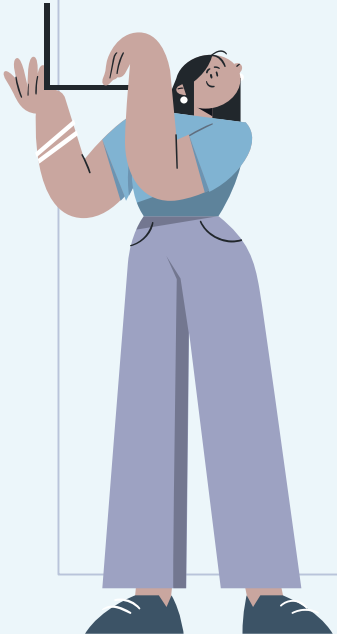





# STATISTIKA



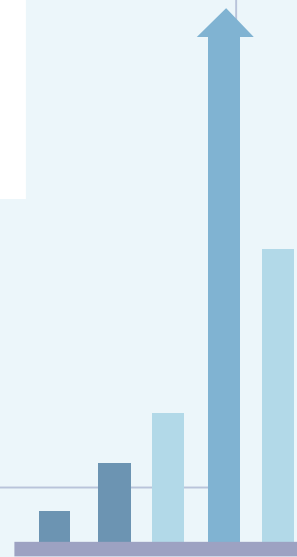
Maša Jevtić  
Aleksandra Krstić  
Tea Korelec  
Jovana Stefanović  
Anđela Čomagić



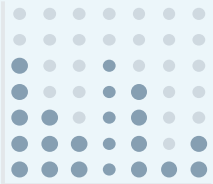


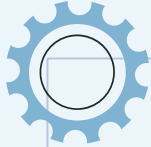
# Sadržaj prezentacije

1. Kratka istorija statistike
2. Mere deskriptivne statistike
3. Istraživanje i uzimanje uzoraka
4. Korelacija
5. Regresija ka sredini



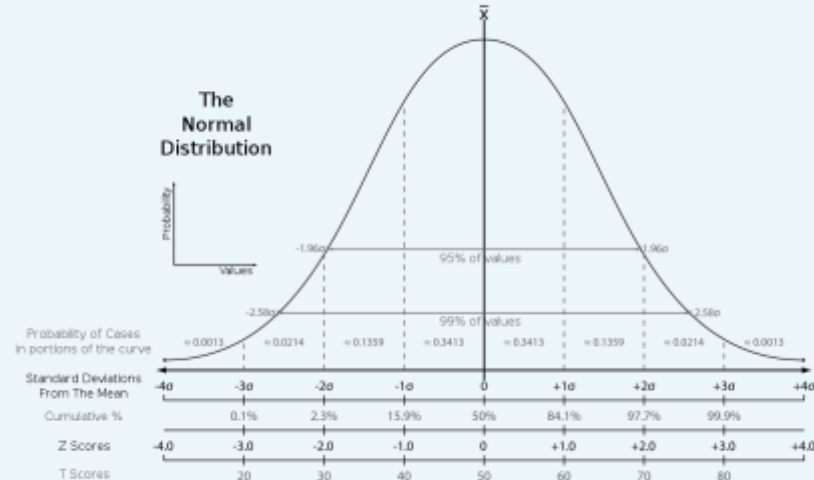
# ŠTA JE STATISTIKA?





# Definicija statistike

Statistika je funkcija uzoraka  $x_1, x_2 \dots x_n$  čiji analitički izraz ne zavisi od nepoznatih parametara raspodele obeležja populacije iz koje je uzorak uzet. Dobija matematičku osnovu u 17.veku.



primer: Gausova kriva



# Istorija statistike



Džona Granta

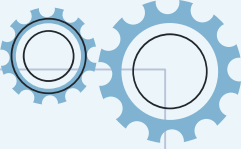


Đirolamo Kardano



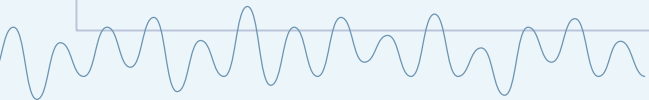
Karl Pirson





# Podela statistike

- Deskriptivna i inferencijalna
- Mere centralne tendencije, mere varijabiliteta, grafička i tabelarna prikazivanja osnovnih statističkih vrednosti
- Proveravanje postavljenih hipoteza uz pomoć statističkih testova, koeficijenata...

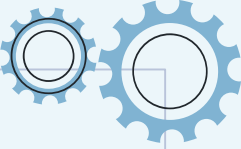




# Deskriptivna statistika

- Deskriptivna statistika nam omogućava da opišemo posmatranja o bilo čemu - ljudima, objektima ili događajima - populacijama.
- Međutim, nekad moramo smanjiti skup podataka na jednu vrednost. Mere centralne tendencije nam to omogućavaju, jer pokazuju tipične ili centralne tendencije podataka. One su široko korišćene i veoma korisne.





