# OSNOVE METODE UZORAKA

## Uvod

U svakidašnjem govoru često se upotrebljava riječ uzorak. Tako se govori o robnim uzorcima, uzorcima na izložbenim prostorima i slično. Pojam uzorka u sklopu statistike ima značenje koje se razlikuje od onoga u svakidašnjem govoru. Statistički pojam spomenut je prije pri opisu podjele statistike kao metode proučavanja pojava i procesa. Istakli smo da se statistika dijeli na deskriptivnu i inferencijalnu.

Dodatni sadržaj veže se uz sadržaj u udžbeniku Šošić, I., *Statistika*, Školska knjiga. Zagreb, 2014. na stranici 30.

Deskriptivna statistika sastoji se od skupa metoda kojima se sažimaju obavijesti na temelju prikupljenih podataka o odabranoj pojavi ili procesu. Zaključci dobiveni uređivanjem, grafičkim prikazivanjem i brojčanom obradbom odnose se isključivo na jedinice skupa o kojima su prikupljeni podatci. Metode deskriptivne statistike ne služe poopćivanju zaključaka dobivenih promatranjem obilježja manjeg broja jedinica (uzorka) na sve jedinice (statistički skup). U sklopu te statistike jednako se provode analitički postupci bez obzira na to odnosi li se skup podataka na sve jedinice ili dio njih. Inferencijalna statistika[1] sastoji se od primjene postupaka kojima se na temelju podataka o dijelu jedinica (slučajnom uzorku) donose sudovi o cijelom skupu.

**UZORAK** je podskup statističkog skupa. S pomoću podataka iz uzorka zaključuje se o svojstvima statističkog skupa. Skup iz kojeg se izabire uzorak još se naziva osnovnim skupom ili populacijom Zaključci o osobitostima skupa temelje se na podatcima o dijelu jedinica, stoga sadržavaju pogrešku koja se naziva pogreškom nastalom zbog primjene uzorka.

Uzorci se primjenjuju u ispitivanju mnogih pojava. Televizijske kompanije uzorkom ispituju gledanost pojedinih programa, proizvodna poduzeća uzorkom prate proizvodne procese, istraživanje tržišta kao po pravilu temelji se na uzorcima potrošača. Isto je s ispitivanjem javnog mišljenja. U prvom smo poglavlju naveli primjer ispitivanja biračkog tijela prije izbora. Ono se ispituje tako da se prikupe podatci za manji broj birača. S pomoću podataka koji čine uzorak izračunava se proporcija birača za jednog kandidata u uzorku. Proporcija birača u uzorku jest *procjena* proporcije *svih* birača za danoga kandidata. Ta veličina sadržava pogrešku koja izvire iz primjene uzorka, to jest uporabe dijela, a ne svih podataka.

Uzmimo drugi primjer. Proizvođač baterija dužan je ispitivati kakvoću svojih proizvoda. Deklarirana je trajnost baterija u prosjeku 200 sati. Ispitivanje proizvodne serije provodi se s pomoću uzorka baterija. Rezultati iz uzorka upućuju na zaključak da je moguće prihvatiti pretpostavku da je prosječna trajnost baterija proizvodne serije 200 sati.

U prvom je primjeru bila riječ o procjeni, a u drugome o *ispitivanju pretpostavke* o osobitosti (parametru) osnovnog skupa. I u drugom je primjeru mo-

[1] Inferencijalna statistika polazi od slučajnog uzorka i temelji se na teoriji vjerojatnosti. Poznavanje pojmova kao što su vjerojatnost, slučajna varijabla i distribucije vjerojatnosti nužno je da bi se prosudili rezultati dobiveni primjenom metode uzoraka.

guća pogreška u zaključivanju zbog primjene uzorka. Moguće je da se prihvati lažna pretpostavka kao istinita ili pak da se odbaci istinita pretpostavka. Pogreške ne bi postajale kad bi se prikupili podatci za sve birače ili ispitala trajnost svih baterija serije.

Budući da sudovi doneseni na temelju uzorka sadržavaju pogrešku, možemo postaviti pitanje o razlozima primjene metode uzoraka. Poželjno je (ako je to moguće) da se u istraživanju obuhvate sve jedinice statističkog skupa jer se tako dolazi do pokazatelja koji ne sadržavaju pogrešku navedenog izvora. No ima više faktora koji uvjetuju primjenu metode uzorka u ispitivanju pojava.

Statistički skupovi u pojedinim slučajevima sadržavaju golem broj jedinica. Skupovi su televizijskih gledatelja, kućanstava ili skupovi potrošača jedne zemlje višemilijunski. Provedba statističkog istraživanja takvih skupova vrlo je skupa. Visoki su troškovi organizacije i obradbe. Istraživanje s pomoću uzorka zahtijeva *manja financijska sredstva*.

Do rezultata na temelju podataka za sve jedinice skupa dolazi se s vremenskim zakašnjenjem, što može negativno djelovati na njihovu uporabnu vrijednost. Istražuje li se struktura potrošnje kućanstava kojih je više milijuna, za prikupljanje podataka i njihovu analizu valja utrošiti mnogo sredstava i vremena. Dobiveni podatci i rezultati u manjoj ili većoj mjeri zastarijevaju s vremenom jer se struktura potrošnje mijenja. Za poslovno odlučivanje i u drugim djelatnostima vrlo je važno u kratkom roku doći do potrebnih informacija. S pomoću uzorka u *kraćem se vremenu* dobivaju obavijesti o traženim obilježjima.

Pojedina istraživanja zahtijevaju uz velika financijskih sredstva vrlo stručno osoblje i potrebnu opremu. S obzirom na to da uzorak sadržava manji broj jedinica, njegovom uporabom pruža se mogućnost provedbe *zamršenijih i opsežnijih istraživanja*.

Tijekom prikupljanja podataka pojavljuju se pogreške različite vrste. Točnost podataka vrlo je važna jer statističko-analitičke veličine računane na temelju netočnih podataka nisu valjane. Zbog toga se prije prikupljanja podataka, tijekom i nakon prikupljanja provode kontrole točnosti i uklanjaju uočene pogreške. Kontrola točnosti vrlo je zahtjevan i opsežan posao jer je kadšto riječ o kontroli milijuna podataka. Rabi li se uzorak, broj podataka koje valja kontrolirati znatno je smanjen, a kontrola uspješnija. Stoga je moguće s pomoću uzorka postići i *veći stupanj točnosti* nego što je to slučaj pri iscrpnom promatranju.

Na uporabu uzorka utječe i priroda objekata istraživanja. Ispituje li se prosječna trajnost proizvedenih baterija odabrane vrste, nužno ih je staviti u uporabu i evidentirati vrijeme trajanja. Takvo ispitivanje ne može se provesti na temelju svih baterija dane proizvodne serije jer bi to dovelo do njihova uništenja. Prema tome, *metoda uzoraka primjenjuje se kada iscrpno promatranje dovodi do uništenja stvari, objekata, proizvoda čija se svojstva ispituju*.

Navedeni primjeri u svezi su s primjenom metode uzoraka koja se odnosi na *konačne realne statističke skupove*. To su skupovi s konačnim brojem članova koji u vremenu istraživanja postoje. Broji li srednjoškolski centar 2 789 učenika, riječ je o konačnome realnom skupu čije su jedinice učenici upisani dane školske godine. Iz tog skupa radi procjene prosječnog vremena provedenog u gledanju televizijskih programa izabire 125 učenika koji tvore uzorak jedinica.

No nisu svi statistički skupovi konačni i realni. Postoje *beskonačni skupovi*. Ispitivanje svojstava takvih skupova moguće je samo metodom uzoraka. Beskonačan skup ima beskonačno mnogo članova. Takvi skupovi u svezi su s nekim postupkom, procesom koji se teorijski zbiva neprekidno i ravna po zakonima vjerojatnosti ili je riječ o mjerenjima koja se mogu (teorijski promatrano) ponavljati beskonačno mnogo puta.

Proizvodnja jednog proizvoda na automatskoj liniji može se smatrati statističkim procesom koji stvara elemente beskonačnog skupa i čiju kakvoću valja ispitati. Tijekom proizvodne smjene kontrolor izabire radi kontrole težine uzorak i s pomoću mjerenja težine proizvoda u uzorku donosi zaključak o prosječnoj težini proizvoda. Zaključne cijene dionica na burzi tijekom odabranog razdoblja također se mogu smatrati uzorkom zamišljene populacije. Kretanje budućih cijena nije matematički predvidivo: one se razvijaju statistički, odnosno po zakonima vjerojatnosti. Da bismo primijenili metodu, potrebno je poznavati svojstva rasporeda populacije, odnosno oblik distribucije vjerojatnosti prema kojoj se zbiva statistički proces.

Pozornost ćemo usmjeriti na metodu koja polazi od uzorka iz konačnih realnih skupova. Zadatci analize podataka na temelju uzoraka iz konačnih i beskonačnih skupova po pravilu su jednaki.

**ZADAĆE METODE UZORAKA.** Dvije su osnovne zadaće metode uzoraka. Prva se zadaća sastoji od procjenjivanja nepoznatih parametara, a druga od testiranja pretpostavki o parametrima i oblicima rasporeda osnovnih skupova. Parametar je funkcija vrijednosti obilježja svih jedinica osnovnog skupa.

Parametri su, primjerice, prosječna plaća u siječnju 2014. godine *svih* zaposlenika u Republici Hrvatskoj i prosječno odstupanje od prosječne plaće. Riječ je o aritmetičkoj sredini i standardnoj devijaciji koje se računaju s pomoću prije predočenih formula, i to na temelju podataka *za svaku* jedinicu statističkog skupa. Prema navedenome, razvidno je da je parametar funkcija vrijednosti osnovnog skupa. Aritmetička sredina skupa koja broji $N$ članova jest parametar dan izrazom $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.

**PARAMETAR, PROCJENITELJ PARAMETRA.** Izabere li se iz osnovnog skupa $N$ elemenata uzorak n elemenata ($n < N$), parametar se procjenjuje, a izraz (funkcija, formula) s pomoću koje se procjenjuje naziva se procjeniteljem. Procjenitelj je funkcija vrijednosti iz uzorka.

Procjenitelj aritmetičke sredine osnovnog skupa jest aritmetička sredina uzorka $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Očito je da je riječ o funkciji vrijednosti iz uzorka koja dovodi do procjene parametra jednim brojem.

**PROCJENA PARAMETRA BROJEM I INTERVALOM.** S pomoću procjenitelja i vrijednosti iz uzorka parametar se procjenjuje brojem i intervalom. Dobivene vrijednosti nazivaju se procjenama. Intervalni procjenitelj dovodi nas do raspona vrijednosti u kojem se očekuje da će se nalaziti parametar. I granice intervala procjene funkcije jesu vrijednosti iz uzorka.

Pretpostavimo da se s pomoću uzorka procjenjuje prosječna vrijednost izvanpansionske potrošnje turista u Republici Hrvatskoj 2013. godine. Parametar je u tom primjeru aritmetička sredina, to jest prosječna vrijednost izvanpansionske potrošnje *svih* turista. Vrijednost parametra mogli bismo izračunati samo kad bismo raspolagali podatkom o izvanpansionskoj potrošnji *svakog* turista.

Prosječna vrijednost potrošnje može se procijeniti. Pođemo li od *uzorka izabrana* prema načelima statističke teorije, *procjenitelj* parametra jest funkcija uzorka određena izrazom aritmetičke sredine uzorka. Na temelju podataka iz uzorka izračunat ćemo njegovu aritmetičku sredinu. Neka je ta vrijednost 200 američkih dolara. Ta je vrijednost *procjena* jednim brojem izvanpansionske potrošnje *svih* turista. S pomoću tog rezultata ne možemo zaključivati o pogrešci nastaloj zbog primjene uzorka ni o preciznosti procjene. Primjenom intervalnog procjenitelja dolazimo do ovog rezultata: s vjerojatnošću 95 % očekujemo da se prosječna izvanpansionska potrošnja turista kreće između 190 i 210 američkih dolara. Očito je da se s pomoću intervalne procjene može zaključivati o preciznosti procjene. Preciznost procjene to je veća što su granice intervala procjene bliže.

Osim spomenutih, u sklopu metode uzoraka procjenjuju se i drugi parametri, kao što su proporcija, medijan, parametri u regresijskome modelu, koeficijent korelacije i dr.

Drugi se zadatak metode uzoraka sastoji od ispitivanja pretpostavki (hipoteza) o osobitostima jednoga ili više osnovnih skupova (populacija).

> **STATISTIČKA PRETPOSTAVKA (HIPOTEZA)** jest tvrdnja o veličini parametra ili obliku rasporeda jednoga ili više osnovnih skupova čija se vjerodostojnost ispituje s pomoću uzorka. Postupci ili pravila kojima se dolazi do odluke o prihvaćanju ili odbacivanju tvrdnje na temelju podataka iz uzorka čine područje ispitivanje hipoteza.

Svaki postupak ispitivanja statističkih hipoteza počiva na dvjema tvrdnjama koje su sadržajno u suprotnosti. Jedna je tvrdnja sadržana u *nultoj pretpostavci*, a druga u *alternativnoj pretpostavci*. Postupkom testiranja s pomoću podataka iz uzorka donosi se statistička odluka o tome može li se prihvatiti nulta pretpostavka kao istinita ili se ona odbacuje. Sadržaj i oblici pretpostavki ovise o slučaju primjene.

**PRIMJER 1.**

Služba za istraživanje tržišta ispituje starosnu karakteristiku potrošača novog napitka. Pretpostavlja se da je prosječna starost potrošača danog napitka 30 godina. Stvarna prosječna starost svih potrošača napitka (parametar, aritmetička sredina populacije) nije poznata. Navedena pretpostavka ispitat će se s pomoću slučajnog uzorka potrošača. S $\mu$ će se označiti nepoznata aritmetička sredina populacije, a s $\mu_0$ njezina pretpostavljena vrijednost, koja iznosi 30 godina, Pretpostavke su za taj test sljedeće.

*Nulta pretpostavka* ($H_o$): prosječna starost potrošača *je* 30 godina.

*Alternativna pretpostavka* ($H_1$): prosječna starost potrošača *različita* je od 30 godina.

Neka je u slučajni uzorak izabrano 265 potrošača. Prosječna starost potrošača u uzorku je 28.9 godina (aritmetička sredina uzorka, $\overline{x}$ ). Razlika aritmetičke sredine uzorka i pretpostavljene aritmetičke sredine jest: $\overline{x} - \mu_0 = 28.9 - 30 = -1.1$.

Sada se postavlja pitanje je li ta razlika značajna ili slučajna, odnosno izvire li ona iz uporabe uzorka ili tvrdnja u nultoj pretpostavci nije istinita? Do odgovora na to pitanje dolazi se primjenom postupaka u sklopu metode uzoraka. Ako razlika nije statistički značajna, prihvatit će se tvrdnja sadržana u nultoj pretpostavci: prihvatit će se pretpostavka da je prosječna starost potrošača napitka 30 godina, a razlika će se smatrati slučajnom (nesignifikantnom). U protivnom, nulta se pretpostavka ne prihvaća, to jest rezultat govori u prilog prihvaćanja suprotne tvrdnje (tvrdnje u alternativnoj pretpostavci, prosječna starost potrošača nije 30 godina, nego se razlikuje od te starosti, razlika je značajna, signifikantna).

Je li prihvaćena tvrdnja o starosti potrošača *doista* istinita? Budući da se zaključak o pretpostavkama donosi s pomoću dijela podataka, u zaključivanju je moguće pogriješiti. Pogriješit će se ako se odbaci pretpostavka da je prosječna starost potrošača 30 godina, a prosječna starost potrošača je doista toliko godina. Pogriješit će se u testiranju prihvati li se pretpostavka da je prosječna starost 30 godina, a ona to nije.

**VRSTE POGREŠAKA U STATISTIČKOM TESTU.** U postupku ispitivanja pretpostavki s pomoću statističkih uzoraka moguće je počiniti dvije vrste pogrešaka. Ako se odbaci istinita nulta pretpostavka, počini se pogreška I. vrste. Pogreška II. vrste počini se prihvati li se tvrdnja u nultoj pretpostavci premda je ta tvrdnja lažna.

