Vladimir Kolesarić
Jasmina Tomašić Humer

Veličina učinka

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FILOZOFSKI FAKULTET

Nakladnik
Sveučilište Josipa Jurja Strossmayera u Osijeku
Filozofski fakultet

Za nakladnika
prof. dr. sc. Loretana Farkaš

Autori
professor emeritus Vladimir Kolesarić
dr.sc. Jasmina Tomašić Humer

Recenzenti
dr. sc. Dinka Čorkalo Biruški, red. prof.
dr. sc. Dragutin Ivanec, izv. prof.

Lektor
Ivan Martinčić, prof.

ISBN 978-953-314-087-2

Objavljivanje ove sveučilišne skripte odobrio je Senat Sveučilišta Josipa Jurja Strossmayera u Osijeku br. 2/16 od 26. siječnja 2016.

Sadržaj

Uvod

Glavni cilj ovoga teksta jest približiti studentima (ali i psiholozima praktičarima i istraživačima) smisao i korist od dodatnih statističkih vrijednosti koje su se do sada relativno rjeđe koristile.

Dodatni je cilj pokazati da je uvijek dobro ponovno raspravljati o svemu, pa tako i o upotrebi statistike u psihologiji, zapravo, o njezinoj nedovoljno racionalnoj upotrebi. Nažalost, vrlo se često zaboravlja da je statistika samo alat i ništa drugo. Statistika ne odgovara na pitanja korisnika, ona samo pomaže u interpretaciji rezultata i štiti od vrlo pogrešnih i subjektivnih zaključaka. Ali, statistika nije svemoguća i – što se često zaboravlja – ima svojih nedostataka koje svakako valja poznavati da ne bismo napravili još veće i teže pogreške u zaključivanju.

Na primjer, „značajno“ i „statistički značajno“ nije isto. Statistički značajno ima svoju jasnu statističku podlogu u logici postavljanja i testiranja nul-hipoteze, ali to ne mora biti, a često i nije u skladu sa stvarnom praksom i značenjem. Poznati primjer: tjelesna visina ljudi ujutro i navečer statistički je značajno različita, ali nema nikakvo praktično značenje. Uvijek, dakle, treba imati na umu da statistika nije nedodirljiva „sveta krava“ i da stoga treba u svakom pojedinačnom slučaju razmisliti o primjerenoj upotrebi statističkih postupaka i o njihovoj interpretaciji u zadanom kontekstu. Statistika samo relativno objektivno potkrepljuje postavljene hipoteze, ništa više od toga.

Prelistavajući ne sasvim svježi broj časopisa *Theory & Psychology* naišli smo na članak koji se bavi veličinom učinka (Vacha-Hasse, Nilsson, Reetz, Lance i Thompson, 2000). Na početku toga članka autori su citirali naputak *Američke psihološke asocijacije*, objavljen 1994. godine, a koji se odnosi na statističku opremu članaka objavljivanih u časopisima te asocijacije. U tom se naputku ističe da sama *p* vrijednost pri statističkom testiranju nul-hipoteze nije dovoljno informativan pokazatelj djelotvornosti učinka koji upućuje na djelovanje nekog ispitivanog faktora. Međutim, autori u tome članku upozoravaju na to da se psiholozi tvrdoglavo drže stare prakse u kojoj nema podatka o veličini učinka.

Navodimo dva vrlo ilustrativna citata dvojice vrlo uglednih autora u području korištenja statistike u psihologiji.

G. V. Glass (1984): „Statistička značajnost je najmanje interesantan podatak o rezultatima. Treba opisati rezultate kao mjere veličine – ne samo da je neki tretman utjecao na ljude, nego koliko je utjecao“.

J. Cohen (1988): „Primarni produkt istraživanja su jedna ili više mjera veličine učinka, ne *p* vrijednosti“.

D. Lakens (2013) ističe da standardizirana veličina učinka daje razumljiv i upotrebljiv podatak bez obzira na skalu mjerenja zavisne varijable. Takva standardizirana veličina učinka omogućuje da se pokaže koje je praktično značenje dobivenih rezultata – tj. koje su praktične posljedice prikupljenih rezultata u svakidašnjem životu.

Što nam pokazuje testiranje nul-hipoteze?

Valja se prisjetiti: u postupku testiranja nul-hipoteze zanima nas razlikuju li se aritmetičke sredine dvaju uzoraka (dva skupa podataka) statistički značajno, tj. možemo li ostati na nul-hipotezi (H_0: nema razlike, odnosno, dobivena razlika je rezultat slučajnih čimbenika), ili nam statističke vrijednosti (npr. *t*-test) pokazuju da je razlika statistički značajna pa više ne ostajemo na nul-hipotezi. Jedan od početnih koraka u tom postupku jest i određivanje razine rizika, tj. donošenje odluke o vrijednosti alfa (α), koja pokazuje vjerojatnost odbacivanja nul-hipoteze kad je ona uistinu točna, ili vjerojatnost pogrešnosti odluke. Najčešće istraživači odabiru $\alpha = 0.05$ ili, u novije vrijeme, $\alpha = 0.01$. Odabirom α vrijednosti istraživač pokazuje do koje granice prihvaća rizik pogrešne odluke (da je odbacio nul-hipotezu, a ona je točna). Istraživači odabiru vrijednosti α vlastitom odlukom, koja se, dakako, treba temeljiti na nekim razlozima.

Vrijednost *p* jest vrijednost realno dobivene razine rizika koja može biti manja ili veća od odabrane vrijednosti α, a pokazuje vjerojatnost kojom je dobivena konkretna statistička veličina; ta vjerojatnost vrijedi ako je nul-hipoteza zaista točna. Ako u nekoj realnoj prilici dobijemo, recimo, $p = 0.007$, to će značiti da postoji vrlo mala vjerojatnost da je nul-hipoteza točna. Ako smo dobili, recimo, $p = 0.62$, to pokazuje kako postoji velika vjerojatnost da je H_0 točna.

Bez obzira na to do koje razine se u nekom istraživačkom postupku koristi statističko-matematička obrada rezultata, uvijek je prijeko potrebno najprije prikazati temeljne podatke i temeljne ishode osnovne statističke obrade rezultata, a među te temeljne podatke pripada, svakako, i veličina učinka. Ti temeljni podaci – osim aritmetičkih sredina, standardnih devijacija, minimalnog i maksimalnog postignutog rezultata – jesu određivanje intervala pomoću pogreške aritmetičke sredine u kojemu se nalazi prava aritmetička sredina i podaci o testiranju nul-hipoteze: veličina *t*-testa i *p*-vrijednost.

Navest ćemo nekoliko upozorenja koja se odnose na upotrebu statistike u psihologiji.

Počet ćemo s općim upozorenjem, koje se odnosi na vrlo raširenu upotrebu jednostavnijih statističkih postupaka. Statistika je, nedvojbeno, neobično korisna i u psihologijskim istraživanjima i u psihologijskoj praksi. Međutim, nažalost, mnogi psiholozi-istraživači, pa i psiholozi-praktičari, gledaju na statistiku kao na demijurga. Misle da je rezultat dobiven statističkom obradom „nedodirljiv“, bez mane, bez pogreške. No, mnogo je puta to, zapravo, skrivanje iza dobivenih statističkih vrijednosti: ako je statistička obrada pokazala da postoji statistički značajna razlika, onda je to tako i nikako drugačije, tome nema prigovora. Ili, ako rezultat statističke obrade rezultata pokaže da razlika među aritmetičkim sredinama dviju skupina ispitanika nije statistički značajna, uzima se kao definitivna činjenica da je nul-hipoteza točna.

No, svaki statistički rezultat treba uzeti *cum grano salis*. Poznati je primjer s ispitivanjem korisnosti aspirina u prevenciji srčanog napadaja. Vrlo velika skupina ispitanika podijeljena je po slučaju u dvije skupine: jedna skupina uzimala je jednu pilulu aspirina dnevno, a druga placebo. Zavisna varijabla bila je pojava srčanog napadaja: je li ga netko imao ili ga nije imao. Statistička obrada rezultata pokazala je da nema statistički značajne razlike u broju srčanih napadaja između skupine koja je uzimala aspirin i one koja je uzimala placebo. Izračunata je korelacija između zavisne i nezavisne varijable i ona je iznosila r = 0.034, što je objašnjavalo samo 0.12% varijance. Međutim, ipak dovoljno velik broj (iako ne statistički značajan) pripadnika skupine koja je uzimala aspirin nije imao napadaj, pa su istraživači ipak preporučili uzimanje aspirina (ako nema nekih idiosinkratičnih neželjenih nuspojava), tj. prekinuli su istraživanje i onima koji su uzimali placebo preporučili uzimanje aspirina (prema Rosnow i Rosenthal, 1989).

Testiranje nul-hipoteze, koje počiva na *t*-testu, postupak je koji daje dihotomni rezultat. Tu je vrijednost *p* ključna. Na temelju nje odlučujemo: ili ostajemo pri nul-hipotezi ili ne prihvaćamo nul-hipotezu i prihvaćamo našu radnu, prethodno postavljenu hipotezu. A *p* izravno ovisi o veličini uzoraka: što je veći *N*, to je automatski manje *p*, pa, prema tome, bez obzira na sve drugo, što je veći *N*, veća je vjerojatnost da će razlika među aritmetičkim sredinama biti statistički značajna. I obrnuto, što je manji *N*, razlika ima manju šansu da bude proglašena statistički značajnom. Psiholozi-statističari misle da takva strategija ne omogućuje napredak ni znanosti ni prakse.

Cohen (1994) upozorava na glavni problem testiranja nul-hipoteze. Istraživač je zainteresiran za odgovor na pitanje kolika je vjerojatnost da je nul-hipoteza točna. No, testiranje nul-hipoteze odgovara na to koliko je vjerojatno da se dobije određeni skup podataka ako je nul-hipoteza točna. A Nickerson (2000), pak, upozorava na to da je testiranje nul-hipoteze velika zabluda jer, gotovo redovito, istraživači na temelju dobivene *p* vrijednosti interpretiraju točnost ili vjerojatnost nul-hipoteze.

Prvo, ostajanje na nul-hipotezi ne mora značiti da je nul-hipoteza zaista točna. Neki, malo stroži statističari, tvrde da nul-hipoteza u zbilji nikada nije točna. Zaboravlja se, međutim, da nul-hipoteza ima ustvari samo radno obilježje jer je tako postavljena procedura testiranja značajnosti razlike.

Drugo, prihvaćanje radne hipoteze nema nikakve veze sa statističkim ishodom. Taj ishod samo kaže kolika je vjerojatnost pogreške, ali ništa ne govori o tome je li točna postavljena radna hipoteza ili je točna neka druga, alternativna, hipoteza. Točnost radne (alternativne) hipoteze ovisi samo o znanju i domišljatosti njezina autora.

Treće, istraživanje se poduzima zato da bi se saznalo ne samo postoji li djelovanje neke nezavisne varijable, nekoga čimbenika, nekoga tretmana, nego i zato da se utvrdi koliko je to djelovanje. Kako poznavanje veličine djelovanja nije nevažno pokazuje dobro poznati primjer s tjelesnom visinom ujutro i navečer: tjelesna visina navečer statistički je značajno manja nego ujutro, ali to, kako nam pokazuje svakodnevno iskustvo, nema nikakvih praktičnih posljedica.

Ali, zamislimo i ovakav primjer: želimo provjeriti neki novi psihologijski postupak, recimo za odvikavanje od pušenja, i to provjeravamo na dvije skupine pušača; na jednoj skupini primijenjujemo uobičajeni postupak, na drugoj neki novi postupak. Rezultat mjerimo pomoću dobro konstruiranog upitnika. I, recimo, da je u tom upitniku skupina s novim postupkom postigla statistički značajno bolji rezultat, tj. veći rezultati u upitniku pokazuju bolje odvikavanje od pušenja. Taj nam podatak u praktičnom smislu gotovo ništa ne znači; značit će nam podatak o tome ***koliko*** je ispitanika u skupini s novim postupkom postiglo bolje rezultate u upitniku od ispitanika u skupini s uobičajenim postupkom. Ako je od stotinjak ispitanika samo desetak postiglo bolje rezultate, onda usprkos značajnom rezultatu, nije opravdano (ni praktično ni teorijski) zaključiti da je nova metoda odvikavanja od pušenja uspješna. Kad bi barem pedesetak ispitanika imalo bolji rezultat, tada bi uvođenje nove metode imalo smislenog značenja.

Četvrto, kad se dobije statistički značajna razlika između dviju skupina ispitanika, prešutno se prihvaća da se – što je redovito vidljivo u daljnjem opisu i interpretaciji takva rezultata – svi ispitanici u tim skupinama međusobno razlikuju, a obično se razlikuje samo manji broj njih, jer se redovito distribucije međusobno prekrivaju.

