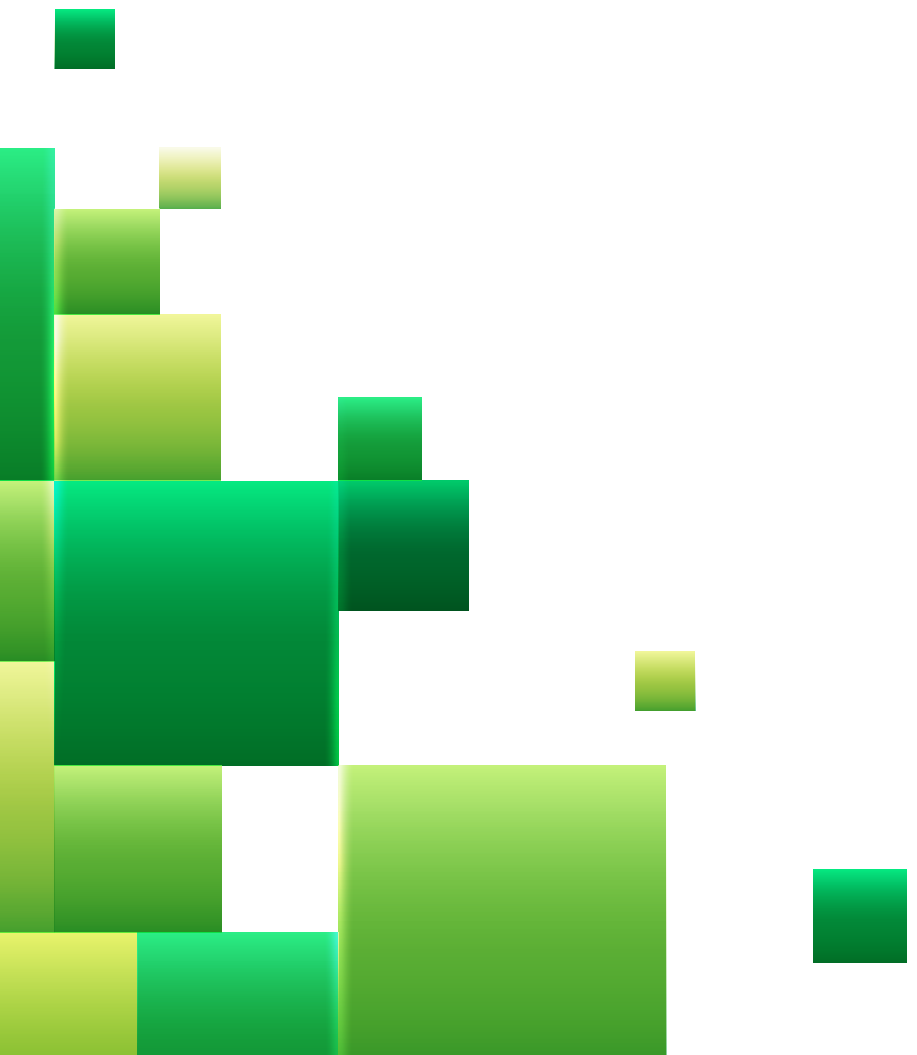
Zoran Maunaga, Vojislav Dukić, Srđan Bilić

ŠUMARSKA BIOMETRIKA



УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
UNIVERSITY OF BANJA LUKA
ШУМАРСКИ ФАКУЛТЕТ
FACULTY OF FORESTRY

Zoran Maunaga, Vojislav Dukić, Srđan Bilić

ŠUMARSKA BIOMETRIKA

POMOĆNI UDŽBENIK

Banja Luka, 2025.

Naslov: Šumarska biometrika

Autori: Zoran Maunaga, Vojislav Dukić, Srđan Bilić

Izdavač: Šumarski fakultet, Univerzitet u Banjoj Luci

Tehnička priprema: Branka Ružičić

Slike: Srđan Bilić

Ilustracije: Ognjen Stojnić

Štampa: Mikro print

Tiraž: 100 primjeraka

CIP - Каталогизација у публикацији
Народна и универзитетска библиотека
Републике Српске, Бања Лука

630:519.2(075.8)

МАУНАГА, Зоран, 1958-
Šumarska biometrika / Zoran Maunaga, Vojislav Dukić, Srđan Bilić. - Banja Luka : Šumarski fakultet, 2025 (Banja Luka : Mikroprint). - 175 стр. : илустр. ; 24 cm

Тираж 100. - Библиографија: стр. 175.

ISBN 978-99938-56-60-3

COBISS.RS-ID 143206913

Naučno-nastavno vijeće Šumarskog fakulteta u Banjoj Luci, na sjednici 20. juna 2025. godine, dalo je saglasnost da se rukopis „Šumarska biometrika“, čiji su autori Zoran Maunaga, Vojislav Dukić i Srđan Bilić, objavi kao pomoćni udžbenik.

PREDGOVOR

Postojeći udžbenik Šumarska statistika (Koprivica, 2015) odgovara nastavnom programu, ali sadrži materiju viših kurseva, izmiješanu sa programom za prvi ciklus. To ovaj udžbenik čini preobimnim i preteškim za studente. Osim toga, u njemu nedostaju neki novi pojmovi, na primjer *boxplot* i *p-vrijednost*. Činjenica je da studenti statistiku mogu pronaći i u nekoj drugoj literaturi, kao i na internetu. Međutim, sve te "statistike" ne odgovaraju studentima šumarskog fakulteta. One su namijenjene različitim oblastima i pisane su svaka "na svoj način". U njima se šumarstvo najčešće i ne spominje kao polje moguće primjene statistike. To su samo neki od razloga za nastanak ove knjige.

Cilj nam je bio da pojmove i statističke postupke objasnimo na jednostavan način, prilagođen studentima koji uče zanat šumarskog inženjera. U knjizi ima nekih novina, ali i razlika u odnosu na postojeći udžbenik. Mi smo nastojali jasnije razdvojiti skup i uzorak. Regresionu analizu prebacili smo na kraj i dali joj veći značaj u odnosu na korelaciju.

Znamo da studenti ne vole preduge i guste tekstove. Dosljedno smo koristili kratke rečenice, uobičajene izraze i simbole, kao i jednostavne grafičke prikaze. Često smo se oslanjali na osnovnu knjigu, obično navodeći broj strane za detaljnija objašnjenja. Iz Knjige smo preuzimali formule i grafikone. Koristili smo iste, ali i nove primjere iz šumarstva, postepeno uvodeći šumarsku terminologiju.

Pokušali smo zainteresovati čitaoca za ono što ga očekuje. Unijeli smo u knjigu neke citate, koji su nam se učinili zanimljivim i vrlo korisnim, kao i par originalnih ilustracija. Govorili smo o šumi, kao masovnoj pojavi (statističkoj masi), dajući pojmu „šuma" opšte (šire) značenje. Ukazivali smo, manjim fontom, na različite pristupe u literaturi i praktične aspekte problema.

Za studente je posebno važna povezanost Šumarske biometrike sa stručnim predmetima (Dendrometrija, Prirast šuma, Uređivanje šuma) i Metodologijom istraživačkog rada (izradom stručnih, seminarskih i završnih radova). Istraživanje (nauka) u šumarstvu, kao i premjer (inventura) šuma u praksi, ne mogu se ni zamisliti bez statističkih metoda.

Knjiga ima tri glavna poglavlja:

- Opšti dio,
- Deskriptivna statistika i
- Inferencijalna statistika.

U poglavlju I (Opšti dio) proći ćemo kroz materiju u vidu jednog sažetka, bolje reći najave onoga što nas očekuje. Nakon uvoda, objasnićemo osnovne pojmove (statistički skup i druge), a potom navesti važnije statističke metode. Opštem (uvodnom) dijelu posvetili smo posebnu pažnju.

U poglavlju II (Deskriptivna statistika) vidjećemo kako statistika, pomoću različitih parametara, numerički opisuje masovne pojave. Materija se odnosi na stvarne pojave sa konkretnim podacima (stvarni rasporedi frekvencija). Drugim riječima, u ovom poglavlju bavimo se statističkim skupovima.

U poglavlju III (Inferencijalna statistika) čeka nas složeniji (teži) dio statistike, a to je statističko zaključivanje o masovnim pojavama na bazi uzorka. Jedan dio se odnosi na procjenjivanje parametara masovne pojave (uzorci), drugi dio na statističko testiranje parametara (testovi) i treći na utvrđivanje zavisnosti jedne pojave od drugih pojava (regresija). Na početku ovog poglavlja obradili smo teorijske rasporede frekvencija.

Nadamo se da će ova knjiga, osim u nastavi, pomoći studentima da uvide koristi od statističkih metoda, ali i „opasnosti“ u njihovoj primjeni.

Septembar, 2025. godine

A u t o r i

SADRŽAJ

I. OPŠTI DIO

II. DESKRIPTIVNA STATISTIKA

III. INFERENCIJALNA STATISTIKA

I. OPŠTI DIO

Karl Pirson
(1857-1936)
Osnivač matematičke statistike

1. UVOD

„Ne želimo da vidimo pojedinačno drveće, već čitavu šumu“ (4)

1.1. Istorijski uvod

Tek početkom XIX vijeka dolazi do spoznaje da se varijacije javljaju, ne samo kod društvenih već i kod prirodnih pojava, čime statistika postaje opšti naučni metod. Prethodno je dva puna vijeka vladalo mišljenje da je statistika disciplina nauke o društvu. Poznate su dvije škole statistike iz tog vremena. To su njemačka i engleska škola. S ciljem vođenja državne politike njemački profesori su sve podatke o državi (teritoriji, stanovništvu, privredi, finansijama) sistematizovano prikazivali u tzv. Državopisu. Tada je nastao naziv „statistika“ (od ital. *stato* = država). Dao ga je profesor Ahenval (Gottfried Achenvall) 1748. godine. U isto vrijeme u Londonu službe administracije, prateći registre umrlih, došle su do tzv. „*zakona smrtnosti*“, po kome nema promjena u strukturi uzroka smrtnosti (bolesti, nesreće, samoubistva, ...). U novije vrijeme, razvojem teorije vjerovatnoće, put moderne statistike trasirali su Karl Pirson i Ronald Fišer. Oni se smatraju najpoznatijim statističarima.

U šumarstvu Evrope statistika je primijenjena prvo u Skandinavskim zemljama, gdje su početkom XX vijeka izvedene nacionalne inventure šuma. U BiH, nakon Drugog svjetskog rata, izvršena su obimna istraživanja šuma (1952-1958), kao i nacionalna inventura šuma (1964-1968). Prvo uređivanje šuma na statističkim principima provedeno je 1962. godine.

1.2. Čime se bavi statistika?

Najveća korist od nauke je što omogućuje predviđanje (planiranje). Da bismo predviđali pojave, moramo znati zakone (pravila) po kojima se one dešavaju. Te zakone otkrivamo posmatranjem jedinica pojave u velikom broju (masi slučajeva), a ne pojedinačno. Takve pojave nazivamo masovne pojave (skupovi). One su u prirodi česte.

Statistika je nauka o masovnim pojavama. Pomoću statističkih parametara (aritmetičke sredine i drugih), statistika kvantitativno opisuje pojave, istražuje varijabilitet i utvrđuje zakonitosti u ponašanju pojava. Statistika se u biologiji naziva *biometrika* (lat. *bios* – život, *metrein* – mjeriti), a za statističke metode u oblasti šumarstva najbolje odgovara naziv š*umarska biometrika*.

Statističko istraživanje ima dva dijela, odnosno dvije etape. To su *deskriptivna statistika* (posmatranje i opis pojava) i *inferencijalna statistika* (statistička analiza pojava i zaključivanje). Deskriptivna statistika opisuje pojavu na osnovu svih njenih individualnih slučajeva ispoljavanja. Ona se sastoji od prikupljanja podataka o svakom pojedinom slučaju (jedinici), grupisanju jedinica prema nekom obilježju, tabelarnom i grafičkom prikazu statističkih

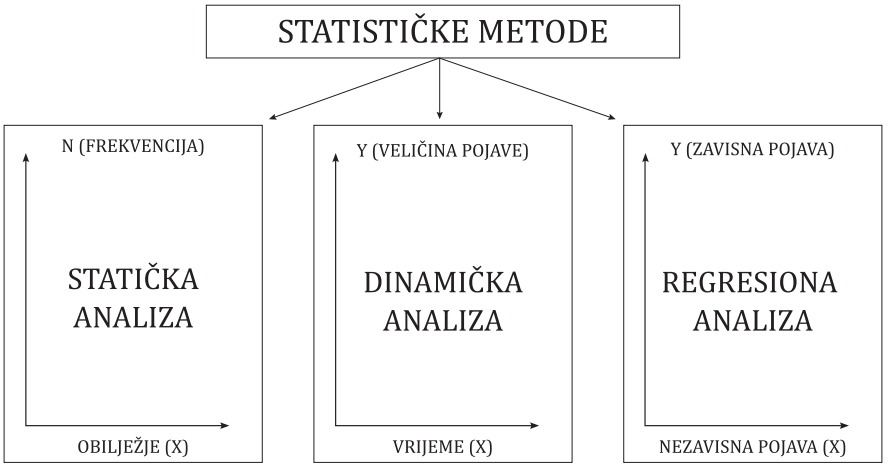
serija i izračunavanju sumarnih statističkih parametara (srednjih vrijednosti, mjera varijabiliteta i drugih). Statistička analiza obuhvata različite metode. O njima ćemo govoriti u nastavku.

Statistika

Statistika opisuje pojave (deskriptivna statistika) i istražuje pojave (inferencijalna statistika).

Statističke metode

Sve statističke metode svrstavaju se u tri grupe: statičke, dinamičke i regresione *(Slika 1)*. U *statičkoj analizi* na apscisi se nalazi obilježje, a na ordinati frekvencija; u *dinamičkoj analizi* na apscisi je vrijeme, a na ordinati veličina pojave; dok u *regresionoj analizi* imamo dvije pojave u koordinatnom sistemu.



Slika 1. Tri grupe statističkih metoda

Statičko ispitivanje ne uzima vrijeme kao faktor, tj. tada ne utvrđujemo promjene pojave tokom vremena, već utvrđujemo njeno stanje (strukturu mase). Na primjer, pri redovnom uređivanju šuma prikazujemo stanje šuma u doba uređivanja po više obilježja: po vrstama drveća, debljini stabala, kvalitetu stabala i drugim karakteristikama. Vrijeme služi samo za definisanje mase. Poređenje stanja šuma sa prethodnim uređajnim periodima već ima karakter dinamičke analize.

U nastavku (sitnim fontom) prenosimo originalan tekst iz udžbenika *Osnovi statističke analize* (6, str. 23), koji „na svoj način“ govori o navedenim statističkim metodama.

Statistička analiza obuhvata mnogobrojne i vrlo različne probleme. Ako bismo hteli da te probleme jasnije sagledamo, morali bismo ih grupisati na neki način. Uzimajući vreme kao faktor promena za merilo, mogli bismo probleme statističke analize podeliti u dve velike grupe: prvo, istraživanje strukture pojave u jednom momentu ili periodu, pri čemu se vreme ne uzima kao faktor, i istraživanje razvitka pojave tokom vremena, pri čemu se promene pojave ispituju kao funkcije vremena. Prva grupa istraživanja obeležava se kao statička a druga kao dinamička.

Varijacija kao karakteristika masovnih pojava već sama po sebi implicira kretanje. Ali varijacija kod izvjesnih, naročito prirodnih pojava, zbiva se oko istog proseka, dok kod drugih, naročito društvenih pojava, dolazi vremenom i do menjanja proseka, što znači do pomeranja u strukturi mase. U statističkom smislu samo se ovi procesi, koji izražavaju razvoj, označuju kao dinamički. Prema tome, za upoznavanje većine prirodnih pojava dovoljna je upotreba statičkih metoda istraživanja.

Dvojna podela na statičke i dinamičke probleme iscrpljuje sve probleme statističke analize. Metodima statičke analize možemo ispitivati ne samo strukturu pojave u datom vremenu nego i međuzavisnost karakteristika te strukture, kao i njihovu uslovljenost raznim faktorima. Dinamički možemo ispitivati ne samo tendenciju razvitka pojave tokom vremena nego i naporedne promene raznih faktora koji utiču na takav razvitak. Pa ipak, ispitivanje veza između raznih pojava, bilo statičkim ili dinamičkim metodima, predstavlja jedan složeniji problem koji, i zbog njegove važnosti za tumačenje pojava, treba posebno izučavati. Zbog toga ćemo probleme statističke analize obuhvatiti u tri grupe: statički problemi, dinamički problemi i problemi korelacije ili regresije.

Kada je u pitanju nastavni program prvog ciklusa studija upoznaćemo se sa svim statističkim metodama, ali će težište svakako biti na statičkoj analizi. U dinamičkoj analizi obradićemo trend. Treća grupa metoda, poznata kao regresiona analiza, za nas je važna, jer je primijenjena u istraživanju zalihe i prirasta šuma, odnosno prilikom izrade odgovarajućih tablica u šumarstvu.

1.3. Osnovni principi statistike

Statistika se može podijeliti na dva dijela. Prvi dio, *teorijska statistika* (otkrivanje novih metoda), nije predmet našeg interesovanja (zanimanja). Drugi dio je *praktična (primijenjena) statistika*. Nju posmatramo na dva načina: kao *metod* istraživanja (postupci i radnje) i kao *rezultat* istraživanja (parametri skupa, statistika uzorka). Statistika je alat, bez kojeg nema istraživačkog (naučnog), ali ni stručnog (praktičnog) rada. Za njenu primjenu neophodno je poznavanje oblasti u kojoj se primjenjuje.

Statističko istraživanje je *masovnog karaktera*, jer obuhvata sve jedinice (slučajeve) mase (a ne samo jedan ili njih nekoliko). Osim toga, ono je i *kvantitativnog karaktera*, jer značaj karakteristika (obilježja) pojave određuje prema njihovoj frekvenciji (brojnom stanju) i dalje ih matematički obrađuje. Na primjer, u šumi obuhvatamo sva stabla po debljini (masovni karakter), a značaj tankih stabala određen je njihovim brojem, tj. učešćem u ukupnom broju (kvantitativni karakter).

Induktivni način rada

Prvi cilj statističkih istraživanja je prikaz obima i strukture masovnih pojava. U opisivanju takvih pojava polazi se od pojedinačnih elemenata i na osnovu njih donosi zaključak o cijeloj pojavi. Samo u velikom broju slučajeva (u masi) možemo otkriti ono što je zajedničko (opšte, generalno) za neku pojavu. Utvrđena zakonitost važi za masu, ali ne i za individue iz te mase. Statistika se, dakle, zasniva na principu *indukcije*. Iz pojedinačnih informacija (podataka) induktivnim načinom dolazi se do opštih (generalnih) karakteristika pojave. Zaključivanje se oslanja na zakon velikih brojeva.

Zakon velikih brojeva (ZVB)

Navedimo jednu opštepoznatu statističku zakonitost. Na svjetskom nivou rađa se oko 50% muške i 50% ženske djece (po literaturi 106:100). Ova statistička zakonitost ne važi za individualne slučajeve. U nekim višečlanim porodicama sva rođena djeca mogu biti istog pola. Zakon važi samo u masi slučajeva i vjerodostojniji je što je broj posmatranih slučajeva veći (ovdje, što više rođene djece posmatramo bliži smo navedenom odnosu). Ova zakonitost poznata je u statistici kao *Zakon velikih brojeva (ZVB)*. Zakon može da glasi ovako: slučajni događaj u masi (velikom broju) slučajeva postaje zakon.

ZVB se zasniva na pretpostavci da u velikom broju slučajeva ili ponavljanja *srednja vrijednost* prestaje da bude slučajna veličina. Ona, na kraju, postaje jedna fiksna (matematička) vrijednost. Prava aritmetička sredina skupa bi bila poznata tek onda kad bismo uzeli u obzir sve elemente ispitivane pojave. U praksi, uzimanjem uzorka, te prave (stvarne) vrijednosti skupa predviđamo s velikom pouzdanošću.

Rad sa uzorcima

Statistika uglavnom radi sa uzorcima. Ako izmjerimo visine određenog broja stabala, njihova srednja visina odnosi se samo na ta stabla, tj. ona ne važi za cijelu šumu. Međutim, statistika ima način da, uz odgovarajuću pouzdanost i preciznost, procijeni srednju visinu svih stabala u šumi. Ako smo, takođe uzorkom, u nekoj drugoj šumi dobili prosječnu visinu stabala veću za dva metra, postavlja se pitanje da li ista tolika razlika postoji između svih stabala ove dvije šume. Statistika omogućuje i takvo zaključivanje uz veliku sigurnost. Uobičajena vjerovatnoća koja se koristi u šumarstvu iznosi 95%.

Napomena: Statistika može obraditi netačne (izmišljene) podatke. Takvi podaci u šumarstvu, koje prikupljamo na terenu, često su problem. Da bismo to dovoljno jako istakli, stavili smo na kraju knjige ilustraciju: škart ulazi – škart izlazi. Treba da znamo da statistika uvijek polazi od pretpostavke da su podaci koji se obrađuju tačni.

"Pogibija jednog čovjeka je tragedija,
a hiljade ljudi statistika"

Erik Marija Remark (1898 – 1970)

2. STATISTIČKI SKUP

2.1. Šta je masovna pojava?

Statistika se, kako već rekosmo, bavi masovnim pojavama. Kakve su to pojave? Obično kad se nešto pojavljuje (ispoljava) u velikom broju (masovno) kažemo da se radi o masovnoj pojavi. Na primjer, u šumi govorimo o masovnoj pojavi potkornjaka (vrsta šumske štetočine) ili u društvu o masovnoj pojavi oboljelih od nekog virusa.

Masovne pojave često nemaju granicu i zahvataju širi prostor. Na primjer, šuma ili stanovništvo pokrivaju veliki dio naše planete. Često nije moguće ili nismo zainteresovani za istraživanje ovakvih masovnih pojava na cijeloj njihovoj površini. Šta je statistički skup i da li ima razlike između masovne pojave i statističkog skupa?

2.2. Definisanje statističkog skupa

Ako želimo masovnu pojavu posmatrati statistički, tj. kao statističku masu – skup podataka, onda to zahtijeva njeno prostorno, vremensko i pojmovno definisanje. Tako dolazimo do pojma *statistički skup* ili *populacija*. Naziv *skup* potiče otuda što se radi o skupovima (skupinama) jedinica (podataka), dok je naziv *populacija* vezan za biološki (živi) sistem jedinki neke vrste. Skup se definiše u skladu sa svrhom (ciljem) istraživanja. Termini u upotrebi su: statistički skup, osnovni skup, skup, populacija, statistička masa.

Statistički skup

Masovna pojava definisana prostorno, vremenski i pojmovno predstavlja statistički skup.

Prostorni aspekt

Masovnu pojavu čine sve njene jedinice, dok statistički skup obuhvata samo one jedinice masovne pojave koje se nalaze unutar svoje (definisane) granice. Na primjer, dok masovnu pojavu predstavlja cijela šuma, dotle skup može biti neki dio te šume, npr. šumskoprivredno područje (ŠPP) ili samo jedan odjel.

Definisanje prostornog okvira uslov je za statističko posmatranje masovne pojave kao statističkog skupa. Od toga kako smo definisali skup zavisi broj elemenata skupa. Znači, skup ne mora uvijek imati jako veliki broj jedinica (elemenata).

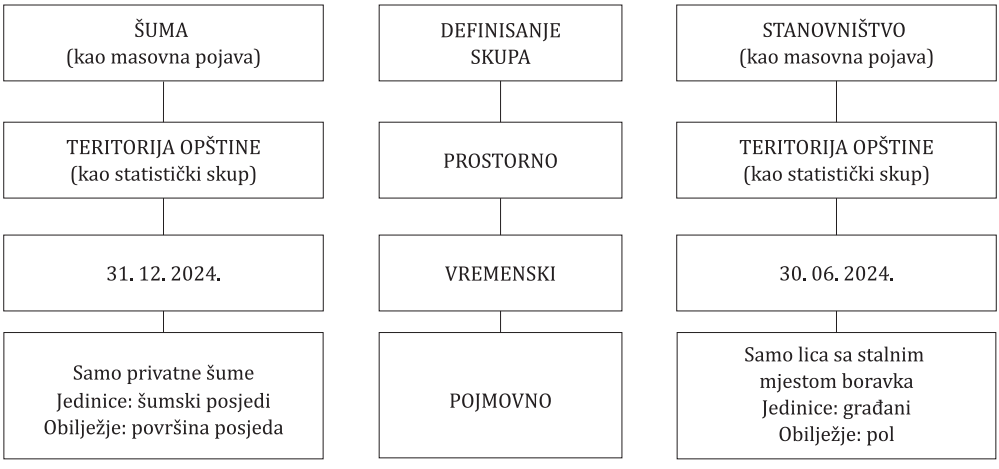
Vremenski aspekt

Osim prostorno, skup definišemo vremenski. Uvijek se mora navesti vrijeme posmatranja pojave. Ako smo utvrdili stanje šuma mora se navesti kalendarsko vrijeme za koje to stanje važi, a ako se radi o proizvodnji šumskih drvnih sortimenata (skraćeno ŠDS) potrebno je navesti vremenski interval na koji se ta proizvodnja odnosi (godina, 10 godina).

Pojmovni aspekt

Skup još treba definisati pojmovno (sadržajno). Ovdje se opisuju jedinice skupa i definiše obilježje. Na primjer, ako je u pitanju mjerenje prečnika stabala mora biti naglašeno: koja stabla dolaze u obzir, da li sve vrste drveća, da li samo živa stabla, koliko iznosi granični prečnik (taksacioni prag), način mjerenja (npr. jedan prečnik s gornje strane) i drugo. Ovaj opis daje se u skladu s ciljem istraživanja. Često je vrlo opširan, ponekad u vidu uputstva ili metodike rada.

Navešćemo, skraćeno, dva primjera definisanja skupa, jedan šumarski, a drugi iz društvenih oblasti *(Slika 2)*. U prvom primjeru zanima nas struktura privatnih šuma po veličini posjeda na nekoj opštini. Cilj može biti rješavanje problema ustitnjenosti šumskih posjeda u gazdovanju privatnim šumama, koji nastaje zbog različitih interesa šumovlasnika. U drugom primjeru posmatrana je polna struktura stanovništva, gdje skup čine sva lica sa stalnim mjestom boravka u nekoj opštini.



Slika 2. Definisanje statističkog skupa

2.3. Karakteristike statističkog skupa

Statistički skup čine jedinice, ali ne bilo kakve jedinice. Da bi skup imao smisla moraju biti ispunjena dva uslova. Prvo, jedinice moraju biti istovrsne i drugo, sve jedinice moraju imati neko zajedničko varijabilno obilježje.

Karakteristike jedinica skupa

Skup čine istovrsne (jednorodne) jedinice, koje su međusobno različite (varijabilne) po nekom obilježju.

Istovrsnost jedinica

Ovo je osnovna, posve logična, odrednica statističkog skupa. Istovrsnost znači da skup čine samo oni podaci koji se mogu sabirati, tj. za koje se može izračunati aritmetička sredina. U školi nas prvo uče da se kruške i jabuke ne mogu sabirati. Ili, primjer kupusa i mesa, figurativno. Ako ljudi jedu različito, jedni samo kupus, a drugi samo meso, to ne znači da oni u prosjeku jedu sarmu.

Istovrsnost se naglašava (precizira) kod pojmovnog definisanja skupa. Stabla u šumi kao elemente čini istovrsnim to što pripadaju šumskim vrstama drveća. U zavisnosti od cilja istraživanja, istovrsnost se može odnositi (ograničiti) na samo jednu vrstu drveća. Na primjer, kod izrade zapreminskih tablica bukvu treba posmatrati odvojeno od jele, jer se ove dvije vrste drveća razlikuju po obliku debla. U tom slučaju to su dva statistička skupa, pa ćemo imati posebne tablice – jedne za bukvu, a druge za jelu.

Varijabilnost obilježja

Masovne pojave imaju više različitih karakteristika. Na primjer, studenti se međusobno razlikuju po polu, mjestu rođenja, opštem znanju itd. Ovo su različita obilježja. U statistici obično posmatramo samo jedno obilježje (oznaka „X"). S obzirom na to da obilježje „X", koje se nalazi na apscisi (po tome je i dobilo oznaku), ima različite vrijednosti, govorimo o njegovoj varijabilnosti (promjenljivosti).

Zapravo, fokus statistike je na varijabilitetu. Proizvodnja cigle u nekoj fabrici, gdje su sve cigle iste (identične), nije predmet statistike. Težina samo jedne cigle je dovoljna da znamo težinu svih ostalih (tu se praktično nema šta računati). Statistika se bavi samo pojavama sa varijabilnim obilježjima. Stabla u šumi se diferenciraju po prečniku, pa kažemo da je prečnik varijabilan (promjenljiv, tj. različit od stabla do stabla). Upravo ta varijabilnost jedinica omogućava da se analizira kvalitativna struktura, a ne samo obim (veličina) pojave.

Za varijabilnost računamo različite mjere. Na osnovu njih donosimo zaključak o homogenosti skupa. Ako skup ima malu varijabilnost kažemo da je

homogen, dok je u suprotnom *heterogen*. Statistika ima različit pristup (metode) u ovim slučajevima. Zato je homogenost važna.

Homogenost, odnosno heterogenost u statistici ima specifično značenje. S obzirom na to da se skup sastoji od istovrsnih elemenata, ne možemo govoriti o homogenosti (heterogenosti) prema opštem značenju ovih pojmova (heterogen grč. znači „različita roda"). Po ovome, svi skupovi bili bi homogeni, jer su po definiciji sastavljeni od istorodnih jedinica.

Posmatrajmo skup čije su jedinice jabuke, a obilježje njihova težina. Ako su razlike u težini jabuka male skup je homogen, a ako su velike onda je heterogen. Mjera varijabiliteta je koeficijent varijacije. Ako izmiješamo kruške sa jabukama to više nije statistički skup, jer jedinice više ne pripadaju istoj vrsti. Dakle, moramo razlikovati raznovrsnost od heterogenosti. *Raznovrsnost* govori o različitim vrstama jedinica (jabuke i kruške), a *heterogenost* o razlikama unutar jedinica iste vrste (razlike među jabukama).

2.4. Jedinice (elementi) skupa

Svaka masovna pojava sastavljena je od jedinki, odnosno pojedinačnih elemenata. Njih nazivamo *jedinice* ili *elementi skupa*. To mogu biti ljudi (npr. zaposleni u šumarstvu), životinjska bića (divljač po lovištima), biljne vrste (stabla neke šume), predmeti (mehanizacija u šumskim gazdinstvima) ili događaji (šumski požari). Između termina jedinice skupa i elementi skupa nema suštinske razlike. Prvi termin je bliži upotrebi u prirodi, a drugi kasnije u fazi obrade. *Napomena*: u šumi, često iz praktičnih razloga, uzimamo određene površine kao vještačke jedinice skupa.

Jedinice (elementi) skupa

Jedinični (elementarni) dijelovi skupa nazivaju se jedinice (elementi) skupa.

Elementi izvorno (sami po sebi) iako čine skup oni ga ne predstavljaju. Da bismo formirali skup, odnosno statistički niz podataka, uzimamo vrijednosti (modalitete) posmatranog obilježja svih elemenata skupa. Na primjer, ako je prečnik nekog izmjerenog stabla 39 cm, u daljem statističkom radu, ova vrijednost će predstavljati taj element skupa. Tako će svako stablo (svaki element) imati svoju vrijednost obilježja. *Podvlačimo*: vrijednosti (modaliteti) obilježja nisu elementi skupa.

Šuma je specifična kao masovna pojava, pa tako i šumarstvo kao oblast. Šumu statistički posmatramo na dva načina. Prvo, kao skup stabala, gdje su prirodne jedinice stabla i drugo, kao skup malih elementarnih površina, gdje su one vještački elementi skupa. U statističkom radu, u šumi prvo biramo površine, a onda na tim površinama mjerimo stabla. O ovome više u poglavlju *Statistički uzorci*.

2.5. Statističko obilježje

Pojam obilježja

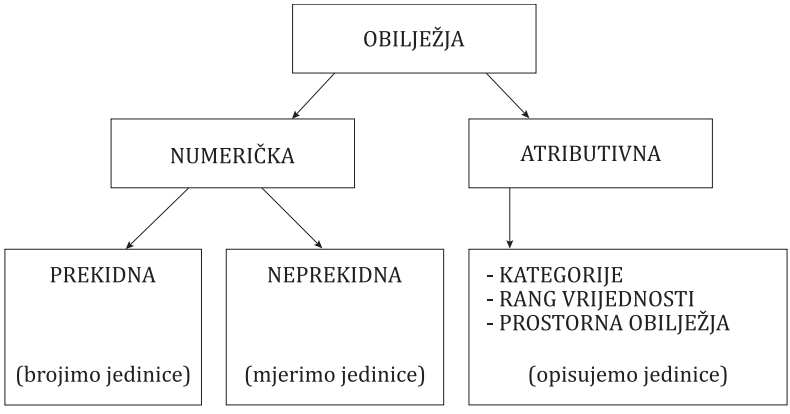
Masovne pojave (statistički skupovi) uvijek imaju više karakteristika, koje se ispoljavaju na njihovim elementima. U statistici te karakteristike (svojstva, osobine, odlike) elemenata skupa nazivamo *obilježjima*. Naziv obilježje potiče od toga što te karakteristike obilježavaju skup, tj. važne su za njegov opis. U statističkom radu posmatramo obično ona obilježja koja su važna (interesantna). Pojavni oblici (vrijednosti, modaliteti) obilježja su promjenljivi, varijabilni od jedne do druge jedinice, pa se obilježje još naziva *promjenljiva* ili *varijabla*. Promjenljiva je više domaći izraz, a varijabla strani. Ovi termini više se koriste kasnije u drugom dijelu statistike. Za vrijednosti (modalitete) obilježja *X* koriste se i izrazi: rezultati, podaci, opažanja, opservacije i drugi.

Statističko obilježje

Zajednička (opšta) karakteristika svih jedinica skupa, koja na neki način obilježava skup, naziva se obilježje (promjenljiva, varijabla).

Vrste obilježja

Prema načinu izražavanja, postoje dvije grupe obilježja: numerička i atributivna *(Slika 3)*. Za nas su manje važne druge podjele obilježja, kao što su po svojoj prirodi (biološka, ekonomska, socijalna obilježja itd) ili prema skalama kojima se mjere (nominalna i druga).



Slika 3. Podjela obilježja

Numerička obilježja

Numerička (brojčana) obilježja kvantitet (veličinu) pojave iskazuju brojevima. Te brojeve nazivamo *vrijednosti* obilježja. Numerička obilježja mogu biti:

- prekidna i
- neprekidna.

Prekidna obilježja. Veličina pojave sa prekidnim obilježjem iskazuje se cijelim brojevima. Statistički niz se dobija brojanjem jedinica. Na primjer: broj djece u porodici, broj radnih jedinica (šumskih radilišta) ili broj šumskih požara. Broj stabala po hektaru takođe predstavlja prekidno obilježje.

Neprekidna obilježja. Podaci za skup sa neprekidnim obilježjem dobijaju se mjerenjem. Na primjer: prečnik stabala u centimetrima, visina stabala u metrima ili zapremina drvne mase po hektaru u kubnim metrima. Iskazuju se cijelim brojevima ili na decimale.

Neprekidna obilježja, unutar svog intervala variranja, mogu imati bilo koju vrijednost, jer ih ima teorijski beskonačno mnogo. Kod ovih obilježja nema prekida, što grafički izgleda kao kontinuelna linija. Između susjednih vrijednosti uvijek se može ubaciti neka nova vrijednost (ona se povećavaju za beskonačno male iznose). *Napomena*: za neprekidna obilježja, teorijski gledano, treba uzimati intervale umjesto pojedinačnih vrijednosti, jer je vjerovatnoća pojedinačnih vrijednosti jednaka nuli.

Atributivna obilježja

Atributivna (opisna) obilježja kvalitet pojave iskazuju opisno (riječima). Ova obilježja javljaju se u različitim oblicima, koje nazivamo modaliteti (za razliku od vrijednosti kod numeričkih obilježja). U atributivna obilježja spadaju obilježja kao što su: kategorijska obilježja (npr. kategorije šuma), rang vrijednosti (npr. bonitet staništa) i prostorna obilježja (npr. opštine, kao geografske jedinice). Mogućnosti statistike kod atributivnih obilježja su mnogo manje nego kod numeričkih, jer nisu pogodna za matematičke operacije.

2.6. Taksacioni elementi šume kao statistička obilježja

Karakteristike koje mjerimo u šumi, bilo na pojedinačnim stablima ili po hektaru, obično nazivamo *taksacioni elementi*. U statistici su to statistička obilježja. Najvažniji taksacioni elementi (obilježja) stabla su: prečnik, visina, zapremina i zapreminski prirast. Na osnovu prečnika utvrđujemo osnovne karakteristike šume, kao što su debljinska struktura i zaliha drvne mase.

Iako je istraživanje pojedinačnih stabala važno, u šumarstvu se fokus stavlja na šumsku proizvodnju (taksacione elemente) po hektaru. Upravo je to pogodnost za primjenu statistike. Glavni taksacioni elementi šume (u literaturi se obično govori sastojine) su: prečnik, visina, temeljnica i njen prirast, zapremina i njen prirast, te broj stabala. Ovi elementi se statistički međusobno suštinski razlikuju. Svrstavamo ih u tri grupe:

- srednje vrijednosti,
- agregati i
- specifični elementi.

Srednje vrijednosti

Kao srednje vrijednosti računaju se *srednji prečnik* i *srednja visina*. One predstavljaju sva stabla, odnosno šumu u cjelini. Razlog je jednostavan. Bilo bi besmisleno sabirati prečnike, odnosno visine svih stabala i iskazivati taj zbir po hektaru. Srednji prečnik i srednja visina računaju se uglavnom za šume u kojima su stabla iste starosti (jednodobne sastojine).

Agregati

Kao agregati (zbirne vrijednosti svih stabala po hektaru) računaju se *temeljnica* i *zapremina*, te prirast temeljnice i prirast zapremine. Veća važnost daje se zapremini i prirastu drvne mase po hektaru. Statistička obrada agregata ne odvija se direktno preko stabala (prirodnih jedinica skupa), već se koriste površine kao vještački elementi skupa.

Specifični elementi

Prilikom premjera, šumu dijelimo na manje površine, koje tako smatramo elementima skupa. Na tim površinama mjerimo stabla i utvrđujemo taksacione elemente. Jedan od njih je *broj stabala po hektaru*. Svrstavamo ga u specifične elemente, jer se dobija izbrajanjem (prebrojavanjem) stabala, a ne sumiranjem pojedinačnih vrijednosti, kao što je to kod agregata (na primjer zapremine).

Ako posmatramo jednu šumu kao skup, gdje su jedinice skupa stabla, postavlja se pitanje da li je u ovom slučaju broj stabala statističko obilježje? Između stabala nema razlike po broju, svako stablo je samo jedna jedinka - jedinica (stablo od 15 cm predstavlja jedno stablo, isto kao i stablo od 70 cm). U ovom slučaju broj stabala ne može biti obilježje, jer nije ispunjen drugi uslov, a to je varijabilnost obilježja.

Jedan od specifičnih taksacionih elemenata (statističkih obilježja) šume je *procent prirasta*. Dobija se iz odnosa dva agregata (prirasta zapremine i zapremine). Procent prirasta je računska veličina. Izražava se u procentima.

2.7. Podjela statističkih skupova

Statističke skupove dijelimo po više osnova:

- po uređenosti,
- po broju elemenata i
- po broju obilježja.

Skupovi po uređenosti

Po uređenosti skupovi mogu biti *neuređeni* i *uređeni*. Podaci prikazani onako kako su prikupljeni, bez ikakvog reda, čine neuređen skup. Sređivanjem, tj. grupisanjem podataka dolazimo do uređenog skupa. Taj postupak se naziva *uređivanje statističkih skupova*.

Skupovi po broju elemenata

S obzirom na broj elemenata, skupovi mogu biti *konačni* i *beskonačni*. Broj stabala u jednom odjelu šume ili broj građana u jednom kvartu predstavljaju konačne skupove. Podaci nekog mjerenja ili potomstvo neke vrste predstavljaju beskonačne skupove. Beskonačni skupovi imaju veliki značaj u statističkoj teoriji. Jedan od primjera je Tablica slučajnih brojeva.

Skupovi po broju obilježja

Po ovoj podjeli postoje *jednodimenzionalni* i *višedimenzionalni* skupovi. Posmatranjem skupa po jednom obilježju (jednoj dimenziji) dobijamo jednodimenzionalan skup. Na primjer, skup stabala po visini. Ako istovremeno posmatramo visinu i prečnik stabala radi se o dvodimenzionalnom skupu, a ako uključimo zapreminu o trodimenzionalnom skupu.

U deskriptivnoj statistici uobičajeno je da posmatramo jednodimenzionalne skupove (skupove sa jednim obilježjem). U regresionoj analizi ispituju se veze između dva ili više obilježja, pa tada govorimo o dvodimenzionalnim ili višedimenzionalnim skupovima.

„On koristi statistiku kao pijanac uličnu svjetiljku – više za oslonac nego za rasvjetu“

Endru Lang (1844 – 1912)
Škotski književnik

3. STATISTIČKI RAD

Upoznajmo se na početku sa brojevima, odnosno vrstama brojeva. To je važno jer njima obično iskazujemo karakteristike (obilježja). Nekada umjesto brojeva koristimo riječi. Sve te radnje nazivamo mjerenjima, bilo da se radi o kvantitetu ili kvalitetu pojave. Mjerenja se izvode na različitim nivoima, koji se definišu skalama mjerenja. Statistički skup možemo posmatrati u cjelini (sve njegove jedinice) ili samo dio skupa (uzorak). O svemu ovome, u vidu najave onoga što nas očekuje kasnije, govorimo u nastavku ovog poglavlja.

3.1. Šta su brojevi, a šta podaci?

Broj je kratak, jasan i precizan način izražavanja. U matematici se koriste apstraktni (čisti, neimenovani) brojevi, dok je statistika nauka empirijskih (iskustvenih, imenovanih) brojeva. Napišemo li 35, za matematiku to je broj, dok u statistici nema značenje. Međutim, ako kažemo 35 cm onda to može biti prečnik stabla, odnosno podatak.

Podaci su brojevi s odgovarajućim značenjem, tj. brojevi koji nešto predstavljaju. Osim brojem, podaci se iskazuju riječima. Takvi su modaliteti kod atributivnih obilježja, na primjer kvalitet debla: dobar, srednji i loš.

Podaci

Podaci su brojevi sa određenim značenjem. To mogu biti i riječi (modaliteti atributivnih obilježja ili razne informacije, zapažanja i slično).

3.2. Vrste brojeva u statistici

U statistici razlikujemo dvije vrste brojeva:

- apsolutni brojevi i
- relativni brojevi.

Apsolutni brojevi

Apsolutni brojevi pokazuju *količinu pojave*. Dobijaju se mjerenjem ili brojanjem, a predstavljaju osnovni (izvorni) materijal (podatke). Oni daju uvid u

stvarno stanje pojave. Količina (intenzitet, veličina) pojave iskazuje se najčešće u jedinicama mjere (npr. prečnik stabla u *cm*, visina stabla u *m*). Obilježja sa različitim jedinicama mjere (npr. visina u metrima i zapremina stabla u metrima kubnim), kao i obilježja, sa istim jedinicama mjere, a različita po veličini (npr. zapremina i prirast, gdje je zapremina mnogostruko veća od prirasta) ne mogu se upoređivati u apsolutnom iznosu. To je najveći nedostatak ovih brojeva. *Napomena*: apsolutno znači bezuslovno, nepromjenljivo.

Relativni brojevi

Relativni brojevi pokazuju *količinske odnose*. Nastaju dijeljenjem dva apsolutna broja, kao izraz neke veličine mjeren drugom veličinom, koja služi kao baza (mjerilo za poređenje). Pošto imaju istu bazu relativni brojevi se mogu upoređivati bez ograničenja. Zato ih statistika često koristi kao svoje parametre. *Napomena*: relativno znači da to vrijedi samo pod određenim uslovima ili u odnosu na nešto drugo.

Relativni brojevi iskazuju se na više načina, kao:

1. Decimalni brojevi,
2. Procenti,
3. Standardizovane vrijednosti i
4. Koeficijenti.

Razlikujemo više vrsta relativnih brojeva. Nazive im dajemo u zavisnosti od toga kako su nastali. Navešćemo neke od njih.

Relativne frekvencije. Kod uređivanja statističkih skupova utvrđujemo broj elemenata po klasama ili grupama. Taj broj predstavlja apsolutne frekvencije, a njihov zbir je ukupan broj elemenata skupa. Ako stavimo u odnos apsolutne frekvencije sa ukupnim brojem dobijamo relativne frekvencije. Njihov zbir je jedan. Ako ih množimo sa 100, relativne frekvencije biće iskazane u procentima. Tada će njihov zbir biti 100. Relativne frekvencije jasnije i upadljivije ističu strukturu mase (skupa) od apsolutnih frekvencija.

Proporcije. Proporcije su specifičan tip odnosa gdje je brojilac dio imenioca. Takav je odnos neke apsolutne frekvencije i ukupnog broja elemenata. Na primjer, broj oboljelih stabala prema ukupnom broju stabala. Tada kažemo proporcija zaraženih stabala je npr. 7%. Proporcije se nalaze u intervalu od nula do jedan ili u odgovarajućim procentima.

Indeksi. Indeksi (indeksni brojevi) služe za poređenje smjera i intenziteta varijacija. Oni pokazuju odnose između članova nekog statističkog niza. Baza (mjerilo) za poređenje bira se prema cilju. Ovdje se jedna veličina mjeri drugom. Na primjer, kod uticaja proreda (sječe) na prirast, kroz životni vijek šume, baza može biti slaba proreda. Indeksi za jaku proredu veći od jedan pokazivaće pozitivan, a manji od jedan negativan uticaj ove prorede. Vrijednosti indeksa u

blizini jedinice znače da nema uticaja ili je vrlo slab. Indeksi se često iskazuju u procentima.

Stope. Kad se prati frekvencija (npr. broj zaraženih stabala po godinama) u odnosu na dugogodišnji prosjek govorimo o stopi. Ona se obično iskazuje u procentima. Kao primjer, često čujemo da se govori o stopi rasta BDP-a (bruto društvenog proizvoda).

Standardizovane vrijednosti. One su česte u statistici. Kako bismo mogli porediti pojave i donositi zaključke, potrebno je njihove originalne vrijednosti standardizovati. Najbolji primjer imamo kod teorijskih rasporeda. Kod primjene normalnog rasporeda standardizujemo (transformišemo) originalne vrijednosti obilježa *X* u standardizovane vrijednosti *Z*.

Koeficijenti. Koeficijenti su posebna vrsta standardizovanih vrijednosti. Na primjer, *Koeficijenti asimetrije* i *koeficijenti izduženosti*, kao neimenovani relativni brojevi, omogućuju poređenja oblika rasporeda različitih pojava. *Koeficijent varijacije* pokazuje relativni varijabilitet (apsolutno variranje u odnosu na aritmetičku sredinu). *Zapreminski koeficijent* ($f_{1,30}$) pokazuje odnos zapremine stabla i zapremine valjka. *Koeficijent međusobnog prekrivanja krošnji* pokazuje relativno prekrivanje krošnji.

Relativne veličine u šumarstvu

Šumovitost je odnos površine pod šumom i ukupne površine, iskazana u procentima. *Stepen sklopa* je odnos površine prekrivene krošnjama stabala i ukupne površine. *Obrast* je odnos stvarne temeljnice i tablične temeljnice, iskazan kao decimalni broj. *Omjer smjese* predstavlja udio neke vrste drveća u ukupnoj drvnoj masi. *Procent prirasta* je relativan prirast, tj. prirast drvne mase u odnosu na drvnu masu.

U statistici zaključke obično donosimo na bazi relativnih brojeva. Uzmimo jedan primjer iz naše prakse. Šumska gazdinstva imaju različitu strukturu šuma. Ako želimo uporediti dva šumska gazdinstva po „snazi" (šumskom bogatstvu) prvo što ćemo uporediti je površina visokih (ekonomskih) šuma, s kojom ta gazdinstva raspolažu. Prema apsolutnim brojevima (površini u hektarima) zaključak može biti pogrešan. Šumsko gazdinstvo koje ima 20.000 ha visokih (ekonomskih) šuma nalazi se u povoljnijem položaju, ako one zauzimaju 70% ukupne površine šuma, od gazdinstva sa 25.000 ha visokih šuma, čije je učešće 20% u ukupnoj površini šuma.

3.3. Skale (nivoi) mjerenja

U prirodi i društvu postoje različite masovne pojave, čije karakteristike (obilježja) trebamo izmjeriti. U nekim slučajevima biće moguće izmjeriti jedinice skupa i to mjerenje iskazati brojevima (kvantitativne razlike), dok ćemo često morati razlike između jedinica po nekom drugom obilježju opisivati riječima

(kvalitativne razlike). U statistici i uopšte za sve slučajeve mjerenja postoje četiri skale mjerenja. To su:

- nominalna,
- rang (ordinalna),
- intervalna i
- numerička (omjerna) skala.

Nominalna skala

Nominalna skala se koristi za mjerenje kvaliteta jedinica po nekom atributivnom obilježju. Ovo je zapravo skala po kojoj se samo daju imena (nazivi) modalitetima (*nomen* = ime). Na primjer, kvalitet šumskih drvnih sortimenata mjeri (određuje) se po ovoj skali. Pa tako imamo standardom definisane sortimente kao što su F i L trupci, trupci za rezanje, ogrevno drvo itd. Nominalna klasifikacija podrazumijeva dodjeljivanje modaliteta atributima, kategorizaciju ili klasifikaciju obilježja. Ovo je najgrublja skala.

Rang (ordinalna) skala

Ova skala je nešto viši nivo od nominalne skale. U njoj se određuje redoslijed u kvalitetu (*ordinalni* = redni). Na primjer, trupci za rezanje rangiraju se po kvalitetu na I, II i III klasu. Prva klasa je kvalitetnija od druge, druga od treće. Ovdje izrazi klasa i klasiranje imaju svoje posebno značenje, različito od onoga kod numeričkih obilježja. Uslovi za rad u šumi se određuju (mjere) uglavnom prema konfiguraciji terena. Na primjer, vrlo teški uslovi, teški uslovi, srednji uslovi i povoljni uslovi. Kao što vidimo po ovoj skali nije određeno kolika je razlika između modaliteta, već samo redoslijed u njihovom značaju (kvalitetu). Nominalna i rang skala primjenjuju se za atributivna obilježja.

Napomenimo da postoje neka obilježja koja pripadaju ovoj skali, a za njih ipak računamo prosječne vrijednosti. Takav primjer je bonitet staništa. To se pokazalo korisnim, a opravdanje je u tome što je jednak razmak između modaliteta (npr. između prvog i drugog, drugog i trećeg boniteta itd). Slično je i sa ocjenama na ispitu.

Intervalna skala

Primjenjuje se za numerička obilježja, gdje obilježja imaju jedinicu mjere. To znači da se uspostavlja funkcionalni odnos (jednake razlike) između susjednih vrijednosti. Zato su kod ove skale moguće računske operacije. Obično se kao primjeri navode temperatura i kalendarsko vrijeme. Položaj nule i mjerna jedinica za ovu skalu određeni su dogovorno (konvencijom). Nulta tačka nije prava vrijednost nule (temperatura na nuli ne znači da nema temperature). Intervalna skala se primjenjuje za vremenske serije, odnosno dinamička istraživanja.

Numerička (omjerna) skala

Ova skala naziva se još mjerna, apsolutna ili skala odnosa. Ona predstavlja najviši nivo mjerenja. Omjerna skala pokazuje jednakost između uzastopnih vrijednosti, ali za razliku od intervalne skale, ona ima pravu nultu vrijednost (ako vaga pokazuje nulu to znači da nema mase – masa ne postoji). Mjerenja numeričkih obilježja u šumarstvu, npr. prečnika i visine stabala, obavljaju se po ovoj skali.

