



# Statistika

---

Stefan Šljivić 229/2023

Mihailo Damnjanović 235/2022

Vuk Stamenković 187/2023



# Statistika

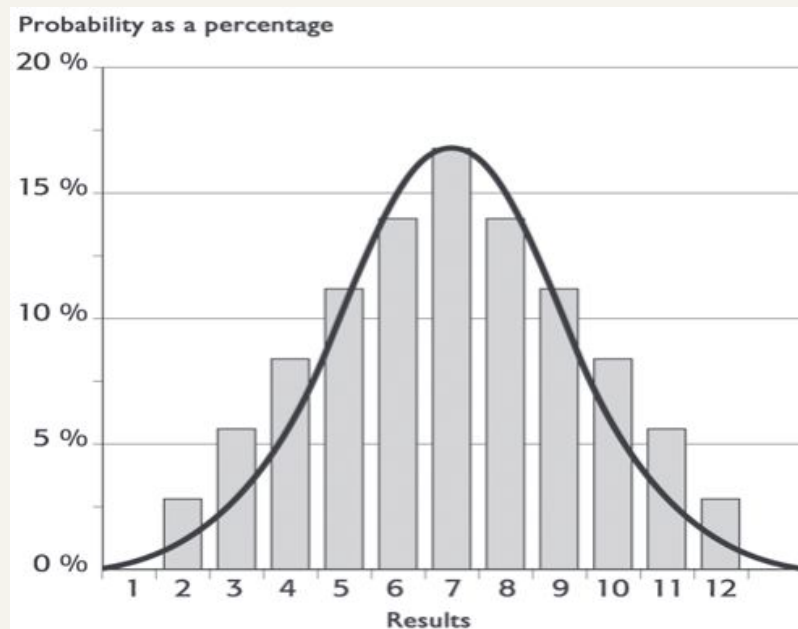
Statistika je grana matematike koja se bavi razvojem metoda za prikupljanje, analizu i prezentaciju podataka.

Deskriptivna statistika je manja grana statistike za opisivanje bilo kakvih opservacija.

# Gausova kriva

Gausova kriva (normalna distribucija) predstavlja veliki broj nasumičnih ljudskih, odnosno prirodnih fenomena.

Primer: zbir dobijen bacanjem dve kockice.



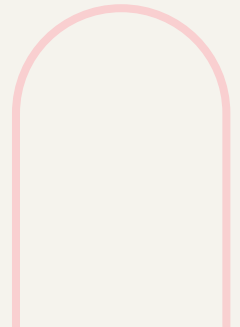
# Mere centralne tendencije

Mere centralne tendencije su:

- aritmetička sredina - zbir vrednosti svih podataka podeljen sa brojem podataka
- medijana - vrednost koja razdvaja gornju polovinu podataka od donje kada se podaci poređaju u rastućem redosledu
- modus - vrednost koja se najčešće ponavlja

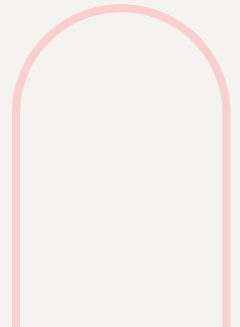
# Mere centralne tendencije

- Imaju slične vrednosti samo kod normalne distribucije.
- Oslanjanje na mere centralne tendencije u nekim slučajevima može biti obmanljivo.
- Svaka mera ima prednosti i mane.



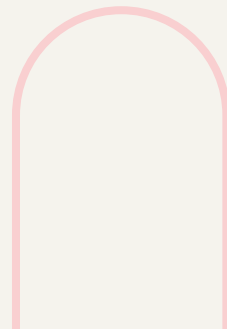
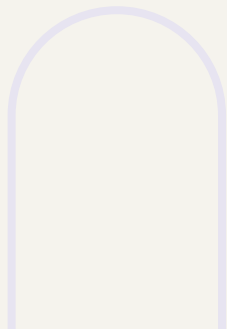
# Aritmetička sredina

- Uvek postoji
- Uzima sve vrednosti u obzir
- Osetljiva na ekstremne vrednosti


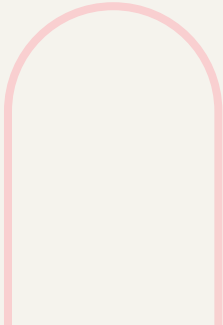


# Medijana

- Uvek postoji
- Ne uzima sve vrednosti u obzir (osim količine podataka)
- Nije osetljiva na ekstremne vrednosti



