1 FORMIRANJE STATISTIČKIH NIZOVA

Statistički skup – skup elemenata kojima proučavamo jedno ili više obilježja čije se vrijednosti mijenjaju od elementa do elementa.

S obzirom na broj elemenata mogu biti: **beskonačni** i **konačni**.

Osnovni skup ili **populacija** – skup iz kojeg se može izvući podskup ili uzorak.
Stratifikacija – postupak dijeljenja osnovnog skupa na disjunktne podskupove.
Statistički skup se definira:

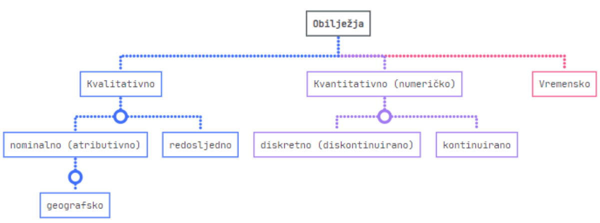
- Pojmovno
- Prostorno
- Vremenski

Studenti Ekonomskog fakulteta u Splitu upisani u akademskoj godini 23/24

Pojmovna definicija – odrediti točno koja obilježja moraju imati elementi da bi bili uvršteni u statistički skup. (BDP)
Prostorna definicija – određuje prostor na koji se odnose, tj. kojem pripadaju elementi statističkog skupa. (RH)
Vremenska definicija – odnosi se na razdoblje ili trenutak kojim će se vremenski obuhvatiti jedinice koje ulaze u statistički skup. (2023. godina)

Statistička obilježja – opće karakteristike elemenata statističkog skupa, po kojima su elementi jedni drugima slični i po kojima se međusobno razlikuju.

kvalitativna | **kvantitativna**



Nominalno obilježje – izražavaju se opisno, pa se nazivaju još i atributivnim obilježjem. (spol)

- **Prostorno (zemljopisno) obilježje** – označava prostor s kojim je element statističkog skupa u vezi.
- **Atributivno obilježje** – npr. oblik udruživanja (Trgovačko društvo)

Redoslijedno obilježje – vrsta obilježja koja fluktuira prema intenzitetu ili **rangu**. (Ocjena 1–5, stručna sprema)
Numeričko obilježje – izražava se brojevima čiji odnos ne ovisi o jedinici mjere.

- **Kontinuirano** – može poprimiti **neprebrojivo** beskonačno mnogo vrijednosti. (cijelim/decimalnim – staž)
- **Diskontinuirano** – može poprimiti prebrojivo beskonačno mnogo vrijednosti. (cijelim brojem – djeca)

Vremensko obilježje – označava trenutak ili vremenski interval s kojim je element statističkog skupa u vezi.
Grupiranje – vrši se dijeljenjem osnovnog skupa na dijunktne dijelove prema oblicima u kojima se obilježje koje se izučava javlja.
Apsolutna frekvencija – broj elemenata statističkog skupa koji pripadaju određenoj grupi.
Statistički niz – niz apsolutnih frekvencija svih grupa statističkog skupa određenog obilježja.
Kumulativni niz – dobiva se postupnim zbrajanjem apsolutnih frekvencija. (ne može se formirati kod atributivnih i geografskih nizova).
Statističke tablice moraju sadržavati: naslov, tekstualni dio, brojčani dio i izvor podataka.

2 GRAFIČKO PRIKAZIVANJE STATISTIČKIH NIZOVA

Koriste se za zorniji pregled opsega i strukture statističkih nizova, te vizualna usporedba frekvencija unutar jednog ili više statističkih nizova.

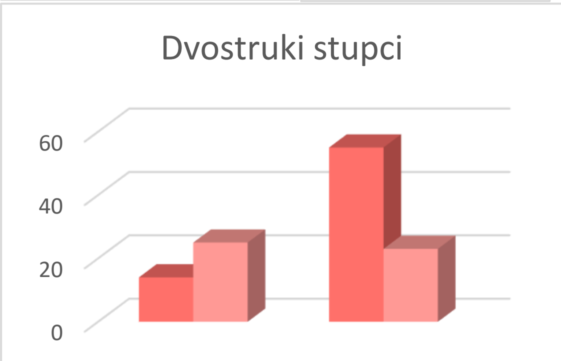
Grafički prikazi:

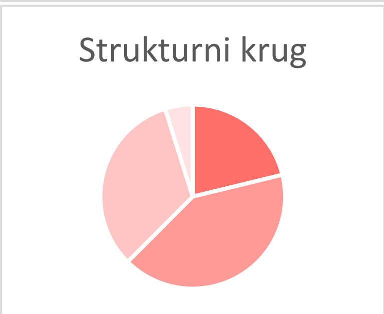
- Površinski grafikoni
- Kartogrami
- Linijski grafikoni

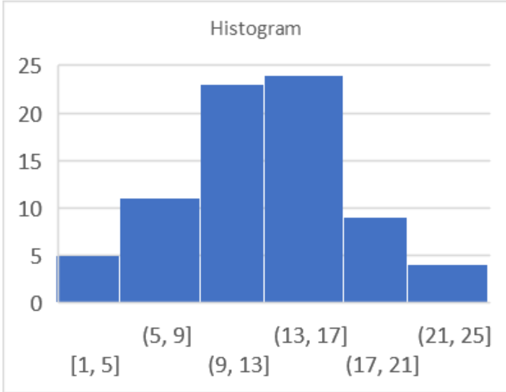
Površinski grafikoni: jednostavni stupci, dvostruki stupci, razdijeljeni stupci, proporcionalno strukturni krugovi, polukrugovi, kvadrati i histogrami.

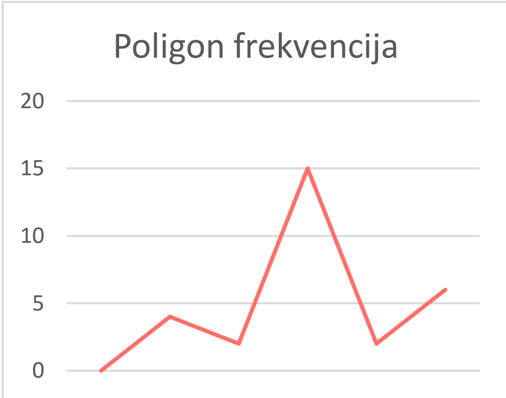
Kartogrami: dijagramske karte, piktogrami, statističke karte.