3.4. Opis masovnih pojava

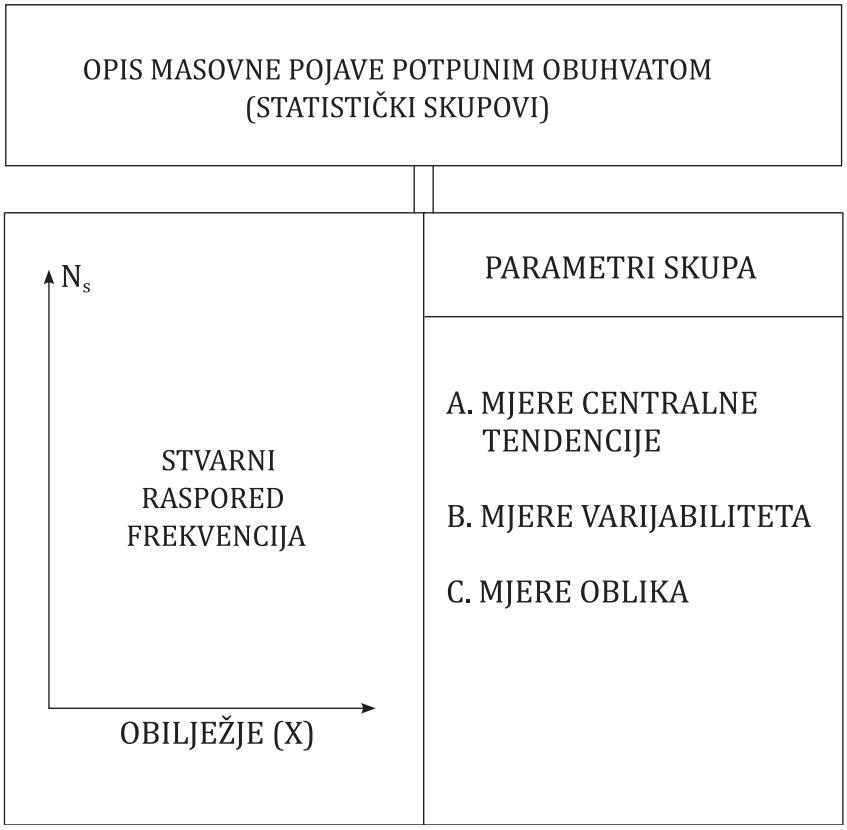
„Pojave u prirodi, specijalno u onoj oblasti prirode koja je objekat šumarskih istraživanja, jesu masovne. Njihovo ispitivanje vrši se, pored ispitivanja samih individua, na *skupovima* određenih individua (skup stabala sastojine, kulture ili plantaže, skup određene vrste gljiva, insekata i slično). Ovakav skup, koji se statistički zove još i kolektiv, populacija ili statistička masa, posjeduje izvjesna karakteristična svojstva, na osnovu kojih se razlikuje jedan skup od drugoga, a čije je poznavanje od velike teoretske i praktične koristi. Osnovna karakteristična svojstva skupa su *raspored (razdioba) elemenata skupa, centralna tendencija i variranje – rasturanje elemenata skupa*. Na primjer: za šumarskog stručnjaka ima teoretski i praktični značaj poznavanje rasporeda, srednje vrijednosti (mjera centralne tendencije) i srednjeg rasturanja (mjera variranja) skupa stabala jedne sastojine, u stvari poznavanje karakterističnih svojstava taksacionih elemenata: debljine, visine, zapremine, prirasta, kvaliteta stabala i dr. U matematičkoj statistici taksacioni elementi zovu se obilježja elemenata skupa, pri čemu su pojedina stabla elementi skupa“ (11, str. 473).

Statistički rad obuhvata tri faze rada. To su prikupljanje podataka, obrada podataka i analiza (interpretacija) rezultata *(Slika 4)*. Dva su osnovna načina prikupljanja podataka: potpuni obuhvat svih jedinica skupa i djelimični obuhvat. U šumarstvu za potpuni obuhvat kažemo *potpuni premjer*, a za djelimični obuhvat *reprezentativni metod*. Premjer šuma se još naziva taksacija (lat. *taxatio* = procjena) ili inventura šuma. U društvenim oblastima koriste se nazivi *popis*, za potpuni obuhvat i *anketa*, za djelimični obuhvat.

Slika 4. Koraci u statističkom radu

Potpuni obuhvat

Statističko posmatranje (mjerenje) jedinica skupa je najvažnija faza rada. Ako su prikupljeni podaci nepotpuni ili netačni svaki statistički metod daće pogrešan rezultat. Potpuni (totalni) premjer podrazumijeva obuhvat svih jedinica skupa. U šumi gdje su jedinice stabla, a čiji je broj praktično beskonačan, potpuni premjer nije moguć (osim u rijetkim slučajevima). Takav premjer bi pratile teškoće, kao što je nemogućnost kontrole radova. S obzirom da su kod ovog načina rada obuhvaćene sve jedinice skupa statistički rad praktično se završava sa deskriptivnom statistikom *(Slika 5)*. Tada su svi izračunati parametri, koji opisuju pojavu, tačni (pravi, stvarni).

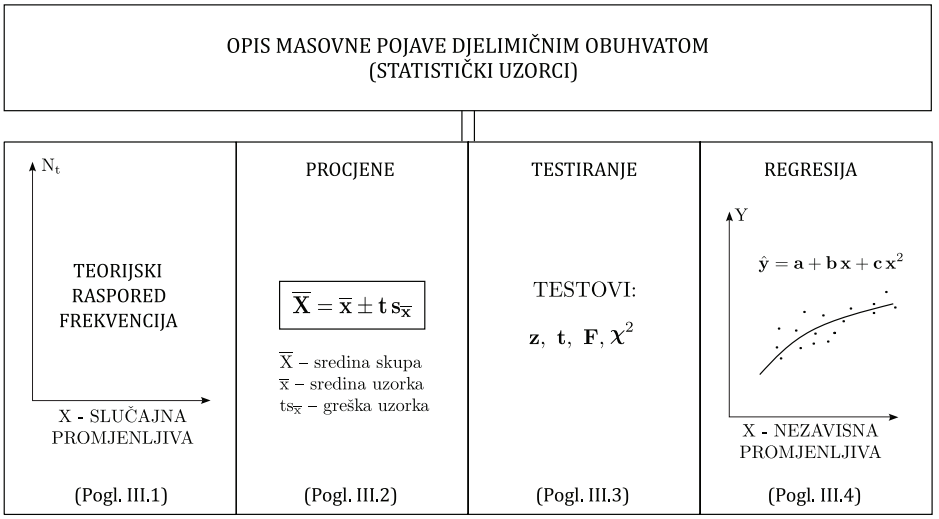


Slika 5. Opis masovne pojave potpunim obuhvatom

Djelimični obuhvat

Ako je pojava (skup) velikog obima, u statističkom radu obuhvatamo (uzimamo) samo određeni broj jedinica. Tada govorimo o *uzorcima*. Za uzorke računamo: srednje vrijednosti, mjere varijabiliteta, proporcije i na osnovu njih

određujemo vrijednosti istih parametara u skupu. U stvari, radi se o njihovim procjenama, na bazi vjerovatnoće i teorijskog modela *(Slika 6)*. Statistika pri tome određuje pouzdanost i preciznost takvih procjena. Metode koje na osnovu uzorka donose zaključke o skupu, odnosno objašnjavaju pojave u cjelini spadaju u tzv. inferencijalnu statistiku. Ona obuhvata još testiranja različitih hipoteza i regresionu analizu.



Slika 6. Opis masovne pojave djelimičnim obuhvatom

Plan istraživanja

U statističkom radu mora da postoji plan istraživanja (plan rada). Taj plan u šumarstvu često je obiman pa nastaju uputstva za prikupljanje podataka ili metodike prikupljanja podataka. Kao što u prodavnicu ne idemo po jednu namirnicu tako ni u šumu ne idemo radi jednog obilježja. Cilj nam je uvijek prikupiti što više informacija (podataka). To je racionalan pristup.

Plan istraživanja je zapravo spoj statistike i metodologije. On treba da sadrži: potpunu definiciju skupa, jasno definisane modalitete atributivnih i numeričkih obilježja i cilj (svrhu) rada. Manual (upitnik) mora biti jasan za mjerače/popisivače i podesan za obradu, sa podacima u vidu brojeva i šifri, zatim odgovora DA/NE, ima/nema, kao i odgovora na zaokruživanje. Podaci su ranije na terenu upisivani u manuale, a danas se već prelazi na direktan elektronski unos podataka.

Velika prirodna varijabilnost i složenost bioloških pojava (rad sa živim organizmima u promjenljivim stanišnim uslovima) otežavaju zaključivanje (utvrđivanje zakonitosti). Osim toga, treba imati u vidu da šumari rade sa puno podataka, na velikom prostoru i dugim vremenskim periodima.

3.5. Upotreba računara

Statistika se danas koristi računarima, što joj daje mogućnosti obrade velike mase padataka. Za osnovne statističke obrade najviše se koristi EXCEL, a dostupni su razni statistički programi, kao što su: SPSS, MINITAB, SAS, STATISTICA ili STATGRAPHICS. Bez računara praktično nije moguća primjena novih statističkih metoda. Statistika je povezana sa metodologijom, a u novije vrijeme sve više sa informacionim sistemima (GIS i drugi).

3.6. Pojam „greške“ u statistici

U statistici se pojam greške pojavljuje u više različitih značenja, što može da zbuni čitaoca. Osim klasičnog značenja, „greška“ u statistici označava odstupanje vrijednosti od nekog parametra, najčešće od aritmetičke sredine (tako standardnu devijaciju u nekim slučajevima nazivamo standardnom greškom). Sve greške u statistici možemo svrstati u:

- tehničke greške,
- greške u statistici i
- greške izvan statistike.

Tehničke greške

Greške koje nastaju u radu (mjerenju) nazivaju se *tehničke greške*. One mogu biti:

- slučajne,
- sistematske i
- grube greške.

Slučajne greške su male, pozitivne ili negativne, greške koje se poništavaju. One se u velikom broju raspoređuju normalno. Njih nije moguće izbjeći, pa se još nazivaju neizbježne greške. Variranje koje je rezultat slučajnih grešaka smatra se statistički slučajnim. One su dobile naziv po tome što se javljaju od slučaja do slučaja.

Sistematske greške su najčešće posljedica neispravnog instrumenta ili metoda rada. One su opasne, jer ih ne vidimo, a daju sistematski više ili niže rezultate (nagomilavaju se). Prirodni faktori ili vještački tretmani takođe izazivaju sistematsko djelovanje (djelovanje u jednom smjeru). To statistika prepoznaje (koristi) i tim faktorima određuje statistički značaj.

Grube greške nastaju uglavnom zbog nepažnje, bilo prilikom mjerenja (brojanja) ili unosa podataka. Otklanjaju se povećanom pažnjom u radu, kontrolama (nadzorom) i naknadnim otkrivanjem među ekstremnim vrijednostima.

Greške u statistici

Standardne devijacije u regresiji i uzorcima nazivaju se standardnim greškama (regresije, odnosno uzorka). „Greškom“ se nazivaju zato što standardna devijacija regresije izravnava odstupanja (pogrešne vrijednosti) od regresione jednačine, a standardna greška uzorka odstupanja (pogrešne vrijednosti) od prave vrijednosti skupa.

Greške izvan statistike

U statistici se slučajevi čija je vjerovatnoća manja od 5% smatraju greškom. To je preneseno u svakodnevni život, pa se za rijetke događaje kaže da su na nivou *statističke greške*.

U radu s brojevima nailazimo i na upotrebu izraza „češka greška“. Naziv je nastao u Njemačkoj. Naime, Nijemci izgovaraju brojeve obrnutim redom od drugih. Oni 21 izgovaraju „einundzwanzig“ (jedan i dvadeset). Vidjevši da Česi rade suprotno, zamjenu brojeva nazvali su češkom greškom. Na primjer: ako umjesto 632 napišemo 623 napravili smo češku grešku.

„Ispitanici“ u šumi.

II. DESKRIPTIVNA STATISTIKA

Ronald Fišer
(1890-1962)
Osnivač moderne statističke nauke

Deskriptivna statistika (opis masovnih pojava) obuhvata prikupljanje podataka, obradu podataka i prikazivanje rezultata pomoću tabela, grafikona i sumarnih mjera (statističkih parametara). Sva četiri poglavlja deskriptivne statistike odnose se na stvarne skupove. To su poglavlja: Uređivanje statističkih skupova, Mjere centralne tendencije, Mjere varijabiliteta i Mjere oblika rasporeda frekvencija.

Deskriptivna statistika može se odnositi i na uzorke. Tada deskriptivne mjere (parametre uzorka) nazivamo Statistika uzorka. Kompjuterski izlazi uobičajeno koriste izraz Deskriptivna statistika.

1. UREĐIVANJE STATISTIČKIH SKUPOVA

Obradu podataka danas radimo pomoću računara i gotovih statističkih programa. Međutim, kako bismo razumjeli statistiku i znali izabrati odgovarajući statistički metod, te tumačiti izlaze iz računara neophodno je osnovne statističke operacije savladati „ručno".

1.1. Redukcija podataka

Kako u velikoj masi podataka praktično vidimo samo brojeve, njih je potrebno redukovati. Redukciju vršimo grupisanjem elemenata. Tako lakše dolazimo do prostih brojeva (parametara, koeficijenata), koji služe da opišemo pojavu. Grupisanje predstavlja temelj statističkog rada.

Statistički skup čine sve njegove jedinice. Od tih jedinica počinje statističko istraživanje. Pojava se ispituje njihovim brojanjem ili mjerenjem. Mjerenje jedinica obavljamo, u stvari, preko njihovih obilježja. Ona mogu biti jako različita, a neka od njih zahtijevaju detaljan opis. Ti opisi nazivaju se *klasifikacije*.

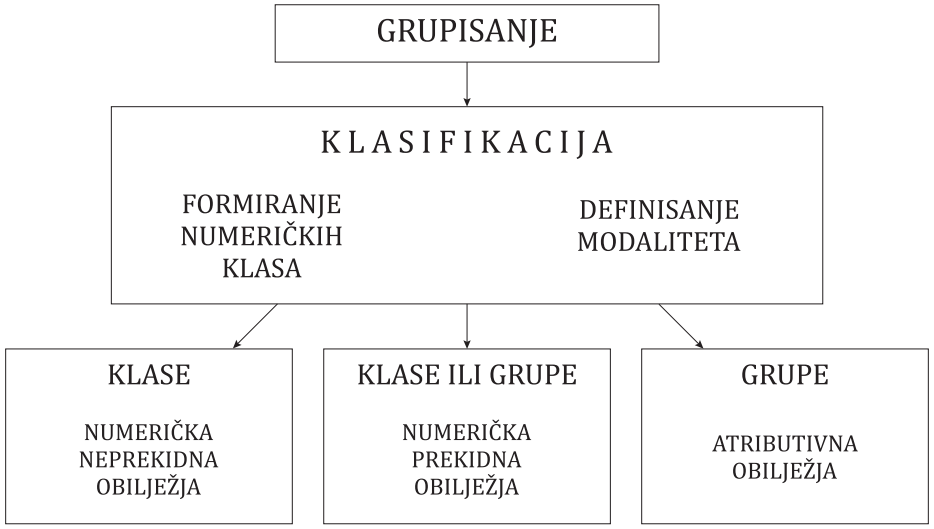
1.1.1. Klasifikacija kao pojam

Klasifikacija je širok pojam koji podrazumijeva različite vrste podjela. Načelno, to je dokument (papir) po kome (na osnovu kojeg) se izvodi grupisanje podataka u razne svrhe. U šumarstvu su posebno važne klasifikacije šuma, klasifikacije živih stabala i klasifikacije šumskih drvnih sortimenata. Klasifikacija, pored razrade kriterijuma kod atributivnih obilježja, takođe podrazumijeva formiranje klasa kod numeričkih obilježja (po njihovoj veličini).

1.1.2. Grupisanje elemenata

Kako se vrši grupisanje zavisi o kakvom se obilježju radi *(Slika 7)*. Kod numeričkih obilježja brojčane vrijednosti razvrstavamo u određene intervale.

Te intervale nazivamo *klase*. Kod atributivnih obilježja elemente razvrstavamo po modalitetima. Tako dobijamo *grupe*. Klase i grupe treba razlikovati. U klasi se nalaze različiti (nejednaki) elementi (npr. u debljinskoj klasi stabla su različitog prečnika), dok su u grupi svi elementi isti (npr. u grupi četinari sva stabla su jednako četinarske vrste drveća). Osim toga, klase imaju širinu (interval), a grupe nemaju. U slučaju malog broja elemenata grupisanje se ne vrši.



Slika 7. Grupisanje elemenata

Ispravno bi bilo reći *klasiranje* numeričkih obilježja i *grupisanje* atributivnih obilježja. Međutim, u literaturi se obično govori samo o grupisanju.

Grupisanje u klase (neprekidna obilježja)

Grupisanje u klase (klasiranje) se uglavnom primjenjuje za numerička neprekidna obilježja. Na primjer, prečnike stabala razvrstavamo po debljinskim klasama, a visine po visinskim klasama. Grupisanje u klase za nas je posebno važno. O njemu ćemo više u nastavku.

Grupisanje u klase ili grupe (prekidna obilježja)

Grupisanje u klase izvodi se i za numerička prekidna obilježja, ali rijetko. Radi se o slučajevima ako se javljaju u velikom broju oblika (različitih vrijednosti). Na primjer, broj radnika u šumskim gazdinstvima. Da bismo ispitali strukturu gazdinstava (po broju radnika) pri uređivanju skupa bolje je uzeti klase određene širine, a ne pojedinačne vrijednosti. Tako dobijamo bolju preglednost.

Ako se *numerička prekidna obilježja,* pojavljuju u samo nekoliko (malom broju) oblika nastaju grupe. Na primjer, kad posmatramo šumska gazdinstva po broju privrednih jedinica zadržaćemo pojedinačne vrijednosti, pa ćemo imati šumska gazdinstva sa jednom, dvije itd privrednih jedinica. Dakle, za prekidna numerička obilježja, kod većeg broja modaliteta primjenjuje se klasiranje, a grupisanje kad je njihov broj mali.

Grupisanje u grupe (atributivna obilježja)

Grupisanje u grupe primjenjuje se uglavnom za *atributivna obilježja*. Grupe se prave prema modalitetima atributivnog obilježja. Na primjer: svrstavanje stabala u grupe vrsta drveća (četinari i lišćari), kategorizacija šuma po porijeklu (sjemenske i izdanačke) ili rangiranje staništa po kvalitetu (boniteti od prvog do petog). U slučaju velikog broja modaliteta nekog obilježja (pri čemu ne smijemo brkati broj modaliteta sa brojem elemenata) prave se stalne šeme grupisanja atributivnih obilježja, kao što je nomenklatura proizvoda ili klasifikacija zanimanja.

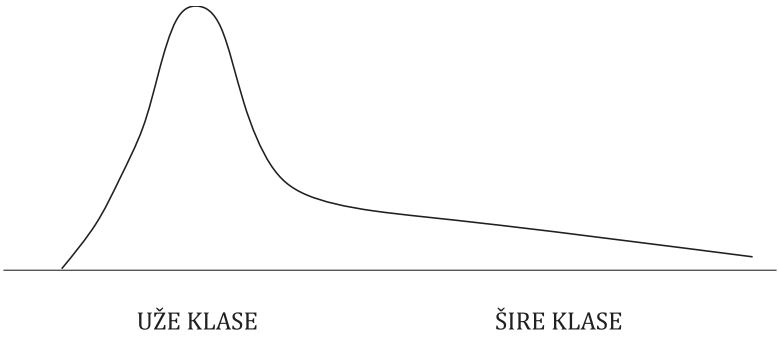
1.1.3. Broj klasa/grupa

Prilikom grupisanja obilježja nailazimo na dva problema: koliko klasa (grupa) formirati i kako ih razgraničiti. Kod atributivnih obilježja, gdje su modaliteti jasno određeni ili prekidnih obilježja (kod kojih se ostaje na pojedinačnim vrijednostima) taj problem ne postoji.

Broj klasa. Broj klasa, odnosno grupa zavisi od više faktora. Klase (grupe) su često unaprijed određene, bilo iskustveno (kao pravilo) ili zakonski (kao obaveza). Serije podataka koje dobijamo grupisanjem podataka moraju biti *pregledne* (što je manji broj klasa/grupa preglednost je bolja) i u isto vrijeme *informativne* (što je veći veći broj klasa/grupa imamo više informacija). U praktičnom radu traži se kompromis između ova dva zahtjeva.

Širina klasa. Kad god je moguće treba formirati klase *jednake širine* i koristiti uobičajene (standardizovane) modalitete grupa. Razlog je jednostavan, trebaju nam statističke serije u kojima će klase/grupe biti uporedive međusobno, ali i sa drugim serijama. Međutim, to neće uvijek biti moguće. Jako asimetrična raspodjela frekvencija „zahtijeva" uže klase tamo gdje su elementi više skoncentrisani *(Slika 8)*. Uže klase treba formirati tako da se mogu objedinjavati.

Otvorene granice. Kod formiranja klasa pojavljuje se još jedan problem. To su otvorene granice krajnjih klasa. Nekada se pojedinačni podaci toliko rasipaju (udaljavaju od ostalih) da ih praktično nije moguće obuhvatiti postojećim klasama. U tim slučajevima krajnje klase se ostavljaju otvorenim. Kod atributivnih obilježja možemo formirati grupu „ostali", koja bi obuhvatala sve rijetke slučajeve (modalitete).



Slika 8. Nejednaka širina klasa kod asimetričnog rasporeda

Pogledajmo kakva je situacija u pogledu formiranja klasa u našem šumarstvu. Već smo rekli da su klasifikacije polazište za grupisanje, odnosno za statistički rad. Osnovne klasifikacije, u našem šumarstvu, propisane su odgovarajućim pravilnicima. Šume podliježu dvjema klasifikacijama: ekološko-proizvodnoj klasifikaciji (Šira kategorija šuma - ŠKŠ, Uža kategorija šuma – UKŠ i Gazdinska klasa - GK) i prostornoj klasifikaciji (Šumskoprivredno područje – ŠPP, Privredna jedinica – PJ i odjel). Klasifikacije stabala urađene su po više osnova: po vrstama drveća ili grupama (četinari i lišćari), po debljini (debljinske klase) i po kvalitetu stabala (uzgojne i tehničke klase). Šumski drvni sortimenti (ŠDS) propisani su standardom.

U našoj praksi više od 50 godina primjenjuje se ista klasifikacija stabala po debljini. Važeći Pravilnik, za visoke (sjemenske) šume, predviđa sljedeće debljinske klase:

5 – 10; 10 – 20; 20 – 30; 30 – 50; 50 – 80, > 80 cm.

Možemo zapaziti da su ove klase različite širine. One se međusobno ne mogu upoređivati. Kod grafičkog prikazivanja za njih ne važi linearna razmjera, pa se mora koristiti površinska razmjera. Takođe, uočavamo postojanje jedne otvorene klase. Ranije, u našoj praksi, ekstremno debela stabla (100 cm i više) upisivana su u manual kao prečnik 99 cm. Za to je postojalo više razloga. Izbjegavane su greške mjerenja koje nastaju kod jako deformisanih debala i svjesno se išlo na smanjivanje zalihe.

1.1.4. Razgraničenje klasa/grupa

Određivanje granica klasa (grupa) je posebno važno pitanje. Posmatraćemo ga odvojeno za numerička neprekidna, numerička prekidna i atributivna obilježja.

Granice kod numeričkih neprekidnih obilježja

Formiranje klasa neprekidnih obilježja zavisi od preciznosti mjerenja i načina zaokruživanja izmjerenih vrijednosti. Kao primjer, u dvije varijante zaokruživanja (na niže i na više) i preciznosti mjerenja (na cijeli cm i na jednu decimalu), date su u *Tabeli 1*, granice za klase prečnika širine 5 cm.

Tabela 1. Granice debljinskih klasa (cm)

Način zaokruživanja	Radne granice (cijeli brojevi)	Prave granice	Radne granice (na decimalu)	Prave granice
1	2	3	4	5
	15 – 19	15 – 20	15,0 – 19,9	15,0 – 20,0
Na niže	20 – 24	20 – 25	20,0 – 24,9	20,0 – 25,0
	25 – 29	25 – 30	25,0 – 29,9	25,0 – 30,0
	...	...		...
	16 – 20	15 – 20	15,1 – 20,0	15,0 – 20,0
Na više	21 – 25	20 – 25	20,1 – 25,0	20,0 – 25,0
	26 – 30	25 – 30	25,1 – 30,0	25,0 – 30,0
	...	...	...	

Ako radimo ručno granice pišemo različito prije grupisanja (radne granice) i nakon grupisanja (prave ili stvarne granice). Radne granice klasa potrebne su nam samo kod razvrstavanja. One se pišu tako da nema preklapanja između vrijednosti susjednih klasa, da se zna kojoj klasi elementi pripadaju (kol. 2 i kol. 4). Prave granice su precizne (oštre) i one ne zavise od načina zaokruživanja i preciznosti mjerenja (kol. 3 i kol. 5). One se najbolje vide na histogramu frekvencija. Više u primjeru na kraju Poglavlja II. Postoje i drugi načini razgraničenja klasa. Na primjer, dodavanjem decimala, koje se kasnije u obradi zanemaruju.

Ovdje nismo prikazali varijantu zaokruživanja na bliže (prema sredini). Ona nije praktična, jer se dobijaju granice klasa koje nisu cijeli brojevi. Zaokruživanjem na bliže, npr. prečnik od 15 cm obuhvatao bi vrijednosti od 14,5 do 15,5 cm, a prva klasa bila bi od 14,5 do 19,5 cm.

Granice kod numeričkih prekidnih obilježja

Vratimo se našem primjeru, koji se odnosi na broj radnika u šumskim gazdinstvima. Napišimo, na dva načina, formiranje klasa širine 50 radnika:

A: 1 – 50; 51 – 100, 101 – 150; 151 – 200.

B: 0 – 50; 50 – 100; 100 – 150; 150 – 200.

Odgovorimo na pitanja: Koji je način ispravan? U koju klasu ćemo svrstati gazdinstvo koje ima 50 radnika? Postoje li ovdje radne i prave granice?

Klase kod prekidnih obilježja, po pravilu, pišu se tako da svaka naredna klasa počinje brojem većim za jedan. Tako je urađeno u varijanti A, gdje se jasno zna da šumsko gazdinstvo sa 50 radnika pripada prvoj klasi, a gazdinstvo sa 51 radnikom drugoj klasi. Klase u slučaju B nisu pravilno formirane.

Osim na ovaj način, klase se mogu razgraničiti i korištenjem znakova veće, manje, jednako. Na primjer ovdje, > 0 – 50, > 50 – 100 itd. Prekidna obilježja imaju samo cijele brojeve, tako da nema zaokruživanja. Dakle, kod prekidnih obilježja nema potrebe za formiranjem radnih granica.

Granice kod atributivnih obilježja

Razgraničenje modaliteta kod atributivnih obilježja vrši se pri definisanju statističkog skupa, opisom modaliteta ili pozivanjem na neku već postojeću klasifikaciju (npr. Matićevu klasifikaciju stabala po kvalitetu). Kod nekih atributivnih obilježja granice nije baš jednostavno definisati. Na primjer, kod boje kore stabla, jer ima mnogo nijansi (prelaza). Ako se radi o prostornim (geografskim) obilježjima granice su određene na terenu administrativno (npr. opštine) ili prostornom podjelom šuma (npr. šumskoprivredna područja).

1.2. Frekvencije (učestalost)

U statistici uglavnom radimo s brojevima. Govorili smo o tome u Opštem dijelu (Pogl. 3). Ovdje govorimo o jednoj vrsti brojeva, koji nemaju jedinice mjere ali imaju svoje značenje. To su oni brojevi koji stoje uz obilježje, odnosno uz vrijednosti (modalitete) obilježja. Nazivamo ih *frekvencijama*.

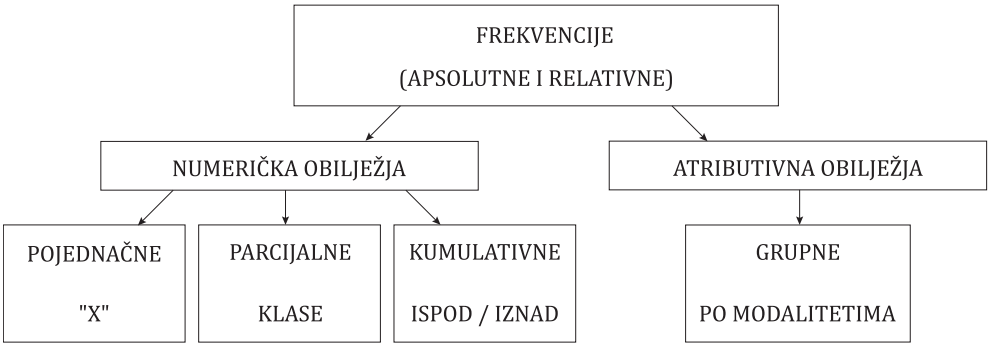
Frekvencije (učestalost)

Frekvencije su brojevi koji pokazuju koliko se puta vrijednosti (modaliteti) nekog obilježja ponavljaju.

1.2.1. Vrste frekvencija

Već na prvi pogled zapazićemo da se u skupu neki podaci javljaju više puta. To su njihove *pojedinačne frekvencije*. Nakon što grupišemo elemente u klase dobijamo *parcijalne frekvencije* (broj elemenata po klasama). Njih možemo sabirati, od najmanje do najveće klase i obrnuto. Tako nastaju *kumulativne frekvencije* (kumulanta *ispod* i kumulanta *iznad*).

Kod atributivnih obilježja postoje samo *frekvencije grupa* (broj elemenata određenog modaliteta). Ne postoje parcijalne i kumulativne frekvencije, jer ova obilježja ne podliježu matematičkim operacijama. Frekvencije se mogu iskazivati na dva načina, apsolutno i relativno. Vrste frekvencija prikazane su na *Slici 9*.



Slika 9. Vrste frekvencija

Apsolutne frekvencije

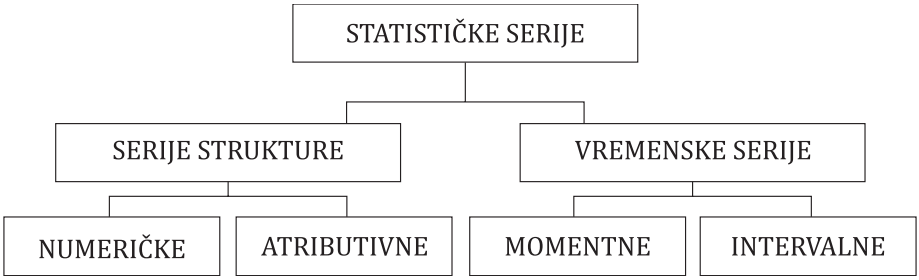
Broj koji pokazuje koliko se puta neka vrijednost (modalitet) obilježja pojavljuje (ponavlja) u skupu nazivamo *apsolutna frekvencija*. Apsolutna frekvencija je uvijek cijeli broj, jer se dobija brojanjem. Ona daje težinu (značaj) vrijednostima (kod numeričkih obilježja), odnosno modalitetima (kod atributivnih obilježja). Iako se frekvencija obično označava sa „f“ ili „f_r“ mi u praktičnom radu često koristimo oznaku „en“ (veliko „N“ u skupu, a malo „n“ u uzorku).

Relativne frekvencije

Dijeljenjem apsolutne frekvencije klase (grupe) sa ukupnim brojem elemenata dobija se *relativna frekvencija*. Ona u nekim slučajevima predstavlja vjerovatnoću klase (grupe), pa joj dajemo oznaku „p“. Relativna frekvencija se iskazuje kao decimalni broj ili u procentima (decimalni brojevi pomnoženi sa 100). Zbir za sve klase (grupe) u prvom slučaju iznosi jedan, a u drugom 100%.

1.2.2. Statističke serije

Sve frekvencije u jednom skupu čine niz ili statističku seriju frekvencija. Zavisno od toga o kakvom se obilježju radi, postoje različite serije *(Slika 10)*.



Slika 10. Vrste statističkih serija

Serije strukture

Statičkim metodama utvrđujemo stanje pojave, odnosno njenu strukturu. Zato te serije nazivamo serije strukture. One mogu biti numeričke i atributivne.

Numeričke serije. Za numeričke serije obično koristimo naziv raspored (distribucija, razdioba) frekvencija. Rasporedi frekvencija mogu se predstaviti empirijskom funkcijom, gdje je obilježje (x) nezavisna, a frekvencija (y) zavisna promjenljiva.

Rasporedi frekvencija, bilo apsolutnih bilo relativnih, obično se prikazuju tabelarno i grafički. Rasporedi otkrivaju karakteristike skupa, kao što su koncentracija elemenata, njihov varijabilitet i oblik rasporeda. Struktura skupa se i ovdje bolje uočava na osnovu relativnih frekvencija, posebno kad su apsolutni brojevi veliki. Zato se niz frekvencija iskazan u procentima često naziva „*struktura*". Strukture su pogodne za razna poređenja.

Raspored frekvencija

Raspored frekvencija je serija (niz) svih frekvencija nekog numeričkog obilježja.

Atributivne serije. Raspodjela mase (struktura skupa) po atributivnom obilježju sastoji se od niza modaliteta i njihovih frekvencija. Najjednostavnije su dihotomne serije sa samo dva modaliteta. Na primjer, stanovništvo po polu (pol je atributivno obilježje, a muški i ženski pol su modaliteti), stabla grupisana po vrstama drveća (četinari i lišćari) ili stabla po porijeklu (sjemensko i vegetativno). Atributivne serije su složenije što je broj modaliteta veći. Na primjer, broj vrsta drveća je veliki: jela, smrča, bukva itd. U šumarstvu postoji opšta ekološko-proizvodna klasifikacija šuma (ŠKŠ, UKŠ, GK). Takođe, veliki broj modaliteta postoji u proizvodnji šumskih drvnih sortimenata (trupci, ogrev).

Kada se na mjestu atributa nalaze teritorijalne jedinice govorimo o geografskoj (prostornoj) raspodjeli statističke mase, tzv. *geografskim* serijama. U šumarstvu je iz organizacionih i ekonomskih razloga posebno važna prostorna podjela šuma (ŠPP, PJ, odjeli). Atributivne serije ne mogu se matematički obrađivati, zbog čega se mnogo veći značaj daje numeričkim serijama.

Vremenske serije

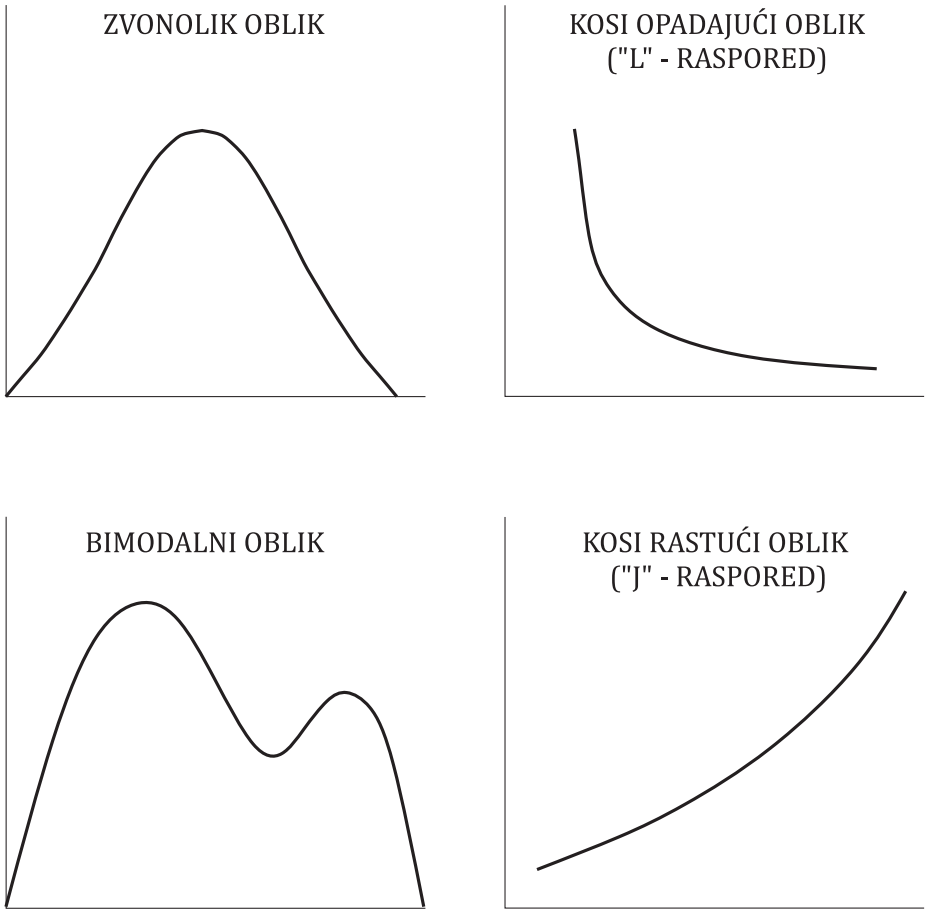
Dinamička analiza pojava ima za cilj da ispita promjenu pojave tokom vremena. Tada se napušta posmatranje frekvencija po obilježjima, jer više nije cilj struktura. Ove serije se nazivaju *vremenske serije*. One pokazuju veličinu pojave u nekim momentima (momentne serije) ili u nekim periodima (intervalne serije). U redovnom gazdovanju šumama, svakih 10 godina, utvrđujemo zalihu i prirast šuma. Zalihu možemo posmatrati kao momentnu seriju, dok prirast u nekom vremenu može biti intervalna serija. Kod momentne serije podaci se ne mogu sabirati (jer zbir nema smisla), dok kod intervalne mogu (zbir tri desetogodišnja prirasta predstavlja prirast za 30 godina).

Kod vremenskih serija radi se o dvodimenzionalnim skupovima. Pored kalendarskog vremena na x-osi (formalno uzetom kao nezavisna promjenljiva) imamo i drugu pojavu, čija se veličina prikazuje na y-osi. Ova veza može se izraziti funkcijom, koja u dužem vremenskom periodu pokazuje razvojnu tendenciju, odnosno *trend* pojave.

1.2.3. Oblici rasporeda frekvencija

Stvarni (empirijski) rasporedi nastaju uređivanjem konkretnih skupova. S obzirom na to da na skupove (masovne pojave) utiču brojni faktori to će i rasporedi frekvencija imati različite oblike *(Slika 11)*. Najčešći (tipičan) oblik rasporeda je zvonolik raspored, kome odgovara teorijski model normalnog rasporeda. Većina prirodnih i društvenih pojava ima ovakav oblik. U šumarstvu važna karakteristika šume je raspodjela njenih stabala po debljini. Zvonolik oblik te raspodjele karakterističan je za šume u kojima su stabla iste starosti (jednodobne sastojine), dok je u prirodnim (prebornim) šumama raspored kosi opadajući (naziva se „L“ – el raspored). Ovo su dva osnovna oblika debljinske strukture šuma.

Ako u šumi postoje dva sprata drveća (gornja i donja etaža) raspodjela ima dva vrha (bimodalni oblik). Po kosom rastućem, tzv. *jot* („*J*“) rasporedu povećava se smrtnost (broj uginulih) insekata u zavisnosti od jačine insekticida. U prirodi postoje i drugi, nepravilni (netipični, prelazni) oblici.



Slika 11. Različiti oblici rasporeda frekvencija

1.3. Tabelarni prikaz rasporeda frekvencija

Statističke serije se obično prikazuju u tabelama, gdje se nastavlja njihova kvantitativna analiza. Po namjeni treba razlikovati tri vrste tabela: *radne tabele* (koristimo ih dok radimo statistiku), *analitičke tabele* (služe za prikaz statističke dokumentacije) i *izvještajne tabele* (koriste se uglavnom u publikacijama). Po sadržaju mogu biti proste, složene i kombinovane. Tabele predstavljaju koristan alat. Pri njihovoj upotrebi treba se držati metodoloških pravila.

1.4. Načini grafičkog prikazivanja rasporeda frekvencija

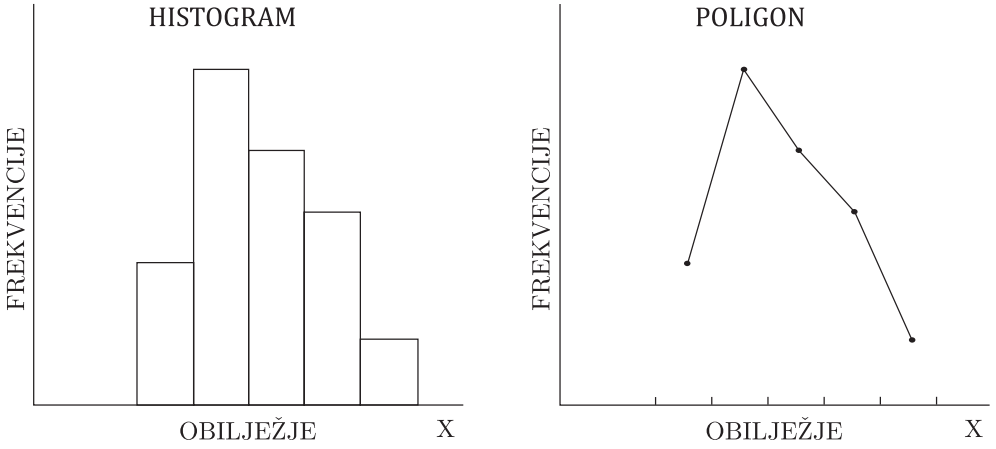
Grafički način prikazivanja rasporeda frekvencija omogućuje brz i jednostavan uvid u nivo (veličinu) pojave, njenu strukturu (koncentraciju i varijabilnost elemenata, oblik rasporeda) ili promjene u vremenu. Takođe, grafikoni omogućavaju različita poređenja. Mogu se prikazivati u koordinatnom sistemu i izvan njega. Po geometrijskom obliku grafikoni (dijagrami) mogu biti: *tačkasti* (korelacioni), *linijski* (poligonalna linija) i *površinski* (stubičasti, histogram, kružni). Primjena računara i statističkih programa omogućava grafičko prikazivanje rasporeda u mnoštvu različitih oblika i figura. Na primjer, česti oblici su tzv. „pite", koje nastaju uvođenjem debljine krugovima kao treće dimenzije (3D efekat).

N*umerička neprekidna* obilježja uglavnom se prikazuju u pravouglom koordinatnom sistemu. Na apscisu se nanose vrijednosti obilježja, a na ordinatu frekvencije. Najčešći oblici njihovog grafičkog prikaza su: histogram frekvencija i poligon frekvencija *(Slika 12)*. Grafički se još prikazuju kriva frekvencija i kumulante.

Histogram frekvencija. U koordinatnom sistemu prikazuju se pravougaonici čija je osnova na x-osi jednaka širini klasa, a visina odgovara frekvenciji. Ukupna površina predstavlja cijelu masu, a pravougaonici frekvencije klasa.

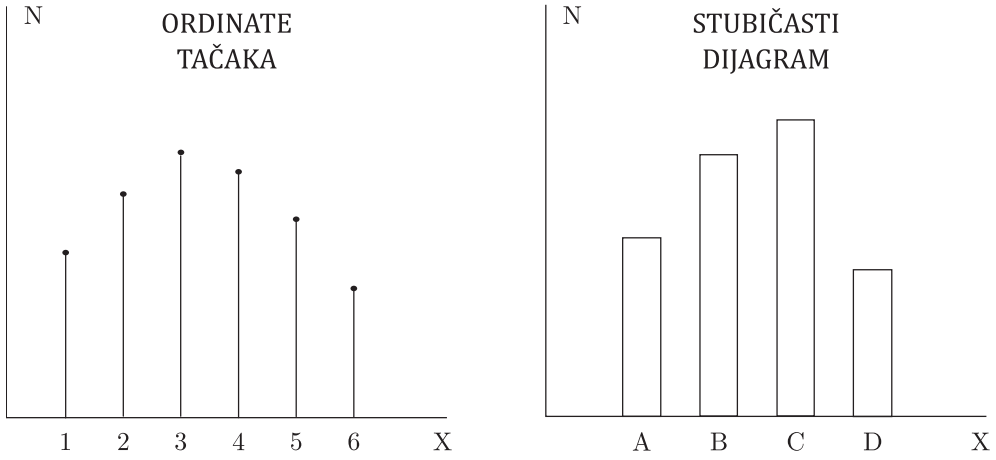
Poligon frekvencija. Poligon se predstavlja izlomljenom linijom, koja spaja tačke nanesene na sredinama klasa. Površina ispod krive linije nije u potpunosti srazmjerna frekvencijama klasa. Zatvoreni poligon predstavlja ukupnu frekvenciju. Kod njega krajnje tačke treba spojiti sa sredinama susjednih klasa, čija je frekvencija jednaka nuli. To vrijedi samo za zvonolike oblike.

Kriva frekvencija. Izravnanjem poligonalne linije dobija se kriva frekvencija, kojom nastojimo približiti se opštoj pravilnosti. To je, u suštini, teorijska linija. Često nam podaci ne dozvoljavaju crtanje ove krive, pa se dešava, ako to izravnanje prepustimo računaru, da dobijemo neodgovarajuće pa i besmislene krive.



Slika 12. Histogram i poligon frekvencija (neprekidno obilježje)

Serije sa *prekidnim numeričkim* obilježjem prikazuju se tačkama, ordinatama tačaka, štapićastim dijagramom ili linijskim dijagramom, a za *atributivna* obilježja koriste se najčešće površinski dijagrami u obliku stubova, pravougaonika ili krugova *(Slika 13)*.



Slika 13. Ordinate tačaka (prekidno obilježje) i stubičasti prikaz (atributivno obilježje)

> *Pojavu smo upoznali i analizirali nakon formiranja niza, tako da smo najprije upoznali kvantitativni izražaj pojave s pomoću apsolutnih frekvencija a potom smo grafičkim prikazom povećali preglednost pojave. Relativnim frekvencijama mogu se vršiti usporedbe raznorodnih distribucija i utvrditi odnosi među njima. U svim navedenim postupcima trebalo je uzimati u obzir čitavu razdiobu, no brojčani je oblik numeričkog obilježja vrlo prikladan i za daljnju primjenu statističke metode. Daljom primjenom statističke metode mogu se izračunati i brojčano odrediti neke karakteristike distribucije frekvencija koje služe za uspoređivanje. Nekoliko karakteristika, od kojih je svaka izražena jednim brojem, u vrlo sažetom obliku prikazuje čitavu distribuciju. S pomoću tih karakteristika upoznajemo pojavu sa svega nekoliko brojeva, pa nije potrebno uzimati u obzir sve frekvencije distribucije*. (10, str. 61)

2. MJERE CENTRALNE TENDENCIJE (SREDNJE VRIJEDNOSTI)

U prethodnom poglavlju govorili smo o uređivanju statističkih skupova. Krajnji rezultat tog postupka kod numeričkih obilježja bio je raspored frekvencija. On će biti predmet našeg interesovanja u nastavku. Rasporedi frekvencija omogućavaju da upoznamo strukturu posmatrane masovne pojave, ali nam ne daju mogućnosti računanja i poređenja. Ako se raspored frekvencija prikaže pomoću brojeva (parametara) opis pojave je mnogo efektniji. Te brojeve (parametre) nazivamo *deskriptivne statističke mjere*. Tu spadaju: mjere centralne tendencije, mjere varijabiliteta i mjere oblika rasporeda. Prilikom opisa pojava služimo se i nekim drugim relativnim brojevima.

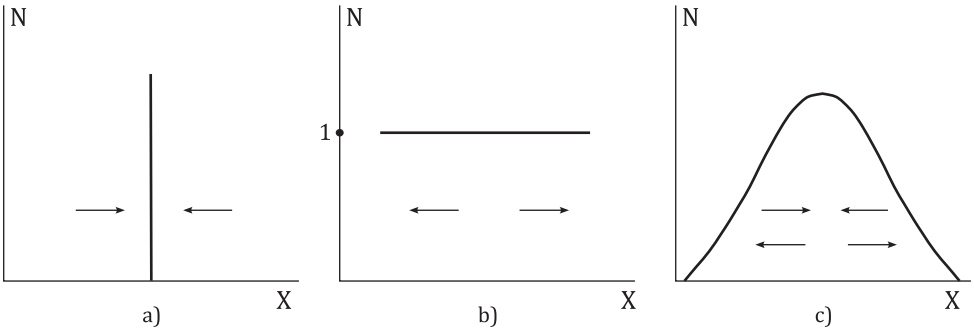
Srednje vrijednosti

Srednje vrijednosti jednim brojem sintetizovano (sumarno) opisuju cijelu pojavu.

Deskriptivne mjere objasnićemo na statističkim skupovima. Na razlike između skupa i uzorka ukazaćemo manjim fontom. Sada kažimo samo da se deskriptivne mjere skupa obično nazivaju parametri skupa (prave vrijednosti), a uzorka parametri uzorka ili „statistika uzorka“. Sotveri koriste izraz „deskriptivna statistika“. Treba imati na umu da statistika uglavnom radi sa uzorcima.

Opšte karakteristike rasporeda frekvencija

Pogledajmo dva krajnja (teorijska) slučaja raspoređivanja frekvencija i oblik koji se najčešće javlja *(Slika 14)*. Na slici *(a)* svi podaci mjerenja su jednaki (visina ordinate pokazuje broj mjerenja). Ovdje je došla do potpunog izražaja težnja (tendencija) elemenata skupa da se grupišu oko jedne vrijednosti (centra). Kao primjer možemo navesti skup cigli. U slučaju *(b)* sve vrijednosti obilježja su različite (frekvencija svakog podatka jednaka je jedan). Ovo je slučaj jedne druge, suprotne težnje (tendencije) koja se javlja u prirodnim pojavama. To je težnja rasipanja (udaljavanja) od centra. Kod svih izmjerenih stabala prečnik je različit.



Slika 14. Tendencije raspoređivanja elemenata

Ove dvije tendencije javljaju se istovremeno. Rezultat njihovog djelovanja je raspored frekvencija koji obično ima oblik zvona *(slučaj c)*. U teoriji je ovaj oblik poznat pod nazivom *normalni* raspored. To je najvažniji teorijski raspored u statistici. O njemu će biti govora u poglavlju „*Teorijski rasporedi*".

Nastojanje jedinica skupa da se grupišu oko jedne vrijednosti (centra) naziva se *centralna tendencija*. Centralnu tendenciju opisujemo sa više parametara. To su, u stvari, njene mjere, koje još nazivamo srednje vrijednosti (sredine). Svrstavamo ih u dvije grupe. Prvu grupu čine računske sredine (aritmetička, harmonijska, geometrijska i kvadratna), a drugu pozicione sredine (medijana i mod).

Prirodno (imanentno) svojstvo jedinica koje čine masovne pojave je da se međusobno razlikuju. To nazivamo *varijabiltet*. I varijabilitet ima svoje mjere. To su: raspon varijacije, interkvartilna razlika, srednje apsolutno odstupanje, varijansa i standardna devijacija, kao apsolutne mjere i koeficijent varijacije, kao relativna mjera.

Za opis rasporeda stvarnih frekvencija potrebne su još mjere za *oblik rasporeda*, tj. mjere odstupanja od normalnog oblika. Odstupanje se javlja u vidu asimetrije (mjera je koeficijent alfa 3) i izduženosti (mjera je koeficijent alfa 4).

Iz praktičnih razloga dalje u statističkoj analizi umjesto cijelih rasporeda koristimo statističke parametre. To se odnosi samo na rasporede frekvencija (numerička obilježja), jer se ostale serije ne mogu matematički obrađivati.

Prvo ćemo obraditi srednje vrijednosti, koje se dijele u dvije grupe:

Računske srednje vrijednosti:

- Aritmetička sredina,
- Harmonijska sredina,
- Geometrijska sredina i
- Kvadratna sredina.

Pozicione srednje vrijednosti:

- Medijana i
- Mod.

2.1. Računske srednje vrijednosti

S obzirom da se vrijednosti mjerenja najviše grupišu oko sredine sve mjere centralne tendencije dobile su naziv srednje vrijednosti ili sredine. One koje računamo iz svih vrijednosti skupa nazivaju se *računske* (*matematičke*, *izračunate*) *srednje vrijednosti*. Među njima najpoznatija i najviše upotrebljavana je aritmetička sredina. To je zbog njene jednostavnosti i matematičkih svojstava. U specifičnim slučajevima koriste se harmonijska, geometrijska i kvadratna sredina.