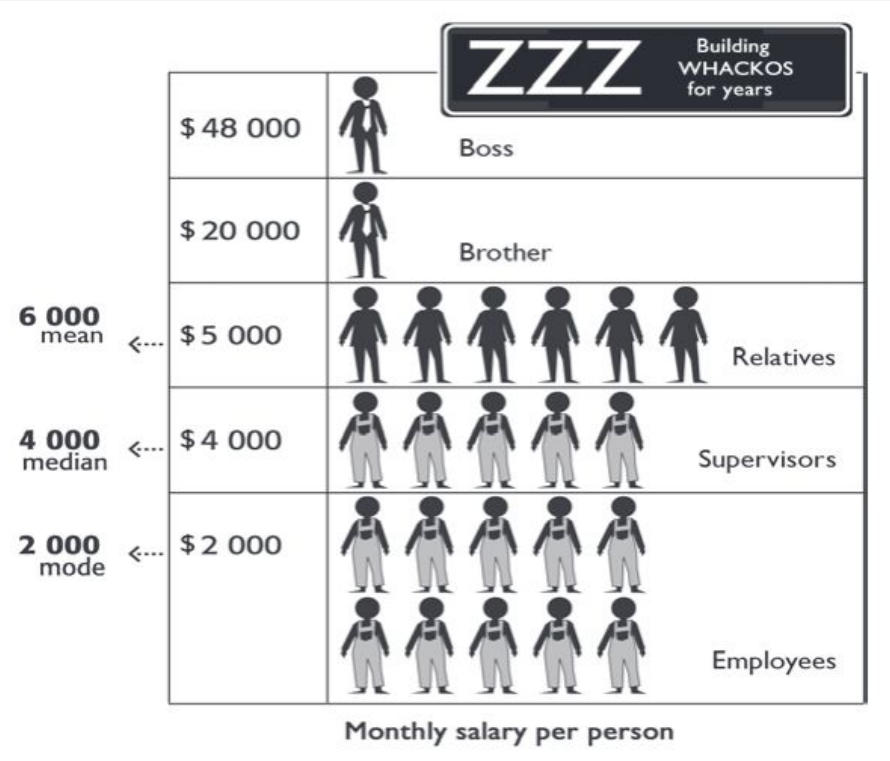
# Modus

- Može da postoji jedan ili više modusa, ili da ne postoji
  - Ne uzima sve vrednosti u obzir
  - Koristi se za opisivanje nominalnih (promenljive opisane imenom) ili diskretnih (promenljive koje imaju ograničen broj vrednosti) promenljivih
- 
- 



# Primer manipulacije

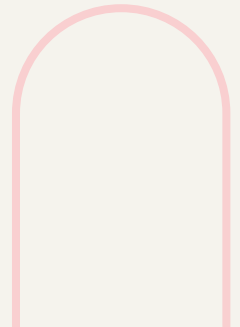
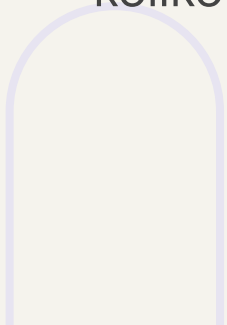
Firma koja zapošljava radnike tvrdi da ima prosečnu platu 6000\$.



# Standardna devijacija

Primer: Odlučujete da li da pojedete ribu iz zagađenog jezera. Toksičnost ribe je data Gausovom krivom gde je prosečna količina toksina u ribi 4mg. Za čoveka je opasno da unese 7mg toksina.