# Mere deskriptivne statistike

**01**

Aritmetička sredina

**03**

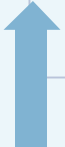
Modus

**02**

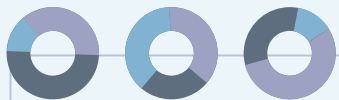
Medijana

**04**

Standardna devijacija

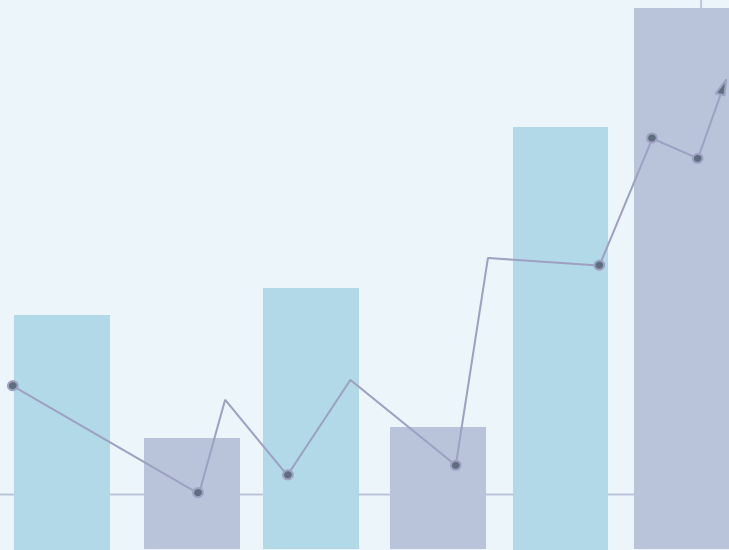
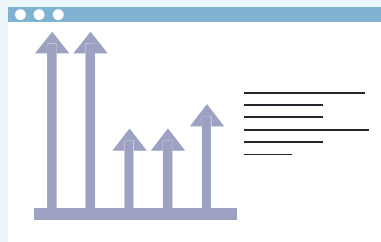






01

# Aritmetička sredina





# Aritmetička sredina

Prosečna vrednost svih podataka uključenih u uzorak. Dobija se sabiranjem vrednosti svih podataka, a zatim deljenjem broja unosa u uzorku.

$$\bar{X} = \frac{\sum x_i}{n}$$

- $X$  - konvencionalni matematički simbol srednje vrednosti  $x_i$
- $x_i$  - posmatrana vrednost,  $\sum x_i$  - zbir svih vrednosti  $x$  koje su posmatrane
- $n$  - broj posmatranja koji čine skup podataka





## Najčešće korišćeno sredstvo statistike

pr. 1: mesečnu platu 10 zaposlenih u firmi: 2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400.

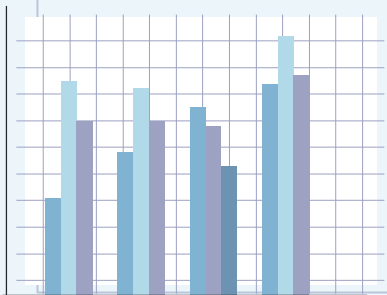
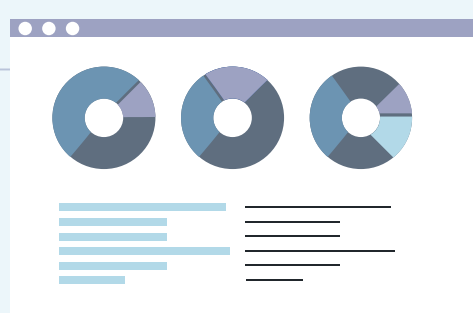
Aritmetička sredina je:

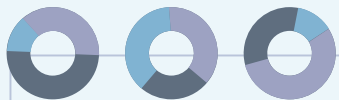
$$\frac{2500 + 2700 + 2400 + 2300 + 2550 + 2650 + 2750 + 2450 + 2600 + 2400}{10} = 2530.$$

pr. 2: 47 radnika zarađuje 1.000 € mesečno, a gazda zarađuje 1.000.000 € mesečno.

