorka i populacije, pokazat ćemo na jednostavnom primjeru što se događa ako iz jedne populacije uzastopno uzimamo uzorke jednake veličine i izračunavamo njihove karakteristike.

Pretpostavimo da imamo jednu veliku kutiju s nama nepoznatim brojem crvenih i zelenih kuglica. Na dnu kutije postoji uređaj s 20 rupica kroz koji može ispasti 20 kuglica. Potresanjem kutije dobro promiješamo kuglice, uzmemo uzorak od 20 kuglica, registriramo njihov sadržaj (broj crvenih kuglica), vratimo kuglice natrag u kutiju, ponovno dobro izmiješamo, izvadimo novi uzorak od 20 kuglica, i tako dalje, ukupno pedeset puta. Rezultati ovog pokusa prikazani su u tablici 9.1.

TABLICA 9.1.
REZULTATI DOBIVENI OD 50 UZORAKA IZ POPULACIJE I.
VELIČINA SVAKOG UZORKA $N = 20$

Red. broj uzorka	Broj crv. kuglica	Red. broj uzorka	Broj crv. kuglica	Red. broj uzorka	Broj crv. kuglica
1	12	18	12	35	12
2	12	19	8	36	11
3	9	20	12	37	10
4	12	21	9	38	10
5	9	22	11	39	9
6	9	23	9	40	10
7	11	24	10	41	11
8	8	25	14	42	9
9	11	26	11	43	11
10	8	27	12	44	12
11	11	28	14	45	12
12	10	29	10	46	12
13	7	30	14	47	13
14	8	31	10	48	10
15	13	32	18	49	10
16	13	33	8	50	13
17	13	34	15	Ukupno 548	

Kako se vidi, ti uzorci *variraju*, i distribucija njihove frekvencije prikazana je u tablici 9.2.

Kako se vidi, uzorci variraju *po slučaju*, ali veličina i oblik te varijacije *zavisi od populacije i njezinog sastava*. Da bi to dokazali, učinimo isti pokus, samo prije toga dademo nekom da promijeni odnose crvenih i zelenih kuglica u kutiji. Ponovno uzimamo 50 uzoraka veličine $N = 20$, i dobivamo, recimo, rezultate prikazane u tablici 9.3.

Prikažemo li distribuciju frekvencije tih uzoraka, dobivamo rezultate prikazane na tablici 9.4.

Prikažemo li dobivene rezultate iz oba pokusa grafički (vidi sliku 9.1.), upadaju u oči dvije stvari:

TABLICA 9.2.
DISTRIBUCIJA FREKVENCIJE UZORAKA POPULACIJE I.

Broj crvenih kuglica u uzorku	f
manje od 7	0
7	1
8	5
9	7
10	9
11	8
12	10
13	5
14	3
15	1
16	0
17	0
18	1
više od 18	0
	Ukupno 50

TABLICA 9.3.
REZULTATI DOBIVENI OD 50 UZORAKA IZ POPULACIJE II.
VELIČINA SVAKOG UZORKA $N = 20$

Red. broj uzorka	Broj crv. kuglica	Red. broj uzorka	Broj crv. kuglica	Red. broj uzorka	Broj crv. kuglica
1	3	18	4	35	2
2	4	19	4	36	5
3	1	20	4	37	4
4	2	21	5	38	3
5	1	22	3	39	4
6	2	23	5	40	4
7	3	24	2	41	3
8	1	25	2	42	2
9	5	26	2	43	3
10	1	27	2	44	4
11	3	28	2	45	2
12	3	29	2	46	3
13	4	30	4	47	7
14	3	31	2	48	3
15	3	32	2	49	2
16	3	33	4	50	4
17	3	34	3	Ukupno 152	

TABLICA 9.4.
DISTRIBUCIJA FREKVENCIJE UZORAKA POPULACIJE II.

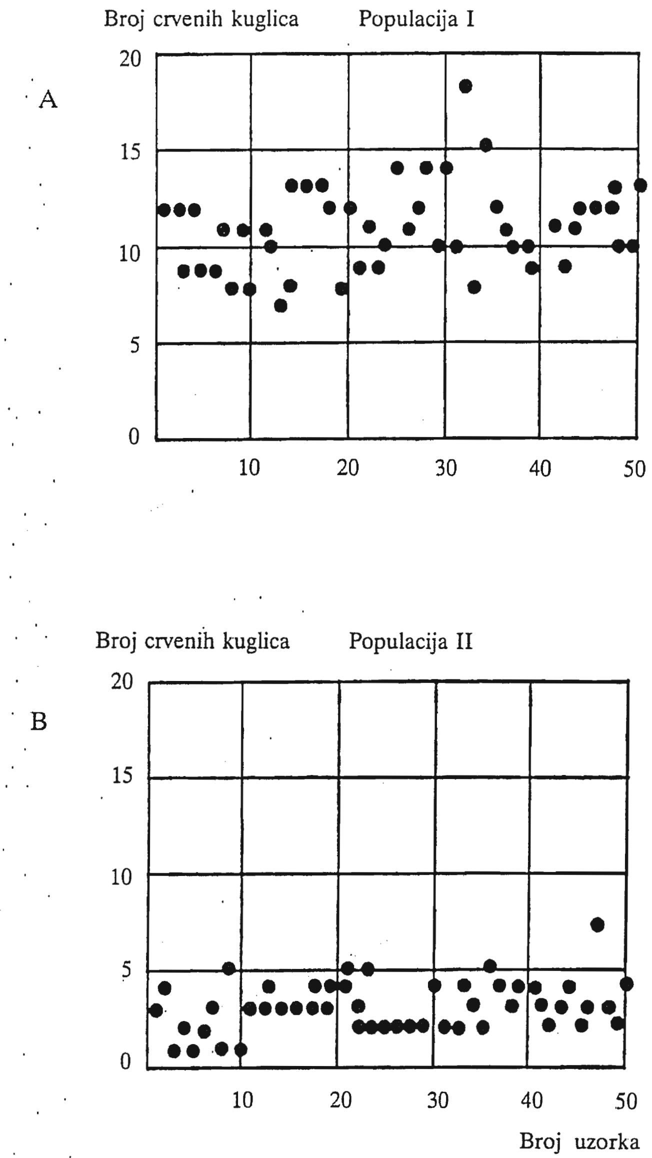
Broj crvenih kuglica u uzorku	f
0	0
1	4
2	14
3	15
4	12
5	4
6	0
7	1
više od 7	0
	Ukupno 50

1. U pokusu s populacijom I. uzorci se grupiraju kao veće vrijednosti nego uzorci u pokusu s populacijom II.
2. Grupiranje uzoraka je "gušće" u pokusu s populacijom II, tj. raspršenje rezultata je manje nego u pokusu s populacijom I.

Iz dobivenih se rezultata dade zaključiti da uzorak nije "minijaturni duplikat" populacije, već — ako na populaciju zaključujemo iz uzorka — moramo uzeti u obzir slučajne varijacije koje se mogu dogoditi.

Vidjeli smo iz opisana dva pokusa da rezultati uzoraka variraju po slučaju, ali i to da *veličina te slučajne varijacije ovisi o stanju u populaciji:* što je veći varijabilitet rezultata u populaciji, to će biti i veći varijabilitet uzoraka uzetih iz te populacije. U našem primjeru možemo reći da je varijabilitet populacije to manji što više prevladava jedna vrsta kuglice: kad bi sve kuglice bile crvene, tj. kad u populaciji ne bi bilo nikakvog varijabiliteta — ne bi bilo varijabiliteta ni među uzorcima (tj. svi uzorci sastojali bi se samo od crvenih kuglica). Naprotiv, ako se populacija sastoji od 50% crvenih i 50% zelenih kuglica, varijabilitet je maksimalan, pa će i varijabilitet uzoraka biti najveći. U našoj populaciji I. iz grupiranja uzoraka (pretežno na vrijednostima 9-12) zaključujemo da se radi o populaciji u kojoj je vjerojatno odnos crvenih i zelenih kuglica približno 50% : 50%, dok je znatno manji postotak crvenih kuglica u populaciji II; to ujedno potvrđuje i varijabilitet uzoraka: on je znatno veći u pokusu s populacijom I. nego u pokusu s populacijom II.

Potpuno je jasno da bismo dobili analogne rezultate kad bismo uzimali uzorke iz jedne populacije rezultata koji daju *normalnu* raspodjelu. Kad bi, na primjer, 1024 rezultata mjerenja visine djece na 1024 ispitanika napisali na žetone koje bi stavili u kutiju i tada vadili uzorke određene veličine, aritmetičke sredine uzoraka kretale bi se u blizini prave aritmetičke sredine svih rezultata, a varijabilitet tih aritmetičkih sredina bio bi to veći što je veći i stvarni varijabilitet originalnih 1024 rezultata.



Slika 9.1. Uzorci se grupiraju oko aritmetičke sredine populacije i variraju to manje što je varijabilitet populacije manji

No pođimo sada korak dalje i pogledajmo što se zbiva ako pri takvim eksperimentima uzimanja uzoraka iz populacije nakon svakog završenog eksperimenta *promijenimo veličinu uzorka* i ponovno izvršimo pokus. Takav eksperiment svojedobno sam izveo radi demonstracije. Eksperiment je izgledao ovako:

Iz jedne *poznate* populacije veličine $N = 1024$, koja je bila normalno distribuirana, i prenesena na žetone, po slučaju su iz kutije vađeni uzorci različite veličine, tj. veličine $N = 1$, $N = 2$, $N = 5$ i $N = 25$. Za svaku od tih grupa uzoraka izvučeno je 20 uzoraka. Rezultati tog pokusa prikazani su u tablici 9.5.

Kako se iz tablice vidi, standardna devijacija *aritmetičkih sredina* uzoraka bila je to manja što je uzorak (dakle broj mjerenja) bio veći. Također je i zajednička aritmetička sredina svih uzoraka to bliže *pravoj* aritmetičkoj sredini (tj. aritmetičkoj sredini populacije koja iznosi 135,0) što su uzorci bili veći.

Ako zamislimo jednu populaciju, kojoj je aritmetička sredina $\mu = 50$, a standardna devijacija $\sigma = 10$, onda se aritmetičke sredine *uzoraka* različite veličine raspoređuju oko prave aritmetičke sredine tako kako je to prikazano na slici 9.2.

Kako se vidi, što su uzorci veći to je raspršenje aritmetičkih sredina uzoraka oko prave aritmetičke sredine manje.

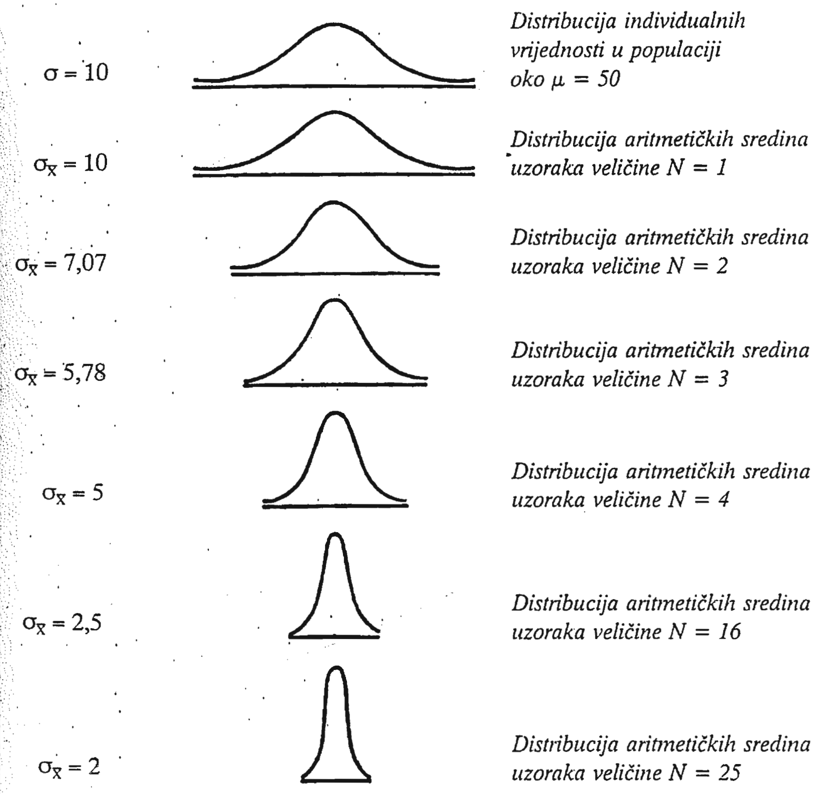
TABLICA 9.5.
DISTRIBUCIJA ARITMETIČKIH SREDINA 20 UZORAKA IZ NORMALNE POPULACIJE KOJOJ ARITMETIČKA SREDINA IZNOSI 135,0, A STANDARDNA DEVIJACIJA 4,74, AKO SU VELIČINE UZORAKA: $N = 1$, $N = 2$, $N = 5$ i $N = 25$

Razredi	Uzorci			
	$N = 1$	$N = 2$	$N = 5$	$N = 25$
124 – 125	1			
126 – 127	1			
128 – 129	6	1		
130 – 131	2	2		
132 – 133	1	2	3	2
134 – 135	3	4	9	9
136 – 137	2	9	6	9
138 – 139	1	1	2	
140 – 141	2	1		
142 – 143				
144 – 145	1			
Ukupno	20	20	20	20
Aritmetička sredina od 20 aritmetičkih sredina*	132,85	135,5	135,21	135,06
Standardna devijacija 20 aritmetičkih sredina*	5,55	3,00	1,27	1,20

* izračunato iz negrupiranih rezultata

Dakle, iz dosadašnjih primjera zaključak je ovaj:

Ako iz jedne populacije uzimamo *mnogo* uzoraka iste veličine i od svakog uzorka izračunamo aritmetičku sredinu ($\bar{X}$), pojedine aritmetičke sredine većinom će se više ili manje razlikovati od *prave* aritmetičke sredine populacije, ali će se te aritmetičke sredine uzoraka ipak *grupirati* oko *prave aritmetičke sredine* i to tako da će ih najviše biti u neposrednoj blizini aritmetičke sredine, a sve manje bit će ih koje znatno odstupaju od prave aritmetičke sredine. Ako je broj uzoraka vrlo velik, ustanovit ćemo *da se aritmetičke sredine uzoraka iste veličine grupiraju oko prave aritmetičke sredine (tj. oko aritmetičke sredine populacije) po jednakom zakonu, kao što se individualni rezultati grupiraju oko svoje aritmetičke sredine, tj. po zakonu NORMALNE RASPODJELE!* Standardna devijacija te raspodjele je to manja što su uzorci veći.



Slika 9.2. Distribucija aritmetičkih sredina uzoraka različite veličine oko prave aritmetičke sredine

Učinimo jedan mali pokus kojim ćemo *dokazati* da se aritmetičke sredine uzoraka grupiraju oko prave aritmetičke sredine po normalnoj raspodjeli.

Zamislimo da imamo jednu *populaciju*, koja je — zbog jednostavnosti — vrlo mala, i koja se sastoji od samo 5 rezultata, tj. od rezultata 4, 5, 6, 7 i 8. Njezina (dakle *prava*) aritmetička sredina iznosi 30/5 = 6. Zamislimo nadalje da smo odlučili iz te populacije izvući *sve moguće* uzorke veličine $N = 2$. Prema teoretskim (matematičkim) pravilima, prema kojima su stvoreni svi zakoni u vezi s uzorcima, uzimanje uzoraka treba obavljati po tzv. sistemu "uz povrat", a to znači da treba prvo izvući jedan rezultat, zabilježiti ga, *vratiti taj rezultat natrag u populaciju* i nakon toga izvući drugi rezultat i zabilježiti ga. Ta dva rezultata sačinjavaju *prvi* uzorak veličine $N = 2$. Na isti način izvukli bismo i sve druge uzorke. Naravno, da smo odlučili da uzorak bude veći (recimo veličine $N = 4$) i tada bismo morali raditi sistemom "uz povrat".

Već mogu zamisliti kako su se mnogi čitaoci pobunili i (opravdano) postavili ova dva prigovora:

1. Na taj način može se dogoditi da npr. u uzorku veličine $N = 2$ dobijemo *dva puta isti rezultat;*

2. U životu se uzorci ne uzimaju na ovaj način, nego donekle čak obratno, tj. ako nam po tablici slučajnih brojeva slučajno u uzorak padne rezultat koji se već nalazi u uzorku, mi taj rezultat preskačemo i uzimamo idući rezultat.

Točan je i prvi i drugi prigovor! No, evo, zašto se tako radi: kao što je rečeno, sve statističke formule u vezi s uzorcima izrađene su po sistemu uzimanja uzoraka "uz povrat". No važno je da tako radimo samo kod vrlo *malih* populacija kakvima se služimo upravo za primjere radi demonstracije (npr. naša populacija sastoji se od samo pet rezultata). U stvarnom životu populacija je obično vrlo velika, katkad čak neizmjerno velika, tj. neograničena (npr. ispitivanje utjecaja tjelesnog napora na brzinu pulsa: "populaciju" predstavlja neizmjerno velik broj mjerenja!). Stoga u stvarnim životnim situacijama ne činimo praktički nikakvu pogrešku kada iz *velike* populacije uzimamo uzorke bez sistema "uz povrat".

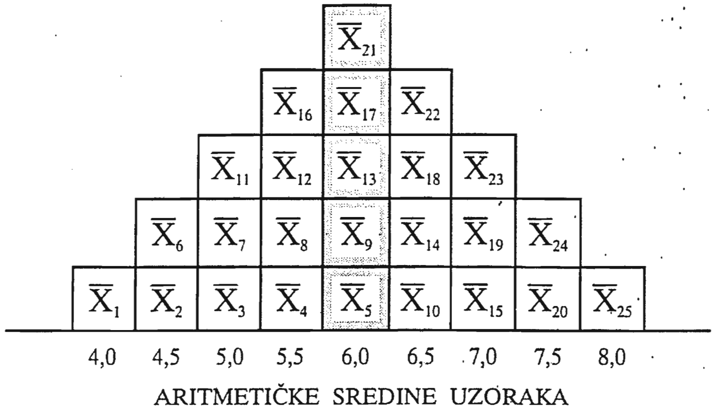
No u našem slučaju, budući da radimo s vrlo malom populacijom, moramo se tog pravila pridržavati ako želimo dokazati ono što namjeravamo. U tablici 9.6. navedeni su *svi mogući* uzorci veličine $N = 2$ iz naše populacije, koja se sastoji iz rezultata 4, 5, 6, 7 i 8. Za svaki smo uzorak izračunali njegovu aritmetičku sredinu i numerirali je rednim brojem.

Dobili smo, dakle, 25 aritmetičkih sredina, i ako ih sada u obliku kvadratića prikažemo u koordinatnom sustavu, *dobit ćemo distribuciju koja već ozbiljno podsjeća na normalnu distribuciju* (vidi sliku 9.3). A kada bi se populacija sastojala od, recimo, 100 rezultata, i kada bi iz nje vadili sve uzorke veličine $N = 30$ (zamislite koliko bi to silno mnogo uzoraka bilo s obzirom nato da se uzorci izvlače "uz povrat"!), onda bismo dobili praktički već *potpuno pravilnu normalnu distribuciju aritmetičkih sredina.*

Iz slike 9.3, naravno, vidi se i to da aritmetička sredina svih 25 aritmetičkih sredina uzoraka odgovara pravoj aritmetičkoj sredini populacije, koja iznosi 6.

TABLICA 9.6.
PREGLED SVIH MOGUĆIH UZORAKA VELIČINE $N = 2$ IZ POPULACIJE KOJA SE SASTOJI IZ REZULTATA 4, 5, 6, 7 i 8

Svi mogući uzorci veličine $N = 2$	Aritmetička sredina uzoraka	Naziv aritmetičke sredine
4,4	4,0	$\overline{X}_1$
4,5	4,5	$\overline{X}_2$
4,6	5,0	$\overline{X}_3$
4,7	5,5	$\overline{X}_4$
4,8	6,0	$\overline{X}_5$
5,4	4,5	$\overline{X}_6$
5,5	5,0	$\overline{X}_7$
5,6	5,5	$\overline{X}_8$
5,7	6,0	$\overline{X}_9$
5,8	6,5	$\overline{X}_{10}$
6,4	5,0	$\overline{X}_{11}$
6,5	5,5	$\overline{X}_{12}$
6,6	6,0	$\overline{X}_{13}$
6,7	6,5	$\overline{X}_{14}$
6,8	7,0	$\overline{X}_{15}$
7,4	5,5	$\overline{X}_{16}$
7,5	6,0	$\overline{X}_{17}$
7,6	6,5	$\overline{X}_{18}$
7,7	7,0	$\overline{X}_{19}$
7,8	7,5	$\overline{X}_{20}$
8,4	6,0	$\overline{X}_{21}$
8,5	6,5	$\overline{X}_{22}$
8,6	7,0	$\overline{X}_{23}$
8,7	7,5	$\overline{X}_{24}$
8,8	8,0	$\overline{X}_{25}$



Slika 9.3. Distribucija aritmetičkih sredina svih mogućih uzoraka veličine $N = 2$ iz populacije, koja se sastoji iz rezultata: 4, 5, 6, 7 i 8

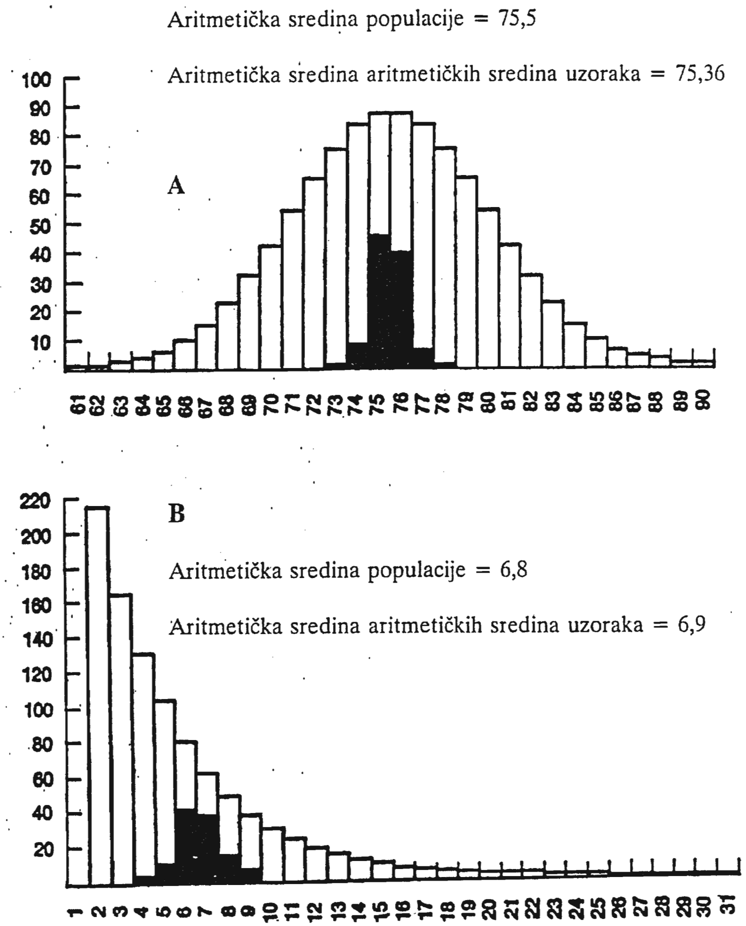
Zanimljivo je da će distribucija aritmetičkih sredina uzoraka iste veličine — pod pretpostavkom da uzorci nisu odveć mali, tj. manji od 30 — *tendirati normalnoj raspodjeli čak i onda ako populacija, iz koje uzorke uzimamo, nije normalno distribuirana!*

Ta je pojava vrlo važna u statistici i poznata je pod nazivom "teorem centralne granice" ("central limit theorem"), a važna je zato što nam omogućuje — kako ćemo vidjeti — stvaranje nekih zaključaka i onda kada radimo s populacijama koje nisu normalno distribuirane.

Evo dokaza "teorema centralne granice" na slici 9.4; na gornjem crtežu crno je prikazana distribucija 100 aritmetičkih sredina slučajnih uzoraka veličine $N = 30$, izvučenih iz jedne normalne populacije (koja je na slici također prikazana). Na donjoj slici prikazana je distribucija 100 aritmetičkih sredina uzoraka također veličine 30, uzetih iz populacije koja je — kako se vidi — totalno asimetrična.

9.2. STANDARDNA POGREŠKA ARITMETIČKE SREDINE

Spomenuli smo da aritmetička sredina, koju smo izračunali na našem konkretnom uzorku, *nije prava* aritmetička sredina populacije, već samo njezina procjena. Iz dosadašnjih demonstriranih pokusa, kao i iz dosadašnje diskusije o tome problemu, jasno je da će "procjena" neke vrijednosti (u našem slučaju aritmetičke sredine) biti to točnija što je uzorak veći i što je pojava koju mjerimo manje varijabilna. Kada neka pojava uopće ne bi varirala, onda bi nam i uzorak, koji se sastoji od samo *jednog jedinog* rezultata, već dao točnu "aritmetičku sredinu". Međutim, kako znamo, u biologiji, fiziologiji, psihologiji, sociologiji i mnogim drugim znanostima pojave koje mjerimo imaju svoj karakteristični varijabilitet.



Slika 9.4. Ako uzorci nisu suviše mali, distribucija njihovih aritmetičkih sredina dat će približno normalnu raspodjelu i onda ako su uzeti iz populacije koja nije normalno distribuirana

Prema tome, što neka pojava više varira, to se izlažemo većoj pogrešci kad iz aritmetičke sredine uzorka zaključujemo na *pravu* aritmetičku sredinu. S druge strane, kod pojava koje variraju izlažemo se, također, većoj pogrešci kad aritmetičku sredinu računamo iz *malog* broja rezultata (mali uzorak), nego kad je računamo iz *većeg* broja rezultata (veliki uzorak).

Prema tome, *pogreška* koja se veže uz svaku aritmetičku sredinu uzorka, bit će to veća što je pojava koju mjerimo više varijabilna i što je uzorak (broj mjerenja) manji.

Kako na varijabilnost neke pojave ne možemo djelovati, a na broj mjerenja možemo, to je jasno da možemo *povećanjem* broja mjerenja smanjiti pogrešku koja se povezuje uz naše mjerenje. No pogreška ne opada proporcionalno povećanju broja mjerenja, nego proporcionalno *drugom korijenu* broja mjerenja. Drugim riječima, aritmetička sredina dobivena iz 25 rezultata nije 25 puta "vjernija" od aritmetičke sredine dobivene iz *jednog* rezultata, nego je ona $\sqrt{25} = 5$ puta vjernija od aritmetičke sredine iz jednog rezultata.

Formula za izračunavanje pogreške, koja se veže uz neku aritmetičku sredinu uzorka, glasi:

$$s_{\overline{x}} = \frac{s}{\sqrt{N}}, \qquad (9.1)$$

pri čemu izraz $s_{\overline{x}}$ treba čitati: *standardna pogreška* aritmetičke sredine.

Matematički je potpuno jasno da je izraz (9.1) to veći što je brojnik veći, i što je nazivnik manji pa je, prema tome, i standardna pogreška aritmetičke sredine to veća što je standardna devijacija pojave koju mjerimo veća i što je broj mjerenja manji.

N a p o m e n a. Izraz s u brojniku formule (9.1) stavljen je ovamo zbog nužde, jer bi tu trebao stajati izraz σ , tj. standardna devijacija *populacije,* ali nam je ona, na žalost, vrlo rijetko poznata, pa zato ovamo uvrštavamo standardnu devijaciju *uzorka (s).* Zbog toga u tom slučaju ne dobivamo *pravu* standardnu pogrešku, već *procjenu* standardne pogreške. (*Prava* standardna pogreška pisala bi se:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{N}} .) \qquad (9.2)$$

Ta standardna pogreška aritmetičke sredine *nije ništa drugo nego nama već poznata standardna devijacija aritmetičkih sredina uzoraka oko prave aritmetičke sredine.* Drugim riječima, "standardna pogreška" također je standardna devijacija, ali dok je standardna devijacija (s) mjera variranja *individualnih* razultata oko njihove aritmetičke sredine, dotle je standardna pogreška ($s_{\overline{x}}$) mjera variranja *aritmetičkih sredina uzoraka* oko prave aritmetičke sredine populacije (μ). Stoga, kako ćemo vidjeti, za standardnu pogrešku vrijede isti zakoni kao i za standardnu devijaciju.

Ako, dakle, u nekom konkretnom slučaju nađemo da standardna pogreška neke aritmetičke sredine iznosi, primjerice, 2, onda to znači da se aritmetičke sredine uzoraka iste veličine (tj. one veličine koliko je velik naš uzorak) grupiraju oko *prave* aritmetičke sredine populacije po *normalnoj raspodjeli kojoj standardna devijacija iznosi 2.* No mi, na žalost, *ne znamo* pravu aritmetičku sredinu populacije, pa moramo "okrenuti" logiku zaključivanja (jer udaljenost od B do A jednaka je udaljenosti od A do B!), pa kazati ovako: ako standardna pogreška iznosi 2, onda naša dobivena aritmetička sredina (u uzorku) može od prave aritmetičke sredine odstupati praktički najviše za $3 \cdot 2 = 6$, pa bi prava aritmetička sredina gotovo sigurno morala biti negdje u intervalu "naša aritmetička sredina" ± 6.

Kako se vidi, služimo se logikom standardne devijacije: oko 68% rezultata nalazi se u intervalu $\overline{X} \pm 1s$, oko 95% rezultata u intervalu $\overline{X} \pm 2s$ i gotovo svi rezultati u intervalu $\overline{X} \pm 3s$.

Evo jednog primjera: uzmimo da smo pri nekom mjerenju dobili aritmetičku sredinu $\overline{X} = 90$, i standardnu devijaciju $s = 10$; mjerenje je izvršeno na 100 ispitanika. Standardna pogreška naše aritmetičke sredine iznosi:

$$s_{\overline{x}} = \frac{s}{\sqrt{N}} = \frac{10}{\sqrt{100}} = \frac{10}{10} = 1.$$

Ako je standardna pogreška 1, onda je 68% vjerojatno da naša aritmetička sredina ne odstupa od *prave* aritmetičke sredine više od plus ili minus 1, oko 95% je vjerojatno da od nje ne odstupa više od plus ili minus 2 (2 puta $s_{\overline{X}}$), a praktički je sigurno (ne posve, nego 99,7%!) da ne odstupa više od plus ili minus 3 (3 puta $s_{\overline{x}}$), pa se prava aritmetička sredina, dakle, gotovo sigurno nalazi negdje *u intervalu između 87 i 93* (90 ± 3).

Taj i takav interval, koji izvodimo iz vrijednosti standardne pogreške aritmetičke sredine, naziva se u statistici *"granicama pouzdanosti"*: 68%-tne granice pouzdanosti za našu aritmetičku sredinu 90 su 89 — 91, 95%-tne granice pouzdanosti su 88 — 92, i, konačno, 99,7%-tne granice pouzdanosti su — kako smo već izračunali — 87 i 93.

— — —

Čitalac se do sada vjerojatno već uvjerio da je bilo točno ono što smo rekli za normalnu raspodjelu, tj. da je ona temelj za statistički pojam vjerojatnosti, i da tek iz dobrog razumijevanja značenja normalne raspodjele možemo razumjeti i ono što se iz njezine logike dalje izvodi. Evo, sada upravo govorimo o toj logici normalne raspodjele! A kako je potrebno protumačiti još neke važne principe, to još bolje dokazati već iznesene principe, i dokazati neke tvrdnje koje su u tekstu bile spomenute, ali nisu bile dokazane — poslužit ćemo se još jednim primjerom. U tom primjeru *bit će ponavljanja,* tj. bit će sličnoga s dosadašnjim primjerima. No molim čitaoca neka se zbog toga ne ljuti. Moja (i tuđa) dugotrajna praksa u nastavi statistike pokazala je da je zapravo jedini ozbiljni "kamen smutnje" u razumijevanju statističkog načina rezoniranja — upravo pojam "standardne pogreške".

Zamislimo ponovno jednu vrlo malu populaciju, ovaj put od samo 4 rezultata: 2, 10, 4 i 8. U tablici 9.7. prikazani su svi postupci koje ćemo s tom populacijom izvoditi, pa upućujem čitaoca da uz tekst pažljivo prati i tablicu. Aritmetička sredina te populacije $\mu = 6$ (kako se radi o populaciji, aritmetičku sredinu bilježimo s μ), a standardna devijacija $\sigma = \sqrt{10}$. (Važna napomena: Budući da se radi o populaciji, standardna devijacija računata je s N, a ne $N-1$ u nazivniku.)

Ako sada — jednako kao što smo to učinili i s prijašnjom malom populacijom - iz nje "uz povrat" uzimamo sve moguće uzorke veličine $N = 2$, dobit ćemo uzorke prikazane u tablici u stupcu 4. Njihove aritmetičke sredine prikazane su u stupcu 5. Aritmetičkih sredina ima 16, jednako koliko i uzoraka. To je sada *populacija* uzoraka i njihovih aritmetičkih sredina (tj. to su svi mogući uzorci veličine 2; svaki daljnji bio bi samo ponavljanje onoga što već posjedujemo). Ta populacija aritmetičkih sredina uzoraka, prikazana u stupcu 5, ima naravno i *svoju* aritmetičku sredinu.

kao i svoju standardnu devijaciju. Aritmetička sredina tih 16 vrijednosti ($\mu_{\overline{x}}$) iznosi $96/16 = 6$, iz čega proizlazi *prvo pravilo*, nama već poznato:

1. *Aritmetička sredina svih mogućih aritmetičkih sredina uzoraka iste veličine jednaka je "pravoj" aritmetičkoj sredini, tj. aritmetičkoj sredini populacije.*

TABLICA 9.7.
KARAKTERISTIČNA SVOJSTVA ARITMETIČKIH SREDINA I STANDARDNIH DEVIJACIJA (VARIJANCA) UZORAKA

Originalna populacija			Uzorci veličine $N = 2$	Populacija arit. sredina			Populacija varijanci
1	2	3	4	5	6	7	8
X	$X - \mu$	$(X - \mu)^2$		$\overline{X}$	$\overline{X} - \mu_{\overline{x}}$	$(\overline{X} - \mu_{\overline{x}})^2$	s^2
2	−4	16	2,2	2	−4	16	0
10	4	16	2,10	6	0	0	32
4	−2	4	2,4	3	−3	9	2
8	2	4	2,8	5	−1	1	18
			10,2	6	0	0	32
$\Sigma = 24$		$\Sigma = 40$	10,10	10	4	16	0
$\mu = 24/4 = 6$			10,4	7	1	1	18
$\sigma^2 = 40/4 = 10$			10,8	9	3	9	2
			4,2	3	−3	9	2
			4,10	7	1	1	18
			4,4	4	−2	4	0
			4,8	6	0	0	8
			8,2	5	−1	1	18
			8,10	9	3	9	2
			8,4	6	0	0	8
			8,8	8	2	4	0
				$\Sigma = 96$	$\Sigma = 0$	$\Sigma = 80$	$\Sigma = 160$
				$\mu_{\overline{x}} = 96/16 = 6$		$\mu_{s^2} = 160/16 = 10$	
				$\sigma^2_{\overline{x}} = 80/16 = 5$			

U stupcima 6 i 7 proveden je račun za dobivanje standardne devijacije tih aritmetičkih sredina oko "prave" aritmetičke sredine. (I ovdje je u nazivniku uzeto $N = 4$, a ne $N - 1 = 3$, jer se radi o *populaciji*, dakle o svim mogućim slučajevima.) Kvadrirana standardna devijacija (varijanca) iznosi $80/16 = 5$. Usporedimo li tu varijancu (5) s varijancom u populaciji (koja iznosi 10), slijedi drugo pravilo:

2. *Varijanca populacije aritmetičkih sredina uzoraka jednaka je varijanci originalne populacije, podijeljenoj s veličinom uzorka.* (U našem slučaju to je $10/2 = 5$.) (To se može vidjeti i iz slike 9.2: ako se iz standardnih devijacija, koje pišu uz pojedinu krivulju, kvadriranjem dobije varijanca, ta vrijednost uvijek odgovara varijanci originalne populacije ($\sigma = 100$) podijeljenoj veličinom uzorka.)

Taj drugi princip služi nam da *razumijemo* formulu kojom izračunavamo standardnu pogrešku aritmetičke sredine: kao što je rečeno, standardna pogreška aritmetičke sredine nije ništa drugo nego standardna devijacija aritmetičkih sredina uzoraka oko prave aritmetičke sredine populacije:

$$\text{varijanca aritmetičkih sredina uzoraka} = \frac{\text{varijanca populacije}}{\text{veličina uzorka}}$$

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{N} \tag{9.3}$$

Dakle, $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{N}}$, a to je formula 9.1, koju već poznajemo.

U stupcu 8 izračunate su varijance svih 16 uzoraka. No budući da se sada radi o *uzorcima*, varijancu računamo s $N - 1$ u nazivniku. Tako imamo u prvom uzorku $s^2 = 0$ (jer pojedinačni rezultati 2 i 2 ne odstupaju od aritmetičke sredine uzorka, koja je 2), u drugom uzorku, u kojem je aritmetička sredina 6, razlike su 4, 4, pa je prema tome $\dfrac{4^2 + 4^2}{2 - 1} = 32$, u trećem uzorku $s^2 = \dfrac{1 + 1}{1} = 2$, u četvrtom uzorku $\dfrac{9 + 9}{1} = 18$, itd. *Prosječna* varijanca $= 160/16 = 10$, iz čega proizlazi treće pravilo:

3. *Varijance uzoraka čine takvu raspodjelu oko prave varijance da im aritmetička sredina odgovara pravoj varijanci:*

$$\mu_{s^2} = \sigma^2$$

Sada se ujedno i vidi razlog zašto pri izračunavanju standardne devijacije *uzorka* treba u nazivniku imati $N - 1$; *jedino u tom slučaju* izračunata varijanca bit će "nepristrana" procjena prave varijance, jer bi — u velikom broju mjerenja — prosjek tako dobivenih varijanci odgovarao stvarnoj varijanci u populaciji. Naprotiv, prosječna varijanca, dobivena iz varijanci uzoraka koje su računate s N u nazivniku, *ne bi* odgovarala pravoj varijanci populacije (ona bi bila *manja*), i u tom slučaju imali bismo tzv. "pristranu" ("biased") procjenu varijance.

A sada ukratko rezimirajmo sve ono što je potrebno za razumijevanje ovog područja: aritmetička sredina uzorka ($\overline{X}$) rijetko će imati istu vrijednost kao i aritmetička sredina populacije (μ). Ali pod pretpostavkom da smo naš uzorak dobili nepristranom metodom, njegova aritmetička sredina neće pokazati nikakvu sistematsku tendenciju bilo da bude veća ili manja od aritmetičke sredine populacije. Dakle, aritmetička sredina uzoraka je relativno najbolje moguće nagađanje o aritmetičkoj sredini populacije koje možemo učiniti. Naprotiv, ako bismo standardnu devijaciju uzorka izračunavali onako kako je izračunavamo kod populacije (tj. s N u nazivniku), onda ona ima sistematsku tendenciju da bude *manja* od standardne devijacije populacije kojoj uzorak pripada. Prema tome, ako bismo takvu standardnu devijaciju uzorka koristili kao nagađanje i "procjenu" standardne devijacije populacije, gotovo je sigurno da ćemo *podcijeniti* standardnu devijaciju populacije, i to tim više što je uzorak manji. Kako smo iz primjera vidjeli, to podcjenjivanje neće se dogoditi ako pri izračunavanju standardne devijacije uzorka računamo s $N - 1$ u nazivniku.

A sada da još samo rastumačimo *zašto* bi prosječna standardna devijacija (odnosno varijanca), izračunata iz populacije standardnih devijacija uzoraka, koje bi bile računate s N u nazivniku, bila *manja* od prave standardne devijacije. Evo zašto: pri računanju standardne devijacije, mi — kako je poznato — računamo razlike između svakog rezultata i aritmetičke sredine. No, ta aritmetička sredina je aritmetička sredina *uzorka*, a ne prava aritmetička sredina populacije. Poznato je da je suma kvadriranih razlika između niza rezultata i *njihove* aritmetičke sredine manja od sume kvadriranih razlika između tih istih rezultata i bilo koje druge vrijednosti. (Aritmetička sredina stoga se često i definira tako da se kaže da je to vrijednost s najmanjom mogućom sumom kvadrata razlika.) Praktički je sigurno da aritmetička sredina našeg uzorka nije jednaka pravoj aritmetičkoj sredini, već od nje manje ili više odstupa. Iz toga proizlazi da bi suma kvadrata razlika između svakoga pojedinog rezultata i prave aritmetičke sredine bila veća nego što je suma kvadriranih razlika između svakoga pojedinog rezultata i njihove aritmetičke sredine! Prema tome, standardna devijacija, izračunata na osnovi razlika prema pripadnoj aritmetičkoj sredini, manja je od one koju bismo trebali dobiti. Da bismo tu pogrešku korigirali, mi smanjujemo vrijednost nazivnika (tj. oduzimamo 1 od N), i tako povećavamo vrijednost standardne devijacije.

9.3. RAZLIKA IZMEĐU ARITMETIČKIH SREDINA VELIKIH NEZAVISNIH UZORAKA

Jedan od najčešćih slučajeva pri eksperimentiranju i statističkoj obradi podataka jest ***uspoređivanje dviju aritmetičkih sredina*** i testiranje *razlike* među njima. Vjerojatno među čitaocima nema ni jednoga tko nije barem nekoliko puta u životu čuo podatak "da je neka razlika ***statistički značajna***". Mnogi su možda znali što to znači, ali mnogi i nisu, nego su vjerojatno mislili da "statistički značajna razlika" znači neku "veliku razliku". No pojam "statistički značajne razlike" ima u statistici sasvim određen i definiran smisao: ako kažemo da je neka razlika statistički značajna, onda smo zapravo ustvrdili da razlika koja je nađena (bez obzira na veličinu razlike!) *nije slučajna, već da razlika vrlo vjerojatno postoji i među populacijama*. Naprotiv, ako ustvrdimo da neka razlika nije statistički značajna, to drugim riječima znači da razlika koju smo prilikom našeg mjerenja dobili, može biti i *slučajna* posljedica variranja uzoraka, a da među populacijama, kojima ti uzorci pripadaju, možda i nema nikakve razlike.

Ubrzo ćemo potanje opisati smisao "statistički značajne" razlike, a sada ćemo najprije riješiti jedan praktični problem.

Uzmimo da smo jednim testom znanja fizike testirali 900 studenata prilikom primanja na elektrotehnički fakultet i da smo dobili ovaj rezultat:

$$\overline{X}_1 = 120 \text{ bodova}, \quad s_1 = 10 \text{ bodova},$$

a 865 studenata arhitektonskog fakulteta na tom istom testu postiglo je ove rezultate:

$$\overline{X}_2 = 123 \text{ boda}, \quad s_2 = 11 \text{ bodova}.$$

Pitanje glasi: Je li razlika (u ovom slučaju razlika iznosi samo 3 boda) statistički značajna, tj. bismo li dobili razliku u korist studenata arhitekture i onda kad bismo testirali još mnogo drugih uzoraka, dakle još mnogo drugih grupa studenata arhitekture i studenata elektrotehnike?

Budući da smo mi ispitali samo dva *uzorka* (a ne sve potencijalne studente obaju fakulteta), sigurno je da se uz naš rezultat vezuje neka pogreška, koju ćemo, analogno standardnoj pogrešci aritmetičke sredine, nazvati ***standardnom pogreškom razlike*** između dviju aritmetičkih sredina.

Standardna pogreška između dviju aritmetičkih sredina kod velikih uzoraka računa se prema formuli:

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{s_{\overline{x}_1}^2 + s_{\overline{x}_2}^2}. \tag{9.4}$$

Budući da umjesto $s_{\overline{x}}$ možemo pisati $\frac{s}{\sqrt{N}}$, to formulu (9.4) možemo pisati i ovako:

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}. \tag{9.5}$$

Izraženo riječima, standardna pogreška razlike između dvije aritmetičke sredine jednaka je drugom korijenu iz sume kvadriranih standardnih pogrešaka obiju aritmetičkih sredina.

Ako su uzorci *veliki* (blizu 100 ili nekoliko stotina), ta je formula posve pouzdana, pa makar oba uzorka nisu slične veličine. Naprotiv, ako su uzorci *mali* (naročito ako su veličine ispod 30), kao i kod većih uzoraka, ali ako se značajno *razlikuju* po veličini (npr. ako je jedan veličine 30, a drugi 60), dobit ćemo preciznije rezultate ako radimo prema formuli za male *uzorke*, što će biti obrađeno u posebnom potpoglavlju. No već i sada možemo pripomenuti nešto o čemu će kasnije još biti govora, da je ***bolje ako su uzorci jednake ili slične veličine.***

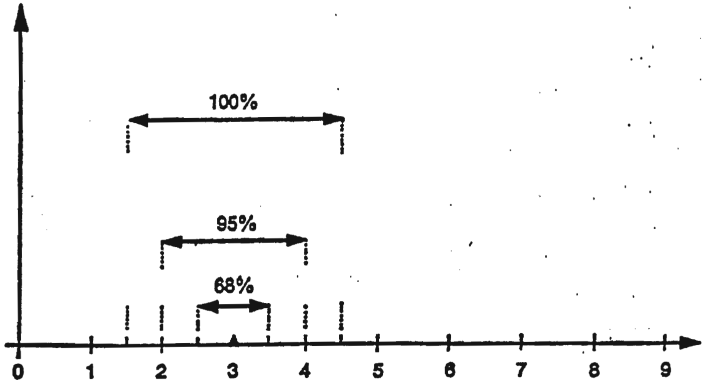
U našem primjeru dobivamo:

$$s_{\overline{x}_1} = \frac{s_1}{\sqrt{N_1}} = \frac{10}{\sqrt{900}} = 10/30 = 0,33$$

$$s_{\overline{x}_2} = \frac{s_2}{\sqrt{N_2}} = \frac{11}{\sqrt{865}} = 11/29,4 = 0,37$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{0,33^2 + 0,37^2} = \sqrt{0,1089 + 0,1369} = \sqrt{0,2458} = 0,5$$

Malo prije smo ustanovili da za standardnu pogrešku vrijede isti zakoni kao i za standardnu devijaciju, pa, prema tome, možemo zaključiti da imamo oko 68% vjerojatnosti da naša dobivena razlika ($\overline{X}_2 - \overline{X}_1 = +3$) ne odstupa od *prave* razlike više od $\pm 0,5$, 98% vjerojatnosti, da ne odstupa više od $\pm 1,0$ (iz 2 puta 0,5) i praktički 100% vjerojatnosti da ne odstupa više od ± 3 puta 0,5 = 1,5. Dakle, 68%-tne granice pouzdanosti naše razlike su 2,5 — 3,5, 95%-tne granice pouzdanosti su 2,0 — 4,0 i gotovo 100%-tne granice pouzdanosti su 1,5 — 4,5. Te su granice prikazane na slici 9.5.



Slika 9.5. Granice pouzdanosti neke dobivene razlike od −3, ako standardna pogreška te razlike iznosi 0,5

Kako se iz slike vidi, u *svakom* slučaju *prava razlika je veća od nule,* i zato ćemo zaključiti da je razlika ***statistički značajna***. Iz slike se naime vidi da se praktički ne može dogoditi da razlika među populacijama bude nula, tj. da obje skupine studenata budu posve jednake po uspjehu u tome testu. Zato možemo sa sigurnošću reći da su studenti arhitektonskog fakulteta sigurno bolji u tom testu od studenata elektrotehničkog fakulteta. (Druga je stvar da ta razlika, iako ***statistički*** značajna, ne mora uopće biti ***praktički*** značajna, jer razlika od 3 boda možda je praktički bez značenja!)

Evo drugoga primjera:

Mjereći nekim testom znanje 36 dječaka u jednom i 36 dječaka u drugom razredu dobili smo ove rezultate:

$\overline{X}_1 = 120$ bodova $\qquad \overline{X}_2 = 123$ boda
$s_1 = 20$ bodova $\qquad s_2 = 20$ bodova
$s_{\overline{x}_1} = \sqrt{400/36} = 3,33 \qquad s_{\overline{x}_2} = \sqrt{400/36} = 3,33.$

Je li razlika u uspjehu između te dvije skupine statistički značajna?

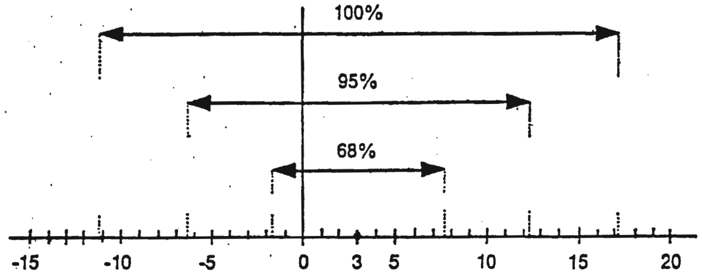
Izračunamo li standardnu pogrešku ove razlike, dobivamo:

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{s^2_{\overline{x}_1} + s^2_{\overline{x}_2}} = \sqrt{3,33^2 + 3,33^2} = \sqrt{22,18} = 4,71.$$

Dakle, oko 68% je vjerojatno da nađena razlika (+3) ne odstupa od prave razlike više od ±4,71, oko 95% je vjerojatno da ne odstupa više od $2 \cdot 4,71 = 9,42$, a gotovo 100% je sigurno da ne odstupa više od $3 \cdot 4,71 = 14,13$.

Ako to prikažemo grafički, dobivamo situaciju prikazanu na slici 9.6.

Kako se iz slike vidi, već 68%-tne granice pouzdanosti zahvaćaju *nulu,* što, drugim riječima, znači da možda među populacijama i nema razlike, ili čak (vidi 68%-tne, 95%-tne i 100%-tne granice na slici!) da je razlika među populacijama "*obratna*" (to su na slici vrijednosti u zoni negativnih bodova), tj. da je zapravo možda *prva* grupa ispitanika bolja od druge! Prema tome, ovu razliku *ne možemo* smatrati statistički značajnom.



Slika 9.6. Granice pouzdanosti razlike +3, ako standardna pogreška te razlike iznosi 4,71

Ovaj izneseni način prikazivanja smisla standardne pogreške razlike između dvije aritmetičke sredine, iako je logički opravdan, ipak nam ne dopušta neke finije zaključke u vezi sa značajnošću razlike. Stoga ćemo sada protumačiti jedan drugi, statistički mnogo više uobičajen, način prikazivanja rezoniranja pri testiranju značajnosti razlike između dvije aritmetičke sredine.

Zamislite jednu kutiju sa žetonima, koji predstavljaju jednu veliku populaciju distribuiranu po normalnoj raspodjeli. Kada bismo lijevom rukom zagrabili u kutiju i izvukli iz nje 30 žetona, mogli bismo iz tog uzorka izračunati aritmetičku sredinu. Ako nakon toga desnom rukom također izvadimo 30 žetona (ako je populacija vrlo velika, nije naročito važno da radimo metodom "uz povrat"), i iz ovog uzorka možemo izračunati aritmetičku sredinu. Iako oba uzorka pripadaju *istoj* populaciji, nije baš mnogo vjerojatno da će obje aritmetičke sredine biti istovjetne, jer će zbog pukog slučaja jedan od uzoraka imati veću, a drugi manju aritmetičku sredinu. Prema tome, u prvom paru uzoraka imamo neku *razliku* među aritmetičkim sredinama, i tu ćemo razliku zabilježiti; na primjer, ako je razlika u korist uzorka iz desne ruke, veličinu razlike označit ćemo oznakom plus, a ako je razlika u korist uzorka lijeve ruke, onda oznakom minus. Nastavimo li takvo vađenje u parovima nekoliko tisuća puta, dobit ćemo nekoliko tisuća razlika između aritmetičkih sredina uzoraka lijeve i desne ruke. Prema nama već poznatim zakonima, tih nekoliko tisuća razlika, ako ih prikažemo grafički, distribuirat će se po *normalnoj raspodjeli,* kojoj će aritmetička sredina biti *prava* razlika. A budući da smo parove uzoraka vadili iz *iste* populacije, jasno da među aritmetičkim sredinama populacije "lijeve ruke" i populacije "desne ruke" nema nikakve razlike, pa je prema tome *prava* razlika između obiju populacija — nula. No kako se vidi, i kada među populacijama nema razlike, mogu se među uzorcima naći i veće razlike (čas u korist jedne, čas u korist druge grupe). Koje su karakteristike te normalne raspodjele razlika među aritmetičkim sredinama? Za aritmetičku sredinu znamo: to je prava razlika među populacijama, u našem slučaju to je nula. A standardna devijacija te raspodjele?

Vjerujem da već pogađate! Jednako kao što je standardnu devijaciju raspodjele aritmetičkih sredina uzoraka oko prave aritmetičke sredine populacije predstavljao izraz koji smo nazvali "standardna pogreška aritmetičke sredine", jednako tako standardnu devijaciju razlika među aritmetičkim sredinama uzoraka predstavlja izraz koji nazivamo "standardna pogreška razlike među aritmetičkim sredinama", i čiju formulu već znamo.

Vratimo se sada našim primjerima: u prvom primjeru imali smo 900 studenata u jednom i 865 studenata u drugom uzorku, i među prosjecima smo našli razliku u bodovima od 3 boda u korist grupe 2. Standardna pogreška te razlike, izračunata prema formuli (9.4), iznosila je 0,5.

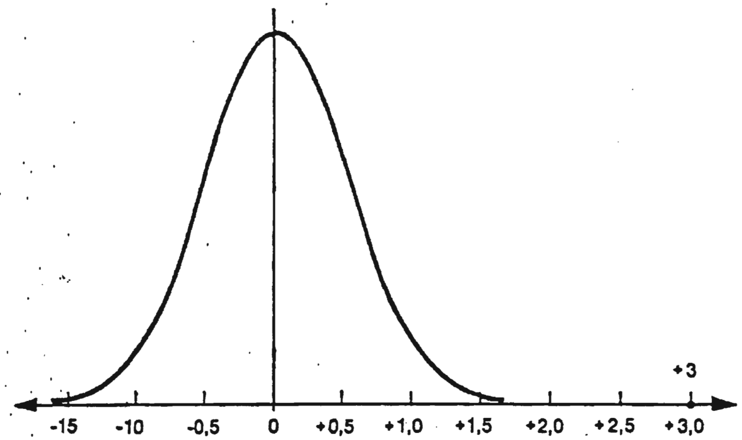
Kako ta vrijednost 0,5 predstavlja standardnu *devijaciju* razlika među aritmetičkim sredinama uzoraka, to rezoniramo — u skladu s maloprije iznesenom logikom — ovako:

Kada među populacijama *ne bi bilo nikakve razlike* (dakle kada bi "prava" razlika među njima bila 0), pa kada bi iz tih populacija vadili slučajne uzorke veličine $N_1 = 900$ i $N_2 = 865$, dobili bismo (nakon mnogo izvlačenja uzoraka) normalnu raspodjelu razlika, kojoj bi aritmetička sredina bila nula, a standardna devijacija 0,5. Na slici 9.7. ta je distribucija prikazana, a prikazana je i nađena razlika od +3.

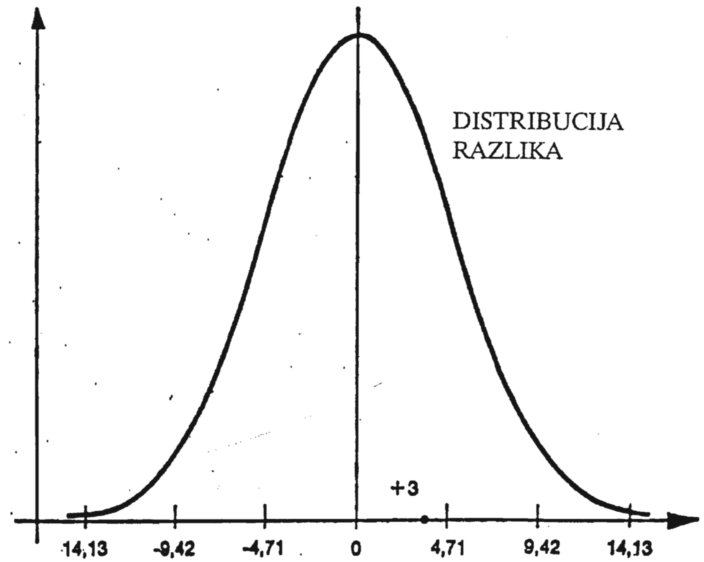
Kako se iz slike vidi, kada među populacijama ne bi bilo razlike, *slučajno* bismo mogli naići na razlike kod uzoraka sve do oko + ili $-1,5$ (tri puta standardna pogreška), ali teško je vjerovati da bismo *slučajno* mogli dobiti razliku veličine 3, jer se ona nalazi 6 standardnih devijacija udaljena od aritmetičke sredine! Stoga ćemo smatrati da dobivena razlika *nije slučajna*, nego da je *statistički značajna*, tj. naša dobivena razlika pripada populaciji razlika kojima aritmetička sredina *nije nula*.

U našem drugom primjeru, s uzorcima veličine 36, nađena je također razlika od 3 boda, ali je standardna pogreška te razlike (zato što su raspršenja rezultata u uzorku bila veća, i što je mjerenje izvršeno na malom broju) iznosila 4,71.

Ako tu situaciju prikažemo grafički, dobivamo normalnu raspodjelu, prikazanu na slici 9.8.



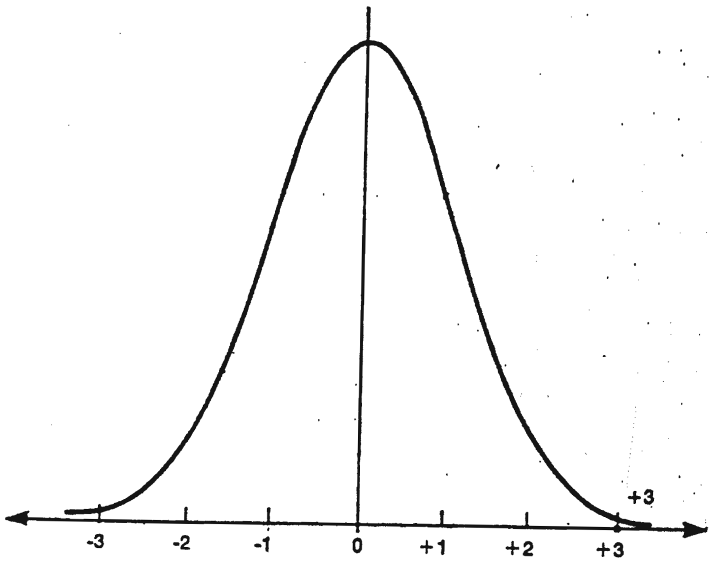
Slika 9.7. Ako standardna pogreška razlike iznosi 0,5, nađena razlika od +3 nalazi se *izvan* distribucije mogućih razlika, koje bi se slučajno mogle dogoditi kada među populacijama ne bi bilo nikakve razlike



Slika 9.8. Ako ustvari nema razlike među populacijama, ipak bismo po zakonu slučaja mogli naći i razlike koje idu do iznad 14. Naša razlika od 3 prema tome može vrlo lako biti i posve slučajna

Kako se na slici vidi, naša dobivena razlika od 3 posve dobro "pristaje" u ovu distribuciju razlika, tj. vrlo je lako moguće da je ona slučajno dobivena, jer — kako nam distribucija pokazuje — čak bismo mogli (iako među populacijama *nema* razlike) slučajno dobiti i razlike koje su veće od 14!

Prema tome, kao osnovno *logičko* pravilo nametao bi se zaključak: ako je neka razlika između dviju aritmetičkih sredina barem tri puta veća od svoje vlastite pogreške, onda je možemo smatrati statistički značajnom, jer je vrlo malo vjerojatno da će se tako velika razlika dogoditi slučajno. To ilustrira i slika 9.9.



Slika 9.9. Razlika, koja je 3 puta veća od svoje vlastite pogreške, pada posve na kraj distribucije razlika, kojima je aritmetička sredina nula

Koliko je puta neka razlika veća od svoje pogreške možemo, naravno, ustanoviti vrlo jednostavno tako što ćemo neku *razliku podijeliti njezinom pogreškom.* Taj odnos poznat je u statistici pod nazivom t-odnos (nekada su ga zvali i C. R. odnos, tj. "critical ratio", dakle "kritički odnos") i računa se ovako:

$$t = \frac{\text{razlika}}{\text{standardna pogreška te razlike}}$$

$$= \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{x}_1 - \overline{x}_2}} \tag{9.6}$$

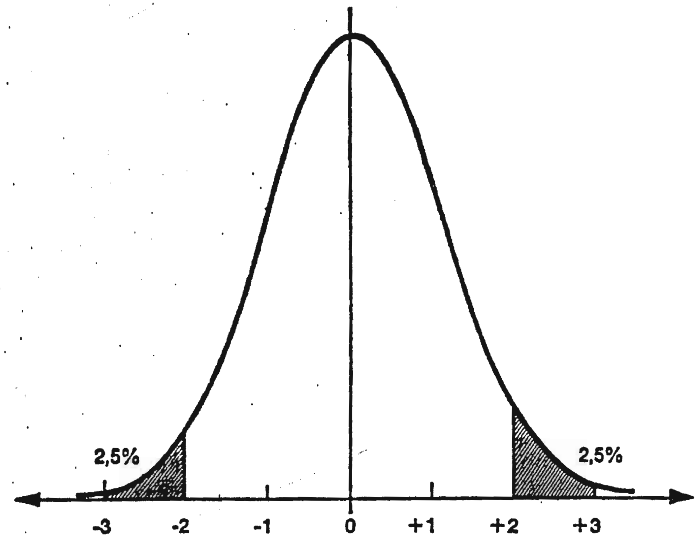
Iz formule će pažljivi čitalac brzo ustanoviti da t zapravo predstavlja jednu vrstu z-vrijednosti: kao i z, i t je jedno *odstupanje* izraženo u terminima *standardne devijacije*; kao što 1 z znači da se neki rezultat nalazi točno jednu standardnu devijaciju

iznad prosjeka, tako i $t = 1$ znači da je neka nađena razlika među aritmetičkim sredinama na jednoj standardnoj devijaciji raspršenja svih slučajnih razlika koje se mogu dogoditi, pa makar među aritmetičkim sredinama populacija nema razlike.

Međutim, suvremeni statističari smatraju da je zahtjev da razlika bude tri puta veća od svoje pogreške suviše strog kriterij, i da time često proglašujemo statistički neznačajnim i one razlike koje su zapravo značajne, pa je zato taj kriterij danas nešto ublažen te se većinom traži da razlika bude oko *dva puta* (točnije 1,96 puta) veća od svoje pogreške. Rečeno statističkom terminologijom, uzima se tzv. *razina značajnosti* od 5%, a to znači ovo: ako zapravo (u populaciji) *n e p o s t o j i nikakva razlika između dviju aritmetičkih sredina, onda bi se takva konkretna razlika koju smo dobili, mogla s l u č a j n o dogoditi samo pet puta u 100 mjerenja (ili jedanput u 20), a to je m a l o v j e r o j a t n o*, pa zato možemo uzeti da je razlika statistički značajna.

Prema tome, razina slučajnosti od 5% znači zapravo "šansu od 5%, da smo pogriješili".

Izraženo grafički, to bi izgledalo kao na sl. 9.10. Ako dobivena razlika padne u interval $0 \pm 1,96\ s_{\overline{x}_1 - \overline{x}_2}$ nećemo je smatrati statistički značajnom.



Slika 9.10. Razliku koja padne izvan 1,96 (praktički 2) svoje pogreške, možemo smatrati statistički značajnom na "razini značajnosti" od 5%

Ako padne *izvan* tog intervala, možemo je smatrati statistički značajnom na razini značajnosti od 5%. Dakle, ako je razlika 1,96 (okruglo dva) puta veća od svoje pogreške (jer se 95% rezultata nalazi u intervalu $\overline{X} \pm 1,96s$), ona se smatra značajnom "na razini značajnosti od 5%". Prema tome, kod velikih uzoraka razlika je značajna ako je t veći od 1,96.

Treba priznati da izraz "razina značajnosti" (ili "razina signifikantnosti") nije baš sam po sebi najsretnije izabran izraz, i stoga neki statističari, u želji da značenje izraza bude što jasnije, tu istu stvar nazivaju "nivo rizika". Prema tome, kada kažemo da je neka razlika "statistički značajna na nivou od 5%", time zapravo mislimo da smatramo "da među populacijama stvarno postoji razlika, pri čemu riskiramo oko 5% da smo ipak izveli pogrešan zaključak".

Ako ustanovimo da je neka razlika statistički značajna na nivou od 5% ili manjem (dakle ako je $t = 1,96$ ili više), napisat ćemo na kraju računa visinu dobivenog izraza t, i iza toga izraz $P < 0,05$ što — "prevedeno" na govorni jezik — znači: "vjerojatnost da smo pogriješili (P) je manja od 5%".

Razina značajnosti od 5% samo je *najčešća* razina, koja se u većini slučajeva upotrebljava, ali je razumljivo da nam je u nekim slučajevima potrebna i veća sigurnost, pa onda uzimamo stroži kriterij; na primjer razinu značajnosti od 1%, što znači da imamo šansu od samo 1% da smo pogriješili (u tom slučaju t mora biti najmanje 2,58; vidi tablicu A, ali ne zaboravi da nam tablica pokazuje samo *jednu* stranu krivulje, pa, prema tome, kad uz z od 2,60 čitamo 0,00466, što znači 0,466%, za obje strane krivulje, to je oko 1%!).

Odluka o tome koju ćemo razinu značajnosti upotrijebiti (5%, 1% ili niže) ovisi o mnogo faktora, a jedan od glavnih je važnost *posljedica* u slučaju pogrešnog zaključka. Tako, na primjer, ako bismo testirali razliku u utjecaju dvaju medikamenata od kojih se jedan pokazao vrlo opasan (pa je zato isključen iz upotrebe), i kad bismo htjeli biti uvjereni da se novi (drugi) medikament *sigurno* razlikuje od prvoga u tome što je manje opasan, morali bismo uzeti veći stupanj sigurnosti, dakle, recimo, nivo značajnosti od 0,01 (1%), ili još stroži (0,001). Nasuprot tome, ako ispitujemo eventualno postojanje nekog "nuz-efekta" kod jednog sredstva za umirenje, bolje je priznati da taj efekt postoji (dakle proglasiti ga značajnim), nego ga previdjeti, tj. smatrati da on zapravo ne postoji. U tom je slučaju opravdaniji *blaži* nivo značajnosti (0,05, pa čak po mišljenju nekih i 0,10).

Ako razlika (prema upotrijebljenom kriteriju) *nije* statistički značajna, iza vrijednosti dobivenog t-izraza napisat ćemo: $P > 0,05$, što znači da je šansa za pogrešku veća od kriterija koji smo odabrali (na primjer 5%).

U *svakom* slučaju, tj. bilo uz veći ili manji oprez (niži ili viši nivo značajnosti), izlažemo se riziku da učinimo jednu od pogrešaka koje su u statistici poznate pod nazivima alfa i beta pogreške, ili pogreška tipa 1 (alfa) i pogreška tipa 2 (beta): uz stroži kriterij (npr. $P = 0,01$) izlažemo se riziku da neku razliku ne proglasimo značajnom (tj. smatramo da dvije aritmetičke sredine pripadaju istoj populaciji), iako među populacijama zapravo razlika *postoji* (pogreška tipa 2), a uz blaži kriterij (npr. $P = 0,05$) izlažemo se riziku da proglasimo da se dvije aritmetičke sredine *razlikuju*, a zapravo među aritmetičkim sredinama populacije nema razlike (pogreška tipa 1). Obje ove situacije prikazane su u tablici 9.8.

TABLICA 9.8.
POGREŠKE PRI ZAKLJUČIVANJU IZ UZORKA NA POPULACIJU

Odluka	Stanje u populaciji	
	Nema razlike između dvije aritmetičke sredine	*Postoji* razlika između dvije aritmetičke sredine
Odbacujemo nul-hipotezu	Pogreška tipa 1 (alfa)	Nema pogreške
Prihvaćamo nul-hipotezu	Nema pogreške	Pogreška tipa 2 (beta)

Iz toga proizlazi da ne možemo znati kojoj se pogrešci izlažemo, a u konkretnom slučaju treba preuzeti rizik one pogreške čije su posljedice manje štetne (npr. treba uzeti niski P u opisanom testiranju medikamenata i prema tome riskirati pogrešku 2, jer bi pogreška 1 mogla biti štetnija u praksi).

Tu se — kako kaže statističar Rowntree — radi o sličnoj dilemi kao kod procjene krivnje optuženoga: ako smo skloni prihvatiti i problematične dokaze o optuženikovoj krivnji, riskiramo da kaznimo mnogo nedužnih ljudi. A ako pak ne uzimamo u obzir sve, nego samo najjače dokaze o nečijoj krivnji, tada će mnogi ljudi, koji *jesu* krivi, ostati nekažnjeni.

U novije vrijeme neki statističari kritiziraju preveliku ***fiksiranost*** istraživača u konvencionalne i standardne kriterije razine značajnosti, tj. uz kriterije $P = 0,05$ (5%) i $P = 0,01$ (1%). Dobro je, doduše, da među istraživačima postoji određena dogovorena metodologija statističkog rada, ali — misle neki — tu se često pretjeruje i kriterij od 5% prelazi u ***opsesiju!***

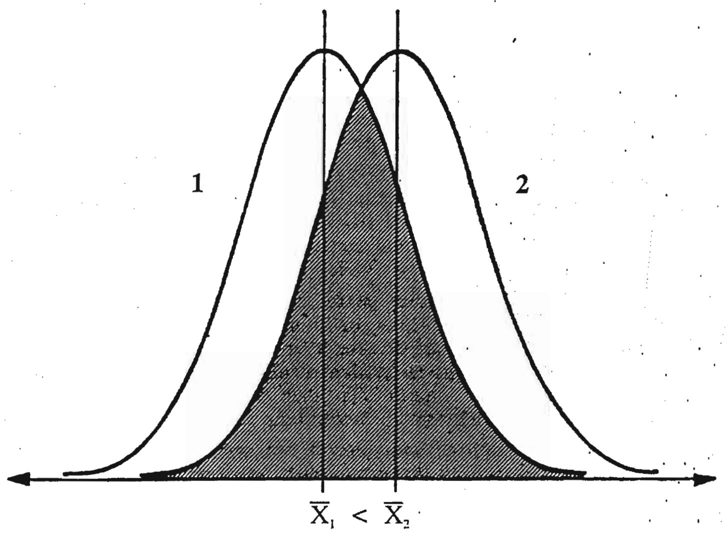
Skipper i suradnici u jednom svom radu na temu "nepovredivosti" nivoa od 0,05 kažu ovo: "Ne bi bilo teško naći dokaze koliko se neki istraživač raduje kada kod F-testa ili t-testa ustanovi značajnost na nivou od 0,05, a kolik je njegov užas ako iz tablice pročita značajnost od 'samo' 0,10 ili 0,06. Gotovo kao da razlika između 0,05 i 0,06 znači razliku između 'dobra' i 'zla', 'časnog' i 'nečasnog', 'uspjeha' i 'neuspjeha'."

Ovdje, dakle, treba ponoviti ono što je već rečeno o izboru nivoa značajnosti: želimo li biti praktički potpuno sigurni da neki postupak ima efekta (jer se u protivnom možda taj postupak, koji je vrlo skup, ne isplati), onda treba uzeti ***stroži*** nivo, primjerice 1%, ili čak još niže. Naprotiv, radi li se o tome da neki novi postupak nije ništa kompleksniji ni skuplji, ni opasniji, itd., od dosadašnjega, a ***čini se*** da bi mogao biti bolji od dosadašnjega, bilo bi besmisleno imati prestrogi kriterij i u tom ćemo slučaju naravno uzeti "blaži" nivo značajnosti, na primjer nivo 0,05, pa možda čak i viši. Kada bi, pretpostavimo, neki neurokirurg tvrdio da određena modifikacija dosadašnje operativne tehnike tumora (a ta modifikacija *sigurno* ne može bolesniku nanijeti nikakve štete) uzrokuje veći postotak ozdravljenja, onda bismo na tu metodu vjerojatno pristali čak kada bi razlika između dosadašnje i nove metode bila statistički značajna samo na nivou od 20%, pa možda čak i 30%, jer je ***još uvijek nešto vjerojatnije da je nova metoda bolja od dosadašnje nego što je vjerojatno da među njima nema razlike!***

N a p o m e n a. Dakako, naše "privatno" ponašanje vjerojatno bismo regulirali po tom principu, jer — kada bi se, uzmimo, radilo o nadi u spasavanje života — pristali bismo na novu metodu čak kada bi bilo samo 1% vjerojatnije da je ona bolja od stare metode! No, naravno, u tim slučajevima, kada smo nivo strogosti spustili značajno iznad 5%, više ne možemo *tvrditi* da je nova metoda bolja, nego samo *vjerovati* u to!

— — —

Ako je neka razlika statistički značajna, onda to, naravno, znači samo to da se *aritmetičke sredine* populacije, iz kojih su uzorci, razlikuju, ali nikako ne i to da su *svi* individualni rezultati jedne varijable u populaciji ili u uzorku veći (ili manji) od *svih* individualnih razlika druge varijable. Prikazano grafički, statistički značajna razlika, uzmimo, izgleda kao crtež na slici 9.11, a *nipošto nije nužno da bude kao na slici 9.12.*

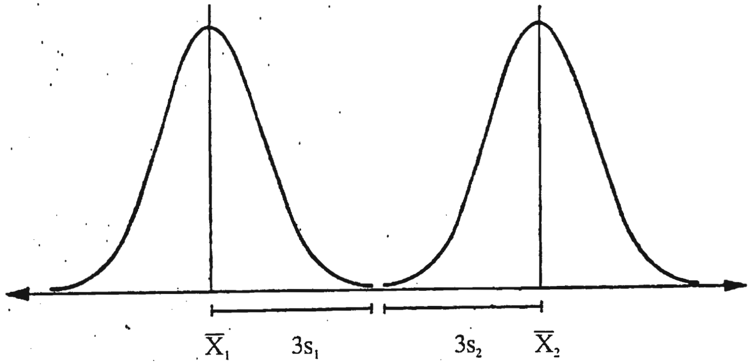


Slika 9.11. Ako je razlika između dviju aritmetičkih sredina statistički značajna, to znači da se *aritmetičke sredine* značajno razlikuju. To međutim ne znači da se svi individualni rezultati jedne populacije razlikuju od individualnih rezultata druge populacije, jer — kako vidimo — ima mnogo prekrivanja između njih. Zajednička površina obiju populacija je zatamnjena

Na primjer, ako nađemo između dvije grupe ljudi statistički značajnu razliku u visini od 3 cm, to znači da je u *prosjeku* jedna od tih grupa *viša* (no, naravno, ne znamo koliko, jer nađena 3 cm mogu biti slučajan rezultat variranja uzoraka), ali

je razumljivo da u prosječno višoj grupi ima pojedinaca koji su *niži* od pojedinaca iz druge grupe (vidi sliku 9.11!).

Kada bismo tražili da između dvije skupine mjerenja bude tako velika razlika da *gotovo nijedan* pojedinac iz više skupine ne bude niži ni od jednog pojedinca iz niže skupine, onda bi razlika među aritmetičkim sredinama populacija morala biti veća od zbroja između 3 standardne devijacije jedne i tri standardne devijacije druge populacije (vidi sliku 9.12).



Slika 9.12. U ovom slučaju gotovo svi rezultati jedne populacije razlikuju se od rezultata druge populacije. Razlika između obiju aritmetičkih sredina iznosi najmanje 3 standardne devijacije jedne raspodjele plus 3 standardne devijacije druge raspodjele

9.4. NUL-HIPOTEZA

Prije nego što pođemo još malo dalje u tumačenju testiranja značajnosti razlika, treba spomenuti jedan izraz koji se često u znanosti uopće, a osobito u statistici, čuje. To je izraz "nul- hipoteza". Mi ćemo ga poslije često u knjizi spominjati, pa je dobro da ga već sada rastumačimo.

U literaturi prividno postoje dva značenja nul-hipoteze (i zbog toga dolazi do određene konfuzije): prema jednom značenju (koje je mnogo više prihvaćeno) "nul-hipoteza" znači *"nema razlike"* među pojavama koje mjerimo. Na primjer, ako nas zanima razlikuju li se u prosječnom uspjehu u nekom testu muškarci od žena, onda "nul-hipoteza", koju u istraživanju postavljamo, glasi: Među njima nema statistički značajne razlike. Upotrebom t-testa (ili kojega drugog) mi *provjeravamo* postavljenu hipotezu: ako nam t- test pokaže da razlika među aritmetičkim sredinama nije statistički značajna, time smo *potvrdili* nul-hipotezu. Naprotiv, ako je razlika statistički značajna, *oborili* smo nul-hipotezu.

Prema drugom značenju, koje joj daje Fisher, nul-hipoteza je svaka hipoteza koju želimo provjeriti ("nul" je u smislu "nulificiranja", poništavanja). Iako se čini da su to dvije različite definicije, to ipak nije tako. Na primjer, možemo se pitati je li neka razlika između dvije aritmetičke sredine statistički značajna u tom smislu da je *statistički značajno veća od, uzmimo, 10.* Sada nam nul-hipoteza ne znači da "nema razlike", nego da "razlika iznosi 10", pa ako tu hipotezu ne oborimo, *također smo prihvatili nul-hipotezu* koju možemo formulirati ovako: nema statistički značajne razlike između naše dobivene razlike i zamišljene razlike od 10.

Uostalom, vrlo ćemo brzo imati upravo jedan takav primjer.

9.5. OPASNOST OD POGREŠNOG ZAKLJUČKA U VEZI S ODBACIVANJEM NUL-HIPOTEZE

U vezi s uobičajeno usvojenim razinama značajnosti od 1 i 5% katkada se u stručnoj literaturi i u komentarima nekih istraživanja može naići na jednu zabludu, koja u pojedinim slučajevima može biti i dosta opasna. Konkretno, radi se o procjeni postotka znanstvenih "otkrića" koja zapravo nisu otkrića.

Kod testiranja značajnosti razlike npr. između dvije aritmetičke sredine, ako nađemo da je razlika statistički značajna na razini manjoj od 5% ($P < 0,05$), mi (ispravno) rezoniramo ovako: ako među populacijama iz kojih smo uzeli uzorke zapravo nema razlike, onda se razlika, kakvu smo našim istraživanjem dobili, mogla slučajno dogoditi samo u manje od 5% slučajeva; stoga (uz rizik od 5%) odbacujemo nul-hipotezu, i zaključujemo da razlika među populacijama postoji. Ako se sada postavi npr. ovakvo pitanje: koliko ima pogrešnih zaključaka na 1 000 istraživanja, koja su sva provedena s nivoom značajnosti od 5% (tj. u koliko smo poprilici slučajeva zaključili da je razlika statistički značajna, makar ona to ustvari nije) veoma velik broj istraživača sklon je (neispravno) ustvrditi da ih ima oko 50 (jer to je 5% od 1 000). Takav zaključak može se npr. naći u nekim farmakološkim tekstovima o testiranju lijekova, gdje se npr. kaže poprilici ovako: budući da je efikasnost svih ispitivanih lijekova testirana na nivou značajnosti od 1 (ili 5) %, to možemo smatrati da oko 1 (ili 5) posto prihvaćenih lijekova nisu efikasni.

No nažalost taj je zaključak pogrešan! Iz logike nul-hipoteze može se izvesti samo ovakav zaključak: ako radimo 1 000 eksperimenata na populacijama koje se ne razlikuju, a pri tome koristimo nivo značajnosti od 5%, onda možemo očekivati da ćemo u oko 50 takvih eksperimenata dobiti rezultat, koji bi bio "statistički značajan", tj. koji bi pokazao da među populacijama postoji razlika. U tih 1 000 eksperimenata mi bismo dakle prihvatili nul-hipotezu u 950 tih eksperimenata a kod njih 50 odbacili bismo nul-hipotezu, i zaključili (uz rizik od 5%) da je razlika statistički značajna. No dakako, ni jedan od tih 50 prihvaćenih rezultata nije ispravan!

Iz toga proizlazi da odgovor na pitanje koliko znanstvenih otkrića su zaista otkrića, a koliko nisu, *ovisi o tome s kakvim populacijama radimo*: ako u 1 000 pokusa uspoređujemo po dvije populacije, koje se zaista razlikuju, onda su svi naši rezultati, kojima ustvrdimo da se populacije razlikuju, zaista točni. A ako radimo s populacijama, koje se ne razlikuju, onda *nijedan* naš rezultat, koji bi pokazao da se one razlikuju (a takvih će biti 1 – 5%, već prema tome na kojem nivou značajnosti radimo) nije točan.

Na tu pogrešku upozorio je svojedobno zagrebački liječnik B. Sorić, a detaljnije je opisana u radu Sorić i Petz, 1987. (vidi popis literature).

9.6. TESTIRANJE ZNAČAJNOSTI RAZLIKE IZMEĐU JEDNE ARITMETIČKE SREDINE I NEKE UNAPRIJED FIKSIRANE VRIJEDNOSTI

Katkada se događa da moramo ustanoviti razlikuje li se neka dobivena aritmetička sredina od jedne druge, *unaprijed "fiksirane"* vrijednosti, koja ne mora biti dobivena mjerenjem, pa nema niti svog varijabiliteta. To je čest slučaj onda ako neku vrijednost iz iskustva ili tradicije smatramo "normalnom", pa želimo provjeriti odstupaju li dobiveni rezultati statistički značajno od te vrijednosti.

Na primjer, pretpostavimo da smo u nekom mjerenju na 144 djece-ispitanika dobili ovu prosječnu težinu i standardnu devijaciju:

$$\overline{X} = 34,0\,\text{kg}$$

$$s = 4,8\,\text{kg}.$$

Nas u ovom slučaju može zanimati razlikuje li se dobivena aritmetička sredina od 34 kg statistički značajno od težine 32 kg, koju smatramo normalnom za djecu te dobi.

Budući da broj 32 u ovom slučaju ne predstavlja rezultat nekog sadašnjega konkretnog mjerenja, već jednu zamišljenu fiksnu vrijednost, ne možemo, naravno, testirati značajnost razlike standardnim postupkom, jer uz broj 32 nemamo standardnu devijaciju niti standardnu pogrešku.

Taj problem možemo vrlo lako riješiti uz pomoć već spomenutih *"granica pouzdanosti"* aritmetičke sredine. Standardna pogreška aritmetičke sredine 34 iznosi:

$$s_{\bar{x}} = \frac{s}{\sqrt{N}} = \frac{4,8}{\sqrt{144}} = \frac{4,8}{12} = 0,4.$$

Iz toga proizlazi da imamo 95% vjerojatnosti da naša dobivena aritmetička sredina ne odstupa od prave aritmetičke sredine više ili manje od $1,96 \cdot 0,4 = 0,78$, pa se, prema tome, "prava" aritmetička sredina nalazi vjerojatno u intervalu $34 \pm 0,78$, dakle između 33,22 kg i 34,78 kg. Budući da donja od tih vrijednosti (33,22) *ne "zahvaća"* vrijednost od 32, možemo, dakle, smatrati da se dobivena aritmetička sredina 34 statistički značajno (na nivou od 5%) razlikuje od 32.

Međutim, ako imamo izvršena dva mjerenja, i dobivene dvije aritmetičke sredine, ali zbog nekih razloga ne zanima nas je li *dobivena* razlika statistički značajna, već jedna *manja* razlika, postupak će opet biti nešto drukčiji (iako u načelu sličan).

Na primjer, ako neki proizvođač reklamira jedno prehrambeno sredstvo za bolji razvoj pilića, ali je to sredstvo toliko skupo da ekonomska računica pokaže kako bi se ono isplatilo jedino kad bi tim sredstvom hranjene kokoši bile *barem* 500 g teže od ostalih, problem se rješava ovako:

Kontrolna skupina kokoši, hranjena standardnom hranom	Eksperimentalna skupina, hranjena reklamiranim sredstvom
$N_1 = 400$	$N_2 = 361$
Prosj. težina $\overline{X}_1 = 2350\,\text{g}$	Prosj. težina $\overline{X}_2 = 3040\,\text{g}$
$s_1 = 200\,\text{g}$	$s_2 = 220\,\text{g}$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

$$= \sqrt{\frac{40\,000}{400} + \frac{48\,400}{361}} = \sqrt{234,1} = 15,3.$$

Budući da nas ne zanima razlikuje li se dobivena razlika od 690 g statistički značajno od *nule*, nego je li dobivena razlika statistički značajno veća od 500 g, postavit ćemo ovakav t-test:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - 500}{s_{\bar{x}_1-\bar{x}_2}} = \frac{690 - 500}{15,3} = 12,4.$$

Prema tome, praktički smo posve sigurni da razlika u hranjenju uzrokuje dobitak u težini koji je veći od 500 g.

— — —

U ovom opsežnom poglavlju vidjeli smo da značajnost neke razlike možemo provjeravati, bilo uz pomoć ***"granica pouzdanosti"*** (vidi slike 9.5. i 9.6) ili uz pomoć ***"t-testa"*** (vidi slike 9.7. do 9.10). Rekli smo već da su oba načina jednako ispravna, iako većina statističara koristi drugi način, tj. t-test. A i meni se čini da neke "finese" u zaključivanju možemo bolje shvatiti iz logike t-testa nego iz logike "granica pouzdanosti". No ima statističara (npr. Peatman), koji misle upravo obratno: Peatman smatra da logika "granica pouzdanosti" jasnije pokazuje da prihvaćanje "nul-hipoteze" (dakle stvaranje zaključka da neka razlika nije statistički značajna) *ne znači* da među populacijama *zaista* nema razlike (dakle da je razlika među njima nula!), već da bi možda mogla postojati i neka stvarna razlika, ali u njezino postojanje ne možemo biti dovoljno sigurni.

Prema tome, čitalac neka sâm odluči koju logiku rezoniranja može lakše usvojiti!

A ako mu *nije drago da usvoji nijednu od ovih logika rezoniranja,* onda neka svakako pročita ovo što *slijedi!*

Iz iskustva znam da će se među čitaocima naći i oni koji — zbog nedovoljne strpljivosti, premalo vremena, ili zbog drugih razloga — nisu uspjeli razumjeti logiku testiranja značajnosti razlika između aritmetičkih sredina.

Ako ste vi jedan od njih, *nemojte se demoralizirati!* Kao što je u prvom poglavlju već bilo rečeno, statistička metodologija dade se *potpuno uspješno* primjenjivati i onda ako zapravo ne razumijemo princip po kojem ona "funkcionira". Sjetite se, spomenuli smo da se automobil može dobro voziti a da se pri tome gotovo ništa ne zna o motoru i mehaničkim zakonima upravljanja vozilom.

To vrijedi i ovdje. Prema tome, ako niste svladali iznešeni slijed mišljenja (u čemu vjerojatno ima i moje krivnje, jer vam zacijelo nisam dovoljno jasno uspio protumačiti ono što sam želio), možete ipak uspješno raditi na području testiranja razlika među aritmetičkim sredinama.

Dovoljno je da zapamtite ovo:

Budući da mi gotovo nikada ne mjerimo populaciju, nego samo uzorke, to aritmetička sredina koju smo dobili prilikom mjerenja, sigurno nije *prava* aritmetička sredina, tj. aritmetička sredina *populacije*. Dakle, svaka aritmetička sredina uzorka vezana je uz neku pogrešku koju nazivamo "standardna pogreška aritmetičke sredine". Ta je pogreška to veća što je uzorak manji, i što je varijabilitet pojave koju mjerimo, veći. Ista logika vrijedi i za *razlike* između dvije aritmetičke sredine: dobivena razlika, s obzirom na to da je dobivena na uzorcima, nije *prava* razlika između obje populacije, nego i ona ima svoju pogrešku koja se naziva "standardna pogreška razlike između dvije aritmetičke sredine".

Da bi neka razlika bila statistički značajna (tj. da bismo koliko-toliko bili sigurni da ona nije *slučajna*) ona — kod velikih uzoraka — treba biti barem 1,96 (dakle oko dva) puta *veća od vlastite pogreške*. Izraz t, koji se računa prema principu

$$t = \frac{\text{razlika}}{\text{pogreška razlike}},$$

pokazuje nam koliko je puta neka razlika veća od svoje pogreške, pa, prema tome, ako je t veći od 1,96 (kod velikih uzoraka!), razliku možemo smatrati "statistički značajnom".

— — —

I, na kraju ove rasprave o smislu pojma "statistički značajne razlike" treba ponoviti nešto što je već bilo rečeno u prvom poglavlju: ako prilikom nekog mjerenja, izvršenog, pretpostavimo, na skupini muškaraca i skupini žena — nađemo u nekom svojstvu razliku u aritmetičkoj sredini, na primjer, u korist ženskih ispitanika, onda — ako smo pošteno i savjesno mjerili — *nema nikakve sumnje* da su ženski ispitanici, *koje smo izmjerili,* u prosjeku bolji od *izmjerenih* muških ispitanika. No istraživača to zapravo ne zanima, već ga zanima može li na temelju toga zaključiti da su u tom svojstvu žene *općenito* bolje od muškaraca.

Dakle, u svim tim slučajevima mi iz konkretnog zaključujemo na *općenito* — i stoga se cijeli ovaj dio statistike, koji iz *uzorka nastoji stvoriti zaključak o populaciji,* naziva "statistika zaključivanja" ili "inferencijalna statistika" (za razliku od "opisne" ili "deskriptivne statistike", koju zanimaju jedino karakteristike konkretnog uzorka).

9.7. RAZLIKA IZMEĐU ARITMETIČKIH SREDINA VELIKIH ZAVISNIH UZORAKA

Ako su dvije varijable, između kojih smo našli neku određenu razliku, u *korelaciji* (o izračunavanju korelacije vidi 13. pogl.), onda se formula za izračunavanje standardne pogreške razlike nešto mijenja i glasi:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 - 2r_{1.2}s_{\bar{x}_1}s_{\bar{x}_2}} \qquad (9.7)$$

pri čemu $r_{1.2}$ znači korelacija između obje varijable.

