

# STATISTIKA U SOCIOLOGIJI

Filozofski fakultet, UCG

Lazar Obradović<sup>1</sup>

<sup>1</sup>Prirodno-matematički fakultet,  
Univerzitet Crne Gore

Zimski semestar 2020/21

# Literatura

Statistika (Dr Beba Mutavdžić, Mr Emilija Nikolić-Đorić)

Osnovne statističke metode za nematematičare (Boris Petz)

# Osnovni pojmovi

## Jedinica posmatranja

Određeni subjekat ili objekat o kojem se prikupljaju podaci (informacije).

## Statistički skup/populacija

Skup jedinica (subjekata, objekata) na osnovu kojih se ispituje jedno ili više obilježja.

## Obilježje/promjenljiva

*Promjenljiva* (obilježje, varijabla) je karakteristika koja se izučava i koja uzima različite vrijednosti među jedinicama posmatranja.

## Opservacija/podatak

Vrijednost promenljive koja se odnosi na jednu jedinicu posmatranja.

# Parametar skupa i statistika uzorka

## Statistički uzorak

Dio statističkog skupa na osnovu kojeg donosimo statističke zaključke o populaciji.

## Reprezentativni uzorak

Uzorak je reprezentativan ako svojim osobinama "dobro" opisuje osobine populacije

## Parametar skupa/populacije

Parametar skupa je neka sumarna numerička karakteristika tog skupa.

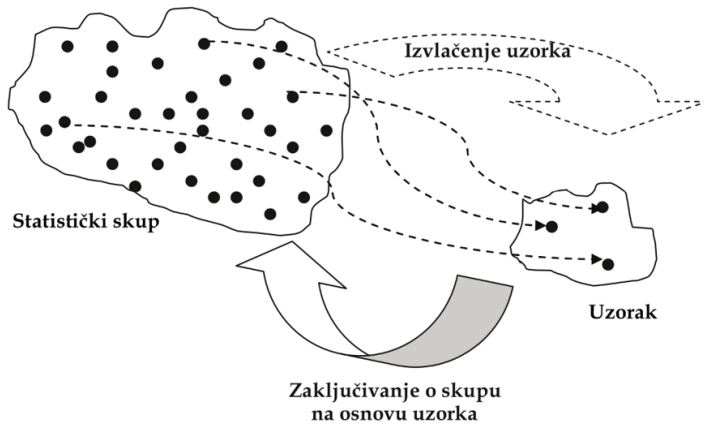
(Npr. prosjek ili maksimum)

## Statistika uzorka

*Statistika uzorka* je neka sumarna karakteristika reprezentativnog uzorka.

## Cilj:

Na osnovu statistike uzorka donijeti zaključke o parametru skupa.



# Promjenljive i podaci

## Obilježje/promjenljiva

*Promjenljiva* (obilježje, varijabla) je karakteristika koja se izučava i koja uzima različite vrijednosti među jedinicama posmatranja.

- ▶ *Kvalitativna/kategorička/atributivna promjenljiva*  
Promjenljiva čija moguće vrijednosti nisu brojevi: pol, marka automobila, državljanstvo...
- ▶ *Kvantitativna/numerička promjenljiva*  
Promjenljiva koja uzima brojne vrijednosti.
  - ▶ *Diskretna promjenljiva* Promjenljiva koja uzima brojne vrijednosti koje možemo staviti na spisak: broj zaposlenih u firmi, broj djece u porodici...
  - ▶ *Neprekidna promjenljiva*  
Promjenljiva koja može uzeti bilo koju vrijednost sa brojne prave i rezultat je merenja: nivo alkohola u krvi, visina djece u školi,...

# Promjenljive i podaci

## Opservacija/podatak

Vrijednost promenljive koja se odnosi na jednu jedinicu posmatranja.

## Negrupisani podaci

Podaci zapisani redosljedom kojim se prikupljaju pre nego što se urede po veličini ili grupišu.