Aritmetička sredina je  $(47 \times 1.000 + 1.000.000) / 48 = 21.182,50\text{€}$  mesečno

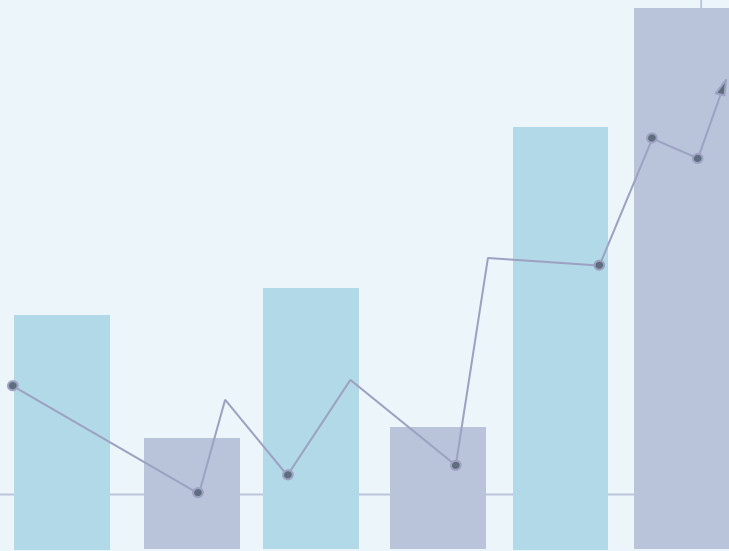
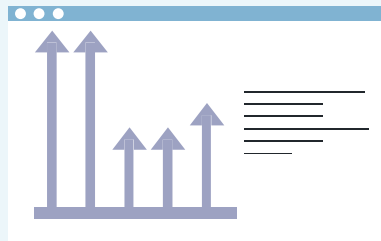
Iz nje se ne može izvući podatak o raspodeli vrednosti!





02

# Medijana



# Medijana

Vrednost koja deli gornju polovinu podataka od donje polovine podataka. Veća je od jedne polovine predstavljenih vrednosti, a manja od druge polovine podataka.

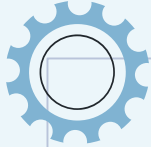
1, 3, 3, **6**, 7, 8, 9

Median = **6**

1, 2, 3, **4**, **5**, 6, 8, 9

Median =  $(4 + 5) \div 2$   
= **4.5**





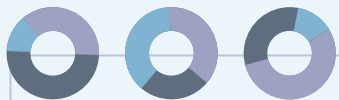
# Medijana

Ne uzima u obzir sve vrednosti i nije osetljiva na ekstremne vrednosti. Ponekad može bolje da opiše podatke od aritmetičke sredine.

pr. 1, 2, 2, 2, 3, 14.

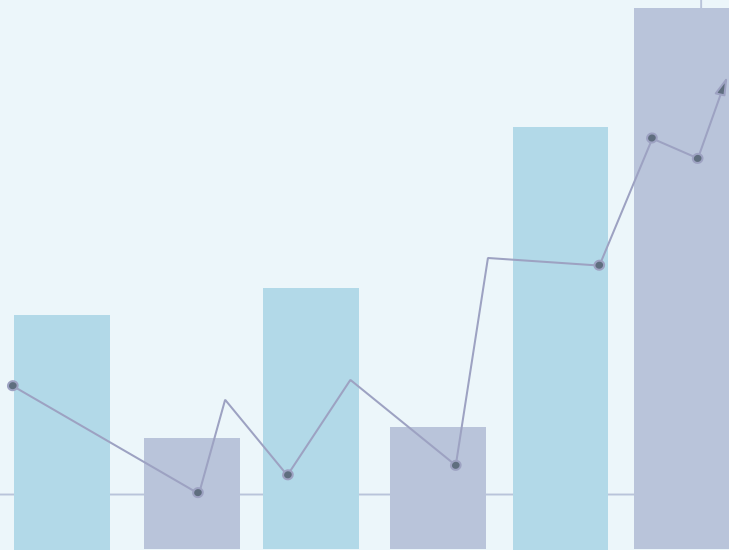
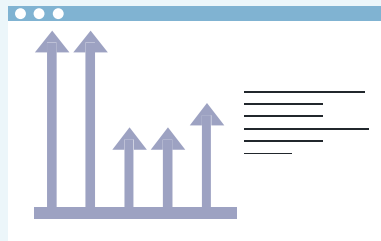
Medijana je u ovom slučaju 2 i može se posmatrati kao bolja indikacija proseka od aritmetičke sredine od 4, koja je veća od svih vrednosti osim jedne.