Peto, proces i konačni ishod testiranja nul-hipoteze daje samo informaciju o tome ostajemo li na nul-hipotezi ili ne ostajemo, tj. kolika je vjerojatnost pogreške pri napuštanju nul-hipoteze ili kolika je vjerojatnost ostajanja uz nul-hipotezu.

Šesto, na kraju, što osobito vrijedi za praktični rad, zanimat će nas pojedinci, a ne samo cijela skupina. Podaci o cijeloj skupini kojoj neki pojedinac pripada, samo su kontekst za mjerenu varijablu, a on nam pokazuje gdje je, u odnosu na ostale članove skupine, smješten taj pojedinac.

Veličina učinka

Veličina učinka koristi se u ovim slučajevima: korelacija između dvije varijable; regresijski koeficijent (koji pokazuje nagib regresijskoga pravca); razlika između dvije aritmetičke sredine; jednostavna i složena analiza varijance; određivanje snage statističkoga testa; planiranje veličine uzorka; meta-analiza.

Veličina učinka jest procjena stupnja u kojemu je ispitivani fenomen prisutan, odnosno u kojem stupnju postoji u populaciji, ili stupanj u kojemu je nul-hipoteza nije točna. Može se reći i ovako: istraživanje, ili konačni rezultat istraživanja ima smisla ako razlika ili stupanj povezanosti dosegne određenu veličinu. To je podatak o tome koliki je bio učinak nezavisne varijable, a ne samo o tome je li postojao neki učinak ili to, dakle, nije samo binarni podatak. Veličina učinka može biti izražena kao razlika između dva populacijska parametra, kao odstupanje od neke konstante odnosno neke poznate vrijednosti te kao veličina povezanosti među varijablama (koeficijent korelacije) (Petz, Kolesarić i Ivanec, 2012).

Postoje barem, tri važna razloga zbog kojih treba odrediti *veličinu učinka* u psihologijskim istraživanjima, kao i u praktičnom radu psihologa (Vacha-Hasse i sur., 2000).

Prvo, *p* vrijednosti same za sebe ne mogu se upotrijebiti kao pokazatelj veličine učinka, jer su pod pomiješanim međusobno povezanim utjecajem mnogih svojstava istraživanja, uključujući i veličinu učinka i veličinu uzorka (*N*).

Drugo, malo vjerojatni rezultati (tj. rezultati s malom *p* vrijednošću) nisu nužno bitno interesantni u istraživanjima. Neki jako nevjerojatni događaji (npr. vjerojatnost da će netko umrijeti čitajući ovaj članak) ili pak jako vjerojatni događaji (npr. da će podloga na koju pada kiša biti mokra), dakle oni događaji koji se ekstremno konsekventno pojavljuju, zapravo su u statističkom smislu potpuno beznačajni. Dakle, test statističke značajnosti ne može se razumno upotrijebiti za razlikovanje važnosti rezultata hinjenjem objektivnosti. Tome se najbolje narugao J. Cohen u članku „The earth is round ($p < 0.05$)“ (1994) (Zemlja je okrugla, p < 0.05).

Treće, *p* vrijednosti ne upućuju izravno na kritično pitanje ponovljivosti (replikabilnosti) rezultata, jer statistički testovi ne testiraju vjerojatnost događanja rezultata uzorka u populaciji. Umjesto toga, statistička značajnost pretpostavlja da nul-hipoteza točno opisuje populacijske parametre (tj. populacijske aritmetičke sredine, populacijske koeficijente korelacije) i tako procjenjuje vjerojatnost rezultata uzorka (tj. aritmetičke sredine uzorka ili koeficijente korelacije uzorka) ili ekstremnijih rezultata u uzorku određene veličine, uz uvjet da su uzorci iz pretpostavljene populacije.

Veličina učinka može se koristiti kao dopuna i potkrepljenje statističkim testovima (kakav je npr. *t*-test ili analiza varijance) ili kao opća, generalizirana procjena veličine djelovanja neke nezavisne varijable ili nekoga tretmana ili, općenito, veličine djelovanja nekoga čimbenika.

Pokušat ćemo prikazati najčešće korištene indekse veličine učinka: što veličina učinka jest, kako se računa i kako se može interpretirati[1]. Određivanje snage statističkoga testa i s tim povezano planiranje veličine uzorka te meta-analiza posebne su teme pa se ovdje njima nećemo baviti.

Veličina učinka jednostavan je način kvantitativnog izražavanja razlike između rezultata dviju skupina ispitanika. Taj način, tj. veličina učinka ima neke prednosti u usporedbi s testiranjem statističke značajnosti, ili statističkog testiranja nul-hipoteze.

Većina autora koji govore o veličini učinka stavljaju taj indeks kao dopunu testiranju nul-hipoteze, uglavnom kada govore o *t*-testu, o analizi varijance ili o koeficijentu korelacije. R. Coe (2002), međutim, prikazuje veličinu učinka potpuno neovisno o testiranju nul-hipoteze pa o tom indeksu govori kao o samostalnoj statističkoj veličini.

d porodica veličine učinka

Razlikuju se dvije vrste odnosno dvije porodice veličine učinka. Jedna se temelji na razlici između statističkih vrijednosti rezultata (najčešće su to aritmetičke sredine) dviju ili više skupina ispitanika. Druga se temelji na asocijaciji, korelaciji među varijablama.

U prvoj porodici najčešće se koristi *d*-indeks, obično nazvan Cohenov *d*, jer potječe od J. Cohena. Formula za određivanje tog indeksa ima ovaj oblik:

$$d = \frac{M_E - M_K}{SD} \qquad (1)$$

M_E je aritmetička sredina „eksperimentalne“ skupine ispitanika, M_K je aritmetička sredina „kontrolne“ skupine, a SD je standardna devijacija. Ovdje je izrazom „eksperimentalna“ označena svaka skupina na kojoj je primijenjena neka nezavisna varijabla, neki tretman, neka varijabla za koju se pretpostavlja da bi mogla djelovati na neko ponašanje, na neku zavisnu varijablu koja je predmet mjerenja. Izrazom „kontrolna“ označena je svaka skupina ispitanika na kojoj nije ništa posebno

[1] Spominjemo kao kuriozitet da smo samo na jednom mjestu našli podatak da postoji između 50 i 100 mjera veličine učinka?! U ovom ćemo se tekstu, ipak, zadržati na njihovu puno manjem broju.

primijenjeno, skupina koja se nalazi u svojem relativno normalnom, uobičajenom stanju. Istraživač želi vidjeti ima li ta „eksperimentalna“ varijabla takvo djelovanje da se može razlikovati od uobičajene situacije – s obzirom na mjerenu varijablu, odnosno s obzirom na neki aspekt ponašanja koji je predmet opažanja ili mjerenja – u kojoj se nalazi „kontrolna“ skupina.

Budući da se u nazivniku koristi standardna devijacija *d*-indeks postaje standardizirana, usporediva vrijednost, neovisna o vrsti mjernih jedinica. Ali, postavlja se pitanje o tome koju standardnu devijaciju upotrijebiti. Logičnim se čini da to bude standardna devijacija kontrolne skupine jer je to „nedirnuta“ skupina s ispitanicima u uobičajenom stanju i u uobičajenim okolnostima. Može se, međutim, koristiti i zajednička standardna devijacija obje skupine, koja se lako izračuna po formuli:

$$SD_z = \sqrt{\frac{SD_1^2(N_1-1)+SD_2^2(N_2-1)}{(N_1-1)+(N_2-1)}} \qquad (2)$$

SD_z je zajednička standardna devijacija, a subskripti „1“ i „2“ odnose se, dakako, na dvije skupine kojima se proizvoljno oni pridjeljuju.

d-indeks je uvijek pozitivnoga predznaka, odnosi se, dakle, na veličinu razlike između dvije aritmetičke sredine. Općenito, *d*-indeks ne pokazuje pozitivnu ili negativnu razliku, nego jednostavno razliku.

Evo primjera. R. Coe (2002) prikazuje rezultate, koje je realno dobio jedan istraživač. Istraživačko pitanje bilo je: uče li učenici djelotvornije ujutro ili poslije podne? Skupina od 38 učenika po slučaju je podijeljena u dvije brojčano jednake podskupine. Jedna podskupina slušala je neku priču prije podne (u 9 sati), a druga tu istu priču poslije podne (u 15 sati). Razumijevanje sadržaja priče bilo je provjereno nizom od dvadeset pitanja (točan odgovor na svako pitanje nosio je 1 bod). Prosječni rezultat skupine koja je slušala priču prije podne iznosio je M_9 = 15.2, a prosječni rezultat skupine koja je slušala priču poslije podne M_{15} = 17.9. Razlika je, dakle, 2.7 bodova. Ovo je eksperiment u kojemu nema eksperimentalne i kontrolne skupine.

Standardna devijacija jutarnje skupine bila je SD_9 = 4.12, a poslijepodnevne SD_{15} = 2.10. Primjenom formule za zajedničku standardnu devijaciju (2) dobije se SD_z = 3.3. Ako razliku između aritmetičkih sredina 2.7 podijelimo s 3.3 dobit ćemo d = 0.82 (formula 1).

Uz odluku o tome koju ćemo standardnu devijaciju u ovom primjeru koristiti kao nazivnik pri određivanju veličine učinka potreban je određeni komentar.

Ako su razlike između standardnih devijacija velike, onda nije opravdano računati zajedničku (prosječnu) standardnu devijaciju. Što je velika razlika? Općenito statističari toleriraju odnos najviše do 1:3, tj. veća standardna devijacija ne smije biti više od tri puta veća od manje standardne devijacije. Veća se razlika ne prihvaća, a stroži statističari neće tolerirati ni taj odnos, već samo manji.

Ako se aritmetičke sredine znatno razlikuju, i distribucije rezultata dislocirane su na apscisi, pa bi zajednička distribucija bila neka bimodalna distribucija, a računanje standardne devijacije za takvu distribuciju nije opravdano, a ako bi se i računala, bila bi vrlo velika. Zato se nikada zajednička standardna devijacija ne računa iz tako skupljenih zajedničkih bruto rezultata, nego pomoću formule navedene pod (2).

Naglašavamo, polazi se od pretpostavke da je mjerenje (primjena nekoga testa, upitnika ili nekog drugoga psihologijskog mjerenja) obavljeno u obje skupine u prihvatljivo sličnim uvjetima. Eventualna razlika u aritmetičkim sredinama uzrokovana je nezavisnom varijablom primijenjenom u jednoj skupini (eksperimentalnoj), ili u obje skupine (ili više njih), ili primjenom nekoga tretmana u jednoj skupini, ili nekim drugim sustavnim čimbenikom koji je apliciran (ili prisutan) samo u jednoj skupini. Dakle, mjerenje u obje skupine (ili više skupina) mora biti obavljeno u prihvatljivo sličnim uvjetima (osim aplikacija nezavisne varijable, ili nekoga tretmana) da bi bilo opravdano uspoređivati ih.

U gornjem primjeru omjer standardnih devijacija iznosi 4.12 : 2.10 = 1.96, dakle razlika je podnošljiva. No, u ovom slučaju postoji jedna druga komplikacija. Prosječni rezultat u poslijepodnevnoj skupini iznosio je 17.9 od 20 mogućih bodova. Ovdje se, dakle, radi o tzv. stropnom efektu, pa je zbog njega raspršenje rezultata u toj skupini fizički smanjeno. (Usput govoreći, distribucija rezultata poslijepodnevne skupine sigurno je negativno asimetrična, tj. veće je grupiranje rezultata oko većih vrijednosti, bližih najvećoj mogućoj vrijednosti: 20.) Stoga je upitno je li opravdano uzeti standardnu devijaciju te skupine za određivanje zajedničke standardne devijacije.

Ipak, ako se ne raspolaže s većim brojem rezultata i ako standardne devijacije nisu jako različite, preporučuje se računanje zajedničke standardne devijacije zbog općeg statističkoga pravila da su izračunate statističke vrijednosti stabilnije ako su određene na temelju većeg broja rezultata. Za donošenje zaključaka na temelju statističkih vrijednosti jako je važno da su te vrijednosti što stabilnije.

Međutim, preporučuje se, u slučajevima kad je teško donijeti jasnu, nedvojbenu odluku, prikazati rezultate statističke obrade obavljene na različite načine. Tako je u spomenutom primjeru korektno prikazati i *d*-indeks izračunat pomoću zajedničke standardne devijacije i *d*-indeks izračunat sa standardnom devijacijom samo jutarnje skupine ispitanika. Ako se koristi samo standardna devijacija jutarnje skupine, koja iznosi 4.12, tada će *d*-indeks biti nešto manji: $d = 2.7 : 4.12 = 0.66$. Budući da je korištena veća standardna devijacija, razumljivo je da je *d*-indeks u tom slučaju manji.