2.1.1. Aritmetička sredina ($\bar{X}$)

Prosta aritmetička sredina

Aritmetička sredina predstavlja prosječnu (srednju) vrijednost svih podataka u skupu. Poznata je kao „prosjek". To je jedan broj (konstanta), koja predstavlja (zamjenjuje, reprezentuje) sve vrijednosti u skupu. Formula za računanje proste aritmetičke sredine glasi:

$$\bar{X} = \frac{\sum X_i}{N}$$

gdje je:

$\bar{X}$ - aritmetička sredina

X - obilježje

Indeks „i" – redni broj elemenata

N - ukupan broj elemenata

($\sum$ - sigma, znači zbir/sumu)

Ova formula se koristi samo u slučajevima kad sredinu računamo iz pojedinačnih (individualnih, negrupisanih) vrijednosti obilježja. Dakle, prosta aritmetička sredina dobija se tako što se zbir svih vrijednosti podijeli s njihovim brojem. Naziv aritmetička dobila je po pridjevu *aritmetički*, što znači *računski* (*Aritmetika* je grana matematike, koja proučava računske operacije s brojevima).

Aritmetička sredina

Aritmetička sredina je prosječna (srednja) vrijednost svih podataka (mjerenja) u skupu.

Složena aritmetička sredina

Numerička obilježja masovnih pojava imaju mnogo vrijednosti. Zato takve skupove uređujemo, odnosno redukujemo broj podataka. U tim slučajevima za aritmetičku sredinu koristimo prilagođenu formulu, u kojoj se pojavljuju frekvencije (Ni), tj. broj elemenata po klasama (indeksi „i" označavaju redni broj klase). Radi se o složenoj aritmetičkoj sredini. Formula je:

$$\overline{X} = \frac{\sum N_i X_i}{N}$$

Klase imaju svoju širinu (interval) i s njima kao takvim ne možemo dalje matematički raditi. Šta onda da uzmemo za X u brojniku? Logično je da se uzme neka srednja vrijednost koja će predstavljati klasu. Pošto nije praktično izračunavati aritmetičke sredine klasa iz njihovih podataka, uzima se centar, koji se računa kao sredina između donje (X_d) i gornje (X_g) granice klase:

$$\frac{X_d + X_g}{2}$$

Pretpostavka za ovu formulu je normalna raspodjela podataka u klasi. Taj uslov obično nije ispunjen pa tom prilikom činimo grešku. Međutim, iskustvo je pokazalo da ona nema praktičnog značaja. *Napomena*: Redukcijom podataka (formiranjem klasa) trajno gubimo informacije o pojedinačnim vrijednostima.

Ako aritmetičku sredinu računamo *za uzorak* koristimo mala slova u istim formulama:

$$\overline{x} = \frac{\sum x_i}{n} \qquad \overline{x} = \frac{\sum n_i x_i}{n}$$

Važno je ovdje istaknuti da se u literaturi koriste različiti simboli (oznake). Mi smo kod aritmetičke sredine odabrali oznaku $\bar{X}$ (čita se iks nadvučeno ili iks srednje). Česta oznaka za aritmetičku sredinu skupa je μ (malo grčko slovo *mi*). Koristi se još oznaka *M* (od engl. *Mean*). Nju koriste uglavnom softveri. Još se može naći oznaka AS, kao skraćenica od „aritmetička sredina". Pregled simbola dat je na kraju Poglavlja II.

U Formuli za složenu aritmetičku sredinu, kao i kod proste sredine, u računu učestvuju svi podaci. Oni ovdje na rezultat utiču indirektno preko frekvencije klasa, dajući različitu „težinu" svakoj klasi u računu. Tako se podaci na neki način vagaju, pa se ova aritmetička sredina naziva još *vagana sredina*. Ona ima značenje slično kao težište vage u mehaničkom sistemu. Frekvencije služe kao ponderi (lat. *pondus* = teg), pa otuda čest naziv *ponderisana sredina*. Naravno da nije ista „težina" klase ako se u njoj nalazi jedan ili pet elemenata. U drugom slučaju taj podatak (sredina klase) se množi sa pet. Mi ćemo uglavnom koristiti termin ponderisana sredina. Osim apsolutnih frekvencija, kao ponderi mogu se koristiti i neki drugi elementi. O tome više u nastavku.

Kad se koriste ponderi?

Ponderi se koriste u sljedećim slučajevima:

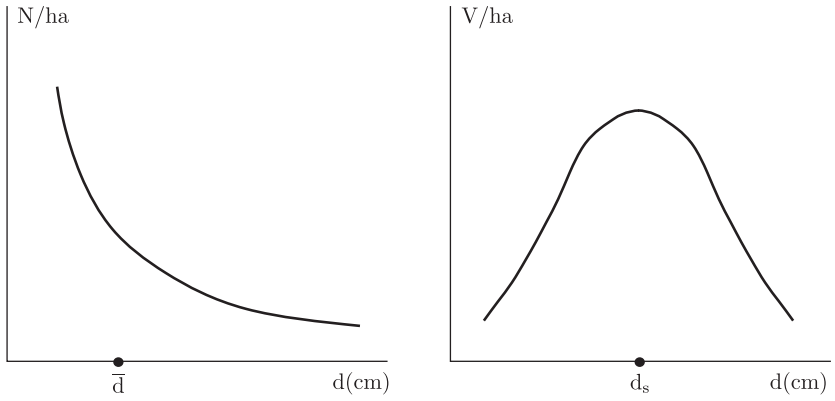
Prvi slučaj. Uvijek kad su podaci grupisani, tj. ako imamo uređene skupove (rasporede frekvencija). Ponderi su frekvencije (broj elemenata po klasama).

Drugi slučaj. Ako treba izračunati zajedničku aritmetičku sredinu iz više podskupova. Ponderi su broj elemenata podskupova. Primjer u knjizi; 3, str. 39.

Treći slučaj. Ako računamo aritmetičku sredinu relativnih brojeva. Razlog je vrlo jednostavan. Svaki relativni broj ima svoju bazu, na osnovu koje je izračunat. Te baze se koriste kao ponderi. Neka je, na primjer, procentualno učešće jele u tri mješovite šume 10%, 15% i 20%. Da li je prosječno učešće jele u ove tri šume 15%? Naravno da nije! Osim u jednom slučaju. Kom? Pogledajte u knjizi (3, str. 40).

Ponderi u šumarstvu

Najčešći raspored frekvencija u šumarstvu je raspored stabala po debljinskim klasama, koji nazivamo *debljinska struktura*. Kakav oblik ovog rasporeda očekujemo u prirodnim šumama? U njima najviše ima tankih stabala, srednje debelih manje, a debelih najmanje. Dakle, oblik raspodjele stabala po debljini je kosi opadajući *(Slika 15a)*. Ako bismo računali aritmetičku sredinu kao pondere trebali bismo uzeti broj stabala po debljinskim klasama. Međutim, ova srednja vrijednost ne bi mogla predstavljati sva stabla u šumi, jer ne postoji centralna tendencija. Za takvu šumu treba nam neka druga srednja vrijednost, koja će je predstavljati. Pogledajmo kako su šumari u BiH riješili ovaj problem.



Slika 15. Broj stabala (a) i drvna masa (b) po debljinskim klasama (u prirodnim šumama)

Kao proizvod u šumarstvu posmatramo drvnu masu (zapreminu) stabala, a ne njihov broj. Ako se uzme u obzir zapremina stabala i njihov broj po debljinskim klasama dolazimo do debljinske strukture zapremine. Ona ima zvonolik (normalni) oblik *(Slika 15b)*. Ovdje vidimo da postoji centralna tendencija. Izračunaćemo *srednji prečnik po zapremini*, koji će dobro predstavljati strukturu zapremine. Koristi se sljedeća formula:

$$d_s = \frac{\sum V_i d_i}{V}$$

u kojoj su:

d_s – srednji prečnik po zapremini

V_i – zapremine po debljinskim klasama

d_i – sredine debljinskih klasa

V – ukupna zapremina ($V=\sum V_i$)

Ovo je, u stvari, formula za ponderisanu aritmetičku sredinu. Kao ponderi uzete su apsolutne vrijednosti zapremine po klasama. Prilikom izrade tablica za naše šume jele, smrče i bukve u BiH, iz praktičnih razloga, uzimane su relativne vrijednosti zapremina po klasama, tj. u procentima. Ta formula glasi:

$$d_s = \frac{\sum p_i d_i}{100}$$

gdje p_i označava procentualno učešće zapremine po klasama. Rezultat je naravno isti. Umjesto zapremina mogu se kao ponderi uzeti temeljnice po

debljinskim klasama. Ovaj srednji prečnik, koji je vezan za temeljnicu, može se dobiti i preko kvadratne sredine. O tome više kasnije.

Matematičke osobine aritmetičke sredine

Iz formule za računanje aritmetičke sredine proizlazi da svi članovi skupa učestvuju u njenom računanju i drugo, zbir svih vrijednosti jednak je proizvodu aritmetičke sredine i ukupnog broja elemenata skupa.

U statistici su posebno važne sljedeće dvije matematičke osobine aritmetičke sredine (dokazi za njih dati su u knjizi; 3, str. 40).

1. Suma odstupanja svih vrijednosti od aritmetičke sredine jednaka je nuli.

$$\sum\left(X_i - \bar{X}\right) = 0$$

2. Suma kvadrata odstupanja od aritmetičke sredine je minimalna. Šta to znači? To znači ako umjesto aritmetičke sredine uzmemo bilo koji drugi broj suma će biti veća.

$$\sum\left(X_i - \bar{X}\right)^2 = \text{MIN}$$

Ovaj princip nazvan je *Metod najmanjih kvadrata*. Otkrio ga je *Gaus* (1795. godine), ispitujući raspoređivanje slučajnih grešaka mjerenja. On je to ovako formulisao: „*Suma kvadrata grešaka mora biti minimalna (ako pod greškom podrazumijevamo odstupanje vrijednosti od aritmetičke sredine*)". Metod najmanjih kvadrata u statistici ima veliku važnost, posebno u regresionoj analizi.

Nedostaci aritmetičke sredine

Aritmetička sredina je najčešća mjera centralne tendencije, iako za njenu primjenu treba biti ispunjeno više uslova. Opišimo ih u nastavku.

1. Aritmetička sredina zahtijeva normalan oblik rasporeda frekvencija. Za kose rasporede nije pogodna kao mjera, jer u tim rasporedima ne postoji centralna već neka druga tendencija. Računanje sredine u takvim slučajevima ima formalni karakter.

2. Drugi uslov je homogenost skupa. Što je variranje podataka veće, to će aritmetička sredina slabije reprezentovati skup. Smatra se da je za koeficijente varijacije veće od 30% aritmetička sredina slaba (nepouzdana) mjera. Takvi skupovi su heterogeni.

3. Na aritmetičku sredinu utiču svi podaci. To je njena dobra osobina. Međutim, ako se u skupu nađu ekstremne vrijednosti one značajnije mogu „povući" prosjek u svoju stranu. Onda takva sredina daje pogrešnu sliku o skupu.

4. Može se desiti da na jednom kraju rasporeda ima više podataka, tj. da je raspored asimetričan. To takođe utiče na reprezentativnost aritmetičke sredine. Granična vrijednost koeficijenta asimetrije je orijentaciono 0,5 (umjerena asimetrija). Sa povećanjem asimetrije aritmetička sredina je sve lošija srednja vrijednost.

5. Računanje aritmetičke sredine je „otežano" ako u rasporedu postoje otvorene klase. Da bismo uopšte mogli izračunati sredinu, te klase moramo stručnom procjenom zatvoriti (odrediti im granicu).

6. U posebnim slučajevima umjesto aritmetičke sredine koriste se harmonijska, geometrijska i kvadratna sredina.

7. Ako radimo sa skupovima aritmetička sredina je tačna bez obzira na veličinu (broj elemenata) skupa. Međutim, kad radimo sa uzorcima, aritmetička sredina iz malog broja podataka je nesigurna. Kao što se slučajne greške mjerenja raspoređuju normalno samo u velikom broju mjerenja tako se i prirodni varijabilitet elemenata (slučajne razlike) raspoređuje normalno samo u velikom broju slučajeva (elemenata). Smatra se da je aritmetička sredina stabilna ako je $n > 30$.

2.1.2. Harmonijska sredina (X_H)

Harmonijska sredina (X_H) koristi se kada pojava i obilježje nisu u skladu (harmoniji). Vjerovatno i naziv ove sredine potiče otuda. Neki autori pak kažu da je naziv vezan za muziku (harmoniju tonova). Nama je najvažnije da prepoznamo situacije kad se ova sredina koristi. Pogledajmo jedan primjer.

Ako sadimo sadnice na nekoj šumskoj površini tada obično govorimo o učinku sadnje. U ovom slučaju kao pojavu posmatramo učinak. Do njega možemo doći na dva načina. Preko pravih vrijednosti, tj. broja zasađenih sadnica (u jedinici vremena) i preko recipročnih vrijednosti, tj. utrošenog vremena za sadnju (po jednoj sadnici). U ovom drugom slučaju, kada se pojava (učinak) smanjuje s povećanjem vrijednosti obilježja (utrošenog vremena), ispravno je računati sredinu tih (recipročnih) vrijednosti. A to je harmonijska sredina.

Harmonijska sredina

Harmonijska sredina se upotrebljava onda kad se pojava smanjuje sa povećanjem obilježja.

Ako se, u našem primjeru, utrošeno vrijeme odnosi na pojedinačne radnike računamo prostu sredinu, a ako su podaci grupisani, tj. ako se odnose na više radnika, onda računamo ponderisanu sredinu. Evo njihovih formula:

$$X_H = \frac{N}{\sum \frac{1}{X_i}} \qquad X_H = \frac{N}{\sum \frac{N_i}{X_i}}$$

Uzmimo da su četiri radnika imali različite utroške vremena (za sadnju po jednoj sadnici), odnosno različit broj zasađenih sadnica (za jedan dan, tj. 8 sati = 480 minuta). Prosječno utrošeno vrijeme po formuli za prostu harmonijsku sredinu je 5,58 minuta.

Radnik:	A	B	C	D
Vrijeme (minuta):	10	6	5	4
Broj sadnica:	48	80	96	120

$$X_H = \frac{N}{\sum \frac{1}{X_i}} = \frac{4}{\frac{1}{10}+\frac{1}{6}+\frac{1}{5}+\frac{1}{4}} = 5,58 \text{ min.}$$

Ako 480 minuta podijelimo s prosječnim utrošenim vremenom dobićemo po harmonijskoj sredini 86 sadnica. Da zaista u prosjeku ova četiri radnika zasade 86 sadnica dnevno potvrđuje prosta aritmetička sredina broja zasađenih sadnica. Da smo izračunali aritmetičku sredinu utrošenog vremena rezultat bi bio 6,25 minuta, a prosjek zasađenih sadnica po radniku 77 (detaljnije u knjizi 3, str. 42).

2.1.3. Geometrijska sredina (X_G)

Ako imamo pred sobom niz podataka, koji pokazuju približno geometrijsku progresiju računaćemo *geometrijsku sredinu* (X_G) . Kod geometrijske progresije svaki naredni član u nizu dobijen je tako što se prethodni množi sa fiksnim brojem. Te odnose između brojeva izravnava geometrijska sredina. Najvažnije je prepoznati takve situacije. Formule za prostu i ponderisanu geometrijsku sredinu, gdje oznaka Π – pi znači proizvod niza brojeva, glase:

$$X_G = \sqrt[N]{\Pi X_i} \qquad X_G = \sqrt[N]{\Pi X_i^{N_i}}$$

Uzmimo opet primjer iz struke. Pretpostavimo proizvodnju trupaca u jednom kratkom periodu od šest godina:

Godina:	prva	druga	treća	četvrta	peta	šesta
Proizvedeno trupaca (u hiljadama m^3):	11,3	14,6	17,5	21,4	29,7	36,2

Ovdje vidimo jedan niz brojeva koji se povećava iz godine u godinu. Njemu odgovara geometrijska sredina. Primijenimo li prostu formulu geometrijska sredina iznosiće 20.120 m^3. Rast proizvodnje može se prikazati i preko indeksa rasta. Njihov prosjek računa se također preko geometrijske sredine (više u knjizi 3, str. 46). Ova sredina izravnava odnose, a ne originalne vrijednosti.

Geometrijska sredina

Ako računamo sredinu niza brojeva koji odgovara geometrijskoj progresiji ispravno je primijeniti geometrijsku sredinu.

U ispravnost primjene geometrijske sredine možemo se uvjeriti na jednom prostom hipotetičkom primjeru. Neka smo neki proizvod u jednom periodu prodavali u pola cijene (relativni iznos 0,5). Nakon toga cijenu smo podigli dva puta (relativni iznos 2,0). Pitanje: da li se cijena našeg proizvoda promijenila? Aritmetička sredina ovih odnosa (0,5 i 2,0) bila bi 1,25. To bi značilo da se cijena povećala za 25%. Međutim, to nije tačno. Cijena je ostala ista, tj. nepromijenjena. Tačan rezultat daje nam geometrijska sredina. Drugi korijen iz 0,5 x 2,0 jednak je jedan.

2.1.4. Kvadratna sredina (X_K)

Kvadratna sredina (X_K), kako i sam naziv kaže, predstavlja sredinu izračunatu iz kvadriranih vrijednosti obilježja (računa se kao drugi korijen). Koristi se u slučajevima kada je pojava povezana sa kvadratima obilježja. Kvadratna sredina daje veći značaj (težinu) većim vrijednostima. Formule glase:

$$X_K = \sqrt{\frac{\sum X_i^2}{N}} \qquad\qquad X_K = \sqrt{\frac{\sum N_i X_i^2}{N}}$$

Ova sredina za nas šumare je posebno važna. Mi kao pojavu često posmatramo temeljnicu šume, a prečnike njenih stabala kao obilježje. Temeljnicu jednog stabla (površinu poprečnog presjeka oblika kruga, na 1,30 m visine) dobijamo računski preko kvadrata prečnika:

$$g = \frac{\pi}{4} d^2$$

Temeljnica pet stabala čiji su prečnici 20, 25, 30, 35 i 40 cm iznosi G = 0,37287 m^2. Ako treba izračunati srednju vrijednost čija će temeljnica pomnožena sa pet dati ovu temeljnicu računamo kvadratnu sredinu:

$$X_K = \sqrt{\frac{\sum X_i^2}{N}} = \sqrt{\frac{20^2 + 25^2 + 30^2 + 35^2 + 40^2}{5}} = 30,82 \text{ cm}$$

Temeljnica prečnika 30,82 cm^2 iznosi 0,074575 m^2. Ona pomnožena sa pet daće ukupnu temeljnicu. Da smo računali aritmetičku sredinu ne bismo dobili ovaj rezultat.

Kvadratna sredina

Kad je pojava vezana sa kvadratima obilježja, a ne s originalnim (pravim) vrijednostima, računa se kvadratna sredina.

Ako temeljnicu šume (G) podijelimo s brojem stabala u šumi (N) dobićemo prosječnu temeljnicu svih stabala u šumi ($\bar{g}$). Srednji prečnik koji odgovara ovoj temeljnici nazivamo srednji prečnik po temeljnici (d_g). Ovaj prečnik korišćen je prilikom izrade tablica za naše borove i hrastove šume. Veza izgleda ovako:

$$\frac{G}{N} = \bar{g} \rightarrow d_g$$

2.1.5. Veza između računskih sredina

Između računskih sredina uvijek postoji stalan odnos, ako se računaju iz istog niza brojeva. Najmanja je harmonijska, a slijede geometrijska i aritmetička sredina. Najveća je uvijek kvadratna sredina.

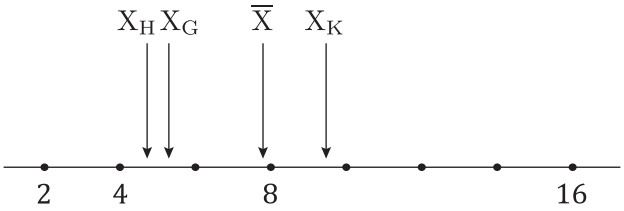
$$X_H < X_G < \bar{X} < X_K$$

Izračunajmo ove sredine iz niza od četiri broja: 2, 4, 8 i 16 i izaberimo sredinu koja bi najbolje odgovarala ovom nizu brojeva.

$$\bar{X} = \frac{\sum X_i}{N} = \frac{2+4+8+16}{4} = 7,50 \qquad X_G = \sqrt[N]{\Pi X_i} = \sqrt[4]{2 \times 4 \times 8 \times 16} = 5,66$$

$$X_H = \frac{N}{\sum \frac{1}{X_i}} = \frac{4}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}} = 4,27$$

$$X_K = \sqrt{\frac{\sum X_i^2}{N}} = \sqrt{\frac{2^2 + 4^2 + 8^2 + 16^2}{4}} = 9,22$$



Harmonijska sredina izravnava recipročne vrijednosti. Zašto je ona najmanja? Zato što su recipročne vrijednosti malih brojeva veće (npr. 1/2 je veća vrijednost od 1/16). One „vuku" prosjek na niže.

Geometrijska sredina izravnava odnose (proporcionalne promjene). U ovom nizu postoji takav odnos po kome je svaki naredni broj dva puta veći od prethodnog (4 je dva puta veći broj od 2, 8 od 4, a 16 od 8). Na ovu sredinu ne utiče veličina brojeva i zato se ona nalazi u pravoj sredini geometrijskog niza. Za ovaj niz odgovara geometrijska sredina, kao najbolja srednja vrijednost.

Aritmetička sredina izravnava prave vrijednosti (apsolutne razlike). Ovdje broj 16 „vuče" prosjek na više, pa je aritmetička sredina veća od geometrijske. On djeluje kao ekstremna vrijednost.

Kvadratna sredina izravnava kvadrate. Kad se brojevi kvadriraju razlika između velikih i malih brojeva se povećava, npr. između 16^2 i 2^2, tj. 256 prema 4 . To još jače povlači kvadratnu sredinu na više. Zato je ona najveća od svih sredina.

2.2. Pozicione srednje vrijednosti

2.2.1. Medijana (X_m)

Medijana (X_m) je centralna vrijednost u skupu. To je vrijednost obilježja koje se nalazi na takvom mjestu da skup dijeli na dva jednaka dijela (jedna polovina elemenata ima manje, a druga veće vrijednosti od medijane). Iskazuje se u jedinicama obilježja, kao i sve druge srednje vrijednosti. Može se računati samo ako su podaci poredani po veličini, odnosno ako je skup uređen.

Medijana je pogodna u slučajevima kad nisu ispunjeni uslovi za primjenu aritmetičke sredine. To su situacije: kada je asimetrija velika, kada u skupu ima ekstremnih vrijednosti i kod otvorenih klasa, posebno kad su u njima velike frekvencije. Međutim, medijana nije podesna za daljnju statističku obradu. Osim toga, ona ne zavisi od promjene vrijednosti obilježja, već samo od njihovog broja.

Medijana

Medijana je centralna vrijednost obilježja, koja dijeli skup na dva jednaka dijela.

Mjesto, odnosno redni broj elementa na kome se nalazi centralna vrijednost daće nam jednostavna formula

$$\frac{N+1}{2}$$

Ako se radi o negrupisanim, po veličini poredanim, podacima imaćemo dva slučaja: sa neparnim brojem podataka (jedan element će biti u sredini) i sa parnim brojem (medijana će biti sredina između dva elementa). Na primjer, ako imamo niz od pet podataka (prečnika stabala): 20, 25, 30, 35 i 40 cm, medijana je 30 cm (treći podatak po redu). Ako dodamo tom nizu prečnik od 40 cm medijana će biti 32,5 cm (sredina između trećeg i četvrtog elementa). Ispod i iznad te vrijednosti nalaze se po tri elementa. Ona je u centru.

U praksi obično radimo sa grupisanim podacima. Prvo, u kumulativnom nizu pronađemo klasu u kojoj se nalazi dobijeni broj po gornjoj formuli. Ta klasa naziva se *medijalna klasa*. Približno, za medijanu se uzima njena sredina. Međutim, nekad se traži njeno preciznije izračunavanje. U tom slučaju koriste se formule, koje suštinski predstavljaju interpolaciju sa susjednim klasama. Tada se pretpostavlja pravougaona raspodjela elemenata unutar klasa. Te formule pogledajte u knjizi (3, str. 53). U našoj praksi one se rijetko koriste.

Medijanu možemo odrediti i grafički. Ako na bilo kojoj kumulanti povučemo paralelu sa X-osom iz N/2 i spustimo okomicu dobićemo vrijednost obilježja koje predstavlja medijanu. Ili, ako na istom grafikonu imamo obje kumulante njihovo presijecište, spušteno na X-osu je medijana.

Da bismo razumjeli suštinsko značenje medijane poslužimo se jednostavnim primjerima. U šumarstvu organizujemo proizvodnju tako da posječenu drvnu masu dovozimo na jedno stovarište. To stovarište treba da zauzme centralno mjesto nekog regiona, kako bi troškovi prevoza bili najmanji. Zato se ono i naziva centralno stovarište. To odgovara značenju medijane.

Ili primjer iz običnog života. Ako u nekom preduzeću iskažemo prosječnu platu svih radnika to ne mora biti realna slika (uspjeha) poslovanja. Može se desiti da veliki broj radnika prima veoma malu platu, a da taj prosjek zapravo „podiže“ manja grupa sa ekstremno velikim platama. Zato bi ovdje medijana bila bolja mjera, jer bi ukazala na tu pojavu.

2.2.2. Mod (X_M)

Mod (X_M) je dominantna vrijednost obilježja u skupu. To je obilježje koje se najčešće javlja, tj. koje ima najveću frekvenciju. Na grafikonu je to najveća ordinata. Ako pogledamo prečnike pet stabala: 20, 25, 30, 35 i 40 cm, koja je njihova modalna vrijednost (mod)? Svaki od tih podataka ima frekvenciju jedan (pojavljuje se samo jednom). U ovom slučaju mod ne postoji. Ako prethodnom nizu dodamo prečnik od 40 cm, ova vrijednost će se pojaviti dva puta. Znači, mod će biti 40 cm.

Mod

Mod je obilježje s najvećom frekvencijom.

U uređenim skupovima ako je obilježje prekidno, a nije grupisano u klase, onda je mod vrijednost obilježja s najvećom frekvencijom. Na primjer, ako najviše ima porodica sa dvoje djece mod je dva. Kod neprekidnih obilježja, prvo se pronađe klasa s najvećom frekvencijom. Ona se naziva *modalna klasa*, a njena sredina će približno predstavljati mod. I ovdje, kao kod medijane, mod se može izračunati preciznije po nekoj od formula (vidi 3, str. 55). *Napomena*: Precizno izračunata vrijednost moda je samo teorijska vrijednost, jer može se desiti da ni jedna vrijednost obilježja u skupu ne bude tolika.

Mod se približno može odrediti na histogramu frekvencija. Spajanjem gornje granice modalne klase s gornjom granicom prethodne klase i donje granice modalne klase s donjom granicom naredne klase nastaje presijecište, čijim spuštanjem na apscisu dobijamo mod.

U šumarstvu nema puno smisla tražiti mod tako precizno. Dovoljno je odrediti modalnu klasu. Na primjer, kažemo da je debljinska klasa 30–35 cm najzastupljenija. Zapravo kod neprekidnih obilježja mod gubi smisao tipične (najčešće) vrijednosti. Zato je najbolje ostati kod modalne klase.

Mod je dobra mjera za jako asimetrične rasporede. Kao i medijana nije osjetljiv na promjene vrijednosti obilježja. Mod ostaje isti sve dok se ne promijeni klasa s najvećom frekvencijom. On zavisi od širine klasa, a ako nije izražen kažemo da je mod neizvjestan. Ponekad konstatujemo postojanje dva vrha na grafičkom prikazu rasporeda frekvencija. Tada kažemo da se radi o bimodalnom skupu. Takav skup je nehomogen. Primjer su dvoetažne (dvospratne) sastojine.

Mod se može određivati i za atributivne serije. Njime se ističe modalitet koji je dominantan. Na primjer, možemo konstatovati da je u nekoj šumi učešće stabala prve klase kvaliteta dominantno. Ili, ako najviše ima studenata s ocjenom osam onda će mod biti osam.

Srednja vrijednost karakteristika je svih vrijednosti numeričkog obilježja koje variraju od jedinice do jedinice u jednoj masi. Srednjom vrijednosti zamjenjuje se svaka individualna vrijednost obilježja, pa, prema tome, srednja vrijednost predstavlja svaku individualnu vrijednost obilježja i skup svih vrijednosti obilježja. Srednja će vrijednost to uspješnije zamijeniti neku individualnu vrijednost obilježja ili je predstavljati što se više ta individualna vrijednost približava vrijednosti sredine, ili, drugim riječima, što manje ta individualna vrijednost odstupa od srednje vrijednosti. Ako se uzmu sumarno sve individualne vrijednosti obilježja, srednja će ih vrijednost to bolje predstavljati što se te vrijednosti manje razlikuju od srednje vrijednosti, što se one više gomilaju oko srednje vrijednosti. (10, str. 93)

3. MJERE VARIJABILITETA

Mjere centralne tendencije (srednje vrijednosti) nisu dovoljne da opišu statistički skup. Zato se uz njih uvijek daju mjere varijabiliteta (raspršenosti, disperzije). Postoji više mjera varijabiliteta. Svrstavamo ih u dvije grupe: apsolutne mjere i relativne mjere.

Apsolutne mjere varijabiliteta su:

- Raspon varijacije,
- Interkvartilni raspon,
- Srednje apsolutno odstupanje,
- Varijansa i standardna devijacija.

Relativne mjere varijabiliteta su:

- Koeficijent varijacije i
- Standardizovano odstupanje

Pogledajmo skupove A, B i C. To su hipotetički podaci, gdje je obilježje (X) prečnik stabala u cm. Ova tri skupa, iako istih aritmetičkih sredina, već na prvi pogled razlikuju se po po varijabilitetu. Ovaj primjer ćemo koristiti da pokažemo računanje navedenih mjera varijabiliteta.

Skup A: 35, 36, 39, 40, 42, 43, 45 ($\bar{X}$ = 40 cm)

Skup B: 1, 5, 10, 40, 70, 75, 79 ($\bar{X}$ = 40 cm)

Skup C: 1, 38, 39, 40, 41, 42, 79 ($\bar{X}$ = 40 cm)

3.1. Apsolutne mjere varijabiliteta

3.1.1. Raspon varijacije (RV)

Raspon varijacije (interval varijacije) predstavlja razliku najveće (X_{MAX}) i najmanje (X_{MIN}) vrijednosti obilježja u nizu.

$$RV = X_{MAX} - X_{MIN}$$

Izračunajmo raspon varijacije, za tri slučaja (A, B, C), u našem primjeru:

RV (A) = 45–35 = 10 cm; RV (B) = 79–1 = 78 cm; RV (C) = 79–1 = 78 cm

Najmanji raspon varijacije ima skup A. Skupovi B i C imaju isti raspon varijacije, ali očigledno je variranje elemenata u njima različito. Raspon varijacije je vrlo gruba mjera varijabiliteta, koja uzima u obzir samo krajnje dvije vrijednosti. One često mogu biti ekstremne, pa dobijamo previsoke vrijednosti raspona varijacije. Dobar primjer za to je skup C, gdje se na oba kraja nalaze ekstremne vrijednosti (to su vrijednosti 1 i 79, koje se jako puno razlikuju od drugih vrijednosti).

Raspon varijacije

Raspon varijacije predstavlja razliku najveće i najmanje vrijednosti u skupu.

Kako u uređenom skupu izračunavamo raspon varijacije? Logika nam nalaže da za X_{MAX} uzmemo gornju granicu najviše klase, a za X_{MIN} donju granicu najniže klase. Tada smo sigurni da nećemo ispustiti neku od vrijednosti, što bi se moglo desiti ako bismo uzeli sredine krajnjih klasa. Nedostaci raspona varijacije su subjektivnost u slučajevima otvorenih klasa, kao i zavisnost od širine klasa.

Raspon varijacije gotovo redovno nalazimo u statističkim izlazima, zajedno sa drugim mjerama varijabiliteta. Uz njega, kao korisne informacije, daju se X_{MAX} i X_{MIN}.

3.1.2. Interkvartilni raspon (IQR)

Interkvartilni raspon (IQR) je zamišljen kao mjera varijabiliteta koja će eliminisati one nedostatke raspona varijacije, koji se odnose na ekstremne vrijednosti i otvorene klase. IQR se računa pomoću kvartila. *Kvartili* dijele skup na četiri jednaka dijela. Ima ih ukupno tri. Prvi (donji) kvartil (Q_1) je vrijednost obilježja ispod koje se nalazi 25%, a iznad 75% elemenata. Treći (gornji) kvartil (Q_3) ima obrnute procente (ispod 75%, a iznad 25%). Drugi (centralni)

kvartil je zapravo jednak medijani (ispod i iznad se nalazi po 50% elemenata). *Interkvartilni raspon* se dobije kao razlika između trećeg i prvog kvartila.

$$IQR = Q_3 - Q_1$$

Interkvartilni raspon se može odrediti na dva vrlo praktična načina: preko medijane i grafički. Kad se određuje preko medijane, prvo se nađe medijana cijelog skupa, a zatim medijane iz nastalih polovina. One predstavljaju kvartile. Grafički se IQR može odrediti iz kumulanti. Povlačenjem paralela iz N/4 i 3N/4 i spuštanjem okomica na apscisu dobiju se Q_1 i Q_3.

Interkvartilni raspon

Interkvartilni raspon je razlika između trećeg i prvog kvartila.

Interkvartilni raspon se kao mjera varijabiliteta obično daje uz medijanu. IQR obuhvata 50% elemenata oko centralne vrijednosti (medijane). Manja njegova vrijednost upućuje na veće grupisanje elemenata, tj. manje variranje. Neki autori umjesto interkvartilnog raspona preporučuju upotrebu *poluinterkvartilnog raspona*, koji obuhvata 25% elemenata oko centralne vrijednosti. Međutim, kod ovih mjera i dalje ostaje kao glavni nedostatak to što ne uzimaju u obzir vrijednosti svih elemenata skupa.

3.1.3. Srednje apsolutno odstupanje ($\bar{D}$)

Prvi pokušaj da se nađe mjera varijabiliteta koja će uzeti u obzir sve vrijednosti u skupu bio je da se izračunaju odstupanja elemenata od aritmetičke sredine. S obzirom da je njihov zbir uvijek jednak nuli, uzeta su odstupanja u apsolutnom iznosu, bez predznaka. Međutim, ova suma zavisi od broja podataka. Što ima više vrijednosti suma je veća. Da bi se otklonio taj nedostatak suma svih odstupanja se dijeli s brojem podataka. Tako se dobije *srednje apsolutno odstupanje*. Formule za negrupisane i grupisane podatke glase:

$$\bar{D} = \frac{\sum \left| X_i - \bar{X} \right|}{N} \qquad \bar{D} = \frac{\sum N_i \left| X_i - \bar{X} \right|}{N}$$

Srednje apsolutno odstupanje za naš skup A pokazuje da prečnici stabala odstupaju od aritmetičke sredine u prosjeku 2,86 cm:

$$\bar{D} = \frac{|35 - 40| + \cdots + |45 - 40|}{7} = 2,86 \text{ cm}$$

Srednje apsolutno odstupanje

Srednje apsolutno odstupanje predstavlja prosječno odstupanje elemenata od aritmetičke sredine, ne uzimajući u obzir njihov predznak.

Postupak izračunavanja srednjeg apsolutnog odstupanja algebarski nije ispravan, tako da se ne može dalje računski koristiti. Zbog toga ova mjera nije našla primjenu u praksi. Ipak, ostaje činjenica da upravo srednje apsolutno odstupanje predstavlja aritmetičku sredinu (pravi prosjek) odstupanja elemenata od aritmetičke sredine.

3.1.4. Varijansa (S^2) i Standardna devijacija (S)

Ako se umjesto apsolutnih odstupanja uzmu kvadrati odstupanja dolazi se do novih mjera varijabiliteta. To su standardna devijacija i varijansa. Standardnu devijaciju formulisao je *Pirson* 1893. godine. Ona se i danas koristi kao osnovna apsolutna mjera varijabiliteta. Pojam varijanse *Fišer* je uveo kasnije, 1918. godine. Ona se pokazala kao pogodna (neophodna) za statističku teoriju.

Varijansa

Ako odstupanja elemenata od aritmetičke sredine kvadriramo sve vrijednosti će biti pozitivne. Pošto i ova suma zavisiti od broja podataka (biće veća što ima više podataka, dodatno i zbog kvadriranja) ona se dijeli brojem podataka. Tako dobijamo prosječno kvadratno odstupanje. Ova mjera varijabiliteta nazvana je *varijansa* (S^2). Ona se može računati iz negrupisanih ili grupisanih podataka. Formule, po definiciji, glase:

$$S^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{N} \qquad S^2 = \frac{\sum N_i\left(X_i - \overline{X}\right)^2}{N}$$

Izračunajmo varijansu po definiciji za skup A:

$$S^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{N} = \frac{(35-40)+\cdots+(45-40)^2}{7} = \frac{80}{7} = 11,42$$

Varijansa nije klasična mjera varijabiliteta. Njen značaj je više teorijski. Ako pogledamo formulu za računanje varijanse zapazićemo da se rezultat dobija u jedinicama obilježja dignutim na kvadrat. Ona ovdje iznosi 11,42 cm^2. Pošto to nema smisla, najčešće izračunatu varijansu ostavljamo kao neimenovan broj. Drugi njen nedostatak za praktičnu upotrebu su suviše visoke vrijednosti.

Varijansa

Varijansa je prosjek kvadriranih odstupanja elemenata od aritmetičke sredine skupa.

Iz prethodnih formula za varijansu mogu se matematički izvesti radne formule. One su pogodne kad se radi ručno, jer ne zahtijevaju računanje odstupanja. Evo jedne takve formule za računanje varijanse, iz negrupisanih i grupisanih podataka:

$$S^2 = \frac{\sum X_i^2}{N} - \bar{X}^2 \qquad S^2 = \frac{\sum N_i X_i^2}{N} - \bar{X}^2$$

Standardna devijacija

Ako vratimo prosječno kvadratno odstupanje na prvi stepen dobićemo standardnu mjeru za varijabilitet (devijaciju). Po tome je ona i dobila naziv *standardna devijacija* (S). Standardna devijacija se iskazuje u originalnim jedinicama obilježja, tj. predstavlja apsolutnu mjeru varijabiliteta. Računa se samo uz aritmetičku sredinu, što se vidi iz njenih formula po definiciji:

$$S = \sqrt{\frac{\sum \left(X_i - \bar{X}\right)^2}{N}} \qquad S = \sqrt{\frac{\sum N_i \left(X_i - \bar{X}\right)^2}{N}}$$

Varijansa i standarna devijacija su matematički povezane. Ako je poznata jedna, druga se lako izračuna. Veza je:

$$S = \sqrt{S^2}$$

Za skup A, za koji imamo izračunatu varijansu standardna devijacija iznosi 3,38 cm. Standardna devijacija je, u stvari kvadratna sredina. Zato je veća od srednjeg apsolutnog odstupanja, koje je iznosilo 2,86 cm.

$$S = \sqrt{S^2} = \sqrt{11,42} = 3,38 \text{ cm}$$

Standardna devijacija

Standardna devijacija je prosječno (linearno) odstupanje elemenata od aritmetičke sredine skupa, računato kao kvadratna sredina.

Standardnu devijaciju možemo grubo procijeniti već nakon prikupljanja podataka. Prema normalnom rasporedu, praktično svi podaci obuhvaćeni su

intervalom od šest standardnih devijacija. Dakle, ako razliku između najvećeg i najmanjeg podatka, tj. raspon varijacije podijelimo sa šest dobićemo približno standardnu devijaciju. Neki autori predlažu da se RV dijeli sa četiri.

Statistiku uglavnom radimo na bazi uzorka. Tada se formule nešto razlikuju. Umjesto sa „N" suma kvadrata se dijeli sa „n-1". O tome će biti više govora kasnije. Sada recimo samo da standardna devijacija izračunata iz uzorka bude nešto manja od one iz skupa. S obzirom da nju koristimo kao procjenu standardne devijacije u skupu uvedena je ova korekcija, za njihovo približavanje. Ako se radi o velikim uzorcima ona nema uticaja na rezultat.

3.2. Relativne mjere varijabiliteta

3.2.1. Koeficijent varijacije (KV)

Pogledajmo sljedeća tri slučaja. Radi se o tri šume koje imaju istu standardnu devijaciju prečnika, ali različite aritmetičke sredine. Trebamo uporediti varijabilitet.

$\bar{X}_1 = 20$ cm $\quad \bar{X}_2 = 40$ cm $\quad \bar{X}_3 = 100$ cm $\quad \bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3$

$S_1 = 10$ cm $\quad S_2 = 10$ cm $\quad S_3 = 10$ cm $\quad S_1 = S_2 = S_3$

Poređenjem standardnih devijacija zaključili bismo da nema razlike u varijabilitetu prečnika u ove tri šume. To je pogrešno. U prvom slučaju aritmetička sredina je mala, tako da standardna devijacija iznosi 50% od njene vrijednosti. U drugom slučaju iznosi 25%, a u trećem svega 10%. Sa povećanjem sredine apsolutno variranje od 10 cm postaje relativno sve manje. U ovakvim slučajevima, kada su apsolutne razlike velike, poređenje varijabiliteta se vrši preko koeficijenta varijacije. Mi smo ga već praktično izračunali.

$\underline{\quad\quad}$ $\quad \underline{\quad\quad}$ $\quad \underline{\quad\quad}$ $\quad \underline{\quad\quad}$

$KV_1 = 50\%$ $\quad KV_2 = 25\%$ $\quad KV_3 = 10\%$ $\quad KV_1 > KV_2 > KV_3$

Najmanje variranje stabala po prečniku je u trećoj šumi, koja ima srednji prečnik 100 cm. Koeficijent varijacije iznosi 10%. *Koeficijent varijacije* je relativna mjera koja pokazuje koliko je variranje u odnosu na aritmetičku sredinu. Koeficijent varijacije (KV) iskazuje se u procentima, zbog lakšeg tumačenja. Jedan procenat varijacije ima uvijek isto značenje (težinu), pa se ovi koeficijenti mogu porediti u svim situacijama. Formula za koeficijent varijacije glasi:

$$KV = \frac{S}{\bar{X}} \times 100$$

Za naš skup A, u ranijem primjeru gdje je s = 3,38 cm, a sredina 40 cm, koeficijent varijacije iznosi 8,45%.

Koeficijent varijacije

Koeficijent varijacije predstavlja prosječno odstupanje elemenata od aritmetičke sredine, iskazano relativno, tj. u procentima.

Za poređenje varijabiliteta koristimo koeficijent varijacije i kada se radi o različitim obilježjima. Na primjer, ako nas zanima da li stabla više variraju po prečniku ili po prirastu drvne mase. Na osnovu standardne devijacije u centimetrima i kubnim metrima to nije moguće zaključiti.

Na šta ukazuje koeficijent varijacije? Što je KV manji grupisanje (zbijenost) elemenata oko sredine je veće. A to znači veću homogenost. Načelno, ako je koeficijent varijacije manji od 30% skupove smatramo *homogenim*, dok za skupove sa većim koeficijentom varijacije od 30% kažemo da su *heterogeni*. Koeficijent varijacije preko 50% ukazuje na odstupanje od normalnog rasporeda, dok koeficijent varijacije manji od 5% može sugerisati na upitnu istinitost podataka.

Već smo rekli da aritmetička sredina, sa povećanjem varijabilnosti (smanjenjem homogenosti), gubi na reprezentativnosti. U statistici je pojam homogenosti skupa posebno važan. Često se za primjenu nekog metoda u statistici kao uslov postavlja homogenost skupova.

3.2.2. Standardizovano odstupanje (z)

Standardizovano odstupanje (z) je relativna mjera varijabiliteta koja se koristi za ocjenjivanje varijabiliteta individualnih (pojedinačnih) podataka u skupu. Dobija se kad odstupanje nekog podatka od aritmetičke sredine podijelimo sa standardnom devijacijom. Formula je:

$$z = \frac{X_i - \bar{X}}{S}$$

Na primjer, zanima nas podatak 45 cm, u skupu A. Ova vrijednost je udaljena od aritmetičke sredine 1,48 standardnih devijacija. Kao neimenovan broj Z je uporediv s drugim podacima. Ako je podatak udaljen od svoje aritmetičke sredine više od tri standardne devijacije može se smatrati ekstremnim podatkom ili grubom greškom. Takvi su npr. podaci 1 i 79, u skupu C.

$$z = \frac{X_i - \bar{X}}{S} = \frac{45 - 40}{3,38} = 1,48$$

4. MJERE OBLIKA RASPOREDA FREKVENCIJA

Ako smo izračunali srednju vrijednost, koja reprezentuje sve podatke u skupu i odredili mjeru varijabiliteta (kvalitet srednje vrijednosti), još uvijek nismo u potpunosti opisali pojavu. Naime, raspored frekvencija neke pojave može da ima različit oblik, što se ogleda u asimetriji i izduženosti rasporeda. Za ove karakteristike rasporeda takođe postoje mjere. One se određuju u odnosu na normalni raspored, kao najčešći opšti oblik (model) raspoređivanja frekvencija.

Centralni momenti

Kod određivanja mjera oblika statistika računa prosječna odstupanja podataka od aritmetičke sredine. Ta odstupanja računaju se sukcesivno, dignuta na r-ti stepen (r = 0, 1, 2, ...). To su *statistički momenti*. S obzirom da se računaju u odnosu na aritmetičku sredinu (centar) nazivaju se *centralni momenti*. Tako postoje momenti: nultog, prvog, drugog itd reda (stepena). Opšta formula za njihovo računanje glasi:

$$M_r = \frac{\sum N_i \left(X_i - \bar{X}\right)^r}{N}$$

Centralni momenat nultog reda (M_0) uvijek je jedan, a prvog reda (M_1) nula. Oni su konstante, pa nisu korisni za dalju upotrebu. Po formuli za momenat drugog reda (M_2) računali smo kvadrate odstupanja podataka od sredine. To je bila varijansa. Dolazimo do momenta trećeg reda (M_3). On služi za mjerenje asimetrije. To je prosjek odstupanja podataka od sredine dignutih na treći stepen. Odstupanja dignuta na četvrti stepen (centralni momenat četvrtog reda, M_4) služi za mjerenje izduženosti rasporeda. Centralne momente nazivamo još prvi momenat, drugi momenat, treći momenat itd. Naziv „*momenat*“ uzet je iz mehanike, gdje pokazuje opadanje sile sa udaljavanjem od njenog centra.

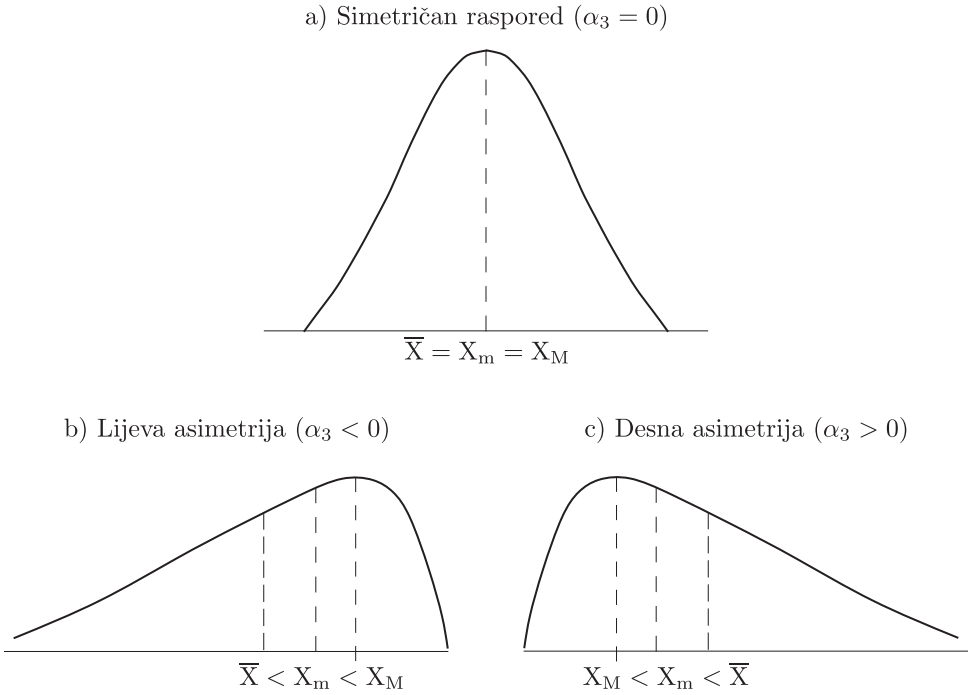
Statistički centralni momenti nisu pogodni za direktno mjerenje oblika rasporeda frekvencija. Razlog su jedinice mjere (dignute na treću i četvrtu), zavisnost od broja podataka i visoke vrijednosti. Zato ih je potrebno relatizovati (standardizovati, transformisati). Tako dobijamo relativne mjere, koje nazivamo koeficijenti alfa 3 i alfa 4.

4.1. Mjera asimetrije (α_3)

Asimetričnost rasporeda frekvencija posmatramo u odnosu na osu simetrije, koju podižemo iz aritmetičke sredine normalnog rasporeda. Kod normalnog (teorijskog) rasporeda negativna odstupanja (s lijeve strane) i pozitivna odstupanja (s desne strane) u odnosu na osu simetrije biće jednaka. U

svakom drugom slučaju, tj. kod stvarnih rasporeda javiće se razlika. Ona će biti u minusu ili plusu. Ova odstupanja dignuta na treći stepen u zbiru predstavljaju *treći momenat* (M_3). Kada njegovu vrijednost podijelimo sa standardnom devijacijom dignutom na treći stepen eliminišemo jedinice mjere. Dobijena relativna mjera naziva se *koeficijent asimetrije* α_3 (alfa tri). Pogledajmo formule i kako izgleda asimetrija grafički *(Slika 16)*.

$$\alpha_3 = \frac{M_3}{S^3} \qquad M_3 = \frac{\sum N_i \left(X_i - \overline{X}\right)^3}{N}$$



Slika 16. Asimetričnost rasporeda

Koeficijent asimetrije (α_3) pokazuje jačinu i smjer asimetrije. On je u stvari relativna vrijednost trećeg centralnog momenta. Predznak koeficijenta α_3 govori o smjeru asimetrije, negativna (lijeva) ili pozitivna (desna), a njegova vrijednost o jačini asimetrije. Obično se smatra umjerenom (srednjom) asimetrijom ako je alfa tri između –0,5 i +0,5.

Koeficijent asimetrije

Koeficijent asimetrije (α_3) pokazuje kakva je asimetrija u odnosu na normalni raspored.

Veza između aritmetičke sredine, medijane i moda

Odnos između ovih srednjih vrijednosti može se iskorititi za grubu ocjenu asimetrije. Kod potpune simetrije (teorijskog oblika rasporeda) aritmetička sredina, medijana i mod su jednaki *(Slika 16a)*. Drugim riječima, prosječna vrijednost svih podataka zauzima centralno mjesto i ujedno ima najveću ordinatu.

$$\bar{X} = X_m = X_M$$

Šta će se desiti ako prethodnom rasporedu dodajemo male vrijednosti? Kriva će se razvući u lijevu stranu *(Slika 16b)*. Mod će ostati isti, medijana će se malo smanjiti (ako bismo dodali dva podatka ona bi se pomjerila samo za jedno mjesto), a aritmetička sredina će se smanjiti više (male vrijednosti „vuku" prosjek naniže). Medijana se uvijek nalazi između sredine i moda. Dobićemo njihov sljedeći odnos:

$$\bar{X} < X_m < X_M$$

Dodavanjem elemenata s desne strane situacija će biti drugačija, medijana se pomjera malo, a aritmetička sredina više (velike vrijednosti „vuku" prosjek naviše). Nastaje desna (pozitivna) asimetrija *(Slika 16c)*. Odnos ovih srednjih vrijednosti je:

$$X_M < X_m < \bar{X}$$

4.2. Mjera izduženosti (α_4)

Ovu mjeru, kao i kod asimetrije, određujemo u odnosu na normalni raspored. Koristimo centralni momenat četvrtog reda. Formula je:

$$M_4 = \frac{\sum N_i \left(X_i - \bar{X}\right)^4}{N}$$

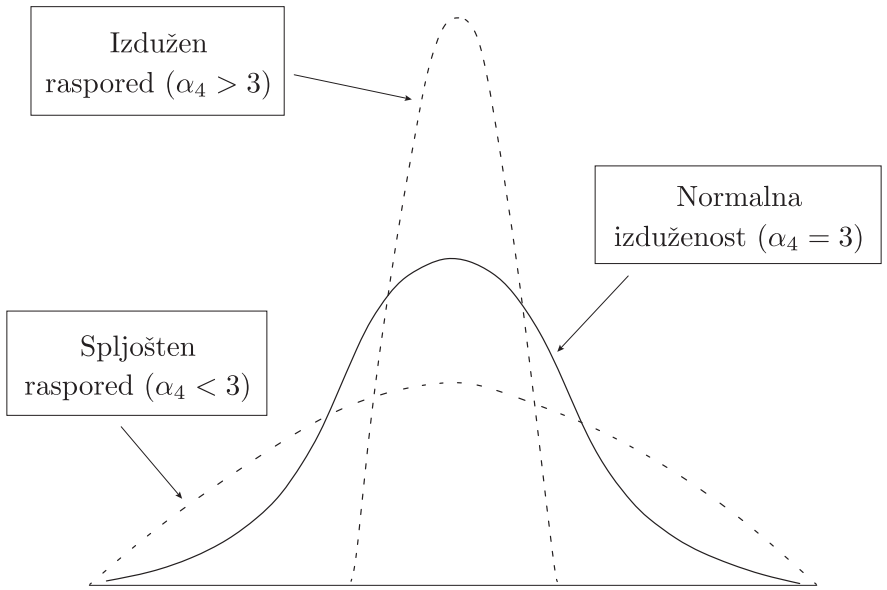
S obzirom da je eksponent paran broj dobićemo uvijek pozitivan zbir. I on je vezan za jedinicu mjere obilježja. Takođe, biće veći što ima više podataka u skupu. Zato ga standardizujemo na sličan način, tj. dijelimo sa standardnom devijacijom dignutom na četvrti stepen. Tako dobijamo *koeficijent izduženosti* α_4 (alfa četiri). On praktično pokazuje visinu rasporeda u odnosu na normalni, što se vidi kao zaobljenost oko moda. Zato se ovaj koeficijent naziva još koeficijent *spljoštenosti* ili *zaobljenosti*. Formula glasi:

$$\alpha_4 = \frac{M_4}{S^4}$$

Kod normalnog rasporeda alfa četiri iznosi tri. Ako je veći od tri to znači da je konkretni raspored izdužen (viši), a ako je manji od tri onda je spljošten (niži), u odnosu na normalni. Pogledajmo to na *Slici 17*. U novije vrijeme softveri u navedenoj formuli oduzimaju tri, kako bi se ovaj koeficijent mogao porediti s nulom. Ovo je važna napomena, jer zaključak iz kompjuterskog izlaza može biti pogrešan.