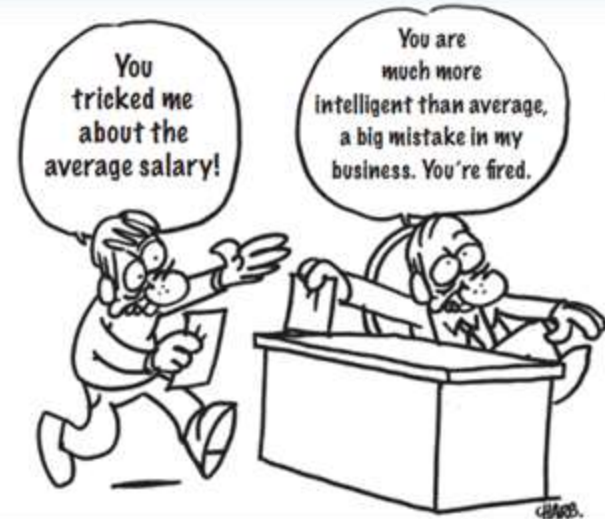
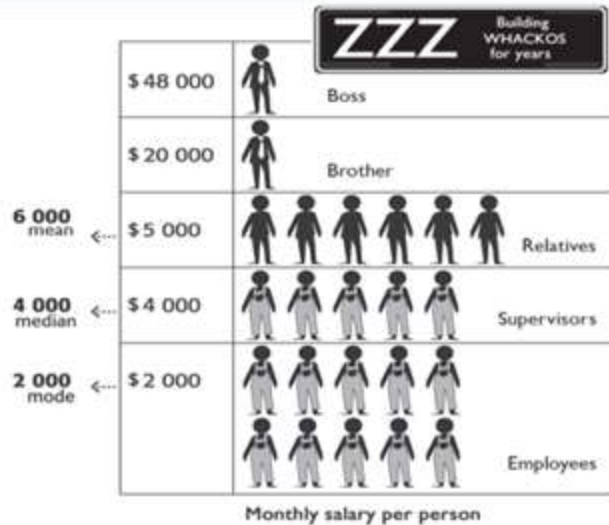
03

# Modus

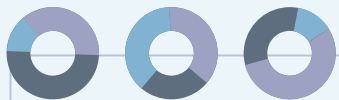


# Modus

Vrednost koja se u uzorku ili grupi podataka pojavljuje najčešće.

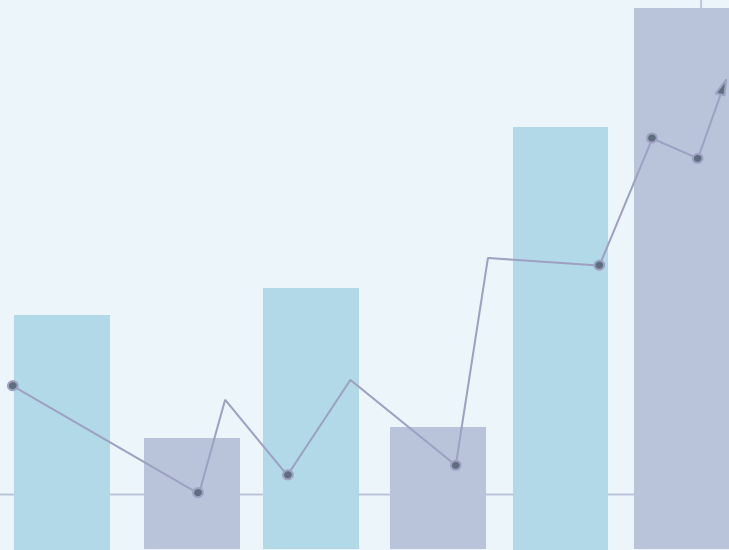
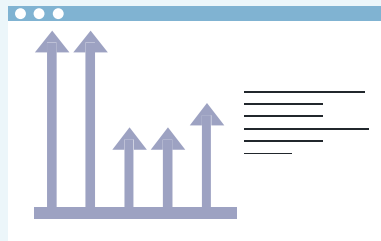






04

# Standardna devijacija





# Standardna devijacija

Standardna devijacija je mera raspršenosti podataka u odnosu na srednju vrednost.

Obeležava se sigma ( $\sigma$ ) i računa se:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Apsolutna mera disperzije  
u osnovnom skupu.

$x_i$  -  $i$ -ti član uzorka

$\mu$  - aritmetička sredina uzorka

$N$  - broj elemenata u uzorku





## Primer:

Otišli ste da pecate na jezero, ali Vam je rečeno da su se ranije u jezeru izlivali toksini iz obližnje fabrike. Rekli su Vam da je za čoveka štetno ako u ribi koju pojedete ima 7mg toksina. Takođe su Vam rekli da u proseku u ribi ima 4mg toksina.



Da li biste pojeli ribu?

Da li biste pojeli ribu ako znate da je standardna devijacija 1mg?

A u slučaju da je standardna devijacija 4mg?