Općenito, standardna devijacija selekcioniranih skupina – kao što su hospitalizirane osobe, recimo alkoholičari u klinici za odvikavanje od ovisnosti, ili pacijenti neke druge klinike, ili, recimo, izdvojeni učenici bilo zbog osobito dobrog ili osobito lošeg uspjeha u školi itd. – redovito je manja nego standardna devijacija neselekcionirane populacije.

Što nam izračunati *d*-indeks pokazuje, kakve nam informacije pruža? Odgovor na to pitanje vrlo je jasan: *d*-indeks nam daje informacije o dobivenim rezultatima kakve nam druge statističke obrade rezultata ne daju, a koje mogu biti vrlo korisne svakom istraživaču, i teoretičaru i praktičaru, a to je *prekrivanje distribucija dvije skupine rezultata*. Pokazuje koliko ispitanika, prema njihovim rezultatima u mjerenoj varijabli, doista pripada samo jednoj ili samo drugoj distribuciji.

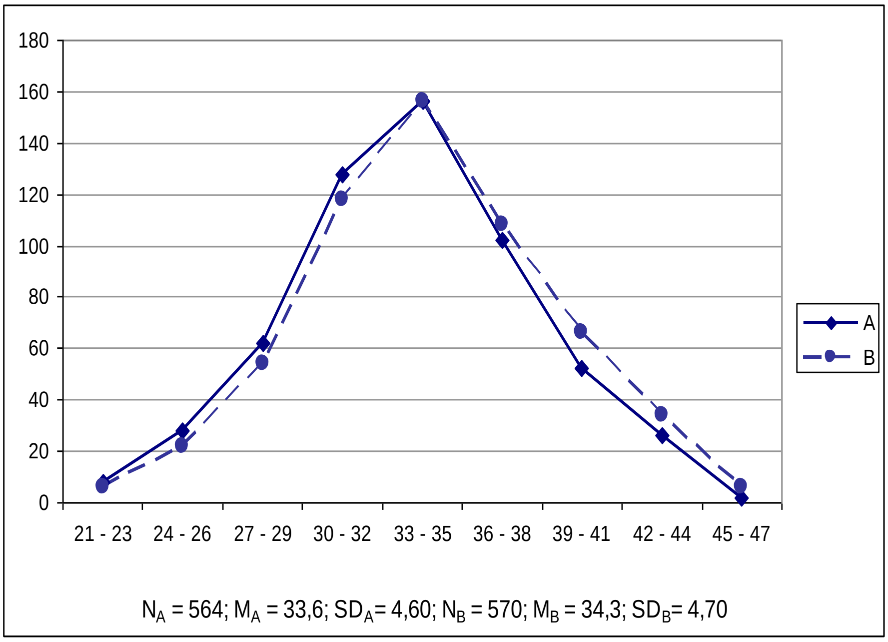
Pri testiranju nul-hipoteze pomoću, primjerice, *t*-testa može se utvrditi statistički značajna razlika, uz odabranu α vrijednost, i kad su razlike među aritmetičkim sredinama vrlo male, praktično zanemarive i kad se distribucije dviju skupina ispitanika u vrlo velikoj mjeri međusobno prekrivaju, kao na slici 1.

Na slici 1. prikazane su dvije distribucije rezultata dviju skupina učenika sedmog i osmog razreda u dvije osnovne škole, a rezultati su dobiveni primjenom perceptivnog testa.

Škola A: $M_A = 33.6$, $SD_A = 4.60$, $N_A = 564$

Škola B: $M_B = 34.3$, $SD_B = 4.70$, $N_B = 570$

$t = 2.53$, $p < 0.01$



Slika 1.
Distribucije rezultata dviju skupina učenika. Rezultati su dobiveni primjenom perceptivnog testa.

Prosječna bodovna razlika učenika te dvije škole manja je od 1 boda, samo 0.7, ali razlika je statistički značajna zbog velikoga broja rezultata. Ima li ta statistički značajna razlika ikakav stvarni smisao? Koliki broj učenika po svom rezultatu pripada samo jednoj ili samo drugoj distribuciji? Možemo li ustvrditi da su učenici škole B doista bolji u perceptivnom testu?

Budući da *d*-indeks ima, zapravo, smisao *z*-vrijednosti, omogućuje nam odrediti u kojem se postotku dvije distribucije prekrivaju.

Prema tzv. Cohenovoj konvenciji (koja je općenito prihvaćena) veličine *d*-indeksa imaju sljedeće značenje:

• *mala veličina učinka:* ***0.2*** *– prekrivanje distribucijâ je oko* ***85%***
• *srednja veličina učinka:* ***0.5*** *– prekrivanje distribucijâ je oko* ***67%***
• *velika veličina učinka:* ***0.8*** *– prekrivanje distribucijâ je oko* ***53%***

Može se interpretirati i ovako: kod d = 0.20 objašnjeno je 1% varijance, kod d = 0.50 objašnjeno je 10% varijance i kod d = 0.80 objašnjeno je 25% varijance.

Pogledajmo kako to izgleda u slučaju primjene perceptivnoga testa. Najprije ćemo odrediti zajedničku standardnu devijaciju prema formuli (2) i zatim ćemo pomoću formule (1) odrediti d. Tako ćemo dobiti d = 0.20, dakle, malu veličinu učinka. Prema

Cohenovoj konvenciji u tom se slučaju distribucije prekrivaju oko 85%. Prekrivanje distribucija na slici 1. čini se većim, međutim, treba kumulativno gledati neslaganja koje postoje i na lijevoj i na desnoj strani distribucije.

Očito je, dakle, da veličina učinka i statistička značajnost imaju različit smisao. Veličina učinka ne ovisi o tome je li neka razlika statistički značajna ili nije, a vrijedi i obrnuto.

Vratimo se primjeru s učenjem prije i poslije podne. U tom je slučaju, kako smo izračunali, d = 0.82 (kad smo koristili zajedničku standardnu devijaciju). Dakle, distribucije se, prema Cohenovoj konvenciji, prekrivaju nešto manje od 53%.

Prekrivanje, ili razlikovanje distribucija, možemo, koristeći činjenicu da *d*-indeks ima zapravo svojstva *z*-vrijednosti, prikazati i malo drugačije. Ako pogledamo u tablicu normalne distribucije (takva se tablica nalazi u svakom statističkom udžbeniku), u kojoj možemo očitati postotak slučajeva koji se nalaze u pojedinim segmentima normalne distribucije (u tablicama se čita postotak između aritmetičke sredine i određene *z*-vrijednosti), vidimo da je od *z*-vrijednosti 0.82 oko 79% rezultata niže (50%+29%). U tome primjeru to znači da je ispod aritmetičke sredine poslijepodnevne skupine (M_{15} = 17.9) oko 79% učenika prijepodnevne skupine.

Postoji tablica koja pokazuje, za određene vrijednosti *d*-indeksa, koliki je postotak ispitanika u skupini s nižom aritmetičkom sredinom ispod više aritmetičke sredine, i koliki je postotak *neprekrivanja* dviju distribucija (tablica 1.).

Cohenova konvencija pokazuje u kojem se postotku dvije distribucije prekrivaju, a u trećem stupcu tablice 1. nalazi se postotak *neprekrivanja*. Primjerice, prema Cohenovoj konvenciji za d = 0.20 prekrivanje distribucija je 85%. U tablici 1. za d = 0.20 neprekrivanje je distribucija 14.7%, ili okruglo 15%. To su, dakle, kongruentni podaci, a mogu se koristiti jedni ili drugi, ili i jedni i drugi, prema osobnoj preferenciji ili prema percepciji jasnoće prikazivanja dobivenih vrijednosti. Možemo gledati i ovako: u slučaju rezultata s perceptivnim testom samo je 58% (stupac „Postotak“ u tablici 1.) učenika u školi A ispod prosječne vrijednosti u školi B jer je u tom slučaju d iznosio samo 0.20.

Kad je d = 0, tada se dvije distribucije potpuno prekrivaju – ako su obje normalne i imaju jednake, ili barem podjednake, standardne devijacije. To je, uostalom, uvjet za korištenje tablice 1. – ona je napravljena prema normalnoj distribuciji. Ako ne možemo pretpostaviti da su dobivene distribucije normalne, ili je testiranje normaliteta distribucije pokazalo da statistički značajno odstupaju od normalne distribucije, onda će se postoci iz tablice razlikovati od stvarnih postotaka to više što dobivena distribucija više odstupa od normalne.

Tablica 1.

Veličine učinka (d) i korespondentni postoci rezultata u skupini s nižom aritmetičkom sredinom od više aritmetičke sredine, postotak neprekrivanja distribucija, koeficijent korelacije *r* te kvadrirani koeficijent korelacije (iz Becker, 2000)

Veličina učinka (d)	Postotak (%)	Postotak neprekrivanja	r	r^2
0.0	50%	0.0%	.000	.000
0.1	54%	7.7%	.050	.002
0.2	58%	14.7%	.100	.010
0.3	62%	21.3%	.148	.022
0.4	66%	27.4%	.196	.038
0.5	69%	33.0%	.243	.059
0.6	73%	38.2%	.287	.083
0.7	76%	43.0%	.330	.109
0.8	79%	47.4%	.371	.138
0.9	82%	51.6%	.410	.168
1.0	84%	55.4%	.447	.200
1.1	86%	58.9%	.482	.232
1.2	88%	62.2%	.514	.265
1.3	90%	65.3%	.545	.297
1.4	92%	68.1%	.573	.329
1.5	93%	70.7%	.600	.360
1.6	95%	73.1%	.625	.390
1.7	96%	75.4%	.648	.419
1.8	96%	77.4%	.669	.448
1.9	97%	79.4%	.689	.474
2.0	98%	81.1%	.707	.500

Ako je *d* veći od 2.0, jasno je da se tada mogu interpolirati vrijednosti koje se približavaju 100%.

U tablici 1. uz *d* = 0 stoji 0.50. Budući da se u tom slučaju distribucije potpuno prekrivaju i aritmetičke su sredine jednake, i ima 50% rezultata manjih od aritmetičke sredine. Kako se distribucije dviju skupina razdvajaju – što pokazuje povećanje *d*-indeksa – i distribucija se s većom aritmetičkom sredinom u koordinatnom sustavu na apscisi pomiče udesno, tako sve veći postotak rezultata u distribuciji s manjom aritmetičkom sredinom ostaje lijevo od veće aritmetičke sredine.

Statističari, ipak, upozoravaju da nije uputno koristiti veličine *d*-indeksa kao neke apsolutne vrijednosti, koje univerzalno vrijede. Valja ih koristiti samo u usporedbi s istovrsnim pojavama i mjerenjima, odnosno, u konkretnom značenju u nekom području. Katkad i mali *d*-indeks može upućivati na praktičnu korist. Primjerice, pri primjeni nekog novog terapijskoga postupka, recimo, za uklanjanje ili ublažavanje kronične glavobolje, ako se vidi poboljšanje kod samo nekoliko osoba, premda *d*-indeks nije velik, tim je osobama bolje. To znači da uvijek treba gledati dalje od samih statističkih vrijednosti.

Ali, vratimo se opet primjeru s prijepodnevnim i poslijepodnevnim učenjem. Ako se u određivanju veličine učinka (d) uzme zajednička standardna devijacija, dobije se d = 0.82. Međutim, ako se uzme standardna devijacija samo jutarnje skupine (SD_9 = 4.60) dobit ćemo d = 0.66 i, to je prema Cohenovoj konvenciji, veličina učinka između srednje i velike, a prekrivanje distribucija iznosi oko 60%. Razlika nije dramatična, ali nije ni zanemariva, pa se opravdano postavlja pitanje koji od ta dva *d*-indeksa treba koristiti. Budući da je distribucija u poslijepodnevnoj skupini najvjerojatnije asimetrična jer se radi o stropnom efektu, standardna devijacija izračunata iz takvih rezultata nije pravi pokazatelj raspršenja rezultata, ili, točnije rečeno, nije ju sasvim opravdano računati, pa ju je zato bolje ignorirati i uzeti samo standardnu devijaciju jutarnje skupine.

Još dvije napomene. Prvo, vidimo da numerička veličina *d*-indeksa nije ograničena, tj. najmanja je vrijednost nula kad su aritmetičke sredine potpuno jednake, a gornje granice nema. Numerička veličina *d*-indeksa ovisi o razlici između aritmetičkih sredina i o raspršenju rezultata u jednoj i drugoj varijabli. Ako su distribucije dvije skupine rezultata potpuno odvojene, uopće se ne prekrivaju i ne dodiruju se, tad će *d*-indeks biti velik, ovisno o tome koliko se razlikuju aritmetičke sredine tih sasvim razdvojenih distribucija. Treba li u takvom slučaju uopće računati d? Ako imamo podatak da su distribucije doista potpuno odvojene, određivanje *d*-indeksa ne donosi nikakvu novu informaciju, ali, dakako, može se računati (što bi zadrti statističari i zahtijevali).