Linijski grafikon (poligon frekvencija) – prikazuje se linijama.









HISTOGRAM

Potrebno je korigirati frekvencije zbog nejednakih veličina razreda.

Korigiranje se vrši tako da se svaka frekvencija podijeli odgovarajućom veličinom razreda.

Na grafikon se unose korigirane frekvencije, a ne originalne.

POLIGON FREKVENCIJA (LINIJSKI GRAFIKON)

Frekvencije kumulativnog niza „manje od" ucrtavaju se na gornjim granicama razreda, dok se frekvencije kumulativnog niza „više od" ucrtavaju na donjim granicama razreda.

3 RELATIVNI BROJEVI I NJIHOVO GRAFIČKO PRIKAZIVANJE

Relativni brojevi su:

- Relativni brojevi strukture
- Relativni brojevi koordinacije
- Indeksi

RELATIVNI BROJEVI STRUKTURE (PROPORCIJE)

Pokazuju odnos dijela prema cjelini.
Relativne frekvencije pokazuju odnos apsolutnih frekvencija prema ukupnom broju elemenata u statističkom skupu.
Zbroj svih relativnih brojeva u statističkom nizu jednak je **1**, odnosno **100**.
Mogu se prikazati pomoću strukturnih stupaca i krugova, polukrugova, itd. Svi stupci imaju istu visinu jer je zbroj relativnih brojeva strukture jednak 100.

RELATIVNI BROJEVI KOORDINACIJE

Pokazuju odnos dviju pojava ili odgovarajućih frekvencija u različitim statističkim nizovima, koji mogu biti potpuno nezavisni jedan od drugog.

$$\boxed{R_i = \frac{A_i}{B_i}}$$

A_i – veličina koja se uspoređuje
B_i – veličina s kojom se A_i uspoređuje.

INDEKSI

Njima uspoređujemo smjer i intenzitet varijacija frekvencija nekog statističkog niza s takvim varijacijama drugog statističkog niza.
Za bazu indeksa može se odabrati, npr. jedan od članova statističkog niza, te se računaju relativni brojevi stavljajući u odnos svaki član niza prema odabranoj bazi.
Prikazuju se preko stupaca koji imaju istu bazu.

4 SREDNJE VRIJEDNOSTI

Mjere centralne tendencije

Izračunava se uvijek kao prosječna vrijednost obilježja elemenata iz kojih se izračunava.

Najčešće su u upotrebi sljedeće:

- Aritmetička sredina
- Harmonijska sredina
- Geometrijska sredina
- Medijan
- Mod

Aritmetička, harmonijska i geometrijska sredina spadaju u tzv. **izračunate srednje vrijednosti**, dok medijan i mod spadaju u tzv. **položajne srednje vrijednosti** – njihove vrijednosti se ne računaju nego se određuju položajem u statističkom skupu.

ARITMETIČKA SREDINA

Onaj jednaki dio vrijednosti numeričkog obilježja koji otpada na jedan element skupa.

Računa se kao omjer zbroja vrijednosti obilježja svih elemenata statističkog skupa i broja elemenata u statističkom skupu.

Veličina u brojniku izraza za aritmetičku sredinu – **total**.

Ako su elementi statističkog niza grupirani u frekvencije onda se upotrebljava izraz za **vaganu (ponderiranu) aritmetičku sredinu**.

3 svojstva:

a) Vrijednost aritmetičke sredine se nalazi između najveće i najmanje vrijednosti obilježja: $x_{min} \leq \bar{x} \leq x_{max}$
b) Zbroj odstupanja vrijednosti obilježja pojedinih elemenata statističkog skupa od njihove aritmetičke sredine jednak je nuli.
c) Zbroj kvadrata odstupanja vrijednosti obilježja pojedinih elemenata statističkog skupa ima minimalnu vrijednost.

HARMONIJSKA SREDINA

Predstavlja recipročnu vrijednost aritmetičke sredine recipročnih vrijednosti iz kojih se ona izračunava.

GEOMETRIJSKA SREDINA

Srednja stopa promjene vremenskih nizova, ukoliko je ta stopa približito konstantna.

Uvijek je manja od aritmetičke sredine.

MEDIJAN

Srednja vrijednost koja statistički niz dijeli na dva jednaka dijela.

Ako je broj članova paran, onda se uzima aritmetička sredina vrijednosti obilježja dvaju članova koji se nalaze u sredini statističkog niza.

Kod grupiranog statističkog niza treba najprije izračunati frekvencije kumulativnog niza „manje od" ili „više od", te pronaći srednji član.

„***50% … ima 3 zaposlena i manje, a 50% ima 3 zaposlena i više***."

Svojstvo – zbroj apsolutnih vrijednosti odstupanja pojedinih vrijednosti obilježja jedinica statističkog skupa od medijana jednak je minimumu.

Medijan je kao srednja vrijednost bolji reprezentant centralne tendencije od aritmetičke sredine u slučaju postojanja ekstremnih vrijednosti koje djeluju na aritmetičku sredinu, a ne djeluju na medijan jer on predstavlja pozicijsku srednju vrijednost.

MOD

Ona vrijednost obilježja koja se **najčešće** pojavljuje.

Može se računati za sve vrste nizova.

Ima ga smisla računati samo kod **unimodalnih distribucija**.

Kod bimodalne distribucije postoji glavni mod i lokalni mod.

U takvoj situaciji mod je manje reprezentativan kao srednja vrijednost.

Ima logičnu interpretaciju kod **diskontinuiranih numeričkih nizova**, kod **redoslijednih nizova** te **nominalnih nizova**.

Kod **kontinuiranih numeričkih nizova** mod predstavlja vrijednost u kojoj funkcija dostiže svoj maksimum.