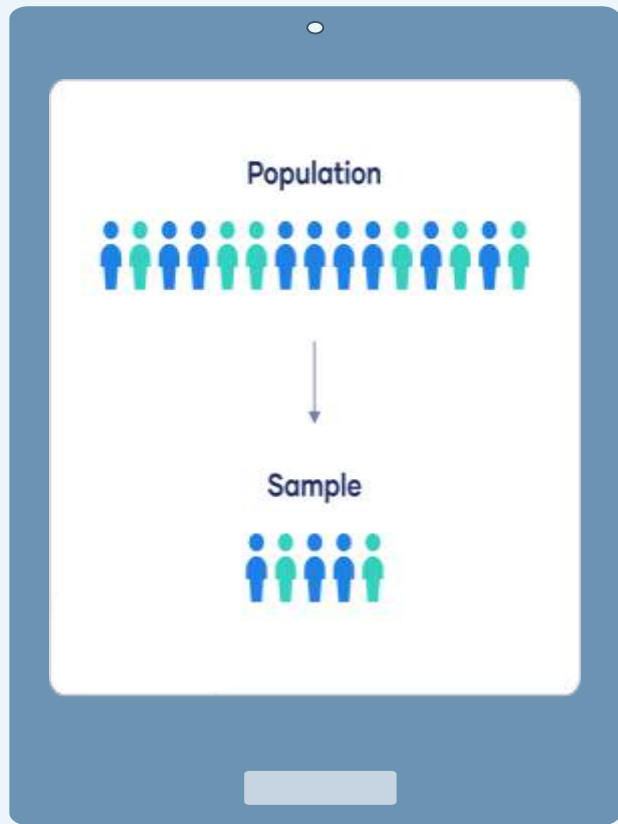


# Izračunavanje standardne devijacije

1. Odredite raspršenost svake vrednosti od srednje vrednosti koju već imate izračunatu;
2. Kvadrirajte svaku razliku i saberite ih;
3. Podelite zbir brojem vrednosti: to je varijansa;
4. Uzmite kvadratni koren ove varijanse i dobićete standardnu devijaciju.

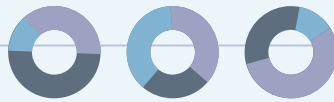
**primer:** 2, 2, 3, 5, 7, 9, 14.

Otkrićete da je varijansa 16,57, a standardna devijacija 4,07.



# Istraživanje i uzimanje uzoraka





- Statistika nam omogućava da zaključujemo o osobinama neke populacije na osnovu malog dela te populacije koji nazivamo uzorak.
- Uglavnom nije moguće ispitati celu populaciju (zbog vremena ili novca koji je za to potreban), zato uzimamo uzorak.
- Kreiranje uzoraka i pravljenje zaključaka na osnovu tih uzoraka je najrasprostranjenija i najvažnija primena statistike.



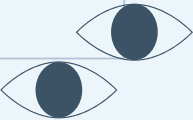


# Validnost uzoraka



Definisanje uzoraka obuhvata važan i kompleksan deo statistike. Da bi sud o populaciji bio validan, uzorak koji se analizira mora biti reprezentativan za tu populaciju (veliki i nepristrasan). Ovaj kriterijum je ključan tako da naš uzorak mora biti dovoljno veliki i nepristrasan.

- *Literary Digest* je od 1920. sprovodio istraživanje o pobedniku Američkih izbora. Iako je nekoliko godina bio uspešan, 1936. rezultat istraživanja je bio pogrešan. Časopis je svojim pretplatnicima slao anonimne upitnike i poslao bi oko 10 miliona upitnika, a dobio bi oko 2 miliona odgovora. Istovremeno je druga agencija sprovela istraživanje nad manjim skupom ljudi (4.500) i oni su imali ispravnu predikciju. Žašto?





# The Literary Digest



GRANDER LISTENING IN ON THE WORLD'S  
**New York FUNK & WAGNALLS COMPANY** *Literary*  
PUBLIC OPINION (New York combined with THE LITTEGOLD DIGEST)

Vol. 46, No. 2 Whole No. 1117

FEBRUARY 21, 1936

Price 10 Cents



HOLIDAY TIME  
(See Travel)







# Veličina uzoraka

Kako odrediti pravu veličinu uzorka?

To zavisi od niza faktora, uključujući populaciju koja se proučava, stepen preciznosti koju želimo postići, iznos novca koji imamo, pitanja na koja je naša istraga usmerena i mnoge druge.

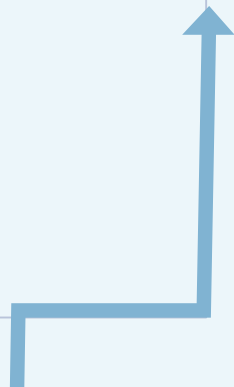
Većina istraživanja o mišljenju o nekoj temi ima između 1.000 i 2.000 ispitanika. Preciznost dobijena uzimanjem većeg skupa uglavnom nije vredna troška.