Drugo, vidimo da se zapravo u dosadašnjem tekstu govorilo o izravnom određivanju *d*-indeksa, bez prethodnoga testiranja nul-hipoteze, odnosno utvrđivanja statičke značajnosti razlike među aritmetičkim sredinama. To je moguće, ali ima li smisla? Ako nam je, primjerice, *t*-test pokazao da nije opravdano napuštati nul-hipotezu, to znači da je dobivena razlika među aritmetičkim sredinama posljedica djelovanja slučajnih čimbenika, određivanje *d*-indeksa u tom slučaju ima dvojbenu statističku vrijednost. Određivanja *d*-indeksa, ima smisla tek kad smo utvrdili da glavni uzrok razlici među aritmetičkim sredinama nije slučajni varijabilitet. Tek kad vidimo da je razlika statistički značajna, na odabranoj razini rizika, α, odnosno da su osim slučajnih čimbenika razliku među aritmetičkim sredinama uzrokovali i neki sustavni čimbenici (npr. nezavisna varijabla ili primijenjeni tretman), ima smisla koristiti *d*-indeks koji će pokazati koliko je bilo to djelovanje, a što ne omogućuje sama primjena *t*-testa.

Jedno od važnih svojstava *d*-indeksa jest to da on nije ovisan o *N*-u. Još jedno njegovo važno svojstvo jest i to da se normalno distribuira. Ta činjenica omogućuje određivanje standardne devijacije te distribucije, pa onda i interval u kojemu se s određenom (unaprijed odabranom) sigurnošću nalazi dobivena vrijednost *d*. Standardna devijacija *d*-indeksa može se izračunati pomoću ove formule:

$$SD_d = \sqrt{\frac{N_1 + N_2}{N_1 \cdot N_2} + \frac{d^2}{2(N_1 + N_2)}} \tag{3}$$

N_1 i N_2 označavaju, kako je to uobičajeno, broj ispitanika ili broj rezultata u jednoj i drugoj skupini. Podatak o standardnoj devijaciji distribucije *d*-indeksa može se iskoristiti za određivanja intervala u kojemu se s određenom vjerojatnošću nalazi prava vrijednost *d*-indeksa. Ako dobivenu vrijednost SD_d pomnožimo s 1.96 (granična vrijednost vjerojatnosti za rizik od 5% odnosno za sigurnost od 95%) i taj umnožak pribrojimo i oduzmemo od dobivenoga *d*, dobit ćemo odgovarajući interval u kojemu se s vjerojatnošću od 95% nalazi pravi *d*-indeks, odnosno da se uz rizik od 5% nalazi izvan tog intervala. Ako tako dobiveni interval obuhvaća nulu, to znači da se *d*-indeks ne razlikuje od slučajne veličine. Dakako, možemo uzeti rizik od 1% ili 2%, prema vlastitoj odluci.

Na primjer, kad za eksperiment s jutarnjim i poslijepodnevnim učenjem izračunamo SD_d, dobit ćemo 0.34; taj broj pomnožimo s 1.96 i dobivamo 0.67, a taj iznos odbijemo i dodamo na 0.82 i tako dobijemo interval 0.15 ↔ 1.49. Dakle, u tom se intervalu s vjerojatnošću od 95% nalazi pravi *d*.

Ako je potrebno, tj. ako je prikazan samo *t*-test, možemo ipak odrediti *d*-indeks:

$$d = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (4)$$

Pogledajmo još jedan primjer. Problem: utječu li intenzivno čitanje literarnih i stručnih tekstova i raspravljanje o njima na rezultate u testovima inteligencije?

Napravljena su dva istraživanja; u jednom je istraživanju korišten verbalni test, a u drugom istraživanju neverbalni test. U oba su istraživanja dobiveni statistički značajni rezultati.

Prvo istraživanje: s verbalnim testom. Provedeno je na dvije skupine prethodno izjednačene po inteligenciji. Jedna skupina (A) tijekom dva tjedna intenzivno je čitala i raspravljala o pročitanim različitim tekstovima (literarnim i stručnim). Druga skupina (B) nije bila posebno angažirana u sličnoj aktivnosti. Nakon toga obje su skupine ispitane jednim verbalnim testom (70 zadataka; minimalni mogući rezultat 0; maksimalni mogući rezultat 70). Dobiveni su ovi rezultati:

A (eksperimentalna)	B (kontrolna)
$M_A = 38.1$	$M_B = 35.0$
$SD_A = 6.02$	$SD_B = 5.81$
$N_A = 78$	$N_B = 80$
$SD_{M_A} = 0.681$	$SD_{M_B} = 0.650$

$$t = \frac{38.1 - 35.0}{\sqrt{0.681^2 + 0.650^2}} = \frac{3.1}{0.9414} = 3.29 \quad p < 0.01$$

Drugo istraživanje obavljeno je na gotovo jednak način na drugim skupinama ispitanika, ali je korišten neverbalni test inteligencije (48 zadataka; minimalni mogući rezultat 0; maksimalni mogući rezultat 48) i bilo je nešto više ispitanika. Dobiveni su ovi rezultati:

A (eksperimentalna)	B (kontrolna)
$M_A = 25.5$	$M_B = 23.4$
$SD_A = 6.91$	$SD_B = 6.87$
$N_A = 158$	$N_B = 162$
$SD_{M_A} = 0.550$	$SD_{M_B} = 0.540$

$$t = \frac{25.5 - 23.4}{\sqrt{0.55^2 + 0.54^2}} = \frac{2.1}{0.771} = 2.72 \quad p < 0.01$$

Na temelju *t*-testova možemo zaključiti da postoji utjecaj čitanja i raspravljanja i na verbalni i na neverbalni test inteligencije: u oba slučaja dobivena je statistički značajna razlika između eksperimentalne i kontrolne skupine.

Odredimo sada *d*-indekse za oba dijela istraživanja. Koristit ćemo zajedničku standardnu devijaciju.

Zajednička SD za **prvo** istraživanje: $\overline{SD}_{prv} = \sqrt{\frac{6.02^2(77) + 5.81^2(79)}{78 + 80}} = 5.88$

Zajednička SD za **drugo** istraživanje: $\overline{SD}_{drug} = \sqrt{\frac{6.91^2(157) + 6.87^2(161)}{158 + 162}} = 6.87$

Pa će *d*-indeksi biti:

$$d_{prv} = \frac{38.1 - 35.0}{5.88} = 0.53$$

$$d_{drug} = \frac{25.5 - 23.4}{6.87} = 0.31$$

Na temelju provedenih *t*-testova možemo zaključiti samo to da i u verbalnom i u neverbalnom testu postoji statistički značajna razlika u prosječnim rezultatima skupina koje su čitale i koje nisu čitale. Veličine razlika u testovnim bodovima ne mogu se koristiti jer je bodovna skala za verbalni i neverbalni test različita. Zaključivanje o veličini razlike pomoću *p*-vrijednosti statistički, kako smo vidjeli, nije ni opravdano, ni dopušteno, a zapravo nije ni moguće.

Iz izračunatih *d*-indeksa vidimo da je taj zaključak o statističkoj značajnosti u oba slučaja nepotpun, čak bismo mogli reći površan, jer implicira podjednak utjecaj čitanja na rezultate u verbalnom i u neverbalnom testu.

Prema Cohenovoj konvenciji *d* od 0.31 pripada među male veličine učinka, a 0.53 ide prema većim veličinama učinka. Ako sada pogledamo u tablicu 1. vidjet ćemo da prekrivanje distribucija za $d = 0.31$ iznosi oko 78%, odnosno, neprekrivanje oko 22%. Za $d = 0.53$ prekrivanje je distribucija 65%, odnosno, neprekrivanje 35%. Taj nam podatak daje prilično drugačiju sliku o rezultatima provedenog eksperimenta.

Za slučaj zavisnih uzoraka (mjerenje obavljeno na istim ispitanicima u dvije situacije), poslužimo se ovim, ne sasvim izmišljenim, primjerom.

Vrijeme jednostavne senzomotorne reakcije vrlo često služi kao zavisna varijabla pri mjerenju doživljajnog intenziteta osjeta. Odnos između intenziteta podražaja i vremena reakcije zakrivljen je, ali u određenom opsegu postoji jasna nelinearna relacija: što je veći intenzitet podražaja, kraće je vrijeme reakcije. Mnogi eksperimenti pokazali su da postoji prilično jasan utjecaj podražajnoga konteksta na verbalne iskaze o jačini osjetnoga doživljaja. Postoji li utjecaj podražajnoga konteksta i na vrijeme reakcije?

U eksperimentu u kojemu je provjeravan utjecaj podražajnoga konteksta na vrijeme jednostavne senzomotorne reakcije sudjelovalo je 11 dobro uvježbanih muških mlađih ispitanika (to su, naravno, bili studenti), a njihova je zadaća bila da reagiraju što brže na zvučne podražaje različitih intenziteta odmakom kažiprsta s tipkala. Bila su zadavana dva niza od po četiri podražaja različitih intenziteta: jači niz i slabiji niz. Kritični podražaj bio je najslabiji u nizu jačih podražaja i najjači u nizu slabijih podražaja. Eksperimentatore je zanimalo hoće li vrijeme reakcije na taj podražaj biti dulje u jačem nizu od vremena reakcije na taj isti podražaj u slabijem nizu. Ako podražajni kontekst ima utjecaja na vrijeme reakcije, onda bi vrijeme reakcije moglo biti dulje na najslabiji podražaj u nizu jačih podražaja od vremena reakcije na podražaj tog istog intenziteta kad je on najjači u slabijem nizu podražaja.

Dobiveni rezultati prikazani su u tablici 2.

Tablica 2.
Vrijeme jednostavne senzomotorne reakcije (u milisekundama) 11 ispitanika na zvučni podražaj istog intenziteta (X) koji se nalazi u jačem nizu podražaja (VR_j) ili u slabijem nizu podražaja (VR_s).

VR_s	VR_j
169	174
156	162
173	177
178	176
170	172
181	180
167	171
160	168
159	160
164	167
166	165
$M_j = 167.6$	$M_s = 170.2$
$SD_j = 7.79$	$SD_s = 6.38$

U postupku „metode diferencija“, kojom ćemo utvrditi postoji li statistički značajna razlika u prosječnom vremenu reakcije na kritični podražaj, najprije smo odredili razliku D = VR_s – VR_j, zatim M_D = 2.6, potom SD_D = 3.17, onda $SD_{M_D} = 0.957$ i, napokon, $t = \frac{M_D}{SD_{M_D}} = 2.72$.

U tablicama za *t*-test nalazimo da je za stupnjeve slobode ss = 10 i α = 0.05 granična vrijednost 2.23. Naš *t*-test veći je pa ne ostajemo na nul-hipotezi (koja je glasila da se ne razlikuje vrijeme reakcije na kritični podražaj u podražajnim nizovima različitog intenziteta). Zaključujemo da postoji utjecaj podražajnoga konteksta na vrijeme jednostavne senzomotorne reakcije, ali na temelju *t*-testa ne možemo odgovoriti na pitanje koliki je taj utjecaj. Tu će nam pomoći *d*-indeks.

d-indeks za zavisne uzorke možemo odrediti prema formuli:

$$d = \frac{M_{poslije} - M_{prije}}{SD_{prije}} \qquad (5)$$

U našem slučaju nemamo „prije“ i „poslije“ pa je svejedno kojim ćemo redom postaviti aritmetičke sredine u gornjoj formuli. No, što ćemo sa standardnom devijacijom: koju ćemo od njih uzeti? To pitanje se postavlja zato što je u nacrtu ovog eksperimenta korištena rotacija: jedan ispitanik je najprije imao jači niz i zatim slabiji, a drugi najprije slabiji i zatim jači. Dakle, 5 ispitanika imalo je „prije“ jači niz, a 6 ispitanika imalo je „prije“ slabiji niz. Odgovor se nameće sam po sebi. Treba izračunati zajedničku standardnu devijaciju. Budući da je sudjelovao isti broj ispitanika, tj. isti ispitanici, u jednoj i drugoj eksperimentalnoj situaciji, zajedničku standardnu devijaciju izračunat ćemo jednostavno tako da zbrojimo varijance, zbroj podijelimo s 2 i iz tako dobivene numeričke veličine izvadimo drugi korijen. Bit će, dakle:

$$d = \frac{170.2 - 167.6}{\sqrt{\frac{7.79^2 + 6.38^2}{2}}} = 0.37$$

Ova veličina *d*-indeksa od 0.37 pokazuje da se distribucije **ne prekrivaju** samo oko 26% (tablica 1.). Iako je razlika statistički značajna, ipak ona nije vrlo velika i ne može se govoriti o utjecaju podražajnoga konteksta na vrijeme reakcije kao o nekakvoj izrazitoj pojavi.

