UDK 37.012:311.2
Pregledni članak
Primljeno: 12. 2. 2011.
Prihvaćeno: 20. 5. 2011.

TESTIRANJE NORMALNOSTI DISTRIBUCIJE U ISTRAŽIVANJIMA ODGOJA I OBRAZOVANJA

doc. dr. sc. Siniša Opić
sinisa.opic@ufzg.hr

Sažetak: *Testiranje normalnosti distribucije (uz homogenost varijance) neizostavan je parametar u istraživanjima odgoja i obrazovanja s obzirom na odabir određenih statističkih parametrijskih ili neparametrijskih testova za testiranje hipoteza. Međutim, neopravdano se izostavlja, što u konačnici može polučiti distorzirane rezultate (generalizacije). U radu su opisane statističke vrijednosti normalne distribucije te njezini modaliteti: spljoštenost (kurtosis) i simetričnost/asimetričnost (skewness). Isto tako, opisane su i druge teorijske distribucije koje se često susreću u istraživanjima odgoja i obrazovanja: Poisson, uniform, t-distribucija, hi-kvadrat i U distribucija. Također su elaborirani testovi za testiranje normalnosti distribucije: Kolmogorov-Smirnov test (K-S), Lilliefors test, Shapiro-Wilks test i Anderson-Darling Normality Test. Kao empirijska provjera statističke značajnosti vrijednosti parametrijske i neparametrijske statistike, uspoređene su vrijednosti statističke značajnosti ANOVA-e i njezina neparametrijskog pandana, Kruskal-Wallis testa. Parcijalno su naglašeni i opisani testovi normalizacije, odnosno transformacije vrijednosti s ciljem poboljšanja normalnosti distribucije. U inferencijalnoj statistici postoji teorem koji pridonosi izvođenju valjanih zaključaka primjenom parametrijske statistike i kad nije slučaj normalne distribucije: centralni granični teorem. Njegova uloga u istraživanjima odgoja i obrazovanja posebno je opisana.*

Ključne riječi: *distribucija podataka, normalna distribucija, kurtosis, skewness, testovi provjere normalnosti, transformacije, centralni granični teorem*

1. Normalnost distribucije

Jedan od vrlo važnih preduvjeta za odabir pristupa obradi podataka u kvantitativnom (i kvalitativnom) pristupu znanstvenim istraživanjima na određenom uzorku ispitanika jest provjera normalnosti distribucije (također i testiranje homogenosti varijance). Distribucija u statistici znači raspodjelu rezultata u nekom skupu. Prema statističkom rječniku (Kolesarić i Petz, 2003.), termin distribucija upotrebljava se i za frekvencije određenih statističkih jedinica koje pripadaju različitim kategorijama određene varijable; npr. sociodemografska obilježja. Provjera normalnosti distribucije implicira razumijevanje glavnih statističkih pojmova vjerojatnosti, a primjenjuje se pri odabiru određenih pristupa obradi podataka (parametrijske ili neparametrijske statistike) te donošenju znanstveno relevantnih zaključaka koji slijede iz rezultata takvih istraživanja. Zapravo, riječ je o donošenju zaključaka izvedenih iz istraživanja na određenom uzorku koji trebaju vrijediti na cijeloj populaciji. Zbog potrebe generalizacije zaključaka dobivenih iz uzorka na cjelokupnu populaciju, normalnost distribucije osnovni je preduvjet, no treba naglasiti da uvijek treba biti oprezan u izvođenju generalizacija u društvenim znanostima iako je distribucija u djelomično normalnoj raspodjeli (zbog parazitarnih faktora.) To je svojevrsni teret u istraživanjima društvenih znanosti, ali ujedno i izazov. U istraživanjima odgoja i obrazovanja često se zamjećuje da se primjenjuje parametrijska statistika iako nije provjerena normalnost distribucije, što je osnovni preduvjet. "Snaga" parametrijskih veća je od "snage" neparametrijskih testova, a tu snagu možemo interpretirati kao vjerojatnost odbacivanja nul hipoteze, u slučaju kad je alternativna hipoteza istinita (tj. kad je mala vjerojatnost pogreške tipa II). Naime, s povećanjem snage testa opada vjerojatnost pogreške tipa II.

Zapravo, snaga testa sastoji se u njegovoj sposobnosti da otkrije određenu razliku, ako ona postoji (Petz, 1997.). To znači da ćemo razliku između dva uzorka, ako ona postoji, preciznije utvrditi pomoću parametrijskih nego pomoću neparametrijskih testova.

Za bolje razumijevanje normalne distribucije najbolje je promotriti Galtonovu mašinu za grah (dasku s čavlima), poznatu kao *quincunx*. Riječ je o uređaju (lijevku) koji se sastoji od ploče s umetnutim redovima čavlića. Kad se kuglice puste s vrha kroz lijevak, one se grupiraju oko sredine te raspršuju s lijeve i desne strane (disperzija). Smisao je Galtonove daske da se kuglice odbijaju od čavlića po slučaju. To je zapravo dokaz slučajnog variranja (na čemu se temelji normalna distribucija).

Takva normalna distribucija, gdje se najviše podataka grupira oko sredine a tada se rasipaju simetrično s obje strane, čest je slučaj u prirodi. Međutim, ni kod nekih prirodnih pojava, procesa često nemamo normalnost distribucije iako bismo je logično očekivali. To je vrlo indikativno jer otežava provedbu istraživanja i odabir statističkih testova obrade, te, štoviše, naglašava se opreznost u donošenju zaključaka.