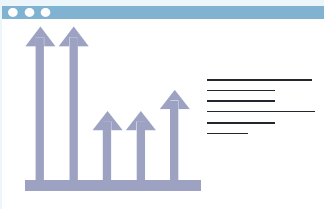


# Biranje uzoraka

Takođe, proces selekcije je ključan ako ne želite da uzorak bude pristrasan; pojedinci moraju biti odabrani nasumično kako bi bili uključeni. Međutim, ova teorija je teška za primenu, zbog čega su razvijene različite metode uzorkovanja:

- uzorkovanje po stratifikaciji,
- uzorkovanje po skupinama (klasterima),
- uzorkovanje po kvotama.







# Stratifikacija

populacija se podeli na grupe i iz svake grupe se bira nasumično

# Kastovanje


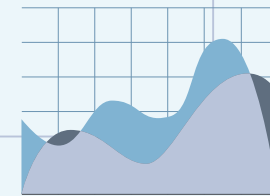
populacija se podeli na grupe i nasumično se izabere nekoliko grupa

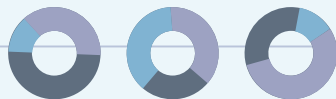




Primer 1: radio stanica sprovodi istraživanje o legalizaciji marihuane. Ukupno 3.636 slušatelja odgovara, a 78% njih kaže da su za to. Stoga stanica objavljuje da je vreme za legalizaciju marihuane i pritiska vladu da deluje.

Primer 2: sprovodimo anketu telefonom među dve hiljade ljudi koji su pozvani nasumično preko računara, iz liste svih brojeva u upotrebi. Međutim, još uvek postoji pristranost, jer najsiromašniji ljudi - koji ne poseduju svi telefone - i beskućnici nisu pravilno zastupljeni u tome.

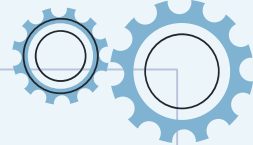




# Margina greške

U dobrim istraživanjima piše kolika je margina greške.

Margina greške je statističko izražavanje količine slučajne greške prilikom uzimanja uzorka. Obično se definiše i kao radijus intervala pouzdanosti za određenu statistiku istraživanja. To je cena koju plaćamo jer ne možemo anketirati celu populaciju .



**Primer:** U januaru su rezultati ankete za nekog predsedničkog kandidata pokazivali da ima 54% podrške. Margina greške je 5%. U junu rezultati ankete pokazuju da je podrška 56% i novinari zaključuju da se podrška povećala.

**Da li je ovaj zaključak dobar?**

Dakle, možemo tvrditi da je u januaru, 95 od 100 puta, popularnost predsednika bila između 48% i 58%, a u junu, 95 od 100 puta, bila je između 51% i 61%.





# Formulisanje pitanja u anketi

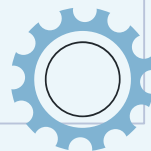
Mera greške koju ovde razmatramo zavisi od dva faktora: kako se uzorak uzima i kako su postavljena pitanja. Dobro postavljeno pitanje nije ni nejasno ni pristrasno.

Postavljeno na isti način svakoj osobi koja je ispitana, trebalo bi da bude shvaćeno na isti način od strane svakog.

Da li pijete puno alkohola?



Da li čitate *Politiku*?

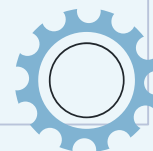
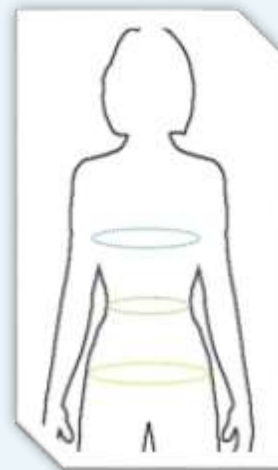




Takođe, da bi jedna anketa bila dobra, pitanja se prvo testiraju na manjem uzorku!

Međutim, kada se podaci dobro prikupe, statistika nam omogućava analizu, posebno davajući značaj traženju veza između određenih karakteristika.