Kako se vidi, pod korijenom se od zbroja kvadrata obiju standardnih pogrešaka oduzima dvostruka korelacija između obje varijable, pomnožena s obje standardne pogreške. Na taj način standardna pogreška razlike postaje *manja* nego što bi bila da nismo upotrijebili korekturu zbog korelacije.

Primjer: Grupa od 64 bolesnika kojima snaga stiska šake na dinamometru na živi iznosi $\overline{X}_1 = 45,0$ cm, $s_1 = 6,0$ cm, podvrgnuta je specijalnom vježbanju, i nakon tjedan dana dobiveni su ovi rezultati snage stiska šake $\overline{X}_2 = 46,5$ cm, $s_2 = 5,0$ cm. Je li ta razlika statistički značajna, tj. je li snaga stiska šake tih bolesnika značajno porasla?

	Prvo mjerenje	Drugo mjerenje
N	64	64
$\overline{X}$	45,0	46,5
s	6,0	5,0
$s_{\overline{x}}$	0,75	0,63
$dif = (\overline{X}_2 - \overline{X}_1)$	$= +1,5$	

Korelacija r između prvog i drugog mjerenja $r = +0,60$.

Uvrstimo li dobivene vrijednosti u formulu (9.7), dobivamo:

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{0,75^2 + 0,63^2 - 2 \cdot 0,60 \cdot 0,75 \cdot 0,63} = 0,63;$$

$$t = \frac{\overline{X}_2 - \overline{X}_1}{s_{\overline{x}_2-\overline{x}_1}} = \frac{+1,5}{0,63} = +2,38;$$

$$P < 0,05.$$

Prema tome, razlika je statistički značajna.

N a p o m e n a 1. Da nismo uzeli u obzir korelaciju, dobili bismo

$$s_{\overline{x}_1-\overline{x}_2} = 0,98, \qquad t = 1,53,$$

iz čega bismo izveli *pogrešan* zaključak da razlika nije statistički značajna!

N a p o m e n a 2. Korelaciju između dva niza mjerenja očekujemo uvijek kada *ista* skupina ispitanika služi ujedno i kao kontrolna grupa. Stoga se upravo opisani postupak i često naziva "metoda jedne grupe".

9.8. t-RASPODJELA I TESTIRANJE RAZLIKA IZMEĐU ARITMETIČKIH SREDINA MALIH NEZAVISNIH UZORAKA

Već u početku treba reći da neki moderni statističari ne diferenciraju postupke kod velikih i kod malih uzoraka, jer i za jedne i za druge upotrebljavaju osnovne formule, koje uključuju u sebi sve potrebne računske operacije. Međutim, *pojednostavnimo* li te računske operacije kod velikih uzoraka, učinili smo doduše određeno "nasilje" nad rezultatima, ali je pogreška u računu tako mala da ona u krajnjem rezultatu nema praktički nikakvog efekta. Zato smo mi kod velikih uzoraka radili tim jednostavnijim metodama. Ali kod *malih* uzoraka (tj. kad je N manji od 30, odnosno — prema nekima — od 50) potrebno je u svakom slučaju upotrijebiti izvorne formule. Prema tome, pravilo, koje u statistici vrijedi u svim slučajevima, jest ovo: kod velikih uzoraka možemo upotrijebiti, bilo formule za velike ili formule za male uzorke; kod *malih* uzoraka *moramo* upotrijebiti formule za male uzorke.

Kako znamo, distribucija aritmetičkih sredina uzoraka oko "prave" aritmetičke sredine (tj. oko aritmetičke sredine populacije) jest *normalna* distribucija, a standardnu devijaciju takve distribucije nazvali smo standardnom pogreškom aritmetičke sredine ($s_{\overline{x}}$). (Ako su uzorci veći od $N = 30$, distribucija aritmetičkih sredina je normalna čak i onda ako osnovna populacija *nije* normalno raspodijeljena!)

Standardna devijacija aritmetičkih sredina uzoraka oko prave aritmetičke sredine populacije izražena je formulom $\sigma/\sqrt{N}$, ali budući da praktički nikada ne znamo standardnu devijaciju *populacije* (σ), mi u tu formulu redovito stavljamo u brojnik standardnu devijaciju našeg *uzorka* (s), i tako — kao što smo rekli — dobivamo *procjenu* standardne pogreške aritmetičke sredine. Ako je uzorak relativno velik, time smo učinili posve malu pogrešku, koju možemo zanemariti. No ako je uzorak malen, pogreška postaje ozbiljnija, i to sve teža, što je uzorak manji.

Mi ne možemo, naravno, radi ustanovljavanja kako se zapravo distribuiraju aritmetičke sredine malih uzoraka oko prave aritmetičke sredine (ili kako se distribuiraju razlike među aritmetičkim sredinama malih uzoraka oko prave razlike) izvoditi tisuće pokusa, kojima bismo eventualno tu distribuciju mogli ustanoviti, nego se moramo služiti podacima koje jedino *imamo*, tj. moramo se služiti računom u kojem koristimo — kao što je rečeno — standardnu pogrešku izračunatu na *uzorku*.

Na sreću, engleski službenik i matematičar, koji je radio u jednoj pivovari, po imenu Gosset, suočen u svom svakodnevnom poslu neprestano s malim uzorcima i testiranjem razlika među njima, znajući za nepreciznost računskog postupka u tim slučajevima, razradio je matematičkim postupkom distribucije t- vrijednosti, koje se dobivaju kada se radi o malim uzorcima. Budući da se on potpisivao pseudonimom "Student", ta je distribucija — uz naziv "t-distribucija" — dobila još i ime "Studentova distribucija".

Evo u čemu je osnovna razlika između logike rezoniranja i zaključivanja kod velikih i kod malih uzoraka:

Kada kod velikih uzoraka izračunamo, na primjer, standardnu pogrešku razlike između dvije aritmetičke sredine, onda znamo da smo time — *unatoč* tome što smo u računu koristili standardnu devijaciju *uzorka* a ne *populacije* — dobili uglavnom točnu procjenu standardne devijacije razlika mnoštva uzoraka oko "prave" razlike među populacijama. Drugim riječima, kada bismo pokus na ovako velikim uzorcima ponavljali nekoliko tisuća puta, onda bi nam se dobivene razlike među aritmetičkim sredinama, *a također i izračunate t-vrijednosti,* distribuirale po *normalnoj* raspodjeli, kojoj standardna devijacija približno odgovara izračunatoj standardnoj pogrešci razlike. (Da smo u formulu mogli uključiti standardnu devijaciju *populacije,* onda bi standardna devijacija tih t-vrijednosti *točno* iznosila koliko i izračunata standardna pogreška razlike, koja bi, naravno, u svakom od tih nekoliko tisuća pokusa morala biti istovjetna.)

Kod *malih* uzoraka, naprotiv, *iako se i kod njih razlike između aritmetičkih sredina uzoraka distribuiraju po normalnoj raspodjeli oko "prave" razlike, izračunati* t-odnosi *ne* distribuiraju se po normalnoj, već po t-raspodjeli, koja je to šira što je uzorak manji.

Da bismo to, što je rečeno, još jasnije protumačili (jer — istini za volju — treba priznati da se ipak radi o dosta zamršenoj temi!), pokušajmo to protumačiti na osnovi konkretnog pokusa.

Zamislimo ponovno (jer o tom je pokusu već bilo riječi) da imamo jednu veliku populaciju, čije smo rezultate ubilježili na žetone koji se nalaze u jednoj vreći. Ako lijevom rukom vadimo jedan uzorak, a desnom drugi, sve "razlike" koje ćemo među aritmetičkim sredinama uzoraka nalaziti, samo su *slučajne razlike, uzrokovane slučajnim varijacijama uzoraka.*

Zamislimo sada da s ovom velikom populacijom učinimo ova dva (vrlo dugotrajna) pokusa:

1. Izvučemo nekoliko desetaka tisuća parova *velikih* uzoraka (npr. svakom rukom od po $N = 200$), i kod svakog para ustanovimo razliku između aritmetičkih sredina, i za svaki par izračunamo t-odnos, tj. razliku između aritmetičkih sredina podijelimo s njezinom standardnom pogreškom, koju naravno svaki put izračunamo iz standardnih devijacija oba uzorka.

2. Izvučemo nekoliko desetaka tisuća parova *malih* uzoraka (na primjer, svakom rukom po 5), i za svaki par izračunamo *iste podatke* kao i kod velikih uzoraka, dakle razlike između aritmetičkih sredina i t-odnose.

Sada raspolažemo s nekoliko desetaka tisuća t-odnosa iz eksperimenta s velikim uzorcima i nekoliko desetaka tisuća t- odnosa iz eksperimenta s malim uzorcima.

Budući da ih ima tako mnogo, možemo dosta precizno izraditi *distribuciju* tih t-odnosa. Ako nakon toga *analiziramo* dvije dobivene t-distribucije, brzo ćemo ustanoviti da je kod velikih uzoraka ta distribucija praktički sasvim *normalna,* tj. t-odnosi se distribuiraju uglavnom od $- 3\ t$ do $+ 3\ t$ (kao što se i normalna raspodjela proteže od $- 3$ do $+ 3$ standardne devijacije). (Treba se sjetiti da t-vrijednosti po svom smislu odgovaraju z- vrijednostima: to su *razlike,* podijeljene standardnom devijacijom tih razlika.)

Takvu smo normalnu raspodjelu dobili dakle *usprkos činjenici da smo pri računanju t-odnosa uvijek koristili standardnu devijaciju uzorka umjesto standardne devijacije populacije.*

Naprotiv, t- odnosi *malih* uzoraka distribuirat će se *mnogo šire,* tj. t-odnosi kretat će se od $-$ 3,5 t preko nule do + 3,5 t. Makar su se dakle *stvarne razlike* među uzorcima u tom pokusu distribuirale oko prave razlike (koja je 0) po *normalnoj* raspodjeli, t-vrijednosti se distribuiraju *šire* jer smo u formuli koristili standardnu devijaciju uzorka, a ne standardnu devijaciju populacije.

Iz toga proizlazi da, ako se u toj distribuciji jedan određeni t nalazi, uzmimo, na udaljenosti od 2 t (što je kod velikih uzoraka bilo dovoljno da znamo da *većih* od njega ima još samo oko 5%), u ovoj t-distribuciji naći ćemo da ih je većih od 2 još možda 7 — 8%.

Drugim riječima, *kod malih uzoraka normalna raspodjela više ne pokazuje sasvim točno vjerojatnost koliko neka izračunata vrijednost ($\overline{X}$) odstupa od prave vrijednosti (μ).*

t-raspodjela je, doduše, slična normalnoj, tj. ona je također simetrična i ima "zvonast" oblik, ali je ona pri "dnu" šira od normalne raspodjele. Drugim riječima, kod normalne raspodjele mi točno znamo da se u intervalu $\overline{X} \pm 1\,s$ nalazi približno 68% rezultata, u intervalu $\overline{X} \pm 2\,s$ oko 95% rezultata, itd., pa, prema tome, i znamo

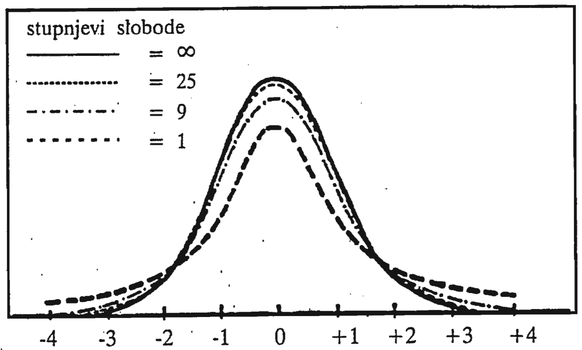
da neka razlika između aritmetičkih sredina mora biti oko dva puta (točnije: 1,96 puta) veća od svoje pogreške da je možemo smatrati značajnom na razini od 5%.

Kod t-raspodjele imamo, međutim, to šire "krajeve" raspodjele što je uzorak manji, i zato ako imamo uzorak od 10 rezultata, pa računamo standardnu pogrešku aritmetičke sredine toga uzorka ($s_{\overline{x}}$), onda se 95% aritmetičkih sredina uzoraka te veličine *neće* nalaziti u intervalu od $\pm 1,96\, s_{\overline{x}}$ oko prave aritmetičke sredine, već u *širem* intervalu, tj. u intervalu $\mu \pm 2,26\, s_{\overline{x}}$. Kod uzorka veličine $N = 4$, 95% aritmetičkih sredina uzoraka te veličine nalazit će se u intervalu $\mu \pm 3,18\, s_{\overline{x}}$, itd.

Kako se vidi, ovdje se *kriterij mijenja* već prema broju rezultata, i zato pri izračunavanju t-vrijednosti moramo uvijek uzeti u obzir broj rezultata, odnosno točnije rečeno t- vrijednosti ovisne su o tzv. *"stupnjevima slobode".* Stupnjevi slobode su na određeni način *korigirani* broj rezultata. U velikom broju računa, pa tako i ovdje, stupnjevi slobode su broj rezultata smanjen za 1, dakle $N - 1$.

Na slici 9.13. pokazano je kako se t-distribucija širi smanjivanjem broja rezultata (tj. "stupnjeva slobode") u uzorku.

Dok smo kod velikih uzoraka mogli i "napamet" znati da je razlika između dvije aritmetičke sredine statistički značajna (na razini značajnosti od 5%) ako t iznosi barem 1,96, kod malih uzoraka moramo se bezuvjetno koristiti *t-tablicom* da bismo iz nje mogli očitati graničnu vrijednost t, tj. koliko puta — uz određenu veličinu uzorka — mora razlika biti veća od svoje pogreške da bismo je mogli smatrati statistički značajnom.



Slika 9.13. t-raspodjela je to šira, što je manji uzorak. Veličinu uzorka, smanjenu za 1, kod t-raspodjele nazivamo "stupnjevima slobode"

Tablica B u Dodatku je t-tablica. U tablici, na primjer, piše uz 18 stupnjeva slobode na nivou značajnosti od 1% ($P = 0,01$) broj 2,88. To je "granična t-vrijednost", i znači ovo: kada između dvije populacije ne bi zapravo postojala nikakva razlika, a mi bismo iz tih populacija uzimali uzorke, pretpostavimo, veličine $N_1 = 10$ i $N_2 = 10$ ($10 - 1$ plus $10 - 1 = 18$) i *izračunavali* svaki put t- vrijednost, onda bismo t, veličine 2,88 i veće mogli *slučajno* dobiti samo u jednom od 100 izračunavanja (1%).

Budući da u t-raspodjeli postoji mnogo nesigurnosti među konzumentima statističke metodologije, ukratko ćemo rezimirati što t-raspodjela *nije*, a što ona *jest*.

1. t-raspodjela *nije* distribucija rezultata u malom uzorku (što neki početnici misle).
2. t-raspodjela također *nije* distribucija aritmetičkih sredina ili distribucija razlika među aritmetičkim sredinama *malih* uzoraka oko prave aritmetičke sredine ili prave razlike (što misle katkad čak i neki napredniji u statistici), jer ta distribucija je *normalna*.
3. t-raspodjela *jest distribucija t-izraza* (izraza koji znači odnos između neke razlike i njezine pogreške) kakva se dobiva kada se računski određuje standardna pogreška aritmetičke sredine, odnosno standardna pogreška razlike između aritmetičkih sredina, a pri tome se u računu koristi standardna devijacija uzorka. (Kada bismo u računu mogli koristiti standardnu devijaciju *populacije*, izračunate t-vrijednosti tvorile bi jednako normalnu raspodjelu kao što ju tvore i stvarne razlike među aritmetičkim sredinama uzoraka.)

– – –

Značajnost razlike između aritmetičkih sredina malih uzoraka računa se prema nešto drukčijem postupku. Princip je jednak kao i kod velikih uzoraka, ali računska se operacija razlikuje utoliko što — pod pretpostavkom da oba uzorka potječu iz iste populacije — izračunavamo *zajedničku* standardnu devijaciju za oba uzorka.

Međutim, zajedničku standardnu devijaciju smijemo izračunati samo onda ako se obje standardne devijacije zaista *značajno ne razlikuju*, pa to treba najprije provjeriti.

Značajnost razlike među standardnim devijacijama malih uzoraka izračunava se pomoću tzv. F-testa, koji stavlja u omjer veću varijancu prema manjoj varijanci (varijanca $= s^2$). Prema tome, F- test glasi:

$$F = \frac{\text{veća } s^2}{\text{manja } s^2}. \tag{9.8}$$

N a p o m e n a: U 20. poglavlju biti će više govora o F-testu.

Prvi primjer: Mjereći na gastroknemijusu dvije žabe vrijeme reakcije u tisućinkama sekunde, dobiveni su ovi rezultati:

1. žaba	2. žaba
160, 160, 140, 190	117, 145, 147, 120, 150, 120
$N = 4$	$N = 6$
$\overline{X}_1 = 162,5$	$\overline{X}_2 = 133,2$

$$\overline{X}_1 - \overline{X}_2 = 29,3.$$

Pitamo se je li ta razlika statistički značajna.

Prije nego što izračunamo zajedničku standardnu devijaciju, treba ustanoviti da li se standardne devijacije značajno ne razlikuju:

Prvi uzorak	Drugi uzorak
160	117
160	145
140	147
190	120
$\Sigma = 650$	150
$\overline{X}_1 = 162,5$	120
	$\Sigma = 799$
	$\overline{X}_2 = 133,2$

$X - \overline{X}$	$(X - \overline{X})^2$	$X - \overline{X}$	$(X - \overline{X})^2$
−2,5	6,25	−16,2	262,44
−2,5	6,25	11,8	139,24
−22,5	506,25	13,8	190,44
27,5	756,25	−13,2	174,24
$\Sigma(X - \overline{X})^2 =$	1 275,00	16,8	282,24
		−13,2	174,24
		$\Sigma(X - \overline{X})^2 =$	1 222,84

$$s_1 = \sqrt{\frac{1275}{3}} = \sqrt{425}$$

$$s_2 = \sqrt{\frac{1222,84}{5}} = \sqrt{244,57}$$

Sada možemo izračunati F-test:

$$F = \frac{\text{veća varijanca}}{\text{manja varijanca}} = \frac{425}{244,57} = 1,74.$$

Tablica C u Dodatku daje nam granične vrijednosti F za pojedine stupnjeve slobode veće i manje varijance. Tablica C je za razinu značajnosti od 2,5%, ali se F-tablice uvijek prikazuju samo za jednu stranu krivulje (o jednosmjernom testiranju vidi 10. poglavlje), pa, želimo li znati je li *razlika* među varijancama značajna (a ne *smjer* razlike!), onda je ta tablica za nas na razini značajnosti od 5%.

Stupnjeve slobode za veću varijancu očitavat ćemo s odgovarajućeg stupca, a za manju varijancu s odgovarajućeg reda. Kako se iz tablice C vidi, granična vrijednost F za 3 stupnja slobode (veća varijanca) i 5 stupnjeva slobode (manja varijanca) iznosi 7,76. Budući da je naš F manji, možemo smatrati da se obje varijance *ne razlikuju* značajno, te smijemo pristupiti izračunavanju zajedničke standardne devijacije, koja se izračunava prema formuli:

$$\text{Zajednička } s = \sqrt{\frac{\Sigma(X - \overline{X}_1)^2 + \Sigma(X - \overline{X}_2)^2}{(N_1 - 1) + (N_2 - 1)}}. \tag{9.9}$$

U našem primjeru dobivamo:

$$\text{Zajednička } s = \sqrt{\frac{1275 + 1222,84}{3+5}} = \sqrt{312,23} = 17,67$$

Dalje postupamo jednako kao i prije, tj. standardnu pogrešku razlike računamo prema formuli (9.5):

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{312,23}{4} + \frac{312,23}{6}} = \sqrt{78,04 + 52,04} = \sqrt{130,1} = 11,41$$

Kada znamo zajedničku s, možemo $s_{\bar{x}_1-\bar{x}_2}$ jednostavnije izračunati i prema formuli:

$$s_{\bar{x}_1-\bar{x}_2} = \text{Zajednička } s\sqrt{\frac{N_1+N_2}{N_1N_2}} \tag{9.10}$$

U našem primjeru dobivamo:

$$s_{\bar{x}_1-\bar{x}_2} = 17,67\sqrt{\frac{4+6}{24}} = 11,41$$

Ako u našem primjeru izračunamo t, dobivamo:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\bar{x}_1-\bar{x}_2}} = \frac{29,3}{11,41} = 2,57.$$

Broj stupnjeva slobode u ovom se slučaju izračunava prema formuli $(N_1 - 1) + (N_2 - 1)$, pa, prema tome, za naš primjer to iznosi $3 + 5 = 8$. Ako smo unaprijed odlučili da značajnost razlike računamo na razini značajnosti od 5%, onda ćemo na tablici graničnih t- vrijednosti očitati uz odgovarajući broj stupnjeva slobode kolika je granična vrijednost t. Kako se vidi iz tablice B, granična vrijednost t na razini značajnosti od 5%, a uz 8 stupnjeva slobode iznosi 2,31 (što konkretno znači da u tom slučaju razlika mora biti barem 2,31 puta veća od svoje pogreške), a kako je naš $t = 2,57$, to je naša razlika statistički značajna na razini značajnosti od 5%, i konačan rezultat treba pisati ovako:

$$\overline{X}_1 - \overline{X}_2 = 29,3 \qquad s_{\bar{x}_1-\bar{x}_2} = 11,41$$

$$t = 2,57 \qquad P < 0,05$$

U različitim statistikama možemo naići i na različite formule izračunavanja značajnosti razlika između aritmetičkih sredina malih uzoraka. Sve su one *matematički* potpuno ekvivalentne i razlikuju se samo po tome što su u pojedinim situacijama neke od njih praktičnije i skraćuju pojedine račune. Tako, primjerice, za standardnu pogrešku razlike između aritmetičkih sredina možemo upotrijebiti i ovu formulu:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(N_1+N_2)[\Sigma(X-\overline{X}_1)^2 + \Sigma(X-\overline{X}_2)^2]}{N_1N_2(N_1+N_2-2)}}. \tag{9.11}$$

U našem slučaju dobivamo:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{10(1275 + 1222,84)}{24\cdot 8}} = 11,4$$

Ili, izravno možemo t izračunati pomoću formule:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\text{Zajed. } s}\cdot\sqrt{\frac{N_1N_2}{N_1+N_2}}, \tag{9.12}$$

ili

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\text{Zajed. } s\sqrt{1/N_1 + 1/N_2}}. \tag{9.13}$$

U našem slučaju dobivamo:

$$t = \frac{29,3}{17,67}\sqrt{\frac{4\cdot 6}{4+6}} = 2,57.$$

Ako ne posjedujemo individualne rezultate, nego samo aritmetičke sredine i standardne devijacije, možemo, ako smo prethodno testirali F i dokazali da se varijance značajno ne razlikuju, također izračunati zajedničku standardnu devijaciju, i to prema formuli:

$$\text{Zajednička } s \doteq \sqrt{\frac{s_1^2(N_1-1) + s_2^2(N_2-1)}{(N_1-1)+(N_2-1)}}. \tag{9.14}$$

Ova je formula dobivena ovako:

$$s = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N-1}}$$

Dakle, $s^2 = \dfrac{\Sigma(X-\overline{X})^2}{N-1}$. Prema tome, $\Sigma(X-\overline{X})^2 = s^2(N-1)$.

N a p o m e n a: Ako je standardna devijacija računata prema formuli:

$$s = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N}},$$

onda naravno u brojniku formule (9.14) dolazi $s_1^2N_1 + s_2^2N_2$.

Drugi primjer. Na grupu od 18 bolesnika primijenjen je penicilin kod gnojenja te je promatrano kako je dugo gnojenje još trajalo nakon prve primjene penicilina. Kod druge grupe od 16 bolesnika nije primijenjen penicilin, nego neko drugo, klasično sredstvo. Dobiveni su ovi rezultati:

Grupa liječena penicilinom	Grupa liječena drugim sredstvom
$(N_1 = 18)$	$(N_2 = 16)$
Prosječno trajanje gnojenja u danima:	Prosječno trajanje gnojenja u danima: $\overline{X}_2 = 13,8$
$\overline{X}_1 = 6,3$	$\overline{X}_2 = 13,8$
$s_1 = 3,5$	$s_2 = 12,0$

$$\overline{X}_2 - \overline{X}_1 = 13,8 - 6,3 = 7,5$$

Najprije ćemo F-testom provjeriti razlike među varijancama:

$$F = \frac{12^2}{3,5^2} = 11,76$$

Iz F-tablice (tablica C) možemo ustanoviti da za 15 i 17 stupnjeva slobode granična vrijednost F iznosi 2,72. Budući da je naš F *znatno veći*, znači da obje standardne devijacije ne spadaju u istu populaciju, pa stoga ne smijemo izračunavati zajedničku standardnu devijaciju.

Ako se to dogodi, *ne može se na standardni način t-testom testirati značajnost razlika između dvije aritmetičke sredine*, nego treba upotrijebiti neke druge, aproksimativne metode, od kojih ćemo izložiti metodu Cochrana i Cox-ove (1950).

Prema Cochranu i Cox-ovoj najprije treba izračunati standardnu pogrešku razlike između dvije aritmetičke sredine, i to prema nama već poznatoj formuli (9.5):

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}.$$

U našem slučaju dobivamo:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{3,5^2}{18} + \frac{12^2}{16}} = \sqrt{9,68} = 3,11,$$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\bar{x}_1-\bar{x}_2}} = \frac{7,5}{3,11} = 2,41.$$

Jedan uzorak vezan je na $N_1 - 1$ stupnjeva slobode, a drugi na $N_2 - 1$ stupnjeva slobode. Iz t-tablice treba pronaći granične vrijednosti za oba stupnja slobode. U našem slučaju stupnjevi slobode su $18 - 1 = 17$ i $16 - 1 = 15$. Granična vrijednost t uz 17 stupnjeva slobode iznosi (na razini značajnosti od 5%) 2,11, a uz 15 stupnjeva slobode 2,13. Nazovimo te granične vrijednosti t_1 i t_2. Prema tome, $t_1 = 2,11$, a $t_2 = 2,13$. Aproksimativna granična vrijednost t na razini značajnosti od 5%, koja se traži za ovaj slučaj, izračunava se prema formuli:

$$t = \frac{s_{\bar{x}_1}^2 t_1 + s_{\bar{x}_2}^2 t_2}{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}. \qquad (9.15)$$

Uvrstimo li poznate vrijednosti u formulu (9.15), dobivamo:

$$t = \frac{0,68 \cdot 2,11 + 9 \cdot 2,13}{0,68 + 9} = 2,13.$$

Ako je prije izračunati t *jednak* ili *veći* od ovoga, možemo smatrati da je razlika statistički značajna.

Budući da je naš dobiveni t *veći* od ovako izračunate granične vrijednosti ($2,41 > 2,13$), možemo, dakle, zaključiti da se te dvije aritmetičke sredine značajno razlikuju na razini značajnosti od 5%.

U ovakvim slučajevima, kada se varijance razlikuju, može se uspješno upotrijebiti neparametrijski "test sume rangova" (vidi poglavlje 21, 2.3).

Razumljivo je da t-raspodjela vrijedi i za "granice pouzdanosti" jednog uzorka. Dobijemo li, na primjer, $\overline{X} = 20$, $s = 3$ na uzorku veličine $N = 9$, standardna

pogreška aritmetičke sredine ($s_{\bar{x}}$) iznosit će $s/\sqrt{N} = 3/3 = 1$. 95%-tne granice pouzdanosti te aritmetičke sredine neće se naravno kretati u intervalu $\overline{X} \pm 1,96\, s_{\bar{x}}$ (kao što smo to računali kod velikih uzoraka), već u intervalu koji možemo očitati s t-tablice uz $9 - 1 = 8$ stupnjeva slobode. Iz tablice vidimo da imamo 95% vjerojatnosti da naša aritmetička sredina ne odstupa od prave aritmetičke sredine više od $2,31\, s_{\bar{x}}$, pa se vjerojatno nalazi u intervalu $17,69 - 22,31$.

Kao što je već spomenuto, testiranje razlika među aritmetičkim sredinama u novije je vrijeme postalo kompliciranije nego što je svojedobno bilo, jer su suvremeni statističari, u želji za što većom preciznošću, postavili strože zahtjeve za male uzorke (gdje je mogućnost pogrešnog zaključka veća), pa s tim u vezi i kompleksnije račune, koje smo upravo iznijeli.

No, na sreću, ima i nekih promjena koje predstavljaju *pojednostavljenje* postupka. Konkretno, u većini se udžbenika može naći upozorenje da za upotrebu t-testa moraju biti ispunjeni ovi osnovni uvjeti:

1. uzorci treba da budu slučajni uzorci iz *normalnih* populacija;
2. varijance obiju populacija trebaju biti podjednake.

Zbog toga smo i mi u primjerima koje smo obrađivali, testirali postoji li razlika među varijancama i savjetovali što da se radi ako razlika zaista postoji.

Međutim, statističar Boneau je uz pomoć kompjutorski proizvedenih brojnih uzoraka dokazao da će t-test (i njegovi srodnici) u mnogim slučajevima dati relativno točne rezultate *unatoč tome što je prekršen uvjet homogenosti varijance* (tj. unatoč tome što varijance oba uzorka nisu jednake) *i unatoč tome što uzorci nisu uzeti iz normalno distribuiranih populacija.*

To su ove situacije:

1. *Ako su oba uzorka jednaka, ili barem vrlo slična, po veličini.*
2. *Ako matične populacije imaju jednaku ili sličnu formu* (npr. ako su obje na jednak način asimetrične).

Iz toga proizlazi i vrlo važno praktično pravilo za onoga tko radi s t-testom: *eksperiment treba po mogućnosti tako planirati da su oba uzorka jednako velika.* Ako je ispunjen taj uvjet, nije ni potrebno testirati F-testom da li im se varijance statistički značajno razlikuju ili ne, jer će — ako se one razlikuju — pogreška u računu biti neznatna!

Ako *ne uspijemo* postići jednaku (ili sličnu) veličinu oba uzorka (a to se, na žalost, u praksi, kod tzv. "prigodnih" uzoraka, često događa, npr. kliničarima i sličnim strukama), onda se *moramo* pridržavati pravila navedenih na posljednjih nekoliko stranica, tj. moramo prethodno testirati da li se varijance razlikuju; ako se razlikuju, onda ne preostaje drugo nego upotreba *aproksimativne* metode Cochrana i Cox-ove.

9.9. RAZLIKA IZMEĐU ARITMETIČKIH SREDINA MALIH ZAVISNIH UZORAKA ("METODA DIFERENCIJE")

Ako radimo s malim uzorcima koji su u međusobnoj *korelaciji*, služimo se tzv. "metodom diferencije", koja je vrlo jednostavna i praktična zbog toga što isključuje potrebu računanja korelacije između obje varijable, tj. u rezultat koji dobivamo, već je samim postupkom uključena korelacija.

Princip "metode diferencije" sastoji se u tome da se individualne *razlike* parova rezultata uzmu kao uzorak koji obrađujemo kao i svaki drugi uzorak, tj. nađemo njegovu aritmetičku sredinu, standardnu devijaciju i standardnu pogrešku. Budući da je to uzorak *razlika*, standardna pogreška aritmetičke sredine toga uzorka znači, dakle, standardnu pogrešku razlike između obje aritmetičke sredine.

Primjer. Izmjeren je bazalni metabolizam osamaestorici muških ispitanika u normalnom bazalnom stanju (BM_1) i 90 minuta nakon uzimanja 15 mg fenamina (BM_2). Vrijednosti BM izražene su u kilo kalorijama na sat. Dobiveno je prosječno povišenje bazalnog metabolizma od 2 kalorije. Je li ta razlika statistički značajna, tj. je li fenamin značajno promijenio BM?

Rezultati prvog i drugog mjerenja BM za svakog ispitanika, kao i postupak izračunavanja pomoću "metode diferencije", prikazani su u tablici 9.9.

TABLICA 9.9.
"METODA DIFERENCIJE"

Ispitanici	BM_1	BM_2	Diferencija $BM_2 - BM_1$	d	d^2
1	70	72	+ 2	0	0
2	86	99	+13	+11	121
3	81	84	+ 3	+ 1	1
4	78	79	+ 1	−1	1
5	70	71	+ 1	−1	1
6	63	68	+ 5	+ 3	9
7	78	82	+ 4	+ 2	4
8	82	98	+16	+14	196
9	66	68	+ 2	0	0
10	72	76	+ 4	+ 2	4
11	81	98	+17	+15	225
12	108	102	−6	−8	64
13	80	86	+ 6	+ 4	16
14	82	78	−4	−6	36
15	74	72	−2	−4	16
16	76	80	+ 4	+ 2	4
17	78	67	−11	−13	169
18	99	80	−19	−21	441
			Σ dif = +36		$\Sigma d^2 = 1308$

$$\overline{X}_{\text{dif}} = +36/18 = +2,0$$

$$s = \sqrt{\frac{\Sigma d^2}{N-1}} = \sqrt{\frac{1308}{17}} = \sqrt{76,94} = 8,77$$

$$s_{\overline{x}} = \frac{s}{\sqrt{N}} = \frac{8,77}{4,24} = 2,07$$

$$t = \frac{\text{razlika}}{\text{stand. pogr. razlike}} = \frac{2,0}{2,07} = 0,97.$$

Razlika, dakle, nije statistički značajna jer je t manji od 2,11 (što je granična vrijednost t za 17 stupnjeva slobode).

Kako se vidi, najprije treba izračunati aritmetičku sredinu svih diferencija, a nakon toga tražiti razlike između svake pojedine diferencije i aritmetičke sredine diferencije (pritom treba naročito paziti na predznake: na primjer, kod ispitanika br. 14 razlika između diferencije i aritmetičke sredine diferencija iznosi 6, jer je razlika između -4 i $+2 = 6$). Kad se te razlike kvadriraju (d^2), može se izračunati s.

Budući da prvi i drugi uzorak predstavljaju *isti* ispitanici, to je $N = 18$ (a ne 36!), prema tome, je broj stupnjeva slobode $(N - 1) = 17$. Iz *t-tablice* vidi se da je uz 17 stupnjeva slobode a uz razinu značajnosti od 5%, granična vrijednost $t = 2,11$. Budući da je dobiveni t *manji*, možemo zaključiti da razlika nije statistički značajna, pa se, prema tome, ne može iz ovog eksperimenta ustvrditi da fenamin djeluje na bazalni metabolizam.

Izračunavanje pomoću "metode diferencije" može se nešto skratiti ako t izračunamo izravno iz dobivenih vrijednosti u tablici diferencija, i to pomoću formule:

$$t = \frac{\overline{X}_{\text{dif}}}{\sqrt{\frac{\Sigma d^2}{N(N-1)}}}. \tag{9.16}$$

U našem primjeru dobivamo:

$$t = \frac{2}{\sqrt{\frac{1308}{306}}} = 0,97.$$

Još jednostavnije to možemo učiniti poslužimo li se formulom 5.4, koja koristi bruto-rezultate za izračunavanje standardne devijacije. U tom slučaju formula glasi:

$$t = \frac{\overline{X}_{\text{dif}}}{\sqrt{\frac{\Sigma_{\text{dif}^2} - \frac{(\Sigma_{\text{dif}})^2}{N}}{N(N-1)}}}. \tag{9.17}$$

U našem primjeru dobivamo:

$$t = \frac{2}{\sqrt{\frac{1380 - \frac{1296}{18}}{18 \cdot 17}}} = 0,97.$$

Tom tehnikom izbjegli smo izračunavanje d (razlika između svake diferencije i prosječne diferencije).

Svojedobno je psiholog Sandler predložio jedan *još kraći* postupak, koji je nazvao A-test, izveden iz t-odnosa, po kojem je dovoljno izračunati jedino razlike između prvog i drugog mjerenja i kvadrate tih razlika, te u formulu uvrstiti ove vrijednosti.

$$A = \frac{\Sigma_{\text{dif}^2}}{(\Sigma_{\text{dif}})^2}. \tag{9.18}$$

Dobiveni izraz A usporedi se s vrijednostima u tablici 9.10. pod određenom vrijednosti P i uz odgovarajući broj stupnjeva slobode. Ako je dobivena A vrijednost *manja* od granične vrijednosti (dakle obratno nego kod t!), razlika je statistički značajna.

TABLICA 9.10.
GRANIČNE A-VRIJEDNOSTI

Stup. slobode $(N-1)$	P 0,05	P 0,01	Stup. slobode $(N-1)$	P 0,05	P 0,01
1	0,5031	0,50012	18	0,267	0,167
2	0,369	0,340	19	0,267	0,166
3	0,324	0,272	20	0,266	0,165
4	0,304	0,238	21	0,266	0,165
5	0,293	0,218	22	0,266	0,164
6	0,286	0,205	23	0,266	0,163
7	0,281	0,196	24	0,265	0,163
8	0,278	0,190	25	0,265	0,162
9	0,276	0,185	26	0,265	0,162
10	0,274	0,181	27	0,265	0,161
11	0,273	0,178	28	0,265	0,161
12	0,271	0,176	29	0,264	0,161
13	0,270	0,174	30	0,264	0,160
14	0,270	0,172	40	0,263	0,158
15	0,269	0,170	60	0,262	0,155
16	0,268	0,169	120	0,261	0,153
17	0,268	0,168	∞	0,260	0,151

Ako račun izvedemo na našem primjeru, dobivamo da je suma kvadriranih diferencija $(\Sigma_{\text{dif}^2}) = 2^2 + 13^2 + 3^2 ... + 19^2 = 1380$, a $(\Sigma_{\text{dif}})^2 = 36^2 = 1296$.

$$A = \frac{1380}{1296} = 1,065.$$

Budući da je naš A veći od granične vrijednosti A na razini značajnosti od 0,05, i uz 17 stupnjeva slobode (granična vrijednost A = 0,268), zaključujemo da razlika nije statistički značajna.

9.10. KOMBINACIJA DVAJU TESTOVA ZNAČAJNOSTI

U istraživačkom radu kadikad se događa da se neki problem ispituje *nekoliko puta*. Tako, na primjer, neki nastavnik može svake godine primijeniti neki test znanja na svoje studente, i svaki put može naći da su ženski ispitanici nešto bolji u prosječnom rezultatu od muških, ali ni jednom prilikom ta razlika *nije statistički značajna*. No, već sama činjenica da je nađena razlika uvijek u istom smjeru, sugerira nam da razlika među populacijama vjerojatno stvarno *postoji*, ali je nijedanput nismo mogli statistički dokazati, vjerojatno zbog nedovoljno velikih uzoraka.

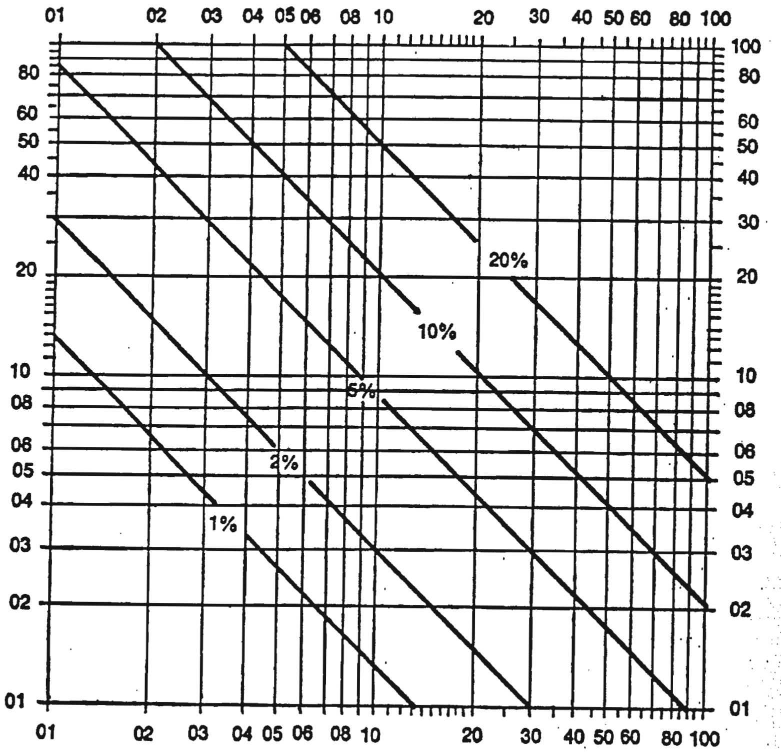
Za takve slučajeve pojedini statističari predlažu neke tehnike izračunavanja značajnosti razlike za sve rezultate zajedno, no mi ćemo prikazati samo jednu od tih tehnika, i to najjednostavniju, tj. tehniku kombiniranja *dviju* razina značajnosti u jednu razinu značajnosti. Radi se dakle o slučajevima kada imamo *dvije* razlike u istom smjeru, dobivene dvaput. Postupak potječe od P. C. Bakera, a iznosi ga McGuigan kao rezultat osobnog saopćenja autora. Postupak je vrlo jednostavan, jer se radi o *nomogramu* s kojega vrlo jednostavno očitavamo rezultat.

Na slici 9.14. prikazan je nomogram. Na njegovu horizontalnom i vertikalnom rubu nalaze se razine značajnosti, dobivene separatno u svakom od dva eksperimenta.

Pretpostavimo da je taj nastavnik u jednom testiranju dobio razliku u korist ženskih ispitanika, kojoj razini značajnosti iznosi $P = 0,08$ (dakle razlika nije statistički značajna), a u drugom pokusu dobio je ponovno razliku u istom smjeru, kojoj razina značajnosti iznosi $P = 0,10$ (pa prema tome ni ta razlika nije statistički značajna).

Vjerojatnost P za oba slučaja uzeta *zajedno* nađe se tako da se uz jedan rub nomograma nađe P koji je vezan za prvi eksperiment, a uz drugi rub nomograma P vezan za drugi eksperiment, te se na *sjecištu* tih vrijednosti u samom nomogramu aproksimativno iz *dijagonala* odredi kombinirani P. U našem slučaju P od 0,08 i 0,10 sijeku se na razini koja je *nešto niža* od $P = 0,05$, te na temelju toga zaključujemo da je razlika u oba kombinirana pokusa statistički *značajna*.

Treba, međutim, upozoriti da rezultat koji na taj način dobijemo, *ne mora nužno biti jednak* rezultatu koji bismo dobili da smo saželi zajedno rezultate prvog i drugog eksperimenta i testirali klasičnim putem (t-testom) značajnost razlike. Drugim riječima, eksperimentator bi u tom slučaju našao aritmetičku sredinu muških ispitanika iz oba navrata zajedno, a isto tako aritmetičku sredinu ženskih ispitanika, i testirao bi razliku. Razlog činjenici da se time mogu dobiti i drukčiji rezultati nego nomogramom jest u tome što se može dogoditi (a to se često u praksi i događa) da zbog različitih razloga u drugom eksperimentu dobijemo, doduše sličnu, *razliku* između jedne i druge aritmetičke sredine, ali su obje aritmetičke sredine na *drugoj razini*: tako je, na primjer, taj nastavnik prilikom prve primjene testa mogao dobiti da muškarci postižu prosječno 30, a žene 33 boda, a u drugom eksperimentu iduće godine muški su ispitanici postigli prosječno 41 bod, a žene 43 boda. Spajanje *takvih* rezultata u računu testiranja značajnosti razlike moglo bi vrlo lako dovesti do toga da razlika ni sada ne bude statistički značajna jer se značajno povećao *varijabilitet* jedne i druge varijable.



Slika 9.14. Nomogram za testiranje dvaju testova značajnosti (prema P. C. Bakeru)

Čini se da je u tom slučaju opravdanije i logičnije upotrijebiti nomogram koji *nije osjetljiv* na takve promjene u varijabilitetu. To, naravno, vrijedi samo za one slučajeve kada su takve promjene u varijabilitetu uzrokovane drugim razlozima, a ne stvarnim varijabilitetom mjerene pojave.

— — —

N a p o m e n a: U statističkim udžbenicima ne može se naići na jedno upozorenje, koje u nekim slučajevima može biti vrlo važno: ako mjerimo dvije *populacije,* pa nađemo razliku među aritmetičkim sredinama, onda je ta razlika stvarna činjenica, i *nećemo testirati je li ona statistički značajna.* To potpuno logički proizlazi iz onoga što znamo o smislu testiranja značajnosti razlika: neka nađena razlika među aritmetičkim sredinama je razlika među *uzorcima,* pa testiranjem značajnosti želimo provjeriti postoji li razlika i među populacijama. Međutim u ovom slučaju mi mjerimo populacije, pa nemamo više što testirati! Takvi se slučajevi mogu dogoditi ako smo populaciju *definicijom znatno ograničili:* na primjer, ako nas zanima da li se učenici 5.a razreda neke škole 1990. godine razlikuju po prosječnoj težini od

učenika 5.b razreda, pa ako smo izmjerili sve učenike u oba razreda, onda je razlika u aritmetičkim sredinama koju smo dobili (pod pretpostavkom da nismo pogriješili u mjerenju) sigurno i *statistički značajna,* tj. *populacije* se razlikuju! (Populacije su u našem primjeru: svi učenici 5.a razreda te škole u 1990. godini, i svi učenici 5.b razreda te škole u 1990. godini)

Budući da sada vjerojatno ima podosta čitalaca koji su se morali dobrano napregnuti da shvate sve što je u ovom opsežnom poglavlju bilo tumačeno, a vjerojatno ima i onih kojima ipak mnogo toga nije posve jasno (jer ova tema, tj. tema "distribucije uzoraka" i testiranja značajnosti razlika je tema na kojoj otpada najviše onih koji su se upustili u proučavanje statistike!), to smatram potrebnim da nešto napomenem:

Mnogi su se sigurno poveselili kada su na početku knjige nekoliko puta pročitali da matematika (osim 4 osnovna računa) nije potrebna za svladavanje gradiva ove knjige. Da, točno je (vjerojatno ste se već do sada uvjerili), matematika nije potrebna. No, ako nema matematike, to ne znači da smo zato oslobođeni svih problema! Jer, kako vidite, statistika nije baš posve jednostavna i posve lako shvatljiva. Moglo bi se čak reći da su neka područja (na primjer ovo!) vrlo kompleksna.

No, ako vas to može utješiti, činjenica je da su gotovo svi koji su se bilo kada u životu počeli baviti statistikom, naišli na slične barijere, ali su ipak svladali statistički način pristupa, a to znači da teškoće prvenstveno proizlaze iz *neobičnosti sadržaja.* Jer, prije nego što ste počeli proučavati statistiku, sigurno se niste susretali s problemima ove vrste. Ta *neobičnost* problema i načina rezoniranja vjerojatno je glavni razlog konfuziji koja se možda u ponekoj glavi stvorila.

Za utjehu čitaocima mogu pripomenuti da problema ove vrste uglavnom u knjizi više neće biti. Ako ste pojam normalne raspodjele i osnovne pojmove t-testa razumjeli i "apsorbirali", ostali dio gradiva ne bi smio više predstavljati teži problem.

ZADACI ZA VJEŽBU

1. Vitalni kapacitet skupine od 80 radnika u jednom kamenolomu iznosio je 3520 ccm ($s = 300$), a skupine od 120 radnika jedne tvornice u istom mjestu bio je 3615 ($s = 310$). Možemo li razliku u vitalnom kapacitetu smatrati statistički značajnom?
2. Prosječno trajanje bolovanja od neke bolesti kod 20 necijepljenih bolesnika iznosilo je 8,5 dana ($s = 1$), a 20 drugih bolesnika, koji su bili cijepljeni, bolovali su u prosjeku 7 dana ($s = 1$). Je li razlika u trajanju oboljenja statistički značajna?
3. Jedan je istraživač želio ustanoviti ima li neka droga utjecaja na koordinaciju pokreta. Devet ispitanika eksperimentalne skupine i deset ispitanika kontrolne skupine (koji su umjesto droge primili "placebo") postigli su na uređaju za mjerenje psihomotorike ove rezultate (veći broj znači bolji rezultat):

E grupa	K grupa
12	21
14	18
10	14
8	20
16	11
5	19
3	8
9	12
11	13
	15

Je li razlika u prosječnom uspjehu tih grupa statistički značajna?

4. Pedeset dječaka i pedeset djevojčica istih godina testirani su jednim testom. Dječaci su u prosjeku postigli 78 bodova, a djevojčice 81, sa standardnim devijacijama 7 (dječaci) i 9 (djevojčice). Je li razlika među njima statistički značajna?
5. Na 25 ljudi izmjerena je sedimentacija u prvom satu. Dobiven je prosjek od 10 mm i standardna devijacija 2 mm. Razlikuje li se ovaj prosjek statistički značajno od vrijednosti 12 mm?
6. Na Olimpijadi u Insbrucku 1976. godine 40 skakača postiglo je u prvom i drugom skoku daljine prikazane u donjoj tablici. Jesu li u prvom ili drugom skoku skakali prosječno dalje i je li razlika statistički značajna?

Skakač	1. skok	2. skok	Skakač	1. skok	2. skok
1.	79	78,8	21.	92,5	89
2.	84	79	22.	81	76
3.	93,5	92,5	23.	102,5	91
4.	68	73	24.	88	86
5.	90	87,5	25.	86,5	83
6.	85	81	26.	89,5	93
7.	81	77	27.	85	82,5
8.	85	80	28.	85,5	78,5
9.	84,5	81,5	29.	102	89,5
10.	79	76	30.	74	71
11.	88	85	31.	88,5	80
12.	95	91	32.	78,5	74,5
13.	84	84	33.	85	82,5
14.	85	78	34.	97,5	97
15.	91	97	35.	89	85
16.	88	79	36.	90	84,5
17.	81	75,5	37.	86	86
18.	85	78	38.	81	81,5
19.	83	74	39.	89,5	84,5
20.	78,5	71	40.	88	85

10.

"DVOSMJERNO" ILI "JEDNOSMJERNO" TESTIRANJE RAZLIKE

U modernijim statistikama sve se češće susrećemo s pojmom "dvosmjernog" ("two-tailed") i "jednosmjernog ("one-tailed") testiranja razlike, pa ćemo razliku između oba ova način ukratko rastumačiti.

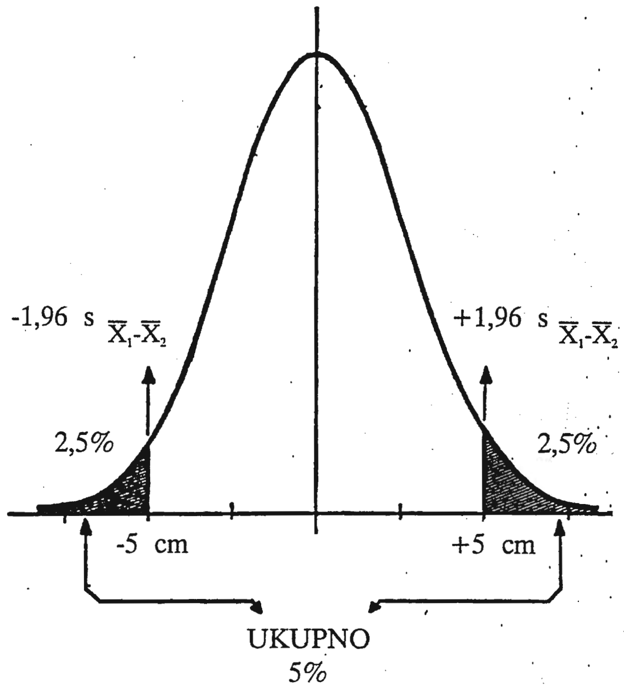
U svim našim dosadašnjim primjerima testiranja značajnosti razlike između dvije aritmetičke sredine naše je pitanje glasilo: "Je li *razlika* između obje aritmetičke sredine dovoljno velika da bismo je mogli smatrati statistički značajnom?" Nas je prema tome zanimala razlika "kao takva", a ne i *smjer* razlike. Drugim riječima, ako odbacimo nul-hipotezu (tj. da se obje aritmetičke sredine ne razlikuju značajno), dokazali smo da se one *razlikuju*. Statističkim simbolima izraženo, dokazali smo ovo:

$$\mu_1 \neq \mu_2$$

(Čitalac će se sjetiti da simbolom μ označujemo aritmetičku sredini *populacije*, a čitav gornji izraz treba čitati ovako: μ_1 *razlikuje se* od μ_2.)

Ako nas je, na primjer, zanimalo postoji li *razlika* u visini djece iz dvije različite pokrajine, mi smo ih izmjerili i ustanovili da su djeca u kraju A u prosjeku 5 cm viša od djece istih godina u kraju B, i testiranjem značajnosti ustanovili smo da je ta razlika od 5 cm statistički značajna na razini od 5%. Kao što znamo, to znači da bi se — u slučaju kad među populacijama *ne bi* postojala razlika — takva (ili veća) razlika od 5 cm mogla *slučajno* pojaviti najviše u 5% slučajeva. Kao što pokazuje slika 10.1, razlike koje su *apsolutno* veće od 5 cm (dakle veće od *plus* 5, kao i veće od *minus* 5), padaju već u regiju od 2,5% na svakoj strani krivulje, što ukupno iznosi 5%. Prema tome, "dvosmjerni" test pokazuje da je u ovom slučaju *razlika* od 5 cm statistički značajna, a iz naših je rezultata jasno da su djeca iz kraja A statistički značajno *viša*.

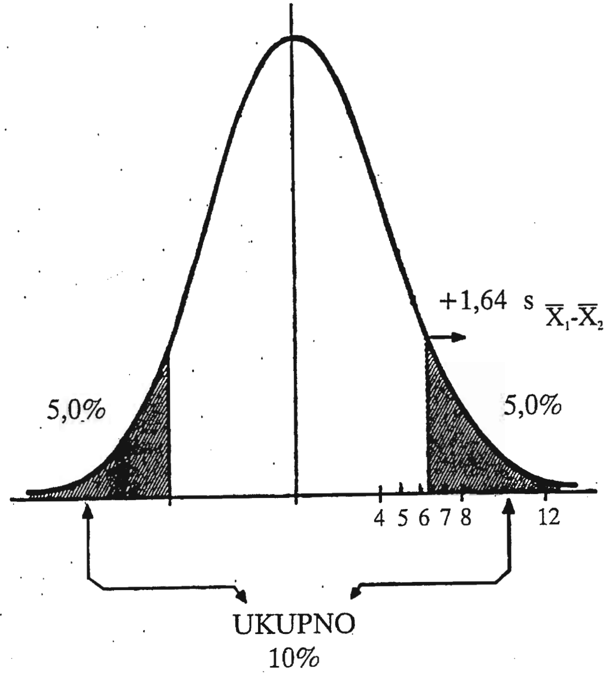
Kad bi nas međutim već unaprijed zanimala ne samo razlika, nego i *smjer* razlike, mogli bismo upotrijebiti "jednosmjerni" test. Na primjer, ako na skupinu ispitanika primijenimo neko sredstvo, koje bi — prema poznavanju njegova kemijskog sastava — trebalo djelovati tako da *povećava* frekvenciju pulsa, pa dobijemo zaista određeno povećanje čiju statističku značajnost želimo testirati, zanima nas samo je li došlo do statistički značajnog *povećanja* frekvencije pulsa. (Da smo dobili smanjenje pulsa, ne bismo razliku niti testirali!) Uzmimo da smo ustanovili da se puls prosječno povisio za 7 udaraca u minuti, a da je standardna pogreška te razlike $s_{\bar{x}_1-\bar{x}_2} = 4$; u krivulji koja nam pokazuje distribuciju razlika (pod hipotezom da



Slika 10.1. "Dvosmjerno" testiranje značajnosti razlike među aritmetičkim sredinama

razlike među populacijama zapravo nema), *povećanje* pulsa od 7 udaraca pada na +1,75 standardne devijacije te distribucije (vidi sliku 10,2), a iz tablice normalne raspodjele (tablica A u Dodatku) vidimo da 5% rezultata na jednoj strani krivulje pada izvan 1,64 standardne devijacije. Prema tome, naša dobivena razlika (jer je $t = 1,75$) već pada u zonu od 5%, i možemo je smatrati statistički značajnom. U ovom slučaju, odbacivši nul- hipotezu, mi smo dokazali, da je druga aritmetička sredina statistički značajno *veća* od prve: $\mu_2 > \mu_1$. Značajnost se u tom slučaju piše: $P/2 < 0,05$.

Treba, međutim, upozoriti da je potrebno *unaprijed* odlučiti hoćemo li razliku među aritmetičkim sredinama u jednom eksperimentu testirati "dvosmjernim" ili "jednosmjernim" testom. Ako se odlučimo za "jednosmjerni" test, to treba biti na temelju jakih argumenata. Pojedini eksperimentatori teško mogu odoljeti napasti da neku razliku koja nije dokazana kao značajna u "dvosmjernom" testu, nakon toga testiraju "jednosmjernim" testom i tako eventualno "postignu" značajnost. Naime, kod dvosmjernog testiranja za statističku značajnost od 5% potreban je t od najmanje 1,96, a kod jednosmjernog testiranja samo 1,64!



Slika 10.2. "Jednosmjerno" testiranje značajnosti razlike među aritmetičkim sredinama

Zbog toga se traži da se način testiranja odredi unaprijed, tj. prije nego su nam poznati rezultati mjerenja. U slučaju da je eksperiment takav da daje *opravdanje* jednosmjernom testiranju (kao npr. u maloprije spomenutom eksperimentu s pulsom), onda se unaprijed moramo za takvo testiranje odlučiti. U slučaju neizvjesnosti treba bezuvjetno raditi s dvosmjernim testom.

Uopće, kao praktičan savjet moglo bi se kazati ovo: *kod testiranja značajnosti razlika treba u pravilu upotrebljavati dvosmjerni test, i samo u iznimnim, temeljito opravdanim slučajevima, jednosmjerni test.*

Evo nekoliko primjera kada bi se, po mišljenju nekih statističara, moglo upotrijebiti jednosmjerno testiranje:

1. Ako testiramo bakterije u vodi za piće, neku ćemo vodu smatrati neprikladnom samo tada ako je broj bakterija *previsok*.
2. Testirajući jedno novo gnojivo s obzirom na prihod usjeva, odbacit ćemo staro gnojivo samo onda ako je novo gnojivo dalo značajno *bolje* rezultate.
3. Neki građevni materijal možemo testirati s obzirom na čvrstoću i odbaciti ga jedino ako je on značajno *slabiji* od drugih.
4. Sadržaji konzervi s hranom, kojima je označena težina, mogu se testirati s obzirom na težinu i treba ih odbaciti kao nedovoljno punjene samo ako su značajno *lakše* od propisane težine.
5. Neki ćemo lijek protiv hipertonije smatrati boljim od drugih ako on značajno više *snizuje* krvni tlak od ostalih lijekova.
6. Neku novu metodu u nastavi proglasit ćemo značajno boljom od drugih ako ona značajno *povećava* uspjeh učenika.

No neki moderni autori (npr. Langley) ne slažu se ni s tim, tj. da se ni u takvim situacijama koristi jednosmjerno testiranje, jer to — između ostaloga — dovodi i do nepotrebne konfuzije u interpretaciji rezultata. Treba se zaista složiti s Langleyem koji kaže da "... u svakom istraživanju mora postojati makar i mala sumnja u vezi sa smjerom razlike, pa problem zato ne treba smatrati 'kompliciranim', već ga treba jednostavno rješavati rutinskom upotrebom dvosmjernog testiranja".

Dobru kritičku opasku dao je Rowntree: "Ako o nekoj razlici znate čak i to u kojem smjeru ona postoji, zašto uopće trebate testirati njezinu značajnost? Razlika koja je toliko predvidiva, *mora* biti statistički značajna".

Uostalom, kada testiramo bilo koju razliku, ako je ona statistički značajna, to ujedno znači da značajno odstupa u *određenom smjeru* (za koji prije nismo znali), pa ako je testiranje rađeno na razini $P = 0,05$, razlika u dobivenom *smjeru* sada je značajna na razini $P = 0,025$. Ako se kod dvosmjernog testiranja ne zadovoljavamo sa $P = 0,10$ (pa *makar* bi u tom slučaju razlika u jednom smjeru bila značajna na razini od $P = 0,05$), ne smijemo se kod jednosmjernog testiranja zadovoljiti sa $P = 0,05$ — a time ujedno otpada "prednost" jednosmjernog testiranja!

11. TESTIRANJE RAZLIKE IZMEĐU PROPORCIJA

Vrlo se često događa da moramo obrađivati podatke koji nisu rezultati kvantitativnog mjerenja jer svojstvo koje registriramo ne možemo mjeriti u kvantitativnim jedinicama, već jedino registriramo u kojoj se frekvenciji pojavljuje. U tom slučaju radimo s postocima odnosno proporcijama (ili pak hi-kvadrat testom /vidi pogl. 15). Na primjer, registriramo da li neki ljudi imaju ili nemaju smeđe oči, da li su neki oboljeli od neke bolesti, itd. U svim tim slučajevima ne možemo se služiti izračunavanjem aritmetičke sredine, nego računamo *proporciju* pojave. Na primjer, ako od 100 ljudi oboli od neke bolesti 19, onda je proporcija oboljelih $19/100 = 0,19$. Pomnožimo li proporciju sa 100, dobivamo postotke: $p = 0,19 = 19\%$.

11.1. STANDARDNA POGREŠKA PROPORCIJE

Jednako kao i aritmetička sredina uzorka tako i proporcija uzorka ima svoju standardnu pogrešku. Dok standardna pogreška aritmetičke sredine dolazi rjeđe u praktičnu upotrebu, standardna pogreška proporcije može biti korisna i praktična pomoć u praktičnom radu.

Primjer. Prilikom neke epidemije u jednoj pokrajini koja obuhvaća 10 000 stanovnika, ekipa koja je poslana na teren pregledala je uzorak od 100 stanovnika i našla da ih je 12 zaraženo. Proporcija zaraženih u uzorku iznosi dakle $p = 12/100 = 0,12$.

Međutim, da li 12% zaraženih predstavlja *stvarni* (pravi) postotak zaraženih u čitavom kraju? Kolikoj se pogrešci izlažemo ako iz proporcije uzorka zaključujemo na proporciju populacije?

O odgovoru na ovo pitanje može, na primjer, ovisiti odluka o tom kako veliku ekipu treba poslati na suzbijanje te zaraze.

Ako N uzorka nije vrlo malen, ili ako p nije vrlo nizak (npr. ispod 0,10), ili vrlo visok (npr. iznad 0,90), možemo pretpostaviti približno normalnu raspodjelu proporcija uzoraka oko prave proporcije populacije. Standardna pogreška proporcije računa se u tom slučaju prema formuli:

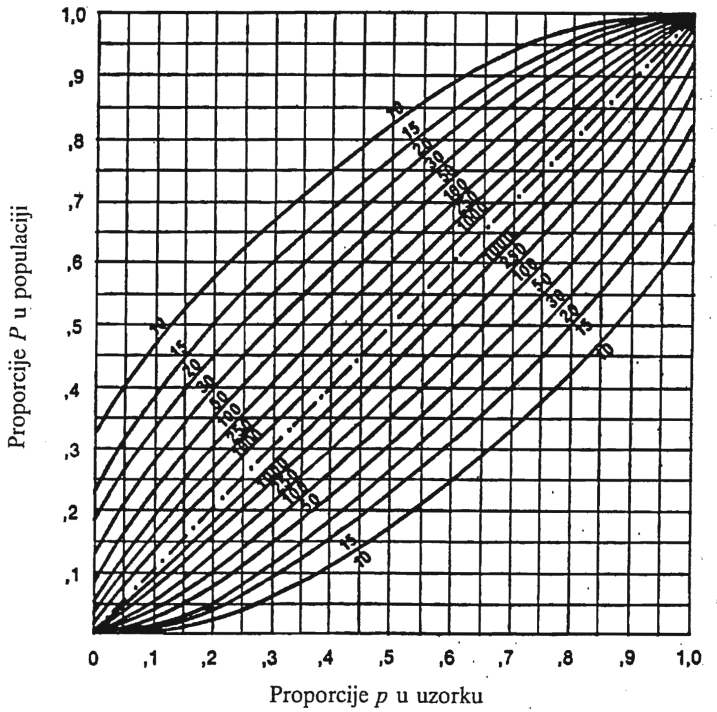
$$s_p = \sqrt{\frac{p \cdot q}{N}} \tag{11.1}$$

pri čemu $q = 1 - p$, tj. proporcija, u kojoj se svojstvo *nije* pojavilo. U našem slučaju dobivamo:

$$s_p = \sqrt{\frac{0,12 \cdot 0,88}{100}} = \sqrt{\frac{0,1056}{100}} = 0,032.$$

Ako je raspodjela normalna, vrijedi isto pravilo kao i kod standardne pogreške aritmetičke sredine, tj. da smo 68% sigurni da naša proporcija ne odstupa od prave proporcije više ili manje od $\pm 0,03$, da smo 95% sigurni da od nje ne odstupa više od $\pm 0,06$, a praktički 100% sigurni da ne odstupa više od $\pm 0,09$.

NOMOGRAM ZA OČITAVANJE 95-POSTOTNIH GRANICA POUZDANOSTI PROPORCIJE
za uzorke veličine N = 10, 15, 20, 30, 50, 100, 250 i 1000



Slika 11.1. Nomogram za očitavanje 95-postotnih granica pouzdanosti proporcije

Želimo li prema tome računati na *najgoru* mogućnost, možemo pretpostaviti da stvarna zaraza stanovništva iznosi možda $p = 0,21$, odnosno 21% (iz 0,12 + 0,09). Već prema tome hoćemo li uzeti 95%-tne ili veće granice pouzdanosti, ovisi kako ćemo veliku ekipu poslati na teren da suzbija spomenutu zarazu.

95%-tne granice pouzdanosti za različito velike proporcije mogu se relativno lako očitati iz nomograma na slici 11.1. Prednost je takvog očitavanja u tome što možemo uzeti u obzir i one slučajeve kada distribucija nije više normalna (tj. kada je N vrlo mali ili p vrlo ekstreman, pa je distribucija asimetrična i klasičnim se načinom više ne može ispravno izračunati standardna pogreška proporcije).

Kako iz nomograma vidimo, na apscisi je označena proporcija uzorka, a na ordinati su 95%-tne granice pouzdanosti različito velikih uzoraka. 95%-tne granice pouzdanosti očitavaju se s *obiju* krivulja koje se odnose na zadanu veličinu uzorka. Na primjer, ako u uzorku od $N = 20$ dobijemo proporciju $p = 0,20$, onda smo 95% sigurni da je prava proporcija između 0,05 i 0,45.

11.2. RAZLIKA IZMEĐU VELIKIH NEZAVISNIH UZORAKA

Kao što kod proporcija nismo mogli izračunavati standardnu pogrešku onako kako smo to radili kod aritmetičke sredine, jednako tako ne možemo ni na do sada opisane načine testirati značajnost razlika između dvije proporcije, nego ovdje upotrebljavamo jedan drugi način računanja koji se, međutim, u svojim principima ne razlikuje od načina računanja značajnosti razlika između dvije aritmetičke sredine.

Primjer. U grupi od 230 muškaraca, 90 su oboljeli od neke bolesti, a u grupi od 260 žena, 170 ih je oboljelo od te iste bolesti. Postoji li značajna razlika između muškaraca i žena u frekvenciji oboljenja od te bolesti?

Najprije ćemo izračunati postotak odnosno proporciju oboljenja kod muškaraca i žena:

Muškarci:	$N_1 = 230$,	$p_1 = 90/230 = 0,391$
Žene:	$N_2 = 260$,	$p_2 = 170/260 = 0,654$
	$q_1 = 1 - p_1 = 0,609$	
	$q_2 = 1 - p_2 = 0,346.$	

Ako su N_1 i N_2 dovoljno veliki (npr. 100 ili više), i ako proporcije nisu ekstremne (npr. veće od 0,90 ili manje od 0,10), možemo za standardnu pogrešku razlike proporcija upotrijebiti formulu koja je analogna (po smislu) formuli za izračunavanje standardne pogreške razlike između dvije aritmetičke sredine, tj. drugi korijen iz sume kvadriranih standardnih pogrešaka svake proporcije:

$$s_{p_1 - p_2} = \sqrt{s_{p_1}^2 + s_{p_2}^2}. \tag{11.2}$$

Budući da se standardna pogreška proporcija računa prema formuli (11.1), to se formula (11.2) može svesti na:

$$s_{p_1 - p_2} = \sqrt{\frac{p_1 \cdot q_1}{N_1} + \frac{p_2 \cdot q_2}{N_2}}. \tag{11.3}$$

Uvrstimo li vrijednosti iz našeg primjera u gornju formulu, dobivamo:

$$s_{p_1-p_2} = \sqrt{\frac{0,391 \cdot 0,609}{230} + \frac{0,654 \cdot 0,346}{260}} =$$

$$= \sqrt{\frac{0,238}{230} + \frac{0,226}{260}} \doteq \sqrt{0,0010 + 0,0009} =$$

$$= \sqrt{0,0019} = 0,0436.$$

Razlika među proporcijama je 0,654 - 0,391 = 0,263, a ako tu razliku podijelimo njezinom standardnom pogreškom, dobivamo:

$$t = \frac{0,263}{0,0436} = 6,03.$$

Budući da se razlike među proporcijama ne raspoređuju potpuno jednako kao razlike među aritmetičkim sredinama (distribucija proporcija i distribucija razlika među proporcijama odstupaju to više od normalne distribucije što je N manji, i što je p bliži 0 ili 1), neki statističari preporučuju da treba — za svaku sigurnost — postaviti kriterij da je razlika među proporcijama statistički značajna, ako je tri puta veća od svoje pogreške. U našem primjeru uočena razlika je prema tome statistički značajna.

Međutim, ako su uzorci veliki, a p ili q nije ekstremno mali ili velik, može se pretpostaviti normalna raspodjela, pa je prema tome razlika među proporcijama značajna na razini od 5%, ako je t veći od 1,96.

Kao pojednostavnjeno pravilo može nam poslužiti ovaj princip: ako Np i Nq kod oba uzorka iznose više od 5, može se pretpostaviti normalna raspodjela.

N a p o m e n a. Neki autori (Guilford) smatraju da normalnu raspodjelu možemo pretpostaviti samo ako Np ili Nq nisu manji od 10. Neki (Siegel) čak traže da izraz $N\,p\,q$ bude najmanje 9. To je možda razlog zašto se u praksi znatno više upotrebljava hi-kvadrat test nego račun proporcija, iako bismo i uz pomoć proporcija neke probleme mogli lako riješiti. Osim toga to je i prilika da vidite da se ni profesionalni statističari (iako su oni u svojoj osnovi *matematičari*) ne slažu uvijek u svom mišljenju!

11.3. RAZLIKA IZMEĐU VELIKIH ZAVISNIH UZORAKA

Ako su varijable u *korelaciji* (na primjer, ako su obje proporcije dobivene na *istim* ispitanicima), treba provesti korekturu pri izračunavanju značajnosti razlike među proporcijama. U načelu ta je korektura potpuno analogna korekturi koju obavljamo pri testiranju značajnosti razlika među aritmetičkim sredinama, ako su varijable u korelaciji (vidi formulu 9.7!). Prema tome, u tom slučaju računamo prema formuli:

$$s_{p_1-p_2} = \sqrt{s_{p_1}^2 + s_{p_2}^2 - 2r_{1,2}s_{p_1}s_{p_2}} \qquad (11.4)$$

N a p o m e n a. O tom kako se u ovom slučaju izračunava korelacija vidi pogl. 13.13. o izračunavanju ϕ koeficijenta korelacije.

Međutim, predložen je jedan jednostavniji postupak koji se sastoji u tome da rezultate iz obje varijable unesemo u tablicu, odakle možemo standardnu pogrešku razlike izračunati bez izračunavanja korelacije (tj. račun korelacije već je uključen!). Kod tog je računa potrebno poznavati za svakog ispitanika njegove individualne rezultate u prvoj i u drugoj varijabli.

Primjer. 150 mladića podvrgnemo testu trčanja na 100 m, u kojem je "norma" 12,0 sek. 75 mladića postiglo je "normu" ($p_1 = 0,500$). Tada ih sve podvrgnemo treningu trčanja na kratke pruge i ponovno ih testiramo te ustanovimo da su sada 100 mladića postiglo "normu" ($p_2 = 0,666$).

Tablicu postavljamo tako da vertikalno unosimo rezultate jednog, a horizontalno drugog trčanja:

		Drugo trčanje		
		Nisu postigli "normu"	Postigli "normu"	Ukupno
Prvo trčanje	Postigli "normu"	10	65	75
	Nisu postigli "normu"	40	35	75
	Ukupno	50	100	150

Sada ćemo svaku vrijednost u tablici pretvoriti u proporcije *ukupnog* broja ($N = 150$) i ujedno ćemo pojedine ćelije u tablici označiti slovima a, b, c i d:

		Drugo trčanje		
		Nisu postigli "normu"	Postigli "normu"	Ukupno
Prvo trčanje	Postigli normu	$0,067_a$	$0,433_b$	$0,500_{p_1}$
	Nisu postigli "normu"	$0,267_c$	$0,233_d$	0,500
	Ukupno	0,334	$0,666_{p_2}$	1,000

Standardnu pogrešku razlike proporcija izračunat ćemo prema formuli:

$$s_{p_1-p_2} = \sqrt{\frac{(d+a)-(d-a)^2}{N}}. \qquad (11.5)$$

Treba imati na umu da ta formula vrijedi samo za *ovako* postavljenu tablicu u kojoj se u ćelijama "a" i "d" nalaze oni koji su *promijenili* svoj uspjeh u kriteriju. Ako margine tablice postavljamo na drugi način, naravno da i formula u tom slučaju treba imati drugi oblik.

U našem slučaju, služeći se vrijednostima iz tablice, dobivamo:

$$s_{p_1-p_2} = \sqrt{\frac{(0,233+0,067)-(0,233-0,067)^2}{150}} =$$

$$= \sqrt{\frac{0,300-0,028}{150}} = \sqrt{0,0018} = 0,043.$$

Kako iz naše tablice vidimo, razlika u proporciji ($p_2 - p_1$) iznosi 0,666 - 0,500 = 0,166. Prema tome,

$$t = \frac{0,166}{0,043} = 3,86.$$

Razlika je statistički značajna.

V a ž n a n a p o m e n a. Formula (11.5) ima jedno ograničenje: izraz $(d+a)$ u tablici *frekvencija* (izvorna tablica) mora biti najmanje 10!

11.4. RAZLIKA IZMEĐU MALIH NEZAVISNIH UZORAKA

Ako su N_1 i N_2 mali (manji od 100), a pogotovu ako su proporcije prilično ekstremne (veće od 0,9 ili manje od 0,1), moramo upotrijebiti jednu drugu formulu, u kojoj ćemo kombinirati proporcije obiju grupa (slično kao što smo kod malih uzoraka kombinirali standardnu devijaciju, tj. izračunavali *zajedničku* standardnu devijaciju).

Primjer. U skupini od 50 dječaka, 24 su visine preko 122 cm, dok su u skupini od 60 djevojčica njih 18 više od 122 cm. Može li se ta razlika u proporciji visoke djece smatrati statistički značajnom?

Dječaci: $N_1 = 50$; $p_1 = 24/50 = 0,48$

Djevojčice: $N_2 = 60$; $p_2 = 18/60 = 0,30$

$$\text{Zajednička proporcija} = \frac{p_1 N_1 + p_2 N_2}{N_1 + N_2}. \tag{11.6}$$

U našem primjeru dobivamo:

$$\text{Zajednička } p = \frac{(0,48 \cdot 50) + (0,30 \cdot 60)}{50+60} = \frac{24+18}{110} = 0,38$$

$$q = 1 - p = 1 - 0,38 = 0,62.$$

Standardna pogreška razlike računa se dalje prema već poznatom postupku, tj. prema formuli (11.2) ili (11.3), ali, kako su p i q sada isti za obje grupe, ta se

formula svodi na ovaj oblik:

$$s_{p_1-p_2} = \sqrt{p \cdot q \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}. \tag{11.7}$$

Prema tome dobivamo:

$$s_{p_1-p_2} = \sqrt{0,38 \cdot 0,62 \left(\frac{1}{50} + \frac{1}{60}\right)} =$$

$$= \sqrt{0,2356 \cdot 0,03667} = \sqrt{0,00864} = 0,093.$$

Razlika u proporcijama je $p_1 - p_2 = 0,48 - 0,30 = 0,18$.

$$t = \frac{0,18}{0,093} = 1,94.$$

Kako se vidi, razlika ipak nije statistički značajna, jer nije 1,96 puta veća od svoje pogreške.

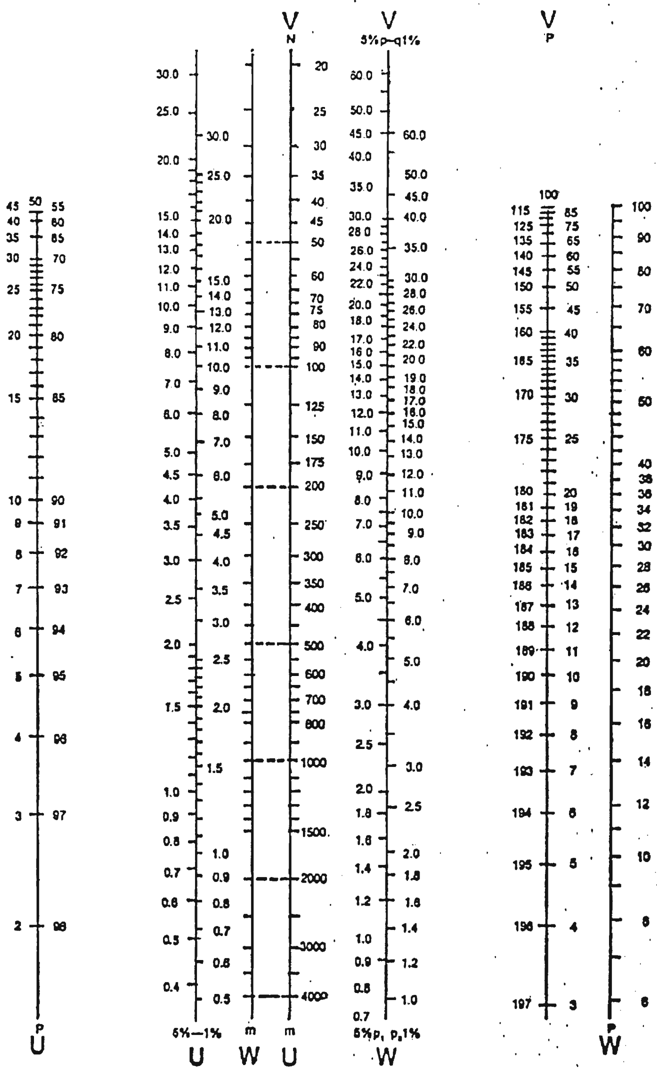
U novije vrijeme (1957) predložio je Brayer jedan jednostavniji postupak za izračunavanje značajnosti razlika među proporcijama. Taj se postupak, primijenjen na naš primjer, sastoji u ovim operacijama:

		Grupa 1.	Grupa 2.	Obje grupe
1.	Totalni broj opažanja	50	60	110
2.	Totalna frekvencija karakteristike	24	18	42
3.	(1) - (2) = 110 - 42			68
4.	(2) . (3) = 42 . 68			2856
5.	(4) : (1) = 2 856 : 110			25,963
6.	(5) podijeljeno s produktom totalnog broja opažanja u svakoj grupi = 25,963 : (50 . 60)			0,008654
7.	$s_{p_1-p_2} = \sqrt{(6)} = \sqrt{0,008654}$			0,093

Kako se vidi, rezultat je potpuno jednak rezultatu dobivenom pomoću standardnog postupka.

Važno je, međutim, zapamtiti da se formule (11.2) i (11.7) mogu upotrijebiti samo onda ako obje proporcije dolaze od dva nezavisna uzorka. Ako, međutim, proporcije potječu od *istog* uzorka (ovdje se ne misli na iste ispitanike kao kod proporcije u korelaciji!), te se formule ne mogu upotrijebiti. Tako, na primjer, ako u skupini od 100 djece nađemo da je 10% djece prelagano, a 12% djece preteško po težini, ne možemo opisanim postupcima testirati razliku između te dvije proporcije, jer obje potječu od *istog* uzorka od 100 djece. (Osim toga ovdje ne uspoređujemo isto s istim, već prelaku djecu s preteškom.)

U novije je vrijeme jedan statističar (Rosenbaum) izradio relativno jednostavan nomogram za rješavanje upravo takvih problema.



Slika 11.2. Nomogram za očitavanje granica pouzdanosti i testiranje značajnosti razlika među proporcijama (vidi tekst!)

Budući da se tim nomogramom mogu testirati i problemi koje smo upravo opisali (standardna pogreška proporcije i značajnost razlike između dvije proporcije), dat ćemo opis svih tih postupaka.

Nomogram (slika 11.2) služi, kako rekosmo, za tri vrste problema i sastoji se od tri grupe skala: U, V i W. Svaka grupa ima 3 skale: dvije vanjske i jednu srednju. U svim slučajevima jedna vanjska skala pokazuje veličinu uzorka, druga vanjska skala pokazuje postotak (dakle proporciju · 100), dok linija koja spaja vrijednosti vanjskih skala, daje rješenje na sjecištu sa srednjom skalom.

U-skala služi za dobivanje "granica pouzdanosti" nekog postotka. Desna skala "m" označuje veličinu uzorka; lijeva skala "p" daje postotak koji smo našli u uzorku; centralna skala daje 95 i 99%- tne granice pouzdanosti.

Primjer. U uzorku od $N = 350$ građana ustanovljeno je da ih 63 posjeduje motorno vozilo. To je 18%. Koje su granice pouzdanosti toga postotka? Položeno ravnalo koje spaja $m = 350$ na desnoj skali U sa 18% na lijevoj U skali siječe srednju U skalu otprilike na približno 4 za razinu od 5%, a približno na 5,5 za razinu od 1%, iz čega slijedi da su 95%-tne granice pouzdanosti $18\% \pm 4\%$, a 99%-tne granice pouzdanosti $18\% \pm 5,5\%$. (Da smo stanardnu pogrešku računali pomoću formule [11.1], dobili bismo da ona iznosi 0,0202, dakle 2· standardna pogreška odgovara našem dobivenom postotku 4.)

V-skala služi za testiranje značajnosti razlike među postocima. Veličinu obiju skupina ($N_1 + N_2$) očitavamo na lijevom V stupcu, *zbrojene* postotke obiju skupina očitavamo na desnom V stupcu (autor te postotke zove p i q, ali q u ovom slučaju zapravo znači p_2, dok $p = p_1$), a srednja V skala daje potrebnu razliku postotka $(p - q)$, koja je značajna na razini od 5 ili 1%.

Primjer. Poslužimo se našim prvim primjerom iz poglavlja 11.2. i testirajmo ga pomoću nomograma:

Muškarci	$N_1 = 230$	$p_1 = 90/230 = 0,39 = 39\%$
Žene	$N_1 = 260$	$p_2 = 170/260 = 0,65 = 65\%$

Je li razlika od 26% (65% − 39%) statistički značajna?

Spojimo li ravnalom 490 (230 + 260) na lijevoj V skali sa 104 (39 + 65) na desnoj V skali, dobivamo da bi ispod 9% razlike bilo značajno na razini od 5%, a nešto ispod 12% na razini od 1%.

W-skala je novost koju nam pruža ova tehnika testiranja, za razliku od standardnih tehnika: lijeva W skala ("m") daje veličinu uzorka, desna W skala daje *zbrojene* postotke oba svojstva, a na srednoj W skali označene su razlike u postocima koje su značajne na razini od 5 ili 1%.

V a ž n a n a p o m e n a: Nitko ne smije posjedovati *oba* svojstva!

Primjer. Poslužimo se primjerom koji smo malo prije spomenuli: ako u skupini od 100 djece nađemo da je 10% djece prelagano, a 12% preteško za svoje godine, možemo li tvrditi da ima više preteških nego prelaganih?

Povučemo li liniju s vrijednosti 100 na lijevoj W skali prema 22% (10 + 12) na desnoj W skali, možemo na srednjoj W skali ustanoviti da bi razlika morala biti veća od 9% (za razinu značajnosti od 5%), a kako naša razlika iznosi samo 2%, nju nikako ne možemo smatrati statistički značajnom.

Drugi primjer. Na skupini od 1 700 djece nađemo da ih 800 ima modre oči (47%), a 700 smeđe (41%). Je li ta razlika od 6% statistički značajna?

Nomogram pokazuje da bi oko 4,5% bilo dovoljno za razinu značajnosti od 5%, a da je razlika od 6% još značajna i na razini od 1%.

W-skala može se naravno koristiti i u slučajevima ako obje klase iscrpljuju sve mogućnosti, tj. daju sumu postotka 100, ali je u tom slučaju bolje raditi sa skalom U i testirati granice pouzdanosti *jednog* (bilo kojeg) od postotaka: ako granice pouzdanosti prelaze granicu od 50%, razlika se — razumljivo — ne može smatrati statistički značajnom.

Primjer. Od 200 studenata na prvom ispitu palo ih je 90 (45%), a prošlo 110 (55%). Je li razlika od 10% statistički značajna?

Testirat ćemo na U-skali značajnost jednog postotka, uzmimo postotka palih na ispitu (45%). Spojimo li $N = 200$ s vrijednosti 90 na lijevoj U skali, vidimo da su 95%-tne granice pouzdanosti 45%± cca 4,2% — dakle razlika je još uvijek značajna na razini od 5% (ali ne na razini od 1%).

11.5. RAZLIKA IZMEĐU MALIH ZAVISNIH UZORAKA

Ako su mali uzorci proporcija u *korelaciji* (kad eksperimentalnu i kontrolnu skupinu sačinjavaju isti ispitanici), treba, kao i kod velikih uzoraka, unijeti najprije rezultate u tablicu.

Primjer. 100 ispitanika podvrgnuli smo ispitivanju psihomotorike pomoću dva testa, koji su takve prirode da daju samo alternativne rezultate: "zadovoljava" i "ne zadovoljava". U prvom testu 60 ispitanika zadovoljava, u drugom 70. Je li ta razlika statistički značajna tj. je li drugi test stvarno lakši od prvoga?

Drugi test

		Nisu zadovoljili	Zadovoljili	Ukupno
Prvi test	Zadovoljili	5	55	60
	Nisu zadovoljili	25	15	40
	Ukupno	30	70	100

Ako sve vrijednosti u tablici pretvorimo u proporcije ukupnog broja N, dobivamo:

Drugi test

		Nisu zadovoljili	Zadovoljili	Ukupno
Prvi test	Zadovoljili	$0,05_a$	$0,55_b$	$0,60_{p_1}$
	Nisu zadovoljili	$0,25_c$	$0,15_d$	$0,40$
	Ukupno	0,30	$0,70_{p_2}$	1,00

Kod malih ćemo uzoraka standardnu pogrešku razlike izračunati prema formuli:

$$s_{p_1-p_2} = \sqrt{\frac{d+a}{N}}. \tag{11.8}$$

U našem slučaju dobivamo:

$$s_{p_1-p_2} = \sqrt{\frac{0,15+0,05}{100}} = \sqrt{0,002} = 0,0447.$$

Razlika u proporcijama $(p_2 - p_1) = 0,70 - 0,60 = 0,10$, pa je prema tome

$$t = \frac{0,10}{0,0447} = 2,24.$$

Prema kriteriju koji smo usvojili, zaključit ćemo da je razlika statistički značajna.

Važna napomena. Ograničenje formule (11.8) sastoji se u tome što frekvencija $(d+a)$ u tablici frekvencija mora biti najmanje 10!

Kako se vidi, kod proporcija i postotaka *uvijek je važno da navedemo i apsolutne brojeve* na kojima smo izvršili mjerenje, a ne samo proporcije, jer ako apsolutni brojevi nisu navedeni ne možemo uopće primijeniti račune koji su malo prije spomenuti. Koliku ulogu imaju apsolutni brojevi u konačnom rezultatu (iako *proporcija* ostaje jednaka!), najbolje može ilustrirati primjer, ako u našem ispitivanju visine dječaka i djevojčica (vidi str. 168) brojeve opažanja povećamo deset puta:

Dječaci: $N_1 = 500$; 240 visokih dječaka; $p_1 = 240/500 = 0,48$;

Djevojčice: $N_2 = 600$; 180 visokih djevojčica: $p_2 = 180/600 = 0,30$;

Zajednička $p = \frac{240+180}{1100} = 0,38$;

$q = 0,62$.

(N a p o m e n a. Čitalac će se sjetiti da je kod velikih uzoraka svejedno upotrijebimo li formulu za velike ili formulu za male uzorke!)

$$s_{p_1-p_2} = \sqrt{p\cdot q\left(\frac{1}{N_1}+\frac{1}{N_2}\right)} = \sqrt{0,38\cdot 0,62\left(\frac{1}{500}+\frac{1}{600}\right)} =$$

$$= \sqrt{0,2356\cdot 0,00367} = \sqrt{0,000865} = 0,029.$$

Diferencija proporcija = 0,18.

$$t = \frac{018}{0,029} = 6,21.$$

Prema tome, iako je proporcija ostala *jednaka* kao i u prijašnjem primjeru, razlika u proporciji sada je statistički *značajna*, dok prije nije bila!

Kako se vidi, značajnost razlike među proporcijama (jednako kao i među aritmetičkim sredinama) raste ako je broj opažanja veći. To je i logički potpuno razumljivo, jer nije svejedno da li smo npr. od 10 bacanja jednog komada novca dobili osam puta "glavu" (a dva puta "pismo"), ili smo od 100 bacanja osamdeset puta dobili "glavu" (a dvadeset puta "pismo"). Proporcija je u oba slučaja jednaka ($p = 0,80$), ali dok u prvom slučaju tu proporciju možemo još smatrati — s obzirom na mali broj mjerenja — *slučajnom*, dotle u drugom primjeru ne možemo pretpostaviti da bi čisti slučaj mogao uvjetovati takvu razliku, nego je ona vjerojatno uzrokovana nekim drugim sistematskim faktorom (npr. neispravnim komadom novca ili "pristranim" bacanjem novca). Zato je u tom slučaju ta proporcija od 0,80 mnogo "značajnija". O tom vidi završetak 3. poglavlja, gdje smo opisali kako povećanjem broja pokušaja *proporcija* nekog ishoda postaje sve bliža očekivanoj.

