TESTIRANJE STATISTIČKIH HIPOTEZA

8. glava

CILJEVI POGLAVLJA

Nakon čitanja ovoga poglavlja bićete u stanju da:

1. *shvatite logiku, značaj i ograničenja statističkog testiranja hipoteza*
2. *razumete razliku između testiranja na osnovu kritičnih vrednosti i p-vrednosti*
3. *primenite najpoznatije parametarske testove u slučaju jednog i dva uzorka*
4. *protumačite računarski izlaz testiranja hipoteza ako koristite statistički softver ili Excel*

U ovom poglavlju upoznaćemo se sa jednim od najznačajnijih, najkorisnijih, ali ujedno i najkontroverznijih statističkih metoda, testiranjem hipoteza. Testiranje hipoteza ima veliku primenu kako u praktičnim tako i u teorijskim istraživanjima. Danas je, na primer, u SAD nezamislivo dobiti odobrenje za distribuciju novog leka od FDA (Food and Drug Administration) pre detaljnog ispitivanja efektivnosti i sigurnosti tog leka, u kome ključnu ulogu ima testiranje hipoteza. Takođe, publikovanje radova u referentnim stručnim časopisima iz različitih oblasti (medicine, farmacije, psihologije, pedagogije, biologije, poljoprivrede, istraživanja tržišta) često podrazumeva da su tvrđenja i hipoteze potkrepljeni rezultatima nekog statističkog testa.

Podsetimo se da je cilj statističkog zaključivanja da se na osnovu informacije iz uzorka dođe do što je moguće pouzdanije i preciznije informacije o populaciji. U prethodnom poglavlju smo detaljnije govorili o ocenjivanju nepoznatih parametara skupa. Rekli smo da se ono koristi kada istraživač nema nikakvo prethodno saznanje o relevantnim karakteristikama posmatranog skupa. Krajnji proizvod ocenjivanja je interval poverenja u kome se sa odgovarajućom pouzdanošću nalazi nepoznati parametar.

Drugа oblast statističkog zaključivanja, testiranje statističkih hipoteza, koristi se kada o skupu imamo neku prethodnu informaciju. **Cilj testiranja je**

da se ispita prihvatljivost nekog tvrđenja ili pretpostavke koja se tiče karakteristike jednog ili više osnovnih skupova. Na primer, testiranjem se može proveriti :

- da li je osnovano tvrđenje proizvođača da prosečno trajanje baterija za mobilni telefon Nokia N95, pre dopunjavanja, iznosi najmanje 380 časova,
- da li 35% mladih u Srbiji smatra da je najveći problem u našoj zemlji alkoholizam,
- da li je prosečna dužina studiranja na privatnim i državnim univerzitetima jednaka,
- da li visina studenata Vašeg univerziteta približno sledi normalan raspored,
- postoji li veza između broja nezaposlenih i broja kriminalnih dela,
- da li uzimanje antioksidanata usporava starenje, i sl.

Primetimo da na svako od postavljenih pitanja postoje samo dva moguća odgovora: da ili ne.

Na osnovu rečenog možemo dati generalnu preporuku da se testiranje hipoteza primenjuje ako su istovremeno ispunjena sledeća tri uslova:

a) Raspolažemo nekim prethodnim saznanjem (pretpostavkom) o karakteristici skupa;
b) Na istraživačko pitanje (problem) postoje samo dva moguća odgovora, da ili ne;
c) Problem se rešava korišćenjem slučajnog uzorka.

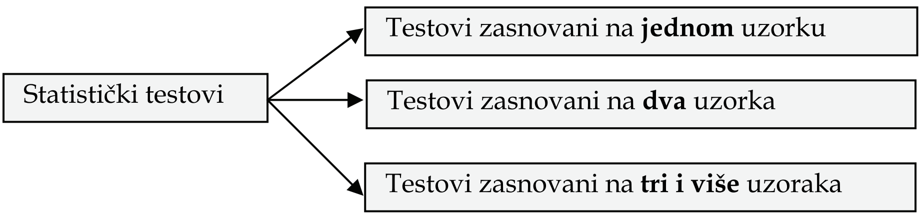
Kao što smo više puta naglasili, tačnu vrednost parametra skupa možemo odrediti samo na osnovu svih elemenata skupa. Zato smo zaključak ocenjivanja uvek donosili sa rezervom, tj. tvrdili smo sa pouzdanošću manjom od 100% da interval poverenja sadrži nepoznati parametar. Analogno, da bismo bili potpuno sigurni u tačnost određene pretpostavke ili tvrđenja o karakteristici populacije, morali bismo da izvršimo popis. Kako iz već poznatih razloga to najčešće ne činimo, rezultat testiranja koji donosimo na osnovu informacije iz uzorka ne može biti da je polazna pretpostavka tačna ili pogrešna. Umesto toga, u zaključku sa odgovarajućim rizikom tvrdimo da je pretpostavka prihvatljiva ili da nije prihvatljiva.

Statistička hipoteza je precizno formulisana pretpostavka (iskaz) o nekoj važnoj osobini jednog ili više skupova. Statistički metod koji na sistematski način empirijske podatke uzorka koristi radi utvrđivanja prihvatljivosti hipoteze naziva se **testiranjem hipoteze**, a sama procedura **statističkim testom**.

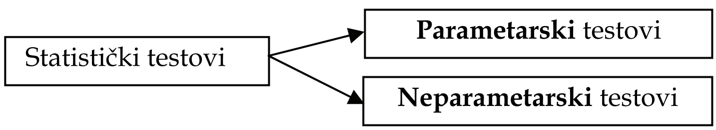
8.1 KLASIFIKACIJA TESTOVA I IZBOR OPTIMALNOG TESTA

Sve statističke testove možemo klasifikovati po tri osnova:

- Prema **prirodi problema** koje rešavamo razlikujemo: *testove o parametrima skupa* (aritmetičkoj sredini, proporciji, medijani, varijansi itd.), o *obliku rasporeda skupa*, o *nezavisnosti dva obeležja*, o *slučajnosti uzorka*, itd.
- Prema **broju uzoraka** na kojima zasnivamo testiranje testove delimo u tri grupe:



- Prema **vrsti i jačini** preduslova na kojima se zasnivaju razlikujemo:



Zadržimo se za trenutak na trećoj podeli.

Parametarski testovi se ponekad nazivaju klasičnim testovima, jer su formulisani znatno pre neparametarskih, početkom XX-og veka. Njihova teorijska podloga se bazira na tadašnjem uverenju istraživača da u prirodi i društvu preovladava samo jedan, normalan raspored. Otuda **parametarski testovi imaju zajednički preduslov da osnovni skup** (ili skupovi) kome pripada analizirani uzorak (odnosno uzorci) **sledi normalan raspored**. Nasuprot njima, neparametarski testovi, koji su nastali uglavnom sredinom XX veka, predstavljaju logičnu reakciju statističara na činjenicu da se pretpostavka o „normalnosti“ svih pojava pokazala netačnom. Ovi testovi se zasnivaju na mnogo blažim uslovima o obliku osnovnog skupa, a mogu se primenjivati i na kvalitativna obeležja.

Zbog velikog broja različitih testova, svakom korisniku statistike neminovno se nameće jednostavno pitanje: Koji test treba da primenim da bih rešio svoj istraživački problem? Odgovor nije trivijalan, jer je u statističkoj teoriji formulisan ne samo veliki broj testova sa različitom namenom, već se ponekad ista hipoteza može proveravati primenom više različitih testova. Budući da svaki test počiva na specifičnim preduslovima koji moraju biti ispunjeni da bi rezultat bio validan, treba izabrati onaj test koji je optimalan u datoj situaciji, a to znači test za koji su preduslovi primene zadovoljeni. Recimo, ipak, da je ovakav pristup testiranju relativno nov. Naime, tokom većeg dela XX veka korišćeni su isključivo parametarski testovi, pri čemu je uslov normalnosti skupa retko proveravan zbog komplikovane i nepouzdane procedure. Situacija se potpuno izmenila u poslednjih deset godina. Sa opštom dostupnošću računara i statističkog softvera, preduslovi primene testova se mogu lako proveriti, a sam test primeniti na podatke iz uzorka. Zbog toga naglasak u modernom testiranju hipoteza više nije puka primena nekog testa, već pre svega:

1) *ispitivanje da li su preduslovi na kojima se test zasniva ispunjeni* i

2) *korektna interpretacija rezultata*.

U ovom poglavlju razmatraćemo *parametarske testove* kojima se testira *aritmetička sredina* i *proporcija jednog* i *dva* osnovna skupa. U sledećem poglavlju posmatraćemo parametarski test zasnovan na više uzoraka kojim se testira hipoteza o jednakosti aritmetičkih sredina više osnovnih skupova. Parametarske testove ćemo, takođe, primenjivati i u regresionoj i korelacionoj analizi, koje su predmet razmatranja u 12. i 13. poglavlju. Neparametarskim testovima ćemo se baviti u 11. poglavlju.

8.2 POSTUPAK TESTIRANJA STATISTIČKIH HIPOTEZA

Pre nego što pređemo na proceduru testiranja daćemo dva primera, da bi se shvatile mogućnosti primene i logika samog testiranja.

PRIMER 8.1[1]: U jednom od najpoznatijih sudskih procesa u prošlom veku, poznati američki fudbaler i glumac O. J. Simpson je optužen za ubistvo svoje bivše supruge Nikol i njenog dvadesetpetogodišnjeg poznanika Ronalda Goldmena. Proces je trajao 133 dana, koštao je preko 15 miliona dolara i kontinuirano je prenošen na televiziji. Iako je policija smatrala da je prikupila dovoljno dokaza za njegovu krivicu, porota je 03. 10. 1995. presudila da Simpson nije kriv. Istoga dana Galup je sproveo anketu i procenio da je čak 56% američkih građana smatralo presudu pogrešnom, a samo 33% je podržalo.

Pravni sistem u velikom broju zemalja optuženog smatra nevinim sve dok se ne nađe dovoljno dokaza da se on proglasi krivim. Ovo ćemo statistički nazvati nultom hipotezom: optuženi nije kriv za počinjeno delo. Nultu hipotezu smatramo tačnom sve dok ne sakupimo dokaze da je ona pogrešna, odnosno da je tačna alternativna hipoteza: optuženi je kriv. Pri svemu tome,

[1] O suđenju O. J. Simpsona pogledati, na primer, http://www.law.umkc.edu/faculty/projects/ftrials/Simpson/simpson.htm

nastojimo da koliko je moguće smanjimo (poželjno na 0) verovatnoću da osudimo nevinu osobu, tj. da odbacimo istinitu nultu hipotezu. Nažalost, kao posledica, povećava se verovatnoća da načinimo drugačiju grešku: da oslobodimo optuženog iako je on stvarno kriv. Statistički rečeno, postoji verovatnoća da ne odbacimo pogrešnu nultu hipotezu. Za shvatanje koncepcije testiranja napomenimo da porota nikada ne donosi odluku da je optuženi nevin, već samo da nije kriv. Iako ovo izgleda kao finesa, razgraničenje je jako bitno u proceduri testiranja. Odmah napomenimo da **se u statistici nikada ne zaključuje da je nulta hipoteza stvarno istinita, već samo da nemamo dovoljno dokaza da je odbacimo**. Da li je O. J. Simpson stvarno bio nevin? Odgovor je verovatno negativan, ali porota nije našla dovoljno dokaza da ga proglasi krivim za ubistvo bivše supruge.

PRIMER 8.2: Državni tužilac u Indijanapolisu u SAD je dobio anonimnu prijavu da je građevinska firma koja je gradila puteve u gradu prevarila gradsku vladu. Na osnovu sklopljenog ugovora, debljina asfalta je u proseku trebalo da bude 10 inča. Tužiocu je bio potreban dokaz da ubedi porotu da je prosečna debljina puta manja od 10 inča. Naravno, otkopavanje čitavog puta nije dolazilo u obzir, jer osim ogromnih troškova zaključak je mogao biti da je kompanija ispoštovala ugovor. Zbog toga je uzet uzorak od 100 delova, svaki prečnika jednog inča. Pokazalo se da je prosečna debljina u uzorku bila 9,5 inča. Da li je ovo bilo dovoljno da se tretira kao ubedljiv dokaz protiv kompanije? Porota je smatrala da jeste i nekoliko službenika te firme je završilo u zatvoru. U ovom primeru nulta hipoteza se može postaviti u vidu da je prosečna debljina puta 10 inča, a alternativna da je manja od 10 inča.