Upozorenja i još neki indeksi d-tipa

Postoji i nekoliko specifičnijih upozorenja koja se odnose na korištenje *d*-indeksa.

1. Budući da je *d*-indeks standardizirana vrijednost neimenovanih jedinica, dakle, lišena konkretnih mjernih jedinica, katkada se neoprezno misli da to omogućuje usporedbu svega i svačega. Oni koji se koriste meta-analizom dobro znaju da je potreban veliki oprez pri skupljanju *d*-vrijednosti i pri njihovoj usporedbi.

2. Ima poteškoća s uzorcima ograničenoga ranga, tj. kad su to selekcionirane grupe koje imaju manji opseg rezultata u mjerenoj varijabli, jer su one samo izdvojeni dio populacije, kakve su, primjerice, darovita djeca ili djeca s poteškoćama u učenju, ili samo osobe muškoga spola ili samo osobe ženskoga spola, ili su dio populacije zbog nekog posebnog obilježja poput narkomana itd.

3. Potreban je oprez u interpretaciji kad se ne može pretpostaviti normalna distribucija ili kad distribucija dobivenih rezultata jako odstupa od normalne distribucije.

4. Oprez je potreban i kad nije dovoljno pouzdano mjerenje kojim su dobiveni rezultati.

5. Osobiti je oprez potreban pri usporedbi veličina učinaka koji se temelje na rezultatima dobivenim različitim vrstama mjerenja, ili na različitim operacionalizacijama istovrsnog ponašanja koje je predmet mjerenja, ili kad se temelje na različitim tretmanima u kliničkoj psihologiji ili pak pri usporedbi veličina učinaka dobivenih na različitim populacijama.

6. Izraz „učinak“ u sintagmi „veličina učinka“ implicira uzročno-posljedični odnos. Kako je poznato, ne radi se uvijek o uzročno-posljedičnu odnosu, već samo o veličini razlike, a veličina učinka samo omogućuje interpretaciju eventualnog uzročno-posljedičnog odnosa na temelju znanja o korištenim varijablama.

Ima slučajeva kad je korisnije zadržati se na nestandardiziranoj veličini učinka, tj. na samoj razlici u prosječnoj vrijednosti ili čestini pojavljivanja nečega. Primjerice, u nekom tretmanu za smanjenje pušenja, kao veličina učinka može se uzeti broj manje popušenih cigareta. Ako netko, ili skupina osoba, nakon tretmana dnevno popuši deset cigareta manje nego prije, onda je taj broj manje popušenih cigareta sam za sebe koristan pokazatelj veličinu učinka. Značenje toga smanjenja procjenjuje se prema tome je li ono pridonosi poboljšanju općega zdravstvenoga stanja ili ne pridonosi. Drugim riječima, tu veličinu učinka ne treba standardizirati i pretvarati je u neimenovanu vrijednost dijeljenjem s nekom pogreškom.

Standardizirana veličina učinka (a to je *d*-indeks i njemu slični indeksi) koristi se uvijek onda kada je mjerena varijabla izražena na nekoj manje-više arbitrarnoj brojčanoj ljestvici (kao što su redovito brojčane ljestvice dobivene primjenom testova ili upitnika), pa dobiveni brojevi nemaju sami za sebe neko značenje. Osobito se koristi u slučajevima u kojima se uspoređuju veličine učinka iz različitih istraživanja.

* * *

Poznati su još Glassov *delta*-indeks (označava se grčkim slovom delta: Δ) koji uzima samo standardnu devijaciju kontrolne skupine, dakle skupine na koju nije djelovala nikakva nezavisna varijabla jer se pretpostavlja da je ta standardna devijacija najbolja procjena populacijske standardne devijacije. Određuje se kao i *d*-indeks: razlika među aritmetičkim sredinama dijeli se sa standardnom devijacijom kontrolne skupine.

$$\Delta = \frac{M_1 - M_2}{SD_{kontrol.skupine}} \qquad (6)$$

Prednost je toga indeksa što ne postoji pitanje jednakosti (sličnosti) varijanci u različitim skupinama ispitanika, a to je osobito korisno kad postoji više eksperimentalnih skupina. Gledajući s druge strane, može se i ovako reći: budući da se redovito polazi od pretpostavke jednakosti varijanci, opravdanije je odrediti prosječnu varijancu dviju (ili više) skupina. Osim toga, u praktičnom radu, pa i u mnogim istraživanjima, često ne postoje kontrolna i eksperimentalna skupina, nego usporedne skupine koje su pod pretpostavljenim utjecajem različitih varijabli ili različitih razina iste varijable, npr. pri usporedbi skupine muških ispitanika i skupine ženskih ispitanika. Tada delta indeks nije primjenjiv.

Spominje se i Hedgesov *g*-indeks koji za standardnu devijaciju u nazivniku uzima sve standardne devijacije eksperimentalne i kontrolne skupine (odnosno svih skupina koje su u istraživanju) pa je tako pogodan za jednostavnu analizu varijance. Međutim, uz njega se veže pogreška koja se može ublažiti pomoću ove korekcije:

$$g^* = g\left[1 - \frac{3}{4(n_1 + n_2) - 9}\right] \qquad (7)$$

Ta korekcija osobito se preporučuje za male nezavisne uzorke. Analognom formulom može se korigirati i *d*-indeks.

r porodica veličine učinka

Druga porodica veličine učinka počiva na korelaciji, ili, kako to neki statističari kažu, na snazi asocijacije. Veličina učinka određena na temelju korelacije između dvije ili više varijabli razlikuje se od veličine učinka koji se temelji na razlici među aritmetičkim sredinama (*d*) jer predstavlja procjenu veličine zajedničke varijance među varijablama.

Sam koeficijent korelacije, Pearsonov *r*, predstavlja veličinu učinka, a njegov kvadrat (r^2) pomnožen sa sto, pokazuje postotak zajedničke varijance dviju varijabli. Jednako vrijedi i za multiplu korelaciju: *R* i R^2.

Ovdje je konvencija o interpretativnoj kvalifikaciji veličine učinka drugačija:
0.10 je mala veličina učinka, 0.30 srednja i 0.5 i više (do 1.0) velika veličina učinka.

Dakle, ako koeficijent korelacije ima veličinu $r = 0.10$, postotak zajedničke varijance, ili prekrivanje varijanci dviju varijabli, iznosit će 1%. Ako je korelacija $r = 0.50$, taj će postotak iznositi 25%. Tih 25% jest postotak zajedničke varijance koji dijele dvije varijable. Taj podatak naziva se i koeficijent determinacije i bilježi se s $r^2(100)$.

U radovima koji se bave tom veličinom učinka mogu se pronaći prijedlozi da se najmanja veličina učinka koja se temelji na *r* bude 0.20 ili čak 0.30 (a ne 0.10 kao u navedenoj konvenciji).

U tablici 1. (str. 13) u stupcima označenim s *r* i r^2 nalaze se korespondentne vrijednosti u odnosu na *d*.

Postoji još nekoliko koeficijenata asocijacije među varijablama koji se koriste kao veličine učinka: *kvadrirana eta, kvadrirana omega, Cohenov f* i *kvadrirani f* (f^2). O njima se govori nešto više u vezi s analizom varijance. Koeficijente asocijacije koji se koriste za dihotomne varijable spomenut ćemo u vezi s hi-kvadrat testom.

χ^2 – test

U psihologiji se vrlo često koristi χ^2-test, kojim se utvrđuje postoji li statistički značajna razlika između distribucija dviju varijabli – obično izraženih kvalitativnim (ili nekim bastardnim kvalitativno-kvantitativnim) kategorijama. Temelj za računanje χ^2-testa jest tablica vezanih frekvencija dviju varijabli. Te tablice mogu biti veličine 2x2, što znači da su i jedna i druga varijabla izražene samo pomoću dvije kategorije (to su obično dihotomne ili dihotomizirane varijable), ali mogu, naravno, biti i veličine 2x3, 3x3, 3x4, 4x4 itd., kad su varijable izražene u kvalitativnim kategorijama.

Ovdje se veličina učinka može izraziti pomoću *Cramerova V* koeficijenta ili pomoću *Cohenova w* koeficijenta. Za 2x2 tablicu može se koristiti *Φ* (fi) koeficijent (formula za Φ koeficijent: $\Phi = \sqrt{\frac{\chi^2}{N}}$). Može se koristiti i koeficijent kontingencije C, ali u zadnje se vrijeme, zbog priličnih nedostataka, taj koeficijent izbjegava.

Da bi slika o χ^2-testu, kao bazi za određivanje asocijacije među varijablama, bilo nešto jasnija, pogledajmo sljedeći primjer.

Jednom, davno, kad je još bilo muških studenata u studiju psihologije, u nekoliko generacija provedena je mala anketa među studentima koji su slušali statistički kolegij. Anketa se sastojala od samo jednoga pitanja: Molim vas da izrazite svoj stav prema statistici opredjeljujući se za jedan od ovih odgovora *volim statistiku*; *niti je volim niti ne volim*; *mrzim statistiku*. Naravno, pitanje je koliko su studenti iskreno odgovarali, ali budući da su odgovarali anonimno, ipak se donekle može vjerovati njihovim odgovorima. Rezultata su navedeni u tablici 3.

Tablica 3.
Rezultati ankete o stavu studenata psihologije prema statistici. Odgovor V: volim statistiku; odgovor nV/nM: niti je volim, niti je mrzim; odgovor M: mrzim statistiku. S f_o su označene opažene, a s f_t teoretske frekvencije.

		V	nV/nM	M	Σ
Studentice	f_o	29	16	101	146
	f_t	33.44	27.01	85.54	
Studenti	f_o	23	26	32	81
	f_t	18.56	14.99	47.46	
	Σ	52	42	133	227

Vizualna inspekcija tablice 3. pokazuje da je vrlo velik broj studentica, 116 (oko 79%), izabrao odgovor „M". Taj odgovor izabrao je manji broj studenata, „samo" 31 (oko 38%). Takva distribucija mogla bi, ugrubo, odgovarati realnosti. Iako po uspjehu u tom predmetu studenti nisu bili bolji od studentica, ipak su manje mrzili statistiku – to vjerojatno ima svoju psihološku podlogu, no tu interpretaciju ostavimo za druge prilike.

Letimičan pogled na opažene i teoretske frekvencije otkriva nam da su najveće razlike među njima u kategoriji „V" i „M", a to upućuje na postojanje povezanosti između varijable „spol studenata" i varijable „stav prema statistici".

Najprije će nas zanimati razlikuje li se distribucija odgovora studentica i studenata na postavljeno pitanje. Odgovor na to pitanje dobit ćemo pomoću *hi-kvadrat* testa kojim testiramo H_0 da nema razlike u distribuciji odgovora između studentica i studenata. Dakle

$$\chi^2 = \sqrt{\Sigma \frac{\left(f_o - f_t\right)^2}{f_t}} = 22.06 \qquad (8)$$

Uz dva stupnja slobode [za *hi-kvadrat* ss = (broj kategorija jedne varijable – 1) x (broj kategorija druge varijable – 1)] i pogledom u tablicu za *hi-kvadrat* vidimo da je $p < 0.01$. Granični hi-kvadrat uz razinu rizika od 1% i 2 stupnja slobode iznosi 9.21. Dakle, ne ostajemo na nul-hipotezi i konstatiramo da se distribucije odgovora studentica i studenata statistički značajno razlikuju. I to je sve što možemo zaključiti; možemo još reći da su najveće razlike u distribucijama – kako smo već vidjeli – u kategorijama „V" i „M".

Podatak o razlici u distribucijama i podatak o asocijaciji među varijablama dva su različita podatka.

Odredimo najprije Cramerov V kao pokazatelj povezanosti za rezultate iz tablice 3.

$$CramerovV = \sqrt{\frac{\chi^2}{N(k-1)}} = \sqrt{\frac{22.0567}{227 \cdot 1}} = 0.3117 \qquad (9)$$

Simbol *k* u Cramerovu *V* označuje manji broj kategorija u tablici kontingencije. U našem slučaju manji broj kategorija ima varijabla „spol": dvije. Značajnost toga koeficijenta najlakše je utvrditi pomoću *hi-kvadrat* testa: ako je *hi-kvadrat* značajan, značajan je i Cramerov *V*. Postoji, dakle, značajna asocijacija između varijable „spol studenata“ i „stav prema statistici“. Kvadrirani Cramerov V iznosi 0.097, a to znači da te dvije varijable imaju približno 10% zajedničkih faktora.