Na kraju ovog poglavlja treba reći ovo: rad s proporcijama *prvenstveno* je koristan pri određivanju standardne pogreške proporcije (sjetite se primjera s proporcijom zaraženih, gdje je trebalo odrediti veličinu ekipe). Što se tiče testiranja *razlike* među proporcijama, većinom se test računanja standardne pogreške razlike malo primjenjuje, jer — kako smo vidjeli — statističari nisu baš posve složni u tome kada možemo smatrati da je distribucija uzoraka normalna, a kada nije. Stoga nisu potpuno jednakog mišljenja niti o graničnoj vrijednosti t: neki misle da je dovoljno (kao i kod aritmetičkih sredina) da t bude najmanje 1,96, dok drugi smatraju da bi trebao biti — za svaki slučaj — čak i do 3. Čitalac će se uskoro uvjeriti da za tu svrhu postoji mnogo praktičniji, jednostavniji i "elegantniji" test, tj. hi- kvadrat test.

ZADACI ZA VJEŽBU

1. Pregledom uzorka veličine $N = 100$ djece utvrđeno je da ih 60 ima karijes. U kojim se granicama sa 99,7% sigurnosti može smatrati da se nalazi prava proporcija djece s karijesom?
2. Za neki narod je poznato da njegovi pripadnici imaju u 70% slučajeva svijetlu kosu. Jedan autobus sa 111 izletnika iz te zemlje stigao je u vaš grad. Ustanovili ste da 69 putnika ima svijetlu kosu. To je 62%. Možete li na temelju te činjenice utvrditi da uzorak nije reprezentativan što se tiče boje kose?
3. Od 300 studenata, 200 ih je položilo neki ispit, a od 500 studentica položilo ih je 300. Je li razlika u proporciji onih koji su položili statistički značajna?
4. Očitajte s nomograma na slici 11.1. kolike su 95%-tne granice pouzdanosti proporcije 0,70 kod uzoraka veličine:
   a. $N_1 = 10$
   b. $N_2 = 50$
   c. $N_3 = 100$.

12. TEŠKOĆE I "OPASNOSTI" PRI RADU S POSTOCIMA

U statističkom računanju treba, koliko se najviše može, izbjegavati račune s postocima jer se u tome lako pogriješi. Navest ćemo nekoliko tipičnih izvora konfuzije:

1. Kod prevelikih postotaka često imamo teškoća da kažemo koliko je puta nešto veće od drugoga.

Primjer. Populacija X je za 5 000% veća od populacije Y, koja broji 20 000 ljudi. Kolika je populacija X?

Odgovor: 5 000% od 20 000 $= 50 \cdot 20\,000$

$X = 1\,000\,000$.

2. Postoci od *premalih* uzoraka ne daju nam nikakvu pouzdanu predodžbu situacije. Ako pregledamo 10 ljudi, i kod 8 nađemo određeno oboljenje, ne valja taj rezultat pretvoriti u 80%.

Poznat je slučaj iz Drugoga svjetskog rata kada je jedan zapovjednik radarske stanice na Mediteranu, koji je imao 7 radarista, uzaludno neprestano tražio da mu se broj radarista poveća na 40-50. Jednog dana jedan od sedmorice radarista doživio je slom živaca zbog napornog rada. U svom idućem mjesečnom izvještaju zapovjednik je javio u bazu "da je prošlog mjeseca duševno oboljelo preko 14% radarista zbog prenapornog rada". Kada je izvještaj stigao u Washington, hitno je poslano još 35 radarista!

Godine 1976. javljeno je iz kneževine Lichtenstein da je nezaposlenost u 4 mjeseca narasla gotovo 20%. No kad se pogledaju stvarni brojevi, onda je situacija ovakva: broj nezaposlenih povećao se s 51 na 61!

Pogrešno je da se navode samo postoci, a ne i apsolutni broj, iz kojeg je taj postotak dobiven. Mnogi autori smatraju da u uzorku treba biti najmanje 100 slučajeva pa da se neka proporcija izrazi u postotku.

3. Ako smo neki broj povećali (ili smanjili) za određeni postotak i tako ga pretvorili u novi broj, nećemo ponovno dobiti početni broj ako novi broj smanjimo (ili povećamo) za isti postotak.

Primjer a): Plaća od 5 000 kn povećana je za 50%. Budući da su sada troškovi postali preveliki, nakon nekog vremena nova je plaća smanjena za 50%.

Je li konačna plaća jednaka prvoj plaći? Odgovor. Ako smo plaću od 5 000 kn povisili za 50%, nova je plaća iznosila 7 500 kn. Ako sata tu novu plaću smanjimo za 50%, konačna će plaća biti 3 750 kn.

Primjer b): Ako cijenu nekoj namirnici, koja je stajala 50 kn jedan kilogram, snizimo za 10%, pa se pokaže da dolazi do gubitka, te cijenu ponovno povisimo za 10%, hoće li sadašnja cijena biti jednaka prvoj cijeni?

Odgovor: 10% od 50 kn = 5 kn. Snižena cijena bila je, dakle, 45 kn. 10% od 45 kn = 4,5 kn. Poskupljenje za 10% dovelo je, dakle, do konačne cijene od 45 + 4,5 = 49,5 kn.

U vezi s ovim primjerima treba spomenuti da *snižavanje* postotka ne može biti veće od 100%. Primjer: Cijena nekog materijalu, koji je stajao 10 kn, povišena je za 200%. Prema tome, sadašnja je cijena 10 + 20 = 30 kn. Želeći vratiti cijenu na prijašnju cijenu, odredimo pojeftinjenje za 200%. 200 posto od 30 = 60; 30 - 60 = -30! Izlazi prema tome da bi trgovac morao svakom kupcu tog materijala uz materijal još — isplatiti 30 kn!

4. *Prosječni postotak* je vjerojatno najčešći izvor pogreške, jer se obično ne vodi računa o broju slučajeva kod svakog postotka.

Primjer: Ispitujući letalitet od neke bolesti (letalitet = broj umrlih na 100 oboljelih od neke bolesti), dobili smo ovu sliku:

Starost	Postotak umrlih	Broj bolesnika
Ispod 20 godina	47,5	40
20-39 godina	15,0	120
40-59 godina	22,4	250
60 i više godina	51,1	91

Koji je *opći* (prosječni) letalitet od te bolesti?

Odgovor. Vrlo se često događa da pojedini ispitivači jednostavno zbroje sve postotke i podijele ih njihovim brojem (136%:4 = 34%), no taj je postupak *neispravan*, a dopušten je samo onda ako je svaki od postotaka dobiven iz jednog broja opažanja (vidi diskusiju o zajedničkoj aritmetičkoj sredini, stranica 52.).

Ispravan se postupak sastoji u tom da nađemo za svaku grupu korigirane postotke s obzirom na N (postotak puta N):

Starost	Postotak umrlih	Broj bolesnika	Korigirani postotak
Ispod 20 godina	47,5	40	1900,0
20-39 godina	15,0	120	1800,0
40-59 godina	22,4	250	5 600,0
60 i više godina	51,1	91	4 650,1
Suma		501	13 950,1

$$\text{Prosječni postotak} = \frac{\text{suma korigiranih postotaka}}{\text{suma } N} \tag{12.1}$$

$$\text{Prosječni postotak} = \frac{13\,950,1}{501} = 27,8\%$$

Dakako, još je jednostavnije (i logički razumljivije), ako prosječni postotak izračunamo tako da ustanovimo koliko je ukupno ljudi umrlo od ukupnog broja bolesnika.

5) Na najveće teškoće pri izračunavanju u postocima nailazimo kad promatramo postotne promjene dvaju ili više brojeva u toku vremena.

Primjer. Cijene za 1/2 l mlijeka i 1 komad peciva bile su u nekom gradu:

	1990. god.	1991. god.
mlijeko	20 bodova	10 bodova
pecivo	5 bodova	10 bodova

Kako se vidi, mlijeko je pojeftinilo za 50%, a pecivo je poskupjelo za 100%. Jesu li cijene za mlijeko i pecivo *zajedno* pale, porasle ili ostale iste?

Odgovor: Ako kao osnovu uzmemo 1990. godinu, dobivamo situaciju, prikazanu na slici 12.1. A. Ta slika pokazuje da su cijene *porasle*. Ako, međutim, uzmemo kao osnovu 1991. godinu, dobivamo situaciju prikazanu na slici 12.1. B, a to je, kako se vidi, potpuno obratni rezultat! Naime, prema toj slici, cijene su pale.

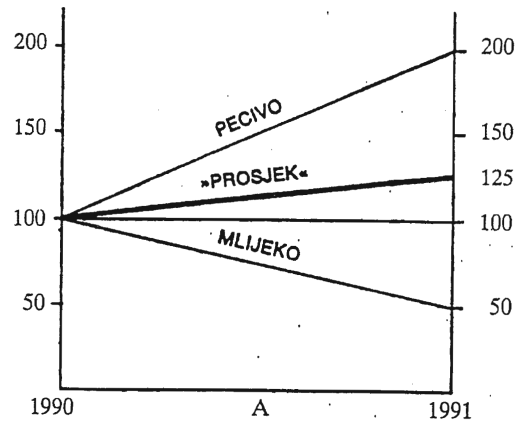
Prema tome, ovaj način prikazivanja ne smijemo upotrijebiti. Jednako tako bilo bi pogrešno uzeti u račun fiksirane količine namirnica, tj. rezonirati ovako: ako je 1/2 l mlijeka i 1 komad peciva 1990. stajalo ukupno 25 jedinica, a 1991. ukupno 20 jedinica, cijene su pale. Međutim, ako je netko 1990. kupovao 3 komada peciva uz $\frac{1}{2}$ litre mlijeka, on je plaćao 20 + 15 = 35 jedinica, a 1991. je za iste artikle plaćao 10 + 30 = 40 jedinica, pa je prema tome 1991. plaćao više!

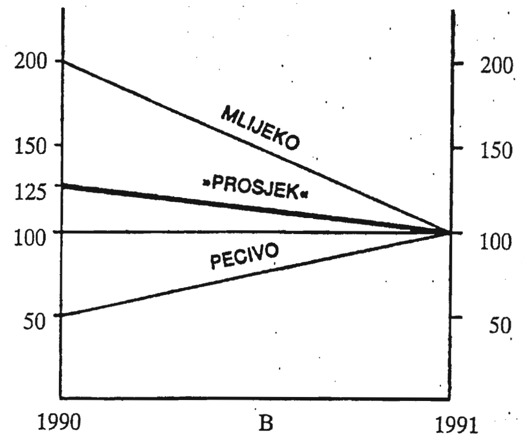
Jedini ispravan način, koji se ovdje smije primijeniti, je izračunavanje geometrijske sredine (vidi formulu 4.7).

Budući da mi imamo u našem primjeru dva podatka (cijena mlijeka i cijena peciva), to će naša formula biti drugi korijen iz umnoška oba postotka.

Račun ćemo izvesti tako da uzmemo kao *bazu* bilo koju od godina. Ako uzmemo bazu godinu 1990., postupak je ovaj:

1990. godina = baza = 100
Cijena mlijeka = 20 = 100%
Cijena peciva = 5 = 100%
Geometrijska sredina = $\sqrt{100 \cdot 100} = 100$
1991. godina:
Cijena mlijeka = 10 = 50% prema cijeni u 1990.
Cijena peciva = 10 = 200% prema cijeni u 1990.
Geometrijska sredina = $\sqrt{50 \cdot 200} = 100$
Prema tome, cijene su u *cjelini* ostale iste.





Slika 12.1. "Poskupljenje" ili "pojeftinjenje" često je prividno i ovisi o načinu na koji izračunavamo promjenu. Oba načina, prikazana na slici, su pogrešna (vidi tekst!)

Evo, ovih nekoliko primjera bilo je dovoljno da upozori čitaoca na opasnosti u radu s postocima. Zbog toga razloga vjerojatno najviše pogrešaka (a katkad i namjernih falsificiranja!) ima u području postotaka. Na neke od njih čitaoci će još biti upozoreni na drugom mjestu u ovoj knjizi, tj. u poglavlju "Zaključivanje u statistici".

ZADACI ZA VJEŽBU

1. Na 5 različitih skupina športaša ispitano je koliko je među njima profesionalaca te je nađeno:

Skupina	Velič. skupine	Postotak profesionalaca
Nogometaši	150	75
Košarkaši	200	50
Boksači	40	80
Lakoatletičari	80	20
Vozači trk. automobila	20	90

Koliki je prosječni postotak profesionalaca?

2. Ako je televizijska pretplata poskupjela 40%, a tarife za tramvajski prijevoz poskupjele za 20%, koliko posto iznosi prosječno poskupljenje?

13. KORELACIJA

13.1. SMISAO I PRINCIP KORELACIJE

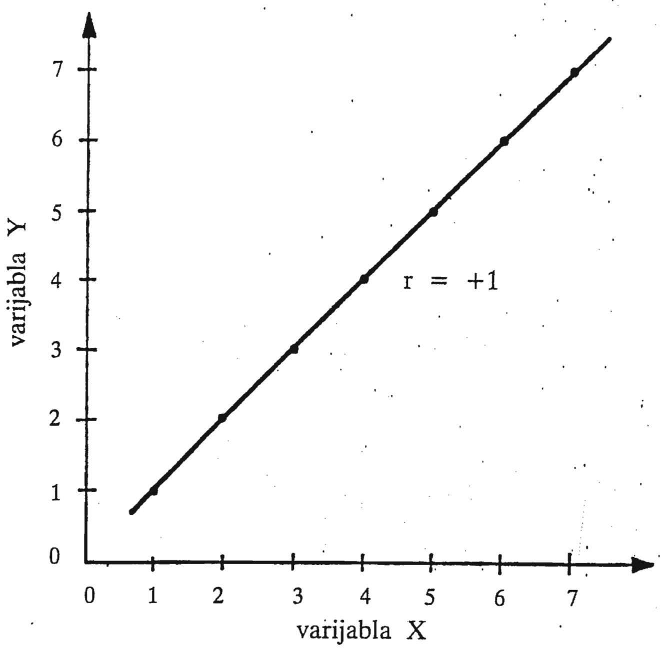
Često u životu opažamo da dvije pojave pokazuju neku međusobnu *zavisnost*, *povezanost*, odnosno *asocijaciju*: znamo da postoji određena povezanost između visine i težine (viši ljudi u prosjeku su teži od nižih), povezanost između ekonomskog statusa i zdravstvenog stanja, između starosti i krvnog tlaka, između rezultata na nekim ispitima i uspjeha na radnom mjestu, između količine oborina i gustoće vegetacije, između količine konzumirane hrane i tjelesne težine, itd.

Engleski matematičar Karl Pearson (1857-1936) razradio je računski postupak za izračuvanje stupnja povezanosti, i izrazio stupanj povezanosti brojem, koji je nazvao *koeficijentom korelacije*. No već prije njega, jedan drugi Englez, tj. sir Francis Galton (spominjali smo "Galtonovu oživu"!) (1822-1911), inače bratić Charlesa Darwina, razradio je osnovnu *logiku* takvog računanja. Galton se naime mnogo zanimao pitanjem u kakvom su odnosu visina očeva i njihovih sinova, inteligencija očeva i sinova, itd. Pri tome je našao da u visini postoji vrlo visoka povezanost (tj. da su sinovi vrlo često slične visine kao njihovi očevi), dok je ta povezanost znatno niža što se tiče inteligencije.

Pokušajmo najprije *grafički*, uz pomoć uglavnom poznatih primjera, protumačiti smisao korelacije.

1. Ako linearnom porastu jedne varijable odgovara također linearni porast druge varijable, i to tako da je jedna određena vrijednost jedne varijable uvijek povezana s *jednom* korespondentnom vrijednošću druge varijable (npr. odnos između polumjera i opsega kruga), onda je korelacija a) *pozitivna* (jer porastu jedne odgovara porast druge varijable), i b) potpuna, tj. *maksimalna* (jer veće slaganje od toga ne može postojati), i bilježi se izrazom $r = +1$ (vidi sliku 13.1).

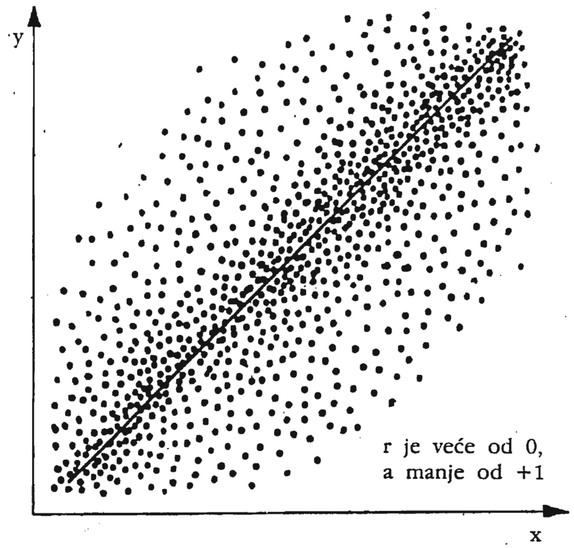
2. Ako linearnom porastu jedne varijable uglavnom odgovara linearni porast druge varijable, i to tako da je jedna određena vrijednost jedne varijable povezana s *više* vrijednosti druge varijable (npr. odnos između visine i težine ljudi: viši ljudi u prosjeku su i teži, ali jedna *određena* visina nije vezana upravo za jednu određenu težinu), onda je korelacija *pozitivna*, ali *nije maksimalna*, i bilježi se izrazom koji pokazuje je li povezanost manja ili veća, tj. izrazom koji je veći od 0, ali manji od +1 (vidi sliku 13.2).



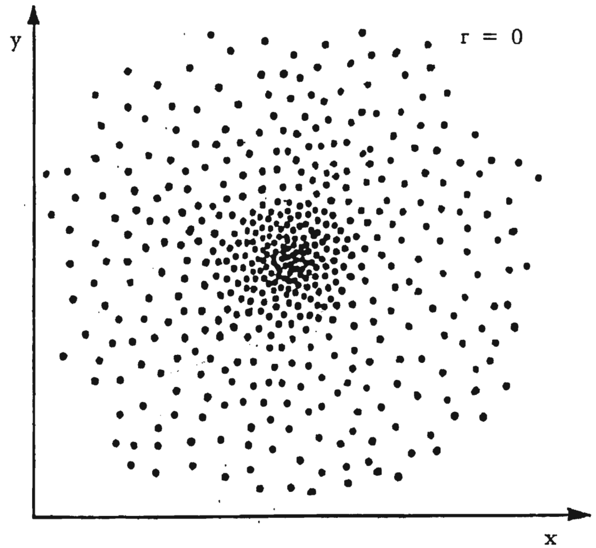
Slika 13.1. Pri maksimalnoj pozitivnoj korelaciji linearnom porastu jedne varijable odgovara linearni porast druge varijable i, osim toga, jednom rezultatu jedne varijable odgovara samo jedan rezultat druge varijable

3. Ako iz određene vrijednosti jedne varijable ne možemo ništa zaključiti na vrijednost druge varijable, tj. ako jednoj određenoj vrijednosti jedne varijable odgovara *bilo koja* (od mogućih) vrijednosti druge varijable (npr. odnos između duljine nosa i krvnog tlaka), onda *nema korelacije* između dvije pojave, i to se bilježi izrazom $r = 0$ (vidi sliku 13.3).

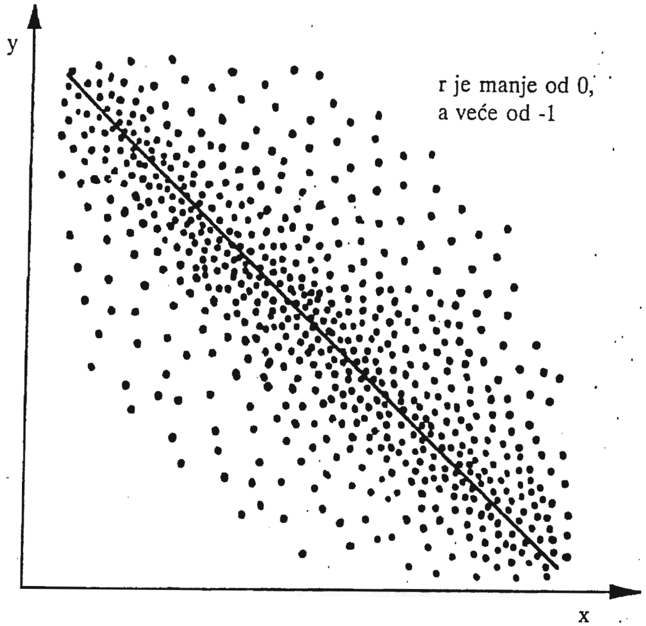
4. Ako linearnom *porastu* jedne varijable odgovara linearno *opadanje* druge varijable, ali je povezanost takva da je jedna vrijednost jedne varijable povezana s više vrijednosti druge varijable (npr. odnos između stupnja treniranosti i frekvencije pulsa u prvoj minuti oporavka: što je stupanj treniranosti veći, to je u prosjeku puls oporavka niži), onda je korelacija *negativna i nepotpuna*, i bilježi se izrazom koji je manji od 0, a veći od -1 (slika 13.4).



Slika 13.2. Porastom jedne varijable raste u prosjeku i druga, ali jednom rezultatu jedne varijable odgovara više rezultata druge varijable. Korelacija je pozitivna, ali je manja od +1



Slika 13.3. Ne postoji nikakva povezanost između obiju varijabli. Nekom rezultatu u jednoj varijabli odgovara bilo koji rezultat u drugoj varijabli. Korelacija iznosi 0

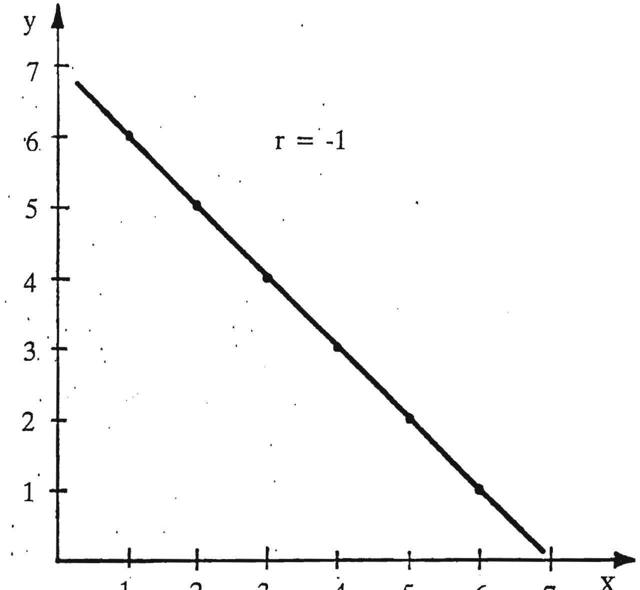


Slika 13.4. Porastom jedne varijable druga varijabla u prosjeku opada. Jednom rezultatu jedne varijable odgovara više rezultata druge varijable. Korelacija je negativna, ali ne maksimalno negativna

5. Ako linearnom *porastu* jedne varijable odgovara linearni *pad* druge varijable, i to tako da je jedna određena vrijednost jedne varijable povezana s *jednom* korespondentnom vrijednosti druge varijable (npr. odnos između vremena proteklog od ispaljivanja metka vertikalno uvis i brzine tog metka), onda je korelacija a) *negativna* (jer porastu jedne varijable odgovara pad druge, i obratno) i b) potpuna, tj. *maksimalna* (jer ne može biti veća, tj. ne može biti potpunije povezanosti), i bilježi se izrazom $r = -1$ (slika 13.5).

N a p o m e n a. U svim dosadašnjim primjerima govoreno je samo o *linearnoj* povezanosti između dvije varijable, a to znači o takvoj povezanosti koja se — kada je prikažemo grafički — može prikazati *ravnom* crtom, tj. pravcem. Naravno da postoje i takve povezanosti koje se ne daju prikazati pravcem, nego *zakrivljenom* linijom: na primjer, kada bismo promatrali odnos između broja ponavljanja nekog materijala, koji učimo, i količine upamćenog gradiva, onda bismo dobili zakrivljeni odnos, tj. u početku, s prvim ponavljanjima, količina naučenog materijala naglo bi rasla, a kasnije bi porast bio sve sporiji. Ili, ako bismo istraživali povezanost između intenziteta rasvjete i radnog učinka u nekom preciznom poslu, našli bismo da u početku radni učinak raste s porastom intenziteta rasvjete, da kasnije (kod već jakih intenziteta) promjene u intenzitetu rasvjete nemaju praktički nikakvog efekta na radni učinak, ali kada bismo rasvjetu i dalje povećavali, došlo bi konačno

do toga da učinak — zbog *zaslijepljenosti* radnika — počinje sve više i više *padati*. U svim takvim i sličnim slučajevima *povezanost* između dvije varijable može biti i *vrlo visoka*, ali ta povezanost *nije linearna*. Na slici 13.6. prikazano je nekoliko tipova takve nelinearne povezanosti.



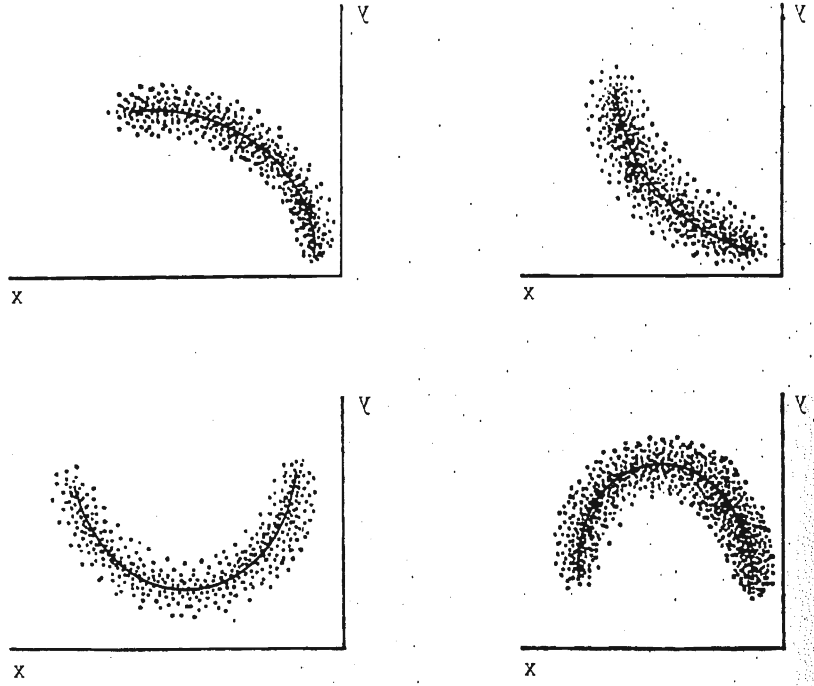
Slika 13.5. Linearnom porastu jedne varijable odgovara linearno opadanje druge varijable. Jednom rezultatu jedne varijable odgovara samo jedan razultat druge varijable. Korelacija je negativna i maksimalna, tj. iznosi -1

Mi ćemo se kod r koeficijenta korelacije baviti *samo linearnom* povezanosti, i to zbog dva razloga: prvo, linearna povezanost vrlo je česta i, drugo, precizno izračunavanje drugih oblika povezanosti mnogo je složenije od izračunavanja linearne povezanosti.

Zbog toga jedan praktičan savjet: kada vas zanima povezanost između dvije varijable, onda — prije nego što se odlučite za *izračunavanje* korelacije, *prikažite grafički* dobivene rezultate (taj grafički prikaz naziva se "scatter-diagram"), pa tek onda, ako je povezanost u cjelini manje-više linearna, možete provesti izračunavanje r koeficijenta.

U biologiji, medicini, psihologiji, sociologiji, i uopće u znanostima koje se bave *živim bićima*, praktički je nemoguće dobiti *potpunu* povezanost (bilo linearnu, bilo zakrivljenu). Tome je razlog jaki varijabilitet unutar mjerenih pojava. Prema tome, mjereći na živom materijalu obično dobivamo povezanosti koje se nižu od jedinice.

Na primjer, tražeći zavisnost između brzine rada osmorice radnika na jednoj preši i količine otpadaka, dobiveni su rezultati prikazani u tablici 13.1.

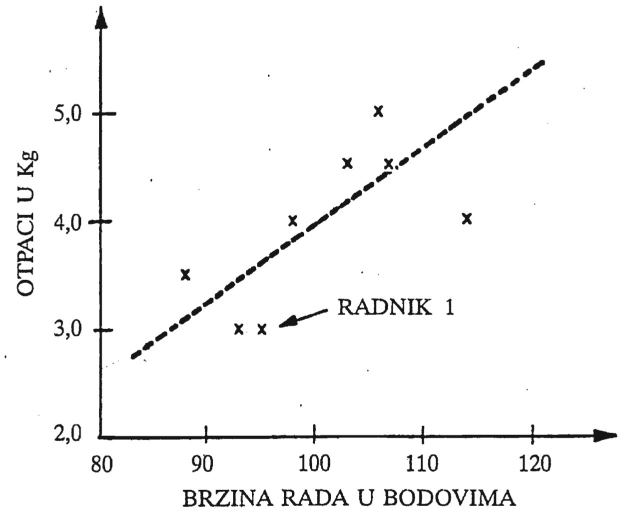


Slika 13.6. Različiti oblici nelinearne povezanosti između dvije varijable

TABLICA 13.1.
ODNOS IZMEĐU BRZINE RADA I KOLIČINE OTPADAKA ZA 8 RADNIKA

Radnik	Brzina rada (u bodovima)	Otpaci (u kg)
1	95	3,0
2	103	4,5
3	88	3,5
4	98	4,0
5	93	3,0
6	107	4,5
7	114	4,0
8	106	5,0

Prikažemo li rezultate s te tablice grafički, dobivamo situaciju prikazanu na slici 13.7.



Slika 13.7. Postoji izvjesna pozitivna korelacija između brzine rada i količine otpadaka: u prosjeku, brži radnici imaju više otpadaka

Kako se iz slike vidi, porastom brzine rada (koja je izražena bodovima: veći broj bodova znači i veću brzinu) u *prosjeku* raste i količina otpadaka, ali povezanost nije tako besprijekorna da bismo iz nečije brzine rada posve točno mogli znati kolika je njegova količina otpadaka. Tako se, na primjer, iz tablice i iz slike vidi da radnici br. 4 i 7 imaju jednaku količinu otpadaka, a brzina rada im se razlikuje. Korelacija je, dakle, pozitivna (veća od 0), ali nije maksimalna (manja je od +1).

N a p o m e n a. Kada se korelacija prikazuje grafički, odnosno u koordinatnom sustavu, obično se na apscisu unose vrijednosti tzv. "nezavisne" varijable, a na ordinatu rezultati "zavisne" varijable. Nezavisna varijabla je ono što možemo samovoljno mijenjati (ako se radi o eksperimentu) ili birati vrijednosti koje nas zanimaju (na primjer, broj ponavljanja kod učenja, broj popušenih cigareta, količina davanja nekog sredstva, brzina rada, itd.), a zavisna varijabla je ono što istraživanjem želimo ustanoviti (na primjer, količina naučenog gradiva, stupanj nekoga zdravstvenog poremećaja, jakost djelovanja nekog sredstva, količina otpadaka, itd.). Naravno, katkada je teško reći što je nezavisna, a što zavisna varijabla: ako nas, primjerice, zanima odnos između količine konzumiranog alkohola i brzine reagiranja, onda je jasno da količina alkohola predstavlja nezavisnu, a brzina reagiranja zavisnu varijablu; ali ako nas zanima odnos između visine i težine, i jedna i druga varijabla mogu biti i nezavisne i zavisne, što ovisi o tome od koje varijable polazimo, tj. o tome što nas zapravo zanima: ako nas zanima kakvu visinu možemo očekivati uz neku određenu težinu, onda je visina zavisna, a težina nezavisna varijabla.

Smisao i osnovne principe izračunavanja koeficijenta korelacije prikazat ćemo postupno i pomoću primjera.

Iz poglavlja "Položaj pojedinog rezultata u grupi" sjećamo se da je uspoređivanje rezultata *različitih* vrsta mjerenja moguće uz pomoć z-vrijednosti. Ako je, na primjer, neki ispitanik u testu A, u kojem je aritmetička sredina 100, a standardna devijacija 10, postigao 110 bodova, on se u tom testu nalazi na $(110 - 100)/10 = 10/10 = 1$ z iznad prosjeka. Ako je u testu B, kome je aritmetička sredina 15, a standardna devijacija 2, isti ispitanik postigao 19 bodova, on je u tom testu 2 standardne devijacije iznad prosjeka ($[17-15]/2 = 2$), pa je prema tome *relativno bolji* u testu B.

Kada bi povezanost između dvije varijable bila besprijekorna, tj. maksimalno moguća, svaki ispitanik bio bi u *obje varijable na jednakim mjestima*, tj. koliko je u jednoj varijabli iznad (ili ispod) prosjeka, *upravo toliko* morao bi biti i u drugoj varijabli iznad (ili ispod) prosjeka: dakle, ako se on u jednoj varijabli nalazi na + 1,75 z, onda bi i u drugoj varijabli morao biti točno na + 1,75 z. Za maksimalnu, ali negativnu korelaciju vrijedilo bi isto, samo što bi z-vrijednosti tog ispitanika morale biti *protivnog predznaka*: upravo toliko koliko je u jednoj varijabli iznad prosjeka, u drugoj varijabli morao bi biti ispod prosjeka.

Kako se vidi, veličina *razlika* između sparenih z-vrijednosti ovisi o visini povezanosti između obje varijable: kad je stupanj *povezanosti* maksimalan, razlike nema; što je povezanost slabija, razlike među z-vrijednostima su veće. Prema tome, neka *prosječna* razlika među svim korespondentnim z-vrijednostima u grupi ispitanika trebala bi nam pružiti neku mjeru o tome koliko su obje varijable povezane. Budući da je aritmetička sredina razlika među z-vrijednostima nužno 0 (nula), to moramo razlike kvadrirati. Prema tome, prosjek sume svih kvadriranih razlika među z-vrijednostima bit će nam neki indeks koji će pokazivati na visinu povezanosti između obje varijable:

$$\frac{\Sigma(z_x - z_y)^2}{N-1}.$$

(N a p o m e n a. $N-1$ u nazivniku potreban je zato što smo i s računali s $N-1$ u nazivniku!) Mali rezultat značit će *visoku* povezanost (jer su male razlike među pripadnim z-vrijednostima), veći rezultat slabiju povezanost (razlike među pripadnim z- vrijednostima su veće), i, konačno, *vrlo velik* rezultat značit će *negativnu* povezanost (jer su razlike između npr. $+2\,z$ i $-2\,z$ velike).

No prosjek, dobiven ovom formulom, nije spretan za interpretaciju, jer se kreće od nule (besprijekorna povezanost) do 4 (besprijekorna *negativna* povezanost). Mnogo je lakše interpretirati stupanj i smjer povezanosti ako polovicu tog prosjeka oduzmemo od jedinice:

$$r = 1 - \frac{1}{2}\frac{\Sigma(z_x - z_y)^2}{N-1}.$$

Taj koeficijent ima ove karakteristike:

1. Vrijednost 0 (nula) označuje da nema nikakve linearne povezanosti između obje varijable.
2. *Veličina* koeficijenta upućuje na *količinu* povezanosti: apsolutno veći broj znači i veću povezanost, a mali broj slabu povezanost.

3. *Predznak* koeficijenta označuje *smjer* povezanosti: + (plus) znači pozitivnu povezanost (tj. porast varijable X povezan je s porastom varijable Y); – (minus) znači negativnu povezanost (tj. porast varijable X povezan je s padom varijable Y i obratno).
4. Najveća moguća pozitivna vrijednost tog koeficijenta iznosi + 1, a najveća negativna vrijednost −1.

Evo jednog primjera. Pretpostavimo da su osmorica ispitanika testirana uz pomoć testova X i Y, i da su postigli rezultate prikazane u tablici 13.2.

TABLICA 13.2.
HIPOTETIČKI REZULTATI OSMORICE ISPITANIKA U TESTOVIMA X i Y

Ispitanici	X	Y	z_x	z_y	$z_x - z_y$
1	7	19	1,5	1,5	0
2	6	17	1	1	0
3	5	15	0,5	0,5	0
4	4	13	0	0	0
5	4	13	0	0	0
6	3	11	-0,5	-0,5	0
7	2	9	-1	-1	0
8	1	7	-1,5	-1,5	0
	$\Sigma X = 32$	$\Sigma Y = 104$		$\Sigma(z_x - z_y) = 0$	
	$\overline{X} = 32/8 = 4$	$\overline{Y} = 104/8 = 13$			
	$s = 2$	$s = 4$			

Uvrstimo li dobivene vrijednosti u prethodnu formulu, dobivamo:

$$r = 1 - 1/2\frac{0}{N-1} = 1 - 0 = 1.$$

Matematički se dosta jednostavno može izvesti (onaj koga to zanima, neka pogleda knjigu J. Welkowitza, R. Ewena, J. Cohena, Introductory Statistics for the Behavioral Sciences, Academic Press, 1971, str. 171 – 172) da je izraz

$$r = 1 - 1/2\frac{\Sigma(z_x - z_y)^2}{N-1}$$

identičan izrazu:

$$r = \frac{\Sigma(z_x \cdot z_y)}{N-1}. \tag{13.1}$$

Ovom je naime formulom lakše pratiti računski postupak pri izračunavanju korelacije, kao i logiku pozitivne ili negativne korelacije: kako se iz formule može zaključiti, visina korelacije je prosječan umnožak između z-vrijednosti obiju varijabli. ("Prosječni" u smislu da je suma podijeljena s $N-1$, a ne s N.) Stoga ćemo rezultate iznesene u tablici 13.2. ponovno obraditi, samo ovaj puta uz pomoć formule 13.1. Rezultati te obrade prikazani su u tablici 13.3.

TABLICA 13.3.
REZULTATI IZ TABLICE 13.2. OBRAĐENI NA DRUGI NAČIN

Ispitanici	X	Y	z_x	z_y	$z_x \cdot z_y$
1	7	19	1,5	1,5	2,25
2	6	17	1,0	1,0	1,00
3	5	15	0,5	0,5	0,25
4	4	13	0	0	0,00
5	4	13	0	0	0,00
6	3	11	−0,5	−0,5	0,25
7	2	9	−1,0	−1,0	1,00
8	1	7	−1,5	−1,5	2,25

$\Sigma X = 32$ $\quad \Sigma Y = 104$ $\quad \Sigma z_x \cdot z_y = 7,00$

$\overline{X}_x = 32/8 = 4$ $\quad \overline{X}_y = 104/8 = 13$ $\quad r = \dfrac{\Sigma z_x \cdot z_y}{N-1}$

$s_x = 2$ $\quad s_y = 4$ $\quad r = \dfrac{7}{7} = +1,0$

Ako je neki rezultat X veći od aritmetičke sredine $\overline{X}_x$, a korespondentni rezultat Y veći od $\overline{X}_y$, onda će i z_x i z_y biti ***pozitivnog*** predznaka, pa prema tome i njihov produkt $(z_x \cdot z_y)$ također pozitivnog predznaka. Ako je neki X ispod $\overline{X}_x$ i korespondentni Y ispod $\overline{X}_y$, onda će i z_x i z_y biti negativnog predznaka, ali njihov produkt će biti pozitivan. Osim toga, suma $z_x \cdot z_y$ bit će maksimalna ako su svi parovi (z_xi z_y) ***numerički jednaki*** (tj. ako, na primjer, jednom rezultatu od + 2 z u varijabli X odgovara rezultat od upravo + 2 z u varijabli Y). Prema tome, suma $z_x z_y$ bit će ***maksimalno pozitivna*** ako su oba člana para $z_x z_y$ (1) numerički jednaki, i (2) istog predznaka. U tom će slučaju koeficijent korelacije biti najveći (i pozitivan), i iznosit će + 1,0. Izračunamo li za naš primjer korelaciju prema formuli 13.1, dobivamo:

$$r = \frac{\Sigma(z_x \cdot z_y)}{N-1} = \frac{7}{7} = 1.$$

Ako su korespondentne z vrijednosti z_x i z_y *pretežno* istog predznaka, ali ne uvijek jednake numeričke vrijednosti, r će biti pozitivan, ali ne više maksimalan, tj. bit će manji od + 1.

Ako među varijablama postoji *potpuno negativni* odnos, to znači da će nekoj vrijednost X, koja je *iznad* $\overline{X}_x$, odgovarati korespondentna vrijednost Y, koja je ispod $\overline{X}_y$, i to upravo toliko ispod $\overline{X}_y$ koliko je X iznad $\overline{X}_x$. To je prikazano u tablici 13.4.

TABLICA 13.4.

Ispitanici	X	Y	z_x	z_y	$z_x \cdot z_y$
1	7	7	1,5	−1,5	−2,25
2	6	9	1,0	−1,0	−1,0
3	5	11	0,5	−0,5	−0,25
4	4	13	0	0	0
5	4	13	0	0	0
6	3	15	−0,5	0,5	−0,25
7	2	17	−1,0	1,0	−1,00
8	1	19	−1,5	1,5	−2,25
$\overline{X}_x = 4$ $\overline{X}_y = 13$	$s_x = 2$ $s_y = 4$	$r = \dfrac{-7}{7} = -1$		$\Sigma z_x z_y = -7,00$	

U tom će slučaju produkt $z_x z_y$ biti *uvijek negativan* (jer množimo plus s minusom i minus s plusom), i to *maksimalno* negativan (jer su korespondentne vrijednosti z_x i z_y jednakih numeričkih vrijednosti), a korelacija će iznositi $r = -1$.

Ako su korespondentne z-vrijednosti pretežno protivnog predznaka, ali ne jednake numeričke vrijednosti, r će biti negativan, ali ne više maksimalno negativan, nego će se kretati između 0 i −1.

Ako prema tom postupku izračunamo korelaciju između brzine rada i količine otpadaka iz prije spomenutog primjera, dobit ćemo ove rezultate:

Radnik	Brz.rada (x)	Otpaci (y)	z_x	z_y	$z_x z_y$
1	95	3,0	−0,65	−1,29	+0,8385
2	103	4,5	+0,29	+0,77	+0,2233
3	88	3,5	−1,47	−0,60	+0,8820
4	98	4,0	−0,29	+0,08	−0,0232
5	93	3,0	−0,88	−1,29	+1,1352
6	107	4,5	+0,76	+0,77	+0,5852
7	114	4,0	+1,59	+0,08	+0,1272
8	106	5,0	+0,65	+1,45	+0,9425

$\Sigma = 804$ $\quad \Sigma = 31,5$ $\quad \Sigma = +4,7107$

$\overline{X} = 100,5$ $\quad \overline{X} = 3,94$ $\quad r = \dfrac{\Sigma z_x z_y}{N-1} = \dfrac{+4,7107}{7}$

$s = 8,5$ $\quad s = 0,73$ $\quad r = +0,67$

13.2. IZRAČUNAVANJE KOEFICIJENTA KORELACIJE r

Takvo izračunavanje koeficijenta korelacije, iako je logički shvatljivo, bilo bi vrlo naporno i dugotrajno u svim slučajevima kada bi N bio veći, i kada aritmetičke sredine i standardne devijacije ne bi bile cijeli brojevi, pa bi trebalo mnogo strpljenja i pažnje da se ispravno izračunaju sve z-vrijednosti. Zato se koeficijent korelacije izračunava različitim drugim postupcima koji su računski jednostavniji i kraći. Ti se postupci naoko dosta međusobno razlikuju (tj. formule su različite), ali se oni — naravno — matematički svi mogu svesti na već opisanu logiku računanja.

Prikazat ćemo tzv. "skraćeni postupak" za izračunavanje koeficijenta r iz negrupiranih rezultata (tj. rezultata koji nisu grupirani u razrede).

Formula za izračunavanje r glasi:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y^2)]}} \tag{13.2}$$

ΣXY = suma umnožaka pojedinih parova rezultata

N = broj parova

ΣX^2 i ΣY^2 = suma kvadriranih rezultata varijable X i varijable Y.

Primjer. Mjereći visinu i težinu 12-godišnjih dječaka u jednoj zagrebačkoj školi dobili smo rezultate prikazane u tablici 13.5. (X = visina, Y = težina); kolika je korelacija između visine i težine kod ove grupe djece? Izračunavanje je provedeno u tablici.

TABLICA 13.5.
IZRAČUNAVANJE KOEFICIJENTA KORELACIJE r

Ispitanik	X	Y	X^2	Y^2	XY
1	139,5	28,4	19 460,25	806,56	3 961,80
2	148,5	37,6	22 052,25	1 413,76	5 583,60
3	138,0	28,5	19 044,00	812,25	3 933,00
4	142,0	39,4	20 164,00	1 552,36	5 594,80
5	141,5	31,8	20 022,25	1 011,24	4 499,70
6	137,0	32,7	18 769,00	1 069,29	4 479,90
7	134,0	26,5	17 956,00	702,25	3 551,00
8	137,0	30,2	18 769,00	912,04	4 137,40
9	143,0	34,4	20 449,00	1 183,36	4 919,20
10	135,5	29,6	18 360,25	876,16	4 010,80
11	142,0	34,0	20 164,00	1 156,00	4 828,00
12	137,0	29,0	18 769,00	841,00	3 973,00
13	147,5	35,8	21 756,25	1 281,64	5 280,50
14	134,0	28,8	17 956,00	829,44	3 859,20
15	144,5	36,2	20 880,25	1 310,44	5 230,90
16	146,5	39,4	21 462,25	1 552,36	5 772,10
17	130,0	26,9	16 900,00	723,61	3 497,00
18	142,5	30,3	20 306,25	918,09	4 317,75
19	148,5	42,5	22 052,25	1 806,25	6 311,25
20	145,5	36,8	21 170,25	1 354,24	5 354,40

$\Sigma X = 2\,814,0$ $\Sigma X^2 = 396\,462,5$ $\Sigma XY = 93\,095,3$

$N = 20$ $\Sigma Y = 658,8$ $\Sigma Y^2 = 22\,112,34$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} =$$

$$= \frac{20 \cdot 93\,095,3 - 2\,814 \cdot 658,8}{\sqrt{[20 \cdot 396\,462,5 - 2\,814^2][20 \cdot 22\,112,34 - 658,8^2]}} =$$

$$= \frac{8\,042,8}{9\,363,5} =$$

$$r = +0,86.$$

Kako se vidi, korelacija između visine i težine prilično je visoka. Ona je u mladih ljudi viša nego kod starijih, jer su mladi ljudi (relativno) proporcionalnije građeni, tj. među njima još nema mnogo debelih ljudi.

Zbog relativno velikih brojeva računanje ovog primjera dosta je mukotrpno. No koeficijent korelacije r neće se ništa promijeniti ako varijabli X ili Y, ili objema varijablama dodamo ili od njih oduzmemo neku konstantu. (To slijedi na temelju toga što se ni standardna devijacija nekih rezultata neće promijeniti ako rezultatima dodamo neku konstantu; a vidjeli smo da se osnovna logika koeficijenta korelacije zasniva na standardnoj devijaciji, tj. na z-vrijednostima.) U našem primjeru mogli bismo varijabli X (visini) oduzeti konstantu 135, a varijabli Y (težini) konstantu 30. U tom slučaju dobili bismo rezultate prikazane u tablici 13.6.

Ako iz rezultata tablice izračunamo r, dobit ćemo:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} =$$

$$= \frac{8\,042,8}{9\,363,5} = +0,86.$$

Kako se vidi, dobili smo potpuno jednaki rezultat kao i prije.

N a p o m e n a. Kada radimo različite statističke račune, pa makar ih radili možda i uz pomoć džepnog elektronskog računala, znadu se (često!) dogoditi pogreške u računu, nastale bilo omaškom pri množenju ili dijeljenju, bilo pri prepisivanju brojeva ili drugih razloga. Stoga račune valja *kontrolirati*. To predstavlja praktički dvostruki posao — naravno pod pretpostavkom da kontrolom dobijemo *isti* rezultat. A ako dobijemo različit rezultat (što se također često događa!), treba račun obaviti ponovno, tj. dotle dok ne dobijemo dva puta jednaki rezultat.

TABLICA 13.6.

Nešto skraćeno izračunavanje koeficijenta korelacije r na osnovi podataka iz tablice 13.5. Skraćenje se sastoji u tome da smo od varijable X odbili 135, a od varijable Y 30.

Ispitanik	X	Y	X^2	Y^2	XY
1	−0,5	−1,6	0,25	2,56	0,8
2	8,5	7,6	72,25	57,76	64,6
3	−2	−1,5	4	2,25	3
4	2	9,4	4	88,36	18,8
5	1,5	1,8	2,25	3,24	2,7
6	−3	2,7	9	7,29	−8,1
7	−6	−3,5	36	12,25	21
8	−3	0,2	9	0,04	−0,6
9	3	4,4	9	19,36	13,2
10	−4,5	−0,4	20,25	0,16	1,8
11	2	4	4	16	8
12	−3	−1	9	1	3
13	7,5	5,8	56,25	33,64	43,5
14	−6	−1,2	36	1,44	7,2
15	4,5	6,2	20,25	38,44	27,9
16	6,5	9,4	42,25	88,36	61,1
17	−10	−3,1	100	9,61	31
18	2,5	0,3	6,25	0,09	0,75
19	8,5	12,5	72,25	156,25	106,25
20	5,5	6,8	30,25	46,24	37,4
Σ	14	58,8	542,5	584,34	443,3

Stoga je korisno znati za neke postupke koji *olakšavaju* kontrolu računa. O nekim kontrolama (pri računanju aritmetičke sredine skraćenim postupkom) bilo je već govora. Pri računanju korelacije taj se postupak kontrole sastoji u tome da se tablici, uz pomoć koje izračunavamo korelaciju, dodaju još dva stupca, i to stupci $(X + Y)$ i $(X + Y)^2$. — Dakle, za one koji žele kontrolu obaviti na ovaj način, evo što treba učiniti: u tablici 13.7. prikazana su dva *dodatna* stupca, izvedena na osnovi podataka tablice 13.5.

Kontrola se sastoji u tome da stupci $\Sigma X + \Sigma Y$ (vidi tablicu 13.5) moraju odgovarati stupcu $\Sigma(X - Y)$, a izraz $\Sigma X^2 + \Sigma Y^2 + 2\Sigma XY$, mora odgovarati stupcu $\Sigma(X + Y)^2$.

Evo toga računa: $2\,814,0 + 658,8 = 3\,472,8$.

$396\,462,5 + 22\,112,34 + 2(93\,095,3) = 604\,765,44$.

Dakle, naš je račun bio točan.

TABLICA 13.7.

Kontrolu računanja korelacije r provodite tako da se stupcima tablice 13.5. dodaju još dva stupca.

Ispitanik X Y X^2 Y^2 XY		$X - Y$	$(X - Y)^2$
O ovom dijelu tablice vidi podatke u tablici 13.5!	1	167,9	28 190,41
	2	186,1	34 633,21
	3	166,5	27 722,25
	4	181,4	32 905,96
	5	173,3	30 032,89
	6	169,7	28 798,09
	7	160,5	25 760,25
	8	167,2	27 955,84
	9	177,4	31 470,76
	10	165,1	27 258,01
	11	176	30 976
	12	166	27 556
	13	183,3	33 598,89
	14	162,8	26 503,84
	15	180,7	32 652,49
	16	185,9	34 558,81
	17	156,9	24 617,61
	18	172,8	29 859,84
	19	191	36 481
	20	182,3	33 233,29
		3 472,8	604 765,44

13.3. TESTIRANJE ZNAČAJNOSTI KOEFICIJENTA KORELACIJE r

Da bismo ustanovili je li dobiveni koeficijent korelacije r značajan (tj. razlikuje li se značajno od nule bez obzira na predznak), možemo izračunati t, i to prema formuli:

$$t = r\frac{\sqrt{(N-2)}}{\sqrt{1-r^2}}. \tag{13.3}$$

U našem primjeru, dakle, dobivamo:

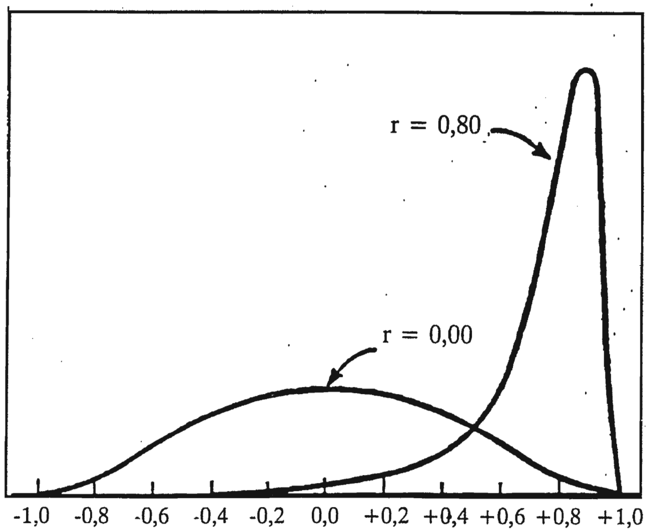
$$t = 0,86 \cdot \frac{\sqrt{18}}{\sqrt{1-0,74}} = 0,86 \cdot \frac{4,24}{\sqrt{0,26}} = 7,15.$$

U ovom slučaju (kod koeficijenta korelacije r) broj stupnjeva slobode računa se: stupnjevi slobode $= N - 2$, pri čemu N znači broj parova. Prema tome, na tablici t-vrijednosti očitamo graničnu vrijednost t na željenoj razini značajnosti. Ako smo odabrali razinu značajnosti od 5%, granična vrijednost t uz 18 stupnjeva slobode iznosi 2,101. Budući da je naš dobiveni t znatno veći, zaključujemo da je ta korelacija statistički značajna.

Testiranje značajnosti korelacije može se provesti i drugačije, tj. očitavanjem s tablice koja nam pokazuje koliko mora najmanje iznositi r uz određeni broj stupnjeva slobode ($N-2$) da bi bio značajan na određenoj razini značajnosti. To je tablica D u Dodatku. Tablica je "dvosmjerna", tj. pokazuje graničnu vrijednost r bez obzira na predznak.

Kako se iz tablice vidi, na razini značajnosti od 5%, a uz 18 stupnjeva slobode (20−2), koeficijent korelacije r mora iznositi najmanje 0,444, a na razini značajnosti od 1% najmanje 0,561. Prema tome, naš r je čak značajan na razini koja je manja od 1% ($P < 0,01$).

Postanak te tablice može se protumačiti vrlo jednostavno, a oni čitaoci koji su proučili pojmove distribucije aritmetičkih sredina uzoraka, ne bi tu više smjeli imati teškoća u razumijevanju: kada između neke dvije populacije podataka ne bi postojala *nikakva* povezanost (na primjer visina ljudi i njihovi rezultati na nekom testu inteligencije), pa kada bismo iz tih dviju populacija po slučaju izvlačili parove rezultata (za svakog čovjeka njegov podatak o visini i o njegovoj inteligenciji), i to uzorke veličine 8 (dakle uvijek po 8 parova), i svaki put nakon izvlačenja za svaki uzorak veličine 8 izračunali koeficijent r, dobili bismo — nakon vrlo velikog broja takvih izvlačenja — simetričnu distribuciju, prikazanu lijevom krivuljom na slici 13.8.



Slika 13.8. Distribucija r koeficijenta korelacije uzoraka veličine 8 parova, kada među populacijama *nema* korelacije (lijeva krivulja), i kada među populacijama postoji korelacija od + 0,80

Ta bi distribucija imala naravno svoju aritmetičku sredinu, koja bi odgovarala pravoj korelaciji između obje varijable, dakle 0 (nula). Kako se iz slike vidi, i *slučajno* bismo mogli dobiti korelacije koje idu sve do gotovo + 1 ili −1. Zato iz tablice D i čitamo da kod 6 stupnjeva slobode (8 parova −2) koeficijent r treba biti najmanje 0,707 (plus ili minus, svejedno), pa da bi bio statistički značajan na razini od 5%.

Što bi veličina uzorka bila veća, ta distribucija korelacija među uzorcima bivala bi naravno sve užom i užom. Kada bi među populacijama *postojala* korelacija, distribucija korelacija uzoraka veličine 8 parova bila bi drugačija, i to drugačija za svaku veličinu korelacije među populacijama. Na slici 13.8. prikazana je desno distribucija korelacija kod uzoraka veličine 8 parova, ako stvarna korelacija među populacijama iznosi 0,80.

13.4. IZRAČUNAVANJE KORELACIJE r NA VELIKIM UZORCIMA

Ako je N veći od 80, izračunavanje koeficijenta korelacije r prema ovom postupku postaje dugotrajno i naporno, te u tom slučaju treba rezultate sažeti u razrede, kao što smo to radili pri skraćenom izračunavanju aritmetičke sredine i standardne devijacije. Tako sažeti rezultati unose se u tzv. "tablicu s dva ulaza", koja služi kao osnova daljnjeg računanja.

Postoji nekoliko različitih načina izračunavanja korelacije grupiranih rezultata, ali oni naravno matematički predstavljaju potpuno ekvivalentne postupke. Mi ćemo protumačiti jedan možda nešto manje uobičajen način, ali je on relativno jednostavan i uključuje već poznate računske operacije.

Primjer. Kod 9 941 žene izmjerena je težina (u funtama) i opseg kukova (u inčima) te je dobiveno 9 941 težina i isto toliko podataka o opsegu kukova. Kad težinu i opseg kukova podijelimo u razrede, te u "tablicu s dvostrukim ulazom" unesemo rezultate, dobivamo situaciju prikazanu na tablici 13.8.

Kad je tablica ispunjena, registriramo frekvenciju varijable X i varijable Y (f_x i f_y) i dalje izvršimo iste operacije koje smo radili pri skraćenom izračunavanju aritmetičke sredine i standardne devijacije: unesemo intervalne udaljenosti svakoga pojedinog razreda od provizorne aritmetičke sredine (dakle izračunavamo x' i y'), pomnožimo te udaljenosti s frekvencijama (fx' i fy') i dobivene rezultate pomnožimo s x' i y' (fx'^2 i fy'^2).

Na jednak način konstruiramo frekvencije dijagonalnih ćelija, i to uvijek tako da nam je smjer dijagonale od kuta gdje obje varijable imaju niske vrijednosti u kut gdje obje varijable imaju visoke vrijednosti (kada bi smjer bio suprotan, dobili bismo koeficijent korelacije pogrešnog predznaka!). Tako, na primjer, u dijagonali koja ide od razreda 170 u varijabli Y do razreda 42 – 43 u varijabli X imamo samo 1 rezultat, u dijagonali 160 → 44 – 45 imamo frekvencije 1, 3, 3, 9, dakle 16 rezultata, itd. Tako dobivamo frekvencije po dijagonali (f_d), koje dalje obrađimo na jednak način kao i frekvencije varijabli X i Y, tj. izračunamo d', fd' i fd'^2.

(N a p o m e n a. Kod d' je svejedno u kojem ćemo smjeru bilježiti pozitivne, a u kojem negativne brojeve.)

Nakon toga izračunamo ove izraze (vidi tablicu!):

TABLICA 13.8
IZRAČUNAVANJE KOEFICIJENATA KORELACIJE r
IZ GRUPIRANIH REZULTATA

Opseg kukova X

Težina tijela Y

	32-33	34-35	36-37	38-39	40-41	42-43	44-45	46-47	48-49	50-51	f_x	y'	fy'	fy'^2
220						1	9	13	20	10	53	8	424	3.392
210						3	15	30	18	6	72	7	504	3.528
200					3	13	36	33	20	2	107	6	642	3.852
190					8	51	74	55	14	1	203	5	1.015	5.075
180			1	1	29	112	126	37	7	1	314	4	1.256	5.024
170			1	11	111	204	81	10			418	3	1.254	3.762
160			2	63	286	257	61	6			675	2	1.350	2.700
150		1	21	242	450	135	16				865	1	865	865
140		5	134	683	434	48	1				1.305	0		
130		29	565	936	156	9					1.695	−1	−1.695	1.695
120	10	252	1.225	390	14						1.891	−2	−3.782	7.564
110	30	786	636	70	2						1.524	−3	−4.572	13.716
100	169	445	63	2							679	−4	−2.716	10.864
90	102	38									140	−5	−700	3.500
f_x	311	1.556	2.648	2.398	1.493	833	419	184	79	20	9.941		Σ = −6.155	Σ = 65.537
x'	−3	−2	−1	0	1	2	3	4	5	6				
fx'	−933	−3.112	−2.648		1.493	1.666	1.257	736	395	120	Σ = 1.026			
fx'^2	2.799	6.224	2.648		1.493	3.332	3.771	2.944	1.975	720	Σ = 25.906			

f_d	d'	fd'	fd'^2
1	−6	−6	36
16	−5	−80	400
51	−4	−204	816
180	−3	−540	1.620
447	−2	−894	1.788
1.112	−1	−1.112	1.112
2.471	0		
3.764	1	3.764	3.764
1.800	2	3.600	7.200
195	3	585	1.755
4	4	16	64
		Σ = 5.129	Σ = 18.555

$$A = \Sigma(fy'^2) - \frac{(\Sigma fy')^2}{N}$$

$$B = \Sigma(fx'^2) - \frac{(\Sigma fx')^2}{N}$$

$$C = \Sigma(fd'^2) - \frac{(\Sigma fd')^2}{N}$$

$$A = \Sigma(fy'^2) - \frac{(\Sigma fy')^2}{N}$$

$$B = \Sigma(fx'^2) - \frac{(\Sigma fx')^2}{N}$$

$$C = \Sigma(fd'^2) - \frac{(\Sigma fd')^2}{N}$$

dobivene rezultate uvrstimo u formulu

$$r = \frac{A + B - C}{2\sqrt{(AB)}}. \tag{13.4}$$

Prema tome, u našem primjeru dobili bismo ove vrijednosti:

$$A = 65\,537 - \frac{37\,884\,025}{9\,941} = 61\,726,11$$

$$B = 25\,906 - \frac{1\,052\,676}{9\,941} = 25\,800,11$$

$$C = 18\,555 - \frac{26\,306\,641}{9\,941} = 15\,908,72$$

$$r = \frac{A + B - C}{2\sqrt{(AB)}}$$

$$= \frac{61\,726,11 + 25\,800,11 - 15\,908,72}{2 \cdot \sqrt{61\,726,11 \cdot 25\,800,11}}$$

$$= \frac{71\,617,50}{2 \cdot \sqrt{1\,592\,539,553}} = \frac{71\,617,5}{79\,813} = +0,897.$$

13.5. IZRAČUNAVANJE RANG KORELACIJE "Ro" (ρ)

Ako su jedna ili obje varijable dane u *rangu*, tj. rezultati nisu mjerene vrijednosti, već su dani samo u redoslijedu, računa se tzv. "rang-korelacija", tj. korelacija među rangovima. Kod rangova ne znamo stvarne rezlike među pojedinim rezultatima, nego jedino razlike u rangu. Ako, na primjer, znamo da je trkač A stigao na cilj prvi, trkač B drugi i trkač C treći, oni se međusobno razlikuju *za jedan rang* (A prema B i B prema C) ili *dva ranga* (A prema C), no stvarna razlika između, primjerice, prvog i drugog može biti mnogo veća nego između drugog i trećeg, jer su drugi i treći možda na cilj stigli gotovo zajedno, dok je prvi trkač stigao mnogo prije njih. Čim dakle ne znamo stvarne razlike među pojedinim rezultatima, nikakvu *preciznu korelaciju ne možemo izračunati*, a takvu preciznu korelaciju predstavlja npr. koeficijent r. Čitalac će se sjetiti da se koeficijent r temelji na umnošku z-vrijednosti, dakle upravo na *udaljenosti* pojedinog rezultata od aritmetičke sredine.

Kao što rekosmo, kod rang-korelacije uzimamo u račun samo rangove i razlike među njima, a te su razlike podjednake: razlika između ranga 2 i 3 jednaka je razlici između rangova 6 i 7, makar su *stvarne* razlike među rezultatima možda veoma različite. Za razliku od Pearsonova r koeficijenta, kod rang-korelacija *nije potrebno* da varijable budu u linearnom odnosu.

Najpoznatija rang-korelacija je korelacija ρ (čitaj: ro), koja se izračunava prema formuli:

$$Ro = 1 - \frac{6 \cdot \Sigma D^2}{N(N^2 - 1)} \tag{13.5}$$

pri čemu D znači diferenciju između rangova prve i druge varijable.

Rang-korelacija daje samo *približnu* indikaciju asocijacije između dvije varijable i opravdano ju je upotrijebiti samo onda ako se ne može izračunati r-korelacija.

Primjer. 15 dječaka testirano je jednim testom inteligencije, a njihov nastavnik poredao ih je po rangu uspjeha na prijamnom ispitu. Kolika je korelacija između uspjeha u testu i uspjeha na prijamnom ispitu?

Na ispitu i u testu dječaci su postigli ove rezultate:

Dječak:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Rang na prijamnom ispitu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bodovi u testu	22	19	6	18	20	16	11	9	15	12	10	7	13	12	8

Treba, dakle, poredati dječake po rangu i u testu inteligencije. Ispitanici koji su postigli jednaki rezultat, dobivaju i jednaki rang, koji se izračunava tako da se zbroje rangovi koje bi oni zauzimali, i podijele brojem tih rangova. Na primjer, ako ispitanik X treba da dođe u 7. rang, ali i ispitanik Y ima jednaki rezultat, onda će svaki od njih zauzeti $(7 + 8)/2 = 7{,}5$-i rang. Ako npr. iza 9 ranga slijede 3 ispitanika s jednakim rezultatom, svaki od njih će zauzeti $(10 + 11 + 12)/3 = 11$–i rang. (Pazi, idući ispitanik iza ove trojice zauzeti će *trinaesti* rang!)

Najbolju kontrolu, da u rangiranju nismo preskočili neki rang, odnosno koji rang računali dva ili više puta, postižemo tako da zbrojimo sve rangove. Njihova suma mora biti: $\frac{N(N+1)}{2}$.

Postupak izračunavanja rang-korelacije prikazan je u tablici 13.9.

TABLICA 13.9.
IZRAČUNAVANJE RANG KORELACIJE ρ

Dječak	Rang prijamn. ispita	Rang u testu	D	D^2
A	1	1	0	0
B	2	3	1	1
C	3	15	12	144
D	4	4	0	0
E	5	2	−3	9
F	6	5	−1	1
G	7	10	3	9
H	8	12	4	16
I	9	6	−3	9
J	10	8,5	−1,5	2,25
K	11	11	0	0
L	12	14	2	4
M	13	7	−6	36
N	14	8,5	−5,5	30,25
O	15	13	−2	4

$\Sigma D^2 = 265,50$

Uvrstimo li dobivene vrijednosti u formulu (13.5), dobivamo:

$$Ro = 1 - \frac{6 \cdot 265,5}{15 \cdot 224} = 1 - 0,474 = +0,526.$$

Zbog jednostavnijeg i kraćeg računanja može se pri izračunavanju korelacije koristiti tablica 13.10. u kojoj se nalaze recipročne vrijednosti izraza $6/N\ (N^2 - 1)$. Da bi se dobio ρ, potrebno je izraz ΣD^2 podijeliti brojem u tablici, i rezultat odbiti od 1. U našem primjeru vrijednost u tablici iznosi 560 ($N = 15$), pa, prema tome, dobivamo:

$$Ro = 1 - \frac{265,5}{560} = 1 - 0,474 = 0,526.$$

Ako postoji relativno velik broj zajedničkih rangova (tzv. "vezanih" rangova), tj. ako ima mnogo slučajeva da dva ili više ispitanika zauzimaju isti rang, potrebno je izvršiti određenu korekturu u računu. (Neki statističari tvrde da je to potrebno učiniti ako je više od 1/4 svih rezultata vezano u zajedničke rangove. No i to se čini prestrogim zahtjevom).

Ne izvršimo li korekturu, to može dovesti u krajnjim slučajevima do besmislenog rezultata. Evo dokaza: pretpostavimo da jedna varijabla daje rezultate u rangovima, koji idu od 1 do N, a u drugoj su varijabli *svi rezultati jednaki*, pa prema tome svi zauzimaju *isti rang*. Ako su svi rezultati jednaki, njihov srednji rang iznosi $(N+1)/2$. Označimo li rezultate tim prosječnim rangom, i izračunamo li koeficijent ρ, dobit ćemo korelaciju od $+0,5$, što je očito besmislica, jer je korelacija zapravo 0. (Ako u tom slučaju izračunamo r, dobit ćemo $r = 0$.) Taj slučaj očito pokazuje da "vezani" rangovi umjetno povećavaju rang-korelaciju: što je vezanih rangova više, to je distorzija rezultata veća.

TABLICA 13.10.

Da bi se dobio ρ, dobiveni izraz ΣD^2 treba podijeliti odgovarajućim brojem u tablici i dobiveni rezultat oduzeti od 1

N		N		N	
10	165	30	4 495	50	20 825
11	220	31	4 960	51	22 100
12	286	32	5 456	52	23 436
13	364	33	5 984	53	24 802
14	455	34	6 545	54	26 235
15	560	35	7 140	55	27 720
16	680	36	7 770	56	29 260
17	816	37	8 436	57	30 856
18	969	38	9 139	58	32 509
19	1 140	39	9 880	59	34 220
20	1 330	40	10 660	60	35 990
21	1 540	41	11 480	61	37 820
22	1 771	42	12 341	62	39 711
23	2 024	43	13 244	63	41 664
24	2 300	44	14 190	64	43 680
25	2 600	45	15 180	65	45 760
26	2 925	46	16 215	66	47 905
27	3 276	47	17 296	67	50 116
28	3 654	48	18 424	58	52 394
29	4 060	49	19 600	69	54 740

Stoga, kao što je već rečeno, u takvim slučajevima primjenjujemo korekturu u računu. Postoji nekoliko predloženih načina korekture, no neki od njih u pojedinim slučajevima dovode do apsurdnih rezultata, te sam odlučio prikazati jedan od tih načina koji se čini relativno najboljim (sličan je način u nas predložio i Krković u "Alternativna formula za izračunavanje koeficijenta korelacije među rangovima", "Revija za psihologiju", Vol 3, br. 1 – 2, 1973, str. 31).

Da bi se korektura mogla provesti, potrebno je najprije izvesti neka preračunavanja za zajedničke rangove. Preračunavanje se sastoji u tome da se za sve zajedničke ("vezane") rangove izračuna izraz "A", koji je računski definiran ovako:

$$A = \frac{k(k^2-1)}{12} \tag{13.6}$$

pri čemu k = broj rezultata vezanih u isti rang.

Ako su npr. 2 rezultata vezana u isti rang, $A = 2(4-1)/12 = 0,5$; ako su vezana 3 rezultata, $A = 3(9-1)/12 = 2$, itd. Pošto smo izračunali sve A za jednu i drugu varijablu, zbrojimo sve A vrijednosti jedne varijable (ΣA_1) i sve A vrijednosti druge varijable (ΣA_2), i rang-korelaciju izračunamo prema formuli:

$$\text{Korig. } Ro = \frac{\frac{N(N^2-1)}{6} - \Sigma D^2 - \Sigma A_1 - \Sigma A_2}{\sqrt{\left[\frac{N(N^2-1)}{6} - 2\Sigma A_1\right]\left[\frac{N(N^2-1)}{6} - 2\Sigma A_2\right]}}. \tag{13.7}$$

N a p o m e n a. Ako *nema* zajedničkih rangova, A_1 i A_2 su nule, i formula (13.7) se svodi na izvornu formulu 13.5:

$$\frac{\frac{N(N^2-1)}{6} - \Sigma D^2 - 0 - 0}{\sqrt{\left[\frac{N(N^2-1)}{6} - 2\cdot 0\right]\left[\frac{N(N^2-1)}{6} - 2\cdot 0\right]}} = \frac{N(N^2-1) - 6\Sigma D^2}{N(N^2-1)} =$$

$$= \frac{N(N^2-1)}{N(N^2-1)} - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6\Sigma D^2}{N(N^2-1)}.$$

Niže je naveden jedan izmišljeni primjer u kojem svi rezultati jedne varijable zauzimaju isti rang:

X	Y	Rang X	Rang Y	D	D^2
9	10	1	3	2	4
8	10	2	3	1	1
7	10	3	3	0	0
6	10	4	3	1	1
5	10	5	3	2	4
					$\Sigma D^2 = 10$

$$Ro = 1 - \frac{6\cdot 10}{5\cdot 24} = 1 - 0,5 = 0,5.$$

Kako se vidi, *Ro* bez korekture iznosi 0,50, što — kako je već upozoreno — nije logično. Provedemo li korekturu prema opisanom postupku, dobit ćemo:

"A" za varijablu $X = 0$

"A" za varijablu $Y = \frac{5\cdot 24}{12} = 10$.

Prema tome, već u brojniku formule 13. 7. dobivamo: $20 - 10 - 0 - 10 = 0$, dakle rang-korelacija je 0.

Ako situacija nije toliko ekstremna kao u iznesenom primjeru, upotrebom formule s korekturom za zajedničke (vezane) rangove koeficijent rang-korelacije samo *nešto malo* se smanjuje, pa se primjena korekture zapravo — *nije isplatila* (jer male razlike u koeficijentu, koji je ionako dosta gruba aproksimacija stupnja povezanosti, nemaju praktične vrijednosti.

Evo nekoliko primjera *Ro* koeficijenta korelacija s dosta vezanih rangova. Navedeni su koeficijenti bez korekture i s korekturom:

N	Broj vez. rangova	*Ro* bez kor.	*Ro* sa korekt.
4	$2\cdot 2$	0,95	0,949
8	$2\cdot 4$	0,88	0,87
8	$2\cdot 4$	0,31	0,22
16	$2\cdot 7$	0,876	0,868
16	$2\cdot 7$	0,218	0,108
20	$5\cdot 3$	0,984	0,983
20	$2\cdot 3$	0,678	0,668
20	$2\cdot 10$	0,876	0,867
20	$2\cdot 10$	0,199	0,087

Kako vidimo, promjene su vrlo male, pogotovo kod visokih koeficijenata. Kao osnovno pravilo vrijedi ovo: razlika između korigiranog i nekorigiranog koeficijenta postaje to veća što su razlike među rangovima jedne i druge varijable veće, dakle što je korelacija niža.

– – – –

Rang-korelacija ρ nije zapravo ništa drugo nego korelacija r primijenjena na *rangove*! Evo dokaza:

Uzmimo da su dva nastavnika ovako rangirali svojih 10 učenika:

Učenik	Rang 1. nastavnika	Rang 2. nastavnika	D	D^2
A	3	3	0	0
B	5	7	2	4
C	7	4	3	9
D	8	10	2	4
E	6	5	1	1
F	1	2	1	1
G	10	8	2	4
H	4	6	2	4
I	2	1	1	1
J	9	9	0	0
				$\Sigma D^2 = 28,$

$$Ro = 1 - \frac{6D^2}{N(N^2-1)} = 1 - \frac{168}{990} = 1 - 0,17 = 0,83.$$

Kada bismo na tim istim rezultatima izračunali koeficijent korelacije r, dobili bismo ovo (tko ne vjeruje, neka provjeri!):

$\Sigma X = 55$; $\Sigma Y = 55$; $\Sigma X^2 = 385$; $\Sigma Y^2 = 385$; $\Sigma XY = 371$.

Unesemo li te podatke u formulu 13.2, dobivamo:

$$r = \frac{(10 \cdot 371) - (55 \cdot 55)}{\sqrt{[(10 \cdot 385) - 55^2][(10 \cdot 385) - 55^2]}} = \frac{685}{825} = 0,83.$$

Kako se vidi, rezultat je jednak.

Ako imamo "vezanih" (zajedničkih) rangova, doći će do određene razlike između rezultata izračunatog uz pomoć formule za r i formule za ρ, tj. dobiveni r točno će odgovarati rezultatu samo ako za izračunavanje ρ upotrijebimo formulu (13.7), tj. formulu za *korigiranu* rang-korelaciju. (Iz toga proizlazi koristan praktični zaključak: ako raspolažete džepnim kalkulatorom, koji izračunava r, računajte pomoću njega i ρ korelaciju, jer ćete dobiti ρ koji je već korigiran s obzirom na vezane rangove.)

No, razumljivo je da tako izračunani r *nema ono isto značenje* kao r izračunan na *mjerenim* vrijednostima neke intervalne ili omjerne skale (o intervalnim i omjernim skalama bit će riječi u posebnom poglavlju!). Drugim riječima, takav r je znatno manje precizan, pa ga zato i ne nazivamo r-koeficijentom, nego je dobio drugo ime, tj. rang-korelacija. Dobio je ujedno i mnogo jednostavniju formulu, koja, naravno, više ne mora voditi računa o relativnom položaju pojedinog rezultata među drugim rezultatima (dakle o z- vrijednosti nekog rezultata), nego vodi računa jedino o razlikama među rangovima.

Upotreba formule za r na rangovnim rezultatima odgovarala bi donekle situaciji u kojoj biste se našli da kao iskusan vozač automobila sjednete za volan nekoga dječjeg automobila na "autodromu" te se u vožnji *ponašate* vrlo precizno: prebacujete mjenjač brzina (a on je u vozilu samo ukras i ne služi ničemu), precizno okrećete volan (a on možda samo reagira na velike pokrete), dodajete i oduzimate "gas" (iako on možda radi samo po principu "sve ili ništa") itd., dakle ponašate se kao da ozbiljno vozite. Tako i koeficijent korelacije r uzima u obzir finese koje u rezultatima ne postoje, i stoga, naravno, u tom slučaju niti ne može biti ništa precizniја mjera povezanosti od jednostavne mjere koeficijenta rang-korelacije.

Pearsonov koeficijent korelacije r — osim što pretpostavlja mjerene vrijednosti, a ne rangove — pretpostavlja i *simetričnost* raspodjele obiju varijabli (čak, prema nekim statističarima, pretpostavlja *normalnu* raspodjelu varijabli). Ako je neka od raspodjela asimetrična, koeficijent r (koji vodi računa o z- vrijednostima, a one imaju smisla samo kod manje-više normalne raspodjele!) dao bi nam iskrivljene rezultate i u tim slučajevima valja upotrijebiti koeficijent rang-korelacije. Ako tako učinimo, onda, dakako, treba prethodno *rangirati* postojeće rezultate, i tek na rangovima izvršiti računanje. Rangiranje se može provesti u bilo kojem smjeru, ali mora biti provedeno na isti način u obje varijable: ako *najvišem* rezultatu u varijabli X damo rang 1, moramo i u varijabli Y rang 1 dati najvišem rezultatu. Ako, naprotiv, rang 1 dajemo najnižem rezultatu, moramo to učiniti jednako u obje varijable. Nepridržavanje tih pravila neće doduše djelovati na *visinu* koeficijenta rang-korelacije, ali će izmijeniti *predznak*, što nas može navesti na potpuno pogrešan zaključak!

N a p o m e n a: Treba ipak upozoriti da ima (doduše rijetkih) situacija, u kojima *Ro* koeficijent korelacije ima *prednost* pred r koeficijentom! To je slučaj onda, ako kod *mjerenih* rezultata postoji neki *veoma ekstreman* rezultat. U jednom dijagramu raspršenja (skater-dijagramu), rezultati dviju varijabli mogu tvoriti površinu koja *očito* upućuje na *negativnu* korelaciju, ali ako se jedan izričito ekstreman rezultat nalazi na *visokim* vrijednostima varijable X i varijable Y, on može "pretvoriti" r koeficijent korelacije u — *pozitivnu* povezanost!

13.6. TESTIRANJE ZNAČAJNOSTI RANG KORELACIJE Ro

Budući da rang-korelacija predstavlja samo aproksimaciju koeficijenta povezanosti između dvije varijable, to neki statističari smatraju da nema pravog smisla izračunavati značajnost rang-korelacije ρ. Ipak, ako N nije vrlo mali ($N < 10$), značajnost rang-korelacije može se izračunati prema formuli:

$$t = Ro\sqrt{\frac{N-2}{1-\rho^2}}. \qquad (13.8)$$

U našem primjeru iz tablice 13.9, dakle, dobivamo:

$$t = 0,526\sqrt{\frac{13}{0,723324}} = 2,24.$$

Budući da kod ρ korelacije broj stupnjeva slobode iznosi $N-2$, iz t-tablice uz

13 stupnjeva slobode vidimo da na razini značajnosti od 5% granična vrijednost t iznosi 2,16, pa, prema tome, zaključujemo da je naš ρ statistički značajan.

Kao i kod r korelacije, i kod rang-korelacije možemo također iz tablice očitati koliko najmanji mora biti ρ (uz određeni N) da bi bio statistički značajan (tj. da bi se značajno razlikovao od nule). Najmanje vrijednosti koeficijenta rang-korelacije, potrebne da bi, uz određeni N, ρ bio značajan, označene su u tablici E1 u Dodatku. Značajnost je navedena na razini značajnosti od 5% i 1%.

Nastanak tablice E1 relativno je lako protumačiti. Zamislimo da je $N = 2$. Ako u osnovi (u populaciji) nema nikakve korelacije između dvije varijable, ipak kod $N = 2$ možemo pukim *slučajem* dobiti ove kombinacije:

Rang varijable X	Mogući rangovi varijable Y	
1	1	2
2	2	1
Ro korelacija =	+1	−1

Kako se vidi, kod $N = 2$ uopće ne možemo dobiti $\rho = 0$, nego imamo 50% vjerojatnosti da ćemo slučajno dobiti $\rho = +1$, i također 50% vjerojatnosti da će ρ biti −1. Iz toga proizlazi da ρ korelacija kod $N = 2$ nipošto ne može biti statistički značajna. Ako je $N = 3$, onda imamo 6 mogućih slučajnih kombinacija rangova prema "zakonu permutacije", tj. $3 \cdot 2 \cdot 1 = 6$. To se, kako znamo (vidi 3. poglavlje), bilježi izrazom 3! (a čita se "tri faktorijel"). Evo tih 6 mogućnosti:

Rang varijable X	Mogući rangovi varijable Y					
1	1	1	2	2	3	3
2	2	3	1	3	1	2
3	3	2	3	1	2	1
Ro =	+1,0	+0,5	+0,5	−0,5	−0,5	−1,0

Kako se vidi, vjerojatnost da će rang-korelacija (ako među populacijama nema korelacije) *slučajno* biti + 1, iznosi 1/6, odnosno oko 16,7%. No kako mi radimo s "dvosmjernim" tablicama, vjerojatnost da će ρ koeficijent slučajno biti +1 ili −1 (dakle po veličini 1) je 2/6 ili oko 33,3%. Tek kod $N = 5$ imamo $5! = 120$ mogućih kombinacija, od kojih dvije sačinjavaju korelaciju visine 1, pa možemo tada korelaciju od 1 smatrati statistički značajnom na razini $P < 0,05$.

13.7. KENDALLOV KOEFICIJENT RANG KORELACIJE "Tau" (τ)

Danas se mnogo upotrebljava još jedan koeficijent rang korelacije, tj. Kendallov *Tau* (τ). Njegovo izračunavanje nešto je kompleksnije od izračunavanja *Ro* koeficijenta, a najjednostavnije se provodi na način koji ćemo opisati.

Ako rangove varijable X i Y ispišemo jedne ispod drugih, i to tako da varijablu X *poredamo pravilno po rangu,* dobit ćemo (pod pretpostavkom da se radi samo o pet rangova) recimo, ovakav rezultat:

Rang X:	1	2	3	4	5
Rang Y:	1	4	3	5	2

Da bismo izračunali rang korelaciju, potrebno je prethodno izračunati izraz S, a on se računa tako da se u Y varijabli svaki rang usporedi s ostalima, koji su desno od njega, pa ako je onaj drugi veći, onda se upiše +1, a ako je manji, označi se sa −1. Konkretno, naš prvi rang u varijabli Y je 1. Idući rang je 4, pa stoga upišemo oznaku +1. Iza toga je rang 3, dakle opet +1, pa 5 (također +1), i konačno rang 2 (+1). Nakon što smo tako broj 1 usporedili sa svima ostalima, prelazimo na idući rang, tj. na broj 4. Kad ga usporedimo s idućim, tj. s brojem 3, stavljamo znak −1 (jer 3 je manje od 4), kod usporedbe s 5 oznaka je +1, i kod usporedbe s 2 oznaka je −1. Na redu je rang 3, itd, te konačno imamo sve podatke za izračunavanje izraza S:

$$+1 \quad +1 \quad +1 \quad +1 \quad -1 \quad +1 \quad -1 \quad +1 \quad -1 \quad -1 \quad = 2$$

(N a p o m e n a: budući da imamo uspoređivanje svakog ranga sa svakim, postoji $N(N-1)/2$ usporedbi, što u našem slučaju iznosi 10.) Kada bi rang u varijabli Y bio potpuno pravilan, i kretao se (kao i u varijabli X) od 1 do 5, onda bismo imali 10 puta oznaku +1, što znači da *maksimalna* vrijednost S iznosi $N(N-1)/2$.

Koeficijent *Tau* je odnos između *nađene* vrijednosti S i *maksimalne* vrijednosti S, dakle:

$$Tau = \frac{S}{1/2N(N-1)} \qquad (13.9)$$

U našem slučaju *Tau* dakle iznosi $2/10 = 0,2$

Tau može poprimiti sve vrijednosti od −1 do +1, što je potpuno logično: kod −1 izračunati S je maksimalan, dakle $N(N-1)/2$, ali negativan, a kod +1, tj. kada je i rang u varijabli Y identičan rangu u varijabli X, S je maksimalan i pozitivan.

Kod *vezanih* (zajedničkih) rangova postupa se na jednak način kao što smo radili i kod *Ro* koeficijenta korelacije, tj. svi elementi istog ranga dobivaju jedan *prosječni* rang. Usporedba dvaju jednakih rangova u Y varijabli donosi rezultat 0, a ako postoje vezani rangovi u X varijabli, to povlači za sobom vrijednost 0 za korespondentne usporedbe rangova u varijabli Y.

Evo dvaju primjera:

(a)	X	1	2	3	4	5	6
	Y	2	3	4,5	4,5	1	6

$$S = +1 \ +1 \ +1 \ -1 \ +1 \ +1 \ +1 \ -1 \ +1 \ 0 \ -1 \ +1 \ -1 \ +1 \ +1 \ = 6$$

(b)	X	1,5	1,5	3	5	5	5
	Y	2	3	4,5	4,5	1	6

$$S = 0 \ +1 \ +1 \ -1 \ +1 \ +1 \ +1 \ -1 \ +1 \ 0 \ -1 \ +1 \ 0 \ 0 \ 0 \ = 4$$

Formula za izračunavanje *Tau* kod vezanih rangova nešto je drugačija, i ona glasi:

$$Tau = \frac{S}{\sqrt{[1/2\ N(N-1) - T_x][1/2\ N(N-1) - T_y]}} \qquad (13.10)$$

pri čemu $T_x = 1/2$ sume $t_x(t_x - 1)$, a $T_y = 1/2$ sume $t_y(t_y - 1)$, a t_{xy} i t označuju *broj* rezultata vezanih u jedan rang. (N a p o m e n a: ako nema vezanih rangova, očito je da se formula /13.10 / svodi na formulu /13.9/).

U primjeru (b) za varijablu X imamo najprije 2 rezultata u vezanom rangu (1,5 i 1,5) , nakon toga 3 rezultata u vezanom rangu (5, 5, 5).

Dakle:

$$T_x = 1/2 \;\; 2\,(2-1) + 3\,(3-1) = 1/2 \;\; 8 = 4$$

U varijabli Y imamo:

$$T_y = 1/2 \;\; 2\,(2-1) = 1$$

Dakle *Tau* iznosi:

$$= \frac{4}{\sqrt{[1/2 \;\; 6\,(6-1) \; - 4][1/2 \;\; 6\,(6-1) \; - 1]}} =$$

$$= \frac{4}{\sqrt{(15-4)\,(15-1)}} = \frac{4}{\sqrt{154}} = 0,32$$

13.8. TESTIRANJE ZNAČAJNOSTI KOEFICIJENTA KORELACIJE *Tau*

Testiranje *značajnosti* koeficijenta *Tau* lako se provodi kod računa bez vezanih rangova, a znatno je kompliciranije kod vezanih rangova.

Ako je N veći od 10, i ako nema vezanih rangova testiranje se provodi tako da se najprije testira statistička značajnost izraza S. To se izvodi pomoću donje formule (kojom se procjenjuje *varijanca* distribucije uzoraka izraza S):

$$Varijanca\ S = \frac{N(N-1)\,(2N+5)}{18} \qquad (13.11)$$

Nakon toga se apsolutna vrijednost prethodno dobivenog S smanji za 1, i podijeli *standardnom devijacijom* S (dakle korijenom formule 13.11):

$$z = \frac{|S-1|}{\sqrt{N(N-1)\,(2N+5)\,/18}} \qquad (13.12)$$

Kao što iz naziva formule vidimo, radi se o jednoj z-vrijednosti, pa prema tome i o *normalnoj* raspodjeli. Iz toga slijedi, ako je z veći od 1,96, smatramo da je *Tau* statistički značajan na nivou značajnosti nižem od 5%, a ako je z veći od 2,58, onda je *Tau* značajan na nivou nižem od 1%. U našem primjeru ($S = 2$) imamo dakle ovu situaciju:

$$z = \frac{2-1}{\sqrt{5\,(5-1)\,(10+5)/18}} = \frac{1}{\sqrt{20 \cdot 0,83}} = \frac{1}{4,07} = 0,25$$

Korelacija nije dakle statistički značajna (što je i logično, jer na tako malom N korelacija iznosi samo 0,2).

Kada bi kod ovih 5 rezultata postojala maksimalna korelacija ($Tau = +1$), onda bi — kako znamo — izraz S bio maksimalan, tj. $N(N-1)/2$, dakle 10. U tom slučaju značajnost bi bila:

$$z = 9/4,08 = 2,20$$

pa bi takva korelacija bila statistički značajna na nivou od 5%. Ako pogledamo Tablicu E u Dodatku, koja nam pokazuje granične vrijednosti *Ro* koef. korelacije, vidjet ćemo da kod $N = 5$ *Ro* mora iznositi 1, da bi bio statistički značajan. Kod *Tau* imamo dakle jednaku situaciju, što je i logično.