Odluka u postupku testiranja ispravna je prihvati li se istinita nulta pretpostavka ili odbaci li se lažna nulta pretpostavka. Ako je riječ o slučajnom uzorku, pogreške u zaključivanju izražavaju se s pomoću vjerojatnosti. Pogreška I. vrste jest vjerojatnost odbacivanja istinite nulte pretpostavke, a naziva se razinom značajnosti (signifikantnosti). Pogreška II. vrste jest vjerojatnost prihvaćanja lažne nulte pretpostavke kao istinite. U nacrtu ispitivanja s pomoću uzorka određuje se veličina tih pogrešaka, odnosno vjerojatnosti, a one ovise o posljedicama za dani problem.

## *PITANJA I ZADATCI ZA VJEŽBU*

1. Što je osnovni skup (populacija), a što uzorak? Navedite primjere realnih konačnih skupova i objasnite što bi bio uzorak. Koje je opće obilježje zaključivanja s pomoću uzoraka?
2. Opišite prednosti i nedostatke istraživanja s pomoću uzorka u usporedbi s istraživanjem na temelju podataka dobivenih iscrpnim promatranjem (cenzusom).
3. U kojim se slučajevima statistička analiza uvijek oslanja na uzorak? Navedite primjere.
4. Definirajte pojmove: *parametar, procjenitelj brojem, intervalni procjenitelj*.
5. Osnovni skup sastoji se od devet zaposlenika ($N = 9$), čiji je radni staž (u godinama) prikazan u tablici.

| $x_i$ : | 2 | 8 | 4 | 2 | 4 | 10 | 6 | 8 | 7 |
|---|---|---|---|---|---|---|---|---|---|

Primjenom izraza $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ i $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$ dobivene su ove vrijednosti: $\mu = 5.66667, \sigma = 2.66667$ . Kako se nazivaju navedeni izrazi (formule) i dobivene veličine? (b) Iz osnovnog skupa od devet članova moguće je izabrati bez ponavljanja 84 različita uzorka veličine $n = 3$. Pretpostavimo da su u uzorak izabrani zaposlenici s ovim radnim stažem: 8, 4, 6 godina. Kako se naziva ovaj izraz: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ? Izračunana vrijednost prethodnog izraza jest $\overline{x} = 6$ godina. Što predočuje ta vrijednost? (c) Dan je ovaj izraz: $\overline{x} - pogrješka < \mu < \overline{x} + pogrješka$. Što predočuje navedeni izraz? Primjenom odgovarajućih veličina dobivena je vrijednost prethodnog izraza: $3.493 < \mu < 7.841$ . Kako se naziva navedeni izraz? Napomena: ovaj je zadatak u svezi sa zadatkom 5. i svrha mu je da se ilustriraju dani pojmovi. U praksi, dakako, nećemo primijeniti metodu uzoraka ako je riječ o konačnom realnom skupu s malim brojem članova.

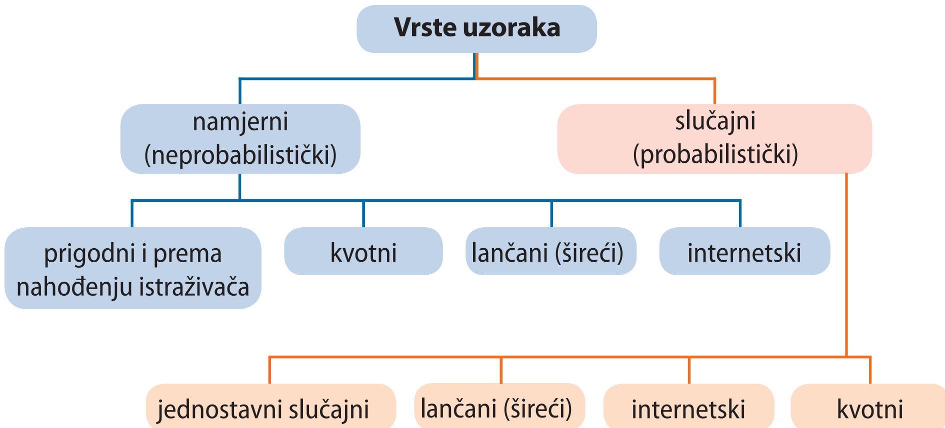
**6.** Rukovodstvo tvrtke Voće pretpostavlja da je proporcija svih srednjoškolaca koji redovito kupuju njihov sok od naranče 0.50, odnosno 50 %. Radi ispitivanja te pretpostavke izabran je slučajni uzorak 200 učenika. Proporcija učenika u uzorku koji kupuju sok od naranče navedene tvrtke je 0.52. (a) Kako se nazivaju sljedeće tvrdnje:

$H_0 ... p = 0.50;\ H_1 ... \ p \neq 0.50$ ?

Protumačite im sadržaj. (b) Odgovarajućim postupkom utvrđeno je da razlika pretpostavljene vrijednosti parametra (proporcije populacije) i proporcije uzorka $(p_0 - \hat{p} = 0.50 - 0.52 = -0.02)$ nije statistički značajna i da se prihvaća tvrdnja sadržana u nultoj hipotezi. Što u ovom primjeru znači pogreška I. vrste i pogreška II. vrste?

## Vrste uzoraka

Istraživanje pojava uzorkom (procjene i ispitivanje pretpostavki) u uskoj je vezi s planom uzorka. Najvažnija svrha plana jest da se osigura izbor reprezentativnog uzorka. Uzorak će biti reprezentativan ako svojim osobitostima naliku-je skupu iz kojeg je izabran. Drugim riječima, uzorak bi trebao biti umanjena slika osnovnog skupa. Reprezentativnost uzorka postiže se u načelu ispravnim postavljanjem plana uzorka, koji uključuje i način izbora jedinica u uzorak. Uzorci se razlikuju i u svezi sa svojim oblikom (nacrtom, modelom).




**Slika 1.** *Odabrane vrste uzoraka*

**NAMJERNI I SLUČAJNI UZORAK.** S obzirom na način izbora jedinica, razlikuje se namjerni od slučajnog uzorka. U namjerni uzorak izabiru se jedinice prema odluci istraživača (anketara, analitičara). Ako svaka jedinica osnovnog skupa može biti izabrana u uzorak s određenom vjerojatnosti, onda se takav uzorak naziva slučajnim.

Među namjernim uzorcima je *prigodni uzorak*. Podatci u uzorku dobiveni su ispitivanjem jednostavno dostupnih članova skupa. Mišljenja osoba zatečenih u robnoj kući o degustiranom proizvodu koja je zabilježio anketar u jednom danu, tvore prigodni uzorak. Istoj vrsti uzorka pripadaju i podatci o stajalištima građana o nekom pitanju koja je prikupio radijski ili televizijski izvjestitelj na ulici ili trgu.

Namjeran je i *kvotni uzorak*. U postupku izbora anketari se po volji opredjeljuju za jedinice u sklopu kvota. Kvotu čine jedinice određenog svojstva. Ako se ispituje npr. mišljenje kupaca o jednom proizvodu i ako je poznato da dani proizvod kupuje 30 % muškaraca i 70 % žena, uzorak će se sastojati od dviju kvota, tj. u njemu će biti zastupljeno približno 30 % muških i 70 % ženskih osoba. Izbor kupaca prepušta se anketaru.

*Lančani uzorak* (šireći) jest namjerni uzorak. Nastaje, primjerice, tako da određen broj osoba koje je izabrao istraživač u prvom koraku uključuje u drugom koraku uzorak osoba njihovih poznanika sličnih svojstava. Osobe izabrane u uzorak u drugom koraku povećavaju uzorak poznatim osobama sličnih svojstava itd. Postupak formiranja lančanog uzorka završava izborom onoliko osoba kolika je planirana veličina uzorka.

*Internetski uzorak (web-uzorak)* u svezi je s prikupljanjem podataka utvrđenih upitnikom na internetskoj stranici. Riječ je o ispitivanju javnog mišljenja, ocjene sadržaja ponuđenih informacija na internetskim stranicama institucija ili tvrtki i slično. Dobiveni podatci registriraju se, a rezultati se na stranicama predočuju u obliku tablica, grafikona i jednostavnih brojčanih pokazatelja. Podatci dobiveni anketiranjem gledatelja na poticaj voditelja televizijskog programa imaju obilježja slična podatcima internetskog uzorka.

Istraživanja s pomoću namjernih uzoraka po pravilu su jednostavna i iziskuju relativno male troškove, do rezultata se dolazi u kratkom vremenu, što uvjetuje njihovu relativno čestu primjenu. Kakvoća zaključaka uvelike ovisi o poznavanju pojava koje se ispituju s pomoću uzorka.

S pomoću podataka koji tvore uzorak i pokazatelja izračunanih s pomoću uzorka zaključujemo o svojstvima osnovnog skupa. Kako smo nekoliko puta istaknuli, ti zaključci sadržavaju pogrešku koja je posljedica uporabe dijela, a ne svih podataka. Za prosudbu kakvoće zaključaka s pomoću podataka iz uzorka veoma je važno brojčano izraziti tu pogrešku. Primjenjuje li se u istraživanju namjerni uzorak, *ta se pogreška ne može izračunati*.

Istraživanja s pomoću namjernih uzoraka po pravilu su jednostavna i iziskuju relativno male troškove, do rezultata se dolazi u kratkom vremenu, što uvjetuje njihovu relativno čestu primjenu. Kakvoća zaključaka uvelike ovisi o poznavanju pojava koje se ispituju s pomoću uzorka.

*Slučajni uzorak* izabire se tako da svaki element osnovnog skupa može biti uvršten u uzorak. Drugim riječima, svaki član skupa ima vjerojatnost izbora u uzorak veću od nule. Podatci dobiveni s pomoću uzorka analiziraju se prema načelima inferencijalne statistike. Za pokazatelje izračunate s pomoću uzorka, koji označuju *procjene* svojstava osnovnog skupa, *mogu se izračunati pogreške nastale primjenom uzorka*, što je vrlo važno za prosudbu kakvoće zaključivanja.

Državni zavodi za statistiku, instituti za istraživanje javnog mišljenja i druge organizacije kojima je to predmet djelatnosti, kada je moguće, u istraživanju polaze od slučajnih uzoraka. Uz rezultate na temelju slučajnih uzoraka priopćuju se i pogreške zbog primjene uzorka. Procjenjuje li se postotak birača za jednoga kandidata na nadolazećim izborima s pomoću slučajnog uzorka, onda bi rezultat mogao biti ovakav: procjena postotka svih birača za danoga kandidata s 95 % pouzdanosti jest 37 % i pogreškom $\pm 3\%$ .

Do sada je bilo riječi o podjeli uzoraka na namjerne i slučajne. Među namjernim uzorcima spomenuti su prigodni, prosudbeni i kvotni. I slučajni uzorci različitih su oblika. S tim u svezi govori se o različitim nacrtima (modelima, dizajnima) slučajnih uzoraka. Izbor nacrta uzorka sastavni je dio plana, a ovisi o cilju ispitivanja, prirodi skupa koji se istražuje s pomoću uzorka, raspoloživim financijskim sredstvima i drugim čimbenicima.

## PITANJA

**1.** Za koji se uzorak kaže da je namjerni, a za koji da je slučajni?
**2.** Što je: (a) prigodni uzorak, (b) prosudbeni uzorak, (c) kvotni uzorak?

## Osnovna obilježja odabranih modela slučajnih uzoraka

Među nacrtima slučajnih uzoraka iz *konačnih* skupova jesu i ovi: jednostavni slučajni uzorak, stratificirani uzorak i uzorak skupina.

**JEDNOSTAVNI SLUČAJNI UZORAK**. Izabere li se iz skupa od *N* elemenata n elemenata ($n < N$) tako da svaki element ima jednaku vjerojatnost izbora, onda se takav uzorak naziva jednostavnim slučajnim uzorkom.

Postupak izbora jedinica statističkog skupa u jednostavni slučajni uzorak može se obaviti na više načina. U praksi se izbor u uzorak iz konačnoga stvarnog skupa gotovo uvijek provodi s pomoću *tablica slučajnih brojeva*, odnosno pomoću odgovarajućeg *programa za računalo* koji stvara slučajne brojeve. Izvadak iz tablice slučajnih brojeva prikazan je u tablici 1.2.

**Tablica 1.** Izvadak iz tablice slučajnih brojeva

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 5 | 9 | 1 | 5 | 1 | 7 | 6 | 4 | 1 |
| 8 | 0 | 0 | 0 | 0 | 9 | 4 | 2 | 7 | 9 |
| 0 | 2 | 5 | 3 | 0 | 0 | 0 | 9 | 2 | 4 |
| 1 | 9 | 3 | 2 | 7 | 2 | 3 | 5 | 9 | 0 |
| 3 | 1 | 1 | 3 | 7 | 2 | 0 | 8 | 4 | 6 |
| 7 | 3 | 4 | 8 | 9 | 4 | 3 | 3 | 5 | 5 |
| 7 | 1 | 8 | 3 | 7 | 0 | 2 | 0 | 1 | 8 |
| 0 | 1 | 9 | 0 | 8 | 0 | 8 | 2 | 9 | 1 |

Ako je riječ o jednostavnome slučajnom uzorku, to jest uzorku koji se formira tako da svaki element skupa ima *jednaku* vjerojatnost izbora, primjenjuju se tablice slučajnih brojeva u kojima svaka znamenka i skupina od istog broja znamenaka ima jednaku vjerojatnost pojavljivanja (jednaku relativnu frekvenciju).

Prvi korak u izboru slučajnog uzorka sastoji se od uređivanja *okvira izbora*. Okvir izbora sastoji se od popisa brojčano označenih članova statističkog skupa, a to može biti birački popis, platna lista u poduzeću, telefonski imenik i sl. Uzmimo da su brojčane oznake jedinica skupa od 1 do *N*. Zatim se iz tablica izabiru skupine znamenaka. Skupina ima onoliko znamenaka koliko ih ima i broj koji predočuje opseg skupa. Pročitana kao broj, svaka skupina znamenaka jest redni broj člana statističkog skupa u sklopu izbora, ako takav redni broj postoji.

Neka statistički skup ima 80 članova ($N = 80$), a iz njega se izabire uzorak od $n = 8$. S obzirom na to da je opseg osnovnog skupa dvoznamenkasti broj (80), iz tablica će se prema odabranom redu po volji, izdvojiti skupine od dvije znamenke. Kako se izabire osam članova, izdvaja se njih osam. Uzimaju li se iz prva dva stupca dvije po dvije znamenke iz tablice 6.1., dobit će se ovi brojevi: 08, 01, 37, 70, 50, 29, 13, 11. Oni su ujedno redni brojevi, s pomoću kojih se elementi lako izdvajaju u uzorak. Ako se, na primjer, pojavi skupina 89, nju izostavljamo jer redni broj 89 u sklopu izbora ne postoji. Provodi li se izbor uzorka bez ponavljanja, od više skupina s *jednakim* znamenkama uzimamo samo jednu. Umjesto uporabe tablica, uobičajeno se primjenjuje program za računalo kojim se dobivaju slučajni brojevi te se prema njima u posebnu listu izdvajaju članovi skupa u uzorak.

Katkad se primjenjuje *sistematski izbor jedinica u uzorak*. Ako skup ima 500 000 članova, a planira se uzorak veličine 1 000, izabrat ćemo svaki 500-ti. Koji će biti prvi član u uzorku? Budući da je riječ o slučajnom uzorku, nužno je odrediti *slučajni početak*. Stoga se izračunava korak[2] izbora: $N/n$, uoči broj znamenaka u njemu i slučajno izabere skupina s jednakim brojem znamenaka. Ta skupina, pročitana kao broj, označava redni broj prvog člana, a ostali se lako određuju prema prvome. U primjeru je korak izbora: $N/n = 500000/1000, N/n = 500$ pa se izabire svaka 500. jedinica. U tablicama se po volji odabere troznamenkasta skupina. Ako je to 085, prvi će biti član skupa s rednim brojem 85, drugi 585, treći 1 085 i tako dalje, sve dok se ne izabere i posljednji od njih 1 000. Sistematski je izbor vrlo jednostavan i brz. Za njegovu je primjenu bitno da položaj elementa u sklopu izbora nije u svezi sa svojstvom koje se istražuje s pomoću uzorka, odnosno da ne postoji periodičnost (trend) njihova poretka.