Izračunajmo sada Cohenov w, prema formuli:

$$w = \sqrt{\Sigma \frac{(p_t - p_o)^2}{p_t}} \tag{10}$$

p_o = vjerojatnost svake dobivene, tj. opažene frekvencije = f_o/N; p_t = vjerojatnost svake teoretske frekvencije u tablici kontingencije iz koje se računa *hi-kvadrat* test = f_t/N. Da bi to bilo jasnije evo dio računa na temelju tablice 3:

f_o	$p_o=f_o/N$	f_t	$p_t=f_t/N$	$(p_t-p_o)/p_t$
29	.127753	33.44	.147313	.002597
16	.070485	27.01	.118987	.019771
101	.444934	85.54	.376828	.012319
23	.101322	18.56	.081762	.004679
26	.114537	14.99	.066035	.035624
32	.140969	47.46	.209075	.022185
N = 227	Σ 1.000000	Σ 227.00	Σ 1.000000	Σ .097175

Dakle $w = \sqrt{.097175} = 0.3117$. U ovom slučaju Cramerov *V* i Cohenov *w* imaju identičnu numeričku veličinu (na četiri decimalna mjesta). To bi moglo značiti da je svejedno koji od njih upotrijebimo. Cramerov *V* nešto je jednostavniji za računanje; ne moraju se računati vjerojatnosti, a njihovo računanje produžuje posao – ali, dakako, svejedno nam je ako to radi neki statistički računalni program, recimo, SPSS. Cohenov *w* možda je nešto „statističkiji“ jer barata s vjerojatnostima opaženih i teoretskih frekvencija.

Da bismo vidjeli korist od računanja tih dodatnih pokazatelja, zamislimo da smo ponovili takvu anketu na novih 227 studenata, 146 studentica i 81 studentu i da smo dobili rezultate prikazane u tablici 4. (N.B. Ovdje su to, zapravo, samo preuređeni rezultati iz tablice 3).

U ovom slučaju – tablica 4. – vidimo da ima još više studentica koje mrze statistiku i još više studenata koji vole statistiku (u odnosu na frekvencije u tablici 3). Izračunati hi-kvadrat iznosi χ^2 = 45.54. Za tu veličinu hi-kvadrat testa sigurno je *p* puno manji nego za *hi-kvadrat* iz tablice 3, iako možemo reći da je u oba slučaja *p* dosta manji od .01. Budući da nam vrijednost *p* govori samo o vjerojatnosti kojom možemo očekivati hi-kvadrat određene veličine (jednako kao i kod *t*-testa), ne možemo ništa drugo zaključiti nego da se u oba slučaja (i u slučaju iz tablice 3 i u slučaju iz tablice 4) distribucije odgovora studentica i studenata statistički značajno razlikuju. Ali ne možemo ništa zaključivati o tome razlikuje li se slučaj iz tablice 3. i slučaj iz tablice 4. Ipak, možemo dobiti odgovor na to pitanje ako usporedimo bilo *Cramerov V*, bilo *Cohenov w* izračunat je za oba slučaja.

Tablica 4.
Rezultati ankete o stavu studenata psihologije prema statistici. Odgovor V: volim statistiku; odgovor nV/nM: niti je volim, niti je mrzim; odgovor M: mrzim statistiku. S f_o označene su opažene, a s f_t teoretske frekvencije. Malo promijenjeni rezultati iz tablice 3.

		V	nV/nM	M	Σ
Studentice	f_o	15	15	116	146
	f_t	34.73	16.72	94.55	
Studenti	f_o	39	11	31	81
	f_t	19.27	9.28	52.45	
	Σ	54	26	147	227

Cramerov V iz tablice 3., kao i *Cohenov w* iznosi 0.31, a iz tablice 4. *Cramerov V* iznosi 0.45, a toliko iznosi i *Cohenov w*.

Konvencija o interpretaciji veličina tih indeksa jednaka je interpretaciji za korelacije:

mala veličina učinka 0.10; srednja veličina učinka 0.30; velika veličina učinka 0.50.

Cramerov V i *Cohenov w* za rezultate iz tablice 3. bliži su srednjoj veličini učinka (oba iznose 0.31), a iz tablice 4. bliži su velikoj veličini učinka (oba iznose 0.45). Na temelju tih podataka može se ustvrditi da postoji razlika u asocijaciji rezultata iz tablice 3. i iz tablice 4. Postotak zajedničke varijance u prvom slučaju iznosi približno 9.6%, a u drugom slučaju približno 20.2%.

Analiza varijance

Veličina učinka može se – i treba – odrediti i u korištenju analize varijance. Glavna vrijednost dobivena analizom varijance jest F-omjer, koji pokazuje može li se ostati na nul-hipotezi, prema kojoj nema razlika među aritmetičkim sredinama različitih skupina ispitanika koji su bili pod utjecajem različitih razina jedne ili više nezavisnih varijabli. Prema nul-hipotezi, ako su eventualno i dobivene neke razlike, one su nastale samo pod utjecajem slučajnih čimbenika. Tih čimbenika koji djeluju po slučaju, ali uvijek, ima mnogo i djeluju pri svakom mjerenju, a imaju različita izvorišta: slučajni odabir uzoraka, slučajne pogreške pri mjerenju, slučajne varijacije samih ispitanika, slučajne promjene u neposrednoj okolini mjerenja itd.

Kako je, međutim, značajan F-omjer *neusmjeren*, tj. pokazuje samo da postoje statistički značajne razlike, ali ne pokazuje gdje su te razlike, između kojih aritmetičkih sredina (od najmanje tri skupine ispitanika koji se koriste, primjerice, u jednostavnoj analizi varijance), potrebna su dodatna provjeravanja. To su tzv. usporedbe parova aritmetičkih sredina za što se najčešće koristi Schefféov test[2]. Tim se postupkom dobije podatak o tome postoji li statistički značajna razlika između aritmetičkih sredina dvije skupine ispitanika (od tri ili više skupina) koji su bili podvrgnuti različitim razinama nezavisne varijable (ili različitim istraživačkim ili praktičnim prilikama odnosno uvjetima), pa ako ta razlika postoji, zaključujemo o specifičnom ili različitom djelovanju tih razina.

Obično se najprije određuje „opća“ veličina učinka. U jednostavnoj analizi varijance Cohen (1988) nudi *f*-indeks ako su sve podskupine podjednake po broju ispitanika:

$$f = \frac{SD_m}{SD}, \qquad (11)$$

gdje je SD_m standardna devijacija aritmetičkih sredina, a *SD* bi trebala biti standardna devijacija populacije, koju redovito zamjenjujemo standardnom devijacijom uzorka. U ovom slučaju može se uzeti standardna devijacija bilo koje skupine (budući da su *n*-ovi u podskupinama podjednaki i, osim toga, analizu varijance opravdano je provoditi samo ako su standardne devijacije u svim skupinama podjednake, tj. međusobno se ne razlikuju statistički značajno) ili pak zajednička standardna devijaciju svih skupina.

A SD_m računa se po ovoj formuli:

$$SD_m = \sqrt{\frac{\Sigma\left(m_g - m_z\right)^2}{g}} \qquad (12)$$

[2] Provedba tih usporedbi nije sasvim jednostavna pa treba konzultirati tekstove koji se detaljnije bave tim statističkim procesima.

U toj formuli m_g označava svaku pojedinu aritmetičku sredinu, aritmetičku sredinu svake podskupine u jednostavnoj analizi varijance, m_z zajedničku aritmetičku sredinu svih skupina, i g označava broj skupina (grupa) odnosno razina nezavisne varijable.

f-indeks je, da se tako izrazimo, sinkretičan, prikazuje općenito kolika je veličina učinka, poput *F*-omjera u analizi varijance. Interpretativna konvencija za *f*-indeks je ova:

0.10 mala veličina učinka, 0.25 srednja i 0.40 (i više) velika veličina učinka.

Za jednostavnu analizu varijance postoji još takvih indeksa, poput onoga koji navodi Howel (2010):

$$d = \sqrt{\frac{1}{g-1}\left(\frac{\Sigma(M - M_{tot})^2}{SD_{tot}{}^2}\right)} \qquad (13)$$

g označuje broj grupa, M pojedinačne aritmetičke sredine svake grupe, M_{tot} ukupnu, totalnu aritmetičku sredinu i SD_{tot}, tj. standardnu devijaciju svih ispitanika.

Nažalost različiti indeksi ne daju identične ishode, čak mogu biti vrlo različiti, pa unaprijed treba izabrati jedan od njih.

No, postoje dodatni indeksi koji daju vrlo korisne informacije o stupnju povezanosti varijabli korištenih u analizi varijance, zapravo, o snazi odnosa među varijablama.

Najčešće se koristi *kvadrirana eta*: $\eta^2 = \frac{ZK_{iz}}{ZK_{tot}}$ (14)

ZK_{iz} je „zbroj kvadrata“ koji se koristi u analizi varijance; ZK_{tot} ukupan je „zbroj kvadrata“ svih rezultata odnosno rezultata svih ispitanika.

Kvadrirana eta može se odrediti i pomoću *F*-omjera i stupnjeva slobode:

$$\eta^2 = \frac{F \cdot ss_{iz}}{ss_{iz}F + ss_{un}} \qquad (15)$$

ss_{iz} su stupnjevi slobode koji se vežu uz zbroj kvadrata između grupa (ZK_{iz}) i ss_{un} stupnjevi su slobode koji se vežu uz zbroj kvadrata unutar grupa (ZK_{un})[3].

[3] U statističkoj literaturi anglo-američkog podrijetla obično se umjesto subskripta „un“ koristi subskript „error“, ili skraćeno „e“, jer, doista, varijabilitet „unutar grupa“ rezultat je slučajnih pogrešaka pri mjerenju i pri uzorkovanju, a koje se distribuiraju normalno.

Citat iz Petzove statistike kaže (str. 360): „η^2, kvadrirano eta, proporcija je varijance odgovorna za veličinu učinka. Ili, preciznije, η^2 proporcija je totalnog (ukupnog) varijabiliteta rezultata oko zajedničke (opće) aritmetičke sredine koji je odgovoran za varijabilitet aritmetičkih sredina grupa čiji su rezultati podvrgnuti analizi varijance. Ili: koliki dio ukupne varijance rezultata otpada na utjecaj nezavisne varijable."

Kvadrirana eta poznata je i kao *korelacijski omjer* koji se koristi za određivanje stupnja povezanosti između varijabli kojima crta regresije nije pravac nego neki stupanj zakrivljenog odnosa pa se stoga naziva još i koeficijent zakrivljene korelacije. Važno svojstvo kvadrirane ete jest, dakle, neosjetljivost na zakrivljenje odnosa među varijablama, pa u takvim slučajevima nije opravdano računati Pearsonov *r*, koji je koeficijent linearne povezanosti među varijablama.

Prema Gamstu i suradnicima (2008) veličina *kvadriranog eta* kao
veličina učinka od 0.09 je mala, 0.14 srednja, a 0.22 (i veća) je velika.

Pogledajmo kako to izgleda na jednom primjeru posuđenom iz knjige Analiza varijance u psihologijskim istraživanjima (Kolesarić, 2006). To je primjer koji je pripravljen na temelju provedenih istraživanja, ali je malo prilagođen kao udžbenički primjer.

Kako je već u jednom prethodnom primjeru spomenuto, psiholozi vrlo često kao zavisnu varijablu koriste vrijeme reakcije – vrijeme koje je potrebno ispitaniku da reagira na neki jednostavniji ili složeniji podražaj (unaprijed dogovoren ili nedogovoren). Podražaji mogu biti vrlo različiti – vidni, slušni, taktilni, verbalni, slikovni itd. (psiholozi su smislili različite jednostavnije ili složenije podražaje), ali i reakcije mogu biti različite, a najčešće su motorne ili verbalne. Najjednostavnije korištenje vremena reakcije kao zavisne varijable jest *vrijeme jednostavne senzomotorne reakcije* – to je vrijeme koje protekne između prezentacije nekog jednostavnog senzornoga podražaja (npr. svjetlosnoga bljeska ili kratkotrajnoga zvučnoga signala) i dogovorene motorne reakcije (npr. podizanje prsta s nekog tipkala ili tipke). To je bila, a još uvijek i jest, među psiholozima popularna zavisna varijabla jer se općenito misli da je to objektivniji pokazatelj ili mjera psiholoških procesa nego, recimo, verbalni odgovor ispitanika. Međutim, još krajem 19. stoljeća bilo je poznato da se na vrijeme jednostavne senzomotorne reakcije može djelovati na različite načine, a jedan od njih je i formuliranje upute ispitanicima.