Međutim, opisani postupak provodi se kada je N veći od 10, a naš prvi primjer sastoji se samo od 5 rangova. Ako je N manji od 10, koriste se tablice, koje su u različitim knjigama različitog oblika. Između njih izabrali smo jednu koja spada među najjednostavnije i najrazumljivije, i to samo za slučajeve kada nema vezanih rangova. To je tablica E2. (Za slučajeve vezanih rangova tablice su velike po opsegu, i nismo ih ovdje navodili; čitalac, koji je posebno zainteresiran za te tablice, može ih naći u knjizi Harsbarger,T.R: Introductory statistics: A decision map /vidi literaturu). U tablici E2 u Dodatku navedene su granične (kritične) vrijednosti S (dakle brojnika formule Tau koeficijenta)za uzorke veličine 3 do 10. Ako je naš dobiveni S jednak ili veći od onoga navedenog u tablici, *Tau* je statistički značajan (na nivou navedenom na vrhu tablice) Primijenimo li ovu tablicu na naš prvi primjer sa pet rangova (u kojem smo dobili $Tau = 0,2$), vidimo da kod $N = 5$ S za jednosmjerni test mora biti najmanje 8 (a za dvosmjerni test 10), pa iz toga proizlazi da naš *Tau* nije statistički značajan.

Uostalom, već iz iskustva s *Ro* koeficijentom znamo da tako niska korelacija na tako malom uzorku ne može biti statistički značajna. — Ako međutim postoje vezani rangovi, izračunavanje značajnosti *Tau* koeficijenta znatno je kompliciranije. U dilemi da li u ovaj udžbenik unijeti i taj postupak, ili ga izostaviti,sama *formula* za testiranje značajnosti kod vezanih rangova nametnula je odluku: formula je dugačka četiri puna reda! Stoga je odlučeno da to područje ne uđe u tekst, a ako je nekom čitaocu upotreba te formule posebno važna, onda može podatke o njoj naći u knjizi: Ferguson: Statistical analaysis in psychology and education, McGraw-Hill, 1966, str. 223.

13.9. USPOREDBA IZMEĐU *Tau* I *Ro* KOEFICIJENATA

Ako se upitamo koja je razlika između rang korelacije *Ro* i rang korelacije *Tau*, možemo reći ovo: što se tiče krajnjeg rezultata računa na istom materijalu, razlike praktički uvijek postoje i gotovo uvijek je izraz *Tau niži* od izraza *Ro*. Statističari smatraju *Tau* boljim od Ro, jer je aproksimacija normalne raspodjele (kod testiranja statističke značajnosti koeficijenta korelacije) mnogo bolja kod *Tau* koeficijenta. No, s druge strane, kod *Ro* koeficijenta posjedujemo *tablice* koje nam pokazuju graničnu vrijednost koeficijenta, značajnu na nekom određenom nivou značajnosti, pa nam aproksimacija normalnosti nije niti potrebna!

Ako je zbog posebnih razloga važno naglasiti *velike* razlike među rangovima, onda je *Ro* koeficijent pogodniji, jer on *kvadrira* razlike, čime se više naglašavaju veće nego manje razlike.

Inače, *Tau* ima još i ove prednosti:

(a) Kendall je razradio postupak kojim se uz pomoć *Tau* koeficijenta može izračunavati i *parcijalna* korelacija, a to nije moguće računati uz koeficijent *Ro*. (Izračunavanje parcijalne korelacije uz *Tau* bit će opisano u poglavlju o parcijalnoj korelaciji.)

(b) *Tau* koeficijent korelacije može se koristiti i za koreliranje jedne ordinalne varijable (rangovi) s jednom *dihotomnom nominalnom* varijablom (npr. "prošao" — "pao", "živ" — "mrtav", "muškarac" — "žena"). (To će također biti opisano, i to u posebnom poglavlju.)

(c) Ako se *Tau* računa na nešto drugačiji način, nego što je opisano, moguće je već dobivenom rezultatu korelacije dodati jedan ili više novih rezultata, *a da pri tome nije potrebno čitav računski postupak ponavljati*, već jedino treba nešto dodatno izračunati. Taj ćemo postupak sada rastumačiti.

Taj se postupak može primijeniti kada ne unosimo *rangove* u račun, već *originalne mjerene rezultate*. Osim toga, varijabla X se *ne rangira*, već se rezultati upisuju bilo kojim redom. Evo primjera:

Recimo da u jednom ispitivanju dobijemo ove rezultate u dvije varijable za sedam ispitanika:

	ISPITANICI						
VARIJABLE	A	B	C	D	E	F	G
X	45	51	52	47	60	48	61
Y	13	15	12	10	16	13	18

U ovom postupku moramo *za obje varijable* izračunati sve razlike među rezultatima (pri čemu ništa ne smeta što ti rezultati nisu rangirani).

Evo tih razlika (vidi o tome u prijašnjem računu)

	A prema ostalima					
X	+1	+1	+1	+1	+1	+1
Y	+1	−1	−1	+1	0	+1

	B prema ostalima				
X	+1	−1	+1	−1	+1
Y	−1	−1	+1	−1	+1

	C prema ostalima			
X	−1	+1	−1	+1
Y	−1	+1	+1	+1

	D prema ostalima		
X	+1	+1	+1
Y	+1	+1	+1

	E prema ostalima	
X	−1	+1
Y	−1	+1

	F prema ostalima
X	+1
Y	+1

Sada treba *međusobno pomnožiti* korespondentne vrijednosti za X i Y varijable, pa ćemo dobiti:

+1 −1 −1 +1 0 +1 / −1 +1 +1 +1 +1 / +1 +1 −1 +1 / +1 +1 +1 / +1 +1 / +1 = 12 = S.

(N a p o m e n a: sada čitalac može vidjeti prednost u početku opisanog postupka kada se rangovi varijable X poredaju po rastućem rangu: tada se u toj varijabli sve razlike rangova bilježe sa +1, pa ih zbog toga ne treba niti računati, jer ukupni rezultat ovisi samo o sumi jedinica u varijabli Y!)

Budući da u varijabli X nema vezanih rangova $Tx = 0$. U varijabli Y imamo 2 vezana ranga, pa Ty iznosi $1/2 \cdot 2(2-1) = 1$

Primjenom formule (13.10) dobivamo:

$$Tau = \frac{S}{\sqrt{[1/2\,N\,(N-1) - T_x][1/2\,N\,(N-1) - T_y]}}$$

$$= \frac{12}{\sqrt{[(42/2) - 0][(42/2) - 1]}} = 0,59$$

Ako se sada pojavi neki *novi* rezultat za novog ispitanika H, koji, recimo, glasi 57 u varijabli X, i 17 u varijabli Y, potrebno je izračunati "jedinice" jedino između svih ostalih ispitanika i ispitanika H: u varijabli X između A i H je +1, između B i H +1, između C i H +1, itd. Za obje varijable to je dolje navedeno:

X	+1	+1	+1	+1	−1	+1	−1	
Y	+1	+1	+1	+1	+1	+1	−1	
Umnožak	+1	+1	+1	+1	−1	+1	+1	= 5

Taj dobiveni broj treba pribrojiti prethodno dobivenom S, pa dobivamo $S = 12 + 5 = 17$. Druge karakteristike, osim što se *povećao* N, nisu se promijenile, pa stoga završni račun glasi:

$$Tau = \frac{17}{\sqrt{[(56/2) - 0][(56/2) - 1]}} = \frac{17}{\sqrt{750}} = 0,62$$

13.10. INTERPRETACIJA KOEFICIJENTA KORELACIJE

Jedna od čestih pogrešaka u interpretaciji koeficijenta korelacije sastoji se u tome da se visina koeficijenta korelacije interpretira kao ***postotak*** zajedničkih faktora; drugim riječima, ako korelacija iznosi, pretpostavimo, 0,70, smatra se da to znači da u obje varijable postoji 70% zajedničkih faktora. *To nije točno.* Postotak zajedničkih faktora je manji od broja izraženog u korelaciji, i to sve manji što je korelacija niža. *Približno* možemo odrediti količinu zajedničkih faktora uz pomoć tzv. "koeficijenta determinacije", tj. kvadriranjem koeficijenta korelacije. Dakle, ako korelacija iznosi, na primjer, 0,70, postoji samo oko $0,70^2$, dakle 0,49, tj. 49% zajedničkih faktora, kod korelacije 0,90 postoji oko 81% zajedničkih faktora, itd. Jedino kod korelacije 1,0, sto posto faktora su zajednički faktori.

Iz ovog primjera vidimo da npr. dva puta veća korelacija znači znatno *više* od dva puta veće povezanosti. Naprimjer, $r = 0,30$ i $r = 0,60$ znače u prvom slučaju 9% zajedničkih faktora, a u drugom 36% zajedničkih faktora. Dakle, dvostruko veći koeficijent korelacije znači ***četiri puta*** veću povezanost, trostruko veći koeficijent korelacije znači ***devet*** puta veću povezanost.

Hoćemo li neku korelaciju proglasiti "visokom" ili "niskom", ovisi o nizu faktora, kao, primjerice, koje varijable mjerimo, kolika je značajnost toga koeficijenta, koliki je varijabilitet grupe, itd., a najviše zavisi o "kontekstu", tj. o konkretnoj situaciji. Kad bi npr. u razmaku od jednog sata dva puta mjerili težinu istim ljudima, pa kad bi između prvog i drugog mjerenja dobili korelaciju +0,70, zaključili bismo da je to vrlo *niska* korelacija, jer u ovom slučaju opravdano očekujemo da bi slaganje moralo biti znatno veće. Naprotiv, kad bismo tražili povezanost između krvnog tlaka i veličine podlaktice, i kad bismo dobili korelaciju +0,40, zaključili bismo da je to ***visoka*** korelacija, jer u ovom slučaju ne očekujemo (već iz iskustva) slaganje između te dvije varijable. (Eventualno bismo mogli očekivati vrlo nisku ***negativnu*** korelaciju, jer je veličina podlaktice u pozitivnoj korelaciji s tjelesnom visinom, a tjelesna visina u niskoj negativnoj korelaciji s krvnim tlakom.)

Ako je mjerenje provedeno na velikom broju slučajeva, onda kao *gruba* aproksimacija visine povezanosti između dvije varijable može poslužiti donja tablica za koeficijent korelacije r:

r od 0,00 do $\pm$ 0,20 znači nikakvu ili neznatnu povezanost
r od $\pm$ 0,20 do $\pm$ 0,40 znači laku povezanost
r od $\pm$ 0,40 do $\pm$ 0,70 znači stvarnu značajnu povezanost
r od $\pm$ 0,70 do $\pm$ 1,00 znači visoku ili vrlo visoku povezanost.

Pitanje interpretacije koeficijenta korelacije ne mora se, međutim, svoditi samo na pitanje ***što*** međusobno koreliramo, i kakav rezultat očekujemo, nego ono može biti i kompliciranje. Visina korelacije nije naime *samo* odraz stupnja povezanosti između dvije varijable, nego može biti i posljedica različitih drugih utjecaja, od kojih ćemo neke spomenuti:

a) *Grupiranje rezultata.* Ako grupiramo rezultate u razrede, to neće značajno mijenjati koeficijent korelacije samo onda ako broj razreda nije mali. Ako tom uvjetu ne možemo udovoljiti, to može imati za posljedicu iskrivljene koeficijente korelacije, to veće što je broj razreda manji.

b) *Zakrivljeni odnos.* Ako odnos između dvije varijable nije linearan nego zakrivljen, koeficijent korelacije r može biti toliko iskrivljen da njegovo izračunavanje nema nikakvog smisla. Tako, na primjer, dobivamo vrlo zakrivljen odnos između starosti i oštrine vida: otprilike do dvadesete godine oštrina vida *raste*, nakon toga neko vrijeme *stagnira*, a onda najprije sporo, a kasnije (otprilike od 45. godine) naglo *opada*. Kad bismo za taj odnos računali koeficijent korelacije r, dobili bismo vrijednost koja je vrlo blizu nuli!

U takvim slučajevima upotrebljava se tzv. "odnos korelacija", o kojem se mogu naći podaci u opširnijim statističkim udžbenicima.

c) *Eliminiranje vrijednosti oko aritmetičke sredine.* Pojedini istraživači, želeći dobiti podatke o odnosu između dvije varijable, katkad iz obrade rezultata isključe "srednju" skupinu. Tako, na primjer, nekoga može zanimati da li stav radnika prema poduzeću postaje sve pozitivniji što su radnici stariji, ali ako mjerenu skupinu radnika podijeli u dvije kategorije starosne dobi tako da uzme natprosječne i ispodprosječne po godinama, razlike u stavu sada mogu biti maskirane, jer i jedna i druga kategorija obuhvaćaju znatan broj "prosječno" starih ispitanika. Zbog toga se u obradu uzmu samo "ekstremne" grupe, tj. samo skupina mladih i skupina starih. Na taj način može se pokazati da se te dvije skupine statistički značajno razlikuju u stavu. Računati pak *korelaciju* (npr. r ili koeficijent kontingencije C) između stava i starosti kod ovako okrnjene grupe, značilo bi *umjetno je povećavati*, jer će korelacija sigurno biti niža ako ne eliminiramo vrijednosti oko prosjeka.

d) *Podskupine s različitim aritmetičkim sredinama.* Jednom prilikom jedan je istraživač ispitao profesore i studente jednog sveučilišta (skupina A) pomoću 3 testa, i to: 1. testa slušnog pamćenja riječi, 2. testa slušnog pamćenja povezanih poglavlja, i 3. testa vidnog prepoznavanja oblika. Te iste testove primijenio je i na skupinu nekvalificiranih radnika (skupina B) te je dobio ove korelacije između uspjeha u pojedinim testovima:

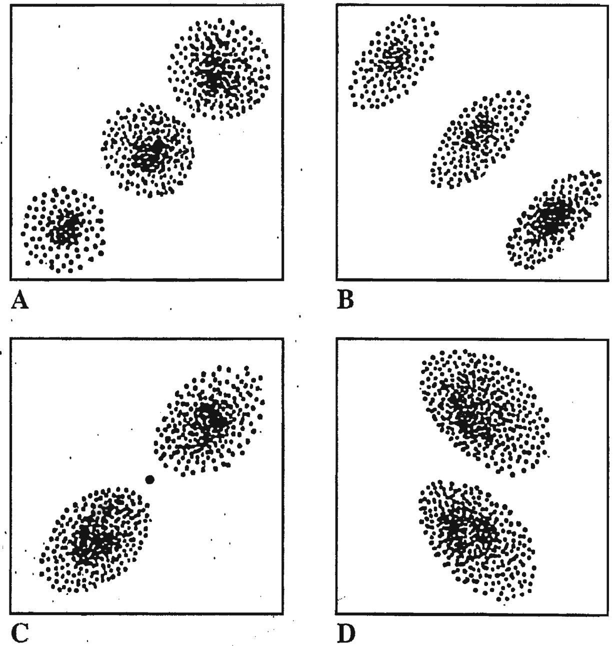
	Skupina A	Skupina B	Obje skupine zajedno
$r_{1,2}$	0,23	0,73	0,82
$r_{1,3}$	0,37	0,15	0,71
$r_{2,3}$	0,14	0,06	0,66

Kako se vidi, rezultat je na prvi pogled vrlo čudan, tj. korelacije za obje skupine zajedno u svim su slučajevima *veće*, nego za svaku skupinu posebno, što je naročito markantno u korelaciji između testa 2. i 3. Kako je to moguće?

Uzrok toj pojavi je u činjenici što se te skupine u jednoj ili u obje varijable *razlikuju u prosjeku*. Ako se ponovno poslužimo grafičkim prikazom, vrlo je lako razjasniti tu pojavu. Na slici 13.9. prikazano je nekoliko takvih mogućnosti: u primjeru A svaka od 3 podskupine ne pokazuje nikakvu korelaciju između varijable X i varijable Y, ali uzete *zajedno*, sve tri skupine daju *pozitivnu* korelaciju. U primjeru B svaka od 3 podskupine daje pozitivnu korelaciju, ali uzete zajedno daju *negativnu* korelaciju. U primjeru C svaka skupina daje pozitivnu ali nisku korelaciju, a zajedno daju pozitivnu i visoku korelaciju. U primjeru D svaka skupina daje negativnu korelaciju, ali zajednička korelacija je oko nule!

Želimo li u takvim slučajevima dobiti "prosječni" koeficijent korelacije koji odražava odnose između obje varijable u svim podskupinama, i nije pod utjecajem razlika među skupinama, moramo se poslužiti ovom formulom za izračunavanje korelacije r (formula je za dvije podskupine, ali ju je posve jednostavno transformirati na veći broj podskupina):

$$r = \frac{\Sigma X_1 Y_1 + \Sigma X_2 Y_2 - \frac{(\Sigma X_1)(\Sigma Y_1)}{N_1} - \frac{(\Sigma X_2)(\Sigma Y_2)}{N_2}}{\sqrt{\Sigma X_1^2 + \Sigma X_2^2 - \frac{(\Sigma X_1)^2}{N_1} - \frac{(\Sigma X_2)^2}{N_2}} \cdot \sqrt{\Sigma Y_1^2 + \Sigma Y_2^2 - \frac{(\Sigma Y_1)^2}{N_1} - \frac{(\Sigma Y_2)^2}{N_2}}} \tag{13.13}$$



Slika 13.9. Utjecaj razlika u aritmetičkim sredinama podgrupa na visinu korelacije. Vidi objašnjenje u tekstu

e) *Utjecaj raspona.* Ako je raspon ograničen u jednoj varijabli, on je zbog toga nužno ograničen i u drugoj varijabli, što značajno *smanjuje* visinu korelacije. Najpoznatiji primjeri ovoga tipa su slučajevi korelacije između uspjeha prilikom *selekcije* kandidata (npr. na jednom prijamnom ispitu) i njihovoga kasnijeg uspjeha u zvanju ili školovanju: ako su prilikom selekcije primljeni samo *najbolji*, time je raspon varijable uspjeha na ispitu *umjetno smanjen*, a posljedica toga je ta da je korelacija između uspjeha na ispitu i uspjeha na studiju znatno niža nego što bi bila da su svi kandidati (bez obzira na uspjeh na ispitu) primljeni na studij ili u zvanje.

Ipak postoji mogućnost da se aproksimativno odredi stvarna korelacija između uspjeha na prijamnom ispitu i uspjeha na studiju. U tu svrhu potrebno je poznavati standardnu devijaciju rezultata ispita (npr. testa znanja) za selekcioniranu grupu (s_x), korelaciju te grupe s kriterijem (npr. s uspjehom na studiju) (r_{xy}) i standardnu devijaciju u testu svih kandidata koji su polagali ispit (S_x).

Procijenjena korelacija između uspjeha na ispitu i uspjeha na studiju za cijelu skupinu može se nakon toga izračunati ovako:

$$R_{xy} = \frac{r_{xy}S_x}{\sqrt{r_{xy}^2 S_x^2 + s_x^2(1-r_{xy}^2)}}. \tag{13.14}$$

f) *Kauzalno intepretiranje korelacije.* Taj je problem, međutim, toliko važan da ćemo mu posvetiti posebno potpoglavlje 13.11.

13.11. KORELACIJA I UZROČNA VEZA

Kod pojedinih ljudi koji dolaze u priliku da istražuju korelaciju između pojedinih pojava, opaža se kadšto tendencija da se korelacija interpretira *kauzalno*, tj. interpretira se kao da je jedna od varijabli uzrok drugoj. Katkada je zaista tako da su dvije pojave koje su u međusobnoj korelaciji, ujedno i u uzročnoj vezi: količina oborina i bujnost vegetacije su u visokoj korelaciji, a ujedno je sigurno i to da oborine *uzrokuju* rast biljnog svijeta. No sama *činjenica* da između dvije pojave postoji korelacija još nam ne daje *nikakvo pravo* da te pojave povežemo *uzročnom* vezom.

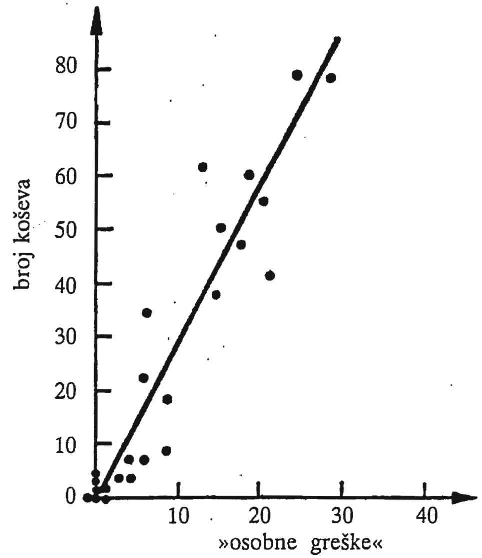
Statističar Freund kaže da koeficijent korelacije nije samo najčešće upotrebljavan, nego i najčešće *zloupotrebljavan* statistički postupak, a kao jedan od razloga navodi upravo spomenutu kauzalnu interpretaciju korelacije.

Kada bismo jednom bolesniku istodobno mjerili temperaturu pomoću dva toplomjera, jasno je da bismo dobili visoku korelaciju između niza rezultata jednog i niza rezultata drugog toplomjera, ali ni na pamet nam ne bi palo da smatramo da je jedan od tih toplomjera djelovao na onaj drugi — jer nam je jasno da su vrijednosti na oba toplomjera uzrokovane *trećim* faktorom, tj. temperaturom tijela. Jednako tako na osnovi činjenice da se naš ručni sat redovito dobro slaže s nekim javnim satom u gradu ne zaključujemo da bi bilo koji od tih satova *utjecao* na drugi. Ako u nekim zemljama usporedimo broj električnih aparata u kućanstvu s brojem djece u tim istim domaćinstvima, obično nalazimo dosta visoku korelaciju, ali ne zato što bi, recimo, broj djece djelovao na broj električnih aparata (ili još manje, što bi broj električnih aparata utjecao na broj djece u obitelji!), već zato što su obje te varijable uzročno povezane s ekonomskim standardom: što je ekonomski standard obitelji viši, ta obitelj može sebi "dopustiti" veći tehnički standard, kao i — veći broj djece.

Poznata je šala među statističarima (neki čak tvrde da se to negdje zaista dogodilo) koja glasi ovako: glavni dokaz da rode donose djecu je taj što postoji visoka pozitivna korelacija između broja roda u selima i broja novorođene djece. No uzrok povezanosti je i opet treći faktor: veća sela imaju i više dimnjaka, pa

prema tome i više roda koje žive na tim dimnjacima, a imaju — zato što su veća — i više djece od malih sela. Zbog istih razloga možemo, primjerice, u europskim gradovima naći pozitivnu korelaciju između broja svećenika i broja prostitutki, između broja kinematografa i broja štakora, itd., itd.

Zanimljiv primjer daje statističar Ingram, koji navodi jedno istraživanje na 25 igrača košarke, i povezanost između postignutih koševa i učinjenih prekršaja tih igrača. Rezultati su prikazani na slici 13.10.



Slika 13.10. Odnos između broja "osobnih pogrešaka" i postignutih koševa u jednoj košarkaškoj momčadi

Kako se vidi, povezanost je vrlo visoka, a koeficijent korelacije r za tih 25 igrača iznosi čak + 0,93. No, Ingram ispravno upozorava da činjenica što neki igrač ima mnogo prekršaja, očito nije *razlog* što on postiže i više koševa, nego su vjerojatno i količina prekršaja i količina koševa uvjetovani nekim trećim faktorom, npr. *vremenom*, što ga neki igrač provede u igri ili *agresivnošću* tog igrača i sl.

Upravo školski primjer vrlo problematičnoga kauzalnog tumačenja korelacije je historijat odnosa između pušenja majke i nedovoljne težine novorođenčeta. Godine 1957. "American Journal of Obstetrics and Gynecology" objavio je podatak da žene koje u toku graviditeta puše, rađaju u većem postotku djecu premalene težine (ispod 2 500 g). U 15 godina koje su slijedile, više od 30 istraživanja te pojave *potvrdilo* je povezanost između pušenja majke i premale težine djeteta u porodu. Ti su nalazi zabrinuli istraživače, jer djeca rođena s premalom težinom imaju više od dvadeset puta veći rizik da će umrijeti u prvom mjesecu života nego djeca normalne težine. Uzročna veza između pušenja i nedonoščadi smatrana je posve razumljivom, jer se držalo da pušenje interferira s intrauterinim razvojem čeda. Autor Yerushalmi odlučio je da ozbiljno ispita tu povezanost te je u tu svrhu ispitao 5 466 graviditeta kod 3 422 žene, koje su sve rodile do najkasnije 25. godine života. Neke su od tih žena pušile, a neke nisu. No radi detaljnije kauzalne analize Yerushalmi je žene podijelio u 4 kategorije:

a) žene-nepušači koje nikada nisu pušile;

b) žene stalni pušači, tj. one koje su pušile u toku graviditeta, a i kasnije, u doba kada su intervjuirane;

c) žene "budući" pušači, tj. one koje su počele pušiti tek nakon poroda i

d) žene "bivši" pušači, tj. one koje su nakon poroda prestale pušiti.

Postotak poroda djece premalene težine za svaku od ovih skupina žena prikazan je u tablici 13.11.

TABLICA 13.11.

POSTOTAK DJECE S PREMALOM TEŽINOM U PORODU U RAZLIČITIM SKUPINAMA MAJKI (vidi tekst)

	Broj poroda	Postotak djece s premalom težinom
Nepušači	2 529	5,3
Stalni pušači	2 076	8,9
"Budući" pušači	210	9,5
"Bivši" pušači	651	6,0

Ako pažljivo analiziramo ovu tablicu, vidjet ćemo da usporedba nepušača sa stalnim pušačima daje očekivanu povezanost: majke-pušači imaju više poroda nedonoščadi nego majke-nepušači. Ali, usporedba "budućih" i "bivših" pušača pokazuje da su majke koje su počele pušiti tek *nakon* poroda imale visok postotak djece s premalom težinom. I, napokon, majke "bivši" pušači, u razdoblju *prije* nego što su napustile pušenje, rodile su znatno manje nedonoščadi nego stalni pušači, i manje od "budućih" pušača.

Na temelju tih rezultata Yerushalmi zaključuje da "činjenice zapravo podržavaju hipotezu da je povećani postotak nedonoščadi uvjetovan *pušačem a ne pušenjem*". On, naime, smatra da žene-"pušači" kao skupina predstavljaju takvu skupinu, koja — bilo zbog određenog načina života ili kojega drugog razloga — ima tendenciju da rodi nedonošče *bez obzira na to da li stvarno puši ili ne*! Pušenje — prema njemu — dakle ne *uzrokuje* defektni porod, nego su i pušenje i defektni porod pod utjecajem *trećeg faktora* (ili trećih faktora). Yerushalmi je to donekle i dokazao jednim svojim drugim istraživanjem na više od 4 000 žena-pušača. Između žena-pušača i žena-nepušača našao je *dosta razlika* u načinu života i ponašanja: pušačice su manje koristile kontraceptivna sredstva, u manjem postotku su planirale rođenje djeteta, u većem su postotku pile alkohol i kavu, i u znatno su većem

postotku *pretjerano* pile alkohol; dakle, u cjelini su — kako kaže Yerushalmi — žene-pušači ekstremnije u užicima i slobodnije u načinu života od žena-nepušača.

Sličan ovom problemu je još poznatiji problem odnosa između pušenja i karcinoma pluća. Ono što je već odavno *nedvojbeno dokazano*, to je prilično visok stupanj povezanosti između prosječne količine dnevno popušenih cigareta i učestalosti karcinoma pluća: najmanji postotak karcinoma pluća nalazimo kod nepušača, a kod pušača je učestalost karcinoma pluća to veća, što veću količinu cigareta dnevno popuše: kod jakih pušača smrtnost od karcinoma pluća dvadest četiri puta je veća nego kod nepušača. (Za utjehu pušačima: to još uvijek ne znači da jako mnogo pušača umire od raka pluća; kod teških pušača 1,66 ljudi na 1 000, a kod nepušača 0,07 na tisuću ljudi umire od raka pluća.) Na osnovi te činjenice naravno da je vrlo plauzibilno zaključiti da pušenje *uzrokuje* rak pluća. Već je Fisher svojedobno opravdano upozorio da se treba čuvati takvih kauzalnih interpretacija, jer bi se — kazao je on — korelacija između pušenja i karcinoma pluća mogla objasniti i djelovanjem nekoga trećeg faktora (možda biološkoga), koji ima utjecaja kako na razvoj karcinoma pluća, tako i na potrebu za pušenjem.

Neka suvremena istraživanja na jednojajčanim blizancima, od kojih je jedan član para bio pušač, a drugi nepušač, pokazuju dosta velik postotak oboljenja od raka pluća kod *oba* blizanca, i to često u približno jednako vrijeme. To kao da ide u prilog Fisherovoj tezi, a donekle bi bilo u skladu i s istraživanjima Yerushalmija.

Kao što je poznato, uzročna veza između dvije pojave može se definitivno dokazati jedino ***eksperimentom***. Eysenck je svojedobno u jednoj svojoj publikaciji o pušenju i raku pluća naveo kako bi teoretski trebao izgledati eksperimentalni nacrt jednog takvog istraživanja: valjalo bi uzeti vrlo velik uzorak (npr. nekoliko tisuća) novorođene djece, te bi *svu* djecu toga uzorka trebalo u toku puberteta naučiti da puše, i tu naviku morali bi održavati cijelog života. Kod druge, kontrolne, također velike skupine, morali bismo imati djecu, kojoj smo potpuno onemogućili da u životu ikada puše. Sve članove ovih obaju uzoraka trebalo bi pratiti *kroz cijeli njihov život*, i registrirati koliko je pojedinaca u svakoj skupini dobilo rak pluća.

No dakako, zbog posve razumljivih etičkih razloga takav eksperiment nije moguće provesti na ljudima. Stoga su godinama rađeni slični eksperimenti na laboratorijskim životinjama (pretežno miševima), te je tako konačno i *dokazano* da postoji *uzročna veza* između pušenja i raka pluća, iako se istovremeno sve češće u modernoj medicinskoj literaturi nailazi i na tvrdnje, da i genetski faktori kod nekih vrsta karcinoma — pa i kod karcinoma pluća — igraju ulogu.

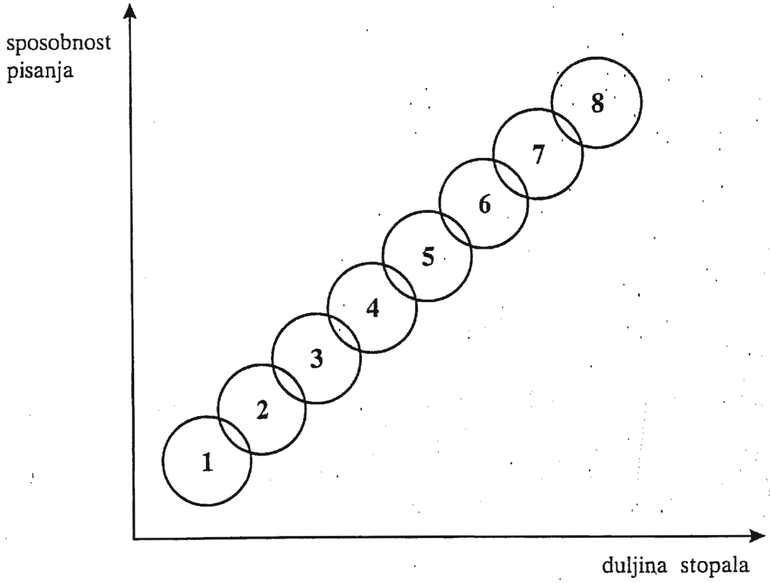
Kao što se iz ovih primjera vidi, problemi su vrlo kompleksni i teški, i zaista smo katkada naprosto izgubljeni u rezultatima koje daju istraživanja. Možda pri tome kao mala utjeha može poslužiti tzv. "Andersonov zakon" (Anderson je pisac znanstveno-fantastičnih romana), koji je rekao: *"Nema problema, pa ma koliko on bio kompleksan, koji ne bi nakon savjesne analize postao još kompleksniji."*

13.12. PARCIJALNA KORELACIJA

Maloprije u diskusiji o interpretaciji koeficijenta korelacije upozorili smo na to kako stvarna korelacija može biti *iskrivljena* ako u jednoj ili u obje varijable skupljamo zajedno skupine koje imaju različite aritmetičke sredine.

Kako upravo iz toga problema proistječe potreba za tzv. parcijalnom korelacijom, poslužit ćemo se ponovno jednim primjerom:

Pretpostavimo da smo uzeli uzorak školske djece dobi od 7 do 15 godina i u svakom razredu izmjerili kod te djece duljinu njihovih *stopala* i sposobnost *pisanja*. Vrlo je malo vjerojatno da su u biti ta dva svojstva povezana, tj. korelacije među njima vjerojatno nema. *No u oba ta svojstva prosječni rezultat raste s godinama djeteta*, pa će prema tome prosječna duljina stopala starije djece biti veća nego mlade djece, a prosječna sposobnost pisanja bit će također bolja što su djeca starija. Ako koreliramo te dvije varijable *za svu djecu zajedno*, dobit ćemo rezultat koji grafički možemo prikazati slikom 13.11.



Slika 13.11. Iako u svakom razredu posebno nema korelacije između duljine stopala djece i sposobnosti pisanja, za svih 8 razreda zajedno korelacija je vrlo visoka!

Brojevi u krugovima označuju pojedine školske razrede, a krugovi (dakle ne elipse!) pokazuju da *nema* korelacije između te dvije varijable. ali budući da i jedna i druga varijabla rastu s porastom djetetovih godina, *zajednička* slika nužno dovodi do toga da za svu djecu zajedno nađemo relativno visoku korelaciju između duljine stopala i sposobnosti pisanja — što je očita besmislica!

U ovom slučaju trebalo bi dakle izračunati *stvarnu* korelaciju između varijable "duljina stopala" i varijable "sposobnost pisanja" — *isključivši utjecaj starenja* (odnosno — što je isto — držeći ga konstantnim). Takvu korelaciju između dviju varijabli, kod koje isključimo utjecaj jednog (ili više) faktora koji nam smetaju, nazivamo *parcijalnom korelacijom*. Ona se računa prema formuli:

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \qquad (13.15)$$

pritom simboli znače:

$r_{12\cdot 3}$ = korelacija između varijable 1 i 2, isključivši utjecaj varijable 3 (u našem slučaju varijabla 1 = duljina stopala, varijabla 2 = sposobnost pisanja, varijabla 3 = starost).

Uzmimo da smo u prvom pristupu tom problemu našli korelaciju od + 0,69 između duljine stopala i sposobnosti pisanja (r_{12}) (sve naravno u rasponu 7 – 15 godina), ali da nam naknadna provjeravanja pokažu da korelacija između duljine stopala i starosti (r_{13}) iznosi + 0,90, a između sposobnosti pisanja i starosti (r_{23}) iznosi + 0,75, onda u našem slučaju dobivamo da je parcijalna (*dakle stvarna*) korelacija između duljine stopala i sposobnosti pisanja:

$$r_{12\cdot 3} = \frac{0,69 - 0,90 \cdot 0,75}{\sqrt{1-0,81}\sqrt{1-0,56}} = \frac{0,015}{0,289} = 0,05.$$

Kako, dakle, vidimo, korelacije između te dvije varijable zapravo nema.

Ispuštanje iz vida činjenice da je možda neka *treća* varijabla dovela do "prividne" korelacije između druge dvije varijable, može dovesti još i do veće zablude nego što je ova, prikazana primjerom s duljinom stopala i sposobnosti pisanja.

Uzmimo, na primjer, da je jedan industrijski psiholog u svom poduzeću na ukupno 450 radnika izračunao korelaciju između broja izostanaka, što ih je svaki radnik do sada ukupno u poduzeću imao, i visine prosječne sadašnje plaće tih radnika, te je dobio korelaciju od + 0,40. Vrlo bi teško bilo interpretirati takvu pozitivnu, i čak relativno visoku korelaciju između broja izostanaka i visine plaće, no odmah nam se nameće misao da vjerojatno ukupno više izostanaka imaju *oni koji se dulje nalaze u poduzeću*. Izvršimo li ispitivanje odnosa između radnog staža i ukupnog broja izostanaka, dobijemo, na primjer, korelaciju od + 0,80, a između radnog staža i visine plaće korelaciju od + 0,70.

Prema tome, parcijalna korelacija između broja izostanaka i visine plaće, isključujući utjecaj radnog staža, iznosi:

$$r_{12\cdot 3} = \frac{0,40 - 0,80 \cdot 0,70}{\sqrt{1-0,80^2}\sqrt{1-0,70^2}} = -0,37,$$

Dakle, korelacija je zapravo *negativna*, što je i razumljivo (jer se može smatrati da u cjelini više izostaju "slabiji" radnici, a ti su i manje plaćeni od onih boljih).

Na temelju ovih primjera možda bi bilo opravdano zaključiti da je parcijalna korelacija postupak kojim smo se prisiljeni poslužiti kada smo pogrešno, neznanstveno i bez iskustva planirali neke eksperimente ili istraživanje!

Značajnost parcijalne korelacije može se testirati t-testom, i to uz pomoć formule:

$$t = \frac{r_{12\cdot 3}}{\sqrt{(1-r_{12\cdot 3}^2)/(N-1)}}. \qquad (13.16)$$

Značajnost se očitava iz t-tablice, uz stupnjeve slobode $N - 3$.

Testiramo li ovom formulom značajnost korelacije u našem drugom primjeru:

$$t = \frac{-0,37}{\sqrt{(1-0,37^2)/449}} = -8,41.$$

Dobivena se parcijalna korelacija, dakle, statistički značajno razlikuje od nule:

Kao što je to već u poglavlju o Kendallovu koeficijentu *Tau* korelacije bilo spomenuto, za razliku od *Ro* koeficijenta korelacije, *Tau* koeficijent može se koristiti i za izračunavanje parcijalne korelacije.

Formula je po smislu *identična* formuli za parcijalnu r korelaciju (13.15) no — kako tvrdi sam Kendall — radi se o slučajnoj sličnosti, i ona glasi:

$$Tau = \frac{Tau_{12} - Tau_{13} - Tau_{23}}{\sqrt{1 - Tau_{13}^2}\,\sqrt{1 - Tau_{23}^2}} \tag{13.17}$$

Nažalost, još nisu izrađeni postupci za testiranje značajnosti parcijalne *Tau* korelacije.

13.13. KOEFICIJENT MULTIPLE KORELACIJE

Najčešće je u psihološkoj praksi, a kadikad i u praksi drugih struka (medicina, sociologija i dr.), potrebno znati kakva je korelacija između *nekoliko* "prediktora" (npr. nekoliko psiholoških ili medicinskih testova) i jednog "kriterija" (npr. uspjeha na radnom mjestu ili uspješnosti dijagnoze zdravstvenog stanja). Ako, na primjer, psiholog prilikom selekcije personala za neko radno mjesto upotrijebi *nekoliko* testova, ili liječnik prilikom određivanja statusa u nekom području koristi *nekoliko* podataka — a pod pretpostavkom da svaki od podataka ima određenu "dijagnostičku" ili "prognostičku" valjanost — postavlja se pitanje kolika je korelacija između *ukupnog* rezultata u prediktorima, i rezultata u kriteriju. Pri tome se ujedno traži onakva "kombinacija" zbrojenih rezultata svih "prediktora" koja će dati *najveću moguću* korelaciju s kriterijem. Na primjer, ako psiholog prilikom selekcije primijeni jedan test inteligencije i jedan test psihomotorike, on želi znati da li su te dvije sposobnosti jednako važne za uspjeh, ili je možda jedna od njih važnija — pa će u tom slučaju rezultatu te važnije sposobnosti dati veće značenje u ukupnom rezultatu iz oba testa.

Prije nego što nastavimo raspravu o tom problemu, treba čitaoca podsjetiti na poglavlje 8.1, u kojem smo govorili o položaju pojedinog rezultata u grupi ostalih rezultata. U tom poglavlju pokazali smo kako uz pomoć z-vrijednosti taj položaj (u normalnoj raspodjeli) možemo točno odrediti. Ujedno smo tom prilikom spomenuli da *skupnu* ocjenu iz nekoliko različitih mjerenja treba dobiti ne na temelju zbrajanja bruto-rezultata, nego na temelju zbrajanja z-vrijednosti, što ih je pojedinac u svakom mjerenju postigao. Pritom smo upozorili da bi pri jednostavnom zbrajanju bruto-rezultata u ukupnoj sumi najveću ulogu imao test s *najvećim varijabilitetom*.

Iz toga slijedi da pri zbrajanju bruto-rezultata mi već (nehotice!) "otežavamo" ("ponderiramo") rezultate, tj. *dajemo veću "težinu" rezultatima s većim varijabilitetom*, a za to nemamo nikakvog opravdanja. Zbog toga je potrebno bruto-rezultate pretvoriti u z-vrijednosti i tek tako dobivene rezultate zbrajati: sada smo svakom rezultatu dali *jednaku* "težinu" ("ponder").

No maloprije smo spomenuli da često upravo *želimo* dati veću važnost nekim rezultatima nego drugima, jer su neki rezultati *važniji* od nekih drugih.

Da bismo to mogli postići, tj. da bismo uspjeli doznati koji su rezultati važniji, moramo se poslužiti posebnim oblikom korelacije, poznatim pod nazivom *multipla korelacija*. Prema manje-više slobodnoj definiciji multipla korelacija je *maksimalno mogući* koeficijent korelacije između dva ili više prediktora i jednog kriterija. Tu maksimalno visoku korelaciju možemo postići samo onda ako veću "težinu" damo važnijim, a manju "težinu" manje važnim prediktorima. Posao izračunavanja multiple korelacije sastoji se u tome da se nađe koja je *maksimalna* korelacija između grupe prediktora i jednog kriterija, a posebnim se računom ustanovi u kojim se uvjetima ta maksimalna korelacija može postići, tj. koliko treba povećati ili smanjiti važnost pojedinog prediktora. (Taj isti posao dao bi se teoretski obaviti i bez postupka multiple korelacije, a sastojao bi se u tome da zbrajamo rezultate pojedinih prediktora, i zbroj — uz pomoć r-korelacije — koreliramo s kriterijem. Pritom bi trebalo davati različite "težine" rezultatima pojedinih prediktora, zbrajati tako "otežane" rezultate i uvijek ponovno korelirati s kriterijem, tražeći onu kombinaciju koja će dati najvišu korelaciju. Broj takvih kombinacija praktički je *neizmjeran*! Multipla korelacija omogućuje da *bez traženja* nađemo *optimalnu* "težinu" svakoga pojedinog prediktora.)

Prikazat ćemo postupak multiple korelacije samo za slučaj s 3 varijable, tj. za slučaj dva prediktora i jednog kriterija. Iako je u načelu postupak za veći broj varijabli jednak, tehnika *računanja* je toliko komplicirana (i dugotrajna) da prelazi okvire ovoga elementarnog udžbenika.

Pretpostavimo da je psihologu, koji obavlja selekciju za jedno radno mjesto, a pri tome primijeni na uzorku od $N = 80$ jedan test inteligencije i jedan test psihomotorike, poznato da njegov test inteligencije ima korelaciju $+\,0{,}40$ s uspjehom u zvanju, a test psihomotorike ima korelaciju $+\,0{,}28$. Međusobna korelacija između ta dva testa, recimo, iznosi $+\,0{,}13$. Budući da želi koristiti *oba* testa, on postavlja pitanje kolika je multipla korelacija između ta dva testa, s jedne, i uspjeha na radnom mjestu, s druge strane. Drugim riječima, uz pomoć multiple korelacije on će dobiti odgovor na pitanje koju "težinu" treba dati rezultatima svakog od ova dva testa da bi *ukupni rezultat* iz oba testa bio relativno najbolji prediktor uspjeha na radnom mjestu. (Unaprijed je jasno da će trebati dati veću težinu testu inteligencije, jer on ima veću korelaciju s kriterijem nego test psihomotorike. Međutim, samo iz tih podataka ne bismo još mogli odrediti preciznije "težine", jer one ovise i o interkorelaciji između oba prediktora.)

Ako varijablu "kriterij" označimo s "0" (nula), a s 1 i 2 oba prediktora (recimo da 1 = test inteligencije, a 2 = test psihomotorike), onda formula multiple korelacije glasi:

$$R_{0\cdot 12} = \sqrt{\frac{r_{01}^2 + r_{02}^2 - 2r_{01}\cdot r_{02}\cdot r_{12}}{1 - r_{12}^2}}, \tag{13.18}$$

pri čemu $R_{0\cdot 12}$ znači: multipla korelacija između kriterija (0) i dva prediktora (1 i 2).

Uvrstimo li naše podatke u tu formulu, dobivamo:

$$R_{0\cdot 12} = \sqrt{\frac{0,40^2 + 0,28^2 - 2\cdot 0,40\cdot 0,28\cdot 0,13}{1 - 0,13^2}}$$

$$= \sqrt{\frac{0,2093}{0,9831}} = \sqrt{0,2129} = +0,46.$$

Dakle, s *oba* testa zajedno korelacija se povisila na 0,46 (test inteligencije ima korelaciju 0,40), ali, naravno, uz uvjet da dademo optimalnu "težinu" rezultatima svakog testa. Ti optimalni "ponderi" nazivaju se obično "beta-koeficijenti" i računaju se, za svaki od dva prediktora, ovako:

$$\text{beta}_1 = \frac{r_{01} - r_{02}r_{12}}{1 - r_{12}^2}, \tag{13.19}$$

$$\text{beta}_2 = \frac{r_{02} - r_{01}r_{12}}{1 - r_{12}^2}. \tag{13.20}$$

Budući da su nam svi elementi poznati, možemo beta-koeficijente lako izračunati:

$$\text{beta}_1 = \frac{0,40 - 0,28 \cdot 0,13}{1 - 0,13^2} = 0,37,$$

$$\text{beta}_2 = \frac{0,28 - 0,40 \cdot 0,13}{1 - 0,13^2} = 0,23.$$

Pomnožimo li beta-koeficijent testa 1 (0,37) sa z-vrijednostima rezultata u testu 1, a beta-koeficijent testa 2 (0,23) sa z- vrijednostima rezultata u testu 2 i tako dobivene rezultate *zbrojimo*, dobit ćemo rezultate koji će s kriterijem imati maksimalno moguću korelaciju, tj. korelaciju od + 0,46. (Želimo li izbjeći preračunavanje svakoga individualnog rezultata u z- vrijednost, možemo beta-koeficijente za svaki test podijeliti standardnom devijacijom toga testa, i tako dobivene "relativne" beta-koeficijente množiti *bruto*-rezultatima u svakom testu.)

Napomena. Izračunavanje beta-koeficijenata može u praksi biti korisno i bez izračunavanja koeficijenta multiple korelacije R. Pretpostavimo da neki liječnik ili psiholog želi *rangirati* grupu ispitanika koristeći dva prediktora koji su u određenoj korelaciji s kriterijem, onda će mu poznavanje beta-koeficijenata omogućiti dobivanje "najispravnije" sume rezultata iz oba prediktora koje će sume rangirati.

Ako su nam poznati beta-koeficijenti, multipla korelacija može se izračunati i prema formuli:

$$R_{0\cdot 12} = \sqrt{\text{beta}_1 r_{01} + \text{beta}_2 r_{02}}. \tag{13.21}$$

U našem slučaju dobivamo:

$$R_{0\cdot 12} = \sqrt{0,37 \cdot 0,40 + 0,23 \cdot 0,28} = 0,46.$$

Značajnost koeficijenta multiple korelacije može se testirati uz pomoć F-testa. (Vidi tablicu L. Ta će tablica biti protumačena tek u poglavlju 20.)

$$F = \frac{R^2}{1 - R^2} \cdot \frac{N - k - 1}{k}. \tag{13.22}$$

gdje je R = koeficijent multiple korelacije
N = broj opažanja
k = broj prediktora (broj nezavisnih varijabli).

Tablica se čita tako da se kao stupnjevi slobode$_1$ uzme k, a stupnjevi slobode$_2$ = $= N - k - 1$.

U našem primjeru dobili bismo:

$$F = \frac{0,46^2}{1 - 0,46^2} \cdot \frac{80 - 2 - 1}{2} = 10,33.$$

Kako je 10,32 znatno veći od 3,10 (aproksimacija iz tablice, stupnjevi slobode$_2$ očitani između 60 i 120), dobivena multipla korelacija statistički se značajno razlikuje od nule.

13.14. "POINT-BISERIJALNI" KOEFICIJENT KORELACIJE

Često se događa da nas zanima korelacija između jedne *kontinuirane* varijable (npr. visine, težine, bodova u testu, itd.) i jedne *dihotomne* varijable, tj. takve koja se dijeli u dvije *jasno odvojene* kategorije (npr. živi-mrtvi, muško-žensko, prošao-pao, itd.).

Poslije ćemo vidjeti da se taj problem može rješavati i nekim drugim postupcima (hi-kvadrat test, i pomoću njega izračunati jedan koeficijent korelacije), no, izravno izračunavanje korelacije vrlo je jednostavno pa ćemo ga stoga opisati. Koeficijent korelacije ove vrste zove se "point-biserijalni r", i piše se r_{pb}.

Prije nego što prijeđemo na računanje, potrebno je "numerirati" dihotomnu varijablu i to tako da se jednoj karakteristici dade jedan, a drugoj karakteristici drugi broj. Svejedno je koji su to brojevi, no najviše je uobičajeno da se daju brojevi 0 i 1.

Pošto smo to učinili, *treba jednostavno izračunati Pearsonov koeficijent korelacije r*, i svakome tko ima računalo koje može izračunavati koeficijent korelacije r, treba preporučiti da tako radi jer je to najjednostavnije i najbrže.

No, kao što znamo, *bez* elektronskog računala izračunavanje r- koeficijenta prilično je dugotrajno i naporno, pa u tom slučaju za ovu vrstu korelacije treba koristiti drugi postupak, izražen ovom formulom:

$$r_{pb} = \frac{N\Sigma Y_1 - N_1 \Sigma Y}{\sqrt{N_1 N_0 [N\Sigma Y^2 - (\Sigma Y)^2]}}. \tag{13.23}$$

Simboli formule znače ovo:

X = dihotomizirana varijabla (koju bilježimo s 0 i 1)
Y = kontinuirana varijabla
ΣY_1 = suma Y vrijednosti vezanih uz vrijednost X varijable, označene sa 1
ΣY = suma svih Y vrijednosti
ΣY^2 = suma svih kvadriranih Y vrijednosti
N_1 = broj svih opservacija kod kojih je $X = 1$
N_0 = broj svih opservacija kod kojih je $X = 0$
N = totalni broj svih slučajeva (opservacija) = $N_1 + N_0$.

Za primjer uzmimo rezultate skupine studenata u jednom testu aritmetike. Rezultati su prikazani u tablici 13.12. U tablici su muški ispitanici u varijabli "Spol" označeni s 1, a ženski s 0.

TABLICA 13.12.
BODOVI U TESTU ARITMETIKE ZA 20 STUDENATA I STUDENTICA

Spol (X)	Bodovi u testu (Y)
1	10
1	15
0	30
0	20
0	25
1	15
0	20
0	25
0	30
1	20
1	5
0	5
1	10
0	10
0	20
1	10
0	30
0	35
1	5
0	10
$\Sigma X = 8$	$\Sigma Y = 350$
	$\Sigma Y^2 = 7\,800$

$$\begin{aligned}
\Sigma Y_1 &= 10 + 15 + 15 + 20 + 5 + 10 + 10 + 5 = 90 \\
\Sigma Y &= 350 \\
N_1 &= 8 \\
N_0 &= 12 \\
\Sigma Y^2 &= 7\,800 \\
(\Sigma Y)^2 &= 350^2 = 122\,500
\end{aligned}$$

$$r_{pb} = \frac{(20 \cdot 90) - (8 \cdot 350)}{\sqrt{8 \cdot 12[(20 \cdot 7\,800) - (122\,500)]}}$$

$$= \frac{1\,800 - 2\,800}{\sqrt{96(156\,000 - 122\,500)}}$$

$$= -0,56.$$

Prema tome, postoji prilično visoka *negativna* korelacija između spola i uspjeha u ovom testu, tj. muškarci imaju slabiji uspjeh od žena. No, u ovom momentu treba biti pažljiv! S obzirom na to da smo muškarcima dali oznaku 1, a ženama 0, negativna korelacija znači da postoji negativan odnos između učinka u testu aritmetike i spola "muškarac", a pozitivan između rezultata testa i spola "žena".

Da smo ženama dali oznaku 1, a muškarcima 0, korelacija bi naravno bila jednaka, samo *pozitivna*.

Kao što je rečeno, da smo na rezultatima tablice 13.12. računali koeficijent r, dobili bismo istovjetan rezultat.

Značajnost koeficijenta r_{pb} testira se tablicama na isti način kao i značajnost koeficijenta r. Dakle, u našem slučaju to je tablica D u Dodatku, gdje uz 18 stupnjeva slobode čitamo da granična vrijednost na razini značajnosti od 5% iznosi 0,444. Naš koeficijent je, dakle, značajan.

U p o z o r e n j e. Ako želimo računati korelaciju između jedne kontinuirane varijable i jedne *umjetno dihotomizirane* varijable, ne smijemo koristiti opisani postupak. Pod umjetno dihotomiziranom varijablom možemo, na primjer, zamisliti radni uspjeh skupine ljudi (koji manje-više daje normalnu distribuciju), pa smo ga dihotomizirali, na primjer, uz pomoć Centralne vrijednosti u rezultate ispod i iznad Centralne vrijednosti. U takvim slučajevima koristi se takozvani *biserijalni* r, a kako je njega kompliciranije računati, nećemo ga obraditi u ovoj knjizi.

13.15. KOEFICIJENT KONKORDANCIJE W

Rang-korelacija *Ro* pokazuje korelaciju (aproksimativnu) između dva niza rangova. No, u životu se katkad događaju situacije u kojima nas zanima slaganje između više nizova rangova: može se npr. dogoditi da nekoliko ocjenjivača dade svaki svoj rang za nekoliko učenika, pa nas zanima koliko se ocjenjivači međusobno slažu — a to se računa pomoću koeficijenta konkordancije W.

Pretpostavimo da su četvorica ocjenjivača ($m = 4$) rangirali šestoricu učenika ($N = 6$), i da su dobiveni ovi rezultati:

	Kandidati (N)					
Ocjenjivači (m)	a	b	c	d	e	f
A	6	4	1	2	3	5
B	5	3	1	2	4	6
C	6	4	2	1	3	5
D	3	1	4	5	2	6
Totali rangova (T_i)	20	12	8	10	12	22

Kada bi postojalo *potpuno* slaganje među ocjenjivačima, onda bi kandidati morali davati ove sume rangova (u bilo kojem redoslijedu); 4, 8, 12, 16, 20 i 24. Naprotiv, kada bi suci kandidate rangirali "bilo kako", tj. kada među njima ne bi bilo nikakvog slaganja, sume rangova za sve kandidate tendirale bi *sličnoj vrijednosti*, tj. ukupnoj sumi rangova svih ocjenjivača, podijeljenoj brojem kandidata koji se rangiraju.

Po svom logičkom smislu, koeficijent konkordancije W testira odnos između *stvarnog* slaganja ocjenjivača i *maksimalno mogućeg* slaganja.

Kao što je rečeno, pod nul-hipotezom (koja glasi: suci su po slučaju dali ocjene, među njima nema nikakva slaganja) svaki rangirani pojedinac morao bi imati jednaku sumu rangova, koja se računa prema formuli:

$$\frac{m(N+1)}{2}. \tag{13.24}$$

U našem primjeru to iznosi $\frac{4 \cdot 7}{2} = 14$.

Da bi se izračunao W, treba naći razlike između svake sume rangova i očekivane sume pod nul-hipotezom. Dakle, u našem primjeru treba naći razlike između svake sume rangova (totala rangova T_i) i 14. Te razlike treba kvadrirati i zbrojiti, a suma tih kvadriranih razlika obično se naziva S. Ona dakle znači ovo:

$$S = \sum \left(\begin{array}{l} \text{suma rangova} \\ \text{prosuđivanih} \\ \text{elemenata} \end{array} - \begin{array}{l} \text{očekivana suma} \\ \text{rangova pod} \\ \text{nul-hipotezom} \end{array} \right)^2 \quad \text{ili,}$$

$$= \sum \left(T_i - \frac{T_i}{N} \right)^2.$$

Budući da S predstavlja mjeru *stvarnog* slaganja među sucima, osnovna formula za W glasi:

$$W = \frac{S}{S_{\max}}.$$

Ono što nam još nije poznato, to je podatak koliko iznosi maksimalno slaganje ($S_{\max}$). Ako postoji potpuno slaganje među sucima, onda:

$$S_{\max} = \frac{m^2(N^3 - N)}{12},$$

pa prema tome $W = \dfrac{S}{\dfrac{m^2(N^3 - N)}{12}}$.

Sredimo li tu formulu, dobivamo konačnu formulu za W:

$$W = \frac{12\,S}{m^2(N^3 - N)}. \tag{13.25}$$

Iz toga slijedi da kod potpunog slaganja sudaca (kada je dakle $S = S_{\max}$) W iznosi 1.

U našem primjeru "srednji rang", očekivan pod nul-hipotezom, iznosi 14 (taj se broj može dobiti i tako da se suma rangova podijeli brojem rangiranih elemenata, dakle $84/6 = 14$).

$$S = (20 - 14)^2 + (12 - 14)^2 + (8 - 14)^2 + (10 - 14)^2 + (12 - 14)^2 + (22 - 14)^2 = 160.$$

$$W = \frac{12\,S}{m^2(N^3 - N)} = \frac{12 \cdot 160}{4^2(6^3 - 6)} = \frac{1\,920}{3\,360} = 0,571.$$

Koeficijent W ne može biti negativnog predznaka, što je logički i razumljivo: za *dvojicu* sudaca moguće je zamisliti da su u negativnoj korelaciji (ako imaju suprotne rangove), ali već za *trojicu ili više* to nije moguće, nego se oni "više ili manje slažu"; ako se nikako ne slažu, W iznosi 0, a ako se potpuno slažu, W je 1.

Značajnost koeficijenta W može se očitati iz tablice F u Dodatku. U toj tablici vrijednosti lijevo od crte sasvim su točne, a desno su aproksimativne, ali sasvim zadovoljavajuće za signifikantnost od 0,05. Vrijednost W treba da bude najmanje kao vrijednost u tablici da bismo ga mogli proglasiti statistički značajnim. U našem primjeru u tablici pod $N = 6$ i $m = 4$ čitamo 0,51, a kako je naš dobiveni W veći, smatramo ga statistički značajnim.

Ako je N veći od 7, značajnost W može se izračunati i pomoću hi-kvadrat testa (o tom testu vidi poglavlje "Hi-kvadrat test"), i to prema formuli: hi-kvadrat $= m(N-1)W$, s $N - 1$ stupnjeva slobode. Ako je hi-kvadrat značajan, i W je značajan.

13.16. JOŠ NEKI KOEFICIJENTI KORELACIJE

Fi (ϕ) koeficijent. — Ako radimo s varijablama koje se raspoređuju u *dvije* očito odijeljene karakteristike (npr. živi-umrli, muškarci- žene, i sl.), ili je karakteristike nemoguće izmjeriti, pa ih podijelimo u dvije skupine (dobri-loši), služimo se tzv. ϕ (čitaj: fi) koeficijentom korelacije, koji je u čvrstoj vezi s izračunavanjem hi-kvadrat testa, pa ćemo ga zato samo ukratko spomenuti, jer se koeficijent ϕ može izravno izračunati iz χ^2 prema formuli:

$$Fi = \sqrt{\frac{\chi^2}{N}}. \tag{13.26}$$

Pretpostavimo da ispitujemo problem, postoji li i kolika je zavisnost između bračnog stanja i intelektualne defektnosti kod muškaraca i da dobijemo ove rezultate:

Inteligencija

		Normalni	Intel. defektni	Ukupno
Bračno stanje	Oženjeni	111_a	84_b	195 (a+b)
	Neoženjeni	95_c	122_d	217 (c+d)
	Ukupno	206 (a+c)	206 (b+d)	412

Primjenom formule (13.21), pri čemu izračunati χ^2 (bez korekture) iznosi 7,1, dobili bismo:

$$Fi = \sqrt{\frac{7,1}{412}} = 0,1313.$$

Koeficijent korelacije Fi možemo iz ove tablice izračunati i prema formuli:

$$Fi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}. \tag{13.27}$$

Uvrstimo li vrijednosti iz tablice u formulu (13.27), dobivamo:

$$Fi = \frac{13\,542 - 7\,980}{\sqrt{195 \cdot 206 \cdot 206 \cdot 217}} = \frac{5\,562}{\sqrt{1\,795\,679\,340}} = \frac{5\,562}{42\,375} = 0,1313.$$

Značajnost koeficijenta ϕ može se ustanoviti iz značajnosti χ^2: ako je χ^2 značajan (vidi χ^2 test, pogl. 14), značajan je i ϕ. χ^2, se iz koeficijenta ϕ može izračunati prema formuli:

$$\chi^2 = N\,\phi^2. \tag{13.28}$$

U našem slučaju $\chi^2 = 412 \cdot 0,0172 = 7,086$.

Fi koeficijent izveden je zapravo iz formule za r koeficijent korelacije, samo što u tom slučaju rezultati varijable X i varijable Y mogu poprimiti samo dvije vrijednosti, u našem primjeru "oženjeni" — "neoženjeni" ili "normalni" — "intelektualno defektni"; a tim vrijednostima daju se numeričke oznake, npr. 0 i 1.

Fi koeficijent može postići vrijednost +1 ili −1 samo u nekim slučajevima: +1 može postići samo ako totali (a+b) odgovaraju totalima (a+c), a −1 samo ako je (a+b) jednako kao i (b+d).

Koeficijent kontingencije. Ako u jednoj ili u obje varijable imamo *više* razreda, možemo korelaciju izračunati s pomoću tzv. *koeficijenta kontingencije C*, koji se, jednako kao i ϕ, dade izračunati iz χ^2, i to prema formuli:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}. \tag{13.29}$$

Primjer. Zanima nas u kakvoj je korelaciji boja očiju muške djece s bojom očiju njihovih očeva i na 1 000 mjerenja dobijemo ove rezultate:

		Boja očiju očeva				
		Modra	Siva	Maslinasta	Smeđa	Ukupno
Boja očiju sinova	Modra	194	70	41	30	335
	Siva	83	124	41	36	284
	Maslinasta	25	34	55	23	137
	Smeđa	56	36	43	109	244
	Ukupno	358	264	180	198	1 000

Budući da χ^2 izračunat iz ovog primjera (vidi str. 263) iznosi 265,554, to naš koeficijent kontingencije C iznosi — prema formulu (13.29) — $C = \sqrt{265,554/1\,265,554} = 0,46$. Ako je χ^2 značajan, značajan je i koeficijent kontingencije C.

Prednost koeficijenta C je u tome što on ne zahtijeva simetričnu raspodjelu varijabli, koje su u međusobnoj korelaciji. Njegov je nedostatak što maksimalna vrijednost C ovisi o broju kategorija (ćelija) u tablici i što ne može praktički doseći visinu od 1, pa zato ne samo da se ne može izravno usporediti s koeficijentom korelacije r, nego se teško uspoređuju i pojedine vrijednosti C, ako su dobivene iz različitog broja kategorija.

Zbog toga su svojedobno statističari izradili metode, kojima se mogao *korigirati* C, tj. približiti smislu r koeficijenta, no ta je korektura vrijedila samo za *kvadratične* tablice (dakle $2 \cdot 2$, $3 \cdot 3$ i sl.)

No danas se koeficijent C uglavnom napušta i umjesto njega statističari sve više upotrebljavaju Cramerov koeficijent, koji je u literaturi poznat pod nazivom

Cramerov Fi. Taj se koeficijent također izračunava iz hi-kvadrata, i to pomoću formule:

$$Cramerov\ Fi\ (\phi) = \sqrt{\frac{\chi^2}{N(s-1)}} \tag{13.30}$$

pri čemu s znači *manji* broj stupaca ili redova: ako se radi o $3 \cdot 2$ tablici kontingencije, $s = 2$, kod $4 \cdot 6$ tablice $s = 4$, itd. (Ako je tablica simetrična, tj. "kvadratična", npr. $3 \cdot 3$, s je naravno 3.)

Naš prijašnji primjer boje očiju otaca i sinova riješili bismo dakle ovako:

$$Cramerov\ Fi = \sqrt{\frac{265,554}{100 \cdot 3}} = \sqrt{0,088518} = 0,30$$

Dakle, dosta je niži od koeficijenta C.

Ako radimo s $2 \cdot 2$ tablicama *Cramerov Fi* svodi se na klasični *Fi* (jer $s - 1$ u nazivniku formule iznosi 1).

Za razliku od *Fi*, *Cramerov Fi* ne može se u potpunosti svesti na r-koeficijent, pa se zbog toga ne može niti interpretirati jednako kao r. No zbog činjenice da može imati sve vrijednosti između 0 i 1 (ali ne može biti negativan), lakše ga je interpretirati od koeficijenta C.

13.17. KORELACIJA IZMEĐU NOMINALNE I ORDINALNE VARIJABLE

Postoji vrlo malo statističkih postupaka kojima se može izračunavati povezanost između jedne dihotomne nominalne varijable (npr. "muškarci" — "žene") i jedne ordinalne varijable (npr. nekog ranga).

Tako bi, na primjer, mogli nekim testiranjem muškaraca i žena proučavati postoji li u uspjehu u testu razlika među spolovima. Pretpostavimo li da smo neku razliku našli, i t-test ili neki drugi test nam je pokazao da je razlika statistički značajna, sada nas dakako može zanimati *kolika* je povezanost između te dvije varijable: "spol" i "uspjeh u testu". Drugim riječima, čim nađemo da je razlika između obje skupine ispitanika statistički značajna, znači da *nije svejedno* pripada li neka testirana osoba jednoj ili drugoj skupini; a ako to nije svejedno, znači da postoji *povezanost* između spola i uspjeha u testu.

Tu korelaciju možemo izračunati uglavnom samo na dva načina, i to Kendallovim *Tau* koeficijentom i Freemanovim *Teta* koeficijentom.

Kendallov Tau. Za ovu svrhu koristi se formula za vezane rangove (13.10).

Evo primjera: Pretpostavimo da skupina od ukupno 25 muškaraca i žena (10 m i 15 ž) postigne u nekom testu ovaj rang uspješnosti:

Rang	(X)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Spol	(Y)	ž	ž	ž	ž	ž	ž	ž	ž	ž	ž	ž	m	ž	ž
		15	16	17	18	19	20	21	22	23	24	25			
		ž	m	ž	m	m	m	m	m	m	m	m			

"Prosječni" rang u varijabli "spol" računa se tako da se zbroje prvi i zadnji rang, što bi ta kategorija zauzimala kada bi rezultati obiju kategorija bili potpuno odvojeni i rezultat podijeli sa 2. Budući da imamo 15 žena, kod žena bi to dakle bilo $(1 + 15)/2 = 8$, a kod muškaraca $(16 + 25/2) = 20{,}5$. (Napomena: potpuno je svejedno koju grupu ćemo rangirati prvu: mogli smo muškarcima dati prosječni rang $(1 + 10)/2$, a ženama $(11 + 25)/2$; korelacija bi bila jednaka, samo obrnutog predznaka!).

Prema tome, sada — ako postupamo kao što se postupa kod *Tkau* korelacije — imamo ovu situaciju:

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Y	8	8	8	8	8	8	8	8	8	8	8	20,5	8	8

15	16	17	18	19	20	21	22	23	24	25
8	20,5	8	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5

Nakon toga ćemo prema već opisanom postupku izračunati S:

$S = 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 - 4 + 9 +$
$+ 9 + 9 - 1 + 8 + 0 = 140.$

Upotrijebit ćemo formulu (13.10) za izračunavanje *Tau:*

$$Tau = \frac{140}{\sqrt{[1/2\ 25\ (25-1) - T_x][1/2\ 25\ (25-1) - T_y]}}$$

Naš $T_x = 0$ (jer u varijabli X nemamo vezanih rangova), a T_y je $1/2(15 \cdot 14) + 1/2\ (10 \cdot 9) = 105 + 45 = 150$. Uvrstimo li te vrijednosti u formulu, dobivamo:

$$Tau = \frac{140}{\sqrt{300\ (300-150)}} = \frac{140}{212,13} = 0,66$$

Evo sada *kontrole* toga računa, a ujedno i dokaza da je svejedno kojim redoslijedom grupe rangiramo: prosječni rang muškaraca je $(1 + 10)/2 = 5{,}5$; prosječni rang žena je $(11 + 25)/2 = 18$.

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	18	18	18	18	18	18	18	18	18	18	18	5,5	18	18	18

16	17	18	19	20	21	22	23	24	25
5,5	18	5,5	5,5	5,5	5,5	5,5	5,5	5,5	5,5

$S = -10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 + 4 - 9 - 9$
$+ 1 - 8 + 0 = -140.$

Kako vidimo S je jednak, samo je protivnog predznaka. A predznaku ovom računu ionako nema nikakvog značenja.

Korelacija je dakle dosta visoka.

Freemanov koeficijent Teta (Θ). Poslužimo se istim primjerom od malo prije, tj. rezultatima 10 muškaraca i 15 žena u jednom testu. Postupak Freemanova testa relativno je jednostavan, i izgleda ovako:

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y	ž	ž	ž	ž	ž	ž	ž	ž	ž	ž	ž	m	ž	ž	ž	m

17	18	19	20	21	22	23	24	25
ž	m	m	m	m	m	m	m	m

Sada ćemo usporediti rang svakog pripadnika jedne od grupa (recimo žena) s rangovima svih pripadnika druge grupe. Ako započnemo s prvom ženom (rang 1) vidimo da je 10 muškaraca po rangu ispod nje (rang 12, 16, i 18-25), a nijedan muškarac nema viši rang. Rezultat za tu ženu je dakle 10 "ispod", 0 "iznad". Taj postupak nastavit ćemo s idućom ženom (10 "ispod", 0 "iznad"), trećom ženom (10 "ispod" 0 "iznad") itd.,sve do ranga 11. Žene u rangu 12 nemamo, te prelazimo na rang 13, i tu imamo 9 "ispod", 1 "iznad", itd.

Kada zbrojimo sve rezultate "ispod" i sve "iznad", dobivamo: Ukupno "iznad": $10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10$
$+ 10 + 9 + 9 + 9 + 8 = 145.$

Ukupno "ispod": $1 + 1 + 1 + 2 = 5$

Budući da nam brojnik formule (13.31) govori da treba uzeti *apsolutnu* razliku među brojevima (ne vodeći računa o predznaku), razlika između 145 i 5 je 140.

Kao što vidimo, dobili smo nama poznati izraz S iz Kendallova koeficijenta *Tau* (samo ponešto drugačijom logikom obrade rezultata).

Ako tu razliku podijelimo ukupnim brojem usporedbi, dobivamo formulu koeficijenta *Teta:*

$$Teta = \frac{|\text{ukupno "ispod"} - \text{ukupno "iznad"}|}{\text{ukupni broj usporedbi}} \qquad (13.31)$$

To iznosi $\frac{140}{115} = 0,93$.

N a p o m e n a: rezultat bi naravno bio jednak da smo započeli od druge skupine, tj. od skupine muškaraca.

Kako taj rezultat interpretirati? Ne možemo ga interpretirati analogno koeficijentu korelacije, već ovako: u 93 *posto* usporedbi ženski ispitanici postigli su bolji uspjeh od muških ispitanika.

Važno je konstatirati da se rezultati, dobiveni metodom Kendallova *Tau* i Freemanova *Teta značajno razlikuju:* dok smo u prvom slučaju dobili koef. korelacije 0,66, sada imamo 0,93 (pa ga stoga i nešto drugačije interpretiramo).

Razlika je velika, a i neki drugi računi, što ih je autor ovog teksta izveo, pokazuju da ta dva postupka daju brojčano i znatno različite rezultate, *čas u korist jednog, a čas u korist drugog izraza!* Ta činjenica, koliko god bila zbunjujuća, donekle je za neprofesionalnog statističara možda i utješna: ona pokazuje da čak i u području koje laiku izgleda izvanredno egzaktno, postoje još uvijek postupci, koji — iako je logika obrade kod svakog od njih prihvatljiva i opravdana — ipak mogu dati i značajno različite rezultate.(Ima još i važnijih teorijskih područja statistike u kojima se statističari ne slažu, i dosta se razlikuju u predloženim postupcima.)

Ako su nam rezultati grupirani u *razrede,* tj. ako u varijabli ranga imamo nekoliko vezanih rangova, dakle po nekoliko slučajeva u istom rangu, formula je drugačija (ali ona vrijedi i za slučajeve kada *nema* vezanih rangova), i glasi:

$$Teta = \frac{|\text{ukupno "ispod"} - \text{ukupno "iznad"}|}{N_1 \cdot N_2} \qquad (13.32)$$

Primjer: 8 učenika i 8 učenica napisali su zadaću iz matematike, i distribucija ocjena izgledala je ovako:

	OCJENE				
	5	4	3	2	1
Učenici	0	1	1	2	4
Učenice	2	1	2	0	3

Postupak traženja ukupno "ispod" i ukupno "iznad" bit će jednak kao i u prošlom primjeru, samo ćemo dobivene podatke morati još množiti s frekvencijom u svakom "razredu", tj. kod svake ocjene. Učinimo to s podacima našeg primjera, polazeći, recimo, od učenika:

Ukupno ispod

Za ocjenu 5	0 (nema učenika s ocjenom 5)
za ocjenu 4	$(2+3)\cdot 1 = 5$ (5 učenica je slabije od jednog učenika)
za ocjenu 3	$3\cdot 1 = 3$ (3 učenice su slabije od 1 učenika s ocjenom 3)
za ocjenu 2	$3\cdot 2 = 6$ (3 učenice su slabije od 2 učenika s ocjenom2)

Ukupno "ispod" = 14

Ukupno iznad

za ocjenu 4	$2\cdot 1 = 2$
za ocjenu 3	$(1+2)\cdot 1 = 3$
za ocjenu 2	$(2+1+2)\cdot 2 = 10$
za ocjenu 1	$(2+1+2)\cdot 4 = 20$

Ukupno "iznad" = 35

$$Teta = \frac{|\text{ukupno "ispod"} - \text{ukupno "iznad"}|}{N_1 \cdot N_2} = \frac{|14-35|}{8\cdot 8}$$

$$Teta = 21/64 = 0,33.$$

Inspekcija tablice pokazuje da učenice tendiraju boljim ocjenama u toj zadaći od učenika, tj. da su one u 33% slučajeva imale bolju ocjenu, nego što su imale lošiju ocjenu.

Primijenimo li formulu 13.32 na naš prvi primjer, dobivamo:

$$Teta = \frac{145-5}{15\cdot 10} = 0,93.$$

N a p o m e n a: Ako je nekom čitaocu možda upalo u oči da smo kod druge formule (13.32) postupili prividno drugačije, nego kod prve (13.31) tj. dok je brojnik u maloprijašnjem primjeru bio 14 – 35, nazivnik *nije* glasio 14 + 35 (=49) kao u prvom primjeru, onda valja kazati ovo: u formuli (13.32) brojnik se u svakom slučaju može izračunati kao i kod formule (13.31), jer to je stvarno frekvencija rezultata koji su "iznad" i "ispod"; ali u nazivniku u kojem treba uzeti *ukupan broj usporedbi*, jednostavnim *zbrajanjem* "iznad" i "ispod" mi smo *izostavili* određeni broj usporedbi, tj. sve one, gdje u obje skupine ispitanika rezultati padaju u istu kategoriju: to su *jedna* usporedba kod ocjene 4 (jedan učenik s jednom učenicom), *dvije* usporedbe kod ocjene 3 (jedan učenik s dvije učenice), i *dvanaest* usporedbi kod ocjene 5 (4 učenika s 3 učenice), dakle ukupno 15 usporedbi. A 49 + 15 = 64, dakle točno toliko koliko smo dobili i množenjem broja učenika s brojem učenica: $8\cdot 8 = 64$.

Ako se radi o izboru između *Tau i Teta,* čini se da ipak treba preferirati upotrebu Freemanova *Teta*. Usporedbe između ta dva postupka pokazuju ovo: kod *potpune* povezanosti *Teta* daje očekivani rezultat, tj. 1, bez obzira na veličinu uzoraka. Naprotiv, *Tau* se smanjuje što je uzorak veći (čak do oko 0,70) , a najviša vrijednost (0,82) postignuta je kod $N = 4$. A kod *nikakve* povezanosti oba su načina jednako loša: kod malih uzoraka rezultat je vrlo iskrivljen, i tek kod velikih uzoraka rezultat se približava nuli (koliko bi zapravo moralo biti).

13.18. RAZLIKA IZMEĐU DVA KOEFICIJENTA KORELACIJE r

Katkada nas može zanimati razlikuju li se dva koeficijenta korelacije statistički značajno. Na primjer, ako korelacija između visine i težine kod 20 dvanaestogodišnjih dječaka (vidi primjer na str. 192 iznosi + 0,86, a kod 40 odraslih muškaraca korelacija između visine i težine = + 0,51, može nas zanimati je li ta razlika statistički značajna, jer tek ako ona to jest, možemo pokušati analizirati uzroke toj razlici.

Testiranje značajnosti razlike između dva koeficijenta korelacije r relativno je jednostavno i izvodi se ovako:

1. Najprije treba oba koeficijenta korelacije pretvoriti u korespondentne vrijednosti z_r, koje — po njihovu autoru R. A. Fisheru — nazivamo i Fisherove z_r vrijednosti. Distribucija uzoraka koeficijenta r nije, naime, normalna nego asimetrična distribucija, a pretvaranjem u z_r vrijednosti, ona postaje približno normalna. Iz tablice G u Dodatku možemo očitati da koeficijentu korelacije $r_1 = 0,86$ odgovara $z_{r_1} = 1,293$, a koeficijentu $r_2 = 0,51$ odgovara $z_{r_2} = 0,563$.

2. Standardna pogreška z_r vrijednosti je, kako rekosmo, približno normalno distribuirana i ona se računa prema formuli:

$$s_{z_r} = \frac{1}{\sqrt{N-3}}. \tag{13.33}$$

Prema tome, standardnu pogrešku razlike između obje z_r vrijednosti izračunat ćemo prema već dobro poznatom principu, tj. ona je drugi korijen iz zbroja obiju kvadriranih standardnih pogrešaka:

$$s_{z_{r_1}-z_{r_2}} = \sqrt{s^2_{z_{r_1}} + s^2_{z_{r_2}}} \tag{13.34}$$

$$= \sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}. \tag{13.35}$$

U našem slučaju dobivamo:

$$s_{z_{r_1}-z_{r_2}} = \sqrt{\frac{1}{17} + \frac{1}{37}} = 0,293.$$

3. Podijelimo li razliku između z_{r_1} i z_{r_2} sa standardnom pogreškom te razlike, dobivamo:

$$t = \frac{z_{r_1} - z_{r_2}}{\sqrt{1/(N_1-3) + 1/(N_2-3)}} = \frac{1,293-0,563}{0,293} = 2,49.$$

Taj rezultat interpretiramo jednako kao što smo interpretirali i razlike među aritmetičkim sredinama: budući da je t *veći* od 1,96, možemo smatrati da je razlika između oba koeficijenta korelacije statistički značajna. Tek kad smo to dokazali, možemo pokušati tu razliku interpretirati. Značajno veću korelaciju između težine i visine možemo kod djece — za razliku od odraslih — naći zato što su djeca većinom građena proporcionalnije, pa je u velikom broju slučajeva dijete koje je više, ujedno i teže. Naprotiv, kod odraslih je proporcionalnost tijela znatno rjeđa (jer nalazimo veći broj debelih ljudi), pa je zbog toga i korelacija manja.

Međutim, ako mjerimo *istu* grupu ispitanika i dobijemo dvije različite korelacije, pa želimo testirati razlikuju li se one značajno, postupak je drugačiji.

Na primjer, ustanovili smo na skupini od 100 ispitanika da je korelacija između uspjeha u jednom dugačkom i teško primjenjivom testu X i uspjeha u zvanju Z $r_{xz} = +0{,}56$, dok je kod tih istih 100 ispitanika korelacija između uspjeha u jednom kratkom i jednostavnom testu Y i uspjeha u istom zvanju Z $r_{yz} = +0{,}43$. Kad bi testovi bili jednako lako primjenjivi, naravno da bi se odlučili za prvi test, bez obzira na to je li razlika u korelaciji značajna. Ali u ovoj situaciji skloni smo da se odlučimo za prvi komplicираniji test jedino ako je njegova korelacija s uspjehom u zvanju statistički značajno veća od korelacije drugog testa.

Za ovaj račun potrebno je znati i korelaciju između oba testa. Uzmimo da ona iznosi $r_{xy} = +{,}52$.

Testiranje značajnosti razlike među korelacijama u ovim se slučajevima izračunava ovako:

$$t = (r_{xz} - r_{yz})\sqrt{\frac{(N-3)(1+r_{xy})}{2(1-r_{xy}^2 - r_{xz}^2 - r_{yz}^2 + 2r_{xy}r_{xz}r_{yz})}}. \qquad (13.36)$$

Uzvrstimo li naše vrijednosti u ovu formulu, dobivamo:

$$t = (0,56 - 0,43)\sqrt{\frac{97 \cdot 1,52}{2(1 - 0,2704 - 0,3136 - 0,1849 + 1,04 \cdot 0,56 \cdot 0,43)}}$$

$$= 0,13\sqrt{153,10} = 1,61.$$

Ovaj t ima t-raspodjelu s $N-3$ stupnja slobode. Granična vrijednost t za 97 stupnjeva slobode iznosi oko 1,99. Budući da je naš dobiveni t manji, zaključujemo da ne možemo razliku između korelacija testova X i Y i uspjeha u zvanju Z smatrati statistički značajnom.

ZADACI ZA VJEŽBU

1. Izračunajte r koeficijent korelacije za rezultate iz tablice 13.1. Provedite kontrolu računa stupcima $(X-Y)$ i $(X-Y)^2$.

2. 20 studenata testirano je jednim testom znanja i jednim testom verbalnog faktora; dobiveni su ovi rezultati:

Stud.	Test znanja (X)	Verb. test (Y)	Stud.	Test znanja (X)	Verb test (Y)
A	52	49	K	64	53
B	49	49	L	28	17
C	26	17	M	49	40
D	28	34	N	43	41
E	63	52	O	30	15
F	44	41	P	65	50
G	70	45	R	35	28
H	32	32	S	60	55
I	49	29	T	49	37
J	51	49	U	66	50

Izračunajte r korelaciju između ova dva niza podataka.

3. Kod 25 košarkaša uspoređen je broj koševa koje su dali, s brojem prekršaja ("osobnih grešaka"), koje je svaki od njih imao.

Igrač	Broj koševa	Broj prekrš.	Igrač	Broj koševa	Broj prekrš.	Igrač	Broj koševa	Broj prekrš.
1	0	0	9	2	4	17	37	16
2	0	0	10	3	0	18	42	24
3	0	1	11	6	6	19	46	20
4	0	1	12	7	3	20	48	17
5	1	0	13	9	9	21	57	22
6	2	0	14	18	9	22	59	18
7	2	1	15	21	7	23	60	15
8	2	3	16	35	5	24	75	24
						25	75	31

Prikažite ove rezultate grafički (na apscisu stavite broj koševa) i izračunajte r koeficijent.

4. U jednoj skupini od 10 radnika ispitano je upitnikom njihovo zadovoljstvo na poslu i stupanj informiranosti. Dobiveni su ovi rezultati:

Radnik	Zadovoljstvo	Informiranost
1	45	46
2	41	37
3	40	45
4	36	21
5	34	37
6	33	32
7	25	30
8	21	19
9	18	26
10	17	17

Izračunajte za ove podatke r koeficijent i ρ koeficijent.

5. Izračunajte rang-korelaciju iz podataka iz tablice 13.5.

6. Na jednoj skupini školske djece korelacija između dobi i težine iznosila je $r = 0,80$, između dobi i rezultata u jednom aritmetičkom testu 0,60, i između težine i rezultata u testu 0,50. Kolika je korelacija između težine i rezultata u testu ako isključimo utjecaj godina?

7. U jednom turističkom mjestu r-korelacija između broja turista i količine prodanih vrućih čokoladnih napitaka iznosila je — 0,30. Korelacija između prodanih količina čokoladnih napitaka i okolne temperature iznosila je — 0,70, a između broja turista i temperature + 0,80. Kolika je stvarna korelacija između broja turista i količine prodanih vrućih čokoladnih napitaka, pod uvjetom konstantne temperature?

8. Na jednom ispitu korelacija između bodova i nekog testa iznosila je 0,50. Korelacija između bodova i jednoga drugog testa bila je 0,60. Međusobna korelacija između oba testa bila je 0,40. Kolika je multipla korelacija između oba testa zajedno i bodova na ispitu?

9. Deset radnika rangirano je prema ovim kriterijima: marljivost (M), produktivnost (P), točnost u radu (T), discipliniranost (D) i odnos prema drugovima (O). Dobiveni su ovi rangovi:

Radnici	M	P	T	D	O
A	2	1	2	3	4
B	1	3	1	2	2
C	3	4	4	1	3
D	5	5	5	5	1
E	4	2	6	7	6
F	7	8	3	4	7
G	6	6	8	6	5
H	8	7	7	8	9
I	9	10	10	9	8
J	10	9	9	10	10

Koliko je slaganje između ovih 5 kriterija?

10. Skupina od 20 studenata rješavala je najprije test znanja iz statistike, a nakon toga je svaki polagao usmeni ispit. Na tom ispitu mnogi su pali. Dolje su navedeni za svakog studenta rezultati u testu i rezultat ispita: 1 = prošao, 0 = pao. Izračunajte point-biserijalnu korelaciju između ove dvije varijable.

Stud.	Test	Ispit	Stud.	Test	Ispit	Stud.	Test	Ispit
1	50	1	8	43	0	15	44	1
2	46	0	9	42	0	16	42	0
3	37	0	10	46	1	17	41	0
4	48	1	11	42	0	18	45	0
5	47	1	12	41	0	19	46	0
6	40	0	13	43	0	20	43	0
7	39	0	14	43	0			

11. U poglavlju 9.9. rastumačena je "metoda diferencije", tj. testiranje značajnosti razlike između dvije aritmetičke sredine kod malih zavisnih uzoraka. Kao primjer uzeti su rezultati za bazalni metabolizam 18 ispitanika prije i nakon uzimanja fenamina. *Korelacija* između prvoga i drugoga mjerenja bazalnog metabolizma iznosi $r = +0,69$. Izračunajte uz pomoć standardne metode, tj. uz pomoć formule 9.7, vodeći brigu o tome da su to *mali* uzorci (zajednička standardna devijacija!), značajnost razlike uz pomoć t-testa.

14. PROGNOZA IZ JEDNE VARIJABLE U DRUGU

14.1. PRAVAC REGRESIJE

Kada jednom znamo da između dviju pojava postoji korelacija, to u praksi može vrlo mnogo koristiti. Na primjer, kada je jednom ustanovljeno da postoji (negativna) povezanost između količine fluora u pitkoj vodi i "pokvarenih" zubi kod djece, tj. da je broj karijesa to manji, što je veća koncentracija fluora u vodi, to je dovelo (kad je provjereno da među varijablama stvarno postoji *uzročna* veza!) do dodavanja floura pitkoj vodi u gradskim vodovodima.

No, često nije dovoljno znati samo to da korelacija postoji, i kolika je, već želimo iz podataka jedne varijable zaključiti koji joj rezultat najvjerojatnije odgovara u drugoj varijabli, tj. želimo da iz podataka jedne varijable *prognoziramo rezultat u drugoj varijabli*. Na primjer, ako znamo da između visine i težine postoji korelacija određenog stupnja, zanima nas koja je najvjerojatnija težina čovjeka, koji je, primjerice, visok 180 cm; ili, ako znamo stupanj povezanosti između količine proljetnih kiša i količine jesenske žetve, može nas zanimati prognoza o tome kakva će biti žetva neke godine s obzirom na registriranu količinu oborina u proljeće.

Ako je povezanost maksimalna (+ 1 ili - 1), onda nema problema, tj. korespondentnu vrijednost zavisne varijable (Y) izračunat ćemo bez teškoća iz neke vrijednosti nezavisne varijable (X).

Primjer. Znamo da je opseg kruga $2\,r\,\pi$. Dakle, opseg kruga, kojemu polumjer iznosi 5 cm, bit će $2 \cdot 5 \cdot 3,14 = 31,4$. Ako je odnos ujedno i linearan (a u opsegu kruga je tako), onda se - kako znamo - sve vrijednosti obiju varijabla nalaze na *pravcu* od kojeg nema nikakvih odstupanja.

Na primjer, u spomenutom primjeru s opsegom kruga za ovih 5 polumjera imamo ove opsege:

Polumjer	Opseg	Porast
1	$2 \cdot 1 \cdot 3,14 = 6,28$	
2	$2 \cdot 2 \cdot 3,14 = 12,56$	6,28
3	$2 \cdot 3 \cdot 3,14 = 18,84$	6,28
4	$2 \cdot 4 \cdot 3,14 = 25,12$	itd.
5	$2 \cdot 5 \cdot 3,14 = 31,40$	

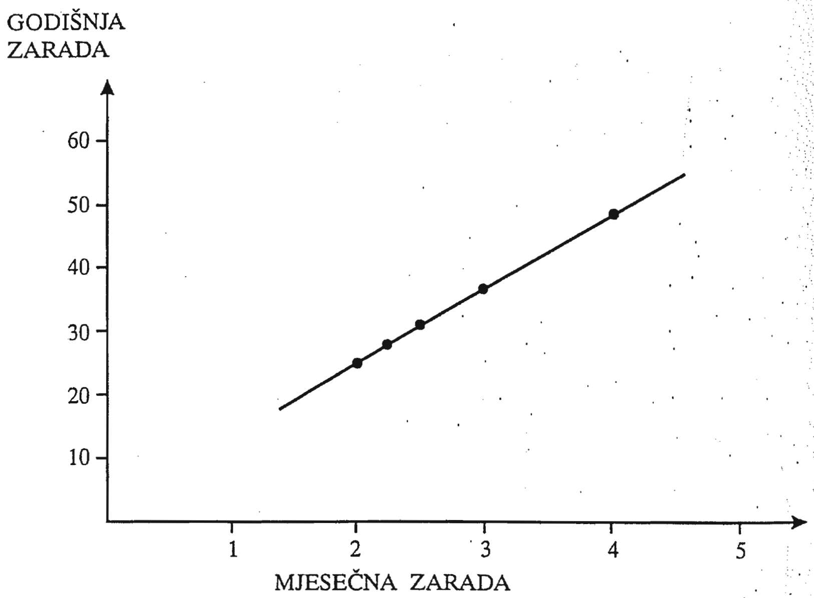
Kako se vidi, linearnom porastu polumjera odgovara linearni porast opsega koji neprestano raste za istu vrijednost, tj. za 2π, ili 6,28.