## Statističke serije

Uređivanjem statističkih podataka nastaju statistički nizovi, odnosno statističke serije:

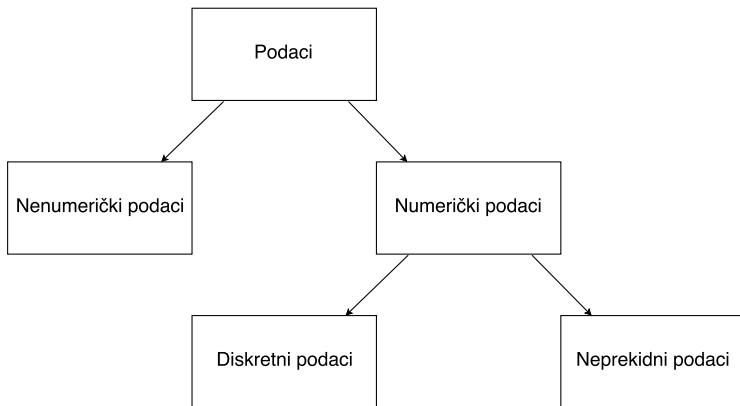
- ▶ Serije strukture: (eng. cross section data) pokazuju raspored (strukturu) statističkog skupa ili uzorka prema vrijednostima obilježja
- ▶ Vremenske serije: statističke serije koje su uređene po hronologiji.

# Promjenljive i podaci

- ▶ *Kvalitativni/kategorički podaci*  
Podaci dobijeni od kvalitativne promjenljive.
- ▶ *Kvantitativni/numerički podaci*  
Podaci dobijeni od kvantitativne promjenljive.
- ▶ *Diskretna promjenljiva*  
Podaci dobijeni od diskretne promjenljive.
- ▶ *Neprekidna promjenljiva*  
Podaci dobijeni od neprekidne promjenljive.



# Promjenljive i podaci



# Grupisanje podataka

Radi bolje preglednosti, pogotovo ako je broj prikupljenih podataka veliki, podaci se grupišu u klase ili grupe i određuje se broj podataka (frekvencija) u svakoj klasi odnosno grupi.

## (Apsolutna) frekvencija

Frekvencija, ili apsolutna frekvencija neke vrijednosti je broj pojavljivanja te vrijednosti unutar posmatrane statističke serije. Ako imamo  $k$  različitih mogućih vrijednosti, frekvencija  $i$ -te vrijednosti se označava sa  $f_i$ ,  $i = 1, \dots, k$ .

# Grupisanje podataka

## Relativna frekvencija

Relativna frekvencija je količnik apsolutne frekvencije i ukupnog broja jedinica posmatranja:  $p_i = \frac{f_i}{N}$ .

Relativne frekvencije se mogu izraziti i u procentima:  $p_i \cdot 100$

## Kumulativna frekvencija

Kumulativna frekvencija  $F_i$  određene vrednosti obilježja dobija se sabiranjem apsolutnih frekvencija svih prethodnih vrijednosti i apsolutne frekvencije te vrijednosti:

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j$$

# Grupisanje podataka

## Distribucija frekvencija (Raspored frekvencija)

Tabela sa dva niza (dvije kolone) podataka:

- 1: Vrijednosti promjenljive (prikazane po pojedinačnim vrijednostima ili grupnim intervalima)
- 2: Frekvencija odgovarajuće vrijednosti

Često distribucija frekvencija ima i dodatne kolone s relativnim frekvencijama i kumulativnim frekvencijama.

# Distribucija frekvencija, primjer

Grupisanje po jednoj promjenljivoj, neintervalna tabela frekvencija

## Primjer

Broj televizora za 50 nasumično izabranih domaćinstava dat je u tablici.

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

# Distribucija frekvencija, primjer

Grupisanje po jednoj promjenljivoj, neintervalna tabela frekvencija

Tabela distribucije frekvencija i relativnih frekvencija

Tabela distribucije frekvencija i relativnih frekvencija:

<b>Number of TVs</b>	<b>Frequency</b>	<b>Relative frequency</b>
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	50	1.00

# Grupisanje podataka

- ▶ Kada promjenljiva ima veliki broj različitih vrednosti one se grupišu u unaprijed određene intervale.
- ▶ Broj i veličina (širina) intervala zavise od broja podataka ( $N$ ) i od prirode samog obeležja.
- ▶ Kada se formiraju intervali, frekvencije prebrojavaju broj pojavljivanja neke vrijednosti u okviru određenog intervala.