Pretpostavke su za normalnu distribuciju (Petz, 1997.):

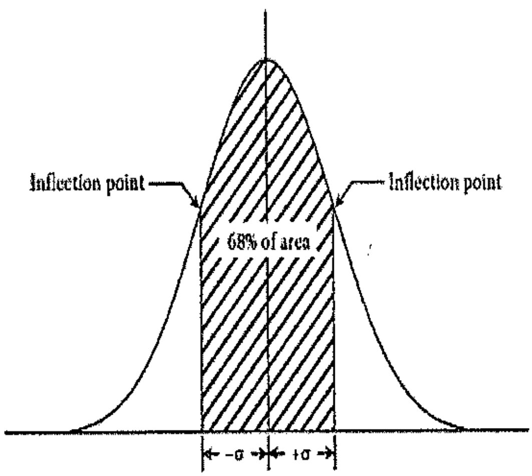
- Normalna distribucija slijedi pretpostavku da su podaci distribuirani slučajno, odnosno da se ono što mjerimo u populaciji distribuira po normalnoj raspodjeli.

- Treba uzeti dovoljno velik broj mjerenja.
- Sva mjerenja trebaju biti provedena istom metodom i u što sličnijim uvjetima.
- Uzorak mjerenja treba biti homogen prema nekim svojstvima, a heterogen prema onom svojstvu koje mjerimo.

Otežavajuća je okolnost da je za razliku npr. od prirodnih znanosti u društvenim znanostima česta pojava da je distribucija podataka suprotna normalnoj, što zahtijeva primjenu normalizacije, ili primjenu metoda neparametrijske statistike. Analogno takvu stajalištu, Micceri (1989.) naglašava da su rijetko varijable normalno distribuirane, osobito u društvenim znanostima. Autor u svojoj studiji opisuje rezultate 440 psihologijskih studija o testovima kojima se mjere postignuća i sposobnosti gdje su distribucije bile sve osim normalne.

Kod normalne distribucije aritmetička sredina i varijanca nisu ovisne jedna o drugoj, što znači da će varijanca povećamo li aritmetičku sredinu u normalnoj distribuciji ostati nepromijenjena. To nije slučaj kod nekih drugih distribucija.

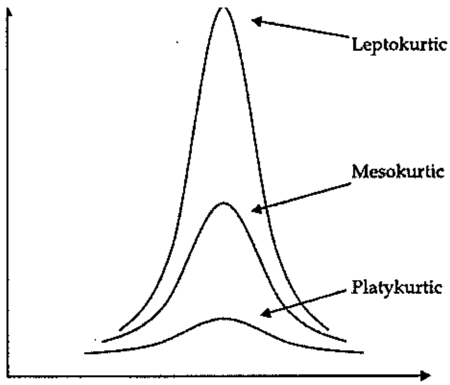
Normalna distribucija (slika 4) naziva se još Gaussova distribucija, prema Carlu Gustavu Gaussu, matematičaru i astronomu koji je 1803. godine uočio da su pogreške ponovljenih mjerenja objekata često normalno distribuirane. Iako je normalna distribucija dobila ime po njemu, neki istraživači zasluge za razvoj distribucije pripisuju prije svega francuskom matematičaru Abrahamu de Moivreu (1667.-1754.), koji je utvrdio da se binomna raspodjela približava normalnoj distribuciji kao granica (prema Black, 2010.). Također, u statističkoj se literaturi u definiranju normalne distribucije naglašava uloga francuskog matematičara Pierre-Simona de Laplacea. Štoviše, u nekim frankofonskim područjima normalna distribucija zove se po njemu: Laplace distribucija.



Slika 1. Infleksija krivulje normalne distribucije (slika prema: Ghilani, 2010., str. 41)

Jedna od karakteristika normalne distribucije mjesto je infleksije, odnosno točka s obje strane krivulje gdje je nagib krivulje najstrmiji, gdje iz konveksne prelazi u konkavnu (slika 1). Te su točke na krivulji mjesta infleksije i kod normalne distribucije udaljenost od aritmetičke sredine (sa svake strane do te točke) jednaka je ± 1 standardnoj devijaciji.

Posljedica je infleksije spljoštenost normalne distribucije (eng. kurtosis, slika 2).

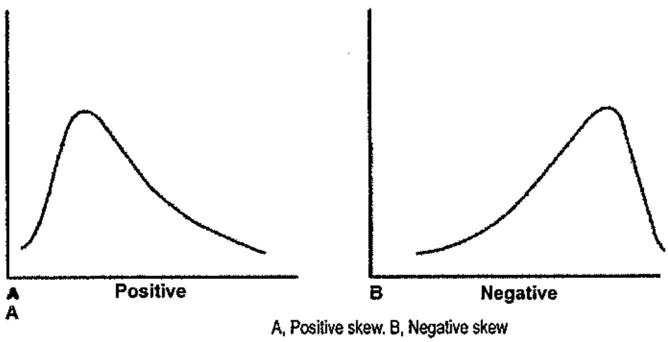


2. Spljoštenost krivulje /kurtosis/
(slika prema: Gaur i Gaur, 2009., str. 40)

Moors (1986.) opisuje kurtosis kao stupanj disperzije između točaka označenih na apscisi (x os) koje odgovaraju: aritmetička sredina ± standardna devijacija. Tako postoje tri vrste distribucije: *leptokurtična*, *mesokurtična (normalna)* i *platikurtična*. Raspon vrijednosti na apscisi koji pada između aritmetičke sredine i jedne standardne devijacije iznad i ispod aritmetičke sredine bit će veći za platikurtičnu, a najmanji za leptokurtičnu distribuciju. Direktni pokazatelj spljoštenosti distribucije standardna je devijacija: što je veća standardna devijacija, krivulja je spljoštenija, dok manja standardna devijacija uzrokuje šiljatost krivulje. Ta činjenica istraživačima odgoja i obrazovanja već iz pokazatelja deskriptivne statistike pruža mogućnost uvida u varijabilitet normalnosti distribucije.

Međutim, precizna i generalno upotrebljiva mjera zaobljenosti četvrti je centralni normirani moment – moment četvrtog reda. U statistici postoji više momenata, a određuju se razlikom između svakog pojedinog rezultata i aritmetičke sredine svih rezultata. Za statističko testiranje zaobljenosti – je li odstupanje od zaobljenosti statistički značajno, što rezultira nenormalnom distribucijom – predlaže se omjer između zaobljenosti i njezine standardne pogreške. Omjer ne smije biti veći od 3 (kod nekih autora i 2).