Model jednostavnoga slučajnog uzorka primjenjuje se kada svojstva jedinica skupa čija su svojstva predmet ispitivanja s pomoću uzorka nisu izrazito promjenjiva. Skup s malom promjenjivošću obilježja jedinica naziva se homogenim.[3] U pojedinim statističkim skupovima postoji znatan stupanj varijabilnosti obilježja jedinica. Zna li se o tome, umjesto modela jednostavnoga slučajnog uzorka primjenjuje se model *stratificiranog*[4] *uzorka*.

> **STRATIFICIRANI UZORAK** nastaje slučajnim izborom elemenata osnovnog skupa iz njegovih homogenih dijelova koji se nazivaju stratumima. Postupku izbora uzorka prethodi razvrstavanje elemenata u stratume, koji predočuju podskupove osnovnog skupa koji se međusobno ne preklapaju. Osnovni skup $N$ elemenata raščlanjuje se na $L$ dijelova ($L \geq 2$) tako da vrijedi: $N = N_1 + N_2 + \cdots + N_h + \cdots + N_L$.

Ispituje li se prosječna veličina izdataka za kozmetiku, skup potrošača čine žene i muškarci. Iz iskustva je poznato da postoje razlike veličine izdataka ovisno o spolu. Stoga je uputno skup potrošača podijeliti na dva stratuma (podskupa): u jednom će se stratumu nalaziti žene, a u drugome muškarci. Unutar svakog stratuma manji je stupanj varijabilnosti izdataka za kozmetiku. Razdioba elemenata osnovnog skupa potrošača u stratume ovdje je prema oblicima varijable spol.

Varijable prema čijim se oblicima provodi stratifikacija nazivaju se *faktorima stratifikacije*. Dobiveni stratumi imaju manji stupanj varijabilnosti u usporedbi s varijabilnošću cijelog skupa. Izbor faktora stratifikacije ovisi o danom slučaju. Primjerice, stratumi mogu biti temeljeni na geografskim jedinicama (županijama, općinama), pogonima tvrtke, starosnim skupinama, prihodnim skupinama…

Postupak stratifikacije ima smisla samo ako se razvrstavanjem jedinica dolazi do podskupova s manjim stupnjem varijabiliteta obilježja jedinica. S pomoću podataka stratificiranog uzorka po pravilu se postižu bolji rezultati procjena parametara i primjene drugih postupaka.

Za uporabu modela jednostavnoga i stratificiranog uzorka nužno je raspolagati okvirom izbora, odnosno brojčano označenim popisom elemenata osnovnog

---

[2] Omjer *N/n* je cijeli ili decimalni broj. Kada je omjer decimalni broj, korak izbora je prvi cijeli broj. Ako je broj elemenata osnovnog skupa velik, zaokruživanje na cijeli broj nema nepoželjnih posljedica na svojstva uzorka.

[3] *Homogen* – istovrstan, u čijem se sastavu jasno ne zapažaju znatnije razlike.

[4] *Stratifikacija,* od lat. *stratum* – sloj, slojevitost. U statistici, raspodjela elemenata osnovnog skupa u homogene podskupove koji se međusobno ne preklapaju. Ti se podskupovi nazivaju *stratumima*. Za njih je svojstveno da sadržavaju elemente s relativno malim stupnjem disperzije ispitivanog obilježja.

skupa. Kadšto se do jedinica izabranih za uzorak dolazi s poteškoćama ili velikim troškom, a kadšto je i veliki trošak sastaviti okvir izbora. Ispituje li se raspoloženje birača s pomoću jednostavnoga slučajnog uzorka, okvir je izbora birački popis. U biračkom popisu birači su navedeni (raspršeni) po ulicama, zatim kućama i stanovima. Da bi anketirao birače izabrane u uzorak, anketar mora utrošiti mnogo vremena u obilasku ulica, kuća i stanova. U jednoj ulici moguće je da se nalazi samo član uzorka. Umjesto opisanog načina izbora, moguće je primijeniti ekonomičniji model uzorka. Takav je model uzorka koji se naziva *uzorkom skupina*.

**PRIMJER 1.**

Podijelimo osnovni skup na skupine s jednakim brojem elemenata. Neka je ukupan broj birača u biračkom popisu 8 000. Birači u popisu označeni su rednim brojevima od 1 do 8 000. Ako uzmemo da svaka skupina broji deset birača, tada će se osnovni skup sastojati od 800 skupina od po deset elemenata. Izabire li se u uzorak 160 birača, valja od 800 skupina izabrati njih 16. Označimo skupine rednim brojem od 0 do 799. S pomoću tablica slučajnih brojeva valja izabrati 16 troznamenkastih brojeva između 000 i 799. Kad bi prva izabrana skupina od po tri znamenke u tablici slučajnih brojeva bila 005, tada bi prva skupina uzorka bila peta. U toj su skupini birači u biračkom popisu s rednim brojevima od 50 do 59. Ako je skupina znamenaka 752, riječ je o 752. skupini. U uzorak će biti izabrani birači u popisu s rednim brojevima od 7 520 do 7 529. Prednost izbora skupina očita je jer su birači u skupini prostorno bliži jedni drugima, što ubrzava postupak izbora jedinica u uzorak i smanjuje troškove.

Navedeni primjer odnosi se na uzorak jednakih skupina. U praksi skupine često nisu jednake veličine, to jest ne sadržavaju jednak broj elemenata. Ako se kao skupina uzme razred učenika ili stambena zgrada, broj jedinica u skupinama različit je. U takvu slučaju valja prilagoditi postupak izbora elemenata u uzorak, to jest primijeniti model uzorka nejednakih skupina.

## *PITANJA I ZADATCI ZA VJEŽBU*

**1.** Koji se uzorak naziva jednostavnim slučajnim uzorkom? Kada se u praksi primjenjuje?

**2.** Kako se provodi izbor jedinica u slučajni uzorak? Što su tablice slučajnih brojeva? Koja svojstva ima tablica slučajnih brojeva s pomoću kojih se izabire jednostavni slučajni uzorak? Ako je popis jedinica osnovnog skupa dan u obliku datoteke, na koji će se način izdvojiti jedinice skupa u uzorak?

**3.** Osnovni skup sastoji se od sedam poljoprivrednih kućanstava označenih slovima A, B, C, D, E, F, G. Površina zemljišta kućanstava iznosi u ha: 8, 12, 11, 4, 6, 14, 15. Koliko se različitih uzoraka od po dva člana može izabrati iz navedenoga osnovnog skupa? Za jednostavni slučajni uzorak svojstveno je da svaka jedinica ili svaka kombinacija istog broja jedinica ima jednaku vjerojatnost izbora. Kolika je ta vjerojatnost?

**4.** Što je sistematski izbor jedinica u uzorak? Kada se provodi? Koja svojstva mora imati popis jedinica skupa na temelju kojeg se provodi sistematski izbor? Opišite postupak sistematskog izbora jedinica iz skupa koji broji 20 000, a u uzorak se bira njih 50.

**5.** Kada se u praksi primjenjuje stratificirani uzorak? Na koji se načinu utvrđuju stratumi? Koja svojstva imaju stratumi? Navedite primjere primjene postupka stratifikacije.

**6.** U kojim će se prilikama primijeniti uzorak skupina? Koja su osnovna svojstva te vrste uzorka? Navedite primjere.

## Elementi plana istraživanja s pomoću uzorka

Podatci dobiveni s pomoću uzorka ubrajaju se među primarne podatke. O postupcima prikupljanja primarnih podataka iz konačnih skupova bilo je riječi u prvom poglavlju. Pretežan dio postupaka izloženih u tom dijelu odnosi se i na istraživanje s pomoću uzorka. Dakako da postoje posebnosti koje izviru iz osobitosti primjene metode uzoraka.

Da bi podatci iz uzorka mogli poslužiti kao osnova procjenjivanja parametara i ispitivanja pretpostavki o njima, nužno je da su dobiveni prema jasnome i preciznom planu. Plan uzorka ovisi o više čimbenika, a sastavljaju ga stručnjaci statističari i istraživači. Plan sadržava, primjerice, ciljeve istraživanja, određivanje statističkih skupova i okvira izbora, podatke koje treba prikupiti, model uzorka (nacrt, dizajn), raspoloživa sredstva i postupke provođenja prikupljanja podataka.

Vrlo je važno precizno ustanoviti što je *svrha istraživanja* s pomoću uzorka. U svezi sa svrhom provodi se izbor varijabli i njihovih oblika, odnosno pitanja u anketnim listovima. Kadšto se postavlja velik broj pitanja, od kojih dio nije bitan za postavljeni predmet istraživanja. To može loše utjecati na kakvoću dobivenih rezultata. Jasnim određenjem svrhe istraživanja izbjegavaju se moguće manjkavosti podataka i nepotrebni troškovi.

U planu je nužno odrediti skupove iz kojih se izabire uzorak. Potrebno je utvrditi što je jedinica skupa, što je jedinica izbora s pomoću koje se dolazi do podataka o jedinicama osnovnog skupa u uzorku (ako nije riječ o istim jedinicama), koliki mu je opseg. *Skupove (populacije)* iz kojih se izabire uzorak valja *definirati* pojmovno, prostorno i vremenski. O tim definicijama bilo je riječi u prvom poglavlju.

Definiranje nekih skupova iz kojih se bira uzorak jednostavno je, ali ima ih velik broj za koje to nije. Procjenjujemo li s pomoću uzorka proporciju neispravnih proizvoda isporučenih u jednoj pošiljci od 100 000 komada, statistički skup čine isporučeni (istovrsni) proizvodi prema proizvodnoj specifikaciji. Ispitujemo li s pomoću uzorka osobitosti nezaposlenih, mora se jasno utvrditi koja se osoba smatra nezaposlenom. Analiziraju li se uzorkom poljoprivredna kućanstva, nedvojbeno se mora znati koje se kućanstvo smatra poljoprivrednim, što nije uvijek tako. U pripremi provođenja prikupljanja podataka valja utvrditi pravila kojima se donosi odluka za sve dvojbene slučajeve, to jest s pomoću kojih se odlučuje pripada li neka jedinica skupu ili mu ne pripada.

Za izbor slučajnog uzorka iz konačnog skupa nužno je raspolagati *okvirom izbora*. Već je rečeno da okvir izbora čini popis elemenata koji se izabiru u uzorak.

> **OKVIR IZBORA** sastoji se od popisa jedinica osnovnog skupa iz kojeg se izabire uzorak ili jedinica izbora uzorka koje obuhvaćaju više elemenata osnovnog skupa.

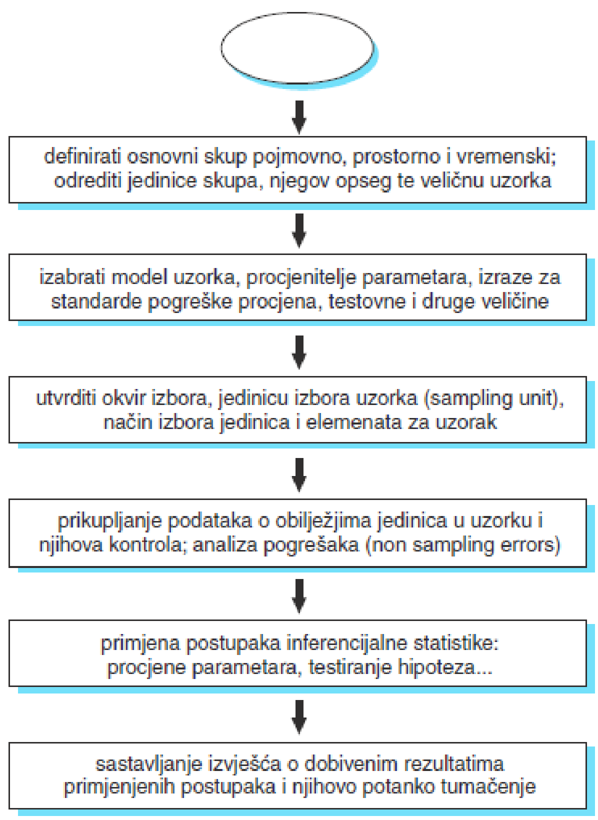
Najjednostavniji okvir izbora sastoji se od popisa elemenata osnovnog skupa. Ispituju li se s pomoću slučajnog uzorka stajališta učenika velikoga srednjoškolskog centra u jednom gradu, riječ je o statističkom skupu u kojem je jedinica jedan učenik upisan dane školske godine. Ukupan broj učenika predočuje opseg skupa. Okvir izbora jest popis svih učenika označenih brojevima od 1 do *N*. U uzorak će se birati na slučajni način učenici pa je jedinica osnovnog skupa jednaka jedinici izbora. Učenik izabran u uzorak pronaći će se prema brojčanoj oznaci u popisu. Takav je okvir izbora platna lista poduzeća. I telefonski imenik može poslužiti kao okvir izbora. Po pravilu telefonski imenik ne obuhvaća sve jedinice skupa jer ima kućanstava bez telefona i kućanstava koja ne dopuštaju unošenje broja u imenik. Nedostatak mu je i to što ima kućanstava s više telefonskih brojeva.

Okvir izbora slučajnog uzorka dionica na velikim burzama su izvješća na financijskim stranicama publikacija ili internetskim stranicama u kojima su navedene sve dionice koje kotiraju na burzi. Okvir izbora su i registri poslovnih subjekata koje vode državni zavodi za statistiku, zemljovidi, knjigovodstveni konti.

Izbor jedinica osnovnog skupa u uzorak često se provodi na bazi *jedinica izbora uzorka* koji sadržava više jedinica osnovnog skupa. Prije izbora uzorka osnovni se skup podijeli u jedinice izbora, i to tako da se svaka jedinica osnovnog skupa nalazi samo u jednoj jedinici izbora. Okvir izbora u takvu slučaju čini popis jedinica izbora uzorka. U navedenim primjerima jedinice osnovnog skupa i jedinice izbora jednake su. Istražujemo li s pomoću uzorka stajališta punoljetnog pučanstva, jedinica izbora može biti osoba, ali i kućanstvo ili stambena zgrada. U kućanstvu ili stambenoj zgradi više je osoba, odnosno jedinica statističkog skupa. Okvir izbora sastoji se od adresa kućanstava ili zemljovida s naznačenim položajem zgrada (blokova zgrada).

Sastavljanje okvira izbora vrlo je zahtjevan posao o kojem u velikoj mjeri ovisi postupak izbora uzorka, a i kakvoća dobivenih podataka. Po pravilu, okvir izbora trebao bi sadržavati sve jedinice statističkog skupa, odnosno sve jedinice izbora uzorka. U okviru ne bi smjelo biti ponavljanja, to jest višestrukog navođenja istog elementa. Velik dio vremena i napora u provođenju nekih istraživanja s pomoću uzorka utroši se na konstrukciju okvira izbora.

Među najvažnijim zadatcima u sklopu planiranja uzorka jest *izbor vrste i nacrta* (*dizajna*) uzorka. Taj izbor ovisi o više čimbenika, među kojima su troškovi, osobitosti osnovnih skupova i dr. Može se primijeniti nacrt jednostavnoga slučajnog uzorka, stratificiranog, uzorka skupina ili kojega drugog tipa. Za svaki izbor precizno se utvrđuje način izbora elemenata u uzorak. Utvrđuju se i izrazi za statističko-analitičke veličine iz uzorka, među kojima su i izrazi za veličine pogrešaka zbog primjene uzorka. Sa stajališta statističke metode glavni su koraci prikazani na sl. 6.2.




**Slika 2.** *Glavni koraci istraživanja s pomoću uzorka*

## PITANJA I ZADATCI

**1.** Navedite osnovne elemente plana uzorka.

**2.** Definirajte pojam okvira izbora. Navedite primjere.

**3.** Analizirajte ovaj navod : „Okvir izbora sredstvo je pristupa osnovnom skupu... Okvir izbora sastoji se od jedinica izbora uzorka... Jedinica izbora uzorka može se sastojati od stambenih jedinica, osoba, trgovačkih društava, farmi, poslovnih udruga, industrijskih ili poljoprivrednih proizvoda, knjigovodstvenih zapisa ili od drugih stvari... Bez okvira izbora nije moguće iscrpno promatranje ni izbor slučajnog uzorka... (E. W. Deming).