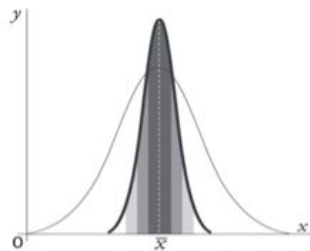
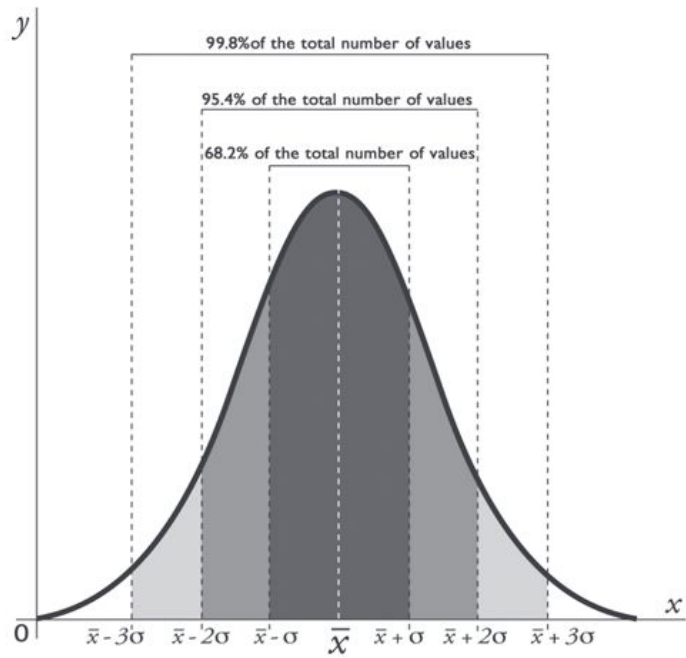
- Odgovor bi trebalo da zavisi od standardne devijacije koja će vam reći koliko vrednosti odstupaju od aritmetičke sredine.



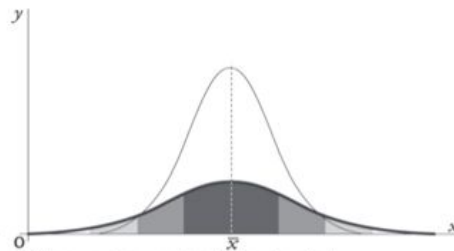
# Standardna devijacija

Standardna devijacija je mera disperzije podataka u odnosu na aritmetičku sredinu.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$



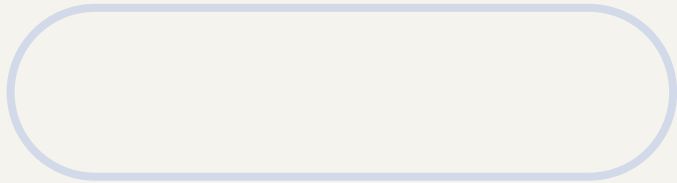
The curve is narrow, the standard deviation  $s$  is small, and precision is high



The curve is spread out, the standard deviation  $s$  is big, and precision is low.



# Ankete

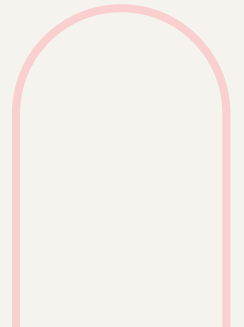


# Uzorci

Formiranje uzoraka je jedna od najvažnijih primena statistike.

Uzorak je deo populacije na osnovu kog se donosi zaključak o celoj populaciji.

Uzorci rešavaju dosta problema kod istraživanja populacije.



# Uzorci

Uzorak mora da bude predstavnik populacije da bi bio validan.

Uzorak mora da bude dovoljno **veliki i nepristrasan**.

Literary Digest - primer u kom je pristrasnost uzorka dovela do pogrešnog rezultata istraživanja populacije.

# Veličina uzorka

Koja je optimalna veličina uzorka?

Odgovor zavisi od populacije koja se istražuje, željenog stepena preciznosti, budžeta, teme istraživanja, itd.

Najčešća veličina uzorka je od 1000 do 2000 ljudi zbog odnosa cene i preciznosti.





# Biranje uzorka

Elementi iz populacije moraju biti odabrani nasumično.

„Nasumično biranje” - teoretski proces gde se iz populacije  $P$  nasumično odabere  $n$  elemenata. Vrlo često je nepraktičan.

Metode biranja uzorka: stratifikacija, klastering, kvota.



# Primer 1 - neadekvatno biranje uzorka

Radio stanica vrši anketu o legalizaciji marihuane.

- Veličina uzorka: 3636
- Rezultat: 78% ljudi glasalo za.

Problem: Uzorak nije bio nasumičan, sastojao se samo od slušalaca te stanice koji su se uz to verovatno javljali jer im je data tema važna.

# Primer 2 - neadekvatno biranje uzorka

Američki časopis vrši anketu o znanju međunarodnog sistema mernih jedinica.

- Rezultat: 98% ljudi zna

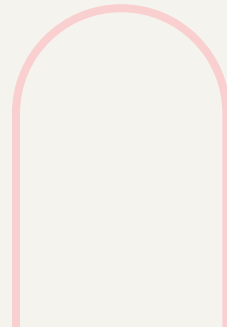
Problem: Isključivo ljudi koji su znali sistem su odgovarali na pitanja.