Osim spljoštenosti normalne distribucije, jedna je od njezinih karakteristika i simetričnost odnosno asimetričnost (eng. skewness, slika 3).



Slika 3. Asimetrija krivulje (skewness)
(slika prema: Lobiondo-Wood i Haber, 2006., str. 366)

Asimetričnost (skewness) mjera je stupnja asimetrije raspodjele. Asimetrija distribucije podrazumijeva nagnutost distribucije na lijevu ili desnu stranu; ako je jedan kraj krivulje distribucije razvučeniji od drugog, riječ je o asimetriji. Postoji pozitivna (desna) ili negativna (lijeva) asimetrija. Kad nema asimetrije, vrijednost je skewnessa 0.

Budući da su empirijske distribucije rijetko simetrične, koeficijent asimetrije, tj. centralni moment trećeg reda obično se kreće u intervalu [-2, 2]. U izrazito asimetričnim distribucijama poprima veće vrijednosti.

Simetričnost/asimetričnost distribucije može se promatrati prema odnosu aritmetičke sredine i medijana. Tako je distribucija pozitivno asimetrična kad je aritmetička sredina veća nego medijan te negativno asimetrična kada je aritmetička sredina manja od njega (Gaur i Gaur, 2009.).

Karl Pearson razvio je metodu za mjerenje skewnessa: Pearsonov koeficijent skeweneesa. Koeficijent uspoređuje aritmetičku sredinu i mod podijeljeno sa standardnom devijacijom: SKp = $\frac{\bar{X}-Mo}{\sigma}$. U teoriji ne postoji ograničenje za izračunavanje vrijednosti Pearson koeficijenta asimetrije, međutim formula ima određena ograničenja (Bajpai, 1971.). U praksi rijetko formula prezentira visoke vrijednosti i općenito se rezultati nalaze između ± 1 standardne devijacije. Treba istaknuti činjenicu da kod simetrične distribucije gdje je mod = medijan = aritmetička sredina vrijednost Pearsonova koeficijenta asimetrije jednaka je 0. Tako će za pozitivno asimetričnu distribuciju vrijednost biti u plusu, a za negativnu distribuciju vrijednost koeficijenta asimetrije bit će u minusu.

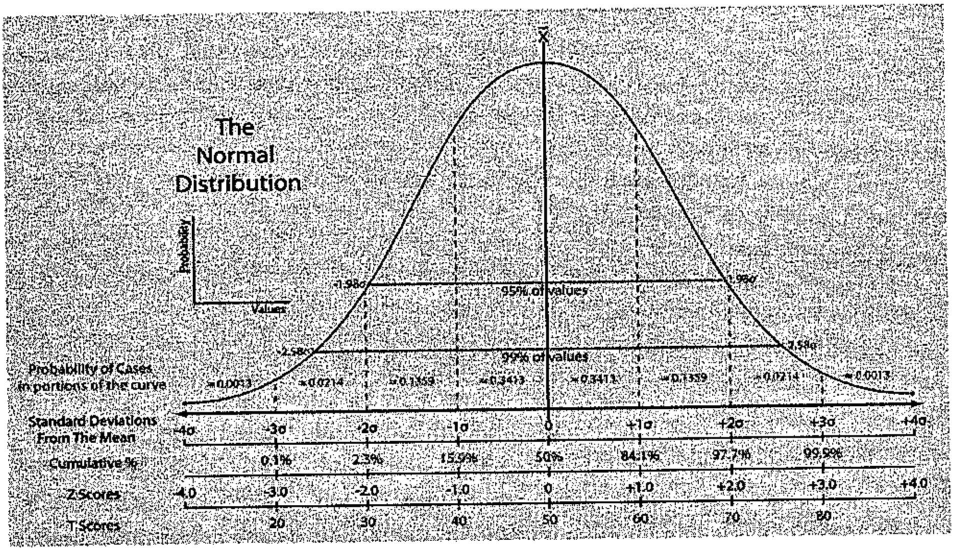
Pearsonov koeficijent skewnessa nije baš reprezentativna mjera asimetrije, a i neupotrebljiva je kod multimodalnih distribucija. Preciznija mjera asimetrije treći je centralni normirani moment. Za statističko testiranje asimetričnosti predlaže se da omjer između asimetrije i njezine standardne pogreške ne bude veći od 3 (2).

Karakteristike normalne distribucije (slika 4) možemo sumirati ovako:

1. Ima kontinuiranu distribuciju ili distribucije kontinuirane slučajne varijable X.
2. Zvonolike je krivulje. Normalna učestalost (gustoće vjerojatnosti) krivulje raste postupno, a zatim se maksimalno smanjuje na isti način. Simetrična je oblika oko aritmetičke sredine.
3. Krivulja je asimptotična: približava se X osi, ali je zapravo nikad ne dotiče (asimptotičnost).
4. Funkcija gustoće vjerojatnosti f(x) X ima formu

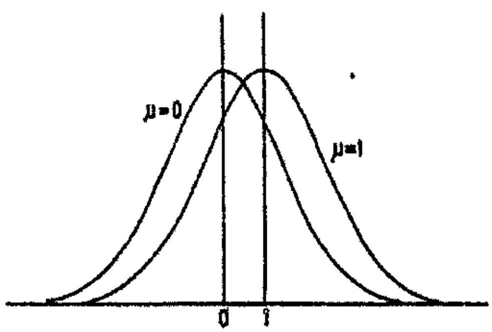
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

5. Uključuje parametre aritmetičke sredine i varijance.
6. Za vjerojatnost da se X nalazi u intervalu oko aritmetičke sredine vrijedi (slika 4):
   - **P ($\mu - \sigma < X < \mu + \sigma$)= 0.683**
   - **P ($\mu - 2\sigma < X < \mu + 2\sigma$)= 0.954**
   - **P ($\mu - 3\sigma < X < \mu + 3\sigma$)= 0.997**



Slika 4. Krivulja normalne distribucije (http://wapedia.mobi/en/File: The_Normal_Distribution.svg)

7. Promjena μ prema većoj vrijednosti (pod uvjetom da σ ostaje ista) uvjetuje pomicanje μ od centra u desno, a da visina krivulje ostaje ista (slika 5).