Da sumiramo: prilikom testiranja neke tvrdnje ili pretpostavke postavljamo dve suprotstavljene hipoteze, nultu i alternativnu. Zatim polazimo od pretpostavke da je nulta hipoteza istinita, biramo uzorak i tražimo empirijske dokaze protiv nulte hipoteze. Ako ih pronađemo, onda ćemo odbaciti nultu hipotezu i usvojiti alternativnu. Ali, ako ne nađemo dovoljno čvrste dokaze, možemo reći samo jedno: **nismo našli dovoljno dokaza da odbacimo nultu hipotezu**. Bilo bi pogrešno tvrditi da je nulta hipoteza tačna.

Postupak testiranja statističkih hipoteza sprovodimo u šest etapa, kao na Dijagramu 8.1.

Za prosečnog korisnika statistike verovatno je najteža druga etapa, zbog čega smo u Dodatku priložili posebno "stablo izbora statističkog metoda" koje čitaocu pomaže da, ispitujući ispunjenost uslova na kojima se testovi zasnivaju, na metodičan način izabere odgovarajući test.

Kao što smo već naglasili, najvažnije etape su treća i šesta. Naime, ako preskočimo treću etapu, moguća posledica je upotreba neodgovarajućeg testa, čijom primenom možemo dobiti nevalidne rezultate. Nekorektno tumačenje rezultata u šestoj etapi je isto tako opasno, jer može biti uzrok pogrešnog istraživačkog zaključka, pa posledično i neosnovane poslovne odluke.

1. Formulišu se **nulta** i **alternativna** hipoteza i bira **nivo značajnosti**, α.

↓

2. Uzima se slučajan uzorak, i na osnovu analiziranja postavljenog problema **bira po istraživačevom mišljenju optimalni test**.

↓

3. Proverava se da li su ispunjeni preduslovi na kojima se izabrani test zasniva. Ako nisu, prekida se proces testiranja i bira neki drugi test, koji počiva na blažim, ispunjenim uslovima.

↓

4. Izračunava se tzv. **statistika testa**. Određuje se ***p*-vrednost** (alternativno, određuju se tzv. **kritične vrednosti** iz tablica odgovarajućeg rasporeda).

↓

5. Postavlja se **pravilo odlučivanja**. Na osnovu *p*-vrednosti (alternativno kritičnih vrednosti) **donosi se odluka** o odbacivanju ili neodbacivanju nulte hipoteze.

↓

6. Formuliše se **zaključak**, odnosno daje odgovor na postavljeni problem.

Dijagram 8.1 Etape u testiranju statističkih hipoteza

U nastavku ćemo najpre detaljno da objasnimo faze testiranja, a zatim ćemo prikazati primenu najpoznatijih parametarskih testova u slučaju jednog i dva uzorka.

8.2.1 Postavljanje nulte i alternativne hipoteze

Nulta i alternativna hipoteza predstavljaju dva precizna, među sobom isključujuća iskaza ili tvrđenja.

Nulta hipoteza

Nulta hipoteza (H_0) je iskaz (tvrđenje) o karakteristici jednog ili više skupova. Kod parametarskih testova koje razmatramo u ovom poglavlju nulta hipoteza se odnosi na vrednost nepoznatog parametra jednog ili dva skupa. Konkretna nulta hipoteza varira od problema do problema, ali se po pravilu postavlja u vidu **status quo**-a, tj. da "nema razlike", "nema uticaja", ili da "nema promene".

Nulta hipoteza može biti prosta i složena. Nulta hipoteza je **prosta** ako se njom tvrdi da je parametar jednak tačno jednoj, unapred poznatoj numeričkoj vrednosti, tzv. **hipotetičnoj vrednosti**. Nultu hipotezu možemo napisati rečima ali je, po pravilu, pišemo simbolima, na primer, H_0: μ=300 € i čitamo: "nulta hipoteza glasi da je aritmetička sredina skupa jednaka 300 evra". Ako nulta hipoteza obuhvata veći broj mogućih vrednosti, na primer, H_0: $\mu \leq 300$ €, ona je **složena**. Bez obzira kako je postavljena, **nulta hipoteza uvek sadrži znak jednakosti**.

Svakoj nultoj hipotezi suprotstavljamo (pridružujemo) **alternativnu hipotezu** i označavamo je sa H_1 ili H_a.

Alternativna hipoteza

Alternativna hipoteza (H_1 ili H_a) je iskaz (tvrđenje) o karakteristici jednog ili više skupova koje je suprotno tvrđenju sadržanom u nultoj hipotezi. Najčešće se postavlja u vidu da "ima razlike ", da "postoji uticaj" ili da je "došlo do promene".

Alternativna hipoteza sadrži sve vrednosti koje parametar može imati, a koje nisu obuhvaćene nultom hipotezom. Zbog toga je data u obliku složene hipoteze. H_1 se u naučnim istraživanjima često naziva i **istraživačkom hipotezom**, jer njom istraživač izražava mišljenje koje postupkom testiranja nastoji da potvrdi. To je i razlog zbog kojeg u praksi često prvo postavljamo alternativnu hipotezu, a tek onda nultu. Napomenimo da nije uvek jednostavno na osnovu istraživačkog pitanja odrediti odgovarajući oblik H_0 i H_1. Pogledajmo nekoliko primera.

PRIMER 8.3: Pretpostavimo da postoji sumnja da prosečna potrošnja benzina automobila Lexus SC430 u gradskoj vožnji (na 100 km) odstupa od deklarisane, koja iznosi 12*l*[2]. Osnovanost ove sumnje proverićemo postavljanjem hipoteza u vidu:

$$H_0: \mu = 12\ l \qquad\qquad H_1: \mu \neq 12\ l$$

Ovde je H_0 prosta, a H_1 je složena hipoteza. Ako H_0 odbacimo, prihvatićemo da je prosečna potrošnja različita od 12 *l*, pri čemu smer odstupanja nije određen.

Alternativnu hipotezu, kojom moguća odstupanja stvarne od hipotetičke vrednosti parametra pratimo **u oba smera** nazivamo **dvosmernom** ili **dvostranom** hipotezom. Test koji se primenjuje u ovakvoj situaciji nazivamo **dvosmernim** ili **dvostranim testom**.

[2] www.greenhouse.gov.au/fuelguide

PRIMER 8.4: Prema podacima Američke trgovinske komore (AMCHAM) proporcija ilegalnog softvera u Srbiji iznosi najmanje 80%, odnosno softverske kompanije su oštećene za približno 10,6 miliona dolara[3]. Ako vladina agencija sumnja u tačnost podatka, tj. smatra da je ovaj procenat preterano visok, ona će najpre postaviti alternativnu hipotezu u vidu $\pi <$ 0,80. Tek nakon toga suprotstaviće joj odgovarajuću nultu hipotezu, $\pi \geq$ 0,80. Dakle:

$$H_0: \pi \geq 0{,}80 \qquad\qquad H_1: \pi < 0{,}80$$

Ne zaboravimo: možemo potvrditi (uz odgovarajući rizik) samo istinitost alternativne hipoteze. Tek ako analiza pokaže da H_0 treba da se odbaci, agencija će zaključiti da je učešće piratskog softvera manje od 80%.

PRIMER 8.5: Prema podacima Gradskog zavoda za zaštitu zdravlja[4] prosečna godišnja vrednost sumpordioksida u Beogradu u 2001. godini iznosila je 12 mg/m^3. Ako smo uvereni da se u međuvremenu situacija pogoršala, odnosno da se prosečna zagađenost u Beogradu povećala, kako da postavimo nultu i alternativnu hipotezu? Postavićemo najpre alternativnu hipotezu H_1: $\mu > 12$ mg/m^3. Sada je lako postaviti nultu hipotezu: H_0: $\mu \leq 12$ mg/m^3. Dakle:

$$H_0: \mu \leq 12 \qquad\qquad H_1: \mu > 12$$

Hipoteze postavljene u poslednja dva primera nazivamo **jednosmernim** ili **jednostranim hipotezama**, a testove kojima ispitujemo prihvatljivost H_0 nazivamo **jednosmernim** ili **jednostranim testovima**. *Primetimo da alternativna hipoteza ukazuje na smer odstupanja koji je kritičan*. U Primeru 8.4 test se zato naziva i *levostranim*, dok se hipoteze u Primeru 8.5 testiraju *desnostranim testom.*

Postavlja se logično pitanje, kada koristiti dvosmerne, a kada jednosmerne testove?

Izbor između jednosmernog i dvosmernog testa

Ako testirana vrednost parametra ne sme odstupati u bilo kom smeru od standarda prihvaćenog u praksi (hipotetičke vrednosti), ili ako unapred ništa ne znamo o potencijalnom odstupanju parametra od njegove hipotetične vrednosti, onda koristimo dvosmerni test.

Napomenimo da neki najpoznatiji statistički softveri (kao, recimo, SPSS) nemaju opcije za primenu jednosmernih testova, kao i da jedan broj autora smatra da treba isključivo koristiti dvosmerne testove. Ipak, gornji primeri pokazuju da u realnim situacijama postoji česta potreba i za jednosmernim

---

[3] http://www.amcham.yu

[4] http://www.zdravlje.org.yu/ekoatlas/08t1.htm

testovima. Pogledajte još jednom Primer 8.2 i postavite nultu i alternativnu hipotezu. Da li ste odabrali jednosmernu alternativnu hipotezu H_1: $\mu < 10$ inča?

Ponekad, za istu hipotetičku vrednost parametra i na osnovu istog uzorka, primenom dvosmernog i jednosmernog testa donosimo različite zaključke; tada u prvom slučaju nultu hipotezu ne možemo da odbacimo, dok je u drugom slučaju odbacujemo i prihvatamo alternativnu hipotezu. Ono što se u praksi nikako ne sme raditi je da primenimo, recimo, dvosmerni test, a da nakon toga, nezadovoljni rezultatom, na istim podacima koristimo jednosmerni test. Veliki broj autora smatra da je ovakav postupak neetički i da vodi falsifikovanju rezultata, i mi se sa njima potpuno slažemo.

Da sumiramo: **mada naše uverenje formulišemo u obliku alternativne hipoteze, u postupku testiranja proveravamo samo nultu hipotezu. Polazimo od pretpostavke da je H_0 istinita i nastojimo da ovu pretpostavku osporimo. Alternativnu hipotezu ne proveravamo, već je automatski prihvatamo kao istinitu ako podaci ubedljivo "svedoče" protiv nulte hipoteze**.

8.2.2 Greške pri testiranju i nivo značajnosti testa

Nulta hipoteza, kao tvrđenje o vrednosti nepoznatog parametra osnovnog skupa, može u stvarnosti biti ili istinita ili neistinita. Sa druge strane, podaci slučajnog uzorka mogu biti ili saglasni sa H_0 ili joj protivrečiti. To znači da kod testiranja, kao i kod ocenjivanja, postoji mogućnost da donesemo pogrešan zaključak. U Tabeli 8.1 prikazali smo sve ishode testiranja u pogledu njihove tačnosti.

Tabela 8.1 Ishodi testiranja hipoteze i njihove verovatnoće

		Stvarna situacija	
		H_0 istinita	H_0 neistinita
Odluka	H_0 ne odbacujemo	Pravilna odluka	**Greška II vrste**
	H_0 odbacujemo	**Greška I vrste**	Pravilna odluka

Na osnovu gornje tabele vidimo da ćemo ispravno postupiti u sledeća dva slučaja:

1. Ako je H_0 istinita, a informacija iz uzorka saglasna sa njom, nultu hipotezu nećemo odbaciti.
2. Ispravno ćemo postupiti i ako odbacimo H_0 koja je neistinita.

Međutim, kako uzorak nikada nije savršeno reprezentativan, moguća su i sledeća dva ishoda: da informacija iz uzorka protivreči istinitoj nultoj hipotezi, ili da se saglasi sa neistinitom nultom hipotezom. Očigledno je da u oba slučaja donosimo pogrešnu odluku. Iz Tabele 8.1 vidimo da realno postoji mogućnost da napravimo dve različite greške u odlučivanju:

1. Prva je da odbacimo istinitu nultu hipotezu. Takva greška se naziva **greškom prve vrste.**
2. Drugu grešku u testiranju činimo ako ne odbacimo netačnu nultu hipotezu; to je **greška druge vrste**.

Jasno je da u zaključku možemo da načinimo samo jednu grešku, a ne istovremeno obe.

Nivo značajnosti testa

Nivo značajnosti testa (naziva se još **rizikom greške prve vrste**) je verovatnoća da ćemo odbaciti istinitu nultu hipotezu. Obeležava se sa $\boldsymbol{\alpha}$.

Verovatnoća da nećemo odbaciti netačnu nultu hipotezu naziva se **rizikom greške II vrste** i obeležava se sa $\boldsymbol{\beta}$.

Vodite računa da *rizik greške* nije isto što i *greška*. Rizik greške je *verovatnoća* da ćemo napraviti grešku. Takođe, u zbiru, rizik greške I vrste i rizik greške II vrste nisu jednaki 1.