Koeficijent izduženosti

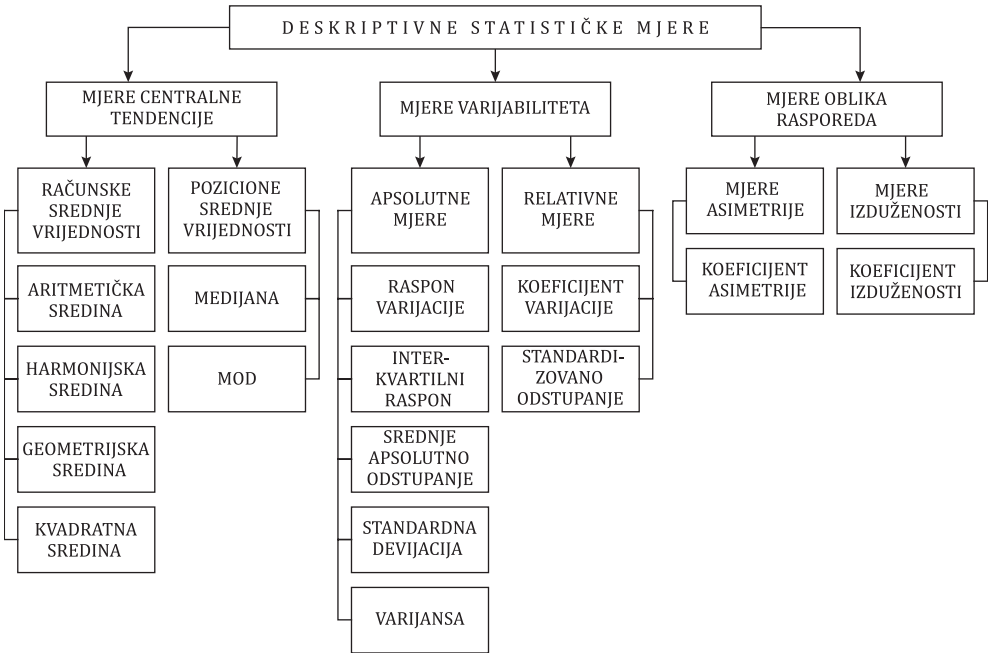
Koeficijent izduženosti (α_4) pokazuje kolika je visina rasporeda u odnosu na normalni raspored.



Slika 17. Izduženost (visina) rasporeda

5. PREGLED DESKRIPTIVNIH MJERA

U opisivanju (deskripciji) pojava, kao što smo vidjeli, koristimo tri vrste statističkih mjera. Šematski smo ih prikazali na *Slici 18*. Pošto smo na kraju poglavlja Deskriptivna statistika osvrnućemo se na najvažnije simbole, skraćenice i sinonime.



Slika 18. Deskriptivne statističke mjere

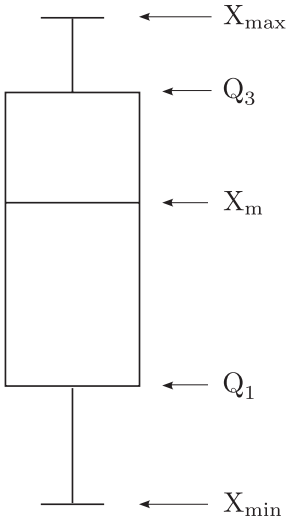
Izbor deskriptivnih mjera

U praksi se nikada ne koriste sve deskriptivne mjere istovremeno. U zavisnosti od karaktera podataka i oblika rasporeda frekvencija biraju se samo one najbolje. U početku uvijek treba dobro pregledati statistički materijal i provjeriti ima li ekstremnih vrijednosti. Načelno, onaj podatak koji je udaljen od aritmetičke sredine za više od tri standardne devijacije, tj. čije je standardizovano odstupanje „z“ veće od tri može se smatrati ekstremnim. Zašto? Zato što je vjerovatnoća pojavljivanja takvog odstupanja (greške) zanemarljivo mala. Takvi podaci zahtijevaju provjeru i eventualno odstranjivanje iz skupa podataka, ako se ispostavi da je u pitanju gruba greška.

Za izbor deskriptivnih mjera presudna je simetričnost rasporeda, pa je na početku potrebno izračunati koeficijent asimetrije (umjerenom

asimetrijom smatra se ako je $-0,5 < \alpha_3 < 0,5$). Za srednju vrijednost najčešće biramo aritmetičku sredinu. Nju uvijek prati standardna devijacija, kao mjera reprezentativnosti (preciznosti). Ukoliko je kao srednja vrijednost izabrana medijana, kao mjera varijabiliteta, računa se interkvartilni raspon.

Ako statističku obradu radimo pomoću računara, većina softvera nudi nam grafikon kutiju, tj. *boxplot (Slika 19)*, koji opisuje raspored sa svega pet brojeva: X_{MIN}, X_{MAX}, Q_1, Q_3 i X_m. Osim toga, on identifikuje ekstremne vrijednosti. Približavanje medijane jednom od krajeva kutije ukazuje nam na odstupanje od normalnog rasporeda. Grafikon kutija se koristi na početku statističke obrade.



Slika 19. *Boxplot*

Oznake (simboli), skraćenice i sinonimi

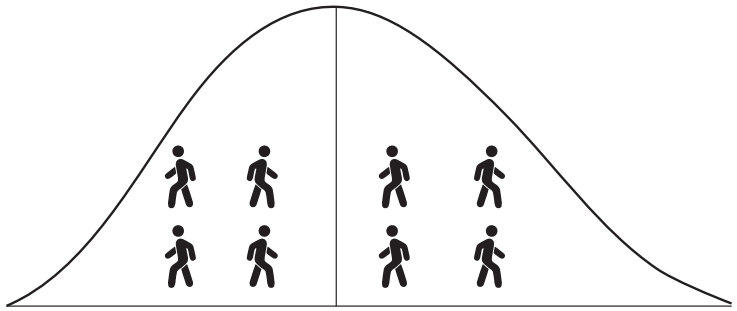
Različite *oznake,* za statističke parametre (deskriptivne mjere), koje se koriste u literaturi često nam otežavaju korištenje (učenje) statistike. Mi smo se odlučili da, umjesto grčkih slova μ (mi) za aritmetičku sredinu i σ (sigma) za standardnu devijaciju, koristimo „naše" uobičajene oznake $\bar{X}$ i S (velika slova za skup, a mala za uzorak). Softverske oznake su M (engl. *Mean*) i SD (engl. *Standard deviation*). Sve srednje vrijednosti „povezali" smo oznakom „X", jer nam je to oznaka za obilježje. Pored aritmetičke sredine, ostale oznake su: X_H – harmonijska, X_G – geometrijska, X_k – kvadratna sredina, X_m – medijana i X_M – mod. Koriste se još oznake od početnih slova: AS – aritmetička, HS – harmonijska, GS – geometrijska, KS – kvadratna sredina, medijana M_e i mod M_o.

Oznake

Aritmetička sredina: $\bar{X}$, μ, M, AS; Standardna devijacija: S, σ, SD

Uobičajene *skraćenice* pojmova koje koristimo u tekstu mogu biti korisne. Takve su: MNK = metod najmanjih kvadrata, CGT = centralna granična teorema, ZVB = zakon veliki brojeva, VRA = višestruka regresiona analiza, ANOVA = analiza varijanse.

Treba imati u vidu i korištenje *sinonima* kao što su: statistički skup = osnovni skup, skup, populacija; obilježje = promjenljiva, varijabla; matematička vjerovatnoća = a priori, unaprijed, teorijska, klasična, logička; statistička vjerovatnoća = a posteriori, unazad, empirijska; objektivnost = nepristrasnost; devijacija = odstupanje; raspored = distribucija; metod uzorka = reprezentativni metod.



Raspored frekvencija je rezultat dvije tendencije:
grupisanja oko centra i udaljavanja od centra

6. PRIMJER

DESKRIPTIVNA STATISTIKA (1)

Uređivanje statističkog skupa

Primjer se odnosi na podatke premjera prečnika stabala hrasta u jednom šumskom odjelu (3, str. 9). Deskriptivnu statistiku počećemo sa uređivanjem skupa, a nastavićemo sa računanjem srednjih vrijednosti, mjera varijabiliteta i mjera oblika rasporeda frekvencija.

Polazni podaci (prečnik u cm) su:

Neuređen skup

63, 46, 37, 42, 28, 34, 26, 50, 29, 48, 46, 38, 32, 42, 41, 43, 46, 39, 42, 48, 53, 60, 52, 39, 38, 42, 40, 39, 45, 37, 38, 48, 44, 46, 46, 51, 38, 34, 32, 58, 37, 40, 43, 41, 24, 27, 39, 33, 42, 41, 40, 31, 26, 42, 33, 39, 53, 47, 39, 38, 34, 44, 37, 39, 17, 23, 39, 50, 51, 49, 37, 49, 59.

Ovo je neuređen statistički skup, koga čine 73 stabla. Svakom od tih stabala (stabla su ovdje elementi skupa) izmjeren je prečnik (prečnik je ovdje obilježje).

Prečnik je, kao neprekidno obilježje, iskazan cijelim brojevima. Kako smo ovdje došli do cijelih brojeva? Jednostavno, izmjerene vrijednosti prečnika smo zaokružili. Zaokruživanje može biti: na niže, na više i na bliže (prema sredini). Pogledajmo sada prvi podatak, koji iznosi 63 cm. Pri premjeru, ovaj prečnik ako je zaokružen „na niže“ obuhvata sve vrijednosti od 63,00 do 64,00 cm (mogli smo uzeti i neki drugi broj decimala, jer to ovdje ništa ne znači). Zaokruživanjem „na više“ ovaj prečnik bi obuhvatao sve vrijednosti od 62,00 do 63,00 cm, dok bi zaokruživanjem „na bliže“ bili obuhvaćeni prečnici od 62,50 do 63,50 cm.

U našem primjeru prečnici su zaokruživani na niže. To je pravilo koje je naša šumarska struka uvela davno, želeći tako doprinijeti očuvanju šuma. Naime, na osnovu nižih prečnika, prilikom utvrđivanja stanja šuma, dobijamo manju zalihu, koja za sobom povlači manji obim sječa.

Prečnik je, često to naglašavamo, osnovni taksacioni element, tj. glavna karakteristika (obilježje) stabala. On nije važan samo za određivanje debljinske strukture šume („krvne slike“ šume), već i za računanje veličine zalihe, debljinske strukture drvne mase i prirasta drvne mase. Prečnici iskazani na cijele centimetre predstavljaju najniže (najmanje, jedinične) brojčane vrijednosti obilježja, pa ih možemo smatrati *debljinskim stepenima*. Objedinjavanjem debljinskih stepena nastaju *debljinske klase*.

Statistički rad nastavljamo korak po korak (kompjuter ove korake obično preskače). Sređivanje podataka možemo započeti tako da ih poredamo po veličini (njihovim vrijednostima).

Serija strukture

17, 23, 24, 26, 26, 27, 28, 29, 31, 32, 32, 33, 33, 34, 34, 34, 37, 37, 37, 37, 37, 38, 38, 38, 38, 38, 39, 39, 39, 39, 39, 39, 39, 39, 40, 40, 40, 41, 41, 41, 42, 42, 42, 42, 42, 42, 43, 43, 44, 44, 45, 46, 46, 46, 46, 46, 47, 48, 48, 48, 49, 49, 50, 50, 51, 51, 52, 53, 53, 58, 59, 60, 63.

Dobili smo jedan niz brojeva (statistički niz) ili seriju (statističku seriju). Kod statičke analize ovakve serije se nazivaju *serije strukture*, jer pokazuju prve karakteristike unutrašnjosti (strukture, sadržaja) skupa. Tako ovdje vidimo najnižu vrijednost (prečnik 17 cm, kao najtanje stablo) i najvišu vrijednost (prečnik 63 cm, kao najdeblje stablo). Još možemo zapaziti da više ima stabala oko sredine nego onih bliže krajevima serije. Nastavljamo uređivanje skupa tako što ćemo uz svaki prečnik napisati broj njegovog ponavljanja. To nazivamo učestalost ili frekvencija.

Raspored frekvencija pojedinačnih vrijednosti

Xi (cm):	17	23	24	26	27	28	29	31	32	33	34	37	38	39	40	41
Ni:	1	1	1	2	1	1	1	1	1	2	2	3	5	5	3	3
	42	43	44	45	46	47	48	49	50	51	52	53	58	59	60	63
	6	2	2	1	5	1	3	2	2	2	1	2	1	1	1	1

Pošto ova statistička serija sadrži sve frekvencije možemo je nazvati rasporedom (distribucijom) frekvencija. Iz nje još bolje uočavamo strukturu skupa. Na primjer, vidimo da se prečnik od 42 cm pojavljuje šest puta (najviše), a krajnji prečnici 17 i 63 cm samo jednom. Ove frekvencije, iskazane brojevima, nazivaju se *apsolutne frekvencije*.

Ovaj raspored frekvencija ne daje konačnu (potpunu) sliku skupa. Naša serija još uvijek nije pregledna, jer sadrži puno (pojedinačnih) vrijednosti obilježja. Njih obično ima više nego u našem primjeru. Zato pristupamo formiranju klasa (klasifikaciji numeričkog neprekidnog obilježja). Ovo je važan korak uređivanja skupova, jer od njega zavisi konačan rezultat. Nakon formiranja klasa slijedi grupisanje (razvrstavanje, klasiranje) elemenata. Uradili smo dvije varijante: klase sa širinom 5 cm (prva varijanta) i klase od 10 cm (druga varijanta).

Prva varijanta — Raspored frekvencija

Xi (cm)	Frekvencije	Ni
15 – 19	/	1
20 – 24	//	2
25 – 29	/////	5
30 – 34	///// ///	8
35 – 39	///// ///// ///// ///	18
40 – 44	///// ///// ///// /	16
45 – 49	///// ///// //	12
50 – 54	///// //	7
55 – 59	//	2
60 – 64	//	2
Suma		73

Druga varijanta — Raspored frekvencija

Xi (cm)	Frekvencije	Ni
15 – 24	///	3
25 – 34	///// ///// ///	13
35 – 44	///// ///// ///// ///// ///// ///// ////	34
45 – 54	///// ///// ///// ////	19
55 – 64	////	4
Suma		73

Frekvencije smo dobili ručnim razvrstavanjem, gdje je svaki podatak predstavljen kosom crtom. Postavlja se sada logično pitanje: koja je varijanta bolja? Vidimo da druga varijanta ima svega pet klasa. U klasi 35–44 cm nalaze se čak 34 elementa, za koje ne znamo kako su raspoređeni unutar klase. Izgubili smo taj dio informacije. Zato ćemo ovu varijantu odbaciti. Prihvatićemo prvu varijantu, koja ima 10 klasa širine 5 cm. Tu smo malo više izgubili na preglednosti, ali zato imamo više informacija (bolju sliku strukture). Da smo uzeli klase uže od 5 cm samo bismo dodatno izgubili na preglednosti. Šire klase od 10 cm nikako ne dolaze u obzir.

Obratimo još pažnju na granice klasa. One su napisane (razgraničene) tako da se tačno zna kojoj klasi svaki od elemenata pripada. Na primjer, prečnik 19 cm bez dileme pripada prvoj klasi, a prečnik 20 cm drugoj. Dalje, prečnici 21, 22, 23 i 24 cm pripadaju drugoj, a naredni (25 cm) trećoj klasi. Važno je imati na umu da su ovo radne granice klasa, koje koristimo samo prilikom razvrstavanja. Već u sljedećem koraku, u tabelarnom prikazu uređenog skupa, prelazimo na prave (tačne) granice klasa.

Da vidimo sada koje su to tačne granice klasa? Prva klasa, 15–19 cm, obuhvata pet brojeva. To su zapravo debljinski stepeni: 15, 16, 17, 18 i 19 cm. Svaki od njih ima širinu jedan centimetar. Na primjer, broj 15 ima širinu od 15,00 do 16,00, dok broj 19 obuhvata sve vrijednosti od 19,00 do 20,00. Donja granica ove klase je 15 cm, a gornja 20 cm. Granice naredne klase su 20 i 25 cm itd. Granice klasa su „oštre“, a samo naizgled se preklapaju. Najbolje se vide na histogramu frekvencija.

Prikažimo konačno naš raspored frekvencija u tabeli, sa pravim granicama klasa. Pored apsolutnih frekvencija (kol. 2) izračunali smo i relativne frekvencije u procentima (kol. 3). Zapazimo da je suma apsolutnih frekvencija jednaka ukupnom broju elemenata ($\sum Ni = N$), a suma relativnih frekvencija je 100%. Izračunali smo, takođe, kumulativne frekvencije ispod i iznad; u apsolutnim iznosima (kolone 4 i 6) i procentima (kolone 5 i 7).

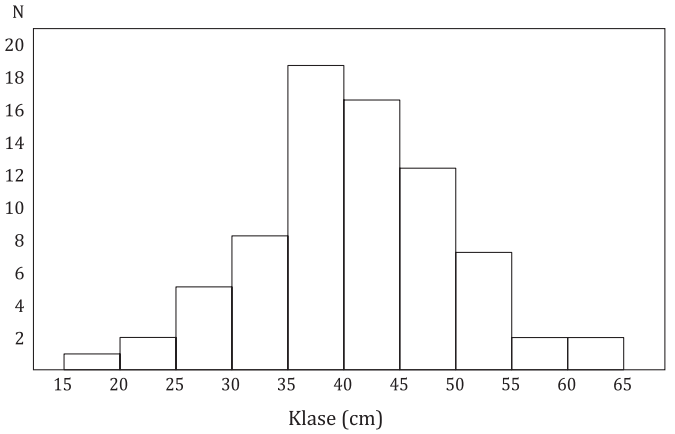
Uređen skup

Klase (cm)	Frekvencije					
	Ni	%	Kumulanta *ispod*		Kumulanta *iznad*	
			Ni	%	Ni	%
1	2	3	4	5	6	7
15 – 20	1	1,36	1	1,36	73	100,00
20 – 25	2	2,74	3	4,10	72	98,63
25 – 30	5	6,85	8	10,96	70	95,89
30 – 35	8	10,96	16	21,92	65	89,04
35 – 40	18	24,66	34	46,57	57	78,08
40 – 45	16	21,92	50	68,49	39	53,42
45 – 50	12	16,44	62	84,93	23	31,51
50 – 55	7	9,59	69	94,52	11	15,07
55 – 60	2	2,74	71	97,26	4	5,48
60 – 65	2	2,74	73	100,00	2	2,74
Suma	73	100,00	-	-	-	-

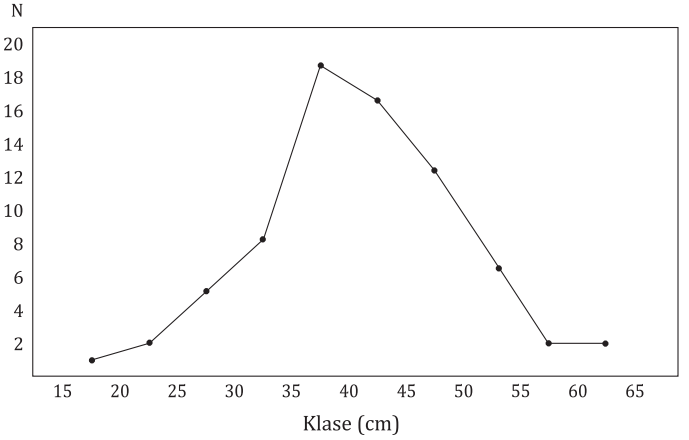
Kumulante se dobijaju sabiranjem frekvencija do određene vrijednosti. Kumulanta *ispod* (rastuća kumulanta) počinje od prve (najniže) klase, a predstavlja sumu frekvencija ispod gornje granice klase na koju se odnosi. Na primjer za klasu 40–45 cm, ispod 45 cm ima 50 elemenata ili 68,49%. Kumulanta *iznad* (opadajuća kumulanta) računa se obrnuto, od zadnje (najviše) klase. Tako u klasi 45–50 cm, čitamo iznad 45 cm imaju 23 elementa ili 31,51%. U zbiru ispod i iznad 45 cm moraju biti svi elementi (73, odnosno 100%). To znači da svaki član kumulativnog niza predstavlja određenu proporciju.

Slijedi nam još grafički prikaz skupa u vidu histograma frekvencija, poligona frekvencija i kumulanti ispod i iznad. *Napomena*: poligon se često „zatvara“ spajanjem sa susjednim klasama, u kojima je frekvencija nula. Slično važi i za kumulante.

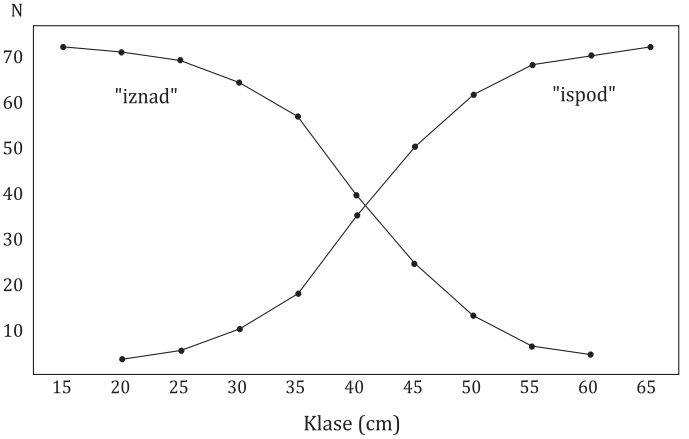
Histogram frekvencija



Poligon frekvencija



Kumulante „*ispod*“ i „*iznad*“



DESKRIPTIVNA STATISTIKA (2)

Srednje vrijednosti skupa

Ovdje nastavljamo obradu našeg primjera. Od računskih srednjih vrijednosti izračunaćemo aritmetičku i kvadratnu sredinu. Medijanu i mod odredićemo samo približno. *Napomena*: U ručnom radu uvijek su nam korisne radne tabele.

Radna tabela

X_i (cm)	N_i	Kumulanta (ispod)	N_iX_i	$N_iX_i^2$
1	2	3	4	5
17,5	1	1	17,50	306,25
22,5	2	3	45,00	1.012,50
17,5	5	8	137,50	3.781,25
32,5	8	16	260,50	8.450,00
37,5	18	34	675,00	25.312,50
42,5	16	50	680,00	28.900,00
47,5	12	62	570,00	27.075,00
52,5	7	69	367,50	19.293,75
57,5	2	71	115,00	6.612,50
62,5	2	73	125,00	7.812,50
Suma	73	-	2.992,50	128.556,25

Aritmetička sredina

Za računanje aritmetičke sredine koristićemo formulu za ponderisanu sredinu, jer imamo uređen skup. Klase u računu predstavljaće njihove sredine (kol. 1). Kao ponderi poslužiće apsolutne frekvencije, tj. broj elemenata po klasama (kol. 2).

$$\overline{X} = \frac{\sum N_i X_i}{N} = \frac{2.992,50}{73} = 40,99 \text{ cm}$$

Aritmetička sredina ovdje nije u potpunosti tačna, jer smo umjesto pojedinačnih vrijednosti obilježja koristili sredine klasa. Da smo izračunali prostu aritmetičku sredinu na osnovu svih podataka (frekvencije bi tada bile jedan, uz svaki broj) dobili bismo srednji prečnik 40,86 cm. Razlika je svega 0,13 cm ili u procentima 0,32%. To je mala greška, koja se u praksi zanemaruje.

Kvadratna sredina

$$X_K = \sqrt{\frac{\sum N_i X_i^2}{N}} = \sqrt{\frac{128556,25}{73}} = 41,96 \text{ cm}$$

Kvadratna sredina u našem rasporedu frekvencija veća je od aritmetičke sredine za oko 1 cm. U šumarstvu se za računanje srednjeg prečnika često koristi ova sredina.

Medijana

Za određivanje medijane poslužiće nam kumulanta ispod (kol. 3). Mjesto (pozicija) medijane, tj. redni broj elementa koji se nalazi u centru, određuje se po formuli:

$$\frac{N+1}{2} = \frac{73+1}{2} = 37$$

Potražićemo klasu u kojoj se nalazi 37. element po redu. U kumulanti „ispod" čija je sredina 37,5 cm, a gornja granica 40,0 cm, nalaze se 34 elementa. Znači da se redni broj 37 nalazi u narednoj (višoj) klasi. To je klasa od 40,0 cm do 45,0 cm. U njoj se nalaze redni brojevi od 35 do 50. Nju uzimamo za medijalnu klasu, a njenu sredinu (X_m = 42,5 cm) kao medijanu.

Mod

Najveću frekvenciju ima klasa 35–40 cm. U njoj ima ukupno 18 elemenata (stabala). Sredinu ove klase uzimamo kao mod (X_M = 37,5 cm). Preciznije određivanje medijane i moda pogledajte u knjizi (3, str. 53 i 55).

Aritmetička (vagana) sredina

DESKRIPTIVNA STATISTIKA (3)

Mjere varijabiliteta

Raspon varijacije

Najmanja vrijednost (donja granica najniže klase) je 15 cm, a najveća (gornja granica najviše klase) je 65 cm. Raspon varijacije iznosi 45 cm. Svi prečnici se nalaze u ovom intervalu. U šumi nema stabala tanjih od 15 cm, niti debljih od 65 cm.

$$RV = X_{MAX} - X_{MIN} = 65{,}0 - 15{,}0 = 45 \text{ cm}.$$

Varijansa i standardna devijacija

Varijansu i standardnu devijaciju izračunali smo na dva načina, po definiciji i radnoj formuli. Prečnici odstupaju od aritmetičke sredine u prosjeku (izračunato kao kvadratna sredina) 8,98 cm.

a) po definiciji

$$S^2 = \frac{\sum N_i \left(X_i - \bar{X}\right)^2}{N} = \frac{5.884{,}26}{73} = 80{,}61$$

$$S = \sqrt{S^2} = \sqrt{80{,}61} = 8{,}98 \text{ cm}$$

b) po radnoj formuli

Radna tabela

X_i	N_i	$N_i X_i^2$	$N_i (X_i - \bar{X})^2$
1	2	3	4
17,5	1	306,25	551,78
22,5	2	1.012,50	683,76
17,5	5	3.781,25	909,90
32,5	8	8.450,00	576,64
37,5	18	25.312,50	219,24
42,5	16	28.900,00	36,48
47,5	12	27.075,00	508,56
52,5	7	19.293,75	927,36
57,5	2	6.612,50	545,16
62,5	2	7.812,50	925,36
Suma	73	128.556,25	5.884,26

$$S^2 = \frac{\sum N_i X_i^{\,2}}{N} - \overline{X}^2 = \frac{128.556,25}{73} - 40,99^2; \quad S^2 = 80,86$$

$$S = \sqrt{S^2} = \sqrt{80,86} = 8,99 \text{ cm}$$

Koeficijent varijacije

Koeficijent varijacije iznosi 21,91%. Toliko u prosjeku odstupaju elementi od svoje sredine. Izračunati koeficijent je manji od 30%, pa zaključujemo da je skup homogen.

$$KV = \frac{S}{\overline{X}} \times 100 = \frac{8,98}{40,99} \times 100 = 21,91\%$$

DESKRIPTIVNA STATISTIKA (4)

Mjere oblika rasporeda frekvencija

Koeficijent asimetrije

$$M_3 = \frac{\sum N_i \left(X_i - \bar{X}\right)^3}{N} = \frac{-594,60}{73} = -8,15$$

$$\alpha_3 = \frac{M_3}{S^3} = \frac{-8,15}{8,98^3} = -0,01$$

Koeficijent izduženosti

$$M_4 = \frac{\sum N_i \left(X_i - \bar{X}\right)^4}{N} = \frac{1.552.376,53}{73} = 21.265,43$$

$$\alpha_4 = \frac{M_4}{S^4} = \frac{21.265,43}{8,98^4} = 3,27$$

Raspored je neznatno lijevo asimetričan (može se reći gotovo idealno simetričan) i malo izdužen u odnosu na normalni raspored. Koeficijente asimetrije i izduženosti izračunali smo po definiciji, uz pomoć radne tabele.

Radna tabela

X_i	N_i	$N_i (X_i - \bar{X})^3$	$N_i (X_i - \bar{X})^4$
1	2	3	4
17,5	1	-12.961,31	304.461,22
22,5	2	-12.642,72	233.763,94
17,5	5	-12.274,55	165.583,69
32,5	8	-4.895,67	41.564,27
37,5	18	765,15	2.670,37
42,5	16	55,08	83.178,05
47,5	12	3.310,73	21.552,82
52,5	7	10.673,91	122.856,75
57,5	2	9.000,59	148.599,76
62,5	2	19.904,49	428.145,66
Suma	73	-594,60	1.552.376,53

III. INFERENCIJALNA STATISTIKA

Karl Friedrich Gauss

(1777-1855)

Opisao statističku teorijsku distribuciju neprekidne promjenljive i metod najmanjih kvadrata

Đirolamo Kardano
(1501- 1576)
Pionir teorije vjerovatnoće

Njegovo djelo "Knjiga o igrama kockom" štampano je 85 godina nakon njegove smrti . Bio je svestran. Bavio se astrologijom i prorekao tačan datum svoje smrti. Ostalo je zapisano da je izvršio samoubistvo taj dan, kako bi dokazao da je bio u pravu.

1. TEORIJSKI RASPOREDI

1.1. Vjerovatnoća

Da smo u stanju izmjeriti svu šumu imali bismo podatke o svim njenim jedinicama. To znači da bi tada utvrđena zaliha i prirast, ali i druga obilježja, odgovarali pravim (stvarnim) vrijednostima. O vjerovatnoći (pouzdanosti) dobijenih rezultata tada uopšte ne bismo govorili.

Međutim, u praksi uglavnom radimo sa statističkim uzorcima, čiji se parametri razlikuju od pravih vrijednosti skupa. Zahvaljujući računu vjerovatnoće omogućeno je prenošenje rezultata sa uzorka na cijeli skup. Vjerovatnoća je iskorištena za izradu teorijskih rasporeda (modela, opštih oblika pojava), pomoću kojih je moguće opisati masovne pojave. O njima govorimo u ovom poglavlju.

Neobični povodi naveli su teoretičare da se počnu baviti vjerovatnoćom. Jedan od njih su igre *kockom* (otuda naziv *kockanje*), a drugi su *greške*, koje nastaju prilikom mjerenja. U vezi s tim poznata imena iz tog vremena (prelaz XVIII u XIX vijek) su Laplas i Gaus. *Laplas* je razvijao teoriju vjerovatnoće na bazi igara kockom, dok su *Gausa* greške mjerenja, na koje je stalno nailazio u geodetskim i astronomskim radovima, natjerale da ih istražuje.

1.1.1. Pojam vjerovatnoće

Ako šest puta bacimo kocku rijetko će se desiti da svaki broj „padne" po jednom. Pad nekog broja dešava se od slučaja do slučaja. Zato takav događaj nazivamo *slučajnim događajem*. Njegovo pojavljivanje povezano je s vjerovatnoćom.

Vjerovatnoća

Vjerovatnoća je mjera za slučajnost događaja. Njoj je pripisan određeni broj, s kojim se dalje može računati kao i s drugim brojevima.

Vjerovatnoću označavamo sa *p*, što potiče od latinske riječi *probabilitas* (vjerovatnoća). Ako broj povoljnih (očekivanih, traženih) slučajeva obilježimo sa „m", a broj ukupno mogućih slučajeva sa „n" formula za računanje vjerovatnoće glasi:

$$p = \frac{m}{n}$$

Vjerovatnoća, pri bacanju kocke, da na gornjoj strani bude broj tri je:

$$p = \frac{m}{n} = \frac{1}{6} = 0{,}17 \text{ ili } 17\%$$

Ukupna vjerovatnoća jednaka je jedan. Prema tome, ako postoji vjerovatnoća da će se desiti neki događaj (npr. da će pasti broj tri), u isto vrijeme postoji vjerovatnoća da se neće desiti taj događaj (da neće pasti broj tri). Tu drugu vjerovatnoću nazivamo *suprotna vjerovatnoća*. Označavamo je sa q. Dakle, možemo napisati:

$$p + q = 1$$

Slijedi da je $q = 1 - p$. Suprotna vjerovatnoća u našem primjeru, da neće pasti broj tri, je q = 1 – 0,17 = 0,83 ili 83%. Nije teško zaključiti da se sve vrijednosti vjerovatnoće nalaze u intervalu od nula do jedan ili od nula do 100%:

$$0 \leq p \leq 1$$

Za događaj čija je vjerovatnoća oko 0,5 kažemo da je neizvjestan. Kad je p = 1 događaj je siguran, a kad je p = 0 događaj je nemoguć. Što je vjerovatnoća bliža jedinici događaj je sve vjerovatniji, a što je bliža nuli događaj je sve manje vjerovatan.

1.1.1.1. Matematička vjerovatnoća

Prethodno smo govorili o jednostavnom događaju, kod koga je vjerovatnoća bila poznata unaprijed. Ona se zato naziva *vjerovatnoća unaprijed* ili *vjerovatnoća a priori* (lat. *a priori* - ono što dolazi prije). Ova vjerovatnoća se još naziva *klasična* ili *logička* vjerovatnoća. Ipak, najčešći naziv je *matematička* vjerovatnoća.

Matematička vjerovatnoća

Matematička vjerovatnoća (p) je količnik broja povoljnih slučajeva (m) i ukupnog broja slučajeva (n):

$$p = \frac{m}{n}$$

Vjerovatnoća pojedinačnih brojeva na kocki je matematička vjerovatnoća. Ako posmatramo vjerovatnoće svih šest brojeva istovremeno, tada već govorimo o rasporedu vjerovatnoća. Radi se o najjednostavnijem teorijskom rasporedu.

1.1.1.2. Statistička vjerovatnoća

Vratimo li se prirodnim pojavama imamo problem kako da u praksi primijenimo matematičku vjerovatnoću. U prirodi (stvarnosti) nalazimo događaje, čiju vjerovatnoću dešavanja ne znamo unaprijed. Mi zapravo nju tek treba da utvrdimo. Ovdje govorimo o jednoj drugoj vrsti vjerovatnoće, koja se naziva *vjerovatnoća unazad* ili *vjerovatnoća a posteriori* (lat. ono što dolazi poslije). Najčešći njen naziv je *statistička vjerovatnoća*. Formula je u suštini ista kao i za matematičku vjerovatnoću.

Statistička vjerovatnoća

Statistička vjerovatnoća (p) je količnik broja ostvarenih slučajeva (apsolutnih frekvencija) (n_i) i njihovog ukupnog broja (n):

$$p = \frac{n_i}{n}$$

Kako dolazimo do statističke vjerovatnoće?

U statističkom radu broj povoljnih slučajeva (n_i) predstavlja apsolutnu frekvenciju događaja, za određeni broj ponavljanja (n). Ako stavimo u odnos n_i/n dobijamo relativnu frekvenciju (oznaka f ili f_r). Ona predstavlja relativnu učestalost događaja, ali se još uvijek ne može uzeti kao mjera njegove slučajnosti, odnosno vjerovatnoća pojavljivanja.

Relativna frekvencija važi samo za *n* (čita se „en“) ponavljanja. S povećanjem *n* relativna frekvencija je sve stabilnija. Ona u jednom momentu postaje *statistička vjerovatnoća*. U praksi se pravi granica između malog i velikog uzorka. U slučaju povećanja uzorka, kad *n* teži u beskonačno, statistička vjerovatnoća postaje fiksna vrijednost, tj. postaje *matematička vjerovatnoća*.

Pogledajmo na jednom hipotetičkom primjeru vezu između relativne frekvencije, statističke vjerovatnoće i matematičke vjerovatnoće *(Tabela 2)*. Uzmimo tri slučaja: kocku bacamo 10 puta (n<30), 60 puta (n>30) i jako mnogo puta ($n \to \infty$). Kod malog broja bacanja relativne frekvencije su jako promjenljive, tj. nepouzdane. To se uočava kod brojeva 2 i 3. Svako novo bacanje utiče na njihove vrijednosti, pa ne znamo sa sigurnošću kakve frekvencije možemo očekivati. Nakon većeg broja bacanja, u našem primjeru 60, relativne frekvencije se donekle stabilizuju. Tek tada se one mogu uzeti kao očekivane, tj. kao statistička vjerovatnoća. Kako vidimo ona je još uvijek daleko od matematičke vjerovatnoće, koja iznosi 0,1667 (za sve brojeve od jedan do šest).

Tabela 2. Primjer vjerovatnoće kod bacanja kocke

	1	2	3	4	5	6	Suma	Naziv
n_i	2	3	0	2	2	1	10	Stvarne frekvencije (n=10)
$\mathbf{f = \frac{n_i}{n}}$	0,2000	**0,3000**	**0,0000**	0,2000	0,2000	0,1000	1	Relativna frekvencija
n_i	8	10	9	12	11	10	60	Stvarne frekvencije (n=60)
$\mathbf{p = \frac{n_i}{n}}$	0,1333	0,1667	0,1500	0,2000	0,1833	0,1667	1	Statistička vjerovatnoća
m	10	10	10	10	10	10	60	Teorijske frekvencije
$\mathbf{p = \frac{m}{n}}$	0,1667	0,1667	0,1667	0,1667	0,1667	0,1667	1	Matematička vjerovatnoća

Ako bismo uporedili statističku vjerovatnoću sa matematičkom vjerovatnoćom kod velikog broja bacanja, npr. 600 bacanja, razlika bi bila veoma mala. Potpuno izjednačavanje ove dvije vjerovatnoće možemo očekivati kod još većeg broja bacanja. Teorijski to kažemo kad *n* teži u beskonačno. Ovo je, u stvari, osnovni zakon u teoriji vjerovatnoće, poznat kao *zakon velikih brojeva*. On glasi ovako: slučajnost u velikom broju slučajeva (ponavljanja) postaje zakon. Šta to u primjeru s kockom znači? Vjerovatnoća svakog od brojeva, iako se oni svaki put slučajno pojavljuju, kod veoma velikog broja bacanja postaće fiksna i iznosiće 0,1667 ili 16,67%. Zakon velikih brojeva se može formulisati i ovako: Razlika između statističke i matematičke vjerovatnoće, pri povećanju broja slučajeva, ima tendenciju smanjivanja.

1.1.2. Vjerovatnoća složenih događaja

Do sada smo govorili o jednostavnim događajima. Više jednostavnih događaja čini jedan složeni događaj. Složeni događaji mogu biti sastavljeni od zavisnih (isključivih) događaja ili od nezavisnih događaja. U prvom slučaju za računanje vjerovatnoće koristi se teorema sabiranja, a u drugom teorema množenja.

1.1.2.1. Teorema sabiranja

Vjerovatnoća složenih događaja, sastavljenih od više jednostavnih međusobno zavisnih (isključivih) događaja, može se predstaviti matematičkim izrazom:

$$P = p_1 + p_2 + \ldots + p_n = \sum_{i=1}^{n} p_i$$

Ovaj izraz poznat je kao *teorema (pravilo) sabiranja* vjerovatnoća. Vjerovatnoća ovakvih složenih događaja naziva se *totalna vjerovatnoća*. Veza između događaja definiše se veznicima *„ili – ili"*. Na primjer, parni broj na kocki je složeni događaj. On je sastavljen od tri jednostavna događaja. To su brojevi 2, 4 i 6, koji se međusobno isključuju (zavisni su jedan od drugog). Ako padne broj dva isključena je mogućnost da padne broj 4 ili broj 6. Totalna vjerovatnoća parnog broja (pri jednom bacanju) kocke je:

$$P = p_2 + p_4 + p_6 = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 0,5 \text{ ili } 50\%$$

Teorema sabiranja

Vjerovatnoća složenog događaja, koji se sastoji od više međusobno isključivih događaja, jednaka je zbiru vjerovatnoća tih događaja.

1.1.2.2. Teorema množenja

Bacanje kocke i novčića zajedno predstavlja složeni događaj, sastavljen takođe od dva jednostavna događaja. U čemu je razlika? Ovdje se radi o nezavisnim događajima. Koji broj na kocki će pasti ne zavisi od novčića i obrnuto. Jedan događaj ne isključuje drugi. Da se ostvari neka kombinacija jednostavnih događaja vjerovatnoća je manja od pojedinačnih vjerovatnoća. Za ovu vezu događaja koriste se veznici *„i – i"*.

U ovakvim slučajevima, kada su događaji nezavisni, do vjerovatnoće se dolazi primjenom *teoreme (pravila) množenja,* čiji matematički izraz glasi:

$$P = p_1 \cdot p_2 \cdot \ldots \cdot p_n = \prod_{i=1}^{n} p_i$$

Teorema množenja

Vjerovatnoća složenog događaja, koji se sastoji iz više međusobno nezavisnih događaja, jednaka je proizvodu vjerovatnoća tih događaja.

Na primjer, vjerovatnoća kombinacije P5, tj. da padne pismo i broj pet, je:

$$P = p_5 \cdot p_p = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

Kombinacija P5 je samo jedna od 12 kombinacija. Njena prosta vjerovatnoća je p = m/n = 1/12. Rezultat potvrđuje teoremu. Moguće kombinacije ovog složenog događaja su:

P1, P2, P3, P4, **P5**, P6, G1, G2, G3, G4, G5, G6

Jedan primjer iz šume

Od 100 pregledanih stabala u jednoj šumi broj registrovanih grešaka (zakrivljenost, račva i druge) po stablu bio je sljedeći:

Broj grešaka	0	1	2	3	≥4	Ukupno
Broj stabala	28	30	24	10	8	100

Odgovorimo na dva pitanja:

a) Kolika je vjerovatnoća da neko stablo ima manje od dvije greške?

Po teoremi sabiranja (veznici „ili-ili“) trebamo sabrati vjerovatnoće pojavljivanja stabala bez greške i stabala s jednom greškom. Radi se o totalnoj vjerovatnoći. Ona iznosi 58%.

$$P = p_0 + p_1 = 0,28 + 0,30 = 0,58 \text{ ili } 58\%$$

b) Kolika je vjerovatnoća da će prva dva slučajno odabrana stabla biti bez greške?

Vjerovatnoća da je prvo odabrano stablo bez greške je 28/100. Da je drugo stablo bez greške vjerovatnoća je također 28/100. Po teoremi množenja (veznici „*i – i*„) vjerovatnoća da će prva dva slučajno odabrana stabla biti bez greške je 7,84%.

$$P = p_0 \cdot p_0 = 0,28 \cdot 0,28 = 0,0784 \text{ ili } 7,84\%$$

Ostali (specifični) događaji

Može se govoriti i o nekim drugim vjerovatnoćama. Na primjer, o *subjektivnoj vjerovatnoći*. Ona se odnosi na specifične događaje i nije egzaktna. Donose je određena lica (eksperti) na bazi znanja, iskustva i dostupnih informacija. Na primjer, iskusni šumari mogu, nakon „loše“ zime (snjegoloma i drugog) dati procjene (vjerovatnoću) broja oštećenih stabala. *Uslovna vjerovatnoća* odnosi se na događaje čija realizacija zavisi od nekog drugog događaja. *Vjerovatnoća uzroka* podrazumijeva određivanje vjerovatnoće nekog događaja na bazi poznate vjerovatnoće njegovih uzroka.

Kakve su nam šanse na lotu 7 od 39?

Da će prvi izvučeni broj biti jedan od naših sedam vjerovatnoća je 7/39 = 17,95%, a da ćemo imati drugi broj je 6/38 = 15,79%. Šanse da bude izvučen prvi broj, kao i drugi broj, nisu tako male. Međutim, po teoremi množenja vjerovatnoća da ćemo imati oba izvučena broja (teorema "*i – i*") je:

$$P = p_1 \cdot p_2 = 0{,}1795 \cdot 0{,}1579 = 0{,}0284 \text{ ili } 2{,}84\%$$

Na kraju, vjerovatnoća da ćemo imati svih 7 brojeva, dobijena množenjem sedam pojedinačnih vjerovatnoća, iznosi 0,000000065 ili 0,0000065%.

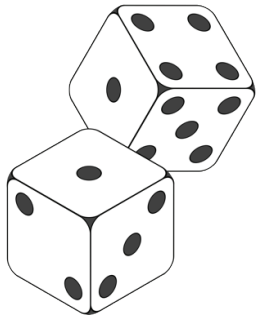
Pitanje za vas. Da li ćete i dalje igrati LOTO?

Zanimljiva (neobična) pitanja o vjerovatnoći

1. Ako u igri "Čovječe ne ljuti se" 20 puta zaredom ne dobijete broj šest, da li ste „ubijeđeni" da naredni put konačno mora pasti šestica? Zašto tako mislite? (Odgovor: vjerovatnoća je uvijek ista i iznosi 1/6)

2. Da li su u pravu oni koji, nakon tri kćerke, konačno očekuju sina? Zašto? (Odgovor: vjerovatnoća je uvijek ista, tj. 50%)

3. Da li na lotu treba izbjegavati brojeve koji su izvučeni u prošlom kolu? Da li biste nekada zaokružili kombinaciju brojeva: 1, 2, 3, 4, 5, 6, 7? Zašto? (Odgovor: svi brojevi imaju istu vjerovatnoću)



1.1.3. Slučajna promjenljiva

Statistički skup (populaciju) definišemo kao skup svih jedinica, a uzorak kao reprezentativni dio skupa. Ovdje ćemo pokušati razjasniti vezu između uzorka i skupa, odnosno njihovih parametara. Na osnovu podataka mjerenja dobijamo raspored frekvencija u uzorku, a njemu odgovarajući teorijski raspored predstavljaće nam statistički skup. Ponovimo još jednom, parametri skupa (pravi, stvarni) ostaju zauvijek nepoznati. Statistika traži put i naučne osnove za najbolju moguću njihovu procjenu. O tome govorimo u nastavku.

Vratimo se našem primjeru s bacanjem kocke *(vidi Tabelu 2)*. Kod bacanja kocke n = 60 puta dobili smo raspored apsolutnih frekvencija, a iz njega raspored relativnih frekvencija. Važno je da znamo da dobijene frekvencije važe samo za ovaj uzorak. Ako bismo uzeli neki drugi, bilo veći ili manji uzorak, relativne frekvencije bi bile drugačije. Kako je uvijek naš cilj skup (populacija), postavlja se pitanje s kojom vjerovatnoćom na bazi uzorka možemo očekivati pojavljivanje brojeva na kocki.

Prvo treba da vidimo šta bi ovdje bio skup. Kocku možemo baciti neograničen broj puta, pa možemo zaključiti da se radi o jednom beskonačnom skupu. Raspored vjerovatnoća u ovom slučaju mi već znamo. Svi brojevi imaju jednaku (istu) vjerovatnoću: p = 1/6 = 0,1667 ili 16,67%. Ovo je matematička vjerovatnoća, kojoj sada možemo pripisati još jedan naziv, a to je *teorijska vjerovatnoća* (nasuprot tome statističkoj vjerovatnoći odgovara naziv *empirijska vjerovatnoća*). Zapravo, pred nama je jedan teorijski raspored (model), koji ima pravougaoni (uniformni, ravnomjerni) oblik. On se ostvaruje kod beskonačno mnogo bacanja, odnosno kad n teži u beskonačno. Dakle, model predstavlja rješenje, tj. zamjenu za skup, jer, kako vidimo, obuhvata sve slučajeve (elemente).

Statističari su razradili veliki broj modela, pomoću kojih opisujemo kvantitativne pojave. Modeli daju (određuju) vjerovatnoće svim mogućim vrijednostima pojave, pri čemu nije moguće unaprijed predvidjeti koja će vrijednost biti ostvarena u nekom konkretnom slučaju. Na primjer, kod kocke znamo vjerovatnoću svih šest brojeva, ali ne možemo predvidjeti koji će broj pasti prilikom narednog bacanja. Dakle, brojevi na kocki su promjenljive veličine. One se javljaju „od slučaja do slučaja". Zato „X" više ne nazivamo obilježje, već *slučajna promjenljiva*. Najpoznatiji teorijski raspored je svakako normalni raspored.

Slučajna promjenljiva

Slučajna promjenljiva je naziv za obilježje u teorijskom modelu.
Njene vrijednosti određene su matematičkom vjerovatnoćom.

Slučajne promjenljive mogu biti prekidne (diskontinuirane) i neprekidne (kontinuirane), slično kao i statistička obilježja. Za njih su razrađeni odgovarajući teorijski rasporedi u vidu prekidnih i neprekidnih funkcija. Prekidne funkcije iskazuju vjerovatnoće pojedinačnih slučajnih promjenljivih, a neprekidne funkcije vjerovatnoće intervala (a,b) u kojima se nalazi slučajna promjenljiva.

1.1.3.1. Prekidna slučajna promjenljiva

Vjerovatnoće (p_i) slučajne promjenljive (x_i) mogu se prikazati kao parovi $x_i p_i$:

$$x_1,\ x_2,\ x_3 \ldots x_n$$

$$p_1,\ p_2,\ p_3 \cdots p_n$$

Ovaj niz zapravo predstavlja raspored prekidne slučajne promjenljive, koji se može prikazati odgovarajućom prekidnom funkcijom. Tako dobijamo teorijski model u kome je svim vrijednostima slučajne promjenljive dodijeljena odgovarajuća vjerovatnoća. Teorijski model može da se prikaže i pomoću kumulativne (zbirne) funkcije, koja daje vjerovatnoću za bilo koju manju (≤) vrijednost slučajne promjenljive. Do totalne (ukupne) vjerovatnoće, koja je jednaka jedan, dolazi se sabiranjem vjerovatnoća pojedinačnih slučajnih promjenljivih, tj.

$$\sum p_i = 1$$

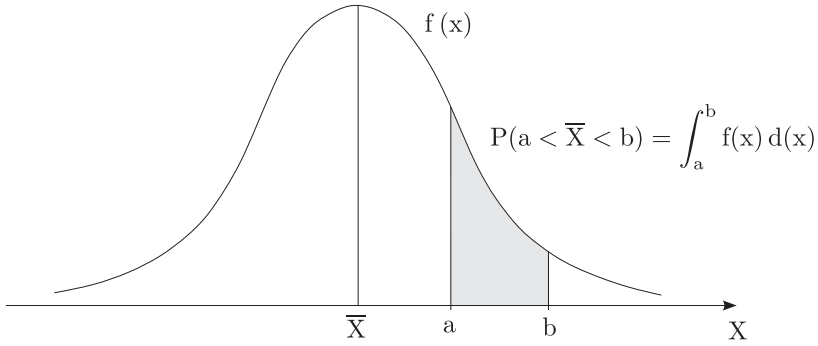
Analogno stvarnim rasporedima frekvencija formulišu se deskriptivne mjere rasporeda vjerovatnoća slučajnih promjenljivih. Značaj se pridaje aritmetičkoj sredini i varijansi. Aritmetička sredina tada odgovara očekivanoj vrijednosti.