Slika 5. Pomicanje aritmetičke sredine

8. Mjere centralne tendencije: aritmetička sredina, medijan i mod podudaraju se u normalnoj krivulji (kao i za bilo koju simetričnu krivulju).
9. Ako je X normalno distribuiran s aritmetičkom sredinom, varijancom i standardnom devijacijom, tada je $z = \frac{x-\mu x-\mu}{\sigma \quad \sigma}$ normalno distribuirana s vrijednostima $\mu = 0$, $\sigma = 1$.
10. Točka je infleksije udaljena $\pm 1\sigma$ od aritmetičke sredine.

Osim normalne distribucije, postoje i druge teorijske distribucije, koje se mogu svrstati u dvije kategorije: distribucije diskontinuiranih ili diskretnih varijabli i distribucije kontinuiranih varijabli. Neke od njih aproksimativno su približne normalnoj, a druge su u potpunosti suprotne traženoj normalnosti. Jedna je od najučestalijih diskontinuiranih distribucija (uz binomnu) *Poisson distribucija*. Poissonova raspodjela proširenje je binomne distribucije i može se upotrebljavati kao aproksimacija. Jedno je od njezinih neobičnih svojstava to da je njezina standardna devijacija jednaka kvadratnom korijenu aritmetičke sredine. Otkrio ju je 1837. godine francuski matematičar Simeon-Davis Poisson (1781.–1840.). Najčešće je pozitivno asimetrična. Za razliku od normalne distribucije koja je određena aritmetičkom sredinom i varijancom, Poissonova je distribucija potpuno određena aritmetičkom sredinom. Kad je uzorak vrlo velik, a p blizu 0, Poissonova raspodjela približava se binomnoj distribuciji (binomna distribucija na velikom uzorku aproksimacija je normalne Gaussove distribucije). Razlika je u tome da kod binomne raspodjele znamo koliko se puta neki događaj zbio i koliko se puta nije zbio, a kod Poissonove distribucije znamo samo broj pojavljivanja određenog slučaja.

U istraživanju odgoja i obrazovanja često se susreće i pravokutna ili ravnomjerna distribucija (uniform distribution), u kojoj gotovo sve kategorije rezultata imaju iste frekvencije. To zna biti slučaj u pedagoškim istraživanjima gdje na ordinalnoj skali slaganja s određenom tvrdnjom ispitanici zaokružuju više od jednog odnosno sve slučajeve (što je pogrešno!) ili svi ispitanici daju iste odgovore (nema disperzije).

Isto tako, susrećemo *U distribuciju*, koja nalikuje velikom slovu U. Kod nje najmanje rezultata ima u sredini, a najveći su rezultati na krajevima distribucije. To također zna biti slučaj u istraživanjima odgoja i obrazovanja, jer se vrlo često ispituju nečiji stavovi. Tako naprimjer kad je riječ o ispitivanjima pozitivnog i negativnog stava o nekoj eksponiranoj stručno-znanstvenoj ili medijskoj pojavi i/ili anomaliji, ispitanici se često opredjeljuju za krajnje pozitivan ili negativan stav, što npr. na Likertovoj skali od pet stupnjeva znači da najmanje ima onih ispitanika koji nemaju stav, što je numerička vrijednost 3 na dotičnoj skali.

T-distribucija unimodalna je simetrična distribucija čija je aritmetička sredina 0, ali standardna devijacija nije 1. Dobila je ime po William Gossetu, radniku u pivovari u Dublinu koji je pod pseudonimom Student otkrio neadekvatno normalne krivulje za male uzorke. Naime, radeći u pivovari i testirajući kvalitetu piva, ustanovio je da se vjerojatnost distribucije kod malih uzoraka razlikuje od normalne distribucije. Kako je svoj rad publicirao pod pseudonimom Student, ta se distribucija u njegovo ime zove i *Student distribucija.* Kod t-distribucije indikativno je da standardna devijacija varira, pa je kod malih slučajnih uzoraka koji su prikazani grafom t-distribucija više šiljasta – leptokurtična. Kako veličina uzorka raste, t-distribucija počinje poprimati oblik standardne normalne distribucije. Schumacker i Akers (2001.) upozoravaju da zapravo t vrijednosti i z vrijednosti postaju identične kad je veličina uzorka 10000, što je unutar 0,001 pogreške aprokismacije. Istraživači često primjenjuju t-distribuciju za male i velike uzorke, jer kako raste veličina uzorka, dotične distribucije postaju identične. Jedan od problema kod istraživanja odgoja i obrazovanja u tomu je što su uzorci često male i/ili srednje veličine. Ta veličina uzorka vrlo je diskutabilna, identično kao i kod centralnog graničnog teorema. Tako npr. Šošić i Serdar (2002.) opisujući *t-distribuciju* naglašavaju da veličina uzorka mora biti veća od 30 kako bi se *t-distribucija* približila normalnoj. Jedan je od neslužbenih neempirijskih savjeta istraživačima odgoja i obrazovanja kad imaju veliku populaciju u istraživanju da im i uzorak (bez obzira koje vrste) bude iznad 100 ili više. U tom slučaju zbog već navedenih razloga vjerojatnost približnosti normalnoj distribuciji veća je nego kod malih uzoraka.

Hi-kvadrat distribucija bazira se na utvrđivanju statistički značajnih razlika između empirijskih i teorijskih frekvencija, odnosno vrijednosti varijabli. Hi--kvadrat distribucija ima vrijednosti u rasponu od 0 do ∞, i vrijednosti su uvijek pozitivne. Kod hi-kvadrat distribucije aritmetička je sredina jednaka broju stupnjeva slobode. To je izraženo asimetrična distribucija za mali broj stupnjeva slobode, a s povećanjem broja stupnjeva slobode ona se polako približava simetričnosti. Kod nje je karakteristično da se varijance ne pokoravaju normalnoj raspodjeli.