Prilikom testiranja hipoteza nastojimo da smanjimo oba rizika greške, što ni malo nije lako. Iako je $\alpha + \beta \neq 1$, ovi rizici su među sobom povezani, tako da smanjenje rizika greške I vrste prati porast rizika greške II vrste, i obrnuto.

Postupak testiranja sprovodimo tako što unapred fiksiramo rizik greške I vrste, tj. nivo značajnosti α. Pri tome biramo relativno mali nivo značajnosti, ali ne toliko mali da on onemogućava odbacivanje svake nulte hipoteze. Koliko mali? Većina istraživača koristi uglavnom samo dva nivoa značajnosti: 0,05 i 0,01. Ako odaberemo najčešće korišćeni nivo značajnosti $\alpha = 0{,}05$[5], onda svesno *unapred* prihvatamo rizik da ćemo u 5% slučajeva odbaciti H_0 iako je ona

[5] Često se studenti i istraživači pitaju zašto se najčešće koristi nivo značajnosti od 5%, a ne neki drugi? Odgovor se nalazi u uticaju Ronalda Fišera koji je, kao vodeći autoritet u statistici, predložio i koristio ovaj nivo značajnosti kao optimalan. Naime, Fišer je smatrao da ako pri testiranju već moramo da se suočimo sa mogućnošću javljanja greške, onda je potrebno da se negde povuče demarkaciona linija, a on je lično preferirao 5%. Drugačije rečeno, po Fišeru, statističari treba da prihvate rizik da u proseku na osnovu jednog od dvadeset uzoraka naprave grešku. Podsetimo se, sličnu logiku smo imali i kod intervala poverenja. I tamo smo tražili optimalni balans između koeficijenta poverenja i širine intervala i najčešće koristili 95% koeficijent poverenja; kod takvog intervala postoji 5% šansi da smo napravili grešku, odnosno, da smo izabrali jedan od 5% nereprezentativnih uzoraka kada formirani interval poverenja ne „pokriva" nepoznati parametar.

istinita. To će se dogoditi ako izaberemo jedan od 5% nereprezentativnih uzoraka (u proseku jedan od dvadeset mogućih uzoraka) koji po svojim karakteristikama značajno odstupaju od karakteristika skupa. Šta bi značilo ako bismo, recimo, sveli rizik greške I vrste na nulu? Tada nikada ne bismo odbacili nultu hipotezu, bila ona tačna ili pogrešna. U čemu bi onda bio smisao testiranja?

Kada se nivo greške I vrste unapred odredi (videli smo da je to najčešće 0,05), ostaje pitanje kako kontrolisati rizik greške II vrste? Nažalost, izračunavanje rizika β u konkretnoj situaciji je skoro nemoguće jer na ovu verovatnoću, pored faktora koje možemo kontrolisati, utiče i razlika između stvarne i hipotetičke vrednosti posmatranog parametra. Kako nam stvarna vrednost testiranog parametra nije poznata, jasno je da ne možemo precizno odrediti ni rizik β.

Na koji način možemo istovremeno smanjiti oba rizika pri testiranju, ili preciznije, kako ćemo za unapred odabrani rizik α*, smanjiti rizik* β? Da bismo našli odgovor, vratimo se Primeru 8.1. Pretpostavimo da ste član porote i da ne biste želeli da osudite nevinu osobu, ali ni da oslobodite krivu. Šta biste uradili da smanjite rizik donošenja pogrešnog suda? Naravno, tražili biste dodatnu istragu, odnosno dodatne informacije. Statistički rečeno, tražili biste da se **poveća uzorak**, što je ujedno i odgovor na postavljeno pitanje.

Umesto da vodi računa o smanjivanju rizika β, istraživač se najčešće rukovodi komplementarnom verovatnoćom, 1 - β. Ova verovatnoća naziva se jačinom testa.

Jačina testa

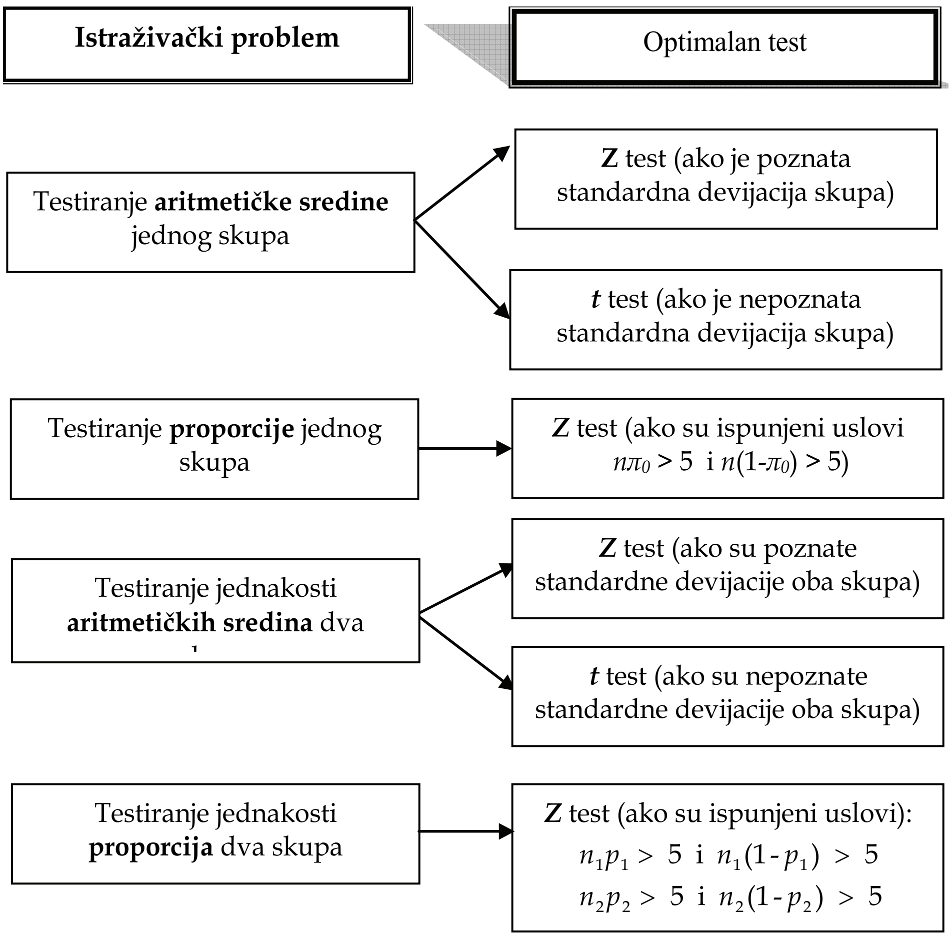
Jačina testa je verovatnoća da se odbaci pogrešna nulta hipoteza. Obeležava se sa π.

8.2.3 Izbor testa

Na osnovu prethodno rečenog zaključujemo da je **optimalan test onaj koji za fiksni nivo značajnosti α ima najveću jačinu**. Međutim, jačina testa zavisi od identičnih faktora kao i rizik greške β pa je jasno da njen nivo ne možemo precizno izračunati za konkretnu hipotezu. Šta nam onda preostaje, koji test da izaberemo?

Ranije smo već rekli da se ista hipoteza može testirati primenom različitih testova (parametarskih i neparametarskih). Pri izboru testa treba da izaberemo onaj koji najviše odgovara empirijskim podacima (tj. test za koji su zadovoljeni polazni uslovi), ali i test koji će za izabrani nivo greške I vrste imati najveću jačinu, i to za unapred izabranu veličinu uzorka, *n*. Rešenje problema izbora nam daje teorijska statistika: ako je raspored skupa normalan, najveću jačinu imaju parametarski testovi.

U ovom poglavlju bavićemo se parametarskim testovima koje smo prikazali u Dijagramu 8.2. Posmatraćemo ukupno četiri istraživačka problema i za njihovo rešavanje koristićemo testove zasnovane na jednom i na dva uzorka. Na osnovu testova zasnovanih na jednom uzorku testiraćemo hipoteze o vrednosti aritmetičke sredine i proporcije osnovnog skupa, dok ćemo primenom testova zasnovanih na dva uzorka testirati hipoteze o jednakosti aritmetičkih sredina i proporcija dva osnovna skupa.



Dijagram 8.2 Optimalni parametarski testovi u zavisnosti od istraživačkog problema

Na osnovu Dijagrama primećujemo da i u okviru grupe parametarskih testova za isti problem imamo više raspoloživih testova. Tako, na primer, hipotezu o aritmetičkoj sredini skupa možemo testirati primenom tzv. Z i t testa.

Zadovoljen uslov normalnosti skupa u konkretnom slučaju nije dovoljan za izbor testa, već moramo uvesti dodatni uslov koji se odnosi na standardnu devijaciju skupa. Ako je standardna devijacija skupa poznata, onda ćemo koristiti Z test, a ako nije (što je najčešći slučaj u praksi), onda ćemo primeniti t test. (Podsetimo sa da smo iste uslove proveravali i kod ocenjivanja nepoznate aritmetičke sredine skupa). Shodno tome, zaključujemo da za svaki od četiri navedena problema, i svaku kombinaciju uslova, postoji jedan optimalan parametarski test.

Navedeni testovi su dobili naziv po odgovarajućem teorijskom rasporedu koji koristimo u procesu testiranja. To znači da ćemo u posmatranim situacijama testiranje sprovoditi primenom standardizovanog normalnog rasporeda ili Studentovog t rasporeda. Ipak, u zavisnosti od hipoteze koju testiramo, promenljive Z i t biće definisane na različite načine, odnosno, biće prikazane različitim formulama. Vrednosti ovih promenljivih, koje ćemo izračunati na osnovu izabranog uzorka, biće osnova za donošenje konačnog zaključka o odbacivanju ili neodbacivanju nulte hipoteze. Drugim rečima, promenljiva koju koristimo u konkretnoj situaciji predstavlja zapravo kriterijum testiranja i nazivamo je statistikom testa.

Statistika testa

Statistika testa je kriterijum na osnovu kojeg vršimo testiranje.

8.2.4 Logika statističkog testiranja hipoteza

Objasnimo logiku kojom se služimo pri testiranju hipoteza, odnosno, kako proveravamo održivost nulte hipoteze da je vrednost parametra skupa jednaka hipotetičkoj vrednosti. Rekli smo da testiranje zasnivamo na kriterijumu – statistici testa, čiju vrednost izračunavamo na osnovu empirijske evidencije iz uzorka. Kako definišemo ovaj kriterijum i od čega zavisi njegova vrednost?

Formulacija kriterijuma zavisi, pre svega, *od predmeta testiranja*, ali i *od ispunjenosti polaznih uslova*. Ako je predmet testiranja vrednost parametra osnovnog skupa, koje informacije iz uzorka ćemo koristiti za izračunavanje vrednosti statistike testa? Na osnovu dosadašnjeg znanja iz oblasti ocenjivanja jasno se nameće sledeći zaključak: Prilikom testiranja hipoteze o aritmetičkoj sredini skupa, μ, empirijsku evidenciju iz uzorka predstavljaće njena ocena, tj. aritmetička sredina uzorka, $\overline{X}$, dok će pri testiranju hipoteze o proporciji skupa, π, to biti proporcija uzorka, Pr.

U postupku testiranja polazimo od pretpostavke da je nulta hipoteza istinita. Drugim rečima, pretpostavljamo da uzorak biramo iz skupa u kome je stvarna vrednost parametra jednaka hipotetičnoj vrednosti. Hipotetičku vrednost aritmetičke sredine skupa obeležavamo sa μ_0, a proporcije skupa sa π_0. Kao što već znamo, ocene $\overline{X}$ i Pr su slučajne promenljive, pa njihove vrednosti u pojedinim uzorcima odstupaju od stvarne vrednosti odgovarajućeg parametra skupa. Samim tim, ne možemo očekivati da će realizovana vrednost ocene u izabranom uzorku biti jednaka hipotetičnoj vrednosti parametra, čak i kada je nulta hipoteza istinita. Zato ćemo meriti razliku između vrednosti ocene i hipotetične vrednosti parametra sadržane u nultoj hipotezi, tj. ($\overline{X} - \mu_0$), odnosno, (Pr -π_0), i na osnovu nje donositi zaključak.