1.1.3.2. Neprekidna slučajna promjenljiva

Kod prekidne slučajne promjenljive do rasporeda vjerovatnoća došli smo jednostavno formiranjem liste (niza) pojedinih vrijednosti slučajne promjenljive i odgovarajućih vjerovatnoća. Međutim, kod neprekidne slučajne promjenljive to nije moguće, jer je broj njenih vrijednosti beskonačan, a njihove pojedinačne vjerovatnoće su jednake nuli:

$$p = \frac{1}{\infty} = 0$$

Zbog toga, kod neprekidne slučajne promjenljive ima smisla određivati samo vjerovatnoću da će se njena vrijednost naći u nekom intervalu (a,b). Grafički prikaz kod neprekidne slučajne promjenljive neće biti ordinate (štapići) kao kod prekidne promjenljive, već glatka kriva linija *(Slika 20)*. Ona se iskazuje nekom neprekidnom funkcijom.



Slika 20. Raspored vjerovatnoće neprekidne slučajne promjenljive

Umjesto znaka sigma za sabiranje ovdje se koristi znak integrala (sabiraju se beskonačno male vrijednosti). Ukupna vjerovatnoća jednaka je jedan i srazmjerna je površini ispod krive.

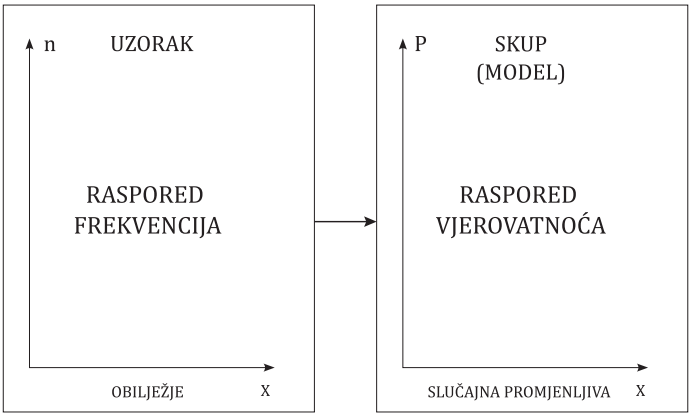
$$\int f_{(x)} d_{(x)} = 1$$

Treba ovdje istaći problem mjerenja neprekidnih obilježja, zbog nedovoljne preciznosti instrumenata i naših čula. Na primjer, mi na terenu ne možemo egzaktno izmjeriti prečnike stabala, već smo prinuđeni da uzimamo njihove približne vrijednosti, obično zaokružene na cijele centimetre. Zato prečnik izgleda kao prekidno obilježje, iako po svojoj prirodi (suštini) to nije. Granice između debljinskih klasa ispravno je pisati na sljedeći način: 5-10, 10-15, 15-20 cm.

Veza između uzorka i skupa

Raspored frekvencija iz uzorka služi nam da dođemo do relativnih frekvencija, tj. empirijske (statističke) vjerovatnoće. Za statističko procjenjivanje parametara skupa, ali i testiranje statističkih hipoteza, koristimo osobine izabranog modela *(Slika 21)*. Prethodno treba da se upoznamo sa najvažnijim teorijskim rasporedima. Za njih se kaže da predstavljaju „ključ" koji otvara vrata inferencijalne statistike.

Stvarni rasporedi se odnose na uzorke. Granični slučaj rasporeda stvarnih vjerovatnoća (relativnih frekvencija) je raspored matematičkih vjerovatnoća (po nekoj funkciji). Izjednačavanje nastupa kad *n* teži beskonačnosti. Zato teorijske rasporede (modele) smatramo populacijom (skupom).



Slika 21. Statističko zaključivanje

1.2. Prekidni rasporedi

Za prekidna obilježja utvrđeni su prekidni rasporedi, kao teorijski modeli koji služe za opisivanje pojava. Svaki od rasporeda (modela) ima svoju formulu po kojoj se računaju vjerovatnoće slučajne promjenljive. Za neke rasporede izrađene su tablice vjerovatnoća (binomski, Poasonov). Teorijske frekvencije (N_t) kod svih rasporeda dobijaju se množenjem vjerovatnoća P(X) sa ukupnim brojem elemenata (N):

$$N_t = P_x \cdot N$$

Najpoznatiji prekidni rasporedi su:

1. Binomski raspored
2. Poasonov raspored
3. Hipergeometrijski raspored
4. Uniformni raspored

1.2.1. Binomski raspored

Binomski raspored je razrađen za prekidne slučajne promjenljive, koje imaju samo dva ishoda. Na primjer: bacanje novčića (pismo i glava), proizvodi (ispravni i neispravni) ili sadnice (primljene i uginule). Obično se jedan ishod uzima kao „uspjeh", a drugi kao „neuspjeh". Vjerovatnoća uspjeha označava se sa „p", a neuspjeha sa „q". Pri tome je p + q = 1. Za otkriće binomskog rasporeda zaslužan je švajcarski matematičar *Bernuli* (Jacob Bernoulli, 1654-1705), pa se po njemu ovaj raspored naziva još Bernulijev raspored.

Primjer binomskog rasporeda

Posmatrajmo tri zasijane sjemenke neke biljke. Svaka sjemenka ima dva isključiva ishoda: da proklija ili da ne proklija. Prvi ishod uzmimo kao *uspjeh*, a drugi kao *neuspjeh*. Uspjeh označimo sa 1, a neuspjeh sa 0. Pretpostavimo dva slučaja klijavosti: u prvom 50%, a u drugom 80%.

U ovom primjeru može nas zanimati recimo kolika je vjerovatnoća da proklijaju sve tri sjemenke. To je jedna od osam mogućih kombinacija (dva ishoda u tri ponavljanja: 2 x 2 x 2 = 8). Sve kombinacije (kol. 1) sa njihovim vjerovatnoćama (kol. 2) prikazane su u *Tabeli 3*. Može se desiti da ne proklija ni jedna sjemenka (X = 0), da proklija jedna (X = 1), dvije (X = 2) ili sve tri sjemenke (X = 3). Prekidna slučajna promjenljiva (X) je broj uspjeha (kol. 3).

Vjerovatnoće slučajne promjenljive (X = 0, 1, 2, 3) date su u opštem obliku, kolona 4. One su dobijene množenjem vjerovatnoća jednostavnih događaja iz kolone 2 (po teoremi množenja). Ove vjerovatnoće su, kao što vidimo u dnu tabele, članovi razvijenog binomnog obrasca $(q + p)^3$. Binomski koeficijenti u ovom slučaju su: 1, 3, 3, 1. Oni pokazuju koliko ima slučajeva određenog broja uspjeha. Izračunali smo vjerovatnoće za klijavost sjemena 50% (kolona 5) i 80% (kol. 6). To su dva odvojena prekidna rasporeda vjerovatnoća. Jasno je već iz ovoga da, u zavisnosti od *n* i *p*, postoji čitava familija binomskih rasporeda.

Tabela 3. Vjerovatnoće klijanja tri sjemenke (n = 3)

Kombinacije elementarnih događaja	Vjerovatnoće kombinacija	Mogući broj uspjeha (X)	Vjerovatnoća uspjeha	P(X) za: p = 0,5 q = 0,5	P(X) za: p = 0,8 q = 0,2
1	2	3	4	5	6
000	qqq	0	q^3	0,125	0,008
001 010 100	qqp qpq pqq	1	$3pq^2$	0,375	0,096
011 101 110	qpp pqp ppq	2	$3p^2q$	0,375	0,384
111	ppp	3	P^3	0,125	0,512
-	-	-	1	1	1
$q^3 + 3pq^2 + 3p^2q + p^3 = (q + p)^3 = 1^3 = 1$					

Vjerovatnoća P(3), tj. da će sve tri sjemenke proklijati, kod klijavosti 50%, je svega 0,125 ili 12,5%. To je, kako smo već rekli, samo jedan od osam mogućih slučajeva (p = 1/8). Isto tolika je i vjerovatnoća da neće proklijati niti jedna sjemenka. U ovom slučaju raspored je simetričan, jer su vjerovatnoće uspjeha i neuspjeha jednake. Kod klijavosti 80%, tri uspjeha (sve tri proklijale sjemenke) možemo očekivati sa vjerovatnoćom od 51,2%, dok je vjerovatnoća za X = 0 svega 0,8%. Ovdje je izražena lijeva asimetrija.

Binomski raspored

Binomski raspored daje vjerovatnoće prekidne slučajne promjenljive (X), u zavisnosti od vjerovatnoće uspjeha *p* u svakom elementarnom događaju od *n* ponavljanja.

Za različite slučajeve prekidnih obilježja definisana je formula koja predstavlja binomski raspored. Ona glasi:

$$P(X) = \binom{n}{x} p^x \; q^{n-x}$$

U formuli je:

$P(x)$ – vjerovatnoća slučajne promjenljive X, gdje je X = 0, 1, 2, ... n

n – veličina uzorka (broj elementarnih događaja - slučajeva, ponavljanja)

p – vjerovatnoća „uspjeha“ u svakom elementarnom događaju

q – vjerovatnoća „neuspjeha“; p + q = 1

$\binom{n}{x}$ – binomski koeficijenti

Izračunavanje vjerovatnoće

Binomski koeficijenti mogu se odrediti po formuli ili očitati iz Paskalovog trougla (pogledajte u knjizi 3, str. 86). *Napomena:* Vjerovatnoća P(X) se može direktno očitati iz tablica binomske raspodjele vjerovatnoća. Ulazi u tablice su *n* i *p* (1, str. 429). Izračunajmo, sada po formuli, vjerovatnoće broja uspjeha (slučajne promjenljive X) u našem primjeru, gdje je n = 3 i p = 0,5.

$$P(X) = \binom{n}{x} p^x \; q^{n-x}$$

$$P(0) = \binom{3}{0} 0,5^0 \, 0,5^{3-0} = 1 \cdot 1 \cdot 0,125 = 0,125$$

$$P(1) = \binom{3}{1} 0,5^1\, 0,5^{3-1} = 3 \cdot 0,5 \cdot 0,5^2 = 0,375$$

$$P(2) = \binom{3}{2} 0,5^2\, 0,5^{3-2} = 3 \cdot 0,5^2 \cdot 0,5 = 0,375$$

$$P(3) = \binom{3}{3} 0,5^3\, 0,5^{3-3} = 1 \cdot 0,125 \cdot 1 = 0,125$$

Teorijske frekvencije

Nakon određivanja vjerovatnoće, lakši dio posla je izračunavanje teorijskih frekvencija. Veza (formula) je ista kod svih drugih rasporeda. Izračunate teorijske vjerovatnoće se jednostavno množe sa ukupnom frekvencijom. Ako smo u našem primjeru imali 40 posudica, sa tri sjemenke u svakoj, onda možemo očekivati pet posudica u kojima nema proklijalih sjemenki, 15 sa jednom, 15 sa dvije i pet sa sve tri proklijale sjemenke.

$$N_0 = P(0) \cdot N = 0,125 \cdot 40 = 5$$

$$N_1 = P(1) \cdot N = 0,375 \cdot 40 = 15$$

$$N_2 = P(2) \cdot N = 0,375 \cdot 40 = 15$$

$$N_3 = P(3) \cdot N = 0,125 \cdot 40 = 5$$

Uslovi binomskog rasporeda

Aritmetička sredina (očekivana vrijednost) i varijansa binomskog rasporeda računaju se po formulama:

$$\bar{X} = n\,p$$

$$s^2 = n\,p\,q$$

Iz navedenih formula proističu uslovi koji se postavljaju, kada se ispituje saglasnost empirijskih frekvencija sa modelom binomskog rasporeda (više u knjizi pod 3, str. 91). Opšte karakteristike binomskog rasporeda su:

1. Radi se o prekidnom obilježju, u čijem nizu posmatranja elementarni događaji (ponavljanja) imaju dva isključiva ishoda (uspjeh i neuspjeh).

2. Vjerovatnoća uspjeha u svakom elementarnom događaju je konstantna.

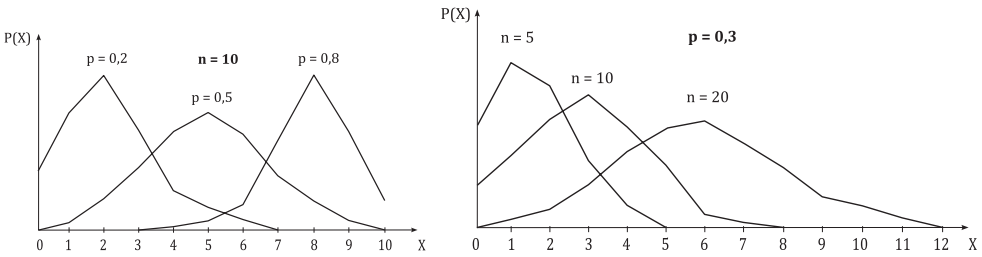
3. Elementarni događaji nastaju na slučajan način i međusobno ne zavise jedan od drugog.

4. Vjerovatnoća uspjeha (p) treba da bude veća od 5%, a veličina uzorka (n) manja od 20, tj.

$$p > 5\%, \quad n < 20$$

Oblik binomskog rasporeda

Ako je vjerovatnoća uspjeha (p) jednaka vjerovatnoći neuspjeha (q) binomski raspored će biti simetričan, bez obzira na veličinu uzorka *n*. Ako je p < 0,5 raspored će biti desno asimetričan, dok će u slučaju p > 0,5 asimetrija biti negativna *(Slika 22a)*.



a) u zavisnosti od *p* b) u zavisnosti od *n*

Slika 22. Oblik binomskog rasporeda

Asimetrija je izraženija što je broj slučajeva (n) manji. Ako se vjerovatnoća uspjeha (p) ne razlikuje puno od 0,5, sa malim povećanjem uzorka binomski raspored dobija normalan oblik. Uglavnom, sa povećanjem n, svaki binomski raspored će biti sve manje asimetričan. Kad je *n* veliki broj, ovaj raspored se približava normalnom rasporedu *(Slika 22b)*. Normalni raspored je granični slučaj binomskog rasporeda, koji kad *n* teži u beskonačno praktično postaje neprekidan raspored.

1.2.2. Poasonov raspored

Binomskim rasporedom, čiji su uslovi: p > 0,05 i n < 20, nisu obuhvaćeni događaji sa jako malom vjerovatnoćom. Takvi rijetki događaji zahtijevaju da u uzorku bude veliki broj elemenata, jer u malom uzorku mogu potpuno da izostanu. Za računanje njihove vjerovatnoće koristi se Poasonov raspored, kojeg je opisao francuski matematičar Poason 1837. godine. Dakle, ako su ispunjeni uslovi:

$$p < 0,05 \qquad n > 20$$

umjesto binomskog može se primijeniti *Poasonov raspored*. Takođe, Poasonovim rasporedom mogu se opisivati rijetke pojave u prostoru i vremenu, kao što su broj kvarova na nekoj mašini (traktoru, žičari), broj povreda šumskih radnika u radu s motornom pilom i druge.

Određivanje vjerovatnoće

Vjerovatnoće se mogu računati na tri načina: direktno po formuli, pomoću radne formule i očitavanjem iz tablica vjerovatnoće Poasonovog rasporeda. Ovaj raspored određen je jednim parametrom *m*. Uzima se da je to aritmetička sredina. Prema tome, u zavisnosti od ovog parametra postoji čitava familija ovog rasporeda.

Formula Poasonovog rasporeda glasi:

$$P(X) = \frac{m^x}{X!} e^{-m}$$

U formuli je:

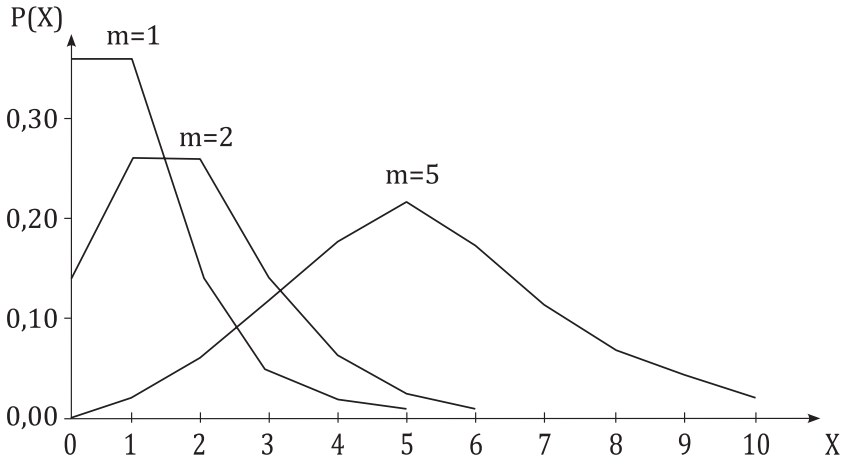
P(X) – vjerovatnoća slučajne promjenljive, X = 0, 1, 2, ...

m – aritmetička sredina

e – baza prirodnih logaritama (e = 2,7182)

Oblik rasporeda

Grafički izgled Poasonovog rasporeda u zavisnosti od *m* dat je na *Slici 23*. Poasonov raspored je desno asimetričan. Sa povećanjem parametra *m* asimetrija se smanjuje, a oblik se približava normalnom rasporedu. Aritmetička sredina i varijansa su približno jednake, što se uzima kao uslov za primjenu ovog rasporeda u praktičnom radu (primjer u knjizi 3, str. 96).



Slika 23. Poasonov raspored u zavisnosti od *m*

1.2.3. Hipergeometrijski raspored

Hipergeometrijski raspored se primjenjuje umjesto binomskog rasporeda u slučajevima kad nije ispunjen uslov nepromjenljivosti vjerovatnoće uspjeha. To se dešava kad su uzorci relativno veliki, tj. kad je stopa izbora veća od 5%. Pomoću ovog rasporeda mogu se takođe utvrđivati vjerovatnoće neke odabrane karakteristike pojava klasifikovanih u dvije grupe elemenata, npr. ispravni i neispravni proizvodi ili zaposleni i nezaposleni.

U vezi s ovim rasporedom preuzeli smo iz knjige s Ekonomskog fakulteta sljedeći citat.

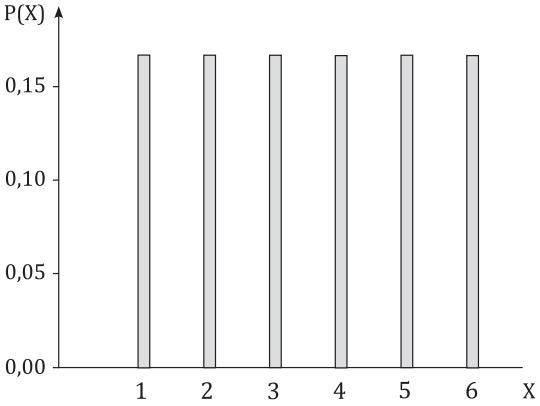
„Vidjeli smo da binomski model zahtijeva da vjerovatnoća uspjeha p mora ostati konstantna iz opita u opit, odnosno opiti moraju biti međusobno nezavisni. Strogo uzevši, elementi uzorka biće međusobno nezavisni (tj. vjerovatnoća p biće konstantna) samo u slučajevima kada uzorak veličine n uzimamo iz beskonačnog osnovnog skupa ili ako iz konačnog skupa uzimamo uzorak sa ponavljanjem. U ovom drugom slučaju svaki element uzorka nakon ispitivanja vraćamo u osnovni skup i pružamo mu mogućnost da ponovo uđe u uzorak. Ukoliko iz konačnog skupa uzimamo uzorak bez ponavljanja uzastopna izvlačenja biće nezavisna i vjerovatnoća p mijenjaće se nakon izbora svakog novog elementa, jer se broj i struktura preostalih elemenata mijenjaju. U praksi se najveći broj istraživanja sprovodi korišćenjem uzorka bez ponavljanja, pa na prvi pogled izgleda da ne bismo smjeli koristiti binomski raspored kada je posmatrana populacija konačna. Međutim, ako veličina uzorka n nije suviše velika u odnosu na veličinu populacije N, binomski raspored se, ipak, može upotrijebiti, jer predstavlja dobru apoksimaciju pravom rasporedu rezultata uzorka. Smatra se da će apsoksimacija biti zadovoljavajuća ako je odnos n/N manji od 5%. Ali u slučaju da primjenjujemo uzorak bez ponavljanja koji sadrži 5% ili više elemenata konačne populacije, moramo koristiti hipergeometrijski raspored.“ (4, str. 143/144).

1.2.4. Uniformni raspored

Uniformni (pravougaoni, ravnomjerni) *raspored* je najjednostavniji prekidni raspored. Kod ovog rasporeda sve vrijednosti slučajne promjenljive imaju istu vjerovatnoću (p = 1/n). Imali smo primjer sa kockom. Ovaj raspored uglavnom se veže za igre na sreću.

Bacanjem pravilne kocke, pri čemu je slučajna promjenljiva X broj tačaka na jednoj strani kocke, dolazimo do uniformnog rasporeda vjerovatnoća. Teorijski oblik ovog rasporeda ostvaruje se tek u jako velikom broju bacanja kocke, prema zakonu velikih brojeva. Pogledajmo njegovu formulu i grafički izgled *(Slika 24)*:

$$P(X) = 1/6 \quad X = 1, 2, \ldots 6$$



Slika 24. Uniformni raspored kod bacanja kocke

1.3. Neprekidni rasporedi

Za neprekidna obilježja koriste se neprekidne funkcije vjerovatnoća. Teorijske frekvencije dobijamo uvijek na isti način, tako što množimo vjerovatnoće sa ukupnim brojem elemenata. Opisaćemo sljedeća četiri rasporeda.

1. Normalni (z) raspored
2. Studentov (t) raspored
3. Fišerov (F) raspored
4. Hi-kvadrat (χ^2) raspored

1.3.1. Normalni (z) raspored

Posmatrajući šta se dešava sa binomskim rasporedom ako se povećava broj ponavljanja, Movr je još 1733. godine pronašao normalni raspored. Ipak, otkriće

normalnog rasporeda pripisuje se Gausu, njemačkom geodeti i astronomu. Proučavajući slučajne greške mjerenja Gaus je 1809. godine došao do funkcije (zakona) njihovog raspoređivanja. Ta funkcija glasi:

$$Y = f(X) = \frac{1}{S\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{X}{S}\right)^2}$$

U vezi s ovim raporedom prenosimo sljedeći citat:

„Gaus je zakonitosti koje se javljaju pri mjerenju neke veličine matematski objedinio jednom krivom linijom, koja je poznata pod nazivom Gausova kriva (vjerovatnoće). Zakonitosti u javljanju slučajnih grešaka su takve da se mjerene vrijednosti gomilaju oko jedne srednje vrijednosti i da ih je sve manje ukoliko se ide dalje od te srednje vrijednosti na jednu ili drugu stranu. Zbog toga pri mjerenju ne nailazimo često na vrijednosti koje se mnogo razlikuju od srednje vrijednosti, pa je i vjerovatnoća da ćemo naići na velika odstupanja od sredine vrlo mala, utoliko manja ukoliko su ta odstupanja veća“ (8, str. 13).

Dvije formule

Direktna veza između biometrike i računa vjerovatnoće je Gausova kriva, Naime, ispitivanja mnogih skupova živih bića su pokazala da se pojedina njihova obilježja ponašaju na isti način kao što se ponašaju slučajne greške mjerenja: gomilaju se oko jedne srednje vrijednosti, a sve ih je manje ukoliko se ide dalje od sredine – bilo na jednu bilo na drugu stranu. Razlika je jedino u tome što su kod grešaka odstupanja vrlo mala, a kod raznih obilježja prirodnih skupova ta odstupanja su veća (8, str. 18).

Ako se Gausova funkcija prilagodi prirodnim pojavama, gdje se neko obilježje mjeri na više individua, dobija se novi oblik funkcije. U njoj se računaju odstupanja individualnih vrijednosti od aritmetičke sredine, a ne od nule kao kod slučajnih grešaka. Formula sada glasi:

$$Y = f(X) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{S}\right)^2}$$

U formuli je:

f (X) –funkcija vjerovatnoće (ordinata za X)

S – standardna devijacija

X – slučajna promjenljiva

$\bar{X}$ – aritmetička sredina

π – 3,1416 (matematička konstanta)

e - 2,7182 (matematička konstanta)

Kada se radi o neprekidnom rasporedu vjerovatnoće se utvrđuju za intervale, a ne za pojedinačne vrijednosti slučajne promjenljive. Sabiranje beskonačno malih vrijednosti u nekom intervalu je zapravo njihovo integrisanje, pa će se u novom obliku funkcije vjerovatnoće normalnog rasporeda pojaviti integral. Ta funkcija izgleda ovako:

$$P(X_1 < X < X_2) = \frac{1}{S\sqrt{2\pi}} \int_{X_1}^{X_2} e^{-\frac{1}{2}\left(\frac{X-\bar{X}}{S}\right)^2} dX$$

Ova funkcija daje vjerovatnoću da će se X naći u određenom intervalu (X_1 – X_2). Geometrijski to odgovara površini intervala ispod normalne krive. Integral od $-\infty$ do $+\infty$ obuhvata ukupnu površinu, odnosno ukupnu vjerovatnoću. Rezultat takve funkcije jednak je jedan.

Napomena: Kako ukupna površina ispod krive predstavlja ukupan broj slučajeva (elemenata) to znači da dijelovi (intervali) predstavljaju relativne frekvencije. Zato možemo površinu ispod krive izjednačiti sa vjerovatnoćom.

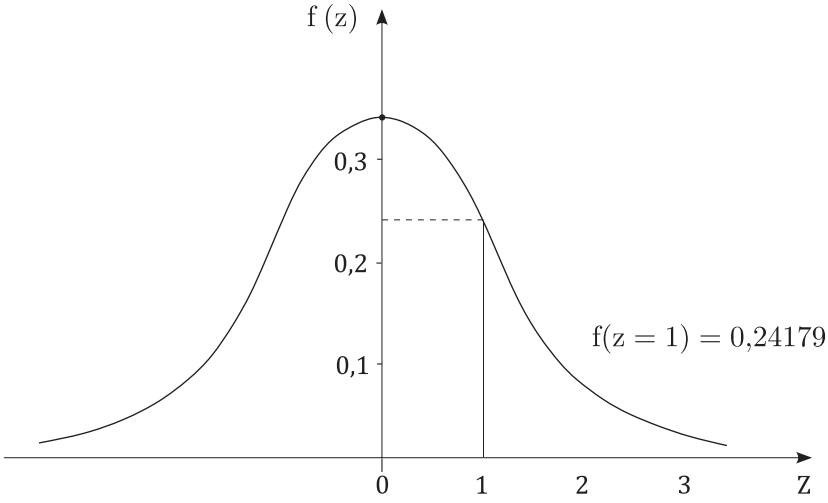
Standardizacija rasporeda

Obje navedene funkcije vjerovatnoće prilagođavaju se za praktičnu upotrebu na način da se originalna promjenljiva *X* standardizuje (transformiše) u promjenljivu *Z*. Time se eliminiše uticaj različitih mjernih jedinica. Osim toga, izbjegava se zamorno računanje vjerovatnoća, jer su aritmetička sredina i standardna devijacija u svakom konkretnom slučaju drugačije. Iz ovoga možemo zaključiti da zapravo postoji čitava familija normalnih (nestandardizovanih) rasporeda. Standardizacijom dobijamo samo jedan normalni raspored, čija je aritmetička sredina nula, a standardna devijacija jedan. Veza je sljedeća:

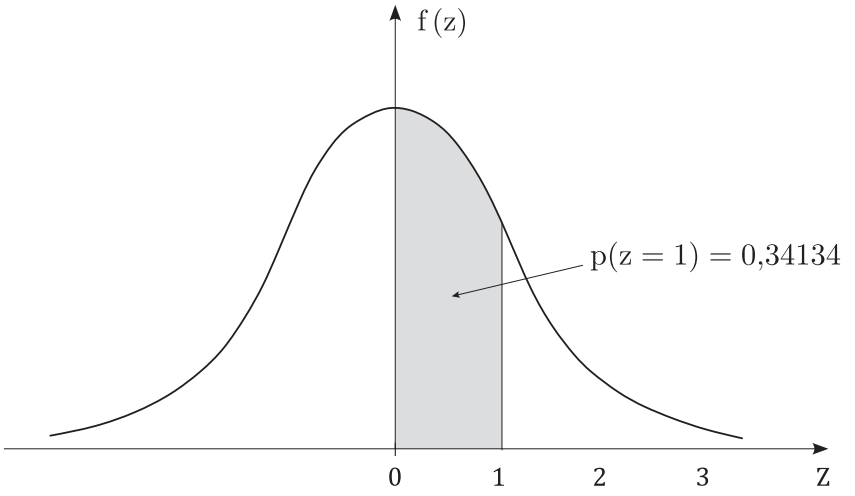
$$Z = \frac{X - \bar{X}}{S}$$

Standardizovana slučajna promjenljiva *Z* pokazuje koliko je odstupanje originalne slučajne promjenljive *X* od aritmetičke sredine, izraženo u jedinicama (broju) standardnih devijacija. To je neimenovan pozitivan ili negativan broj.

Ako u prethodnim funkcijama *X* zamijenimo sa *Z* dobijamo nove oblike funkcije normalnog rasporeda, čije su vrijednosti izračunate i utabličene (3, str. 365/366). Prva funkcija daje ordinate standardne normalne krive *(Slika 25)*, a druga površine standardne normalne krive *(Slika 26)*.

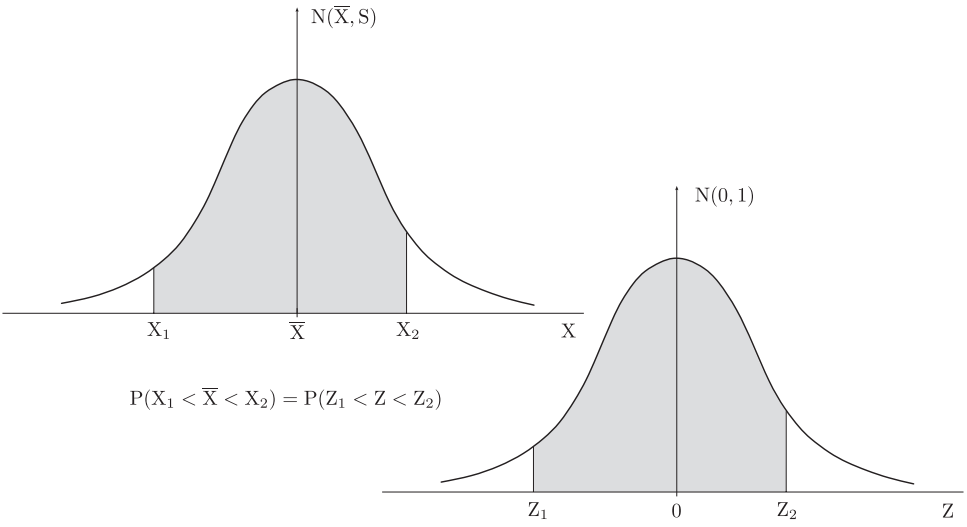


Slika 25. Ordinate standardne normalne krive



Slika 26. Površine standardne normalne krive

S obzirom da postoje dva oblika funkcije vjerovatnoće postoje i dva načina (metoda) računanja teorijskih vjerovatnoća. To su: *Metod ordinata* i *Metod površina*. Oni su detaljno opisani u knjizi (3, str. 104/106). Primjer praktične primjene normalnog rasporeda takođe možete naći u Knjizi (3, str. 109). Pogledajmo kako grafički izgleda povezanost nestandardizovanog i standardizovanog normalnog rasporeda *(Slika 27)*.

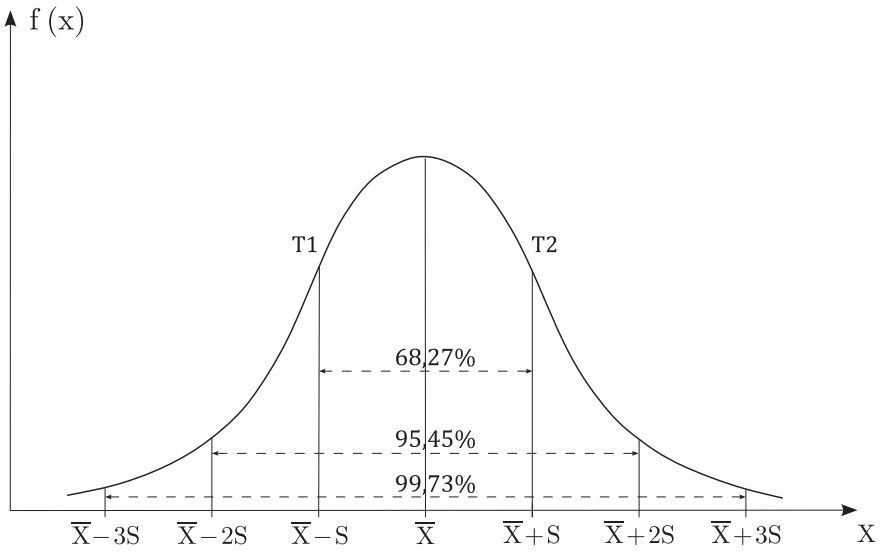


Slika 27. Povezanost normalnog rasporeda N($\bar{X}$, S) i standardizovanog normalnog rasporeda N(0,1)

Karakteristike normalnog rasporeda

Karakteristike normalnog rasporeda najbolje se vide na *Slici 28*. Iste karakteristike ima i standardizovani normalni raspored.

1. Kriva normalnog rasporeda je zvonolikog oblika, unimodalna i potpuno simetrična u odnosu na ordinatu povučenu iz $\bar{X}$. Pozitivna i negativna odstupanja su jednaka; f(-X) = f(X), odnosno P(-X) = P(X). Mala odstupanja su vjerovatnija od velikih.
2. Ukupna površina (vjerovatnoća) ispod krive iznosi jedan, odnosno u procentima 100%. U intervalu jedne, dvije i tri standardne devijacije oko aritmetičke sredine nalazi se: 68,27%; 95,45% i 99,73% površine (vjerovatnoće). To važi za svaki normalni raspored, bez obzira na aritmetičku sredinu i standardnu devijaciju.
3. Aritmetička sredina, medijana i mod su jednaki ($\bar{X} = X_m = X_M$).
4. Koeficijent asimetrije $\alpha_3 = 0$ i koeficijent izduženosti $\alpha_4 = 3$.
5. Normalni raspored je definisan sa aritmetičkom sredinom i standardnom devijacijom. Označava se sa N($\bar{X}$, S), a njegov standardizovani oblik sa N(0, 1).
6. Kriva se asimptotski približava X-osi, što znači da raspon varijacije nije ograničen. Prevojne tačke na krivoj (T1 i T2) udaljene su po jednu standardnu devijaciju lijevo i desno od aritmetičke sredine. Maksimum krive nalazi se iznad aritmetičke sredine.



Slika 28. Normalni raspored

Značaj normalnog rasporeda

1. Veliki broj pojava u prirodi i društvu ima približno normalan raspored.
2. Normalni raspored predstavlja teorijsku osnovu za statističko (parametarsko) zaključivanje. Smatra se najvažnijim teorijskim rasporedom.
3. U mnogim situacijama normalni raspored se može primijeniti kao aproksimacija (približnost) nekih prekidnih rasporeda, koji pod određenim uslovima teže normalnom rasporedu (npr. binomski, Poasonov).
4. Iz normalnog rasporeda izvedeni su važni neprekidni rasporedi, kao Studentov, Fišerov i Hi – kvadrat raspored.
5. Normalni raspored uvijek ostaje teorijski oblik, kome se samo približavamo. To nam dopušta da koristimo njegova svojstva.

Uslovi normalnog rasporeda

Empirijskom rasporedu odgovara normalni raspored ako su ispunjena tri uslova:

1. Da je obilježje neprekidno.
2. Da su aritmetička sredina, medijana i mod približno jednaki.
3. Da je približno koeficijent asimetrije $\alpha_3 = 0$, a koeficijent izduženosti $\alpha_4 = 3$.

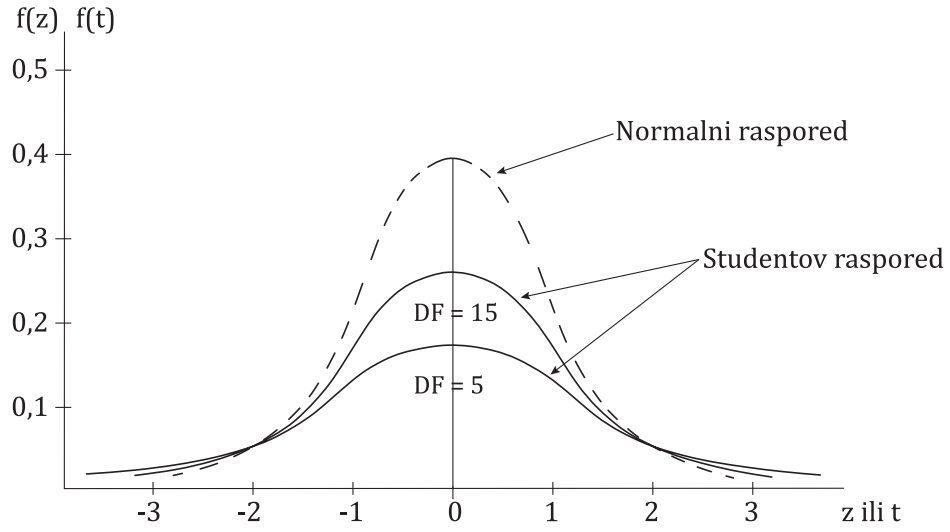
1.3.2. Studentov (t) raspored

Ako posmatramo uzimanje više uzoraka, neke određene veličine od *n* jedinica, njihove sredine (kao i drugi parametri) će se grupisati oko aritmetičke sredine skupa. Raspored će biti normalan, ako uzorak ima više od 30 jedinica, a imaće oblik t-rasporeda u slučaju manjih uzoraka. Ove rasporede (*z* i *t*) najčešće koristimo kao modele za statistička zaključivanja.

Raspored malih uzoraka oko prave vrijednosti (stvarne vrijednosti u skupu) opisao je i sačinio tablicu površina engleski statističar *Goset* (William Gosset, 1876-1937). On je svoje radove potpisivao pod pseudonimom *Student*, pa se ovaj raspored još naziva Studentov raspored.

Karakteristike t-rasporeda

Ovaj raspored je simetričan u odnosu na svoju sredinu, koja je jednaka nuli. Od standardizovanog normalnog rasporeda razlikuje se po varijabilitetu. Kod z-rasporeda standardna devijacija je fiksna i iznosi jedan, dok je kod t-rasporeda veća i zavisi od veličine uzorka *(Slika 29)*. Kod uzorka od 30 elemenata smatra se da nema razlike između ova dva rasporeda. Što je uzorak manji ovaj raspored je sve niži (spljošteniji) u odnosu na normalni raspored. Koristi se kao osnovna baza za testiranje razlike parametara.



Slika 29. Studentov (t) raspored

Tablica t-rasporeda

Tablica površina t-rasporeda data je u Knjizi (3, str. 368). Za razliku od normalnog rasporeda ova tablica ima dva ulaza. To su veličina uzorka (iskazana

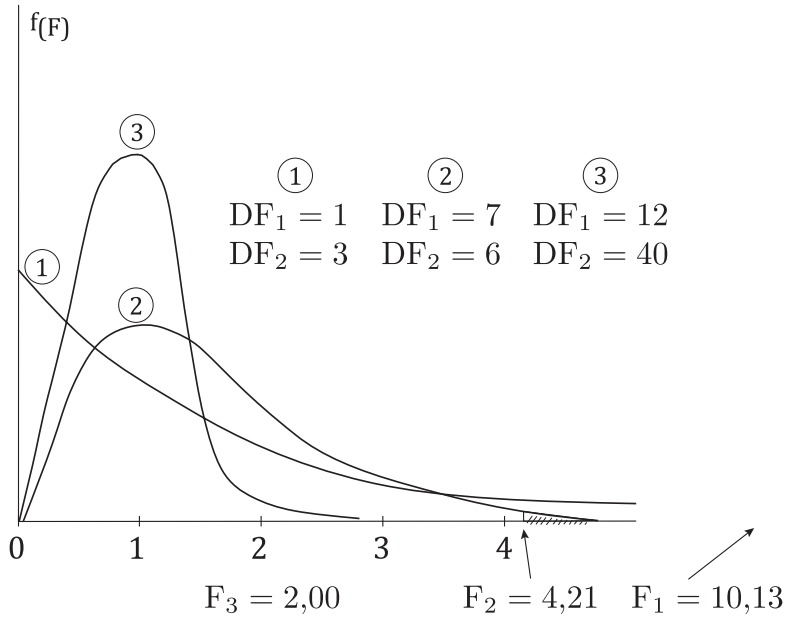
brojem stepena slobode *n–1*, u pretkoloni) i vjerovatnoća (iskazana greškom – kao „vjerovatnoća za veće vrijednosti", u zaglavlju). *Napomena*: pojmove koje pominjemo ovdje objasnićemo kod primjene rasporeda u narednim poglavljima.

Ako je uzorak jako mali, recimo tri elementa (n–1=2 stepena slobode) t-vrijednost je 4,30, kod n=15 vrijednost je 2,14, dok kod n=30 iznosi 2,04. Ona se približava normalnom rasporedu. Teoretski, kad *n* teži beskonačnoj vrijednosti ova dva rasporeda se potpuno poklapaju, tj. z = t = 1,96. Izvan ovog intervala nalazi se 5% površine, ravnomjerno raspoređene na krajevima krive (po 2,5%).

1.3.3. Fišerov (F) raspored

Ispitujući odnos (količnik) dvije varijanse u masi slučajeva Fišer je došao do zakona njegovog raspoređivanja. Ipak, ovaj raspored u konačnom obliku definisao je Džordž Snedekor. U znak poštovanja prema Ronaldu Fišeru nazvao ga je *Fišerov* ili *F-raspored*. Primjenjuje se za male uzorke, najčešće kod *analize varijanse*, pri izvođenju statističkih eksperimenata.

F-raspored je desno asimetričan *(Slika 30)*. Sa povećanjem veličine uzorka asimetrija se smanjuje i približava normalnom obliku. Vrijednosti *f(F*) su pozitivne, tj. kriva počinje od nule. U Knjizi (3, str. 369/370) date su tablice za vjerovatnoću 95% i 99%, odnosno za rizik 5% i 1%. Ulazi u tablice su stepen slobode uzorka sa većom varijansom u zaglavlju i stepen slobode uzorka sa manjom varijansom u pretkoloni.



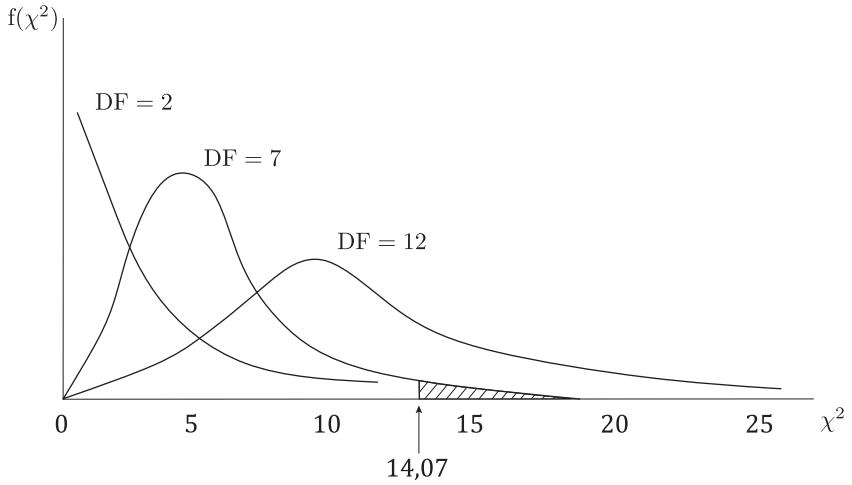
Slika 30. F-raspored

U našem primjeru, gdje su dati stepeni slobode 7 i 6, vrijednost slučajne promjenljive *F* u tablicama iznosi 4,21. Površina ispod krive lijevo od ove vrijednosti iznosi 95%, a desno 5%. Drugim riječima, samo u 5% slučajeva možemo očekivati količnik varijansi veći od 4,21. Takve vrijednosti smatramo statistički značajnim.

Napomena: Treba zapaziti da je najmanja vrijednost u tablicama jedan. Naime, prilikom izrade tablica u brojniku je uvijek uzimana veća varijansa, iz praktičnih razloga. Tablice F-rasporeda su izrađene samo za neke slučajeve. Danas su sve statističke tablice sastavni dio softvera.

1.3.4. Hi-kvadrat (χ^2) raspored

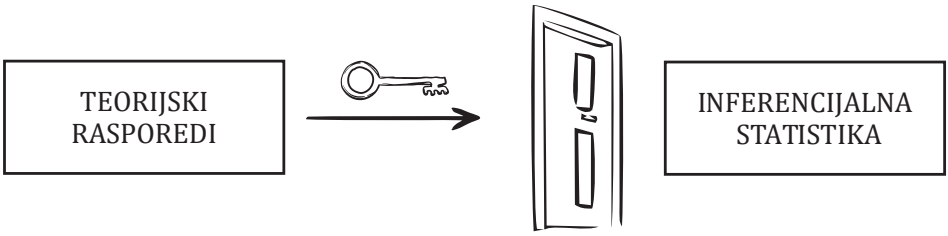
Slučajna promjenljiva χ^2 ima desno asimetričan raspored, koji se također približava normalnom obliku sa povećanjem broja stepena slobode *(Slika 31)*. Ovdje se stepen slobode određuje prema broju klasa, a ne prema broju elemenata uzorka. Najčešće se koristi za testiranje razlike rasporeda stvarnih i teorijskih frekvencija.



Slika 31. Hi-kvadrat raspored

Tablice ovog rasporeda nalaze se u Knjizi (3, str. 371). Pored vjerovatnoće, ulaz u tablice je stepen slobode, u pretkoloni. Ako stepen slobode iznosi 7, za uobičajenu vjerovatnoću od 95% (u zaglavlju piše „vjerovatnoće za veće vrijednosti", tj. za 0,05) vrijednost hi-kvadrata je 14,07. Značenje je slično kao i kod *F*-rasporeda. Površina ispod krive lijevo od te vrijednosti iznosi 95%. Veće vrijednosti od 14,07 se očekuju u manje od 5% slučajeva, tj. vjerovatnoća im je manja od 5%.. To znači da su tako velike razlike vrlo rijetke i ne mogu se smatrati slučajnim, već statistički značajnim.

Svaki od neprekidnih teorijskih rasporeda ima svoju funkciju (krivu) vjerovatnoće. Te funkcije su složene i nismo ih ovdje navodili (osim kod normalnog rasporeda). Površina ispod krive svakog rasporeda odgovara vjerovatnoći. Ukupna površina, odnosno ukupna vjerovatnoća jednaka je jedan. Normalni i t-raspored su dvostrani i potpuno simetrični, dok su F i χ^2 desno asimetrični. Standardizovane slučajne promjenljive Z, t, F i χ^2 su neimenovani brojevi (standardizovane, tj. relativne vrijednosti). One su na x osi, a računamo ih po odgovarajućim formulama i očitavamo iz tablica. O ovim rasporedima više u narednim poglavljima, gdje ćemo govoriti o njihovoj primjeni.



Teorijski rasporedi su "ključ" za vrata inferencijalne statistike

Prof. dr Ostoja Stojanović
(1926-2016)

Zaslužan za uvođenje statistike u nastavu i šumarsku praksu BiH

2. STATISTIČKI UZORCI

U poglavlju II govorili smo o statističkim skupovima. Ako je populacija mala mjere se svi elementi, a rasporedom frekvencija i statističkim parametrima opisujemo pojavu. Tada smo napomenuli da postoji mogućnost procjene karakteristika skupa na osnovu određenog broja jedinica. Takav pristup u statistici nazivamo *Metod uzorka* ili *Reprezentativni metod*.

2.1. Osnovno o uzorcima

Načelno, statističke skupove proučavamo:

1. Potpunim obuhvatom (potpuni premjer, popis)
2. Djelimičnim obuhvatom (reprezentativni uzorak, anketa)
3. Procjenom (pomoću tablica, okularno)

Potpuni obuhvat podrazumijeva ispitivanje (mjerenje) svih jedinica skupa. To je jedini način na koji možemo saznati prave (stvarne, tačne) karakteristike skupa. *Djelimičnim obuhvatom* posmatraju (mjere) se samo određene jedinice skupa, tj. uzima se uzorak. Na osnovu njega procjenjuju se parametri skupa. *Grube procjene* parametara skupa mogu se vršiti pomoću odgovarajućih tablica (npr. tablice debljinskog prirasta stabala), a u izuzetnim slučajevima okularno (odoka).

Napomena: Proučavanje skupova pod 2 i 3 podrazumijeva procjene. Procjene pomoću uzorka (pod 2) vrše se na osnovu mjerenja određenih jedinica skupa, dok se grubim procjenama (pod 3) ne vrše nikakva mjerenja. U šumarstvu je uobičajeno da kažemo *procjene* na osnovu uzorka, dok se u drugoj literaturi koristi termin *ocjene*. Uzorci imaju naučni karakter. Pomoću njih takođe *testiramo* razne pretpostavke u vezi parametara skupa (statistički testovi), te ispitujemo *zavisnosti* između pojava (regresiona analiza).

2.1.1. Pojam uzorka

Uzorak ima dva značenja. Prvo, u *tehničkom* smislu uzorak predstavlja materijalni dio nečega. Na primjer, uzimamo uzorak zemlje da bismo odredili tip zemljišta ili komad stijene da odredimo geološki supstrat. U *statističkom* smislu uzorak predstavlja broj jedinica (elementarnih dijelova) uzetih iz neke masovne pojave, koja je definisana kao statistički skup. Naravno, te jedinice se ne uzimaju „bilo kako". Za to postoje statistička pravila. Statistički uzorak mora da zadovolji dva uslova:

- da je reprezentativan i
- da se može utvrditi greška uzorka.

Reprezentativnost uzorka

Elementi skupa koji čine uzorak treba da predstavljaju (reprezentuju) strukturu skupa, po ispitivanom obilježju (uzorak važi samo za određeno obilježje). Na primjer, ako su u nekoj šumi zastupljena stabla jele, smrče i bukve, očekujemo u uzorku sličan omjer ovih vrsta drveća ili ako izračunavamo srednji prečnik u nekoj šumi u uzorku se trebaju naći stabla svih debljina srazmjerno njihovom učešću u cijeloj šumi. Kako je uzorak dio skupa, njegov rezultat se uvijek razlikuje od skupa. To je greška reprezentativnosti, koju nazivamo jednostavno greškom uzorka. Kada bi uzorak bio 100% reprezentativan tada bi se statističko zaključivanje svelo na uzorak (mjerenje, računanje) i proglašavanje to parametrima skupa. Reprezentativnost uzorka postiže se objektivnim (slučajnim) izborom jedinica.

Greška uzorka

Razlika između stvarne (prave) vrijednosti parametra u skupu i vrijednosti parametra u uzorku naziva se g*reška uzorka*. Da bismo je mogli odrediti potrebno je poznavati parametar skupa. Nekad je to bio jedini način da se ispita valjanost uzorka. Danas se greška uzorka utvrđuje na drugi način, pomoću statističkih formula na bazi vjerovatnoće teorijskih rasporeda. Važnost greške je izuzetno velika, pa joj se posvećuje posebna pažnja.

Prave parametre skupa mi zapravo ne možemo tačno utvrditi, ali možemo odrediti koliko se vrijednost iz uzorka udaljava od prave vrijednosti iz skupa. Drugim riječima, možemo dati interval u kome će se nalaziti parametar skupa. To činimo s određenom vjerovatnoćom, najčešće 95%. Nastojimo uvijek da interval procjene bude što uži, tj. da greška uzorka bude što manja. To će zavisiti od izbora tipa uzorka i njegove veličine. Pri tome ćemo morati uzeti u obzir i druge faktore, kao što su potrebna sredstva i vrijeme.

2.1.2. Opravdanost primjene uzoraka

Uzorak je jedino moguće rješenje u slučaju beskonačnih skupova (npr. ponavljanje nekog mjerenja ili eksperimenta) i skupova sa praktično beskonačnim brojem jedinica (npr. stabala u šumi, neke vrste divljači, šumskih mrava ili štetnih šumskih insekata).