Temeljem tog staroga podatka proveden je eksperiment (Kolesarić, Krizmanić i Špehar, 1986) u kojemu je u tri skupine ispitanika mjereno jednostavno vrijeme senzomotorne reakcije na zvučni podražaj. Svaka skupina imala je malo izmijenjenu uputu. Jedna skupina dobila je standardnu uputu koja glasi otprilike ovako: „Molimo vas da reagirate što brže možete na zvučni podražaj“. Skupina s tom uputom označena je s *N* (neutralna uputa). Druga skupina dobila je „senzornu“ uputu koja je glasila otprilike ovako: „Molimo vas da reagirate što brže možete, ali budite sigurni da ste čuli podražaj“. Ta je skupina označena sa *S* (senzorna uputa). Treća je skupina dobila „motornu“ uputu koja je glasila otprilike ovako: „Molimo vas da reagirate doista što brže možete čim čujete zvuk“ (označena s M). Motorna reakcija bila je što brže odmicanje kažiprsta s tipkala. Na temelju rezultata tog eksperimenta složen je primjer, malo prilagođen potrebi udžbenika, te je, u nekoj mjeri i nekom obliku, korišten i u knjizi *Analiza varijance u psihologijskim istraživanjima* (Kolesarić, 2006) i u *Petzovoj* statistici (Petz, Kolesarić i Ivanec, 2012).

Osnovni rezultati prikazani su u tablici 5.

Tablica 5.
Vrijeme jednostavne senzomotorne reakcije na zvučni podražaj tri skupine ispitanika izraženo u tisućinkama sekunde. Skupine su označene s N (neutralna uputa), sa S („senzorna“ uputa) te s M („motorna“ uputa).

	N	S	M	
M	177.0	183.9	164.1	$\overline{M}$ = 175.0
SD	14.92	18.84	16.99	
n	18	18	18	

U tablici 6. prikazani su osnovni podaci potrebni u analizi varijance.

Tablica 6.
Osnovni podaci potrebni u analizi jednostavne analize varijance u primjeru s jednostavnim vremenom senzomotorne reakcije za rezultate prikazane u tablici 5. ZK = zbroj kvadrata; ss = stupnjevi slobode; PK = prosječni kvadrat ili varijanca.

Izvor varijabiliteta	ZK	ss	PK	F
Između grupa	3668.111	(g – 1) = 2	1834.0556	6.35
Unutar grupa	14727.889	(N – g) = 51	288.7821	
Ukupno (total)	18396.000	N – 1 = 53		

PK je prosječni kvadrat ili varijanca, dobije se dijeljenjem zbroja kvadrata *(ZK)* s odgovarajućim stupnjevima slobode. Češće se upotrebljava izraz „prosječni kvadrat" zbog izbjegavanja konfuzije jer su to varijance na različitim razinama.

Uz 5% rizika dobiveni *F*-omjer statistički je značajan, tj. $p < 0.05$. Granična vrijednost u tablicama *F*-omjera za 5% rizika te 2 i 51 stupanj slobode iznosi 3.18, a dobiveni je 6.35. Prema tome, ne ostajemo na nul-hipotezi.

Pojedinačne usporedbe, pomoću Schefféova testa, pokazale su da je razlika između aritmetičke sredine neutralne skupine (M_N = 177.0) i aritmetičke sredine „motorne" skupine (M_M = 164.1) statistički značajna na razini rizika od 5%, kao i razlika između „senzorne" (M_S = 183.9) i „motorne" (M_M = 164.1) skupine.

Na temelju tih statističkih vrijednosti možemo zaključiti, dakle, da postoji utjecaj sadržaja upute na vrijeme jednostavne senzomotorne reakcije. Ali, nemamo ni jedan podatak o tome kolika je, zapravo, razlika, odnosno, kolike su razlike. Podatak da je pri odustajanju od nul-hipoteze $p \leq 0.01$ ne govori nam praktično ništa o veličini razlike, a to je, ipak, podatak koji treba zanimati svakog istraživača, a osobito praktičara.

No, pođimo korak dalje. Izračunajmo najprije Cohenov *f*, koji pokazuje veličinu učinka u jednostavnoj analizi varijance. Izračunat ćemo ga pomoću formula (11) i (12) i dobit ćemo f = 0.48. Viši je od 0.40. pa je to, dakle, velika veličina učinka (granične vrijednosti interpretativne konvencije nalaze se na stranici 28).

Izračunajmo i *d* koji predlaže Howel – formula (13), str. 28. Dobije se nešto manja vrijednost, d = 0.41, jer je, očito, ovaj indeks konzervativniji od *f*-indeksa, ali, također, predstavlja veliku veličinu učinka.

Koji od njih odabrati? Najbolje oba, ali to ovisi o korisniku, o tome želi li stroži ili malo manje strog kriterij, a ovisi i o vrsti varijabli i cilju istraživanja.

Sad odredimo *kvadrirano eta* prema formuli (14): $\eta^2 = \frac{3668.11}{18396} = 0.20$. Na str. 29 nalazi se konvencija o interpretativnoj veličini *kvadriranog eta*. Misli se da je 0.22 velika veličina, a ovdje je dobiveno 0.20, dakle, blizu velike veličine. Proporcija nezavisne varijable (u ovom slučaju vrsta upute ispitanicima) koja je odgovorna za razlikovanje aritmetičkih sredina tri skupine ispitanika iznosi 0.20 ili 20%.

Zašto nam je važan takav podatak? Redovito smo skloni statistički značajnu razliku (dobivenu korištenjem *t*-testa ili analize varijance) generalizirati i reći otprilike ovako: skupina ispitanika koja je imala „motornu“ uputu ima kraće vrijeme jednostavne senzomotorne reakcije. A taj zaključak nije sasvim točan. Ta tvrdnja vrijedi za prosječan rezultat, a to znači da ima onih ispitanika koji doista imaju kraće vrijeme, ali i onih koji nemaju kraće vrijeme reakcije. Međutim, dodatni podaci, Cohenov *f* i Howellov *d*, upućuju na to kolika je zajednička varijanca. *Kvadrirano eta* pokazuje nam koliko je snažan utjecaj nezavisne varijable na zavisnu varijablu – što se prosječne vrijednosti u zavisnoj varijabli različitih skupina ispitanika više međusobno razlikuju, to je veći utjecaj nezavisne varijable. Prema tome, nakon ove složenije obrade rezultata možemo zaključiti da ne samo da postoje prosječne razlike između skupina ispitanika koje su imale različite upute, nego i da su te razlike velike. Odnosno, očito postoji značajan i relativno velik utjecaj uputa ispitanicima na njihovo vrijeme reakcije.

Kvadrirano eta ima još jedno korisno svojstvo: može se koristiti kao pokazatelj parcijalne korelacije pa se tada zove *parcijalno kvadrirano eta*, a bilježi se s *parc.*η^2. Koristi se u složenoj analizi varijance kad se želi utvrditi veličina učinka pojedinih nezavisnih varijabli. U jednostavnoj analizi varijance *kvadrirano eta* i *parcijalno kvadrirano eta* jedno su te isto.

Kao primjer određivanja veličine učinka kod složene analize varijance posudit ćemo opet primjer iz *Analize varijance u psihologijskim istraživanjima* (Kolesarić, 2006, str. 131). Taj je primjer proširenje prethodno opisanog eksperimenta s vremenom reakcije i različitim uputama, ali ovaj put ispitanici su podijeljeni na mlađe i starije. Kao što je poznato, mlađe i starije osobe razlikuju se, u prosjeku, po brzini reagiranja (stariji su, naravno, sporiji), pa je svaka skupina od 18 ispitanika podijeljena na mlađe i starije ispitanike, a to znači da sada imamo dvije nezavisne varijable: jedna je vrsta upute, koja ima tri kategorije, a druga je dob ispitanika, koja ima dvije kategorije. Njihovi su rezultati, malo prilagođeni edukacijskim potrebama, prikazani u tablici 7.

Tablica 7.
Rezultati u 2x3 tablici složene analize varijance

	Mlađi N	Mlađi S	Mlađi M	Mlađi – svi
M	175.8	181.0	149.9	168.9
SD	15.62	20.78	9.13	15.90
n	9	9	9	27
	Stariji N	Stariji S	Stariji M	Stariji – svi
M	178.2	186.9	178.2	181.1
SD	15.03	17.40	8.86	14.23
n	9	9	9	27
	Svi N	Svi S	Svi M	
M	177.0	183.95	164.05	175.0
SD	15.33	19.16	9.0	13.50
n	18	18	18	54

Završni podaci složene analize varijance 2x3 nalaze se u tablici 8.

Tablica 8.
Završni rezultati složene analize varijance

Izvor varijabiliteta	ZK	ss	PK	F
Dob (ml., st.)	2016.6667	$g_{Ml,St}$ - 1 = 1	2016.6667	8.85 $p < 0.01$
Upute (N, S, M)	3668.1111	$g_{N,S,M}$ -1 = 2	1834.0556	8.05 $p < 0.01$
Interakcija	1768.7778	($g_{Ml,St}$-1)($g_{N,S,M}$ -1)=2	884.3889	3.88 $p < 0.05$
Unutar grupa	10932.4445	N – g = 48	227.7593	
Ukupno (total)	18386.0001	N – 1 = 53		

Za oba glavna efekta, tj. za obje varijable *F*-omjeri statistički su značajni; značajan je i *F*-omjer za interakciju. Postojanje razlike među skupinama koje su imale različite upute već smo utvrdili jednostavnom analizom varijance. Novi podatak, koji smo dobili složenom analizom varijance, jest značajna razlika u brzini reakcije između skupine mlađih i skupine starijih ispitanika, a kvalitativno novi podatak jest podatak o interakciji. Opisno, interakcija se u ovom eksperimentu očituje u tome da mlađi ispitanici imaju značajno kraće vrijeme s „motornom“ uputom, ali ne i sa „senzornom“ uputom.

Na temelju podataka iz tablice 8. možemo odrediti *parcijalne kvadrirane ete* pomoću formula preuzetih iz *Petzove statistike* (2012). *Parcijalne kvadrirane ete* određuju se posebno za svaku varijablu i posebno za interakciju.

Za varijablu „upute“:

$$parc.\eta^2_{upute} = \frac{ZK_{upute}}{ZK_{total} - ZK_{dob} - ZK_{interakcija}} \qquad (16)$$

Za varijablu „dob“:

$$parc.\eta^2_{dob} = \frac{ZK_{dob}}{ZK_{total} - ZK_{upute} - ZK_{interakcija}} \qquad (17)$$

Za interakciju dob * upute:

$$parc.\eta^2_{inter} = \frac{ZK_{interakcija}}{ZK_{total} - ZK_{upute} - ZK_{dob}} \qquad (18)$$

ZK jesu odgovarajući zbrojevi kvadrata koji se mogu očitati iz tablice 8.

Koristeći gornje formule dobili smo:

$parc.\eta^2_{upute} = 0.25$ $\qquad$ $parc.\eta^2_{dob} = 0.16$ $\qquad$ $parc.\eta^2_{inter} = 0.14$

Proporcija totalnog varijabiliteta koji je odgovoran za veličinu učinka u varijabli „upute“ iznosi 0.25, a to je velika veličina učinka, veća od granične (koja iznosi 0.22, str. 29). Za dob iznosi 0.16, što je nešto veće od srednje veličine učinka (0.14), a za interakciju iznosi upravo 0.14. Ti se podaci mogu izravno interpretirati kao proporcija ili postotak (ako dobivene brojeve pomnožimo sa 100) zajedničke varijance, poput r^2 (koeficijent determinacije).

Složenom analizom varijance utvrdili smo da se skupine koje su imale različite upute pri provedbi mjerenja vremena jednostavne senzomotorne reakcije međusobno razlikuju (detaljnijim usporedbama utvrdili smo i koje se skupine međusobno statistički značajno razlikuju), utvrdili smo i da se mlađi i stariji ispitanici međusobno razlikuju po brzini reakcije (mlađi su općenito brži), te da postoji interakcija: najniži prosječni rezultat imala je mlađa skupina s motornom uputom.

Dodatni uvid omogućilo nam je računanje *parcijalnoga kvadriranog eta* koji pokazuje da je varijabla „upute“ odredila relativno najveći dio totalne varijance, a varijabla „dob“ i interakcija to su odredile puno manje. Taj podatak, da u totalnoj varijanci dominira način prezentiranja upute i njezin sadržaj, može imati sasvim praktičnih posljedica u provedbi ispitivanja u kojima se koristi vrijeme reakcije kao zavisna varijabla, a koristi se i u eksperimentalnim i u praktičnim prilikama. Ali, s prilično velikom vjerojatnošću može se doći i do općenitijega zaključka: nije svejedno kako – formom i sadržajem – prezentiramo uputu svojim ispitanicima. Ako različito prezentiranje upute ima učinka na vrijeme reakcije, kakav tek utjecaj ima na, recimo, verbalne reakcije, koje su pod znatno većom – namjernom ili nenamjernom – kontrolom ispitanika?!