Da bismo eliminisali jedinicu mere u kojoj je izražena vrednost pojave koju posmatramo, razliku između ocene i hipotetičke vrednosti parametra standardizujemo, tj. delimo sa standardnom greškom ocene, i izražavamo je u broju standardnih grešaka. Statistike testa koje ćemo upoznati u ovoj knjizi najčešće imaju upravo takav oblik:

$$\text{Statistika testa} = \frac{\text{Ocena parametra - Hipotetična vrednost parametra}}{\text{Standardna greška ocene}} \qquad \textbf{(8.1)}$$

Tako, na primer, statistika testa za testiranje hipoteze o aritmetičkoj sredini skupa, u zavisnosti od ispunjenosti polaznih uslova, može biti:

$$Z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \quad \text{ili} \quad t = \frac{\overline{X} - \mu_0}{S_{\overline{X}}}.$$

Budući da statistika testa predstavlja funkciju barem jedne slučajne promenljive, jasno je da je i ona slučajna promenljiva. Samim tim, statistika testa će uzimati različite vrednosti od uzorka do uzorka i te vrednosti je nemoguće unapred predvideti. Ako nam je raspored statistike testa poznat, tada možemo izračunavati verovatnoće da će statistika testa biti manja ili veća od neke vrednosti, pa i od one dobijene u konkretnom uzorku.

Logika formule (8.1), kao i samog postupka testiranja, je u sledećem. Ako je nulta hipoteza istinita, tada očekujemo da razlika između ocenjene i hipotetične vrednosti parametra bude mala, pa će i vrednost statistike testa biti mala. U suprotnom, očekujemo relativno veliku (bilo pozitivnu, bilo negativnu) vrednost statistike testa. Vidimo da je značaj statistike testa u tome da ona predstavlja *meru usaglašenosti između podataka iz uzorka i nulte hipoteze*.

Da sumiramo: mala vrednost statistike testa nam ne daje osnove da odbacimo nultu hipotezu. Nasuprot tome, njene velike vrednosti nam sugerišu da nulta hipoteza nije tačna. *Što je veća vrednost statistike testa, to su uverljiviji dokazi da je nulta hipoteza pogrešna*. Ostaje samo da se odredi koje vrednosti statistike testa ćemo smatrati malim a koje velikim, odnosno, kolika treba da bude vrednost statistike testa da bi se nulta hipoteza odbacila.

Postoje **dva načina** da donesemo konačan zaključak o odbacivanju ili neodbacivanju nulte hipoteze. To su **Nimen-Pirsonov** (Neyman-Pearson) pristup, i **Fišerov** pristup, zasnovan na primeni tzv. p-vrednosti. Mi ćemo u ovoj knjizi primeniti kombinovani pristup.

8.2.5 *p*-vrednost

Kao što smo već rekli, testiranje hipoteza se u praksi najčešće sprovodi primenom računara, kada ćemo *p*-vrednost naći u „izlazu“ svakog statističkog softvera. Drugim rečima, nećemo imati potrebe da je sami izračunavamo. Ipak, da bismo razumeli njeno značenje i logiku testiranja, pokazaćemo kako određujemo *p*-vrednost i kako je interpretiramo.

Vratimo se Primeru 8.3, i pretpostavimo da je standardna devijacija skupa poznata i jednaka σ=1*l*, i da biramo uzorak od n=25 elemenata (odnosno, da je $\sigma_{\overline{X}}$ =0,2). Radi jednostavnosti objašnjenja polazimo od jednosmernih hipoteza datih oblikom B:

$$H_0\text{: } \mu \le 12\ l, \quad H_1\text{: } \mu > 12\ l.$$

Pretpostavimo da je aritmetička sredina u izabranom uzorku $\overline{x}$ =12,4*l*, odnosno, da je vrednost statistike testa z=2. Postavimo sledeće pitanje:

Ako je H_0 istinita, koliko iznosi verovatnoća da se javi vrednost statistike testa koja je jednaka upravo realizovanoj, z=2, ili je veća od ove vrednosti (odnosno, **još više protivreči tvrđenju datom u nultoj hipotezi**), tj. kolika je $P(Z \ge 2)$?

Traženu verovatnoću lako izračunavamo na osnovu Tablice 1 i ona je jednaka $P(Z \ge z) = P(Z \ge 2)=1-0{,}9772= 0{,}0228$. Možemo je interpretirati na sledeći način: Ako je H_0 istinita (tj. ako uzorak potiče iz skupa sa prosečnom potrošnjom koja je (najviše) jednaka 12 litara), onda se ovako velike vrednosti statistike testa mogu javiti samo u 2,28% uzoraka. Dobijeni rezultat bismo mogli da tumačimo na dva načina:

a) Nulta hipoteza je tačna, i prisustvujemo veoma retkom događaju, odnosno izvukli smo uzorak koji je ipak moguće izvući, bez obzira što je malo verovatan, ili
b) Nulta hipoteza je pogrešna.

Budući da moramo doneti odluku, mi smo pre skloni da odbacimo nultu hipotezu, nego da poverujemo da prisustvujemo nekom retkom događaju. *Da smo dobili još veću prosečnu vrednost u uzorku*, na primer, $\overline{x}$ =12,6*l* (*tj.* **ekstremniju** *u odnosu na hipotetičku vrednost sadržanu u* H_0), izračunata verovatnoća bi bila još manja (iznosila bi 0,0013), a mi bismo imali još jače argumente protiv nulte hipoteze. Ova verovatnoća je zapravo *p*-vrednost.

p-vrednost

p-vrednost je **verovatnoća** da **statistika testa** uzme vrednost jednaku **ili još ekstremniju** od vrednosti koja se upravo realizovala u uzorku, pod uslovom da je **nulta hipoteza tačna**. **Što je manja *p*-vrednost, jači su dokazi protiv nulte hipoteze**.

Pokažimo sada kako izračunavamo p-vrednost u slučaju da su jednosmerne hipoteze date u obliku C:

$$H_0:\ \mu \geq 12\ l, \quad H_1: \mu < 12\ l.$$

Ako pretpostavimo da je aritmetička sredina u izabranom uzorku $\bar{x}$ =11,48 l, odgovarajuća vrednost statistike testa jednaka je z=-2,6. Koliko sada iznosi p-vrednost? Ako je H_0 istinita, kolika je verovatnoća da se javi vrednost statistike testa koja je jednaka ili manja od -2,6 (tj. ekstremnija u odnosu na hipotetičku vrednost sadržanu u H_0)? Ta verovatnoća je P(Z ≤ z) = P(Z ≤ -2,6)=0,0047, što znači da se ovako male vrednosti statistike testa mogu javiti samo u 0,47% uzoraka, ako oni potiču iz skupa čija je prosečna potrošnja (najmanje) jednaka 12 litara. Zaključujemo da nam dobijena p-vrednost pruža veoma jake argumente protiv nulte hipoteze.

Pokažimo, na kraju, kako se izračunava p-vrednost u slučaju dvosmernog testa. U našem primeru hipoteze su date u obliku (A):

$$H_0:\ \mu=12\ l \quad \text{i} \quad H_1: \mu\neq 12\ l.$$

Ako je aritmetička sredina u izabranom uzorku, na primer, $\bar{x}$ =12,35 l, odgovarajuća vrednost statistike testa jednaka je z=1,75, pa je P(Z ≥ z) = P(Z ≥ 1,75)=1-0,9599= 0,0401. Podsetimo se da dvosmerni testovi prate odstupanja u oba smera. Zato, **ako je nulti raspored simetričan, *p*-vrednost kod dvosmernih testova određujemo tako što dobijenu verovatnoću množimo sa 2**. Nulti raspored u našem primeru je standardizovan normalan raspored, pa budući da je on simetričan, odgovarajuća p-vrednost je jednaka 0,0401×2=0,0802. To je verovatnoća da vrednost statistike testa Z bude veća od 1,75, ili da bude manja od -1,75, tj. P(|Z| ≥ z) = P(|Z| ≥ 1,75) = 0,0802.

Na osnovu gornjih primera zaključujemo sledeće: što je vrednost statistike testa Z više udaljena od 0, to je p-vrednost manja, odnosno, podaci su manje konzistentni sa nultom hipotezom. U tom smislu kažemo da p-vrednost **meri jačinu dokaza protiv nulte hipoteze**. Ako u kompjuterskom „izlazu“ dobijemo da je p-vrednost=0, onda možemo biti skoro sasvim sigurni da je H_0 neistinita i prihvatiti H_1. Ali, šta ćemo zaključiti ako je p-vrednost jednaka, na primer, 0,03, 0,006 ili 0,08? Postavlja se pitanje, koliko mala (bliska nuli) treba da bude p-vrednost da bismo odbacili nultu hipotezu. Budući da nivo značajnosti α predstavlja verovatnoću da ćemo odbaciti istinitu nultu hipotezu, logično se nameće zaključak da odluku možemo doneti upoređivanjem dobijene p-vrednosti sa unapred izabranim rizikom α.

Pravilo odlučivanja na osnovu p-vrednosti glasi:

> **Ako je *p*-vrednost manja od nivoa značajnosti α, nulta hipoteza se odbacuje.** U suprotnom, kazaćemo da nemamo dovoljno argumenata da odbacimo H_0.

Kao što vidimo, u slučajevima kada koristimo statistički softver, odnosno, kada raspolažemo sa *p*-vrednošću, pravilo odlučivanja postaje veoma lako. Ako je *p*-vrednost manja od nivoa značajnosti α, kažemo da je dobijeni rezultat **statistički značajan** ili **signifikantan**, i nultu hipotezu odbacujemo sa rizikom α. U suprotnom, kažemo da rezultat nije statistički značajan.

Statistička značajnost

Ako je *p*-vrednost manja od nivoa značajnosti α, kažemo da je rezultat statistički značajan na nivou α.

Vodimo strogo računa da pojam "statistički značajan" ne mora nužno da znači i "praktično važan". U statistici se ovaj termin koristi samo da ukaže da su dokazi protiv nulte hipoteze dostigli standard označen postavljenim nivoom značajnosti α. **Praktična značajnost**, međutim, **zavisi od veličine razlike** između hipotetične i stvarne vrednosti parametra. Zato se može dogoditi da statistički značajna razlika bude toliko mala da je sa stanovišta posmatranog praktičnog problema smatramo irelevantnom.

Dobijene *p*-vrednosti se često (naročito u stručnim časopisima) tumače kao u sledećoj tabeli:

Tabela 8.2 Interpretacija jačine argumenata protiv nulte hipoteze

p-vrednost $\geq 0{,}05$	Nemamo dovoljno dokaza protiv H_0
p-vrednost $< 0{,}05$	Jaki dokazi da je H_0 pogrešna
p-vrednost $< 0{,}01$	Veoma jaki dokazi da je H_0 pogrešna
p-vrednost $< 0{,}001$	Izuzetno jaki dokazi da je H_0 pogrešna

Na kraju razmatranja *p*-vrednosti ukažimo da u slučaju ručne obrade podataka ona ima jedan krupan nedostatak. Naime, bez pomoći kompjutera njenu **tačnu vrednost** možemo izračunati samo kod *Z* testa, na osnovu Tablice 1. Kod svih ostalih testova to nije moguće, jer se oni zasnivaju na teorijskim rasporedima za koje ne raspolažemo kompletnim tablicama. U takvim situacijama možemo samo izračunati **interval** u kome se nalazi *p*-vrednost.

8.3 TESTIRANJE HIPOTEZE ZASNOVANO NA JEDNOM UZORKU

Sve statističke testove u ovom poglavlju, radi preglednosti i lakšeg razumevanja, tretiraćemo na identičan način. Najpre ćemo tabelarno prikazati uslove koji moraju biti ispunjeni za njihovu primenu, oblike nulte i alternativne hipoteze, i statistiku testa koja se koristi. Zatim ćemo ukratko objasniti logiku testiranja i na primeru pokazati kako se ono sprovodi. Konačno, kod svih analiziranih testova prikazaćemo i izlaz iz nekog statističkog softvera.

8.3.1 Testiranje hipoteze o aritmetičkoj sredini osnovnog skupa

U Dijagramu 8.2, a i tokom prethodnog izlaganja, videli smo da se za testiranje hipoteze o aritmetičkoj sredini osnovnog skupa mogu koristiti dva parametarska testa, Z i t.

Z test

Osnovne informacije vezane za primenu Z testa u slučaju testiranja aritmetičke sredine jednog skupa prikazane su u Tabeli 8.3.

Tabela 8.3 Testiranje hipoteze o aritmetičkoj sredini skupa primenom Z testa

Uslovi za primenu Z testa:

1) Osnovni skup ima normalan raspored ili može se primeniti Centralna granična teorema ($n \geq 30$).
2) Standardna devijacija skupa, σ, je poznata.

Izbor oblika nulte i alternativne hipoteze:

A) $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$

B) $H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$

C) $H_0 : \mu \geq \mu_0 \quad H_1 : \mu < \mu_0$

gde je μ_0 hipotetična vrednost aritmetičke sredine skupa.

Statistika testa: $$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \qquad \textbf{(8.2)}$$

ako je H_0 istinita ima standardizovan normalan raspored, $Z : N(0, 1)$.