Također, postoji još cijeli niz distribucija koje su više definirane za određena matematička određenja: beta, Weibull, eksponencijalna, lognormal, step, triangular itd. Treba istaknuti F-distribuciju (F-omjer), bimodalnu i binomnu distribuciju koje su ponekad prisutne u istraživanjima odgoja i obrazovanja.

Testiranje normalnosti distribucije

Postoje dva glavna načina za provjeru normalnosti distribucije: grafičkim i/ili statističkim postupkom. Grafička procjena distribucija opravdana je ako je veličina uzorka iznad 100. Međutim, kod nje uvijek postoji latentna opasnost od donošenja pogrešnog zaključka o normalnosti, jer je Gaussova krivulja matematički definirana. Grafički prikaz distribucije određene varijable indikator je istraživaču o možebitnoj normalnosti, ali je daleko pouzdanije provesti određeni statistički test, iako i kod njih treba biti oprezan.

Jedan je od najpoznatijih postupaka testiranja normalnosti distribucije u društvenim znanostima Kolmogorov-Smirnov test (K-S). Riječ je o neparametrijskom testu za testiranje pripadaju li dva uzorka istoj populaciji, odnosno je li se dva skupa slučajno razlikuju. Za primjenu K-S testa potrebno je imati parametre aritmetičke sredine i varijance. Zasniva se na usporedbi dviju kumulativnih frekvencija oba skupa, a testira se uz pomoć hi-kvadrata (kod većih uzoraka) ili pomoću posebnih tablica (Kolesarić i Petz, 2003.). Može se primijeniti i za testiranje jednog uzorka u slučaju da želimo istražiti je li uzorak slučajan.

Primjena Kolmogorov-Smirnov testa u istraživanju odgoja i obrazovanja treba biti neizostavan preduvjet za odabir statističkih metoda obrade. To se prije svega odnosi na odluku o primjeni parametrijskih ili neparametrijskih statističkih metoda. Neparametrijska statistika i neparametrijske metode nisu ovisne o normalnosti distribucije te su ujedno slabije osjetljive u određenim osobinama koje se mjere. Kolmogorov-Smirnov test interpretira se s obzirom na p vrijednost, i to:

- ako je $p < 0{,}05$, zaključujemo da se podaci (ili uzorci) značajno razlikuju od normalne distribucije odnosno odbacujemo nul hipotezu jer je manje od pet posto mogućnosti da je nul hipoteza točna;
- kad je $p > 0{,}05$, zaključujemo da su podaci (ili uzorci) normalno distribuirani i ujedno potvrđujemo nul hipotezu o nepostojanju razlika.

Treba istaći da pomoću (nijednoga) stat. testa ne dokazujemo točnost/istinitost postavljene hipoteze nego samo utvrđujemo uz koju je vjerojatnost pogreške možemo prihvatiti ili odbaciti.

U istraživanjima odgoja i obrazovanja K-S test uživa popularnost iako danas postoje i drugi testovi za testiranja normalnosti distribucije, koji se upotrebljavaju s obzirom na određene specifičnosti:

- **Lilliefors test** (modifikacija Kolmogorov-Smirnov testa): bolje ga je upotrebljavati ako se parametri gustoće, aritmetička sredina i varijanca varijable, ocjenjuju na osnovi podataka dobivenih na uzorku (prema Tenjević, 2002.). Test mjeri maksimalnu udaljenost između promatrane distribucije i normalne distribucije s istom standardnom devijacijom i aritmetičkom sredinom, i procjenjuje je li ta udaljenost veća nego što se može objasniti slučajnošću.
- **Shapiro-Wilks test**: najbolje ga je upotrebljavati za testiranje normalnosti za male i srednje uzorke (Canover, 1999.). Rezultat ovog testa nije jasan dokaz o normalnosti ili ne-normalnosti nego samo korisna indicija za dalje testiranje.

- **Anderson-Darling Normality Test**: upotrebljava se za određivanje može li se krivulja koja opisuje neki skup podataka adekvatno aproksimirati normalnom distribucijom. Test uspoređuje kumulativnu normalnu distribucijsku funkciju koja odgovara promatranim podacima sa stvarnom kumulativnom funkcijom (Rust i sur., 1989.). Anderson-Darlingov test upotrebljava specifičnu distribuciju u izračunavanju kritičkih vrijednosti. Prednost ovakva izračuna u tomu je što je test osjetljiviji, a nedostatak je u tomu što se kritičke vrijednosti moraju računati za svaku distribuciju. To je modifikacija Kolmogorov-Smirnov (K-S) testa i daje više težine na rep krivulje nego K-S test. Anderson-Darling test bazira se na izračunu kritičnih vrijednosti koje se nalaze u tablicama.

Odabir odgovarajućeg testa za testiranje normalnosti distribucije ovisi o veličini uzorka i o istraživačevoj opredijeljenosti za određenu statističku paradigmu: konzervativnu ili suvremenu. Tako Shapiro-Wilks test tendira odbacivanju nul hipoteze više što bi netko "želio", a hi-kvadrat i K-S test više su konzervativni (Riffenburgf, 2006.). Međutim, bez obzira na to za koji se test opredijelili, treba biti oprezan u donošenju zaključaka s obzirom na normalnost distribucije, timviše kad je riječ o graničnim vrijednostima.