Vrijednost moda se nalazi u razredu kojem pripada najveća korigirana frekvencija.

5 MJERE DISPERZIJE

Disperzija – raspršenost vrijednosti numeričkog obilježja.
Služe za ocjenjivanje reprezentativnosti srednje vrijednosti obilježja.

Apsolutne mjere disperzije:
Raspon varijacije obilježja
Prosječno apsolutno odstupanje
Varijanca i standardna devijacija
Interkvartil
Relativne mjere disperzije:
Koeficijent varijacije
Koeficijent kvartilne devijacije

RASPON VARIJACIJE OBILJEŽJA

Razlika između najveće i najmanje vrijednosti numeričkog obilježja.
Gruba mjera disperzije obilježja.

PROSJEČNO APSOLUTNO ODSTUPANJE (MAD)

Dobiva se kao aritmetička sredina apsolutnih vrijednosti odstupanja od aritmetičke sredine vrijednosti obilježja.

VARIJANCA (σ)

Srednje kvadratno odstupanje numeričkih vrijednosti obilježja od aritmetičke sredine.
„***… prosječno kvadratno odstupa …***"

STANDARDNA DEVIJACIJA (σ)

Pozitivni korijen iz varijance i predstavlja apsolutnu mjeru disperzije u prvom stupnju.
„***… prosječno odstupa od prosječne…***"

INTERKVARTIL (I_q)

Predstavlja mjeru raspona vrijednosti obilježja srednjih 50% jedinica u distribuciji, odnosno razliku između gornjeg i donjeg kvartila.
Kvartili zajedno s medijanom dijele distribuciju na **četiri** jednaka dijela.
Prva frekvencija u kumulativnom nizu „manje od", koja je veća od četvrtine zbroja svih frekvencija pripada razredu u kojem se nalazi **donji kvartil**.
Prva frekvencija u kumulativnom nizu „manje od", koja je veća od tri četvrtine zbroja svih frekvencija, pripada razredu u kojem se nalazi vrijednost **gornjeg kvartila**.

Donji kvartil – distribuciju dijeli 1:3, „**25% zaposlenih** ***…Imaju 4 djece i manje, a ostalih 75% imaju 3 djece i više.***" **Q_1**
Gornji kvartil – distribuciju dijeli 3:1, „***75% zaposlenih … imaju 4 djece i manje, a ostalih 25% imaju 3 djece i više.***" **Q_3**

KOEFICIJENT VARIJACIJE (V)

Relativna mjera disperzije, služi za mjerenje i uspoređivanje disperzije u različitim distribucijama.
Omjer između standardne devijacije i aritmetičke sredine. Taj omjer se može pomnožiti sa 100 i onda predstavlja postotak standardne devijacije u odnosu na aritmetičku sredinu.
„***Postotni omjer standardne devijacije i prosječne …***"

KOEFICIJENT KVARTILNE DEVIJACIJE (V_q)

Relativna mjera disperzije srednjih 50% jedinica u statističkom nizu.
Njegova vrijednost se kreće **između 0 i 1.**
„***Raspon varijacije središnjih 50% …***"
Velika disperzija podataka >0,3 (30%)

7 MJERE ASIMETRIJE I MJERA ZAOBLJENOST

MJERE ASIMETRIJE

KOEFICIJENT ASIMETRIJE α_3

Vrijednost **prvog momenta** oko aritmetičke sredine uvijek jednaka **0**.
Drugi moment oko aritmetičke sredine je uvijek pozitivan kao i svi parni momenti.
Za simetrične distribucije koeficijent asimetrije α_3 jednak je **nuli**, međutim može biti jednak nuli i za neke asimetrične distribucije.
$-2 \leq \alpha_3 \leq +2$

$\alpha_3 = 0 \rightarrow$ simetrična distribucija
$\alpha_3 > 0 \rightarrow$ pozitivno asimertična
$\alpha_3 < 0 \rightarrow$ negativno asimetrična

PEARSONOVA MJERA ASIMETRIJE S_k

Desnostrana asimetrija – Mo < $\bar{x}$
Ljevostrana asimetrija – Mo > $\bar{x}$

$S_k = 0 \rightarrow$ simetrična distribucija
$S_k > 0 \rightarrow$ pozitivno asimetrična
$S_k < 0 \rightarrow$ negativno asimetrična

Za simetrične distribucije Pearsonova mjera asimetrije jednaka je **nuli**.
Kreće se između **−3 i 3**.

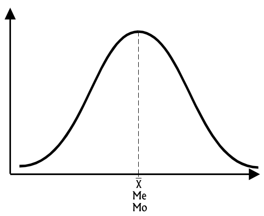
BOWLEYEVA MJERA ASIMETRIJE

Razlika između kvartila i medijana.
Za simetrične distribucije vrijednost Bowleyeve mjere asimetrije jednaka je **nuli**.
Kreće se između **−1 i 1**.

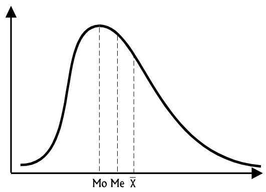
MJERA ZAOBLJENOSTI

Zaobljenost vrha – **kurtoza**
U tu svrhu se upotrebljava koeficijent zaobljenosti vrha distribucije – α_4

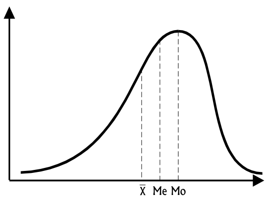
$\alpha_4 = 3 \rightarrow$ normalna distribucija
$\alpha_4 > 3 \rightarrow$ šiljastija distribucija
$\alpha_4 < 3 \rightarrow$ plosnatija distribucija
$\alpha_4 \approx 1{,}8 \rightarrow$ pravokutna distribucija
$\alpha_4 < 1{,}8 \rightarrow$ U- distribucija



Simetrična



Pozitivno asimetrična
Desnostrana asimetrija



Negativno asimetrična
Ljevostrana asimetrija

8 VJEROJATNOST. TEORIJSKE DISTRIBUCIJE

KLASIČNA DEFINICIJA VJEROJATNOSTI

Teorija vjerojatnosti nastala je iz igara na sreću u 17. stoljeću.