# Distribucija frekvencija, primjer

## Grupisanje po intervalima

### Primjer

Dati su podaci o cijenama 37 luksuznih automobila s Frankfurtskog sajma automobila (u hiljadama EUR):

129.2 185.3 218.1 182.5 142.8 155.2 170.0 151.3 187.5 145.6 167.3  
161.0 178.7 165.0 172.5 191.1 150.7 187.0 173.7 178.2 161.7 170.1  
165.8 214.6 136.7 278.8 175.6 188.7 132.1 158.5 146.4 209.1 175.4  
182.0 173.6 149.9 158.6

Grupisati podatke u 8 grupa širine 20 000 EUR, tako da donja granica prve grupe bude 120 000, a zatim napraviti tabelu distribucije frekvencija i relativnih frekvencija po grupama.



# Distribucija frekvencija, primjer

## Grupisanje po intervalima

*Rješenje:*

Cijena (u hiljadama EUR)	Frekvencija	Relativna frekvencija
120 - do 140	3	0.081
140 - do 160	9	0.243
160 - do 180	14	0.378
180 - do 200	7	0.189
200 - do 220	3	0.081
220 - do 240	0	0.000
240 - do 260	0	0.000
260 - do 280	1	0.027

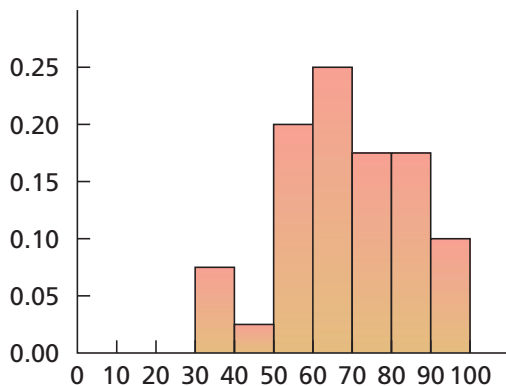
# Grafičko predstavljanje numeričkih podataka

- ▶ Nakon grupisanja možemo da pravimo tabele distribucija.
- ▶ Histogram - "slijepljeni štapićasti dijagram".
- ▶ Tačkasti dijagram.
- ▶ Dijagram stablo-list (stem and leaf).

# Grafičko predstavljanje numeričkih podataka

## Histogram

Histogram - "slijepljeni štapićasti dijagram".



- ▶  $O_x$ -osa → grupe.
- ▶  $O_y$ -osa → (relativne) frekvencije.

# Grafičko predstavljanje numeričkih podataka

## Histogram

### Grupisanje po jednoj vrijednosti

- ▶ Svakom štapiću odgovara jedna vrijednost.
- ▶ Vrijednost koja odgovara štapiću se nalazi u središtu osnovice štapića.

### Grupisanje po intervalima

- ▶ Svakom štapiću odgovara jedan interval.
- ▶ Osnovica štapića je odgovarajući interval.

# Grafičko predstavljanje numeričkih podataka

## Histogram

### Procedura: crtanje histograma

- ▶ Izabrati grupe i napraviti tablicu (relativnih) distribucija.
- ▶ Nacrtati koordinatni sistem, označiti ose.
- ▶ Označiti grupe na  $Ox$  osi, označiti vrijednosti na  $Oy$  osi.
- ▶ Iznad svake grupe povući vertikalnu duž širine grupe na visini koja odgovara (relativnoj) frekvenciji. Nacrtati štapić.

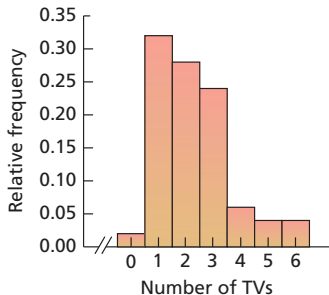
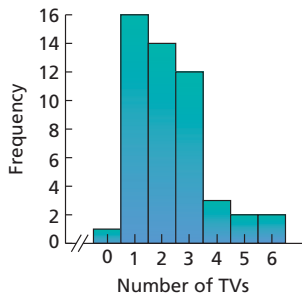
# Grafičko predstavljanje numeričkih podataka

Primjer - Grupisanje po jednoj vrijednosti

Number of TVs	Frequency	Relative frequency
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	50	1.00

Data je tabela distribucija (tv primjer):

Nacrtati histogram frekvencija i relativnih frekvencija.