# Primer 3 - neadekvatno biranje uzorka

Kompjuter je iz baze svih brojeva telefona nasumično odabrao 2000 brojeva za anketu.

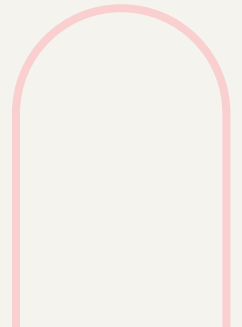
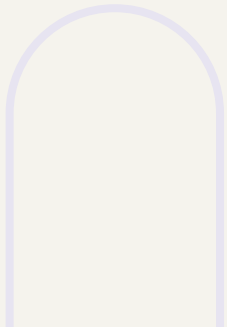
Problem: Iz ankete su isključeni najsiromašniji koji ne mogu da priušte telefon.



# Margina greške

Dobre ankete će vam reći stepen tačnosti.

Primer: U 95% slučajeva, margina greške je 5%.



# Primer neuračunate margine greške

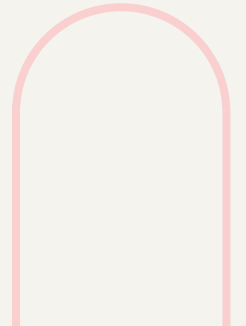
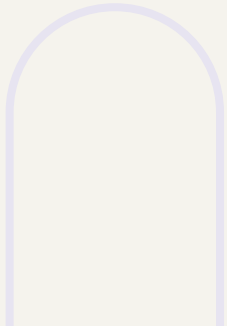
U januaru anketa tvrdi da je predsednikova popularnost 53%, a u martu 56%. Novine pišu o tome kako predsednikova popularnost raste.

Problem: Margina greške od 5% nije rečena, u januaru je popularnost 48%-58%, a u martu 51%- 61%.

# Formulacija pitanja

Pitanje mora da bude jednosmisleno i nepristrasno.

Dobre ankete prvo testiraju svoja pitanja na manjem uzorku.



# Primer 1 - pristrasno formulisano pitanje

Policija tvrdi da 86% populacije podržava kupovinu novih pendreka.

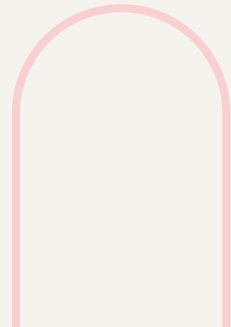
Pitanje: Uzevši u obzir rast broja anarhista i efektivnosti koje je nova generacija pendreka pokazala sa njihovim suzbijanjem da li podržavate zamenu starih, istrošenih pendreka povoljnijim i ergonomičnijim novim pendrecima?



# Primer 2 - dvosmisleno formulisano pitanje

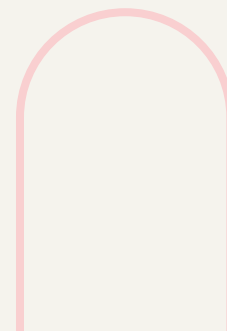
Da li čitate „New York Times“?

Pitanje zvuči precizno, ali može da se protumači na više načina u zavisnosti od učestalosti čitanja i količine teksta koje bi neko mogao da čita.



# Dobra anketa

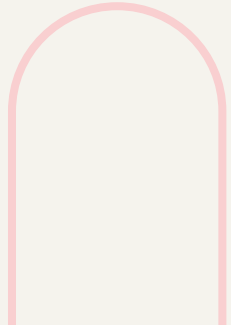
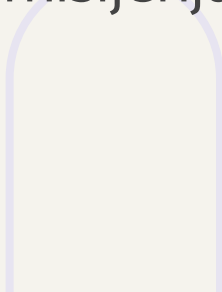
- Uzorak je dovoljno veliki i nasumičan.
- Data je margina greške.
- Pitanje formulisano jednosmisleno i nepristrasno.