Budući da smo već pokazali kako se kod Z testa izračunava p-vrednost, ovde ćemo posebnu pažnju posvetiti etapama testiranja, koje smo prikazali u Dijagramu 8.1.

PROBLEM 8.1: U deklaraciji uz pošiljku od 5000 litijumskih Energizer L92 baterija navedeno je da neprekidna dužina njihovog trajanja (kada se koriste na digitalnim audio aparatima) iznosi 13 časova,[6] sa dozvoljenim prosečnom odstupanjem od 2 časa. Prilikom preuzimanja pošiljke, slučajnim putem je izabrano 50 baterija i ustanovljeno da njihov prosečan vek trajanja iznosi 12,6 časova. Da li pošiljka zadovoljava navedeni standard po pitanju trajanja baterija?

REŠENJE

Rešavanje problema zahteva, pre svega, da pažljivom analizom precizno

[6] http://data.energizer.com/PDFs/l92.pdf

definišemo pitanje, a zatim da izaberemo odgovarajući statistički metod kojim ćemo doći do traženog odgovora. Ova faza pokazuje da je statistički način istraživanja i kvalitativan, a ne samo kvantitativan, tj. da se ne svodi na puko primenjivanje formula i izračunavanje.

Prvo što treba utvrditi je koji se statistički parametar ispituje. Pažljivim čitanjem (analizom Problema 8.1) uočavamo da je predmet analize **aritmetička sredina** skupa.

Da bismo utvrdili da li kvalitet baterija odgovara navodima u deklaraciji moramo odabrati odgovarajući statistički metod. Vidimo da raspolažemo uzorkom od 50 baterija i da na postavljeno pitanje "Da li pošiljka zadovoljava navedeni standard" postoje samo dva moguća odgovora, da ili ne. Primenjujući preporuku datu na početku ovog poglavlja, zaključujemo da treba izabrati **testiranje hipoteza**.

Tek sada možemo da pokrenemo proceduru testiranja po etapama sugerisanim u Dijagramu 8.1.

Etapa 1. *Postavljanje nulte i alternativne hipoteze, i određivanje nivoa značajnosti.*

Odredimo najpre da li postavljeni problem treba prikazati u obliku dvosmernog ili jednosmernog testa. Vidimo da su dozvoljena odstupanja *u oba smera*, zbog čega se opredeljujemo za *dvosmerni* test. Nivo značajnosti biramo sami, pa ćemo po Fišerovoj preporuci izabrati $\alpha = 0{,}05$.

$H_0 : \mu = 13^h$ $\qquad$ $H_1 : \mu \neq 13^h$ $\qquad$ $\alpha = 0{,}05$

Etapa 2. *Uzima se slučajan uzorak, bira odgovarajući test i proveravaju uslovi za primenu izabranog testa.*

Budući da već raspolažemo uzorkom, ostaje da odredimo optimalan test, rukovodeći se Dijagramom 8.2. Izbor između Z i t testa zavisi od toga da li je poznata standardna devijacija skupa. U deklaraciji pošiljke je navedeno da dozvoljeno "prosečno odstupanje" iznosi 2 časa, iz čega zaključujemo da je $\sigma = 2^h$. Samim tim, biramo Z test, sa statistikom testa datom izrazom 8.2.

Etapa 3. *Proveravanje ispunjenosti uslova izabranog testa.*

Z test, kao što vidimo u Tabeli 8.3, zasniva se na dva uslova. Proverimo da li su oni ispunjeni.

1) *Osnovni skup ima normalan raspored, ili može se primeniti CGT (n > 30).* $\Longrightarrow$ Ispunjeno, jer je $n = 50$

2) *Standardna devijacija skupa, σ, je poznata.* $\Longrightarrow$ Ispunjeno, jer je $\sigma = 2^h$

Zaključujemo da su oba uslova ispunjena i nastavljamo sa testiranjem.

Etapa 4. *Izračunavanje statistike testa i p-vrednosti.*

Realizovana vrednost statistike testa je:

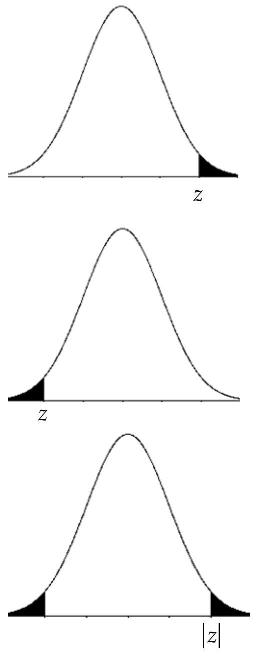
$$z = \frac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} = \frac{12{,}6 - 13}{2/\sqrt{50}} = -1{,}414$$

Vidimo da je aritmetička sredina izabranog uzorka, $\overline{x}$ =12,6, manja od hipotetičke aritmetičke sredine skupa od 13 časova za 1,414 standardne greške. Da li je ova razlika dovoljno velika da sugeriše odbacivanje H_0? Odgovor na ovo pitanje daće nam p-vrednost, koju za dvosmerne testove izračunavamo na sledeći način:

$$2P(Z \geq |z|) = 2P(Z \geq |-1,414|) = 2P(Z \geq 1,414) = 2\left[1 - P(Z < 1,414)\right]$$
$$= 2\left[1 - 0,9207\right] = 0,1586$$

Generalno, p-vrednost kod svakog Z testa izračunavamo na osnovu objašnjenja u odeljku 8.2.6 koji je grafički ilustrovan na Slici 8.2.

Alternativna hipoteza	**p-vrednost**
$H_1 : \mu > \mu_0$	$P(Z \geq z)$
$H_1 : \mu < \mu_0$	$P(Z \leq z)$
$H_1 : \mu \neq \mu_0$	$2P(Z \geq \|z\|)$



Slika 8.2 Izračunavanje p-vrednosti kod Z testa

Etapa 5. *Postavljanje pravila odlučivanja i donošenje odluke o nultoj hipotezi.*

Kao što smo videli pravilo odlučivanja pri korišćenju p-vrednosti glasi:

Ako je p-vrednost manja od nivoa značajnosti, H_0 se odbacuje; u suprotnom se ne odbacuje.

Naša p-vrednost je jednaka 0,1586 i kao takva veća od odabranog nivoa značajnosti α = 0,05. Usled toga nemamo dovoljno razloga da odbacimo nultu hipotezu.

Za dobijeni rezultat kažemo da *nije statistički značajan*, odnosno da je *nesignifikantan*.

Etapa 6. *Donošenje zaključka.*

Nemamo dovoljno dokaza da tvrdimo da kvalitet pošiljke (meren prosečnim trajanjem baterija) odstupa od vrednosti u deklaraciji. Konsekventno, prihvatićemo pošiljku od 5000 baterija.

Na kraju razmatranja Z testa prikažimo rezultate analize primenom *EduStata*. Primećujemo da postoje neznatne razlike u rezultatima, koje su posledica zaokrugljivanja.

Z-test za aritmetičku sredinu jednog skupa			
H_0: μ = 13			
H_1: μ <> 13			
Varijabla	Veličina uzorka (n)	Aritmetička sredina uzorka	Stand. devijacija skupa
	50	12,6	2
	Standardna greška	Z statistika	***P***
	0,283	-1,4142	0,15735
Zaključak			
Nemamo dovoljno razloga da tvrdimo da se prosečna vrednost u populaciji razlikuje od 13			

Studentov *t* test

Osnovne informacije vezane za primenu verovatno **najpoznatijeg** statističkog testa, *t* testa, u slučaju testiranja aritmetičke sredine jednog skupa prikazane su u Tabeli 8.4.

Tabela 8.4 Testiranje hipoteze o aritmetičkoj sredini skupa primenom *t* testa

Uslovi za primenu *t* testa:

1) Osnovni skup je normalno raspoređen, ili
2) Važi Centralna granična teorema (uzorak ima najmanje 30 elemenata)

Izbor oblika nulte i alternativne hipoteze:

A) $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ B) $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$ C) $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$

gde je μ_0 hipotetična vrednost aritmetičke sredine skupa

Statistika testa: $$t = \frac{\overline{X} - \mu_0}{S_{\overline{X}}} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \qquad \textbf{(8.3)}$$

ako je H_0 istinita ima Studentov raspored sa n-1 stepeni slobode.

U praksi nam standardna devijacija populacije skoro nikada nije poznata. Zato, kao i kod ocenjivanja, σ ocenjujemo na osnovu standardne devijacije uzorka S, i na osnovu nje dobijamo ocenu standardne greške, $S_{\bar{X}}$. Tako dobijamo novu statistiku testa, tzv. **statistiku *t* testa** datu izrazom 8.3.

PROBLEM 8.2: Prema jednom istraživanju[7] samo 11% studenata u Srbiji završi studije u predviđenom roku, a prosečna dužina studiranja je 8 godina. Mi, međutim, verujemo da studenti smera Bankarstvo na Ekonomskom fakultetu u Beogradu po ovom pitanju ostvaruju bolje rezultate. Da bismo to proverili uzeli smo od rukovodioca studentske službe slučajan uzorak od 16 studenata koji su na tom smeru diplomirali školske 2005/2006. godine. Dobili smo sledeće podatke (dužina studiranja u godinama):

8,5	6,5	4,5	9,5	11,5	4,5	5,5	4,5
7,5	7,5	6,5	5,5	7,5	4,5	6,5	5,5

Da li je prosečna dužina studiranja na smeru Bankarstvo kraća od proseka studiranja u Srbiji? Iz iskustva znamo da je dužina studiranja normalno raspoređena.

REŠENJE

Budući da istraživačko pitanje glasi "da li je prosečna dužina studiranja *kraća*" testiranje ćemo sprovesti primenom jednosmernog (levostranog) testa:

$$H_0\colon \mu \geq 8 \qquad H_1 : \mu < 8,$$

i izabraćemo nivo značajnosti $\alpha = 0{,}05$.

Pri izboru testa, uzimamo u obzir da je osnovni skup normalno raspoređen i da nepoznatu σ moramo da ocenimo na osnovu S. Zato koristimo statistiku ***t*** (8.3) i Studentov raspored sa n - 1 = 16 - 1 = 15 stepeni slobode.

Aritmetička sredina uzorka jednaka je

$$\bar{x} = \frac{\sum x_i}{n} = \frac{106}{16} = 6{,}625$$

a ocenjena vrednost standardne greške iznosi:

$$s_{\bar{X}} = \sqrt{\frac{\sum x_i^2 - n \cdot \bar{x}^2}{n\,(n-1)}} = \sqrt{\frac{11236 - 16 \cdot 6{,}625^2}{16 \cdot 15}} = 0{,}499$$

pa zamenom u (8.3) izračunavamo vrednost statistike ***t***:

$$t = \frac{6{,}625 - 8}{0{,}499} = -2{,}76 .$$

Primetimo da Tablica 2 sadrži kritične vrednosti Studentovog

[7] Slobodanka Turajlić, "Higher education in Serbia – Reform strategy", dostupno na sajtu www.see-educoop.net/education_in/pdf/heser_reform-ser_enl_t02.pdf

rasporeda za nekoliko izabranih verovatnoća. Ovde smo prikazali samo red koji se odnosi na 15 stepeni slobode, zajedno sa zaglavljem:

Verovatnoća	0,10	0,05	0,025	0,01	0,005
Kritična t rednost	1,341	1,7531	2,1315	2,602	2,9467

Budući da ne raspolažemo svim verovatnoćama Studentovog rasporeda za 15 stepeni slobode, bez primene nekog statističkog softvera ne možemo odrediti tačnu p-vrednost. Moguće je odrediti samo interval u kojem se nalazi p-vrednost. Bez obzira što u praksi nikada nećemo koristiti tablice i formirati interval za p-vrednost da bismo do kraja shvatili koncept testiranja pokazaćemo kako se taj interval određuje.

Po definiciji p-vrednost je verovatnoća da statistika testa bude jednaka ili ekstremnija (u ovom slučaju manja) od realizovane vrednosti, -2,76. Budući da zbog simetričnosti t rasporeda važi relacija:

$$P\,(\,\boldsymbol{t}_{15} < -2{,}76) = P\,(\boldsymbol{t}_{15} > 2{,}76),$$

p-vrednost možemo odrediti **zanemarujući predznak** vrednosti statistike testa. Dakle naša statistika testa iznosi 2,76, i nalazi se između 2,602 i 2,9467. Ovim tabličnim vrednostima odgovaraju verovatnoće u zaglavlju od 0,01 i 0,005. Samim tim zaključujemo da je 0,005 < p-vrednost < 0,01. Ovaj rezultat potvrđuju i rezultati testiranja u statističkom paketu *Minitab*, verzija 14, koji sadrže tačnu p-vrednost= 0,007.