Jacob Bernoulli – Vještina naslućivanja

Slučajni događaj – vjerojatnost realizacije slučajnog događaja A jednaka je omjeru broja povoljnih ishoda i svih mogućih ishoda.

Broj povoljnih ishoda ne može premašiti broj svih mogućih ishoda, zaključujemo da je vjerojatnost mjera slučaja koja se kreće **između 0 i 1**.

Sigurnom događaju odgovara vjerojatnost **jedan**.

Nemogućem događaju odgovara vjerojatnost **nula**, ako vjerojatnost iznosi **nula**, ne znači da je događaj nemoguć.

Vjerojatnost „a priori“ – unaprijed poznat broj svih povoljnih i ukupnih ishoda.

Vjerojatnost „a posteriori“ – vjerojatnost realizacije slučajnog događaja A nije poznata unaprijed. Jednaka je graničnoj vrijednosti relativne frekvencije kada broj pokusa teži u beskonačnost.

TEOREM O ZBRAJANJU VJEROJATNOSTI (ADICIJSKI TEOREM)

Ako dva slučajna događaja ne mogu nastupiti istodobno, kažemo da se ti događaji međusobno **isključuju**.

Vjerojatnost realizacije jednog ili drugog događaja jednaka je zbroju vjerojatnosti realizacije jednog i vjerojatnosti realizacije drugog događaja. P (A ili B) = P(A) + P(B)

Ukoliko se slučajni događaji A i B ne isključuju, tada se vjerojatnost realizacije slučajnog događaja A ili događaja B dobiva:

P(A ili B) = P(A) + P(B) – P(AB)

P(AB) – vjerojatnost da slučajni događaji A i B nastupe istovremeno.

TEOREM O MNOŽENJU VJEROJATNOSTI (MULTIPLIKACIJSKI TEOREM)

Ukoliko se događaji A i B međusobno ne isključuju, i ako vjerojatnost realizacije jednog događaja ne zavisi o vjerojatnosti realizacije drugog događaja, tada je vjerojatnost istovremene realizacije događaja A i događaja B jednaka umnošku vjerojatnosti realizacije događaja A i događaja B: P(A i B) = P(A) * P(B)

UVJETNA (KONDICIONALNA) VJEROJATNOST. BAYESOV TEOREM

Ukoliko je realizacija događaja A uvjetovana prethodnom realizacijom događaja B, radi se o uvjetnoj ili **kondicionalnoj** vjerojatnosti događaja A.

Ovdje se pita kolika je vjerojatnost da će se dogoditi događaj A ako se dogodi događaj B, vjerojatnost **da posljedica A ima za uzrok B**.

DISKONTINUIRANE SLUČAJNE VARIJABLE

Slučajna varijabla koja može poprimiti najviše prebrojivo beskonačno mnogo vrijednosti s određenom vjerojatnošću.

Zakon po kojem svakoj vrijednosti slučajne varijable X pripada vjerojatnost $P(x_i)$ – **zakon vjerojatnosti** slučajne varijable X.

Funkcija distribucije slučajne varijable X predstavlja vjerojatnost da slučajna varijabla X ne premaši neku određenu vrijednost $x_k \in \mathbf{R}$.

Funkcija distribucije slučajne varijable X je **monotono neopadajuća** funkcija.
Očekivanje slučajne varijable jednako je zbroju umnožaka vrijednosti varijable X i odgovarajućih vjerojatnosti $P(x_i)$.

TEORIJSKE DISTRIBUCIJE DISKONTINUIRANE SLUČAJNE VARIJABLE

Binomna distribucija
Poissonova distribucija
Jednolika distribucija
Hipergeometrijska distribucija

BINOMNA DISTRIBUCIJA

Ako je vjerojatnost da nastupi neki slučajni događaj poznata i uvijek ista tijekom izvođenja pokusa može se izračunati vjerojatnost da se slučajna varijabla X realizira x puta u *n* pokusa. U tom slučaju kažemo da se diskontinuirana slučajna varijabla X ravna prema tzv. binomnoj distribuciji.
Očekivanje slučajne varijable X koja se distribuira prema binomnoj distribuciji izračunava se kao i kod ostalih diskontinuiranih distribucija:
$E(X) = n * p$

Ukoliko je p=q=0,5 onda binomna distribucija ima **simetričan oblik**.

POISSONOVA DISTRIBUCIJA

Ako je vjerojatnost slučajnog događaja veoma malena i konstantna tijekom izvođenja pokusa koristi se Poissonova distribucija.
U tom slučaju broj pokusa raste u beskonačnost, ali očekivana vrijednost $\mu = n * p$ ostaje konstantna.
$n \geq 50 \qquad p \leq 0{,}10$
Što je vrijednost od n veća, a vrijednost od p manja, to je aproksimacija bolja.
Uvijek je **pozitivno asimetrična**.

Asimetričnost distribucije se **smanjuje porastom** parametra μ.

JEDNOLIKA DISTRIBUCIJA

Ako slučajna varijabla X poprima s istom vjerojatnošću bilo koju od n vrijednosti, kažemo da X ima jednoliku distribuciju.

DVODIMENZIONALNA DISKONTINUIRANA DISTRIBUCIJA

Diskontinuirana slučajna varijabla X može poprimiti vrijednosti $X_1, X_2, X_3, \ldots, X_k,$ dok diskontinuirana slučajna varijabla Y može istovremeno poprimiti vrijednost $y_1, y_2, y_3, \ldots, y_m$.
Vjerojatnost da slučajna varijabla X poprimi vrijednost x_i, a istovremeno slučajna varijabla Y poprimi vrijednost y_j: $P(X = x_i, Y = y_j) = P(x_i, y_j)$
Skup svih uređenih parova $\{(x_i,y_j);P(x_i,y_j)\}$ sačinjava dvodimenzionalnu distribuciju slučajne varijable (X,Y).