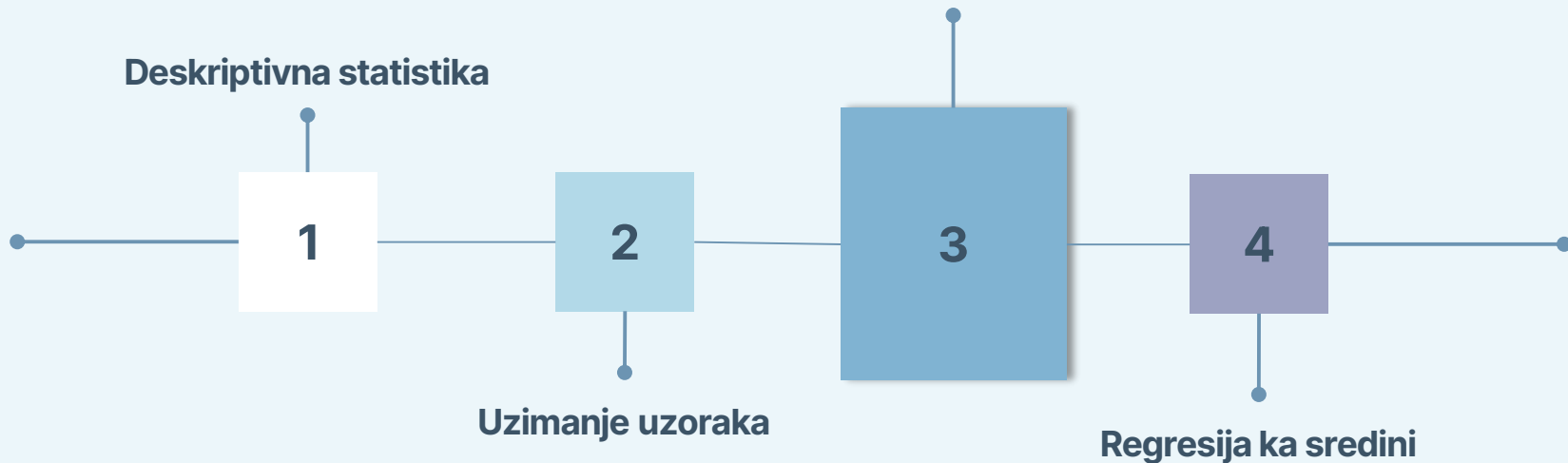
Razvijene su sofisticirane metode koje pružaju rigorozna objašnjenja u kojoj meri je jedna karakteristika povezana s drugom.







# KORELACIJA





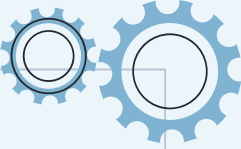
# Korelacija

Korelacija je međuodnos ili međusobna povezanost između različitih pojava predstavljenih vrednostima dve promenljive.

Uspostavljanje korelacije ne znači da ste otkrili uzročno-posledični odnos. Nerazumevanje ova dva pojma jedan je od glavnih izvora konfuzije.

Kada utvrdimo da su dve promenljive, A i B, u korelaciji, to ne znači nužno da postoji uzročni odnos između njih!

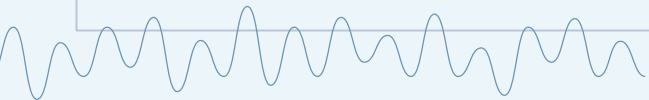


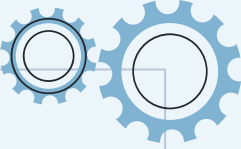


# Korelacija

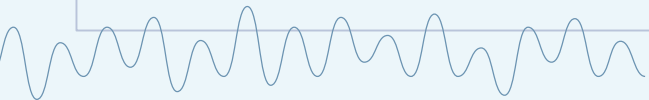
A i B su u korelaciji može da znači:

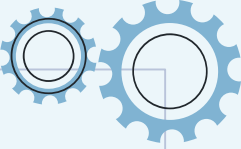
1. A je uzrok pojave B
2. B je uzrok pojave A
3. A i B su slučajno povezani bez bilo kakve uzročne veze između njih
4. A i B su oboje zavisni od nekog trećeg faktora C



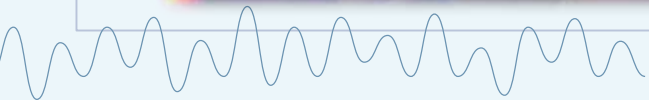


# Primeri



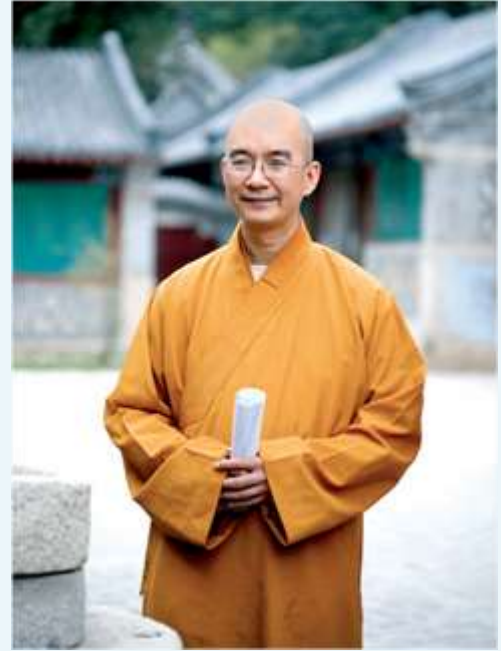
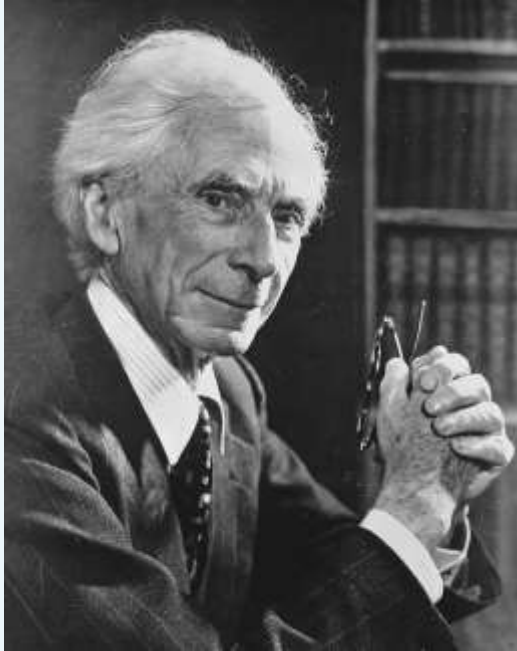