Evo još jednog primjera:

Mjesečne zarade petorice ljudi prikazane su u varijabli X, a godišnje zarade u varijabli Y:

Službenik	Mjesečna zarada (u nekim arbitrarnim jedinicama)	Godišnja zarada
A	2	24
B	2,25	27
C	2,50	30
D	3	36
E	4	48

Ako odnos između obiju varijabli prikažemo grafički, dobivamo rezultat prikazan na slici 14.1.



Slika 14.1. Odnos između mjesečne i godišnje zarade petorice službenika

Budući da godišnja zarada iznosi dvanaest puta iznos mjesečnih zarada, godišnju zaradu možemo jednostavno izraziti kao

$$Y = 12\,X.$$

Prema tome, neki novi službenik, s mjesečnom zaradom od 2,75, imao bi godišnje

$$Y = 12 \cdot 2,75 = 33.$$

I naš prvi primjer s opsegom kruga također bismo mogli prikazati analognom jednadžbom, koja bi glasila:

$$Y = 6,28\,X.$$

Ovi izneseni rezultati dadu se - kako smo vidjeli na slici 14.1. — grafički prikazati pravcem, kome je ishodište u $X = 0$ i $Y = 0$ (jer ako je mjesečna zarada 0 i godišnja je zarada 0 ili ako je polumjer 0 i opseg kruga je 0).

Međutim, ako svaki od spomenutih 5 službenika dobiva godišnje nagradu, koja je za sve njih jednaka, i iznosi 3 jedinice, onda tu vrijednost od 3 treba pribrojiti svakoj godišnjoj plaći, pa ćemo tako dobiti formulu:

$$Y = 3 + 12\,X, \text{ ili u simbolima:}$$

$$Y = a + bX, \tag{14.1}$$

a to je ujedno i *"jednadžba pravca regresije"*, u kojoj su a i b konstante: a označuje *odsječak* na osi Y, a b označuje *nagib* pravca. Nagib pravca je porast (ili pad) u varijabli Y za promjenu varijable X za jednu jedinicu. Ili, nešto egzaktnije, nagib je

$$\frac{\text{vertikalna udaljenost}}{\text{horizontalna udaljenost}} = \frac{Y_2 - Y_1}{X_2 - X_1}.$$

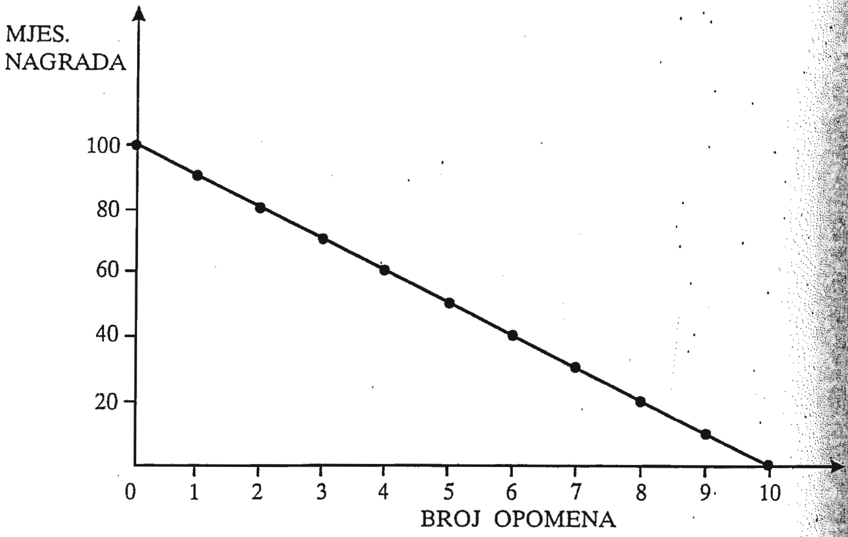
U našem primjeru s plaćom nagib pravca je 12, tj. za porast mjesečne plaće za 1 (horizontalna udaljenost) godišnja plaća poraste 12 tih istih jedinica, prema tome, imamo izraz $12/1 = 12$. A u slučaju s opsegom kruga, nagib je 6,28, jer za porast polumjera od 1 opseg raste za 6,28.

Početnike (pa i naprednije!) u statistici obično zbunjuje *naziv* ovog pravca, tj. nije baš posve jasno zašto se taj pravac često tajanstveno naziva pravac *regresije*. Uzrok tome je međutim vrlo banalan: već smo spomenuli da se Galton bavio pitanjem odnosa između visine otaca i sinova, i da je time postao zapravo začetnik računa korelacije. Prilikom svojih istraživanja Galton je ustanovio — što se i moglo očekivati — da visoki očevi imaju (u prosjeku) i visoke sinove; no istodobno on je primijetio i nešto neobično: sinovi visokih očeva bili su manje visoki nego što bi se to moglo očekivati, a isto tako sinovi niskih očeva bili su manje niski od očekivanog. Drugim riječima, očevi koji su, na primjer, bili 15 cm viši od prosjeka očeva, imali su sinove koji su (u prosjeku) bili samo 8 cm viši od prosječne visine sinova, a očevi koji su bili oko 10 cm niži od prosječne visine očeva, imali su sinove koji su (u prosjeku) bili oko 5 cm niži od prosječne visine sinova.

Na osnovi tih rezultata Galton je pretpostavio postojanje "zakona o regresiji (vraćanju) prema prosječnosti": s *ekstremnim* vrijednostima varijable X povezane su *manje ekstremne* vrijednosti varijable Y; dakle, kao da Y-vrijednosti tendiraju vraćanju (regresiji) prema svom prosjeku.

Eto, i tako je nastao naziv pravca "regresije", koji se - bez obzira na točnost ili netočnost tvrdnje o "vraćanju prema prosjeku" — zadržao u statistici.

No vratimo se našem nagibu krivulje! On može naravno biti i *negativan* (u slučaju negativne korelacije), tj. ako porastu vrijednosti X odgovara *pad* vrijednosti Y. Na primjer, ako u jednom kolektivu broj *opomena* donosi sa sobom smanjenje mjesečne plaće, uzmimo, za 10 jedinica, a mjesečna plaća iznosi 100 jedinica, onda će službenik koji u toku mjeseca nije imao ni jednu opomenu dobiti naknadu od 100, a netko tko je imao 10 opomena, neće dobiti nikakvu naknadu. Taj odnos grafički izgleda kao na slici 14.2.



Slika 14.2. Odnos između broja opomena i mjesečne nagrade

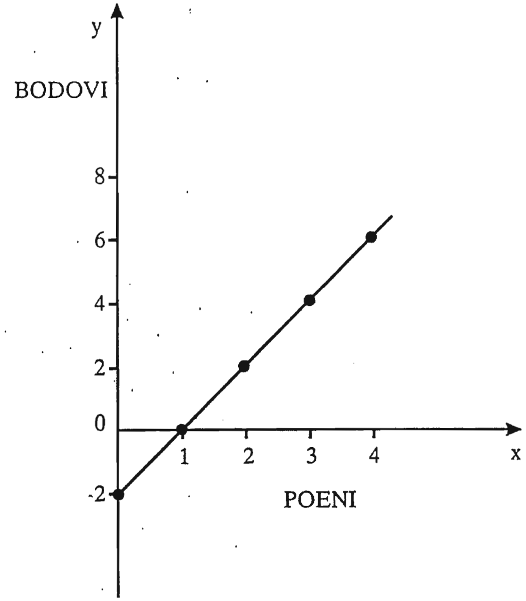
Jednadžba pravca regresije za ovaj će slučaj iznositi:

$$Y = a + bX = 100 + (-10)X = 100 - 10X.$$

Dakle, nagib iznosi - 10. To proizlazi iz prije spomenute formule: $\frac{80 - 90}{3 - 2} = -10.$
Iz jednadžbe, dakle, slijedi da će netko, tko je imao 4 opomene, dobiti nagradu u visini od 60:

$$Y = 100 - 10 \cdot 4 = 60.$$

Odsječak na ordinati može biti i *negativan*. Na primjer, ako se u nekoj igri postizanje *iznad* 1 poena nagrađuje s 2 boda za svaki poen, 1 poen ne donosi nagrade, a *nula* poena se kažnjava s 2 *negativna* boda, onda imamo situaciju prikazanu na slici 14.3.



Slika 14.3. Odnos između postignutih poena i nagrade odnosno kazne

Ovu situaciju odražava donja formula pravca regresije:

$$X = -2 + 2X.$$

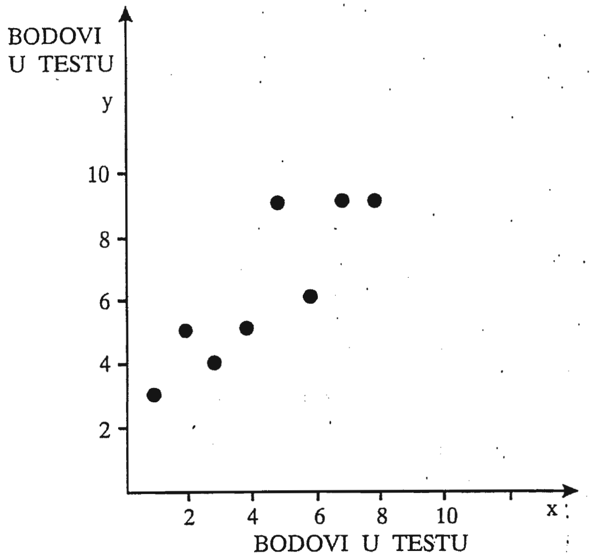
– – –

Sve ovo do sada bilo je lako izračunati jer su svi rezultati bili *točno na pravcu*. No u praksi nikada nije tako (osim naravno kod matematičkih zakona). Stoga se možemo zapitati kako ćemo postaviti prognozu iz neke vrijednosti varijable X za varijablu Y.

Pretpostavimo da 8 ispitanika u testu X i testu Y postignu ove rezultate:

Ispitanici	X	Y
A	1	3
B	2	5
C	3	4
D	4	5
E	5	9
F	6	6
G	7	9
H	8	9

Prikažemo li te rezultate grafički, dobit ćemo rezultat prikazan na slici 14.4.



Slika 14.4. Grafički prikaz rezultata osmorice ispitanika u testovima X i Y.

Kako svi rezultati *nisu* na pravcu, sada je teško reći koji je rezultat u varijabli Y najvjerojatniji za vrijednost X od, recimo, 3. U ovom konkretnom slučaju vrijednosti $X = 3$ odgovara rezultat $Y = 4$, ali te su vrijednosti dobivene na malom uzorku, i logički slijedi već iz slike 14.4. da bismo uz $X = 3$ mogli imati i koji drugi rezultat u varijabli Y.

Kako dakle nacrtati pravac koji bi "pošteno" reprezentirao odnos između obje varijable? Ako bismo to učinili "odoka", to bi moglo biti i prilično pogrešno, jer bi neka druga osoba vjerojatno ucrtala neki drugi pravac, treća osoba opet neki drugi, itd.

No matematičari su naravno našli rješenje ovog problema: "najpošteniji" pravac regresije je onaj *koji ima najmanju sumu kvadrata odstupanja* pojedinačnih Y rezultata od tog pravca. Ako vrijednosti na tom pravcu nazovemo $\tilde{Y}$ (zovu ga "procijenjeni", ili "predviđeni", ili "regresijski" Y), onda ćemo za svaki individualni rezultat (tj. za svaku točku u scatter dijagramu) dobiti neku razliku $(Y - \tilde{Y})$. Suma kvadrata tih razlika (tj. udaljenosti od pravca) treba biti najmanja od svih mogućih. Stoga se metoda kojom se to postiže naziva "metoda najmanjih kvadrata".

Jasno je da takav pravac nećemo tražiti metodom "pokušaja i pogrešaka", jer bismo na to mogli utrošiti dobar dio života, već postoje mnogo jednostavniji i kraći

načini za izračunavanje konstanti a i b u jednadžbi pravca $Y = a + bX$ (odnosno, sada je bolje pisati $\tilde{Y} = a + bX$).

Matematički je nađeno da se konstanta b može izračunati uz pomoć formule:

$$b = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}. \tag{14.2}$$

a konstanta a izračunava se uz pomoć formule:

$$a = \overline{Y} - b \cdot \overline{X} \tag{14.3}$$

($\overline{Y}$ = aritmetička sredina varijable Y, a $\overline{X}$ = aritmetička sredina varijable X).

Ako naših 8 rezultata unesemo u tablicu za izračunavanje potrebnih vrijednosti (to su *iste* vrijednosti koje smo računali i kod koeficijenta korelacije r!), dobit ćemo podatke prikazane na lijevoj strani tablice 14.1.

TABLICA 14.1.
POSTUPAK ZA IZRAČUNAVANJE PODATAKA ZA PRAVAC REGRESIJE I STANDARDNU POGREŠKU PROGNOZE

Ispitanik	X	Y	X^2	Y^2	XY	$\tilde{Y}$	$Y - \tilde{Y}$	$(Y - \tilde{Y})^2$
A	1	3	1	9	3	3,25	−0,25	0,06
B	2	5	4	25	10	4,11	0,89	0,79
C	3	4	9	16	12	4,96	−0,96	0,92
D	4	5	16	25	20	5,82	−0,82	0,67
E	5	9	25	81	45	6,68	2,32	5,38
F	6	6	36	36	36	7,54	-1,54	2,37
G	7	9	49	81	63	8,39	0,61	0,37
H	8	9	64	81	72	9,25	-0,25	0,06
Σ	36	50	204	354	261	50,0	0,0	10,62
$r = 0,862$	$\overline{X} = 36/8 = 4,5$ $\overline{Y} = 50/8 = 6,25$							

Unesemo li dobivene vrijednosti u formulu (14.2), dobivamo:

$$b = \frac{(8 \cdot 261) - (36 \cdot 50)}{(8 \cdot 204) - (36 \cdot 36)} = \frac{288}{336} = 0,857.$$

Konstantu a izračunat ćemo prema formuli (14.3):

$$a = 6,25 - (0,857 \cdot 4,5) = 2,39.$$

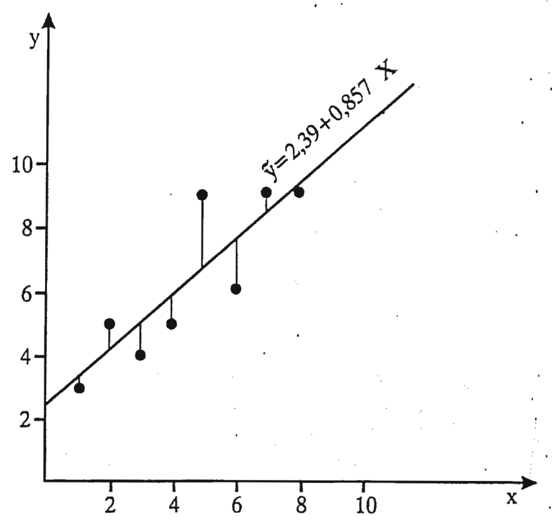
Naša jednadžba pravca regresije, prema tome, glasi:

$$\tilde{Y} = 2,39 + 0,857X.$$

Sada možemo izračunati nekoliko vrijednosti $\tilde{Y}$ (dovoljno bi bilo dvije, dosta udaljene) da bismo mogli nacrtati pravac:

X	$\tilde{Y}$	
0	2,39	(iz 2,39 +0)
2	4,11	(iz 2,39 +0,857 · 2)
6	7,54	(iz 2,39 + 0,857 · 6)
9	10,11	(iz 2,39 + 0,857 · 9).

Ako kroz dobivene točke povučemo pravac, dobit ćemo pravac regresije, prikazan na slici 14.5.



Slika 14.5. Pravac regresije za rezultate iz tablice 14.1.

Sve izračunate vrijednosti $\tilde{Y}$ za svaki pojedini X su dakako na tome pravcu, i te su vrijednosti u tablici 14.1, prikazane na desnoj strani tablice, u stupcu $\tilde{Y}$. Na slici 14.5. i u tablici 14.1. nacrtane su, tj. u stupcu $Y - \tilde{Y}$ izračunate, razlike između stvarnih rezultata varijable Y i "prognoziranog" $\tilde{Y}$. Iz tablice se ujedno vidi da suma $\tilde{Y}$ vrijednosti odgovara sumi Y, a također i da je suma razlika ravna *nuli*. (N a p o m e n a. Zbog decimalnog zaokruživanja to ne mora uvijek biti *točno* nula!)

Uz bilo koji drugi pravac, zadnji stupac (tj. suma kvadriranih razlika, dakle $\Sigma(Y - \tilde{Y})^2$) bio bi *veći* od dobivene sume 10,62, jer 10,62 je najmanja moguća suma kvadrata razlika.

Budući da se računa s istim podacima kao i koeficijent korelacije r, pravac regresije možemo izračunati uvijek kada računamo i korelaciju.

N a p o m e n a. Jednako tako kao što smo računali pravac regresije za varijablu Y, jer želimo iz nezavisne varijable X prognozirati vrijednosti varijable Y, možemo, dakako, na analogan način izračunati i pravac regresije za varijablu X, ako želimo prognozirati u obratnom smjeru, tj. iz Y u X. *To nisu dva jednaka pravca* (osim u slučaju korelacije +1), i oni se sijeku pod to većim kutom, što je korelacija između obje varijable niža. (Kod korelacije 0, pravci su jedan na drugom okomiti.)

Na ovom mjestu možda je umjesno postaviti pitanje: zašto za prognozu vrijednosti Y iz neke vrijednosti X koristimo pravac regresije, a ne prosječan rezultat, što ga uz određeni X imaju sve vrijednosti varijable Y? Tomu je nekoliko razloga: prvo, često se uz neki konkretni rezultat X ne nalazi niti jedan rezultat Y, pa u tom slučaju ne bismo mogli prognozirati najvjerojatniji Y. Drugo, i mnogo važnije: sa stanovišta *stabilnosti* uzorka pravac regresije je mnogo stabilnija mjera od bilo koje druge vrijednosti, koja označuje odnos između obje varijable, i to zato što se pravac regresije zasniva na *svim* rezultatima scatter dijagrama. Prema tome, prognoza, osnovana na pravcu regresije, neće biti toliko pod utjecajem slučajnih fluktuacija u uzorku koliko bi to bila prosječna vrijednost svih rezultata Y uz neki određeni X.

N a p o m e n a. Da smo imali na raspolaganju cijelu populaciju, onda bi bilo opravdano da neki određeni $\tilde{Y}$ računamo iz *prosječne* vrijednosti svih Y uz neku određenu vrijednost X.

Još nekoliko riječi o tome kako izračunavamo pravac regresije ako imamo mnogo rezultata, kao i onda ako imamo rezultate grupirane u razrede.

U tom je slučaju znatno jednostavnije koristiti jednu drugu formulu, koja se koristi podacima o korelaciji između obje varijable, kao i podacima o aritmetičkim sredinama i standardnim devijacijama. Ta formula glasi:

$$\tilde{Y} = r\frac{s_y}{s_x}(X - \overline{X}) + \overline{Y}. \tag{14.4}$$

Na primjer, kada bi u nekom istraživanju povezanosti između dvije varijable imali ove rezultate:

$$\overline{X} = 119 \qquad \overline{Y} = 1,30$$
$$s_x = 10 \qquad s_y = 0,55$$
$$r = 0,70$$
$$N = 100,$$

onda bi prognoziranu vrijednost $\tilde{Y}$ za, primjerice, $X = 130$ izračunali primjenom formule (14.4) ovako:

$$\begin{aligned}\tilde{Y} &= 0,70\frac{0,55}{10}(130 - 119) + 1,30\\ &= (0,0385 \cdot 11) + 1,30\\ &= 0,42 + 1,30\\ &= 1,72.\end{aligned}$$

14.2. POGREŠKA PROGNOZE

Posve je razumljivo da prognoza koju smo izveli uz pomoć pravca regresije, ne može biti sasvim točna; naprotiv, ona će biti to manje točna, što je niža korelacija između obje pojave, pa su, prema tome, individualni rezultati *jako raspršeni* oko pravca regresije. Iz toga slijedi da pri računanju pogreške prognoze moramo uzeti u obzir raspršenje rezultata oko pravca regresije. Pri tome moramo pretpostaviti (a to ne mora uvijek biti tako!) da je *raspršenje rezultata* oko pravca regresije manje-više podjednako uz čitavu duljinu pravca (to se naziva "homoscedascitet").

Kako se i u ovom slučaju radi o sličnom fenomenu kao kod standardne devijacije, to je i formula za izračunavanje pogreške prognoze (koja se naziva "standardna pogreška prognoze") slična formuli za standardnu devijaciju. Ona glasi:

$$\text{Standardna pogreška prognoze} = \sqrt{\frac{\Sigma(Y-\tilde{Y})^2}{N-2}}. \qquad (14.5)$$

N a p o m e n a. Ako znamo korelaciju između obje varijable, pogrešku možemo izračunati i uz pomoć ove formule:

$$\text{Standardna pogreška prognoze} = s_y\sqrt{(1-r^2)\frac{N-1}{N-2}}. \qquad (14.6)$$

U našem primjeru iz tablice 14.1. možemo se npr. zapitati koji je najvjerojatniji $\tilde{Y}$ za $X = 4$, pa ćemo iz jednadžbe pravca dobiti:

$$\tilde{Y} = 2,39 + 0,857 \cdot 4 = 5,818,$$

a standardna pogreška te prognoze je:

$$\sqrt{\frac{10,62}{6}} = \sqrt{1,77} = 1,33.$$

Vidjeli smo da ova "standardna pogreška" po svom *smislu* odgovara zapravo *standardnoj devijaciji* i ne treba je miješati s izrazom "standardna pogreška" kod aritmetičke sredine, jer se tamo radi o variranju aritmetičkih sredina *uzoraka* oko prave aritmetičke sredine, a kod standardne devijacije radi se o variranju originalnih rezultata oko njihove aritmetičke sredine. Upravo ovdje imamo istu stvar: nas zanima variranje konkretnih rezultata oko pravca regresije.

Kako ćemo, dakle, interpretirati standardnu pogrešku prognoze, za koju smo izračunali da iznosi 1,33? Jednako kao i standardnu devijaciju, a to znači da 68% rezultata Y varijable varira za $\pm 1,33$ oko vrijednosti $\tilde{Y}$, 95% rezultata varira za $\pm 2 \cdot 1,33 = \pm 2,66$ oko pravca, itd. Dakle, *najvjerojatniji* rezultat varijable Y uz rezultat $X = 4$ je rezultat 5,818, ali je 95% vjerojatno da se on nalazi u intervalu između 3,16 i 8,48 (što je $5,818 \pm 2 \cdot 1,33$).

Pogrešku prognoze bilo bi naporno računati kod velikog broja rezultata kada bismo morali računati svaku razliku između Y i $\tilde{Y}$, pa se izraz $\Sigma(Y-\tilde{Y})^2$ u tom slučaju izračunava pomoću formule:

$$\Sigma(Y-\tilde{Y})^2 = 1/N\left\{N\Sigma Y^2 - (\Sigma Y)^2 - \frac{[N\Sigma XY - (\Sigma X)(\Sigma Y)]^2}{N\Sigma X^2 - (\Sigma X)^2}\right\}. \qquad (14.7)$$

Primijenimo li tu formulu na naš primjer iz tablice 14.1 (za koji smo već izračunali pogrešku prognoze), dobivamo:

$$\Sigma(Y-\tilde{Y})^2 = 1/8\left\{(8 \cdot 354) - 50^2 - \frac{[(8 \cdot 261) - (36 \cdot 50)]^2}{(8 \cdot 204) - 36^2}\right\} = 10,64.$$

(U računu "pješice" dobili smo 10,62.)

Na kraju treba još jednom posebno upozoriti da prognoziranje nekog rezultata Y iz jednog određenog rezultata X (jednako kao i računanje koeficijenta korelacije r), ima svoje opravdanje *jedino ako su obje varijable u linearnom odnosu*, a donekle i onda ako postoji "homoscedascitet" (tj. ako je varijabilitet rezultata Y oko pravca regresije podjednak duž cijelog pravca.)

Zato je koristan savjet (ako nemamo odveć mnogo rezultata) da kako pri računanju korelacije, tako i pri računanju pravca regresije radi prognoze nacrtamo i scatter dijagram rezultata: na toj ćemo slici zaista vidjeti "više nego na 1 000 brojeva", tj. vidjet ćemo: 1. je li odnos linearan i 2. nije li ozbiljno deformiran homoscedascitet (a on se može ozbiljno poremetiti ako bilo jedna, bilo druga, bilo obje varijable nisu normalno distribuirane ili barem simetrično distribuirane).

ZADACI ZA VJEŽBU

1. Nađite jednadžbu pravca regresije za ove rezultate:

X	1	2	3	4	5
Y	5	4	3	2	1

2. 15 studenata bilo je upitano koliko su se sati pripremali za jedan kolokvij iz statistike. Njihovi odgovori na to pitanje uspoređeni su s bodovima što su ih dobili na kolokviju (maks. broj bodova = 100). Dobiveni su ovi rezultati:

(X)	Sati učenja	0,50	0,75	1,00	1,25	1,50	1,75	2,00
(Y)	Bodovi	57	64	59	68	74	76	79.

(X)	2,25	2,50	2,75	3,00	3,25	3,50	3,75	4,00
(Y)	83	85	86	88	89	90	94	96.

a. Nađite jednadžbu pravca regresije za ove rezultate.

b. Ako se neki student pripremao 0,25 sati, koji je njegov najvjerojatniji rezultat na kolokviju?

c. Izračunajte standardnu pogrešku prognoze.

3. Pretpostavimo da smo u jednom istraživanju varijabli X i Y dobili ove rezultate:

$\overline{X} = 600$	$\overline{Y} = 4,8$	$r = 0,58$
$s_x = 100$	$s_y = 0,4$	$N = 200.$

a. Koliko iznosi prognozirani $\tilde{Y}$ ako je $X = 70$?

b. Koliko iznosi prognozirani $\tilde{Y}$ ako je $X = 350$?

c. Kolika je pogreška prognoze u oba slučaja?

15. HI-KVADRAT TEST

Razlike između aritmetičkih sredina, neki računi korelacije itd., mogu se primijeniti samo na *kvantitativne* brojčane podatke, koji su ili *normalno* raspoređeni ili bar *simetrično* raspoređeni. Međutim, ako su podaci *kvalitativni* ili ako im distribucija značajno odstupa od normalne; onda se velik broj do sada opisanih postupaka (osim računa proporcija, nekih koeficijenata korelacije ne mogu upotrijebiti, nego se većinom upotrebljava postupak nazvan χ^2-test (čitaj: hi-kvadrat). Već u početku treba naglasiti da se hi-kvadrat test računa ***samo s frekvencijama***, pa, prema tome, nije dopušteno u račun unositi nikakve mjerne *jedinice*! ***Osnovni podaci*** istraživanja dakako mogu biti i mjerene vrijednosti, ali u hi-kvadrat unose se samo njihove frekvencije.

Hi-kvadrat test je vrlo praktičan test, koji može osobito poslužiti onda kad želimo utvrditi da li neke dobivene (opažene) frekvencije odstupaju od frekvencija koje bismo očekivali pod određenom hipotezom. On je često utoliko sličan računu korelacije, što i kod hi-kvadrat testa katkada tražimo postoji li *povezanost* između dvije varijable, ali i u tim slučajevima postoji bitna razlika između računa korelacije i hi-kvadrat testa, jer nam račun korelacije pokazuje *stupanj* povezanosti između dvije varijable, dok nam hi-kvadrat test pokazuje *vjerojatnost* povezanosti. O tome će još biti riječi na kraju ovog poglavlja.

Gotovo se u svim slučajevima hi-kvadrat izračunava na jednak način (uz ograničenje da katkada treba unijeti neke dodatne korekcije, ili je pak praktičnije upotrijebiti neku drugu formulu koja skraćuje računanje), i to prema formuli:

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}. \tag{15.1}$$

pri čemu f_o znači opažene frekvencije, a f_t očekivane (teoretske) frekvencije, tj. frekvencije koje bismo očekivali pod nekom određenom hipotezom.

Najčešće upotrebljavamo hi-kvadrat test u ovim slučajevima:

1. Kad imamo frekvencije *jednog* uzorka pa želimo ustanoviti odstupaju li te frekvencije od frekvencija koje očekujemo uz neku hipotezu.

2. Kad imamo frekvencije *dvaju* ili više *nezavisnih* uzoraka te želimo ustanoviti razlikuju li se uzorci u opaženim svojstvima.
3. Kad imamo frekvenciju *dvaju zavisnih* uzoraka, koji imaju *dihotomna* svojstva, te želimo ustanoviti razlikuju li se uzorci u mjerenim svojstvima, tj. je li došlo do *promjene*.

15.1. JEDAN UZORAK

Prvi primjer. 48 liječnika iznijelo je mišljenje o tome treba li ženi u porodu dati analgeziju. Dobiveni su ovi odgovori: 26 odgovora "da", 12 odgovora "ne znam" i 10 odgovora "ne". Da li ti odgovori pokazuju neko značajno odstupanje od onoga što bismo očekivali kad bi odgovori bili dani "nasumce", tj. posve slučajno?

Postavit ćemo "nul-hipotezu": nema razlike između *dobivenih* odgovora i *slučajno* raspoređenih odgovora. Kad bi odgovori bili dani potpuno slučajno, svaki bi od njih imao jednaku vjerojatnost, pa bismo prema tome svaki odgovor očekivali $48/3 = 16$ puta. Dakle, očekivana frekvencija za svaki odgovor bila bi 16.

Najprije ćemo rezultate tabelirati:

	"Da"	"Ne znam"	"Ne"	Ukupno
f_o	26	12	10	48
f_t	16	16	16	48

Kad smo dobili očekivane frekvencije, možemo izračunavati podatke potrebne za formulu (15.1):

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
26	16	10	100	$100/16 = 6,25$
12	16	-4	16	$16/16 = 1,00$
10	16	-6	36	$36/16 = 2,25$
				$\Sigma = 9,50$

$$\sum \frac{(f_o - f_t)^2}{f_t} = \chi^2 = 9,50.$$

Prije nego što interpretiramo dobiveni $\chi^2 = 9,50$, rastumačit ćemo princip njegove interpretacije: kad ne bi našli *nikakve* razlike između opaženih i očekivanih frekvencija, izraz χ^2 bi bio 0. Što su razlike između opaženih i očekivanih frekvencija veće, to je veći i definitivni izraz χ^2. Prema tome, što je hi-kvadrat manji (bliži nuli) (do neke određene granice, vidi o tome završetak poglavlja o hi-kvadrat testu), to je vjerojatnije da treba *prihvatiti* postavljenu hipotezu, a što je hi-kvadrat veći, to je vjerojatnije da postavljenu hipotezu treba *odbaciti*, jer se opaženi rezultati znatno

razlikuju od onih koje bismo pod određenom hipotezom očekivali. *Tablica* graničnih vrijednosti χ^2 (tablica H u Dodatku) pokazuje nam *do koje* vrijednosti (uz određeni broj stupnjeva slobode) moramo smatrati da je hi-kvadrat još uvijek dovoljno visok, a da bismo mogli odbaciti hipotezu, odnosno, drugim riječima, koliko mora *najmanje* iznositi vrijednost hi-kvadrat pa da odbacimo hipotezu. Naravno da i ovdje (kao i kod svih dosadašnjih testiranja značajnosti) možemo postaviti blaže ili strože zahtjeve, tj. možemo tražiti značajnost na razini od 5%, od 1%, itd.

Kao praktično pravilo može poslužiti činjenica da centralna vrijednost hi-kvadrata uz neki stupanj slobode iznosi po prilici toliko koliko imamo stupnjeva slobode. Prema tome, nul-hipotezu sigurno možemo prihvatiti (bez uvida u tablicu hi-kvadrata) ako je dobiveni hi-kvadrat manji ili jednak broju stupnjeva slobode.

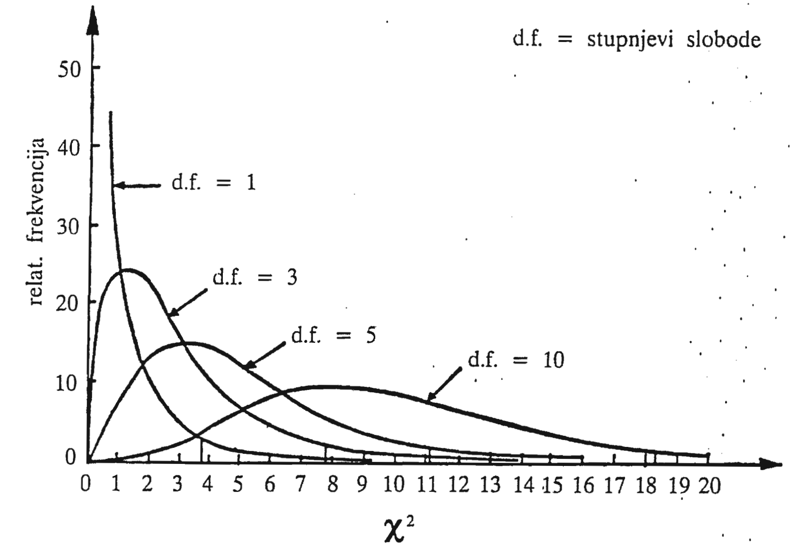
Postanak tablice H možemo relativno jednostavno protumačiti: Zamislimo da smo 100 ispravnih komada novca bacili u zrak (ili 1 novčić 100 puta), i da smo dobili 46 "glava" i 54 "pisma". Kao što znamo, očekivane su frekvencije: 50 "glava" i 50 "pisama". Izračunamo li hi-kvadrat, dobit ćemo:

	f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
Glava	46	50	-4	16	0,32
Pismo	54	50	4	16	0,32
					$\chi^2 = 0,64.$

Nastavimo li bacanjem tih 100 komada novaca i dalje, dobit ćemo i dalje često određene razlike između broja "glava" i "pisma", a ako su novci potpuno ispravni (tj. nemaju pojedini komadi novca tendenciju da pretežno padaju na jednu stranu), *sigurno je da su sva takva odstupanja potpuno slučajna*. Budući da je dovoljno znati koliko je palo "glava" pa da time odmah znamo koliko je palo "pisama" (jer su obje ćelije zavisne jedna od druge), to je broj stupnjeva slobode = 1. Na slici 15.1. prikazana je distribucija hi-kvadrata uz različite stupnjeve slobode. Među njima je i distribucija hi-kvadrata uz 1 stupanj slobode, tj. distribucija rezultata koje bismo dobili kad bismo zaista bacali 100 komada novca mnogo puta. Prema tome, sve su te vrijednosti hi-kvadrata *slučajne*.

(N a p o m e n a: Da smo umjesto 100 komada novčića bacali recimo 20 komada, pa registrirali ishode "pismo" i "glava", i uspoređivali ih s očekivanim ishodima, te izračunavali velik broj hi-kvadrata, dobili bismo jednaku distribuciju hi-kvadrata).

Međutim, one vrijednosti hi-kvadrata koje toliko jako odstupaju od očekivanog da je njihovo *slučajno* pojavljivanje moguće samo u 1% ili u 5% slučajeva, možemo već smatrati tolikim odstupanjem da s pravom možemo pretpostaviti da vjerojatno *nisu* slučajne. Na slici 15.1. uz krivulju distribucije hi-kvadrata uz 1 stupanj slobode označena je na apscisi 5%-tna granica, iza koje površina krivulje nadesno iznosi 5%. Kako se vidi iz slike (i čitamo iz tablice H), ta je vrijednost 3,84.



Slika 15.1. Distribucija uzoraka hi-kvadrata uz različite stupnjeve slobode

Ako umjesto 100 komada novca bacimo 100 igraćih kocaka, također možemo promatrati koliko odstupanje imamo kod svakog broja od 1 do 6, prema očekivanim frekvencijama (1/6 kocaka morala bi pasti na broj 1, 1/6 na broj 2, itd.). U tom slučaju imamo $6 - 1 = 5$ stupnjeva slobode. Velikim brojem bacanja dobili bismo distribuciju hi-kvadrata, prikazanu na slici 15.1, uz 5 stupnjeva slobode. Granična vrijednost hi-kvadrata (na razini značajnosti od 5%) ovdje iznosi 11,07. Na slici su još prikazane i distribucije uzoraka hi-kvadrata za 3 i 10 stupnjeva slobode.

Kad imamo samo jednu varijablu s jednim nizom rezultata, broj stupnjeva slobode računa se prema formuli $N - 1$, pri čemu N znači ukupan broj *ćelija* (a ne ukupan broj frekvencija). Kako u našem primjeru imamo samo 3 ćelije ("da", "ne znam", "ne"), broj stupnjeva slobode $= 3 - 1 = 2$. Želimo li testirati značajnost na razini od 5%, očitat ćemo u tablici graničnu vrijednost χ^2 uz 2 stupnja slobode, a na razini značajnosti $P = 0,05$ (= 5%). Kako se iz tablice vidi, granična vrijednost χ^2 uz 2 stupnja slobode na razini od %5 = 5,991. Kako je naš hi-kvadrat veći od 5,991, zaključujemo da treba *odbaciti* postavljenu hipotezu, tj. dobiveni se odgovori statistički značajno razlikuju od odgovora koje bismo očekivali kad bi oni bili dani posve slučajno.

Razumljivo je da postavljena hipoteza ne mora uvijek biti takva kao u prošlom primjeru. U tome i jest prednost hi-kvadrat testa da možemo postaviti hipotezu kakvu želimo. Na primjer, možemo postaviti hipotezu da bismo u nekom slučaju morali očekivati "normalnu raspodjelu", što ćemo pokazati u idućem primjeru.

Drugi primjer. S pomoću jednog testa psihomotorike testiramo 200 ljudi. Test je takve prirode da daje samo tri kategorije rezultata: A = slab, B = prosječan, C = dobar.

Kao rezultat mjerenja dobijemo ove frekvencije:

	A	B	C
f_o	40	110	50.

Odstupa li taj rezultat značajno od rezultata koji bismo očekivali da je svojstvo *normalno* raspoređeno među ispitanicima?

Budući da imamo 3 kategorije, najopravdanije je pretpostaviti da bi — po toj hipotezi — trebalo biti 50% prosječnih, a po 25% loših i dobrih:

	A	B	C
f_t	50	100	50.

Prema tome, račun će izgledati ovako:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
40	50	−10	100	2
110	100	10	100	1
50	50	0	0	0
				$\chi^2 = 3,00.$

Taj je broj *manji* od 5,991, pa ćemo, prema tome, *prihvatiti* hipotezu i zaključiti da se dobiveni rezultati ne razlikuju statistički značajno od onih koje bismo očekivali pod pretpostavkom da je mjereno svojstvo normalno distribuirano u skupini.

Treći primjer. Međutim, postoji mogućnost da mi neku očekivanu frekvenciju *već unaprijed znamo* jer je ona *poznata* u populaciji. Tako, na primjer, možemo ispitati da li se uzorak u kojem imamo 50 ljudi, i to 40 s tamnom kosom (80%) i 10 sa svjetlom kosom (20%), značajno razlikuje od omjera koji je *poznat* u nekoj populaciji, tj. da 75% ljudi imaju tamnu, a 25% svjetlu kosu.

Prema tome, možemo postaviti ovu tablicu:

	Tamna kosa	Svjetla kosa	Ukupno
f_o	40	10	50
f_t	37,5	12,5	50

i izračunati χ^2:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
40	37,5	2,5	6,25	0,17
10	12,5	−2,5	6,25	0,50
				$\chi^2 = 0,67.$

Broj stupnjeva slobode = 2−1 = 1. Dobiveni χ^2 znatno je manji od granične vrijednosti 3,84, pa stoga zaključujemo da naš uzorak ne odstupa statistički značajno od stvarne proporcije tamne i svijetle kose u populaciji.

N a p o m e n a. Kada imamo više od 2 ćelije, ako je više od 20% očekivanih frekvencija manje od 5, treba spajati susjedne ćelije zajedno. Kad radimo samo s 2 ćelije, većina statističara smatra (ali neki nisu tako strogi) da ne smije ni jedna očekivana frekvencija biti manja od 5.

Četvrti primjer. Uzmimo da smo analizirali nesreće kod 398 ljudi u jednom poslu i našli da su one među tim ljudima raspoređene kao što je prikazano u tablici 15.1.

TABLICA 15.1.
FREKVENCIJA LJUDI S RAZLIČITIM BROJEM NESREĆA

Broj nesreća	Broj ljudi
0	14
1	37
2	76
3	70
4	64
5	53
6	31
7	19
8	14
9	9
10	5
11	5
12	1
	$N = 398$

Zanima nas da li su nesreće među tim ljudima raspoređene prema "slučaju" tj. prema zakonu "rijetkih događaja" (to je tzv. Poissonova raspodjela).

Poissonova se raspodjela može izračunati ovako:

1. Ukupan broj nesreća podijelimo brojem ljudi te tako dobijemo "prosječan" broj nesreća;
2. izračunamo logaritam iz broja ljudi;
3. prosječan broj nesreća (1) pomnožimo izrazom 0,4343;
4. izvršimo operaciju (2) − (3);
5. izračunamo antilogaritam izraza pod (4). Tako dobivamo frekvenciju ljudi s 0 nesreća. Ovaj i daljnje račune treba računati na nekoliko decimala, a kad smo sve izračunali, možemo u tablicu očekivanih frekvencija unositi rezultate s manje (npr. 1 − 2) decimala;
6. izvedemo operaciju (5) · (1), i tako dobivamo frekvenciju ljudi s 1 nesrećom;
7. $\frac{(6)\cdot(1)}{2}$ = broj ljudi s 2 nesreće;

8. $\frac{(7)\cdot(1)}{3}$ = broj ljudi s 3 nesreće;
9. $\frac{(8)\cdot(1)}{4}$ = broj ljudi s 4 nesreće;

itd.

Jednostavnije i brže možemo izračunati očekivanu Poissonovu distribuciju uz pomoć Poissonovih tablica (vidi tablicu J u Dodatku). Ta tablica daje očekivanu *proporciju* u razredu 0 (nula) kod Poissonovih raspodjela s različitom aritmetičkom sredinom. U našem primjeru aritmetička sredina iznosi 1 549/398 = 3,89. U tablici možemo očitati da proporcija u razredu 0 (tj. očekivana proporcija ljudi bez nesreća) iznosi 0,0204. Pomnožimo li tu proporciju s brojem ljudi, dobivamo $0,0204 \cdot 398 = 8,12$. Od tog momenta dalje radimo prema već opisanom postupku, tj. očekivan broj ljudi s 1 nesrećom dobivamo tako da očekivan broj s 0 nesreća pomnožimo s aritmetičkom sredinom, itd.

Ako ovako izračunamo Poissonovu raspodjelu, dobit ćemo očekivane frekvencije, prikazane u tablici 15.2:

TABLICA 15.2.
OČEKIVANA FREKVENCIJA LJUDI S RAZLIČITIM BROJEM NESREĆA

Broj nesreća	Broj ljudi
0	8,12
1	31,59
2	61,44
3	79,67
4	77,48
5	60,28
6	39,08
7	21,72
8	10,56
9	4,56
10	1,77
11	0,63
12	0,20
13	0,06
14	0,02
15	0,01
	$\Sigma = 397,19$

Kontrola rezultata sastoji se u tome da suma očekivanih (teoretskih) frekvencija (uz dopuštene manje razlike zbog zaokruživanja decimalnih brojeva) mora odgovarati sumi opaženih frekvencija.

Izračunavanje će nakon toga imati tok prikazan u tablici 15.3.

TABLICA 15.3.
IZRAČUNAVANJE HI-KVADRATA ZA PODATKE IZ TABLICA 15.1. i 15.2.

Broj nesreća	f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
0	14	8,12	5,88	34,57	4,26
1	37	31,59	5,41	29,27	0,93
2	76	61,44	14,56	211,99	3,45
3	70	79,67	−9,67	93,51	1,17
4	64	77,48	−13,48	181,71	2,35
5	53	60,28	−7,28	53,00	0,88
6	31	39,08	−8,08	65,29	1,67
7	19	21,72	−2,72	7,40	0,34
8	14	10,56	3,44	11,83	1,12
9 i više	20	7,25	12,75	162,56	22,42
					$\Sigma = \chi^2 = 38,59$

Vidljivo je da su u tablici *spojeni* rezultati od razreda 9. nadalje. To je učinjeno zato što kod te vrste hi-kvadrat računa statističari zahtijevaju da ni jedna *očekivana* frekvencija ne bude manja od 5.

U testiranju Poissonove raspodjele broj stupnjeva slobode računa se po principu: broj razreda −2. (Jedan "stupanj slobode" izgubljen je na zajednički N kod opažene i teoretske krivulje, a drugi na zajedničku aritmetičku sredinu. Kod Poissonove raspodjele aritmetička sredina jednaka je varijanci, pa stoga na zajedničku varijancu ne gubimo daljnji, treći stupanj slobode.) Dakle, u našem slučaju imamo $10 - 2 = 8$ stupnjeva slobode. Iz tablice hi-kvadrata možemo očitati da uz 8 stupnjeva slobode granična vrijednost hi-kvadrat iznosi (na razini značajnosti od 5%) 15,507. Kako je naš dobiveni hi-kvadrat veći, odbacujemo nul-hipotezu i zaključujemo da vrlo vjerojatno (tj. uz rizik od 5%) naša distribucija nesreće *nije* Poissonova distribucija. (Taj zaključak ima dakako vrlo dalekosežno značenje, jer on govori da u distribuciji nesreća nije *slučaj* onaj jedini faktor koji je odgovoran za to da različiti ljudi imaju različit broj nesreća! No da bismo taj zaključak smjeli izvesti, mora biti manje-više ispunjen uvjet da ljudi kojima smo registrirali nesreće, na svojim radnim mjestima budu uglavnom podjednako eksponirani.)

N a p o m e n a. Za one čitaoce, koji posjeduju bolje džepno elektronsko računalo evo metode da se bez tablica izračunaju sve očekivane proporcije Poissonove raspodjele: treba, naime, primijeniti originalnu matematičku formulu za Poissonovu raspodjelu, koja glasi:

$$P_x = \frac{\lambda^x e^{-\lambda}}{X!},$$

pri čemu P_x znači vjerojatnost da će se pojaviti x, tj. neka *određena* frekvencija nesreća (npr. 4 nesreće), λ = aritmetička sredina, tj. prosječan broj nesreća, e = baza prirodnih logaritama = 2,7182818. Na primjer, vjerojatnost da će se u našem primjeru dogoditi 8 nesreća je:

$$P_8 = \frac{3,89^8 \cdot e^{-3,89}}{8!} = 0,027.$$

0,027 je *proporcija*. Ako želimo dobiti *frekvenciju*, treba taj broj pomnožiti s N, dakle s 398, i dobivamo 10,75. (U tablici 15.3. očekivana frekvencija u razredu 8. iznosi 10,56, no razlika je posljedica različitog načina računanja i efekta zaokruživanja na dvije decimale.)

Peti primjer. Mjereći visinu 135 20-godišnjih zagrebačkih mladića (vidi tablicu 6.2, str. 74), dobiveni su rezultati prikazani ponovno u tablici 15.4. u stupcima 1 i 2.

Želimo li testirati odstupa li dobivena distribucija značajno od *normalne* distribucije, treba izvesti ove operacije:

1. Izračunati aritmetičku sredinu i standardnu devijaciju rezultata.
2. Izračunati koliko su prava donja i gornja granica svakog razreda udaljene od aritmetičke sredine, i to izraziti u z- vrijednostima (stupci 4 i 5). Pritom ćemo ići 1-2 razreda i više i niže od razreda u kojima se nalaze opažene frekvencije. Budući da je prava donja granica nekog razreda ujedno i gornja granica nižeg razreda, to je dovoljno izračunati udaljenost samo do jedne od njih; u tablici 15.4. u stupcu 4 prikazana je udaljenost do donje granice svakog razreda.
3. Iz tablice normalne raspodjele (vidi tablicu A u Dodatku) izračunati površinu između z-vrijednosti, koje predstavljaju donju i gornju granicu svakog razreda (stupac 6 u tablici 15.4).

TABLICA 15.4.
TESTIRANJE NORMALNOSTI RASPODJELE

1	2		3	4	5	6	7	
Razred	Opažena frekvencija		Donja granica razreda	Odstupanje od aritm. sredine	Odstupanje u z-vrijednosti	Proporcija rezultata od donje do gornje gran. razreda	Očekivana frekvencija	
			ispod 153,5			0,0001		
154-156			153,5	-19,97	-3,72	0,0002	0,2	
157-159	1		156,5	-16,97	-3,16	0,0039	0,4	8,9
160-162	2	12	159,5	-13,97	-2,60	0,0160	2,0	
163-165	9		162,5	-10,97	-2,04	0,0487	6,3	
166-168	15		165,5	-7,97	-1,48	0,1068	14,7	
169-171	25		168,5	-4,97	-0,93	0,1795	24,3	
172-174	28		171,5	-1,97	-0,37	0,2197	29,6	
175-177	20		174,5	1,03	0,19	0,1980	26,7	
178-180	16		177,5	4,03	0,75	0,1315	17,6	
181-183	13		180,5	7,03	1,31	0,0644	9,0	
184-186	5	19	183,5	10,03	1,87	0,0232	3,0	13,1
187-189	1		186,5	13,03	2,43	0,0061	0,9	
190-192			189,5	16,03	2,98	0,0012	0,2	
193-195			192,5	19,03	3,54	0,0002		

$\Sigma = 135$ $\Sigma = 134,9$

$\overline{X} = 173,47$

$s = 5,37$

4. Budući da površina ispod normalne krivulje predstavlja *frekvenciju*, to ćemo očekivane frekvencije (stupac 7) dobiti tako da proporciju površine (stupac 6) pomnožimo s N.

5. Budući da pri krajevima raspodjele očekivanih frekvencija imamo male brojeve, spojit ćemo krajnje razrede tako da ukupna frekvencija iznosi najmanje 5. Iste ćemo razrede spojiti i u opaženim frekvencijama.

6. Izračunat ćemo hi-kvadrat uz broj stupnjeva slobode koji se računa: broj ćelija −3.

Ako izvedemo hi-kvadrat račun, dobivamo:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
12	8,9	3,1	9,61	1,080
15	14,7	0,3	0,09	0,006
25	24,3	0,7	0,49	0,020
28	29,6	−1,6	2,56	0,086
20	26,7	−6,7	44,89	1,681
16	17,6	−1,6	2,56	0,145
19	13,1	5,9	34,81	2,657
Stupnjevi slobode = 7 − 3 = 4				$\chi^2 = 5{,}675$

Budući da je naš hi-kvadrat manji od granične vrijednosti hi- kvadrata uz 4-stupnja slobode $(5,675 < 9,488)$, *prihvaćamo* postavljenu nul-hipotezu da se dobivena distribucija visine ne razlikuje od normalne distribucije.

15.2. DVA ILI VIŠE NEZAVISNIH UZORAKA

Prvi primjer. U jednoj tvornici provedena je anketa među 23 radnika i 26 radnica te je ispitivan stav radnika prema liječniku u ambulanti. Iz dobivenih odgovora moglo se zaključiti je li stav prema liječniku u cjelini "pozitivan" ili "negativan". Budući da je liječnik u toj ambulanti bila žena, postavljeno je pitanje razlikuju li se muškarci od žena u stavu prema toj liječnici. Dobiveni su ovi rezultati:

Muškarci ($N = 23$)	Žene ($N = 26$)
Pozitivan stav 14	Pozitivan stav 9
Negativan stav 9	Negativan stav 17.

Najprije ćemo unijeti rezultate u tzv. 2·2 tablicu u kojoj će apscisa predstavljati jednu varijablu (stav), a ordinata drugu varijablu (spol):

Stav prema liječniku

		Negativan	Pozitivan	Ukupno
Spol	Muškarci	9_a	14_b	23_{a+b}
	Žene	17_c	9_d	26_{c+d}
	Ukupno	26_{a+c}	23_{b+d}	49

Pod pretpostavkom da *nema* značajne razlike između muškaraca i žena, proporcija negativnog (ili pozitivnog) stava morala bi biti jednaka kod muškaraca i kod žena. Budući da u čitavoj grupi imamo 26 ljudi s negativnim stavom, znači da je proporcija tih ljudi u uzorku 26/49, pa stoga frekvencija *muškaraca* s negativnim stavom treba biti: $23 \cdot 26/49$ (jer imamo ukupno 23 muškarca), a frekvencija žena s negativnim stavom treba da bude: $26 \cdot 26/49$. Kako se vidi, *očekivane* frekvencije u svakoj ćeliji dobivamo jednostavno tako da *pomnožimo sumu reda sa sumom stupca i rezultat podijelimo totalnom sumom frekvencija.* Na taj ćemo način dobiti očekivane frekvencije:

Stav prema liječniku

		Negativan	Pozitivan	Ukupno
Spol	Muškarci	$23 \cdot 26/49 = 12,2$	$23 \cdot 23/49 = 10,8$	23,0
	Žene	$26 \cdot 26/49 = 13,8$	$26 \cdot 23/49 = 12,2$	26,0
	Ukupno	26,0	23,0	49,0

Većina statističara preporučuje da uvijek kad radimo s 2·2 tablicama (a također i onda kad radimo s drugim tablicama, npr. 2·3, itd., a u bilo kojoj ćeliji imamo *očekivanu* frekvenciju manju od 5), upotrijebimo tzv. Yates-ovu korekciju, koja se sastoji u tome da se za 0,5 *smanji* svaka opažena frekvencija, koja je *veća* od očekivane, a za 0,5 *poveća* svaka opažena frekvencija, koja je manja od očekivane. Drugim riječima, svaka se *razlika* između očekivane i opažene frekvencije smanji za 0,5.

Primijenimo li, dakle, tu korekciju na naš primjer (jer radimo s 2·2 tablicom), računat ćemo ovako:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
9,5	12,2	−2,7	7,29	0,598
13,5	10,8	2,7	7,29	0,675
16,5	13,8	2,7	7,29	0,528
9,5	12,2	−2,7	7,29	0,598
				$\chi^2 = 2,399$

U tablicama koje imaju redove i stupce, broj stupnjeva slobode izračunava se: (broj redova − 1) · (broj stupaca − 1). Budući da mi imamo 2 · 2 tablicu (jer imamo 2 reda i 2 stupca), broj stupnjeva slobode $=(2-1)\cdot(2-1)=1$. Iz tablice χ^2 možemo očitati da je granična vrijednost χ^2 uz 1 stupanj slobode na razini značajnosti od 5%, $\chi^2 = 3,841$. Budući da je naš hi-kvadrat manji, prihvatit ćemo hipotezu, tj. zaključit ćemo da se muškarci ne razlikuju statistički značajno od žena u stavu prema konkretnom liječniku.

Ovo je gotovo "školski primjer" kako statistički postupci *"kažnjavaju"* mali broj mjerenja. To je i potpuno opravdano, jer na 23 muška i 26 ženskih ispitanika zaista bi se i potpuno *slučajno* moglo dogoditi to, što se dogodilo tj. da žene u relativno većem postotku imaju negativan stav prema liječnici, nego muškarci. No budući da za takav rezultat postoje i izvjesna moguća psihološka ili sociološka opravdanja (tj. možda je liječnica bila atraktivna osoba, pa se više sviđala muškarcima), bilo bi zanimljivo znati radi li se možda zaista o jednom takvom fenomenu. *Pod pretpostavkom* da bi *odnosi* između pozitivnog i negativnog stava kod velikog broja ispitanika ostali *jednaki* (tj. da preko 65% žena ima negativan stav, a samo 39% muškaraca također negativan stav prema toj liječnici), uz 10 puta veće uzroke imali bismo 230 muškaraca i 260 žena. Muškaraca bi bilo 90, a žena 170 s negativnim stavom. Kada bismo sada računali hi-kvadrat, dobili bismo da je on ***deset puta veći***, tj. da iznosi gotovo 40, što je dakako (jer i sada imamo 1 stupanj slobode) statistički potpuno značajno. Iz toga bi u praksi bilo potpuno neopravdano, pa čak i *nedozvoljeno* izvesti zaključak: "dakle, kada bi uzorak bio 10 puta veći, razlika bi bila statistički značajna". Za svakoga, tko je do sada naučio "statistički misliti", bit će jasno, da bi taj zaključak bio točan samo *pod pretpostavkom da odnosi ostanu jednaki.* A to ***nikad*** ne možemo znati, jer možda bi se kod velikih uzoraka *promijenio* postotak zadovoljnih ili nezadovoljnih ispitanika različitih spolova.

Postoji međutim jedan jednostavniji postupak za izračunavanje hi- kvadrata kod 2 · 2 tablice, a jednostavniji je u tome što pomoću tog postupka nije uopće potrebno izračunavati razlike između opaženih i očekivanih frekvencija. Ako, naime, ćelije označimo slovima a, b, c, d, onda se χ^2 (uključujući i Yatesovu korekciju) može izračunati prema formuli:

$$\chi^2 = \frac{N\left(|ad-bc| - \frac{N}{2}\right)^2}{(a+c)(b+d)(c+d)(a+b)}. \tag{15.2}$$

N a p o m e n a. Znak | | oko izraza $|ad - bc|$ znači da treba uvijek uzeti *pozitivnu* razliku između ad i bc, tj. uvijek treba oduzeti manji izraz od većega.

U našem primjeru dobivamo ove rezultate (vidi prvu tablicu):

$$\chi^2 = \frac{49(|238-81| - 24,5)^2}{26\cdot 23\cdot 26\cdot 23} = \frac{49\cdot 132,5^2}{357\,604} = \frac{860\,256,25}{357\,604} = 2,406.$$

Kako se vidi, rezultat je praktički jednak rezultatu koji smo dobili prije.

(Malu razliku treba pripisati tome što smo izraze $\dfrac{(f_o - f_t)^2}{f_t}$ sveli na samo 3 decimale.)

Drugi primjer. Medicinski centar u Osijeku izvršio je 1967. godine analizu oboljenja od epidemije influence A-2 u poduzećima, od kojih su kolektivi nekih bili necijepljeni, kolektivi nekih cijepljeni 11 mjeseci prije epidemije, a kolektivi nekih neposredno prije epidemije. Dobiveni su ovi rezultati:

	Oboljeli	Nisu oboljeli	
Necijepljeni	402	2 497	2 899
Cijepljeni 11 mjeseci prije epidemije	378	3 789	4 167
Cijepljeni neposredno prije epidemije	131	2 009	2 140
	911	8 295	9 206

Izračunamo li već spomenutim postupkom očekivane frekvencije (suma stupca puta suma reda, podijeljeno ukupnom sumom), možemo postaviti donju tablicu izračunavanja hi-kvadrata:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\dfrac{(f_o - f_t)^2}{f_t}$
402	286,9	115,1	13 248,01	46,176
2 497	2 612,1	−115,1	13 248,01	5,072
378	412,3	−34,3	1 176,49	2,853
3 789	3 754,7	34,3	1 176,49	0,313
131	211,8	−80,8	6 528,64	30,825
2 009	1 928,2	80,8	6 528,64	3,386
				$\chi^2 = 88,625$

Broj stupnjeva slobode je 1 · 2 = 2. Dobiveni hi-kvadrat je znatno veći od 5,991, pa zaključujemo da *postoji* statistički značajna razlika u frekvenciji oboljenja između te tri grupe.

No, kao što se vidi iz formulacije gornjeg zaključka, ***takva informacija još nije dovoljna***, jer treba znati *u čemu se sastoji razlika:* jesu li cijepljeni obolijevali *manje* ili *više* od necijepljenih? Budući da su veličine skupina u svakoj kategoriji dosta različite, to je *interpretaciju* rezultata najlakše provesti ako vrijednosti u tablici pretvorimo u postotke. Pretvaranje u postotke treba obaviti u onom "smjeru" koji ispitujemo: budući da nas zanima da li više obolijevaju necijepljeni od cijepljenih, pretvorit ćemo naše frekvencije u tablici u postotke tako da će nam ukupne kategorije "necijepljenih", "cijepljenih prije 11 mjeseci" i "cijepljenih neposredno prije" iznositi 100%. Prema tome, nova tablica izgledat će ovako:

	Oboljeli	Nisu oboljeli	
Necijepljeni	13,9	86,1	100
Cijepljeni 11 mjeseci prije epidemije	9,1	90,9	100
Cijepljeni neposredno prije epidemije	6,1	93,9	100

Sada se iz tablice lijepo vidi da je najmanji postotak oboljelih među cijepljenim neposredno prije epidemije (6,1%), a najveći među necijepljenima (13,9%), pa prema tome zaključak iz prethodnog računa treba glasiti otprilike ovako: Postoji statistički značajna razlika u frekvenciji oboljenja između cijepljenih i necijepljenih, s tim da među cijepljenima ima *najmanje* oboljelih.

N a p o m e n a . Ova nam tablica, dakle, govori da postoje statistički značajne razlike u frekvencijama među pojedinim grupama. Međutim, hi-kvadrat ne govori ništa o tome među *kojim* grupama je razlika signifikantna. U našem slučaju značajnost razlike mogla bi se odnositi samo na grupe "necijepljeni" i "cijepljeni neposredno prije". Ako nas izričito zanima postoji li statistički značajna razlika između grupe "cijepljeni prije 11 mjeseci" i grupe "cijepljeni neposredno prije", morali bismo izračunati poseban hi-kvadrat samo za te dvije skupine. (Da smo to izračunali dobili bismo hi-kvadrat veći od 16, što znači da bi i ta razlika bila statistički značajna, tj. da najmanje oboljevaju oni, koji su cijepljeni neposredno prije epidemije).

U vezi s pretvaranjem rezultata tablice u postotke treba posebno naglasiti da se to radi samo radi lakše interpretacije rezultata, a iz te postotne tablice nikako ne smijemo *računati* hi-kvadrat, nego se on računa jedino iz tablice s originalnim frekvencijama.

Treći primjer. Uzmimo primjer koji smo spomenuli kod koeficijenta kontingencije C (str. 228), tj. postoji li zavisnost između boje očiju sinova i očeva. Ako rezultate unesemo u tzv. tablicu kontingencije, i ujedno u svaku ćeliju prema već spomenutom principu (suma reda puta suma stupca podijeljena ukupnom sumom) unesemo očekivane frekvencije (pod pretpostavkom da nema asocijacije između boje očiju sinova i otaca), dobivamo ove rezultate (očekivane frekvencije navedene su u zagradama):

		Boja očiju očeva				
		Modra	Siva	Maslinasta	Smeđa	Ukupno
Boja očiju sinova	Modra	194 (119,93)	70 (88,44)	41 (60,30)	30 (66,33)	335
	Siva	83 (101,67)	124 (74,98)	41 (51,12)	36 (56,23)	284
	Maslinasta	25 (49,05)	34 (36,17)	55 (24,66)	23 (27,13)	137
	Smeđa	56 (87,35)	36 (64,42)	43 (43,92)	109 (48,31)	244
	Ukupno	358	264	180	198	1000

Izračunavanje:

$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
74,07	5 486,36	45,746
−18,44	340,03	3,845
−19,30	372,49	6,177
−36,33	1 319,87	19,899
−18,67	348,57	3,428
49,02	2 402,96	32,048
−10,12	102,41	2,003
−20,23	409,25	7,278
−24,05	578,40	11,792
−2,17	4,71	0,130
30,34	920,52	37,328
−4,13	17,06	0,629
−31,35	982,82	11,252
−28,42	807,70	12,538
−0,92	0,85	0,019
60,69	3 683,28	76,243
		$\chi^2 = 270,355$.

Broj stupnjeva slobode = $(4-1)\cdot(4-1) = 9$. Iz tablice se vidi da granična vrijednost χ^2 uz 9 stupnjeva slobode, a na razini značajnosti od 5%, iznosi 16,919. Naš je hi-kvadrat znatno veći čak i od granične vrijednosti χ^2 na razini značajnosti od 1%, pa zato odbacujemo hipotezu (tj. da nema asocijacije između boje očiju sinova i otaca) i postavljamo zaključak da su te dvije varijable posve sigurno povezane.

V a ž n a   n a p o m e n a. Hi kvadrat kod $2 \cdot 2$ tablica, kao i formula (15.2), smije se upotrijebiti uvijek ako je N veći od 40. Kad je N manji od 40, ali veći od 20, smijemo računati samo ako ni jedna očekivana frekvencija nije manja od 5. U tablicama kontingencije, kad je broj stupnjeva slobode veći od 1, hi-kvadrat test može se još računati ako manje od 20% ćelija imaju očekivanu frekvenciju manju od 5, a ako ni jedna ćelija nema očekivanu frekvenciju manju od 1. Ako taj uvjet nije postignut, moramo neke kategorije (ćelije) spajati zajedno da bismo tako povećali očekivanu frekvenciju. No već smo kazali da neki statističari smatraju da nije neophodno pridržavati se tih pravila.

Za slučajeve vrlo malog N postoji tzv. Fisherov "egzaktni test" (koji uključuje dosta opsežno računanje), no mi ćemo ovdje izložiti jednu sasvim jednostavnu metodu, i to samo za one slučajeve kada se radi o dvije *jednako velike* skupine.

Uzmimo da imamo dvije skupine od po 15 ispitanika; eksperimentalna skupina primila je jedno sredstvo protiv morske bolesti, a kontrolna skupina primila je "placebo", tj. nedjelotvorne pilule (to je potrebno učiniti zato da bi se izjednačilo eventualno djelovanje sugestije na rezultate). Svi su ispitanici podvrgnuti vestibularnim stresovima vrtnje, i nakon toga 2 ispitanika eksperimentalne skupine pokazala su znakove "morske bolesti", a iz kontrolne skupine 8 ispitanika ih je pokazalo jednake znakove. Rezultate eksperimenta mogli bismo, dakle, prikazati u tablici ovako:

	Imaju simptome	Nemaju simptome	
Eksperimentalna skupina	2	13	15
Kontrolna skupina	8	7	15

Može li se razlika između eksperimentalne i kontrolne skupine smatrati značajnom?

Tablica I u Dodatku daje odgovor na to pitanje. Da bi se tablica mogla koristiti, treba naći *najmanju* frekvenciju u rezultatima (to je u našem primjeru frekvencija 2), kao i frekvenciju koja u *drugoj* grupi njoj korespondira (u našem primjeru to je frekvencija 8). U glavi tablice I nalaze se brojevi koji označuju *najmanju* frekvenciju u rezultatima, a uz lijevi rub tablice nalazi se veličina jednog od uzoraka ($N_1 = N_2$). Rješavajući naš primjer, treba u glavi tablice naći broj 2 (naša najmanja frekvencija), a na lijevom rubu broj 15 (veličina uzorka): u sjecištu stupca 2 i reda 15 u tablici čitamo brojeve 9 i 10. To su *najmanje* frekvencije koje bi *morala* imati korespondentna ćelija, i to za razine značajnosti od 5% (9) i 1% (10). Budući da naša korespondentna ćelija ima frekvenciju 8, zaključujemo da razliku *ne možemo* smatrati statistički značajnom.

Ali da smo na primjer dobili rezultat:

	Imaju simptome	Nemaju simptome	
Eksperimentalna skupina	0	15	15
Kontrolna skupina	5	10	15

prema podacima iz tablice I ta bi razlika bila značajna na razini od 5% (ali ne i na razini od 1%).

15.3. DVA ZAVISNA UZORKA (McNemarov test)

Ako uspoređujemo rezultate jedne te *iste* grupe "prije" i "poslije", ili uspoređujemo istu grupu u dvije različite aktivnosti, onda vjerojatno postoji *korelacija* između prvih i drugih rezultata.

Primjer. Uzmimo isti primjer koji smo upotrijebili pri izračunavanju značajnosti razlike u proporcijama koje su u korelaciji (vidi str. 172): 100 ispitanika ispitani su testom 1 i testom 2. Dobili smo ove rezultate:

		Test 2	
		Nisu zadovoljili	Zadovoljili
Test 1	Zadovoljili	5_A	55_B
	Nisu zadovoljili	25_C	15_D

Postoji li značajna razlika između rezultata u 1. i 2. testu?

Kako se iz tablice vidi, *razlike* između 1. i 2. testa nalaze se u ćelijama A i D, dok su u ćelijama B i C navedeni samo oni koji su ili uspjeli ili nisu uspjeli u *oba* testa. Prema tome, A + D predstavlja totalni broj onih kod kojih se *ne slaže* uspjeh prvog i drugog mjerenja.

Budući da A + D predstavljaju ukupan broj ispitanika koji su promijenili svoj uspjeh, očekivali bismo pod nul-hipotezom da bi se 1/2 (A + D) slučajeva promijenilo u jednom, a 1/2 (A + D) u drugom smjeru. Drugim riječima, pod nul-hipotezom očekivane frekvencije u ćeliji A iznose 1/2 (A + D), a jednako toliko u ćeliji D. Zanimaju nas samo ćelije A i D (jer B i C pokazuju poklapanje), pa su, prema tome, opažene frekvencije one koje se nalaze u A i D, a očekivane su frekvencije: 1/2 (A + D).

Dakle,

$$\chi^2 = \Sigma\frac{(f_o - f_t)^2}{f_t} = \frac{\left(A - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} + \frac{\left(D - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}}. \tag{15.3}$$

Izvršimo li potrebne računske operacije u gornjoj formuli, dobivamo na kraju:

$$\chi^2 = \frac{(A-D)^2}{A+D}, \tag{15.4}$$

a uz Yatesovu korekciju (ako je $(A+D) < 20$) konačna fromula glasi:

$$\chi^2 = \frac{(|A-D|-1)^2}{A+D}. \tag{15.5}$$

U našem primjeru dobivamo:

$$\chi^2 = \frac{(10-1)^2}{20} = \frac{81}{20} = 4,05.$$

Broj stupnjeva slobode (2 · 2 tablica!) = 1. Granična vrijednost χ^2 za 1 stupanj slobode je 3,841, a kako je naš hi-kvadrat veći, odbacit ćemo nul-hipotezu (tj. da nema razlike u težini testova) i zaključiti da razlika postoji, tj. da je drugi test lakši.

N a p o m e n a. U ovom se računu zapravo radi o testiranju značajnosti razlike između dviju proporcija $p_1 = (A+B)/N$; $p_2 = (B+D)/N$, dakle jednako kao i u primjeru na strani 172, samo ga sada izračunavamo drugačije.

N a p o m e n a. Treba paziti na smisao ćelija A i D. To su ćelije koje predstavljaju one ispitanike koji su se *promijenili.* Ako je tablica formirana drukčije treba analogno tome preurediti i formule 15.3. do 15.5.

(N a p o m e n a. Ako su očekivane frekvencije u ćelijama A i D manje od 5, taj se račun ne može upotrijebiti.)

Valja uočiti da bi primjena *standardnog* postupka za izračunavanje očekivanih frekvencija , koji inače koristimo kod kontingencijskih tablica hi-kvadrata (suma reda puta suma stupca, podijeljeno s ukupnom sumom) *dala potpuno nelogične i neupotrebljive rezultate.* Evo primjera, koji će to dokazati: pretpostavimo da u našem prijašnjem primjeru nije došlo *ni do kakvih promjena* između prvog i drugog mjerenja, i da je rezultat recimo bio ovakav:

		Test 2: Nisu zadovoljili		Test 2: Zadovoljili		
Test 1	Zadovoljili	0	A	70	B	70
	Nisu zadovoljili	30	C	0	D	30
		30		70		100

Kako vidimo, od ukupno 100 ispitanika isti ispitanici koji nisu zadovoljili u prvom testiranju (njih 30), nisu zadovoljili ni u drugom, a također istih 70 ispitanika oba je puta zadovoljilo. Izračunamo li standardnim postupkom očekivane frekvencije, dobili bi ove frekvencije: 21, 49, 9, 21. Izračunati hi-kvadrat (uz Yatesovu korekciju) iznosio bi 95,29, no taj bi rezultat bio potpuno besmislen, jer *ništa* se nije promijenilo. McNemarov test (ovaj puta bez Yatesove korekture, jer je u brojniku nula) dao bi naprotiv potpuno točan rezultat: hi-kvadrat = 0, tj. *nema promjene*.

Za neke (rijetke) situacije može McNemarov test ipak biti izričito nepogodan: ako neki postupak, primijenjen na grupu ispitanika, može kod njih proizvesti *suprotne* učinke (npr. neki ispitanici se od nekog sredstva uzbude, a neki umire, ili pak neki postupak kod jednih ispitanika dovodi do povećanja, a kod drugih do smanjenja agresivnosti), onda se dakako može dogoditi da ih bude podjednako ili sličan broj u ćelijama A i D, i McNemarov test će dati *malu* vrijednost (što bi trebalo značiti da nije došlo do promjene), a do *značajnih* promjena je došlo!

15.4. NEKI OSNOVNI UVJETI ZA UPOTREBU HI-KVADRAT TESTA

Kao što smo vidjeli, hi-kvadrat test je stvarno vrlo jednostavan test, jer je njegova logika jasna, a izračunavanje vrlo jednostavno. No upravo se u tome vjerojatno i krije opasnost da se njegova jednostavnost *precijeni*, pa se tako u stručnoj i naučnoj literaturi najviše pogrešaka u primjeni statističkih postupaka nalazi upravo kod primjene hi-kvadrat testa. Dok se mnogi drugi statistički postupci dadu često primijeniti dosta mehanički i bez posebnog opreza, kod hi-kvadrat testa uvijek je potrebno dobro promisliti kako ćemo rezultate prikazati u tablici.

Prije nego što iznesemo neke osnovne uvjete, koji moraju biti ispunjeni da bi se smio računati hi-kvadrat test, navest ćemo jednu praktičnu stranu hi-kvadrata. To je test koji posjeduje tzv. aditivna svojstva, a to znači da imamo pravo zbrojiti nekoliko hi-kvadrata iz istih istraživanja, i na značajnost dobivenog rezultata zaključivati iz tablice, s tim da zbrojimo i stupnjeve slobode. Tako je, na primjer, poznato da su svojedobno, u doba ispitivanja cjepiva protiv kolere, izvršena brojna istraživanja djelovanja cjepiva. Iz Indije je bilo poznato 5 izvještaja o 5 manjih ispitivanja, koja, ako se rezultati izraze hi-kvadrat testom, daju ovakvu situaciju:

	χ^2
Regimenta pokrajine ist. Lancashire	2,04
Britanske trupe u Cownporeu	1,83
Britanske trupe u Dinaporeu	1,60
Gya Jail	5,90
Durbhanga Jail	3,18.

Svi ti rezultati bili su vezani svaki za 1 stupanj slobode. Kako se vidi, samo jedan od njih bio je statistički značajan. No, ako sve te rezultate *zbrojimo*, dobivamo $\chi^2 = 14,55$, a iz tablice ustanovljujemo da je uz 5 stupnjeva slobode taj rezultat statistički značajan ($P < 0,05$).

Pri takvim situacijama zbrajanja rezultata hi-kvadrata treba paziti da se zbroje svi raspoloživi rezultati (a ne samo pozitivni!). Osim toga, potpuno je razumljivo da smijemo zbrajati samo one hi-kvadrate koji svi pokazuju devijaciju u "istom smjeru". Budući da je "smjer" devijacije kod hi-kvadrata vidljiv samo iz *inspekcije tablice* (a ne iz samog broja, jer je broj uvijek pozitivan!), pri tom poslu treba biti vrlo oprezan.

Evo na kraju sažetih glavnih uvjeta, koji moraju biti ispunjeni da bi se smio računati hi-kvadrat test:

1. *hi-kvadrat test može se računati samo s frekvencijama.* Prema tome, u ćelije hi-kvadrat testa ne smijemo unositi aritmetičke sredine, kao ni postotke, ni proporcije. Ako u ćelije unesemo postotke, sveli smo na taj način N svake grupe na 100, što, naravno, nije dopušteno.
2. *Suma očekivanih frekvencija mora biti jednaka sumi opaženih frekvencija.* Toleriraju se minimalne razlike u vezi sa zaokruživanjem decimalnih brojeva.
3. Kad god u hi-kvadrat testu radimo s nekim svojstvom koje se *pojavilo* ili se *nije* pojavilo, treba u računu staviti i frekvencije u kojima se to svojstvo nije pojavilo. Ako to ne učinimo, može nam se u nekim slučajevima dogoditi da suma opaženih frekvencija ne odgovara sumi očekivanih frekvencija.

I kad suma očekivanih frekvencija potpuno odgovara sumi opaženih frekvencija, treba se pridržavati pravila da u računu navedemo i frekvencije u kojima se svojstvo nije pojavilo.

Primjer. Zanima nas postoje li razlike u frekvenciji ozljeđivanja među radnicima različite starosti, i izvršimo registraciju nesreća u jednom poduzeću te dobijemo ove rezultate:

Starost radnika	20-29 god.	30-49 god.	50 i više god.
Broj radnika	200	500	300
Broj radnika sa dvije ili više nesreća	70	100	30

Ako nema razlike u frekvenciji ozljeđivanja među radnicima različite starosne dobi (nul-hipoteza), možemo uzeti sve radnike zajedno, pa tako dobivamo da je od ukupno 1 000 radnika njih 200 imalo nesreće. To iznosi 20%, pa bismo stoga morali očekivati jednak postotak u svim dobnim skupinama; to su ove očekivane frekvencije: 40, 100, 60. Kako se vidi, suma očekivanih frekvencija iznosi 200, jednako kao i suma opaženih frekvencija. Međutim, izračunamo li samo iz tih rezultata hi-kvadrat, dobit ćemo $\chi^2 = 37,5$. Naprotiv, unesemo li u tablicu hi-kvadrata i frekvencije radnika *bez* nesreća (tj. s manje od dvije nesreće), hi-kvadrat će iznositi 46,9. U graničnim slučajevima, tj. kada je hi-kvadrat upravo u blizini granične vrijednosti prema tablici, upotreba ispravnog postupka može imati odlučujuće značenje za krajnji rezultat.

Još bi, naravno, teža pogreška bila da se uopće ne osvrćemo na stvarni broj slučajeva u svakoj kategoriji, nego da očekivane frekvencije izračunamo samo na temelju prosjeka opaženih frekvencija. Ako su opažene frekvencije: 70, 100, 30 (ukupno 200), onda bi očekivane frekvencije trebale biti 200/3 = 66,7, 66,7, 66,7. I takvi su slučajevi mogu katkada u praksi naći, ali to već prestavlja potpuno nerazumijevanje hi-kvadrat postupka.

4. *Frekvencije u pojedinim ćelijama moraju biti u tom smislu nezavisne* da svaka frekvencija u pojedinoj ćeliji mora pripadati *drugom* individuumu. Na primjer, ne smijemo u tablicu unositi *nekoliko* odgovora jednog ispitanika; također se N ne smije povećati tako da se na svakom ispitaniku učini nekoliko pokusa pa se svaki pokus unese u tablicu.

5. *Nijedna očekivana frekvencija ne smije biti odveć mala.* U tom se treba pridržavati ovih pravila:

   a) Kad imamo više od dvije ćelije, ako je više od 20% očekivanih frekvencija manje od 5, treba spajati susjedne ćelije. Kad radimo samo s dvije ćelije, ne smije ni jedna očekivana frekvencija biti manja od 5.

   b) Kod $2 \cdot 2$ tablica hi-kvadrat smije se upotrijebiti uvijek ako je N veći od 40. Ako je N manji od 40, ali veći od 20, ne smije ni jedna očekivana frekvencija biti manja od 5.

   c) U tablicama kontingencije kad je broj stupnjeva slobode veći od 1, hi-kvadrat se smije računati ako manje od 20% ćelija ima očekivanu frekvenciju manju od 5, a ni jedna ćelija manju od 1. Ako to nije postignuto, treba spajati ćelije u kojima su očekivane frekvencije odveć malene. (Naravno da raditi takvo spajanje ima smisla samo onda ako se time ne upropasti svrha samog ispitivanja, tj. ako fenomen koji ispitujemo, ostaje i dalje vidljiv.)

V a ž n a   n a p o m e n a: U novije vrijeme pojavile su se međutim rasprave koje dokazuju da nije *naročito važno* pridržavati se pravila 5.

6. *Kada postoji samo 1 stupanj slobode, potrebno je provesti korekciju za kontinuitet* (Yatesova korekcija). Ako su razlike između opaženih i očekivanih frekvencija vrlo male, tako da primjenom Yatesove korekcije dobijemo razliku koja je numerički *veća* (bez obzira na predznak), *onda upotreba te korekcije nema opravdanja!* No i ovdje valja primijetiti da ta korekcija ima smisla samo kod *malih* frekvencija u ćelijama, jer će kod velikih frekvencija s korekcijom doći samo do malih razlika u završnom rezultatu.

15.5. JOŠ O HI-KVADRAT TESTU

Na kraju rasprave o hi-kvadrat testu dodajemo još tri napomene, od kojih smo prvu spomenuli već na početku ovog poglavlja, druga je manje poznata, ali je vrlo zanimljiva, a treća je posebno važna za one koji *nedovoljno razmišljaju* prilikom korištenja toga testa.

1. Prva se napomena odnosi na spomenuto svojstvo hi-kvadrat testa da uz njegovu pomoć možemo ustanoviti i *vjerojatnost povezanosti* između dvije varijable (ne dakle *visinu* povezanosti koju nam daje koeficijent korelacije).

Budući da u tom pogledu kadšto vlada kod početnika određena konfuzija, razjasnit ćemo na jednom jednostavnom primjeru o čemu se zapravo radi.

Uzmimo da nas zanima razlikuju li se muškarci od žena u svom stavu prema boksačkim borbama i da anketom dobijemo podatke da od 200 anketiranih žena samo njih 50 izjavljuje da odobrava boksačka natjecanja, a od 300 anketiranih muškaraca njih 200 izjasnilo se u prilog boksu. U donjoj tablici prikazani su dobiveni rezultati:

		Stav		
		za	protiv	
Spol	žene	50	150	200
	muškarci	200	100	300
		250	250	500

Izračunamo li na osnovi tih podataka hi-kvadrat, dobit ćemo da on iznosi 83,4, i prema tome visoko je statistički značajan, pa smo, dakle, dokazali da se muškarci od žena statistički značajno razlikuju u stavu prema tom pitanju, tj. da muškarci imaju znatno povoljniji stav prema boksu. No istodobno mi smo time dokazali i *postojanje povezanosti između varijable "stav prema boksačkim natjecanjima" i varijable "spol"*. Drugim riječima, nije svejedno anketiramo li o tom pitanju muškarce ili žene, a ako nije svejedno, onda znači da postoji *korelacija* između stava i spola! To, naravno, *ne mora značiti* da je korelacija visoka, već samo to da ona postoji i da je statistički značajna. A kolika je aproksimativno ćemo ustanoviti ako uz pomoć hi-kvadrata izračunam *Cramerov Fi* koeficijent (formula 13.28):

$$Cramerov\ Fi = \sqrt{\frac{\chi^2}{N(s-1)}} = \sqrt{\frac{83,4}{583,4}} = 0,38$$

(Kao što se može lako ustanoviti, jednak rezultat dobili bismo da smo računali i kritizirani koeficijent kontingencije C. Razlog tome je činjenica da kod $2 \cdot 2$ kontingencijskih tablica nema razlike između C i *Cramerova Fi.*).

(N a p o m e n a. Hi kvadrat ima za korelaciju otprilike ono isto značenje što ga ima i testiranje značajnosti korelacije: ako je hi-kvadrat značajan, i korelacija - bila ona niska ili visoka - statistički je značajna.)

2. Druga napomena o hi-kvadrat testu pripada među rijetko poznate, a radi se o *malim* vrijednostima hi-kvadrata. Evo u čemu se ta napomena sastoji:

Iz t-testa naučeni smo da neku razliku smatramo statistički značajnom ako je dobiveni t *veći* od granične t-vrijednosti u tablici. Jednako postupamo i kod hi-kvadrat testa, tj. smatramo da je razlika između opaženih i teoretskih frekvencija statistički značajna ako je dobiveni hi-kvadrat veći od granične vrijednosti u tablici hi-kvadrata, uz određeni broj stupnjeva slobode. U tumačenju logike hi-kvadrat testa na strani 250, zato smo i rekli: "Što je hi-kvadrat manji (bliži nuli), to je vjerojatnije da treba prihvatiti postavljenu hipotezu...". Međutim, malo pažljiviji uvid u distribuciju uzoraka hi-kvadrata (vidi sliku 15.1) pokazat će da to *ne mora biti doslovno tako.*

Uzmimo primjer s bacanjem igraće kocke. Ako je kocka ispravna, vjerojatnost pojavljivanja svakog ishoda je jednaka i iznosi $P = 1/6$. Pretpostavimo da u jednom izvještaju iz jedne igračnice čitamo da je izvršeno testiranje igraće kocke tako da je kocka bačena šesto puta, te je zabilježeno koliko je puta dobiven rezultat 1, koliko puta 2, 3, 4, itd., te je nakon toga izračunat hi-kvadrat test. Uzmimo da objavljeni rezultati izgledaju ovako:

Ishod	f_o	f_t
1	98	100
2	99	100
3	101	100
4	102	100
5	99	100
6	101	100
Ukupno	600	600

Izračunamo li na temelju tih podataka hi-kvadrat, dobivamo $\chi^2 = 0,12$. Budući da je granična vrijednost hi-kvadrata uz 5 stupnjeva slobode 11,070, u prvi mah bez ikakve sumnje *prihvaćamo* nul-hipotezu, tj. zaključujemo da je kocka potpuno ispravna, jer se rezultati tek minimalno razlikuju od očekivanih frekvencija.

No je li to zaista tako? Pogledajmo *distribuciju* uzoraka hi-kvadrata uz 5 stupnjeva slobode na slici 15.1. Tu možemo vidjeti da bi se — i kod najispravnije kocke — hi-kvadrati distribuirali tako da im se *dominantna vrijednost* kreće negdje oko 4, a vrijednosti koje od dominantne značajno odstupaju rijetke su na *obje* strane krivulje. Drugim riječima, iz krivulje jasno vidimo da u ovom slučaju i izričito *mali* hi-kvadrat ne možemo smatrati sigurno slučajnim, jer bi se on *slučajno* mogao pojaviti samo izvanredno rijetko. Vjerojatnost slučajnog pojavljivanja hi-kvadrata *većeg* od 15,086 je jednaka ($P = 0,01$) kao i vjerojatnost slučajnog pojavljivanja hi-kvadrata *manjeg* od 0,554 ($P = 0,99$, što znači da je vjerojatnost 99% da će slučajni hi-kvadrat biti *veći* od 0,554).

Dakle, u t-testu — što je t manji, to smo *sigurniji* da nema razlike između dvije populacije; a kod hi-kvadrat testa i *suviše mali* hi-kvadrat možemo smatrati da nije *slučajno* nastao!

Što u ovom slučaju treba zaključiti? Samo jedno: rezultati izvještaja vrlo su vjerojatno *izmišljeni* (jer su "predobri" da bi mogli biti istiniti!), a izmislio ih je netko tko zna računati hi-kvadrat i poznaje njegovu osnovnu logiku, ali ga ne razumije dokraja.

Neka usputno bude spomenuto i to da su naknadna provjeravanja eksperimenata osnivača genetike Gregora Mendela (prvi je te podatke provjeravao R. A. Fisher, jedan od najvećih statističara do sada) pokazala da se stvarno dobivene frekvencije nekih nasljednih karakteristika u njegovim pokusima toliko dobro slažu s teoretski očekivanim frekvencijama Mendelovih zakona da su svi hi-kvadrati (koje Mendel, naravno, u ono vrijeme nije znao računati) "previše visoki da bi izgledali istiniti"! Stručnjaci danas na različite načine tumače tu pojavu (jer Mendelovi zakoni su točni i nitko u njih ne sumnja), pa se između ostaloga spominje i mogućnost da su njegovi mladi suradnici, želeći mu ugoditi, ponešto "frizirali" rezultate eksperimenata kako bi se oni još bolje slagali s očekivanim frekvencijama. No, bilo kako bilo, Mendelovo otkriće — kako kažu Hodges, Krech i Crutchfield — koje mnogi smatraju jednim od najvećih trijumfa ljudskog uma, bilo je dovoljno snažno da odoli čak i kritici "odveć dobrih rezultata".

N a p o m e n a. Spomenutu mogućnost očitavanja hi-kvadrat distribucije i s *lijevog kraja* (tj. za one hi-kvadrat vrijednosti koje su "suviše male da bi izgledale istinite") ne treba miješati s "jednosmjernim" ili "dvosmjernim" testiranjem značajnosti razlike kod *t-testa!* "Dvosmjerno" testiranje — kao što smo rekli — znači testirati je li neka razlika — *bez obzira na smjer te razlike* — statistički značajna ili nije. Drugim riječima, ako nađemo da je npr. grupa djece A viša od grupe djece B, onda dvosmjernim testiranjem značajnosti razlike mi samo odgovaramo na pitanje je li moguće da se razlika, koju smo među uzorcima dobili, dogodila slučajno ili ne. Pri tome je potpuno svejedno je li ta razlika u korist grupe A ili grupe B, jer zanima nas samo *veličina* razlike, bez obzira na predznak, tj. na njezin smjer. (A ako nas opravdano zanima samo jedan smjer razlike, onda — eventualno — možemo koristiti samo jednu stranu normalne odnosno t-distribucije, i provesti "jednosmjerno" testiranje.)

Granične vrijednosti hi-kvadrat distribucije, koje se nalaze u hi-kvadrat tablici, makar se odnose na "desnu stranu" hi-kvadrat distribucije, jesu vrijednosti *dvosmjernog* testiranja, jer pomoću njih testiramo značajnost razlike *bez obzira na njezin smjer!* (Kao što znamo, prilikom računanja hi-kvadrata smjer razlike u računu nema nikakvu ulogu, jer se razlike između opaženih i očekivanih frekvencija kvadriraju!)

Prema tome, upozorenje da i suviše mali hi-kvadrat može biti sumnjiv (tj. da možda nije nastao potpuno slučajno), i da se to može provjeriti s *lijeve* strane krivulje distribucije hi-kvadrata — problem je sasvim druge vrste od problema "jednosmjernog" ili "dvosmjernog" testiranja značajnosti razlike.