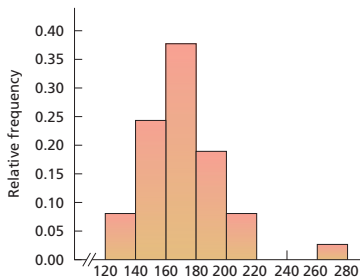
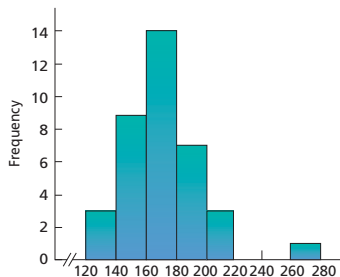
# Grafičko predstavljanje numeričkih podataka

## Histogram primjer - Grupisanje po intervalima.

Data je tabela distribucija (luks.auto):

Cijena (u hiljadama EUR)	Frekvencija	Relativna frekvencija
120 - do 140	3	0.081
140 - do 160	9	0.243
160 - do 180	14	0.378
180 - do 200	7	0.189
200 - do 220	3	0.081
220 - do 240	0	0.000
240 - do 260	0	0.000
260 - do 280	1	0.027

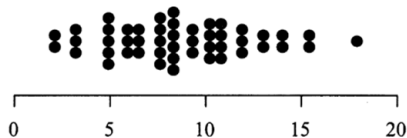
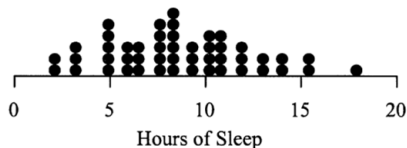
Nacrtati histogram frekvencija i relativnih frekvencija.



# Grafičko predstavljanje numeričkih podataka

## Tačkasti dijagram

- ▶ Horizontalna ose sa tačkicama koje odgovaraju vrijednostima.
- ▶ Korisni za prikazivanje međusobnog odnosa podataka.
- ▶ Korisni za upoređivanje 2 skupa podataka.





# Grafičko predstavljanje numeričkih podataka

## Tačkasti dijagram

### Procedura: crtanje tačkastog dijagrama

1. Horizontalna osa sa mogućim vrijednostima.
2. Svaka opservacija - tačka iznad ose.
3. Ako se vrijednosti ponavljaju "slagati tačkice".

# Grafičko predstavljanje numeričkih podataka

## Tačkasti dijagram - primjer

### Primjer

Istraživanjem tržišta DVD plejera iz 2011. godine došlo se do uzorka od 16 uređaja čije su cijene (u američkim dolarima):

210 219 214 197 224 219 199 199

208 209 215 199 212 212 219 210

Nacrtati tačkasti dijagram cijena plejera iz datog uzorka:

*Rješenje:*



# Grafičko predstavljanje numeričkih podataka

## Stem and leaf

- ▶ Dijagram *stablo-list*
- ▶ Koristi se za diskretne cjelobrojne podatke.
- ▶ Svaka opservaciju interpretiramo kao stablo i list.
- ▶ Stablo - sve cifre osim posljednje.
- ▶ List - posljednja cifra.
- ▶ Dijagram: Kolona sa 'stablima' u rastućem poretku i kolona sa "lišćem" u rastućem poretku pored svakog stabla.

Stems	Leaves
3	6 8 9
4	7
5	0 1 1 3 5 5 6 7
6	0 2 3 4 4 5 6 7 8 9
7	0 0 0 1 5 8 9
8	0 1 3 5 6 7 9
9	5 8 9 9

# Grafičko predstavljanje numeričkih podataka

## Stem and leaf

### Procedura: crtanje dijagrama stablo-list

1. Poređati podatke u rastući niz.
2. Svaku opservaciju podijeliti na stablo i list.
3. Napraviti kolonu stabala, od najvećeg ka najmanjem.
4. Za svako stablo, dopisati listove (u rastućem poretku).