**4.** Razmotrite sadržaj sljedećeg navoda: „Jednom kada je donesena odluka o istraživanju s pomoću uzorka, kada je jasna svrha i okvir izbora, tada se sam plan istraživanja uzorkom sastoji od pet odrednica. To su: (1) okvir izbora, (2) postupak izbora uzorka jedinica izbora iz okvira izbora (sa stratifikacijom ili bez nje, ali uvijek s pomoću slučajnih brojeva), (3) formula s pomoću koje će se izračunati procjena na temelju rezultata iz uzorka... (4) formula za izračunavanje standardnih pogrešaka svake procjene... (5) postupak revizije i kontrole za procjenu neuzoračkih pogrešaka i njihovih efekata na konačne vrijednosti procjena." (Prema W. E. Demingu.)

**5.** Srednja škola broji 1 536 učenika, čiji je raspored po razredima prikazan u tablici.

| I. razredi | II. razredi | III. razredi | IV. razredi |
|---|---|---|---|
| 451 | 395 | 362 | 328 |

Detaljno opišite na koji biste način izabrali kvotni uzorak veličine 96 učenika. (b) Na temelju kojeg će se okvira provesti izbor slučajnog uzorka iste veličine? Na koji biste način izabrali (1) jednostavni slučajni uzorak, (2) stratificirani uzorak? (c) Kako biste proveli sistematski izbor uzorka?

**6.** Odredite sve elemente za procjenu proporcije neispravnih proizvoda iz jedne pošiljke od 50 000 komada proizvoda s pomoću slučajnog uzorka veličine 500 komada. Proizvodi su pakirani u kutije od po 100 komada.

# Procjena aritmetičke sredine i proporcije osnovnog skupa

Procjenjivanje parametara jedna je od glavnih zadaća metode uzoraka. Riječ je o procjeni aritmetičke sredine, standardne devijacije, proporcije, parametara u regresijskome modelu, koeficijenta korelacije i drugih statističko-analitičkih veličina. Od mnogobrojnih postupaka procjene izdvojit će se postupci procjene aritmetičke sredine i proporcije osnovnog skupa s pomoću jednostavnog slučajnog uzorka kako bismo se ponajprije upoznali s temeljnim obilježjima dobivenih procjena i načinom njihova tumačenja. Shvatimo li te postupke, jednostavnije ćemo prosuđivati i rezultate procjenjivanja drugih parametara.

## Procjena aritmetičke sredine osnovnog skupa jednostavnim slučajnim uzorkom

Želimo li procijeniti prosječnu mjesečnu veličinu izdataka za prehranu kućanstava na području jednoga grada s pomoću podataka prikupljenih za dio kućanstava, primijenit ćemo postupke u sklopu metode uzoraka. Riječ je o procjeni parametra (aritmetičke sredine osnovnog skupa). Kako smo već istaknuli, procjena parametra može biti brojem i intervalom, a izrazi s pomoću kojih se procjenjuje na temelju podataka iz uzorka nazivaju se procjeniteljima.

**PROCJENA ARITMETIČKE SREDINE OSNOVNOG SKUPA.** Procjenjuje li se nepoznata aritmetička sredina osnovnog skupa $\mu$ s pomoću jednostavnog slučajnog uzorka $n$ jedinica, tada je procjenitelj aritmetičke sredine osnovnog skupa jednim brojem aritmetička sredina uzorka, to jest $\hat{\mu} = \bar{x}$, odnosno

(1) $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ .

Kada su podatci iz uzorka grupirani u obliku distribucije frekvencija, umjesto izraza za jednostavnu primijenit ćemo izraz za vaganu aritmetičku sredinu.

Uvrstimo li u izraz (1) vrijednosti iz uzorka, dobit ćemo broj koji predočuje procjenu aritmetičke sredine osnovnog skupa. Kako je riječ o veličini dobivenoj s pomoću dijela (uzorka), a ne svih podataka, ta veličina sadržava pogrešku. Veličina pogreške očito ovisi o veličini uzorka. Što je veći uzorak, manja je pogreška procjene zbog primjene uzorka. Ona je jednaka nuli kada je veličina „uzorka” jednaka veličini osnovnog skupa, to jest kada je $n = N$.

Pogreška procjene ovisi i o stupnju varijabilnosti obilježja u osnovnom skupu. Ako zamislimo osnovni skup u kojemu su svi proizvodi jednake težine, za zaključivanje o skupu bio bi dovoljno izabrati u uzorak samo jedan proizvod. Za danu veličinu uzorka pogreška procjene veća je što je veći stupanj varijabilnosti obilježja (što je veća disperzija). Stupanj raspršenosti obilježja mjeri se standardnom devijacijom pa se može zaključiti da pogreška procjene ovisi i o standardnoj devijaciji obilježja osnovnog skupa.

Pogreška procjene aritmetičke sredine smanjuje se s povećanjem veličine uzorka n, a povećava s povećanjem standardne devijacije osnovnog skupa σ.

**POGREŠKA PROCJENE ARITMETIČKE SREDINE** osnovnog skupa s pomoću jednostavnoga slučajnog uzorka veličine *n* izabranoga iz konačnog skupa veličine *N* elemenata dana je ovim izrazom:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \quad (2)$$

Ako standardna devijacija osnovnog skupa σ nije poznata, nego njezina procjena uzorkom $\hat{\sigma}$, izraz je: $\sigma_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}}$.

Izraz (2) još se naziva standardnom pogreškom procjene aritmetičke sredine, a faktor $\sqrt{\frac{N-n}{N-1}}$ faktorom korekcije za konačne skupove. Ako je uzorak izabran iz beskonačnoga osnovnog skupa ili je izbor iz konačnog skupa s ponavljanjem taj je faktor jednak jedan, a izraz (2) postaje: (2 a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ ili (2 b) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Da bismo objasnili značenje pogreške procjene, potrebno je imati na umu da se iz konačnoga osnovnog skupa veličine *N* elemenata može izabrati bez ponavljanja onoliko različitih uzoraka veličine *n* koliko je kombinacija. Za svaki uzorak može se izračunati aritmetička sredina. I njih je koliko i uzoraka. Te sredine razlikuju se od sredine osnovnog skupa. Stupanj varijabilnosti sredina uzoraka prema sredini skupa mjeri se standardnom devijacijom, koja nije ništa drugo nego standardna pogreška procjene aritmetičke sredine osnovnog skupa. Iz navedenoga slijedi i tumačenje pogreške procjene.

**STANDARDNA POGREŠKA PROCJENE ARITMETIČKE SREDINE** jest prosječno odstupanje sredina uzoraka od aritmetičke sredine osnovnog skupa.

## PRIMJER 2.

Pretpostavimo da se osnovni skup sastoji od četiri učenika: A, B, C, D. Učenik A gleda televiziju jedan sat na dan, učenik B četiri sata, učenik C tri sata, a učenik D dva sata. Prosječno vrijeme gledanja televizijskih programa jest aritmetička sredina tog skupa i ona je:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{4}(1+4+3+2), \mu = 2.5 .$$

Ta se veličina naziva parametrom. Za navedene podatke izračunana je i standardna devijacija osnovnog skupa $\sigma = 1.11803$. I ta je veličina parametar jer je izračunana na temelju podataka za svaku jedinicu osnovnog skupa.

Iz navedenog se skupa mogu izabrati različiti uzorci učenika veličine $n = 2$. Broj tih uzoraka je 6 i jednak je broju kombinacija, to jest:

$$k = \frac{N!}{n!(N-n)!} = \frac{4!}{2!(4-2)!} = \frac{4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 2\cdot 1}, k = 6 .$$

U tablici 2. navedeni su uzorci, vrijednosti varijable učenika u uzorku te aritmetičke sredine uzoraka (prosječno vrijeme gledanja programa učenika u uzorku).

**Tablica 2.** Sredine uzoraka i razlike sredina uzoraka i sredine osnovnog skupa

| Učenici u uzorku | Sati gledanja programa | Sredine uzoraka | Razlike sredina uzoraka i sredine osnovnog skupa | Kvadrati razlika |
|---|---|---|---|---|
| | | $\overline{x}_i$ | $(\overline{x}_i - \mu)$ | $(\overline{x}_i - \mu)^2$ |
| 1 | 2 | 3 | 4 | 5 |
| A, B | 1, 4 | $\overline{x}_1 = \frac{1+4}{2} = 2.5$ | 2.5 – 2.5 = 0 | 0 |
| A, C | 1, 3 | $\overline{x}_2 = \frac{1+3}{2} = 2.0$ | 2.0 – 2.5 = -0.5 | 0.25 |
| A, D | 1, 2 | $\overline{x}_3 = \frac{1+2}{2} = 1.5$ | 1.5 – 2.5 = -1.0 | 1 |
| B, C | 4, 3 | $\overline{x}_4 = \frac{4+3}{2} = 3.5$ | 3.5 – 2.5 = 1 | 1 |
| B, D | 4, 2 | $\overline{x}_5 = \frac{4+2}{2} = 3.0$ | 3.0 – 2.5 = 0.5 | 0.25 |
| C, D | 3, 2 | $\overline{x}_6 = \frac{3+2}{2} = 2.5$ | 2.5 – 2.5 = 0 | 0 |
| – | – | Zbroj | 0 | 2.5 |

Sredine uzoraka u trećem stupcu nisu jednake jer su računane za različite uzorke. Aritmetička sredina aritmetičkih sredina uzoraka jest:

$$\overline{\overline{x}} = \frac{\overline{x}_1 + \overline{x}_2 + \cdots + \overline{x}_6}{6} = \frac{2.5 + 2.0 + 1.5 + 3.5 + 3.0 + 2.5}{6}, \overline{\overline{x}} = 2.5, \overline{\overline{x}} = \mu .$$

Aritmetička sredina aritmetičkih sredina svih uzoraka jednaka je 2.5 sati, kolika je i aritmetička sredina osnovnog skupa. Aritmetičke sredine uzoraka odstupaju od aritmetičke sredine osnovnog skupa. Prosječno odstupanje sredina uzoraka od sredine osnovnog skupa predočeno je standardnom devijacijom, koja se, kako smo istaknuli, naziva standardnom pogreškom sredine. Ona je:

$$\sigma_{\overline{x}} = \sqrt{\frac{\sum_{i=1}^{6}(\overline{x}_i - \mu)^2}{6}} = \sqrt{\frac{(2.5-2.5)^2 + (2.0-2.5)^2 + \cdots + (2.5-2.5)^2}{6}}, \sigma_{\overline{x}} = 0.64549 .$$

Do istog se rezultata dolazi primjenom formule (6.2):

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.11803}{\sqrt{2}}\sqrt{\frac{4-2}{4-1}}, \sigma_{\bar{x}} = 0.64549\ .$$

Standardna pogreška govori nam da je prosječno odstupanje aritmetičkih sredina uzoraka od aritmetičke sredine osnovnog skupa 0.64549 sati. Standardnu pogrešku sredine moramo razlikovati od standardne devijacije osnovnog skupa.

Standardna devijacija osnovnog skupa govori nam o raspršenosti vrijednosti varijable prema aritmetičkoj sredini osnovnog skupa, a *standardna pogreška upućuje na raspršenost sredina uzoraka prema aritmetičkoj sredini osnovnog skupa.* S povećanjem veličine uzorka standardna pogreška približava se nuli.

U postupcima procjenjivanja (i testiranja hipoteza) važnu ulogu ima distribucija vrijednosti procjenitelja parametra (*sampling*-distribucija). Tako je riječ, primjerice, o distribuciji aritmetičkih sredina uzoraka. U primjeru se upućuje na pojam i svojstva te distribucije polazeći od podataka prethodnog primjera.

## PRIMJER 3.

U primjeru 2. navedene su vrijednosti aritmetičkih sredina svih uzoraka za navedeni osnovni skup.

Vrijednosti sredina uzoraka uredit ćemo u obliku distribucije[5] i prikazati tablicom.

**Tablica 3.** Distribucija sredina uzoraka

| Sredine uzoraka, $\bar{x}_i$ | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
|---|---|---|---|---|---|
| Relativna frekvencija, $p_i$ | 1/6 | 1/6 | 2/6 | 1/6 | 1/6 |

Aritmetička sredina distribucije sredina uzoraka jednaka je aritmetičkoj sredini osnovnog skupa, a standardna devijacija jednaka je standardnoj pogrešci sredine.

Što se može reći o obliku distribucije? Uzmemo li da se osnovni skup sastoji od 50 učenika i da se formiraju uzorci veličine pet učenika, broj različitih uzoraka jednak je broju kombinacija i iznosi 2 118 760. Izračunamo li sredinu svakog uzorka, dobit ćemo 2 118 760 sredina. Osnovni skupovi u praksi broje vrlo velik broj jedinica pa je broj kombinacija ili mogućih uzoraka golem. Golem je i broj sredina uzoraka. Grupiramo li vrijednosti sredina uzoraka, doći ćemo do distribucije čiji je oblik sličan obliku normalne distribucije ili joj je jednak. U svezi s tim postoji više teorijskih rezultata. Među njima[6] je i sljedeći.

[5] Distribucija sredina svih mogućih uzoraka naziva se *sampling-distribucijom sredina;* engl. *sample* – uzorak.

[6] Pri procjeni aritmetičke sredine velikim se uzorkom smatra uzorak s više od 30 jedinica. Uzorak koji broji 30 i manje jedinica naziva se malim uzorkom. Za mali uzorak vrijedi posebno pravilo procjenjivanja sredine.

**OBLIK I SVOJSTVA DISTRIBUCIJE VRIJEDNOSTI PROCJENITELJA SREDINE OSNOVNOG SKUPA (SAMPLING-DISTRIBUCIJE).** Ako iz bilo kojega osnovnog skupa sa sredinom ∝ i standardnom devijacijom σ izaberemo slučajni uzorak veći od 30 jedinica, distribucija sredina svih mogućih uzoraka približno je oblika normalne distribucije sa sredinom koja je jednaka sredini osnovnog skupa μ i standardnom devijacijom koja je jednaka standardnoj pogrešci sredina $\sigma_{\bar{x}}$.

Ako se uzme u obzir da se pri procjeni primjenjuje model jednostavnoga slučajnog uzorka, tada normalna distribucija ima svojstva distribucije vjerojatnosti. Koje su posljedice navedenog svojstva distribucije sredina? Budući da su svojstva normalne distribucije poznata, možemo zaključivati o mogućem odstupanju sredina uzorka od aritmetičke sredine osnovnog skupa. Time se čini važan korak k određenju intervalnog procjenitelja aritmetičke sredine osnovnog skupa na temelju *jednog uzorka*. Izaberemo li jedan uzorak, vjerojatnost je 0.68 da će sredina tog uzorka odstupati od sredine osnovnog skupa za $\pm\sigma_{\bar{x}}$ , odnosno za ± jednu standardnu pogrešku, to jest da će se naći u intervalu:

$$\mu - \sigma_{\bar{x}} < \bar{x} < \mu + \sigma_{\bar{x}} .$$

Aritmetička sredina osnovnog skupa ∝ nepoznata je, a navedeni interval[7] procjene uz vjerojatnost 0.68 jednak je ovomu: $\bar{x} - \sigma_{\bar{x}} < \mu < \bar{x} + \sigma_{\bar{x}}$ .