3. Iskustvo pokazuje, da pojedini korisnici hi-kvadrat testa katkada nedovoljno razmišljaju o tome, *što ih zapravo zanima u njihovu istraživanju.* To ćemo najbolje objasniti jednim primjerom, koji se katkada događa.

Recimo da je neki istraživač sakupio podatke o broju samoubojstava u toku svih 12 mjeseci neke godine u jednom velikom gradu, i da ga zanima postoje li razlike između muškaraca i žena u smislu frekvencije samoubojstava, tj. da li muškarci ili žene čine veše samoubojstava. Pretpostavimo da je dobio ove rezultate:

	Siječ.	Velj.	Ož.	Trav.	Svib.	Lip.	Srp.	Kol.	Ruj.	List.	Stud.	Pros.	Ukupno
Mušk.	5	7	8	4	12	10	12	9	10	6	10	9	102
Žene	4	7	10	8	7	10	9	8	4	5	8	12	92

Ako on sada na *sve* te rezultate primijeni pravila izračunavanja očekivanih frekvencija (tj. učini i sume stupaca, pa množi sume reda sa sumom stupca i dijeli ukupnom sumom), on je zapravo računao razlikuju li se muškarci od žena po broju samoubojstava *u toku pojedinih mjeseci*, a nije dobio odgovor na svoje pitanje da li muškarci ili žene imaju više samoubojstava. Ako je to problem, koji on želi riješiti uz pomoć hi-kvadrat testa, onda ga uopće ne zanima stanje samoubojstava po mjesecima, već jedino ukupni broj samoubojstava kod žena i muškaraca u toku godine dana. Budući da u populaciji postoji uglavnom jednaki broj žena i muškaraca, on bi mogao postaviti jednostavnu tablicu:

	f_o	f_t
Muškarci	102	97
Žene	84	97

pa sada upotrijebiti hi-kvadrat.

Kao što vidimo, glavna opasnost od hi-kvadrat testa je u tome što se on *lagano izračunava*, ali treba prethodno dobro promisliti *što nas zapravo zanima*, pa tek tada ići na izračunavanje teoretskih (očekivanih) frekvencija, jer one ovise o hipotezi, koju smo postavili.

ZADACI ZA VJEŽBU

1. Jedan je nastavnik tvrdio da među njegovih 50 studenata tridesetorica moraju pasti, da su 15 prosječni, a 5 vrlo dobri. Odstupa li takva raspodjela statistički značajno od onoga što bismo mogli očekivati pod vidom normalne raspodjele, tj. da je 50% učenika u srednjoj kategoriji?

2. Četvorica nastavnika istog predmeta imali su ovaj rezultat ispita na kraju godine:

	A	B	C	D
Broj "palih" učenika	8	5	4	7
Broj "prošlih" učenika	48	40	35	43.

Razlikuje li se proporcija prošlih (ili palih) učenika kod ove četvorice nastavnika?

3. Na jednom tečaju statistike, na kojem je bilo 40 muškaraca i 30 žena, na završnom ispitu postignuti su ovi rezultati:

	nedovoljan	dovoljan i dobar	v. dobar i odličan
Muškarci	8	24	8
Žene	5	16	9

Je li razlika između muškaraca i žena statistički značajna?

4. Jedan je sociolog ispitivao postoje li razlike u vrsti kriminalnih čina između 3 grada i dobio je ove rezultate:

	Silovanje	Krađa automob.	Krađe i džepar.	Ostalo
Grad A	76	112	87	102
Grad B	64	184	77	98
Grad C	39	131	48	82

Postoji li statistički značajna razlika među gradovima?

5. U jednoj zemlji anketirano je nekoliko desetaka pripadnika različitih političkih stranka pitanjem: odobravaju li smrtnu kaznu. Rezultati su prikazani u donjoj tablici:

	Republikanci	Demokrati	Nezavisni
Da	19	26	13
Ne	7	27	8

a. Je li razlika među grupama statistički značajna?

b. Povećajte sve brojeve deset puta, izračunajte hi-kvadrat test te prokomentirajte rezultat.

6. U toku 8-godišnjeg razdoblja 17 952 američka pilota imala su distribuciju nesreća u službi kao što je prikazano dolje:

Broj nesreća	Broj pilota
0	12 475
1	4 117
2	1 016
3	269
4	53
5	14
6	6
7	2
	17 952

Je li ta distribucija nesreća slučajna (Poissonova) distribucija?

16.

STANDARDNA POGREŠKA STANDARDNE DEVIJACIJE I GRANICE POUZDANOSTI STANDARDNE DEVIJACIJE

Čitalac se možda zapitao zašto problem standardne pogreške standardne devijacije nismo spominjali odmah nakon standardne pogreške aritmetičke sredine i standardne pogreške razlike između aritmetičkih sredina. No ubrzo će biti jasno zašto je to tek sada učinjeno.

Kao što je poznato, uzimajući velik broj uzoraka iste veličine iz neke populacije, dobit ćemo niz aritmetičkih sredina uzoraka, koje će se distribuirati po *normalnoj* raspodjeli, kojoj se standardna devijacija (nazvana "standardna pogreška aritmetičke sredine") računa prema formuli: $s_{\overline{X}} = s/\sqrt{N}$. Jednako tako i uzorci *razlika* između dvije aritmetičke sredine uglavnom se distribuiraju po normalnoj raspodjeli. Ako se pak radilo o vrlo malim uzorcima, izračunata raspodjela uzoraka davala je tzv. t-raspodjelu. Prema tome, *logika* rezoniranja pri izračunavanju standardne pogreške aritmetičke sredine ili standardne pogreške razlike bila je vezana uz *simetričnost* normalne raspodjele, ili simetričnost t-raspodjele.

Kada bismo radili s *velikim* uzorcima (statističari se potpuno ne slažu u tome koliko veliki, neki misle već od 30, a drugi izričito tvrde da moraju biti veći od 100), distribucija varijanci odnosno standardnih devijacija uzoraka (oko standardne devijacije populacije), uglavnom je *normalna*, dakle simetrična, pa se možemo poslužiti nama već poznatom logikom rezoniranja, tj. da je oko 68% slučajeva u intervalu $s\pm$ jedna standardna pogreška, 95% u intervalu 2 standardne pogreške, itd.

U tim slučajevima, dakle kod *velikih* uzoraka, nema problema sa standardnom pogreškom standardne devijacije, i ona se izračunava prema formuli:

$$s_s = \sqrt{\frac{s}{2N}}. \tag{16.1}$$

Uzmimo da na 450 učenika, primijenivši jedan test znanja, dobijemo neki prosječan rezultat (koji u ovoj raspravi nije važan) i standardnu devijaciju $s = 30$ bodova. Standardna pogreška te standardne devijacije bila bi:

$$\frac{s}{\sqrt{2N}} = \frac{30}{\sqrt{900}} = 1.$$

Budući da u ovom slučaju, kako rekosmo, možemo smatrati da se standardne devijacije tako velikih uzoraka distribuiraju po aproksimativno normalnoj raspodjeli, upotrijebit ćemo istu logiku zaključivanja kao i kod standardne pogreške aritmetičke sredine. Postoji oko 68% vjerojatnosti da naša izračunata standardna devijacija ne odstupa od "prave" standardne devijacije (σ) više od jedne standardne pogreške, dakle više od 1, postoji oko 95% vjerojatnosti da od nje ne odstupa više od $2 \cdot 1 = 2$, i gotovo smo potpuno sigurni da ne odstupa više od $3 \cdot 1 = 3$. Ili, drugim riječima, *95%-tne granice pouzdanosti* za našu standardnu devijaciju iznose $30 \pm 2 = 28 - 32$.

Međutim, radimo li s *malim* uzorcima, distribucija standardnih devijacija nije više normalna distribucija, *pa makar uzorke uzimali iz posve normalne populacije!* Ta je distribucija *asimetrična* (nešto manje za standardnu devijaciju nego za varijancu), pa stoga nije moguće granice pouzdanosti računati na ovaj način, jer bismo tim načinom dodavali i oduzimali *jednako* na lijevoj i desnoj strani od naše standardne devijacije, a to nije ispravno ako je distribucija asimetrična. (Slično se o tome raspravljalo kod testiranja standardne pogreške malih proporcija, vidi nomogram na sl. 11.1, str. 164).

Stoga u takvim slučajevima i ne možemo drugo, nego jedino izračunati — i to na nov način — *granice pouzdanosti* naše standardne devijacije. (Standardnu pogrešku nema smisla računati, jer treba različite vrijednosti dodati lijevo i desno.)

Kad bismo iz velikog broja jednako velikih uzoraka izračunavali varijance, i kad bi nam ujedno bila poznata prava standardna devijacija populacije (σ), onda bi se izraz

$$\frac{\Sigma(X-\overline{X})^2}{\sigma^2}$$

distribuirao po — *hi-kvadrat distribuciji* (!) s $(N-1)$ stupnjeva slobode. (Sada je jasno zašto tu temu tumačimo tek sada: razlog je taj što su čitaoci sada već upoznati s hi-kvadrat raspodjelom.)

Kako se vidi, navedeni izraz sadrži u brojniku sumu kvadriranih razlika između svakog rezultata i aritmetičke sredine uzorka, a u nazivniku pravu varijancu populacije. (To što je ta prava varijanca nepoznata, kako će se brzo vidjeti, za samo računanje nije smetnja.) Kao što znamo, izraz $\Sigma(X-\overline{X})^2$ može se pisati i $(N-1)s^2$, pa stoga taj razlomak može glasiti i:

$$\frac{(N-1)s^2}{\sigma^2}.$$

Na jednom primjeru pokazat ćemo izračunavanje granica pouzdanosti standardne devijacije. Uzmimo da imamo uzorak veličine $N = 16$, s varijancom $s^2 = 20\,392$ grama (radi se o jedinicama težine grupe djece). Iz toga slijedi da je standardna devijacija $s = \sqrt{20\,392} = 142,8$.

Za ustanovljivanje granica pouzdanosti potrebna je hi-kvadrat tablica, sada već poznata čitaocima. Pretpostavimo da nas zanimaju 95%-tne granice pouzdanosti dobivene standardne devijacije. U tablici H u Dodatku očitavat ćemo uz $N-1$ stupnjeva slobode, dakle u našem slučaju uz $16-1=15$ stupnjeva slobode. Za ovu svrhu treba očitavati s oba kraja, tj. 95%-tne granice pouzdanosti očitat ćemo u stupcu $P = 0,975$ i u stupcu $P = 0,025$ (2,5% sa svake strane hi-kvadrat distribucije). Te vrijednosti su 6,26 i 27,5. To znači da bi — među svim mogućim uzorcima iz naše populacije — izraz $15\,s^2/\sigma^2$ u 2,5% slučajeva pao ispod 6,26, a u 2,5% slučajeva iznad 27,5. Drugim riječima, u 95% uzoraka taj bi izraz bio između 6,26 i 27,5.

Imamo dakle ovu situaciju:

$$6,26 < \frac{15\,s^2}{\sigma^2} < 27,5.$$

Takav je izraz u matematici poznat pod nazivom "nejednakost" (ili "nejednadžbe"). Da bismo ga mogli računski riješiti, potrebno je spomenuti dva od više pravila u vezi s "nejednakostima":

1. Ako uzmemo recipročne vrijednosti članova nejednakosti, nejednakosti mijenjaju smjer. Na primjer:

$$3 < 5 < 7.$$

Recipročne vrijednosti: $1/3 > 1/5 > 1/7$, ili, što je isto,

$$1/7 < 1/5 < 1/3.$$

2. Nejednakosti ne mijenjaju smjer ako sve članove pomnožimo istim pozitivnim brojem. Na primjer:

$$3 < 5 < 5/\text{puta}10.$$

$$30 < 50 < 70$$

Postupimo li tako u našem primjeru, primijenit ćemo najprije pravilo (1), tj. uzet ćemo recipročne vrijednosti članova nejednakosti. Time će se promijeniti smjer nejednakosti:

$$\frac{1}{6,26} > \frac{\sigma^2}{15\,s^2} > \frac{1}{27,5},\ \text{ili što je isto,}$$

$$\frac{1}{27,5} < \frac{\sigma^2}{15\,s^2} < \frac{1}{6,26}.$$

Želeći se riješiti nazivnika u srednjem izrazu, pomnožit ćemo nejednakosti s $15\,s^2$, a time — u skladu s pravilom (2) — smjer nejednakosti ostaje jednak:

$$\frac{15\,s^2}{27,5} < \sigma^2 < \frac{15\,s^2}{6,26}.$$

Uvrstimo li u taj izraz vrijednost naše varijance $s^2 (s^2 = 20\,392)$, dobivamo:

$$\frac{15 \cdot 20\,392}{27,5} < \sigma^2 < \frac{15 \cdot 20\,392}{6,26}$$

$$11\,123 < \sigma^2 < 48\,863.$$

To su 95%-tne granice pouzdanosti *varijance*, a drugi korijen iz tih vrijednosti su 95%-tne granice pouzdanosti *standardne devijacije:*

$$105,5 < \sigma < 221,0.$$

Dakle, 95%-tne granice pouzdanosti za našu standardnu devijaciju 142,8 su 105,5 i 221,0. Kako se vidi, granice *nisu* (zbog asimetričnosti) jednako udaljene od naše standardne devijacije: donja granica udaljena je 37,3, a gornja 78,2.

Evo još jednog primjera:

Ako na 30 djece dobijemo neki prosječni kvocijent inteligencije (I Q) od recimo $\overline{X} = 118$, sa standardnom devijacijom $s = 15$, koje su 99%-tne granice pouzdanosti dobivene standardne devijacije?

Rješenje: u hi-kvadrat tablici očitamo vrijednosti uz 29 stupnjeva slobode pod $P = 0,995$ i $P = 0,005(1/2\% + 1/2\% = 1\%)$, a to su vrijednosti: 13,1 i 52,3.

Dakle imamo ovu situaciju:

$$13,1 < \frac{29 \cdot 15^2}{\sigma^2} < 52,3.$$

Recipročne vrijednosti:

$$\frac{1}{52,3} < \frac{\sigma^2}{6\,525} < \frac{1}{13,1}.$$

Pomnožimo li nejednakost sa 6 525, dobivamo:

$$\frac{6\,525}{52,3} < \sigma^2 < \frac{6\,525}{13,1}$$

$$124,8 < \sigma^2 < 498,1.$$

Dakle, 99%-tne granice pouzdanosti za standardnu devijaciju su:

$$11,2 < \sigma^2 < 22,3.$$

Još je teža situacija ako uzorci nisu iz *normalnih* populacija. Čitalac će se prisjetiti da smo u poglavlju "Razlika između dvije aritmetičke sredine" spominjali "teorem centralnih granica", koji kaže da će — pod pretpostavkom da su uzorci dovoljno veliki — aritmetičke sredine uzoraka formirati normalnu raspodjelu, čak i onda ako populacija nije distribuirana prema normalnoj raspodjeli. Taj teorem, na žalost, ne vrijedi kod varijance odnosno standardne devijacije: ako matična populacija nije normalno distribuirana, onda *ni kod velikih* uzoraka distribucija standardnih devijacija uzoraka neće dati normalnu raspodjelu (a vidjeli smo, ako su uzorci mali, a matična populacija normalno distribuirana, standardna devijacija dat će hi-kvadrat raspodjelu).

Raspodjela standardnih devijacija uzoraka kod velikih uzoraka uzimanih iz ne-normalne populacije dala bi se za svaki pojedini slučaj doduše matematički izračunati, međutim ti postupci naravno ne dolaze u obzir za nekoga tko nije matematičar i profesionalni statističar, već samo "konzument" statistike.

Stoga se moramo pomiriti s činjenicom da nam računanje "granica pouzdanosti" standardne devijacije — ako nam nije poznato da li se populacija distribuira po normalnoj raspodjeli — može dati i *dosta pogrešne rezultate!* Statističari još nisu pronašli postupak kojim bi doskočili toj teškoći.

Što se tiče testiranja značajnosti *razlika* između dvije standardne devijacije (odnosno dvije varijance), čitalac će se sjetiti da smo taj postupak već obradili, i to u poglavlju o testiranju razlika između aritmetičkih sredina malih uzoraka, kada je trebalo računati zajedničku standardnu devijaciju, ali se ona smjela računati samo ako ne postoji statistički značajna razlika između obje varijance (vidi poglavlje 9.7). To testiranje izvodi se pomoću tzv. F-testa, što je znatno jednostavnije nego testiranje granica pouzdanosti jedne standardne devijacije i upućujemo čitaoca da taj postupak prouči u poglavlju 9.7.

17. OSNOVNI PRINCIPI UZIMANJA UZORAKA

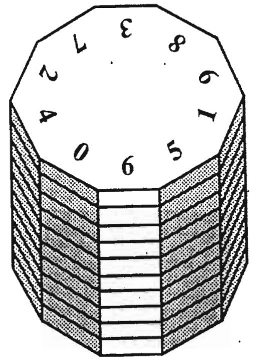
Pri različitim istraživanjima, a naročito kod takozvanih "terenskih istraživanja", stručnjaci različitih struka dolaze u situaciju da iz jedne relativno velike populacije odaberu *uzorak* na kojem će izvršiti mjerenje. Tako npr. antropolog želi ispitati neke biometrijske karakteristike ljudi u nekom kraju, epidemiolog želi ustanoviti postotak neke zaraze među stanovništvom, psiholog želi ispitati stav industrijskih radnika prema nekom faktoru radne okoline, inženjer želi ispitati karakteristike nekog proizvoda, itd.

Iako je u principu važno da uzorak bude što je moguće veći, u mnogim slučajevima uzorak se ipak ograničuje po veličini, i to uglavnom zato što a) povećavanjem uzorka istraživanje postaje sve skuplje i b) što se katkada uzorci (samim tim što su uzorci) uništavaju: npr. ispitivanje čvrstoće materijala, kemijska analiza nekih prehrambenih artikala, itd.

Slučajni uzorak. Među većim brojem različitih uzoraka, najpoznatiji je tzv. *slučajni uzorak*, tj. onaj u kojem *svaki individuum populacije ima jednaku vjerojatnost da bude izabran u uzorak.* Ako neki članovi populacije imaju veću šansu od drugih da budu izabrani, uzorak više nije slučajan, nego se naziva *pristranim* uzorkom (biased sample). Tako bi npr. bio pristrani uzorak kad bi — ispitujući frekvenciju neke zaraze u jednoj pokrajini — uzeli u obzir samo stanovništvo iz gradova, a ne i seosko stanovništvo; ili, kad bi za ispitivanje neke nove metode liječenja uzeli samo dobrovoljce (tj., one ljude koji imaju povjerenje u tu metodu); ili, kad bi sociolog — želeći dobiti mišljenje jedne populacije o nekom problemu — odabrao svoj uzorak iz telefonskog imenika (a u telefonskom imeniku nema — kako je poznato — ljudi iz nekih društvenih grupa), itd. Drugim riječima, uzorak ne smije biti *selekcioniran*, nego mora biti *reprezentativan*. Jedan od načina, koji doduše ne garantira, ali ipak daje veliku vjerojatnost reprezentativnog uzorka, jest uzimanje "slučajnog uzorka".

Slučajnom se uzorku daje u statistici vrlo velika važnost, jer — kako smo vidjeli — mi iz uzorka zaključujemo na populaciju, a to je moguće jedino onda ako je uzorak zaista slučajan. Drugim riječima, gotovo sve vrste uzoraka imaju neku svoju "raspodjelu uzoraka", ali "slučajni uzorci" su jedini koji imaju *poznatu* raspodjelu (većinom normalnu).

Poznati zakoni u vezi s normalnom raspodjelom i vjerojatnošću (npr. pravila u standardnoj pogrešci aritmetičke sredine i standardnoj pogrešci razlike) vrijede samo onda ako je uzorak zaista slučajan — i to je zapravo glavni razlog zbog kojega je "slučajni uzorak" tako važan u statistici.



Slika 17.1. Deseterostrana prizma za izradu tablice slučajnih brojeva.

U vezi s pojmom "slučajni uzorak" postoji katkad nesporazum, kao da se tu radi o "bilo kako izabranom" slučajnom uzorku. To bi npr. bio slučaj kad bismo željeli "bez sistema" ispisati, recimo, stotinu "bilo kojih" brojeva između 1 i 1 000, i na taj način označiti one individuume koji će ući u uzorak. Zbog različitih — nama često nesvjesnih — preferiranja pojedinih brojeva (npr. onih koji sadrže cifru 3 ili 7 ili koju drugu), dolazi do toga da ipak svi brojevi od 1 do 1 000 nemaju jednaku šansu, a to je — kako znamo — uvjet za slučajni uzorak. Ili, ne bi odgovaralo slučajnom uzorku kad bismo odlučili da iz jednog popisa ljudi u uzorak uzmemo one koji nam u letimičnom pregledu "padnu u oko", jer i tu djeluju mnogi, vjerojatno sistematski, faktori (npr. dužina imena, poznatost imena, itd.).

Slučajni uzorak sastavlja se zapravo prema određenim *principima* (dakle ne "bilo kako"), a ti principi odgovaraju zakonu slučaja. Najbolji način sastavljanja slučajnog uzorka je upotreba "tablice slučajnih brojeva". Njezin postanak je lako opisati: bacanjem jedne igraće "kocke", koja zapravo predstavlja deseterostranu prizmu s brojevima od 0 do 9 (vidi sliku 17.1) dobivaju se brojevi koji se redom zapisuju (i to obično u skupinama od po 2 do 5 brojeva), prema tome, kako su slučajno padali. Na primjer, prilikom 40 bacanja jedne takve deseterostrane prizme, dobiveno je ovih 40 brojeva (koji su zbog preglednosti skupljeni u skupine od po 4):

7766 7520 1607 6048 2771 4733 8558 8681 5204 3806.

Brojevi dobiveni na ovaj način zovu se slučajni brojevi, a dobiveni su iz populacije brojeva 0 do 9, u kojoj svaki broj ima jednaku vrijednost da će se pojaviti, i to $p = 0,1$.

Želimo li odabrati uzorak od, recimo, 350 ljudi (na primjer, prilikom nekog intervjuiranja ili sistematskog pregleda u jednoj manjoj tvornici), označit ćemo brojevima ime svakog radnika (uzmimo da ih ima 780), a onda ćemo pomoću "tablice slučajnih brojeva" izabrati uzorak od 350. To ćemo učiniti tako da ćemo najprije tablicu otvoriti na bilo kojem mjestu i od bilo kojeg broja početi "stvarati" uzorak (neki traže da se i te predradnje učine strogo slučajno, tj. da žmireći stavimo vrh olovke na stranicu s brojevima, a i tu stranicu izabrali smo bacanjem novca, itd.). Ako imamo 780 radnika, treba da od mjesta gdje ćemo početi čitamo redom brojeve (bilo u redovima, bilo u stupcima, bilo dijagonalno — to je svejedno) u skupini od *tri* (jer 780 ima 3 znamenke). Učinimo li tako s našim primjerom slučajnih brojeva, prva kombinacija je broj 776, i radnik s tim brojem postaje prvi radnik u našem uzorku. Drugi radnik je broj 675, treći 201, četvrti 607, peti 604. Iduća kombinacija (827) otpada; idući radnik je 714, iza njega 733, 204, itd. Tako ćemo iz popisa vaditi popis ljudi tako dugo dok ne skupimo uzorak od 350. Ako se koji broj pojavi ponovno, više ga naravno ne uzimamo u obzir.

Tablica K u Dodatku je jedna od manjih tablica slučajnih brojeva.

N a p o m e n a. Neka džepna elektronska računala imaju mogućnost da prema određenom algoritmu izbacuju po dvije do tri znamenke slučajnih brojeva. U većini slučajeva to je posve dobra i "bezopasna" tehnika proizvodnje slučajnih brojeva, ali je ipak dobro da se vlasnik takvog računala prethodno uvjeri ne pojavljuje li se neki od brojeva možda prečesto ili odveć rijetko. To se postiže vrlo jednostavno tako da se producira nekoliko stotina slučajnih brojeva, i da se hi-kvadrat testom testira odstupa li frekvencija svakoga pojedinog broja od teoretske frekvencije (koja za svaku znamenku mora iznositi 1/10 od ukupnog broja izbačenih znamenki).

Sistematski uzorak. Umjesto slučajnog uzorka u praktičnom se radu često primjenjuje jedan drugi tip uzorka, tzv. *sistematski uzorak*, koji većinom daje jednako reprezentativan uzorak kao i slučajni uzorak. Sistematski se uzorak sastoji u tome da se prema jednom popisu (npr. popis ljudi prema abecednom redu) odabere slučajnim izborom jedan, a nakon toga se u uzorak uzima svaki n-ti (npr. svaki drugi, ili svaki peti, deseti, itd., već prema tome kako velik uzorak želimo). Takav će uzorak po svom efektu biti sličan slučajnom uzorku samo onda ako je *lista* zaista sastavljena bez nekog smislenog sustava (a to je npr. slučaj ako je lista sastavljena prema abecedi).

Stratificirani uzorak. Osim slučajnog i sistematskog uzorka, poznat je još i tzv. *stratificirani uzorak* (slojevit uzorak), koji se sastoji od toga da se populacija podijeli u "potpopulacije" ili "slojeve" ("stratume"), prema nekim karakteristikama, te da se nakon toga iz svake od grupa uzme slučajni uzorak. U takvim je slučajevima reprezentativnost uzorka često bolja nego kod slučajnog uzorka. Na primjer, želimo li ustanoviti morbiditet od tuberkuloze u nekom kraju, opravdano je stanovništvo podijeliti u nekoliko "stratuma" (prema godinama, spolu, socijalnom sastavu i dr.) te onda iz tih grupa slučajnim odabiranjem stvarati uzorke. Obično je veličina svakoga slučajnog uzorka iz svake grupe proporcionalna veličini grupe u cijeloj populaciji: na primjer, ako među 10 000 ljudi imamo 60% mladih ljudi, 30% ljudi srednje starosne dobi, a 10% starih, onda se i naš uzorak treba sastojati od 60% mladih, 30% srednje starih i 10% starih. (To su tzv. "proporcionalni" stratificirani uzorci). Aritmetička sredina *svih stratuma zajedno* nije — kako znamo — vrijednost koju dobijemo ako aritmetičke sredine svih stratuma zbrojimo, i rezultat podijelimo brojem stratuma, već treba voditi računa o *veličini* svakog stratuma, i izračunati zajedničku ili *ponderiranu* aritmetičku sredinu. Ako npr. neka populacija sadrži milijun članova, a sastoji se od ova tri sloja: sloj A, veličine 500 000, sloj B veličine 300 000, i sloj C veličine 200 000, a njihove aritmetičke sredine iznose: $\overline{X}_A = 100$, $\overline{X}_B = 80$ i $\overline{X}_C = 60$, onda aritmetička sredina za cijelu populaciju iznosi:

$$\overline{X} = \frac{(100 \cdot 500\,000) + (80 \cdot 300\,000) + (60 \cdot 200\,000)}{1\,000\,000}$$

= 86 (a ne 80, koliko bismo dobili da smo 240 podijelili s 3). Ako iz takve populacije uzimamo *vrlo velik* slučajni uzorak, onda nema ozbiljne opasnosti, jer će i u tome velikom uzorku proporcije pojedinih stratuma biti vrlo slične stvarnim proporcijama u populaciji. No kod malih uzoraka *potrebno je* namjerno sastavljati uzorak od poznatih postotaka, što ga svaki stratum u populaciji zauzima.

Neki statističari preporučuju i neke drugačije postupke. Tako se npr. "neproporcionalni" stratificirani uzorci sastavljaju tako da su odnosi veličina pojedinih stratuma u istom omjeru kao što su *produkti standardne devijacije i veličine uzorka u pojedinom stratumu.* Na primjer, kada bi stratum 1, veličine 1 000, imao stand. devijaciju 5, a stratum 2, veličine 100, stand. devijaciju 20, onda bi odnos umnožaka veličine uzorka i stand. devijacije bio

$$(1\,000 \cdot 5)/(100 \cdot 20) = 5/2.$$

Dakle, optimalni odnos veličina uzoraka treba biti 5 : 2; u uzorku veličine 70, 50 podataka mora biti iz uzorka 1 , a 20 iz uzorka 2. Zajednička aritmetička sredina izračuna se nakon toga primjenom formule (4.5) za ponderiranu aritmetičku sredinu. (Vidi pobliže o tome u: Edgington, E.S: Statistical inference: the distribution-free approach. McGraw-Hill, 1969.)

Statističari upozoravaju i na povremenu opasnost od takozvanih "indirektnih" uzoraka: ako bismo recimo željeli doznati prosječnu starost otaca školske djece, mogli bismo učiniti i dosta ozbiljnu pogrešku kada bismo uzorak otaca birali *pomoću uzorka djece,* tj. kada bismo po slučaju odabrali izvjestan broj školske djece i tražili od njih podatke o starosti njihovih otaca. Pogreška se naime sastoji u tome, što očevi s više djece imaju *i u školi* više djece, a ti očevi u prosjeku su *stariji* od očeva sa samo jednim djetetom. Dakle, u slučajnom uzorku školske djece može biti *više djece iz većih obitelji,* pa će kod starosti otaca prema tome prevladavati podaci od *starijih* očeva!

Klaster uzorci. Ti uzorci predstavljaju *lošiju* varijantu slučajnog uzorka (djelomično i stratificiranog uzorka), a vrlo se često koriste u velikim ekonomskim, političkim ili tržišnim istraživanjima, jer bi uzimanje slučajnog uzorka bilo izvanredno skupo. Konkretno, ako npr. treba dobiti mišljenje stanovnika jednog velikog grada o nekom pitanju, onda se populacija toga grada (npr. plan grada) podijeli na 50 ili više manjih blokova, pa se *po slučaju* odabere određen broj tih blokova, i pošalju se anketari da intervjuiraju stanovnike u njima. Važno je da se intervjuiraju svi stanovnici toga bloka (koji dolaze u obzir za intervju, dakle uglavnom odrasli), i intervjuer se mora toliko dugo vraćati na adrese ljudi, koje nije pronašao, dok ih konačno ne pronađe. Takav klaster uzorak naziva se *jednostupanjski* klaster uzorak, no mogu postojati i *dvostupanjski, trostupanjski, višestupanjski* uzorci: npr. u nekom velikom univerzitetskom gradu od svih fakulteta mogu se po slučaju izabrati samo neki, u tim fakultetima po slučaju samo pojedini odsjeci, u odsjecima po slučaju samo pojedini studenti.

Kvotni uzorci. To je još lošiji uzorak, jer on u stvari predstavlja *"neslučajni"* stratificirani uzorak, a najčešće se koristi kod različitih "ad hoc" organiziranih istraživanja za potrebe tržišta, za prikupljanje mišljenja građana o nekom pitanju, i sl. Takvi se uzorci formiraju na taj način, da organizator istraživanja, poznajući strukturu stanovništva s obzirom na predmet istraživanja (npr. postotak mladih majki, postotak umirovljenika i dr.) unaprijed izabere *broj ljudi* svakog pojedinog stratuma, koje intervjuer mora intervjuirati (zato se uzorak naziva kvotni). No glavni problem takvog uzorka je u tome što intervjuer, obilazeći gradom, zaustavlja i moli za suradnju ljude *koje on odabere,* a većinom se događa da i ne prilazi ljudima koju mu ne izgledaju simpatični ili pripravni na razgovor. Osim toga na taj način praktički su potpuno isključeni iz ankete građani koji se prevoze javnim prometnim sredstvima, oni koji se voze na posao svojim vozilom, kao i oni koji u to vrijeme rade na radnim mjestima. Poznat je slučaj s intervjuerom koji je imao zadatak da intervjuira unaprijed određeni broj muškaraca o njihovu stavu prema klađenju, pa je — putujući lokalnim vlakom prema centru grada — iskoristio priliku da na željezničkoj stanici intervjuira veću skupinu ljudi koji su čekali vlak. Iz rezultata se ustanovilo da su gotovo svi intervjuirani imali pozitivno mišljenje o klađenju, a provjeravanje situacije i mjesta na kojem je intervju izvršen, pokazalo je da je intervjuer slučajno intervjuirao ljude koji su čekali na peronu odakle je odlazio vlak prema trkalištu (a to su upravo oni koji se većinom klade prije konjskih utrka!).

Prigodni uzorak. To je onaj uzorak, koji nam se "nađe pri ruci", jer drugoga nemamo. To su npr. *momentano prisutni* bolesnici u nekom kliničkom odjelu, *postojeći studenti* neke godine studija, i sl. Nije potrebno naglasiti da prigodni uzorci *mogu biti ekstremno pristrani:* tipičan primjer takvog uzorka je uzorak žena koje odlaze u psihijatrijsku ambulantu zbog različitih razloga, a psihijatar nakon nekoliko stotina pregledanih slučajeva objavi koliki postotak žena je — frigidno.

No dakako, prigodni uzorci *ne moraju* biti uvijek pristrani: ako se recimo radi o ispitivanju trajanja vidne paslike nakon 10 sekundi promatranja svjetlosnog izvora, *nema nikakvog razumnog razloga* zbog kojeg bismo mogli pomisliti da se u tom pogledu razlikuju studenti različitih struka, ili ljudi različitog stupnja obrazovanja, kvalifikacije i sl. Prema tome, ako nam je zavisna varijabla (dakle ono što istražujemo) takve vrste da nije moguće pretpostaviti utjecanje nekih drugih situacijskih faktora na zavisnu varijablu, već možemo opravdano smatrati da jedino naša nezavisna varijabla, čije djelovanje ispitujemo, može imati utjecaja — prigodni uzorci mogu biti posve upotrebljivi, i — kao što je poznato — u brojnim znanstvenim istraživanjima oni se i koriste jer su to uzorci do kojih najlakše dolazimo.

Pogreške nekog istraživanja mogu se podijeliti na dvije glavne skupine:

(a) na pogreške *izvan uzorka* (to su pogrešna metoda, pogrešna obrada rezultata, itd.) i

(b) pogreške *uzorka.*

Obje te pogreške zajedno dovode do *ukupne pogreške* istraživanja. Ako u nekom istraživanju zaista postoje obje vrste pogrešaka, onda dakako smanjivanje samo *jedne* od njih ne poboljšava u većoj mjeri situaciju. Neki autori to uspoređuju s pravokutnim trokutom kod kojeg svaka kateta predstavlja jednu od ovih pogrešaka, a hipotenuza je ukupna pogreška: ako jednu od kateta značajno smanjimo, hipotenuza se *neće bitno skratiti*.

17.1. VELIČINA UZORKA

Jedno od čestih pitanja koja se postavljaju statističarima jest pitanje *koliko velik mora biti uzorak*. Obično dolazi do razočaranja kad statističar odgovori da je na to pitanje vrlo teško dati precizan odgovor. Iako u tome ima za *neke specijalne slučajeve* već gotovih pravila, tablica i nomograma iz kojih se može po prilici odrediti potrebna veličina uzorka (o tome vidi udžbenike koji govore o planiranju eksperimenta), ipak u velikom broju slučajeva na to pitanje *nije moguće odgovoriti konkretnom brojkom* jer potrebna veličina uzorka ovisi prvenstveno *o varijabilnosti pojave koju mjerimo, a potom o preciznosti kojom želimo pojavu izmjeriti*. Ako je pojava malo varijabilna, bit će dovoljan i manji uzorak, a kod jako varijabilne pojave, potreban je veliki uzorak. S druge strane, ako nam nije stalo do velike preciznosti, možemo se zadovoljiti i s manjim uzorkom; što veću preciznost želimo, to nam je veći uzorak potreban. Želimo li pogrešku aritmetičke sredine smanjiti na polovicu, treba, kako znamo, uzeti otprilike četiri puta veći uzorak.

Primjer. Mjereći visinu djece određenih godina u nekom kraju, na uzorku od $N = 49$ dobili smo $\overline{X} = 135,5$, $s = 14$, $s_{\overline{X}} = 2$. Prema tome, 95%-tne "granice pouzdanosti" naše aritmetičke sredine su $131,58 - 139,42$ (iz $\overline{X} \pm 1,96\ s_x$). Kako se vidi, oduzmemo li i pribrojimo aritmetičkoj sredini 1,96 standardne pogreške, dobili smo "razmak" unutar kojeg će se — s 95% sigurnosti — kretati prava aritmetička sredina. Taj razmak dakle iznosi: $2 \cdot 1,96\ s_x$.

Ako smatramo da je taj "raspon" od gotovo 8 cm prevelik, tj. da bismo morali imati *precizniju* aritmetičku sredinu, znamo da moramo povećati uzorak. Uzmimo da želimo raspon "granice pouzdanosti" smanjiti na 3 cm. Taj broj treba dakle uvrstiti u jednadžbu:

$$3 = 2 \cdot 1,96 \cdot \frac{s}{\sqrt{N}}.$$

Riješimo li ovu jednadžbu, proizlazi da moramo imati uzorak veličine oko 334 djeteta da bi se "granice pouzdanosti" kretale u rasponu od oko 3 cm.

Ako malo bolje pogledamo formulu za izračunavanje standardne pogreške uzorka ($s/\sqrt{N}$), može nam se učiniti da ona zapravo ne odgovara potpuno onome što u praksi o uzorku mislimo, jer ta formula ništa ne govori o tome koliko se pouzdanost uzorka povećava time što uzorak predstavlja veći *dio* (frakciju, postotak) populacije. Kadikad se čak misli i ističe kako uzorak treba predstavljati barem oko 10% populacije da bi bio reprezentativan. *To, međutim, nije točno*. Iz formule o standardnoj pogrešci uzorka vidimo da ona ovisi jedino o varijabilitetu pojave koju mjerimo i o *apsolutnoj veličini* uzorka, a ne, dakle, o njegovoj *relativnoj veličini*. To praktički znači da ćemo — mjereći, pretpostavimo, visinu studenata na različitim fakultetima — s uzorcima od $N = 50$ dobiti *jednako pouzdane* podatke na fakultetima s 10 000 studenata, kao i na fakultetima s 1 500 studenata. S *većim* uzorkom dobit ćemo *u svim* tim slučajevima *jednako bolje* i pouzdanije rezultate. Pitati statističara: "Koliki postotak populacije treba uzeti u uzorak?" bilo bi — kako kažu Wallis i Roberts — jednako kao i pitati kuhara: "Koliki postotak brašna iz posude treba staviti u kolač?" U oba slučaja odgovor je isti, tj. da je potrebna određena *količina*, a koliko, to ovisi o tome što se želi postići, o tome kolik je varijabilitet pojave koju mjerimo, bez obzira na to koliko je velika populacija (ili "koliko brašna ima u kutiji").

Proporcija populacije uključena u uzorku ima samo *vrlo blag* utjecaj na standardnu pogrešku aritmetičke sredine, dakle na preciznost uzorka. Ako je populacija *vrlo mala* (a to se događa izvanredno rijetko), onda, naravno, što je uzorak veći, on ujedno predstavlja i veću proporciju populacije (primjerice, kod populacije veličine 50, uzorak od 25 je već 50% populacije). Međutim, kod velikih populacija (a to je gotovo uvijek) proporcija ima — kako rekosmo — neznatnu ulogu.

Korigirana formula za standardnu pogrešku aritmetičke sredine, s obzirom na proporciju populacije u uzorku, glasi:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} \cdot \sqrt{\frac{N_p - N}{N_p - 1}}, \tag{17.1}$$

pri čemu N_p znači veličinu populacije, a N je veličina uzorka. Prema tome, desni izraz pod korijenom uvijek je nešto manji od 1 (osim ako je N_p neizmjerno velik), i to nešto malo smanjuje vrijednost $s_{\overline{X}}$, ali čak kad je uzorak 20% populacije, $s_{\overline{X}}$ bio bi smanjen samo oko 10%.

Kao dokaz svemu izloženom neka posluži jedan primjer uzimanja uzorka u industriji. Pretpostavimo da se u jednoj tvornici električnih žarulja uzimaju uzorci kojima se provjerava kvaliteta proizvoda. Pretpostavimo, nadalje, da je 20% žarulja u produkciji defektno (dakle 1/5), dok je 80% (4/5) bez pogreške. Pretpostavimo također da iz svake "šihte" produkcije uzimamo uzorke veličine $N = 2$.

Ako se "šihte" sastoje od 5 žarulja, od kojih je (u prosjeku) 1/5 defektnih, znači da u tim skupinama postoje u prosjeku 4 ispravne žarulje i jedna defektna. Koja je vjerojatnost da ćemo uzorkom od $N = 2$ (koji iznosi 40% populacije, ili "šihte") dobiti obje žarulje ispravne? Podsjetimo se jednog od osnovnih zakona multiplikacije (vidi str. 33), pa ćemo lako ustanoviti da ta vjerojatnost iznosi:

$$4/5 \cdot 3/4 = 0,6,$$

što znači da će prosječno od 1 000 "šihta" njih oko 600 "proći" kao ispravne.

Ako su "šihte" veće i sadrže 20 žarulja, naš uzorak od $N = 2$ predstavlja sada samo 10% populacije. Vjerojatnost da ćemo izvući dvije ispravne žarulje iz populacije, iznosi:

$$16/20 \cdot 15/19 = 0,632.$$

U "šihtama" od 100 žarulja bit će u prosjeku 20 defektnih, a 80 ispravnih. Vjerojatnost da ćemo uzorkom veličine $N = 2$ (2% populacije!) dobiti 2 ispravne žarulje iznosi:

$$80/100 \cdot 79/99 = 0,638.$$

Kad bi "šihte" bile goleme, na primjer 10 000 žarulje, uzorkom od $N = 2$ (0,02% populacije!), imali bismo ovu vjerojatnost:

$$8\,000/10\,000 \cdot 7\,999/9\,999 = 0,640.$$

Kako se, dakle, vidi, razlike su veoma male i postaju tek *nešto* značajnije ako uzorak iznosi relativno velik postotak populacije (u našem prvom slučaju on je 40% populacije).

Za uzorak je mnogo važnije da bude *reprezentativan* nego da bude velik, a reprezentativnost uzorka postiže se pažljivo skupljenim slučajnim uzorkom. Vrlo dobar primjer kako i *vrlo velik* uzorak, ako nije reprezentativan, dovodi do pogrešnih rezultata jest slučaj s ispitivanjem javnog mnijenja u SAD 1936. godine prije izbora predsjednika. Jedan je časopis organizirao anketiranje upravo golemog uzorka: anketirano je 2 300 000 ljudi o tome za koga će glasati. Prema rezultatima te ankete dana je prognoza da će na izborima pobijediti republikanski kandidat London. No, dogodilo se suprotno, tj. na izborima je pobijedio demokratski kandidat F. D. Roosevelt. Razlog pogrešnoj prognozi bio je "pristrani" uzorak, tj. uzorak koji uopće nije bio reprezentativan: ljudi koji su uzeti u uzorak, birani su iz *telefonskih imenika,* a u ono vrijeme telefon su posjedovali samo imućniji slojevi građana; na taj način u uzorku je bilo vrlo malo radnika i poljoprivrednika!

Ekonomski statističar Weber daje za neke specifične slučajeve ovaj praktični savjet za veličinu uzorka: ako možemo "grubo" predvidjeti u kojem postotku je neko svojstvo zastupljeno u populaciji, onda veličinu uzorka možemo donekle odrediti tako da taj postotak pomnožimo s postotkom koji nedostaje do 100%. Na primjer, ako smatramo da oko 5% populacije posjeduje neku karakteristiku koju bismo željeli detaljnije ispitati, onda treba uzeti uzorak veličine $5 \cdot 95 = 475$; naprotiv, ako je postotak te karakteristike u populaciji oko 50%, potrebni uzorak bit će veličine $50 \cdot 50 = 2\,500$. (U ovom primjeru lijepo se može vidjeti kako kod postotaka vrijednosti bliže nuli ili 100 znače mali, a vrijednosti oko 50 veliki varijabilitet. Stoga kod 50% trebamo i veliki uzorak. Vidi o tome eksperiment na slici 9.1)

Danas postoje čitave biblioteke knjiga o uzorcima i njihovom sastavljanju. Detaljna pitanja o različitim vrstama uzoraka pripadaju već u opširnu problematiku *planiranja eksperimenta,* pa čitaoci koji o tome žele dobiti potanke informacije, treba da konzultiraju poglavlja o planiranju eksperimenata u nekim većim statistikama (primjerice, Wallis i Roberts, Statistics, a New Approach).

ZADACI ZA VJEŽBU

1. Pretpostavimo da iz skupine od 200 ljudi želite izabrati slučajni uzorak veličine $N = 10$, i da se pri tome služite tablicom slučajnih brojeva u ovoj knjizi (tablica K), te da pri izboru tablicu počnete čitati na prvoj stranici *vertikalno prema dolje,* počevši od prvoga vertikalnog stupca (7, 9, 4, 6, 8 ... itd.). Napišite od kojih će se brojeva sastojati vaš uzorak veličine 10.

18. ZAKLJUČIVANJE U STATISTICI

Jedna od čestih pogrešaka početnika u statističkom radu sastoji se u tome da se na temelju dobivenih rezultata donose suviše dalekosežni zaključci, ili čak takvi koji se ne mogu ili ne smiju donositi na osnovi dobivenih rezultata. Već smo u prvom poglavlju upozorili da statistička obrada rezultata *ne garantira* znanstvenu vrijednost nekog istraživanja jer se statistički mogu obraditi i površno prikupljeni podaci, nadalje podaci koji uopće nisu relevantni za problem i, konačno — izmišljeni podaci! Ako bi se, na primjer, netko sjetio da prikupi sve kućne brojeve desne strane jedne ulice i da iz dobivenih brojeva izračuna "prosječni kućni broj" te strane ulice, ne bismo u tom pothvatu vidjeli baš mnogo smisla, iako su rezultati — statistički obrađeni!

Čak i u slučajevima kada su podaci ispravno i savjesno prikupljeni, osnovni je problem, o kojem zapravo ovisi hoće li rezultati imati ili neće imati znanstvenu vrijednost, u tome kako su ti rezultati interpretirani. Engleski statističar Moroney kaže da je vrlo teško interpretirati rezultate, a da je, naprotiv, izvanredno lako provesti na dobivenim brojevima statističke račune. "Aritmetičke se sredine mogu začudno lako izračunati na devetnaest decimala. Kada je to učinjeno, rezultat se čini da je vrlo točan. Ali ništa nije lakše i štetnije nego iz toga zaključiti da je točnost izračunavanja ekvivalentna točnosti našeg znanja o problemu koji istražujemo."

N a p o m e n a. Ovdje usputno možemo spomenuti pojavu da su ljudi obično prilično impresionirani nekim rezultatima, ako su ti rezultati prikazani *decimalnim brojevima!* Uz decimalni broj kao da se obavezno asocira i povjerenje u rezultate, koji nam se — zbog decimalnog broja — čine vrlo precizni i pouzdani. No, to, dakako, ne mora biti točno, o čemu ima dosta svakodnevnih primjera: vrlo često u dnevnoj, pa i stručnoj, štampi možemo pročitati da "prosječna ocjena" nekog ispita iznosi, recimo, 4,135, i ta preciznost gotovo da nam ulijeva strahopoštovanje! Međutim, onaj tko znade koliko je sama *ocjena,* što je nastavnik daje na ispitu, subjektivna, nepouzdana i podložna nizu slučajnih i neslučajnih faktora, toga će naravno taj decimalni broj mnogo manje impresionirati. Poznato je pravilo kod računanja: ako prethodne operacije računamo na točnost jedne decimale, onda nema nikakvog smisla konačan rezultat izraziti, na primjer, na 3 ili 4 decimale! (Ipak, to, na žalost, mnogi rade!) Jednako tako, prosjek od brojeva koji imaju samo donekle karakteristike mjerne ljestvice, ne treba izražavati s velikom preciznošću, jer ta je preciznost — fikcija.

U ovoj smo knjizi već nekoliko puta, a posebice u poglavlju o korelaciji, upozoravali na opasnost pogrešne interpretacije rezultata. Kod korelacije specijalno je upozoreno na opasnost *kauzalnog* interpretiranja povezanosti između dviju varijabli. Neće škoditi da jednim novim primjerom čitaoce podsjetimo na vrstu te opasnosti.

Svojedobno je jedan medicinski članak izazvao prilično veliku uzbunu objavljivanjem podataka da navodno s povećanom količinom potrošnje mlijeka raste i učestalost oboljenja od raka. U članku je naime objavljeno da je rak mnogo češći u Novoj Engleskoj, Minnesoti, Wisconsinu i Švicarskoj, nego na Ceylonu, a poznato je da stanovnici spomenutih država mnogo više piju mlijeko nego stanovnici na Ceylonu. Tako se u članku nadalje tvrdilo da engleske žene (koje piju mnogo mlijeka) dobivaju rak osamnaest puta češće od japanskih žena, koje rijetko piju mlijeko. Analizom tih rezultata moglo bi se pronaći dosta različitih tumačenja toj pojavi, no jedna od interpretacija je dovoljna da obori *kauzalno* tumačenje povezanosti između trošenja mlijeka i frekvencije oboljenja od raka: rak je — kako znamo — bolest koja se pretežno pojavljuje kod ljudi *srednje i starije* dobi; Švicarska i ostale spomenute države imaju prosječno *stariju* populaciju od Ceylona, a engleske žene u vrijeme tog istraživanja prosječno su živjele 12 godina duže od japanskih žena.

Sada ću pokušati navesti još nekoliko skupina primjera pogrešnog zaključivanja. Neki su od njih izmišljeni, a neki su istiniti — ali i jedni i drugi su jednako mogući da se u stvarnosti dogode. Neki od tih primjera predstavljaju već toliko iskrivljavanje činjenica da ih se može svrstati u namjerno ili nenamjerno "laganje", jer toliko pogrešno interpretiraju rezultate da zaključak glasi upravo obratno od onog kakav bi zapravo trebao biti.

Prva grupa primjera:

Jedan je istraživač zarazio 10 laboratorijskih miševa nekom opasnom bolešću i nakon toga dao im je određenu količinu nekog lijeka što ga je sam izmislio. Nakon određenog vremena dva su miša uginula, a osam ih je ozdravilo. Eksperimentator je iz tog pokusa zaključio ovo: lijek koristi u 80% slučajeva.

Analiziramo li detaljnije opisani (izmišljeni) primjer, brzo možemo ustanoviti da je eksperimentator učinio barem dvije bitne pogreške:

a. On je izračunavao postotak iz vrlo malog broja mjerenja, gdje dobiveni rezultat može biti i posljedica čestog slučaja. (Kada biste bacanjem 10 komada novca dobili 1 "glavu" i 9 "pisama" biste li sigurno mogli tvrditi da novac nije ispravan?)

b. Eksperimentator nije imao *kontrolnu grupu* miševa (iz istog legla, istog spola i starosne dobi), koja je također morala biti istodobno zaražena; tek iz razlike između onoga što se dogodilo u eksperimentalnoj i onoga što se dogodilo u kontrolnoj grupi (pod pretpostavkom da su obje grupe dovoljno velike), mogli bi se donositi sigurniji zaključci. Ovako, bez kontrolne skupine, gotovo jednakom logikom moglo bi se zaključiti da taj lijek *ubija* u 20% slučajeva! Drugim riječima, da nisu primili lijek, možda bi svi miševi ostali živi.

Druga grupa primjera:

Dva kaveza miševa zaražena su jednom bolešću (eksperimentalna i kontrolna grupa). Miševi iz prvog kaveza liječeni su novom metodom, a miševi iz drugog kaveza standardnim načinom. Pretpostavimo da je konačan rezultat eksperimenta ovaj: u prvom kavezu uginulo je 5% miševa, u drugom 28%. Statističkim postupkom našli smo da je razlika u proporciji uginulih *statistički značajna*. Idući korak obično je zaključak: nova metoda liječenja bolja je od stare. To je, međutim, točno samo onda ako su u *svim ostalim faktorima* (osim u metodi liječenja) obje grupe miševa bile potpuno jednake, a to znači: iz istog legla, jednako stari, jednako ishranjeni, držani inače pod jednakim prilikama, itd.

Jer, statistički značajna razlika u proporciji uginulih može biti uzrokovana i drugim faktorima, recimo time što su u kavezu, koji je liječen novom metodom, bili stariji, pa prema tome otporniji miševi.

Prema brojnim istraživanjima proporcija duševnih bolesti kod muškaraca statistički je značajno veća od proporcije duševnih bolesti kod žena. Iako je razlika statistički značajna, ona još ne govori o tome kako ćemo tu razliku interpretirati. Jer, neki, na primjer, opravdano smatraju da način života muškaraca, tj. njihov rad i namještenje omogućuje više *otkrivanja* duševnih bolesti kod njih nego kod žena.

Jedan je istraživač svojedobno našao značajne anatomske promjene u nekim unutarnjim organima miševa koji su bili izloženi buci, a kod miševa koji nisu bili u buci, takvih promjena uopće nije bilo. Istraživač je pokušao nađene anatomske promjene pripisati specifičnom djelovanju buke. Međutim, pokazalo se da su jedni i drugi miševi bili u kavezima, i da su oni koji su bili izloženi buci do iscrpljenosti trčali uplašeno po kavezu, dok su se oni drugi kretali kao obično. Nađene anatomske promjene vrlo su vjerojatno bile posljedice tjelesne aktivnosti, a ne specifičan efekt buke.

Godine 1925. jedan je istraživač na temelju velikog broja slučajeva našao da je mortalitet djece koja su rođena unutar jedne godine nakon poroda prijašnjeg djeteta, oko jedan i pol puta veći od mortaliteta djece rođene nakon intervala od 2 ili više godina. Autor je iz toga zaključio da je tome uzrok kratki razmak među porodima, te se počelo ozbiljno govoriti o važnosti dovoljnoga vremenskog raspona između sukcesivnih poroda.

Taj zaključak međutim *nije* opravdan (iako je možda i zaista važno da među porodima bude dovoljan razmak), i to zato što *oba uspoređena uzorka djece nisu bila komparabilna:* u uzorku djece rođene unutar jedne godine nakon prethodnog poroda, nužno je moralo biti više nedonošene djece nego u drugom uzorku. Naime, samo djeca koja su rođena barem 9 mjeseci nakon poroda prijašnjeg djeteta nisu bila nedonoščad, a ostala su bila. A poznato je da je mortalitet nedonošene djece znatno veći od mortaliteta ostale djece.

Statistički značajna razlika sama nam po sebi ništa ne govori o tome čemu tu razliku treba pripisati! Kod pogrešno planiranih eksperimenata razlika može biti uzrokovana *nekomparabilnim uzorcima.*

Treća grupa primjera:

U našem primjeru na str. 259 opisali smo slučaj različitog stava radnika i radnica prema tvorničkoj liječnici: od 23 muškarca negativan stav je imalo njih 9 ($p = 0,39$), a pozitivan 14 ($p = 0,61$), dok je kod žena situacija bila obratna: od 26 žena negativan stav imalo je njih 17 ($p = 0,65$), a pozitivan stav njih 9 ($p = 0,35$). Tako velika razlika između muškaraca i žena sugerirala je da je ta razlika značajna. Međutim, hi-kvadrat dao nam je vrijednost 2,399, što je manje od granične vrijednosti uz 1 stupanj slobode, pa smo zaključili da razlika nije statistički značajna. Pogrešno bi međutim bilo zaključiti da razlike nema. Ako bi iste razlike u proporciji bile nađene uz veći N, razlika bi bila statistički značajna. Prema tome, jedini ispravan zaključak iz takvog rezultata mogao bi se formulirati ovako: Iz dobivenog materijala *ne može se zaključiti* da postoji razlika između muškaraca i žena. Ili, kako kaže Mainland: "*Nije značajno*, to zapravo znači *nije dokazano*".

Prema tome: *prihvaćanje nul-hipoteze još nije garancija da je ta hipoteza jedina istinita i ispravna.* Njezino prihvaćanje znači samo da iz evidencije koju posjedujemo, za sada ne možemo izvesti drugi zaključak. O tome R. A. Fisher, najveći statističar do danas, kaže ovo "... Jedino što test značajnosti daje, i želi da daje, jest odgovor na izravno pitanje: 'Bi li ova dva uzorka mogla pripadati istoj populaciji?' Test značajnosti izračunava vjerojatnost. Ako je vjerojatnost vrlo mala odgovor je 'ne'. Ako nije tako mala, da bi dosegla traženu razinu značajnosti, odgovor je 'da', mogla bi pripadati istoj populaciji." Odgovor nikada nije: "Da, sigurno su oba uzorka iz iste populacije."

Na ovom mjestu možda je korisno spomenuti jednu posebnost znanstvenih istraživanja uopće, pa prema tome i posebnost statističkih zaključivanja koja se gotovo nikada ne spominje u literaturi i statističkim udžbenicima: naime, znanost uopće, pa tako i statistička obrada rezultata mjerenja, može dokazati (uz određeni rizik pogreške, obično 1% ili 5%) da neki fenomen *postoji* (primjerice, da neka razlika između dvije aritmetičke sredine postoji i među populacijama ili da neka korelacija postoji i u populaciji), ali *ne može dokazati da neki fenomen ne postoji.*

Drugim riječima, ako između dvije aritmetičke sredine *nađemo* neku razliku i statističkim putem dokažemo da je ta razlika *statistički značajna*, onda smo prilično sigurni da ta razlika stvarno među populacijama i postoji. Jednako tako ako *nađemo* neku korelaciju između dvije varijable, pa ustanovimo da je ona statistički značajna, to znači da je gotovo sigurno da korelacija zaista i postoji. No, ako između dvije aritmetičke sredine *ne nađemo* razliku (ili je razlika tako neznatna da nema govora o tome da bi bila statistički značajna), ili pak ako među nekim varijablama ne nađemo korelaciju, to nam *ne dopušta* da zaključimo da razlike odnosno korelacije u populaciji nema! Jedino što smijemo zaključiti — kako je već rečeno — jest to da razliku (ili korelaciju) *nismo mogli ustanoviti.*

Možda će jedan banalan primjer razjasniti zašto je to tako. Ako pomoću nekog teleskopa *vidimo* neko nebesko tijelo, onda gotovo uopće nema sumnje da ono postoji (pod pretpostavkom da nismo imali halucinaciju!) Ali ako tim istim teleskopom na jednom drugom mjestu neba *ne vidimo* nikakvo tijelo — da li to znači da nema nikakvoga nebeskog tijela u tom prostoru?! Naravno da ne znači. Jer:

a. neko nebesko tijelo možda je suviše daleko da bi ga naš teleskop mogao registrirati, ili

b. neko tijelo je možda suviše malo da bismo ga našim teleskopom mogli uočiti, ili

c. neko tijelo možda je "maskirano" neprozirnim plinovima pa je za nas "nevidljivo", itd., itd.

Statističar Runyon u jednoj svojoj popularnoj statističkoj knjižici naveo je velik broj primjera za ovu pojavu, i to osobito iz područja medicine. Tako, na primjer, mi možemo s dosta velikom vjerojatnošću da nismo pogrešno zaključili dokazivati da neko sredstvo *ima* neki određeni efekt (npr. na puls, krvni tlak, krvnu sliku, itd.), ali *ne možemo dokazati* da to sredstvo *nema* nikakvog efekta! Jedino što možemo reći jest to da na onom području na kojem smo promatrali *nismo mogli ustanoviti* nikakav efekt tog sredstva (a to je posve drugo nego tvrditi da sredstvo *nema* efekta!). No, koliko stotina tisuća ili milijuna različitih sistema postoji u nekom organizmu? Za mnoge od njih medicina još i ne zna, pa kako bismo onda mogli tvrditi da neko sredstvo nema efekta! Runyon kao primjer spominje talidomid i pita: "Tko je mogao misliti da će jedno umirujuće sredstvo imati takav poguban utjecaj na neke fetuse u određenoj fazi njihova razvoja?" Čak kao primjer navodi i antibebi-pilule: "Prije desetak godina smatralo se da su bezopasne; prije pet godina smatralo se da su bezopasne za većinu žena; a danas se savjetuje oprez i česti liječnički pregledi."

Četvrta grupa primjera:

U jednom eksperimentu tzv. "ekstrasenzorne percepcije" dano je ispitaniku ("mediju") da zatvorenim očima pogađa kakvi se znakovi nalaze na pojedinim karticama. U grupi od 25 kartica koje su upotrijebljene, postoji pet različitih simbola (kvadrat, krug, križ, itd.), i to svaki simbol na 5 kartica. Prema zakonu slučaja, vjerojatnost da će "medij" slučajno pogoditi simbol na kartici, $P = 1/5$. Iz toga slijedi da je za čitavu grupu najvjerojatnije da će pogoditi $25 \cdot 1/5 = 5$ kartica. Ako ispitanik pogodi *znatno* više od 5 kartica, te statističkim postupkom ustanovimo da tu razliku ne možemo pripisati slučaju, moramo *odbaciti* nul-hipotezu. Međutim, time što smo odbacili nul- hipotezu ne znači da smo priznali da postoji ekstrasenzorna percepcija. Jer, znatno veći broj ispravnih pogodaka nego što bismo mogli slučajno očekivati, može biti uzrokovan različitim faktorima (npr. time da je ispitanik primao signale od nekog prisutnog, da je na neki način ipak vidio što je nacrtano, da je opipom poznavao neke kartice, itd.).

Svojedobno je u dnevnom tisku bilo objavljeno da jednoga određenog dana nije u jednom europskom gradu radio veći broj taksista, jer da su im njihovi "bioritmovi" toga dana bili na minimumu pa je postojala opasnost nesreće. Pretpostavimo da netko želi ispitati je li to istina, i da grupi od nekoliko desetaka vozača izradi prognozu za njihove "kritične dane", te da prati prometne nesreće tih vozača određeno razdoblje, i, uzmimo, da ustanovi da su oni u prosjeku imali zaista povećani broj sudara upravo u svoje "kritične dane". Hi-kvadrat test je dakle odbacio nul-hipotezu. Jesmo li time dokazali da "kritični dani" postoje? Naravno da nismo, jer eksperiment nije bio valjano planiran: morala je biti predviđena kontrolna skupina taksista, koji *ne znaju* za svoje kritične dane. Uz pomoć kontrolne grupe mogli bismo, naime, ustanoviti postoji li *sugestivno* djelovanje znanja o kritičnim danima na nesreće i nezgode u prometu. (Kad smo već kod ovog primjera, spomenimo da *ispravno* provedena istraživanja ove vrste *ne pokazuju* da je broj nezgoda povećan u "kritične dane").

Čitalac će se možda sjetiti slučaja spomenutog u poglavlju o "Osnovnim pojmovima vjerojatnosti", gdje smo spomenuli G. Cardanoa, koji je izvršio samoubojstvo na dan kada je sam sebi prorekao svoju smrt. Takvo "ispunjavanje" proročanstva nazvali smo "proročanstvo koje samo sebe ispunjava". I u ovom bi se, naime, slučaju moglo raditi o sličnom fenomenu: ljudi koji su *očekivali* toga dana nesreću, bili su zbog toga u vožnji nesigurniji pa su zaista i doživjeli prometnu nesreću.

Prema tome, *odbacivanje nul-hipoteze ne mora nikako značiti da treba prihvatiti suprotno stanovište od nul-hipoteze.* (U naša dva primjera suprotno bi stanovište bilo: postoji ekstrasenzorna percepcija, tj. postoje "kritični dani".)

Peta grupa primjera:

Kao što je već isticano, ***statistički*** značajna razlika ne mora uvijek biti i ***praktički*** značajna. Jedan od najpoznatijih primjera na tom području je ljudska visina: poznato je da su svi ljudi navečer otprilike od 1 do 3 mm niži nego ujutro, dakle razlika je, nedvojbeno, statistički značajna, jer ona ***postoji u populaciji***. A praktično značenje te razlike je — nikakvo! Slično, razlika od nekoliko stotina tisuća eritrocita u krvnoj slici može biti statistički značajna, ali bez praktičnog značenja.

Ovaj tip pogrešnog zaključivanja vrlo je čest kod početnika, koji gotovo redovito riječ "značajno" interpretiraju kao "veliko". To se čak donekle i može očekivati, zapravo je normalno jer se u svakodnevnom govoru ta dva pojma izmjenično upotrebljavaju u situacijama kada se želi naglasiti da je nešto ***veliko***. Tako, na primjer, kažemo da je A "značajno" inteligentniji od B, i pri tome mislimo da je on *mnogo* inteligentniji.

No u statistici riječ "značajno" ne znači "veliko", nego znači "nije slučajno".

Pogreška ove vrste najopasnija je pri interpretiranju visine korelacije. Ako negdje pročitamo ili čujemo da je između dvije varijable nađena "statistički značajna" korelacija (kadšto statističari čak napišu samo "značajna"), skloni smo odmah pomisliti da se radi o ***visokoj*** korelaciji. No, naravno, ni to nije točno, jer korelacija, od, primjerice, +0,10 (pa i niža!) može biti "statistički značajna" (ako je uzorak vrlo velik), ali ona je zapravo vrlo ***niska***. Prema tome, ako kažemo da je korelacija od +0,10 statistički značajna, to samo znači ovo: možemo biti gotovo potpuno sigurni da u populaciji između te dvije varijable ***postoji*** povezanost koja je *vrlo niska*.

— — —

U svakodnevnom životu ima naravno i potpuno pogrešnih i primitivnih zaključaka koji se izvode iz nekih podataka što nam ih je pružilo jedno opažanje. Kao školski primjer pogrešnog zaključivanja navest ćemo primjer iz jednih naših dnevnih novina, koje su u listopadu g. 1977. objavile podatak da se na magistralnoj cesti Zagreb-Ljubljana od svih nesreća 60% dogodilo na suhoj, a 40% na mokroj cesti. Iz tih podataka zaključeno je (ispod slike jedne nesreće): "Suh kolnik opasniji je od mokrog." (!) To bi otprilike odgovaralo tvrdnji da su trijezni vozači opasniji od pijanih, jer kod nesreća nalazimo više trijeznih nego pijanih vozača! (Svi

ste vjerojatno čuli za sličnu logiku u šali koja tvrdi da je za život čovjeka krevet najopasniji, jer golema većina ljudi umire u krevetu!)

Jedan je statističar pogreške u interpretiranju i zaključivanju (koje mogu biti nehotične i hotimične) svrstao u tri kategorije:

a. bezazlene zloupotrebe statistike,
b. manje bezazlene zloupotrebe i
c. opasne zloupotrebe.

Među bezazlene zloupotrebe možemo npr. ubrojiti statistički "račun" nekog Weirusa, liječnika koji je živio koncem 16. stoljeća, i koji je izračunao da na svijetu ima točno 7 405 926 demona.

(N a p o m e n a: Podaci iz literature se ne slažu: neki autori nazivaju tog liječnika imenom Weyer, i navodno je tvrdio da postoji 7 409 127 vještica. No usputno valja spomenuti da je on pripadao rijetkima koji se javno usprotivio "pretjerivanjima" i zloupotrebama u proganjanju vještica).

Među manje bezazlene pripada npr. jedan svojedobni propagandni oglas automobilske tvornice "Volvo" (među propagandnim oglasima često nalazimo zloupotrebu statističkih podataka!). Godine 1968. jedan oglas te firme objavio je ovo: Prosječan Amerikanac vozi 50 godina, i svake treće godine nabavlja novi auto; on je, dakle, u toku svoga vozačkog staža kupio 15,1 auto. No s "volvom" trebao bi ih samo 4,5, jer u Švedskoj "volvo" traje u prosjeku 11 godina.

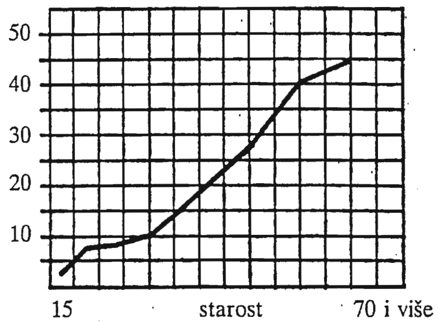
U spomenutom oglasu zaključak o tome da bi Amerikanac trebao samo 4,5 "volva" je, dakako, nedopušten, jer a) ništa se ne zna koliko kilometara vozi prosječan Amerikanac, a koliko prosječan Švеđanin, b) ne zna se ***zašto*** Amerikanac mijenja svake treće godine auto: da li zato što mu je auto dotrajao ili zato što je zaželio imati novi model (pa bi u tom slučaju mijenjao i "volvo"!), c) ne zna se kako vozi prosječan Amerikanac i prosječan Šveđanin, tj. da li vozi tako da troši motor i gume, itd.

Među opasne zloupotrebe pripadaju oni pogrešni zaključci koji mogu imati za posljedicu promjene čovjekova ponašanja, povećanje troškova ili bilo koju drugu djelatnost važnu za čovjeka. Među do sada navedenim primjerima u ovoj knjizi čitalac će sam prepoznati one slučajeve kada pogrešan zaključak može biti opasan. Jedan od njih spomenuli smo i u poglavlju "Teškoće pri radu s postocima", gdje smo pokazali kako se za istu situaciju može "dokazati" da cijene rastu ili pak da cijene padaju.

Evo još jednog primjera. Neka tvornica ima proizvodnju visine 60 tona, i u petogodišnjem planu ona predviđi povećanje proizvodnje na 100 tona. Međutim, u 5 godina tvornica postigne samo proizvodnju od 65 tona. S koliko je posto plan ostvaren? Ako se rezonira ovako: cilj nam je bio 100 tona, a ostvarili smo 65 tona, znači da smo plan ostvarili s 65% — takav je zaključak naravno pogrešan. Po toj bi logici i proizvodnja od 60 tona (što znači da je *plan* ostvaren s 0%) bila "ostvarenje plana" za 60%. Jedini ispravan zaključak bi morao biti ovaj: od predviđenog povećanja za 40 tona ostvarili smo 5 tona, dakle ostvarenje plana iznosi $5/40 = 12,5\%$.

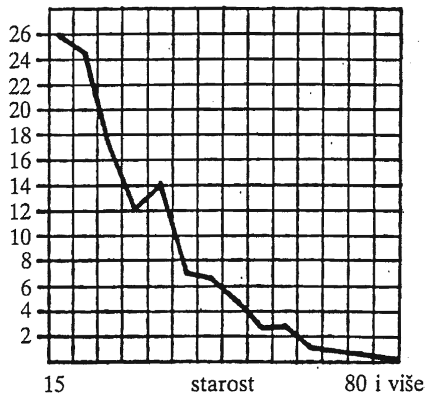
Među štetne zaključke pripada zasigurno i maloprije spomenuta tvrdnja da je vožnja po mokroj cesti "sigurnija od vožnje po suhom", a mogli bismo naravno nabrojati još niz tvrdnji, koje su i te kako značajne za društveni život, a koje nisu dokazane ili su nastale kao pogrešni zaključci iz nekih istraživanja (npr. "žene su slabiji vozači od muškaraca", "saharin izaziva rak kod čovjeka", "85% žena je frigidno", "tko dugo kašlje, dugo živi", itd.).

Nehotične ili namjerne zloupotrebe i pogrešne interpretacije lako se dogode kod različitih načina *grafičkog prikazivanja* rezultata. U poglavlju koje je bilo posvećeno toj temi, već smo na neke opasnosti upozorili. A sada razmotrimo još jedan zanimljiv primjer koji će pokazati kako način grafičkog prikazivanja iste pojave može navoditi na različite zaključke (koji su čak međusobno sasvim suprotni). Radi se o prikazivanju broja samoubojstava kod ljudi različite starosne dobi. Ako godišnji broj samoubojstava na 100 000 stanovnika različite dobi prikažemo grafički, dobit ćemo sliku koja će izgledati po prilici kao crtež prikazan na slici 18.1.



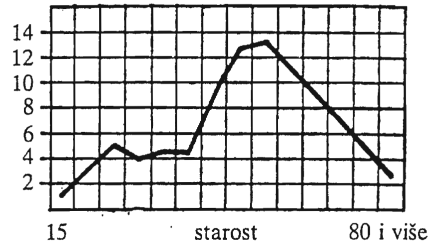
Slika 18.1. Broj samoubojstava na 100 000 stanovnika različite dobi

Naprotiv, ako ta ista samoubojstva prikažemo na svakih 100 smrti kod ljudi različite starosne dobi, dobivamo situaciju prikazanu na slici 18.2.



Slika 18.2. Broj samoubojstava na 100 smrti pojedinih dobnih skupina

I, konačno, ako uzmemo sama samoubojstva kao osnovicu, pa prikažemo koliko od 100 samoubojstava otpada na pojedine godine starosti samoubojica, dobit ćemo situaciju prikazanu na slici 18.3.



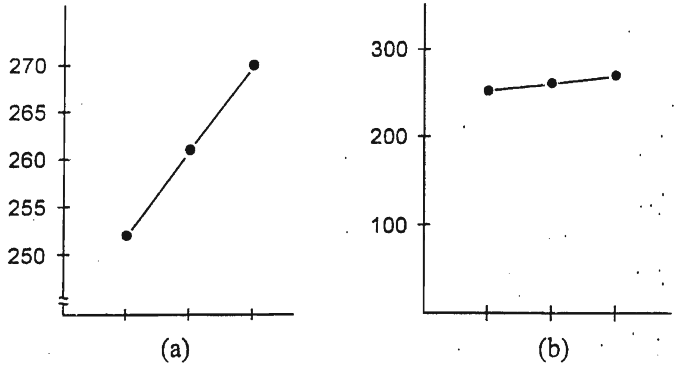
Slika 18.3. Od ukupno 100 samoubojstava, na pojedine godine starosti samoubojica otpadaju ovi brojevi

Što je prema tome istina?

Svaki od ovih prikaza je istinit samo ga treba ispravno interpretirati. Prvi prikaz najtočnije prikazuje da na 100 000 stanovnika određene dobi broj samoubojstava godinama sve više raste: dok kod mladih ljudi u životnoj dobi od oko 15 godina imamo na njih 100 000 nešto ispod 5 samoubojstava godišnje, kod sedamdeset i osamdesetgodišnjaka imamo već oko 45 samoubojstava na 100 000 ljudi tih godina. Drugi crtež, iako je točan, *zavodi* na pogrešan zaključak kao da tobože među mladima ima najviše samoubojica. No u tom su crtežu samoubojstva prikazana na 100 *smrtnih slučajeva*. Od 100 smrtnih slučajeva mladih ljudi naravno da ima mnogo samoubojstava, jer oni u toj životnoj dobi još uglavnom ne umiru od drugih bolesti, od kojih boluju stariji, te su uzroci smrti kod njih uglavnom nesreće i samoubojstva. Naprotiv, od 100 smrti kod *starih* ljudi, manje od jedne smrti otpada na samoubojstvo, a sve ostalo su smrti zbog različitih drugih oboljenja od kojih stari ljudi najčešće umiru.

Treći crtež navodi na misao da ljudi srednjih godina imaju najviše samoubojstava, što naravno također nije točno, jer slika prikazuje starosnu dob samoubojica na svakih 100 *samoubojstava*. Drugim riječima, ta slika pokazuje da među samoubojicama ima najviše ljudi srednjih-starijih godina, ali iz toga ne možemo izvoditi nikakve daljnje zaključke tako dugo dok ne znamo strukturu stanovništva po dobi i broj ljudi različite starosti u populaciji neke zemlje. Iz iskustva znamo da je mladih ljudi obično najviše, a starih najmanje. Budući da na 100 000 stanovnika svake dobne skupine broj samoubojstava *raste* (vidi sliku 18.1), među samoubojicama ipak neće biti mnogo mladih, jer oni imaju i najmanji indeks samoubojstava, ali neće biti ni mnogo starih (makar kod njih ima najviše samoubojstava u usporedbi s drugim dobnim skupinama), jer njih ima *malo* u ukupnom broju stanovnika!

Dakle, iako su sva tri crteža zapravo ispravna, jedino crtež na slici 18.1. prikazuje ono što nas u većini slučajeva i najviše zanima, tj. razlikuju li se dobne skupine po broju samoubojstava. Odgovor glasi: razlikuju se, najviše samoubojstava nalazimo u starih ljudi.



Slika 18.4. Promjena nekih rezultata koji iznose 252, 261, 270 izgleda impresivno na lijevoj slici (a), ali se zapravo radi o maloj promjeni koja je ispravno prikazana desnom slikom (b).

Već je spomenuto, da se u području ekonomske propagande često nailazi na *zloupotrebu* statistike. Karakterističan primjer možemo naći u *grafičkim prikazima* pojedinih promjena. Ako se npr. prizvedena količina nekog artikla mijenja od 252, preko 261 na 270, onda se to recimo može prikazati lijevim ili desnim crtežem na slici 18.4. Na lijevom crtežu ta promjena izgleda vrlo velika, ali samo zato, jer je prikazan samo mali. ali jako povećan, dio ordinate. Ispravno bi tu promjenu trebalo prikazati desnim cetežem, gdje ordinata započinje od nule, a iz toga se ceteža vidi da porast baš i nije tako senzacionalan. Uvijek kada se pokazuje samo jedan dio ordinate to valja jasno označiti na crtežu: ordinata mora biti vidljivo prekinuta.

— — —

Možda se poneki čitalac i obeshrabrio čitajući o različitim "zamkama" i "smicalicama" u koje može upasti pri interpretaciji statistički dobivenih podataka. Ako je tako, to je samo *pohvalno* za njega, jer znači da je zaista osjetio s kakvom moralnom i znanstvenom odgovornošću treba donositi zaključke iz dobivenih rezultata.

A zaključci su naravno upravo ono što je *bitno*. Ako u nekom eksperimentu nađemo da je razlika između aritmetičkih sredina, dobivenih pomoću dva različita eksperimentalna postupka, statistički značajna, to je vrlo *zanimljivo*; ali *bitno* je kako ćemo tu razliku *interpretirati!*

Statistika nam pomaže samo u tome da ustanovimo da li razlika zaista *postoji* (tj. nije li dobivena razlika možda rezultat slučaja), ali nam više ne pomaže u *interpretaciji* rezultata. Nije stoga čudno što je jedan londonski profesor statistike svoja statistička predavanja obično započinjao ovom usporedbom:

"Statistika je kao kupaći kostim bikini: ono što *pokazuje*, to je vrlo *zanimljivo*, ali ono što nam *ne* pokazuje, to je — *bitno!*"

19.

SKALE MJERENJA

Smatramo da je na ovom mjestu potrebno dati neka pravila o upotrebi pojedinih statističkih računa navedenih do sada u ovoj knjizi. U ovom se području vrlo često griješi, tj. u pojedinim se slučajevima upotrebljavaju statistički postupci, koje — s obzirom na mjernu skalu — nemamo prava upotrijebiti.

Najtipičniji slučaj takve pogreške je 1. izračunavanje aritmetičke sredine iz rangova (npr. "prosječne ocjene" u školskom ocjenjivanju), i 2. izračunavanje koeficijenta varijabilnosti V u skalama koje nemaju apsolutnu nulu.

Pri različitim mjerenjima služimo se različitim skalama od kojih su najpoznatije nominalne, ordinalne, intervalne i tzv. "omjerne skale".

1. *Nominalne skale.* Kod nominalnih skala označujemo istu stvar istim brojem, tj. umjesto imena predmeta navodimo njegov *broj* (npr. brojevi igrača u športu, brojevi automobila itd.). Nominalne skale ustvari i nisu nikakve skale, jer nam brojevi služe samo za *identifikaciju.* Od statističkih postupaka navedenih do sada, uz nominalne skale smijemo upotrebljavati samo dominantnu vrijednost, račun proporcija, χ^2-test, ϕ, *Cramerov Fi*, i koeficijent kontingencije C.

2. *Ordinalne skale* služe za označivanje *redoslijeda.* Tipičan primjer ordinalne skale su brojevi kuća u ulici, rangovi, školske ocjene itd. Karakteristika je tih skala da određuju samo je li nešto veće ili manje od drugoga, ali *razlike* između pojedinih jedinica skale nisu jednake.

Od do sada opisanih statističkih postupaka smijemo pri obradi ordinalnih skala, uz postupke navedene pod 1, upotrebljavati: centralnu vrijednost, koeficijent korelacije ρ, *Tau*, *Teta* i koeficijent W.

3. *Intervalne skale.* Kod njih znamo ne samo *redoslijed*, već i *razliku* među brojevima na skali, jer je u tim skalama neka definirana razlika jednaka na svakom dijelu skale. Na primjer, razlika od 1°C je uvijek jednaka, bez obzira radi li se o razlici između 0° i 1°, ili o razlici između 153° i 154°.

Međutim, te skale nemaju *apsolutnu nulu*, i zato ne možemo reći da je npr. temperatura od 100° F dva puta veća od temperature od 50°F, jer ako te vrijednosti pretvorimo u stupnjeve Celzijusa, dobivamo 38°C i 10°C!

Osim svih računa navedenih pod brojevima 1. i 2, kod intervalnih skala možemo računati i ove parametre: aritmetičku sredinu, standardnu devijaciju, z-vrijednosti i koeficijente korelacije r (uključujući i parcijalnu i multiplu korelaciju).

4. *"Omjerne" skale* ("Ratio-skale") imaju sva svojstva intervalnih skala, a k tome još i svojstvo da jednaki brojčani odnosi (omjeri) znače i jednake odnose u mjerenoj pojavi. To je moguće zato što te skale imaju *apsolutnu nulu*. To su, primjerice: dužina, težina, otpor, itd. Zato možemo npr. utvrditi da je težina od 90 kg *tri puta* veća od težine od 30 kg, jer je to točno u svakom sustavu mjerenja, tj. u bilo kojim težinskim jedinicama.

Osim računa navedenih pod 1, 2. i 3, kod "omjernih" skala smijemo upotrijebiti i geometrijsku sredinu i koeficijent varijabilnosti.

N a p o m e n a. Koeficijent varijabilnosti V može se — uz oprez — upotrijebiti i kod *intervalnih* skala, ali samo ako uspoređujemo varijabilnost u *istoj* varijabli: na primjer, zanima nas variraju li u nekom testu inteligencije više muški ili ženski ispitanici. Pri tome bi, naravno, aritmetičke sredine obiju grupa morale biti slične — a u tom slučaju nije nam ni potreban koeficijent V, nego možemo *izravno* usporediti standardne devijacije.

No, ako se radi o dva *različita* testa, koeficijent varijabilnosti ne smijemo koristiti. Na primjer, ne bi imalo smisla ispitivati variraju li neki ispitanici više u jednom testu inteligencije ili jednom testu psihomotorike.

N a p o m e n a. U ovoj knjizi ipak je naveden jedan koeficijent varijabilnosti, koji nije izračunan na omjernoj, nego na intervalnoj skali, a to je varijabilitet čovjekove tjelesne temperature, kojim smo željeli pokazati kako čovjek mnogo više varira npr. u težini i visini nego u temperaturi. Iako taj primjer ne udovoljava upravo spomenutom pravilu, on je ipak ostavljen, jer čovjekova temperatura *zaista vrlo malo varira,* bez obzira da li V računamo sa Celsiusima, Fahrenheitima, Reaumirima ili Kelvinima. (U svim tim slučajevima dobit ćemo doduše različit, ali uvijek veoma malen V.)

20. UVOD U ANALIZU VARIJANCE

20.1. KAKO TESTIRATI RAZLIKU IZMEĐU VIŠE ARITMETIČKIH SREDINA?

Analiza varijance upotrebljava se onda kada se želi utvrditi postoje li razlike između *nekoliko* aritmetičkih sredina. Na primjer, ako imamo 4 skupine ispitanika, i svaka skupina primi drugu dozu alkohola, tražimo da li i do kakvih će promjena doći u nekim psihomotoričkim vještinama. Ako dobijemo različite aritmetičke sredine kod svake skupine, zanima nas pripadaju li te sredine istoj populaciji ili ne. Drugim riječima, zanima nas je li razlika među njima statistički značajna. Mi bismo doduše mogli izračunavati t-test za svaki par aritmetičkih sredina, ali ima nekoliko razloga zašto to nije preporučljivo, a u nekim slučajevima (kad je broj aritmetičkih sredina velik) čak ni dopušteno:

1. Povećavanjem broja t-odnosa zapravo povećavamo razinu značajnosti.

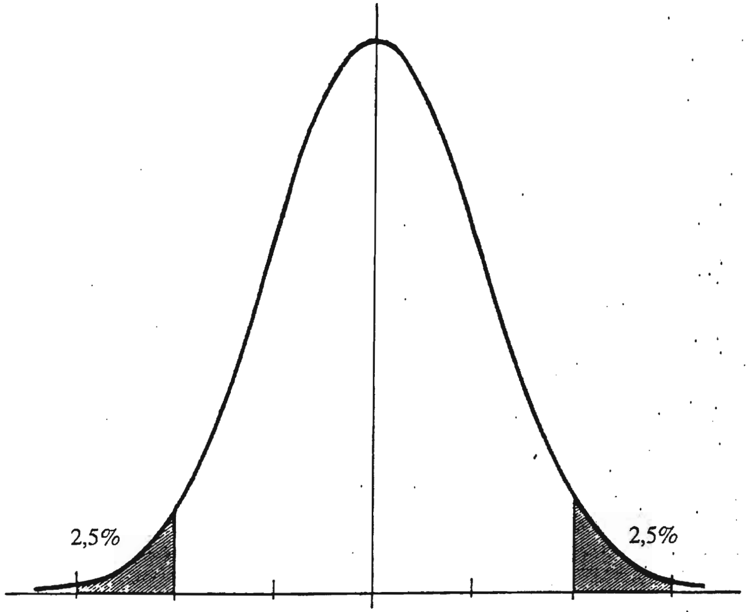
Kad testiramo značajnost razlike između *dvije* aritmetičke sredine, onda rezoniramo ovako (vidi sliku 20.1): razlika koja padne izvan granica $\pm 1,96$ standardne pogreške razlike, smatra se da je statistički značajna (na razini od 5%), tj. ako zapravo (u populaciji) *nema* razlike, *šansa* da ćemo u našem mjerenju dobiti tako veliku razliku je neznatna ($P = 5\%$).

Uzmemo li, međutim, *velik* broj mjerenja na populacijama među kojima nema razlike, nužno ćemo dobiti čitavu *distribuciju razlika* kojima je doduše aritmetička sredina nula, ali među kojima će 5% razlika biti "statistički značajne". Ali u ovom slučaju očito je da one nisu značajne, jer ih po pukom slučaju na velikom broju mjerenja *moramo* očekivati u 5% slučajeva.

Drugim riječima, povećanjem broja t-omjera u nekom eksperimentu, *povećava se vjerojatnost* pojavljivanja *slučajno* značajnih t-omjera. Dakle, što je broj aritmetičkih sredina veći, to je i razina značajnosti više povećana, ako računamo pojedinačne t-omjere.

2. Zakon t-testa vrijedi za *slučajne* uzorke. A ako iz niza aritmetičkih sredina *izaberemo* upravo najmanju i najveću, pa njihovu razliku testiramo t-testom, onda te dvije aritmetičke sredine *nisu* izabrane po zakonu slučaja.

3. Kad prigovor (1) i ne bi vrijedio, povećanjem t-omjera znatno bi se povećao i posao izračunavanja.

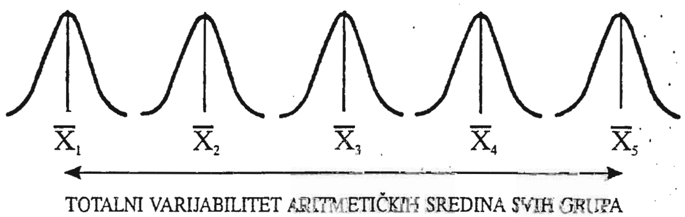


Slika 20.1. Testiranje značajnosti razlike između dvije aritmetičke sredine

4. Uzimajući u obzir samo po dvije aritmetičke sredine, gubimo na preciznosti izračunavanja varijance, koja je uvjetovana varijabilitetom *svih* grupa, a ne samo varijabilitetom dviju grupa koje upravo računamo.

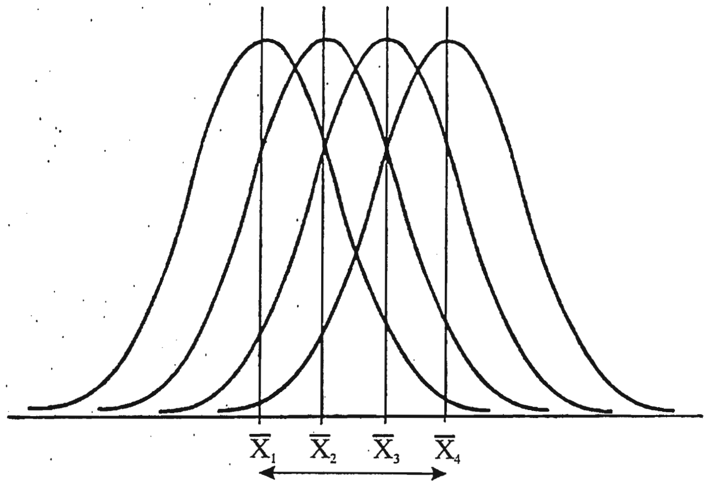
Osnovna misao koju sadrži analiza varijance je ova: treba dokazati je li varijabilitet *među* grupama veći od varijabiliteta *unutar* grupa. Ako je on statistički značajno veći, onda su to zaista grupe koje ne pripadaju istoj populaciji ili različitim populacijama, ali s jednakom aritmetičkom sredinom.

Pojednostavljeno, problem bi se mogao prikazati kao što je prikazan na slici 20.2.



Slika 20.2. Varijabilitet među grupama znatno je veći od varijabiliteta unutar grupa

Ovdje vidimo da su varijabiliteti pojedinih grupa relativno mali prema varijabilitetu aritmetičkih sredina tih grupa (dakle prema varijabilitetu među grupama). Prema tome, možemo smatrati da su to zaista grupe koje ne pripadaju istoj populaciji.



Slika 20.3. Varijabilitet među grupama manji je od varijabiliteta unutar grupa

Na slici 20.3, naprotiv, vidimo da svaka grupa posebno varira čak i više nego što variraju aritmetičke sredine tih grupa. Prema tome, možemo pretpostaviti da sve te grupe pripadaju istoj populaciji.