One-Sample T: Dužina studiranja

Test of mu = 8 vs < 8

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
Dužina	16	6,62500	1,99583	0,49896	7,49970	-2,76	0,007

To znači da bismo postavljenu nultu hipotezu mogli da odbacimo uz rizik greške I vrste od samo 0,007. Ipak, ako zaključak donosimo poređenjem p-vrednosti sa rizikom α, na osnovu odnosa $p<\alpha$, nultu hipotezu ćemo odbaciti uz rizik greške od 0,05.

Važna napomena: Prilikom izračunavanja p-vrednosti, u slučaju kada se koristi dvosmerni test, dobijene vrednosti verovatnoća u zaglavlju tablice potrebno je pomnožiti sa 2, kao kod Z testa.

8.3.2 Testiranje hipoteze o proporciji osnovnog skupa

Osnovne informacije vezane za primenu Z testa u slučaju testiranja proporcije jednog skupa prikazane su u Tabeli 8.5.

Tabela 8.5 Testiranje hipoteze o proporciji skupa primenom *Z* testa

Uslovi za primenu *Z* testa:
$n \cdot \pi_0 \geq 5$, $n \cdot (1 - \pi_0) \geq 5$ i $n \geq 30$
Izbor oblika nulte i alternativne hipoteze:
A) $H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$ B) $H_0 : \pi \leq \pi_0$ $H_0 : \pi > \pi_0$ C) $H_0 : \pi \geq \pi_0$ $H_1 : \pi < \pi_0$
gde je π_0 hipotetična proporcija skupa.
Statistika testa: $Z = \dfrac{p - \pi_0}{S_p}$ $S_p = \sqrt{\dfrac{\pi_0 (1 - \pi_0)}{n}}$
pod uslovom da je H_0 istinita ima standardizovan normalan raspored.

Logika statistike testa o proporciji skupa je analogna onoj kod aritmetičke sredine. Razlika je u tome što proporcije malih uzoraka nemaju normalan raspored. Međutim, u ovoj knjizi mi ćemo testiranje hipoteze o proporciji skupa sprovoditi samo primenom velikih uzoraka. Tada se raspored verovatnoće proporcije *Pr*, ako su ispunjeni uslovi dati u Tabeli 8.5, može aproksimirati normalnim rasporedom, *Pr*: *približno* $\mathrm{N}(\pi_0, \sigma_p)$. Generalna formula za statistiku testa (8.1) sugeriše i oblik koji će statistika testa imati:

$$Z = \frac{\text{Proporcija uzorka - Hipotetična vrednost proporcije}}{\text{Standardna greška proporcije}} = \frac{p - \pi_0}{\sigma_p} \qquad \textbf{(8.4)}$$

Statistika 8.4, pod uslovom da je H_0 istinita, ima standardizovan normalan raspored, *Z*: *N*(0, 1).

Primetimo da u praksi u formuli (8.4) umesto nepoznate standardne greške proporcije (σ_p) koristimo njenu ocenu (S_p). Važno je da napomenemo da ocenu standardne greške proporcije kod testiranja izračunavamo primenom hipotetične vrednosti parametra (π_0), a ne kao kod ocenjivanja, na osnovu realizovane vrednosti proporcije (*p*) u uzorku:

$$\text{Standardna greška proporcije} \qquad S_p = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n}} \qquad \textbf{(8.5)}$$

PROBLEM 8.3: Na osnovu istraživanja sprovedenih 2003. godine poznato je da je procenat stanovnika koji podržavaju ulazak Srbije u EU iznosio 76%. U decembru 2006. godine sprovedena je nova anketa[8] i u slučajnom uzorku od 1700 ispitanika njih 1139 je izjavilo da su za pridruživanje Srbije EU. Da li je došlo do promene stavova stanovništva Srbije o ulasku u EU?

REŠENJE

Pročitajte pažljivo problem i pronađite koji je parametar u pitanju i koliko se uzoraka koristi. Da li ste utvrdili da je parametar proporcija, i da se koristi jedan uzorak? Uočimo da postavka problema zahteva primenu dvosmernog testa. U prvoj etapi postavimo nultu i alternativnu hipotezu:

$$H_0 : \pi = 0{,}76 \qquad\qquad H_1 : \pi \neq 0{,}76$$

i nivo značajnosti specifikujmo na nivou od 0,05.

Ispitajmo sada da li su ispunjeni uslovi za njegovu validnu primenu; drugim rečima da li se može primeniti normalna aproksimacija.

$$n \cdot \pi_0 = 1700 \cdot 0{,}76 = 1292 > 5 \quad \text{i} \quad n \cdot (1 - \pi_0) = 1700 \cdot 0{,}24 = 408 > 5$$

Oba uslova su ispunjena, a kako je uzorak veliki, prelazimo na izračunavanje statistike testa i p-vrednosti.

Proporcija ispitanika koji podržavaju pridruživanje je:

$$p = \frac{1139}{1700} = 0{,}67\ ,$$

a ocenjena vrednost standardne greške, na osnovu (8.5) iznosi:

$$S_p = \sqrt{\frac{0{,}76(1 - 0{,}76)}{1700}} = 0{,}01\,.$$

Realizovana vrednost statistike Z testa, na osnovu formule (8.4), jednaka je:

$$z = \frac{p - \pi_0}{s_p} = \frac{0{,}67 - 0{,}76}{0{,}01} = -9\,.$$

Budući da primenjujemo Z test, p-vrednost se identično računa kao i prilikom testiranja aritmetičke sredine skupa (Slika 8.2). Budući da je test dvosmeran

$$\begin{aligned} 2P(Z \geq |z|) &= 2P(Z \geq |-9|) = 2P(Z \geq 9) = 2\left[1 - P(Z < 9)\right] \\ &= 2\left[1 - 1\right] = 0 \end{aligned}$$

Kako je p-vrednost manja od nivoa značajnosti rezultat je signifikantan, tj. statistički značajan. To znači da nultu hipotezu odbacujemo. Uz rizik od 0,05 tvrdimo da se proporcija stanovništva koje podržava ulazak Srbije u EU razlikuje od 76%.

Na kraju prikažimo izlaz iz statističkog softvera *Statgraphics Plus*.

---

[8] Anketu je sprovela novosadska agencija "SCAN" a rezultati su objavljeni u Politici 7 decembra 2006.

Hypothesis Tests
Sample proportion = 0,67
Sample size = 1700

Approximate 95,0% confidence interval for p: [0,646001;0,691529]

Null Hypothesis: proportion = 0,76
Alternative: not equal
P-Value = 0,0
Reject the null hypothesis for alpha = 0,05.

8.4 TESTIRANJE HIPOTEZE ZASNOVANO NA DVA UZORKA

Na kraju ovog poglavlja objasnićemo parametarske testove kojima poredimo parametre dva skupa. Pri tome zadržaćemo se samo na poređenju aritmetičkih sredina dva skupa. Neki autori smatraju da ovi metodi spadaju među najkorisnije statističke metode u eksperimentalnoj praksi. Na primer, prilikom ispitivanja efikasnosti nekog novog leka uobičajeno je da se oforme dve grupe. Jedna (**eksperimentalna**), prima novi lek, a druga (**kontrolna**) prima standardni lek ili pilulu bez ikakvog aktivnog sastojka (tzv. placebo). Pri tome je od posebne važnosti da se a) svaki subjekt slučajnim putem dodeli jednoj od dve grupe, i b) da niti učesnici u eksperimentu (pacijenti) ni medicinski personal koji im daje lekove ne znaju ko prima lek a ko placebo. Nakon određenog vremena upoređuje se zdravstveno stanje pacijenata u obe grupe i testira da li je dobijena razlika statistički značajna. Usled ovako postavljenog eksperimentalnog plana, ako eksperimentalna grupa u proseku pokaže bolje rezultate od kontrolne, može se zaključiti da je novi lek efikasniji od standardnog, odnosno da ima pozitivan efekat.

Statističke testove u slučaju dva uzorka možemo koristiti i u drugim situacijama. Na primer, radi provere pretpostavke da su LCD televizori dva proizvođača istog kvaliteta u pogledu prosečnog trajanja, da su prosečna primanja zaposlenih u jednom gradu veća za 100 evra u odnosu na primanja u drugom, da nema razlike u prosečnoj dužini studiranja na državnim i privatnim fakultetima, da je u proizvodnji MP3 plejera učešće neispravnih kod jednog proizvođača manje nego kod drugog i slično. Vidimo da predmet testiranja ne mora biti sama vrednost jednog parametra u dva skupa, već **razlika** između ovih vrednosti. Ipak zadržaćemo se samo na hipotezama o jednakosti parametara dva skupa (preciznije njihovih aritmetičkih sredina). Ako dva skupa obeležimo sa 1 i 2, a vrednosti aritmetičkih sredina u njima sa μ_1 i μ_2, nulte a alternativne hipoteze možemo postaviti na sledeće načine::

A. $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$ dvosmerna alternativna hipoteza

B. $H_0 : \mu_1 \leq \mu_2 \quad H_1 : \mu_1 > \mu_2$ jednosmerna alternativna hipoteza

C. $H_0 : \mu_1 \geq \mu_2 \quad H_1 : \mu_1 < \mu_2$ jednosmerna alternativna hipoteza

Nultu hipotezu testiramo po identičnom postupku kao do sada. Statistika testa i ovde u suštini ima generalni oblik (8.1), s tim što se umesto vrednosti jednog parametra ovde testira razlika između parametara dva skupa.

$$\text{Statistika testa} = \frac{\text{Ocena razlike - Hipotetična vrednost razlike}}{\text{Standardna greška ocene razlike}}$$

Ograničićemo se samo na slučaj kada su oba slučajna uzorka međusobno nezavisna, odnosno kada izbor bilo kog elementa u jedan uzorak ne zavisi od izbora elemenata u drugi uzorak.

8.4.1 Testiranje hipoteze o jednakosti aritmetički sredina dva skupa

Z test

Osnovne informacije vezane za primenu Z testa u slučaju testiranja jednakosti aritmetičkih sredina dva skupa prikazane su u Tabeli 8.6.

Tabela 8.6 Testiranje hipoteze o razlici aritmetičkih sredina dva skupa primenom Z testa

Uslovi za primenu Z testa:

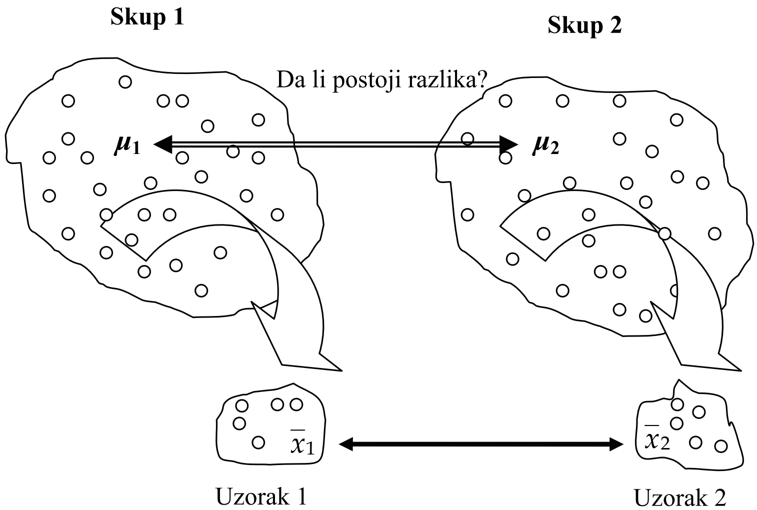
1) Osnovni skupovi imaju normalan raspored ili može da se primeni Centralna granična teorema, odnosno oba uzorka imaju najmanje 30 elemenata.
2) Standardne devijacije (ili varijanse) oba osnovna skupa su poznate.

Statistika testa: $$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \qquad \textbf{(8.7)}$$

pod uslovom da je H_0 istinita ima standardizovan normalan raspored .

Identična ograničenja vezana za primenu Z i t-testa pri testiranju aritmetičke sredine jednog skupa važe i ovde. Dakle, Z test se zasniva na nerealnoj pretpostavci da su standardne devijacije oba skupa poznate, pa u praksi preporučujemo primenu t testa.

Testiranje hipoteze o jednakosti aritmetičkih sredina dva skupa sprovodimo analogno testiranju u slučaju jednog uzorka. Iz oba skupa uzećemo po jedan prost slučajan uzorak, kao na Slici 8.3.



Slika 8.3 Grafički prikaz testiranja jednakosti aritmetičke sredine dva skupa

Ideja testiranja je u sledećem: ako je nulta hipoteza istinita, tj. ako su aritmetičke sredine dva skupa među sobom jednake, onda očekujemo da razlike između aritmetičkih sredina uzoraka, izabranih iz dva skupa, bude relativno male. Takve razlike ćemo pripisati slučajnom kolebanju uzoraka. U suprotnom, ako su razlike između aritmetičkih sredina uzoraka relativno velike, onda ćemo ih tretirati kao argumente protiv nulte hipoteze, koju ćemo odbaciti uz izabrani nivo značajnosti.