**PROCJENA ARITMETIČKE SREDINE INTERVALOM.** Procjena aritmetičke sredine osnovnog skupa brojem je aritmetička sredina uzorka (6.1), a procjena intervalom dana je izrazom:

$$\bar{x} - z\sigma_{\bar{x}} < \mu < \bar{x} + z\sigma_{\bar{x}} . \qquad (3)$$

Kako je vidljivo, interval ima svoju donju i gornju granicu i simetričan je. Vjerojatnost s kojom se procjenjuje naziva se *razinom povjerenja procjene*, a sam interval *intervalom povjerenja procjene aritmetičke sredine osnovnog skupa*. Faktor[8] $z$ uz standardnu pogrešku ovisi o razini povjerenja procjene i zove se koeficijentom povjerenja (pouzdanosti) procjene. Za različite razine sredina osnovnog skupa procjenjuje se sljedećim izrazima.

| Intervalni procjenitelj | Vjerojatnost da će interval sadržavati aritmetičku sredinu osnovnog skupa |
|---|---|
| $\bar{x} - \sigma_{\bar{x}} < \mu < \bar{x} + \sigma_{\bar{x}}$ | 0.68 |
| $\bar{x} - 1.65\sigma_{\bar{x}} < \mu < \bar{x} + 1.65\,\sigma_{\bar{x}}$ | 0.90 |
| $\bar{x} - 2\sigma_{\bar{x}} < \mu < \bar{x} + 2\sigma_{\bar{x}}$ | 0.95 |
| $\bar{x} - 2.58\,\sigma_{\bar{x}} < \mu < \bar{x} + 2.58\,\sigma_{\bar{x}}$ | 0.99 |

[7] Interval $\mu - \sigma_{\bar{x}} < \bar{x} < \mu + \sigma_{\bar{x}}$ jest dvostruka nejednakost. Ako od svakog člana oduzmemo ∝, dolazimo do izraza $-\sigma_{\bar{x}} < \bar{x} - \mu < +\sigma_{\bar{x}}$ , koji množenjem s (-1) postaje $\sigma_{\bar{x}} > \mu - \bar{x} > -\sigma_{\bar{x}}$ , a uređenjem $\bar{x} - \sigma_{\bar{x}} < \mu < \bar{x} + \sigma_{\bar{x}}$ .

[8] Faktor se određuje s pomoću tablica površina normalne distribucije, odnosno računalnim programom. Često se daje s većim brojem decimalnih mjesta.

Izrazom $\bar{x}-2\sigma_{\bar{x}}<\mu<\bar{x}+2\sigma_{\bar{x}}$ procjenjujemo aritmetičku sredinu osnovnog skupa s pomoću jednostavnoga slučajnog uzorka s 95 % vjerojatnosti. Drukčije rečeno, kad bismo izabrali sve moguće uzorke i formirali sve moguće intervale procjene, tada bi u njih 95 % bila obuhvaćena nepoznata aritmetička sredina osnovnog skupa. Postupak procjene aritmetičke sredine osnovnog skupa s pomoću jednostavnog slučajnog uzorka ilustrirat ćemo primjerom.

### PRIMJER 4.

Tijekom sezone preko turističke agencije Sun u hotelu Jadran boravile su na sedmodnevnom odmoru skupine s istoga turističkog tržišta. Ukupno je u hotelu bio 3 201 turist. S pomoću uzorka želi se procijeniti brojem i intervalom prosječna izvanpansionska potrošnja turista u objektima hotela (teniska igrališta, restoran, bar i drugi objekti), i to na razini povjerenja 95 %. Sva su plaćanja u sklopu hotela bezgotovinska („na broj sobe”). Radi procjene izabran je 1 % uzorak turista iz datoteke turista, odnosno njih 32. Uvidom u račune turista izabranih u uzorak dobiveni su sljedeći podatci o izvanpansionskoj potrošnji tijekom sedmodnevnog boravka objektima hotela u američkim dolarima.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 344 | 319 | 209 | 337 | 269 | 295 | 329 | 402 |
| 282 | 376 | 294 | 356 | 257 | 190 | 357 | 263 |
| 230 | 294 | 380 | 221 | 232 | 244 | 302 | 322 |
| 273 | 281 | 335 | 338 | 248 | 368 | 317 | 279 |

Postupak procjene provodi se u ovim koracima: (1) izračunati aritmetičku sredinu uzorka. Aritmetička sredina uzorka jest procjena brojem aritmetičke sredine osnovnog skupa, to jest prosječne izvanpansionske potrošnje *svih* turista. (2) Izračunati standardnu pogrešku procjene sredine. (3) Odrediti koeficijent 95 % pouzdanosti procjene. (4) Odrediti donju i gornju granicu procjene.

(1) Procjenitelj aritmetičke sredine osnovnog skupa brojem jest aritmetička sredina uzorka veličine $n = 32$. Uzorak je veći od 30 jedinica i smatra se velikim uzorkom. Sredina uzorka je:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{344+319+209+\cdots+279}{32},\ \bar{x}=298.21875\,.$$

Prosječna izvanpansionska potrošnja turista u objektima hotela iznosila je (zaokruženo) 298 američkih dolara. To je ujedno procjena prosječne potrošnje *svih* turista. O pogrešci zbog primjene uzorka i preciznosti procjene zaključuje se na temelju standardne pogreške i intervala procjene.

(2) Standardna pogreška procjene aritmetičke sredine osnovnog skupa (prosječne potrošnje *svih* turista) jest prema izrazu (2):

$$\sigma_{\bar{x}}=\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}},\quad \hat{\sigma}=\sqrt{\frac{\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2}{n-1}},\ n=32,\ N=3201\,.$$

U navedenom izrazu $\hat{\sigma}$ je procjenitelj standardne devijacije osnovnog skupa jer ona nije poznata. Procjena standardne devijacije osnovnog skupa jest:

$$\hat{\sigma}=\sqrt{\frac{344^2+319^2+\cdots+279^2-\frac{(344+319+\cdots+279)^2}{32}}{32-1}},\ \hat{\sigma}=53.74896,$$

a standardna pogreška procjene aritmetičke sredine:

$$\sigma_{\bar{x}}=\frac{53.74896}{\sqrt{32}}\sqrt{\frac{3201-32}{3201-1}},\ \sigma_{\bar{x}}=9.45543.$$

Slijedi određivanje koeficijenta pouzdanosti.

(3) Sredina osnovnog skupa procjenjuje se na temelju velikog uzorka, i to s vjerojatnošću 0.95, odnosno na 95 % razini povjerenja. Toj razini povjerenja (vjerojatnosti) odgovara koeficijent pouzdanosti $z = 2$.

(4) Intervalni procjenitelj aritmetičke sredine jest: $\bar{x}-2\sigma_{\bar{x}}<\mu<\bar{x}+2\sigma_{\bar{x}}$, a vrijednosti granica su:

$$298.21875-2\cdot 9.45543<\mu<298.21875+2\cdot 9.45543,$$

to jest:

$$279.30789<\mu<317.12961.$$

Vrijednosti granica zaokružit ćemo na cijeli broj pa je donja granica intervala 279, a gornja 317 američkih dolara. Dobiveni interval tumači se ovako: s vjerojatnošću 0.95 prosječna izvanpansionska potrošnja *svih* turista u objektima hotela kretala se između 279 i 317 američkih dolara. Sadržava li navedeni interval doista aritmetičku sredinu osnovnog skupa? Na to pitanje ne može se sigurno odgovoriti jer se zaključak temelji na dijelu podataka, to jest uzorku. Kad bismo iz skupa izabrali sve moguće uzorke iste veličine i odredili isto toliko intervala procjene, u 95 % slučajeva u intervalima bi se nalazila aritmetička sredina osnovnog skupa.

Određivanjem granica moguće je donijeti sud o preciznosti procjene. Procjena je to preciznija što je interval procjene uži. Na preciznost procjene može se utjecati povećanjem veličine uzorka ili, ako za to postoje razlozi, primjenom prikladnijeg nacrta (dizajna) uzorka.

## *PITANJA I ZADATCI*

**1.** Kako glasi procjenitelj aritmetičke sredine osnovnog skupa brojem s pomoću jednostavnoga slučajnog uzorka? Na koji se način dolazi do procjene?

**2.** Što se mjeri standardnom pogreškom procjene aritmetičke sredine? Koja je razlika između standardne devijacije osnovnog skupa, standardne devijacije uzorka i standardne pogreške procjene aritmetičke sredine? Kako se tumači standardna pogreška?

**3.** O kojim veličinama ovisi veličina standardne pogreške procjene sredine konačnoga osnovnog skupa jednostavnim slučajnim uzorkom? Kolika je standardna pogreška ako je: (a) $n = N$, (b) $\sigma = 0$? Objasnite odgovor.

**4.** Navedite sve elemente za izračunavanje standardne pogreške procjene sredine iz konačnog skupa s pomoću jednostavnoga slučajnog uzorka. Kako ćete postupiti ako nije poznata standardna devijacija osnovnog skupa?

**5.** Osnovni skup sastoji se od sedam poljoprivrednih kućanstava označenih slovima A, B, C, D, E, F, G. Površina zemljišta kućanstava iznosi u ha: 8, 12, 11, 4, 6, 14, 15. (a) Kolika je prosječna površina kućanstava? Koliko je prosječno odstupanje od prosjeka? (b) Izaberite (bez ponavljanja) sve moguće uzorke veličine $n = 2$ Za svaki uzorak izračunajte: aritmetičku sredinu i aritmetičku sredinu aritmetičkih sredina. Što zaključujete? (c) Koliko je prosječno odstupanje aritmetičkih sredina uzoraka od aritmetičke sredine osnovnog skupa? Kako se naziva ta standardna devijacija (d) Uredite aritmetičke sredine uzoraka u obliku distribucije frekvencija (distribucije vjerojatnosti).

**6.** Navedite opći izraz za intervalno procjenjivanje aritmetičke sredine osnovnog skupa s pomoću velikog uzorka i modela jednostavnoga slučajnog uzorka. Što je: (a) razina povjerenja(pouzdanosti), (b) koeficijent povjerenja (pouzdanosti), (c) interval povjerenja?

**7.** Kako se tumači interval procjene aritmetičke sredine osnovnog skupa? Može li se tvrditi sa "sigurnošću" da se u *jednom* intervalu procjene *doista* nalazi nepoznata aritmetička sredina osnovnog skupa?

**8**. Izračunajte vrijednost standardne pogreške procjene aritmetičke sredine osnovnog skupa ako su dane ove veličine:
(1) $N = 126542$, $n = 95$, $\sigma = 25$, (2) $N = 25832$, $n = 54$, $\hat{\sigma} = 17.2$.

**9.** Aritmetička sredina jednostavnoga slučajnog uzorka 225 jedinica izabranoga iz skupa od 38 795 jedinica iznosi 300. Standardna devijacija osnovnog skupa je 75. Odredite granice 95 % intervala povjerenja procjene aritmetičke sredine osnovnog skupa.

**10.** Tvrtka Metal isporučuje vijke tipa AM2 u pakiranjima od po stotinu komada. Kupac je primio pošiljku od 10 000 pakiranja. Radi kontrole kakvoće izabran je slučajni uzorak 60 pakiranja. Kontrolom je ustanovljen sljedeći broj neispravnih vijaka po pakiranju.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 3 | 3 | 1 | 3 | 2 | 0 | 0 | 2 | 1 | 1 | 2 |
| 3 | 4 | 3 | 1 | 2 | 2 | 2 | 3 | 0 | 1 | 1 | 3 |
| 3 | 3 | 3 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 3 | 1 |
| 1 | 2 | 2 | 1 | 1 | 3 | 2 | 3 | 3 | 3 | 0 | 1 |
| 2 | 2 | 2 | 2 | 4 | 0 | 3 | 2 | 1 | 2 | 1 | 1 |

Procijenite brojem i intervalom prosječan broj neispravnih vijaka po pakiranju za *cijelu* pošiljku. Razina povjerenja procjene 95 %. $\sum_{i=1}^{60} x_i = 117, \sum_{i=1}^{60} x_i^2 = 293$.

**11.** U uzorku 40 trgovina na malo živežnim namirnicama zabilježena je cijena beskofeinske kave marke GOLD od 250 g. Prosječna cijena kave u trgovinama u uzorku iznosila je 60 kn. U kojim se granicama može očekivati da će se naći cijena kave *u trgovini na malo*? Pouzdanost procjene 95 %. $\hat{\sigma} = 2.5$ kn. Pri računanju odgovarajuće veličine zanemarite faktor korekcije za konačne skupove.

**12.** Protumačite sljedeći djelomični ispis obradbe s pomoću programske potpore.
Mean (*sredina*) 6806.435 Std Err (*standardna pogreška*) 144.6042
95 % CI for Mean (95 % *interval povjerenja za sredinu*) (6522.289; 7090.581)

**13.** prikazan je sljedeći dio ispisa rezultata obradbe programskom potporom Excel *(Data Analysis, Summary Statistics)*.

| | |
|---|---|
| Mean (*sredina*) | 1.95 |
| Standard Error (*standardna pogreška*) | 0.135348473 |
| Standard Deviation (*standardna devijacija* ) | 1.048404761 |
| Skewness (*koeficijent asimetrije*) | -0.08010925 |
| Count (*broj podataka, veličina uzorka*) | 60 |
| Confidence Level(95,0 %) (*razina povjerenja* (95 %)) | 0.270831953 |

U ovom ispisu posljednja vrijednost je umnožak koeficijenta pouzdanosti i standardne pogreške procjene sredine, to jest $z\sigma_{\bar{x}}$ . Odredite vrijednosti granica procjene aritmetičke sredine osnovnog skupa. Protumačite navedene rezultate.

**14.** Kako se određuje veličina jednostavnoga slučajnog uzorka iz konačnog skupa za intervalnu procjenu aritmetičke sredine?
(a) Ako se u procjeni tolerira odstupanje od sredine skupa veličine *d*, a procjenjuje se na razini povjerenja kojoj odgovara koeficijent povjerenje *z* , veličina *d* jednaka je $d = z\sigma_{\bar{x}}$ , odnosno:

$$d = z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{n-1}}\,.$$

U prethodnom izrazu σ je standardna devijacija osnovnog skupa. Ako je nepoznata, njezina se veličina prosuđuje na temelju poznavanja pojave. Veličina uzorka *n* je rješenje navedene jednadžbe, to jest:

$$n = \frac{Nz^2\sigma^2}{d^2(N-1)+z^2\sigma^2}\,.$$

Ako je riječ o uzorku iz beskonačne populacije ili je *n* manji od 5 % od *N*, veličina uzorka određuje se izrazom:

$$n = \left[\frac{z\sigma}{d}\right]^2.$$

(b) S pomoću jednostavnoga slučajnog uzorka procjenjuje se prosječna veličina mjesečnih izdataka za reklamu u trgovini na malo s približno istom strukturom ponude. Koliko će se trgovina izabrati u uzorak od njih 3 689 ako se tolerira odstupanje od nepoznate sredine skupa od 10 tisuća kuna, a procjenjuje na razini povjerenja 95 %? Standardna devijacija osnovnog skupa je 40 tisuća kuna.

## Procjena proporcije osnovnog skupa jednostavnim slučajnim uzorkom

Osim procjene aritmetičke sredine, u praksi se često procjenjuje proporcija. Proporcija je relativni broj koji nastaje diobom broja jedinica s određenim oblikom varijable (obilježja) i ukupnog broja jedinica. Ako su u danom poduzeću od 1 456 zaposlenih njih 732 žene, proporcija žena je omjer 732/1 456 i iznosi 0.5027, odnosno 50.27 % zaposlenih žena. Proporcija neispravnih proizvoda u isporučenoj robi jest omjer broja neispravnih proizvoda i ukupnog broja proizvoda. Moglo bi se navesti velik broj primjera koji govore o raširenosti tog pokazatelja strukture podataka.

Proporcija osnovnog skupa procjenjuje se s pomoću podataka iz uzorka na način sličan procjeni aritmetičke sredine. Pretpostavit ćemo da se procjenjuje s pomoću jednostavnog slučajnog uzorka. Procjena je brojem i intervalom.

**NEPOZNATA PROPORCIJA** jedinica osnovnog skupa koje imaju određeni oblik varijable (obilježja) definira se izrazom:

(4) $$p = \frac{M}{N},$$

a proporcija jedinica koje nemaju dani oblik obilježja jest:

(5) $$q = \frac{N - M}{N} = 1 - \frac{M}{N}, \; q = 1 - p, \; p + q = 1.$$

U navedenim izrazima $M$ je broj jedinica osnovnog skupa s određenim oblikom obilježja, a $N$ je broj elemenata osnovnog skupa.