Komentar: Nekad se svi "redovi listova" podijele na dva reda po "stablu". Tada svakom stablu odgovaraju dva reda: gornji, sa ciframa čije su vrijednosti 0, 1, 2, 3, 4, i donji, sa ciframa, 5, 6, 7, 8, 9. (Vidi primjer.)

# Grafičko predstavljanje numeričkih podataka

## Stem and leaf - primjer

### Primjer

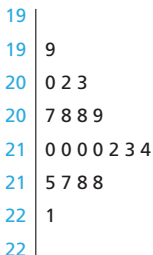
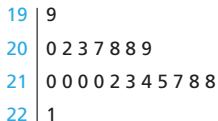
Istraživanjem uzoraka krvi za 20 djece (uzrasta od 4 do 19 godina) sa povišenim nivoom holesterola dobijene su sljedeće vrijednosti holesterola:

210 209 212 208 217 207 210 203 208 210

210 199 215 221 213 218 202 218 200 214.

Predstaviti date podatke koristeći dijagram stablo-grana na dva načina: (a) Sa jednim redom po stablu. (b) Sa dva reda po stablu.

*Rješenje:*



# Distribucija

## Distribucija skupa podataka

- ▶ Tabela, grafik, formula.
- ▶ Opisuje vrijednosti opservacija i njihove frekvencije.

## Distribucija $\leftrightarrow$ uzorak

- ▶ Skup podataka o populaciji  $\rightarrow$  distribucija populacije.
- ▶ Skup podataka o uzorku  $\rightarrow$  distribucija uzorka.
- ▶ Distribucija uzorka zavisi od izbora uzorka.
- ▶ Distribucija populacije  $\neq$  distribucija uzorka.
- ▶ Distribucija populacije  $\approx$  distribucija uzorka.

# Deskriptivne mjere

Broj koji izračunamo iz podataka je *deskriptivna mjera*.

## Primjeri:

- ▶ Prosjek; mjere centralne tendencije.
- ▶ Max/Min, raspon.
- ▶ Standardna devijacija; mjere raspršenja.

## Uzorak ↔ Populacija

- ▶ Deskriptivna mjera populacije  $\neq$  deskriptivna mjera uzorka.
- ▶ Kada je:  
Deskriptivna mjera uzorka  $\approx$  Deskriptivna mjera populacije?

# Mjere centralne tendencije

Mjere centralne tendencije daju informaciju o onome što intuitivno nazivamo "prosjek".

## Tri vrste mjera centralne tendencije

- ▶ Aritmetička sredina,  $\bar{x}$ : ("pravi") prosjek.
- ▶ Medijan,  $Q_2$ : srednja vrijednost.
- ▶ Mod: najčešća vrijednost.



# Mjere centralne tendencije

## Aritmetička sredina

*Aritmetička sredina* je prosjek opservacija: zbir svih opservacija podjeljen sa brojem opservacija.

- ▶ Definisana samo za numeričke podatke.
- ▶ Prosjek, mean.
- ▶ Oznaka:  $\bar{x}$ .
- ▶ Ako imamo  $n$  opservacija,  $x_1, x_2, \dots, x_n$ , onda:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

## Prelaki primjer

Ako imamo 4 opservacije  $x_1 = 0.4$ ,  $x_2 = 2.8$ ,  $x_3 = -1.2$  i  $x_4 = 0.2$  onda je:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{0.4 + 2.8 - 1.2 + 0.2}{4} = \frac{2.2}{4} = 0.55.$$

# Mjere centralne tendencije

## Medijan

Ako su podaci poređani po veličini, *medijan* je opservacija koji se nalazi tačno "po sredini" podataka.

- ▶ Broj koji "razdvaja gornju i donju polovinu podataka".
- ▶ Definisano samo za numeričke podatke.
- ▶ Oznaka:  $Q_2$ .

Procedura: računanje medijana od  $n$  podataka,  $x_1, x_2, \dots, x_n$

1. Posložiti podatke u spisak po veličini:  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ .
2. Neparno  $n$ : medijan je vrijednost koja je tačno na sredini spiska.
  - ▶  $n = 3, Q_2 = x_{(2)}$ ;  $n = 19, Q_2 = x_{(10)}$ ;  $n = 21, Q_2 = x_{(11)}$ .
3. Parno  $n$ : medijan je aritmetička vrijednost dva podatka koja su tačno na sredini spiska.
  - ▶  $n = 4, Q_2 = \frac{x_{(2)} + x_{(3)}}{2}$ ;  $n = 20, Q_2 = \frac{x_{(10)} + x_{(11)}}{2}$ .