MARGINALNE DISTRIBUCIJE DISKONTINUIRANE SLUČAJNE VARIJABLE

Primjenjuje se adicioni teorem iz teorije vjerojatnosti.
Traži se vjerojatnost da slučajna varijabla X poprimi neku vrijednost X_i bez obzira na to koju će vrijednost poprimiti slučajna varijabla Y_j.
Marginalna distribucija slučajne varijable Y predstavlja vjerojatnost da varijabla Y poprimi vrijednost Y_j bez obzira koju vrijednost poprima slučajna varijabla X.

UVJETNE DISTRIBUCIJE DISKONTINUIRANE SLUČAJNE VARIJABLE

Vjerojatnost da slučajna varijabla X poprimi neku vrijednost x_i pod uvjetom da je varijabla Y poprimila neku vrijednost y_j.

TEORIJSKE DISTRIBUCIJE KONTINUIRANE SLUČAJNE VARIJABLE

Kontinuirana slučajna varijabla X može poprimiti neprebrojivo beskonačno mnogo vrijednosti.

Zbog toga se ne računa vjerojatnost u određenoj točki, nego nad određenim intervalom vrijednosti slučajne varijable X.

Vjerojatnost da će neka kontinuirana slučajna varijabla X poprimiti neku određenu vrijednost x jednaka je **nuli**, ali to ne znači i nemogući događaj.

Funkcija vjerojatnosti kontinuirane slučajne varijable X ima sljedeća **svojstva**:

a) $f(x) \geq 0$

b) $\int_{+\infty}^{+\infty} f(x)dx = 1$, površina ispod krivulje jednaka je 1.

c) $\int_{x_1}^{x_2} f(x)dx = P(x_1 < X \leq x_2,\ x_2 > x_1$

Funkcija distribucije kontinuirane slučajne varijable X:

$$F(x) = \int_{-\infty}^{x} f(x)dx\,, \frac{dF(x)}{dx} = f(x)$$

F(x) predstavlja vjerojatnost da slučajna varijabla X ne premaši neku unaprijed zadanu vrijednost x.

NORMALNA DISTRIBUCIJA

Parametri normalne distribucije su njeno očekivanje μ i varijanca σ^2.

Normalna distribucija je potpuno simetrična distribucija, pa svi koeficijenti asimetrije iznose **nula**, dok je vrijednost **koeficijenta zaobljenosti** jednaka **3**.

STUDENTOVA DISTRIBUCIJA

Česta u primjenama kod procjene parametara i kod testiranja hipoteza na osnovu uzorka.

Područje vrijednosti varijable „t“ je interval $(-\infty; +\infty)$.

Studentova distribucija je simetrična s obzirom na t=0. Spljoštenija je od normalne distribucije.

Studentova distribucija teži jediničnoj normalnoj distribuciji.

v – broj stupnjeva slobode

$\boxed{v > 30}$, pogreška aproksimacije je manja od 0,3

HI-KVADRAT DISTRIBUCIJA

Ako su $X_1, X_2, X_3, \ldots, X_n$ nezavisne normalne varijable koje imaju jednaka očekivanja, $E(X_1)=\ldots=E(X_n)=\mu$ i jednake varijance $V(X_1)=\ldots=V(X_n)=\sigma^2$

Područje vrijednosti varijable χ^2 je $(0;+\infty)$.

Za različit broj stupnjeva slobode distribucija ima različit oblik.

Za $\boxed{v \geq 3}$ distribucija je **pozitivno asimetrična**.

Što je broj stupnjeva slobode veći, distribucija je bliža obliku normalne distribucije. Kod v > 30 upotrebljava se aproksimacija normalnom distribucijom.

9 OCJENE PARAMETARA OSNOVNOG SKUPA NA OSNOVU UZORKA

UZORAK

Uzorak – podskup osnovnog statističkog skupa koji se uzima u svrhu ispitivanja obilježja elemenata osnovnog skupa.
Da bi tu svrhu mogao ispuniti, potrebno je da bude **reprezentativan** i da je **izbor jedinica** izvršen na **slučajan način**.
Frakcija odabiranja – omjer broja jedinica u uzorku i broja jedinica u osnovnom skupu.
Recipročna vrijednost frakcije odabiranja naziva se **korakom izbora**, a upotrebljava se kod sistemskog izbora jedinica u uzorak.

PROCJENA ARITMETIČKE SREDINE OSNOVNOG SKUPA

$\boxed{\Pr\{\hat{\bar{x}} - z \cdot Se(\hat{\bar{x}}) < \bar{x} < \hat{\bar{x}} + z \cdot Se(\hat{\bar{x}})\}}$

Distribucija aritmetičkih sredina svih mogućih uzoraka naziva se ***sampling distribucija***.
Ukoliko uzorci potječu iz osnovnih skupova koji su normalno distribuirani, onda je sampling distribucija također normalnog oblika.
Sampling distribucija aritmetičkih sredina teži normalnom obliku kada veličina uzorka teži k beskonačnosti i u slučaju kada uzorci ne potječu iz normalno distribuiranih osnovnih skupova.
n≤30 – studentova t distribucija.
Nepristranost – očekivana vrijednost sampling distribucije aritmetičkih sredina jednaka je očekivanoj vrijednosti osnovnog skupa.
Pristranost – očekivana vrijednost sampling distribucije može biti jednaka očekivanoj vrijednosti osnovnog skupa. Razlika između očekivane vrijednosti parametra u osnovnom skupu i u sampling distribuciji.
Pristranost procjene na bazi uzorka predstavlja razliku između očekivane vrijednosti parametra u sampling distribuciji i vrijednosti populacijskog parametra (parametra osnovnog skupa).
Ocjena aritmetičke sredine na osnovu uzorka je nepristrana ako se uzima jednostavni slučajni uzorak.
Konzistentnost – za neki procjenitelj $\hat{\Theta}$ kažemo da konzistentno procjenjuje parametar Θ ako $\hat{\Theta}$ teži prema Θ.
Aritmetička sredina uzorka konzistentno procjenjuje aritmetičku sredinu osnovnog skupa.
Standardna greška – standardna devijacija sampling distribucije
Ukoliko standardna devijacija osnovnog skupa nije poznata, upotrebljavamo procjenu na osnovu uzorka. $\boxed{Se(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}}$ ili
$\boxed{Se(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$ ako je **f > 0,05**.
Procjena standardne devijacije na osnovu uzorka je pristrana.
Procjena aritmetičke sredine na osnovu uzroka naziva se **točkastom procjenom aritmetičke sredine osnovnog skupa**.
Određivanje veličine uzorka:
$\boxed{n' = \left[\frac{z \cdot \sigma}{greška}\right]^2}$, ako je **f > 0,05** $\boxed{n = \frac{n'}{1+\frac{n'}{N}}}$
1 – α – nivo pouzdanosti procjene
n≤30 → studentova t distribucija.