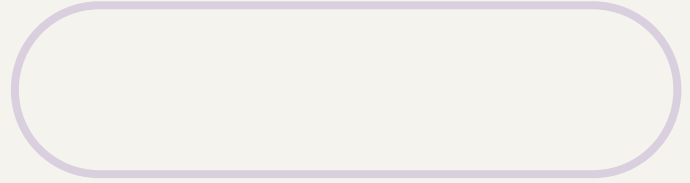


# Ankete o stavovima

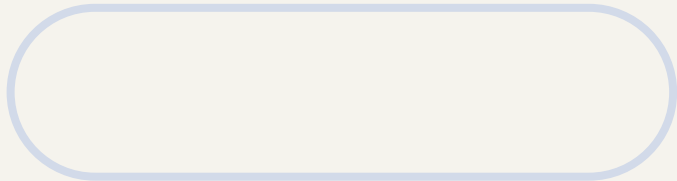
Ankete o stavovima nekada mogu biti nerelevantne.

„Mišljenje populacije iskazano anketom je obmana čija je funkcija da u bilo kom trenutku razdvoji stanje mišljenja, sistema sila i tenzija, i ništa nije neadekvatnije za prikazivanje stanja mišljenja od procenta.” - Pjer Burdje, sociolog





# Korelacija



# Korelacija

Korelacija znači da su vrednosti dve promenljive povezane na neki način.

**Korelacija ne mora da znači da promenljive imaju uzročno-posledične veze.**

A i B su u korelaciji može da znači više stvari:

- A uzrokuje B
- B uzrokuje A
- A i B su slučajno povezani bez uzročno-posledične veze
- A i B su zavisni od C

# Korelacija - primeri

Studenti koji puše imaju lošije ocene.

Cena kafe u Oregonu je u korelaciji sa količinom padavina u nekom delu sveta.

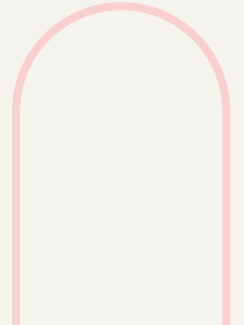
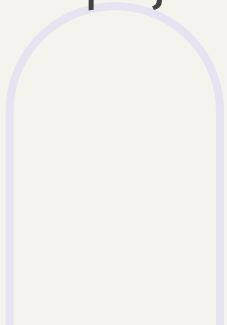
U nekim delovima sveta gnezda roda na krovovima su u korelaciji sa brojem dece koja žive u kući.

Količina kose kod muškaraca je u korelaciji sa godinama njihovih baba.

# Korelacija - primeri

Monasi u Kini su verovali da moraju da udare u gong kako bi oterali nebeskog psa da ne pojede mesec.

Mešanje korelacije i uzročno-posledičnog odnosa može dovesti do pojava sujeverja i pojave pod nazivom regresija ka sredini.





# Sujeverje i regresija ka sredini





# Regresija ka sredini

Regresija ka sredini je pojava gde dve promenljive koje zavise od mnogo faktora nisu u savršenoj korelaciji.

Ekstremna vrednost jedne promenljive je u korelaciji sa manjom ekstremnom vrednošću druge.

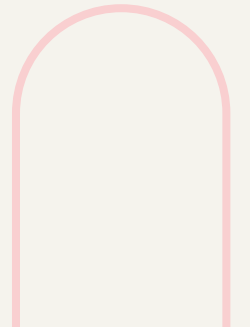
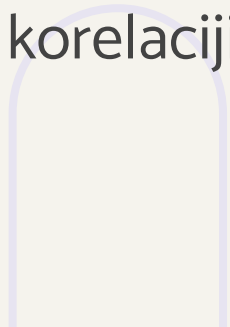
Regresija ka sredini može dovesti do sujeverja ako se odnos promenljivih shvati kao uzročno-posledični.

# Regresija ka sredini - primer

Frensis Galton je istraživao povezanost visine očeva sa visinom njihovih sinova.

Uvideo je da veoma visoki očevi imaju sinove koji su niži od njih.

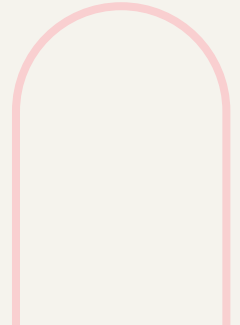
Visina zavisi od mnogo faktora i promenljive su u nesavršenoj korelaciji.



# Sujeverje

Neki sportisti ne žele da se pojavljuju na naslovnim stranama časopisa jer veruju da će nakon toga imati slabiji učinak.

Objašnjenje: Pozvani su nakon izuzetnog učinka koji zavisi od mnogo faktora i nije u savršenoj korelaciji sa narednim učinkom.



**Hvala na pažnji!**