# Mjere centralne tendencije

## Medijan

### Prelaki primjer 1

Ako imamo 4 opservacije  $x_1 = 0.4$ ,  $x_2 = 2.8$ ,  $x_3 = -1.2$  i  $x_4 = 0.2$   
onda je "spisak":  $-1.2, 0.2, 0.4, 2.8$ , pa je  $Q_2 = \frac{0.2 + 0.4}{2} = 0.3$

### Prelaki primjer 2

Ako imamo 7 opservacija:

$x_1 = 4$ ,  $x_2 = 2$ ,  $x_3 = 2$ ,  $x_4 = 0$ ,  $x_5 = 3$ ,  $x_6 = 1$ ,  $x_7 = 0$

onda je "spisak":  $0, 0, 1, 2, 2, 3, 4$ , pa je  $Q_2 = 2$ .

# Mjere centralne tendencije

## Mod

*Mod* je najčešća vrijednost koja se pojavljuje u skupu podataka: vrijednost s najvećom frekvencijom.

- ▶ Definirano i za numeričke i nenumeričke podatke.
- ▶ Ako se ni jedna vrijednost ne pojavljuje više od jednom onda mod ne postoji.
- ▶ Ako više vrijednosti imaju jednaku, najveću frekvenciju onda je mod bilo koja od tih vrijednosti.
- ▶ Najlakše se određuje kada je na raspolaganju distribucija frekvencija.

## Prelaki primjer (bez tabele)

Ako imamo 7 opservacija:

$$x_1 = 4, x_2 = 2, x_3 = 2, x_4 = 0, x_5 = 3, x_6 = 1, x_7 = 0$$

onda je "spisak": 0, 0, 1, 2, 2, 3, 4, pa je *mod* = 2.

*Komentar1*: Trebalo je napraviti distribuciju (tabelu) frekvencija.

*Komentar2*: Tačko bi bilo i *mod* = 0.

# Mjere centralne tendencije

## Primjer

### Primjer

U prvoj polovini godine, revizorska kuća je angažovala 13 eksternih konsultanata čiji su honorari bili (u EUR):

300 300 300 940 300 300 400 300 400 450 800 450 1050.

Ista revizorska kuća je u drugoj polovini godine angažovala 10 konsultanata, i njihovi honorari su bili:

300 300 940 450 400 400 300 300 1050 300.

1. Napraviti tačkaste dijagrame za oba skupa podataka.
2. Izračunati aritmetičku sredinu oba skupa podataka.
3. Odrediti medijan oba skupa podataka.
4. Odrediti mod oba skupa podataka.

*Rješenje:* Na tabli!

Računska pomoć:

$$6290 = 300 + 300 + 300 + 940 + 300 + 300 + 400 + 300 + 400 + 450 + 800 + 450 + 1050$$

$$4740 = 300 + 300 + 940 + 450 + 400 + 400 + 300 + 300 + 1050 + 300$$

# Mjere centralne tendencije

## Izbor mjere centralne tendencije

Kad koristiti koju mjeru centralne tendencije?

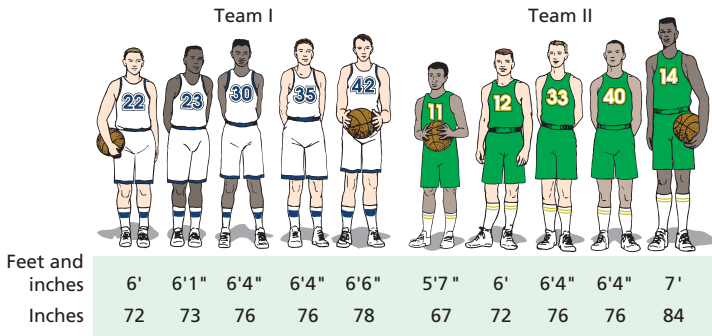
- ▶ Generalno:  $\bar{x} \neq Q_2 \neq \text{mod}$ .
- ▶ Medijan je otporan na ekstremume:
  1. 0, 1, 2, 4, 7 :  $Q_2 = 2, \bar{x} = \frac{0+1+2+4+7}{5} = 2.8$
  2. 0, 1, 2, 4, 107 :  $Q_2 = 2, \bar{x} = \frac{0+1+2+4+107}{5} = 22.8$
- ▶ Kod nenumeričkih podataka možemo koristiti samo *mod*.