**PROCJENITELJ PROPORCIJE OSNOVNOG SKUPA BROJEM.** Izabere li se iz osnovnog skupa uzorak veličine $n$ elemenata, od kojih $m$ elemenata ima određeni oblik varijable (obilježja), procjenitelj proporcije osnovnog skupa brojem jest:

(6) $$\hat{p} = \frac{m}{n}.$$

**INTERVALNI PROCJENITELJ PROPORCIJE.** Ako je uzorak slučajan i dovoljno velik, intervalni je procjenitelj proporcije osnovnog skupa dan izrazom:

(7) $$\hat{p} - z\sigma_{\hat{p}} < p < \hat{p} + z\sigma_{\hat{p}}.$$

Interval ima donju i gornju granicu. Vrijednosti granica ovise o procjeni proporcije $\hat{p}$, koeficijentu povjerenja procjene $z$ te standardnoj pogrešci procjene proporcije.

Standardna pogreška procjene proporcije tumači se na sličan način kao i standardna pogreška procjene aritmetičke sredine. Kad bismo iz konačnog skupa

[9] Navedeni intervalni procjenitelj može se rabiti samo ako granice procjene dobivene njegovom primjenom ne uključuju vrijednosti 0 i manje te vrijednosti 1 i veće. Nadalje, dovoljno velikim uzorkom smatra se (prema praktičnom pravilu) uzorak za koji vrijedi $np \geq 5, nq \geq 5$.

izabrali bez ponavljanja sve moguće uzorke veličine $n$ i za svaki uzorak izračunali proporciju, uočili bismo da se proporcije uzoraka međusobno razlikuju, a njihova je aritmetička sredina proporcija jednaka proporciji osnovnog skupa.

**STANDARDNOM POGRJEŠKOM PROCJENE** $\sigma_{\hat{p}}$ proporcije mjeri se prosječno odstupanje proporcija uzoraka od proporcije osnovnog skupa. Dana je izrazom:

(8 a)
$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)},$$

odnosno:

(8 b)
$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{N-n}{N-1}\right)}.$$

Uredimo li proporcije svih mogućih uzoraka u distribuciju, uz određene će uvjete ta distribucija biti približno oblika normalne distribucije. Koristeći se svojstvima te distribucije, odredit ćemo interval procjene proporcije s pomoću podataka iz *jednog* uzorka. Za različite razine povjerenja intervalni su procjenitelji sljedeći.

| Intervalni procjenitelj | Vjerojatnost da će interval sadržavati proporciju osnovnog skupa |
|---|---|
| $\hat{p}-\sigma_{\hat{p}} < p < \hat{p}+\sigma_{\hat{p}}$ | 0.68 |
| $\hat{p}-1.65\,\sigma_{\hat{p}} < p < \hat{p}+1.65\,\sigma_{\hat{p}}$ | 0.90 |
| $\hat{p}-2\sigma_{\hat{p}} < p < \hat{p}+2\sigma_{\hat{p}}$ | 0.95 |
| $\hat{p}-2.85\,\sigma_{\hat{p}} < p < \hat{p}+2.85\,\sigma_{\hat{p}}$ | 0.99 |

I tumačenje navedenih intervala slično je onomu za aritmetičku sredinu. Uzmimo, primjerice, intervalni procjenitelj $\hat{p}-3\sigma_{\hat{p}} < p < \hat{p}+3\sigma_{\hat{p}}$. Izračunaju li se vrijednosti granica, tada se može reći da se s vjerojatnošću 0.99, odnosno 99 %, očekuje da će se unutar tih granica nalaziti proporcija osnovnog skupa.

No valja istaknuti da nije poznato sa sigurnošću sadržava li dani interval procjene doista proporciju osnovnog skupa jer se procjena temelji na dijelu podataka, to jest uzorku. Kad bismo iz skupa izabrali sve moguće uzorke iste veličine i odredili isto toliko intervala procjene, u 99 % slučajeva u intervalima bi se nalazila proporcija osnovnog skupa.

Postupak procjene proporcije brojem i intervalom provodi se u istim koracima kao i postupak procjene aritmetičke sredine s očitim razlikama. Koraci su sljedeći. (1) Izračunati proporciju uzorka. Proporcija uzorka jest procjena brojem proporcije osnovnog skupa. (2) Izračunati standardnu pogrešku procjene proporcije. (3) Odrediti koeficijent pouzdanosti procjene $z$. (4) Odrediti donju i gornju granicu procjene.

## PRIMJER 5.

Procjenjuje se procjena proporcije rukovoditelja o očekivanom stanju likvidnosti u prvom tromjesečju 2013. godine s pomoću jednostavnoga slučajnog uzorka i na razini povjerenja 90 %. Od ukupno 5 689 rukovoditelja u poslovnim subjektima srednje veličine i djelatnosti proizvodnja hrane i pića anketirano je njih 225. Anketno je pitanje ovo: „Što predviđate glede stanja likvidnosti u prvom tromjesečju 2014.?“ Odgovori su: poboljšanje stanja, nepromijenjeno stanje, pogoršanje stanja. Od ukupnog broja anketiranih njih 81 je odgovorilo da u prvom tromjesečju 2014. očekuju poboljšanje stanja likvidnosti.

Procjena proporcije *svih* rukovoditelja koji očekuju poboljšanje stanja likvidnosti provodi se u navedenim koracima.

(1) Procjenitelj proporcije osnovnog skupa brojem (proporcija uzorka) i procjena jesu:

$$\hat{p} = \frac{m}{n},\ m = 81,\ n = 225;\ \hat{p} = 0.36,\ \hat{q} = 1 - \hat{p},\ \hat{q} = 0.60\ .$$

Prema tome, procjena rukovoditelja koji očekuju poboljšanje likvidnosti u spomenutom razdoblju je 0.36 ili 36 %. Proporcija rukovoditelja s drugim odgovorima je $\hat{q} = 0.64$, odnosno 64 %.

(2) Da bi se stekao uvid u preciznost procjene, potrebno je izračunati vrijednost standardne pogreške na temelju ovih podataka: $N$=5689, $n$=225, $\hat{p} = 0.36, \hat{q} = 0.64$ i to primjenom izraza (6.8 a):

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{N-n}{N-1}\right)} = \sqrt{\frac{0.36 \cdot 0.64}{225-1}\left(\frac{5689-225}{5689-1}\right)},\ \sigma_{\hat{p}} = 0.03143\,.$$

Slijedi određivanje koeficijenta pouzdanosti.

(3) Neka se proporcija osnovnog skupa procjenjuje vjerojatnošću 0.90, odnosno na 90 % razini povjerenja. Toj razini povjerenja (vjerojatnosti) odgovara koeficijent pouzdanosti $z = 1.65$.

(4) Intervalni je procjenitelj $\hat{p} - 1.65\,\sigma_{\hat{p}} < p < \hat{p} + 1.65\,\sigma_{\hat{p}}$, a vrijednosti granica su:

$$0.36 - 1.65 \cdot 0.03143 < p < 0.36 + 1.65 \cdot 0.03143\,,$$

to jest:

$0.30814 < p < 0.41186$.

Neka je zaokružena vrijednost donje granice 0.31, a gornje 0.41. Interval se tumači ovako: s vjerojatnošću 0.90 proporcija *svih* rukovoditelja koji očekuju poboljšanje stanja likvidnosti jest između 0.31 i 0.41, odnosno između 31 % i 41 %. I ovdje valja ponovo naglasiti da nije riječ o tvrdnji da se proporcija osnovnog skupa doista nalazi u navedenom intervalu. Kad bismo iz skupa izabrali sve moguće uzorke iste veličine i odredili isto toliko intervala procjene, u 90 % slučajeva u intervalima bi se nalazila proporcija osnovnog skupa.

## PITANJA I ZADATCI

**1.** Kako glasi procjenitelj proporcije osnovnog skupa brojem s pomoću jednostavnoga slučajnog uzorka? Na koji se način dolazi do procjene? Koji uvjeti moraju biti ispunjeni da bi se mogao primijeniti intervalni procjenitelj (6.7)?

**2.** Što se mjeri standardnom pogreškom procjene proporcije? Kako se tumači standardna pogreška?

**3.** O kojim veličinama ovisi veličina standardne pogreške procjene proporcije konačnoga osnovnog skupa jednostavnim slučajnim uzorkom? Kolika je standardna pogreška ako je $n = N$? Objasnite odgovor.

**4.** Izračunajte vrijednost standardne pogreške procjene proporcije ako su zadane ove veličine: (a) $N = 26543$, $n = 352$, $p = 0.5$, $q = 0.5$, (b) $N = 97\,572$, $n = 128$, $\hat{p} = 0.28$.

**5.** Intervalni procjenitelj proporcije osnovnog skupa dovoljno velikim uzorkom jest $\hat{p} - z\sigma_{\hat{p}} < p < \hat{p} + z\sigma_{\hat{p}}$. S pomoću podataka iz uzoraka dobiveni su ovi rezultati: (a) $0.73264 < p < 0.79241$, (b) $-0.0241 < p < 0.1254$, (c) $0.9585 < p < 1.0123$, (d) $n = 89, \hat{p} = 0.025$, (e) $n = 272, \hat{p} = 0.3724, \hat{q} = 0.6276$.
Koji su od navedenih rezultata u skladu s uvjetima primjene procjenitelja proporcije osnovnog skupa (7)? Odgovor obvezatno obrazložite.

**6.** U jednostavni slučajni uzorak izabrano je 40 učenika završnih razreda srednjih škola jednoga grada, kojih je ukupno 1 856, kako bi se procijenila proporcija svih učenika završnih razreda koji su pušači (puše 10 i više cigareta na dan). Od ukupnog broja učenika izabranih u uzorak, njih 15 su pušači. (a) Procijenite brojem i 95 % intervalom proporciju pušača završnih razreda srednjih škola danoga grada. (b) Što se može reći o preciznosti procjene? (c) Na koji se način može povećati preciznost procjene?

**7.** Područnom uredu upućeno je 43 696 poreznih prijava za 1999. godinu. S pomoću slučajnog uzorka procjenjuje se proporcija neispravnih prijava. U slučajnom uzorku 450 prijava njih 90 je neispravno. Procijenite brojem i 90 % intervalom proporciju neispravno popunjenih poreznih prijava.

**8.** Kako se određuje veličina jednostavnoga slučajnog uzorka iz konačnog skupa za intervalnu procjenu proporcije?
(a) Ako se u procjeni tolerira odstupanje od proporcije osnovnog skupa veličine *d*, a procjenjuje se na razini povjerenja kojoj odgovara koeficijent povjerenja *z*, veličina *d* jednaka je $d = z\sigma_{\hat{p}}$, odnosno:

$$d = z\sqrt{\frac{pq}{n}\left(\frac{N-n}{n-1}\right)}.$$

U prethodnom izrazu *p* je nepoznata vrijednost proporcije, čija se vrijednost radi određivanja veličine uzorka prosuđuje na temelju poznavanja pojave. U najnepovoljnijem slučaju uzima se da je $p = 0.5$, kada je produkt *pq* najveći. Veličina uzorka *n* jest rješenje navedene jednadžbe, to jest:

$$n = \frac{Nz^2pq}{d^2(N-1) + z^2pq}.$$

Ako je riječ o uzorku iz beskonačne populacije ili je $n$ manji od 5 % od $N$, veličina uzorka određuje se izrazom:

$$n = \left[\frac{z\sqrt{pq}}{d}\right]^2.$$

(b) S pomoću jednostavnoga slučajnog uzorka procjenjuje se proporcija kućanstava jednoga grada čiji članovi redovito čitaju dnevnik Jutarnje novine. Koliko će se kućanstava izabrati u jednostavni slučajni uzorak od njih 2 695 ako se tolerira odstupanje od nepoznate proporcije osnovnog skupa od 0.05 (5 %) kuna, a procjenjuje na razini povjerenja 90 %? (Uputa: pretpostavite da je $p = q = 0.5$).

**9.** Protumačite sadržaj sljedećeg ispisa obradbe programskom potporom.

**Interval procjene proporcije populacije (veliki uzorci)**

| | |
|---|---|
| **Veličina uzorka....................................................** | **256** |
| **Broj elemenata s danim modalitetom...............** | **121** |
| **Empirijska proporcija.........................................** | **0.4732** |
| **Standardna pogreška proporcije......................** | **0.03121** |
| **Teorijska vrijednost normalne distribucije........** | **1.960** |

**Interval procjene proporcije populacije (veliki uzorci) (95.00 %):**
**(0.4121..... 0.5344)**

# Testiranje hipoteze o pretpostavljenoj sredini i proporciji osnovnog skupa

Među temeljnim postupcima u sklopu metode uzoraka su oni o ispitivanju pretpostavki (hipoteza) o osobitostima jednoga ili više osnovnih skupova (populacija). O prirodi statističkih testova već je bilo riječi. Tako smo istaknuli da se oni odnose na tvrdnje o veličini parametra ili obliku rasporeda jednoga ili više osnovnih skupova. Ispitivati se mogu različite pretpostavke. Ovdje će se predočiti najprije najjednostavniji test o pretpostavljenoj sredini osnovnog skupa, a zatim test o pretpostavljenoj proporciji osnovnog skupa.

## Testiranje hipoteze o pretpostavljenoj sredini osnovnog skupa

Svaki postupak ispitivanja statističkih hipoteza počiva na dvjema tvrdnjama koje su sadržajno u suprotnosti. Tvrdnje su navedene u nultoj i alternativnoj pretpostavci. Postupkom testiranja s pomoću podataka iz uzorka donosi se odluka o tome može li se prihvatiti nulta pretpostavka kao istinita ili se ona odbacuje.

Postupak testiranja o pretpostavljenoj aritmetičkoj sredini osnovnog skupa s pomoću slučajnog uzorka provodi se u sljedećim koracima.

> **POSTUPAK TESTIRANJA.** (1) određivanje nulte i alternativne pretpostavke, (2) utvrđivanje razine značajnosti (pogreške tipa I., vjerojatnosti odbacivanja istinite nulte hipoteze), (3) izračunavanje veličina s pomoću kojih se donosi odluka (granica, odnosno granice prihvaćanja nulte hipoteze ili druge odgovarajuće veličine), (4) donošenje odluke.

(1) Pretpostavke o veličini aritmetičke sredine osnovnog skupa su ove: aritmetička sredina osnovnog skupa jednaka je pretpostavljenoj vrijednosti, koju ćemo označiti s $H_0$ (nulta pretpostavka); aritmetička sredina osnovnog skupa razlikuje se od pretpostavljene vrijednosti ( $H_1$ alternativna pretpostavka). Simbolički:

$$H_0 \ldots \mu = \mu_0 ; H_1 \ldots \mu \neq \mu_0 .$$

Ovaj test naziva se *dvosmjernim testom o pretpostavljenoj aritmetičkoj sredini osnovnog skupa.*[10] To je stoga što alternativna pretpostavka sadržava tvrdnju da sredina osnovnog skupa može biti veća i manja od $\mu_o$ .

Prije smo naveli zadatak ispitivanja pretpostavke o prosječnoj starosti potrošača napitka. Pretpostavlja se da je prosječna starost 30 godina, to jest $\mu_o$ = 30, pa će test o toj tvrdnji počivati na ovim hipotezama:

$$H_0 \ldots \mu = 30 ; H_1 \ldots \mu \neq 30 .$$

(2) Slijedi određivanje razine značajnosti. Razinu značajnosti opisali smo kao vjerojatnost neprihvaćanja istinite nulte hipoteze, odnosno pogrešku tipa I. Uobičajeno se uzima da je ta vjerojatnost 0.01, 0.05, 0.10 ili postotno izražena 1 %, 5 %, 10 %. Njezina se veličina određuje od slučaja do slučaja i dio je plana provođenja testa.

[10] Ima i drugih oblika hipoteza o pretpostavljenoj sredini, ali o njima ovdje neće biti riječi.