Moguće je, naravno, i spekuliranjem doći do sličnih ili jednakih zaključaka, ali naše će inventivno rezoniranje biti uvjerljivije ako je potkrijepljeno objektivnim statističkim argumentima.

I ovdje možemo odrediti, nazovimo ga tako, ukupni *kvadrirani eta* pomoću formule (14), a iznosi 0.41, dakle vrlo je velika veličina učinka (granični za veliki učinak jest 0.22, str. 29). Napomena: ovdje ćemo ZK_{iz} dobiti zbrajanjem vrijednosti iz tablice 8, $ZK_{upute}+ZK_{dob}+ZK_{interakcija}$. Taj *kvadrirani eta*, međutim, u ovom slučaju nema neku osobito informativnu vrijednost, jer smo računali parcijalne indekse koji su nam važniji.

Ponavljamo upozorenje statističara: dobivene vrijednosti nikada se ne smiju smatrati apsolutnim podacima, koji jednako vrijede u svakom slučaju. Naprotiv, njihovo i teorijsko i praktično značenje ovisi o kontekstu, o korištenim varijablama i ciljevima istraživanja.

Howell (2010) nudi, prema njegovu mišljenju, bolji indeks veličine učinka iz korelacijske porodice, ω^2 (*kvadrirana omega*). Osim njega, i mnogi drugi autori tvrde da se taj indeks najčešće koristi u analizi varijance. *Kvadrirana omega* pokazuje kolika je proporcija varijabiliteta u zavisnoj varijabli u odnosu na određenu nezavisnu varijablu. Uobičajeno je da se *kvadrirana omega* interpretira samo za statistički značajne rezultate. Statističari kažu da je *kvadrirana omega* bolja od *kvadrirane ete* zato što u formuli uzima u obzir stupnjeve slobode. Formula za određivanje *kvadrirane omege* u jednostavnoj analizi varijance:

$$\omega^2 = \frac{ZK_{iz} - (g-1)PK_{un}}{ZK_{tot} + PK_{un}} \tag{19}$$

ZK_{iz} je zbroj kvadrata za određivanje varijabiliteta između grupa; (g – 1) stupnjevi su slobode s kojima se dijeli ZK_{iz}, tj. broj grupa minus 1; PK_{un} prosječni je kvadrat (varijanca) i pokazuje variranje unutar grupa, a ima značenje pogreške mjerenja i pogreške uzorkovanja (u engl. jeziku najčešće se označuje s „error“); ZK_{tot} ukupan je zbroj kvadrata za sve rezultate[4].

Koristeći rezultate, kao i za kvadriranu *etu*, iz tablice 6: $\omega^2 = 0.10$. Taj je indeks konzervativniji od *kvadrirane ete* i redovito je manji, što se vidi i u našem slučaju. Interpretira se prema konvenciji kao i *kvadrirana eta* ili se pomnoži sa 100 i tako se dobije postotak zajedničke varijance, dakle, u ovom slučaju, zajednička varijanca iznosi 10%.

Kvadriranu omegu smije se koristiti samo ako su veličine skupina jednake ili barem vrlo slične.

Formule za parcijalnu *kvadriranu omegu* za dvosmjernu analizu varijance (jednu nezavisnu varijablu označit ćemo s A, drugu s B, a interakciju s AB) jesu:

$$parc.\omega_A^2 = \frac{ss_{izA}\left(PK_{izA} - PK_{un}\right)}{ZK_{izA} + \left(N_{tot} - ss_{izA}\right)PK_{un}} \qquad (20)$$

$$parc.\omega_B^2 = \frac{ss_{izB}\left(PK_{izB} - PK_{un}\right)}{ZK_{izB} + \left(N_{tot} - ss_{izB}\right)PK_{un}} \qquad (21)$$

$$parc.\omega_{AB}^2 = \frac{ss_{AB}\left(PK_{AB} - PK_{un}\right)}{ZK_{AB} + \left(N_{tot} - ss_{AB}\right)PK_{un}} \qquad (22)$$

Za rezultate iz tablice 8. dobivamo ove vrijednosti parcijalnih kvadriranih omega:

$$parc.\omega_{upute}^2 = 0.21 \qquad parc.\omega_{dob}^2 = 0.13 \qquad parc.\omega_{\mathrm{int}\,er}^2 = 0.09$$

Te su vrijednosti manje nego one vrijednosti koje smo dobili računajući *parcijalne kvadrirane ete*, ali zaključci mogu biti jednaki, možda malo stroži.

Koji od tih indeksa koristiti? Čini nam se da se *kvadriranu etu* češće može susresti, ali u novije vrijeme *kvadrirana omega*, sve više prodire. Ipak, vrlo često autori prikazuju oba indeksa – za svaki slučaj!

Postoji još takvih indeksa. Jedan je od njih *kvadrirani epsilon:*

$$\varepsilon^2 = \frac{ZK_{iz} - ss_{iz}PK_{un}}{ZK_{tot}} \qquad (23)$$

[4] U Petzovoj statistici, nažalost, u formuli za *kvadriranu omegu* na str. 361. u nazivniku pogrešno stoji minus, a treba biti plus, dakle, točno je $ZK_{tot} + PK_{un}$. Međutim, na str. 362. gdje se računaju vrijednosti iz primjera, formula je točno napisana.

Višestruka (multipla) regresija

U modelu višestruke regresije pogodan indeks za određivanje veličine učinka jest Cohenov f^2.

$$f^2 = \frac{R^2}{1-R^2} = \frac{\eta^2}{1-\eta^2} \qquad (24)$$

$R^2 = \eta^2$ (η^2 je kvadrirana eta koju smo već spominjali). To je globalna veličina učinka i pokazuje proporciju objašnjene varijance nasuprot neobjašnjenoj varijanci.

No, važnija je specifičnija veličina učinka kad se promatra višestruka korelacija jedne varijable i skupa nekoliko varijabli. Ako se ta jedna varijabla označi s *K*, a skup ostalih varijabli s *P* formula za određivanje veličine učinka pomoću f^2 bit će:

$$f^2 = \frac{R_{KP}^2 - R_P^2}{1-R_{KP}^2} \qquad (25)$$

R_{KP}^2 je proporcija varijance odgovorna za *K* i *P* zajedno, a R_P^2 je proporcija varijance odgovorna za *P* (skup varijabli u multiploj regresiji koji obično nazivamo prediktorima). Stoga brojnik u gornjoj formuli ($R_{KP}^2 - R_P^2$) pokazuje proporciju varijance odgovornu samo za varijablu *K* u odnosu na sve druge varijable (tj. *P*).

* * *

U **Dodatku A** nalaze se konvencionalne granične vrijednosti za različite veličine učinaka, u **Dodatku B** uobičajena tablica graničnih vrijednosti t-testa i u **Dodatku C** granične vrijednosti χ^2 testa.

Literatura

Bachmann, C., Luccio, R. i Salvadori, E. (2005). Statistična pomembnost in njen pomen. *Psihološka obzorja, 14(3),* 7-40.

Becker. L.A. (2000). http://web.uccs.edu/lbecker/Psy590/es.htm

Coe, R. (2002). *It's the Effect Size, Stupid. What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: LEA.

Cohen, J. (1994). The earth is round (p < 0.05). *American Psychologist, 49(12)*, 997-1003.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Resarch and Practice, 40(5)*, 532-538.

Gamst, G., Meyers, L.S. i Guariono, A.J. (2008). *Analysis of Variance Designs*. Cambridge: Cambridge University Press.

Glass, G.V. & Hopkins K.D. (1984). *Test Bank for Statistical Methods in Education & Psychology*. Englewood Cliffs, NJ: Prentice Hall.

Howell, D.C. (2010). Statistical Methods for Psychology. Belmont, CA: Wadsworth, Cengage Learning.

Kolesarić, V. (2006). *Analiza varijance u psihologijskim istraživanjima*. Osijek: Filozofski fakultet u Osijeku.

Kolesarić, V., Krizmanić, M. i Špehar, B. (1986). Uputom inducirane individualne razlike u "senzornom" i "motornom" stavu pri ispitivanju vremena jednostavne senzomotorne reakcije. *Primijenjena psihologija, 7(1-4),* 275-280.

Kühberger, A., Fritz, A. i Scherndl, T. (2014). Publication bias in Psychology: A diagnosis based on the correlation between effect size and sample size. DOI:10.1371/*Journal.pone.0105825*.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

Nickerson, R.S. (2000). Null hypothesis significance testing: A review of an old and continuing controcersy. *Psychological Methods, 5(2),* 241.

Petz, B., Kolesarić, V. i Ivanec. D. (2012). *Petzova statistika*. Jastrebarsko: Naklada Slap.

Rosnow, R. i Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psyhological science. *American Psychologist, 44*, 1276-1284.

Vacha-Haase, T., Nilsson, J.E., Reetz, D.R., Lance, T.S. i Thompson, B. (2000). Repoting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology, Vol. 10(3)*, 413-425.

DODATAK A

Granične vrijednosti za interpretaciju različitih veličina učinka

Veličina učinka	***Mala***	***Srednja***	***Velika***
Cohenov d	0.20	0.50	0.80
r	0.10	0.30	0.50
r^2	0.01	0.09	0.25
f	0.10	0.25	0.40
Φ, Cramerov V, Cohenov w	0.10	0.30	0.50
Cohenov f^2	0.02	0.15	0.35
η^2, ω^2 i R^2	0.01	0.06	0.14

Primjedba: U različitim statistikama mogu se naći donekle različite vrijednosti. Tako, primjerice, Gamst i suradnici (2008) za interpretaciju veličine *kvadrirane ete* daju ove vrijednosti:

veličina učinka od 0.09 je mala, 0.14 je srednja, a 0.22 (i veća) je velika.

DODATAK B

Granične vrijednosti t-testa

	razine rizika			
ss	**10%**	**5%**	**2%**	**1%**
1	6.34	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.90	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.06
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.72	2.09	2.53	2.84
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.50	2.81
24	1.71	2.06	2.49	2.80
25	1.71	2.06	2.48	2.79
26	1.71	2.06	2.48	2.78
27	1.70	2.05	2.47	2.77
28	1.70	2.05	2.47	2.76
29	1.70	2.04	2.46	2.76
30	1.70	2.04	2.46	2.75
35	1.69	2.03	2.44	2.72
40	1.68	2.02	2.42	2.71
45	1.68	2.02	2.41	2.69
50	1.68	2.01	2.40	2.68
60	1.67	2.00	2.39	2.66
70	1.67	2.00	2.38	2.65
80	1.66	1.99	2.38	2.64
90	1.66	1.99	2.37	2.63
100	1.66	1.98	2.36	2.63
125	1.66	1.98	2.36	2.62
150	1.66	1.98	2.35	2.61
200	1.65	1.97	2.35	2.60
300	1.65	1.97	2.34	2.59
400	1.65	1.97	2.34	2.59
500	1.65	1.96	2.33	2.59
1000	1.65	1.96	2.33	2.58
$\propto$	1.65	1.96	2.33	2.58

DODATAK C

Granične vrijednosti χ^2 testa uz različite razine rizika

ss↓ / p→	0.30	0.20	0.10	0.05	0.02	0.01
1	1.074	1.642	2.706	3.841	5.412	6.635
2	2.408	3.219	4.605	5.991	7.824	9.210
3	3.665	4.642	6.251	7.815	9.837	11.431
4	4.878	5.989	7.779	9.488	11.668	13.277
5	6.064	7.289	9.236	11.070	13.388	15.086
6	7.231	8.558	10.645	12.592	15.033	16.812
7	8.383	9.803	12.117	14.067	16.622	18.475
8	9.524	11.030	13.362	15.507	18.168	20.090
9	10.656	12.242	14.684	16.919	19.679	21.666
10	11.781	13.442	15.987	18.307	21.161	23.209
11	12.899	14.631	17.275	19.675	22.618	24.725
12	14.011	15.812	18.549	21.026	24.054	26.217
13	15.199	16.985	19.812	22.362	25.472	27.688
14	16.222	18.151	21.064	23.685	26.873	29.141
15	17.322	19.311	22.307	24.996	28.259	30.578
16	18.418	20.465	23.542	26.296	29.633	32.000
17	19.511	21.615	24.769	27.587	30.995	33.409
18	20.601	22.760	25.989	28.869	32.346	34.805
19	27.689	23.900	27.204	30.144	33.687	36.191
20	22.775	25.038	28.412	31.410	35.020	37.566
21	23.858	26.171	29.615	32.671	36.343	38.932
22	24.939	27.301	30.813	33.924	37.659	40.289
23	26.018	28.429	32.007	35.172	38.968	41.638
24	27.096	29.553	33.196	36.415	40.270	42.980
25	28.172	30.675	34.382	37.652	41.566	44.314