PROCJENA VARIJANCE OSNOVNOGA SKUPA NA OSNOVU UZORKA

$\boxed{Pr\left\{\frac{n \cdot S^2}{{\chi_{\alpha/2}}^2} < \sigma^2 < \frac{n \cdot S^2}{{\chi_{1-\alpha/2}}^2}\right\}}$

Sampling distribucija varijanci ima oblik Hi–kvadrat distribucije.

Ukoliko uzorak raste u beskonačnost, sampling distribucija teži k normalnom obliku.

$$\chi_{\alpha/2}{}^2 = \frac{1}{2} \cdot \left(z_{\alpha/2} + \sqrt{2 \cdot v - 1}\right)^2$$
$$\chi_{1-\alpha/2}{}^2 = \frac{1}{2} \cdot \left(-z_{\alpha/2} + \sqrt{2 \cdot v - 1}\right)^2$$

Intervalna procjena standardne devijacije dobiva se tako da se izračuna pozitivni korijen iz donje i gornje granice intervalne procjene varijance osnovnoga skupa.

PROCJENA TOTALA OSNOVNOG SKUPA

$$\Pr\left\{N \cdot \hat{\bar{x}} - z \cdot Se(\bar{x}) < \sum_{i=1}^{N} x_i < N \cdot \hat{\bar{x}} + z \cdot Se(\bar{x})\right\} = 1 - \alpha$$

Procjena totala vrši se na osnovu uzorka pomoću procjene aritmetičke sredine.

Procjena totala na osnovu uzorka također je **nepristrana** jer je procjena aritmetičke sredine nepristrana.

$Se(\sum x_i) = N \cdot Se(\bar{x})$ standardna greška

PROCJENA PROPORCIJE (RELATIVNE FREKVENCIJE) OSNOVNOGA SKUPA

$$Pr\{\hat{p} - z \cdot Se(p) < P < \hat{p} + z \cdot Se(p)\} = 1 - \alpha$$

Određivanje veličine uzorka:

$n' = \left[\frac{z \cdot \sqrt{P \cdot Q}}{greška}\right]$ P*Q – **varijanca osnovnog skupa**

Standardna greška:

$Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$ a ako je **n ≤ 30** $Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}}$

$\hat{q} = 1 - \hat{p}$

Ako je **f > 0,05** onda se standardna greška korigira s $\sqrt{\frac{N-n}{N-1}}$

Nikako se ne radi s malim uzorcima i preporučuje se da veličina uzorka ne smije biti manja od **100**.

10 TESTIRANJE HIPOTEZA

TESTIRANJE HIPOTEZE O NEPOZNATOJ ARITMETIČKOJ SREDINI OSNOVNOG SKUPA

Postavljaju se sljedeće hipoteze:

$H_0: \ldots \bar{x} = \overline{x_0}$
$H_1: \ldots \bar{x} \neq \overline{x_0}$

Interval prihvaćanja hipoteze H_0 glasi:

$\overline{x_0} \pm z \cdot Se(\bar{x})$

Ako je veličina uzorka n≤30, tada se upotrebljava „t" iz Studentove distribucije umjesto koeficijenta „z" iz normalne distribucije.

Ako se aritmetička sredina uzorka nalazi u navedenom intervalu, **prihvaćamo hipotezu H_0.**

Nulta hipoteza – izraz je potpuno odgovarajući onda kada testiramo hipotezu da je vrijednost nekog parametra jednaka nuli.

Ukoliko se aritmetička sredina uzorka nalazi izvan navedenog intervala, prihvaća se **alternativna hipoteza H_1**.

Ako je $n > 30 \rightarrow z^*$

Da bi se izvršilo testiranje, potrebno je pronaći kritičnu vrijednost za „t", odnosno „z" iz tablica.

Ako je uzorak mali, onda se uzima t iz Studentove distribucije sa (n–1) stupnjeva slobode i uz signifikantnost testa α.

Pri testiranju na **donju granicu** postavlja se alternativna hipoteza da je aritmetička sredina osnovnog skupa manja od neke pretpostavljene vrijednosti.

Hipoteza H_0 glasi da je aritmetička sredina osnovnog skupa veća ili jednaka nekoj pretpostavljenoj vrijednosti.

Dakle, kod hipoteze H_0 postoji familija pretpostavljenih vrijednosti aritmetičke sredine, koja se testira nasuprot familiji aritmetičkih sredina sadržanih u alternativnoj hipotezi.

$H_0: \ldots \bar{x} \geq \overline{x_0}$
$H_1: \ldots \bar{x} < \overline{x_0}$

Donja granica prihvaćanja nulte hipoteze:

$DG = \overline{x_0} - z \cdot Se(\bar{x})$

Ako aritmetička sredina uzorka ima ***vrijednost manju*** od donje granice prihvaćanja nulte hipoteze, prihvaća se alternativna hipoteza da je aritmetička sredina manja od neke pretpostavljene vrijednosti.

Testiranje na gornju granicu:

$H_0: \ldots \bar{x} \leq \overline{x_0}$
$H_1: \ldots \bar{x} > \overline{x_0}$

Hipotezom H_0 tvrdi se da vrijednost aritmetičke sredine osnovnog skupa **ne premašuje** pretpostavljenu vrijednost $\overline{x_0}$. Alternativnom hipotezom tvrdi se da je aritmetička sredina osnovnog skupa veća od pretpostavljene vrijednosti.