(3) Da bismo bolje razumjeli prethodni korak i način odlučivanja, potrebno je imati na umu da se testiranje provodi na temelju *jednoga* slučajnog uzorka. Kako smo pokazali, iz konačnog skupa $N$ jedinica sa sredinom $\mu_o$ moguće je izabrati onoliko različitih uzoraka veličine $n$ jedinica koliko ima kombinacija. Sredine uzoraka razlikuju se međusobno jer se odnose na različite uzorke. Uredimo li sve sredine, doći ćemo do distribucije sredina uzoraka. Ako je riječ o velikom uzorku, distribucija će biti približno oblika normalne distribucije sa sredinom $\mu_o$ i standardnom devijacijom koja se naziva standardnom pogreškom sredine $\sigma_{\bar{x}}$. Ta se standardna pogreška određuje na isti način kao i pogreška procjene sredine. Budući da se test oslanja na slučajni uzorak, normalna distribucija ima svojstva distribucije vjerojatnosti.

U postupku testiranja pretpostavit će se da je nulta hipoteza istinita, to jest da joj je sredina $\mu_o$. Koristeći se vrijednošću pretpostavljene sredine, podatke iz uzorka i svojstva normalne distribucije moguće je odrediti granice unutar kojih se s odabranom vjerojatnošću može naći aritmetička sredina uzorka.

**GRANICE PRIHVAĆANJA NULTE HIPOTEZE** o pretpostavljenoj vrijednosti aritmetičke sredine osnovnog skupa s pomoću velikoga jednostavnoga slučajnog uzorka jesu:

(9) $\mu_0 - z\sigma_{\bar{x}} < \bar{x} < \mu_0 + z\sigma_{\bar{x}}$ .

Navedene granice razdvajaju područje prihvaćanja od područja neprihvaćanja nulte pretpostavke. $\mu_o$ je pretpostavljena aritmetička sredina osnovnog skupa, $\sigma_{\bar{x}}$ je standardna pogreška sredine. $z$ je koeficijent koji govori koliko je najveće dopušteno odstupanje aritmetičke sredine uzorka $\bar{x}$ od pretpostavljene sredine $\mu_o$ izraženo u jedinicama standardne pogreške sredine. Koeficijent je ovisan o razini značajnosti pa se i naziva koeficijentom značajnosti. Određuje se s pomoću tablica površina normalne distribucije, odnosno odgovarajućim računalnim programom. Granice prihvaćanja nulte hipoteze za odabrane razine značajnosti prikazane su u tablici.

| Razina značajnosti (vjerojatnost odbacivanja istinite hipoteze) | Teorijske vrijednosti koeficijenta $z$ | Granice prihvaćanja nulte pretpostavke |
|---|---|---|
| 0.32 | ±1 | $\mu_0 - \sigma_{\bar{x}} < \bar{x} < \mu_0 + \sigma_{\bar{x}}$ |
| 0.10 | ±1.65 | $\mu_0 - 1.65\ \sigma_{\bar{x}} < \bar{x} < \mu_0 + 1.65\ \sigma_{\bar{x}}$ |
| 0.05 | ±2 | $\mu_0 - 2\sigma_{\bar{x}} < \bar{x} < \mu_0 + 2\sigma_{\bar{x}}$ |
| 0.01 | +2.58 | $\mu_0 - 2.58\,\sigma_{\bar{x}} < \bar{x} < \mu_0 + 2.58\,\sigma_{\bar{x}}$ |

(4) Odluka se donosi s pomoću testovne veličine. Testovna veličina jest aritmetička sredina uzorka $\bar{x}$ .

**NULTA HIPOTEZA PRIHVAĆA SE** na danoj razini značajnosti ako se aritmetička sredina uzorka $\bar{x}$ nalazi unutar granica prihvaćanja nulte hipoteze, a ne prihvaća ako je sredina uzorka veća od donje ili manja od gornje granice prihvaćanja nulte hipoteze.

Postupak donošenja odluke provodi se i usporedbom standardiziranog odstupanja sredine uzorka od pretpostavljene sredine osnovnog skupa i teorijske vrijednosti koeficijenta *z*. Oba načina dovode do istog rezultata. Testovna veličina *z** je odstupanje sredine uzorka od pretpostavljene sredine osnovnog skupa izraženo u jedinicama standardne pogreške.[11] Odluka je sljedeća.

**TESTOVNA VELIČINA** definira se izrazom:

(10) $$z^* = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}.$$

Nulta pretpostavka prihvaća se na danoj razini značajnosti ako je z* omjer prema apsolutnoj vrijednosti manji od z, to jest ako je $|z^*| < z$. Nulta pretpostavka ne prihvaća se ako je z* omjer prema apsolutnoj vrijednosti veći od vrijednosti z, to jest ako je $|z^*| > z$.

U koraku (1) naveli smo hipoteze o pretpostavljenoj sredini potrošača. Nulta hipoteza sadržava tvrdnju da je prosječna starost potrošača 30 godina, odnosno $\mu_o = 30$.

## PRIMJER 6.

Test (1) $H_0 ....\mu = 30; H_1 ....\mu \neq 30$ provodi se razini značajnosti 5 %, kojoj odgovara koeficijent značajnosti $z = 2$. (2) Sa svrhom provođenja testa izabran je slučajni uzorak 265 potrošača. U uzorak je izabrano manje od 5 % potrošača. Na temelju podataka iz uzorka izračunana je prosječna starost potrošača u uzorku $\bar{x} = 28.9$ godina. Procijenjena je i standardna devijacija osnovnog skupa $\hat{\sigma} = 5.2$ godine.

(3) Granice prihvaćanja nulte hipoteze (razina značajnosti 5 %) jesu:

$$\mu_0 - 2\sigma_{\bar{x}} < \bar{x} < \mu_0 + 2\sigma_{\bar{x}}.$$

U navedenom izrazu $\mu_o$ =30, a standardnu pogrešku sredine $\sigma_{\bar{x}}$ valja izračunati. Standardna pogreška sredine jest:

$$\sigma_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

[11] Treći način zaključivanja o ishodu testa temelji se na usporedbi razine značajnosti, koju ćemo označiti s α i razine značajnosti izračunane s pomoću *z** i površina ispod normalne krivulje. Razina značajnosti α utvrđuje se u nacrtu testa. Izračunana razina značajnosti naziva se *p-vrijednost* (*p-value, significance*) i temelji se na rezultatima iz uzorka. Način odlučivanja je ovaj: ako je α > *p-vrijednost*, prihvaća se nulta hipoteza. Kada je α < *p-vrijednost*, nulta hipoteza se ne prihvaća. Sva tri načina zaključivanja dovode do istog rezultata. Kada se navodi da je $p < 0.01$ ili $p << 0.01$, to upućuje na odbacivanje nulte hipoteze.

Kako je u uzorak izabrano iz osnovnog skupa manje od 5 % jedinica, faktor $\sqrt{\frac{N-n}{N-1}}$ može se zanemariti, odnosno uzeti da je približno jednak jedan. Standardna pogreška sredine jest:

$$\sigma_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{5.2}{\sqrt{265}}, \sigma_{\bar{x}} = 0.31943\,,$$

a granice prihvaćanja nulte hipoteze:

$30\text{-}2 \cdot 0.31943 < \bar{x} < 30+2 \cdot 0.31943$,

to jest:

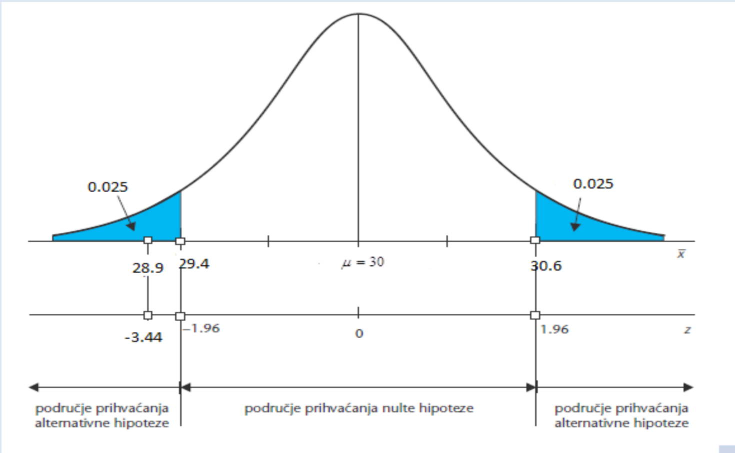
$29.4 < \bar{x} < 30.6$.

(4) Odluka je ova: aritmetička sredina uzorka je $\bar{x} = 28.9$ i ne pada u područje prihvaćanja nulte pretpostavke. Prema tome, na danoj razini značajnosti ne prihvaća se pretpostavka da je uzorak izabran iz skupa sa sredinom 30. Drugim riječima, na temelju rezultata uzorka ne prihvaća se pretpostavka da je prosječna starost potrošača 30 godina.

Do iste se odluke dolazi primjenom *z* testa. Ako je nulta hipoteza istinita, a test se provodi na danoj razini značajnosti, aritmetička sredina uzorka može odstupiti od pretpostavljene sredine za najviše ±2 standardne pogreške. Testovna veličina jest:

$$z^* = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{28.9 - 30}{0.31943}, \; z^* = -3.44\,.$$

Aritmetička sredina uzorka odstupa od pretpostavljene sredine za –3.44 standardne pogreške. Ako je nulta pretpostavka istinita, najveće dopušteno odstupanje je –2, što upućuje na neprihvaćanje nulte pretpostavke. Prema tome, na temelju rezultata uzorka ne prihvaća se pretpostavka da je prosječna starost potrošača 30 godina.




**Slika 3.** *Dvosmjerni test o pretpostavljenoj sredini osnovnog skupa (veliki uzorak)*

Je li prihvaćena tvrdnja koja je ishod statističkog testa *doista* istinita? Budući da se zaključak o pretpostavkama donosi s pomoću dijela podataka, kako smo već nekoliko puta naveli, u zaključivanju je moguće pogriješiti. Moguće je da je uzorak izabran iz skupa s aritmetičkom sredinom $\mu_o = 30$ premda smo na temelju rezultata iz uzorka zaključili suprotno. Pogreška takve vrste govori o neprihvaćanju istinite nulte pretpostavke i ona je izražena u obliku razine značajnosti. U primjeru razina značajnosti je 0.05, odnosno 5 %. To znači da bi od ukupnog broja intervala prihvaćanja nulte hipoteze, kojih ima koliko i različitih uzoraka, njih 5 % ne bi sadržavalo aritmetičku sredinu uzorka.

Uzmimo sada da je aritmetička sredina slučajnog uzorka $\bar{x} = 30.2$. Prema opisanom pravilu zaključivanja, prihvatili bismo pretpostavku da uzorak potječe iz skupa potrošača s prosječnom starosti 30 godina. No uzorak s navedenom sredinom i istom pogreškom sredine mogao je potjecati iz skupa sa sredinom jednakom 32 godine. Prihvaćajući nultu pretpostavku kao istinitu, a ona je lažna, počinili smo pogrešku tipa II. U postupku testiranja pretpostavki izračunava se vjerojatnost pojavljivanja te pogreške.

Kao što proizlazi iz dosad rečenoga, zaključak u postupku testiranja statističke hipoteze ima obilježje suda koji je *vjerojatno* istinit. O tome valja voditi računa pri tumačenju rezultata testa.

Uzmimo još jedan primjer testiranja pretpostavke o pretpostavljenoj sredini osnovnog skupa.

**PRIMJER 7.**

Mesna pašteta tvrtke GAMA u plastičnom ovitku ima deklariranu prosječnu netotežinu 5 dag, odnosno 50 g. Radi provjere težine proizvoda iz pošiljke 10 000 komada izabran je slučajni uzorak 50 ovitaka. Mjerenjem težine dobivene su sljedeće vrijednosti (u gramima).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 54 | 50 | 48 | 49 | 48 | 48 | 52 | 50 | 48 | 49 |
| 50 | 50 | 50 | 53 | 50 | 49 | 53 | 50 | 51 | 49 |
| 48 | 46 | 50 | 50 | 52 | 50 | 50 | 48 | 48 | 51 |
| 48 | 49 | 54 | 50 | 51 | 51 | 52 | 50 | 51 | 53 |
| 48 | 50 | 46 | 48 | 53 | 48 | 48 | 50 | 49 | 50 |

Može li se prihvatiti pretpostavka da je prosječna neto-težina proizvoda u pošiljci 50 grama? Razina značajnosti je 5 %.

(1) Nulta pretpostavka jest da je prosječna neto-težina proizvoda u pošiljci 50 g, a alternativna da je prosječna težina proizvoda u pošiljci različita od 50 g. Pretpostavljena je sredina $\mu_o = 50$. Simbolički, hipoteze su:

$$H_0....\mu = 50; H_1....\mu \neq 50 .$$

(2) Razina značajnosti je 5 %, odnosno vjerojatnost odbacivanja istinite nulte hipoteze je 0.05. Toj razini značajnosti odgovara koeficijent značajnosti $z = 2$. Uzorak je velik jer broji više od 30 elemenata.

(3) Odlučivanje o nultoj pretpostavci je na temelju granica $\mu_0 - 2\sigma_{\bar{x}} < \bar{x} < \mu_0 + 2\sigma_{\bar{x}}$ i testovne veličine. Budući da standardna devijacija

osnovnog skupa nije poznata, procijenit će se kako bi se izračunala standardna pogreška sredine. Testovna veličina je aritmetička sredina uzorka, to jest prosječna težina proizvoda u uzorku.

Aritmetička sredina uzorka jest:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{54+50+48+\cdots+50}{50},\ \overline{x} = 49.86,$$

a procjena standardne devijacije osnovnog skupa:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1}} = \sqrt{\frac{124471 - \frac{2493^2}{50}}{50-1}},\ \hat{\sigma} = 1.86274\,.$$

Računanje standardne pogreške sredine polazi od ovih veličina: $N = 10\,000$, $n = 50$, $\hat{\sigma} = 1.86274$. Standardna pogreška sredine jest:

$$\sigma_{\overline{x}} = \frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.86274}{\sqrt{50}}\sqrt{\frac{10000-50}{10000-1}},\ \hat{\sigma}_{\overline{x}} = 0.26278\,.$$

Granice prihvaćanja nulte pretpostavke (razina značajnosti 5 %) su:

$\mu_0 - 2\sigma_{\overline{x}} < \overline{x} < \mu_0 + 2\sigma_{\overline{x}}$, to jest:

$50 - 2\cdot 0.26278 < \overline{x} < 50 + 2\cdot 0.26278$,

*odnosno:*

$49.47444 < \overline{x} < 50.52556$.

Odluku je moguće donijeti i s pomoću *z*-omjera, koji je:

$$z^* = \frac{\overline{x} - \mu_0}{\sigma_{\overline{x}}} = \frac{49.86 - 50}{0.26278},\ z^* = -0.53\,.$$

(4) Odluka je ova: aritmetička sredina uzorka (49.86) veća je od donje granice prihvaćanja nulte hipoteze (49.47444), manja od gornje granice (50.52556), to jest sredina uzorka pada u područje prihvaćanja nulte hipoteze.

Isti je zaključak i s pomoću *z*-omjera. Sredina uzorka odstupa od pretpostavljene sredine za −0.53 standardne pogreške. Nulta se pretpostavka prihvaća ako je izračunani omjer u intervalu ±2. Vrijednost izračunanog omjera nalazi se unutar spomenutog intervala pa se prihvaća nulta pretpostavka.

Prihvaćajući nultu pretpostavku, prihvaća se (uz danu razinu značajnosti) tvrdnja da je prosječna težina proizvoda u pošiljci 50 g, to jest da je u skladu s deklariranom težinom.

## PITANJA I ZADATCI

**1.** Od čega se sastoji ispitivanje statističkih hipoteza? Opišite korake ispitivanja hipoteze o pretpostavljenoj sredini osnovnog skupa s pomoću velikoga jednostavnoga slučajnog uzorka.

**2.** Kako se naziva interval $\mu_0 - z\sigma_{\bar{x}} < \bar{x} < \mu_0 + z\sigma_{\bar{x}}$? Objasnite značenje svakog simbola. Na koji se način donosi odluka u postupku testiranja s pomoću navedenog intervala?

**3.** Dan je sljedeći omjer: $z^* = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$. Objasnite značenje svakog simbola u navedenom izrazu. Za koje se svrhe rabi i na koji način? Jesu li rezultati postupka odlučivanja s pomoću intervala u prethodnom zadatku i ovog omjera jednaki?