# Mjere varijabilnosti

## Uvod

- ▶ Mjere centralne tendencije nisu dovoljne. Primjer?
- ▶ Mjere raspršenja.
- ▶ Mjere varijabilnosti daju (djelimičan) odgovor na dva pitanja:
  - ▶ Koliko su opservacije blizu jedna drugoj?
  - ▶ Koliko su opservacije blizu prosjeku?

Isto pitanje?

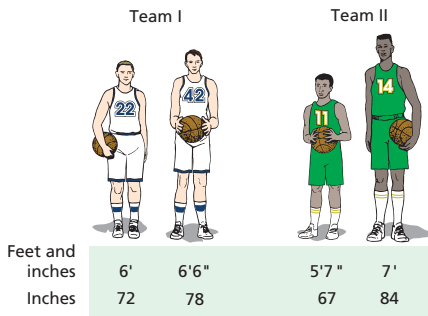


# Mjere varijabilnosti

## Raspon

*Raspon* je razlika između najveće i najmanje opservacije.

- ▶ Najmanja opservacija:  $\min, x_{(1)}$
- ▶ Najveća opservacija :  $\max, x_{(n)}$ .
- ▶ Raspon:  $x_{(n)} - x_{(1)}$



Team I:  $78 - 72 = 6'$

Team II:  $84 - 67 = 17'$



# Sigma notacija

## Tehnička priprema za standardnu devijaciju

- ▶ Matematička oznaka: "kutija za zbir".

- ▶ Definicija: 
$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

- ▶ Primjer: 
$$\sum_{k=3}^7 a_k = a_3 + a_4 + a_5 + a_6 + a_7$$

- ▶ Primjer: 
$$\sum_{j=1}^n x_j^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

- ▶ Primjer: 
$$\sum_{i=2}^5 (2i - 1) = 3 + 5 + 7 + 9$$

- ▶ Primjer: 
$$\sum_{i=1}^{2n} 3 = 3 + 3 + \dots + 3 = 2n \cdot 3 \text{ (2n sabiraka!)}$$

# Mjere varijabilnosti

## Standardna devijacija - uvod

- ▶ Devijacija = odstupanje.
- ▶ Ideja: Izmjeriti "prosječno odstupanje od prosjeka".
- ▶ Tražimo deskriptivnu mjeru koja raste s raspršenjem.

# Mjere varijabilnosti

## Standardna devijacija - uvod, nastavak

Team I (Uzorak): 72, 73, 76, 76, 78,  $\bar{x} = \frac{72+73+76+76+78}{5} = 75$ .

▶ Devijacija od prosjeka:  $x_i - \bar{x}$ . [Team I: -3, -2, 1, 1, 3]

▶ Prosjek devijacija? [Team I:  $\frac{-3-2+1+1+3}{5} = 0!$ ]

Ne valja: veće raspršenje  $\nrightarrow$  veći prosjek devijacija.

▶ Kvadratna devijacija:  $(x_i - \bar{x})^2$ . [Team I: 9, 4, 1, 1, 9]

▶ *Disperzija* - "prosjek" kvadratnih devijacija:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}. \quad [\text{Team I: } s^2 = \frac{9+4+1+1+9}{4} = \frac{24}{4} = 6]$$

▶ Zašto  $n - 1$ , a ne  $n$ ? Komplikovani tehnički razlozi.

▶ *Standardna devijacija* - korijen disperzije:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}. \quad [\text{Team I: } s = \sqrt{6}]$$

▶ Veće raspršenje  $\Rightarrow$  veća disperzija, veća standardna devijacija.

# Mjere varijabilnosti

## Standardna devijacija

Definicija: Disperzija

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Definicija: Standardna devijacija

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Standardna devijacija: alternativna formula

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i)^2 - n(\bar{x})^2}{n - 1}}$$

# Mjere varijabilnosti

## Standardna devijacija

### Primjer

Izračunati standardnu devijaciju visine u prvom timu.