Primjena uzorka je nužna u situacijama kada dolazi do *oštećenja* ili *uništenja* jedinica. Navedimo nekoliko slučajeva iz naše oblasti: a) Kod utvrđivanja starosti neke šume broje se godovi na panjevima posječenih stabala. Potpunim obuhvatom morali bismo posjeći cijelu šumu, b) Rast stabala u toku njihovog života utvrđujemo dendrometrijskom analizom, koja podrazumijeva da posječena stabla trebamo prerezati na svaka dva metra. To bi značilo ne samo uništenje živih stabala, već i uništenje posječene drvne mase, tj. sortimenata, c) Bušenjem stabala pomoću Preslerovog svrdla mjerimo debljinski prirast. Tom prilikom stabla se oštećuju. Ako bismo to radili na svim stablima u šumi šteta bi bila velika.

U većini slučajeva primjena uzorka je racionalna i ekonomična. Kod premjera šuma (taksacije, inventure), koji podrazumijeva utvrđivanje stanja šuma na velikom prostoru, primjena uzorka daje čak bolje rezultate od potpunog premjera (stručnija i sigurnija mjerenja). Kod potpunog premjera morao bi se angažovati veliki broj nekvalifikovane radne snage, a radove bi bilo nemoguće kontrolisati. Osim toga, radovi bi trajali dugo i puno bi koštali.

2.1.3. Koje karakteristike skupa procjenjujemo pomoću uzorka?

Opis masovne pojave pomoću uzorka svodi se uglavnom na dva parametra. To su *aritmetička sredina* i *proporcija*. Vrijednosti ovih parametara u skupu nazivaju se prave vrijednosti, a u uzorku parametri uzorka ili statistika uzorka. Naime, da bi napravio razliku od skupa Fišer je parametre uzorka nazvao „statistika uzorka", što zapravo i jesu statističke a ne prave vrijednosti.

Aritmetička sredina i agregat. Mi očekujemo da aritmetička sredina uzorka bude jednaka aritmetičkoj sredini skupa. Na primjer, da uzorkom utvrđena prosječna zapremina drvne mase po hektaru važi za cijelu šumu (skup). Na osnovu takve procjene utvrđujemo zalihu drvne mase na cijeloj površini šume. Takva procjena naziva se *agregat* (A) ili *total* (T). *Napomena:* u nastavku ćemo vidjeti da se procjene uzorkom ne daju kao jedna vrijednost, već intervalno.

Proporcija i kontingent. Dok aritmetička sredina pokazuje veličinu (intenzitet) pojave, proporcija nas upoznaje sa strukturom mase. Proporcija može biti po atributivnom obilježju, npr. učešće četinara 60% ili numeričkom obilježju, npr. od ukupne drvne mase 25 % otpada na stabla prečnika preko 50 cm. Ako znamo sa koliko drvne mase raspolažemo, tj. koliki je total, onda

na osnovu prethodnih proporcija možemo saznati koliko otpada na četinare, odnosno koliko je učešće drvne mase preko 50 cm. Takav rezultat naziva se *kontingent*.

2.1.4. Veliki i mali uzorak

U statistici postoji formalna podjela uzoraka na velike i male. Kod proporcije se postavlja i poseban uslov. Osim toga, potrebno je voditi računa i o relativnoj veličini uzoraka.

Veliki uzorci. Prikupljanje širokih informacija o nekoj velikoj masovnoj pojavi zahtijeva uzimanje jako velikog broja jedinica u uzorak, ponekad i više hiljada. Ako je apsolutna veličina uzorka preko 30 jedinica uzorci se smatraju velikim ($n > 30$). Tada se zaključivanje izvodi na bazi zakona vjerovatnoće normalnog rasporeda. Uzorci sa manje od 30 jedinica smatraju se malim uzorcima.

Mali uzorci. Mali uzorci se obično koriste kod eksperimentalnih istraživanja. Tako, pri utvrđivanju uticaja nekog faktora, npr. đubrenja na prinos poljoprivredne ili šumske kulture dovoljno je uzeti svega nekoliko parcela. Procjene koje se u statistici izvode na bazi malih uzoraka ($n < 30$) ne zasnivaju se na normalnom već na t-rasporedu. Ovaj raspored je simetričan i spljošten (razvučen) u odnosu na normalni raspored. Sa povećanjem uzorka on se približava normalnom rasporedu. Već kod $n = 30$ gubi se razlika između ova dva rasporeda.

Podjela uzoraka po veličini:

Veliki uzorci $n > 30$, mali uzorci $n < 30$

Uzorci kod proporcije. Za procjenu proporcije „traže se" veliki uzorci po kriterijumu da su proizvodi *np* i *nq* veći od pet. Kad je $p = q$ dovoljna veličina uzorka je $n = 10$. Što je veća razlika između p i q potreban je veći uzorak. Na primjer, za $p = 0,9$ i $q = 0,1$ potreban uzorak je $n > 50$. Primjenjuje se normalni raspored.

Uslov za velike uzorke kod proporcije:

$$np > 5 \quad i \quad nq > 5$$

Relativno veliki uzorci. Kod konačnih skupova, ako se uzimaju relativno veliki uzorci, u statističkim formulama potrebno je koristiti tzv. faktor konačnosti (FK). Relativno velikim smatraju se uzorci čija je stopa izbora veća od 5%.

Relativno veliki uzorci

Ako je stopa izbora veća od 5% (n/N > 0,05) koristi se faktor $FK = \sqrt{\frac{N - n}{N - 1}}$

Faktor konačnosti (popravni faktor) ima za cilj da smanji grešku. Približavanjem veličine uzorka skupu ovaj faktor teži nuli. U slučaju izjednačenja uzorka i skupa greška potpuno nestaje. Naravno, tada više nema ni uzorka.

2.2. Izbor elemenata uzorka

Različite situacije i ciljevi nameću u praksi primjenu različitih načina izbora elemenata uzorka. Osnovni princip na kome je zasnovana teorija statističkih uzoraka je princip slučajnosti. Međutim, postoje i subjektivni načini. Oni se obično koriste u naučnim istraživanjima. Dakle, izbor elemenata uzorka u principu može se podijeliti u dvije grupe. To su:

- objektivni (slučajni) način izbora i
- subjektivni (namjerni) način izbora.

2.2.1. Objektivni način izbora

Uzorci svojom strukturom elemenata treba da predstavljaju strukturu skupa po ispitivanom obilježju. Zapravo uzorci treba da budu umanjena „slika“ skupa. Sam izbor elemenata zasniva se na vjerovatnoći izbora. U objektivne izbore spadaju:

- slučajni izbor i
- sistematski izbor.

2.2.1.1. Slučajni izbor

Slučajni izbor podrazumijeva da svaka jedinica u skupu ima istu vjerovatnoću (šansu) da bude izabrana u uzorak. Ipak, ovako definisana slučajnost izbora važi samo za jednostavne uzorke. U širem smislu, slučajni izbor znači da svaka jedinica ima unaprijed poznatu vjerovatnoću da bude izabrana u uzorak. Slučajni izbor se obavlja na dva načina:

- lutrijski izbor i
- Izbor pomoću Tablice slučajnih brojeva.

Lutrijski izbor

Lutrijski izbor u početku je obavljan izvlačenjem *iz* šeširа (posude). Kasnije je šešir zamijenjen *bubnjem*, a papirići lopticama. Kod ovog izbora sve jedinice

skupa obilježavaju se rednim brojevima na terenu i u kancelariji. Izvlačenjem dobijamo redne brojeve elemenata koje kasnije pronalazimo na terenu i premjeravamo.

Primjena slučajnog izbora kod velikih skupova rijetko dolazi u obzir. Pogodan je jedino ako raspolažemo odgovarajućom bazom podataka. Kod nas u šumarstvu obilježavanje i pronalaženje stabala u šumi praktično je neizvodljiv posao. Sličan problem je i kad koristimo Tablicu slučajnih brojeva.

Izbor pomoću Tablice slučajnih brojeva

Provođenje slučajnog izbora elemenata znatno je olakšano izumom Tablice slučajnih brojeva (1930. godine). *Tablica slučajnih brojeva* je sastavljena od brojeva koji su poredani potpuno slučajno, tako da se u velikom nizu postiže jednaka vjerovatnoća svakog od njih. Slučajnost brojeva u tablicama jednako se ispoljava u svim smjerovima: horizontalno (lijevo ili desno), vertikalno (odozgo ili odozdo), dijagonalno, ...). Danas se ove tablice koriste u elektronskoj formi, tako što brojeve generiše (proizvodi) računar.

Vjerovatnoća u Tablici slučajnih brojeva

Jednocifreni brojevi (0-9)	Dvocifreni brojevi (00-99)	Trocifreni brojevi (000-999)
p = 1/10	p = 1/100	p = 1/1000

Sami možemo sastaviti Tablicu slučajnih brojeva, tako što izvlačimo jednocifrene brojeve „iz šešira“ i pišemo ih jedan pored drugoga. Brojeve svaki put vraćamo u šešir, jer je u pitanju izbor sa ponavljanjem.

Primjer upotrebe tablice. Ovdje smo uzeli jedan mali dio Tablice iz Knjige (3, str. 367), kako bismo prikazali njenu upotrebu *(Slika 32)*. Pretpostavimo da skup ima 80 jedinica (N = 80) i da pomoću Tablice slučajnih brojeva treba izabrati uzorak od osam elemenata (n = 8).

20414	73116	53396	45514	41258	59982	90259	55178	08458	71774
00086	50389	28113	93592	82332	09778	50310	78893	84070	30868
57604	62449	59431	66212	57537	33199	84221	64024	14336	69938

Slika 32. Tablica slučajnih brojeva

Postupak je sljedeći. Ukupan broj elemenata (*N*) određuje koliko cifara gledamo, a broj elemenata uzorka (*n*) koliko brojeva (elemenata) uzimamo. Ovdje pratimo dvocifrene brojeve, gdje je najveći broj 80. Biramo osam brojeva. Prvo treba odrediti polazno mjesto na neki nepristrasan način. Na primjer, uzimaćemo zadnja dva broja, u grupi od pet brojeva i ići horizontalno udesno, od početka tablice iz gornjeg lijevog ugla. Na taj način čitamo sljedeće brojeve: prvi broj je 14, a drugi 16. Broj 96 preskačemo, jer je veći od 80. Naredni broj 14 se ponavlja pa i njega preskačemo. Slijede brojevi 58, 59, 78, 74, 13 i 32. Izvučeni brojevi su boldirani.

Napomena: Nekad, zbog „nedostatka" brojeva može se oduzimati ukupan broj (N) od očitanog broja (npr. 16 bi uslovno bio naš treći broj, 96 - 80 = 16). O primjeni ovog postupka odlučuje se unaprijed.

Varijante slučajnog izbora

Kod slučajnog izbora elemenata u uzorak postoje dvije varijante:

- Izbor bez ponavljanja i
- Izbor sa ponavljanjem.

Izbor bez ponavljanja. U praksi se uglavnom koristi izbor bez ponavljanja, jer nema smisla ponovo uzimati iste elemente. To znači da preskačemo ponovljene brojeve u tablici, odnosno izvučene ceduljice ne vraćamo u šešir.

Vjerovatnoća izbora elemenata bez ponavljanja nakon svakog izvlačenja postaje veća. Početna vjerovatnoća u našem primjeru je bila 1/80, nakog prvog izvučenog broja ona je 1/79, zatim 1/78 itd. Kad su u pitanju veliki uzorci, kakvi su najčešće u šumarstvu, promjena vjerovatnoće nema praktični značaj.

Izbor sa ponavljanjem. Izbor sa ponavljanjem znači da svaki izvučeni broj vraćamo u šešir, odnosno ako se pojavi u tablici uzimamo ga ponovo. Imali smo jedan takav primjer kod sastavljanja tablice slučajnih brojeva. Izbor sa ponavljanjem ima više teorijski značaj.

2.2.1.2. Sistematski izbor

Objektivnost izbora elemenata sistematskim načinom obezbjeđuje se po nekom sistemu, šemi ili redu, bilo da se uzima svaki n-ti element ili svaki element na nekom određenom rastojanju.

Izbor po sistemu „svaki n-ti element"

Ova varijanta sistematskog izbora takođe zahtijeva obrojčavanje svih jedinica na terenu. Skup dijelimo na jednake grupe, a onda iz svake grupe uzimamo po jedan element. U prethodnom primjeru, u kome je N = 80, n = 8, dobijamo osam grupa po 10 elemenata. Svaki deseti element ulazi u uzorak. To nazivamo *korak izbora*.

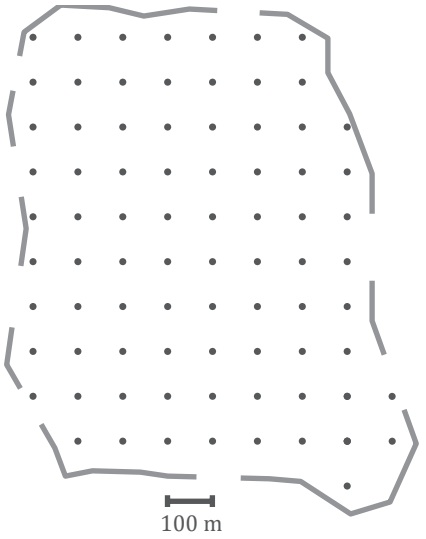
Broj grupa = Korak izbora

$$\text{Broj grupa (korak izbora)} = \frac{N}{n} = \frac{80}{8} = 10$$

Element iz prve grupe izvlači se slučajno. Recimo da je „iz šešira“, u kome se nalaze brojevi od 1 do 10, izvučen broj 3. Svaki sljedeći broj dobijamo „koračanjem“: 3 + 10 = 13, zatim 13 + 10 = 23 itd. Ovakav sistem osigurava objektivnost, a statistika se dalje izvodi kao da se radi o slučajnom izboru.

Izbor po sistemu “na određenom rastojanju“

Ovaj sistem ne zahtijeva obrojčavanje elemenata skupa, što omogućuje njegovu primjenu pri redovnim inventurama šuma. Kako taj sistem izgleda na primjeru jednog odjela, koji je dio jedinstvene šire mreže krugova prikazuje *Slika 33*. *Krug* je uobičajeni naziv za elementarnu površinu (jedinicu skupa). Krugovi se nalaze na rastojanju svakih 100 metara i tako čine kvadratnu mrežu. Oni se prenose sa karte na teren.



Slika 33. Sistematski izbor elemenata uzorka

2.2.2. Subjektivni način izbora

Subjektivni izbor je način izbora elemenata uzorka koji nije zasnovan na vjerovatnoći. Njegova subjektivnost potiče od pristrasnosti svakog lica. Postoji više načina subjektivnog izbora, kao što su stručni izbor, metod kvota, izbor nasumice ili po nahođenju. Kod subjektivnog izbora ne postoji mogućnost određivanja greške uzorka.

Ipak, u šumarstvu, za potrebe naučnih istraživanja stručni izbor ima veliki značaj. Upravo je ovaj metod korišten pri obimnim istraživanjima naših šuma. Ogledne plohe su tada postavljane radi utvrđivanja prirasta i drugih obilježja šuma u različitim uslovima (uticaj različitih faktora). To nije bilo moguće uraditi slučajnim ili sistematskim izborom. Elementi uzorka bile su različite ogledne plohe. One su postavljanje po stručnim kriterijumima, prema cilju i zadacima istraživanja. Objektivnost izbora ploha nije mogla biti značajnije narušena. Naime, mogućnost sistematskog (namjernog) biranja ploha sa većim ili manjim prirastom bila je isključena, pošto se na njima javljaju različite kombinacije faktora. Time je opravdana primjena statistike u daljem radu. Važno je također napomenuti da stručnim uzorkom nije moguće utvrditi stanje, kao ni strukturu šuma.

2.3. Tipovi (vrste) uzoraka

Prirodne i društvene pojave se razlikuju po masovnosti, strukturi i mnogim karakteristikama. Njihovo ispitivanje pomoću uzorka zahtijeva različite pristupe. Tako su nastale različite vrste (tipovi, planovi) uzoraka. Možemo ih svrstati u dvije grupe:

- jednostavni uzorci i
- složeni (kombinovani) uzorci.

2.3.1. Jednostavni uzorci

Jednostavni uzorci nemaju posebna ograničenja. To podrazumijeva jednak pristup svim jedinicama skupa, odnosno svim dijelovima skupa iz koga se jedinice uzimaju. Govorili smo o slučajnom i sistematskom načinu izbora elemenata u uzorak. Ako se ostane kod izbora jedinica u uzorak na jedan od ova dva načina imamo:

- jednostavni slučajni uzorak i
- jednostavni sistematski uzorak.

2.3.1.1. Jednostavni slučajni uzorak

Jednostavni slučajni uzorak nastaje ako se jedinice biraju na slučajan način. To znači da sve jedinice skupa imaju istu (jednaku) vjerovatnoću da će biti izabrane u uzorak. Iz praktičnih (tehničkih) razloga primjena ovog uzorka ograničava se na male skupove. Drugi uslov za primjenu jednostavnog slučajnog uzorka je jednoličnost skupa. U šumarstvu takve skupove predstavljaju rasadnici ili šumske kulture, odnosno plantaže.

2.3.1.2. Jednostavni sistematski uzorak

Jednostavni sistematski uzorak nastaje sistematskim načinom izbora jedinica. U šumarstvu se primjenjuje sistem kvadratne mreže. Ovaj način ne zahtijeva obrojčavanje na terenu, pa se može koristiti i za veće skupove (veće površine šuma). Ipak, kad se radi o velikim skupovima primjenjuje se neki drugi tip uzorka, npr. stratifikovani (ako se radi o nejednoličnim skupovima) ili etapni (ako se radi o jednoličnim skupovima).

2.3.2. Složeni (kombinovani) uzorci

Nekad je korisno da se izbor jedinica umjesto jednostavnog slučajnog vrši planski, na neki način kontrolišući vjerovatnoću izbora. Kod svih ovih uzoraka zadržan je princip slučajnosti. Složeni (kombinovani, kontrolisani, ograničeni) uzorci su:

- stratifikovani uzorak
- etapni uzorak
- fazni uzorak
- uzorak grupa i
- uzorak nejednake vjerovatnoće izbora.

2.3.2.1. Stratifikovani uzorak

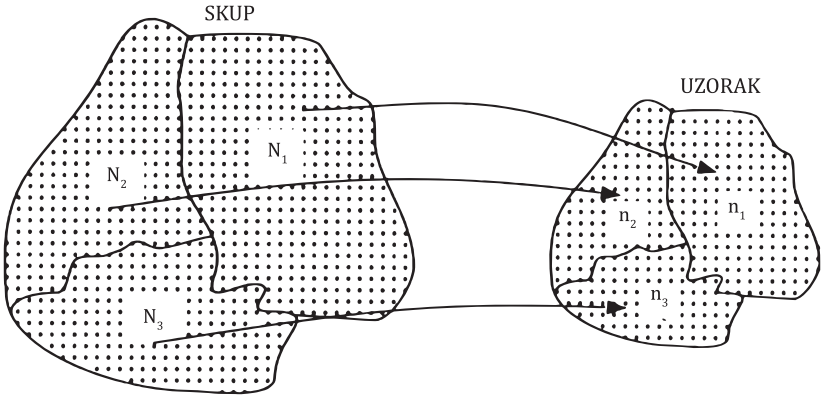
Stratifikovani uzorak se primjenjuje na velike skupove u kojima se mogu izdvojiti dijelovi koji se između sebe razlikuju po strukturi (homogenosti, ujednačenosti). Upravo zbog takvih razlika greška procjene bila bi velika ako bi se primijenio jednostavni uzorak (povećanje standardne devijacije u skupu dovodi do povećanja standardne greške uzorka). Ti dijelovi skupa nazivaju se *stratumi*, a postupak njihovog izdvajanja naziva se *stratifikacija*.

Stratifikacija ima dva cilja. Prvi smo već spomenuli. To je da se dobije bolja (preciznija) procjena skupa u cjelini, tj. manja greška. Drugi cilj je da se dobiju podaci za stratume. Na primjer, jedno šumskoprivredno područje, tj. veliki šumski kompleks, statistički gledano predstavlja veliki skup. Logično je da se on sastoji od različitih šuma. Te šume, npr. kategorije šuma (KŠ) mogu se uzeti kao stratumi. Tako jedan stratum mogu biti mješovite šume, drugi bukove, a treći hrastove šume. Stratifikacija može da se radi na više nivoa. U našem slučaju mogli bismo unutar već postojećeg stratuma, recimo bukovih šuma, posmatrati pojedine tipove bukovih šuma. Njih u stručnoj praksi nazivamo gazdinske klase. One bi predstavljale nove stratume. Osim za šumu u cjelini (ŠPP – šumskoprivredno područje) nama trebaju podaci i za pojedine vrste šuma (KŠ – kategorije šuma i GK – gazdinske klase).

Zavisno od cilja stratifikacije, izbor elemenata, koji se najčešće vrši sistematskim načinom, može biti:

- proporcionalan i
- neproporcionalan.

Proporcionalan izbor. Kod ovog načina izbora broj elemenata uzorka koji se uzima iz pojedinih stratuma je proporcionalan (srazmjeran) broju elemenata u njima *(Slika 34)*. Na primjer, ako je stopa izbora 3%, onda će u uzorak biti uzeto po 3% elemenata od svakog stratuma. Ovaj izbor ima najširu primjenu.



Slika 34. Stratifikovani uzorak

Neproporcionalan izbor. Nekad možemo imati takvu situaciju da se stratumi međusobno razlikuju po značaju, npr. da je šuma u nekom stratumu vrednija od šume u drugom stratumu. Ako nju želimo bolje procijeniti, uzimamo veći uzorak u tom stratumu.

Kad pored veličine stratuma uzmemo u obzir i varijabilitet dobićemo poseban vid neproporcionalnog izbora koji se naziva *optimalan izbor*. Pogledajmo jedan primjer. Neka jedan stratum ima 1000 jedinica, a drugi 100 jedinica. Kod proporcionalnog izbora odnos broja elemenata uzorka bio bi 10:1. Međutim, ako je standardna devijacija (u nekim jedinicama mjere) u prvom stratumu iznosila 5, a u drugom 20, tada se odnos bitno mijenja. Kad se pomnože veličina stratuma i standardna devijacija dobija se odnos 5:2. Formule i primjer za stratifikovani uzorak možete pronaći u Knjizi (3, str. 221).

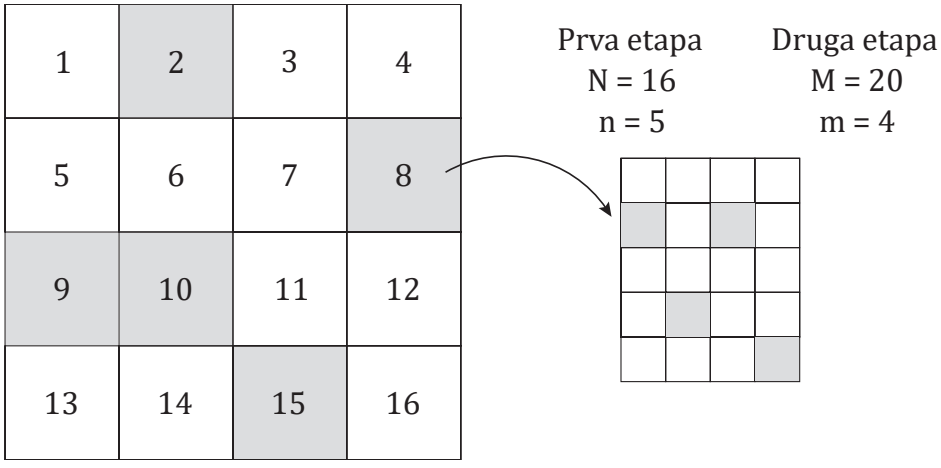
2.3.2.2. Etapni uzorak

Etapni uzorak se primjenjuje za velike jednolične skupove. Takav primjer u šumarstvu su „nepregledne" površine ravničarskih šuma. Primjena jednostavnog uzorka u ovakvim situacijama nije opravdana, jer bismo očigledno trošili puno

vremena i sredstava za premjer jednoličnе mase. Pokazalo se da je efikasniji etapni uzorak, čija preciznost ne zaostaje značajnije za slučajnim uzorkom. Etapni uzorak može biti dvoetapni ili višeetapni.

Kod dvoetapnog uzorka biramo dvije vrste jedinica. U prvoj etapi biraju se tzv. *primarne jedinice*, a u drugoj *sekundarne jedinice*. Kod ovog uzorka svaka primarna jedinica može podjednako dobro reprezentovati skup (one su međusobno slične „kao jaje jajetu"). Što su razlike između primarnih jedinica manje dvoetapni uzorak je efikasniji. Kod stratifikovanog uzorka bilo je suprotno (tamo je svaki stratum morao biti zastupljen u uzorku).

Pogledajmo jedan hipotetički primjer dvoetapnog uzorka *(Slika 35)*. Neka se radi o ravničarskoj šumi hrasta, koja je već prostorno podijeljena na 16 odjela. Odjeli ovdje mogu predstavljati primarne jedinice. Pretpostavimo da su one jednake i da svaka ima površinu 20 hektara. Uzmimo dalje da svaki hektar predstavlja jednu sekundarnu jedinicu. Neka smo slučajnim izborom izvukli u prvoj etapi pet primarnih jedinica, a iz svake od njih, u drugoj etapi, po četiri sekundarne jedinice. *Napomena*: mjerenja vršimo samo na izabranim sekundarnim jedinicama. Na slici smo označili sve izvučene jedinice. U Knjizi (3, str. 227) možete vidjeti kako se pomoću dvoetapnog uzorka procjenjuju aritmetička sredina i proporcija.



Slika 35. Dvoetapni uzorak

Plan uzorka može se postaviti tako da u etapama dobijamo podatke o različitim obilježjima, što predstavlja novi kvalitet informacija. Takav uzorak dobija naziv *stepeni uzorak* (dvostepeni ili višestepeni). Na našem prethodnom primjeru mogli smo na primarnim jedinicama premjeravati prečnike stabala, na sekundarnim jedinicama osim prečnika i visine, a debljinski prirast mjeriti na još manjim površinama, unutar sekundarnih jedinica. To bi bio trostepeni uzorak.

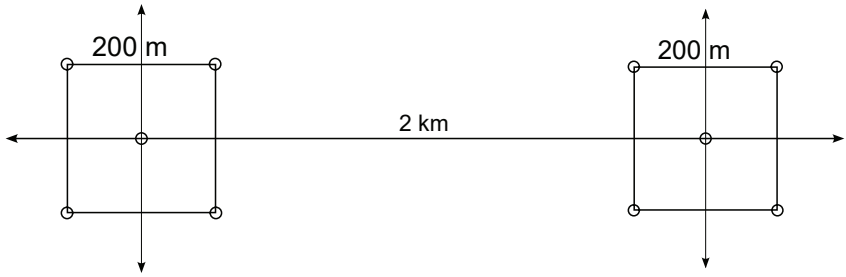
2.3.2.3. Fazni uzorak

Ima li razlike između etapa i faza? *Etape* se provode u isto vrijeme na različitim jedinicama, dok se *faze* izvode na istim jedinicama vremenski odvojeno. Fazni uzorak se također primjenjuje za velike skupove, s ciljem da se smanje troškovi. Na primjer, kod inventure šuma na velikim površinama snime se tačke na terenu nekom daljinskom metodom (npr. aviosnimcima). Kasnije, u drugoj fazi, izvrši se korekcija rezultata na terenu na svim tačkama.

Također, utvrđivanje zapremine drvne mase nekog područja može se izvesti u dvije faze. U prvoj fazi mjerimo sve prečnike i visine stabala. Nakon toga, na manjem broju stabala nekom dendrometrijskom metodom tačno utvrdimo njihovu zapreminu. Rezultat toga mogu biti zapreminske tablice, koje u drugoj fazi iskoristimo za određivanje zapremine svih stabala. Ovaj uzorak može biti dvofazni ili višefazni.

2.3.2.4. Uzorak grupa

Uzorak grupa karakteriše vrlo racionalan pristup. Kod njega je tipično to da se mjerenja (ispitivanja) koncentrišu na određena mjesta, koja se nazivaju *grupe*. U šumarstvu takve grupe koriste se u državnim inventurama. Tako je bilo u I i II državnoj inventuri šuma u BiH. Grupe u prvoj inventuri nazivane su *traktovi*, a u drugoj *klasteri*. Princip racionalnosti ostvaruje se tako da se tačke (mjesta) na kojima se formiraju grupe nalaze na prilično velikom rastojanju *(Slika 36)*. Grupe najčešće čine po četiri kruga, što je ujedno dnevna norma. Tako se znatno smanjuju troškovi.

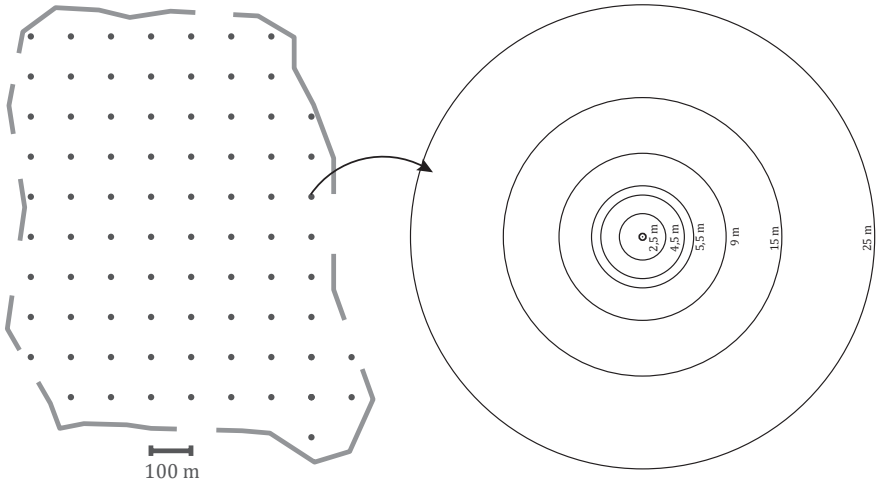


Slika 36. Uzorak grupa

2.3.2.5. Uzorak nejednake vjerovatnoće izbora

Ovdje govorimo o jednom specifičnom uzorku, koji se više od 50 godina koristi u inventuri šuma u BiH. On se naziva *metod koncentričnih krugova*. Kako sam naziv kaže radi se o više krugova različitog radijusa sa zajedničkim centrom *(Slika 37)*. Primjenjuje se za izbor stabala na elementarnoj površini (krugu). Krugovi su ovdje složene elementarne jedinice.

Kao primjer, u našoj praksi, na krugu radijusa 2,5 m u uzorak se uzimaju stabla prečnika 5-10 cm, radijusa 4,5 m stabla prečnika 10-20 cm, itd do najvećeg radijusa od 25 m na kome se mjere sva stabla preko 80 cm debljine. Metod koncentričnih krugova detaljno se izučava u dendrometriji. Metod su opisali profesori Ostoja Stojanović i Petar Drinić (rad u literaturi pod brojem 14).



Slika 37. Uzorak nejednake vjerovatnoće izbora

Osim variranja obilježja (agregatnih taksacionih elemenata) šume po površini postoji i varijabilitet između prirodnih jedinica, tj. stabala. On se najčešće posmatra po debljini stabala. Varijabilitet debljih stabala (viših debljinskih klasa) po zapremini i prirastu veći je nego kod tankih stabala (nižih debljinskih klasa). Zbog toga, ali i zbog veće vrijednosti drvne mase, debljim stablima dajemo veću šansu (vjerovatnoću) da uđu u uzorak. U prirodnim (prebornim) šumama najviše ima tankih stabala, a najmanje debelih. Kod iste veličine uzorka (istog radijusa) u uzorku bi bila previše zastupljena tanka stabla. Tada bi greška procjene kod debelih stabala bila velika.

2.4. Teorijska osnova uzoraka

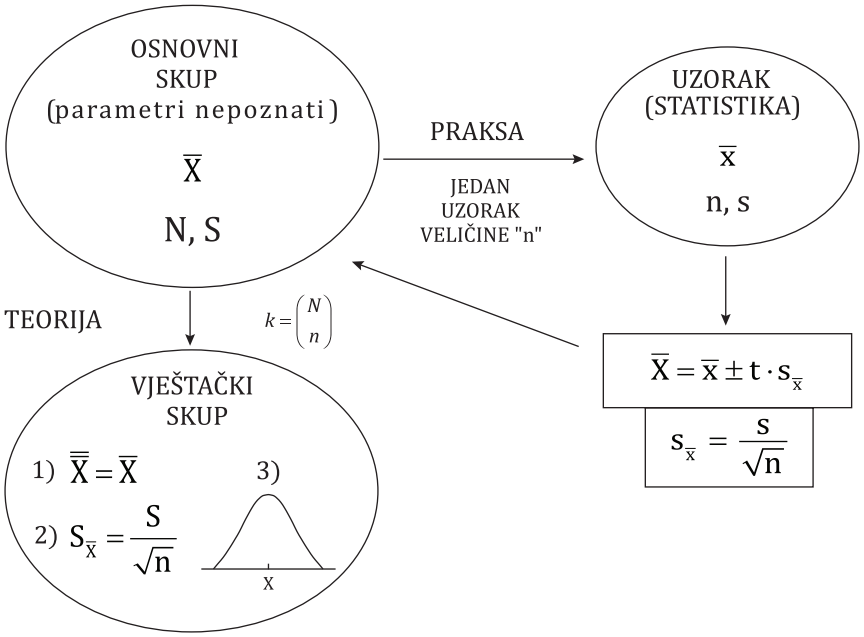
Na osnovu aritmetičke sredine uzorka ($\overline{x}$) procjenjujemo aritmetičku sredinu skupa ($\overline{X}$). Slučajni uzorak čine elementi koji se biraju slučajno, pa je i statistika uzorka slučajna promjenljiva. Postavlja se ključno pitanje kako uspostaviti vezu između sredine uzorka, kao slučajne promjenljive i sredine skupa, kao jedne konstante. Teorija uzoraka je jedno od najtežih pitanja u statistici. Zasniva se na teoriji vjerovatnoće.

Teorija vjerovatnoće

Procjena greške uzorka moguća je zahvaljujući teoriji vjerovatnoće, koja je zasnovana na slučajnom kombinovanju elemenata.
Zato se traži izbor na principu slučajnosti.

2.4.1. Skup sredina uzoraka

Iz nekog skupa možemo uzeti veliki broj uzoraka veličine (n). Svaki od parametara tih uzoraka (aritmetička sredina, proporcija i drugi) formirali bi svoj raspored. Poznavanje karakteristika tih rasporeda je od suštinske važnosti u teoriji uzoraka *(Slika 38)*.



Slika 38. Veza sredine uzorka ($\overline{x}$) i sredine skupa ($\overline{X}$)

Teorijski broj uzoraka (k)

Postavlja se pitanje koliko se teorijski može uzeti uzoraka veličine *n* iz jednog skupa. Taj broj zavisi od toga da li se uzorci uzimaju bez ponavljanja ili sa ponavljanjem.

Ako skup ima samo tri elementa (N = 3), a uzorak dva (n = 2) imaćemo *bez ponavljanja* ukupno tri uzorka (k = 3). To su kombinacije brojeva: 1 2, 1 3 i 2 3. Taj broj uzoraka dobija se po formuli:

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

Ako se uzima uzorak *sa ponavljanjem* broj kombinacija je mnogo veći. U našem primjeru ima devet kombinacija (k = 9): 1 2, 1 3, 2 3, 2 1, 3 1, 3 2, 1 1, 2 2 i 3 3. Njihov broj dobija se po formuli:

$$k = N^n = 3^2 = 9$$

Ako se poveća broj elemenata skupa ili uzorka broj *k* će naglo rasti. Na primjer, kod N = 100 i n = 5, broj mogućih uzoraka (bez ponavljanja) biće 75 miliona (k = 75.287.520). Sa ponavljanjem taj broj je mnogo veći. To su, naravno, samo teorijske mogućnosti. Zato naglašavamo da se radi o vještačkim skupovima.

Karakteristike vještačkog skupa

Vještački skup, kao i svaki prirodni skup, ima karakteristike, kao što su:

- aritmetička sredina,
- standardna devijacija i
- oblik rasporeda.

Aritmetička sredina. Aritmetička sredina vještačkog skupa (skupa sredina uzoraka) $\overline{\overline{X}}$ jednaka je sredini osnovnog skupa ($\overline{X}$):

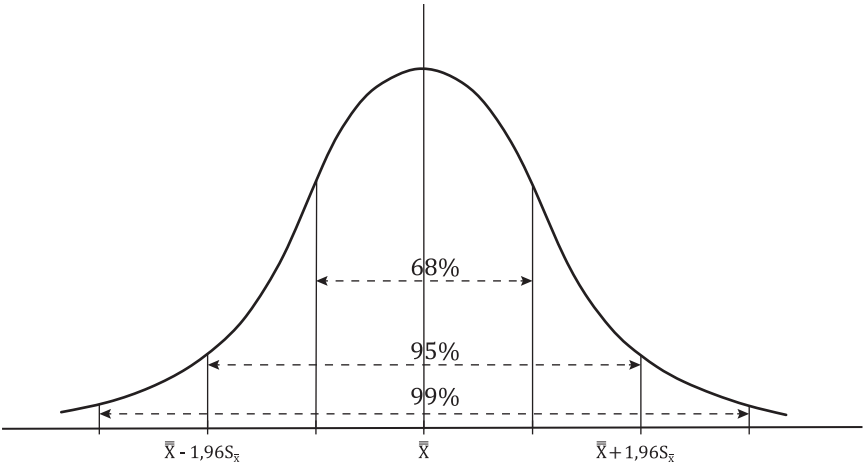
$$\overline{\overline{X}} = \overline{X}$$

Standardna devijacija. Varijabilitet vještačkog skupa je manji od varijabiliteta osnovnog skupa (aritmetičke sredine su manje varijabilne od individualnih podataka). To je logično, jer ekstremne pojedinačne vrijednosti više dolaze do izražaja u osnovnom skupu, nego u uzorcima gdje se kombinuju s ostalim vrijednostima. U statistici je izvedena (dokazana) sljedeća veza između standardne devijacije vještačkog skupa ($S_{\overline{X}}$) i osnovnog skupa (S):

$$S_{\overline{X}} = \frac{S}{\sqrt{n}}$$

Standardna devijacija vještačkog skupa predstavlja prosječno odstupanje sredina uzoraka od prave sredine. Kako ta odstupanja smatramo pogrešnim vrijednostima, standardnu devijaciju nazivamo *standardnom greškom*. Ona zavisi od standardne devijacije u osnovnom skupu (S) i veličine uzorka (n).

Oblik rasporeda. Ako se radi o velikom uzorku (n > 30) raspored sredina biće normalan *(Slika 39)*. Normalni raspored je osnova za procjenu parametara skupa (sredine i drugih), a kasnije ćemo vidjeti i za statistička testiranja hipoteza. Kod ovog rasporeda u intervalu 1,96 standardnih grešaka lijevo i desno od aritmetičke sredine nalazi se 95% uzoraka (njihovih sredina). To znači da u praksi jedan uzeti uzorak ima vjerovatnoću 95% da se nađe u tom intervalu, odnosno 5% da bude izvan intervala.



Slika 39. Normalni raspored sredina uzoraka

Veza osnovnog skupa i skupa sredina uzoraka

Uzmimo jedan primjer (1, str. 62). Skup ima svega šest elemenata (N = 6), a uzorak dva (n = 2). Jedinice mjere su zanemarene. Prvo ćemo izračunati aritmetičku sredinu i standardnu devijaciju osnovnog skupa *(Tabela 4)*. To su prave (stvarne) vrijednosti skupa. Njih u praksi ne računamo.

Tabela 4. Osnovni skup

X	$X - \bar{X}$	$(X - \bar{X})^2$
16	-2	4
20	2	4
17	-1	1
19	1	1
13	-5	25
23	5	25
108	0	60

$$\bar{X} = \frac{\Sigma X}{N} = \frac{108}{6} = 18,00$$

$$S^2 = \frac{\Sigma\left(X - \bar{X}\right)^2}{N} = \frac{60}{6} = 10,00$$

$$S = \sqrt{S^2} = \sqrt{10} = 3,16$$

Ako izvlačimo uzorke veličine n = 2, iz našeg skupa može se izvući ukupno 15 slučajnih uzoraka *(Tabela 5)*. To nam daje formula:

$$\binom{N}{n} = \frac{N!}{n!\left(N-n\right)!} = \frac{6!}{2!\left(6-2\right)!} = \frac{6\cdot5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot\left(4\cdot3\cdot2\cdot1\right)} = 15$$

Radi poređenja sa osnovnim skupom, izračunaćemo aritmetičku sredinu i standardnu devijaciju vještačkog skupa. Elementi ovog skupa su uzorci, koji nastaju kombinovanjem vrijednosti po dva elementa (kolona 2). Njihove aritmetičke sredine su obilježje (kolona 3).

Tabela 5. Vještački skup

Redni broj uzorka	Kombinacije vrijednosti		Sredine uzoraka ($\bar{x}_i$)	$\overline{x_i} - \overline{\overline{X}}$	$\left(\overline{x_i} - \overline{\overline{X}}\right)^2$
1	2		3	4	5
1	16	20	18,0	0,0	0,00
2	16	17	16,5	-1,5	2,25
3	16	19	17,5	-0,5	0,25
4	16	13	14,5	-3,5	12,25
5	16	23	19,5	1,5	2,25
6	20	17	18,5	0,5	0,25
7	20	19	19,5	1,5	2,25
8	20	13	16,5	-1,5	2,25
9	20	23	21,5	3,5	12,25
10	17	19	18,0	0,0	0,00
11	17	13	15,0	-3,0	9,00
12	17	23	20,0	2,0	4,00
13	19	13	16,0	-2,0	4,00
14	19	23	21,0	3,0	9,00
15	13	23	18,0	0,0	0,00
Ukupno			270,0	0,0	60,00

$$\overline{\overline{X}} = \frac{\sum \overline{X}_i}{N} = \frac{270,0}{15} = 18,00$$

$$S^2_{\overline{X}} = \frac{\sum\left(\overline{X}_i - \overline{\overline{X}}\right)}{N} = \frac{60}{15} = 4,00$$

$$S_{\overline{X}} = \sqrt{S^2_{\overline{X}}} = \sqrt{4} = 2,00$$

Ako uporedimo izračunate sredine i standardne devijacije uvjerićemo se u tačnost navedenih osobina vještačkog skupa. Prvo, da je aritmetička sredina vještačkog skupa jednaka sredini osnovnog skupa, što potvrđuje reprezentativnost slučajnog uzorka. I drugo, standardna devijacija vještačkog skupa manja je od standardne devijacije osnovnog skupa (2,00 < 3,13). Nju smo ovdje izračunali iz originalnih podataka, što u praksi nije moguće. Umjesto toga primjenjuje se izvedena formula. Za računanje standardne greške ovdje ćemo koristiti faktor korekcije (FK), jer je stopa izbora veća od 5%. *Napomena*: Dobijen je isti rezultat što potvrđuje ispravnost ove formule.

$$S_{\overline{X}} = \frac{S}{\sqrt{n}} \cdot FK = \frac{3,16}{\sqrt{2}} \cdot 0,89 = 2,00$$

$$FK = \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{6-2}{6-1}} = 0,89$$

Da je uzorak bio n = 3 ukupno bi bilo moguće uzeti 20 uzoraka. Njihova sredina ostala bi ista, a standardna greška bila bi manja i iznosila bi 1,41. Sredine uzoraka sve više se grupišu oko zajedničke sredine što je uzorak veći. To nam govori da se povećanjem uzorka sve više približavamo pravoj vrijednosti.

2.4.2. Skup proporcija uzoraka

Proporcije svih uzoraka činiće vještački skup, koji ima slične osobine kao i skup sredina.

Aritmetička sredina. Proporcija vještačkog skupa ($\overline{\overline{P}}$) jednaka je proporciji osnovnog skupa (P):

$$\overline{\overline{P}} = P$$

Standardna devijacija. Standardna devijacija skupa koga čine proporcije uzoraka je:

$$S_{\overline{P}} = \sqrt{\frac{PQ}{n}}$$

S obzirom da su proporcije iz uzoraka pogrešne ova standardna devijacija naziva se *standardna greška proporcije.* Ona pokazuje prosječno odstupanje proporcija uzoraka od proporcije osnovnog skupa.

Oblik rasporeda. Raspored proporcija uzoraka ima oblik binomskog rasporeda ako se uzimaju uzorci s ponavljanjem, a hipergeometrijskog ako su uzorci bez ponavljanja. Međutim, kod velikih uzoraka oba rasporeda teže normalnom rasporedu. I ovdje se potvrđuje važnost normalnog rasporeda, kroz njegovu široku primjenu u praksi.

2.5. Procjena parametara skupa

U ovom poglavlju govorimo o statističkim procjenama parametara skupa na bazi jednostavnog (slučajnog ili sistematskog) uzorka. Obično procjenjujemo aritmetičku sredinu i proporciju. Parametre u skupu nazivamo prave vrijednosti, a parametre u uzorku statistika uzorka. Procjene su zasnovane na vjerovatnoći teorijskog modela, uz odgovarajuću pouzdanost. *Napomena*: Iz skupa čije parametre želimo upoznati uzima se samo jedan uzorak.

Šta procjenjujemo?

1. Aritmetičku sredinu (kvantitet) i 2. Proporciju (kvalitet - strukturu)

Izračunata vrijednost parametra iz uzorka, kao što smo vidjeli, samo je jedna od velikog broja mogućih vrijednosti iz vještačkog skupa. Kako se sve te sredine smatraju pogrešnim, tako je i sredina uzetog uzorka pogrešna. Inače, takva procjena sa jednim brojem (tačkom) naziva se *tačkasta procjena*. Taj pristup bi očigledno bio pogrešan. Zato umjesto toga određujemo interval u kome se nalazi parametar skupa. Takva procjena naziva se *intervalna procjena.* U literaturi se *interval procjene* još naziva *interval pouzdanosti* ili *interval povjerenja*.

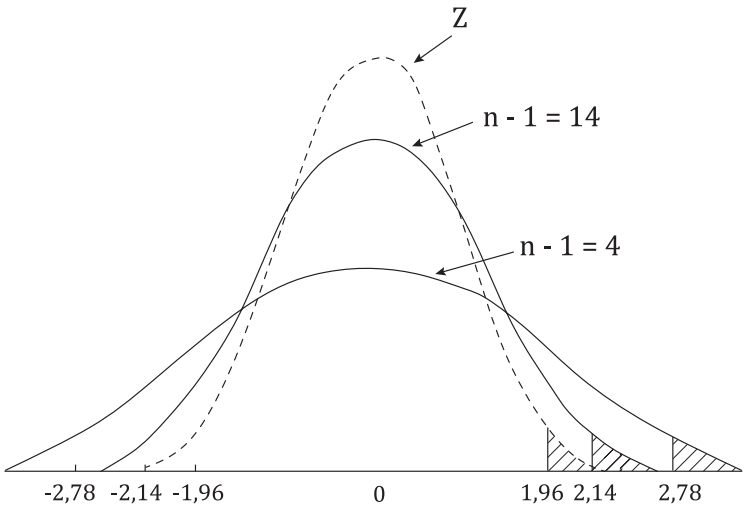
Intervalna procjena

Intervalno procjenjivanje podrazumijeva formiranje intervala u kojem se očekuje vrijednost parametra skupa, s određenom vjerovatnoćom.

2.5.1. Procjena aritmetičke sredine

Raspored sredina uzoraka odgovara normalnom obliku samo pod određenim uslovima. U stvari, raspored sredina uzoraka ima jedan poseban oblik, koji je opisan kao t-raspored. On je sličan normalnom rasporedu, ali se od njega razlikuje kod malih uzoraka ($n < 30$). Što je uzorak veći t-raspored je sve bliži normalnom rasporedu. Kod normalnog (z-rasporeda) 95% površine

ispod krive obuhvaćeno je intervalom 1,96 standardnih grešaka, lijevo i desno od sredine, dok je kod t-rasporeda taj interval širi (za uzorak n = 15 interval je 2,14, a za n = 5 broj standardnih grešaka je 2,78). Procjene na osnovu malih uzoraka su nesigurnije. Zato je interval širi nego kod velikih uzoraka *(Slika 40)*.



Slika 40. Interval procjene u zavisnosti od veličine uzorka (iskazan brojem standardnih grešaka)

Jednačina reprezentativnog metoda

Osnovna jednačina za procjenu sredine skupa na bazi uzorka glasi:

$$\overline{X} = \overline{x} \pm t \cdot s_{\overline{x}}$$

U jednačini su:

$\overline{X}$ – aritmetička sredina skupa

$\overline{x}$ – aritmetička sredina uzorka

t – broj standardnih grešaka (iz tablice t-rasporeda, za stepen slobode n – 1 i rizik α)

$s_{\overline{x}}$ – standardna greška (sredine uzorka)

$t\,s_{\overline{x}}$ – greška uzorka

Stepen slobode. Ulaz u tablicu t-rasporeda je veličina uzorka umanjena za jedan (n – 1), što se naziva *stepen slobode*. Stepen slobode je već korišten kod računanja standardne devijacije uzorka. Sam naziv „stepen slobode" potiče od broja slobodnih vrijednosti u uzorku. Na primjer, kod računanja aritmetičke sredine u uzorku od tri broja samo dva su slobodna dok je treći tačno određen

(ograničen aritmetičkom sredinom). Ako su to brojevi 5 i 10, a aritmetička sredina 10, treći broj nije slobodan (mora biti 15). Stepen slobode ima značaj samo kod malih uzoraka, jer umanjenje za jedan ne utiče na rezultat kad je *n* veliki broj.

Standardna greška. U praksi nije poznata standardna devijacija skupa (S), a ona nam je neophodna za određivanje intervala procjene. Kao njena logična zamjena, u formuli za računanje standardne greške, koristi se standardna devijacija iz uzorka (s):

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Iz formule vidimo da veća standardna devijacija daje veću standardnu grešku, kao i to da je možemo smanjiti povećanjem uzorka. Greška uzorka je zapravo mjera reprezentativnosti uzorka. Ova formula ima posebnu važnost u inferencijalnoj statistici, na samo kod procjena već i kod statističkih testova.

Standardna devijacija uzorka. Kod ocjenjivanja standardne devijacije mase (skupa) očekujemo da nam standardna devijacija uzorka bude njena zamjena. Međutim, ispitivanja odnosa ove dvije standardne devijacije pokazala su da formula za standardnu devijaciju uzorka pristrasno daje niže vrijednosti, tj. manju standardnu devijaciju od standardne devijacije u skupu. Da bi se otklonio taj nedostatak suma kvadrata odstupanja dijeli se sa stepenom slobode *n–1*, umjesto sa *n*. Formula za računanje standardne devijacije uzorka glasi:

$$s = \sqrt{\frac{\sum n_i \left(X_i - X\right)^2}{n-1}}$$

Korekcija standardne greške. Ako se povećava uzorak onda je logično da se njegova greška smanjuje i da na kraju nestane kad se uzorak izjednači sa skupom. Međutim, primjenom formule za računanje greške to se ne ostvaruje. Ovaj nedostatak otklanja se pomoću korekcionog faktora, koji se još naziva *faktor konačnosti* (FK). Kad je stopa izbora manja od 5% ovaj faktor je približno jednak jedan, pa se tada izostavlja. Formula za računanje standardne greške sa korekcionim faktorom glasi:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot FK$$

gdje je:

$$FK = \sqrt{\frac{N-n}{N-1}}$$

Procjena agregata (totala)

Procjena agregata, tj. ukupne (zbirne) vrijednosti svih jedinica skupa ima smisla samo kod obilježja koja se mogu sabirati, drugim riječima čija prosječna vrijednost pomnožena sa ukupnim brojem jedinica skupa daje ukupnu masu.

Kod procjene agregata (A) ili totala (T), kao i kod procjene aritmetičke sredine, imaćemo također interval procjene:

$$A = \left(\overline{x} \pm t \cdot s_{\overline{x}}\right) \cdot N$$

gdje je *N* ukupan broj elemenata. U šumarstvu to je najčešće broj hektara (površina šume), jer je i agregatno obilježje iskazano kao prosječna vrijednost po hektaru. Tako nam agregat daje ukupno raspoloživu drvnu masu nekog šumskog prostora ili ukupno ostvarenu proizvodnju drvne mase u šumskom gazdinstvu.