U istraživanjima odgoja i obrazovanja najčešće se upotrebljava nominalna i ordinalna skala. Klasična Likertova skala najčešće se temelji na skali procjene s pet kategorija, tj. stupnjeva. Broj kategorija utječe na varijabilnost – veći broj kategorija na skali procjene (npr. 7 ili 9) rezultira većom varijabilnosti. Međutim, značajniju posljedicu na normalnost podataka ima uravnoteženost skale. Ako kategorije na skali procjene nisu ekvidistantne, podaci više nisu ordinalni nego su nominalni (što nije u skladu s postavkama Likertove skale). U tom slučaju postoji opasnost da pojedine kategorije na skali procjene imaju različite pondere (u semantičkom smislu) prema predmetu mjerenja, što u konačnici dovodi do nepravilnosti (distorzije) distribucije. Treba naglasiti da mnogi parametrijski statistički testovi naglašavaju kao preduvjet omjernu i intervalnu skalu, iako se to često u istraživanjima odgoja i obrazovanja zanemaruje pa se čak multivarijatnim obradama testiraju hipoteze na nominalnoj skali, što je pogrešno i daje iskrivljene rezultate i zaključke koji slijede. Iako je osnovni preduvjet za provedbu parametrijskih testova normalnost distribucije, koja se, kako je opisano, rijetko pojavljuje u istraživanjima odgoja i obrazovanja, ni sama približnost normalnoj distribuciji neće dati krive podatke. Tako npr. Marusteri i Bacarea (2010.) ustvrđuju da će ANOVA testovi, t-testovi i ostali statistički testovi ispravno ispitivati čak i u slučaju da je raspodjela samo približna Gaussovoj (posebno kod velikih uzoraka, npr. $N > 100$) i ti se testovi rutinski primjenjuju na mnogim poljima znanstvenih istraživanja.

Za indikativnu provjeru takva stajališta služio sam se rezultatima istraživanja o preferenciji prema agresivnom ponašanju (Bionda i Opić, 2010.). Cilj tog istraživanja bio je usporediti razlikuju li se učenici četvrtoga i osmog razreda s obzirom na varijable preferencije prema agresivnom ponašanju. Iako je distribucija odstupala od normalne (za dotične prikazane varijable), paralelno je primijenjen parametrijski test analize varijance (ANOVA) i njegov neparametrijski pandan Kruskal-Wallis test s ciljem utvrđivanja diferencija među subuzorcima na odabranim varijablama (tablica 1). S obzirom na to da je relativno velik uzorak (N = 208), oba testa pokazuju u većini varijabli iste razine traženih statističkih diferencija na subuzorcima s obzirom na manifestne varijable (x1-x10), odnosno i zaključci su u većini varijabli isti.

Tablica 1. Usporedba primjene parametrijskog testa (ANOVA) i njegova neparametrijskog pandana (Kruskal-Wallis test)

Varijable	Testovi	Sig.
X1	ANOVA	,096
	KRUSKAL-WALLIS	,136
X2	ANOVA	,292
	KRUSKAL-WALLIS	,067
X3	ANOVA	,000*
	KRUSKAL-WALLIS	,000*
X4	*ANOVA*	***,046****
	KRUSKAL-WALLIS	,051
X5	ANOVA	,292
	KRUSKAL-WALLIS	,288
X6	ANOVA	,000*
	KRUSKAL-WALLIS	,000*
X7	ANOVA	,000*
	KRUSKAL-WALLIS	,000*
X8	ANOVA	,521
	KRUSKAL-WALLIS	,427
X9	ANOVA	,095
	KRUSKAL-WALLIS	***,041****
X10	ANOVA	,822
	KRUSKAL-WALLIS	,663

*p < 0,05

Kao što je vidljivo iz tablice 1, većina varijabli upućuje na podudaranja u diferencijama na subuzorcima s obzirom na stipulirane varijable (X3, X6, X7), i to kod parametrijskog testa ANOVA i njegova neparametrijskog pandana, Kruskal-Wallis testa. Međutim, varijable X9 i X4 upućuju na to da kod primjene neparametrijskog testa postoje diferencije, za razliku od ANOVA-e. To je signifikantno s obzirom na donošenje zaključaka koji slijede primjenom parametrijske i neparametrijske statistike, koja nije podložna preduvjetu normalnosti distribucije.

Međutim, u inferencijalnoj statistici postoji teorem koji pridonosi izvođenju valjanih zaključaka primjenom parametrijske statistike i kad se ne radi o normalnoj distribuciji. Riječ je o centralnom graničnom teoremu koji pokazuje da se aritmetičke sredine velikog broja uzoraka iste veličine, a uzetih iz neke populacije, distribuiraju po normalnoj raspodjeli čak i onda kad populacija i značajno odstupa od normalne (Kolesarić i Petz, 2003.). To je stoga što standardna pogreška sampling distribucije opada kako raste veličina uzorka. Centralni granični teorem utemeljio je Laplace 1810. godine. Kao dio inferencijalne statistike pruža istraživačima u odgoju i obrazovanju možebitne zaključke i onda kad rade s uzorcima (populacijama) koji nisu normalno distribuirani. Petz (1997.) naglašava da će procjena neke vrijednosti biti točnija što je uzorak veći i što je pojava varijabilnija. Sukladno važnosti navedenog teorema, Micceri (1989.) naglašava da bi bez centralnog graničnog teorema inferencijalna statistika koja pretpostavlja normalnost distribucije (npr. ANOVA, Two sample t-test, regresijska analiza...) bila u potpunosti neupotrebljiva, posebno u društvenim znanostima, gdje većina mjerenja nije normalno distribuirana. Naravno, nameće se pitanje koliko je dovoljno velik uzorak da centralni granični teorem vrijedi. Tako npr. Smith i Wells (2006.) sumnjaju u statističku sugestiju gdje čak uzorak veličine N = 30 uzrokuje aproksimativnu normalnu distribuciju. Razlog ovakvu statističkom stajalištu leži u činjenici da u inferencijalnoj statistici uzorci veći od 30 (kod nekih 50) ulaze u kategoriju većih uzoraka.

Iako se u statističkoj literaturi vode rasprave gdje je donja granica veličine takva uzorka, zaključit ću raspravu zanimljivim stajalištem prof. Arshama Hosseina sa Sveučilišta u Baltimoreu, koji kaže da nije moguće decidirano utvrditi koja je donja granica veličine uzorka kad centralni granični teorem djeluje (Arsham, 2005.). U istraživanju odgoja i obrazovanja npr. kad je riječ o agresivnosti, motivaciji, dokimologiji, procesu učenja, kurikulumu i sl., gdje imamo na raspolaganju veliku populaciju, treba težiti povećanju uzorka jer se izlažemo manjoj pogrešci s obzirom na donošenje zaključaka kad je riječ o distribuciji podataka. U onim specifičnim područjima istraživanja odgoja i obrazovanja gdje neminovno imamo mali uzorak (N < 30 [50]) primjenjuju se robusni statistički testovi ili je riječ o neparametrijskim metodama.