Kao što znamo, ako osnovni skup ima normalan raspored, onda i aritmetičke sredine uzoraka iz tog skupa slede normalan raspored. U statistici je pokazano da **razlika aritmetičkih sredina uzoraka izabranih iz dva normalno raspoređena osnovna skupa ima normalan raspored**.

$$\overline{X}_1 - \overline{X}_2 : N(\mu_1 - \mu_2, \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}) .$$

Pošto razlika aritmetičkih sredina uzoraka sledi normalan raspored, testiranje hipoteze o jednakosti aritmetičkih sredina dva skupa ćemo sprovesti primenom Z testa. Statistika Z testa glasi:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}},$$

Na osnovu statistike testa vidimo da hipotezu o jednakosti aritmetičkih sredina dva skupa testiramo na osnovu razlike između aritmetičkih sredina izabranih uzoraka.

$\sigma_{\overline{X}_1 - \overline{X}_2}$ je standardna greška razlike aritmetičkih sredina dva uzorka i izračunavamo je po formuli:

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{8.8}$$

Dakle, statistiku Z možemo napisati u sledećem obliku:

Statistika Z testa $$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{8.9}$$

Pod uslovom da je H_0 istinita, statistika Z ima standardizovan normalan raspored.

PROBLEM 8.4: Radi ispitivanja da li novi lek, Topiramejt, ima efekta pri lečenju migrene, formirane su dve grupe (eksperimentalna i kontrolna) od po osam pacijenata. Rezultati o mesečnom broju migrena su objavljeni u februaru 2004. godine u časopisu Science News[9]:

Topiramejt	3	3	2	4	3	4	3	5
Placebo	3	3	5	5	7	5	5	4

Da li Topiramejt uzrokuje smanjenje prosečnog broja javljanja migrena? (pretpostavimo da oba skupa imaju normalan raspored i da su varijanse skupova poznate i da iznose u eksperimentalnoj grupi 1 a u drugoj 1,7).

REŠENJE

Postavimo najpre nultu i alternativnu hipotezu. Budući da istraživačko pitanje glasi da li lek uzrokuje smanjenje, jasno je da se zahteva primena jednosmernog testa. Označavajući Topiramejt grupu sa 1, a kontrolnu (placebo) sa 2, nulta i alternativna hipoteza glase :

$$H_0 : \mu_1 \geq \mu_2 \quad H_1 : \mu_1 < \mu_2$$

Budući da nivo značajnosti nije dat uzećemo $\alpha = 0,05$.

Kako su varijanse poznate, a traži se poređenje aritmetičkih sredina, odabraćemo Z test.

Proverimo da li su ispunjeni preduslovi:

1) Osnovni skupovi su normalno raspoređeni $\Rightarrow$ ispunjeno

2) Varijanse oba skupa su poznate $\Rightarrow$ ispunjeno

Da bismo izračunali statistiku testa moramo najpre odrediti aritmetičke

[9] The Science News, 165, No. 9, 28 februar 2004.

sredine uzoraka:

$$\bar{x}_1 = 27/8 = 3{,}375 \qquad \bar{x}_2 = 37/8 = 4{,}625$$

Na osnovu formule (8.9) realizovana vrednost statistike Z testa jednaka je:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{3{,}375 - 4{,}625}{\sqrt{1/8 + 1{,}7/8}} = -2{,}15$$

p-vrednost izračunavamo kao kod svakog Z testa (Slika 8.2):

$$P(Z \le z) = P(Z \le -2{,}15) = 0{,}0158\,.$$

Pošto je p-vrednost manja od 0,05 nultu hipotezu odbacujemo. Zaključujemo, uz rizik od 0,05, da Topiramejt statistički značajno smanjuje prosečan broj dana u mesecu u kojima se javlja migrena.

Na kraju prikažimo rezultate navedenog problema pomoću kompjuterskog izlaza iz Excela.

z-Test: Two Sample for Means

	Topiramejt	*Placebo*
Mean	3,375	4,625
Known Variance	1	1,7
Observations	8	8
Hypothesized Mean Difference	0	
z	-2,151657415	
P(Z<=z) one-tail	0,015712173	
z Critical one-tail	1,644853627	
P(Z<=z) two-tail	0,031424347	
z Critical two-tail	1,959963985	

Primetimo da izborom nivoa značajnosti od 0,01 ne bismo mogli da odbacimo nultu hipotezu, pa bismo zaključili da novi lek nije efikasan! To je još jedan primer koji govori u prilog mišljenju velikog broja autora da p-vrednost ne treba porediti sa rizikom α, već da treba prepustiti donosiocu odluke da na osnovu realizovanog nivoa značajnosti odredi smer svoje akcije.

Studentov *t* test

Studentov test kojim se porede aritmetičke sredine dva skupa je jedan od **najvažnijih i u praksi najčešće korišćenih testova**. Kao i kod Z testa zadržaćemo se samo na slučaju kada je hipotetična vrednost razlike aritmetičkih sredina dva skupa jednaka nuli

Osnovne informacije vezane za primenu t testa u slučaju testiranja jednakosti aritmetičkih sredina dva skupa prikazane su u Tabeli 8.7.

Tabela 8.7 Testiranje hipoteze o jednakosti aritmetičkih sredina dva skupa primenom *t* testa (nepoznate varijanse skupova su među sobom jednake)

Uslovi za primenu t testa:

1) Oba osnovna skupa imaju normalan raspored ili može da se primeni Centralna granična teorema (oba uzorka imaju najmanje 30 elemenata).
2) Nepoznate varijanse dva skupa su među sobom jednake ($\sigma_1^2 = \sigma_2^2$)

Statistika testa: $$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{S_{\overline{X}_1 - \overline{X}_2}} \qquad \textbf{(8.10)}$$

$$S_{\overline{X}_1 - \overline{X}_2} = S_P \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad S_P = \sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}} \qquad \textbf{(8.11)}$$

pod uslovom da je H_0 istinita ima Studentov raspored sa $n_1 + n_2 - 2$ stepeni slobode.

. Primetimo, takođe, da je pored uslova normalnosti uveden i dopunski uslov da su varijanse dva skupa među sobom jednake. Ovaj uslov se u statistici naziva pretpostavkom **homogenosti varijansi**, a *t* test koji se na njemu zasniva naziva se **ponderisani *t* test** (eng. pooled *t* test**)**. Problem sa praktičnom primenom ovog testa je što ne postoji pouzdan način provere ispunjenosti pretpostavke homogenosti varijansi[10]. Ipak, radovi objavljeni u drugoj polovini XX-og veka pokazuju da se ponderisani *t* test može koristiti i kada je pretpostavka homogenosti varijansi narušena, ali pod uslovom da **oba uzorka imaju jednak broj elemenata**. Zbog toga postoji **preporuka istraživačima da istraživanje sprovode na uzorcima jednake veličine**.

Razlog za primenu ponderisanog *t* testa se nalazi u problemu[11] koji sobom nosi primena *t* statistike, koju dobijamo kada u formuli statistike Z testa (8.10), nepoznate varijanse skupova, σ_1^2 i σ_2^2, samo zamenimo njihovim ocenama, tj. sa S_1^2 i S_2^2. Pored toga što ova statistika ima samo približno *t* raspored, broj stepeni slobode se u ovom slučaju izračunava primenom veoma složenog obrasca, i često se dobija decimalan broj[12]. Upravo iz tog razloga, ovde smo se opredelili za ponderisani *t* test, mada se on poslednjih godina, sa sve

---

[10] Testovi koji ispituju jednakost varijansi dva skupa (recimo *F* test) su neprecizni, jer strogo zahtevaju da oba skupa imaju normalan raspored.

[11] Ovaj problem poznat je u literaturi pod nazivom Fišer-Berensov problem.

[12] U literaturi je predloženo više predloga kako da se odredi broj stepeni slobode. Najpoznatiji predlog je tzv. Velčova (Bernard Lewis Welch 1911.-1989.) korekcija broja stepeni slobode koja se koristi, između ostaloga u Excelu, Minitabu i EduStatu.

širom upotrebom statističkih softvera polako napušta.

Ponderisani t test je dobio naziv po tome što se standardna greška računa na osnovu ponderisane aritmetičke sredine varijansi dva uzorka, koja predstavlja ponderisanu ocenu jednakih varijansi dva skupa. Takva, ponderisana, ocena varijanse S_P^2 izračunava se po formuli:

$$S_P^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$$

(indeks p ukazuje da je reč o ponderisanoj oceni iste varijanse u oba skupa, pa je ne treba dovoditi u vezu sa ocenom varijanse proporcije uzorka, S_p^2).

Usled toga, statistika ponderisanog t testa glasi:

Statistika *t* testa	$t = \dfrac{\overline{X}_1 - \overline{X}_2}{S_P \sqrt{(1/n_1 + 1/n_2)}}$	**(8.12)**

Pod pretpostavkom da je H_0 istinita, statistika testa (8.12) ima Studentov raspored sa $n_1 + n_2 - 2$ stepeni slobode.

PROBLEM 8.5: Pretpostavimo da varijanse dva skupa u prethodnom problemu (8.4) nisu poznate, ali da su jednake. Odaberite optimalni test i ispitajte da li je Topiramejt efikasniji od placeba.

REŠENJE

Nulta i alternativna hipoteza imaju identični oblik kao i kod Z testa. Dakle:

$$H_0 : \mu_1 \geq \mu_2 \quad H_1 : \mu_1 < \mu_2 \qquad \alpha = 0{,}05$$

Budući da testiramo razliku između aritmetičkih sredina dva skupa, da su skupovi normalno raspoređeni i da su njihove varijanse nepoznate ali jednake, kao optimalni test biramo ponderisani t test.

Lako je proveriti da smo u postavci problema naveli da su ispunjena oba preduslova za primenu ponderisanog t-testa.

Da bismo izračunali vrednost statistike testa datu formulom (8.12), potrebno je da izračunamo varijanse oba uzorka (aritmetičke sredine uzoraka smo već izračunali prilikom rešavanja problema 8.4.):

$$s_1^2 = \frac{\sum x_i^2 - n_1 \cdot \overline{x}^2}{n_1 - 1} = \frac{97 - 8 \cdot 3{,}375^2}{8 - 1} = 0{,}84$$

$$s_2^2 = \frac{\sum x_i^2 - n_2 \cdot \overline{x}^2}{n_2 - 1} = \frac{183 - 8 \cdot 4{,}625^2}{8 - 1} = 1{,}69$$

Na osnovu dobijenih vrednosti izračunavamo ponderisanu ocenjenu vrednost varijansi u oba skupa na osnovu formule (8.11):

$$s_P^2 = \frac{(8-1) \times 0,84 + (8-1)\times 1,69}{8+8-2} = 1,265,$$

odakle je s_P = 1,09. Vrednost statistike ponderisanog t testa izračunavamo po formuli (8.12):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_P\sqrt{(1/n_1 + 1/n_2)}} = \frac{3,375 - 4,625}{1,125 \cdot \sqrt{1/8 + 1/8}} = -2,22 .$$

Testiranje ćemo sprovesti primenom p-vrednosti, odnosno, odredićemo interval u kome se ona nalazi, vodeći računa da statistika testa sledi t raspored sa 8 + 8 - 2 = 14 stepeni slobode. U Tablici 2 prvo nalazimo red koji odgovara broju stepeni slobode, ν=14. Kao što smo objasnili, koristićemo apsolutnu vrednost statistike t testa, tj. 2,22, i naći kritične vrednosti između kojih se ona nalazi. Vidimo da je 2,1448 < 2,2, < 2,624. Verovatnoće u zaglavlju koje odgovaraju ovim kritičnim vrednostima su 0,025 i 0,01, iz čega sledi: 0,01 < p-vrednost < 0,025.

Da bismo odredili preciznu p-vrednost, ovde ćemo primeniti statistički paket *SPSS*, sa sledećim izlazom.

Independent Samples Test

		t-test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
TOPIRAME	Equal variances assumed	-2,220	14	,043	-1,2500	,56300
	Equal variances not assumed	-2,220	12,564	,045	-1,2500	,56300

Ali, kao što smo već rekli, primenom ovog paketa nije moguće testirati jednosmerne hipoteze, već samo dvosmerne, pa u „izlazu" ne možemo ni dobiti p-vrednost jednosmernog testa. Da bismo ipak utvrdili tačnu vrednost ove verovatnoće, potrebno je da p-vrednost dvosmernog testa, koja iznosi 0,043, podelimo sa 2, na osnovu čega dobijamo P=0,0215. Budući da je $P<\alpha$ nultu hipotezu odbacujemo uz rizik α = 0,05. Samim tim, zaključujemo da Topiramejt statistički značajno smanjuje prosečan broj dana u mesecu u kojima se javlja migrena.