Gornja granica prihvaćanja:

$GG = \overline{x_0} + z \cdot Se(\bar{x})$

Ukoliko aritmetička sredina ***premaši gornju granicu*** prihvaćanja nulte hipoteze, **prihvaća se alternativna hipoteza**, tj. hipoteza da je aritmetička sredina osnovnog skupa veća od neke pretpostavljene vrijednosti.

GREŠKA TIPA I

Vjerojatnost da odbacimo nultu hipotezu ukoliko je ona istinita. Ta vjerojatnost je jednaka nivou signifikantnosti testa α (5%).

GREŠKA TIPA II (β)

Vjerojatnost da prihvatimo nultu hipotezu premda ona nije istinita.

Veličina greške tipa II je funkcija istinite vrijednosti aritmetičke sredine osnovnog skupa. Što se više razlikuju prava i pretpostavljena aritmetička sredina, to je veličina greške tipa II manja.

Ako je vrijednost prave aritmetičke sredine jednaka pretpostavljenoj vrijednosti, greška tipa II iznosi 1–α.

Snaga testa – funkcija istinite vrijednosti aritmetičke sredine osnovnog skupa. Veličina snage testa predstavlja vjerojatnost da ne prihvatimo lažnu nultu hipotezu. **1–β**

Snaga testa je veća što je greška tipa II manja. Po tome se može zaključiti da **jednosmjerni testovi imaju veću snagu od dvosmjernih testova**.

Ukoliko se smanji greška tipa I, povećava se greška tipa II, i obratno.

Odluka u svezi nul-hipoteze	Stvarno stanje nul-hipoteze	
	Istinita	**Lažna**
Odbacujemo nul-hipotezu	Greška tipa I (α)	Ispravan zaključak
Prihvaćamo nul-hipotezu	Ispravan zaključak	Greška tipa II (β)

TESTIRANJE HIPOTEZE O RAZLICI ARITMETIČKIH SREDINA DVAJU NEZAVISNIH OSNOVNIH SKUPOVA

Postavlja se nulta hipoteza da nema značajne razlike između aritmetičkih sredina dvaju nezavisnih osnovnih skupova.

Nezavisni uzorci potječu iz različitih osnovnih skupova te među njima ne postoji povezanost.

Ukoliko se na istom uzorku vrše različita mjerenja pod različitim uvjetima, onda govorimo o **zavisnim** uzorcima.

Kod zavisnih uzoraka postoji korelacija između rezultata prije i poslije eksperimenta.

Kod nezavisnih uzoraka standardna greška je veća negoli kod zavisnih.

$$\boxed{\begin{array}{l} H_0:\ldots \overline{x_1}-\overline{x_2}=0 \\ H_1:\ldots \overline{x_1}-\overline{x_2}\neq 0 \end{array}}$$

Interval prihvaćanja nulte hipoteze:

$\boxed{0 + z \cdot Se(\overline{x_1}-\overline{x_2})}$

Standardna greška $\boxed{Se(\overline{x}_1-\overline{x}_2)=\sigma\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Ako se radi o velikom uzorku:

$\boxed{Se(\overline{x}_1-\overline{x}_2)=\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}}$

Pretpostavlja se da su varijance osnovnih skupova **jednake**, ali su nepoznate.

Treba naći apsolutnu vrijednost razlike aritmetičkih sredina uzoraka.

Ukoliko ta razlika **premaši interval** prihvaćanja nulte hipoteze, tada prihvaćamo **alternativnu hipotezu kao istinitu**, tj. zaključujemo da se aritmetičke sredine dvaju osnovnih skupova značajno razlikuju.

TESTIRANJE HIPOTEZE O NEPOZNATOJ PROPORCIJI (RELATIVNOJ FREKVENCIJI) OSNOVNOG SKUPA

Nultom hipotezom pretpostavlja se da je proporcija (relativna frekvencija) osnovnog skupa jednaka nekoj pretpostavljenoj vrijednosti:

$$\boxed{\begin{array}{l} H_0:\ldots P=P_0 \\ H_1:\ldots P\neq P_0 \end{array}}$$

Interval prihvaćanja nulte hipoteze:

$\boxed{P_0 \neq z \cdot Se(p)}$

Standardna greška: $\boxed{Se(p)=\sqrt{\frac{P_0\cdot Q_0}{n}}}$ a ako je **$n \leq 30$** $\boxed{Se(p)=\sqrt{\frac{P_0\cdot Q_0}{n-1}}}$

Ako se proporcija iz uzroka nalazi u **gornjem intervalu** prihvaćanja, tada **prihvaćamo nultu hipotezu kao istinitu**.

Ako se proporcija uzorka nalazi **izvan** intervala prihvaćanja, tada prihvaćamo kao moguću **alternativnu hipotezu**, tj. zaključujemo na nivou signifikantnosti α da proporcija osnovnog skupa **nije** jednaka nekoj pretpostavljenoj vrijednosti.

TESTIRANJE HIPOTEZE O RAZLICI PROPORCIJA (RELATIVNIH FREKVENCIJA) DVAJU NEZAVISNIH OSNOVNIH SKUPOVA

Nultom hipotezom pretpostavlja se da nema razlike u proporcijama dvaju nezavisnih skupova, odnosno da uzroci koje smo odabrali potječu iz istih populacija.

$H_0:...P_1=P_2$
$H_1:...P_1 \neq P_2$

Interval prihvaćanja nulte hipoteze:

$0+z \cdot Se(p_1-p_2)$

$\hat{\bar{p}}$ – prosječna proporcija za oba uzorka zajedno

Standardna greška:

$$Se(p_1-p_2)=\sqrt{\hat{\bar{p}}\cdot(1-\hat{\bar{p}})\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\hat{\bar{p}}=\frac{m_1+m_2}{n_1+n_2}$$

Ukoliko se razlika proporcija uzoraka nalazi **u granicama** intervala prihvaćanja nulte hipoteze **prihvaćamo nultu hipotezu kao istinitu**.