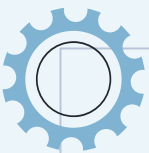
# Primeri





# Korelacija pomračenja meseca i psa





# REGRESIJA KA SREDINI





# Regresija ka sredini

Kada su dve promenljive, čije vrednosti zavise od velikog broja faktora u nepotpunoj korelaciji, ekstremne vrednosti jedne će težiti da budu u korelaciji sa manje ekstremnim vrednostima druge.

Pojam je uveo engleski naučnik Frensis Galton (statističar, sociolog, psiholog, antropolog, ...).

Galton je želeo da prouči odnos između visine očeva i njihovih sinova.







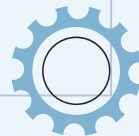
# Regresija ka sredini

Otkrio je da visoki očevi imaju visoke sinove, a niski očevi imaju niske sinove.

Veoma visoki očevi imaju sinove koji su niži od njih.

Veoma niski očevi imaju sinove koji su viši od njih.

Ovo je upravo slučaj nesavršene korelacije između dve promenljive - visine očeva i njihovih sinova. Veliki broj faktora igra ulogu u određivanju visine osobe: visina njegove majke i oca, način odrastanja, ishrana, mnogi geni...





# Regresija ka sredini

Veliki broj ovih faktora mora da se poklopi da bi osoba bila posebno visoka (ili niska). To objašnjava zašto, kada se to dogodi, obično je povezana sa manje izuzetnim događajem: manje ekstremnim vrednostima u našem slučaju, koji su visoki sinovi koji nisu toliko visoki kao njihovi očevi.

To je predvidljivo i poznato kao regresija ka srednjoj vrednosti, međutim, nekada može biti prikazana kao sujeverje.





# Rezime

Statistika je funkcija uzoraka  $x_1, x_2, \dots, x_n$  čiji analitički izraz ne zavisi od nepoznatih parametara raspodele obeležja populacije iz koje je uzorak uzet.

Podela statistike na inferencijalnu i deskriptivnu.

Aritmetička sredina je tehnika koju najčešće koristimo, ali ne daje nam podatke o raspodeli vrednosti.

Standardna devijacija je mera raspršenosti podataka u odnosu na srednju vrednost.





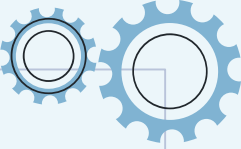
# Rezime

Proces selekcije je ključan ako ne želite da uzorak bude pristrasan. Pojedinci moraju biti odabrani nasumično kako bi bili uključeni. Razvijene različite metode uzorkovanja, od kojih smo mi pomenuli stratifikaciju i kastovanje.

Korelacija je međusobna povezanost između različitih pojava predstavljenih vrednostima dve promenljive, a regresija ka sredini se javlja kada kod dve vrednosti koje nisu u savršenoj korelaciji, a ekstremne vrednosti jedne promenljive su u korelaciji sa neekstremnim vrednostima druge promenljive.

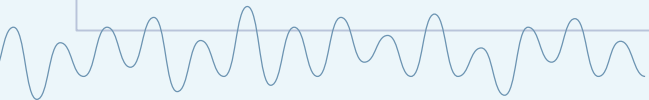
PITANJA?

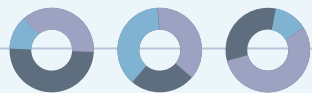




# Literatura

- Normand Baillargeon, *A Short Course in Intellectual Self-Defense*, UQAM, Seven Stories, 2008.
- John Allen Paulos, *A mathematician reads the newspaper*, New York: Anchor Books, 1995
- <https://sh.wikipedia.org/wiki/Statistika>
- <https://www.stat.gov.rs/>
- [Zavod za statistiku Srbije](#)





**HVALA  
VAM NA  
PAŽNJI!**