*Rješenje:* Tabela. Popunjavamo po kolonama.

$i$	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	72	-3	9
2	73	-2	4
3	76	1	1
4	76	1	1
5	78	3	9
$\Sigma$	375		24
	$\bar{x} = \frac{375}{5} = 75$		$s^2 = \frac{24}{4} = 6$ $s = \sqrt{6}$

# Mjere varijabilnosti

## Kvartili

Ideja: Podijeliti podatke na četiri dijela (kao što medijan dijeli podatke na "gornju i donju polovinu")

- ▶ "Donja polovina podataka":  
Opservacije do medijana (uključujući i medijan - neparni slučaj)
- ▶ "Gornja polovina podataka":  
Opservacije od medijana (uključujući i medijan - neparni slučaj)
- ▶ *Prvi kvartil*: medijan donje polovine podataka. Oznaka:  $Q_1$ .
- ▶ *Drugi kvartil*: medijan svih podataka. Oznaka:  $Q_2$ .
- ▶ *Treći kvartil*: medijan gornje polovine podataka. Oznaka:  $Q_3$
- ▶ Interkvartilni raspon:  $IQR = Q_3 - Q_1$ .

# Mjere varijabilnosti

## Kvartili - primjeri

### Primjer 1

Podaci: 3, 5, 5, 6, 9, 9, 11, 15, 16

Rješenje:  $Q_1 = 5$ ,  $Q_2 = 9$ ,  $Q_3 = 11$ ,  $IQR = Q_3 - Q_1 = 11 - 5 = 6$ .

### Primjer 2

Podaci: 2, 5, 7, 9, 11, 12

Rješenje:  $Q_1 = 5$ ,  $Q_2 = \frac{7+9}{2} = 8$ ,  $Q_3 = 11$ ,  
 $IQR = Q_3 - Q_1 = 11 - 5 = 6$ .

### Primjer 3

Podaci: 0, 0, 0, 0, 4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 13, 14

Rješenje:  $Q_1 = \frac{0+4}{2} = 2$ ,  $Q_2 = \frac{7+7}{2} = 7$ ,  $Q_3 = \frac{9+10}{2} = 9.5$ ,  
 $IQR = Q_3 - Q_1 = 9.5 - 2 = 7.5$ .

# Mjere varijabilnosti

## Box plot

- ▶ Kutijasti dijagram.
- ▶ Pet podataka: min,  $Q_1$ ,  $Q_2$ ,  $Q_3$ , max.

## Procedura

1. Poređati podatke po veličini.
2. Izračunati pet podataka.
3. Nacrtati korizontalnu osu, označiti pet podataka dužima iznad ose.
4. Nacrtati "kutiju" oko  $Q_1$ ,  $Q_2$  i  $Q_3$ .
5. Spojiti min i max sa kutijom.

Postoje varijacije dijagrama sa drugim izborima pet vrijednosti.



# Mjere varijabilnosti

## Box plot

### Primjer

Nacrtati kutijasti dijagram (box plot) za sljedeći skup podataka:  
4.3, 5.1, 3.9, 4.5, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6, 4.4, 4.3, 4.8, 4.4, 4.2.

*Rješenje:*

Podaci poređani po veličini:

3.9, 4.1, 4.2, 4.3, 4.3, 4.4, 4.4, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1

▶ 15 opservacija - medijan je 8. opservacija:  $Q_2 = 4.4$ .

▶ Donja polovina podataka:

3.9, 4.1, 4.2, 4.3, 4.3, 4.4, 4.4, 4.4

▶  $Q_1 = (4.3 + 4.3)/2 = 4.3$

▶ Gornja polovina podataka:

4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1

▶  $Q_3 = (4.7 + 4.8)/2 = 4.75$

▶  $\min = 3.9$ ,  $\max = 5.1$ .

# Mjere varijabilnosti

## Box plot

$\min = 3.9$ ,  $Q_1 = 4.3$ ,  $Q_2 = 4.4$ ,  $Q_3 = 4.75$ ,  $\max = 5.1$ .