U prvom dijelu skripte govorili smo o podjeli obilježja (karakteristika) šume. Rekli smo da obilježja kao što su zapremina drvne mase ili prirast drvne mase po hektaru spadaju u tzv. agregatna obilježja, dok npr. srednji prečnik i srednja visina sastojine predstavljaju srednje vrijednosti. Prečnike ili visine svih stabala u šumi bilo bi besmisleno sabirati. Za takva obilježja agregat se ne računa.

2.5.2. Procjena proporcije

Osim procjene prave sredine (i agregata) skupa kao zadatak često imamo i procjenu prave proporcije (i kontingenta). Dok aritmetička sredina daje kvantitativnu, proporcija daje kvalitativnu informaciju o skupu, tj. informaciju o strukturi mase po nekom numeričkom ili atributivnom obilježju.

Proporcije se u praksi svode na dvoklasnu (dihotomnu) podjelu mase, čime se uprošćuje problem istraživanja. Skup se dijeli na jedinice koje imaju neku vrijednost kod numeričkog obilježja ili modalitet kod atributivnog obilježja (oznaka *p*) i one koje to nemaju (oznaka *q*). Odatle slijedi da je p + q = 1.

Da bismo na osnovu proporcije dobijene u uzorku ocijenili proporciju u skupu pozivamo se na teorijski raspored proporcija. Obično se pribjegava velikim uzorcima koji ispunjavaju uslove np > 5 i nq > 5, za koje se onda može primijeniti normalni raspored. *Kontingent* se dobija množenjem granica intervala procjene sa ukupnim brojem jedinica (N). Jednačina za procjenu proporcije glasi:

$$P = p \pm z s_p$$

gdje je:

$$s_p = \sqrt{\frac{p \cdot q}{n - 1}}$$

gdje je:

P – proporcija skupa

p – proporcija uzorka

z – broj standardnih grešaka (prema normalnom rasporedu)

s_p – standardna greška proporcije

2.6. Određivanje (planiranje) veličine uzorka

Planiranje uzoraka obuhvata:

1. Izbor tipa uzorka i
2. Određivanje veličine uzorka

O tipovima uzoraka već smo govorili. Ostalo je da nešto, više informativno, kažemo o određivanju veličine uzorka. Radi se o kompleksnoj problematici.

Određivanje veličine jednostavnog uzorka

Polazeći od osnovne jednačine reprezentativnog metoda možemo doći do formule za određivanje veličine uzorka za procjenu aritmetičke sredine skupa. Dakle, ako u formuli

$$\overline{X} = \overline{x} \pm t \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

zanemarimo faktor konačnosti i prebacimo aritmetičku sredinu uzorka na lijevu stranu dobijamo izraz

$$\overline{X} - \overline{x} = \pm t \cdot \frac{s}{\sqrt{n}}$$

Razlika između aritmetičke sredine skupa ($\overline{X}$) i aritmetičke sredine uzorka ($\overline{x}$) predstavlja *apsolutnu grešku uzorka (d)*. Jednostavnim računom dolazimo do formule za određivanje veličine uzorka (n):

$$n = \frac{t^2 \cdot s^2}{d^2}$$

U šumarskoj praksi obično se uzima da je t = 2, dok standardnu devijaciju (ako nije poznata) grubo procjenjujemo. Greška uzorka, koja se može tolerisati, obično bude propisana. U knjizi možete pogledati primjere (3, str. 239).

Umjesto prethodne često se koristi praktičnija formula sa koeficijentom varijacije (KV%) i relativnom greškom uzorka (d%):

$$n = \frac{t^2 \cdot KV_{\%}^2}{d_{\%}^2}$$

Pojmovi koje koristimo kod planiranja uzoraka

Osnovni cilj kod planiranja uzoraka je da uzorak bude što efikasniji, a to podrazumijeva preciznost i ekonomičnost, uz odgovarajuću pouzdanost. Potrebno je naravno da razjasnimo ove pojmove. Njima treba dodati još tačnost rezultata.

Preciznost. Preciznost procjene parametara skupa na osnovu uzorka određena je intervalom procjene $\left(\pm ts_{\bar{x}}\right)$, odnosno greškom uzorka $\left(ts_{\bar{x}}\right)$. Što je interval procjene uži, tj. greška uzorka manja preciznost je veća. Pri istoj pouzdanosti, osim izborom tipa uzorka na preciznost se može uticati povećanjem uzorka. Pri tome treba imati u vidu ekonomičnost uzorka.

Ekonomičnost. Uzorak je ekonomičniji ako ima manje troškove, pri istoj preciznosti. Pitanje troškova odnosno izvodljivosti uzorka u šumarstvu je posebno važno: da li za premjer šume na velikom prostoru možemo obezbijediti dovoljno stručne radne snage, koliko nam treba vremena i hoćemo li imati dovoljno sredstava na raspolaganju?

Efikasnost. Ako se istovremeno vodi računa o preciznosti i troškovima govorimo o efikasnosti uzorka.

Efikasnost procjene

Preciznost + Ekonomičnost = Efikasnost

Pouzdanost. Kada govorimo o grešci uzorka uvijek moramo naglasiti na bazi koje vjerovatnoće je određena ili drugim riječima, kolika je pouzdanost naše procjene. U šumarstvu se najčešće koristi pouzdanost 95%. Ako bismo povećali pouzdanost na 99% interval procjene bi bio širi, odnosno procjena bi bila manje precizna.

Pouzdanost procjene

Vjerovatnoća nakon izvršene procjene prelazi u pouzdanost.

Tačnost rezultata

U statističkom radu polazi se od pretpostavke da su podaci uzorka tačni, tj. da se greške nalaze u granicama slučajnih (neizbježnih) odstupanja. Međutim, posebno u šumarstvu, mogu se pojaviti sistematske i grube greške u mjerenjima na terenu. To su *greške izvan uzorka*, tzv. *tehničke greške*. Ove greške se otklanjaju povećanom pažnjom i kontrolama. Dakle, važno je razlikovati pojmove tačnost i preciznost. Statistika je „odgovorna" samo za preciznost. Naš interes je, naravno, da tačnost bude zadovoljavajuća.

Tačnost procjene

Greška uzorka + Greška izvan uzorka = Tačnost procjene

Planiranje uzoraka je složena problematika kako sa stanovišta statistike tako i sa stručnog aspekta. U šumarstvu se Pravilnikom za uređivanje šuma propisuje dozvoljena greška procjene najvažnijih obilježja (taksacionih elemenata) šuma. To je rezultat obimnog naučnog i stručnog rada. Greška se iskazuje relativno, tj. u procentima. *Napomena*: Treba razlikovati grešku uzorka, koja je jednaka približno dvostrukoj standardnoj grešci i grešku u zaključivanju, koja je obično određena unaprijed i iznosi $\alpha = 0{,}05$.

U našoj praksi za širu kategoriju šuma (ŠKŠ) 1000, odnosno sve njene uže kategorije šuma (UKŠ) i gazdinske klase (GK) primjenjuje se stratifikovani proporcionalni uzorak u vidu jedinstvene mreže krugova 100 x 100 metara. Jedan krug „pada" na jedan hektar, što znači da broj krugova odgovara (jednak je) broju hektara. Na ovaj način veće kategorije šuma imaju više krugova, tj. apsolutno veći uzorak. To praktično znači da je njihova greška procjene uvijek manja.

Teorija vjerovatnoće je "most" između uzorka i skupa

Uzorci u šumarstvu

Šuma posmatrana kao statistički skup je vrlo složena i specifična. Ne ulazeći u sve segmente toga navešćemo samo tipove uzoraka, koji se kombinuju u našoj praksi uređivanja, pri taksacionoj procjeni šuma. Osnovu procjene glavnih karakteristika šume čini *stratifikovani sistematski uzorak* sa proporcionalnim načinom izbora krugova. Stratume predstavljaju šire kategorije šuma, uže kategorije šuma i gazdinske klase. S obzirom da postoje obični krugovi (za procjenu zalihe drvne mase) i detaljni krugovi (za procjenu prirasta drvne mase) uzorak ima karakter *dvostepenog uzorka*. Izbor stabala na krugovima vrši se metodom koncentričnih krugova, što predstavlja *uzorak nejednake vjerovatnoće izbora* elemenata. Treba imati na umu da se pri premjeru šuma prikuplja veliki broj informacija. O tome više na drugim predmetima studija.

Vilijam Goset
(1876 – 1937)

Engleski statističar, koji je opisao t-raspored vjerovatnoće kod malih uzoraka. Kako se potpisivao pod pseudonimom Student, t-test je nazvan Studentov test.

3. STATISTIČKI TESTOVI

3.1. Opšti dio

3.1.1. Osnovno o statističkim testovima

Statističke procjene na bazi uzoraka su uobičajeno sredstvo naše informacije o različitim problemima u prirodi i društvu. Ipak, svaka takva procjena ili zaključivanje je manje ili više varijabilna (nesigurna), pa zahtijeva provjeru. Vršeći te provjere polazimo od tvrdnji ili pretpostavki koje nazivamo *statističke hipoteze*. Na primjer: da li je jedan način sadnje bolji od drugog, postoji li razlika u debljini stabala dvije šume, odgovara li debljinskoj strukturi neke šume normalan raspored, da li je neki preparat u zaštiti šuma bolji od drugoga itd. U vezi s tim, na postavljena pitanja moguća su samo dva odgovora, *da* ili *ne*. Odgovor na ta pitanja daju statistički testovi.

Kod statističkih testova postoje samo dva odgovora:

DA ili *NE*.

Nulta hipoteza (H_0)

Uzroci prirodnog (imanentnog) varijabiliteta u statistici se nazivaju slučajnim uzrocima. Rezultati mogu varirati i do beskonačnosti, ali je uobičajeno da se krajevi koji iznose površinski 5% smatraju područjem gdje variranje više nije slučajno. Na tom principu i nultoj hipotezi zasnovano je statističko testiranje.

Nulta hipoteza se postavlja tako da je razlika koja se testira jednaka nuli. Ona se prihvata ili odbacuje s određenom vjerovatnoćom. To znači: ako nultu hipotezu prihvatamo razlike su statistički slučajne, a ako je odbacujemo razlike

su statistički značajne. Statistička značajnost ne podrazumijeva i praktičnu značajnost, jer ona zavisi od veličine apsolutne razlike.

Nulta hipoteza glasi:

Između testiranih vrijednosti nema razlike, tj. razlika je jednaka nuli. Nultu hipotezu prihvatamo (ne odbacujemo) ili odbacujemo, uz rizik greške α.

Nulta hipoteza (H_0), može se napisati na dva načina. Ako testiramo aritmetičku sredinu to bi izgledalo ovako:

$$H_0 : \ \bar{x} - \bar{X} = 0$$

ili

$$H_0 : \ \bar{x} = \bar{X}$$

Dva tipa greške

Prilikom testiranja možemo napraviti dvije vrste grešaka. Ako smo odbacili H_0, a ona je tačna (istinita) napravili smo *grešku tipa I* (oznaka α), a ako smo prihvatili H_0, a ona je pogrešna, učinili smo *grešku tipa II* (oznaka β). U praktičnom radu vodimo računa samo o grešci tipa I.

Treba imati u vidu da mi zapravo nikad ne (sa)znamo da li je naša polazna pretpostavka (hipoteza H_0) tačna. Zato se u literaturi upozorava da nije ispravno reći „prihvatamo H_0", jer se njena tačnost ne može dokazati, već „ne odbacujemo H_0". U čemu je razlika? Pogledajmo na primjeru suđenja. Sud ako nema dovoljno dokaza kaže optuženi „nije kriv", a ne optuženi „je nevin". Time se naglašava samo da sud nije uspio skupiti dovoljno dokaza da optuženog proglasi krivim, ali isto tako nije ni dokazao da je optuženi nevin. I ovdje, na sudu, moguće su dvije vrste grešaka: može se desiti da optuženi bude osuđen iako je nevin ili da bude oslobođen iako je stvarno kriv. To su statistički gledano greške tipa I i tipa II.

Dva koncepta testiranja

Postoje dva koncepta (načina) testiranja. To su:

1. Klasični način testiranja
2. Testiranje pomoću p-vrijednosti

Mi ćemo se prvo upoznati sa klasičnim načinom testiranja, a onda, na samom kraju, sa novim konceptom testiranja pomoću p-vrijednosti.

3.1.2. Logika i postupak testiranja

Logika testiranja

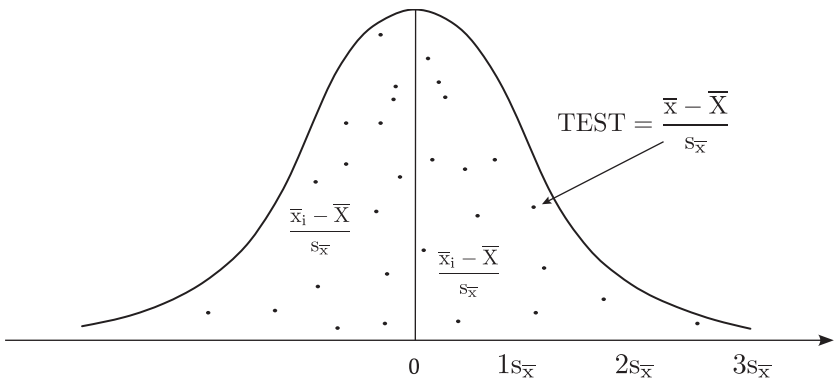
Ako testiramo razliku između sredine uzorka i sredine skupa, treba nam neki indikator koji bi mjerio tu razliku. Ako bismo za taj indikator uzeli apsolutnu razliku, ona bi bila iskazana u mjernim jedinicama obilježja i mijenjala bi vrijednost sa promjenom mjerne jedinice. To znači da bi takav indikator bio beskoristan. Do rješenja se dolazi tako što apsolutnu razliku podijelimo sa standardnom greškom, u ovom slučaju standardnom greškom sredine uzorka. Dobijamo relativnu vrijednost, koja je zapravo rezultat statističkog testa.

Logika testiranja

Ako je H_0 tačna (istinita) izračunata vrijednost testa će biti mala, dok će u suprotnom biti velika.

Formule za računanje statistike (vrijednosti) testova su definisane (kreirane) upravo tako da njihov rezultat bude relativan broj. Te statistike imaju karakter slučajne promjenljive, jer se razlikuju od uzorka do uzorka. One imaju svoj raspored (model), na osnovu koga mi izvodimo zaključke o veličini rezultata testa (statističkom značaju razlike). Dakle, svaki test slijedi (ima) neki teorijski raspored vjerovatnoće. Taj raspored pokazuje vjerovatnoće za konkretne vrijednosti testa. Radi se o tzv. p-vrijednosti, o kojoj ćemo govoriti kasnije.

Logika testiranja na bazi jednog uzorka je sljedeća. Teorijski, kad se iz jednog skupa uzmu svi slučajni uzorci i izračunaju statistike testa, one će imati svoj raspored kao na *Slici 41*. Ako je nulta hipoteza tačna te statistike će se grupisati oko centra rasporeda, tj. oko nule.



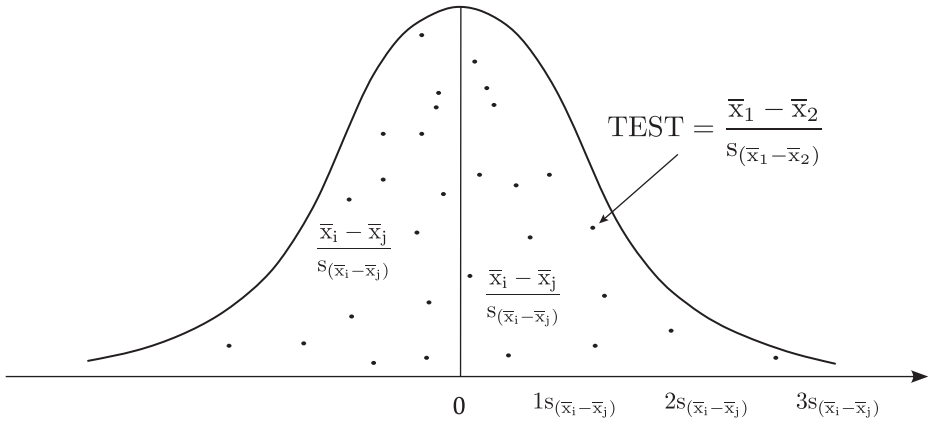
OPŠTI OBLIK TESTA:

$$\text{TEST} = \frac{\text{parametar uzorka} - \text{parametar skupa}}{\text{standardna greška parametra uzorka}}$$

Slika 41. Nulti raspored statistike testa – jedan uzorak

Drugim riječima, taj raspored koji nazivamo *nulti raspored* će imati normalni oblik ili oblik t-rasporeda. To znači da možemo, polazeći od pouzdanosti 95%, prihvatiti svaku nultu hipotezu čija se vrijednost testa nalazi unutar 95% površine.

Kao što smo napravili vještački skup od svih mogućih razlika sredina uzoraka od sredine skupa tako isto možemo doći do vještačkog skupa razlika sredina između dva uzorka. Razlike se ponašaju kao slučajne varijable (promjenljive), jer su i sredine uzoraka slučajne varijable. Oblik njihovog raspoređivanja, tj. statistike testova (gdje svaka tačka predstavlja rezultat jednog testa) slijedi oblik t-rasporeda, pa se primjenjuje t-test *(Slika 42)*. Ova slika, zapravo, ilustruje nultu hipotezu. Provodimo test i izvodimo zaključak o udaljavanju rezultata od nule.



OPŠTI OBLIK TESTA:

$$\text{TEST} = \frac{\text{parametar prvog uzorka} - \text{parametar drugog uzorka}}{\text{standardna greška razlike parametara}}$$

Slika 42. Nulti raspored statistike testa – dva uzorka

Veća vrijednost testa (izračunata po formuli) ukazivaće na veću razliku, odnosno da smo bliže odbacivanju H_0. Postavlja se samo pitanje koja je to granica, koju izračunata vrijednost testa treba da pređe da bi H_0 bila odbačena. Obično se uzima $\alpha = 0{,}05$. Tu graničnu (kritičnu) vrijednost nalazimo u odgovarajućim tablicama (t-rasporeda ili nekog drugog rasporeda). Nju smo koristili i kod uzoraka, gdje je predstavljala broj standardnih grešaka prilikom određivanja intervala procjene.

Veza između testova i intervala procjene. Između testiranja hipoteza i intervala procjene postoji očigledna povezanost. Ako se hipotetička vrijednost dvosmjerne hipoteze nalazi unutar intervala procjene ona prilikom testiranja neće biti odbačena, dok će svaka druga vrijednost, koja se nalazi izvan intervala, biti odbačena.

Berksonov paradoks. Statistički induktivni način razmišljanja kaže da je s povećanjem uzorka ocjena parametra skupa preciznija. Međutim, kod dvosmjernih testova povećanje uzorka dovodi do toga da i beznačajno mala razlika može postati statistički značajna. Razlog je smanjenje standardne greške.

Postupak (etape, koraci) kod klasičnog testiranja

1. Na osnovu konkretnog problema definiše se nulta hipoteza.
2. Odabere se nivo značajnosti (rizik greške α).
3. Uzima se slučajni uzorak i računaju njegovi parametri (statistika uzorka).
4. Bira se test i izračunava vrijednost (statistika) testa (izračunate vrijednosti imaju indeks nula: z_0, t_0, F_0 ili χ_0^2).
5. Očitava se granična (kritična) vrijednost iz tablica teorijskog rasporeda (tablične vrijednosti imaju indeks jedan: z_1, t_1, F_1 ili χ_1^2).
6. Uporedi se izračunata vrijednost testa sa tabličnom.
7. Donosi odluka o prihvatanju ili odbacivanju nulte hipoteze.
8. Izvodi zaključak.

Napomena: Pri testiranju treba imati u vidu preduslove za primjenu nekog testa i voditi računa o zaključivanju (interpretaciji rezultata).

3.1.3. Vrste i podjela testova

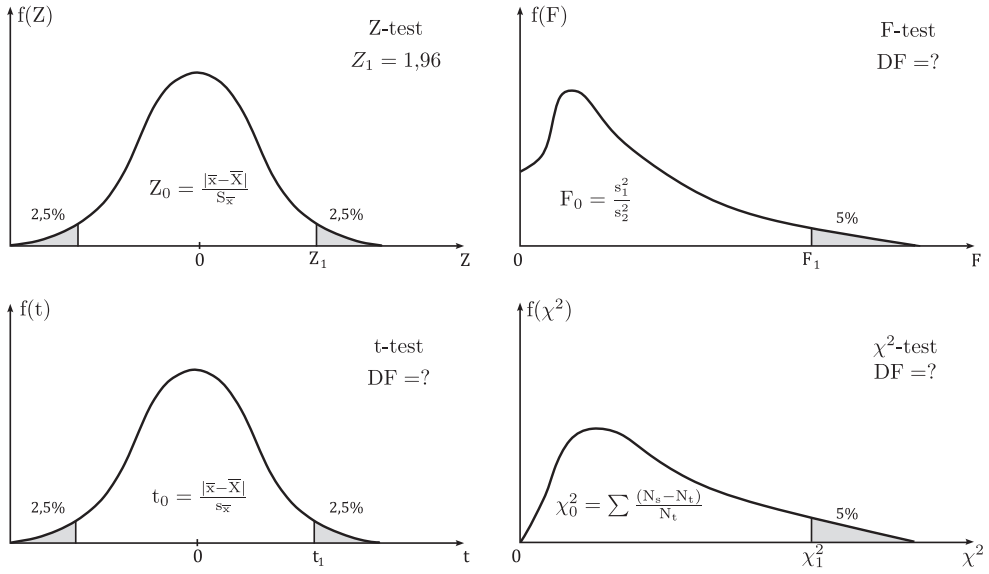
Vrste statističkih testova

Testovi koji se koriste za testiranje različitih karakteristika skupova (sredine i drugih) baziraju se na istoimenim teorijskim rasporedima. Po njima su i dobili nazive *(Slika 43)*. Najčešće se upotrebljavaju:

1. z-test (cet test)
2. t-test (Studentov test)
3. F-test (Fišerov test)
4. χ^2-test (hi-kvadrat test)

Svi testovi (z-test, t-test, F-test i χ^2-test) imaju svoje formule za računanje testa. Ti rezultati su standardizovane vrijednosti (relativni, neimenovani brojevi). Nalaze se na x osi. Z-test i t-test su dvostrani testovi, bazirani na simetričnom rasporedu. T-test se koristi za testiranje razlike parametara, dok se z-test, osim kao školski test, koristi kao zamjena za t-test kod velikih uzoraka.

F-test i χ^2*-test* su jednostrani testovi, a njihovi teorijski rasporedi su desno asimetrični. F-test se koristi za testiranje varijabiliteta preko količnika varijansi, pa se naziva *test količnika*. Primjenjuje se u eksperimentalnim istraživanjima, a metod je poznat pod nazivom analiza varijanse. χ^2-test ima široku upotrebu, a najčešće služi za testiranje rasporeda frekvencija.



Slika 43. Teorijski rasporedi (z, t, F, χ^2) i statistički testovi (z-test, t-test, F-test i χ^2-test)

Zaključke kod testiranja donosimo uobičajeno na bazi rizika od 5%, koji se kod z i t rasporeda ravnomjerno raspoređuje na lijevoj i desnoj strani (po 2,5%), dok je kod F i χ^2 testa sav rizik na desnoj strani. Kod z-testa granična vrijednost je fiksna (z_1 = 1,96), a kod ostalih testova zavisi od stepena slobode (*DF*, od engl. *Degress of Freedom*).

Podjela statističkih testova

Postoje dvije klasifikacije testova: na bazi oblika rasporeda skupa i prema broju uzoraka. Prema obliku rasporeda, testovi se svrstavaju (klasifikuju) u dvije grupe:

1. *Parametarski testovi*. To su klasični testovi koji zahtijevaju poznavanje oblika rasporeda skupa (npr. z, t i F). Oni su nastali još u vrijeme kada se mislilo da praktično postoji samo jedan, tj. normalni raspored. Dakle, oni polaze od uslova da osnovni skup iz koga se uzima uzorak ima normalni raspored.

2. *Neparametarski* testovi. Ovi testovi su novijeg datuma. Oni ne zavise od oblika rasporeda osnovnog skupa. Njih ima veliki broj, npr: hi-kvadrat, test medijane, Fridmanov test, Kruskal-Valisov test. Mi ćemo obraditi samo hi-kvadrat test. Ovaj test jednim dijelom pripada klasičnim testovima.

Po drugoj klasifikaciji, prema broju uzoraka, testovi se svrstavaju u tri grupe:

1. Testovi na bazi jednog uzorka (z-test, t-test)
2. Testovi na bazi dva uzorka (z-test, t-test, F-test)
3. Testovi na bazi tri ili više uzoraka (F-test).

U našoj knjizi (3) testovi su obrađeni po parametrima: testiranje sredina, testiranje proporcija, testiranje varijansi itd. U drugoj literaturi testovi se obrađuju uglavnom prema broju uzoraka koji se koriste (testovi na bazi jednog uzorka, testovi na bazi dva uzorka itd). Mi smo testove obradili po vrstama (tipovima) testova: z-test, t-test, F-test i χ^2-test. Za drugi nivo smo uzeli broj uzoraka, a za treći parametre. Može se još praviti razlika između testova na bazi nezavisnih i zavisnih uzoraka, kao i između dvosmjernih i jednosmjernih testova.

3.2. Z-test i t-test

Od svih klasičnih testova najviše se upotrebljava *t-test*. S obzirom na to da se koristi za testiranje razlike parametara naziva se još „*test razlike*". Obično se testiraju aritmetička sredina i proporcija, a rjeđe drugi parametri, kao što su standardna devijacija, koeficijent varijacije ili koeficijent korelacije. Umjesto t-testa može se primjeniti *z-test* ako su uzorci veliki i ako je poznata standardna devijacija osnovnog skupa.

Nulta hipoteza kod ovih testova može se postaviti na više načina, zavisno od problema koji se rješava. Obično se postavlja kao *dvosmjerna,* kada se ispituje samo apsolutna razlika, tj. kad nije važan predznak razlike. Složeniji oblik je *jednosmjerna* hipoteza. Mi ćemo obraditi samo dvosmjerne hipoteze. Prvo na bazi jednog uzorka, a potom na bazi dva uzorka.

3.2.1. Testovi na bazi jednog uzorka

3.2.1.1. Testiranje aritmetičke sredine

Formule za računanje vrijednosti (statistike) z-testa i t-testa izvedene su iz jednačine procjene sredine skupa na bazi uzorka, tj. jednačine reprezentativnog metoda. Razlika između ova dva testa je samo u tome što se standardna greška kod z-testa računa na bazi standardne devijacije skupa, a kod t-testa na bazi standardne devijacije iz uzorka. Oznake u narednim formulama su poznate od ranije.

$$\bar{X} = \bar{x} \pm zS_{\bar{x}} \qquad \bar{X} = \bar{x} \pm ts_{\bar{x}}$$

$$\downarrow \qquad \downarrow$$

$$z = \frac{\bar{x} - \bar{X}}{S_{\bar{x}}} \qquad t = \frac{\bar{x} - \bar{X}}{s_{\bar{x}}}$$

Uzmimo jedan hipotetički primjer. Neka je, prema tablicama, temeljnica neke šume 42 m^2/ha. Da li ona stvarno može biti tolika provjeravamo t-testom. Uzimamo uzorak i računamo njegove parametre: veličina uzorka n = 15, aritmetička sredina $\bar{x}$ = 40,72 m^2/ha i standardna devijacija s = 11,54 m^2/ha.

Nulta hipoteza (H_0) glasi:

$$H_0 : \ \bar{x} = \bar{X}$$

Slijedi računanje vrijednosti testa (t_0), gdje standardnu grešku sredine ($s_{\bar{x}}$) računamo po poznatoj formuli. Potom očitavamo tabličnu vrijednost (t_1) iz tablice t-rasporeda, gdje su ulazi rizik greške (α) i stepeni slobode (DF). Uspoređujemo dvije t-vrijednosti i izvodimo zaključak *(Slika 44)*.

$$t_0 = \frac{\bar{x} - \bar{X}}{s_{\bar{x}}} = \frac{40,72 - 42,00}{2,98} = 0,43$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{11,54}{\sqrt{15}} = 2,98 \ m^2 / ha$$

$$\alpha = 0,05; \ \ DF = (n-1) = 14$$

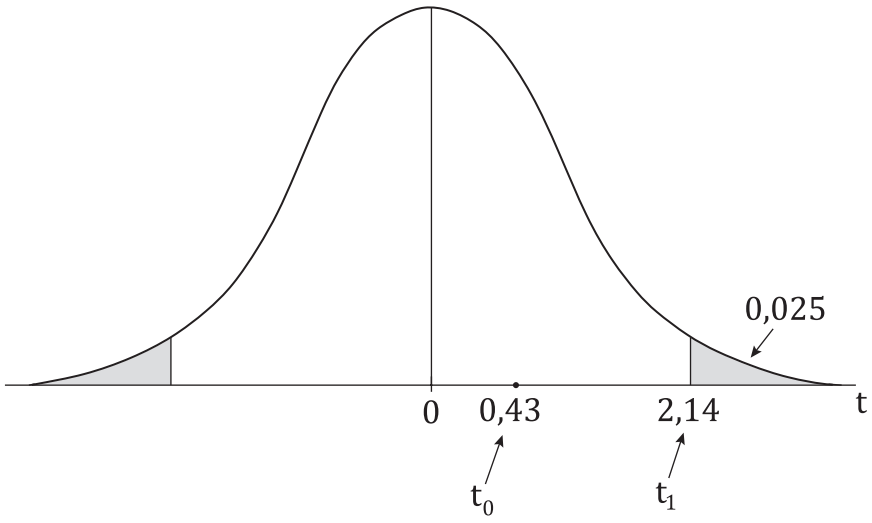
$$t_1 = 2,14$$

$$t_0 < t_1$$

Pošto je izračunata vrijednost testa manja od granične vrijednost iz tablica t-rasporeda konstatujemo da nema statistički značajne razlike između sredine uzorka i očekivane sredine skupa. Drugim riječima, prihvatamo H_0 i zaključujemo da temeljnica na terenu odgovara temeljnici prema šumarskim tablicama.

Zaključivanje

Ako je $t_0 < t_1$ prihvatamo H_0, a ako je $t_0 > t_1$ odbacujemo H_0, uz α=0,05.



Slika 44. t-test na bazi jednog uzorka

3.2.1.2. Testiranje proporcije

Procjena proporcije na malom uzorku je nesigurna. Na primjer, ako iz bubnja u kome se nalaze bijele i crvene loptice izvučemo mali uzorak nećemo moći sa sigurnošću zaključiti kakav je odnos loptica u bubnju. Može se desiti da od tri izvučene loptice sve tri budu bijele. Tada bismo zaključili da se u bubnju nalaze samo bijele loptice. Zato se kod proporcije kao uslov postavlja uzimanje velikih uzoraka. Kako smo već rekli, taj uslov je:

$$np > 5 \text{ i } nq > 5$$

Dakle, ako želimo testirati proporciju (*P*) u nekom skupu mi ćemo uzeti veliki uzorak iz tog skupa i odrediti njegovu proporciju (*p)*. Primjenićemo z-test i provesti identičan postupak kao kod testiranja aritmetičkih sredina. Ovdje dajemo samo formule, a primjer pogledajte u knjizi (3, str. 281).

Nulta hipoteza za test proporcije glasi:

$$H_0: \ p = P$$

Vrijednost z-testa računa se po formuli:

$$z_0 = \frac{p - P}{s_p}$$

gdje je:

z_0 – izračunata vrijednost testa

p – proporcija uzorka

P – proporcija skupa

s_p – standardna greška proporcije uzorka:

$$s_p = \sqrt{\frac{pq}{n-1}}$$

Zaključivanje

Ako je $z_0 < 1{,}96$ prihvatamo H_0, a ako je $z_0 > 1{,}96$ odbacujemo H_0 ($\alpha = 0{,}05$).

3.2.1.3. Testiranje korelacije

Testiranje koeficijenta korelacije izvodi se t-testom. Testira se (utvrđuje) da li postoji korelacija u skupu (*R*) između dva obilježja. Uzima se uzorak i računa koeficijent korelacije (*r*). Nulta hipoteza se postavlja tako da nema razlike između koeficijenta korelacije uzorka i koeficijenta korelacije u skupu (R), koji se izjednačava s nulom (zato što testiramo postojanje korelacije u skupu). Ako se utvrdi da je izračunato t_0 manje od t_1 tablično prihvatićemo H_0 i zaključiti da između dva obilježja ne postoji korelacija u skupu. U suprotnom odbacujemo H_0 i konstatujemo da se koeficijent korelacije (*r*) iz uzorka značajno razlikuje od nule.

Postupak računanja testa:

$$H_0: r = 0$$

$$t_0 = \frac{r - R}{s_r} = \frac{r}{s_r} \qquad (R = 0)$$

Standardna greška koeficijenta korelacije u uzorku:

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Zaključivanje

Ako je $t_0 < t_1$ prihvatamo H_0, što znači da u skupu ne postoji korelacija, a ako je $t_0 > t_1$ obacujemo H_0; u skupu postoji korelacija.

3.2.2. Testovi na bazi dva uzorka

3.2.2.1. Testiranje aritmetičkih sredina dva uzorka

Za testiranje razlike aritmetičkih sredina dva skupa potrebna su nam dva uzorka. Za primjer ćemo uzeti uzorak koji već imamo (pod 3.2.1.1), dok ćemo za drugi pretpostaviti parametre. Neka oba uzorka imaju po 15 elemenata. Postavljamo nultu hipotezu tako da nema razlike između sredina dva uzorka, računamo standardnu grešku razlike sredina i statistiku testa, biramo rizik (α = 0,05), očitavamo tabličnu vrijednost (t_1) i izvodimo zaključak.

PRVI UZORAK

$$\bar{x}_1 = 40,72 \text{ m}^2;\ s_1^2 = 133,22$$

DRUGI UZORAK

$$\bar{x}_2 = 35,60 \text{ m}^2;\ s_2^2 = 67,90$$

$$H_0 :\ \bar{x}_1 = \bar{x}_2$$

$$t_0 = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{(\bar{x}_1 - \bar{x}_2)}} = \frac{|40,72 - 35,60|}{3,66} = \frac{5,12}{3,66} = 1,40$$

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{133,22}{15} + \frac{67,90}{15}} = 3,66 \text{ m}^2$$

t_1 → $\alpha = 0,05$

t_1 → $= 2,04$

t_1 → $DF = n_1 + n_2 - 2 = 18$

$$t_0 < t_1$$

Izračunata vrijednost testa manja je od granične iz tablica. Prihvatamo H_0 i konstatujemo da nema statistički značajne razlike između sredina dva uzorka, odnosno između temeljnica dvije šume.

Zaključivanje

Ako je $t_0 < t_1$ prihvatamo H_0, a ako je $t_0 > t_1$ odbacujemo H_0, uz α=0,05.

3.2.2.2. Testiranje proporcija dva uzorka

Za testiranje proporcija dva skupa potrebna su nam dva velika uzorka. Postupak je sličan testiranju sredina. Nakon što uzmemo uzorke i utvrdimo njihove proporcije postavljamo nultu hipotezu (primjer, u knjizi 3, str. 283):

$$H_0: \ p_1 = p_2$$

Formula za z-test glasi:

$$Z_0 = \frac{p_1 - p_2}{s_{(p_1 - p_2)}}$$

gdje je:

z_0 – izračunata vrijednost testa

p_1 - proporcija prvog uzorka

P_2 - proporcija drugog uzorka

$s_{(p_1 - p_2)}$ – standardna greška razlike proporcija, računa se po formuli:

$$s_{(p_1 - p_2)} = \sqrt{\frac{p_1 q_1}{n_1 - 1} + \frac{p_2 q_2}{n_2 - 1}}$$

Zaključivanje

Ako je $z_0 < 1{,}96$ prihvata se H_0, a ako je $z_0 > 1{,}96$ odbacuje se H_0 (α=0,05).

3.2.2.3. Test na bazi dva zavisna uzorka (test parova)

Ovo je specifičan t-test. Uzorci se biraju po nekom planu, pa spada u planiranje eksperimenata. Za dva zavisna uzorka uzimaju se parovi podataka i računaju razlike svakog para (otuda naziv metod parova). Ne računaju se sredine uzoraka, već aritmetička sredina razlika između svih parova. Testiranje razlike te sredine od nule je nulta hipoteza kod ovog testa:

$$H_0: \ \bar{d} = 0$$

Formula za računanje testa glasi:

$$t_0 = \frac{\bar{d} - 0}{s_{\bar{d}}} = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

gdje je:

t_0 – izračunata vrijednost testa

$\bar{d}$ – aritmetička sredina odstupanja

$s_{\bar{d}}$ – standardna greška sredine odstupanja (srednjeg odstupanja)

Kao i kod drugih oblika t-testa izračunata vrijednost se uspoređuje s tabličnom (t_1). Ulaz u tablice, pored uobičajenog ririka od 5%, je stepen slobode DF = n – 1, gdje je n broj parova.

Zaključivanje

Ako je $t_0 < t_1$ prihvata se H_0, a ako je $t_0 > t_1$ odbacuje se H_0 (α=0,05).

U čemu je prednost ovog testa u odnosu na obični t-test pogledajmo na primjerima. U rasadniku za proizvodnju šumskih sadnica često ispitujemo uticaj đubrenja na njihov rast. Kako bismo to utvrdili treba da isključimo uticaje drugih faktora, npr. plodnost zemljišta. Parcele iste plodnosti podijelićemo na dva dijela, jedan đubrimo, a drugi ostavljamo bez tretmana. Uzimajući parove podataka dobijamo zavisne uzorke.

Test parova se takođe primjenjuje u slučajevima kad testiramo dva instrumenta za mjerenje visine (uporedićemo ih mjerenjem na istim stablima) ili dva različita dendrometrijska metoda (provjeravamo ih na istoj šumi). Konkretne primjere ovog testa pogledajte u knjizi (3, str. 278).

3.3. F-test

3.3.1. F-test na bazi dva uzorka

Testiranje varijabiliteta malih uzoraka izvodi se preko njihovih varijansi. Nulta hipoteza kod ovog testa glasi:

$$H_0: s_1^2 = s_2^2$$

Koristi se F-test. Izračunate vrijednosti kod ovog testa su uvijek veće od jedan ($F_0 > 1$), jer se iz praktičnih razloga uzima da je u brojniku veća varijansa, tj. $s_1^2 > s_2^2$. Formula testa glasi:

$$F_0 = \frac{s_1^2}{s_2^2}$$

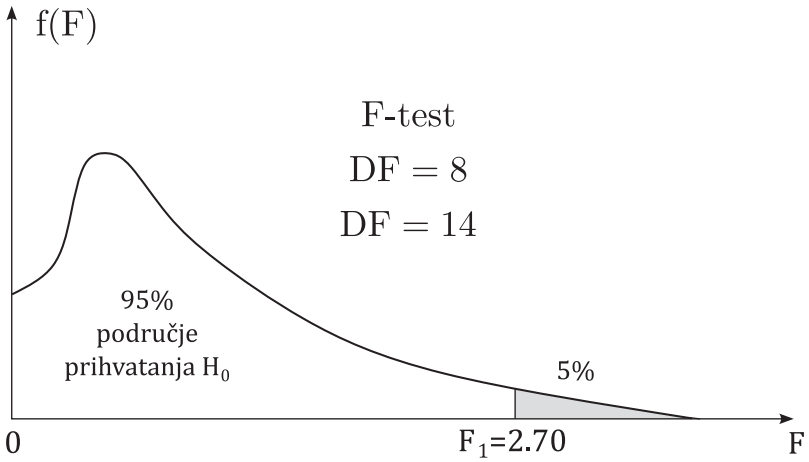
gdje je:

F_0 – izračunata vrijednost testa

s_1^2 – veća varijansa

s_2^2 – manja varijansa

Kritičnu vrijednost F_1 čitamo iz tablice F-rasporeda, koja se odnosi na rizik $\alpha = 0,05$. Na primjer, ako je stepen slobode veće varijanse 8 (čita se u zaglavlju), a manje varijanse 14 (čita se u predkoloni), granična vrijednost je $F_1 = 2,70$ *(Slika 45)*.



Slika 45. F-test

Nulta hipoteza je i ovdje polazna baza za rezonovanje. Ako je ona tačna onda male vrijednosti $F_{0,}$ koje se nalaze sasvim blizu nule, ukazuju na to da se te razlike javljaju često i da nastaju slučajno. Sa povećanjem vrijednosti testa razlike su sve rjeđe i manje vjerovatne. Kad neka razlika pređe graničnu vrijednost (ovdje $F_1 = 2,70$), njena vjerovatnoća pada ispod 5% i tada se za svako $F_0 > F_1$ ona smatra značajnom.

Zaključivanje

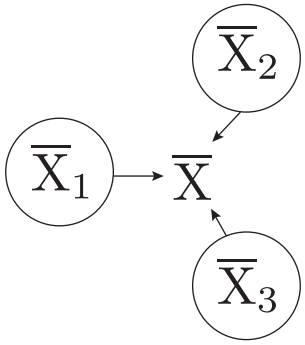
Ako je $F_0 < F_1$ prihvata se H_0, a ako je $F_0 > F_1$ odbacuje se H_0 (α=0,05).

3.3.2. F-test na bazi tri ili više uzoraka (analiza varijanse)

Testiranje hipoteze o jednakosti aritmetičkih sredina na bazi tri ili više uzoraka ne izvodi se pomoću t-testa. U tu svrhu u statistici se koristi metod koji se naziva *analiza varijanse*, skraćeno *ANOVA* (od engl. *analysis of variance*).

Analiza varijanse se primjenjuje u eksperimentalnim istraživanjima. Da bismo razjasnili suštinu ovog metoda uzmimo jedan primjer. Ako u šumskom rasadniku želimo da utvrdimo da li različita količina đubriva utiče na rast biljaka postavićemo eksperiment (ogled) sa tri različite količine đubriva (tri tretmana). Na tri površine, koje se u startu neće bitno razlikovati po drugim uslovima za rast biljaka (prije svega po plodnosti zemljišta), primjenićemo različite količine

đubriva, npr: ploha 1 najmanja, 2 srednja i 3 najveća količina. Nakon nekog vremena (perioda rasta) provjerićemo rezultate. Uzećemo po jedan slučajni nezavisni uzorak sa svake plohe i izmjeriti visine biljaka. Šematski izgled eksperimenta sa tri uzorka dat je na *Slici 46*.



Slika 46. Analiza varijanse

Imaćemo dakle tri uzorka (k = 3). Na svakoj od ovih ploha biljke će se po visini razlikovati od svoje aritmetičke sredine. Varijabilitet za sva tri uzorka (plohe) zajedno izračunaćemo kao ponderisanu sredinu, gdje kao pondere koristimo broj elemenata svakog uzorka (n_1, n_2, n_3). Ovdje se radi o varijabilitetu unutar ploha pa ga nazivamo *unutrašnji varijabilitet*. Iskazujemo ga unutrašnjom varijansom.

U ovom eksperimentu potrebno je još izračunati variranje između ploha. Taj varijabilitet nazivamo *vanjski varijabilitet*, a predstavljamo ga vanjskom varijansom. Računamo prvo zajedničku sredinu ($\overline{X}$) za tri plohe, a onda odstupanja sredina uzoraka od te zajedničke sredine. I ovdje vodeći računa o broju elemenata uzoraka.

Varijabilitet računamo preko kvadrata odstupanja, dobijajući na kraju njihove sume. Tako ovdje imamo dvije sume kvadrata, unutrašnju i vanjsku. Njihov zbir će predstavljati ukupni varijabilitet, odnosno ukupnu sumu kvadrata *(Tabela 6)*. Ukupni varijabilitet je odstupanje elemenata od zajedničke sredine.

UKUPNI VARIJABILITET =
VARIRANJE IZMEĐU UZORAKA + VARIRANJE UNUTAR UZORAKA

Mi treba da utvrdimo da li je đubrenje uticalo na rast biljaka, tj. da li se aritmetičke sredine naših uzoraka međusobno razlikuju. Nulta hipoteza u ovom slučaju glasi:

$$H_0: \ \overline{X}_1 = \overline{X}_2 = \overline{X}_3$$

Razliku između sredina nećemo ispitivati direktno, već indirektno preko varijansi, uz uslov njihove homogenosti. Logika nam kaže, ako je vanjski varijabilitet veći od unutrašnjeg to znači da će se i sredine uzoraka međusobno razlikovati. Značaj tih razlika utvrđujemo putem F-testa, iz odnosa vanjske i unutrašnje varijanse.

Ako je $F_0 < F_1$ prihvatićemo nultu hipotezu i konstatovati da ne postoje statistički značajne razlike između ploha. U suprotnom, ako je $F_0 > F_1$ zaključićemo da je vanjska varijansa značajno veća od unutrašnje, tj. da postoji uticaj đubrenja.

Zaključivanje

Ako je $F_0 < F_1$ nultu hipotezu usvajamo, a ako je $F_0 > F_1$ odbacujemo, uz α=0,05.

Tabela 6. F-test - analiza varijanse

Izvor varijacije	Stepeni slobode (DF)	Suma kvadrata	Varijansa (S^2)	F_0	F_1
Između uzoraka	k – 1	Vanjska	S_v^2	S_v^2 / S_u^2	F_1 (tablično)
Unutar uzoraka	k (n -1)	Unutrašnja	S_u^2		
Ukupno (total)	k n - 1	Ukupna	-	-	-

Ako je $F_0 > F_1$ test je pokazao statistički značajnu razliku, ali nije dao odgovor na pitanje između kojih uzoraka (tretmana) postoji ta razlika. U tu svrhu ranije se primjenjivao t-test, a danas postoje testovi, kao što je Dankanov ili Takijev test, pomoću kojih se istovremeno testiraju sve razlike. Više u knjizi (3, str. 307).

Planiranje eksperimenata

Postoje dva osnovna naučna pristupa istraživanju, posmatranjem (mjerenjem) pojava i putem eksperimenata. Mi prirodne pojave posmatramo uglavnom izvorno, onako kako se dešavaju, bez našeg uticaja. Izvođenje (odvijanje) prirodne pojave pod kontrolom istraživača naziva se *eksperiment*. To zahtijeva planski pristup. Time se bavi posebna naučna oblast, koja se naziva *planiranje eksperimenata*.

Potreba za eksperimentima u šumarstvu javlja se zbog toga što postoje različiti izvori (uzroci) varijabiliteta u živom biljnom svijetu (prirodi). To su:

1. Varijabilitet koji je imanentan (svojstven) prirodnoj pojavi, tzv. slučajni varijabilitet. On se ne može izbjeći.

2. Uslovi, tj. prirodni faktori pod kojima se dešava pojava.

3. Varijabilitet koji nastaje zbog našeg namjernog (eksperimentalnog) djelovanja, npr. tretman đubrenjem.

Plan (projektovanje) eksperimenata podrazumijeva izbor tipa eksperimenta, veličine i broja eksperimentalnih jedinica (uzoraka). Jedan od najjednostavnijih eksperimenata (*metod parova*) smo ranije opisali. Najčešći tipovi eksperimenata u šumarstvu su: *jednostavni slučajni eksperiment, slučajni blok sistem* i *latinski kvadrat.* Za više od jednog faktora koriste se tzv. *faktorijalni eksperimenti*. *Napomena*: planiranje eksperimenata spada u materiju II i III ciklusa studija.

3.4. Hi–kvadrat test

Hi-kvadrat test ima višestruku primjenu. Njime se testiraju frekvencije rasporeda, tj. podudarnost stvarnih i teorijskih (očekivanih) frekvencija. Nulta hipoteza glasi:

H_0: Između dva rasporeda frekvencija nema razlike

Formula za računanje testa po definiciji glasi:

$$\chi_0^2 = \sum \frac{(n_s - n_t)^2}{n_t}$$

gdje je:

χ_0^2 – izračunata vrijednost testa

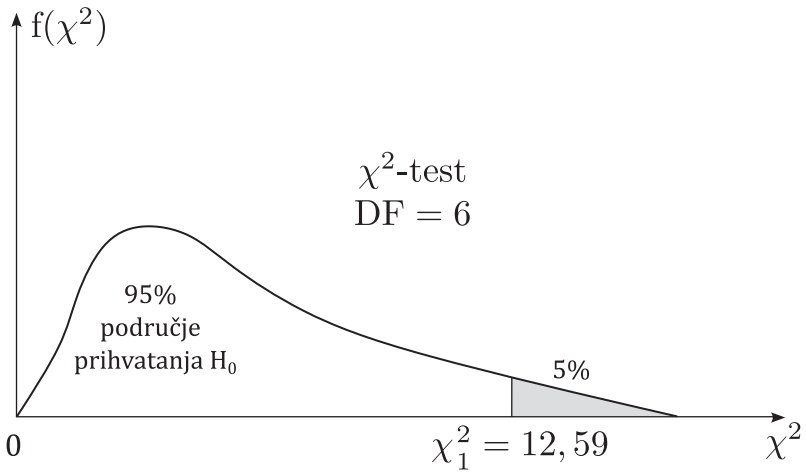
n_s – stvarne frekvencije

n_t – teorijske frekvencije

Pretpostavimo da je broj stepena slobode *DF = 6*. Iz tablice, za $\alpha = 0{,}05$ čitamo $\chi_1^2 = 12{,}59$ *(Slika 47)*. Svaka naša izračunata vrijednost testa koja je manja od granične znači prihvatanje nulte hipoteze, dok veća znači njeno odbacivanje. Primjer u knjizi (3, str. 293).

Zaključivanje

Ako je $\chi_0^2 < \chi_1^2$ prihvata se H_0, a ako je $\chi_0^2 > \chi_1^2$ odbacuje, uz $\alpha=0{,}05$.



Slika 47. Hi-kvadrat test

Kod primjene ovog testa treba obratiti pažnju na sljedeće:

- Prvo, ukoliko se razlikuju sume stvarnih i teorijskih frekvencija, razliku treba dodati krajnjim klasama.
- Zbog činjenice da je u krajnjim klasama često stvarna frekvencija velika, a teorijske frekvencije male, izračunata vrijednost Hi-kvadrat testa bude pristrasno velika. Zato se prije računanja testa sažimaju krajnje klase i to tako da frekvencija bude najmanje 5 (po nekim autorima 10).
- Graničnu vrijednost očitavamo iz tablice hi-kvadrat rasporeda. Ovdje treba obratiti pažnju na stepene slobode, koji zavise od toga kako su izračunate teorijske frekvencije. Na primjer, kod normalnog rasporeda imaju dva dodatna ograničenja (sredina i standardna devijacija). Broj stepena slobode biće *DF = k – 2 - 1*, gdje je k broj klasa nakon sažimanja. Ako se testiraju grupe biće *DF = k – 1*.

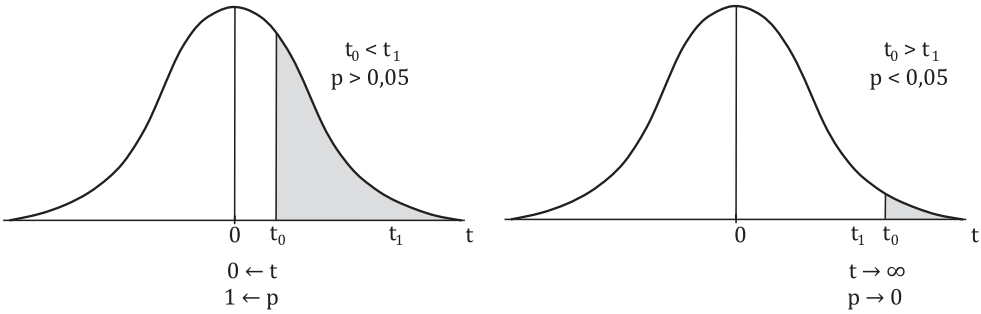
3.5. Softversko testiranje (p–vrijednost)

Još od 1930. godine *Fišer* (Ronald Fisher, 1890-1962) je zagovarao testiranje na osnovu p-vrijednosti. Pune tri decenije vodile su se rasprave između njega i grupe statističara koje su predvodili zagovornici klasičnog načina testiranja *Pirson* (Egon Pearson, 1895-1980) i *Nimen* (Jerzy Newman, 1894-1981). Nakon smrti Fišera 1962. godine prevladao je klasični način. Međutim, ulaskom u 21. vijek ovaj način počeo je da gubi na značaju. Računari su postali dostupni svima, a sa njima i računanje tzv. p-vrijednosti. Ona je zamijenila tablice i formule koje se koriste kod klasičnog testiranja, što je znatno pojednostavilo postupak testiranja.

Šta je p-vrijednost?