Također, već postojeći uzorak određenim statističkim programima možemo normalizirati na način da podatke transformiramo u nove vrijednosti kako bismo poboljšali distribuciju. Svrha je takvih transformacija stabilizacija varijance, odnosno izjednačivanje varijanci subuzoraka i eventualno linearizacija odnosa između varijabli. Riječ je o metodološki opravdanom postupku koji ponekad djelomično približava distribuciju normalnosti te omogućuje primjenu parametrijske statistike (Sokal i Rohlf, 1995.). Riječ je npr. o arcsine (angular – kutna transformacija, posebno je korisna za postotke i proporcije), arcsine square, root, angular, inverse sin transformacijama i sl. (De Muth, 2006.):

- Logarithms transformacija – normalizira eksponencijalni rast vrijednosti i prikladna je ako varijanca raste paralelno s aritmetičkom sredinom
- Square roots – transformiraju podatke iz Poissonove u normalnu distribuciju
- Inverse transformation – računaju se inverzne vrijednosti varijabli 1/x. To zapravo velike brojeve čini malima dok male brojeve velikim vrijednostima (Osborne, 2002.).

Naravno da se kod nekih istraživača nameće opravdano pitanje je li normalizacija podataka navedenim transformacijama fiktivna te će i rezultati koji slijede biti takvi. U tom smjeru Sokal i Rohlf (1995.) naglašavaju opravdanost primjene takvih transformacija.

Umjesto zaključka

U društvenim znanostima, istraživanjima odgoja i obrazovanja, stvarnu normalnost distribucije, kako je navedeno, teško je postići. Međutim, upravo testiranje normalnosti određenim statističkim testovima (K-S, Shapiro-Wilks i dr.) daje nam uvid u distribuciju podataka, odnosno mogućnost za njezinu statističku transformaciju te odabir određenih pandana neparametrijskih statističkih testova.

Upuštati se u generalizacije u istraživanjima odgoja i obrazovanja treba oprezno, zbog cijelog niza razloga, a jedan je od njih zadovoljenje preduvjeta o normalnosti distribucije. Provjera normalnosti distribucije, uz npr. testiranje pouzdanosti upitnika, preduvjet je za odabir odgovarajućih statističkih metoda obrade podataka (statističkih testova), iako je česta pojava da se ti postupci neopravdano izostavljaju. S obzirom na rijetko zastupljenu normalnu distribuciju, istraživači odgoja i obrazovanja nalaze se pred svojevrsnim izazovom, što je određeni teret. Međutim, taj teret zapravo je dio znanstvene predikcije koja iz određenih parametara uzorka prejudicira određene generalizacije na cjelokupnu populaciju. I bez problema zadovoljenja normalnosti distribucije tu treba biti posebno oprezan zbog cijelog niza parazitarnih faktora koji djeluju više ili manje na svako istraživanje u društvenim znanostima. Kako bismo smanjili mogućnost donošenja pogrešnih znanstvenih zaključaka, testiranje normalnosti distribucije preduvjet je za odabir testova obrade podataka. Međutim, nakon testiranja i uvida u normalnost distribucije, te ako ona i nije normalna, ostaje metodološki opravdana statistička mogućnost transformacije koja može poboljšati normalnost. Ako i tada distribucija ostaje izvan normalnosti, treba pristupiti neparametrijskim testovima (tablica 2).

Tablica 2. Parametrijski i neparametrijski pandan testovi (De Vaus, 2004.)

Svrha analize	Normalna distribucija *Parametrijski testovi	Ne-normalna distribucija *Neparametrijski testovi
Diferencije između dvije nezavisne skupine	• t-test	• Wald-Wolwofitz test • Mann-Whitney test • Kolmogorov-Smirnov (2 uzorka Z test)
Razlike između dviju ili više nezavisnih skupina	• Analiza varijance • F test	• Kruskal-Wallis (analize ranga) • Median test
Razlike između dva zavisna uzorka	• t-test za zavisne uzorke	• Sign test • Wilcoxon odgovarajući pair-test • McNemarov hi-kvadrat test
Dvije varijable mjerene na istom uzorku	• Ponovljeno mjerenje ANOVA	• Friedmanova two-way analiza varijance • Cochran Q test
Povezanosti između varijabli	• Pearsonov r (za intervalne varijable)	• Spearman Rho • Kendall's Tau • Gamma • Hi-kvadrat, Phi, Fisherov egzaktni test

Dobra je okolnost u istraživanjima odgoja i obrazovanja granični teorem koji omogućuje primjenu parametrijskih testova kad je uzorak veći od 100 iako je distribucija nenormalno distribuirana (De Vaus, 2004., 295). Međutim, u literaturi je to kod raznih autora diskutabilno te treba biti oprezan u takvu tumačenju. Ali analogno takvu zaključku, činjenica je da treba težiti većim uzorcima u istraživanjima odgoja i obrazovanja. Neki istraživači čak iz opravdane znanstvene radoznalosti za veći broj varijabli istovremeno primjenjuju parametrijske i neparametrijske testove, kako bi anulirali određene varijable koje se na razini pandan testova razlikuju s obzirom na tražene diferencije.

Međutim, iako je cilj ovog rada naglasiti važnost normalne distribucije, ona ipak nije uvijek poželjna u istraživanjima odgoja i obrazovanja. Npr. kad je riječ o distribuciji postignuća (školski uspjeh). Na razini razrednog odjeljenja ne želimo dobiti normalnu distribuciju, nego negativnu lijevo asimetričnu distribuciju što više pomaknutu udesno, dočim na razini cijele škole (nažalost) moramo očekivati normalnost distribucije. Isto tako, možemo biti zadovoljni ako je distribucija postignuća na nekom školskom testu lijevo asimetrična (pod uvjetom da test nije prelagan).