R E Z I M E

Testiranje hipoteza je statistička procedura u kojoj se podaci uzorka koriste da bi se donela odluka da li neku tvrdnju (ili postavku) o nekoj karakteristici skupa treba odbaciti ili ne odbaciti. U ovom poglavlju izučavali smo dva parametarska testa (***Z*** i ***t***) i njihovu primenu u slučaju jednog i dva uzorka. Pri tome smo kao karakteristike skupa posmatrali njegove parametre, konkretno aritmetičku sredinu i proporciju. **Parametarski testovi se zasnivaju na pretpostavci da**

izučavana pojava, odnosno osnovni skup, ima normalan raspored.

Z test smo koristili 1) pri testiranju aritmetičke sredine jednog skupa u (nerealnom) slučaju da je poznata standardna devijacija toga skupa, σ, 2) pri testiranju jednakosti aritmetičkih sredina dva skupa uz pretpostavku da su standardne devijacije oba skupa poznate, 3) prilikom testiranja proporcije jednog skupa ako statistika testa ima približno normalan raspored.

Studentov *t*-test se nikada ne koristi pri testiranju proporcije skupa. Koristili smo ga u dva slučaja: 1) pri testiranju aritmetičke sredine skupa kada je standardna devijacija skupa nepoznata, i 2) kod testiranja jednakosti aritmetičkih sredina dva skupa (uz uslov da oba skupa imaju nepoznate (ali jednake) standardne devijacije).

Logika testiranja hipoteza se zasniva na pronalaženju dovoljno dokaza u uzorku kako bi se nulta hipoteza odbacila. Budući da ne raspolažemo sa svim podacima o skupu, odluku o odbacivanju ili neodbacivanju nulte hipoteze uvek donosimo uz izvestan rizik da smo doneli pogrešnu odluku. Ukoliko odbacimo istinitu nultu hipotezu činimo grešku prve vrste, a verovatnoća da se takva greška javi naziva se **nivo značajnosti testa** i obeležava sa α.

Da bismo doneli odluku o "sudbini" nulte hipoteze moramo da postavimo pravilo odlučivanja. U statistici postoje dva koncepta po ovom pitanju. Po jednom odluka se donosi na osnovu kritičnih vrednosti i formiranja oblasti odbacivanja. Po drugom potrebno je da se izračuna *p*-vrednost. ***P*-vrednost je verovatnoća da statistika test uzme vrednost koja je jednaka ili još više ekstremna od realizovane, pod uslovom da je nulta hipoteza istinita. Pravilo odlučivanja na osnovu *p*-vrednosti glasi: nulta hipoteza se odbacuje ako je *p*-vrednost manja od odabranog nivoa značajnosti, α**. Kada odbacimo nultu hipotezu automatski usvajamo alternativnu kao tačnu, ali obavezno uz odgovarajući rizik greške. Ako je *p*-vrednost veća od nivoa značajnosti nemamo dovoljno dokaza da odbacimo nultu hipotezu. Nultu hipotezu nikada ne usvajamo kao tačnu.

Prilikom testiranja hipoteza vodimo računa o sledećem:

1. Nulta hipoteza kod parametarskih testova se uvek odnosi na određenu vrednost parametra skupa (recimo μ), a ne na statistiku uzorka (recimo $\overline{X}$).

2. Iskaz o nultoj hipotezi uvek sadrži znak jednakosti, bilo da se radi o prostoj ili složenoj hipotezi. Alternativna hipoteza nikada ne sadrži znak jednakosti).

3. Uvek prilikom tumačenja koristimo izraz "statitička značajnost", a ne samo značajnost.

4. **Nikada rezultate testiranja ne tumačimo kao kod ocenjivanja rečima "uz pouzdanost"**.

5. U statistici je pogrešno pri neodbacivanju nulte hipoteze interpretirati rezultat rečima da se nulta hipoteza usvaja. **U statistici nikada ne usvajamo nultu hipotezu**, možemo je samo odbaciti ili ne odbaciti.

KLJUČNI NOVI POJMOVI

Nulta hipoteza	Alternativna hipoteza
Greška prve vrste	Greška druge vrste
Nivo značajnosti	Jačina testa
Statistika testa	*p*-vrednost
Dvosmerni test	Statistička značajnost
Parametarski test	Jednosmerni test
Studentov t-test	Z test

KONTROLNA PITANJA I ZADACI

1. Kada u statističkoj analizi primenjujemo ocenjivanje, a kada testiranje?

2. Objasnite razliku između nulte i alternativne hipoteze.

3. Kada primenjujemo jednosmerni, a kada dvosmerni test?

4. U čemu je razlika između parametarskih i neparametarskih testova?

5. Koji test primenjujemo ako testiramo hipotezu o aritmetičkoj sredini skupa, a standardna devijacija tog skupa je nepoznata?

6. Koji su neophodni uslovi da bi se primenio Z test u ispitivanju hipoteze o proporciji jednog skupa?

7. Kada odbacujemo nultu hipotezu primenom *p*-vrednosti?

8. Kada u statistici usvajamo nultu hipotezu kao tačnu?

9. U čemu je razlika između greške prve i druge vrste? A rizika greške prve i druge vrste?

10. Proizvođač automobilskih guma tvrdi da je prosečan vek trajanja veći od 40000 kilometara, a da je standardna devijacija 5000 km. Ako je u uzorku od 150 guma prosečna dužina trajanja 40700 km, ispitajte, uz rizik α = 0,05, da li se može prihvatiti tvrdnja proizvođača?

11. U slučajnom uzorku od 25 pakovanja jednog proizvoda prosečna težina iznosi 489 grama. Propisana težina pakovanja je 500 grama, a standardna devijacija iznosi 15 gr. Ako težina pakovanja ima normalan raspored, utvrdite da li se na osnovu informacije iz uzorka može zaključiti da težina pakovanja odstupa od standarda? (rizik α = 0, 01).

12. U reklami jednog servisa se tvrdi da prosečno vreme opsluživanja jednog klijenta iznosi manje od 30 minuta, a da je standardna devijacija 7 minuta. Ako smo u slučajnom uzorku od 36 klijenata zabeležili prosečno vreme opsluživanja od 28 minuta, ispitajte, uz rizik od 0,01, da li možemo prihvatiti tvrđenje servisa kao tačno?

13. U slučajnom uzorku od 100 vozača, 77 vozača je koristilo sigurnosni pojas. Da li na osnovu uzorka možemo prihvatiti hipotezu da više od 75% vozača koristi sigurnosni pojas?

14. Nabavljač katodnih cevi će prihvatiti pošiljku ako je prosečno vreme trajanja veće od 5000^h. U uzorku od devet cevi dobijeni su sledeći rezultati: $\overline{x} = 5060$ i $s^2 = 2500$. Pod pretpostavkom da trajanje cevi ima normalan raspored, zaključite, primenom p-vrednosti, da li pošiljku treba prihvatiti.

15. Propisani prečnik kugličnih ležaja je 5cm, sa standardnom devijacijom 0,lcm. U slučajnom uzorku od 100 kugličnih ležaja prosečna dužina prečnika je 5,07 cm. Zaključite da li proizvod odgovara propisanom standardu; rizik α= 0,05.

16. Iz prispele pošiljke od 1000 pakovanja kafe, čija je ugovorena težina 200 gr, a varijansa 49 gr^2 formiran je uzorak od 80 pakovanja. Prosečna težina u uzorku je 198,5 gr. Uz rizik α= 0,05 proverite da li pošiljka zadovoljava ugovorom predviđenu težinu.

17. Od 500 kupaca u jednoj prodavnici u toku jednog dana, slučajnim putem smo odabrali 100 i zabeležili da je prosečna vrednost njihovih kupovina 758 dinara. Prodavac tvrdi da je prosečni dnevni pazar manji od 780 dinara, a da varijansa iznosi $\sigma^2 = 10000$. Ispitajte, na osnovu p-vrednosti, da li možemo prihvatiti tvrđenje prodavca. Šta ćete zaključiti ako je $\alpha = 0,05$, a šta ako je $\alpha = 0,01$?

18. Ako je u slučajnom uzorku od 60 elemenata aritmetička sredina jednaka 25, testirajte, uz rizik α =0,05 sledeće hipoteze:

$H_0 : \mu = 28,$ $\qquad H_1 : \mu \neq 28$

$H_0 : \mu \leq 23,$ $\qquad H_1 : \mu > 23$

$H_0 : \mu \geq 27,$ $\qquad H_1 : \mu < 27$

Varijansa osnovnog skupa je 150.

19. Proizvod odgovara standardima ako je njegov radni vek duži od godinu dana. U uzorku od 200 proizvoda prosečno vreme trajanja bilo je 11,5 meseci, tj. 0,958 godina. Odredite, uz rizik α=0,05, da li se proizvođač pridržava standarda, ako je varijansa u uzorku 2,l godina.

20. Standardna devijacija osnovnog skupa je 500. Ako je u uzorku od 35 elemenata aritmetička sredina jednaka 8800, ispitajte, uz rizik greške I vrste od 0,01 da li se može prihvatiti pretpostavka da je aritmetička sredina skupa manja od 9000? Izračunajte p-vrednost.

21. Ako je u uzorku od 12 studenata prosečan broj položenih ispita u jednom ispitnom roku 2,7, a standardna devijacija 0,4, ispitajte da li se može prihvatiti pretpostavka da je prosečan broj položenih ispita u skupu od 100 studenata manji od 3? Rizik α je 0,01.

22. Na osnovu informacije uzorka od 100 anketiranih odraslih osoba, od kojih je 85 izjavilo da prati političke informativne emisije na TV, ispitajte da li se mogu prihvatiti sledeće pretpostavke:

a) da je gledanost veća od 80%;

b) da je gledanost manja od 90%.

23. Ako je broj "uspeha" u uzorku od 100 elemenata jednak 73, testirajte sledeće hipoteze:

a) $H_0: \pi \leq 0{,}71$, $H_1: \pi > 0{,}71$

b) $H_0: \pi = 0{,}70$, $H_1: \pi \neq 0{,}70$

c) $H_0: \pi \geq 0{,}745$, $H_1: \pi < 0{,}745$

Uporedite rezultate i objasnite ih.

24. U uzorku od 30 studenata društvenih fakulteta prosečno vreme učenja u toku dana je 150 minuta, a standardna devijacija 14 minuta. U slučajnom uzorku od 40 studenata tehničkih fakulteta, prosečno vreme učenja je 160 minuta, sa standardnom devijacijom 10 minuta. Testirajte nultu hipotezu da studenti dve grupe fakulteta u proseku provode isto vreme u učenju; pretpostavka je da su varijanse oba skupa jednake medu sobom.

25. Ako je u uzorku od 10 prodavnica štampe u gradu A, prosečna mesečna prodaja jednaka 4000 dnevnih listova i časopisa, a u uzorku od 8 novinarnica u gradu B ona iznosi 4300, ispitajte da li se gradovi razlikuju u kupovini štampe. Standardna devijacija u prvom uzorku je 200 komada, a u drugom 250; pretpostavlja se da su varijanse skupova među sobom jednake.

26. U jednoj auto kući, koja se bavi prodajom stranih automobila, anketirano je 400 potencijalnih kupaca i konstatovano da je kod 250 potencijalnih kupaca prethodni automobil bio domaće proizvodnje.

a) Ispitati tvrdnju agenta ove auto kuće da je kod 70% potencijalnih kupaca prethodni automobil bio automobil domaće proizvodnje.

b) Uz rizik greške 0,05, ispitati tvrdnju agenta ove auto kuće da je proporcija potencijalnih kupaca čiji je prethodni automobil bio automobil domaće proizvodnje manja od 75%.

c) Uz rizik greške 0,05, ispitati tvrdnju agenta ove auto kuće da je proporcija potencijalnih kupaca čiji prethodni automobil nije bio automobil domaće proizvodnje veća od 20%.

27. Na osnovu anketiranja 30 nezaposlenih lica koja se nalaze na birou za zapošljavanje izračunata je njihova prosečna starost 30,77 godina i standardna devijacija 8,62 godine.

a) Uz nivo značajnosti testa 0,05, ispitati da li se može prihvatiti pretpostavka da prosečna starost nezaposlenih lica značajno odstupa od 26 godina.

b) Uz nivo značajnosti testa 0,05, ispitati da li se može prihvatiti pretpostavka da je prosečna starost nezaposlenih lica veća od 28 godina.

c) Uz nivo značajnosti testa 0,05, ispitati da li se može prihvatiti pretpostavka da je prosečna starost nezaposlenih lica manja od 34 godine.