P-vrijednost je vjerovatnoća izračunate vrijednosti (statistike) nekog testa, koja se dobija konvertovanjem (preračunavanjem) vrijednosti testa na skalu od nula do jedan. Pogledajmo logiku i značenje ovog indikatora.

Ako je zaista nulta hipoteza tačna, statistike (rezultati) testa će se skoncentrisati oko nule (centra rasporeda), a vrijednosti udaljavanjem od centra postaju sve rjeđe. Na *Slici 48* predstavljena su osjenčenom površinom dva slučaja: kada je vrijednost testa mala i kada je ona velika. Ta osjenčena površina je u stvari *p-vrijednost*.



Slika 48. Statistika testa i p-vrijednost

Vidjeli smo ranije, ako se povećava apsolutna razlika između testiranih parametara vrijednost testa (relativna razlika) se povećava. Te razlike se smatraju slučajnim sve do neke granice, koja se uobičajeno uzima iz tablica za rizik $\alpha = 0,05$. U isto vrijeme, dok se vrijednost testa povećava p-vrijednost se smanjuje. Koliko ona treba da bude mala da bi se razlika smatrala značajnom? Logika nam govori da p-vrijednost treba izjednačiti sa $\alpha = 0,05$, tj. da granična vrijednost bude 5%. Pogledajte raniju *Sliku 43* na kojoj su ove granične vrijednosti, za četiri klasična testa, predstavljene zatamnjenim površinama.

P-vrijednost

P-vrijednost je vjerovatnoća izračunate vrijednosti testa,
tj. vjerovatnoća date (tolike) razlike.

Zaključivanje na bazi p-vrijednosti u principu je isto za sve testove. Ako je p-vrijednost veća od 5% razlike se smatraju statistički slučajnim, odnosno ako je p-vrijednost manja od 5% razlike se smatraju statistički značajnim.

Zaključivanje

Ako je $p > 0,05$ nulta hipoteza se prihvata, a ako je $p < 0,05$ ona se odbacuje.

Primjer kako se računa p-vrijednost

Računanje p-vrijednosti ručno moguće je samo kod z-testa. Kod drugih testova možemo doći jedino do intervala u kojima se nalazi p-vrijednost. Zato je njena upotreba bila dugo ograničena. Uzećemo jedan primjer (4, str. 273).

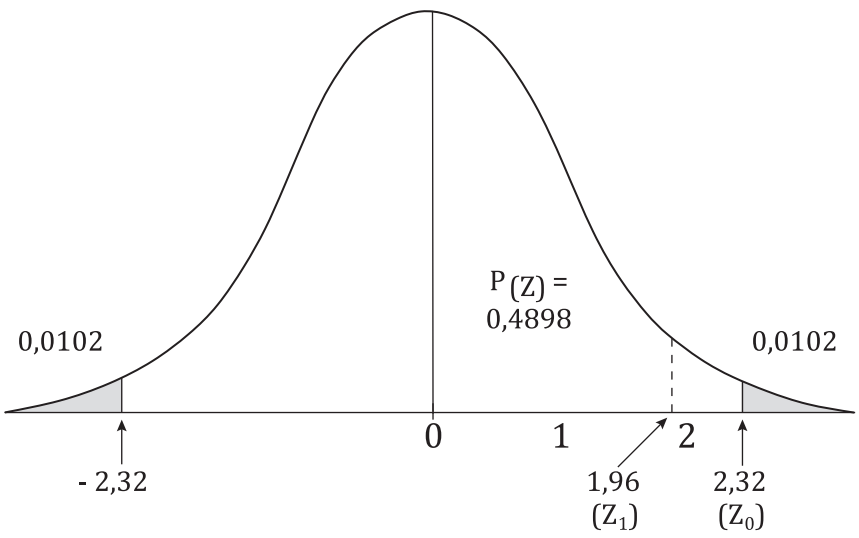
Prilikom preuzimanja pošiljke od 4.000 kutija keksa, deklarisane težine 1 kg, sa dozvoljenim odstupanjem od 20 g, slučajnim načinom je odabrano 60 kutija i ustanovljeno da njihova prosječna težina iznosi 0,994 kg. Treba provjeriti da li se proizvođač pridržava date obaveze, odnosno da li prosječna težina pošiljke (kutije) stvarno iznosi 1 kg. Imamo sljedeće podatke:

$$N = 4.000 \text{ kutija}, \; n = 60 \text{ kutija};$$

$$\overline{X} = 1{,}000 \text{ kg}, \; \overline{x} = 0{,}994 \text{ kg}, \; S = 0{,}020 \text{ kg}$$

Primjer ćemo uraditi na dva načina: a) klasičnim načinom (z-testom) i b) pomoću p-vrijednosti, uz zajednički grafički prikaz *(Slika 49)*. Postavljamo nultu hipotezu, koja važi za oba načina testiranja:

$$H_0 : \overline{x} = \overline{X}$$



Slika 49. Računanje p-vrijednosti (α = 0,05)

a) Klasični način testiranja

Za klasično testiranje biramo z-test, jer su ispunjeni uslovi za njegovu primjenu (veliki uzorak, n > 30 i poznata standardna devijacija skupa - S).

$$Z_0 = \frac{\left|\overline{x} - \overline{X}\right|}{S_{\overline{x}}} = \frac{\left|0,994 - 1,000\right|}{0,0026} = 2,32$$

$$S_{\overline{x}} = \frac{S}{\sqrt{n}} = \frac{0,020}{\sqrt{60}} = 0,0026$$

$$Z_1 = 1,96 \rightarrow Z_0 > Z_1$$

Zaključak: Izračunata vrijednost testa veća je od tablične, tj. odbacujemo H_0. Statistički značajna razlika upućuje na to da proizvođač ne poštuje deklarisanu težinu pakovanja keksa. Na stručnoj analizi i poslovnoj politici je da utvrde da li postoji i praktično značajna razlika.

b) Testiranje pomoću p-vrijednosti

Izračunatu vrijednost testa, z = 2,32 prevešćemo u p-vrijednost. Rekli smo da p-vrijednost odgovara površini (vjerovatnoći), koja kod z-testa ostaje na krajevima (izvan intervala izračunate vrijednosti testa). Dakle, trebamo samo od ukupne površine, koja uvijek iznosi jedan, oduzeti dvostruku površinu, za izračunato z = 2,32. Nju čitamo iz tablice površina normalnog rasporeda.

$$Z = 2,32 \rightarrow P_{(Z)} = 0,4898$$

$$P = 1 - \left(2 \cdot 0,4898\right) = 0,0204$$

$$P = 0,0204$$

$$P < 0,05 \rightarrow H \quad \text{se odbacuje}$$

Dobili smo p-vrijednost p = 0,0204 (grafički su to dvije zatamnjene površine, 2 x 0,0102). Vjerovatnoća da se može desiti ovolika razlika u težini pakovanja manja je od 5%, što nas upućuje na isti zaključak: razlika je statistički značajna.

Napomena: U statističkom zaključivanju na bazi uzorka, bez obzira da li se radi o procjeni ili testiranju, ne možemo izbjeći mogućnost nastajanja greške, drugim riječima, u naše rezultate nikada nismo 100% sigurni.

Prof. dr Miloš Koprivica
(1948 – 2021)

Autor knjige Šumarska statistika (2015)

4. REGRESIONA ANALIZA

4.1. Uvod u regresionu analizu

U prethodnim poglavljima razmatrali smo jednodimenzionalne skupove. Međutim, prirodne pojave su obično međusobno povezane, pa se ne mogu izolovano posmatrati. Na primjer, rast šume zavisi od dubine zemljišta, klimatskih prilika, načina sječe šume itd. Statistika zajedno sa strukom otkriva te faktore i utvrđuje njihove uticaje. Metod koji tome služi naziva se *regresiona analiza*. To je jedan od najvažnijih statističkih metoda u oblasti šumarstva.

Dakle, u ovom poglavlju prelazimo na dvodimenzionalne skupove. To mogu biti skupovi sa dva obilježja na istim elementima (npr. prečnik i visina stabala) ili združeni parovi podataka (prinos i vlaga) – *Slika 50*. Kasnije ćemo vidjeti slučajeve sa više od dvije promjenljive, tj. višedimenzionalne skupove.

Slika 50. Prelaz sa jednodimenzionalnih na dvodimenzionalne skupove

Kod jednodimenzionalnih skupova, u deskriptivnoj statistici, posmatrali smo samo jedno obilježje (X). To obilježje nalazilo se na x-osi, dok su na ordinati bile frekvencije. Kad smo prešli na uzorke oslanjali smo se na teorijske rasporede frekvencija, u kojima je naše obilježje (X) posmatrano kao slučajna promjenljiva. Sada, kod regresije obilježje (X) predstavlja nezavisnu promjenljivu, a na ordinati

se pojavljuje drugo obilježje, kao zavisna promjenljiva (*Y*). Na početku ćemo se upoznati s osnovnim pojmovima: matematička i statistička veza, korelacija i regresija, višestruka i neto regresija.

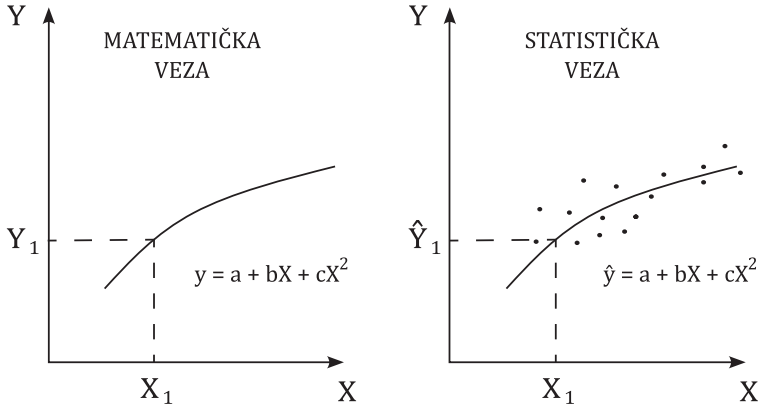
4.1.1. Matematička i statistička veza

Veza između pojava može biti:

1. Matematička (funkcionalna) i
2. Statistička (stohastička)

Matematička veza

Matematička (funkcionalna) veza je takva veza gdje se jedna pojava (zavisna promjenljiva) u potpunosti može odrediti na osnovu druge pojave (nezavisne promjenljive). To znači da uvijek određenoj vrijednosti *X* odgovara tačno određena vrijednost *Y (Slika 51)*.



Slika 51. Matematička i statistička veza

Statistička veza

Statistička (stohastička) veza je takva veza u kojoj nezavisna promjenljiva (*X*) ne određuje u potpunosti zavisnu promjenljivu (*Y*). Za neko *X* ne dobija se tačno određeno *Y*, već približno – procijenjeno, najvjerovatnije, srednje, izravnato. To je zato što na *Y* utiču i drugi neobuhvaćeni faktori, prisustvo grešaka mjerenja, kao i aproksimacija (nesavršenost) funkcije. Zato se ova veza naziva i stohastička (*stohastika* grč. = ono što je vjerovatno). Osim toga, u literaturi se još naziva korelaciona veza. Za matematičku vezu dovoljna je funkcija, dok statističku vezu, pored funkcije, određuju još standardna greška i jačina veze. Između ove dvije veze razlika je u rezultatu.

4.1.2. Korelacija i regresija

Nas prvenstveno zanimaju statističke veze. One mogu biti:

1. Čiste statističke veze (korelacija) i
2. Uzročne statističke veze (regresija).

Korelacija

Sve one veze između pojava u kojima se ne može utvrditi šta je uzrok, a šta posljedica nazivamo korelacijom. Pogledajmo nekoliko primjera.

Debljina kore stabala na osojnoj (sjevernoj) i prisojnoj (južnoj) strani se prate, ali ne možemo reći da debljina kore na jednoj strani zavisi od debljine kore na drugoj strani ili obrnuto. Mjerenjem prirasta na dva različita načina dobijamo podudarne (saglasne) rezultate. Dužina ruku i dužina nogu kod ljudi se prati, oni koji imaju duže ruke po pravilu imaju i duže noge. Međutim, ne možemo reći da dužina ruku zavisi od dužine nogu ili obrnuto.

U navedenim primjerima vidimo da između pojava postoji međuzavisnost (lat. "*correlatio"* = međuzavisnost). Korelacijom se utvrđuje samo jačina veze (stepen slaganja) između pojava. Njome se ne objašnjavaju pojave. Uzroke veza treba tražiti u drugim naukama (fiziologija i slično). Korelacija se izražava koeficijentom korelacije (r), koji može imati vrijednosti između –1 i +1.

Nonsens korelacija. Treba imati u vidu da između pojava u prirodi često postoji visok stepen slaganja (korelacije), iako one ne zavise jedna od druge. Radi se o nonsens ili nelogičnoj korelaciji. Na primjer topljenje asvalta i rojenje pčela su pojave povezane sa visokom temperaturom, a ne međusobno. U područjima gdje su intervenisali vatrogasci obično su veće štete od požara, nego tamo gdje nisu. Zaključak da vatrogasci prouzrokuju veću štetu bio bi besmislen. U selima gdje ima više roda rađa se više djece. Zaključak da rode donose djecu takođe nema smisla.

Specifičan primjer korelacije iz šumarstva. U šumarskoj praksi često se koristi veza između visine (h) i prečnika (d) stabala. Radi se o klasičnoj korelaciji, jer ne možemo reći šta od čega zavisi, visina od prečnika ili obrnuto. Međutim, mi vezu ova dva obilježja posmatramo kao regresiju. Za zavisnu uzimamo visinu, a za nezavisnu prečnik. To činimo iz praktičnih razloga, jer prečnik lakše i sigurnije mjerimo. Osim toga, prečnik služi kao zamjena za starost.

Regresija

Sve one veze u kojima se može utvrditi šta je uzrok, a šta posljedica nazivaju se regresione veze (lat. *regressio* = uzvraćanje, nazadovanje). Na primjer, rast biljaka i vlaga zemljišta. Znamo (logično je) da rast biljaka zavisi od vlage, a ne obrnuto. Kakvog je oblika ta zavisnost određuje se funkcijom (regresionom jednačinom), a koliko je izabrana jednačina „dobra" (kvalitetna) pokazuje njena standardna greška (s_t) i koeficijent determinacije (r^2).

Regresiona analiza

Regresiona analiza je metod za procjenu i predviđanje neke pojave na osnovu jedne ili više drugih pojava.

Prvi put pojam *regresija* upotrijebio je engleski naučnik Galton 1885. godine, prilikom ispitivanja nasljednih osobina. Otkrio je da visina sinova prema visini očeva pokazuje nazadovanje (regresiju). Početkom XX vijeka Pirson je počeo primjenjivati statističke metode zasnovane na Galtonovoj teoriji i od tada se pojam regresije koristi za sve metode koje ispituju zavisnost između pojava (bez obzira da li se radi o progresiji ili regresiji).

4.1.3. Višestruka i neto regresija

Regresiona analiza (regresija) služi za utvrđivanje zavisnosti jedne pojave (Y) od neke druge pojave (X) ili više drugih pojava (X_1, X_2, X_3, ...). Pojavu Y nazivamo *zavisna promjenljiva*, a X_1, X_2, X_3 ... *nezavisne promjenljive* (faktori). Ako se uzima u obzir samo jedna nezavisna promjenljiva (X) radi se o *jednostavnoj regresiji*, a ako se u jednačinu regresije uključe dvije ili više nezavisnih promjenljivih (X_1, X_2, X_3, ...) govorimo o *višestrukoj regresiji*.

U višestrukoj regresiji često se posmatra uticaj samo jedne (nezavisne) promjenljive na zavisnu promjenljivu (neto uticaj). Pri tome se ostale nezavisne promjenljive isključuju (izoluju), tako što se uzimaju kao konstantne (najčešće kao prosječne vrijednosti iz uzorka). Taj postupak se naziva *neto regresija*.

4.2. Podjela statističkih veza

Statističke veze dijelimo po više osnova: po matematičkom obliku, po smjeru veze i po broju nezavisnih promjenljivih. Postoje i specijalni oblici veza.

Po matematičkom obliku:

- linearne (pravolinijske)
- krivolinijske (sve ostale, najčešće parabola drugog reda).

Po smjeru veze:

- rastuće (s povećanjem nezavisne promjenljive raste zavisna promjenljiva)
- opadajuće (s povećanjem nezavisne promjenljive opada zavisna promjenljiva).

Po broju nezavisnih promjenljivih:

- jednostavne (Y zavisi od jedne promjenljive - X, pri čemu su ostale zanemarene)
- višestruke (Y zavisi od najmanje dvije promjenljive – X_1, X_2, ..., pri čemu su ostale zanemarene).

Specijalni oblici statističkih veza:

- kvalitativna korelacija (odnosi se na atributivna obilježja)
- korelacija ranga (veze između rangiranih, a ne mjerenih vrijednosti)
- trend (vremenske serije u kojima se posmatra pojava tokom vremena).

4.3. Dijagram rasturanja

Na osnovu stručno-logičkog zaključivanja (kvalitativne analize) i dijagrama rasturanja (kvantitativne analize) utvrđujemo:

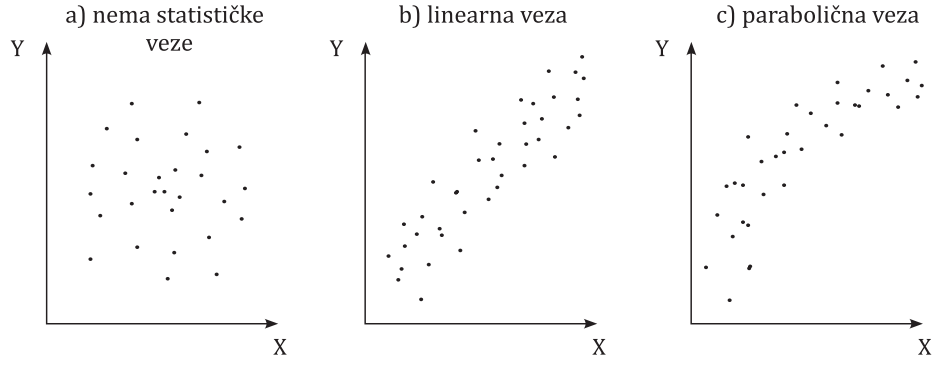
- da li postoji veza i
- koji se matematički oblik funkcije može prilagoditi toj vezi.

Ako posmatramo istovremeno dva obilježja, bilo na istim jedinicama (prečnik i visina stabla na primjer) ili uparena obilježja (visina biljaka i vlaga u zemljištu) imaćemo nove jedinice skupa (uzorka) u vidu parova podataka. Nanošenjem tih parova podataka *XY* u koordinatni sistem dobijamo grafički prikaz koji se naziva *dijagram rasturanja*.

Dijagram rasturanja

Dijagram rasturanja je prikaz svih parova podataka u koordinatnom sistemu.

Ako su tačke nepravilno razbacane ili u obliku kruga, tj. tako da se kroz njih ne može povući (interpolisati) neka funkcija, zaključujemo da ne postoji statistička veza. Ako u rasturanju tačaka pak uviđamo neku pravilnost zaključujemo da između promjenljivih postoji veza *(Slika 52)*. Ona može biti u obliku prave ili krive linije. U šumarstvu se uglavnom koristi parabola drugog reda. Radi jednostavnosti i lakšeg razumijevanja statističkih veza u nastavku ćemo govoriti samo o linearnim vezama.



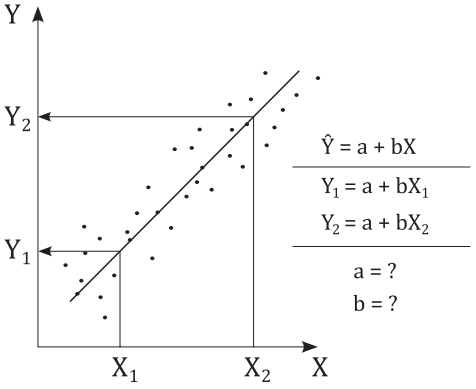
Slika 52. Dijagram rasturanja

4.4. Određivanje parametara regresione jednačine

Nakon što smo utvrdili da postoji veza i odabrali njen oblik sljedeći korak u regresionoj analizi je određivanje parametara odabrane jednačine. U principu postoje dva načina. To su grafičko-računski način i analitički (računski) način određivanja parametara, poznat kao metod najmanjih kvadrata.

4.4.1. Grafičko-računski način

Grafičko-računski metod uglavnom je korišten dok nisu postojali računari. Metod je jednostavan i razumljiv, ali subjektivan. Naime, nakon što nacrtamo dijagram rasturanja liniju izravnanja povlačimo proizvoljno („od oka"). Pri tome nastojimo da odstupanja iznad i ispod linije budu što manja i približno jednaka *(Slika 53)*.



Slika 53. Grafičko-računski način određivanja parametara

Nakon toga slijedi jednostavan postupak određivanja parametara. Za pravu liniju uzimamo dvije proizvoljne vrijednosti X_1 i X_2 (obično malo razmaknute) i za njih očitamo vrijednosti Y_1 i Y_2. Kad uvrstimo sve te vrijednosti u dvije jednačine sa dvije nepoznate, parametre *a* i *b* dobićemo rješenjem jednačina.

4.4.2. Metod najmanjih kvadrata (MNK)

Jednačina (linija) regresije mora da ispuni dva uslova:

$$\sum\left(Y_i - \widehat{Y}_i\right) = 0$$

$$\sum\left(Y_i - \widehat{Y}_i\right)^2 = \text{MIN}$$

Ovo su poznata svojstva aritmetičke sredine, kod koje smo računali odstupanja pojedinačnih vrijednosti obilježja od jedne konstante. Ovdje se radi o odstupanjima stvarnih vrijednosti (Y_i) od više sredina povezanih linijom

($\hat{Y}_i$). Po drugom uslovu, da suma kvadrata odstupanja mora biti minimalna, ovaj način računanja parametara je dobio naziv *metod najmanjih kvadrata*. Tako dobijena linija ima najmanju standardnu devijaciju, odnosno najmanju standardnu grešku.

Metod najmanjih kvadrata

Metod najmanjih kvadrata daje parametre funkcije koja ima najmanju standardnu grešku.

U linearnoj regresiji *Y=f(X)*, odnosno *Y=a+bX* govori se o zavisnosti *Y* od *X*. Parametar *"a"* pokazuje koliko iznosi zavisna promjenljiva (*Y*) kad je nezavisna promjenljiva (*X*) jednaka nuli. Geometrijski to je tačka u kojoj linija regresije presijeca *Y* osu. Parametar *"b"* pokazuje koliko se mijenja zavisna promjenljiva (*Y*) kad se nezavisna promjenljiva (*X*) poveća za jedinicu. Geometrijski „*b*" predstavlja nagib prave. Ako je ovaj parametar pozitivan radi se o rastućoj, a ako je negativan o opadajućoj regresiji.

Primjer

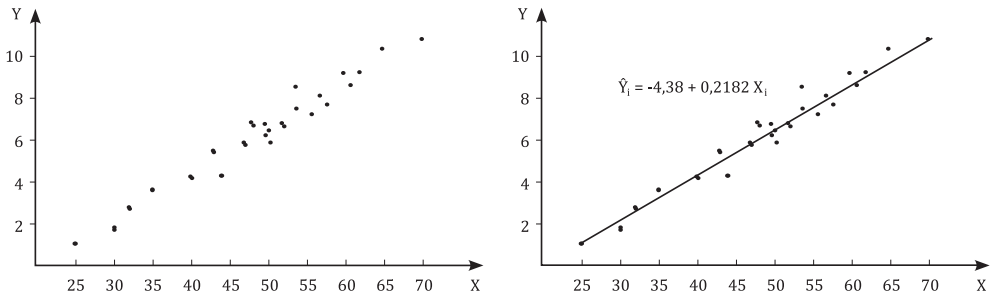
U cilju ispitivanja zavisnosti vremena potrebnog za kresanje grana (*Y*) od prečnika stabla (*X*) na terenu su prikupljeni podaci za 22 stabla smrče *(Tabela 8)*. Primjer je iz naše knjige (3, str. 126). Isti primjer koristićemo u nastavku za računanje standardne greške i koeficijenta determinacije.

Tabela 8. Regresija – radna tabela za računanje parametara (MNK)

$\mathbf{X_i}$ (cm)	$\mathbf{Y_i}$ (min)	$\mathbf{X_i^2}$	$\mathbf{X_iY_i}$	$\mathbf{Y_i^2}$	$\widehat{\mathbf{Y}}_\mathbf{i}$	$\left(\mathbf{Y_i}-\widehat{\mathbf{Y}}_\mathbf{i}\right)$	$\left(\mathbf{Y_i}-\widehat{\mathbf{Y}}_\mathbf{i}\right)^2$
25	1,13	625	28,25	1,2769	1,08	0,05	0,0025
30	1,80	900	54,00	3,2400	2,17	-0,37	0,1369
32	2,71	1024	86,72	7,3441	2,60	0,11	0,0121
40	4,26	1600	170,40	18,1476	4,35	-0,09	0,0081
44	4,35	1936	191,40	18,9225	5,22	-0,87	0,7569
47	5,85	2209	274,95	34,2225	5,88	-0,03	0,0009
50	6,19	2500	309,50	38,3161	6,53	-0,34	0,1156
50	6,68	2500	334,00	44,6224	6,53	0,15	0,0225
52	6,73	2704	349,96	45,2929	6,97	-0,24	0,0576
54	8,48	2916	457,92	71,9104	7,41	1,07	1,1449
56	7,17	3136	401,52	51,4089	7,84	-0,67	0,4489
58	7,56	3364	438,48	57,1536	8,28	-0,72	0,5184
60	9,11	3600	546,60	82,9921	8,71	0,40	0,1600
62	9,17	3844	568,54	84,0889	9,15	0,02	0,0004
65	10,30	4225	669,50	106,0900	9,81	0,49	0,2401
61	8,60	3721	524,60	73,9600	8,93	-0,33	0,1089
54	7,45	2916	402,30	55,5025	7,40	0,05	0,0025
35	3,65	1225	127,75	13,3225	3,26	0,39	0,1521
48	6,77	2304	324,96	45,9329	6,10	0,67	0,4489
57	8,00	3249	456,00	64,000	8,06	-0,06	0,0036
43	5,42	1849	233,06	29,3764	5,00	0,42	0,1764
70	10,80	4900	756,00	116,6400	10,90	-0,10	0,0100
1093	142,18	57247	7706,41	1063,6632	142,18	0,00	4,5282

Na osnovu dijagrama rasturanja *(Slika 54, lijevo)* zaključili smo da postoji linearna statistička veza, između vremena kresanja grana (zavisne promjenljive) i prečnika stabala (nezavisne promjenljive). Opšti oblik te veze je:

$$\widehat{Y}_i = a + bX_i$$



Slika 54. Dijagram rasturanja (lijevo) i regresiona jednačina (desno)

Uslov metoda najmanjih kvadrata glasi:

$$\sum\left(Y_i - \widehat{Y}_i\right)^2 = \text{MIN}$$

$$\sum\left(Y_i - a - bX_i\right)^2 = \text{MIN}$$

Ovaj izraz će biti u minimumu kad prvi parcijalni izvod po parametru *a* i prvi parcijalni izvod po parametru *b* izjednačimo s nulom:

$$2\sum\left(Y_i - a - bX_i\right)(-1) = 0$$

$$2\sum\left(Y_i - a - bX_i\right)(-X_i) = 0$$

Sređivanjem ovih izraza dobijamo dvije *normalne jednačine*, gdje je *n* broj parova:

$$na + b\sum X_i = \sum Y_i$$

$$a\sum X_i + b\sum X_i^2 = \sum X_i Y_i$$

Vrijednosti parametara izračunaćemo uvrštavanjem dobijenih suma iz radne tabele:

$$22a + 1,093b = 142,18$$

$$1,093a + 52,247b = 7.706,41$$

Rješenjem ove dvije jednačine sa dvije nepoznate dolazimo do vrijednosti parametara:

$$a = -4,379535;\ b = 0,218234$$

Napišimo jednačinu regresije, s tim da ćemo parametre zaokružiti na logičan broj decimala:

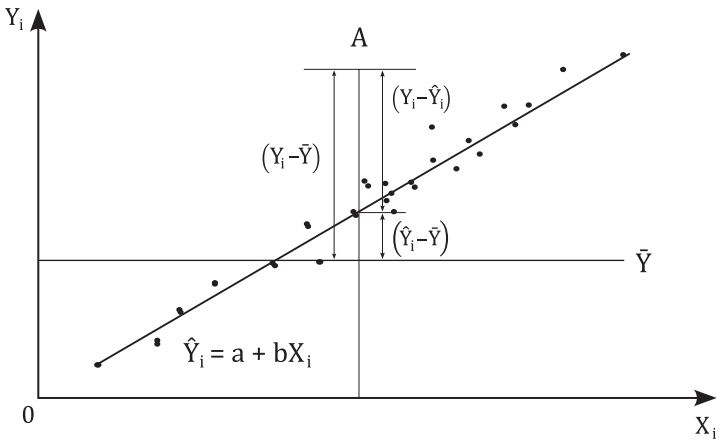
$$\widehat{Y}_i = -4,38 + 0,2182X_i$$

Sada možemo ucrtati liniju regresije na naš dijagram rasturanja *(Slika 54, desno)*. Dovoljno je da izračunamo procijenjenje vrijednosti zavisne promjenljive za dvije vrijednosti nezavisne promjenljive i te tačke spojimo pravom linijom. Time smo napravili novi korak u postupku regresione analize. Slijedi nam utvrđivanje kvaliteta procjene, odnosno mjera reprezentativnosti.

4.5. Standardna greška regresije (s_t)

Nakon izračunavanja parametara regresione jednačine potrebno je još utvrditi *standardnu grešku regresije* (s_t) i koeficijent determinacije (r^2). Ove dvije mjere kvaliteta regresije koriste se zajedno.

Standardna greška regresije je apsolutna mjera variranja stvarnih podataka oko linije procjene. To je zapravo standardna devijacija, definisana kao prosječno odstupanje stvarnih (mjerenih, originalnih) vrijednosti Y_i od procjenjenih vrijednosti $\hat{Y}_i$. Računa se kao kvadratna sredina. Svako pojedinačno odstupanje od linije je u stvari pogrešna vrijednost, pa se ova standardna devijacija naziva standardna greška *(Slika 55)*.



Slika 55. Standardna greška

Aritmetička sredina, kao što od ranije znamo, je jedna vrijednost (konstanta). Aritmetička sredina naše zavisne promjenljive, pojave koju ispitujemo, ako je prikažemo grafički biće horizontalna linija ($\overline{Y}$). Udaljenost svake tačke od linije $\overline{Y}$ predstavlja njeno ukupno odstupanje (ukupni varijabilitet). To važi i za tačku A, čije odstupanje se sastoji iz dva dijela: odstupanja procijenjene vrijednosti ($\widehat{Y}_i$) od aritmetičke sredine ($\overline{Y}$) i odstupanja stvarne vrijednosti (Y_i) od procijenjene vrijednosti ($\widehat{Y}_i$). Prvi dio smatramo objašnjenim, a drugi neobjašnjenim varijabilitetom. To se može napisati ovako:

$$\left(Y_i - \overline{Y}\right) = \left(\widehat{Y}_i - \overline{Y}\right) + \left(Y_i - \widehat{Y}_i\right)$$

Da bismo dobili mjeru preciznosti moramo uzeti u obzir sva odstupanja. Pošto je zbir linearnih odstupanja jednak nuli prvo ćemo ih kvadrirati, a potom sabrati. Može se lako dokazati da je ukupna suma kvadrata jednaka zbiru objašnjene i neobjašnjene sume kvadrata. Ako ove sume podijelimo s ukupnim brojem parova u skupu (*N*) dobijamo tri varijanse, za koje takođe važi aditivno svojstvo:

$$\frac{\sum\left(Y_i - \overline{Y}\right)^2}{N} = \frac{\sum\left(\widehat{Y}_i - \overline{Y}\right)^2}{N} + \frac{\sum\left(Y_i - \widehat{Y}_i\right)^2}{N}$$

$$s_y^2 = s_{\hat{y}}^2 + s_t^2$$

s_y^2 – varijansa ukupnog odstupanja,

$s_{\hat{y}}^2$ – varijansa objašnjenog odstupanja i

s_t^2 – varijansa neobjašnjenog odstupanja

Varijansa neobjašnjenog odstupanja (s_t^2) naziva se *varijansa regresije*, a njen korijen *standardna greška* regresije (s_t). Standardna greška predstavlja prosječno odstupanje stvarnih vrijednosti od linije regresije. Ako je ona manja znači da je procjena preciznija.

Standardna greška

Standardna greška je apsolutna mjera kvaliteta regresione jednačine. Dobija se kao korijen iz varijanse neobjašnjenog odstupanja.

Standardnu grešku možemo izračunati po definiciji ili po radnoj formuli. Iskoristićemo naš prethodni primjer *(Tabela 8)*. *Napomena*: Kod računanja varijansi u skupu se sume dijele sa N, a u uzorku s brojem stepena slobode (u jednostavnoj regresiji sa *n-2*).

a) Po definiciji

$$s_t^2 = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}$$

$$s_t^2 = \frac{4,5282}{20} = 0,22641$$

$$s_t = 0,46 \text{ min}$$

b) Po radnoj formuli

$$s_t^2 = \frac{\sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i}{n-2}$$

$$s_t^2 = \frac{1063,6632 + 4,379535 \cdot 142,18 - 0,218234 \cdot 7706,41}{20} = \frac{4,5282}{20} = 0,22641$$

$$s_t = 0,46 \text{ min}$$

Zaključujemo da utrošeno vrijeme za kresanje grana u prosjeku odstupa od linije procjene 0,46 minuta.

4.6. Koeficijent determinacije (r^2)

Mjera statističke (stohastičke) veze treba da bude neki broj, koji u nekim granicama pokazuje stepen veze – od odsustva do potpune veze. Postoje dvije takve mjere. To su koeficijent korelacije (*r*) i koeficijent determinacije (r^2). U regresiji govorimo o koeficijentu determinacije.

Prethodno smo vidjeli da se ukupni varijabilitet s_y^2 sastoji iz dva dijela: objašnjenog $s_{\hat{y}}^2$ i neobjašnjenog s_t^2 varijabiliteta. Iskazali smo to preko varijansi. Ako u odnos stavimo objašnjeni prema ukupnom varijabilitetu dobićemo relativni broj koji se naziva *koeficijent determinacije* (r^2). Ovaj koeficijent determiniše (pokazuje) u kojoj mjeri (procentu) je *Y* određena sa *X*. Vrijednost ovog koeficijenta može se naći u intervalu od nula do jedan, odnosno 0 – 100% *(Slika 56)*. Formula glasi:

$$r^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

Ako od ukupnog varijabiliteta, tj. od jedinice, oduzmemo neobjašnjeni dio varijabiliteta dobićemo drugu formulu za računanje koeficijenta determinacije. Nju ćemo iskoristiti da izračunamo koeficijent determinacije u našem primjeru. Varijansu greške imamo izračunatu (ona iznosi 0,22641), a varijansu ukupnog variranja uzećemo iz knjige, str. 158 (ona iznosi 6,58183).

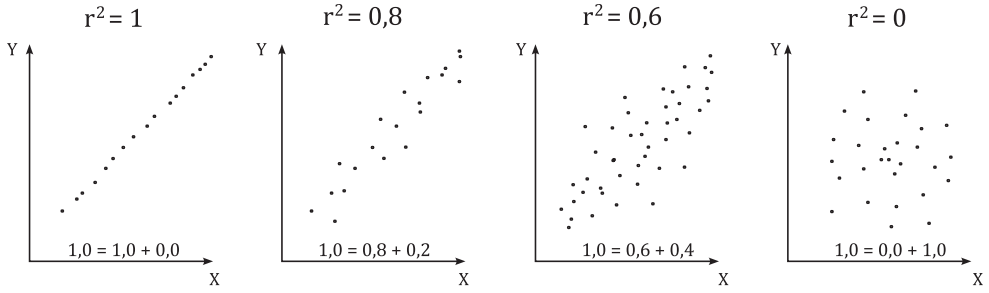
$$r^2 = 1 - \frac{s_t^2}{s_y^2} = 1 - \frac{0,22641}{6,58183} = 1 - 0,0344 = 0,9656$$

Zaključak: Zavisna promjenljiva (vrijeme za kresanje grana) može se objasniti sa 96,56% pomoću nezavisne promjenljive (prečnikom stabala), dok se ostatak od 3,44% pripisuje drugim neobuhvaćenim faktorima. Zavisnost je ovdje veoma visoka (radi se o hipotetičkim podacima). U praksi se smatra visokom stohastičkom vezom ako je koeficijent determinacije iznad 50%.

Napomena: Koeficijent determinacije može se definisati i na druge načine. Takođe, postoje i radne formule za njegovo ručno računanje (o tome u knjizi, 3).

Koeficijent determinacije

Koeficijent determinacije je relativna mjera kvaliteta regresije, koja pokazuje stepen određenosti (objašnjenosti) zavisne promjenljive (ispitivane pojave).



Slika 56. Koeficijent determinacije

Koeficijent determinacije smatra se najboljom mjerom kvaliteta regresione jednačine, jer:

- je razumljiv za tumačenje,
- predstavlja relativnu mjeru, što omogućava različita poređenja,
- na osnovu njega lako se izračunava koeficijent korelacije.

Koeficijent korelacije (r) pokazuje stepen linearne međuzavisnosti X i Y, u odnosu na funkcionalnu vezu. Može iznositi od -1 do 1 (od potpune negativne r = -1, preko potpune nezavisnosti r = 0, do potpune pozitivne međuzavisnosti r = 1). Koeficijent korelacije obuhvata sve faktore koji utiču na promjenljive *X* i *Y*, dok koeficijent determinacije predstavlja zajednički varijabilitet *X* i *Y*. Kod r = 0,20 koeficijent determinacije r^2 iznosi svega 0,04, odnos 4% . Koeficijent korelacije računa se samo u slučajevima kada želimo utvrditi koliko se pojave slažu, tj. kolika je korelacija (međuzavisnost) između njih. Tada nam nije važno šta je Y, a šta X. Na primjer, postignuti rezultati studenata na različitim predmetima.

Kako se tumači veličina koeficijenta korelacije? Tumači se u zavisnosti od situacije. Na primjer, ako bismo dobili r = 0,7 kod korelacije između težine studenata danas i prije 10 dana to je jako mali koeficijent. Očekivali bismo gotovo potpunu saglasnost (koeficijent vrlo blizu jedan), jer je u pitanju težina kod istih studenata.

4.7. Višestruka regresiona analiza (*VRA*)

Do sada je bilo govora o jednostavnoj regresiji. Međutim, rijetko možemo objasniti neku pojavu na osnovu samo jednog faktora. Na primjer, na debljinski prirast stabala utiče više faktora. To su: prečnik ili starost stabla, kvalitet (bonitet) staništa, stepen sklopa (gustina šume), srednji prečnik sastojine (debljinska struktura) i drugi. Ako u regresionu analizu uzmemo dva ili više faktora govorimo o *višestrukoj regresionoj analizi*. Pri tome se postavlja više pitanja: koliko faktora uzeti, koje faktore odabrati i kakvog je oblika njihov pojedinačni uticaj. To se rješava od slučaja do slučaja, na osnovu stručnog znanja i statističkih pokazatelja. Neko pravilo je da broj faktora, koje u regresiji nazivamo nezavisne promjenljive (varijable), bude između tri i pet *(Slika 57)*.

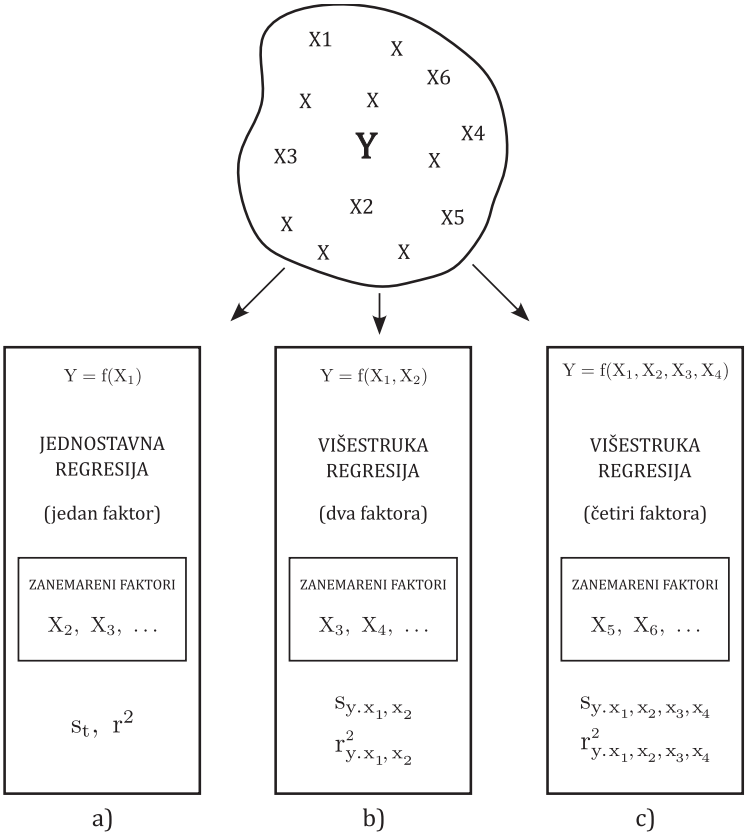
Obično se u početnoj fazi istraživanja metodom regresione analize koristi *korelaciona matrica*. Softveri omogućavaju tabelarni prikaz koeficijenta korelacije (r) između velikog broja potencijalnih varijabli (promjenljivih). Tako saznajemo koje pojave su povezane i koliko je ta veza jaka. Treba napomenuti da postoje dva tipa regresionih jednačina. Ako između nezavisnih promjenljivih ne postoji ili je mala korelacija, u regresionoj jednačini njihovi uticaji se sabiraju. To je tzv. *aditivni* model. Međutim, ako su nezavisne promjenljive međusobno zavisne onda one u regresionoj jednačini jedna drugu zamjenjuju ili nadopunjuju. Tada se radi o tzv. *supstitutivnom modelu*.

Višestruka regresiona analiza

Ako neku pojavu statistički objašnjavamo sa dvije ili više drugih pojava taj metod nazivamo višestruka regresiona analiza.

Postupak kod *VRA*, kao i kod jednostavne regresije, sastoji se od određivanja parametara jednačine, standardne greške regresije (s_t) i koeficijenta determinacije (r^2). Glavni kriterijum kod ocjene kvaliteta regresione jednačine je koeficijent determinacije. On pokazuje koliko zavisnu promjenljivu (*Y*)

određuju (objašnjavaju) nezavisne promjenljive (X_1, X_2, X_3 itd). Metodom višestruke regresije ocjenjujemo (procjenjujemo) i predviđamo zavisnu pojavu u različitim uslovima.



Slika 57. Višestruka regresiona analiza

Metod višestruke regresione analize korišten je, kao glavni metod, u istraživanjima prirasta šuma u BiH. Rezultat toga su Tablice taksacionih elemenata šuma, koje su doživjele tri izdanja (1963, 1980. i 1990. godine). Ovdje ćemo navesti kao primjer regresionu funkciju za površinu projekcije krošnje stabala jele prečnika 12,5 cm (*Y*). Rezultat se dobija u m^2.

$$Y = 14{,}6 - 0{,}257X_1^2 + 1{,}06X_1 - 6{,}00 \cdot X_2^2 + 8{,}37X_2 + 0{,}01380X_3^2 - 0{,}7914X_3$$

gdje su: X_1- bonitet staništa, X_2 – stepen sklopa, X_3 – srednji prečnik sastojine.

Dobijen je $r^2 = 0{,}63$, što znači da navedene tri nezavisne promjenljive (zajedno) objašnjavaju zavisnu promjenljivu sa 63%, dok se preostalih 37% uticaja pripisuje drugim neobuhvaćenim faktorima.

4.8. Neto regresija

Kako smo vidjeli kod jednostavne regresije uzima se u obzir samo jedan faktor, dok se svi ostali faktori zanemaruju, kao da ne postoje. Glavni cilj je da se utvrdi (procjeni) vrijednost zavisne *Y*, što znači da jednostavna regresija ima samo jednu vrijednost za rezultat.

Koristeći jednačinu višestruke regresije, kojom je iskazan *bruto uticaj* na zavisnu promjenljivu, možemo ispitati kakav je pojedinačni uticaj svakog od obuhvaćenih faktora. To se čini isključivanjem ostalih faktora, tako što se oni drže nepromijenjenim. U računu se samo mijenja onaj faktor čiji uticaj tražimo. Na ovaj način omogućava se sagledavanje "čistog" (neto) uticaja jednog faktora na zavisnu promjenljivu. To i jeste osnovni cilj neto regresije. Dakle, kod neto regresije imaćemo više rezultata.

Neto regresija

Utvrđivanje pojedinačnih uticaja nezavisnih promjenljivih (faktora) na zavisnu promjenljivu naziva se neto regresija.

Napišimo jedan hipotetički primjer opšteg oblika jednačine višestruke regresije sa tri nezavisne promjenljive i prikažimo njihove jednačine neto uticaja:

$$\hat{Y} = a + b_1X_1 + b_2X_1^2 + cX_2 + dX_3$$

Neto uticaj X_1:

$$\hat{Y} = b_1X_1 + b_2X_1^2 + k_1; \quad k_1 = a + cX_2 + dX_3$$

Neto uticaj X_2:

$$\hat{Y} = cX_2 + k_2; \quad k_2 = a + b_1X_1 + b_2X_1^2 + dX_3$$

Neto uticaj X_3:

$$\hat{Y} = dX_3 + k_3; \quad k_3 = a + b_1X_1 + b_2X_1^2 + cX_2$$

Jednačina neto regresije uzima iz višestruke regresije sumande sa faktorom čiji uticaj tražimo. Slobodni član *k* obuhvata ostali dio jednačine, kad se za faktore uzmu konstantne vrijednosti (najčešće prosječne iz uzorka).

Neto uticaj se izračunava jednostavno uvrštavanjem u neto jednačinu logično izabranih vrijednosti tog faktora, uz konstantnu vrijednost *k*. Na primjer, ako je X_1 bonitet staništa uzećemo vrijednosti boniteta od jedan do pet i uvrštavati u prvu jednačinu. Dobićemo pet rezultata, tj. vrijednosti zavisne promjenljive

na prvom, drugom, trećem, četvrtom i petom bonitetu. Naravno, uz prosječne vrijednosti X_2 i X_3. To onda možemo prikazati grafički. Za neto regresiju (neto uticaj) koriste se još izrazi djelimična i parcijalna regresija.

4.9. Specijalni oblici statističkih veza

Spomenimo trend, korelaciju ranga i kvalitativnu korelaciju, kao neke od specijalnih oblika statističkih veza.

Trend

Istraživanje dinamičkih pojava (vremenskih serija) je specifično po tome što se posmatra ponašanje pojave tokom nekog vremenskog perioda (engl. *trend* – tendencija, težnja, tok). Zavisna pojava (Y) pri tome ne zavisi od vremena, ali se u računu vrijeme formalno uzima kao nezavisna promjenljiva (X). Tokom vremena mijenjaju se drugi faktori, tj. uslovi u kojima se pojava (Y) odvija, što utiče na nju. Grafički gledano kod trenda nema zone (dijagrama) rasturanja tačaka. Svakoj vrijednosti X_i odgovara samo jedna vrijednost Y_i. Postupak (analiza) kod ovakvih zavisnosti provodi se na isti način kao kod jednostavne regresije.

Procjena zavisne unutar intervala posmatranja (raspona variranja stvarnih podataka) naziva se *interpolacija*, a izvan *ekstrapolacija*. Ekstrapolacija se obično koristi za predviđanje pojave u bliskom budućem vremenu.

U šumarstvu imamo još jednu specifičnu vrstu veze, koja je vrlo česta. Radi se o rastu stabala ili sastojina u zavisnosti od starosti. Tu postoji poznata dinamika ili tok rasta. Iako je na x-osi vrijeme, ne radi se o trendu. Nisu u pitanju kalendarske godine, već godine koje se odnose na starost. Zato kod pojedinačnih stabala ili sastojina nema dijagrama rasturanja, iako se suštinski radi o regresiji (zavisnosti).

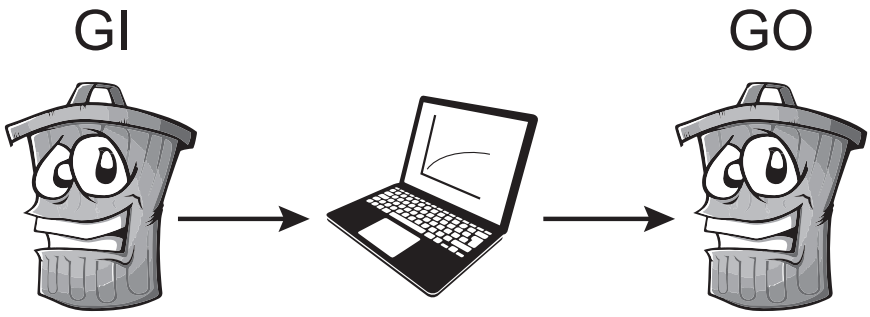
Kvalitativna korelacija

Kada su u pitanju kvalitativna (atributivna) obilježja ispitujemo postojanje veze između njih, ne postavljajući pitanje šta od čega zavisi. Radi se dakle o korelaciji (međuzavisnosti). Takvi primjeri u šumarstvu su: ispitivanje korelacije između boje lišća i boje ploda, boje lišća i veličine ploda, boje i debljine kore neke vrste drveća. Mjera jačine kvalitativne korelacije naziva se *koeficijent kontingencije*. U postupku analize umjesto dijagrama rasturanja koristi se dvoulazna tabela (primjer u knjizi, str. 175).

Korelacija ranga

Obilježja koja se ne mogu precizno izmjeriti kvantitativno iskazuju se u vidu ranga. Rang je jednostavno redoslijed po veličini. Na primjer u šumarstvu takva obilježja su: bonitet staništa (proizvodna sposobnost, kvalitet staništa),

kvalitet stabala, intenzitet prirasta stabala. Pri rangiranju boniteta najbolje stanište označava se sa *I* (prvi bonitet), a najlošije sa *V* (peti bonitet). Slično je i kod klasifikacije stabala po kvalitetu. Može da nas zanima postoji li korelacija između kvaliteta staništa i kvaliteta stabala ili u redoslijedu rasta stabala različitih provenijencija nakon određenog vremena (primjer u knjizi, str. 178).



KORIŠTENA LITERATURA

1) Hadživuković, S. (1973). Statistički metodi. Univerzitet „Radivoj Ćirpanov“. Novi Sad.
2) Hadživuković, S. (1979). Statistika. Rad. Beograd
3) Koprivica, M. (2015). Šumarska statistika. Univerzitet u Banjoj Luci. Šumarski fakultet.
4) Lovrić, M., Komić, J. i Stević, S. (2006). Statistička analiza - metodi i primjena. Ekonomski fakultet. Banja Luka.
5) Mićić, N. (2011). Eksperimentalna biometrika. Poljoprivredni fakultet Univerziteta u Banjoj Luci, Naučno voćarsko društvo Republike Srpske.
6) Obradović, S. i Sentić, M. (1967). Osnovi statističke analize. Naučna knjiga. Beograd.
7) Petz, B., Kolesarić, V. i Ivanec, D. (2012). Petzova statistika. Naklada slap. Hrvatska.
8) Popović, B. (1962). Matematsko-statističke metode u poljoprivredi i šumarstvu. Univerzitet u Sarajevu.
9) Pranjić, A. (1986). Šumarska biometrika. Sveučilište u Zagrebu. Šumarski fakultet.
10) Serdar, V. (1966). Udžbenik statistike. Školska knjiga. Zagreb.
11) Stojanović, O. (1962). Statističke metode u naučnoistraživačkom radu i u praksi šumarstva. Narodni šumar XVI, Sarajevo.
12) Stojanović, O. (1963). Savremena inventarizacija šuma zahteva i savremeniju koncepciju dendrometrije. Narodni šumar XVII, Sarajevo.
13) Stojanović, O. (1964). Primjena reprezentativnog metoda pri taksacionoj procjeni šuma. Narodni šumar XVIII, Sarajevo.
14) Stojanović, O. i Drinić, P. (1974): Istraživanje veličine koncentričnih kružnih površina za taksacionu procjenu šuma. Radovi Šumarskog fakulteta i Instituta za šumarstvo u Sarajevu. Sarajevo.
15) Stojanović, O. (1985). Kontrola terenskih radova i testiranje rezultata taksacione procjene šuma. Šumarstvo i prerada drveta. Poseban otisak. Sarajevo.