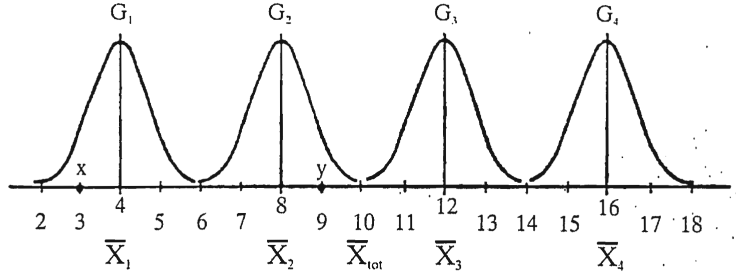
Već je iz ovoga do sada vidljivo da se analiza varijance zapravo sastoji u tome da se varijabilitet svih dobivenih rezultata *rastavi* na dijelove od kojih je sastavljen, tj. na interni varijabilitet *unutar* svake pojedine grupe rezultata i na varijabilitet *između* pojedinih grupa. Iz odnosa tih dvaju varijabiliteta može se zaključiti radi li se o grupama koje su među sobom različite (dakle ne pripadaju istoj populaciji), ili su njihove razlike možda samo slučajne, pa sve (grupe) potječu zapravo iz iste "matične populacije".

Pogledamo li položaj nekog rezultata u masi drugih rezultata i drugih grupa, možemo ustanoviti da se njegovo *odstupanje* od zajedničke aritmetičke sredine (dakle, njegov varijabilitet) može podijeliti na dvije komponente:

a) na odstupanje tog rezultata od vlastite aritmetičke sredine, tj. od aritmetičke sredine grupe kojoj taj rezultat pripada (varijabilitet unutar grupe) i

b) na odstupanje aritmetičke sredine kojoj pripada taj rezultat od zajedničke ("totalne") aritmetičke sredine (varijabilitet među grupama).

Evo primjera prikazanog na slici 20.4.



Slika 20.4. Udaljenost rezultata X i Y od aritmetičke sredine svih rezultata ($\overline{X}_{tot}$) može se odrediti: a) izračunavanjem razlike između tog rezultata i $\overline{X}_{tot}$ ili b) zbrajanjem odstupanja tog rezultata od njemu pripadne aritmetičke sredine, s odstupanjem te aritmetičke sredine od $\overline{X}_{tot}$

Na primjer, rezultat X grupe G_1 udaljen je za 7 jedinica od zajedničke aritmetičke sredine ($X = 3 - 10 = -7$). Taj isti rezultat dobivamo ako zbrojimo odstupanje rezultata X od $\overline{X}_1$ s odstupanjem $\overline{X}_1$ od $\overline{X}_{tot}$:

$(3-4)+(4-10) = (-1)+(-6) = -7.$

Analogno dobivamo za rezultat Y :

$$Y = 9 - 10 = -1$$

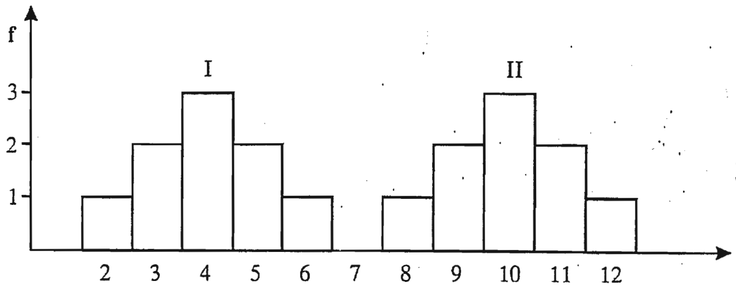
$$(9-8)+(8-10) = 1+(-2) = -1.$$

Ako odstupanja pojedinih rezultata od tot. arit. sredine kvadriramo (kvadrirane devijacije) i te kvadrate sumiramo (to su poznate "sume kvadrata" u analizi varijance), ustanovit ćemo da se ta suma, nazovimo je "totalna suma kvadrata", dade također rastaviti na dvije "pod-sume" kvadrata:

a) na sumu kvadrata unutar grupe i

b) na sumu kvadrata između grupa.

Smisao tih naziva najbolje ćemo uočiti iz primjera koji slijede.



Slika 20.5. Totalna suma kvadrata dade se rastaviti na a) sumu kvadrata unutar grupe i b) sumu kvadrata između grupa. — Detaljnije tumačenje vidi u tekstu!

Na slici 20.5. dan je primjer koji je zbog jednostavnosti izveden samo na dvije grupe podataka (grupe su crtane histogramom da se lakše vidi frekvencija):

Ako izračunamo devijacije svih rezultata od $\overline{X}_{tot}$, ako te devijacije kvadriramo, i ovako kvadrirane ih sumiramo, dobit ćemo:

$$\text{Grupa I} \begin{cases} 1\cdot 5^2 = 25 \\ 2\cdot 4^2 = 32 \\ 3\cdot 3^2 = 27 \\ 2\cdot 2^2 = 8 \\ 1\cdot 1^2 = 1 \end{cases} \qquad \text{Grupa II} \begin{cases} 1\cdot 1^2 = 1 \\ 2\cdot 2^2 = 8 \\ 3\cdot 3^2 = 27 \\ 2\cdot 4^2 = 32 \\ 1\cdot 5^2 = 25 \end{cases}$$

$$\Sigma = 186 = \text{totalna suma } (SS_{tot})$$

Izračunamo li sada sumu kvadrata unutar grupe I i II, dobivamo:

Grupa I; $\overline{X}_1 = 4$	Grupa II; $\overline{X}_2 = 10$
$1\cdot 2^2 = 4$	$1\cdot 2^2 = 4$
$2\cdot 1^2 = 2$	$2\cdot 1^2 = 2$
$3\cdot 0^2 = 0$	$3\cdot 0^2 = 0$
$2\cdot 1^2 = 2$	$2\cdot 1^2 = 2$
$1\cdot 2^2 = 4$	$1\cdot 2^2 = 4$
Suma =12	Suma =12

Prema tome, suma kvadrata unutar grupe (SS_{wg}) = 12 + 12 = 24. Konačno još treba izračunati ponderiranu sumu kvadrata *između* grupa, tj. ponderiranu sumu kvadriranih razlika između svake aritmetičke sredine i $\overline{X}_{tot}$. Drugim riječima, pretpostavit ćemo da svaki od 9 rezultata u grupi I iznosi koliko i $\overline{X}_1$, tj. 4, a isto tako i svaki od 9 rezultata u grupi II kao da iznosi 10.

Dakle, suma kvadrata između grupa (SS_{bg}) iznosit će:

I grupa		II grupa
$9\cdot(4-7)^2 = 81$	plus	$9\cdot(10-7)^2 = 81$

Prema tome, $SS_{bg} = 81 + 81 = 162$.

Ako sada zbrojimo SS_{wg} i SS_{bg}, dobivamo $24 + 162 = 186$, a to je, kako vidimo, broj koji predstavlja *totalnu sumu kvadrata* (SS_{tot}).

Da reziminiramo:

Totalna suma kvadrata izračunava se prema formuli:

$$SS_{tot} = \Sigma(X - \overline{X}_{tot})^2. \tag{20.1}$$

Suma kvadrata unutar grupa računa se:

$$SS_{wg} = \Sigma(X_g - \overline{X}_g)^2. \tag{20.2}$$

Suma kvadrata između grupa računa se prema formuli:

$$SS_{bg} = \Sigma[N_g(\overline{X}_g - \overline{X}_{tot})^2]. \tag{20.3}$$

Pokušajmo sada na jednom primjeru izračunati sve spomenute vrijednosti. Uzmimo da smo kod 4 skupine mjerenja (svaka od po 5 mjerenja) dobili ove rezultate:

	I	II	III	IV
	2	3	6	5
	3	4	8	5
	1	3	7	5
	3	5	4	3
	1	0	10	2
Suma	10	15	35	20
$\overline{X}$	2	3	7	4

Totalna suma = 10 + 15 + 35 + 20 = 80.
Totalni N =5 · 4 = 20.

$$\overline{X}_{\text{tot}} = \frac{80}{20} = 4.$$

a) Želimo li izračunati totalnu sumu kvadrata (SS_{tot}), treba kvadrirati razliku između svakog rezultata (X) i $\overline{X}_{\text{tot}}$.

	I	II	III	IV
	4	1	4	1
	1	0	16	1
	9	1	9	1
	1	1	0	1
	9	16	36	4
Suma	24	19	65	8

Prema tome, $SS_{\text{tot}} = 24 + 19 + 65 + 8 = 116$.
b) Sumu kvadrata unutar grupe (SS_{wg}) izračunat ćemo ovako:

	I	II	III	IV
	$\overline{X} = 2$	$\overline{X} = 3$	$\overline{X} = 7$	$\overline{X} = 4$
	0	0	1	1
	1	1	1	1
	1	0	0	1
	1	4	9	1
	1	9	9	4
Suma	4	14	20	8

Prema tome, SS_{wg} iznosi:

$$\Sigma(X_{\text{g}} - \overline{X}_{\text{g}})^2 = 4 + 14 + 20 + 8 = 46.$$

c) I konačno ćemo izračunati sumu kvadrata između grupa (SS_{bg}) :

I	II	III	IV
$5 \cdot (2-4)^2$	$5 \cdot (3-4)^2$	$5 \cdot (7-4)^2$	$5 \cdot (4-4)^2$

$$SS_{\text{bg}} = 20 + 5 + 45 + 0 = 70.$$

Kako se vidi, totalna suma kvadrata ($SS_{\text{tot}} = 116$) sastoji se od zbroja sume kvadrata unutar grupe ($SS_{\text{wg}} = 46$) i sume kvadrata između grupa ($SS_{\text{bg}} = 70$).

Sada, kada znamo princip ovakvog računanja, možemo, naravno, odustati od ovog "drugog postupka" i poslužiti se "skraćenim metodama" za izračunavanje analize varijance, jednako kao što to činimo i pri izračunavanju aritmetičke sredine i standardne devijacije. Formule za skraćeno izračunavanje nama potrebnih vrijednosti glase ovako:

$$SS_{\text{tot}} = (\Sigma X^2) - (N_{\text{tot}} \cdot \overline{X}^2_{\text{tot}}), \tag{20.4}$$

$$SS_{\text{bg}} = [\Sigma(N_{\text{g}}\overline{X}^2_{\text{g}})] - (N_{\text{tot}}\overline{X}^2_{\text{tot}}), \tag{20.5}$$

$$SS_{\text{wg}} = SS_{\text{tot}} - SS_{\text{bg}}. \tag{20.6}$$

Primijenimo li te formule na naš primjer, dobivamo:

$$SS_{\text{tot}} = 436 - 320 = 116$$

$$SS_{\text{bg}} = 390 - 320 = 70$$

$$SS_{\text{wg}} = 116 - 70 = 46.$$

Međutim, time što smo dobili *sumu* kvadrata unutar grupa i između grupa, još ne možemo dobro ocijeniti je li veći *varijabilitet* unutar ili između grupa, jer veličina sume kvadrata, naravno, ovisi o *broju rezultata* (npr. suma kvadrata od 100 slučajnih odnosno neselekcioniranih rezultata bit će oko dva puta veća od sume kvadrata 50 rezultata koji pripadaju istoj populaciji). Zato je SS neprikladna vrijednost za ocjenu varijabiliteta, nego treba uzeti vrijednost koja se i inače uzima kao mjera varijabiliteta, a to je varijanca (s^2). Tu ćemo vrijednost dobiti tako *da svaku SS podijelimo s pripadajućim brojem stupnjeva slobode.* Iako se to, kako znamo, inače zove

$$varijanca \left(\text{tj. } s^2 = \frac{\Sigma(X - \overline{X})^2}{N - 1}\right),$$

u računu analize varijance obično se taj izraz naziva *srednjim kvadratom (MS)*.

Stupnjevi slobode za sumu kvadrata unutar grupa izračunavaju se tako da od ukupnog broja rezultata oduzmemo broj grupa, dakle $df_{\text{wg}} = N_{\text{tot}} - k$ (jer u svakoj grupi imamo $N - 1$ stupnjeva slobode).

Stupnjevi slobode za sumu kvadrata između grupa računaju se tako da se od broja grupa oduzme 1, dakle $df_{\text{bg}} = k - 1$.

U našem primjeru imamo, dakle:

$$df_{\text{wg}} = 20 - 4 = 16$$

$$df_{\text{bg}} = 4 - 1 = 3.$$

Prema tome, varijanca (ili srednji kvadrat) rezultata unutar grupa iznosi:

$$MS_{\text{wg}} = \frac{46}{16} = 2,875$$

a varijanca rezultata između grupa iznosi:

$$MS_{\text{bg}} = \frac{70}{3} = 23,3.$$

Tek sada jasno vidimo da je varijabilitet među grupama veći od varijabiliteta unutar grupa (23,3 je veće od 2,875), i sada se još samo radi o tome da ustanovimo je li varijabilitet među grupama *dovoljno* veći od varijabiliteta unutar grupa, pa da bismo mogli proglasiti da se grupe zaista statistički značajno razlikuju. Jednako kao što smo kod t-omjera iz t-tablice očitavali koliko najmanje puta mora razlika biti veća od svoje pogreške, da bi bila značajna na razini značajnosti od 5% ili 1%, *jednako tako i u ovom slučaju uz pomoć Snedecorovih F tablica možemo ustanoviti koliko najmanje puta mora varijabilitet među grupama biti veći od varijabiliteta unutar grupa* da bi razlika između oba varijabiliteta bila statistički značajna.

Prema tome,

$$F = \frac{MS_{\text{bg}}}{MS_{\text{wg}}}. \qquad (20.7)$$

U našem primjeru $F = 23,3/2,875 = 8,1$. Iz F-tablice (tablica L u Dodatku) ustanovljujemo da za 3 i 16 stupnjeva slobode granična vrijednost F na razini od 5% iznosi 3,24. Budući da je naš F veći, to znači da naše grupe ne pripadaju istoj populaciji ($P < 0,05$).

N a p o m e n a. F tablica se uvijek čita tako da se stupnjevi slobode brojnika čitaju na *gornjem* rubu tablice, a stupnjevi slobode nazivnika na njezinom *lijevom* rubu. To je potrebno naglasiti zato što u mnogim udžbenicima kod stupnjeva slobode na gornjem rubu piše "za veću varijancu", a kod stupnjeva slobode na lijevom rubu piše "za manju varijancu"; *većinom* je u brojniku veća varijanca pa je to zato tako navedeno; ali ako je slučajno u brojniku manja varijanca, *također treba stupnjeve slobode brojnika čitati na gornjem rubu.* No praktički to uopće nije potrebno činiti, jer ako je brojnik manji od nazivnika, onda je i F manji od 1, a u tom slučaju *nikako ne može biti značajan!*

Kompletan pregled računa analize varijance obično se postavlja u "tablicu analize varijance" (tablica 20.1):

TABLICA 20.1.
ZAVRŠNA TABLICA ANALIZE VARIJANCE

Izvor varijabiliteta	Suma kvadrata (*SS*)	Stupnjevi slobode (*df*)	Srednji kvadrat (varijanca) (*MS*)	F
Između grupa	70	3	70/3 =23,3	23,3/2,875 = 8,1
Unutar grupa	46	16	46/16 =2,9	
Total	116	19		

Kao što smo rastumačili postanak t-tablice i hi-kvadrat tablice, rastumačit ćemo i postanak F-tablice.

Zamislimo da u nekoj kutiji imamo žetone s rezultatima jedne vrlo velike populacije s normalnom raspodjelom i odlučimo da ćemo iz te populacije vaditi uzorke veličine $N_1 = 21$ i $N_2 = 25$. Prvo izvučemo 21 žeton i izračunamo varijancu tog uzorka (varijanca $= s^2$). Pretpostavimo da ona iznosi 35. Nakon toga izvučemo drugi uzorak veličine 25 i njegova varijanca iznosi, na primjer, 20. Podijelimo li prvu varijancu s drugom, dobivamo:

$$F = \frac{35}{20} = 1,75.$$

Nakon toga ponovimo postupak, i, recimo, u prvom uzorku ($N = 21$) dobijemo varijancu 10, a u drugom ($N = 25$) varijancu 30. Za taj uzorak F iznosi:

$$F = 10/30 = 0,33.$$

Kod trećeg para uzoraka, recimo, dobijemo

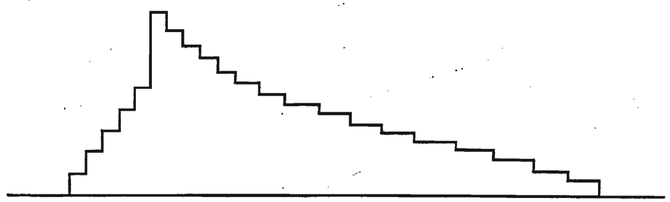
$$F = 4/20 = 0,20,$$

kod četvrtog

$$F = 25/15 = 1,67$$

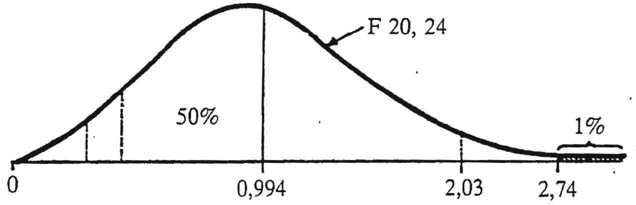
itd.

Ako vađenje parova uzoraka izvedemo petsto puta, i svaki put izračunamo F-omjer, dobit ćemo distribuciju F-omjera koja će izgledati po prilici kao distribucija na slici 20.6.



Slika 20.6. Približna distribucija 500 F-omjera iz iste populacije, sa stupnjevima slobode 20 i 24

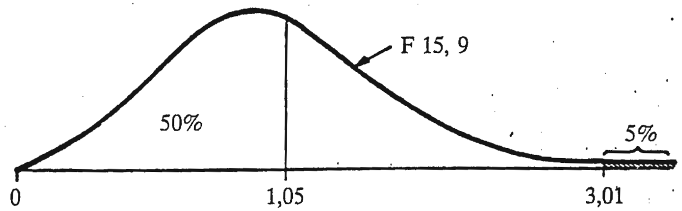
Za razliku od normalne raspodjele, ova raspodjela F-omjera, kako se vidi, nije simetrična, već je manje ili više asimetrična nadesno: ni jedan F-omjer ne može biti manji od nule, dok neki ekstremni mogu biti i vrlo visoki. Konačno, kod neizmjerno velikog broja pokusa, krivulja bi postala kontinuirana i pravilna, izgledala bi kao krivulja na slici 20.7.



Slika 20.7. F-distribucija sa stupnjevima slobode $N_1 = 20$ i $N_2 = 24$

To je F-distribucija sa stupnjevima slobode $N_1 - 1 = 20$ i $N_2 - 1 = 24$. Kao što se iz slike vidi, polovica očekivanih F-omjera bit će manja od 0,994, a polovica veća; 5% omjera bit će veće od 2,03, a to je upravo broj koji nalazimo u F-tablici uz stupnjeve slobode 20 i 24. (U tablici su samo brojevi iznad 1.)

Da smo isti pokus učinili s uzorcima veličine 16 i 10, dobili bismo krivulju prikazanu na slici 20.8.



Slika 20.8. F-distribucija sa stupnjevima slobode 15 i 9

Na toj krivulji 50% F-vrijednosti je ispod 1,05, a 50% iznad te vrijednosti. 5% vrijednosti je iznad 3,01, a ta se vrijednost nalazi i u tablici uz stupnjeve slobode 15 i 9.

Dakle, ako među populacijama *nema* razlike (a kod našeg je primjera sigurno nema, budući da smo uzorke vadili iz *iste* populacije!), F-omjeri mogu se kretati od nule, preko 1,5, pa sve do preko 3. Oni preko 3,01 *slučajno* se mogu pojaviti samo u 5% slučajeva. Po logici kojom smo se i do sada koristili, smatrat ćemo u našim mjerenjima, ako u analizi varijance dobijemo odnos F veći od 3,01 (uz ove stupnjeve slobode) da je razlika među grupama *statistički značajna* (jer bi se takav ili veći F-odnos — kada među grupama u populaciji ne bi bilo razlike — *slučajno* mogao dogoditi samo u 5% slučajeva).

Opisani postupak analize varijance pripada u najjednostavnije slučajeve analize, tj. u takve gdje samo jedna varijabla varira (npr. varira samo doza primljenog lijeka kod više eksperimentalnih skupina). To je tzv. "jednosmjerna klasifikacija". Nešto su kompliciraniji računski postupci kod "dvosmjerne" ili "višesmjerne" klasifikacije (npr. može istodobno varirati doza primljenog lijeka, dob ispitanika, spol ispitanika), a također kod zavisnih rezultata, tj. ako su *isti* ispitanici testirani u nekoliko eksperimentalnih situacija. Opisani postupak uglavnom je nastojao protumačiti *logiku* analize varijance, i iz te logike izvedene računske postupke.

Sada ćemo dati jedan "korak-po-korak" postupak za taj isti račun, ali taj će se postupak *formalno* razlikovati od do sada opisanoga: neki izrazi dobivat će se na drugi način, a također će se koristiti neki termini koji prije nisu korišteni. Nije potrebno dokazivati da se radi o *ekvivalentnom* računskom postupku (kao npr. kada izračunavamo standardnu devijaciju iz bruto rezultata, umjesto iz razlika između svakog rezultata i aritmetičke sredine), i da konačan rezultat mora biti identičan. A taj drugačiji postupak na ovom je mjestu usvojen zato što ćemo kasnije, u poglavljima 20.3, u kojima ćemo iznositi neke kompleksnije metode analize varijance, koristiti jednaku terminologiju koristeći se također "korak-po-korak" receptom.(Sve izložene "korak-po-korak" postupke dali su autori Cohen i Holliday u svojoj veoma korisnoj i praktičnoj knjizi "Statistics for education and physical education. — Harper i Row, 1979).

Evo tog postupka:

1. Izračunaj tzv. GRAND TOTAL (GT) po formuli

$$GT = \frac{\Sigma X^2}{N}$$

pri čemu X = suma svih rezultata u svim skupinama

2. Izračunaj TOTALNU SUMU KVADRATA (SS_{tot}) po fomuli:

$$SS_{tot} = \Sigma X^2 - GT$$

3. Izračunaj sumu kvadrata MEĐU GRUPAMA (SS_{bg}) po formuli:

$$SS_{bg} = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \dots \frac{(\Sigma X_n)^2}{N_n}$$

4. Izračunaj sumu kvadrata UNUTAR GRUPA (SS_{wg}) po formuli:

za prvu grupu: $SS_{wg} = \Sigma X_1^2 - \dfrac{(\Sigma X_1)^2}{N_1}$

za drugu grupu: $SS_{wg} = \Sigma X_2^2 - \dfrac{(\Sigma X_2)^2}{N_2}$

itd, i konačno za n-tu grupu:

$$SS_{wg} = \Sigma X_n^2 - \frac{(\Sigma X_n)^2}{N_n}$$

SS_{wg} = zbroj svih dobivenih vrijednosti za svaku grupu

5. Odredi STUPNJEVE SLOBODE za svaku sumu kvadrata, po formulama:

$$df \text{ za } SS_{bg} = (k - 1)$$

pri čemu k = broj grupa

$$df \text{ za } SS_{wg} = (N - k)$$

6. Odredi VARIJANCU, po formuli:

$$\text{Varijanca} = \frac{\text{suma kvadrata}}{df}$$

$$\text{Varijanca}_{bg} = \frac{SS_{bg}}{df_{bg}}$$

$$\text{Varijanca}_{wg} = \frac{SS_{wg}}{df_{wg}}$$

7. Izračunaj F po formuli:

$$F = \frac{\text{Varijanca}_{bg}}{\text{Varijanca}_{wg}}$$

8. Unesi rezultate u TABLICU ANALIZE VARIJANCE.

Ako taj "recept" primijenimo na naš primjer, koji smo postupno logičkim putem razrađivali, dobivamo ovo:

1. $GT = \frac{(\Sigma X)^2}{N} = \frac{6\,400}{20} = 320$
2. $SS_{tot} = \Sigma X^2 - GT = 436 - 320 = 116$
3. $SS_{bg} = \frac{\Sigma X^2}{N_1} + \frac{\Sigma X^2}{N_2} + \frac{\Sigma X^2}{N_3} + \frac{\Sigma X^2}{N_4} - 320 =$

$$= \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} + \frac{(\Sigma X_3)^2}{N_3} + \frac{(\Sigma X_4)^2}{N_4} = \frac{100}{5} + \frac{225}{5} + \frac{1\,225}{5} +$$

$$+ \frac{400}{5} - GT = 390 - 320 = 70$$

4. $SS_{wg} =$
   za 1. grupu: $24 - 20 = 4$
   za 2. grupu: $59 - 45 = 14$
   za 3. grupu: $265 - 245 = 20$
   za 4. grupu: $88 - 80 = 8$
   $SS_{wg} = 46$
5. df za $SS_{tot} = 20 - 1 = 119$
   df za $SS_{bg} = 4 - 1 = 3$
   df za $SS_{wg} = 20 - 4 = 16$
6. $\text{Varijanca}_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{70}{3} = 23,3$

$$\text{Varijanca}_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{46}{16} = 2,875$$

7. $F = \frac{\text{Varijanca}_{bg}}{\text{Varijanca}_{wg}} = \frac{23,3}{2,875} = 8,1$
   Kao što vidimo, rezultat je identičan.

20.2. ŠTO NAKON ZAVRŠENOG RAČUNA ANALIZE VARIJANCE?

Ako analiza varijance pokaže da možemo smatrati da svi uzorci potječu iz iste populacije (dakle ako naš F nije statistički značajan), onda nas naravno dalje više ne zanimaju pojedinačne razlike između nekih aritmetičkih sredina.

No ako F-test *odbaci* nul-hipotezu, tj. ako dokažemo da uzorci *ne pripadaju* istoj populaciji, često nas može zanimati *koji* se uzorci među sobom statistički značajno razlikuju. Tako, na primjer, ako smo u pokusu imali 3 eksperimentalne skupine (svaka skupina pod različitom "dozom" nezavisne varijable) i jednu kontrolnu skupinu, može nas zanimati *koje* se sve eksperimentalne grupe statistički značajno razlikuju od kontrolne grupe. (Jer F- test — ako je značajan — govori nam samo to da sve grupe ne pripadaju istoj populaciji, pa je, prema tome, jedino u što možemo biti sigurni samo to da se barem dvije grupe međusobno statistički značajno razlikuju: najvjerojatnije je da su to grupa s najmanjom i grupa s najvećom aritmetičkom sredinom.

Već na početku poglavlja o analizi varijance prezentirani su argumenti zašto ne dolazi u obzir upotreba niza t-testova; iz argumenata je vidljivo da ima mnogo razloga, no glavni je razlog taj što upotrebom većeg broja t-testova (kod 4 grupe već imamo $\frac{4 \cdot 3}{2} = 6$ t-testova, kod 10 grupa 45 t-testova) povećavamo šansu da ćemo *slučajno* dobiti t, koji je statistički značajan, pa prema tome povećavamo šansu da učinimo "pogrešku tipa I", tj. da proglasimo značajnom razliku koja zapravo među populacijama ne postoji. (Vidi izlaganje o pogrešci tipa I i tipa II na str. 135) Konkretno, ako bismo na primjer, imali 7 grupa (7 eksperimentalnih situacija), u tom slučaju postoji 21 mogući *par* aritmetičkih sredina, tj. 21 razlika između dvije aritmetičke sredine. U tom slučaju već postoji dosta visoka vjerojatnost da će — usprkos činjenici da možda i svi uzorci pripadaju istoj populaciji — barem 1 od 21 t-testova pokazati "statistički značajnu" razliku na nivou od 5 %, jer — kao što znamo — ako i *nema* razlike među populacijama, u prosjeku će se od 100 mjerenja u 5% slučajeva pojaviti "statistički značajna" razlika. A 5 od 100 je isto kao i 1 od 20 — dakle na 20 mjerenja 1 razlika "statistički značajna".

Svojedobno (1955) je statističar C. W. Dunnett tome pokušao doskočiti tako da je izradio tablice slične t-tablicama, u kojima se sistematski ***povećavala*** kritična (granična) t-vrijednost što je broj skupina, koje se uspoređuju, bio veći. Tako, na primjer, dok u klasičnoj t-tablici granična t-vrijednost na razini značajnosti od 0,05 za 60 stupnjeva slobode iznosi 2,00 u Dunnettovim tablicama za 4 grupe, kod kojih međusobno uspoređujemo sve aritmetičke sredine, granična t-vrijednost iznosi 2,55.

U novije vrijeme različiti autori (Tuckey, Duncan, Scheffe, Newman-Keuls i dr.) predložili su različite tehnike za sistematsko smanjivanje rizika pogreške tipa I kod povećanja broja usporedbi između dviju aritmetičkih sredina. Neki od tih autora naročito su oprezni te razlikuju postupke "a priori" od postupaka "a posteriori", tj. daju različite formule za izračunavanje statističke značajnosti između dvije po dvije aritmetičke sredine, već prema tome vršimo li usporedbu *prije* ili *nakon* računa analize varijance! (Jer nakon završenog računa, mi već *znamo* pripadaju li sve aritmetičke sredine istoj populaciji.)

Čini se, međutim, da situacija u tom pogledu u statistici nije još posve raščišćena, jer se u literaturi nailazi na različite savjete, koji, doduše, preporučuju neki od tih postupaka, ali — iako se navodi ista formula — daju se *različite* upute za izračunavanje stupnjeva slobode, što naravno dovodi do prilično različitoga definitivnog rezultata!

Stoga sam se u ovoj knjizi odlučio za jednog od najviše citiranih autora, tj. za Scheffea, te ću ukratko iznijeti jednu njegovu varijantu za izračunavanje razlika među aritmetičkim sredinama *nakon* završenog F-testa.

Postupak u biti nije kompliciran i može se sažeti u ove "korake":

1. Nakon izračunanog F-omjera u analizi varijance, za svaki par aritmetičkih sredina, koje želimo usporediti, primijeniti ovu formulu:

$$F = \frac{(\overline{X}_a - \overline{X}_b)^2}{MS_{wg}(N_a + N_b)/N_a N_b}. \qquad (20.8)$$

2. Iz F-tablice očitamo granični F uz željenu razinu značajnosti (uzmimo $P = 0,05$), za $(k-1)$ i $(N-k)$ stupnjeva slobode. (To smo već očitavali kod analize varijance.)

3. Očitana granična vrijednost F pomnoži se sa $(k-1)$, i tako se dobije *nova* granična vrijednost, nazovimo je F'.

4. Za sve parove aritmetičkih sredina, za koje to želimo, izračunamo izraz F prema formuli (20.8), i usporedimo ga sa F'. Ako je F veći od F', razliku možemo smatrati statistički značajnom (na razini na kojoj smo očitavali F u 2. "koraku").

Primijenimo li taj postupak na naš prvi primjer (vidi stranicu 304), dobivamo rezultate prikazane u tablici 20.3.

Granični F na razini 0,05 za analizu varijance za stupnjeve slobode 3 i 16 iznosi 3,24. Pomnožimo li tu vrijednost sa $(k-1)$, dobivamo 3,24 (4 − 1) = 9,72. Samo su *dvije* razlike između aritmetičkih sredina dakle statistički značajne (one su u tablici označene zvjezdicom), i to razlika između prve i treće, i razlika između druge i treće aritmetičke sredine.

N a p o m e n a. Scheffeova metoda dosta je rigorozna u pogledu pogreške tipa I, tj. rjeđe će nam se dogoditi da odbacimo nul-hipotezu, iako među populacijama *nema* razlike. Drugim riječima, metoda više tendira tome da *prihvatimo* nul-hipotezu, pa makar među populacijama razlika i postojala. Sam Scheffe stoga je predložio da se eventualno umjesto razine značajnosti 0,05 uzme blaža razina, recimo 0,10. (Za upotrebu te blaže razine treba koristiti detaljne F-tablice, koje se ne nalaze u ovoj knjizi.)

TABLICA 20.3.
TESTIRANJE RAZLIKA MEĐU POJEDINIM ARITMETIČKIM SREDINAMA NAKON ZAVRŠENE ANALIZE VARIJANCE

Razlika između		F
$\overline{X}_1$ i $\overline{X}_2$	$\longrightarrow \frac{(2-3)^2}{2,9(10)/25} = 1/1,16 =$	0,86
$\overline{X}_1$ i $\overline{X}_3$	$\longrightarrow \frac{(2-7)^2}{2,9(10)/25} = 25/1,16 =$	21,55*
$\overline{X}_1$ i $\overline{X}_4$	$\longrightarrow \frac{(2-4)^2}{2,9(10)/25} = 4/1,16 =$	3,45
$\overline{X}_2$ i $\overline{X}_3$	$\longrightarrow \frac{(3-7)^2}{2,9(10)/25} = 16/1,16 =$	13,79*
$\overline{X}_2$ i $\overline{X}_4$	$\longrightarrow \frac{(3-4)^2}{2,9(10)/25} = 1/1,16 =$	0,86
$\overline{X}_3$ i $\overline{X}_4$	$\longrightarrow \frac{(7-4)^2}{2,9(10)/25} = 9/1,16 =$	7,76

Treba spomenuti da se gotovo u svim udžbenicima nalazi zahtjev, koji smo već spominjali i kod t-testa, tj. da je potrebno da su populacije, iz kojih su uzorci uzimani, normalno distribuirane i da varijance tih populacija moraju biti približno jednake.

No, isto što smo rekli za t-test, vrijedi i ovdje: Boneauova su ispitivanja pokazala da ti uvjeti mogu biti prekršeni ako su uzorci jednake ili slične veličine, i ako su populacije međusobno slične u odstupanju od normalne raspodjele.

N a p o m e n a. Mnogi nešto upućeniji čitaoci čuli su vjerojatno za izraz "analiza kovarijance". Obično na tom području postoji potpuno mutna slika o tome što to znači, no radi se o jednoj u biti vrlo jednostavnoj stvari: u praktičnom terenskom radu često nije moguće *izjednačiti* aritmetičke sredine pojedinih eksperimentalnih podgrupa, pa prema tome već *započinjemo* istraživanje s aritmetičkim sredinama, koje se statistički značajno razlikuju! Klasična analiza varijance pokazuje nam da li se aritmetičke sredine različitih podgrupa statistički značajno razlikuju, ali pri tome polazimo od pretpostavke da — u slučaju da naša nezavisna varijabla nema utjecaja — nema razlika među skupinama. Ako te razlike već u početku postoje, onda se — da bi se ustanovilo je li nezavisna varijabla imala utjecaj i koliki je on u pojedinim slučajevima bio — koristi račun "analize kovarijance". Sam postupak opisan je u nekim većim i kompleksnijim statističkim udžbenicima.

20.3. DVOSMJERNA ANALIZA VARIJANCE

A. Nezavisni rezultati

Istraživači katkada u svojim istraživanjima imaju više od jedne nezavisne varijable. U primjeru jednosmjerne analize varijance koji smo obradili u prošlom poglavlju, 4 skupine ispitanika bile su mjerene svaka u drugim uvjetima, i zanimalo nas je da li se te skupine međusobno razlikuju. Ali kada bi se, recimo, jedan istraživač zanimao za problem kakva je povezanost između *količine pušenja* (prva nezavisna varijabla) i dubine disanja, ali ujedno ga zanima ima li *starost* u tom odnosu neku ulogu, onda starost predstavlja *drugu* nezavisnu varijablu, i sada analiza varijance postaje kompliciranija. Kao što smo obećali u prethodnom poglavlju, dat ćemo "korak-po-korak" metodu za obradu takvih rezultata.

Pretpostavimo da je taj istraživač imao ovakav nacrt istraživanja, izražavajući dubinu disanja u nekim jedinicama. (Uzet je mali broj ispitanika, a rezultati su mali brojevi, kako bi se lakše pratio računski postupak):

		Nezavisna varijabla A (faktor A, starost)		
		30g	40g	50g
Nezavisna varijabla B (faktor B, pušenje)	Jaki pušači	4 5 6	5 6 10	7 9 11
	Slabi pušači	3 3 6	2 4 6	3 5 4
	Nepušači	2 2 5	5 3 4	3 6 6

Postupak:

1. Kvadriraj svaki rezultat $(X)^2$ i zbroji sve rezultate (ΣX) i njihove kvadrate (ΣX^2) unesi u tablicu:

	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	Suma X za redove
	4	16	5	25	7	49	16
	5	25	6	36	9	81	20
	6	36	10	100	11	121	27
Suma	15	77	21	161	27	251	63
	3	9	2	4	3	99	8
	3	9	4	16	5	25	12
	6	36	6	36	4	16	16
Suma	12	54	12	56	12	50	36
	2	4	5	25	3	9	10
	2	4	3	9	6	36	11
	5	25	4	16	6	36	15
Suma	9	33	12	50	15	81	36
Suma X za stupce	36		45		54		

$$\Sigma X_{\text{tot}} = 15 + 21 + 27 + 12 + 12 + 12 + 9 + 12 + 15 = 135$$

$$\Sigma X_{\text{tot}}^2 = 77 + 161 + 251 + 54 + 56 + 50 + 33 + 50 + 81 = 813$$

2. Izračunaj GRAND TOTAL (GT)

$$GT = \frac{(\Sigma X)^2}{N_{\text{tot}}} = \frac{135^2}{27} = 675$$

$$N_{\text{tot}} = \text{ukupni broj rezultata}$$

3. Izračunaj TOTAL SUME KVADRATA (SS_{tot})

$$SS_{\text{tot}} = \Sigma X_{\text{tot}}^2 - GT = 813 - 675 = 138$$

4. Izračunaj sumu kvadrata MEĐU SKUPINAMA (SS_{bg})

$$SS_{\text{bg}} = \left[\frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_2} \ldots + \ldots \frac{(\Sigma X_n)^2}{N_n}\right] - GT$$

$$= \frac{15^2}{3} + \frac{21^2}{3} + \frac{27^2}{3} + \frac{12^2}{3} + \frac{12^2}{3} + \frac{12^2}{3} +$$

$$+ \frac{9^2}{3} + \frac{12^2}{3} + \frac{15^2}{3} - 675 =$$

$$+ \frac{2277}{3} - 675 = 759 - 675 = 84$$

5. Izračunaj sumu kvadrata FAKTORA A (SS_{fA})

$$SS_{\text{fA}} = \frac{(\text{Suma stupca 1})^2}{N \text{ u stupcu 1}} + \frac{(\text{Suma stupca 2})^2}{N \text{ u stupcu 2}} + \ldots$$

$$\frac{(\text{Suma stupca n})^2}{N \text{ u stupcu n}} - GT =$$

$$= \left(\frac{36^2}{9} + \frac{45^2}{9} + \frac{54^2}{9}\right) - 675 =$$

$$= 693 - 675 = 18$$

6. Izračunaj sumu kvadrata za FAKTOR B (SS_{fB})

$$SS_{\text{fB}} = \frac{(\text{Suma redova})^2}{\text{ukupni broj redova}} - GT$$

$$= \frac{(\text{suma } X \text{ jakih pušača})^2}{\text{broj jakih pušača}} + \frac{(\text{suma } X \text{ slabih pušača})^2}{\text{broj slabih pušača}} +$$

$$+ \frac{(\text{suma } X \text{ nepušača})^2}{\text{broj nepušača}} - GT =$$

$$= \left(\frac{63^2}{9} + \frac{36^2}{9} + \frac{36^2}{9}\right) - 675 = 729 - 675 = 54$$

7. Izračunaj sumu kvadrata za INTERAKCIJU između faktora A i Faktora B (SS_{ab})

$$SS_{\text{ab}} = SS_{\text{bw}} - SS_{\text{fA}} - SS_{\text{fB}} =$$

$$= 84 - 18 - 54 = 12$$

8. Izračunaj sumu kvadrata UNUTAR GRUPA (SS_{wg})

$$SS_{wg} = SS_{tot} - SS_{wg} = 138 - 84 = 54$$

9. Odredi stupnjeve slobode za svaku sumu kvadrata

$$df \text{ za } SS_{tot} = (N-1) = 27 - 1 = 26$$

$$df \text{ za } SS_{bg} = (g-1) = 9 - 1 = 8$$

$$g = \text{broj grupa}$$

$$df \text{ za } SS_{fA} = (s-1) = 3 - 1 = 2$$

$$s = \text{broj stupaca}$$

$$df \text{ za } SS_{fB} = (r-1) = 3 - 1 = 2$$

$$r = \text{broj redova}$$

$$df \text{ za } SS_{AB} = (r-1)(s-1) = (3-1)(3-1) = 4$$

$$df \text{ za } SS_{wg} = (N - rs) = (27) - (3 \cdot 3) = 27 - 9 = 18$$

10. Procijeni varijance

$$\text{Varijanca} = \frac{\text{suma kvadrata}}{\text{stup. slobode}}$$

$$\text{Varijanca Faktora A} = \frac{SS_{fA}}{df_{fA}} = 18/2 = 9$$

$$\text{Varijanca Faktora B} = \frac{SS_{fB}}{df_{fB}} = 50,67/2 = 25,34$$

$$\text{Varijanca } WG = \frac{ss_{wg}}{df_{wg}} = 54/18 = 3$$

$$\text{Varijanca interakcije A} \cdot \text{B} = \frac{ss_{AB}}{df_{AB}} = 15,33/4 = 3,83$$

Izračunaj F vrijednosti za glavne efekte tablice varijanci:

$$F_{\text{faktor A}} = \frac{\text{Varijanca}_{\text{fakt. A}}}{\text{Varijanca}_{wg}} = 9/3 = 3$$

$$F_{\text{faktor B}} = \frac{\text{Varijanca}_{\text{fakt. B}}}{\text{Varijanca}_{wg}} = 25,34/3 = 8,45$$

$$F_{\text{interakc.}} = \frac{\text{Varijanca}_{AB}}{\text{Varijanca}_{wg}} = 3,83/3 = 1,28$$

12. Unesi rezultate u tablicu:

Izvor varijabiliteta	Suma kvadrata	Stup. slob.	Varijance	F
(Među grupama)	(84)	(18)		
Faktor A (godine)	18	2	$\frac{18}{2} = 9$	9/3=3
Faktor B (pušenje)	54	2	$\frac{54}{2} = 27$	27/3=9
Interakcija ($A \cdot B$)	12	4	$\frac{12}{4} = 3$	3/3=1
Unutar grupa	54	18	$\frac{54}{18} = 3$	
TOTAL	138	26		

Kako vidimo, F među grupama *nije* statistički značajan ($P > 0,05$), F faktora B (pušenje) *jest* statistički značajan (dakle količina pušenja *ima* efekta na dubinu disanja), ($P < 0,05$), F interakcije nije značajan ($P > 0,05$), što znači da nema interakcije između starosti i pušenja u pogledu dubine disanja.

B. Zavisni rezultati

Za ovo ćemo uzeti primjer sa samo *dvije* skupine istih ispitanika koji su ispitani u dva navrata (prije i poslije djelovanja neke nezavisne varijable). Čitalac koji je upamtio neke postupke iz testiranja statističke značajnosti razlika među aritmetičkim sredinama zavisnih grupa, sjetit će se da se kod *dvije* skupine taj problem veoma elegantno rješava uz pomoć *"metode diferencije"*. No primjer koji je sada na redu ipak je obrađen postupkom analize varijance za zavisne rezultate, jednostavno zato što je cijeli postupak lakše pratiti na samo dvije skupine, a rad sa tri ili više skupina samo povećava količinu računanja, a ne mijenja tehniku rada.

Pretpostavimo da su šest ispitanika ispitani u jednoj karakteristici prije i poslije primjene nekog postupka i da su dobiveni ovi rezultati:

Ispitanik	Prije	Poslije
1	23	32
2	27	25
3	31	40
4	32	31
5	26	38
6	25	29

Postupak metodom "korak-po-korak" izgleda ovako:
Kvadriraj svaki rezultat:

	X	X^2	X	X^2	X_1 i X_2
	23	529	32	1 024	55
	27	729	25	625	52
	31	961	40	1 600	71
	32	1 024	31	961	63
	26	676	38	1 444	64
	25	625	29	841	54
Suma	164	4 544	195	6 495	

2. Zbroji sve bruto rezultate (ΣX) i izračunaj GRAND TOTAL (GT):

$$GT = \frac{(\Sigma X)^2}{N} = \frac{359^2}{12} = \frac{128881}{12} = 10740,08$$

3. Izračunaj TOTALNU SUMU KVADRATA (SS_{tot}):

$$SS_{tot} = \Sigma X^2 - GT = 11\,039 - 10\,740,08 = 298,92$$

4. Izračunaj sumu kvadrata MEĐU GRUPAMA (SS_{bg}):

$$SS_{bg} = \frac{(\Sigma X_1)^2 + (\Sigma X_2)^2}{N} - GT$$

$$= \frac{164^2 + 195^2}{6} - GT =$$

$$= 10\,820,17 - 10\,740,08 = 80,09$$

5. Izračunaj sumu kvadrata MEĐU ISPITANICIMA (SS_{wg}):

$$SS_{wg} = \frac{(X_1)^2}{N} + \frac{(X_2)^2}{N} + \ldots + \frac{(X_n)^2}{N} - GT$$

X_1 = totalni rezultati ispitanika 1 u svim situacijama
N = broj pokušaja

$$SS_{wg} = \frac{55^2}{2} + \frac{52^2}{2} + \frac{71^2}{2} + \frac{63^2}{2} + \frac{64^2}{2} + \frac{54^2}{2} - GT =$$
$$10\,875,5 - 10\,740,08 = 135,42$$

6.Izračunaj INTERAKCIJU (ili rezidual) sume kvadrata ($SS_{interakc}$):

$$SS_{interakc} = SS_{tot} - (SS_{wg} + SS_{bg}) = 298,92 - 215,51 = 83,41$$

7. Izračunaj STUPNJEVE SLOBODE za svaku sumu kvadrata:
df za $SS_{tot} = (N-1) = 12 - 1 = 11$
N = broj rezultata
df za $SS_{bg} = (s-1) = 2 - 1 = 1$
s = broj pokušaja (stupaca)
df za $SS_{wg} = (r-1) = 6 - 1 = 5$
r = broj redova ispitanika
df za $SS_{interakc} = (r-1)(s-1) = (6-1)(2-1) = 5$

8. Odredi VARIJANCE

$$\text{Varijanca} = \frac{\text{suma kvadrata}}{\text{stup. slob.}}$$

$$\text{Varijanca}_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{80,09}{1} = 80,09$$

$$\text{Varijanca}_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{135,42}{5} = 27,08$$

$$\text{Varijanca}_{interakc.} = \frac{SS_{interakc.}}{df_{interakc}} = \frac{83,41}{5} = 16,68$$

9. Izračunaj F vrijednosti između situacija i ispitanika:

$$F_{situac} = \frac{\text{Varijanca}_{bg}}{\text{Varijanca}_{interakc}} = \frac{80,09}{1} = 80,09$$

$$F_{ispitanici} = \frac{\text{Varijanca}_{ispitanici}}{\text{Varijanca}_{interakc}} = \frac{27,08}{16,68} = 1,62$$

10. Unesi rezultate u TABLICU ANALIZE VARIJANCE:

Izvor varijacije	Suma kvadrata	Stupnj. slobode	VARIJANCA (sred. kvadrat)	F
Među pokušajima	80,09	1	80,09/1= 80,09	4,80
Među ispitanicima	135,42	5	135,42/5 = 27,08	(1,62)
Interakcija	298,92	11		

Zaključak: nema statistički značajne razlike niti među eksperimentalnim situacijama (granični $F = 6,61$, a naš F je manji), niti među ispitanicima (granični $F = 5,05$, a naš F je manji).

ZADACI ZA VJEŽBU

1. Jedan proizvođač igračaka želio je ustanoviti ima li boja neke igračke utjecaj na njezinu atraktivnost pa je na 4 uzorka od po 10-ero djece mjerio minute koliko se pojedino dijete zadržalo u igri tom igračkom. Dobio je ove rezultate:

Crveni slonić	Žuti slonić	Zeleni slonić	Modri slonić
1	2	2	5
2	3	4	3
5	6	2	1
7	3	1	2
6	2	2	1
1	8	3	3
2	7	4	4
2	5	1	2
4	6	3	3
4	8	2	1

Mogu li se razlike između aritmetičkih sredina smatrati statistički značajnima?

Ako mogu, koje su razlike značajne?

2. Izvršite račun analize varijance za donje podatke:

I	II	III	IV
3	4	6	7
5	9	8	5
8	6	3	6
3	10	5	6
6	4	5	
	2	7	
		7	

21. IZBOR IZ NEPARAMETRIJSKIH TESTOVA

21.1. PARAMETRIJSKA I NEPARAMETRIJSKA STATISTIKA

Kao što znamo, znatan dio testova koje smo do sada spominjali, zahtijeva *normalnu* raspodjelu rezultata u populaciji. U praksi, međutim, često nailazimo na situacije kad nam ništa nije poznato o distribuciji populacije, ili čak znamo da populacija *nije* normalna. To je npr. slučaj kad je vjerojatnost pojavljivanja nekog događaja vrlo mala ($P < 0,10$), pa dobivamo tzv. Poissonovu distribuciju, ili kad postoji donja ili gornja granica preko koje rezultati ne mogu ići. Na primjer, u jednom vrlo lakom testu znanja većina će ispitanika postići maksimalan rezultat, pa će i krivulja biti asimetrična, a jednako će tako biti i kod jednog preteškog testa, gdje će većina postići minimalan broj bodova. U prvom slučaju gornju granicu predstavlja maksimalan broj bodova, a u drugom slučaju donju granicu predstavlja nula bodova.

Nadalje, čest je slučaj u nekim društvenim znanostima (npr. sociologiji) da se neke pojave distribuiraju upravo "suprotno" od normalne raspodjele. To je npr. slučaj sa stavovima koji često daju tzv. "U-raspodjelu", tj. najveće frekvencije nalaze se na *ekstremnim* vrijednostima apscise.

U svim tim slučajevima ne možemo primijeniti neke metode koje smo do sada spominjali.

Jednako tako ne možemo veći broj do sada poznatih statističkih računa primijeniti ako rezultati nisu izraženi *mjernim jedinicama,* nego *kvalitetama* ("zdrav", "bolestan", "mlad", "star", itd.). One metode koje se služe *mjerljivim* podacima, koji se distribuiraju *normalno,* nazivaju se *parametrijskim metodama.* Naprotiv, one metode kod kojih nije važno je li populacija normalno distribuirana, a katkada čak rezultati uopće nisu izraženi u mjernim jedinicama, nego u *frekvencijama* nekih kvaliteta, naziva se *neparametrijskim metodama* (katkad ih nazivaju i "statistika slobodna od distribucije"). Među takvim metodama mi smo već spominjali izračunavanje centila, koeficijenta rang-korelacije, hi-kvadrat testa i dr. Neparametrijske se metode upotrebljavaju prema tome prvenstveno kod podataka izraženih *nominalnim i ordinalnim* skalama, a parametrijske metode kod rezultata izraženih *intervalnim i omjernim* skalama (o skalama vidi poglavlje 19).

Razumljivo je da je neparametrijske metode *dopušteno* i moguće primijeniti i pri obradi podataka koji su izraženi intervalnim i omjernim skalama (npr. mi možemo izračunati *rang*-korelaciju između težine i visine), ali bi takav postupak bio neracionalan, jer na taj način namjerno gubimo niz informacija (npr. kod rang-korelacije, gdje registriramo samo rang, a ne i *razlike* između pojedinaca), i precizniji postupak zamjenjujemo *aproksimativnim* postupkom: ako rangovi u obje varijable korespondiraju, onda će rang-korelacija iznositi + 1, ali da smo računali koeficijent korelacije r, on bi bio manji od 1.

Zbog toga statističari kažu da je "snaga" ("power") neparametrijskih testova manja od "snage" parametrijskih testova. "Snagu" testa možemo definirati kao vjerojatnost odbacivanja nul-hipoteze, ako je ta hipoteza zaista pogrešna. Drugim riječima, snaga testa sastoji se u njegovoj sposobnosti da otkrije neku razliku, *ako ta razlika zaista postoji*. Dakle, ako između dvije populacije *postoji* razlika, to ćemo obično preciznije i uspješnije ustanoviti pomoću parametrijskih nego neparametrijskih testova.

To, doduše, ne mora uvijek biti tako, jer se može dogoditi da iz nekih "formalnih" razloga parametrijski test "zakaže", tj. njime ne uspijemo dokazati nešto što je inače potpuno evidentno i logično. Na primjer, ako imamo dva niza podataka od *istih* ispitanika i svi su podaci drugoga mjerenja po brojčanoj vrijednosti veći od korespondentnih podataka prvog mjerenja, primjena "metode diferencije" (vidi poglavlje 9.8) većinom će potvrditi da je razlika statistički značajna. Ali ako su jedan ili dva rezultata u drugoj seriji mjerenja vrlo aberantni (tj. vrlo različiti od korespondentnih rezultata prve serije mjerenja), može se dogoditi da "metodom diferencije" — iako je sada prosječna razlika između prve i druge seerije mjerenja *još i veća nego u prvom slučaju* — nećemo uspjeti odbaciti nul-hipotezu (jer se povećao varijabilitet rezultata, pa je time i standardna pogreška razlike postala veća). Naprotiv, primjenom neparametrijskih testova rázlika će biti potvrđena usprkos pojedinačnim aberantnim rezultatima — što je zapravo i logičnije, jer su ti rezultati aberantni u *istom* smjeru u kojem odstupaju i drugi rezultati.

Da bi se izbjegli određeni nesporazumi u vezi s pojavom da se kod mnogih neparametrijskih testova često susrećemo s nama poznatim z-vrijednostima (dakle s pojmom poznatim iz *normalne* raspodjele), treba reći da je neparametrijska statistika "slobodna" od bilo kakve pretpostavke o distribuciji *populacije,* ali nije, naravno, slobodna od pretpostavke o obliku očekivane varijacije i distribucije *uzoraka.* Kao što znamo, ako uzorci nisu odveć mali, od *bilo kako* distribuirane populacije uzorci aritmetičkih sredina distribuirat će se uglavnom po *normalnoj* raspodjeli — i otuda z- vrijednosti i u neparametrijskoj statistici.

Osim već do sada spomenutih, izložit ćemo još neke najpoznatije neparametrijske testove. Kako različiti autori nekima od njih daju različite nazive, uz neke testove označit ćemo ta različita imena.

21.2. DVA NEZAVISNA UZORKA

21.2.1. *Test homogenog niza (Run test, Wald-Wolfowitzov test)*

Uzmimo da jedan istraživač želi ispitati da li djeca koja su frustrirana pokazuju agresivno ponašanje; on u tu svrhu izvede ovaj eksperiment. Skupinu od 20 djece slučajnim izborom podijeli u dvije skupine po 10, od kojih eksperimentalnu skupinu (E) podvrgne postupcima koji dovode do frustracije (oduzimanje igračaka, ometanje u igri i dr.), dok je u kontrolnoj skupini (K) dopušteno da se igra u posve normalnim prilikama. Tada su djeca stavljena zajedno u jednu prostoriju, a jedan uvježbani opažač rangira ih prema stupnju agresivnosti koji ona u igri pokazuju (npr. od najslabije prema najvećoj agresivnosti).

Uzmimo nekoliko primjera, tj. nekoliko različitih mogućih situacija uz pomoć kojih ćemo protumačiti smisao testa homogenog niza.

1. Ako su sva djeca eksperimentalne grupe agresivnija od djece kontrolne skupine, opažač bi ih rangirao ovako (rang 1 = najmanje agresivno dijete; K = kontrolna skupina, E = eksperimentalna skupina):

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Dijete	K	K	K	K	K	K	K	K	K	K	E	E	E	E	E	E	E	E	E	E.

2. Kad bi frustracija dovela do toga da neka djeca postanu pojačano agresivna, a neka povučena i plašljiva, onda bi rang mogao izgledati ovako:

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Dijete	E	E	E	E	E	K	K	K	K	K	K	K	K	K	K	E	E	E	E	E.

3. Ako frustracija nije dovela ni do kakve promjene u agresivnosti, ne bismo mogli očekivati pojavljivanje nikakvih pravilnosti, pa bi rang prema tome bio slučajan, na primjer ovakav:

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Dijete	K	E	E	K	K	K	E	E	K	E	K	E	E	E	K	K	E	K	K	E.

Ako svaku skupinu *jednakih* znakova nazovemo homogenim nizom, pa te nizove zbrojimo u naša tri primjera, dobivamo u prvom slučaju *dva* niza, u drugom slučaju *tri* niza, a u trećem slučaju *dvanaest* nizova (i *jedan* znak računa se u niz!). Ako u nekoj listi postoji N članova, m od jedne a n od druge vrste, ne može biti manje od 2 niza, niti više od N nizova. Da bismo dobili 2 niza, moraju svi članovi iste vrste biti zajedno (naš prvi primjer), a da bi dobili N nizova, morali bi se članovi oba niza *pravilno* izmjenjivati: K, E, K, E, K, E itd. (to naravno samo onda ako su oba niza jednako velika ili se razlikuju samo za 1). *Ni za jednu* od te dvije kombinacije *nije mnogo vjerojatno da bi se dogodila slučajno*. Prema tome, ako ustanovimo da je broj nizova *vrlo mali* ili *vrlo velik,* pretpostavit ćemo da je to uzrokovano nečim drugim, a ne pukim slučajem. U praksi — kad želimo testirati pripadaju li oba uzorka istoj populaciji — obično nas zanima samo to je li broj nizova *manji* od onoga koji bismo još mogli očekivati slučajno. Da bismo ustanovili koliko nizova možemo smatrati statistički značajno premalim brojem, služimo se tablicom M u

Dodatku. U toj tablici m je broj članova jedne, a n je broj članova druge grupe; u samoj tablici nalaze se granične vrijednosti broja nizova. Svaki broj nizova koji je *manji* od broja koji očitamo u tablici, ili *jednak* kao taj broj, mogao bi se slučajno dogoditi samo u 5% ili manje slučajeva, ako zapravo ne bi postojala razlika među uzorcima. Dakle, svaki broj nizova jednak ili manji od broja u tablici značajan je na razini od 5%. U našem slučaju mi tablicu M čitamo pod $m = 10$ i $n = 10$, i dobivamo broj 6. Prema tome, u prvom primjeru (2 niza) i u drugom primjeru (3 niza) možemo odbaciti nul-hipotezu i zaključiti da obje skupine ne pripadaju u istu populaciju, pa da se zato po agresivnosti frustrirana djeca *razlikuju* od nefrustrirane djece. Naprotiv, u trećem primjeru (12 nizova) ne možemo odbaciti nul- hipotezu.

Test homogenog niza možemo naravno upotrijebiti i kad su rezultati izraženi u intervalnoj skali. Na primjer, ako smo u grupi A ($m = 4$) i grupi B ($n = 5$) dobili ove rezultate:

Grupa A:	12	16	8	10	
Grupa B:	11	6	6	3	4,

možemo ih po rangu poredati ovako:

3	4	6	6	8	10	11	12	16
B	B	B	B	A	A	B	A	A.

U ovom slučaju imamo 4 niza, a iz tablice M uz $m = 4$ i $n = 5$ čitamo broj 2. Prema tome, *prihvaćamo* nul-hipotezu, tj. zaključujemo da se ta dva uzorka ne razlikuju statistički značajno.

Ako su m ili n veći od 20, test homogenog niza računa se ovako: Najprije treba izračunati teoretsku, tj. pravu aritmetičku sredinu nizova prema formuli:

$$\overline{X}_{\text{niza}} = \frac{2mn}{m+n} + 1 \qquad (21.1)$$

i nakon toga "standardnu devijaciju" niza prema formuli:

$$S_{\text{niza}} = \sqrt{\frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}}. \qquad (21.2)$$

Ako smo, na primjer, imali uzorak od ukupno 100 podataka, i to 40 podataka jedne i 60 podataka druge vrste, te smo našli 39 nizova, dobivamo ove vrijednosti:

$$\overline{X}_{\text{niza}} = \frac{2 \cdot 40 \cdot 60}{40 + 60} + 1 = 49$$

$$S_{\text{niza}} = \sqrt{\frac{2 \cdot 40 \cdot 60(2 \cdot 40 \cdot 60 - 40 - 60)}{(40+60)^2(40+60-1)}} = 4,77.$$

Tada izračunamo vrijednost z:

$$z = \frac{\text{Dobiveni broj nizova} - \overline{X}_{\text{niza}}}{S_{\text{niza}}}. \qquad (21.3)$$

U našem slučaju to iznosi: $z = \dfrac{39 - 49}{4,77} = -2,10.$

Ako je iznos dobivenog z (bez obzira na predznak) veći od 1,654, možemo odbaciti nul-hipotezu i smatrati da oba uzorka ne pripadaju istoj populaciji. (Razlika je značajna na razini od 5%.)

Ako se dogodi da su dva (ili više rezultata iz *suprotnih* grupa jednaki i da prema tome zauzimaju isti rang, nastaju kod ovog testa određene teškoće: ako su to dva rezultata, možemo ih poredati AB i BA, a ako tri rezultata zauzimaju isti rang, možemo ih poredati: ABA, AAB i BAA. Raspored tih rezultata zacijelo će utjecati na broj nizova. U takvim slučajevima preporučuje se izračunati z za svaku od tih kombinacija: ako je z u svim slučajevima značajan, onda takvi zajednički rangovi ne predstavljaju problem. Ali ako neke kombinacije daju značajni, a neke neznačajni z, preporučuje se odrediti vjerojatnost za svaki z, i izračunati *prosječnu* vjerojatnost, pa tek na temelju tako dobivene vrijednosti odlučiti hoćemo li prihvatiti ili odbaciti nul-hipotezu.

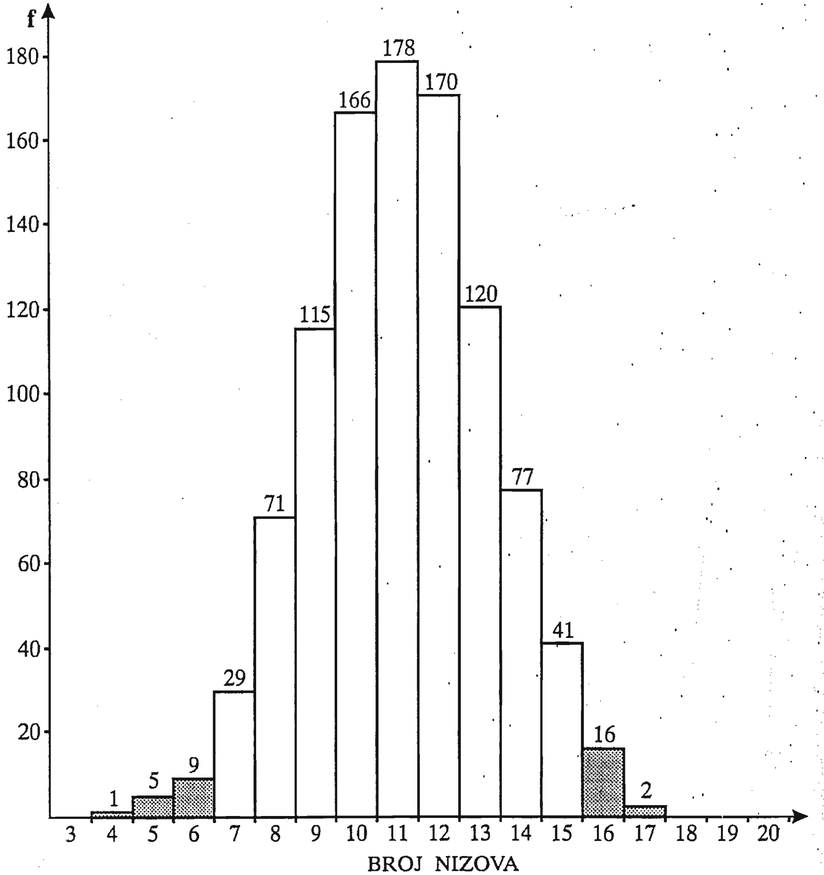
Test homogenog niza — ako uzorak nije vrlo mali — omogućuje odbacivanje nul-hipoteze ako se oba uzorka međusobno razlikuju u *bilo kojem* pogledu: u centralnoj tendenciji, u varijabilnosti, u simetričnosti i dr. Ako nas upravo zanima razlikuju li se uzorci samo po svom *medijanu,* bolje je upotrijebiti medijan test.

Kao što smo već rekli, i *suviše velik* broj nizova može biti "sumnjiv". Naime, pokušajte uzeti 10 crvenih i 10 crnih karata, i dobro ih izmiješajte, te ustanovite koliko ste nizova dobili. Ako pokus ponovite nekoliko desetaka puta, lako ćete ustanoviti da ste najčešće dobivali otprilike 9-13 nizova, a manje ili više od toga već rjeđe. Praktički gotovo uopće nije se moglo očekivati da ćete dobiti 19 ili čak 20 nizova, jer (kod 20 nizova) to bi značilo da bi dvadeset karata moralo ovako biti poredano: crna, crvena, crna, crvena, crna, crvena, crna, crvena, itd., još dvanaest puta! A to je — vjerujem da ćete se složiti — gotovo nemoguće.

Svojedobno sam učinio takav pokus i u 1 000 miješanja 10 crnih i 10 crvenih karata dobio sam prilično simetričnu distribuciju nizova, sličnu normalnoj distribuciji, pri čemu je najmanji broj dobivenih nizova iznosio 4 (dobiven je jedanput u 1 000), a najveći broj nizova 17 (dobiven dvaput). Deset nizova dobiveno je 166 puta, 11 nizova 178 puta i 12 nizova 170 puta. Ti su rezultati prikazani na slici 21.1.

Gledajući distribuciju na slici, jasno je ujedno kako su nastale tablice za očitavanje prevelikog ili premalog broja nizova, tj. onog broja čija je vjerojatnost *slučajnog* pojavljivanja manja od 5% ili 1%. (Naša tablica M u Dodatku je samo za lijevu stranu distribucije, tj. za dokazivanje da oba uzorka *ne* pripadaju istoj populaciji.)

Kao što smo kod hi-kvadrat testa rekli da odveć mali hi-kvadrat može biti sumnjiv, jer je njegovo *slučajno* pojavljivanje malo vjerojatno, tako i kod testa niza *prevelik* broj nizova nije baš vrlo vjerojatan, pa ako netko — želeći dokazati da se uzorci ne razlikuju, tj. da pripadaju istoj populaciji — navede prevelik broj nizova, možemo posumnjati u istinitost podataka.



Slika 21.1. Rezultat pokusa, pri kojem je 1 000 puta bilo pomiješano 10 crnih i 10 crvenih karata, te je promatran broj nizova nakon svakog miješanja

21.2.2. *Medijan test*

To je vrlo jednostavan test koji se zapravo svodi na hi-kvadrat test, a kojim ispitujemo pripadaju li dva uzorka populaciji s istim medijanom. U parametrijskoj

statistici njemu djelomično odgovara t-test, kojim ispitujemo značajnost razlika između dvije aritmetičke sredine.

Uzmimo da smo na dva uzorka (koji mogu po veličini biti jednaki ili različiti) dobili u nekom mjerenju ove rezultate, koje smo zbog preglednosti poredali prema veličini:

Uzorak I 8 9 9 10 10 10 12 13 15 17 17 18 19 19 21 23 24 25 26 28 28 29 31 31;

Uzorak II 3 6 7 7 8 8 8 10 12 16 19 22 24 27 30 32.

Princip medijan testa sastoji se u tome da nađemo centralnu vrijednost (tj. medijan) ***iz svih rezultata*** zajedno i da ih unesemo u $2 \cdot 2$ tablicu. Budući da u našem primjeru imamo ukupno 41 rezultat, medijan je dvadeset prvi rezultat po veličini (o izračunavanju medijana vidi poglavlje 4.3.1), a to je 17. Ako sve rezultate koji su iznad medijana, označimo oznakom "plus", a rezultate na medijanu ili ispod njega oznakom "minus", dobivamo:

Uzorak I − − − − − − − − − − − + + + + + + + + + + + + +
Uzorak II − − − − − − − − − − + + + + + + +

Unesemo li frekvencije tih rezultata u tablicu, dobivamo:

	+	−	
Uzorak I	13	11	24
Uzorak II	7	10	17
	20	21	41 .

Iz ove tablice sada izračunamo hi-kvadrat test, vodeći računa o svim pravilima koja vrijede za hi-kvadrat, pa stoga u ovom slučaju moramo (jer se radi o $2 \cdot 2$ tablici) upotrijebiti Yatesovu korekturu. Izračunani hi-kvadrat iznosi 0,258, pa zato *prihvaćamo* hipotezu da se medijani obaju uzoraka statistički značajno ne razlikuju.

Ako je ukupan broj rezultata paran, medijan je aritmetička sredina između dva rezultata koji se nalaze u sredini niza svih rezultata poredanih po veličini. U tom slučaju će nam svi rezultati biti ili iznad ili ispod medijana, a niti jedan na samom medijanu.

21.2.3. *Test sume rangova (Wilcoxonov T-test, Mann-Whitneyev U- test)*

Taj je test donekle sličan testu homogenih nizova, ali on koristi više informacija (tj. koristi rangove, a ne samo podjelu u dvije kategorije) i zato se može smatrati boljim i "snažnijim". Kao i medijan testom, testom sume rangova testiramo pripadaju li dva uzorka u populaciju s istim medijanom.

Uzmimo da imamo dvije skupine ispitanika od kojih smo prvoj skupini ($N_1 = 9$) davali neki vitaminski preparat, a drugoj skupini ($N_2 = 8$) nismo. U ostalim faktorima obje su skupine naravno slične. Zanima nas pokazuje li "vitaminska" grupa bolje "opće zdravstveno stanje".

Ispitanici obiju skupina označeni su slovima:

Eksperimentalna skupina	Kontrolna skupina
A	J
B	K
C	L
D	M
E	N
F	O
G	P
H	R
I	

Postupak se sastoji u tome da uzmemo sve ispitanike zajedno i da ih rangiramo prema općem zdravstvenom stanju, s tim da rang 1 damo najzdravijem ispitaniku. Uzmimo da smo dobili ovaj rezultat:

Rang	Ispitanik	Rangovi eksp. grupe	Rangovi kontr. grupe
1	B	1	
2	H	2	
3	P		3
4	C	4	
5,5	M		5,5
5,5	F	5,5	
7	R		7
8	K		8
9	G	9	
10	D	10	
11	A	11	
12	J		12
13,5	O		13,5
13,5	N		13,5
15	D	15	
16	L		16
17	I	17	
		$T_1 = 74,5$	$T_2 = 78,5$

Za obje skupine posebno zbrojimo rangove. Zbog kontrole dobro je provjeriti sumu rangova formulom:

$$T_1 + T_2 = \frac{N(N+1)}{2}$$

$$153 = 153.$$

Nakon toga izračunamo izraz:

$$z = \frac{|2T_i - N_i(N+1)| - 2}{\sqrt{\frac{N_1 N_2 (N+1)}{3}}}, \tag{21.4}$$

gdje T_i znači *bilo koju* od suma rangova, a N_i znači broj ispitanika u skupini iz koje smo uzeli T_i. (Podsjećamo čitaoca da znak || znači da se uzima u obzir apsolutna vrijednost izraza, pa, prema tome, brojnik te formule govori da izraz treba *smanjiti* za 2; ako je izraz u zagradi pozitivan, oduzet ćemo 2, a ako je negativan, matematički ćemo pribrojiti 2.)

Izračunamo li z za eksperimentalnu skupinu, dobivamo:

$$z = \frac{|2 \cdot 74,5 - 9 \cdot 18| - 2}{\sqrt{\frac{9 \cdot 8 \cdot 18}{3}}} = 0,53.$$

Dobit ćemo isti rezultat ako z izračunamo za kontrolnu grupu:

$$z = \frac{|2 \cdot 78,5 - 8 \cdot 18| - 2}{\sqrt{\frac{8 \cdot 9 \cdot 18}{3}}} = 0,53.$$

Ako je broj ispitanika u svakoj skupini barem 8, tada izračunani z daje normalnu distribuciju s aritmetičkom sredinom 0 i standardnom devijacijom 1, tj. izračunana vrijednost nije ništa drugo nego nama već dobro poznata z-vrijednost (pa, prema tome, i t-vrijednost). Budući da jedino rezultate koji su veći od 1,96 možemo smatrati statistički značajnim (na razini značajnosti od 5%), zaključujemo da je naš z premalen, pa prema tome prihvaćamo nul-hipotezu: nismo dokazali da se ta dva uzorka statistički značajno razlikuju. (To je "dvosmjerni" test. Za "jednosmjerni" test granična vrijednost z bila bi — kako znamo — 1,64.)

Znatno je jednostavnije testirati značajnost razlike među uzorcima uz pomoć tablice N u Dodatku. Ako su uzorci manji od 8, tablicu *moramo* upotrijebiti. Međutim, upotreba tablice zahtijeva jedan dodatni račun koji ćemo ukratko rastumačiti. Kad značajnost očitavamo s tablice, treba uzeti u obzir samo *manji* uzorak, tj. njegovu sumu rangova. No, međutim, njegova bi suma rangova bila drukčija (veća ili manja) da smo rezultate rangirali obratnim redom. Zato tu *drugu* sumu rangova treba prethodno izračunati, i to prema formuli:

$$T' = N_{\text{manjeg uzorka}}(N_1 + N_2 + 1) - T_{\text{manjeg uzorka}}. \tag{21.5}$$

U našem primjeru dobivamo:

$$T' = 8(9 + 8 + 1) - 78,5 = 144 - 78,5 = 65,5.$$

Za tablicu se uzima *manja* od te dvije sume rangova T'. Budući da naš prvi T' iznosi 78,5, a drugi T' iznosi 65,5, u daljnji postupak uzimamo broj 65,5. U tablici, na sjecištu broja 8 (N manjeg uzorka), s brojem 9 (N većeg uzorka), čitamo 51, što znači da *manja* suma rangova manjeg uzorka smije biti najviše 51. Kako je naša suma veća ($65,5 > 51$), zaključujemo da ne možemo odbaciti nul-hipotezu — dakle zaključak je isti kao i računanjem z-vrijednosti.

Kao što smo iz našeg primjera vidjeli, jednakim rezultatima daje se zajednički rang (o izračunavanju zajedničkog ranga vidi str. 200). Ako imamo *mnogo* zajedničkih rangova, u račun testa sume rangova treba unijeti određenu korekturu, tako da formula za z glasi:

$$z = \frac{|T_i - O\check{c}(T_i)| - 1}{\sqrt{\left[\frac{N_1 N_2}{N(N-1)}\right]\left[\frac{N(N^2-1)}{12} - \sum T\right]}}, \tag{21.6}$$

gdje $O\check{c}\ (T_i)$ = očekivana suma rangova T_i a $N = N_1 + N_2$

$T = \frac{n(n^2-1)}{12}$, gdje je n broj rezultata koji čine zajednički rang.

Očekivana suma rangova računa se prema formuli:

$$O\check{c}(T_i) = \frac{N_i(N_1 + N_2 + 1)}{2}. \tag{21.7}$$

Mi u našem primjeru imamo dva puta po dva rezultata koji su vezani zajedničkim rangom (rezultati MF i ON), pa naš ΣT iznosi:

$$\frac{2 \cdot 3}{12} + \frac{2 \cdot 3}{12} = 1.$$

Izračunamo li naš primjer s tom korekturom, dobivamo:

$$O\check{c}(T_i) = \frac{9 \cdot 18}{2} = 81$$

$$z = \frac{|74,5 - 81| - 1}{\sqrt{\left[\frac{72}{17 \cdot 16}\right]\left(\frac{4\,913 - 17}{12} - 1\right)}} = 0,53.$$

Kako se vidi, zbog malog broja zajedničkih ranogva, u ovom se slučaju z nije promijenio.

21.2.4. *Siegel-Tukeyev test*

Siegel-Tukeyev test pogodan je *jedino* za testiranje značajnosti razlika u *varijabilitetu* (to svojstvo ima i test homogenog niza). Po upotrijebljenim *formulama* on je jednak upravo opisanom testu sume rangova, ali se razlikuje po *načinu rangiranja* rezultata. Dok smo u testu sume rangova rangirali rezultate na standardan, uobičajen način, ovdje je rangiranje malo neobično, što će se vidjeti iz primjera.

Uzmimo da imamo dva uzorka s ovim rezultatima:

Uzorak I:	25	5	14	19	0	17	15	8	8
Uzorak II:	12	16	6	13	13	3	10	10	11.

Rezultate ćemo rangirati tako da ćemo rang 1 dati najnižem rezultatu, rangove 2 i 3 *najvišim* rezultatima, rangove 4 i 5 idućim najnižim rezultatima, 6 i 7 idućim najvišim, itd. Pritom ćemo, kao i u testu sume rangova, bilježiti pripadnost uzorku. U našem slučaju, dakle, imamo:

Rezultat:	0	3	5	6	8	8	10	10	11	12	13	13	14
Uzorak:	I	II	I	II	I	I	II	II	II	II	II	II	I
Rang:	1	4	5	8	9	12	13	16	17	18	15	14	11
Rezultat:	15	16	17	19	25								
Uzorak:	I	II	I	I	I								
Rang:	10	7	6	3	2.								

Ako je totalni broj rangova *neparan,* ne uzima se srednji rezultat u daljnjem računu.

Ako se obje populacije *ne* razlikuju u varijabilitetu (nul- hipoteza), sume rangova jednog i drugog uzorka bit će slične. Naprotiv, ako se populacije u varijabilitetu razlikuju, uzorak iz populacije s većim varijabilitetom tendirat će ekstremima sekvence rangova, i stoga će biti označen nižim rangovima. Naprotiv, uzorak iz populacije s manjim varijabilitetom, tendirat će prema sredini sekvence rangova, i zato će biti označen višim rangovima. Dakle, uzorak s većom varijacijom imat će manju sumu rangova. To se lako može dokazati na jednom ekstremnom primjeru:

Rezultati uzorka I:	1	2	3	15	20	25
Rezultati uzorka II:	4	5	6	7	10	12 .

Rezultati:	1	2	3	4	5	6	7	10	12	15	20	25
Uzorak:	I	I	I	II	II	II	II	II	II	I	I	I
Rang:	1	4	5	8	9	12	11	10	7	6	3	2

Suma rangova uzorka I = 21
Suma rangova uzorka II = 57 .

Kako se vidi, uzorak I, s većom varijacijom, ima značajno nižu sumu rangova.

U našem primjeru suma rangova uzorka I (T_1) iznosi 59, a suma rangova uzorka II (T_2) iznosi 112.

Kontrola računa:

$$T_1 + T_2 = \frac{N(N+1)}{2}$$

$$171 = 171.$$

Dalje računamo jednako kao u testu sume rangova, tj. primijenimo formulu (21.4):

$$z = \frac{|2\,T_i - N_i(N+1)| - 2}{\sqrt{\frac{N_1 N_2 (N+1)}{3}}}.$$

Ako u formulu umjesto T_i uvrstimo vrijednosti prvog uzorka, dobivamo:

$$z = \frac{|2 \cdot 59 - 9 \cdot 19| - 2}{\sqrt{9 \cdot 9 \cdot 19/3}} = 2,25.$$

(Da smo u formulu uzeli T_2, dobili bismo isti rezultat.)
Dakle, uzorci se statistički *značajno* razlikuju po varijabilitetu ($P < 0,05$).
Ako u rezultatima ima *zajedničkih rangova*, valja upamtiti ovo:

a) Ako se zajednički ("vezani") rangovi nalaze samo unutar istog uzorka, to neće utjecati na sumu rangova T, pa je zato svejedno kojem od istih rezultata damo viši, kojem niži rang.

b) Ako postoje "vezani" rangovi između *oba* uzorka, potrebno je u postupku izračunavanja rangova, kao i u formuli za izračunavanje z, provesti neke promjene, koje ćemo ukratko opisati poslužívši se primjerom iz jednog eksperimenta.

U jednom laboratorijskom pokusu ispitanici, ispitani u dvije različite eksperimentalne situacije, dali su ove rezultate (rezultati su poredani prema veličini):

Rezultat:	1,5	3,2	3,4	3,6	4,4	4,4	5,2	5,2	5,4	5,4	5,6	5,6	5,8
Uzorak:	2	1	1	2	1	2	1	2	1	2	1	1	2
Rang:	1	4	5	8	9	12	13	16	17	20	21	24	25
					10,5	10,5	14,5	14,5	18,5	18,5	22,5	22,5	

Rezultat:	6,2	6,2	6,2	6,4	6,6	7,0	7,2	7,6	7,6	8,0	8,0	8,0	8,8
Uzorak:	1	1	2	2	2	2	1	1	2	1	1	2	1
Rang:	28	29	32	33	36	36	35	34	31	30	27	26	23
	29,7	29,7	29,7					32,5	32,5	27,7	27,7	27,7	

Rezultat:	9,0	9,2	9,6	9,8	9,8	10,2	10,2	10,2	11,0	11,2	14,6
Uzorak:	2	2	2	1	2	1	2	2	1	2	2
Rang:	22	19	18	15	14	11	10	7	6	3	2
				14,5	14,5	9,3	9,3	9,3			

Pri rangiranju u početku ćemo postupiti prema standardnom postupku — kako je to učinjeno i u ovom primjeru — tj. redom ćemo napisati rezultate i rangirati ih na opisan način, ne vodeći, zasad, računa kod vezanih rangova koji smo uzorak stavili prije a koji kasnije. U primjeru su vezani rezultati podcrtani.

Budući da ukupno imamo neparan broj rezultata (37), srednji rezultat nećemo dalje uzimati u obzir (srednji rezultat je 7,0).

Sada ćemo vezanim rezultatima dati i vezane rangove. Ti su rangovi ispisani ispod vezanih rezultata.

Suma rangova $T_1 = 332,6$
$T_2 = 333,5.$

Sada treba izračunati izraze S_1 i S_2, pri čemu:
S_1 = suma onih kvadriranih rangova koji su tvorili vezane rangove,
S_2 = suma kvadriranih vezanih rangova.

U našem primjeru, dakle, dobivamo:

$$S_1 = 9^2 + 12^2 + 13^2 + 16^2 + 17^2 + 20^2 + 21^2 + 24^2 + 28^2 + 29^2 + 32^2 + 34^2 + \\ +31^2 + 30^2 + 27^2 + 26^2 + 15^2 + 14^2 + 11^2 + 10^2 + 7^2 = 10\,118;$$

$$S_2 = 10,5^2 + 10,5^2 + 14,5^2 + 14,5^2 + 18,5^2 + 18,5^2 + 22,5^2 + 22,5^2 + 29,7^2 + \\ +29,7^2 + 29,7^2 + 32,5^2 + 32,5^2 + 27,7^2 + 27,7^2 + 27,7^2 + \\ +14,5^2 + 14,5^2 + 9,3^2 + 9,3^2 + 9,3^2 = 10\,078,6.$$

Formula za izračunavanje z također se mijenja, tj. mijenja se njezin nazivnik i ona sada glasi:

$$z = \frac{|2T_1 - N_1(N+1)| - 2}{\sqrt{\dfrac{N_1 N_2 (N+1)}{3} - \dfrac{4 N_1 N_2 (S_1 - S_2)}{N(N-1)}}}. \qquad (21.8)$$

Uvrstimo li naše vrijednosti u tu formulu, dobivamo:

$$z = \frac{|2 \cdot 332,6 - 17 \cdot 37| - 2}{\sqrt{\dfrac{17 \cdot 19 \cdot 37}{3} - \dfrac{4 \cdot 17 \cdot 19 \cdot 39,4}{36 \cdot 35}}} = 0,54.$$

Kako se vidi, razliku u varijabilitetu ne možemo smatrati statistički značajnom.

21.3. DVA ZAVISNA UZORKA

Kao i kod parametrijskih testova, i kod neparametrijskih testova metode su nešto drukčije ako radimo sa *zavisnim* uzorcima, dakle ili dva puta s *istom* skupinom ispitanika ili s dvije skupine ispitanika, u kojima svaki ispitanik jedne skupine ima svoj *par* u drugoj skupini.

21.3.1. *Test predznaka (Sign test)*

Uzmimo da je izveden eksperiment s 15 pari jednojajčanih blizanaca. Eksperiment se sastojao u tome da je jedan član svakog para blizanaca bio hranjen prvih 8 mjeseci od majke, dok je drugi član para već od drugog tjedna hranjen na bočicu. Kad je svaki par blizanaca imao 5 godina, izvršen je zdravstveni pregled razvoja. Nul-hipoteza koju treba testirati je ova: nema razlike u tjelesnom razvoju između blizanaca othranjenih od majki i onih othranjenih na bočicu (osim, dakako, razlika koje možemo pripisati slučaju).

Postupak testa predznaka sastoji se u tome da ćemo svakom *paru*, kod kojeg je bolje razvijen od majke othranjeni blizanac dati predznak (oznaku) "plus", a svakom paru gdje je bolje razvijen na bočicu othranjeni blizanac predznak "minus". Parove u kojima nismo našli nikakvu razliku označit ćemo s "nulom" *i nećemo ih kasnije uzeti u račun.* Time će se N, tj. broj parova, smanjiti. Dakle, u testu predznaka N = broj parova koji se razlikuju. (Ako imamo mnogo parova bez razlike, test predznaka se ne može upotrijebiti!)

Uzmimo da smo kod naših 15 pari blizanaca dobili ovaj rezultat:

Parovi	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Predznak	+	+	+	+	+	−	−	+	+	+	−	+	+	−	+

Kako se vidi, u 11 od 15 pari bio je bolje razvijen blizanac koji je othranjen od majke, a u 4 para bio je bolje razvijen blizanac othranjen na bočicu. Dakle, u računu imamo 4 minusa i 11 plusova. Daljnji postupak radimo s *manjim* od ta dva broja, dakle s brojem 4. U tablici O u Dodatku uz N = 15 (15 parova) nalazimo brojeve 2, 3, 3 i 4. Ti brojevi znače granične vrijednosti, tj. najveći dopušten broj razlika koji još možemo dopustiti uz razinu značajnosti od 1%, 5%, 10% i 25%. Naš broj razlika je 4, što pokazuje da bi se takva razlika mogla i slučajno dogoditi u 25% slučajeva. Tablica je međutim "dvosmjerna", tj. računa razlike u bilo kojem smjeru, a mi smo u ovom slučaju već od početka zainteresirani hoće li blizanci koje je othranila majka biti kasnije *bolje* razvijeni, pa prema tome možemo razliku testirati "jednosmjernim" testom. U tom slučaju treba vjerojatnost u tablici "raspoloviti", tj. podijeliti s 2. Dakle, u našem primjeru dobivamo da bi se ta razlika (u tom smjeru) mogla slučajno dogoditi u 12,5% slučajeva — što je naravno još uvijek nedovoljno, jer ne želimo ići na razine značajnosti, koje su veće od $P = 0,05$, dakle u tablici čitati pod 10%).

Evo još jednog primjera za test predznaka:

Na treningu košarke imali smo 10 ispitanika, kojima smo izmjerili koliko su puta od 50 bacanja lopte pogodili "koš". Nakon toga sve smo ispitanike podvrgnuli napornoj igri koja ih je umorila i ponovno im izmjerili broj pogodaka u 50 bacanja. Zanima nas je li umor *smanjio* (jednosmjerni test!) preciznost pogodaka.

Pretpostavimo da smo dobili ove rezultate:

Ispitanici	A	B	C	D	E	F	G	H	I	J
Broj pogodaka kod prvog mjerenja	15	19	31	36	10	11	19	15	10	16
Broj pogodaka kod drugog mjerenja	17	20	16	8	10	6	7	8	12	8
Predznak	+	+	−	−	0	−	−	−	+	−

Označimo li povećanje broja pogodaka s +, smanjenje broja pogodaka s −, a nepromijenjeni rezultat s 0, dobivamo 3 plusa, 6 minusa i jednu nulu. Nula se izbacuje iz računa te tako broj parova postaje $N = 9$. Od brojeva 3 i 6 broj 3 je

manji, i zato u tablici O kod $N = 9$ čitamo 0, 1, 1, 2. Iz toga zaključujemo da bi na razini značajnosti od 5% (čitamo pod 10%, jer nas zanima smjer razlike!) naš manji broj morao iznositi najviše 1. Budući da mi imamo 3, zaključujemo da razlika nije statistički značajna.

U ovom i u prijašnjem primjeru mogli smo uočiti jedan bitan nedostatak testa predznaka: on ne uzima u obzir *veličinu* razlike, nego samo njezin *smjer*. Konkretno, u prvom našem primjeru moglo se dogoditi kod 11 parova blizanaca da je od majke othranjeni blizanac bio *znatno* bolje razvijen od svog para, a kod 4 preostala para da je na bočicu othranjeni blizanac bio samo *neznatno* bolje razvijen. Za test predznaka to je, na žalost, "svejedno". Zato za taj test neki statističari (Senders) kažu da je to vrlo prikladan postupak štednje vremena kada su razlike velike, i vrlo koristan instrument da bi se aproksimativno ustanovilo "odakle vjetar puše", tj. postoji li uopće neki fenomen ili ne. Ali kad nam je potrebna veća preciznost, gdje želimo iskoristiti i ostale podatke koje imamo, a to su *veličine* pojedinih razlika, mnogo je prikladniji Wilcoxonov test.

21.3.2. *Wilcoxonov test ekvivalentnih parova*

Taj test zahtijeva *mjerene* vrijednosti (dakle intervalnu ili omjernu skalu). Postupak se sastoji u tome da izračunamo razlike (d) između oba člana u svakom paru. Ako razlike nema, ona je 0, i taj se par ispušta iz daljnje obrade. Razlike mogu biti pozitivne i negativne. Te se razlike *rangiraju i to bez obzira na predznak. Najmanja razlika* dobiva rang 1, iduća rang 2 itd. Ako su dvije ili više razlika jednake veličine, dobivaju zajednički rang. Nakon toga svakom se rangu daje onaj isti predznak koji je imala i razlika: ako je razlika pozitivna, i rang je pozitivan, a ako je razlika negativna, i rang je negativan.

Pod pretpostavkom nul-hipoteze (tj. da nema razlike među uzorcima) postojat će tendencija da suma pozitivnih i suma negativnih rangova budu jednake ili slične. Ako postoji značajna razlika u sumi, to već govori u prilog odbacivanja nul-hipoteze. Iz tablice možemo očitati koliko najviše smije iznositi *manja* suma rangova (T) uz određeni broj parova (N).

Primjer 1:

Uzmimo da imamo u nekom ispitivanju ovih 10 parova rezultata:

Parovi	A	B	C	D	E	F	G	H	I	J
Rezultat jednog člana para	15	19	31	36	10	11	19	15	10	16
Rezultat drugog člana para	19	30	26	8	10	6	17	13	22	8
Razlike (d)	−4	−11	5	28	0	5	2	2	−12	8
Rang razlika	−3	−7	4,5	9		4,5	1,5	1,5	−8	6.