**4.** Usporedite intervale: (1) $\mu_0 - z\sigma_{\bar{x}} < \bar{x} < \mu_0 + z\sigma_{\bar{x}}$ i (2) $\bar{x} - z\sigma_{\bar{x}} < \mu < \bar{x} + z\sigma_{\bar{x}}$. Objasnite njihove sadržaje, razlike i sličnosti.

**5.** Dane su ove veličine: $N = 15\,879$, $n = 225$, $\bar{x} = 98, \sigma = 15$. Testirajte ove hipoteze: $H_0 .... \mu = 100; H_1 .... \mu \neq 100$. Razina značajnosti je 5 %. Objasnite odluku. U postupku se koristite obama načinima donošenje odluke o ishodu testa.

**6.** Iz osnovnog skupa od 6 000 jedinica izabran je jednostavni slučajni uzorak 64 jedinice. Na temelju vrijednosti iz uzorka dobivene su ove veličine: $\sum_{i=1}^{64} x_i = 62.8, \sum_{i=1}^{64} (x_i - \bar{x})^2 = 9.24$. Može li se prihvatiti pretpostavka da je uzorak izabran iz osnovnog skupa sa sredinom koja je jednaka 1.1? Testirati na razini 5 % i 1 % značajnosti. Odluku donesite s pomoću granica prihvaćanja nulte hipoteze i izračunanoga odgovarajućeg omjera.

**7.** Zračni prijevoznik TRANS-AIR s pomoću uzorka ispituje prosječnu težini prtljage putnika na liniji AB200. U uzorku prtljage 70 putnika zabilježena je sljedeća težina prtljage.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 20.2 | 15.5 | 19.2 | 21.8 | 19.7 | 19.1 | 23.6 | 21.9 | 17.9 |
| 22.0 | 17.8 | 23.3 | 17.4 | 16.4 | 19.7 | 23.7 | 20.6 | 19.5 |
| 25.3 | 19.3 | 22.2 | 20.2 | 19.9 | 22.0 | 20.7 | 22.8 | 20.3 |
| 18.0 | 20.6 | 21.6 | 18.6 | 17.2 | 18.3 | 19.9 | 21.5 | 19.5 |
| 22.2 | 19.3 | 21.2 | 17.9 | 19.5 | 21.8 | 19.2 | 18.2 | 23.6 |
| 19.6 | 16.4 | 17.8 | 20.0 | 21.5 | 19.4 | 17.2 | 21.3 | 18.7 |
| 20.2 | 19.4 | 20.4 | 18.5 | 21.6 | 19.4 | 20.2 | 17.4 | 19.7 |

Može li se prihvatiti pretpostavka da je prosječna težina prtljage svih putnika na liniji AB200 jednaka 20 kg? Testirati na razini 5 % značajnosti. Pri računanju standardne pogreške sredine zanemarite faktor korekcije za konačne skupove. $\sum_{i=1}^{70} x_i = 1402.8, \sum_{i=1}^{70} x_i^2 = 28417.7$.

**8.** Protumačite sljedeći ispis obradbe programskom potporom:
**Test o pretpostavljenoj sredini populacije**

**Varijabla : prodaja**
**Veličina uzorka.................................................... 125**
**Aritmetička sredina uzorka................................ 46.853**
**Procjena standardne devijacije populacije......... 25.659**
**Standardna pogreška sredine............................ 2.295**

| | |
|---|---|
| **Pretpostavljena sredina populacije....................** | **50.000** |
| **Razina signifikantnosti.......................................** | **0.05000** |
| **Teorijska vrijednost normalne distribucije........** | **1.960** |
| **Empirijski *z*-omjer..............................................** | **–1.371** |

**H i p o t e z e :**

**Nulta hipoteza................ Sredina populacije = 50.000**

**Alternativna hipoteza..... Sredina populacije <> 50.000**

**O d l u k a - Na dvije granice :**

**Na danoj razini signifikantnosti prihvaća se nulta hipoteza.**

## Testiranje hipoteze o pretpostavljenoj proporciji osnovnog skupa

Postupak ispitivanja pretpostavke o proporciji osnovnog skupa sličan je postupku ispitivanja pretpostavke o pretpostavljenoj vrijednosti aritmetičke sredine osnovnog skupa. Razmotrit će se samo slučaj kada se testiranje provodi s pomoću velikog uzorka. Postupak se također provodi u ovim koracima:

(1) određivanje nulte i alternativne hipoteze, (2) utvrđivanje razine značajnosti (pogreške tipa I., vjerojatnosti odbacivanja istinite nulte hipoteze), (3) izračunavanje veličina s pomoću kojih se donosi odluka (granica odnosno granice prihvaćanja nulte hipoteze ili druge odgovarajuće veličine), (4) donošenje odluke.

(1) Pretpostavke o veličini proporcije osnovnog skupa su ove: proporcija osnovnog skupa jednaka je pretpostavljenoj vrijednosti, koju ćemo označiti s $p_o$ (nulta pretpostavka); proporcija osnovnog skupa razlikuje se od pretpostavljene vrijednosti (alternativna pretpostavka). Simbolički:

$$H_0....p = p_0; H_1....p \neq p_0 .$$

Ovaj test naziva se *dvosmjernim testom o pretpostavljenoj proporciji osnovnog skupa*.

(2) Slijedi određivanje razine značajnosti. Razinu značajnosti opisali smo kao vjerojatnost neprihvaćanja istinite nulte hipoteze, odnosno pogrešku tipa I. Uobičajeno se uzima da je ta vjerojatnost 0.01,0.05,0.10 ili postotno izražena 1 %, 5 %, 10 %.

(3) U postupku testiranja pretpostavit će se da je nulta hipoteza istinita, to jest da joj je proporcija skupa $p_0$. Koristeći se vrijednošću pretpostavljene proporcije, podatcim iz uzorka i svojstvima normalne distribucije, moguće je odrediti granice unutar kojih se s odabranom vjerojatnosti može naći proporcija uzorka izabranoga iz skupa s pretpostavljenom proporcijom.

**GRANICE PRIHVAĆANJA NULTE HIPOTEZE** o pretpostavljenoj vrijednosti proporcije osnovnog skupa s pomoću velikoga jednostavnoga slučajnog uzorka jesu:

(11) $$p_0 - z\sigma_{\hat{p}} < \hat{p} < p_0 + z\sigma_{\hat{p}}.$$

Standardna pogreška u navedenom izrazu jest:

(12) $$\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}\left(\frac{N-n}{N-1}\right)}.$$

Navedene granice razdvajaju područje prihvaćanja od područja neprihvaćanja nulte hipoteze. $p_o$ je pretpostavljena proporcija osnovnog skupa, $\sigma_{\hat{p}}$ je standardna pogreška proporcije. $z$ je koeficijent koji govori koliko je najveće dopušteno odstupanje proporcije uzorka $\hat{p}$ od pretpostavljene proporcije $p_o$ izraženo u jedinicama standardne pogreške proporcije. Koeficijent je ovisan o razini značajnosti. Granice prihvaćanja nulte hipoteze za odabrane razine značajnosti i test s pomoću velikog uzorka prikazane su u tablici.

| Razina značajnosti (vjerojatnost odbacivanja istinite pretpostavke) | Teorijske vrijednosti koeficijenta *z* | Granice prihvaćanja nulte pretpostavke |
|---|---|---|
| 0.32 | ±1 | $p_0 - \sigma_{\hat{p}} < \hat{p} < p_0 + \sigma_{\hat{p}}$ |
| 0.10 | ±1.65 | $p_0 - 1.65\,\sigma_{\hat{p}} < \hat{p} < p_0 + 1.65\,\sigma_{\hat{p}}$ |
| 0.05 | ±2 | $p_0 - 2\sigma_{\hat{p}} < \hat{p} < p_0 + 2\sigma_{\hat{p}}$ |
| 0.01 | ±2.58 | $p_0 - 2.58\,\sigma_{\hat{p}} < \hat{p} < p_0 + 2.58\,\sigma_{\hat{p}}$ |

Postupak donošenja odluke provodi se i usporedbom standardiziranog odstupanja proporcije uzorka od pretpostavljene proporcije osnovnog skupa i teorijske vrijednosti koeficijenta $z$. Testovna veličina jest omjer: $z^* = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$. Testovna veličina $z^*$ je odstupanje proporcije uzorka od pretpostavljene proporcije osnovnog skupa izraženo u jedinicama standardne pogreške.[12] Odluka je sljedeća.

[12] Treći način zaključivanja o ishodu testa temelji se na usporedbi razine značajnosti, koju ćemo označiti s $\alpha$, i razine značajnosti izračunane s pomoću $z^*$ i površina ispod normalne krivulje. Razina značajnosti $\alpha$ utvrđuje se u nacrtu testa. Izračunana razina značajnosti naziva se *p-vrijednošću* (*p-value*) ili „izračunanom razinom značajnosti" (*significance*) i temelji se na rezultatima iz uzorka. Način odlučivanja je ovaj: ako je $\alpha$ > *p-vrijednost*, prihvaća se nulta hipoteza. Kada je $\alpha$ < *p-vrijednost*, nulta hipoteza se ne prihvaća. Sva tri načina zaključivanja dovode do istog rezultata.

**NULTA PRETPOSTAVKA SE PRIHVAĆA** na danoj razini značajnosti ako je $z^*$ omjer prema apsolutnoj vrijednosti manji od z, to jest ako je $|z^*| < z$. Nulta pretpostavka ne prihvaća se ako je $z^*$ omjer prema apsolutnoj vrijednosti veći od vrijednosti $z$, to jest ako je $|z^*| > z$.

Postupak testiranja hipoteze o pretpostavljenoj proporciji ilustrirat ćemo jednim primjerom.

### PRIMJER 8.

Pretpostavlja se da od 35 928 kućanstava jedne regije njih 75 % posjeduje perilicu za rublje. U slučajnom uzorku 1 000 kućanstava ustanovljeno je da 765 kućanstava posjeduje perilicu. Može li se prihvatiti navedena pretpostavka na razini 5 % značajnosti?

U ovom primjeru dane su sljedeće veličine: $N = 35\,928$, $n = 1\,000$, $m = 765$,

$\hat{p} = \frac{m}{n} = \frac{765}{1000}$, $\hat{p} = 0.765$, pretpostavljena proporcija $p_0 = 0.75$.

(1) Nulta pretpostavka jest: proporcija kućanstava koja posjeduju perilicu jest 75 %, odnosno proporcija tih kućanstava je 0.75. Alternativna je pretpostavaka da se proporcija takvih kućanstava razlikuje od 0.75. Pretpostavljena proporcija $p_0 = 0.75$. Simbolički, pretpostavke su:

$$H_0 \ldots p = 0.75; H_1 \ldots p \neq 0.75 .$$

(2) Razina značajnosti je 5 %, odnosno vjerojatnost odbacivanja istinite nulte hipoteze je 0.05. Toj razini značajnosti odgovara koeficijent značajnosti $z = 2$.

(3) Odlučivanje o nultoj hipotezi je na temelju granica $p_0 - 2\sigma_{\hat{p}} < \hat{p} < p_0 + 2\sigma_{\hat{p}}$ i testovne veličine. Računanje standardne pogreške proporcije polazi od ovih veličina: $N = 35\,928$, $n = 1\,000$, $p_0 = 0.75$, $q_0 = 0.25$. Standardna pogreška proporcije jest:

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}\frac{N-n}{N-1}} = \sqrt{\frac{0.75 \cdot 0.25}{1000}\left(\frac{35928-1000}{35928-1-1}\right)},\ \hat{\sigma}_{\hat{p}} = 0.01350 .$$

Granice prihvaćanja nulte hipoteze (razina značajnosti 5 %) su:

$p_0 - 2\sigma_{\hat{p}} < \hat{p} < p_0 + 2\sigma_{\hat{p}}$, to jest:

$$0.75 - 2 \cdot 0.01350 < \hat{p} < 0.75 + 2 \cdot 0.01350 ,$$

*odnosno:*

$$0.723 < \hat{p} < 0.777 .$$

$z$-omjer je: $z^* = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.765 - 0.75}{0.01350}$, $z^* = 1.11$.

(4) Odluka je ova: proporcija uzorka (0.765) veća je od donje granice prihvaćanja nulte hipoteze (0.723), manja od gornje granice (0.777), to jest proporcija uzorka pada u područje prihvaćanja nulte hipoteze.

Isti je zaključak i s pomoću *z*- omjera. Proporcija uzorka odstupa od pretpostavljene proporcije za 1.11 standardnih pogrješaka. Nulta pretpostavka se prihvaća ako je izračunani omjer u intervalu ±2. Vrijednost izračunanog omjera nalazi se unutar spomenutog intervala pa se prihvaća nulta hipoteza.

Prihvaćajući nultu hipotezu prihvaća se (uz danu razinu značajnosti) pretpostavka da je proporcija kućanstava koja posjeduju perilicu za rublje 0.75 (odnosno 75 % ). I ovaj rezultat valja prosuđivati imajući u vidu pogreške koje se pojavljuju u postupcima ispitivanja pretpostavki s pomoću uzorka.

## PITANJA I ZADATCI

**1.** Opišite korake ispitivanja hipoteze o pretpostavljenoj proporciji osnovnog skupa s pomoću velikoga jednostavnoga slučajnog uzorka.

**2.** (a) Kako se naziva interval $p_0 - z\sigma_{\hat{p}} < p < p_0 + z\sigma_{\hat{p}}$? Objasnite značenje svakog simbola. Na koji se način donosi odluka u postupku testiranja s pomoću navedenog intervala? (b) Dan je sljedeći omjer: $z^* = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$. Objasnite značenje svakog simbola u navedenom izrazu. Kako se donosi odluka s pomoću navedenog omjera?

**4.** Usporedite intervale: (1) $p_0 - z\sigma_{\hat{p}} < \hat{p} < p_0 + z\sigma_{\hat{p}}$ i (2) $\hat{p} - z\sigma_{\hat{p}} < p < \hat{p} + z\sigma_{\hat{p}}$. Objasnite njihove sadržaje, razlike i sličnosti.

**5.** Dane su ove veličine: *N* = 7 892, *n* = 235, *m* = 115. Testirajte ove hipoteze: $H_0 .... p = p_0; H_1 .... p \neq p_0, p_0 = 0.5$. Razina značajnosti je 5 %. Objasnite odluku. U postupku se koristite obama načinima donošenja odluke o ishodu testa.

**6.** Pretpostavlja se da je postotak pušača među učenicima IV. razreda srednjih škola 25 %. Radi ispitivanja navedene pretpostavke izabran je slučajni uzorak 585 učenika, među kojima je 140 pušača. U uzorak je izabrano manje od 5 % učenika. Može li se prihvatiti navedena pretpostavka na razini 10 % značajnosti? (Uputa: pri računanju standardne pogreške proporcije zanemarite faktor $\left(\dfrac{N-n}{N-1}\right)$).

**7.** Iz jedne pošiljke 10 000 otvarača limenki u uzorak je izabrano 2 % otvarača. U uzorku je pronađeno osam neispravnih otvarača. Hoće kupac li prihvatiti pošiljku ako je ugovorom dopušten škart od 5 %? Razina značajnosti je 5 %.

**8.** Protumačite sljedeći ispis obradbe programskom potporom.

**Test o pretpostavljenoj proporciji populacije (veliki uzorci)**

**Varijabla : potrošači**

| | |
|---|---|
| **Veličina uzorka.................................................** | **535** |
| **Broj elemenata s danim modalitetom...............** | **302** |
| **Empirijska proporcija........................................** | **0.5650** |
| **Standardna pogreška proporcije.....................** | **0.02160** |
| **Teorijska vrijednost normalne distribucije (*z*)...** | **1.960** |
| **Razina signifikantnosti.......................................** | **0.05000** |
| **Empirijski *z*-omjer...............................................** | **2.083** |
| **Empirijska razina signifikantnosti (*p*)................** | **0.03722** |

**H i p o t e z e :**

**Nulta hipoteza............... Proporcija uzorka = 0.5200**

**Alternativna hipoteza... Proporcija uzorka <> 0.5200**

**O d l u k a - Na dvije granice:**

**Na danoj razini signifikantnosti prihvaća se alternativna hipoteza.**