Ukoliko se ta razlika nalazi **izvan granica** intervala prihvaćanja **prihvaćamo alternativnu hipotezu**, tj. zaključujemo da se proporcije dvaju osnovnih skupova razlikuju, uz rizik da smo pogrešno odbacili nultu hipotezu koji je jednak razini signifikantnosti testa α.

TESTIRANJE HIPOTEZE O RAZLICI PROPORCIJA TRIJU ILI VIŠE OSNOVNIH SKUPOVA

Može se testirati pomoću Hi-kvadrat testa. Ako su proporcije u jednoj distribuciji jednake radi se o jednolikoj razdiobi, pa ovaj test predstavlja **testiranje hipoteze ima li distribucija oblik jednolike distribucije**.

Postavljaju se hipoteze za testiranje **jednolike distribucije**:

$H_0:...P_1=P_2=...=P_k=P$
$H_1=\exists P_1 \neq P$

Empirijska vrijednost Hi-kvadrat testa:

$$\chi^{2*}=\sum_{i=1}^{k}\frac{(m_i-e_i)^2}{e_i}$$

Tablična vrijednost testa se traži iz tablica Hi-kvadrat distribucije ($\chi_{tab}{}^2=[\alpha, df]$), uz odgovarajuću razinu signifikantnosti α i stupnjeve slobode *df*, ***df=k-1***

$\chi^{2*} < \chi_{tab}{}^2$ → H_0 što znači da zadana distribucija ima oblik jednolike distribucije.

HI-KVADRAT TEST

Spada u neparametrijske testove, ne pretpostavlja se oblik distribucije.

Spada u testove koji se zasnivaju na *rasporedu frekvencije* unutar tabele kontingence.

Pretpostavlja se da podaci potječu iz slučajno odabranog uzroka.

a) Testiranje hipoteze da distribucija ima određeni oblik
b) Testiranje hipoteze o nezavisnosti obilježja

TESTIRANJE HIPOTEZE DA DISTRIBUCIJA IMA ODREĐENI OBLIK

Može se testirati hipoteza da distribucija slučajne varijable X ima oblik neke teorijske distribucije.

Testiranja ima li distribucija oblik **jednolik**, **binomne**, **Poissonove** ili **normalne distribucije**.

Kod testiranja hipoteze da distribucija ima **normalni oblik** ne računa se vjerojatnost u točki nego vjerojatnost nad određenim intervalom i ta vjerojatnost također množi sa zbrojem frekvencija.

Empirijska vrijednost Hi-kvadrata se uspoređuje s kritičnom vrijednošću Hi-kvadrata iz tablica na određenom nivou signifikantnosti testa α i uz određeni broj stupnjeva slobode df=v.

Broj stupnjeva slobode se izračunava:
Za jednoliku distribuciju **df=k-1**
Za binomnu distribuciju **df=k-2**
Za Poissonovu distribuciju **df=k-2**
Za normalnu distribuciju **df=k**
K – broj frekvencija
Ukoliko empirijska vrijednost Hi-kvadrata **ne premaši** kritičnu vrijednost iz tablica, **prihvaćamo hipotezu da empirijska distribucija ima pretpostavljeni oblik**.
Ukoliko je empirijska vrijednost Hi-kvadrata **veća** od kritične vrijednosti iz tablica, **prihvaćamo alternativnu hipotezu da distribucija nema pretpostavljeni oblik**.
Očekivane (teorijske) frekvencije ne smiju biti male, i **ne bi smjele biti manje od 5**.

TESTIRANJE HIPOTEZE O NEZAVISNOSTI OBILJEŽJA ELEMENATA OSNOVNOG SKUPA

Nulta hipoteza glasi da ne postoji zavisnost dvaju obilježja.
Test-veličina:

$$\chi^{2*} = \sum_{i=1}^{r}\sum_{i=1}^{c}\frac{(m_{ij}-e_{ij})^2}{e_{ij}}$$

m_{ij} – originalne frekvencije (empirijske)
e_{ij} – očekivane frekvencije (teorijske)
Hipoteze glase:

$H_0:\ldots P_{ij}=P_{i\cdot}\cdot P_{\cdot j}$
$H_1:\ldots \exists P_{ij}\neq P_{i\cdot}\cdot P_{\cdot j}$

$$e_{ij} = \frac{m_{i\cdot}\cdot m_{\cdot j}}{n}$$

$m_{i\cdot}$ – marginalna frekvencija i-tog retka
$m_{\cdot j}$ – marginalna frekvencija j-tog stupca

Navedena veličina uspoređuje se s kritičnom vrijednošću Hi-kvadrata iz tablica.
Broj stupnjeva slobode jednak je
df = (r-1)(c-1),
r – broj redaka
c – broj stupaca

Ukoliko empirijska vrijednost Hi-kvadrata **ne premaši** kritičnu vrijednost Hi-kvadrata iz tablica, **prihvaćamo nultu hipotezu kao istinitu**, tj. zaključujemo da nema ovisnosti obilježja elemenata osnovnog skupa.
Ako vrijednost empirijskog Hi-kvadrata **premaši** kritičnu vrijednost **prihvaćamo alternativnu hipotezu**, tj. zaključujemo da postoji zavisnost dvaju obilježja elemenata osnovnog skupa.

Hi-kvadrat testom ustanovljava se samo **vjerojatnost povezanosti dviju varijabli**, ali ne i visina povezanosti.
Pearsonov koeficijent kontigencije – aproksimativna visina povezanosti
Najmanja vrijednost koeficijenta kontingence iznosi **nula**.
Hi-kvadrat test može se izvesti samo s **apsolutnim frekvencijama**.
To znači da nije dopušteno rabiti postotke ili bilo koje druge izvedene veličine.
Ukoliko je broj stupnjeva slobode veći od 1, potrebno je da 20% ćelija ima očekivane frekvencije manje od 5, a ni jednu očekivanu frekvenciju manju od 1.
Kod tablica koje imaju samo jedan stupanj slobode potrebno je provesti korekciju za kontinuitet.