Izračunali smo razlike d, i te razlike rangirali od najmanje do najveće (bez obzira na predznak), a rangovima smo dali isti predznak koji je imala i razlika.

Suma negativnih rangova iznosi 18 (zapravo −18, ali budući da je važna samo veličina broja, sada možemo predznak izostaviti). Suma pozitivnih rangova iznosi 27. Manja suma rangova = $T = 18$. $N = 9$ (jer smo jedan par odbacili). Tablica P u Dodatku pokazuje da bi na razini značajnosti od 5% (u "dvosmjernom" testu) manja suma rangova smjela iznositi najviše 6. Budući da je naš $T = 18$, zaključujemo da razliku ne možemo smatrati statistički značajnom.

Ako nas unaprijed zanima *smjer* razlike, ravnat ćemo se u tablici prema razini značajnosti označenoj za "jednosmjerni" test. Ekstrapolirajući naš $T = 18$ na razinu $P/2 = 0,05$ (što u tablici nije navedeno), možemo sa sigurnošću zaključiti da bi i u tom slučaju naš T bio prevelik.

Primjer 2:

Pokušajmo naš primjer s bacanjem lopte (drugi primjer iz Testa predznaka, str. 334) obraditi Wilcoxonovim testom:

	A	B	C	D	E	F	G	H	I	J
	15	19	31	36	10	11	19	15	10	16
	17	20	16	8	10	6	7	8	12	8
Razlike (d)	2	1	−15	−28	0	−5	−12	−7	2	−8
Rang razlike	2,5	1	−8	−9		−4	−7	−5	2,5	−6

Suma negativnih rangova = 39.
Suma pozitivnih rangova = 6.
$T = 6$.
$N = 9$.

Iz tablice P možemo ustanoviti da je razlika između oba uzorka u "dvosmjernom" testu *značajna* na razini od 5%, odnosno da je u "jednosmjernom" testu (koji smijemo upotrijebiti jer nas je zanimalo samo je li umor *smanjio* preciznost u gađanju) razlika značajna na razini od 2,5%.

Kao što se iz ovog primjera očito vidi, "snaga" Wilcoxonova testa znatno je veća od "snage" Testa predznaka, jer smo Wilcoxonovim testom dokazali *postojanje* jedne razlike koju nismo mogli dokazati Testom predznaka.

N a p o m e n a. Taj smo primjer mogli još preciznije obraditi parametrijskim testom, koji smo nazvali "metoda diferencije" (vidi str. 151). Da smo to učinili, dobili bismo:

$$\overline{X}_{\text{dif}} = -7$$
$$s = 9,49$$
$$s_{\overline{X}_{\text{dif}}} = 3,00$$
$$t = \frac{-7}{3} = -2,33.$$

Za 9 stupnjeva slobode ta je razlika značajna na razini koja je nešto manja od 5%. Kako se vidi, usprkos tome što za neparametrijske testove vrijedi pravilo da su manje "snažni" od parametrijskih, neki od njih daju rezultate koji su *vrlo bliski* rezultatima što bismo ih dobili parametrijskim testom.

Ako je uzorak velik (veći od 25), T ima približno normalnu raspodjelu, te se može izračunati z, koji glasi:

$$z = \frac{T - \dfrac{N(N+1)}{4}}{\sqrt{\dfrac{N(N+1)(2N+1)}{24}}}. \tag{21.9}$$

Budući da se radi o normalnoj raspodjeli, kod "dvosmjernoga" testa dovoljna je vrijednost z od 1,96 da bismo na razini značajnosti od 5% mogli smatrati da je razlika značajna, odnosno z od 1,64 da bismo kod "jednosmjernog" testa razliku mogli smatrati značajnom.

21.4. VIŠE NEZAVISNIH UZORAKA

21.4.1. *Prošireni medijan test*

Ako imamo *više* nezavisnih skupina pa želimo testirati pripadaju li one ili ne pripadaju populaciji s istim medijanom, možemo se poslužiti Proširenim medijan testom na sličan način kao što smo to učinili kod dva uzorka.

Postupak se sastoji u tome da nađemo medijan *svih* rezultata i da rezultate, koji su iznad medijana, označimo s "plus", a one ispod medijana s "minus". Ako je broj rezultata neparan, medijan postaje jedan ili više postojećih rezultata; u tom slučaju i rezultati koji predstavljaju medijan dobivaju oznaku "minus". Rezultati se nakon toga uvrste u $2 \cdot k$ tablicu (k = broj uzoraka) i izračuna se hi-kvadrat test.

Primjer. Na 4 nezavisna uzorka dobili smo ove rezultate (zbog preglednosti oni su poredani prema veličini):

I	8	12	13	14	20	21	25	33	43	45	47		
II	10	15	19	25	30	38	40	45	47	48	51	54	55
III	16	16	17	21	29	33	44	45	46	53	62	67	
IV	22	41	49	54	59	60	65	69	71	75.			

Kako imamo 46 rezultata, medijan je sredina između 23. i 24. rezultata (tj. između vrijednosti 41 i 43), i iznosi 42.

Označimo li u svakom uzorku rezultate već prema tome padaju li ispod ili iznad medijana, dobivamo:

I	−	−	−	−	−	−	−	−	+	+	+		
II	−	−	−	−	−	−	−	+	+	+	+	+	+
III	−	−	−	−	−	−	+	+	+	+	+	+	
IV	−	−	+	+	+	+	+	+	+	+			

Te rezultate unesemo u tablicu:

	I	II	III	IV	
Iznad medijana	3	6	6	8	23
Ispod medijana	8	7	6	2	23
	11	13	12	10	46

Očekivane frekvencije za svaku ćeliju iznose u ovom slučaju polovicu broja slučajeva u toj grupi. (Ali ako je medijan jedan od rezultata, onda obje desne sume u tablici nisu jednaki brojevi, pa očekivane frekvencije treba *izračunati* prema standardnom postupku. Očekivana frekvencija svake ćelije je suma reda · suma stupca, podijeljeno ukupnom sumom). Tako dobivamo ovaj hi-kvadrat račun:

f_o	f_t	$f_o - f_t$	$(f_o - f_t)^2$	$\frac{(f_o - f_t)^2}{f_t}$
3	5,5	−2,5	6,25	1,136
8	5,5	2,5	6,25	1,136
6	6,5	−0,5	0,25	0,038
7	6,5	0,5	0,25	0,038
6	6	0	0	0
6	6	0	0	0
8	5	3	9	1,800
2	5	−3	9	1,800
				$\chi^2 = 5,948$.

Broj stupnjeva slobode je (2 - 1) (4 - 1) = 3. Granična vrijednost χ^2 iznosi 7,815. Budući da je naš dobiveni χ^2 manji, prihvaćamo nul-hipotezu i zaključujemo da ne možemo smatrati da ti uzorci pripadaju populacijama s različitim medijanom.

N a p o m e n a. Za Prošireni medijan test vrijede pri izračunavanju hi-kvadrata ista pravila koja smo spomenuli na str. 268 u vezi s dopuštenim brojem ćelija, u kojima je očekivana frekvencija manja od 5. (Toga se pravila doduše ne pridržavaju svi autori.)

21.4.2. *Kruskal-Wallisov test*

Taj test zapravo predstavlja test analize varijance, samo se umjesto brojčanih mjernih podataka služi rangovima. On donekle predstavlja prošireni test sume rangova.

Postupak se kod tog testa može sažeti u ovih nekoliko operacija:

1. Svi se rezultati rangiraju, i to tako da najniži rezultat dobije rang 1.

Uzmemo li u račun naš prijašnji primjer iz Proširenog medijan testa, imamo ovu situaciju:

I. uzorak Rezultat	Rang	II. uzorak Rezultat	Rang	III. uzorak Rezultat	Rang	IV. uzorak Rezultat	Rang
8	1	10	2	16	7,5	22	14
12	3	15	6	16	7,5	41	23
13	4	19	10	17	9	49	33
14	5	25	15,5	21	12,5	54	36,5
20	11	30	18	29	17	59	39
21	12,5	38	21	33	19,5	60	40
25	15,5	40	22	44	25	65	42
33	19,5	45	27	45	27	69	44
43	24	47	30,5	46	29	71	45
45	27	48	32	53	35	75	46
47	30,5	51	34	62	41		
		54	36,5	67	43		
		55	38				
T_i	153,0		292,5		273,0		362,5
N_i	11		13		12		10

2. Izračunamo sume rangova u svakom uzorku (T_i). Broj rezultata u svakom uzorku označit ćemo s N_i. Dobivene brojeve korisno je kontrolirati: suma T_i mora iznositi:

$$\Sigma T_i = \frac{N(N+1)}{2}.$$

U našem primjeru dobivamo:

$$1\,081 = \frac{46 \cdot 47}{2} = 1\,081$$

3. Izračunamo izraz H prema formuli:

$$H = \frac{12}{N(N+1)} \sum \frac{T_i^2}{N_i} - 3(N+1), \tag{21.10}$$

gdje je:

T_i = suma rangova u jednom uzorku,
N = ukupan broj opažanja,
N_i = broj opažanja u jednom uzorku.

Ako kvadriramo svaki T_i i rezultat podijelimo korespondentnim N_i, dobivamo:

$$\frac{153^2}{11} = 2\,128,1,$$

$$\frac{292,5^2}{13} = 6\,581,2,$$

$$\frac{273^2}{12} = 6\,210,8,$$

$$\frac{362,5^2}{10} = 13\,140,6.$$

Zbrojimo li sve te vrijednosti, dobivamo:

$$\sum \frac{T_i^2}{N_i} = 28\,060,7.$$

Uvrstimo li vrijednosti u formulu (21.10), dobivamo:

$$H = \frac{12}{46 \cdot 47} \cdot 28\,060,7 - 3 \cdot 47 = 14,75.$$

4. Ako su uzorci dovoljno veliki (kod ovog se računa smatra da su uzorci dovoljno veliki ako svaki uzorak sadrži više od 5 rezultata), H ima jednaku distribuciju kao i hi-kvadrat, pa zato možemo značajnost očitati iz χ^2 tablice uz $k-1$ stupnjeva slobode (k = broj grupa).

Budući da naš broj stupnjeva slobode iznosi 3, a granična vrijednost $\chi^2 = 7,815$, morali bismo zaključiti da se uzorci statistički značajno *razlikuju*, tj. da *ne pripadaju* istoj populaciji.

Kao što se vidi, dok medijan testom nismo dokazali da uzorci pripadaju različitim populacijama, Kruskal-Wallisovim testom smo to dokazali. Budući da Kruskal-Wallisov test koristi više informacija od medijan testa (rangove umjesto jednostavnu podjelu rezultata u dvije skupine), on je "snažniji" od medijan testa.

Ako imamo veći broj zajedničkih rangova, a H je nešto ispod granice značajnosti, treba upotrijebiti drugu, korigiranu, formulu za izračunavanje H, u kojoj formulu (21.10) dijelimo s $1 - \dfrac{\Sigma T}{N(N^2-1)}$:

$$H = \frac{\dfrac{12}{N(N+1)} \sum \dfrac{T_i^2}{N_i} - 3(N+1)}{1 - \dfrac{\Sigma T}{N(N^2-1)}}, \qquad (21.11)$$

gdje je $T = n(n^2-1)$, a n = broj rezultata koji čine zajednički rang.

Taj račun treba posebno učiniti za *sve* zajedničke rangove: na primjer, ako imamo dva puta po 2 "vezana" ranga, dva puta po 3 vezana ranga i po jedanput 4, 5, 7 i 10 vezanih rangova, onda će vrijednost T iznositi:

$$\begin{array}{rlr}
2 & (2^2-1) = & 6 \\
2 & (2^2-1) = & 6 \\
3 & (3^2-1) = & 24 \\
3 & (3^2-1) = & 24 \\
4 & (4^2-1) = & 60 \\
5 & (5^2-1) = & 120 \\
7 & (7^2-1) = & 336 \\
10 & (10^2-1) = & 990 \\
\hline
 & \Sigma T = & 1\,566\,.
\end{array}$$

Taj će postupak nešto *povećati* vrijednost H.

U našem slučaju tu korekturu ne treba upotrijebiti, jer je 1. vrijednost H već ionako veća od granične vrijednosti, i 2. zbog relativno malog broja zajedničkih rangova konačan rezultat bio bi samo neznatno promijenjen (možda tek u drugoj decimali).

Kada broj mjerenja u pojedinim uzorcima nije dovoljno velik (dakle kad se grupe sastoje od 5 ili manje rezultata, H se ne može interpretirati kao χ^2, i u tu svrhu postoje posebne tablice iz kojih se može očitavati vjerojatnost. Te se tablice mogu naći u nekim priručnicima neparametrijskih metoda.

21.5. VIŠE ZAVISNIH UZORAKA

21.5.1. *Friedmanov test*

Ako na *istoj* grupi ispitanika vršimo mjerenje u *različitim* uvjetima, onda su rezultati dobiveni u svakom od tih uvjeta u *korelaciji* s ostalim rezultatima, pa se zbog toga više ne možemo služiti Kruskal-Wallisovim testom.

U tom slučaju Friedmanov test "dvostruke analize varijance rangova" predstavlja vrlo korisnu i upotrebljivu metodu kojoj u parametrijskoj statistici odgovara "dvostruka analiza varijance", a koja se upotrebljava, između ostaloga, i pri testiranju razlika između aritmetičkih sredina više zavisnih uzoraka.

Postupak Friedmanova testa sastoji se u tome da se rezultati najprije razvrstaju u tablicu s N redova i k kolona. Redovi odgovaraju pojedinim ispitanicima (ili grupama ispitanika), a kolone predstavljaju eksperimentalne uvjete. Rezultati u svakom *redu* (dakle za svakog ispitanika posebno) pretvore se u *rangove*. U slučaju jednakih rezultata, dobivamo naravno zajedničke rangove, ali to — prema Friedmanu — ne utječe na vrijednost testa.

Rangovi se u svakoj *koloni* (eksperimentalnoj situaciji) zbroje (T). Kada ne bi bilo razlika u rezultatima među uzorcima iz različitih eksperimentalnih uvjeta (tj. kad bi svi uzorci bili iz iste populacije), sume rangova tendirale bi sličnim vrijednostima. Ako se te sume značajno razlikuju, možemo odbaciti nul-hipotezu. Da bismo izmjerili relativnu veličinu tih razlika, zbrojit ćemo kvadrirane *sume* rangova (suma rangova = T_i), i nakon toga ćemo izračunati:

$$\chi_r^2 = \frac{12}{N\,k(k+1)} \Sigma(T_i)^2 - 3\,N(k+1). \qquad (21.12)$$

Ako su broj ispitanika (N) i broj eksperimentalnih uvjeta dovoljno veliki, izraz χ^2 ima približno jednaku distribuciju kao i hi-kvadrat sa $k-1$ stupnjeva slobode, pa stoga značajnost očitavamo iz tablice graničnih vrijednosti hi-kvadrata.

Primjer. Jedan je istraživač ispitivao kako na radni učinak utječe više odmora i je li u toku rada racionalnije uzeti jedan dulji odmor ili više kraćih odmora. Mjerio je ukupan radni učinak kod rada od 4 minute bez odmora (eksperimentalna situacija "a"), kod rada od ukupno 3 minute s jednim odmorom od 60 sek u sredini rada (eksperimentalna situacija "b") kod rada od ukupno 3 minute s 2 odmora od po 30 sek u toku rada (eksperimentalna situacija "c") i kod rada od ukupno 3 minute s 3 odmora od po 20 sek (eksperimentalna situacija "d"). Na ukupno 11 ispitanika dobio je ove rezultate, koje je za *svakog ispitanika posebno* pretvorio u rangove (rang je uz svaki rezultat označen u zagradi):

EKSPERIMENTALNE SITUACIJE

Ispitanici	a	b	c	d
1	991 (4)	1 157 (3)	1 232 (1)	1 217 (2)
2	1 139 (2)	1 055 (4)	1 057 (3)	1 173 (1)
3	762 (4)	775 (3)	931 (1)	890 (2)
4	1 074 (4)	1 121 (3)	1 220 (2)	1 260 (1)
5	544 (4)	596 (3)	655 (2)	671 (1)
6	765 (2)	728 (3)	840 (1)	637 (4)
7	904 (1)	839 (2)	746 (4)	774 (3)
8	862 (4)	916 (2)	881 (3)	1 157 (1)
9	725 (4))	886 (3)	925 (2)	992 (1)
10	1 079 (2)	894 (4)	1 130 (1)	1 009 (3)
11	833 (3)	844 (3)	890 (2)	963 (1)
T_i	35	33	22	20

Zbog kontrole treba izračunati sumu rangova:

$$\Sigma T_i = \frac{N\,k(k+1)}{2}$$

$$110 = \frac{44 \cdot 5}{2} = 110.$$

Izračunamo li zbroj kvadriranih suma rangova, dobivamo:

$$\Sigma T_i^2 = 35^2 + 33^2 + 22^2 + 20^2 = 3\,198.$$

Uvrstimo li dobivene vrijednosti u formulu (20.11), dobivamo:

$$\chi_r^2 = \frac{12}{44 \cdot 5} \cdot 3\,198 - 33 \cdot 5 = 9,44.$$

Uz $(k-1) = 3$ stupnja slobode, granična vrijednost χ^2 iznosi 7,815. Budući da je $9,44 > 7,815$, zaključujemo da uzorci *ne pripadaju* istoj populaciji, i da se zato statistički značajno razlikuju.

Ako su N i k mali, postoje posebne tablice za očitavanje značajnosti izraza χ_r^2, na razini značajnosti od 5% i 1%. Tablica R u Dodatku daje vjerojatnosti povezane uz različite χ_r^2, za slučajeve kada je $k = 3$ i kada je $k = 4$.

Kad bismo, na primjer, mjereći 6 ispitanika u 3 eksperimentalne situacije dobili $\chi_r^2 = 4,30$, zaključili bismo da ne možemo odbaciti nul-hipotezu (tj. da ne možemo dokazati da se uzorci značajno razlikuju), jer iz tablice se vidi da bi na razini značajnosti od $P = 0,05$ granična vrijednost χ_r^2 morala iznositi barem 6,4.

Zanimljivo je da Friedmanov test — iako koristi jedino rangove, a ne stvarne izmjerene vrijednosti — ima gotovo jednaku "snagu" kao i analiza varijance zavisnih uzoraka. Izračunamo li primjer s prethodne stranice uz pomoć "analize varijance zavisnih rezultata" dobit ćemo $F = 3,69$, što je značajno na razini $P < 0,05$.

Za ilustraciju gotovo jednake "snage" Friedmanova testa i parametrijskog F testa navodimo rezultate 56 nezavisnih analiza, u kojima je Friedman usporedio značajnost dobivenu F testom i Friedmanovim testom (vidi tablicu 21.1).

TABLICA 21.1.

Vjerojatnost χ_r	Vjerojatnost F: Veća od 0,05	Između 0,05 i 0,01	Manja od 0,01	Ukupno
Veća od 0,05	28	2	0	30
Između 0,05 i 0,01	4	1	4	9
Manja od 0,01	0	1	16	17
Ukupno	32	4	20	56

Kako se iz tablice vidi, slaganje je vrlo veliko.

21.5.2. *Cochranov Q test*

Ako na istoj skupini ispitanika (ili na različitim skupinama, ali one moraju biti "mečovane", tj svaki ispitanik u svakoj skupini mora imati svog "dvojnika" u drugim skupinama koji mu je u svim relevantnim faktorima veoma sličan) vršimo mjerenje u različitim uvjetima (kao što je npr. bio slučaj u Friedmanovu testu u poglavlju 21.5.1.) , ali su rezultati ispitanika *dihotomni*, tj. svrstani u samo dvije kategorije (pao-prošao, zdrav-bolestan, i sl.), onda je za testiranje postoje li razlike između pojedinih situacija pogodan Cochranov Q test. Tim testom zapravo se testira razlika između proporcija neke karakteristike u različitim uvjetima.

Pretpostavimo da je 20 studenata polagalo tri ispita, i da smo na svakom ispitu registrirali je li student prošao ili pao, pa nas zanima postoji li razlika u proporciji uspješnosti polaganja tih ispita. Na donjoj tablici pokazani su rezultati (+ = prošao. − = pao):

Studenti	Ispiti a	b	c	X_R	X_R^2
1	+	+	+	3	9
2	+	−	+	2	4
3	+	−	−	1	1
4	−	−	−	0	0
5	+	+	−	2	4
6	+	+	−	2	4
7	+	−	−	1	1
8	+	−	−	1	1
9	−	−	+	1	1
10	−	+	+	2	4
11	+	+	−	2	4
12	+	−	+	2	4
13	+	+	−	2	4
14	+	−	−	1	1
15	+	+	+	3	9
16	+	−	−	1	1
17	+	−	−	1	1
18	+	−	+	2	4
19	+	+	−	2	4
20	+	+	−	2	4
Suma	17	9	7	33	65

Postupak za izračunavanje može se podijeliti u tri "koraka":

1. Nađi sumu svakog eksperimentalnog uvjeta (situacije) (dakle ΣX_1, ΣX_2, ΣX_3... itd. U našem slučaju to su ove sume: $\Sigma X_1 = 17$, $\Sigma X_2 = 9$, $\Sigma X_3 = 7$.

2. Sumiraj rezultate svakog ispitanika u svim eksperimentalnim uvjetima (situacijama), tj. izračunaj sumu za svaki *red* (ΣX_R) i tu sumu kvadriraj (ΣX_R^2).

3. Sumiraj oba *stupca*, tj. sumiraj stupac ΣX_R i stupac $(\Sigma X_R)^2$.

Formula za Q glasi:

$$Q = \frac{k-1[k\Sigma(\Sigma X)^2 - (\Sigma X_R)^2]}{(k\Sigma X_R) - \Sigma X_R^2} \qquad (21.13)$$

Pri čemu k = broj situacija (eksperimentalnih uvjeta)
U našem slučaju dakle imamo:

$$Q = \frac{2[3(17^2+9^2+7^2)-33^2]}{(3\cdot 33)-65} =$$

$$= \frac{2(1257-1089)}{99-65} = \frac{336}{34} = 9,88$$

Q test distribuira se približno kao i hi-kvadrat test, sa stupnjevima slobode $k-1$, pa u našem slučaju imamo 2 stupnja slobode. Granična vrijednost hi-kvadrata (uz nivo značajnosti od 5%) iznosi 5,99. Budući da je dobiveni Q veći, odbacujemo nul-hipotezu i zaključujemo da se rezultati ispita statistički značajno *razlikuju*. (To nam dakako ništa ne kaže kako takav rezultat valja *interpretirati:* možda je razlog razlikama u težini predmeta, možda u različitoj strogosti pojedinih nastavnika, možda u manjem interesu studenata za neki predmet, itd. Vjerojatno se čitalac sjeća da smo već u ovoj knjizi spomenuli da je ono što statistički račun pokaže veoma zanimljivo, ali da on ne pokazuje ono što je često *bitno:* vrlo je zanimljivo znati da se uspjeh ova tri ispita razlikuje, ali bitno bi bilo znati zašto se razlikuje!)

Kao što iz ovog primjera vidimo, Q test nam — slično kao i analiza varijance — daje podatak da rezultati različitih ispitivanih uvjeta ne pripadaju istoj populaciji, ali nam ne kaže *koji* se rezultati među sobom statistički značajno razlikuju. Ako bi nas to posebno zanimalo, mogli bismo (iako je to samo aproksimativna metoda) po dvije situacije međusobno testirati testom dvaju zavisnih uzoraka sličnog tipa, a to je test predznaka. (Vidi str. 333).

21.5.3. *Fergusonov test monotonije trenda*

U novije vrijeme Ferguson je razradio neparametrijske postupke za testiranje *trenda* u eksperimentima tipa opisanog kod Friedmanova testa. (Postoje i parametrijski testovi trenda, ali oni u ovom priručniku nisu spomenuti.) Nas, naime, može zanimati ne samo to da li se eksperimentalne situacije statistički značajno razlikuju, već i to postoji li određena pravilnost u porastu (ili padu) rezultata od jedne eksperimentalne situacije do druge. Tako bi, primjerice, u maloprije obrađenom primjeru uz pomoć χ^2_r testa bilo posve logično očekivati da — u slučaju da broj odmora *ima* utjecaja na radni učinak — taj utjecaj bude to veći,što je veći broj kraćih odmora u toku radnog vremena. Friedmanov test ne daje nam mogućnost da odgovorimo i na to pitanje trenda.

Fergusonov test prikazat ćemo ukratko, i to samo za slučajeve kada nema zajedničkih rangova, jer u takvim slučajevima (što je inače u eksperimentima ovog tipa vrlo rijetko) računanje je znatno kompliciranije. Zbog istog razloga nismo prikazali Fergusonov test monotonije trenda za nezavisne rezultate — računanje je znatno kompliciranije. U vezi s tim problemima upućujemo čitaoce na knjigu: Ferguson, "Statistical Analysis in Psychology and Education", 2. izdanje, McGraw-Hill, 1966.

Poslužit ćemo se našim rezultatima, obrađenim u Friedmanovu testu, a metodu ćemo iznijeti u 5 "koraka":

1. Rangiraju se rezultati svakog ispitanika posebno, za sve eksperimentalne situacije.

Kao što znamo, to je već u našem primjeru učinjeno.

2. Za svakog ispitanika izračuna se izraz S, koji se računa ovako: usporedi se svaki rang sa svakim (dakle imamo $N(N-1)/2$ usporedbi rangova za svakog ispitanika); ako je par rangova, koji se uspoređuje, u "prirodnom" redu (npr. 1,4), zabilježi se + 1, a ako je red izvrnut (npr. 4,1), zabilježi se −1. Rezultati se za svakog ispitanika zbroje.

Na donjoj tablici dan je prikaz rangova i vrijednosti S za svakog ispitanika.

Ispitanici	Rangovi				Minusi	Plusovi	S
1	4	3	1	2	−5	+1	−4
2	2	4	3	1	−4	+2	−2
3	4	3	1	2	−5	+1	−4
4	4	3	2	1	−6	0	−6
5	4	3	2	1	−6	0	−6
6	2	3	1	4	−2	+4	+2
7	1	2	4	3	−1	+5	+4
8	4	2	3	1	−5	+1	−4
9	4	3	2	1	−6	0	−6
10	2	4	1	3	−3	+3	0
11	4	3	2	1	−6	0	−6
						Suma svih $S = -32$	

Pokazat ćemo računanje S za prva da ispitanika:

Ispitanik 1

par $4:3 \to -1$
$4:1 \to -1$
$4:2 \to -1$ Ukupno ispitanik 1:
$3:1 \to -1$ $-5+1=-4.$
$3:2 \to -1$
$1:2 \to +1$

Ispitanik 2

par $2:4 \to +1$
$2:3 \to +1$
$2:1 \to -1$ Ukupno ispitanik 2:
$4:3 \to -1$ $-4+2=-2.$
$4:1 \to -1$
$3:1 \to -1$

3. Zbroje se sve vrijednosti S da bi se dobio izraz ΣS. U našem slučaju $\Sigma S = -32$.

4. Izračuna se izraz σ_s^2(to je varijanca distribucije uzoraka S) prema formuli:

$$\sigma_s^2 = \frac{k(k-1)(2k+5)}{18}, \qquad (21.14)$$

pri čemu je k = broj eksperimentalnih situacija, i dobiveni se izraz pomnoži s N (broj ispitanika) kako bi se dobila varijanca distribucije izraza ΣS. Drugi korijen iz toga izraza je standardna devijacija izraza ΣS.

Dakle, u našem primjeru imamo:

$$\sigma_s^2 = \frac{4 \cdot 3 \cdot (8+5)}{18} = 8,67.$$

$$\sigma_{\Sigma S}^2 = 8,67 \cdot 11 = 95,37;$$

$$\sigma_{\Sigma S} = \sqrt{95,37} = 9,77.$$

5. Izraz $|\Sigma S| - 1$ podijeli s izrazom $\sigma_{\Sigma S}$, i tako se dobije odstupanje u terminima normalne distribucije, dakle z :

$$z = \frac{-31}{9,77} = -3,17.$$

Ako je z veći od 1,96 (za razinu značajnosti od 5%), ili veći od 2,58 (za razinu značajnosti od 1%), odbacit ćemo nul-hipotezu.

Dakle, u našem slučaju zaključujemo da naši rezultati *pokazuju* statistički značajan trend ($P < 0,01$).

ZADACI ZA VJEŽBU

1. Donja dva niza podataka (X i Y) potječu iz dva uzorka. Testirajte "testom niza" postoji li među njima statistički značajna razlika.

X	Y
12	7
8	9
6	16
18	17
14	20
15	19
3	2
5	10
4	22
1	21

2. Dvije skupine ispitanika postigle su na jednom testu računanja ove rezultate:

X	Y
12	7
16	12
18	14
7	18
6	5
4	16
11	9
12	10
8	14
20	3
18	18
16	9
10	7
	4

Testirajte "medijan testom" pripadaju li oba uzorka istoj populaciji.

3. Jedan je dreser dresirao 20 lavova za neku točku. Lavovi su bili podijeljeni u dvije grupe: grupa A bila je nagrađivana u toku dresure, a grupa B nije. Dolje su prikazani dani potrebni za svakog lava da nauči točku:

A	78	95	82	69	111	65	73	84	92	110
B	121	132	101	79	94	88	102	93	98	127.

Da li se vrijeme dresure između obje te grupe statistički značajno razlikuje? Za provjeru upotrijebite "test sume rangova".

4. Neke nove pilule protiv debljanja iskušane su na 15 ljudi tijekom 3 tjedna. Navedene su njihove težine prije i poslije završene terapije. Uz pomoć "testa predznaka" provjerite jesu li pilule imale efekta.

Prije	Poslije
59,4	56,7
57,6	58,1
52,6	53,5
69,4	70,3
80,7	81,2
91,6	90,7
87,1	88,5
83,0	81,6
77,6	81,6
82,6	81,6
76,7	78,9
70,3	68,0
73,9	76,7
77,6	78,0
94,3	90,7

5. Deset tipkačica-početnica na prvoj stranici pisanja učinile su od 10 do 22 pogreške (stupac X); na drugoj stranici učinile su 4 - 18 pogrešaka (stupac Y). Može li se razlika smatrati statistički značajnom? Upotrijebite "Wilcoxonov test ekvivalentnih parova", a *također* i t-test ("metoda diferencije"), i izvedite iz rezultata zaključak.

X	Y
16	4
12	18
22	10
16	14
14	12
10	14
20	10
18	12
10	4
22	12

6. Na donje tri skupine podataka primjenom "Kruskal-Wallisova testa" provjerite pripadaju li uzorci istoj populaciji.

Rezultati se odnose na vrijeme reagiranja (u minutama) laboratorijskih miševa na neki preparat.

Skupine	A	B	C
	8,6	9,3	8,5
	10,5	9,6	9,1
	11,4	8,4	7,4
	9,4	10,1	6,9
	9,0	10,2	9,0
	10,8	7,6	8,8
		8,8	9,2
		7,6	

7. Osmorica ispitanika ispitivana su u 4 eksperimentalne situacije: ispitivana je količina upamćenog materijala nakon 4 različito duge pauze. Rezultati su dolje prikazani (broj u tablici označuje količinu upamćenog materijala).

Vremenski intervali

Ispit.	I	II	III	IV
1	4	5	9	3
2	8	9	14	7
3	7	13	14	6
4	16	12	14	10
5	2	4	7	6
6	1	4	5	3
7	2	6	7	9
8	5	7	8	9

Postoji li statistički značajna razlika između količine upamćenog materijala u te 4 eksperimentalne situacije? Upotrijebite "Friedmanov test".

8. Može li se za rezultate iz 7. zadatka tvrditi da postoji neki trend s obzirom na eksperimentalne situacije? Upotrijebite "Fergusonov test monotonije trenda".

Rješenja zadataka za vježbu

Poglavlje 4.

1. a) $\overline{X} = 4,7$ Medijan = 4,5 Domin. vrijednost = 8
   b) $\overline{X} = 5,0$ Medijan = 5,0 Domin. vrijednost = 5,0
   c) $\overline{X} = 17,5$ Medijan = 4 Domin. vrijednost = 4

   Aritmetička sredina u slučaju c) je neprikladna, jer ju je jedan ekstremni rezultat "navukao" na visoke vrijednosti.

2. Treba izračunati zajedničku aritmetičku sredinu. Ona iznosi 58,55.

3. $\overline{X} = 173,47$.

Poglavlje 5.

1. a) $s = 3,71$
   b) $s = 2,45$
   c) $s = 41,45$

2. $s = 5,94$; $V = 3,42\%$

3. a) $V = 78,9\%$
   b) $V = 49\%$
   c) $V = 236,9\%$

   Iako su i prva dva koeficijenta varijabilnosti visoka, treći je nenormalno visok zbog jednog aberantnog rezultata.

Poglavlje 7.

1. 2 "pisma" past će u 36/512 = 7% slučajeva.

2. a. 34,13%
   b. 2,28%
   c. 13,59%.

3. 65,6%.

Poglavlje 8.

1. a. oko 250
   b. oko 154
   c. oko 9
   d. oko 92
   e. oko 309
   f. oko 4
   g. oko 290
   h. oko 220
   i. oko 207
   j. 497-498.
2. Vjerojatnost iznosi 7% ($p = 0,07$).
3. Šansa je vrlo mala, tj. 0,35%.
4. Neće, jer ih u cijeloj grupi ima oko 40!
5. u 91,5 centil.
6. a. 51. centil
   b. 85.
   c. 31.
   d. 12.
   e. 3.
   f. 1.

Poglavlje 9.

1. $t = 2,16$; $P < 0,05$.
2. $t = 4,75$; $P < 0,05$.
3. $t = 2,76$; $P < 0,05$.
4. $t = 1,86$; $P > 0,05$, razlika nije statistički značajna.
5. Da, jer 95%-tne granice pouzdanosti aritmetičke sredine iznose 9,2 do 10,8.
6. Skokovi su u drugom pokušaju bili za 3,7 m kraći;

   $t = 6,08$; $P < 0,05$.

Poglavlje 11.

1. 99,7%-tne granice pouzdanosti su $0,06 \pm 3 \cdot 0,05 = 0,45$ do 0,75.
2. Ne može se tvrditi da uzorak nije reprezentativan. Njegova standardna pogreška iznosi 0,046, pa 95%-tne granice iznose 0,53 do 0,71, a to prelazi vrijednost 0,70.
3. $t = \dfrac{0,07}{0,03} = 2,33$; $P < 0,05$.
4. $N_1 = 0,33$ do 0,94; $N_2 = 0,55$ do 0,83; $N_3 = 0,59$ do 0,79.

Poglavlje 12.

1. 56,8%.
2. 28,3%.

Poglavlje 13.

1. $r = 0,67$;
2. $r = 0,85$;
3. $r = 0,93$;
4. $r = 0,80$; $\rho = 0,83$;
5. $\rho = 0,86$;
6. $r_{23\cdot 1} = 0,04$;
7. $N_{12,3} = 0,61$;
8. $R_{1,23} = 0,66$;
9. $W = 0,82$.
10. $r_{pb} = 0,67$.
11. $\overline{X}_x = 79,11$; $\overline{X}_y = 71,11$; $dif = 2,0$; zajednička $s = \sqrt{123,63}$; zajednička $s_{\overline{X}} = 2,62$; $s_{\overline{X}_1-\overline{X}_2} = 2,06$; $t = 0,97$, dakle isti rezultat kao i pomoću "metode diferencije".

Poglavlje 14.

1. $\tilde{Y} = 6 - 1X$.
2. a. $\tilde{Y} = 54,77 + 10,86X$;
   b. oko 57,4;
   c. 3,01.
3. a. 5,0;
   b. 4,2;
   c. 0,33.

Poglavlje 15.

1. Nastavnikova raspodjela statistički značajno odstupa od normalne raspodjele: hi-kvadrat = 33,0.
2. Ne; hi-kvadrat = 0,53.
3. Nije; hi-kvadrat = 0,93.
4. Razlika među gradovima statistički je značajna: hi-kvadrat iznosi 24,49.
5. a. Nije, hi-kvadrat = 4,24.

   b. Sada je razlika statistički značajna, hi-kvadrat = 42,4. Tu razliku uglavnom uzrokuju demokrati koji su relativno više protiv smrtne kazne od drugih.

6. Ta je distribucija statistički visoko značajno različita od Poissonove distribucije jer hi-kvadrat iznosi 531,11. (Ako niste dobili takav rezultat, prekontrolirajte jeste li spojili posljednja 4 razreda u jedan razred!)

Poglavlje 17.

1. Uzorak će sadržavati ove ljude: 29, 154, 7, 14, 63, 103, 75, 22, 59, 4 (drugi broj 154 je ispušten!).

Poglavlje 20.

1. $F = 14,487/3,258 = 4,45$; $P < 0,05$.

   Statistički su značajne jedino razlike između druge i treće te druge i četvrte aritmetičke sredine. Njihove F-vrijednosti iznose 10,4, odnosno 9,62, što je veće od granične vrijednosti 8,52.

2. $F = 1,03/4,76$; $P > 0,05$, dakle razlika nije značajna.

Poglavlje 21.

1. Ne, jer postoji čak 10 nizova.
2. Možemo smatrati da smo potvrdili nul-hipotezu, tj. da uzorci pripadaju istoj populaciji, $\chi^2 = 0,034$; $P > 0,05$.
3. Razlikuje se, $z = 2,04$; $P < 0,05$.
4. Pilule nisu statistički značajno djelovale. $P > 0,05$.
5. U Wilcoxonovu testu razlika je "upravo" značajna, jer P iznosi točno 0,05. — "Metodom diferencije" $t = 2,35$, a $P < 0,05$, što pokazuje veću "snagu" parametrijskog testa, odnosno potvrđuje pravilo: ako neparametrijski test (osim rijetkih iznimaka) pokaže da je razlika statistički značajna, onda će parametrijski to pogotovu pokazati.
6. Skupine se ne razlikuju statistički značajno: $\chi^2 = 5,2$: $P > 0,05$.
7. Razlika je statistički značajna: $\chi_r^2 = 9,45$; $P < 0,05$.
8. Ne možemo tvrditi da trend postoji: $z = 1,56$; $P > 0,05$.

Literatura

Avenoso, F.J. and *Cheifetz, Ph. M.*, Elementary Statistics Throught Problem Solving, Williams & Wilkins Co., 1974.

Balsley, H. L., Basic Statistics for Business and Economics, Grid Inc., 1978.

Besag. F.P. and *Besag. P. L.*, Statistics for the helping professions, 1985.

Blalock, H. M., Social Statistics, McGraw-Hill, 1960.

Blejec, M., Statistične metode za psihologe, Ljubljana, 1959.

Brown, F. L., Amos, J. R., Mink, O. G., Statistical Concepts, Harper & Row, 1975.

Bujas, Z., Uvod u metode eksperimentalne psihologije, Zagreb, 1967.

Bujas, R. i Bujas, Z., Dobivanje psihologijskih podataka i njihovo računsko obrađivanje, Zagreb, 1942.

Castle, W. M. and North, P. M., Statistics in Small Doses, Churchill Livingstone, 1995.

Chambers, E. G., Statistical Calculations for Beginners, Cambridge University Press, 1952.

Champion, D. J., Basic Statistics for Social Research, Chandler, 1970.

Castle, W. M. , Statistics in operation, Churchill Livingstone, 1979.

Clayton, Keith, N., An introduction to statistics for psychology and education, Charles E. Merrill Publ. Co. , 1984.

Clegg, F, Simple statistics, Cambridge Univ. Press, 1987.

Cohen, L. and *Holliday M.* , Statistics for education and physical education, Harper and Row, 1979.

Connor, L. R. and *Morrell, A. J. H.*, Statistics in Theory and Practice, Pitman, 1977.

Conover, W. J. , Practical nonparametric statistics, John Wiley and Sons, 1971.

Dayton, C. M. and *Stunkard, C. L.*, Statistics for Problem Solving, McGraw-Hill, 1971.

Dixon, W. J. and *Massey, F. J.*, Introduction to Statistical Analysis, McGraw-Hill, 1957.

Dowie, J. and *Lefrere, P.* (edit): Risk and chance, The Open University Press, 1980.

Downie, N. M. and *Heath, R. W.*, Basic Statistical Methods, Harper & Row, 1974.

Du Bois, Ph., An Introduction to Psychological Statistics, Harper & Row, 1965.

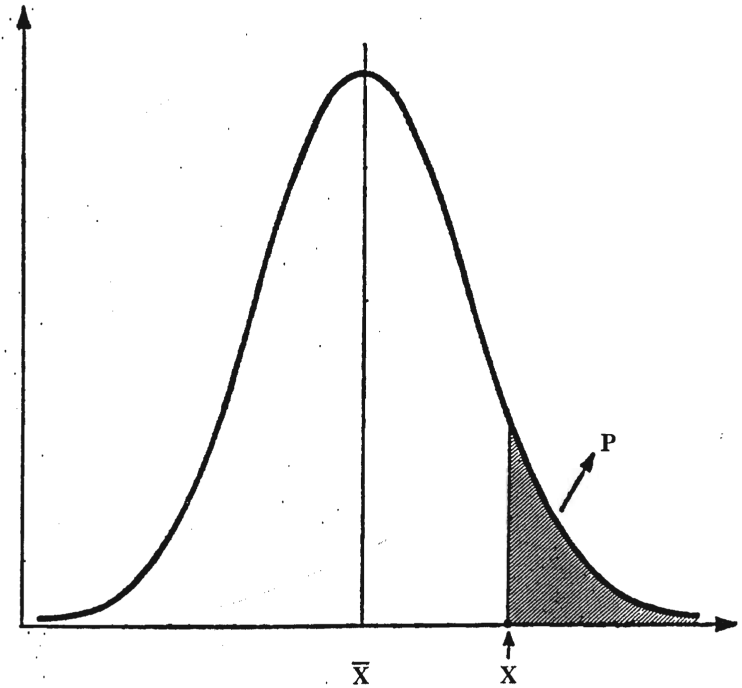
Dunn, Olive, J., Basic Statistics, a Primer for the Biomedical Sciences, John Wiley, 1964.

Edgington, E. S., Statistical Inference: The Distribution-free Approach, McGraw-Hill, 1969.
Edwards, A. L., Statistical Analysis, Rinehart & Co., 1958.
Ellis, R. B., Statistical Inference, Prentice-Hall, 1975.
Ferguson, G. A., Statistical Analysis in Psychology and Education, McGraw-Hill, 1966.
Freeman, L. C., Elementary Applied Statistics, John Wiley, 1965.
Freund, J. E., Modern Elementary Statistics, Prentice-Hall, 1973.
Freund, J. F. and Williams, F. J., Dictionary/Outline of Basic Statistics, McGraw-Hill, 1966.
Games, P. A. and *Klare, G. R.*, Elementary Statistics, McGraw- Hill, 1967.
Garrett, H. E., Statistics in Psychology and Education, Lingmans, Green, 1954.
Gilbert, Norma, Statistics, Saunders, 1976.
Graham, A., Statistics, Teach Yourself Books, 1994.
Guilford, J. P. and *Fruchter, B.* , Fundamental statistics in psychology and education, McGraw-Hill, 1987.
Hamburg, M. , Basic statistics: a modern approach, Harcourt Brace Jovanovich, 1979.
Hammerton, M., Statistics for the Human Sciences, Longman, 1975.
Hardyck, C. D. and *Petrinovich, L. F.*, Introduction to Statistics for the Behavioral Sciences, Saunders, 1969.
Harsbarger, Th. R. , Introductory statistics; a decision map, Macmillan, Collier, Macmillan, 1977.
Hayslett, H. T. , Statistics, made simple, Heinemann, 1981.
Henderson, N. K., Statistical Research Methods, Hong Kong Univ. Press, 1964.
Hill, A. B., Principles of Medical Statistics, The Lancet Ltd., 1952.
Hodges, J. L., *Krech, D.* and *Crutchfield, R. S.*, StatLab, McGraw-Hill, 1975.
Huff, D., How to Lie with Statistics, Penguin, 1974.
Huff, D., How to Take Chance, Penguin, 1965.
Ingram, J. A., Introductory Statistics, Cummings, 1974.
Kendall, M. G., Rank Correlation Methods, Ch. Griffin, 1955.
Kenny, D. A. , Statistics for the social and behavioral sciences, Litlle, Brown and Co, 1987.
Kohout, F. J., Statistics for Social Scientist, John Wiley, 1974.
Kolstoe, R. H., Introduction to Statistics for the Behavioral Sciences, Dorsey, 1969.
Koosis, D. J., Business Statistics, John Wiley, 1978.
Langley, R., Practical Statistics, Pan Books, 1968.
Leach, C. , Introduction to statistics. A nonparametric approach for the social sciences, John Wiley and Sons, 1979.
Leonard, M., Understanding Statistics, St Paul's House, 1971.
Levinw, G. , Introductory statistics for psyhology; the logic and methods, Academic Press, 1981.
Levinson, H. C. , Chance, luck and statistics, Dover Publications, 1963.
Lieberman, B., (edit.): Contemporary Problems in Statistics, Oxford Univ. Press, 1971.

Lindgren, B. W., Basic Ideas of Statistics, McMillan, 1975.
Lindquist, E. F., Statistical Analysis in Education Research, Houghton Mifflin, 1940.
Lindquist, E. F., A First Cours in Statistics, Houghton Mifflin, 1942.
Mainland, D., Elementary Medical Statistics, Saunders, 1952.
McCall, R. B., Fundamental Statistics for Psychology, Harcourt Brace Jovanovich, 1975.
McGuigan, F. J., Experimental Psychology, Prentice-Hall, 1961.
McNemar, Q., Psychological Statistics, John Wiley, 1969.
Mendenhall, W., Introduction to Probability and Statistics, Duxbury Press, 1975.
Minium, E. W., Statistical Reasoning in Psychology and Education, John Wiley, 1970.
Moroney, M. J., Facts from Figures, Penguin, 1970.
Mosteller, F. and *Rourke, R. E. K.*, Sturdy Statistics, Addison-Wesley, 1973.
Mueller, J. H. and *Schuessler, K. F.*, Statistical Reasoning in Sociology, Houghton Mifflin, 1961.
Naiman, A., *Rosenfeld, R.* and *Zirkel, G.*, Understanding Statistics, McGraw-Hill, 1972.
Newmark, J., Statistics and Probability in Modern Life, Holt, Rinehart and Winston, 1977.
Noether, G. E., Introduction to Statistics, a Fresh Approach, Houghton Mifflin, 1971.
O'Toole, A. L., Elementary Practical Statistics, Mcmillan, 1964.
Peatman, J. G., Descriptive and Sampling Statistics, Harper, 1947.
Peatman, J. G., Introduction to Applied Statistics, Harper, 1963.
Petz, B. , O nekim slučajevima kada su rezultati "predobri" da bi izgledali istiniti. Statisti, revija, 18 (1968) 129-137.
Petz, B. , "Jedanput na tisuću slučajeva" — što to ustvari znači. Primijenj. psihologija. 10 (1989) 73-75.
Petz, B., Statistika za praksu, MUP Rep. Hrv., 1994.
Phillips, J. L., Statistical Thinking, Freeman, 1973.
Pirc, B. i Milat, D., Osnove istraživanja u zdravstvu, Zagreb, 1970.
Quenouille, M. H., Introductory Statistics, Pergamon, 1966.
Reichmann, W. J., Use and Abuse of Statistics, Methneu, 1963.
Rowntree, J. , Statistics without tears — a primer for non-mathematitian, Penguin, 1981.
Runyon, R. P., Winning with Statistics, Addison-Wesley, 1977.
Runyon, R. P. and *Haber, A.*, Fundamentals of Behavioral Statistics, Addison-Wesley, 1967.
Runyon, R. P. , Nonparametric statistics. A contemporary approach. Addison-Wesley Publ. Co. , 1977.
Sanders, D. H., *Murph, A. F.* and *Eng, R. J.*, Statistics — a Fresh Approach, McGraw-Hill, 1976.
Senders, Virginia, Measurement and Statistics, Oxford Univ. Press, 1958.
Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, 1956.

Smith, G. M., A Simplified Guide to Statistics for Psychology and Education, Holt, Rinehart and Winston, 1970.
Snedecor, G. W. and *Cochran, W. G.*, Statistical Methods, Yowa State Univ. Press, 1967.
Sorić, B. i *Petz, B.* , Koliki postotak znanstvenih otkrića nisu otkrića? Arh. hig. rada, *38* (1987) 251-260.
Spence, J. T., Underwood, B. J., Duncan, C. P. and *Cotton, J. W.*, Elementary Statistics, Appleton-Century-Crofts, 1968.
Spiegel, M. R., Theory and Problems of Statistics, Schaum Publishing Co., 1961.
Sprent, P., Quick Statistics, Penguin, 1981.
Swoboda, H., Knaurs Buch der modernen Statistik, Droemer Knaur, 1971.
Šošić, I., i *Serdar, V.*, Uvod u statistiku, Školska knjiga, 1992.
Tate, M. W. and *Clelland, R. C.*, Nonparametric and Shortcut Statistics, Interstate Printers and Publishers, 1957.
Vesselo, I. R., How to Read Statistics, Harrap, 1962.
Vogt, W. P., Dictionary of Statistics and Methodology, Sage Publications, 1993.
Vouk, V. V., Statistika u medicini (skripta), Zagreb, 1956.
Walker, Helen, Elementary Statistical Methods, Holt, 1955.
Walker, Helen, and *Lev, J.*, Statistical Inference, Holt, 1953.
Walker, Helen and *Lev, J.*, Elementary Statistical Methods, Holt, 1958.
Wallis, W. A. and *Roberts, H. V.*, Statistic, a New Approach, The Free Press, 1958.
Walpole, R. E., Introduction to Statistics, Macmillan, Collier, 1974.
Weber, P., Elements of Statistics for Market Research, Gosby Lockwood, 1968.
Weinberg, G. H. and *Schumaker, J. A.*, Statistics, an Intuitive Approach, Brooks/Cole, 1969.
Welkowitz, J., Ewen, R. B. and *Cohen, J.*, Introductory Statistics for the Behavioral Sciences, Academic Press, 1971.
Wike, E. L., Data Analysis, Aldine/Atherton, 1971.
Williams, F. , Reasoning with statistics, Holt, Rinehart and Winston, 1979.
Woitschach, M., Moderne Matametik: Wahrscheinlichkeit und Zufall, Moderne Industrie, 1973.
Wright, R. L. D. , Understanding statistics, Harcourt Brace Jovanovich. 1976.
Zirkel, G. and *Rosenfeld, R.*, Beginning Statistics, McGraw-Hill, 1976.
Zuwaylif, F. H., General Applied Statistics, Addison-Wesley, 1972.

Dodatak



Ova slika odnosi se na Tablicu A na idućoj stranici.
O načinu očitavanja tablice i o smislu tog očitavanja vidi poglavlje 8.1!

TABLICA A

POVRŠINE ISPOD NORMALNE KRIVULJE

P =površina standardne normalne raspodjelne krivulje od zadane vrijednosti X do kraja krivulje. Površina cijele krivulje = 1

$$z = \frac{X - \overline{X}}{s}$$

z	P	z	P	z	P
0,00	0,5000	0,44	0,3300	0,88	0,1894
0,01	0,4960	0,45	0,3264	0,89	0,1867
0,02	0,4920	0,46	0,3228	0,90	0,1841
0,03	0,4880	0,47	0,3192	0,91	0,1814
0,04	0,4840	0,48	0,3156	0,92	0,1788
0,05	0,4801	0,49	0,3121	0,93	0,1762
0,06	0,4761	0,50	0,3085	0,94	0,1736
0,07	0,4721	0,51	0,3050	0,95	0,1711
0,08	0,4681	0,52	0,3015	0,96	0,1685
0,09	0,4641	0,53	0,2981	0,97	0,1660
0,10	0,4602	0,54	0,2946	0,98	0,1635
0,11	0,4562	0,55	0,2912	0,99	0,1611
0,12	0,4522	0,56	0,2877	1,00	0,1587
0,13	0,4483	0,57	0,2843	1,05	0,1469
0,14	0,4443	0,58	0,2810	1,10	0,1357
0,15	0,4404	0,59	0,2776	1,15	0,1251
0,16	0,4364	0,60	0,2743	1,20	0,1151
0,17	0,4325	0,61	0,2709	1,25	0,1056
0,18	0,4286	0,62	0,2676	1,30	0,0968
0,19	0,4247	0,63	0,2643	1,35	0,0885
0,20	0,4207	0,64	0,2611	1,40	0,0808
0,21	0,4168	0,65	0,2578	1,45	0,0735
0,22	0,4129	0,66	0,2546	1,50	0,0668
0,23	0,4090	0,67	0,2514	1,55	0,0606
0,24	0,4052	0,68	0,2483	1,60	0,0548
0,25	0,4013	0,69	0,2451	1,65	0,0495
0,26	0,3974	0,70	0,2420	1,70	0,0446
0,27	0,3936	0,71	0,2389	1,75	0,0401
0,28	0,3897	0,72	0,2358	1,80	0,0359
0,29	0,3859	0,73	0,2327	1,85	0,0322
0,30	0,3821	0,74	0,2296	1,90	0,0287
0,31	0,3783	0,75	0,2266	1,95	0,0256
0,32	0,3745	0,76	0,2236	2,00	0,0228
0,33	0,3707	0,77	0,2206	2,10	0,0179
0,34	0,3669	0,78	0,2177	2,20	0,0139
0,35	0,3632	0,79	0,2148	2,30	0,0107
0,36	0,3594	0,80	0,2119	2,40	0,00820
0,37	0,3557	0,81	0,2090	2,50	0,00621
0,38	0,3520	0,82	0,2061	2,60	0,00466
0,39	0,3483	0,83	0,2033	2,70	0,00347
0,40	0,3446	0,84	0,2005	2,80	0,00256
0,41	0,3409	0,85	0,1977	2,90	0,00187
0,42	0,3372	0,86	0,1949	3,00	0,00135
0,43	0,3336	0,87	0,1922	3,50	0,000233
				4,00	0,0000317

TABLICA B

GRANIČNE VRIJEDNOSTI t UZ ZADANI BROJ STUPNJEVA SLOBODE

Stupnjevi	P				
slobode	0,50	0,10	0,05	0,02	0,01
1	1,000	6,34	12,71	31,82	63,66
2	0,816	2,92	4,30	6,96	9,92
3	,765	2,35	3,18	4,54	5,84
4	,741	2,13	2,78	3,75	4,60
5	,727	2,02	2,57	3,36	4,03
6	,718	1,94	2,45	3,14	3,71
7	,711	1,90	2,36	3,00	3,50
8	,706	1,86	2,31	2,90	3,36
9	,703	1,83	2,26	2,82	3,25
10	,700	1,81	2,23	2,76	3,17
11	,697	1,80	2,20	2,72	3,11
12	,695	1,78	2,18	2,68	3,06
13	,694	1,77	2,16	2,65	3,01
14	,692	1,76	2,14	2,62	2,98
15	,691	1,75	2,13	2,60	2,95
16	,690	1,75	2,12	2,58	2,92
17	,689	1,74	2,11	2,57	2,90
18	,688	1,73	2,10	2,55	2,88
19	,688	1,73	2,09	2,54	2,86
20	,687	1,72	2,09	2,53	2,84
21	,686	1,72	2,08	2,52	2,83
22	,686	1,72	2,07	2,51	2,82
23	,685	1,71	2,07	2,50	2,81
24	,685	1,71	2,06	2,49	2,80
25	,684	1,71	2,06	2,48	2,79
26	,684	1,71	2,06	2,48	2,78
27	,684	1,70	2,05	2,47	2,77
28	,683	1,70	2,05	2,47	2,76
29	,683	1,70	2,04	2,46	2,76
30	,683	1,70	2,04	2,46	2,75
35	,682	1,69	2,03	2,44	2,72
40	,681	1,68	2,02	2,42	2,71
45	,680	1,68	2,02	2,41	2,69
50	,679	1,68	2,01	2,40	2,68
60	,678	1,67	2,00	2,39	2,66
70	,678	1,67	2,00	2,38	2,65
80	,677	1,66	1,99	2,38	2,64
90	,677	1,66	1,99	2,37	2,63
100	,677	1,66	1,98	2,36	2,63
125	,676	1,66	1,98	2,36	2,62
150	,676	1,66	1,98	2,35	2,61
200	,675	1,65	1,97	2,35	2,60
300	,675	1,65	1,97	2,34	2,59
400	,675	1,65	1,97	2,34	2,59
500	,674	1,65	1,96	2,33	2,59
1 000	,674	1,65	1,96	2,33	2,58
∞	,674	1,64	1,96	2,33	2,58

TABLICA C

F RASPODJELA (ZA TESTIRANJE RAZLIKE MEĐU VARIJANCAMA)

(Razina značajnosti = 5%)

stup. slob.₂ \ stup. slob.₁	1	2	3	4	5	6	7	8	9
1	648	800	864	900	922	937	948	957	963
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,12
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63
28	5,61	4,22	3,63	2,29	3,06	2,90	2,78	2,69	2,61
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11

TABLICA C

(nastavak)

10	12	15	20	24	30	40	60	120	∞
969	977	985	993	997	1 001	1 006	1 010	1 014	1 018
39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
14,4	14,3	14,3	14,2	14,1	14,1	14,0	14,0	13,9	13,9
8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14
4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
3,06	2,96	2,86	2,76	2,70	2,64	2,58	2,52	2,46	2,40
2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25
2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48
2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31
2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00

TABLICA D

GRANIČNE VRIJEDNOSTI KOEFICIJENTA KORELACIJE r

Razina značajnosti za dvosmjerno testiranje

Stupnjevi slobode $N-2$	P 0,05	P 0,01	Stupnjevi slobode $N-2$	P 0,05	P 0,01
1	0,997	1,000	24	0,388	0,496
2	0,950	0,990	25	0,381	0,487
3	0,878	0,959	26	0,374	0,478
4	0,811	0,917	27	0,367	0,470
5	0,754	0,874	28	0,361	0,463
6	0,707	0,834	29	0,355	0,456
7	0,666	0,798	30	0,349	0,449
8	0,632	0,765	35	0,325	0,418
9	0,602	0,735	40	0,304	0,393
10	0,576	0,708	45	0,288	0,372
11	0,553	0,684	50	0,273	0,354
12	0,532	0,661	60	0,250	0,325
13	0,514	0,641	70	0,232	0,302
14	0,497	0,623	80	0,217	0,283
15	0,482	0,606	90	0,205	0,267
16	0,468	0,590	100	0,195	0,254
17	0,456	0,575	125	0,174	0,228
18	0,444	0,561	150	0,159	0,208
19	0,433	0,549	200	0,138	0,181
20	0,423	0,537	300	0,113	0,148
21	0,413	0,526	400	0,098	0,128
22	0,404	0,515	500	0,088	0,115
23	0,396	0,505	1 000	0,062	0,081

TABLICA E1

GRANIČNE VRIJEDNOSTI KOEFICIJENTA KORELACIJE ρ

Razine značajnosti za jednosmjerno i dvosmjerno testiranje

N (broj parova)	Razina značajnosti za jednosmj. testiranje			
	0,05	0,025	0,01	0,005
	Razina značajnosti za dvosmj. testiranje			
	0,10	0,05	0,02	0,01
5	0,900	1,000	1,000	—
6	0,829	0,886	0,943	1,000
7	0,714	0,786	0,893	0,929
8	0,643	0,738	0,833	0,881
9	0,600	0,683	0,783	0,833
10	0,564	0,648	0,746	0,794
12	0,506	0,591	0,712	0,777
14	0,456	0,544	0,645	0,715
16	0,425	0,506	0,601	0,665
18	0,399	0,475	0,564	0,625
20	0,377	0,450	0,534	0,591
22	0,359	0,428	0,508	0,562
24	0,343	0,409	0,485	0,537
26	0,329	0,392	0,465	0,515
28	0,317	0,377	0,448	0,496
30	0,306	0,364	0,432	0,478

TABLICA E2

GRANIČNE VRIJEDNOSTI S U KOEFICIJENTU RANG-KORELACIJE τ

	P				
N	0,05	0,025	0,01	0,005	jednosmjerno
3	-	-	-	-	
4	6	-	-	-	
5	8	10	10	-	
6	11	13	13	15	
7	13	15	17	19	
8	16	18	20	22	
9	18	20	24	26	
10	21	23	27	29	

TABLICA F

GRANIČNE VRIJEDNOSTI KOEFICIJENTA KONKORDANCIJE W

m \ N	3	4	5	6	7	8	9	10
3	1,00	0,82	0,71	0,65	0,62	0,60	0,58	0,56
4	,81	,65	,54	,51	,48	,46	,45	,44
5	,64	,52	,44	,41	,39	,38	36	,35
6	,58	,42	,37	,35	,33	,32	,31	,30
7	,51	,36	,32	,30	,29	,27	,26	,26
8	,39	,32	,29	,27	,25	,24	,23	,23
9	,35	,28	,26	,24	,23	,22	,21	,20
10	,31	,25	,23	,21	,20	,20	,19	,18
12	,25	,21	,19	,18	,17	,16	,16	,15
14	,21	,18	,17	,16	,15	,14	,14	,13
16	,19	,16	,15	,14	,13	,12	,12	,12
18	,17	,14	,13	,12	,11	,11	,11	,10
20	,15	,13	,12	,11	,10	,10	,10	,09
25	,12	,10	,09	,09	,08	,08	,08	,07
30	,10	,09	,08	,07	,07	,07	,07	,06

TABLICA G

PRETVARANJE r U z_r

r	z_r	r	z_r	r	z_r	r	z_r	r	z_r
,000	,000	,200	,203	,400	,424	,600	,693	,800	1,099
,005	,005	,205	,208	,405	,430	,605	,701	,805	1,113
,010	,010	,210	,213	,410	,436	,610	,709	,810	1,127
,015	,015	,215	,218	,415	,442	,615	,717	,815	1,142
,020	,020	,220	,224	,420	,448	,620	,725	,820	1,157
,025	,025	,225	,229	,425	,454	,625	,733	,825	1.172
,030	,030	,230	,234	,430	,460	,630	,741	,830	1.188
,035	,035	,235	,239	,435	,466	,635	,750	,835	1,204
,040	,040	,240	,245	,440	,472	,640	,758	,840	1,221
,045	,045	,245	,250	,445	,478	,645	,767	,845	1,238
,050	,050	,250	,255	,450	,485	,650	,775	,850	1,256
,055	,055	,255	,261	,455	,491	,655	,784	,855	1,274
,060	,060	,260	,266	,460	,497	,660	,793	,860	1,293
,065	,065	,265	,271	,465	,504	,665	,802	,865	1,313
,070	,070	,270	,277	,470	,510	,670	,811	,870	1,333
,075	,075	,275	,282	,475	,517	,675	,820	,875	1.354
,080	,080	,280	,288	,480	,523	,680	,829	,880	1,376
,085	,085	,285	,293	,485	,530	,685	,838	,885	1,398
,090	,090	,290	,299	,490	,536	,690	,848	,890	1,422
,095	,095	,295	,304	,495	,543	,695	,858	,895	1,447
,100	,100	,300	,310	,500	,549	,700	,867	,900	1,472
,105	,105	,305	,315	,505	,556	,705	,877	,905	1,499
,110	,110	,310	,321	,510	,563	,710	,887	,910	1,528
,115	,116	,315	,326	,515	,570	,715	,897	,915	1,557
,120	,121	,320	,332	,520	,576	,720	,908	,920	1,589
,125	,126	,325	,337	,525	,583	,725	,918	,925	1,623
,130	,131	,330	,343	,530	,590	,730	,929	,930	1,658
,135	,136	,335	,348	,535	,597	,735	,940	,935	1,697
,140	,141	,340	,354	,540	,604	,740	,950	,940	1,738
,145	,146	,345	,360	,545	,611	,745	,962	,945	1,783
,150	,151	,350	,365	,550	,618	,750	,973	,950	1,832
,155	,156	,355	,371	,555	,626	,755	,984	,955	1,886
,160	,161	,360	,377	,560	,633	,760	,996	,960	1,946
,165	,167	,365	,383	,565	,640	,765	1,008	,965	2,014
,170	,172	,370	,388	,570	,648	,770	1,020	,970	2,092
,175	,177	,375	,394	,575	,655	,775	1,033	,975	2,185
,180	,182	,380	,400	,580	,662	,780	1,045	,980	2,298
,185	,187	,385	,406	,585	,670	,785	1,058	,985	2,443
,190	,192	,390	,412	,590	,678	,790	1,071	,990	2,647
,195	,198	,395	,418	,595	,685	,795	1,085	,995	2,994

TABLICA H

GRANIČNE VRIJEDNOSTI χ^2

Stup.	P												
slob.*	,995	,990	,975	950	.900	,750	,500	250	,100	,050	,025	,010	,005
1	,00004	,00016	,0010	,0039	.016	,102	,455	1,32	2,71	3,84	5,02	6,63	7,83
2	,0100	,0201	,0506	,103	,211	,575	1.39	2,77	4,61	5,99	7,38	9,21	10,6
3	,0717	,115	,216	,352	,584	1.21	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	,207	,297	,484	,711	1.06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	,412	,554	,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	,676	,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	,989	1.24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4.17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3.94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6.30	8,44	11,3	14,8	18,5	21.0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7.04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24.8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10.9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13.1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,4	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44.3	46,5
26	11,2	12,2	13,8	15,4	17,8	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27.3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53.7

* Ako je broj stupnjeva slobode veći od 30, značajnost se može testirati formulom:

$$\sqrt{(2 \cdot \chi^2)} - \sqrt{(2 \cdot \text{stup. slobode}) - 1}.$$

Ako taj izraz premašuje 1,65, χ^2 je značajan (na razini od 5%), tj. hipotezu treba odbaciti.

TABLICA I

POTREBNA FREKVENCIJA DA BI HI-KVADRAT BIO ZNAČAJAN NA 5% (obični brojevi) I 1% (debelo otisnuti brojevi) ZA SLUČAJEVE KADA JE $N_1 = N_2$

N	Najmanja frekvencija																									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
4	4	–	–																							
	–	–	–																							
5	4	5	–	–																						
	5	–	–																							
6	5	6	–	–																						
	6	–	–																							
7	5	6	7	–																						
	6	**7**	–	–																						
8	5	6	7	8	–																					
	6	**8**	**8**	–	–																					
9	5	6	8	8	9																					
	6	**8**	**9**	**9**	–																					
10	5	7	8	9	10	10																				
	7	**8**	**9**	**10**	–	–																				
11	5	7	8	9	10	11																				
	7	**8**	**9**	**10**	**11**	–																				
12	5	7	8	9	10	11	12																			
	7	**8**	**10**	**11**	**11**	**12**	–																			
13	5	7	8	9	10	11	12																			
	7	**9**	**10**	**11**	**12**	**13**	**13**																			
14	5	7	8	10	11	12	12	13																		
	7	**9**	**10**	**11**	**12**	**13**	**14**	**14**																		
15	5	7	9	10	11	12	13	14																		
	7	**9**	**10**	**11**	**12**	**13**	**14**	**15**																		
16	5	7	9	10	11	12	13	14	15																	
	7	**9**	**10**	**12**	**13**	**14**	**14**	**15**	**16**																	
17	5	7	9	10	11	12	13	14	15																	
	7	**9**	**11**	**12**	**13**	**14**	**15**	**16**	**16**																	
18	5	7	9	10	11	12	13	14	15	16																
	7	**9**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**17**																
19	5	7	9	10	11	12	14	14	15	16																
	7	**9**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**																
20	5	7	9	10	11	13	14	15	16	16	17															
	7	**9**	**11**	**12**	**13**	**15**	**16**	**16**	**17**	**18**	**19**															
30	6	8	9	11	12	13	15	16	17	18	19	20	21	22	23	24										
	8	**10**	**12**	**13**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**										
40	6	8	9	11	12	14	15	16	18	18	20	21	22	23	24	25	26	27	28	29	30					
	8	**10**	**12**	**14**	**15**	**17**	**18**	**19**	**20**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**32**					
50	6	8	10	11	13	14	15	17	18	19	20	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	8	**10**	**12**	**14**	**15**	**17**	**18**	**20**	**21**	**22**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
N_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

TABLICA J

POISSONOVA DISTRIBUCIJA

Proporcija slučajeva u razredu 0

$\overline{X}$	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0		,9900	,9802	,9704	,9608	,9512	,9418	,9324	,9231	,9139
0,1	,9048	,8958	,8869	,8781	,8694	,8607	,8521	,8437	,8353	,8270
0,2	,8187	,8106	,8025	,7945	,7866	,7788	,7711	,7634	,7558	,7483
0,3	,7408	,7334	,7261	,7189	,7118	,7047	,6977	,6907	,6839	,6771
0,4	,6703	,6637	,6571	,6505	,6440	,6376	,6313	,6250	,6188	,6126
0,5	,6065	,6005	,5945	,5886	,5828	,5770	,5712	,5655	,5599	,5543
0,6	,5488	,5434	,5380	,5326	,5273	,5221	,5169	,5117	,5066	,5016
0,7	,4966	,4916	,4868	,4819	,4771	,4724	,4677	,4630	,4584	,4538
0,8	,4493	,4449	,4404	,4361	,4317	,4274	,4232	,4190	,4148	,4107
0,9	,4066	,4025	,3985	,3946	,3906	,3867	,3829	,3791	,3753	,3716
1,0	,3679	,3642	,3606	,3570	,3535	,3499	,3465	,3430	,3396	,3362
1,1	,3329	,3296	,3263	,3230	,3198	,3166	,3135	,3104	,3073	,3042
1,2	,3012	,2982	,2952	,2923	,2894	,2865	,2837	,2808	,2780	,2753
1,3	,2725	,2698	,2671	,2645	,2619	,2592	,2567	,2541	,2516	,2491
1,4	,2466	,2441	,2417	,2393	,2369	,2346	,2322	,2299	,2276	,2254
1,5	,2231	,2209	,2187	,2165	,2144	,2123	,2101	,2081	,2060	,2039
1,6	,2019	,1999	,1979	,1959	,1940	,1921	,1901	,1882	,1864	,1845
1,7	,1827	,1809	,1791	,1773	,1755	,1738	,1721	,1703	,1686	,1670
1,8	,1653	,1637	,1620	,1604	,1588	,1572	,1557	,1541	,1526	,1511
1,9	,1496	,1481	,1466	,1452	,1437	,1423	,1409	,1395	,1381	,1367
2,0	,1353	,1340	,1327	,1313	,1300	,1287	,1275	,1262	,1249	,1237
2,1	,1225	,1212	,1200	,1188	,1177	,1165	,1153	,1142	,1130	,1119
2,2	,1108	,1097	,1086	,1075	,1065	,1054	,1044	,1033	,1023	,1013
2,3	,1003	,0993	,0983	,0973	,0963	,0954	,0944	,0935	,0925	,0916
2,4	,0907	,0898	,0889	,0880	,0872	,0863	,0854	,0846	,0837	,0829
2,5	,0821	,0813	,0805	,0797	,0789	,0781	,0773	,0765	,0758	,0750
2,6	,0743	,0735	,0728	,0721	,0714	,0707	,0699	,0693	,0686	,0679
2,7	,0672	,0665	,0659	,0652	,0646	,0639	,0633	,0626	,0620	,0614
2,8	,0608	,0602	,0596	,0590	,0584	,0578	,0573	,0567	,0561	,0556
2,9	,0550	,0545	,0539	,0534	,0529	,0523	,0518	,0513	,0508	,0503
3,0	,0498	,0493	,0488	,0483	,0478	,0474	,0469	,0464	,0460	,0455
3,1	,0450	,0446	,0442	,0437	,0433	,0429	,0424	,0420	,0416	,0412
3,2	,0408	,0404	,0400	,0396	,0392	,0388	,0384	,0380	,0376	,0373
3,3	,0369	,0365	,0362	,0358	,0354	,0351	,0347	,0344	,0340	,0337
3,4	,0334	,0330	,0327	,0324	,0321	,0317	,0314	,0311	,0308	,0305
3,5	,0302	,0299	,0296	,0293	,0290	,0287	,0284	,0282	,0279	,0276
3,6	,0273	,0271	,0268	,0265	,0263	,0260	,0257	,0255	,0252	,0250
3,7	,0247	,0245	,0242	,0240	,0238	,0235	,0233	,0231	,0228	,0226
3,8	,0224	,0221	,0219	,0217	,0215	,0213	,0211	,0209	,0207	,0204
3,9	,0202	,0200	,0198	,0196	,0194	,0193	,0191	,0189	,0187	,0185
4,0	,0183	,0181	,0180	,0178	,0176	,0174	,0172	,0171	,0169	,0167
4,1	,0166	,0164	,0162	,0161	,0159	,0158	,0156	,0155	,0153	,0151
4,2	,0150	,0148	,0147	,0146	,0144	,0143	,0141	,0140	,0138	,0137
4,3	,0136	,0134	,0133	,0132	,0130	,0129	,0128	,0127	,0125	,0124
4,4	,0123	,0122	,0120	,0119	,0118	,0117	,0116	,0114	,0113	,0112
4,5	,0111	,0110	,0109	,0108	,0107	,0106	,0105	,0104	,0103	,0102
4,6	,0101	,0100	,0099	,0098	,0097	,0096	,0095	,0094	,0093	,0092
4,7	,0091	,0090	,0089	,0088	,0087	,0087	,0086	,0085	,0084	,0083
4,8	,0082	,0081	,0081	,0080	,0079	,0078	,0078	,0077	,0076	,0075
4,9	,0074	,0074	,0073	,0072	,0072	,0071	,0070	,0069	,0069	,0068

TABLICA J

(nastavak)

Proporcija slučajeva u razredu 0

Ispred svakog broja u tablici nalaze se dvije nule

$\bar{X}$	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
5	6 738	6 097	5 517	4 992	4 517	4 087	3 698	3 346	3 028	2 739
6	2 479	2 243	2 029	1 836	1 662	1 503	1 360	1 231	1 114	1 008

Ispred svakog broja u tablici nalaze se tri nule

7	912	825	747	676	611	553	500	453	410	371
8	335	304	275	249	225	203	184	167	151	136
9	123	112	101							

Ispred svakog broja u tablici nalaze se četiri nule

9				914	827	749	677	613	555	502
10	454	411	372	336	304	275	249	225	204	185
11	167	151	137	124	112	101				

Primjer. Ako je aritmetička sredina $\bar{X} = 10,7$, proporcija slučajeva u razredu 0 (prvi razred u Poissonovoj seriji) je ,0000225.

TABLICA K

MALA TABLICA SLUČAJNIH BROJEVA

7766	7520	1607	6048	2771	4733	8558	8681	5204	3806
9627	5293	3569	0457	4426	2857	3666	9156	6931	6157
4594	2563	6826	8102	2543	4032	6897	2012	0945	0709
6668	4104	4018	4544	8117	7664	5270	3014	0420	4232
8874	0822	0949	8697	7550	4154	9697	9045	4916	1235
8009	5708	7072	8045	8451	5777	1613	0399	2069	7909
7271	5633	6025	0745	9804	3333	7160	5150	7743	5221
6450	6850	0602	9518	2275	9221	6441	8899	4640	7742
0598	0564	9655	3988	5620	3286	6319	6392	5743	1111
6546	4417	4453	5125	1356	6011	5965	9253	1486	7503
2806	6217	4278	3170	1626	1746	9731	9289	7667	5209
6901	9464	9302	6404	8049	3653	8101	4498	8558	6238
3625	0749	5025	7327	3984	1635	5963	0970	7357	2033
2222	9942	1706	2907	6304	8022	7972	7852	6242	6269
7224	3014	3943	5982	4052	4243	5306	1530	7537	3233
7160	6043	0767	0230	6082	3637	4556	6654	8972	9697
7965	7435	3397	9741	6207	2297	6491	7961	0243	6897
6708	0600	2765	1911	0813	2268	3554	7976	4102	0414
4159	6804	3838	4255	9664	7044	3067	6720	7416	4748
6592	1846	2269	9136	7107	0676	9782	8016	2715	3932
2805	7999	3743	1655	7812	7223	0954	4397	7427	9120
9501	0400	8056	4148	5585	7497	7421	0640	6695	6127
3346	6596	1997	9417	0164	9718	5671	9765	7091	1920
4447	3427	6134	9130	4763	2301	2892	4251	4491	5772
0610	4363	0705	0969	4684	4202	5274	6660	0468	1814
2131	4792	1418	0080	9763	7306	0167	9688	6959	2250
9569	9413	5681	9632	8505	8948	6475	2934	6046	9640
1412	7690	5615	1776	8568	7209	9907	3541	8847	8752
5064	7408	1951	1033	7817	2626	2441	3795	3275	1319
4193	2082	0412	5519	4108	3333	5546	0177	9345	5260
6414	5111	4003	3695	2976	4939	7555	7374	2913	2705
2672	8616	7005	5736	0172	7472	2033	6308	8779	1270
0758	3869	9288	2397	6264	8352	8617	7869	2459	8591
4502	2535	2434	5018	1202	9081	2674	2467	2532	9689
4823	3965	2801	6179	8592	6763	6567	1016	5801	9288
3011	0939	7162	4443	3849	9142	2922	9191	6029	7631
6611	9238	2160	9339	8177	2180	3905	2977	9234	3434
0378	8311	0623	4299	2335	7044	5855	0186	5895	5642
9905	4972	6907	5633	6548	3412	8469	0559	8878	8671
9424	4750	8325	3871	1831	7268	1863	9963	1905	7484
7004	3469	1159	4841	8681	8751	9214	1145	4394	1160
5658	2963	5798	4691	8653	7427	7826	9971	2622	9886
9327	2129	3459	1165	1011	4805	1821	7999	2136	9308
1161	2217	1797	3906	5304	4087	6766	3063	1747	3836
6002	3340	3648	3765	1565	8483	6353	8232	4942	5721
4311	3087	1756	6612	3277	1269	6573	3096	0898	1103
5237	1667	5941	2504	6213	5797	9326	3079	8796	4220
0163	7150	0894	9009	7858	4812	7678	0835	8447	1524
0437	7497	0187	4907	2202	2318	5339	3290	4342	9375
0974	9130	4974	9757	8802	8514	6564	5485	0793	5675

TABLICA K

nastavak

3754	7829	9473	8264	8502	0364	5146	0609	4708	5229
9278	1828	8171	8788	3821	0923	8249	8431	6516	0911
9152	6396	7516	2959	4988	0943	6070	8342	5643	7476
0306	8452	1326	8892	2571	4860	1907	4843	0248	5283
1775	3205	8496	0201	6864	3375	0599	7516	8592	9823
4448	1897	3406	1429	8153	3408	1136	9173	9582	2866
3406	4332	0083	1214	5107	0912	8257	4015	5933	5520
4869	7491	5786	3633	9450	4572	6046	7844	2536	9502
5042	6524	1138	4001	6957	7220	8715	5082	8909	2384
0371	1656	8756	3369	3347	3534	0519	7230	2516	2674
2969	0056	8199	9383	4840	4135	7713	6317	4188	8073
4680	0551	7807	9470	9460	2253	0146	6082	9037	1862
1979	1845	0247	4813	2052	2758	6032	8288	6840	2677
3463	7252	3753	1178	2766	3207	2332	8262	8499	4501
0698	8601	2945	6077	3785	4647	4226	8959	9006	0964
2709	2447	0580	3375	1775	2038	3797	5163	7845	9397
6014	1671	2362	2315	8297	3930	6686	5835	9464	0916
7219	3355	3933	9312	3808	7579	6254	7075	7818	0295
6900	7276	4131	5402	3263	4026	5185	2862	8450	7749
0652	9020	6533	5737	6390	8723	8240	6442	4775	6040
3559	8683	0358	0118	0825	3360	7913	1403	4016	0202
1133	5094	3564	9818	0188	6367	2887	5038	1039	1658
1066	2065	4018	9132	3343	6165	1351	1312	7876	8452
8099	2678	7288	1970	9523	4070	7258	7276	3138	6818
5599	5836	0212	7112	8857	5894	6647	1660	3518	5780
6204	6540	1791	3190	3727	4500	5370	5231	8629	6291
8288	1891	5014	8442	9712	3435	4570	9493	1563	9165
7590	9691	1601	6615	0848	2885	1863	5682	1666	3398
7162	9599	9286	2819	2867	6533	9931	9217	4987	7722
9948	6283	0839	4175	8654	2005	6128	1306	6879	3152
5187	9791	4301	8481	5699	2522	0394	1538	8492	1812
5330	8112	2323	3056	1282	0543	4135	5819	6172	1017
6454	8783	7254	5267	9809	9964	9835	1111	5988	8017
8771	0872	6538	9975	4349	4106	6047	9630	4211	3234
1804	3896	2518	5665	8766	7161	0755	0886	3256	3198
8109	0020	3347	9221	6511	7593	6133	6123	2128	2735
9371	0132	4794	3110	5357	7242	4790	8002	9268	9733
6062	6416	7311	1167	5131	9955	9738	6038	1119	4832
7072	3929	8902	8062	6898	5499	5278	3407	0544	8772
5867	5384	8700	8017	5235	4094	9441	2381	8478	0981
1390	8293	7525	7188	8218	0131	3543	1679	8610	5737
4974	9904	7964	6038	0910	9364	4842	3873	3495	5511
9086	9898	1529	8544	7800	8523	1353	3312	5255	3096
8786	4498	5476	6266	9636	1897	3924	7298	3764	0906
7215	2019	6780	1005	4812	0787	8463	3784	6072	0940
2701	2584	8904	7799	9877	9015	0310	9330	0037	8215
9830	7090	3878	7553	7460	2845	9183	6429	9249	0246
0008	1130	3811	1862	1670	6389	9179	8571	7621	2169
5338	0351	6437	6148	5015	6174	5761	4690	0799	3291
6508	4163	0794	5801	1272	2814	0989	1130	3918	8596

TABLICA L

F-RASPODJELA (ZA ANALIZU VARIJANCE):
granične vrijednosti F na razini značajnosti od 5%

stup. slob. brojnika / stup. slob. nazivnika	1	2	3	4	5	6	7	8	9
1	161	200	216	225	230	234	237	239	241
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,32	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,37	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

TABLICA L

(nastavak)

10	12	15	20	24	30	40	60	120	∞
242	244	246	248	249	250	251	252	253	254
19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
4,47	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,00	2,97	2,93
3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
2,98	2,91	2,84	2,77	2,74	2,70	2,66	2,62	2,58	2,54
2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
2,27	2,20	2,13	2,05	2,00	1,96	1,91	1,86	1,81	1,76
2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

TABLICA M

GRANIČNE VRIJEDNOSTI BROJA NIZOVA ZA TEST HOMOGENOG NIZA

$P = 0,05$

n \ m	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

TABLICA N

TEST SUME RANGOVA

Maksimalna vrijednost T ili T' (uzeti onaj broj koji je m a n j i) za manji uzorak u testu sume rangova (na razini od 5%)

N_1 *(manji uzorak)*

N_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3																				
4				(4) 10																
5			6	11	(5) 17															
6			7	12	18	(6) 26														
7			7	13	20	27	(7) 36													
8		3	8	14	21	29	38	(8) 49												
9		3	8	14	22	31	40	51	(9) 62											
10		3	9	15	23	32	42	53	65	(10) 78										
11		3	9	16	24	34	44	55	68	81	(11) 96									
12		4	10	17	26	35	46	58	71	84	99	(12) 115								
13		4	10	18	27	37	48	60	73	88	103	119	(13) 136							
14		4	11	19	28	38	50	62	76	91	106	123	141	(14) 160						
15		4	11	20	29	40	52	65	79	94	110	127	145	164	(15) 184					
16		4	12	21	30	42	54	67	82	97	113	131	150	169	190	(16) 211				
17		5	12	21	32	43	56	70	84	100	117	135	154	174	195	217	(17) 240			
18		5	13	22	33	45	58	72	87	103	121	139	158	179	200	222	246	(18) 270		
19		5	13	23	34	46	60	74	90	107	124	143	163	182	205	228	252	277	(19) 303	
20		5	14	24	35	48	62	77	93	110	128	147	167	188	210	234	258	283	309	(20) 337

(Nastavak na idućoj stranici)

TABLICA N

(nastavak)

TEST SUME RANGOVA

Maksimalna vrijednost T ili T' (uzeti onaj broj koji je m a nj i) za manji uzorak u testu sume rangova (na razini od 1%)

N_1 *(manji uzorak)*

N_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3																				
4																				
5					(5) 15															
6				10	16	(6) 23														
7				10	16	24	(7) 32													
8				11	17	25	34	(8) 43												
9			6	11	18	26	35	45	(9) 56											
10			6	12	19	27	37	47	58	(10) 71										
11			6	12	20	28	38	49	61	73	(11) 87									
12			7	13	21	30	40	51	63	76	90	(12) 105								
13			7	14	22	31	41	53	65	79	93	109	(13) 125							
14			7	14	22	32	43	54	67	81	96	112	129	(14) 147						
15			8	15	23	33	44	56	69	84	99	115	133	151	(15) 171					
16			8	15	24	34	46	58	72	86	102	119	136	155	175	(16) 196				
17			8	16	25	36	47	60	74	89	105	122	140	159	180	201	(17) 223			
18			8	16	26	37	49	62	76	92	108	125	144	163	184	206	228	(18) 252		
19		3	9	17	27	38	50	64	78	94	111	129	147	168	189	210	234	258	(19) 283	
20		3	9	18	28	39	52	66	81	97	114	132	151	172	193	215	239	263	289	(20) 315

TABLICA O

GRANIČNE VRIJEDNOSTI ZA TEST PREDZNAKA

N	1%	5%	10%	25%	N	1%	5%	10%	25%
1					46	13	15	16	18
2					47	14	16	17	19
3				0	48	14	16	17	19
4				0	49	15	17	18	19
5			0	0	50	15	17	18	20
6		0	0	1	51	15	18	19	20
7		0	0	1	52	16	18	19	21
8	0	0	1	1	53	16	18	20	21
9	0	1	1	2	54	17	19	20	22
10	0	1	1	2	55	17	19	20	22
11	0	1	2	3	56	17	20	21	23
12	1	2	2	3	57	18	20	21	23
13	1	2	3	3	58	18	21	22	24
14	1	2	3	4	59	19	21	22	24
15	2	3	3	4	60	19	21	23	25
16	2	3	4	5	61	20	22	23	25
17	2	4	4	5	62	20	22	24	25
18	3	4	5	6	63	20	23	24	26
19	3	4	5	6	64	21	23	24	26
20	3	5	5	6	65	21	24	25	27
21	4	5	6	7	66	22	24	25	27
22	4	5	6	7	67	22	25	26	28
23	4	6	7	8	68	22	25	26	28
24	5	6	7	8	69	23	25	27	29
25	5	7	7	9	70	23	26	27	29
26	6	7	8	9	71	24	26	28	30
27	6	7	8	10	72	24	27	28	30
28	6	8	9	10	73	25	27	28	31
29	7	8	9	10	74	25	28	29	31
30	7	9	10	11	75	25	28	29	32
31	7	9	10	11	76	26	28	30	32
32	8	9	10	12	77	26	29	30	32
33	8	10	11	12	78	27	29	31	33
34	9	10	11	13	79	27	30	31	33
35	9	11	12	13	80	28	30	32	34
36	9	11	12	14	81	28	31	32	34
37	10	12	13	14	82	28	31	33	35
38	10	12	13	14	83	29	32	33	35
39	11	12	13	15	84	29	32	33	36
40	11	13	14	15	85	30	32	34	36
41	11	13	14	16	86	30	33	34	37
42	12	14	15	16	87	31	33	35	37
43	12	14	15	17	88	31	34	35	38
44	13	15	16	17	89	31	34	36	38
45	13	15	16	18	90	32	35	36	39

TABLICA P

GRANIČNE VRIJEDNOSTI T ZA WILCOXONOV TEST EKVIVALENTNIH PAROVA

N	Razina značajnosti za "jednosmjerni" test		
	,025	,01	,005
	Razina značajnosti za "dvosmjerni" test		
	,05	,02	,01
6	0	—	—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

TABLICA R

GRANIČNE VRIJEDNOSTI χ_r^2 ZA FRIEDMANOV TEST

Broj stupaca k	Broj redova N	Minimalna vrijednost χ_r^2 potrebna za razinu značajnosti od: 5%	1%
3	3	5,8	—
	4	6,4	7,8
	5	6,2	8,3
	6	6,4	8,7
	7	6,1	8,7
	8	6,2	9,0
	9	6,2	8,7
	10	6,1	9,0
	11	6,2	9,0
	12	6,2	8,9
	13	6,0	9,0
	14	6,1	9,0
	15	6,2	8,9
	16 i više	5,99	9,21
4	2	5,95	—
	3	7,1	8,3
	4	7,6	9,4
	5	7,7	9,9
	6	7,5	10,2
	7	7,6	10,3
	8	7,6	10,4
	9 i više	7,81	11,34

TABLICA S

FAKTORIJELI

n	$n!$
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5.040
8	40.320
9	362.880
10	3,628.800
11	39,916.800
12	479,001.600
13	6,227,020.800
14	87,178,291.200
15	1,307,674,368.000
16	20,922,789,888.000
17	355,687,428,096.000
18	6,402,373,705,728.000
19	121,645,100,408,832.000
20	2,432,902,008,176,640.000

TABLICA T

BINOMNI KOEFICIJENTI $\frac{n!}{x!(n-x)!}$

n \ X	2	3	4	5	6	7	8	9	10
2	1								
3	3	1							
4	6	4	1						
5	10	10	5	1					
6	15	20	15	6	1				
7	21	35	35	21	7	1			
8	28	56	70	56	28	8	1		
9	36	84	126	126	84	36	9	1	
10	45	120	210	252	210	120	45	10	1
11	55	165	330	462	462	330	165	55	11
12	66	220	495	792	924	792	495	220	66
13	78	286	715	1287	1716	1716	1287	715	286
14	91	364	1001	2002	3003	3432	3003	2002	1001
15	105	455	1365	3003	5005	6435	6435	5005	3003
16	120	560	1820	4368	8008	11440	12870	11440	8008
17	136	680	2380	6188	12376	19448	24310	24310	19448
18	153	816	3060	8568	18564	31824	43758	48620	43758
19	171	969	3876	11628	27132	50388	75582	92378	92378
20	190	1140	4845	15504	38760	77520	125970	167960	184756