Literatura

1. Arsham, H. (2005.): *System Simulation: The Shortest Route to Applications*, Version 9.1. Retrieved 5/31/06 from http://home.ubalt.edu/ntsbarsh/simulation/sim.htm.
2. Bajpai. N. (1971.): *Business Statistics*. Upper Saddle River NJ:Pearson education.
3. Bionda, M., Opić, S. (2010.): *Preferencije učenika osnovne škole prema agresivnom ponašanju*. 4. međunarodna konferencija o naprednim i sustavnim istraživanjima. 11. Dane Mate Demarina: *Očekivanja, postignuća i perspektive u teoriji i praksi ranog i primarnog odgoja i obrazovanja*. U Zagrebu 11.-13. studenog – u objavi.
4. Black, K. (2010.): *Business Statistics: for Contemporary* Decision Making. 6th edition. John Willey & Sons. ISBN 13 978-0470-4091-3.
5. Conover, W. J. (1999.): *Practical Nonparametric Statistics* (3rd edition). Wiley.
6. De Muth, E. James (2006.): *Basic Statistics and Pharmaceutical Statistical Application*. 2nd edition. Chapman and Hall/CRC.
7. De Vaus, D. (2004.): *Analyzing Social Science Data*: 50 Key Problems in Data Analysis. London: Sage publications.
8. Gaur, S. Ajai; Gaur, S. Sanjaya (2009.): *Statistical Methods for Practice and Research*: A guide to dana analysis using SPSS (second edition). New Delhy: Mohan Cooperative Industrial Area (Business books from Sage).
9. Ghilani, C. D. (2010.): Adjustment Computations Spatial Data Analysis (5th edition). New Jersey: John Willey & Sons.
10. Kolesarić, V., Petz, B. (2003.): *Statistički rječnik: tumač statističkih pojmova. Jastrebarsko*: Naklada Slap.
11. Lobiondo-Wood, G; Haber, J. (2006.): *Nursing Research; Methods and Critical Appraisal for Evidence-Based Practice.* /6th edition/ Missouri: Mosby Elsevior.
12. Marusteri, M.; Bacarea, V. (2010.): *Kako odabrati pravi test za procjenu statističke značajnosti razlike između skupina?* Biochemia Medica 20 (1), 15-32.
13. Medhi, J. (2005.): Statistical Methods: An Introductory Text. New Delhy: New Age International (p) Limited, Publishers.
14. Micceri, T. (1989.): The Unicorn, The Normal Curve, and Other Improbable Creatures. *Psychological Bulletin,* 105 (1), 156-166.

15. Moors, J. J. A. (1986.): The meaning of kurtosis: Darlington reexamined. *The American Statistician*, 40, 283-284.
16. Osborne, J. (2002.): Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 8 (6).
17. Petz, B. (1997.): *Osnovne statističke metode za nematematičare*. Jastrebarsko: Naklada Slap.
18. Riffenburgf, H. Robert (2006.): Statistics in Medicine: 3rd edition. Burlington, MA, 01803 USA: Elsevier Academic Press.
19. Rust, W. Steven; Todt, W.; Frederic, R. K.; Harris, B.; Heal, D.; Vangel, M. (1989.): *Statistical Methods for Calculating* Material Allow Ablesfor MIL-HDBK-17. In: The Methods and Design Allowables for Fibrous Composites (Ed. Chaims, C. Christos), 3rd edition, ASTM special technical publication.
20. Schumacker, E., Rundall; Akers, A. (2001.): Understanding Stasitical Concept Using S Plus.mahvah NJ: Lawrence Erlbaum Associates.
21. Smith, R. Zachary; Wells, S. Craig. (2006.): *Central Limit Theorem and Sample Size*. University of Massachusetts Amherst, Paper presented at the annual meeting of the Northeastern Educational Research, Association, Kerhonkson, New York, October 18-20, 2006.
22. Sokal, R. R; Rohlf, F. J. (1995.): Biometry: *The principles and practice of statistics in biological research*. 3rd edition. New York; Freeman.
23. Šošić, I., Serdar, V. (2002.): *Uvod u statistiku*. Zagreb: Školska knjiga.
24. Tenjević, L. (2002.): *Statistika u psihologiji*. Beograd: Centar za primjenjenjenu psihologiju, Društva psihologa Srbije.
25. (http://wapedia.mobi/en/File:The_Normal_Distribution.svg)

UDC 37.012:311.2
Review article
Accepted: 12th February 2011
Confirmed: 20th May 2011

DISTRIBUTION NORMALITY TESTING IN THE STUDIES OF EDUCATION AND UPBRINGING

Siniša Opić, Assistant Professor
E-mail address: sinisa.opic@ufzg.hr

Summary: *Distribution normality testing, including the homogeneity of variance, is an inevitable parameter in the studies of education and upbringing due to the choice of certain statistic parametric or non-parametric tests for hypotheses testing. However, it has been unjustifiably excluded which can eventually cause distorted results (generalizations). The work describes statistic values of normal distribution and its modalities: kurtosis and skewness. It also describes other theoretical distributions which are often seen in the studies of education and upbringing: Poisson, uniform, t-distribution, hi-square and U-distribution. The tests for distribution normality have been as well elaborated: Kolmogorov-Smirnov test (K-S), Lilliefors test, Shapiro-Wilks test Anderson-Darling Normality Test.*

As an empirical evaluation of statistically significant values of parametric and non-parametric statistics, the statistical values of ANOVA and its non-parametric double Kruskal-Wallis test have been compared. The normalization tests, i.e. the transformation values due to the improvement of distribution normality have been partially emphasized. There is a theorem in inferential statistics which contributes to making conclusions through the application of parametric statistics, even if normal distribution does not occur, i.e. the central border theorem. Its role in the study of education and upbringing has been specifically described.

Key words: *data distribution, normal distribution, kurtosis, skewness, normality evaluation tests, transformations, central border*